К. К. Абгарян, Д. Л. Ревизников

ЧИСЛЕННЫЕ МЕТОДЫ
в моделировании электронных свойств наноразмерных гетероструктур

Учебное пособие

МОСКВА
МАКС Пресс
2017

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ
(национальный исследовательский университет)» (МАИ)
ФЕДЕРАЛЬНЫЙ ЦЕНТР «ИНФОРМАТИКА И УПРАВЛЕНИЕ» РАН

К. К. Абгарян, Д. Л. Ревизников

ЧИСЛЕННЫЕ МЕТОДЫ
в моделировании электронных свойств наноразмерных гетероструктур

Учебное пособие

МАКС ПРЕСС

МОСКВА – 2017

УДК 519.7 (075.8)
ББК 22.18 (я73)
А13

Работа выполняется при финансовой поддержке РФФИ по гранту 16-08-01178 А

*Печатается по рекомендации Ученого Совета
факультета Информационных технологий и прикладной математики
Московского авиационного института (национального исследовательского университета)*

Рецензенты:
В. А. Харченко, доктор технических наук, ведущий научный сотрудник
Федерального исследовательского центра «Информатика и управление» РАН;
И. Э. Иванов, кандидат физико-математических наук,
доцент физического факультета МГУ имени М. В. Ломоносова

Абгарян К. К., Ревизников Д. Л.
А13 Численные методы в моделировании электронных свойств наноразмерных гетероструктур: учебное пособие / К. К. Абгарян, Д. Л. Ревизников. – Москва: МАКС Пресс, 2017. – 108 с.
ISBN 978-5-317-05705-3

В учебном пособии представлены базовые математические модели, численные методы и алгоритмы, используемые при проектировании и оптимизации современных полупроводниковых наноструктур.

Рассмотрены методы численного расчета распределения электронной плотности в наноструктурах на основе математической модели, включающей систему взаимосвязанных уравнений Шредингера и Пуассона, методы моделирования подвижности электронов с учетом различных механизмов рассеяния, задачи и методы оптимизации многослойных наноразмерных гетероструктур. Основное внимание уделено построению эффективных вычислительных алгоритмов.

Представленные методы и средства математического моделирования открывают перспективы решения задач, актуальных для развития современной наноэлектроники.

Учебное пособие предназначено для студентов старших курсов, аспирантов технических университетов, научных работников, специализирующихся в области прикладной математики, вычислительной электроники и наноструктурного материаловедения.

Ключевые слова: численные методы и алгоритмы, математическое моделирование, оптимизация, плотность электронов, подвижность электронов, вычислительные алгоритмы.

УДК 519.7 (075.8)
ББК 22.18 (я73)

ISBN 978-5-317-05705-3

© Абгарян К. К., Ревизников Д. Л., 2017
© Оформление. ООО «МАКС Пресс», 2017

ОГЛАВЛЕНИЕ

Предисловие 5

Введение 8

Глава 1. Моделирование распределения носителей заряда в наноразмерных полупроводниковых гетероструктурах 12

1.1. Математическая модель. Система уравнений Шрёдингера–Пуассона 12

1.2. Глобальный итерационный процесс для получения самосогласованного решения 17

1.3. Методы решения уравнения Шрёдингера 22

1.3.1. Конечно-разностная аппроксимация 23

1.3.2. Неявный QR-алгоритм со сдвигом 28

1.3.3. Метод бисекции 31

1.3.4. Метод Ланцоша 33

1.3.5. Метод обратных итераций 35

1.4. Метод сопряженных градиентов для решения уравнения Пуассона 36

1.5. Приближенная модель для экспресс-расчета концентрации носителей заряда в гетероструктуре AlGaN/GaN 39

1.6. Примеры решения задач 40

Глава 2. Моделирование подвижности электронов в полупроводниковых гетероструктурах 49

2.1. Модели динамики электронов в полупроводниках 50

2.2. Расчет подвижности электронов в гетроструктурах 56

2.3. Методы решения уравнения переноса (дрейфа–диффузии) носителей заряда 64

Глава 3. Задачи оптимизации наноразмерных полупроводниковых гетероструктур 80
3.1. Постановка задач оптимизации 81
3.2. Вычисление градиента функционала. Сопряженная задача 84
3.3. Методы решения задач оптимизации 88
3.4. Бессеточный алгоритм решения прямых и обратных задач математической физики 93
3.5. Примеры решения задач 96

Библиографический список 102

ПРЕДИСЛОВИЕ

В последнее десятилетие двадцатого века среди физико-химических процессов и систем произошло выделение в отдельный класс таких объектов и связанных с ними явлений, как нанокластеры и наноструктуры. Это обусловлено, главным образом, значительным прогрессом в получении и исследовании нанообъектов, новых материалов, создании нанотехнологий и наноустройств. Нанометровый диапазон 1–100 нм открывает ряд новых физических и химических свойств веществ, а также новые подходы к изучению вещества. Одной из наиболее активных отраслей, в которых нашли применение наноразмерные элементы, стала электроника. Отличительной особенностью электронных наноструктур является доминирующая роль квантово-механических явлений, определяющих их электронные, оптические, магнитные и др. свойства. Это факт необходимо учитывать при математическом моделировании таких структур.

В настоящем учебном пособии даются базовые математические модели и численные методы, используемые при проектировании и оптимизации современных полупроводниковых наноструктур.

Учебное пособие состоит из введения и 3 глав.

Во введении кратко рассмотрены некоторые физические аспекты, характерные для данного направления.

В главе 1 основное внимание уделено вопросам численного расчета распределения электронной плотности в наноструктурах. Базовая математическая модель представляет собой систему взаимосвязанных уравнений Шрёдингера и Пуассона. На данном уровне локальные вычислительные процедуры, служащие для решения уравнений Шрёдингера и Пуассона, объеди-

нены глобальным итерационным процессом, направленным на согласование решений. Принципиальным моментом является сильная взаимосвязь уравнений, характерная для рассматриваемого класса задач, что приводит к проблеме сходимости глобальных итераций. В особой мере эта проблема проявляется для вюрцитных структур, когда на гетероинтерфейсах появляется мощный локализованный источник заряда. С целью ускорения вычислительного процесса при расчете электронной плотности в гетероструктуре рекомендован подход, основанный на аппроксимации нелинейной зависимости электронной плотности от потенциала в сочетании с линеаризацией уравнения Пуассона. Показана эффективность данного подхода в задачах рассматриваемого класса. Как ясно из изложения материала, вычислительные алгоритмы носят комплексный характер. Важными элементами вычислительных алгоритмов являются методы решения базовых задач математической физики – спектральной задачи Штурма–Лиувилля (уравнение Шрёдингера) и краевой задачи для уравнения Пуассона. Поскольку решение базовых задач осуществляется многократно в ходе вычислительного процесса, важным требованием к алгоритмам их решения является экономичность вычислений. В этой связи дано описание эффективных численных методов, отражающих специфику проблемной области.

В главе 2 рассмотрены вопросы моделирования динамики электронов в полупроводниковых структурах. Описаны модели различной степени сложности: представляемые кинетическим уравнениям Больцмана и более простые «гидродинамические» модели. Основное внимание уделено расчету подвижности носителей заряда с учетом различных механизмов рассеяния. Рассмотрены разностные схемы решения уравнений переноса носителей заряда. Проанализированы типичные проблемы, связанные с диссипативными и дисперсионными схемными эффектами, приведены алгоритмы построения квазимонотоннных схем повышенного порядка точности.

Глава 3 посвящена вопросам оптимизации наноразмерных полупроводниковых структур. Рассмотрен класс задач, связанный с определением оптимального легирования барьерного слоя. Построены эффективные алгоритмы оптимизации, основанные на градиентных методах. Приведены результаты решения задач оптимального легирования барьерного слоя, обеспечивающего заданную концентрацию электронов в канале проводимости в полупроводниковых гетероструктурах. Полученные в ходе вычислительного эксперимента результаты согласуются с современной тенденцией к переходу от однородного профиля легирования к планарному δ-легированию в технологиях изготовления полевых транзисторов.

Проведенные с применением данного подхода исследования могут быть использованы на практике для создания новых полупроводниковых гетероструктур с прогнозируемыми свойствами для авиационной и ракетно-космической отрасли.

Представленные в учебном пособии модели, методы и алгоритмы позволяют оперативно проводить многовариантный анализ многослойных наноразмерных полупроводниковых структур, что создает основу для решения ряда оптимизационных задач, актуальных для современной электроники.

ВВЕДЕНИЕ

Цель настоящего пособия – ознакомить студентов с современными подходами к моделированию и оптимизации полупроводниковых наноструктур, основными математическими моделями и численными методами, используемыми в этой области.

Во введении кратко рассмотрим некоторые физические аспекты, характерные для данного направления.

Одной из современных тенденций развития высокочастотной полупроводниковой техники является стремление к достижению максимальных концентраций носителей заряда при максимально возможной их подвижности [1]. Традиционный путь создания в полупроводниках зон с заданным типом носителей заряда связан с легированием соответствующих областей донорными или акцепторными примесями. При этом повышение концентрации образующихся свободных носителей заряда приводит к снижению их подвижности вследствие усиления кулоновского рассеяния на ионизированных примесных центрах. Таким образом, стремление к достижению высокой концентрации носителей заряда вступает в противоречие с достижением высокой подвижности. Использование многослойных полупроводниковых гетероструктур позволяет отделить области формирования носителей заряда от зоны, где осуществляется их перенос. Это достигается созданием потенциальных барьеров, ограничивающих движение носителей заряда в одном или более направлениях.

Различают структуры с квантовым ограничением за счет внешнего и внутреннего электрического поля [1].

В качестве примера систем первого типа можно привести МОП (металл/окисел/полупроводник) структуры полевых транзисторов. Типичная МОП структура состоит из монокристаллической кремниевой подложки с p-типом проводимости, слоя ди-

оксида кремния и верхнего металлического электрода (затвора). Положительное смещение, приложенное к затвору, отделенному от полупроводника диэлектрическим слоем, притягивает электроны, обедняя приповерхностный слой полупроводника дырками. По мере увеличения внешнего напряжения в приповерхностном слое полупроводника образуется тонкий слой, обогащенный электронами, который называется инверсионным слоем в связи изменением типа носителей в полупроводнике. Таким образом, образуется квантовый колодец, стенки которого формируются границей диэлектрик/полупроводник и электрическим полем, прижимающим электроны к этой границе. Движение электронов ограничивается в направлении, нормальном к гетерогранице. В связи с малой шириной колодца (толщиной инверсного слоя) состояние и подвижность электронов в нем могут быть описаны в терминах двумерного электронного газа.

Ярким примером систем с квантовым ограничением за счет внутреннего электрического поля являются модуляционно-легированные структуры. Они образуются полупроводниками с различной шириной запрещенной зоны. При этом в полупроводник с большей шириной запрещенной зоны обычно вводится донорная примесь. Электроны донорных атомов вследствие тепловой активации переходят в соседнюю область с более низкой потенциальной энергией, где оказываются захваченными в приграничной области, поскольку, потеряв энергию, не способны преодолеть потенциальный барьер. Они индуцируют электростатический потенциал, который наряду с притяжением положительно заряженными ионами примеси, прижимает их к гетерогранице. Таким образом, в окрестности гетерограницы в полупроводнике с меньшей шириной запрещенной зоны (узкозонный слой) образуется квантовый колодец для электронов шириной порядка нескольких нанометров. Движение электронов в направлении, нормальном к гетерогранице, ограничено, энергетические уровни квантованы. При этом электроны на этих уровнях могут свободно двигаться в плоскости гетерограницы. Образуется двумерный электронный газ. Один из широко распространенных вариантов модуляционно-легированной

структуры – комбинация легированного слоя AlGaAs (барьерный слой с большей шириной запрещенной зоны) и нелегированного слоя GaAs (слой с меньшей шириной запрещенной зоны). Такая структура обеспечивает высокую подвижность носителей, но слоевая концентрация электронов при этом сравнительно невысока – до $5 \cdot 10^{11}$ см$^{-2}$.

Весьма перспективными с точки зрения повышения концентрации и подвижности носителей заряда являются структуры с наличием поверхностного заряда на гетерогранице. Примером может служить система AlGaN/GaN (кристаллическая структура вюрцита). Возникающий вследствие спонтанной и пьезоэлектрической поляризации вюрцитных структур поверхностный заряд на интерфейсе создает условия для образования квантового колодца с близким к треугольному профилем потенциала. Ситуация очень похожа на описанную выше для модуляционно-легированных структур, но здесь нет необходимости в легировании (хотя оно возможно), т.е. практически исключается механизм рассеяния носителей на примесных центрах. Подробный анализ таких структур приведен в работе [2]. Достигаемый уровень концентрации носителей в ДЭГ – порядка 10^{13} см$^{-2}$.

Таким образом, для проектирования и оптимизации современных полупроводниковых гетероструктур большое значение имеет разработка математических моделей и методов, отражающих квантово-механический характер поведения носителей заряда.

В работах авторов настоящего пособия реализована следующая схема многомасштабного моделирования полупроводниковых гетероструктур [3].

Выделены три характерных масштабных уровня. Описание системы на атомарном уровне осуществляется с использованием кристаллографической информации и квантово-механической модели, в основе которой лежит теория функционала электронной плотности [4, 5]. Квантово-механическое моделирование позволяет определить зонную структуру и базовые свойства гетеросистемы, выявить поляризационные эффекты и рассчитать плотности зарядов на интерфейсах между слоями [6]. Полученная информация

передается в модель наномасштабного уровня, где используется для расчета распределения носителей заряда в гетероструктуре. На данном уровне также используется квантово-механическое описание, математическая модель представляет собой систему уравнений Шрёдингера и Пуассона [7]. Полученные в результате решения этих уравнений данные о волновых функциях и распределении плотности носителей заряда поперек слоистой структуры поступают в модель следующего масштабного уровня, где осуществляется расчет подвижности носителей заряда в продольном направлении. При этом учитывается широкий спектр механизмов рассеяния электронов – рассеяние на оптических и акустических фононах, шероховатостях гетерограницы, на заряженных центрах и дислокациях, пьезоэлектрическое рассеяние [8, 9]. Сравнение результатов расчетов с экспериментальными данными показало достаточно высокую точность моделирования как в плане расчета концентрации носителей в двумерном электронном газе, так и касательно подвижности электронов.

Таким образом, применение методов математического моделирования позволяет достаточно точно предсказывать свойства полупроводниковых наноструктур, что создает основу для решения ряда оптимизационных задач, актуальных для современной электроники [10].

Глава 1

МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ НОСИТЕЛЕЙ ЗАРЯДА В НАНОРАЗМЕРНЫХ ПОЛУПРОВОДНИКОВЫХ ГЕТЕРОСТРУКТУРАХ

В настоящей главе основное внимание уделено вопросам численного моделирования полупроводниковых гетероструктур на наномасштабном уровне. На данном уровне локальные вычислительные процедуры, служащие для решения уравнений Шрёдингера и Пуассона, объединены глобальным итерационным процессом, направленным на согласование решений. Принципиальным моментом является сильная взаимосвязь уравнений, характерная для рассматриваемого класса задач, что приводит к проблеме сходимости глобальных итераций. В особой мере эта проблема проявляется для вюрцитных структур, когда на гетероинтерфейсах появляется мощный локализованный источник заряда.

1.1. Математическая модель. Система уравнений Шрёдингера–Пуассона

Типичная схема наноразмерной полупроводниковой гетероструктуры представлена на рис. 1.1, заимствованном из работы [11].

Показана гетероструктура, выращенная в Институте физики полупроводников СО РАН на основе нитрида галлия и тройных растворов [11]. Наличие слоев, выполненных из полупроводников с различной шириной запрещенной зоны, в сочетании с поляризационными эффектами обеспечивает создание в окрестности гетерограницы в слое с меньшей шириной запрещенной зоны

(GaN) квантового колодца для электронов шириной порядка нескольких нанометров. Движение электронов в направлении, нормальном к гетерогранице, ограничено, энергетические уровни квантованы. При этом электроны на этих уровнях могут свободно двигаться в плоскости гетерограницы. Образуется двумерный электронный газ (схематично область образования двумерного электронного газа показана точками на рис. 1.1).

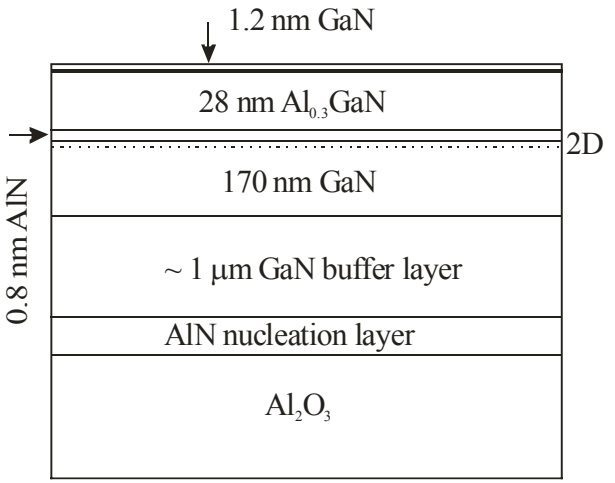


Рис. 1.1. Многослойная полупроводниковая гетероструктура [11]

Математическая модель, описывающая распределение электронов в таких структурах, представляет собой систему уравнений Шрёдингера и Пуассона [12, 13]. Отметим, что в полупроводниковых приборах движение носителей заряда может ограничиваться не только в одном (двумерный электронный газ), но и в двух (квантовая нить), а также в трех (квантовая точка) направлениях [1]. Как правило, это достигается приложением напряжения в соответствующих зонах. Таким образом, в общем случае задача моделирования полупроводниковых гетероструктур ставится в пространственной постановке.

Математическая модель имеет вид:

$$-\frac{\hbar^2}{2}\nabla\cdot\left(\frac{1}{m^*}\nabla\psi\right)+(-e\varphi+\Delta E_c)\psi=E\psi \qquad (1)$$

$$\nabla\cdot(\varepsilon\nabla\varphi)=-e(N_D(z)-N_A(z)-n(\mathbf{r}))+\sum_l \sigma_l\delta(z-z_l) \qquad (2)$$

$$n(\mathbf{r})=\sum_i(\psi_i(\mathbf{r}))^2 n_i \qquad (3)$$

$$n_i=k_B T\frac{m^*}{\pi\hbar^2}\ln\left(1+\exp\left(\frac{E_F-E_i}{k_B T}\right)\right). \qquad (4)$$

Здесь E_i и $\psi_i(\mathbf{r})$ – энергетические уровни и соответствующие им волновые функции, $n(\mathbf{r})$ – электронная плотность, $\hbar$ – постоянная Планка, e – заряд электрона, m^* – эффективная масса электрона, E_F – положение уровня Ферми, $\varphi(\mathbf{r})$ – электростатический потенциал, N_d, N_A – концентрации донорной и акцепторной примесей, σ_l – плотности зарядов на интерфейсах, δ – дельта функция, z_l – местоположения интерфейсов, ε – диэлектрическая постоянная материала, ΔE_c – сдвиг зоны проводимости материала, k_B – постоянная Больцмана, T – температура.

В модели (1)–(4) отражен тот факт, что значения $m*$, ε и ΔE_c могут меняться от слоя к слою. При этом температура в гетероструктуре предполагается постоянной.

При записи уравнений (1)–(4) выделено направление поперек слоев (координата z), что отражает послойный рост структуры и зависимость ряда характеристик только этой координаты.

Задача (1)–(4) решается, как правило, в прямоугольной области Ω (пример двумерной области представлен на рис. 1.1), что характерно для большинства полупроводниковых устройств. Постановка условий на границе Γ области Ω отражает специфику функционирования гетероструктуры. Область Ω в расчетах обычно значительно больше области, где реализуется квантовое ограничение, так что справедливым является постановка однородного условия Дирихле для волновой функции на границах –

$\psi|_{\Gamma} = 0$. Для электростатического потенциала в зависимости от условий функционирования гетероструктуры на различных участках границы могут задаваться как значения приложенных напряжений (граничные условия Дирихле, $\varphi|_{\Gamma} = \varphi_g$), так и значения напряженности электрического поля (граничные условия Неймана, $\left.\frac{\partial\varphi}{\partial\eta}\right|_{\Gamma} = f_g$, где $\frac{\partial\varphi}{\partial\eta}$ – производная по нормали к границе). Первый вариант характерен, например, для задач, в которых квантовые ограничения достигаются за счет внешнего электрического поля (квантовые нити в двумерном случае и квантовые точки в трехмерном случае), а также в зонах контакта полупроводника с металлами. Второй вариант часто используется на границе со стороны подложки, где задается нулевое значение напряженности (однородное условие Неймана).

Решение спектральной задачи (1) (уравнение Шрёдингера) зависит от электростатического потенциала φ, распределение которого в структуре $\varphi(\mathbf{r})$ определяется уравнением Пуассона (2). При этом в правую часть уравнения Пуассона входит электронная плотность $n(\mathbf{r})$, которая, в свою очередь, определяется энергетическими уровнями E_i и волновыми функциями $\psi_i(\mathbf{r})$ согласно статистике Ферми-Дирака (3), (4). Задача, таким образом, имеет ярко выраженный сопряженный характер.

Самосогласованное решение этой системы дает искомые энергетические уровни E_i и соответствующие им волновые функции $\psi_i(\mathbf{r})$, профиль потенциальной ямы $V(\mathbf{r})$, а также распределение электронной плотности в гетероструктуре $n(\mathbf{r})$.

Кратко остановимся на том, как в математической модели отражены основные факторы, влияющие на возникновение выделенных каналов переноса носителей в окрестности гетерограницы.

Выражение для потенциальной энергии электрона в уравнении (1) $V = -e\varphi + \Delta E_c$ включает величину сдвига зоны проводимости полупроводниковых материалов ΔE_c, что обусловливает со-

здание потенциального барьера, удерживающего электроны в квантовом колодце.

Для полупроводников с кристаллической структурой сфалерита важнейшим фактором является концентрация доноров в барьерном слое N_D, входящая в правую часть уравнения Пуассона. Для полупроводников с вюрцитной кристаллической структурой определяющую роль играют поляризационные эффекты и связанные с ними поверхностные заряды на гетерограницах. Этот фактор представлен в математической модели слагаемым $\sum_l \sigma_l \delta(z - z_l)$ в уравнении Пуассона. Здесь двумерный электронный газ возникает даже без введения доноров в барьерный слой. Эффекты спонтанной и пьезоэлектрической поляризации описываются квантово-механическими моделями на атомарном уровне, что позволяет получить плотности поверхностного заряда на интерфейсах [6].

Разумеется, на распределение электронной плотности в слоистой гетероструктуре влияют толщины и расположение слоев, а также другие параметры математической модели.

Многие задачи расчета и оптимизации слоистых полупроводниковых структур решаются в пространственно-одномерной постановке, позволяющей определить основные характеристики двумерного электронного газа. Математическая модель в этом случае выглядит следующим образом:

$$-\frac{\hbar^2}{2}\frac{d}{dz}\left(\frac{1}{m^*(z)}\frac{d\psi}{dz}\right) + V(z)\psi(z) = E\psi(z)$$

$$\frac{d}{dz}\left(\varepsilon(z)\frac{d\varphi}{dz}\right) = -e(N_D(z) - N_A(z) - n(z)) + \sum_l \sigma_l \delta(z - z_l)$$

$$V(z) = -e\varphi(z) + \Delta E_c(z)$$

$$n(z) = \sum_i (\psi_i(z))^2 n_i(z)$$

$$n_i(z) = k_B T \frac{m^*(z)}{\pi\hbar^2} \ln\left(1 + \exp\left(\frac{E_F - E_i}{k_B T}\right)\right).$$

Отметим, что сложности, связанные со сходимостью глобальных итераций для получения самосогласованного решения уравнений Шрёдингера и Пуассона, отчетливо проявляются при решении задачи в одномерной постановке. Поэтому анализ, проводимый в настоящей работе, базируется на этой модели.

Остановимся подробнее на постановке граничных условий.

На границах системы (z=0, z=L, где L – общая толщина слоистой структуры) должны выполняться условия равенства нулю волновых функций:

$$\psi(0) = 0\,, \quad \psi(L) = 0\,.$$

На левой границе (z=0), как правило, задается потенциальный барьер φ_b, образующийся в приконтактном слое полупроводника, граничащего с металлом (барьер Шоттки). Кроме того, возможно задание смещения φ_s вследствие приложенного напряжения. На правой границе (z=L) поведение потенциала определяется положением зоны проводимости соответствующего материала относительно уровня Ферми. В большинстве случаев естественным условием здесь представляется отсутствие электрического поля. Таким образом, граничные условия для уравнения Пуассона имеют вид:

$$\varphi(0) = \varphi_b + \varphi_s\,, \quad \varphi'(L) = 0\,.$$

1.2. Глобальный итерационный процесс для получения самосогласованного решения

Как отмечалось выше, сопряженный характер задачи обусловливает необходимость итерационного согласования решений уравнений Шрёдингера и Пуассона. С этой целью локальные вычислительные процедуры, служащие для решения уравнений, объединены глобальным итерационным циклом. Вычислительный процесс можно схематично представить следующим образом.

В результате выполнения очередной итерации имеется распределение электронной плотности в системе $n^{(k)}(\mathbf{r})$.

С использованием этой информации решается уравнение Пуассона:

$$\nabla \cdot (\varepsilon \nabla \varphi^{(k+1)}) = -e(N_D - N_A - n^{(k)}) + \sum_l \sigma_l \delta(z - z_l) .$$

Полученное распределение потенциала в системе $\varphi^{(k+1)}$ используется при последующем решении уравнения Шрёдингера:

$$-\frac{\hbar^2}{2} \nabla \cdot (\frac{1}{m^*} \nabla \psi^{(k+1)}) + (-e\varphi^{(k+1)} + \Delta E_c)\psi^{(k+1)} = E^{(k+1)} \psi^{(k+1)} .$$

В результате получаются значимые (вносящие вклад в повышение концентрации носителей) уровни энергии и соответствующие им волновые функции $E_i^{(k+1)}$, $\psi_i^{(k+1)}$.

После этого с использованием статистики Ферми–Дирака рассчитывается новое распределение электронной плотности в системе:

$$n^{(k+1)} = \sum_i (\psi_i^{(k+1)})^2 k_B T \frac{m^*}{\pi \hbar^2} \ln\left(1 + \exp\left(\frac{E_F - E_i^{(k+1)}}{k_B T}\right)\right) .$$

Для рассматриваемого класса задач характерна сильная взаимосвязь уравнений, что приводит к проблеме сходимости последовательных приближений. Для достижения сходимости возможно использование нижней релаксации:

$$\overline{n}^{(k+1)} = \omega n^{(k+1)} + (1 - \omega) n^{(k)} ,$$

где $\overline{n}^{(k+1)}$ – итоговое распределение электронной плотности на данной итерации.

При этом, согласно вычислительным экспериментам, проведенным для широкого класса полупроводниковых гетероструктур, приходится использовать очень малый параметр релаксации ω (типичный уровень значений параметра ω – 0.025), что существенно сказывается на скорости сходимости глобальных итераций. В ряде задач удается построить алгоритм с адаптивным релаксационным параметром, однако трудно добиться универсальности такого подхода. Отметим, что подобные сложности применительно к моделированию структур сфалеритного типа

отмечались в работе [14]. Положение существенно осложняется при рассмотрении полупроводников с вюрцитной кристаллической структурой, когда появляются локализованные на гетерограницах заряды.

В настоящей работе для решения данного класса задач применяется подход, основанный на локальной аппроксимации неявной зависимости электронной плотности от потенциала [14]. Этот подход хорошо зарекомендовал себя при моделировании распределения носителей заряда в гетероструктурах сфалеритного типа [15, 16]. Аппроксимационное выражение для электронной плотности получено с использованием теории возмущений и имеет следующий вид:

$$n(\varphi) \approx \sum_i \psi_i^2 k_B T \frac{m^*}{\pi\hbar^2} \ln\left(1+\exp\left(\frac{E_F - E_i + e(\varphi-\varphi^{(k)})}{k_B T}\right)\right), \qquad (5)$$

где $\varphi^{(k)}$ – потенциал, полученный на предыдущей глобальной итерации.

Так как использование зависимости (5) позволяет существенно повысить эффективность вычислительных алгоритмов, приведем здесь схему ее вывода.

Рассмотрим приращение потенциала $\delta\varphi = \varphi - \varphi^{(k)}$. Ему соответствует приращение электронной плотности, обусловленное изменением волновых функций $\delta\psi_i$ и энергетических уровней δE_i:

$$\delta n = A\sum_i (\psi_i + \delta\psi_i)^2 \ln\left(1+\exp\left(F_i - \frac{\delta E_i}{k_B T}\right)\right) - A\sum_i (\psi_i)^2 \ln(1+\exp F_i),$$

где $A = k_B T \dfrac{m^*}{\pi\hbar^2}$, $F_i = \dfrac{E_F - E_i}{k_B T}$.

В линейном приближении δn имеет вид:

$$\delta n = -A\sum_i \psi_i^2 \frac{\delta E_i}{k_B T} \frac{\exp F_i}{1+\exp F_i} + 2A\sum_i \psi_i \delta\psi_i \ln(1+\exp F_i).$$

Согласно линейной теории возмущений [17]

$$\delta E_i = -e(\psi_i, \delta\varphi\psi_i)\,,\ \delta\psi_i = -e\sum_{j\neq i}\psi_j \frac{(\psi_j, \delta\varphi\psi_i)}{E_i - E_j}\,,$$

где $(\psi_j, \delta\varphi\psi_i) = \int\limits_{\Omega} \psi_j \delta\varphi\psi_i dv$.

С учетом симметричности второго слагаемого по индексам выражение для приращения электронной плотности примет вид:

$$\delta n = Ae\sum_i \psi_i^2 \frac{(\psi_i, \delta\varphi\psi_i)}{k_B T}\frac{\exp F_i}{1+\exp F_i} -$$

$$-Ae\sum_i\sum_{j\neq i}\psi_i\psi_j(\psi_j, \delta\varphi\psi_i)\frac{\ln(1+\exp F_i) - \ln(1+\exp F_j)}{E_i - E_j} \approx$$

$$\approx \frac{Ae}{k_B T}\sum_i \psi_i^2(\psi_i, \delta\varphi\psi_i)\frac{\exp F_i}{1+\exp F_i} + \frac{Ae}{k_B T}\sum_i\sum_{j\neq i}\psi_i\psi_j(\psi_j, \delta\varphi\psi_i)\frac{\exp F_i}{1+\exp F_i}\,.$$

Свойство ортонормированности волновых функций дает $\sum\limits_{j\neq i}\psi_j(\psi_j, \delta\varphi\psi_i) = \delta\varphi\psi_i - \psi_i(\psi_i, \delta\varphi\psi_i)$, что приводит к следующему выражению:

$$\delta n = \frac{Ae}{k_B T}\sum_i \psi_i^2 \frac{\exp F_i}{1+\exp F_i}\delta\varphi\,.$$

Это позволяет получить искомую аппроксимацию зависимости электронной плотности от потенциала:

$$n(\varphi) \approx n^{(k)} + \delta n = A\sum_i \psi_i^2 \ln\left(1+\exp(\frac{E_F - E_i}{k_B T})\right) +$$

$$+\frac{Ae}{k_B T}\sum_i \psi_i^2 \frac{\exp F_i}{1+\exp F_i}\delta\varphi \approx A\sum_i \psi_i^2 \ln\left(1+\exp\left(\frac{E_F - E_i + e\delta\varphi}{k_B T}\right)\right).$$

Использование такого подхода приводит к модифицированному уравнению Пуассона с явно заданным нелинейным слагаемым $n(\varphi)$:

$$\nabla\cdot(\varepsilon\nabla\varphi) = -e(N_D - N_A - n(\varphi)) + \sum_l \sigma_l \delta(z - z_l)\,.$$

Для решения модифицированного уравнения Пуассона эффективным приемом является линеаризация (метод Ньютона) [18]:

$$\varphi = \varphi^{(p)} + \xi$$

$$\nabla \cdot (\varepsilon \nabla \xi) = g\xi + q$$

$$g = e\frac{\delta n}{\delta \varphi}(\varphi^{(p)})$$

$$q = -\nabla \cdot (\varepsilon \nabla \varphi^{(p)}) + (-e(N_D - N_A - n(\varphi^{(p)})) + \sum_l \sigma_l \delta(z - z_l)\,.$$

Здесь индекс (p) указывает на функцию, полученную на предыдущей локальной итерации.

Для решения уравнений Шрёдингера и Пуассона в настоящей работе применяются конечно-разностные методы. Аппроксимация производных с использованием центрально-разностных отношений позволяет свести задачу на каждой глобальной итерации к решению спектральной задачи для симметричной разреженной матрицы ленточной структуры (дискретный аналог уравнения Шрёдингера) и к решению системы алгебраических уравнений (дискретный аналог уравнения Пуассона). Учитывая, что размерности матриц могут быть достаточно велики в связи с необходимостью должного разрешения относительно тонкого слоя, содержащего двумерный электронный газ, а также то, что вычисления многократно повторяются в итерационном процессе, важным требованием к алгоритмам является экономичность вычислений. Наибольшие вычислительные затраты связаны с решением дискретного аналога уравнения Шрёдингера. В этой связи существенную роль играет учет особенностей расположения значимых энергетических уровней. Согласно статистике Ферми-Дирака вклад в повышение концентрации носителей в двумерном электронном газе вносят электроны, находящиеся на нижних энергетических уровнях. Информация о нижней границе поступает из решения уравнения Пуассона на текущей глобальной итерации: $E_i^{(k)} > \min V^{(k)}(\mathbf{r})$. Верхний уровень может быть оценен с использованием зависимости (4). Здесь принци-

пиально то, что в силу особенностей распределения Ферми-Дирака количество электронов, обладающих энергией, превышающей уровень Ферми более, чем на k_bT, убывает практически экспоненциально. Такая локализация позволяет использовать экономичные методы поиска собственных значений в сочетании с обратными итерациями для определения сеточных волновых функций.

1.3. Методы решения уравнения Шрёдингера

Фундаментальным уравнением, описывающим поведение волновой функции частицы, в квантовой механике является уравнение Шрёдингера:

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi ,$$

где $\hat{H}$ – оператор Гамильтона, представляющий собой сумму операторов кинетической и потенциальной энергий, $\hat{H} = \hat{T} + \hat{U} = -\frac{\hbar^2}{2m}\Delta + U$.

Как правило, интерес представляют стационарные состояния, когда вероятность местонахождения частицы не зависит от времени. Такие состояния описываются стационарным уравнением Шрёдингера:

$$\hat{H}\psi = E\psi .$$

Здесь E – энергия частицы.

Данное уравнение определяет волновую функцию ψ, как собственную функцию оператора Гамильтона, а соответствующий уровень энергии E как собственное значение этого оператора.

С решением уравнения Шрёдингера связаны, как правило, наибольшие вычислительные затраты, при расчетах электронной плотности в гетероструктурах. Для решения этой задачи в совре-

менной практике математического моделирования весьма распространенным является метод стрельбы [см., напр., 18]. Согласно этому методу решение задачи сводится к последовательности решения задач Коши, и итерационному определению значений параметра E для которых одновременно выполняются оба однородных условий на границах. При этом второе начальное условие (накладываемое на производную волновой функции) может быть выбрано произвольно, т.к. в дальнейшем волновая функция подлежит нормировке. К достоинствам метода следует отнести экономичность вычислений и возможность использования неравномерной сетки. К недостаткам метода стрельбы относится, в первую очередь, ограниченность его применения одномерными задачами. В этой связи в настоящей работе используются более универсальные методы решения уравнения Шрёдингера.

1.3.1. Конечно-разностная аппроксимация

Одним из таких универсальных подходов является применение конечно-разностных методов.

В наиболее простой, пространственно – одномерной постановке, задача имеет вид:

$$-a\frac{d^2\psi}{dz^2}+V(z)\psi(z)=E\psi(z)\,,\ \ \psi(0)=\psi(L)=0\,,\ \ a=\frac{\hbar^2}{2m}\,.$$

Центрально-разностная аппроксимация на равномерной сетке $\{z_j, j=0,...,N\}$ с шагом $h=L/N$ дает:

$$-a\frac{\psi_{j-1}-2\psi_j+\psi_{j+1}}{h^2}+V(z_j)\psi_j=E\psi_j\,,\ \ j=1,\ldots,N-1,$$

$$\psi_0=\psi_n=0\,.$$

Задача может быть представлена в традиционном для спектральных задач линейной алгебры матрично-векторном виде:

$$A\boldsymbol{\psi}=\lambda\boldsymbol{\psi}\,,$$

где матрица **A** имеет вид

$$A=\begin{pmatrix} 2a/h^2+V_1 & -a/h^2 & & & \\ -a/h^2 & 2a/h^2+V_2 & -a/h^2 & & \\ \ldots\ldots & \ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots & & \\ & -a/h^2 & 2a/h^2+V_{N-2} & -a/h^2 & \\ & & -a/h^2 & 2a/h^2+V_{N-1} & \end{pmatrix}.$$

Матрица $A_{N-1*N-1}$ симметрична и трехдиагональна, может быть представлена ее ненулевыми элементами:

$$\alpha_j=\frac{-a}{h^2},\ j=2,\ldots N-1,$$

$$\beta_j=\frac{2a}{h^2},\ j=1,\ldots N-1,$$

$$\gamma_j=\frac{-a}{h^2},\ j=1,\ldots N-2.$$

В гетерогенных (слоистых) структурах эффективная масса электрона не является постоянной величиной. В этом случае постановка задачи принимает вид:

$$-\frac{d}{dz}\left(a(z)\frac{d\psi}{dz}\right)+V(z)\psi(z)=E\psi(z),$$

$$\psi(0)=\psi(L)=0,$$

$$a(z)=\frac{\hbar^2}{2m(z)}.$$

Аппроксимация центральными разностями дает следующий дискретный аналог:

$$\frac{(-a_{j+1/2}(\psi_{j+1}-\psi_j)+a_{j-1/2}(\psi_j-\psi_{j-1}))}{h^2}+V(z_j)\psi_j=E\psi_j,$$

где $a_{j\pm1/2}=\dfrac{2a_ja_{j\pm1}}{a_j+a_{j\pm1}}$.

Этот результат может быть получен с помощью интегро-интерполяционного метода (метод контрольного объема) при

условии расположения граней контрольного объема посредине между узлами сетки и в предположении о постоянстве функции $a(z)$ в пределах ячейки, что характерно для рассматриваемых в работе слоистых гетероструктур.

Матрица системы, как и в случае постоянной эффективной массы электрона, является при этом трехдиагональной и симметричной. Ее ненулевые элементы равны:

$$\alpha_j = \frac{-a_{j-1/2}}{h^2}, \quad \beta_j = \frac{a_{j-1/2} + a_{j+1/2}}{h^2} + V_j, \quad \gamma_j = \frac{-a_{j+1/2}}{h^2}.$$

Если задача рассматривается в двумерной или трехмерной постановках, то матрицы сохраняют разреженную ленточную структуру, но не являются трехдиагональными. Для примера рассмотрим стационарное двумерное уравнение Шрёдингера в прямоугольной области $[0, L_x] \times [0, L_y]$:

$$-a\frac{\partial^2\psi}{\partial x^2} - a\frac{\partial^2\psi}{\partial y^2} + V(x,y)\psi(x,y) = E\psi(x,y),$$

$$\psi(0,y) = \psi(L_x,y) = \psi(x,0) = \psi(x,L_y) = 0.$$

Аппроксимируя уравнение Шрёдингера на сетке $\omega_{h_1,h_2} = \{x_i = ih_x, i = \overline{0,N_x};\ y_j = jh_y, j = \overline{0,N_y}\}$, получим следующий дискретный аналог:

$$-a\frac{\psi_{ij-1} - 2\psi_{ij} + \psi_{ij+1}}{h_x^2} - a\frac{\psi_{ij-1} - 2\psi_{ij} + \psi_{ij+1}}{h_y^2} + V(z_{ij})\psi_{ij} = E\psi_{ij},$$

$$i = \overline{1, N_x - 1}, \quad j = \overline{1, N_y - 1}.$$

Если, занумеровав узлы сетки (например, $l = i + j - 1$), представить неизвестные ψ_{ij} в виде вектора $\boldsymbol{\psi}$, можно получить систему линейных алгебраических уравнений с пятидиагональной матрицей $A_{M\times M}$, $M = (N_x - 1)\times(N_y - 1)$. Для наглядности приведем вид матрицы для $N_x = N_y = 4$:

$$A=\begin{pmatrix} \beta_1 & \alpha_x & 0 & \alpha_y & & & & & \\ \alpha_x & \beta_2 & \alpha_x & 0 & \alpha_y & & & & \\ 0 & \alpha_x & \beta_3 & 0 & 0 & \alpha_y & & & \\ \alpha_y & 0 & 0 & \beta_4 & \alpha_x & 0 & \alpha_y & & \\ & \alpha_y & 0 & \alpha_x & \beta_5 & \alpha_x & 0 & \alpha_y & \\ & & \alpha_y & 0 & \alpha_x & \beta_6 & 0 & 0 & \alpha_y \\ & & & \alpha_y & 0 & 0 & \beta_7 & \alpha_x & 0 \\ & & & & \alpha_y & 0 & \alpha_x & \beta_8 & \alpha_x \\ & & & & & \alpha_y & 0 & \alpha_x & \beta_9 \end{pmatrix},$$

где $\beta_l = 2a\left(\frac{1}{h_x^2}+\frac{1}{h_y^2}\right)+V_l$, $l=1,...,M$, $\alpha_x = -\,a/h_x^2$, $\alpha_y = -\,a/h_y^2$.

В матрице показаны только пять диагоналей, остальные элементы равны нулю. Отчетливо прослеживается ленточная структура матрицы. Матрица, как и ранее, является разреженной и симметричной.

В задачах моделирования полупроводниковых гетерогенных структур могут возникать ситуации, когда определенные зоны требуют высокого пространственного разрешения (например, зона формирования двумерного электронного газа), тогда как в остальной части расчетной области можно ограничиться довольно грубой сеткой (например, дальняя зона буферного слоя). В таких случаях предпочтительным может оказаться применение неравномерных сеток. При этом, однако, нарушается симметричность получающихся матриц, что утяжеляет алгоритмы решения спектральной задачи.

В одномерном случае эффективным приемом, позволяющим сочетать аппроксимацию уравнения Шрёдингера на неравномерной сетке с возможностью применения алгоритмов поиска собственных значений и собственных векторов симметричных матриц, является простая симметризация дискретного аналога.

В случае неравномерной сетки будем иметь:

$$\alpha_j = \frac{-2a}{h_j(h_j + h_{j+1})}, \quad j = 2,...N-1,$$

$$\beta_j = 2a\left(\frac{1}{h_j} + \frac{1}{h_{j+1}}\right)\frac{1}{(h_j + h_{j+1})}, \quad j = 1,...N-1,$$

$$\gamma_j = \frac{-2a}{h_{j+1}(h_j + h_{j+1})}, \quad j = 1,...N-2.$$

Источником несимметричности матрицы является наличие множителя $\frac{1}{(h_j + h_{j+1})}$ в коэффициентах α_j, γ_j. Если дважды домножить матрицу A слева на диагональную матрицу D с элементами $(h_j + h_{j+1})^{1/2}$, то полученная матрица $B = DDA$, будет симметричной. При этом матрица $\tilde{A} = D^{-1}BD^{-1}$ будет подобна исходной матрице A ($\tilde{A} = D^{-1}DDAD^{-1} = DAD^{-1}$). Таким образом, задача нахождения собственных значений матрицы A свелась к задаче нахождения собственных значений симметричной матрицы $\tilde{A}$. Собственные векторы матриц при этом будут связаны простым соотношением: $\boldsymbol{\psi}_A = D^{-1}\boldsymbol{\psi}_{\tilde{A}}$, т.к. $\tilde{A}D\boldsymbol{\psi}_A = D^{-1}DDAD^{-1}D\boldsymbol{\psi}_A = \lambda D\boldsymbol{\psi}_A$. Отметим, что все преобразования могут быть выполнены экономично, поскольку матрица D является диагональной.

Завершая описание методов дискретизации уравнения Шрёдингера, обратим внимание на следующее обстоятельство. Во многих случаях спектр дифференциального оператора состоит из дискретной (связанные состояния) и непрерывной частей. При конечно-разностной аппроксимации это может приводить к плохой обусловленности матрицы. Общим приемом для улучшения обусловленности является сдвиг потенциальной энергии на величину V' [19], что приводит к видоизменной спектральной задаче $A'\boldsymbol{\psi}' = \lambda'\boldsymbol{\psi}'$, $A' = A + V'E$, где E – единичная матрица. Собственные значения матриц A и A' связаны соотношением $\lambda = \lambda' - V'$, а собственные векторы одинаковы $\boldsymbol{\psi} = \boldsymbol{\psi}'$.

Перейдем к рассмотрению методов нахождения собственных векторов и собственных значений матриц, получающихся в результате дискретизации уравнения Шрёдингера. Учитывая, что размерности матриц могут быть достаточно велики в связи с необходимостью должного разрешения зон повышенной электронной плотности в гетероструктурах, а также то, что вычисления обычно многократно повторяются в итерационном процессе (глобальные итерации), важным требованием к алгоритмам является экономичность вычислений.

1.3.2. Неявный QR-алгоритм со сдвигом

Начнем рассмотрение с трехдиагональной симметричной матрицы. Для нахождения полного спектра (всех собственных значений) таких матриц эффективным подходом является неявный QR-алгоритм с одинарным сдвигом. Неявным такой алгоритм называется, т.к. в отличие от традиционного QR-алгоритма, где на каждом шаге строится QR-разложение матрицы, здесь матрица Q строится неявно как произведение матриц вращения.

Напомним, что в традиционном QR-алгоритме [напр., 20], каждая итерация реализуется в два этапа. На первом этапе осуществляется разложение матрицы $A^{(k)}$ (k – номер итерации) в произведение ортогональной $Q^{(k)}$ и верхней треугольной $R^{(k)}$ матриц $A^{(k)} = Q^{(k)} R^{(k)}$, а на втором – полученные матрицы перемножаются в обратном порядке $A^{(k+1)} = R^{(k)} Q^{(k)}$. При этом в силу ортогональности $Q^{(k)}$ $A^{(k+1)} = R^{(k)} Q^{(k)} = Q^{(k)T} Q^{(k)} R^{(k)} Q^{(k)} = Q^{(k)T} A^{(k)} Q^{(k)}$, т.е. матрицы $A^{(k+1)}$ и $A^{(k)}$ подобны.

Существенным недостатком рассмотренного выше алгоритма является большое число операций ($O(n^3)$, где n – размерность матрицы), необходимое для QR-факторизации матрицы на каждой итерации. Эффективность QR-алгоритма может быть повышена, если предварительно с помощью преобразования подобия привести матрицу к верхней Хессенберговой форме, в которой равны нулю все элементы, находящиеся ниже главной диагонали

за исключением элементов первой поддиагонали. Здесь принципиально то, что в дальнейшем, в ходе QR-итераций, матрицы $A^{(k)}$ сохраняют верхнюю Хессенбергову форму, что позволяет более экономно проводить их QR-разложение. Очевидно, что трехдиагональная матрица уже имеет Хессенбергову форму и дальнейшее повышение эффективности вычислительного процесса заключается в оптимизации вычислений на итерации $A^{(k+1)} = Q^{(k)T} A^{(k)} Q^{(k)}$. Рассмотрим последовательность шагов, позволяющих экономично провести такое преобразование [21, 22].

Изначально матрица $A_0 = A^{(k)}$ имеет следующую структуру (знак x обозначает ненулевые элементы):

$$A_0 = \begin{pmatrix} x & x & & & & \\ x & x & x & & & \\ & x & x & x & & \\ & & \dots & \dots & \dots & \\ & & & x & x & x \\ & & & & x & x \end{pmatrix}..$$

Определим матрицу вращения Q_1 так, чтобы матрица $A_1 = Q_1^T A_0 Q_1$ имела структуру

$$A_1 = \begin{pmatrix} x & x & & & & \\ x & x & x & & & \\ + & x & x & x & & \\ & & \dots & \dots & \dots & \\ & & & x & x & x \\ & & & & x & x \end{pmatrix}, \quad Q_1^T = \begin{pmatrix} c_1 & s_1 & & & & \\ -s_1 & c_1 & & & & \\ & & 1 & & & \\ & & \dots & \dots & \dots & \\ & & & & 1 & \\ & & & & & 1 \end{pmatrix}.$$

Знаком + обозначен появившейся в первом столбе матрицы выступ, который перемещается во второй столбец на следующем шаге:

$$A_2 = Q_2^T A_1 Q_2 ,$$

$$A_2 = \begin{pmatrix} x & x & & & & \\ x & x & x & & & \\ & x & x & x & & \\ & + & \dots & \dots & \dots & \\ & & & x & x & x \\ & & & & x & x \end{pmatrix}, \quad Q_2^T = \begin{pmatrix} 1 & & & & & \\ & c_2 & s_2 & & & \\ & -s_2 & c_2 & & & \\ & & \dots & \dots & \dots & \\ & & & & 1 & \\ & & & & & 1 \end{pmatrix}.$$

Продолжая процесс, получим после выполнения (n–1) шагов матрицу $A_{n-1} = Q^T A_0 Q$, имеющую трехдиагональный вид.

Здесь $Q = Q_1 Q_2 \dots Q_{n-1} = \begin{pmatrix} c_1 & x & x & x & x & x \\ s_1 & x & x & x & x & x \\ & s_2 & x & x & x & x \\ & & \dots & \dots & \dots & \dots \\ & & & s_{n-2} & x & x \\ & & & & s_{n-1} & x \end{pmatrix}$

Согласно известной Q-теореме (напр., [21]) первый столбец матрицы Q $\mathbf{q}_1 = (c_1, s_1, 0, \dots, 0)^T$ однозначно определяет остальные столбцы. Если выбрать $\mathbf{q}_1$, пропорциональным первому столбцу матрицы A_0, то матрица Q совпадет с ортогональной матрицей в QR-разложении матрицы A_0.

Таким образом, неявно выполнена QR-итерация $A^{(k+1)} = A_{n-1} = Q^{(k)T} A^{(k)} Q^{(k)}$. Принципиальным с точки зрения экономичности вычислений является то, что производимые в ходе выполнения итерации матричные умножения $\tilde{A}_l = Q_l^T A_{l-1}$, $A_l = \tilde{A}_{l-1} Q_l$ могут выполнены за фиксированное, не зависящее от размера матрицы, количество операций, что обусловлено структурой матриц Q и A. В итоге, количество операций, затрачиваемое на одну итерацию пропорционально размеру матрицы. Отметим, что при вычислении собственных значений нет необходимости явно вычислять матрицу Q на каждой итерации. Она становится нужной для вычисления собственных векторов, однако в настоящей

работе собственные векторы определялись методом обратных итераций, о чем будет говориться ниже.

Как и для традиционного QR-алгоритма, эффективным приемом повышения скорости сходимости итерационного процесса является применение сдвигов:

$$A^{(k)} - \sigma_k E = Q^{(k)} R^{(k)},$$

$$A^{(k+1)} = R^{(k)} Q^{(k)} + \sigma_k E,$$

где E – единичная матрица, σ – сдвиг.

В ходе неявного алгоритма сдвиг осуществляется путем выбора вектора $\mathbf{q}_1$, пропорциональным первому столбцу матрицы $A_0 - \sigma_k E$.

При этом в качестве значения сдвига σ_k выбирается то из собственных значений подматрицы $\begin{pmatrix} a_{n-1n-1}^{(k)} & a_{n-1n}^{(k)} \\ a_{nn-1}^{(k)} & a_{nn}^{k} \end{pmatrix}$, которое ближе нижнему угловому элементу $a_{nn}^{(k)}$ (стратегия Уилкинсона). Обоснование такого выбора можно найти, напр., в [21]).

Другим эффективным приемом, снижающим вычислительные затраты, является дефляция матрицы (понижение размерности) по мере сходимости нижнего углового элемента к соответствующему собственному значению.

В среднем на вычисление одного собственного значения требуется 2 итерации неявного QR-алгоритма. Учитывая, что на каждой итерации вычислительные затраты пропорциальны размеру матрицы *n*, суммарные вычислительные затраты составляют $O(n^2)$.

1.3.3. Метод бисекции

В задачах моделирования полупроводниковых гетероструктур часто не требуется вычислять весь спектр, а достаточно найти только собственные значения, лежащие в определенном диапазоне. Как правило, такой диапазон соответствует нижним энерге-

тическим уровням, на которых располагается большая часть носителей заряда. Одним из эффективных методов решения таких задач применительно к трехдиагональным матрицам является метод бисекции.

Метод основан на свойстве преобразования $S^T AS$ сохранять такие свойства матрицы, как число отрицательных, нулевых и положительных собственных значений (в совокупности эти три величины называются инерцией матрицы). Если несложно построить разложение матрицы $A - aE$ вида $A - aE = SDS^T$, где S – невырожденная, а D – диагональная матрица, для которой вычисление инерции не вызывает проблем, то легко определить число N_a^b собственных значений матрицы A, попадающих в интервал $[a,b)$:

$N_a^b = N_b - N_a$, где N_a – число собственных значений матрицы A, меньших a (или, что то же самое, число отрицательных собственных значений матрицы $A - aE$),

N_b – число собственных значений матрицы A, меньших b.

Далее, следуя методу бисекции, производится дробление интервала, в результате чего с необходимой точностью определяются все собственные значения матрицы A, лежащие в интервале $[a,b)$. Важным моментом в применении данного метода является возможность экономичного построения разложения $A - aE = SDS^T$.

Для трехдиагональной симметричной матрицы с ненулевыми элементами α_j, β_j, γ_j, ($\alpha_j = \gamma_{j-1}, j = 2,..,n$) такое разложение строится особенно просто, диагональные элементы матрицы D определяются следующим образом [21]:

$$d_j = (\beta_j - a) - \frac{\gamma_{j-1}}{d_{j-1}}, \; j = 2,...,n, \; d_1 = (\beta_1 - a).$$

Таким образом, вычислительные затраты на нахождение m собственных значений методом бисекции составляют $O(mn)$.

1.3.4. Метод Ланциоша

В более общем случае симметричных разреженных матриц большой размерности эффективным алгоритмом является алгоритм Ланцоша. Он объединяет метод Ланцоша построения ортонормированного базиса Крыловского подпространства $Q_{n\times m} = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m]$ и процедуру Рэлея–Ритца, согласно которой собственные значения матрицы $B_{m\times m} = Q^T_{m\times n} A_{n\times n} Q_{n\times m}$ интерпретируются как приближения к собственным значениям матрицы A.

Напомним, что Крыловское подпространство размерности m, порожденное матрицей A и вектором $\mathbf{q}$, строится как линейная оболочка векторов, получаемых с помощью последовательности матрично-векторных произведений, в виде $\operatorname{span}(\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m)$, $\mathbf{q}_1 = \mathbf{q}, \mathbf{q}_j = A\mathbf{q}_{j-1}, j = 2,..m$. Такая процедура аналогична известному степенному методу, с ростом m вектор $\mathbf{q}_m$ все более приближается к собственному вектору матрицы A, соответствующему максимальному по модулю собственному значению. Таким образом, векторы в базисе становятся почти параллельными. Чтобы избежать этой негативной тенденции, на каждом шаге производится ортогонализация вектора по отношению к векторам, полученным на предыдущих шагах: $\mathbf{q}_m = \mathbf{q}_m - (\mathbf{q}_j, \mathbf{q}_m)\mathbf{q}_j$, $j = 1,...,m-1$ (модифицированный алгоритм Грама–Шмидта). В результате алгоритм Ланцоша для вычисления собственных значений (и собственных векторов) симметричной матрицы A в точной арифметике записывается следующим образом [21, 22]:

Задаются начальный вектор $\mathbf{q}_1$, такой что $\|\mathbf{q}_1\|_2 = 1$, нулевой вектор $\mathbf{q}_0 = 0$, и число $\theta_1 = 0$.

1. Осуществляется цикл $j = 2,...,m$

$\mathbf{q} = A\mathbf{q}_{j-1}$

$\mu = \mathbf{q}^T_{j-1}\mathbf{q}$

$\mathbf{q} = \mathbf{q} - \mu\mathbf{q}_{j-1} - \theta_{j-1}\mathbf{q}_{j-2}$

$\theta_j = \|\mathbf{q}\|_2$

Если $\theta_j = 0$, то выполнение алгоритма прекращается

$\mathbf{q}_j = \mathbf{q} / \theta_j$.

Завершение цикла. В результате построена ортогональная матрица $Q_{n\times m} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m)$.

2. Вычисление собственных значений $\lambda_i(B)$ и собственных векторов $\mathbf{v}_i(B)$ матрицы $B_{m\times m} = Q^T_{m\times n} A_{n\times n} Q_{n\times m}$, $Q_{n\times m} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m)$ и соответствующих оценок погрешности ε_i:

$$\varepsilon_i = \theta_m \left| v_i^m \right|,$$

где v_i^m – последняя m-ая компонента собственного вектора $\mathbf{v}_i$, $i = 1, \ldots, m$.

Если погрешности искомых собственных значений достаточно малы, завершить алгоритм.

Если погрешности велики, то размерность m должна быть увеличена. При этом нет необходимости начинать выполнение алгоритма с начала, можно продолжать выполнение с текущей позиции.

Разумеется, возможна реализация алгоритма с вычислением собственных значений, собственных векторов и оценок погрешности на каждом шаге. Здесь принципиально, что матрица B является трехдиагональной и симметричной. Учитывая, что, как правило, ее размерность сравнительно невелика (решается частичная проблема собственных значений), решение спектральной задачи для нее довольно экономично (см. неявный QR-алгоритм и метод бисекции, рассмотренные выше). Другим важным моментом является особенность сходимости алгоритма Ланцоша, заключающаяся в том, что быстрее осуществляется сходимость к собственным значениям, прилежащим к краям спектра. Это соответствует логике решения рассматриваемых в данной работе задач, в которых, как уже отмечалось, основной интерес представляют нижние энергетические уровни. Благодаря тому, что вычисление ортонормированного базиса Крыловского под-

пространства удается свести к коротким рекурсиям, выполнение m шагов алгоритма Ланцоша требует всего $O(mn)$ операций (с учетом разреженности матрицы A). Однако в представленном виде алгоритм может оказаться неработоспособным, т.к. при его компьютерной реализации вследствие наличия погрешностей вычислений векторы $\mathbf{q}_j$ могут терять ортогональность, и становиться практически линейно зависимыми. С целью подавления этого эффекта на каждом шаге может применяться полная переортогонализация векторов:

$$\mathbf{q} = \mathbf{q} - \sum_{l=1}^{j-1} (\mathbf{q}_l, \mathbf{q}) \mathbf{q}_l ,$$

что, однако, существенно повышает вычислительные затраты – до $O(m^2 n)$ операций.

Разумной альтернативой является алгоритм Ланцоша с выборочной переортогонализацией [21], основанный на том факте, что векторы $\mathbf{q}_j$ теряют ортогональность вследствие приобретения больших компонент в направлениях векторов $Q_j \mathbf{v}_i$, соответствующих уже сошедшимся числам Ритца λ_i. При этом сходимость характеризуется вычисленной погрешностью $\varepsilon_i = \theta_m \left| v_i^m \right|$. Когда эта погрешность становится достаточно малой, целесообразно провести дополнительную ортогонализацию вектора $\mathbf{q}$ к векторам $\mathbf{g}_i = Q_j \mathbf{v}_i$: $\mathbf{q} = \mathbf{q} - (\mathbf{g}_i, \mathbf{q}) \mathbf{g}_i$, $i = 1,..,I$, где количество векторов $\mathbf{g}_i$ определяется условием малости ε_i. Ясно, что такой подход требует вычисления собственных значений, собственных векторов и оценок погрешности на каждом шаге.

1.3.5. Метод обратных итераций

Рассмотренные выше методы позволяют эффективно определять собственные значения в требуемом диапазоне. Для вычисления собственных векторов, соответствующих найденным собственным значениям, используется метод обратных итераций.

В этом методе на каждой итерации решается система уравнений $(A-\lambda E)\mathbf{y}^{(k)}=\mathbf{z}^{(k-1)}$, $\mathbf{z}^{(k-1)}=\mathbf{y}^{(k-1)}/\left\|\mathbf{y}^{(k-1)}\right\|_2$. Итерационный процесс стартует с произвольного вектора $\mathbf{y}^{(0)}$. Вектор $\mathbf{y}^{(k)}$ сходится к собственному вектору, соответствующему собственному значению λ. Обратим внимание на следующее обстоятельство. Найденное итерационным методом приближенное собственное значение λ очень близко к точному собственному значению λ^*, при котором определитель матрицы $(A-\lambda^* E)$ равен нулю. Таким образом, матрица $(A-\lambda E)$ оказывается плохо обусловленной. Однако, в данном случае плохая обусловленность способствует ускорению сходимости, так как вектор погрешности вычислений выстраивается вдоль искомого собственного вектора. В результате метод отличается высокой скоростью сходимости – как правило, для достижения сходимости достаточно нескольких итераций. Следует отметить высокий потенциал распараллеливания вычислений, присущий методу обратных итераций. Каждый собственный вектор можно вычислить независимо от других. Особо привлекательной является в этом плане комбинация метода бисекции и метода обратных итераций.

1.4. Метод сопряженных градиентов для решения уравнения Пуассона

Для нахождения поля электростатического потенциала и энергетических барьеров по заданному (как правило, определяемому из решения уравнения Шрёдингера) распределению носителей заряда в гетероструктурах используется уравнение Пуассона

$$\Delta\varphi = f(\mathbf{r})$$

с соответствующими граничными условиями.

Поскольку структура дифференциального оператора в уравнении Пуассона идентична структуре дифференциального оператора уравнения Шрёдингера, здесь применимы те же подходы к конечно-разностной аппрокимации. Они были подробно рас-

смотрены в предыдущем разделе. В результате задача сводится к решению системы линейных алгебраических уравнений с разреженной, как правило, ленточной матрицей большой размерности

$$A\mathbf{x} = \mathbf{b}\ .$$

Для решения таких задач в одномерной постановке (система с трехдиагональной матрицей) используется хорошо известный метод прогонки. В более общих случаях эффективным инструментом решения являются итерационные методы, использующие свойства Крыловских подпространств. Вопросы построения таких подпространств рассматривались выше, когда шла речь о методе Ланцоша для решения спектральной задачи.

В методах рассматриваемого класса система уравнений $A\mathbf{x} = \mathbf{b}$ решается построением ортогонального базиса крыловского подпространства $[\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m]$ и поиском приближенного решения задачи в этом подпространстве $\mathbf{x}^{(k)} = \sum_{j=1}^{k} c_j \mathbf{q}_j$.

Возможны различные способы определения коэффициентов разложения c_j [21].

1. Вектор $\mathbf{x}^{(k)}$ минимизирует квадратичную норму невязки $\left\|\mathbf{r}^{(k)}\right\|_2$, $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$. Такой выбор реализуется в известных алгоритмах MINRES (minimum residual) для симметричных матриц и GMRES (generalized minimum residual) для общего случая несимметричных матриц.

2. Вектор $\mathbf{x}^{(k)}$ обеспечивает ортогональность невязки Крыловскому подпространству $(\mathbf{r}^{(k)}, \mathbf{q}_j) = 0$, $j = 1,..,k$.

3. Вектор $\mathbf{x}^{(k)}$ минимизирует А-норму невязки $\left\|\mathbf{r}^{(k)}\right\|_{A^{-1}} = (\mathbf{r}^{(k)T} A^{-1} \mathbf{r}^{(k)})^{1/2}$, $\mathbf{r}^{(k)} = \mathbf{b} - A\mathbf{x}^{(k)}$. Такой способ реализуется в алгоритме сопряженных градиентов (*Conjugate Gradient, CG*) для симметричных положительно определенных матриц и его обобщении на случай несимметричных матриц (*Biconjugate Gradient, BiCG* – метод бисопряженных градиентов и его стабилизированный вариант *BiCGStab*) [23, 24].

Последний метод (CG) привлекателен возможностью построения алгоритма с помощью коротких рекурсий. Алгоритм сопряженных градиентов состоит из следующих операций.

Выбирается вектор начального приближения $\mathbf{x}^{(0)}$ и вычисляется вектор невязки $\mathbf{r}^{(0)} = \mathbf{b} - A\mathbf{x}^{(0)}$. Задается вектор $\mathbf{p}_0$: $\mathbf{p}_0 = \mathbf{r}^{(0)}$.

Осуществляется итерационный процесс $k = 1, \ldots$

$\mathbf{q}_k = A\mathbf{p}_{k-1}$,

$\nu_k = \left(\mathbf{r}^{(k-1)}, \mathbf{r}^{(k-1)}\right) \Big/ \left(\mathbf{q}_k, \mathbf{p}_{k-1}\right)$,

$\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \nu_k \mathbf{p}_{k-1}$ – уточняется приближенное решение,

$\mathbf{r}^{(k)} = \mathbf{r}^{(k-1)} - \nu_k \mathbf{q}_k$ – вычисляется невязка для нового решения,

$\mu_k = \left(\mathbf{r}^{(k)}, \mathbf{r}^{(k)}\right) \Big/ \left(\mathbf{r}^{(k-1)}, \mathbf{r}^{(k-1)}\right)$,

$\mathbf{p}_k = \mathbf{r}^{(k)} + \mu_k \mathbf{p}_{k-1}$ – вычисляется новое направление поиска.

Завершение итерационного процесса по достижении сходимости (достаточной малости $\left\|\mathbf{r}^{(k)}\right\|_2$).

Таким образом, в данном алгоритме на каждой итерации с помощью коротких рекурсий строятся новое направление поиска $\mathbf{p}_k$, новое приближенное решение $\mathbf{x}^{(k)}$ и соответствующая ему невязка $\mathbf{r}^{(k)}$. Векторы $\mathbf{p}_k$ являются A-сопряженными (ортогональными по отношению к скалярному произведению, определяемому матрицей A: $\mathbf{p}_i^{\mathrm{T}} A \mathbf{p}_j = 0, i \neq j$).

Скорость сходимости метода сопряженных градиентов зависит от распределения собственных значений матрицы A и, в частности, от ее числа обусловленности. Для задач, рассматриваемых в настоящей работе, характерны довольно высокие числа обусловленности, которые возрастают с увеличением размера матриц. Поэтому часто оказывается предпочтительным решать не исходную систему линейных алгебраических уравнений, а её предобусловленный вариант. Предобуславливание осуществляется путем перехода от системы $A\mathbf{x} = \mathbf{b}$ к системе $C^{-1}A\mathbf{x} = C^{-1}\mathbf{b}$, где

произведение матриц C^{-1} и A обладает лучшими спектральными свойствами. Одним из главных критериев оптимальности выбора матрицы C является относительная легкость её обращения. В качестве простых, но достаточно эффективных предобуславливателей, могут использоваться матрицы диагональной или трёхдиагональной ленточной структуры, образованные соответствующими элементами матрицы **A**.

1.5. Приближенная модель для экспресс-расчета концентрации носителей в гетероструктуре AlGaN/GaN

Для приближенного расчета волновых функций электронов в потенциальной яме в окрестности интерфейса полупроводниковой гетероструктуры можно использовать приближение Фанга-Ховарда [25], в котором волновая функция записывается в виде:

$$\psi(z) = \left(b^3/2\right)^{1/2} z \exp\left(-bz/2\right),$$

где b – нормировочная константа, а z – расстояние от барьера. Здесь принято, что в барьерном слое $z<0$, в буферном слое – $z>0$. Хорошей аппроксимацией пространственного распределения потенциала в окрестности интерфейса является треугольный профиль. Для треугольной потенциальной ямы константа b может быть записана в виде:

$$b = \left(\frac{33\pi m^* e^2 N_s}{2\hbar\varepsilon_s}\right)^{1/3},$$

где m^* – эффективная масса носителей заряда, N_s – концентрация электронов в ДЭГ.

Как видно из приведенных формул, приближенный расчет волновой функции требует знания концентрации электронов в двумерном электронном газе. Эта величина также может быть рассчитана приближенно. Рассмотрим вариант такого расчета применительно к гетерогранице между барьерным и узкозонным слоями (например, AlGaN и GaN) с ненулевой поверхностной плотностью заряда (вюрцитная структура) [2].

$$N_s = \frac{\sigma}{e} - \frac{E_F}{e^2}\left(\frac{\varepsilon_1}{d_1} + \frac{\varepsilon_2}{d_2}\right) - \frac{\varepsilon_1}{e^2 d_2}\left(e\Phi_1 + \Delta - \Delta E_1^C\right),$$

где σ – плотность связанных зарядов на интерфейсе (данная величина получена из первопринципных расчетов), ε_1, ε_2 – диэлектрические проницаемости слоев, d_1, d_2 – толщины слоев, E_F – положение уровня Ферми по отношению к краю зоны проводимости узкозонного слоя, $e\Phi_1$ – высота барьера Шоттки, ΔE_1^C – разрыв зон проводимости между слоями, Δ – проникновение края зоны проводимости ниже уровня Ферми на гетерогранице.

Последняя величина может быть приближенно определена следующим образом:

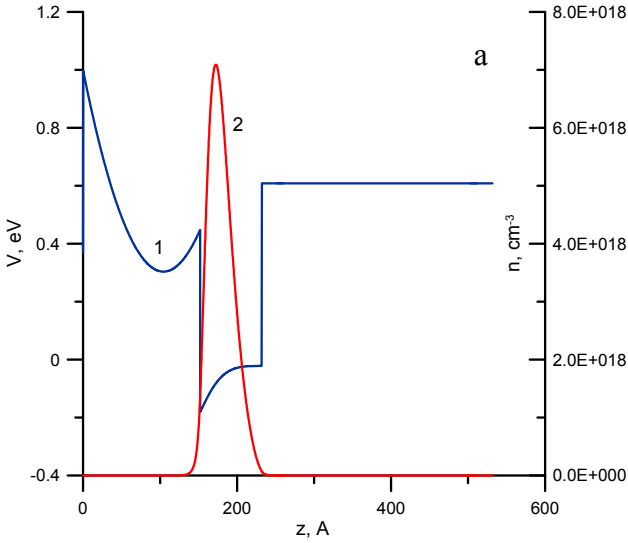
$\Delta = E_0 + \frac{\pi\eta^2}{m_2^*} N_s$, где E_0 – нижний уровень зоны проводимости,

$$E_0 = \left(\frac{9\pi\eta^2 e^2 N_s}{8\varepsilon_2\sqrt{8m_2^*}}\right)^{2/3}.$$

Два последних выражения и формула для концентрации электронов в ДЭГ образуют систему нелинейных уравнений, которая решается методом последовательных приближений.

1.6. Примеры решения задач

Сначала рассмотрим вариант, когда полупроводниковые материалы имеют кристаллическую структуру сфалерита. В качестве примера возьмем трехслойную гетероструктуру $Al_{0.3}GaN/GaN/Al_{0.3}GaN$ [26]. Толщины слоев следующие – барьерный слой $Al_{0.3}GaN$ – 15,2 нм, слой GaN – 8 нм, слой $Al_{0.3}GaN$ – 30 нм. Остальные исходные данные следующие [26]. Диэлектрические проницаемости материалов равны $8.9\,\varepsilon_0$ для GaN и $8.5\,\varepsilon_0$ для $Al_{0.3}GaN$, где ε_0 – электрическая постоянная. Эффективные массы электрона принимались равными $0.228\,m_0$ для обоих материалов (m_0 – масса покоя электрона). Сдвиг зоны проводимости ΔE_c полагался равным 0.63 эВ. Потенциал на внешней поверхности гетероструктуры определялся из условия $-e\phi(0) = 1$ эВ (отсчет энергии ведется от уровня Ферми).



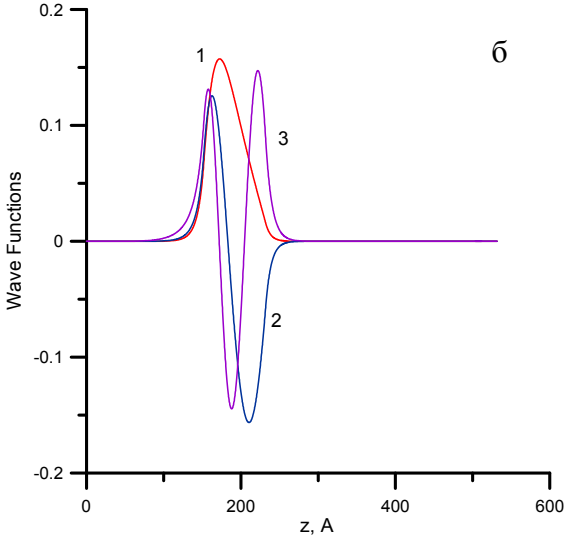


Рис. 1.2

а – распределение потенциальной энергии (1, левая шкала) и электронной плотности (2, правая шкала) по толщине гетероструктуры, б – волновые функции, соответствующие трем низшим энергетическим уровням. Вариант кристаллической структуры сфалерита

Как уже отмечалось, в случае сфалеритной кристаллической структуры создание надлежащих потенциальных барьеров осуществляется с использованием легирования барьерного слоя донорной примесью. В рассматриваемом варианте объемная концентрация примеси равна $6 \cdot 10^{18}$ см$^{-3}$. Здесь и далее в этом разделе задача решается в одномерной постановке.

На рис. 1.2 приведены распределения по толщине гетероструктуры потенциальной энергии (рис. 1.2а, кривая 1, отнесена к левой шкале), электронной плотности (рис. 1.2а, кривая 2, отнесена к правой шкале) и волновых функций (рис. 1.2б, нумерация кривых соответствует возрастанию энергетических уровней от самого нижнего). На рис. 1.2а отчетливо видны изгиб на кривой потенциальной энергии вследствие легирования барьерного слоя и потенциальная яма в слое GaN, что приводит к образованию локализованной зоны повышенной концентрации электронов (двумерный электронный газ). Суммарная концентрация электронов в ДЭГ равна $3 \cdot 10^{12}$ см$^{-2}$. Полученные результаты хорошо согласуются с данными, приведенными в статье [26].

На рис. 1.3 проиллюстрирована сходимость глобальных итераций для рассматриваемого варианта. Здесь кривая 1 соответствует методу последовательных приближений с нижней релаксацией, кривая 2 – локальной аппроксимации зависимости электронной плотности от потенциала (нижняя релаксация не требуется). В качестве погрешности рассматривалась разность между наименьшими энергетическими уровнями (минимальное собственное значение для уравнения Шрёдингера), полученными на смежных итерациях. Видно, что ускорение сходимости в последнем случае значительное. В сочетании с линеаризацией уравнения Пуассона данный подход позволяет, таким образом, существенно сократить вычислительные затраты. Отметим, что выбор в качестве величины, характеризующей погрешность, разности между наименьшими энергетическими уровнями обусловлен тем, что согласно многочисленным вычислительным экспериментам сходимость по наименьшим значениям энергии, практически гарантирует сходимость по другим возможным критериям, так как электроны, обладающие такой энергией, вносят решающий вклад в распре-

деление электронной плотности, что в конечном итоге отражается и на распределении электростатического потенциала.

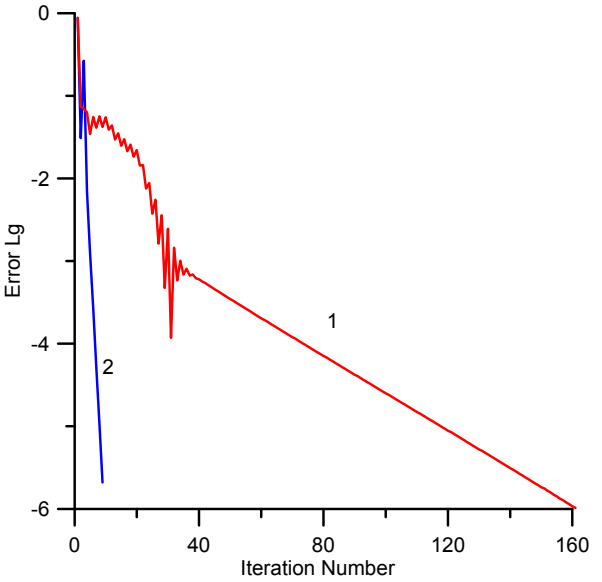


Рис. 1.3. Сходимость глобальных итераций:
1 – метод последовательной нижней релаксации, 2 – аппроксимация нелинейной зависимости электронной плотности от потенциала. Вариант кристаллической структуры сфалерита

Перейдем к полупроводниковым материалам с вюрцитной кристаллической структурой. Рассмотрим двухслойный вариант $Al_{0.3}GaN/GaN$ (толщина слоя $Al_{0.3}GaN$ 30 нм), хорошо изученный экспериментально и теоретически [2]. Здесь и далее исходные данные для расчетов определялись по аппроксимационным зависимостям [2]:

$$\varepsilon_{AlGaN}(x) = (0.03x + 10.28)\varepsilon_0 ,$$

$$\Delta E_c(x) = 0.7\left(E_g(x) - E_g(0)\right),$$

$$E_g(x) = 6.13x + 3.42(1 - x),$$

$$-e\varphi_b = 1.3x + 0.84 ,$$

где x – мольная доля алюминия в сплаве AlGaN. Использовалось граничное условие для потенциала $\varphi(0) = \varphi_b$.

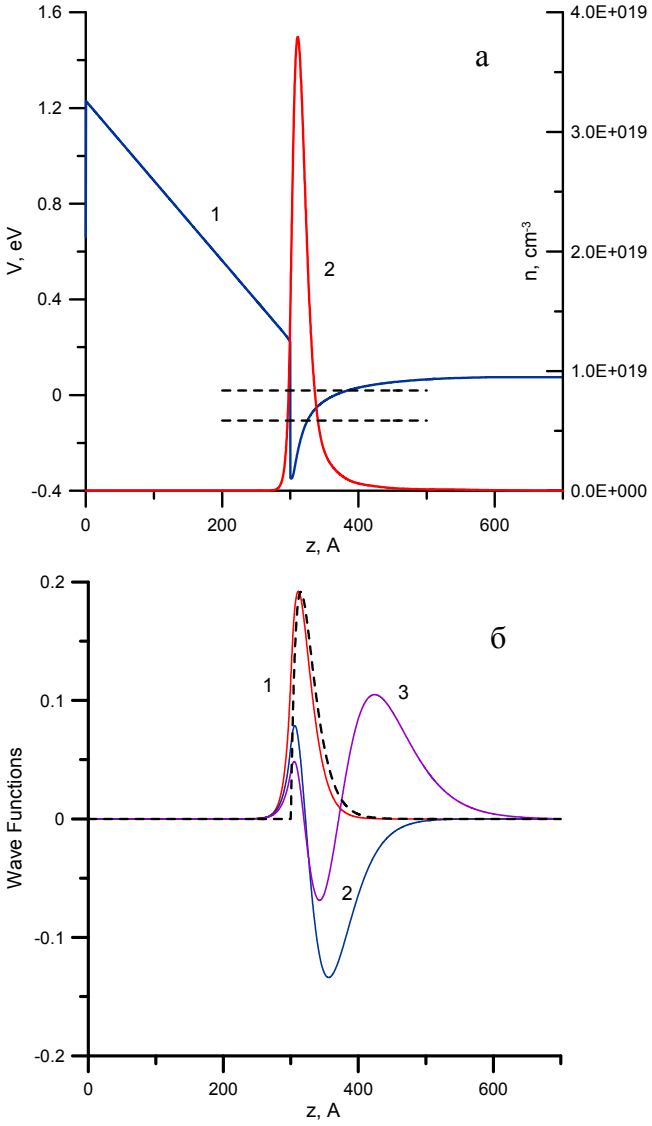


Рис. 1.4

а – распределение потенциальной энергии (1, левая шкала) и электронной плотности (2, правая шкала) по толщине гетероструктуры, б – волновые функции, соответствующие трем низшим энергетическим уровням. Вариант кристаллической структуры вюрцита

На рис. 1.4 представлены распределения потенциальной энергии и электронной плотности по толщине гетероструктуры, а также волновые функции. Нумерация кривых аналогична рис. 1.2. Здесь профиль потенциальной ямы и образование двумерного электронного газа обусловлены наличием поверхностного заряда на гетерогранице. Плотность заряда согласно проведенным ab initio квантово-механическим расчетам [6] равна $1.34 \cdot 10^{13}$ е Кл/см^2. Полученная в данном варианте суммарная концентрация электронов в ДЭГ равна $1.14 \cdot 10^{13}$ см$^{-2}$, что существенно выше, чем аналогичная величина для материалов с кристаллической структурой сфалерита.

Рис. 1.5 иллюстрирует сходимость глобальных итераций (нумерация кривых аналогична рис. 1.3).

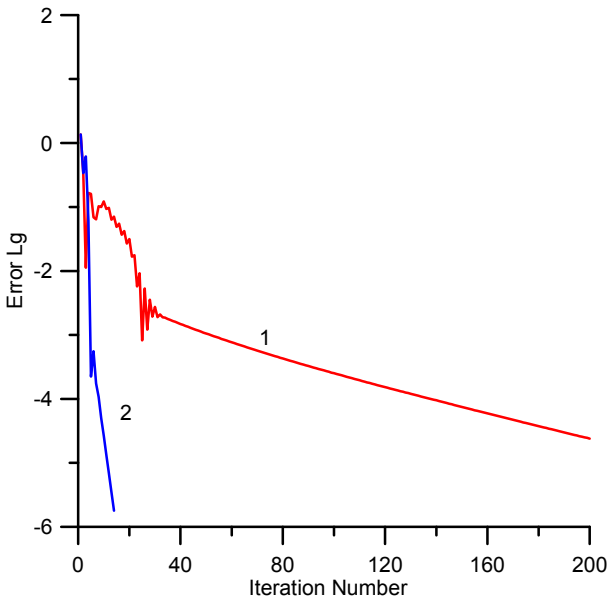


Рис. 1.5. Сходимость глобальных итераций:

1 – метод последовательной нижней релаксации, 2 – аппроксимация нелинейной зависимости электронной плотности от потенциала. Вариант кристаллической структуры вюрцита

Видно, что по сравнению с вариантом кристаллической структуры сфалерита сходимость последовательных приближений в данном случае замедляется. Использование аппроксимации нелинейной зависимости электронной плотности от потенциала позволяет существенно (в рассмотренном случае практически на порядок) ускорить сходимость глобальных итераций.

На рис. 1.6 представлены зависимости суммарной концентрации электронов в ДЭГ от содержания Al в слое Al_xGaN. Здесь величина концентрации отнесена к значению 10^{-13} $см^{-2}$. Сплошная кривая соответствует результатам численного моделирования, маркеры – аппроксимации экспериментальных данных [2]. Видно, что расчетные и экспериментальные результаты хорошо согласуются.

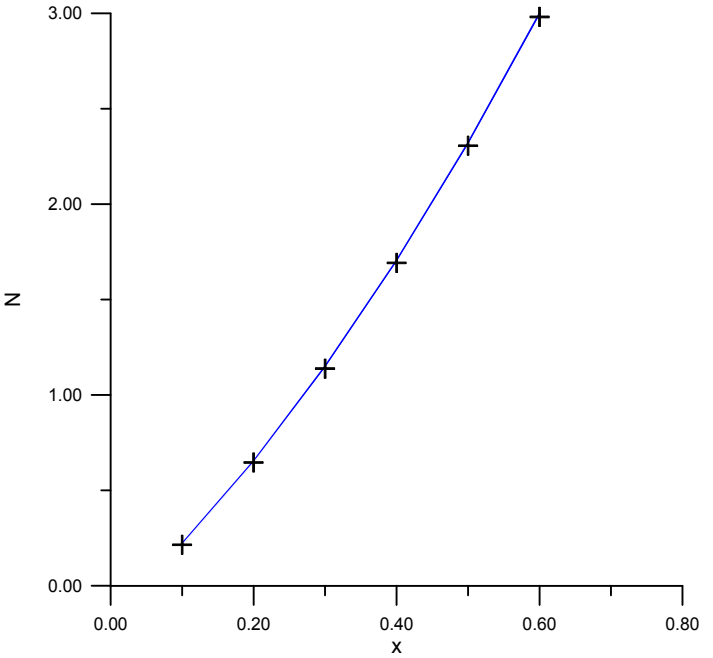


Рис.1.6. Зависимость суммарной концентрации электронов в двумерном электронном газе от концентрации Al в барьерном слое: сплошная кривая – численное решение, маркеры – аппроксимация экспериментальных данных [2]

В заключение, на рис. 1.7 приведены результаты расчета многослойной структуры, представленной на рис. 1.1. Нумерация кривых здесь аналогична рис. 1.2а, 1.4а. В данной структуре слой GaN на поверхности барьерного слоя предотвращает его окисление. Роль тонкого слой AlN можно проследить на рис.1.7. Видно небольшое различие в распределении концентрации носителей в окрестности гетероинтерфейса на рис. 1.4а и 1.7. Наличие прослойки AlN приводит к менее выраженному прониканию электронов в барьерный слой, что способствует более высокой подвижности электронов в ДЭГ. При этом различие в суммарных концентрациях электронов незначительно.

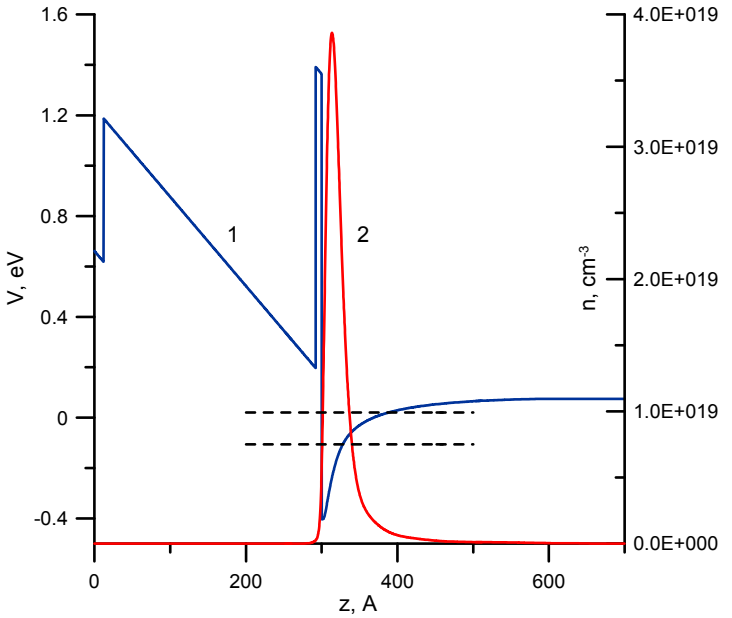


Рис. 1.7. Распределение потенциальной энергии (1, левая шкала) и электронной плотности (2, правая шкала) поперек многослойной гетероструктуры [11]

Таким образом, представленные методы и средства численного моделирования позволяют оперативно проводить многовариантный анализ многослойных наноразмерных гетероструктур с учетом поляризационных эффектов. В сочетании с возможностью моделирования подвижности носителей заряда в таких структурах (см. главу 2) это создает основу для решения ряда оптимизационных задач, актуальных для современной СВЧ-электроники. К таким задачам можно отнести синтез гетероструктур, обеспечивающих максимальную концентрацию при максимальной подвижности носителей (максимальную проводимость), обратные задачи по определению недоступных в прямом измерении характеристик выращенных гетероструктур на основе экспериментальных данных по концентрации и подвижности носителей в двумерном электронном газе.

Глава 2

МОДЕЛИРОВАНИЕ ПОДВИЖНОСТИ ЭЛЕКТРОНОВ В ПОЛУПРОВОДНИКОВЫХ ГЕТЕРОСТРУКТУРАХ

В предыдущей главе были рассмотрены вопросы численного моделирования распределения носителей заряда в наноразмерных полупроводниковых гетероструктурах. Благодаря наличию потенциальных барьеров, ограничивается движение электронов в поперечном (нормальном к гетероинтерфейсам) направлении и формируется локализованная в окрестности интерфейса область повышенной электронной плотности – двумерный электронный газ (ДЭГ). Информация о концентрации электронов в ДЭГ, уровнях энергии и волновых функциях, особенностях гетероинтерфейса используется для моделирования переноса носителей заряда в плоскости, параллельной слоям гетероструктуры.

Динамика электронов при этом определяется, с одной стороны, электрическим полем, создаваемым внешним воздействием со стороны электродов полупроводникового прибора (традиционно, это источник, сток и управляющий затвор), а также неоднородностью распределения самих носителей заряда, а с другой стороны, рассеянием электронов вследствие взаимодействия с колебаниями решетки, примесными центрами, дефектами решетки и др.

Применительно к рассматриваемым в настоящей работе гетероструктурам принципиальными являются следующие механизмы рассеяния: рассеяние на акустических и полярных оптических фононах (колебаниях решетки), рассеяние на неоднородностях сплава, на дислокациях и на шероховатостях гетерограницы, пьезоэлектрическое рассеяние, рассеяние на заряженных центрах,

локализованных в барьерном слое. При сильных электрических полях существенную роль играет также междолинное рассеяние, определяемое спецификой зонной структуры полупроводниковых материалов. В результате рассеяния меняются направление движения (волнового вектора), а в случае неупругого взаимодействия и энергия (модуль волнового вектора) электрона.

2.1. Модели динамики электронов в полупроводниках

При моделировании динамики двумерного электронного газа электроны обычно рассматриваются как классические частицы. Такое приближение справедливо, если длина свободного пробега электрона значительно меньше размера структуры в продольном направлении. В противном случае перенос носителей заряда может протекать без их рассеяния (баллистический транспорт) и могут стать существенными интерференционные эффекты [1, 27].

Наиболее полное описание динамики электронов в квазиклассическом приближении возможно с использованием кинетического уравнения Больцмана [напр., 12, 28]:

$$\frac{\partial f}{\partial t} + \mathbf{v}\nabla_r f + \frac{1}{\hbar}\mathbf{F}\nabla_k f = \frac{1}{4\pi^3}\int_{\Omega_B} W(\mathbf{k}',\mathbf{k})\big(f(\mathbf{k}') - f(\mathbf{k})\big)d\omega .$$

Здесь $f(\mathbf{r},\mathbf{k},t)$ – зависящая от времени функция распределения электронов по координатам и импульсам ($\mathbf{p} = \hbar\mathbf{k}$), $\mathbf{v}$ – скорость, $\mathbf{F}(\mathbf{r},t)$ – внешняя сила, действующая на электроны, $W(\mathbf{k}',\mathbf{k})$ – вероятность перехода из состояния $\mathbf{k}$ в состояние $\mathbf{k}'$ (предполагается, что имеет место кратковременное взаимодействие, которое не вызывает изменения координат электрона), интеграл столкновений берется по объему зоны Бриллюэна.

В общем случае уравнение Больцмана должно решаться совместно с уравнениями Шрёдингера и Пуассона (см. предыдущую главу), дающими распределение потенциала (и, следовательно, силы $\mathbf{F} = -\nabla_r\varphi$) в системе. Однако, часто при расчете динамики электронов квантованием их энергетического спектра в попереч-

ном направлении пренебрегают, в результате чего задача сводится к совместному решению уравнений Больцмана и Пуассона.

Одним из мощных инструментов решения уравнения Больцмана является метод Монте-Карло.

Согласно этому методу рассматривается движение и рассеяние отдельных электронов с последующим осреднением параметров по ансамблю и вычислением функции распределения. С точки зрения согласования динамики электронов и электрического поля, описываемого уравнением Пуассона, такой подход соответствует комбинации частица – сетка [29].

Описание движения электронов в полупроводнике осуществляется на основе информации о зонной структуре, определяющей соотношение между энергией электрона и волновым вектором $E(\mathbf{k})$. Эта информация может быть получена с помощью первопринципных квантово-механических расчетов (моделирование на атомарном уровне).

Уравнения движения электрона имеют следующий вид [29]:

$$\frac{d\mathbf{r}}{dt} = \frac{1}{\hbar}\nabla_k E,$$

$$\frac{d\mathbf{k}}{dt} = \frac{1}{\hbar}\mathbf{F}.$$

Первое уравнение отражает тот факт, что электрон движется с групповой скоростью волнового пакета, второе уравнение выражает закон изменения импульса.

Данные уравнения могут быть преобразованы к виду, традиционному для описания динамики классических частиц:

$$\frac{d\mathbf{r}}{dt} = \mathbf{v},$$

$$\frac{d\mathbf{v}}{dt} = M^{*-1}\mathbf{F},$$

где M^* – эффективная масса электрона, которая представляется тензором второго ранга:

$$m_{ij}^* = \hbar^2\left(\frac{\partial^2 E}{\partial k_i \partial k_j}\right).$$

Из последнего выражения видно, что эффективная масса электрона пропорциональна кривизне энергетической поверхности в k-пространстве и отражает квантово-механическое воздействие полупроводящей среды на движение электрона.

Как правило, большая часть электронов располагается около минимальных уровней энергии в каждой энергетической зоне. В таких случаях функция $E(\mathbf{k})$ может быть приближена квадратичной. Тогда эффективная масса электрона становится постоянной величиной:

$$m_{ij}^* = \frac{\hbar^2}{2C}, \quad E(\mathbf{k}) = C(k_1^2 + k_2^2 + k_3^2).$$

В последнем случае энергия электрона может рассматриваться как классическая кинетическая энергия: $E(\mathbf{k}) = \frac{p^2}{2m^*} = \frac{m^* v^2}{2}$.

Вдали от минимума энергетической зоны параболическое приближение теряет точность, однако возможно использование следующих аппроксимаций, учитывающих эффект непараболичности [29]:

$$\frac{p^2}{2m^*} = \frac{\hbar^2 k^2}{2m^*} = E(1 + \alpha E), \quad \alpha = \frac{1}{E_g}\left(1 + \frac{m^*}{m_e}\right)^2,$$

где E_g – расстояние между валентной зоной и зоной проводимости в энергетическом пространстве (ширина запрещенной зоны). При этом скорость электрона связана с волновым вектором следующим образом:

$$\mathbf{v} = \frac{\hbar\, \mathbf{k}}{m^*(1 + 2\alpha E)}.$$

Движение частицы периодически прерывается элементарными актами рассеяния, в результате которых мгновенно меняется направление движения (направление волнового вектора) и, при неупругих взаимодействиях, энергия (модуль волнового вектора) электрона.

Каждый механизм рассеяния характеризуется вероятностью рассеяния, зависящей от энергии электрона, и правилом изменения волнового вектора в результате рассеяния.

Логику метода Монте-Карло можно описать следующим образом. Зная энергию электрона в текущий момент времени и полную вероятность рассеяния, можно определить время свободного пробега электрона, динамика которого описывается приведенными выше уравнениями. Далее на основе соответствующих вероятностей выбирается конкретный механизм рассеяния и разыгрывается элементарный акт рассеяния – вычисляется новое значение энергии, модуль нового волнового вектора и, исходя из индикатрисы рассеяния, выбирается новое направление волнового вектора.

Интенсивность рассеяния выражается следующим образом [12, 30]:

$$W(\mathbf{k}) = \frac{2\pi}{\hbar}\sum_{k'}(\mathbf{k}', |V_s(\mathbf{r})|\mathbf{k})^2 \delta(E(k') - E(k) \mp \hbar w),$$

где V_s – рассеивающий потенциал, дельта-функция выражает закон сохранения энергии, $\hbar w$ – поглощенная (верхний знак) или выделенная (нижний знак) в процессе рассеяния энергия.

Конкретные выражения для интенсивности различных видов рассеяния электронов в гетероструктурах будут приведены ниже.

Подробное описание метода Монте-Карло и результатов, полученных с его помощью, выходит за рамки настоящей работы. Авторы планируют посвятить этому вопросу специальное издание.

Наряду с полным описанием динамики носителей заряда на основе уравнения Больцмана в практике математического моделирования используются и более простые («гидродинамические») модели, вытекающие из общего кинетического уравнения.

Наиболее простым подходом является использование уравнения неразрывности носителей заряда:

$$\frac{\partial n}{\partial t} + \nabla Q = 0,$$

Здесь n – концентрация электронов, Q – поток электронов, состоящий из дрейфовой и диффузионной составляющих:

$$Q = n\mathbf{v} - D\nabla n.$$

Скорость электронов связана с потенциалом электрического поля следующим образом:

$$\mathbf{v} = \mu\nabla\varphi ,$$

где μ – подвижность электронов, при определении которой интегрально учитываются различные виды рассеяния, ее ограничивающие. Для вычисления коэффициента диффузии D обычно используется соотношение Эйнштейна: $D = \mu\frac{k_B T}{e}$, где, как и ранее k_B – постоянная Больцмана, T – температура, e – заряд электрона.

Уравнение неразрывности дополняется уравнением Пуассона для распределения электростатического потенциала $\varphi(\mathbf{r})$:

$$\nabla\cdot(\varepsilon\nabla\varphi) = -e\left(N_D(\mathbf{r}) - N_A(\mathbf{r}) - n(\mathbf{r})\right),$$

где N_d, N_A – концентрации донорной и акцепторной примесей, ε – диэлектрическая постоянная материала.

В слабых электрических полях подвижность практически постоянна, дрейфовая скорость электронов пропорциональна напряженности электрического поля $\nabla\varphi$ (закон Ома). При усилении электрического поля энергия электронов возрастает и становится существенным эффект рассеяния электронов из центральной в сателлитные долины, где эффективная масса электрона существенно выше. В результате подвижность значительно уменьшается.

Самосогласованное решение уравнений неразрывности и Пуассона описывает динамику электронов в системе. Недостатком этой модели является пренебрежение временами релаксации импульса и энергии (считается, что скорость подстраивается к

электрическому полю мгновенно $\mathbf{v} = \mu\nabla\varphi$). Такое ограничение может быть существенным при моделировании наноразмерных структур, когда время релаксации становится соизмеримым с характерными временами динамики электрона в системе. С целью учета этих эффектов применяются модифицированные гидродинамические модели, включающие, наряду с уравнениями неразрывности и Пуассона, уравнения переноса импульса и энергии хаотического движения (температуры) электронов [29]:

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\mathbf{v} = \frac{e}{m^*}\nabla\varphi - \frac{\mathbf{v}}{\tau_v},$$

$$\frac{\partial w}{\partial t} + \mathbf{v}\cdot\nabla w = e\mathbf{v}\nabla\varphi - \frac{w - w_s}{\tau_e}.$$

Здесь τ_v, τ_e – времена релаксации импульса и энергии, w_s – тепловая энергия решетки $w_s = \frac{3}{2}k_B T$.

«Гидродинамические» уравнения часто записывают для каждой из неэквивалентных долин, различающихся эффективной массой электрона. Применительно к рассматриваемым гетероструктурам ряд важных результатов был получен с использованием квазидвумерной модели [31], которая включает одномерные (вдоль доминирующего продольного направления) «гидродинамические» уравнения для узкозонного и широкозонного материалов с источниковыми слагаемыми, описывающими обмен между слоями. Такой подход позволяет достаточно экономично учитывать поперечный пространственный перенос носителей заряда в гетероструктурах.

Численные методы решения уравнений переноса рассмотрены ниже. Здесь остановимся на одном из ключевых вопросов моделирования динамики носителей заряда в полупроводниковых гетероструктурах – определении коэффициента подвижности. Будем рассматривать наиболее перспективные на сегодняшний день структуры AlGaN/GaN.

2.2. Расчет подвижности электронов в гетроструктурах

Как уже отмечалось, при расчете подвижности необходимо учитывать различные механизмы рассеяния, такие как рассеяние на акустических и полярных оптических фононах (колебаниях решетки), рассеяние на неоднородностях сплава, на дислокациях и на шероховатостях гетерограницы, пьезоэлектрическое рассеяние, рассеяние на заряженных центрах, локализованных в барьерном слое.

Наиболее простой подход основан на использовании понятия времени релаксации, т.е. времени установления равновесного состояния после прекращения действия внешних полей. Так как установление равновесного состояния происходит за счет столкновений и при этом достаточно нескольких соударений, то время релаксации примерно соответствует времени свободного пробега электрона.

Суммарное время релаксации τ, определяющее подвижность электронов, определяется по правилу Матиссена:

$$1/\tau = \sum_i 1/\tau_l \ ,$$

где τ_l – время релаксации по импульсу для l-го механизма рассеяния. Подвижность электронов выражается через время релаксации следующим образом:

$$\mu = e\tau/m^* \ .$$

Обоснование такого подхода дается, например, в [11]. В гетероструктурах AlGaN/GaN концентрация эектронов в двумерном электронном газе достигает значений порядка $N_s = 10^{13}$ см$^{-2}$ (см. предыдущий раздел). ДЭГ является, таким образом, вырожденным вплоть до комнатных температур, так как выполняется условие $N_s >> m^* k_b T/\pi\hbar^2$. Для таких концентраций ДЭГ энергия Ферми значительно превышает величину тепловой энергии $k_b T$. Так как в вырожденном электронном газе рассеиваются только те электроны, энергия которых отличается от энергии Ферми не более

чем на k_bT, то можно считать, что энергия рассеивающихся электронов одинакова и время релаксации по импульсу не зависит от энергии. Это обусловливает справедливость приведенных выражений для подвижности.

В приближенных моделях данного класса используется следующая информация, поступающая из расчета распределения электронов в гетероструктуре: концентрация электронов в ДЭГ и параметр b, входящий в аппроксимационное выражение для волновой функции $\psi(z)=\left(b^3/2\right)^{1/2} z\exp(-bz/2)$ (функция Фанга–Ховарда, см. гл. 1, $z\geq 0$, здесь и далее в этом разделе $z=0$ соответствует границе барьерного и активного слоев). Для его определения можно использовать формулу $b=\left(\dfrac{33\pi m^* e^2 N_s}{2\hbar\varepsilon_s}\right)^{1/3}$.

Самосогласованное решение уравнений Шрёдингера и Пуассона (см. предыдущую главу) показало, что функция Фанга–Ховарда очень хорошо согласуется с вычисленной главной волновой функцией (при условии использования в выражении для параметра b вычисленной концентрации N_s). Это свидетельствует об обоснованности применяемого подхода.

Выражения для определения времени релаксации, соответствующего каждому из механизма рассеяния (в системе СГС) приведены ниже согласно работе [11].

Рассеяние на акустических фононах

Время релаксации электронов при рассеянии на акустических фононах вычисляется на основе теории деформационного потенциала:

$$\frac{1}{\tau_{DP}}=\frac{3m^* a_C^2 b k_b T}{16\pi\hbar^3 c_L}\int_0^{\pi} S(q)^2\left(1-\cos\theta\right)d\theta\,,$$

где a_c – константа потенциала деформации, c_L – продольная постоянная упругости, $q=2k_F\sin\left(\theta/2\right)$ – изменение волнового век-

тора электрона при рассеянии, $k_F = \sqrt{2\pi N_s}$ – модуль вектора Ферми, $S(q)$ – экранирующий множитель:

$$S(q) = \frac{q}{q + q_s \dfrac{b\left(8b^2 + 9bq + 3q^2\right)}{8\left(b+q\right)^3}}, \qquad q_s = 2m^* e^2 / \varepsilon_s \hbar^2 .$$

Рассеяние на полярных оптических фононах

Интенсивность рассеяния электронов на полярных оптических фононах существенно зависит от температуры. При низких температурах вкладом этого механизма рассеяния можно пренебречь, тогда как при комнатной температуре он является доминирующим и в значительной мере определяет итоговую подвижность электронов.

Рассеяние на оптических фононах не является упругим, поэтому приближение времени релаксации в этом случае является несправедливым. Однако, если энергия оптических фононов $\hbar\omega_0$ значительно превышает тепловую энергию $k_b T$ (для GaN $\hbar\omega_0$ = = 90.5 мэВ), то в качестве времени релаксации для вычисления подвижности можно брать характерное время поглощения фонона:

$$\frac{1}{\tau_{PO}} = \frac{2\pi e^2 \omega_0 m^* N G\left(k_0\right)}{\varepsilon^* k_0 \hbar^2} \cdot \frac{1}{1 + \dfrac{1 - \exp\left(-y\right)}{y}},$$

где $N = \left(\exp\left(\dfrac{\hbar\omega_0}{k_B T}\right) - 1\right)^{-1}$ – функция распределения фононов,

$k_0 = \sqrt{2m^* \omega_0 / \hbar}$ – модуль волнового вектора,

$G\left(k_0\right) = \dfrac{b\left(8b^2 + 9k_0 b + 3k_0^2\right)}{8\left(k_0 + b\right)^3}$ – форм-фактор,

$y = \pi\hbar^2 N_s / m^* k_B T$, $1/\varepsilon^* = 1/\varepsilon_\infty - 1/\varepsilon_0$ – приведенная диэлектрическая проницаемость,

ε_∞ и ε_0 – высокочастотная и статическая диэлектрические проницаемости, соответственно.

Пьезоэлектрическое рассеяние

В полупроводниках, обладающих пьезоэлектрическими свойствами, продольные акустические фононы приводят к появлению пьезоэлектрического потенциала, который рассеивает носители заряда. Выражение для времени релаксации при пьезоэлектрическом рассеянии имеет вид [11]:

$$\frac{1}{\tau_{PE}} = \frac{1}{\tau_L} + \frac{1}{\tau_T},$$

где $\frac{1}{\tau_{L,T}} = \frac{k_B T \alpha_{L,T}}{\pi \hbar k^2} \int_0^{\pi} qS(q)^2 f_{L,T}(q) d\theta$,

$$\alpha_L = (eh_{14})^2 \frac{m^*}{4\hbar^2 c_L} \frac{9}{32}, \quad \alpha_T = (eh_{14})^2 \frac{m^*}{4\hbar^2 c_T} \frac{13}{64},$$

h_{14} – компонента пьезоэлектрического тензора,

c_L – поперечная компонента тензора упругости,

$f_L(q)$, $f_T(q)$ – безразмерные форм-факторы для продольных и поперечных фононов:

$$f_T(q) = \frac{13 + 78(q/b) + 72(q/b)^2 + 82(q/b)^3 + 36(q/b)^4 + 6(q/b)^5}{13(1+(q/b))6}.$$

$$f_L(q) = \frac{1 + 6(q/b) + 12(q/b)^2 + 2(q/b)^3}{(1+(q/b))^6}$$

Рассеяние на удаленных заряженных центрах

Как отмечалось, для гетероструктур рассматриваемого класса характерно отделение области повышенной концентрации электронов (двумерного газа) от области легирования. Это позволяет избежать прямого рассеяния на заряженных центрах и способствует повышению подвижности. Однако, в случае легирования барьерного слоя возможно рассеяние электронов на ионизированных донорах в барьере. Время релаксации при рассеянии на удаленных заряженных центрах описывается выражением:

$$\frac{1}{\tau_{imp}} = \frac{\pi\hbar}{2m^*}\int_0^{\pi}(1-\cos\theta)\left(\frac{q_s}{q}\right)^2 S^2(q)\int F^2(q,z)N(z)\,dz\,d\theta\,,$$

где $N(z)$ – описывает распределение легирующей примеси (для заряда на гетерогранице $N(z)=N_0\delta(0)$), $F(q,z)$ – форм-фактор.

Выражение для форм-фактора при $z\leq 0$ (барьерный слой) имеет вид:

$$F(q,z) = \frac{b^3}{(b+q)^3}\exp(-q|z|)\,.$$

Для $z>0$ возможны два варианта:

$$F(q,z) = \frac{b^3}{(b-q)^3}\left[\exp(-qz) - \left(a_0 + a_1 z + a_2 z^2\right)\exp(-bz)\right],\ q \neq b$$

$$F(q,z) = \frac{1}{8}\left[1 + 2bz + 2b^2z^2 + \frac{4}{3}b^3z^3\right]\exp(-bz)\,,\ q = b$$

где $a_0 = \dfrac{2q\left(3b^2+q^2\right)}{(b+q)^3}$, $a_1 = \dfrac{4bq(b-q)}{(b+q)^2}$, $a_2 = \dfrac{q(b-q)^2}{b+q}$.

Рассеяние на дислокациях

Вследствие значительного рассогласования постоянных решеток AlGaN и GaN с подложкой (сапфиром), плотность дислокаций N_{disl}, проходящих через ДЭГ, может достигать величины 10^{10} см$^{-2}$. Время релаксации при рассеянии на дислокациях описывается выражением:

$$\frac{1}{\tau_{disl}} = \frac{2\pi m^* e^2 \rho_L^2 N_{disl}}{\hbar^3 \varepsilon_s^2 k_F^2}\cdot\int_0^{\pi}\frac{d\theta}{(q+q_s)^2}\,,$$

где k_F – модуль волнового вектора на поверхности Ферми и ρ_L – линейная плотность заряда дислокации. Если все акцепторные состояния на дислокации заполнены, то $\rho_L = e/c_0$, где c_0 – постоянная решетки GaN в направлении (0001).

Рассеяние на неоднородностях сплава

В твердом растворе AlGaN вследствие случайного характера распределения бинарных компонент AlN и GaN мольный состав x может случайно меняться по образцу. Так как ширина запрещенной зоны E_g зависит от мольной концентрации алюминия x, то появляется случайная модуляция E_g, что приводит к рассеянию носителей заряда. В исследуемых гетероструктурах рассеиваются только те электроны, которые проникают в AlGaN вследствие конечной высоты потенциального барьера. Аппроксимация волновой функции формулой Фанга-Ховарда в этом случае неприменима, так как она получена для бесконечно высокого потенциального барьера. Поэтому для расчета времени релаксации отдельно находится часть волновой функции электронов $\chi'(z)$, проникающая в барьер. Время релаксации находится из выражения:

$$\frac{1}{\tau_{all}} = \frac{m^* x(1-x)\Omega(\Delta E_c)^2}{\hbar^3} \int_{-\infty}^{0} \left|\chi'(z)\right|^4 dz ,$$

где ΔE_c – потенциал рассеяния, равный разности запрещенных зон AlN и GaN, Ω – объем элементарной ячейки, x – мольная доля Al в барьере. Часть волновой функции, приходящаяся на барьерный слой AlGaN может быть приближенно описана следующим выражением:

$$\chi'(z)^2 = \frac{2\pi e^2 N_s}{\varepsilon_s \Delta E_c(x)} \exp\left(\sqrt{8m^* \Delta E_c(x)}\frac{z}{\hbar}\right).$$

Рассеяние на шероховатостях гетерограницы

Наличие шероховатости гетерограницы приводит к изменению ширины квантовой ямы, и, следовательно, к изменению положения энергетических уровней. Данная флуктуация потенциала приводит к рассеянию носителей заряда. Вклад этого механизма рассеяния сильно зависит от технологического совер-

шенства гетероструктур. Время релаксации при рассеянии на шероховатостях гетерограницы определяется выражением:

$$\frac{1}{\tau_{IR}} = \left(\frac{\pi e^2 N_s \Delta\Lambda}{\varepsilon_s} \right)^2 \frac{m^*}{2\hbar^3} \int_0^{\pi} \exp\left(-q^2\Lambda^2/4\right) S(q)^2 \left(1-\cos\theta\right) d\theta ,$$

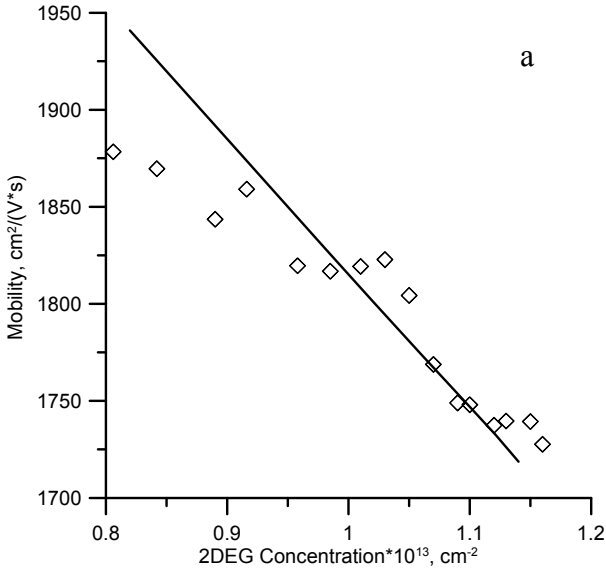
где Δ – размер шероховатости, Λ – корреляционная длина между неровностями.

Ниже на рис. 2.1, 2.2 приведены результаты сравнения расчетных и экспериментальных данных для гетероструктуры, приведенной на рис. 1.1. Используемые в расчетах параметры материала $Al_xGa_{1-x}N$ задавались согласно [11]. Они приведены в табл. 2.1.

Таблица 2.1

Параметры, использованные в расчетах концентрации и подвижности

Высокочастотная диэлектрическая проницаемость	$\varepsilon_\infty = 5.47$
Низкочастотная диэлектрическая проницаемость	$\varepsilon_s = 10.4$
Энергия продольных оптических фононов, мэВ	$\hbar\omega_0 = 90.5$
Плотность, г/см	$\rho = 6.1$
Скорость акустических фононов, см/с	$u = 6.59 \times 10^5$
Пьезоэлектрическая постоянная, SGS_V/см	$h_{14} = 5 \times 10^7/300$
Деформационный потенциал, эВ	$a_C = 9.2$
Постоянная упругости, Дин/см^2	$c_L = 2.66 \times 10^{12}$, $c_T = 6.2 \times 10^{11}$
Объем единичной ячейки, см^3	$\Omega = 2.28 \times 10^{-23}$
Ширина запрещенной зоны, эВ	$E_g(x) = x6.13 + (1-x)3.42 - x(1-x)$
Эффективная масса электрона	$m^* = 0.22 m_0$



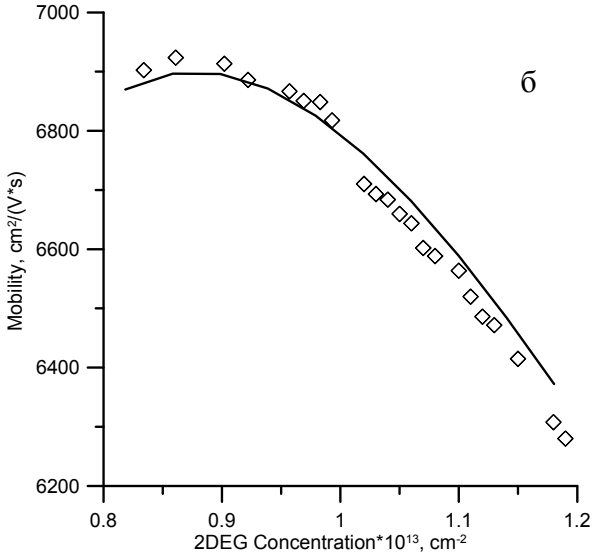


Рис. 2.1 Зависимость подвижности от концентарции:
(а) – при температуре 300 К, (б) – при температуре 77 К

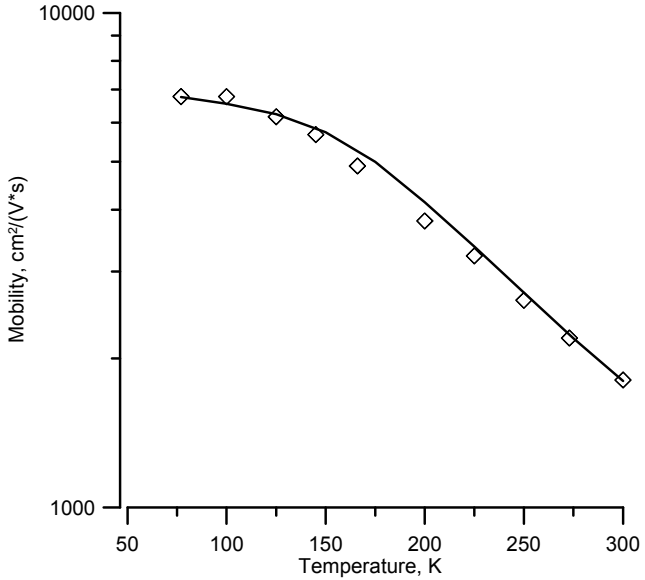


Рис. 2.2. Зависимость подвижности от температуры

На рисунках представлены зависимости подвижности от концентрации носителей в ДЭГ при различных температурах (рис. 2.1 а, б) и зависимость подвижности от температуры при концентрации 10^{13} см$^{-2}$ (рис. 2.2). Ясно, что здесь необходимо применение подгоночных процедур, поскольку ряд параметров (например, концентрация дислокаций, данные по шероховатости интерфейса) не поддаются прямому измерению. Однако применение фитинга в одной точке (при одной температуре и концентрации) позволяет достаточно точно воспроизвести экспериментальные зависимости в исследуемом диапазоне температур и концентраций.

2.3. Методы решения уравнения переноса (дрейфа–диффузии) носителей заряда

В данном разделе рассмотрим вопросы численной реализации математической модели переноса носителей заряда в полупроводниковых структурах. Как было показано в предыдущем разде-

ле, уравнение переноса может быть записано в виде обобщенного уравнения:

$$\frac{\partial \varphi}{\partial t} + div\mathbf{q} = f(\varphi, \mathbf{r}, t),$$

где φ – искомая величина (в нашем случае – концентрация носителей заряда), $\mathbf{q}$, f – плотность потока этой величины и плотность объемного источника в данной точке пространства.

В одномерном приближении уравнение принимает вид:

$$\frac{\partial \varphi}{\partial t} + \frac{\partial q}{\partial x} = f(\varphi, x, t).$$

Начнем рассмотрение с простейшего уравнения дрейфа носителей заряда. В этом случае $q = U\varphi$, $f \equiv 0$ и

$$\frac{\partial \varphi}{\partial t} + U\frac{\partial \varphi}{\partial x} = 0.$$

Это уравнение имеет аналитическое решение в виде бегущей волны:

$$\varphi(x,t) = \varphi^0(x - Ut),$$

где $\varphi^0(x) = \varphi(x,0)$ – начальное распределение величины.

В этой связи важно, чтобы численный метод максимально точно воспроизводил такой перенос, не привнося существенных искажений в форму бегущей волны.

Рассмотрим простейшую разностную схему решения уравнения переноса.

Вводя разностную сетку $\omega_{h\tau} : \{x_j = jh,\ \ j = \overline{0,N};\ \ t^k = k\tau,\ \ k = \overline{0,K}\}$ с пространственным шагом h и шагом по времени τ, и аппроксимируя производные, входящие в уравнение направленными разностями, получим следующую явную схему:

$$\frac{\varphi_i^{k+1} - \varphi_i^k}{\tau} + U\frac{\varphi_i^k - \varphi_{i-1}^k}{h} = 0. \qquad (1)$$

Значения искомой функции на новом временном слое φ_i^{k+1} могут быть явно выражены через известные значения на предыдущем слое φ_i^k:

$$\varphi_i^{k+1} = \varphi_i^k - U\frac{\tau}{h}\frac{\varphi_i^k - \varphi_{i-1}^k}{h}.$$

Здесь предполагается, что $U > 0$, т.е. движение среды осуществляется слева направо. При этом пространственная производная аппроксимирована разностью против потока. Такие схемы называются противопоточными. Из курса численных методов известно, что данная схема имеет первый порядок аппроксимации по времени и пространству и устойчива при выполнении условия $\sigma = \frac{U\tau}{h} \le 1$. Отметим, что аппроксимация пространственной производной разностью вперед приведет к абсолютно неустойчивой разностной схеме: $\frac{\varphi_i^{k+1} - \varphi_i^k}{\tau} + U\frac{\varphi_{i+1}^k - \varphi_i^k}{h} = 0$.

В качестве примера применения явной противопоточной схемы рассмотрим задачу в бесконечной области со следующим начальным условием:

$$\varphi^0(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & x < 0, x > 1 \end{cases}$$

Физически такая постановка задачи соответствует локализованной на участке $[0,1]$ зоне нахождения носителей заряда и последующему распространению носителей в условиях отсутствия диффузии и источников.

Распределения φ по продольной координате в различные моменты времени для $U = 1$, $h = 0.04$, $\sigma = 0.5$ приведены на рис. 2.3 (все величины даны в безразмерном виде). Сплошной кривой показано численное решение, штриховой – аналитическое.

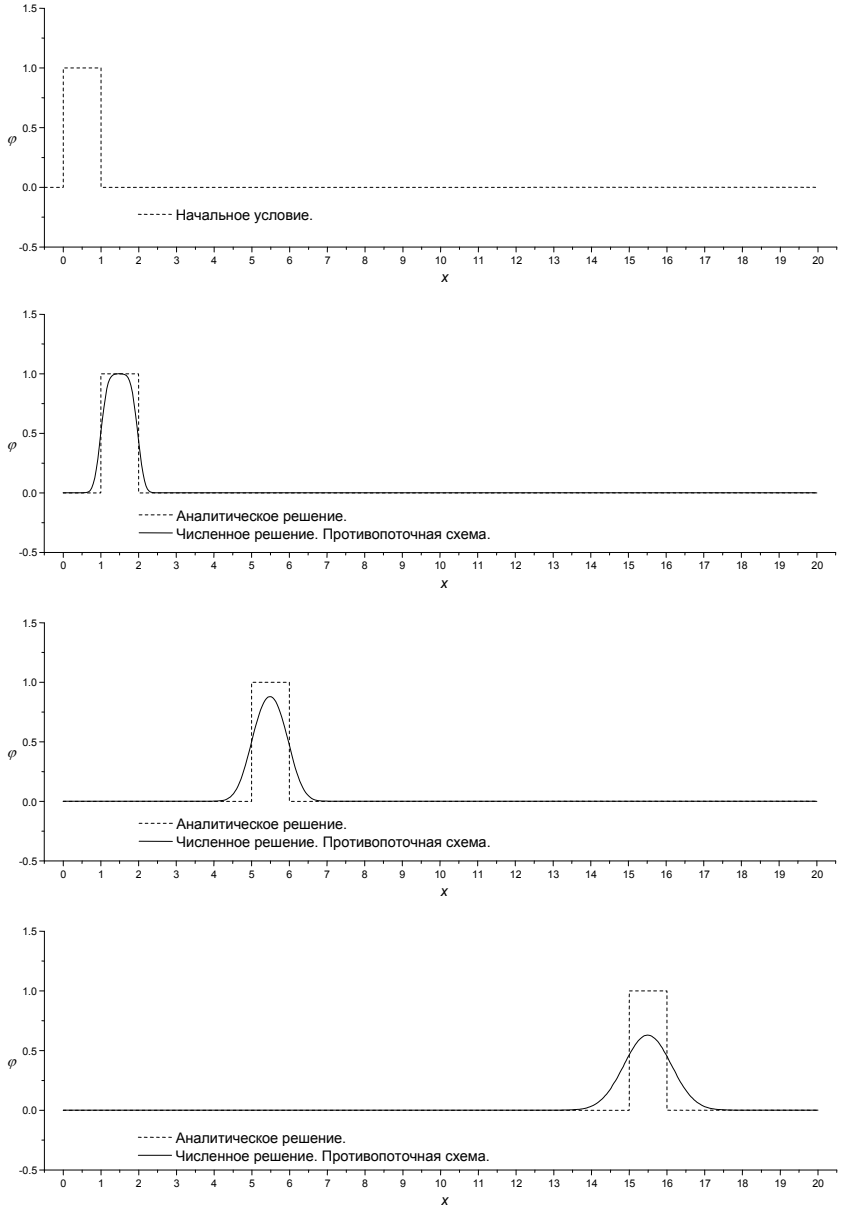


Рис. 2.3. Решение задачи переноса
в различные моменты времени $t = 1, 5, 15$

Видно, что со временем численное решение начинает заметно отличаться от точного. Профиль становится более пологим и распространяется на все большую область. Происходит постепенное «размазывание» начального профиля, подобное тому, которое имеет место при наличии диффузии. В такой ситуации создается иллюзия, что по мере удаления от зоны начальной локализации носителей заряда их концентрация убывает. Это может привести к принципиально ошибочным выводам относительно распределения носителей в системе. В действительности же наблюдаемый эффект связан с неудачным вычислительным алгоритмом.

Рассмотрим этот эффект более подробно. Построим дифференциальное приближение разностной схемы (1), выражая φ_i^{k+1} и φ_{i-1}^k в виде ряда Тейлора (предполагая существование всех необходимых производных):

$$\frac{(\varphi_i^k + \tau\varphi_t + \frac{1}{2}\tau^2\varphi_{tt} + \frac{1}{6}\tau^3\varphi_{ttt} + \ldots) - \varphi_i^k}{\tau} +$$

$$+U\frac{\varphi_i^k - (\varphi_i^k - h\varphi_x + \frac{1}{2}h^2\varphi_{xx} - \frac{1}{6}h^3\varphi_{xxx} + \ldots)}{h} = 0.$$

После несложных преобразований уравнение приводится к виду:

$$\frac{\partial\varphi}{\partial t} + U\frac{\partial\varphi}{\partial x} = -\frac{\tau}{2}\varphi_{tt} + U\frac{h}{2}\varphi_{xx} - \frac{\tau^2}{6}\varphi_{ttt} - \frac{h^2}{6}\varphi_{xxx} + \ldots$$

или

$$\varphi_t + U\varphi_x = -\frac{\tau}{2}\varphi_{tt} + U\frac{h}{2}\varphi_{xx} + O(\tau^2 + h^2). \qquad (2)$$

Выразим вторую производную по времени через вторую производную по пространству. Дифференцируя полученное уравнение по времени, получим:

$$\varphi_{tt} + U\varphi_{xt} = -\frac{\tau}{2}\varphi_{ttt} + U\frac{h}{2}\varphi_{xxt} - \frac{\tau^2}{6}\varphi_{tttt} - \frac{h^2}{6}\varphi_{xxxt} + \ldots, \qquad (3)$$

Аналогично, дифференцируя то же уравнение по x, получим:

$$\varphi_{tx} + U\varphi_{xx} = -\frac{\tau}{2}\varphi_{ttx} + U\frac{h}{2}\varphi_{xxx} - \frac{\tau^2}{6}\varphi_{tttx} - \frac{h^2}{6}\varphi_{xxxx} + \dots . \qquad (4)$$

Умножая последнее уравнение на $-U$ и прибавляя к нему (3), находим:

$$\varphi_{tt} = U^2\varphi_{xx} + O(\tau) .$$

Подставляя это выражение в (2), получим дифференциальное приближение явной противопоточной разностной схемы:

$$\varphi_t + U\varphi_x = (-\frac{\tau}{2}U^2 + U\frac{h}{2})\varphi_{xx} + O(\tau^2 + h^2). \qquad (5)$$

Отсюда хорошо видно, что разностная схема аппроксимирует исходное дифференциальное уравнение $\varphi_t + U\varphi_x = 0$ с первым порядком по времени и пространству. Однако более точно при этом аппроксимируется несколько иное дифференциальное уравнение:

$\varphi_t + U\varphi_x = \nu\varphi_{xx}$, где $\nu = (-\frac{\tau}{2}U^2 + U\frac{h}{2})$.

Здесь слагаемое в правой части имеет вид, в точности совпадающий с диффузионным слагаемым. При этом в качестве коэффициента диффузии выступает величина ν, зависящая от параметров разностной сетки τ, h. Данный эффект, характерный для схем первого порядка точности по пространству, называется схемной диффузией или в более общей формулировке – схемной диссипацией.

Приведенные выше рассуждения наглядно демонстрируют причину существенного расхождения численного и точного решений и, кроме того, указывают пути преодоления этого негативного эффекта. Первый путь связан с измельчением сетки, что позволяет снизить коэффициент схемной диссипации. При этом, однако, существенно увеличиваются вычислительные затраты, что делает такой подход малоэффективным, а зачастую и вовсе неприемлемым.

Другой более эффективный подход состоит в использовании разностных схем более высокого порядка точности по пространству. В качестве примера таких схем рассмотрим известную схему Лакса–Вендрофа:

$$\frac{\varphi_i^{k+1}-\varphi_i^k}{\tau}=-U\frac{\varphi_{i+1}^k-\varphi_{i-1}^k}{2h}+\frac{U^2\tau}{2}\frac{\varphi_{i-1}^k-2\varphi_i^k+\varphi_{i+1}^k}{h^2}. \qquad (6)$$

Схема может быть получена с использованием разложения искомой функции в ряд Тейлора:

$$\varphi(x,t+\tau)=\varphi(x,t)+\tau\varphi_t(x,t)+\frac{\tau^2}{2}\varphi_{tt}(x,t)+...$$

Заменяя φ_t на $-U\varphi_x$, а φ_{tt} на $U^2\varphi_{xx}$ получаем

$$\varphi(x,t+\tau)=\varphi(x,t)-U\tau\varphi_x(x,t)+U^2\frac{\tau^2}{2}\varphi_{xx}(x,t)+...$$

Аппроксимация производных с использованием центральных разностей дает схему Лакса–Вендрофа:

$$\varphi_i^{k+1}=\varphi_i^k-U\tau\frac{\varphi_{i+1}^k-\varphi_{i-1}^k}{2h}+\frac{U^2\tau^2}{2}\frac{\varphi_{i-1}^k-2\varphi_i^k+\varphi_{i+1}^k}{h^2}.$$

Схема имеет первый порядок точности по времени и второй по пространству. Как и противопоточная схема, она условно устойчива при $\sigma=\frac{U\tau}{h}\le 1$.

Заметим, что второе слагаемое в правой части (6) аналогично слагаемому, входящему в выражение для схемной вязкости в противопоточной схеме, но взято с обратным знаком, и поэтому может быть интерпретировано как антидиффузионное слагаемое.

Результаты решения задачи о локализованном вбросе по схеме Лакса–Вендрофа приведены на рис. 2.4.

Видно, что численное решение существенно точнее воспроизводит аналитический профиль. В то же время заметно появление ложных осцилляций в окрестности резкого изменения решения. Такое явление называется схемной дисперсией. Оно характерно для схем второго порядка по пространству и связано с искажением соотношения фаз различных волн при переходе от одного временного слоя к другому.

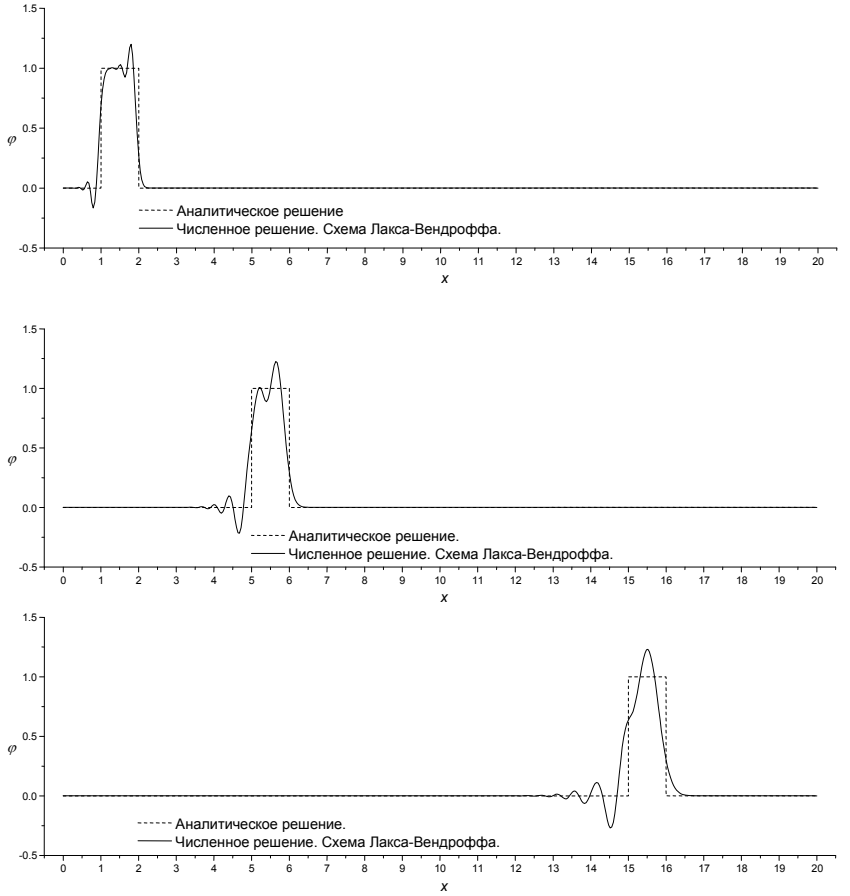


Рис. 2.4. Решение задачи переноса
в различные моменты времени $t = 1, 5, 15$

В результате изначально монотонный профиль распределения искомой функции по пространственной координате приобретает со временем осциллирующий характер. В то же время в отсутствие источниковых слагаемых в уравнении переноса монотонность должна сохраняться. Разностные схемы, обладающие свойством сохранять монотонность профиля, называются монотонными. Первая из рассмотренных выше схем (противопоточная схема первого порядка) является монотонной, в то время как более точ-

ная схема Лакса–Вендрофа таким свойством не обладает, что приводит к появлению осцилляций в решении.

Таким образом, построение эффективного вычислительного алгоритма расчета переноса вещества в подвижной среде требует с одной стороны, использование схем повышенного порядка точности (второго порядка и выше), позволяющих избежать схемной диффузии, а с другой – схем, обладающих свойством монотонности, не допускающим появления ложных осцилляций, присущих, например, многим схемам второго порядка.

В настоящее время наиболее распространенным подходом к монотонизации является использование схем, перестраивающихся в зависимости от решения. К ним относятся, в частности, TVD-схемы, при построении которых вместо довольно сильного требования монотонности используется условие невозрастания полной вариации решения (TVD – total variation diminishing) [32].

Для сеточной функции φ_i полная вариация определяется следующим образом:

$$TV(\varphi) = \sum_i \left|\varphi_i - \varphi_{i-1}\right|.$$

Основная идея TVD-подхода состоит в том, что расчет ведется всюду со вторым порядком точности, кроме окрестностей экстремумов. При приближении к экстремуму схема автоматически переключается на первый порядок точности. Этот переход обеспечивается с помощью специальных функций – лимитеров (ограничителей).

Рассмотрим применение такого подхода в сочетании с методом контрольного объема (интегро – итнтерполяционный метод) для построения разностных схем.

Введем разностную сетку $\omega_{h\tau} : \{x_j = jh,\ \ j = \overline{0,N};\ \ t^k = k\tau,\ \ k = \overline{0,K}\}$ с пространственным шагом h и шагом по времени τ.

Запишем уравнение в дивергентном виде

$$\frac{\partial \varphi}{\partial t} + \frac{\partial q}{\partial x} = 0$$

и проинтегрируем его по ячейке (контрольному объему) $[x_{i-1/2}, x_{i+1/2}]$ содержащей узел x_i:

$$\frac{\partial}{\partial t}\left(\int\limits_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \varphi dx\right) + q_{i+\frac{1}{2}} - q_{i-\frac{1}{2}} = 0.$$

Среднее по ячейке значение искомой функции $\overline{\varphi}_i = \frac{1}{h}\int\limits_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \varphi dx$,

отсюда:

$$\frac{d\overline{\varphi}_i}{dt} = \frac{1}{h}(-q_{i+1/2} + q_{i-1/2}).$$

Определим потоки $q_{i-1/2} = U\varphi_{i-1/2}$ и $q_{i+1/2} = U\varphi_{i+1/2}$ на границах ячейки. Для этого необходимо по значениям искомой функции в узлах сетки восстановить значения на границах ячейки. Принимая во внимание направление тока ($U > 0$) и используя предположение о постоянстве искомой функции в ячейке, получим:

$$\varphi_{i-1/2} = \varphi_{i-1}, \quad \phi_{i+1/2} = \phi_i.$$

В результате схема запишется следующим образом:

$$\frac{d\overline{\varphi}_i}{dt} = \frac{1}{h}U(-\varphi_{i-1} + \varphi_i).$$

Если при этом используется стандартная аппроксимация производной по времени, то получается противопоточная схема первого порядка точности, обладающая сильными диффузионными свойствами. С целью повышения точности схемы будем использовать предположение о линейном характере изменения искомой функции в ячейке. Тогда выражения для значений функции слева и справа от границы ячейки можно записать в виде:

$$\varphi^R_{i-\frac{1}{2}} \approx \varphi_i - \frac{1}{2}h\left(\frac{\partial\varphi}{\partial x}\right)_i, \qquad \varphi^L_{i-\frac{1}{2}} \approx \varphi_{i-1} + \frac{1}{2}h\left(\frac{\partial\varphi}{\partial x}\right)_{i-1}.$$

$$\varphi^R_{i+\frac{1}{2}} \approx \varphi_{i+1} - \frac{1}{2}h\left(\frac{\partial\varphi}{\partial x}\right)_{i+1}, \qquad \varphi^L_{i+\frac{1}{2}} \approx \varphi_i + \frac{1}{2}h\left(\frac{\partial\varphi}{\partial x}\right)_i.$$

Если определить производные в этих выражениях стандартным образом $\left(\frac{\partial\varphi}{\partial x}\right)_i = \frac{(\varphi_i - \varphi_{i-1})}{h}$, то получим разностную схему второго порядка точности. Однако такая схема не будет обладать свойством монотонности. Процедура монотонизации осуществляется путем ограничения производных с использованием специальных функций-ограничителей (limiter):

$$\left(\frac{\partial\varphi}{\partial x}\right)_i = \frac{1}{h}\text{limiter}\left((\varphi_{i+1} - \varphi_i), (\varphi_i - \varphi_{i-1})\right).$$

При этом возможно использование различных ограничителей. Здесь приведем один из широко распространенных лимитеров – minmod:

$$\min\text{mod}(a,b) = \begin{cases} 0, & ab \le 0 \\ a, & |a| \le |b|, ab > 0 \\ b, & |b| \le |a|, ab > 0 \end{cases}$$

где $a = \varphi_i - \varphi_{i-1}$, $\quad b = \varphi_{i+1} - \varphi_i$.

Таким образом, если производные (точнее, конечные разности) в соседних ячейках имеют одинаковый знак, выбирается минимальное по модулю из двух значений. Если же знаки разные (т.е. возможно возникновение ложных экстремумов), производная принимается равной нулю, т.е. схема переключается на противопоточную. Такой подход позволяет получить приемлемые диффузионные и дисперсионные свойства разностной схемы.

Окончательно выражения для потоков представляются в виде:

$$q_{i+1/2} = \frac{1}{2}(q^L_{i+1/2} + q^R_{i+1/2} - |U_{i+1/2}|(\varphi^R_{i+1/2} - \varphi^L_{i+1/2})).$$

Для аппроксимации производных по времени удобно использовать явные схемы Рунге–Кутты. Приведем один из возможных вариантов (метод Рунге–Кутты второго порядка).

Запишем схему в виде:

$$\frac{d\varphi}{dt} = L(\varphi)\,,\ L(\varphi) = \frac{1}{h}\left(q_+ - q_-\right).$$

Тогда переход с k-го временного слоя на k+1-й осуществляется в два этапа:

$$\varphi^* = \varphi^k + \tau L(\varphi^k)\,,$$
$$\varphi^{k+1} = 0.5\left(\varphi^k + \varphi^* + \tau L(\varphi^*)\right).$$

На рис. 2.5 представлены результаты решения задачи о локализованном выбросе с использованием приведенной выше схемы. Как видно на графиках, TVD-монотонизация позволила существенно снизить диссипативный эффект (сравн. с рис. 2.3) и исключить появление ложных осцилляций (сравн. с рис. 2.4).

Описанная выше разностная схема является явной, т.е. искомые переменные на новом временном слое в каждом узле пространственной сетки определяются простыми вычислениями по известным значениям на предыдущем слое. При этом, однако, требование устойчивости схемы накладывает ограничения на временной шаг ($\sigma < 1$), что может стать существенным при расчете продолжительных временных отрезков. Подобных ограничений можно избежать, если использовать неявные схемы. Простейшая неявная противопоточная схема первого порядка точности выглядит следующим образом:

$$\frac{\varphi_i^{k+1} - \varphi_i^k}{\tau} + U\,\frac{\varphi_i^{k+1} - \varphi_{i-1}^{k+1}}{h} = 0\,.$$

Она безусловно устойчива, однако, как и ее явный аналог, обладает сильными диссипативными свойствами. Возможно построение монотонизированных неявных схем второго порядка точности.

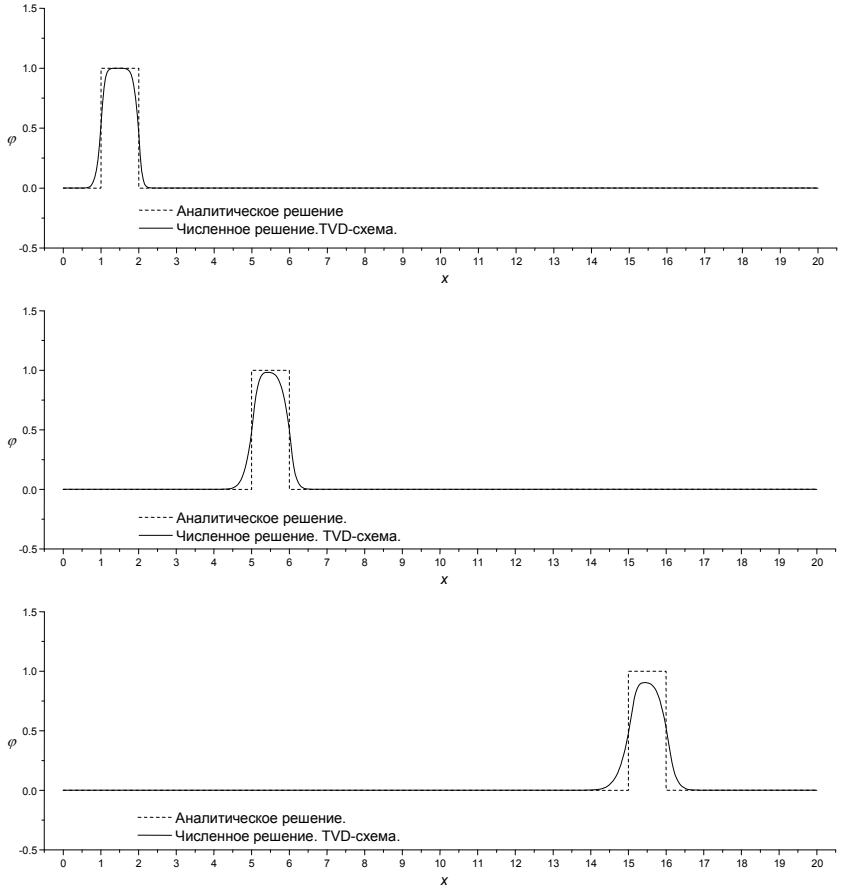


Рис. 2.5. Решение задачи переноса
в различные моменты времени t = 1, 5, 15

Выше рассматривались вопросы построения разностных схем для уравнения дрейфа носителей заряда. Перейдем к алгоритмам решения задачи о дрейфе-диффузии носителей с учетом источников заряда. Уравнение переноса имеет вид:

$$\frac{\partial \varphi}{\partial t} + \frac{\partial q}{\partial x} = f(\varphi, x, t) ,$$

где плотность потока величины φ складывается из дрейфовой и диффузионной составляющих: $q = U\varphi - D\dfrac{\partial \varphi}{\partial x}$.

При постоянных U, D уравнение дрейфа–диффузии можно записать следующим образом:

$$\frac{\partial \varphi}{\partial t} + U\frac{\partial \varphi}{\partial x} = D\frac{\partial^2 \varphi}{\partial x^2} + f(\varphi, x, t).$$

Аппроксимируя конвективное слагаемое разностью против потока, а диффузионное – центральной разностью, получим простую явную разностную схему:

$$\frac{\varphi_i^{k+1} - \varphi_i^k}{\tau} + U\frac{\varphi_i^k - \varphi_{i-1}^k}{h} = D\frac{\varphi_{i-1}^k - 2\varphi_i^k + \varphi_{i+1}^k}{h^2} + f(\varphi_i^k, x_i, t^k).$$

Схема имеет первый порядок аппроксимации и условно устойчива при $\sigma = \frac{U\tau}{h} \le 1$, $r = \frac{D\tau}{h^2} \le \frac{1}{2}$. Здесь σ, r – числа Куранта соответственно для дрейфовой и диффузионной задач.

Как отмечалось выше, первый порядок аппроксимации конвективного слагаемого приводит к возникновению схемной диффузии. Анализ дифференциального приближения для этой схемы показывает, что в действительности решается задача с коэффициентом диффузии, равным $D + \nu$, где $\nu = (-\frac{\tau}{2}U^2 + U\frac{h}{2})$ – коэффициент схемной диффузии.

Следовательно, схема пригодна для решения задач с преобладающим влиянием физической диффузии. Если использовать центральную разность для аппроксимации дрейфового слагаемого, получим схему второго порядка точности по пространству:

$$\frac{\varphi_i^{k+1} - \varphi_i^k}{\tau} + U\frac{\varphi_{i+1}^k - \varphi_{i-1}^k}{2h} = D\frac{\varphi_{i-1}^k - 2\varphi_i^k + \varphi_{i+1}^k}{h^2} + f(\varphi_i^k, x_i, t^k).$$

Отметим, что в отсутствие диффузионного слагаемого (уравнение дрейфа), такая схема была бы абсолютно неустойчивой. В данном случае схема оказывается устойчивой при выполнении условий: $\sigma = \frac{U\tau}{h} \le 1$, $r = \frac{D\tau}{h^2} \le \frac{1}{2}$, $2\sigma \le R_h \le 2/\sigma$.

Здесь $R_h = \frac{Uh}{D} = \frac{\sigma}{r}$ выражает соотношение дрейфовой и диффузионной составляющих переноса. Данная схема, а также ее неявный аналог:

$$\frac{\varphi_i^{k+1} - \varphi_i^k}{\tau} + U\frac{\varphi_{i+1}^{k+1} - \varphi_{i-1}^{k+1}}{2h} = D\frac{\varphi_{i-1}^{k+1} - 2\varphi_i^{k+1} + \varphi_{i+1}^{k+1}}{h^2} + f(\varphi_i^{k+1}, x_i, t^{k+1})$$

и их комбинации имеют довольно широкое распространение при решении задач рассматриваемого класса. К сожалению, эти схемы не обладают свойством монотонности, поэтому их применение может приводить к появлению ложных осцилляций. В особой мере осциллирующий характер решения проявляется в задачах с преобладающим влиянием дрейфа носителей. В этой связи актуальным оказывается применение монотонизирующих процедур, аналогичных рассмотренным выше применительно к задаче дрейфового переноса.

Рассмотрим здесь подход, позволяющий использовать уже имеющийся опыт монотонизации. Он основан на идее расщепления вычислений по физическим процессам. На каждом шаге по времени последовательно решаются две задачи. Сначала решается задача дрейфа носителей (диффузионное и источниковое слагаемые опускаются) $\frac{\partial\varphi}{\partial t} + U\frac{\partial\varphi}{\partial x} = 0$, а затем решается задача диффузии: $\frac{\partial\varphi}{\partial t} = D\frac{\partial^2\varphi}{\partial x^2} + f(\varphi, x, t)$.

Например, в случае применения явной противопоточной схемы для первого уравнения и явной центрально-разностной схемы для второго – алгоритм вычислений можно записать следующим образом:

$$\frac{\varphi_i^p - \varphi_i^k}{\tau} + U\frac{\varphi_i^k - \varphi_{i-1}^k}{h} = 0,$$

$$\frac{\varphi_i^{k+1} - \varphi_i^p}{\tau} = D\frac{\varphi_{i-1}^p - 2\varphi_i^p + \varphi_{i+1}^p}{h^2} + f(\varphi_i^p, x_i, t^p).$$

Такие схемы называются аддитивными схемами или схемами расщепления. Они не аппроксимируют исходное дифференциальное уравнение в точности на каждом из этапов, однако обладают свойством суммарной аппроксимации.

Если теперь на первом этапе для вычисления промежуточных значений φ_i^p вместо противопоточной схемы применить рассмотренную выше TVD-схему, итоговая разностная схема будет обладать приемлемыми диффузионными и дисперсионными свойствами.

Глава 3

ЗАДАЧИ ОПТИМИЗАЦИИ НАНОРАЗМЕРНЫХ ПОЛУПРОВОДНИКОВЫХ ГЕТЕРОСТРУКТУР

Выходные характеристики гетероструктурных полупроводниковых СВЧ устройств определяются многообразием факторов, отражающих как конструктивные, так и технологические особенности изготовления гетероструктуры. При этом чрезвычайно важен выбор оптимальных параметров, определяющих основные электрофизические характеристики структуры – концентрацию и подвижность носителей заряда в каналах двумерного электронного газа (ДЭГ) [33].

Вопросам математического моделирования таких структур рассмотрены в предыдущих главах 1, 2.

В третьей главе были представлены результаты многовариантного компьютерного анализа плотности и подвижности электронов в наноразмерных нитридных гетероструктурах AlGaN/GaN. Вычислительные эксперименты позволили определить зависимости концентрации и подвижности электронов от мольного содержания алюминия в барьерном слое, от толщины барьерного слоя, а также ряда других характеристик. В ряде случаев построение таких зависимостей дает возможность определить оптимальные с точки зрения концентрации и подвижности носителей параметры гетероструктуры. Однако часто возникают задачи, когда необходимо применение аппарата теории оптимизации. К таким задачам относятся, в частности, обратные задачи по определению недоступных в прямом измерении характеристик выращенных гетероструктур на основе экспериментальных данных по концентрации и подвижности носителей в двумерном

электронном газе. Другой класс задач связан с определением оптимального легирования барьерного слоя, состоящего из ряда подслоев. Такие задачи можно классифицировать как задачи оптимального управления. Этот класс задач рассматривается в настоящей главе.

3.1. Постановка задач оптимизации

Математическая модель, описывающая распределение электронов в наноразмерных полупроводниковых гетероструктурах подробно описана в главе 1. Здесь для придания цельности материалу приведем основные уравнения в одномерной постановке, которая используется далее в оптимизационной задаче:

$$-\frac{\hbar^2}{2}\frac{d}{dz}\left(\frac{1}{m^*(z)}\frac{d\psi}{dz}\right)+V(z)\psi(z)=E\psi(z) \tag{1}$$

$$\frac{d}{dz}\left(\varepsilon(z)\frac{d\varphi}{dz}\right)=-e\left(u(z)-n(z)\right)+\sum_{l=1}^{M}\sigma_l\delta(z-z_l) \tag{2}$$

$$V(z)=-e\varphi(z)+\Delta E_c(z) \tag{3}$$

$$n(z)=\sum_{i=1}^{I}\left(\psi_i(z)\right)^2 n_i(z) \tag{4}$$

$$n_i(z)=k_B T\frac{m^*(z)}{\pi\hbar^2}\ln\left(1+\exp\left(\frac{E_F-E_i}{k_B T}\right)\right) \tag{5}$$

Здесь E_i и $\psi_i(z)$ – энергетические уровни и соответствующие им волновые функции, I – количество значимых с точки зрения вклада в сумму (4) энергетических уровней, $n(z)$ – электронная плотность, $\hbar$ – постоянная Планка, e – заряд электрона, m^* – эффективная масса электрона, E_F – положение уровня Ферми, $\varphi(z)$ - электростатический потенциал, σ_l – плотности зарядов на интерфейсах, δ – дельта функция, z_l – местоположения интерфейсов, M – число гетероинтерфейсов с наличием нескомпенсированного заряда, ε – диэлектрическая постоянная матери-

ала, ΔE_c – сдвиг зоны проводимости материала, k_B – постоянная Больцмана, T – температура. Функция $u(z)$ описывает распределение легирующих примесей в системе, $u(z) = N_d(z) - N_A(z)$, где $N_d(z), N_A(z)$ – концентрации донорной и акцепторной примесей. Эта функция в дальнейшем будет рассматриваться как функция управления. В модели (1)–(5) отражен тот факт, что значения m^*, ε и ΔE_c могут меняться от слоя к слою, т.е. функции $m^*(z)$, $\varepsilon(z)$, $\Delta E_c(z)$ принадлежат классу кусочно-постоянных функций. При этом температура в гетероструктуре предполагается постоянной.

На границах системы ($z = 0$, $z = L$, где L – общая толщина слоистой структуры) должны выполняться условия равенства нулю волновых функций:

$$\psi(0) = 0, \quad \psi(L) = 0 \tag{6}$$

На левой границе ($z = 0$), задается потенциальный барьер φ_b, образующийся в приконтактном слое полупроводника, граничащего с металлом (барьер Шоттки). Кроме того, возможно задание смещения φ_s вследствие приложенного напряжения. На правой границе (z=L) естественным условием является отсутствие электрического поля. Таким образом, граничные условия для уравнения Пуассона имеют вид:

$$\varphi(0) = \varphi_b + \varphi_s, \quad \varphi'(L) = 0 \tag{7}$$

Перейдем к формулировке задачи оптимизации. Наиболее важным показателем эффективности гетероструктуры является концентрация электронов в канале двумерного электронного газа $N = \int_0^L n(z)dz$. Как показано в главе 2, эта величина во многом является определяющей и для подвижности электронов. Хотя с ее увеличением подвижность, как правило, падает, однако произведение концентрации и подвижности (проводимость) возрастает (см. главу 2), что определяет стремление к повышению N. Необ-

ходимо отметить, что излишнее повышение концентрации может сопровождаться и негативными эффектами [33]. Поэтому разумным компромиссом обычно является задание желаемого уровня концентрации электронов N^*. Основным управляющим фактором, воздействующим на величину N^*, является концентрация доноров в барьерном слое – области, расположенной слева от канала двумерного электронного газа (см. рис. 1.1), $0 \le z \le z_b$, z_b – толщина барьерного слоя, $z_b < L$.

При этом естественным является желание уменьшить средний уровень легирования (или суммарную ионизацию) барьерного слоя, чтобы избежать излишнего рассеяния электронов на удаленных заряженных центрах. Таким образом, минимизируемый функционал можно записать в виде:

$$J(u) = \left(\int n(z)dz - N^*\right)^2 + w\left(\int u(z)dz\right)^2, \tag{8}$$

где w – весовой параметр. Здесь и далее пределы интегрирования опущены, так как всегда подразумевается интегрирование по всей области.

Могут существовать технологические ограничения по легированию отдельных слоев при выращивании гетероструктуры, что отражается в ограничениях на функцию управления:

$$0 \le u(z) \le \xi(z). \tag{9}$$

Здесь левое ограничение означает, что рассматривается только донорное легирование. Это связано с тем, что именно такое легирование способствует повышению концентрации электронов в двумерном электронном газе. Кроме того, при стремлении к достижению максимально высокой концентрации электронов (высокое значение N^*) может стать существенным ограничение на состояние системы:

$$V(z) - E_f \ge 0, \quad z \le z_b. \tag{10}$$

Данное ограничение позволяет избежать образования параллельного канала проводимости в барьерном слое ($z \le z_b$).

Таким образом, задача нахождения распределения легирующей примеси в барьерном слое, обеспечивающего заданную концентрацию электронов в канале ДЭГ (проводимости) поставлена. Ясно, что во многих случаях решение такой задачи не является единственным – заданной концентрации электронов могут соответствовать различные варианты легирования. В этом случае проектировщику гетероструктуры могут быть представлены на выбор различные варианты. При этом выбор существенно сужается за счет введенных ограничений и влияния второго слагаемого функционала.

3.2. Вычисление градиента функционала. Сопряженная задача

Для решения задачи (1)–(10) эффективными являются градиентные методы.

Остановимся на вопросе определения градиента функционала. Здесь принципиальным моментом является установленная ранее связь между приращениями электронной плотности и потенциала (см. главу 1):

$$\delta n(z) = R(z)\delta\varphi(z)\,, \quad R(z) = \frac{m^*(z)e}{\pi\hbar^2}\sum_i \psi_i^2(z)\frac{\exp\left(\dfrac{E_F - E_i}{k_B T}\right)}{1+\exp\left(\dfrac{E_F - E_i}{k_B T}\right)} \qquad (11)$$

Применение выражения (11) существенно упрощает процедуру определения градиента функционала.

Следуя традиционной схеме вариационного исчисления [34], домножим уравнение (2) на множитель Лагранжа $p(z)$, проинтегрируем результат по толщине гетероструктуры и прибавим к функционалу (8):

$$\begin{aligned} J(u) &= \left(\int n(z)dz - N^*\right)^2 + w\left(\int u(z)dz\right)^2 + \\ &+\int p(z)\left(\frac{d}{dz}\left(\varepsilon(z)\frac{d\phi}{dz}\right) + e\big(u(z)-n(z)\big) - \sum_l \sigma_l\delta(z-z_l)\right)dz. \end{aligned} \qquad (12)$$

Вариация функционала (12) выглядит следующим образом:

$$\delta J = 2(N - N^*)\int \delta n(z)dz + 2wU\int \delta u(z)dz + \\ + \int p(z)\left(\frac{d}{dz}\left(\varepsilon(z)\frac{d\delta\varphi}{dz}\right) + e\left(\delta u(z) - \delta n(z)\right)\right)dz, \tag{13}$$

где $U = \int u(z)dz$.

Проинтегрируем слагаемое $\int p(z)\left(\frac{d}{dz}\left(\varepsilon(z)\frac{d\delta\varphi}{dz}\right)\right)dz$ дважды по частям. С учетом граничных условий (7) получим:

$$\int p(z)\left(\frac{d}{dz}\left(\varepsilon(z)\frac{d\delta\varphi}{dz}\right)\right)dz = \\ = \int \frac{d}{dz}\left(\varepsilon(z)\frac{dp}{dz}\right)\delta\varphi dz - p(0)\varepsilon(0)\frac{d\delta\varphi}{dz}(0) - \frac{dp}{dz}(L)\varepsilon(L)\delta\varphi(L).$$

Тогда (13) примет вид:

$$\delta J = 2(N - N^*)\int \delta n(z)dz + 2wU\int \delta u(z)dz + \int \frac{d}{dz}\left(\varepsilon(z)\frac{dp}{dz}\right)\delta\varphi + \\ + p(z)e\left(\delta u(z) - \delta n(z)\right)dz - p(0)\varepsilon(0)\frac{d\delta\varphi}{dz}(0) - \frac{dp}{dz}(L)\varepsilon(L)\delta\varphi(L)$$

или с учетом выражения (11):

$$\delta J = 2(N - N^*)\int R(z)\delta\varphi(z)dz + 2wU\int \delta u(z)dz + \int \frac{d}{dz}\left(\varepsilon(z)\frac{dp}{dz}\right)\delta\varphi + \\ + p(z)e(\delta u(z) - R(z)\delta\varphi(z))dz - p(0)\varepsilon(0)\frac{d\delta\varphi}{dz}(0) - \frac{dp}{dz}(L)\varepsilon(L)\delta\varphi(L).$$

Составим систему уравнений для определения $p(z)$ так, чтобы занулить все слагаемые, не содержащие δu:

$$\frac{d}{dz}\left(\varepsilon(z)\frac{dp}{dz}\right) - eR(z)p(z) + 2R(z)(N - N^*) = 0 \tag{14}$$

$$p(0) = 0, \quad \frac{dp}{dz}\Big|_{z=L} = 0 \tag{15}$$

В результате получим следующее выражение для вариации функционала:

$$\delta J = \int p(z) e \delta u(z) dz + 2wU \int \delta u(z) dz = \int (p(z)e + 2wU) \delta u(z) dz \qquad (16)$$

Отсюда определяется искомый градиент функционала:

$$J' = p(z)e + 2wU\ . \qquad (17)$$

Отметим, что такой подход позволил построить сравнительно простую (сопряженную) задачу (14)–(15) для определения функции $p(z)$. Таким образом, при определении градиента функционала удается избежать решения наиболее вычислительно затратной спектральной задачи. Это позволяет построить эффективные алгоритмы оптимизации, основанные на градиентных методах. В частности, результаты, представленные в следующем разделе, были получены с использованием простейшего метода градиентного спуска. Учет ограничений в настоящей работе проводился с использованием проекции решения на каждой итерации на подпространство ограничений.

Остановимся подробнее на методе вычисления оптимального шага спуска. Согласно логике точного линейного поиска шаг выбирается из условия минимизации нового значения функционала.

Пусть на некотором шаге оптимизационного процесса определена функция $u(z)$, направление дальнейшего спуска $\delta u(z)$ и требуется вычислить шаг в этом направлении из условия $\min\limits_{\alpha} J\left(u(z) + \alpha \delta u(z)\right)$.

Не ограничивая общности, будем рассматривать следующий функционал $J(u) = \left(\int n(z) dz - N^*\right)^2$. Запишем функционал в виде $J = \left(\int n\left(\varphi(u + \alpha\delta u)\right) dz - N^*\right)^2$ и, удерживая линейные слагаемые, получим

$$J \approx (\int n(\varphi(u + \alpha\delta u)(u) + \alpha\delta\varphi(u + \alpha\delta u))) dz - N^*)^2 \approx$$

$$\approx (\int (n(\varphi(u + \alpha\delta u)(u)) + \alpha\delta n) dz - N^*)^2.$$

Дифференцирование по α дает

$$\frac{\partial J}{\partial \alpha} = 2\left(\int (n(\varphi(u)) + \alpha\delta n)dz - N^*\right)\int \delta n dz.$$

Отсюда, приравнивая производную нулю, получаем выражение для определения шага спуска:

$$\alpha = \frac{N^* - \int n(\varphi(z))dz}{\int \delta n dz}. \qquad (18)$$

Выше была отмечена связь между приращениями δn и $\delta\varphi$: $\delta n(z) = R(z)\delta\varphi(z)$. Таким образом, достаточно получить уравнение для определения $\delta\varphi$. Оно может быть получено с использованием линейной теории чувствительности. Запишем уравнение (2) с учетом приращений δu, $\delta\phi$:

$$\frac{d}{dz}\left(\varepsilon\frac{d(\varphi + \delta\varphi)}{dz}\right) = -e(u + \delta u - n - R\delta\varphi) + \sum_{l=1}^{M}\sigma_l\delta(z - z_l).$$

Вычитая уравнение (2), получим уравнение, связывающее приращения:

$$\frac{d}{dz}\left(\varepsilon\frac{d\delta\varphi}{dz}\right) = -e(\delta u - R\delta\varphi). \qquad (19)$$

Данное уравнение дополняется граничными условиями:

$$\delta\varphi(0) = 0, \qquad \delta\varphi'(L) = 0.$$

(20)

Таким образом, для нахождения приращения $\delta\varphi(z)$ по известному приращению $\delta u(z)$, и соответственно, для определения шага спуска согласно (18), требуется решить задачу (19)–(20), аналогичную по типу задаче (14)–(15) для нахождения градиента функционала. В результате на каждом шаге оптимизационного процесса, осуществляемого с использованием градиентных методов, решаются три задачи – прямая задача (1)–(5) для нахождения

значения функционала, сопряженная задача (14)–(15) для нахождения градиента функционала и задача чувствительности (19)–(20) для нахождения шага спуска. Некоторые варианты градиентных методов оптимизации рассматриваются ниже.

3.3. Методы решения задач оптимизации

Некоторые постановки задач оптимизации наноразмерных полупроводниковых гетероструктур были в предыдущем разделе. Как правило, такие задачи сводятся к минимизации функционала $J(u(\mathbf{r}))$ (или после дискретизации – функции многих переменных $J(\mathbf{u})$), причем для вычисления значения функционала при определенной функции управления $u(\mathbf{r})$ необходимо решение прямой задачи, например, как было показано выще, решение системы уравнений Шрёдингера–Пуассона. Таким образом, в оптимизационном цикле, направленном на поиск минимума функционала, многократно решается прямая задача. Поскольку процедура решения такой задачи может быть вычислительно затратной, актуальным вопросом является экономичность оптимизационного процесса. С точки зрения технологии построения вычислительного процесса (и программного обеспечения) наиболее простым вариантом является применение методов оптимизации нулевого порядка, не требующих вычисления градиента функционала. В этом случае на каждом шаге оптимизационного цикла вычисляются лишь значения функционала. Как правило, подобные методы обладают довольно низкой скоростью сходимости. Этот недостаток в особой мере проявляется в задачах с большой размерностью искомого вектора (более 100). Учитывая, что дискретизация характерных для настоящей работы задач должна осуществляться на достаточно мелких сетках с целью должного разрешения особенностей распределения электронов в гетероструктурах, непосредственное применение методов нулевого порядка к поиску минимума функционала является проблематичным даже при решении задач в одномерной постановке. Однако эта проблема может быть частично решена путем параметриза-

ции искомой функции управления, что существенно понижает размерность искомого вектора. К настоящему времени создано большое количество методов нулевого порядка, с которыми можно познакомиться в учебных пособиях и монографиях, посвященных методам оптимизации. Здесь приведем метод Нелдера–Мида (деформируемого многогранника), который хорошо зарекомендовал себя при решении рассматриваемых в работе задач оптимизации в параметризованной форме.

Метод Нелдера–Мида заключается в последовательном перемещении и деформировании симплекса вокруг точки экстремума. Симплекс состоит из $n_u + 1$ вершин, где n_u – размерность искомого вектора. Алгоритм строится следующим образом.

1. Задается начальный симплекс с координатами $\mathbf{u}_1$, $\mathbf{u}_2$, …, $\mathbf{u}_{n_u+1}$, коэффициенты отражения α, сжатия β, растяжения γ, редукции σ, точность $\varepsilon > 0$.

2. Определяется лучшая точка симплекса $\mathbf{u}_b$ (доставляющая минимальное значение функционалу $J(\mathbf{u})$), худшая $\mathbf{u}_h$ (с максимальным значением функционала) и вторая после худшей точка $\mathbf{u}_s$.

3. Вычисляется центр тяжести всех вершин, кроме худшей $\mathbf{u}_m = \dfrac{1}{n_u}\left(\sum_{j=1}^{n_u+1} \mathbf{u}_j - \mathbf{u}_h\right)$.

4. Если $\sqrt{\dfrac{1}{n_u+1}\sum_{j=1}^{n_u+1}\left(J(\mathbf{u}_j) - J(\mathbf{u}_m)\right)^2} \le \varepsilon$, то выполнение алгоритма завершается.

5. Осуществляется отражение худшей вершины относительно центра тяжести $\mathbf{u}_r = \mathbf{u}_m + \alpha(\mathbf{u}_m - \mathbf{u}_h)$.

6. Если $\mathbf{u}_b \le \mathbf{u}_r < \mathbf{u}_s$, то $\mathbf{u}_h = \mathbf{u}_r$, осуществляется переход на шаг 2.

7. Если $\mathbf{u}_r < \mathbf{u}_b$, то направление выбрано удачно, делается растяжение многогранника в этом направлении $\mathbf{u}_e = \mathbf{u}_m + \gamma(\mathbf{u}_m - \mathbf{u}_r)$. Если $\mathbf{u}_e < \mathbf{u}_r$, то $\mathbf{u}_h = \mathbf{u}_e$, в противном случае $\mathbf{u}_h = \mathbf{u}_r$, осуществляется переход на шаг 2.

8. Если $\mathbf{u}_r \geq \mathbf{u}_s$, то осуществляется сжатие $\mathbf{u}_c = \mathbf{u}_m + \beta(\mathbf{u}_m - \mathbf{u}_h)$. Если $\mathbf{u}_c < \mathbf{u}_r$, то $\mathbf{u}_h = \mathbf{u}_c$, осуществляется переход на шаг 2.

9. Формируется новый симплекс с уменьшенными сторонами (сжатие симплекса относительно лучшей вершины $\boldsymbol{\psi}_l$) $\mathbf{u}_j = \mathbf{u}_b + \sigma(\mathbf{u}_j - \mathbf{u}_l), j = 1,...,n_u + 1$, осуществляется переход на шаг 2.

Типичные значения параметров для этого алгоритма $\alpha = 1$, $\beta = 0.5$, $\gamma = 2$, $\sigma = 0.5$. В ряде случаев полезной оказывается модификация метода деформируемого многогранника применительно к решению задач нелинейного программирования с ограничениями-неравенствами (комплексный метод Бокса).

Существенного ускорения оптимизационного процесса можно добиться с помощью методов оптимизации первого порядка. При этом важным элементом становится вычисление градиента функционала. Подробное изложение этого вопроса применительно к задачам оптимального легирования дано в предыдущем разделе. Здесь дадим описание одного из наиболее эффективных методов первого порядка – метода сопряженных градиентов (этот метод уже рассматривался нами в контексте решения разреженных систем линейных алгебраических уравнений). При использовании данного метода для решения задач оптимизации важен тот факт, что минимум квадратичной нормы теоретически (без учета погрешностей различного рода) находится за n (размерность пространства) шагов при поиске вдоль сопряженных относительно матрицы Гессе направлений (**H**-сопряженных направлений). На практике, в окрестности минимума большой класс функций может быть достаточно точно представлен в виде квадратичной аппроксимации, что позволяет эффективно применять метод для различных функций.

Вычислительный процесс строится следующим образом:

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} + \alpha_k \mathbf{p}_k, \quad \mathbf{p}_{k+1} = -\nabla J(\mathbf{u}^{(k+1)}) + \beta_k \mathbf{p}_k .$$

В качестве начального приближения $\mathbf{p}_0$, как правило, выбирается шаг метода наискорейшего градиентного спуска.

С точки зрения эффективности вычисления принципиальными являются три момента: экономичный расчет градиента функционала $\nabla J(\mathbf{u}^{(k+1)})$, выбор коэффициента β_k, определяющего направление спуска, выбор коэффициента α_k, определяющего шаг спуска вдоль выбранного направления.

Метод вычисления градиента функционала путем построения сопряженного уравнения рассмотрен в разделе 3.2. Для расчета коэффициента β предложен целый ряд подходов, подробный обзор которых представлен, например, в работе [35]. Здесь приведем исторически первую для данного метода формулу Флетчера–Ривса и формулу Хагера–Чанга (Hager, Zhang) [36], которая хорошо зарекомендовала себя при решении широкого класса оптимизационных задач.

Формула Флетчера–Ривса имеет вид:

$$\beta_k = \frac{\|\mathbf{g}_{k+1}\|^2}{\|\mathbf{g}_k\|^2}, \quad \mathbf{g}_k = \nabla J(\mathbf{u}^{(k)}) .$$

Наилучшим образом эта формула проявляет себя при оптимизации квадратичных функций.

Более универсальной является формула Хагера–Чанга [36]:

$$\beta_k = \max\left\{\tilde{\beta}_k, \eta_k\right\},$$

$$\tilde{\beta}_k = \left(\mathbf{y}_k - 2\mathbf{p}_k \frac{\|\mathbf{y}_k\|^2}{\mathbf{p}_k^{\mathrm{T}}\mathbf{y}_k}\right)^T \frac{\mathbf{g}_{k+1}}{\mathbf{p}_k^{\mathrm{T}}\mathbf{y}_k}, \quad \mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k ,$$

$$\eta_k = \frac{-1}{\|\mathbf{p}_k\| \min\{\eta, \|\mathbf{g}_k\|\}},$$

где $\eta > 0$ – заданная константа. Авторы предлагают значение константы $\eta = 0,01$. Такой подход позволяет избежать негативных эффектов, связанных с потерей свойства **H**-сопряженности направлений вследствие ошибок вычислений. Реализация метода сопряженных градиентов с таким выбором направления спуска получила название CG_decent (метод сопряженных градиентов с гарантированным спуском). Обоснование метода и детали его применения можно найти в [36].

Одним из важных элементов реализации методов сопряженных градиентов является эффективная процедура подбора оптимального шага (коэффициента α). Выбор шага, по сути, должен удовлетворять двум противоречащим условиям: с одной стороны, учитывая, что каждое вычисление функционала связано с решением прямой задачи, важно, чтобы данный подбор был осуществлен с минимальными накладными расходами, с другой стороны, требуется, чтобы значение целевой функции существенно уменьшалось. В этом плане можно отметить методы неточного линейного поиска. Типичный алгоритм подобного типа состоит из 2 фаз: во-первых, определяются интервалы допустимых величин шагов (*bracketing*), во-вторых, выполняются процедуры бисекции и/или интерполяции для выбора конкретного значения квазиоптимального шага из интервалов. Окончание процедуры неточного поиска шага вдоль направления определяется, как правило, условиями Вульфа (Wolfe):

$$\varphi(\alpha_k) \le \varphi(0) + c_1 \alpha_k \varphi'(0), \quad \varphi'(\alpha_k) \ge c_2 \varphi'(0),$$

$$0 < c_1 < c_2 < 1, \quad \varphi(\alpha_k) = J(\mathbf{u}^{(k)} + \alpha_k \mathbf{p}_k).$$

Первое условие требует, чтобы при выбранном шаге α_k существенно уменьшалось значение целевой функции. Второе условие соответствует существенному уменьшению наклона целевой функции. Типичные значения для методов сопряженных градиентов: $c_1 = 10^{-4}$, $c_2 = 0.1$.

Однако, как показывает опыт решения задач рассматриваемого в работе класса, наиболее эффективным является подход, основанный на сочетании точного линейного поиска и функций чувствительности. Этот подход был описан в предыдущем разделе.

Рассмотренные методы позволяют найти локальный минимум целевой функции, для поиска глобального минимума применяются методы глобальной оптимизации. В настоящий работе применялся один из простейших подходов – метод рестартов со случайным выбором начального приближения. Существенным достоинством данного метода является высокая степень распараллеливания.

Отметим, что в плане глобальной оптимизации заслуживают внимания получающие все большее развитие метаэвристические методы, такие, как генетические алгоритмы, метод имитации отжига, метод частиц в стае и т.д. [37].

3.4. Бессеточный алгоритм решения прямых и обратных задач математической физики

В заключение настоящей главы дадим описание универсального бессеточного метода, который позволяет в рамках единой алгоритмической структуры решать как прямые, так и обратные задачи (задачи идентификации и оптимального управления) [38]. Метод успешно применялся авторами для решения широкого класса задач математической физики.

Рассмотрим прямую задачу в общей постановке. Ищется функция $\varphi(\mathbf{x})$, удовлетворяющая уравнению

$$A\varphi = f(\mathbf{x}), \quad \mathbf{x} \in \Omega \subset R^d,$$

и граничным условиям на границе Γ

$$B\varphi = g(\mathbf{x}), \quad \mathbf{x} \in \Gamma,$$

где A, B дифференциальные операторы, f, g – некоторые заданные функции пространственных координат. Специальных требований к границе не предъявляется.

Будем искать решение в виде разложения по адаптивному базису, состоящему из нормализованных радиальных базисных функций:

$$\tilde{\varphi}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \boldsymbol{\omega}_i \xi(\| \mathbf{x} - \mathbf{x}_i^c \|_2, \boldsymbol{\sigma}_i)}{\sum_{i=1}^{n} \xi(\| \mathbf{x} - \mathbf{x}_i^c \|_2, \boldsymbol{\sigma}_i)},$$

где $\boldsymbol{\omega}, \boldsymbol{\sigma}, \mathbf{x}^c$ – веса, «ширины» и координаты центров базисных функций, ξ – радиальная базисная функция, в качестве которой можно использовать различные варианты, например, $\xi(r) = \exp(-r^2)$, $\xi(r) = (1 + r^2)^{-1}$ и др.

Один из возможных подходов для подбора параметров разложения заключается в минимизации квадратичного функционала ошибки, который формируется при подстановке разложения в исходное уравнение и граничные условия. Достоинством данного подхода является его универсальность, т.к. он не накладывает ограничений на конфигурацию области, на гладкость коэффициентов и т.д.

Функционал ошибки имеет вид:

$$J(\tilde{\varphi}) = \int_{\Omega} w_{\Omega} \left[A(\tilde{\varphi}(\mathbf{x}, \boldsymbol{\rho})) - f(\mathbf{x}) \right]^2 d\mathbf{x} + \int_{\Gamma} w_{\Gamma} \left[B(\tilde{\varphi}(\mathbf{x}, \boldsymbol{\rho})) - g(\mathbf{x}) \right]^2 d\mathbf{x},$$

где $\boldsymbol{\rho}$ – искомый вектор параметров, включающий $\boldsymbol{\omega}, \boldsymbol{\sigma}, \mathbf{x}^c$, w_{Ω}, w_{Γ} – коэффициенты, которые «выравнивают» вклад внутренней и граничной составляющей.

Как правило, вычислить интегралы аналитически довольно сложно или невозможно, поэтому используется дискретный аналог функционала:

$$J(\tilde{\varphi}) = \sum_{j=1}^{m_{\Omega}} w_{\Omega} \left[A(\tilde{\varphi}(\mathbf{x}_j^{\Omega}, \boldsymbol{\rho})) - f(\mathbf{x}_j^{\Omega}) \right]^2 + \sum_{j=1}^{m_{\Gamma}} w_{\Gamma} \left[B(\tilde{\varphi}(\mathbf{x}_j^{\Gamma}, \boldsymbol{\rho})) - g(\mathbf{x}_j^{\Gamma}) \right]^2,$$

где m_{Ω}, m_{Γ} – число контрольных точек в области и на границе, соответственно, $\{\mathbf{x}\}_j^{\Omega} \in \Omega$ – набор контрольных точек в области,

$\{\mathbf{x}\}_j^{\Gamma} \in \Gamma$ – на границе. Подбор параметров $\boldsymbol{\rho}$ аппроксимирующей функции $\tilde{\varphi}$ осуществляется путем минимизации сформированного функционала:

$$J(\tilde{\varphi}(\mathbf{x}, \boldsymbol{\rho})) \xrightarrow{\boldsymbol{\rho}} \min .$$

Таким образом, алгоритмы этого класса не требуют построения сеток и использования разностных аппроксимаций, основная нагрузка ложится на процессы оптимизации. При этом в ходе оптимизации определяются как весовые коэффициенты разложения, так и параметры базисных функций (местоположения центров и «ширины»), что позволяет адаптировать базис к особенностям решения. Разумеется, ключевым моментом здесь является эффективность алгоритмов оптимизации.

Данный подход естественным образом распространяется на обратные задачи. При этом алгоритмическая структура практически не меняется, что выгодно отличает его от других известных подходов.

Дадим краткое изложение метода применительно к характерным для настоящей работы задачам.

Пусть рассматривается обратная задача в следующей постановке: требуется найти решение $\varphi(\mathbf{x})$ уравнения

$$A\varphi = f(\mathbf{x}),\ \mathbf{x} \in \Omega,$$

удовлетворяющее условию на границе:

$$B\varphi = g(\mathbf{x}),\ \mathbf{x} \in \Gamma,$$

причем функции $f(\mathbf{x})$ или $g(\mathbf{x})$ могут быть не заданы или заданы не полностью (соответственно задачи идентификации правой части и граничных условий).

При этом вместо отсутствующей информации известен дополнительный набор из m_s точечных значений искомой функции (результатов измерений), измеренные значения содержат некоторую погрешность ζ:

$$\mathrm{v}_j = \varphi(\mathbf{x}_j^{\mathrm{v}}) + \zeta, \quad j = 1, m_s .$$

Решение, как и для прямой задачи, ищется в виде разложения по адаптивному базису, состоящему из нормализованных радиальных функций. Причем используется пара разложений – первое аппроксимирует искомую функцию, а второе, в зависимости от типа задачи, приближает правую часть или граничное условие. Подбор параметров разложений ведется путем оптимизации дискретного квадратичного функционала, который получается при подстановке разложения в исходное уравнение, граничные условия и точечные (измеренные) значения функции:

$$J(\tilde{\varphi}) = \sum_{j=1}^{m_\Omega} w_\Omega \left[A(\tilde{\varphi}(\mathbf{x}_j^\Omega, \boldsymbol{\rho}_u)) - f(\mathbf{x}_j^\Omega) \right]^2 +$$

$$+ \sum_{j=1}^{m_\Gamma} w_\Gamma \left[B(\tilde{\varphi}(\mathbf{x}_j^\Gamma, \boldsymbol{\rho}_u)) - g(\mathbf{x}_j^\Gamma) \right]^2 + \sum_{j=1}^{m_s} w_s \left[\tilde{\varphi}(\mathbf{x}_j^\nu, \boldsymbol{\rho}_u) - \nu_j \right]^2 \xrightarrow{\boldsymbol{\rho}} \min,$$

где $\boldsymbol{\rho}$ – вектор параметров разложения, $\tilde{\varphi}$ – аппроксимация решения краевой задачи.

Если рассматривается задача идентификации правой части, то источниковый член $f(\mathbf{x}_j^\Omega)$ в первом слагаемом функционала заменяется аппроксимацией $\tilde{f}(\mathbf{x}_j^\Omega, \boldsymbol{\rho}_f)$, где $\boldsymbol{\rho}_f$ – вектор параметров разложения функции $f(\mathbf{x}_j^\Omega)$ по адаптивному базису нормализованных радиальных функций, определяемый в ходе оптимизации.

Аналогично, при идентификации граничного условия неизвестное слагаемое $g(\mathbf{x}_j^\Gamma)$ заменяется на $\tilde{g}(\mathbf{x}_j^\Gamma, \boldsymbol{\rho}_g)$.

В результате оптимизации функционала на выходе получаются приближения решения краевой задачи $\tilde{\varphi}(\mathbf{x})$ и идентифицируемых функций $\tilde{f}(\mathbf{x})$ или $\tilde{g}(\mathbf{x})$ (в зависимости от задачи). Аналогичным образом метод может быть применен к решению задач оптимального управления.

3.5. Примеры решения задач

В качестве примера рассмотрим гетероструктуру $Al_{0.25}GaN/GaN$ с суммарной толщиной барьерного слоя $Al_{0.25}GaN$ 30 нм. Для иллюстрации влияния легирующих добавок приведем

варианты нелегированного и равномерного легированного донорной примесью барьерного слоя. Исходные данные для расчетов следующие.

Эффективные массы электрона принимались равными $0.228\,m_0$ для обоих материалов (m_0 – масса покоя электрона). Плотность заряда на гетероинтерфейсе $Al_{0.25}GaN/GaN$, определенная из квантово-механического расчета на атомарном уровне, равна $1.08\cdot10^{13}e$ Кл/см^2. Отметим, что здесь рассматривается задача с одним каналом двумерного электронного газа, или, иными словами, с одним гетероинтерфейсом, на котором образуется нескомпенсированный заряд ($M=1$). Остальные данные определялись по аппроксимационным зависимостям [2]:

$$\varepsilon_{AlGaN}(x)=(0.03x+10.28)\varepsilon_0\,,$$

$$\Delta E_c(x)=0.7(E_g(x)-E_g(0))\,,\quad E_g(x)=6.13x+3.42(1-x)\,,$$

$$-e\varphi_b=1.3x+0.84\,,$$

где ε_0 – электрическая постоянная x – мольная доля алюминия в сплаве AlGaN. Использовалось граничное условие для потенциала $\varphi(0)=\varphi_b$.

Результаты расчетов приведены на рис. 3.1.

Здесь показаны распределения по толщине гетероструктуры потенциальной энергии (кривые 1, 2 отнесены к левой шкале) и электронной плотности (кривые 3, 4 отнесены к правой шкале). Штриховые кривые соответствуют нелегированному барьерному слою, сплошные – легированному донорами с постоянной концентрацией $2.61\cdot10^{18}$ см$^{-3}$. Отметим, что это максимально допустимая концентрация, при которой удовлетворяется ограничение (10). Видно, что во втором случае профиль потенциальной энергии приобретает характерный изгиб и повышается электронная плотность. В результате слоевая концентрация электронов в двумерном электронном газе N возрастает с $8.3\cdot10^{12}$ до $1.27\cdot10^{13}$ см$^{-2}$.

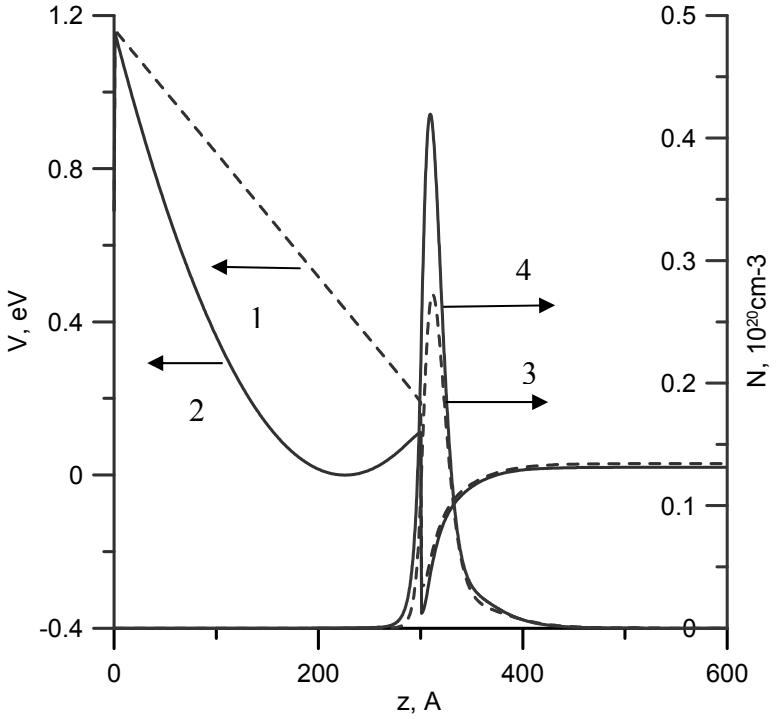


Рис. 3.1. Распределение потенциальной энергии (левая шкала) и электронной плотности (правая шкала) по толщине гетероструктуры для случаев нелегированного (штриховые кривые 1, 3) и легированного (сплошные кривые 2, 4) барьерного слоя

Перейдем к результатам решения задач оптимизации.

Пусть требуется обеспечить концентрацию электронов в ДЭГ $N^* = 1.1 \cdot 10^{13}$ см$^{-2}$. Решение будем искать в классе кусочно-постоянных функций, что соответствует технологии изготовления слоистых структур. Рассмотрим барьерный слой, состоящий из пяти подслоев. Для наглядности все свойства слоев за исключением легирования будем считать одинаковыми, включая их толщины. Наложим следующие ограничения.

$0 \le u_m \le 2.5 \cdot 10^{18}$ см$^{-3}$, $m = 1,..4,\ u_5 = 0$.

Последнее ограничение означает запрет на легирование непосредственно прилегающего к каналу ДЭГ слоя. Это довольно распространённое ограничение, которое диктуется необходимостью уменьшить рассеяние на неоднородностях сплава. Такие подслои обычно называются спейсерами. Результаты решения этой задачи представлены на рис. 3.2.

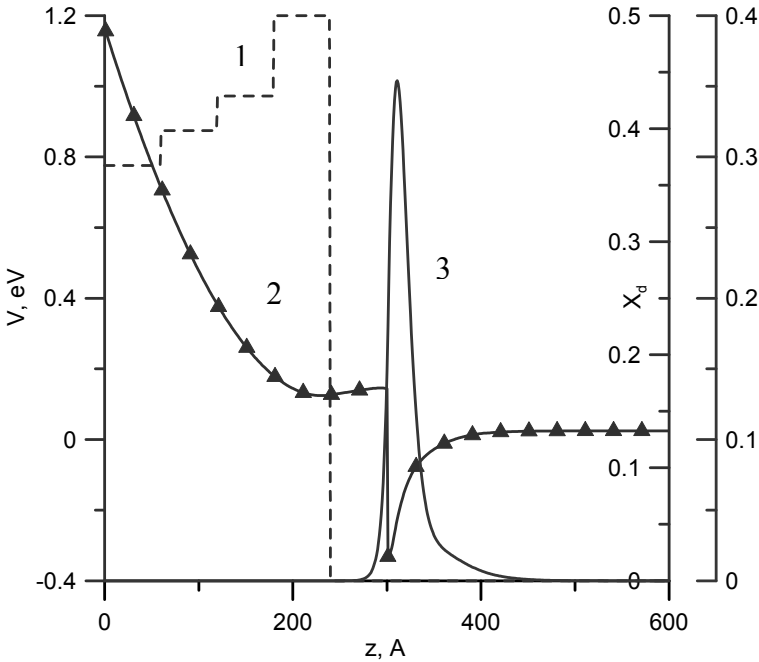


Рис. 3.2. Распределение оптимальной концентрации доноров (штриховая кривая 1, правая шкала 1), потенциальной энергии (маркированная кривая 2, левая шкала) и электронной плотности (сплошная кривая 3, правая шкала 2) по толщине гетероструктуры

Здесь штриховая кривая (правая шкала 1) иллюстрирует полученное в результате оптимизации распределение легирующей примеси по подслоям барьерного слоя $x_d(z) = n_d(z)/\overline{n}_d$ (концентрация примеси отнесена к значению $\overline{n}_d = 5 \cdot 10^{18}$ см$^{-3}$), маркиро-

ванная кривая (отнесена к левой шкале) показывает профиль потенциальной энергии, а сплошная кривая (отнесена к правой шкале 2) демонстрирует распределение электронной плотности в гетероструктуре (плотность отнесена к значению 10^{20} см$^{-3}$).

Далее рассмотрим несколько видоизменную задачу. Снимем ограничения на подслои 1–4, но при этом усилим влияние второго слагаемого в минимизируемом функционале, т.е. попытаемся достигнуть заданной концентрации электронов при минимальной суммарной ионизации барьерного слоя. Результаты решения показаны на рис. 3.3. Обозначения аналогичны рис. 3.2.

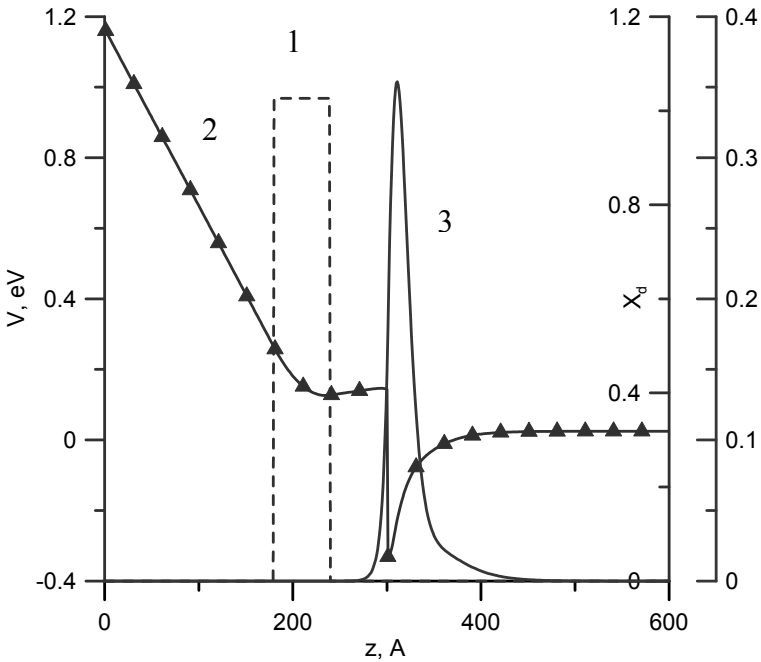


Рис. 3.3. Распределение оптимальной концентрации доноров (штриховая кривая 1, правая шкала 1), потенциальной энергии (маркированная кривая 2, левая шкала) и электронной плотности (сплошная кривая 3, правая шкала 2) по толщине гетероструктуры

Видно, что наиболее выгодным оказывается максимальное легирование слоя 3. При этом суммарная ионизация барьерного слоя U во втором варианте оказывается существенно ниже, чем при относительно равномерном легировании.

Этот результат согласуется с современной тенденцией к переходу от однородного профиля легирования к планарному δ-легированию в технологиях изготовления полевых транзисторов [1, 33].

Представленные методы математического моделирования и оптимизации могут применяться для автоматизированного проектирования наноразмерных полупроводниковых гетероструктур.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Борисенко В.Е., А.И. Воробьева, Е.А. Уткина*. Наноэлектроника. – М.: Бином. Лаборатория знаний, 2009.

2. *Ambacher O., J. Majewski, C. Miskys, A. Link, M Hermann, M. Eickhoff, M. Stutzmann, F. Bernardini, V. Fiorentini, V. Tilak, B. Schaff and L.F. Eastman*. Pyroelectric properties of Al(In)GaN/GaN hetero- and quantum well structures. // J. Phys.: Condens. Matter 14 (2002), pp. 3399–3434.

3. *Абгарян К.К., Ревизников Д.Л.* Вычислительные алгоритмы в задачах моделирования и оптимизации полупроводниковых гетероструктур. М.: МАКС Пресс, 2016, 120 с.

4. *Kohn W., L.J. Sham.* Self-consistent equations including exchange and correlation effects // Phys. Rev. 1965, V. 140, A1133.

5. *Kresse G., Furthmuller J.* Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set // Phys. Rev. 1996. B 54. 11169.

6. *Supryadkina I., K. Abgaryan, D. Bazhanov, I. Mutigullin.* Study of the Polarizations of (Al,Ga,AlGa)N Nitride Compounds and the Charge Density of Various Interfaces Based on Them // Semiconductors, 2013, Vol. 47, No. 12, pp. 1621–1625.

7. *Абгарян К.К., Д.Л. Ревизников.* Численное моделирование распределения носителей заряда в наноразмерных полупроводниковых гетероструктурах с учетом поляризационных эффектов // Журнал вычислительной математики и математической физики, 2015, т. 56, № 1, с. 153–164.

8. *Abgaryan K.K., I.V. Mutigullin, D.L. Reviznikov.* Computational model of 2DEG mobility in the AlGaN/GaN heterostructures // Physica Status Solidi (c), 2015, V. 12, No. 4–5, 460–465.

9. *Abgaryan K.K., I.V. Mutigullin, D.L. Reviznikov*. Theoretical investigation of 2DEG concentration and mobility in the AlGaN/GaN heterostructures with various Al concentrations // Physica Status Solidi (c), 2015, V. 12, No. 12, pp. 1376–1382.

10. *Абгарян К.К.* Задачи оптимизации наноразмерных полупроводниковых гетероструктур // Известия высших учебных заведений. Материалы электронной техники. 2016, Т. 19, № 2. С. 112–118.

11. *Протасов Д.Ю., Т.В. Малин, А.В. Тихонов, А.Ф. Цацульников, К.С. Журавлев*. Рассеяние электронов в гетероструктурах AlGaN/GaN с двумерным электронным газом // Физика и техника полупроводников, 2013, том 47, вып. 1, с. 36–47.

12. *Vasileska D., S.M. Goodnick, S. Goodnick.* Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation, CRC Press, 2010.

13. *Зубков В.И.* Моделирование вольт-фарадных характеристик гетероструктур с квантовыми ямами с помощью самосогласованного решения уравнений Шрёдингера и Пуассона // Физика и техника полупроводников, 2006, т. 40, вып. 10, с. 1236–1240.

14. *Trellakis A., A.T. Galick, A. Pacelli, U. Ravaioli*. Iteration scheme for solution of the two-dimensional Schrodinger-Poisson equations in quantum structures. J. Appl. Phys. 81(12), 1997, pp. 7880–7884.

15. *Trellakis A., U. Ravaioli*. Computational issues in the simulation of semiconductor quantum wires // Comput. Methods Appl. Mech. Engrg. 181 (2000), pp. 437–449.

16. *Trellakis A., A.T. Galick, A. Pacelli, U. Ravaioli*. Comparison of Iteration Schemes for the Solution of the Multidimensional Schrodinger-Poisson Equations // VLSI DESIGN, 1998, Vol. 8, Nos. (1–4), pp. 105–109.

17. *Коэн-Таннуджи К., Б. Диу, Ф. Лалоэ*. Квантовая механика. Т. 2. М.: УРСС, 2015.

18. *Федоренко Р.П.* Введение в вычислительную физику. – М.: Изд. дом Интеллект, 2008 с.

19. *Кашурников В.А., А.В. Красавин*. Численные методы квантовой статистики. М.: Физматлит, 2010.

20. *Пирумов У.Г., Гидаспов В.Ю., Иванов И.Э., Ревизников Д.Л., Стрельцов В.Ю., Формалев В.Ф*. Численные методы. Теория и практика. – М.: Изд. Юрайт, 2012.

21. *Деммель Дж.*. Вычислительная линейная алгебра. Теория и приложения. – М.: Мир, 2001.

22. *Голуб Дж., Ч. Ван Лоун*. Матричные вычисления. М.: Мир, 1999.

23. *Saad Y*. Iterative Methods for Sparse Linear Systems. PWS Publishing Company. Boston, 1996.

24. *Van der Vorst H.A*. Krylov subspace iteration // Computing in Science & Engineering. 2000, V. 2. N 1. P. 32–37.

25. *Fang F.F., V. E. Howard*. Negative Field-Effect Mobility on (100) Si Surfaces. // Phys. Rev. Lett., 16, 797 (1966).

26. *Yarar Z., B. Ozdemir, M. Ozdemir*. Mobility of electrons in a AlGaN/GaN QW: Effect of temperature, applied field, surface roughness and well width // Phys. Stat. Sol. (b) 242, 2005, No. 14, pp. 2872–2884.

27. *Ferry D.K., S.M. Goodnick, J. Bird*. Transport in Nanostructures. – Cambrige University Press, 2009.

28. *Шалимова К.В.* Физика полупроводников. – СПб.: Лань, 2010.

29. *Хокни Р., Дж. Иствуд*. Численное моделирование методом частиц. – М.: Мир, 1987.

30. *Tomizawa K.*. Numerical simulation of submicron semiconductor devices. Artech House Inc., Japan, 1993.

31. *Климова А.В., В.М. Лукашин, А.Б. Пашковский*. Поперечный пространственный перенос в полевых транзисторах на гетероструктурах с селективным легированием и границы применимости квазигидродинамических моделей. // Физика и техника полупроводников, 2009, т. 43, вып. 1, с. 113–118.

32. *Куликовский А.Г., Погорелов Н.В., Семенов А.Ю.* Математические вопросы численного решения гиперболических систем уравнений. – М.: Физматлит, 2001.

33. *Лукашин В.М., А.Б. Пашковский, К.С. Журавлев, А.И. Торопов, В.Г. Лапин, Е.И. Голант, А.А. Капралова.* Перспективы развития мощных полевых транзисторов на гетероструктурах с донорно-акцепторным легированием // Физика и техника полупроводников, 2014, Т. 48, № 5. С. 684–692.

34. *Евтушенко Ю.Г.*. Оптимизация и быстрое дифференцирование. – М.: Вычислительный центр им. А.А. Дородницина РАН, 2013.

35. *Hager W.W., H. Zhang.* A survey of nonlinear conjugate gradient methods // Pacific Journal of Optimization, 2 (2006), pp. 35–58.

36. *Hager W.W., H. Zhang.* A new conjugate gradient method with Guaranteed descent and an efficient line search // Siam J. Optim., 2005, Vol. 16, No. 1, pp. 170–192.

37. *Пантелеев А.В.* Метаэвристические алгоритмы поиска глобального экстремума. – М.: МАИ-Пресс, 2009.

38. *Vasilyev A.N., Kolbin I.S., Reviznikov D.L.* Meshfree computational algorithms based on normalized radial basis functions // Lecture Notes in Computer Science, 2016, V. 9719, pp. 583–591.

Abgaryan K. K., Reviznikov D. L.

Numerical methods in modeling of electronic properties of nanoscale heterostructures: textbook / K. K. Abgaryan, D. L. Reviznikov. – Moscow: MAKS Press, 2017. – 108 p.

ISBN 978-5-317-05705-3

The textbook presents basic mathematical models, numerical methods and algorithms used in the design and optimization of modern semiconductor nanostructures.

Methods for calculating the electron density distribution in nanostructures based on a mathematical model involving a system of joint Schrödinger and Poisson equations, methods for modeling the mobility of electrons with accounting for various scattering mechanisms, and optimization problems for multilayer nanosized heterostructures are considered.

The main attention is paid to the construction of effective computational algorithms.

The presented methods and tools of mathematical modeling open up prospects for solving problems that are relevant for the development of modern nanoelectronics.

The manual is intended for graduate and post-graduate students of technical universities, researchers specializing in applied mathematics, computational electronics and nanostructured materials science.

Keywords: numerical methods and algorithms, mathematical modeling, optimization, electron density, electron mobility, computational algorithms.

Учебное издание

АБГАРЯН Каринэ Карленовна
РЕВИЗНИКОВ Дмитрий Леонидович

ЧИСЛЕННЫЕ МЕТОДЫ
В МОДЕЛИРОВАНИИ ЭЛЕКТРОННЫХ СВОЙСТВ
НАНОРАЗМЕРНЫХ ГЕТЕРОСТРУКТУР

Учебное пособие

Подготовка оригинал-макета:
Издательство «МАКС Пресс»
Главный редактор: *Е. М. Бугачева*
Компьютерная верстка: *Н. С. Давыдова*
Обложка: *М. А. Еронина*

Подписано в печать 22.11.2017 г.
Формат 60x90 1/16. Усл.печ.л. 6,75. Тираж 500 экз. Заказ 282.

Издательство ООО “МАКС Пресс”.
Лицензия ИД N 00510 от 01.12.99 г.
119992, ГСП-2, Москва, Ленинские горы,
МГУ им. М.В. Ломоносова, 2-й учебный корпус, 527 к.
Тел. 8(495)939-3890/91. Тел./Факс 8(495)939-3891.

Отпечатано в ППП «Типография «Наука»
121099, Москва, Шубинский пер., 6
Заказ №