

Московский физико-технический институт

Учебное пособие

**«Математическое и компьютерное
моделирование наносистем»**

Вьюрков В.В., Орликовский А.А., Семенихин И.А.,
Негров Д.В. Озерин А.Ю. Свинцов Д.А.

Москва — 2011

УДК 001.57

ISBN 978-5-8493-0212-6

Основной целью учебного пособия «Математическое и компьютерное моделирование наносистем» является ознакомление читателя с методами решения уравнений квантовой механики в сложных системах. Главы 1-5 пособия посвящены приближенным методам решения уравнения Шредингера и определения энергетических спектров систем многих частиц. Главы 6-8 и приложения посвящены применению полученных результатов к моделированию приборов нанoeлектроники и расчету свойств нанообъектов с помощью программного пакета Gaussian.

Для студентов старших курсов физических специальностей и аспирантов.

Авторы:

Негров Д.В. (гл. 1-4, приложение Б)

Озерин А.Ю. (гл. 5)

Вьюрков В.В., Орликовский А.А., Семенихин И.А. (гл. 6-7)

Свинцов Д.А. (гл. 5, 8 приложение А)

Данное пособие издано на средства государственного контракта №16.647.12.2016 от 25 ноября 2010 г. в рамках работ по направлению 2 Федеральной целевой программы «Развитие инфраструктуры нанопромышленности в Российской Федерации на 2008—2011 годы»

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	5
1. УРАВНЕНИЕ ШРЕДИНГЕРА ДЛЯ СИСТЕМ МНОГИХ ЧАСТИЦ И ПРИБЛИЖЕНИЕ БОРНА—ОППЕНГЕЙМЕРА	6
2. МЕТОД ХАРТРИ—ФОКА	12
3. МЕТОД ФУНКЦИОНАЛА ЭЛЕКТРОННОЙ ПЛОТНОСТИ	21
4. СТРУКТУРА МЕТОДОВ КВАНТОВОЙ ХИМИИ	34
5. МЕТОДЫ РАСЧЕТА ЭЛЕКТРОННЫХ СПЕКТРОВ ТВЕРДЫХ ТЕЛ	49
5.1. Приближение сильной связи	49
5.2. Электронный спектр графитового слоя	53
5.3. Электронный спектр двухслойного графена	62
5.4. Электронные свойства нанотрубок	68
6. ПРИБОРЫ НАНОЭЛЕКТРОНИКИ И МЕТОДЫ ИХ МОДЕЛИРОВАНИЯ	73
6.1. Физические принципы работы различных типов транзисторов	74
6.2. Эволюция моделей МОП транзисторов	86
6.3. Основные квантовые эффекты в полевых нанотранзисторах	88
6.4. Обзор квантовых методов моделирования нанотранзисторов	91

7. РЕШЕНИЕ ЗАДАЧИ РАССЕЯНИЯ ДЛЯ УРАВНЕНИЯ ШРЕДИНГЕРА С ПОМОЩЬЮ T- И S-МАТРИЦ (МОДЕЛИРОВАНИЕ КНИ-ТРАНЗИСТОРА)	96
7.2. Результаты моделирования	103
7.3. Выводы о результатах моделирования	107
8. МОДЕЛИРОВАНИЕ НОВЫХ ЭЛЕКТРОННЫХ ПРИБОРОВ НА ОСНОВЕ ГРАФЕНА	108
8.1. Квантовая гидродинамическая модель транзистора на основе графена	108
8.2. Баллистическая модель транзистора на двухслойном графене	120
ПРИЛОЖЕНИЕ А	133
Алгоритм, реализующий моделирование характеристик транзистора на двухслойном графене в среде Wolfram Mathematica 7.0	
ПРИЛОЖЕНИЕ Б	138
Работа с квантово-химическим пакетом Gaussian	
ЛИТЕРАТУРА	149

Введение

Основной целью учебного пособия по курсу «Математическое и компьютерное моделирование наносистем» является ознакомление читателя с методами решения уравнений квантовой механики в сложных системах. Все пособие условно можно разделить на две части: в первой рассказывается о математических основах численных методов решения уравнений квантовой механики (главы 1-4); во второй — рассказывается о применениях данных методов к моделированию наносистем (главы 5-8). В приложениях к пособию приведены примеры технических реализаций некоторых задач по моделированию.

Целью большинства приближенных методов квантовой механики сложных систем является определение электронной волновой функции (или электронной плотности) и энергетических спектров. Сами по себе эти величины, однако, не дают практически полезных и проверяемых данных о свойствах изучаемых объектов, однако влияют на эти свойства непосредственно. Данный курс построен таким образом, чтобы «фундаментальные» знания о методах расчета многочастичных систем не остались без практического приложения. В частности, различные методы расчета энергетических спектров кластеров находят применение при моделировании оптических свойств наночастиц с помощью программной среды Gaussian (Приложение Б);

модельные спектры электронов в периодическом потенциале, вычисленные в приближении сильной связи (Глава 5) находят применение в моделировании транзисторов на основе новых углеродных материалов (Глава 8).

Конечно, охватить широкий круг задач нанотехнологий, в которых используется моделирование квантовых систем, не представляется возможным в одном пособии. При выборе тем, касающихся приложений методов математического моделирования, авторы руководствовались соображениями наглядности и практической значимости. Так, вычисление оптических спектров молекул и нанокластеров с помощью программы Gaussian позволяет сравнить особенности различных приближенных методов квантовой химии. Разделы о моделировании транзисторов на новых физических принципах знакомят читателя с физическими основами наноэлектроники; более того, моделирование подобных структур важно с точки зрения оптимизации их параметров. Изучение пособия предполагает знакомство читателя с квантовой механикой и основами физики твердого тела.

1. Уравнение Шредингера для систем многих частиц и приближение Борна—Оппенгеймера

Квантовая химия — раздел химии, использующий аппарат квантовой механики для исследования свойств и строения химических соединений, их реакционной способности, кинетики и механизма химических реакций. Такой подход часто называют *ab-initio* (от начала), поскольку задача решается исходя из основополагающих принципов без использования каких-либо эмпирических предположений. Несмотря на это, почти всегда используется какой-либо набор упрощений и приближений, так как решение исходных уравнений потребовало бы слишком большого количества вычислительных мощностей.

В следующих главах будут описаны основы методов квантовой вычислительной химии.

Как известно, поведение частицы в пространстве в квантовой механике описывается стационарным уравнением Шредингера, которое фактически является задачей на собственные значения и собственные функции оператора Гамильтона:

$$\hat{H}\phi(\mathbf{r}) = E\phi(\mathbf{r}), \quad (1.1)$$

где \hat{H} — оператор Гамильтона, описывающий систему, E — энергия состояния, $\phi(\mathbf{r})$ — пространственная часть волновой функции частицы. Оператор Гамильтона может быть записан в виде суммы операторов кинетической и потенциальной энергий:

$$\hat{H} = \hat{T} + \hat{U} = -\frac{\hbar^2}{2m} \Delta + U(\mathbf{r}), \quad (1.2)$$

где m — масса частицы, $U(\mathbf{r})$ — потенциальная энергия частицы.

Для атома водорода уравнение (1.1) (в системе СИ) запишется в виде:

$$\left[-\frac{\hbar^2}{2m_e} \Delta - \frac{e^2}{4\pi\epsilon_0 r} \right] \phi(\mathbf{r}) = E\phi(\mathbf{r}), \quad (1.3)$$

Уравнение Шредингера для систем многих частиц и приближение Борна—Оппенгеймера

где m_e — масса электрона, e — элементарный заряд, ϵ_0 — электрическая постоянная. Произведем следующую замену $x = a_0 x'$, $y = a_0 y'$, $z = a_0 z'$ в уравнении (1.3) для приведения его к безразмерному виду. Получим:

$$\left[-\frac{\hbar^2}{2m_e} \frac{\Delta'}{a_0^2} - \frac{e^2}{4\pi\epsilon_0 a_0 r'} \right] \phi' = E \phi', \quad (1.4)$$

Принимая $a_0 = \frac{4\pi\epsilon_0 \hbar^2}{m_e e^2}$, $E = \mathcal{E} \mathcal{E}_a$, $\mathcal{E}_a = \frac{m_e e^4}{16\pi^2 \epsilon_0^2 \hbar^2}$ получим уравнение:

$$\left[-\frac{\Delta'}{2} - \frac{1}{r'} \right] \phi' = \mathcal{E} \phi', \quad (1.5)$$

которое является уравнением Шредингера, записанным в атомных единицах. Постоянная a_0 называется атомной единицей длины или Боровским радиусом, а \mathcal{E}_a — атомной единицей энергии или энергией Хартри. С помощью этих величин можно сформировать атомную систему единиц, которой удобно пользоваться в расчетах электронных систем. В таблице 1.1 перечислены фундаментальные атомные единицы и коэффициенты их перевода. В дальнейшем, для упрощения записей, будет использоваться именно атомная система единиц.

Таблица 1.1. Фундаментальные атомные единицы.

Единица	Выражение	Значение в СИ
Масса	m_e	$9,1093826(16) \times 10^{-31}$ кг
Длина	$a_0 = \frac{\hbar}{m_e c \alpha}$	$5,291772108(18) \times 10^{-11}$ м
Время	$\frac{\hbar^3}{8\pi^2 m_e e^4}$	$2,418884326505(16) \times 10^{-17}$ сек
Энергия	$m_e c^2 \alpha^2$	$4,35974417(75) \times 10^{-18}$ Дж
Угловой момент	\hbar	$1,05457168(18) \times 10^{-34}$ Дж×сек
Электрический заряд	e	$1,60217653(14) \times 10^{-19}$ Кл

Уравнение Шредингера для систем многих частиц и приближение Борна—Оппенгеймера

Рассмотрим систему, состоящую из N электронов и M атомных ядер. Обозначим положения, массы и заряды ядер \mathbf{R}_i , M_i и Z_i соответственно, положения электронов \mathbf{r}_i , а операторы Лапласа, дифференцирующие по координатам i -го электрона и i -го ядра, — Δ_i и Δ_i^n соответственно. Тогда оператор Гамильтона такой системы запишется как

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{U}(\mathbf{r}, \mathbf{R}) = -\sum_{i=1}^N \frac{1}{2} \Delta_i - \sum_{i=1}^M \frac{1}{2M_i} \Delta_i^n - \sum_{i=1}^N \sum_{j=1}^M \frac{Z_j}{|\mathbf{R}_j - \mathbf{r}_i|} + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{Z_i Z_j}{|\mathbf{R}_j - \mathbf{R}_i|}. \quad (1.6)$$

Очевидно, что процедура решения уравнения Шредингера для такого гамильтониана крайне сложна и прямое вычисление решения потребовало бы огромное количество вычислительных ресурсов даже для простейших систем. Поэтому на практике часто используется набор приближений, сводящий эту задачу к более простой.

Одним из важнейших приближений, используемых в задачах квантовой химии и физики твердого тела, является приближение Борна—Оппенгеймера, являющееся разновидностью адиабатического приближения. Поскольку ядра атомов гораздо тяжелее электронов, характерные времена изменения состояния ядер значительно больше таковых для электронов. Таким образом, можно считать, что электроны движутся в поле неподвижных ядер. При этом в (1.6) можно считать кинетическую и потенциальную энергию взаимодействия ядер постоянной, и, поскольку константа, добавленная к оператору Гамильтона, не влияет на волновую функцию, можно утверждать, что электронная волновая функция определяется следующим уравнением:

$$\hat{H}_e \Phi_e = \mathcal{E}_e \Phi_e, \quad (1.7)$$

где \mathcal{E}_e — энергия электронов, которую также называют адиабатическим потенциалом или адиабатическим электронным термом молекулы, а \hat{H}_e — электронный гамильтониан, равный

$$\hat{H}_e = -\sum_{i=1}^N \frac{1}{2} \Delta_i - \sum_{i=1}^M \sum_{j=1}^M \frac{Z_j}{|\mathbf{R}_j - \mathbf{r}_i|} + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|}. \quad (1.8)$$

Φ_e является функцией \mathbf{r}_i и параметризована координатами \mathbf{R}_i , то есть разным наборам \mathbf{R}_i соответствуют различные Φ_e . Это верно также в отношении \mathcal{E}_e .

Уравнение Шредингера для систем многих частиц и приближение Борна—Оппенгеймера

В приближении неподвижных ядер полную энергию можно записать как энергию электронов и потенциальную энергию взаимодействия ядер:

$$\mathcal{E}_{tot} = \mathcal{E}_\diamond + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{Z_i Z_j}{|\mathbf{R}_j - \mathbf{R}_i|}. \quad (1.9)$$

Решив уравнение Шредингера для электронов, возможно решить уравнение, описывающее поведение ядер. Поскольку, как было сказано выше, электроны двигаются значительно быстрее ядер, возможно заменить члены связанные с электронами, их средними значениями в полном гамильтониане (1.6), после чего ядерный гамильтониан \hat{H}_n возможно записать следующим образом:

$$\begin{aligned} \hat{H}_n &= -\sum_{i=1}^M \frac{1}{2M_i} \Delta_i^n + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{Z_i Z_j}{|\mathbf{R}_j - \mathbf{R}_i|} + \left\langle -\sum_{i=1}^N \frac{1}{2} \Delta_i - \sum_{i=1}^N \sum_{j=1}^M \frac{Z_j}{|\mathbf{R}_j - \mathbf{r}_i|} + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{1}{|\mathbf{r}_j - \mathbf{r}_i|} \right\rangle = \\ &= -\sum_{i=1}^M \frac{1}{2M_i} \Delta_i^n + \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{Z_i Z_j}{|\mathbf{R}_j - \mathbf{R}_i|} + \mathcal{E}_e(\{\mathbf{R}_i\}) = -\sum_{i=1}^M \frac{1}{2M_i} \Delta_i^n + \mathcal{E}_{tot}(\{\mathbf{R}_i\}) \end{aligned} \quad (1.10)$$

Таким образом, можно говорить, что ядра атомов движутся в потенциале, определяемом решением задачи для электронов. Решения уравнения Шредингера для ядер

$$\hat{H} \Phi_n = \mathcal{E} \Phi_n \quad (1.11)$$

описывают степени свободы молекулы, такие, как вибрации, вращения и перемещения. При этом, \mathcal{E} является полной энергией молекулы, включающей в себя как энергию, связанную со степенями свободы молекулы (энергию вибраций, вращений и поступательного движения), так и электронную энергию. В результате, общая волновая функция может быть записана как

$$\Phi = \Phi_e \Phi_n. \quad (1.12)$$

Замена потенциальных энергий в (1.10) на их средние возможна только в том случае, если скорость изменения электронной волновой функции при изменении координат \mathbf{R}_i настолько мала, что можно пренебречь ее первой и второй производной по этим координатам.

Уравнение Шредингера для систем многих частиц и приближение Борна—Оппенгеймера

М. Борном и Р. Оппенгеймером в 1927 году было показано, что обычно это условие выполняется.

Существует простой критерий применимости приближения Борна—Оппенгеймера для устойчивых многоатомных молекул, который записывается следующим образом:

$$\frac{\omega}{\mathcal{E}_{en} - \mathcal{E}_{em}} \ll 1, \quad (1.13)$$

где \mathcal{E}_{en} и \mathcal{E}_{em} — энергии двух соседних электронных состояний ω — наибольшая частота колебаний ядер. Этот критерий выполняется для многих молекул, поэтому расчеты, основанные на приближении Борна—Оппенгеймера, хорошо согласуются с экспериментальными данными, и ошибка, вносимая в результат этим приближением, обычно значительно меньше, чем ошибки, вносимые другими допущениями. Для электронных состояний, отличных от основного, ошибка приближения выше, но и в этом случае ей можно пренебречь по сравнению с ошибками, вносимыми в результат приближенным методом решения электронного уравнения (1.7). Все это обычно позволяет ограничиться решением только электронного уравнения без решения ядерного.

Одним из важных понятий, используемых в дальнейшем, является орбиталь. Орбиталь — это одноэлектронная волновая функция. Пространственной орбиталью называется функция $\phi(\mathbf{r})$, модуль квадрата которой определяет плотность вероятности нахождения электрона в конкретной точке. Часто пространственные орбитали выбирают так, чтобы они образовывали ортонормированный базис:

$$\int \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}) d\mathbf{r} = \delta_{ij}. \quad (1.14)$$

Полным базисом называется такой набор $\phi_i(\mathbf{r})$, что произвольную функцию $f(\mathbf{r})$ можно представить в виде линейной комбинации базисных волновых функций:

$$f(\mathbf{r}) = \sum_{i=1}^{\infty} \alpha_i \phi_i(\mathbf{r}), \quad (1.15)$$

где α_i — некоторые коэффициенты. Вообще говоря, для того, чтобы базис был полон, обычно необходимо, чтобы он был бесконечным, так как произвольная функция имеет бесконечное число степеней свобо-

ды. На практике в расчетах использовать бесконечный базис не представляется возможным.

Кроме пространственных степеней свободы электрон также обладает спином. В нерелятивистской квантовой теории существование спина предсказать невозможно, но, тем не менее, он оказывает существенное влияние на поведение частиц. Поэтому его необходимо ввести в теорию искусственно. Поскольку спин электрона равен $\frac{1}{2}$, существуют только 2 собственных спиновых состояния. Обозначим их $\alpha(\omega)$ и $\beta(\omega)$; одна из этих функций отвечает состоянию «спин вверх», а другая — «спин вниз».

Теперь из каждой пространственной орбитали можно сформировать две ортогональные базисные функции (спин-орбитали):

$$\begin{aligned}\psi(x) &= \phi(\mathbf{r})\alpha(\omega) \\ \psi(x) &= \phi(\mathbf{r})\beta(\omega)\end{aligned}\tag{1.16}$$

Из набора N орбиталей можно построить набор из $2N$ спин-орбиталей (1.9), при этом, если набор исходных орбиталей был ортонормированным, то набор спин—орбиталей будет также ортонормирован.

2. Метод Хартри—Фока

Несмотря на то, что приближение Борна—Оппенгеймера значительно упрощает задачу расчета, она все равно остается очень сложной. Действительно, в случае молекулы, содержащей N валентных электронов, электронная волновая функция Φ_e зависит от $3N$ переменных, что даже для случая двух электронов крайне много для численного расчета. Таким образом, необходимость дополнительного упрощения очевидна.

Для начала рассмотрим систему, состоящую из невзаимодействующих электронов, гамильтониан которой является просто суммой гамильтонианов каждого электрона:

$$\hat{H} = \sum_{i=1}^N \hat{h}_i. \quad (2.1)$$

В этом случае мы, фактически, пренебрегаем взаимным отталкиванием электронов, но энергия отталкивания может быть включена в \hat{h}_i в виде некоторого эффективного потенциала, чем мы воспользуемся в дальнейшем. Для каждого \hat{h}_i существует набор собственных функций $\psi_j(x_i)$, являющихся одноэлектронными волновыми функциями, и соответствующих им собственных значений энергий \mathcal{E}_j :

$$\hat{h}_i \psi_j(x_i) = \mathcal{E}_j \psi_j(x_i). \quad (2.2)$$

Поскольку все электроны находятся в одинаковых условиях, \hat{h}_i отличаются только координатами, на которые они действуют. Таким образом, набор собственных волновых функций для каждого электрона одинаков. Полная волновая функция будет просто произведением волновых функций отдельных электронов, а полная энергия — суммой соответствующих собственных энергий:

$$\begin{aligned} \Psi &= \psi_{j_1} \psi_{j_2} \dots \psi_{j_N} \\ E &= \mathcal{E}_{j_1} + \mathcal{E}_{j_2} + \dots + \mathcal{E}_{j_N} \end{aligned} \quad (2.3)$$

Такая многоэлектронная волновая функция называется произведением Хартри. Произведение Хартри описывает систему электронов, в которой отсутствуют взаимные корреляции, то есть вероятность обнаружить электрон 1 в некоторой точке пространства не зависит от того, в каких

точках были обнаружены другие электроны. В реальности это не так, поскольку частицы взаимно отталкиваются. К тому же такое решение не удовлетворяет принципу тождественности частиц, так как явно указывает на то, что электрон i занимает орбиталь j_i , то есть при перестановке двух электронов местами волновая функция изменит свой вид, хотя, согласно принципу Паули, она может лишь изменить свой знак.

Для решения этой проблемы можно применить специальную процедуру симметризации. Для простоты рассмотрим систему двух электронов. Произведение Хартри (2.3) будет иметь вид:

$$\Psi_{12} = \psi_{j_1}(x_1)\psi_{j_2}(x_2). \quad (2.4)$$

При перестановке электронов получим:

$$\Psi_{21} = \psi_{j_1}(x_2)\psi_{j_2}(x_1). \quad (2.5)$$

Легко видеть, что обе волновые функции не соответствуют принципу тождественности. Поскольку уравнение Шредингера линейно, то линейная комбинация его решений, соответствующая одному собственному числу, также будет решением. Это дает нам возможность записать антисимметричную волновую функцию в виде комбинации (2.4) и (2.5):

$$\Psi = \frac{1}{\sqrt{2}}(\Psi_{12} - \Psi_{21}) = \frac{1}{\sqrt{2}}(\psi_{j_1}(x_1)\psi_{j_2}(x_2) - \psi_{j_1}(x_2)\psi_{j_2}(x_1)), \quad (2.6)$$

где $\frac{1}{\sqrt{2}}$ — нормировочный множитель, а знак минус выбран для того, чтобы получить именно антисимметричную функцию:

$$\Psi(x_1, x_2) = -\Psi(x_2, x_1). \quad (2.7)$$

Благодаря антисимметричности волновой функции две частицы не могут иметь одинаковый набор координат. То есть антисимметричность эквивалентна принципу Паули. Выражение (2.6) можно переписать в виде:

$$\Psi = \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_{j_1}(x_1) & \psi_{j_2}(x_1) \\ \psi_{j_1}(x_2) & \psi_{j_2}(x_2) \end{vmatrix}. \quad (2.8)$$

Такая запись называется детерминантом Слэтера и удобна для записи антисимметричной волновой функции системы N частиц:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{j_1}(x_1) & \psi_{j_2}(x_1) & \dots & \psi_{j_N}(x_1) \\ \psi_{j_1}(x_2) & \psi_{j_2}(x_2) & & \psi_{j_N}(x_2) \\ \dots & & & \dots \\ \psi_{j_1}(x_N) & \psi_{j_2}(x_N) & \dots & \psi_{j_N}(x_N) \end{vmatrix}. \quad (2.9)$$

Для краткости в дальнейшем для детерминанта Слэтера будет использоваться следующее обозначение:

$$\Psi = |\psi_{j_1} \psi_{j_2} \dots \psi_{j_N} \rangle. \quad (2.10)$$

Стоит также отметить, что состояние, описываемое детерминантом Слэтера, уже не является нескоррелированным, так как частицы с одинаковыми спинами уже не могут находиться в одной точке пространства. Рассмотрим корреляцию частиц подробнее. Для начала рассмотрим состояния с одинаковым спином. Положим:

$$\begin{aligned} \psi_1(x_1) &= \phi_1(\mathbf{r}_1)\beta(\omega_1) \\ \psi_2(x_1) &= \phi_2(\mathbf{r}_2)\beta(\omega_2) \end{aligned} \quad (2.11)$$

Запишем вероятность $P(\mathbf{r}_1, \mathbf{r}_2)$ обнаружения частиц в точках с координатами $\{\mathbf{r}_1, \mathbf{r}_2\}$, проинтегрировав квадрат волновой функции по спиновым координатам:

$$\begin{aligned} P(\mathbf{r}_1, \mathbf{r}_2) &= \int d\omega_1 d\omega_2 |\psi_1 \psi_2|^2 d\mathbf{r}_1 d\mathbf{r}_2 = \frac{1}{2} [|\psi_1(\mathbf{r}_1)|^2 |\psi_2(\mathbf{r}_2)|^2 + |\psi_1(\mathbf{r}_2)|^2 |\psi_2(\mathbf{r}_1)|^2 + \\ &+ \psi_1(\mathbf{r}_1) \psi_2^*(\mathbf{r}_1) \psi_2(\mathbf{r}_2) \psi_1^*(\mathbf{r}_2) - \psi_1^*(\mathbf{r}_1) \psi_2(\mathbf{r}_1) \psi_2^*(\mathbf{r}_2) \psi_1(\mathbf{r}_2)]. \end{aligned} \quad (2.12)$$

Видим, что кроме суммы квадратов плотностей вероятности в выражении появляется обменный член, что приводит к скоррелированности частиц. Легко заметить, что $P(\mathbf{r}, \mathbf{r}) = 0$.

В случае различных спиновых функций, когда:

$$\begin{aligned} \psi_1(x_1) &= \phi_1(\mathbf{r}_1)\alpha(\omega_1) \\ \psi_2(x_1) &= \phi_2(\mathbf{r}_2)\beta(\omega_2) \end{aligned} \quad (2.13)$$

$P(\mathbf{r}_1, \mathbf{r}_2)$ будет выглядеть как:

$$P(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2} [|\psi_1(\mathbf{r}_1)|^2 |\psi_2(\mathbf{r}_2)|^2 + |\psi_1(\mathbf{r}_2)|^2 |\psi_2(\mathbf{r}_1)|^2], \quad (2.14)$$

то есть поведение электронов действительно не скоррелировано.

В дальнейшем нам понадобится вычислять матричные элементы операторов от волновых функций, заданных детерминантами Слэтера. Поскольку для системы N электронов детерминант содержит $N!$ элементов, необходима специальная техника работы с многоэлектронными интегралами.

Пусть $\hat{h}(i)$ — некоторый оператор, действующий на один электрон, а $\hat{g}(i, j)$ — двухэлектронный оператор. Нас будут интересовать значения $\langle \Psi_1 | \Psi_2 \rangle$, $\langle \Psi_1 | \hat{h} | \Psi_2 \rangle$ и $\langle \Psi_1 | \hat{g} | \Psi_2 \rangle$. Сначала рассмотрим первое выражение. По определению детерминанта:

$$\langle \Psi_1 | \Psi_2 \rangle = \frac{1}{N!} \sum_{i,j=1}^N (-1)^{i+j} \langle \psi_{1i} | \psi_{2j} \rangle \times M_{1i}(2,3\dots N) | M_{2j}(2,3\dots N) \rangle, \quad (2.15)$$

где M_{1i} и M_{2j} — соответствующие миноры. Поскольку $\langle \psi_{1i} | \psi_{2j} \rangle = 1$ только в том случае, если $\psi_{1i} = \psi_{2j}$, а в противном случае $\langle \psi_{1i} | \psi_{2j} \rangle = 0$, то если Ψ_1 и Ψ_2 отличаются хотя бы одной спин-орбиталью, $\langle \Psi_1 | \Psi_2 \rangle$ будет равно 0. В противном случае получим $\langle \Psi_1 | \Psi_2 \rangle = 1$.

Перейдем к $\langle \Psi_1 | \hat{h} | \Psi_2 \rangle$. Подобно (2.15) получим:

$$\langle \Psi_1 | \hat{h}(1) | \Psi_2 \rangle = \frac{1}{N!} \sum_{i,j=1}^N (-1)^{i+j} \langle \psi_{1i} | \hat{h}(1) | \psi_{2j} \rangle \times M_{1i}(2,3\dots N) | M_{2j}(2,3\dots N) \rangle. \quad (2.16)$$

Подобно предыдущему случаю, если M_{1i} равен M_{2j} или отличается от него только порядком нумерации спин-орбиталей, то $\langle M_{1i}(2,3\dots N) | M_{2j}(2,3\dots N) \rangle = (N-1)!$.

Иначе $\langle M_{1i}(2,3\dots N) | M_{2j}(2,3\dots N) \rangle = 0$.

Матричный элемент $\langle \psi_{1i} | \hat{h}(1) | \psi_{2j} \rangle$ уже не обязательно обращается в нуль при $\psi_{1i} \neq \psi_{2j}$. Поэтому возможны два случая:

$\Psi_1 = \Psi_2$, тогда

$$\langle \Psi_1 | \hat{h}(1) | \Psi_1 \rangle = \frac{1}{N} \sum_{i=1}^N \langle \psi_{1i} | \hat{h}(1) | \psi_{1i} \rangle. \quad (2.17)$$

Ψ_1 и Ψ_2 отличаются лишь одной спин-орбиталью, скажем, $\psi_{1i} \neq \psi_{2i}$. Тогда:

$$\langle \Psi_1 | \hat{h}(1) | \Psi_1 \rangle = \frac{1}{N} \langle \psi_{1i} | \hat{h}(1) | \psi_{2i} \rangle. \quad (2.18)$$

В случаях, отличных от этих, рассматриваемый матричный элемент обращается в нуль. Рассмотрим сумму одноэлектронных операторов $\sum_{i=1}^N \hat{h}(i)$. Если в матричном элементе заменить переменные i -го электрона на переменные 1-го, то, согласно принципу тождественности электронов, матричный элемент не изменится, а оператор $\hat{h}(i)$ превратится в $\hat{h}(1)$. Таким образом, при нахождении матричного элемента для всей системы электронов достаточно умножить соответствующие элементы вида (2.17) и (2.18) на число электронов.

Аналогичные рассуждения могут быть проведены для матричных элементов двухэлектронных операторов. Не останавливаясь на вычислениях подробно, выпишем соотношения:

$$\langle \Psi | \hat{g} | \Psi \rangle = \sum_{i < j} \langle \psi_i \psi_j | \hat{g}(1,2) | \psi_i \psi_j \rangle - \langle \psi_i \psi_j | \hat{g}(1,2) | \psi_j \psi_i \rangle, \quad (2.19)$$

$$\langle \Psi_i^j | \hat{g} | \Psi \rangle = \sum_{k \neq i} \langle \psi_j \psi_k | \hat{g}(1,2) | \psi_i \psi_k \rangle - \langle \psi_j \psi_k | \hat{g}(1,2) | \psi_k \psi_i \rangle, \quad (2.20)$$

где Ψ_i^j — детерминант Слэтера, в котором ψ_i заменена на ψ_j и

$$\langle \Psi_{ik}^{jl} | \hat{g} | \Psi \rangle = \langle \psi_j \psi_l | \hat{g}(1,2) | \psi_i \psi_k \rangle - \langle \psi_j \psi_l | \hat{g}(1,2) | \psi_k \psi_i \rangle. \quad (2.21)$$

где Ψ_{ik}^{jl} — детерминант Слэтера, в котором ψ_i заменена на ψ_j , а ψ_k заменена на ψ_l . В остальных случаях рассматриваемый матричный элемент равен нулю.

Рассмотрим теперь вид решения для некоторой системы с учетом взаимодействия электронов. Запишем гамильтониан системы:

$$\hat{H} = \sum_{i=1}^N \hat{h}_i + \sum_{i=1}^{N-1} \sum_{j=i}^N \frac{e^2}{|r_j - r_i|}, \quad (2.22)$$

где

$$\hat{h}_i = -\frac{\Delta_i}{2} + U_n(\mathbf{r}_i), \quad (2.23)$$

где $U_n(\mathbf{r}_i)$ — потенциал ядерного поля.

Пусть основное состояние системы, описываемой гамильтонианом \hat{H} , выражается волновой функцией:

$$\Psi_0 = |\psi_1 \psi_2 \dots \psi_N\rangle. \quad (2.24)$$

Согласно вариационному принципу, эта функция должна минимизировать функционал энергии:

$$E_0 = \langle \Psi_0 | \hat{H} | \Psi_0 \rangle, \quad (2.25)$$

то есть, варьируя ψ_n так, чтобы они оставались ортонормированными до тех пор, пока минимум E_0 не будет достигнут можно получить набор спин-орбиталей, описывающий систему наилучшим образом. Опуская математические выкладки, запишем интегродифференциальное уравнение на орбитали, получающееся из вариационного принципа (уравнение Хартри-Фока):

$$\hat{h}\psi_a(x_1) + \left[\sum_{b \neq a} \frac{|\psi_b(x_2)|^2}{r_{12}} dx_2 \right] \psi_a(x_1) - \left[\sum_{b \neq a} \frac{\psi_b^*(x_2) \psi_a(x_2)}{r_{12}} dx_2 \right] \psi_b(x_1) = \mathcal{E}_a \psi_a(x_1), \quad (2.26)$$

где \mathcal{E}_a — энергия спин-орбитали a .

Легко заметить, что без двух интегральных членов уравнение (2.26) представляло бы собой просто уравнение Шредингера для электрона, находящегося в поле ядер. Интегральные члены, в свою очередь, описывают межэлектронное взаимодействие. Первый из них определяет электростатическое отталкивание электронов, второй — возникает из-за антисимметричности волновой функции.

Рассмотрим кулоновский член подробнее. Для удобства можно ввести кулоновский оператор:

$$\hat{J}_b(x_1) = \int \frac{\psi_b^*(x_2) \psi_b(x_2)}{r_{12}} dx_2, \quad (2.27)$$

который, фактически, определяет вклад электрона на спин-орбитали ψ_b в потенциал в точке x_1 .

Второй член не имеет простой классической интерпретации, поскольку возникает из-за чисто квантового эффекта. Его так же удобно представить в виде суммы операторов, называемых обменными:

$$\hat{\mathcal{K}}_b(x_1)\psi_a(x_1) = \left[\int \frac{\psi_b^*(x_2)\psi_a(x_2)}{r_{12}} dx_2 \right] \psi_b(x_1). \quad (2.28)$$

Теперь уравнение (2.26) можно переписать в виде:

$$\left[h_i + \sum_{b \neq a} \mathcal{J}_b(x_i) - \sum_{b \neq a} \mathcal{K}_b(x_i) \right] \psi_a(x_i) = \mathcal{E}_a \psi_a(x_i). \quad (2.29)$$

Уравнение (2.29) является уравнением на собственные функции некоторого оператора, но форма оператора зависит от вида самой собственной функции, так как он содержит неполные суммы по спин-орбиталям.

Из (2.27) и (2.28) следует, что

$$(\mathcal{J}_a - \mathcal{K}_a)\psi_a = 0, \quad (2.30)$$

что позволяет заменить суммы в (2.29) на полные. Определим оператор Фока, как

$$\hat{F} = \hat{h} + \sum_b (\mathcal{J}_b - \mathcal{K}_b) = \hat{h} + v^{HF}, \quad (2.31)$$

где v^{HF} называют потенциалом Хартри-Фока. Тогда уравнение Хартри-Фока (2.26) можно переписать в виде:

$$\hat{F}\psi_a = \mathcal{E}_a\psi_a. \quad (2.32)$$

Обычно точное решение этого уравнения можно получить лишь для атомов. Для более сложных систем обычно используют некоторый набор базисных функций. В этом случае можно перейти к набору матричных уравнений, решив которые, можно получить приближенное решение. Если бы было возможно использовать полный базис, то такой метод дал бы точное решение уравнения Хартри-Фока.

Ввиду того, что оператор Фока сам по себе зависит от решения уравнения Хартри-Фока через кулоновский и обменный операторы, то уравнение (2.32) не является задачей на собственные функции, и, вообще говоря, нелинейно. Поэтому его численное решение требует некой итеративной процедуры.

Рассмотрим теперь, какие свойства молекулы можно получить, используя метод Хартри-Фока. Одним из основных параметров является

полная энергия молекулы. Ввиду того, что молекулярные орбитали являются лишь приближением, а в реальности волновые функции более сложны (из-за электронных корреляций невозможно разделить переменные и представить полную волновую функцию в виде произведения спин-орбиталей), энергия, получаемая с помощью метода Хартри-Фока будет несколько выше реальной (реальная энергия основного состояния является, очевидно, абсолютным минимумом). В большинстве задач разница очень мала и абсолютные энергии, вычисленные таким методом, очень точны, что весьма неплохо для такого относительно простого приближения.

Однако для вычисления относительных энергий метод Хартри-Фока является уже значительно менее точным. В частности, вычисление энергии связи, являющейся разностью между полной энергией и энергией атомов, из которых состоит молекула, метод Хартри-Фока не всегда является достаточно точным и по этой причине необходимы другие методы вычисления этого параметра.

Кроме полной энергии также можно вычислить орбитальные энергии (матричные элементы оператора Фока на спин-орбиталях). Эти энергии полезны для построения приближенной диаграммы уровней и вычисления приближенной энергии ионизации.

Энергия ионизации равна разности энергии молекулы и иона. В первом приближении для иона можно пренебречь изменением кулоновского и обменного оператора а также изменением положения ядер. Таким образом, за энергию ионизации можно примерно принять разность между энергией молекулы и энергией спин-орбитали удаленного электрона. Этот результат называется теоремой Купманса.

Дипольный момент молекулы также может быть вычислен. Оператор дипольного момента записывается как $q \mathbf{r}$, где q — заряд. Полный дипольный момент, соответственно, можно записать в виде:

$$\mu = \sum_{electrons} \langle \Psi | -\mathbf{r}_i | \Psi \rangle + \sum_{nuclei} \mathbf{R}_i Z_i, \quad (2.33)$$

где \mathbf{R}_i — положения ядер, а Z_i — их заряды.

Зная волновую функцию, также можно вычислить атомные заряды и распределение электрического потенциала.

Кроме вычислений, использующих исключительно электронные оболочки, возможно также вычислить геометрию самой молекулы и ее колебательный спектр. Как было сказано выше, в приближении Борна-Оппенгеймера электронная энергия есть функция положения ядер. Эта

зависимость называется поверхностью потенциальной энергии. Зная волновые функции в некоторой точке, можно вычислить производные этой поверхности в этой точке, после чего можно применить методы математической оптимизации и найти положение ядер, в котором энергия молекулы достигает минимума. Таким образом можно искать стабильные структуры для больших молекул. Зная производные, можно также рассчитать частоту колебаний, что требуется для вычисления колебательного спектра молекулы.

С помощью вычисленной поверхности потенциальной энергии можно также вычислить реактивность. Идея такого вычисления заключается в нахождении наиболее низкого потенциального барьера между двумя устойчивыми состояниями и траектории, которая пересекает этот барьер. Это позволяет получить информацию о механизме реакции и ее скорости.

Вычислительная сложность метода Хартри—Фока быстро возрастает с ростом числа атомных электронов, так что соответствующее расчетное время для тяжелых атомов оказывается слишком большим даже при использовании современных компьютеров. Другим недостатком данного метода является то, что в нем не учитывается корреляционное взаимодействие между электронами, зависящее от разности их координат. Учет корреляционного взаимодействия находится за рамками приближения самосогласованного поля, которое является физической основой метода Хартри—Фока.

Указанные недостатки до некоторой степени удается преодолеть с использованием подходов, в которых производится локализация обменного потенциала. Одним из таких методов является приближение локальной электронной плотности, в котором плотность электронов $n(\mathbf{r})$ рассматривается как базовая величина, определяющая свойства основного состояния атома. В рамках данного приближения удастся в известной мере учесть и корреляционное межэлектронное взаимодействие путем введения единого обменно-корреляционного потенциала.

3. Метод функционала электронной плотности

Теория функционала плотности (DFT, density functional theory) — один из методов расчета электронной структуры сложных систем, состоящих из многих частиц. Вместе с методом самосогласованного поля он является одним из наиболее универсальных и часто применяемых методов.

DFT является *ab initio* методом приближенного решения уравнения Шредингера для системы многих частиц. Так же как и метод Хартри-Фока, метод функционала плотности использует приближение Борна—Оппенгеймера. Благодаря своей универсальности, в настоящее время метод начинает использоваться не только в вычислительной химии и физике твердого тела (например, для вычисления энергий связи или зонной структуры), но и в областях, считавшихся весьма далекими от квантовой механики, например, в минералогии и биологии.

Универсальность метода связана с тем, что в его основу положены весьма общие предположения. Как и метод Хартри-Фока, теория функционала плотности пытается свести задачу многих частиц, которая требует вычисления функции, зависящей от очень большого числа переменных, к задаче, описывающей поведение одной частицы. Центральной идеей метода является использование наблюдаемой величины, называемой электронной плотностью:

$$n(\mathbf{r}) = N \int \Psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) d\mathbf{r}_2 \dots d\mathbf{r}_N, \quad (3.1)$$

в качестве основной вычисляемой функции.

Методу теории функционала плотности предшествовала модель Томаса — Ферми, развитая Л. Томасом и Энрико Ферми в 1927 г. Они рассчитали энергию атома как сумму его кинетической энергии, представленной в виде функционала электронной плотности, и потенциальной энергии взаимодействия электронов с ядром и друг с другом; энергия взаимодействия также была выражена через электронную плотность.

Теория основана на предположении о том, что все электроны занимают одинаковый объем в фазовом пространстве, при этом каждый элементарный объем h^3 занят двумя электронами. Тогда можно записать, что:

$$n(\mathbf{r}) = \frac{8\pi}{3h^3} p_f^3(\mathbf{r}), \quad (3.2)$$

где $p_f(\mathbf{r})$ — импульс Ферми. Это уравнение можно разрешить относительно импульса и записать выражение для максимальной кинетической энергии электрона в месте, где электронная плотность равна $n(\mathbf{r})$:

$$\frac{p_0^2}{2} = \frac{1}{2} (3\pi^2 n)^{2/3}. \quad (3.3)$$

Если $\varphi(\mathbf{r})$ — электрический потенциал в данной точке пространства, то полная энергия электрона запишется в виде

$$\varepsilon = \frac{p^2}{2} - \varphi; \quad (3.4)$$

ноль потенциала выбран на бесконечности. Полная энергия электрона в атоме должна быть отрицательной, причем максимальному импульсу соответствует отрицательная энергия, которую мы обозначим $-\varphi_0$, то есть

$$\frac{p_0^2}{2} = \varphi - \varphi_0. \quad (3.5)$$

Теперь мы можем получить, комбинируя (3.3) и (3.5), уравнение, связывающее локальную электронную плотность и локальный потенциал:

$$n(\mathbf{r}) = \frac{[2(\varphi(\mathbf{r}) - \varphi_0)]^{3/2}}{3\pi^2}. \quad (3.6)$$

При $\varphi(\mathbf{r}) = \varphi_0$ мы имеем нулевую электронную плотность — это место определяет границу атома. Электрическое поле вне нейтрального атома также отсутствует (по теореме Гаусса), поэтому нулевое значение потенциала простирается от бесконечности до границы атома. Поэтому мы приходим к выводу о том, что константа φ_0 для нейтральных атомов должна быть положена равной нулю

$$n(\mathbf{r}) = \frac{[2\varphi(\mathbf{r})]^{3/2}}{3\pi^2}. \quad (3.7)$$

Подставив (3.7) в уравнение Пуассона, мы получим окончательное уравнение, описывающее поведение потенциала в атоме:

$$\Delta\varphi = \frac{8\sqrt{2}}{3\pi}\varphi^{3/2} \quad (3.8)$$

с граничными условиями

$$\begin{aligned} \varphi(r) &\xrightarrow{r \rightarrow 0} \frac{Z}{r}, \\ \varphi(\infty) &= 0. \end{aligned} \quad (3.9)$$

Это уравнение может быть численно проинтегрировано, причем распределение электронной плотности в атомах с разным порядковым номером оказывается подобным. На самом деле это, конечно же, не так. Применимость метода Томаса-Ферми ограничена квазиклассическим приближением, согласно которому в каждой клетке фазового пространства объемом $(2\pi\hbar)^3$ находится по 2 электрона.

Несмотря на заметную роль, которую модель Томаса — Ферми сыграла в развитии квантовой механики, её точность была недостаточной, поскольку не учитывалось обменное взаимодействие, в отличие, например, от метода Хартри — Фока. В 1928 г. Поль Дирак уточнил функционал энергии в модели Томаса — Ферми, добавив к нему слагаемое, описывающее обменное взаимодействие (это слагаемое также имело вид функционала электронной плотности).

Несмотря на это, для ряда применений модель Томаса — Ферми — Дирака не давала удовлетворительного результата. Основным источником погрешности являлось выражение кинетической энергии, приводящее к погрешности в вычислении обменной энергии. Кроме того, не учитывалась энергия электронной корреляции.

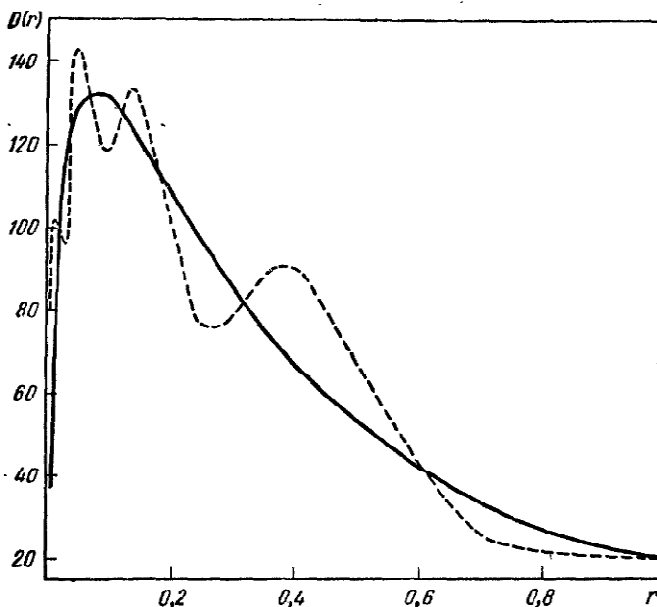


Рис. 3.1. Распределение электронной плотности $D(r) = 4\pi r^2 n(r)$ в атоме ртути как функция расстояния до ядра (в атомных единицах длины). Для сравнения штриховой линией показан результат расчета по методу Хартри-Фока.

Теория функционала плотности базируется на ставшей классической модели Томаса—Ферми, но надёжное теоретическое обоснование под нее было подведено только с формулировкой теорем Хоэнберга—Кона. Эти теоремы составляют основу метода.

Первая теорема утверждает, что существует взаимно однозначное соответствие между плотностью основного состояния электронной подсистемы, находящейся во внешнем потенциале атомных ядер, и самим потенциалом ядер. Первая теорема является теоремой существования и не дает метода построения такого соответствия.

Вторая теорема представляет собой вариационный принцип квантовой механики, сформулированный для функционала плотности, и утверждает, что энергия электронной подсистемы, записанная как функционал электронной плотности, имеет минимум, равный энергии основного состояния.

Первоначально теоремы Хоэнберга — Кона были сформулированы только для основного состояния электронной подсистемы в отсутствие магнитного поля. Они могут быть обобщены путём введения зависимости от времени, что позволяет использовать этот формализм для расчета состояний возбуждённых электронов.

Традиционные методы определения электронной структуры, в частности, метод Хартри — Фока и производные от него, описывают систему с помощью многоэлектронной волновой функции. Основная цель теории функционала плотности — при описании электронной подсистемы заменить многоэлектронную волновую функцию электронной плотностью. Это ведет к существенному упрощению задачи, поскольку многоэлектронная волновая функция зависит от $3N$ переменных — по 3 пространственных координаты на каждый из N электронов, в то время как плотность — функция лишь трёх пространственных координат.

Как правило, метод теории функционала плотности используется совместно с формализмом Кона — Шэма, в рамках которого трудно-разрешимая задача об описании нескольких взаимодействующих электронов в статическом внешнем поле (атомных ядер) сводится к более простой задаче о независимых электронах, которые движутся в некотором эффективном потенциале. Этот эффективный потенциал включает в себя статический потенциал атомных ядер, а также учитывает кулоновские эффекты, в частности, обменное взаимодействие и электронную корреляцию.

Описание двух последних взаимодействий и представляет собой основную сложность метода теории функционала плотности в формулировке Кона — Шэма. Простейшим приближением здесь является приближение локальной плотности, основанное на точном расчёте обменной энергии для пространственно однородного электронного газа, который может быть выполнен в рамках модели Томаса — Ферми и из которого можно получить также и корреляционную энергию электронного газа.

Метод теории функционала плотности широко применяется для расчётов в физике твёрдого тела с 1970-х годов. В ряде случаев даже использование простого приближения локальной плотности дает удовлетворительные результаты, соответствующие экспериментальным данным, причём вычислительная сложность метода невысока относительно других подходов к проблеме многих частиц в квантовой механике. Тем не менее, долгое время метод был недостаточно точен для расчётов в области квантовой химии, пока в 1990-х годах не про-

изошёл заметный сдвиг в описании обменного и корреляционного взаимодействий. В настоящее время метод теории функционала плотности является главным подходом в обеих областях. Впрочем, несмотря на прогресс в теории, все ещё имеются проблемы в применении метода к описанию межмолекулярных сил, в особенности Ван-дер-Ваальсовых сил и дисперсионного взаимодействия, а также в расчётах ширины запрещённой зоны в полупроводниках.

Сложности с расчётом дисперсионного взаимодействия в рамках теории функционала плотности (которые возникают, как минимум, в том случае, когда этот метод не дополняется другими) делают метод теории функционала плотности малоприменимым для систем, в которых дисперсионные силы являются преобладающими (например, при рассмотрении взаимодействия между атомами благородных газов) или систем, в которых дисперсионные силы имеют тот же порядок, что и другие взаимодействия (например, в органических молекулах). Решение этой проблемы является предметом современных исследований.

Рассмотрим уравнение Шредингера для системы электронов:

$$(\hat{T} + \hat{U} + \hat{V})\Psi = E\Psi, \quad (3.10)$$

где \hat{T} — оператор кинетической энергии электронов:

$$\hat{T} = -\frac{1}{2} \sum_{i=1}^N \Delta(i), \quad (3.11)$$

\hat{U} — оператор, описывающий энергию взаимодействия электронов:

$$\hat{U} = \sum_{i < j} U(\mathbf{r}_i, \mathbf{r}_j), \quad (3.12)$$

а \hat{V} — оператор потенциальной энергии, описывающий взаимодействие электронов с ядрами:

$$\hat{V} = \sum_{i=1}^N V_n(\mathbf{r}_i). \quad (3.13)$$

Очевидно, что первые два оператора универсальны для всех систем и каждая конкретная система отличается от остальных только оператором \hat{V} .

Фактически, теорема Хоэнберга—Кона означает, что выражение (3.1) может быть обращено, то есть по заданной плотности частиц в основном состоянии, можно найти соответствующую волновую функцию основного состояния. Можно записать, что:

$$\Psi_0 = \hat{\Psi}_0[n(\mathbf{r})], \quad (3.14)$$

Благодаря этому можно утверждать, что все наблюдаемые физические величины также являются функционалами $n(\mathbf{r})$. Если

$$O = \langle \Psi_0 | \hat{O} | \Psi_0 \rangle, \quad (3.15)$$

то

$$O(n_0) = \langle \Psi_0[n_0] | \hat{O} | \Psi_0[n_0] \rangle. \quad (3.16)$$

Тогда, согласно теореме Хоэнберга—Кона, для основного состояния электронной системы, находящейся в ядерном потенциале $V_n(\mathbf{r})$, существует универсальный функционал $\hat{F}[n(\mathbf{r})]$, не зависящий от системы, действующий на электронную плотность $n(\mathbf{r})$, который определяет полную энергию системы как

$$E = F[n(\mathbf{r})] + \int V_n(\mathbf{r})n(\mathbf{r})d\mathbf{r}. \quad (3.17)$$

Используя это выражение, можно было бы найти $n(\mathbf{r})$, используя вариационный принцип. К сожалению, так как форма функционала $\hat{F}[n(\mathbf{r})]$ точно не известна, необходимы дальнейшие предположения.

Наиболее часто используемой аппроксимацией является приближение локальной электронной плотности (local density approximation, LDA). Впервые оно было сформулировано Коном и Шэмом в 1965 году. В этом приближении функционал \hat{F} записывается в виде:

$$\hat{F}[n(\mathbf{r})] = T[n(\mathbf{r})] + \frac{1}{2} \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + E_{xc}[n(\mathbf{r})], \quad (3.18)$$

где $T[n(\mathbf{r})]$ — электронная кинетическая энергия, а $E_{xc}[n(\mathbf{r})]$ — функционал обменно-корреляционной энергии, выражающийся как:

$$E_{xc}[n(\mathbf{r})] = \int n(\mathbf{r})\epsilon_{xc}(n)d\mathbf{r}, \quad (3.19)$$

где $\epsilon_{xc}(n)$ — обменная и корреляционная энергия на одну частицу электронного газа. Именно потому, что в этом приближении обменно-корреляционная энергия зависит от локальной плотности заряда в точке, приближение носит такое название.

Корреляционный функционал вообще говоря является суммой корреляционного и обменного вкладов:

$$E_{xc} = E_c + E_x. \quad (3.20)$$

В приближении локальной электронной плотности обмен E_x вычисляется как для однородного газа с плотностью, равной плотности в точке (то есть LDA подходит для медленно меняющихся плотностей):

$$E_x = -\frac{3e^2}{4\pi} \int n(\mathbf{r})(3\pi^2 n(\mathbf{r}))^{\frac{1}{3}} d\mathbf{r}. \quad (3.21)$$

Применяя к (3.17) вариационный принцип и учитывая, что электронная плотность нормирована на число частиц ($\int n(\mathbf{r})d\mathbf{r} = N$), получим уравнение:

$$\int \left[\frac{\delta T[n]}{\delta n} + V_n(\mathbf{r}) + \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \frac{\delta E_{xc}[n]}{\delta n} - \mu \right] \delta n(\mathbf{r})d\mathbf{r} = 0, \quad (3.22)$$

где μ — множитель Лагранжа. Можно определить электронную плотность с помощью волновых функций уровней:

$$n(\mathbf{r}) = \sum_{i=1}^N |\psi_i(\mathbf{r})|^2, \quad (3.23)$$

тогда кинетическую энергию можно определить следующим образом:

$$T[n] = -\frac{1}{2} \sum_{i=1}^N \int \psi_i^*(\mathbf{r}) \Delta \psi_i(\mathbf{r}) d\mathbf{r}. \quad (3.24)$$

Тогда решение уравнения (3.22) является решением одноэлектронного уравнения Шредингера:

$$\left[-\frac{1}{2} \Delta + V_n(\mathbf{r}) + \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \frac{\delta E_{xc}[n]}{\delta n} \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}). \quad (3.25)$$

Эти уравнения называются уравнениями Кона-Шэма. Соответственно их решения являются одноэлектронными волновыми функциями, а собственные значения — энергиями соответствующих состояний.

Уравнение (3.25), вообще говоря, нелинейно, так как оператор Гамильтона сам зависит от решения, поэтому для получения решения необходимо применять самосогласованный итерационный метод.

Стоит отметить, что ψ_i не являются спин-орбиталями в приближении Хартри-Фока, поэтому детерминант Слэтера, составленный из них, не будет волновой функцией системы в приближении Хартри-Фока. Несмотря на это, плотность заряда, выражаемая через ψ_i , является правильной.

Для того чтобы рассчитать полную энергию и электронную плотность методом DFT, необходимо вычислить орбитали ψ_i , которые являются решением уравнения (3.25). Для того, чтобы записать это уравнение необходимо знать ядерный потенциал $V_n(\mathbf{r})$ (который мы задаем при формулировке задачи), электронный потенциал и корреляционный вклад. Электронный потенциал $U_e(r) = \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}d\mathbf{r}'$ можно определить, решив уравнение Пуассона:

$$\Delta U_e(\mathbf{r}) = -4\pi n(\mathbf{r}). \quad (3.26)$$

Также необходимо уравнение для вычисления корреляционного вклада. Таким образом мы получим систему уравнений, решив которую, мы получим электронную плотность.

Поскольку система является нелинейной, основным методом ее решения является итерационный. На первом шаге каким-либо образом выбирается и фиксируется начальная электронная плотность $n_0(\mathbf{r})$. После этого для нее решается система уравнений, что дает набор ψ_i . Затем, используя (3.23), можно вычислить новую электронную плотность $n_1(\mathbf{r})$ и повторить процесс. Эти итерации необходимо повторять до тех пор, пока $n_i(\mathbf{r})$ будет отличаться от $n_{i-1}(\mathbf{r})$ не больше, чем на некоторую допустимую погрешность.

Полученная электронная плотность и будет искомой. После этого можно вычислить полную энергию по формуле:

$$E = -\frac{1}{2} \sum \left[\int \psi_i^* \Delta \psi_i d\mathbf{r} + \int V_n(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \int \epsilon_{xc}(n) n(\mathbf{r}) d\mathbf{r} \right]. \quad (3.27)$$

Используя вычисленные величины можно применить процедуры, описанные в предыдущей главе, для вычисления различных свойств.

Как было сказано выше, приближение LDA рассчитано на системы с медленно меняющейся электронной плотностью. Несмотря на это, она довольно точна и для систем с быстро меняющейся плотностью, например, твердых тел. К сожалению, для задач квантовой химии LDA оказывается недостаточно точной. К тому же она не учитывает спиновые корреляции.

Вторая проблема может быть решена довольно просто привлечением приближения локальной спиновой плотности, в которой корреляционная энергия определяется как:

$$E_{xc}(n_{\uparrow}, n_{\downarrow}) = \int \epsilon_{xc}(n_{\uparrow}, n_{\downarrow}) n(\mathbf{r}) d\mathbf{r}. \quad (3.28)$$

Для этого приближения значение $\epsilon_{xc}(n_{\uparrow}, n_{\downarrow})$ было вычислено достаточно точно с помощью стохастических методов (квантовый метод Монте-Карло) при расчетах систем свободных электронов.

Первая проблема приводит к тому, что молекулярные связи, вычисляемые при оптимизации геометрии, оказываются слишком короткими.

Для решения этой проблемы может использоваться обобщенная градиентная аппроксимация (generalized gradient approximation, GGA). Помимо значения локальной электронной плотности, она использует также градиент плотности в точке. Таким образом, GGA можно считать следующим порядком приближения по сравнению с LDA.

Корреляционный функционал GGA записывается в виде:

$$E_{xc}(n_{\uparrow}, n_{\downarrow}) = \int \epsilon_{xc}(n_{\uparrow}, n_{\downarrow}, \vec{\nabla} n_{\uparrow}, \vec{\nabla} n_{\downarrow}) n(\mathbf{r}) d\mathbf{r}. \quad (3.29)$$

Вообще говоря, GGA — это семейство методов, которые являются полужемпирическими. Например в методе PBE обменная энергия записывается как

$$E_x(n) = \int n(\mathbf{r}) \epsilon_x^{unif}(n) F_x(s) d\mathbf{r}, \quad (3.30)$$

где

$$\begin{aligned}
 \epsilon_x^{unif}(n) &= -\frac{3k_F}{4\pi}, \\
 k_F &= (3\pi^2 n)^{\frac{1}{3}}, \\
 F_x &= 1 + k - \frac{k}{1 + \frac{\mu s^2}{k}}, \\
 s &= \frac{|\nabla n|}{2k_F n}
 \end{aligned} \tag{3.31}$$

k и μ — некоторые эмпирические коэффициенты. Их значения равны, соответственно, 0.804 и 0.21951.

Другим вариантом полуэмпирического GGA приближения является функционал Беке:

$$E_x(n) = E_x^{LDA} - \gamma \int \frac{n^3 x^2}{1 + 6\gamma \text{sh}^{-1} x} d\mathbf{r}, \tag{3.32}$$

где $x = n^{\frac{4}{3}} |\nabla n|$, а γ — эмпирический параметр, подбираемый так, чтобы воспроизвести обменные энергии атомов инертных газов. ($\gamma = 0.0042$).

Обычно корреляционный потенциал вычисляется из значений электронной плотности на некоторой конечной сетке, соответственно градиент вычисляется с использованием конечно-разностных формул. Конечная сетка строится как набор сферических оболочек вокруг каждого атома, каждая из которых содержит набор точек для разностной схемы. Сетки могут быть равномерными и неравномерными (усеченными). В первом случае на каждой сфере содержится одинаковое число точек, во втором используется более редкая сетка в местах, где для достижения необходимой точности достаточно меньшего числа точек. То есть усеченные сетки строятся так, чтобы быть наиболее плотными в областях, в которых свойства меняются наиболее быстро.

Существуют и более точные приближения, которые в значительной степени позволяют решить проблему вычисления функционала обменно-корреляционной энергии. Локальные методы типа GGA и LDA не всегда дают точность, достаточную для химических вычислений. В настоящее время разрабатываются более точные полуэмпирические методы (meta GGA, MGGA).

В частности, одним из направлений улучшения метода является добавление нелокальной информации в подынтегральное выражение (3.29).

Формализм метода теории функционала плотности нарушается в условиях наличия векторного потенциала, в частности, в присутствии магнитного поля, что усложняет расчеты, связанные с магнитными резонансами. В этом случае по значению электронной плотности невозможно восстановить волновую функцию, т.к. последняя является комплексной величиной. Попытки обобщения формализма для учёта эффектов, связанных с магнитным полем, вылились в две разных теории: в теорию функционала плотности с учётом вектора плотности тока, и в теорию функционала плотности с учётом магнитного поля. В обоих случаях функционал обменно-корреляционной энергии обобщается и становится зависящим не только от электронной плотности. В первом подходе, развитом Vignale и Rasolt, помимо электронной плотности, аргументом функции полной энергии является ещё и плотность тока. Во втором подходе (Salsbury, Grayce, Harris) дополнительным аргументом функционала служит магнитное поле, и вид функционала зависит от вида магнитного поля. Для обоих методов вычисление обменно-корреляционной энергии за рамками приближения локальной плотности (вернее, его обобщения на случай магнитного поля) оказалось весьма сложным.

На практике метод Кона — Шэма может быть применён несколькими различными способами в зависимости от цели исследования. В расчётах для физики твёрдого тела до сих пор широко используется приближение локальной плотности вкупе с базисом плоских волн. Для расчётов электронной структуры молекул требуются более сложные выражения для функционалов. Так, большое число приближенных функционалов для расчёта обменно-корреляционного взаимодействия было развито для задач химии. Некоторые из них противоречат приближению пространственно однородного электронного газа, тем не менее, в пределе при переходе к электронному газу должны сводиться к приближению локальной плотности.

Для расчётов физических задач наиболее часто применяется уточнённая обменная модель Perdew—Burke—Ernzerhof, однако известно, что она приводит к ошибкам в калориметрических параметрах, будучи приложенной к расчётам молекул в газовой фазе.

В расчётах квантовой химии одним из распространённых является вид обменного функционала, называемый BLYP (Becke, Lee, Yang, Parr). Ещё более широко распространено приближение B3LYP, которое основано на гибридном функционале, в котором обменная энергия рассчитывается с привлечением точного результата, полученного методом Хартри — Фока.

В целом, текущее состояние метода теории функционала плотности таково, что невозможно оценить погрешность расчёта, не сравнивая его результаты с другими подходами или с результатами экспериментов.

4. Структура методов квантовой химии

Как уже было сказано выше, получить численное решение много-частичного уравнения Шредингера невозможно ввиду огромной вычислительной сложности задачи. Поэтому для решения задач вычислительной химии и физики твердого тела применяются различные приближения. Некоторые из таких приближений были рассмотрены в предыдущих главах.

Большинство *ab initio* квантово-механических расчетов опираются сразу на несколько методов и приближений. Например, все вычислительные методы используют приближение Борна-Оппенгеймера в качестве основы. Большинство основывается также на методе самосогласованного поля.

Рассмотрим уравнение Хартри—Фока. Введем набор базисных функций (например, атомных спин-орбиталей) $\chi_\mu(x)$. Тогда произвольную функцию можно записать в виде

$$\psi_i(x) = \sum_{\mu=1}^{\infty} C_{\mu i} \chi_\mu. \quad (4.1)$$

С некоторой точностью можно ограничиться конечным набором базисных функций. Тогда

$$\psi_i(x) \cong \sum_{\mu=1}^K C_{\mu i} \chi_\mu. \quad (4.2)$$

Подставляя это выражение в уравнение Хартри-Фока получим:

$$\hat{F} \left(\sum_{\mu=1}^K C_{\mu i} \chi_\mu \right) = \varepsilon_i \sum_{\mu=1}^K C_{\mu i} \chi_\mu. \quad (4.3)$$

Умножая (4.3) справа на χ_ν^* и интегрируя по всему пространству, получим:

$$\sum_{\mu=1}^K C_{\mu i} \int \chi_\nu^* \hat{F} \chi_\mu dx = \varepsilon_i \sum_{\mu=1}^K C_{\mu i} \int \chi_\nu^* \chi_\mu dx. \quad (4.4)$$

Введем матрицы:

$$\begin{aligned} F_{\nu\mu} &= \int \chi_\nu^* \hat{F} \chi_\mu dx \\ S_{\nu\mu} &= \int \chi_\nu^* \chi_\mu dx \end{aligned} \quad (4.5)$$

Эти матрицы называются молекулярными интегралами. Тогда уравнение (4.4) можно переписать в виде:

$$\sum_{\mu} F_{\nu\mu} C_{\mu i} = \varepsilon_i \sum_{\mu} S_{\nu\mu} C_{\mu i}, \quad (4.6)$$

или, для краткости:

$$\mathbf{FC} = \mathbf{SCE}. \quad (4.7)$$

Уравнение в таком виде называется уравнением Хартри-Фока-Рутана. На первый взгляд оно является уравнением на собственные вектора для матрицы \mathbf{F} в базисе, в котором \mathbf{S} диагональна. На самом деле это не так, поскольку матрица \mathbf{F} определяется оператором \hat{F} , который, в свою очередь, зависит от решения уравнения.

Определим выражение для $F_{\nu\mu}$. Запишем оператор Фока:

$$\hat{f} = \hat{h} + \sum_a \hat{\mathcal{J}}_a - \hat{\mathcal{K}}_a. \quad (4.8)$$

Подставляя это выражение в (4.5) получим:

$$\begin{aligned} F_{\mu\nu} &= \int \chi_\mu^*(x) \hat{f} \chi_\nu(x) dx = \int \chi_\mu^*(x) \hat{h} \chi_\nu(x) dx + \sum_a \int \chi_\mu^*(x) [\hat{\mathcal{J}}_a - \hat{\mathcal{K}}_a] \chi_\nu(x) dx = \\ &= H_{\mu\nu}^{core} + \sum_a (2 \langle \mu\nu | aa \rangle - \langle \mu a | a\nu \rangle). \end{aligned} \quad (4.9)$$

Молекулярные интегралы $H_{\mu\nu}^{core}$ определяют энергию взаимодействия электрона с ядрами и его кинетическую энергию. В свою очередь он состоит из интегралов кинетической энергии:

$$T_{\mu\nu} = -\frac{1}{2} \int \chi_\mu^* \Delta \chi_\nu dx \quad (4.10)$$

и интегралов ядерного притяжения:

$$V_{\mu\nu}^{nuc1} = - \int \chi_{\mu}^* \left[\sum_a \frac{Z_a}{r - R_a} \right] \chi_{\nu} dr. \quad (4.11)$$

В итоге:

$$H_{\mu\nu}^{core} = T_{\mu\nu} + V_{\mu\nu}^{nuc1}. \quad (4.12)$$

Стоит обратить внимание на то, что, в отличие от полной Фоковской матрицы $F_{\nu\mu}$, $H_{\mu\nu}^{core}$ зависит только от выбранного базиса и ее необходимо вычислить только один раз.

Рассмотрим теперь двухэлектронную часть оператора Фока. Запишем (4.9), подставив в него (4.12). Получим:

$$\begin{aligned} F_{\mu\nu} &= H_{\mu\nu}^{core} + \sum_a \sum_{\lambda\sigma} C_{\lambda a} C_{\sigma a}^* [2 \langle \mu\nu | \sigma\lambda \rangle - \langle \mu\lambda | \sigma\nu \rangle] = \\ &= H_{\mu\nu}^{core} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[\langle \mu\nu | \sigma\lambda \rangle - \frac{1}{2} \langle \mu\lambda | \sigma\nu \rangle \right] = H_{\mu\nu}^{core} + G_{\mu\nu} \end{aligned} \quad (4.13)$$

Именно такое выражение Фоковской матрицы используется для вычислений. Отметим, что его двухэлектронная часть состоит из двух элементов: матрицы плотности \mathbf{P} , определяемой решением, и двухэлектронными молекулярными интегралами:

$$\langle \mu\nu | \sigma\lambda \rangle = \int \chi_{\mu}^* \chi_{\nu} r_{12}^{-1} \chi_{\lambda}^* \chi_{\sigma} dr_1 dr_2. \quad (4.14)$$

Ввиду огромного количества этих интегралов, их вычисление является одной из наиболее вычислительно сложных операций в методе Хартри-Фока.

Ввиду того, что $F_{\nu\mu}$ само по себе зависит от $C_{\mu i}$, для решения уравнения Хартри-Фока-Рутана необходим итеративный метод самосогласования, подобный описанному в предыдущей главе для теории функционала плотности. Примерная процедура решения выглядит следующим образом:

1. Задание геометрии молекулы (набор координат ядер $\{\mathbf{R}_i\}$, их зарядов $\{Z_i\}$ и числа электронов N) и базиса χ_i .
2. Вычисление молекулярных интегралов $S_{\nu\mu}$, $H_{\mu\nu}^{core}$ и $\langle \mu\nu | \sigma\lambda \rangle$.
3. Диагонализация матрицы $S_{\nu\mu}$ (нахождение матрицы перехода \mathbf{X}).
4. Определение первого приближения матрицы плотности \mathbf{P} .
5. Вычисление матрицы $G_{\mu\nu}$ и сложение ее с $H_{\mu\nu}^{core}$ для определения матрицы Фока $F_{\nu\mu}$.

6. Вычисление преобразованной матрицы Фока $\mathbf{F}' = \mathbf{X}^+ \mathbf{F} \mathbf{X}$.
7. Вычисление \mathbf{C}' и \mathbf{E} и переход к исходному базису ($\mathbf{C} = \mathbf{X} \mathbf{C}'$).
8. Вычисление новой матрицы плотности и нормы ее разности с матрицей плотности предыдущей операции. Если норма больше, чем допустимое отклонение, то возврат к пункту 5 с новой матрицей плотности.

9. Определение интересующих значений с помощью вычисленных $C_{\mu i}$.

Рассмотрим теперь вопрос базиса. Выбор базиса является одним из важнейших факторов, влияющих на точность и скорость вычислений. Выбор правильного базиса является необходимым требованием для получения правильного результата расчета. Например, для двухатомных молекул можно использовать наборы атомных орбиталей обоих атомов.

Каждая атомная орбиталь может быть представлена в виде так называемых орбиталей слэтеровского типа (STO). Радиальная часть слэтеровской орбитали замисывается как

$$R_n(r) = r^{(n-1)} \exp(-\zeta r), \quad (4.15)$$

где n — эффективное квантовое число, ζ — так называемая орбитальная экспонента, выражаемая следующим образом:

$$\zeta = \frac{Z - s}{n}, \quad (4.16)$$

где Z — заряд ядра, s — константа экранирования. Полное выражение для орбитали слэтеровского типа можно записать в виде

$$\chi_{nlm}(\theta, \phi) = N Y_{lm}(\theta, \phi) R_n(r), \quad (4.17)$$

где $Y_{lm}(\theta, \phi)$ — сферическая гармоника, а N — коэффициент нормализации, выражаемый как

$$N = (2\zeta)^n \sqrt{\frac{2\zeta}{(2n)!}}. \quad (4.18)$$

Для многоатомных молекул использование STO в качестве базиса требует использования набора слэтеровских орбиталей, имеющих центры во всех атомах. Это приводит к тому, что вычисление молекуляр-

ных интегралов становится крайне затратным с вычислительной точки зрения: для трехатомных молекул необходимо вычислять интегралы функций с центрами в одной, двух и трех точках, что довольно неудобно. Для черытех- и более атомных молекул необходимо так же вычислять четырехцентровые интегралы.

Легко подсчитать, что для работы с базисом размера N необходимо вычислить $\frac{N^4}{8}$ интегралов вида $\langle \mu\nu | \sigma\lambda \rangle$. Вычисления молекул средних размеров обычно требуют базиса размером примерно от 20 до 1000 функций. Таким образом, необходимо вычислить от $2 \cdot 10^4$ до $1.25 \cdot 10^{11}$ интегралов, при этом вычисление трех- и четырехцентровых интегралов по STO требует достаточно большого количества времени.

Для того чтобы ускорить вычисления, можно использовать орбитали гауссова типа (GTO) вместо STO в качестве атомных орбиталей. Использование гауссовых атомных орбиталей было предложено Бойсом в 1950 году. В декартовых координатах Гауссова орбиталь может быть выражена в виде:

$$g_{ijk} = N(x - x_n)^i (y - y_n)^j (z - z_n)^k \exp(-\alpha |r - r_n|^2), \quad (4.19)$$

где координаты с индексом n принадлежат ядру, для которого строится орбиталь, а i, j, k — некоторые неотрицательные целые числа. Нормировочную константу N можно записать как:

$$N = \left(\frac{2\alpha}{\pi} \right)^{\frac{3}{4}} \left[\frac{(8\alpha)^{i+j+k} i! j! k!}{(2i)!(2j)!(2k)!} \right]^{\frac{1}{2}}. \quad (4.20)$$

Если $i + j + k = 0$, то орбиталь называется гауссовой орбиталью s-типа, при $i + j + k = 2$ орбиталь называется орбиталью p-типа.

Существует 6 орбиталей p-типа: $x_b^2, y_b^2, z_b^2, x_b y_b, x_b z_b, y_b z_b$. Удобно перейти к следующим их линейным комбинациям: $x_b y_b, x_b z_b, y_b z_b, x_b^2 - y_b^2, 3z_b^2 - r_b^2$. Такие линейные комбинации ведут себя подобно слэтеровским атомным орбиталям. Шестая комбинация ведет себя так же, как s-орбиталь и обычно опускается.

Подобные рассуждения можно провести и для орбиталей f-типа. Всего существует 10 гауссовых орбиталей этого типа из которых можно образовать 7 линейных комбинаций, ведущих себя так же, как реальные атомные орбитали f-типа.

Отметим, что в выражении для GTO отсутствует главное квантовое число n . Орбиталь с некоторым главным квантовым числом может быть представлена в виде линейной комбинации GTO с различными коэффициентами α .

Одна из проблем, связанная с использованием GTO — их неправильное поведение около атомного ядра. Поэтому для того, чтобы получить хорошее представление атомной орбитали необходимо использовать линейную комбинацию нескольких GTO. Из-за этого вычисление молекулярных интегралов для гауссовых орбиталей требует интегрирования значительно большего числа функций. Несмотря на это, вычисление интегралов от GTO значительно проще, поэтому использование их вместо STO значительно ускоряет расчет. По этой причине большинство современных квантовохимических расчетных программ использует тот или иной набор гауссовых орбиталей.

В качестве примера можно привести часто используемые базисы STO- n G. Это минимальные базисы, состоящие из линейных комбинаций n гауссовых орбиталей, которыми аппроксимируется орбиталь слэтеровского типа. Под минимальным набором понимается набор орбиталей, достаточный для расположения всех электронов в нейтральном атоме (1 для водорода, 2 для гелия и т.п.). Впервые эти базисы были предложены Джоном Поплом.

Обычно используются базисы с n от 2 до 6. Наиболее широко используется STO-3G. Для сравнения точности аппроксимации в таблице 2 приведены энергии $1s$ электрона в атоме водорода. Точное значение энергии составляет -0.5 Хартри.

Таблица 2. Сравнение энергий $1s$ электрона в атоме водорода для различных аппроксимаций STO- n G

Базис	Энергия, Хартри
STO-3G	-0.49491
STO-4G	-0.49848
STO-5G	-0.49951
STO-6G	-0.49983

Ввиду важности валентных электронов, часто разумно использовать больше одной базисной функции для валентных орбиталей, каждая из

которых, в свою очередь, образована линейной комбинацией ГТО. Часто для обозначения таких базисов используется нотация Попла. В этой нотации обозначение базиса состоит из числа, дефиса и набора нескольких чисел. Первое число является количеством ГТО, используемых для построения базиса. Числа, следующие за дефисом, указывают на то, из скольких ГТО образована каждая валентная базисная функция.

Например, запись 6-311G означает, что для формирования базиса используется 6 гауссовых орбиталей, а валентные орбитали составляются из трех базисных функций, образованных, соответственно, комбинациями трех, одной и одной ГТО.

Рассмотрим основные методы, применяемые в квантовой химии. Метод Хартри—Фока использует несколько упрощений:

1. Приближение Борна—Оппенгеймера: в реальности же волновая функция зависит не только от электронных координат, но и от ядерных;
2. Обычно не учитываются релятивистские эффекты;
3. Используется конечный набор базисных функций: базис, из которого образуется решение, неполон;
4. Полная волновая функция вычисляется как линейная комбинация произведений одноэлектронных волновых функций (спин-орбиталей): не учитываются эффекты электронной корреляции (кроме антисимметричности волновой функции, вводимой искусственно с помощью детерминанта Слэтера).

Из всех этих предположений обычно наиболее существенным является четвертое. Для того чтобы учесть электронные корреляции существует множество методов, которые обычно называют пост-Хартри—Фоковскими. Учет эффектов электронной корреляции позволяет получать более точные решения, но при этом вычислительные методы, позволяющие это делать, имеют более высокую вычислительную сложность.

На рис. 4.1 показана иерархия некоторых методов квантовой химии. Методы UHF и ROHF являются модификациями метода Хартри—Фока, позволяющие производить расчеты молекул с открытой (не до конца заполненной, занимаемой только одним электроном) электронной оболочкой.

Метод ROHF (restricted open-shell Hartree-Fock, ограниченный метод Хартри—Фока для открытых оболочек). В этом методе сначала двумя электронами заполняются все нижние орбитали настолько, насколько это возможно. Затем заполняются орбитали, содержащие один электрон.

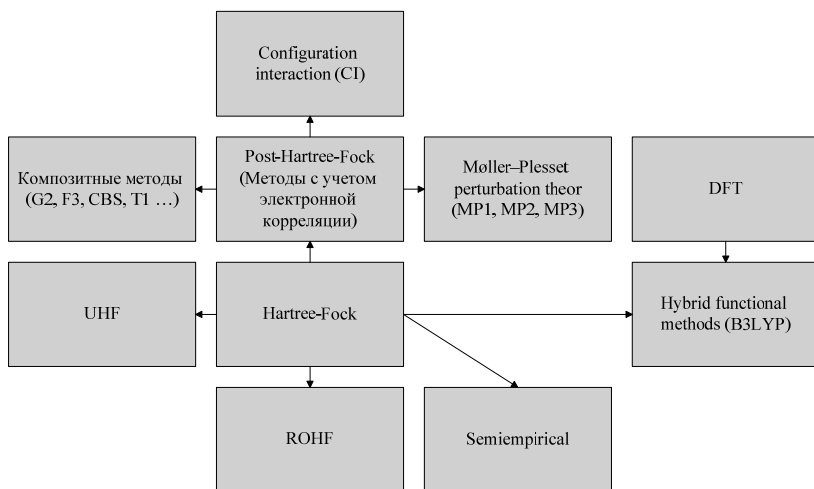


Рис. 4.1. Иерархия некоторых методов квантовой химии

Впервые этот метод был предложен Рутаном. Вид уравнения Хартри-Фока-Рутана в этом методе такой же, как и для обычного метода Хартри—Фока:

$$FC = SCE, \quad (4.21)$$

Но, в отличие от него, фокиан F определен неоднозначно. Различный выбор этой матрицы будет приводить к различным орбиталям и энергиям орбиталей, но к одним и тем же наблюдаемым величинам, значим как электронная плотность, полная энергия и т.п.

Метод *UHF (unrestricted Hartree-Fock, неограниченный метод Хартри-Фока)* используется для молекул с открытыми оболочками значительно чаще, чем ROHF. Идея метода заключается в том, чтобы разделить орбитали, занимаемые электронами с разными спинами (со спиновыми частями спин орбиталей α и β). Этот метод также называется DODS (different orbitals for different spins).

Применение этого метода дает два связанных уравнения Рутана, называемых уравнениями Поппла—Нэсбита:

$$\begin{aligned} \mathbf{F}^{\alpha} \mathbf{C}^{\alpha} &= \mathbf{S} \mathbf{C}^{\alpha} \mathbf{E} \\ \mathbf{F}^{\beta} \mathbf{C}^{\beta} &= \mathbf{S} \mathbf{C}^{\beta} \mathbf{E} \end{aligned} \quad (4.22)$$

где \mathbf{F}^{α} и \mathbf{F}^{β} — матрицы для орбиталей с различными спинами. Соответственно их решение даст два набора спин-орбиталей, соответствующих разным спинам.

Недостатком этого метода, по сравнению с ROHF, является то, что результирующий детерминант Слэтера не является собственной функцией оператора полного спина $\hat{\mathbf{S}}^2$. Это приводит к тому, что в основное состояние подмешиваются возбужденные состояния. Среднее значение спина становится выше того, которое должно было бы быть в основном состоянии. Поэтому при вычислениях этим методом следует следить за средним значением спина. Сильное отклонение будет означать, что метод не пригоден в данной ситуации.

Несмотря на этот недостаток, UHF используется чаще ROHF, поскольку он проще и результатом его работы будет однозначный набор молекулярных спин-орбиталей.

Метод CI (конфигурационного взаимодействия), в отличие от HF, использует в качестве решений линейные комбинации детерминантов Слэтера, называемых в данном случае CSF (configuration state functions):

$$\Psi = \sum_{i=0} c_i \Psi_i^{SO}, \quad (4.23)$$

где Ψ_i^{SO} — детерминанты Слэтера.

Если (4.23) включает все возможные CSF, то в результате решения соответствующих уравнений будет получено точное решение уравнения Шредингера. Количество функций конфигурационных состояний является основной характеристикой метода. Обычно первый член в разложении (4.23) является детерминантом Слэтера, получаемым при решении Хартри—Фока. Остальные функции определяются количеством спин-орбиталей, замененных на незанятые спин-орбитали.

Если заменена только одна орбиталь, то такая CSF называется однократно возбужденной и т.п. Ограничение количества CSF является критическим для ускорения расчетов. В основном используются только однократно и двукратно возбужденные функции. Обычно метод, использующий только двукратно возбужденные функции, называется CID, а использующий однократно и двукратно возбужденные функции — CISD. Метод CISD занижает энергию конфигурационного взаимодействия. Су-

существует простой метод коррекции для примерного учета вклада более возбужденных состояний, называемый поправкой Дэвидсона:

$$E_{corrected} \cong E_{CISD} + (1 - a_0^2)(E_{CISD} - E_{HF}), \quad (4.24)$$

где a_0 — коэффициент перед волновой функцией, являющейся решением уравнения Хартри—Фока в выражении (4.23).

Одной из основных проблем использования ограниченного пространства функций конфигурационных состояний заключается в нарушении строгой отделимости. То есть энергия системы двух удаленных на бесконечное расстояние частиц не будет равной сумме энергий отдельных частиц.

Большинство квантово-механических программ реализуют методы CID и CISD. Метод конфигурационного взаимодействия крайне сложен вычислительно, поэтому его применение ограничено относительно малыми системами.

Теория возмущений Моллера—Плессета является пост-хартри-фоковским методом, основанным на учете электронных корреляций с помощью теории возмущений Релея—Шредингера. Обычно используется второй (MP2), третий (MP3) и четвертый (MP4) порядок теории.

В MP-теории в качестве невозмущенного оператора используется оператор Фока \hat{F} . Соответственно возмущением является корреляционный потенциал.

Полный гамильтониан системы может быть записан в виде:

$$\hat{H} = \hat{H}_0 + \lambda \hat{V}. \quad (4.25)$$

Тогда, согласно теории возмущений, собственные функции и значения для оператора \hat{H} могут быть записаны в виде рядов:

$$\begin{aligned} \Psi &= \sum_{i=0}^{\infty} \lambda_i \Psi_i \\ E &= \sum_{i=0}^{\infty} \lambda_i E_i \end{aligned} \quad (4.26)$$

Корреляционный потенциал \hat{V} можно записать в виде:

$$\hat{V} = \hat{H} - \left(\hat{F} + \left\langle \Psi_0 \right| \hat{H} - \hat{F} \right| \Psi_0 \right), \quad (4.27)$$

где Ψ_0 — детерминант Слэтера, являющийся решением уравнения Хартри-Фока. В качестве невозмущенного оператора \hat{H}_0 используется смещенный оператор Фока $\hat{F} + \langle \Psi_0 | \hat{H} - \hat{F} | \Psi_0 \rangle$.

Согласно теореме Моллера—Плессета, вклад корреляционного возмущения в первом порядке равен нулю. Согласно теории возмущения, поправка во втором порядке запишется в виде:

$$E_{MP2} = \sum_{i,j,a,b} \langle \varphi_i(1) \varphi_j(2) | r_{12}^{-1} | \varphi_a(1) \varphi_b(2) \rangle \times \\ \times \frac{2 \langle \varphi_a(1) \varphi_b(2) | r_{12}^{-1} | \varphi_i(1) \varphi_j(2) \rangle - \langle \varphi_a(1) \varphi_b(2) | r_{12}^{-1} | \varphi_j(1) \varphi_i(2) \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b}, \quad (4.28)$$

где φ_i и φ_j — занятые орбитали, φ_a и φ_b — виртуальные орбитали, а $\varepsilon_i, \varepsilon_j, \varepsilon_a, \varepsilon_b$ — соответствующие энергии. Полная энергия во втором порядке точности запишется как:

$$E \cong E_{HF} + E_{MP2}. \quad (4.29)$$

Методы MP2, MP3 и MP4 являются часто используемыми и реализованы в большинстве вычислительных пакетов. Иногда используется также метод MP5, но обычно его использование не оправдано из-за вычислительной сложности.

Известно, что метод Моллера—Плессета, вообще говоря, не всегда является сходящимся в высоких порядках. Даже если метод сходится, предсказать характер сходимости крайне затруднительно ввиду очень сложной зависимости от рассчитываемой системы и базиса. Поэтому для некоторых физических величин порядки MP3 и MP4 не всегда более точны, чем MP2.

Для молекул с незаполненной электронной оболочкой метод Моллера—Плессета может быть непосредственно применен к решению, полученному методом UHF (применение для ROHF проблематично ввиду того, что его электронные состояния не всегда являются собственными для оператора Фока). Однако такой подход часто приводит к сильному искусственному перемешиванию спиновых функций, и, в свою очередь, к большим погрешностям.

Для применения теории возмущения в подходе ROHF также были созданы методы, подобные MPn. К сожалению, ввиду неоднозначности фокиана и волновых функций, таких методов существует очень большое количество, в частности ROHF-MBPT2, OPT1, OPT2, ZAPT, ROMP, HCPT и множество других.

Некоторые такие методы также подвержены погрешностям, связанным с перемешиванием спиновых частей спин-орбиталей.

Полуэмпирические методы обычно опираются на формализм Хартри-Фока, но при этом применяют множество приближений, основанных на экспериментальных данных. Одним из простейших полуэмпирических методов является метод Хюккеля. Это довольно простой метод, используемый для определения энергий молекулярных орбиталей π -электронов для углеводородных систем.

Этот метод является методом LCAO MO, то есть молекулярные орбитали строятся как линейная комбинация атомных орбиталей.

Например, для этилена решение строится в виде линейной комбинации 2π орбиталей углерода:

$$\Psi = c_1\psi_1 + c_2\psi_2. \quad (4.30)$$

Подставляя это значение в уравнение Шредингера получим:

$$c_1\hat{H}\psi_1 + c_2\hat{H}\psi_2 = Ec_1\psi_1 + Ec_2\psi_2. \quad (4.31)$$

Умножая это уравнение на ψ_1^* и интегрируя получим:

$$\begin{aligned} c_1(H_{11} - ES_{11}) + c_2(H_{12} - ES_{12}) &= 0 \\ c_1(H_{21} - ES_{21}) + c_2(H_{22} - ES_{22}) &= 0 \end{aligned} \quad (4.32)$$

где

$$\begin{aligned} H_{ij} &= \int \psi_i^* \hat{H} \psi_j dx \\ S_{ij} &= \int \psi_i^* \psi_j dx \end{aligned} \quad (4.33)$$

Метод предполагает, что $H_{11} = H_{22} = \alpha$ и $H_{12} = H_{21} = \beta$, а так же, что матрица S единична. Тогда (4.32) можно переписать в виде:

$$\begin{aligned} c_1(\alpha - E) + c_2\beta &= 0, \\ c_1\beta + c_2(\alpha - E) &= 0. \end{aligned} \quad (4.34)$$

Эта система имеет нетривиальные решения только при $E = \alpha \pm \beta$.

При этом $\Psi = \frac{1}{\sqrt{2}}(\psi_1 \pm \psi_2)$.

Естественно этот метод имеет крайне ограниченную применимость, но он крайне хорошо демонстрирует подход.

Методы гибридных функционалов — это методы аппроксимации обменно-корреляционного функционала теории функционала электронной плотности. Эти методы используют сумму обменной энергии, вычисленной по методу Хартри-Фока и обменно-корреляционной энергии, полученной каким-либо другим способом (это может быть *ab initio* метод типа LDA или какой-либо эмпирический метод).

Функционал обменной энергии в этом случае выражается через орбитали Кона-Шэма, а не через электронную плотность, поэтому такой функционал называется неявным.

Впервые такой подход был применен Акселем Беке в 1993 году. Использование решений, полученных методом Хартри—Фока позволяет улучшить точность вычисления многих параметров, которые плохо описываются *ab initio* функционалами ввиду своей нелокальности. Примерами таких величин являются колебательные частоты, геометрия и энергия атомизации.

Обычно такие функционалы строятся как линейная комбинация обменного функционала Хартри—Фока E_x^{HF} и какого-либо набора обменно-корреляционных функционалов, оперирующих электронной плотностью. Значения констант в этих линейных комбинациях обычно подбираются по экспериментальным данным. Таким образом, эти методы являются полуэмпирическими.

Наиболее часто используемым функционалом этого типа является B3LYP (трехпараметрический функционал Беке, Ли-Янга-Парра). Он определяется следующим выражением:

$$E_{xc}^{B3LYP} = E_{xc}^{LDA} + a_0(E_x^{HF} - E_x^{LDA}) + a_x(E_x^{GGA} - E_x^{LDA}) + a_c(E_c^{GGA} - E_c^{LDA}), \quad (4.35)$$

где $a_0 = 0.2$, $a_x = 0.72$, $a_c = 0.81$, E_x^{GGA} — функционал Беке, E_c^{GGA} — корреляционный функционал Ли, Янга и Парра.

Композитные методы используют результаты нескольких вычислений для повышения точности. Обычно используется комбинация вычислительно сложных подходов с использованием малого набора базисных функций и более простых подходов с большим базисом. Обычно эти методы используются для вычисления термодинамических параметров, таких, как энергии ионизации, сродство к электрону, энергия атомизации, энтальпия образования и т.п. Погрешность вычисления этих величин с помощью композитных методов может быть меньше 5 кДж/моль.

Первый подобный метод, Gaussian-1 (G1) был предложен Джоном Поплом. После был предложен широко используемый Gaussian-2 (G2) и Gaussian-3 (G3).

Метод G2 использует значения, полученные семью различными методами:

1. Сначала производится оптимизация геометрии с помощью метода MP2 и базиса 6-31F(d), причем учитывается возмущение для всех электронов. Полученная геометрия используется для всех последующих расчетов.
2. Используется метод QCI (модификация CI, решающая проблему нарушения строгой отделимости) с учетом тройных возбуждений QCISD(T) с базисом 6-311G(d). Также вычисляются энергии для MP2 и MP4.
3. MP4 с базисом 6-311G(2df, p).
4. MP4 с базисом 6-311+G(d, p).
5. MP2 с базисом 6-311+G(3df, 2p).
6. Производится оптимизация геометрии с помощью метода Хартри-Фока с базисом 6-31G(d)
7. На геометрии, полученной в 6 производится вычисление колебательных мод, для оценки энергий нулевых колебаний (ZPVE).

Также вычисляется эмпирическая поправка, равная:

$$HLC = -0.0048N_p - 0.00019N_u, \quad (4.36)$$

где N_p и N_u — количество парных и непарных валентных электронов. Итоговая энергия вычисляется как

$$E = E_{QCISD(T)} + (E_{MP4,3} - E_{MP4,2}) + (E_{MP4,4} - E_{MP4,2}) + (E_{MP2,5} + E_{MP2,2} - E_{MP2,3} - E_{MP2,4}) + 0.8929ZPVE + HLC \quad (4.37)$$

Для некоторых молекул также добавляется энергия спин-орбитального взаимодействия. Существует также несколько упрощенный этого метода, не использующих часть шагов и использующих DFT методы, такие как B3LYP, для оптимизации геометрии на шаге 1.

G3 похож на G2. Основным отличием является использование базиса 6-31G вместо 6-311G и на этапе 7 рассчитываются поправки для всех электронов, а не только для валентных. Также эмпирическая поправка HLC использует другие константы.

Совсем недавно был предложен метод G4, который также отличается используемыми базисами и использует теорию функционала плотности на некоторых шагах.

Альтернативными методами являются correlation consistent Composite Approach (ccCA) и набор complete basis set (CBS) методов.

Весьма простым и при этом достаточно точным методом является T1. Этот метод полезен для вычислений для молекул, состоящих из следующих элементов: H, C, N, O, F, S, Cl и Br. Метод состоит из следующих шагов:

1. Оптимизация геометрии с помощью HF на базисе 6-31G*.
2. Вычисление энергии методом RI-MP2 на базисе 6-311+G(2d, p) [6-311G*].
3. Набор эмпирических поправок на число атомов и порядок связей.

Метод дает результат, сходный с G3, но быстрее на 3 порядка. При сравнении вычисляемых результатов с экспериментальными данными по органическим молекулам из базы данных NIST среднеквадратичное отклонение составляет примерно 11 Дж/моль.

Обычно чем точнее метод, тем больше вычислительных ресурсов требуется для расчета с его использованием. Самыми грубыми и быстрыми методами являются методы молекулярной динамики. Они используют классические законы движения (закон Ньютона) для анализа поведения молекулярных систем. Фактически, молекулярная динамика представляет молекулярную систему в виде набора взаимодействующих точечных частиц. Для больших систем типа белков они часто оказываются единственным возможным методом расчета ввиду огромной размерности задачи.

Ввиду существования огромного количества методов в вычислительной химии иногда бывает довольно сложно выбрать правильную модель, но часто это оказывается критичным для решения задачи. Поэтому крайне важно понимать, как работают те или иные методы и какими преимуществами и недостатками они обладают.

5. Методы расчета электронных спектров твердых тел

В предыдущих главах, говоря о разных вычислительных методах, мы имели в виду в основном их приложения к расчету характеристик атомов и молекул. Однако область их применимости значительно шире и простирается от отдельных атомов до макроскопических тел. На первый взгляд может показаться, что последовательное увеличение числа частиц в системе только усложняет задачу определения электронного спектра. Это, однако, не совсем так.

Идеальные кристаллы представляют собой структуры, инвариантные относительно параллельного переноса системы координат на вектор $n\vec{a}_1 + l\vec{a}_2 + m\vec{a}_3$, где n, l, m — целые числа, а векторы \vec{a}_1, \vec{a}_2 и \vec{a}_3 называются векторами решетки, они определяются типом кристаллической структуры. Именно эта инвариантность накладывает свои ограничения на вид волновой функции электрона в кристалле (теорема Блоха), благодаря чему вычисление матричных элементов гамильтониана по этим функциям (а значит, и электронных спектров твердых тел) становится не такой уж сложной задачей.

В данном разделе мы не будем касаться обоснования адиабатического приближения и приближения самосогласованного поля в кристаллических телах — читатель, ознакомившийся с этим материалом на примере атомных систем, без труда самостоятельно разберет данный вопрос по литературе. Более того, мы не будем касаться вида функции потенциальной энергии электрона в твердом теле — нам будет важно только то, что она обладает пространственной периодичностью.

5.1. Приближение сильной связи

Задача приближения сильной связи ставится следующим образом: зная электронную волновую функцию изолированного атома, найти закон дисперсии электрона в кристалле $\varepsilon_l(p)$, где l — номер зоны, p — импульс электрона в периодическом потенциале. В этом приближении предполагается, что одноэлектронные волновые функции в кристалле являются сильно локализованными около «собственных» атомов и быстро убывают при отдалении от них.

Рассмотрим сначала одномерную цепочку атомов. Будем считать, что электронные оболочки мало перекрываются, а в нулевом приближении каждый электрон «принадлежит своему атому».

Потенциал кристалла имеет вид

$$V(x) = \sum_n U(x - na). \quad (5.1)$$

Запишем уравнение Шредингера для электрона

$$-\frac{\hbar^2}{2m} \frac{d^2\psi}{dx^2} + V(x)\psi(x) = \varepsilon\psi(x). \quad (5.2)$$

Точными решениями этого уравнения будут трансляционно-инвариантные блоховские функции $\psi_p(x) = e^{ipx/\hbar} u_p(x)$, где $u_p(x)$ — периодическая функция с периодом решетки. Блоховские функции для этой задачи неудобны, поскольку «размазаны» по всему кристаллу, однако можно составить из них комбинации следующего типа:

$$w_n(x) = \frac{1}{\sqrt{N}} \sum_p e^{-ipna/\hbar} \psi_p(x) = \frac{1}{\sqrt{N}} \sum_p e^{ip(x-na)/\hbar} u_p(x). \quad (5.3)$$

Это так называемые функции Ванье, суммирование по p производится в первой зоне Бриллюэна $-\pi\hbar/a \leq p \leq \pi\hbar/a$, N — число атомов в цепочке. Как можно видеть, $w_n(x)$ отличны от нуля лишь вблизи n -ого иона.

Обратное преобразование, выражающее блоховские функции через функции Ванье, имеет вид:

$$\psi_p(x) = \frac{1}{\sqrt{N}} \sum_n e^{ipna/\hbar} w_n(x). \quad (5.4)$$

Функции Ванье с разными n ортогональны и обладают интересным свойством: из-за периодичности $u_p(x)$ все w_n являются по сути одной и той же функцией $w_0(x) \equiv w(x)$ со сдвинутым аргументом $w_n(x) \equiv w(x - na)$. Описанные свойства функций Ванье выявляют некоторую аналогию их с локализованными электронными функциями отдельных атомов, однако это не совсем одно и то же.

Подставив в уравнение Шредингера представление (5.4) для ψ_p , получим

$$\sum_n \left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + V(x) \right] e^{ipna/\hbar} w_n(x) = \varepsilon(p) \sum_n e^{ipna/\hbar} w_n(x). \quad (5.5)$$

Выделим из потенциальной энергии $V(x)$ слагаемое, отвечающее потенциалу n -го иона, и подставим это представление в уравнение Шредингера:

$$V(x) = U(x - na) + h(x),$$

$$\sum_n \left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x - na) \right] e^{ipna/\hbar} w_n(x) + \sum_n h(x) e^{ipna/\hbar} w_n(x) = \varepsilon(p) \sum_n e^{ipna/\hbar} w_n(x) \quad (5.6)$$

Слагаемое с $h(x)$ будет мало в силу локализации функций Ванье около «своих» атомов, его можно рассматривать как малое возмущение. В нулевом приближении, пренебрегая этим членом, мы видим, что решением уравнения (5.6) будет волновая функция электрона в изолированном атоме, $w^{(0)} = \varphi(x)$. Соответствующая энергия совпадет с уровнем энергии изолированного атома $\varepsilon^{(0)}(p) = \varepsilon_0$.

Найдем приближение следующего порядка. Для поиска первой поправки к энергии по теории возмущений мы могли бы воспользоваться стандартной формулой квантовой механики:

$$\varepsilon(p) = \varepsilon_0 + \left\langle \sum_n e^{ipna/\hbar} w_0(x) \left| h(x) \right| \sum_n e^{ipna/\hbar} w_0(x) \right\rangle - \quad (5.7)$$

первой поправкой к энергии является просто диагональный матричный элемент гамильтониана с «вырезанным» центральным атомом по трансляционно-инвариантной комбинации атомных функций. Этим приближением мы будем пользоваться в дальнейшем, например, при расчете электронного спектра графена. Но приближение сильной связи позволяет не ограничиваться первой поправкой, а получить более точную формулу для энергии. Подставив в уравнение Шредингера (5.6) волновую функцию в виде $w = w^{(0)} + w^{(1)}$ и энергию $\varepsilon(p) = \varepsilon_0 + (\varepsilon(p) - \varepsilon_0)$, а затем сократив члены нулевого порядка, мы получим

$$\sum_n \left[-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} + U(x - na) - \varepsilon(p) \right] e^{ipna/\hbar} w_n^{(1)}(x) = [\varepsilon(p) - \varepsilon_0 - h(x)] \sum_n e^{ipna/\hbar} w_n^{(0)}(x). \quad (5.8)$$

Это линейное неоднородное уравнение относительно $w^{(1)}$. Условием его разрешимости является ортогональность правой части и решения однородного уравнения с теми же граничными условиями (теоре-

ма Фредгольма). Умножим скалярно правую часть на $(w_k^{(0)})^*$ и проинтегрируем от $-\infty$ до $+\infty$. Из условия ортогональности получаем

$$\varepsilon(p) = \varepsilon_0 + \frac{\sum_n H_n e^{ipna/\hbar}}{\sum_n I_n e^{ipna/\hbar}}, \quad (5.9)$$

где H и I — интегралы перекрытия:

$$H_n = \int \varphi^*(x) h(x) \varphi(x - na) dx, \\ I_n = \int \varphi^*(x) \varphi(x - na) dx. \quad (5.10)$$

Поскольку $h(x)$ и $\varphi(x)$ быстро убывают с расстоянием, то при суммировании в формуле (5.9) обычно ограничиваются несколькими ближайшими соседями. Отсюда видно, что $\varepsilon(p)$ — периодическая функция с периодом $2\pi\hbar/a$.

В случае объемных тел вычисления становятся сложнее. Дополнительные сложности возникают, если на одну элементарную ячейку приходится более одного атома, в этом случае надо искать преобразования, относительно которых кристалл остается симметричным. Также атомное состояние может не быть сферически симметричным, т.е. соответствовать некоторому ненулевому моменту, уровень энергии ε_0 будет вырожден и расчет значительно усложнится. В простейшем случае, когда на элементарную ячейку приходится один атом, содержащий только один s-электрон, обобщение (5.9) имеет следующий вид:

$$\varepsilon(p) = \varepsilon_0 + \frac{\sum_n H_n e^{ipna/\hbar}}{\sum_n I_n e^{ipna/\hbar}}. \quad (5.11)$$

В частности, для объемно-центрированной кубической решетки в приближении ближайших соседей (их координаты $(\pm a/2, \pm a/2, \pm a/2)$, а — сторона куба) мы имеем закон дисперсии в виде

$$\varepsilon(p) = \varepsilon_0 + H_0 + 8(H_1 - H_0 I_1) \cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_z a}{2}\right). \quad (5.12)$$

Одним из основных результатов приближения сильной связи является расщепление отдельных атомных энергетических уровней в зоны разрешенных значений энергии. Этим качественным выводом полезность метода не ограничивается — метод позволяет сделать вывод об энергетических спектрах электронов в кристаллах на основании знания одного лишь типа симметрии кристалла. Вычисление матричных элементов требует знания атомных волновых функций и потенциалов отдельных атомов, поэтому значения констант в приближении сильной связи зачастую берутся из эксперимента.

5.2. Электронный спектр графитового слоя

Развитая в предыдущем параграфе теория сильной связи находит свое непосредственное применение в расчете электронных спектров новых материалов на основе углерода — однослойного и двухслойного графена и углеродных нанотрубок. Экспериментальные данные о спектрах таких материалов находятся в хорошем согласии с результатами расчетов в данном приближении (форма кривых воспроизводится хорошо, а на точное вычисление констант метод сильной связи и не претендует). Задача о вычислении спектров графена и нанотрубок является актуальной, т.к. именно необычные свойства законов дисперсии носителей заряда в этих материалах определяют характеристики нанoeлектронных приборов на их основе.

Двумерный материал графен привлек внимание исследователей, занимающихся современной электроникой, своими необычными транспортными свойствами. Теоретически было предсказано, что подвижность электронов в графене может достигать $2 \cdot 10^6 \text{ см}^2/(\text{В} \cdot \text{с})$, что превышает подвижность электронов в кремнии в 1500 раз. Реально в эксперименте наблюдались значения подвижности в 10 раз меньшие ($230\,000 \text{ см}^2/(\text{В} \cdot \text{с})$, [4]), тем не менее, это внушительная цифра по сравнению со значениями для полупроводниковой электроники. Высокую электронную подвижность связывают с необычным линейным спектром электронов в графене (в отличие от квадратичного спектра в полупроводниках). Исследованием этого спектра мы и займемся.

В некоторых научных работах *линейный энергетический спектр* электронов в графене преподносится как нечто необычное, свойственное лишь этому материалу. Теперь, познакомившись с различными подходами к изучению электронных спектров кристаллических тел, мы можем сделать некоторые выводы о поведении электронов в графене, основываясь лишь на знании его симметрии. Данные вычисле-

ния были проделаны еще в 1949 году британским физиком П. Вэллимсом, занимавшимся проблемами электронных свойств графита. Будучи знакомым с кристаллической структурой графита (атомные плоскости, состоящие из шестигранников со стороной $a/\sqrt{3} = 0,14$ нм, расположены параллельно друг другу на расстоянии 0,67 нм), он предложил сначала рассчитать закон дисперсии электрона, движущегося по отдельной плоскости, а затем учесть слабую связь между плоскостями.

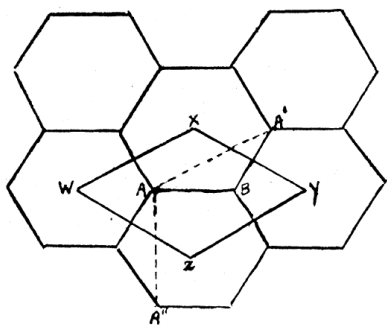


Рис. 5.1. Элементарная ячейка графитовой плоскости WXYZ и векторы трансляций AA' и AA''

вой плоскости, они же дают основной вклад в проводимость¹.

Элементарная ячейка графитовой плоскости содержит 2 атома А и В (Рис. 5.1), они являются неэквивалентными, т.к. не могут быть совмещены параллельным переносом — для совмещения требуется еще и зеркальное отражение. Именно наличие более чем одного атома в элементарной ячейке не позволяет нам напрямую применить формулу (5.11) приближения сильной связи, в связи с чем процедуру поиска энергетического спектра придется повторить, слегка модифицировав. Внимательный читатель заметит, что выполняемая процедура во многом напоминает метод ЛКАО, рассмотренный в предыдущем разделе.

В качестве базисных волновых функций в приближении сильной связи мы выберем волновые функции $2p_z$ -электронов, локализованных у атомов А и В, обозначим эти функции как $X_A(r)$ и $X_B(r)$.

При формировании графитовой плоскости три валентных электрона углерода принимают участие в образовании sp^2 -гибридных орбиталей, они ответственны за прочные σ -связи в плоскости. Четвертый электрон атома углерода не затронут гибридизацией, он находится в состоянии $2p_z$ и слабо связан со своим атомом. Именно электроны в $2p_z$ -состоянии делокализуются при формировании графитовой

¹ $2p_z$ -электроны также отвечают за формирование «полуторных» связей в ароматическом кольце бензола C_6H_6 .

Волновая функция электрона, удовлетворяющая условию трансляционной инвариантности, имеет вид:

$$\psi = \sum_A X_A(\mathbf{r} - \mathbf{r}_A) \exp(i\mathbf{k}\mathbf{r}_A) + \lambda \sum_B X_B(\mathbf{r} - \mathbf{r}_B) \exp(i\mathbf{k}\mathbf{r}_B) = \psi_1 + \lambda \psi_2; \quad (5.13)$$

в первом слагаемом ведется суммирование по всем атомам типа А, во втором — по атомам типа В. Константа λ будет найдена нами позднее после определения энергии из вариационного принципа. Введенная таким образом волновая функция еще не обязана быть собственной для уравнения Шредингера и отвечать определенному значению энергии; для вычисления собственных значений подставим волновую функцию (5.13) в уравнение Шредингера $\hat{H}\psi = E\psi$ и спроецируем его последовательно на ψ_1 и ψ_2 :

$$\begin{cases} H_{11} + \lambda H_{12} = ES, \\ H_{21} + \lambda H_{22} = \lambda ES, \end{cases} \quad (5.14)$$

где H_{ij} есть матричные элементы гамильтониана по отношению к линейно волновым функциям ψ_1 и ψ_2 :

$$\begin{aligned} H_{ij} &= \int \psi_i^*(\mathbf{r}) \hat{H} \psi_j(\mathbf{r}) d^3\mathbf{r}, \\ S &= \int \psi_1^*(\mathbf{r}) \psi_1(\mathbf{r}) d^3\mathbf{r} = \int \psi_2^*(\mathbf{r}) \psi_2(\mathbf{r}) d^3\mathbf{r}. \end{aligned} \quad (5.15)$$

Применяя очевидные симметрии $H_{11} = H_{22}$ и $H_{21} = H_{12}^*$, сразу можно выписать выражение для энергии:

$$E = \frac{1}{S} \{ H_{11} \pm |H_{12}| \}. \quad (5.16)$$

Для получения закона дисперсии, то есть зависимости $E(\mathbf{k})$, необходимо вычислить матричные элементы гамильтониана по отношению к блоховским функциям в зависимости от \mathbf{k} . Воспользуемся для этого приближением «ближайших соседей»: в выражении для матричных элементов гамильтониана мы положим

$$\begin{aligned} \int X_A^*(\mathbf{r} - \mathbf{r}_A) \hat{H} X_A(\mathbf{r} - \mathbf{r}_{A'}) d^3\mathbf{r} &= 0 \quad \text{при } |\mathbf{r}_A - \mathbf{r}_{A'}| > a, \\ \int X_B^*(\mathbf{r} - \mathbf{r}_B) \hat{H} X_A(\mathbf{r} - \mathbf{r}_A) d^3\mathbf{r} &= 0 \quad \text{при } |\mathbf{r}_A - \mathbf{r}_B| > a/\sqrt{3}; \end{aligned} \quad (5.17)$$

при вычислении же скалярных произведений волновых функций достаточно положить

$$\int X_A^*(\mathbf{r}-\mathbf{r}_A)X_{A'}(\mathbf{r}-\mathbf{r}_{A'})d^3\mathbf{r}=\delta_{AA'}; \quad (5.18)$$

тогда величина S окажется равна числу атомов в графитовой плоскости:

$$S=\sum_{A,A'}\exp[ik(\mathbf{r}_A-\mathbf{r}_{A'})]\int X_A^*(\mathbf{r}-\mathbf{r}_A)X_{A'}(\mathbf{r}-\mathbf{r}_{A'})d^3\mathbf{r}=\sum_{A,A'}\exp[ik(\mathbf{r}_A-\mathbf{r}_{A'})]\delta_{AA'}=N. \quad (5.19)$$

Чтобы отвлечься от количества частиц в системе, сделаем замену $H_{ij}=NH'_{ij}$ (выражение для энергии электрона приобретает совсем простой вид $E=H'_{11}\pm|H'_{12}|$). При вычислении матричных элементов удобно перейти от суммирования по всевозможным парам атомов к суммированию по всем атомам и для каждого атома — по ближайшим соседям (положение ближайших соседей задается вектором ρ):

$$H'_{11}=\frac{1}{N}\sum_{A,A'}\exp[ik(\mathbf{r}_A-\mathbf{r}_{A'})]\int X_A^*(\mathbf{r}-\mathbf{r}_A)\hat{H}X_{A'}(\mathbf{r}-\mathbf{r}_{A'})d^3\mathbf{r}=E_0-\gamma'_0\sum_{\rho}\exp[ik\rho]. \quad (5.20)$$

В данном выражении E_0 и γ'_0 есть интегралы перекрытия:

$$\begin{aligned} E_0 &= \int X^*(\mathbf{r})\hat{H}X(\mathbf{r})d^3\mathbf{r}, \\ \gamma'_0 &= \int X^*(\mathbf{r}-\rho)\hat{H}X(\mathbf{r})d^3\mathbf{r}, \end{aligned} \quad (5.21)$$

а вектор ρ , соединяющий ближайших соседей среди атомов типа А, может принимать в гексагональной решетке значения

$$(0,\pm a), \left(\frac{a\sqrt{3}}{2}, \pm\frac{a}{2}\right), \left(-\frac{a\sqrt{3}}{2}, \pm\frac{a}{2}\right). \quad (5.22)$$

Выполняя суммирование по ближайшим соседям, мы получаем:

$$H'_{11}=E_0-2\gamma'_0\left(\cos(k_y a)+2\cos\left(\frac{k_y a}{2}\right)\cos\left(\frac{k_x a\sqrt{3}}{2}\right)\right). \quad (5.23)$$

Совершенно аналогично вычисляются матричные элементы $H'_{12} = H'^*_{21}$, только в этом случае ближайших соседей остается лишь трое, а их радиус-векторы имеют вид

$$\rho = \left(\frac{a}{\sqrt{3}}, 0 \right), \left(-\frac{a}{2\sqrt{3}}, \pm \frac{a}{2} \right). \quad (5.24)$$

Выполняя суммирование по соседям, мы получим:

$$\begin{aligned} H'_{12} &= \frac{1}{N} \sum_{AB} \exp(ik(\mathbf{r}_A - \mathbf{r}_B)) \int X_A^*(\mathbf{r} - \mathbf{r}_A) \hat{H} X_B(\mathbf{r} - \mathbf{r}_B) d^3\mathbf{r} = \\ &= \gamma_0 \sum_{\rho} \exp(ik\rho) = \gamma_0 \left(e^{-ik_y a} + 2e^{ik_y a/2} \cos\left(\frac{k_x a \sqrt{3}}{2}\right) \right), \quad (5.25) \\ \gamma_0 &= \int X_A^*(\mathbf{r}) \hat{H} X_B(\mathbf{r} - \rho) d^3\mathbf{r}. \end{aligned}$$

Из структуры вышеприведенных выражений видно, что неизвестный гамильтониан электрона в двумерной периодической структуре «скрылся» в трех параметрах γ_0 , γ'_0 и E_0 . Последняя величина, определяющая энергию связи p_z -электрона в изолированном атоме, нас вообще не будет интересовать, вычисление остальных величин остается делом эксперимента или более сложных теоретических расчетов. Самым важным является тот факт, что зависимость энергии от волнового вектора определяется лишь структурой кристаллической решетки — и если полученный нами закон дисперсии для графена действительно окажется линейным, то он останется таковым и для любого другого двумерного вещества с атомами в вершинах шестиугольников. Итак, подставляя выражения для матричных элементов (5.23) и (5.25) в (5.16) и отбрасывая E_0 , для закона дисперсии электрона в графитовой плоскости мы получаем выражение:

$$\begin{aligned} E(\mathbf{k}) &= -2\gamma'_0 \left(\cos(k_y a) + 2 \cos\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_x a \sqrt{3}}{2}\right) \right) \pm \\ &\pm \gamma_0 \sqrt{1 + 4 \cos\left(\frac{\sqrt{3} k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) + 4 \cos^2\left(\frac{k_y a}{2}\right)}. \quad (5.26) \end{aligned}$$

Пока полученный закон дисперсии далек от линейного, но спектр электронов в твердом теле и не может быть линейным во всем диапазоне квазиимпульсов — по крайней мере, должна наблюдаться его периодичность с периодом обратной решетки. Обратной решеткой для рассмотренной нами структуры графена также является гексагональная решетка. Лучше всего это видно при численном построении закона дисперсии в виде трехмерного графика (Рис. 5.2).

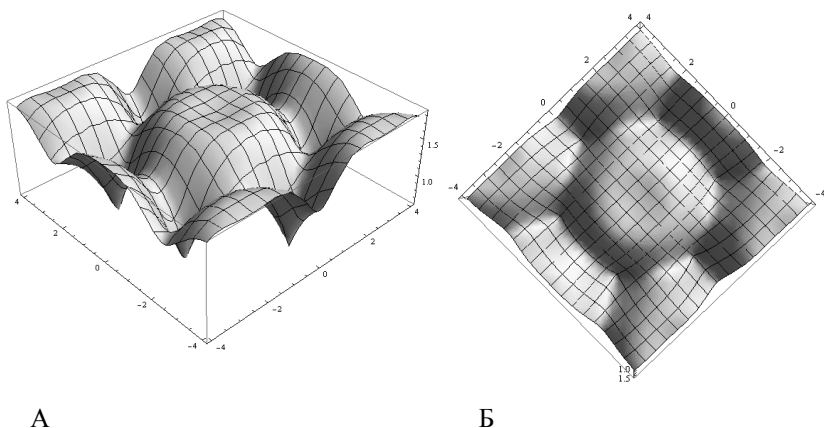


Рис. 5.2. Закон дисперсии электронов в графене: А — трехмерный график $E(k_x, k_y)$, Б — он же при виде сверху (видно, что ячейка обратной решетки также гексагональная)

Особыми точками для обратной решетки являются вершины шестиугольника. Разложим это выражение по степеням волнового вектора, например, в окрестности точки $(0, \pm 4\pi/3a)$:

$$E(k) = -3\gamma'_0 \pm \frac{\sqrt{3}a\gamma_0}{2} \sqrt{\left(k_y \pm \frac{4\pi}{3a}\right)^2 + k_x^2}. \quad (5.27)$$

Мы видим, что в окрестности этой точки спектр является линейной функцией модуля волнового вектора $|k| = \sqrt{\left(k_y \pm 4\pi/3a\right)^2 + k_x^2}$, отсчитанного от вершины шестиугольника. Сокращенно спектр такого вида записывают как

$$E(k) = \pm v_F \hbar |k|, \quad (5.28)$$

отвлекаясь от наличия интеграла перекрытия γ_0' (который предполагается малым по сравнению с γ_0). Величина $v_F = \sqrt{3}a\gamma_0/\hbar$ называется скоростью Ферми, экспериментальные исследования показывают, что она равна 10^6 м/с (не так уж много, в большинстве металлов скорость Ферми имеет приблизительно то же значение). В окрестности остальных экстремумов зависимость $E(k)$ линейна с тем же значением скорости v_F . Линейность спектра в окрестности минимума зоны Бриллюэна подтверждена экспериментально по данным фотоэлектронной спектроскопии с угловым разрешением (Рис. 5.3).

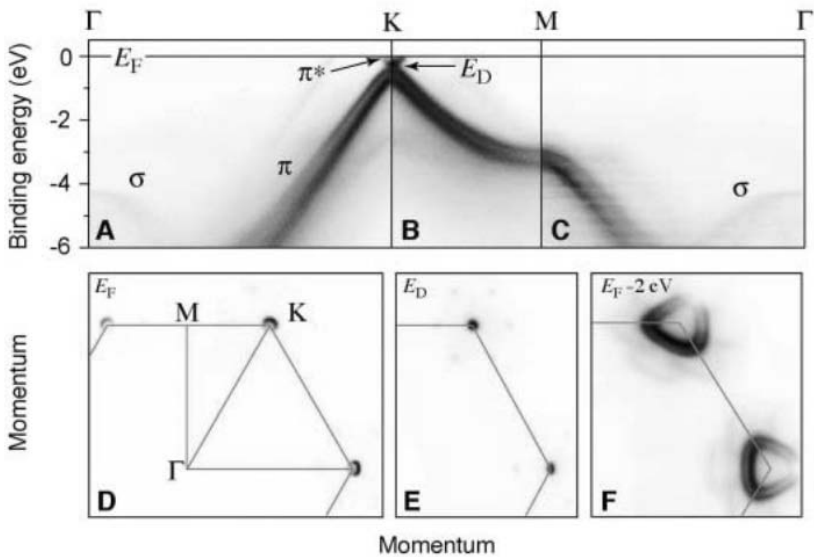


Рис. 5.3. Энергетический спектр электронов в графене, полученный методом фотоэлектронной спектроскопии с угловым разрешением (сверху) и линии равной энергии в импульсном пространстве (снизу) [5]

При вычислении электронных свойств графена нас будут интересовать, однако, не все экстремумы, а только неэквивалентные минимумы в k -пространстве, в качестве которых мы и выберем $(0, \pm 4\pi/3a)$. Т.к. элементарная ячейка графена в координатном пространстве содержит 2 неэквивалентных атома, то и в импульсном пространстве следует учитывать только 2 экстремума. При вычислении термодинамических величин это приводит к дополнительному факто-

ру вырождения электрона $g_V = 2$, называемому долинным фактором (первым фактором вырождения является спиновый с $g_S = 2$).

Важным вопросом для вычисления термодинамических величин графена является вопрос о том, где в изолированном графене находится уровень химического потенциала (уровень Ферми). С учетом того, что в атоме углерода имеется один (из двух возможных) $2p_z$ — электронов, то уровень Ферми будет проходить посередине $2p_z$ -зоны, закон дисперсии для которой мы только что получили. Это приводит к существенному упрощению задач: теперь можно считать химический потенциал равным нулю, а закон дисперсии считать линейным и вообще забыть о константах γ'_0 и E_0 . При этом частицы с энергиями выше уровня Ферми мы будем, как это принято в физике полупроводников, называть электронами, а ниже уровня Ферми — дырками.

Для иллюстрации применения полученных соотношений вычислим равновесную концентрацию электронов (на единицу площади) в графене при комнатной температуре:

$$n_e = \frac{g_s g_V}{(2\pi\hbar)^2} \int_0^\infty \frac{2\pi p dp}{1 + \exp(p v_F / T)} = \frac{4 \cdot 8\pi}{(2\pi\hbar)^2} \left(\frac{T}{v_F} \right)^2 \int_0^\infty \frac{t dt}{1 + e^t}. \quad (5.29)$$

Подстановка численных значений ($T = 300$ К, значение интеграла равно $\pi^2/12$) приводит к результату:

$$n_e = \frac{2\pi}{3\hbar^2} \left(\frac{T}{v_F} \right)^2 = 2.3 \cdot 10^{11} \text{ см}^{-2}. \quad (5.30)$$

Концентрацией электронов и дырок в графене можно эффективно управлять, помещая лист графена в конденсатор и прикладывая внешнее напряжение. Например, если слой графена заземлен, а на противоположную обкладку подано напряжение $+V_0$, то концентрацию носителей заряда (электронов) можно оценить, вовсе не прибегая к статистическим формулам. Достаточно записать определение емкости конденсатора

$$C = \frac{q}{V_0} = \frac{S\varepsilon}{4\pi d} \quad (5.31)$$

и выразить заряд через концентрации носителей ($q = |e|(n_e - n_h)$), тогда

$$n_e = \frac{\varepsilon V_0}{4\pi d} \quad (5.32)$$

(мы считаем, что напряжение велико и вкладом дырок в полный заряд можно пренебречь). Если толщина диэлектрика равна 30 нм, а диэлектрическая проницаемость $\varepsilon = 4$ (диоксид кремния), то

$$n_e [\text{см}^{-2}] = 2.4 \cdot 10^{13} V_0 [\text{В}]. \quad (5.33)$$

Таким образом, при приложении внешнего напряжения концентрацией носителей заряда можно управлять в довольно широких пределах. Отметим, что вблизи нуля напряжений формула (5.33) дает неверно нулевую концентрацию носителей — в этом случае вкладом дырок в полный заряд по сравнению с электронами пренебрегать нельзя. В отличие от типичного полупроводника кремния, графен нельзя сколь угодно сильно обеднить носителями заряда — при нулевом приложенном напряжении концентрации электронов и дырок будут определяться формулой (5.30). Это накладывает свои ограничения на диапазон изменения проводимости графена и его применения в электронике, о которых речь пойдет чуть позже.

Отметим, что минимальную концентрацию носителей заряда можно было бы существенно понизить, если бы в электронном спектре графена имелаась запрещенная зона. В собственных полупроводниках наличие запрещенной зоны с шириной Δ приводит к тому, что равновесная концентрация носителей имеет вид:

$$n = n_i \exp \left\{ -\frac{\Delta}{2kT} \right\}, \quad (5.34)$$

где n_i имеет смысл эффективного числа уровней в зоне проводимости. В связи с проблемой эффективного управления концентрацией носителей заряда стал актуален поиск материалов на основе графена, в электронном спектре которых имеется щель между электронными и дырочными состояниями.

5.3. Электронный спектр двухслойного графена

Упомянутую в конце предыдущего пункта проблему с отсутствием запрещенной зоны у однослойного графена пытались решить несколькими способами, наиболее продуктивными из которых оказались:

5. Создание нанолент — узких полосок из графита; ограниченность их размера в поперечном направлении приводит к появлению запрещенной зоны; ее ширина достигает десятых долей электрон-вольта при ширине полоски в единицы нанометров;

6. Получение двухслойного графена — материала, представляющих собой два слоя графита, наложенных друг на друга.

Вычисление электронного спектра нанолент сводится, по сути, к правильной постановке граничных условий на волновую функцию. Аналогичный подход применяется и при вычислении спектров нанотрубок, где граничные условия являются циклическими. Мы рассмотрим эти подходы чуть позже, в главе 5.4. Сейчас обратимся к исследованию электронного спектра двухслойного графена.

В двухслойном графене атомы углерода, относящиеся к соседним плоскостям, не лежат непосредственно друг над другом. Соседние плоскости сдвинуты друг относительно друга в одном направлении на вектор, равный стороне шестиугольной ячейки. Такое расположение схематически показано на

Рис. 5.4. В результате часть атомов имеет «соседей снизу», под другими же находится пустой центр шестиугольной ячейки.

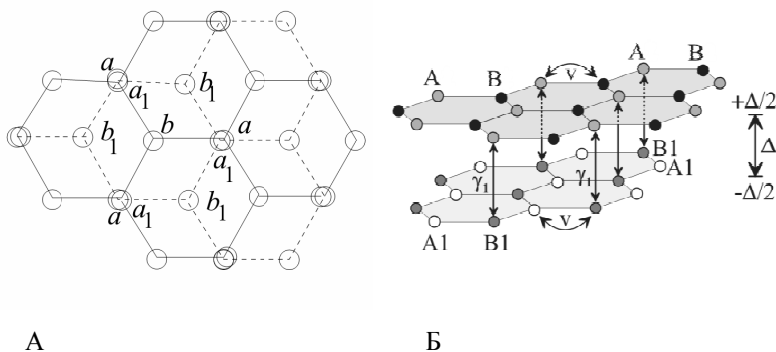


Рис. 5.4. Расположение атомов в соседних плоскостях графена: А — вид сверху, Б — вид сбоку с обозначением констант связи между атомами

Элементарная ячейка для двухслойного графена будет содержать четыре атома, вслед за обозначениями

Рис. 5.4 А мы назовем их a, b, a_1, b_1 . Из волновых функций атомных π -электронов мы «сконструируем» трансляционно-инвариантные блоховские функции:

$$\begin{aligned}\psi_a &= \frac{1}{\sqrt{N}} \sum_j \exp[ik r_j] X(r - r_j), \\ \psi_b &= \frac{1}{\sqrt{N}} \sum_j \exp[ik(r_j + a)] X(r - (r_j + a)), \\ \psi_{a_1} &= \frac{1}{\sqrt{N}} \sum_j \exp[ik(r_j + c)] X(r - (r_j + c)), \\ \psi_{b_1} &= \frac{1}{\sqrt{N}} \sum_j \exp[ik(r_j + c + a)] X(r - (r_j + c + a)).\end{aligned}\tag{5.35}$$

По сравнению с предыдущим пунктом, мы внесли в запись волновых функций некоторые изменения — в частности, нормировали их на число частиц в слое, а также от суммирования по отдельным атомам перешли к суммированию по элементарным ячейкам, которые нумеруются индексом j . Вектор a соединяет соседние атомы в шестиугольной ячейке плоскости, а вектор c — лежащие друг над другом атомы смежных плоскостей.

Для нахождения собственных функций гамильтониана, составленных из линейных комбинаций функций (5.34), запишем уравнение Шредингера в матричном виде:

$$H_{ij} \psi = E(k) \psi, \tag{5.36}$$

в котором ψ является столбцом: $\psi = (\psi_a \ \psi_b \ \psi_{a_1} \ \psi_{b_1})^T$, а H_{ij} — гамильтоновой матрицей с элементами:

$$H_{ij} = \begin{pmatrix} \langle \psi_a | \hat{H} | \psi_a \rangle & \langle \psi_a | \hat{H} | \psi_b \rangle & \langle \psi_a | \hat{H} | \psi_{a_1} \rangle & \langle \psi_a | \hat{H} | \psi_{b_1} \rangle \\ \langle \psi_a | \hat{H} | \psi_b \rangle^* & \langle \psi_b | \hat{H} | \psi_b \rangle & \langle \psi_b | \hat{H} | \psi_{a_1} \rangle & \langle \psi_b | \hat{H} | \psi_{b_1} \rangle \\ \langle \psi_a | \hat{H} | \psi_{a_1} \rangle^* & \langle \psi_b | \hat{H} | \psi_{a_1} \rangle^* & \langle \psi_{a_1} | \hat{H} | \psi_{a_1} \rangle & \langle \psi_{a_1} | \hat{H} | \psi_{b_1} \rangle \\ \langle \psi_a | \hat{H} | \psi_{b_1} \rangle^* & \langle \psi_b | \hat{H} | \psi_{b_1} \rangle^* & \langle \psi_{a_1} | \hat{H} | \psi_{b_1} \rangle^* & \langle \psi_{b_1} | \hat{H} | \psi_{b_1} \rangle \end{pmatrix}. \quad (5.37)$$

Матричные элементы, находящиеся в левом верхнем и правом нижнем углах матрицы, уже были вычислены в предыдущем пункте. Осталось только учесть взаимодействие между электронами, локализованными на разных графитовых плоскостях. Ясно, что наиболее сильным будет взаимодействие между электронами a и a_1 , т.к. они расположены непосредственно друг над другом. Соответствующий матричный элемент обозначим как

$$\gamma_1 = \langle \psi_a | \hat{H} | \psi_{a_1} \rangle = 0.39 \text{ eV}. \quad (5.38)$$

Он не зависит от направлений, т.к. при его вычислении мы учитываем только одного «ближайшего соседа», оба находятся на одной вертикали. Другие матричные элементы, описывающие взаимодействие между слоями

$$\gamma_4 f(k) = \langle \psi_a | \hat{H} | \psi_{b_1} \rangle = \langle \psi_b | \hat{H} | \psi_{a_1} \rangle, \quad \gamma_3 f^*(k) = \langle \psi_b | \hat{H} | \psi_{b_1} \rangle \quad (5.39)$$

уже будут зависеть от направления, и эта зависимость заключена в множителе $f(k)$. Действительно, у атома b есть три ближайших соседа типа a_1 , расположенных в соседней плоскости (аналогично у a есть три соседа типа b_1). Множитель $f(k)$, на самом деле, уже был вычислен нами при учете взаимодействия трех равноудаленных соседей — он равен

$$f(k) = \sum_{\rho} \exp(ik\rho) = e^{-ik_y a} + 2e^{ik_y a/2} \cos\left(\frac{k_x a \sqrt{3}}{2}\right). \quad (5.40)$$

Мы будем интересоваться поведением электронов вблизи минимумов зоны Бриллюэна¹, в связи с этим волновой вектор (и квазиимпульс

¹ Только эти состояния заполнены при небольших температурах и малых напряжениях на затворах транзисторных структур.

электрона) будем отсчитывать от точек $(0, \pm 4\pi/3a)$. Все выражения мы линеаризуем в окрестности этих минимумов:

$$f(k) \approx \frac{v_F (ip_y - p_x)}{\gamma_0}. \quad (5.41)$$

В гамильтониане (5.37) мы можем пренебречь слагаемыми (5.39) — соответствующие им константы связи малы по сравнению с γ_0 , тем более они умножены на $f(k)$, малое вблизи экстремумов зоны Бриллюэна.

В пленке двухслойного графена, находящейся в изотропном пространстве, мы могли бы положить диагональные матричные элементы равными друг другу — ни один из базисных атомов ничем не выделен перед другими. Однако мы предположим, что графен может быть помещен в перпендикулярное его плоскости электрическое поле, тогда энергия состояний на нижнем слое будет отличаться от энергии состояний на верхнем. Мы запишем эту разницу в виде

$$\langle \psi_a | \hat{H} | \psi_a \rangle = \langle \psi_b | \hat{H} | \psi_b \rangle = E_0 - \frac{\Delta}{2}, \quad \langle \psi_a | \hat{H} | \psi_{a_1} \rangle = \langle \psi_b | \hat{H} | \psi_{b_1} \rangle = E_0 + \frac{\Delta}{2}, \quad (5.42)$$

а энергии состояний будем отсчитывать от энергии π -электрона в изолированном атоме E_0 . В простых оценках можно предположить, что разность энергий между слоями Δ равна разности электростатических энергий электрона на верхнем и нижнем слоях:

$$\Delta = Fed, \quad (5.43)$$

F — напряженность поперечного электрического поля. На самом деле под действием поля сами волновые функции π -электронов на различных слоях подвергаются деформации, при этом происходит перераспределение электронной плотности и выражение (5.43) оказывается неверным. Мы получаем, что волновые функции зависят от величины Δ , но и сама величина Δ по определению равна $\langle \psi_{a_1} | \hat{H} | \psi_{a_1} \rangle - \langle \psi_a | \hat{H} | \psi_a \rangle$! В таком случае значение Δ ищется самосогласованно путем минимизации полной энергии системы «электроны в графене + внешнее электрическое поле» [7].

С учетом всех упрощающих предположений, гамильтониан в окрестности экстремумов зоны Бриллюэна запишется в виде:

$$H_{ij} = \begin{pmatrix} \Delta/2 & v_F(ip_x - p_y) & \gamma_1 & 0 \\ v_F(-ip_x - p_y) & \Delta/2 & 0 & 0 \\ \gamma_1 & 0 & -\Delta/2 & v_F(ip_x - p_y) \\ 0 & 0 & v_F(-ip_x - p_y) & -\Delta/2 \end{pmatrix}. \quad (5.44)$$

Нахождение собственных значений такого гамильтониана сводится к решению биквадратного уравнения и дает четыре ветви электронного спектра:

$$E(p) = \pm \sqrt{\frac{\gamma_1^2}{2} + \frac{\Delta^2}{4} + v_F^2 p^2} \pm \sqrt{\frac{\gamma_1^2}{4} + v_F^2 p^2 (\gamma_1^2 + \Delta^2)}. \quad (5.45)$$

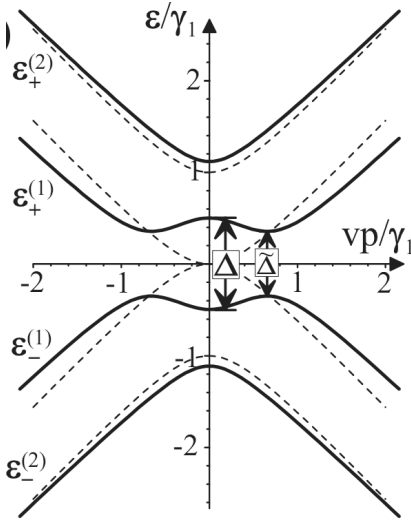


Рис. 5.5. Электронный спектр в двухслойном графене

График данной зависимости изображен на Рис. 5.5, из которого видно, что разность энергий электронов на соседних слоях в точности равна расстоянию между дисперсионными кривыми при $p = 0$, величина запрещенной зоны имеет чуть меньшее значение:

$$\tilde{\Delta} = \Delta \frac{\gamma_1}{\sqrt{\gamma_1^2 + \Delta^2}}, \quad (5.46)$$

которому соответствует квазиимпульс

$$p_{\min} = \frac{\Delta}{2v_F} \sqrt{\frac{2\gamma_1^2 + \Delta^2}{\gamma_1^2 + \Delta^2}}. \quad (5.47)$$

Подобную зависимость энергии от импульса в двухслойном графене часто называют «мексиканской шляпой» за внешнее сходство графика спектра с контуром головного убора.

Из структуры данных выражений видно, что запрещенная зона является следствием разной энергии электронных состояний на соседних атомных плоскостях; если бы мы предположили $\Delta = 0$, то никакой

щели не существовало бы, а электронный спектр выродился в показанный на Рис. 5.5 штриховой линией.

Управление шириной запрещенной зоны является одной из актуальных проблем физики графена. Ясно, что с увеличением напряженности поперечного электрического поля ширина щели будет расти, в работе [6] показано, что в двухслойном графене, помещенном симметрично между двумя электродами с расстоянием 600 нм между ними, может наблюдаться щель до 90 мэВ. Для этого между электродами необходимо приложить напряжение порядка 20 В (предполагается, что межэлектродное пространство заполнено диоксидом кремния).

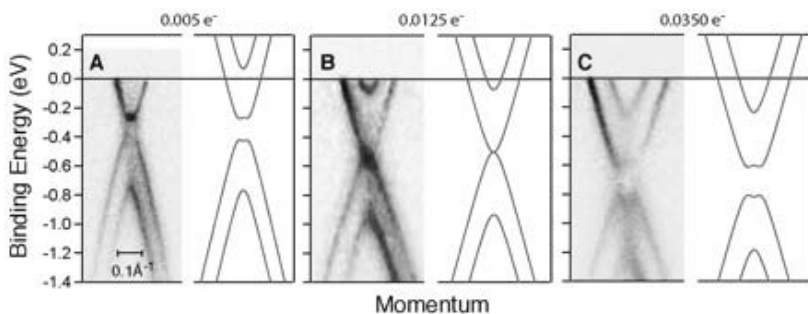


Рис. 5.6. Эволюция электронного спектра двухслойного графена при нанесении атомов натрия на его поверхность. Сверху выписано число легирующих электронов на 1 элементарную ячейку [5]

Альтернативой приложению высоких межэлектродных напряжений является легирование двухслойного графена. В работе [5] двухслойный графен был выращен на подложке карбида кремния в результате пиролизического разложения (часть атомов кремния была испарена с поверхности, оставив несколько верхних слоев графита). Измерения электронной структуры с помощью фотоэлектронной спектроскопии с угловым разрешением показали, что даже при нулевом внешнем поле в графене существует запрещенная зона. Это неудивительно, т.к. подложка карбида кремния вносит свой вклад в асимметрию слоев — из-за разной работы выхода графита и карбида кремния у поверхности формируется сильное поперечное поле, вызванное контактной разностью потенциалов. Выращенный на карбиде кремния графен изначально наполнен электронами, сама подложка заряжена положительно. Асимметрию можно увеличить еще сильнее, нанося на поверхность графена атомы натрия, валентные электроны которых

слабо связаны. При слабом легировании асимметрия пропадает, а при более сильном — возвращается, вызванная появлением положительно заряженных ионов натрия на поверхности (натрий легко отдает свои валентные электроны). Таким образом были достигнуты значения ширины запрещенной зоны вплоть до 200 мэВ (Рис. 5.6).

Заметим, что мы могли бы предположить наличие асимметрии между матричными элементами $\langle \psi_a | \hat{H} | \psi_a \rangle$ и $\langle \psi_b | \hat{H} | \psi_b \rangle$, соответствующим электронам, находящимся на соседних атомах А и В в одной плоскости. Это привело бы к существованию запрещенной зоны уже в однослойном графене. Конечно, продольное электрическое поле, прикладываемое в обычных схемах, очень слабое для создания такой асимметрии (напряжения порядка единиц вольт падают на расстояниях в сотни межатомных). Однако существуют вещества с гексагональной решеткой в плоскости, где А и В являются разными атомами — это, к примеру, нитрид бора. В нем асимметрия создается уже внутренним полем.

Рассмотрев электронный спектр двухслойного графена, мы займемся моделированием транзистора на его основе. Нас будут интересовать такие величины, как проводимость, крутизна характеристики и отношение токов во включенном и выключенном состояниях. Именно ради улучшения последней величины стараются увеличить ширину запрещенной зоны в графене.

5.4. Электронные свойства нанотрубок

Закон дисперсии для электрона в углеродной нанотрубке можно получить из закона дисперсии для графитовой плоскости путем наложения граничных условий. Действительно, описанная выше процедура «склеивания» графеновой ленты обязывает волновую функцию электрона быть одинаковой на прямых l и l' . Аналогично это условие можно сформулировать в виде:

$$\psi(r) = \psi(r + C_h). \quad (5.48)$$

Записывая теперь волновую функцию в виде, который требуется теоремой Блоха

$$\psi(r) = e^{ikr} u_k(r), \quad (5.49)$$

можно получить, что наше граничное условие эквивалентно условию:

$$C_h k = 2\pi m, m \in \mathbb{Z}. \quad (5.50)$$

В пространстве волновых векторов (5.50) задает систему параллельных плоскостей, находящихся на равном расстоянии друг от друга. Пересечение дисперсионной поверхности графита с этими плоскостями как раз дает всевозможные дисперсионные кривые для электронов в нанотрубках.

Закон дисперсии для электрона в графитовой плоскости был найден П. Уэллисом в 1949 году, согласно его вычислениям:

$$E(k) = \pm \gamma_0 \sqrt{1 + 4 \cos\left(\frac{\sqrt{3}k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) + 4 \cos^2\left(\frac{k_y a}{2}\right)}. \quad (5.51)$$

$\gamma_0 \sim 3$ эВ — интеграл перекрытия, характеризующий энергию взаимодействия электрона с ближайшим атомом.

Прежде чем подставить (5.50) в (5.51), необходимо заметить, что элементарной ячейкой нанотрубки уже не будет элементарная ячейка графитовой плоскости. Различие происходит оттого, что нанотрубка переходит сама в себя при параллельных переносах вдоль своей оси (назовем ее y) и поворотах вокруг этой оси на угол $2\pi m/N_x$ — соответственно, нам необходимо переопределить базисные векторы. На Рис. 5.7 приведено сравнение элементарных ячеек и зон Бриллюэна для нанотрубки и графитовой плоскости. По сравнению с графеном, в нанотрубке размер элементарной ячейки удваивается, а размер зоны Бриллюэна, соответственно, уменьшается в 2 раза. «Отброшенную» половину зоны Бриллюэна необходимо отобразить на оставшуюся, это эквивалентно добавлению в закон дисперсии двух дополнительных ветвей.

Выразим теперь k_x через k_y из формулы (5.50), записанной в развернутом виде:

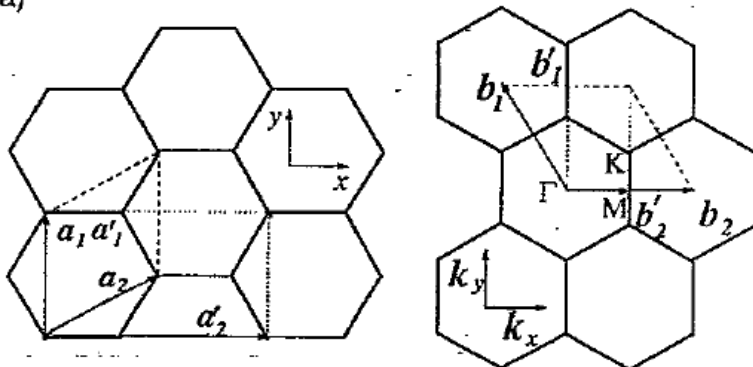
$$\sqrt{3}N_x k_x a + N_y k_y a = 2\pi m, \quad (5.52)$$

и подставим в закон дисперсии; тогда он будет приведен к виду:

$$E_m(k) = \pm \gamma_0 \sqrt{1 \pm 4 \cos\left(\frac{\pi m}{N_x} - \frac{N_y}{N_x} \frac{k_y a}{2}\right) \cos\left(\frac{k_y a}{2}\right) + 4 \cos^2\left(\frac{k_y a}{2}\right)}. \quad (5.53)$$

Два возможных знака в выражении под корнем появились как раз из-за отображения части зоны Бриллюэна.

(a)



(b)

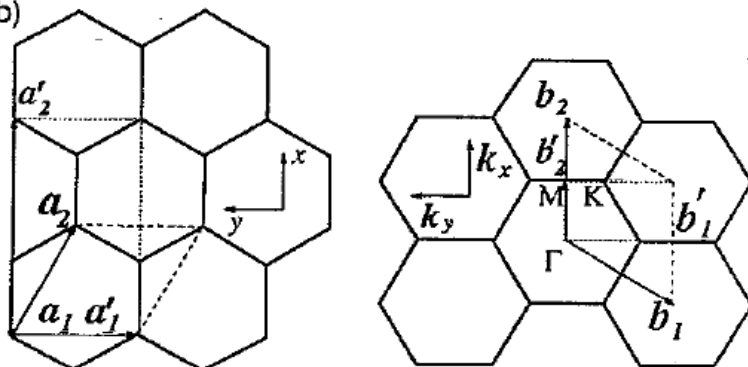


Рис. 5.7. Элементарная ячейка (слева) и зона Бриллюэна (справа) для нанотрубки типа «кресло» (a) и «зигзаг» (б). Штрихованные векторы (a_1' , a_2') соответствуют нанотрубкам, нештрихованные- графитовой плоскости

В частном случае УНТ типа «кресло» вектор хиральности сонаправлен с a_1 , поэтому $N_y = 0$ и закон дисперсии принимает вид:

$$E_m(k) = \pm \gamma_0 \sqrt{1 \pm 4 \cos\left(\frac{\pi m}{N_x}\right) \cos\left(\frac{k_y a}{2}\right) + 4 \cos^2\left(\frac{k_y a}{2}\right)}. \quad (5.54)$$

Если посмотреть на график данной зависимости (Рис. 5.8), можно заметить пересечение двух ветвей при некотором k_y , не находящемся на границе зоны Бриллюэна. Это приводит к тому, что в электронном спектре отсутствует щель между валентной зоной и зоной проводимости — такие трубки называются металлическими.

В УНТ типа «зигзаг» $N_x = 0$, для них закон дисперсии имеет вид

$$E(k) = \pm \gamma_0 \sqrt{1 + 4 \cos\left(\frac{\sqrt{3}k_x a}{2}\right) \cos\left(\frac{\pi m}{N_y}\right) + 4 \cos^2\left(\frac{\pi m}{N_y}\right)}. \quad (5.55)$$

В силу наличия щели между состояниями в валентной зоне и зоне проводимости такие УНТ называются полупроводниковыми.

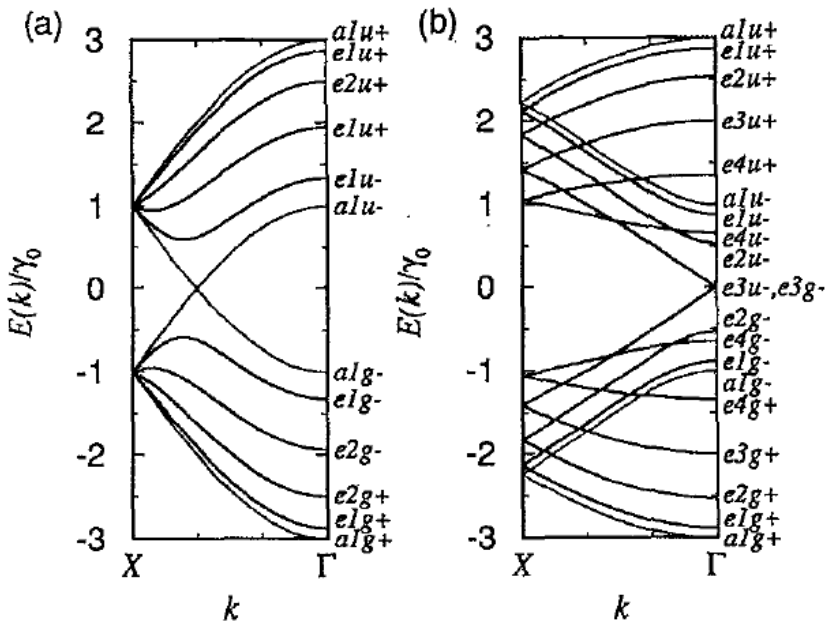


Рис. 5.8. Законы дисперсии металлических (А) и полупроводниковых (Б) нанотрубок

В общем случае хиральной нанотрубки (5.51) щель между валентной зоной и зоной проводимости будет отсутствовать, если разность $N_x - N_y$ равна трем, поэтому металлическим типом проводимости будет обладать лишь треть нанотрубок, полученных при неконтролируемом дуговом распылении.

Экспериментально измерить параметры электронного спектра очень трудно, поэтому измеряется величина, функционально зависящая от закона дисперсии — это плотность электронных состояний. Такие измерения проводятся, например, методами туннельной спектроскопии, где производная сигнала dI/dV пропорциональна плотности состояний $g(E)$. В одномерных проводниках с квадратичным законом дисперсии эта величина на краю зоны ведет себя по закону $g(E) \sim E^{-1/2}$ (т.е. имеет сингулярность).

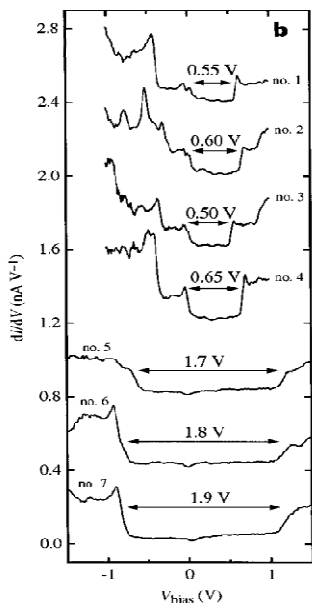


Рис. 5.9. Зависимость плотности электронных состояний в УНТ от энергии (реально изображены производная тока СТМ dI/dV от напряжения)

В одномерных проводниках с линейным законом дисперсии плотность состояний является константой. Данные предсказания хорошо подтверждаются экспериментально (Рис. 5.9) — на этом рисунке приведены результаты для различных углеродных нанотрубок. К сожалению, по этим данным невозможно отделить полупроводниковые УНТ, где вблизи нуля энергий плотность состояний нулевая, от металлических, где в силу линейного закона дисперсии она является константой.

6. Приборы нанoeлектроники и методы их моделирования

Вычисленные нами электронные спектры новых перспективных материалов нанoeлектроники — графена и нанотрубок — вскоре найдут свое законное применение при моделировании транзисторных структур. Перед тем, как заняться собственно моделированием, необходимо познакомиться с физическими принципами работы различных типов транзисторов. Материал следующего параграфа уже далек от тематики квантовой химии и охватывает широкий круг явлений, которые необходимо учитывать при моделировании структур нанoeлектроники, т.к. для корректного моделирования приборов недостаточно знания одного лишь электронного спектра.

Как и в микроэлектронике, так и в нанoeлектронике транзисторы являются основными элементами структур обработки информации: аналоговых схем (усилители), цифровых схем (процессоры в компьютерах), схем памяти (оперативная память в компьютерах). Характеризуя произвольный транзистор, можно сказать, что это трехполюсный прибор, сопротивлением между двумя контактами которого можно управлять, подавая напряжение (или ток) на третий электрод.

Первым управляемым прибором подобного типа был триод — вакуумная лампа, представляющая собой диод с добавленным третьим электродом — сеткой. Напряжение на сетке меняло ВАХ диода, а тока сетка практически не потребляла — именно в силу своей «дырявости». Говоря простыми словами, триод был эквивалентен подстраиваемому резистору. Именно на основе вакуумных триодов в 1943 году Атанасовым и Берри в США был сконструирован первый компьютер.

В 1947 году в лаборатории Белла под руководством Вильяма Шоккли был создан первый биполярный транзистор — усилитель тока, управляемый током. В нем сила тока между двумя контактами — эмиттером и коллектором была усилена по сравнению с током управляющего электрода — базы (в одном из типичных режимов подключения). Модификация биполярных транзисторных схем позволила создать логических элементы, выполненные по технологии ТТЛ (транзистор-транзисторная логика). Сейчас эта технология почти полностью ушла в прошлое, а заменила ее КМОП-логика (построение логических схем на базе комплементарных транзисторов металл-оксид-полупроводник). Рассмотрением полевых транзисторов, выполненных по этой технологии мы займемся подробнее.

6.1. Физические принципы работы различных типов транзисторов

Полевые транзисторы

Полевой транзистор является трехполюсным прибором, проводимостью между двумя контактами которого (истоком и стоком) можно управлять, подавая напряжение на третий контакт — затвор. Существуют два основных типа полевых транзисторов — с управляющим p - n -переходом и с изолированным затвором. Мы более подробно рассмотрим вторые, т.к. они являются основой логических элементов в современных процессорах.

Схема полевого транзистора с изолированным затвором (МОП-транзистор) приведена на Рис. 6.1. Управляющий электрод (затвор) этого транзистора изолирован слоем диэлектрика от объемной кремниевой подложки, в которой выполнено местное легирование областей стока и истока. При отсутствии напряжения на затворе проводимость между контактами стока и истока близка к нулю, т.к. эффективно эта часть цепи представляет собой два p - n -перехода, включенные во встречных направлениях.

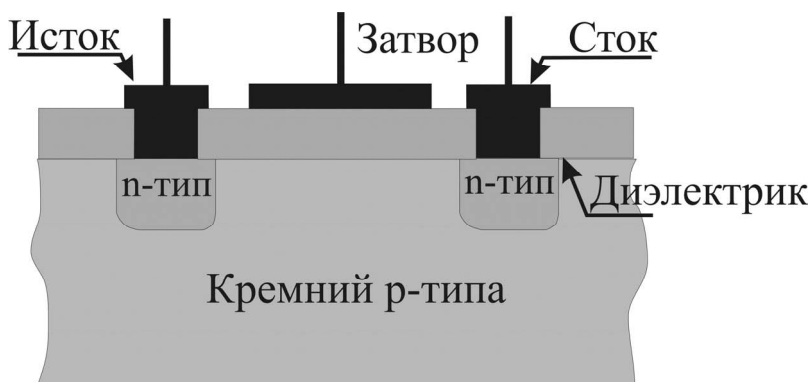


Рис. 6.1. Схематичное изображение полевого транзистора с изолированным затвором (дырочного типа проводимости)

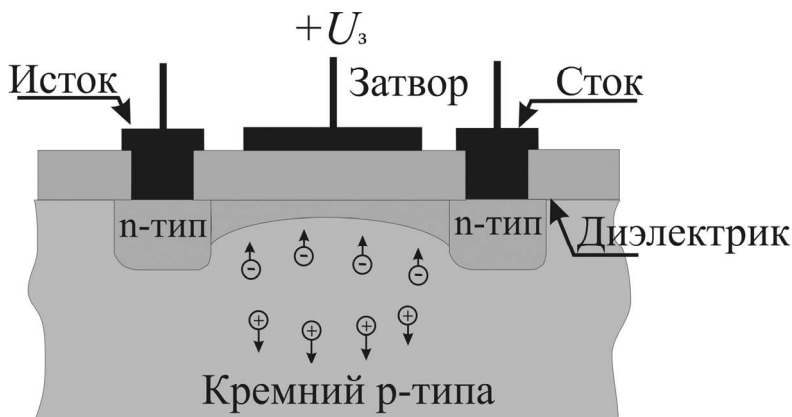


Рис. 6.2. Схематичное изображение полевого транзистора
(образование индуцированного канала)

Физическим эффектом, позволяющим в широких пределах менять проводимость между контактами стока и истока полевого транзистора, является эффект проникновения электрического поля вглубь объема полупроводника. Действительно, при подаче положительного напряжения на затвор транзистора начинается перераспределение зарядов внутри подложки: неосновные носители заряда (электроны) скапливаются ближе к затвору, а основные — дырки — мигрируют вглубь подложки. В результате такого процесса в области вблизи затвора формируется контакт с инверсным (электронным) типом проводимости, и в области исток-сток p - n -переходы уже отсутствуют (транзистор переходит в открытое состояние).

Рассмотренный нами тип транзистора называется «МОП-транзистор с индуцированным каналом» — он является закрытым при нулевом напряжении на затворе. Возможно также создание МОП-транзистора со встроенным каналом — для этого область между истоком и стоком легируется противоположно типу легирования подложки (в рассмотренном нами случае n -канального транзистора — легируется донорами).

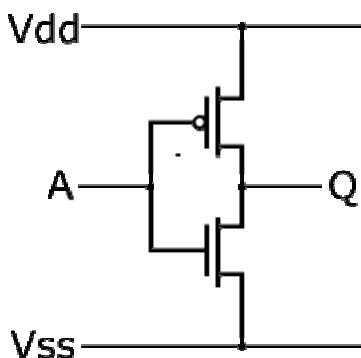


Рис. 6.3. КМОП-инвертор: V_{dd} — шина положительного напряжения, V_{ss} — шина отрицательного напряжения, сверху — р-канальный транзистор, снизу — н-канальный

Наконец, возможно создание МОП-транзисторов, где роль носителей заряда выполняют дырки — они, естественно, переходят в открытое состояние при отрицательных напряжениях на затворе. Комбинация двух таких типов транзисторов (n- и р- канального) позволяет реализовать базовый логический элемент — КМОП-инвертор (Рис. 6.3).

При подаче отрицательного напряжения на вход схемы А верхний (р- канальный) транзистор начинает проводить, а нижний

находится в закрытом состоянии. Фактически при этом выход схемы Q имеет потенциал, равный потенциалу «положительной шины» V_{dd} , то есть выходной сигнал противоположен входному по знаку. Аналогично осуществляется инверсия положительного сигнала.

В современных процессорах важна высокая скорость выполнения арифметических (а поэтому и логических операций). Это требование сводится к увеличению быстродействия логических элементов, а следовательно, и составляющих их транзисторов. Негативное влияние на быстродействие оказывает наличие паразитных емкостей в транзисторах (в частности, между контактами стока и подложкой). Но как бы ни мала была паразитная емкость, предельное быстродействие полевого транзистора определяется временем пролета электрона от области истока до стока. Именно этот факт заставляет стремиться к нанометровым размерам в электронике.

Длина канала полевого транзистора не может быть сделана сколько угодно малой. Во-первых, при длине канала, сравнимой с дебройлевской длиной волны электрона в полупроводнике, могут стать существенными квантовые эффекты, в том числе туннелирование электронов между стоком и истоком при фактически закрытом состоянии транзистора. Однако есть и другие, более технические причины, препятствующие безграничному уменьшению размера транзистора. При малых расстояниях сток-исток возможно смыкание их облас-

тей пространственного заряда и увеличение токов утечки в закрытом состоянии; для предотвращения этого факта необходимо более сильное легирование областей стока, истока и подложки. Но более сильное легирование приводит к сильному рассеянию электронов на заряженных центрах в канале транзистора и к увеличению пролетного времени (следовательно, к снижению быстродействия).

Именно по этой причине в настоящее время производительность компьютеров увеличивается за счет совершенствования их архитектуры, например, путем использования многоядерных процессоров. Однако отказаться от объемной подложки и перейти к подложкам «кремний на изоляторе» (КНИ) заставляет, прежде всего, высокое энерговыделение: схемы по технологии меньше 45 нм уже невозможно охладить.

Для решения данной проблемы было предложено множество альтернативных приборов, основным принципом построения которых является отсутствие легирования контакта и создание резких переходов «сток-канал» и «исток-канал».

Кремниевый полевой нанотранзистор на подложке «кремний на изоляторе»

Подложка КНИ-транзистора (SOI FET) так же, как и у обычного МОП-транзистора, представляет собой монокристалл кремния. На подложке выращен слой изолятора — кристаллической двуокиси кремния SiO_2 . Над изолятором располагается монокристаллический канал из монокристаллического кремния толщиной менее 10 нм. Области истока и стока фактически представляют собой продолжения канала, но отличаются от него высоким уровнем легирования.

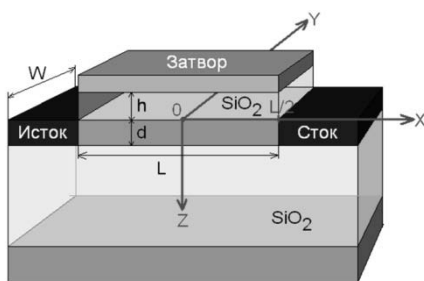


Рис. 6.4. Схематическое изображение КНИ-транзистора

Главные отличия КНИ-транзистора со сверхтонким слоем кремния от субмикронного МОП-транзистора на объемной подложке заключаются в следующем.

1) Канал КНИ-транзистора не содержит примесей. Это обеспечивает баллистический (без рассеяния) режим переноса носителей в канале, что, в

свою очередь, влечет за собой увеличение тока и уменьшение времени пролета канала. Оба эти обстоятельства повышают быстродействие;

2) Поскольку канал тонкий и изолирован от подложки слоем диэлектрика, в закрытом состоянии транзистора малы токи утечки;

3) Отделение контактов стока и истока от кремниевой подложки заметно уменьшает паразитную емкость и, соответственно, приводит к увеличению быстродействия и уменьшению потребляемой мощности.

Важно отметить, что первые структуры со сверхтонким кремнием были изготовлены и исследованы группой В.Попова (Институт физики полупроводников СО РАН). Создание механического напряжения в слое кремния позволяет регулировать пороговое напряжение транзистора, что особенно важно при использовании металлических затворов.

Согласно International Technology Roadmap for Semiconductor (ежегодно обновляемый ведущими экспертами международный прогноз развития техники интегральных схем) основной прогресс до 2021 года будет связан именно с КНИ-технологией. Продвижению к меньшей длине канала препятствует большой ток прямого туннелирования электронов между истоком и стоком в закрытом состоянии транзистора. Очевидным преодолением этого ограничения является применение полупроводников с большей шириной запрещенной зоны, например, нитрида галлия или алмаза.

Полевой транзистор с высокой подвижностью электронов

В полевом транзисторе с высокой подвижностью электронов (НЕМТ) используется гетеропереход между полупроводником с большой запрещенной зоной и полупроводником с малой шириной запрещенной зоны. Обычно используют полупроводники группы A_3B_3 . В такой структуре впервые был получен двумерный электронный газ. На гетерогранице возникает потенциальная яма, имеющая треугольный профиль. Носители занимают низшее состояние пространственного квантования и образуют двумерный газ, который и служит каналом транзистора. Подвижность электронов в канале транзистора достигает рекордных значений, что и обуславливает его название. Высокая подвижность обеспечивается значительным удалением (сотни нанометров) легированной области, поставляющей электроны в канал транзистора, от самого канала. Приближение этой области к каналу, неизбежное при стремлении к нанометровым размерам, может вызвать снижение подвижности из-за сильного кулоновского рассеяния носителей на заряженных примесных центрах. Кроме того, в транзисторах микронных размеров высокая проводимость контактов

истока и стока была обусловлена их большой площадью. Создание хороших контактов в нанометровой области размеров пока представляет значительную технологическую проблему.

Сравнение КНИ-транзистора с альтернативными приборами

В этом разделе мы указываем на некоторые направления поиска новых перспективных нанотранзисторных структур, применение которых могло бы продлить закон Мура после 2020 г. Эти структуры часто называются в качестве конкурентов или приемников кремниевых транзисторов. При этом предполагается, что кремниевая подложка остается основой в крупномасштабных производствах. Пока все вновь предлагаемые приборы на иных материалах являются полевыми транзисторами, в которых потенциал затвора управляет током между истоком и стоком. Новые материалы должны быть встроены в канал транзистора. Как будет показано ниже, в настоящее время создание и использование альтернативных приборов сталкивается не только с очень серьезными технологическими трудностями, но и принципиальными ограничениями. Весьма вероятно, что обычная конструкция полевых транзисторов с использованием новых материалов так и не позволит превзойти по характеристикам кремниевые транзисторы. Возможно, только новые принципы работы элементов и организации логических схем позволят нанoeлектронике продвинуться в технологические размеры ниже 5-10 нм.

Туннельные транзисторы

Как мы отмечали ранее, одной из проблем, препятствующих возможности масштабирования полевых транзисторов в нанометровую область, является смыкание областей пространственного заряда стока и истока. Ее можно обойти, если сделать границу областей с различным типом проводимости очень резкой — в пределе необходимо взять два разных вещества вместо одного полупроводника с разным типом легирования. В туннельном транзисторе контакты стока и истока выполнены из металла, а канал — из полупроводника (на

Рис. 6.5 — из кремния).

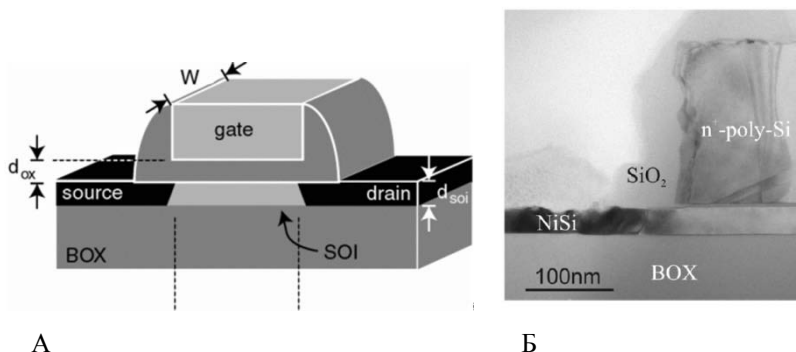


Рис. 6.5. Схема туннельного транзистора на нелегированной кремниевой подложке (А) и изображение его скола в просвечивающем электронном микроскопе

Различное положение уровня Ферми в металле и полупроводнике приводит к перераспределению зарядов (см. зонную диаграмму на Рис. 6.6) и образованию потенциального барьера высотой Φ_B (барьер Шоттки). Для контакта нелегированного кремния с большинством металлов эта величина порядка полуширины запрещенной зоны кремния.

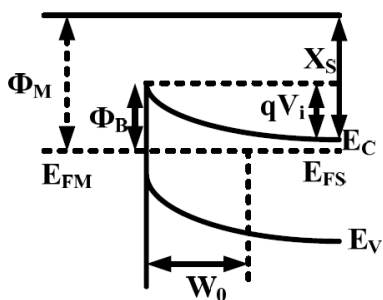


Рис. 6.6. Изгиб зон при контакте полупроводника и металла (работа выхода полупроводника X_S меньше работы выхода металла Φ_M)

В туннельном транзисторе при подаче различного напряжения на затвор можно регулировать положение дна зоны проводимости на Рис. 6.6, а следовательно, и ширину барьера Шоттки. Вместе с шириной меняется вероятность туннелирования через барьер, а значит, и ток в системе сток-исток. Тот факт, что зависимость прозрачности барьера от его ширины практически экспоненциальна, позволяет регулировать ток транзистора в очень широких пределах.

Интерференционные транзисторы

Одним из мотивов поиска новых принципов работы транзисторов является уменьшение энергии их переключения из открытого состояния в закрытое состояние в логической схеме.

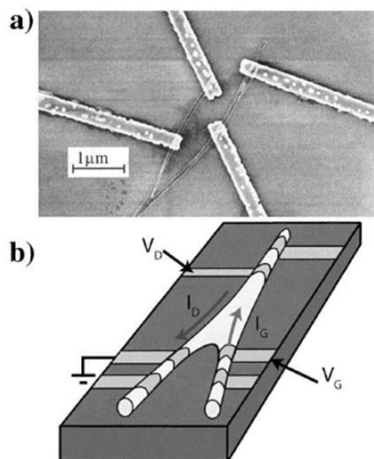


Рис. 6.7. Транзистор на Y-образной нанотрубке [34]:
а) Структура интерференционного транзистора;
б) Схематическое изображение интерференционного транзистора.

Во всех используемых в настоящее время транзисторах такое переключение осуществляется за счет перемещения электронов из одного места пространства в другое. Эти положения разделяются потенциальным барьером, изменяя высоту которого можно делать такой переход возможным или невозможным. Фактически, это означает управление термоэмиссией. Энергия переключения в расчете на один электрон не может быть меньше тепловой энергии (kT). В этом и состоит фундаментальное ограничение. Интерференционные и

спиновые транзисторы работают совсем на других принципах, и это ограничение для них не действует. Интерференционные транзисторы являются квантовыми приборами. Их работа основана на управлении интерференцией носителей тока с помощью потенциалов внешних электродов.

Все предложенные до сих пор конструкции интерференционных транзисторов пока оказываются нереализуемыми. Возможно, наиболее перспективными являются Y-образные нанотрубки. Потенциалом затвора, расположенным в месте разветвления, можно направлять пучок электронов либо в одно русло, либо в другое. Важно понять, как формировать и размещать подобные объекты в нужных местах структуры, ведь, как правило, Y-образные нанотрубки возникают случайным образом среди прочих обычных нанотрубок. Однако уже удалось изготовить трехвыводное устройство усилителя тока, выращивая регулярным образом именно Y-образные разветвления на нанотрубке.

Частным случаем интерференционных приборов являются резонансно-туннельные приборы. Эффект резонансного туннелирования имеет оптический аналог — это интерферометр Фабри-Перо. Если энергия налетающего электрона совпадает с энергией резонансного уровня в структуре, то происходит значительное увеличение прозрачности структуры. Следствием этого эффекта является наличие N-образной вольтамперной характеристики с участком отрицательной дифференциальной проводимости, которая может быть использована для генерации СВЧ излучения.

Спиновые транзисторы

Спинтроника является новой перспективной областью электроники, которая начала развиваться сравнительно недавно. В ее основе лежит управление состоянием спина электрона. Ожидаемая практическая польза от применения спиновых транзисторов — уменьшение энергии переключения. В отсутствие внешнего магнитного поля на изменение состояния спина электрона (спин направлен вверх или вниз) в результате прецессии не тратится энергия. Конечно же, при этом энергия тратится в цепи на создание определенных потенциалов на электродах.

В 1990 году S. Datta и B. Das предложили спиновый полевой транзистор. Исток, которым может быть, например, ферромагнетик, в идеале поставляет поток электронов со 100% спиновой поляризацией в канал транзистора. Сток служит спиновым фильтром, он пропускает только электроны с соответствующей спиновой поляризацией. Спин-орбитальное взаимодействие внутри канала, управляемое напряжением затвора, производит прецессию спинов. При повороте на 180° электроны уже не проходят в сток, ток равен нулю. Это и есть закрытое состояние транзистора. Потенциал затвора создает перпендикулярное электрическое поле в канале. При переходе в систему отсчета электрона это поле приобретает магнитную компоненту, с которой и взаимодействует магнитный момент электрона, возникает прецессия магнитного момента (спина).

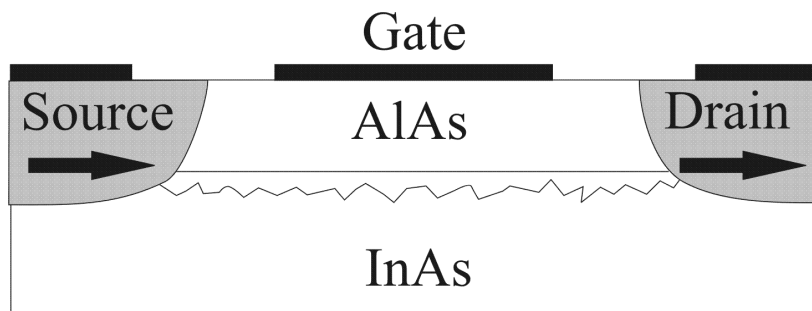


Рис. 6.8. Структура спинового транзистора. Стрелками показано направление спина в контактах истока и стока. Затвор управляет прецессией спина при прохождении канала

В предложенном транзисторе используется особый вид спин-орбитального взаимодействия, свойственный полупроводникам группы A_3B_5 . Это взаимодействие впервые было изучено советскими теоретиками Ю. А. Бычковым и Э. И. Рашбой в 1984 году [36]. Они произвели расчет спин-орбитального взаимодействия в многозонной модели, т.е. с учетом замешивания состояний электронов и дырок, и получили величину этого взаимодействия на 5 порядков выше, чем дает однозонная модель. Разница энергий различной ориентации спина в реальных структурах, обусловленная спин-орбитальным взаимодействием Рашбы, составляет несколько мэВ, что уже открывает перспективы практического использования этого эффекта.

Однако пока надежды, связанные со спиновым транзистором, а именно, низкое управляющее напряжение, малое энергопотребление и высокое быстродействие, оказываются противоречивыми. Возможно, в будущем будут созданы структуры, в которых имеет место гораздо более сильное спин-орбитальное взаимодействие. Кроме того, требуется создать инжектор спинового тока и спиновый фильтр с эффективностью близкой к 100%. Применение т.н. разбавленных магнитных полупроводников (полупроводников с внедренными магнитными примесями) может привести к решению этих проблем.

В последнее время собственно магнитные полупроводники вновь привлекли к себе внимание. В них переход из ферромагнитного состояния в немагнитное при повышении температуры сопровождается увеличением сопротивления на 13 порядков величины. Это рекордное по величине изменение сопротивления при переходе металл-изолятор. Первое объяснение этого явления принадлежит советскому ученому Э. Л. Нагаеву, который и ввел в обиход понятие «магнитного полярно-

на» в 1967 году. Магнитный полярон представляет собой маленькую каплю ферромагнитной фазы размером 0.5 нм в парамагнитном окружении. Когда образуются магнитные поляроны, электроны проводимости локализуются и уходят из процесса проводимости, при этом возникает диэлектрическое состояние. Сотрудники лаборатории В. Сторчака (Курчатовский НБИК Центр) уже много лет используют метод мюонной спиновой релаксации для исследования магнитных поляронов в магнитных полупроводниках. Для применения в наноэлектронике требуется возможность управления состоянием магнитного полярона с помощью электрического поля.

Молекулярные транзисторы

Привлекательность молекул и атомов, прежде всего, состоит в том, что они являются идеальными объектами и воспроизводятся со 100% точностью. Однако их применение в микроэлектронных схемах сталкивается с проблемой формирования воспроизводимых контактов. Действительно, в качестве канала транзистора могут использоваться молекулы, металлические гранулы, фуллерены и даже одиночные атомы. Однако пока общим недостатком таких структур является невозможность их характеристик именно из-за контактов. Возможно, со временем будут созданы молекулярные контакты, а пока, по-видимому, подобные транзисторы найдут свое ближайшее применение в качестве чувствительных сенсоров, причем, особенности вольтамперных характеристик, в принципе, могут позволить идентифицировать ту или иную молекулу.

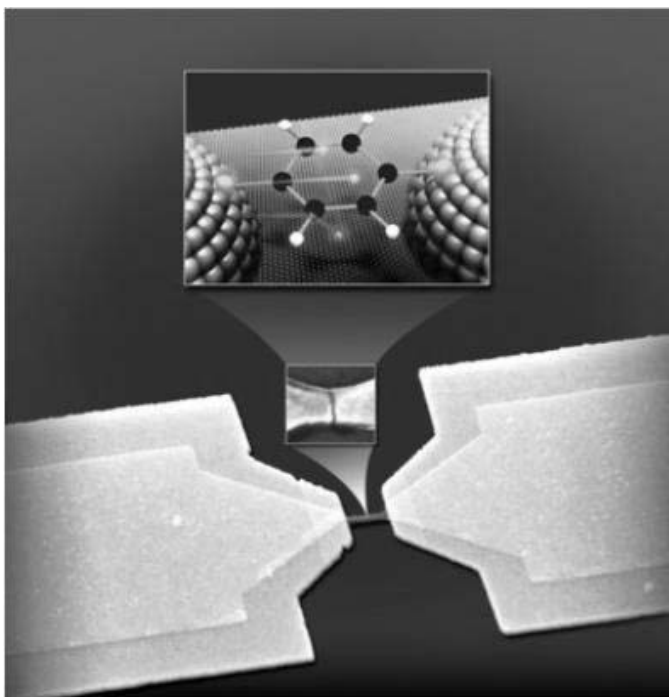


Рис. 6.9. Молекулярный транзистор на основе молекулы бензола

Полевые транзисторы на основе графена

Использование графена в качестве проводящего канала полевых транзисторов представляет принципиально новый подход к решению проблемы быстродействия. Если ранее пролетное время стремились уменьшить за счет уменьшения длины канала транзистора, то в графене это достигается за счет увеличения скорости самих носителей заряда. Даже если режим движения носителей является баллистическим, то в полупроводниках соответствующую скорость можно оценить как $\sqrt{kT/m_e} \sim 70$ км/с.

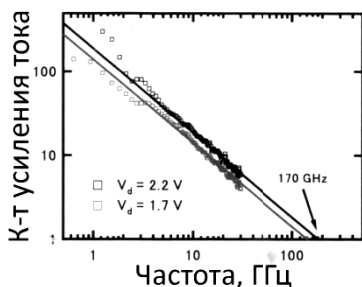


Рис. 6.10. Частотная зависимость к-та усиления наиболее быстрых графеновых транзисторов

полупроводников. Более того, баллистический режим переноса носителей заряда в графене реализуется уже при микронных длинах каналов (для подвешенного графенового листа). Это делает его перспективным материалом для новых быстродействующих транзисторов.

К 2010 году наиболее перспективными в плане быстродействия являются транзисторы на основе графена с частотой единичного усиления порядка 170 ГГц [1, Рис. 6.10].

6.2. Эволюция моделей МОП транзисторов

Моделирование полевых транзисторов, несомненно, важно как для понимания физических процессов, происходящих в них, так и для оптимизации их конструкции с целью достижения наилучших характеристик. Именно моделирование отражает эволюцию представлений о том, как работает транзистор при различных его размерах (Рис. 6.11).

Принципиально расчет характеристик полевых транзисторов (зависимости тока от напряжения сток-исток, зависимости тока от напряжения на затворе) осуществляется следующим образом:

1. Вычисляется распределение электрического потенциала, создаваемого напряжением на затворе транзистора. Это делается, например, решением уравнения Пуассона;
2. При вычислении малосигнальных параметров (тока при малом напряжении сток-исток) достаточно решить задачу о движении электрона в заданном потенциале;
3. При вычислении отклика транзистора на напряжение, сравнимое с напряжением сток-исток, необходимо рассчитывать потенциальный рельеф с учетом этих двух напряжений. Фактически задача сводится к решению динамических (кинетических) уравне-

В металлах роль скорости хаотического движения играет фермиевская, по порядку величины равная 1000 км/с.

Графен объединяет в себе высокую электронную скорость, унаследованную от металла ($v_F = 8 \cdot 10^5$ м/с) и возможность управления концентрацией носителей заряда, унаследованную от

ний с заданными граничными условиями в контактах транзистора (потенциал истока обычно берется равным нулю, потенциал истока — V_D).

Для микронных размеров транзисторов, когда длина свободного пробега была гораздо меньше длины канала, использовалось представление о том, что в канале транзистора протекает заряженная жидкость, и для его описания использовались гидродинамические модели, в частности, дрейфово-диффузионная. Для субмикронных размеров, когда длина свободного пробега уже оказывалась сопоставимой с длиной канала, более реалистичными оказываются представления о частицах, пролетающих через канал транзистора. Для описания использовалось кинетическое уравнение Больцмана для функции распределения. Эффективным методом решения подобного уравнения оказался метод Монте-Карло. Следует заметить, что применение диффузионно-дрейфовой модели для описания переноса в баллистическом режиме приводит к совершенно неправильному результату. Эта модель дает бесконечное нарастание тока при уменьшении длины канала. В то же время, проводимость баллистического канала ограничена квантом проводимости и перестает зависеть от длины канала. Представления о процессах, происходящих в канале транзистора, радикально изменяются, когда ширина и длина канала становятся близкими к де-Бройлевской длине волны носителей. Реально это происходит при длине канала 10-50 нм. В этом случае канал ведет себя как волновод и для его описания необходимо использовать уравнение Шредингера. Такой транзистор становится квантовым.

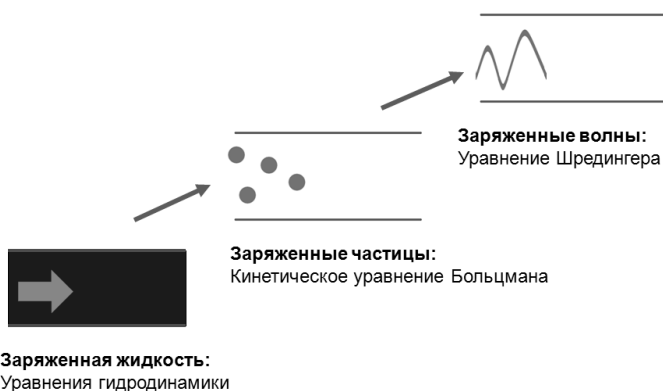


Рис. 6.11. Эволюция физических моделей, описывающих работу транзисторов

6.3. Основные квантовые эффекты в полевых нанотранзисторах

Прежде чем рассмотреть квантовые методы моделирования КНИ транзисторов, назовем основные квантовые эффекты, которые должны быть приняты во внимание.

Поперечное квантование.

Движение носителей поперек канала ограничено высокими потенциальными стенками, что отвечает слабому проникновению «хвостов» волновых функций вглубь окисла. Это означает, что мы можем рассматривать канал как волновод, а энергия носителей (электронов или дырок), отвечающая их поперечному движению, квантуется, т.е. принимает дискретные значения. Энергия поперечного квантования электрона в тонком слое КНИ вносит вклад в величину порогового напряжения. Величина поправки для носителей (электронов и дырок) может быть оценена как энергия низшего состояния поперечного квантования в канале

$$\varepsilon_0 = \frac{\hbar^2}{2m} \left[\frac{\pi}{d_{Si}} \right]^2, \quad (6.1)$$

где m — соответствующая эффективная масса (например, для электронов это наибольшая эффективная масса зоны проводимости $m = 0,98m_0$, m_0 — масса свободного электрона), d_{Si} — толщина слоя кремния (рис. 6.12).

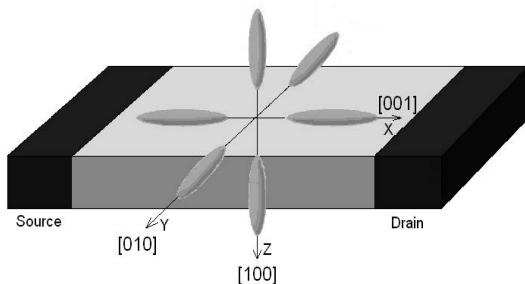


Рис. 6.12. Структура зоны проводимости в канале КНИ транзистора

Для слоя кремния толщиной $d_{Si}=5$ нм эта поправка еще невелика по сравнению с типичными пороговыми напряжениями (0,5 В) и составляет 20 мВ. Однако соответствующий сдвиг порогового напряжения уже может вызвать существенное изменение величины подпорогового тока. Для толщины слоя кремния $d_{Si}=2$ нм сдвиг порогового напряжения уже настолько велик (0.12 мВ), что его необходимо компенсировать выбором материала затвора. В столь тонком канале в переносе тока, фактически, участвуют только электроны двух долин из шести (рис. 6.12). Интересно, что само по себе это обстоятельство не означает уменьшение тока транзистора в 3 раза, и даже изменение тока вообще.

Вторым следствием поперечного квантования является снижение емкости системы затвор-канал. В классическом рассмотрении предполагалось, что электронная плотность в открытом режиме равномерно «размазана» по толщине канала, при этом рассматривался плоский конденсатор с толщиной диэлектрика, равной толщине подзатворного окисла. С учетом поперечного квантования ясно, что из-за малой величины амплитуды волновой функции на стенках канала электронная плотность «отодвинута» от стенок (Рис. 6.13). В этом случае уже следует рассматривать конденсатор с большей толщиной диэлектрика. Емкость такого конденсатора меньше и ее принято называть квантовой емкостью.

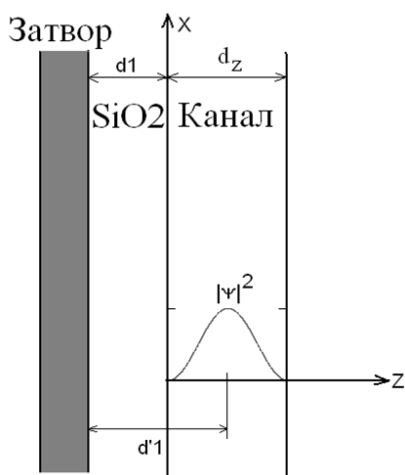


Рис. 6.13. Распределение электронной плотности по сечению канала.

Поперечное квантование может играть существенную роль при учете поверхностного рассеяния [11, 12, 13]. Квантовый расчет дает исключительно высокие степени зависимости интенсивности рассеяния от d_{Si} вплоть до d_{Si}^{-6} в зависимости от характера шероховатости [14]. Интуитивное обоснование этого явления может быть следующим. Классическая частица «чувствует» поверхность только в момент непосредственного касания. Волновая функция «чувствует» поверхность во всем объеме канала.

Поверхностное рассеяние может существенно уменьшить токи КНИ-нанотранзисторов. Чтобы избежать этого, нужны совершенные атомно-слоевые технологии. Впрочем, они нужны не только для получения тонких слоев кремния, но и подзатворных диэлектриков.

Продольное квантовое движение в канале

Квантовый транспорт вдоль канала транзистора резко отличается от классического, в частности, при прохождении потенциального барьера. Следствием этого эффекта может быть прохождение через барьер частиц с энергией ниже высоты этого барьера и отражение частиц, энергия которых превышает высоту барьера.

Наиболее существенным фактором, который ограничивает снизу возможную длину канала транзистора, является прямое туннелирование электронов между истоком и стоком, когда транзистор работает в подпороговом режиме. Оценки показывают, что при длине канала 12 нм туннельный ток начинает превосходить термоэмиссионный [15]. Заметим, что наблюдавшаяся в эксперименте величина $I_{\text{выкл}}$ при длине канала 5 нм составила 10^{-6} А/мкм [16], что на несколько порядков превосходит минимальную величину термоэмиссионного тока 10^{-9} А/мкм.

Следующим квантовым эффектом, который необходимо учитывать при расчете баллистического нанотранзистора, является интерференция электронов на случайно расположенных дискретных заряженных примесях или других заряженных центрах внутри диэлектрика или на его границе со слоем кремния. В самом деле, при столь малых размерах канала уже необходимо принимать во внимание дискретность заряженных центров. Неоднородность (шероховатость) стенок канала тоже может существенно сказаться на переносе носителей.

Статистика Ферми—Дирака

Высокая степень легирования контактов истока и стока (10^{20} – 10^{21} см⁻³) приводит к вырождению электронного или дырочного газа. Это вызывает необходимость использовать в расчетах квантовую функцию распределения Ферми—Дирака.

6.4. Обзор квантовых методов моделирования нанотранзисторов

К числу основных квантовых методов, которые предлагается использовать для моделирования объектов нанoeлектроники можно отнести метод функций Вигнера [17, 18], метод неравновесных функций Грина [19-21] и метод, основанный на формализме Ландауэра-Бюттикера [22-28].

Метод функций Грина

Ключевым понятием в этом методе является функция Грина, а точнее, причинная функция Грина. Введем ее для системы тождественных квантовых частиц. Пусть каждая частица может находиться в одном из квантовых состояний $\psi_i(x)$. Тогда так называемые ψ -операторы записываются в виде:

$$\hat{\psi}(x) = \sum_i \hat{a}_i \psi_i(x), \quad \hat{\psi}^+(x) = \sum_i \hat{a}_i^+ \psi_i^*(x) \quad (6.2)$$

где \hat{a}_i^+ и \hat{a}_i — канонические операторы рождения и уничтожения частиц, вводимые в представлении вторичного квантования.

Представим гамильтониан системы в виде

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}} \quad (6.3)$$

где \hat{H}_0 является невозмущенной частью гамильтониана, а \hat{H}_{int} — возмущение. Далее ψ -операторы преобразуются следующим образом в представлении взаимодействия:

$$\tilde{\psi}_\alpha(\vec{r}, t) = e^{\frac{i}{\hbar} \hat{H}_0 t} \hat{\psi}_\alpha(\vec{r}) e^{-\frac{i}{\hbar} \hat{H}_0 t}, \quad \tilde{\psi}_\alpha^+(\vec{r}, t) = e^{\frac{i}{\hbar} \hat{H}_0 t} \hat{\psi}_\alpha^+(\vec{r}) e^{-\frac{i}{\hbar} \hat{H}_0 t} \quad (6.4)$$

где α обозначает спиновый индекс.

Введем понятие хронологического произведения операторов (Т-произведения) для некоммутирующих операторов A и B следующим образом

$$T \hat{A}(\vec{r}, t) \hat{B}(\vec{r}, t') = \begin{cases} \hat{A}(\vec{r}, t) \hat{B}(\vec{r}, t'), & t > t', \\ \pm \hat{B}(\vec{r}, t') \hat{A}(\vec{r}, t), & t < t'. \end{cases} \quad (6.5)$$

Знак «+» соответствует статистике Бозе, а «-» — статистике Ферми. Теперь можно определить саму причинную функцию Грина:

$$G_{\alpha\beta}(\vec{r}, \vec{r}', t, t') = -i \langle \hat{S}_0 \rangle^{-1} \langle 0 | T \tilde{\psi}_\alpha(\vec{r}, t) \tilde{\psi}_\beta(\vec{r}', t') | 0 \rangle \quad (6.6)$$

где $\langle 0 | \dots | 0 \rangle$ обозначает матричный элемент, взятый по основному состоянию системы с невозмущенным гамильтонианом \hat{H}_0 . Множитель $\langle \hat{S}_0 \rangle^{-1} = \langle 0 | \exp \left[-\frac{i}{\hbar} \hat{H}_0 t \right] | 0 \rangle^{-1}$ взят для нормировки.

Если функция Грина становится известной из расчетов, то с ее помощью легко найти значения различных важных физических величин. Так, например, одночастичная матрица плотности равна

$$\rho_{\alpha\beta}(\vec{r}, \vec{r}', t) = \pm i \lim_{t' \rightarrow t+0} G_{\alpha\beta}(\vec{r}, \vec{r}', t, t') \quad (6.7)$$

Плотность частиц и плотность их потока в свою очередь выражаются следующим образом:

$$\begin{aligned} n(\vec{r}, t) &= \pm i \lim_{\substack{t' \rightarrow t+0 \\ \vec{r}' \rightarrow \vec{r}}} Tr G_{\alpha\beta}(\vec{r}, \vec{r}', t, t') \\ \vec{j}(\vec{r}, t) &= \pm \frac{\hbar}{2m} \lim_{\substack{t' \rightarrow t+0 \\ \vec{r}' \rightarrow \vec{r}}} (\vec{\nabla}_{\vec{r}} - \vec{\nabla}_{\vec{r}'}) Tr G_{\alpha\beta}(\vec{r}, \vec{r}', t, t') \end{aligned} \quad (6.8)$$

(знак «+» снова соответствует статистике Бозе, а «-» — статистике Ферми.), m — масса частицы, взятие следа Tr (суммирование) производится по спиновым индексам.

Основным преимуществом метода функций Грина является наличие регулярной процедуры введения рассеяния, как частиц друг на друга, так и на внешних объектах (фононах и дефектах). При рассмотрении баллистического транспорта этот метод не имеет преимуществ перед другими методами. В последнее время метод функций Грина успешно применяется для моделирования транспорта в молекулярных структурах. Есть основания полагать, что именно в этой области он станет лидирующим.

Метод функций Вигнера

Уравнение Вигнера является полным эквивалентом уравнения Шредингера и записывается для "функции распределения" Вигнера $f_w(\vec{r}, \vec{p}, t)$, которая выражается через волновую функцию $\psi(\vec{r}, t)$ или соответствующую матрицу плотности

$$\rho(\vec{r}, \vec{r}', t) = \psi(\vec{r}, t) \psi^*(\vec{r}', t) \quad (6.9)$$

следующим образом:

$$f_w(\vec{r}, \vec{p}, t) = \int_{-\infty}^{+\infty} d\vec{r}' \rho(\vec{r} + \vec{r}'/2, \vec{r} + \vec{r}'/2) e^{-i\vec{p}\vec{r}'/\hbar} \quad (6.10)$$

Функция Вигнера является решением уравнения

$$\frac{\partial f_w}{\partial t} = -\frac{\vec{p}}{m} \frac{\partial f_w}{\partial \vec{r}} - \frac{1}{\hbar} \int_{-\infty}^{+\infty} \frac{d\vec{p}'}{2\pi\hbar} V_w(\vec{r}, \vec{p} - \vec{p}') f_w(\vec{r}, \vec{p}', t) \quad (6.11)$$

где
$$V_w(\vec{r}, \vec{p}) = 2 \int_0^{+\infty} d\vec{r}' \sin(\vec{p}\vec{r}'/\hbar) \left[V(\vec{r} + \frac{1}{2}\vec{r}') - V(\vec{r} - \frac{1}{2}\vec{r}') \right].$$

Здесь $V(\vec{r})$ — потенциальная энергия, включающая электрическую часть $eU(\vec{r})$, величина \vec{p} имеет смысл импульса частицы, тогда $\vec{p} = m\vec{v}$. Легко заметить значительное сходство уравнения Вигнера с кинетическим уравнением Больцмана, точнее, с уравнением Лиувилля, когда интеграл столкновений равен нулю. На первый взгляд, уравнение Вигнера выглядит, как естественное продолжение классического кинетического уравнения в квантовую область. С его помощью особенно удобно сопрягать классические зоны (контакты) с квантовыми (канал транзистора), поскольку граничные условия для обоих уравнений выглядят одинаково: они ставятся на функцию распределения частиц, влетающих в расчетную область. На этом, к сожалению, сходство заканчивается и начинаются существенные отличия, которые, вообще говоря, и следует ожидать, поскольку уравнение Вигнера является просто по-другому записанным уравнением Шредингера. Отметим эти различия:

- уравнение Вигнера обладает исключительно сложным нелокальным полевым слагаемым, которое фактически описывает интерференцию;

- не удастся ввести произвольное рассеяние в уравнение Вигнера, пока это сделано лишь для δ -образного потенциала рассеивающих центров, что далеко от реальной ситуации (фононы, кулоновское рассеяние, поверхностное рассеяние);
- в отличие от истинной функции распределения, функция "распределения" Вигнера не является положительно определенной величиной и, в действительности, не может трактоваться как вероятность или плотность распределения частиц.

Отметим, что последнее обстоятельство делает невозможным применение традиционного метода Монте-Карло для решения этого уравнения. Однако попытки обойти это затруднение предпринимаются. В последнее время уравнение Вигнера успешно привлекается для моделирования резонансно-туннельных диодов, хотя время расчетов оказывается огромным.

Формализм Ландауэра—Бюттикера

На наш взгляд, именно формализм Ландауэра—Бюттикера является наиболее приемлемым для моделирования полевых нанотранзисторов. Во-первых, он обладает ясным физическим смыслом, заключающимся в представлении работы квантового транзистора как распространение электронных волн в волноводе, что соответствует физической реальности. Во-вторых, что наиболее важно, этот метод обладает преемственностью с предшествующими подходами, основанными на классической кинетике.

Согласно этому подходу рассчитывается коэффициент прохождения из одной моды поперечного квантования на входе в канал во все другие моды поперечного квантования на выходе из канала $T(E)$, затем ток между истоком и стоком $I(V_{sd})$ при напряжении между ними V_{sd} получается из соотношения Ландауэра [24, 25]

$$I(V_{sd}) = \frac{2e}{h} \int T(E) [f_s(E) - f_d(E)] dE \quad (6.12)$$

где e — заряд электрона, h — постоянная Планка, f_s и f_d — функции распределения в истоке и стоке, соответствующие разности химических потенциалов eV_{sd} , определяемой напряжением между истоком и стоком V_{sd} . Коэффициент перед знаком интеграла происходит из кванта проводимости $G_0 = 2e^2/h$.

Коэффициенты прохождения электронов из контакта истока в контакт стока получаются путем решения задачи рассеяния для уравнения

Шредингера с учетом распределения электрического поля в канале транзистора. Электрическое поле определяется путем решения уравнения Пуассона с учетом распределения плотности носителей, полученной в результате решения уравнения Шредингера. Совместное итерационное решение обоих уравнений приводит к вычислению тока по формуле (6.12).

Несмотря на простоту и ясность формулировки метода Ландауэра—Бюттикера, он обладает скрытыми недостатками, свойственными всем кинетическим подходам, описывающим баллистический или квазibalлистический перенос, как квантовый, так и классический. Действительно, для применения формулы (6.12) необходимо знать функцию распределения частиц, влетающих в расчетную область. Вообще говоря, ее можно считать близкой к равновесной только в глубине контактов, но там присутствует очень сильное рассеяние, которое совершенно не учитывается в модели. Как правило, в работах эта проблема обходится вниманием. Нам, однако, удалось обосновать применимость метода даже в том случае, когда сильное рассеяние в контактах не принимается во внимание. Наше обоснование основано на наличии высокого потенциального барьера вблизи контактов истока и стока, который возникает из-за натекания носителей в нелегированный канал транзистора из сильнолегированных областей истока и стока. Большая часть частиц (волн) отражается от этого барьера и возвращается в контакт. Только небольшая доля частиц, имеющая достаточную энергию, проникает в контакт, поэтому можно считать эти частицы как находящиеся в равновесии с резервуаром и имеющие равновесное распределение. Это утверждение справедливо независимо от интенсивности рассеяния в контактах.

7. Решение задачи рассеяния для уравнения Шредингера с помощью T- и S-матриц (моделирование КНИ-транзистора)

7.1. Описание численного алгоритма решения

Основной задачей при моделировании транзистора типа «кремний на изоляторе» является вычисление коэффициента прохождения электрона через потенциальный барьер, сформированный затвором транзистора в области канала. Вычисление формы барьера является задачей классической и сводится к решению уравнения Пуассона с заданными граничными условиями, мы же остановимся на чисто квантовой задаче — вычислении прозрачности барьера.

Рассмотрим модифицированный метод T-матрицы (трансфер-матрицы) [29]. Главным достоинством этого метода является простота и малое время расчета по сравнению, например с методом матрицы рассеяния (S-матрицы) [30]. Это особенно важно при расчетах трехмерных (3D) структур, например, при полностью квантовом моделировании полевых нанотранзисторов. Тем не менее, существенным недостатком, сдерживавшим использование этого метода для расчета приборов, являлась катастрофическая потеря точности при работе с «затухающими» модами (evanescent modes), иногда их также называют «нераспространяющимися» модами. Затухающим модам соответствует отрицательная энергия продольного движения или, другими словами, мнимое значение продольного волнового вектора. Таким образом, затухающие моды соответствуют решениям, которые в одну сторону экспоненциально затухают, а в другую экспоненциально возрастают вдоль канала транзистора. По названной причине в расчетах, как правило, используют намного более сложный метод матрицы рассеяния (S-матрицы), где возрастающие и затухающие моды разделяются и при расчетах S-матрицы взаимно компенсируются.

Описываемый алгоритм позволяет при достаточно умеренном увеличении времени расчетов устранить вышеупомянутый недостаток. Идея метода заключается в точном учете экспоненциально возрастающих и затухающих мод (evanescent modes) посредством чисел произвольной заданной нами точности. Существо метода заключается в следующем. Если предположить, что для длины канала Δl нам для правильного расчета коэффициентов прохождения и отражения достаточно 16 значащих цифр после запятой, то для канала в n раз большей

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

длины нам потребуются числа с n раз большим количеством значащих цифр. Таким образом, для одновременной работы как с экспоненциально возрастающими, так и с экспоненциально убывающими коэффициентами при модах в разложении проходящей электронной волны необходимо линейное увеличение количества значащих цифр в используемых числах. На сегодняшний день существуют эффективные алгоритмы, позволяющие проводить математические операции над числами произвольной точности за время порядка $n \log n$, где n — число значащих цифр в числе. В результате этого увеличение длины канала в n раз приводит к увеличению времени работы нашего алгоритма также лишь в $n \log n$ раз, что соответствует практически линейному росту.

Перейдем теперь непосредственно к описанию предлагаемого метода. Мы решаем стационарное уравнение Шредингера:

$$\left[-\frac{\partial^2}{m_x \partial x^2} - \frac{\partial^2}{m_y \partial y^2} - \frac{\partial^2}{m_z \partial z^2} + U(x, y, z) \right] \psi(x, y, z) = E \psi(x, y, z). \quad (7.1)$$

Уравнение записано в атомных единицах, m_x , m_y и m_z есть отношения эффективной массы в соответствующем направлении к массе свободного электрона. Волна распространяется в направлении оси x , геометрические характеристики канала не зависят от координаты x .

Раскладываем волновую функцию в плоскости перпендикулярной оси x по собственным модам канала $\varphi_i(y, z)$:

$$\psi(x, y, z) = \sum_{i=1}^{\infty} c_i(x) * \varphi_i(y, z). \quad (7.2)$$

Полная система функций $\varphi_i(y, z)$ является решением двумерного уравнения Шредингера в плоскости перпендикулярной оси x в отсутствии потенциала:

$$\left[-\frac{\partial^2}{m_y \partial y^2} - \frac{\partial^2}{m_z \partial z^2} \right] \varphi_i(y, z) = \varepsilon_i \varphi_i(y, z). \quad (7.3)$$

В случае прямоугольного сечения канала с размерами $d_y \times d_z$, $\varphi_i(y, z)$ — есть просто произведение синусов:

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

$$\varphi_i(y, z) = \frac{2}{\sqrt{d_y d_z}} \sin\left(\frac{\pi i_y y}{d_y}\right) \sin\left(\frac{\pi i_z z}{d_z}\right). \quad (7.4)$$

Соответствующая энергия есть

$$\varepsilon_i = \frac{1}{m_y} \left(\frac{\pi i_y}{d_y}\right)^2 + \frac{1}{m_z} \left(\frac{\pi i_z}{d_z}\right)^2. \quad (7.5)$$

Если взять конечное число M мод $\varphi_i(y, z)$, получим в разложении (7.2) приближенное равенство. Подставим разложение (7.2) в уравнение (7.1), умножим слева на волновую функцию $\varphi_j^*(y, z)$ и проинтегрируем по z и по y . Поменяв местами индексы i и j , получим:

$$-\frac{\partial^2 c_i(x)}{m_x \partial x^2} + \sum_j U_{i,j}(x) c_j(x) = [E - \varepsilon_i] c_i(x). \quad (7.6)$$

На входе, где потенциал отсутствует, $E = k_x^2 / m_x + \varepsilon_j$, где ε_j - энергия j -ой моды. Будем решать уравнение (7.6) на сетке с равноотстоящими узлами x_k , $k=1..N$. Соответствующее разностное уравнение запишется как:

$$c_i(x_{k-1}) = -c_i(x_{k+1}) + 2c_i(x_k) + m_x \cdot \Delta x^2 \left(\sum_j U_{i,j}(x_k) c_j(x_k) - [E - \varepsilon_i] c_i(x_k) \right) \quad (7.7)$$

или в векторной форме:

$$\mathbf{c}(x_{k-1}) = -\mathbf{c}(x_{k+1}) + \mathbf{c}(x_k) \left[2 + m_x \Delta x^2 \left(\mathbf{U}(x_k) - [E - \mathbf{e}] \right) \right] = -\mathbf{c}(x_{k+1}) + \mathbf{A}(x_k) \mathbf{c}(x_k) \quad (7.8)$$

Уравнение можно переписать в следующем виде:

$$\begin{pmatrix} \mathbf{c}(x_{k-1}) \\ \mathbf{c}(x_k) \end{pmatrix} = \begin{pmatrix} \mathbf{A}(x_k) & -\mathbf{I} \\ \mathbf{I} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{c}(x_k) \\ \mathbf{c}(x_{k+1}) \end{pmatrix} = \mathbf{T}(x_k) \begin{pmatrix} \mathbf{c}(x_k) \\ \mathbf{c}(x_{k+1}) \end{pmatrix}. \quad (7.8)$$

Таким образом,

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

$$\begin{pmatrix} \mathbf{c}(x_{-1}) \\ \mathbf{c}(x_0) \end{pmatrix} = \prod_{k=0}^{N+1} \mathbf{T}(x_k) \begin{pmatrix} \mathbf{c}(x_{N+1}) \\ \mathbf{c}(x_{N+2}) \end{pmatrix} = \mathbf{T} \begin{pmatrix} \mathbf{c}(x_{N+1}) \\ \mathbf{c}(x_{N+2}) \end{pmatrix}. \quad (7.9)$$

Выделим из коэффициентов \mathbf{c} части соответствующие модам, которые при $E > \varepsilon_i$ распространяются в прямом направлении — $c_i^+(x)$ и в обратном направлении — $c_i^-(x)$ вдоль оси x . В случае $E < \varepsilon_i$ (evanescent modes) это будут моды соответственно экспоненциально возрастающие и затухающие.

$$c_i(x) = c_i^+(x) + c_i^-(x). \quad (7.10)$$

Из уравнения (7.6) следует, что в начале канала, где отсутствует потенциал U , изменение коэффициентов $c_i(x)$ при изменении координаты x на Δx будет:

$$c_i(x + \Delta x) = c_i^+(x) e^{i\sqrt{m_x(E - \varepsilon_i)}\Delta x} + c_i^-(x) e^{-i\sqrt{m_x(E - \varepsilon_i)}\Delta x}. \quad (7.11)$$

Выразим коэффициенты $c_i^+(x)$ и $c_i^-(x)$ через значения коэффициентов c_i в точках x и $x + \Delta x$:

$$\begin{aligned} c_i^+(x) &= b_{11}^i c_i(x) + b_{12}^i c_i(x + \Delta x) \\ c_i^-(x) &= b_{21}^i c_i(x) + b_{22}^i c_i(x + \Delta x) \end{aligned} \quad (7.12)$$

Где

$$\begin{aligned} b_{11}^i &= \left(1 - e^{-2i\sqrt{m_x(E - \varepsilon_i)}\Delta x}\right)^{-1}, \quad b_{12}^i = \left(e^{-i\sqrt{m_x(E - \varepsilon_i)}\Delta x} - e^{i\sqrt{m_x(E - \varepsilon_i)}\Delta x}\right)^{-1} \\ b_{21}^i &= \left(1 - e^{2i\sqrt{m_x(E - \varepsilon_i)}\Delta x}\right)^{-1}, \quad b_{22}^i = \left(e^{i\sqrt{m_x(E - \varepsilon_i)}\Delta x} - e^{-i\sqrt{m_x(E - \varepsilon_i)}\Delta x}\right)^{-1} = -b_{12}^i \end{aligned} \quad (7.13)$$

Следует отметить, что при $\Delta x \rightarrow 0$ коэффициенты b^i неограниченно возрастают и имеет смысл только их линейная комбинация в выражении (7.12).

Запишем также выражения коэффициентов c_i в точках x и $x + \Delta x$ через $c_i^+(x)$ и $c_i^-(x)$ в конце канала, где потенциал постоянен и равен U_{drain} :

$$\begin{aligned} c_i(x) &= \tilde{b}_{11}^i c_i^+(x) + \tilde{b}_{12}^i c_i^-(x) \\ c_i(x + \Delta x) &= \tilde{b}_{21}^i c_i^+(x) + \tilde{b}_{22}^i c_i^-(x) \end{aligned} \quad (7.14)$$

где

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

$$\tilde{b}_{11}^i = 1, \quad \tilde{b}_{12}^i = 1, \\ \tilde{b}_{21}^i = \exp\left[i\sqrt{m_x(E - \varepsilon_i - U_{drain})}\Delta x\right], \quad \tilde{b}_{22}^i = \exp\left[-i\sqrt{m_x(E - \varepsilon_i - U_{drain})}\Delta x\right]. \quad (7.15)$$

Перепишывая (7.12) и (7.14) в векторной форме и учитывая (7.9), получим связь коэффициентов $c_i^+(x)$ и $c_i^-(x)$ на входе и выходе из канала:

$$\begin{pmatrix} c^+(x_{-1}) \\ c^-(x_{-1}) \end{pmatrix} = \mathbf{B} \begin{pmatrix} c(x_{-1}) \\ c(x_0) \end{pmatrix} = \mathbf{B} \cdot \mathbf{T} \begin{pmatrix} c(x_{N+1}) \\ c(x_{N+2}) \end{pmatrix} = \mathbf{B} \cdot \mathbf{T} \cdot \tilde{\mathbf{B}} \begin{pmatrix} c^+(x_{N+1}) \\ c^-(x_{N+1}) \end{pmatrix}. \quad (7.16)$$

Матрица \mathbf{B} имеет вид:

$$\mathbf{B} = \begin{pmatrix} b_{11}^1 & 0 & 0 & b_{12}^1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & b_{11}^M & 0 & 0 & b_{12}^M \\ b_{21}^1 & 0 & 0 & b_{22}^1 & 0 & 0 \\ 0 & \ddots & 0 & 0 & \ddots & 0 \\ 0 & 0 & b_{21}^M & 0 & 0 & b_{22}^M \end{pmatrix}. \quad (7.17)$$

Для матрицы $\tilde{\mathbf{B}}$ все формулы выглядят аналогично.

Представим себе, что мы посылаем в канал l -ю моду с энергией $E = k_x^2/m_x + \varepsilon_l$, тогда на входе получим распространяющиеся моды с $\varepsilon_i < E, i = 1..i_{MPleft}$, из которых одна налетающая и i_{MPleft} отраженных, а также экспоненциально возрастающие и затухающие моды, т.е. эванесцентные моды, у которых $\varepsilon_i > E, i = i_{MPleft}..M$. На выходе также будем иметь $i_{MPright}$ прошедших распространяющихся мод с $\varepsilon_i < E + U_{drain}, i = 1..i_{MPright}$ и эванесцентные моды с $\varepsilon_i > E + U_{drain}, i = i_{MPright}..M$, где потенциал в конце канала U_{drain} уже не зависит от координат. В нашем случае он равен напряжению между истоком и стоком, умноженному на заряд электрона. Коэффициенты c_i^+ и c_i^- на входе и выходе из канала будут выглядеть следующим образом:

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

$$\mathbf{c}_{in}^+ = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1_l \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{c}_{in}^- = \begin{pmatrix} \mathbf{c}_{tr,1}^- \\ \vdots \\ \mathbf{c}_{tr,i_{MPleft}}^- \\ \mathbf{c}_{incr,i_{MPleft}+1}^- \\ \vdots \\ \mathbf{c}_{incr,M}^- \end{pmatrix}, \quad \mathbf{c}_{out}^+ = \begin{pmatrix} \mathbf{c}_{tr,1}^+ \\ \vdots \\ \mathbf{c}_{tr,i_{MPright}}^+ \\ \mathbf{c}_{decr,i_{MPright}+1}^+ \\ \vdots \\ \mathbf{c}_{decr,M}^+ \end{pmatrix}, \quad \mathbf{c}_{out}^- = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (7.18)$$

где на входе мы положили равными нулю коэффициенты при экспоненциально возрастающих при $x \rightarrow -\infty$ модах, а на выходе при модах возрастающих на $x \rightarrow +\infty$. Решая уравнение (7.16) относительно $2M$ неизвестных можно вычислить оставшиеся коэффициенты $\{c\}$. Заметим, что трансформация мод, т. е. изменение по модулю коэффициентов $\{c\}$ происходит только в неоднородном потенциале, поэтому на входе и выходе из канала, т.е. там, где потенциал постоянен, граничные условия на коэффициенты можно ставить в произвольной точке x . Потенциал фактически выравнивается внутри сильнолегированной области контактов истока и стока.

Коэффициенты прохождения и отражения получаются из соотношений

$$T = \frac{\sum_{i=1..i_{MPright}} |\mathbf{c}_{tr,i}^+|^2 k_i}{k_l} = 1 - R = 1 - \frac{\sum_{i=1..i_{MPleft}} |\mathbf{c}_{ref,i}^-|^2 k_i}{k_l}. \quad (7.19)$$

где коэффициенты c_i нормированы так, чтобы на входе коэффициент при налетающей волне содержащей l -ю моду, был равен единице.

Подходя к описанию практической реализации метода, заметим, что коэффициенты \mathbf{c} в случае нераспространяющихся мод состоят из экспоненциально возрастающей и экспоненциально убывающей частей. Поэтому для получения правильного результата, мы должны следить, чтобы разность порядков $c_i^+(x)$ и $c_i^-(x)$ как минимум не превышала количества значащих цифр в числах, используемых при расчетах на конкретной вычислительной машине. При двойной точности это 15-16 десятичных цифр. Нетрудно привести условие на максимальную длину части канала Δl , на которой еще можно работать с двойной точностью. Для этого необходимо, чтобы $e^{2\sqrt{m_s(\epsilon_{imax}-E)}\Delta l} < 10^{16}$. Чтобы проводить расчеты в канале длины большей Δl , мы используем числа произвольной точности. В этом случае каждый раз после получения

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

Т матрицы очередного участка из $m < \Delta l / \Delta x$ слоев, на котором еще не происходит катастрофической потери точности, мы умножаем результат на результирующую матрицу, состоящую из чисел повышенной точности. При таком умножении обычные числа умножаются на числа произвольной точности, каждое из которых представлено n обычными числами, поэтому временные затраты на такое умножение больше только в n раз. При моделировании транзистора нам обычно требовалось 5-7 таких перемножений. Мантисса чисел произвольной точности должна содержать $N_d > 2 \lg e \sqrt{m_x (\varepsilon_{i_{\max}} - E)} L$ значащих цифр.

В общей структуре алгоритма только решение системы линейных уравнений (7.16) требует операций умножения и деления чисел произвольной точности. Как уже было сказано выше, в настоящее время существуют эффективные алгоритмы, позволяющие проводить математические операции над числами произвольной точности за время порядка $N \log N$, где n — число значащих цифр в числе. Это увеличивает времени работы алгоритма также лишь в $N \log N$ раз. Таким образом, мы имеем алгоритм, с помощью которого можно рассчитывать рассеяние электронной волны с учетом произвольного числа нераспространяющихся мод, в канале произвольной длины. Время работы алгоритма, как показали наши расчеты, по сравнению с обычным методом Т-матрицы, который не позволяет проводить реальные расчеты, увеличивается лишь линейно от длины канала.

Представленный метод решения задачи рассеяния для трехмерного (3D) уравнения Шредингера может найти широкое применение при моделировании различных нанoeлектронных структур, в частности, интерференционных структур с длинными каналами. Метод оказался применимым для широкого класса задач, основанных на решении волнового уравнения в неоднородных средах, в частности, для моделирования наноструктурированных солнечных элементов.

По сравнению с известным методом S-матриц предложенный нами метод дает значительное (в десятки раз) ускорение расчетов. Это вполне обосновано. Метод S-матриц предполагает решение задачи на собственные значения в каждом сечении неоднородного волновода. Только после этого можно выделить затухающие и нарастающие моды, и учитывать их отдельно. В методе Т-матриц используется разложение по поперечным модам идеального волновода, а матричные элементы потенциала в канале описывают взаимодействие мод.

В настоящей главе мы остановимся только на моделировании КНИ транзисторов, которые имеют самую близкую перспективу использования в больших интегральных схемах.

7.2. Результаты моделирования

На основе описанного выше метода в Физико-технологическом институте РАН была разработана программа моделирования для расчета характеристик КНИ-транзистора [31, 32]. Программа основана на расчете коэффициентов прохождения электронов из контакта истока в контакт стока путем решения уравнения Шредингера с учетом распределения электрического поля в канале транзистора. Для обеспечения устойчивости решения уравнения Шредингера без потери точности применяется разложение волновой функции электронов в канале по конечному числу поперечных мод, а также использование чисел произвольной точности, как это описано в предыдущем разделе. Для расчета самосогласованного поля решается уравнение Пуассона. Ток рассчитывается согласно соотношению Ландауэра-Бюттикера (6.12), которое, фактически, основано на представлении канала нанотранзистора как квантового провода.

В настоящей работе для демонстрации расчета коэффициентов прохождения были выбраны следующие геометрические параметры структуры: длина затвора 10нм, спейсеры по 5нм, толщина кремния 2нм, ширина канала 5нм, эффективная толщина подзатворного окисла 1.5нм. Легирование контактов истока и стока 10^{20}см^{-3} . Расчетная область включала 5нм контактов, что оказалось достаточным для правильного описания самосогласованного потенциального барьера на границах канала.

Рассчитанные коэффициенты прохождения для различных долин зоны проводимости кремния при нулевом напряжении на стоке $V_D = 0$ представлены на Рис. 7.1. Интерференционные пики на зависимостях для 4 случайных примесей в канале более отчетливы для электронов из долин [100] и [010], чем из долин [001]. Это объясняется значительно меньшей массой продольного движения электронов в канале (Рис. 6.12), следовательно, значительно большей длиной волны. Следует отметить, что при указанных размерах структуры преимущественно электроны только двух долин [100] дают вклад в ток, поскольку именно эти долины имеют наименьшую энергию поперечного квантования в канале. Надо особо отметить, что в конечном итоге это обстоятельство весьма мало сказывается на токе транзистора в открытом состоянии.

Полученный немонотонный характер зависимости коэффициента прохождения как раз отражает квантовую природу процессов внутри канала транзистора, а именно, интерференцию волн. При классическом поведении электронов в канале коэффициент прохождения при

Решение задачи рассеяния для уравнения Шредингера с помощью T- и S-матриц (моделирование КНИ-транзистора)

некоторой энергии электронов резко бы изменялся от 0 до 1. Следует заметить, что, несмотря на то, что зависимости коэффициента прохождения электронов изломанные, итоговые выходные вольт-амперные характеристики транзистора оказываются вполне монотонными. В этом заключается, пожалуй, главный результат моделирования.

Другим предназначением разработанной программы моделирования является выяснение влияния интерференции на самосогласованном потенциале и случайных заряженных центрах в канале на характеристики транзистора. Действительно, не раз высказывались опасения, что случайные заряженные центры (дефекты и примеси) в канале транзисторов могут привести к значительному разбросу характеристик транзисторов. Это могло бы стать препятствием на пути создания больших быстродействующих логических схем в будущем. Аналогичные опасения вызывали и неровности (шероховатости) стенок канала.

Расчеты, проведенные с помощью программы моделирования, позволили выяснить степень влияния реалистических неоднородностей канала на зависимости тока транзистора от напряжения на стоке (Рис. 7.2, Рис. 7.3). Как видно из Рис. 7.2, наличие ступеньки снижает ток открытого состояния транзистора, что означает сдвиг порогового напряжения. Особенно эта неоднородность канала сказывается на токе транзистора, когда она находится вблизи контакта истока. Одиночный положительно заряженный центр увеличивает ток транзистора (Рис. 7.3), что также связано со сдвигом порогового напряжения. Предварительные расчеты показали, что разброс характеристик нанотранзисторов, связанный с указанными выше неоднородностями канала, составляет 5-20%. Хотя эти исследования пока носят предварительный характер, они указывают на то, что разброс не является катастрофическим, как предполагалось раньше [37, 38], однако, по всей видимости, потребует совершенствования технологии процессов изготовления интегральных схем на основе нанотранзисторов.

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

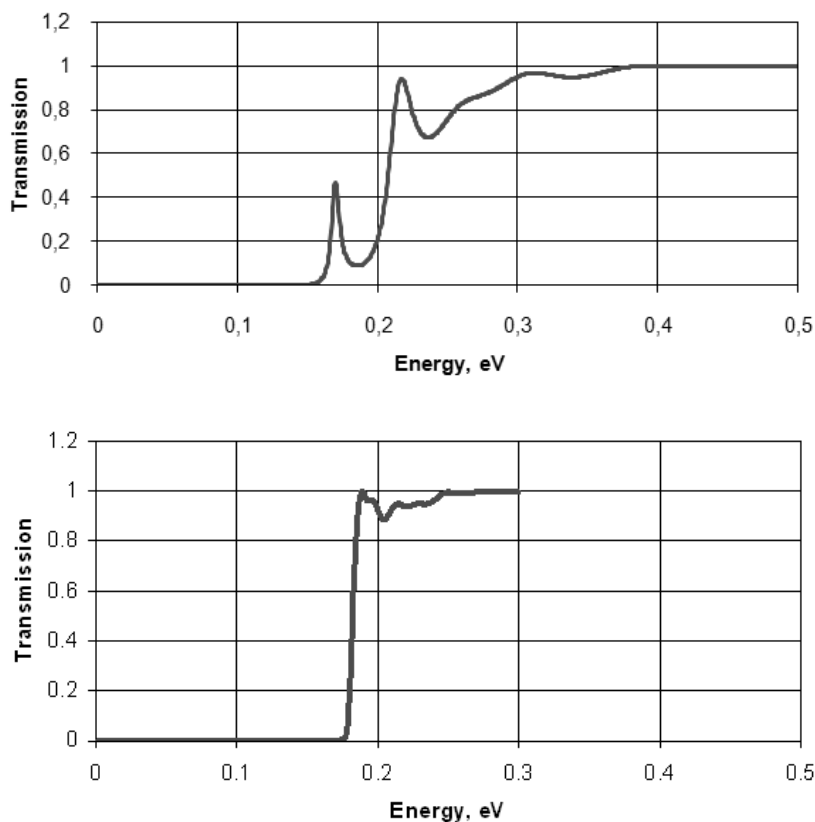


Рис. 7.1. Зависимость коэффициента прохождения от продольной энергии электронов из долин [100] и [010] (сверху) и долины [001] (снизу) в канале транзистора. Присутствуют 4 случайные примеси в канале [31, 32].

Решение задачи рассеяния для уравнения Шредингера с помощью Т- и S-матриц (моделирование КНИ-транзистора)

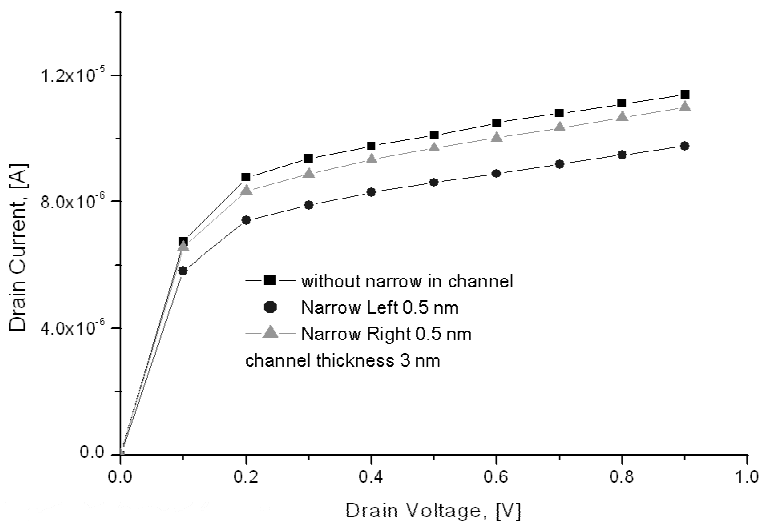


Рис. 7.2. Зависимость тока транзистора от напряжения на затворе для гладких стенок канала и при различном расположении ступеньки (вблизи контакта истока или стока) [31, 32].

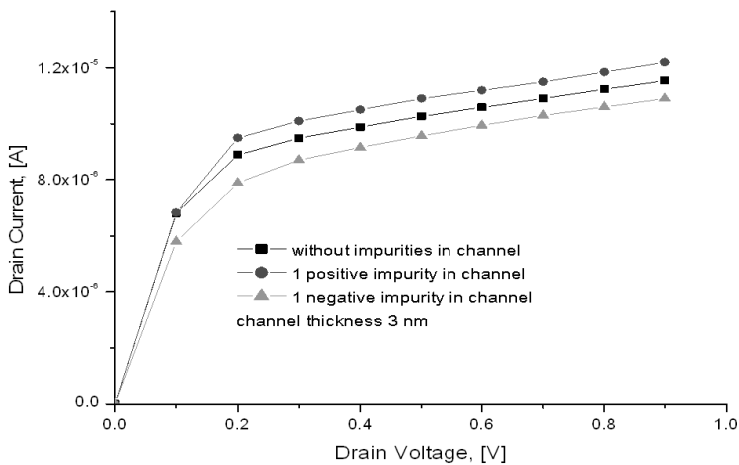


Рис. 7.3. Зависимость тока транзистора от напряжения на затворе для идеального канала и при различном расположении одиночного заряженного центра (вблизи контакта истока или стока). [31, 32]

7.3. Выводы о результатах моделирования

Кремниевый полевой транзистор еще долгое время будет сохранять свое лидерство и в области нанометровых размеров как основной элемент сверхбольших интегральных схем. При нанометровых размерах этот транзистор становится квантовым объектом. Его работа более похожа на распространении заряженных волн в волноводе, чем на пролет частиц или протекание жидкости, как было прежде. Моделирование нанотранзисторов требует применения методов квантовой механики.

Что касается создания и использования альтернативных приборов, то пока все они сталкиваются с очень серьезными трудностями, поскольку в значительной степени уступают кремниевым полевым транзисторам по совокупности своих характеристик, превосходя их только в отдельных показателях.

8. Моделирование новых электронных приборов на основе графена

8.1. Квантовая гидродинамическая модель транзистора на основе графена

Кинетическое уравнение Больцмана для носителей заряда в графене

Об удивительных электронных свойствах графена и нанотрубок уже было рассказано в предыдущих главах данного пособия. Сейчас перед нами встает задача построения математической модели приборов на основе графена, в частности, полевого транзистора на графене (англ. GFET). В первую очередь желательно получить модель для расчета статических зависимостей «ток стока — напряжение на затворе» и «ток стока — напряжение на стоке».

Отправной точкой для расчета кинетических характеристик является уравнение Больцмана; далее, в зависимости от роли столкновений, строятся либо баллистическое (бесстолкновительное), либо гидродинамическое приближения. В гидродинамическом приближении предполагается, что частота столкновений между носителями заряда настолько велика, что в каждой точке успевает установиться локальное равновесие, т.е. в каждой точке можно определить свои значения температуры, концентрации и среднего импульса частиц.

Мы остановим свой выбор именно на гидродинамическом приближении, т.к. согласно экспериментальным данным и строгим расчетам, частота электрон-электронных столкновений в графене действительно очень высока (по сравнению, например, с частотой столкновения с фононами). Сильное электрон-электронное рассеяние не приведет к уменьшению проводимости, т.к. столкновения между ними не могут поменять суммарного импульса электронной системы¹. Иное дело — столкновения электронов с дырками, они могут привести к существенному снижению проводимости. Гидродинамика графена будет существенно двухжидкостной — мы выпишем кинетические уравнения для функции распределения электронов и дырок, а затем учтем взаимодействие между ними, добавив полуфеноменологическое слагаемое с «силой трения».

¹ За исключением так называемых процессов Пайерлса, когда суммарный импульс сталкивающихся частиц оказывается в соседней зоне Бриллюэна.

Кинетическое уравнение Больцмана для функции распределения f частиц с произвольным спектром имеет вид

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial \varepsilon}{\partial \vec{p}} \frac{\partial f}{\partial \vec{r}} + \vec{F} \frac{\partial f}{\partial \vec{p}} = \text{St}\{f\}, \quad (8.1)$$

где символы d/dt и $\partial/\partial t$ означают полную и частную производные по времени, соответственно, $\vec{v} = \partial \varepsilon / \partial \vec{p}$ — есть скорость частицы, \vec{F} — сила, действующая на частицу. В случае энергетический спектр является линейным вблизи минимумов зоны Бриллюэна

$$\varepsilon(p) = v_F |p|, \quad (8.2)$$

поэтому скорость частицы равна

$$\vec{v} = \frac{\partial \varepsilon(p)}{\partial \vec{p}} = v_F \frac{\vec{p}}{p}. \quad (8.3)$$

Соотношение (8.3) символизирует тот факт, что в графене все частицы движутся с постоянной по модулю скоростью, равной скорости Ферми.

Уравнение непрерывности

Для того чтобы перейти от уравнения Больцмана к уравнениям гидродинамики, содержащим только зависимость макроскопических параметров (скорости и плотности) от координаты и времени, необходимо усреднить (8.1) по фазовому пространству — проинтегрировать по $d\Gamma = g d^2 \vec{p} / (2\pi \hbar)^2$:

$$\frac{\partial}{\partial t} \left(\int f d\Gamma \right) + \frac{\partial}{\partial \vec{r}} \left(\int \vec{p} v_F \frac{\partial f}{\partial \vec{p}} d\Gamma \right) + \int \vec{F} \frac{\partial f}{\partial \vec{p}} d\Gamma = \int \text{St}\{f\} d\Gamma, \quad (8.4)$$

Как и в классической гидродинамике, мы должны получить при интегрировании уравнение непрерывности, выражающее сохранение числа частиц в системе вида

$$\frac{\partial n}{\partial t} + \text{div} \vec{J} = 0. \quad (8.5)$$

Моделирование новых электронных приборов на основе графена

Уравнение (8.4) действительно приводится к такому виду, если учесть, что

$\int f d\Gamma$ есть концентрация частиц в расчете на единицу площади;

Третье слагаемое можно преобразовать в поверхностный интеграл по формуле $\int_V \nabla f dV = \oint_S f d\vec{S}$. Выбирая поверхность интегрирования в импульсном пространстве достаточно удаленной (где $f(p) \rightarrow 0$), можно добиться обнуления этого слагаемого;

$\int \text{St}\{f\} d\Gamma = 0$ в силу сохранения числа частиц в системе.

Назовем теперь потоком частиц величину, стоящую под знаком пространственной производной:

$$\vec{J} = \int f v_F \frac{\vec{p}}{p} d\Gamma, \quad (8.6)$$

где усреднение ведется по локально равновесной функции распределения. Мы можем получить более явное выражение для потока, если будем считать, что искомая функция распределения есть функция Ферми с ненулевым средним импульсом¹:

$$f = \left(1 + \exp \left[\frac{\varepsilon(\vec{p} - \vec{P}(\vec{r})) - \mu(\vec{r})}{T(\vec{r})} \right] \right)^{-1}. \quad (8.7)$$

С помощью замены переменной $\vec{p} - \vec{P}(\vec{r}) \rightarrow \vec{p}$ можно перейти к усреднению по сферически симметричной функции распределения (элемент объема при этой замене не поменяется):

$$\vec{J} = v_F \left\langle \frac{\vec{p} + \vec{P}}{|\vec{p} + \vec{P}|} \right\rangle, \quad (8.8)$$

¹ Такое приближение исторически называется пятимоментным – по числу характеризующих его неизвестных параметров: три компоненты импульса, температура и химический потенциал. В графене, конечно, у среднего импульса всего две компоненты – в силу двумерности.

В данном случае мы рассчитываем средние величины на единицу площади, поэтому, например, $\langle p \rangle$ будет измеряться в кг/м·с. Предполагая, что средний импульс макроскопического движения жидкости \bar{P} мал, можно разложить выражение для потока в ряд Тейлора и проинтегрировать по p по формуле:

$$J_i \approx v_F \left\langle \frac{p_i + P_i}{p} (1 - p_k P_k) \right\rangle = \frac{v_F P_i}{2} \left\langle \frac{1}{p} \right\rangle. \quad (8.9)$$

При вычислении мы воспользовались тем фактом, что в двумерном пространстве среднее от произведения координат единичных векторов $\langle n_i n_k \rangle = \delta_{ik} / 2$. Усреднение обратного импульса в (8.9) ведется уже по равновесной функции распределения!

Итак, первое из уравнений гидродинамики для ферми-частиц с линейным спектром принимает вид:

$$\frac{\partial n}{\partial t} + \frac{v_F}{2} \operatorname{div} \left(\bar{P} \left\langle \frac{1}{p} \right\rangle \right) = 0. \quad (8.10)$$

Уравнения Эйлера для электронов и дырок

Далее необходимо получить еще два уравнения, описывающие собственно динамику жидкости под действием внешних сил (уравнение Эйлера) и распределение энергии в жидкости (уравнение теплопроводности). Их вывод принципиально не отличается от сделанного выше: для получения уравнения Эйлера необходимо умножить кинетическое уравнение Больцмана на компоненту импульса p_i и проинтегрировать по $d\Gamma$, для получения уравнения теплопроводности — умножить на ε и проинтегрировать по $d\Gamma$. Существенно, что при этом правая часть с интегралом столкновений обратится в нуль в силу закона сохранения импульса и энергии при электрон-электронных столкновениях.

Мы не будем заострять внимание на дальнейших математических выводах уравнений, т.к. уже указали принципиальный путь; выпишем лишь окончательный результат:

$$\frac{\partial (n P_i)}{\partial t} + \frac{\partial \Pi_{ij}}{\partial x_j} - n F_i = 0, \quad \Pi_{ij} \approx v_F \langle p \rangle \frac{\delta_{ij}}{2} \quad \text{— уравнение Эйлера} \quad (8.11)$$

Моделирование новых электронных приборов на основе графена

$$\frac{\partial \langle \varepsilon \rangle}{\partial t} + v_F^2 \operatorname{div} (n \vec{P}) - (\vec{F}, \vec{J}) = 0 \quad \text{— уравнение теплопроводности.} \quad (8.12)$$

Напомним, что концентрация в данных уравнениях считается на единицу площади и имеет размерность см^{-2} , а под знаком усреднения подразумевается величина, взятая на единицу площади.

В дальнейшем мы не будем интересоваться распределением температуры по образцу, поэтому в нашей гидродинамической системе останутся лишь уравнения непрерывности и Эйлера.

Но уравнения (8.10) и (8.11) выписаны для существенно одножидкостной модели, реально в графеновых структурах существуют и электроны, и дырки, причем с сильным взаимным трением. Для различения электронов и дырок снабдим величины, относящиеся к ним, символами e и h . В первом приближении сила взаимного трения будет пропорциональна разности импульсов электронной и дырочной подсистем:

$$\vec{F}_{fr\,eh} = v_{eh} \frac{n_h}{n_e + n_h} (\vec{P}_h - \vec{P}_e). \quad (8.13)$$

Здесь мы ввели частоту электронно-дырочных столкновений v_{eh} , значение которой может быть определено из эксперимента или из микроскопического рассмотрения процесса рассеяния. Выбор безразмерного множителя, составленного из концентраций электронов и дырок соответствует сохранению полного импульса электронно-дырочной системы при таком виде трения.

Аналогичным образом можно учесть и рассеяние электронов на дефектах решетки и фононах. «Сила трения», отвечающая за этот процесс, запишется в виде

$$\vec{F}_{fr\,i} = -v_i \vec{P}_e, \quad (8.14)$$

где v_i — частота столкновений с внутренними дефектами решетки (индекс i — от англ. *intrinsic* — внутренний).

Теперь, добавив в (8.11) к обычной электрической силе $\vec{F}_e = |e| \nabla \varphi$ две силы трения, мы получим модифицированное уравнение Эйлера:

$$\frac{\partial (n_e P_{ei})}{\partial t} + \frac{\partial \Pi_{eij}}{\partial x_j} = n_e \left(|e| \frac{\partial \varphi}{\partial x_i} - v_i P_{ei} - v_{eh} \frac{n_h}{n_e + n_h} (P_{ei} - P_{hi}) \right). \quad (8.15)$$

Для дырок можно получить совершенно аналогичное уравнение — следует лишь заменить индекс e на h и поменять знак заряда:

$$\frac{\partial(n_h P_{hi})}{\partial t} + \frac{\partial \Pi_{hij}}{\partial x_j} = n_h \left(-|e| \frac{\partial \varphi}{\partial x_i} - v_i P_{hi} - v_{eh} \frac{n_e}{n_e + n_h} (P_{hi} - P_{ei}) \right). \quad (8.16)$$

Посчитав количество уравнений (2 — непрерывности и 4 — Эйлера) и количество неизвестных (2 концентрации, 2 двухкомпонентных импульса и электрический потенциал) мы приходим к необходимости добавления еще одного, замыкающего уравнения, в систему. Им будет уравнение Пуассона, причем мы сразу выпишем его решение в предположении и о том, что систему «графен — верхний затвор» можно рассматривать как плоский конденсатор. В этом случае связь между концентрацией носителей и потенциалом будет иметь вид:

$$|e|(n_e - n_h) = C(V_G - \varphi(r)), \quad (8.17)$$

где C есть емкость системы на единицу площади:

$$C = \frac{\varepsilon}{4\pi d}, \quad (8.18)$$

d — расстояние между затвором и слоем графена.

Теперь нам осталось выбрать неизвестные функции в уравнениях (8.10), (8.15), (8.16) и (8.18) и поставить на них корректные граничные условия. Мы сделаем это двумя способами: первый будет предназначен для численных решений, второй — для полуаналитических оценок.

Способ 1: Неизвестные функции μ_e , μ_h , P_e , P_h , φ .

Полученную нами систему уравнений следует немного упростить: во-первых, двумерная задача фактически сводится к одномерной в силу наличия выделенного направления — направления электрического поля, совпадающего с направлением тока. Во-вторых, интересуясь стационарным режимом работы транзистора, мы можем пренебречь членами с временными производными. Тогда система приводится к виду:

$$\begin{aligned}
 \frac{\partial}{\partial x} \left(P_e \left\langle \frac{1}{p_e} \right\rangle \right) &= 0, \quad \frac{\partial}{\partial x} \left(P_h \left\langle \frac{1}{p_h} \right\rangle \right) = 0, \\
 \frac{v_F}{2} \frac{\partial \langle p_e \rangle}{\partial x} &= n_e \left(|e| \frac{\partial \varphi}{\partial x} - v_i P_e - v_{eh} \frac{n_h}{n_e + n_h} (P_e - P_h) \right), \\
 \frac{v_F}{2} \frac{\partial \langle p_h \rangle}{\partial x} &= n_e \left(-|e| \frac{\partial \varphi}{\partial x} - v_i P_e - v_{eh} \frac{n_e}{n_e + n_h} (P_h - P_e) \right), \\
 |e| (n_e - n_h) &= C (V_G - \varphi(r)).
 \end{aligned} \tag{8.19}$$

Упрощения на этом не заканчиваются: вспомним теперь определения средних величин:

$$\begin{aligned}
 n_e &= \frac{2\pi g}{(2\pi\hbar)^2} \int_0^\infty \left[1 + \exp\left(\frac{pv_F - \mu_e}{T}\right) \right] p dp, \quad n_h = \frac{2\pi g}{(2\pi\hbar)^2} \int_0^\infty \left[1 + \exp\left(\frac{pv_F + \mu_h}{T}\right) \right] p dp, \\
 \langle p_e \rangle &= \frac{2\pi g}{(2\pi\hbar)^2} \int_0^\infty \left[1 + \exp\left(\frac{pv_F - \mu_e}{T}\right) \right] p^2 dp, \quad \langle p_h \rangle = \frac{2\pi g}{(2\pi\hbar)^2} \int_0^\infty \left[1 + \exp\left(\frac{pv_F + \mu_h}{T}\right) \right] p^2 dp.
 \end{aligned} \tag{8.20}$$

Дифференцированием этих выражений по координате легко получить связь производных концентрации и среднего импульса единицы площади с производными химического потенциала:

$$\begin{aligned}
 \frac{\partial n_e}{\partial x} &= \frac{1}{v_F} \left\langle \frac{1}{p_e} \right\rangle \frac{\partial \mu_e}{\partial x}, \quad \frac{\partial n_h}{\partial x} = -\frac{1}{v_F} \left\langle \frac{1}{p_h} \right\rangle \frac{\partial \mu_h}{\partial x}, \\
 \frac{\partial \langle p_e \rangle}{\partial x} &= \frac{2n_e}{v_F} \frac{\partial \mu_e}{\partial x}, \quad \frac{\partial \langle p_h \rangle}{\partial x} = -\frac{2n_h}{v_F} \frac{\partial \mu_h}{\partial x}.
 \end{aligned} \tag{8.21}$$

Обозначив теперь потоки электронов и дырок (не зависящие от координаты), как J_e и J_h , мы приходим к более простому виду системы уравнений гидродинамики:

$$\frac{v_F P_e}{2} \left\langle \frac{1}{p_e} \right\rangle = J_e, \quad \frac{v_F P_h}{2} \left\langle \frac{1}{p_h} \right\rangle = J_h, \tag{8.22}$$

$$\begin{aligned}\frac{\partial}{\partial x}(\mu_e - |e|\varphi) &= -v_i P_e - v_{eh} \frac{n_h}{n_e + n_h} (P_e - P_h), \\ -\frac{\partial}{\partial x}(\mu_h - |e|\varphi) &= -v_i P_h - v_{eh} \frac{n_e}{n_e + n_h} (P_h - P_e), \\ \frac{|e|}{v_F} \left(\left\langle \frac{1}{P_e} \right\rangle \frac{\partial \mu_e}{\partial x} + \left\langle \frac{1}{P_h} \right\rangle \frac{\partial \mu_h}{\partial x} \right) &= -C \frac{\partial \varphi}{\partial x}\end{aligned}$$

В левых частях уравнений Эйлера находится слагаемое, которое можно назвать «обобщенной силой» — оно учитывает градиент и электрического, и химического потенциалов (соответственно, диффузионный и дрейфовый ток). Эта сила приравняется к силе трения. Вся «квантовость» уравнений гидродинамики теперь содержится лишь в определениях концентрации и среднего обратного импульса с помощью функции распределения Ферми. Далее уравнения удобнее всего привести к паре неизвестных функций $\mu_e(x)$ и $\mu_h(x)$, избавившись от компонент вектора импульса с помощью уравнений непрерывности, и от градиента электрического потенциала — с помощью уравнения Пуассона.

Вернемся теперь к моделируемой транзисторной структуре и поставим граничные условия для функций $\mu_e(x)$ и $\mu_h(x)$. В контактах стока ($x=0$) и истока ($x=L$) в силу значительной электронно-дырочной рекомбинации эти функции совпадают:

$$\mu_e(0) = \mu_h(0), \mu_e(L) = \mu_h(L). \quad (8.23)$$

Если считать исток заземленным ($\varphi(0)=0$), а сток — находящимся под потенциалом V_D ($\varphi(L)=V_D$), то уровни химического потенциала в контактах могут быть найдены из численного решения уравнения Пуассона:

$$n_e(\mu(0)) - n_h(\mu(0)) = CV_G/|e|, n_e(\mu(L)) - n_h(\mu(L)) = C(V_G - V_D)/|e|. \quad (8.24)$$

Фактически, система уравнений (8.22) сводится к дифференциальным уравнениям первого порядка относительно μ_e и μ_h с двумя неизвестными алгебраическими параметрами J_e, J_h . К системе добавлены граничные условия (8.24), что позволяет решать ее, к примеру, методом стрельбы. Следует отметить, что введение разных химиче-

Моделирование новых электронных приборов на основе графена

ских потенциалов для электронов и дырок отвечает предположению о слабой рекомбинации между носителями заряда в графене, которое еще необходимо проверить.

Численное решение для хода химических потенциалов в этом случае показано на Рис. 8.1 — ход этих кривых аналогичен поведению квазиуровней Ферми в полупроводниковом диоде. Заметим также, что предположение о слабой электрон-дырочной рекомбинации приводит к возможности создания инверсной населенности в слое графена.

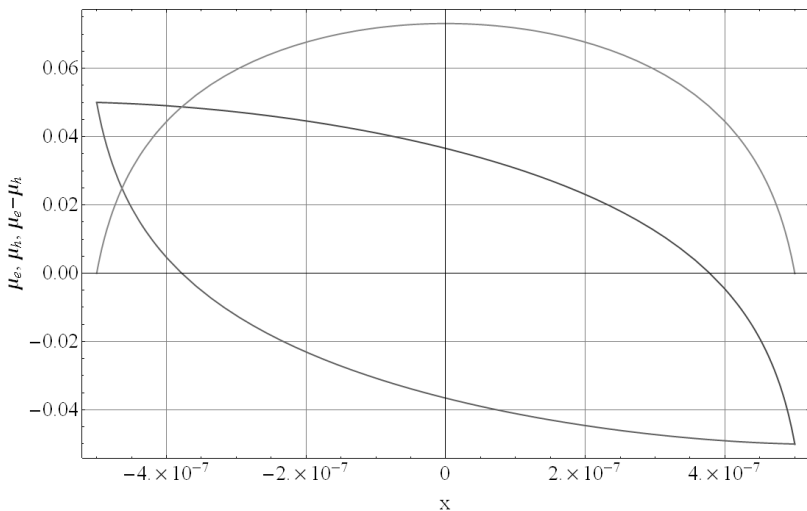


Рис. 8.1. Ход химических потенциалов электронов, дырок и их разности в графеновом транзисторе с длиной канала 1 мкм.

Способ 2: Неизвестные функции n_e , n_h , V_e , V_h , φ

Второй способ выбора неизвестных примечателен тем, что он подчеркивает существенное отличие электронно-дырочной жидкости в графене от жидкости классической. Напомним, что при решении гидродинамических задач неизвестными функциями являются концентрация (плотность частиц) и их скорость. Функции распределения частиц по энергиям в уравнения классической гидродинамики вообще не входят — они используются только на этапе вывода.

В графене мы также можем ввести «гидродинамическую массу» частицы, она, естественно, не будет совпадать с массой свободного электрона. Согласно классическому определению потока

$$J = n\vec{v} = n \frac{\vec{P}}{m}. \quad (8.25)$$

Сравнивая это выражение с определением потока в графене, мы приходим к выражению для гидродинамической массы носителя заряда в виде

$$m_{e(h)} = \frac{2n_{e(h)}}{v_F \langle p_{e(h)}^{-1} \rangle}. \quad (8.26)$$

В чистом графене, где уровень Ферми проходит строго между валентной зоной и зоной проводимости, гидродинамическая масса электрона (и дырки) составляет десятые доли массы свободного электрона. Важным отличием введенной таким образом массы от массы классической частицы является ее зависимость от химического потенциала или, что то же самое, от плотности частиц.

Мы не будем выписывать уравнения гидродинамики явно, используя массу в качестве параметра, т.к. тензор вязких напряжений может быть выражен через массу явно лишь при нулевой температуре. Мы ограничимся приближением дрейфового тока, считая изменение химического потенциала с координатой достаточно медленным. В этом случае уравнения Эйлера оказываются очень похожими на уравнения динамики Ньютона

$$\begin{aligned} |e|E &= -v_i m_e V_e - v_{eh} \frac{n_h}{n_e + n_h} (m_e V_e - m_h V_h), \\ -|e|E &= -v_i m_h V_h - v_{eh} \frac{n_e}{n_e + n_h} (m_h V_h - m_e V_e), \end{aligned} \quad (8.27)$$

с той лишь особенностью, что массы и концентрации рассчитываются согласно квантовой функции распределения Ферми. Полная плотность тока электронов и дырок дается выражением

$$J = |e|(J_h - J_e) = |e|(n_h V_h - n_e V_e). \quad (8.28)$$

Исходя из (8.27) и (8.28), нетрудно получить выражение для локальной проводимости:

$$G = \frac{\partial J}{\partial E} = \frac{e^2}{v_i + v_{eh}} \left[\left(\frac{n_e}{m_e} + \frac{n_h}{m_h} \right) + \frac{v_{eh}}{v_i} \left(\frac{n_e}{m_e} - \frac{n_h}{m_h} \right) \left(\frac{n_e - n_h}{n_e + n_h} \right) \right]. \quad (8.29)$$

При бесконечно малом смещении напряжения между стоком и истоком все величины в правой части равенства (8.29) определяются только локальным потенциалом в данной точке, соответственно и проводимость как функция координаты выглядит как $G(x) \equiv G(V_G - \varphi(x))$. Локальный потенциал меняется от значения $\varphi(0) = 0$ до $\varphi(L) = V_D$. Для вычисления полного тока транзистора (не зависящего от координаты) мы применим следующий трюк: запишем закон Ома в локальной форме

$$I = JW = -G(V_G - \varphi(x)) \frac{d\varphi(x)}{dx} \quad (8.30)$$

и, учтя независимость полного тока от координаты, проинтегрируем (8.30) по dx от нуля до длины канала L :

$$I = -\frac{W}{L} \int_0^{V_D} G(V_G - \varphi) d\varphi. \quad (8.31)$$

В итоге для расчета полного тока транзистора мы избавились от необходимости рассчитывать распределения потенциала по координате. Конечно же, расчет правой части уравнения (8.29) как функции потенциала может быть выполнен только численно, т.к. интегралы в выражениях для концентрации (8.20) не берутся аналитически.

Рассчитанные по формуле (8.31) зависимости полного тока от напряжения на сток-исток приведены на Рис. 8.2. Токи и напряжения выражены в обезразмеренном виде (I/I_0 , V/V_0), причем

$$V_0 = \frac{2|e|d(kT)^2 D}{\hbar^2 v_F^2 \kappa}, \quad I_0 = G_0 \frac{W}{L} V_0. \quad (8.32)$$

В данных формулах D — толщина подзатворного диэлектрика, W и L — ширина и толщина слоя графена, G_0 — проводимость графена в слабом поле, определяемая формулой (8.29) при $n_e = n_h = n(\mu = 0)$.

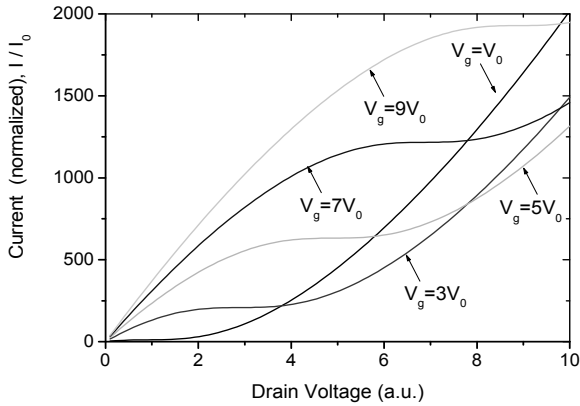


Рис. 8.2. Рассчитанная зависимость $I(V_D)$ (образец предполагается достаточно чистым: $v_{eh}/v_i = 100$)

Аналогичным образом можно построить характеристики «ток-напряжение на затворе» (Рис. 8.3). V-образная форма характеристик находится в хорошем согласии с экспериментальными данными по проводимости графенового слоя [4].

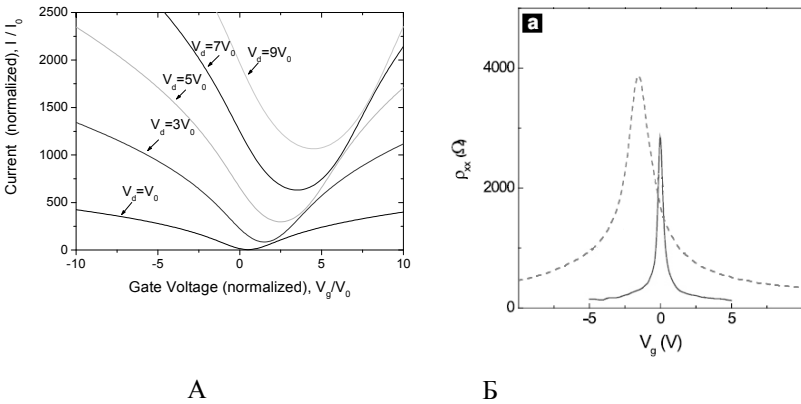


Рис. 8.3. Рассчитанная зависимость $I(V_G)$ ($v_{eh}/v_i = 100$, А) и экспериментальная зависимость сопротивления графенового слоя от напряжения на затворе (Б)

8.2. Баллистическая модель транзистора на двухслойном графене

В рассматриваемом транзисторе двухслойный графен находится между двумя электродами — верхним и нижним затвором, расстояния от которых до слоя графена равны соответственно d_1 и d_2 . Межэлектродное пространство заполнено диэлектриком (в приводимых ниже расчетах — диоксидом кремния, диэлектрическая проницаемость $\kappa = 4$). Для повышения эффективности управления концентрацией носителей заряда слой графена располагается ближе к верхнему электроду, чем к нижнему: $d_1 = 1$ нм и $d_2 = 10$ нм.

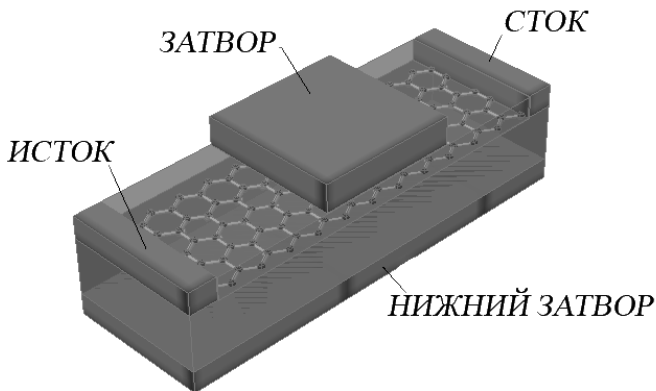


Рис. 8.4. Эскизное изображение моделируемого транзистора

Для начала мы ограничимся вычислением проводимости двухслойного графена — только она может быть вычислена аналитически, и лишь при нулевой температуре. Однако, ставя своей задачей компьютерное моделирование данной системы, впоследствии мы выпишем уравнения, описывающие работу транзистора при ненулевой температуре и при больших напряжениях между стоком и истоком.

В первую очередь для решения данной задачи необходимо найти высоту потенциального барьера, формируемого для электронов напряжением на затворе. В дальнейшем, зная высоту потенциального барьера, мы вычислим потоки надбарьерных

электронов, летящих из контакта истока в контакт стока и обратно, и по их разности определим ток транзистора. При отсутствии приложенного напряжения предполагается заполнение слоя графена электронами, симметризация получаемых характеристик с учетом дырок не представляет особого труда.

Потенциальную энергию электронов в области затвора можно найти из системы уравнений электростатики для транзистора на Рис. 8.4:

$$\begin{cases} E_1 d_1 + E_2 d_2 = V_G, \\ \kappa_1 E_1 - \kappa_2 E_2 = 4\pi e n(V_C, T), \\ E_2 d_2 = V_C, \end{cases} \quad (8.33)$$

в которых E_1 и E_2 — напряженности поперечного электрического поля в области верхнего и нижнего затворов, соответственно, V_G — потенциал верхнего затвора, потенциал нижнего затвора предполагается равным нулю, V_C есть потенциал в канале транзистора (на слое графена), определяющий высоту потенциального барьера для электронов при нулевых напряжениях на стоке и истоке.

Принимая во внимание соотношение между толщинами диэлектриков $d_1 / d_2 \ll 1$, легко получить соотношение между потенциалом затвора и потенциалом слоя графена:

$$V_G = V_C + \frac{4\pi e d}{\kappa_1} n(V_C, T). \quad (8.34)$$

Нашей дальнейшей задачей будет обращение функции $V_G(V_C)$ — потенциал верхнего затвора обычно задан, в то время как потенциал в графене подлежит определению. Вычленение концентрации как функции потенциала может быть произведено по формуле, аналогичной (5.29), с подстановкой спектра двухслойного графена (5.37) вместо pv_F . Нам, однако, необходимо сделать несколько уточнений по поводу пределов интегрирования в импульсном пространстве.

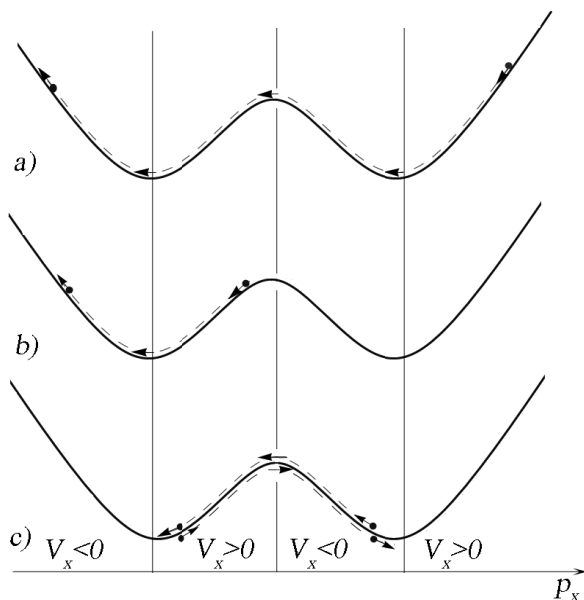


Рис. 8.5. Движение электронов вдоль дисперсионной кривой: а) электрон на «полях мексиканской шляпы» отражается от потенциального барьера б) электрон на «колпаке мексиканской шляпы» отражается от барьера в) электрон на колпаке мексиканской шляпы локализован внутри канала транзистора

Оказывается, что состояния электронов, соответствующие «колпаку мексиканской шляпы» (с импульсами $|p| < p_{\min}$) оказываются локализованными в канале транзистора, несмотря на то, что находятся на вершине потенциального барьера, а не в потенциальной яме — причиной этому является их отрицательная эффективная масса. Схематически движение такого электрона изображено на Рис. 8.5 в: пусть изначально он находится вблизи правого минимума дисперсионной кривой и движется влево ($v_x = dE/dp_x < 0$). Дойдя до истока транзистора, он увеличивает свою кинетическую энергию и одновременно с этим уменьшает модуль скорости, достигнув максимума дисперсионной кривой, он останавливается и разворачивается обратно в сторону канала транзистора. Вблизи левого минимума дисперсионной кривой он выходит из области действия электрической силы и движется с постоянной скоростью вправо и около стока транзистора разворачивается по тому же сценарию. Столь необычный характер

движения электронов свойственен немногим полупроводникам, т.к. наличие ветви с отрицательной эффективной массой вблизи минимума зоны проводимости является редким явлением.

Дальнейший анализ движения электронов показывает, однако, что локализованные состояния на «колпаке мексиканской шляпы» в канале транзистора могут быть заполнены только благодаря рассеянию. «Обычный» электрон, летящий из левого контакта транзистора (Рис. 8.5 а) в области потенциального барьера движется непрерывно по дисперсионной кривой влево, при этом испытывая три последовательных разворота в экстремумах функции $E(p)$ — их итогом опять же является возвращение его в контакт истока. То же самое можно сказать и об элетронах, изначально находившихся на дисперсионной кривой в положении Рис. 8.5 б. Получается, что в область канала транзистора могут проникнуть только электроны с энергиями $\varepsilon(p) - eV_C > \varepsilon(p_{\min})$ — они при своем движении не дойдут до точки минимума и не испытают разворота.

В итоге мы получаем, что для заполнения локализованных состояний в канале транзистора необходимо достаточно большое время, определяемое вероятностью рассеяния, которая в баллистическом приближении предполагается малой. Акцентируя внимание на высокочастотном поведении транзистора, мы будем считать состояния на «колпаке мексиканской шляпы» пустыми и проводить интегрирование при расчете концентрации носителей по области $p > |p_{\min}|$:

$$n(V_C) = e \int_{p_{\min}}^{\infty} \frac{2\pi g p dp}{(2\pi)^2} f_F(\varepsilon(p)), \quad (8.35)$$

где $g = 4$ — фактор вырождения электронов в графене, а f_F — функция распределения Ферми. При нулевой температуре интегрирование выполняется аналитически и дает:

$$n(V_C) = \frac{e^2}{\pi \hbar^2 v^2} \left((\mu - eV_C)^2 - \varepsilon_{\min}^2 + \sqrt{\gamma_1^2 + \Delta^2} \sqrt{(\mu - eV_C)^2 - \varepsilon_{\min}^2} \right). \quad (8.36)$$

Из вышеприведенного уравнения следует, что концентрация носителей заряда в канале становится отличной от нуля, когда дно зоны проводимости «касается» уровня химического потенциала в одном из контактов (см. зонную диаграмму на Рис. 8.6 В).

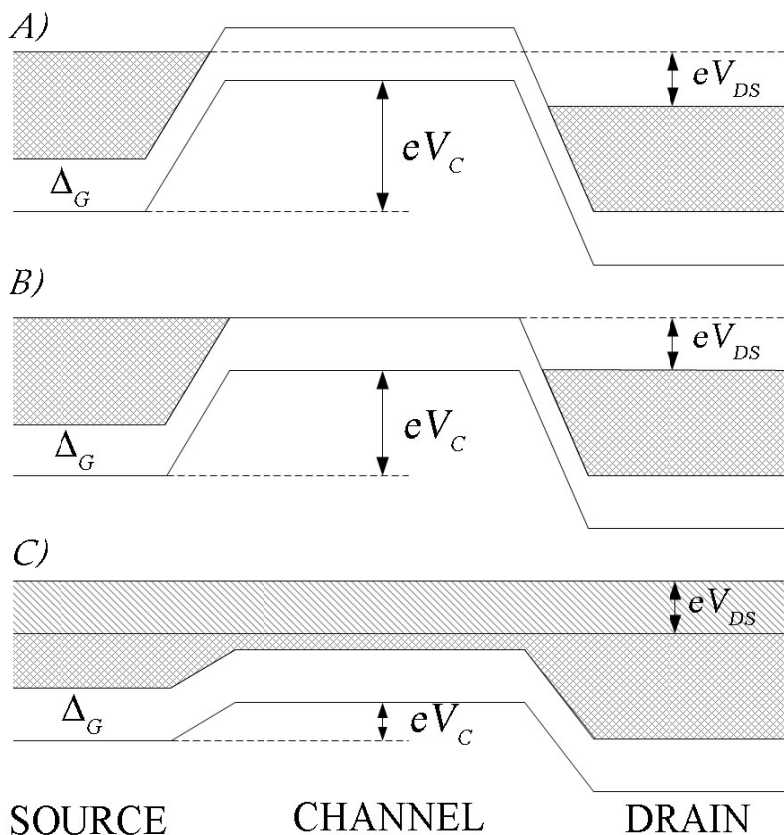


Рис. 8.6. Зонные диаграммы, поясняющие режимы работы транзистора а) выключенное состояние б) напряжение отсечки в) включенное состояние

При $(\mu - eV_C) < \varepsilon_{min}$ в канале отсутствуют носители заряда, и, соответственно, экранирование — тогда решением уравнения (8.35) является очевидное равенство $V_C = V_G$.

Для упрощения дальнейших рассуждений мы сместим уровень отсчета напряжений на величину химического потенциала в контактах — будем считать, что положение уровня Ферми в канале соответствует середине запрещенной зоны при нулевом напряжении на затворе. Необходимо, однако, помнить, что реальные пленки графена на

подложках заряжены благодаря контактной разности потенциалов, и внутреннее электрическое поле приводит к наличию запрещенной зоны (вспомним Рис. 5.6!), поэтому наше предположение является не физическим, а математическим упрощением.

Приступим, наконец, к расчету проводимости — вклад в нее вносят лишь электроны в узком энергетическом интервале между μ_L и μ_R — уровнями химического потенциала в истоке и стоке, разность между этими уровнями равна приложенному напряжению сток-исток: $\mu_L - \mu_R = eV_{DS}$. Полный ток, вызванный бесконечно малым смещением dV_{DS} дается формулой

$$dj = e \langle v_x dn \rangle = e \int_{\substack{v_x > 0 \\ \varepsilon \in [\mu_L; \mu_R]}} g v_x \frac{d^2 p}{(2\pi\hbar)^2}. \quad (8.37)$$

Вводя компоненту групповой скорости

$$v_x = \frac{d\varepsilon}{dp_x} = \frac{d\varepsilon}{dp} \cos \theta, \quad (8.38)$$

для тока легко получить выражение

$$dj = \frac{ge}{(2\pi\hbar)^2} \int_{-\pi/2}^{\pi/2} d\theta \int_{p(\mu_L)}^{p(\mu_R)} \frac{d\varepsilon}{dp} \cos \theta p dp d\theta = \frac{2ge^2 p(eV_C) dV_{DS}}{(2\pi\hbar)^2}. \quad (8.39)$$

Соответствующая проводимость как функция потенциала в канале транзистора дается выражением:

$$G = \frac{e^2}{h} \frac{gk(eV_C)}{\pi}, \quad (8.40)$$

в котором $k(eV_C)$ — функция, обратная к энергетическому спектру $\varepsilon(k)$. Вообще говоря, она является многозначной, но мы считаем состояния на «колпаке мексиканской шляпы» незаполненными и выбираем одну конкретную ветвь. Значения функции $k(eV_C)$ начинаются с величины $k_{\min} = p_{\min}/\hbar$, а не с нуля, что приводит нас к выводу о том, что проводимость транзистора на двухслойном графене как функция напряжения на затворе испытывает скачок при $eV_C = \varepsilon_{\min}$. Величина скачка равна

$$\Delta G = \frac{e^2}{h} \frac{g k_{\min}}{\pi}. \quad (8.41)$$

Для ширины запрещенной зоны $\Delta = 0.2$ эВ волновой вектор $k_{\min} = 4 \text{ нм}^{-1}$, что соответствует скачку проводимости $\Delta G = 1.25 \cdot 10^4 (\text{Ом} \cdot \text{м})^{-1}$. Скачок обратного сопротивления получается умножением величины (8.41) на ширину канала транзистора (ширину графеновой ленты) и может принимать любое значение, в отличие от одномерных систем, где он был равен $e^2/2h$.

Наличие скачка проводимости, вообще говоря, не характерно для двумерных систем. Двухслойный графен является уникальным «полупроводником» — одновременно изотропным и непрямозонным. Именно такой тип спектра приводит к возникновению скачка проводимости. Классические полупроводники — кремний и германий — являются анизотропными и непрямозонными, поэтому в двумерных материалах на их основе скачка проводимости не наблюдалось.

В принципе, формулу (8.40) совместно с (8.34) и (8.36) можно использовать для аналитического расчета проводимости как функции напряжения на затворе при нулевой температуре. Эта задача, однако, сводится к решению алгебраического уравнения четвертой степени, результаты решения не являются наглядными, поэтому мы приступим к описанию методов численного расчета характеристик транзисторов на двухслойном графене.

В случае произвольной температуры в формулы для концентрации и тока начинает входить функция распределения Ферми в явном виде, к тому же мы должны учитывать вклад частично ионизованных дырок. В этом случае вместо (8.35) для концентрации мы должны написать:

$$n(V_C, T) = \frac{2}{\pi \hbar^2 v_F^2} \int_{q_{\min}}^{\infty} (f_F(\varepsilon, V_C) - f_F(\varepsilon, -V_C)) q dq, \quad (8.42)$$

где $q = p v_F$, а второе слагаемое отвечает вкладу дырок. В случае, если напряжение между стоком и истоком V_{DS} уже не бесконечно мало, то концентрация носителей заряда в области канала определяется полусуммой концентраций частиц, летящих из истока ($n(V_C)$) и из стока ($n(V_C - V_{DS})$). Система уравнений электростатики в этом случае модифицируется:

$$\begin{cases} E_1 d_1 + E_2 d_2 = V_G, \\ \kappa_1 E_1 - \kappa_2 E_2 = 4\pi e \cdot \frac{1}{2} (n(V_C) + n(V_C - V_{DS})), \\ E_2 d_2 = V_C, \end{cases} \quad (8.43)$$

Как и во всех классических транзисторных схемах, мы предполагаем потенциал истока равным нулю.

Модифицированное с учетом распределения Ферми выражение для тока принимает вид:

$$j = \frac{2eg}{(2\pi\hbar)^2} \int_{-\infty}^{\infty} (f_F(\varepsilon, V_C) - f_F(\varepsilon, V_C - V_{DS})) p(\varepsilon) d\varepsilon, \quad (8.44)$$

причем устремление нижнего предела интегрирования по энергиям к минус бесконечности отвечает за дырочный вклад в плотность тока. Легко проверить, что при бесконечно малом напряжении сток-исток выражение (8.44) переходит в (8.39). Сделаем замену

$$f_F(\varepsilon, V_C) - f_F(\varepsilon, V_C - V_{DS}) = -\frac{df_F}{d\varepsilon} V_{DS},$$

и поделим на перепад напряжений dV_{DS} :

$$\begin{aligned} G = \frac{dj}{dV_{DS}} &= \frac{2e^2 g p_{\min}}{(2\pi\hbar)^2} \{f_F(\varepsilon_{\min}, V_C) + f_F(-\varepsilon_{\min}, V_C)\} + \\ &+ \frac{2e^2 g}{(2\pi\hbar)^2} \int_{p_{\min}}^{\infty} \{f_F(\varepsilon(p), V_C) + f_F(-\varepsilon(p), V_C)\} dp \end{aligned} \quad (8.45)$$

Полученное выражение состоит из двух чашей — первая строка отчетливо изотропному сдвигу спектра двухслойного графена и приводит к существованию скачка проводимости при нулевой температуре, оно исчезает при $p_{\min}=0$. Второе слагаемое соответствует обычному квадратичному спектру $\varepsilon(p) \sim p^2$.

Задача численного расчета проводимости как функции напряжения на затворе не вызывает вычислительных трудностей, за исключением построения обратной функции $V_C(V_G)$. При ненулевой температуре в выражение (8.34) V_C входит как параметр в несобственном интеграле. При отсутствии программных средств, способных к построению обратных функций подобного рода удобно задать функцию $V_G(V_C)$

таблично, а при нахождении обратной функции просто поменять местами два столбца этой таблицы.

Программная среда Wolfram Mathematica 7.0 позволяет реализовать нахождение обратной функции даже в том случае, если аргумент входит в ее определение как параметр интеграла — с помощью этой среды и было выполнено моделирование транзистора. Алгоритм программы с краткими пояснениями приведен в Приложении А.

Интересно также проследить за эволюцией проводимости и линейной плотности тока как функции напряжения на затворе при изменении температуры — на Рис. 8.7 и Рис. 8.8 приводятся графики для трех температур: гелиевой $T = 4$ К, азотной $T = 77$ К и комнатной $T = 300$ К.

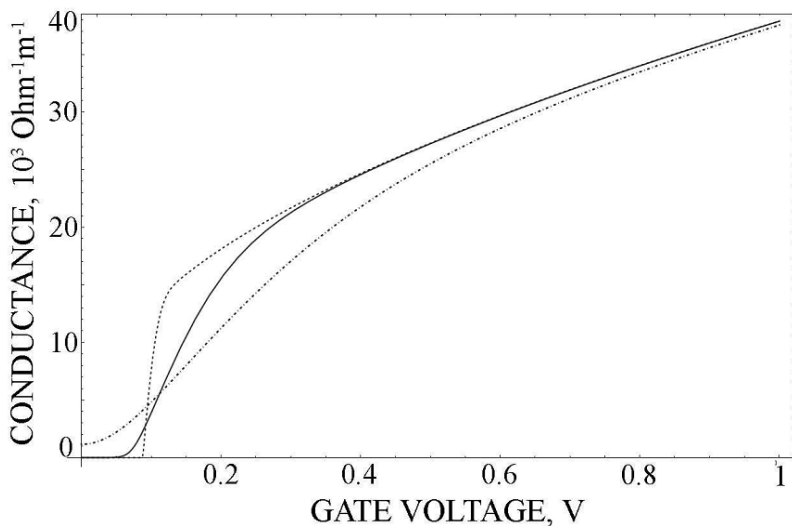


Рис. 8.7. Зависимость проводимости транзистора на 2-слойном графене от напряжения на затворе ($T = 4$ К — точки, $T = 77$ К — сплошная, $T = 300$ К. — штрихпунктир)

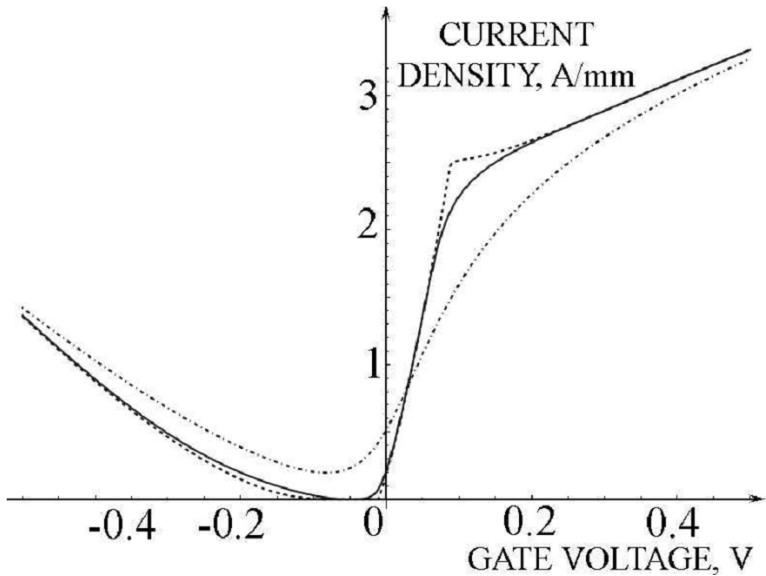


Рис. 8.8. Зависимость плотности тока транзистора на 2-слойном графене от напряжения на затворе ($T = 4$ К — точки, $T = 77$ К — сплошная, $T = 300$ К. — штрихпунктир), напряжение сток-исток равно 0,1 В

Анализируя полученные кривые, мы видим, что отношение токов во включенном и выключенном состояниях для транзистора на двухслойном графене порядка 10 при комнатной температуре и диапазоне изменения напряжения на затворе 0,5 В. Этот результат, конечно, очень далек от достижений кремниевых транзисторов. Именно поэтому методы увеличения ширины запрещенной зоны (*англ. bandgap engineering*) в транзисторах на двухслойном графене привлекают значительное внимание исследователей.

В такой реализации транзистор на основе двухслойного графена не может быть эффективно использован для цифровых схем, однако в качестве аналогового усилителя он имеет очень хорошие перспективы. Определяющим фактором для аналоговых применений является крутизна характеристики $j(V_G)$ (*англ. transconductance*), которая при стремлении температуры к нулю вообще обращается в бесконечность в окрестности напряжения отсечки! Высокая крутизна характеристики рассчитанного нами транзистора при комнатной температуре происходит от скачка проводимости при нулевой температуре.

Моделирование новых электронных приборов на основе графена

Оказывается, что крутизна характеристики может быть оценена аналитически. Если при нулевой температуре скачок проводимости равен ΔG , то производная проводимости в точке скачка при нулевой температуре может быть оценена по формулам

$$\begin{aligned}\frac{dG}{dV_G} &= \frac{\Delta G}{\Delta V_G}, \\ e\Delta V_C &= kT.\end{aligned}\tag{8.46}$$

Вторая формула определяет (оценочно) энергетический интервал, занятый электронами при ненулевой температуре при касании уровня химического потенциала и дна зоны проводимости — это и есть то состояние, в котором крутизна достигает максимального значения. Далее нам остается лишь связать величины ΔV_G и ΔV_C — но у нас уже имеется уравнение связи между ними, это уравнение (8.34)! По сравнению с поиском обратной функции $V_C(V_G)$, сейчас наша задача гораздо легче благодаря наличию соотношения $e\Delta V_C = kT$.

Итак, в уравнении (8.34) мы

Делаем замену $\varepsilon_{\min} + \Delta V_C = kT$;

Раскладываем $\varepsilon(q)$ по степеням $(q - q_{\min})$;

Пренебрегаем вкладом дырок.

После этого выражение для концентрации носителей заряда принимает вид

$$n(V_C, T) = \frac{2}{\pi \hbar^2 v^2} \int_{q_{\min}}^{\infty} \left(1 + \exp \left\{ \frac{(q - q_{\min})^2}{2mv^2 kT} - 1 \right\} \right)^{-1} q dq, \tag{8.47}$$

где m — эффективная масса электрона в окрестности минимума закона дисперсии, определяемая соотношением

$$m = \frac{\gamma_1 (\gamma_1^2 + \Delta^2)^{3/2}}{2v^2 \Delta (2\gamma_1^2 + \Delta^2)}.\tag{8.48}$$

Интеграл (8.47) берется аналитически. После подстановки его значения в формулу для связи потенциалов V_G и V_C мы получаем

$$eV_G = \varepsilon_{\min} + kT + \frac{8e^2 d}{\kappa_1 \hbar^2} \cdot \left\{ \ln(1 + e) m k T + 0.91 q_{\min} \sqrt{\frac{2mkT}{v^2}} \right\}, \quad (8.49)$$

0.91 — это приближенное численное значение интеграла $\int_0^\infty \frac{dt}{1 + e^{t^2 - 1}}$. В кратком виде «тепловое размытие» напряжения на затворе может быть представлено в виде

$$\Delta V_G = aT + b\sqrt{T}, \quad (8.50)$$

где a и b являются функциями как спектра двухслойного графена, так и параметров структуры. Для вышеобозначенных параметров структуры (температура выражена в электрон-вольтах)

$$\begin{aligned} \Delta(eV_G) &= 4.266T + 0.8775\sqrt{T}, \\ \frac{dG}{dV_G} &= \frac{\Delta G(T=0)}{4.266T + 0.8775\sqrt{T}}. \end{aligned} \quad (8.51)$$

Заметим, что наличие слагаемого, пропорционального \sqrt{T} в знаменателе происходит опять же из-за наличия изотропного сдвига по оси импульсов в спектре двухслойного графена. Можно было бы предположить, что для получения крутизны характеристики при ненулевой температуре достаточно было бы поделить значение скачка на kT — такой способ годится, например, при оценке наклона распределения Ферми. В случае с крутизной характеристики транзистора такой простой способ уже не годится.

При комнатной температуре оба слагаемых — aT и $b\sqrt{T}$ — одного порядка величины, при низких температурах второе доминирует.

Вообще говоря, не стоит ожидать точного согласия между расчетом крутизны по формуле (8.3) с точным численным расчетом, т.к. предположение $e\Delta V_G = kT$ является действительно оценочным (можно было с таким же успехом написать $2kT$ или $kT/2$). Поразительно, но оценка оказывается очень точной — на Рис. 8.9 приведено сравнение крутизны, рассчитанной по оценочной формуле (сплошная линия) со значениями, полученными при точном численном расчете.

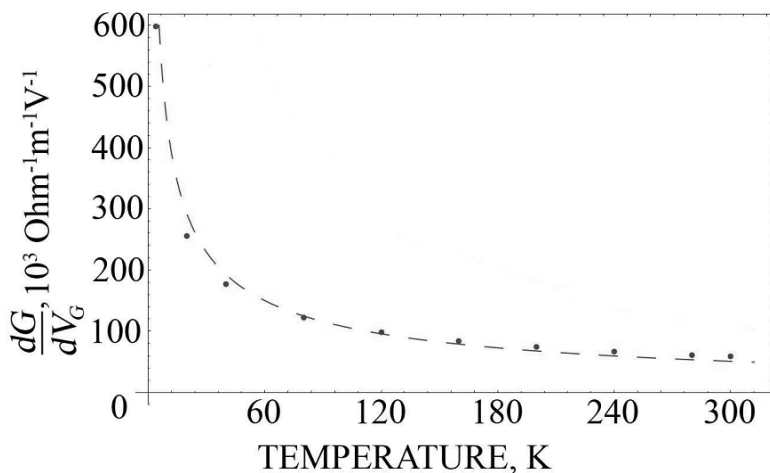


Рис. 8.9. Сравнение оценочной проводимости (штриховая линия)
и точного численного расчета (точки)

Значение крутизны для транзистора на двухслойном графене при обозначенных параметрах структуры и спектра составляет $55 \cdot 10^3 \text{ (Ом} \cdot \text{м} \cdot \text{В)}^{-1}$, что на порядок больше аналогичной величины для транзистора типа «кремний на изоляторе» ($5 \cdot 10^3 \text{ (Ом} \cdot \text{м} \cdot \text{В)}^{-1}$), который рассматривается как один из самых перспективных в кремниевой электронике

Приложение А

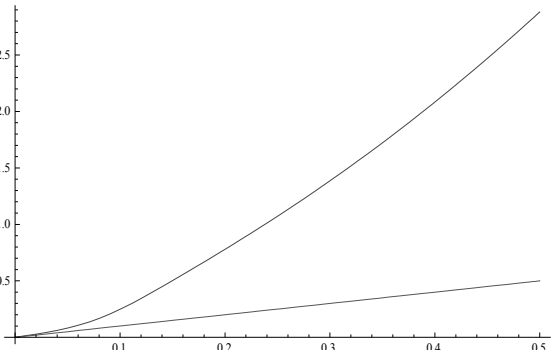
Алгоритм, реализующий моделирование характеристик транзистора на двухслойном графене в среде Wolfram Mathematica 7.0

Левая колонка таблицы содержит код, вводимый в программной среде, а также некоторые выходные данные (графики, числа). Правая колонка — комментарии и пояснения.

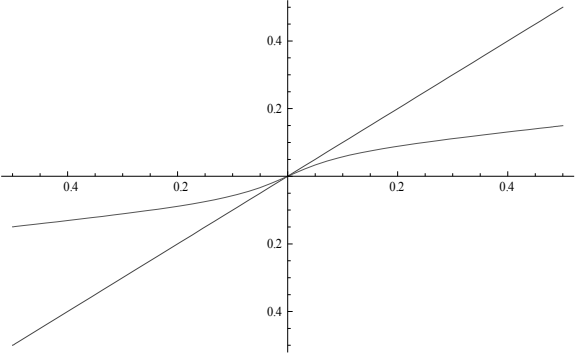
$\varepsilon[q_]:= \sqrt{\frac{\gamma_1^2}{2} + \frac{\Delta^2}{4} + q^2} - \sqrt{\frac{\gamma_1^4}{4} + q^2 (\gamma_1^2 + \Delta^2)}$	<p>Вводим закон дисперсии двухслойного графена (импульс будем измерять в эВ, $q = p v_F$)</p>
$q[x_]:= \sqrt{x^2 + \frac{\Delta^2}{4} + \frac{1}{2} \sqrt{4x^2 \gamma_1^2 + 4x^2 \Delta^2 - \gamma_1^2 \Delta^2}}$	<p>Нам также понадобится аналитически заданная обратная функция $q(\varepsilon)$.</p>
$p_{\min} := \frac{\Delta \sqrt{2\gamma_1^2 + \Delta^2}}{2\sqrt{\gamma_1^2 + \Delta^2}}$	<p>В принципе, среда Mathematica 7.0 может вычислять точку минимума функции и значение в ней, но аналитическое задание ε_{\min} и p_{\min} ускорит расчеты.</p>
$\varepsilon_{\min} := \frac{\gamma_1 \Delta}{2\sqrt{\gamma_1^2 + \Delta^2}}$	

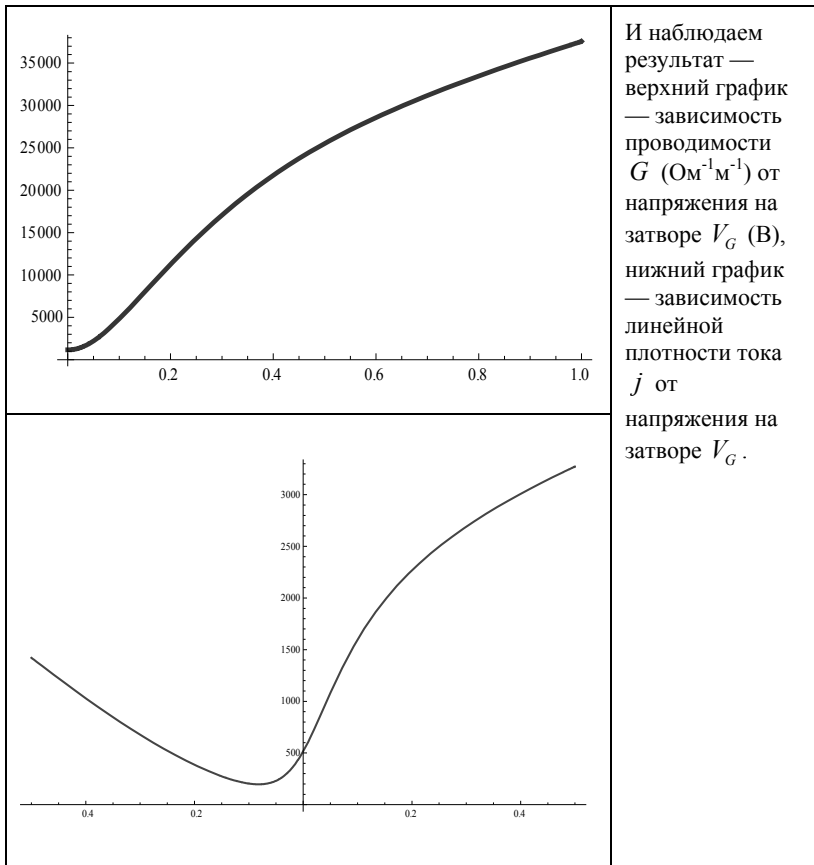
Приложение А

$\gamma_1 := 0.39$ $\Delta := 0.2$ $v := 8 \cdot 10^5$ $\varepsilon_1 := 4$ $h := 1.054 \cdot 10^{-34}$ $d := 1 \cdot 10^{-9}$ $\varepsilon_0 := 8.85 \cdot 10^{-12}$ $\text{charge} := 1.6 \cdot 10^{-19}$	Вводим блок констант, работаем в СИ, все энергии измеряем в эВ
$n[\mu, T] := \frac{2 \text{charge}^2}{\pi (h)^2} \text{Integrate} \left[p \left(\left(1 + \exp \left[\frac{\varepsilon[p] - \mu}{T} \right] \right)^{-1} - \left(1 + \exp \left[\frac{\varepsilon[p] + \mu}{T} \right] \right)^{-1} \right), \{p, p_{\min}, \infty\} \right]$	Вводим функцию расчета концентрации носителей заряда в канале транзистора
$\text{Manipulate}[\text{Plot}[n[\mu, T], \{\mu, 0, 0.2\}], \{T, 0.001, 0.025\}]$	Строим ее график с возможностью варьирования температуры

$t := 0.025 * \left(\frac{300}{300} \right)$	<p>Далее фиксируем температуру в эВ (в данном случае — комнатную)</p>
$Gate[x_] := x + \frac{charge * d}{\epsilon_0 \epsilon_1} n[x, t]$	<p>Задаем функцию напряжения на затворе от потенциала в канале</p>
$Plot[\{Gate[\mu], \mu\}, \{\mu, 0, 0.5\}]$	<p>И строим ее вместе с графиком $V_G = V_C$</p>
	
$inv[f_ , s_] := Function[\{t\}, s /. FindRoot, f - t, \{s, 1\}]$	<p>Задаем операцию построения обратной функции к некоторой функции как решение уравнения $f(x) = 0$</p>
$Vol = inv[Gate[x], x]$	<p>Задаем и строим график обратной функции</p>
$Plot[\{x, Vol[x]\}, \{x, -0.5, 0.5\}]$	

Приложение А

	$V_C(V_G)$ $(Vol \equiv V_C)$
$Cond[\mu_-] := \frac{2charge^3 \cdot 4}{(2\pi h)^2 v} \times$ $\times \left(p_{\min} \left(\left(1 + \text{Exp} \left[\frac{\varepsilon_{\min} - Vol[\mu]}{t} \right] \right)^{-1} + \left(1 + \text{Exp} \left[\frac{\varepsilon_{\min} + Vol[\mu]}{t} \right] \right)^{-1} \right) \right) +$ $+ NIntegrate \left[\left(1 + \text{Exp} \left[\frac{\varepsilon[x] - Vol[\mu]}{t} \right] \right)^{-1} + \left(1 + \text{Exp} \left[\frac{\varepsilon[x] + Vol[\mu]}{t} \right] \right)^{-1} \right], \{x, p_{\min}, \infty\}$	<p>Задаем формулу для расчета проводимости (проинтегрировать, как в (8.44), в пределах $(-\infty, +\infty)$ затруднительно, т.к. $p(\varepsilon)$ в этих пределах — разрывная функция)</p>
$J[\mu_-] := \frac{2charge^3 \cdot 4}{(2\pi h)^2 v} NIntegrate \left[\left(\left(1 + \text{Exp} \left[\frac{x - (Vol[\mu] + 0.1)}{t} \right] \right)^{-1} - \left(1 + \text{Exp} \left[\frac{x - (Vol[\mu])}{t} \right] \right)^{-1} \right) + \right. \\ \left. + \left(1 + \text{Exp} \left[\frac{x + (Vol[\mu] + 0.1)}{t} \right] \right)^{-1} - \left(1 + \text{Exp} \left[\frac{x + (Vol[\mu])}{t} \right] \right)^{-1} \right], q[x], \{x, \varepsilon_{\min}, \infty\}$	<p>Аналогично задаем формулу для расчета плотности тока (напряжение на стоке — 0.1 В)</p>



Приложение Б

Работа с квантово-химическим пакетом Gaussian

В этой главе будут описаны основы работы с одним из наиболее универсальных на сегодняшний день пакетом Gaussian. Несмотря на то, что для некоторых задач существуют более эффективные программные коды, описание их всех вышло бы за рамки данного пособия.

Из других пакетов стоит отметить программу Priroda, позволяющую выполнять DFT-вычисления с GGA функционалами. Priroda также имеет базу элементов и возможность релятивистских вычислений. Отличительной ее особенностью является крайне высокая скорость вычислений. Также возможна ее интеграция с Gaussian.


Gaussian может выполнять множество различных вычислений и рассчитывать большое количество параметров, в частности:

- Молекулярные энергии и орбитали основного и переходных состояний.
- Геометрическая структура молекул
- ИК и комбинационные спектры
- Атомные заряды, дипольные моменты, электронные плотности и потенциалы.
- Поляризуемость.
- Потенциалы ионизации и сродство к электрону.
- Термодинамические свойства.
- Пути реакций.

Пакет может производить вычисления как для молекулярных систем в газовой фазе и растворе, так и для периодических кристаллических структур.

Для того, чтобы выполнить какой-либо расчет в программе необходимо предварительно подготовить файл с описанием задачи. Обычно для таких файлов используется расширение .gjf. Структура файла задания будет описана ниже.

Чтобы запустить файл задания на выполнение, необходимо сначала открыть его. Это можно сделать следующими способами: выбрав

пункт меню File->Open, нажав кнопку  на панели инструментов или перетащив файл задания в окно Gaussian. После выполнения какого-либо из этих действий появится окно редактирования задания, изображенное на рис. Б.2. Для немедленного запуска нужно нажать кнопку



Ниже перечислены функции остальных кнопок окна редактирования задания:



— принять изменения и немедленно запустить задание на выполнение.



— принять изменения и выйти из редактора.



— сохранить изменения в файл.



— закрыть окно редактора без принятия изменений.



— выполнить утилиту check-route.



— выбрать начальный этап.

Эти функции также могут быть вызваны из меню. Для редактирования набора дополнительных этапов вычисления используется кнопка "AdditionalSteps". После ее нажатия появляется окно, похожее на редактор задания. Изменять количество и порядок следования шагов можно с помощью функций Step->AddStep, Step->DeleteStep и Step->Reorder.

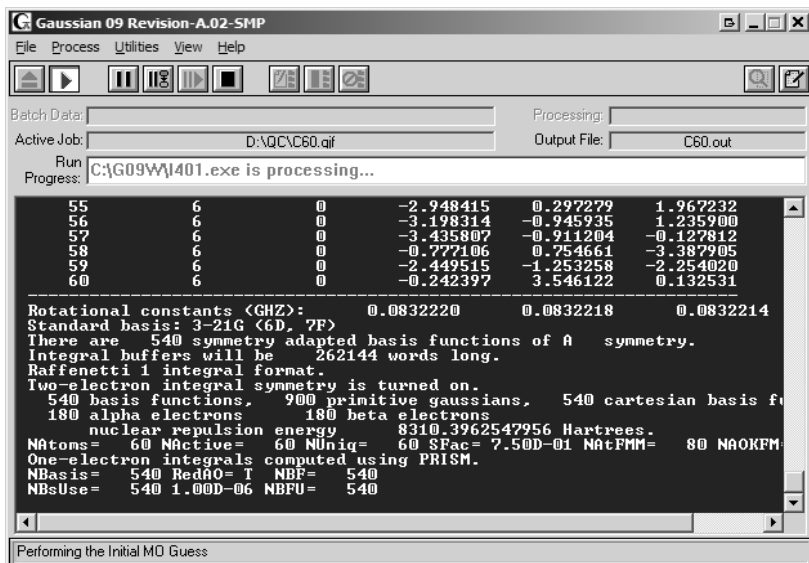


Рис. Б.1. Внешний вид главного окна Gaussian.

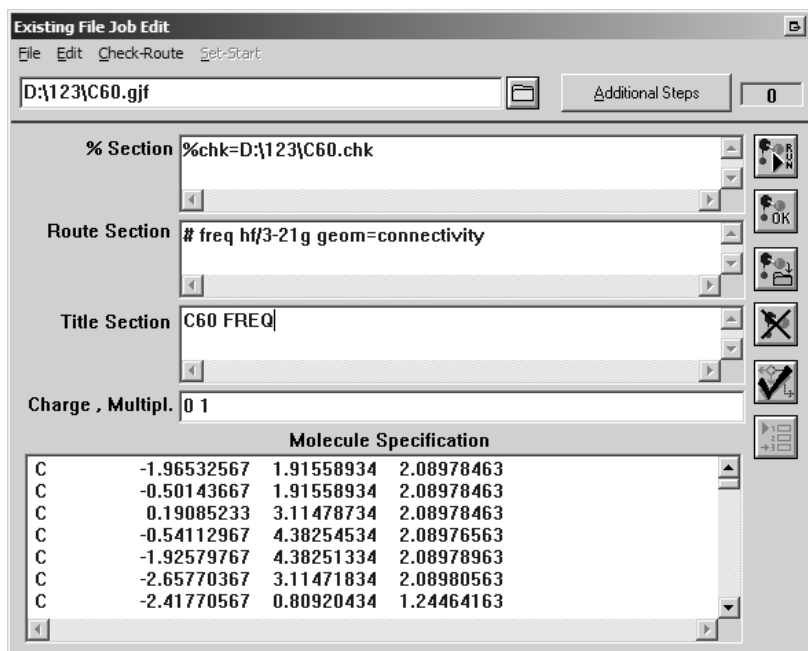






Рис. Б.2. Внешний вид окна редактирования файла задания.

Для изменения редактируемого в данный момент этапа можно пользоваться кнопками  и .

Для вызова check-route можно воспользоваться соответствующим пунктом меню или кнопкой . Для того, чтобы принять изменения и закрыть окно надо нажать кнопку .

Ниже перечислены кнопки главного окна:



— запустить на выполнение загруженный файл задания.



— немедленно приостановить выполнение задания.



— приостановить выполнение после завершения работы текущего звена.



— возобновить выполнение задания.



— прекратить выполнение задания.



— прекратить выполнение пакета заданий после завершения текущего задания.



— немедленно прекратить выполнение пакета заданий.



— запустить текстовый редактор.



— открыть файл результатов в текстовом редакторе.

Процесс расчета в Gaussian состоит из последовательного запуска программ, называемых линками (звеньями). Каждый линк выполняет некоторую задачу (например, выбор начального приближения, подсчет интегралов, решение уравнений и т.п.). Полный перечень линков можно найти в документации. Выполнение всегда начинается с линков L0 и L1. Первый производит инициализацию программы, а второй, обрабатывая файл задания, определяет, какие линки должны быть запущены. Для определения того, какие линки будут запущены, можно использовать утилиту `check-route`, упоминавшуюся выше.

Gaussian может быть настроен таким образом, чтобы в ходе вычислений создавался так называемый файл чекпоинта (контрольной точки). Тогда при каком-либо сбое можно будет перезапустить расчет с места, описываемого этим файлом. Также файл чекпоинта может быть использован в качестве файла результатов. В наборе утилит, поставляемых с Gaussian, есть программа `formchk`, позволяющая преобразовать файл чекпоинта в текстовый формат, который может быть прочитан человеком.

Ниже перечислены все утилиты, поставляемые вместе с Gaussian. Некоторые из этих утилит доступны из меню `Utilities` главного окна Gaussian:

- `c8609` — преобразовывает файлы чекпоинтов из формата предыдущих версий в формат, используемый версией 09.
- `chkchk` — показывает описание и route из файла чекпоинта.
- `cubegen` — утилита для генерации массивов данных (кубов).
- `subman` — утилита для работы с массивами данных.
- `formchk` — конвертирует чекпоинты из бинарного формата в текстовый.
- `freqmem` — определяет объем памяти, необходимый для вычисления частот.

- gauopt — производит оптимизации по переменным, отличным от координат атомов.
- ghelph — справка.
- mm — программа для молекулярно-механических расчетов.
- newzmat — конвертирует между различными форматами описания молекулярной геометрии.
- testrt — проверка синтаксиса route.
- unfchk — конвертирует чекпоинты из текстового формата в бинарный. Это можно использовать, например, для продолжения вычислений с использованием вычислительной системы, имеющей другую архитектуру.

Файл задания состоит из набора секций. Первая секция называется Link 0. Эта секция состоит из команд, начинающихся на %. Их синтаксис обычно следующий: %<имя>=<параметры>. Здесь указываются настройки, управляющие основными аспектами работы программы, например:

- %chk указывает имя используемого чекпоинт-файла.
- %mem задает предел используемой памяти.
- %nprocshared определяет число используемых процессоров в SMPсистемах.

Следующая секция называется route. Она определяет, что будет вычисляться и каким образом и всегда начинается с символа # и заканчивается пустой строкой. В простых случаях данная секция может быть записана в виде:

<тип расчета><метод>/<базис><опции>,

например,

OptHF/6-31G.

Gaussian поддерживает следующие типы расчетов:

- SP — потенциальная энергия молекулы (используется по умолчанию если метод не указан).
- Opt — оптимизация геометрии.
- Freq — анализ колебаний и термохимический анализ (используется для вычисления ИК и комбинационных спектров).
- IRC — анализ пути реакции.
- IRCMax — вычисление высоты потенциального барьера для реакции.
- Scan — вычисление формы поверхности потенциальной энергии.
- Polar — вычисление поляризуемостей и гиперполяризуемостей.
- ADMPи BOMD — молекулярно-динамические расчеты.

- Force — вычисление силы, действующей на ядра.
- Stable — проверка стабильности волновой функции.
- Volume — вычисление объема, занимаемого молекулой.
- NMR — вычисление ЯМР экранирования и химических сдвигов.

После route задается молекула. Первая строка определения содержит два числа — полный заряд (например, 0 для нейтральной молекулы, 1 для однозарядного катиона, -2 для двухзарядного аниона). Второе число определяет мультиплетность молекулы.

Самый простой вариант определения геометрии подразумевает указание координат всех атомов в молекуле. Указание координат атома выглядит следующим образом:

<Обозначение атома>xyz

Например, молекула этилового спирта может быть задана таким образом:

0 1			
C	-3.57765250	-0.18560175	0.54720051
H	-3.22097966	0.31879644	1.42085202
H	-3.22097966	0.31879644	-0.32645099
H	-4.64765250	-0.18558857	0.54720051
C	-3.06433678	-1.63753391	0.54720051
H	-3.42100924	-2.14193196	1.42085225
H	-3.42101007	-2.14193279	-0.32645041
O	-1.63433678	-1.63755081	0.54719936
H	-1.31390120	-2.54249166	0.54895494

Для задания периодических структур необходимо определить элементарную ячейку, а затем определить вектора трансляции. Вектор трансляции задается следующим образом:

TVdxdydz

Кроме указания координат атомов возможно задание геометрии с помощью так называемой Z-матрицы. В этом случае задаются расстояния и углы между самими атомами.

Помимо вышеперечисленных параметров, Gaussian имеет множество других. Их подробное описание можно найти в руководстве. Описание их всех выходит за рамки данного пособия.

После выполнения задания гауссиан создает лог-файл, в котором описан весь процесс решения и результаты. Также результаты могут быть получены из чекпоинт-файла.

Составление файлов заданий и интерпретация результатов вручную с помощью текстовых файлов зачастую очень неудобна. К тому же некоторые часто используемые виды постобработки требуют дополнительных вычислений. Самым простым примером таких вычислений может быть вычисление изоповерхностей электронной плотности молекулярных орбиталей для электрона, с известным разложением по базису. Для упрощения этой задачи существует множество программ-визуализаторов. Вот некоторые из них: GaussView, ChemCraft, ViewMol3D, Molden.

Визуализаторы позволяют отображать геометрию молекулы, моды колебаний, спектры, электронные плотности, заряды и т.п. Некоторые из них также позволяют конструировать файлы заданий для различных расчетных программ. Визуализатор GaussView разрабатывается той же компанией, которая разрабатывает Gaussian и работает исключительно с их пакетом. Кроме очевидных недостатков этот факт имеет также одно преимущество: GaussView позволяет управлять многими аспектами работы Gaussian.

Основное окно GaussView показано на Рис. Б.3. Кроме основного окна обычно открыто одно или несколько окон просмотра молекулярной структуры. Мы не будем здесь описывать все возможности GaussView. В место этого, для примера, мы приведем краткое описание процедуры вычисления ИК-спектра молекулы этилового спирта.

В начале следует создать новую молекулярную группу. Для этого необходимо выбрать в меню File пункт CreateMoleculeGroup. После этого откроется пустое окно, в котором необходимо определить структуру молекулы. Сначала необходимо разместить две молекулы углерода. Чтобы это сделать, необходимо нажать в главном окне кнопку



и затем на кнопку выбора типа атома справа от надписи builder-fragment. Появится окно с таблицей Менделеева, в которой надо выбрать углерод, а внизу указать, что он четырехвалентный. После этого, изображение четырехвалентного углерода появится в главном окне программы.

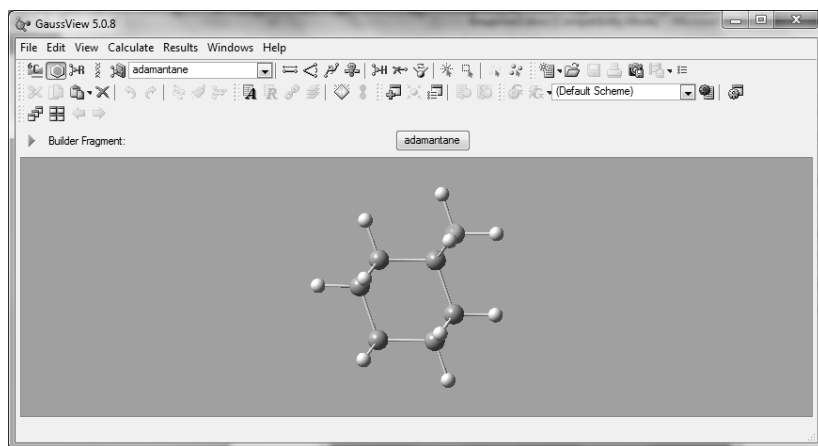



Рис. Б.3. Основное окно GaussView.

Теперь его необходимо разместить в молекуле. Для этого необходимо кликнуть мышью в произвольный участок окна молекулярной группы. Для того, чтобы разместить второй углерод, связанный с первым, необходимо кликнуть мышью на атом водорода, который заменится на еще один углерод. Теперь таким же образом необходимо разместить атом кислорода. После этих действий окно молекулярной структуры должно принять вид, подобный изображенному на Рис. Б.4.

Далее необходимо задать параметры расчета. Для этого надо на-

жать на кнопку . После этого появится окно GaussianCalculation-Setup, изображенное на Рис. Б.5. На вкладке jobtype следует выбрать режим Opt+Freq, на вкладке Method –требуемый метод расчета и базис. После этого надо нажать на кнопку Submit, сохранить файл, и согласиться с передачей этого файла в Gaussian, после чего будет запущен Gaussian и начнется расчет.

После того, как расчет будет закончен, GaussView предложит открыть файл результата. Это может быть как лог-файл, так и чекпоинт. После открытия появится новое окно с геометрией оптимизированной молекулы. Для обработки результатов можно использовать пункт меню Results.

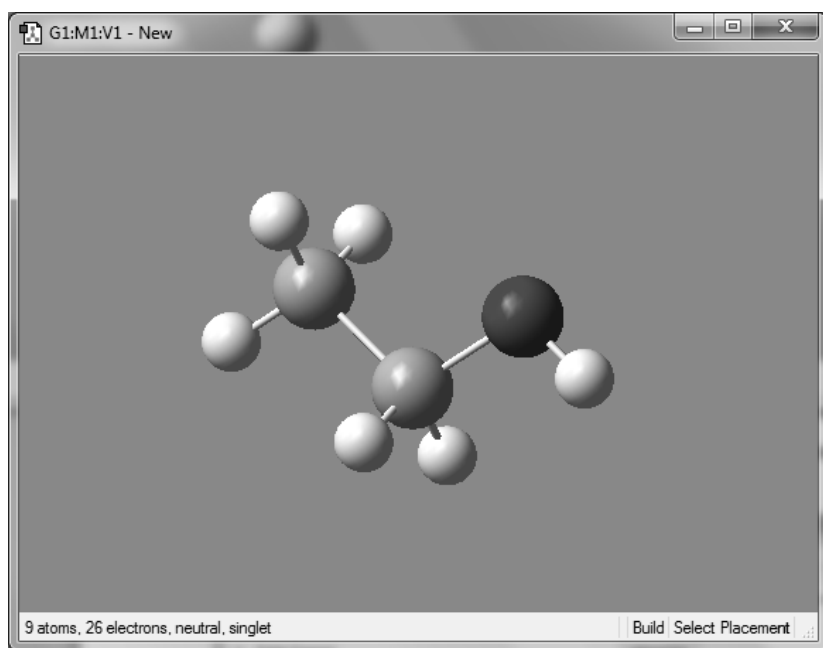


Рис. Б.4. Окно молекулярной структуры с молекулой этилового спирта.

Для того чтобы построить изоповерхности плотности молекулярных орбиталей следует использовать пункт Results->Surfaces/Contours...

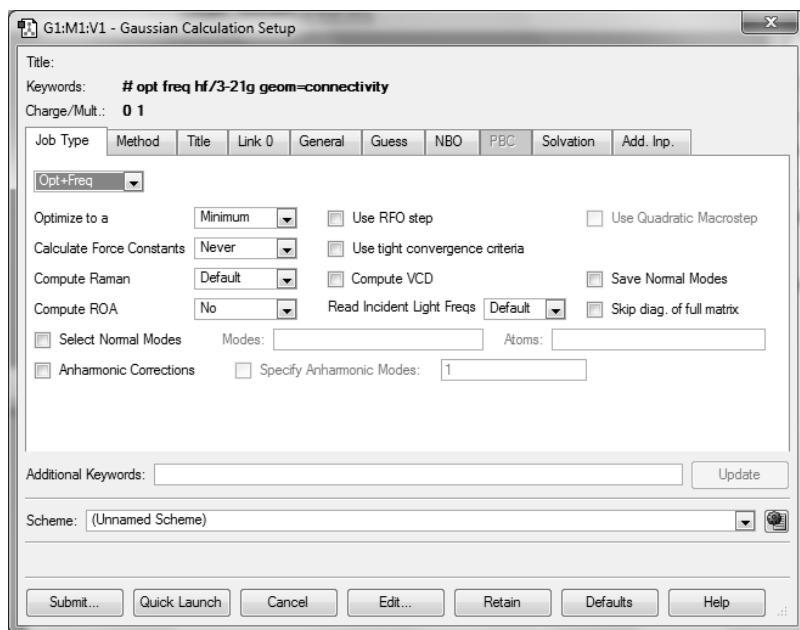


Рис. Б.5. Окно GaussianCalculationSetup с выбранным режимом оптимизации и вычисления спектра.

Для отображения колебательных мод и спектра молекулы следует использовать пункт Results->Vibrations. После выбора этого пункта появится окно со списком всех найденных колебательных мод. Для отображения моды в окне молекулы следует выбрать требуемую моду и нажать кнопку StartAnimation.

Для того чтобы отобразить спектр, необходимо нажать кнопку Spectrum. Полученный результат должен выглядеть примерно так, как на Рис. Б.6.

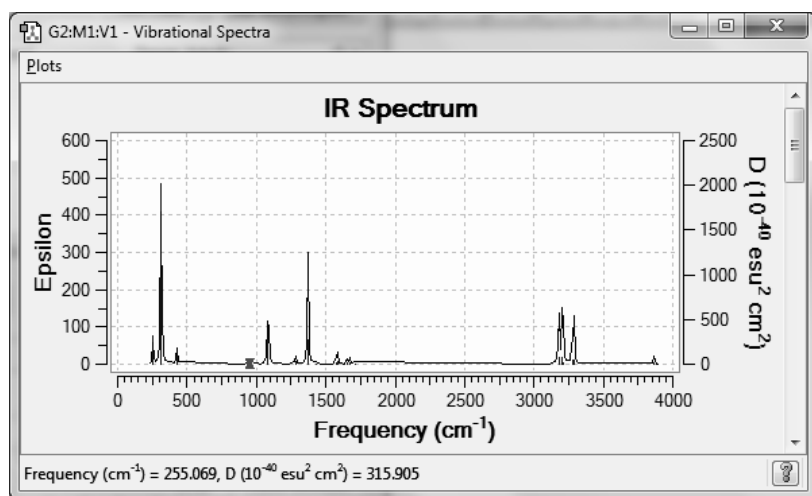


Рис. Б.6. ИК-спектр поглощения молекулы этилового спирта, вычисленный в Gaussian.

Кроме спектров GaussView также может отображать пути реакций, поверхности потенциальной энергии и многое другое.

Литература

1. Зи С. Физика полупроводниковых приборов: М., Мир, 1984;
2. Абрикосов А.А. Основы теории металлов: М., Физматлит, 2009;
3. Yu-Ming Lin. 'Recent progress on high frequency graphene devices'. ISGD 2010 — book of abstracts;
4. Bolotin K. I. *et al.* "Ultrahigh electron mobility in suspended graphene" Solid State Comm. 146, 351 (2008);
5. Taisuke Ohta, Aaron Bostwick, Thomas Seyller, Karsten Horn, Eli Rotenberg « Controlling the Electronic Structure of Bilayer Graphene», Science 313, 951 (2006);
6. E. McCann, D.S.L. Abergel, V.I. Falco, Eur. Phys. J. Special Topics 148 (2007) 91;
7. L. Falkovsky JETP, 110(2), (2010);
8. Kenji Hata, Don N. Futaba, Kohei Mizuno, Tatsunori Naimai, Motoo Yumura and Sumio Iijima « Water-Assisted Highly Efficient Synthesis of Impurity-Free Single-Walled Carbon Nanotubes» Science 19 November 2004: Vol. 306 no. 5700 pp. 1362-1364
9. «Nanodiode based on a multiwall CNx/carbon nanotube intramolecular junction» Y Chai, X L Zhou, P J Li, W J Zhang, Q F Zhang and J L Wu Nanotechnology 16 (2005) 2134-213
10. «Transport through crossed nanotubes» M.S. Fuhrer, Andrew K.L. Lim, L. Shih, U. Varadarajan, A. Zettl, Paul L. McEue, Physica E 6 (2000) 868
11. Ernst T., Munteanu D. *et al.* // Proc. IEEE 1999. Int. SOI Conf., Rohnert Park (California, USA), oct. 1999.
12. Popov V. P., Antonova I. V., Stas V. F. *et al.* // J. Mater. Sci. Eng. 2000. Vol. B73. P. 82-86.
13. Uchida K., Takagi S. // Appl. Phys. Lett. 2003. Vol. 82. P. 2916.
14. Ananiev S.D., V'yurkov S.D., Lukichev V.F. // Proc. SPIE. 2006. Vol. 6260 ;
15. Сидоров А.А., Вьюрков В.В., Орликовский А.А. // Труды ФТИ-АН. 2005. Т. 18. С. 327.
16. . Lolivier J. *et al.* // Intern.SOI Conf.Proc. Hawaii. IEEE. 2005, p.2.6.
17. Shifren L., Ringhofer C., Ferry D.K. // IEEE Trans. El. Devices. 2003. Vol. 50. P. 769.
18. Bordone P., Pascoli M., Brunetti R., Bertoni A., Jacoboni C. // Phys. Rev. B. 1998. Vol. 59. P. 3060.

19. *Mamaluy D., Sabathil M., Vogl P.* // J. Appl. Phys. 2003. V. 93, P. 4628-4633.
20. *Lake R., Klimeck G., Bowen R.C., Javanovic D.* // J. Appl. Phys. 1997. Vol. 81. P. 7845.
21. *Hake A., Khondker A.N.* // J. Appl. Phys. 2000. Vol. 87. P. 2553.
22. *Venugopal R., Paulsson M., Goasguen S., Datta S., Lundstrom M.S.* // J. Appl. Phys. 2003. Vol. 93. P. 5613.
23. *Jimenez D., Saenz J.J., Iniguez B. et al.* // J. Appl. Phys. 2003. Vol. 94. P. 1061.
24. *Landauer R.* Transport as a Consequence of the Incident Carrier Flux // in Localization, Interaction, and Transport Phenomena, G.Bergmann and Y. Buynseraede, Eds., Springer-Verlag. Heidelberg. 1985, P. 38.
25. *Buttiker M.* // IBM J. Res. Dev. 1988. Vol. 32. P. 317.
26. *Venugopal R., Goasguen S., Datta S., Lundstrom M.S.* // J. Appl. Phys. 2004. Vol. 95. P. 292.
27. *Gilbert M.J., Ferry D.K.* // J. Vac. Sci. Technol. 2004. Vol.22, P.2039.
28. *Gilbert M. J. and Ferry D. K.* // J. Appl. Phys. 2004. Vol. 95, P. 7954.
29. *Wu H., Sprung D. W. L., Martorell J., and Klarsfeld S.* // Phys. Rev. B. 1991. Vol. 44. P. 6351.
30. *Cattapan G. and Lotti P.* // Eur. Phys. J. B. 2007. Vol. 60. P. 51
31. *Orlikovsky A., Vyurkov V., Lukichev V., Semenikhin I., Khomyakov A.* All quantum simulation of ultrathin SOI MOSFET in Nanoscaled Semiconductor-on-Insulator Structures and Devices., Springer. 2007. P. 323-340.
32. *Vyurkov V., Semenikhin I., Lukichev V., Burenkov A., and Orlikovsky A.* // Proc. SPIE. 2008. Vol. 7025, P. 70251K.
33. О.В. Наумова, И.В. Антонова, В.П. Попов, Ю.В. Настаушев, Т.А. Гаврилова, Л.В. Литвин, А.Л. Асеев «Нанотранзисторы кремний-на-изоляторе: перспективы и проблемы реализации», ФТП, 2003, том 37, вып.10
34. B.R.Perkins, D.P.Wang, D.Soltman, A.J.Yin, J.M.Xu, and A.Zaslavsky. Differential current amplification in three-terminal Y-junction carbon nanotube devices // Appl.Phys.Lett. 2005, v. 87, p. 123504.
35. S.Bandyopadhyay and M.Cahay. Reexamination of some spintronic field-effect device concepts // Appl. Phys. Lett. 2004, 85, 1433-1435.
36. Ю. А. Бычков, Э. И. Рашба // Письма в ЖЭТФ 39 66 (1984).
37. H. W. Ch. Postma, M. de Jonge, Z. Yao, C. Dekker, Phys. Rev. B 62, R10653 (2000).

Математическое и компьютерное моделирование наносистем

Подписано в печать 12.07.2011.

Формат 60х90 1/16

Гарнитура Times.

Бумага офсетная. Тираж 500 экз.

Заказ № 6051.

Отпечатанов ОАО «Можайский полиграфический комбинат»,
143200, г. Можайск, Московская обл., ул. Мира, д.93,
Тел.: (49638) 2-06-85, (495) 745-84-28
www.oaompk.ru; www.oaompk.pф