

Нижегородский государственный университет им. Н.И. Лобачевского
Национальный исследовательский университет
Программа повышение конкурентоспособности ННГУ им. Н.И. Лобачевского
Стратегическая инициатива 7 «Достижение лидирующих позиций в области суперкомпьютерных технологий и высокопроизводительных вычислений»

Основные образовательные программы
Радиофизика, Информационные технологии,
Информационная безопасность телекоммуникационных систем

Учебно-методическая разработка по дисциплинам
Физическая электроника, Электроника и схемотехника

Е.В.Волкова, С.В.Оболенский

ФИЗИКА ОПТОЭЛЕКТРОННЫХ, СВЧ И КВЧ ДИОДОВ
И ОРГАНИЗАЦИЯ ИХ ПРОЕКТИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ
ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ

Нижний Новгород,
2014 год

УДК 53.082, 538.95

ББК 31.85

В67

В67 Волкова Е.В., Оболенский С.В. Физика оптоэлектронных, СВЧ и КВЧ диодов и организация их проектирования с использованием высокопроизводительных вычислений: Учебное пособие. – Нижний Новгород: Нижегородский госуниверситет, 2014. – 83 с.

Данное пособие является продолжением цикла учебных пособий по полупроводниковой электронике и содержит информацию о физических основах работы полупроводниковых диодов и методах организации их проектирования с использованием высокопроизводительных вычислений.

Пособие предназначено для студентов дневного и вечернего отделений, изучающих курсы «Электроника и схемотехника» и «Физическая электроника».

УДК 53.082, 538.95

ББК 31.85

© Нижегородский государственный
университет им. Н.И. Лобачевского, 2014

СОДЕРЖАНИЕ

Часть 1. Оптоэлектронные диоды	4
1.1. Фотодиоды	4
1.2. Солнечные батареи	12
1.3. Светодиоды и лазеры	17
Часть 2. СВЧ и КВЧ диоды	30
2.1. Диод Ганна	30
2.2. Лавинно-пролетные и инжекционно-пролетные диоды	40
2.1.1. Лавинно-пролетные диоды	40
2.2.2. Инжекционно-пролетные диоды	43
2.3. Диоды на основе эффекта туннелирования. Сверхрешетки	45
2.3.1. Туннельный диод	45
2.3.2. Туннельно-резонансный диод	50
2.3.3. Сверхрешётки	52
Часть 3. Организация проектирования диодов с использованием высокопроизводительных вычислений	56
3.1. Методы моделирования физических процессов в диодах	56
3.2. Организация вычислений	65
3.3. Описание вычислительного алгоритма	69
3.4. Тестирование алгоритма	72
3.5. Анализ производительности алгоритма	74
Список литературы	80

ЧАСТЬ 1

ОПТОЭЛЕКТРОННЫЕ ДИОДЫ

Первая часть учебного пособия посвящена изучению принципов работы базовых диодов полупроводниковой оптоэлектроники. Рассматриваются три основные группы приборов: а) фотодетекторы, т.е. устройства, преобразующие оптические сигналы в электрические за счет протекающих под действием света электронных процессов; б) фотовольтаические приборы, т.е. устройства, преобразующие энергию излучения в электрическую энергию; в) светоизлучающие приборы, т.е. устройства, преобразующие электрическую энергию в оптическое излучение.

1.1. Фотодиоды

Объяснение принципов работы фотодиодов удобно начать с рассмотрения наиболее простого фотодетектора – *фоторезистора*. Он представляет собой пластину полупроводника с омическими контактами, присоединенными к противоположным концам пластины, как показано на рис. 1.1. Принцип работы фотоприемников (в том числе, и фоторезисторов) основывается на явлении *внутреннего фотоэффекта* или *фотопроводимости*: при падении излучения с энергией квантов больше ширины запрещенной зоны на поверхность полупроводника в нем генерируются электронно-дырочные пары, т.е. увеличивается число свободных носителей заряда (собственное возбуждение). Возможно также использование переходов с участием уровней, лежащих в запрещенной зоне (примесное возбуждение). В обоих случаях в соответствии с формулой $\sigma = e(\mu_n n + \mu_p p)$, где n , p – концентрации электронов и дырок, μ_n , μ_p – их подвижности, e – заряд электрона, облучение приводит к увеличению проводимости материала, т.е. появлению фототока, измерение которого позволяет детектировать падающее излучение.

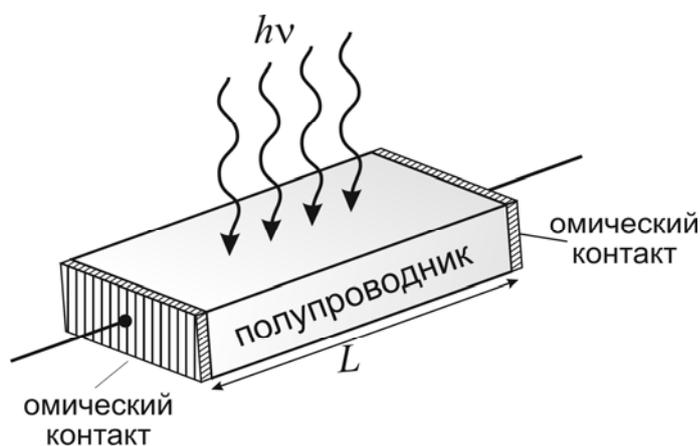


Рис. 1.1. Схематическое изображение фоторезистора

Существенным недостатком фотосопротивлений является значительная величина так называемого темнового тока, т.е. тока, который протекает в структуре в отсутствие излучения. Поэтому на практике в качестве фотоприемников обычно используются фотодиоды, в которых данная проблема снимается. Семейство фотодиодов с обедненным слоем включает в себя $p-n$ переходы, $p-i-n$ диоды, диоды, образованные контактом металл - полупроводник (диоды с барьером Шоттки), гетеропереходы и точечно-контактные диоды.

В простейшем случае фотодиод представляет собой $p-n$ переход, к которому приложено небольшое (по сравнению с напряжением пробоя) обратное смещение¹. Как известно, ток обратносмещенного $p-n$ перехода определяется неосновными носителями заряда, а, значит, в отсутствие излучения ток фотодиода мал. Поскольку толщина ОПЗ $p-n$ перехода достаточно мала, то на практике обычно светят в торец прибора, как показано на рис. 1.2 а. Фотоны поглощаются в области пространственного заряда (ОПЗ) диода и в нейтральном объеме полупроводника. Электронно-дырочные пары, фотогенерированные в ОПЗ перехода, немедленно разделяются его полем и дают вклад в суммарный ток, что позволяет детектировать излучение (рис. 1.2 б). При этом носители, которые генерируются вне ОПЗ, попадают туда за счет диффузии. Область диффузии в этом случае должна быть достаточно тонкой (много меньше диффузионной длины), чтобы основная часть генерированных электронно-дырочных пар без рекомбинации достигала области перехода.

Одной из основных характеристик любого фотодетектора, в том числе, и фотодиода, является *область спектральной чувствительности* – диапазон

¹ Отметим, что в случае так называемых лавинных фотодиодов прикладывается высокое напряжение. Такие приборы будут рассматриваться ниже.

длин волн, в котором может быть генерирован заметный фототок. Одним из основных факторов, определяющих эту область, является коэффициент поглощения α . На рис. 1.3 представлены экспериментальные кривые зависимости коэффициента поглощения от длины волны в собственной области для германия и кремния [3]. Кроме того, указаны длины волн, соответствующие излучению некоторых распространенных лазеров.

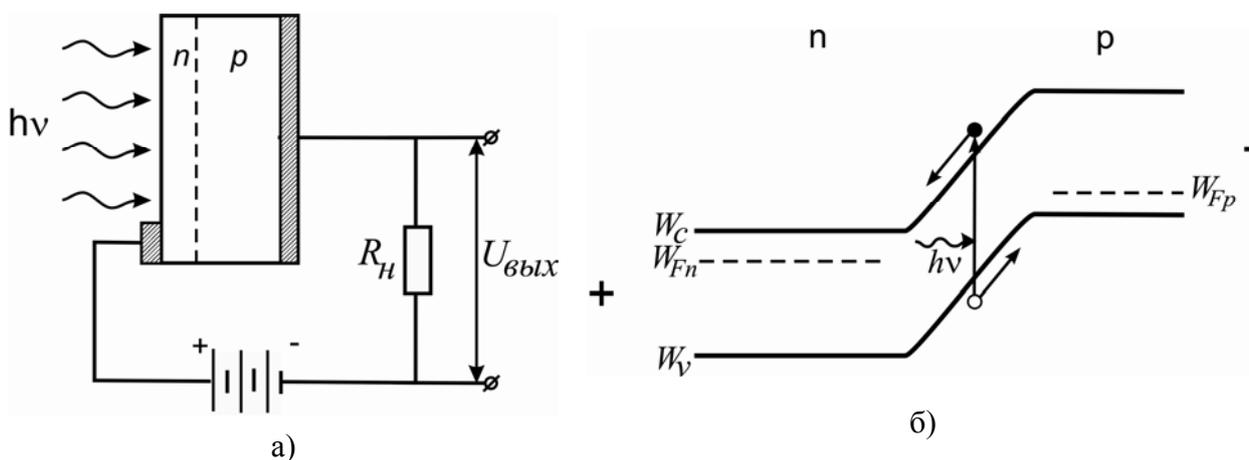


Рис. 1.2. Схематическое изображение фотодиода с подключением (а) и зонной диаграммы прибора при рабочем смещении (б). R_n – сопротивление нагрузки; $U_{вых}$ – напряжение выхода

Вследствие того, что α очень резко зависит от длины волны, для данного полупроводника область длин волн, в которой может быть генерирован заметный фототок, оказывается ограниченной с обеих сторон. Поскольку межзонная генерация электронно-дырочных пар возможна, когда энергия квантов падающего излучения больше ширины запрещенной зоны, существует так называемая *длинноволновая граница фоточувствительности (красная граница фотоэффекта)*:

$$\lambda_c = \frac{hc}{W_g} = \frac{1.24}{W_g[\text{эВ}]}, \text{ мкм.} \quad (1.1)$$

Здесь λ_c - длина волны, соответствующая ширине запрещенной зоны полупроводника W_g , c – скорость света, h – постоянная Планка. Излучение с длиной волны $\lambda < \lambda_c$ поглощается в полупроводнике с образованием пар электрон-дырка. При $\lambda > \lambda_c$ сравнительно небольшая по величине фотопроводимость может возникать в случае примесного поглощения фотона, т.е. при переходе электрона с уровня в запрещенной зоне в зону проводимости (или дырки в

валентную зону). Коротковолновая граница спектральной чувствительности обусловлена тем, что в этой области длин волн коэффициент α очень велик (около 10^5 см^{-1}), т.е. излучение поглощается очень близко к поверхности, где времена рекомбинации малы из-за наличия дефектов. Вследствие этого фотоносители рекомбинируют раньше, чем уходят из поверхностного слоя к области пространственного заряда. На рис. 1.4 представлены типичные кривые зависимости квантовой эффективности от длины волны для кремниевого и германиевого высокочастотных фотодиодов [3].

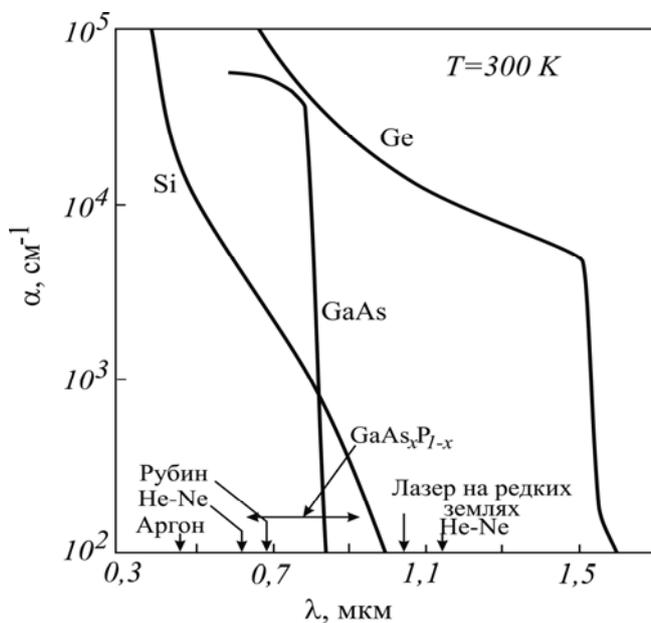


Рис. 1.3. Зависимость коэффициента поглощения от длины волны для Ge и Si при 300 К [3]. Указаны линии излучения некоторых лазеров

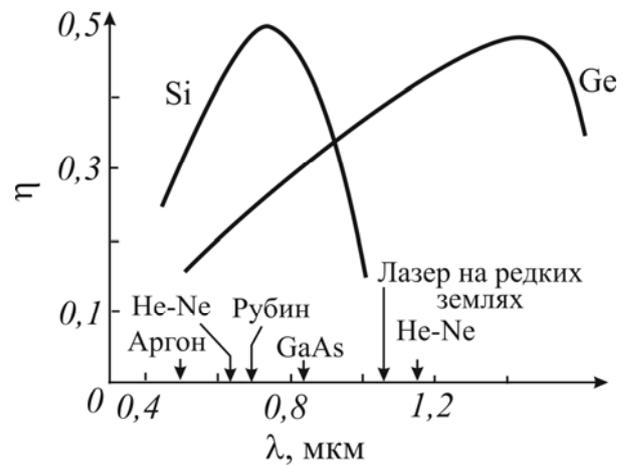


Рис. 1.4. Зависимость эффективного квантового выхода η (количества электронно-дырочных пар на один фотон) от длины волны λ для Ge и Si фоторезисторов [3]

Еще одним важным параметром фотодетектора является его время отклика. Быстродействие фотодиодов определяется тремя факторами: временем диффузии носителей от места генерации к области пространственного заряда, временем дрейфа через обедненный слой и его емкостью. Чтобы свести к минимуму влияние диффузии, переход необходимо формировать очень близко к поверхности. Далее, чтобы основная часть излучения поглощалась непосредственно в ОПЗ, а свет не проникал бы вглубь за ее пределы, толщина области

перехода должна быть порядка $1/\alpha$. При этом носители будут дрейфовать в ОПЗ со скоростями, ограниченными их рассеянием. Однако, область пространственного заряда не должна быть слишком широкой, так как времена переноса будут ограничивать предельную частоту работы прибора. С другой стороны, она не должна быть и слишком узкой, так как в этом случае становится большой емкостью p - n перехода, а значит, и постоянная времени RC , где R - сопротивление базы диода. Оптимальным значением ширины области пространственного заряда считается такое, при котором время переноса примерно равно половине периода модуляции излучения. Например, для частоты модуляции, равной 10 ГГц , оптимальная ширина области пространственного заряда в кремнии (при скорости дрейфа порядка 10^7 см/сек) составляет около 5 мкм .

Типичным детектором с обедненным слоем является фотодиод со структурой p - i - n , где между p и n слоями заключена i -область, т.е. слой собственного полупроводника. Структура и зонная диаграмма такого фотодиода [3] показаны на рис. 1.5.

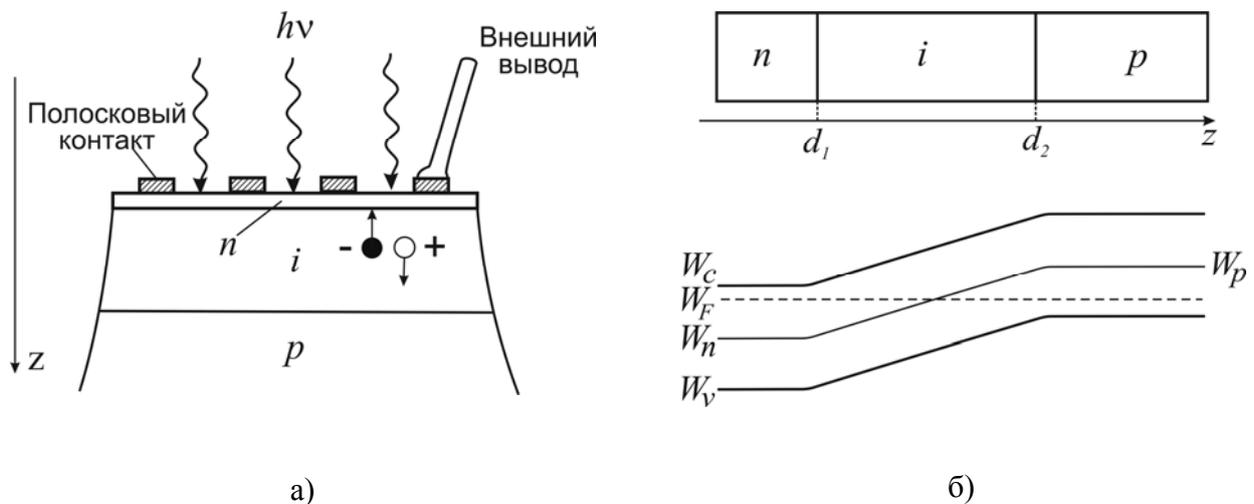


Рис. 1.5. Фотодиод с p - i - n структурой [3]: а) конструкция прибора; б) упрощенная структура и зонная диаграмма вдоль оси z в равновесном случае. Полосковые контакты объединены друг с другом – обычно это тонкая металлическая сетка на поверхности полупроводника

Зависимость потенциальной энергии электронов от координаты $W(z)$ определяется уравнением Пуассона, которое для области обеднения имеет вид:

$$\begin{cases} \frac{d^2W}{dz^2} = \frac{e^2}{\varepsilon\varepsilon_0} N_D, & d_1 - d_n < z < d_1; \\ \frac{d^2W}{dz^2} = 0, & d_1 \leq z \leq d_2; \\ \frac{d^2W}{dz^2} = -\frac{e^2}{\varepsilon\varepsilon_0} N_A, & d_2 < z < d_2 + d_p. \end{cases} \quad (1.2)$$

Здесь e – заряд электрона; N_D, N_A – соответственно, концентрации доноров в n и акцепторов в p областях; ε – диэлектрическая проницаемость материала; d_1, d_2 – координаты границ области собственной проводимости (рис. 1.5 б); d_n, d_p – толщины областей пространственного заряда в n и p областях, соответственно, причем, из закона сохранения заряда следует, что $N_D \cdot d_n = N_A \cdot d_p$. Если отсчитывать энергию от ее величины в n -области, то решение (1.2) будет иметь вид:

$$W(z) = \begin{cases} \frac{e^2 N_D}{2\varepsilon\varepsilon_0} (z - d_1 + d_n)^2, & d_1 - d_n < z < d_1; \\ \frac{e^2 N_D}{2\varepsilon\varepsilon_0} d_n^2 + \frac{e^2 N_D}{\varepsilon\varepsilon_0} d_n (z - d_1), & d_1 \leq z \leq d_2; \\ W_\kappa - \frac{e^2 N_A}{2\varepsilon\varepsilon_0} (d_2 + d_p - z)^2, & d_2 < z < d_2 + d_p. \end{cases} \quad (1.3)$$

Здесь W_κ – энергия, соответствующая контактной разности потенциалов U_κ . В пределе, когда $d_2 - d_1 \gg d_n, d_p$, решение имеет простой вид:

$$W(z) = \begin{cases} 0, & z < d_1; \\ \frac{W_\kappa}{d_2 - d_1} (z - d_1), & d_1 \leq z \leq d_2; \\ W_\kappa, & d_2 < z. \end{cases} \quad (1.4)$$

Таким образом, мы получили линейный закон изменения потенциала между двумя легированными областями n и p -типа. То есть подобная система ведет себя, как простой конденсатор, при этом в области собственной проводимости

мости постоянное электрическое поле определяется простым соотношением $E=eU_{к}/(d_2-d_1)$.

P-i-n диоды характеризуются быстрым фотооткликом и достаточно высокой чувствительностью. Это объясняется тем, что емкость такой структуры значительно меньше емкости *p-n* перехода, а значит, меньше и постоянная времени *RC*. Кроме того, толщина *i*-слоя может быть сделана достаточно большой для того, чтобы поглотить практически все излучение, т.е. поглощение в нейтральных областях сводится к минимуму.

Фотодиоды с барьером Шоттки также могут быть использованы в качестве высокоэффективных фотодетекторов [3]. Типичный пример приведен на рисунке 1.6.

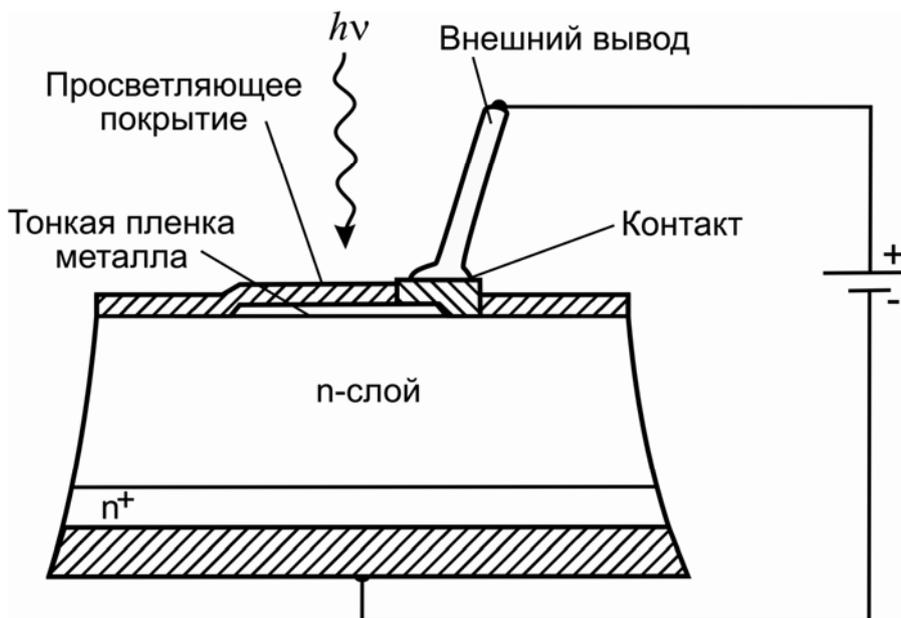


Рис. 1.6. Диод с барьером Шоттки и с просветляющим покрытием [3]

Для того, чтобы исключить большие потери на отражение и поглощение при освещении диода через металлический контакт, слой металла делают очень тонким ($\approx 100 \text{ \AA}$) и применяют просветляющие покрытия [2].

Фотодиоды с обедненным слоем изготавливаются на основе не только гомо-, но и гетероструктур, в которых переход образован между двумя полупроводниками с различной шириной запрещенной зоны. На рис. 1.7 схематически показан фотодиод, содержащий переход между *n-GaAs* и *p-Ge*, предназначенный для высокочастотного детектирования излучения с длиной волны $\lambda = 8450 \text{ \AA}$, испускаемого источником инфракрасного излучения, изготовленным из *GaAs* [3]. На этой длине волны коэффициент поглощения излучения

равен 10 см^{-1} в GaAs и $2.2 \cdot 10^4 \text{ см}^{-1}$ в Ge. Таким образом, *основная доля излучения поглощается в германии в слое толщиной 1 мкм, т.е. там, где расположена обедненная область перехода*. В связи с тем, что носители создаются около перехода, а не близ поверхности, диоды имеют высокое быстродействие и нечувствительны к состоянию поверхности. Граничная частота диода слабо зависит от расстояния от поверхности до перехода, но более сильно зависит от того, какая часть излучения поглощается в той и другой частях обедненной области, образующей переход. Энергетический барьер на зонной диаграмме определяется разрывом дна зоны проводимости на границе раздела между GaAs и Ge. Преодолеть этот барьер носители могут либо за счет туннельного механизма, либо за счет надбарьерной эмиссии, имея достаточно большую энергию.

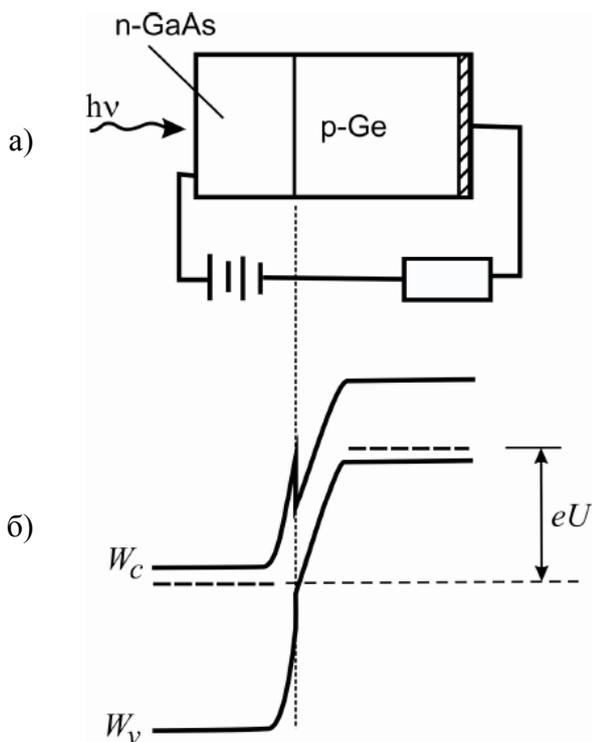


Рис. 1.7. Схематическое изображение фотодиода с гетеропереходом при приложении обратного смещения (а) и диаграмма энергетических зон диода (б)

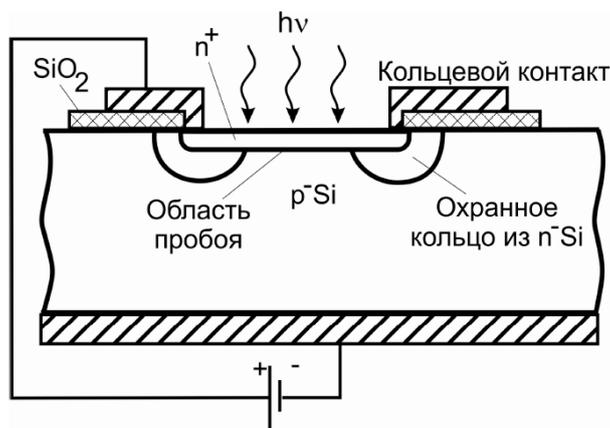
Еще одним примером полупроводникового фотоприемника является *лавинный фотодиод*, т.е. диод, на который подаются большие (достаточные для реализации пробоя) обратные смещения. В такой ситуации фотовозбужденные электроны и дырки участвуют в формировании лавины свободных носителей заряда за счет процесса ударной ионизации полупроводника в сильном электрическом поле. Таким образом, с одного поглощенного фотона в лавинном фотодиоде образуется много свободных носителей за счет процессов вторичной генерации, т.е. прибор является фотодетектором с внутренним усилением тока.

Минусом таких приборов является то, что лавинный процесс вносит дополнительные источники шума. Кроме того, появляются повышенные требования к термостабильности (лавинный пробой не должен переходить в термический).

Основным требованием для получения высокого усиления по току является исключение микроплазм или участков малой площади, пробивное напряжение которых ниже, чем пробивное напряжение всего перехода. Кроме того, для получения однородного по площади умножения необходимо исключить пробой у краев перехода. На рис. 1.8 показана типичная структура лавинного диода с охранным кольцом [3]. Градиент примеси в охранным кольце ниже, а радиус кривизны больше, поэтому центральная активная область пробивается раньше, чем периферийная (область охранным кольца).

Толщина и уровень легирования слоев диода выбираются так, чтобы область пространственного заряда перехода имела величину порядка длины волны излучения. В этом случае весь свет будет поглощаться в самом переходе, что приведет к максимальному увеличению тока, то есть реализации максимального КПД.

Рис. 1.8. Поперечное сечение планарного фотодиода с охранным кольцом



1.2 СОЛНЕЧНЫЕ БАТАРЕИ

Солнечная батарея (фотоэлемент) представляет собой $p\text{-}n$ переход большой площади, в котором энергия фотонов непосредственно преобразуется в электрическую энергию. Высокоэффективные солнечные батареи были впервые созданы в 1954 г. При этом были использованы диффузионные кремниевые $p\text{-}n$ переходы. В настоящее время солнечные батареи изготавливаются из различных полупроводниковых материалов (Si , GaAs , CdS). Подобные

приборы являются наиболее важными источниками питания на спутниках и космических кораблях.

Преобразование мощности излучения в электрическую происходит за счет *фотовольтаического эффекта* следующим образом. При попадании излучения в области пространственного заряда генерируются электронно-дырочные пары. Встроенное электрическое поле перехода будет перемещать электроны и дырки в противоположных направлениях. В результате область перехода *n*-типа зарядится отрицательно, а область *p*-типа – положительно. Возникающая разность потенциалов называется *фото-ЭДС* и направлена навстречу контактной разности потенциалов, т.е. снижает потенциальный барьер перехода. При подключении нагрузки параллельно освещаемому переходу в цепи будет протекать постоянный ток до тех пор, пока не прекратится излучение. Таким образом, работа солнечных батарей основана на тех же физических процессах, что и функционирование фотодиода: в обоих приборах используется разделение фотогенерированных электронов и дырок встроенным электрическим полем перехода.

Природа источника падающего излучения не имеет значения для последующего анализа; это может быть источник фотонов (солнечная энергия, γ -излучение, лампа накаливания, рентгеновское излучение и т. д.), источник частиц с высокой энергией (электронная пушка, β - излучение, α - частицы, протоны, нейтроны и т. п.) или любой другой источник, создающий электронно-дырочные пары, но не изменяющий заметно свойства идеального перехода.

Солнечная батарея, показанная на рис. 1.9, состоит из неглубокого *p-n* перехода, верхнего полоскового омического контакта и нижнего омического контакта, который полностью покрывает всю тыльную поверхность полупроводникового кристалла. Простейшая эквивалентная схема солнечной батареи при ее освещении показана на врезке рис. 1.10.

Вольт-амперная характеристика такого прибора может быть представлена в виде

$$I = I_s \left(e^{eU/kT} - 1 \right) - I_\phi, \quad (1.5)$$

$$I_s = en_i^2 \left[\frac{1}{N_A} \left(\frac{D_n}{\tau_n} \right)^{1/2} + \frac{1}{N_D} \left(\frac{D_p}{\tau_p} \right)^{1/2} \right] \quad (1.6)$$

Здесь, помимо введенных выше стандартных обозначений, I_ϕ – ток, протекающий в структуре за счет генерации электронно-дырочных пар при освещении, т.е. ток в режиме короткого замыкания (сопротивление нагрузки $R=0$). На рис. 1.10 изображена ВАХ, соответствующая значениям $I_\phi=100$ мА, $I_s=10^{-9}$ А и $T=300$ К [3]. Кривая проходит через четвертый квадрант, следовательно, прибор может отдавать мощность во внешнюю цепь. ВАХ позволяет выбрать оптимальный режим работы фотоэлемента, т.е. оптимальное значение сопротивления нагрузки, которому соответствует наибольшая площадь прямоугольника с вершиной, расположенной на ВАХ, при данном значении светового потока Φ ($\frac{\text{фотонов}}{с \cdot \text{см}^2}$). КПД солнечных батарей, изготовленных из Si и GaAs, составляет 10-15%, а на основе тройных соединений с использованием гетеропереходов 20-50%. Основной вклад в уменьшение КПД дают потери на отражение, утечки и влияние сопротивления пассивных слоев полупроводника.

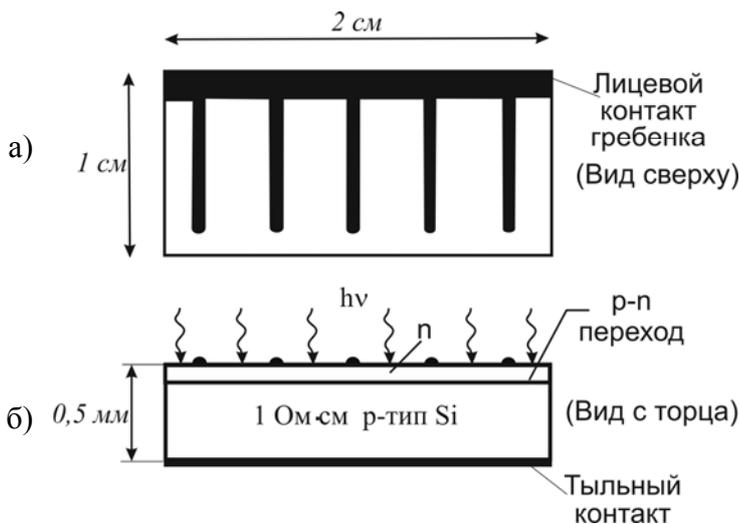


Рис. 1.9. Пример конструкции солнечной батареи:
а) вид сверху; б) вид с торца

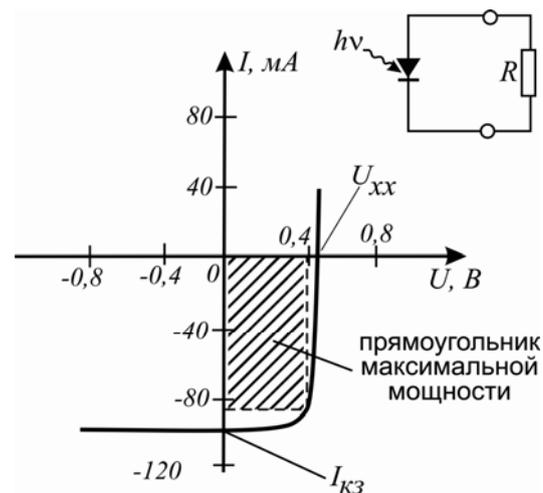
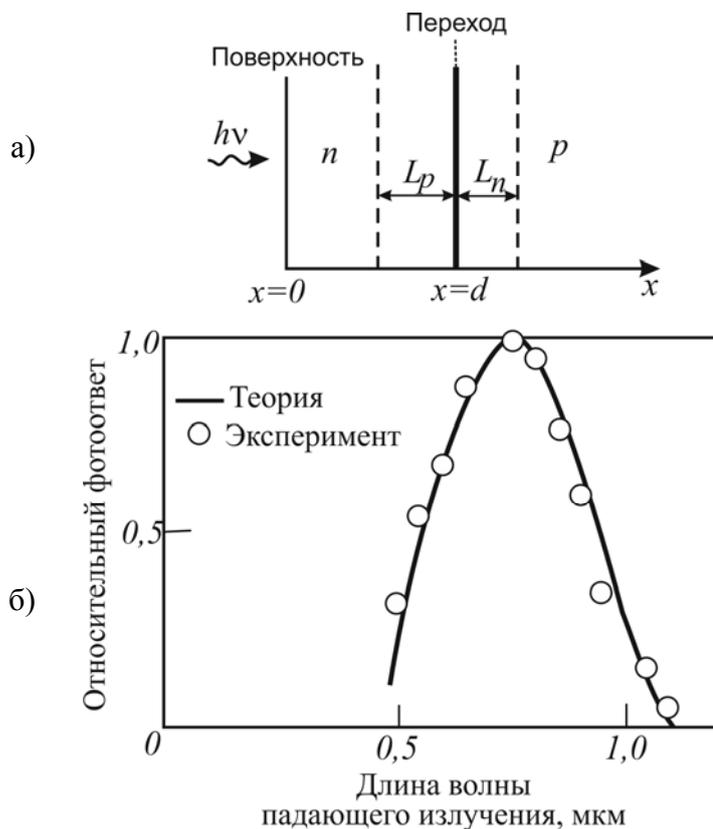


Рис. 1.10. ВАХ солнечной батареи с сопротивлением нагрузки R . На рисунке отмечены напряжение холостого хода $U_{ХХ}$ и ток короткого замыкания $I_{КЗ}$

Для оценки величины тока короткого $I_{КЗ}$ замыкания и анализа спектрального распределения фоточувствительности солнечной батареи (т.е. зависимости $I_{КЗ}(\lambda)$) удобно воспользоваться простой одномерной моделью, изображенной на рис. 1.11 а [3].

Рис. 1.11. Спектральная чувствительность тока солнечной батареи: а) - одномерное изображение солнечной батареи с глубиной залегания перехода d и диффузионной длиной неосновных носителей L_n и L_p для p - и n -областей, соответственно; б) сравнение типичных экспериментальных и расчетных кривых фотоответа для p - на n -солнечной батареи. Экспериментальные точки соответствуют данным, полученным при $d=2.0$ мкм. Сплошная кривая представляет собой расчетные результаты, полученные для $d=2.0$ мкм, $L_n=0.5$ мкм и $L_p=10.0$ мкм [3]



Здесь d – глубина залегания перехода, а L_n и L_p диффузионные длины неосновных носителей заряда в p - и n -областях, соответственно. Предполагается, что ширина обедненного слоя значительно меньше L_n и L_p .

Если энергия квантов излучения превышает ширину запрещенной зоны полупроводника ($h\nu > W_g$), то за счет поглощения плотность потока фотонов Φ в материале изменяется в направлении x вглубь образца: $\Phi(x) = \Phi_0 \cdot \exp(-\alpha x)$, где Φ_0 – плотность падающего потока фотонов, α - коэффициент поглощения, зависящий от длины волны излучения (рис. 1.3). Тогда скорость поглощения фотонов в материале составит $\alpha \Phi_0 \cdot \exp(-\alpha x)$. Если при поглощении одного фотона рождается одна электронно-дырочная пара, то скорость их генерации (т.е. количество пар носителей, генерируемых в единице объема полупроводника за единицу времени) выражается формулой

$$G(x) = \Phi_0 \alpha e^{-\alpha x}. \quad (1.7)$$

Генерированные вне области перехода носители движутся к нему только за счет диффузии, при этом часть носителей рекомбинирует, не успевая до-

браться до области сильного поля. В n -области из общего числа неосновных носителей (дырок), генерированных на расстоянии x от поверхности, перехода достигает часть носителей, пропорциональная $\exp\left[-\frac{(d-x)}{L_p}\right]$. Тогда полное число неосновных носителей, попадающих на единицу площади сечения перехода за единицу времени и появившихся за счет генерации электронно-дырочных пар в n -области, равно:

$$N_p \approx \int_0^d \Phi_0 \alpha \exp(-\alpha x) \exp\left[-\frac{|d-x|}{L_p}\right] dx = \frac{\Phi_0 \alpha}{\alpha - \frac{1}{L_p}} \left[\exp(-d/L_p) - \exp(-\alpha d) \right]. \quad (1.8)$$

Аналогично, число неосновных носителей (электронов), достигающих перехода и обусловленных генерацией электронно-дырочных пар в p -области, равно интегралу (1.8), взятому в пределах от $x=d$ до $x=\infty$ при замене L_p на L_n :

$$N_n \approx \int_d^{+\infty} \Phi_0 \alpha \exp(-\alpha x) \exp\left[-\frac{|d-x|}{L_n}\right] dx = \frac{\Phi_0 \alpha \exp(-\alpha d)}{\alpha + \frac{1}{L_n}}. \quad (1.9)$$

Тогда полное число носителей, проходящих через единичное сечение p - n перехода в единицу времени, равно:

$$N \approx \frac{\Phi_0 \alpha}{\alpha - \frac{1}{L_p}} \left[\exp(-d/L_p) - \exp(-\alpha \cdot d) \right] + \frac{\Phi_0 \alpha \cdot \exp(-\alpha \cdot d)}{\left(\alpha + \frac{1}{L_n} \right)}. \quad (1.10)$$

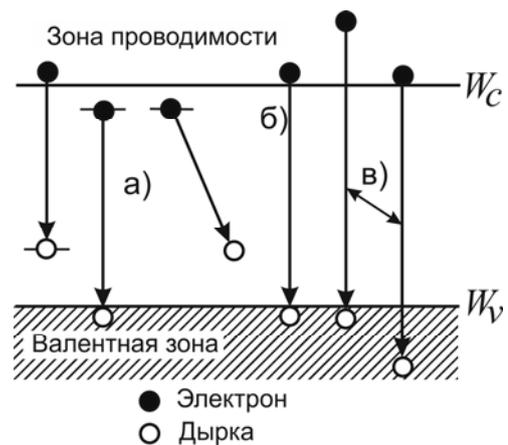
Таким образом, в стационарных условиях через структуру протекает постоянный ток, плотность которого $J_{кз} = eN$. Отметим, что выражение (1.10) получено для солнечных батарей, в которых освещаемая область имеет проводимость n -типа. Очевидно, что при освещении со стороны p -слоя необходимо лишь поменять местами L_p и L_n .

Полученное выражение позволяет также найти спектральное распределение фоточувствительности батареи, однако, при этом необходимо помнить, что, поскольку энергия одного фотона $h\nu$ пропорциональна частоте, то при одной и той же энергии излучения плотность потока фотонов Φ_0 пропорциональна длине волны λ . Рис. 1.11 б иллюстрирует хорошее совпадение расчетных и экспериментальных данных.

1.3 СВЕТОДИОДЫ И ЛАЗЕРЫ

Полупроводниковые светодиоды и лазеры представляют собой оптоэлектронные приборы, преобразующие электрическую энергию в энергию оптического излучения. В классических устройствах такого типа используются межзонные переходы электронов (рис. 1.12), т.е., в основном, генерация фотонов в материале происходит при излучательной рекомбинации электронов и дырок². В такой ситуации энергия излучаемых квантов, как видно из рисунка, определяется шириной запрещенной зоны кристалла: $h\nu > W_g$. В настоящее время разнообразие полупроводниковых материалов позволяет перекрыть широкий спектральный диапазон излучения: от ближнего ультрафиолета до среднего ИК диапазона.

Рис. 1.12. Основные виды переходов в полупроводнике [3]: а) переходы с участием примесей или дефектов (вакансии решетки и т. п.): между зоной проводимости и акцепторным уровнем, между донорным уровнем и валентной зоной, между донорным и акцепторным уровнями (межпримесное излучение); б) межзонные переходы: собственное излучение, энергетически очень близкое к ширине запрещенной зоны; в) излучение с более высокой энергией с участием «горячих» носителей, т.е. носителей, обладающих высокой энергией



Сразу отметим, что не все полупроводниковые материалы эффективно испускают свет при межзонной рекомбинации. Для объяснения данного факта необходимо вспомнить, что волновые числа фотонов в оптическом диапазоне составляют по порядку величины $k_{opt} = 2\pi/\lambda \sim 10^4 \dots 10^6 \text{ см}^{-1}$, а типичные значения волновых чисел носителей заряда в зоне Бриллюэна – $k_e = 2\pi/a \sim 10^8 \text{ см}^{-1}$, где a – постоянная решетки материала. Таким образом, волновой вектор излучения является пренебрежимо малым по сравнению с тем же параметром для

² Хотя в последнее время все большее распространение получают лазеры на внутризонных переходах.

электронов валентной зоны и зоны проводимости. То есть оптические межзонаные переходы электронов происходят практически без изменения волнового числа (импульса). В этом смысле такие переходы называют *прямыми* или *вертикальными* (на зонной диаграмме начальное и конечное состояния электрона находятся на одной вертикали – рис. 1.13).

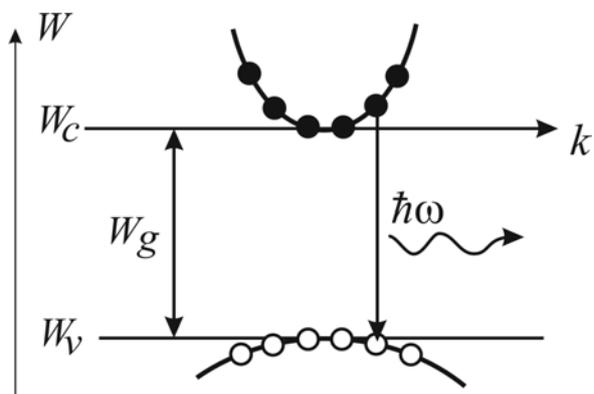


Рис. 1.13. Схематическое изображение вертикального перехода на зонной диаграмме полупроводника

Как известно [7], по структуре энергетических зон полупроводники разделяют на *прямозонные* и *непрямозонные*. К первым относятся кристаллы, у которых нижний минимум зоны проводимости и верхний максимум валентной зоны находятся в одной и той же точке зоны Бриллюэна (т.е. им соответствует одно и то же значение волнового числа). В непрямозонных же полупроводниках экстремумам зон соответствуют различные волновые числа. Из прямозонных полупроводников наиболее широко используются арсенид галлия и фосфид индия. Самыми распространенными непрямозонными материалами являются кремний и германий. Важно, что в прямозонных полупроводниках возможны вертикальные переходы электронов, поэтому коэффициент поглощения и скорость излучательной рекомбинации в таких веществах высоки. В полупроводниках непрямозонных оптические переходы также возможны (рис. 1.14), но законы сохранения энергии и импульса в этом случае требуют, чтобы в таких процессах принимали участие не только фотоны и электроны, но и центры рассеяния (например, фононы), которые могли бы обеспечить существенное изменение импульса носителей. Поскольку в данном случае необходимо участие многих частиц, то, очевидно, вероятность подобных *непрямых переходов* будет ниже. Это отражается на зависимости коэффициента поглощения от длины волны для GaAs, Si и Ge, которую мы приводили на рис. 1.3. Видно, что у непрямозонных Si и Ge вблизи границы фотоэффекта коэффици-

ент поглощения сравнительно мал, поскольку он определяется маловероятными непрямыми переходами. С увеличением энергии фотонов (с уменьшением длины волны излучения) становятся возможны переходы непосредственно в верхние долины, поэтому коэффициент поглощения растет. В то же время, у прямозонного GaAs наблюдается резкий рост коэффициента поглощения практически сразу при переходе через границу фотоэффекта. В итоге, вероятность межзонных излучательных переходов в непрямозонных полупроводниках значительно ниже, чем в материалах с прямой зоной. В связи с этим в настоящее время для изготовления высококачественных светодиодов и лазеров используют, в основном, прямозонные полупроводники.

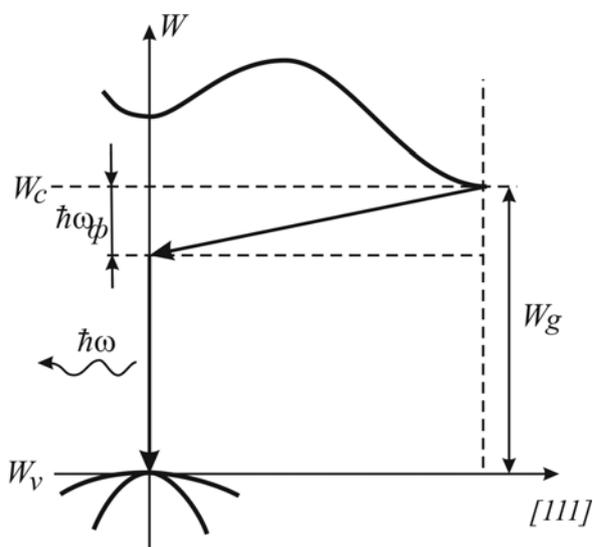


Рис. 1.14. Оптический межзонный переход с участием фона в непрямозонном полупроводнике.

Приведенные выше рассуждения справедливы, однако, лишь для межзонных переходов в чистых полупроводниках. При межподзонных переходах в гетероструктурах или при межпримесных переходах в сильнолегированных полупроводниках ситуация меняется. Сейчас ведутся активные поиски возможностей организации лазерного излучения на непрямозонных материалах. Последнее необходимо, в первую очередь, для кремниевых цифровых СБИС, где важно заменить многоэлектродные шины данных, ограничивающие скорость передачи информации внутри микросхемы, на интегрированные оптические волноводы, имеющие существенно большую пропускную способность.

Перейдем далее непосредственно к рассмотрению принципов работы светоизлучающих диодов и полупроводниковых лазеров. Функционирование подобных приборов основано на явлении *люминесценции*. Напомним, что *люминесценцией* называют оптическое излучение, возникающее в результате

возбуждения электронов в материале (к люминесценции, однако, не относится тепловое излучение). В зависимости от способа возбуждения различают следующие типы люминесценции: *фотолюминесценция*, при которой предполагается возбуждение системы оптическим излучением; *катодолюминесценция* – возбуждение пучком электронов; *радиолюминесценция* – возбуждение другими быстрыми частицами или излучением высокой энергии; *электролюминесценция* – возбуждение электрическим полем или током.

Наиболее широкое распространение получили полупроводниковые источники света, в которых используется механизм *инжекционной электролюминесценции*. Суть его состоит в следующем: при подаче прямого смещения на *p-n*-переход неравновесные избыточные электроны и дырки инжектируются через барьер в приконтактную область; когда электрон встречается с дыркой (оказывается вблизи нее на расстоянии туннелирования), он переходит с уровня W' в зоне проводимости на уровень W в валентной зоне с излучением фотона энергии $h\nu = W' - W$. При малой скорости образования избыточных электронно-дырочных пар излучательная рекомбинация происходит самопроизвольно (спонтанно). Именно *спонтанное излучение* формируется в полупроводниковых светодиодах.

Спецификой светодиодного излучения является его некогерентность. Однако, в ряде задач принципиальным является требование когерентности источника. В настоящее время широкое распространение получили полупроводниковые лазеры, которые, как и твердотельные и газовые лазеры, позволяют получить высокую когерентность излучения, обладая при этом малыми габаритами и высокой эффективностью.

Напомним, что если квантовая система находится в возбужденном состоянии W' , то ее переход в основное состояние W может произойти вынужденно, когда с такой системой «сталкивается» фотон, имеющий энергию $h\nu = W' - W$. Избыточная энергия при этом выделяется в виде кванта света, т.е. происходит *стимулированное излучение*. Важно, что рожденный фотон полностью идентичен исходному, т.е. имеет те же самые энергию $h\nu$ и волновой вектор. Такое излучение получило название *когерентного*.

Как известно, необходимым условием для получения лазерного излучения является создание *инверсии населенности* в излучающей среде. Населенность считается инверсной, когда в некотором возбужденном состоянии с

энергией W' частиц находится больше, чем на основном уровне с энергией $W < W'$. Если фотон с энергией $h\nu = W' - W$ взаимодействует с системой, имеющей инверсную населенность уровней, то стимулированное излучение квантов света будет преобладать над их поглощением. При этом число фотонов с энергией $h\nu$, покидающих систему, будет больше, чем число входящих в нее. Такое явление называется *квантовым усилением*. Как известно, для создания генератора на основе усилителя, необходимо организовать положительную обратную связь. В лазерах она реализуется за счет введения активной среды в оптический резонатор, представляющий собой в простейшем случае два параллельных зеркала (резонатор Фабри-Перо), одно из которых является глухим, а другое – полупрозрачным (для осуществления вывода излучения). За счет таких зеркал часть излучения будет направляться обратно в инвертированную среду, вызывать вынужденное излучение и, таким образом, усиливаться. Когда усиление скомпенсирует суммарные потери на поглощение, рассеяние и вывод излучения, в резонаторе установятся стационарные автоколебания и система превратится в источник когерентного излучения. Рассмотрим далее, как эти общие принципы реализуются в полупроводниковых лазерах.

Для того чтобы рассмотреть условия получения инверсной населенности для полупроводниковых лазеров, обратимся к рисунку 1.15, на котором изображена зависимость энергии носителей от плотности состояний в прямозонном полупроводнике.

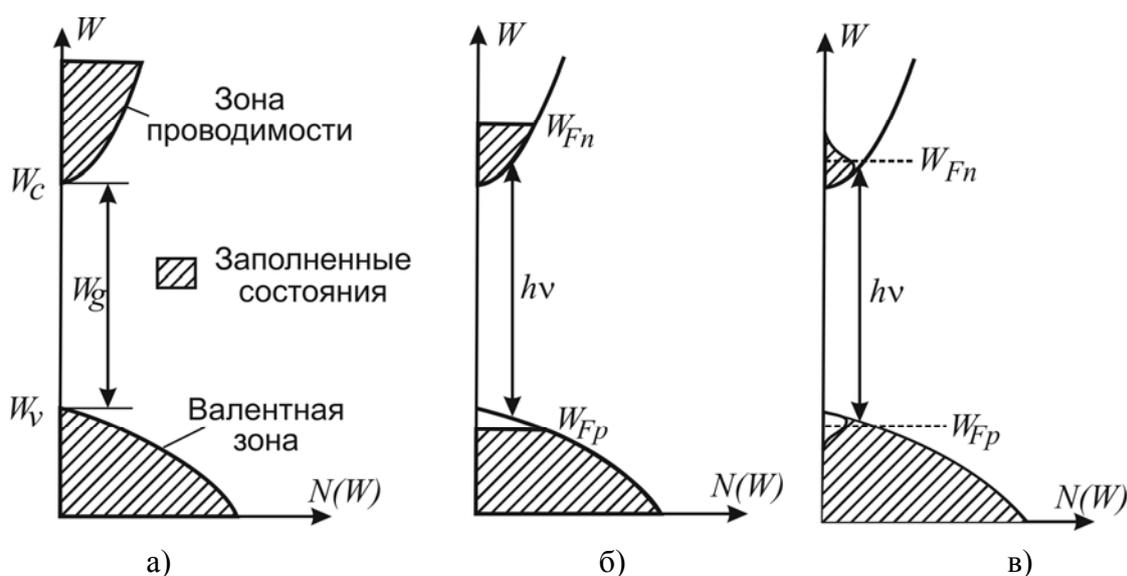


Рис. 1.15. Зависимость плотности состояний от энергии в полупроводнике [3]: а) состояние равновесия, $T=0\text{ K}$; б) инверсия, $T=0\text{ K}$; в) инверсия, $T > 0\text{ K}$

На рис. 1.15, *а* представлены равновесные условия при $T=0$ К для собственного полупроводника. Заштрихованная площадь соответствует заполненным состояниям. Рис. 1.15, *б* соответствует случаю инверсной населенности при $T=0$ К. Валентная зона пуста до энергетического уровня W_{Fp} , а зона проводимости заполнена до уровня W_{Fn} . Фотоны, имеющие энергию $h\nu$, лежащую в интервале $W_g < h\nu < (W_{Fn} - W_{Fp})$, вызывают переходы электронов на нижние уровни, т.е. возникает стимулированная эмиссия.

При отличных от абсолютного нуля температурах распределение носителей «размывается» по энергиям, как показано на рис. 1.15, *в*. Хотя общее тепловое равновесие всей системы отсутствует, можно считать, что носители в данной энергетической зоне находятся в тепловом равновесии друг с другом. В этом случае, как мы видели выше, пользуются распределением Ферми-Дирака, вводя квазиуровни Ферми для электронов и дырок. Тогда условие инверсии для какой-либо пары уровней, один из которых (W') лежит в зоне проводимости, а другой (W) – в валентной зоне, будет выглядеть следующим образом:

$$f(W') = \frac{1}{1 + \exp\left(\frac{W' - W_{Fn}}{kT}\right)} > f(W) = \frac{1}{1 + \exp\left(\frac{W - W_{Fp}}{kT}\right)}, \quad (1.12)$$

где W_{Fn} и W_{Fp} – квазиуровни Ферми для электронов в зоне проводимости и дырок в валентной зоне. Это приводит нас к выражению

$$W_{Fn} - W_{Fp} > W' - W \quad (1.13)$$

или, что то же самое,

$$W_{Fn} - W_{Fp} > h\nu. \quad (1.14)$$

Поскольку наименьшее значение разности $W_{Fn} - W_{Fp}$ – это ширина запрещенной зоны, то условие инверсии можно записать в виде

$$W_{Fn} - W_{Fp} > W_g. \quad (1.15)$$

Это условие, которое необходимо для того, чтобы стимулированное излучение преобладало над поглощением для межзонных переходов в собственном полупроводнике. Если полупроводник содержит примеси и энергетический уровень примеси является либо конечным, либо начальным состоянием, необходимо использовать квазиуровни Ферми для примесных уровней с соответствующими коэффициентами вырождения.

Как упоминалось выше, наиболее распространенным методом возбуждения в полупроводниковых лазерах является инжекция электронов и дырок через прямосмещенный p - n переход. Энергия для генерации излучения в этом случае черпается из электрической батареи, подключенной к диоду. Для увеличения интенсивности свечения p - и n -области перехода легируются до вырождения электронно-дырочного газа. В такой ситуации в условиях равновесия уровень Ферми в p -области лежит в валентной зоне, а в n -области – в зоне проводимости (рис. 1.16 а).

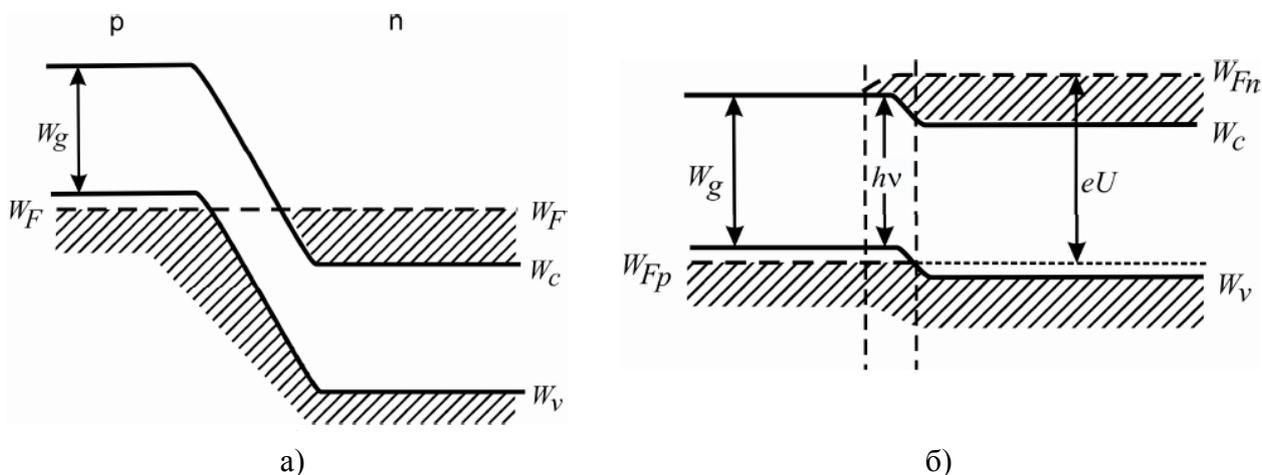


Рис. 1.16. Диаграмма энергетических зон лазера с p - n переходом (подобная диаграмме туннельного диода): а) без внешнего напряжения; б) при прямом смещении. Штриховкой изображены энергетические состояния, занятые электронами

Отметим, что зонная диаграмма, приведенная на рис. 1.16, подобна диаграмме туннельного диода (см. далее). Однако степень легирования областей лазера ниже, чем туннельного диода, поэтому обычно вольтамперная характеристика лазерного диода не имеет области отрицательного сопротивления или эта область пренебрежимо мала (сравните с ВАХ туннельного диода из главы 3).

При подаче определенного смещения, как видно из рисунка 1.16 б, вблизи границы перехода выполняется условие инверсии (1.15), необходимое для обеспечения однопроходного квантового усиления.

Для обеспечения генерации когерентного излучения, как мы помним, необходимо наличие резонатора. В отличие от твердотельных лазеров, в данном случае нет необходимости во внешнем оптическом резонаторе. Зеркалами служат сами грани кристалла, что автоматически обеспечивает их строгую

параллельность при скалывании. Такая методика позволяет избежать проблем юстировки и снизить габариты прибора. Для формирования гладкой отражающей поверхности кристалл скалывается и полируется. Коэффициенты отражения получаемых зеркал сравнительно невысоки:

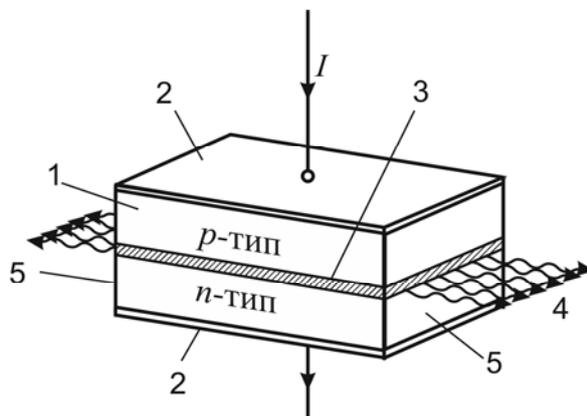
$$R = \frac{(n-1)^2}{(n+1)^2} \approx 20 \dots 40\% , \quad (1.16)$$

в зависимости от значения показателя преломления n в кристалле. Однако, при высоком уровне инжекции носителей коэффициент усиления в полупроводниковых лазерах настолько высок ($\sim 10^2 \text{ см}^{-1}$), что данного значения оказывается достаточно для достижения порога генерации даже при малых длинах активных областей ($\sim 100 \text{ мкм}$).

Общий вид инжекционного лазера представлен на рис. 1.17.

Рис. 1.17. Общий вид полупроводникового лазера с резонатором Фабри-Перо [3]:

- 1 – грубо обработанная поверхность (тыльная поверхность такая же);
- 2 – омические контакты;
- 3 – активная область;
- 4 – когерентное излучение;
- 5 – зеркала резонатора (полированные параллельные поверхности сколов)



При малых токах накачки лазерного диода возникает спонтанное излучение, распространяющееся по всем направлениям. При возрастании тока усиление также возрастает до тех пор, пока не будет достигнут порог генерации, т. е. не начинают выполняться условия, когда световая волна пересекает резонатор без затухания.

Обратите внимание, что активная область лазера сосредоточена непосредственно вблизи границы p и n областей (рис. 1.16 б) и имеет поперечный размер порядка микрометра, что сравнимо с длиной волны излучения. Поэтому при выходе во внешнюю среду излучение сильно дифрагирует. Таким образом,

полупроводниковые лазеры обладают высокой расходимостью ($20^\circ \dots 40^\circ$), которую компенсируют с помощью дополнительной фокусирующей оптики.

Ширина линии спонтанного излучения, характерного для светодиодов, определяется распределением носителей заряда по энергии и составляет несколько единиц kT (см. рис. 1.18). Это приводит к тому, что спектр излучения светодиодов будет иметь ширину³ примерно $4kT$.

Наличие в лазере резонатора Фабри-Перо не только обеспечивает положительную обратную связь, но и влияет на спектр излучения, поскольку в системе зеркал могут поддерживаться только те моды, для которых на удвоенной оптической длине резонатора укладывается целое число длин волн. Такие волны, проходя по резонатору в прямом и обратном направлении, не меняют своей фазы и, следовательно, усиливают друг друга при интерференции. Так как размеры полупроводникового кристалла порядка $100 \dots 1000$ мкм, а длина волны излучения, например, для GaAs лазера имеет величину около 0.9 мкм, то в резонаторе помещаются сотни длин волн. Для излучаемых квантов с энергией из диапазона от Wg до $Wg + 4kT$ условиям существования стоячей волны в резонаторе будут удовлетворять порядка десятка длин волн, т.е. на линии усиления среды будут уместиться порядка десятка продольных мод резонатора. Моды, которые испытывают усиление, достигающее порогового уровня, приводят к лазерной генерации. В зависимости от параметров структуры лазеры могут работать в одномодовом или многомодовом режимах. При этом отметим, что даже в одномодовом режиме длина волны излучения может быть нестабильной, что зависит от множества факторов — силы тока, температуры и т.д.

У лазерного диода есть два главных отличия от светодиода:

1. Лазерный диод имеет встроенный оптический резонатор, который отсутствует у светодиода.
2. Для перехода в режим генерации в лазерном диоде необходимо превысить некоторое пороговое значение тока накачки.

В результате лазеры имеют значительно меньшую ширину спектра излучения (≤ 1 нм) по сравнению со светодиодами (≥ 50 нм).

³ Данная оценка является достаточно грубой и отражает ширину спектра светодиода, измеренную по уровню 0.1 от максимального значения интенсивности.

Важно, что в процессе формирования вынужденного излучения в полупроводниковом лазере участвуют одновременно электрон, дырка и первичный фотон, которые должны геометрически оказаться вблизи друг друга. В обычном *p-n*-переходе это затруднено, так как рекомбинирующие носители заряда «размазаны» в слое, размером порядка 1 мкм, т.е. на расстоянии диффузионной длины от границы перехода, а профиль светового поля существенно шире толщины активной области прибора. В связи с этим для обеспечения генерации при комнатной температуре необходима очень сильная инверсия населенности, что, в свою очередь, требует больших плотностей тока накачки ($\sim 1 \text{ кА/см}^2$). Во избежание сильного разогрева и выхода прибора из строя подобные диоды работают в импульсном режиме либо требуют принудительного охлаждения. Поэтому все современные лазеры изготавливаются не на основе гомоструктур, а с использованием гетеропереходов.

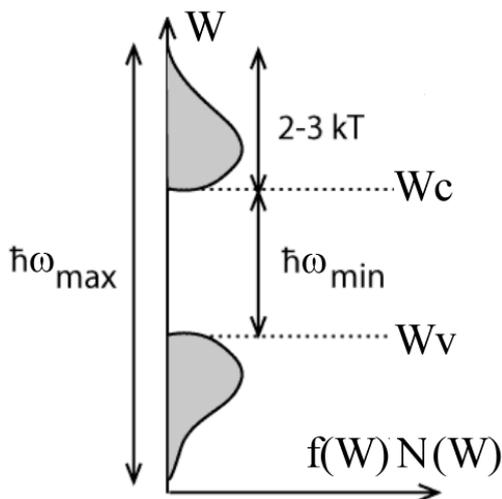


Рис. 1.18. Распределение электронов и дырок по энергии в зоне проводимости и валентной зоне, соответственно. Стрелками указаны минимальная и максимальная энергия кванта, генерируемого при электронно-дырочной рекомбинации

Одной из самых распространенных лазерных конструкций является двойная гетероструктура (ДГС), изображенная на рисунке 1.19. Здесь слой материала с более узкой запрещенной зоной (GaAs) располагается между двумя слоями материала с более широкой запрещенной зоной (AlGaAs). Каждый из двух контактов GaAs–AlGaAs сам по себе является гетероструктурой, поэтому подобную трехслойную композицию и называют двойной (или двусторонней) гетероструктурой.

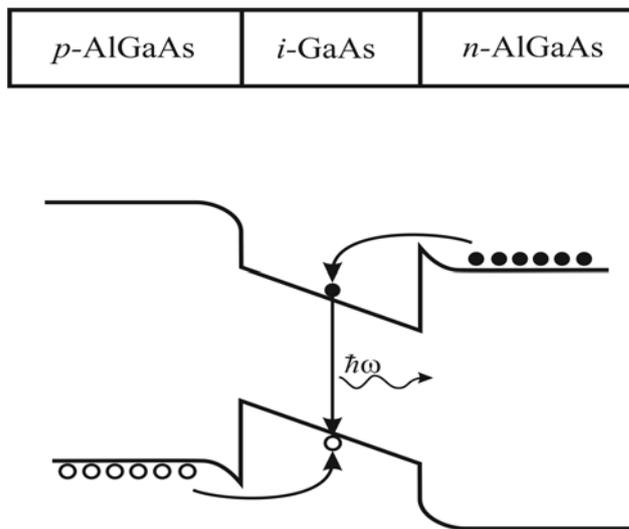


Рис. 1.19. Схематическое изображение конструкции и зонной диаграммы лазера с двумя гетеропереходами, ограничивающими диффузию инжектируемых электронов и дырок

ДГС лазеры имеют ряд преимуществ, основными из которых:

а) *Электронное ограничение*: благодаря наличию потенциальной ямы на границе p - n перехода для одного или обоих типов носителей инжектируемые электроны и/или дырки удерживаются в активной области прибора. При заданном токе концентрация электронно-дырочных пар в гетеропереходе будет намного выше, чем в гомопереходе, где диффузия носителей будет эффективно «выбрасывать» их из активного слоя.

б) *Оптическое ограничение*: поскольку широкозонные материалы обладают меньшим коэффициентом преломления, в гетероструктурных лазерах реализуется каналирование фотонов вдоль границы перехода, как это происходит в световодах. Это позволяет уменьшить потери квантов при их рассеянии на неоднородностях.

В итоге, снижение электронных и оптических потерь в гетероструктурном лазере на порядок уменьшает ток, необходимый для возбуждения генерации когерентного излучения (*пороговый ток накачки*), что позволяет приборам работать в непрерывном режиме при комнатной температуре.

В случае, когда толщина слоя узкозонного материала имеет размер порядка длины волны электрона, в активной области прибора реализуется квантовая яма (рис. 1.19), в которой электроны и дырки размещены на энергетических уровнях. Последнее приводит к резкому увеличению коэффициента усиления и позволяет уменьшить пороговый ток накачки лазерного диода еще на порядок.

Дальнейшее продвижение технологии изготовления лазерных диодов связано с использованием квантовых объектов еще меньшей (нулевой) размерно-

сти, чем квантовая яма – *квантовых точек*. По сути, квантовая точка – это объемное включение узкозонного материала внутри широкозонного, причем размеры этого объекта составляют доли длины волны электрона. Таким образом, по всем трем направлениям формируется квантовая яма для носителей заряда. Используя термины электродинамики, можно говорить, что в случае квантовой точки мы имеем резонатор для одного электрона.

Для справки рассмотрим схему основных этапов технологии изготовления квантовых точек, основанной на *механизме Странского - Крастанова*.

Первоначально на поверхность GaAs с помощью технологического процесса эпитаксии наносят тонкий слой InAs (рис. 1.20 а). Далее начинается рост трехмерных изолированных островков: для минимизации упругой энергии InAs собирается в «пирамидки» (рис. 1.20 б, в). После этого поверх сформированных «пирамидок» наносят следующий слой GaAs, «заращая» образованные островки. При нанесении второго слоя InAs новые островки будут формироваться в поле упругих напряжений, созданных островками первого слоя. Таким образом, новые квантовые точки будут образовываться непосредственно над старыми. Описанная процедура повторяется 20...50 раз, что, в итоге, позволяет вырастить трехмерный массив островков InAs в матрице GaAs.

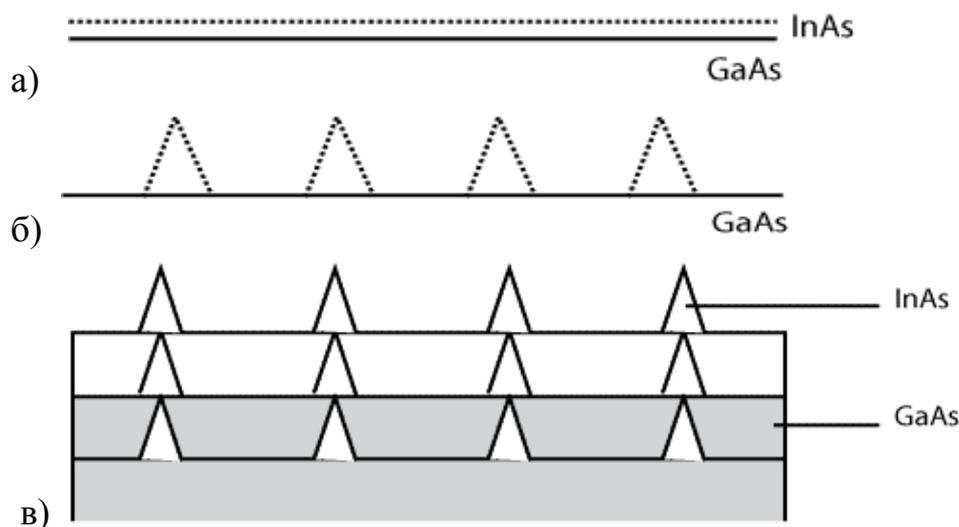


Рис. 1.20. Схема основных этапов технологии изготовления матрицы квантовых точек⁴

⁴ Изображение носит схематический характер. В действительности сформировать абсолютно упорядоченную структуру не удастся.

Поскольку нульмерные объекты обладают дельтаобразной плотностью состояний и значительной степенью геометрической локализации электронно-дырочных пар, использование подобных структур для формирования активной области лазерного диода позволяет еще больше снизить пороговый ток лазерных диодов, т.е., в итоге, повысить их эффективность. Кроме того, спектр энергии электронов в квантовой точке носит дискретный характер, подобно спектру электронов в атоме. Т.е. в такой ситуации излучательные переходы носителей заряда будут происходить между отдельными уровнями, а не зонами, что позволит сузить ширину линии усиления при условии незначительного разброса размеров квантовых точек.

ЧАСТЬ 2

СВЧ И КВЧ⁵ ДИОДЫ

2.1. Диод Ганна

Диодом Ганна называют тип полупроводниковых диодов, использующихся для генерации и преобразования колебаний в СВЧ диапазоне. Конструктивно диод Ганна представляет собой просто однородный полупроводник с нанесенными на него омическими контактами (рис. 2.1). Принцип действия диодов Ганна, в отличие от большинства полупроводниковых приборов, основан не на свойствах различных переходов, а на объёмных свойствах однородного полупроводника. В 1963 году Джон Ганн исследовал вольтамперные характеристики образцов GaAs и InP. Он обнаружил, что когда напряженность электрического поля, приложенного к образцу, которая определяется как

$$E = U/L \quad (2.1)$$

(где U – приложенное напряжение, L – длина образца), превосходит некоторую критическую величину E_t (~ 3 кВ/см для GaAs и ~ 6 кВ/см для InP), в цепи возникают спонтанные колебания тока (рис. 2.2). Частота этих колебаний примерно равнялась величине, обратной времени пролета носителей заряда через образец:

$$f = v_{dr}/L, \quad (2.2)$$

где $v_{dr} = 10^7$ см/с – скорость носителей заряда, L – длина образца.



Рис. 2.1. Структура GaAs диода Ганна. Система слоев n^+ -Au реализует омический контакт

⁵ СВЧ – сверхвысокая частота, КВЧ – крайне высокая частота. Термины введены по аналогии с англоязычной литературой для обозначения частот свыше 1 ГГц (СВЧ) и выше 100 ГГц (КВЧ). Следует отметить, что указанные границы частот весьма условны.

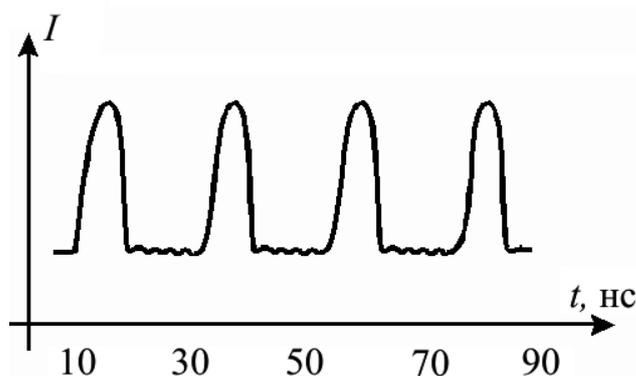


Рис. 2.2. Зависимость тока от времени для GaAs диода Ганна

Позднее Ганн опубликовал результаты детального экспериментального исследования этого эффекта. Используя зондовые измерения распределения потенциала вдоль образца, он установил, что при $E \geq E_p$ в образце формируется область сильного поля (домен). Домен зарождается вблизи катода, движется к аноду со скоростью 10^7 см/с и исчезает вблизи анода. Когда домен формируется, ток падает, при исчезновении домена, ток возрастает.

В 1963г. Ридли предсказал, что доменная неустойчивость должна появляться в полупроводниковом образце в случае, если на вольтамперной характеристике образца имеется участок *отрицательного дифференциального сопротивления (ОДС)* или, что то же самое, *отрицательной дифференциальной проводимости (ОДП) N-типа* (рис. 2.3). ВАХ будет иметь такой вид, если либо скорость носителей, либо их концентрация уменьшается с увеличением приложенного напряжения. Ридли, Уоткинс и Хилсум показали, что скорость носителей в GaAs и InP *n*-типа и в некоторых других полупроводниковых материалах должна уменьшаться с ростом электрического поля, когда напряженность поля превышает некоторое критическое значение. Данный эффект объясняется междолинными переходами электронов: в сильных полях энергия носителей заряда оказывается достаточной для их перехода из основной Γ -долины в более высоколежащие X и L -долины, где эффективная масса электронов увеличивается, что и приводит к снижению их дрейфовой скорости. График зависимости $\nu(E)$ для разных значений подвижности, определяемых уровнем легирования образца, приведен на рис. 2.4. На графике хорошо виден спадающий участок – область ОДП.

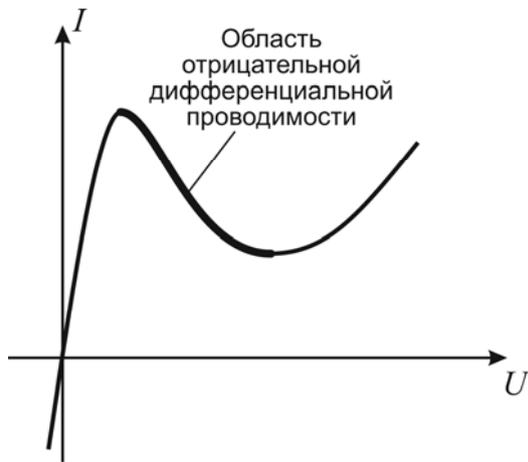


Рис. 2.3. Пример ВАХ N-типа. Жирной линией показан участок отрицательной дифференциальной проводимости, т.е. область, где $\frac{dI}{dU} < 0$

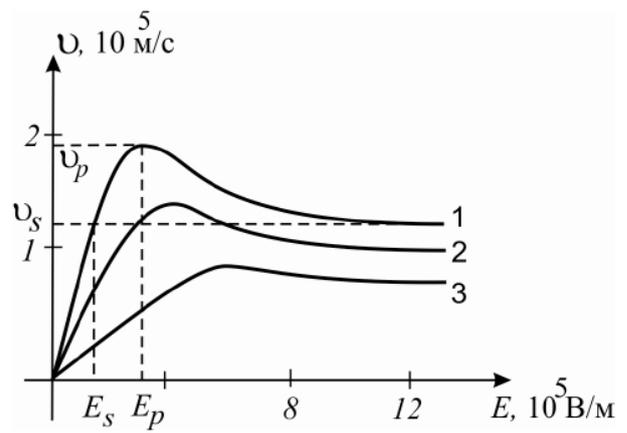


Рис. 2.4. Зависимости дрейфовой скорости электронов v от напряженности электрического поля E в GaAs для различных значений подвижности μ .

1 – $\mu = 0,85 \text{ м}^2/\text{В}\cdot\text{с}$;

2 – $\mu = 0,42 \text{ м}^2/\text{В}\cdot\text{с}$;

3 – $\mu = 0,14 \text{ м}^2/\text{В}\cdot\text{с}$

Рассмотрим подробнее процесс формирования области сильного поля (домена Ганна) в образце (рис. 2.5) и ее влияние на ток, протекающий через структуру.

Предположим, что в начальный момент времени $t=0$ напряженность электрического поля E в образце равна E_p всюду, за исключением небольшой области вблизи катода, где она немного выше вследствие неоднородности легирования. Так как напряженность поля в указанной области превышает E_p , скорость электронов здесь меньше, чем в остальных частях образца, поскольку при $E > E_p$ скорость V падает с ростом напряженности поля E (рис. 2.4). Следовательно, электроны перед областью сильного поля и за этой областью движутся быстрее, чем внутри нее, где они претерпевают междолинный переход и становятся более тяжелыми и медленными. Это ведет к обеднению электронами переднего и обогащению ими заднего фронтов движущейся области сильного поля. Иными словами, быстрые электроны, находящиеся левее домена (рис. 2.5), догоняют его, попадают в область сильного поля и тормозятся. Быстрые электроны, находящиеся справа от домена, наоборот, убегают от него, обнажая положительный заряд ионизованных доноров. Таким образом, концен-

трация электронов в задней части домена превышает концентрацию ионов доноров, а в передней части – наоборот. То есть распределение зарядов в домене представляет собой дипольный слой, а направление электрического поля совпадает с его направлением в остальных частях образца. Тогда напряженность поля в области домена начинает расти и вызывает дальнейшее увеличение междолинных переходов и рост домена.

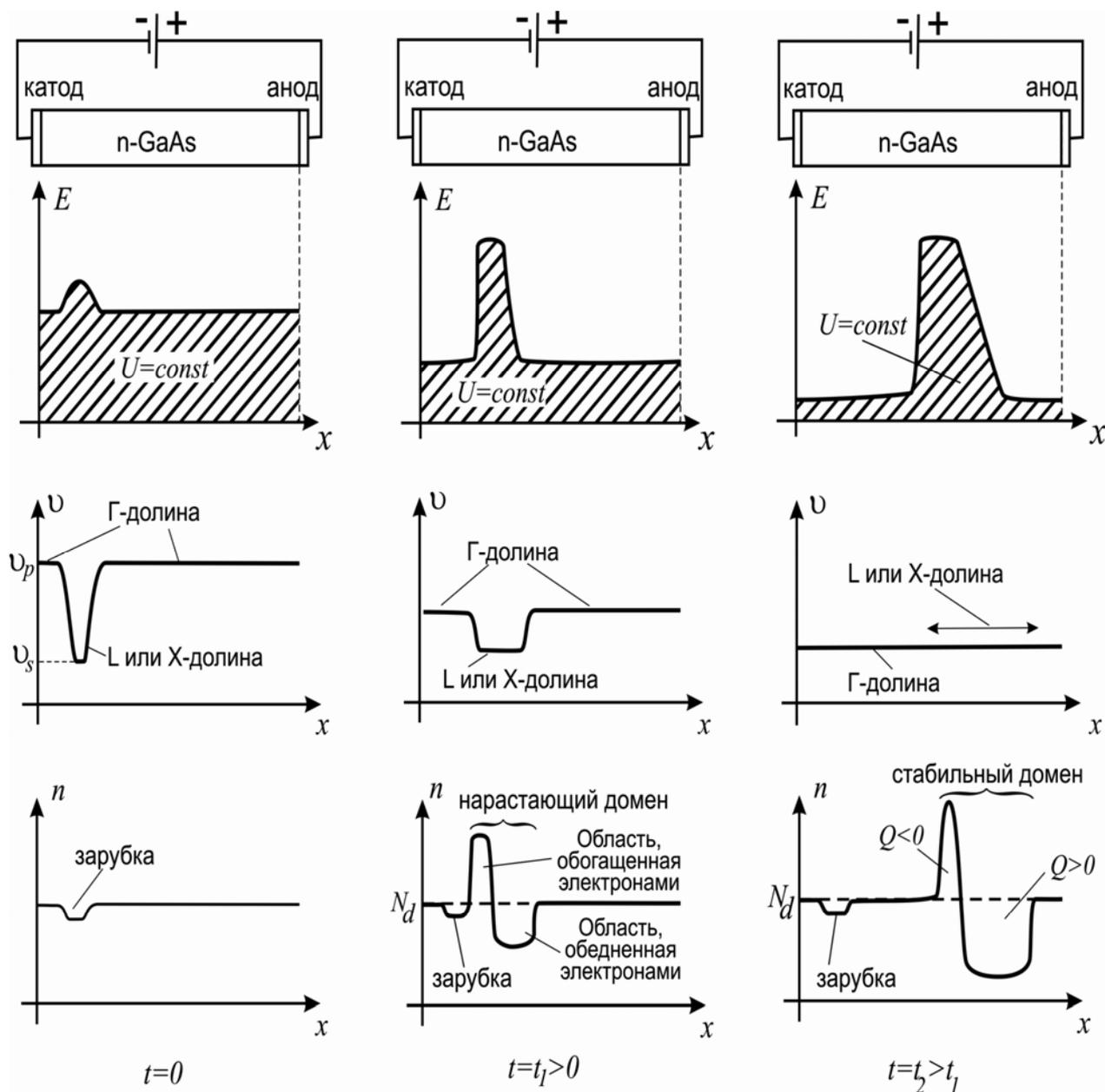


Рис. 2.5. Процесс формирования домена. Здесь Q – суммарный заряд электронов и доноров; E – напряженность электрического поля; V – скорость электронов; n – концентрация носителей заряда; x – координата вдоль образца. Пунктиром показана концентрация доноров N_d в объеме полупроводника. В области зарубки концентрация доноров (и электронов) имеет меньшее значение

Если напряжение смещения U на образце поддерживается постоянным, то нарастание напряженности поля внутри домена происходит за счет уменьшения напряженности поля в остальных областях (площадь под графиком $E(x)$ остается постоянной). При этом как внутри, так и вне домена скорость носителей заряда уменьшается. Через некоторое время скорости электронов, находящихся вне и внутри области сильного поля, уравниваются, а домен перестает расти. В такой ситуации говорят, что он стабилизируется. Другой причиной стабилизации домена является диффузия электронов, вызванная градиентом их концентрации. После стабилизации домен двигается к аноду, не меняя своей формы. Что касается тока диода, то он уменьшается в процессе формирования домена, а затем остается постоянным, пока домен не добежит до анода. Когда домен сильного поля достигает анода, он рассасывается, а ток при этом снова возрастает до исходного значения. Затем вблизи катода зарождается новый домен, и процесс повторяется. Обычно в образце формируется только один домен, поскольку напряженность поля везде, за исключением домена, меньше E_p , т.е. междолинные переходы в области вне домена подавлены. В итоге, мы получаем периодическую зависимость тока образца от времени (рис. 2.2)

В большинстве случаев (но не всегда) домен формируется вблизи катода и движется по образцу со скоростью, близкой к скорости насыщения v_s . Плотность тока в образце, по которому движется домен, в этом случае равна

$$j_s = qn v_s. \quad (2.4)$$

В процессе исчезновения домена и формирования нового плотность тока возрастает до величины

$$j_p = qn v_p. \quad (2.5)$$

Время движения домена (пролетное время) определяется выражением

$$T \approx L/v_s, \quad (2.6)$$

где L – длина образца. Время начального нарастания малой флуктуации t_{gr} определяется *временем диэлектрической релаксации (временем максвелловской релаксации)* τ_m :

$$t_{gr} \approx 3\tau_m = \frac{3\varepsilon}{e|\mu_d|n}, \quad \mu_d = \frac{dv}{dE}. \quad (2.7)$$

Напомним, что по своему физическому смыслу максвелловское время релаксации может быть определено как время, за которое рассасывается флуктуация концентрации электронов Δn , возникшая из-за их случайного теплового движения: $\Delta n = \Delta n_0 e^{-\frac{t}{\tau_m}}$. В случае специальных условий, которые реализуются в диоде Ганна (т.е. при наличии отрицательной дифференциальной подвижности $\mu_d < 0$), флуктуация концентрации электронов будет не рассасываться, а, наоборот, нарастать до тех пор, пока домен не стабилизируется. Для формирования стабильного движущегося домена время t_{gr} должно быть существенно меньше, чем пролетное время

$$t_{gr} < \frac{L}{v_s}. \quad (2.8)$$

Из этого неравенства вытекает так называемый *критерий Кремера*:

$$nL > (nL)_1 = \frac{3\varepsilon v_s}{e|\mu_d|}. \quad (2.9)$$

Для GaAs имеем $\varepsilon = 1,14 \cdot 10^{-10}$ Ф/м, $|\mu_d| \approx 0,07$ м²/В·с, $v_s = 10^5$ м/с и $(nL)_1 \approx 3 \cdot 10^{15}$ м⁻². Если величина параметра $(nL) \geq 10^{16}$ м⁻², то стабильный движущийся домен формируется, когда $E \geq E_p$. При $(nL) \leq 10^{16}$ м⁻² напряженность порогового поля формирования домена E_t больше, чем E_p . Величина E_t зависит от параметра (nL) и формы зависимости $v(E)$.

Процесс формирования домена можно рассматривать как зарядку плоского конденсатора, где одной обкладкой является область, в которой электронов больше, чем доноров, а другой – соседняя область, где концентрация ионов доноров больше, чем электронов (рис. 2.5). Тогда время формирования домена $\tau_m = RC_{\text{домена}}$. Поскольку домен формируется во всем сечении диода Ганна, то при увеличении площади сечения емкость домена будет возрастать, а сопротивление соседних с доменом областей, через которые эта емкость заряжается – снижаться. В результате характерное время зарядки домена останется без изменений. Это говорит о том, что формирование ганновской неустойчивости – процесс, определяемый свойствами самого полупроводникового материала, поэтому характерное время формирования домена должно определяться константами материала ($\tau_m = \varepsilon/\sigma$), как мы и получили выше. Явление доменообразо-

вания⁶ в диоде Ганна, по сути, является механизмом положительной обратной связи, которая необходима для возникновения колебаний в любой колебательной системе.

Важной особенностью ганновской генерации является тот факт, что сформировавшийся домен не исчезает, если во время его движения в диоде напряжение на образце уменьшается до величины, меньшей, чем пороговое напряжение формирования домена. Происходит это потому, что поле внутри домена остается *большим*, чем E_p , даже если средняя напряженность поля по образцу становится ниже E_p . Для значений параметра $(nL) \geq 10^{16} \text{ м}^{-2}$ напряженность поля, необходимая для поддержания домена, близка к величине E_s .

Рассмотрим работу генератора на основе диода Ганна (рис. 2.6). Прибор представляет собой коаксиальный резонатор, который с одной стороны ограничен боковой стенкой коаксиального волновода с металлизированным торцом, к которому припаян диод, а с другой стороны – плунжером – подвижным поршнем с отверстием в центральной части. Перемещая плунжер, т.е. изменяя геометрические размеры резонатора, можно осуществлять его настройку на различные резонансные частоты. Таким образом, в генераторе имеется возможность управлять частотой колебаний выходного сигнала, хотя длина диода Ганна является фиксированной, так что время пролета домена строго задано. Кроме того, некоторые конструкции плунжера позволяют менять диаметр его внутреннего отверстия так, что возможна регулировка мощности выходного сигнала, а также мощности сигнала, который «остается» в резонаторе, детектируется центральным стержнем и прикладывается в виде переменного напряжения к диоду Ганна. Иными словами, диаметр внутреннего отверстия плунжера определяет глубину обратной связи в генераторе.

Сначала рассмотрим наиболее простой случай *пролетного* режима работы генератора. Пусть время пролета домена в диоде в точности равно периоду колебаний электромагнитного поля резонатора. Такая мода колебаний представлена на рис. 2.7.

⁶ Более точно было бы говорить о явлении нарастания волн пространственного заряда, поскольку усиление и генерация в диоде Ганна в некоторых случаях могут реализовываться и без возникновения ярко выраженного домена.

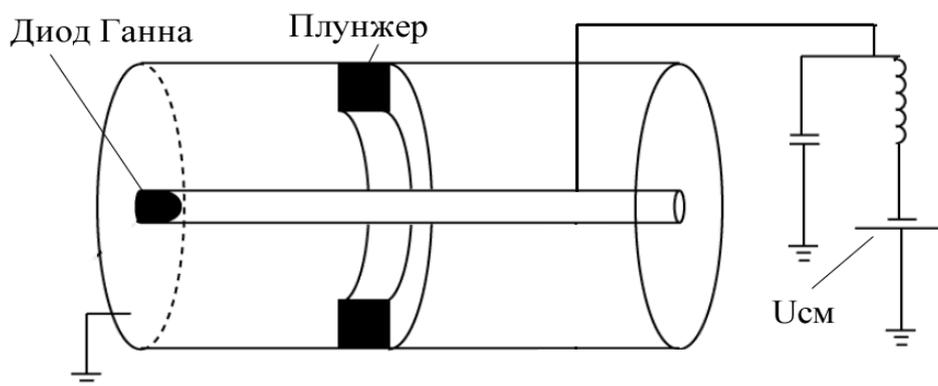


Рис. 2.6. Схема СВЧ генератора на основе коаксиального резонатора и диода Ганна.

$U_{см}$ – напряжение смещения.

- Плунжер может перемещаться вправо/влево.
- Регулирование резонансной частоты резонатора осуществляется перемещением плунжера, а изменение внутреннего диаметра плунжера регулирует как потери резонатора, так и связь с внешним волноводом.
- Левый торец волновода металлизирован, заземлен и к нему припаян диод Ганна; в совокупности с плунжером образуется область резонатора.

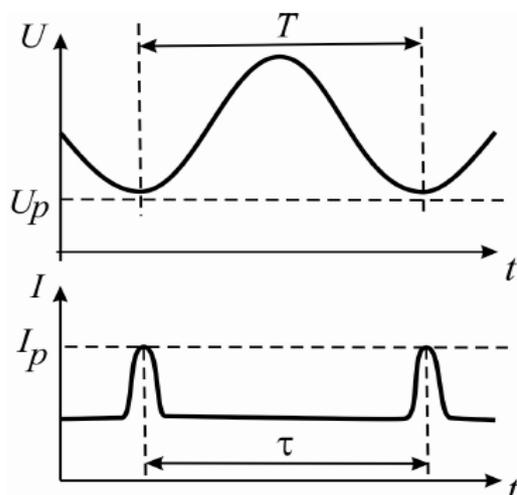


Рис. 2.7. Пролетный режим работы генератора (пролетное время домена в диоде τ равно периоду колебаний электромагнитного поля T)

Важно, что период переменной составляющей напряжения, прикладываемого к диоду, будет соответствовать периоду электромагнитного поля, так как это напряжение порождено самим полем. Характерная форма колебаний тока полностью аналогична рассмотренной выше форме колебаний для постоянного напряжения смещения. Следует отметить, что в данном режиме работы постоянную составляющую напряжения, подаваемого на диод от батареи, выбирают такой, что напряженность поля внутри диода всегда выше полей зарождения и удержания домена. Обозначим такое напряжение U_p . При исчезновении домена

у анода и формировании вблизи катода нового домена поле в образце быстро достигает порогового значения, которое сохраняется, пока не сформирован новый домен. Поэтому генерируются короткие импульсы тока (рис. 2.7).

В режиме *запаздывания* пролетное время τ меньше периода колебаний резонатора T (рис. 2.8).

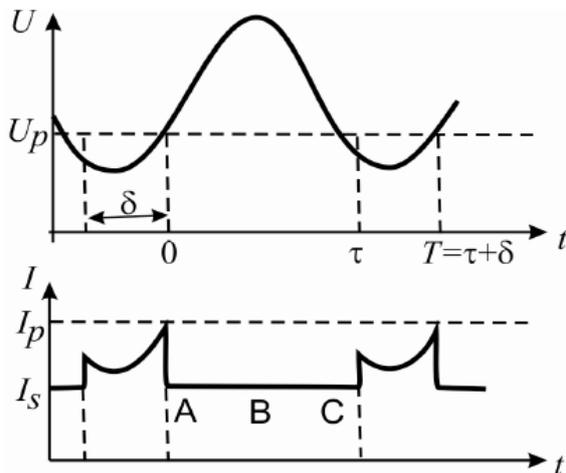


Рис. 2.8. Идеализированный режим запаздывания формирования домена, параллельный резонансный контур. В момент времени А домен формируется, В – распространяется вдоль образца, С – достигает анода

Пусть домен формируется в момент времени $t = 0$. Пусть также величина постоянного смещения и амплитуда СВЧ напряжения таковы, что перед тем, как домен достигнет анода в момент времени $t = \tau$ (τ - пролетное время), напряжение на диоде упадет ниже порогового значения U_p . Далее в некоторый момент времени $t = \tau + \delta$ напряжение снова превысит порог (рис. 2.8), и у катода начнет образовываться новый домен. Таким образом, если предположить, что домен рассасывается на аноде мгновенно, то формирование нового домена запаздывает на время δ . Период колебаний при этом равен $T = \tau + \delta$. Максимальная выходная мощность в рассматриваемом режиме достигается, если домен достигает анода и рассасывается сразу же после того, как напряжение падает ниже порогового значения. Варьируя геометрические размеры резонатора, можно менять период колебаний T за счет изменения задержки δ , в то время как пролетное время τ остается неизменным и определяется длиной диода Ганна. Отметим, что в отсутствие домена форма зависимости тока от времени повторяет форму напряжения, т.е. диод представляет собой обычный резистор и подчиняется закону Ома. После зарождения нового домена ток снова резко падает (рис. 2.8).

Еще одним режимом работы генератора с диодом Ганна является режим с *гашением* домена (рис. 2.9). В этом случае в процессе перемещения по образцу домен разрушается (гасится), не успевая дойти до анода. Для такого разрушения дипольного слоя необходимо настолько сильно уменьшить поданное на диод напряжение, чтобы даже в максимуме напряженность электрического поля внутри домена стала меньше пороговой (рис. 2.5). Напряжение U_s , при котором сформированный домен разрушается, не доходя до анода, называют напряжением гашения, и, как упоминалось выше, оно меньше порогового напряжения U_p . Когда суммарное напряжение на диоде падает до величины, меньшей напряжения гашения, домен исчезает, а прибор возвращается в омическое состояние, в котором он остается до тех пор, пока напряжение вновь не превысит порог образования домена (рис. 2.9). Далее процесс повторяется.

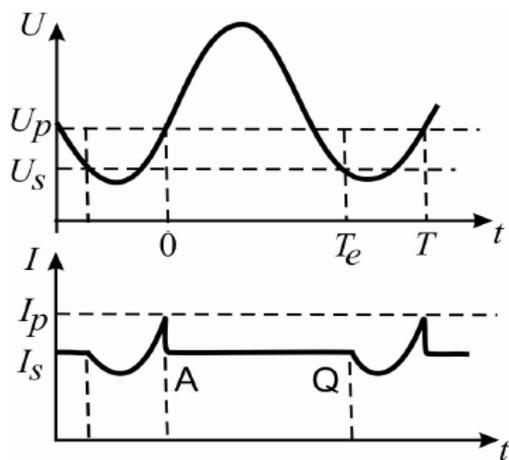


Рис. 2.9. Идеализированный режим гашения домена. В момент времени А домен формируется, а в момент Q – гасится, не дойдя до анода

Так как для реализации этого режима необходима большая величина переменной составляющей напряжения на диоде, диаметр отверстия плунжера необходимо уменьшать, увеличивая тем самым глубину обратной связи в генераторе, но снижая выходную мощность прибора. Отметим, что режим с гашением домена имеет наибольший интервал перестройки по частоте, а пролетный режим – максимальную выходную мощность и КПД.

2.2. ЛАВИННО-ПРОЛЕТНЫЕ И ИНЖЕКЦИОННО-ПРОЛЕТНЫЕ ДИОДЫ

2.1.1. Лавинно-пролетные диоды

В лавинно-пролетных диодах (ЛПД) для получения отрицательного сопротивления в СВЧ-диапазоне используются ударная ионизация в полупроводниковых структурах и пролетные эффекты сгустков электронов и дырок. В режиме ЛПД p - n переход может работать при двух условиях: а) на него подано обратное смещение, достаточное для развития лавинного пробоя; б) он помещен в микроволновый резонатор, например, такой как приведен на рис. 2.6 предыдущего раздела.

ЛПД являются одними из наиболее мощных полупроводниковых источников энергии в СВЧ диапазоне. Отрицательное сопротивление, возникающее в результате пролетных эффектов в полупроводниковых диодах, впервые рассмотрено Тагером в 1957 г. и Шокли. В 1958 г. Первые экспериментальные результаты по получению генерации в режиме ЛПД были получены в СССР в 1959 г. [2].

Основными представителями семейства лавинно-пролетных диодов являются $p^+ - n - i - n^+$ или $n^+ - p - i - p^+$ диод Рида, асимметричный резкий p - n переход, плавный p - n переход с линейным распределением примесей и p - i - n диод.

Рассмотрим принцип действия ЛПД на примере диода Рида. На рис. 2.10 показаны структура и распределения примесей, электрического поля и эффективной плотности ионизации в условиях пробоя для идеализированного диода Рида [3].

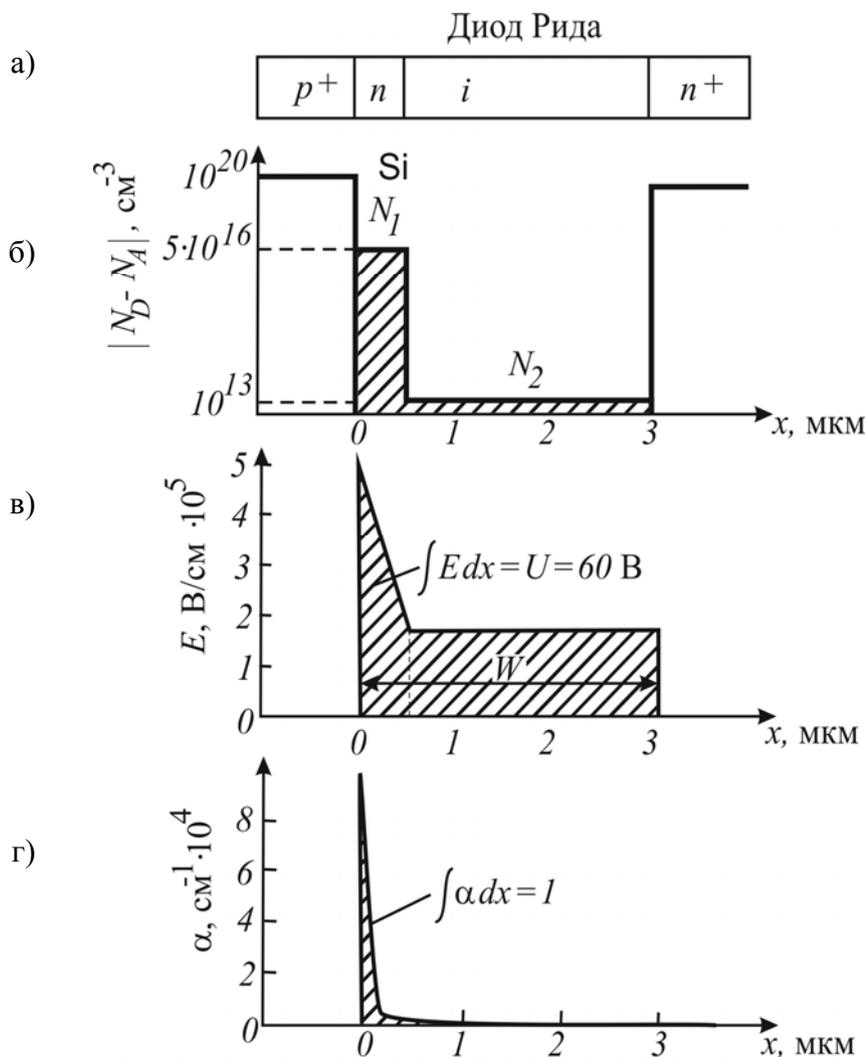
Как упоминалось выше, лавинный пробой развивается в том случае, когда на длине свободного пробега электрон набирает энергию, достаточную для ионизации атомов вещества (т.е. больше ширины запрещенной зоны). Необходимая для этого напряженность электрического поля является характеристикой материала и по порядку величины составляет 100...400 кВ/см для наиболее распространенных полупроводников.

В диоде Рида зависимость напряженности поля от координаты является резко неоднородной (рис. 2.10 в), поэтому область «слоя умножения», т.е. слой, где происходит ударная ионизация, оказывается локализованным вблизи p - n перехода. Иными словами, большая часть актов ударной ионизации

совершается в узком слое вблизи максимума напряженности электрического поля (рис. 2.10 г).

Обычно ЛПД работает при постоянной компоненте напряжения питания чуть меньше, чем напряжение пробоя.

- Рис. 2.10. Диод Рида:
- а) структура диода;
 - б) распределение примесей в образце ($p^+ - n - i - n^+$);
 - в) распределение электрического поля;
 - г) распределение коэффициента ударной ионизации при лавинном пробое [2]



При помещении ЛПД в резонатор аналогично диоду Ганна, к прибору прикладывается переменное напряжение, так что суммарное напряжение в течение небольшой части периода колебаний становится больше напряжения пробоя $U_{проб}$ (рис. 2.11). В течение этого интервала времени вблизи $p-n$ перехода происходит лавинная генерация электронно-дырочных пар, но из-за того, что электроны достигают анода за конечное время (пролетный эффект), максимум тока запаздывает относительно максимума напряжения. Величина запаздывания определяется временем прохождения электрона через пролётную n -область:

$$\tau = \frac{L}{v_{нас}} \approx \frac{T}{2}, \quad (2.10)$$

где L – длина пролетной области, T – период колебаний, $v_{нас}$ – насыщенная дрейфовая скорость носителей заряда.

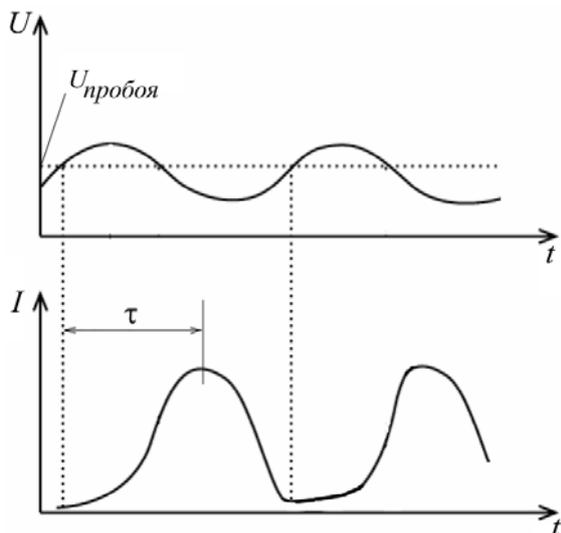


Рис. 2.11. Зависимость напряжения и тока от времени для ЛПД, помещенного в резонатор

Отрицательная дифференциальная проводимость лавино-пролетных диодов, диода Ганна и других генераторных диодов реализуется на тех участках периода колебаний, когда ток и напряжение изменяются в противофазе. Однако, в отличие от диода Ганна, в лавино-пролетных диодах ОДП достигается *только за счет наличия обратной связи в резонаторе*. Другими словами, в отличие от диода Ганна, в ЛПД генерация без резонатора невозможна – вольт-амперная характеристика p - n -перехода *не* имеет участков с отрицательным наклоном.

КПД любого из пролётных диодов можно оценить как отношение мощности переменного сигнала в интервале времени, когда существует ОДП, к рассеиваемой за период колебаний мощности. В связи с этим перестройка частоты генератора приводит к сокращению длительности интервалов ОДП, снижению КПД и выходной мощности. Теоретически период колебаний в резонаторе T может соотноситься со временем пролета τ следующим образом:

$\frac{1}{4}T < \tau < \frac{3}{4}T$. На практике, однако, величина перестройки генератора обычно в

1.5...2 раза меньше.

2.2.2. Инжекционно-пролетные диоды

Принцип работы *инжекционно-пролетного диода* (ИПД), в целом, подобен ЛПД. Конструкция ИПД содержит два *p-n*-перехода и пролетную область (рис. 2.11), за счет которой осуществляется задержка между импульсами тока и напряжения на диоде. Разница заключается в режиме питания – ИПД работают при прямом смещении. На каждом периоде колебаний напряжения ИПД на короткое время отпирается и в пролетную *n*⁻-область вбрасывается некоторое количество дырок, которые затем дрейфуют к аноду и создают импульс тока во внешней цепи. Схематические изображения зависимостей тока и напряжения от времени будут аналогичны приведенным на рис. 2.11. Поскольку правый *p-n*-переход находится при обратном смещении, большая часть пролетной области ИПД в рабочем режиме обеднена электронами. При этом сгусток дырок, который инжектируется в *n*-область из левого, прямо смещенного, *p-n*-перехода, не рекомбинирует в пролетной области, а добегает до анода. Это позволяет получить эффективную задержку импульса тока диода относительно положительного полупериода напряжения, который открыл левый (инжектирующий) переход.

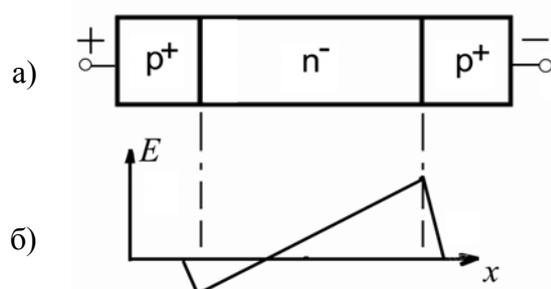


Рис. 2.12. Инжекционно-пролетный диод: а) схематическое изображение структуры диода; б) распределение напряженности электрического поля в образце.

В левой части диода электрическое поле образовано донорами из n^- слоя и акцепторами из p^+ слоя катода, в правой части - донорами из n^- слоя и акцепторами из p^+ слоя анода

Важной особенностью работы ИПД является возможность реализации такого напряжения питания, когда ОПЗ прямо смещенного перехода катода и обратнсмещенного перехода анода смыкаются (рис. 2.12 б), так что во всей пролетной области (кроме одной точки, в которой и происходит смыкание) существует ненулевая напряженность электрического поля. Такое явление получило название «прокола». Данный режим работы диода позволяет существенно снизить эффект накопления инжектированных дырок внутри диода – поле высокой напряженности вблизи анода захватывает дырки и очень быстро переносит их в анод. В связи с этим на вольт-амперной характеристике прибора

(рис. 2.13) участок резкого нарастания тока термоэмиссии сменяется участком ограничения тока пространственным зарядом инжектированных носителей только при очень больших плотностях тока. Это позволяет создавать не только инжекционно-пролетные диоды с относительно высоким КПД, но и использовать подобные приборы в качестве быстродействующих ограничителей напряжения.

В заключение, сравним основные характеристики ЛПД, ИПД и диода Ганна. Так же, как и ЛПД, ИПД отличается от диода Ганна тем, что отрицательная дифференциальная проводимость в приборе реализуется в узком спектральном диапазоне вблизи пролётной частоты. У диода Ганна ОДП реализуется в широком диапазоне частот. Энергия носителей заряда в ИПД меньше, чем в диоде Ганна (так как в ИПД нет междолинных переходов), а в диоде Ганна меньше, чем в ЛПД (так как в диоде Ганна нет пробоя), поэтому наименьшие шумы реализуются в ИПД и в диодах Ганна, а в ЛПД они значительно выше. Наоборот, амплитуды токов и выходные мощности соотносятся следующим образом: самые большие – у ЛПД, а у ИПД – наименьшие. Предельные частоты работы приборов составляют $f_{\text{ЛПД}} \approx 600$ ГГц, $f_{\text{Ганна}} \approx 150$ ГГц, $f_{\text{ИПД}} \approx 50$ ГГц. Сравнение характеристик диодов приведено на рис. 2.14.

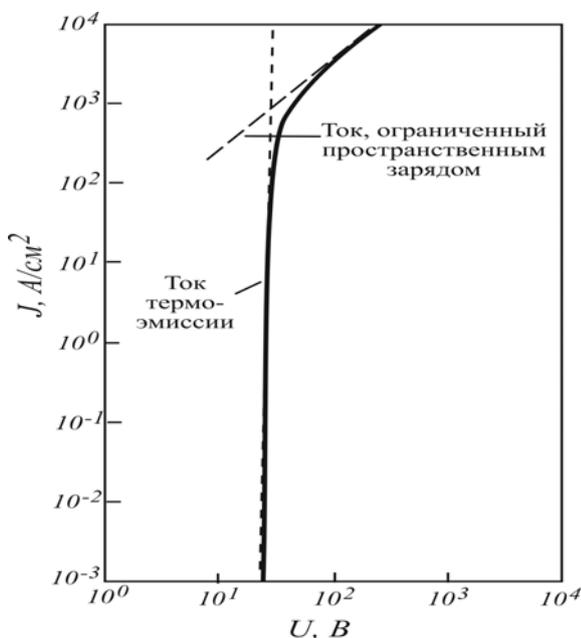


Рис. 2.13. Вольт-амперная характеристика $p^+ - n - p^+$ диода в условиях работы с «проколом» [3]: пунктир - теоретический результат, сплошная линия - эксперимент

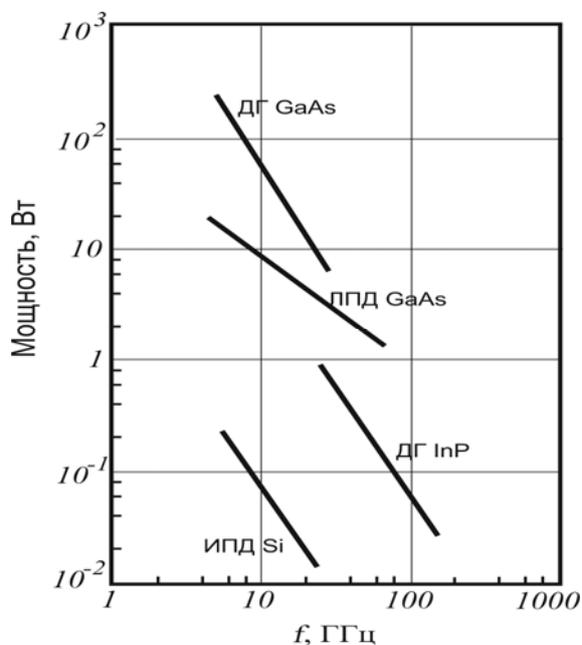


Рис. 2.14. Зависимость выходной мощности диодов Ганна (ДГ), ИПД и ЛПД от частоты генерируемого сигнала [3]

2.3. ДИОДЫ НА ОСНОВЕ ЭФФЕКТА ТУННЕЛИРОВАНИЯ. СВЕРХРЕШЕТКИ

2.3.1. Туннельный диод

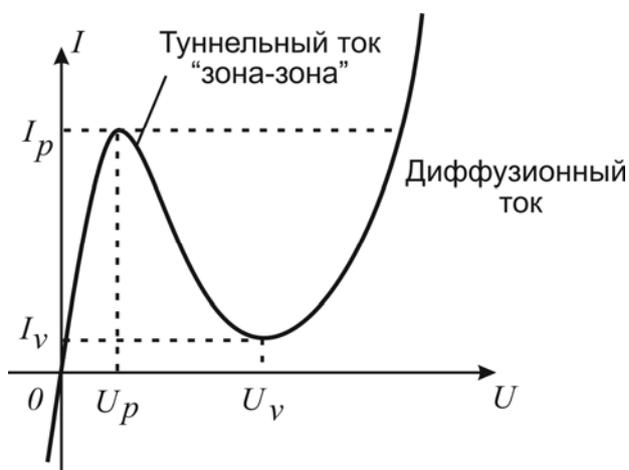
Туннельный диод представляет собой сильно легированный p^+-n^+ -переход с малой толщиной запирающего слоя. Первая статья о туннельном диоде (ТД) (называемом также *диодом Эсаки*) была опубликована в 1958 г [2]. В ней Лео Эсаки описал «аномальную» вольт-амперную характеристику, полученную в процессе изучения вырожденных германиевых $p-n$ переходов при прямом смещении. Аномальные характеристики некоторых $p-n$ переходов отмечались многими исследователями твердотельных приборов и до 1958 г., но эти данные подвергались критике, поскольку они не описывались привычным диодным уравнением. В классическом случае, как описано выше, ток через структуру определяется надбарьерным протеканием носителей заряда. Однако, в случае вырожденного полупроводника, когда концентрация примесей в областях с различными типами проводимости составляет 10^{18} - 10^{20} см^{-3} , толщина запирающего слоя столь мала (≤ 10 нм), что становится возможным туннелирование носителей из одной области в другую «сквозь» барьер. Эсаки впервые объяснил полученные вольт-амперные характеристики с помощью концепции квантового туннелирования носителей заряда и получил удовлетворительное соответствие между теорией и экспериментальными результатами.

Типичная статическая вольт-амперная характеристика туннельного диода показана на рисунке 2.15. На графике можно отметить следующие особенности. При приложении напряжения обратной полярности (отрицательный потенциал к p -области) ток монотонно возрастает. Таким образом, *туннельный диод*, в отличие от классического, *не обладает выпрямляющими свойствами*. При приложении прямого смещения ток сначала увеличивается до максимального значения I_p при напряжении U_p , а затем уменьшается до минимального значения I_V при напряжении U_V . При напряжениях, больших, чем U_V , ток ведет себя классическим образом, т.е. экспоненциально возрастает. Статическая характеристика складывается, в основном, из туннельного тока носителей между зонами и надбарьерного диффузионного тока (рис. 2.15).

Проведем качественное рассмотрение процесса туннелирования, используя упрощенную зонную структуру (рис. 2.16). Отметим, что, поскольку переход сильно легирован, электронно-дырочный газ в структуре является вырож-

денным, т.е. уровни Ферми p и n областей на зонной диаграмме находятся *внутри разрешенных зон*. В условиях теплового равновесия уровень Ферми постоянен по всей структуре (рисунок 2.16 б). Для простоты будем считать, что и с той, и с другой стороны p - n перехода выше уровня Ферми отсутствуют заполненные состояния, а ниже уровня Ферми отсутствуют свободные состояния (вообще говоря, это справедливо только при нулевой температуре, но качественной картины такое упрощение не изменит). Таким образом, в отсутствие внешнего напряжения область, содержащая свободные электроны и область, имеющая свободные уровни, не перекрываются по энергиям, т.е. туннельный ток в такой ситуации протекать не может.

Рис. 2.15. Вольт-амперная характеристика туннельного диода: I_p и U_p - пиковый ток и соответствующее напряжение; I_v и U_v - ток и напряжение в минимуме



При приложении внешнего смещения электроны могут туннелировать из валентной зоны в зону проводимости и наоборот (поэтому такой ток иногда называют током «зона – зона»). При этом носители преодолевают потенциальный барьер, который приближенно можно считать треугольным (рис. 2.16 а, треугольник серого тона). Следующие условия являются необходимыми для реализации процесса туннелирования:

- 1) наличие заполненных состояний с той стороны перехода, откуда электрон туннелирует;
- 2) наличие свободных уровней той же энергии с другой стороны перехода;
- 3) высота и ширина потенциального барьера должны быть достаточно малыми, чтобы существовала конечная вероятность туннелирования электрона сквозь барьер;
- 4) квазиимпульс в процессе туннелирования должен сохраняться.

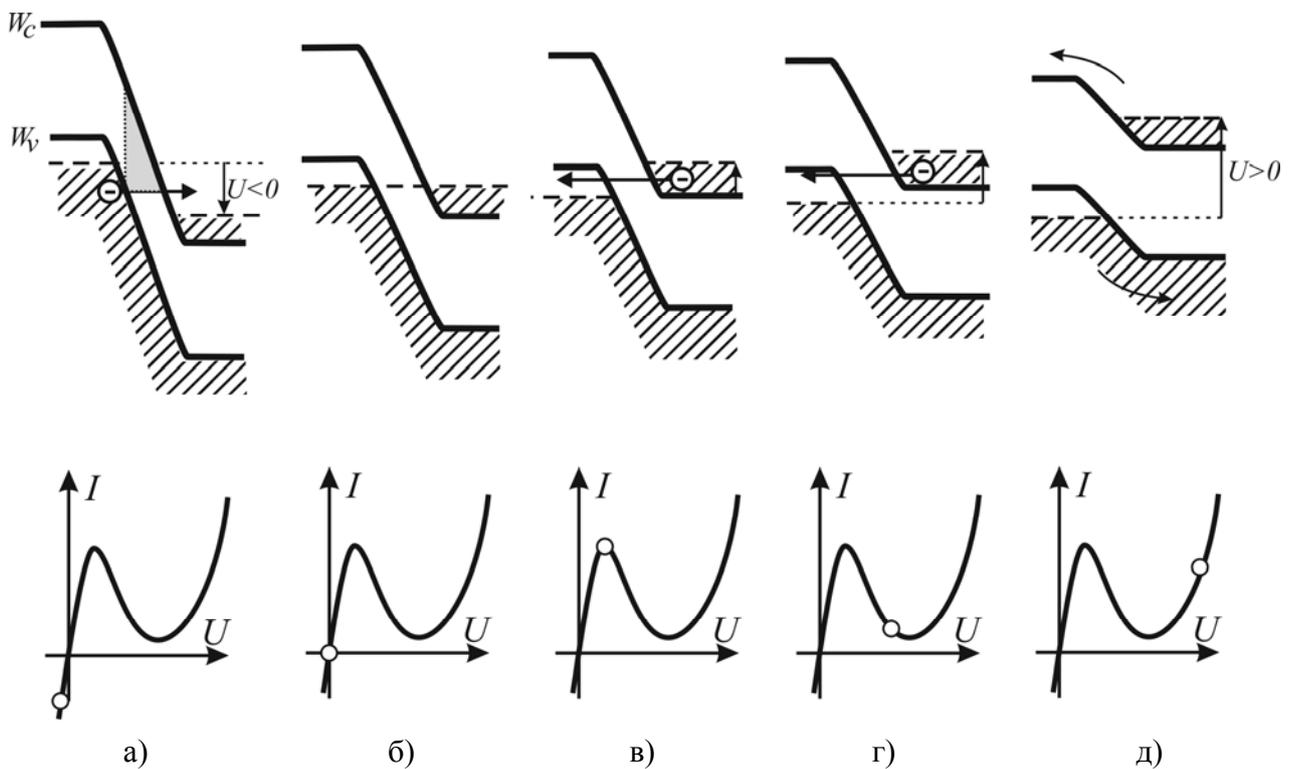


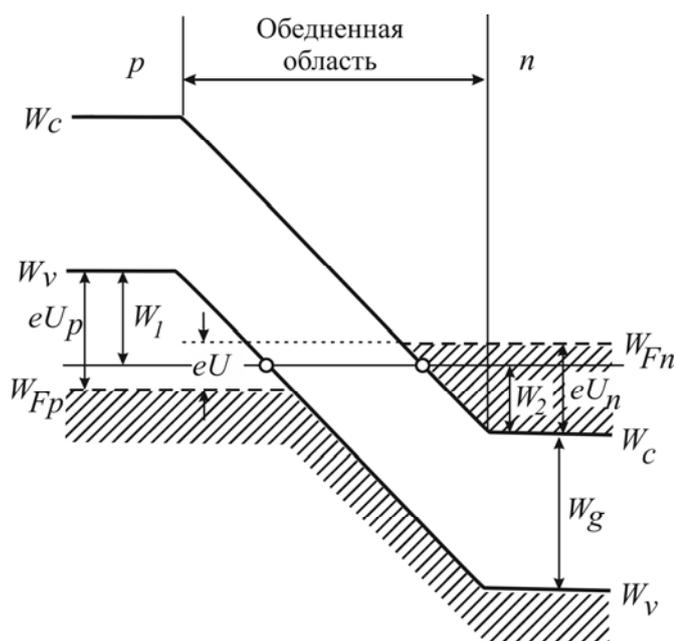
Рис. 2.16. Упрощенная зонная структура туннельного диода: а) - обратное смещение; б) термодинамическое равновесие; в) прямое смещение, при котором возникает пиковый ток; г) прямое смещение, при котором начинается «ток долины»; д) прямое смещение, при котором течет диффузионный ток [2]. Серым тоном выделен треугольник потенциального барьера, преодолеваемого электроном при туннелировании

На рисунке 2.16 а показано туннелирование электрона из валентной зоны в зону проводимости при приложении обратного смещения. Соответствующие этому случаю ток и напряжение отмечены точкой на вольт-амперной характеристике. В случае небольшого прямого смещения существует область значений энергии, для которой имеются заполненные состояния с n -стороны и соответствующие свободные состояния с p -стороны. Причем, когда заполненная и свободная области максимально перекрываются по энергиям, в структуре протекает наибольший ток (рисунок 2.16 в). При дальнейшем увеличении прямого смещения количество свободных уровней с p -стороны, имеющих ту же энергию, уменьшается (рисунок 2.16 г), а в запрещенной зоне состояний для электронов нет, т.е. переход туда невозможен. Если приложено прямое напряжение такой величины, что зоны «не перекрываются», т.е. дно зоны проводимости и вершина валентной зоны расположены на одном уровне, то в этом случае напротив заполненных состояний n -области лежат запрещенные состояния с p -области (наличием состояний в запрещенной зоне мы, для простоты,

пренебрегаем). В этой точке туннелирование носителей прекращается. Однако, полный ток при этом нулю не равняется, поскольку присутствует еще и классическая компонента, обусловленная надбарьерными переходами электронов и дырок. В точке минимума этот ток мал, поскольку потенциальный барьер для носителей еще достаточно высок. При дальнейшем увеличении приложенного напряжения инжекционный ток будет экспоненциально расти (рисунок 2.16 д).

С ростом прямого напряжения туннельный ток увеличивается от нуля до максимума, а затем уменьшается до нуля, когда $U=U_n+U_p$, где U - приложенное напряжение, U_n - степень вырождения с n -стороны, $U_n \equiv (W_{Fn} - W_c)/e$, U_p - степень вырождения с p -стороны, $U_p \equiv (W_v - W_{Fp})/e$, как показано на рисунке 2.17. Обратите внимание (рис. 2.15), что область ВАХ, расположенная между экстремумами тока, является областью отрицательного дифференциального сопротивления. Наличие данной области позволяет использовать туннельный диод для усиления и генерации СВЧ сигналов.

Рис. 2.17. Упрощенная энергетическая диаграмма с постоянным электрическим полем в обедненном слое. Степени вырождения с n - и p -сторон равны eU_n и eU_p . U – напряжение смещения. W_1 и W_2 – энергии электрона, отсчитанные от краев зон в p - и n -областях



Отметим, что туннелирование в структуре может быть прямым и непрямым. Случай прямого туннелирования изображен на рисунке 2.18 а, на котором график зависимости энергии от волнового числа $W(k)$ совмещен с картиной туннельного перехода в W - x координатах. Электрон может туннелировать из окрестности минимума функции, описывающей зависимость энергии от волнового вектора для зоны проводимости, в точку с аналогичным значением квазиимпульса в окрестности максимума функции $W(k)$ валентной зоны. Для того,

чтобы имело место прямое туннелирование, необходимо, чтобы минимум зоны проводимости и максимум валентной зоны имели одинаковое значение импульса. Это условие, как уже упоминалось выше, выполняется в прямозонных полупроводниках.

Непрямое туннелирование происходит тогда, когда абсолютный минимум зоны проводимости и максимум валентной зоны лежат в разных точках зоны Бриллюэна (рисунок 2.18 б). При этом разность импульсов передается рассеивающим агентам, например, фононам. Законы сохранения энергии и импульса при туннелировании с участием фононов: $W_0 + w_{\text{фон}} = W_1$, $p_0 + \hbar k = p_1$, где W_0 – первоначальная энергии электрона, туннелирующего из n -области в p -область, $w_{\text{фон}}$ – энергия фонона, W_1 – конечная энергия электрона после туннелирования, p_0 – начальное значение импульса электрона, $\hbar k$ – импульс фонона, p_1 – импульс электрона после туннелирования. Отметим, что вероятность многофононного непрямого туннелирования много меньше вероятности туннелирования с использованием одного фонона. Кроме того, вероятность непрямого туннелирования существенно меньше, чем вероятность прямого.

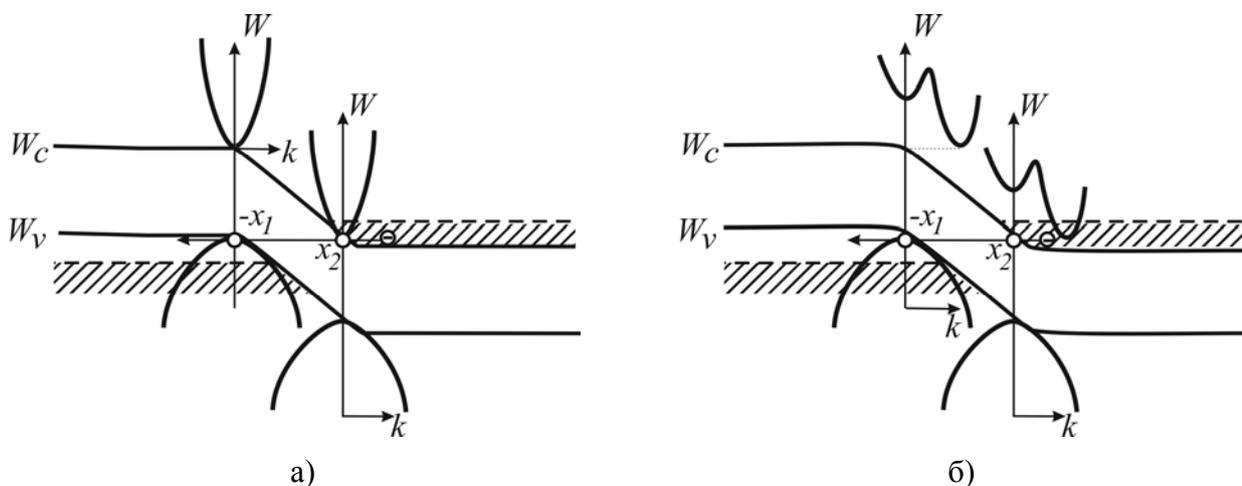


Рис. 2.18. Туннелирование: а) случай прямого туннелирования; б) случай непрямого туннелирования, когда $k_{\text{мин}} \neq k_{\text{макс}}$. Зависимость $W(k)$ наложена в характерных точках ($-x_1$ и x_2) на картину туннельного перехода в $W - x$ координатах

В заключение раздела приведем основную эквивалентную схему туннельного диода (рис. 2.19). Она состоит из четырех элементов: последовательной индуктивности L_S , последовательного сопротивления R_S , диодной емкости C и отрицательного сопротивления диода $-R$. Последовательное сопротивление

R_S включает сопротивление вывода, омических контактов и сопротивление кристалла.

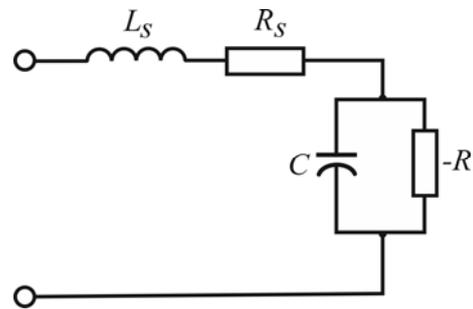


Рис. 2.19. Эквивалентная схема туннельного диода

2.3.2. Туннельно-резонансный диод

Структура туннельно-резонансного диода (ТРД) и его зонные диаграммы в равновесном случае и при смещении показаны на рис. 2.20.

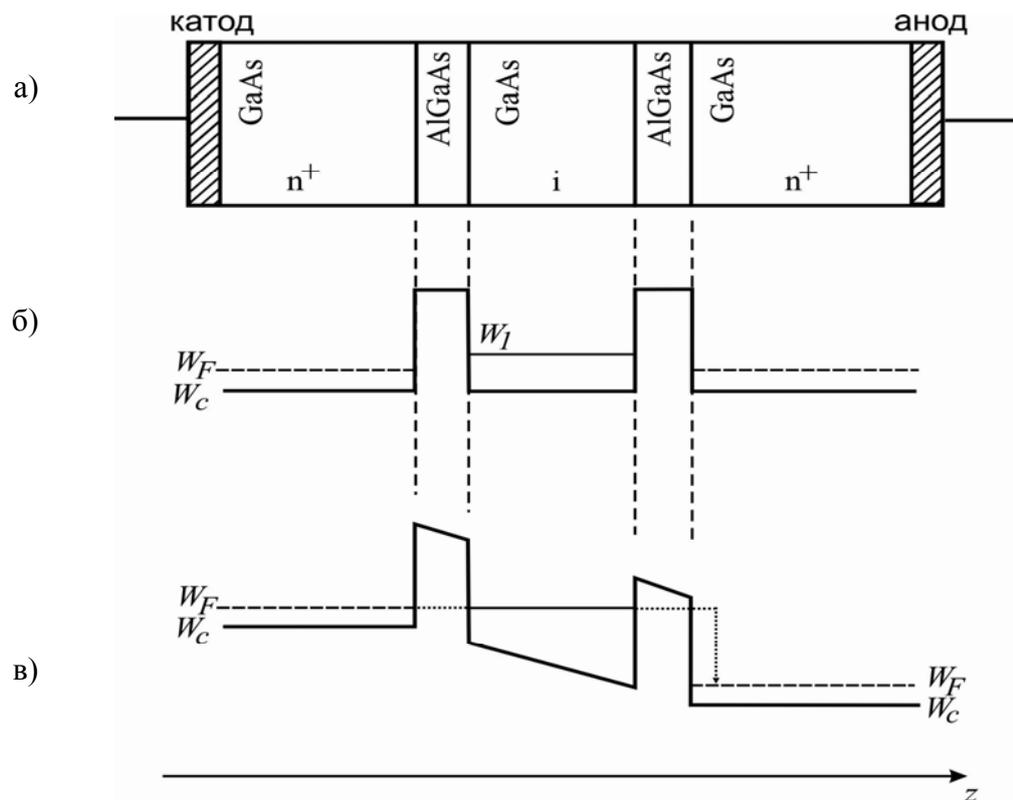


Рис. 2.20. Туннельно-резонансный диод на основе гетероструктуры GaAs/AlGaAs: а) структура диода; б) профиль дна зоны проводимости в отсутствие смещения; в) профиль дна зоны проводимости при подаче внешнего напряжения

ТРД состоит из трех слоев узкозонного полупроводника (на рисунке – GaAs), разделенных двумя слоями широкозонного материала (AlGaAs). Примыкающие к контактам сильно легированные области слева и справа от барьеров представляют собой «резервуары носителей заряда». Эти слои легируются настолько сильно, что электронный газ в таких областях становится вырожденным, а уровень Ферми лежит в зоне проводимости. Активная область диода состоит из двух потенциальных барьеров, между которыми расположена квантовая яма, содержащая один или несколько уровней для электронов (рис. 2.20 б). Прозрачность квантовой системы «барьер-яма-барьер» носит резонансный характер, откуда и берется название прибора. Если энергия налетающих носителей совпадает с энергией дискретного уровня в яме, то вероятность прохождения сквозь два барьера резко возрастает. Такая ситуация возникает при подаче напряжения на диод, когда квантовая яма опускается, как показано на рисунке 2.20 в и ток диода возрастает (см. рис.2.21 а). При напряжениях питания больше, чем высота барьера eU_0 происходит надбарьерная термоэмиссия электронов и ток увеличивается экспоненциально, как и в случае обычного *pn*-перехода.

Когда в яме существует несколько энергетических уровней, на вольт-амперной характеристике будет несколько резонансных пиков тока, каждый из которых соответствует своему уровню (см. рис.2.21 б). Если расстояние между соседними уровнями в яме меньше, чем расстояние от дна зоны проводимости до уровня Ферми в контактных слоях диода, то пики тока на ВАХ будут перекрываться, так как будет возможно одновременное туннелирование электронов через несколько уровней.

Рассмотрим процесс туннелирования более детально. Для того, чтобы происходило туннелирование электронов, необходимо чтобы выполнялись законы сохранения энергии и импульса электронов. Применительно к закону сохранения импульса речь идет только о проекции импульса на ось Z и соответствующей резонансной длине волны электрона, а также о продольной энергии, соответствующей указанной проекции импульса. Для выполнения закона сохранения энергии необходимо, чтобы полная энергия электрона соответствовала энергии уровня в яме.

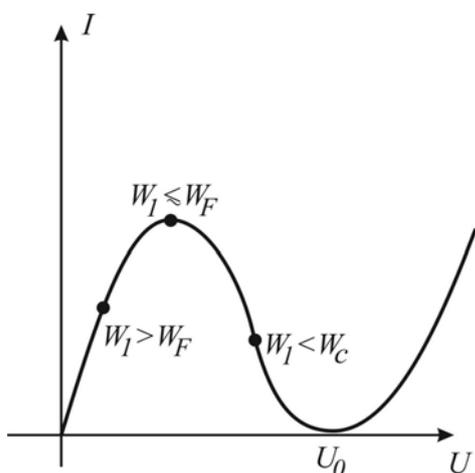


Рис. 2.21 а. ВАХ туннельно-резонансного диода с одним энергетическим уровнем в квантовой яме

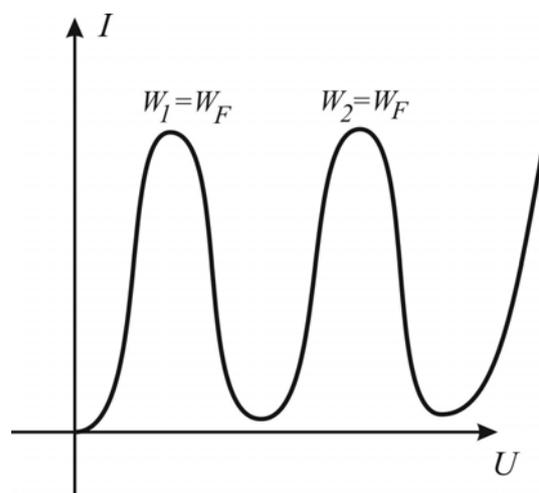


Рис. 2.21 б. ВАХ туннельно-резонансного диода с несколькими энергетическими уровнями в квантовой яме

Эквивалентная схема ТРД аналогична обычному туннельному диоду, но из-за большей толщины туннельной области (примерно полторы длины волны электрона, по сравнению с половиной длины волны для туннельного диода) емкость ТРД в 2...3 раза меньше, чем у ТД. Последнее позволяет увеличить предельную частоту работы ТРД.

Другой особенностью ТРД является наличие симметричной ВАХ, которая объясняется симметричностью конструкции диода. Это позволяет при использовании ТРД в качестве умножителя не подавать постоянное смещение на диод, что позволяет повысить КПД за счет снижения потерь, вызванных наличием цепей смещения по постоянному току. Кроме того в диоде с большим количеством пиков на ВАХ возможна реализация умножителей высокой кратности, которые будут давать умножение частоты сигнала в 4 или даже 6 раз.

2.3.3. Сверхрешётки

Сверхрешетками называют слоистые периодические структуры, образованные путем чередования тонких слоев полупроводников. По способу формирования сверхрешетки делятся на *композиционные* и *легированные*. Первые представляют собой чередующиеся слои полупроводников с различной шириной запрещенной зоны, вторые - чередующиеся слои *n*- и *p*-типов одного и того же полупроводника. В обоих случаях на электроны, кроме обычного потенциа-

ла кристаллической решётки, действует дополнительный одномерный периодический потенциал, называемый *потенциалом сверхрешетки*. Его период d меньше длины свободного пробега электронов, но значительно больше периода самой кристаллической решетки (1...10 нм). Таким образом, мы имеем периодическую систему квантовых ям, разделенных узкими, имеющими значительную туннельную прозрачность, потенциальными барьерами. В качестве примера на рисунке приведена композиционная сверхрешетка, изготовленная на основе широко используемой гетероструктуры GaAs-Al_xGa_{1-x}As.

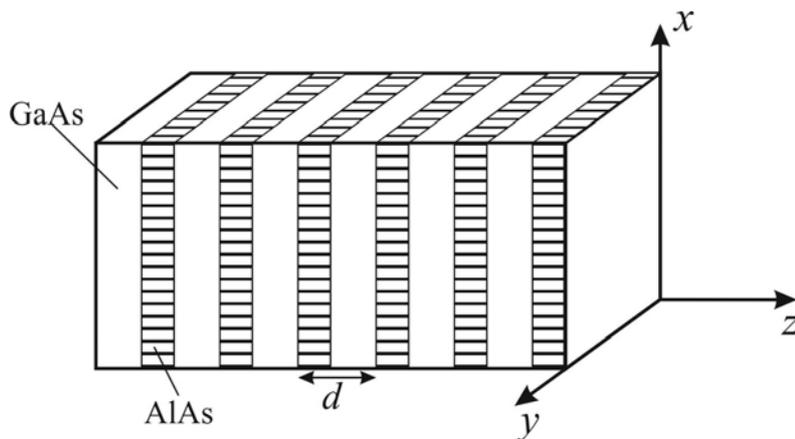


Рис. 2.22. Схематическая структура сверхрешетки GaAs/AlAs. Z – продольная ось сверхрешетки, вдоль которой пропускают ток

Наличие дополнительного периодического потенциала, зависящего только от одной координаты, делает энергетический спектр электронов и дырок резко анизотропным. Поскольку потенциальные барьеры туннельно-прозрачны, энергетические уровни в квантовых ямах расщепляются и возникает одна или несколько разрешенных энергетических *минизон* (рис. 2.23). При этом в плоскости, перпендикулярной оси сверхрешетки, носители ведут себя как свободные частицы с соответствующей эффективной массой. Таким образом, сверхрешетка является искусственным одномерным кристаллом.

Ширина разрешенной минизоны может быть оценена по формуле $\Delta W_m \sim \frac{1}{d} \approx 0.05 \dots 0.5 \text{ эВ}$. Аналогичный расчет для ширины разрешенной зоны объемного полупроводникового материала с постоянной кристаллической решетки c дает значение $\Delta W_c \sim \frac{1}{c} \approx 3 \dots 10 \text{ эВ}$, т.е. на порядок больше, чем у сверхрешетки.

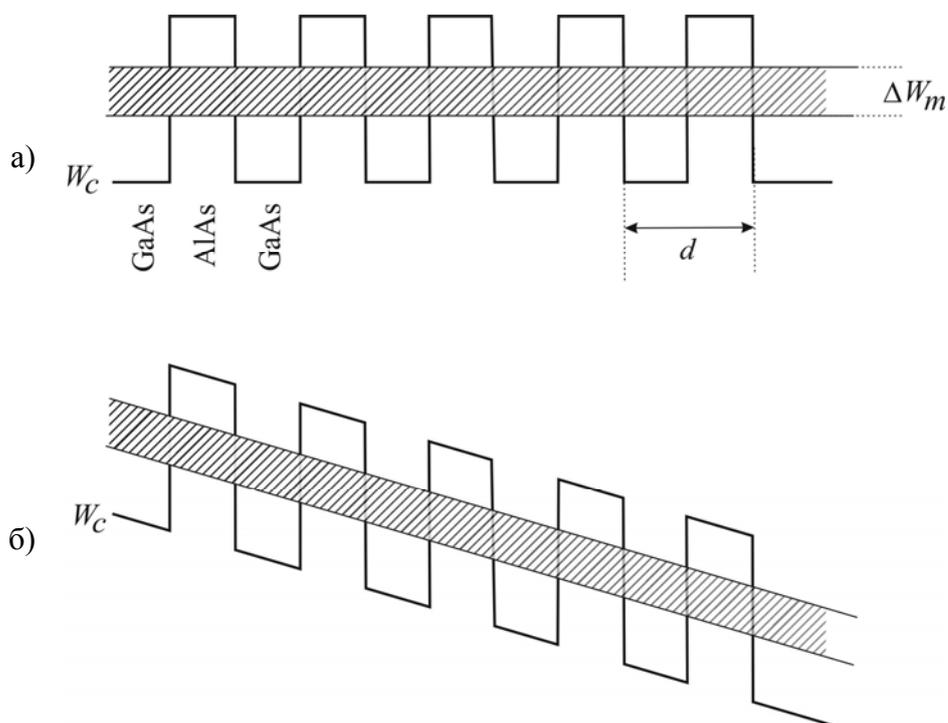


Рис. 2.23. Зонная диаграмма сверхрешетки: а) в отсутствие внешнего напряжения; б) при подаче внешнего напряжения. ΔW_m – ширина минизоны

При подаче на структуру внешнего напряжения минизона наклоняется (рис. 2.23 б) и в ней протекает ток электронов, как и в обычных разрешенных зонах полупроводника. При этом возможна реализация двух вариантов движения электронов.

1) При малых напряжениях носители двигаются вдоль дна минизоны, периодически рассеиваясь на фонах и ионах примесей, как и в обычной разрешённой зоне объемных полупроводников.

2) При больших напряжениях ситуация меняется: когда расстояние Δx (зависящее от внешнего напряжения) становится меньше длины свободного пробега, электрон испытывает колебания в пространстве, отражаясь от потолка минизоны. Смещение его в пространстве возможно только после того, как произойдёт рассеяние. В результате дрейфовая скорость таких электронов будет ниже, чем в первом случае. Такие колебания носят название *блоховских осцилляций*.

Блоховские осцилляции возникают благодаря тому, что в нижней половине минизоны масса электронов положительна, а в верхней – отрицательна, также как и в обычных полупроводниковых кристаллах. В результате электрон отражается от потолка разрешённой зоны, хотя электрическое поле всегда направлено в одну сторону. В нижней половине минизоны электрон ускоряется

электрическим полем, а в верхней – тормозится и разворачивается по полю. Период блоховских осцилляций должен быть меньше времени свободного пробега электронов, иначе осцилляций не будет, а будет только рассеяние, как это показано на рисунке 2.24 а.

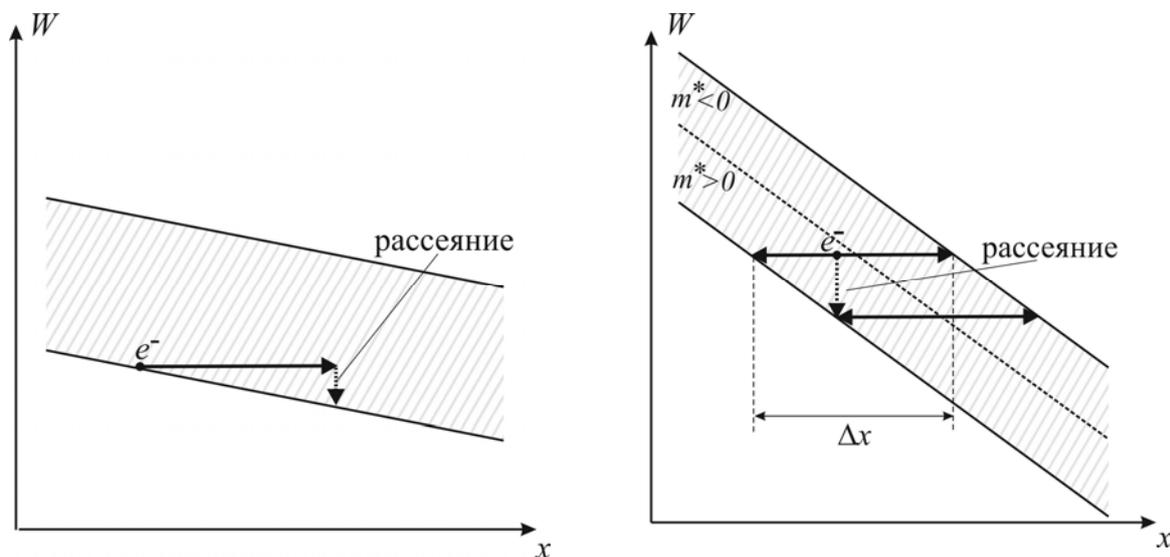


Рис. 2.24. Схема движения электронов в минизоне сверхрешетки: а) при малых напряжениях питания; б) при больших напряжениях питания. Пунктиром со стрелкой изображен процесс рассеяния, переводящий электрон на более низкий энергетический уровень

Основной проблемой реализации генераторов терагерцовых частот на основе процесса блоховских осцилляций является наличие в структуре двух сортов электронов с разной дрейфовой скоростью. В малых электрических полях дрейфовая скорость электронов в сверхрешетках велика, а в больших она уменьшается из-за наличия блоховских осцилляций. Поэтому в сверхрешётках возникают колебания тока, аналогичные колебаниям в диоде Ганна, с образованием доменов, похожих на домены Ганна. Этот эффект в сверхрешетках является паразитным и ограничивает частоту колебаний на уровне предельных частот указанных диодов, т.е. порядка 200 ГГц.

В последние годы было предпринято много попыток подавить паразитные колебания ганновского типа путем уменьшения длины сверхрешеток, использования аперриодических решеток и/или их освещения оптическим излучением. Несмотря на некоторые успехи до сих пор надежного генератора на основе эффекта блоховских осцилляций не создано.

ЧАСТЬ 3

ОРГАНИЗАЦИЯ ПРОЕКТИРОВАНИЯ ДИОДОВ С ИСПОЛЬЗОВАНИЕМ ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ

В качестве пособий для изучения параллельного программирования читателю можно порекомендовать издания [10-24].

3.1 МЕТОДЫ МОДЕЛИРОВАНИЯ ФИЗИЧЕСКИХ ПРОЦЕССОВ В ДИОДАХ

Одним из основных методов моделирования переноса электронов в полупроводниках основан на *квазигидродинамическом приближении*. Этот метод сводится к решению следующей системы уравнений:

$$\begin{aligned}\Delta V &= \frac{e}{\varepsilon_s} (n(W) - N) \\ \frac{\partial m(W) \bar{v}(W)}{\partial t} &= -e \bar{E} - \frac{m(W)}{\tau_p(W)} \bar{v}(W) \\ \frac{\partial n(W)}{\partial t} &= \frac{1}{e} (\nabla \cdot \bar{j}) \\ \frac{\partial (Wn(W))}{\partial t} &= (\nabla \cdot \bar{j}_w) + (\bar{j} \cdot \bar{E}) - \frac{n(W)(W - W_0)}{\tau_w(W)} \\ \bar{j} &= -en \bar{v}(W) + e \nabla (D(W)n(W)) \\ \bar{j}_w &= -n(W)W \bar{v}(W) + \nabla (D(W)n(W)W) \\ \bar{j}_t &= \bar{j} + \frac{\partial \bar{E}}{\partial t} \\ \bar{E} &= -\nabla V\end{aligned}\tag{3.1}$$

где: W – энергия носителей заряда, m – масса носителей заряда, v – скорость носителей заряда, τ_p – время релаксации импульса, n – концентрация носителей заряда, \bar{j}_w – поток энергии носителей заряда, \bar{j} – поток носителей заряда, τ_w – время релаксации энергии, D – коэффициент диффузии, \bar{j}_t – ток смещения.

Первое уравнение определяет связь потенциала с пространственным распределением зарядов. Второе уравнение представляет собой закон сохранения импульса. Третье – это закон сохранения числа частиц в системе. Четвёртое – закон сохранения

энергии и показывает, что энергия электронов, запасённая в единице объёма, может изменяться за счёт втекания или вытекания частиц с различной энергией из данного объёма вследствие их движения в пространстве, увеличиваться за счёт внешнего электрического поля и уменьшаться из-за передачи энергии кристаллической решётке при столкновениях. Последние два уравнения говорят о том, что перенос частиц и их энергии обусловлен двумя процессами: дрейфом под действием электрического поля и диффузией вследствие разности концентраций и неравномерного распределения энергии частиц.

Для решения систем нелинейных уравнений в частных производных применяются методы конечных разностей, а для двумерных областей также и метод конечных элементов. Метод конечных элементов сложен в реализации и не даёт выигрыша в точности и времени решения. Поэтому в данной работе будет применяться метод конечных разностей.

В основе решения уравнений методом конечных разностей лежит конечно-разностная аппроксимация производных. Разностные методы позволяют получить систему алгебраических уравнений, решение которой аппроксимирует искомое решение начально-краевой задачи. Аппроксимация осуществляется в три этапа:

- в области аргументов x, y, t строится пространственно-временная сетка, соответствующая характеру задачи и граничным условиям;

- систему уравнений в частных производных записывают в наиболее удобной системе координат, все функции считают определёнными в узлах сетки и дифференциальные операторы заменяются разностными аналогами. Полученную алгебраическую систему уравнений используют в дальнейшем для описания функциональной связи между соседними узлами сетки;

- формируются разностные аналоги граничных и начальных условий.

В дальнейшем полученную алгебраическую систему уравнений относительно значений неизвестных функций в узлах сетки решают одним из численных методов. Выбор пространственной сетки зависит от формы и характера области и распределения легирующей примеси в структуре прибора. Пространственная сетка в значительной степени определяет точность решения задачи и затраты машинного времени. Сетка может быть как равномерной, так и иметь переменный пространственный шаг.

Дискретизация временной оси может быть равномерной и неравномерной. Выбор временного шага диктуется в одних случаях условиями устойчивости разностной схемы, в других необходимой точностью решения. Использование переменного

временного шага, как правило, необходимо для сокращения затрат времени на ЭВМ при заданной точности решения.

Как показано в [12], применение конечно-разностной аппроксимации к уравнениям плотностей токов приводит к численной расходимости, когда падение напряжения на одной пространственной ячейке превышает $2kT/e$. Это ограничение вынуждает пользоваться мелкой расчётной сеткой, и затраты машинного времени возрастают. Обойти ограничение позволяет конечно-разностная аппроксимация уравнений плотностей токов, предложенная Гумелем и Шарфеттером. Этот метод можно обобщить для уравнений плотностей тока и потока энергии в квазигидродинамической модели.

Разностная аппроксимация граничных условий на омических и барьерных контактах сводится к заданию соответствующих концентраций, потенциалов и энергий на поверхностях контактов.

После построения разностной схемы получаем систему нелинейных уравнений с большим числом неизвестных, и возникает проблема выбора метода её решения. Для решения систем нелинейных уравнений применяются итерационные методы.

Метод простой итерации для решения нелинейной системы уравнений является развитием метода итерации для одного уравнения. Он основан на допущении, что систему уравнений можно привести к виду:

$$X_i^* = f(X_1, X_2, \dots, X_N) \quad (3.2)$$
$$i = 1 \dots N$$

Алгоритм решения состоит в следующем. Выбирают некоторые исходные значения X_i и последовательно вычисляют новые значения X_i^* , при этом в каждом следующем уравнении все переменные заменяют уже вычисленными из предыдущих уравнений новыми значениями. Значения X_i^* сравнивают с предыдущими X_i и если каждая переменная изменилась лишь на допустимую малую величину, счёт прекращается. Если же изменение слишком велико, процесс повторяется до тех пор, пока все переменные не станут слабо меняться от итерации к итерации. Значения X_i^* используются при этом в качестве новых исходных значений неизвестных.

Рассмотренный метод обладает рядом недостатков. Прежде всего, метод имеет очень узкую область сходимости по начальным значениям неизвестных. Вторым важным недостатком является медленная сходимость метода. Число итераций,

необходимое для получения решения с приемлемой точностью превышает число узлов сетки в несколько раз.

Широкая распространённость метода Ньютона обусловлена тем, что по сравнению с методом итераций, он обеспечивает гораздо более быструю сходимость.

Одновременное решение уравнений, входящих в систему, является очень трудоёмкой задачей. Поэтому часто пользуются методами, в которых уравнения решаются последовательно. Основа таких методов - так называемый алгоритм Гуммеля [13].

Сначала решается уравнение Пуассона в предположении известного распределения плотности заряда. Полученные значения узловых потенциалов подставляют в уравнение непрерывности, которое затем решается относительно концентраций. Этот процесс повторяется до тех пор, пока для всех переменных значения в узлах не будут меняться в пределах заданной точности. Один временной шаг требует следующей последовательности вычислений:

1. вычисление плотностей токов и потоков энергии по известным значениям потенциалов, концентраций и энергий в узлах сетки;
2. вычисление новых распределений концентрации и энергии электронов;
3. решение уравнения Пуассона, нахождение распределения потенциала;
4. вычисление полных токов, текущих через электроды.

Явная схема обладает весьма малыми затратами времени на ЭВМ на один временной шаг, поскольку концентрации носителей на $i+1$ шаге находятся непосредственным вычислением. Однако, схема оказывается условно устойчивой по отношению к величине временного шага, то есть шаг должен быть достаточно малым для того, чтобы процесс вычислений сходился к решению. Временной шаг тем меньше, чем меньше размер пространственных ячеек и больше градиенты концентрации активной примеси.

Уравнение Пуассона относится к эллиптическому типу. Методы его численного решения весьма разнообразны и довольно хорошо разработаны. Все имеющиеся методы решения могут быть разбиты на три большие группы. Это методы сеточной релаксации, матричные методы и методы, основанные на быстром преобразовании Фурье. В двумерном случае уравнение Пуассона, записанное в конечно-разностном виде, представляет собой пятидиагональную систему алгебраических уравнений. Эта система уравнений решается методом Гаусса-Зайделя. В отдельных случаях, когда число узлов сетки невелико, можно использовать прямые методы для решения уравнения Пуассона.

Поскольку моделирование происходит в условиях высоких градиентов концентрации, потенциала и энергии носителей заряда, то расчётная сетка в области затвора сильно сгущается (рис. 3.1), что приводит к резкому увеличению времени счёта.

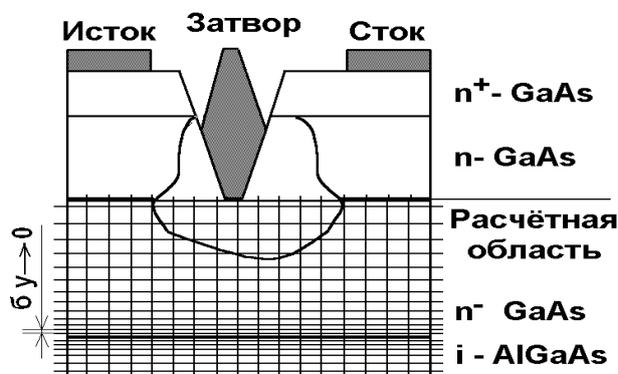


Рис. 3.1. Расчётная сетка прибора

Стандартным методом численного решения уравнения Больцмана является метод Монте-Карло. Он основан на подходе, предложенном в работе [9]. Идея этого

подхода состоит в том, чтобы моделировать движение электрона в \mathbf{k} -пространстве.

Другим возможным методом моделирования потока электронов в полупроводниковых диодах является *метод Монте-Карло*. Рассматривается пролёт свободного электрона [15], прерываемый процессами рассеяния (рис. 3.2). Если следить за одним электроном достаточно долгое время, то распределение времён, которые электрон проводит в окрестностях разных точек \mathbf{k} -пространства, воспроизводит функцию $f(\mathbf{k})$.



Рис. 3.2. Траектория движения электронов

Между актами рассеяния изменение волнового вектора электрона определяется электрическим полем \mathbf{E} :

$$\vec{k}(t) = \vec{k}_0 + \frac{q\vec{E}}{\hbar}t, \quad (3.3)$$

где: k_0 - начальное значение волнового вектора.

Моменты времени, в которые происходят акты рассеяния, определяются случайным числом r , генерируемым в ЭВМ датчиком случайных чисел.

Эти моменты времени зависят от полной частоты рассеяния

$$\lambda(\vec{k}) = \sum_{i=1}^n \lambda_i(\vec{k}), \quad (3.4)$$

где: $\lambda_i(\mathbf{k})$ - частота для разных механизмов рассеяния.

Полная частота $\lambda(\mathbf{k})$ является сложной функцией волнового вектора $\mathbf{k}(t)$, что создаёт трудности при попытке воспроизвести распределение вероятностей актов рассеяния с помощью датчика случайных чисел. Эти трудности могут быть преодолены путём включения в рассмотрение фиктивного процесса (так называемого саморассеяния), который не изменяет волновой вектор. Вероятность рассеяния в таком процессе имеет вид

$$W(\vec{k}, \vec{k}') = \lambda_0(\vec{k}) \delta(\vec{k} - \vec{k}'). \quad (3.5)$$

Функция $\lambda_0(\vec{k}) + \lambda(\vec{k})$ выбирается таким образом, чтобы $\lambda_0(\vec{k}) + \lambda(\vec{k}) = \Gamma$, где Γ - постоянная, а $\lambda_0(\mathbf{k})$ - положительная величина.

Тогда полная вероятность рассеяния электрона в интервале времени от t до $t+dt$ равна

$$P(t)dt = \Gamma e^{-\Gamma t} dt, \quad (3.6)$$

где: t - время, прошедшее после предыдущего акта рассеяния.

Время свободного пролёта t_s можно связать со случайным числом r , равномерно распределённым от 0 до 1, с помощью соотношения

$$r = \int_0^{t_s} P(t) dt, \quad (3.7)$$

Или

$$r = 1 - e^{-\Gamma t_s}, \quad (3.8)$$

Откуда

$$t_s = -\frac{1}{\Gamma} \ln(1 - r). \quad (3.9)$$

где: t - время, прошедшее после предыдущего акта рассеяния.

После того как вычислено время свободного пролёта t_s , необходимо установить процесс рассеяния, ответственный за прерывание пролёта. Поскольку вероятность прерывания пролёта i -м процессом рассеяния пропорциональна $\lambda_i(\vec{k})$ и $\sum_{i=1}^n \lambda_i(\vec{k}) = \Gamma$, следует лишь сгенерировать случайное число s между 0 и Γ и проверить выполнение неравенств

$$s < \sum_{i=1}^m \lambda_i(\vec{k}), \quad (3.10)$$

для $m = 0, 1, 2, \dots, n$. Здесь n - число реальных механизмов рассеяния.

Если это неравенство удовлетворяется при $m = 0$, считается, что реализовался процесс саморассеяния. Если неравенство удовлетворяется при $m = 1$ (числа m последовательно возрастают), то полагают, что сработал первый механизм рассеяния и т.д.

Следующий шаг состоит в определении конечного состояния после рассеяния. Предполагается, что электрон в результате рассеяния исчезает из точки \mathbf{k} фазового пространства и мгновенно появляется вновь в точке \mathbf{k}' , соответствующей конечному состоянию. Это новое положение \mathbf{k}' является начальной точкой для следующей фазы свободного пролёта, то есть \mathbf{k}' заменяет \mathbf{k}_0 в выражении (5).

Для саморассеяния конечное состояние, естественно, известно, так как при этом волновой вектор \mathbf{k} не изменяется.

Для всех же реальных процессов рассеяния необходимо генерировать дополнительные случайные числа, чтобы смоделировать распределение вероятностей конечных состояний. Например, для акустического рассеяния в параболической долине или междолинное рассеяние, все состояния на энергетической поверхности, соответствующей конечному состоянию, равновероятны. Следовательно, вероятность того, что угол между вектором \mathbf{k}' и направлением электрического поля \mathbf{E} равен Θ , пропорциональна числу состояний на кольце радиуса $|\vec{k}'| \sin \theta$ (Рис. 3.11):

$$P(\theta)d\theta = \frac{1}{2} \sin \theta d\theta, \quad (3.11)$$

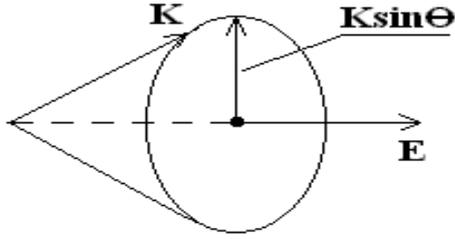


Рис. 3.3. Выбор конечного состояния. Число конечных состояний с одинаковой энергией ($k' = \text{const}$) и одинаковым углом между вектором k' и вектором электрического поля E пропорционально длине окружности $2\pi k' \sin\Theta$.

Здесь угол Θ изменяется от 0 до π , и множитель $1/2$ является нормирующей постоянной. Для определения Θ генерируется случайное число s_1 , равномерно распределённое от 0 до 1 и соотношение $s_1 = \int_0^\theta P(\theta) d\theta$ даёт распределение вероятности конечного состояния, определяемого предыдущим выражением. Подставляя одно выражение в другое, получим, что $s_1 = 1/2(1 - \cos\theta)$.

Для рассеяния на акустических фононах в непараболической долине, такой, как Γ -минимум в GaAs, и для рассеяния на полярных оптических фононах следует использовать более сложную процедуру [14]. Сначала рассчитывается распределение вероятностей $P(\cos\Theta) = P(\Theta)/\sin\Theta$, где Θ - угол между векторами k и k' . Затем генерируются два случайных числа s_2 и s_3 , причём s_2 равномерно распределено между -1 и 1, а s_3 - между 0 и $P(1)$. Если s_3 меньше, чем $P(s_2)$, то в качестве значения $\cos\Theta$ выбирается s_2 . Если же s_3 больше, чем $P(s_2)$, то генерируется вторая пара случайных чисел s_2 и s_3 , и процесс повторяется до тех пор, пока s_3 не окажется меньше, чем $P(s_2)$. Как показано в работе [14], получаемые значения s_2 отвечают требуемому распределению вероятностей.

При рассмотрении GaAs следует учитывать три набора долин, соответствующих Γ -, L - и X - минимумам. Для расчёта функций распределения в Γ -, L - и X - долинах строятся гистограммы в пространстве (k_z, k_p) , где оси k_z и k_p направлены соответственно параллельно и перпендикулярно направлению электрического поля, и фиксируются времена, которые электрон проводит в каждой ячейке гистограммы. Некоторые величины, например, дрейфовая скорость, рассчитываются прямо по начальным и конечным значениям волнового вектора k для каждого свободного пролёта, например, средняя дрейфовая скорость v_d в долине j имеет вид [14]

$$v_j = \frac{1}{\hbar k_j} \sum \varepsilon_f - \varepsilon_i, \quad (3.12)$$

$\therefore \varepsilon_f$ и ε_i - энергии конечного и начального состояний для каждого свободного пролёта, k_j - полная длина траектории электрона в долине j в \mathbf{k} -пространстве, равная $k_j = eE_jT_j/h$, T_j - полное время, которое электрон проводит в этой долине.

Суммирование в выражении (12) ведётся по всем свободным пролётам в долине j .

Этапы описанного алгоритма Монте-Карло:

1. генерируется случайное число r и определяется время свободного пролёта;
2. регистрируется время, проводимое электроном в каждой ячейке \mathbf{k} -пространства в течение свободного пролёта, вычисляются дрейфовая скорость, средняя энергия и т.д.;
3. генерируются случайные числа, определяющие механизм рассеяния и конечное состояние;
4. этапы повторяются до тех пор, пока не наберётся запрограммированное число актов рассеяния;
5. рассчитываются функция распределения, дрейфовая скорость, средняя энергия и т.д.

Естественно, что для оптимизации конструкции прибора необходима максимально возможная детализация траекторий движения носителей заряда, а это приводит к увеличению числа итераций и, соответственно, времени счёта.

В данной работе предлагается использовать одночастичный метод Монте-Карло совместно с квазигидродинамической моделью. Благодаря использованию рассчитанного при помощи квазигидродинамики распределения электрического поля, возможно (при сохранении точности вычислений) применение одночастичного метода Монте-Карло, в результате чего время расчёта сильно уменьшается, по сравнению с использованием многочастичного метода Монте-Карло.

Поскольку времени на моделирование отводится обычно немного, и для оптимизации конструкции прибора вычисления проводятся довольно большое количество раз (от $10^2 \dots 10^5$), то применение для расчётов комбинации методов Монте-Карло и квазигидродинамики оказывается целесообразным. Необходимо также учесть различие во времени счёта методами Монте-Карло и квазигидродинамики. Отметим, что для числа итераций более 10^3 необходимо применение суперкомпьютерных технологий, а при $10^1 \dots 10^2$ итераций возможно использо-

вание локальной вычислительной сети, что и рассматривается в данном случае. Следует понимать, что трудоемкость процедуры моделирования существенно различается в зависимости от числа узлов расчетной сетки, размерности задачи, необходимости моделирования тепловых процессов в приборах и т.п. Поэтому применению суперкомпьютерных технологий чаще всего оправдано, однако в каждом конкретном случае необходимо четко выделять те процедуры оптимизации, которые могут быть выполнены на локальной вычислительной сети. Как говорится «из пушки по воробьям не стреляют». Ниже речь пойдет именно о такой задаче, поэтому в качестве вычислительного инструмента будет фигурировать вычислительная сеть. Однако обсуждаемые ниже методы параллельных вычислений с успехом можно применять и при использовании суперкомпьютеров.

3.2 ОРГАНИЗАЦИЯ ВЫЧИСЛЕНИЙ

Поскольку задача моделирования полупроводниковых приборов распадается на две практически независимых подзадачи (методы Монте-Карло и квазигидродинамики), которые требуют довольно редкого обмена небольшими массивами данных, то можно использовать слабосвязанную вычислительную систему [11] на основе 3-5 компьютеров (рис. 3.4)



Рис.3.4. Слабосвязанная вычислительная система

Так как подзадачи обмениваются данными, размером до 1 мегабайта, 1 раз в 1...10 минут, то требования к быстродействию сети предъявляются самые минимальные⁷.

В работе использование известных методов решения на основе кластерных систем и алгоритмов для решения поставленной задачи признано нерациональным из-за малого объёма передаваемых данных, малого числа используемых ЭВМ и невысоких требований к увеличению быстродействия (2-3 раза). Использование двухпроцессорной ЭВМ признано нерациональным, так как выигрыш по времени счёта (как показано ранее), по сравнению с однопроцессорной ЭВМ, составляет всего 10-30%. Наилучшим решением данной задачи является использование локальной вычислительной сети лаборатории в качестве кластерной вычислительной системы, что требует разработки специального алгоритма. Использование возможностей суперкомпьютера предпочтительно при решении набора указанных задач, например при проведении оптимизации параметров интегральных схем на основе дидов.

Все существующие вычислительные системы можно условно поделить [10] на системы с общей памятью (SMP) и системы с распределенной памятью (кластеры). Сферы их применения можно представить следующим образом (рис. 3.5):



Рис. 3.5. Области применения вычислительных систем (шкала условная, логарифмическая)

⁷ Для задачи, при решении которой используется минимально возможная расчетная сетка 20x30 узлов. При использовании сеток с количеством узлов более 10^4 выгодным становится использование суперкомпьютеров и/или соответствующих вычислительных сетей.

Под «научными», понимается класс задач, который требует степени связанности вычислений меньшей, нежели базы данных, но большей, чем распределенные слабосвязанные вычисления.

Почти в любой ориентированной на параллельное вычисление задаче невозможно избежать необходимости передавать данные от одной подзадачи другой. Соответственно, быстродействие кластерной системы определяется в основном быстродействием узлов и связей между ними. Причем влияние скоростных параметров этих связей на общую производительность системы зависит от характера выполняемой задачи. Если задача требует частого обмена данными с подзадачами, тогда быстродействию коммуникационного интерфейса следует уделять максимум внимания. Естественно, чем меньше взаимодействуют части параллельной задачи между собою, тем меньше времени потребуется для ее выполнения.

Основные проблемы при необходимости обмена данными между подзадачами возникают в связи с тем, что быстродействие передачи данных между центральным процессором и оперативной памятью узла значительно превышает скоростные характеристики систем межкомпьютерного взаимодействия.

Далее приведены результаты тестирования скорости работы однопроцессорных и двухпроцессорных ЭВМ на задаче кодирования аудио- и видеoinформации. Задачи такого класса наиболее соответствуют численным методам решения систем дифференциальных уравнений типа (3).

По результатам тестирования [11] видно (рис. 3.6), что вычислительные системы на базе процессоров Athlon MP 1900+ заметно уступают по производительности системам на базе процессоров Xeon 1700. Несомненное превосходство имеют системы на базе процессоров Xeon 2,2 ГГц. Также можно заметить, что выигрыш от использования любых двухпроцессорных систем составляет всего 10-30%.

Поскольку время моделирования составляет десятки часов, то даже небольшое увеличение производительности даст возможность провести вычисления за один рабочий день. В процессе оптимизации конструкции прибора необходимо проводить несколько циклов моделирования, и возможность провести один цикл вычислений за один рабочий день имеет важное значение для организации всего комплекса работ по проектированию приборов.

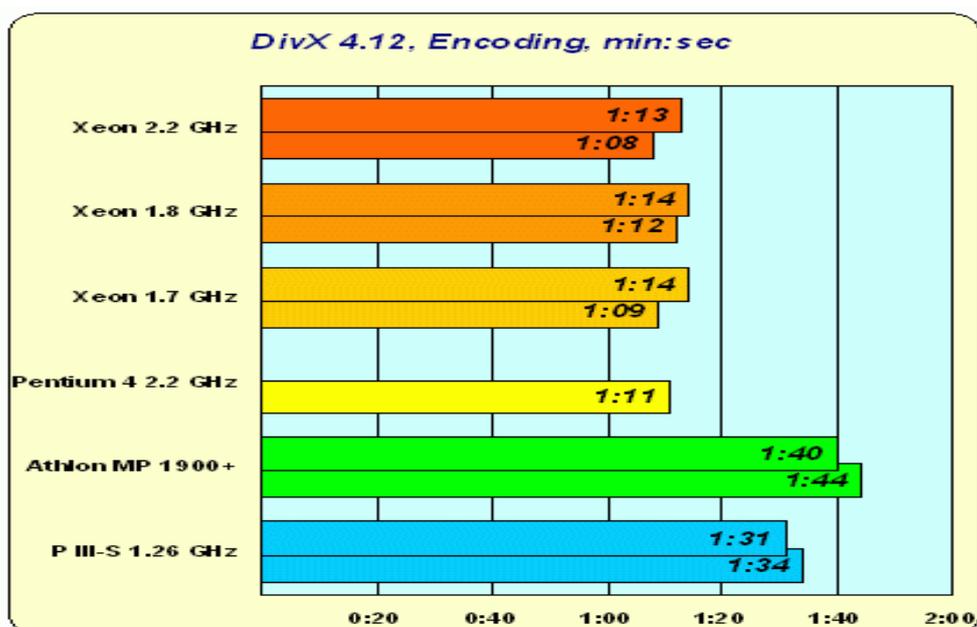


Рис. 3.6. Времена счёта однопроцессорных и двухпроцессорных ЭВМ для различных типов процессоров

Важным является тот факт, что при введении в структуру моделируемого прибора гетеропереходов возникают резкие градиенты концентрации, потенциала и энергии носителей заряда, что приводит к увеличению времени счёта используемых численных методов решения уравнений квазигидродинамики.

Многочастичный метод Монте-Карло рассчитывает самосогласованное движение множества зарядов в электрическом поле, которое сами движущиеся заряды модифицируют, поэтому время решения задачи таким методом примерно на порядок превышает характерные времена вычислений методом квазигидродинамики. Одночастичный метод Монте-Карло рассчитывает траекторию движения одного электрона по известному распределению потенциала электрического поля. Если использовать полученное из квазигидродинамики распределение потенциала электрического поля в качестве исходных данных для метода Монте-Карло, то возможно применение одночастичного метода Монте-Карло, что существенно экономит вычислительные затраты.

Такое объединение предложенных методов, при сохранении точности моделирования, даст уменьшение времени расчётов, по сравнению с использованием многочастичного метода Монте-Карло. Если использовать для расчётов слабосвязанную вычислительную систему на основе нескольких ЭВМ, то время моделирования уменьшится ещё в несколько раз. Это требует разработки специального алгоритма, который учитывал бы ещё и простой ЭВМ, возника-

ющие за счёт различия времён счёта методов квазигидродинамики и Монте-Карло, а также из-за различной производительности ЭВМ.

Постановка задачи сводится к ряду подзадач:

1. Разработка эффективного алгоритма расчёта параметров полупроводниковых приборов. Алгоритм должен реализовать вычисления совмещением методов квазигидродинамики и Монте-Карло, учитывать различные времена решения задач и организовать вычисления на нескольких ЭВМ различной вычислительной мощности, минимизируя их времена простоев и, тем самым, увеличивая общую производительность алгоритма.
2. Тестирование алгоритма для проверки его работоспособности необходимо проводить при различных параметрах расчетной задачи. При оптимизации конструкции полупроводниковых приборов обычно требуется провести несколько циклов моделирования, изменяя те или иные параметры задачи. Может быть, "выгоднее" не рассчитывать несколько параметров в данном цикле моделирования, за счет чего возможно увеличение общей производительности алгоритма. Подобное увеличение достигается не только за счет снижения «объема» задачи, но и минимизации времени простоев отдельных ЭВМ локальной вычислительной сети, что даст дополнительный выигрыш. Поэтому необходимо провести тестирование алгоритма перед началом вычислений для оценки отличия его производительности от теоретически возможной. Алгоритм должен вычислять время моделирования для различных вариантов соотношения производительности ЭВМ, времен решения и количества задач.
3. При организации работ по проектированию полупроводниковых приборов необходимо проводить апробацию вычислительных алгоритмов с целью изучения влияния гетеропереходов и высокого градиента легирования на параметры приборов. Как было сказано ранее, при введении в структуру прибора гетеропереходов, параметры приборов существенно изменяются, а само наличие указанной конструктивной особенности приводит к росту числа оптимизируемых параметров. Алгоритм должен позволять проводить оптимизацию структуры приборов за минимальное время, что имеет важное значение для организации работ.

3.3 ОПИСАНИЕ ВЫЧИСЛИТЕЛЬНОГО АЛГОРИТМА

Поскольку в рассматриваемой задаче моделирование предполагается проводить на нескольких компьютерах, объединённых в вычислительную сеть, то первоочередной задачей алгоритма является минимизация времён простоя ЭВМ в ожидании расчётных данных. Необходимо учитывать различие во временах счёта методами Монте-Карло и квазигидродинамики. Алгоритм должен также обеспечивать "автоматическое" распределение подзадач на разные ЭВМ.

В результате применения такого алгоритма должна быть известна полная информация о работе прибора: вольт-амперная, вольт-фарадная характеристики, коэффициент усиления, коэффициент шума и т.д. Пользователь должен иметь возможность настраивать алгоритм для решения какой-либо другой, подобной этой, задачи моделирования полупроводниковых приборов.

Алгоритм, блок-схема которого представлена на рис. 3.7, исходя из известной информации о временах счёта методов квазигидродинамики и Монте-Карло, полученных на этапе тестирования вычислительной сети, выстраивает очередь задач на каждую из расчётных ЭВМ.

Программа реализации алгоритма распределяет задачи таким образом, чтобы обеспечить минимальное время простоя более быстрой ЭВМ в ожидании расчётных данных от другой, более медленной, ЭВМ, так как именно её простои оказывают наибольшее влияние на производительность алгоритма в целом. Медленная ЭВМ может простаивать, но должна вовремя обеспечить загрузку новыми расчётными данными быстрой ЭВМ

Если одна из ЭВМ закончила текущие вычисления, программа реализации алгоритма определяет рациональность загрузки этой ЭВМ новыми расчётными данными, в результате чего принимается решение о следующем запуске ЭВМ с новыми исходными данными или о необходимости её простоя.

В работе было предложено использовать локальную вычислительную сеть лаборатории из трёх ЭВМ в качестве кластерной вычислительной системы. Одна из ЭВМ - управляющая, а две других - расчётные (рис. 3.8).

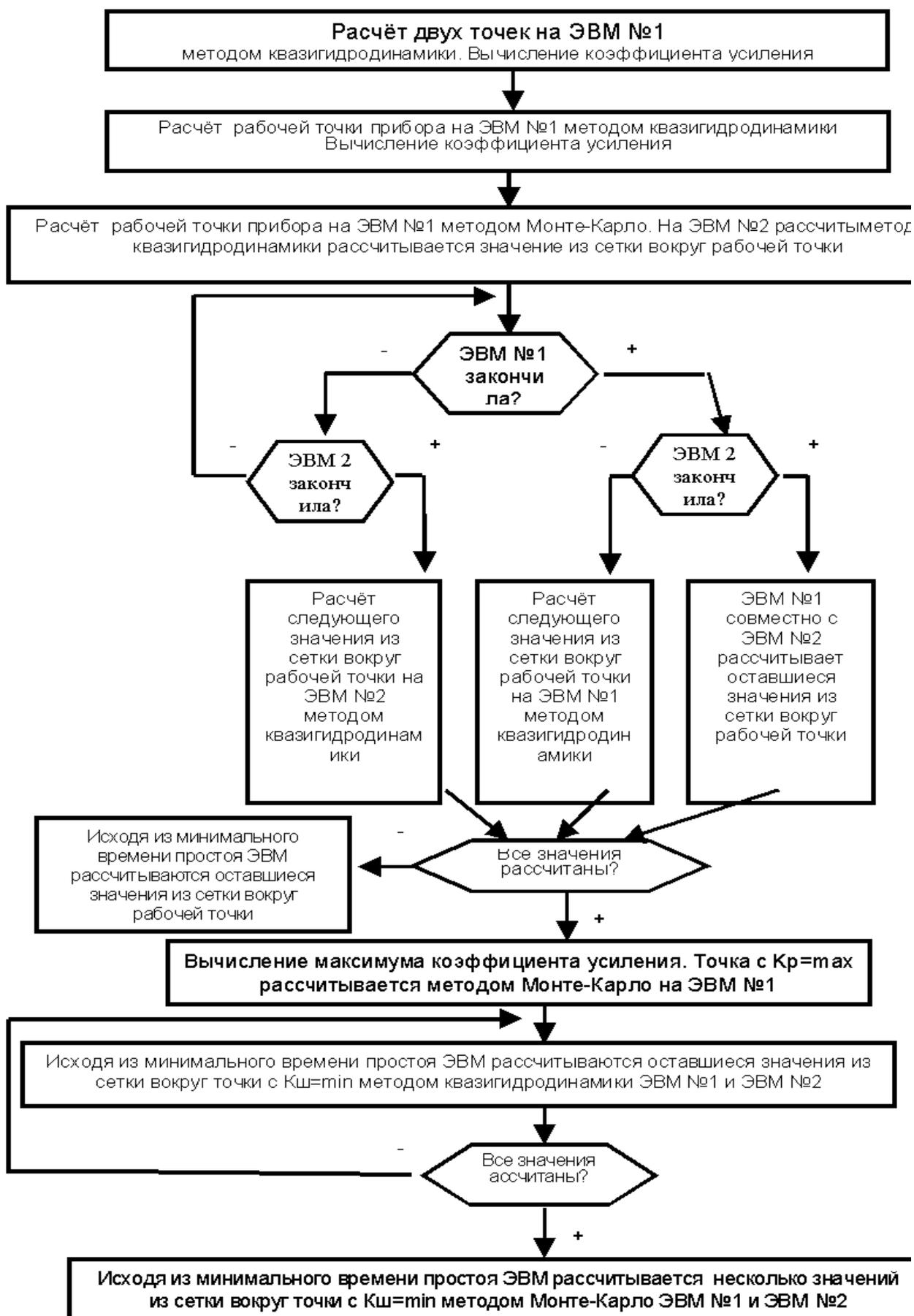


Рис. 3.7. Блок-схема разработанного алгоритма

Управляющая ЭВМ выполняет программу, реализующую предложенный алгоритм. Она постоянно "опрашивает" расчётные ЭВМ об их состоянии. Если какая-то ЭВМ закончила текущие вычисления, то управляющая программа вычисляет рациональность загрузки этой ЭВМ новыми расчётными данными, и, если это необходимо, передаёт ей по сети новые исходные данные.



Рис. 3.8. Структурная схема реализованной вычислительной системы

Предложенная вычислительная система - это типичный представитель небольших вычислительных систем, которые можно организовать в каждой лаборатории. Обычно смена вычислительного оборудования в лабораториях происходит не сразу. Постепенно меняются старые ЭВМ на более мощные. То есть в лаборатории всегда имеется компьютеры с разными вычислительными мощностями, что оправдывает разработку предложенного кластерного алгоритма.

Естественно, что задержки, связанные с передачей расчётных данных по сети отрицательно сказываются на общем времени моделирования. Поэтому при практической реализации алгоритма было произведено тестирование локальной сети лаборатории.

При практической реализации вычислений была использована одноранговая сеть Windows кафедры электроники РФФ ННГУ. Было установлено, что задержки, связанные с передачей расчётных данных по сети, не превосходят 1-3 секунд (рис. 3.9). Время счёта задач составляет 20-30 минут, поэтому задержки передачи данных по сети слабо влияют на общую производительность алгоритма.

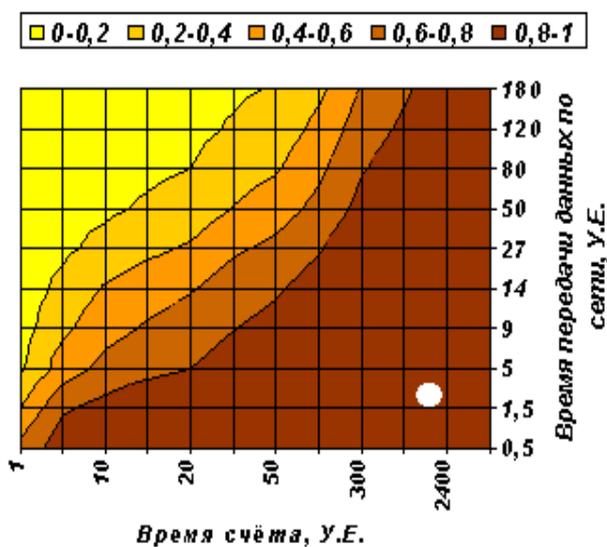


Рис. 3.9. Анализ влияния задержек сети на общую производительность алгоритма

3.4 ТЕСТИРОВАНИЕ АЛГОРИТМА⁸

Понятно, что полностью избавиться от простоев ЭВМ, связанных с ожиданием расчётных данных, в принципе, невозможно. Поэтому в работе также произведено и тестирование алгоритма для выяснения влияния простоев, а также других параметров расчётной задачи на общую производительность. Простои могут быть двух типов (рис. 3.10):

а) Если одна из ЭВМ закончила выполнять свою очередь задач, и оставшееся время расчёта определяется другой ЭВМ. Понятно, что при большом количестве задач время моделирования стремится (рис. 3.10) к теоретически возможному, которое определяется следующим образом:

$$t = \frac{T_{\text{счётаКГД}} + T_{\text{счётаМК}}}{P_1 + P_2} \quad (3.13)$$

где: $T_{\text{счёта КГД}}$ - время счёта методом квазигидродинамики, $T_{\text{счёта МК}}$ - время счёта методом Монте-Карло, P_1 - производительность ЭВМ №1, P_2 - производительность ЭВМ №2.

б) Простои, связанные с самим алгоритмом расчёта ВАХ. Одна из ЭВМ ожидает расчётных данных от другой ЭВМ. Поскольку задача моделирования

⁸ В разделах 3.4 и 3.5 использованы материалы работ М.Н.Логунова и С.В.Оболенского

заключается не только в расчёте вольт-амперных характеристик, но и в вычислении минимального шума приборов, их коэффициента усиления и т.п., то требуется применять процедуру "переключения" задач квазигидродинамики и Монте-Карло. Поэтому одна из ЭВМ может простаивать в ожидании исходных данных от другой ЭВМ.

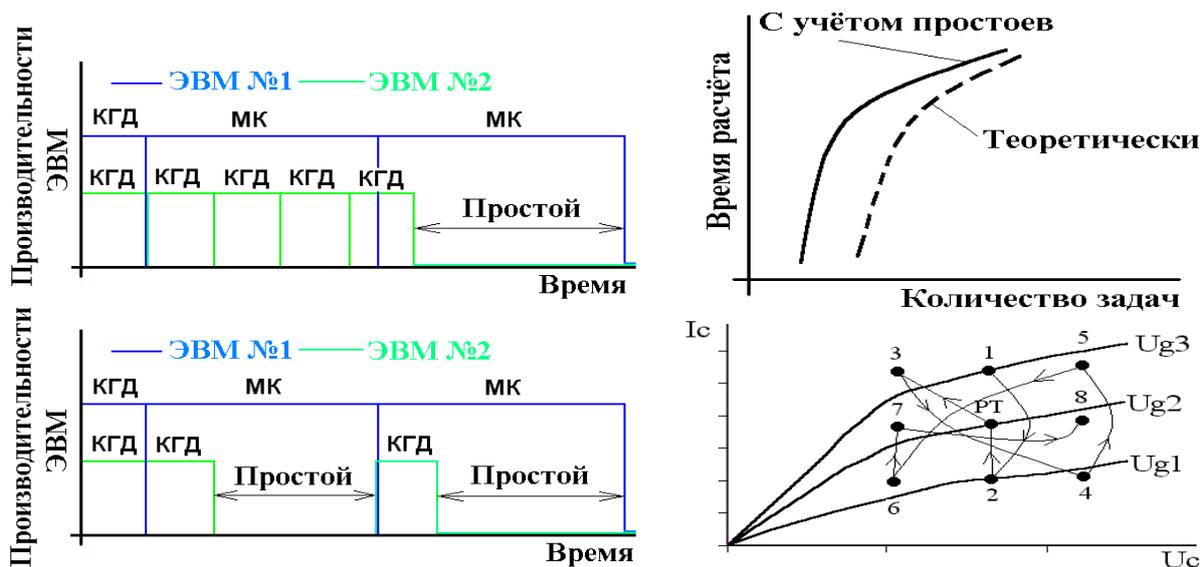


Рис. 3.10. а) Причины возникновения простоев б) Алгоритм расчёта ВАХ

При практической реализации алгоритма было замечено, что различные точки на ВАХ методом Монте-Карло рассчитываются разное количество времени. Это объясняется различием в количестве итераций, необходимых для получения результата. Поэтому было предложено использовать, так называемый адаптивный алгоритм, который учитывает ещё и эту особенность, определяя, какую точку на ВАХ в данный момент времени "выгоднее" рассчитать методом Монте-Карло.

После первого прохода алгоритма времена счёта различных точек на ВАХ методом Монте-Карло известны, и при следующих проходах можно скорректировать последовательность обхода точек, например, как показано на рис. 3.12. Как показывают эксперименты, в этом случае простои ЭВМ уменьшаются (рис. 3.11) и, следовательно, увеличивается общая производительность алгоритма.

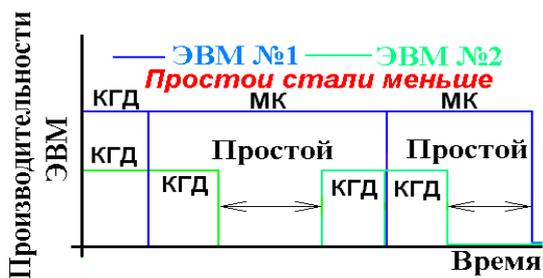


Рис. 3.11. Уменьшение простоев при применении адаптивного алгоритма

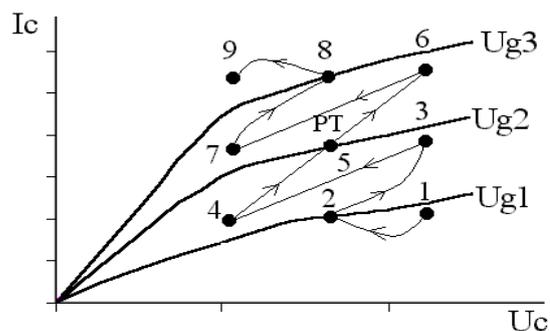


Рис. 3.12. Адаптивный алгоритм

Влияние простоев резко возрастает при уменьшении количества расчётных ЭВМ. Именно поэтому тестирование алгоритма было организовано с использованием всего двух расчётных ЭВМ, чтобы продемонстрировать увеличение производительности даже в таком, наихудшем, случае.

3.5 АНАЛИЗ ПРОИЗВОДИТЕЛЬНОСТИ АЛГОРИТМА

Для описания работы алгоритма предложено ввести набор следующих параметров:

- соотношение производительностей ЭВМ;
- соотношение времён счёта задач методами квазигидродинамики и Монте-Карло;
- количество точек, рассчитываемых методом квазигидродинамики;
- количество точек, рассчитываемых методом Монте-Карло.

В ходе выполнения работы было выяснено влияние предложенных параметров исходной задачи на общую производительность алгоритма. Анализ влияния количества расчётных точек квазигидродинамики и Монте-Карло на общую производительность алгоритма показывает, что при его увеличении максимальное отклонение производительности алгоритма от теоретически возможного уменьшается. Максимальное отличие производительностей при трёх расчётных точках Монте-Карло объясняются увеличением времени простоя ЭВМ, которое связано с тем, что одна из ЭВМ закончила выполнение своей очереди задач, и оставшееся время моделирования определяется другой ЭВМ, которая «досчитывает» третью расчётную точку Монте-Карло. Когда

количество расчётных точек невелико, то такие простои сильно влияют на общую производительность алгоритма, особенно при соотношении времён счёта методами квазигидродинамики и Монте-Карло – как 1 к 10.

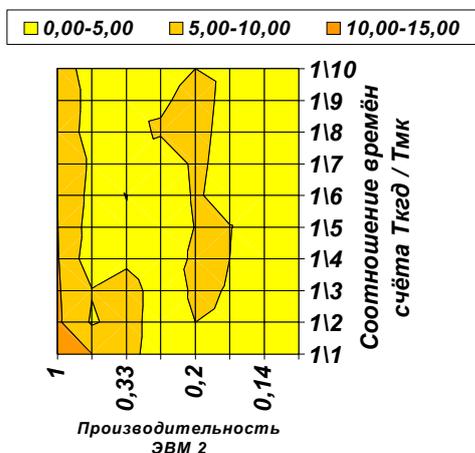


Рис. 3.13. Максимальные отклонения производительности алгоритма от теоретически возможной в зависимости от производительности медленной ЭВМ (производительность быстрой ЭВМ принята за 1) и соотношения времён счёта задач

Количество точек КГД – 34

Количество точек МК – 11

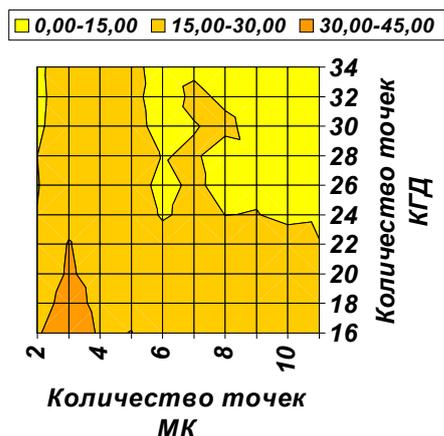


Рис. 3.13. Максимальные отклонения производительности алгоритма от теоретически возможной в зависимости от количества расчётных точек квазигидродинамики и Монте-Карло для различных соотношений времён счёта задач при неоптимальном соотношении производительности ЭВМ (наихудший случай)

При практической реализации алгоритма было установлено, что наиболее выгодное соотношение производительностей (с точки зрения минимального отличия производительности алгоритма от теоретически возможного) компьютеров составляет 1/4. Это объясняется минимальным временем простоев одной ЭВМ в ожидании расчётных данных от другой. Так при выше указанном соотношении производительностей ЭВМ максимальное отличие производительности алгоритма от теоретически возможного составляет не более 17 % (см. рис.3.14).

Предложенный алгоритм позволяет оценить эффективность использования ЭВМ с конкретной производительностью, что в конечном счёте определяет вложение финансовых ресурсов лаборатории. Например, в небольшой лабора-

тории имеется одна расчётная ЭВМ с частотой процессора 2,7 ГГц. Целесообразнее приобрести ещё одну ЭВМ с частотой процессора 700 МГц за \$350, чем ЭВМ с частотой процессора 3,2 ГГц за \$1000. В результате приобретения такой ЭВМ соотношение цена/производительность значительно уменьшается. При увеличении количества расчётных ЭВМ алгоритм также позволяет определить, какую из расчётных ЭВМ потребуется в будущем заменить.

Для анализа эффективности работы алгоритма целесообразно использовать специальную тестовую программу. Эта программа до начала основного цикла моделирования позволяет сделать вывод о рациональности соотношения различных параметров исходной задачи (рис. 3.14).

При конкретном соотношении производительностей компьютеров (рис. 3.15), как это бывает в реальной сети, из-за изменения времён простоя ЭВМ зависимость становится сложной, с большим количеством локальных экстремумов. Наличие локальных экстремумов при нечётном количестве расчётных точек квазигидродинамики и Монте-Карло объясняется увеличением времени простоя ЭВМ, связанного с тем, что одна из ЭВМ закончила выполнение своей очереди задач, и оставшееся время функционирует только другая ЭВМ, которая «досчитывает» последнюю расчётную точку квазигидродинамики или Монте-Карло. При соотношении времён счёта методов квазигидродинамики и Монте-Карло 1/10 наиболее сильное влияние на общую производительность алгоритма оказывает «нечётность» расчётных точек Монте-Карло.

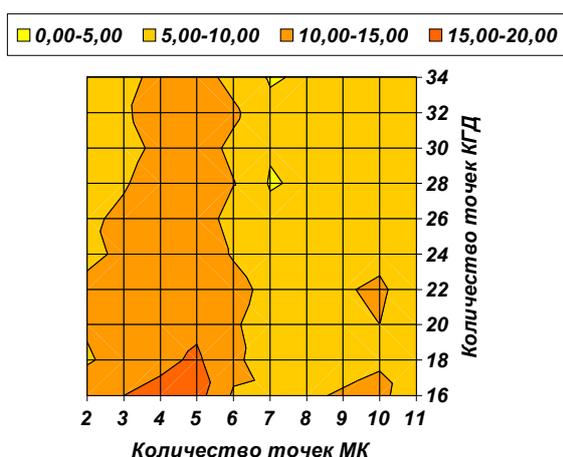


Рис. 3.14. Максимальные отклонения производительности алгоритма от теоретически возможной для соотношения производительностей ЭВМ 1/4 и различных соотношений времён счёта задач (выбирался наихудший случай)

Введение процедуры оценки производительности процесса расчёта имеет очень важное значение, особенно для малого количества задач моделирования в

одном цикле проектирования прибора. В этом случае пользователь алгоритма может до начала вычислений получить информацию о требуемом для вычислений времени. Это позволяет скорректировать объем расчетной задачи (т.е. количество задач КГД и МК) таким образом, чтобы задача в целом была решена за приемлемое для пользователя время.

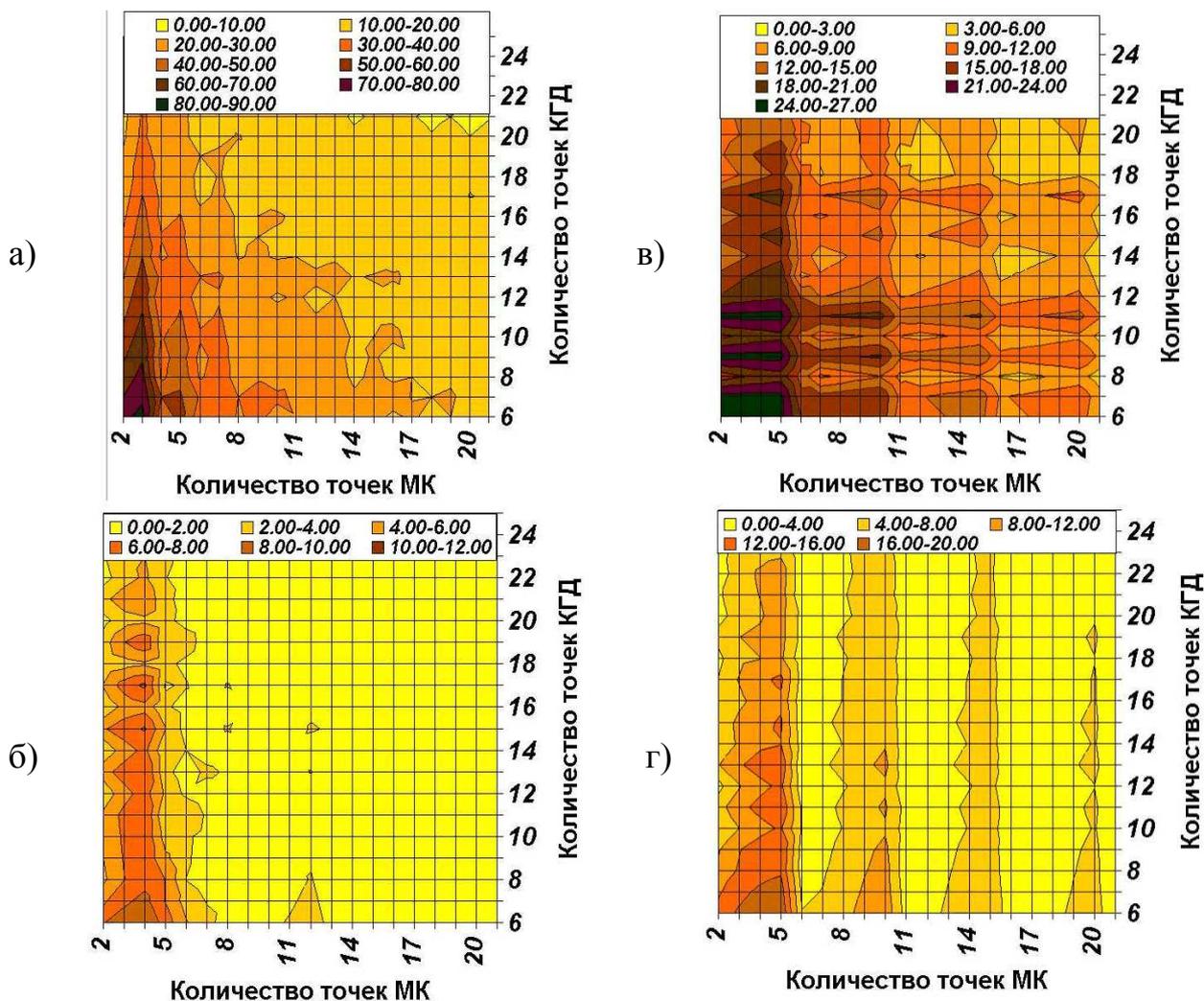


Рис. 3.15.. Анализ эффективности работы алгоритма: а) максимальные отклонения эффективности алгоритма от теоретически возможной для различных соотношений времён счёта задач и производительностей ЭВМ; б) минимальные отклонения эффективности алгоритма от теоретически возможной для различных соотношений времён счёта задач и производительностей ЭВМ; в) максимальные отклонения эффективности алгоритма от теоретически возможной для различных соотношений времён счёта задач и соотношении производительностей ЭВМ равном 1/4; г) минимальные отклонения эффективности алгоритма от теоретически возможной для различных соотношений времён счёта задач и соотношении производительностей ЭВМ равном 1/4.

Последнее чрезвычайно актуально при проведении оптимизационных работ, когда вычисления должны быть проведены в ограниченное время, так чтобы они уложились во временной интервал между двумя практическими итерациями по изготовлению прибора.

Использование предложенного алгоритма приводит к резкому уменьшению времени решения задачи (рис. 3.15 и 3.16). Даже при использовании двух расчётных ЭВМ видно, что время расчёта задачи значительно сокращается, по сравнению с многочастичным методом Монте-Карло.

Выигрыш производительности алгоритма при его реализации на двух и на четырех ЭВМ скромнее, но не следует забывать о том, что реальное время решения оптимизационной задачи при проектировании полупроводниковых приборов составляет десятки часов, а порой и несколько суток. Последнее показывает, что экономия расчетного времени даже порядка 10% на самом деле дает экономию в несколько часов.

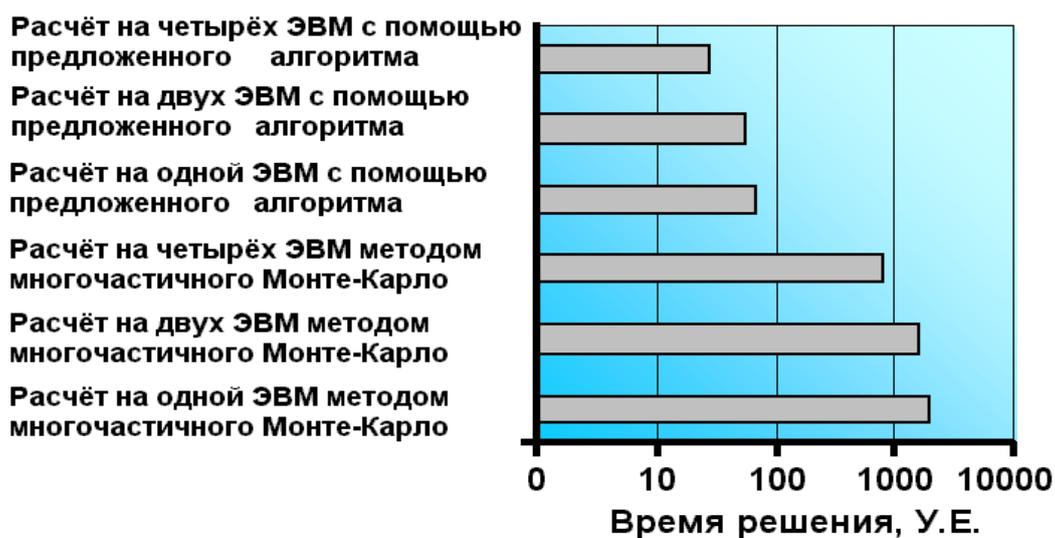


Рис. 3.16. Сравнение алгоритма с многочастичным методом Монте-Карло. Соотношение производительностей ЭВМ 1/4

Последнее важно не только само по себе, но и с точки зрения соотношения вычислительного процесса с организацией работ персонала лаборатории. Минимальное время получения результата достигается не только за счет увеличения производительности ЭВМ и алгоритма в целом, но и за счет организации эффективного ввода исходных данных в ЭВМ. Например, при проведении круглосуточных расчетов, окончание процесса вычислений ночью означает простой ЭВМ до утра, т.е. до момента прихода персонала на работу.

Поэтому оптимизация и распараллеливание процесса вычислений между ЭВМ позволит ликвидировать простои без введения режима многосменной работы персонала.

СПИСОК ЛИТЕРАТУРЫ

1. Оболенский С.В., Демарина Н.В., Волкова Е.В. Основы физики полупроводников. Транспорт носителей заряда в электрических полях: учебное пособие. – Нижний Новгород, Издательство Нижегородского госуниверситета, 2007. – 79 с.
2. Зи С.М. Физика полупроводниковых приборов. – М. Мир, 1981. – 567 с.
3. Пожела Ю. Физика быстродействующих транзисторов. – Вильнюс: Мокслас, 1989. – 264 с.
4. Степаненко И.П. Основы Микроэлектроники. – М.: Лаборатория базовых знаний, 2001. – 488с.
5. Шур М. Современные приборы на основе арсенида галлия. – М.: Мир, 1991. – 632 с.
6. Полевые транзисторы на арсениде галлия / ред. Ди Лоренцо Д.В., Канделуола Д.Д. – М.: Радио и связь, 1988. – 496 с.
7. Бонч-Бруевич В.Л., Калашников С.Г. Физика полупроводников. – М.: Наука, 1990. – 559 с.
8. Технология СБИС / Пирс К., Адамс А., Кац Л., Цай Дж., Сейдел Т., Макгиллис Д.; под редакцией Зи С. – М: Мир, 1986. – 404 с.
9. Боресков А.В., Харламов А.А. Основы работы с технологией CUDA. – М.: ДМК Пресс, 2010. – 232 с.
10. Сандерс Д., Кэндрот Э. Технология CUDA в примерах: введение в программирование графических процессоров. – М.: ДМК Пресс, 2011. – 232 с.
Боресков А.В., Харламов А.А., Марковский А.А., Микушин Д.Н., Мортиков Е.В., Мыльцев А.А., Сахарных Н.А., Фролов В.А. Параллельные вычисления на GPU. Архитектура и программная модель CUDA. – М.: МГУ, 2012. – 336 с.
11. Воеводин В.В. Вычислительная математика и структура алгоритмов. – М.: МГУ, 2010. – 168 с.
12. Гергель В.П. Высокопроизводительные вычисления для многопроцессорных многоядерных систем. – М.: МГУ, 2010. – 544 с.
13. Корняков К.В., Кустикова В.Д., Мееров И.Б., Сиднев А.А., Сысоев А.В., Шишков А.В. Инструменты параллельного программирования в системах

- с общей памятью. – М.: МГУ, 2010. – 272 с.
14. Линева А.В., Боголепов Д.К., Бахраков С.И. Технологии параллельного программирования для процессоров новых архитектур – М.: МГУ, 2010. – 160 с.
 15. Боресков А.В., Харламов А.А. Основы работы с технологией CUDA. – М.: ДМК Пресс, 2011. – 232 с.
 16. Сандерс Дж., Кэндрот Э. Технология CUDA в примерах: Введение в программирование графических процессоров. – М.: ДМК Пресс, 2011. – 232 с.
 17. Гречников Е.А., Михайлов С.В., Нестеренко Ю.В., Поповян И.А. Вычислительно сложные задачи теории чисел. – М.: МГУ, 2012. – 312 с.
 18. Боресков А.В., Харламов А.А., Марковский Н.Д., Микушин Д.Н., Мортиков Е.В., Мыльцев А.А., Сахарных Н.А., Фролов В.А. Параллельные вычисления на GPU. Архитектура и программная модель CUDA. – М.: МГУ, 2012. – 336 с.
 19. Гергель В.П. Современные языки и технологии параллельного программирования. – М.: МГУ, 2012. – 408 с.
 20. Антонов А.С. Технологии параллельного программирования MPI и OpenMP. – М.: МГУ, 2012. – 344 с.
 21. Лыкосов В.Н., Глазунов А.В., Кулямин Д.В., Мортиков Е.В., Степаненко В.М. Суперкомпьютерное моделирование в физике климатической системы. – М.: МГУ, 2012. – 408 с.
 22. Яковлевский М.В. Введение в параллельные методы решения задач. – М.: МГУ, 2013. – 328 с.
 23. Стронгин Р.Г., Гергель В.П., Гришагин В.А., Баркалов К.А. Параллельные вычисления в задачах глобальной оптимизации – М.: МГУ, 2013. – 280 с.
 24. Рутм Г., Фатика М. CUDA Fortran для ученых и инженеров. – М.: ДМК Пресс, 2014. – 364 с.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Каковы характерные пространственные масштабы в физике полупроводниковых диодов?
2. Каковы особенности функционирования современных полупроводниковых диодов?
3. В чем причины необходимости использования высокопроизводительных вычислений для проектирования современных полупроводниковых диодов?
4. Физический смысл терминов когерентное и монохроматическое излучение?
5. Каковы тепловые эффекты в полупроводниковых диодах?
6. Каковы особенности применения высокопроизводительных вычислений для оптимизации параметров приборов?
7. Каковы возможности квазигидродинамического приближения при проведении анализа процессов переноса электронов в полупроводниковых структурах диодов?
8. Каковы возможности использования параллельных вычислений при решении указанных задач?

Екатерина Валерьевна **Волкова**
Сергей Владимирович **Оболенский**

**ФИЗИКА ОПТОЭЛЕКТРОННЫХ, СВЧ И КВЧ ДИОДОВ И
ОРГАНИЗАЦИЯ ИХ ПРОЕКТИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ
ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ ВЫЧИСЛЕНИЙ**

Учебное пособие

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования «Нижегородский государственный
университет им. Н.И. Лобачевского».
603950, Нижний Новгород, пр. Гагарина, 23.

Подписано в печать . Формат 60×84 1/16.
Бумага офсетная. Печать офсетная. Гарнитура Таймс.
Усл. печ. л. . Уч-изд. л.
Заказ № . Тираж 200 экз.

Отпечатано в типографии Нижегородского госуниверситета
им. Н.И. Лобачевского
603600, г. Нижний Новгород, ул. Большая Покровская, 37
Лицензия ПД № 18-0099 от 14.05.01