УДК 621.382
ББК 32.852
Л 33

Лебедев А. И. **Физика полупроводниковых приборов.** — М.: ФИЗМАТЛИТ, 2008. — 488 с. — ISBN 978-5-9221-0995-6.

Рассмотрены физические принципы работы наиболее важных классов современных полупроводниковых приборов: диодов, биполярных и полевых транзисторов, тиристоров, СВЧ приборов с отрицательным дифференциальным сопротивлением (диодов Ганна, лавинно-пролетных и инжекционно-пролетных диодов), приборов с зарядовой связью, оптоэлектронных приборов (фотоприемников, светодиодов, инжекционных лазеров и др.). Выведены основные теоретические соотношения, определяющие характеристики этих приборов. Большое внимание уделено описанию особенностей современных быстродействующих приборов с субмикронными и нанометровыми размерами, в том числе приборов, в работе которых используются гетеропереходы, квантовые ямы и квантовые точки. Помимо этого, в книге рассмотрены основы планарной технологии, описаны возникшие в последнее время технологические проблемы и указаны перспективные пути их решения.

Для студентов старших курсов, аспирантов и научных сотрудников, работающих в области физики полупроводников.

Рекомендовано УМО по классическому университетскому образованию РФ в качестве учебного пособия для студентов ВУЗов, обучающихся по специальностям 010701 — «Физика», 010704 — «Физика конденсированного состояния вещества», 010803 — «Микроэлектроника и полупроводниковые приборы».

ISBN 978-5-9221-0995-6

© ФИЗМАТЛИТ, 2008
© А. И. Лебедев, 2008

ОГЛАВЛЕНИЕ

Введение 7

Глава 1. **Полупроводниковые диоды** 11

1.1. Потенциальный барьер в p–n-переходе 11
1.2. Вольт-амперная характеристика p–n-перехода 19
1.2.1. Вольт-амперная характеристика тонкого p–n-перехода 19
1.2.2. Влияние генерации и рекомбинации в области пространственного заряда на вольт-амперные характеристики p–n-перехода (модель Са–Нойса–Шокли) 28
1.2.3. p–n-переход при высоких уровнях инжекции 33
1.2.4. Вольт-амперная характеристика p–n-диода 39
1.3. Явление пробоя p–n-перехода 44
1.3.1. Лавинный пробой p–n-перехода 45
1.3.2. Туннельный пробой p–n-перехода 55
1.3.3. Тепловой пробой p–n-перехода 57
1.3.4. Стабилитроны 58
1.4. Туннельные диоды 62
1.4.1. Вольт-амперная характеристика туннельного диода . . 64
1.4.2. Избыточный ток в туннельных диодах 74
1.4.3. Выбор материалов для туннельных диодов 76
1.4.4. Обращенные диоды 81
1.5. Диоды с барьером Шоттки 83
1.5.1. Энергетическая диаграмма контакта металл–полупроводник 84
1.5.2. Вольт-амперная характеристика барьера Шоттки 89
1.5.3. Омические контакты к полупроводникам 99
1.6. Гетеропереходы и сверхрешетки 103
1.6.1. Гетеропереходы 103
1.6.2. Квантовые ямы и сверхрешетки 110
1.7. Диод на переменном токе 121
1.7.1. Барьерная емкость 121
1.7.2. Диффузионная емкость 126
1.7.3. Импульсные характеристики и быстродействие диодов 128

1.7.4. Диоды с накоплением заряда 133
1.7.5. Емкостная спектроскопия глубоких уровней 135

Глава 2. **Биполярные транзисторы** 140

2.1. Немного истории. Конструкции биполярного транзистора. . . 140
2.2. Параметры, определяющие коэффициент усиления транзистора. 148
2.2.1. Коэффициент инжекции эмиттера 149
2.2.2. Коэффициент переноса носителей через базу 153
2.2.3. Эффективность инжекции при очень малых и очень больших токах 155
2.2.4. Эффект оттеснения эмиттерного тока 157
2.2.5. Влияние напряжения на коллекторе на коэффициент усиления 158
2.3. Транзистор при высоком напряжении на коллекторе. 160
2.4. Высокочастотные свойства и быстродействие транзисторов. . 165
2.4.1. Частота отсечки и максимальная частота генерации . . 165
2.4.2. Гетеропереходные транзисторы 170
2.5. Транзисторы в схемах усиления сигналов 177
2.5.1. Схемы включения транзистора и выбор рабочей точки 177
2.5.2. Описание транзистора с помощью h-параметров 178
2.6. Шумы в биполярных транзисторах. 182
2.7. Особенности работы транзисторов в импульсном режиме. . . 187
2.8. Элементы интегральных схем на биполярных транзисторах 190
2.8.1. Планарная технология 191
2.8.2. Особенности устройства цифровых ИС на биполярных транзисторах. 209

Глава 3. **Тиристоры и другие многослойные структуры** . . . 215

3.1. Тиристоры. 215
3.1.1. Вольт-амперные характеристики тиристора. 216
3.1.2. Процессы включения и выключения тиристора 231
3.2. Многослойная структура — симистор 236

Глава 4. **Полевые транзисторы** 239

4.1. Полевые транзисторы с изолированным затвором. 241
4.1.1. МОП-конденсатор 241

4.1.2. Вольт-амперная характеристика МОП-транзистора . . . 249
4.1.3. Особенности реальных полевых транзисторов 252
4.1.4. Полевые транзисторы с коротким каналом 255
4.1.5. Быстродействие полевых транзисторов 259
4.1.6. Пути дальнейшего повышения быстродействия МОП-транзисторов 261
4.1.7. Мощные и высоковольтные МОП-транзисторы 264
4.2. Элементы интегральных схем на МОП-транзисторах 269
4.2.1. Интегральные схемы на n-МОП-транзисторах 271
4.2.2. КМОП-структуры 279
4.2.3. Энергонезависимые постоянные запоминающие устройства на МОП-транзисторах 285
4.3. Полевые транзисторы с управляющим p—n-переходом и барьером Шоттки 294

Глава 5. **Приборы с зарядовой связью** 311

Глава 6. **Полупроводниковые СВЧ приборы** 332
6.1. Диоды Ганна 333
6.2. Лавинно-пролетные диоды 355
6.3. Инжекционно-пролетные диоды 368
6.4. Другие способы генерации СВЧ колебаний 376
6.4.1. TRAPATT-режим 376
6.4.2. Туннельно-пролетные диоды (TUNNETT) 376
6.4.3. QWITT-диоды 379

Глава 7. **Оптоэлектронные приборы** 380
7.1. Приемники излучения 381
7.1.1. Механизмы поглощения излучения в полупроводнике 381
7.1.2. Фотоприемники, основанные на явлении фотопроводимости 384
7.1.3. Шумы в фотоприемниках 401
7.1.4. Фотовольтаические приемники (фотодиоды) 405
7.1.5. Лавинные фотодиоды 410
7.1.6. Преобразователи солнечной энергии (солнечные элементы) 414
7.1.7. Детекторы ядерных излучений 423

7.2. Светодиоды и полупроводниковые лазеры 428
7.2.1. Механизмы излучательной рекомбинации 428
7.2.2. Светодиоды 432
7.2.3. Инжекционные полупроводниковые лазеры 445

Приложение 461
Список литературы 463
Предметный указатель 478

Введение

Настоящая книга написана по материалам лекций, которые автор читает на физическом факультете МГУ в течение ряда последних лет. Со времени выхода последней двухтомной монографии М. Шура по физике полупроводниковых приборов прошло уже более 10 лет. Из-за чрезвычайно быстрого темпа развития прикладных разработок в области полупроводниковых приборов — а темпы развития микроэлектроники относятся, пожалуй, к наиболее быстрым — в этой области многое изменилось. Параметры приборов существенно улучшились, сменилось несколько поколений микросхем процессоров для ЭВМ. Появились новые физические идеи, предложены новые принципы работы и реализованы новые конструкции приборов. Технологические приемы, используемые при производстве полупроводниковых приборов, подошли к границам своих возможностей. Некоторые направления, только намечавшиеся 10 лет назад, вышли на магистральный путь, а другие, казавшиеся перспективными, отошли на второй план. Значит, пришло время подумать над изданием новой книги по физике полупроводниковых приборов, которая отразила бы эти изменения.

Развитие физики полупроводниковых приборов неразрывно связано с фундаментальными исследованиями по физике полупроводников. Некоторые из сделанных в этой области открытий, нашедших широкое практическое применение, были признаны важнейшими и отмечены Нобелевскими премиями. Речь идет о работах Дж. Бардина, У. Браттейна, У. Шокли (премия 1956 г. за открытие транзистора), работах Л. Есаки (премия 1973 г. за открытие туннельного эффекта в полупроводниках и сверхпроводниках), работах нашего соотечественника — Ж. И. Алфёрова совместно с Дж. Килби и Г. Кремером (премия 2000 г. за основополагающие работы в области информационной технологии и связи, вклад в изобретение интегральной схемы и создание полупроводниковых гетероструктур для высокоскоростной электроники и оптоэлектроники).

Разработка современного полупроводникового прибора является дорогостоящей задачей, требующей больших затрат труда и времени. Например, цикл изготовления современной сложной

микросхемы может занимать до трех месяцев. Поэтому в настоящее время на первый план выходит задача предварительного моделирования прибора, только после решения которой можно приступать к его созданию. Это требует умения количественно рассчитывать все необходимые параметры приборов. По этой причине к физике полупроводниковых приборов, как области науки, предъявляется требование уметь не только качественно объяснять, но и *количественно предсказывать* поведение рассматриваемой полупроводниковой структуры. Это определяет обилие в книге формул, обсуждений обоснованности тех или иных физических приближений — все это необходимо, чтобы обеспечить требуемый количественный результат.

Дополнительную сложность при решении задачи разработки полупроводниковых приборов создает и то, что эти приборы изготавливают из конкретных полупроводников, реальные свойства которых часто далеки от идеализированных представлений. Поэтому автор счел необходимым добавить в настоящую книгу некоторые особенности полупроводников (например, касающиеся поведения конкретных легирующих примесей), без знания которых создание совершенных приборов оказывается просто невозможным. Кроме того, в книгу включено описание основных приемов планарной технологии и обсуждаются направления их совершенствования, поскольку без понимания взаимосвязи физических принципов работы приборов и технологии их изготовления невозможно в полной мере проникнуться «духом» современной микроэлектроники.

Понимание материала, изложенного в настоящей книге, требует предварительного знакомства с курсом физики полупроводников, основами квантовой механики и радиотехники.

В книге рассмотрены основные классы современных полупроводниковых приборов и физические основы их работы. В первой и наибольшей по объему главе обсуждаются физические явления, возникающие на контакте двух полупроводников разного типа проводимости — в так называемом p–n-переходе. Здесь же рассмотрены явления, проявляющиеся в более сложных структурах с потенциальными барьерами: контактах металл–полупроводник (барьерах Шоттки), гетеропереходах, одиночных квантовых ямах и сверхрешетках. Эта глава закладывает основы, необходимые для понимания материала последующих глав. Вторая глава посвящена изучению биполярных транзисторов и путей дальнейшего улучшения их характеристик. В частности, рассмотрены подходы к созданию быстродействующих транзисторов (гетеропереходные транзисторы, транзисторы на горячих

электронах). В этой главе также изложены основы планарной технологии, которая в настоящее время является основой производства практически всех типов полупроводниковых приборов, а также некоторые приемы схемотехники, позволяющие за счет функциональной интеграции существенно увеличить плотность упаковки элементов в интегральных схемах и подойти к созданию сверхбольших интегральных схем (СБИС). Принципы работы и свойства четырехслойных и еще более сложных биполярных структур, из которых изготавливают крайне необходимые для современной силовой энергетики тиристоры и симисторы, рассмотрены в главе 3. Четвертая глава посвящена полевым транзисторам — наиболее распространенным на сегодня полупроводниковым приборам. Тем, что в настоящее время нас окружают высокопроизводительные компьютеры, быстродействие которых возрастает с головокружительной скоростью, мы обязаны разработке именно этого класса полупроводниковых приборов. Особое внимание в этой главе занимают современные идеи и решения, позволяющие создавать полевые транзисторы, которые способны работать на частотах, относимых к субмиллиметровой области спектра (выше 300 ГГц). В этой же главе обсуждаются актуальные для современной электроники гибридные (биполярные+полевые) структуры типа IGBT и BiCMOS, а также основные типы ИС на полевых транзисторах (n-МОП, КМОП, статические, динамические и перепрограммируемые запоминающие устройства, флэш-память). В пятой главе рассмотрены принципы работы важного класса функционально-интегрированных приборов на основе эффекта поля — приборов с зарядовой связью. Наиболее интересным направлением развития этих приборов является, по-видимому, создание приемников изображения, которые широко используются в таких бытовых приборах, как цифровые фотоаппараты и видеокамеры. В главе 6 книги рассмотрен совершенно другой класс приборов — полупроводниковые СВЧ приборы. В этой главе описаны методы получения в полупроводниках отрицательного дифференциального сопротивления и создание на основе этого явления генераторов электромагнитных колебаний в диапазонах сантиметровых, миллиметровых и субмиллиметровых длин волн. Наконец, седьмая глава книги посвящена физическим принципам работы широкого класса оптоэлектронных приборов. Это — приемники излучения, используемые для регистрации электромагнитных колебаний начиная от дальней инфракрасной области спектра (BIB- и HIWIP-детекторы) до диапазона рентгеновского и гамма-излучения (детекторы ядерных излучений), и полупроводниковые источники

излучения (светодиоды, лазеры). Особое внимание в этой главе уделяется физическим явлениям в новых полупроводниковых объектах (квантовых ямах, нитях и точках) и использованию этих явлений для существенного улучшения параметров оптоэлектронных приборов.

Отличительной чертой настоящей книги является то, что бóльшая часть информации о наиболее важных идеях, разработках и достижениях последних лет в области физики полупроводниковых приборов почерпнута не из журнальных статей, а из Интернета. Интернет сделал общедоступными аналитические обзоры, написанные специалистами ведущих фирм-разработчиков всего мира, он позволяет оперативно следить за последними достижениями в области фундаментальных и прикладных исследований полупроводников и выявлять основные тенденции развития в этой области знаний.

Автор считает своей приятной обязанностью выразить благодарность своим коллегам, профессору А.Э. Юновичу, доценту М.В. Чукичеву, ст. научн. сотр. И.А. Куровой и И.А. Случинской, к.х.н. С.Г. Дорофееву и В.М. Шахпаронову, которые прочли отдельные разделы рукописи и высказали ряд ценных замечаний, способствовавших улучшению содержания книги в целом.

Глава 1

ПОЛУПРОВОДНИКОВЫЕ ДИОДЫ

Работа большинства полупроводниковых приборов основана на использовании специфических свойств контакта полупроводников разного типа проводимости — так называемого p–n-перехода. Эти свойства обусловлены целым рядом физических явлений, происходящих в таком контакте: инжекцией, туннелированием, ударной ионизацией носителей и др. В этой главе мы рассмотрим эти физические явления, установим их роль в конкретных условиях работы полупроводниковых диодов, рассчитаем характеристики p–n-перехода в этих условиях и обсудим, как ими можно управлять, изменяя геометрию прибора и параметры полупроводника.

1.1. Потенциальный барьер в p–n-переходе

Одним из основных физических явлений, происходящих на границе раздела полупроводников p- и n-типа проводимости, является возникновение энергетического барьера и области, обедненной носителями заряда (так называемого *обедненного (истощенного) слоя* или *области пространственного заряда*). Причиной возникновения этого барьера является диффузия свободных носителей заряда (электронов и дырок). Рассмотрим эти явления более подробно.

Из общего курса физики полупроводников [1] известно, что в невырожденном полупроводнике в любой его точке концентрации электронов n и дырок p связаны соотношением

$$n \cdot p = n_i^2 = N_c N_v \exp\left(-\frac{E_g}{kT}\right), \qquad (1.1)$$

где E_g — ширина запрещенной зоны полупроводника, N_c и N_v — эффективные плотности состояний в зоне проводимости и валентной зоне, k — постоянная Больцмана, а T — температура полупроводника, n_i — собственная концентрация носителей заряда.

При достаточно высокой температуре, когда примесные атомы практически полностью ионизованы, в состоянии равновесия в n-области, легированной донорами с концентрацией N_d, концентрация основных носителей (электронов) равна $n_0 \approx N_d$, концентрация неосновных носителей (дырок) равна $p_{n0} = n_i^2/n_0$, причем $n_0 \gg n_i \gg p_{n0}$. В области p-типа, легированной акцепторами с концентрацией N_a, концентрация дырок равна $p_0 \approx N_a$, концентрация электронов равна $n_{p0} = n_i^2/p_0$, причем $p_0 \gg n_i \gg n_{p0}$. Здесь и далее индексом 0 мы будем обозначать равновесные концентрации носителей.

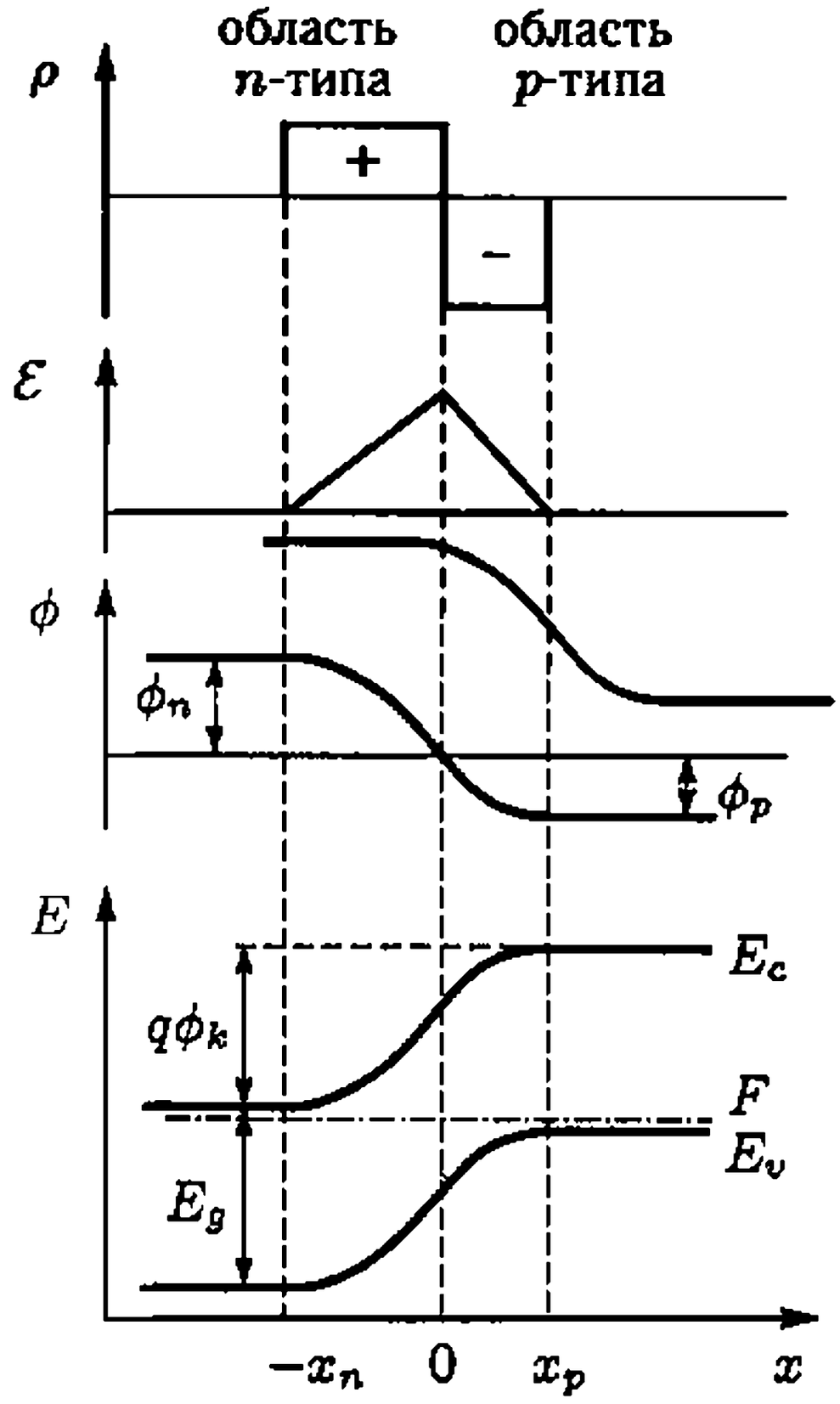


Рис. 1.1. Распределение плотности заряда ρ, электрического поля $\mathcal{E}$ и потенциала ϕ в резком p–n-переходе. Внизу — энергетическая диаграмма p–n-перехода в отсутствие внешнего смещения

Если привести в соприкосновение полупроводники разного типа проводимости (см. рис. 1.1), то в области их контакта начинается диффузия: электроны из n-области диффундируют

в p-область (где их равновесная концентрация $n_{p0} \ll n_0$) и там рекомбинируют с дырками, а дырки диффундируют из p-области в n-область (где $p_{n0} \ll p_0$) и там рекомбинируют с электронами. В результате этого область n-типа заряжается положительно, а p-область — отрицательно. Возникающее при этом электрическое поле сосредоточено вблизи границы p- и n-областей и направлено так, что препятствует диффузии. Поскольку это электрическое поле выталкивает свободные носители из пограничной области, на границе p- и n-областей возникает *обеднённый (свободными носителями) слой*, в котором пространственный заряд формируется оставшимися после ухода свободных носителей положительно заряженными донорами и отрицательно заряженными акцепторами. Перераспределение носителей продолжается до тех пор, пока не установится такой энергетический барьер, при котором в каждой точке p–n-перехода дрейфовые токи носителей в электрическом поле $\mathcal{E}$ не будут точно компенсировать их диффузионные токи, то есть плотности полных токов электронов и дырок не станут равными нулю:

$$\mathbf{J}_n = q\mu_n n\,\mathcal{E} + qD_n \nabla n = 0, \quad \mathbf{J}_p = q\mu_p p\,\mathcal{E} - qD_p \nabla p = 0. \quad (1.2)$$

В этой формуле μ_n и μ_p — подвижности электронов и дырок, D_n и D_p — их коэффициенты диффузии, а q — величина элементарного заряда.

В состоянии термодинамического равновесия, когда к p–n-переходу не приложено никакого внешнего напряжения, положение уровня Ферми F во всем кристалле одинаково (см. рис. 1.1). Это позволяет найти высоту возникающего барьера и связанную с ним *контактную разность потенциалов* ϕ_k. [1]) Используя известные из курса физики полупроводников соотношения между концентрацией электронов и дырок и положением уровней Ферми F_n и F_p относительно края зоны проводимости E_c и края валентной зоны E_v в невырожденном полупроводнике,

$$n = N_c \exp\left(-\frac{E_c - F_n}{kT}\right), \quad p = N_v \exp\left(-\frac{F_p - E_v}{kT}\right), \quad (1.3)$$

[1]) В зарубежной литературе эта величина часто обозначается как V_{bi}, где bi — сокращение от built-in («встроенный» потенциал).

и выражая E_g из формулы (1.1), получаем

$$q\phi_k = F_n - F_p = E_g - (E_c - F_n) - (F_p - E_v) =$$

$$= kT \ln\left(\frac{N_c N_v}{n_i^2}\right) - kT \ln\left(\frac{N_c}{n_0}\right) - kT \ln\left(\frac{N_v}{p_0}\right) = \qquad (1.4)$$

$$= kT \ln\left(\frac{n_0 p_0}{n_i^2}\right) \approx kT \ln\left(\frac{N_a N_d}{n_i^2}\right).$$

Для оценки величины $q\phi_k$ рассмотрим кремниевый p–n-переход с $N_d = N_a = 4 \cdot 10^{15}$ см$^{-3}$. Учитывая, что при комнатной температуре $kT = 0,0259$ эВ, а $n_i \approx 10^{10}$ см$^{-3}$, из формулы (1.4) находим $q\phi_k \approx$ $\approx 0,668$ эВ, что на 0,45 эВ меньше ширины запрещенной зоны кремния ($E_g = 1,12$ эВ). Для диодов, изготовленных из невырожденных полупроводников, высота потенциального барьера обычно оказывается на 0,2–0,5 эВ меньше ширины запрещенной зоны полупроводника, из которого они изготовлены.

Рассчитаем теперь распределение потенциала и напряженности электрического поля в p–n-переходе. Для этого нам надо решить уравнение Пуассона

$$\nabla \cdot \mathcal{E} \equiv -\nabla \cdot (\nabla \phi) = \frac{4\pi\rho}{\varepsilon}, \qquad (1.5)$$

связывающее электростатический потенциал ϕ и напряженность электрического поля $\mathcal{E}$ с плотностью объемного заряда ρ и диэлектрической проницаемостью полупроводника ε. [1])

Для простоты рассмотрим случай плоского p–n-перехода, для которого уравнение (1.5) сводится к одномерному дифференциальному уравнению,

$$-\frac{d^2\phi}{dx^2} = \frac{d\mathcal{E}}{dx} = \frac{4\pi\rho}{\varepsilon} = \frac{4\pi q}{\varepsilon}\left[p(x) - n(x) + N_d^+(x) - N_a^-(x)\right], \qquad (1.6)$$

с граничными условиями $\mathcal{E} = 0$, $\rho = 0$ в нейтральных областях вдали от p–n-перехода (рис. 1.1). В правой части этого уравнения мы учли все подвижные и неподвижные заряды, находящиеся в p–n-переходе. При комнатной температуре в большинстве практически важных полупроводников мелкие донорные и акцепторные уровни полностью ионизованы, и концентрации этих ионизованных уровней равны $N_d^+ \approx N_d$, $N_a^- \approx N_a$.

---

[1]) Здесь и далее мы будем записывать все уравнения в системе СГСЭ. Для перехода в систему СИ во всех формулах необходимо заменить ε на $4\pi\varepsilon\varepsilon_0$, где $\varepsilon_0 = 8,85 \cdot 10^{-14}$ Ф/см — электрическая постоянная.

В обедненном слое ($-x_n < x < x_p$), в котором за счет изгиба зон локальные концентрации электронов и дырок удовлетворяют соотношениям $n \ll N_d^+$ и $p \ll N_a^-$, уравнение (1.6) упрощается и можно записать:

$$-\frac{d^2\phi}{dx^2} \approx \frac{4\pi q}{\varepsilon} N_d \qquad \text{при} \quad -x_n < x < 0,$$

$$-\frac{d^2\phi}{dx^2} \approx -\frac{4\pi q}{\varepsilon} N_a \qquad \text{при} \quad 0 < x < x_p. \tag{1.7}$$

В наиболее простом случае, когда концентрация легирующих примесей в областях p- и n-типа не зависит от координат (такой p–n-переход называется *резким* [1]), интегрирование этих уравнений с учетом граничных условий дает:

$$\mathcal{E}(x) = \frac{4\pi q}{\varepsilon} N_d(x + x_n), \quad \phi(x) = \phi_n - \frac{2\pi q}{\varepsilon} N_d(x + x_n)^2, \tag{1.8}$$

$$\mathcal{E}(x) = -\frac{4\pi q}{\varepsilon} N_a(x - x_p), \quad \phi(x) = \phi_p + \frac{2\pi q}{\varepsilon} N_a(x - x_p)^2, \tag{1.9}$$

где ϕ_n и ϕ_p — значения потенциала в нейтральных n- и p-областях p–n-перехода. «Сшивая» полученные решения для $\mathcal{E}$ и ϕ в точке $x = 0$, из равенства для $\mathcal{E}$ получаем условие электронейтральности образца ($N_a x_p = N_d x_n$), а из равенства для ϕ — уравнение для контактной разности потенциалов:

$$\phi_k = \phi_n - \phi_p = \frac{2\pi q}{\varepsilon}\left(N_d\, x_n^2 + N_a x_p^2\right). \tag{1.10}$$

Выражая теперь с помощью условия электронейтральности величины x_p и x_n как доли полной толщины обедненного слоя $W = x_p + x_n$ и подставляя их в уравнение (1.10), находим

$$\phi_k = \frac{2\pi q}{\varepsilon} N^* W^2, \tag{1.11}$$

---

[1]) Приближение резкого p–n-перехода применимо к структурам, в которых толщина переходной области (в которой разность $N_d - N_a$ заметно меняется) мала по сравнению с толщиной обедненного слоя. На практике такие p–n-переходы получают либо вплавлением примеси в легированный полупроводник, либо наращиванием эпитаксиального слоя полупроводника одного типа проводимости на подложку, изготовленную из полупроводника другого типа проводимости; в обоих случаях диффузионное размытие профиля легирования за время создания p–n-перехода оказывается небольшим.

где $N^* = N_d N_a/(N_d + N_a)$ — так называемая *приведенная концентрация примеси*. Из этой формулы сразу же следует, что толщина области пространственного заряда равна

$$W = \sqrt{\frac{\varepsilon}{2\pi q N^*}\,\phi_k}\ . \qquad (1.12)$$

Таким образом, толщина области пространственного заряда уменьшается с ростом концентраций легирующих примесей.

Если p–n-переход легирован асимметрично (то есть $N_d \neq N_a$), то, как следует из полученных формул, область пространственного заряда в основном располагается в области, которая легирована слабее, а такие характеристики p–n-перехода, как W и $\mathcal{E}$, определяются концентрацией примеси в этой слабо легированной области (базе диода). [1])

Сделаем два замечания, касающихся ранее пренебрегавшегося нами вклада свободных носителей в величину ρ в уравнении (1.6).

Первое замечание касается пренебрежения вкладом свободных носителей на краях p–n-перехода, где изгиб зон мал. На самом деле строгим решением уравнения (1.6) в этой области является экспоненциальное затухание $\mathcal{E}$ по мере удаления от перехода с характерным масштабом, равным *дебаевской длине экранирования* $L_D = \sqrt{\varepsilon kT/(4\pi q^2 n_0)}$ [1]. Можно показать, что если последовательно учесть вклад свободных носителей, то конечная формула (1.12) останется такой же за исключением того, что в ней вместо ϕ_k будет стоять величина $(\phi_k - 2kT/q)$. Однако поскольку обычно $\phi_k \gg kT$, при качественных рассуждениях этой поправкой часто пренебрегают.

Второе замечание относится к асимметрично легированным p–n-переходам. Рассмотрим его на примере p^+—n-перехода. [2]) Если асимметрия легирования велика ($(N_a/N_d)\ln(N_a/N_d) > q\phi_k/kT$), то вблизи границы p–n-перехода со стороны n-области появляется тонкий слой, в котором концентрация дырок $p > N_d^+$ и, таким образом, приближение $\rho \approx qN_d^+$ (формула (1.7)) в этом слое оказывается неприменимым. В этом случае форму потенциального барьера рассчитывают, разбивая p–n-переход не на две, а на три области [2].

---

[1]) Базой диода (от англ. base — основание) обычно называют механически прочную пластинку из слабо легированного полупроводника, на поверхности которой создается полупроводниковая структура. При диффузии или вплавлении примеси в такую пластинку обычно образуются асимметрично легированные p–n-переходы.

[2]) При указании слоев в полупроводниковых структурах принято для обозначения сильно легированной области сразу за типом ее проводимости писать знак плюс (p^+, n^+), а слабо легированные области p- и n-типа обозначать греческими буквами π и ν.

Вклад свободных носителей в формирование потенциального барьера может оказаться существенным и в ситуациях, в которых проявляется туннельный эффект. О влиянии туннелирования носителей на вольт-амперные характеристики барьеров Шоттки мы будем говорить на с. 94.

Рассмотрим теперь как изменится распределение электрического поля и потенциала в случае, когда к p–n-переходу приложено внешнее смещение. Если пренебречь изменением плотности пространственного заряда, которое может возникать при прямом смещении в результате инжекции носителей, [1]) то проведенные выше вычисления можно повторить и для этого случая. Поскольку в уравнениях изменяются лишь граничные условия, то все изменения сводятся к тому, что вместо $\phi_n - \phi_p = \phi_k$ в формуле (1.10) следует писать $\phi_n - \phi_p = \phi_k - V$, где V — напряжение смещения на переходе ($V < 0$ для обратного смещения). В итоге, выражение для толщины обедненного слоя резкого p–n-перехода в общем случае имеет вид

$$W = \sqrt{\frac{\varepsilon}{2\pi q N^*}(\phi_k - V)}\,. \qquad (1.13)$$

Из этого уравнения следует, что при прямом смещении на диоде (плюс к p-области, $V > 0$) высота барьера и толщина обедненного слоя уменьшаются, а при обратном смещении — возрастают.

Оценим характерные значения толщины обедненного слоя и максимальной напряженности электрического поля в p–n-переходе. При $V = 0$ в симметрично легированном p–n-переходе из Si с $N_d = N_a = 4 \cdot 10^{15}$ см$^{-3}$ $W \approx 0{,}66$ мкм. Как следует из рис. 1.1, максимальная напряженность электрического поля в резком p–n-переходе достигается на границе p- и n-областей. Нетрудно показать, что для резкого перехода она равна $\mathcal{E}_{\max} = 2(\phi_k - V)/W(V)$, то есть для рассматриваемого нами примера значение $\mathcal{E}_{\max}$ при $V = 0$ составляет $\approx 2 \cdot 10^4$ В/см. При более низких уровнях легирования и подаче обратного смещения толщина обедненного слоя может достигать сотен мкм.

Рассмотрим теперь другую важную и допускающую аналитическое решение задачу — случай *плавного* p–n-перехода (см. рис. 1.2). Плавным называют переход, в котором разность концентраций $N_d - N_a$ является линейной функцией координаты $N_d - N_a = -ax$. [2])

---

[1]) О явлении инжекции мы будем говорить в п. 1.2.1.

[2]) Такое распределение примесей часто получается в диодах, созданных диффузией, когда характерный масштаб размытия диффузионного профиля намного превышает толщину области пространственного заряда

НАУЧНАЯ БИБЛИОТЕКА

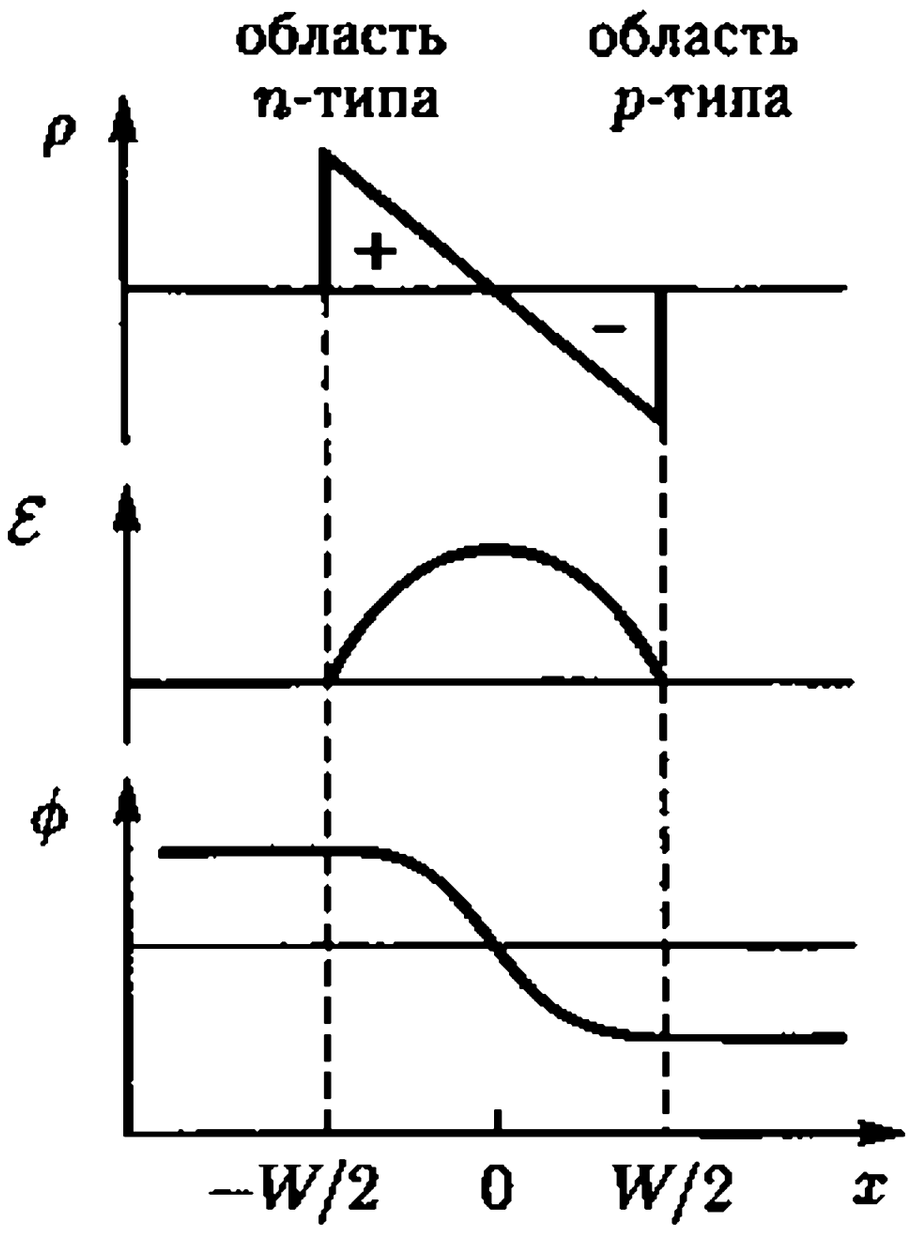


Рис. 1.2. Распределение плотности заряда ρ, электрического поля $\mathcal{E}$ и потенциала ϕ в плавном p–n-переходе

В этом случае уравнение Пуассона (1.6) приобретает вид

$$-\frac{d^2\phi}{dx^2} = \frac{d\mathcal{E}}{dx} = \frac{4\pi\rho}{\varepsilon} = \\ = \frac{4\pi q}{\varepsilon}(p - n - ax), \tag{1.14}$$

где a — градиент разности концентраций легирующих примесей. Как и выше, в области пространственного заряда ($-W/2 < x < W/2$) пренебрежем вкладом свободных носителей ($|p - n| \ll a|x|$) и будем решать уравнение с граничными условиями $\mathcal{E} = 0$ при $x = \pm W/2$. [1])

Интегрирование уравнения (1.14) для случая произвольного напряжения смещения V на переходе дает:

$$\mathcal{E}(x) = \frac{4\pi qa}{\varepsilon}\frac{(W/2)^2 - x^2}{2}, \quad \phi_k - V = \frac{\pi qaW^3}{3\varepsilon}, \tag{1.15}$$

откуда для толщины обедненного слоя в плавном p–n-переходе получаем

$$W = \left(\frac{3\varepsilon(\phi_k - V)}{\pi qa}\right)^{1/3}. \tag{1.16}$$

Контактную разность потенциалов ϕ_k в плавном p–n-переходе можно оценить по формуле (1.4), считая ее равной изменению потенциала в области пространственного заряда:

$$\phi_k \approx \frac{kT}{q}\ln\left(\frac{n(W/2)\,p(W/2)}{n_i^2}\right) = \frac{2kT}{q}\ln\left(\frac{aW}{2n_i}\right). \tag{1.17}$$

В плавных переходах величина ϕ_k обычно оказывается несколько меньше, чем в резких p–n-переходах.

Как видно из рис. 1.2, распределение электрического поля в плавном p–n-переходе более однородно, чем в резком переходе;

---

[1]) Строго говоря, из-за неоднородного легирования полупроводника напряженность электрического поля вдали от плавного p–n-перехода равна $|\mathcal{E}| \approx \approx kT/(q|x|)$ и никогда не обращается в нуль. Тем не менее, мы можем пользоваться граничными условиями $\mathcal{E} = 0$, поскольку это поле намного меньше максимального поля в p–n-переходе.

нетрудно показать, что максимальная напряженность поля в таком p–n-переходе равна $\mathcal{E}_{\max} = 3(\phi_k - V)/2W(V)$.

1.2. Вольт-амперная характеристика p–n-перехода

1.2.1. Вольт-амперная характеристика тонкого p–n-перехода. В этом разделе мы получим аналитическое выражение для вольт-амперной характеристики идеального p–n-перехода, которое было выведено Шокли в 1949 году [3].

Идеальным p–n-переходом называют такой переход, для которого выполняются следующие условия:

1) границы p–n-перехода резкие (то есть электрическое поле обращается точно в нуль на его краях);
2) полупроводник — невырожденный (к электронам и дыркам применима статистика Больцмана);
3) концентрация инжектированных неосновных носителей мала по сравнению с концентрацией основных носителей;
4) процессами генерации и рекомбинации в области пространственного заряда можно пренебречь (переход является *тонким*, то есть толщина области пространственного заряда много меньше диффузионной длины).

В п. 1.1 мы показали, что на границе полупроводников разного типа проводимости возникает потенциальный барьер, который препятствует диффузии носителей, и полный ток через p–n-переход в состоянии равновесия равен нулю. Рассмотрим, что будет происходить с диффузионным и дрейфовым токами, если к p–n-переходу приложить внешнее смещение.

Качественно картину происходящего можно представить следующим образом. В состоянии равновесия (при нулевом напряжении смещения, см. рис. 1.3a) высота потенциального барьера $q\phi_k$ такова, что потоки носителей, протекающие через p–n-переход в обоих направлениях, точно скомпенсированы. Например, поток электронов, движущихся из n- в p-область за счет диффузии и преодолевающих потенциальный барьер, в точности равен потоку неосновных электронов, генерируемых в p-области, которые, подходя к p–n-переходу, затягиваются электрическим полем и дрейфуют в n-область. То же справедливо и для дырок.

Если теперь на p–n-переход подать напряжение смещения, то равновесие нарушится, потоки окажутся нескомпенсированными и через переход потечет электрический ток. Ясно, что значение тока будет зависеть от знака приложенного

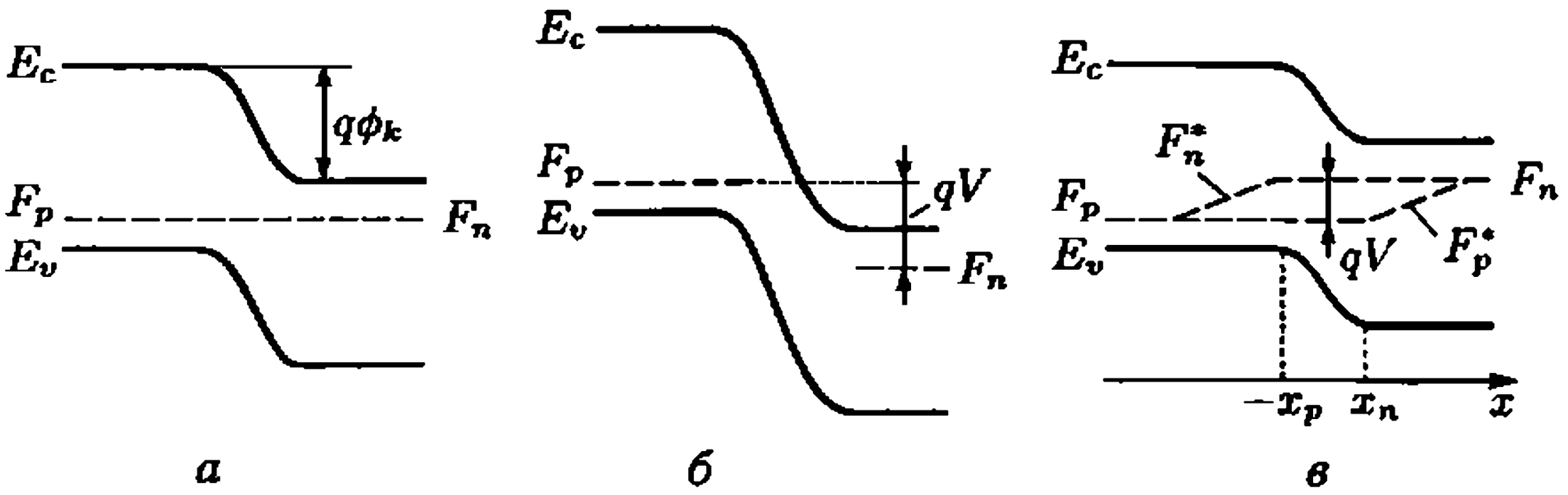


Рис. 1.3. Энергетическая диаграмма p–n-перехода при нулевом (*а*), обратном (*б*) и прямом (*в*) смещении

напряжения. Если смещение отрицательно (минус к p-области, рис. 1.3*б*), то высота потенциального барьера возрастает и поток носителей, преодолевающих потенциальный барьер, резко уменьшается. При этом ток неосновных носителей, затягиваемых электрическим полем p–n-перехода, не изменяется, и полный ток будет определяться этим током. Если же на p–n-переход подать положительное смещение V (плюс к p-области, рис. 1.3*в*), то высота потенциального барьера понижается и ток через p–n-переход будет уже определяться током носителей, преодолевающих потенциальный барьер. Поскольку к электронам применима классическая (больцмановская) статистика и число электронов, энергия которых достаточна для преодоления барьера высотой $q(\phi_k - V)$, изменяется пропорционально $\exp[-q(\phi_k - V)/kT]$, то при $V > 0$ ток будет экспоненциально увеличиваться с ростом приложенного смещения. Преодолевшие потенциальный барьер носители, которые становятся неосновными носителями при переходе в область другого типа проводимости и концентрация которых вблизи p–n-перехода намного превышает равновесную концентрацию неосновных носителей, далее движутся в электрически нейтральных областях в основном за счет диффузии.

Таким образом, на основании проведенных качественных рассуждений можно ожидать, что вольт-амперная характеристика p–n-перехода будет сильно нелинейной. Направление смещения, при котором через p–n-переход протекает большой ток, называют прямым (пропускным), а противоположное направление — обратным (запорным).

Явление, при котором при пропускании электрического тока через смещенный в прямом направлении p–n-переход в полупроводнике создаются избыточные концентрации неосновных носителей, называют *инжекцией неосновных носителей заряда*.

Выведем теперь формулу, описывающую вольт-амперную характеристику идеального (тонкого) p–n-перехода.

Для начала введем понятие *квазиуровней Ферми*. Часто оказывается, что в неравновесных условиях, когда концентрации носителей заряда в полупроводнике отличаются от равновесных, распределение носителей по энергиям тем не менее можно описать распределением Ферми–Дирака или его частным случаем — распределением Больцмана (в невырожденном полупроводнике). [1]) В этом случае все формулы для статистики электронов и дырок остаются такими же, как и в равновесии, но с единственным отличием: в них вместо равновесного уровня Ферми формально стоит некоторая новая величина, которую и называют квазиуровнем Ферми. Тогда для расчета концентраций свободных электронов и дырок можно использовать формулы (1.3), в которых вместо равновесного уровня Ферми стоят соответствующие квазиуровни F_n^* и F_p^*:

$$n = N_c \exp\left(-\frac{E_c - F_n^*}{kT}\right), \quad p = N_v \exp\left(-\frac{F_p^* - E_v}{kT}\right). \qquad (1.18)$$

Другими важными соотношениями, которые могут быть использованы при описании полупроводников в неравновесных условиях, являются соотношения, связывающие полные плотности токов электронов и дырок (1.2) с наклонами соответствующих квазиуровней Ферми [1]:

$$\mathbf{J}_n = q\mu_n n \nabla F_n^*, \quad \mathbf{J}_p = q\mu_p p \nabla F_p^*. \qquad (1.19)$$

В случае, когда плотность токов мала, а концентрации носителей достаточно велики, эти соотношения позволяют обосновать введение единых квазиуровней Ферми в ограниченных областях пространственно неоднородных систем.

В соответствии с рассмотренной выше качественной картиной, протекание тока при подаче прямого смещения на p–n-переход включает два этапа: 1) преодоление носителями потенциального барьера и 2) их диффузия в электрически нейтральных областях.

Для определенности рассмотрим дырочную составляющую тока в p–n-переходе. Если плотность тока не слишком высока и в обедненном слое можно пренебречь генерацией и рекомбинацией носителей, то при произвольном смещении концентрация

---

[1]) Для этого необходимо, чтобы времена жизни неравновесных носителей намного превышали характерные времена релаксации импульса и энергии [1].

дырок в любой точке обедненного слоя будет определяться положением квазиуровня Ферми F_p^*, который совпадает с уровнем Ферми в p-области (см. рис. 1.3*в*). Отсюда сразу же следует, что концентрация дырок на границе n-области, примыкающей к p–n-переходу (в точке $x = x_n$ на рис. 1.3*в*), равна

$$p_n(x_n) = p_0 \exp\left(-\frac{q(\phi_k - V)}{kT}\right) = p_{n0} \exp\left(\frac{qV}{kT}\right), \qquad (1.20)$$

где p_{n0} — равновесная концентрация неосновных дырок в n-области.

Поскольку при $V > 0$ концентрация инжектированных дырок на границе нейтральной n-области ($x = x_n$) превышает их концентрацию в глубине n-области (p_{n0}), то начинается диффузия этих носителей. Величина тока дырок при этом определяется тем, насколько быстро преодолевшие барьер носители уходят от границы p–n-перехода за счет диффузии.

Чтобы рассчитать диффузионный ток, нам надо решить систему уравнений непрерывности с учетом рекомбинации неосновных носителей [1]:

$$\frac{dp}{dt} = -\frac{1}{q}\nabla \cdot \mathbf{J}_p + G - R, \quad \frac{dn}{dt} = \frac{1}{q}\nabla \cdot \mathbf{J}_n + G - R, \qquad (1.21)$$

где G — темп генерации, а R — темп рекомбинации. При линейном законе рекомбинации имеем $R - G = (n - n_0)/\tau$, где τ — время жизни неосновных носителей. Рассмотрим стационарную одномерную задачу и подставим в уравнения (1.21) выражения для токов электронов и дырок (1.2). Если считать, что подвижности и коэффициенты диффузии носителей не зависят от их концентраций, а подвижности также не зависят и от напряженности электрического поля, [1]) то уравнения непрерывности принимают вид

$$D_n \frac{d^2n}{dx^2} + \mu_n n \frac{d\mathcal{E}}{dx} + \mu_n \mathcal{E} \frac{dn}{dx} - \frac{n - n_0}{\tau} = 0, \qquad (1.22)$$

$$D_p \frac{d^2p}{dx^2} - \mu_p p \frac{d\mathcal{E}}{dx} - \mu_p \mathcal{E} \frac{dp}{dx} - \frac{p - p_0}{\tau} = 0. \qquad (1.23)$$

Исключая из этих уравнений слагаемые, пропорциональные $d\mathcal{E}/dx$, и учитывая, что в полупроводнике обычно выполняется условие локальной электронейтральности ($n - n_0 \approx p - p_0$), мы приходим к хорошо известному уравнению, описывающему

[1]) О случаях, в которых этих условия не выполняются, см. с. 39.

совместное (*амбиполярное*) движение (диффузию и дрейф) электронов и дырок:

$$D_a \frac{d^2n}{dx^2} + \mu_a \mathcal{E} \frac{dn}{dx} - \frac{n - n_0}{\tau} = 0, \qquad (1.24)$$

где

$$D_a = \frac{n + p}{n/D_p + p/D_n}, \quad \mu_a = \frac{n - p}{n/\mu_p + p/\mu_n} \qquad (1.25)$$

— коэффициент амбиполярной диффузии и амбиполярная подвижность, соответственно. [1)]

При низком уровне инжекции ($p \ll n$) величина $D_a \approx D_p$, а напряженность электрического поля в нейтральной области n-типа, примыкающей к p–n-переходу, мала [2]. Поэтому в уравнении непрерывности (1.24) можно пренебречь слагаемым, содержащим $\mathcal{E}$, и это уравнение принимает вид

$$D_p \frac{d^2p}{dx^2} - \frac{p - p_{n0}}{\tau} = 0. \qquad (1.26)$$

Решением этого уравнения с граничным условием (1.20) на краю нейтральной области (при $x = x_n$) и граничным условием $p = p_{n0}$ при $x \to \infty$ (будем считать n-область достаточно толстой) является

$$p_n(x) - p_{n0} = p_{n0}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]\exp\left(-\frac{x - x_n}{L_p}\right), \qquad (1.27)$$

где $L_p = \sqrt{D_p\tau}$ — диффузионная длина дырок. Уменьшение концентрации инжектированных дырок с ростом x, описываемое этой формулой, проявляется на рис. 1.3*в* в приближении F_p^* к равновесному уровню Ферми F_n в n-области по мере удаления от p–n-перехода. Дифференцируя уравнение (1.27) по x, находим плотность тока диффузии дырок на границе n-области:

$$J_p = -qD_p \left.\frac{dp}{dx}\right|_{x=x_n} = \frac{qD_p p_{n0}}{L_p}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \qquad (1.28)$$

[1)] Напомним, что благодаря достаточно сильному электростатическому взаимодействию электронов и дырок их диффузия и дрейф в полупроводнике происходят *совместно*, в виде квазинейтрального пакета. Чтобы описать характеристики такого движения, и вводятся амбиполярный коэффициент диффузии и амбиполярная подвижность.

Это и есть плотность дырочной составляющей тока, протекающего через p–n-переход при подаче на него напряжения смещения V.

Аналогичные выкладки можно повторить и для электронной компоненты тока. Конечное выражение для плотности тока диффузии электронов, инжектируемых в p-область, выглядит так:

$$J_n = qD_n \frac{dn}{dx}\bigg|_{x=-x_p} = \frac{qD_n n_{p0}}{L_n}\left[\exp\left(\frac{qV}{kT}\right) - 1\right]. \qquad (1.29)$$

Поскольку в модели тонкого p–n-перехода полный ток складывается из тока инжектированных электронов и дырок, а генерацией и рекомбинацией в области пространственного заряда можно пренебречь, вольт-амперная характеристика тонкого p–n-перехода принимает вид

$$J = J_s\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \qquad (1.30)$$

где

$$J_s = \frac{qD_n n_{p0}}{L_n} + \frac{qD_p p_{n0}}{L_p} = \frac{qL_n n_{p0}}{\tau_n} + \frac{qL_p p_{n0}}{\tau_p} \qquad (1.31)$$

— так называемая *плотность тока насыщения*. Общий вид вольт-амперной характеристики p–n-перехода показан на рис. 1.4.

Оценим плотность тока насыщения германиевого и кремниевого диодов с $N_d = N_a = 4 \cdot 10^{15}$ см$^{-3}$, $\tau_n = \tau_p = 1$ мкс при 300 К. Вычисляя с помощью соотношения Эйнштейна $D = (kT/q)\mu$ коэффициенты диффузии электронов и дырок из известных значений их подвижности (см. табл. 2 в Приложении) и затем рассчитывая диффузионные длины, находим $J_s = 5 \cdot 10^{-4}$ А/см^2 для германиевого диода и $4 \cdot 10^{-11}$ А/см^2 для кремниевого.

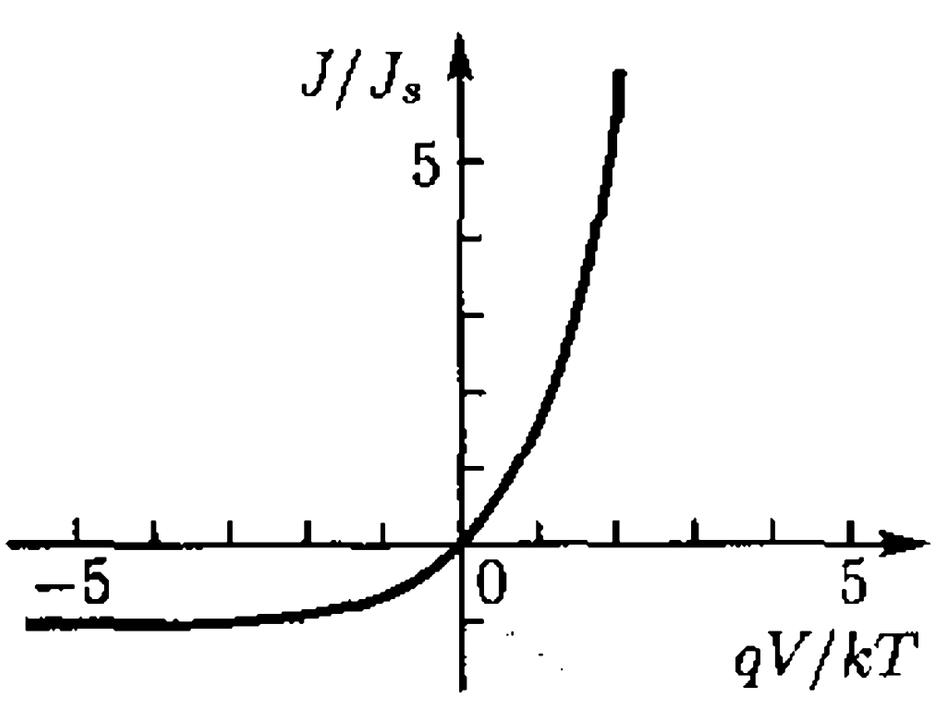


Рис. 1.4. Вольт-амперная характеристика тонкого p–n-перехода

Физический смысл формулы (1.31) заключается в том, что плотность тока насыщения есть ни что иное как полный заряд неравновесных носителей, возбужденных за единицу времени в слое p-области толщиной L_n и слое n-области толщиной L_p, который протекает через p–n-переход единичной площади.

Ток насыщения диода сильно зависит от температуры; наиболее

сильной температурной зависимостью в формуле (1.31) характеризуются величины n_{p0} и p_{n0}. Поскольку $n_{p0} \approx n_i^2/N_a$, $p_{n0} \approx \approx n_i^2/N_d$, а температурная зависимость n_i описывается формулой (1.1), то ток насыщения в тонком p–n-переходе изменяется с температурой приблизительно как $J_s \sim \exp(-E_g/kT)$.

Резкое возрастание тока тока насыщения с увеличением температуры приводит к существованию *максимальной рабочей температуры* p–n-перехода. С повышением температуры собственная концентрация носителей в полупроводнике быстро возрастает, контактная разность потенциалов ϕ_k согласно формуле (1.4) уменьшается и, следовательно, выпрямительные свойства p–n-перехода ухудшаются. Максимальную рабочую температуру $T_{\max}$ можно оценить из условия $q\phi_k(T_{\max}) \approx \approx kT_{\max}$. Для типичной концентрации примесей $N_d \approx N_a \approx 10^{15}$ см$^{-3}$ эта температура составляет ~100 °C для диодов из Ge, ~270 °C для диодов из Si и ~500 °C для диодов из GaAs.

Сопоставляя величины электронного и дырочного вкладов в ток p–n-перехода (формулы (1.28) и (1.29)), нетрудно видеть, что в асимметрично легированном p–n-переходе преобладает ток носителей, инжектируемых из области, которая легирована сильнее, в область, которая легирована слабее. Физическая причина этого достаточно проста: носителям обоих знаков при инжекции приходится преодолевать потенциальный барьер одной и тот же высоты, и поэтому бóльший вклад в ток дают те носители, концентрация которых у p–n-перехода выше. Поэтому изменяя соответствующим образом уровни легирования p- и n-областей, можно *управлять направлением инжекции* носителей в p–n-переходе. Это свойство широко используется в различных полупроводниковых приборах, в частности, оно лежит в основе работы биполярных транзисторов и инжекционных лазеров. Влияние более тонких эффектов сильного легирования на эффективность инжекции будет рассмотрено нами в п. 2.2.1.

Диоды с тонкой базой. До сих пор мы полагали, что размеры областей p- и n-типа проводимости практически бесконечны (во всяком случае, намного больше характерных диффузионных длин L_n, L_p). Рассмотрим теперь другой практически важный случай — случай так называемого *диода с тонкой базой («короткого» диода)*, когда размер одной или обеих областей диода оказывается порядка или меньше диффузионной длины. С такой геометрией p–n-перехода мы встречаемся в биполярных транзисторах (см. гл. 2) и в полупроводниковых структурах, изготавливаемых по планарной технологии (см. п. 2.8.1).

Так, глубина залегания p–n-переходов, создаваемых в современных интегральных схемах, достигает ~0,1 мкм.

Рассмотрим p–n-переход с тонкой n-областью. Особенностью этой структуры является то, что на некотором расстоянии X_n от p–n-перехода находится граница, на которой можно считать $p_n(X_n) \approx p_{n0}$ (это может быть омический контакт или другой p–n-переход как, например, в транзисторе). [1]) Второе граничное условие, как и раньше, задается уравнением (1.20). В этом случае решение уравнения (1.26) следует искать в более общем виде:

$$p(x) - p_{n0} = A \exp\left(\frac{x}{L_p}\right) + B \exp\left(-\frac{x}{L_p}\right), \qquad (1.32)$$

где коэффициенты A и B находятся из граничных условий. В результате несложных вычислений для указанных выше граничных условий получаем:

$$J_p = \frac{qD_p p_{n0}}{L_p}\left[\exp\left(\frac{qV}{kT}\right) - 1\right] \operatorname{cth}\left(\frac{X_n - x_n}{L_p}\right). \qquad (1.33)$$

Сравнивая полученное решение с решением (1.28), нетрудно видеть, что они различаются лишь появлением сомножителя cth(...). Если толщина нейтральной части n-области удовлетворяет условию $X_n - x_n \ll L_p$, то справедливо приближенное равенство

$$\operatorname{cth}\left(\frac{X_n - x_n}{L_p}\right) \approx \frac{L_p}{X_n - x_n}. \qquad (1.34)$$

Из него следует, что чем тоньше n-область, тем сильнее возрастает дырочная компонента тока. Физической причиной этого является увеличение градиента концентрации инжектированных в базу дырок с уменьшением толщины n-области (при фиксированном напряжении смещения) и, следовательно, увеличение их тока диффузии.

Несмотря на ряд приближений, сделанных при выводе выражений для вольт-амперной характеристики p–n-перехода, полученные формулы достаточно хорошо описывают характеристики

---

[1]) В общем случае граничное условие при $x = X_n$ следует записывать так: $D_p(dp/dx) = -s(p - p_{n0})$, где s — коэффициент, называемый *скоростью поверхностной рекомбинации*. Идеальному контакту соответствует условие $s \to \infty$. Проблемы, связанные с созданием омических контактов к полупроводникам, будут обсуждаться подробнее в п. 1.5.3.

диодов из Ge как в области прямых (рис. 1.5), так и обратных смещений (рис. 1.6). Отклонение характеристик этих диодов от расчетных зависимостей может быть связано с проявлением

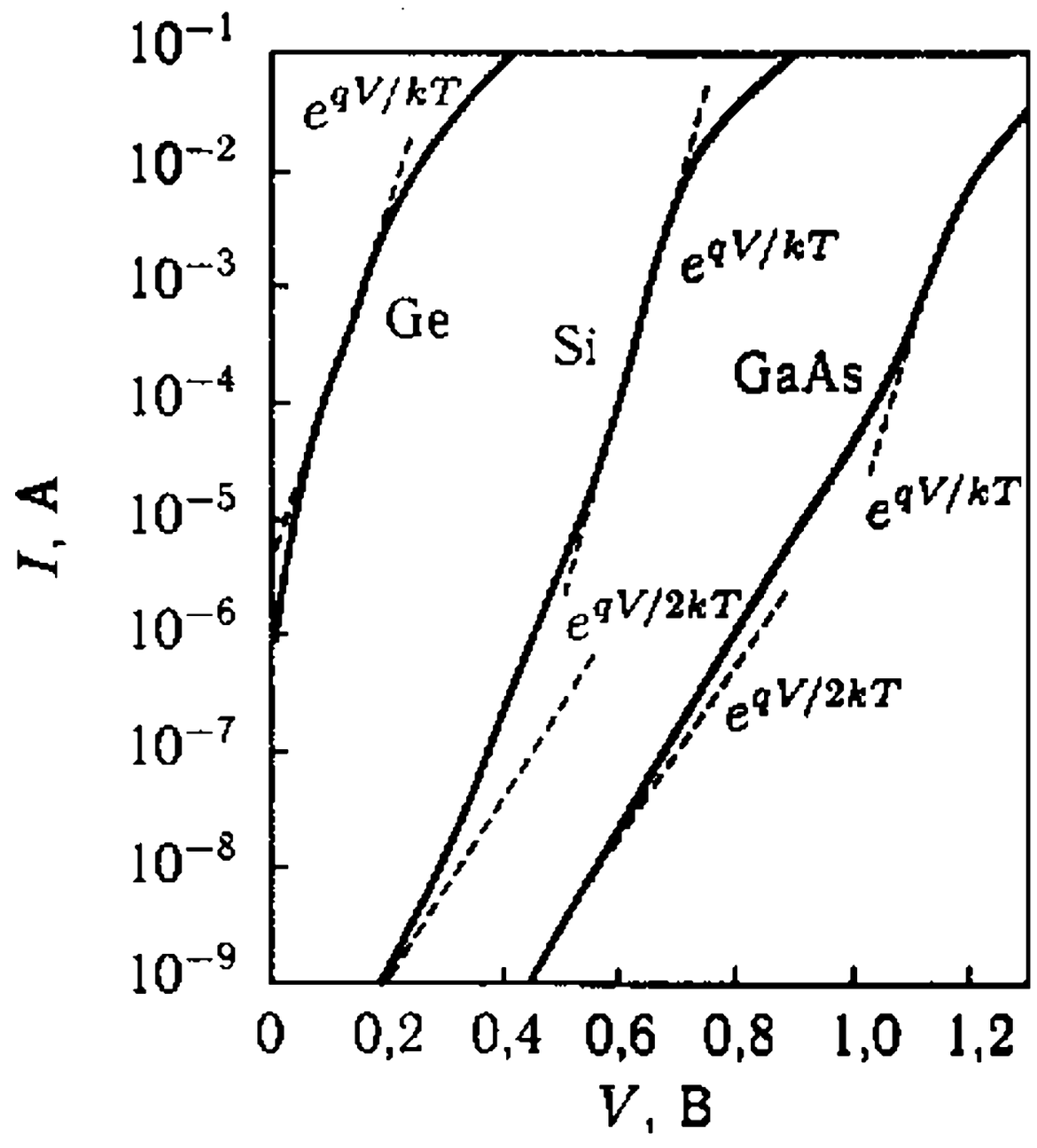


Рис. 1.5. Вольт-амперные характеристики диодов из Ge, Si и GaAs в области прямых токов

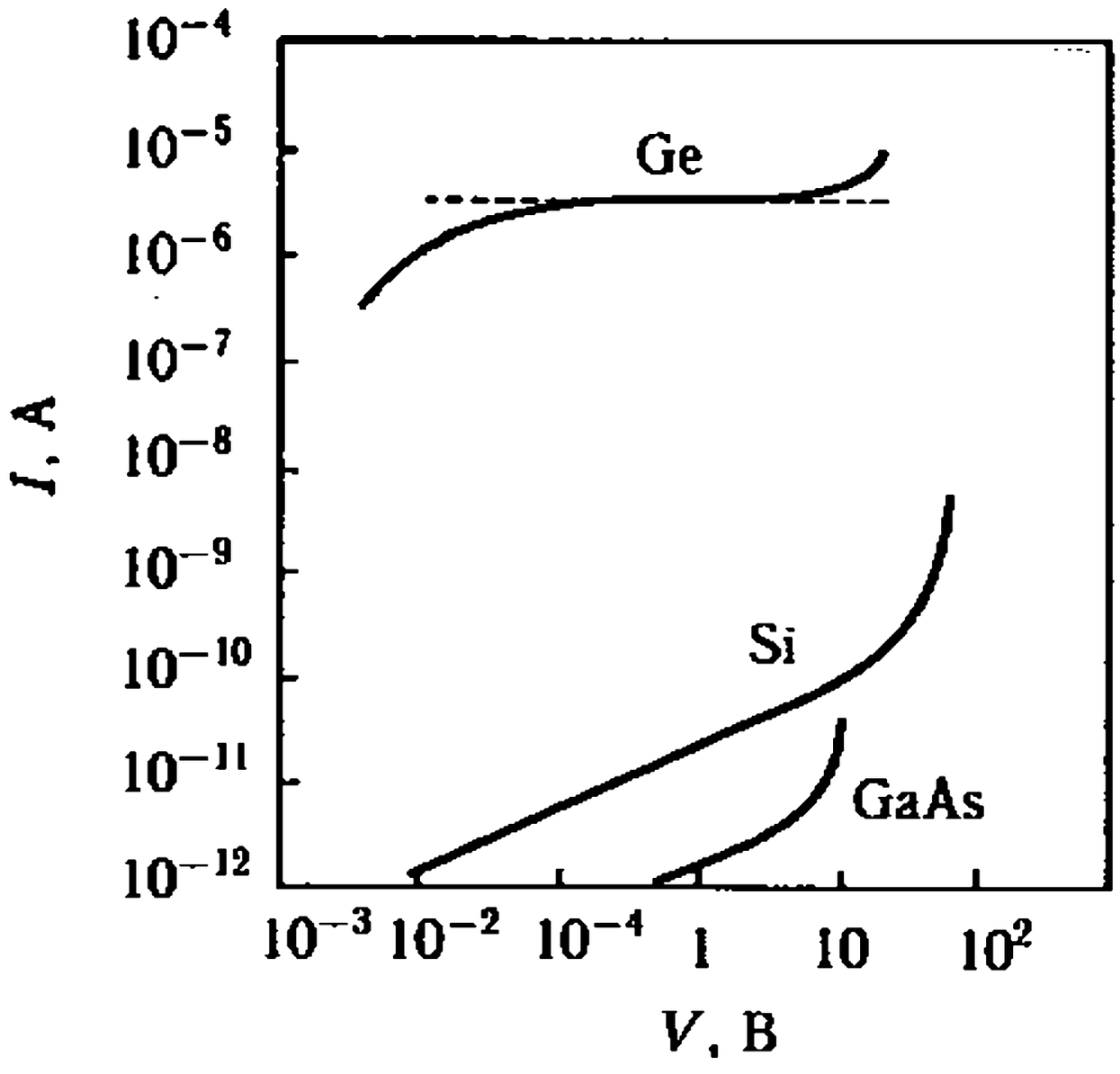


Рис. 1.6. Вольт-амперные характеристики диодов из Ge, Si и GaAs в области обратных токов

последовательного сопротивления областей p–n-перехода и рядом явлений, появляющихся при высоком уровне инжекции

(см. п. 1.2.3); явлениями лавинного умножения и пробоя при высоких обратных напряжениях (см. п. 1.3); явлениями туннелирования (см. п. 1.4); утечками по поверхности приборов.

Исследования вольт-амперных характеристик диодов из более широкозонных полупроводников (Si, GaAs и др.) однако показывают, что эти характеристики заметно отличаются от теоретических: для них характерны более пологий наклон зависимости $\ln I$ от V в прямой ветви (см. рис. 1.5) и отсутствие участка насыщения тока p–n-перехода при обратном смещении (рис. 1.6). Причиной наблюдаемых отклонений является проявление процессов генерации и рекомбинации в области пространственного заряда, которыми мы до сих пор пренебрегали.

1.2.2. Влияние генерации и рекомбинации в области пространственного заряда на вольт-амперные характеристики p–n-перехода (модель Са–Нойса–Шокли). Хорошо известно, что в большинстве полупроводников основным каналом рекомбинации электронов и дырок является рекомбинация с участием *глубоких уровней*, расположенных в запрещенной зоне.

В теории рекомбинации через глубокие уровни, развитой Шокли и Ридом [4] и Холлом [5], выводится следующее общее выражение для разности темпов рекомбинации R и генерации G:

$$R - G = \frac{pn - n_i^2}{\tau_p(n + n_t) + \tau_n(p + p_t)}, \qquad (1.35)$$

где τ_n и τ_p — времена жизни электронов и дырок, а n_t и p_t — концентрации электронов и дырок в полупроводнике, когда уровень Ферми совпадает с положением уровня рекомбинации [1]. Времена жизни носителей связаны с концентрацией глубоких уровней N_t, сечением захвата электронов σ_n и дырок σ_p и скоростью теплового движения носителей $v_T = \sqrt{3kT/m^*}$ следующим образом: $\tau_n = 1/(v_T \sigma_n N_t)$, $\tau_p = 1/(v_T \sigma_p N_t)$. [1]) Из формулы (1.35) следует, что наибольший вклад в генерацию и рекомбинацию дают уровни, для которых выражение в знаменателе оказывается минимальным. Поскольку τ_n и τ_p обычно близки по порядку величины, а $n_t p_t = np = n_i^2$, то этот минимум достигается при $n_t \approx n_i$, то есть наиболее сильное влияние на рекомбинацию носителей оказывают *уровни, расположенные вблизи середины запрещенной зоны.*

---

[1]) Методы определения концентрации глубоких уровней и сечений захвата ими электронов и дырок будут рассмотрены нами в п. 1.7.5.

Са, Нойс и Шокли [6] получили выражения, описывающие вклад процессов генерации и рекомбинации в области пространственного заряда в ток p–n-перехода. Предположим, что в полупроводнике есть только один тип глубоких уровней, и рассмотрим сначала случай обратного смещения на p–n-переходе. Если напряжение смещения велико ($|V| \gg kT/q$), то из-за высокого потенциального барьера можно считать, что свободных носителей в обедненном слое практически нет ($pn \ll n_i^2$, $p \ll p_t$, $n \ll n_t$). Тогда плотность обратного (*генерационного*) тока p–n-перехода определяется темпом тепловой генерации носителей с глубоких уровней в этом слое и может быть записана следующим образом:

$$J_{\text{ген}} = q \int_0^W (R - G)\, dx \approx q \int_0^W \frac{-n_i^2}{\tau_p n_t + \tau_n p_t}\, dx = -\frac{q n_i W}{\tau_{\text{эф}}}, \qquad (1.36)$$

где $\tau_{\text{эф}} = (\tau_p n_t + \tau_n p_t)/n_i$ — эффективное время жизни. Отметим, что интегрирование в (1.36) проводится по *всей* области пространственного заряда толщиной W.

Установим основные закономерности тока генерации. Если предположить, что $\tau_{\text{эф}}$ слабо меняется с температурой, то температурная зависимость обратного тока p–n-перехода будет определяться температурной зависимостью n_i, то есть $|J_{\text{ген}}| \sim \exp(-E_g/2kT)$. Далее, поскольку $|J_{\text{ген}}| \sim W$, а толщина области пространственного заряда при увеличении обратного смещения возрастает (для резкого p–n-перехода $W \sim \sqrt{\phi_k + V}$), то и обратный ток должен соответствующим образом увеличиваться. Как следует из рис. 1.6, зависимость именно такого типа и наблюдается на обратных ветвях вольт-амперных характеристик диодов из Si и GaAs. Наконец, поскольку $|J_{\text{ген}}| \sim 1/\tau_{\text{эф}}$, а $\tau_{\text{эф}} \sim 1/N_t$, то обратный ток p–n-перехода пропорционален концентрации глубоких уровней. Последнее объясняет почему обратный ток p–n-перехода возрастает после облучения диодов ионизирующим излучением, которое создает в материале глубокие уровни *радиационных дефектов*. Нетрудно видеть, что все основные закономерности тока генерации существенно отличаются от закономерностей, полученных в модели тонкого p–n-перехода.

Дополнительный вклад в возрастание обратного тока p–n-перехода при увеличении смещения может также давать *эффект Пула–Френкеля*, заключающийся в уменьшении эффективной энергии ионизации

примесных центров в сильном электрическом поле. Влияние этого эффекта на вольт-амперные характеристики рассмотрено в работе [7].

Рассмотрим теперь прямую ветвь вольт-амперной характеристики. Плотность тока рекомбинации в области пространственного заряда описывается формулой

$$J_{\text{рек}} = q\int_0^W (R-G)\,dx = q\int_0^W \frac{p(x)n(x)-n_i^2}{\tau_p\left[n(x)+n_t\right]+\tau_n\left[p(x)+p_t\right]}\,dx. \quad (1.37)$$

Рассмотрим для простоты симметрично легированный p–n-переход, в котором уровень рекомбинации расположен «посередине» запрещенной зоны ($n_t = p_t = n_i$), [1]) а $\tau_n = \tau_p = \tau$. Если ширина области пространственного заряда мала по сравнению с диффузионными длинами, то изменение концентраций $n(x)$ и $p(x)$ в области пространственного заряда можно описать введением единых квазиуровней Ферми. Для вычисления интеграла (1.37) удобно ввести величину $q\psi$, равную разности энергий между «серединой» запрещенной зоны в каждой точке области пространственного заряда и полусуммой энергий квазиуровней Ферми для электронов и дырок:

$$q\psi(x) = E_c(x) - \frac{E_g}{2} - \left(\frac{kT}{2}\right)\ln\left(\frac{N_c}{N_v}\right) - \frac{F_n^* + F_p^*}{2}.$$

Учитывая, что энергетический зазор между квазиуровнями Ферми равен $F_n^* - F_p^* = qV$, профили концентраций можно представить в виде

$$p(x) = n_i \exp\left[\frac{qV}{2kT} + \frac{q\psi(x)}{kT}\right], \quad n(x) = n_i \exp\left[\frac{qV}{2kT} - \frac{q\psi(x)}{kT}\right].$$

Нетрудно показать, что наибольший вклад в рекомбинацию дает *узкий слой* в области пространственного заряда, в котором $n \approx p$ и значение знаменателя в подынтегральном выражении в формуле (1.37) минимально. При $V \gg kT/q$ слагаемыми n_t, p_t в знаменателе по сравнению с n и p можно пренебречь, и тогда

---

[1]) Под уровнем, расположенным «посередине» запрещенной зоны, здесь мы будем понимать уровень, для которого $n_t = p_t$. В полупроводнике с $N_c \neq N_v$ этот уровень лежит на $E_g/2 + (kT/2)\ln(N_c/N_v)$ ниже края зоны проводимости.

после перехода от интегрирования по x к интегрированию по ψ выражение (1.37) превращается в

$$J_{\text{рек}} \approx \frac{qn_i}{\tau} \exp\left(\frac{qV}{2kT}\right) \int\limits_{\psi(0)}^{\psi(W)} \frac{(dx/d\psi)}{2\,\mathrm{ch}(q\psi/kT)}\, d\psi \approx \tag{1.38}$$

$$\approx \frac{\pi kTn_i}{2\tau\mathcal{E}_0} \exp\left(\frac{qV}{2kT}\right),$$

где $\mathcal{E}_0 = (d\psi/dx)|_{\psi=0}$ — напряженность электрического поля в точке p–n-перехода, в которой в неравновесных условиях $n = p$. Из сопоставления этой формулы с (1.36) следует, что эффективная толщина слоя, дающего основной вклад в ток рекомбинации, равна $\pi kT/q\mathcal{E}_0$, что составляет $\sim 1/20$ от полной ширины области пространственного заряда W. Из формулы (1.38) также следует, что в случае, когда ток в p–n-переходе определяется рекомбинацией в области пространственного заряда, зависимость тока от напряжения смещения остается экспоненциальной, однако наклон этой зависимости в полулогарифмическом масштабе уменьшается до $q/2kT$. Такая зависимость действительно наблюдается в эксперименте на диодах из Si и GaAs при небольших прямых смещениях (см. рис. 1.5). [1]) Как и в случае обратного смещения, при прямом смещении плотность тока рекомбинации пропорциональна N_t.

При выводе формулы (1.38) предполагалось, что уровень рекомбинации расположен посередине запрещенной зоны, а времена жизни равны. Если это не так, то расчет зависимости $J_{\text{рек}}(V)$ становятся более громоздким, а выражения — сложными для анализа. Авторы работы [6] показали, что если положение уровня рекомбинации более, чем на $10kT$ удаляется от середины запрещенной зоны, то коэффициент m в эмпирической формуле

$$J_{\text{рек}} \sim \exp\left(\frac{qV}{mkT}\right) \tag{1.39}$$

становится близким к 1. Таким образом, в зависимости от параметров рекомбинационных центров, величина m может изменяться в пределах от 1 до 2. Эту величину называют *фактором идеальности*.

---

[1]) Наиболее заметную роль в рекомбинации в кремнии играют акцепторные уровни золота с энергией $E_c - 0,54$ эВ, а в арсениде галлия — донорные уровни дефектов решетки EL2 с энергией $E_c - 0,82$ эВ.

становится преобладающим и наклон кривой $\ln J(V)$ изменяется от $q/2kT$ до q/kT.

1.2.3. p–n-переход при высоких уровнях инжекции. Падение напряжения при протекании тока через толщу p- и n-областей диода и контакты всегда немного искажает вольт-амперную характеристику. Иногда эти изменения удается описать введением некоторого последовательного сопротивления R_s, которое при протекании через диод тока I увеличивает падение напряжения на диоде на величину IR_s.

Однако при высоких плотностях тока через p–n-переход на вольт-амперную характеристику начинают влиять *эффекты модуляции проводимости* областей диода, и это влияние уже нельзя учесть простым введением последовательного сопротивления. Наиболее ярко эти эффекты выражены в асимметрично легированных p–n-переходах, в которых база диода слабо легирована. [1])

Рассмотрим в качестве примера p^+–n-переход. Если на него подать большое прямое смещение, то концентрация инжектированных в n-область дырок Δp может превысить равновесную концентрацию электронов n_0 в этой области. Чтобы скомпенсировать заряд инжектированных носителей и обеспечить локальную электронейтральность базы, с контакта в объем n-области подходят избыточные электроны, локальная концентрация которых возрастает на $\Delta n = n - n_0 \approx \Delta p$. [2]) Как мы покажем ниже, в этих условиях вольт-амперные характеристики диода существенно изменяются. Оценка плотности тока, отвечающей условию $\Delta p = n_0$, для n-Si с $n_0 = 4 \cdot 10^{15}$ см$^{-3}$ и $L_p = 50$ мкм дает $J_p = qD_p\Delta p/L_p \approx 1{,}6$ А/см^2. Поскольку выпрямительные диоды обычно работают при плотностях тока 10–100 А/см^2, анализ случая высокого уровня инжекции представляется весьма актуальным.

Рассчитаем вольт-амперную характеристику *короткого* p^+–n-перехода (диода с тонкой базой) в случае, когда концентрация инжектированных дырок сравнима или превышает

---

[1]) Эта ситуация очень характерна для силовых выпрямительных диодов, которые работают при высоких плотностях тока и в которых для получения высокого напряжения пробоя база диода намеренно легируется довольно слабо (см. п. 1.3).

[2]) Здесь и далее в этом разделе мы предполагаем, что в полупроводнике нет глубоких уровней, способных изменять свое зарядовое состояние при инжекции дырок. Случай, когда это условие не выполняется, подробно разобран в [8].

концентрацию равновесных электронов в базе диода. Мы намеренно рассматриваем случай короткого диода, поскольку из-за одностороннего характера инжекции в асимметрично легированном p–n-переходе и несущественности рекомбинации (толщина n-области X_n мала по сравнению с диффузионной длиной дырок) в этом случае можно считать, что ток переносится в основном дырками, а плотность тока электронов $J_n \approx 0$.

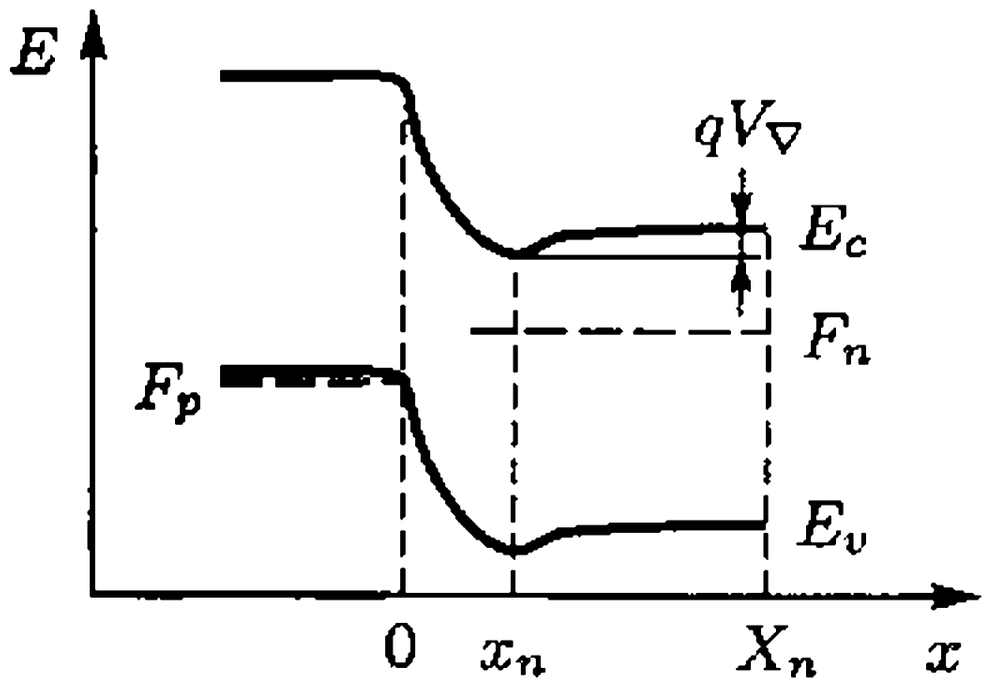


Рис. 1.7. Энергетическая диаграмма p^+–n-диода в условиях высокого уровня инжекции

Из условия

$$J_n = qn\mu_n \mathcal{E} + qD_n \frac{dn}{dx} = 0 \quad (1.41)$$

следует, что в базе диода возникает объемное электрическое поле (см. рис. 1.7), напряженность которого равна

$$\mathcal{E} = -\frac{D_n}{\mu_n}\frac{1}{n}\frac{dn}{dx} = -\frac{kT}{q}\frac{1}{n}\frac{dn}{dx} \approx -\frac{kT}{q}\frac{1}{n}\frac{dp}{dx}, \quad (1.42)$$

где последнее соотношение следует из условия локальной электронейтральности, $\Delta n \approx \Delta p$. Это поле создает в n-области ток дрейфа дырок, направленный в том же направлении, что и их диффузия. При высоком уровне инжекции ($p \approx n \gg n_0$) полная плотность тока дырок равна

$$J_p = q\mu_p p\,\mathcal{E} - qD_p \frac{dp}{dx} = -qD_p\left(1 + \frac{p}{n}\right)\frac{dp}{dx} \approx -2qD_p \frac{dp}{dx}. \quad (1.43)$$

Если исключить из рассмотрения омическое падение напряжения на n-области, то разность потенциалов, связанная с возникновением этого объемного поля, равна

$$V_\nabla = \int\limits_{x_n}^{X_n} \mathcal{E}\,dx = -\frac{kT}{q}\int\limits_{x_n}^{X_n}\frac{1}{n}\frac{dn}{dx}\,dx = -\frac{kT}{q}\ln\left(\frac{n_0}{n(x_n)}\right), \quad (1.44)$$

где $n(x_n)$ — концентрация электронов на границе p–n-перехода. При выводе этого уравнения мы полагали, что на контакте к n-области $n(X_n) = n_0$, то есть достигающие контакта дырки уходят из образца. [1]) Обозначая через V_0 величину смещения

[1]) Для этого необходимо, чтобы скорость поверхностной рекомбинации на контакте была достаточно велика (см. п. 1.5).

на самом p–n-переходе (без учета объемного поля) и учитывая, что при высоком уровне инжекции $n(x_n) \approx p(x_n) = p_{n0}\exp(qV_0/kT)$, преобразуем уравнение (1.44) к виду

$$V_\nabla = \frac{kT}{q}\ln\left(\frac{n(x_n)}{n_0}\right) = \frac{kT}{q}\ln\left(\frac{p_{n0}\,n(x_n)}{n_i^2}\right) \approx \frac{kT}{q}\ln\left(\frac{p_{n0}\,p(x_n)}{n_i^2}\right) =$$

$$= \frac{kT}{q}\ln\left[\left(\frac{p_{n0}^2}{n_i^2}\right)\exp\left(\frac{qV_0}{kT}\right)\right] = \frac{2kT}{q}\ln\left(\frac{p_{n0}}{n_i}\right) + V_0. \qquad (1.45)$$

Поскольку приложенное к диоду смещение равно сумме, $V = V_0 + V_\nabla$, то

$$V = V_0 + V_\nabla = 2V_0 + \frac{2kT}{q}\ln\left(\frac{p_{n0}}{n_i}\right), \qquad (1.46)$$

откуда

$$V_0 = \frac{V}{2} - \frac{kT}{q}\ln\left(\frac{p_{n0}}{n_i}\right). \qquad (1.47)$$

Теперь, выражая плотность диффузионного тока J_p, рассчитываемую по формуле (1.43), через приложенное напряжение V, находим

$$J_p \approx -2qD_p\left.\frac{dp}{dx}\right|_{x=x_n} \approx \frac{2qD_p\,p_{n0}}{X_n - x_n}\exp\left(\frac{qV_0}{kT}\right) =$$

$$= \frac{2qD_p\,n_i}{X_n - x_n}\exp\left(\frac{qV}{2kT}\right). \qquad (1.48)$$

При выводе этой формулы мы предположили, что напряженность электрического поля в базе диода (1.42) не слишком велика, так что распределение концентрации дырок в базе можно описать линейной функцией, как в коротком диоде. [1])

Из формулы (1.48) следует, что при высоком уровне инжекции вольт-амперная характеристика диода действительно изменяется, причем ее наклон в полулогарифмическом масштабе становится вдвое меньше, чем при низком уровне инжекции. Физический смысл этого изменения состоит в том, что при высоком уровне инжекции из-за увеличения локальной концентрации электронов в базе (которые компенсируют заряд

[1]) Проверка условия малости поля (см. с. 134) показывает, что это предположение можно считать еще применимым.

инжектированных дырок) в соответствии с формулой (1.4) контактная разность потенциалов ϕ_k в p–n-переходе *также возрастает*, причем изменяется с ростом тока вдвое медленнее по сравнению с приложенным к диоду напряжением. Очевидно, что этот вывод о возрастании ϕ_k с увеличением уровня инжекции остается верным и для диодов с произвольной толщиной базы. [1])

Расчет вольт-амперных характеристик p–n-перехода с *толстой* базой и омическим тыловым контактом в условиях высокого уровня инжекции был выполнен Стафеевым [9]. Им было показано, что для произвольного отношения толщины базы к амбиполярной длине диффузии [2]) W/L_a значение фактора идеальности m в зависимости $J \sim \exp(qV/mkT)$ равно

$$m = 2\,\frac{b + \mathrm{ch}(W/L_a)}{b+1}\,, \tag{1.49}$$

где $b = \mu_n/\mu_p$ — отношение подвижностей электронов и дырок. Значение $m \approx 2$, следующее из этой формулы при $W/L_a \ll \ll 1$, согласуется с полученным выше результатом. Эта формула предсказывает, что с ростом W/L_a наклон зависимости $\ln J$ от V в прямой ветви должен становиться все более пологим. Расчет вольт-амперных характеристик для случая высокого уровня инжекции при разных граничных условиях и с учетом перезарядки глубоких уровней можно найти в [8].

При еще более высоких плотностях тока электрическое поле в базе диода становится настолько сильным, что изменяется само пространственное распределение инжектированных носителей в n-области [2]. В этой ситуации мы больше не можем пользоваться уравнением (1.26), которое является частным случаем уравнения непрерывности и применимо только в случае слабого электрического поля, и должны искать более общее решение.

К сожалению, решение этой задачи в общем виде слишком сложно, поэтому рассмотрим случай очень высокой плотности тока, когда в уравнении непрерывности можно пренебречь

---

[1]) Еще одним следствием высокого уровня инжекции, следующим из описанной качественной картины, является увеличение отношения токов J_n/J_p с ростом уровня инжекции. В п. 2.2.3 мы увидим, что этот эффект определяет уменьшение коэффициента усиления биполярных транзисторов, работающих в области больших токов.

[2]) При высоком уровне инжекции ($p \approx n$) при вычислении диффузионной длины по формуле $L = \sqrt{D\tau}$ мы должны использовать *амбиполярный* коэффициент диффузии, определяемый формулой (1.25). Полученная величина L_a называется амбиполярной длиной диффузии.

диффузионным током по сравнению с дрейфовым током. [1]) Тогда уравнение непрерывности для дырок (1.21) в одномерном стационарном случае приобретает вид

$$\frac{1}{q}\frac{dJ_p}{dx} = -\frac{p - p_{n0}}{\tau_p}. \qquad (1.50)$$

Если через образец протекает ток плотностью J, то напряженность локального электрического поля $\mathcal{E}(x)$ в квазинейтральной области, в которой локальные концентрации носителей связаны соотношением $p(x) \approx n(x) - n_0$, равна

$$\mathcal{E}(x) = \frac{J}{\sigma(x)} = \frac{J}{q\mu_p\left[(b+1)p(x) + bn_0\right]}, \qquad (1.51)$$

где $b = \mu_n/\mu_p$ — отношение подвижностей электронов и дырок, которое мы будем считать не зависящим от концентрации носителей. Из формулы (1.51) следует, что плотность дырочного тока равна

$$J_p = q\mu_p\, p(x)\mathcal{E}(x) =$$

$$= \frac{Jp(x)}{(b+1)p(x) + bn_0} = \frac{J}{(b+1) + bn_0/p(x)}. \qquad (1.52)$$

Подставляя (1.52) в уравнение (1.50), получаем

$$\frac{Jbn_0}{q[(b+1) + bn_0/p(x)]^2}\frac{1}{p^2}\frac{dp}{dx} = -\frac{p - p_{n0}}{\tau_p}. \qquad (1.53)$$

При высоком уровне инжекции ($n \approx p \gg n_0 \gg p_{n0}$) в левой части этого уравнения можно пренебречь слагаемым bn_0/p по сравнению с $(b+1)$, а в правой части — пренебречь величиной p_{n0} по сравнению с p. Поскольку, в соответствии с формулой (1.35), при высоком уровне инжекции τ_p не зависит от p, то после разделения переменных уравнение непрерывности приводится к виду

$$dx = -\frac{J\tau_p bn_0}{q(b+1)^2}\frac{dp}{p^3}. \qquad (1.54)$$

---

[1]) Условия, в которых можно пренебречь током диффузии по сравнению с током дрейфа, легко реализуются в диодах с очень толстой базой ($W/L_a > > 20$). В таких диодах при достаточном удалении от p–n-перехода градиент концентрации инжектированных носителей становится очень малым и при высокой плотности тока носители в этой части диода переносятся за счет дрейфа.

Интегрирование этого уравнения от $x = x_n$ до $x = X_n$ дает уравнение, связывающее ток, концентрации инжектированных носителей в двух точках n-области и ее длину:

$$d = X_n - x_n = \frac{J\tau_p b n_0}{2q(b+1)^2}\left(\frac{1}{p^2(X_n)} - \frac{1}{p^2(x_n)}\right). \qquad (1.55)$$

Падение напряжения на n-области можно найти интегрированием уравнения (1.51), в котором мы также полагаем, что $n \approx p \gg n_0$:

$$U = U(X_n) - U(x_n) = \int_{x_n}^{X_n} \mathcal{E}\, dx \approx \frac{J}{q\mu_p(b+1)} \int_{x_n}^{X_n} \frac{dx}{p(x)}. \qquad (1.56)$$

Чтобы вычислить последний интеграл, сделаем замену переменных, подставив дифференциал (1.54) в подынтегральное выражение в (1.56) и перейдя тем самым от интегрирования по x к интегрированию по p. В результате получаем

$$U = \frac{J^2\tau_p b n_0}{3q^2\mu_p(b+1)^3}\left(\frac{1}{p^3(X_n)} - \frac{1}{p^3(x_n)}\right). \qquad (1.57)$$

Если скорость поверхностной рекомбинации на контакте и пропускаемый ток таковы, что выполняется соотношение $p(x_n) \gg \gg p(X_n) \gg n_0$, то в уравнениях (1.55) и (1.57) можно пренебречь слагаемыми, содержащими $p(x_n)$, и после несложных преобразований выразить плотность тока через падение напряжения на n-области:

$$J = \frac{9}{8}\,\frac{q\mu_p^2\tau_p b n_0}{d^3}\,U^2. \qquad (1.58)$$

Полученная квадратичная зависимость тока от напряжения называется *законом Мотта* и часто наблюдается в высокоомных полупроводниках при высокой плотности протекающего тока. [1)]

[1)] Следует заметить, что квадратичная зависимость тока от напряжения может иметь различную природу. В рассмотренном нами случае она возникала из-за особенностей пространственного распределения носителей, рекомбинирующих в процессе дрейфа в сильном электрическом поле. Точно такая же зависимость, отличающаяся лишь заменой времени жизни на максвелловское время релаксации τ_M, получается в случае так называемых *токов, ограниченных пространственным зарядом* [1, 10], которые возникают, когда время пролета носителей через образец становится сравнимым с τ_M и распределение электрического поля в образце становится неоднородным. В последнем случае рекомбинация носителей не играет особой роли.

Наконец, при сверхвысоких плотностях тока, когда величина тока ограничена скоростью рекомбинации на контакте, база диода практически однородно «заливается» инжектированными носителями и вольт-амперная характеристика полупроводниковой части структуры на этом участке принимает вид $J \sim U$.

Изучение силовых выпрямительных диодов и тиристоров показывает, что сделанные выше предположения о неизменности μ и τ не выполняются при очень высоком уровне инжекции. При концентрациях $p \approx n > 3 \cdot 10^{16}$ см$^{-3}$ становится существенным *взаимное рассеяние электронов и дырок*, отношение их подвижностей стремится к 1, а сами подвижности изменяются приблизительно как $\mu_p \sim p^{-1/2}$ [11]. Так, в p–i–n-диодах из Ge подвижность носителей при плотности тока 100 А/см^2 падает почти вдвое [12]. Понятно, что усиление рассеяния может изменить рассчитанные выше вольт-амперные характеристики приборов. При еще более высокой концентрации инжектированных носителей (>10^{18} см$^{-3}$) большую роль начинает играть Оже-рекомбинация, которая сильно уменьшает их время жизни.

Таким образом, проведенный нами анализ показывает, что в реальных диодах прямые ветви вольт-амперных характеристик могут иметь до четырех участков с качественно различным поведением (см. рис. 1.8). При низком напряжении смещения V (участок *1*) ток через p–n-переход определяется рекомбинацией в области пространственного заряда. При увеличении V (участок *2*) ток начинает определяться инжекцией неосновных носителей. При достижении условия высокого уровня инжекции (участок *3*) ток продолжает определяться инжекцией, но наклон вольт-амперной характеристики в полулогарифмическом масштабе уменьшается в m раз, где m определяется формулой (1.49). И наконец, на участке *4* (очень высокий уровень инжекции) зависимость тока от напряжения изменяется от экспоненциальной на степенную (закон Мотта) и далее линейную (закон Ома).

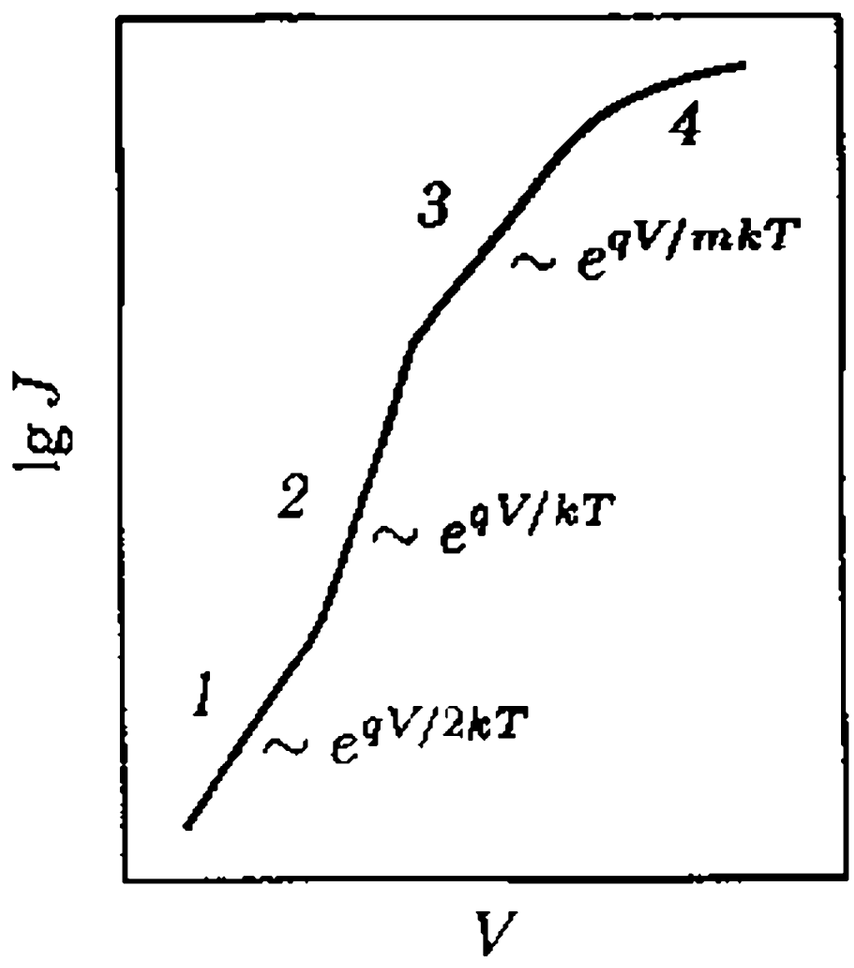


Рис. 1.8. Качественный вид прямой ветви вольт-амперной характеристики диода

1.2.4. Вольт-амперная характеристика p–i–n-диода. По сравнению с p–n-переходом, p–i–n-диод представляет собой несколько более сложную структуру (см. рис. 1.9), в которой

области p- и n-типа проводимости разделены высокоомной областью, проводимость которой близка к собственной (i) [13]. Поскольку технологически создать нелегированную область с собственной проводимостью очень трудно, i-слой обычно представляет собой слабо легированный полупроводник p- или n-типа проводимости; такие слои принято обозначать буквами π и ν, соответственно.

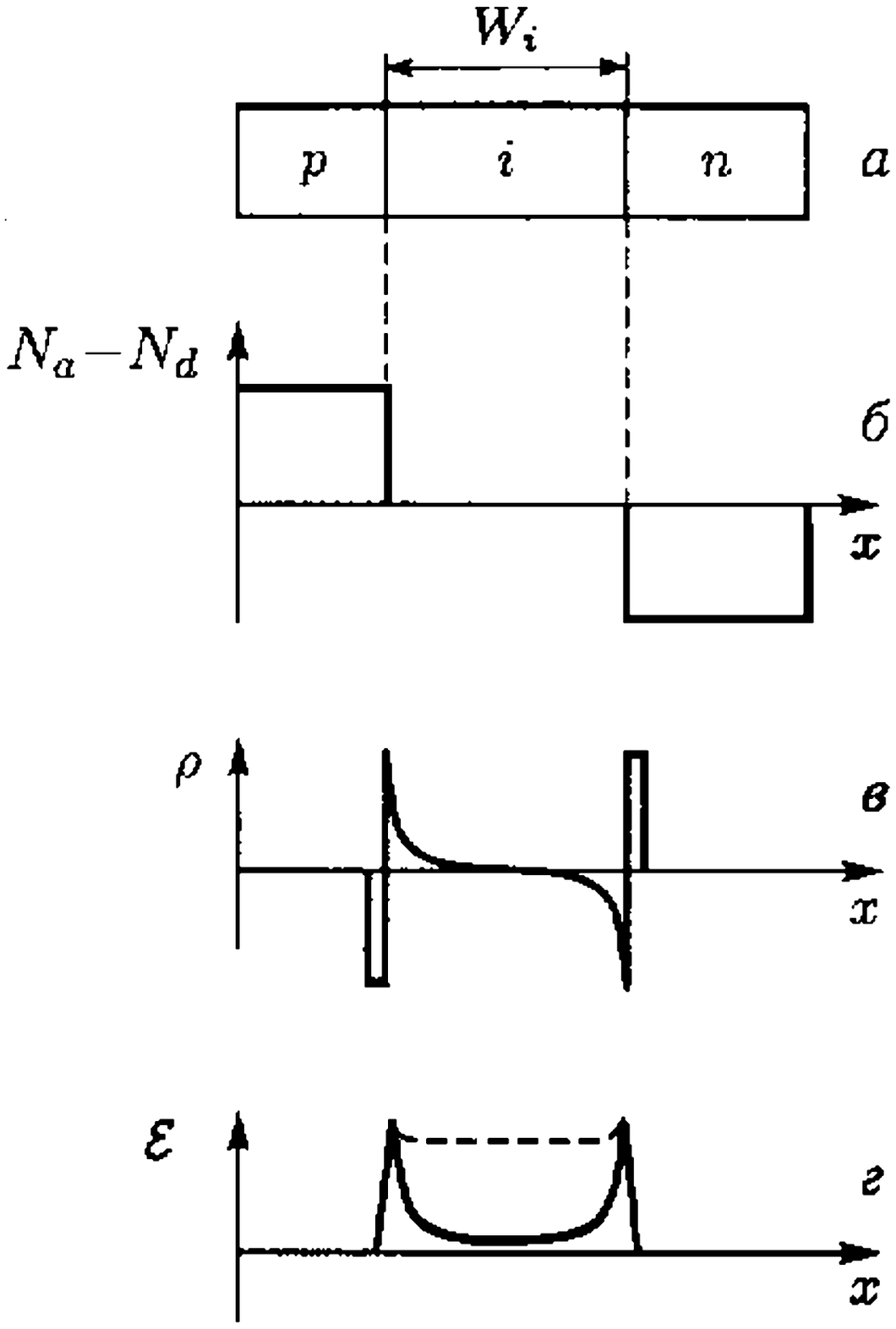


Рис. 1.9. Устройство p–i–n-диода ($а$), его профиль легирования ($б$), распределение плотности заряда ($в$) и электрического поля ($г$) в структуре при нулевом напряжении смещения. Пунктиром показано распределение электрического поля в p–i–n-диоде при обратном смещении

Особенностью энергетической диаграммы p–i–n-диода является наличие в ней двух энергетических барьеров, образующихся на границах i-слоя с сильно легированными областями. Решение уравнения Пуассона показывает, что практически все электрическое поле в структуре сосредоточено в i-области, причем поскольку в этой области концентрация примесей мала, то экранирование электрического поля в ней осуществляется не заряженными примесями, а *подвижными носителями заряда* (электронами и дырками), возникающими в результате тепловой генерации. По этой причине энергетическая диаграмма p–i–n-структуры при нулевом смещении зависит от толщины i-слоя (W_i). Если этот слой достаточно тонок (меньше дебаевской длины экранирования в собственном полупроводнике [1]), то электрическое поле в i-области не спадает до нуля (этот случай показан сплошной линией на рис. 1.9 $г$), а в случае толстого i-слоя — спадает до нуля. При подаче на p–i–n-диод обратного смещения $|V| \gg kT/q$ ситуация меняется, поскольку концентрации свободных носителей в i-области становятся исчезаю-

[1]) Для примера, дебаевская длина экранирования в собственном кремнии равна 24 мкм при 300 К [14].

ще малыми. При этом распределение электрического поля в i-слое становится практически однородным (пунктир на рис. 1.9 *г*), а его напряженность равна $\mathcal{E} = (\phi_k - V)/W_i$, где ϕ_k — контактная разность потенциалов между p- и n-областями p–i–n-диода.

p–i–n-структуры находят практическое применение при изготовлении высоковольтных диодов, модуляторов и аттенюаторов СВЧ излучения, лавинных фотодиодов. Преимущество от использования p–i–n-структур для создания высоковольтных диодов состоит в том, что при заданных геометрических размерах (толщине) кристалла в структурах с близким к однородному распределением поля получаются наибольшие напряжения пробоя. Применение p–i–n-диодов для модуляции СВЧ излучения основано на изменении дифференциального сопротивления диода, которым легко управлять изменяя величину протекающего через диод электрического тока. Примерами отечественных p–i–n-диодов могут служить выпрямительный диод КД529 (2 кВ и 400 А в импульсном режиме) и переключательные СВЧ диоды КА509 и КА520.

Для p–i–n-диодов с тонкой базой, то есть структур, в которых толщина i-слоя много меньше длины амбиполярной диффузии [1]) ($W_i \ll L_a$), вольт-амперная характеристика может быть найдена из следующих простых соображений. Наличие двух барьеров в структуре приводит к тому, что когда напряжение прямого смещения невелико, то инжектированные в i-слой носители обоих знаков оказываются как бы «запертыми» в этом слое (второй барьер препятствует выходу носителей из i-слоя в сильно легированные области). В этом случае ток через структуру можно считать равным току рекомбинации в i-области. Далее, поскольку толщина i-слоя мала по сравнению с диффузионной длиной, то распределение инжектированных носителей в i-слое можно считать однородным. Тогда, по аналогии с подходом, использованным нами при расчете тока рекомбинации в области пространственного заряда p–n-перехода в п. 1.2.2, для уровня рекомбинации, расположенного посередине запрещенной зоны ($n_t = p_t = n_i$) получаем

$$J_{\text{рек}} = q \int_0^{W_i} \frac{pn - n_i^2}{\tau_p(n + n_i) + \tau_n(p + p_i)}\, dx = \frac{qW_i}{\tau_i}(n - n_i), \qquad (1.59)$$

[1]) Определение амбиполярной длины диффузии см. на с. 36.

где $\tau_i = \tau_p + \tau_n$ — время жизни в собственном полупроводнике, а W_i — толщина i-слоя. При вычислении интеграла в (1.59) мы учли, что концентрации носителей в i-слое удовлетворяют соотношению $n \approx p$ (условие электронейтральности).

Если теперь обозначить напряжения смещения на p–i- и i–n-переходах через V_1 и V_2 и выразить концентрации инжектированных носителей через эти напряжения,

$$n = n_i \exp\left(\frac{qV_1}{kT}\right), \quad p = n_i \exp\left(\frac{qV_2}{kT}\right), \tag{1.60}$$

а затем учесть, что $n = p$, то легко видеть, что $V_1 = V_2$. Выражая ток рекомбинации через суммарное падение напряжения на двух переходах $V = V_1 + V_2$, получаем

$$J_{\text{рек}} = \frac{qn_iW_i}{\tau_i}\left[\exp\left(\frac{qV}{2kT}\right) - 1\right]. \tag{1.61}$$

Это и есть вольт-амперная характеристика p–i–n-диода с тонким i-слоем. Сравнение полученной формулы с формулой (1.38) для рекомбинационного тока в модели Са–Нойса–Шокли показывает, что эти две формулы различаются лишь толщиной области, в которой происходит рекомбинация.

Полученная вольт-амперная характеристика не учитывает падения напряжения на i-области, которое может стать существенным при высокой плотности протекающего через диод тока. Поскольку при прямом смещении в p–i–n-диодах наблюдается сильная модуляция проводимости i-области (проводимость этой области определяется инжектированными в нее электронами и дырками), то понятно, что сопротивление i-слоя будет какой-то сложной функцией протекающего тока. Рассчитаем падение напряжения на i-области в p–i–n-диоде с тонкой базой.

В p–i–n-диоде с тонкой базой концентрация инжектированных носителей в i-слое примерно постоянна и поэтому диффузионной компонентой тока можно пренебречь. Полагая, что ток в i-области определяется только дрейфом носителей, его можно записать следующим образом:

$$J = q(\mu_n n + \mu_p p)\mathcal{E} = q(\mu_n + \mu_p)\, n\mathcal{E}. \tag{1.62}$$

Поскольку этот ток равен току рекомбинации, то

$$J = q(\mu_n + \mu_p)n\mathcal{E} = \frac{qW_i}{\tau_i}(n - n_i). \tag{1.63}$$

Из этого уравнения следует, что при $n \gg n_i$ напряженность электрического поля в i-области практически не зависит от протекающего тока, а падение напряжения на ней составляет

$$V_i = \mathcal{E}W_i = \frac{W_i^2}{(\mu_n + \mu_p)\tau_i} = \frac{kT}{q}\frac{2b}{(1+b)^2}\left(\frac{W_i}{L_a}\right)^2. \qquad (1.64)$$

При выводе последней формулы мы использовали соотношение Эйнштейна $D = (kT/q)\mu$, чтобы перейти от подвижностей к коэффициентам диффузии, выразили сумму $D_n + D_p$ через коэффициент амбиполярной диффузии D_a (1.25) и использовали соотношение $D_a\tau_i = L_a^2$, чтобы перейти к длине амбиполярной диффузии L_a.

Полученный результат показывает, что за счет эффекта модуляции проводимости сопротивление i-слоя в p–i–n-структуре изменяется обратно пропорционально протекающему через него току, то есть приближение постоянного последовательного сопротивления, о котором мы говорили на с. 33, к таким структурам совершенно неприменимо. [1])

Более последовательный расчет вольт-амперной характеристики p–i–n-диода с произвольным соотношением W_i/L_a [5, 8] показывает, что решение (1.64) справедливо при $W_i/L_a < 2$. При этом падение напряжения на i-слое не превышает 0,05 В и им во многих случаях можно пренебречь. Однако для структур с более толстым i-слоем зависимость V_i от W_i переходит в экспоненциальную:

$$V_i \approx \frac{kT}{q}\frac{\pi\sqrt{b}}{1+b}\exp\left(\frac{W_i}{2L_a}\right). \qquad (1.65)$$

Этот случай характерен для высоковольтных p–i–n-структур с толстым i-слоем, в которых падение напряжения на i-слое может достигать нескольких вольт. Поскольку в высоковольтных приборах толщина i-слоя выбирается исходя из напряжения пробоя (см. п. 1.3) и не может быть сделана меньше, для уменьшения падения напряжения p–i–n-диоды следует изготавливать из материала с как можно бóльшим временем жизни (бóльшей диффузионной длиной L_a). То же самое относится и к выбору материалов для изготовления тиристоров (см. п. 3.1.1).

---

[1]) Заметим, что при очень высоком уровне инжекции, когда начинает проявляться взаимное рассеяние электронов и дырок, о котором мы говорили на с. 39, полученные выше выражения становятся еще более сложными.

1.3. Явление пробоя p–n-перехода

В этом разделе мы рассмотрим различные механизмы пробоя, возникающего в p–n-переходах и p–i–n-структурах при высоком обратном напряжении.

Упрощенно можно считать, что для каждого полупроводника или диэлектрика существует некоторая критическая напряженность электрического поля $\mathcal{E}_{\text{проб}}$, по достижении которой в нем возникает *электрический пробой* — резкое возрастание протекающего тока, который может приводить к необратимому изменению свойств материала. Для полупроводников характерные значения $\mathcal{E}_{\text{проб}}$ составляют ~100 кВ/см для Si и ~400 кВ/см для GaAs.

Как было показано в п. 1.1, в p–n-переходе всегда существует достаточно сильное электрическое поле, величина которого зависит от напряжения смещения. Так, в резком p–n-переходе максимальная напряженность электрического поля равна

$$\mathcal{E}_{\max} = \frac{2(\phi_k - V)}{W(V)} = \sqrt{\frac{8\pi q}{\varepsilon}\frac{N_a N_d}{N_d + N_a}(\phi_k - V)}\,. \qquad (1.66)$$

Приравнивая $\mathcal{E}_{\max} = \mathcal{E}_{\text{проб}}$ и пренебрегая величиной ϕ_k, которая обычно много меньше $V_{\text{проб}}$, из этой формулы нетрудно оценить напряжение пробоя p–n-перехода. Для асимметрично легированного p^+–n-перехода имеем:

$$V_{\text{проб}} \approx \frac{\varepsilon}{8\pi q N_d}\mathcal{E}^2_{\text{проб}}, \qquad (1.67)$$

то есть если $\mathcal{E}_{\text{проб}}$ не зависит от N_d, то напряжение пробоя меняется обратно пропорционально концентрации легирующей примеси в базе диода. Отсюда можно сделать качественный вывод, что для создания высоковольтных приборов уровень легирования базы диода должен быть достаточно низким.

Из материала предыдущих разделов следует, что в резком, плавном p–n-переходах и p–i–n-структуре максимальная напряженность электрического поля составляет, соответственно, $2(\phi_k - V)/W$, $1{,}5(\phi_k - V)/W$ и $(\phi_k - V)/W$, и поэтому наибольшими напряжениями пробоя при заданной толщине области пространственного заряда W (которая в конечном счете определяется толщиной базы диода) характеризуются p–i–n-структуры.

Сравнение зависимости (1.67) с экспериментом (рис. 1.10) показывает, что на опыте зависимость $V_{\text{проб}}$ от N_d оказывается

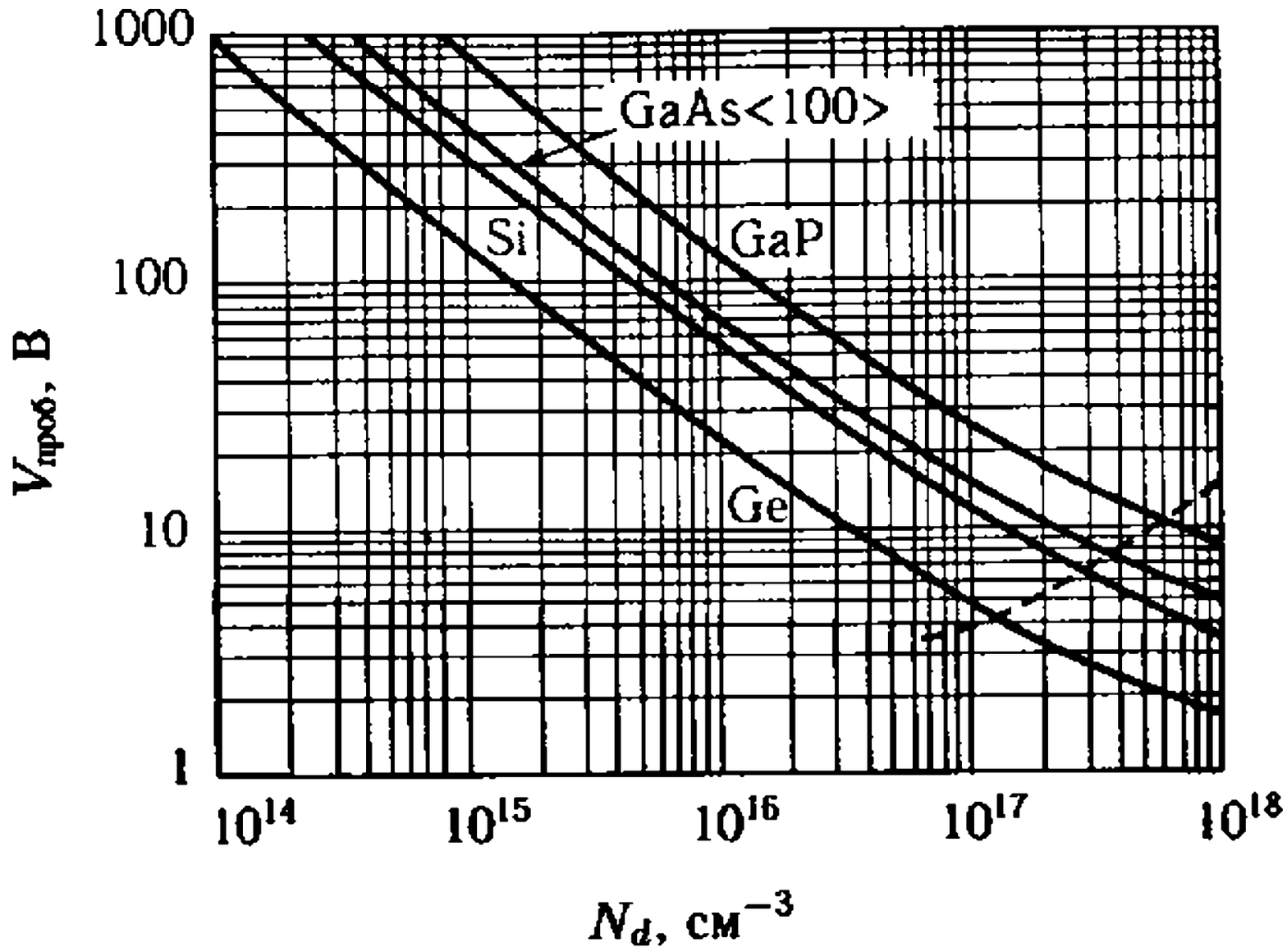


Рис. 1.10. Зависимость напряжения пробоя асимметрично легированных p^+—n-переходов из Ge, Si, GaAs и GaP с ориентацией <100> от уровня легирования n-области диода [14]. Пунктирной линией отмечено начало области туннельного пробоя

несколько более слабой ($\sim N_d^{-3/4}$). Кроме того, эксперимент показывает, что $V_{проб}$ зависит также и от ориентации p–n-перехода относительно кристаллографических осей, а при концентрации примесей выше $\sim 2 \cdot 10^{17}$ см$^{-3}$ вольт-амперные характеристики в области пробоя становятся более плавными, то есть пробой приобретает качественно иной характер. Это означает, что описанные выше представления о пробое являются чисто качественными, и нам следует более детально разобраться в физических процессах, протекающих в полупроводниковых структурах при пробое. Знание этих процессов важно не только для проектирования высоковольтных полупроводниковых приборов, но и имеет принципиальное значение для понимания принципов работы лавинно-пролетных диодов (п. 6.2) и лавинных фотодиодов (п. 7.1.5), в которых активно используется явление пробоя.

1.3.1. Лавинный пробой p–n-перехода. Одной из причин, приводящих к пробою в полупроводниках, является *ударная ионизация*. Физика этого явления по сути достаточно проста: носители, ускоряемые сильным электрическим полем, могут набрать столь большую энергию, что начнут вызывать ионизацию полупроводника (рождение электронно-дырочных пар). [1])

[1]) При низких температурах в полупроводниках также возможно появление *примесного пробоя*, при котором разогретые электрическим полем носители

Поскольку рождающиеся при ионизации электроны и дырки сами начинают ускоряться и рождать другие электронно-дырочные пары, то процесс размножения электронов и дырок в сильном электрическом поле приобретает характер лавины, откуда и происходит название *лавинного пробоя*.

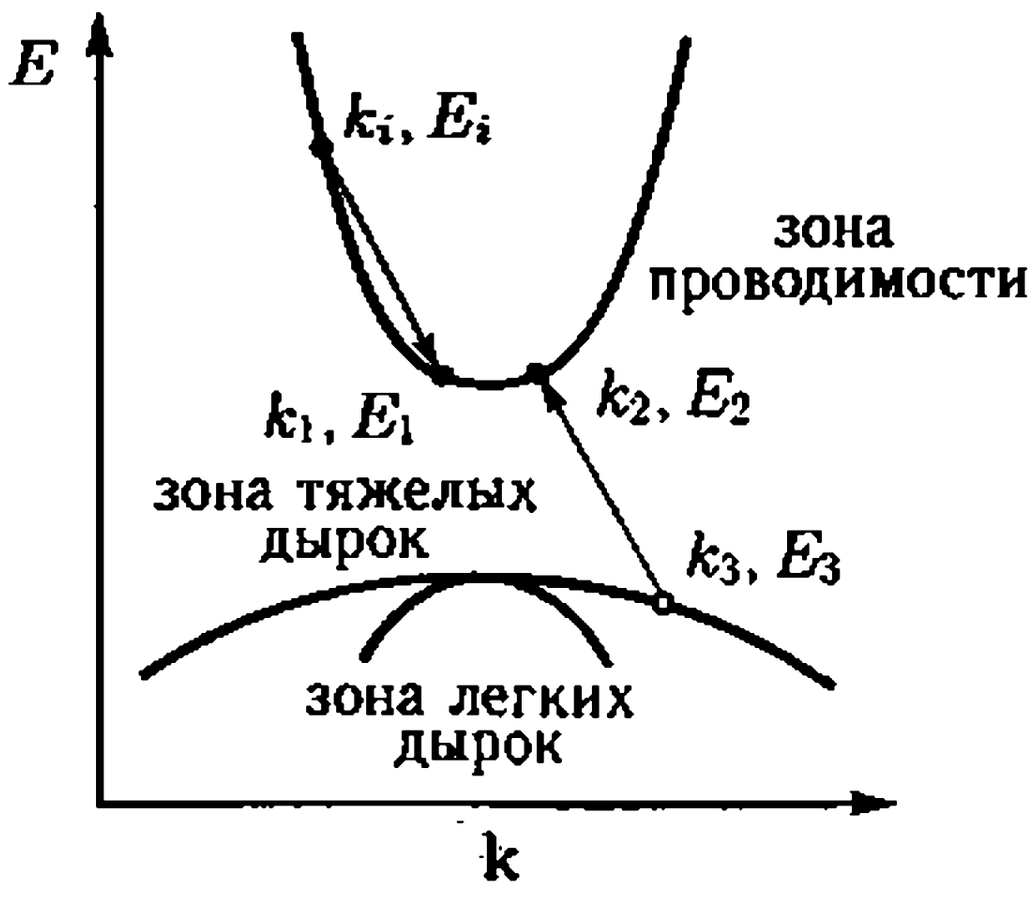


Рис. 1.11. Энергетическая диаграмма, поясняющая выполнение законов сохранения энергии и квазиимпульса при ударной ионизации. Электрон и дырка (частицы с индексами 2 и 3) возникают в результате перехода горячего электрона из состояния с индексом i в состояние с индексом 1

Как и в любом кристалле, в полупроводнике при рождении электронно-дырочной пары в результате ударной ионизации должны выполняться законы сохранения энергии и квазиимпульса ($E_i - E_1 = E_2 - E_3$, $\mathbf{k}_i - \mathbf{k}_1 = \mathbf{k}_2 - \mathbf{k}_3$, см. рис. 1.11). Поэтому скорость ударной ионизации зависит от зонной структуры полупроводника, то есть взаимного расположения ветвей $E(\mathbf{k})$ в области энергий на несколько эВ выше края зоны проводимости и ниже края валентной зоны. Так как в результате акта ударной ионизации в кристалле возникают электрон в зоне проводимости и дырка в валентной зоне, то, во-первых, ясно, что кинетическая энергия первичной (ионизующей) частицы всегда должна превышать E_g. Во-вторых, поскольку зависимости $E(\mathbf{k})$ для электронов и дырок различаются между собой и зависят от направления в кристалле, то скорость ударной ионизации оказывается различной для электронов и дырок и анизотропной. Например, в GaAs пороговая энергия электронов, выше которой они начинают ионизировать полупроводник, составляет $E_{\text{пор}} = 2,05$ эВ при движении вдоль оси <100>, 2,01 эВ при движении вдоль оси <110>, а электроны, движущиеся вдоль направления <111>, вообще не могут вызывать ударную ионизацию из-за близкого расположения экстремумов зоны проводимости в точках L и Γ (см. рис. 4.1). Для дырок

вызывают ионизацию мелких донорных или акцепторных уровней и тем самым резко увеличивают концентрацию свободных носителей, участвующих в проводимости. Примесный пробой в Ge при гелиевой температуре наблюдается уже в полях $\sim$10 В/см [1].

пороговые энергии равны 1,81 эВ при их движении вдоль оси <100> и 1,58 эВ при их движении вдоль направлений <110> и <111> [15]. Все эти энергии больше ширины запрещенной зоны GaAs, которая составляет 1,42 эВ. В кремнии $E_{\text{пор}} \approx$ $\approx$ 1,6 эВ [16].

Для количественного описания процесса ударной ионизации электронами и дырками вводят понятие *коэффициентов ударной ионизации* $\alpha(\mathcal{E})$, которые определяются как вероятность создания соответствующим носителем электронно-дырочной пары при прохождении единицы длины в электрическом поле с напряженностью $\mathcal{E}$:

$$\alpha_n = \frac{1}{n}\frac{dn}{dx}, \quad \alpha_p = \frac{1}{p}\frac{dp}{dx}. \tag{1.68}$$

Согласно модели Шокли [17], коэффициент ударной ионизации можно считать пропорциональным доле носителей, энергия которых превышает $E_{\text{пор}}$. Очевидно, что лишь небольшая доля ускоряемых сильным электрическим полем электронов и дырок способна набрать энергию, достаточную для ионизации полупроводника, поскольку в рассеянии *горячих носителей* (носителей, энергия которых заметно превышает тепловую) преобладает рассеяние с испусканием оптических фононов, при котором носители быстро теряют свою энергию. Предположим, что длина свободного пробега горячего носителя λ слабо зависит от его энергии. Тогда долю носителей, энергия которых превышает $E_{\text{пор}}$, можно оценить, зная λ и расстояние l, которое носитель должен пройти без столкновений, чтобы набрать эту энергию ($l = E_{\text{пор}}/q\mathcal{E}$):

$$w \sim \exp\left(-\frac{l}{\lambda}\right) \sim \exp\left(-\frac{E_{\text{пор}}}{q\mathcal{E}\lambda}\right). \tag{1.69}$$

Зависимость коэффициентов ударной ионизации электронов и дырок от $1/\mathcal{E}$ для ряда полупроводников показана на рис. 1.12. Как следует из рис. 1.12, проведенные выше рассуждения и формула (1.69) качественно правильно описывают наблюдаемые закономерности. Характерное значение λ для горячих носителей в полупроводнике составляет $\sim$100 Å, тогда из формулы (1.69) следует, что при $E_{\text{пор}} \sim 1$ эВ для наблюдения лавинного умножения требуются поля порядка 10^5 В/см. Более последовательная теория, основанная на решении уравнения Больцмана для горячих носителей, была развита Бараффом [18].

Выведем теперь уравнения, описывающие изменение плотностей токов электронов и дырок в области сильного поля в условиях ударной ионизации [19]. Рассмотрим p–n-переход,

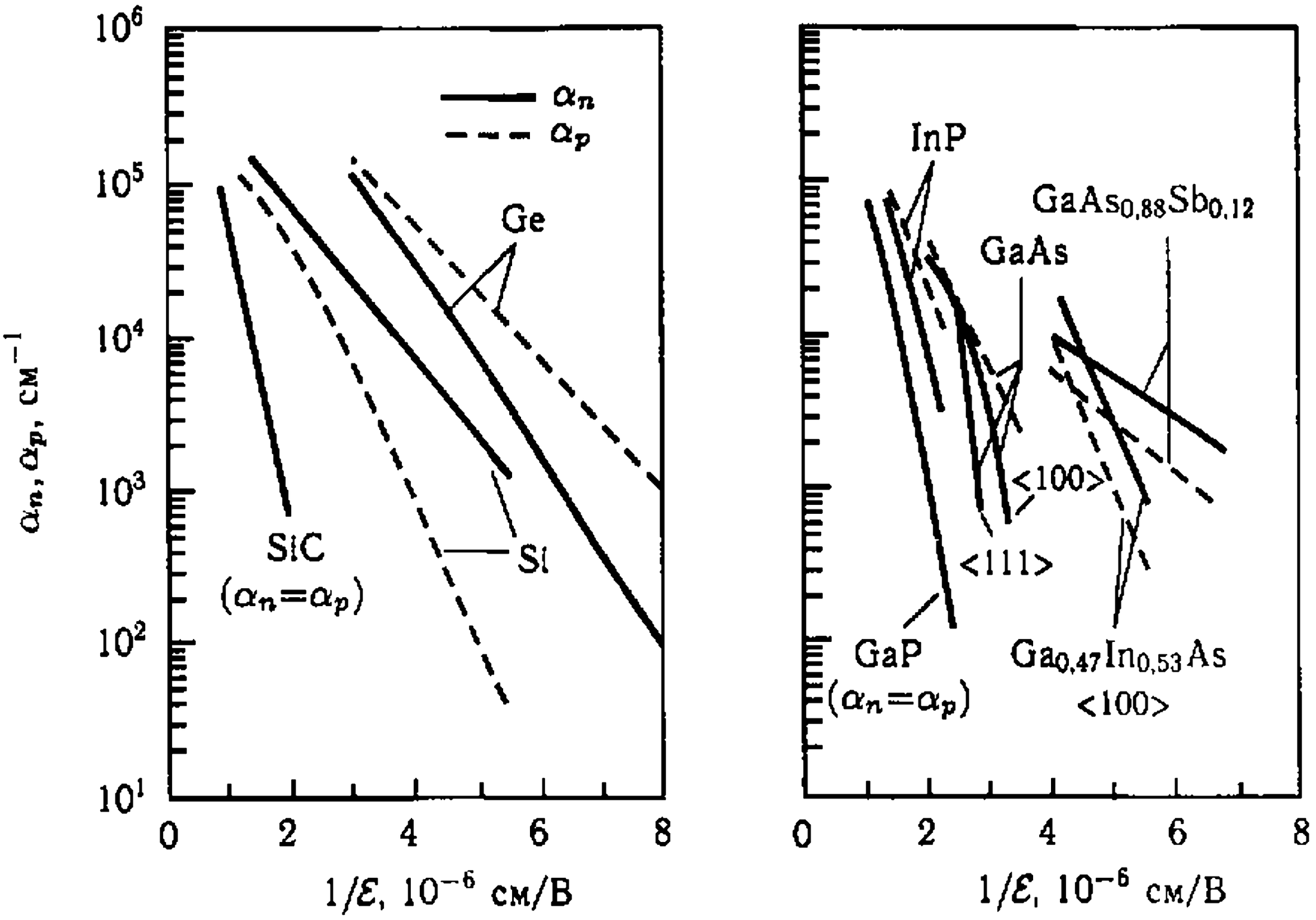


Рис. 1.12. Коэффициенты ударной ионизации электронов и дырок в некоторых полупроводниках при $T = 300$ К [14]

смещенный в обратном направлении, в который из нейтральных областей p- и n-типа втекают токи насыщения неосновных носителей J_{sn} и J_{sp} (см. рис. 1.13). Будем считать, что коэффициенты ударной ионизации зависят только от локальной напряженности электрического поля, а генерацией и рекомбинацией электронно-дырочных пар в области сильного поля можно пренебречь. [1]) Выделим в области сильного поля вблизи точки с координатой x небольшой участок шириной dx (см. рисунок). Поскольку рождение электронов происходит при ионизации полупроводника как горячими электронами, так и горячими дырками, то изменение концентрации электронов и связанное с ним изменение плотности тока электронов J_n равно

$$\frac{dJ_n}{dx} = \alpha_n J_n + \alpha_p J_p, \tag{1.70}$$

где α_n и α_p — коэффициенты ударной ионизации электронов и дырок. Соответствующее уравнение для изменения плотности

[1]) Рекомбинация носителей в области сильного поля действительно несущественна, поскольку характерные времена пролета носителями этой области обычно на несколько порядков меньше времени жизни.

тока дырок J_p может быть получено из (1.70) и условия, что полная плотность тока $J = J_n + J_p$ остается постоянной по всей длине образца.

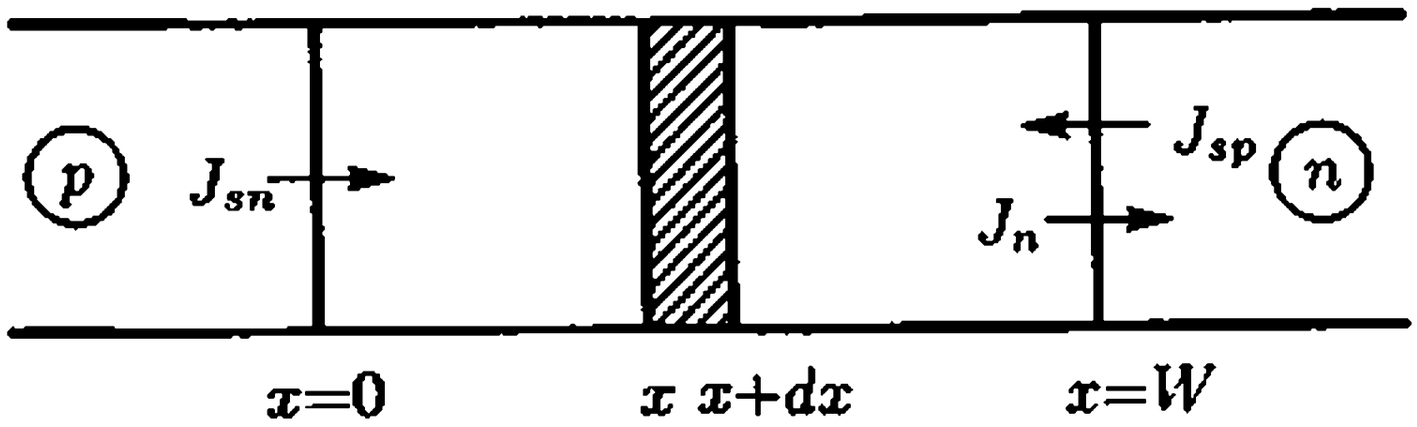


Рис. 1.13. Рисунок, иллюстрирующий решение уравнения (1.70)

Перепишем уравнение (1.70), выразив плотность тока дырок как $J_p = J - J_n$ и перенеся искомую функцию $J_n(x)$ в левую часть:

$$\frac{dJ_n}{dx} - (\alpha_n - \alpha_p) J_n = \alpha_p J. \tag{1.71}$$

Полученное уравнение является линейным дифференциальным уравнением относительно J_n с коэффициентами, зависящими только от x (через зависимость $\mathcal{E} = \mathcal{E}(x)$). Полное решение этого уравнения имеет вид

$$J_n(x) = e^{-F(x)} \left(J_n(0) + J \int_0^x \alpha_p(x') e^{F(x')}\, dx' \right),$$
$$F(x) = -\int_0^x [\alpha_n(x') - \alpha_p(x')]\, dx', \tag{1.72}$$

где $J_n(0) = J_{sn}$ — граничное условие при $x = 0$.

Полученное уравнение слишком сложно для анализа, однако качественное представление о характере распределения токов в области сильного поля мы можем получить из решения этого уравнения в случае $\alpha_n = \alpha_p$. В этом случае функция $F(x) \equiv 0$ и уравнение существенно упрощается:

$$J_n(x) = J_n(0) + J \int_0^x \alpha_n(x')\, dx'. \tag{1.73}$$

Рассмотрим асимметрично легированный p–n^+-переход, для которого выполняется условие $J_{sn} \gg J_{sp}$. Это означает, что в область сильного поля в основном затягиваются неосновные

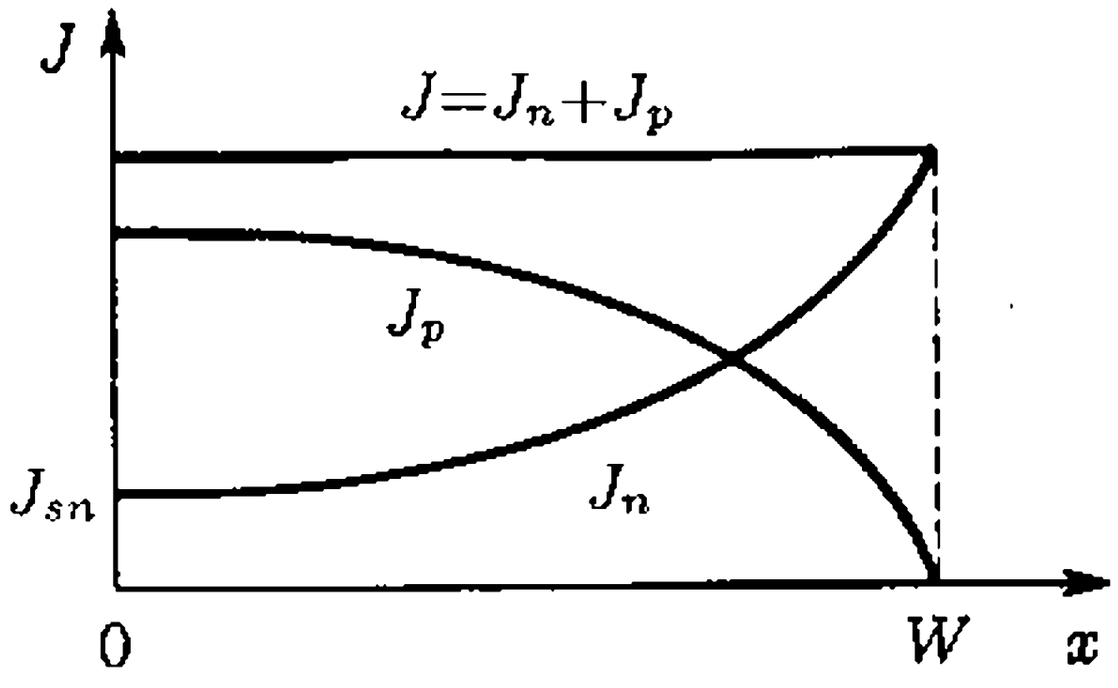


Рис. 1.14. Распределение плотностей токов электронов и дырок в области сильного поля p—n^+-перехода при лавинном пробое

носители (электроны) из p-области, а вклад дырок, приходящих из n-области, пренебрежимо мал. Поэтому граничные условия в точке $x = W$ можно записать так:

$$J_p(W) = 0, \quad J_n(W) = J.$$

Распределение плотностей токов электронов и дырок в области сильного поля показано на рис. 1.14. Максимальная скорость нарастания плотностей тока соответствует максимуму напряженности электрического поля в p—n^+-переходе ($x \approx W$); электроны движутся в электрическом поле слева направо, а дырки — справа налево.

Введем понятие *коэффициента умножения* M, который определим как отношение тока носителей на выходе области сильного поля к току носителей на входе этой области. Для рассматриваемого нами случая умножения электронов $M_n = J_n(W)/J_n(0)$. С учетом того, что при $x = W$ выполняется условие $J_n(W) = J$, из уравнения (1.73) следует уравнение [1])

$$1 - \frac{1}{M_n} = \int_0^W \alpha_n(x)\, dx. \qquad (1.74)$$

Аналогичное выражение для коэффициента умножения дырок M_p может быть получено для случая, когда в обратном токе p–n-перехода преобладает ток дырок.

Поскольку возникновению пробоя соответствует бесконечно сильное возрастание тока электронов при прохождении ими области сильного поля ($M_n \to \infty$), из уравнения (1.74) следует, что условием пробоя является достижение интегралом в правой части (1.74) значения единицы. Зная зависимость $\alpha_n(\mathcal{E})$ и распределение электрического поля в p–n-переходе, нетрудно найти напряжение $V_{\text{проб}}$, при котором будет возникать лавинный

[1]) Впервые это уравнение было выведено МакКеем [20], который использовал его для анализа самых первых данных по исследованию ударной ионизации в кремнии. Формулы для коэффициентов умножения электронов и дырок в случае $\alpha_n \neq \alpha_p$ можно найти в работе [19].

пробой, и рассчитать величину коэффициента умножения в предпробойной области. Например, аппроксимируя зависимость $\alpha_n(\mathcal{E})$ степенной функцией ($\alpha_n(\mathcal{E}) \sim \mathcal{E}^r$) и вычисляя интеграл в правой части уравнения (1.74) для распределения электрического поля в резком p–n-переходе ($\mathcal{E}(x) = \mathcal{E}_{\max}(1 - x/W)$) с учетом ранее полученных выражений для $\mathcal{E}_{\max}$ и W, приходим к формуле

$$M = \frac{1}{1 - \left(\dfrac{\phi_k - V}{\phi_k - V_{\text{проб}}}\right)^m} \approx \frac{1}{1 - \left(\dfrac{V}{V_{\text{проб}}}\right)^m}, \quad m = \frac{r+1}{2}, \quad (1.75)$$

которая часто используется для аппроксимации зависимости коэффициента умножения от напряжения обратного смещения. Экспериментально найденные значения m составляют 4,7–6,6 для электронов и 3–3,4 для дырок в Ge [19] и 1,4–2 для электронов и 3,4–4 для дырок в Si [21]. Заметим, что отношение α_n/α_p может быть как больше единицы (Si), так и меньше единицы (Ge).

Важно иметь в виду, что для возникновения лавинного пробоя необходимо, чтобы *лавина успевала развиваться*, а для этого нужно, чтобы область сильного поля была достаточно протяженной (во всяком случае, намного больше, чем длина свободного пробега). Именно поэтому очень тонкие пленки оказываются электрически более прочными, чем объемные материалы. Из-за того, что с ростом концентрации примеси верхняя граница интегрирования в (1.74) уменьшается, для пробоя p–n-перехода требуются все более сильные электрические поля, и поэтому наблюдаемая зависимость $V_{\text{проб}}(N_d)$ (рис. 1.10) оказывается более слабой по сравнению с предсказаниями упрощенной теории (1.67).

Методы определения коэффициентов ударной ионизации. Определение коэффициентов ударной ионизации основано на обработке полученных в эксперименте зависимостей коэффициента умножения M в p–n-переходе от напряжения обратного смещения [20].

В основе этого метода лежит то, что коэффициент ударной ионизации α является функцией только напряженности электрического поля. Поскольку распределение электрического поля в p–n-переходе обычно имеет вид функции с максимумом, в формуле (1.74), полученной в предположении $\alpha_n = \alpha_p$, можно сделать

замену переменных и переписать ее следующим образом:

$$1-\frac{1}{M_n}=\int_0^W \alpha_n(x)\,dx=\int_0^{\mathcal{E}_{\max}} \alpha(\mathcal{E})\left[\left(\frac{d\mathcal{E}}{dx}\right)^{-1}+\left(\frac{d\mathcal{E}}{dx}\right)^{-1}\right]d\mathcal{E}, \quad (1.76)$$

где $\mathcal{E}_{\max}$ — максимальная напряженность электрического поля в p–n-переходе. Выражение в квадратных скобках описывает вклад в интеграл областей, расположенных слева и справа от максимума электрического поля, и для известного профиля легирования p–n-перехода рассчитывается с помощью уравнения Пуассона, $d\mathcal{E}/dx = 4\pi\rho/\varepsilon$. Дифференцируя левую и правую часть уравнения (1.76) по $\mathcal{E}_{\max}$, нетрудно найти зависимость $\alpha(\mathcal{E})$. Так, для резкого p^+–n-перехода выражение в квадратных скобках приблизительно равно $\varepsilon/(4\pi q N_d)$ и

$$\alpha(\mathcal{E}_{\max})=\frac{4\pi q N_d}{\varepsilon}\frac{d}{d\mathcal{E}_{\max}}\left(1-\frac{1}{M_n}\right)=$$

$$=\frac{4\pi q N_d}{\varepsilon}\left(\frac{d\mathcal{E}_{\max}}{dV}\right)^{-1}\frac{d}{dV}\left(1-\frac{1}{M_n}\right), \quad (1.77)$$

где

$$\mathcal{E}_{\max}=\sqrt{\frac{8\pi q N_d}{\varepsilon}(\phi_k-V)}, \qquad \frac{d\mathcal{E}_{\max}}{dV}=-\sqrt{\frac{2\pi q N_d}{\varepsilon}\frac{1}{\phi_k-V}}. \quad (1.78)$$

Эти формулы позволяют, обработав зависимость $M(V)$ и рассчитав для каждого значения V величину $\mathcal{E}_{\max}$, построить искомую зависимость $\alpha(\mathcal{E})$. Соответствующие формулы для вычисления $\alpha(\mathcal{E})$ в плавном p–n-переходе можно найти в работе [20].

К сожалению, анализ зависимости обратного тока p–n-перехода от напряжения не позволяет надежно определить коэффициент умножения из-за большого вклада в этот ток утечек по поверхности. Поэтому определение коэффициентов умножения обычно основывается на измерении тока неравновесных носителей, создаваемых различными внешними воздействиями: светом, бомбардировкой альфа-частицами, инжекцией носителей через точечный контакт или с помощью p–n-перехода. Для раздельного определения коэффициентов M_n и M_p такие измерения часто проводят на двух типах структур (p^+–n и n^+–p), а при обработке данных используют достаточно сложные формулы, следующие из уравнения (1.72) [19, 21].

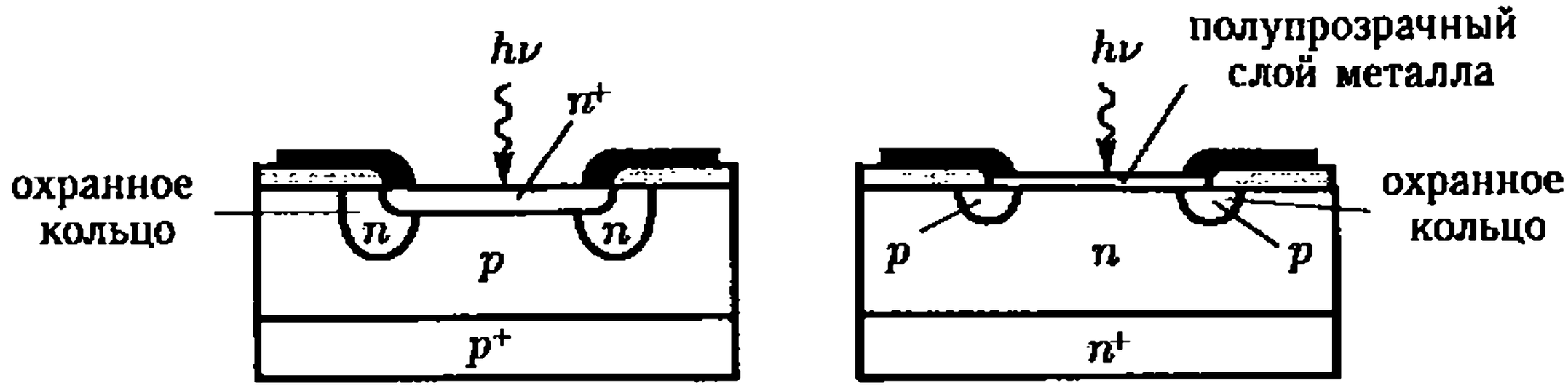
Один из вариантов методики, позволяющей раздельно определить коэффициенты умножения электронов и дырок на одном и том же образце, состоит в измерении фотоотклика диодной структуры с неглубоким p–n-переходом, освещаемой светом перпендикулярно p–n-переходу. Измерения проводят при двух энергиях квантов, для которых глубина проникновения света оказывается много больше и много меньше расстояния до p–n-перехода. При этом неравновесные носители возбуждаются либо в объеме структуры (области одного типа проводимости), либо в приповерхностном слое (области другого типа проводимости). Измеряя зависимость фототока от напряжения смещения для двух энергий квантов, на одном образце удается выполнить измерение коэффициентов умножения и для электронов, и для дырок.

Особенности явления пробоя в реальных p–n-переходах. До сих пор мы неявно полагали, что p- и n-области полупроводника однородны, а граница p–n-перехода — плоская. Однако на практике это часто оказывается не так.

В планарной технологии, используемой в настоящее время для создания большинства полупроводниковых приборов (см. п. 2.8.1), конструкция p–n-переходов предполагает появление в них закругленных (цилиндрических или сферических) участков перехода. Решение уравнения Пуассона показывает, что на этих участках напряженность электрического поля *заметно выше*, чем в плоской части p–n-перехода, и поэтому пробой будет происходить в первую очередь именно на закругленных участках. При радиусе кривизны, равном толщине области пространственного заряда, напряжение пробоя цилиндрического участка приблизительно на 35% ниже, а сферического — на 45% ниже, чем напряжение пробоя плоского p–n-перехода [14].

Поскольку при возникновении пробоя небольшая область p–n-перехода, в которой напряженность электрического поля наиболее высока, сильно разогревается, это чревато необратимым разрушением прибора. Поэтому в приборах, работающих в условиях лавинного пробоя (например, в стабилитронах (см. п. 1.3.4) и лавинных фотодиодах (см. п. 7.1.5)), создают так называемые *охранные кольца*. Эти кольца представляют собой области слабо легированного полупроводника, которые формируются в тех местах, где будут располагаться закругленные участки рабочего p–n-перехода (см. рис. 1.15). Из-за более низкой концентрации примесей в охранных кольцах толщина обедненного слоя около них больше, а напряженность электрического поля —

ниже, чем в плоской части рабочего p–n-перехода, и поэтому лавинное умножение происходит однородно по всей площади плоской части структуры.



а *б*

Рис. 1.15. Конструкции охранных колец в лавинных фотодиодах с p–n-переходом (*а*) и барьером Шоттки (*б*)

В мощных полупроводниковых приборах с плоскими p–n-переходами пробой может возникать и в районе выхода p–n-перехода на боковую поверхность. Дело в том, что присутствие границы раздела искажает распределение электрического поля в объеме полупроводника и в определенных случаях напряженность этого поля вблизи боковой поверхности может оказаться выше, чем в объеме. Чтобы уменьшить напряженность электрического поля вблизи поверхности, границу раздела полупроводник–воздух делают не перпендикулярно плоскости p–n-перехода, а под некоторым углом (создают так называемые *фаски*). [1]) Кроме того, возникновению поверхностного пробоя может способствовать адсорбция электрически активных молекул (например, молекул воды) на поверхности, что вызывает дополнительный поверхностный изгиб зон в полупроводнике. Поэтому поверхность полупроводника всегда защищают слоем диэлектрика (легкоплавким стеклом, полимером). При этом важно, чтобы защитный слой сам бы не имел встроенного заряда, который бы искажал распределение электрического поля вблизи поверхности. Так, из-за высокого встроенного заряда термически выращенные слои SiO_2 не годятся для защиты поверхности высоковольтных приборов из Si [11].

Другой причиной преждевременного пробоя, характерной для реальных приборов, часто оказывается неоднородность полупроводника. Давно замечено, что пробой в полупроводниковых приборах часто происходит вблизи существующих в кристаллах

[1]) Вопрос о выборе угла фаски мы рассмотрим в главе, посвященной тиристорам (на с. 228).

дефектов — дислокаций, пересекающих p–n-переход, включений второй фазы, а также «встроенных» в кристаллы неоднородностей распределения легирующих примесей. [1]) Такие локальные области пробоя получили название *микроплазм*. Рассматривая в микроскоп p–n-переходы, на которые подано обратное смещение, эти области можно наблюдать как яркие желтоватые светящиеся точки очень малого диаметра [22]. Ток, протекающий через микроплазму, носит характер импульсов амплитудой $\sim$50 мкА и длительностью $\sim$1 мкс, включающихся и выключающихся в случайные моменты времени [20]. Такое поведение было объяснено случайным характером захвата и выброса носителей заряда на уровни, связанные с дефектами (дислокациями, включениями). Изменение заряда дефекта меняет напряженность электрического поля в районе дефекта, которое становится то выше, то ниже порогового значения [17]. Очевидно, что существование микроплазм приводит к нестабильности характеристик приборов, работающих в режиме лавинного пробоя, и, в частности, является причиной сильного возрастания шумов стабилитронов.

1.3.2. Туннельный пробой p–n-перехода. Как известно из квантовой механики, волновая природа электрона проявляется в том, что при его движении в потенциальной яме существует конечная вероятность обнаружить электрон даже в области, в которой потенциальная энергия в яме $U(x)$ превышает его кинетическую энергию E и где по законам классической механики появление электрона невозможно. Если толщина барьера невелика, то при этом будет наблюдаться прохождение электронов сквозь потенциальный барьер — явление, получившее название *туннелирования*. В квазиклассическом приближении вероятность туннелирования описывается формулой

$$D \approx \exp\left(-2\int\limits_{x_1}^{x_2} |k|\,dx\right), \qquad (1.79)$$

где $|k|$ — модуль мнимого волнового вектора туннелирующего носителя, а x_1 и x_2 — координаты точек поворота, в которых $k = 0$ [23].

---

[1]) Дислокации имеют свойство окружать себя примесями, образуя так называемые *примесные атмосферы*. Поскольку локальная концентрация примеси в районе дислокации выше средней, то толщина обедненного слоя в этой области оказывается меньше, что и приводит к локальному уменьшению напряжения пробоя.

Представления о туннелировании привлекались для объяснения вольт-амперных характеристик диодов начиная с 30-х годов, однако первые убедительные аргументы в пользу проявления туннельного эффекта при пробое обратно смещенных p–n-переходов были получены только в 1957 г. [24].

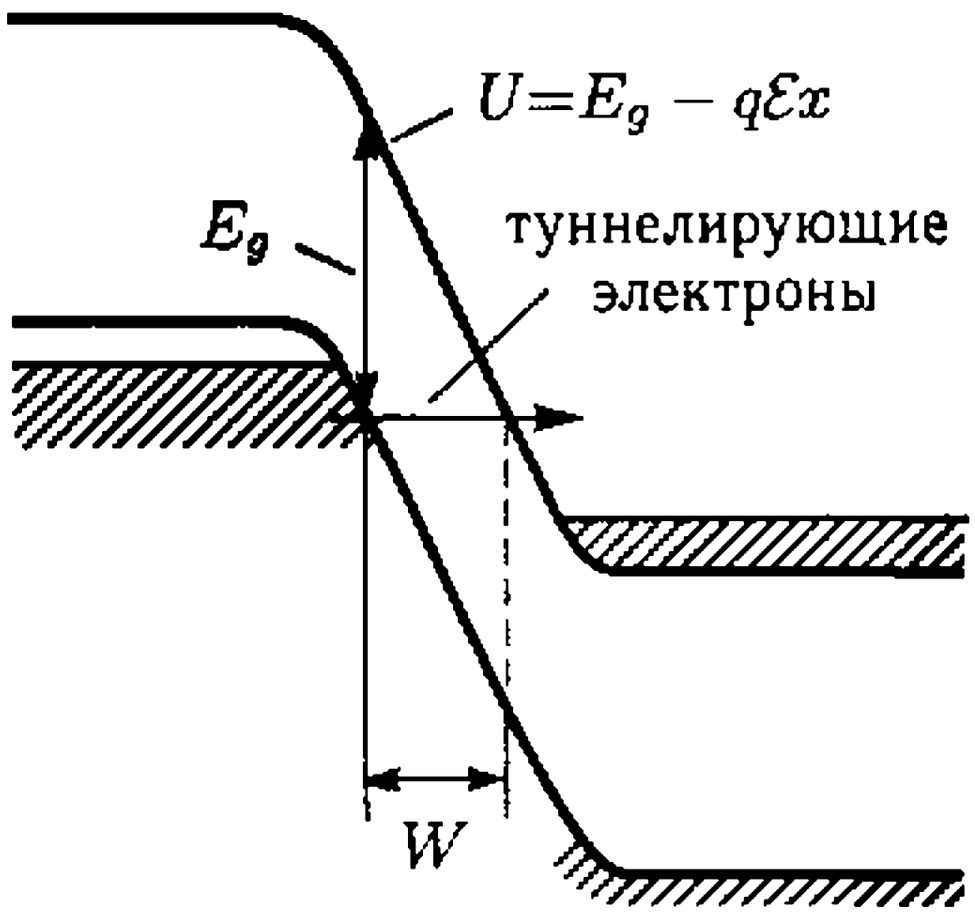


Рис. 1.16. Туннелирование в p–n-переходе при обратном смещении

Чтобы наблюдать явление туннельного пробоя в p–n-переходе, необходимо, чтобы область пространственного заряда была достаточно узкой, для чего области p- и n-типа проводимости должны быть сильно легированы. Ток в таком p–n-переходе определяется электронами, туннелирующими из заполненных состояний валентной зоны p-области на свободные состояния зоны проводимости n-области (стрелка на рис. 1.16).

Рассчитаем теперь величину D. Рассмотрим обратно смещенный p–n-переход и аппроксимируем потенциальный барьер, который преодолевает туннелирующий электрон, треугольником, высота которого равна E_g, а тангенс угла наклона гипотенузы равен средней напряженности электрического поля: $U(x) = E_g - q\mathcal{E}x$ (см. рис. 1.16). Для электронов, описываемых параболическим законом дисперсии с эффективной массой m^*, в области туннелирования

$$|k| = \sqrt{2m^*[U(x) - E]}\,/\hbar.$$

Подставляя эту величину в интеграл (1.79) и проводя интегрирование между точками поворота, получаем

$$D \approx \exp\left[-\frac{4\sqrt{2m^*}}{3\hbar}\frac{E_g^{3/2}}{q\mathcal{E}}\right] = \exp\left(-\frac{\mathcal{E}_0}{\mathcal{E}}\right), \qquad (1.80)$$

где через $\mathcal{E}_0$ обозначена напряженность характеристического поля, численное значение которой выражается через отношение эффективной массы к массе свободного электрона m^*/m_0 и ширину запрещенной зоны (в эВ) следующим образом:

$$\mathcal{E}_0 \approx 5{,}5 \cdot 10^7 \sqrt{m^*/m_0}\; E_g^{3/2} \ \text{В/см}. \qquad (1.81)$$

Плотность туннельного тока p–n-перехода связана с D соотношением

$$J = \frac{\sqrt{2m^*}\, q^3 \mathcal{E} V}{4\pi^2 \hbar^2 E_g^{1/2}} D, \qquad (1.82)$$

где V — приложенное к переходу обратное напряжение [14]. Из формул (1.80), (1.82) следует, что с увеличением обратного смещения, при котором напряженность электрического поля $\mathcal{E}$ возрастает, барьер становится все более прозрачным и это вызывает быстрое увеличение туннельного тока. Это и есть туннельный пробой p–n-перехода.

Для GaAs ($m_n^* \approx 0{,}067 m_0$, $E_g \approx 1{,}42$ эВ) величина $\mathcal{E}_0$ составляет $2{,}4 \cdot 10^7$ В/см. Отсюда следует, что туннельный механизм протекания тока через p–n-переход может стать существенным только при очень высокой напряженности электрического поля, порядка 10^6 В/см. Соответствующая этой напряженности поля толщина потенциального барьера составляет $W \sim 300$ Å.

Здесь может возникнуть вопрос: в каких условиях все-таки можно наблюдать туннельный пробой? Ведь конкурирующий с ним лавинный пробой обычно возникает при более низкой напряженности электрического поля ($\sim 10^5$ В·см$^{-1}$, см. с. 47). Ответ заключается в том, что в сильно легированных p–n-переходах из-за того, что размер области пространственного заряда сравним с длиной свободного пробега горячих носителей, условия для лавинного пробоя реализовать достаточно трудно. В этом случае лавина просто не успевает развиться и тогда на первое место выходит туннельный механизм пробоя. Уровни легирования, необходимые для реализации условий туннельного пробоя, лежат правее пунктирной линии на рис. 1.10; в кремниевых диодах туннельный механизм становится основным механизмом пробоя при концентрации примеси в базе диода $N_d > 3 \cdot 10^{17}$ см$^{-3}$.

1.3.3. Тепловой пробой p–n-перехода. Тепловой пробой не связан, строго говоря, с каким-либо новым механизмом пробоя. Этот термин используется, чтобы обозначить те существенные изменения в вольт-амперных характеристиках, которые возникают в результате выделения большого количества тепла в полупроводниковых структурах, работающих в режиме, близком к пробою.

Выделение тепла вызывает нагрев полупроводниковых приборов, при котором обратные токи p–n-переходов возрастают,

приводя к еще более сильному выделению тепла. Если при этом нарастание выделяющейся мощности превысит возможности отвода тепла от прибора, то разогрев прибора может стать неуправляемым и привести к разрушению прибора.

Тепловой пробой оказывается особенно существенным для p–n-переходов, изготовленных из полупроводников с не очень широкой запрещенной зоной (например, из Ge), поскольку для *таких переходов характерны большие обратные токи.*

Основными методами защиты p–n-перехода от теплового пробоя являются, с одной стороны, разработка такой топологии прибора, которая обеспечивает хороший теплоотвод и, с другой, работа в условиях, далеких от предельно допустимых. Следует добавить, что конструирование силовых полупроводниковых приборов предъявляет повышенные требования к *однородности полупроводника*, поскольку если из-за его неоднородности разогрев прибора начнется в одной точке, то наличие даже хорошего теплоотвода не спасет p–n-переход от локального перегрева и преждевременного выхода из строя. В силовых транзисторах проявление теплового пробоя известно под названием *вторичного пробоя* (см. подробнее п. 2.3).

1.3.4. Стабилитроны. Резкое возрастание обратного тока p–n-перехода в области пробоя (см. рис. 1.17 и 1.18) позволяет использовать эту особенность его вольт-амперной характеристики для создания *стабилитронов* — полупроводниковых приборов, напряжение на которых почти не зависит от протекающего через них тока. [1]) Эти приборы используются в радиоэлектронике в качестве источников стабильного напряжения [26]. Напряжение стабилизации $U_{ст}$, которое получают с их помощью, обычно лежит в пределах 3–200 В.

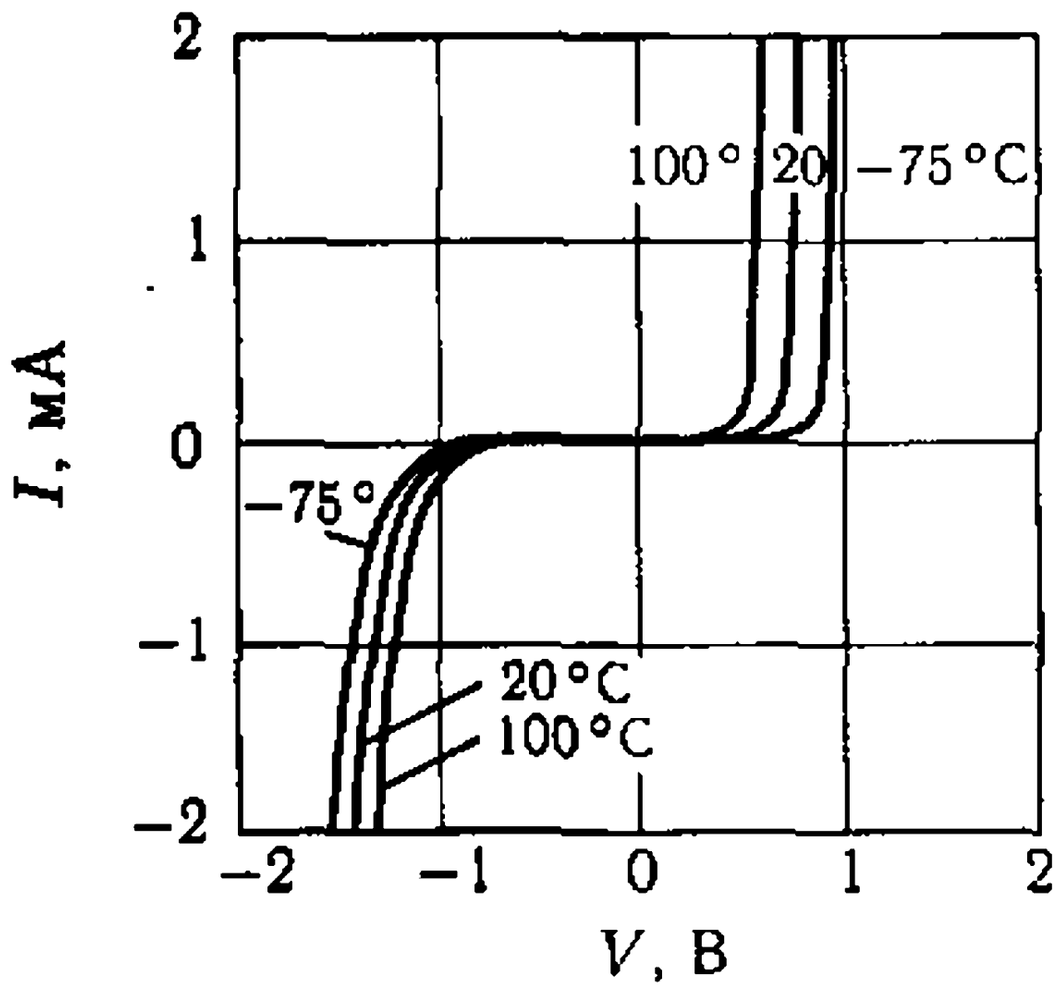


Рис. 1.17. Температурная зависимость вольт-амперных характеристик диодов с туннельным пробоем [12]

[1]) В зарубежной литературе стабилитроны также называют зенеровскими диодами в честь К. Зенера, первым предложившего теоретическую модель туннельного пробоя в диэлектриках [25].

Для практических применений важно знать, как изменение температуры окружающей среды и изменение тока через стабилитрон влияют на напряжение пробоя. Эти зависимости характеризуются такими двумя важнейшими параметрами стабилитрона как температурный коэффициент напряжения стабилизации (ТКН) и коэффициент качества.

Рассмотрим сначала факторы, определяющие температурную стабильность напряжения пробоя. Скорость относительного изменения напряжения на стабилитроне с температурой $(1/U_{ст})(dU_{ст}/dT)$, измеренная при фиксированном токе диода, называется *температурным коэффициентом напряжения*.

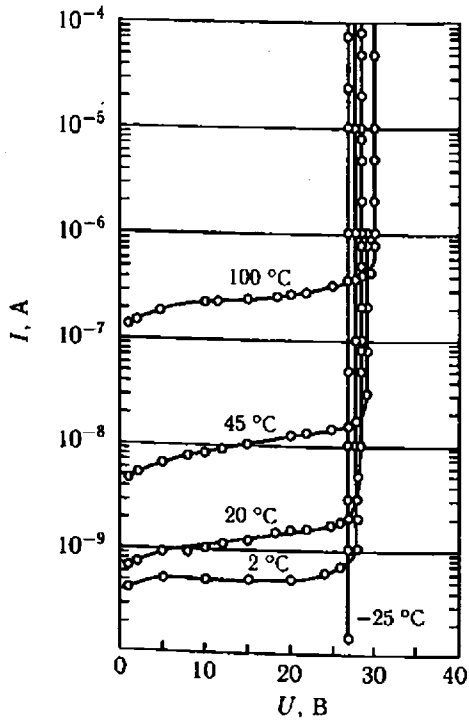


Рис. 1.18. Температурная зависимость в обратной ветви вольт-амперной характеристики кремниевого p–n^+-перехода с концентрацией примесей в базе $N_a = 2,5 \cdot 10^{16}$ см$^{-3}$ [14]

Температура оказывает заметное влияние на механизмы пробоя. Поскольку в большинстве полупроводников ширина запрещенной зоны уменьшается с ростом температуры, то характерный параметр в теории туннельного пробоя $\mathcal{E}_0$ (см.

формулу (1.81)) также уменьшается. Это означает, что с ростом температуры туннельный пробой должен начинаться при более низком обратном смещении, то есть для туннельного пробоя характерен *отрицательный* знак ТКН. Это действительно наблюдается в сильно легированных p–n-переходах, в которых $U_{ст}$ мало (рис. 1.17).

Для механизма ударной ионизации, напротив, характерен *положительный* ТКН (см. рис. 1.18). Увеличение напряжения пробоя с ростом температуры связано с усилением рассеяния носителей на оптических фононах, в результате чего входящая в формулу (1.69) длина свободного пробега уменьшается *по закону*

$$\lambda \sim \mathrm{th}(\hbar\omega_0/2kT),$$

где $\hbar\omega_0$ — *энергия оптического фонона* [14]. *Поэтому с ростом* температуры для достижения пороговой энергии необходимо все более сильное электрическое поле. По этой причине в кремниевых приборах с невысокой концентрацией примеси в базе диода, в которых наблюдается лавинный пробой, напряжение стабилизации увеличивается с ростом температуры с ТКН $\approx 0,1\,\%/\mathrm{K}$.

Разные знаки температурных коэффициентов напряжения стабилизации для туннельного механизма пробоя и ударной ионизации позволяют, подобрав уровень *легирования базы диода* и проварьировав другие технологические параметры, сделать так, что при некотором значении $U_{ст}$ увеличение туннельного тока с ростом температуры будет точно компенсироваться уменьшением тока лавинного пробоя. Тогда напряжение на таком стабилитроне практически не будет зависеть от температуры и он может быть использован как прецизионный источник напряжения («опорный диод»). Эксперимент показывает, что в разных полупроводниках напряжение стабилизации, при котором достигается такая компенсация, составляет $(4\text{–}6)E_g/q$. В кремниевых диодах условию ТКН $= 0$ соответствует $U_{ст} \approx 5,5$ В (см. рис. 1.19).

Другим способом компенсации температурного дрейфа $U_{ст}$ в *стабилитронах является* включение последовательно с обратно смещенным p–n-переходом одного или нескольких p–n-переходов, работающих при прямом смещении. Поскольку из-за уменьшения контактной разности потенциалов с ростом температуры падение напряжения на прямо смещенном кремниевом p–n-переходе уменьшается со скоростью $\approx$2,2 мВ/К, это позволяет скомпенсировать небольшой положительный ТКН обратно смещенного p–n-перехода. Изготовленные таким образом

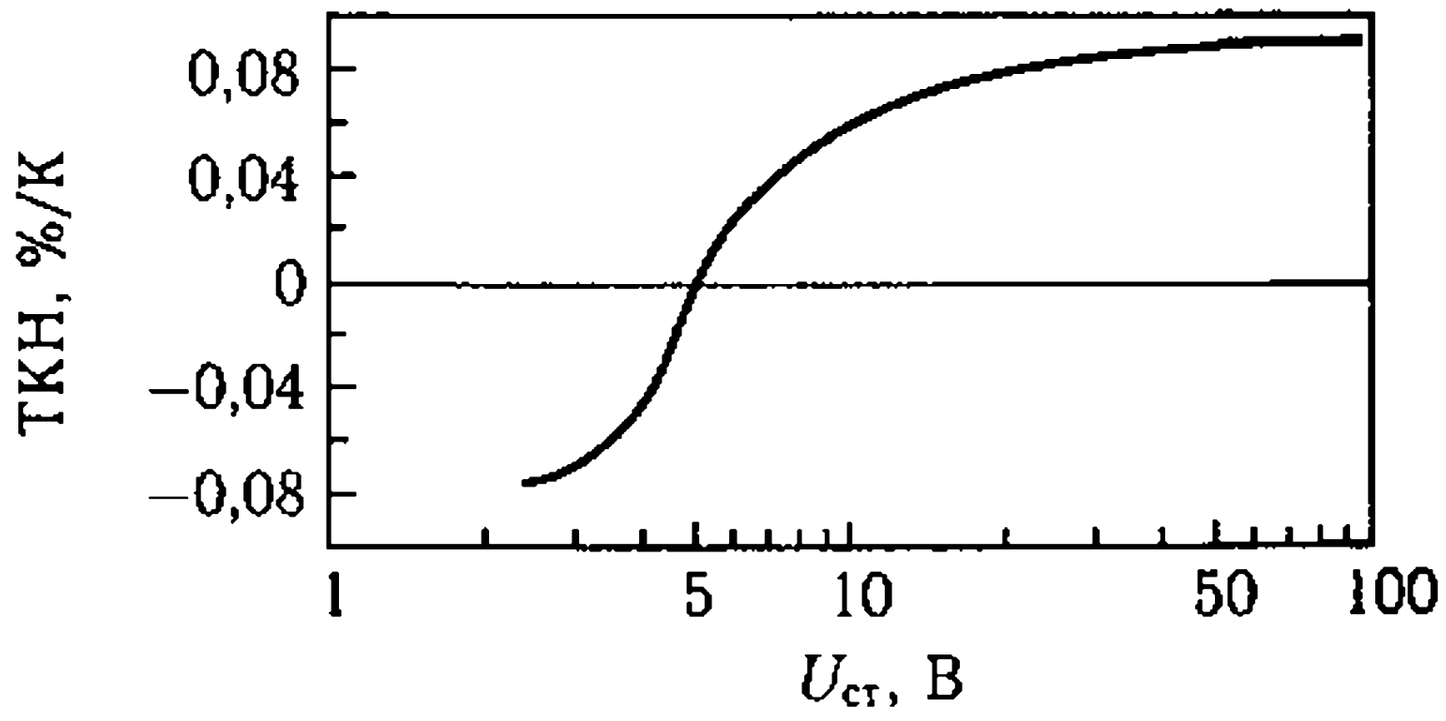


Рис. 1.19. Зависимость ТКН кремниевых стабилитронов от напряжения стабилизации

стабилитроны с ТКН = 0 называются *термокомпенсированными*. Примером термокомпенсированных стабилитронов с одним диодом, включенном в прямом направлении, могут служить отечественные стабилитроны КС170А с $U_{ст} \approx 7$ В и стабилитроны 1N821, 1N829 с $U_{ст} \approx 6{,}2$ В. Среди термокомпенсированных стабилитронов с несколькими диодами наиболее популярны приборы с напряжением стабилизации 9 и 11,7 В.

Современные стабилитроны, используемые как источники опорного напряжения, имеют ТКН, равный $(2–5) \cdot 10^{-6}$ $К^{-1}$ (КС196Г, 1N940), что заметно меньше ТКН электрохимических элементов Вестона, которые в течение многих лет использовались в качестве эталонов напряжения в измерительной технике ($(5–40) \cdot 10^{-6}$ $К^{-1}$). Для прецизионных источников эталонного напряжения оказывается важным еще один параметр — долговременная стабильность напряжения стабилизации. В современных специально разработанных стабилитронах (1N3501, 1N4890) относительный дрейф напряжения стабилизации за 1000 часов работы составляет $5 \cdot 10^{-6}$ или еще меньше. Еще более высокие параметры имеют источники опорного напряжения, в которых стабилитроны входят в состав интегральных схем (LTZ1000 с ТКН $= 5 \cdot 10^{-8}$ $К^{-1}$ и дрейфом $0{,}3 \cdot 10^{-6}$/месяц).

Коэффициент качества Q, определяемый как отношение дифференциального сопротивления стабилитрона $dU_{ст}/dI_{ст}$ при некотором рабочем токе $I_{ст}$ к «статическому» сопротивлению стабилитрона $R_{ст} = U_{ст}/I_{ст}$, характеризует насколько сильно изменение тока через стабилитрон влияет на напряжение стабилизации (изменение $I_{ст}$ может вызываться, например, изменением напряжения питания электронной схемы или изменением тока нагрузки). По сути, величина Q характеризует степень

нелинейности вольт-амперной характеристики в области пробоя. В кремниевых стабилитронах зависимость $Q(U_{ст})$ имеет вид кривой с минимумом, который лежит при $U_{ст} = 6$–8 В, то есть недалеко от напряжения, отвечающего условию ТКН $= 0$. Минимальное значение Q может составлять 0,01 и менее.

В *стабилитронах с очень низким напряжением стабилизации* (1–2 В) использование туннельного пробоя не позволяет получить высокий коэффициент качества из-за слишком плавной вольт-амперной характеристики в обратной ветви. Для таких напряжений стабилизации более высокие параметры удается получить в структурах с двумя p–n-переходами, в работе которых используется явление прокола. Работу этих приборов мы обсудим в п. 6.3.

Для практических применений важна не только стабильность напряжения стабилитрона, но и его уровень собственных шумов. Как мы отмечали в п. 1.3.1, для лавинного пробоя, в особенности в области малых токов, характерны большие флуктуации напряжения, связанные с существованием микроплазм (см. с. 55). [1] При увеличении тока диода пробой обычно становится более «стабильным» и уровень шума понижается. Тем не менее при лавинном пробое уровень шума всегда остается сравнительно высоким, что связано со статистическим характером процессов ударной ионизации. При понижении напряжения стабилизации, когда в приборах все более значительную роль начинает играть туннельный пробой, шумы стабилитронов резко ослабевают. В лучших современных стабилитронах напряжение шума составляет 1–20 мкВ в полосе частот 0,1–10 Гц.

Высокий уровень шума стабилитронов позволяет использовать их в качестве *генераторов шума*; с этой целью были разработаны приборы, плотность шума которых характеризуется высокой стабильностью (например, отечественный диод 2Г401).

1.4. Туннельные диоды

В 1958 г. Есаки, изучая сильно легированные германиевые p–n-переходы, обнаружил необычную вольт-амперную характеристику с падающим участком в прямой ветви (см. рис. 1.20*а*) и объяснил появление этого участка туннельным эффектом [27]. [2]

---

[1] В первых стабилитронах с $U_{ст} > 15$ В при токах менее ~0,3 мА напряжение шума иногда достигало долей вольта.

[2] За экспериментальное обнаружение туннельного эффекта в полупроводниках Лео Есаки был удостоен Нобелевской премии по физике в 1973 г.

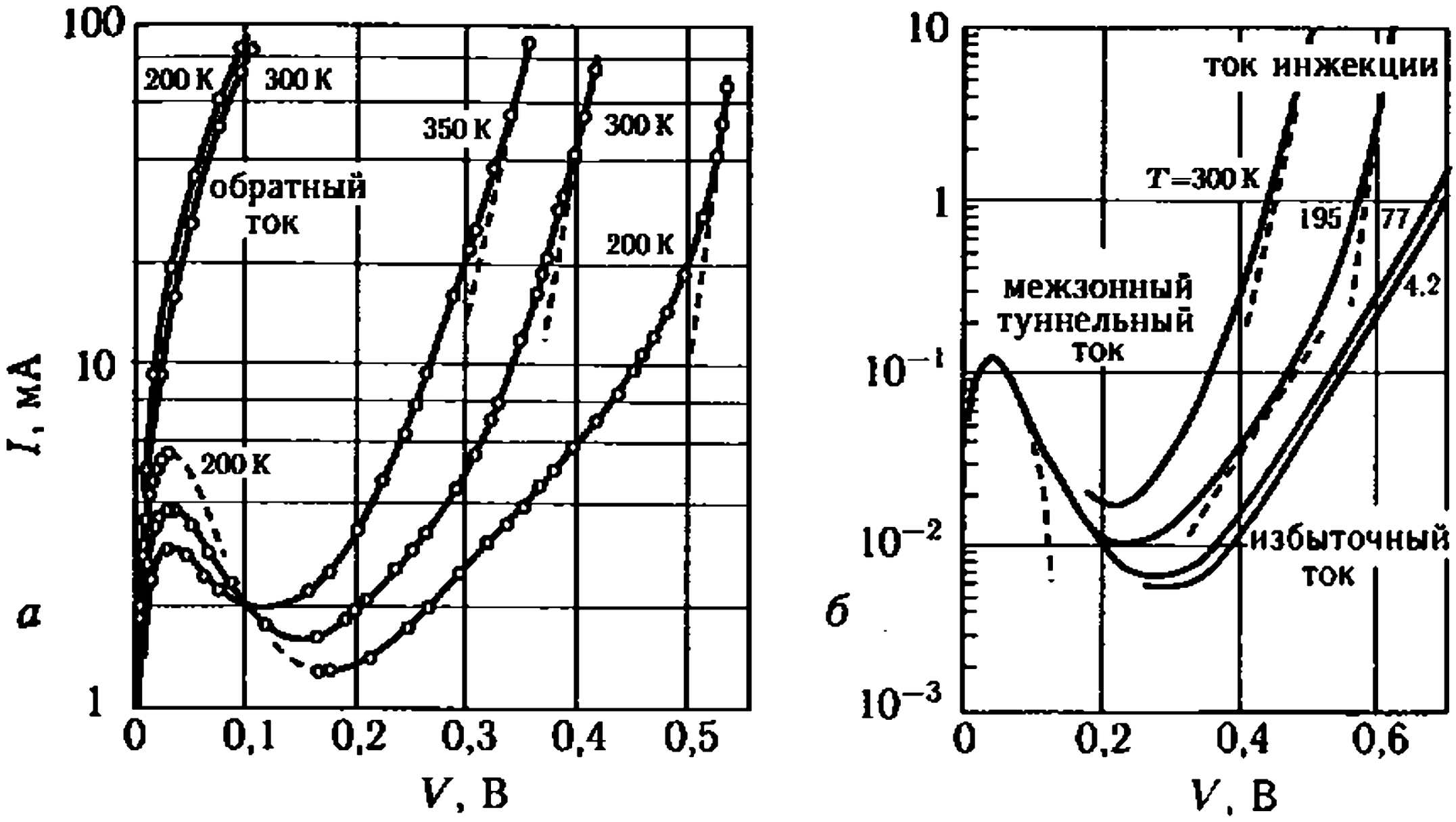


Рис. 1.20. Типичная вольт-амперная характеристика (*а*) [27] и три составляющие полного тока в германиевом туннельном диоде (*б*) [2]

Представление о туннелировании частиц возникло в процессе развития квантовой механики. Коль скоро электрон описывается волновой функцией, он может проникать в классически запрещенную область и туннелировать через достаточно тонкий потенциальный барьер. В 1928 г. Фаулер и Нордгейм [28] на основе этих представлений объяснили основные черты явления холодной электронной эмиссии — явления, которое оставалось необъясненным с момента его открытия Вудом в 1897 г. Почти сразу эти идеи нашли применение в теории α-распада. В 1934 г. Зенер [25] предложил идею *межзонного туннелирования* (внутренней автоэлектронной эмиссии) для объяснения пробоя в диэлектриках. В сущности, он показал, что запрещенную зону можно рассматривать как потенциальный барьер. Далее теория туннельного пробоя развивалась Хьюстоном [29] и Л.В. Келдышем [30]. После открытия туннельных диодов появляется большое число работ, посвященных объяснению их вольт-амперных характеристик. Среди этих работ следует выделить работы Прайса и Рэдклифа [31], Кейна [32, 33], И.И. Иванчика [34], В.Л. Бонч-Бруевича и П.С. Серебренникова [35–37].

О туннельном эффекте мы уже говорили в п. 1.3.2 при обсуждении явления туннельного пробоя и отмечали, что в сильно легированном p–n-переходе потенциальный барьер может оказаться достаточно прозрачным для туннелирования электронов.

То, что работа туннельных диодов основана на переносе основных носителей, создает предпосылки для широкого использования этих приборов в технике СВЧ. Однако особый интерес к туннельным диодам обусловлен тем, что на их вольт-амперной

характеристике имеется *падающий участок* (участок с отрицательным дифференциальным сопротивлением, $dV/dI < 0$), и, следовательно, эти приборы можно использовать не только для детектирования, но и для *усиления и генерации* высокочастотных сигналов [38].

1.4.1. Вольт-амперная характеристика туннельного диода. Обсудим формирование вольт-амперной характеристики туннельного диода сначала на качественном уровне. Рассмотрим p–n-переход, обе области которого вырождены: уровень Ферми в n-области лежит в зоне проводимости на расстоянии Δ_n от края зоны, а уровень Ферми в p-области лежит в валентной зоне на расстоянии Δ_p от ее края (см. рис. 1.21 *а*). В вырожденных полупроводниках уравнение, связывающее концентрацию носителей и энергию Ферми, отличается от уравнения для случая невырожденной статистики (формула (1.3)) и записывается в более общем виде [1]:

$$n = N_c \mathcal{F}_{1/2}(\Delta_n/kT), \quad p = N_v \mathcal{F}_{1/2}(\Delta_p/kT), \tag{1.83}$$

где

$$\mathcal{F}_{1/2}(\eta) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{x^{1/2} dx}{1 + \exp(x - \eta)} \tag{1.84}$$

— так называемый интеграл Ферми. Чтобы носители в зонах были вырожденными, концентрация легирующей примеси должна превышать эффективную плотность состояний в соответствующей зоне, которая по порядку величины составляет 10^{18}–10^{19} см$^{-3}$. Толщина обедненного слоя W при таких уровнях легирования равна всего нескольким сотням ангстрем.

В качестве примера оценим параметры p–n-перехода из Ge с типичными для туннельных диодов уровнями легирования ($N_d = 2 \times 10^{19}$ см$^{-3}$, $N_a = 5 \cdot 10^{19}$ см$^{-3}$) при 300 К. Учитывая, что $N_c = 1{,}04 \cdot 10^{19}$ см$^{-3}$, $N_v = 6{,}0 \cdot 10^{18}$ см$^{-3}$, с помощью уравнений (1.83) и (1.12) находим $\Delta_n = 34$ мэВ, $\Delta_p = 124$ мэВ, $q\phi_k = 0{,}818$ эВ, $W = 100$ Å.

При подаче на p–n-переход небольшого положительного смещения V через диод начинает течь ток электронов, туннелирующих из заполненных состояний n-области в не занятые электронами состояния валентной зоны p-области (рис. 1.21 *б*). При увеличении V туннельный ток при каком-то напряжении проходит через максимум и начинает уменьшаться, поскольку

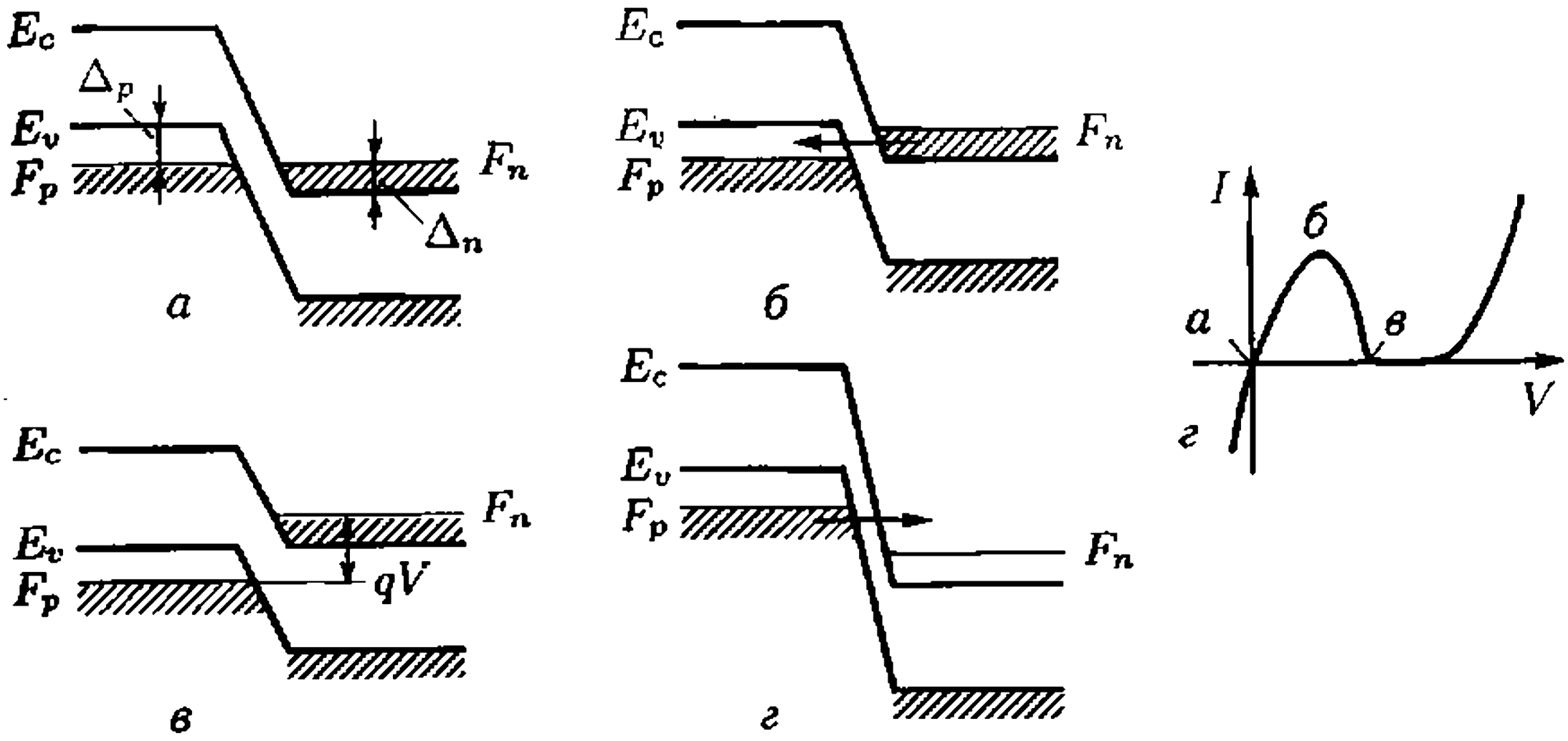


Рис. 1.21. Энергетические диаграммы туннельного диода при нулевом (*а*), прямом (*б*, *в*) и обратном (*г*) смещении. Справа на вольт-амперной характеристике показаны положения точек, которым соответствуют диаграммы

с ростом напряжения смещения «перекрытие» заполненных состояний в зоне проводимости n-области и пустых состояний в валентной зоне p-области становится все меньше и меньше и, кроме того, из-за понижения напряженности электрического поля прозрачность барьера D (см. формулу (1.80)) также уменьшается. Наконец, при напряжении $V = (\Delta_n + \Delta_p)/q$, когда край зоны проводимости в n-области оказывается точно напротив края валентной зоны в p-области (рис. 1.21 *в*), перекрытие разрешенных состояний исчезает и туннельный ток обращается в нуль. Таким образом, на вольт-амперной характеристике формируется падающий участок в прямой ветви. При дальнейшем увеличении V ток снова начинает возрастать за счет обычной инжекции (преодоления потенциального барьера). При обратном смещении на p–n-переходе ток, как и при туннельном пробое, определяется туннелированием электронов из заполненных состояний валентной зоны p-области в пустые состояния зоны проводимости n-области (рис. 1.21 *г*); этот ток быстро нарастает, поскольку одновременно возрастают и прозрачность барьера D (из-за роста напряженности электрического поля в p–n-переходе), и перекрытие зон, между которыми идет туннелирование.

Есаки предположил, что вклад электронов с энергией E в туннельные токи, текущие из зоны проводимости в валентную зону и в обратном направлении, пропорциональны числу заполненных состояний с энергией E в одной из зон, числу пустых состояний с той же энергией в другой зоне и вероятности

перехода. Полные токи, таким образом, равны:

$$J_{c\to v} = A \int_{E_c}^{E_v} f_c(E)\,\rho_c(E)\,D(E)[1 - f_v(E)]\rho_v(E)\,dE, \qquad (1.85)$$

$$J_{v\to c} = A \int_{E_c}^{E_v} f_v(E)\,\rho_v(E)\,D(E)[1 - f_c(E)]\rho_c(E)\,dE. \qquad (1.86)$$

Здесь f_c и f_v — функции распределения Ферми–Дирака, описывающие заполнение электронами состояний зоны проводимости и валентной зоны, ρ_c и ρ_v — плотности состояний в соответствующих зонах, D — прозрачность туннельного барьера, A — некий коэффициент. Интегрирование ведется по всей области энергий, в которой состояния зоны проводимости и валентной зоны перекрываются, то есть от энергии E_c, отвечающей дну зоны проводимости в n-области, до энергии E_v, отвечающей потолку валентной зоны в p-области.

Разность токов, определяемых выражениями (1.85) и (1.86), и представляет собой протекающий через туннельный диод ток:

$$J = J_{c\to v} - J_{v\to c} = A \int_{E_c}^{E_v} [f_c(E) - f_v(E)]D(E)\rho_c(E)\rho_v(E)\,dE. \qquad (1.87)$$

Рассчитав по этой формуле туннельный ток, Есаки получил качественное согласие с экспериментом.

К сожалению, предложенная Есаки модель является одномерной и не учитывает особенностей трехмерного движения электронов и закон сохранения импульса. Поэтому перейдем к строгой теории вольт-амперной характеристики туннельного диода в формулировке, данной Кейном [33].

Рассмотрим прямозонный полупроводник с кейновским законом дисперсии, в котором энергия электрона связана с квазиимпульсом $\hbar\mathbf{k}$ соотношением

$$E(\mathbf{k}) = \pm\sqrt{\left(\frac{E_g}{2}\right)^2 + \frac{\hbar^2\mathbf{k}^2 E_g}{2m^*}}\,, \qquad (1.88)$$

где m^* — эффективная масса в экстремумах зон, а E_g — ширина запрещенной зоны. [1]) Этот закон дисперсии описывает две

[1]) Этот закон дисперсии строго обосновывается в рамках $\mathbf{k}\hat{\mathbf{p}}$-метода расчета зонной структуры в случае, когда расстояние до других ветвей электронного

зеркально-симметричные изотропные зоны. Знак плюс в уравнении (1.88) отвечает состояниям зоны проводимости, а знак минус — состояниям валентной зоны. Кейновский закон дисперсии учитывает *непараболичность зон* (зависимость эффективной массы электронов и дырок от энергии), которая наблюдается во многих полупроводниках.

Пусть p–n-переход лежит в плоскости $z = 0$. Рассмотрим группу электронов с энергией в интервале $(E,\ E + dE)$ и z-компонентой импульса в интервале $(p_z,\ p_z + dp_z) = (\hbar k_z,\ \hbar k_z + \hbar dk_z)$, движущихся в полупроводнике из n-области в направлении p–n-перехода. Вклад этих электронов в туннельный ток равен произведению числа электронов, которые подходят к границе перехода единичной площади за единицу времени, на заряд электрона, прозрачность барьера для этих электронов и вероятность того, что с другой стороны перехода при энергии E есть незаполненное состояние:

$$dJ_{c\to v} = qv_z D(E)[1 - f_v(E)]\, dn(p_z) = \\ = q\frac{\partial E}{\partial p_z} D(E)[1 - f_v(E)]\, f_c(E)\frac{2}{(2\pi\hbar)^3}\, dp_x\, dp_y\, dp_z. \tag{1.89}$$

Для дальнейшего нам будет удобно перейти в цилиндрическую систему координат, ось которой направлена вдоль z. После несложных преобразований выражение (1.89) может быть переписано следующим образом:

$$dJ_{c\to v} = \frac{q}{2\pi^2\hbar^3} f_c(E)[1 - f_v(E)]\, D(E)\, p_\perp\, dp_\perp \frac{\partial E}{\partial p_z}\, dp_z. \tag{1.90}$$

Записывая аналогичное выражение для тока электронов из валентной зоны в зону проводимости $dJ_{v\to c}$ и вычитая его из (1.90), приходим к следующему выражению для вклада рассматриваемых электронов в туннельный ток:

$$dJ = dJ_{c\to v} - dJ_{v\to c} = \\ = \frac{q}{2\pi^2\hbar^3}[f_c(E) - f_v(E)]\, D(E)\, p_\perp\, dp_\perp \frac{\partial E}{\partial p_z}\, dp_z. \tag{1.91}$$

Вычисляя значение мнимой z-компоненты волнового вектора электрона в области туннелирования с помощью закона

спектра велико по сравнению с расстоянием между ближайшими зоной проводимости и валентной зоной [39].

дисперсии (1.88) и подставляя его в уравнение (1.79), получаем для вероятности туннелирования следующее выражение:

$$D(E) \approx \exp\left(-\frac{\pi\sqrt{2m^*}}{4\hbar}\,\frac{E_g^{3/2}}{q\mathcal{E}}\right)\exp\left(-\frac{p_\perp^2}{m^*E_0}\right), \qquad (1.92)$$

где $E_0 = \sqrt{2}\,q\hbar\mathcal{E}/(\pi\sqrt{m^*E_g})$. Появление второго сомножителя в этой формуле отражает тот факт, что на движение *перпендикулярно* границе p–n-перехода приходится лишь часть энергии электрона и дырки. При этом чем выше $p_\perp$, тем меньше оказывается величина p_z, точки поворота удаляются друг от друга к границам нейтральных областей p–n-перехода, и поэтому при туннелировании носителям приходится преодолевать все *более широкий барьер*, что и проявляется в уменьшении D. Отличие множителя в показателе экспоненты от множителя в формуле (1.80) связано с тем, что при выводе формулы (1.92) была учтена непараболичность закона дисперсии. [1)]

Подставляя найденную величину D в уравнение (1.91), теперь можно рассчитать полную плотность туннельного тока:

$$J = \frac{q}{2\pi^2\hbar^3}\exp\left(-\frac{\pi\sqrt{2m^*}}{4\hbar}\,\frac{E_g^{3/2}}{q\mathcal{E}}\right)\times \qquad (1.93)$$
$$\times\iint[f_c(E)-f_v(E)]\exp\left(-\frac{p_\perp^2}{m^*E_0}\right)p_\perp\,dp_\perp\,\frac{\partial E}{\partial p_z}\,dp_z.$$

Здесь интегрирование по $p_\perp$ проводится так, чтобы удовлетворить закону сохранения поперечной компоненты импульса при туннелировании. [2)] Поэтому пределы интегрирования по $p_\perp$ простираются от 0 до величины p_s, равной меньшему из двух максимальных значений импульса $p_{\perp 1}$ и $p_{\perp 2}$, отвечающих кинетическим энергиям $E_1 = E - E_c$ и $E_2 = E_v - E$, участвующих в туннелировании электрона и дырки.

[1)] На необходимость учета непараболичности при туннелировании указывают тонкие эксперименты, проведенные на диодах Шоттки [40]. Они показали, что масса туннелирующего электрона действительно меньше его эффективной массы на дне зоны проводимости и зависит от энергии.

[2)] При движении электрона в области пространственного заряда электрическое поле воздействует *только на продольную* компоненту импульса, оставляя поперечную компоненту неизменной.

Интеграл по $p_\perp$ легко берется, в результате чего получаем

$$J = \frac{qm^*}{2\pi^2\hbar^3}\exp\left(-\frac{\pi\sqrt{2m^*}}{4\hbar}\frac{E_g^{3/2}}{q\mathcal{E}}\right)\left(\frac{E_0}{2}\right)\times$$
$$\times\int_{E_c}^{E_v}[f_c(E)-f_v(E)]\left[1-\exp\left(-\frac{p_s^2}{m^*E_0}\right)\right]dE, \qquad (1.94)$$

где $p_s = \min\{p_{\perp 1}, p_{\perp 2}\}$.

Уравнение (1.94), в котором пределы интегрирования зависят от взаимного положения краев зоны проводимости и валентной зоны, а следовательно, от приложенного напряжения, описывает зависимость туннельного тока от напряжения смещения. Оценки, проведенные Кейном [33], показали, что напряжение, при котором туннельный ток достигает максимума, определяется соотношением

$$qV_p \approx (\Delta_n + \Delta_p)/3, \quad \text{если} \quad \max\{\Delta_n, \Delta_p\} \leqslant 2\min\{\Delta_n, \Delta_p\},$$

и

$$qV_p \approx \min\{\Delta_n, \Delta_p\}$$

в противном случае. Эти зависимости хорошо согласуются с зависимостями, наблюдаемыми на опыте.

Обсудим теперь вопрос о применимости рассмотренной теории к расчету вольт-амперных характеристик туннельных диодов в реальных полупроводниках. Как известно, в большинстве полупроводников эффективные массы электронов и дырок различаются, а энергетический спектр вблизи края валентной зоны даже в наиболее простых полупроводниках состоит из двух энергетических зон, называемых зонами легких и тяжелых дырок [1]. Тем не менее, в прямозонных полупроводниках (InSb, GaAs) эффективная масса легких дырок близка к эффективной массе электронов, а закон дисперсии этих носителей неплохо описывается кейновским законом дисперсии.

Л.В. Келдыш [30] обратил внимание на то, что в первых теориях межзонного туннелирования не учитывалось различие эффективных масс в валентной зоне и зоне проводимости. Он показал, что в прямозонных полупроводниках в уравнение (1.92) для вероятности туннелирования должна входить *приведенная масса*, учитывающая законы дисперсии носителей в обеих зонах. Эта масса выражается через эффективные массы электронов

и дырок следующим образом:

$$\frac{2}{m^*} = \frac{1}{m_n^*} + \frac{1}{m_p^*}. \tag{1.95}$$

Из-за сильной зависимости прозрачности барьера D от m^* (см. формулу (1.92)) переходы, в которых принимают участие *тяжелые дырки*, имеют существенно меньшую вероятность по сравнению с переходами с участием *легких дырок*, и реальный вклад в туннельный ток дают только последние переходы. Поэтому, если учитывать в туннелировании только электроны и легкие дырки, то использование полученных выше уравнений для прямозонных полупроводников оказывается вполне обоснованным.

Туннелирование в непрямозонных полупроводниках. Рассмотренный выше подход применим к прямозонным полупроводникам (например, GaAs, в котором экстремумы зоны проводимости и валентной зоны расположены в Г-точке зоны Бриллюэна). Однако он не годится для *непрямозонных полупроводников*, таких как Si и Ge, [1]) в которых экстремумы зон расположены в разных точках зоны Бриллюэна. Как было показано Келдышем [30], в этих полупроводниках прямое туннелирование чрезвычайно затруднено. Поэтому для наблюдения в непрямозонном полупроводнике туннелирования необходимо, чтобы в процессе туннелирования кроме электрона и дырки участвовала бы еще третья частица, которая «уносила» бы с собой разность квазиимпульсов. Эксперимент показывает, что такими «частицами» могут быть фононы и атомы примеси.

Туннелирование с участием фононов теоретически рассматривалось Келдышем [30], Прайсом и Рэдклифом [31] и Кейном [33]. В полупроводнике, в котором минимум зоны проводимости и потолок валентной зоны находятся в разных точках зоны Бриллюэна, квазиимпульс, равный разности квазиимпульсов электрона в начальном и конечном состояниях, может передаваться решетке либо путем поглощения фонона, либо путем его испускания. Если фонон поглощается, то энергия электрона возрастает на величину энергии фонона, а если фонон испускается — то уменьшается на ту же величину. По оценкам Кейна [33], в Ge вероятность туннельных переходов с участием фононов примерно в 4000 раз меньше, чем для прямого туннелирования.

---

[1]) Список непрямозонных полупроводников можно найти в табл. 2 Приложения.

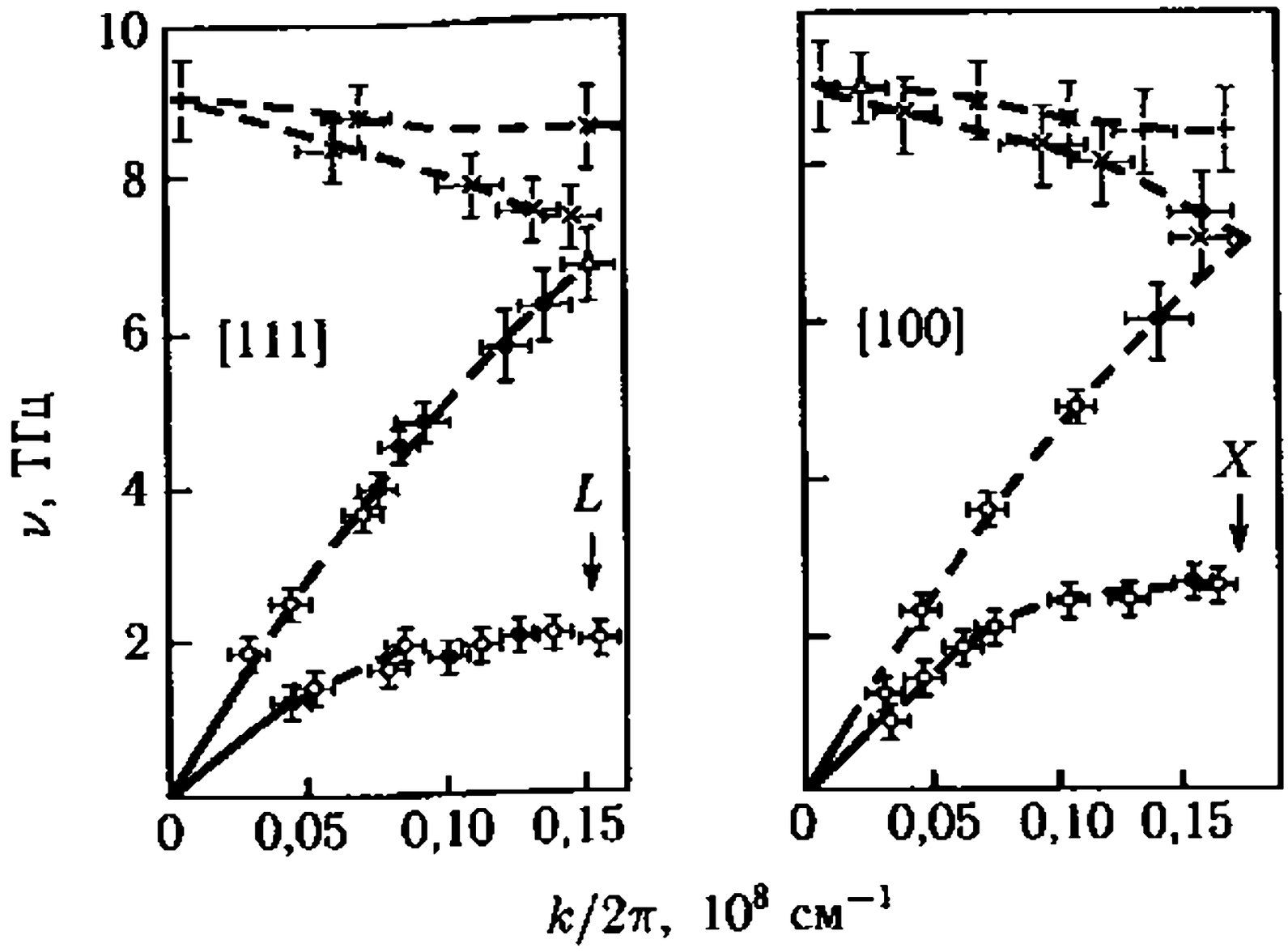


Рис. 1.22. Фононный спектр германия по данным неупругого рассеяния нейтронов [41]. Частоте 1 ТГц соответствует энергия $\approx 4{,}14$ мэВ

Важные данные, проливающие свет на действительную роль фононов и примесей в процессах туннелирования, были получены при исследовании туннельных диодов при низких температурах (4,2 К и ниже). [1]) При низкой температуре, когда фононы практически не возбуждены, туннельные переходы могут идти только с *испусканием фононов*. При этом все состояния выше уровня Ферми пусты, а ниже уровня Ферми — полностью заполнены. Поскольку туннелирование может идти только из заполненных состояний в пустые и при этом должно сопровождаться передачей энергии фонону, это значит, что туннельный ток начнет течь только тогда, когда напряжение на p–n-переходе превысит пороговое значение, связанное с энергией фонона соотношением $qV_{\text{пор}} = \hbar\omega$. В Ge, Si, соединениях $A^{III}B^{V}$ и $A^{II}B^{VI}$ фононный спектр кристалла состоит из двух акустических (LA, TA) и двух оптических (LO, TO) ветвей (см. рис. 1.22). Поэтому можно ожидать, что изменение дифференциальной проводимости dI/dV туннельного диода будет происходить каждый раз, когда значение qV будет превышать энергию очередного фонона

[1]) Следует отметить, что в отличие от обычных диодов, выпрямляющие свойства которых резко ухудшаются при охлаждении до температур ниже 20–50 К из-за вымораживания свободных носителей на примесях, в туннельных диодах этого не происходит и их можно исследовать вплоть до самых низких температур. Это является следствием того, что при высокой концентрации примеси образованная из примесных уровней примесная зона сливается с краем ближайшей разрешенной зоны и примесные атомы при любой температуре остаются полностью ионизованными.

с квазиимпульсом, равным разности квазиимпульсов экстремумов зон.

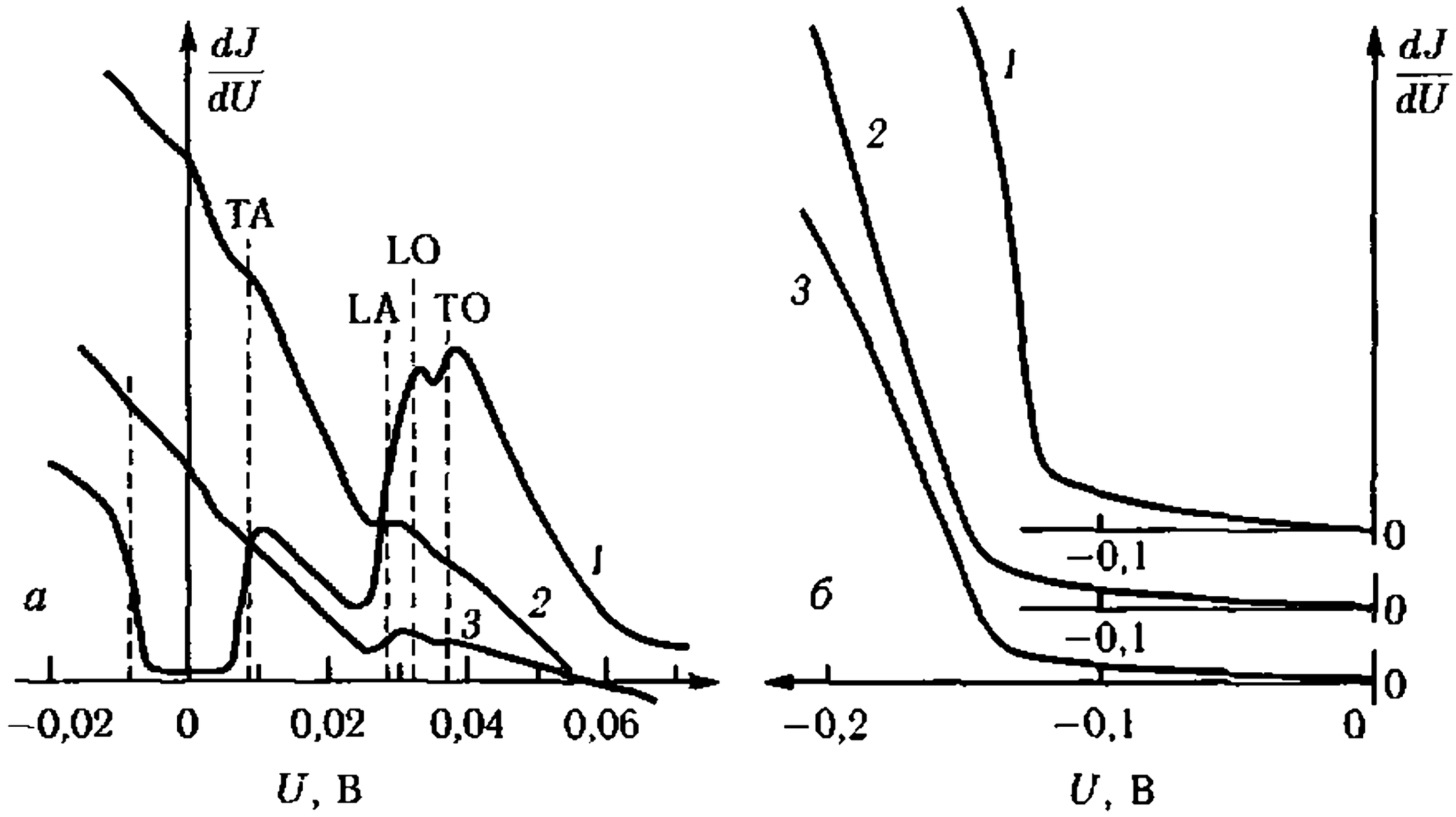


Рис. 1.23. Тонкая структура в дифференциальной проводимости германиевых туннельных диодов при 4,2 К, связанная с участием фононов в туннелировании (*а*) и туннелированием в Г-минимум зоны проводимости (*б*). Цифры на рисунке обозначают германий, легированный примесям Sb (*1*), As (*2*) и P (*3*) [2]

Исследования дифференциальной проводимости dI/dV туннельных диодов при низкой температуре действительно обнаруживают особенности в вольт-амперных характеристиках при напряжениях, соответствующих энергиям акустических и оптических фононов [42]. Так, в германиевых туннельных диодах энергии этих особенностей соответствуют энергиям фононов в L-точке зоны Бриллюэна (см. рис. 1.23*а*). Однако, как следует из рисунка, относительная величина этих особенностей сильно зависит от типа легирующей примеси, использованной при создании туннельного диода. Последний факт ясно указывает на то, что одновременно с участием фононов в туннелировании проявляется и другой механизм туннелирования, который связан с присутствующими в образце *примесями*. Это значит, что как и фононы, атомы примеси также могут выступать как центры рассеяния импульса. Теория туннелирования с участием примесей была развита П.С. Серебренниковым [36].

Исследование вольт-амперных характеристик туннельных диодов при низких температурах позволяет обнаружить и другие особенности зонной структуры полупроводника. Так, например, на рис. 1.23*б* видно, что при обратном напряжении около 0,12 В дифференциальная проводимость германиевых туннельных диодов

начинает резко возрастать. Это связано с прямыми туннельными переходами из валентной зоны в минимум зоны проводимости, который расположен в Г-точке зоны Бриллюэна немного выше основного минимума зоны проводимости в L-точке. Быстрое нарастание туннельного тока при этом объясняется тем, что туннельные переходы в Г-точку являются прямыми и, следовательно, имеют существенно более высокую вероятность по сравнению с непрямыми переходами в L-минимум зоны проводимости.

Совокупность методов, использующих явление туннелирования для исследования материалов, получила название *туннельной спектроскопии* и широко применяется для исследования зонной структуры, фононного спектра, поляронных и других эффектов в твердых телах [40].

До сих пор при обсуждении эффекта туннелирования мы ничего не говорили о спине туннелирующего электрона. В то же время в *магнитных полупроводниках* учет спиновой переменной может приводить к целому ряду физических явлений, на основе которых уже создаются новые типы приборов — *магниторезистивные запоминающие устройства* (MRAM). В основе этих явлений лежит то, что в ферромагнетике из-за снятия крамерсова вырождения зоны разрешенных зон для состояний «спин вверх» и «спин вниз» расположены при разной энергии (см. рис. 1.24). Поэтому вероятность туннелирования, которое происходит с сохранением ориентации спина, в структурах ферромагнетик–немагнитный материал–ферромагнетик зависит от взаимной ориентации намагниченности в соседних слоях. Действительно, как схематично показано на рис. 1.24, при подаче на структуру небольшого напряжения смещения при параллельной ориентации намагниченности для туннелирующих электронов всегда находятся незаполненные конечные состояния с той же ориентацией спина, а при антипараллельной ориентации таких состояний нет. Таким образом, изменение взаимной ориентации намагниченностей должно приводить к заметному (на десятки процентов [43]) изменению туннельной проводимости структуры. Этот эффект, называемый *туннельным магнетосопротивлением*, лежит

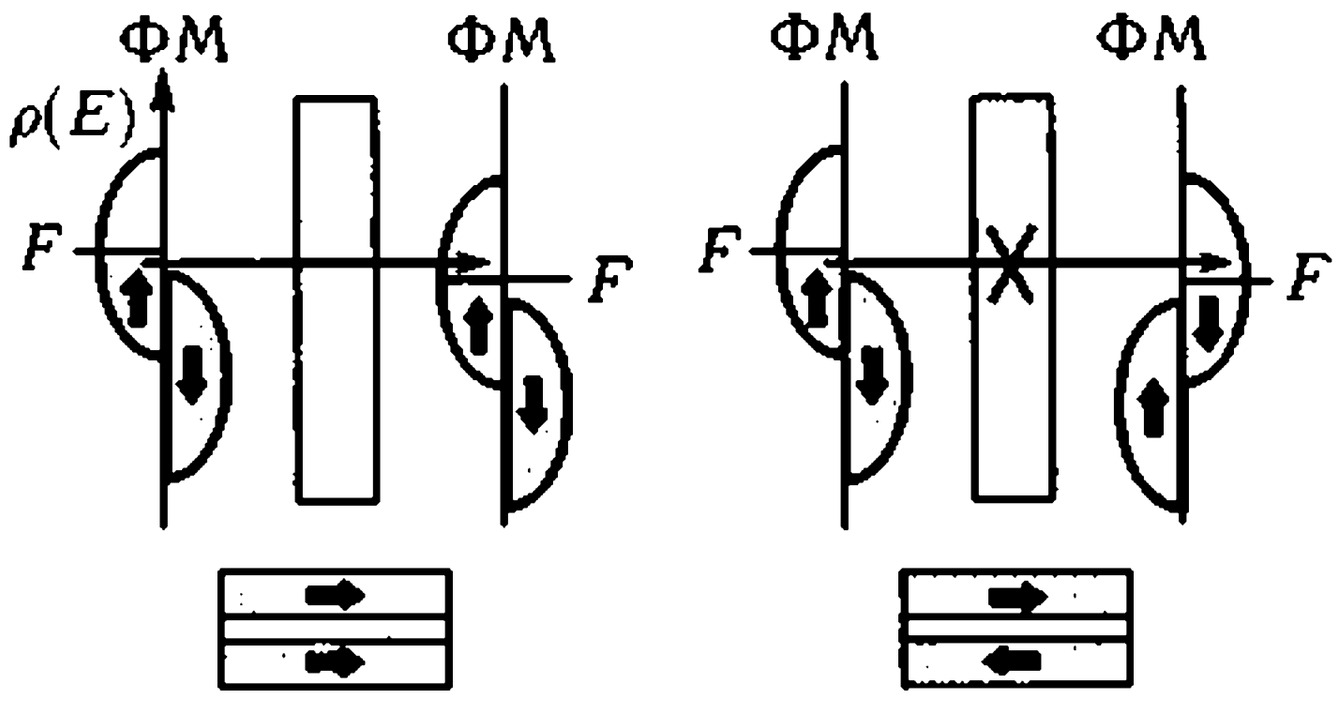


Рис. 1.24. К объяснению магниторезистивного эффекта, возникающего при туннелировании электронов между двумя ферромагнетиками с параллельной и антипараллельной ориентацией намагниченности

в основе работы микросхем MRAM емкостью 4–16 Мбит, выпускаемых в настоящее время фирмами Motorola, IBM и Infineon.

В исследовательских лабораториях ведется активный поиск магнитных полупроводников, которые могли бы заменить ферромагнитные металлы (Co–Fe), используемые в современных конструкциях MRAM. К сожалению, известный магнитный полупроводник $Ga_{1-x}Mn_xAs$ имеет максимальную температуру Кюри около 110 К, которая недостаточна для использования в приборах, работающих при комнатной температуре. Ферромагнитное состояние при 300 К недавно удалось получить в полупроводниковых твердых растворах $Ga_{1-x}Mn_xP$ и $Ga_{1-x}Mn_xN$ [44]. Более подробно прочитать о различных физических явлениях, на основе которых можно построить магниторезистивные запоминающие устройства, можно в обзоре [45].

1.4.2. Избыточный ток в туннельных диодах. Уже первые исследования вольт-амперных характеристик туннельных диодов обнаружили, что в туннельных диодах кроме туннельного тока и тока инжекции существует еще одна компонента тока, которая была названа *избыточным током* [46, 47]. Действительно, если из полного тока диода вычесть расчетные зависимости туннельного тока и тока инжекции (см. рис. 1.20*б*), то на вольт-амперной характеристике становится ясно видна компонента тока, которая обычно экспоненциально зависит от напряжения. Характерной особенностью избыточного тока является слабая зависимость тока и наклона вольт-амперной характеристики $d\ln J/dV$ от температуры.

Исследования показали, что на величину избыточного тока сильно влияет присутствие в образцах примесей, создающих глубокие уровни в запрещенной зоне. Так, намеренное введение таких примесей в полупроводник, используемый для создания туннельных диодов, приводит к сильному увеличению избыточного тока (см. рис. 1.25*а*). Иногда в таких образцах в области избыточного тока удается наблюдать второй максимум, который сопровождается вторым падающим участком на вольт-амперной характеристике (см. рис. 1.26), что ясно указывает на туннельную природу избыточного тока. К сильному увеличению избыточного тока приводит и облучение туннельных диодов быстрыми электронами или нейтронами (облучение вызывает образование в полупроводнике радиационных дефектов) [47]. Совокупность этих фактов позволяет связать появление избыточного тока с *туннелированием через примесные состояния*. Поскольку полупроводники, из которых изготавливают туннельные диоды, содержат высокие концентрации легирующих примесей, то кроме уровней глубоких примесей и дефектов,

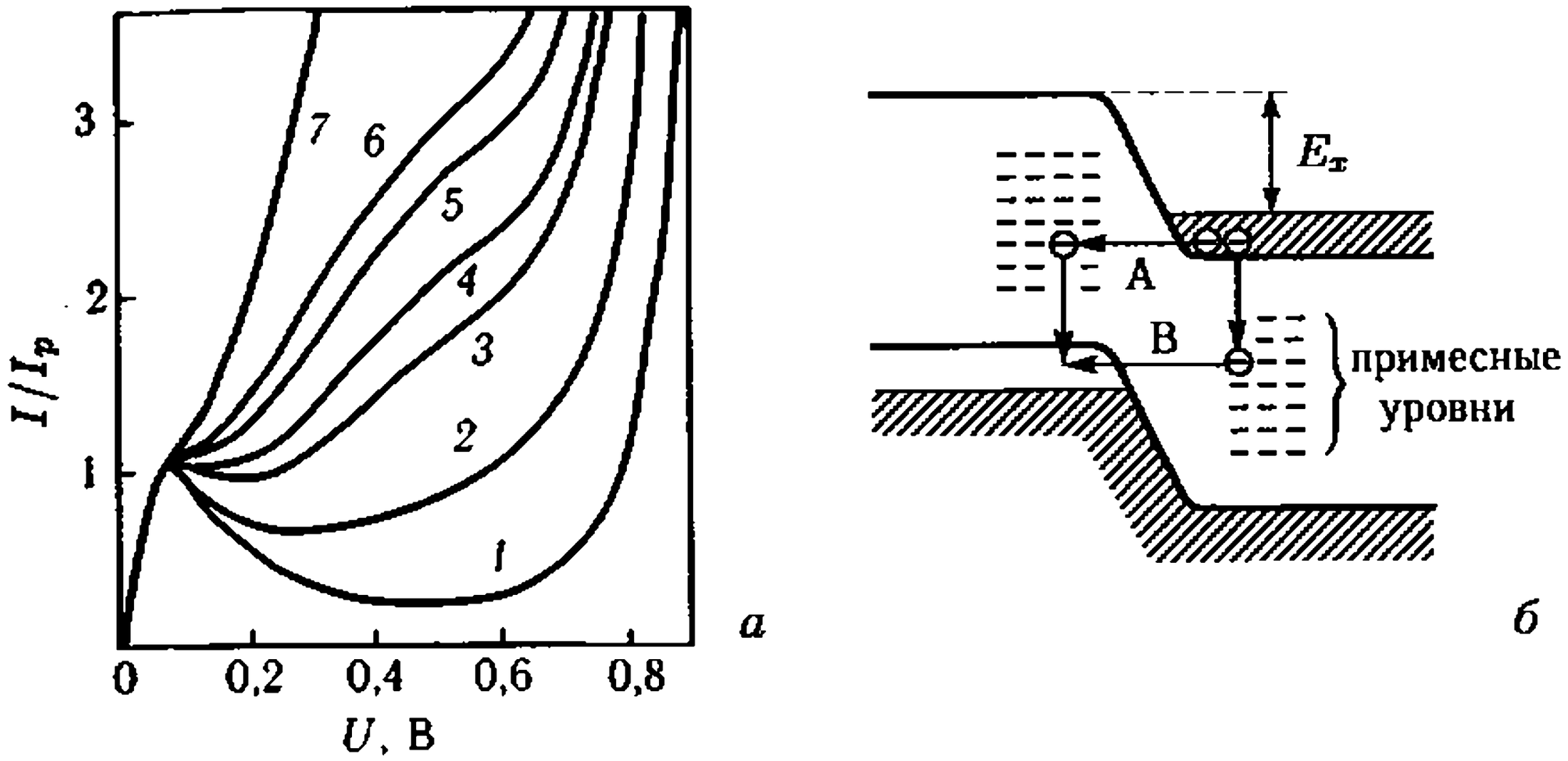


Рис. 1.25. Изменение вольт-амперных характеристик туннельного диода из Si при легировании исходного материала золотом (*а*). Концентрация примеси возрастает от кривой *1* к кривой *7*. Диаграмма, поясняющая механизм образования избыточного тока (*б*)

в избыточный ток могут давать заметный вклад и *хвосты плотности состояний*, возникающие вблизи краев зон в сильно легированных полупроводниках [1].

Рассмотрим энергетическую диаграмму туннельного диода при напряжении смещения, при котором прямое туннелирование электронов уже невозможно (рис. 1.25*б*). Существует два канала протекания избыточного тока: туннелирование электрона из зоны проводимости на уровень в запрещенной зоне, на котором он затем рекомбинирует с дыркой (канал А), и захват электрона на уровень в запрещенной зоне с последующим его туннелированием в незаполненные состояния валентной зоны (канал В). Для этих каналов характерна схожая зависимость тока от напряжения смещения, поэтому рассмотрим только первый канал. Для туннелирования электрона, находящегося на уровне Ферми, ему надо преодолеть потенциальный барьер высотой $E_x = q(\phi_k - V)$ (см. рис. 1.25*б*); для электронов, лежащих ниже уровня Ферми, это равенство можно считать выполненным приближенно. По аналогии с формулой (1.80), в квазиклассическом приближении вероятность такого перехода равна

$$D \approx \exp\left(-\frac{4\sqrt{2m^*}}{3\hbar}\frac{E_x^{3/2}}{q\mathcal{E}}\right), \qquad (1.96)$$

где $\mathcal{E}$ — среднее значение напряженности электрического поля в p–n-переходе. Учитывая, что в резком p–n-переходе величина $\mathcal{E}$

изменяется с напряжением смещения как $\mathcal{E} \sim (\phi_k - V)^{1/2}$, после подстановки выражений для E_x и $\mathcal{E}$ в уравнение (1.96) получаем

$$D(V) \sim \exp[-\alpha(\phi_k - V)] \sim \exp(\alpha V), \qquad (1.97)$$

где коэффициент α определяется электронными параметрами полупроводника и практически не зависит от температуры. Если плотность состояний дефектов в запрещенной зоне, на которые идет туннелирование, не зависит от энергии или изменяется по экспоненциальному закону, то избыточный ток, пропорциональный произведению $D(V)$ на плотность состояний, будет описываться экспоненциальной зависимостью от напряжения. Это соответствует зависимостям, обычно наблюдаемым в туннельных диодах (см. рис. 1.20). Если же в плотности состояний в запрещенной зоне будут проявляться явные особенности (связанные, например, с дискретными энергетическими уровнями примесей или дефектов), то в области избыточного тока могут появиться особенности типа горбов (см. рис. 1.25*а*) или даже дополнительных пиков, сопровождаемых вторым падающим участком на вольт-амперной характеристике (см. рис. 1.26).

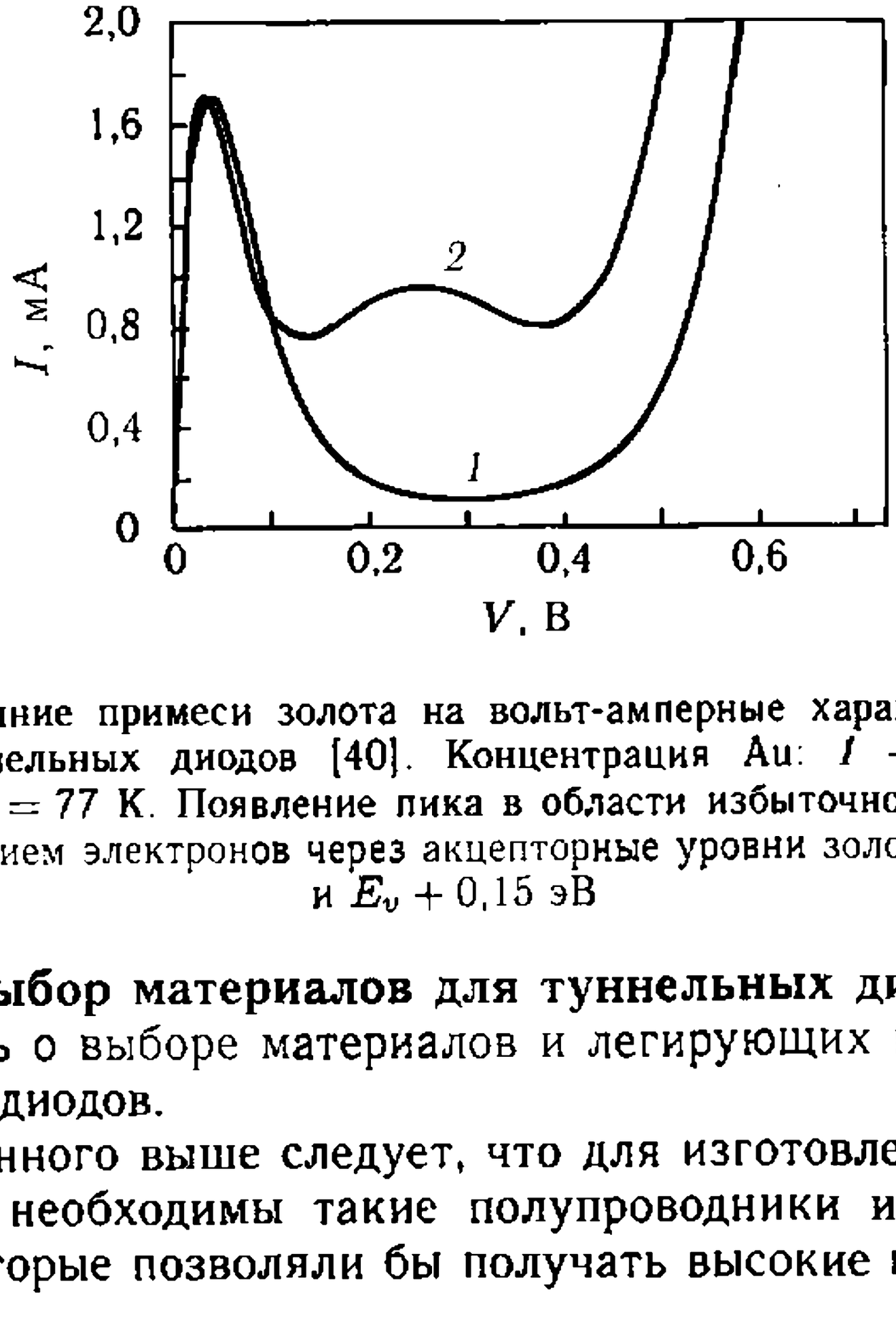


Рис. 1.26. Влияние примеси золота на вольт-амперные характеристики германиевых туннельных диодов [40]. Концентрация Au: *1* — 0, *2* — 2×10^{16} см$^{-3}$, $T = 77$ К. Появление пика в области избыточного тока связано с туннелированием электронов через акцепторные уровни золота $E_c - 0{,}20$ эВ и $E_v + 0{,}15$ эВ

1.4.3. Выбор материалов для туннельных диодов. Поговорим теперь о выборе материалов и легирующих примесей для туннельных диодов.

Из сказанного выше следует, что для изготовления туннельных диодов необходимы такие полупроводники и легирующие примеси, которые позволяли бы получать высокие концентрации

носителей в разрешенных зонах и в то же время в запрещенной зоне при этом не возникало бы высокой плотности глубоких состояний. Параметром, характеризующим то, насколько выполняется это требование, является отношение тока диода в пике J_p к току в минимуме («долине») J_v. Хотя туннельные диоды удалось создать из нескольких десятков различных полупроводников, практическое применение находят только три из них — Ge, GaAs и GaSb, в которых удается получить отношение J_p/J_v более 15. В остальных полупроводниках избыточный ток оказывается слишком большим [48].

От выбора материала для туннельных диодов сильно зависят и их шумовые характеристики. Основным источником шума в туннельном диоде является дробовой шум (природа различных источников шума в полупроводниковых приборах будет рассмотрена нами в п. 2.6). Параметром, характеризующим уровень шума туннельного диода, является *шум-фактор*, который определяется выражением

$$F = 1 + \frac{q}{2kT}|R_d I|_{\min}, \tag{1.98}$$

где $|R_d I|_{\min}$ — минимальное значение произведения модуля отрицательного дифференциального сопротивления на ток диода [49]. Наименьшими значениями шум-фактора характеризуются туннельные диоды из GaSb ($F = 1,9$) и Ge ($F = 2,2$); диоды из GaAs имеют более высокий уровень шума ($F = 3,4$).

Чтобы создать хороший туннельный диод, важен *выбор технологии* его получения, поскольку она существенно влияет на величину избыточного тока. В настоящее время наибольшее распространение получила технология создания туннельных диодов путем *вплавления*. Так, туннельные диоды из Ge получают вплавлением сплава In+(0,5–2)%Ga в материал n-типа или сплава Sn+5%As в p-Ge. Туннельные диоды из GaAs получают вплавлением олова в материал p-типа. Температура, продолжительность вплавления и скорость последующего охлаждения сильно влияют на характеристики получаемых диодов; например, при медленном охлаждении из-за диффузии примесей p–n-переход становится менее резким, что ведет к сильному уменьшению туннельного тока. С другой стороны, слишком быстрое охлаждение кристалла, характерное для другого способа получения туннельных диодов — *импульсной сварки*, не позволяет получить в диодах отношение J_p/J_v более 2,5 из-за высокой концентрации образующихся дефектов.

В туннельных диодах из полупроводников группы $A^{III}B^{V}$ важную роль играет и ориентация поверхности, в которую происходит вплавление. [1]) Так, например, в диодах из GaAs наибольшую плотность тока J_p имеют диоды, полученные вплавлением в поверхность (111)B.

Выбор легирующей примеси также имеет важное значение. Так, например, германиевые туннельные диоды можно создавать, легируя p-Ge донорными примесями P, As и Sb. Однако практическое значение имеет только примесь мышьяка, поскольку при легировании фосфором избыточный ток в диодах оказывается существенно выше, чем при легировании As, а при легировании сурьмой он настолько велик, что падающий участок на вольт-амперной характеристике вообще исчезает. Исследования показали, что появление избыточного тока при легировании германия фосфором и сурьмой связано с образованием примесных комплексов.

В туннельных диодах из GaAs, работающих в переключательных схемах, был обнаружен *эффект старения (деградации)*, который отсутствовал в германиевых диодах [48]. Этот эффект состоял в быстром уменьшении тока J_p при работе диода при прямом смещении, превышающем напряжение в минимуме. Было установлено, что старение диодов ускоряется при работе их в условиях повышенной температуры, однако если просто хранить диоды (не пропуская ток) при той же или даже более высокой температуре, то это не вызывает деградации их характеристик. Старение диодов сопровождалось уменьшением емкости p–n-переходов, что указывало на увеличение ширины области пространственного заряда в процессе деградации. Исследования этого явления установили, что причиной деградации диодов из GaAs (это относится также и к диодам из GaSb) является переход атомов Zn, используемого для легирования p-области диода, в междоузельное положение, в котором эти атомы проявляют донорное действие. Такой переход оказывается возможным в результате передачи энергии, выделяемой при рекомбинации электронно-дырочной пары, примесному атому, энергия связи которого в решетке меньше, чем для основных атомов Ga и As.

---

[1]) Атомарно чистая поверхность (111) в полупроводниках $A^{III}B^{V}$ может заканчиваться либо слоем атомов элемента A, либо слоем атомов элемента B. Эти поверхности имеют различные физические свойства и обозначаются, соответственно, (111)A и (111)B.

Изменение электрического состояния атомов цинка уменьшает концентрацию дырок в p-области диода, увеличивает толщину области пространственного заряда и резко уменьшает величину туннельного тока. То, что деградация диодов происходит только при высоких напряжениях смещениях, объясняется тем, что для создания неравновесных электронно-дырочных пар необходима инжекция. Таким образом, стабильность туннельных диодов может сильно зависеть от выбора типа легирующей примеси.

Обсудим теперь вопрос о выборе концентрации легирующей примеси. Одним из важных требований, предъявляемых к туннельным диодам, является стабильность их вольт-амперной характеристики при изменении температуры. Известно, что в большинстве полупроводников ширина запрещенной зоны с ростом температуры уменьшается. Поэтому, согласно формуле (1.80), прозрачность барьера D с ростом температуры возрастает, приводя к увеличению тока в пике J_p. В непрямозонном полупроводнике вероятность туннелирования с участием фононов также увеличивается с ростом температуры из-за увеличения чисел заполнения фононов. С другой стороны, уменьшение энергий Ферми (Δ_n, Δ_p) за счет увеличения эффективных плотностей состояний N_c и N_v и «размытие» ступеньки в распределении Ферми с ростом температуры уменьшают туннельный ток. Поскольку с понижением концентрации примеси роль последних двух конкурирующих факторов возрастает, то появляется возможность управлять температурным коэффициентом J_p путем выбора концентраций легирующих примесей. При комнатной температуре компенсация двух указанных тенденций происходит при концентрации примесей $\sim 2,5 \cdot 10^{19}$ см$^{-3}$ в туннельных диодах из Ge и при $\sim 6 \cdot 10^{19}$ см$^{-3}$ в диодах из GaAs. Так как диффузионный и избыточный токи возрастают с увеличением температуры, то отношение J_p/J_v с ростом температуры обычно уменьшается, что определяет существование некой верхней границы рабочих температур. Однако в реальных туннельных диодах максимальную рабочую температуру ограничивают значениями 70–85 °C , чтобы предотвратить медленную деградацию характеристик приборов из-за диффузии примесей в очень тонком p–n-переходе.

Другой характеристикой туннельного диода, сильно зависящей от уровня легирования, является его *быстродействие*. В качестве параметра, характеризующего быстродействие туннельного диода, обычно выбирают постоянную времени, равную произведению минимального значения модуля отрицательного

дифференциального сопротивления $|R_{d.\min}|$ на емкость p–n-перехода в минимуме вольт-амперной характеристики C_j. Величину $|R_{d.\min}|$ можно оценить с помощью эмпирической формулы $|R_{d.\min}| \approx 2V_p/J_p$, где V_p — напряжение пика на вольт-амперной характеристике [14]. Поскольку удельная емкость p–n-перехода (см. п. 1.7.1) и напряжение пика V_p возрастают с увеличением концентрации легирующей примеси по степенному закону, а плотность туннельного тока J_p — экспоненциально, то ясно, что для увеличения быстродействия туннельного диода необходимо легировать его p- и n-области как можно сильнее. Кроме того, на быстродействие туннельных диодов влияет и последовательное сопротивление диода R_s, равное сумме сопротивления растекания [1]) и сопротивления омических контактов, которые также уменьшаются с ростом уровня легирования.

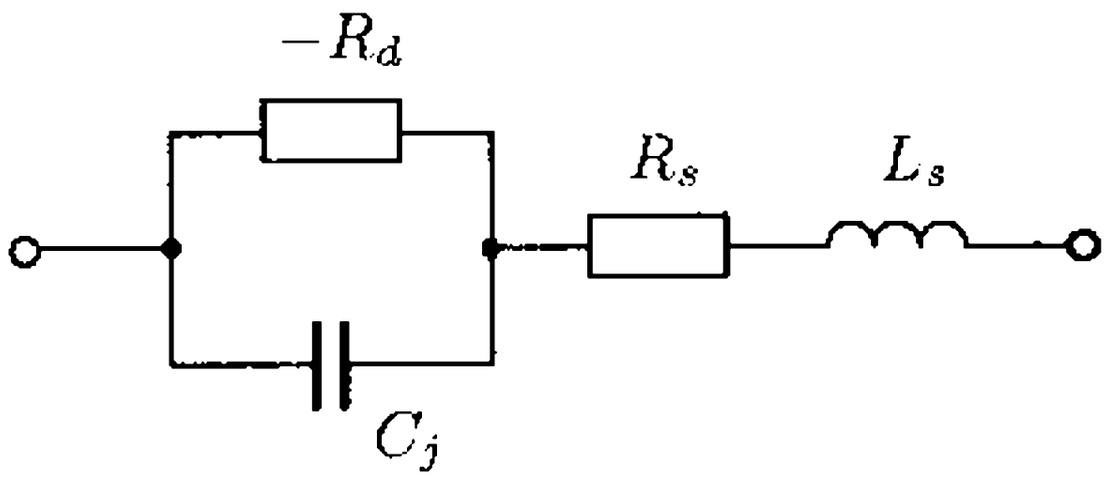


Рис. 1.27. Эквивалентная схема туннельного диода

С точки зрения механизма образования отрицательного дифференциального сопротивления туннельный диод является практически безынерционным прибором. Реальная частота, до которой он может работать, ограничивается барьерной емкостью перехода C_j и последовательным сопротивлением R_s. В соответствии с эквивалентной схемой, показанной на рис. 1.27, частота, до которой действительная часть комплексного сопротивления туннельного диода остается отрицательной, равна

$$f_r = \frac{1}{2\pi R_{d.\min} C_j} \sqrt{\frac{R_{d.\min}}{R_s} - 1}\ . \qquad (1.99)$$

Эта частота называется *частотой отсечки* туннельного диода. В германиевых туннельных диодах f_r достигает 30 ГГц, а в тщательно сконструированных туннельных диодах из GaAs удалось наблюдать генерацию на частоте 103 ГГц [50]. Анализ эквивалентной схемы показывает, что для устойчивой работы

---

[1]) *Сопротивлением растекания* называется эффективное сопротивление толщи образца в условиях, когда линии тока в образце сконцентрированы в одном месте, например, вокруг точечного контакта. Для полусферического контакта радиуса r сопротивление растекания равно $R_s = \rho/2\pi r$, где ρ — удельное сопротивление материала.

(без паразитной генерации на частотах ниже f_r) индуктивность диода L_s должна быть очень мала.

Отечественная промышленность выпускает целый ряд туннельных диодов из GaAs и Ge, предназначенных для работы в усилительных, генераторных или переключающих схемах. Примерами диодов для этих применений могут служить диоды ГИ103, АИ201, ГИ307А, соответственно.

1.4.4. Обращенные диоды. Другим интересным прибором, в котором используется туннельный эффект в p–n-переходе, является *обращенный диод*. В этих приборах уровень легирования областей диода подбирается таким образом, чтобы туннелирование электронов только начинало проявляться при нулевом смещении. Быстрое увеличение прозрачности потенциального барьера с ростом обратного смещения приводит к тому, что обратный ток диода нарастает гораздо быстрее, чем ток в прямом направлении (по крайней мере при невысоких напряжениях смещения), и вольт-амперная характеристика такого диода оказывается как бы «перевернутой» по отношению к вольт-амперной характеристике обычного диода (см. рис. 1.28). По этой причине этот диод и получил название обращенного.

В прямой ветви вольт-амперной характеристики обращенного диода иногда наблюдается слабо выраженный падающий участок (как, например, на рис. 1.28), но в отличие от туннельных диодов, его появление связано не с изменением перекрытия заполненных электронами состояний в зоне проводимости и незаполненных состояний в валентной зоне, а с уменьшением прозрачности потенциального барьера.

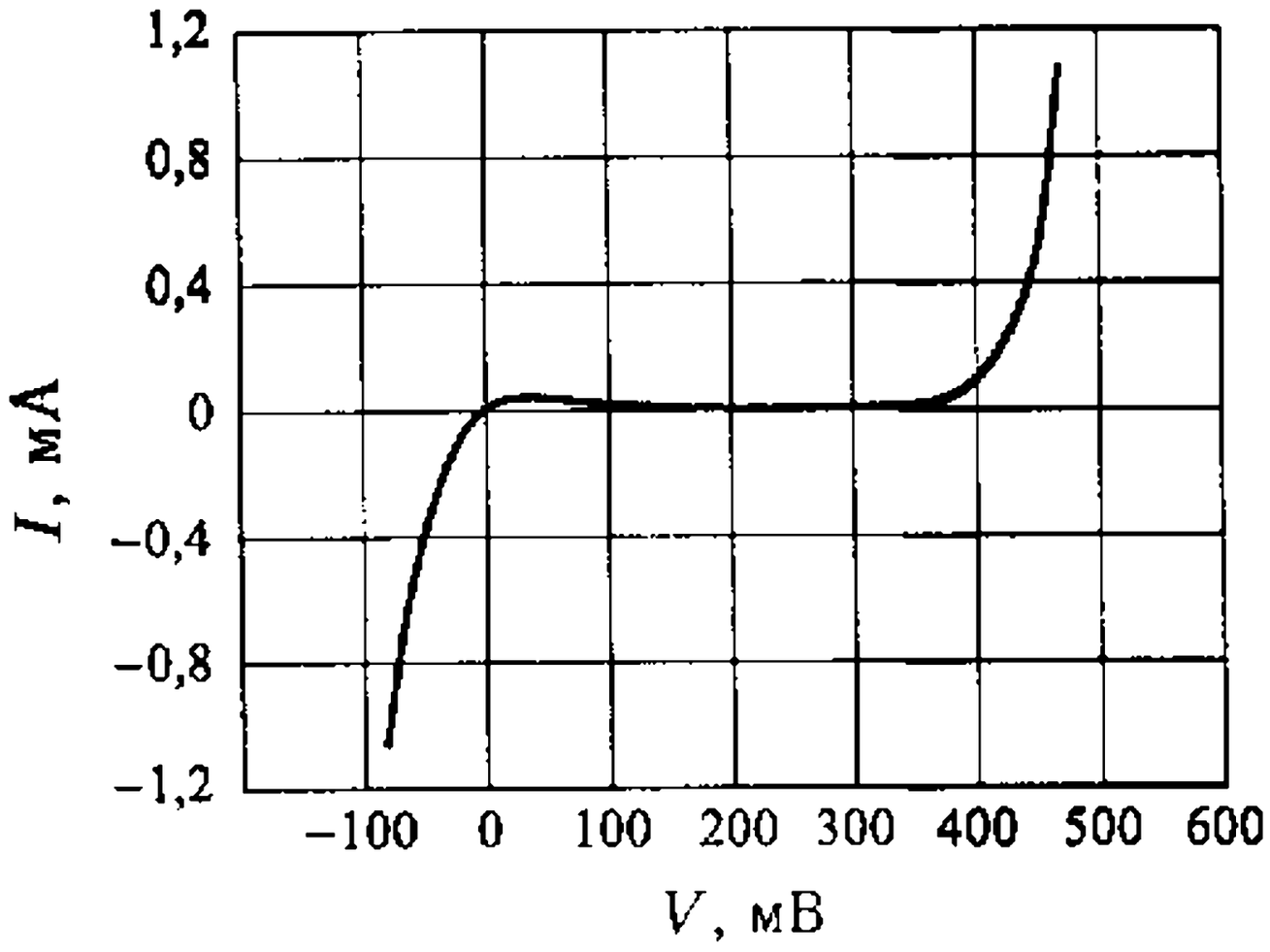


Рис. 1.28. Вольт-амперная характеристика германиевого обращенного диода ГИ401Б

Обращенные диоды находят применение в схемах детектирования СВЧ сигналов, в качестве смесителей и умножителей частоты. Все эти применения основаны на использовании сильной нелинейности вольт-амперной характеристики обращенного диода. Параметром, характеризующим диод как нелинейный

элемент, является величина

$$\gamma = \left.\frac{d^2J/dV^2}{dJ/dV}\right|_{V=0}. \qquad (1.100)$$

Например, при детектировании малых сигналов чувствительность детектора по току (отношение величины выпрямленного тока к мощности поглощенного диодом сигнала) равна $\gamma/2$. Для обычных диодов, вольт-амперная характеристика которых описывается формулами (1.30), (1.39), значение γ не может превышать величины $q/kT \approx 40$ В$^{-1}$. Для обращенных диодов, подбирая соответствующим образом уровни легирования p- и n-областей, удается достигнуть значения $\gamma \approx 70$ В$^{-1}$, причем величина γ к тому же слабо зависит от температуры [51]. Считая, что туннельный ток в p–n-переходе описывается формулой (1.82), можно показать, что при изменении приведенной концентрации примесей N^* величина γ изменяется как $\text{const} + (N^*)^{-1/2}$. Поэтому оптимальная концентрация примесей в p- и n-областях обращенных диодов должна быть невысокой, но такой, чтобы при небольших напряжениях смещения туннельный ток уже преобладал над всеми другими компонентами тока. Для германиевых диодов это условие выполняется при $N^* \sim 3 \cdot 10^{18}$ см$^{-3}$.

При использовании в системах детектирования слабых СВЧ сигналов важным достоинством обращенных диодов является возможность получения при нулевом смещении на диоде невысокого дифференциального сопротивления, значение которого (~100 Ом) согласовано с волновым сопротивлением волноводного тракта.

При использовании обычных диодов для согласования с волноводным трактом через диод приходится пропускать прямой ток, что неизбежно приводит к появлению дополнительных шумов и уменьшению отношения сигнал/шум. [1]

Слабая зависимость характеристик обращенных диодов от температуры, высокая нелинейность вольт-амперной характеристик, легкость согласования с волноводным трактом объясняют, почему обращенные диоды широко используются для детектирования СВЧ сигналов.

---

[1]) Так, в доплеровских радарных системах с приемником СВЧ сигнала на основе обращенных диодов мощность низкочастотного шума оказывается более, чем в 100 раз меньше, чем в приемниках на основе точечных кремниевых диодов [52].

1.5. Диоды с барьером Шоттки

В предыдущих разделах мы изучали работу полупроводниковых приборов на основе p–n-переходов, созданных в одном и том же полупроводнике. В этом и следующем разделах мы рассмотрим физические основы работы приборов, созданных на основе контакта двух различных материалов: металла и полупроводника или двух полупроводников. Мы увидим, что такие структуры позволяют существенно расширить функциональные возможности приборов, в работе которых используются специфические свойства структур с потенциальным барьером.

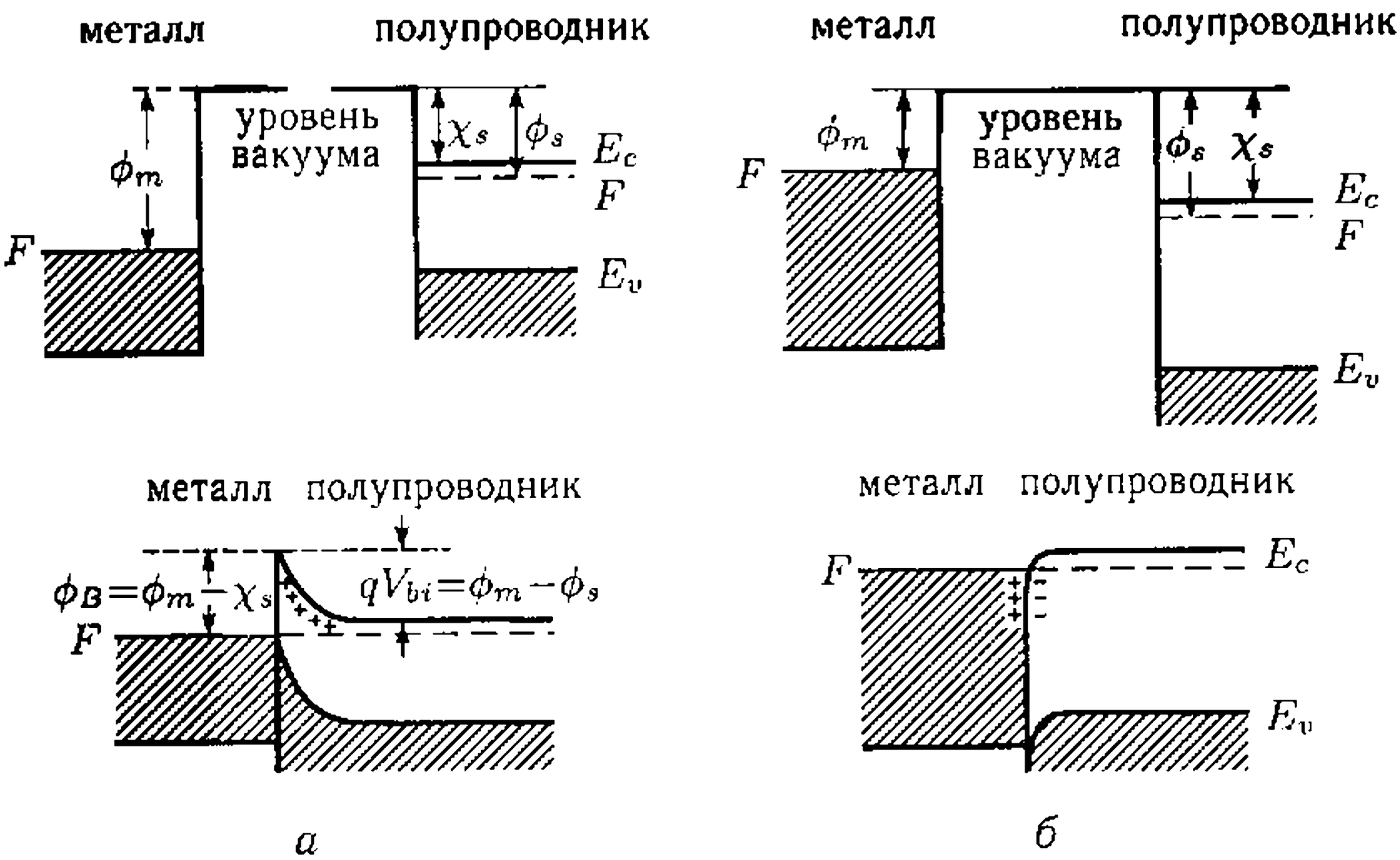
Основным недостатком приборов, в которых используется явление инжекции неосновных носителей заряда, как мы покажем в п. 1.7.3, является их не слишком высокое быстродействие, связанное с конечным временем жизни неравновесных носителей. Этого недостатка нет в полупроводниковых приборах, использующих в своей работе *только основные* носители заряда — туннельных диодах, которые мы рассмотрели в п. 1.4, и диодах с барьером металл–полупроводник (барьером Шоттки).

Нелинейные электрические свойства контакта металла с природными полупроводниками (прежде всего, галенитом PbS) были обнаружены Ф. Брауном еще в 1874 году [53]. На основе этого явления Браун и независимо Бозе в самом конце XIX века разработали полупроводниковые точечные детекторы, которые стали основными детекторами для приема радиоволн. Пикард, который (по некоторым данным) изучил выпрямляющие свойства контакта металла примерно с 30 тысячами различных веществ, установил, что наилучшими характеристиками обладают точечные контакты с кремнием и получил в 1906 г. патент на эту тему. Описываемые детекторы, названные *кристаллическими* (в отличие от «жидких» детекторов, использовавших нелинейные свойства контакта металл–электролит) до конца 20-х годов широко применялись в детекторных приемниках, пока не были вытеснены радиолампами.

В 30-е годы был начат промышленный выпуск сильноточных меднозакисных и селеновых выпрямителей на основе контактов металл–полупроводник к Cu_2O и Se (как выяснилось позже, на самом деле работа селеновых выпрямителей основана на образовании гетероперехода p-Se–n-CdSe). Новый интерес к кристаллическим детекторам возник в начале 40-х годов, когда было осознано, по своим характеристикам эти приборы существенно превосходят электровакуумные приборы при использовании их в качестве детекторов СВЧ сигналов в радиолокации. Работы В. Шоттки по изучению контактов металл–полупроводник и предложенная им в 1938 г. модель потенциального барьера на границе полупроводника и металла [54] имели столь большое значение, что такие барьеры стали называть барьерами Шоттки.

1.5.1. Энергетическая диаграмма контакта металл–полупроводник. Рассмотрим физическую модель контакта металл–полупроводник. Анализ целесообразно начать с идеализированной модели, предложенной в 1938 г. в работах Шоттки [54] и Мотта [55].

Как следует из расчетов зонной структуры полупроводников, уровни энергий, соответствующие краям зоны проводимости E_c и валентной зоны E_v, лежат на несколько эВ ниже уровня вакуума (см. рис. 1.29). Уровень вакуума — это та минимальная энергия электрона, начиная с которой он может выходить за пределы кристалла. В металле зона проводимости частично заполнена электронами и уровень Ферми F лежит в разрешенной зоне. Расстояние между уровнем Ферми и уровнем вакуума, которое мы будем обозначать ϕ_m, называется *работой выхода*. Наиболее важной характеристикой полупроводника является *электронное сродство́* χ_s, равное разности энергий между краем зоны проводимости E_c и уровнем вакуума; работа выхода ϕ_s также является характеристикой полупроводника, но эта величина не является постоянной, а зависит от его легирования.



а *б*

Рис. 1.29. Схема образования и энергетическая диаграмма контакта металл–полупроводник n-типа для случаев $\phi_m > \phi_s$ (*а*) и $\phi_m < \phi_s$ (*б*)

Мотт [55] первым высказал идею о том, что возникающий на границе металла с полупроводником потенциальный барьер обусловлен разностью работ выхода этих материалов. После приведения в контакт полупроводника и металла положение уровня

Ферми во всей системе должно стать одинаковым, для этого электроны переходят из одного материала в другой, и на границе возникает потенциальный барьер. Шоттки [54] предположил, что электрическое поле, возникающее в барьере, создается заряженными примесями, остающимися в приконтактной области полупроводника (обедненном слое) после ухода из нее электронов. В случае, когда работа выхода металла превышает работу выхода полупроводника ($\phi_m > \phi_s$, см. рис. 1.29*а*), на границе металл–полупроводник n-типа возникает потенциальный барьер, препятствующий движению электронов. Этот барьер и называется *барьером Шоттки*. В случае $\phi_m < \phi_s$, напротив, изгиб зон таков, что вблизи контакта в полупроводнике n-типа образуется обогащенный слой и никаких препятствий движению электронов нет. Заметим, что энергия, необходимая для преодоления барьера Шоттки при переходе электрона из металла в полупроводник ($\phi_B = \phi_m - \chi_s$) и называемая *высотой барьера Шоттки*, отличается от энергии, которую должен преодолеть электрон при переходе из зоны проводимости полупроводника в металл ($qV_{bi} = \phi_m - \phi_s$).

Таблица 1.1. Высота барьеров Шоттки некоторых металлов и их силицидов на n-Si [14, 56–58]. *Указан разброс данных для контактов, полученных разными способами.* Жирным шрифтом выделены данные, полученные на сколотых в вакууме поверхностях

Металл	ϕ_B, эВ	Металл	ϕ_B, эВ	Металл	ϕ_B, эВ
Au	0,52–**0,81**	Al	0,50–**0,77**	Pt	0,90
Ag	0,56–**0,79**	Cr	0,57–0,61	Cu	0,58–**0,79**
Hf	0,58	Mo	0,42–0,68	Ni	0,51–**0,70**
Pb	0,41–**0,79**	Ti	0,45–0,62	W	0,45–0,67
Pd	0,71	Sn	0,58		
PtSi	0,84; 0,85	Pd_2Si	0,72–0,75	$TiSi_2$	0,60
WSi_2	0,65; 0,86	$MoSi_2$	0,55	$TaSi_2$	0,59
Ni_2Si	0,7–0,75	NiSi	0,66–0,75	$NiSi_2$	0,7
IrSi	0,93	CoSi	0,68	$CoSi_2$	0,64
RhSi	0,69	$ZrSi_2$	0,55	HfSi	0,53
$CrSi_2$	0,57	MnSi	0,76		

Качественно понятно, что если между полупроводником n-типа и металлом создать разность потенциалов, то в случае $\phi_m > \phi_s$ напряжение будет падать в основном на обедненном слое,

и, поскольку высота барьера для электронов, движущихся из полупроводника в металл, зависит от приложенного напряжения, то можно ожидать, что такой контакт металл–полупроводник, как и обычный p–n-переход, будет обладать выпрямляющими свойствами. В случае $\phi_m < \phi_s$ можно ожидать, что контакт к полупроводнику n-типа будет невыпрямляющим (см. рис. 1.29 *б*).

В соответствии с обсуждаемой моделью, для контакта металла с полупроводником p-типа барьер Шоттки (барьер для движения дырок) должен возникать при $\phi_m < \phi_s$, а невыпрямляющий контакт — при $\phi_m > \phi_s$.

Значения высоты барьера Шоттки для контакта ряда металлов и их силицидов с n-Si приведены в табл. 1; данные для других полупроводников можно найти в [14, 56, 57]. Анализ этих данных показывает, что в действительности рассмотренный выше подход к определению ϕ_B слишком упрощен и часто предсказывает неверные результаты. Так, из табл. 2 следует, что барьер Шоттки на опыте образуется даже с металлами, для которых $\phi_m < \chi_s$ и для которых следовало бы ожидать возникновения невыпрямляющего (омического) контакта.

Таблица 1.2. Сравнение расчетных и экспериментальных значений высоты барьера Шоттки для контакта Al и Au с различными полупроводниками [57]

Полупроводник	Металл	$\phi_{B\,\text{эксп}}$, эВ	$\phi_{B\,\text{расч}}$, эВ
n-Si	Au	0,81	0,57
	Al	0,70–0,77	−0,26 (омич.)
n-Ge	Au	0,45	0,45
	Al	0,48	−0,38 (омич.)
n-GaAs	Au	0,90	0,51
	Al	0,80	−0,33 (омич.)
n-GaP	Au	1,30	0,55
	Al	1,05	−0,29 (омич.)

Измерения показали, что выбор металла влияет на высоту барьера Шоттки (см. рис. 1.30), однако зависимость ϕ_B от ϕ_m оказывается в 3–10 раз слабее, чем предсказывает рассмотренная выше модель ($\phi_B = \phi_m - \chi_s$).

Объяснение слабой зависимости высоты барьера Шоттки от ϕ_m было дано Бардиным [59], который предположил, что в формировании барьера на контакте металл–полупроводник важную роль играют *поверхностные состояния*. Эти состояния, как впервые показал И. Е. Тамм [60], образуются в полупроводнике

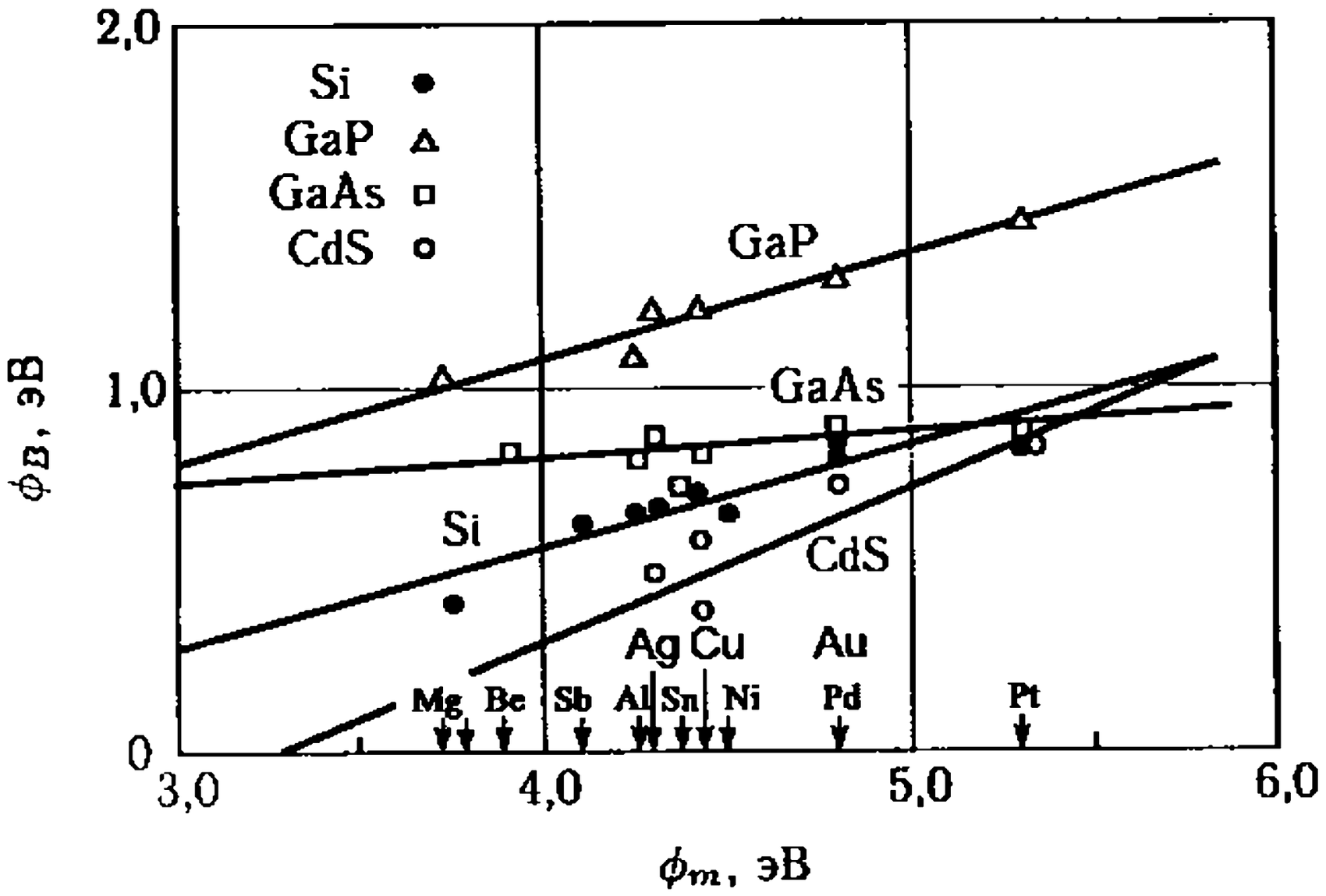


Рис. 1.30. Зависимость высоты энергетического барьера ϕ_B от работы выхода металла ϕ_m в контактах металлов с различными полупроводниками n-типа проводимости [14]

вблизи его поверхности. [1]) Шокли [63], развивая идеи Тамма, пришел к выводу, что поверхностные состояния могут образовывать зону, которая для обеспечения электронейтральности поверхности должна быть заполнена наполовину. Таким образом, существует некоторая энергия, называемая *уровнем нейтральности*, поверхностные состояния ниже которой должны быть заполнены, чтобы поверхность была электрически нейтральной. Во многих полупроводниках уровень нейтральности располагается в запрещенной зоне на расстоянии $\approx E_g/3$ выше края валентной зоны. [2]) Плотность поверхностных состояний достаточно велика; так, в Si она составляет $\sim 5 \cdot 10^{14}$ эВ$^{-1}$см$^{-2}$ для сколотой поверхности и $\sim 10^{13}$ эВ$^{-1}$см$^{-2}$ для травленой поверхности [56]. Поэтому поверхностные состояния, принимая на себя или отдавая заряд, эффективно компенсируют разность работ выхода полупроводника и металла и стабилизируют высоту

---

[1]) На возможность участия таммовских состояний в формировании барьера на контакте металл–полупроводник первым обратил внимание Мотт [55]. В эксперименте о важной роли таких состояний говорил тот факт, что барьер возникал даже на контакте двух *одинаковых* полупроводников (Ge–Ge [61]). В настоящее время общепризнанно, что барьеры, возникающие из-за поверхностных состояний, определяют свойства не только контактов металл–полупроводник, но и свойства гетеропереходов и поликристаллических полупроводников и, в частности, поликристаллического кремния, широко используемого при создании интегральных схем [62].

[2]) Исключением из этого правила является InAs, в котором уровень нейтральности располагается чуть выше края зоны проводимости.

потенциального барьера к полупроводнику n-типа на уровне $2E_g/3$, а к полупроводнику p-типа — на уровне $E_g/3$ (см. рис. 1.31).

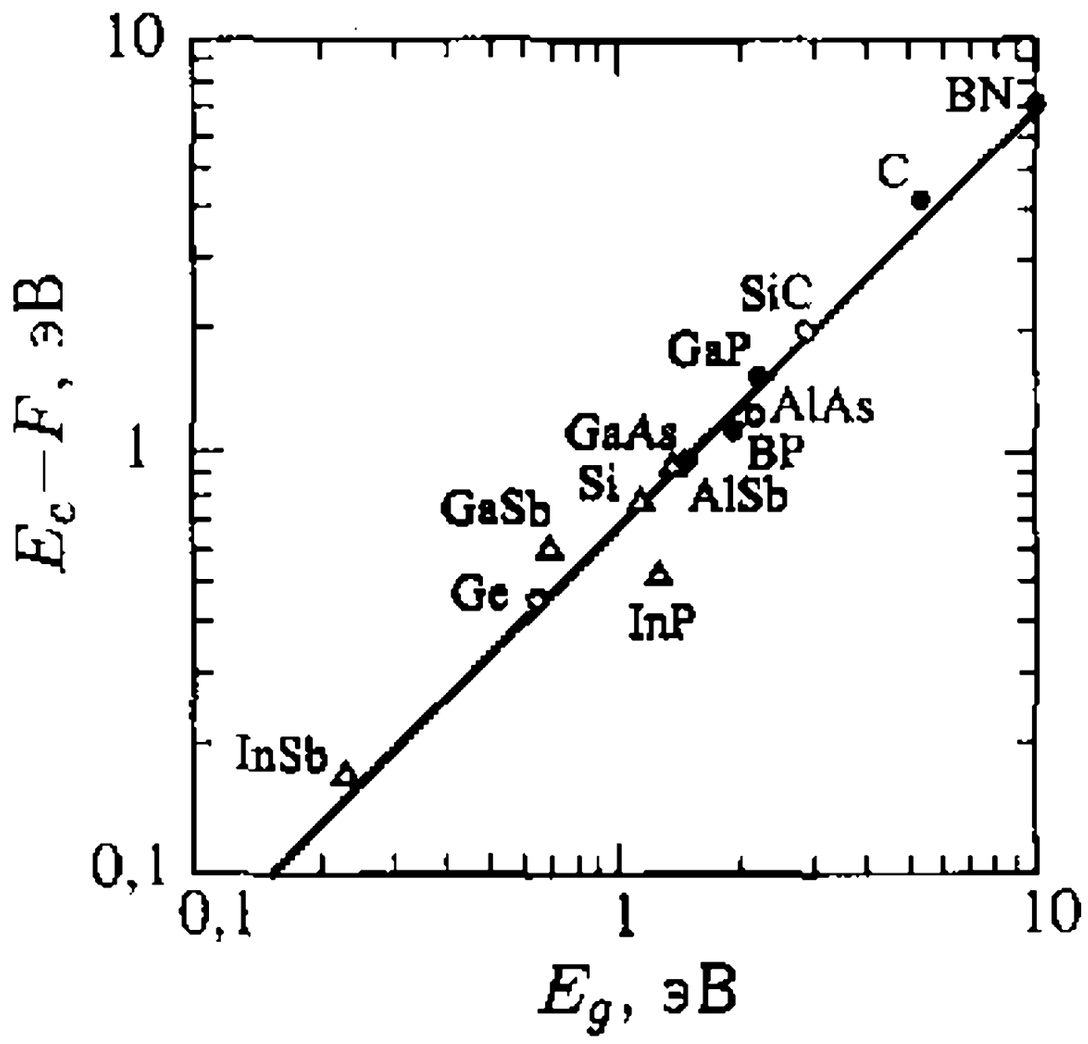


Рис. 1.31. Положение уровня Ферми, отсчитанного от края зоны проводимости, на границе раздела различных полупроводников с золотом [57]. Сплошная линия — зависимость $\phi_B = E_c - F = 2E_g/3$. Точки — данные для кристаллов n-типа (○) p-типа (●) и обоих типов проводимости (△)

Анализ данных по высоте потенциального барьера к различным полупроводникам и диэлектрикам [56, 57] показывает, что роль поверхностных состояний особенно велика в ковалентных полупроводниках и кристаллах с небольшой долей ионной связи; в материалах с большой долей ионной связи и ионных кристаллах высота барьера хорошо описывается рассмотренной выше простой моделью Мотта и Шоттки.

Одним из факторов, осложняющих экспериментальное определение высоты барьера Шоттки, является присутствие на поверхности полупроводника тонкой *окисной пленки*, которая остается там даже после самой тщательной подготовки поверхности перед нанесением металла. Так, например, толщина окисной пленки на свежетравленной в HF поверхности кремния составляет 10–20 Å, а при хранении образца на воздухе она возрастает до ~50 Å. Считается, что эта пленка туннельно прозрачна для электронов и поэтому практически не мешает движению носителей, однако ее существование может заметно (на 0,1–0,2 эВ) понижать высоту барьера. Присутствие окисных пленок в барьерах Шоттки может вызывать деградацию (нестабильность) характеристик этих приборов; так, в барьерах, изготовленных на химически травленных поверхностях, наблюдается дрейф характеристик, который связывается с миграцией заряженных ионов

(прежде всего, ионов щелочных металлов) в слое окисла [56]. Для устранения влияния окисной пленки в современных диодах Шоттки барьер изготавливают из силицидов переходных металлов (см. с. 99).

В заключение следует добавить, что величина ϕ_B может также зависеть и от кристаллографической ориентации барьера. Так, высота барьера Al к кремнию составляет 0,74 эВ для ориентации <111> и 0,81 эВ для ориентации <100>. В полупроводниках $A^{III}B^{V}$ на поверхностях (111)A и (111)B (см. подстрочное замечание на с. 78) барьер также имеет различную высоту. Изменить высоту барьера Шоттки в небольших пределах можно легируя подбарьерный слой полупроводника с помощью ионной имплантации [56].

Для дальнейшего обсуждения характеристик барьера Шоттки нам надо знать распределение электрического поля и потенциала в контакте металл–полупроводник. Поскольку длина экранирования в металле чрезвычайно мала (обычно она составляет несколько ангстрем), то можно считать, что все электрическое поле сосредоточено в полупроводнике. Поэтому граничными условиями при решении уравнения Пуассона (1.5) в полупроводнике n-типа с концентрацией доноров N_d можно считать $\mathcal{E} = 0$ на краю обедненного слоя (при $x = W$, см. рис. 1.32) и значение потенциала $\phi = V_{bi} - V$ при $x = 0$ (здесь V — приложенное к контакту напряжение смещения, а V_{bi} — величина встроенного потенциала). В результате несложных вычислений находим:

$$\mathcal{E}(x) \approx \frac{4\pi q}{\varepsilon} N_d (x - W), \quad \phi(x) = \frac{2\pi q}{\varepsilon} N_d (x - W)^2, \qquad (1.101)$$

откуда следует, что толщина обедненного слоя в барьере Шоттки равна

$$W = \sqrt{\frac{\varepsilon}{2\pi q} \frac{(V_{bi} - V)}{N_d}}\,. \qquad (1.102)$$

1.5.2. Вольт-амперная характеристика барьера Шоттки. Перейдем теперь к расчету вольт-амперной характеристики барьера Шоттки. Очевидно, что ток через барьер определяется балансом двух токов (см. рис. 1.32): тока электронов, которые движутся из полупроводника в металл и которым надо преодолеть барьер высотой $q(V_{bi} - V)$, и тока электронов, которые движутся из металла в полупроводник и которым надо преодолеть барьер высотой ϕ_B.

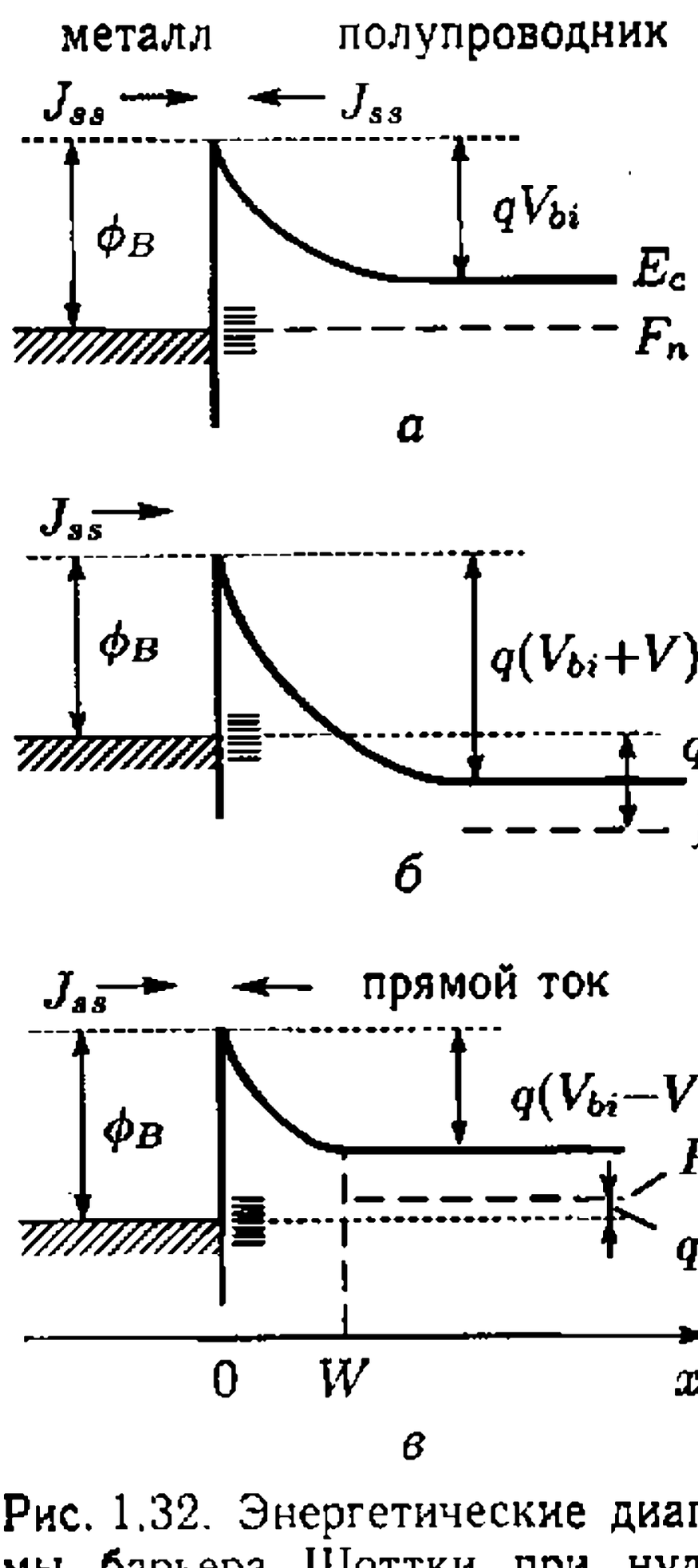


Рис. 1.32. Энергетические диаграммы барьера Шоттки при нулевом (*а*), обратном (*б*) и прямом (*в*) смещении

Рассчитаем ток электронов, текущий из полупроводника в металл, в рамках предложенной Бете в 1942 г. *модели термоэлектронной эмиссии* [64]. [1] Основным предположением этой модели является то, что электроны, движущиеся в полупроводнике в направлении контакта и имеющие достаточную для преодоления барьера энергию, не испытывают сильного рассеяния в области приконтактного изгиба зон. [2]

При нулевом напряжении смещения плотность тока электронов из полупроводника в металл равна заряду, переносимому через площадку единичной площади за единицу времени электронами, энергия которых превышает qV_{bi} (за начало отсчета энергии примем положение края зоны проводимости в нейтральной области полупроводника):

$$J = q\int\limits_{qV_{bi}}^{\infty} v_z\, dn(E) = q\int\limits_{qV_{bi}}^{\infty} v_z \rho(E) f(E)\, dE. \qquad (1.103)$$

Здесь $\rho = [(2m_n^*)^{3/2}/2\pi^2\hbar^3](E - E_c)^{1/2}$ — плотность состояний в зоне проводимости, а $f(E) \approx \exp[-(E - F_n)/kT]$ — функция распределения электронов по энергиям в невырожденном полупроводнике.

---

[1] Заметим, что одна из первых теорий выпрямления на контакте металл–полупроводник была предложена в 1939 г. Б. И. Давыдовым [65].

[2] Точнее, модель термоэлектронной эмиссии справедлива если длина свободного пробега электронов λ заметно больше характерного расстояния, на котором потенциальная энергия электрона меняется на величину kT: $\lambda > kT/q\mathcal{E}_{max}$ (условие Бете).

Для простоты будем рассматривать полупроводник с изотропным законом дисперсии. Энергия E представляет собой кинетическую энергию электрона в зоне проводимости: $E - E_c = m_n^* v^2/2$. Учитывая, что $dE = m_n^* v dv$, подынтегральное выражение в (1.103) можно преобразовать к следующему виду:

$$dn = \frac{m_n^{*3}}{4\pi^3\hbar^3} e^{-\frac{E_c - F_n}{kT}} e^{-\frac{m_n^* v^2}{2kT}} 4\pi v^2 dv. \qquad (1.104)$$

Разлагая вектор скорости на компоненты $v^2 = v_x^2 + v_y^2 + v_z^2$ и учитывая, что $4\pi v^2 dv = dv_x\, dv_y\, dv_z$, интеграл в (1.103) легко берется:

$$J = q\frac{m_n^{*3}}{4\pi^3\hbar^3} e^{-\frac{E_c - F_n}{kT}} \int_{v_{0z}}^{\infty} v_z e^{-\frac{m_n^* v_z^2}{2kT}} dv_z \int_{\infty}^{\infty}\int_{\infty}^{\infty} e^{-\frac{m_n^*(v_x^2+v_y^2)}{2kT}} dv_x\, dv_y =$$

$$= \frac{q m_n^* k^2}{2\pi^2\hbar^3} T^2 e^{-\frac{m_n^* v_{0z}^2}{2kT}} e^{-\frac{E_c - F_n}{kT}}. \qquad (1.105)$$

Здесь через v_{0z} обозначена минимальная скорость электрона, начиная с которой он может преодолеть потенциальный барьер высотой qV_{bi}. Принимая во внимание, что $m_n^* v_{0z}^2/2 = qV_{bi}$ и $(E_c - F_n) + qV_{bi} = \phi_B$, окончательно получаем

$$J = \frac{q m_n^* k^2}{2\pi^2\hbar^3} T^2 \exp\left(-\frac{\phi_B}{kT}\right) = A^* T^2 \exp\left(-\frac{\phi_B}{kT}\right). \qquad (1.106)$$

Сомножитель A^* в этой формуле есть так называемая постоянная Ричардсона, численно равная $120(m_n^*/m_0)$ А/см^{2}К^2. Для многодолинных полупроводников с анизотропным законом дисперсии (например, Si) решение имеет качественно тот же вид за исключением того, что входящая в формулу (1.106) величина m_n^* представляет собой некую комбинацию эффективных масс, зависящую от ориентации границы металл–полупроводник относительно кристаллографических осей [14, 56].

Учитывая, что в состоянии термодинамического равновесия (нулевое напряжение смещения) ток электронов из полупроводника в металл точно компенсируется током из металла в полупроводник, а при приложении напряжения смещения V высота барьера для электронов, движущихся из полупроводника в металл, понижается на величину qV, приходим к следующему

выражению для плотности тока через барьер Шоттки:

$$J = J_{ss}\left[\exp\left(\frac{qV}{kT}\right) - 1\right], \quad J_{ss} = \frac{qm_n^* k^2}{2\pi^2\hbar^3} T^2 \exp\left(-\frac{\phi_B}{kT}\right). \quad (1.107)$$

Хотя полученная формула для вольт-амперной характеристики по своему виду аналогична формуле (1.30) в теории тонкого p–n-перехода, существует качественное различие этих двух моделей. В случае барьера Шоттки электрон, преодолев энергетический барьер, сразу же попадает в металлический контакт, в то время как в p–n-переходе носители, преодолевшие барьер, должны еще продиффундировать в глубь нейтральных областей. Поэтому становится понятным, почему плотность тока насыщения в барьере Шоттки J_{ss} оказывается намного (на 3–4 порядка) выше плотности тока насыщения в p–n-переходе, изготовленном из того же полупроводника, а при равных плотностях тока падение напряжения на диоде с барьером Шоттки на ~0,2 В меньше, чем на p–n-переходе.

Анализ температурной зависимости вольт-амперных характеристик с помощью формулы (1.107) является одним из основных методов экспериментального определения высоты барьера на контакте металл–полупроводник. Другие методы нахождения ϕ_B основаны на определении красной границы в спектрах фотоответа, связанного с фотоэмиссией электронов из металла в полупроводник, и изучении вольт-фарадных характеристик (см. п. 1.7.1) барьеров Шоттки [57].

На самом деле в величину тока через барьер Шоттки небольшой вклад дают и *неосновные носители* (дырки), инжектируемые из металла в полупроводник n-типа. Доля дырочного тока при типичных уровнях легирования полупроводника и невысокой плотности тока составляет всего $\sim 10^{-4}$ от полного тока. Однако с ростом плотности тока эффективность инжекции дырок возрастает, поскольку дрейф дырок в возникающем в базе электрическом поле способствует их более быстрому продвижению в глубь квазинейтральной области. В этой ситуации конечная скорость рекомбинации инжектированных дырок может начать ограничивать быстродействие диодов Шоттки.

В качестве характеристики замедления быстродействия, связанного с инжекцией неосновных носителей, вводят величину *времени накопления*, которая определяется как отношение полного заряда инжектированных дырок к току через диод. Анализ показывает, что для уменьшения времени накопления следует повышать уровень легирования полупроводника и использовать металлы, которые образуют с ним менее высокий барьер [14].

При выводе уравнения (1.107) не был учтен ряд более тонких эффектов, проявляющихся в контактах металл–полупроводник:

1) эффект понижения высоты барьера за счет сил изображения (эффект Шоттки);
2) различные проявления эффекта туннелирования;
3) рассеяние носителей в обедненном слое.

Рассмотрим, как каждый из указанных эффектов влияет на вольт-амперные характеристики барьеров Шоттки.

Начнем с эффекта Шоттки. *Эффектом Шоттки* называется явление понижения высоты потенциального барьера на границе металл–вакуум или металл–полупроводник при приложении к контакту внешнего электрического поля. Рассмотрим электрон, находящийся вблизи металлического контакта (см. рис. 1.33*а*). Заряд электрона индуцирует появление в металле заряда-изображения противоположного знака, к которому он начинает притягиваться. Если граница раздела плоская, то заряд-изображение находится от электрона на расстоянии, равном удвоенному расстоянию между электроном и границей раздела. Рассчитав работу по перемещению электрона в поле притяжения из бесконечности в точку на расстоянии x от границы раздела, нетрудно показать, что понижение потенциальной энергии электрона составляет $q/4\epsilon x$ (рис. 1.33*б*). Если теперь в приконтактной области создать электрическое поле напряженностью $\mathcal{E}$, то энергетическая диаграмма контакта примет вид, показанный на рис. 1.33*в*. Находя максимум функции

$$(-q/4\epsilon x - \mathcal{E}x),$$

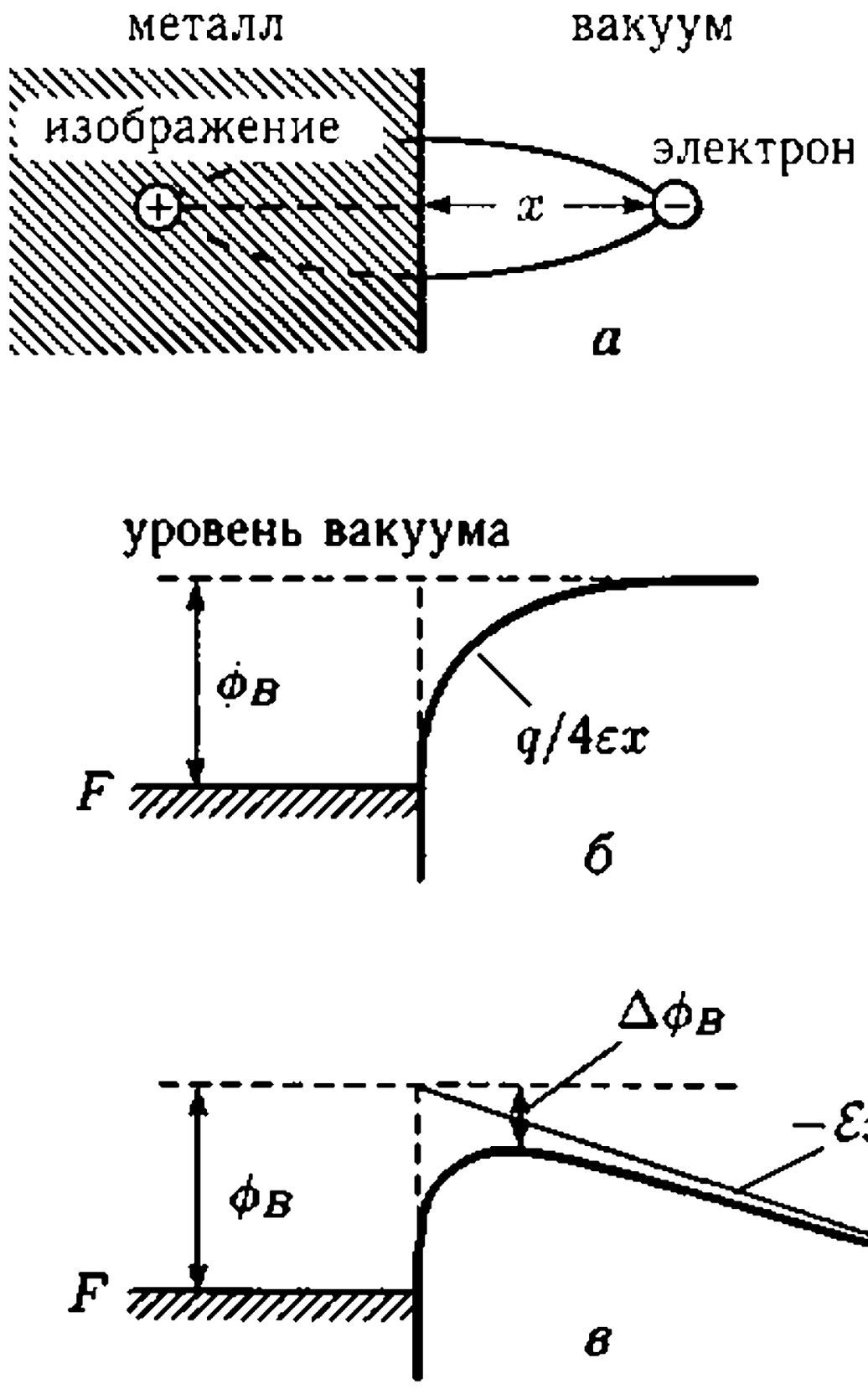


Рис. 1.33. Проявление сил изображения на границе металл–полупроводник: *а* — появление заряда-изображения; *б*, *в* — зависимость потенциальной энергии электрона от расстояния до границы раздела при нулевом и ненулевом значении напряженности электрического поля

получаем, что понижение высоты барьера равно

$$\Delta\phi_B = (q\mathcal{E}/\epsilon)^{1/2}.$$

В кремнии при $\mathcal{E} = 10^5$ В/см величина $\Delta\phi_B$ равна ≈35 мэВ, что превышает характерную энергию kT при комнатной температуре и составляет около 5% от ϕ_B. Доказательством понижения высоты барьера Шоттки за счет сил изображения являются наблюдаемые с увеличением обратного смещения сдвиг красной границы в спектрах фотоответа и изменение величины обратного тока по закону [57]

$$\ln J_{ss} = A + B(V_{bi} - V)^{1/4}.$$

Изменение высоты барьера при изменении напряжения смещения приводит к небольшому отклонению от «идеальности» наклона прямой ветви вольт-амперной характеристики: в эксперименте обычно наблюдается зависимость

$$J \sim \exp(qV/mkT),$$

в которой фактор идеальности m составляет 1,01–1,05.

Более подробные исследования показали, что наблюдаемое в эксперименте понижение высоты барьера Шоттки в электрическом поле оказывается сильнее, чем предсказывается рассмотренной выше теорией этого эффекта. Хейне [66], обсуждая проблему поверхностных состояний, показал, что волновая функция электронов металла может заметно проникать в полупроводник и, следовательно, дополнительное понижение высоты барьера может быть связано с эффектом *туннелирования электронов из металла в полупроводник*. Поскольку при этом в уравнение Пуассона кроме положительного заряда ионизованных доноров входит экспоненциально спадающий с расстоянием отрицательный заряд туннелирующих электронов, то распределение потенциала в барьере сильно изменяется по сравнению с решением, даваемым формулой (1.101). Оценки Хейне показывают, что глубина проникновения электронов в полупроводник может составлять 50–90 Å.

Рассмотрим теперь влияние эффектов рассеяния. Сравнение экспериментальных и рассчитанных по формуле (1.107) вольт-амперных характеристик показывает, что во многих случаях эксперимент хорошо описывается моделью термоэлектронной эмиссии. Как мы уже отмечали, так и должно быть, если выполняется условие Бете. Можно показать, что условие Бете эквивалентно условию

$$\mu\mathcal{E}_{max} > v_R,$$

где $\mathcal{E}_{max}$ — максимальная напряженность электрического поля в барьере Шоттки, а

$$v_R = \frac{A^* T^2}{qN_c}$$

— так называемая эффективная скорость рекомбинации, равная $\approx 1/4$ средней тепловой скорости электронов [56]. Из этого условия следует, что отклонение от модели термоэлектронной эмиссии можно ожидать в полупроводниках с низкой подвижностью или невысокой концентрацией доноров в базе диода ($< 10^{15}$ см$^{-3}$), в которых $\mathcal{E}_{max}$ не очень велико. Вольт-амперная характеристика барьера Шоттки в случае, когда влияние рассеяния существенно и необходимо рассматривать одновременную диффузию и дрейф в области барьера, была получена Кроуэллом и Зи [67]. Они показали, что учет рассеяния приводит к заметному уменьшению плотности тока насыщения J_{ss} в уравнении (1.107).

Термополевая и полевая эмиссия. Эффект туннелирования сквозь потенциальный барьер требует более детального рассмотрения. Как мы уже знаем, туннелирование проявляется в условиях, когда толщина энергетического барьера достаточно мала. Поэтому в барьерах Шоттки возможна ситуация, когда для электронов с энергией меньшей ϕ_B (E_m на рис. 1.34 *а*) толщина барьера оказывается настолько малой, что электроны начинают преодолевать его путем туннелирования. Такой механизм протекания тока через барьер Шоттки, в котором одновременно участвуют эффекты теплового возбуждения и туннелирования, получил название *термополевой эмиссии*.

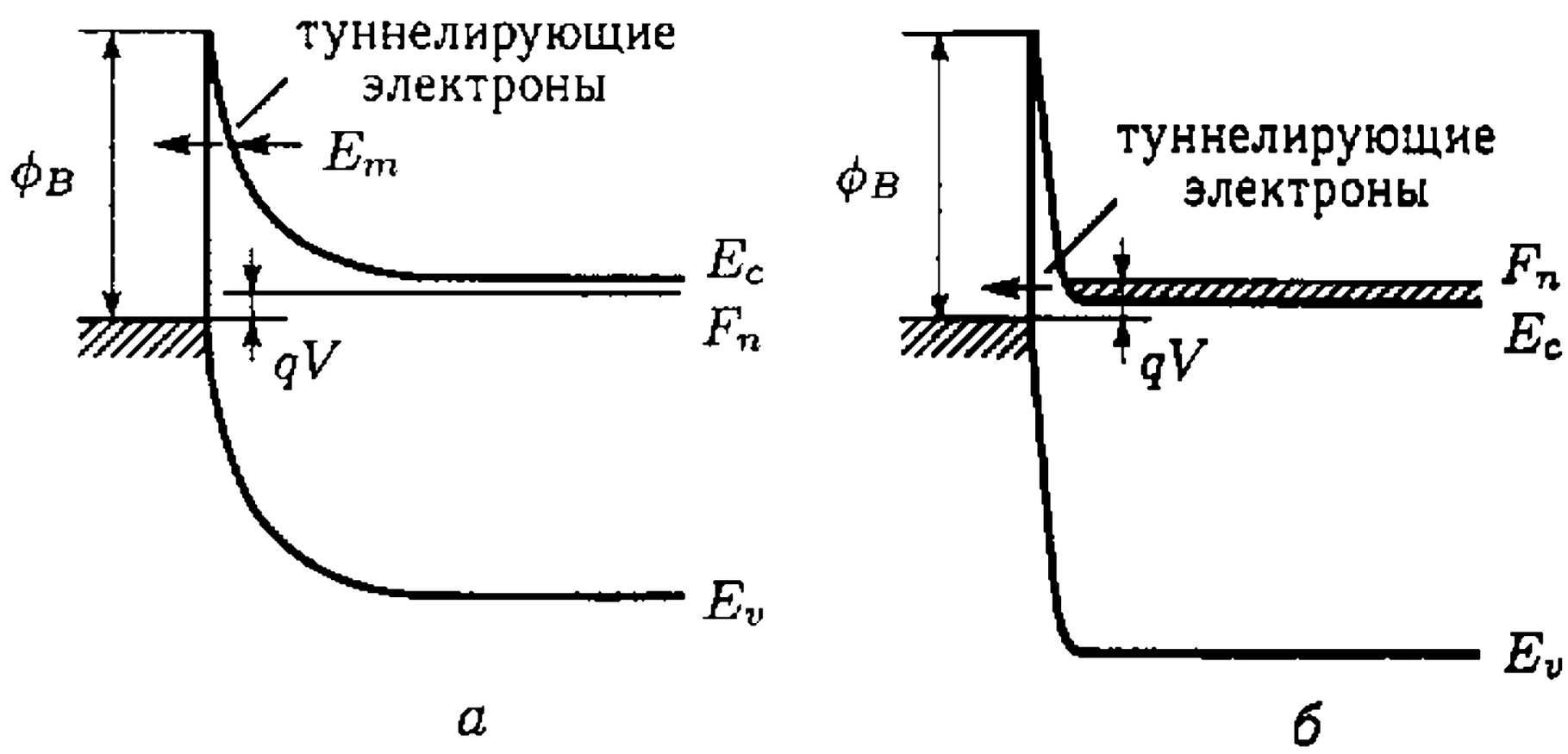


Рис. 1.34. Энергетические диаграммы барьеров Шоттки, в которых проявляется термополевая (*а*) и полевая (*б*) эмиссия

С ростом уровня легирования, когда полупроводник становится вырожденным, барьер может оказаться туннельно прозрачным и для носителей с энергией вблизи уровня Ферми (рис. 1.34 *б*). Механизм протекания тока, при котором носители проходят барьер Шоттки чисто туннельным способом (без тепловой активации), получил название *полевой эмиссии*.

Плотность тока, обусловленную термополевой и полевой эмиссией, можно вычислить, пользуясь формулами, аналогичными выведенными нами для туннельного эффекта (см. п. 1.3.2). Так, плотность тока эмиссии из полупроводника в металл равна

$$J \sim \int_0^{qV_{bi}} \rho_c(E)\, f_c(E)\, D(E)\, [1 - f_m(E)]\, dE, \qquad (1.108)$$

где ρ_c — плотность состояний в зоне проводимости, f_c и f_m — функции распределения Ферми–Дирака, описывающие заполнение электронами состояний в полупроводнике и металле, а D — прозрачность барьера. В этом уравнении мы предполагаем, что плотность состояний в металле не зависит от энергии. Полная плотность тока через барьер Шоттки складывается из вычисленных по формулам (1.107) и (1.108) плотности тока термоэлектронной эмиссии и тока полевой (термополевой) эмиссии.

Падовани и Стреттон [68] рассмотрели барьер Шоттки к вырожденному полупроводнику и показали, что при высоком уровне легирования и низкой температуре, когда в токе барьера Шоттки преобладает туннельная составляющая, плотность тока полевой эмиссии при достаточно больших прямых смещениях описывается формулой

$$J \approx J_{sf} \exp\left(\frac{qV}{E_{00}}\right), \qquad (1.109)$$

где $E_{00} = (q\hbar/2)\sqrt{4\pi N_d/m_n^* \epsilon}$,

$$J_{sf} = \frac{\pi A^* T}{k c_1 \sin(\pi k T c_1)} \exp\left(-\frac{\phi_B}{E_{00}}\right),$$
$$c_1 = \frac{1}{2E_{00}} \ln\left[\frac{4(\phi_B - qV)}{\Delta_n}\right], \qquad (1.110)$$

A^* — постоянная Ричардсона, а $\Delta_n = F_n - E_c$ — энергия Ферми электронов. При $N_d = 10^{18}$ см$^{-3}$ характеристическая энергия E_{00} в арсениде галлия составляет ≈20 мэВ.

При более высоких температурах ток через барьер Шоттки определяется термополевой эмиссией, для которой при достаточно больших прямых смещениях

$$J \sim \exp\left(\frac{qV}{E_0}\right), \quad E_0 = E_{00}\,\mathrm{cth}\left(\frac{E_{00}}{kT}\right), \tag{1.111}$$

а максимальный вклад в ток дают носители с энергией

$$E_m = \frac{\phi_B - qV + \Delta_n}{\mathrm{ch}^2(E_{00}/kT)}.$$

Из формулы (1.111) следует, что зависимость параметра $E_0(T)$ в модели термополевой эмиссии явно отличается от зависимости mkT, что может быть использовано для идентификации указанного механизма протекания тока.

Расчеты показывают, что преобладающий механизм токопереноса в барьерах Шоттки определяется соотношением между величинами kT и E_{00}. При низких температурах ($kT \lesssim E_{00}/3$) преобладает полевая эмиссия, в области промежуточных температур ($E_{00}/3 \lesssim kT \lesssim 4E_{00}$) преобладает термополевая эмиссия, а при совсем высоких температурах ($kT \gtrsim 4E_{00}$) — термоэлектронная эмиссия.

Качественное сходство характеристик диодов с барьером Шоттки и обычных p–n-переходов позволяет использовать их для решения одних и тех же практических задач (за исключением создания диодов с накоплением заряда (см. п. 1.7.4)). Более низкое падение напряжения в прямом направлении в диодах Шоттки по сравнению с p–n-переходами делает эти диоды незаменимыми для создания экономичных сильноточных выпрямителей. Однако то важное отличие, что работа барьеров Шоттки основана на переносе основных носителей, делает их особенно привлекательными для создания *сверхвысокочастотных приборов*. В настоящее время диоды Шоттки являются важными компонентами быстродействующих логических интегральных схем (см. п. 2.8). Кроме того, барьеры Шоттки часто используются в конструкциях других полупроводниковых приборов — полевых транзисторов с барьером Шоттки (см. п. 4.3), приборов с зарядовой связью (см. гл. 5), инжекционно-пролетных диодов (см. п. 6.3), фотодиодов (см. п. 7.1.4).

Для описания частотных характеристик СВЧ диодов Шоттки вводится так называемая *частота отсечки*, которая определяется как частота, на которой отношение выпрямленного тока к подводимой к диоду мощности уменьшается в 2 раза

по сравнению с его значением на низкой частоте. Эта частота рассчитывается по формуле

$$f_c = \frac{1}{2\pi C_j \sqrt{R_s R_j}},$$

где R_s — последовательное сопротивление диода, а C_j и R_j — его барьерная емкость и дифференциальное сопротивление [52]. Для уменьшения R_s СВЧ диоды Шоттки обычно имеют структуру металл–n–n^+, в которой толщина n-слоя делается по возможности более тонкой (порядка 1 мкм). Поскольку, как и в туннельных диодах, в последовательное сопротивление большой вклад дает сопротивление растекания (см. с. 80), уровень легирования n-слоя должен быть достаточно высоким. Частота отсечки характеризует быстродействие диодов Шоттки, накладываемое конструкцией диодов; принципиальное же (физическое) ограничение быстродействия этих диодов определяется временем рассасывания объемного заряда, которое равно времени максвелловской (диэлектрической) релаксации

$$\tau_M = \frac{\varepsilon}{4\pi\sigma},$$

где ε — диэлектрическая проницаемость полупроводника, а σ — его удельная электропроводность. В кремнии с типичным для СВЧ диодов Шоттки удельным сопротивлением n-области $\rho = 0{,}1$ Ом · см, $\tau_M \approx 10^{-13}$ с. В современных промышленных диодах Шоттки из Si и GaAs частота отсечки достигает сотен ГГц. Имеются данные об успешном использовании лабораторных образцов диодов Шоттки из GaAs для детектирования инфракрасного излучения с частотой 7,2 ТГц.

В заключение скажем несколько слов о конструкции диодов Шоттки. Как мы уже отмечали во введении к настоящему разделу, диоды со структурой металл–полупроводник были первым типом диодов и получались механическим прижимом заостренного металлического контакта к полупроводнику (*точечный диод*). Для улучшения характеристик таких диодов использовалась *формовка* (кратковременное пропускание большого тока через контакт). Важным преимуществом точечных диодов являлась малая площадь контакта и, следовательно, его малая емкость. К сожалению, характеристики точечных диодов сложно контролировать и для таких диодов характерны большие токи утечки по поверхности. Современные диоды Шоттки изготавливают по планарной технологии (см. п. 2.8.1) путем напыления металла на тщательно подготовленную (химическим или ионным травлением) поверхность полупроводника. Кроме чистых металлов

(Au, Al, Pt), при изготовлении диодов с барьером Шоттки на кремнии часто используются такие металлы как Rh, Pd, Pt, Ni, Zr и W, которые после нагревания при сравнительно невысоких температурах (300–650 °С) образуют *силициды* в результате твердофазной химической реакции с кремнием. Силициды имеют металлический тип проводимости и низкое удельное сопротивление (13–100 мкОм · см), а в силу того, что в результате химической реакции граница раздела кремний–силицид переходного металла сдвигается в объем кремния, все загрязнения остаются в слое силицида и создаваемые барьеры не имеют промежуточного слоя окисла и обладают хорошей воспроизводимостью. Для предотвращения утечки по поверхности и защиты структуры от преждевременного пробоя в местах концентрации электрического поля в планарной конструкции диодов Шоттки используются охранные кольца (см. рис. 1.15*б*).

Примерами отечественных диодов Шоттки могут служить выпрямительный кремниевый диод 2Д219, характеризуемый падением напряжения всего 0,3 В при токе 2 А, и арсенид-галлиевые высокочастотные диоды 3А110, 3А111, 3А527, 3А529, 3А530.

1.5.3. Омические контакты к полупроводникам. Согласно теоретической модели Мотта и Шоттки, рассмотренной нами в п. 1.5.1, для получения невыпрямляющего контакта к полупроводнику n-типа необходимо, чтобы работа выхода металла была меньше работы выхода полупроводника, а для полупроводника p-типа должно выполняться обратное соотношение. Однако в ковалентных полупроводниках и кристаллах с небольшой долей ионной связи поверхностные состояния играют столь большую роль, что в действительности энергетическая диаграмма контакта практически перестает зависеть от работы выхода металла и контакт любого металла с полупроводником как p-, так и n-типа проводимости вызывает появление барьера Шоттки. При изготовлении контактов к полупроводниковым приборам существование даже небольшого барьера на контакте приводит к целому ряду нежелательных последствий: дополнительному падению напряжения на контакте, дополнительной нелинейности вольт-амперной характеристики, инжекции носителей через контакт. В ряде случаев требования к контактам особенно высоки: так, в диодах с тонкой базой (см. с. 25) необходимо, чтобы контакт был невыпрямляющим как для электронов, так и для дырок (чтобы достигающие контакта инжектированные носители не накапливались в объеме, а уходили через контакт). Для решения задачи создания *омических контактов* к полупроводникам

оказывается очень плодотворным использование рассмотренного на с. 95 эффекта полевой эмиссии.

Идея этого омического контакта состоит в том, чтобы создавать металлический контакт не к самому полупроводнику (в котором концентрация носителей может быть невысокой), а к созданной на его поверхности сильно легированной области, концентрация примеси в которой такова, чтобы в барьере металл–сильно легированный полупроводник происходила полевая эмиссия носителей (рис. 1.35). В этом случае сопротивление контакта будет очень низким, и нелинейность его вольт-амперной характеристики будет практически незаметной. Чтобы на границе слабо- и сильнолегированных областей не возникало выпрямления, обе области должны иметь один и тот же тип проводимости (n^+-слой к полупроводнику n-типа и p^+-слой к полупроводнику p-типа).

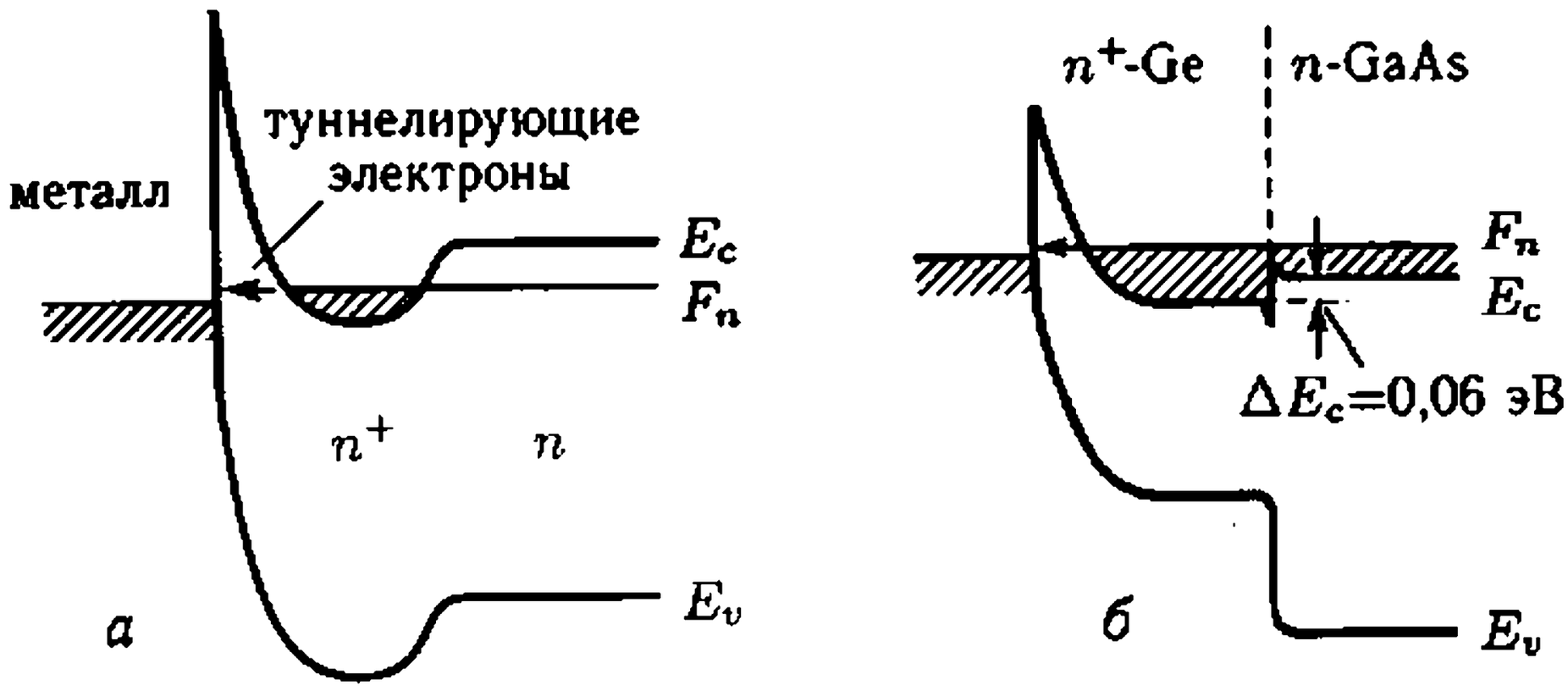


Рис. 1.35. Энергетические диаграммы, иллюстрирующие способы создания омических контактов к полупроводнику: *а* — с помощью сильно легированного слоя, *б* — с помощью гетероперехода

Как следует из формул (1.109) и (1.110), удельное контактное сопротивление при полевой эмиссии изменяется как

$$R_c \sim \exp(\phi_B / E_{00}).$$

Поскольку $E_{00} \sim N_d^{1/2}$, то сопротивление контакта будет тем меньше, чем выше концентрация примеси в сильно легированном слое. Лучшие омические контакты к Si и GaAs имеют удельное контактное сопротивление порядка 10^{-6} Ом · см^2.

Исследование *сплавных* омических контактов к GaAs показало [69], что величина R_c в таких контактах изменяется не по ожидаемому закону $R_c \sim \exp(A/\sqrt{N_d})$, а приблизительно обратно пропорционально концентрации примеси в сильно легированном слое. Объяснение этой зависимости основывается на предположении о неоднородности пограничного слоя. Предполагается, что в сплавных контактах на границе металла и полупроводника

формируются выступы небольшого радиуса. В области этих выступов электрическое поле барьера столь велико, что барьер оказывается туннельно прозрачным, а конечное сопротивление растекания в области выступа и приводит к зависимости

$$R_c \sim \frac{1}{\mu_n N_d}.$$

К сожалению, созданный описанным способом контакт является омическим только для основных носителей заряда; в приконтактной области в нем образуется потенциальный барьер, мешающий движению неосновных носителей (см. рис. 1.35). Чтобы охарактеризовать свойства контакта в отношении неосновных носителей, вводят понятие скорости поверхностной рекомбинации s на контакте (см. подстрочное замечание на с. 26). Типичное значение s для сплавных контактов из Sn+Sb к n-Ge составляет 10^3–10^4 см/с. Чтобы увеличить s, в приконтактной области часто создают высокую концентрацию рекомбинационных центров. Для этого можно использовать шлифовку поверхности, создавать нарушенный слой путем ионной имплантации или в процессе распыления металла.

Метод *вплавления* был одним из первых способов создания омических контактов, реализующих описанную идею. В этом методе на полупроводник кладется кусочек легкоплавкого металла, содержащий легирующую примесь. При нагревании металл плавится, растворяет поверхностный слой полупроводника и после охлаждения на поверхности полупроводника под каплей образуется тонкий рекристаллизованный слой с концентрацией примеси, близкой к пределу растворимости ($\sim 10^{20}$ см$^{-3}$). Этот слой и образует с каплей металла низкоомный омический контакт. Для изготовления сплавных омических контактов к германию используют сплавы Sn+5%Sb (к n-Ge) и In+1%Ga (к p-Ge). В случае кремния можно использовать алюминий (к p-Si) и сплав Au+1%Sb (к n-Si). Заметим, что метод вплавления широко используется не только для получения омических контактов, но и для создания полупроводниковых приборов — *сплавных диодов и транзисторов*. Кроме вплавления, для создания сильно легированной приконтактной области в полупроводнике можно использовать диффузию или ионную имплантацию. Нанесение металла на этот слой осуществляют путем напыления в вакууме, химическим или электролитическим осаждением (*несплавной* контакт).

В настоящее время большой интерес к себе привлекают *гетеропереходные омические контакты*. Идея такого контакта

состоит в том, чтобы на поверхности полупроводника создать тонкий (порядка 200 Å) слой другого полупроводника, высота барьера Шоттки к которому существенно ниже, чем к основному полупроводнику. Примером описываемого контакта может служить контакт n-GaAs–n^+-Ge–Au [70]. На его энергетической диаграмме (см. рис. 1.35*б*) разрыв края зоны проводимости ΔE_c в гетеропереходе GaAs–Ge невелик и при достаточно сильном легировании ($n_{GaAs} \sim 10^{18}$ см$^{-3}$, $n_{Ge} \sim 10^{20}$ см$^{-3}$) он не препятствует движению электронов. В то же время высота барьера Шоттки Au–Ge (0,5 эВ) почти вдвое ниже, чем высота барьера Au–GaAs (0,9 эВ), что и обеспечивает очень низкое контактное сопротивление $R_c < 10^{-7}$ Ом · см^2 [70]. Важной особенностью описываемого контакта является то, что в нем отсутствует барьер для движения неосновных носителей, то есть контакт оказывается омическим для обоих типов носителей. В настоящее время имеются данные об успешном создании гетеропереходных контактов к InP (слои InGaAs и InSb), к GaAs (слой InAs, варизонная структура InGaAs, короткопериодная сверхрешетка InGaAs/InAs), к InGaAs (слой InAs), к GaN (слой InGaN).

Изготовление контактов к полупроводниковым приборам, когда к контактам предъявляются жесткие требования по надежности, стабильности и коррозионной стойкости, оказывается непростой задачей. Например, при создании контактов к Si недостаточно просто напылить на полупроводник хорошо проводящий металл (Al, Au). Нижний слой, непосредственно контактирующий с полупроводником, должен обладать хорошей адгезией к кремнию. Этому условию удовлетворяют Al, Ti, Cr, V, Mo, W, силициды переходных металлов. Золото характеризуется плохой адгезией к кремнию. Из перечисленных материалов Al и Ti легко окисляются, Cr и Al взаимодействуют с SiO_2 (двуокись кремния используется для изоляции элементов в современных транзисторах и интегральных схемах), а все остальные металлы имеют сравнительно высокое удельное сопротивление. Поэтому контакты, изготовленные целиком из одного из этих металлов, оказываются не очень хорошими.

Чтобы получить низкоомный контакт, можно было бы напылить хорошо проводящий металл (например, золото) на тонкий слой металла с хорошей адгезией к кремнию и создать двухслойный контакт. Однако это не всегда возможно из-за химического взаимодействия контактирующих материалов (например, при взаимодействии Al и Au возникает хрупкая интерметаллическая фаза $AuAl_2$, известная под именем «пурпурной чумы», а при нагревании выше 200 °C Al начинает взаимодействовать с некоторыми силицидами и может проникать в кремний). Чтобы подавить взаимную диффузию и взаимодействие металлов, в контактах создают еще один (*барьерный*) слой из Pt, Ti, Mo, TiN, TaN. Поэтому самые лучшие контакты к Si имеют трехслойную структуру, например, Ti–Pt–Au.

Использование Al в качестве хорошо проводящего материала в контактах имеет два недостатка [58]. Во-первых, в алюминиевых контактах при высокой плотности тока (выше 10^5 А·см$^{-2}$) наблюдается явление *электромиграции* (электродиффузии), то есть направленное перемещение ионов алюминия под действием электрического поля, которое при продолжительной эксплуатации может приводить к нарушению целостности соединения (деградации контактов). Для уменьшения электромиграции к алюминию добавляют примесь меди. В последние годы в качестве хорошо проводящего материала в контактах все чаще используется медь (в интегральных схемах это способствует уменьшению задержки распространения сигнала). Во-вторых, алюминий во время термообработки при 450 °C немного растворяет поверхностный слой кремния, что может приводить к образованию на его поверхности глубоких каналов, заполняемых Al. В структурах с неглубокими p–n-переходами эти каналы могут проникать на всю глубину перехода и вызывать замыкание. Во избежание этих явлений в качестве контакта используется не чистый алюминий, а его сплав с добавкой $\sim 1\%$ Si.

Список материалов, используемых в качестве омических контактов к другим различным полупроводникам, можно найти в [14, 57].

1.6. Гетеропереходы и сверхрешетки

1.6.1. Гетеропереходы. *Гетеропереходом* принято называть контакт, образованный двумя различными полупроводниками. На возможность использования специфических свойств такого контакта для повышения эффективности инжекции эмиттеров биполярных транзисторов указывалось еще в патенте Шокли в 1948 г. [71]. В настоящее время гетеропереходы находят столь широкое применение при создании высокочастотных транзисторов (см. п. 2.4.2 и 4.1.6) и оптоэлектронных приборов (см. гл. 7), что работы нашего соотечественника Ж.И. Алфёрова и Г. Крёмера (США) в области исследования гетеропереходов были отмечены присуждением им совместно с изобретателем интегральной схемы Дж. Килби Нобелевской премии по физике в 2000 г.

Качественная модель формирования энергетической диаграммы гетероперехода была развита Андерсоном в 1960 г. на основании идей Мотта, обсуждавшихся нами в п. 1.5.1. [1]) Как мы уже отмечали, зонная структура любого полупроводника устроена так, что уровни энергий, соответствующие краям зоны

---

[1]) Заметим, что сходные представления об энергетической диаграмме гетероперехода использовались А.И. Губановым в цикле работ 1950–52 гг. по расчету вольт-амперных характеристик гетеропереходов [72], которые, в свою очередь, были стимулированы первыми исследованиями гетеропереходов в СССР [73].

проводимости и валентной зоны, лежат на несколько эВ ниже уровня вакуума.

Количественными параметрами, характеризующими полупроводник, являются:

1) ширина запрещенной зоны E_g,
2) *электронное сродство́* χ, равное разности энергий между краем зоны проводимости E_c и уровнем вакуума, и
3) *работа выхода* ϕ, равная разности энергий между уровнем Ферми F и уровнем вакуума (см. рис. 1.36 *а*).

Для GaAs и Ge (пары материалов, из которых были сделаны первые гетеропереходы) значения χ составляют, соответственно, 4,07 эВ и 4,13 эВ, а значения E_g — 1,42 эВ и 0,66 эВ.

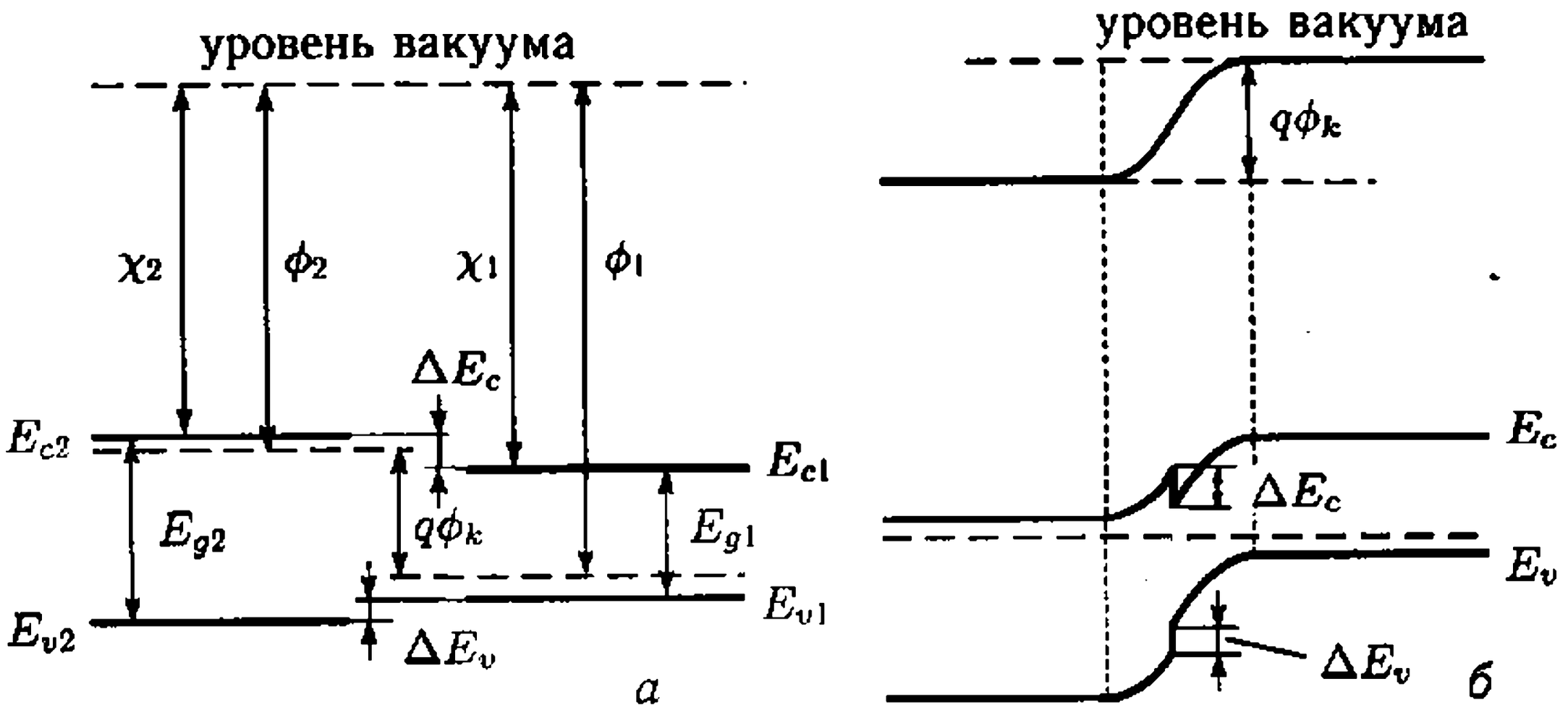


Рис. 1.36. Энергетическая диаграмма двух полупроводников до (*а*) и после (*б*) соприкосновения

В модели Андерсона предполагается, что на границе двух полупроводников не образуется никакого дипольного слоя. [1] Тогда в состоянии равновесия из требования единства уровня Ферми следует, что, как и в обычном p–n-переходе (см. п. 1.1), в гетеропереходе возникают электрическое поле и потенциальный барьер. Изменение энергии уровня вакуума в гетеропереходе, вызванное этим электрическим полем, характеризуется контактной разностью потенциалов ϕ_k (см. рис. 1.36 *б*). Поскольку контактирующие материалы имеют различные энергии электронного сродства́, на их границе возникают еще и дополнительный разрыв в энергии края зоны проводимости, равный

$$\Delta E_c = \chi_2 - \chi_1, \qquad (1.112)$$

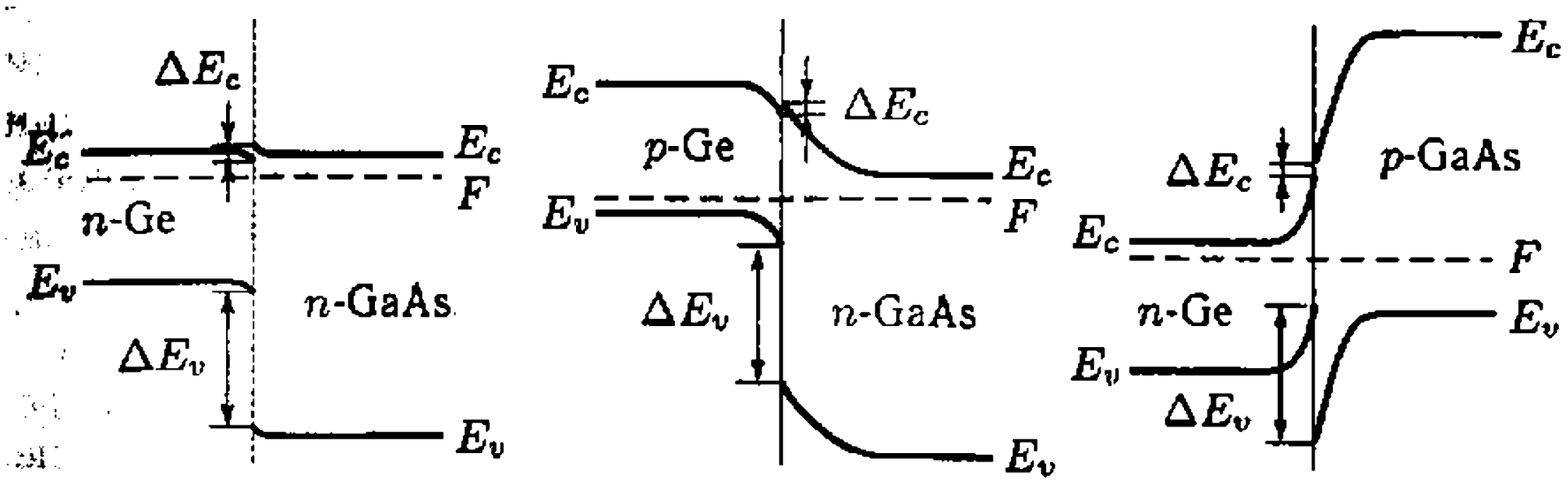
[1]) Возможные механизмы образования такого слоя обсуждаются на с. 106.

и соответствующий разрыв

$$\Delta E_v = E_{g2} - E_{g1} + \chi_2 - \chi_1 \qquad (1.113)$$

в энергии края валентной зоны.

Если в обоих полупроводниках, составляющих гетеропереход, тип проводимости одинаков, то такой гетеропереход называют *изотипным*, а если тип проводимости различен, то — *анизотипным*. Идеализированные энергетические диаграммы изотипных и анизотипных гетеропереходов в системе Ge–GaAs показаны на рис. 1.37. Видно, что разрывы зон на гетерогранице возникают даже в случае изотипных гетеропереходов.



а *б* *в*

Рис. 1.37. Идеализированные энергетические диаграммы гетеропереходов Ge–GaAs. *а* — изотипный n–n-гетеропереход, *б*, *в* — анизотипные p–n- и n–p-гетеропереходы

Следует иметь в виду, что при расчете распределения электрического поля в гетеропереходе необходимо учитывать различие диэлектрических проницаемостей в контактирующих полупроводниках. Если на гетерогранице нет электрически активных дефектов, то условием сшивания решений уравнения Пуассона в каждом из полупроводников должно быть равенство значений потенциала и *электрической индукции* $D = \varepsilon_1 \mathcal{E}_1 = \varepsilon_2 \mathcal{E}_2$ на границе раздела. Так, повторяя вычисления п. 1.1 с учетом указанных условий, для p–n-гетероперехода получаем

$$W = \sqrt{\frac{\varepsilon_1 \varepsilon_2 (\phi_k - V)(N_a + N_d)^2}{2\pi q (\varepsilon_1 N_d + \varepsilon_2 N_a) N_d N_a}}, \qquad (1.114)$$

где ε_1 — диэлектрическая проницаемость полупроводника p-типа, ε_2 — диэлектрическая проницаемость полупроводника n-типа, а ϕ_k — контактная разность потенциалов.

Для экспериментального определения величины разрывов зон в резких гетеропереходах обычно используют измерения

барьерной емкости и вольт-амперных характеристик гетеропереходов, хотя наиболее надежные сведения о разрывах зон получаются методом фотоэлектронной спектроскопии. Исследования показывают, что определяемые в эксперименте разрывы зон обычно отличаются от рассчитанных согласно формулам (1.112) и (1.113) и часто имеют значительный разброс. Так, например, для гетеропереходной пары Ge–GaAs вместо ожидаемого значения $\Delta E_c = 0,06$ эВ эксперимент дает значения 0,09–0,54 эВ [74]. Это, с одной стороны, может быть связано с трудностями определения энергии электронного сродства в полупроводниках (значения χ зависят от качества подготовки поверхности перед измерениями). С другой стороны, оказывается, что величины разрывов зон сильно зависят от технологии изготовления гетероперехода, кристаллографической ориентации границы раздела и даже последовательности выращивания слоев. Разброс значений разрывов зон для одной и той же гетеропереходной пары может достигать ~0,5 эВ.

Причины отклонения значений разрывов зон от предсказаний модели Андерсона могут быть разделены на физические и технологические. Эти причины подробно разобраны в обзоре Крёмера [74].

Главным недостатком модели Андерсона является пренебрежение дипольным слоем, который может возникать на границе двух полупроводников. Дело в том, что работа по переносу электрона из одного полупроводника в другой *не равна* разности их электростатических потенциалов. Реальный электрон чувствует производимое им возмущение в полупроводнике за счет кулоновского отталкивания и квантово-механических обменных эффектов, то есть кроме электростатической энергии взаимодействия необходимо учитывать еще и *корреляционную* энергию.

В упрощенной модели Френсли–Крёмера эта энергия выражалась через разность средних электроотрицательностей составляющих полупроводники атомов (понятно, что перераспределение заряда на границе двух полупроводников за счет разности электроотрицательностей будет приводить к образованию дипольного слоя).

Зависимость разрывов зон от технологии получения может быть связана с взаимопроникновением атомов через границу раздела, химическим взаимодействием на этой границе, а также неконтролируемым загрязнением границы раздела в процессе выращивания. Однако как было показано в работе [75], электростатика даже идеальной гетерограницы полупроводников с частично ионным типом связи такова, что для большинства ориентаций (кроме (110)) на гетерогранице должен накапливаться огромный электрический заряд, делающий ее энергетически неустойчивой. Реконструкция границы раздела, с помощью которой минимизируется этот заряд, происходит очень медленно, в результате чего конфигурация атомов в приграничном слое

оказывается зависящей от условий выращивания. В этом и состоит причина, почему скачок потенциала в дипольном слое на гетерогранице сильно зависит от технологии. Кроме того, на величину разрыва зон могут влиять и механические напряжения в гетероструктурах, которые в кристаллах без центра инверсии (GaAs, GaN и др.) вызывают появление дополнительных *пьезоэлектрических полей.*

Все сказанное выше относилось к случаю *резких* гетеропереходов, в которых толщина переходного слоя между двумя полупроводниками мала по сравнению с другими характерными длинами (в частности, толщиной области пространственного заряда). Существует и другая разновидность гетеропереходов, в которых в процессе изготовления толщина переходного слоя намеренно делается большой и получается так называемая *варизонная структура*. Примером варизонной структуры может служить структура, созданная на основе пары GaAs–$Al_xGa_{1-x}As$, ширина запрещенной зоны в широкозонной части которой ($Al_xGa_{1-x}As$) изменяется с координатой в соответствии с изменением параметра состава x твердого раствора. [1]) В этом случае на энергетических диаграммах вместо разрывов зон появляются участки, на которых наклоны краев зоны проводимости и валентной зоны различны (из-за зависимости E_g от координаты). Так, на рис. 2.13 в п. 2.4.2 показана энергетическая диаграмма варизонной структуры в системе Si–Ge, которая в настоящее время широко используется при создании гетеропереходных биполярных транзисторов.

Важным следствием энергетических диаграмм гетеропереходов является то, что из-за разрывов зон высота потенциальных барьеров, которые при инжекции приходится преодолевать электронам и дыркам, становится различной, что приводит к изменению соотношения токов инжектируемых электронов и дырок. Покажем это на примере энергетической диаграммы гетероперехода, изображенной на рис. 1.36*б*. Действительно, если небольшие пики, образуемые разрывами зон на энергетической диаграмме гетероперехода, не мешают инжекции носителей, то понятно, что сдвиг края зоны проводимости при переходе из широкозонной n-области в узкозонную p-область вниз на величину ΔE_c увеличивает ток инжекции электронов в $\exp(\Delta E_c/kT)$ раз, а сдвиг края валентной зоны при переходе из p-области в n-область вниз на ΔE_v уменьшает ток инжекции дырок

[1]) Зависимость $E_g(x)$ для этого твердого раствора представлена на рис. 7.24*б*.

в $\exp(\Delta E_v/kT)$ раз. Таким образом, отношение тока инжекции электронов к току инжекции дырок возрастает в

$$\exp\left(\frac{\Delta E_v - \Delta E_c}{kT}\right) = \exp\left(\frac{E_{g2} - E_{g1}}{kT}\right)$$

раз. Аналогичные рассуждения могут быть проведены и для других гетеропереходов, энергетические диаграммы которых показаны на рис. 1.37. Общим итогом этих рассуждений является то, что наличие разрывов в положении краев энергетических зон приводит к тому, что *в гетеропереходах имеет место практически односторонняя инжекция из широкозонного в узкозонный полупроводник.* Более того, если из-за разрыва в положении края зоны инжектируемые носители при переходе в узкозонный полупроводник увеличивают свою кинетическую энергию (как это имеет место для электронов на рис. 1.36), то при больших прямых смещениях концентрация носителей, инжектированных в узкозонный полупроводник, может даже превышать концентрацию свободных носителей в широкозонном слое (явление *«сверхинжекции»*). Эти свойства гетероперехода широко используется в полупроводниковых лазерах, некоторых конструкциях транзисторов [1]) и других приборах, принцип работы которых требует определенного направления инжекции носителей.

Оценим изменение отношения токов инжекции J_n/J_p при 300 К для широко используемой гетеропереходной пары $GaAs$–$Al_{0,3}Ga_{0,7}As$, для которой $\Delta E_c \approx 0,62\Delta E_g \approx 0,23$ эВ, $\Delta E_v \approx 0,38\Delta E_g \approx 0,14$ эВ. Из приведенной выше формулы получаем, что J_p/J_n должно измениться в $\exp(\Delta E_g/kT) \approx 2 \cdot 10^6$ раз. Столь значительный эффект дает основание говорить о действительно одностороннем характере инжекции в гетеропереходе.

В отношении сказанного следует сделать одно замечание. «Пички», появляющиеся на энергетических диаграммах многих гетеропереходов (см. рис. 1.36 и 1.37), могут затруднять протекание тока в структуре, и на самом деле вопрос об эффективности

---

[1]) Как мы покажем в гл. 2, для создания биполярных транзисторов со сверхбольшим коэффициентом усиления необходима односторонняя инжекция из эмиттера в базу, которую не удается получить только за счет сильного легирования эмиттера. Идея использовать для этой цели гетеропереход была предложена Шокли в 1948 г. [71], а в 1957 г. Крёмер [76] высказал аналогичную идею и в отношении варизонных структур. В настоящее время гетеропереходные транзисторы на основе резких гетеропереходов и варизонных структур выпускаются в промышленном масштабе (см. п. 2.4.2).

инжекции в узкозонный полупроводник может оказаться более сложным. В зависимости от уровня легирования (и, следовательно, от толщины образуемого «пичком» потенциального барьера), возможны различные ситуации, при которых этот «пичок» может либо образовывать барьер для движения носителей, либо преодолеваться за счет туннелирования. Мы не будем здесь рассматривать, к каким изменениям вольт-амперных характеристик могут привести эти особенности энергетической диаграммы; обсуждение этих вопросов можно найти в [57].

К сожалению, практически неизбежное для гетеропереходов различие параметров кристаллической решетки контактирующих материалов приводит к появлению на гетерогранице оборванных связей, дислокаций несоответствия и других дефектов, энергетические уровни которых выступают как центры рекомбинации. Влияние этих дефектов на вольт-амперные характеристики гетеропереходов аналогично влиянию глубоких уровней в области пространственного заряда на вольт-амперные характеристики обычных p–n-переходов (см. п. 1.2.2). По этой причине задача выбора материалов для создания гетеропереходов оказывается весьма непростой. Требуется не только, чтобы различие параметров решетки двух контактирующих полупроводников при температуре выращивания было невелико, но и чтобы были близки их температурные коэффициенты расширения (чтобы избежать появления дополнительных напряжений на гетерогранице при охлаждении структур после их выращивания). [1)]

В табл. 1.3 представлены свойства полупроводников, входящих в состав некоторых хорошо изученных гетеропереходных пар. К сожалению, рассогласование параметров решетки в гетеропереходах, построенных на основе индивидуальных соединений, достаточно велико. Поэтому более перспективным решением проблемы подбора полупроводников для гетеропереходных

[1)] Эксперимент показывает, что в тонких пленках полупроводника, выращенных на подложке из другого полупроводника, небольшое рассогласование параметров решетки этих материалов может «сдерживаться» (без образования дефектов) за счет упругих напряжений. Структуры с такими упруго напряженными слоями называют *напряженными* (*псевдоморфными*) структурами. При увеличении толщины наращиваемого слоя выше некоторой критической величины упругие напряжения снимаются путем образования сетки дислокаций, что резко ухудшает свойства этих структур. Интерес к псевдоморфным структурам связан с тем, что величина упругих напряжений является еще одним параметром, с помощью которого можно управлять энергетической диаграммой гетероперехода.

пар оказывается использование твердых растворов, согласованных по параметру решетки с индивидуальным соединением.

Таблица 1.3. Ширина запрещенной зоны, энергия электронного сродства и параметр решетки полупроводников, входящих в состав некоторых хорошо изученных гетеропереходных пар

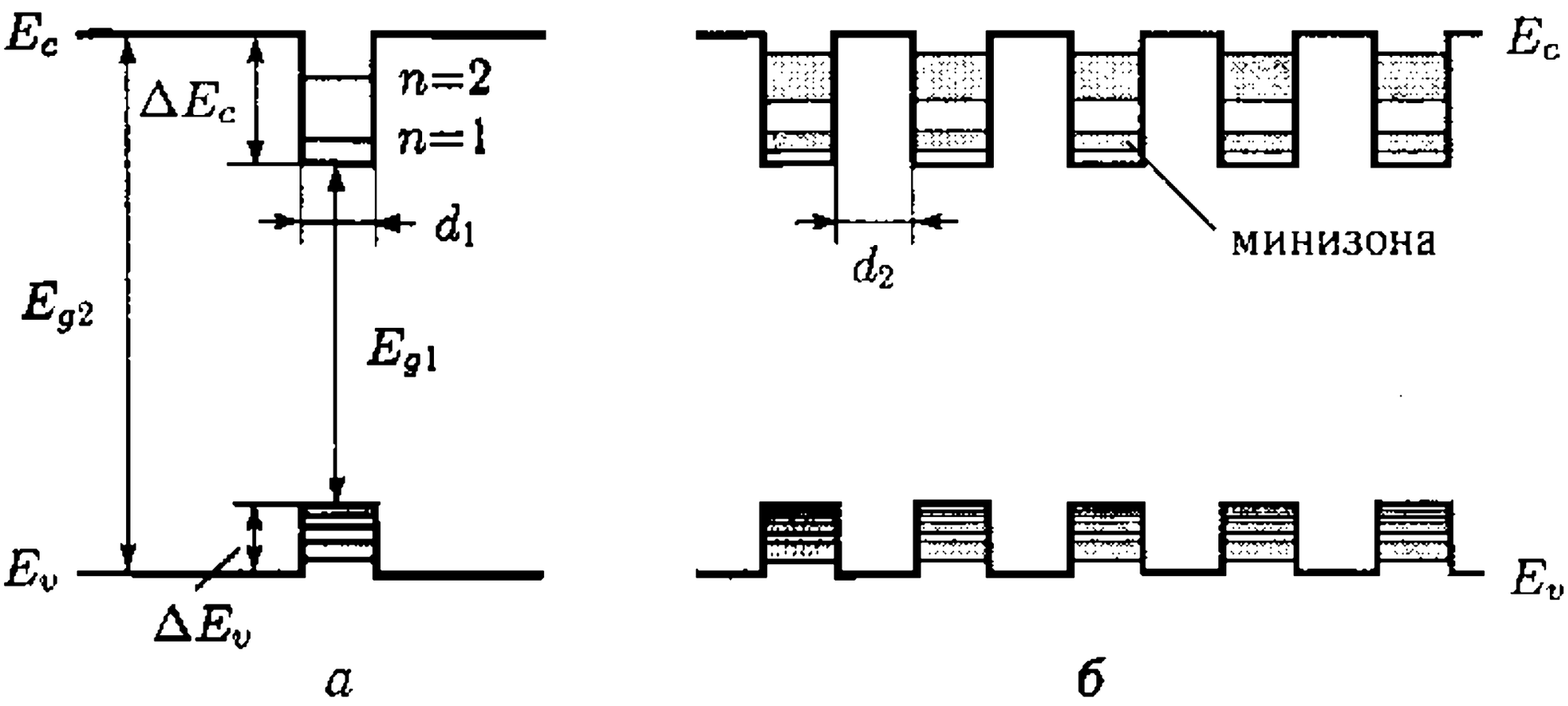
Пара	E_{g1}, эВ	E_{g2}, эВ	χ_1, эВ	χ_2, эВ	a_1, Å	a_2, Å
Ge–GaAs	0,66	1,42	4,13	4,07	5,658	5,654
Ge–ZnSe	0,66	2,70	4,13	4,09	5,658	5,667
GaAs–AlAs	1,42	2,15	4,07	3,5	5,654	5,661
GaAs–ZnSe	1,42	2,70	4,07	4,09	5,654	5,667
InAs–GaSb	0,36	0,68	4,9	4,06	6,058	6,095
GaAs–InAs	1,42	0,36	4,07	4,9	5,654	6,058
Ge–Si	0,66	1,12	4,13	4,01	5,658	5,431
$InP–In_{0,53}Ga_{0,47}As$	1,34	0,75	4,38	4,63[1)]	5,869	5,869
$InP–Al_{0,48}In_{0,52}As$	1,34	1,50	4,38	4,08[1)]	5,869	5,869
$InP–GaAs_{0,49}Sb_{0,51}$	1,34	0,72	4,38	4,20[1)]	5,869	5,869
$GaAs–In_{0,5}Ga_{0,5}P$	1,42	1,87	4,07	3,89[1)]	5,654	5,654

[1)] Оценено из имеющихся в литературе данных по величинам разрывов зон.

Четыре такие наиболее важные для практики пары представлены в последних строках табл. 3. Кроме этого, следует отметить важную пару с четверным твердым раствором $In_{1-x}Ga_xAs_yP_{1-y}$, который согласован по параметру решетки с InP при $x \approx 0,47y$, и четверной твердый раствор $Si_{1-x-y}Ge_xC_y$, который согласован по параметру решетки с Si. Наконец, в качестве пары можно выбрать два твердых раствора, согласованных по параметру решетки ($Pb_{1-x}Sn_xTe–PbTe_{1-y}Se_y$), или индивидуальное соединение и твердый раствор на его основе, если параметр решетки в твердом растворе не сильно зависит от параметра состава твердого раствора ($Al_xGa_{1-x}As–GaAs$, $Al_xGa_{1-x}Sb–GaSb$).

1.6.2. Квантовые ямы и сверхрешетки. В последнее время большое внимание как с фундаментальной, так и с прикладной точек зрения привлекают к себе новые полупроводниковые структуры с гетеропереходами: *одиночные квантовые ямы* и построенные из них периодические структуры — *сверхрешетки* [77]. Интерес к этим объектам обусловлен тем, что в очень тонких (20–100 Å) слоях полупроводника энергетический спектр носителей оказывается квантованным (дискретным)

для движения носителей в направлении перпендикулярно слою и непрерывным для движения в двух других направлениях, то есть характер движения электронов в этом слое по сути носит *двумерный* характер. Поместив тонкий слой из интересующего нас полупроводника между двумя областями из более широкозонного полупроводника (то есть создав для электронов и дырок своеобразную потенциальную яму, ограниченную с двух сторон барьерами ΔE_c и ΔE_v, см. рис. 1.38 *а*), такую квазидвумерную структуру удается реализовать практически.



а *б*

Рис. 1.38. Энергетическая диаграмма одиночной квантовой ямы (*а*) и сверхрешетки (*б*)

В случае одиночной прямоугольной потенциальной ямы с бесконечно высокими стенками из решения уравнения Шрёдингера следует, что для полупроводника с минимумом, расположенным в центре зоны Бриллюэна, и изотропным законом дисперсии энергия электрона в зоне проводимости может быть записана в виде

$$E_n(k_x, k_y) = E_c + \frac{\hbar^2}{2m_n^*}\left(\frac{n\pi}{d_1}\right)^2 + \frac{\hbar^2(k_x^2 + k_y^2)}{2m_n^*}, \qquad (1.115)$$

где d_1 — ширина квантовой ямы, k_x, k_y — компоненты волнового вектора в плоскости слоя, m_n^* — эффективная масса электрона, а $n \geqslant 1$ — номер уровня размерного квантования (см. рис. 1.38 *а*). В реальной квантовой яме высота стенок конечна, волновая функция немного выходит за пределы ямы и сдвиги энергии, связанные с размерным квантованием, оказываются немного меньше. Такое же квантование уровней энергии, зависящее от эффективных масс дырок, возникает и в валентной зоне для легких

и тяжелых дырок. [1]) Таким образом, в одиночной квантовой яме разрешенные зоны состоят из ряда *подзон*, описываемых различными значениями n, движение носителей в каждой из которых носит двумерный характер.

По аналогии с квантовыми ямами можно представить себе другие полупроводниковые структуры пониженной размерности, в которых движение электронов ограничено в двух и трех измерениях. Эти структуры называются, соответственно, *квантовыми нитями* и *квантовыми точками*. В квантовой нити движение носителей носит одномерный характер, а квантовая точка представляет собой аналог атома с полностью дискретным энергетическим спектром.

Используя выражения для закона дисперсии квантовой ямы (1.115), квантовой нити и квантовой точки, нетрудно найти выражения для плотности энергетических состояний [77]. Оказывается, что плотность состояний в низкоразмерных системах сильно отличается от таковой в объемных полупроводниках: в квантовых ямах она носит характер ступенек высотой $m^*/(\pi\hbar^2)$, возникающих при энергиях, отвечающих минимумам подзон E_n. В квантовых нитях она представляет собой сумму острых пиков, спадающих по закону

$$\rho(E) \sim (E - E_n)^{-1/2},$$

а в квантовых точках она представляет собой сумму δ-функций. Такое изменение плотности состояний отражается на всех физических свойствах низкоразмерных систем и, в частности, эти изменения позволяют существенно понизить порог возбуждения лазерной генерации в структурах пониженной размерности (см. п. 7.2.3).

Идея создания сверхрешеток возникла при поиске новых приборов, обладающих отрицательным дифференциальным сопротивлением. Начиная с 1928 г. в литературе широко обсуждалась возможность создания т. н. *блоховского осциллятора*. Суть этого простого прибора заключалась в том, что если к кристаллу, в котором рассеяние электронов невелико, приложить сильное электрическое поле, то можно создать условия, при которых ускоряемый электрическим полем электрон будет совершать периодическое движение, «отражаясь» от границ зоны Бриллюэна

---

[1]) Из-за сильной зависимости энергетического спектра квантовых ям от высоты потенциальных барьеров (ΔE_c, ΔE_v) изучение спектров оптического поглощения квантовых ям сейчас стало одним из наиболее точных методов *определения величин разрывов зон в гетеропереходах*.

и генерируя при этом высокочастотные колебания (см., например, [78]). К сожалению, простые оценки показывали, что в полупроводниках даже с самой высокой подвижностью рассеяние электрона происходит раньше, чем он достигает границы зоны Бриллюэна. Решением задачи создания блоховского осциллятора могло бы быть формирование искусственной длиннопериодной структуры, в которой размеры зоны Бриллюэна намного меньше, чем в исходном кристалле, и в которой можно было бы наблюдать блоховские осцилляции или хотя бы отрицательное дифференциальное сопротивление.

Л.В. Келдыш первым предложил создавать сверхрешетку в кристалле с помощью мощной ультразвуковой волны [79]. Позже были предложены похожие способы создания сверхрешеток с помощью стоячих световых волн, дифракционных решеток и другими способами (см. обзоры [80, 81]). Однако первой достаточно просто реализуемой конструкцией, нашедшей позже широкое применение, оказалась конструкция, предложенная Есаки и Цу [82]. Для реализации требуемой искусственной периодичности они предложили два способа:

1) использовать один и тот же полупроводник, но легировать его попеременно, создавая слои n- и p-типа (так называемые *nipi-сверхрешетки*), и

2) использовать чередующиеся слои двух различных полупроводников, в которых запрещенная зона одного материала перекрывает запрещенную зону другого (так называемые *композиционные сверхрешетки*, см. рис. 1.38 *б*).

Практически создать оба типа сверхрешеток оказалось возможным только после развития технологии молекулярно-лучевой эпитаксии (MBE) и газофазной эпитаксии из паров металлоорганических соединений (MOCVD). Оба метода позволяют осуществлять послойное эпитаксиальное наращивание атомных слоев заданного состава на монокристаллическую подложку. Метод молекулярно-лучевой эпитаксии обычно используется для получения опытных образцов гетероструктур, а метод MOCVD — для их массового производства. Особенностью сверхрешеток является возможность искусственно формировать их электронный спектр. Если при выращивании сверхрешетки расстояние между квантовыми ямами d_2 сделать небольшим, чтобы электроны могли туннелировать из одной ямы в другую через потенциальный барьер, образованный широкозонным полупроводником, то уровни размерного квантования, отвечающие движению электрона перпендикулярно стенкам ямы, размываются в так называемую *мини-зону*

(закрашенные участки на рис. 1.38 *б*). При этом ширина мини-зоны определяется перекрытием волновых функций электронов в соседних ямах, а расстояние между разными мини-зонами — шириной квантовой ямы d_1. Это позволяет, меняя толщины слоев в сверхрешетке, *направленно изменять электронный спектр полупроводника* и тем самым создавать новый искусственный полупроводниковый материал, оптимально подходящий для того или иного практического применения. Более подробно эти возможности рассмотрены в книгах [83, 84]. [1])

Одним из возможных практических применений одиночных квантовых ям и сверхрешеток является создание нового типа материалов для высокочастотных транзисторов. Дело в том, что для получения высокого быстродействия транзисторов их следует изготавливать из полупроводников с высокой концентрацией электронов и высокой подвижностью. В обычных полупроводниках эти требования взаимоисключающи, ибо при легировании кристалла примесями подвижность электронов быстро снижается из-за примесного рассеяния. Оказывается, что в квантовой яме можно создать высокую концентрацию носителей *легируя не этот слой, а прилежащие к нему слои широкозонного полупроводника* [86]. При этом в узкозонной части структуры нет примесных центров, играющих роль центров рассеяния, и подвижность носителей в ней оказывается заметно выше, чем в объемном легированном полупроводнике. То же можно сделать и в *модулированно-легированной сверхрешетке*, построенной из чередующихся слоев узкозонного нелегированного и широкозонного легированного полупроводников. Для ослабления рассеяния электронов на кулоновском потенциале заряженных примесей, находящихся в широкозонной части структуры, прилегающие к квантовой яме слои широкозонного полупроводника оставляют нелегированными (эти слои называют «спейсерами»).

Как следует из рис. 1.39, подвижность носителей в квантовых ямах GaAs/AlGaAs действительно существенно выше, чем в объемном GaAs, и достигает $6,4 \cdot 10^6$ см^2/В$\cdot$с при 4,2 К [85]. Описанная возможность увеличения подвижности носителей

---

[1]) Заметим, что в совершенных сверхрешетках толщины слоев d_1 и d_2 могут меняться только дискретным образом, ибо они построены из атомных слоев полупроводника. Поэтому иногда толщину слоев указывают в числе моноатомных слоев (ML). При выращивании сверхрешетки на основе GaAs в направлении оси (100) один монослой имеет толщину $a/2 \approx 2,8$ Å, где a — параметр решетки полупроводника.

используется в HEMT-транзисторах, которые мы рассмотрим в п. 5.

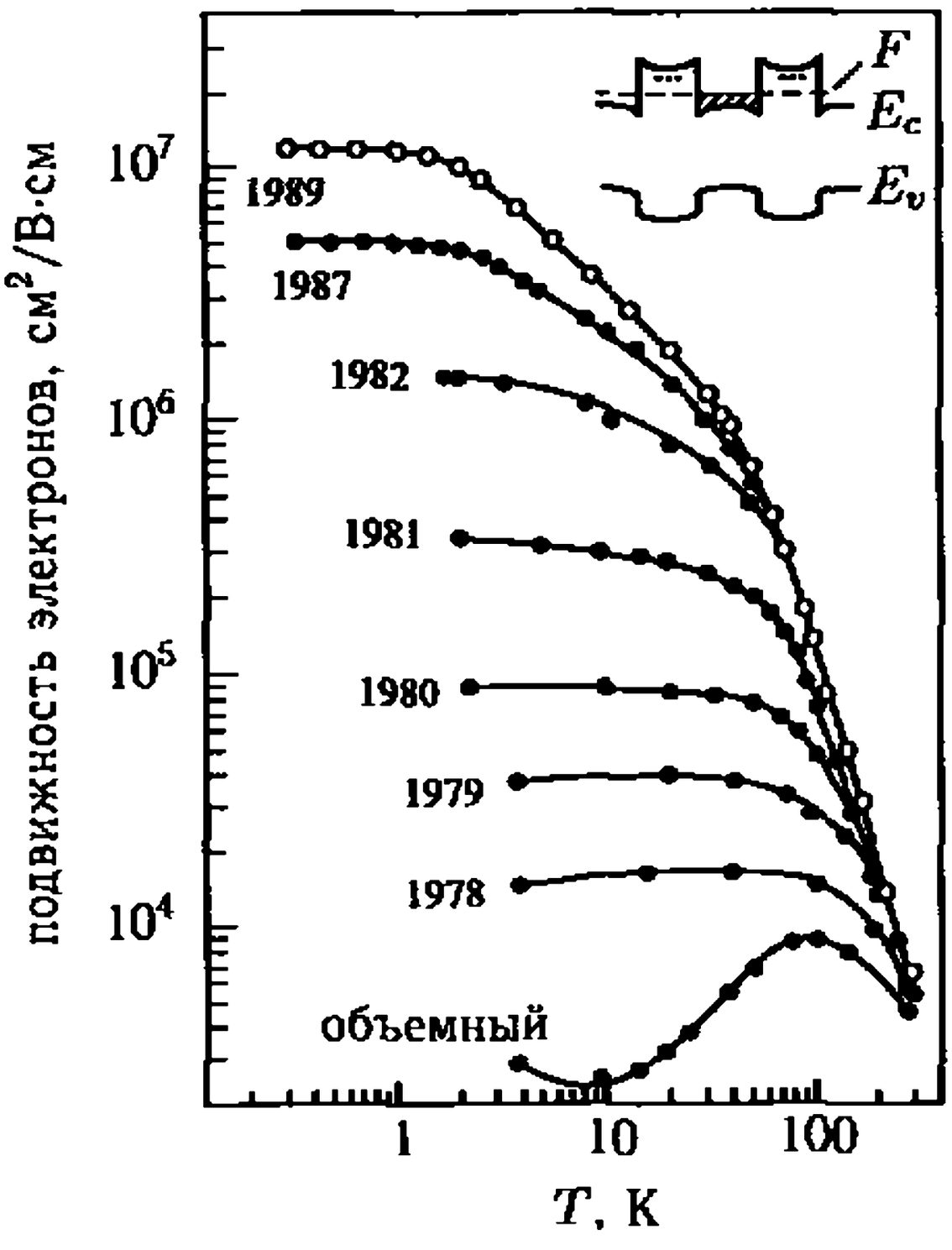


Рис. 1.39. Эволюция температурных зависимостей подвижности электронов в модулированно-легированных слоях и сверхрешетках GaAs/AlGaAs по мере совершенствования технологии их изготовления (цифры рядом с кривыми — год публикации) [85]. Нижняя кривая — подвижность электронов в объемном GaAs. На вставке показана энергетическая диаграмма модулированно-легированной сверхрешетки

Резонансно-туннельные диоды. Примером нового класса полупроводниковых приборов, построенных на основе множественных квантовых ям, может служить *резонансно-туннельный диод* (РТД) — диод с несколькими (обычно двумя) потенциальными барьерами, в котором реализуются условия для *резонансного туннелирования* [84, 87, 88]. Теория резонансного туннелирования была развита в 1963–1964 гг. Дэвисом и Хозаком [89] и Л.В. Иогансеном [90]. [1]) Обобщив эту теорию на случай структур с произвольным числом квантовых ям, Цу и Есаки [92] предложили создавать барьеры из слоев широкозонного полупроводника (см. рис. 1.40*а*).

Физический смысл резонансного туннелирования состоит в том, что электронная волна проникающего через первый барьер

[1]) Заметим, что принцип резонансного туннелирования достаточно подробно разбирался еще в учебниках 50-х годов по квантовой механике [91].

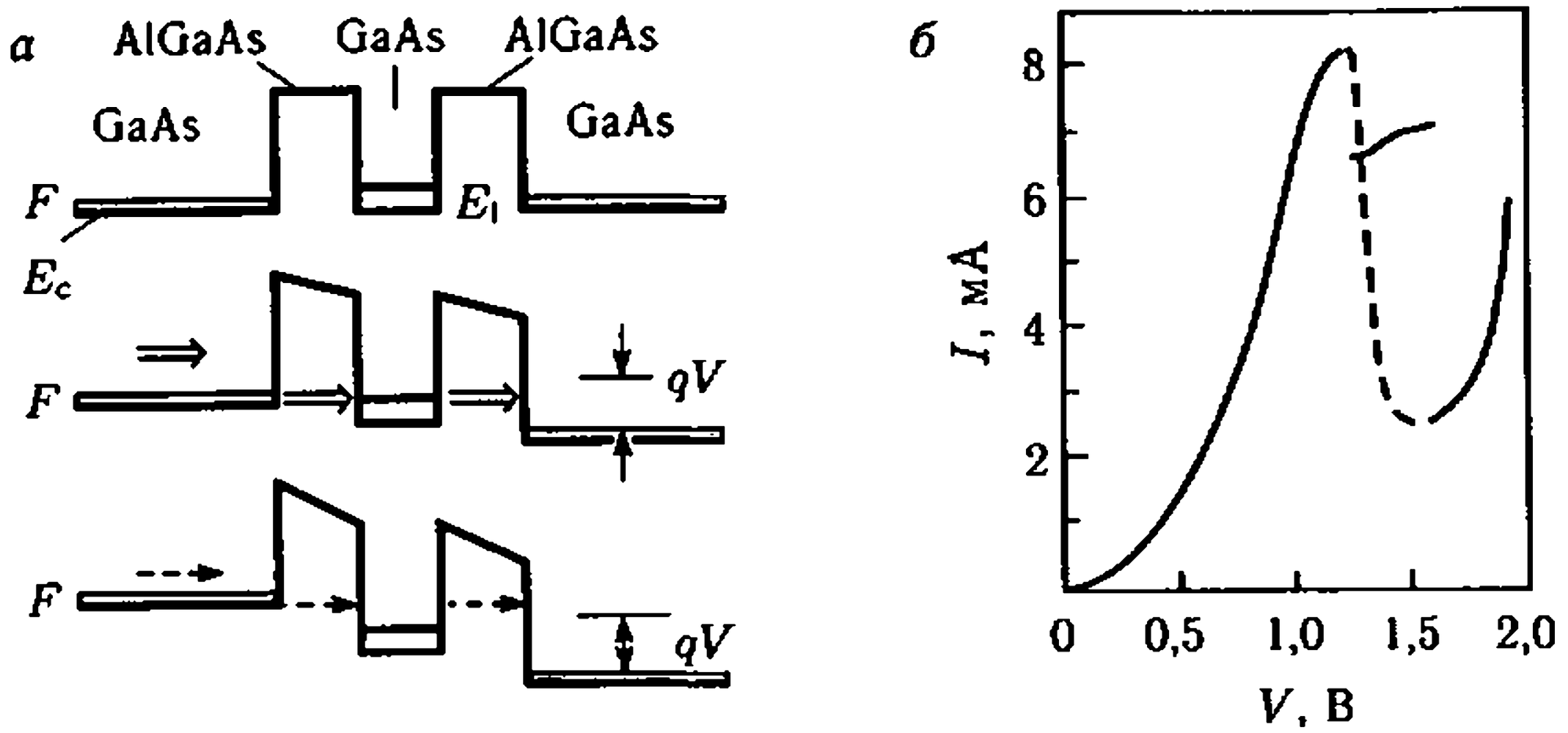


Рис. 1.40. Энергетическая диаграмма резонансно-туннельного диода (*а*) [93] и вольт-амперная характеристика диода на основе гетероструктуры InAs/AlSb (*б*) [94]. На правом рисунке сплошной линией показана экспериментально полученная характеристика диода; разрывы на кривой связаны с выпрямлением возникающих в области отрицательного дифференциального сопротивления паразитных колебаний на частоте 360 ГГц. Истинная вольт-амперная характеристика в этой области показана пунктирной линией

электрона попадает в потенциальную яму, в которой волна практически полностью «отражается» от стенок и интерферирует с волной продолжающего туннелировать электрона. Если фазы падающей и дважды отраженной от стенок волн совпадают, то амплитуда волны в яме резко возрастает, что вызывает резкое (резонансное) возрастание тока, протекающего через структуру. [1]) Можно показать, что условием возникновения резонанса является совпадение энергии туннелирующего электрона с уровнем энергии в квантовой яме.

Решение квантовомеханической задачи о движении электрона с энергией $E_\parallel$, отвечающей его движению перпендикулярно границе раздела, в структуре с двумя одинаковыми барьерами в пренебрежении рассеянием электрона показывает [90, 92], что коэффициент пропускания рассматриваемой структуры определяется формулой

$$T_{\text{tot}}(E_\parallel) = \frac{(1 - |r|^2)^2}{|1 + |r|^2 e^{2i(kL+\phi_t)}|^2}, \qquad (1.116)$$

где r — амплитуда отражения волны от барьера, L — расстояние между барьерами, $k = (2m^* E_\parallel)^{1/2}/\hbar$ — продольная компонента волнового

[1]) Наблюдаемый в структуре с двумя барьерами резонанс при туннелировании электрона является аналогом резонанса, наблюдаемого в оптике в интерферометре Фабри–Перо.

вектора электрона в падающей волне, а ϕ_t — фаза, определяемая соотношением волнового вектора k и волнового вектора электрона в яме q (при $V \neq 0$ энергия дна ямы сдвинута относительно E_c и поэтому $q \neq k$). Из этой формулы следует, что при выполнении условия

$$2(kL + \phi_t) = (2n + 1)\pi$$

прозрачность структуры резонансно увеличивается и становится равной единице, в то время как для других значений энергии прозрачность структуры приблизительно равна произведению прозрачностей двух барьеров. Таким образом, двухбарьерная структура играет роль *фильтра, пропускающего электроны только* с энергией, близкой к энергии уровня в яме. Энергетическая ширина полосы пропускания фильтра определяется двумя факторами: временем жизни электронных состояний в яме (оно связано с конечной вероятностью ухода электрона из ямы путем туннелирования сквозь барьеры) и временем релаксации импульса электрона (когерентность волновой функции электрона нарушается любым актом рассеяния).

Используя выражение (1.116) для коэффициента пропускания двухбарьерной структуры и повторяя вычисления, которые мы проводили в п. 1.4.1 при расчете вольт-амперной характеристики туннельного диода, нетрудно получить вольт-амперную характеристику резонансно-туннельного диода [92]:

$$J = \frac{qm_n^* kT}{2\pi^2\hbar^3} \int_{E_c}^{\infty} T_{\text{tot}}(E_\parallel) \ln\left[\frac{1 + \exp((F - E_\parallel)/kT)}{1 + \exp((F - E_\parallel - qV)/kT)}\right] dE_\parallel. \quad (1.117)$$

Пока напряжение, приложенное к структуре, таково, что положение уровня в квантовой яме соответствует заполненным электронами состояниям в зоне проводимости левой части структуры (см. рис. 1.40 *а*), в структуре протекает резонансный туннельный ток, однако когда уровень в яме опускается ниже края зоны проводимости, ток через структуру резко уменьшается, что приводит к появлению на вольт-амперной характеристике участка с отрицательным дифференциальным сопротивлением (см. рис. 1.40 *б*). Аналогичные резонансные явления могут наблюдаться в структурах с тремя барьерами [90] и в периодических сверхрешетках [92].

Избыточный ток в *резонансно-туннельном диоде*, как и в обычном туннельном диоде (см. п. 1.4.2), определяется туннелированием с участием фононов и примесей. Поскольку шероховатость границы раздела и локальные флуктуации состава твердого раствора (как в области барьера, так и в квантовой яме) также выступают в роли «примесей», отношение тока пика к току в минимуме J_p/J_v в РТД сильно зависит от

совершенства структуры. Кроме того, в избыточный ток большой вклад вносит термополевая эмиссия через барьер и резонансное туннелирование через более высокие квантовые состояния в яме. Наилучшее отношение J_p/J_v, равное 30 при комнатной температуре (которое даже выше, чем в обычных туннельных диодах), было получено в резонансно-туннельных диодах на основе псевдоморфных (напряженных) структур InGaAs/AlAs/InGaAs/InAs/InGaAs/AlAs/InGaAs, выращенных на подложке InP [95].

Преимуществом резонансно-туннельных диодов по сравнению с обычными туннельными диодами является их существенно более высокое быстродействие. Физическими факторами, ограничивающими быстродействие РТД, являются время жизни электронного состояния в яме и время пролета электроном обедненного слоя. [1]) Время жизни состояния в яме τ_1 определяется временем жизни относительно туннельного «просачивания» электрона сквозь стенки ямы и временем сохранения когерентного состояния электрона (временем релаксации импульса). Для ослабления эффектов рассеяния яму стремятся изготовить из полупроводника с возможно более высокой подвижностью, а сама яма, барьеры и прилегающие к барьерам снаружи тонкие области («спейсеры») не легируются. Время жизни относительно туннельного просачивания электрона из ямы контролируется толщиной и высотой барьера. При выборе этой толщины учитывают то, что плотность тока в пике J_p пропорциональна энергетической ширине электронного состояния $\Delta E = \hbar/\tau_1$, и для получения высокой плотности тока (максимального быстродействия) толщину барьеров необходимо уменьшить до 4–5 монослоев полупроводника. При этом характерное время жизни основного состояния в яме составляет $\tau_1 \sim 0,1$ пс, а время пролета истощенного слоя оказывается того же порядка.

Эксперимент показывает, что отрицательное дифференциальное сопротивление в резонансно-туннельном диоде сохраняется по крайней мере до частоты ~2,5 ТГц [93]. Из анализа эквивалентной схемы диода следует, что частота отсечки (частота, на которой отрицательное дифференциальное сопротивление прибора исчезает) также зависит от емкости структуры

---

[1]) При подаче смещения на РТД этот слой образуется со стороны положительного вывода структуры в области, примыкающей к барьеру. Задержка, возникающая при пролете электронов через обедненный слой, сильно влияет на характеристики диодов на высоких частотах и может быть использована для улучшения этих характеристик (см. п. 6.4.3).

и величины последовательного сопротивления. При этом особенно важное значение имеет величина последовательного сопротивления (РТД часто работают при плотностях тока в несколько сотен тысяч А/см^2). В работе [93] продемонстрирована возможность использования резонансно-туннельных диодов в качестве детекторов и смесителей электромагнитных колебаний с частотами до 2,5 ТГц (в субмиллиметровом диапазоне длин волн). Максимальная частота генерации, полученная на РТД из GaAs с двумя барьерами из AlAs толщиной 11 Å (4 монослоя), составляет 420 ГГц [96], а в приборах на основе InAs с барьерами толщиной в 5 монослоев из AlSb максимальная частота генерации достигла 712 ГГц [94].

Говоря о существенно более высоком быстродействии резонансно-туннельных диодов по сравнению с обычными туннельными диодами, следует задаться вопросом: а почему так происходит? На с. 79 при обсуждении туннельных диодов в качестве характеристики их быстродействия мы использовали постоянную времени $|R_{d.\min}|C_j$, которую грубо можно оценить как $C_j\Delta V/J_p$, где ΔV — разность напряжений, отвечающих минимуму и максимуму на вольт-амперной характеристике, J_p — плотность тока в максимуме, а C_j — удельная барьерная емкость структуры. Подобное отношение можно ввести и для резонансно-туннельного диода. Так вот, существенное различие этих двух типов приборов состоит в том, что обычные туннельные диоды работают при плотности тока $J_p = 10^2$–10^3 А/см^2, которая ограничена невысокой прозрачностью туннельного барьера, а в резонансно-туннельных диодах из-за практически 100%-го прохождения барьера электронами в определенном интервале энергий плотность тока может достигать $4 \cdot 10^5$ А/см^2. При этом значения ΔV в двух типах приборов остаются близкими, а удельная барьерная емкость в РТД из-за присутствия истощенного слоя в несколько раз меньше, чем в туннельных диодах. Таким образом, главными причинами более высокого быстродействия РТД является существенное улучшение условий для туннельного преодоления барьера и меньшая емкость структуры. Следует заметить, что использование все более высоких плотностей токов с целью повышения быстродействия характерно и для современных биполярных транзисторов, в которых рабочие плотности тока достигают 10^5 А/см^2.

Резонансно-туннельные диоды обладают одним из самых высоких быстродействий среди полупроводниковых диодов. На их основе созданы логические элементы, работающие на частоте 12 ГГц, делители частоты на 40 ГГц, быстродействующие аналого-цифровые преобразователи ($2 \cdot 10^9$ преобразований в секунду). Время переключения схем на РТД достигает 1,7 пс. На основе РТД можно создавать и экономичные микросхемы статических запоминающих устройств, они хорошо совместимы с современной технологией изготовления интегральных схем (ИС) на HEMT-транзисторах (см. с. 298). Основным препятствием для

использования РТД в сверхбольших ИС является трудность получения сверхтонких слоев одинаковой толщины на большой площади пластин. Для создания резонансно-туннельных диодов в основном используются полупроводники группы $A^{III}B^{V}$, однако уже есть первые результаты по созданию их на основе структур $Si/Si_{1-x}Ge_x$. Предполагается, что скоро будут освоены и чисто кремниевые структуры с сверхтонким диэлектриком из SiO_2.

Другим примером нового класса полупроводниковых приборов, построенных на основе сверхрешеток, могут служить инфракрасные фотоприемники, работающие на переходах между уровнями размерного квантования (QWIP — quantum well infrared photodetector). Работа этих фотоприемников будет рассмотрена нами в гл. 7 на с. 397.

Использование сверхрешеток позволяет улучшить характеристики существующих полупроводниковых приборов. Так, например, для систем волоконно-оптической связи необходимы быстродействующие фотоприемники, в качестве которых обычно используются лавинные фотодиоды (см. п. 7.1.5). В первых системах волоконно-оптической связи, работавших на длине волны $\lambda \approx 0,8$ мкм, использовались малошумящие лавинные фотодиоды на основе p–n-переходов из Si. Однако эти системы связи сейчас практически полностью перешли на более длинные волны (1,3–1,55 мкм), на которых оптические потери в волокне существенно ниже (до 0,2 дБ/км). Для этого диапазона длин волн уже разработаны лавинные фотодиоды на основе p–n-переходов из полупроводников $A^{III}B^{V}$, но, к сожалению, они имеют высокий уровень шума, характерный для полупроводников с близкими коэффициентами ударной ионизации электронов и дырок (см. подробнее п. 7.1.5). Использование сверхрешеток позволяет изменить эффективные коэффициенты ударной ионизации в структурах на основе этих полупроводников и получить материал, более подходящий для лавинных фотодиодов [97].

На рис. 1.41 показана энергетическая диаграмма лавинного фотодиода на основе сверхрешетки $GaAs–Al_xGa_{1-x}As$. Известно, что для указанной пары полупроводников бóльшая часть разрыва зон на гетерогранице приходится на зону проводимости. Горячий электрон, влетая из слоя $Al_xGa_{1-x}As$ в яму из GaAs, увеличивает свою кинетическую энергию на величину разрыва в зоне проводимости ($\Delta E_c \sim$ $\sim 0,5$ эВ). Поскольку пороговая энергия, необходимая для рождения электронно-дырочной пары, в слое GaAs меньше,

чем в $Al_xGa_{1-x}As$, то вероятность ударной ионизации в слое GaAs резко возрастает. Когда электрон вновь оказывается в слое $Al_xGa_{1-x}As$, его кинетическая энергия уменьшается, и в этом слое электрон, практически не ионизуя полупроводник, просто набирает энергию. В результате такого движения эффективное значение $\overline{\alpha}_n$, определяемое как

$$\overline{\alpha}_n = \frac{\alpha_{GaAs} L_{GaAs} + \alpha_{AlGaAs} L_{AlGaAs}}{L_{GaAs} + L_{AlGaAs}}, \quad (1.118)$$

где α_{GaAs} и α_{AlGaAs} — коэффициенты ударной ионизации, а L_{GaAs} и L_{AlGaAs} — толщины соответствующих слоев, заметно возрастает по сравнению с α_{GaAs}. В то же время из-за малой амплитуды модуляции энергии края валентной зоны ΔE_v аналогичный эффект для дырок практически отсутствует. В итоге, отношение коэффициентов ударной ионизации $\overline{\alpha}_n/\overline{\alpha}_p$ в такой сверхрешетке может существенно (в 5 и более раз) возрасти. Кроме того, более сильное рассеяние дырок в сверхрешетках по сравнению с электронами может способствовать дополнительному увеличению отношения $\overline{\alpha}_n/\overline{\alpha}_p$ [97].

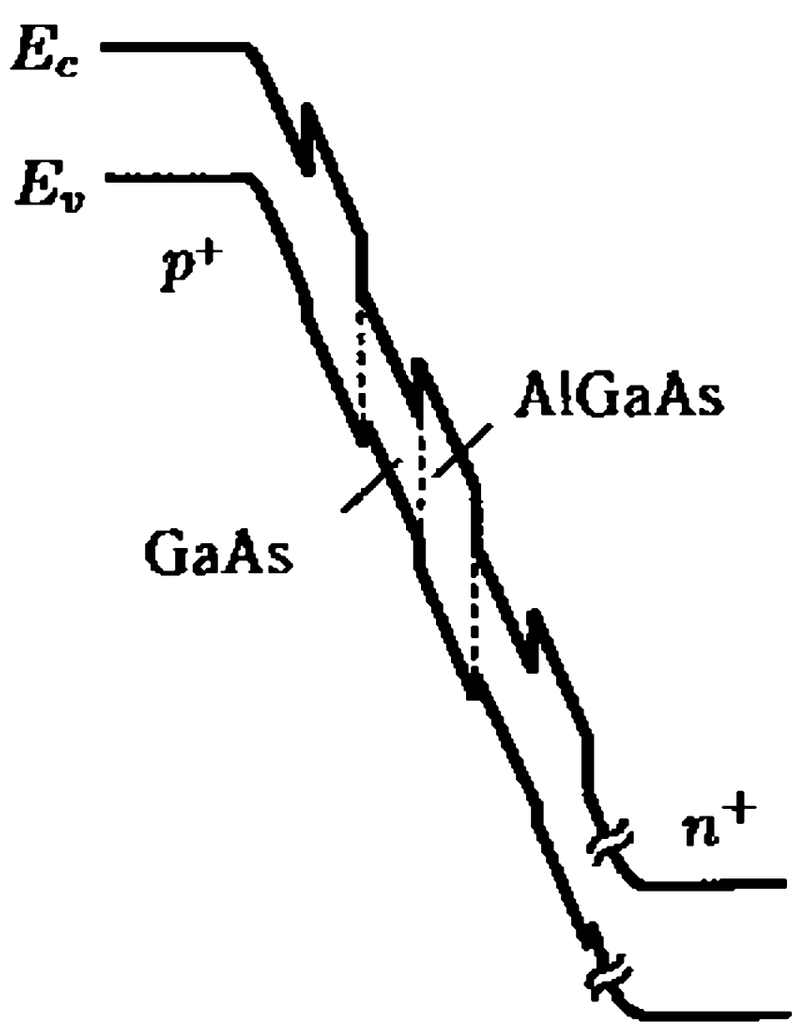


Рис. 1.41. Энергетическая диаграмма сверхрешетки GaAs/AlGaAs при обратном смещении, используемой в качестве лавинного фотодиода [97]

1.7. Диод на переменном токе

В предыдущих разделах этой главы мы рассмотрели физические явления в структурах с потенциальными барьерами и рассчитали вольт-амперные характеристики этих структур на постоянном токе. В этом разделе мы изучим особенности поведения этих структур на переменном токе.

1.7.1. Барьерная емкость. Зависимости толщины обедненного слоя в p–n-переходе и барьере Шоттки от напряжения смещения, полученные нами в п. 1.1 и 1.5.1, позволяют ожидать, что реакция этих структур на подачу переменного напряжения будет иметь емкостную составляющую.

Рассмотрим p–n-переход единичной площади. Пусть на p–n-переход от внешнего источника подается перепад напряжения dV. Поскольку новому значению напряжения смещения отвечает

новое значение толщины обедненного слоя W, то для изменения заряда в обедненном слое через переход должен протечь заряд

$$dQ = N_d dx_n = N_a dx_p,$$

где dx_n и dx_p — доли полного изменения толщины обедненного слоя,

$$dW = dx_n + dx_p,$$

приходящиеся на n- и p-область (см. рис. 1.42). Величину, определяемую как отношение приращения заряда dQ к приращению напряжения dV, и принято называть *барьерной (зарядной) емкостью* p–n-перехода $C = dQ/dV$.

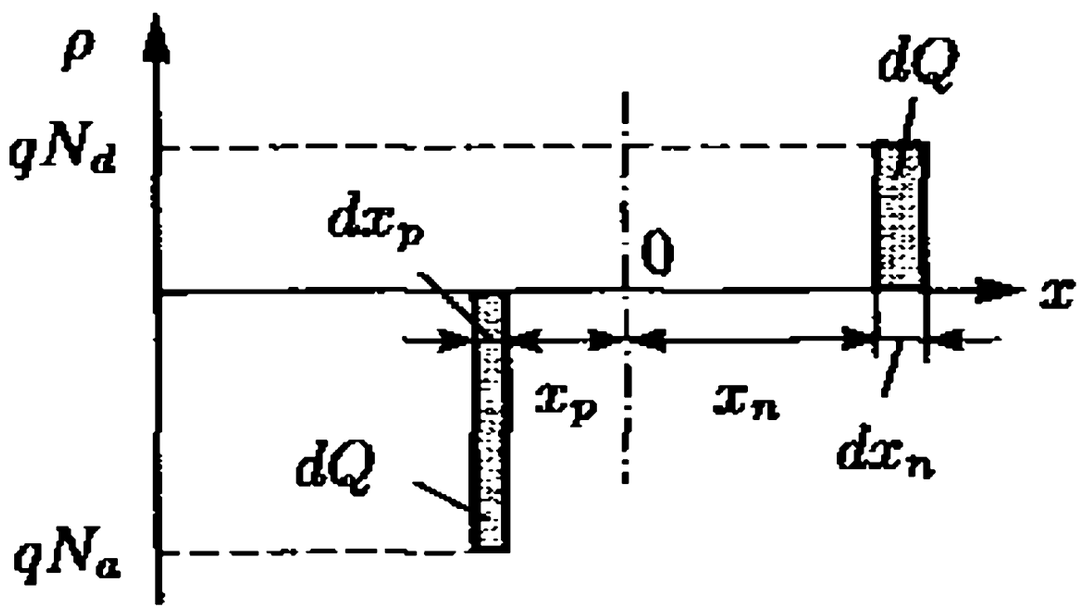


Рис. 1.42. Изменение толщины и заряда в обедненном слое при изменении напряжения смещения на p–n-переходе

Чтобы рассчитать удельную барьерную емкость (емкость p–n-перехода единичной площади), надо выразить величину заряда примесей Q' в таком переходе через толщину области пространственного заряда W и воспользоваться полученными в п. 1.1 зависимостями этой толщины от напряжения смещения $W(V)$.

Рассмотрим сначала случай резкого p–n-перехода. Используя соотношение

$$Q' = qN_d x_n = qN_a x_p = q\frac{N_d N_a}{N_d + N_a} W, \qquad (1.119)$$

приходим к следующему уравнению: [1])

$$C = \frac{dQ}{dV} = \frac{d(-Q')}{dV} = -q\frac{N_d N_a}{N_d + N_a}\frac{dW}{dV}. \qquad (1.120)$$

Подставляя в это уравнение зависимость $W(V)$ для резкого p–n-перехода (формула (1.13)), находим его удельную емкость:

$$C = \sqrt{\frac{\varepsilon q}{8\pi}\frac{N_d N_a}{N_d + N_a}}\sqrt{\frac{1}{\phi_k - V}}. \qquad (1.121)$$

[1]) Условие $dQ = -dQ'$ означает, что приходящий от источника напряжения заряд dQ *уменьшает* (компенсирует) заряд примесей dQ' в области пространственного заряда p–n-перехода.

Заметим, что полученная величина в точности совпадает с удельной емкостью плоского конденсатора, расстояние между обкладками которого равно W ($C = \varepsilon/4\pi W$). Можно доказать, что это утверждение справедливо и для произвольного профиля легирования p–n-перехода [98].

В случае плавного p–n-перехода (см. рис. 1.2) поверхностная плотность заряда примесей определяется интегралом

$$Q' = qa \int_0^{W/2} x\,dx = \frac{qaW^2}{8}, \qquad (1.122)$$

где a — градиент концентрации примеси в p–n-переходе. Учитывая зависимость $W(V)$ для плавного p–n-перехода (формула (1.16)) и проводя несложные вычисления, получаем его удельную емкость:

$$C = \frac{d(-Q')}{dV} = -\frac{qa}{4} W \frac{dW}{dV} = \frac{\varepsilon}{4\pi}\left(\frac{\pi qa}{3\varepsilon}\frac{1}{\phi_k - V}\right)^{1/3}. \qquad (1.123)$$

В барьере Шоттки к полупроводнику n-типа поверхностная плотность заряда примесей в обедненном слое равна $Q' = qN_dW$. Используя формулу (1.102) для зависимости $W(V)$ в барьере Шоттки, находим его удельную емкость:

$$C = \sqrt{\frac{\varepsilon q N_d}{8\pi}}\sqrt{\frac{1}{V_{bi} - V}}. \qquad (1.124)$$

Нетрудно убедиться, что и в случае барьера Шоттки, и в случае плавного p–n-перехода результат совпадает с формулой для удельной емкости плоского конденсатора.

Из полученных формул следует, что с ростом обратного смещения барьерная емкость уменьшается, причем по-разному для разных профилей легирования p–n-перехода (ср. формулы (1.121) и (1.123)). Это позволяет, экспериментально измерив вольт-фарадную характеристику p–n-перехода, определить не только величину контактной разности потенциалов ϕ_k (или высоту потенциального барьера qV_{bi} в контакте металл–полупроводник), но также сделать вывод о характере распределения примесей в p–n-переходе. Метод исследования полупроводниковых структур, основанный на изучении их вольт-фарадных характеристик, получил название *емкостной спектроскопии полупроводников*.

Метод емкостной спектроскопии может быть использован для определения профиля легирования p–n-перехода даже тогда,

когда этот профиль отличается от рассмотренных выше идеализированных случаев. Покажем, как это делается, на примере асимметрично легированного p^+–n-перехода с произвольным профилем распределения примесей в n-области. Дифференцируя выражение для удельной емкости p–n-перехода, $C = \varepsilon/4\pi W$, по V, получаем

$$\frac{dC}{dV} = \frac{\varepsilon}{4\pi}\frac{d}{dV}\left(\frac{1}{W}\right) = -\frac{\varepsilon}{4\pi}\frac{1}{W^2}\frac{dW}{dV}. \qquad (1.125)$$

Чтобы связать приращения ΔV и ΔW, учтем, что при расширении обедненного слоя на ΔW полный заряд ионизованных примесей в нем увеличивается на $qN_d(W)\Delta W$, что приводит, в соответствии с теоремой Остроградского–Гаусса, к возрастанию напряженности электрического поля *в каждой точке* обедненного слоя на величину

$$\Delta\mathcal{E} = \frac{4\pi q N_d(W)\Delta W}{\varepsilon}.$$

Полное изменение высоты потенциального барьера при этом составляет

$$\Delta V = W\Delta\mathcal{E},$$

откуда сразу же получаем

$$\frac{dV}{dW} = \frac{4\pi q W N_d(W)}{\varepsilon}.$$

Подставляя это соотношение в уравнение (1.125) и исключая из него переменную W, окончательно находим

$$N_d(W) = -\frac{8\pi}{q\varepsilon}\left[\frac{d}{dV}\left(\frac{1}{C^2}\right)\right]^{-1}. \qquad (1.126)$$

Эта формула позволяет, обработав экспериментально полученную зависимость $C(V)$, построить зависимость концентрации N_d от расстояния до металлургической границы p–n-перехода ($x_n \approx$ $\approx W = \varepsilon/4\pi C$). Аналогичный подход может быть использован и для определения распределения примеси в барьере Шоттки [56].

Интересная возможность изучения полупроводниковых структур возникает при использовании барьера Шоттки с ртутным контактом. Передвигая заполненный ртутью капилляр по поверхности полупроводниковой пластины и измеряя емкость контакта (и, при необходимости, ее зависимость от напряжения), можно получить информацию о распределении примеси по поверхности пластины [56].

Варикапы и варакторы. Свойство p–n-перехода изменять емкость при изменении приложенного к нему напряжения находит применение в двух типах полупроводниковых приборов: варикапах и варакторах. Эти приборы широко используются в промышленной и бытовой электронике [98].

Варикапами называют p–n-переходы, используемые как емкость, перестраиваемую напряжением, [1]) в схемах с невысокой рассеиваемой мощностью. При малой амплитуде переменного напряжения, когда нелинейность зависимости $C(V)$ практически не проявляется, эти приборы применяются в системах электронной перестройки частоты (например, в системах настройки телевизоров). Как элемент с нелинейной зависимостью $C(V)$, варикапы используются для параметрического усиления слабых сигналов [2].

Основными характеристиками варикапов являются *чувствительность* (относительное изменение емкости при изменении напряжения смещения $(1/C)(dC/dV)$), *кратность перестройки емкости* (отношение максимальной к минимальной емкости $C_{\max}/C_{\min}$) и *добротность* (отношение запасенной в варикапе энергии к энергии, рассеиваемой за один период колебаний). Нетрудно показать, что чувствительность варикапа пропорциональна показателю степени m в зависимости емкости p–n-перехода от напряжения $C \sim (\phi_k - V)^{-m}$. Поэтому резкие p–n-переходы имеют более высокие чувствительность и кратность перестройки емкости по сравнению с плавными p–n-переходами. Чтобы еще больше увеличить значения этих параметров, варикапы создают на основе так называемых *сверхрезких переходов*, в которых концентрация примесей в базе диода убывает по мере удаления от p–n-перехода. При этом показатель m удается увеличить до 1,5–2. Необходимый профиль легирования в таких диодах можно создать, например, путем вплавления акцепторной примеси в диффузионный слой n-типа [98].

Высокая добротность варикапа важна для таких применений, как параметрическое усиление сигналов и перестройка частоты. Так, мощность собственного шума параметрического усилителя обратно пропорциональна добротности. Потери, вносимые варикапом в колебательный контур на высоких частотах, определяются произведением его емкости C_j на величину последовательного сопротивления R_s толщи диода и контактов. *Предельная*

---

[1]) Название этого прибора происходит от английского variable capacitor — переменная емкость.

частота, на которой могут работать варикапы, равна

$$f_c = 1/(2\pi R_s C_j).$$

Тщательная оптимизация конструкции приборов и использование материалов с высокой подвижностью (GaAs) позволяет получить $f_c \approx 400$ ГГц (диод MV20001 компании Microwave Device Technology). Характерные значения емкости варикапов лежат в пределах 0,3–200 пФ. Примером отечественного варикапа, разработанного для систем электронной настройки телевизоров, является кремниевый диод КВ109.

Варакторами называют мощные диоды, нелинейная зависимость емкости которых от напряжения используется для умножения частоты. Работа этих приборов основана на том, что при подаче на диод переменного напряжения большой амплитуды из-за сильной нелинейности зависимости $C(V)$ в токе диода возникает большое число высших гармоник. Варакторы, работающие в диапазоне СВЧ, также называют *умножительными диодами*. Поскольку, в отличие от обращенных диодов (см. п. 1.4.4), в варакторах используется нелинейность *реактивной* составляющей проводимости, к.п.д. преобразователей частоты на варакторах оказывается очень высоким и достигает 85% для генерации второй гармоники на частоте 500 МГц и 60% на 10 ГГц. Выходная мощность преобразованного сигнала при этом измеряется единицами–десятками ватт. Максимальная частота, полученная с помощью четырехкаскадного умножителя на барьерах Шоттки из GaAs, равна 1500 ГГц (при к.п.д. последнего удвоителя частоты 4%) [99]. Примерами варакторов, выпускаемых отечественной промышленностью, могут служить кремниевые диоды КВ106 и КА609.

В настоящее время большой интерес представляют варакторы типа HBV — heterostructure barrier varactor, построенные на основе симметричных гетеропереходных структур типа n-GaAs–i-AlGaAs–n-GaAs [100]. Из-за симметричной вольт-фарадной характеристики с помощью таких структур можно генерировать только нечетные гармоники, но для их работы не требуется подавать на диод напряжение смещения. На таких варакторах в схеме утроителя частоты была получена выходная мощность 9 мВт на частоте 247 ГГц (к.п.д. равен 12%) [101].

1.7.2. Диффузионная емкость. На практике приборы с p–n-переходами (полупроводниковые диоды) используются прежде всего для выпрямления и других преобразований на переменном токе. Поэтому нам важно знать, какие характеристики

полупроводника определяют основные параметры приборов в этих условиях.

Рассмотрим тонкий p–n-переход, на который подано постоянное прямое смещение V_0 и небольшое переменное напряжение амплитудой $\delta V \ll kT/q$, изменяющееся по гармоническому закону:

$$V = V_0 + \delta V\, e^{i\omega t}.$$

Плотность тока, протекающего через диод, будем искать в виде

$$J = J_0 + \delta J\, e^{i\omega t}.$$

Для расчета диффузионного тока нам надо решить уравнения непрерывности (1.21). Для определенности рассмотрим асимметрично легированный p^+–n-переход, ток в котором определяется дырками, инжектируемыми в n-область. На переменном токе в хорошо знакомое нам уравнение (1.26) мы должны добавить слагаемое, описывающее накопление заряда dp/dt:

$$D_p \frac{d^2p}{dx^2} - \frac{p - p_{n0}}{\tau_p} = \frac{dp}{dt}. \qquad (1.127)$$

Решение этого уравнения будем также искать в виде суммы стационарного распределения $p_0(x)$ и возмущения $\delta p(x)$, зависящего от времени:

$$p(x,t) = p_0(x) + \delta p(x)\, e^{i\omega t}. \qquad (1.128)$$

Подставляя это решение в уравнение (1.127), для функции $p_0(x)$ получаем обычное решение (1.27), а для функции $\delta p(x)$ следующее уравнение:

$$D_p \frac{d^2\,\delta p}{dx^2} - \frac{\delta p}{\tau_p} = i\omega\,\delta p. \qquad (1.129)$$

В диоде с достаточно толстой базой n-типа граничными условиями для этого уравнения являются

$$\delta p(x \to \infty) = 0 \quad \text{и} \quad \delta p(x_n) = p_{n0} \exp\left(\frac{qV_0}{kT}\right) \frac{q\,\delta V}{kT}.$$

Заметим, что формально уравнение (1.129) становится эквивалентным уравнению (1.26), если произвести в нем замену

$$\tau_p \to \frac{\tau_p}{1 + i\omega\tau_p}. \qquad (1.130)$$

Тогда, подставляя это комплексное выражение для τ_p в уравнение для диффузионной длины $L_p = \sqrt{D_p\tau_p}$ и далее подставляя

L_p в уравнение (1.28), в конце концов приходим к следующему соотношению:

$$\delta J = \frac{qD_p p_{n0}}{L_p}\sqrt{1+i\omega\tau_p}\,\exp\left(\frac{qV_0}{kT}\right)\frac{q}{kT}\,\delta V. \qquad (1.131)$$

Из уравнения (1.131) следует, что в комплексной проводимости диода на переменном токе можно выделить действительную и мнимую составляющие,

$$Y = \frac{\delta J}{\delta V} = G_d + i\omega C_d,$$

которые при $\omega\tau_p \ll 1$ равны

$$G_d \approx \frac{qJ_0}{kT}, \quad C_d \approx \frac{qD_p\,\tau_p\,p_{n0}}{2L_p}\exp\left(\frac{qV_0}{kT}\right)\frac{q}{kT} \approx \frac{qJ_0}{2kT}\,\tau_p \qquad (1.132)$$

и связаны между собой соотношением

$$C_d \approx \frac{\tau_p}{2}\,G_d. \qquad (1.133)$$

Таким образом, мы видим, что в проводимости p–n-перехода на переменном токе в области прямых смещений появляется компонента, имеющая емкостный характер. Характеризующая ее величина C_d называется *диффузионной емкостью p–n-перехода*. Измерение диффузионной емкости p–n-перехода может служить методом определения времени жизни неосновных носителей заряда.

1.7.3. Импульсные характеристики и быстродействие диодов. То, что в реакции p–n-перехода на изменение напряжения прямого смещения есть емкостная составляющая, означает, что если скачком увеличить напряжение на p–n-переходе, то в первый момент будет наблюдаться всплеск протекающего тока, который затем будет релаксировать к значению, отвечающему новой величине напряжения смещения. Физической причиной такой реакции является то, что при быстром понижении высоты потенциального барьера концентрация преодолевших этот барьер носителей на границе p–n-перехода и нейтральной области быстро увеличивается. При этом градиент концентрации инжектируемых носителей на этой границе становится очень большим, и диффузионный ток резко возрастает. Переходный процесс продолжается до тех пор, пока распределение инжектированных носителей в нейтральных областях не достигнет стационарного распределения, отвечающего новому значению высоты

потенциального барьера. Характерное время установления равновесия, очевидно, равно времени жизни неравновесных носителей.

Рассмотрим два важных случая, которые довольно часто встречаются при практическом использовании диодов:

1) поведение p–n-перехода при внезапном прекращении тока через него и

2) поведение p–n-перехода при быстром переключении полярности поданного на него напряжения.

Обсудим сначала, что происходит в диоде, когда после продолжительного пропускания прямого тока через p–n-переход ток через него внезапно прерывается. Эксперимент показывает, что после небольшого начального спада напряжения, связанного с перезарядкой емкости измерительной части схемы, на диоде в течение некоторого времени продолжает оставаться так называемая *послеинжекционная ЭДС*, которая практически линейно уменьшается со временем (рис. 1.43) [102]. Что же является источником этой ЭДС?

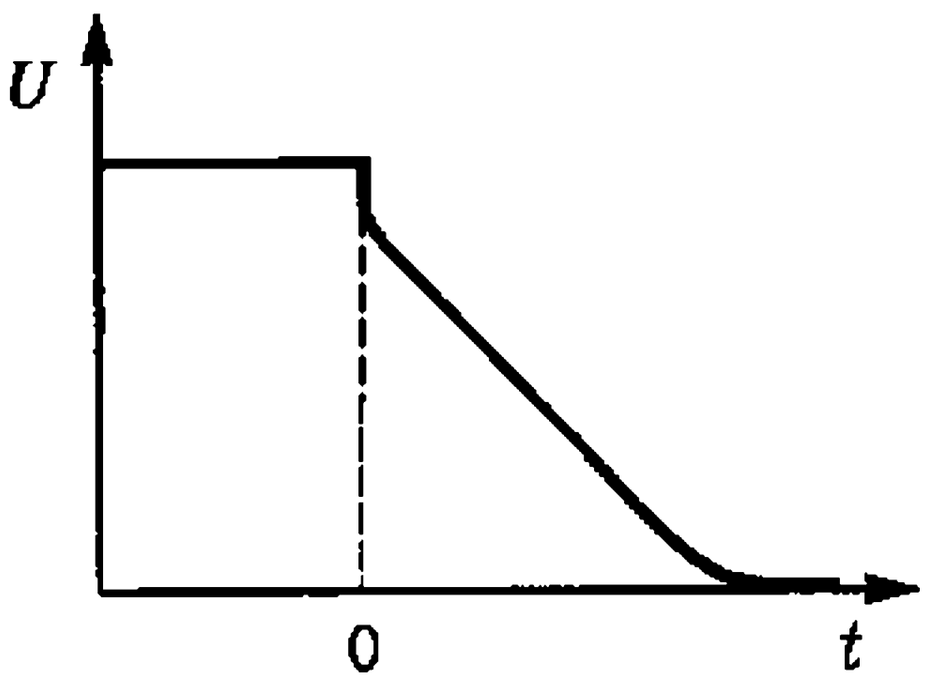


Рис. 1.43. Релаксация послеинжекционной ЭДС после прерывания пропускания прямого тока через p–n-переход

В отсутствии внешней цепи после прекращения перезарядки емкости полный ток через p–n-переход можно считать равным нулю. Как мы показали в п. 1.2.1, этот ток определяется разностью двух встречных потоков — потока основных носителей, пытающихся преодолеть потенциальный барьер, и потока неосновных носителей, затягиваемых электрическим полем p–n-перехода. В рассматриваемых нами *неравновесных условиях* этот последний поток заряжает области диода и изменяет высоту барьера таким образом, чтобы в каждый момент времени суммарный ток через p–n-переход оставался равным нулю. Поэтому в этих условиях высота потенциального барьера и значение послеинжекционной ЭДС будут определяться *мгновенным значением концентрации неравновесных носителей* на краях p–n-перехода. Если пренебречь пространственным перераспределением инжектированных носителей и считать, что их концентрация в результате рекомбинации уменьшается по закону

$$n_p - n_{p0} \sim \exp\left(\frac{-t}{\tau_n}\right),$$

то напряжение на p–n-переходе будет изменяться как

$$U(t) = \frac{kT}{q} \ln\left(\frac{n_p}{n_{p0}}\right) \approx U(0) - \frac{kT}{q} \frac{t}{\tau_n}. \qquad (1.134)$$

Из этого следует, что изучение кинетики спада послеинжекционной ЭДС в диодах может служить одним из методов измерения времени жизни неравновесных носителей заряда. Этот метод широко используется в промышленности для контроля параметров полупроводниковых приборов.

С другим случаем, когда на p–n-переход попеременно подается напряжение разного знака, приходится сталкиваться в любых электронных схемах, где диоды используются в качестве выпрямителей. Что же происходит в этом случае?

Мы только что показали, что сразу после прекращения инжекции и нулевом токе через диод напряжение на диоде уменьшается со скоростью

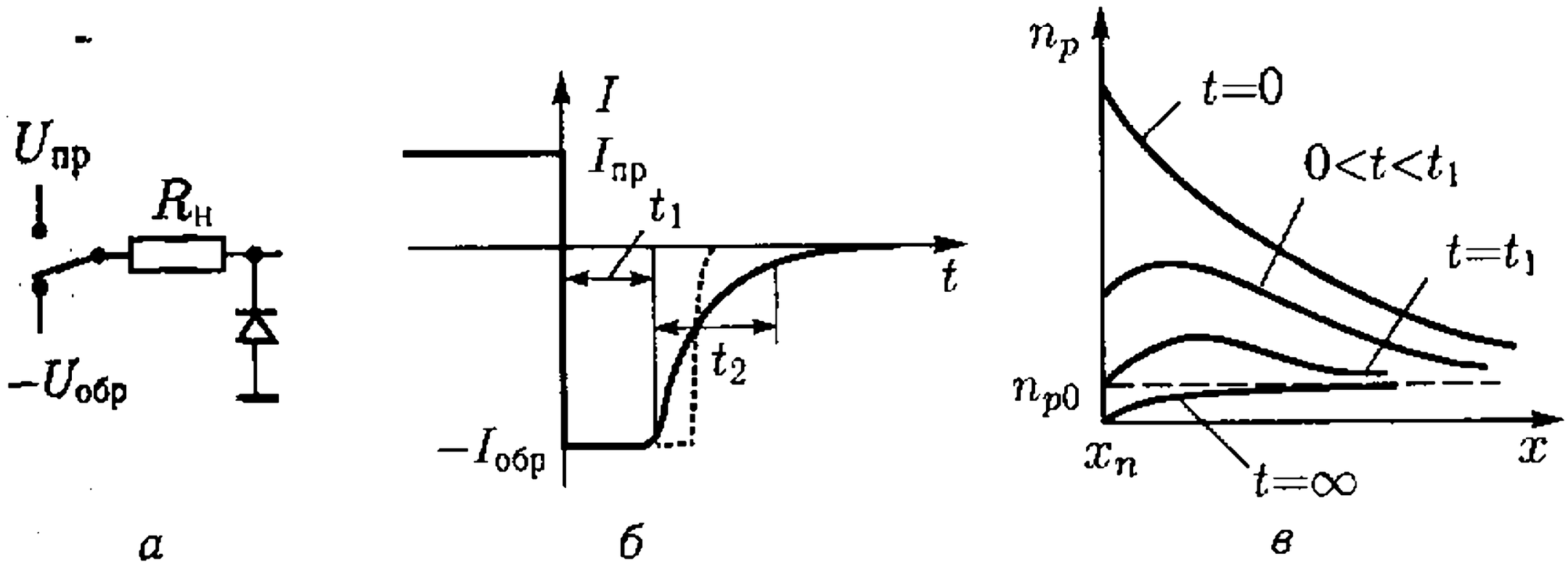
$$\left|\frac{dU}{dt}\right| = \frac{kT}{q\tau_n}.$$

Если же к диоду приложить напряжение, уменьшающееся с более высокой скоростью, то в этой ситуации диод начнет вести себя как *источник тока* и через диод потечет ток обратной полярности. Это — ток неосновных носителей, затягиваемых электрическим полем p–n-перехода, который в предельном случае, когда полярность напряжения на диоде переключается с прямой на обратную мгновенно, просто воспроизводит эволюцию градиента концентрации неосновных носителей на краю p–n-перехода.

Если величина обратного тока $I_{обр}$ ограничена сопротивлением нагрузки (как в электрической схеме, показанной на рис. 1.44 *а*), то после переключения полярности напряжения диод некоторое время находится в проводящем состоянии и падение напряжения на нем близко к нулю; продолжительность этого первого этапа обратного восстановления диода (до обращения напряжения на диоде в нуль в момент времени $t = t_1$, см. рис. 1.44 *б*) определяется двумя процессами:

1) рекомбинацией инжектированных носителей в базе диода и

2) *вытягиванием* (*экстракцией*) этих носителей из базы диода с помощью обратно смещенного p–n-перехода и удалением их через контакты.



а *б* *в*

Рис. 1.44. Переходные процессы в p–n-переходе при переключении полярности подаваемого на него напряжения: *а* — электрическая схема, *б* — изменение тока через диод при переключении полярности напряжения, *в* — эволюция распределения концентрации неосновных носителей в базе диода после переключения. Пунктиром на рис. *б* показана переходная характеристика диода с накоплением заряда.

В p^+–n-переходе с толстой базой продолжительность первого этапа определяется уравнением

$$\operatorname{erf}\sqrt{\frac{t_1}{\tau_n}} = \frac{1}{1 + I_{обр}/I_{пр}}, \qquad (1.135)$$

где

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2}\, dy$$

— функция ошибок, а $I_{пр}$ и $I_{обр}$ — значения прямого и обратного токов [102]. Анализ этой формулы показывает, что пропускание через p–n-переход обратного тока, при котором происходит экстракция инжектированных носителей, позволяет значительно уменьшить продолжительность первого этапа обратного восстановления диода.

На втором этапе обратного восстановления диода величина обратного тока определяется величиной остаточного заряда неосновных носителей в базе диода. Продолжительность этого этапа восстановления (t_2 на рис. 1.44 *б*) обычно определяется как время, в течение которого обратный ток уменьшается в 10 раз. Расчеты показывают, что экстракция носителей и здесь помогает уменьшить время t_2. На рис. 1.44 *в* показана временна́я эволюция профиля распределения накопленных в базе диода неосновных носителей после того, как к нему было приложено отрицательное смещение.

В схемах выпрямителей переменного тока существование большого и продолжительного переходного обратного тока диодов может приводить к заметному уменьшению к.п.д. выпрямителей и увеличению мощности, рассеиваемой диодами. Это

особенно заметно при работе высоковольтных диодов на повышенных частотах (в этих диодах для получения хороших статических вольт-амперных характеристик обычно используется материал с большим временем жизни). Длительное время обратного восстановления диода сильно ограничивает и использование этих приборов в электронных схемах, работающих в импульсном режиме.

Из изложенного выше следует, что для создания *быстродействующих* полупроводниковых диодов необходимо, чтобы полупроводник имел *малое время жизни*. Для уменьшения времени жизни полупроводниковые кристаллы *намеренно* легируются примесями, создающими глубокие примесные уровни вблизи середины запрещенной зоны (как мы видели в п. 1.2.2, именно такие уровни дают наибольший вклад в темп рекомбинации). В кремнии для этой цели используются примеси Au и Pt, а также облучение готовых структур быстрыми электронами [11]. Радиационные дефекты, образующиеся при таком облучении, отжигаются при 300 °С , то есть при температуре, заметно превышающей рабочую.

Другим способом повышения быстродействия может быть использование диодов с тонкой базой, рассмотренных нами на с. 25. Быстродействие этих диодов определяется временем, необходимым для диффузии инжектированных носителей через тонкую базу до контакта. Примером таких диодов могут служить отечественные эпитаксиально-планарные диоды КД509А, КД510А, имеющие время восстановления 2–4 нс при напряжении пробоя 70 В.

Рассмотренные нами явления, приводящие к ограничению быстродействия диодов с p–n-переходами, обязаны своим происхождением тому, что в работе этих диодов используется *инжекция неосновных носителей*, которые имеют конечное время жизни. Чтобы увеличить быстродействие, диоды для импульсных устройств (*диоды с малым временем восстановления*) изготавливают из полупроводников, в которые с целью уменьшения времени жизни вводят примеси, создающие глубокие уровни. Однако возможности этого подхода ограничены обычно невысокой растворимостью таких примесей; в Si, например, достаточно трудно получить время жизни короче нескольких наносекунд. В GaAs, который является прямозонным полупроводником, время жизни может быть уменьшено до 0,1 нс. Поэтому решительное продвижение в область более высоких частот требует отказа от использования приборов, работа которых основана на явлении инжекции.

Приборами, в работе которых используются только основные носители, являются туннельные (см. п. 1.4), резонансно-туннельные диоды (п. 1.6.2.1) и диоды с барьером Шоттки (п. 1.5). Быстродействие туннельных диодов определяется временем жизни электронного состояния относительно его просачивания в другую область диода путем туннелирования. Быстродействие диодов Шоттки ограничено временем рассасывания объемного заряда, которое равно времени максвелловской (диэлектрической) релаксации (оценки этих времен см. на с. 98).

1.7.4. Диоды с накоплением заряда. Понимание явлений, происходящих при рассасывании инжектированных в p–n-переход неосновных носителей (см. п. 1.7.3), используется для создания одного важного типа диодов — *диодов с накоплением заряда* [103]. [1]) Принцип действия этих диодов был предложен и изучен в работе Молла с соавторами [104].

Основной задачей, которая решается в диодах с накоплением заряда, является максимальное уменьшение продолжительности второго этапа обратного восстановления диода. В обычных диодах после окончания первого этапа максимум остаточного инжектированного заряда находится на расстоянии порядка диффузионной длины от p–n-перехода (см. рис. 1.44 *в*), и этот заряд медленно, путем диффузии, экстрагируется из базы диода обратно смещенным p–n-переходом. Чтобы рассасывание заряда на втором этапе протекало быстрее, надо, во-первых, приблизить максимум распределения заряда к p–n-переходу и, во-вторых, сделать так, чтобы диффузии носителей во время экстракции помогало бы тянущее электрическое поле.

Для решения поставленной задачи авторы [104] предложили использовать p–i–n-диод (см. п. 1.2.4). Путем диффузии донорной и акцепторной примесей в противоположные грани тонкой пластинки в диоде создают области, в которых в отсутствие инжекции существует встроенное электрическое поле (см. рис. 1.45).

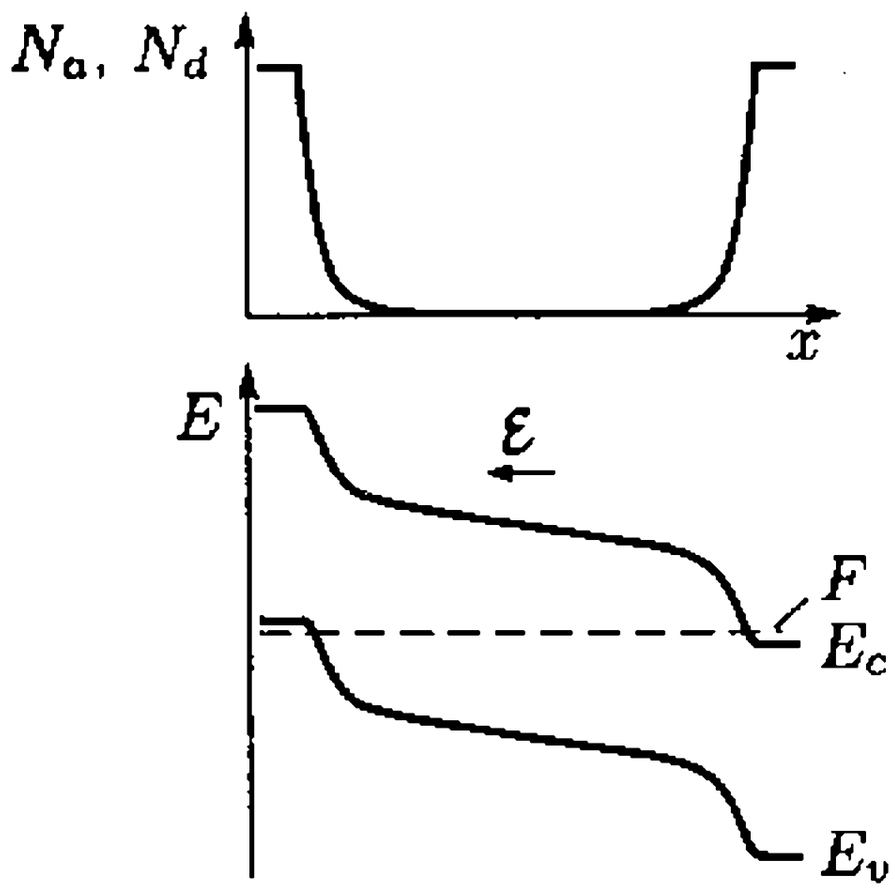


Рис. 1.45. Профиль легирования и энергетическая диаграмма p–i–n-диода с накоплением заряда

[1]) В зарубежной литературе для обозначения этих диодов используется аббревиатура SRD (step-recovery diodes).

Направление этого поля таково, что оно *препятствует* диффузии инжектируемых носителей в глубь пластины и, следовательно, при подаче прямого смещения инжектированные носители остаются «прижатыми» к p–i- и i–n-переходам. Поэтому, когда после пропускания прямого тока, к диоду оказывается приложенным обратное смещение, экстракция инжектированных носителей происходит очень быстро, обеспечивая быстрое восстановление непроводящего (блокирующего) состояния диода.

Для оценки основных параметров диодов с накоплением заряда рассчитаем распределение инжектированных носителей в p^+—n-переходе, в n-области которого за счет неоднородного легирования создано постоянное электрическое поле $\mathcal{E}$. Поскольку в таком переходе происходит преимущественная инжекция дырок, нам надо решить уравнение непрерывности для дырок (1.23), в котором $\mathcal{E} = \text{const} \neq 0$.

Уровень инжекции будем считать невысоким, чтобы экранированием электрического поля инжектируемыми носителями можно было пренебречь. Нетрудно показать, что решение уравнения непрерывности представляет собой сумму нарастающей и убывающей экспонент, характерные масштабы длины которых L_1 и L_2 определяются из решения квадратного уравнения

$$L_{1,2}^2 \pm \mu_p \mathcal{E} \tau_p L_{1,2} - D_p \tau_p = 0, \qquad (1.136)$$

где μ_p, τ_p и D_p — подвижность, время жизни и коэффициент диффузии дырок. Если напряженность электрического поля мала ($\mathcal{E} \ll kT/qL_p$), то значения L_1 и L_2 практически совпадают с диффузионной длиной $L_p = \sqrt{\mu_p \tau_p}$. Если же $\mathcal{E} \gg kT/qL_p$, то значения L_1 и L_2 существенно изменяются: для тормозящего электрического поля характерная длина становится равной

$$L_1 \approx \frac{L_p^2}{\mu_p \mathcal{E} \tau_p} = \frac{kT}{q\mathcal{E}},$$

а для тянущего электрического поля — $L_2 \approx \mu_p \mathcal{E} \tau_p$. Физический смысл L_1 и L_2 хорошо ясен: L_1 — это расстояние, на котором потенциальная энергия электрона изменяется на величину kT, а L_2 — это расстояние, на которое за время жизни успевают продрейфовать носители.

Поскольку в случае сильного тормозящего поля характерный масштаб длины L_1 уменьшается в $qL_p\mathcal{E}/kT$ раз, то характерное время восстановления уменьшается от τ_p до [103]

$$\tau_p \left(\frac{kT}{qL_p\mathcal{E}}\right)^2 = \left(\frac{kT}{q\mathcal{E}}\right)^2 \frac{1}{D_p}.$$

Таким образом, видно, что создание в диоде сильного тормозящего поля для диффузии носителей позволяет существенно уменьшить время восстановления.

Типичная переходная характеристика диода с накоплением заряда показана пунктирной линией на рис. 1.44 *б*. Видно,

что длительность первой фазы восстановления ($t_1 \approx \tau_p \ln(1 + I_{пр}/I_{обр})$ [103]) в них оказывается несколько больше, чем в обычных диодах (см. формулу (1.135)). А вот время второй фазы восстановления (t_2) становится существенно короче. В специально спроектированных диодах с накоплением заряда скорость нарастания обратного напряжения на диоде может достигать 500 В/нс. Дифференцирование этого нарастания с помощью специальных электронных схем (например, короткозамкнутой линии задержки) позволяет с помощью диодов с накоплением заряда формировать импульсы длительностью в *пикосекундном* диапазоне. Поразительно, что длительность этих импульсов оказывается на много порядков меньше времени жизни в используемом материале ($\sim 10^{-6}$ с). Отечественная промышленность давно выпускает диоды с накоплением заряда; так, для диодов типа КД524 и КД528 характерные времена выключения составляют 50–100 пс. Наиболее короткое время t_2, полученное в диодах с накоплением заряда, составляет 12 пс. Недостатком диодов с накоплением заряда является их сравнительно невысокое напряжение пробоя, составляющее 25–75 В. Недавно на специально выращенных структурах $p^+-\pi-\nu-n^+$ из GaAs с напряжением пробоя 220 В удалось получить рекордно высокую скорость нарастания напряжения ($\sim$2000 В/нс) [105].

Диоды с накоплением заряда находят применение в схемах генераторов импульсов наносекундной и субнаносекундной длительности, а также схемах умножения частоты [103]. Максимальная достигнутая в таких умножителях частота составляет $\sim$50 ГГц. Последнее применение связано с тем, что спектральный состав гармоник, возбуждаемый в диодах с накоплением заряда, оказывается существенно шире, чем в умножителях частоты на основе варакторов (см. с. 126).

1.7.5. Емкостная спектроскопия глубоких уровней. Как мы видели в п. 1.2.2, глубокие уровни, расположенные в запрещенной зоне, определяют такую важную характеристику полупроводника как время жизни. Поэтому развитие методов исследования этих уровней представляет огромное значение. В настоящее время для изучения глубоких уровней в полупроводниках широко используется *метод нестационарной спектроскопии глубоких уровней.* [1]) Основы этого метода были заложены Уильямсом [106], а Ланг [107] предложил удобную модификацию

[1]) В зарубежной литературе этот метод называется DLTS (deep-level transient spectroscopy).

этого метода, которая пригодна для изучения ловушек как для неосновных, так и основных носителей, причем работает и в случае одновременного присутствия в образцах нескольких глубоких уровней.

До сих пор, рассматривая барьерную емкость p–n-перехода, мы полагали, что толщина обедненного слоя определяется концентрацией мелких (водородоподобных) доноров и акцепторов, которыми легированы n- и p-области полупроводника. На самом деле толщина обедненного слоя определяется концентрацией *всех* заряженных уровней в p–n-переходе, в том числе и глубоких.

Метод DLTS основан на изучении кинетики изменения емкости p–n-перехода (или барьера Шоттки) при перезарядке (изменении заполнения электронами) глубоких уровней. Рассмотрим асимметрично легированный p^+–n-переход, в n-области которого кроме мелких доноров с концентрацией N_d есть еще глубокие уровни однозарядных акцепторов с концентрацией N_t (см. рис. 1.46,*а*), которые являются ловушками для неосновных носителей (дырок). Подадим на p–n-переход короткий импульс прямого смещения (так называемый *насыщающий импульс инжекции*). При этом дырки, инжектируемые в n-область диода, захватываются на глубокие акцепторы и изменяют их зарядовое состояние с A^- на A^0. Если после этого восстановить начальное смещение на p–n-переходе, то (поскольку полный заряд примесей, приходящийся на единицу объема n-области, увеличился с $N_d - N_t$ до N_d на величину заряда дырок, захваченных глубокими акцепторами) толщина области пространственного заряда уменьшится по сравнению с начальным значением, а емкость p–n-перехода — *увеличится*.

После восстановления начального смещения заполнение глубоких уровней начинает возвращаться к равновесному состоянию. Поскольку уровни, определяющие толщину области пространственного заряда, находятся в обедненном слое (концентрация свободных электронов и дырок в котором очень мала), то можно пренебречь захватом носителей на уровни и считать, что изменение заполнения уровней определяется только процессами выброса носителей. Для глубоких акцепторов *время релаксации заполнения уровня* связано с темпом теплового выброса дырок с акцепторного уровня e_p следующим соотношением [108]:

$$\tau_{\text{рел}} = \frac{1}{e_p} = \frac{1}{\upsilon_p \sigma_p N_t} \exp\left(\frac{E_t - E_v}{kT}\right), \qquad (1.137)$$

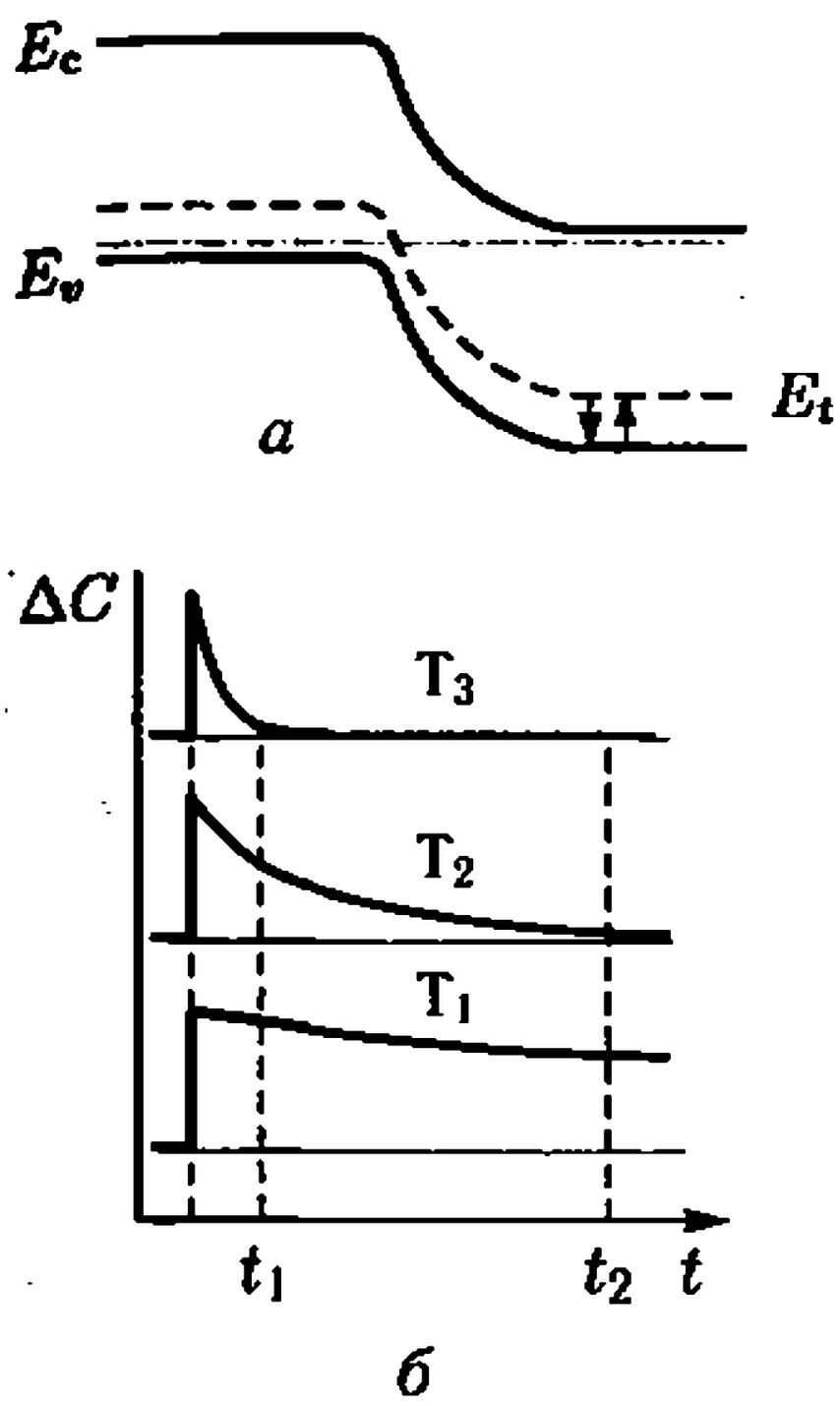


Рис. 1.46. *а* — энергетическая диаграмма p–n-перехода с глубоким акцепторным уровнем, *б* — релаксация сигнала DLTS после выключения импульса инжекции при разных температурах ($T_1 < T_2 < T_3$), *в* — спектры DLTS диодов из Si(V) для темпа выброса, равного $e_p = 87$ с$^{-1}$

где v_p — средняя скорость теплового движения дырок, σ_p — сечение захвата дырок на глубокий уровень, а E_t — энергетическое положение уровня в запрещенной зоне. По мере того, как дырки, захваченные на глубокие акцепторы, выбрасываются в валентную зону, заряд этих центров изменяется и, соответственно, изменяется и емкость p–n-перехода.

Изучая кинетику релаксации емкости после прекращения насыщающего импульса инжекции, можно определить основные характеристики глубоких уровней: их концентрацию, энергетическое положение и сечение захвата. Концентрация глубоких уровней рассчитывается из величины относительного изменения емкости $\Delta C_0/C_0$. Если считать, что за время насыщающего импульса инжекции глубокие акцепторы полностью заполняются дырками, а их концентрация $N_t \ll N_d$, то для резкого p^+–n-перехода из формулы (1.121) получаем

$$N_t \approx 2\frac{\Delta C_0}{C_0} N_d.$$

Для того, чтобы определить энергетическое положение и сечение захвата, измерения скорости релаксации емкости прово-

дятся при различных температурах (см. рис. 1.46 *б*). В соответствии с формулой (1.137), по наклону зависимости $\ln \tau_{\text{рел}}$ от $1/T$ определяется энергетическое положение уровня в запрещенной зоне, а по величине отсечки прямой при $1/T \to 0$ — сечение захвата.

При практической реализации метода DLTS обычно изучаются не кривые релаксации, а измеряется разность емкостей диода при двух значениях времени задержки после окончания импульса инжекции (t_1 и t_2 на рис. 1.46 *б*). Так как изменение емкости происходит приблизительно по экспоненциальному закону, то разность емкостей, измеренных в эти моменты времени, равна

$$\Delta C = \Delta C_0 \left[\exp\left(\frac{-t_1}{\tau_{\text{рел}}}\right) - \exp\left(\frac{-t_2}{\tau_{\text{рел}}}\right)\right],$$

где ΔC_0 — изменение емкости, вызванное импульсом инжекции. Поскольку время релаксации $\tau_{\text{рел}}$ быстро уменьшается с увеличением температуры, то зависимость $\Delta C(T)$ будет иметь вид кривой с максимумом при температуре, при которой эта разность максимальна, то есть когда

$$\tau_{\text{рел}} = \frac{t_2 - t_1}{\ln(t_2/t_1)}.$$

Записав серию кривых $\Delta C(T)$ при разных значениях t_1 и t_2 (например, увеличивая каждый раз значения обеих задержек в 10 раз), мы получим семейство кривых, максимум на которых будет систематически сдвигаться. Обработка этих кривых позволяет по амплитуде пика сделать вывод о концентрации глубоких уровней, а по смещению максимума с температурой — об энергетическом положении уровня и его сечении захвата. Этот подход оказывается особенно полезным для изучения полупроводников с несколькими глубокими уровнями, релаксация емкости в которых описывается суммой нескольких экспонент. Типичные спектры DLTS для кремниевых диодов, легированных ванадием, показаны на рис. 1.46 *в*. Наблюдаемый в них положительный пик C отвечает акцепторному уровню с энергией $E_t = E_v + 0{,}45$ эВ. Напомним, что положительные пики на кривых DLTS отвечают ловушкам для неосновных носителей.

При изучении параметров глубоких уровней, являющимися ловушками для *основных носителей* (уровнями донорного типа в p^+–n-переходе), необходимо иметь в виду, что эти уровни обычно лежат *ниже* уровня Ферми, они полностью заполнены

электронами и практически не захватывают дырки при их инжекции. [1]) Поэтому условия эксперимента надо немного изменить. В этом случае изучение релаксации емкости проводится при обратном смещении на p–n-переходе, а для изменения заполнения глубоких уровней смещение с p–n-перехода на некоторое время снимается. При этом в области пространственного заряда всегда существует слой, в котором глубокие донорные уровни при обратном смещении лежат выше квазиуровня Ферми, а при нулевом смещении — ниже него. При снятии смещения с p–n-перехода эти уровни перезаряжаются из состояния D^+ в состояние D^0 и поэтому после восстановления обратного смещения средний положительный заряд в единице объема n-области оказывается *меньше* на величину заряда электронов, захваченных на глубокие доноры. Это значит, что перезарядка донорных уровней будет проявляться в *уменьшении* емкости p–n-перехода, а в спектрах DLTS будут формироваться отрицательные пики (см. рис. 1.46*в*). Таким образом, ловушки для основных носителей проявляются в спектрах DLTS в виде отрицательных пиков.

Метод DLTS широко используется не только для изучения глубоких уровней, создаваемых в полупроводниках примесями, но также и для изучения уровней радиационных дефектов. Специфика этих дефектов состоит в том, что после облучения они начинают взаимодействовать между собой и с атомами примеси, образуя сложные комплексы, которые оказывают существенное влияние на свойства полупроводников. Понимание этих взаимопревращений позволяет указать пути повышения радиационной стойкости полупроводниковых материалов.

---

[1]) Напомним, что донорный или акцепторный характер глубокого уровня определяется не его энергетическим положением, а отношением сечений захвата электронов и дырок на этот уровень. Так, расположенные практически точно посередине запрещенной зоны в GaAs уровни атомов хрома оказываются акцепторами, а уровни атомов кислорода — донорами [10].

Глава 2

БИПОЛЯРНЫЕ ТРАНЗИСТОРЫ

2.1. Немного истории. Конструкции биполярного транзистора

Изобретение в 1947 г. группой сотрудников из Bell Laboratories транзистора — твердотельного прибора, способного усиливать электрические сигналы, и осознание возможности замены не очень надежных электронных ламп в радиоэлектронных устройствах более надежными транзисторами послужило толчком к исключительно бурному развитию всех направлений физики и технического применения полупроводников. На сегодняшний день транзистор остается одним из наиболее важных дискретных приборов и существенным компонентом интегральных схем. Число изготовленных к настоящему времени транзисторов оценивается фантастической цифрой — несколько единиц на 10^{18}, что составляет почти миллиард транзисторов на каждого жителя Земли. [1)]

Транзистор представляет собой трехэлектродный прибор, который может быть использован как усилитель или переключатель тока. В настоящее время существуют два больших класса транзисторов, в основе работы которых лежат различные

[1)] Эта цифра не будет казаться столь уж фантастической, если учесть, что выпускаемые в настоящее время двухъядерные процессоры фирм Intel и AMD содержат около 230 млн. транзисторов, а типичное для современных ЭВМ оперативное запоминающее устройство емкостью 512 Мбайт — более 4 млрд. транзисторов. По данным электронной биржи DRAMeXchange, в 2006 г. во всем мире было произведено 10,58 млрд. микросхем динамической памяти условной емкостью 256 Мбит, т. е. в запоминающих матрицах только одного этого типа микросхем содержалось $2{,}84 \cdot 10^{18}$ транзисторов.

физические принципы. Это — *биполярные* и *полевые* транзисторы. В этой главе мы рассмотрим физические принципы работы биполярных транзисторов, а работа полевых транзисторов будет рассмотрена в гл. 4. Хотя биполярные транзисторы появились первыми, в настоящее время их доля на рынке полупроводниковых приборов составляет около 5%.

Открытие транзистора было случайным. В 1946 г. в Bell Laboratories была организована группа, занимавшаяся фундаментальными исследованиями в области полупроводников, одним из руководителей которой был Уильям Шокли.

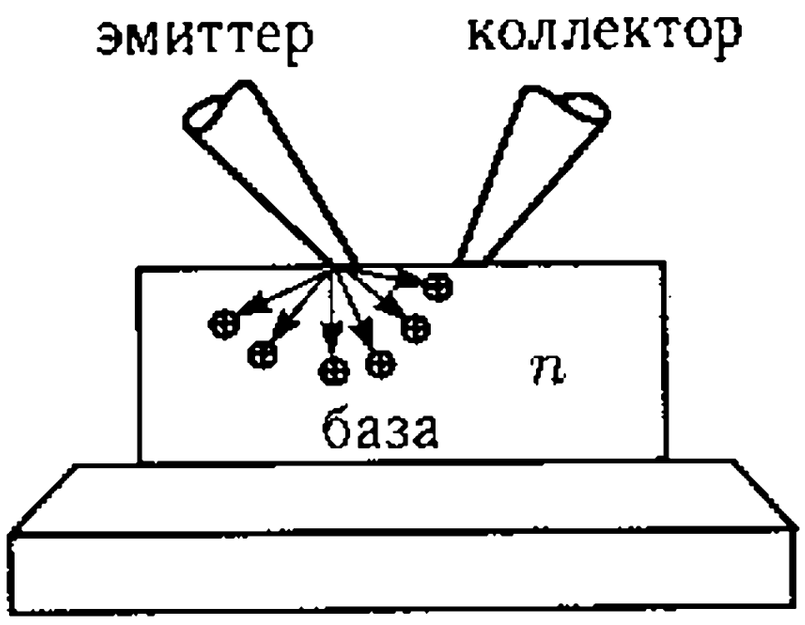
Сотрудники группы пытались создать полевой транзистор. После не очень успешных результатов по изучению эффекта поля в напыленных тонких пленках Ge и Si было решено попробовать реализовать идею полевого транзистора в тонких инверсионных слоях на поверхности этих полупроводников. Уолтер Браттейн и Джон Бардин (теоретик, обсуждавший получаемые в группе результаты) начали исследовать структуры с двумя контактами, созданными на поверхности полупроводника. Один из контактов в этой структуре использовался для управления толщиной инверсионного слоя, а с помощью другого (точечного) контакта измерялся ток носителей, возбуждаемых в этом слое. Пока в качестве первого контакта использовалась капля электролита, ток во втором контакте изменялся в соответствии с ожидаемой зависимостью для толщины инверсионного слоя. Однако когда электролит был заменен на металл, обнаружилось, что при подаче прямого смещения на этот контакт вместо уменьшения ток во втором контакте *увеличивался*.

Браттейн и Бардин поняли, что этот неожиданный эффект связан с инжекцией носителей первым контактом и собиранием их вторым контактом. После нескольких экспериментов они нашли конструкцию, в которой этот эффект был максимален. Это и был точечно-контактный транзистор [109].

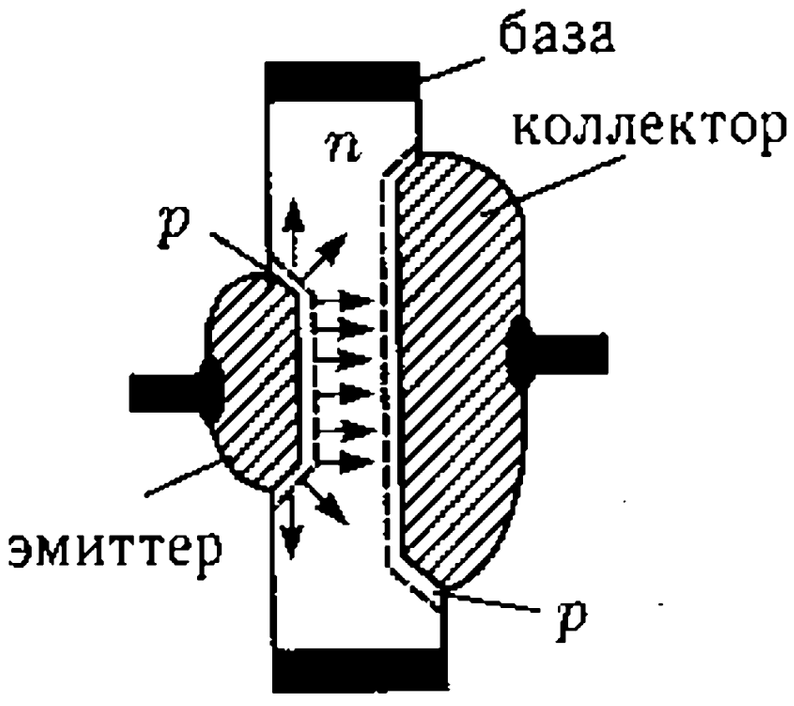
За исследования в области физики полупроводников и открытие транзисторного эффекта Браттейн, Бардин и Шокли были удостоены Нобелевской премии по физике в 1956 г.

Первым типом биполярного транзистора был *точечный (точечно-контактный) транзистор*, созданный в 1947 г. Этот транзистор представлял собой пластинку n-Ge, называемую *базой* (от англ. base — основание), на поверхности которой создавались два точечных контакта из вольфрама или фосфористой бронзы, расположенные рядом на расстоянии около 50 мкм (см. рис. 2.1 *а*). На один контакт, называемый *эмиттером* (от англ. emit — испускать), подавалось положительное (прямое) смещение относительно базы, а на другой контакт, называемый *коллектором* (от англ. collect — собирать), — отрицательное (обратное) смещение. Ток в цепи эмиттера оказывал сильное влияние на ток в цепи коллектора. Так, при достаточно высоком

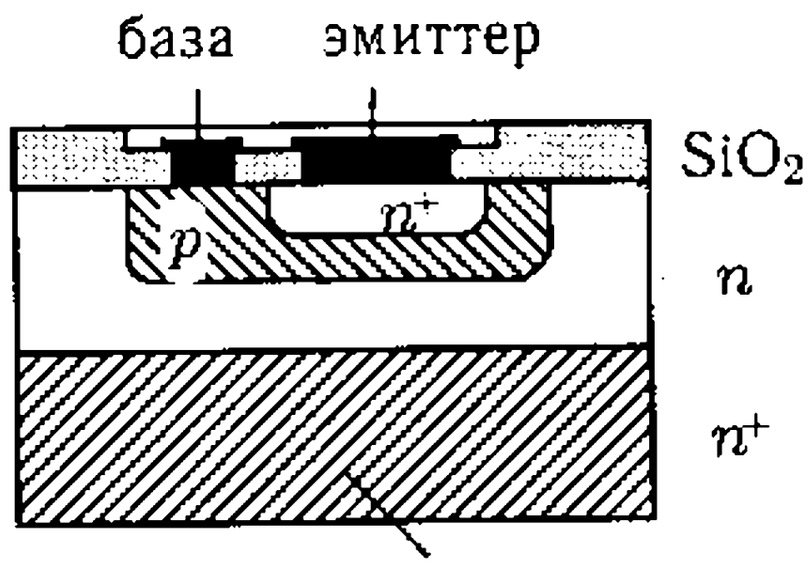
напряжении на коллекторе (5–25 В) в таком транзисторе удавалось получить *коэффициент усиления по току* $\alpha = dI_к/dI_э = 1–2$.



а



б



в

Рис. 2.1. Конструкции биполярных транзисторов: *а* — точечный транзистор, *б* — сплавной транзистор, *в* — эпитаксиальный транзистор

Возможность использования точечного транзистора для усиления электрических сигналов основана на существенном различии дифференциальных сопротивлений эмиттерного и коллекторного переходов в рабочем режиме. При подаче в цепь эмиттера небольшого входного переменного напряжения из-за низкого дифференциального сопротивления эмиттерного перехода при прямом смещении (сотни Ом) удается получить большие изменения тока эмиттера. Этот ток порождает изменение тока в коллекторной цепи, в α раз превышающее изменение тока эмиттера. Поскольку дифференциальное сопротивление коллекторного перехода, работающего при обратном смещении, велико (10–100 кОм), в цепь коллектора можно включить сравнимое с ним сопротивление нагрузки и получить на нем выходное напряжение, в сотни раз превышающее входное напряжение.

Работа точечного транзистора может быть упрощенно объяснена следующим образом. В области контакта металла со слабо легированным полупроводником n-типа образуются p–n-переходы. Эмиттерный p–n-переход, смещенный в прямом направлении, инжектирует дырки в базу транзистора. Если расстояние между эмиттером и коллектором меньше или порядка диффузионной длины, то часть инжектированных дырок, диффундируя в базе, достигает области коллектора и затягивается полем коллекторного p–n-перехода. Эта упрощенная схема однако не объясняет, почему в точечных транзисторах

удается получить коэффициент усиления по току $\alpha > 1$, так как в транзисторе обсуждаемой конструкции лишь малая доля инжектированных дырок может достигнуть коллекторного перехода. Считается, что в работе точечного транзистора заметную роль играет дрейф инжектированных в базу носителей в электрическом поле, создаваемом протекающим коллекторным током (база транзистора намеренно делается из достаточно высокоомного германия с $\rho = 5$–10 Ом · см), а структура коллекторного перехода (создаваемого формовкой) представляет собой n–p–n–p-структуру, в которой происходит внутреннее усиление коллекторного тока [110]. Из-за несовершенства конструкции и высокого уровня собственных шумов точечные транзисторы были полностью вытеснены плоскостны́ми транзисторами в середине 50-х годов.

Чтобы увеличить *коэффициент собирания носителей* (то есть долю инжектированных в базу носителей, которая движется в сторону коллектора), Шокли в 1949 г. предложил конструкцию и развил теорию *плоскостно́го транзистора* [3, 71, 110], в котором эмиттерный и коллекторный p–n-переходы расположены параллельно и разделены тонким слоем базы (см. рис. 2.2*а*). [1])

Идея конструкции плоскостного транзистора состоит в следующем. Смещенный в прямом направлении эмиттерный p–n-переход осуществляет одностороннюю инжекцию неосновных носителей в базу транзистора (см. рис. 2.2). Неосновные носители диффундируют через базу, толщина которой выбирается много меньше диффузионной длины, достигают коллекторного p–n-перехода, на который подано обратное смещение, и, затягиваясь его электрическим полем, попадают в область коллектора. Если база транзистора достаточно тонкая, то бóльшая часть инжектированных носителей проходит базу без рекомбинации и лишь небольшая их часть рекомбинирует в базе. Ток в цепи базы $I_б$ складывается из тока этих рекомбинирующих носителей,

---

[1]) Интересно, что первая статья Шокли была отклонена редакцией Physical Review и была опубликована в «ведомственном» журнале Bell Systems Technical Journal. Тем не менее предложенная идея оказалась очень плодотворной. Первые опытные образцы n–p–n-транзисторов были созданы в Bell Laboratories уже в 1950 году, а транзисторы со структурой p–n–p — в 1951 году [110]. Промышленный выпуск транзисторов был начат фирмой Texas Instruments в 1952 г., и транзисторы из Ge и Si появились на рынке уже в 1954 г. Сразу же, то есть в 1954 г., было выпущено и первое бытовое устройство на транзисторах — радиоприемник Regency TR-1, а первые компьютеры на транзисторах (TRIDAC, изготовленный в Bell Laboratories, и TX-0, изготовленный в Lincoln Laboratory, MIT) появились в 1955 г.

тока носителей, инжектируемых из базы в эмиттер, а также небольшого тока обратно смещенного коллекторного перехода. Важно то, что при оптимальной конструкции транзистора ток базы может быть сделан намного меньше тока $I_к$, протекающего в цепи коллектора, то есть с помощью небольшого тока базы можно управлять большим током коллектора. Поскольку, как и в случае точечного транзистора, дифференциальное сопротивление эмиттерного перехода при прямом смещении намного меньше дифференциального сопротивления коллекторного перехода при обратном смещении, то подавая небольшие напряжение и ток в цепь базы транзистора (относительно эмиттера), можно получать ток и напряжение в цепи нагрузки коллектора, намного превышающие входные ток и напряжение. Таким образом, плоскостной транзистор может одновременно служить и усилителем тока, и усилителем напряжения.

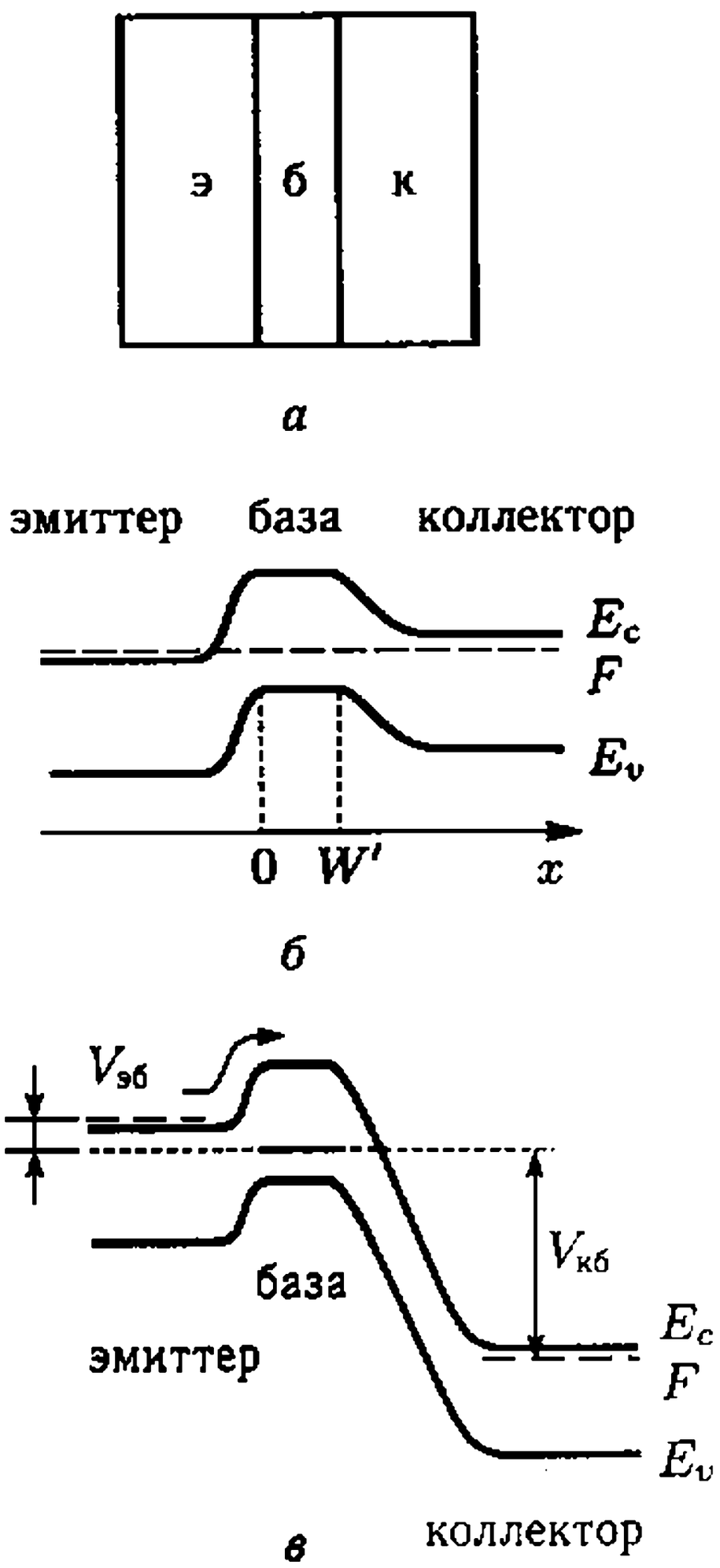


Рис. 2.2. Идеализированная схема плоскостного транзистора ($а$) и его энергетическая диаграмма в отсутствие смещения ($б$) и в рабочем режиме ($в$)

Одним из основных параметров, характеризующих усилительные свойства транзистора, является *коэффициент усиления по току* в схеме с общим эмиттером. [1] Он определяется как отношение приращения тока коллектора к вызвавшему его приращению тока базы: $\beta = dI_к/dI_б$. [2] Ясно, что коэффициент усиления будет тем выше, чем больше инжектированных эмиттером носителей пройдет через базу без рекомбинации. Поэтому очень важно создать такую геометрию прибора, при которой практически все инжектированные в базу носители собирались бы коллектором. В этом отношении

---

[1] О схемах включения транзисторов мы будем говорить в п. 2.5.1.

[2] В ряде случаев усилительные свойства транзистора характеризуют также величиной $\beta_{ст} = I_к/I_б$, называемой *статическим* коэффициентом усиления транзистора по току.

геометрия плоскостного транзистора оказывается намного лучше геометрии точечного транзистора (ср. рис. 2.1 *а* и рис. 2.1 *б*).

В зависимости от того, какой тип проводимости имеет база транзистора, возможны две структуры биполярных транзисторов: *p–n–p* и *n–p–n*. Поскольку подвижности электронов и дырок в полупроводнике различаются, то и параметры этих транзисторов также будут различными. В частности, можно ожидать, что из-за более высокой подвижности электронов (и, следовательно, их более высокого коэффициента диффузии) приборы, в которых осуществляется инжекция электронов (транзисторы со структурой *n–p–n*), будут характеризоваться более высоким коэффициентом усиления и бо́льшим быстродействием.

Идея плоскостного транзистора была первоначально реализована в конструкции *сплавного транзистора*, [1]) в котором эмиттерный и коллекторный переходы создавались вплавлением капелек металла с двух сторон в тонкую пластинку полупроводника (см. рис. 2.1 *б*).

Сплавные германиевые *p–n–p*-транзисторы, примером которых могут служить отечественные приборы серий П13–П16, МП35–МП42, П210, П213–П217 и др., изготавливались вплавлением капелек индия (в мощных транзисторах — сплава In+Ga) в пластинки *n*-Ge толщиной около 200 мкм при температуре 500 °С в вакууме или атмосфере водорода, а для создания *n–p–n*-транзисторов использовались, соответственно, *p*-Ge и сплав Pb+Sb. При нагревании пластинки металлический сплав растворяет часть расположенного под ним германия, а после охлаждения слои рекристаллизованного германия оказываются сильно легированными индием, галлием или сурьмой, образуя эмиттерную и коллекторную области. В полученных таким образом транзисторах минимальная толщина базы составляет ~10 мкм, а характерные диаметры эмиттера и коллектора — 0,35–0,5 мм и 0,7–1 мм. Заметим, что диаметр коллектора в сплавных транзисторах намеренно делается больше диаметра эмиттера (см. рис. 2.1 *б*), чтобы увеличить коэффициент собирания носителей.

Из-за невозможности технологически создать очень тонкую базу в сплавных транзисторах их рабочая частота обычно не превышала 5–10 МГц. Необходимость повышения быстродействия транзисторов привела к разработке *дрейфового транзистора* (см. подробнее в п. 2.2.2), в котором за счет неоднородного легирования в базе транзистора создается *встроенное*

---

[1]) Технология создания сплавных *p–n*-переходов была разработана в начале 50-х годов в фирмах General Electric и RCA. По этой технологии в одной только фирме Texas Instruments было изготовлено более 10^9 транзисторов.

электрическое поле. [1]) Дрейф в этом поле ускоряет перенос инжектированных носителей через базу и тем самым увеличивает быстродействие и коэффициент усиления транзистора. Одним из способов создания дрейфовых транзисторов является технология *диффузионно-сплавных транзисторов*, которая широко использовалась в производстве таких отечественных транзисторов как П401–П403, П416, ГТ308, ГТ313, ГТ322, ГТ806, ГТ905 и др.

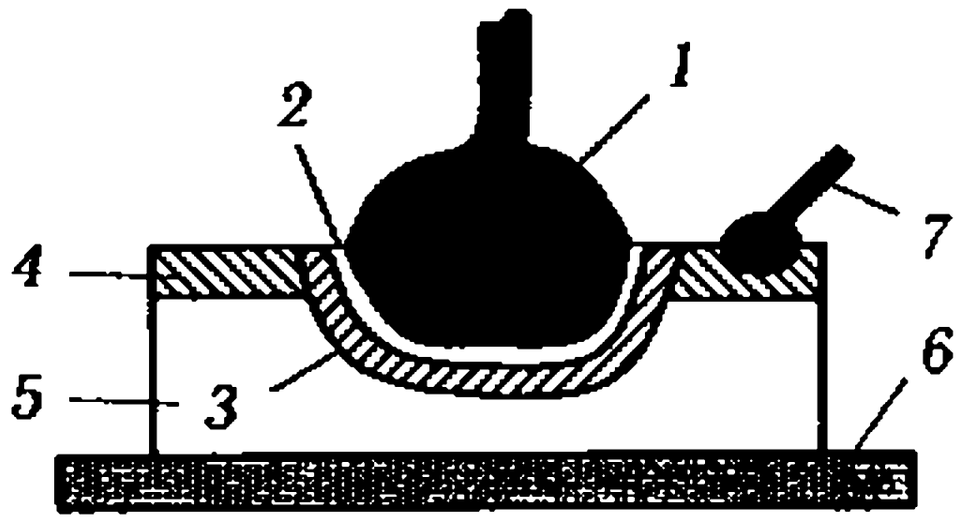


Рис. 2.3. Конструкция диффузионно-сплавного транзистора. *1* — вывод эмиттера, *2* — сильно легированная область эмиттера, *3* — база, *4* — диффузионный слой *n*-типа, *5* — коллектор, *6* — держатель и вывод коллектора, *7* — вывод базы

В этом способе в пластинку *p*-Ge с удельным сопротивлением около 1 Ом·см, на поверхности которого диффузией Sb предварительно создан тонкий *n*-слой, вплавляют металлический сплав Pb+2 % Ga+ +1 % Sb, содержащий *одновременно* и акцепторную (Ga), и донорную (Sb) примеси (см. рис. 2.3). После выдержки пластины в нагретом состоянии при высокой температуре (500 °C, 15–30 минут) и последующего охлаждения в кристалле оказываются сформированными сильно легированная рекристаллизованная область *p*-типа, служащая эмиттером, и тонкий слой базы *n*-типа, возникший за счет быстрой диффузии сурьмы во время выдержки пластины при высокой температуре. Этот слой непосредственно контактирует со слоем *n*-типа на поверхности, который служит электрическим контактом к базе. После вплавления и изготовления контакта к базе лишние части поверхностного *n*-слоя для уменьшения емкости коллекторного перехода удаляются травлением. Рабочая частота полученных таким образом транзисторов достигает ~500 МГц, что существенно выше, чем в сплавных транзисторах.

При изготовлении транзисторов по диффузионно-сплавной технологии необходимо иметь в виду, что в Ge коэффициенты диффузии доноров V группы заметно выше коэффициентов диффузии акцепторов III группы, поэтому таким способом можно создавать только *p–n–p*-транзисторы. В Si соотношение коэффициентов диффузии обратное, и поэтому диффузионно-сплавная технология годится лишь для создания *n–p–n*-транзисторов. Примером кремниевых транзисторов, изготовленных по этой технологии, являются отечественные транзисторы П504 и П505.

---

[1]) Идея дрейфового транзистора была высказана Крёмером в 1953 г. За большой вклад в развитие современной полупроводниковой электроники Герберт Крёмер был одним из ученых, удостоенных Нобелевской премии по физике в 2000 г.

Сплавные и диффузионно-сплавные транзисторы имеют ряд недостатков. Чтобы рабочее напряжение транзистора было высоким, область коллектора необходимо изготавливать из достаточно высокоомного материала. Так как при этом требование механической прочности не позволяет сделать кристалл тоньше ~100 мкм, сопротивление толщи коллектора оказывается достаточно большим, что препятствует созданию мощных транзисторов. Кроме того, довольно длинный путь тока в базе не позволяет получить низкое сопротивление слоя базы, а это, как будет показано в п. 2.4, сильно ограничивает высокочастотные свойства транзистора.

Чтобы уменьшить сопротивление коллектора, в 1960 г. была предложена конструкция так называемых *эпитаксиальных транзисторов*, в которой на достаточно толстой подложке из низкоомного материала (n-Si с удельным сопротивлением ~0,01 Ом·см) выращивается эпитаксиальный слой высокоомного материала, в котором собственно и располагается область пространственного заряда коллекторного перехода (см. рис. 2.1 *в*). Уровень легирования и толщина этого слоя (2–10 мкм) выбираются исходя из максимального рабочего напряжения транзистора. В этом слое путем последовательной диффузии акцепторной и донорной примесей через одну и ту же поверхность (*двойная диффузия*) создаются базовая и эмиттерная области, причем профиль распределения примеси в базе таков, что в ней возникает встроенное электрическое поле, ускоряющее перенос носителей через базу. Заметим, что для транзисторов, изготовленных описанным способом, коэффициент собирания носителей близок к идеальному значению — 100%, поскольку эмиттер и база транзистора полностью «погружены» в коллектор (см. рис. 2.1 *в*).

Технология создания эпитаксиальных транзисторов тесно связана с *планарной технологией*, позволяющей одновременно создавать огромное количество транзисторов на поверхности пластины полупроводника; об этой технологии мы будем говорить подробнее в п. 2.8.1. Эпитаксиально-планарная технология в настоящее время является основой производства дискретных транзисторов из кремния и германия. Существенно более низкое сопротивление коллектора и возможность создания на поверхности полупроводника сложной конфигурации контактов предопределяют широкое использование этой технологии для создания мощных и высокочастотных транзисторов. Примерами отечественных транзисторов, изготовленных по эпитаксиально-планарной технологии, могут служить транзисторы КТ315, КТ342, КТ361, ГТ346, КТ3102, КТ603.

Поскольку быстродействие транзисторов в значительной степени ограничивается емкостью коллекторного перехода, дополнительным способом увеличения их быстродействия является создание так называемых *меза-структур.* [1]) В этом подходе после формирования в эпитаксиальном слое областей базы и эмиттера лишние части коллекторного перехода удаляются травлением (см. рис. 2.4).

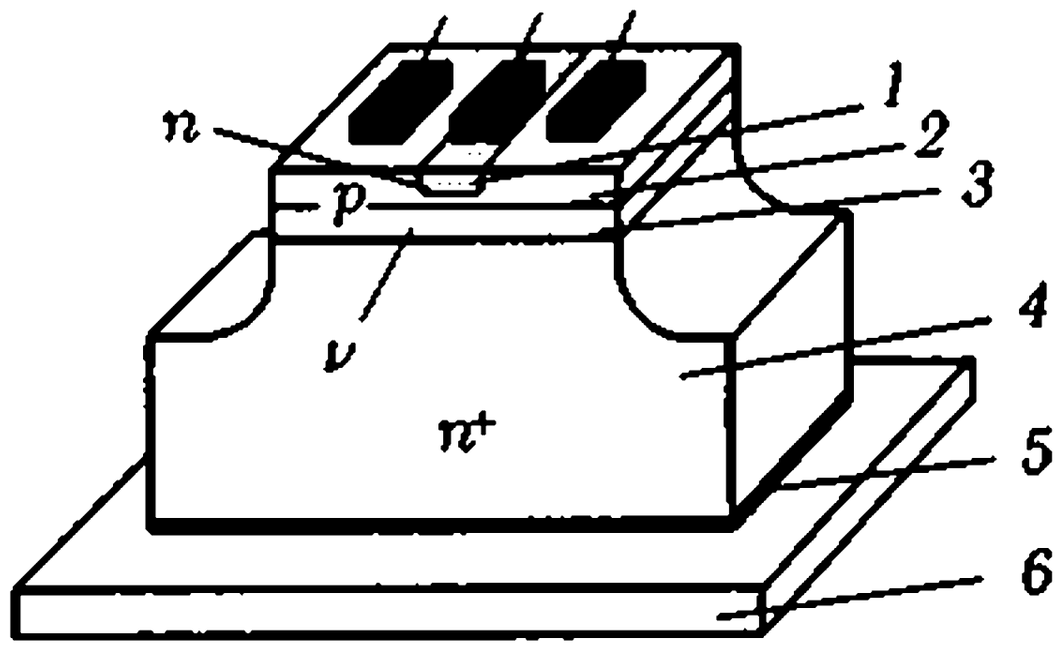


Рис. 2.4. Конструкция эпитаксиального транзистора с меза-структурой. *1*, *2* — диффузионные области эмиттера и базы, созданные в эпитаксиальном слое, *3* — эпитаксиальный слой ν-типа (высокоомная часть коллектора), *5* — омический контакт, *6* — держатель кристалла

Таким образом создаются отечественные транзисторы КТ814–КТ817, КТ903, КТ940.

2.2. Параметры, определяющие коэффициент усиления транзистора

Из описанного выше принципа действия биполярного транзистора следует, что его коэффициент усиления зависит от трех факторов: эффективности инжекции неосновных носителей из эмиттера в базу, эффективности переноса носителей через базу и эффективности собирания продиффундировавших носителей коллектором. Последнее, как мы видели в п. 2.1, определяется геометрией и взаимным расположением эмиттера и коллектора в транзисторе. В этом разделе мы подробно проанализируем, от каких параметров полупроводника и размеров структуры зависят первые два фактора — эффективность инжекции эмиттера и эффективность переноса носителей через базу.

---

[1]) Название структуры происходит от испанского слова mesa — столовая гора. Так геологи называют горы с плоской вершиной, распространенные на юго-западе США, — там, где создавалась «кремниевая долина».

2.2.1. Коэффициент инжекции эмиттера. Для определенности рассмотрим транзистор со структурой n–p–n, находящийся в рабочем режиме, когда на эмиттерный переход подано прямое смещение $V_{бэ} > 0$, а на коллекторный — обратное смещение $V_{бк} < 0$ (см. рис. 2.2 *в*). Чтобы рассчитать коэффициент инжекции эмиттера, нам надо найти стационарное распределение инжектированных носителей в базе транзистора. При низком уровне инжекции это распределение является решением уравнения непрерывности,

$$D_n \frac{d^2 n_б}{dx^2} - \frac{n_б - n_{б0}}{\tau_n} = 0, \qquad (2.1)$$

с граничными условиями

$$n_б(0) = n_{б0} \exp\left(\frac{qV_{бэ}}{kT}\right),$$

на краю эмиттерного p–n-перехода, и

$$n_б(W') = n_{б0} \exp\left(\frac{qV_{бк}}{kT}\right) \approx 0,$$

на краю коллекторного перехода. Здесь $n_{б0} = n_i^2/N_{аб}$ — равновесная концентрация неосновных носителей в базе транзистора, а $N_{аб}$ — концентрация акцепторов в базе. Указанные граничные условия похожи на условия, которые мы использовали при изучении короткого диода (см. п. 1.2.1), и отличаются от них только вторым граничным условием, которое имеет такой вид потому, что на коллектор подано обратное смещение ($V_{бк} < 0$) и он вытягивает все неосновные носители, приближающиеся к коллекторному переходу. Через W' мы обозначили толщину *нейтральной* области базы, которая, очевидно, меньше геометрической толщины базы W на толщину располагающихся в ней областей пространственного заряда эмиттерного и коллекторного p–n-переходов. Таким образом, W' зависит от напряжения смещения на этих переходах. [1])

Решение уравнения (2.1) с указанными граничными условиями имеет вид

$$n_б(x) = n_{б0}\left\{1 - \frac{\mathrm{sh}(x/L_n)}{\mathrm{sh}(W'/L_n)} + \right.$$
$$\left. + \frac{\mathrm{sh}[(W'-x)/L_n]}{\mathrm{sh}(W'/L_n)}\left[\exp\left(\frac{qV_{бэ}}{kT}\right) - 1\right]\right\}, \qquad (2.2)$$

[1]) Как мы покажем в п. 2.2.5, следствием этой зависимости будет изменение коэффициента усиления транзистора при изменении напряжения на коллекторе.

где $L_n = \sqrt{D_n \tau_n}$ — диффузионная длина электронов. В практически важном случае тонкой базы ($W' \ll L_n$) это решение сводится к следующему:

$$n_б(x) \approx n_{б0} \frac{W' - x}{W'} \exp\left(\frac{qV_{бэ}}{kT}\right). \qquad (2.3)$$

В этом случае плотность диффузионного тока электронов, текущего из эмиттера в базу, определяется уравнением, похожим на уравнение (1.33) для диффузионного тока в p–n-переходе с короткой базой:

$$J_{nэ} = qD_n \frac{dn}{dx} \approx -\frac{qD_n n_{б0}}{W'} \exp\left(\frac{qV_{бэ}}{kT}\right). \qquad (2.4)$$

Отрицательный знак тока означает, что ток течет в направлении, противоположном направлению оси x, в соответствии с полярностью приложенного к эмиттерному переходу смещения (плюс на базу транзистора).

Из базы в эмиттерную область толщиной $x_э \ll L_p$ течет ток дырок, плотность которого определяется уравнением (1.33):

$$J_{pэ} = -qD_p \frac{dp_э}{dx} \approx -\frac{qD_p p_{э0}}{x_э} \exp\left(\frac{qV_{бэ}}{kT}\right). \qquad (2.5)$$

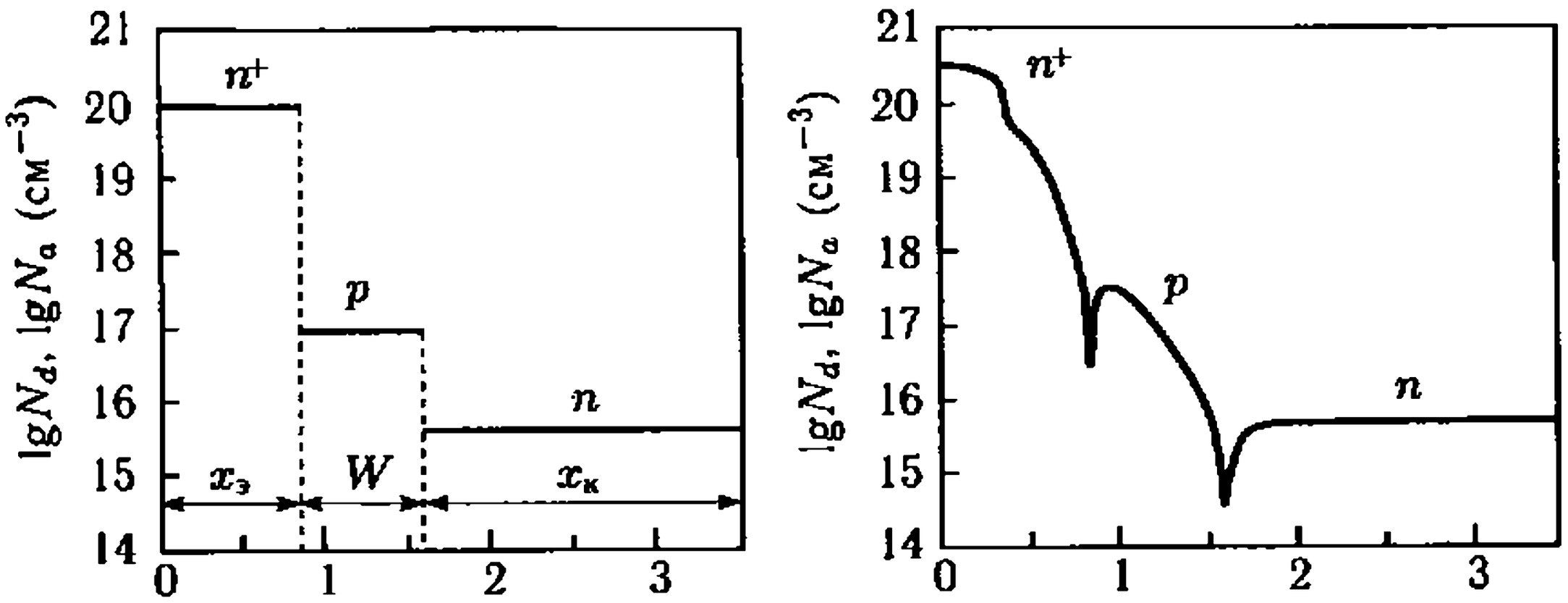


Рис. 2.5. Профили легирования n–p–n-транзистора: *а* — идеальный профиль, *б* — реальный профиль дрейфового транзистора

Отношение тока носителей, инжектируемых из эмиттера в базу, к полному току эмиттерного p–n-перехода называют *коэффициентом инжекции эмиттера* и обозначают буквой γ. Если

пренебречь всеми другими (генерационно-рекомбинационным, туннельным) механизмами протекания тока через переход и рассматривать только диффузионную компоненту, то из уравнений (2.4) и (2.5) следует, что величина γ определяется соотношением концентраций примесей в эмиттере ($N_{d\text{э}}$) и базе ($N_{a\text{б}}$), соотношением коэффициентов диффузии электронов и дырок (D_n, D_p) и соотношением толщин базы и эмиттера (W', $x_\text{э}$):

$$\gamma = \frac{J_{n\text{э}}}{J_{n\text{э}} + J_{p\text{э}}} = \frac{1}{1 + \dfrac{D_p W' N_{a\text{б}}}{D_n x_\text{э} N_{d\text{э}}}}. \qquad (2.6)$$

При конструировании транзисторов толщины эмиттера и базы и их уровни легирования подбираются таким образом, чтобы сделать коэффициент инжекции как можно более высоким. Как следует из формулы (2.6), для этого эмиттер должен быть легирован значительно сильнее, чем база (см. рис. 2.5). Произведение $W' N_{a\text{б}}$, стоящее в этой формуле и представляющее собой полное число акцепторов в базовой области единичной площади, называют *числом Гуммеля* $Q_\text{Г}$. Хотя соотношение (2.6) было получено нами для однородно легированной базы, можно показать [14], что оно справедливо и для произвольного (неоднородного) профиля легирования базы, если вместо $W' N_{a\text{б}}$ в нем использовать число Гуммеля

$$Q_\text{Г} = \int_0^{W'} N_{a\text{б}}(x)\, dx.$$

Поскольку в современных транзисторах коэффициент усиления транзистора по току $\beta = dI_\text{к}/dI_\text{б}$ определяется в основном коэффициентом инжекции γ, то величина β изменяется пропорционально $Q_\text{Г}^{-1}$.

Казалось бы, как следует из формулы (2.6), для увеличения γ следует как можно сильнее легировать область эмиттера. Однако эксперимент показывает, что при высоком уровне легирования эмиттера γ начинает уменьшаться и связано это с проявлением двух побочных физических явлений.

Первое из этих явлений — *сужение запрещенной зоны в сильно легированном полупроводнике*. [1]) Расчеты в работе [111],

---

[1]) Этот эффект возникает из-за того, что в полупроводнике с высокой концентрацией свободных носителей собственная электростатическая энергия заряженных частиц из-за экранирования оказывается меньше, чем в отсутствие экранирования. Поэтому для рождения электронно-дырочной пары в таком

неплохо согласующиеся с экспериментом, показывают, что величина этого сужения составляет

$$\Delta E_g = \frac{3q^2}{4\varepsilon L_D},$$

где L_D — дебаевская длина экранирования (см. формулу на с. 16). При $N_d = 10^{19}$ см$^{-3}$ в кремнии $\Delta E_g \approx 70$ мэВ. В результате этого эффекта эмиттерный переход по своим свойствам становится похожим на гетеропереход, в котором с ростом концентрации примеси $N_{dэ}$ облегчаются условия для нежелательного направления инжекции — из базы в эмиттер, и формула (2.6) будет иметь вид

$$\gamma = \frac{J_{nэ}}{J_{nэ} + J_{pэ}} = \frac{1}{1 + \dfrac{D_p W' N_{аб}}{D_n x_э N_{dэ}} \exp\left(\dfrac{\Delta E_g}{kT}\right)}. \qquad (2.7)$$

Второе явление состоит в уменьшении времени жизни в сильно легированном эмиттере из-за проявления *Оже-рекомбинации.* [1]) С увеличением уровня легирования диффузионная длина дырок в эмиттере уменьшается и когда она сравнивается с толщиной эмиттера, дырочная составляющая тока насыщения начинает быстро возрастать. Максимальное значение γ в кремниевых приборах достигается при уровне легирования эмиттера $\sim 10^{19}$ см$^{-3}$. Поэтому при создании транзисторов со сверхвысоким коэффициентом усиления ($\beta = 1000$ и выше) увеличивать γ можно лишь уменьшая концентрацию легирующих примесей в базе транзистора.

Отметим, что на коэффициент инжекции эмиттера сильное влияние оказывают также материал и способ создания *контакта*

полупроводнике требуется *меньшая энергия*, что и отражается в уменьшении E_g [111].

[1]) Напомним, что Оже-рекомбинация является процессом, при котором энергия, выделяющаяся при рекомбинации электронно-дырочной пары, передается третьей частице — электрону или дырке. Этот процесс по сути является обратным процессу ударной ионизации (здесь рекомбинирующие электрон и дырка рождают горячий носитель, а там горячий носитель рождал электронно-дырочную пару). Темп Оже-рекомбинации пропорционален вероятности одновременной встречи трех частиц, т. е. $R \sim n^2 p$ в образцах n-типа и $R \sim np^2$ в образцах p-типа, откуда следует, что времена жизни неравновесных носителей изменяются как $\tau_p \sim 1/n^2$ и $\tau_n \sim 1/p^2$. В кремнии с концентрацией носителей 10^{19} см$^{-3}$ времена жизни относительно Оже-рекомбинации составляют $\tau_n \approx 10^{-7}$ с и $\tau_p \approx 4 \cdot 10^{-8}$ с [87].

к эмиттеру. [1]) Дело в том, что величина $J_{рэ}$ зависит от граничных условий на контакте, которые различны для разных способов изготовления контактов. Так, при создании контактов к эмиттеру из силицидов металлов (см. с. 99) величина γ обычно уменьшается, поскольку некоторая часть кремния из области эмиттера уходит на химическую реакцию образования силицида и толщина эмиттера $x_э$ уменьшается. При использовании контактов из сильно легированного поликристаллического кремния, наоборот, величина γ обычно возрастает, поскольку граница, на которой концентрация неравновесных носителей обращается в нуль, отодвигается от p–n-перехода. Поэтому в транзисторах с контактами из поликристаллического кремния удается получить коэффициенты усиления, в 3–7 раз более высокие, чем в транзисторах с обычными контактами [58].

Среди других способов увеличения коэффициента инжекции эмиттера следует отметить использование в качестве эмиттера гетероперехода или варизонной структуры, для которых характерен односторонний характер инжекции (см. п. 1.6.1). Реализовать эту идею долгое время мешала довольно высокая плотность дефектов на гетерогранице, которая приводила к большому генерационно-рекомбинационному току и препятствовала эффективной инжекции. Тем не менее, в опытных образцах этих так называемых *гетеропереходных транзисторов* в системе $GaAs$–$Al_xGa_{1-x}As$ удалось получить коэффициент усиления $\beta \approx 50000$ [112]. В настоящее время налажен широкий промышленный выпуск СВЧ транзисторов на основе гетеропереходов Si–$Si_{1-x}Ge_x$, $Al_xGa_{1-x}As$–$GaAs$, $In_{1-x}Ga_xP$–$GaAs$ и InP–$In_xGa_{1-x}As$ (см. подробнее п. 2.4.2).

Другая интересная идея, обсуждаемая Шуром [87], заключается в использовании для создания эффективной инжекции различие в вероятности туннелирования электронов и дырок через потенциальный барьер на границе эмиттера и базы (так называемый *биполярный транзистор с туннельным эмиттером*). По-видимому впервые эта идея была реализована де Граафом и де Гроотом в 1979 г.; авторы при изготовлении кремниевых транзисторов вводили между контактом из поликристаллического кремния и монокристаллическим эмиттером тонкий (20–40 Å) слой диэлектрика, толщина которого подбиралась так, чтобы падение напряжения на барьере не превышало kT. Этот барьер эффективно пропускал электроны с малой эффективной массой, но практически не пропускал дырки с большой эффективной массой, что приводило к существенному увеличению эффективности инжекции эмиттера [62].

---

[1]) При создании интегральных схем материал контакта может одновременно использоваться и для осуществления всех необходимых соединений между элементами.

2.2.2. Коэффициент переноса носителей через базу. Другим фактором, влияющим на коэффициент усиления транзистора, является эффективность переноса носителей через базу. Эта величина характеризуется *коэффициентом переноса носителей через базу* α_T, который определяется как отношение приращения потока инжектированных носителей на выходе базы (при $x = W'$) к вызвавшему его приращению потока на ее входе (при $x = 0$). Дифференцируя полученное выше выражение для распределения электронов в базе (уравнение (2.2)), находим [1]

$$\alpha_T \equiv \frac{\Delta J_n(W')}{\Delta J_n(0)} = \frac{1}{\operatorname{ch}(W'/L_n)} \approx 1 - \frac{1}{2}\left(\frac{W'}{L_n}\right)^2. \qquad (2.8)$$

Коэффициент переноса через базу оказывал существенное влияние на коэффициент усиления только в первых конструкциях транзисторов, в которых база была достаточно толстой. В современных транзисторах с тонкой базой коэффициент α_T очень близок к единице. Например, для $W' = 1$ мкм и $L_n = 30$ мкм из формулы (2.8) получаем $\alpha_T = 0,9995$. Поэтому можно считать, что в современных приборах коэффициент переноса носителей через базу не является фактором, ограничивающим коэффициент усиления транзистора.

Как мы говорили в п. 2.1, для увеличения коэффициента усиления и быстродействия в свое время была предложена конструкция так называемого *дрейфового транзистора*, в котором за счет специфического профиля легирования базы, создаваемого при изготовлении транзистора, в базе возникает *встроенное электрическое поле*. Дрейф в этом поле убыстряет прохождение инжектированных носителей через базу и тем самым дополнительно увеличивает коэффициент α_T, еще более приближая его к единице. [2] Типичный профиль легирования базы дрейфового транзистора показан на рис. 2.5*б*. Создать такой профиль можно различными способами, например, путем одновременной диффузии доноров и акцепторов в полупроводник, используя различие коэффициентов диффузии примесей (об этом методе, называемом двойной диффузией, мы уже говорили на с. 147); в качестве

---

[1]) Определение коэффициента α_T через приращения необходимо, чтобы исключить стационарную неоднородность в распределении неосновных носителей в базе, которая возникает даже в отсутствие инжекции из-за того, что на коллектор подано отрицательное смещение и из базы постоянно происходит экстракция неосновных носителей.

[2]) Оценка степени влияния встроенного электрического поля на движение носителей будет дана нами на с. 166.

примесей для создания кремниевых транзисторов обычно используют As, P (доноры) и B (акцептор). [1]) В настоящее время такой профиль легирования создается практически во всех биполярных транзисторах. Его основной целью является увеличение быстродействия транзисторов (о высокочастотных свойствах транзисторов мы будем говорить в п. 2.4).

2.2.3. Эффективность инжекции при очень малых и очень больших токах. До сих пор мы полагали (следуя Шокли [3]), что весь ток через эмиттерный p–n-переход определяется инжекцией. Однако, как было показано в п. 1.2.2, в полупроводниках с достаточно широкой запрещенной зоной (в частности, в Si) это не так, и при малых прямых напряжениях смещения ток в p–n-переходе определяется рекомбинацией в области пространственного заряда. Поскольку этот ток *не сопровождается инжекцией* носителей, то существование этой компоненты тока в токе эмиттерного перехода будет уменьшать коэффициент инжекции эмиттера γ и, в конечном счете, снижать коэффициент усиления транзистора β [6]. Описать влияние тока рекомбинации на коэффициент инжекции мы можем, добавив его плотность $J_{\text{рек}}$ в знаменатель формулы (2.6).

При малых напряжениях смещения $V_{\text{бэ}}$ зависимость статического коэффициента усиления $\beta_{\text{ст}}$ от тока коллектора можно легко рассчитать. Учитывая, что $J_{\text{рек}} \gg J_{n\text{э}}, J_{p\text{э}}$, а токи коллектора и базы равны

$$J_{\text{к}} \approx \alpha_T J_{n\text{э}} \approx J_{n\text{э}}, \quad J_{\text{б}} = J_{\text{рек}} + (1 - \alpha_T J_{n\text{э}}) + J_{p\text{э}} \approx J_{\text{рек}},$$

из найденных нами ранее зависимостей этих токов от напряжения смещения (формулы (1.29) и (1.39)),

$$J_{\text{рек}} \sim \exp\left(\frac{qV_{\text{бэ}}}{mkT}\right), \quad J_{n\text{э}} \sim \exp\left(\frac{qV_{\text{бэ}}}{kT}\right), \tag{2.9}$$

следует, что

$$\beta_{\text{ст}} = \frac{J_{\text{к}}}{J_{\text{б}}} \approx \frac{J_{n\text{э}}}{J_{\text{рек}}} \sim \exp\left[\frac{qV_{\text{бэ}}}{kT}\left(1 - \frac{1}{m}\right)\right] \sim J_{\text{к}}^{(1-1/m)}. \tag{2.10}$$

---

[1]) Каждая из этих примесей имеет свои индивидуальные особенности диффузии, которые необходимо учитывать при изготовлении приборов. Так, появление ступеньки в области высокой концентрации примеси (см. рис. 2.5 *б*) является характерной особенностью диффузии фосфора и связано с образованием комплексов примесь–вакансия в приповерхностном слое кремния.

Такая зависимость $\beta(J_к)$ действительно наблюдается в биполярных транзисторах в области малых токов (см. рис. 2.6).

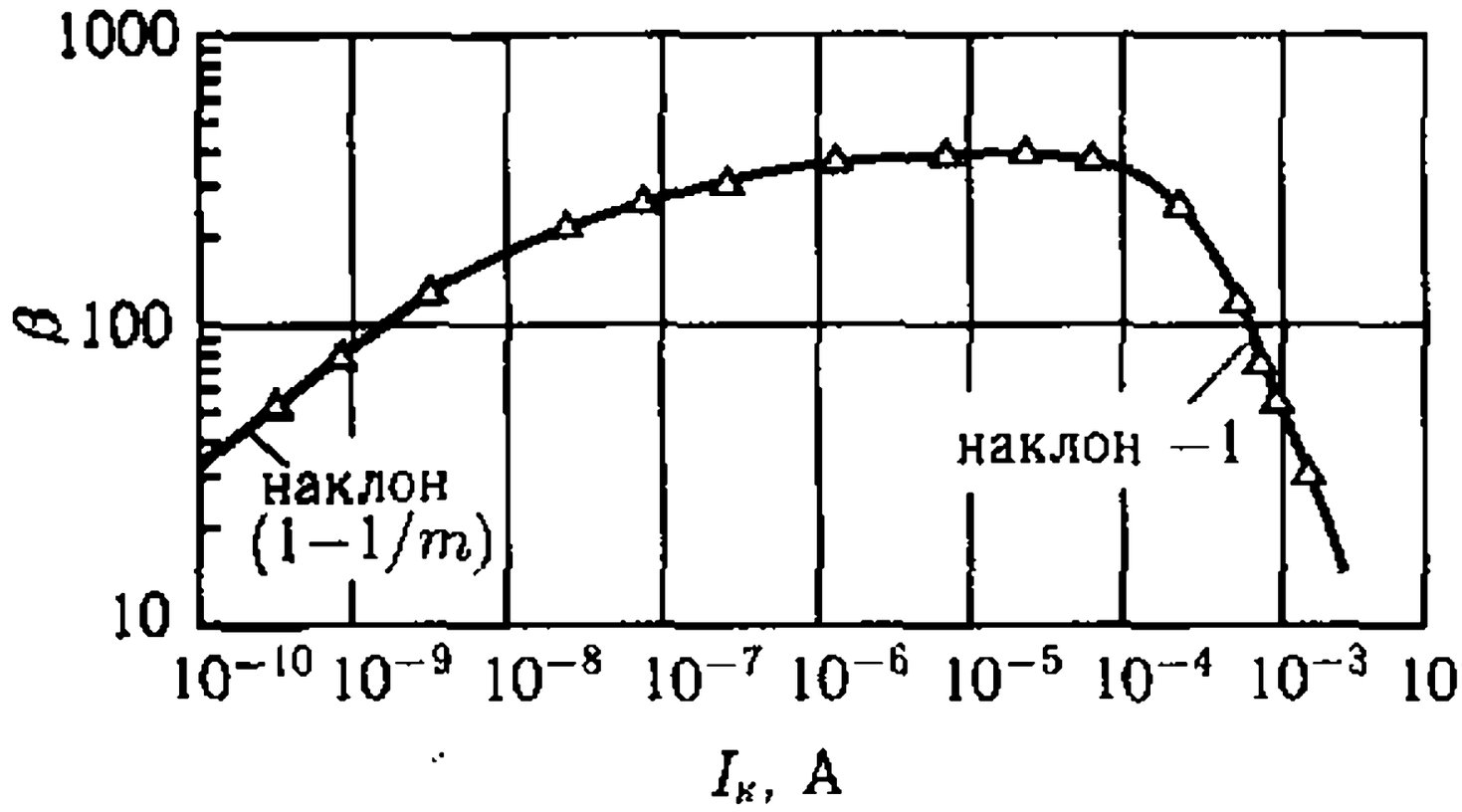


Рис. 2.6. Зависимость коэффициента усиления транзистора β от тока коллектора [14]

При высоком уровне инжекции эффективность инжекции эмиттера также не остается постоянной. Когда концентрация инжектированных в базу носителей оказывается выше концентрации примесей в базе ($n_б(0) > N_{аб}$), для компенсации электрического заряда инжектированных электронов в базу из контакта подходят дырки. Как мы видели в п. 1.2.3, в этом случае ток инжекции электронов из эмиттера в базу (и, следовательно, ток коллектора) изменяется с напряжением как $\exp(qV_{бэ}/2kT)$, а ток базы, который при $\alpha_T \approx 1$ определяется током инжекции дырок из базы в эмиттер, продолжает изменяться как $\exp(qV_{бэ}/kT)$. Поэтому

$$\beta_{ст} = \frac{J_к}{J_б} \sim \exp\left(-\frac{qV_{бэ}}{2kT}\right) \sim I_к^{-1}. \qquad (2.11)$$

Эта зависимость действительно соответствует зависимости, наблюдаемой в эксперименте в области больших токов коллектора (см. рис. 2.6). [1]) Уменьшение $\beta_{ст}$ с ростом уровня инжекции

---

[1]) Как мы уже отмечали на с. 25, высота потенциального барьера, который преодолевают инжектируемые электроны и дырки в p–n-переходе, одинакова, и поэтому отношение токов электронов и дырок пропорционально отношению их концентраций в электронейтральных областях. В условиях высокого уровня инжекции концентрация носителей в базе увеличивается, при этом доля тока, текущего из базы в эмиттер, также возрастает, что и приводит к уменьшению коэффициента инжекции эмиттера γ. Из этого объяснения сразу же следует, что в транзисторах, эмиттерным переходом в которых является гетеропереход, понижение эффективности инжекции с ростом плотности тока должно отсутствовать.

ограничивает максимальную плотность тока, при которой транзистор сохраняет свои усилительные свойства. Детальные расчеты (см. [11]) показывают, что изменение уровня легирования базы не оказывает влияния на этот эффект. Поэтому мощные (сильноточные) транзисторы неизбежно должны иметь большую площадь.

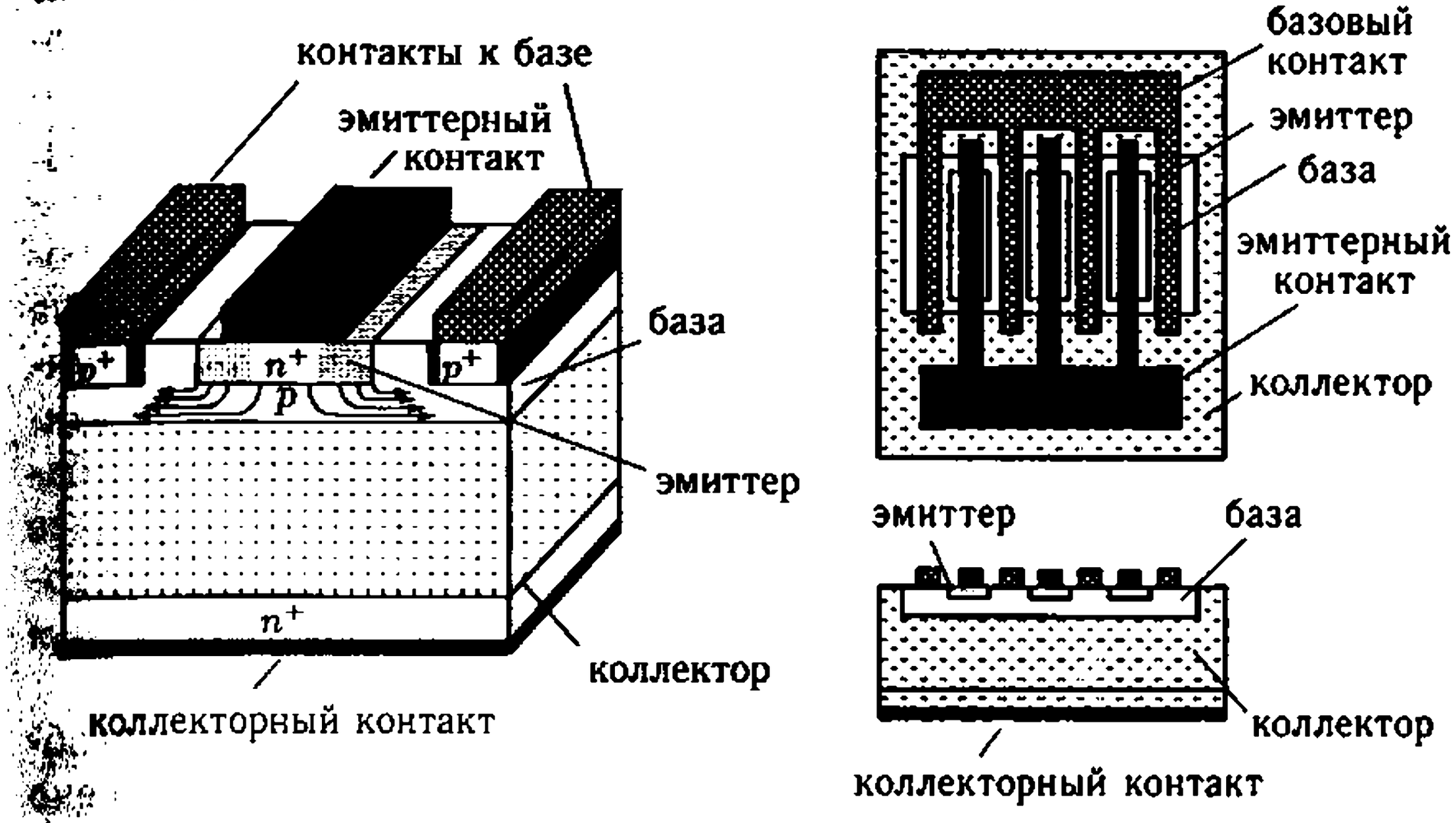


Рис. 2.7. *а* — схема эпитаксиально-планарного n–p–n-транзистора, на которой показаны линии тока дырок в базе; *б* — вид сверху и поперечное сечение конструкции транзистора с гребенчатым расположением выводов эмиттера и базы

На уменьшение коэффициента усиления при высокой плотности тока может влиять и *эффект Кирка* — эффект эффективного увеличения толщины базы в эпитаксиальных транзисторах с ростом плотности тока. Так как этот эффект существенно ухудшает и быстродействие транзисторов, мы рассмотрим его в п. 2.4.1., посвященном высокочастотным свойствам транзисторов.

2.2.4. Эффект оттеснения эмиттерного тока. В приведенных выше расчетах мы пренебрегали падением напряжения при протекании тока через области биполярного транзистора, имеющие конечное сопротивление, и считали, что ток в транзисторе течет однородно по всей площади структуры. Однако в действительности это не так.

На рис. 2.7,*а* показано распределение линий тока дырок в базе типичного эпитаксиального n–p–n-транзистора. Видно, что плотность линий тока в различных частях базы различна и максимальна вблизи краев эмиттера. Это является следствием достаточно большого удельного сопротивления материала

базовой области. Поскольку падение напряжения между контактом к базе и разными точками эмиттерного перехода из-за разного расстояния до этих точек оказывается различным, то и *локальное напряжение смещения эмиттерного перехода в этих* точках различно, что и порождает неоднородность эмиттерного тока. Этот эффект получил название *эффекта оттеснения («сжатия») эмиттерного тока.* Очевидно, что указанный эффект должен проявляться особенно сильно в приборах, работающих при высокой плотности тока — мощных транзисторах [11]. Кроме того, конечное сопротивление базы оказывает существенное влияние и на свойства высокочастотных транзисторов, в которых для получения максимального быстродействия толщина базы делается особенно тонкой (~0,1 мкм) и приходится принимать все меры по уменьшению длины линий тока в базе транзистора (см. п. 2.4). Чтобы в этих транзисторах сопротивление базы оставалось достаточно низким, концентрацию примесей в базе высокочастотных транзисторов приходится увеличивать и поэтому эти транзисторы характеризуются сравнительно небольшим коэффициентом усиления ($\beta \sim 50$).

Для уменьшения сопротивления базовой области и ослабления эффекта оттеснения эмиттерного тока было предложено создавать эмиттер в виде длинных узких полосок и использовать гребенчатое расположение выводов эмиттера и базы (рис. 2.7*б*). В такой геометрии, благодаря большому периметру контактов, эффективную длину линий тока в базе удается существенно сократить. В современных СВЧ транзисторах ширина эмиттерных полосок достигает 0.4–0.5 мкм.

2.2.5. Влияние напряжения на коллекторе на коэффициент усиления. Уже первые исследования плоскостных транзисторов показали, что на их выходных характеристиках наблюдается заметное увеличение тока коллектора с ростом напряжения на коллекторе при фиксированном токе базы (см. рис. 2.8). Эрли [113] первым объяснил причину такого поведения и с тех пор этот эффект носит его имя. *Эффект Эрли* состоит в возрастании статического коэффициента усиления транзистора $\beta_{ст} = J_к / J_б$ с увеличением напряжения на коллекторе. Причиной этого изменения является уменьшение толщины нейтральной области базы W' с ростом напряжения на коллекторе в результате расширения той части области пространственного заряда коллекторного p–n-перехода, которая заходит в базу. Уменьшение W' приводит к двум эффектам: 1) увеличению градиента концентрации инжектированных в базу носителей, что проявляется

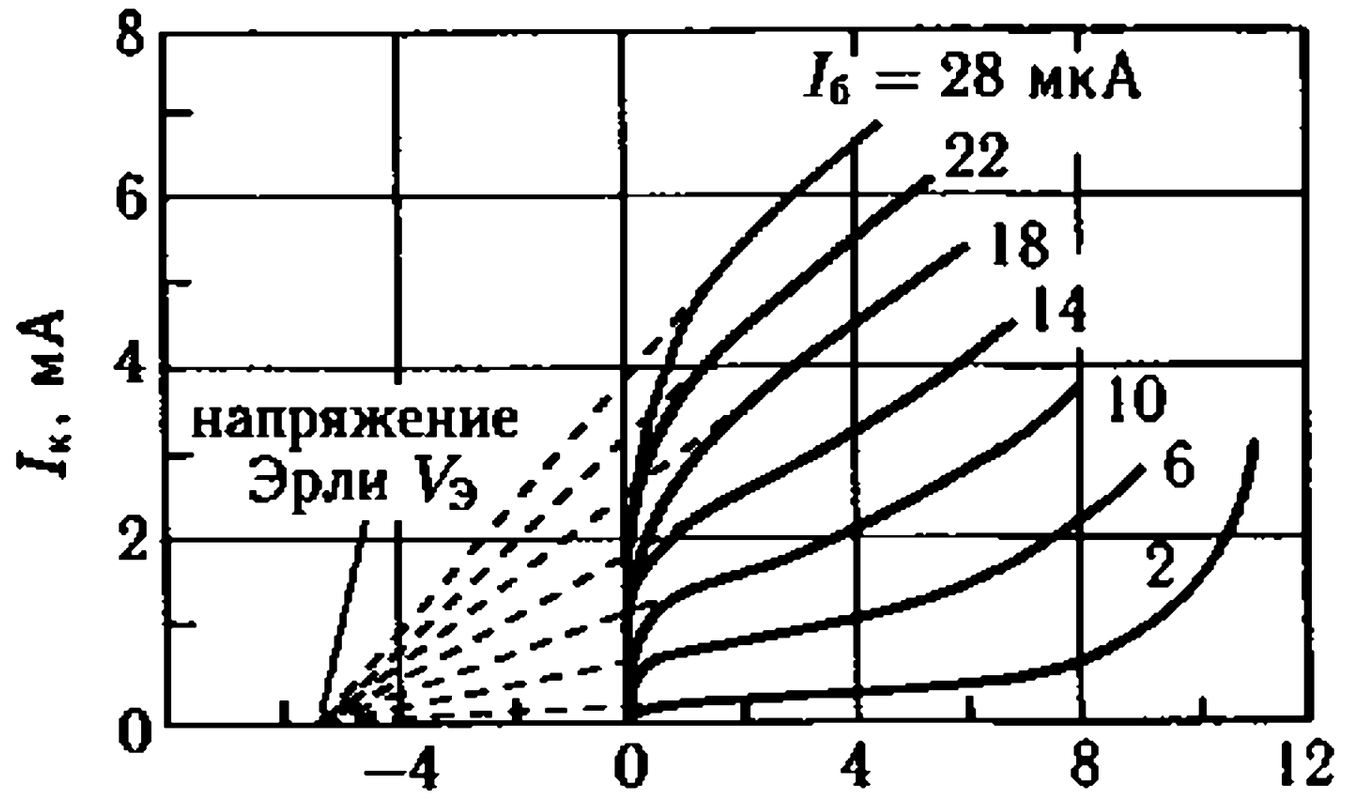


Рис. 2.8. Проявление эффекта Эрли в выходных характеристиках транзистора в схеме с общим эмиттером

в увеличении коэффициента инжекции эмиттера γ (см. формулу (2.6)), и 2) уменьшению времени пролета носителей через базу, что отражается в увеличении коэффициента переноса носителей α_T (см. формулу (2.8)). Оба эффекта приводят к увеличению коэффициента усиления транзистора.

Увеличение градиента концентрации инжектированных в базу носителей при постоянном напряжении смещения $V_{бэ}$ в результате уменьшения W' приводит еще к одному следствию: понижению напряжения $V_{бэ}$, необходимого для поддержания постоянного тока коллектора, при увеличении напряжения на коллекторе.

Очевидно, что наиболее ярко эффект Эрли должен проявляться в транзисторах с тонкой и слабо легированной базой, в которых относительное изменение толщины $\Delta W'/W'$ при изменении напряжения на коллекторе особенно велико. Поскольку эффект Эрли отрицательно влияет на усилительные свойства транзистора (в конечном счете он определяет максимальный коэффициент усиления по напряжению, который можно получить с помощью транзистора), то эффект Эрли обычно стремятся ослабить. Для этого концентрацию примеси в коллекторе стараются сделать *заметно ниже* концентрации примеси в базе с тем, чтобы область пространственного заряда коллекторного перехода располагалась бы в основном в области коллектора. Таким образом, для получения высоких усилительных характеристик транзистора в нем необходимо обеспечить следующий профиль легирования: $N_{dэ} \gg N_{aб} \gg N_{dк}$ (см. рис. 2.5).

Хотя расчеты предсказывают нелинейную зависимость $I_к(V_{кэ})$ при фиксированном токе базы, тем не менее, экстраполируя линии семейства выходных вольт-амперных характеристик транзистора (см. рис. 2.8), можно найти точку, к которой

сходятся все экстраполированные линии. Эта точка называется *напряжением Эрли* $V_Э$. Значение напряжения Эрли близко к напряжению прокола базы (см. п. 2.3).

2.3. Транзистор при высоком напряжении на коллекторе

При высоком напряжении на коллекторе напряженность электрического поля в коллекторном p–n-переходе может достигнуть столь высоких значений, что в нем начнет происходить лавинное умножение носителей. Рассмотрим, как это умножение будет влиять на вольт-амперные характеристики транзистора.

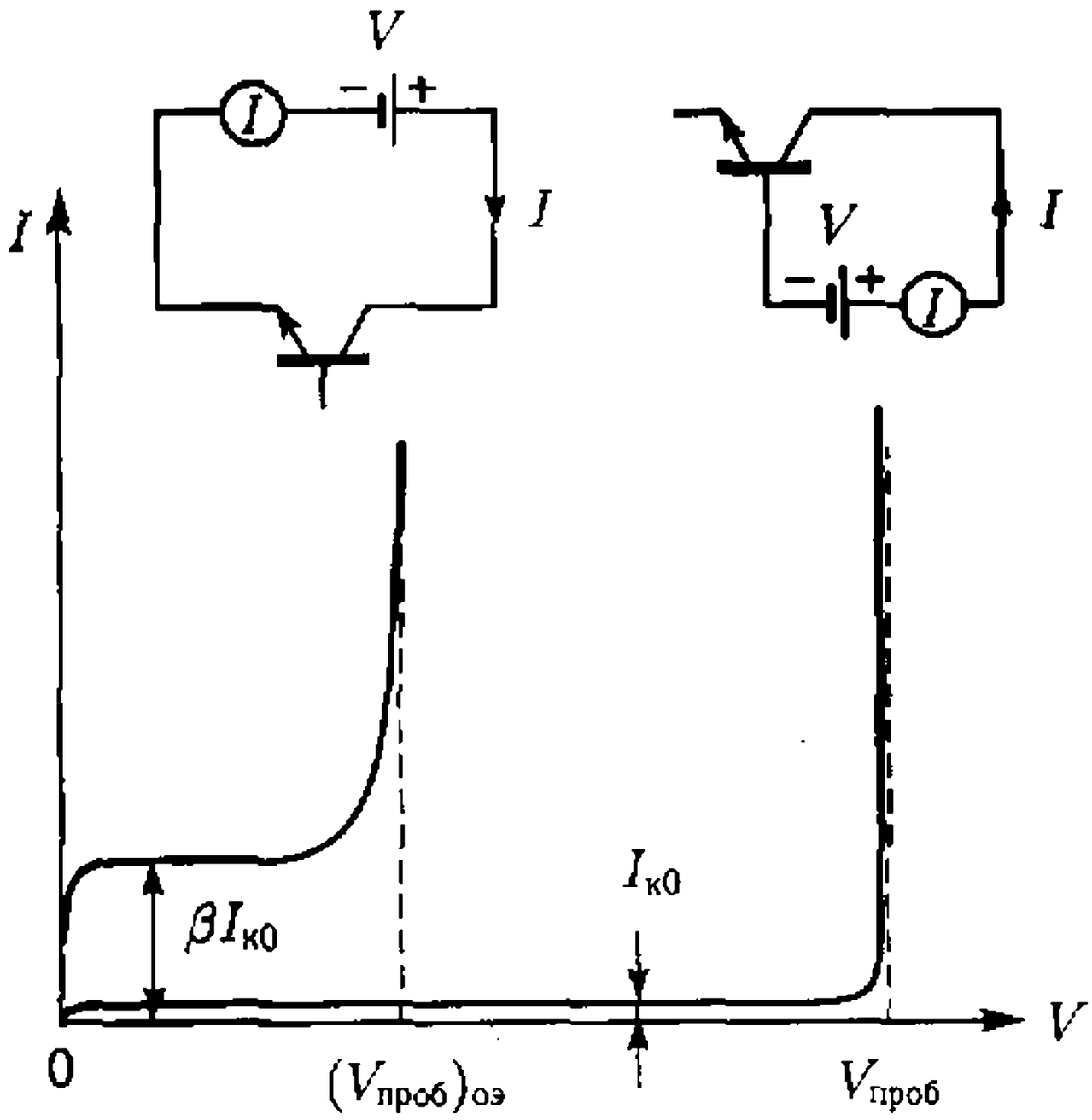


Рис. 2.9. Вольт-амперные характеристики транзистора в схемах с оборванной базой (слева) и оборванным эмиттером (справа).

Если транзистор включен в электрическую схему так, что внешнее напряжение приложено между базой и коллектором, то ток коллекторного p–n-перехода будет изменяться с напряжением точно так же, как в обычном p–n-переходе (правая кривая на рис. 2.9), то есть обратный ток коллектора $I_{к0}$ (создаваемый током неосновных носителей, возбуждаемых в области коллектора и затягиваемых полем коллекторного перехода) в предпробойной области умножается в M раз, а напряжение лавинного пробоя $V_{проб}$ (отвечающее условию $M \to \infty$, см. п. 1.3.1) определяется концентрацией примеси в области коллектора (напомним, что для ослабления эффекта Эрли база транзистора обычно легируется сильнее, чем коллектор).

Для транзистора, включенного в схему так, что внешнее напряжение приложено между эмиттером и коллектором (см. левую часть рис. 2.9), ситуация качественно изменяется, поскольку в этом случае транзистор *усиливает* втекающий в цепь базы обратный ток коллектора. Теперь в область лавинного

умножения попадают не только неосновные дырки, приходящие туда из коллектора (для определенности мы говорим о транзисторе со структурой n–p–n), но и электроны, которые инжектируются эмиттером (прямое смещение на эмиттерном переходе создается втекающим в базу обратным током коллектора). Чтобы найти условия пробоя транзистора в этом случае, запишем выражение для тока коллектора,

$$I_к = M(I_{к0} + \alpha I_э),$$

и учтем, что при оборванной базе $I_к = I_э$. Подставляя одно уравнение в другое, получаем

$$I_к = I_э = \frac{M}{1 - \alpha M} I_{к0}. \qquad (2.12)$$

Из этой формулы следует, что пробой транзистора с оборванной базой будет происходить по достижении условия $M = 1/\alpha$, то есть при намного меньших напряжениях, чем пробой транзистора с оборванным эмиттером (см. рис. 2.9).

Для оценки напряжения пробоя n–p–n-транзистора в схеме с оборванной базой используем полученное нами ранее выражение (1.75) для зависимости коэффициента умножения M от напряжения V,

$$M \approx \frac{1}{1 - (V/V_{проб})^m}, \qquad (2.13)$$

где $V_{проб}$ — напряжение пробоя коллекторного перехода (отвечающего условию $M \to \infty$), а параметр $m \approx 4$ для дырок в Si (см. с. 51). Тогда из условия $M = 1/\alpha$ для напряжения пробоя в рассматриваемой схеме (с общим эмиттером) находим

$$(V_{проб})_{оэ} = V_{проб}(1 - \alpha)^{1/m}. \qquad (2.14)$$

Для транзистора с $\alpha = 0,99$ при $m = 4$ получаем

$$(V_{проб})_{оэ} \approx 0,32 V_{проб},$$

то есть напряжение пробоя транзистора с оборванной базой значительно ниже, чем напряжение пробоя коллекторного перехода.

Чтобы предотвратить пробой транзисторов при их работе в ключевых схемах типа импульсного стабилизатора напряжения, в которых управляющий сигнал подается в цепь базы транзистора, между выводами базы и эмиттера включают небольшое сопротивление утечки, чтобы при выбросах напряжения на коллекторе обратный ток коллектора мог «стекать» через вывод базы, не открывая эмиттерный переход. С похожим решением мы встретимся в п. 3.1.1 при обсуждении конструкции тиристора с закороченным катодом.

Явление прокола. В полупроводниковых структурах, содержащих более одного p–n-перехода, возникает новое интересное явление, называемое *явлением прокола* (сквозного пробоя). Оно заключается в том, что при изменении напряжений на электродах структуры возможна ситуация, при которой области пространственного заряда p–n-переходов начинают перекрываться, и тогда ток в структуре начинает быстро нарастать, как при пробое. В биполярном транзисторе такая ситуация возникает, когда с увеличением напряжения на коллекторе части областей пространственного заряда эмиттерного и коллекторного p–n-переходов, расположенные в базе, смыкаются. Решая уравнение Пуассона (1.5) для транзисторной структуры с однородно легированной базой толщиной W и концентрациями легирующих примесей в эмиттере, базе и коллекторе, удовлетворяющих соотношению $N_{d\text{э}} \gg N_{a\text{б}} \gg N_{d\text{к}}$, нетрудно найти условие возникновения прокола:

$$W = \sqrt{\frac{\varepsilon}{2\pi q}\frac{1}{N_{a\text{б}}}\phi_{k\text{э}}} + \sqrt{\frac{\varepsilon}{2\pi q}\frac{N_{d\text{к}}}{N_{a\text{б}}^2}(\phi_{k\text{к}} + V_{\text{прок}})}, \qquad (2.15)$$

где $\phi_{k\text{э}}$ и $\phi_{k\text{к}}$ — контактные разности потенциалов эмиттерного и коллекторного переходов, соответственно. Если принять, что толщина обедненного слоя эмиттерного перехода (первое слагаемое в (2.15)) мала по сравнению с толщиной той части обедненного слоя коллекторного перехода, которая расположена в базе, и пренебречь величиной $\phi_{k\text{к}}$ по сравнению с приложенным к коллектору внешним напряжением, то напряжение прокола можно оценить по формуле

$$V_{\text{прок}} \approx \frac{2\pi q N_{a\text{б}}^2}{\varepsilon N_{d\text{к}}} W^2. \qquad (2.16)$$

Как показывают расчеты, после достижения напряжения $V_{\text{прок}}$ высота потенциального барьера, который необходимо преодолевать инжектируемым носителям, начинает быстро уменьшаться с ростом напряжения на коллекторе, а ток коллектора — экспоненциально быстро возрастать (вольт-амперная характеристика диода с проколом будет рассчитана нами в п. 6.3, когда мы будем изучать работу инжекционно-пролетного диода).

Явление прокола можно наблюдать в транзисторах с тонкой слабо легированной базой, в которых смыкание областей пространственного заряда происходит при напряжении на коллекторе, недостаточном для развития лавинного пробоя.

Вторичный пробой и область безопасной работы. Кроме описанных выше явлений в мощных транзисторах наблюдается еще один тип пробоя — так называемый *вторичный пробой* [11, 14, 114], который связан с тепловым пробоем, рассмотренным нами в п. 1.3.3. [1]) Опыт эксплуатации транзисторов показывает, что при достаточно высоком напряжении на коллекторе и высокой плотности тока в мощных транзисторах ток коллектора становится неустойчивым: через некоторое время после подачи на транзистор напряжения ток через транзистор резко возрастает, а напряжение на транзисторе — падает. Время, через которое это происходит, называют *временем включения вторичного пробоя*, его характерные значения лежат в пределах 10^{-7}–10^{-3} с.

Исследования показали, что появление вторичного пробоя обычно связано с изначальной неоднородностью протекающего в транзисторе тока, хотя согласно теоретическим расчетам, неустойчивость может возникать и в совершенно однородной структуре. Причиной этой неустойчивости является сильная зависимость плотности тока эмиттера от температуры ($dJ_э/dT > 0$) при фиксированном напряжении эмиттер–база. [2]) В местах, где плотность тока случайно оказывается выше, происходит повышенное тепловыделение и возникает *локальный перегрев*, который способствует еще большей концентрации протекающего тока в этих областях. Когда температура локальных областей достигает критического значения, выше которого протекание тока в p–n-переходе становится неуправляемым (значения этой температуры для разных полупроводников мы оценивали в п. 1.2.1), в структуре возникают «горячие точки», диаметр которых составляет ~10 мкм. Протекающий через прибор ток концентрируется вокруг этих точек и в конце концов начинает течь по узким «шнурам», разогреваемым самим током. При этом средний ток и средняя рассеиваемая транзистором мощность могут быть существенно ниже максимальных значений, рассчитанных в предположении однородно протекающего тока. Если ток коллектора не ограничить, в местах локализации шнура может произойти расплавление кристалла или контактов, что

---

[1]) В принципе явление вторичного пробоя наблюдалось и в других полупроводниковых приборах — мощных диодах, тиристорах, МОП-транзисторах, однако в мощных биполярных транзисторах оно проявляется наиболее часто.

[2]) Сильное увеличение плотности тока эмиттера с ростом температуры при фиксированном $V_{бэ}$ связано с уменьшением высоты потенциального барьера, который приходится преодолевать инжектируемым носителям.

приведет к выходу прибора из строя. Однако даже если этого не произойдет, возникновение сильных температурных градиентов в приборе может вызвать растрескивание кристалла.

Режим работы, при котором возникает вторичный пробой, сильно зависит от геометрии прибора. Для обеспечения однородности протекания тока в мощных транзисторах их выводы базы и эмиттера выполняют в виде гребенчатой структуры (см. рис. 2.7 *б*), причем выводы отдельных эмиттерных полосок подключают к общему выводу эмиттера через стабилизирующие (балластные) резисторы, которые компенсируют любое отклонение тока в отдельном эмиттере [11].

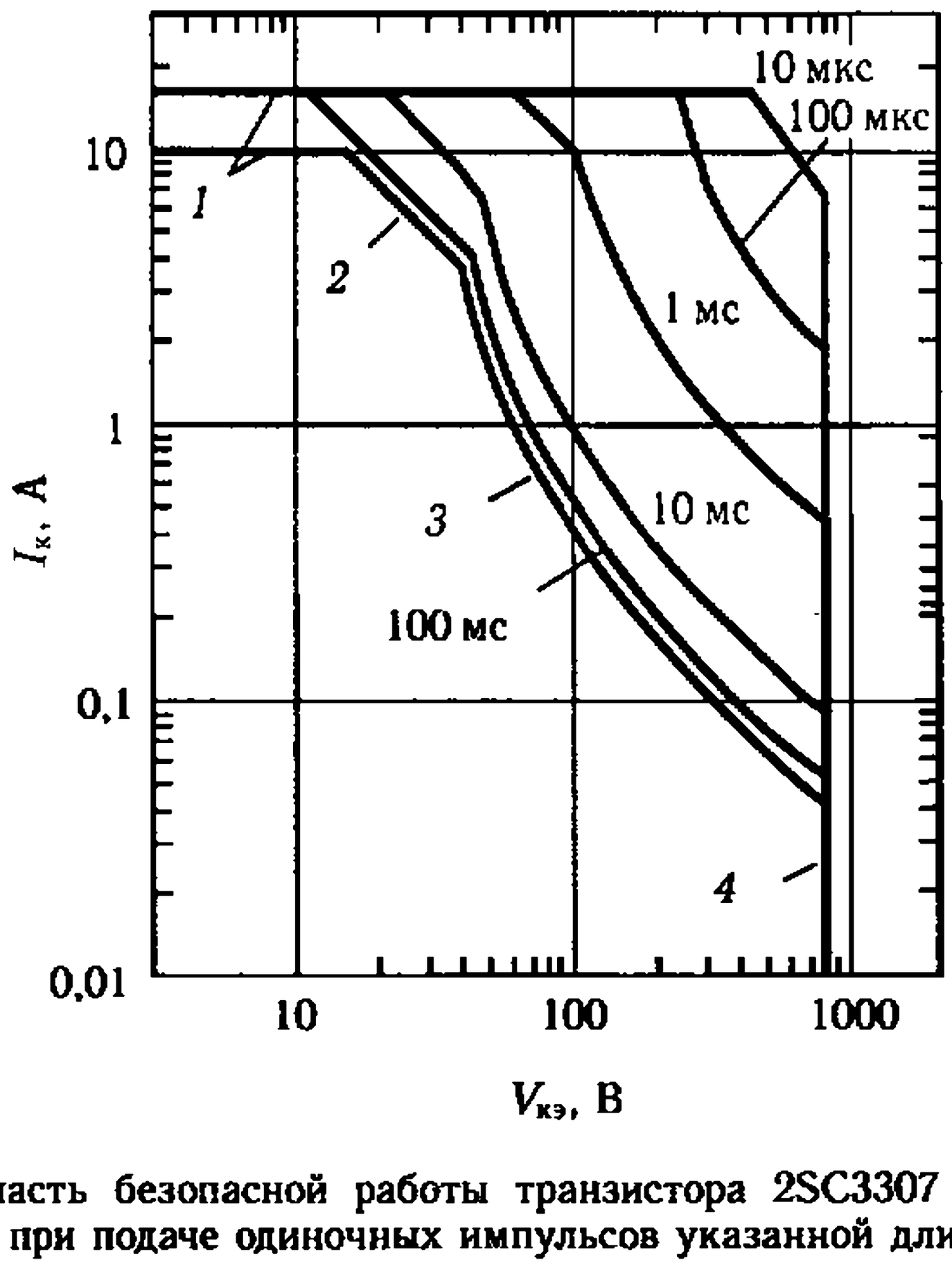


Рис. 2.10. Область безопасной работы транзистора 2SC3307 в статическом режиме и при подаче одиночных импульсов указанной длительности

Область электрических режимов, в которых возможна длительная и надежная эксплуатация мощных транзисторов, называется *областью безопасной работы*. Область безопасной работы ограничена отрезками четырех линий (см. рис. 2.10): линии максимального рабочего тока *1*, линии максимальной рассеиваемой мощности *2*, линии максимального рабочего напряжения *4*, и линии *3*, ограничивающей область, за пределами которой возможно возникновение вторичного пробоя. Максимальный

рабочий ток определяется током, при котором коэффициент усиления транзистора β имеет еще приемлемое значение (как мы показали в п. 2.2.3, при высоких плотностях тока β быстро уменьшается). Максимальная рассеиваемая мощность определяется условиями отвода тепла от прибора и связана с необходимостью ограничить среднюю температуру полупроводникового кристалла значением, при котором транзистор еще остается вполне управляемым, а темп деградации прибора невелик. Для кремниевых приборов эта температура лежит в пределах 150–200 °C. В двойном логарифмическом масштабе ($\lg I_к - \lg V_{кэ}$) линия максимальной рассеиваемой мощности имеет наклон, равный -1. Максимальное рабочее напряжение ограничено напряжением пробоя транзистора в схеме с общим эмиттером $(V_{проб})_{оэ}$.[1]) Наконец, граница области вторичного пробоя определяет ту область, в которой транзистор данной конструкции может работать достаточно стабильно. Положение этой линии для каждого типа транзисторов находится экспериментально и приводится в технической документации. Поскольку, как мы отмечали выше, возникновение вторичного пробоя зависит от продолжительности протекания тока и рассеиваемой мощности, положение границы области вторичного пробоя в импульсном режиме (как и положение линии ограничения по мощности) зависит от длительности импульсов (см. рис. 2.10).

2.4. Высокочастотные свойства и быстродействие транзисторов

2.4.1. Частота отсечки и максимальная частота генерации. Одним из наиболее важных требований, предъявляемых к транзисторам, является их высокое быстродействие. Так, для создания процессоров ЭВМ необходимо, чтобы максимальная рабочая частота транзисторов была примерно в 30 раз выше тактовой частоты процессора. Это означает, что для создания современного процессора с тактовой частотой 3 ГГц необходимы транзисторы, способные работать на частоте 90 ГГц. В этом разделе мы рассмотрим, какие параметры конструкции биполярных транзисторов определяют их высокочастотные свойства.

---

[1]) Напряжение пробоя коллекторного перехода при этом часто указывается в справочниках как максимальное импульсное напряжение между эмиттером и коллектором при закрытом транзисторе.

Одним из важнейших параметров, характеризующих высокочастотные свойства транзистора, является *частота отсечки* f_T. Она определяется как частота, на которой модуль коэффициента усиления по току в схеме с общим эмиттером, $|\beta| = |dI_к/dI_б|$, измеренный в режиме короткого замыкания на выходе [1], становится равным 1.

Частота отсечки непосредственно связана соотношением $f_T = 1/2\pi\tau_{ec}$ с временем задержки распространения сигнала от эмиттера к коллектору τ_{ec}. Это время задержки представляет собой сумму четырех времен, характеризующих последовательные фазы распространения инжектированных носителей от эмиттера к коллектору:

$$\tau_{ec} = \tau_E + \tau_B + \tau_C + \tau'_C. \qquad (2.17)$$

Первое слагаемое в этой сумме представляет собой постоянную времени заряда емкости эмиттера. Оно равно произведению дифференциального сопротивления эмиттера,

$$r_э = dI_э/dV_{бэ} \approx kT/qI_э,$$

на суммарную емкость эмиттера $C_{эб}$, коллектора $C_{кб}$ и паразитную емкость базового вывода $C_{пар}$. Появление трех составляющих емкости связано с тем, что входной сигнал подается в цепь базы транзистора.

Второе слагаемое в (2.17) представляет собой время пролета носителей через базу. Выражение для времени пролета может быть найдено из частотной зависимости коэффициента переноса носителей через базу α_T. Для транзистора с однородно легированной базой эта зависимость может быть легко вычислена в рамках подхода, использованного нами при рассмотрении диффузионной емкости p–n-перехода (см. п. 1.7.2). Подставляя комплексное выражение для τ_n (1.130) в уравнение (2.8), в пределе $\omega\tau_n \ll 1$, $W' \ll L_n$ получаем

$$\alpha_T(\omega) \approx \frac{1}{\operatorname{ch}\left(W'\sqrt{\dfrac{1+i\omega\tau_n}{D_n\tau_n}}\right)} \approx \frac{\alpha_T(0)}{1+i\omega\dfrac{W'^2}{2D_n}}, \qquad (2.18)$$

откуда $\tau_B = W'^2/2D_n$.

---

[1]) Условие короткого замыкания на выходе означает, что выходная цепь транзистора подключается к столь малому сопротивлению нагрузки, что обратной связью со стороны нагрузки коллектора можно пренебречь.

Как мы отмечали в п. 2.2, в дрейфовых транзисторах, в базе которых специально создано встроенное электрическое поле, время пролета носителей через базу заметно меньше, чем в транзисторах с однородно легированной базой. Расчеты показывают, что если напряженность встроенного поля в базе постоянна и равна $\mathcal{E}_{bi}$, то время пролета уменьшается в

$$\eta \approx [1 + (\mathcal{E}_{bi}/\mathcal{E}_0)^{3/2}]$$

раз, где $\mathcal{E}_0 = 2kT/qW'$ (другие аппроксимирующие функции см. в [115]). Величина η может достигать нескольких десятков.

Третье слагаемое в (2.17) представляет собой время пролета носителей через обедненный слой коллекторного p–n-перехода. Полагая, что в сильном электрическом поле p–n-перехода носители движутся со скоростью насыщения v_s (в Si $v_s \approx \approx 10^7$ см/с), [1]) это время составляет

$$\tau_C = \frac{x_{\text{обедн.к}}}{v_s}, \qquad (2.19)$$

где $x_{\text{обедн.к}}$ — толщина обедненного слоя коллекторного перехода.

Наконец, последнее слагаемое в (2.17) представляет собой постоянную времени заряда коллектора, равную произведению барьерной емкости коллекторного перехода $C_{\text{кб}}$ и последовательного сопротивления области коллектора $r_{\text{к}}$:

$$\tau'_C = r_{\text{к}} C_{\text{кб}}. \qquad (2.20)$$

Таким образом, полное время задержки имеет вид:

$$\tau_{ec} = (C_{\text{эб}} + C_{\text{кб}} + C_{\text{пар}})\frac{kT}{qI_{\text{э}}} + \frac{W'^2}{2\eta D_n} + \frac{x_{\text{обедн.к}}}{v_s} + r_{\text{к}} C_{\text{кб}}. \qquad (2.21)$$

Из этой формулы следует, что для увеличения быстродействия транзистора необходимо уменьшать емкости эмиттера, коллектора и паразитную емкость базы транзистора, уменьшать толщину базы и последовательное сопротивление коллектора, а электронные схемы должны работать при достаточно больших токах эмиттера. Толщина базы в современных кремниевых дрейфовых

---

[1]) В сильном электрическом поле скорость дрейфа носителей в полупроводниках перестает линейно зависеть от напряженности электрического поля и стремится к некоторой величине, называемой *скоростью насыщения*. Этот эффект называется *эффектом насыщения скорости дрейфа*. Причиной такого поведения является сильный разогрев электронного газа, при котором основным механизмом рассеяния энергии носителей становится их рассеяние на оптических фононах. Для оценки скорости насыщения можно воспользоваться эмпирической формулой $v_s \approx \sqrt{\hbar\omega_{LO}/m^*}$, где $\hbar\omega_{LO}$ — энергия продольного оптического фонона, а m^* — эффективная масса носителя [2, 84].

транзисторах может быть сделана менее 0,1 мкм; при этом f_T достигает 45 ГГц (транзистор BFP520 фирмы Siemens/Infineon). Для уменьшения $r_к$ в современных транзисторах коллектор делают в виде тонкого (2–5 мкм) слабо легированного эпитаксиального слоя, выращенного на сильно легированной (n^+) подложке. Невысокая концентрация примесей в коллекторе важна для получения низкой емкости коллекторного перехода. Следует заметить, что из-за малой толщины базы и коллектора СВЧ транзисторы обычно имеют невысокое рабочее напряжение.

Расчеты показывают [14], что *коэффициент усиления транзистора по мощности* на высокой частоте определяется не только временем распространения сигнала от эмиттера к коллектору, но также и постоянной времени, действующей внутри транзистора обратной связи, которая равна произведению сопротивления базы транзистора $r_б$ на емкость коллекторного перехода $C_{кб}$:

$$K_P = \frac{\alpha f_T}{8\pi r_б C_{кб} f^2}. \qquad (2.22)$$

Частота, на которой коэффициент усиления по мощности обращается в единицу, называют *максимальной частотой генерации* транзистора:

$$f_{max} = \sqrt{\frac{\alpha f_T}{8\pi r_б C_{кб}}}. \qquad (2.23)$$

На частотах выше f_{max} коэффициент усиления по мощности становится меньше единицы и осуществить режим самовозбуждения генератора на этой частоте невозможно.

Из формулы (2.23) следует, что для получения высокого значения f_{max} необходимо не только уменьшать время задержки распространения сигнала τ_{ec}, но и уменьшать сопротивление базы. К сожалению, это сделать довольно трудно из-за необходимости иметь в высокочастотных транзисторах как можно более тонкую базу. Для понижения $r_б$ база СВЧ транзисторов легируется достаточно сильно (при этом в жертву приносится коэффициент усиления по току), а специальная геометрия транзисторов (полосковая геометрия эмиттера или гребенчатая конструкция в более мощных транзисторах, см. рис. 2.7) способствует укорочению среднего пути тока в области базы. Ширина эмиттерной полоски является, пожалуй, наиболее важным фактором, определяющим величину $r_б$ и максимальную частоту генерации. Нетрудно видеть, что при постоянной площади эмиттера уменьшение ширины эмиттерной полоски $L_э$ сопровождается увеличением ее длины, при этом сопротивление области базы, расположенной

под эмиттером, изменяется как $r_б \sim L_э^2$ [88]. Тогда, если пренебречь изменением емкости коллектора при изменении геометрии эмиттера, из формулы (2.23) следует, что $f_{max} \sim L_э^{-1}$.

Как отмечалось во введении, в настоящее время стоимость производства полупроводниковых приборов довольно велика и поэтому большое значение играет этап предварительного проектирования и моделирования работы этих приборов. В настоящее время для моделирования транзисторов разработаны специальные компьютерные программы (см., например, [116]), позволяющие оптимизировать их конструкцию, добиваясь достижения заданных значений различных параметров, в частности, быстродействия.

Эффект Кирка. На рис. 2.11 показана зависимость частоты отсечки типичного высокочастотного транзистора. Видно, что с ростом тока коллектора частота отсечки сначала монотонно возрастает, а затем резко падает. В то время как начальное возрастание f_T объясняется уменьшением первого слагаемого в формуле (2.21), причина резкого спада f_T некоторое время оставалась неясной. Кирк [117] первым объяснил это явление и с тех пор оно получило название *эффекта Кирка*.

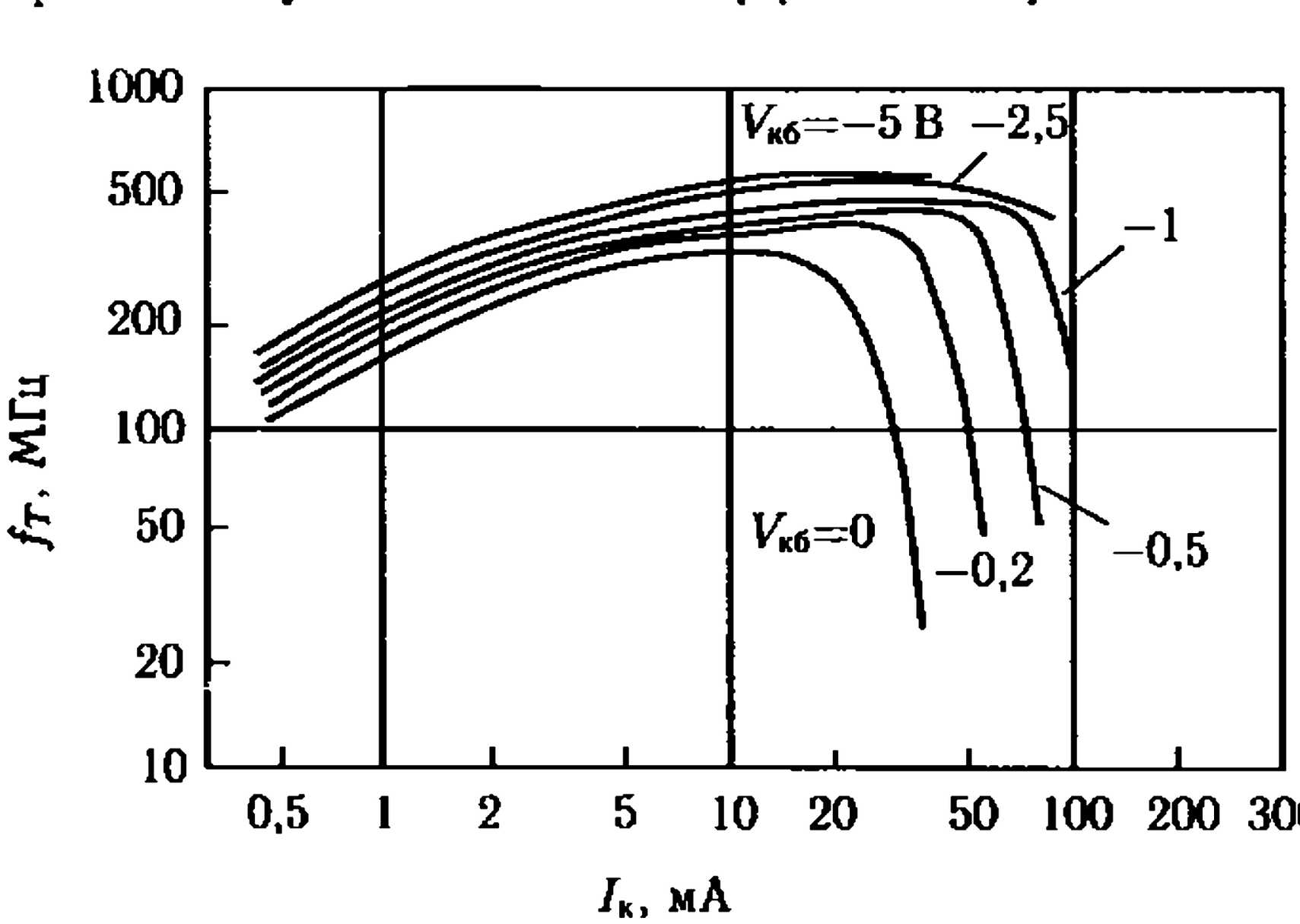


Рис. 2.11. Зависимость частоты отсечки транзистора 2N695 от тока коллектора при различных напряжениях $V_{кб}$ [117]

Оказалось, что причиной резкого спада f_T является то, что с ростом тока коллектора падение напряжения на сравнительно высокоомной области коллектора быстро возрастает, пока при некотором токе это падение напряжения не превысит напряжение, приложенное между коллектором и базой. При этом

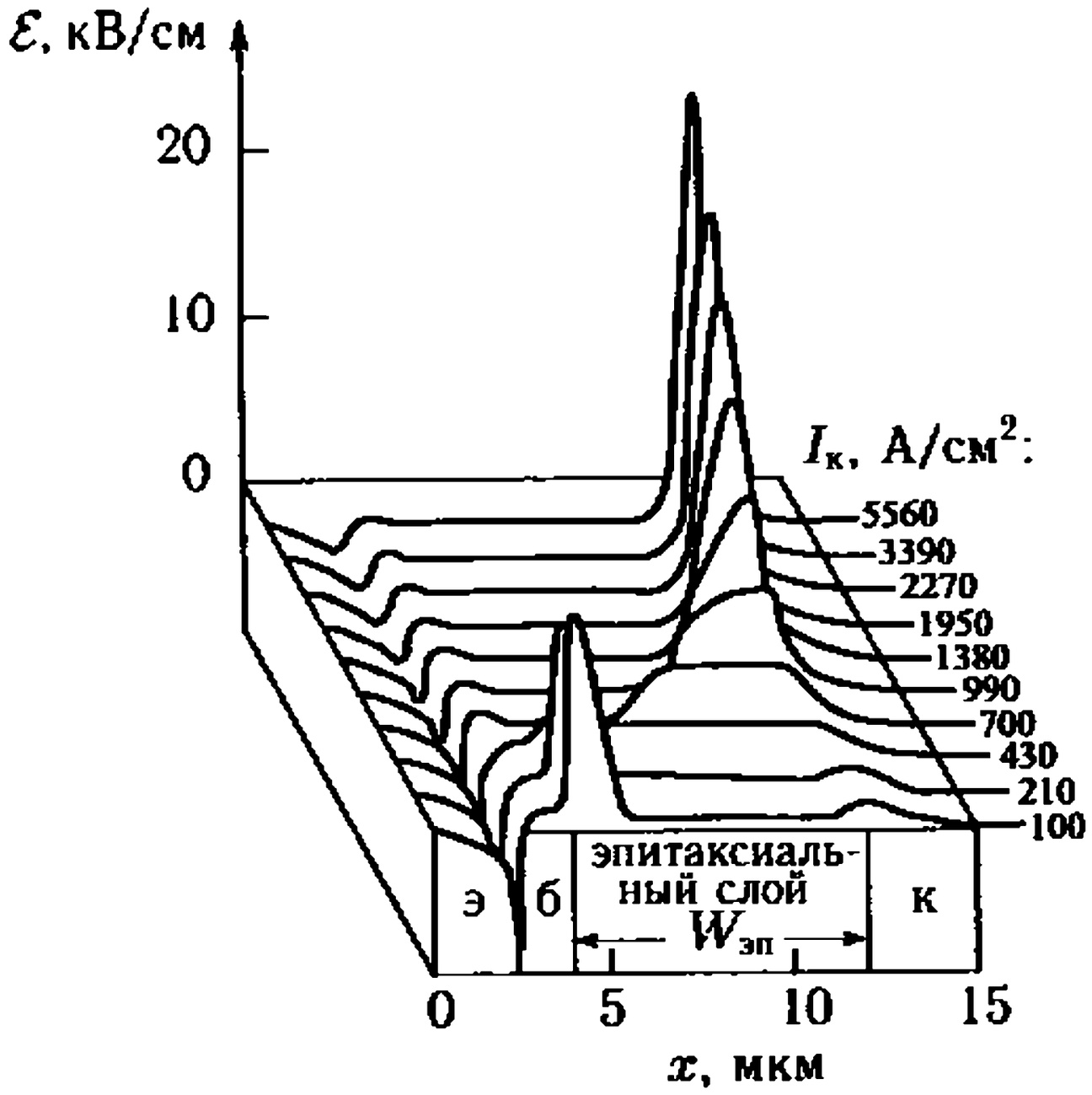


Рис. 2.12. Распределение электрического поля в эпитаксиальном транзисторе при различных плотностях тока коллектора. $n_{эп} = 10^{15}$ см$^{-3}$ [14]

коллекторный переход становится смещенным в *прямом* направлении и, как показывает численное моделирование (см. рис. 2.12), максимум в распределении электрического поля перемещается в область n—n^+-контакта, а эффективная толщина базы (и, соответственно, время пролета носителей через базу) возрастает. Это и вызывает резкое уменьшение f_T. Одновременно с этим уменьшается и коэффициент усиления транзистора, поскольку в результате увеличения эффективной толщины базы эффективное число Гуммеля (см. с. 151) также возрастает.

В эпитаксиальных транзисторах критическую плотность тока J_1, выше которой возникает эффект Кирка, можно оценить из условия равенства падения напряжения на эпитаксиальном слое толщиной $W_{эп}$ сумме приложенного к коллектору напряжения $V_{кб}$ и контактной разности потенциалов $\phi_{kк}$ коллекторного перехода:

$$J_1 \approx q\mu_n n_{эп} \frac{|V_{кб}| + \phi_{kк}}{W_{эп}}. \qquad (2.24)$$

В этой формуле μ_n и $n_{эп}$ — подвижность и концентрация электронов в эпитаксиальном слое.

2.4.2. Гетеропереходные транзисторы. В последнее время в производстве СВЧ транзисторов стали широко использоваться прямозонные полупроводники группы $A^{III}B^V$, подвижность и скорость насыщения электронов в которых намного выше, чем в Si (см. табл. 2 в Приложении). Создание в этих транзисторах гетеропереходов (или варизонных структур) для обеспечения эффективной инжекции в тонкую сильно легированную базу позволило на этих т. н. *гетеропереходных транзисторах* (HBT — heterostructure bipolar transistor) получить удивительные результаты.

Так, в гетеропереходных транзисторах с широкозонным эмиттером из InP, в которых в качестве узкозонного полупроводника был использован согласованный с InP по параметру решетки твердый раствор $In_{0,53}Ga_{0,47}As$ (см. рис. 2.13 *а*), в 2003 г. удалось получить частоту отсечки $f_T = 452$ ГГц [118], а при использовании в базе и коллекторе транзистора варизонных слоев — $f_T = 604$ ГГц [119]. Соответствующее последней частоте время распространения сигнала составляет всего $\tau_{ec} = 0,26$ пс. Максимальная частота генерации в гетеропереходных транзисторах InP/InGaAs достигает $f_{max} = 310$ ГГц [120]. Заметно увеличить быстродействие этих транзисторов удается за счет сильного легирования базы — до концентраций $N_{аб} = (4–6) \cdot 10^{19}$ см$^{-3}$, при которых еще не происходит заметного снижения коэффициента инжекции гетероперехода эмиттера. Это позволяет получить очень малые значения сопротивления базы $r_б$ и постоянной времени внутренней обратной связи $r_б C_к$. Уровень легирования эмиттера в таких транзисторах намеренно делают невысоким ($N_{дэ} \sim 5 \cdot 10^{17}$ см$^{-3} \ll N_{аб}$), чтобы снизить емкость эмиттерного перехода. В настоящее время для создания гетеропереходных транзисторов с параметрами, несколько уступающими приведенным выше для пары $InP–In_{0,53}Ga_{0,47}As$, используются такие пары как $Al_xGa_{1-x}As–GaAs$ и $In_{0,5}Ga_{0,5}P–GaAs$ (первым в паре указан более широкозонный материал). Очень перспективной считается пара AlSb–InAs, поскольку арсенид индия имеет одну из наиболее высоких среди полупроводников $A^{III}B^{V}$ подвижность (33000 см^2/В · с) и скорость насыщения ($3,5 \cdot 10^7$ см·с$^{-1}$).

Основным недостатком гетероперехода транзистора $InP/In_xGa_{1-x}As$ является низкое напряжение пробоя (~2 В) коллектора, который расположен в узкозонной части структуры. Этот недостаток удается преодолеть в транзисторах с двойной гетероструктурой (DHBT — double heterojunction bipolar transistor), в которых коллектор изготавливают из более широкозонного полупроводника, например InP или $Al_xIn_{1-x}As$. Однако возникающий при этом разрыв в положении края зоны проводимости на коллекторном переходе препятствует движению носителей из базы в коллектор и тем самым ухудшает характеристики такого транзистора. Преодолеть эту проблему удалось в гетеропереходных транзисторах со структурой $InP/GaAs_{1-x}Sb_x/InP$, в которой край зоны проводимости в узкозонной базе лежит *выше* края зоны проводимости в коллекторе. В таких транзисторах в 2001 г. была получена частота отсечки 300 ГГц [121].

Поскольку присутствие подложки, на которой выращивается гетероструктура, заметно увеличивает емкость коллекторного перехода и постоянную времени $r_б C_к$, в 1998 г. была предложена технология создания *гетеропереходных транзисторов с удаленной подложкой* (transferred-substrate HBT). На структурах $InP/In_xGa_{1-x}As/InP$ с удаленной подложкой в 1998 г. удалось увеличить максимальную частоту генерации до 425 ГГц, а в 1999 г. появились сообщения о транзисторе с $f_{max} = 1080$ ГГц [122]. Таким образом, в настоящее время устройства на биполярных транзисторах способны генерировать колебания, длина волны которых в вакууме короче 0,3 мм (эта область длин волн отвечает дальней ИК-области спектра).

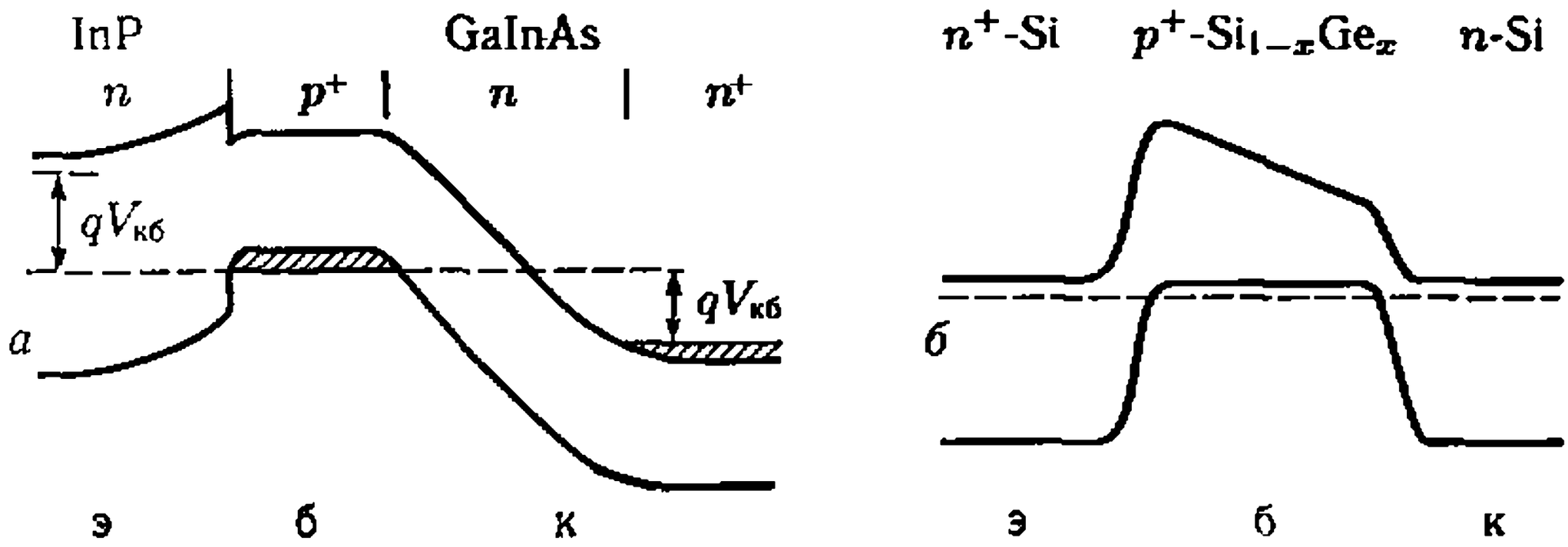


Рис. 2.13. Энергетические диаграммы гетеропереходных транзисторов: *а* — конструкция с широкозонным эмиттером, *б* — конструкция с варизонной базой

Несколько неожиданным результатом, обнаруженным при исследовании гетеропереходных транзисторов $InP/In_xGa_{1-x}As$, оказалось то, что с ростом напряжения на коллекторе время пролета τ_{ec} может возрастать [123]. Причиной этого является то, что электроны, попадающие в коллекторный переход и ускоряемые там электрическим полем, приобретают энергию, достаточную для перехода из основного минимума зоны проводимости в побочные минимумы, в которых подвижность электронов и скорость дрейфа существенно ниже. Этот эффект междолинного рассеяния выражен тем сильнее, чем больше толщина коллекторного перехода и выше напряженность электрического поля в нем. Так, в коллекторе толщиной 3000 Å в работе [123] при увеличении напряжения на коллекторе наблюдалось *двукратное* возрастание времени пролета τ_C (см. формулу (2.19)). Проведенное авторами моделирование показало, что междолинное рассеяние при движении в поле коллектора становится существенным уже при его толщине 2000 Å.

Другой тип структуры биполярного гетеропереходного транзистора имеет энергетическую диаграмму, показанную на рис. 2.13*б*. Идея создания такого транзистора с базой из полупроводника с переменной шириной запрещенной зоны была впервые высказана Крёмером в 1954 году, а Мартин и Стреттон [124] теоретически изучили высокочастотные свойства этого транзистора. Оригинальное название этой структуры (graded bandgap base transistor — транзистор с градиентом ширины запрещенной зоны в базе) почему-то не прижилось, и в современной литературе эти транзисторы тоже называют гетеропереходными биполярными транзисторами (HBT). Идея этой конструкции — создать в базе транзистора очень сильное тянущее электрическое поле для дрейфа носителей за счет использования в качестве базы *варизонной структуры* с непрерывно уменьшающейся по мере удаления от эмиттера шириной запрещенной зоны (см. рис. 2.13*б*). Кроме очевидного уменьшения времени пролета носителей через базу, наличие тянущего поля увеличивает коэффициент инжекции эмиттера, что позволяет более сильно легировать базу транзистора при сохранении высокого коэффициента усиления.

В начале 80-х годов фирма IBM предложила создать описываемый транзистор, используя хорошо отлаженную кремниевую технологию и твердый раствор $Si_{1-x}Ge_x$ в качестве материала базы. Такие транзисторы были получены в 1987 г. и оказались намного дешевле гетеропереходных транзисторов на основе полупроводников группы $A^{III}B^{V}$. Основным препятствием для создания варизонной структуры в базе является большое различие параметров решетки Ge и Si (около 4%, см. таблицу на с. 85). Это может приводить к возникновению дефектов на гетерограницах. Однако, как мы уже отмечали в п. 1.6.1, в тонких слоях небольшое рассогласование решеток может «сдерживаться» за счет упругой деформации слоя (псевдоморфная структура). Для пары $Si–Si_{1-x}Ge_x$ предельные значения x составляют 8–12% Ge при толщине базы 800–1000 Å и 16–24% Ge при толщине базы 400–500 Å. Высокая максимальная частота генерации этих транзисторов получается за счет возможности более сильно легировать базу и уменьшать $r_б$ (типичные концентрации носителей в эмиттере, базе и коллекторе составляют $7 \cdot 10^{18}$, 10^{19} и $4 \cdot 10^{16}$ см$^{-3}$), а сильное тянущее электрическое поле в базе (30–50 кВ/см) существенно увеличивает частоту отсечки.

Первые промышленные низкошумящие гетеропереходные транзисторы $Si/Si_{1-x}Ge_x$ с частотой отсечки $f_T > 15$ ГГц появились в 1999 г. (транзистор 43RF0100 фирмы IBM);

в 2002 г. частота отсечки этих транзисторов уже достигала 350 ГГц [125], а f_{max} — 185 ГГц. Минимальный коэффициент шума, полученный в этих гетеропереходных транзисторах, составляет 0,2 дБ. В настоящее время из транзисторов $Si/Si_{1-x}Ge_x$ уже созданы интегральные схемы, содержащие более 10^5 транзисторов. О современном состоянии разработок приборов на основе этих гетеропереходных транзисторов можно прочитать в [126].

Транзисторы на горячих электронах. Использование гетеропереходов позволяет создать транзисторы, в которых потенциальные барьеры образуются не за счет чередования слоев одного полупроводника разного типа проводимости, а за счет чередования слоев различных полупроводников. При этом все три области транзистора могут иметь один и тот же тип проводимости и транзистор будет работать только на основных носителях. В качестве носителей в таких транзисторах используют электроны, поскольку их подвижность в полупроводниках обычно выше подвижности дырок. Строго говоря, такие транзисторы уже нельзя отнести к *биполярным* приборам, и мы рассматриваем их в этом разделе несколько условно.

Примером такого типа транзисторов может служить *транзистор на горячих электронах* [88, 127, 128], также называемый *баллистическим транзистором.* [1]) Работа этого транзистора происходит следующим образом. Приложенное между эмиттером и базой напряжение модулирует поток электронов, туннелирующих из эмиттера в базу сквозь потенциальный барьер, созданный из широкозонного полупроводника или диэлектрика (см. рис. 2.14,*а*). В базе энергия этих электронов относительно уровня Ферми оказывается равной $E - F_n \approx qV_{бэ}$, а их скорость намного превышает скорость насыщения. Если толщину базы сделать очень тонкой (меньше длины свободного пробега горячих электронов), то происходит их *баллистический перенос*, при котором большинство электронов без рассеяния достигают

[1]) Идея создания транзистора с туннельной инжекцией горячих электронов была впервые высказана и реализована на структурах $Al/Al_2O_3/Al/Al_2O_3/Al$ Мидом [129]. По оценкам автора, потенциальное быстродействие этих транзисторов (10^{12} Гц) может даже превышать быстродействие биполярных транзисторов. Однако из-за невысокого коэффициента усиления перспективы практического использования транзисторов на горячих электронах пока не ясны. В зарубежной литературе для обозначения этих транзисторов используются аббревиатуры HET (hot-electron transistor) и THETA (tunneling hot-electron transfer amplifier).

коллектора [1]). Высокая скорость горячих электронов, баллистический характер их переноса, а также очень низкое сопротивление базы являются теми отличительными особенностями транзисторов на горячих электронах, которые позволяют обеспечить очень малое время задержки распространения в базе ($\tau_B <$ $< 0,1$ пс) и высокую максимальную частоту генерации.

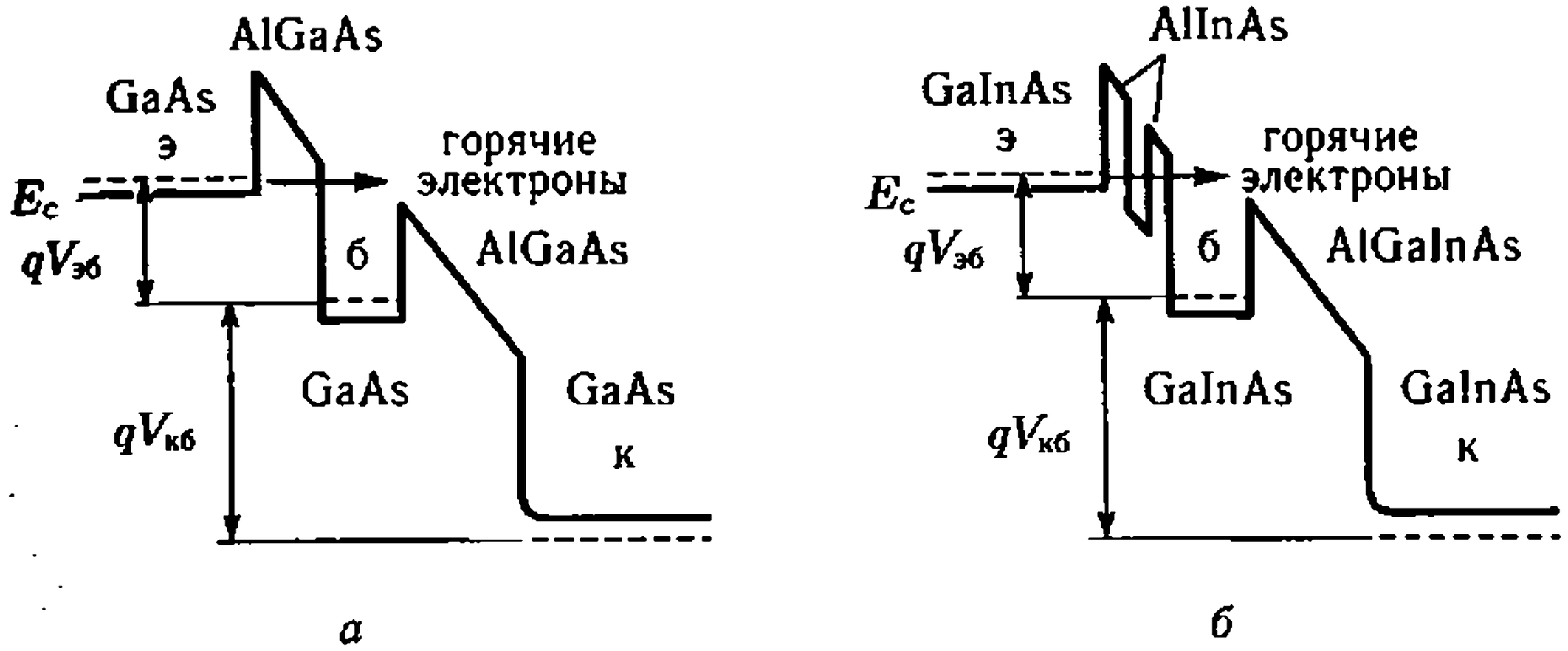


Рис. 2.14. Энергетические диаграммы транзистора на горячих электронах (*а*) и резонансного туннельного транзистора на горячих электронах (*б*)

Задача практической реализации транзисторов на горячих электронах требует поиска подходящих гетероструктурных пар. Так, для повышения рабочей температуры и уменьшения обратного тока коллектора, определяемого термоэлектронной эмиссией электронов из базы, необходимо увеличивать высоту барьера на переходе база–коллектор. В то же время эта высота должна оставаться меньше энергии горячих электронов, поступающих в базу (чтобы баллистически прошедшие базу электроны могли беспрепятственно проходить в коллектор). К сожалению, многие полупроводники $A^{III}B^{V}$, из которых обычно изготавливают обсуждаемые транзисторы, имеют небольшой энергетический зазор между основным и побочными минимумами зоны проводимости, и если энергия поступающих в базу электронов оказывается больше этой величины, то в результате междолинного

[1]) Кроме обсуждавшегося нами в п. 1.3.1 рассеяния горячих носителей на оптических фононах, в сильно легированной базе большую роль играет рассеяние с испусканием *плазмонов* (коллективных колебаний плотности свободных электронов) и *междолинное рассеяние* электронов из основного в побочные минимумы зоны проводимости. Характерные времена этих процессов составляют ~0,1 пс, поэтому при скорости электрона $v \sim 10^8$ см/с его длина свободного пробега не превышает 1000 Å [128]. Эксперимент показывает, что в транзисторах с базой из GaAs толщиной 300 Å при 4,2 К 75% носителей пролетают сквозь базу не испытывая рассеяния [130].

рассеяния длина свободного пробега электронов в базе сильно уменьшается. Это, в частности, относится к GaAs, в котором энергетические зазоры $\Delta E_{\Gamma-L} = 0,28$ эВ и $\Delta E_{\Gamma-X} = 0,48$ эВ слишком малы, чтобы транзистор мог работать при комнатной температуре. В твердом растворе $In_xGa_{1-x}As$ величина зазора $\Delta E_{\Gamma-L}$ возрастает с ростом x, достигая $\sim 1,1$ эВ в InAs [69].

Кроме рассеяния горячих носителей в базе, серьезную проблему при разработке транзисторов на горячих электронах создает *эффект квантовомеханического (надбарьерного) отражения* электронов от барьера, образованного коллектором. Согласно квантовой механике, даже если энергия электрона такова, что возможно классическое преодоление этого барьера, некоторая часть электронов, достигающих коллектора, отражается назад в базу и, рассеиваясь в ней, увеличивает ток базы. Доля отраженных электронов R определяется соотношением волновых чисел электрона k_1 перед барьером и k_2 после него и может составлять $\sim 10\%$ [88] (для прямоугольного барьера $R = (k_1 - k_2)^2/(k_1 + k_2)^2$). Поэтому получение высоких коэффициентов усиления в транзисторах на горячих электронах представляет серьезную проблему. В тщательно сконструированных транзисторах на горячих электронах коэффициент усиления β достигает 25 при 77 К и 17 при 300 К, а частота отсечки — 40 ГГц [127].

Как вариант транзистора на горячих электронах можно рассматривать и конструкцию *транзистора с металлической базой* [14, 127]. В этой конструкции роль базы выполняет тонкий слой металла (толщиной в несколько десятков Å). При такой толщине горячие носители практически не успевают рассеяться в слое металла и легко достигают коллектора. Однако из-за сильного вырождения электронов в металле значения волновых векторов k_1 и k_2 в металле и полупроводнике существенно различаются, и эффект квантовомеханического отражения на границе металл–кремний проявляется очень сильно. Поэтому в монолитных кремниевых транзисторах $Si/CoSi_2/Si$, в которых базой служит тонкий слой силицида кобальта (имеющий металлическую проводимость), максимальный коэффициент усиления по току в структурах без «дыр» в слое базы составляет $\alpha \sim 0,15$ [127].

Сочетание идеи транзистора на горячих электронах с идеей резонансно-туннельного диода позволяет создать *резонансно-туннельный транзистор на горячих электронах* (RHET — resonant-tunneling hot-electron transistor). В этом транзисторе (см. рис. 2.14*б*) для инжекции горячих электронов используется структура резонансно-туннельного диода, описанная нами на с. 115. Наиболее привлекательным в этих транзисторах

оказывается существование падающего участка на их вольт-амперных характеристиках, которое может найти применение при создании различных логических элементов. Частота отсечки в таких транзисторах достигает 40 ГГц.

2.5. Транзисторы в схемах усиления сигналов

2.5.1. Схемы включения транзистора и выбор рабочей точки.

Как мы показали в п. 2.1, характеристики транзистора позволяют использовать его в качестве усилителя электрических сигналов. С точки зрения теории электрических цепей, выводы транзистора можно включить так, чтобы один из выводов относился к входной цепи, другой — к выходной цепи, а третий вывод был общим для обеих цепей (см. рис. 2.16*а* ниже). В зависимости от того, какой из трех выводов транзистора является общим, различают следующие *схемы включения транзисторов*: с *общим эмиттером* (ОЭ), *общей базой* (ОБ) и *общим коллектором* (ОК) [26, 131].

В схеме с ОЭ входной сигнал подается на базу транзистора, а выходной — снимается с коллектора; схема с ОЭ имеет относительно высокие входное и выходное сопротивление и характеризуется высокими коэффициентами усиления по току и напряжению. В схеме с ОБ сигнал подается на эмиттер транзистора и снимается с коллектора; эта схема включения характеризуется низким входным и высоким выходным сопротивлением, а усиление в схеме возможно только по напряжению (напомним, что коэффициент усиления современных транзисторов по току α чуть меньше единицы). В схеме с ОК сигнал подается на базу транзистора и снимается с эмиттера; эта схема включения характеризуется высоким входным и низким выходным сопротивлением, а усиление в схеме происходит только по току (напряжение на эмиттере в схеме с ОК практически точно следует за напряжением на базе и поэтому коэффициент усиления по напряжению близок к единице).

Для получения максимальной амплитуды выходного сигнала и наименьших нелинейных искажений необходимо правильно выбрать положение *рабочей точки* в электрической схеме, то есть значения тока коллектора ($I_к$) и напряжения на нем ($V_{кэ}$) при нулевой амплитуде переменного сигнала. Рабочая точка A (см. рис. 2.15*б*) выбирается так, чтобы она лежала вблизи середины нагрузочной прямой (так, чтобы $V_{кэ}$ составляло приблизительно половину напряжения питания схемы). Для расчета

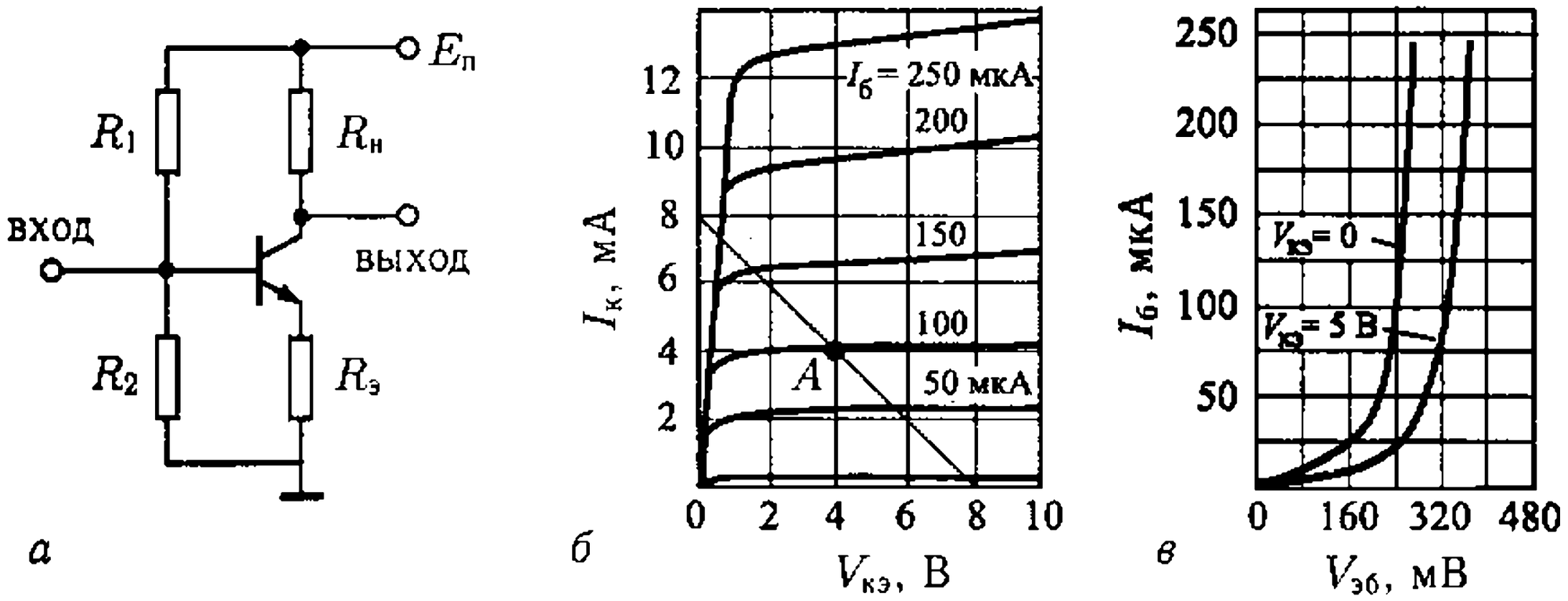
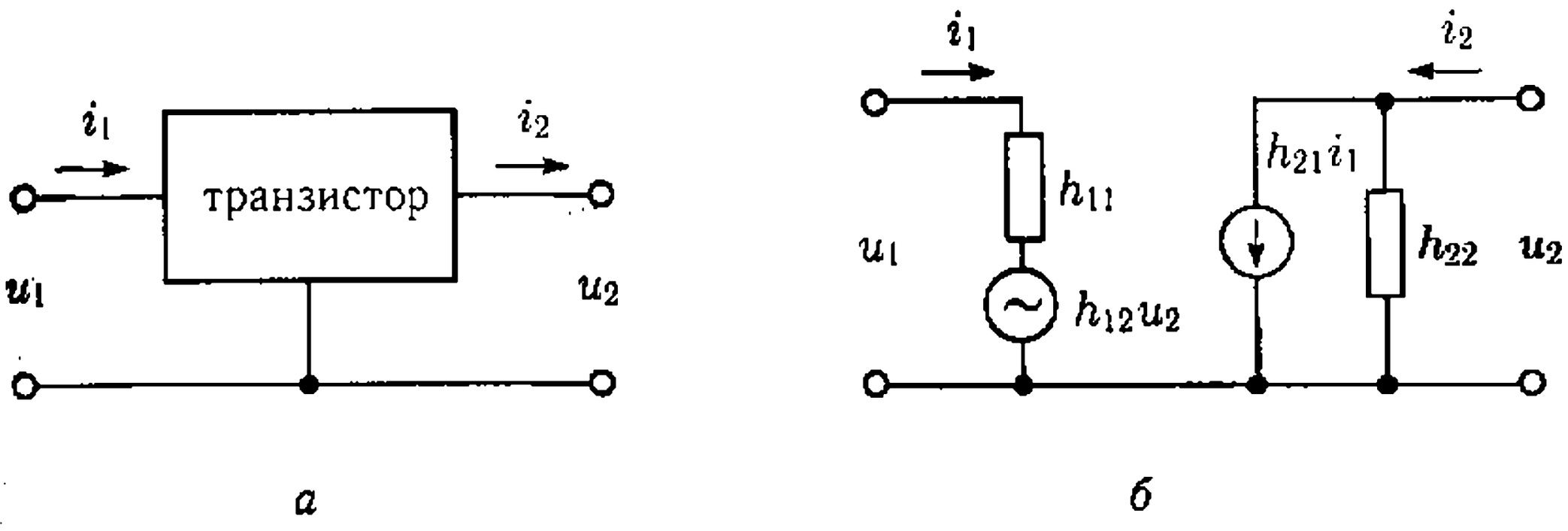


Рис. 2.15. Выбор рабочей точки усилителя на транзисторе: *а* — электрическая схема, *б*, *в* — выходные и входные характеристики германиевого транзистора ГТ322А

электрической схемы усилителя по типовой выходной вольт-амперной характеристике (рис. 2.15*б*) находится ток базы, отвечающий рабочей точке, а по типовой входной вольт-амперной характеристике (рис. 2.15*в*) находится отвечающее этому току напряжение на эмиттерном переходе. После этого по известным напряжениям и токам рассчитываются величины всех сопротивлений в схеме. При расчете схемы необходимо учесть неизбежный разброс параметров транзисторов (прежде всего, коэффициента усиления β) и их изменение с температурой. В схемах на германиевых транзисторах приходится также принимать во внимание довольно большой обратный ток коллектора $I_{к0}$ и его сильную зависимость от температуры. Для обеспечения хорошей температурной стабильности рабочей точки значение тока в цепи делителя напряжения на резисторах R_1 и R_2 обычно выбирают в 5–10 раз больше рассчитанного тока базы, а в цепь эмиттера добавляют резистор $R_э$, осуществляющий отрицательную обратную связь по току. Подробнее о расчете электрических схем усилителей можно прочитать в [26, 131].

2.5.2. Описание транзистора с помощью h-параметров. При расчете электронных схем методами теории электрических цепей транзистор представляют в виде четырехполюсника (см. рис. 2.16*а*). Напряжение и ток на входе этого четырехполюсника будем обозначать буквами с индексом 1 (U_1, I_1), а напряжение и ток на выходе — буквами с индексом 2 (U_2, I_2). При описании четырехполюсников два из этих четырех параметров считаются независимыми переменными, а два других — зависящими от них функциями. Характеристики четырехполюсника можно представить в виде семейства кривых, на которых зависимость каждого

из зависимых параметров как функция одного из независимых параметров откладывается для ряда фиксированных значений второго независимого параметра. В качестве примера таких характеристик на рис. 2.15*б*, *в* представлены зависимость $I_к(V_{кэ})$ для ряда значений $I_б$ и зависимость $I_б(V_{бэ})$ при двух значениях $V_{кэ}$ для отечественного транзистора ГТ322А. Эти зависимости, представленные в параметризованной форме (модель Эберса–Молла [14]), используются при математическом моделировании сложных электронных схем с помощью компьютерных программ, наиболее популярной среди которых является программа SPICE.



а *б*

Рис. 2.16. Представление транзистора в виде четырехполюсника (*а*) и эквивалентная схема транзистора на низких частотах (*б*)

Вольт-амперные характеристики транзистора, вообще говоря, являются нелинейными. Однако когда уровни сигналов невелики, то связь между входными и выходными параметрами можно с хорошей точностью считать линейной. В этом случае расчет электронных схем существенно упрощается, поскольку транзистор можно рассматривать как «черный ящик» и описывать его набором дифференциальных (малосигнальных) параметров, которые равны отношению изменения зависимого параметра к вызвавшему его малому изменению одного из независимых параметров при фиксированном значении другого независимого параметра. В дальнейшем эти небольшие приращения входных и выходных параметров мы будем обозначать маленькими буквами (u_1, i_1, u_2, i_2).

Для описания четырехполюсника с помощью введенных выше дифференциальных параметров наиболее употребительными являются наборы z-, y- и h-параметров. Для набора z-параметров в качестве независимых переменных выбираются входной i_1 и выходной i_2 токи, для набора y-параметров ими служат входное u_1 и выходное u_2 напряжения, а для набора h-параметров независимыми переменными являются входной ток i_1 и выходное напряжение u_2.

Рассмотрим подробнее h-параметры, которые часто используются при расчете электронных схем на транзисторах на низких частотах. Линейные уравнения, выражающие зависимые параметры через независимые переменные, имеют вид

$$\begin{cases} u_1 = h_{11}i_1 + h_{12}u_2, \\ i_2 = h_{21}i_1 + h_{22}u_2. \end{cases} \tag{2.25}$$

Как следует из этих уравнений, параметр h_{11} характеризует входное сопротивление транзистора при постоянном напряжении на выходе ($u_2 = 0$). Параметр h_{12} характеризует коэффициент обратной связи по напряжению при неизменном входном токе ($i_1 = 0$). Параметр h_{21} характеризует коэффициент передачи тока в прямом направлении при постоянном напряжении на выходе ($u_2 = 0$). Параметр h_{22} характеризует выходную проводимость при неизменном входном токе ($i_1 = 0$). Система уравнений (2.25) позволяет представить транзистор на низких частотах в виде эквивалентной схемы, которая показана на рис. 2.16,*б*.

Необходимо отметить, что значения h-параметров различны для разных схем включения. Для обозначения схемы включения, к которой относятся эти параметры, после цифровых индексов в обозначении параметра указывают одну из букв: *э*, *б* или *к*. Зная значения h-параметров для одной из схем включения, можно вычислить соответствующие значения и для любой другой схемы включения [132]. Так, например, коэффициент усиления по току в схеме с ОЭ ($\beta = h_{21э}$) связан с коэффициентом усиления по току в схеме с ОБ ($\alpha = h_{21б}$) следующим соотношением:

$$\beta \approx \frac{\alpha}{1 - \alpha}. \tag{2.26}$$

В справочной литературе для многих транзисторов приводятся зависимости h-параметров от тока коллектора, которые обычно измеряются на частоте 1 кГц при некотором заданном напряжении на коллекторе. Такие зависимости для двух типичных кремниевых транзисторов показаны на рис. 2.17.

В принципе, если рассматривать h-параметры как комплексные величины, зависящие от частоты, то набор этих параметров можно использовать для расчета электронных схем, работающих в широком диапазоне частот. Однако для расчетов высокочастотных схем (>100 МГц) удобнее использовать y-параметры. Переход от h-параметров к наборам y- и z-параметров производится с помощью несложных формул, выводимых в теории электрических цепей [132].

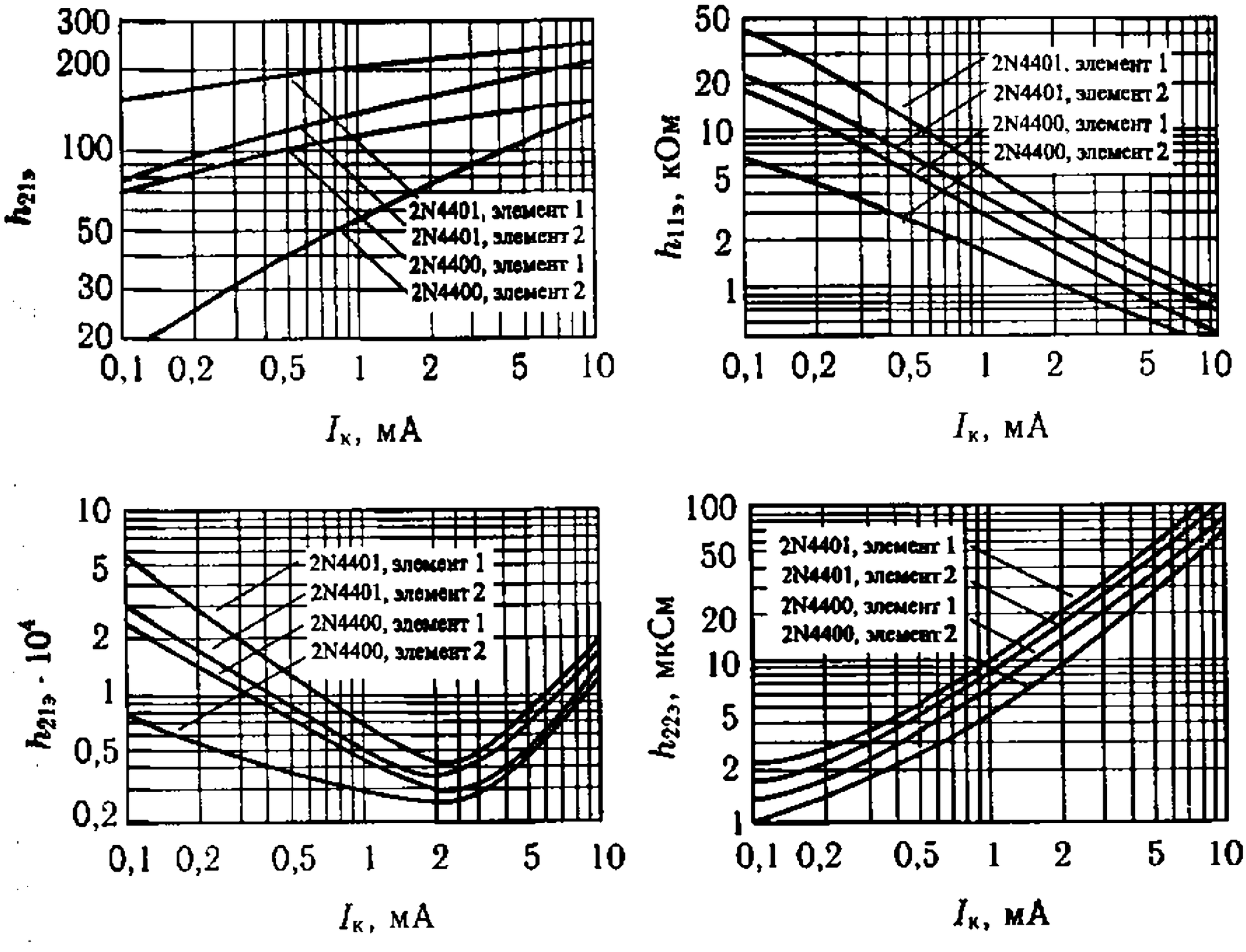


Рис. 2.17. Зависимость h-параметров кремниевых транзисторов 2N4400 и 2N4401 при $V_{кэ} = 10$ В и $T_{окр} = 25$ °С. Кривые, обозначенные элемент 1 и элемент 2, указывают типовой разброс параметров транзисторов [26]

Все h-параметры непосредственно связаны с внутренним устройством и характеристиками транзисторов, о которых мы говорили выше. Так, $h_{11б}$ представляет собой дифференциальное сопротивление эмиттерного p–n-перехода при прямом смещении ($h_{11б} \approx kT/qI_э$), $h_{21э}$ есть ни что иное как коэффициент усиления по току транзистора в схеме с ОЭ, а $h_{21б}$ — то же самое в схеме с ОБ. Выходная проводимость транзистора $h_{22э}$ связана с рассмотренным в п. 2.2.5 эффектом Эрли. Наконец, параметр $h_{12э}$ описывает изменение напряжения на эмиттерном переходе при фиксированном токе базы, вызванное изменением толщины нейтральной области базы W' при изменении напряжения на коллекторе. Как следует из формулы (2.4), уменьшение W' с ростом $V_{кэ}$ вызывает уменьшение $V_{бэ}$, то есть знак $h_{12э}$ должен быть отрицательным, однако из-за возрастания сопротивления базы с уменьшением W' и увеличения падения напряжения на этом сопротивлении при протекании тока в базе знак $h_{12э}$ в реальном транзисторе может измениться [113].

2.6. Шумы в биполярных транзисторах

Широкое использование транзисторов в схемах усиления электрических сигналов поставило задачу определения наименьших величин сигналов, которые еще могут быть измерены или усилены в условиях неизбежных для электронных схем шумов — флуктуаций напряжения и тока. По причинам возникновения шумов и их свойствам шумы принято подразделять следующим образом [49, 133]:

1. *Тепловой шум.* [1]) Он возникает в любом проводнике электрического тока и связан с хаотичным движением подвижных носителей заряда, в результате которого на контактах образца появляются флуктуации напряжения. Средний квадрат напряжения этого шума зависит только от активного сопротивления R и температуры T образца и может быть рассчитан по формуле Найквиста:

$$\bar{e}_{\text{ш}}{}^2 = 4kTR\Delta f, \tag{2.27}$$

где k — постоянная Больцмана, а Δf — полоса частот, в которой проводятся измерения. Спектральная плотность теплового шума $\bar{e}_{\text{ш}}{}^2/\Delta f$ остается постоянной вплоть до частоты $f_m = kT/(2\pi\hbar)$. При 300 К $f_m \approx 6 \cdot 10^{12}$ Гц.

2. *Дробовой шум.* Этот вид шума связан с дискретной природой электрического заряда и обусловлен тем, что при протекании некоторого тока число электронов, пересекающих заданную границу за заданный интервал времени, всегда дискретно и испытывает статистические флуктуации. Если движение отдельных зарядов независимо, то средний квадрат флуктуаций тока определяется только величиной протекающего тока I и рассчитывается по формуле Шоттки:

$$\bar{i}_{\text{ш}}{}^2 = 2qI\Delta f, \tag{2.28}$$

где q — заряд электрона. Спектральная плотность дробового шума $\bar{i}_{\text{ш}}{}^2/\Delta f$ остается постоянной до частоты, равной $f_m = 1/2\pi\tau$, где τ — время пролета. В электронных лампах f_m обычно составляет $\sim$100 МГц; в полупроводниковых приборах с малым временем пролета эта частота еще выше.

3. *Фликкер-шум.* Этот вид шума, также называемый *избыточным* или *мерцающим* шумом (от английского flicker —

---

[1]) В зарубежной литературе этот вид шума также называют *джонсоновским* по имени Дж. Джонсона, который в 1928 г. экспериментально установил закономерности этого шума.

мерцать), отличается от рассмотренных выше шумов сильной частотной зависимостью его спектральной плотности, которая часто имеет вид $1/f^{\alpha}$, где $\alpha \approx 1$. Этот шум проявляется *только при протекании тока через образец*. Причинами возникновения избыточного шума являются процессы генерации–рекомбинации в полупроводнике (флуктуации концентрации свободных носителей приводят к флуктуациям проводимости образца), поверхностные шумы, шумы утечки, контактные явления (флуктуации сопротивления контактов). Рассмотрим некоторые виды избыточного шума более подробно.

Генерационно-рекомбинационный шум вызван флуктуациями концентрации свободных носителей заряда, которые появляются и исчезают в полупроводнике в результате процессов генерации и рекомбинации. Спектральная плотность вызванных этими процессами флуктуаций тока через образец в общем случае имеет вид

$$S_i(\omega) = 4\,\frac{I_0^2}{N_0^2}\,\overline{\Delta N^2}\,\frac{\tau}{1+\omega^2\tau^2}\,, \qquad (2.29)$$

где I_0 — средняя величина тока, N_0 – среднее число носителей в образце, ΔN^2 — мощность их флуктуаций, τ — время жизни носителей, $\omega = 2\pi f$. В ряде практически важных случаев эта формула сводится к формуле

$$S_i(\omega) = \frac{4\beta I_0^2}{N_0}\,\frac{\tau}{1+\omega^2\tau^2}\,, \qquad (2.30)$$

где величина β определяется видом уравнений, описывающих кинетику процессов генерации и рекомбинации [134]. В частности, в собственном полупроводнике без глубоких уровней и в некомпенсированном примесном полупроводнике при низкой температуре $\beta = 1/2$, а в компенсированном полупроводнике при низкой температуре $\beta = 1$.

Поверхностный шум возникает вследствие генерационно-рекомбинационных процессов, идущих с участием поверхностных состояний.

Изучение частотных зависимостей эффекта поля, позволившее исследовать кинетику экранирования внешнего электрического поля поверхностными состояниями, показало, что на поверхности полупроводника имеются два типа уровней, резко различающихся временами релаксации [135]. Первые из них, называемые *«быстрыми»* поверхностными состояниями, имеют характерное время перезарядки $\sim 10^{-8}$ с и наблюдаются даже на атомарно-чистых поверхностях, сколотых

в сверхвысоким вакууме. Поэтому эти состояния связывают с оборванными связями на поверхности кристалла. *«Медленные»* поверхностные состояния, появляющиеся на поверхностях, покрытых окисной пленкой или адсорбированными молекулами, характеризуются временами перезарядки от единиц миллисекунд до нескольких часов. В рекомбинационных процессах медленные поверхностные состояния проявляются в основном как центры захвата (ловушки), а быстрые — как центры поверхностной рекомбинации.

Перезарядка медленных состояний в результате захвата на них и выброса с них носителей заряда приводит к изменению локальных электрических полей, которые вызывают модуляцию проводимости приповерхностного слоя полупроводника и одновременно влияют на скорость поверхностной рекомбинации. Эти изменения проводимости и времени жизни и являются причиной возникновения поверхностного шума. Предполагая, что уровни медленных состояний равномерно распределены по толщине окисла и обмениваются носителями с зонами путем туннелирования, удается объяснить основные закономерности этого шума — его слабую зависимость от температуры и частотную зависимость его спектральной плотности, близкую к $1/f$ [49].

Поверхностный шум очень чувствителен к состоянию атмосферы, окружающей полупроводник (например, шум резко возрастает во влажной атмосфере). Причиной этого являются процессы адсорбции–десорбции, идущие на поверхности полупроводника. Вклад этих процессов в поверхностный шум может быть значительно снижен путем соответствующей обработки поверхности.

Поверхностный шум особенно сильно проявляется в полевых транзисторах с изолированным затвором (см. гл. 4). Однако имеются убедительные доказательства и того, что появление фликкер-шума в диодах и биполярных транзисторах также связано с поверхностным шумом — флуктуациями скорости поверхностной рекомбинации в местах выхода p–n-переходов на поверхность.

В p–n-переходах из-за приповерхностного изгиба зон по периметру перехода может возникать тонкий слой, тип проводимости которого противоположен типу проводимости в объеме полупроводника. Наличие этого слоя приводит к появлению тока утечки, который быстро возрастает с увеличением обратного смещения. Флуктуации тока утечки создают *шум утечки*, который сильно зависит от окружающей атмосферы. Этот шум обычно незаметен при прямых смещениях, но может оказаться очень сильным при высоких обратных напряжениях. Надлежащая обработка поверхности позволяет ослабить шум утечки до такой величины, которую можно не учитывать при невысоких обратных смещениях.

Для того, чтобы найти оптимальный (в смысле уровня шума) режим работы транзистора, рассмотрим эквивалентную схему шумов транзисторного усилителя (см. рис. 2.18). Для простоты ограничимся анализом случая средних частот, когда взаимной корреляцией различных источников шума (которая проявляется

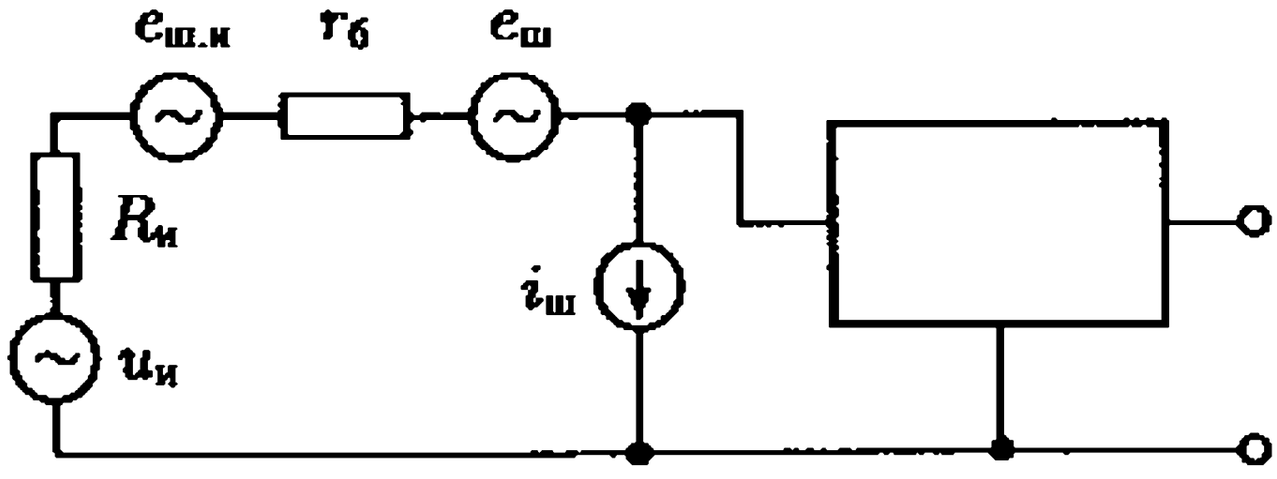


Рис. 2.18. Эквивалентная схема шумов биполярного транзистора в области средних частот

на высоких частотах) можно пренебречь, а уровень фликкер-шума (который проявляется на низких частотах) уже сравнительно мал. [1]) Будем считать, что коэффициент усиления схемы достаточно велик, так что вкладом шумов, возникающих в коллекторной цепи транзистора, и шумов последующих каскадов в общий шум можно пренебречь. В схеме, показанной на рисунке, транзистор характеризуется двумя источниками шума: генератором напряжения $e_ш$, включенным последовательно с выводом входной цепи идеального (бесшумного) усилителя, и генератором тока $i_ш$, включенным параллельно входной цепи. Величина шумового напряжения $e_ш = \sqrt{\bar{e}_ш^{\,2}}$ определяется двумя источниками шума: тепловым шумом, генерируемым в базе транзистора (сопротивление которой равно $r_б$), и дробовым шумом тока эмиттера, флуктуации которого выделяются на дифференциальном сопротивлении эмиттерного перехода $r_э = kT/qI_э$. Средний квадрат амплитуды последнего напряжения равен

$$\bar{e}_{ш.э}^{\,2} = r_э^2 \bar{i}_{ш.э}^{\,2} = r_э^2 \cdot 2qI_э\Delta f,$$

то есть его величина эквивалентна тепловому шуму, возникающему на сопротивлении, равном $r_э/2$. Поэтому

$$\bar{e}_ш^{\,2} = 4kT\left(r_б + \frac{r_э}{2}\right)\Delta f = 4kT\left(r_б + \frac{kT}{2qI_э}\right)\Delta f. \qquad (2.31)$$

Шумовой ток $i_ш = \sqrt{\bar{i}_ш^{\,2}}$ определяется флуктуациями тока во входной цепи. Для транзистора, включенного по схеме с ОЭ, квадрат флуктуаций тока базы равен

$$\bar{i}_ш^{\,2} = 2qI_б\Delta f = 2q\frac{I_э}{\beta_{ст}}\Delta f, \qquad (2.32)$$

где $\beta_{ст}$ — коэффициент усиления транзистора по току в схеме с ОЭ.

---

[1]) Подробный анализ шумовых характеристик транзисторов в широкой области частот можно найти в работе [49].

Протекание этого шумового тока через сопротивление источника сигнала $R_и$ и сопротивление базы $r_б$ вызывает дополнительное падение напряжения на этих сопротивлениях. Поэтому квадрат эффективного напряжения шума транзистора, приложенного ко входу усилителя, равен

$$\bar{e}^2_{эф} = \bar{e}^2_ш + (R_и + r_б)^2 \bar{i}^2_ш. \qquad (2.33)$$

Отношение квадрата напряжения полного шума (определяемого и источником сигнала, и усилителем) к квадрату напряжения теплового шума источника сигнала (который характеризуется напряжением шума $e_{ш.и}$) называется *коэффициентом шума* (*шум-фактором*) усилителя и равно

$$F = \frac{\bar{e}^2_{ш.и} + \bar{e}^2_{эф}}{\bar{e}^2_{ш.и}} = 1 + \frac{\bar{e}^2_{эф}}{4kTR_и\Delta f}. \qquad (2.34)$$

Эта величина характеризует, насколько сильно шумы реального транзистора ухудшают отношение сигнал/шум. Иногда вместо коэффициента шума используется понятие *температуры шума*, которая вычисляется по формуле $300(F-1)$ и выражается в градусах Кельвина.

Найдем теперь условия получения минимального коэффициента шума в усилителе на биполярном транзисторе. Поскольку, в соответствии с формулами (2.31) и (2.32), в уравнении (2.33) одно слагаемое возрастает, а другое уменьшается при изменении $I_э$, то подбирая ток эмиттера, можно найти такой режим, при котором величины $\bar{e}^2_{эф}$ и F достигают минимума. Вычисляя производную выражения (2.33) по $I_э$ и приравнивая ее нулю, нетрудно показать, что минимум достигается при

$$I_э = \frac{kT}{q}\frac{\sqrt{\beta}}{R_и + r_б}. \qquad (2.35)$$

Подставляя найденное значение тока в (2.33) и затем в (2.34), окончательно получаем

$$F = \left(1 + \frac{r_б}{R_и}\right)\left(1 + \frac{1}{\sqrt{\beta}}\right). \qquad (2.36)$$

Из этой формулы следует, что при разработке малошумящих усилителей необходимо использовать транзисторы с возможно более низким сопротивлением базы $r_б$ и максимально высоким коэффициентом усиления β. К сожалению, эти требования противоречат друг другу: как мы показали в п. 2.2.1, для получения

высокого коэффициента усиления необходима низкая концентрация примесей в базе, но это неизбежно приводит к увеличению сопротивления базы. Поэтому при проектировании транзисторов между величинами $r_б$ и β приходится искать компромисс. Уменьшить величину $r_б$ позволяет использование гребенчатой конструкции выводов транзистора, которая была описана нами на с. 159.

В современных транзисторах характерное значение $r_б$ составляет ~100 Ом, однако в специально разработанных транзисторах (например, 2SA1316 и 2SC3329 фирмы Toshiba) $r_б$ может быть снижено до 2 Ом. Поэтому при типичном коэффициенте усиления $\beta \approx 400$ и выходном сопротивлении источника сигнала 1 кОм коэффициент шума усилителя на биполярном транзисторе может составлять $F \approx 1{,}1$. Это означает, что на биполярных транзисторах можно создавать усилители, в которых вклад шумов транзистора в общий уровень шума очень мал.

Следует отметить, что низкий уровень шума должен быть присущ всем гетеропереходным транзисторам. Это является следствием низкого сопротивления базы и высокого коэффициента усиления по току, которые «заложены» в конструкцию этих транзисторов (см. п. 2.4.2).

2.7. Особенности работы транзисторов в импульсном режиме

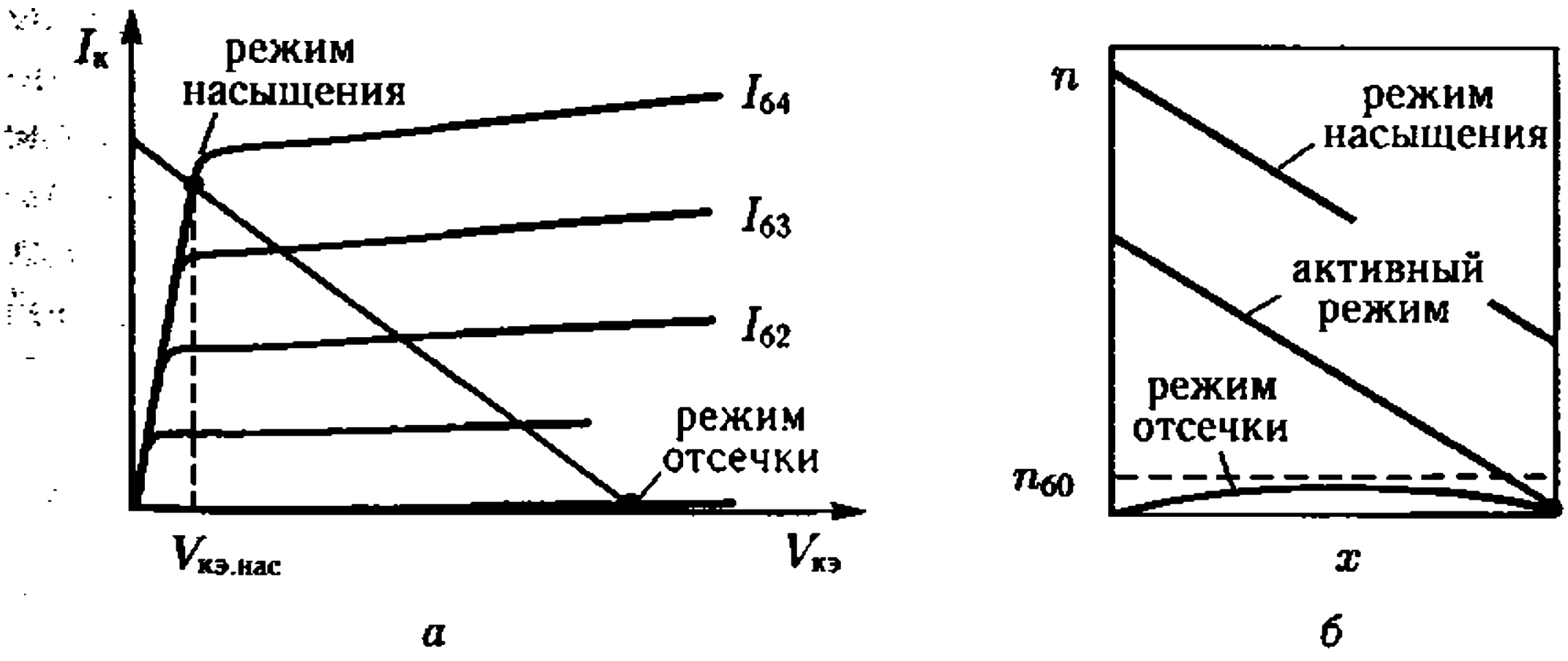


Рис. 2.19. *а* — рабочие точки на нагрузочной прямой, отвечающие режиму отсечки и режиму насыщения; *б* — распределение инжектированных носителей в базе транзистора при различных режимах работы

В электронных схемах, работающих в импульсном режиме (триггерах, одновибраторах, мультивибраторах, элементах

цифровых интегральных схем), транзистор обычно находится в одном из двух состояний: *режиме отсечки*, при котором на эмиттерный переход подается смещение $V_{бэ} \leqslant 0$ и транзистор закрыт, или *режиме насыщения*, при котором через базу транзистора пропускается такой прямой ток, что оба p–n-перехода транзистора смещены в прямом направлении и транзистор открыт. Рабочие точки, отвечающие этим двум состояниям, показаны на рис. 2.19 *а*.

Распределение инжектированных носителей в базе транзистора, находящегося в режиме насыщения, режиме отсечки и активном (усилительном) режиме, показано на рис. 2.19 *б*. Как следует из этого рисунка, в режиме насыщения оба p–n-перехода транзистора смещены в прямом направлении и концентрация неосновных носителей у обеих границ базы превышает $n_{б0}$. При этом напряжение между эмиттером и коллектором (напряжение насыщения $V_{кэ.нас}$) составляет ~0,1 В.

При использовании биполярных транзисторов в импульсных схемах особое значение приобретает вопрос о *быстродействии* этих схем. В качестве характеристик быстродействия вводят понятия времен задержки включения и задержки выключения схемы, которые в значительной мере определяются временами открывания и закрывания транзистора.

Открывание транзистора происходит в три этапа. На первом этапе инжектируемые эмиттером носители достигают коллекторного перехода и ток коллектора начинает возрастать. На втором этапе по мере увеличения концентрации носителей у коллектора происходит быстрое нарастание тока коллектора и уменьшение $V_{кэ}$ до того момента, пока транзистор не войдет в режим насыщения. Наконец, на третьем этапе происходит установление стационарного распределения инжектированных носителей в базе, сопровождаемое накоплением в ней избыточного заряда этих носителей. Чем выше ток базы, тем быстрее протекает второй этап, но при этом на третьем этапе неизбежно происходит накопление все бóльшего избыточного заряда в базе.

Закрывание транзистора происходит в два этапа. На первом этапе идет процесс рассасывания инжектированных в базу носителей за счет их экстракции через эмиттерный и коллекторный переходы и рекомбинации (эти процессы аналогичны процессам, происходящим при переключении диодов, см. п. 1.7.3). Этот этап заканчивается, когда транзистор выходит из состояния насыщения. На втором этапе, называемом *областью динамической отсечки*, происходит уменьшение тока коллектора и возрастание напряжения $V_{кэ}$ в соответствии с темпом экстракции

коллекторным переходом оставшегося заряда инжектированных носителей из середины базы.

Для увеличения быстродействия транзисторов в цифровых интегральных схемах транзисторно-транзисторной логики (ТТЛ) базы транзисторов легируются примесью золота, которое создает в Si глубокие уровни и резко уменьшает время жизни неравновесных носителей, а сама принципиальная схема логического элемента построена так, что позволяет осуществить эффективную экстракцию носителей из баз транзисторов [136]. К сожалению, легирование золотом увеличивает ток рекомбинации в области пространственного заряда, что снижает коэффициент усиления транзисторов в области малых токов.

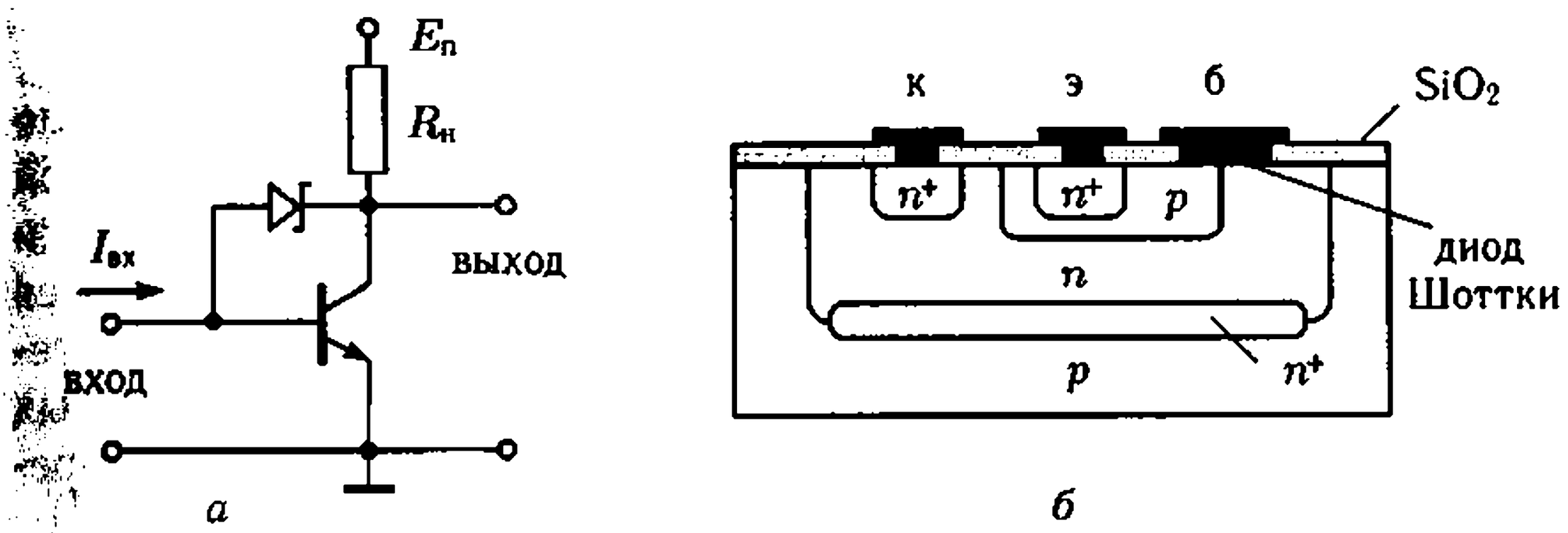


Рис. 2.20. Включение «фиксирующего» диода Шоттки для предотвращения глубокого насыщения транзистора: *а* — эквивалентная схема, *б* — топология прибора

Однако даже при использовании явления экстракции из-за ее ограниченной скорости время закрывания транзистора все равно остается больше времени открывания транзистора. Очевидно, что для уменьшения времени закрывания следует избегать накопления в базе транзистора высокой концентрации инжектированных носителей. Практически это делают, шунтируя переход база–коллектор «фиксирующим» диодом с барьером Шоттки (см. рис. 2.20*а*). При таком включении входной ток схемы $I_{вх}$ может быть довольно большим (чтобы обеспечить быстрое открывание транзистора), а после перехода транзистора в открытое состояние заметная часть входного тока протекает через диод Шоттки, ограничивая ток базы такой величиной, которая достаточна для удержания транзистора в открытом состоянии. При этом используется известная особенность диодов Шоттки открываться при более низком прямом напряжении по сравнению с p–n-переходами (причины этого обсуждались нами в п. 1.5.2), и рабочая точка транзистора находится в области активного

режима вблизи точки, отвечающей режиму насыщения. Для создания описываемой конструкции достаточно лишь немного расширить контакт к базе, наложив его на коллектор (см. рис. 2.20*б*). В качестве материала контактов обычно используют силициды металлов, характеризуемые наибольшей высотой барьера к n-Si, например PtSi [58]. Описываемые «транзисторы Шоттки» широко используются при создании современных цифровых интегральных схем на биполярных транзисторах (в частности, в интегральных схемах транзисторно-транзисторной логики с диодами Шоттки — ТТЛШ [136]).

2.8. Элементы интегральных схем на биполярных транзисторах

Бурное развитие полупроводниковой электроники привело к значительным изменениям ее элементной базы, основу которой в настоящее время составляют интегральные схемы (ИС). Среди современных интегральных схем наиболее сложными являются цифровые (логические) ИС на основе кремния. Уровень интеграции этих схем, который определяется числом эквивалентных логических вентилей (типа «2И-НЕ») в одном корпусе, уже превзошел величину 10^7 и продолжает стремительно нарастать. [1])

Концепция интегральной электронной схемы, элементы которой были бы созданы только из полупроводника и изолирующих слоев, была впервые высказана английским экспертом по радарам Даммером в 1952 г. Предпринятые им в 1956 г. попытки построить интегральную схему не увенчались успехом, и первые интегральные схемы были созданы только в 1958–59 годах независимо Джеком Килби из фирмы Texas Instruments [137] и Робертом Нойсом из Fairchild Semiconductors [138]. Это изобретение сыграло настолько большую роль в развитии современной полупроводниковой электроники, что один из создателей интегральной схемы — Джек Килби — был удостоен Нобелевской премии по физике в 2000 г.

В 50-е годы различные компоненты полупроводниковых электронных схем (диоды, транзисторы, резисторы, конденсаторы) изготавливались из совершенно разных материалов, используя при этом наборы совершенно различных технологических операций. Основной идеей

---

[1]) По уровню интеграции ИС условно подразделяют на схемы с низким, средним, большим (БИС) и сверхбольшим (СБИС) уровнем интеграции. Уровню СБИС отвечают микросхемы, содержащие более 10^4 вентилей или более 10^5 элементов памяти.

разработчиков интегральной схемы было сделать так, чтобы все элементы электронной схемы могли быть изготовлены из одного и того же материала — полупроводника, используя при этом минимальный набор технологических операций. Первая интегральная схема такого типа, содержавшая пять элементов, была представлена Килби в сентябре 1958 года. Это была схема генератора, в котором все элементы, изготовленные в объеме и на поверхности Ge, соединялись между собой тонкими золотыми проволочками. В 1959 г. Нойс предложил, используя найденные при разработке транзисторов технологические решения (локальную диффузию с защитой окислом, фотолитографию, травление, напыление, изоляцию элементов p–n-переходами, а также идею Хорни изготовления транзисторов с планарной геометрией), оставить закрытой окислом всю поверхность полупроводника (кроме контактных выводов) и осуществлять соединение между элементами схемы дорожками из алюминия, которые либо напылялись через маску, либо создавались путем избирательного вытравливания ненужных участков в тонкой пленке Al, напыленной на поверхность структуры. Таким образом Нойсом в 1959 году была создана первая микросхема из кремния.

2.8.1. Планарная технология. Планарная технология основывается на ряде технологических приемов, разработанных в 50-х годах в процессе производства полупроводниковых диодов и транзисторов. Основная идея планарной технологии заключается в том, чтобы, используя минимальный набор стандартных технологических операций, формировать все элементы полупроводникового прибора в приповерхностном слое на одной из сторон пластины полупроводника, причем так, чтобы после каждой из этих операций поверхность пластины оставалась плоской (планарной).

Использование в планарной технологии одного и того же набора многократно повторяемых технологических операций (окисления, травления, диффузии, ионной имплантации, эпитаксиального наращивания, осаждения диэлектрических и металлических пленок), выполняемых в сочетании с литографией, которая позволяет с высокой геометрической точностью переносить рисунок элементов структуры на пластину полупроводника, делает эту технологию достаточно универсальной.

Преимущество планарной технологии состоит в том, что формируя элементы полупроводниковой структуры на пластине большого диаметра (100–300 мм), можно одновременно создавать многие миллионы диодов, транзисторов и других элементов, что существенно удешевляет стоимость их производства. В настоящее время эта технология используется для создания как

одиночных «дискретных» приборов, так и сложных интегральных микросхем, состоящих из миллионов элементов.

Кремний является не единственным материалом, из которого изготавливают интегральные схемы. В настоящее время для этих целей также используются арсенид галлия и фосфид индия, которые выпускаются промышленностью в виде пластин диаметром 100 мм. Подвижность и скорость насыщения электронов в GaAs и InP заметно выше, чем в кремнии, и поэтому эти материалы долго рассматривались как материалы, которые придут на смену кремния в быстродействующих приборах [69]. Однако в последние годы интерес к GaAs как материалу для будущих сверхбыстродействующих ИС стал ослабевать. Причиной этому являются два обстоятельства: 1) примерно втрое более низкая теплопроводность GaAs по сравнению с Si, которая не позволяет эффективно отводить от арсенида галлия достаточно высокую рассеиваемую мощность, характерную для современных ИС, и 2) высокая хрупкость GaAs. Это является серьезным препятствием на пути создания арсенид-галлиевых СБИС с уровнем интеграции, близким к достигнутому в настоящее время на кремнии. Тем не менее, GaAs остается важным материалом для производства сверхбыстродействующих цифровых и аналоговых ИС среднего уровня интеграции. Второй из обсуждаемых материалов — InP — технологически менее развит по сравнению с GaAs, однако созданные на его основе приборы в настоящее время имеют самое высокое быстродействие. Из этих двух материалов изготавливают D-триггеры, делители частоты, аналого-цифровые преобразователи и целый ряд аналоговых сверхвысокочастотных ИС, необходимых для спутниковой и оптоволоконной связи. Использованию германия для создания интегральных схем мешает невысокая максимальная рабочая температура изготавливаемых из него диодов и транзисторов.

Рассмотрим сначала основные технологические приемы, используемые в планарной технологии на кремнии, а затем проиллюстрируем их применение на примере создания биполярных транзисторов (дискретных приборов и элементов интегральных схем). Особенности технологии создания полевых транзисторов с МОП-структурой и ИС на их основе будут рассмотрены нами в п. 4.2.

Фотолитография. Некоторые специалисты полагают, что половиной своих поразительных успехов современная микроэлектроника обязана развитию техники литографии. [1]) *Оптическая*

---

[1]) Литографией называют разработанный в конце XVIII века способ переноса изображения на бумагу с плоской (нерельефной) печатной формы, при котором места на поверхности печатной формы, которые должны дать отпечаток, покрываются специальным составом, который смачивается печатной краской.

литография (*фотолитография*) как метод, позволяющий создавать на поверхности полупроводника защитные маски с окнами любой формы для проведения последующих технологических операций, был разработан Андрусом и Бондом (Bell Laboratories) в 1955–57 гг.

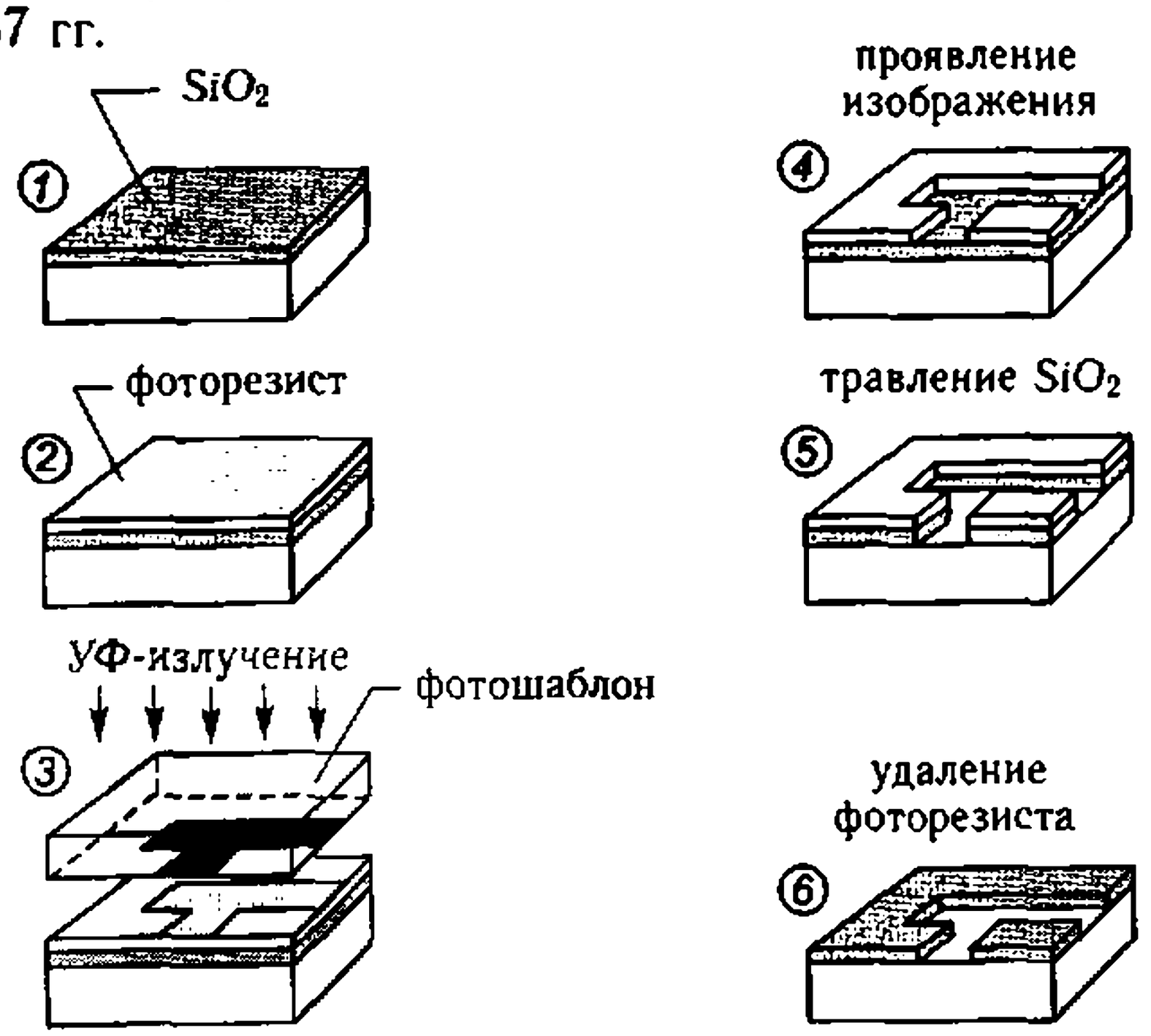


Рис. 2.21. Основные этапы процесса фотолитографии

Рассмотрим основные этапы фотолитографического процесса на примере формирования окон в слое окисла на поверхности кремния для последующей диффузии примесей. Участки, на которых будет проводиться диффузия, задаются на *фотошаблоне* [1]) в виде некого рисунка, и задача литографии состоит в том, чтобы удалить окисел с поверхности кремния точно в указанных местах. Для этого поступают следующим образом (см. рис. 2.21). На предварительно окисленную поверхность кремния наносится тонкий (0,2–1 мкм) слой *фоторезиста* — материала (обычно представляющего собой смесь полимера с сенсибилизатором), который способен изменять свои физические свойства в результате облучения его синим или ультрафиолетовым излучением.

Существуют два способа переноса изображения на фоторезист: *контактный*, при котором фоторезист засвечивается через прижатый к пластине кремния фотошаблон, и *проекционный*,

[1]) Фотошаблон обычно представляет собой кварцеваю пластину, на которой в виде хромового покрытия представлен подлежащий переносу рисунок.

при котором изображение рисунка проецируется на фоторезист с помощью оптической системы. При контактном способе экспонирования для исключения возможности повреждения слоя фоторезиста и рисунка на фотошаблоне последний располагают на небольшом расстоянии от поверхности пластины. Поэтому контактный способ обеспечивает не слишком высокое пространственное разрешение. Проекционный метод с оптическим уменьшением изображения в 4–5 раз позволяет получить более высокое разрешение и поэтому преимущественно используется в настоящее время.

В результате фотохимических реакций, идущих под действием света, фоторезист меняет свои свойства [139]: в так называемых *позитивных фоторезистах* (например, полиметилметакрилате) ультрафиолетовое облучение разрушает полимерные связи и остатки полимера на экспонированных участках растворяются в растворителе (метилизобутилкетоне) более, чем в 10 раз быстрее, чем неэкспонированные участки. В другом, известном еще с 30-х годов, позитивном фоторезисте — смеси диазонафтохинона с новолаком — ультрафиолетовое облучение трансформирует гидрофобные (отталкивающие воду) молекулы диазосоединения в гидрофильную кислоту, и экспонированные участки фоторезиста затем легко смачиваются растворителем (раствор щелочи) и удаляются. В *негативных фоторезистах* (например, смесях циклических полиизопренов с диазидами или в специально синтезированных азидокаучуках), наоборот, под действием света начинается распад азидов, образуются реакционно-способные радикалы, которые осуществляют поперечную сшивку линейных полимерных цепочек с образованием на экспонированных участках нерастворимых объемно сшитых полимеров. В последнее время для инициации полимеризации все чаще используют ониевые соли, в результате фотодиссоциации которых образуются кислоты Льюиса, обладающие каталитическим действием [140]; это позволяет значительно повысить чувствительность негативных фоторезистов.

При помещении экспонированной пластины в растворитель изображение «проявляется» (растворимые участки фоторезиста удаляются) и рисунок с фотошаблона оказывается представленным на поверхности пластины в виде островков полимера. После задубливания резиста, при котором его адгезия к слою окисла и кислотостойкость возрастают, пластину помещают в раствор смеси NH_4F+HF и растворяют участки SiO_2, не защищенные фоторезистом. Наконец, после удаления задубленного фоторезиста с помощью химического или плазменного травления (см. с. 202) получается пластина, на окисленной поверхности которой в соответствии с рисунком на фотошаблоне открыты участки незащищенной поверхности кремния.

На сегодняшний день минимальный размер элементов в интегральных схемах (минимальная длина затвора в случае полевых транзисторов) составляет 0,055–0,09 мкм и продолжает неуклонно уменьшаться, [1]) поэтому возможности оптической литографии приближаются к своему теоретическому пределу, налагаемому дифракцией. Уменьшение размеров элементов заставило в середине 80-х годов перейти от ртутных ламп к использованию источников света с более короткими длинами волн. Сейчас ими являются эксимерные лазеры KrF и ArF, работающие в области вакуумного ультрафиолета ($\lambda = 248$ и 193 нм, соответственно). Для создания элементов, размеры которых лежат вблизи или за дифракционным пределом, создаются специальные фотошаблоны, в которых для коррекции дифракционных искажений вводятся специальные «предыскажения» (фазовые сдвиги), однако изготовление таких фотошаблонов обходится довольно дорого.

Непрерывное совершенствование техники оптической литографии отодвигает тот момент, когда ее возможности будут исчерпаны. Так, в настоящее время быстро развивается техника *иммерсионной литографии*, в которой пространство между линзой и слоем фоторезиста заполняется жидкостью (обычно водой). Уменьшение длины волны при распространении света в жидкости (показатель преломления воды на длине волны 193 нм составляет $n_r \approx 1,44$) и увеличение апертуры позволяют достигнуть более высокого разрешения. Эта техника уже была опробована в промышленности и позволила получить минимальный размер элементов 45 нм. По прогнозам, к 2009 г. с ее помощью в оптической литографии будет достигнуто разрешение 32 нм.

В качестве альтернативных методов литографии для получения более высокого разрешения в настоящее время разработаны методы *электронно-лучевой литографии*, в которой изображение «рисуется» на резисте с помощью электронного луча в сканирующем электронном микроскопе, *рентгеновская*

---

[1]) Минимальный размер 90 нм характерен для основной массы выпускаемых в настоящее время процессоров. Первые процессоры, выполненные по 65 нм технологии, появились в 2006 г., а в 2008 г. должны появиться процессоры с минимальным размером элементов 45 нм. Микросхемы динамической (см. с. 275) и флэш-памяти (см. с. 289) с минимальным размером элементов 90 нм начали производиться еще в 2005 г. По состоянию на 2006 г., наименьший размер элементов (55 нм) имеет флэш-память, выпускаемая фирмой Samsung.

литография, в которой резист экспонируется мягким рентгеновским излучением от источника синхротронного излучения или более дешевого импульсного источника — плазмы, возбуждаемой лазером, и *ионная литография*, при которой фоторезист «засвечивается» пучком высокоэнергичных ионов [58]. Электронно-лучевая литография с непосредственным экспонированием резиста электронным пучком имеет недостаточную для массового производства производительность и поэтому используется в основном для создания экспериментальных структур и фотошаблонов для оптической и рентгеновской литографии.

Наиболее перспективной для создания приборов с субмикронными размерами ($< 0,1$ мкм) представляется *проекционная ионная литография*, первые практические шаги в реализации которой были сделаны в 1985 г. В этом методе широкий пучок ионов водорода или гелия с энергией 10 кэВ падает на кремниевую мембрану толщиной 2–3 мкм и диаметром 150–200 мм, в которой с помощью электронно-лучевой литографии сделаны сквозные окна в соответствии с требуемым рисунком. Прошедшие через окна ионы ускоряются до энергии 50–200 кэВ, фокусируются системой электростатических линз с уменьшением размера изображения в 3–20 раз и направляются на слой резиста. Из-за более тяжелой массы ионов боковое рассеяние ионов в резисте оказывается меньше, чем для электронов, и поэтому в ионной литографии достигается более высокое разрешение по сравнению с электронно-лучевой литографией. Кроме того, в отличие от электронно-лучевой и рентгеновской литографии, ионная литография практически не вносит радиационных повреждений в расположенный под слоем резиста полупроводник.

Литография является наиболее важной операцией планарной технологии, поскольку от ее разрешающей способности и точности совмещения зависит плотность упаковки элементов в ИС. Литография является и наиболее часто используемой в планарной технологии операцией. Так, при производстве микросхем памяти емкостью 1 Гбит (разработка 2002 г.) требуется 31 операция литографии, 18 из которых являются «критическими», то есть требующими максимального разрешения и точности совмещения.

Более подробную информацию о разных методах литографии, изготовлении шаблонов, свойствах фоторезистов можно найти в работах [58, 139–143].

Технология локальной диффузии. Возможности диффузии как способа введения примесей III и V групп в Ge и Si для создания в полупроводнике областей *p*- и *n*-типа проводимости были впервые продемонстрированы Пфанном из Bell Laboratories и Холлом и Данлэпом из General Electric в 1949–1950 гг.

В 1955 г. Дерик и Фрош (Bell Laboratories) запатентовали использование слоя окисла, созданного на поверхности кремния при его термическом окислении, в качестве *маски* для защиты Si от проникновения в него диффундирующих примесей. Основанием для этого является то, что коэффициент диффузии многих важных легирующих примесей (B, P, As) в SiO_2 на порядок меньше, чем в кристаллическом Si, и поэтому слой окисла толщиной 0,5–1 мкм может служить эффективной защитой, не позволяющей указанным примесям проникать с поверхности в лежащий под окислом кремний. Заметим, что этот метод не годится для защиты кремния при легировании его примесями Ga и Al, коэффициент диффузии которых в SiO_2 выше, чем в Si.

Для создания защитного слоя окисла кремниевую пластину нагревают в атмосфере сухого или влажного кислорода или просто в парах воды. Было установлено, что присутствие в кислороде даже небольшой примеси паров воды ($\sim 10^{-3}$ %) вызывает резкое увеличение скорости окисления (на 1–1,5 порядка). Слой образующегося SiO_2 оказывается в 2,2 раза толще слоя израсходованного на окисление кремния, и поэтому на границе Si/SiO_2 возникают сильные механические напряжения, которые могут привести к повреждению структуры. По этой причине окисление обычно проводят при температурах выше 950 °C, при которых возникающие упругие напряжения снимаются за счет вязкого течения образующегося стеклообразного (аморфного) SiO_2. Для получения пленки толщиной 1 мкм время окисления во влажном кислороде при атмосферном давлении и 1000 °C составляет ~4 часа [58].

Следует отметить, что в современной технологии слои SiO_2 используются не только в качестве защитных масок, но и как *источник примеси* при проведении диффузии. Дело в том, что, попадая в SiO_2, примеси бора и фосфора образуют с двуокисью кремния сравнительно легкоплавкие боросиликатное и фосфорно-силикатное стекло. При обычных для диффузии температурах эти стекла становятся жидкими, они прекрасно смачивают поверхность кремния и растекаются по ней. Помимо создания на поверхности однородной содержащей примесь пленки, это покрытие выполняет очень важную функцию — предохраняет поверхность кремния от *эрозии* (газового травления). По этой

причине первый этап операции диффузии (так называемую *загонку примеси*) проводят в проточной системе, добавляя в поток газа, содержащий примесь, еще и кислород. При этом открытые участки поверхности кремния покрываются тонким слоем окисла, который реагирует с находящейся в газовой фазе примесью и образует соответствующее примесно-силикатное стекло. На втором этапе операции диффузии, называемом *разгонкой примеси*, примесные атомы диффундируют из этой пленки стекла в глубь кристалла. Описанная технология позволяет получить диффузионные слои с концентрацией примеси до 10^{18} см$^{-3}$. Типичные температуры загонки и разгонки примеси в кремнии составляют 1000 и 1200 °С, соответственно.

Фосфорно-силикатное стекло выполняет еще одну важную функцию: оно служит *геттером* примесей щелочных и некоторых переходных металлов, которые могут попадать в структуры во время их изготовления. Как мы увидим в п.п. 3.1.1 и 4.1.1, активизирующиеся во время термообработки примеси золота, меди и железа могут существенно уменьшать время жизни в Si, а примеси щелочных металлов — вызывать нестабильность характеристик МОП-транзисторов.

Ионная имплантация. Говоря о легировании полупроводников, нельзя обойти вниманием метод *ионной имплантации* [108, 144], заключающийся в бомбардировке поверхности кристалла ионами легирующих примесей, ускоренными до энергии 5–500 кэВ. В результате этой бомбардировки примеси внедряются в кристалл и распределяются там по глубине так, что их профиль распределения имеет приблизительно гауссову форму с максимумом, находящимся на некотором расстоянии от поверхности. Глубина залегания этого максимума зависит от массы и энергии ионов и обычно составляет 0,03–1 мкм. [1)]

Облучение высокоэнергичными ионами как метод направленного изменения свойств полупроводников был впервые использован в 1950–52 гг. Олом [145] для улучшения характеристик кремниевых диодов, фотопреобразователей и точечных транзисторов. Несколько позже

[1)] При имплантации полупроводника ионами, движущимися в направлении одной из кристаллических осей, рассеяние ионов в кристалле резко ослабляется и наблюдается явление *каналирования*, состоящее в аномально сильном проникновении ионов в кристалл. В кремнии это явление наиболее сильно выражено для ионов, движущихся в направлении {110}. Для устранения нежелательного влияния этого эффекта на профиль легирования, оси кристалла во время имплантации разориентируют на угол 5–10° относительно направления пучка ионов.

Бредов с соавт. [146] изучали влияние облучения ионами K на вольт-амперные характеристики точечных диодов. В 1955 г. Кассинс [147] пытался использовать ионную имплантацию для легирования германия, но возникающие при бомбардировке образцов радиационные дефекты помешали ему добиться успеха. В 1957 г. Шокли [148] получил патент на способ создания с помощью ионной имплантации заглубленного слоя, тип проводимости которого противоположен типу проводимости остальных частей кристалла; в патенте указывалось на необходимость проведения постимплантационного отжига для устранения радиационных дефектов. Первыми изготовленными с помощью ионной имплантации полупроводниковыми приборами были детекторы ядерных излучений [149] и элементы солнечных батарей [150, 151]. С середины 70-х годов ионная имплантация становится стандартной технологической операцией при производстве полупроводниковых приборов.

По сравнению с диффузией основными преимуществами ионной имплантации как метода легирования являются: возможность более тонкого управления профилем и концентрацией вводимой примеси; возможность получения слоев с концентрацией примеси выше 10^{18} см$^{-3}$; возможность создания очень мелких ($\lesssim 0,1$ мкм) легированных слоев и высокая степень чистоты (как следствие использования масс-спектрометрических сепараторов для выделения необходимого изотопа примеси при имплантации). Кроме того, ионная имплантация характеризуется невысоким рассеянием внедряемых ионов в поперечном направлении (20–40% от средней глубины проникновения), что позволяет при использовании этого метода сохранять высокое разрешение, задаваемое фотолитографией, и создавать легированные области очень малых размеров.

Поскольку при ионной имплантации в кристаллах возникает высокая концентрация радиационных дефектов (число которых обычно намного превосходит число внедренных атомов), [1]) то после имплантации необходимо провести термообработку полупроводника, чтобы отжечь эти дефекты, восстановить кристаллическую структуру и перевести легирующие примеси из нейтральных комплексов в электрически активное состояние. В случае кремния эта термообработка проводится при темпера-

---

[1]) При высокой дозе облучения ионами ($>10^{14}$–10^{15} см$^{-2}$) приповерхностный слой полупроводника может вообще аморфизоваться. Избежать аморфизации поверхности можно проводя «горячее внедрение», то есть проводя имплантацию в нагретый кристалл. Это особенно важно для GaAs, поскольку, в отличие от Si, при кристаллизации возникающего в GaAs аморфного слоя получается материал очень низкого качества. Поэтому ионную имплантацию в GaAs и InP проводят при температурах 150–250 °C.

турах 500–1000 °С, причем чем выше доза внедренных ионов, тем более длительная и высокотемпературная термообработка требуется для отжига дефектов. В кристаллах, в которых приповерхностный слой аморфизовался, отжиг при ~600 °С приводит сначала к образованию дислокационной структуры, которая отжигается только при ~1000 °С. Заметим, что при указанных температурах диффузия в кремнии еще незначительна, и профиль распределения внедренных примесей по глубине практически не меняется.

В 1974–1975 гг. как альтернатива термическому отжигу была предложена технология *лазерного отжига* [152]. При лазерном отжиге имплантированные структуры в течение короткого времени облучаются импульсным или непрерывным светом лазера, который при плотности возбуждения 1–10 Дж/см^2 оказывает действие, эквивалентное термообработке при 1000 °С. Оказывается, что действие интенсивного облучения светом не сводится только к нагреву полупроводника: в образцах наблюдаются эффекты, которые отсутствуют при термическом отжиге [108]. Эти особенности связываются с интенсивной генерацией электронно-дырочных пар светом и, по-видимому, являются следствием известного факта, что темп отжига радиационных дефектов возрастает при рекомбинации неравновесных носителей на этих дефектах. Важным преимуществом лазерного отжига является то, что время обработки пластины обычно не превышает одной минуты, в то время как продолжительность термического отжига измеряется часами.

В качестве примера использования ионной имплантации рассмотрим формирование профиля легирования базы транзистора со «скрытым» эмиттером. Начнем с того, что при создании транзисторов по стандартной планарной технологии (см. с. 203) их эмиттеры создаются последними и располагаются на самом верху структуры, а необходимый для создания в базе тянущего электрического поля профиль легирования базы транзистора (с убыванием концентрации по мере удаления в глубь кристалла, см. рис. 2.5*б*) получается за счет диффузии акцепторной примеси. В то же время для изготовления современных интегральных микросхем И^{2}Л (см. п. 2.8.2) необходимы транзисторы, в которых скрытый (лежащий в глубине структуры) n^+-слой служил бы эмиттером, а область n-типа на поверхности — коллектором. Чтобы такой транзистор имел высокий коэффициент усиления и высокое быстродействие, необходимо сформировать

такой профиль легирования базы, при котором концентрация примеси *возрастает* по мере удаления от поверхности. Такой профиль может быть получен только ионной имплантацией. Распределение легирующих примесей в транзисторе со «скрытым» эмиттером, в котором необходимое распределение акцепторов в базе создано ионной имплантацией атомов бора, показано на рис. 2.22.

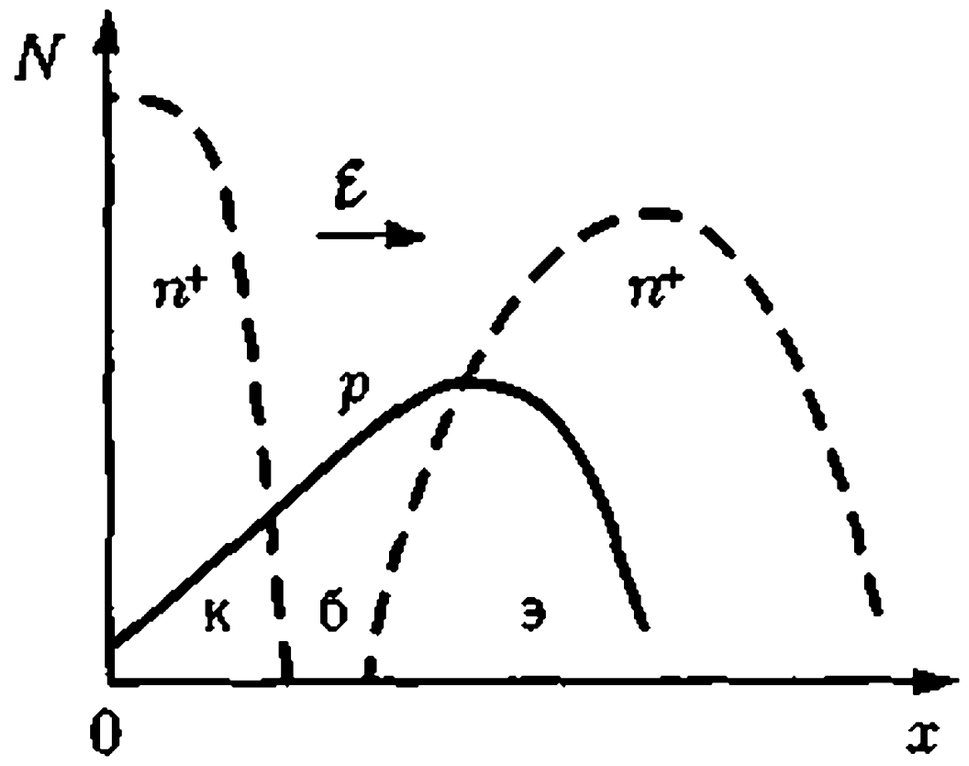


Рис. 2.22. Распределение примесей в транзисторе со «скрытым» эмиттером и ионно-легированной базой. Вектор $\mathcal{E}$ указывает направление встроенного электрического поля, способствующего дрейфу инжектированных электронов от эмиттера к коллектору

В 1987 г. в связи с разработкой методов создания все более мелких p–n-переходов была предложена модификация метода ионной имплантации — метод *ионной имплантации при погружении в плазму* или, как его еще называют, метод *плазменного легирования* [153]. [1]) Этот метод позволил преодолеть недостаток стандартной схемы ионной имплантации — проблему получения большого тока ионов при их малой энергии. Реализовать этот метод можно в обычной камере для плазменной обработки полупроводников, добавив в поток газа-носителя (H_2, He, Ar) небольшое количество газообразных соединений легирующих примесей (B_2H_6, PH_3, AsH_3). При подаче на полупроводниковую подложку коротких импульсов отрицательного напряжения амплитудой 1–10 кВ (длительность 5–20 мкс, частота 0,2–5 кГц) в результате электрического разряда в камере возбуждается плазма, и ускоренные ионы внедряются в незащищенные участки поверхности полупроводника. Профиль получаемого распределения атомов примеси по глубине имеет максимум вблизи поверхности (и, следовательно, похож на профиль, получаемый при диффузии), однако примеси локализованы в очень тонком (200–600 Å) слое, а их концентрация может превышать 10^{20} $см^{-3}$. Различные применения этого метода имплантации рассмотрены в обзорной статье [154].

Ионная имплантация является важным компонентом современной технологии БИС и СБИС, позволяющим создавать неглубокие p–n-переходы, необходимые для современных биполярных и полевых транзисторов субмикронных размеров. С помощью ионной имплантации изготавливаются перспективные для интегральных микросхем подложки со структурой кремний-

---

[1]) Для обозначения этой технологии в зарубежной литературе используют аббревиатуру PIII (plasma immersion ion implantation).

на-изоляторе (см. с. 282). Более подробно об ионной имплантации и лазерном отжиге можно прочитать в работах [58, 108, 143, 144, 152].

Плазменное травление. Недостатком метода химического травления, используемого для вскрытия окон в SiO_2 перед проведением диффузии или ионной имплантации, является заметное боковое подтравливание окисла, которое заметно уменьшает разрешающую способность фотолитографического процесса. Поэтому как альтернатива химическому травлению при производстве БИС начиная с середины 70-х годов стала широко использоваться технология *плазменного травления*, иногда также называемая «сухим» травлением [58, 141, 143].

Электрический разряд в газе низкого давления (10^{-3}–10^{1} мм рт. ст.) приводит к образованию плазмы, состоящей из электронов, ионов, а также заряженных и нейтральных радикалов. Плазма является источником реакционно-способных компонентов, вступающих в химическую реакцию с атомами поверхности твердого тела, в результате которой образуются летучие продукты реакций. Отличительной особенностью плазменного травления, которой нет у химического травления, является его *анизотропия*, то есть различие скоростей вертикального и бокового травления. При определенных режимах плазменного разряда, когда незащищенная поверхность распыляется высокоэнергичными ионами-радикалами, скорость бокового травления может быть уменьшена в ~100 раз по сравнению со скоростью вертикального травления. Это особенно важно для производства современных ИС с субмикронными размерами элементов, поскольку анизотропное травление позволяет сохранить высокое разрешение, достигаемое в фотолитографии. Именно этим способом изготавливаются глубокие «каналы» в поверхности кремния, необходимые для создания накопительных конденсаторов в современных динамических запоминающих устройствах (см. с. 278).

Подбирая состав газовой смеси, можно найти условия для избирательного травления определенного материала. Так, например, для удаления SiO_2 без заметного воздействия на фоторезист и находящийся под слоем SiO_2 кремний используют газовую смесь CF_4+H_2 или CHF_3. В то же время плазменное травление в смеси O_2+CF_4 с высоким содержанием кислорода можно использовать для преимущественного удаления самого резиста [141].

Плазменное травление оказывается эффективным и в тех случаях, когда трудно подобрать подходящий химический травитель.

Так, например, чрезвычайно химически стойкие защитные слои нитрида кремния Si_3N_4, широко используемые при создании ИС, легко травятся в плазменном разряде в смеси CF_4+O_2; для химического стравливания нитрида кремния пластины следовало бы нагревать в H_3PO_4 до температуры 180 °С, которую не выдерживает ни один фоторезист.

Изготовление планарного транзистора. Рассмотрим сначала последовательность операций, используемых для создания дискретного (одиночного) кремниевого транзистора по планарной технологии, а затем обсудим формирование более сложных транзисторных структур, применяемых в современных ИС.



Рис. 2.23. Основные операции при создании кремниевого планарного транзистора

Последовательность основных операций, проводимых при создании планарных n–p–n-транзисторов, показана на рис. 2.23. Исходным материалом для изготовления этих транзисторов является кремний n-типа проводимости или структура, состоящая из сильно легированной подложки n^+-типа, на поверхности которой выращен эпитаксиальный слой n-типа необходимой

толщины. [1]) Пластина, на поверхности которой будут формироваться транзисторы, окисляется по методике, описанной на с. 197 (рис. 2.23 *а*). После этого с помощью фотолитографии в окисле вскрываются окна для формирования области базы (рис. 2.23 *б*). База создается путем диффузии примеси бора. Загонка примеси проводится при ~ 1000 °C из летучего при высоких температурах B_2O_3 или путем добавления паров жидкого BBr_3 в поток азота (аргона) с небольшой примесью кислорода. [2]) Разгонка примеси происходит при ~ 1200 °C в токе азота (аргона) с примесью влажного кислорода. Слой образующегося в процессе разгонки боросиликатного стекла (рис. 2.23 *в*) будет далее использован в качестве маски при создании эмиттера. После вскрытия окон в этой маске с помощью фотолитографии (рис. 2.23 *г*) в эти окна проводят диффузию или имплантацию донорной примеси (P или As) и создают области эмиттера (рис. 2.23 *д*). Наконец, после всех указанных операций с помощью фотолитографии вскрываются окна для контактных площадок к областям базы и эмиттера (рис. 2.23 *е*), напыляются и вжигаются металлические контакты (например, алюминий), пластина кремния разрезается на части и полученные дискретные приборы напаиваются на держатели (рис. 2.23 *ж*).



Рис. 2.24. Поперечное сечение типичного n–p–n-транзистора, используемого в конструкции интегральных схем [156]

Поперечное сечение транзистора, используемого в современных интегральных схемах, показано на рис. 2.24. В отличие от рассмотренного выше дискретного транзистора, конструкция транзисторов для ИС должна быть такова, чтобы все выводы транзистора выходили на поверхность кристалла. Кроме того, как мы отмечали выше, в конструкции высокочастотных и мощных транзисторов создается еще сильно легированный слой коллектора, обозначенный на рис. 2.24 как *скрытый слой*.

---

[1]) Об особенностях и технологии эпитаксиального роста можно прочитать в книгах [143, 155].

[2]) Использование BBr_3 позволяет полностью избавиться от следов воды, неконтролируемо присутствующей в B_2O_3, и тем самым повысить воспроизводимость получения заданной концентрации примеси при диффузии.

Для создания таких транзисторов с целью их дальнейшей интеграции в составе ИС (для чего нам еще понадобится изолировать транзисторы друг от друга) поступают следующим образом [58]. В качестве подложки используется пластина слабо легированного кремния p-типа проводимости ($\rho = 1$–10 Ом·см) толщиной около 300 мкм. На ней с помощью окисления, фотолитографии и последующей имплантации As или Sb в тех местах, где будут располагаться транзисторы, создают слой полупроводника n^+-типа проводимости, который затем «разгоняется» в глубь подложки на глубину 2–8 мкм. Это — тот самый скрытый слой, который в дальнейшем будет служить низкоомным выводом коллектора.

Следующим этапом создания планарного транзистора для ИС является наращивание на поверхности кремниевой пластины высококачественного *эпитаксиального слоя* n-типа, который будет служить коллектором и в котором будут формироваться база и эмиттер транзистора.

Перед началом эпитаксии подложка тщательно очищается и затем помещается в реактор, в котором остатки загрязнений и собственного окисла удаляются с поверхности кремния путем газового травления при ~ 1200 °C в безводном HCl, добавляемом в поток водорода. Затем в поток водорода добавляется $SiCl_4$ и при той же или чуть более низкой температуре методом газовой эпитаксии происходит наращивание эпитаксиального слоя n-типа, толщина и уровень легирования которого выбираются исходя из требуемого максимального рабочего напряжения транзистора (см. п. 2.3). Легирование эпитаксиального слоя осуществляется добавлением в поток водорода газообразных AsH_3 или PH_3. Характерная толщина наращиваемого слоя составляет 1–5 мкм, а его удельное сопротивление — 1 Ом · см ($n \sim 4 \times 10^{15}$ см$^{-3}$).

После этого путем термического окисления на поверхности кремния создают тонкий (~500 Å) слой окисла, на который осаждают пленку Si_3N_4 толщиной около 1000 Å. Выделив с помощью фотолитографии участки на поверхности образца, на которых будет создаваться окисная изоляция между элементами, методом плазменного травления удаляют слой нитрида кремния и проводят окисление кремния на всю глубину эпитаксиального слоя (нитрид кремния практически не окисляется и защищает закрытые им участки поверхности от окисления).

После удаления слоя нитрида в эпитаксиальном слое создают базовую и эмиттерную области аналогично тому, как это делалось выше для дискретного транзистора. Для повышения

быстродействия транзисторов область базы, прилегающая к контакту, легируется заметно сильнее, чем область, располагающаяся непосредственно под эмиттером (см. рис. 2.24). Контакт к коллектору (область n^+-типа) может быть создан одновременно с созданием n^+-области эмиттера (как показано на рис. 2.24), однако в случае, когда сопротивление толщи коллектора должно быть минимальным, перед созданием областей базы и эмиттера область под контактом к коллектору с помощью имплантации ионов фосфора сильно легируют на всю глубину вплоть до скрытого слоя (пунктир на рис. 2.24). Перед созданием контактов на поверхность прибора осаждают тонкий слой Si_3N_4, который защищает структуру от проникновения в нее ионов Na^+ и молекул воды. [1)]

Оказывается, что тип примеси, используемой для создания эмиттера, может существенно влиять на характеристики транзисторов. Хотя примесь фосфора широко применяется для создания эмиттерных n-областей в транзисторах, выяснилось, что при ее использовании для создания неглубоких ($< 0,3$ мкм) p–n-переходов не удается получить высокий градиент концентрации примеси в области эмиттерного перехода, который необходим для высокого коэффициента инжекции эмиттера в области микротоков. Кроме того, при легировании фосфором происходит неконтролируемое оттеснение металлургической границы коллекторного p–n-перехода, лежащего под эмиттером (так называемый «deep-эффект» [141]), что мешает создать очень тонкую базу. Последнее связано с особенностью диффузии P в Si: фосфор начинает диффундировать с поверхности вглубь кристалла в виде комплексов с вакансиями, а когда комплексы распадаются, в объеме возникает высокая концентрация вакансий, которая и стимулирует более быструю диффузию находящейся в базе примеси бора. Напротив, при использовании вместо фосфора мышьяка градиент концентрации примеси удается сделать в $\sim$4 раза выше, что обеспечивает более чем двукратное увеличение коэффициента усиления, а отсутствие deep-эффекта позволяет примерно вдвое увеличить и частоту отсечки транзистора.

При создании транзисторов для ИС большую роль играет совмещение в создаваемой структуре скрытого слоя, базы и эмиттера, которое осуществляется при проведении фотолитографии. Заметим, что в связи с неуклонным уменьшением размеров отдельных элементов (как мы уже отмечали, в настоящее время их размер составляет $\sim$0,1 мкм) проблемы совмещения выходят

[1)] Ионы натрия имеют достаточно высокий коэффициент диффузии в SiO_2 и могут вызывать появление токов утечки p–n-переходов и инверсию типа проводимости поверхностного слоя. Молекулы воды могут стимулировать электрохимическую коррозию контактов и приводить к их деградации.

на первый план, в связи с чем в технологии большое внимание уделяется использованию приемов *самосовмещения*. Основной идеей этих приемов является использование ранее созданных элементов структуры в качестве масок при формировании последующих элементов (см. подробнее на с. 273).

Подробнее об особенностях технологии изготовления различных типов ИС можно прочитать в [58].

Изоляция элементов в ИС. В интегральной схеме, содержащей обычно от нескольких тысяч до нескольких миллионов элементов, необходимо принять меры для уменьшения взаимного влияния элементов друг на друга путем их изоляции. Для изоляции элементов в ИС было разработано много способов, два из которых получили наибольшее распространение [142].

В первом способе, называемом *изоляцией с помощью p–n-перехода*, после наращивания эпитаксиального слоя n-типа в него через окна в окисле диффузией вводят акцепторную примесь (обычно бор), которая разгоняется на глубину до подложки p-типа (см. рис. 2.25 *а*). В этом случае, если подать на подложку отрицательное напряжение, коллекторы отдельных транзисторов оказываются изолированными друг от друга обратно смещенными p–n-переходами. К сожалению, этот способ изоляции (также называемый *разделительной диффузией*) характеризуется большой паразитной емкостью коллекторов и большими непроизводительными потерями площади кристалла, идущей на создание изолирующих областей (из-за боковой диффузии толщина изолирующей области в простых ИС с эпитаксиальным слоем толщиной 10 мкм при ширине окна 10 мкм достигает 35–40 мкм). Поэтому этот способ изоляции использовался только в первых ИС с невысоким уровнем интеграции (примерно до 1970 года).

В настоящее время при создании ИС используется комбинированный способ изоляции, в котором сочетаются боковая *окисная изоляция* и изоляция p–n-переходом, образуемым скрытым слоем и подложкой. Окисную изоляцию можно создать двумя способами. В кристаллах с тонким (порядка 1 мкм) эпитаксиальным слоем перед созданием областей транзистора в защитном слое Si_3N_4 [1]) в необходимых местах вскрываются окна, в них

---

[1]) Слои стехиометрического Si_3N_4 обладают высокой устойчивостью по отношению к окислению и могут быть получены на поверхности кремния химическим осаждением из смеси SiH_4+NH_3 при 700–900 °С. Защитные пленки нестехиометрического нитрида кремния получаются плазмохимическим осаждением (в тлеющем разряде) из той же смеси при 200–350 °С [58].

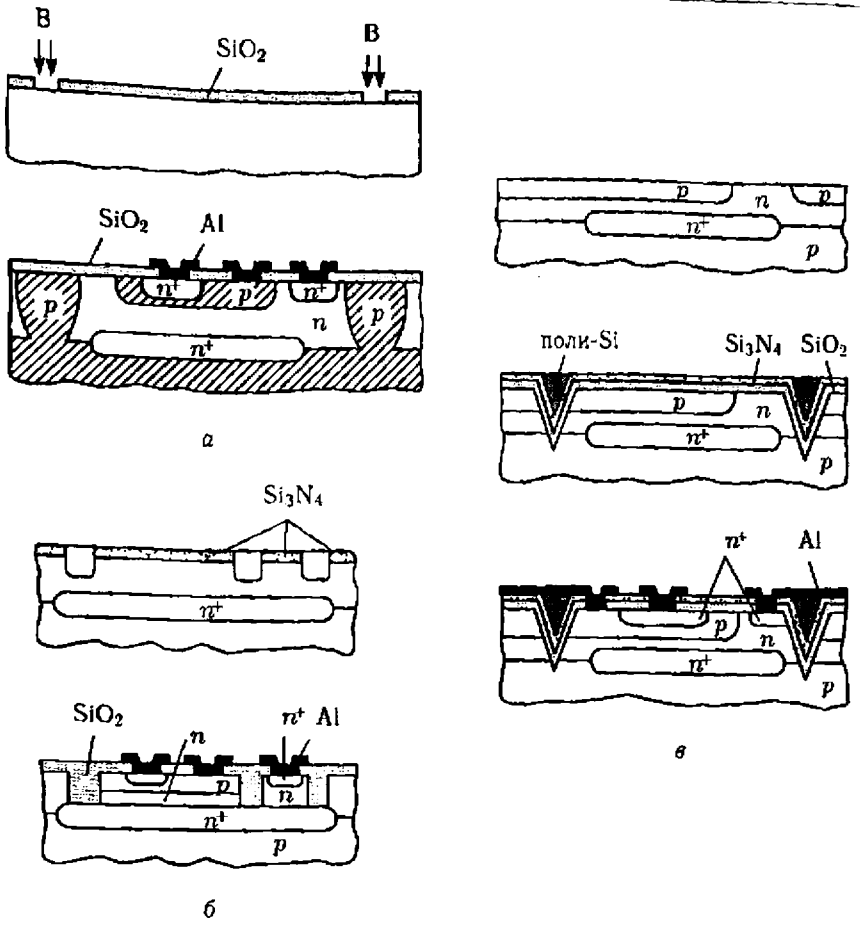


Рис. 2.25. Способы изоляции элементов в ИС: *а* — изоляция p–n-переходом, *б* — окисная изоляция (изопланарный процесс), *в* — окисная изоляция с использованием V-образных канавок

травлением удаляется часть кремния (см. рис. 2.25 *б*), после чего оставшаяся часть эпитаксиального слоя термически окисляется на глубину до скрытого слоя. Удалить травлением часть кремния необходимо для того, чтобы после окисления поверхность структуры вновь стала планарной (напомним, что при окислении объем образующегося окисла в 2,2 раза превышает объем вступившего в химическую реакцию кремния). После этого в структуре можно создавать базу и эмиттер. Описанный способ создания окисной изоляции получил название *изопланарного процесса*.

В кристаллах с толстым эпитаксиальным слоем после создания диффузионных областей базы в них вытравливаются

V-образные канавки (см. рис. 2.25*в*), которые покрываются слоями окисла, нитрида [1]) и затем заполняются каким-либо материалом (обычно поликристаллическим кремнием) для восстановления планарности. Для создания таких канавок используют травление кремния в *селективном травителе* (горячем водном растворе KOH, $(CH_3)_4NOH$ или гидразина) под маской из Si_3N_4. В таком травителе скорость травления поверхности (100) кремния в несколько десятков раз выше, чем поверхности (111), и поэтому на поверхности кремния, ориентированной в направлении (100), образуются V-образные канавки, идущие вдоль направлений <110>. В этих канавках угол между «ограненной» плоскостью стенки (111) и вертикальной плоскостью равен 35,3°.

Структуры, в которых используется окисная изоляция, характеризуются малой паразитной емкостью, они имеют в 3–8 раз бóлее высокую плотность упаковки по сравнению со структурами, созданными разделительной диффузией, и поэтому идеально подходят для создания современных БИС и СБИС. Еще более высокую плотность упаковки можно получить, изолируя элементы с помощью U-образных канавок, получаемых методом анизотропного плазменного травления [141].

Полностью исключить взаимное влияние элементов друг на друга можно только в структурах, в которых тонкий эпитаксиальный слой полупроводника выращен на изолирующей подложке [142]. Наиболее важными среди таких структур являются кремний-на-сапфире и кремний-на-изоляторе (см. с. 282). Микросхемы, созданные в таких структурах, обладают лучшими характеристиками (меньшей паразитной емкостью, более высокой радиационной стойкостью), но в то же время имеют и заметно более высокую стоимость.

2.8.2. Особенности устройства цифровых ИС на биполярных транзисторах. В схемотехнике современных цифровых интегральных схем на биполярных транзисторах широко используются *многоэмиттерные* и *многоколлекторные транзисторы* (см. рис. 2.26 и 2.27). Применение таких транзисторов существенно упрощает реализацию логических функций в этих ИС. Например, с помощью многоэмиттерного транзистора в микросхемах транзисторно-транзисторной логики (ТТЛ, см. рис. 2.26)

---

[1]) Из-за сильных напряжений, создаваемых пленкой Si_3N_4 в расположенном под ней кремнии, в последнем могут возникать дислокации. Чтобы избежать этого, между нитридом кремния и кремнием помещают тонкий (порядка 500 Å) подслой SiO_2.

реализуется функция «И-НЕ». Действительно, низкий уровень напряжения на выходе логического элемента (логический 0) получается только тогда, когда на все эмиттеры многоэмиттерного транзистора VT1 (входы логического элемента) подан высокий уровень напряжения (логическая 1) и соответствующие эмиттерные p–n-переходы закрыты. При подаче на любой из входов элемента напряжения низкого уровня транзистор VT1 входит в режим насыщения, выходной транзистор VT2 закрывается и на выходе логического элемента устанавливается высокий уровень напряжения. Использование многоэмиттерного транзистора для реализации логической функции одновременно решает и проблему электрической развязки входов логического элемента, поскольку в электронной схеме эти входы обычно подключены к выходам разных логических элементов, состояние которых должно оставаться независимым.

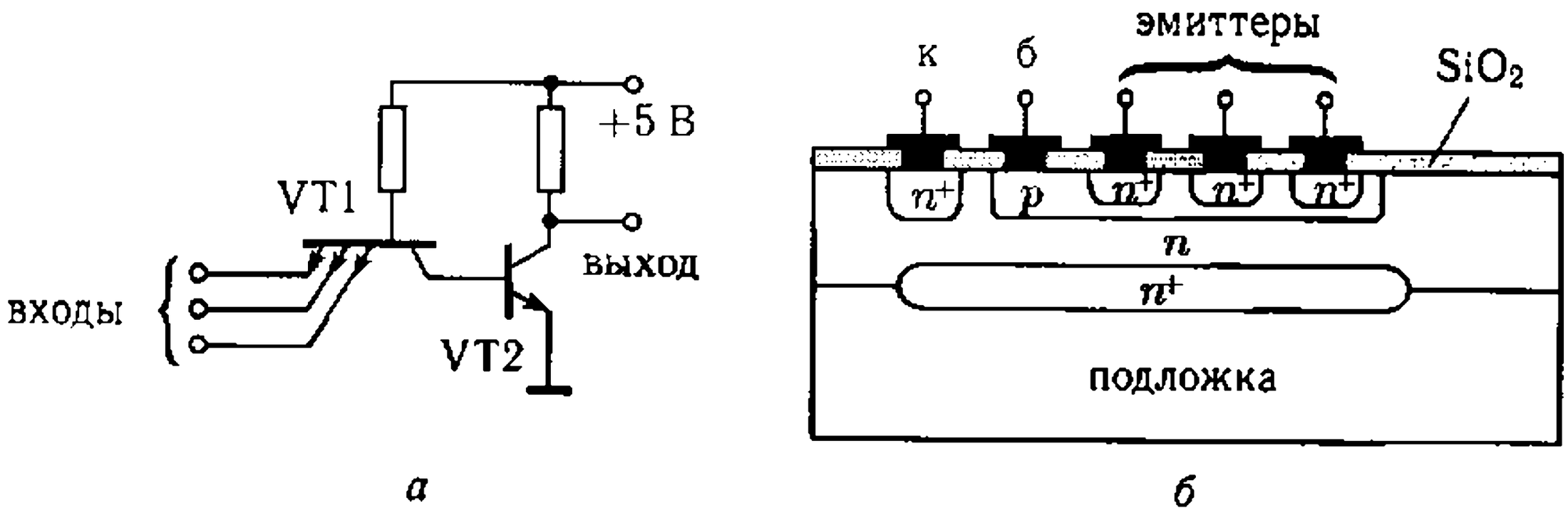


Рис. 2.26. Упрощенная принципиальная схема базовой ячейки ИС транзисторно-транзисторной логики (*а*) и топология используемого в ней многоэмиттерного транзистора (*б*)

В микросхемах интегральной инжекционной логики ($И^2Л$, см. рис. 2.27) выходное состояние инвертора, кодируемое разомкнутым или замкнутым состоянием транзисторного ключа, одновременно выдается на несколько электрически независимых коллекторов многоколлекторного транзистора, а необходимая логическая функция реализуется путем соединения выходов нескольких инверторов (такой способ реализации логической функции в схемотехнике называется «проводно́е ИЛИ»).

Микросхемы, построенные по принципу *интегральной инжекционной логики* ($И^2Л$) нашли широкое применение в современной микроэлектронике [157]. Это связано с тем, что конструкция базовой ячейки этих ИС позволяет достигнуть очень высокой плотности интеграции при низкой потребляемой мощности. При изготовлении этих интегральных схем используется

стандартный набор операций планарной технологии (см. п. 2.8.1), однако общее число этих операций заметно меньше, чем при изготовлении других типов биполярных ИС. Все это определяет низкую стоимость микросхем $И^2Л$ и простоту их интеграции с другими типами биполярных ИС.

Рассмотрим логический элемент $И^2Л$ более подробно. В конструкцию его базовой ячейки заложены две принципиально новые идеи: *новый способ подачи смещения в базу* усилительного транзистора, позволяющий вообще отказаться от использования резисторов (которые обычно занимают значительную площадь на поверхности кристалла), и *функциональная интеграция* рабочих областей транзисторов разного типа проводимости, при которой отпадает необходимость изолировать транзисторы друг от друга и создавать отдельные контакты к каждому их выводу (контакты также требуют дополнительного места на поверхности кристалла). Все это позволило увеличить плотность упаковки в микросхемах $И^2Л$ (число элементов, размещаемых на единице площади кристалла) в 10–30 раз по сравнению с другими типами биполярных ИС.

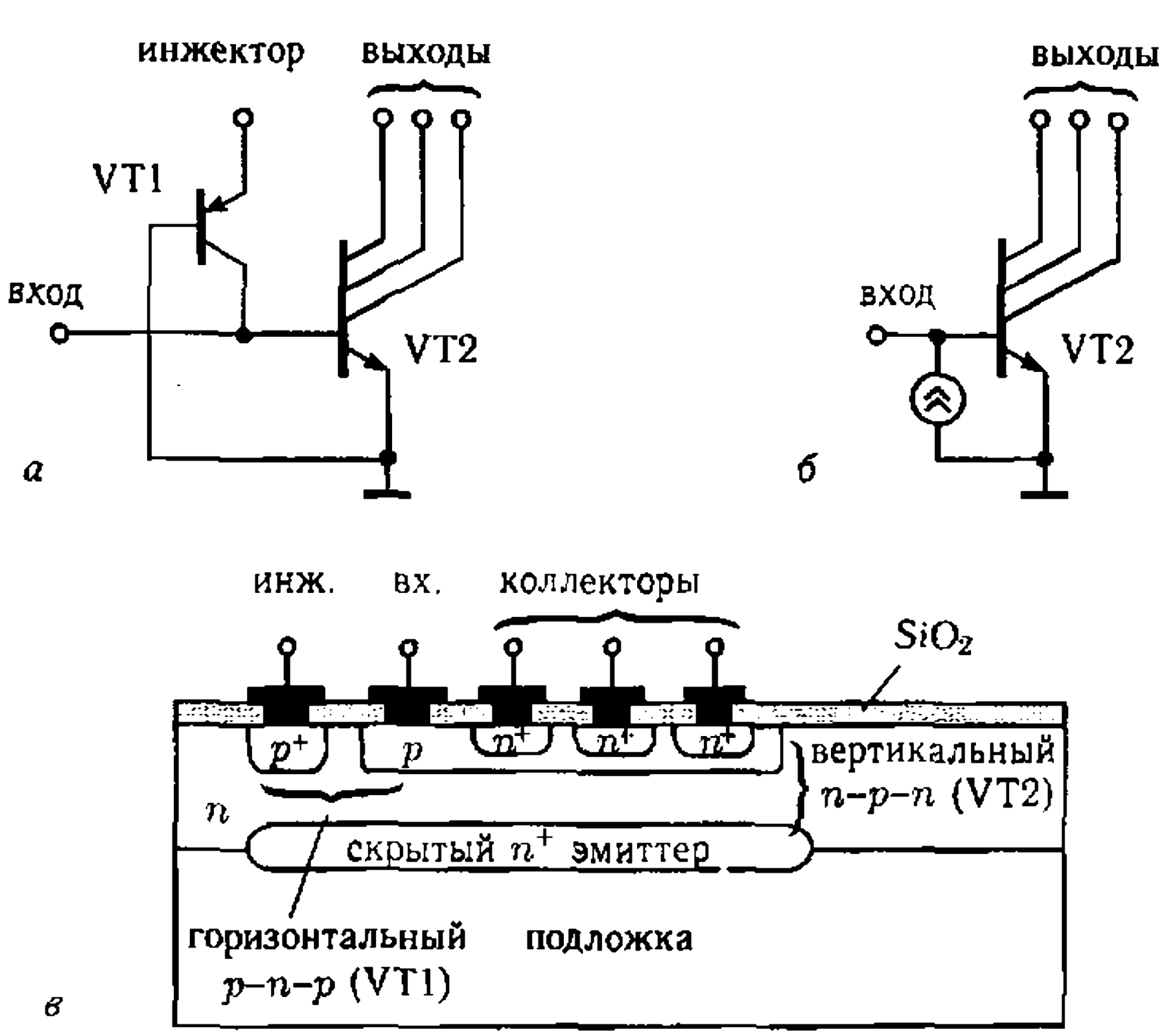


Рис. 2.27. Принципиальная схема базовой ячейки (*а*), ее эквивалентная схема (*б*) и топология ячейки интегральной схемы $И^2Л$ (*в*)

Базовый элемент $И^2Л$ (см. рис. 2.27 *в*) представляет собой комбинацию горизонтального p–n–p-транзистора (так называемого *инжектора*) VT1 и вертикального n–p–n- (усилительного) транзистора VT2, который обычно делается многоколлекторным для реализации необходимых логических функций с помощью «проводно́го ИЛИ». Усилительный транзистор имеет «скрытый» эмиттер, сильно легированный коллектор и базу, профиль легирования которой (для создания тянущего электрического поля) создают ионной имплантацией (см. рис. 2.22 и его обсуждение в тексте). При подаче на эмиттерный переход p–n–p-транзистора прямого смещения инжектируемые из p^+-области дырки попадают в p-область, которая *одновременно* является коллектором транзистора VT1 и базой транзистора VT2. Эти дырки создают в базе n–p–n-транзистора заряд, вызывающий открывание этого транзистора, если только вывод базы транзистора не соединен с общим проводом через выход какого-либо другого логического элемента. Поскольку p–n–p-транзистор включен по схеме с ОБ и имеет высокое выходное сопротивление, его можно рассматривать как генератор тока, что позволяет изобразить эквивалентную схему логического элемента в виде, показанном на рис. 2.27 *б*. Если вход такого элемента замкнуть на общий провод (например, через открытый транзистор другого логического элемента), то транзистор VT2 закроется и на его выходах (к которым подключаются входы других логических элементов) установится состояние логической 1. При разомкнутом входе ток, втекающий в базу n–p–n-транзистора из инжектора, поддерживает транзистор в открытом состоянии и потенциал его коллекторов близок к нулю (состояние логического 0).

Возможности функциональной интеграции элементов в микросхемах $И^2Л$ иллюстрирует пример построения ячейки статического запоминающего устройства, показанный на рис. 2.28. Такие запоминающие устройства широко используются в ЭВМ. Для подачи смещения на все усилительные транзисторы в ячейке используется один общий инжектор — область p_3, которая создана диффузией в эпитаксиальный слой n-типа. Диффузионная область p_1 одновременно служит коллектором горизонтального транзистора VT3 и базой вертикального транзистора VT2, а область p_2 — коллектором VT4 и базой VT1. Эмиттерами транзисторов VT1 и VT2 является скрытая n^+-область, расположенная под эпитаксиальным слоем; контакт к этому слою образует горизонтальная полоска, обозначенная n^+, которая подключается к адресной шине АШ. Перекрестные связи между коллекторами и базами транзисторов VT1 и VT2 осуществляют две

горизонтальные заштрихованные перемычки. Вторые коллекторы транзисторов, расположенные в центрах областей p_1 и p_2 и помеченные n^+, являются выходами ячейки и подключаются к разрядным шинам РШ. Плотность упаковки элементов в запоминающем устройстве на самом деле еще выше, поскольку один инжектор (область p_3) используется одновременно для подачи смещения в две соседние ячейки памяти (вторая из них, геометрически расположенная левее инжектора, на рис. 2.28 не показана).

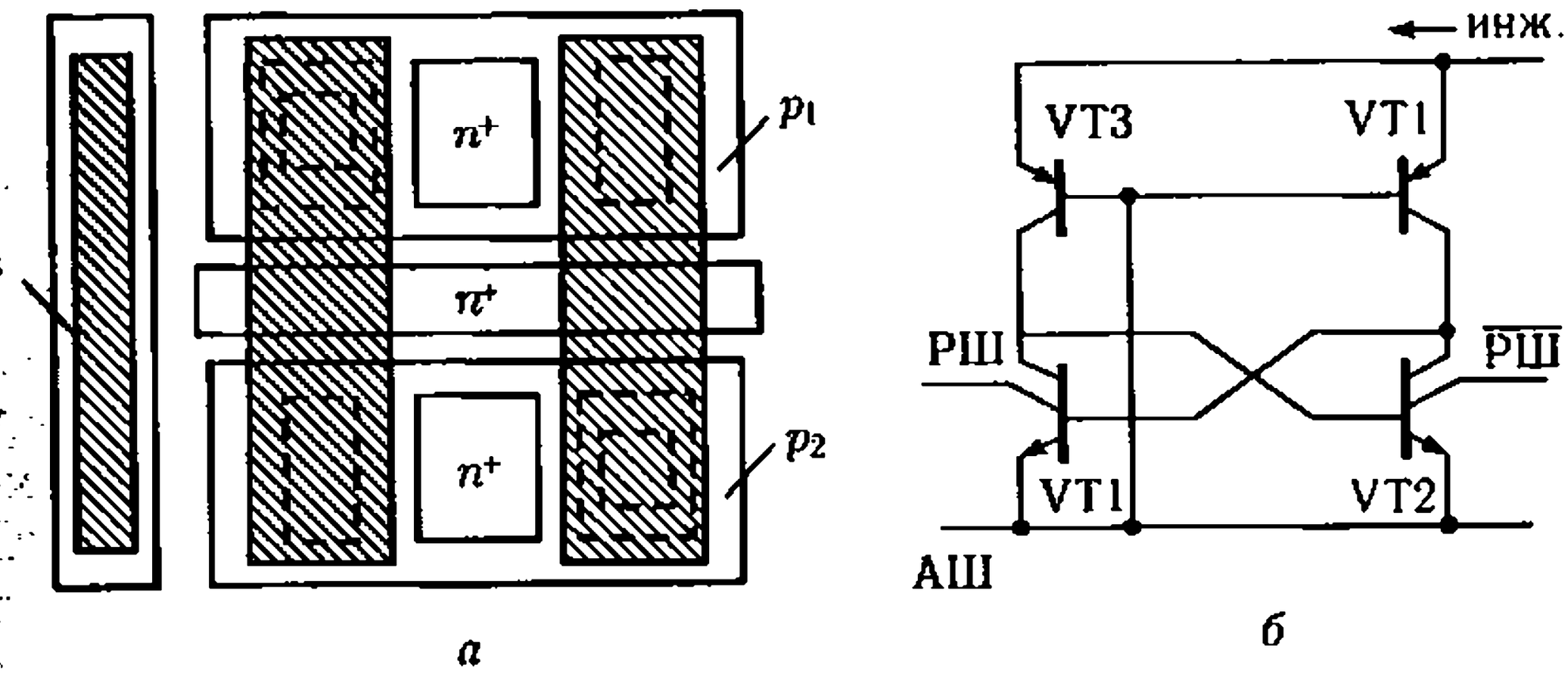


Рис. 2.28. Топология (*а*) и электрическая схема (*б*) ячейки запоминающего устройства на основе $И^2Л$-вентилей [157]

Дальнейшее повышение плотности упаковки и быстродействия микросхем $И^2Л$ связано с двумя изменениями в их конструкции [157]: 1) использованием подложки p-типа в качестве инжектора (такие структуры, называемые SFL (substrate fed logic), позволяют увеличить плотность упаковки в 2–3 раза за счет уменьшения числа межэлементных соединений на поверхности кристалла) и 2) применением диодов Шоттки в качестве развязывающих и шунтирующих элементов. Развязка выходов логического элемента с помощью диодов Шоттки позволяет исключить необходимость создания многоколлекторных транзисторов, а уменьшение напряжения логического перепада до 150–350 мВ позволяет в 2–3 раза уменьшить динамическую мощность рассеяния (см. п. 4.2) и соответственно увеличить быстродействие. Использование фиксирующих диодов Шоттки (см. рис. 2.20) позволяет избежать насыщения усилительных транзисторов и тем самым еще более повышает быстродействие микросхем.

Логические элементы $И^2Л$ нашли широкое применение при создании цифровых ИС на биполярных транзисторах. В качестве примера можно привести выпускаемые отечественной промышленностью микромощные статические запоминающие устройства серии К541 и микропроцессорные комплекты серий К583, К584.

Элементы $И^2Л$ обладают достаточно высоким быстродействием; в современных элементах задержка распространения может достигать 320 пс/вентиль [141], а произведение времени задержки на среднюю мощность потребления (которое иногда называют энергией переключения) [1] — 0,02 пДж [156].

В случае, когда требуется особенно высокое быстродействие, используются логические ИС эмиттерно-связанной логики (ЭСЛ), базовая ячейка которых представляет собой управляемый переключатель тока [136]. Работа транзисторов в этих микросхемах в режиме, исключающем насыщение, и небольшой логический перепад напряжений (около 0,6 В) позволяют строить из таких ячеек наиболее быстродействующие ИС на биполярных транзисторах. Так, в 1999 г. задержка распространения сигнала в микросхемах ЭСЛ, в которых использовались гетеропереходные (HBT) транзисторы $Si/Si_{1-x}Ge_x$ с размерами эмиттера $2\times$ $\times 0{,}2$ мкм, достигла 6,7 пс [158]. К сожалению, являясь наиболее быстродействующими среди биполярных логических ИС, микросхемы ЭСЛ характеризуются и наиболее высокой потребляемой мощностью. В настоящее время производство микросхем ЭСЛ практически свернуто, поскольку более высокие параметры удается получить на кремниевых ИС, изготовленных по BiCMOS-технологии (см. с. 283), и на арсенид-галлиевых полевых транзисторах с барьером Шоттки (см. п. 4.3).

---

[1] Величина этого произведения может служить оценкой энергии, необходимой для выполнения одной логической операции, и широко используется для сравнения различных типов микросхем. Например, типичное значение этого произведения для стандартных микросхем ТТЛ серии SN74 (К155) составляет 100 пДж, для маломощных микросхем ТТЛШ серии SN74ALS (К1533) — 8 пДж, микросхем ЭСЛ серии MC10000 (К500) — 50 пДж. Поэтому сразу становятся понятными преимущества микросхем $И^2Л$, для которых значение энергии переключения не превышает 1 пДж [131].

Глава 3

ТИРИСТОРЫ И ДРУГИЕ МНОГОСЛОЙНЫЕ СТРУКТУРЫ

3.1. Тиристоры

Тиристор представляет собой более сложный по сравнению с биполярным транзистором прибор, основой которого является четырехслойная p–n–p–n-структура. Из-за сильного взаимодействия входящих в эту структуру p–n–p- и n–p–n-транзисторов, тиристоры приобретают новое свойство: они обладают *бистабильными* вольт-амперными характеристиками и могут переключаться из одного состояния в другое. В закрытом состоянии тиристор выдерживает подачу на него достаточно высокого напряжения (от 100 В до 10 кВ), практически не пропуская при этом ток, а перейдя в открытое состояние, может пропускать токи от единиц до тысяч ампер при низком падении напряжения (1–2 В).

На особенности усилительных свойств четырехслойной p–n–p–n-структуры первым обратил внимание Шокли [110], обсуждая причины необычно высокого коэффициента усиления по току точечных транзисторов. В 1955 году Молл с соавт. (Bell Laboratories) заложили основы теории и создали первые четырехслойные структуры с бистабильными характеристиками [159]. Исследования этих авторов позволили увидеть большие перспективы применения тиристоров в силовой электронике. Являясь твердотельными аналогами газоразрядного прибора тиратрона и отличаясь от последнего существенно более низким падением напряжения в открытом состоянии, тиристоры быстро вытеснили тиратроны из сильноточной электроники. Тиристоры оказались идеальными приборами для коммутации больших токов, создания управляемых выпрямителей и преобразователей тока. Они широко

используются в гибких системах передачи электроэнергии как компенсаторы реактивной мощности; в качестве инверторов (преобразователей постоянного тока в переменный) в высоковольтных линиях электропередачи постоянного тока; в системах электропривода (железнодорожный транспорт и др.); в металлургии; для формирования импульсов высокого напряжения в физике высоких энергий. Выпускаемые в настоящее время тиристоры имеют диаметр до 100–150 мм (то есть один прибор может занимать целую пластину кремния!). В общем, тиристоры — это самые крупные и мощные полупроводниковые приборы.

По возможности управления характеристиками описываемых четырехслойных структур их обычно подразделяют на *тиристоры* (управляемые структуры, имеющие третий вывод — управляющий электрод)[1]) и *динисторы* (неуправляемые структуры с двумя выводами). Последние приборы не очень распространены и используются в схемах релаксационных генераторов и формирователей импульсов.

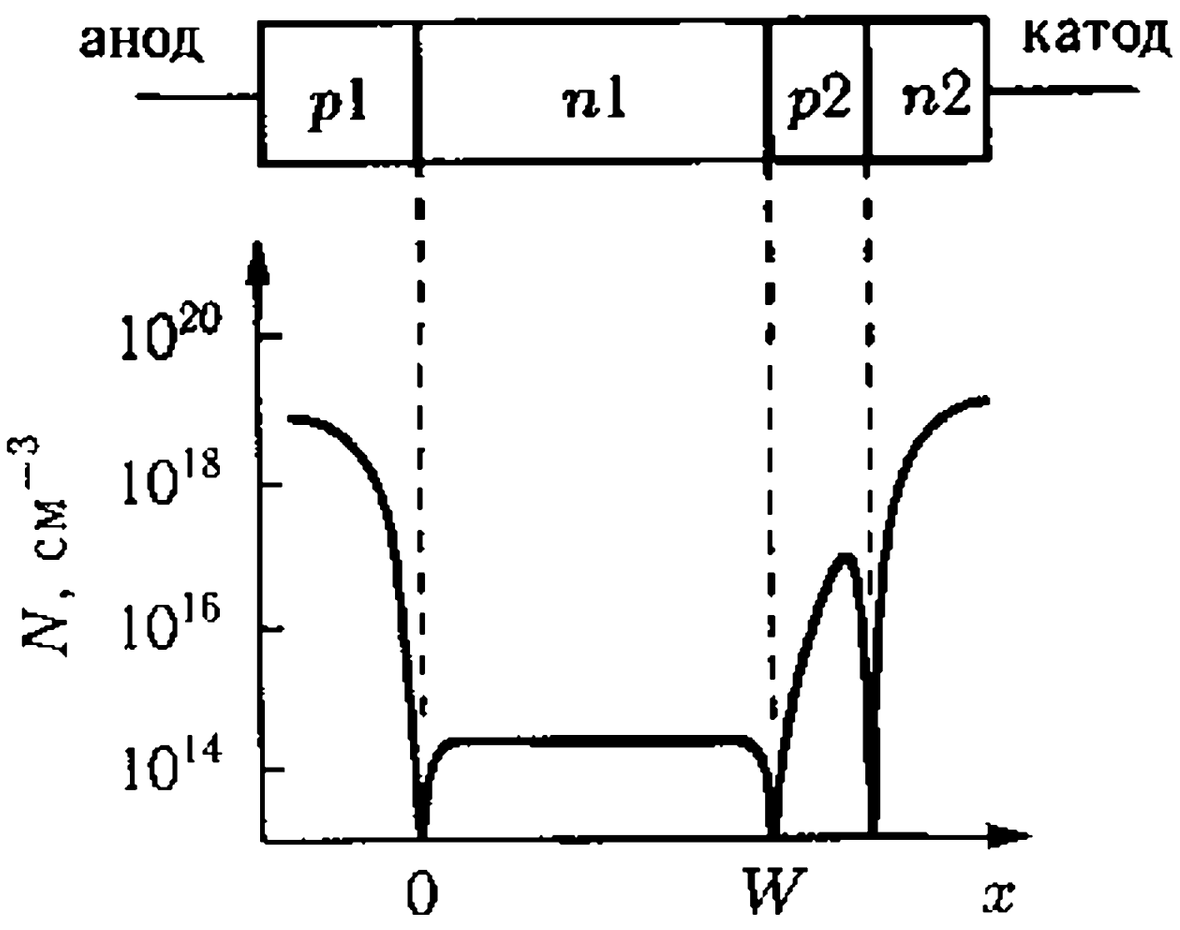


Рис. 3.1. Структура и типичный профиль легирования тиристора

Типичный профиль легирования стандартного диффузионно-сплавного тиристора показан на рис. 3.1. Эту структуру получают следующим образом. Сначала с помощью диффузии акцепторной примеси (B, Al) на противоположных сторонах пластины из слабо легированного n-Si толщиной 0,3–1 мм (область $n1$ на рисунке) создают области $p1$ и $p2$; толщина пластины и ее уровень легирования выбираются исходя из требуемого максимального рабочего напряжения тиристора (об этом мы будем говорить ниже). После этого на одной из только что созданных поверхностей p-типа путем вплавления сплава Au+Sb создают катодную область n-типа (область $n2$ на рисунке). В итоге получается четырехслойная структура с тремя p–n-переходами. Область, обозначенную на рисунке $p1$, называют *анодом* тиристора, а область $n2$ — его *катодом*. *Управляющий электрод* тиристора подключается к области $p2$.

---

[1]) В отечественной литературе такие трехэлектродные приборы также называют *тринисторами* или *кремниевыми управляемыми вентилями*.

3.1.1. Вольт-амперные характеристики тиристора. Вольт-амперная характеристика динистора (и, соответственно, тиристора при нулевом токе через управляющий электрод) показана на рис. 3.2.

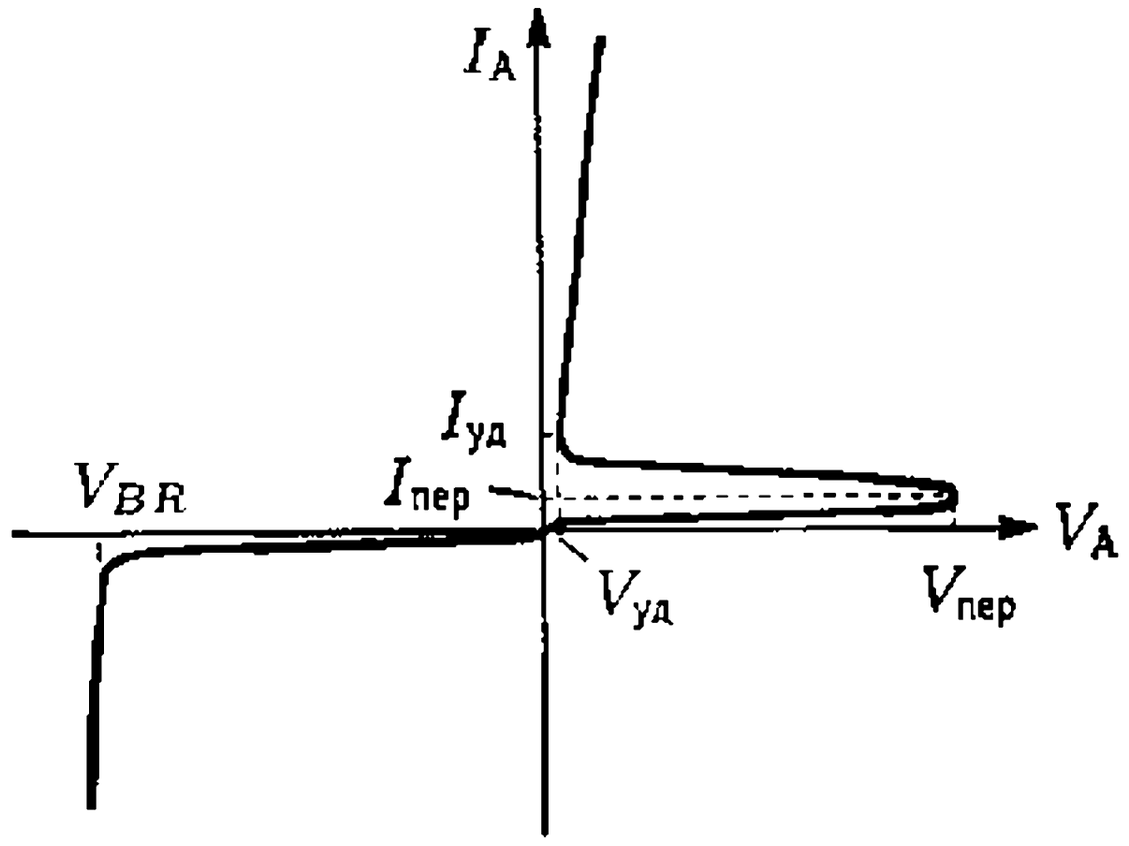


Рис. 3.2. Вольт-амперная характеристика тиристора при нулевом токе через управляющий электрод

При подаче на тиристор обратного напряжения (минус к аноду) переходы $p1$–$n1$ и $p2$–$n2$ смещаются в обратном направлении, а переход $n1$–$p2$ — в прямом. При этом бóльшая часть приложенного напряжения падает на области пространственного заряда перехода $p1$–$n1$, которая располагается в основном в слабо легированной ($n1$) области структуры. Поэтому наиболее важными параметрами, которые определяют максимальное обратное напряжение на тиристоре, оказываются уровень легирования и толщина области $n1$.

Максимальное обратное напряжение на тиристоре ограничивается двумя величинами: напряжением возникновения лавинного пробоя перехода $p1$–$n1$ и напряжением возникновения прокола структуры (когда область пространственного заряда перехода $p1$–$n1$ достигает границы слоя $p2$). Как мы показали в п. 2.3, напряжение пробоя толстого p–n-перехода изменяется примерно обратно пропорционально концентрации примеси в базе (формула (1.67)), а напряжение прокола — прямо пропорционально ей ($V_{\text{прок}} \approx 2\pi q N_d W^2/\varepsilon$). Поэтому для каждого значения толщины базы W тиристора существует некий оптимальный уровень легирования, при котором достигается максимальное напряжение пробоя p–n-перехода (см. рис. 3.3). Этот уровень легирования можно оценить, считая, что лавинный пробой и прокол структуры возникают при одном и том же обратном смещении.

На самом деле при расчете условий возникновения лавинного пробоя необходимо учитывать, что рассматриваемой структуре присуще *внутреннее усиление*, поскольку при высоких обратных напряжениях транзистор, образованный слоями $p1$, $n1$ и $p2$, ведет себя совершенно аналогично транзистору с оборванной базой

(см. п. 2.3). А именно, его напряжение пробоя равно

$$V_{BR} = V_{\text{проб}}(1 - \alpha_1)^{1/m},$$

где $V_{\text{проб}}$ — напряжение пробоя p–n-перехода, $m \approx 4$, а α_1 — коэффициент усиления транзистора. Поскольку толщина электронейтральной области базы W' в рассматриваемом транзисторе довольно велика (при низких обратных напряжениях она сравнима или может даже превышать диффузионную длину), то этот транзистор имеет невысокий коэффициент усиления ($\alpha_1 \sim 0,5$), который в основном определяется эффективностью переноса носителей через базу транзистора в соответствии с формулой (2.8).

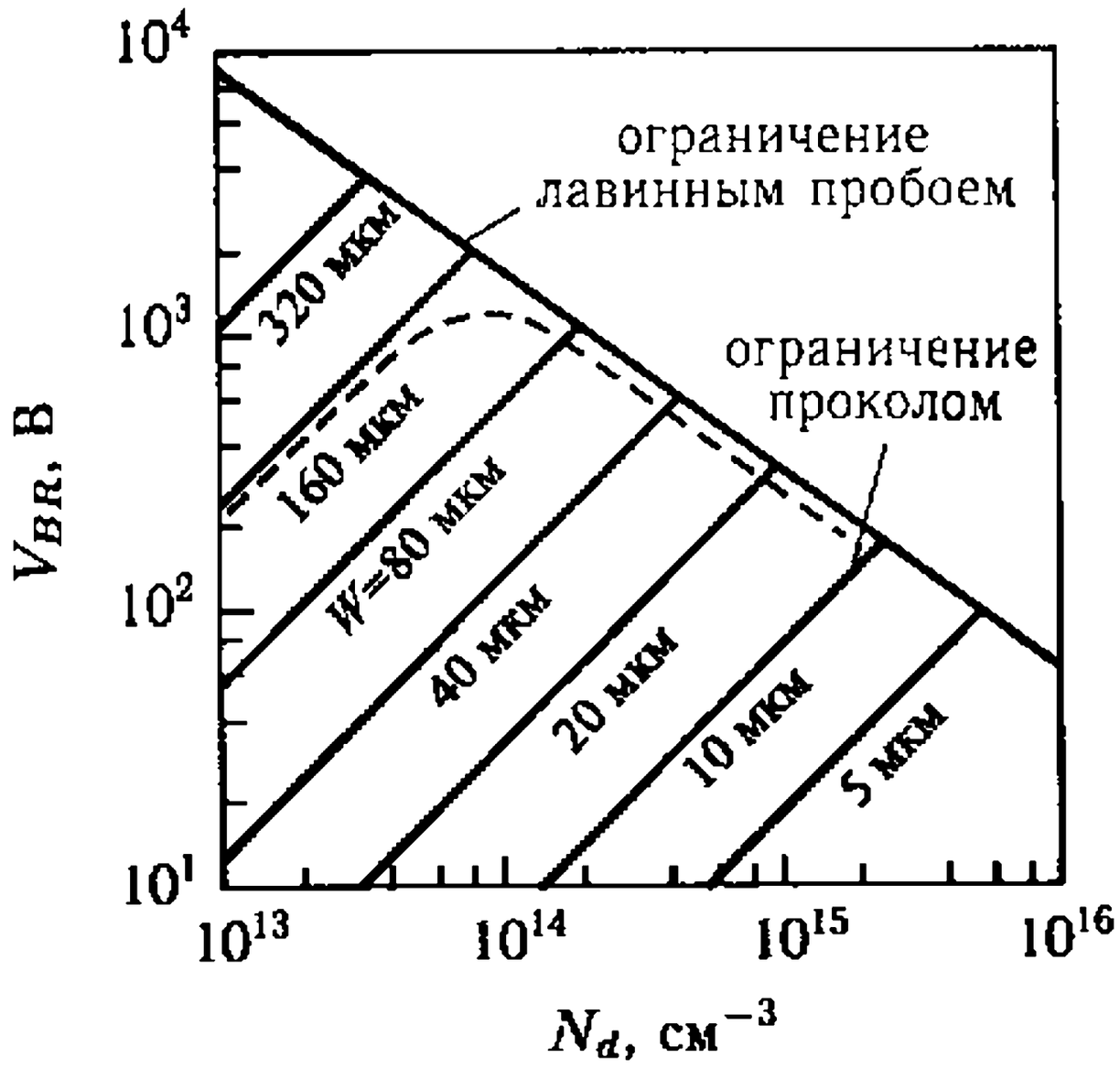


Рис. 3.3. Зависимость максимального обратного напряжения в кремниевых тиристорах при 300 К от уровня легирования и толщины базы прибора (пунктир). Сплошными линиями показаны ограничения, накладываемые лавинным пробоем толстого p–n-перехода и проколом структуры [14]

Рассмотрим теперь, что формирует вольт-амперную характеристику тиристора при подаче на него прямого смещения (плюс к аноду). При такой полярности напряжения переходы $p1$–$n1$ и $p2$–$n2$ структуры смещены в прямом направлении, а переход $n1$–$p2$ — в обратном. При невысоких приложенных напряжениях ток, протекающий через тиристор, определяется обратным током перехода $n1$–$p2$.

Найдем условие включения тиристорной структуры (*напряжение переключения* $V_{\text{пер}}$ на рис. 3.2). Формально это условие можно рассматривать как условие обращения в бесконечность производной тока анода тиристора по напряжению на аноде (потенциал катода будем считать равным нулю). Эквивалентную схему тиристора можно представить в виде двух включенных навстречу друг другу p–n–p- и n–p–n-транзисторов (см. рис. 3.4). В этой эквивалентной схеме для n–p–n-транзистора можно записать:

$$I_{\text{к}2} = \alpha_N I_{\text{К}} + I_{\text{к}02}, \qquad (3.1)$$

где $I_{\text{К}}$ — ток катода тиристора (который служит эмиттером

n–p–n-транзистора), $I_{к02}$ — обратный ток коллектора этого транзистора, α_N — его *статический* коэффициент усиления. При анализе этого уравнения необходимо учесть, что величина α_N сама зависит от протекающего через транзистор тока (см. п. 2.2.3), и поэтому приращение тока коллектора равно

$$dI_{к2} = \left(\alpha_N + I_К \frac{d\alpha_N}{dI_К}\right) dI_К + dI_{к02}. \qquad (3.2)$$

Входящая в это уравнение величина $\alpha_2 = \alpha_N + I_К(d\alpha_N/dI_К)$ представляет собой коэффициент усиления транзистора, измеренный *на малом сигнале* (в том смысле, как он был определен в п. 2.5.2).

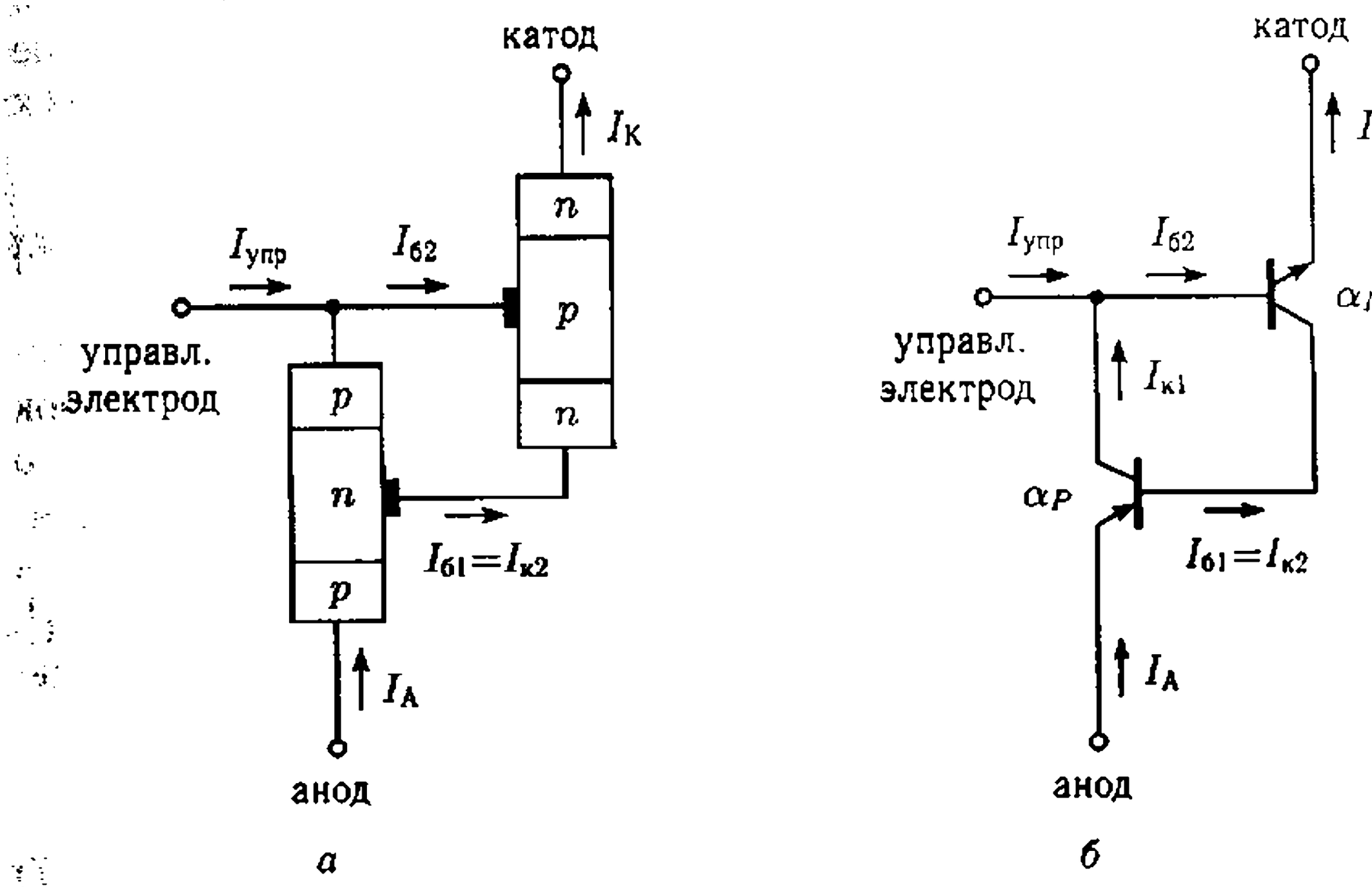


Рис. 3.4. Эквивалентная схема тиристора

Выведем теперь выражение для приращения тока базы p–n–p-транзистора. Исходя из уравнений $I_{к1} = \alpha_P I_A + I_{к01}$ и $I_{б1} = I_A - I_{к1}$ и учитывая, что величина α_P тоже зависит от тока, получаем:

$$dI_{б1} = \left(1 - \alpha_P - I_A \frac{d\alpha_P}{dI_A}\right) dI_A - dI_{к01}. \qquad (3.3)$$

Здесь I_A — ток анода тиристора (эмиттера p–n–p-транзистора), $I_{к01}$ — обратный ток коллектора этого транзистора, а α_P — его *статический* коэффициент усиления. Выражение в скобках

представляет собой разность единицы и коэффициента усиления $\alpha_1 = \alpha_P + I_A(d\alpha_P/dI_A)$ p–n–p-транзистора, измеренного *на малом сигнале*. Поскольку $dI_{б1} = dI_{к2}$, то можно записать, что

$$(1 - \alpha_1)dI_A - dI_{к01} = \alpha_2 dI_K + dI_{к02}. \tag{3.4}$$

Учитывая, что $dI_K = dI_A + dI_{упр}$, окончательно находим

$$dI_A = \frac{\alpha_2 dI_{упр} + d(I_{к01} + I_{к02})}{1 - (\alpha_1 + \alpha_2)}. \tag{3.5}$$

Поскольку при увеличении напряжения на аноде тиристора обратные токи коллекторов обоих входящих в структуру транзисторов возрастают, то условием бесконечно сильного возрастания вызванного ими тока анода при $I_{упр} = 0$ является обращение в единицу суммы малосигнальных коэффициентов усиления двух образующих структуру транзисторов:

$$\alpha_1 + \alpha_2 = 1. \tag{3.6}$$

В проведенном выше расчете мы пренебрегли ударной ионизацией носителей — умножением токов электронов и дырок при прохождении носителями области сильного поля в обратно смещенном p–n-переходе. Учтем этот эффект. Ток дырок, движущихся в p–n–p-транзисторе от анода к катоду, умножается в области сильного поля в M_p раз и составляет $M_p[\alpha_P I_A + I_{к01}]$, а ток электронов в n–p–n-транзисторе, движущихся в направлении от катода к аноду, умножается в M_n раз и составляет $M_n[\alpha_N I_K + I_{к02}]$ [1]). Тогда полный ток в структуре равен

$$I = M_p[\alpha_P I_A + I_{к01}] + M_n[\alpha_N I_K + I_{к02}]. \tag{3.7}$$

Как и выше, рассчитаем приращение тока dI, вызванное увеличением приложенного к структуре напряжения. Изменение этого напряжения вызывает не только изменение токов $I_{к01}$ и $I_{к02}$, но и изменение коэффициентов умножения M_p и M_n:

$$dI = M_p[\alpha_1 dI_A + dI_{к01}] + dM_p[\alpha_P I_A + I_{к01}]$$
$$+ M_n[\alpha_2 dI_K + dI_{к02}] + dM_n[\alpha_N I_K + I_{к02}]. \tag{3.8}$$

---

[1]) Под величинами $I_{к01}$ и $I_{к02}$ выше и здесь мы подразумеваем дырочную и электронную составляющую обратного тока перехода $n1$–$p2$.

При нулевом токе управляющего электрода $I_A = I_K = I$ и тогда, группируя слагаемые в (3.8), получаем, что

$$dI = \frac{M_p\, dI_{к01} + M_n\, dI_{к02} + dM_p[\alpha_P I + I_{к01}] + dM_n[\alpha_N I + I_{к02}]}{1 - M_p\alpha_1 - M_n\alpha_2}. \tag{3.9}$$

Из этой формулы следует, что с учетом ударной ионизации условием переключения тиристора является условие

$$M_p(V_2)\alpha_1 + M_n(V_2)\alpha_2 = 1, \tag{3.10}$$

где V_2 — напряжение на обратно смещенном переходе $n1$–$p2$, которое приблизительно равно приложенному к тиристору напряжению. Входящие в эту формулу коэффициенты усиления α_1 и α_2 зависят от протекающего через тиристор тока (из-за зависимости коэффициента инжекции эмиттера от тока, см. п. 2.2.3) и напряжения V_2 (из-за зависимости коэффициентов переноса носителей в базах транзисторов от толщин электронейтральных областей баз, которые в свою очередь зависят от напряжения смещения, см. п. 2.2.2). Поэтому формула (3.10) представляет собой уравнение, задающее в неявной форме положение точки переключения на прямой ветви вольт-амперной характеристики тиристора.

Если протекающий через тиристор ток станет выше тока переключения ($I_{пер}$ на рис. 3.2), то в тиристоре включается положительная обратная связь. Электроны, инжектируемые переходом $p2$–$n2$ и достигающие базы $n1$, еще сильнее открывают переход $p1$–$n1$, а инжектируемые этим переходом дырки, попадающие в область $p2$, сильнее открывают переход $p2$–$n2$ и стимулируют дальнейшее нарастание тока электронов. Концентрации инжектируемых в области $n1$ и $p2$ дырок и электронов возрастают настолько, что при некоторой величине тока, называемой *током включения* $I_{вкл}$, напряжение на среднем p–n-переходе проходит через нуль (транзисторы входят в режим насыщения). Это происходит при обращении в единицу суммы *статических* коэффициентов усиления, $\alpha_P + \alpha_N = 1$. При $I > I_{вкл}$ все три p–n-перехода в тиристоре смещены в прямом направлении.

Чтобы устойчиво поддерживать тиристор в открытом состоянии, необходимо, чтобы протекающий через него ток был выше некоторой критической величины, называемой *током удержания* ($I_{уд}$ на рис. 3.2). Этот ток определяется из условия обращения в бесконечность производной dI_A/dV_A. Теоретический расчет этой величины представляет большие трудности,

но в идеализированном случае симметричной структуры его значение можно оценить по формуле $I_{уд} \approx 2,5 I_{вкл}$ [160]. Вольт-амперная характеристика тиристора в открытом состоянии при $I > I_{уд}$ практически совпадает с вольт-амперной характеристикой p–i–n-диода (см. п. 1.2.4), поскольку концентрация инжектируемых в области $n1$ и $p2$ («i-область») носителей превышает концентрацию находящихся там легирующих примесей и реализуются условия высокого уровня инжекции. Таким образом, напряжение на открытом тиристоре близко к контактной разности потенциалов в p–n-переходе, а при увеличении тока оно несколько возрастает из-за падения напряжения на «i-области» структуры, сопротивлении толщи областей $p1$ и $n2$ и сопротивлении контактов. Расчет падения напряжения на «i-области» структуры представляет собой отдельную довольно сложную задачу, поскольку при высоких плотностях тока, характерных для тиристоров, существенную роль играют эффекты взаимного рассеяния электронов и дырок и Оже-рекомбинация (см. с. 39).

Для получения низкого падения напряжения на «i-области» структуры, как следует из теории p–i–n-диода (см. формулы (1.64) и (1.65)), необходимо, чтобы диффузионная длина была сравнима с толщиной базы. Для этого в тиристорах стараются использовать полупроводники с большим временем жизни. Однако выбрав такой материал, оказывается трудно достигнуть высокого быстродействия тиристора. Поэтому при разработке тиристоров всегда приходится искать компромисс между параметрами их вольт-амперных характеристик и быстродействием.

Для дополнительного уменьшения падения напряжения на «i-области» структуры в конструкцию современных высоковольтных запираемых тиристоров добавляют еще одну область — сильно легированный буферный n^+-слой между областями $p1$ и $n1$ (см. рис. 3.6). Это позволяет создать, не опасаясь прокола структуры, практически *однородное* электрическое поле в слое $n1$ закрытого тиристора при $V_A > 0$. При этом для получения заданного напряжения пробоя требуется *меньшая* толщина слоя $n1$, что, в соответствии с формулами (1.64) и (1.65), сразу же приводит к заметному уменьшению падения напряжения на «i-области» структуры. Недостатком структур с буферным n^+-слоем является невысокое напряжение пробоя при $V_A < 0$, и поэтому их использование ограничивается работой при одной полярности приложенного напряжения.

При разработке высоковольтных тиристоров исследователи столкнулись с рядом серьезных проблем. Для приборов с рабочим напряжением 10 кВ база тиристора должна иметь концентрацию электронов менее 10^{13} см$^{-3}$ (удельное сопротивление > 500 Ом·см) и толщину не менее 1,2 мм. Попытка создания таких приборов показала, что после нагрева Si до 1250–1300 °С для проведения диффузии происходит заметное снижение удельного сопротивления исходного материала [161], что препятствует созданию высоковольтных приборов. Этот эффект особенно сильно выражен в кристаллах, выращенных по методу Чохральского, которые содержат высокую концентрацию остаточного кислорода. Чтобы уменьшить влияние кислорода, в настоящее время для создания высоковольтных тиристоров используют кремний, полученный бестигельной зонной плавкой, в котором концентрация кислорода существенно ниже, но который часто содержит более высокую концентрацию примесей, создающих глубокие уровни в запрещенной зоне (прежде всего, Cu и Au). Нагревание этих кристаллов почти не изменяет удельное сопротивление, но приводит к заметному уменьшению времени жизни дырок в n-области (до 1–10 мкс). Это вызывает столь сильное уменьшение диффузионной длины, что материал становится практически непригодным для создания тиристоров.

Чтобы повысить время жизни, во время диффузии примеси на поверхность кремния наносится слой *геттера* (добавка кобальта к источнику акцепторной примеси; использование фазы Ni–Co–P–Si или фосфорно-силикатного стекла в качестве источника донорной примеси). Геттер «вытягивает» на поверхность и связывает быстро диффундирующие в Si атомы меди и золота и позволяет увеличить время жизни дырок до 150–400 мкс. После термообработки кристаллов при 400–800 °С (во время создания контактов и припайки вольфрамового термокомпенсатора) время жизни дырок дополнительно увеличивается до ~800 мкс.

Роль управляющего электрода. Наличие управляющего электрода у тиристора позволяет целенаправленно изменять его вольт-амперные характеристики и, в частности, переводить тиристор в открытое состояние, а в запираемых тиристорах (см. ниже) — также и в закрытое состояние. Подавая на управляющий электрод смещение, которое открывает переход $p2$–$n2$ (см. рис. 3.1), мы тем самым вызываем увеличение коэффициента усиления n–p–n-транзистора. Одновременно с этим, как следствие возрастания протекающего через тиристор тока, увеличивается и коэффициент усиления p–n–p-транзистора. В результате этого, в соответствии с уравнением (3.10), напряжение переключения тиристора уменьшается, а при некотором значении тока управляющего электрода, называемом *током спрямления*, тиристор переходит в открытое состояние при любых положительных напряжениях на аноде (на вольт-амперной характеристике

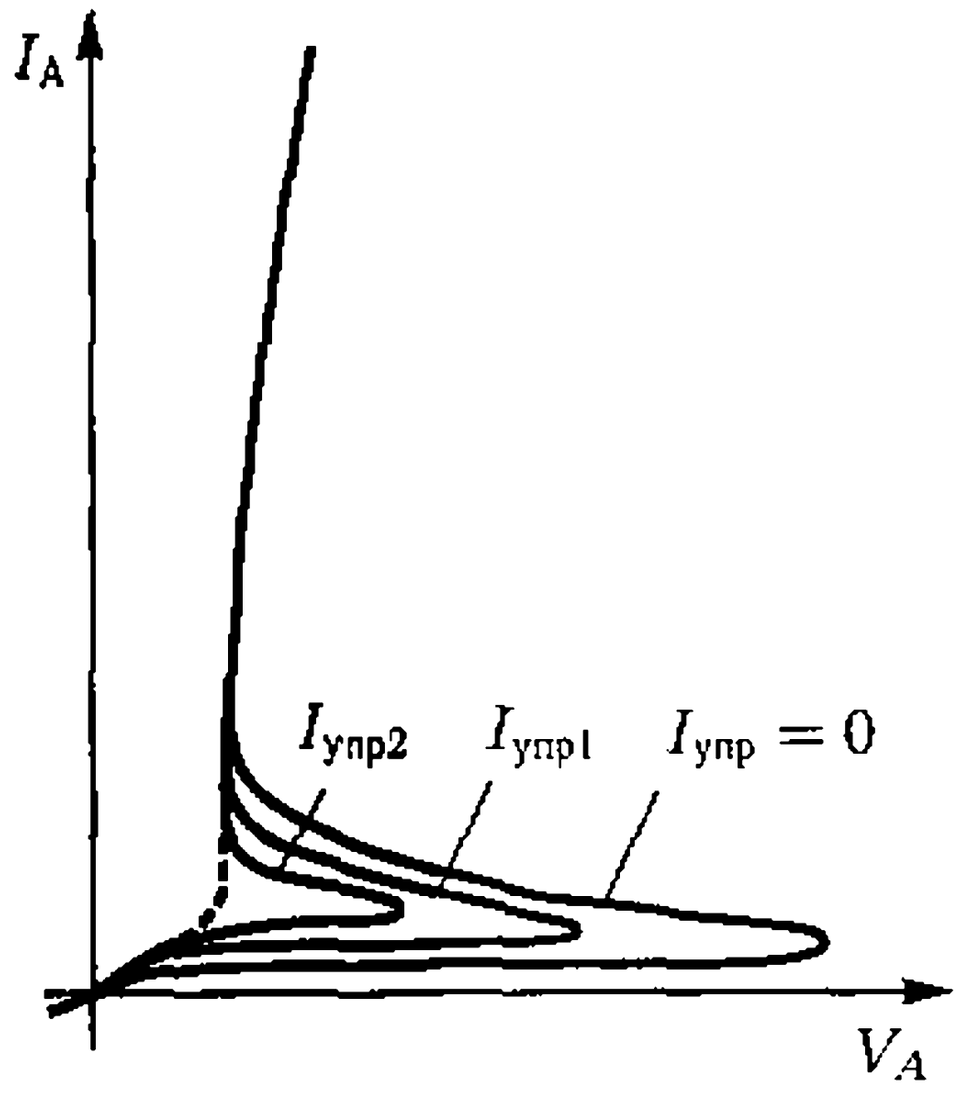


Рис. 3.5. Семейство вольт-амперных характеристик тиристора при различных значениях тока управляющего электрода ($I_{упр.2} > I_{упр.1} > 0$)

исчезает S-образный участок, см. рис. 3.5).

Если при работе тиристора его анодный ток в открытом состоянии будет превышать $I_{уд}$, то для включения тиристора ток через управляющий электрод можно пропускать лишь в течение небольшого времени (10–100 мкс). Через это время описанная выше положительная внутренняя обратная связь в тиристоре включается, и необходимые для поддержания тиристора в открытом состоянии носители обоих типов генерируются уже в самом приборе. Параметры, определяющие время включения тиристора, мы обсудим в п. 3.1.2.

Запираемые тиристоры. Выключение тиристора обычно происходит при уменьшении тока анода ниже $I_{уд}$ или при подаче на анод отрицательного напряжения. Поэтому наиболее удобно использовать тиристоры в цепях переменного тока. При необходимости их применения в цепях постоянного тока особенности выключения тиристоров требуют создания сложных схем коммутации с импульсной подачей отрицательного напряжения на анод тиристора [160]. Однако в принципе тиристор может быть переведен в закрытое состояние и путем подачи на управляющий электрод отрицательного смещения. Это решение используется в так называемых *запираемых тиристорах*.

Основная проблема при создании запираемых тиристоров заключается в том, что для выключения тиристора через управляющий электрод необходимо пропускать большой ток, равный

$$I_{упр} = I_A \cdot (\alpha_P + \alpha_N - 1)/\alpha_N,$$

который обычно составляет 20–40% от тока анода I_A [162]. Из-за трудности получения низкого сопротивления области $p2$ выпуск запираемых тиристоров долгое время ограничивался маломощными приборами. Только с развитием технологии создания многоячеечных структур с малыми размерами элементов (~100 мкм) и разветвленной структурой управляющего

электрода появилась возможность выпускать мощные запираемые тиристоры. Первые такие тиристоры, обозначаемые в зарубежной литературе аббревиатурой GTO (gate turn-off), появились в середине 70-х годов.

Особенностью работы тиристоров GTO является то, что из-за конечного прикладываемого к управляющему электроду обратного напряжения (оно ограничено напряжением пробоя перехода $p2$–$n2$, которое составляет ~20 В) и достаточно высокого сопротивления области $p2$ запирание перехода $p2$–$n2$ начинается в местах, наиболее близко расположенных к выводам управляющего электрода (см. рис. 3.6 *а*), и лишь затем фронт выключения распространяется на всю площадь перехода. Наиболее критическим моментом при запирании тиристора является тот момент, когда площадь области проводящего состояния становится очень малой и плотность протекающего через нее тока резко возрастает. Если при этом запирание происходит слишком медленно, то возникает опасность сильного локального перегрева тиристора и его выхода из строя. Для запирания перехода $p2$—$n2$ за 1 мкс при характерной скорости движения фронта 50 мкм/мкс ширина ячейки не должна превышать 100 мкм. Из этого с очевидностью следует, что мощные запираемые тиристоры можно создавать лишь в виде многоячеечных структур. Поскольку выключение всех ячеек в таких структурах должно происходить одновременно, требуется, чтобы технологический разброс параметров ячеек был мал. Для уменьшения падения напряжения на области $p2$, при пропускании большого тока $I_{упр}$, выводы управляющего электрода делаются несколько «утопленными» (см. рис. 3.6). К недостаткам тиристоров со структурой GTO следует отнести необходимость подключения к тиристору специальных демпфирующих схем (snubber), при работе которых возникают большие непроизводительные потери энергии.

В середине 90-х годов разработчики фирм ABB Semiconductors и Mitsubishi Electric предложили усовершенствованную конструкцию запираемых тиристоров, названную GCT (gate-commutated turn-off) [163]. Основная идея этой конструкции состоит в том, чтобы подав отрицательный импульс напряжения на управляющий электрод, перенаправить весь анодный ток на этот электрод (см. рис. 3.6 *б*). При этом отрицательное смещение на управляющем электроде закрывает переход $p2$–$n2$, тиристор превращается в p–n–p-транзистор с оборванной базой и быстро переходит в закрытое состояние. Для обеспечения высокой скорости нарастания тока управляющего электрода (до 4–6 кА/мкс) используется специальная низкоиндуктивная

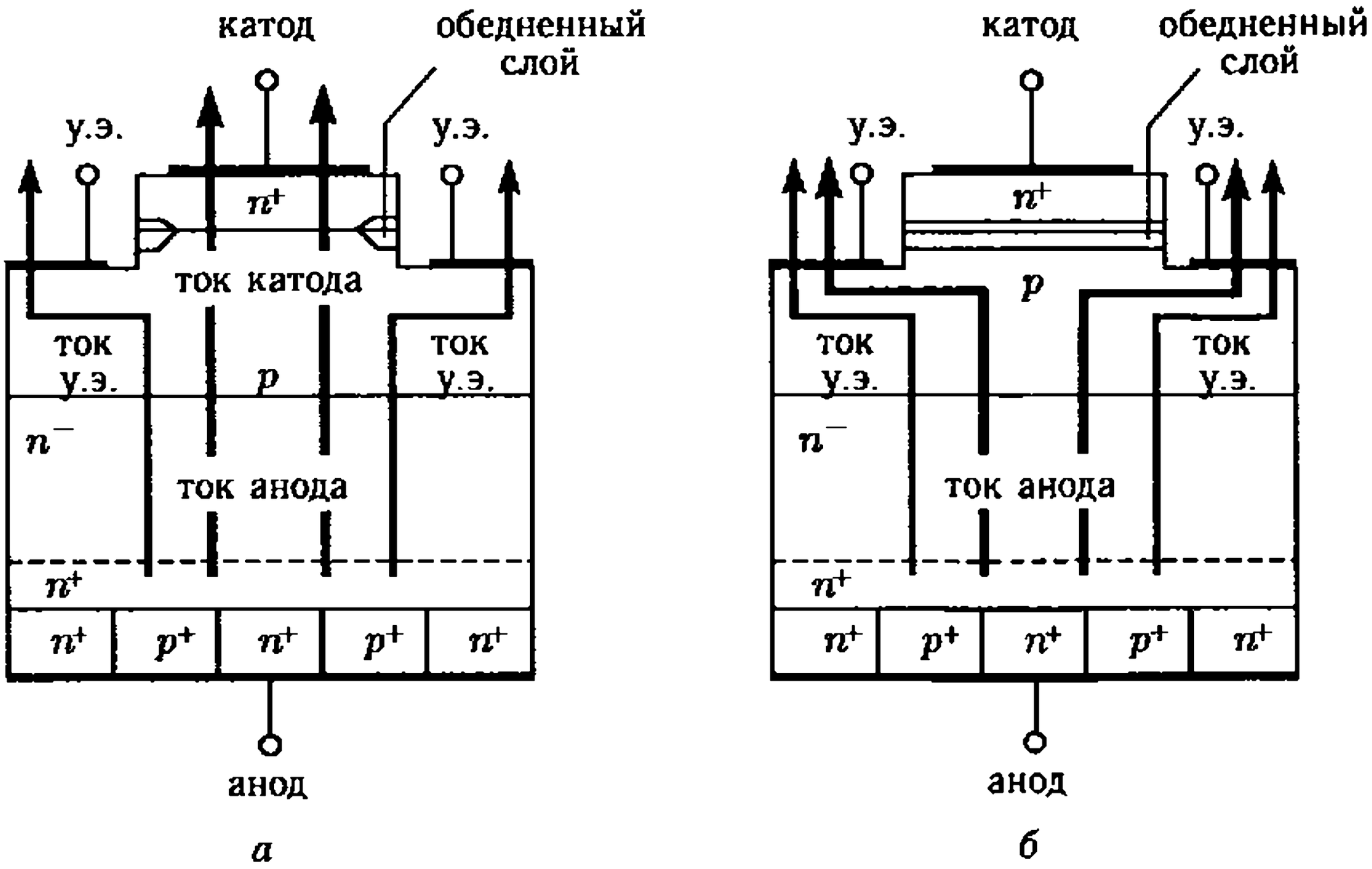


Рис. 3.6. Устройство одной ячейки запираемых тиристоров со структурой GTO (*а*) и GCT (*б*)

конструкция корпуса тиристора и низкоиндуктивная цепь формирователя импульсов управления; эти два блока часто объединяются в единую конструкцию, называемую *GCT с интегрированным блоком управления* (IGCT). Намного более простая конструкция блока управления тиристорами GCT по сравнению с тиристорами GTO и отсутствие необходимости в использовании демпфирующих схем позволяют существенно снизить потери и повысить общий к.п.д. устройства преобразования.

Работа еще одной управляемой сильноточной многослойной структуры IGBT, в которой используется полевой способ управления прибором, будет рассмотрена в п. 4.1.7.

Тиристоры с закороченным катодом. Говоря о вольт-амперной характеристике тиристора, необходимо отметить одну специфическую особенность тиристоров — возможность их открывания быстро нарастающим импульсом напряжения, подаваемым на анод [159]. Причиной этого является протекание в структуре тока заряда барьерной емкости обратно смещенного перехода $n1$–$p2$. Этот ток, суммируясь с током утечки тиристора, приоткрывает транзисторы, вызывает увеличение их коэффициентов усиления и при выполнении условия (3.10) приводит к включению описанного выше механизма положительной обратной связи, которая и переводит тиристор в открытое

состояние. Характерная скорость нарастания напряжения на аноде dV_A/dt, приводящая к этому явлению, составляет ~20 В/мкс. Очевидно, что эта особенность тиристоров, которая может приводить к неконтролируемому открыванию тиристора при попадании на его анод выбросов напряжения, крайне нежелательна, и поэтому Олдрич и Холоньяк решили модифицировать конструкцию тиристора, предложив *тиристор с закороченным катодом* [164].

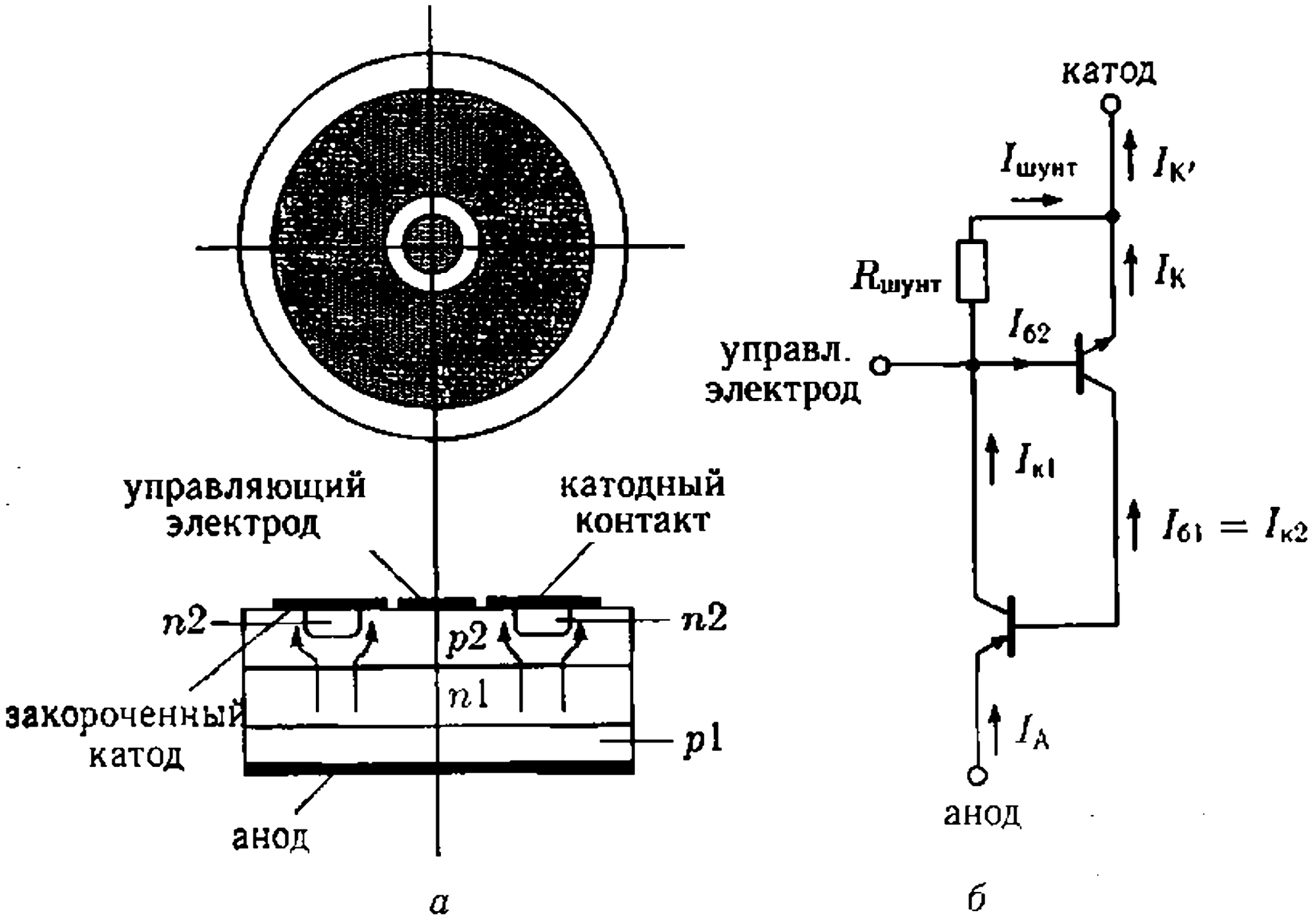


Рис. 3.7. Тиристор с закороченным катодом: *а* — конструкция тиристора, *б* — его эквивалентная схема

Конструкция и эквивалентная схема тиристора с закороченным катодом показаны на рис. 3.7. Тиристоры, рассчитанные на работу с большими токами, традиционно изготавливаются в форме тонких пластин, которые одной стороной (анодом) напаиваются на диск из Mo или W (термокомпенсатор); с другой стороны на поверхности пластины располагаются кольцеобразный вывод катода и вывод управляющего электрода в центре. Особенностью конструкции тиристора с закороченным катодом является то, что при изготовлении контакта к катоду размер контактной площадки делается несколько больше размера области $n2$, и в некоторых местах эта область оказывается электрически соединенной с областью $p2$ управляющего электрода. При этом в эквивалентной схеме эмиттерный переход n–p–n-транзистора оказывается зашунтированным сопротивле-

нием $R_{\text{шунт}}$ (см. рис. 3.7 *б*), и при подаче на анод тиристора быстро нарастающего импульса напряжения емкостный ток, протекающий через структуру, утекает через этот резистор, оставляя n–p–n-транзистор закрытым. Использование закороченного катода позволяет сделать тиристоры устойчивыми по отношению к подаче на анод импульсов напряжения, нарастающих со скоростью до 500–3000 В/мкс. Поскольку коэффициент усиления транзистора с зашунтированным эмиттерным переходом близок к нулю, то условием переключения тиристора, которое следует из формулы (3.10), в этом случае становится равенство $M_p(V_2)\alpha_1 = 1$.

Очень важно, что структура с закороченным катодом позволяет также повысить и максимальную рабочую температуру тиристора, так как ток утечки закрытого тиристора, возрастающий с увеличением температуры, стекает через шунт и не позволяет n–p–n-транзистору открыться.

Структура, аналогичная описанной выше, используется и в запираемых тиристорах. Однако поскольку в этих тиристорах переход $p2$–$n2$ используется для запирания тиристора, в них шунтируется не катодная, а анодная часть структуры (переход $p1$–$n1$, см. рис. 3.6).

Защита тиристоров от пробоя. При создании высоковольтных тиристоров необходимо иметь в виду, что из-за существования встроенного заряда в слое окисла на поверхности полупроводника или загрязнения поверхности ионами вблизи места выхода p–n-перехода на поверхность возникает изгиб границ области пространственного заряда. При этом напряженность электрического поля на поверхности может оказаться выше, чем в объеме полупроводника, и при высоком обратном напряжении на p–n-переходе может возникнуть *поверхностный пробой*. Поверхностный пробой более опасен для

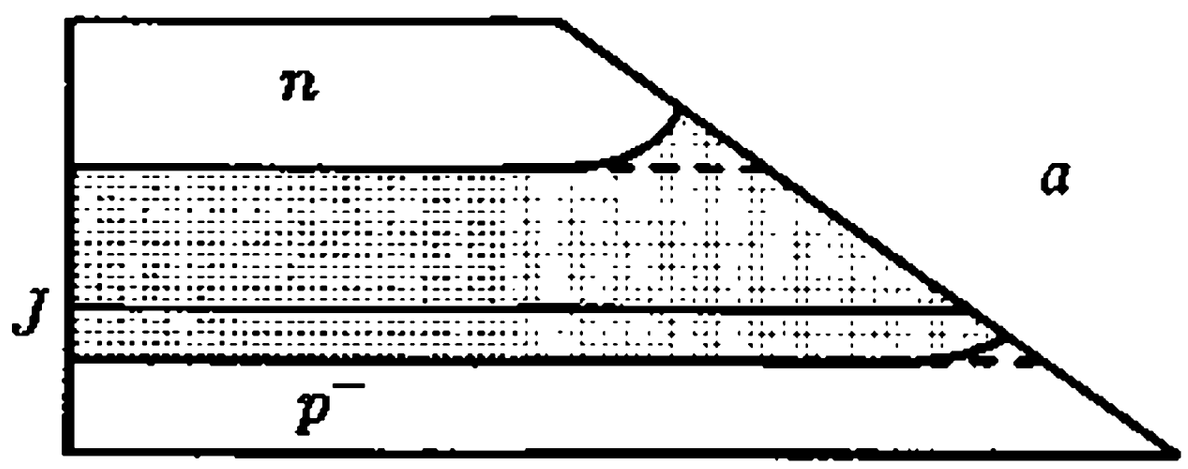


Рис. 3.8. Изгиб границ области пространственного заряда вблизи наклонной границы раздела: *а* — положительная фаска, *б* — отрицательная фаска [162]. J — металлургическая граница p–n-перехода

полупроводникового прибора по сравнению с объемным пробоем, поскольку энергия, которая при поверхностном пробое выделяется в очень тонком приповерхностном слое, легко выводит прибор из строя.

Одним из способом уменьшения поверхностной напряженности электрического поля является создание *фасок*, то есть придание боковой поверхности прибора в месте выхода p–n-перехода наклонной (конической) формы [14, 160, 162]. Результаты численного решения двумерного уравнения Пуассона для структур с фасками приведены на рис. 3.8. На рис. 3.8*а* показан изгиб границ области пространственного заряда для положительной фаски [1]). Появление этого изгиба связано с тем, что p–n-переход в целом должен оставаться нейтральным, а объемы, занимаемые ионизованными донорами и акцепторами, при плоских границах области пространственного заряда и наклонной боковой границе p–n-перехода различны. Для компенсации этой разницы объемов и возникает изгиб границ, в результате которого суммарный заряд в области пространственного заряда становится равным нулю. Предсказываемое расчетом более сильное смещение границы области пространственного заряда в слабо легированной части структуры с положительной фаской гарантирует понижение напряженности электрического поля по мере приближения к поверхности, и поэтому пробой в таких структурах начинается в средней части прибора.

В структурах с отрицательной фаской ситуация несколько сложнее. Выбирая малое значение угла фаски ϕ, в этих структурах также можно добиться уменьшения напряженности электрического поля на поверхности. Однако решение уравнения Пуассона показывает, что максимум напряженности электрического поля в этом случае находится на некотором расстоянии от поверхности в объеме структуры (пунктирная линия на рис. 3.8*б*). Максимальное значение напряженности электрического поля в этой области может на 5–25% превышать ее значение в середине структуры. Хотя пробой структур с отрицательной фаской происходит при более низких напряжениях по сравнению со структурами с положительной фаской, он все же идет в объеме

---

[1]) Положительный знак угла фаски приписывается фаске, для которой площадь поперечного сечения прибора уменьшается при переходе от сильно легированной к слабо легированной области структуры. В отечественной литературе для обозначения знака угла фаски используются термины «прямая» и «обратная» фаски, соответствующие отрицательной и положительной фаскам.

прибора и, следовательно, не так опасен, как поверхностный пробой.

На рис. 3.9 показано поперечное сечение двух конструкций высоковольтных тиристоров с фасками. В первой из них используются положительная и отрицательная фаски, однако из-за того, что отрицательная фаска должна иметь достаточно малый угол, эта конструкция тиристора характеризуется большой непроизводительной потерей площади структуры. Этого недостатка нет во второй конструкции тиристора, в которой обе фаски положительны. К сожалению, создание таких приборов затруднено сложностью формирования профилированной боковой поверхности в достаточно тонкой пластине.

В приборах с планарной геометрией для защиты p–n-переходов от поверхностного пробоя используется целый ряд других приемов (полевые обкладки, охранные и ограничительные кольца), которые подробно описаны в книге [160]. Эти приемы уменьшения поверхностной напряженности электрического поля широко используются не только при создании тиристоров, но также и при создании мощных диодов и транзисторов. Мы ограничимся здесь лишь упоминанием об этих приемах, поскольку они используются в приборах, работающих при напряжениях до нескольких киловольт, а более высокие рабочие напряжения (до 12 кВ) получаются только при использовании фасок.

Как мы отмечали на с. 54, из-за неоднородности полупроводника в p–n-переходах могут возникать локальные области с более высокой напряженностью электрического поля, в которых и будет в первую очередь возникать лавинный пробой. Для тиристоров большой площади, изготавливаемых из целых пластин кремния диаметром 100–150 мм, проблема однородности полупроводника становится особо острой. Поэтому при создании таких тиристоров используется особо однородно легированный кремний, полученный *трансмутационным легированием* [165].

Трансмутационное легирование — явление образования в полупроводнике легирующих примесей под воздействием проникающей радиации — было обнаружено Ларк-Горовицем с сотр. в 1948 г. При захвате нейтронов, дейтронов или альфа-частиц ядрами атомов, из которых состоит полупроводник, происходят ядерные превращения (трансмутации), в результате которых могут возникать атомы легирующих примесей. Например, при облучении нейтронами естественной смеси изотопов Ge в нем образуются изотопы легирующих примесей ^{71}Ga, ^{75}As и ^{77}Se [166]. При трансмутационном легировании кремния единственной реакцией, приводящей к появлению легирующей примеси, является реакция $^{30}\mathrm{Si}(n,\gamma)\rightarrow{}^{31}\mathrm{Si}(\beta^-)\rightarrow{}^{31}\mathrm{P}$ с периодом полурас-

пада промежуточного изотопа ^{31}Si, равным 2,62 часа. Малое сечение захвата нейтронов кремнием (коэффициент поглощения составляет 0,008 $см^{-1}$) гарантирует высокую однородность легирования толстых образцов кремния (изменение концентрации фосфора на 1 % на расстоянии 50 мм) [14, 108, 165].

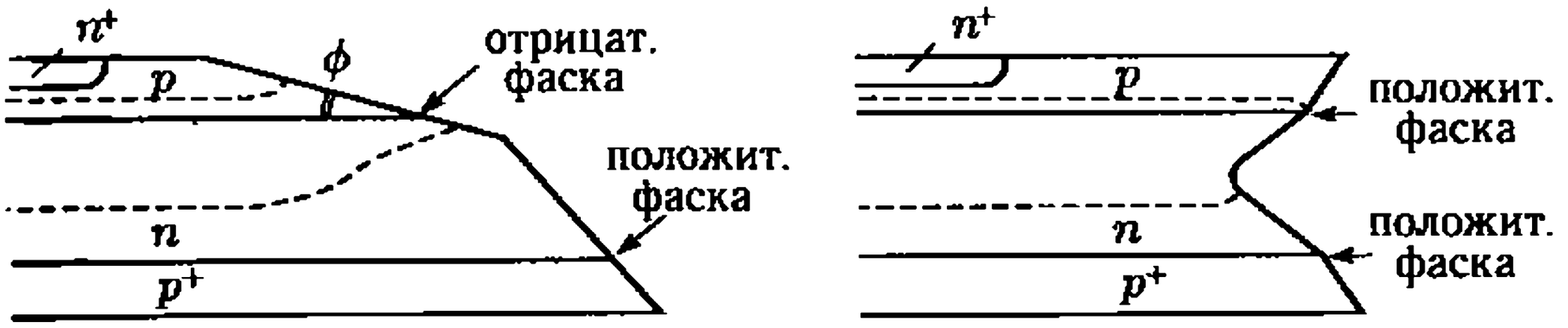


Рис. 3.9. Поперечное сечение двух конструкций высоковольтного тиристора с фасками. Штриховыми линиями показаны границы области пространственного заряда в закрытом тиристоре при положительном напряжении на аноде [52]

Хотя подавляющее большинство тиристоров в настоящее время изготавливается из кремния, все бóльшее внимание исследователей обращается к SiC. Этот материал имеет теплопроводность, которая в 3 раза выше теплопроводности Si и даже превышает теплопроводность меди. Напряженность поля лавинного пробоя в карбиде кремния в 10 раз выше, чем в кремнии, и, наконец, p–n-переходы из SiC способны работать при высоких температурах. В настоящее время уже выпускаются тиристоры из SiC на ток до 250 А и напряжение до 4500 В. Эти приборы сохраняют работоспособность до 350 °С. Широкому использованию карбида кремния мешают пока еще высокая плотность дефектов ($2\ см^{-2}$ на пластинах диаметром 75 мм) и высокая стоимость.

3.1.2. Процессы включения и выключения тиристора. Как мы уже говорили, для перехода тиристора в открытое состояние необходимо, чтобы инжектируемые переходами $p1$–$n1$ и $p2$–$n2$ электроны и дырки успевали пройти всю структуру и реализовали заложенную в конструкции прибора положительную обратную связь. Однако поскольку мы имеем дело со сравнительно медленными процессами диффузии и довольно толстыми областями $n1$ и $p2$, то переходные процессы, связанные с включением тиристора, протекают сравнительно медленно. Характерные времена диффузии носителей через базы соответствующих транзисторов составляют

$$t_1 = \frac{W_{n1}^2}{2D_p}, \qquad t_2 = \frac{W_{p2}^2}{2D_n}. \tag{3.11}$$

За время включения тиристора обычно принимают величину, равную среднему геометрическому из указанных величин, то есть $t_{\text{вкл}} = \sqrt{t_1 t_2}$. Из этих формул следует, что чем с более высоковольтным прибором мы имеем дело (то есть чем толще его база), тем медленнее он включается.

Выключение тиристора обычно требует намного больше времени, чем его включение. Чтобы выключить тиристор, на его анод обычно подают отрицательное напряжение. Кинетика выключения тиристора является достаточно сложной [160, 162], причем из-за сохранения локальной электронейтральности кинетика изменения концентраций электронов и дырок должна анализироваться совместно.

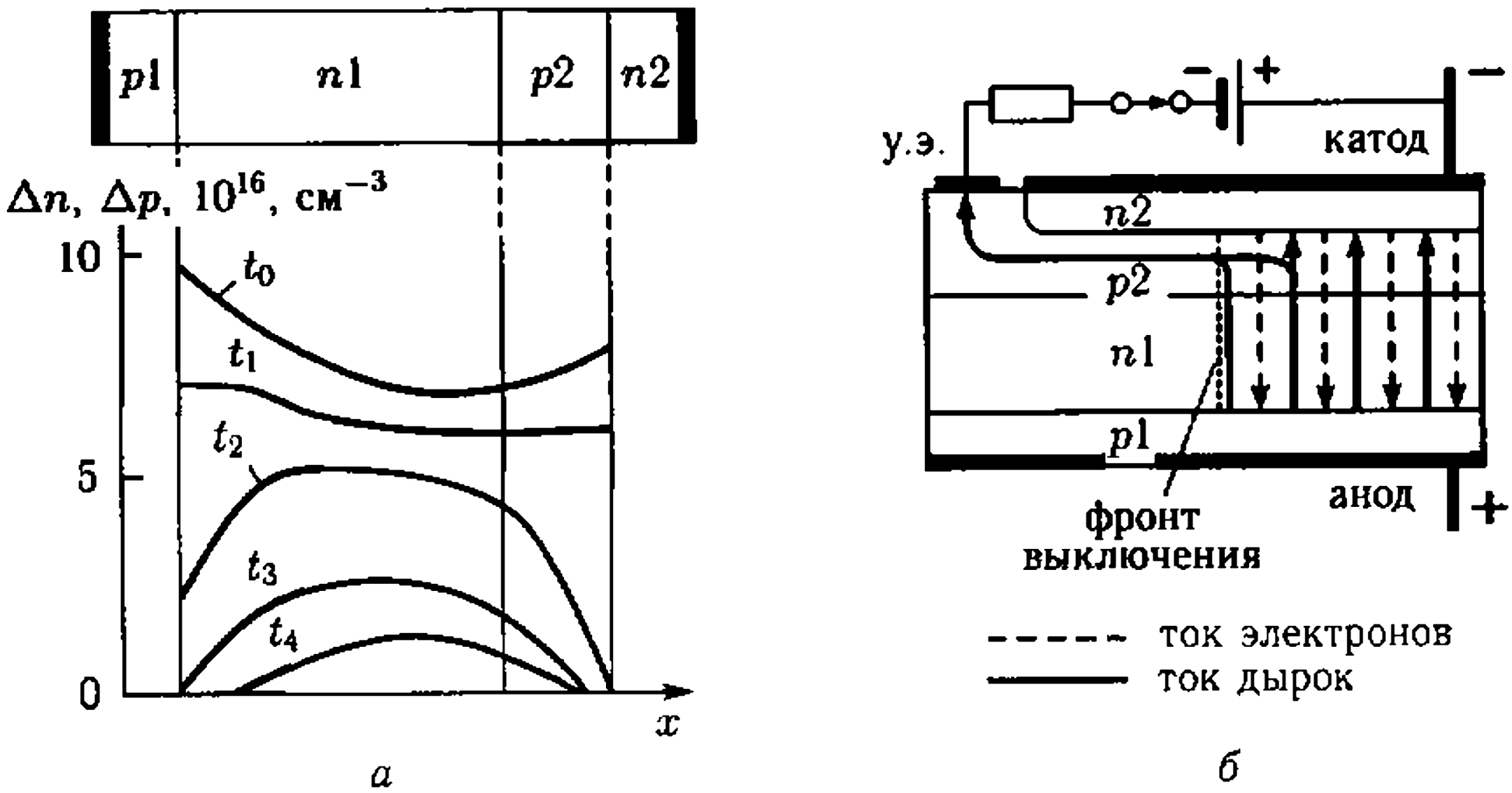


Рис. 3.10. Кинетика изменения концентраций инжектированных носителей в тиристоре на разных этапах его выключения (*а*), линии тока в частично выключенном запираемом тиристоре (*б*) [162]

Рассмотрим процесс выключения тиристора при подаче отрицательного напряжения на его анод. В начальный момент времени, когда полярность напряжения на аноде меняется на противоположную, все три p–n-перехода остаются смещенными в прямом направлении (момент t_0 на рис. 3.10 *а*). Протекающий через тиристор ток отрицательной полярности вызывает экстракцию носителей из баз транзисторов $n1$ и $p2$ переходами $p1$–$n1$ и $p2$–$n2$, соответственно. Одновременно в базах происходит и обычная рекомбинация неравновесных носителей. При этом, поскольку время жизни электронов в области $p2$ обычно меньше времени жизни дырок в области $n1$, то первым закрывается переход $p2$–$n2$ (момент t_2 на рис. 3.10 *а*). На этом завершается

первый этап выключения тиристора. Дальнейшая кинетика зависит от того, достигает ли напряжение на аноде к концу первого этапа напряжения пробоя перехода $p2$–$n2$. Если это так, что встречается довольно часто, то переход $p2$–$n2$ пробивается, и на втором этапе продолжается экстракция дырок из области $n1$ переходом $p1$–$n1$. В какой-то момент времени смещение на этом переходе становится отрицательным (момент t_3 на рис. 3.10*а*), и только с этого момента ток анода начинает быстро уменьшаться. Завершением второго этапа принято считать момент, когда ток анода упадет до величины $I_{\text{уд}}$. Поскольку второй этап намного продолжительнее первого этапа, характерное время выключения тиристора можно оценить по формуле

$$t_{\text{выкл}} \approx \tau_p \ln\left(\frac{I_{\text{прям}}}{I_{\text{уд}}}\right), \qquad (3.12)$$

где $I_{\text{прям}}$ — ток, протекавший в прямом направлении, $I_{\text{уд}}$ — ток удержания, а τ_p — время жизни дырок в базе $n1$. Характерные значения времени выключения составляют 10–200 мкс; для уменьшения времени жизни и ускорения выключения базу тиристора легируют золотом, платиной, а также облучают тиристоры быстрыми электронами или протонами. Следует иметь в виду, что полное восстановление характеристик тиристора требует дополнительного времени, и поэтому, чтобы исключить неконтролируемое включение тиристора, полное рабочее напряжение положительной полярности можно подавать на тиристор только после истечения некоторого времени, называемого *коммутационным временем выключения* t_q. Для грубой оценки этой величины можно использовать формулу $t_q \approx 10\tau_p$ [162]. Максимальная частота, на которой могут работать мощные тиристоры, составляет 1–3 кГц, хотя в некоторых конструкциях прибора она может быть повышена до десятков килогерц.

В запираемых тиристорах GTO процесс выключения протекает несколько иначе. Как мы уже отмечали на с. 225, из-за сравнительно высокого сопротивления области $p2$, кинетические процессы в этих приборах усложнены конечной скоростью движения фронта выключения. Поэтому после начала пропускания тока обратной полярности через управляющий электрод процесс выключения начинается в области, прилегающей к этому электроду (см. рис. 3.10*б*). Отрицательное смещение на переходе $p2$–$n2$, действующее в области левее фронта выключения, прекращает инжекцию электронов в этой части перехода, и вследствие уменьшения неравновесного заряда

электронов в расположенной напротив части области $n1$ ослабляется инжекция дырок переходом $p1$–$n1$. Это приводит к уменьшению потока дырок, поступающих в область $p2$ в районе фронта выключения, и позволяет фронту выключения продвинуться дальше вправо. Когда площадь включенной части структуры становится очень малой, происходит спад анодного тока. Первый этап выключения завершается прекращением инжекции электронов на всей площади перехода $p2$–$n2$. На втором этапе протекают два процесса: достаточно быстрый процесс восстановления области пространственного заряда обратно смещенного перехода $p2$–$n2$ и более медленный процесс экстракции дырок, накопленных в базе $n1$, смещенным в обратном направлении переходом $n1$–$p2$. На этом этапе ток катода уже равен нулю, а ток управляющего электрода и ток анода быстро спадают во времени.

Фототиристоры. Для увеличения скорости включения тиристоров была разработана специальная модификация конструкции тиристора — *фототиристор*. В этой конструкции необходимые для открывания тиристора носители создаются *светом*. Максимум спектральной чувствительности кремниевых фототиристоров лежит около $\lambda = 950$ нм, что соответствует максимуму в спектре излучения светодиодов из GaAs. Возбуждение неравновесных носителей происходит во всем объеме структуры. Электронно-дырочные пары, возникшие в результате поглощения квантов света в области пространственного заряда перехода $n1$–$p2$, через несколько наносекунд оказываются в областях баз составляющих тиристор транзисторов и практически одновременно открывают их. Поэтому при высоких интенсивностях света $t_{вкл}$ в фототиристорах может измеряться десятками наносекунд (против единиц и десятков микросекунд в обычной конструкции).

По данным фирмы Siemens, импульс света мощностью 40 мВт производит эффект, эквивалентный воздействию импульса тока управляющего электрода, равного 50 А (для сравнения: средний ток управляющего электрода в силовых тиристорах составляет 3 А).

Оптическое управление тиристором особенно широко используется при создании высоковольтных приборов (на напряжение выше 4,5 кВ), поскольку этот способ управления не требует сложных схем формирования управляющих сигналов и позволяет одновременно открывать несколько тиристоров (последовательное соединение тиристоров используется для создания высоко-

вольтных схем, например, на напряжение 500 кВ для работы в линиях электропередачи). Конструктивно фототиристоры могут быть изготовлены либо как приборы с окном, через которое по световоду на тиристор поступает импульс света (эта конструкция характерна для мощных высоковольтных приборов), либо как устройства со встроенным в конструкцию светодиодом из GaAs, который электрически изолирован от силовой части прибора (такая конструкция называется *оптронным тиристором* или *оптотиристором*). Благодаря оптической развязке оптотиристоры обладают повышенной помехоустойчивостью по сравнению с обычными тиристорами.

Отметим, что использование идеи фототиристора позволяет преодолеть еще одно серьезное техническое ограничение, существующее у мощных тиристоров: ограничение на скорость нарастания протекающего через прибор тока dI_A/dt. Дело в том, что из-за конечного сопротивления области $p2$ переход тиристоров большой площади в открытое состояние обычно начинается вблизи вывода управляющего электрода и уже потом возбуждение распространяется на весь объем. Скорость поперечного распространения возбуждения достаточно мала (20–200 мкм/мкс), и поэтому если не ограничить скорость нарастания анодного тока, то протекание через структуру высокой плотности тока может вызвать локальный перегрев, образование шнура тока в области первичного канала и проплавление этой области [1]). Чтобы избежать этого, в приборах большой площади для увеличения скорости распространения возбуждения на поверхности структуры создаются специальные конфигурации вывода управляющего электрода, а также используются структуры с усилением сигнала управления, в которых маломощный вспомогательный тиристор формирует мощный управляющий сигнал для быстрого включения основного тиристора [14, 160, 163]. Для ограничения скорости нарастания анодного тока последовательно с тиристором (как часть демпфирующей схемы) обычно включается катушка индуктивности.

---

[1]) Статистика показывает, что при правильной эксплуатации приборов деградация параметров тиристоров и их выход из строя редко происходит из-за проплавления структуры. Гораздо чаще они выходят из строя в результате растрескивания пластины, которое является следствием термических напряжений, возникающих каждый раз после прохождения через тиристор импульса тока. Заметим, что перегрев активной области тиристора до 200–280 °C может приводить к исчезновению запирающей способности тиристора как в прямом, так и в обратном направлениях.

Отечественной промышленностью разработано и выпускается множество тиристоров. Примерами тиристоров малой мощности могут служить приборы КУ202 (средний ток 10 А, максимальное обратное напряжение до 400 В) и запираемый тиристор КУ204 (2 А, 200 В). Тиристоры средней мощности представляют фототиристор ТФ132-25 (25 А, 1 кВ) и оптотиристор ТО142-80 (80 А, 1,2 кВ). К мощным тиристорам относятся Т173-5000 (5000 А, 800 В), Т193-2000 (2000 А, 6 кВ), запираемый тиристор Т3273-2500 (2500 А, 4,5 кВ). Максимально достигнутые в настоящее время параметры для тиристоров составляют 6000 А и 6 кВ, а для фототиристоров — 4000 А и 8 кВ.

3.2. Многослойная структура — симистор

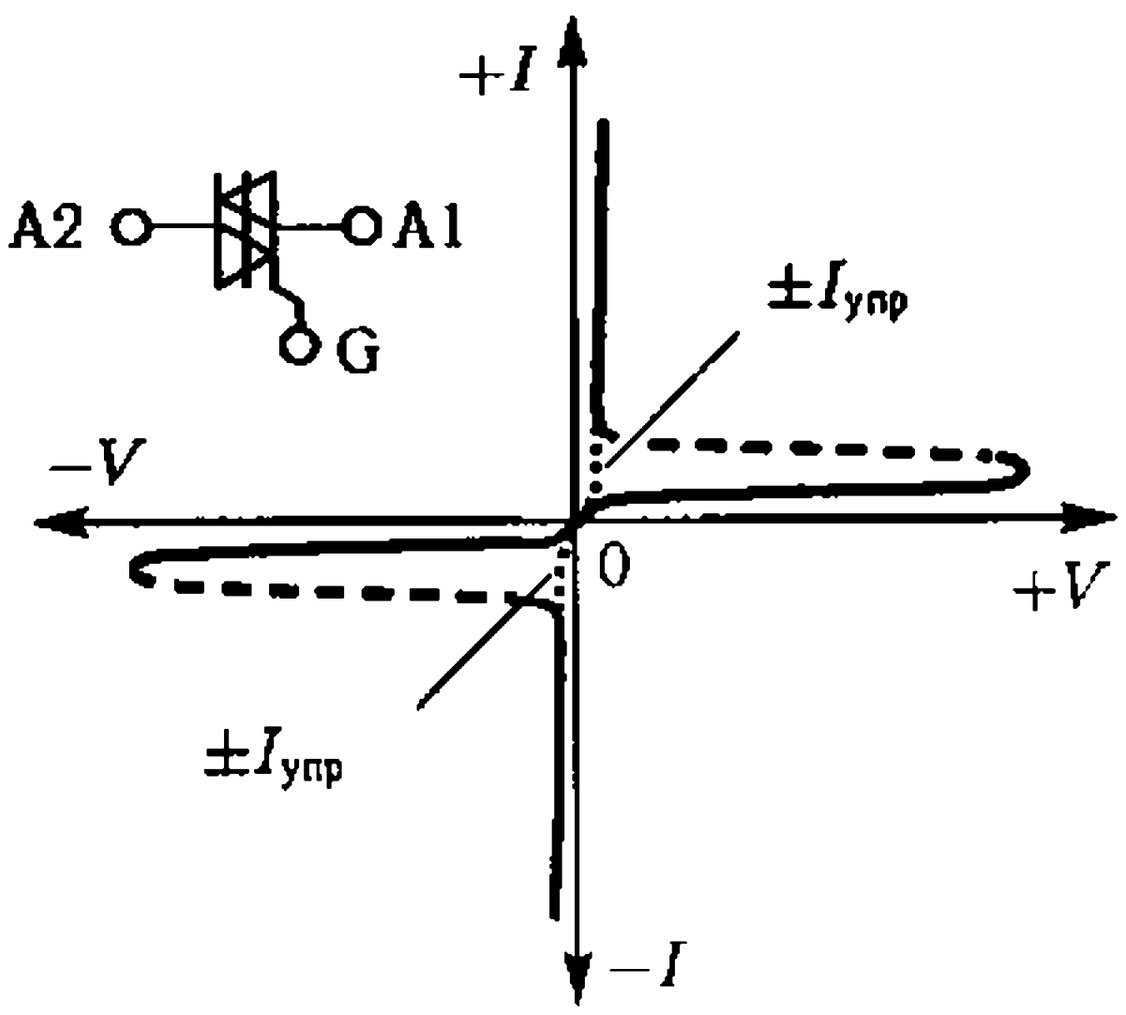


Рис. 3.11. Вольт-амперная характеристика симистора

Еще более сложную структуру по сравнению с тиристором имеет прибор, называемый *симистором* (симметричным тиристором или триаком) [1]. Этот прибор является управляемым *коммутатором переменного тока*, то есть его вольт-амперная характеристика симметрична по отношению к полярности подаваемого на него напряжения (см. рис. 3.11), а открывание прибора может осуществляться подачей на управляющий электрод напряжения как положительной, так и отрицательной полярности. Для получения такой вольт-амперной характеристики в структуре симистора создается шесть областей разного типа проводимости (см. рис. 3.12). Левая и правая половины структуры представляют собой два тиристора с закороченным катодом (см. п. 3.1), включенных параллельно, но навстречу друг другу. Это означает, что при подаче на симистор переменного напряжения ток положительной и отрицательной полуволн будет протекать

[1]) По-английски triac — это сокращение от triode ac switch (трехэлектродный коммутатор переменного тока). Первые триаки были изготовлены в 1965 г. [167].

через разные половины прибора. Для управленияструктурой в ней создается дополнительная область n-типа ($n3$ на рис. 3.12). Для обеспечения возможности открывания тиристора обеими полярностями напряжения на управляющем электроде вывод этого электрода делается «закороченным».

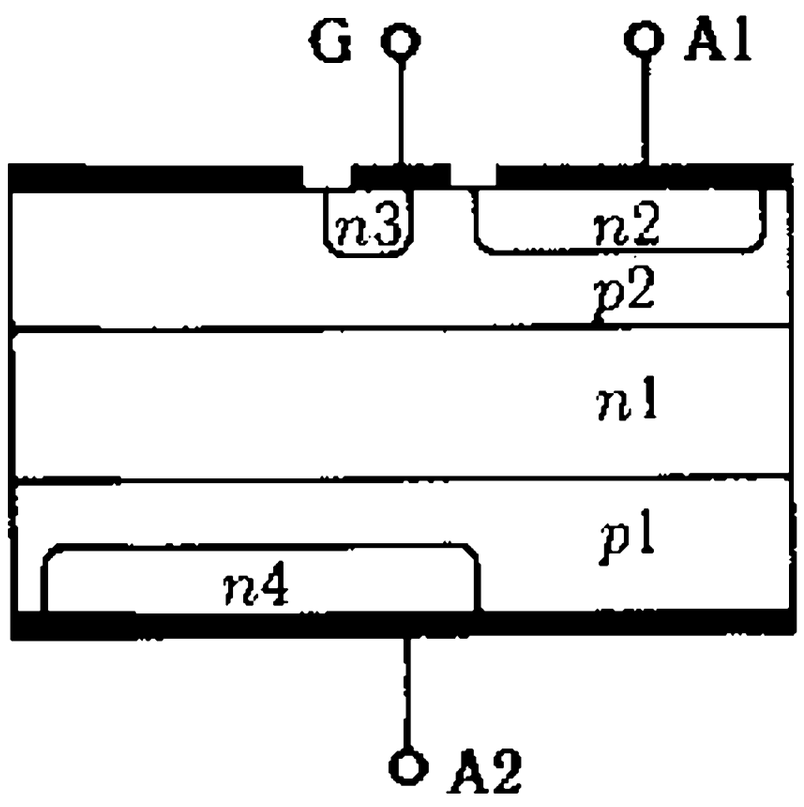


Рис. 3.12. Поперечное сечение симистора (структуры с шестью областями разного типа проводимости и пятью p–n-переходами)

Покажем, что представленная структура позволяет использовать ее в качестве коммутатора переменного тока. Процессы, приводящие к открыванию прибора при четырех возможных комбинациях полярностей напряжения, приложенных к симистору и его управляющему электроду, показаны на рис. 3.13. Отсчет всех напряжений будем проводить относительно электрода A1.

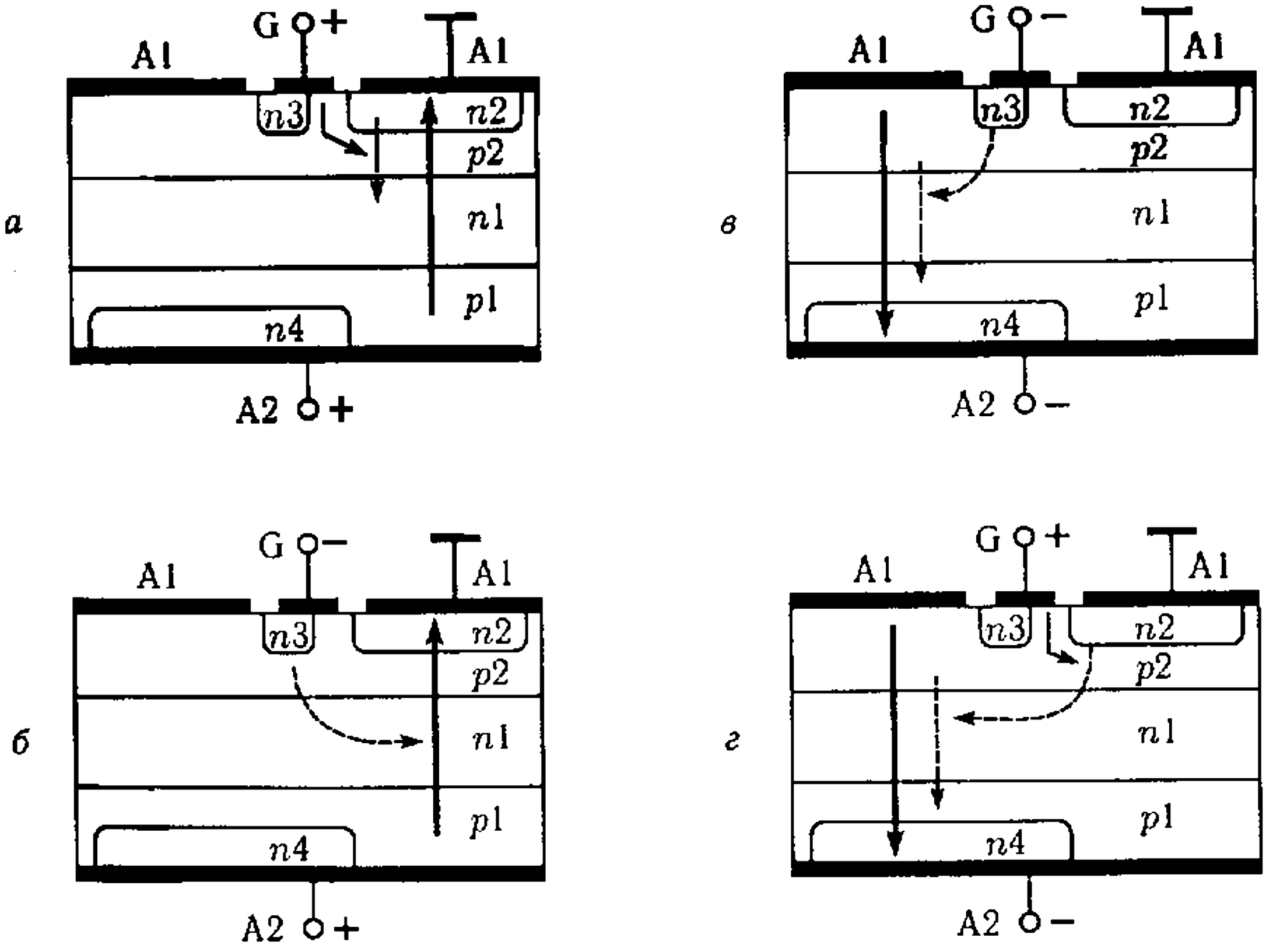


Рис. 3.13. Схема, поясняющая возможность открывания симистора в каждой из четырех возможных комбинаций напряжений, приложенных к симистору и его управляющему электроду. Пунктирной линией обозначен ток первичной инжекции, который вызывает открывание прибора

При подаче на электрод A2 и управляющий электрод G напряжений положительной полярности правая половина симистора работает как обычный тиристор. При этом левая половина прибора представляет собой тиристор, на который подано обратное смещение, и он неактивен. Если при положительной полярности на электроде A2 на управляющий электрод подать напряжение отрицательной полярности (рис. 3.13*б*), то открывается переход $p2$–$n3$ и инжектируемые из области $n3$ электроны попадают в область базы $n1$, вызывая открывание перехода $p1$–$n1$ и последующий переход правого тиристора в открытое состояние. Если на электрод A2 симистора подать напряжение отрицательной полярности, то начинает работать левая половина прибора. Если при этом на управляющий электрод подано напряжение отрицательной полярности, то открывается переход $p2$—$n3$, инжектированные им электроны проникают в базу $n1$ (пунктирная стрелка на рис. 3.13*в*); для обеспечения электронейтральности туда же начинают инжектироваться дырки из области $p2$ и левый тиристор переходит в открытое состояние. Если, при отрицательной полярности напряжения на электроде A2, на электрод G подано напряжение положительной полярности, то открывается переход $p2$–$n2$ и инжектируемые в область базы $n1$ электроны (пунктирная стрелка на рис. 3.13*г*) по описанной выше схеме вызывают открывание левого тиристора. Следует иметь в виду, что для открывания симистора при разных комбинациях полярностей напряжений требуется разный ток управляющего электрода, который может изменяться в 4 раза [160].

Наряду с симисторами с управляющим электродом разработаны и выпускаются *оптронные симисторы* (оптотриаки), в которых для открывания симистора через электрически изолированный от силовой части светодиод, добавленный в конструкцию прибора, пропускается импульс тока.

Примерами отечественных симисторов могут служить широко используемый в бытовой электронике прибор КУ208 (средний ток 5 А, максимальное напряжение до 400 В), силовой симистор ТС171-320 (320 А, 1,2 кВ) и оптотриак ТСО152-125 (125 А, 1,2 кВ).

Глава 4

ПОЛЕВЫЕ ТРАНЗИСТОРЫ

Класс приборов, называемых *полевыми транзисторами*, объединяет то, что работа этих приборов основана на использовании *эффекта поля* — явления изменения электропроводности полупроводника под действием приложенного к его поверхности внешнего поперечного электрического поля. Идея и несколько вариантов конструкции полевого транзистора были предложены Лилиенфельдом в 1926–28 гг. [168]. Этот транзистор представлял собой плоский конденсатор, одной из обкладок которого являлся тонкий слой полупроводника. При изменении напряжения на конденсаторе менялась величина наведенного электрическим полем заряда в полупроводнике и, следовательно, менялась и проводимость слоя. Это позволяло создать на основе этого эффекта прибор, в котором величина протекающего тока управлялась бы внешним напряжением.

Первые попытки реализовать подобную конструкцию полевого транзистора, предпринятые в Bell Laboratories в середине 40-х годов под руководством Шокли, показали, что лишь очень малая часть наведенного заряда (0,1–10%) дает вклад в изменение проводимости [169]. Оказалось, что внешнее электрическое поле в основном экранируется поверхностными состояниями в полупроводнике (см. п. 1.5) и поэтому наблюдаемый эффект оказывается намного меньше ожидаемого.

Реализовать описываемую конструкцию полевого транзистора удалось лишь в 1960 г. в приборах из кремния при использовании в качестве диэлектрика в конденсаторе тонкого слоя SiO_2, полученного термическим окислением кремния (граница раздела $Si–SiO_2$ характеризуется довольно низкой плотностью поверхностных состояний) [170]. Первым же типом полевых транзисторов, созданных в 1953 г. в Bell Laboratories, были полевые транзисторы с управляющим p–n-переходом [171]. В 1966 г. в связи с освоением новых материалов

(GaAs) и невозможностью получения стабильного окисла на их поверхности Мид [172] предложил и реализовал еще одну конструкцию транзистора — полевой транзистор с барьером Шоттки.

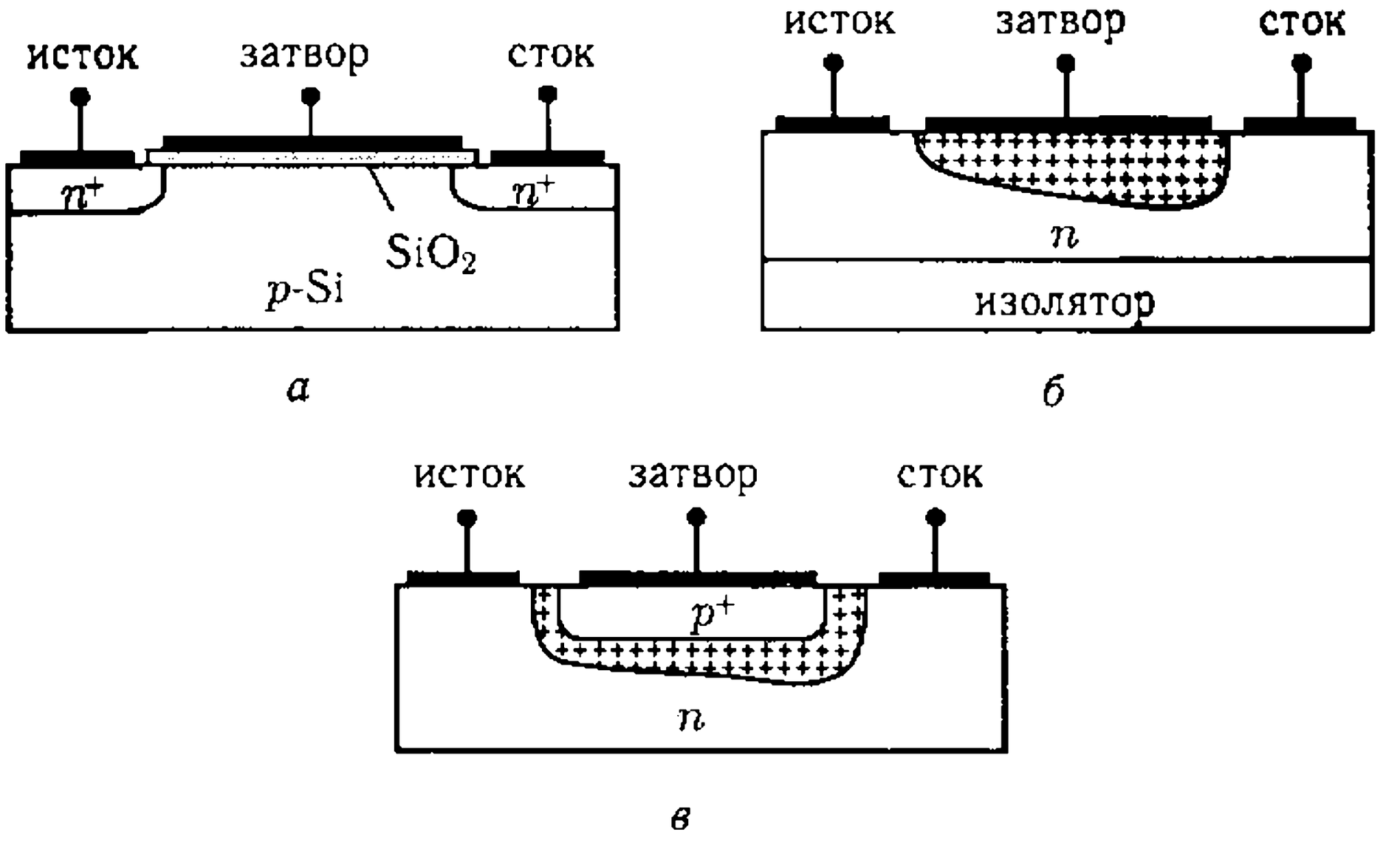


Рис. 4.1. Конструкции полевых транзисторов: *а* — с изолированным затвором и структурой металл–окисел–полупроводник, *б* — с затвором на основе барьера Шоттки, *в* — с затвором на основе p–n-перехода. Заполненные крестиками области на рис. *б* и *в* указывают положение областей пространственного заряда

В настоящее время существуют три основных типа полевых транзисторов, отличающиеся способом создания затвора и методом его изоляции от проводящего канала. Это:

1) полевые транзисторы с изолированным затвором, имеющие структуру металл–окисел–полупроводник (МОП) или металл–диэлектрик–полупроводник (МДП),

2) полевые транзисторы с барьером Шоттки и

3) полевые транзисторы с управляющим p–n-переходом. Устройство этих транзисторов схематически показано на рис. 4.1.

Полевые транзисторы — наиболее распространенный сегодня класс транзисторов. На основе кремниевых МОП-транзисторов в настоящее время создаются наиболее сложные интегральные схемы: микропроцессоры, цифровые сигнальные процессоры, запоминающие устройства ЭВМ. Полевые транзисторы с барьером Шоттки на основе полупроводников группы $A^{III}B^{V}$ используются для создания низкошумящих и мощных усилителей в СВЧ технике. Полевые транзисторы с управляющим p–n-переходом используются в радиоэлектронике в качестве низкошумящих усилителей с высоким входным сопротивлением, электронных ключей, стабилизаторов тока и как резисторы, управляемые напряжением.

4.1. Полевые транзисторы с изолированным затвором

Чтобы определить основные понятия и установить закономерности эффекта поля, рассмотрим сначала так называемый МОП-конденсатор — конденсатор, созданный на поверхности полупроводника, обкладками которого являются полупроводник и металлический электрод, разделенные диэлектрическим слоем из собственного окисла полупроводника.

4.1.1. МОП-конденсатор. На рис. 4.2 показаны энергетические диаграммы идеального МОП-конденсатора (в котором мы пренебрегли существованием электрических зарядов на поверхности полупроводника и в диэлектрике) при различных значениях поданного на металлический электрод потенциала. Рассмотрим их сначала на качественном уровне.

Режим плоских зон, который будет служить нам «точкой отсчета», отвечает той ситуации, когда электрическое поле в приповерхностном слое полупроводника равно нулю (см. рис. 4.2 *а*). Как мы знаем, полупроводник и металл имеют разные работы выхода, и поэтому напряжение $V_з$, которое надо подать на металлический электрод для реализации режима плоских зон, отлично от нуля; это напряжение мы будем называть *потенциалом плоских зон* $V_{пз}$ [1]. Если на металлический электрод подать более высокий (положительный) потенциал по сравнению с $V_{пз}$, то создаваемое электродом электрическое поле проникает в полупроводник и начинает притягивать к поверхности электроны и отталкивать дырки. Возникающее при этом изменение концентраций подвижных носителей экранирует электрическое поле и порождает в полупроводнике приповерхностный изгиб зон (рис. 4.2 *б*). Если исходный полупроводник имеет p-тип проводимости (как показано на рис. 4.2), то концентрация основных носителей вблизи поверхности уменьшается и говорят, что

---

[1]) В идеальном МОП-конденсаторе напряжение плоских зон определяется разностью работ выхода полупроводника и металла. Поэтому, подбирая металл и уровень легирования полупроводника, можно направленно изменять $V_{пз}$. В настоящее время при создании МОП-транзисторов вместо металла широко используется поликристаллический кремний (*поликремний*). Легируя его донорами или акцепторами, можно изменять его работу выхода в пределах 4,15–5,25 эВ. В реальном МОП-конденсаторе из-за наличия встроенного заряда в окисле и поверхностных состояний величина $V_{пз}$ может заметно отличаться от $V_{пз}$ в идеальном случае (см. подробнее текст, набранный мелким шрифтом на с. 247).

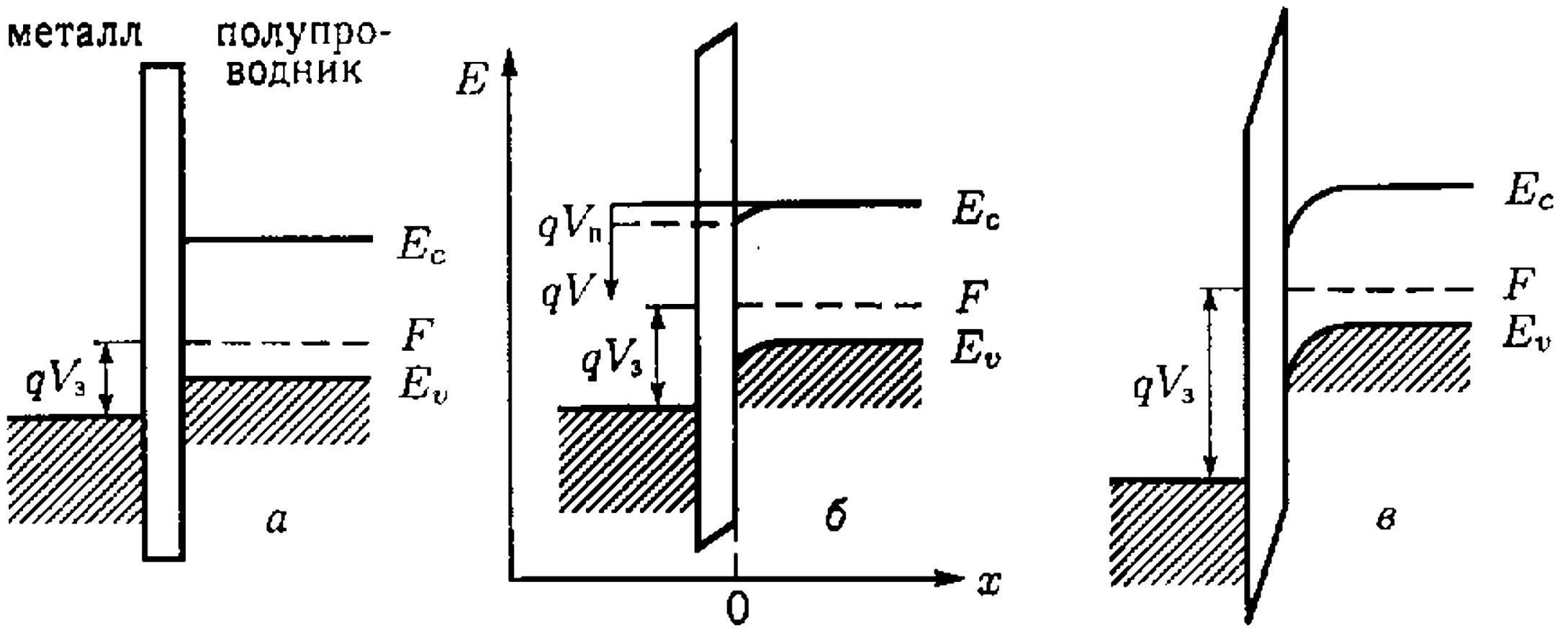


Рис. 4.2. Энергетическая диаграмма идеальной МОП-структуры в режиме плоских зон (*а*), обеднения (*б*) и инверсии (*в*)

происходит *обеднение* приповерхностного слоя. Проводимость полупроводника при этом уменьшается. Если продолжать увеличивать напряжение на металлическом электроде, то начиная с некоторого его значения концентрация электронов вблизи поверхности становится выше концентрации дырок и возникает *инверсионный слой* n-типа проводимости (рис. 4.2*в*). Этот режим называется *режимом инверсии*. При этом проводимость полупроводника вновь начинает возрастать с увеличением напряжения на электроде. Наконец, в случае, когда на электрод подан более низкий (отрицательный) потенциал по сравнению с $V_{пз}$, создаваемое электродом электрическое поле притягивает дырки и вблизи поверхности полупроводника образуется *обогащенный слой* p-типа; этот режим называется *режимом обогащения*. При этом проводимость полупроводника также возрастает.

В режиме инверсии принято различать *слабую* и *сильную* инверсии. Граница между ними отвечает случаю, когда концентрация электронов в инверсионном слое вблизи поверхности равна концентрации дырок в объеме полупроводника. Как мы увидим ниже, различие между этими режимами по сути определяется тем, какие заряды — подвижные или связанные — дают основной вклад в экранирование электрического поля.

Рассчитаем распределение электрического поля в МОП-конденсаторе. Для определенности будем рассматривать структуру, созданную на поверхности невырожденного полупроводника p-типа. Потенциал $V(x)$ в точке x определим как сдвиг положения краев зон вниз по отношению к их положению в объеме полупроводника (см. рис. 4.2*б*). Если напряженность электрического поля вблизи поверхности невелика и полупроводник

остается невырожденным, то локальные значения концентраций электронов и дырок будут определяться выражениями

$$n(x) = n_{p0} \exp\left(\frac{qV(x)}{kT}\right), \quad p(x) = p_{p0} \exp\left(-\frac{qV(x)}{kT}\right), \qquad (4.1)$$

где p_{p0} и n_{p0} — равновесные концентрации дырок и электронов в объеме полупроводника. Последние концентрации связаны с концентрацией акцепторной примеси N_a условием электронейтральности:

$$p_{p0} - n_{p0} = N_a.$$

Для нахождения распределения потенциала надо решить одномерное уравнение Пуассона:

$$-\frac{d^2V}{dx^2} = \frac{d\mathcal{E}}{dx} = \frac{4\pi\rho}{\epsilon} = \frac{4\pi q}{\epsilon}\left[p(x) - n(x) - N_a\right]. \qquad (4.2)$$

К сожалению, система уравнений (4.2) и (4.1) не допускает аналитического решения, однако можно связать напряженность электрического поля $\mathcal{E}$ в точке x с величиной потенциала $V(x)$ в той же точке. Поскольку

$$\frac{d\mathcal{E}}{dx} = \frac{d\mathcal{E}}{dV}\frac{dV}{dx} = -\mathcal{E}\frac{d\mathcal{E}}{dV}, \qquad (4.3)$$

уравнение (4.2) с учетом (4.1) легко преобразуется к виду

$$\mathcal{E}\frac{d\mathcal{E}}{dV} = -\frac{4\pi q}{\epsilon}\left[p - n - N_a\right] =$$

$$= \frac{4\pi q}{\epsilon}\left[n_{p0}(e^{qV/kT} - 1) - p_{p0}(e^{-qV/kT} - 1)\right]. \qquad (4.4)$$

Интегрируя это уравнение по V, находим:

$$\mathcal{E}^2 = \frac{8\pi q}{\epsilon}\int_0^V \left[n_{p0}(e^{qV'/kT} - 1) - p_{p0}(e^{-qV'/kT} - 1)\right] dV' =$$

$$= \frac{8\pi kT}{\epsilon}\Big[p_{p0}(e^{-qV/kT} +$$

$$+qV/kT - 1) + n_{p0}(e^{qV/kT} - qV/kT - 1)\Big]. \qquad (4.5)$$

Это уравнение связывает напряженность электрического поля в каждой точке полупроводника с потенциалом в этой точке и позволяет, в частности, связать напряженность электрического поля на поверхности полупроводника $\mathcal{E}_п = \mathcal{E}(0)$ с поверхностным потенциалом $V_п = V(0)$.

Найдем теперь величину напряжения на металлическом электроде, создающего заданный поверхностный потенциал в полупроводнике. Из условия непрерывности электрической индукции на границе раздела полупроводник–диэлектрик ($\mathcal{D}_{\text{п}} = \mathcal{D}_{\text{д}}$) следует, что напряженность электрического поля в диэлектрике равна $\mathcal{E}_{\text{д}} = \varepsilon\mathcal{E}_{\text{п}}/\varepsilon_{\text{д}}$. Это позволяет, зная толщину диэлектрика $d_{\text{д}}$, рассчитать искомое напряжение:

$$V_{\text{з}} = V_{\text{пз}} + V_{\text{п}} + \mathcal{E}_{\text{д}} d_{\text{д}} \equiv V_{\text{пз}} + V_{\text{п}} + \frac{\varepsilon}{4\pi}\frac{\mathcal{E}_{\text{п}}}{C_{\text{д}}}, \qquad (4.6)$$

где $C_{\text{д}} = \varepsilon_{\text{д}}/4\pi d_{\text{д}}$ — удельная емкость слоя диэлектрика, $V_{\text{пз}}$ — потенциал плоских зон, а $\mathcal{E}_{\text{п}}$ — напряженность электрического поля на поверхности полупроводника, которая связана с $V_{\text{п}}$ уравнением (4.5).

Наведенную электрическим полем поверхностную плотность заряда на металлическом электроде $Q_{\text{з}}$ и противоположную ей по знаку поверхностную плотность заряда в полупроводнике $Q_{\text{пп}}$ можно рассчитать с помощью теоремы Остроградского–Гаусса:

$$Q_{\text{з}} = -Q_{\text{пп}} = \frac{\varepsilon}{4\pi}\mathcal{E}_{\text{п}}. \qquad (4.7)$$

Зависимость $|Q_{\text{пп}}|$ от поверхностного потенциала $V_{\text{п}}$ показана на рис. 4.3. Величина $Q_{\text{пп}}$ складывается из поверхностной плотности находящегося в обедненном слое заряда ионизованных примесей $Q_{\text{обедн}}$ и поверхностной плотности подвижного заряда свободных носителей qn_{ss}:

$$Q_{\text{пп}} = Q_{\text{обедн}} + qn_{ss}.$$

В режимах обеднения и слабой инверсии, когда концентрации электронов и дырок в приповерхностном слое малы по сравнению с концентрацией акцепторов, электрическое поле, создаваемое зарядом на металлическом электроде, экранируется неподвижным зарядом ионизованных акцепторов. При этом пространственное распределение электрического поля и потенциала в области изгиба зон оказывается точно таким же, как и в барьере Шоттки (см. формулу (1.101) с заменой N_d на N_a), а индуцированный заряд $Q_{\text{пп}}$ изменяется пропорционально толщине обедненного слоя $d_{\text{обедн}}$ ($|Q_{\text{пп}}| \approx |Q_{\text{обедн}}| = qN_a d_{\text{обедн}} \sim \sim \sqrt{V_{\text{п}}}$, см. рис. 4.3). В режиме сильной инверсии концентрация электронов вблизи поверхности становится выше N_a и теперь уже основной вклад в экранирование дают свободные электроны. Из формул (4.5) и (4.7) следует, что в режиме сильной инверсии $|Q_{\text{пп}}| \approx |qn_{ss}| \sim \exp(qV_{\text{п}}/2kT)$. Такая же экспоненциальная

зависимость $Q_{\text{пп}}(|V_{\text{п}}|)$ получается и в режиме обогащения, когда электрическое поле экранируется свободными дырками.

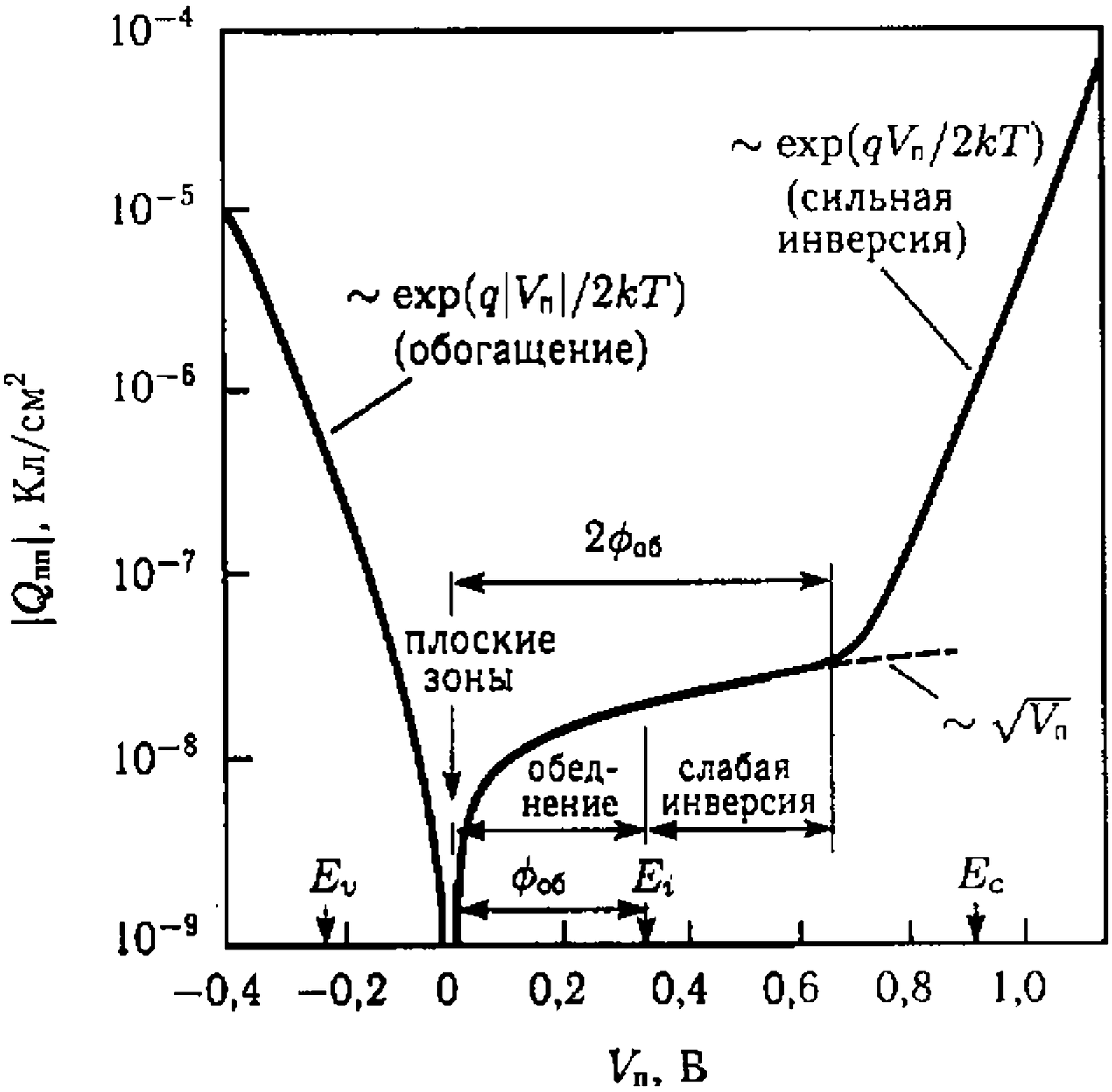


Рис. 4.3. Зависимость поверхностной плотности заряда в полупроводнике от поверхностного потенциала для p-Si с $N_a = 4 \cdot 10^{15}$ см$^{-3}$ при 300 К [14]

Важной характеристикой МОП-конденсатора является *пороговое напряжение*, которое определяется как напряжение на металлическом электроде, при котором происходит переход из режима слабой инверсии в режим сильной инверсии. Как мы говорили выше, условием этого перехода является равенство $n(0) = p_{p0}$. Из формул (4.1) находим, что поверхностный потенциал, отвечающий этому условию, равен

$$V_{\text{п}} = 2\phi_{\text{об}} = \frac{2kT}{q} \ln\left(\frac{p_{p0}}{n_i}\right),$$

а из формул (4.5) и (4.6) следует, что искомое пороговое напряжение равно

$$V_{\text{пор}} = V_{\text{пз}} + 2\phi_{\text{об}} + \frac{1}{C_{\text{д}}}\left(\frac{q\epsilon N_a \phi_{\text{об}}}{\pi}\right)^{1/2}. \qquad (4.8)$$

Важность понятия порогового напряжения состоит в том, что начиная с этого напряжения индуцированный металлическим

электродом заряд в полупроводнике начинает определяться подвижными носителями и, следовательно, в приповерхностном слое полупроводника появляется *проводящий канал*. Характерная толщина этого канала составляет 100 Å.

Вольт-фарадные характеристики МОП-конденсатора. Исследование *вольт-фарадных характеристик* является наиболее простым методом определения параметров МОП-структур (толщины диэлектрика, плотности встроенного заряда в нем, концентрации примеси в полупроводнике и т. д.). Из формул (4.6) и (4.7) следует, что удельную емкость МОП-конденсатора $dQ_з/dV_з$ можно представить в виде последовательно соединенных удельной емкости полупроводника,

$$C_{пп} = dQ_з/dV_п = -dQ_{пп}/dV_п,$$

и удельной емкости диэлектрика

$$C_д = \epsilon_д/(4\pi d_д).$$

Как мы показали выше, индуцированный заряд состоит из двух составляющих (заряда ионизованных примесей в обедненном слое и наведенного заряда подвижных носителей), поэтому и емкость полупроводника $C_{пп}$ включает в себя два слагаемых: емкость обедненного слоя $C_{обедн} = -dQ_{обедн}/dV_п$ и емкость $C_{навед} = -d(qn_{ss})/dV_п$, связанную с подвижными носителями.

В режимах обеднения и слабой инверсии основной вклад в $dQ_{пп}$ дает заряд в обедненном слое и, следовательно, на соответствующих участках вольт-фарадной характеристики $C_{пп} \approx \approx C_{обедн}$. В режимах обогащения и сильной инверсии в $dQ_{пп}$ преобладает вклад подвижных носителей, и на этих участках емкость полупроводника определяется уже ими ($C_{пп} \approx C_{навед}$).

Удельную емкость обедненного слоя можно легко рассчитать, зная его толщину. Так, в практически важном случае начала сильной инверсии ($V_п = 2\phi_{об}$)

$$C_{обедн} = \frac{\varepsilon}{4\pi d_{обедн}}, \quad d_{обедн} \approx \sqrt{\frac{\varepsilon\phi_{об}}{\pi q N_a}}. \qquad (4.9)$$

Эквивалентная схема МОП-конденсатора показана на рис. 4.4. Поскольку в режиме инверсии изменение концентрации наведенных подвижных зарядов вблизи границы полупроводник–диэлектрик при изменении напряжения на металлическом электроде происходит за счет тепловой генерации *неосновных* носителей в объеме полупроводника (то есть медленно),

то емкость $C_{\text{навед}}$ включена в эквивалентной схеме через сопротивление $R_{\text{ген}}$ достаточно большой величины. Сопротивление R_s учитывает сопротивление толщи полупроводника. Из этой эквивалентной схемы следует, что вольт-фарадные характеристики МОП-конденсатора в режиме сильной инверсии должны сильно различаться на низких и высоких частотах, поскольку на высокой частоте ($\omega C_{\text{навед}} R_{\text{ген}} \gg 1$) емкость $C_{\text{навед}}$ не успевает перезаряжаться за период колебаний. Эксперимент (рис. 4.5) полностью подтверждает этот вывод.

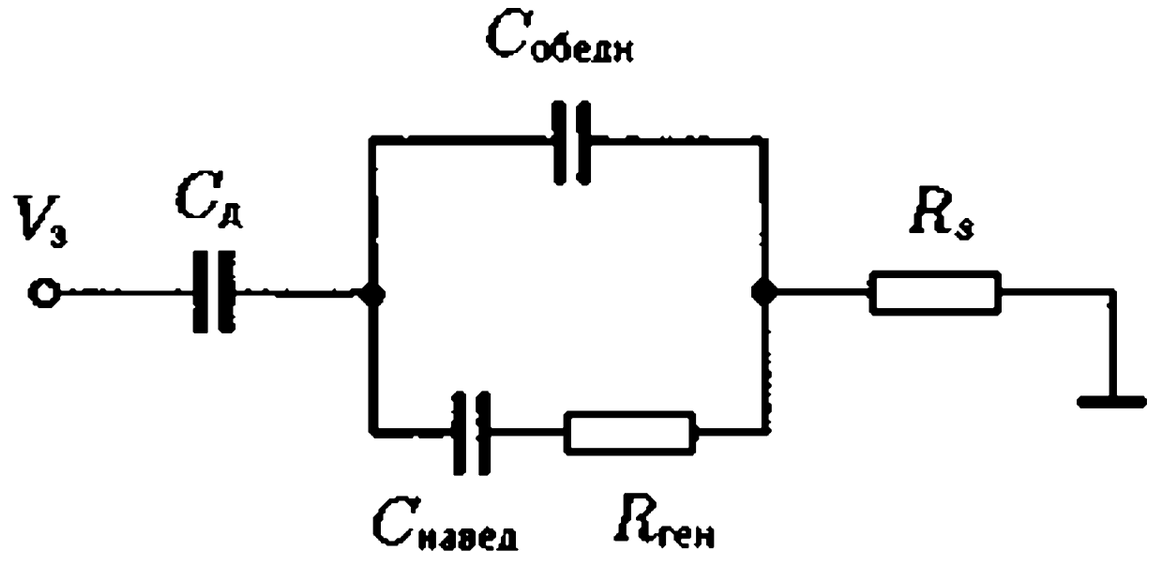


Рис. 4.4. Эквивалентная схема МОП-конденсатора

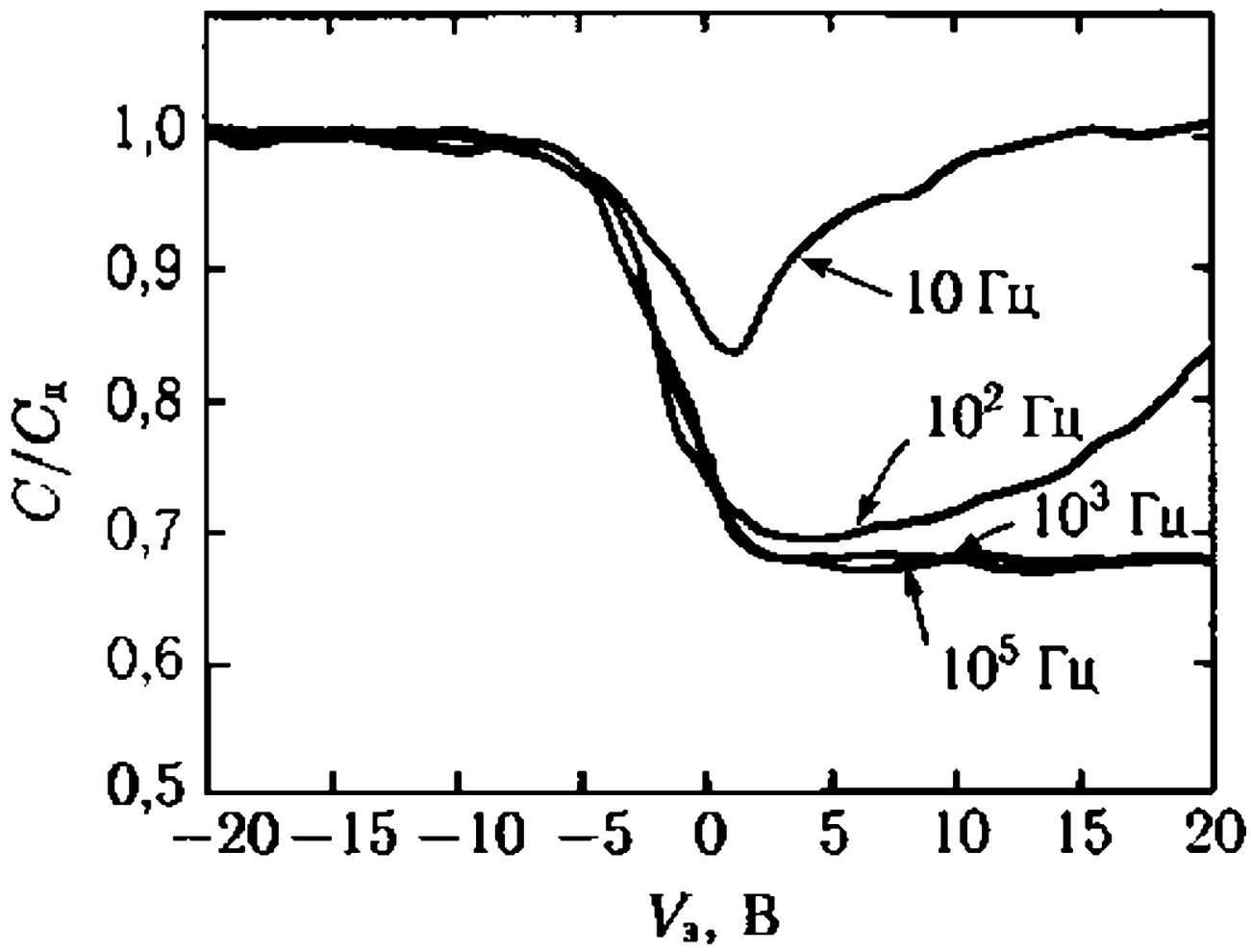


Рис. 4.5. Вольт-фарадные характеристики МОП-конденсатора на различных частотах [87]

Анализ вольт-фарадных характеристик МОП-структур (см. подробнее [87]) позволяет: 1) из значения емкости в режиме обогащения определить толщину диэлектрика; 2) из значения емкости обедненного слоя, выделенной из измеренной на высокой частоте емкости в режиме сильной инверсии, найти концентрацию примеси в полупроводнике; 3)по смещению кривых вдоль оси $V_з$ найти потенциал плоских зон, из которого можно оценить плотность встроенного заряда в диэлектрике.

Проведенный выше анализ относился к идеальной МОП-структуре. Эксперимент показывает, что в реальных структурах существуют дополнительные особенности, которые необходимо учитывать при конструировании полевых транзисторов. Прежде всего это *встроенный заряд в окисле*, который сильно влияет на потенциал плоских зон.

На границе раздела $Si–SiO_2$, полученной термическим окислением кремния, всегда присутствуют четыре (!) различных по своей природе источника заряда [14, 58]. Это — заряд быстрых поверхностных состояний в полупроводнике, постоянный заряд в окисле, заряд на ловушках в слое окисла и заряд подвижных ионов. Энергетическая плотность *быстрых поверхностных состояний*, которые расположены непосредственно на границе $Si–SiO_x$, сильно зависит от режима получения окисла и ориентации поверхности полупроводника; ее удается снизить до $\sim 10^{10}$ см$^{-2}$эВ$^{-1}$ отжигом в атмосфере форминг-газа (смеси 10% H_2+90% N_2) при 450 °С. Поскольку заряд поверхностных состояний меняется при изменении поверхностного потенциала и экранирует внешнее электрическое поле, эти состояния уменьшают величину индуцированного заряда в инверсионном слое и ухудшают характеристики МОП-транзисторов. Кроме того, поверхностные состояния являются источником поверхностного шума в приборах с зарядовой связью.

Постоянный заряд в окисле (обычно положительный) имеет поверхностную плотность 10^{10}–10^{12} см$^{-2}$ и располагается в слое толщиной $\sim$30 Å вблизи границы раздела $Si–SiO_x$. Величина этого заряда зависит от режима окисления, условий отжига и ориентации подложки (наименьшую плотность постоянного заряда имеет кремний с ориентацией <100>).

Заряд на ловушках в слое окисла представляет собой объемный заряд, захваченный на энергетические уровни дефектов в SiO_2; его поверхностная плотность может меняться в пределах 10^9–10^{13} см$^{-2}$. Этот тип заряда ассоциируется с медленными поверхностными состояниями, которые являются одной из главных причин возникновения фликкер-шума. Заряд на ловушках в слое окисла «отжигается» в ходе низкотемпературной термообработки структуры, но может появляться при облучении МОП-структуры (например, при имплантации ионов через тонкий слой подзатворного диэлектрика с целью коррекции порогового напряжения МОП-транзистора) или при лавинной инжекции заряда в ЛИЗМОП-структурах (см. с. 286).

Наконец, *заряд подвижных ионов* связан с присутствием в окисле ионов щелочных металлов (Na^+, K^+, Li^+) и ионов тяжелых металлов, а также отрицательно заряженных ионов (OH^-), попадающих в окисел из окружающей среды и материалов, используемых в технологическом процессе. Поверхностная плотность этих зарядов обычно лежит в пределах 10^{10}–10^{12} см$^{-2}$.

Ионы щелочных металлов (прежде всего, ионы натрия) подвижны в SiO_2 уже при 100 °С и под действием электрического поля могут перемещаться в диэлектрике, вызывая медленный дрейф порогового напряжения. Для связывания эти ионов сотрудники IBM в 1964 г. предложили обрабатывать окисел в P_2O_5; образующееся при последующем нагревании фосфорно-силикатное стекло является хорошим геттером натрия и прочно связывает его ионы. Позже был предложен еще один способ связывания ионов Na — окисление кремния в присутствии хлорсодержащих газов (например, HCl); при этом попавший в окисел хлор прочно связывает ионы натрия в форме NaCl.

4.1.2. Вольт-амперная характеристика МОП-транзистора. Конструкция кремниевого полевого транзистора с изолированным затвором показана на рис. 4.6. На подложке p-типа проводимости диффузией или ионной имплантацией сначала создают две области n^+-типа, которые будут служить *истоком* и *стоком* полевого транзистора. После этого на поверхности кремния создается тонкий (толщиной $d_д = 15$–1200 Å) изолирующий слой из собственного окисла (в транзисторах с МОП-структурой) или другого диэлектрика (в транзисторах с МДП-структурой), на который затем наносится проводящий (металлический или поликремниевый) электрод — *затвор*.

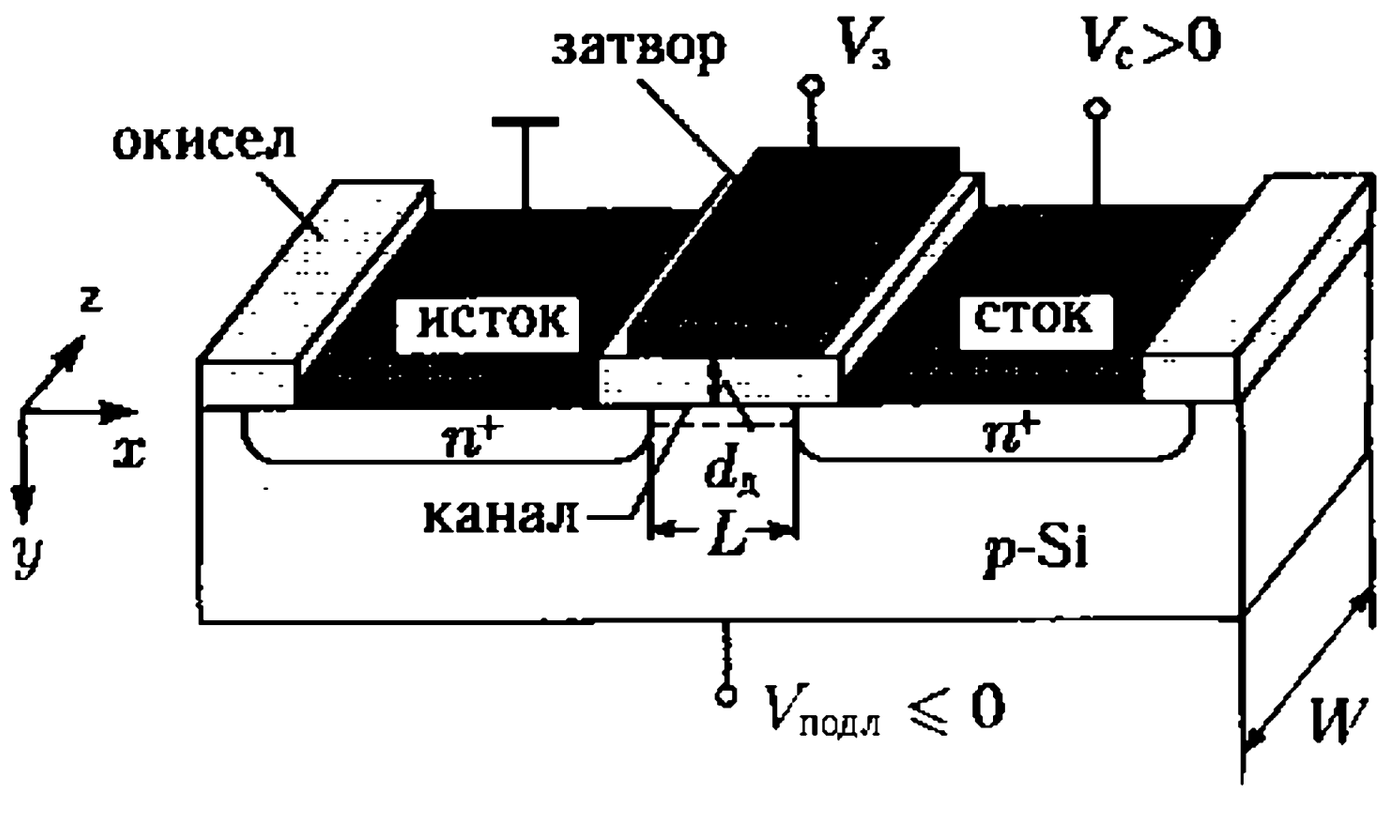


Рис. 4.6. Конструкция полевого транзистора с изолированным затвором и структурой металл-окисел-полупроводник

Если на затвор подать положительное напряжение, величина которого выше порогового, то под затвором в полупроводнике наводится отрицательный заряд подвижных электронов, которые образуют *проводящий канал* n-типа между истоком и стоком. Если при этом между истоком и стоком приложено напряжение, то в канале начинает течь ток. Величина этого тока определяется концентрацией индуцированных в канале носителей, которая, в свою очередь, зависит от напряжения на затворе. Обедненные области между областями истока, стока и канала, с одной стороны, и подложкой p-типа, с другой стороны, обеспечивают изоляцию транзистора от других элементов, созданных на той же подложке (изоляция с помощью p–n-перехода, см. с. 207).

Рассчитаем теперь зависимость тока стока полевого транзистора от напряжений на затворе и стоке.

В отличие от МОП-конденсатора, из-за разности потенциалов, приложенной между истоком и стоком, потенциал в канале полевого транзистора $V_{кан}$ меняется по длине и, поскольку потенциал затвора фиксирован, локальная плотность индуцированного заряда (и, следовательно, проводимость) в разных точках

канала оказывается различной. Если считать, что продольная компонента напряженности электрического поля в канале $\mathcal{E}_x$, создаваемая разностью потенциалов между истоком и стоком, много меньше поперечной компоненты напряженности поля $\mathcal{E}_y$, создаваемой затвором, то искривлением линий электрического поля под затвором можно пренебречь, и плотность наведенно в данной точке заряда определяется разностью потенциалов между этой точкой канала и затвором. Это приближение получило название *приближения плавного канала*.

Будем считать, что напряжение на затворе $V_з$ настолько велико, что во всех точках канала инверсия оказывается сильной. Положим потенциал истока равным нулю ($V_{кан}(0) = 0$), а потенциал на стоке равным $V_с = V_{кан}(L) > 0$. Кроме этого, будем считать, что подложка транзистора соединена с истоком ($V_{подл} = 0$). Учитывая, что появлению проводящего канала отвечает условие $V_п \approx 2\phi_{об}$, из уравнений (4.6) и (4.7) находим локальную плотность поверхностного заряда в полупроводнике:

$$Q_{пп}(x) \approx -C_д[V_з - V_{пз} - 2\phi_{об} - V_{кан}(x)]. \qquad (4.10)$$

Как мы отмечали выше, этот заряд состоит из заряда обедненного слоя и заряда подвижных носителей: $Q_{пп} = Q_{обедн} + qn_{ss}$. Поверхностная плотность заряда обедненного слоя легко находится из решения уравнения Пуассона (4.2) и равна

$$Q_{обедн}(x) = -qN_a d_{обедн}(x) = -\sqrt{\frac{\varepsilon q N_a [2\phi_{об} + V_{кан}(x) - V_{подл}]}{2\pi}}. \qquad (4.11)$$

Это позволяет нам найти изменение поверхностной плотности подвижного заряда qn_{ss} вдоль канала.

Расчет вольт-амперной характеристики заметно упрощается, если пренебречь изменением величины $Q_{обедн}$ вдоль канала (см. уравнение (4.11)) и считать эту величину постоянной. Это приближение получило название *модели управления зарядом*. В этом приближении локальная плотность подвижного заряда связана с напряжением в данной точке канала соотношением

$$qn_{ss}(x) \approx -C_д[V_з - V_{пор} - V_{кан}(x)], \qquad (4.12)$$

где $V_{пор} = V_{пз} + 2\phi_{об} - \overline{Q}_{обедн}/C_д$ есть некоторое эффективное пороговое напряжение, зависящее от среднего значения $\overline{Q}_{обедн}$ [1].

---

[1]) Из-за зависимости qn_{ss} от $Q_{обедн}$ МОП-транзистор оказывается не трехэлектродным, а *четырехэлектродным* прибором, которым можно управлять

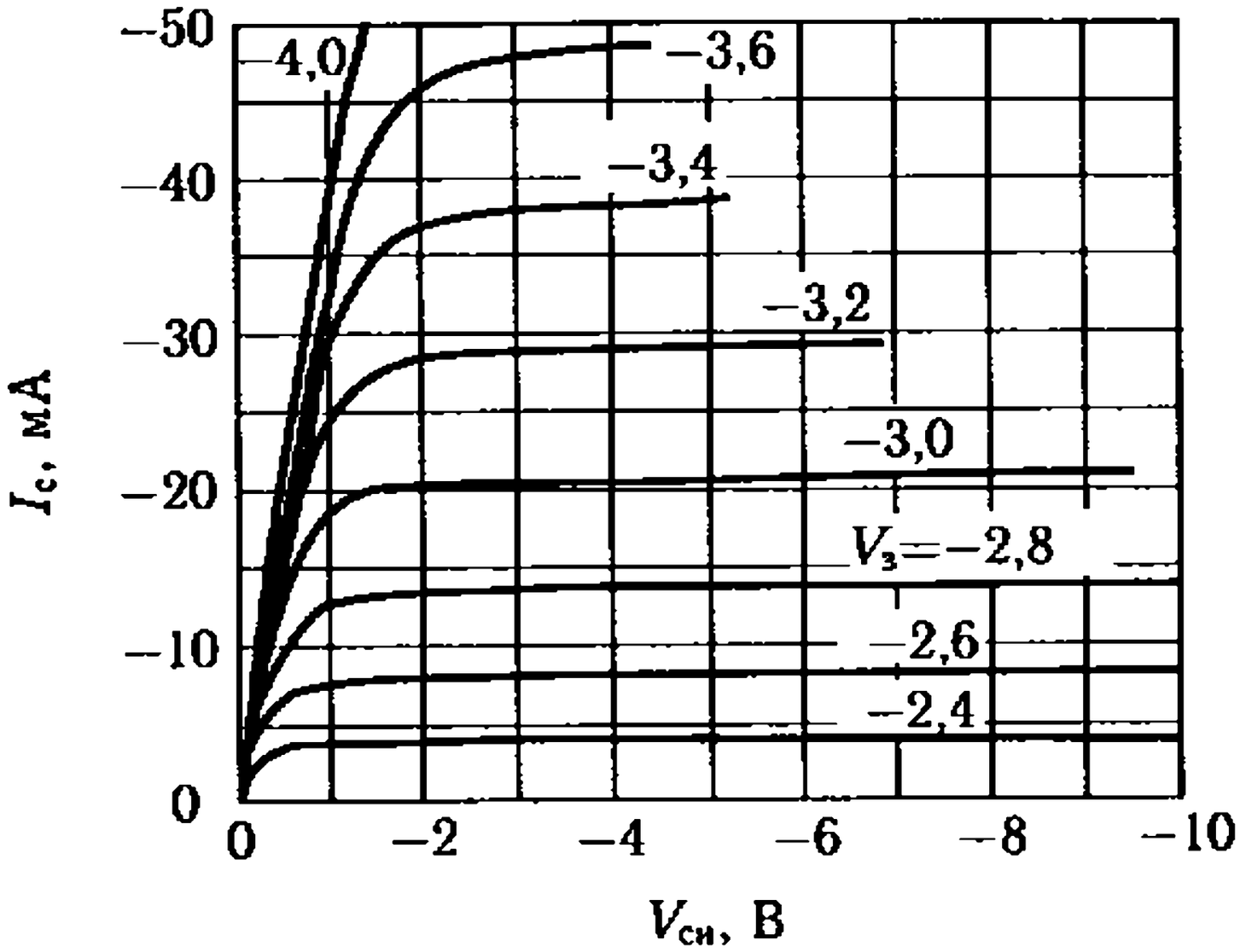


Рис. 4.7. Выходные характеристики промышленного МОП-транзистора 2SJ343 фирмы Toshiba с каналом p-типа

Если считать, что подвижность носителей μ_n не зависит от напряженности электрического поля, то ток в канале шириной W можно записать следующим образом [1]):

$$I_{\text{с}} = |qn_{ss}|\mu_n W \frac{dV_{\text{кан}}}{dx} = \mu_n W C_{\text{д}}[V_{\text{з}} - V_{\text{пор}} - V_{\text{кан}}(x)]\frac{dV_{\text{кан}}}{dx}. \tag{4.13}$$

После разделения переменных это уравнение можно переписать так:

$$dx = \frac{1}{I_{\text{с}}}[\mu_n W C_{\text{д}}(V_{\text{з}} - V_{\text{пор}} - V_{\text{кан}})]\, dV_{\text{кан}}. \tag{4.14}$$

Интегрируя это уравнение по всей длине канала (от $x = 0$ до L), находим:

$$I_{\text{с}} = \mu_n \frac{W}{L} C_{\text{д}} \left[(V_{\text{з}} - V_{\text{пор}})V_{\text{с}} - \frac{V_{\text{с}}^2}{2}\right]. \tag{4.15}$$

Это решение остается справедливым, пока во всех точках канала $|qn_{ss}| > 0$. При $V_{\text{с}} > V_{\text{с.нас}} = V_{\text{з}} - V_{\text{пор}}$, когда, в соответствии с формулой (4.12), в самой «узкой» части канала около стока

и со стороны подложки. Так, подавая отрицательное смещение на подложку, мы увеличиваем заряд $|Q_{\text{обедн}}|$ и, следовательно, уменьшаем $|qn_{ss}|$ и ток в транзисторе.

[1]) В этом уравнении мы считаем ток чисто дрейфовым и пренебрегаем диффузионной составляющей тока, которая может возникать из-за неоднородного распределения в канале поверхностной концентрации подвижных носителей. Оценки показывают, что это приближение является хорошим, пока канал ни в одной точке не перекрывается.

$|qn_{ss}|$ уменьшается до нуля (происходит *отсечка канала*) [1], транзистор переходит в режим *насыщения* и ток в канале стабилизируется на уровне

$$I_{\text{с}} = \frac{1}{2}\,\mu_n \frac{W}{L} C_{\text{д}} (V_{\text{з}} - V_{\text{пор}})^2. \tag{4.16}$$

Выходные вольт-амперные характеристики МОП-транзистора, представленные в виде семейства кривых $I_{\text{с}}(V_{\text{с}})$ для ряда фиксированных значений $V_{\text{з}}$, показаны на рис. 4.7.

Одной из важных для практики характеристик полевого транзистора является *крутизна вольт-амперной характеристики*, определяемая как производная $g_m = dI_{\text{с}}/dV_{\text{з}}$ при фиксированном напряжении на стоке. Из уравнений (4.15), (4.16) легко находим, что

$$g_m = \begin{cases} \mu_n (W/L) C_{\text{д}} V_{\text{с}}, & V_{\text{с}} < V_{\text{с.нас}}, \\ \mu_n (W/L) C_{\text{д}} (V_{\text{з}} - V_{\text{пор}}), & V_{\text{с}} > V_{\text{с.нас}}. \end{cases} \tag{4.17}$$

Из этой формулы следует, что для достижения максимальной крутизны полевые транзисторы следует изготавливать из полупроводников с высокой подвижностью, делать канал как можно короче и шире, а слой диэлектрика — как можно тоньше. Эти требования учтены в конструкциях современных МОП-транзисторов: большинство из них имеют канал n-типа проводимости, длина канала уменьшилась от первоначальных ~50 мкм до ~0,1 мкм, а толщина диэлектрика — от 1200 Å до 15–50 Å. Кроме того, из формулы (4.17) следует, что в области насыщения для получения высокой крутизны следует также увеличивать $V_{\text{з}} - V_{\text{пор}}$, то есть желательно, чтобы транзистор работал в области больших токов.

4.1.3. Особенности реальных полевых транзисторов. В этом разделе мы рассмотрим некоторые особенности реальных полевых транзисторов, которые необходимо иметь в виду при разработке этих приборов.

Существование в канале сильного поперечного электрического поля, прижимающего носители к границе раздела $Si–SiO_2$, приводит к *дополнительному рассеянию носителей*. Эксперимент показывает, что это рассеяние возрастает с увеличением

---

[1]) В этом случае уравнение (4.13) становится неприменимым, поскольку в области стока, где $qn_{ss} \approx 0$, нарушается приближение плавного канала и для нахождения распределения электрического поля необходимо решать двумерное уравнение Пуассона. Кроме этого, в области насыщения основным механизмом протекания тока становится *диффузия* носителей из канала в область пространственного заряда.

напряженности электрического поля и может приводить к 2–5-кратному уменьшению подвижности носителей в инверсионном слое по сравнению с объемным материалом (рис. 4.8).

Физическими причинами дополнительного рассеяния являются [156, 173]: 1) рассеяние на объемных фононах в условиях, когда характер движения носителей в узком канале по сути становится квазидвумерным; 2) рассеяние на поверхностных акустических модах (волнах Рэлея); 3) рассеяние на шероховатостях границы раздела.

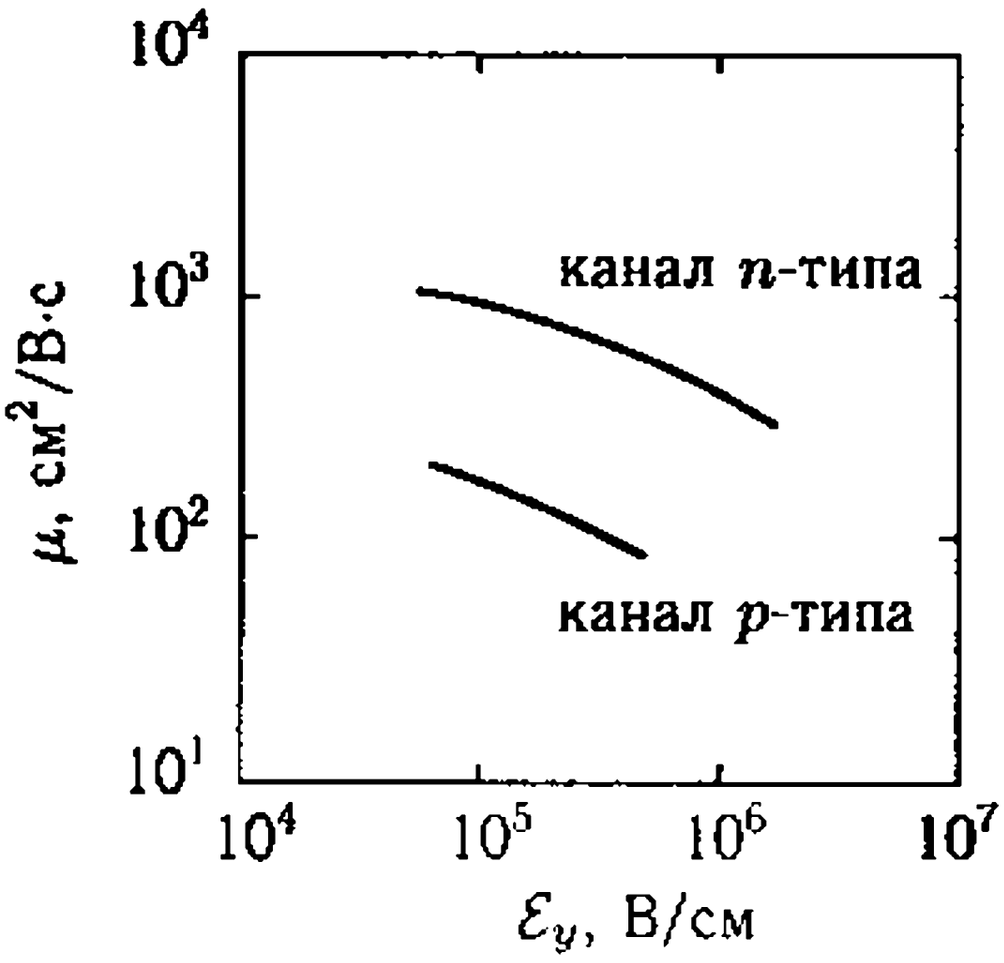


Рис. 4.8. Зависимость подвижности электронов и дырок в инверсионных слоях кремния от напряженности поперечного электрического поля под затвором [87]

Следующим очень важным для современных полевых транзисторов эффектом является *эффект насыщения скорости дрейфа*. Как мы уже отмечали в п. 2.4, с ростом напряженности электрического поля скорость дрейфа носителей отклоняется от закона $v = \mu\mathcal{E}$, справедливого в слабых полях, и насыщается на уровне $v_s \sim 10^7$ см/с (см. рис. 4.9). Причиной этого является сильное увеличение темпа рассеяния электронов на оптических фононах с ростом энергии электронов, разогреваемых электрическим полем. Поэтому часть канала, в которой напряженность поля оказывается выше критической ($\mathcal{E}_x > \mathcal{E}_s = v_s/\mu \sim 10^4$ В/см в случае электронов в Si), становится «узким местом» для движения электронов, что проявляется в заниженных значениях тока стока и крутизны вольт-амперной характеристики транзистора. Нетрудно видеть, что при указанном выше критическом поле и напряжении на стоке $V_c \approx \approx 5$ В эффект насыщения скорости дрейфа начинает проявляться в приборах с длиной канала $L \leqslant 5$ мкм.

Учет эффекта насыщения скорости дрейфа изменяет рассчитанную выше вольт-амперную характеристику МОП-транзистора. Более корректную характеристику можно получить из того же уравнения (4.13), но учтя, что переход в область насыщения происходит теперь при условии, когда электрическое поле в области стока достигает значения $\mathcal{E}_s$. Несложные расчеты (см. [87]) приводят к следующей зависимости тока стока

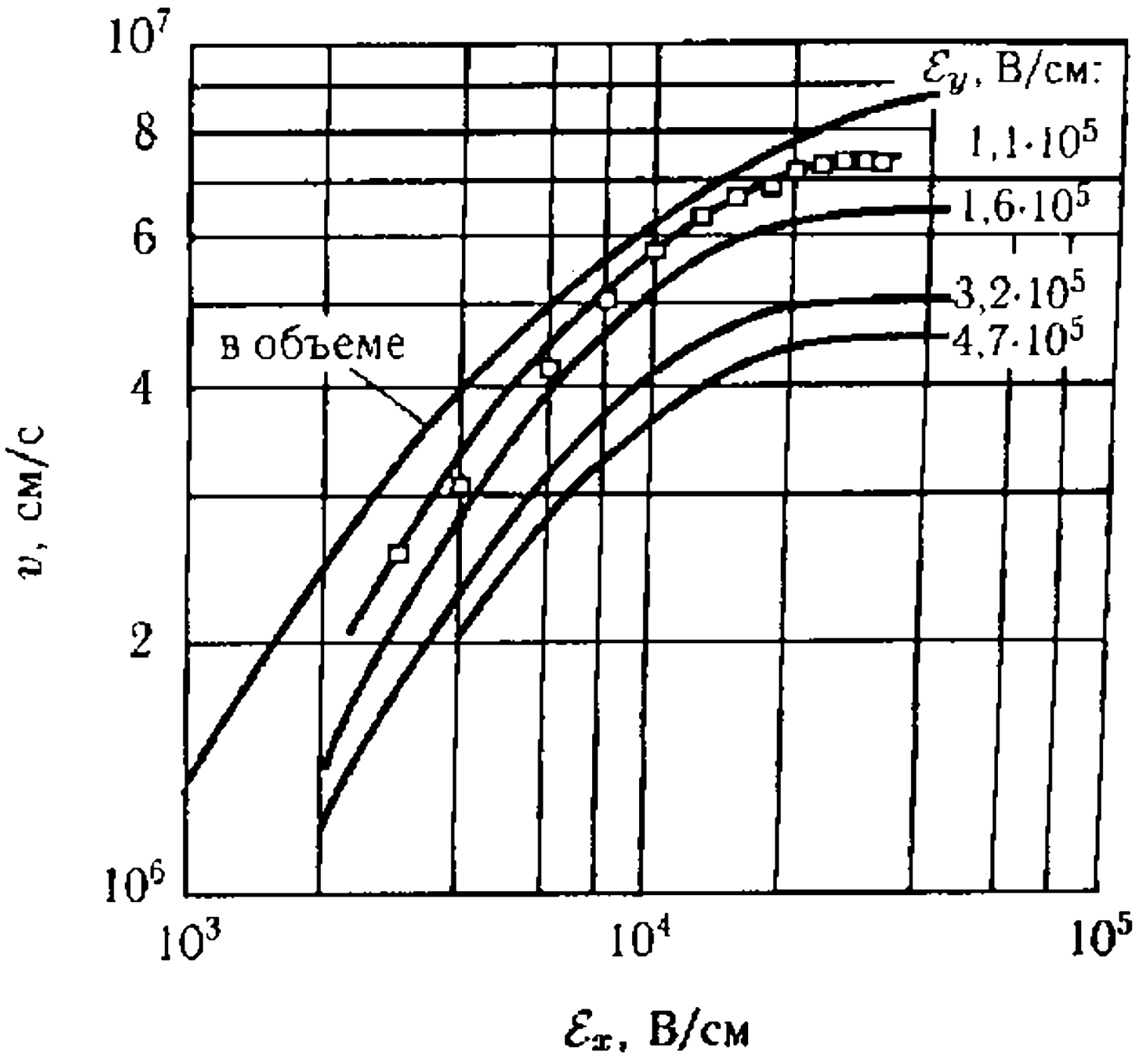


Рис. 4.9. Зависимость скорости дрейфа электронов от напряженности продольного электрического поля $\mathcal{E}_x \parallel \langle 100 \rangle$ в объемном кремнии и в канале МОП-транзистора при нескольких значениях напряженности поперечного электрического поля $\mathcal{E}_y$. $T = 300$ К [14]

в области насыщения от напряжения на затворе:

$$I_{\text{с.нас}} = WC_{\text{д}} v_s (V_{\text{з}} - V_{\text{пор}}).$$

Из этой формулы следует, что соответствующая крутизна вольт-амперной характеристики $g_m = WC_{\text{д}} v_s$ не зависит от длины канала и тока стока. Поэтому в современной литературе для оценки достигнутых параметров полевых транзисторов часто используют величину удельной крутизны — отношение g_m/W.

Стремление повысить крутизну вольт-амперной характеристики, которая изменяется пропорционально $C_{\text{д}} = \varepsilon_{\text{д}}/(4\pi d_{\text{д}})$, заставляет разработчиков использовать все более тонкие слои подзатворного диэлектрика. Однако с уменьшением толщины окисла быстро возрастает ток утечки, связанный с туннелированием электронов сквозь тонкий слой диэлектрика [174, 175]. Допустимая плотность тока утечки затвора различна для разных областей применения МОП-транзисторов: при их использовании в цифровых ИС можно довольно терпимо относиться к плотности тока утечки 1–10 А/см^2 (!), тогда как при их использовании в аналоговых ИС и микросхемах динамических запоминающих устройств плотность тока утечки не должна превышать 10^{-8}–10^{-6} А/см^2. Поэтому в настоящее время возможность улучшения

характеристик МОП-транзисторов за счет уменьшения толщины SiO_2 или часто используемого вместо него SiO_xN_y [176, 177] практически исчерпана, и ведется активный поиск других диэлектрических материалов. Уже найден ряд материалов с *высокой диэлектрической проницаемостью*, в которых при том же значении $C_д$ за счет увеличения толщины диэлектрика удается уменьшить туннельный ток утечки на 4–6 порядков. Некоторые физические свойства этих материалов приведены в табл. 1 в Приложении. Наиболее перспективными диэлектриками среди них считаются соединения гафния (HfO_2 и $HfSiO_4$) [178].

К сожалению, многие материалы с высокой диэлектрической проницаемостью склонны к химическому взаимодействию с кремнием и материалом затвора. Для предотвращения этого между кремниевой подложкой и диэлектрическим слоем помещают барьерный слой толщиной 3–7 Å из оксида или нитрида кремния, а затвор стараются сделать из металла или химически инертных TiN или TaSiN.

4.1.4. Полевые транзисторы с коротким каналом. По мере уменьшения размеров элементов МОП-транзисторов, используемых в интегральных схемах, было обнаружено, что свойства транзисторов с каналом короче некоторой критической длины начинают сильно отличаться от свойств транзисторов с длинным каналом. Эти отклонения — *эффекты короткого канала* — связаны с существенно двумерным характером распределения электрических полей в структуре и высокими напряженностями этих полей. Исследования установили следующую эмпирическую зависимость для длины канала, начиная с которой характеристики транзистора начинают сильно изменяться:

$$L_{\min} = 0,4[r_j d_д (w_и + w_с)^2]^{1/3} \text{ [мкм]}, \qquad (4.18)$$

где r_j — глубина истокового и стокового p–n-переходов, мкм, $d_д$ — толщина подзатворного окисла, Å, $w_и$ и $w_с$ — толщина обедненных областей истока и стока, мкм [14]. Эффекты короткого канала являются серьезным препятствием на пути дальнейшего уменьшения размеров МОП-транзисторов, и поэтому требуют подробного обсуждения.

Одним из проявлений эффектов короткого канала является исчезновение области насыщения тока стока на вольт-амперной характеристике при $V_с > V_з - V_{пор}$ (см. рис. 4.10). Существует несколько причин, приводящих к этому. Наиболее важная из них — проявление *эффекта модуляции длины канала.*

В транзисторах с коротким каналом, в которых толщина областей пространственного заряда вокруг истока и стока ($w_и$, $w_с$) сравнима с геометрической длиной канала L (см. рис. 4.11), увеличение напряжения на стоке вызывает уменьшение эффективной длины проводящего канала, $L' \approx \approx L - w_и - w_с$, что очевидно приводит к возрастанию протекающего через него тока (см. рис. 4.10). Кроме того, с увеличением $V_с$ происходит эффективное уменьшение заряда обедненного слоя (из-за уменьшения площади WL', занимаемой этим слоем), что при фиксированном $V_з$ вызывает возрастание поверхностной плотности подвижных носителей n_{ss} и дополнительно увеличивает ток стока (то есть по сути увеличение $V_с$ приводит к уменьшению порогового напряжения). Очевидно, что для ослабления эффекта модуляции длины канала необходимо уменьшать толщину областей пространственного заряда вокруг истока и стока, для чего подложку МОП-транзисторов следует легировать сильнее.

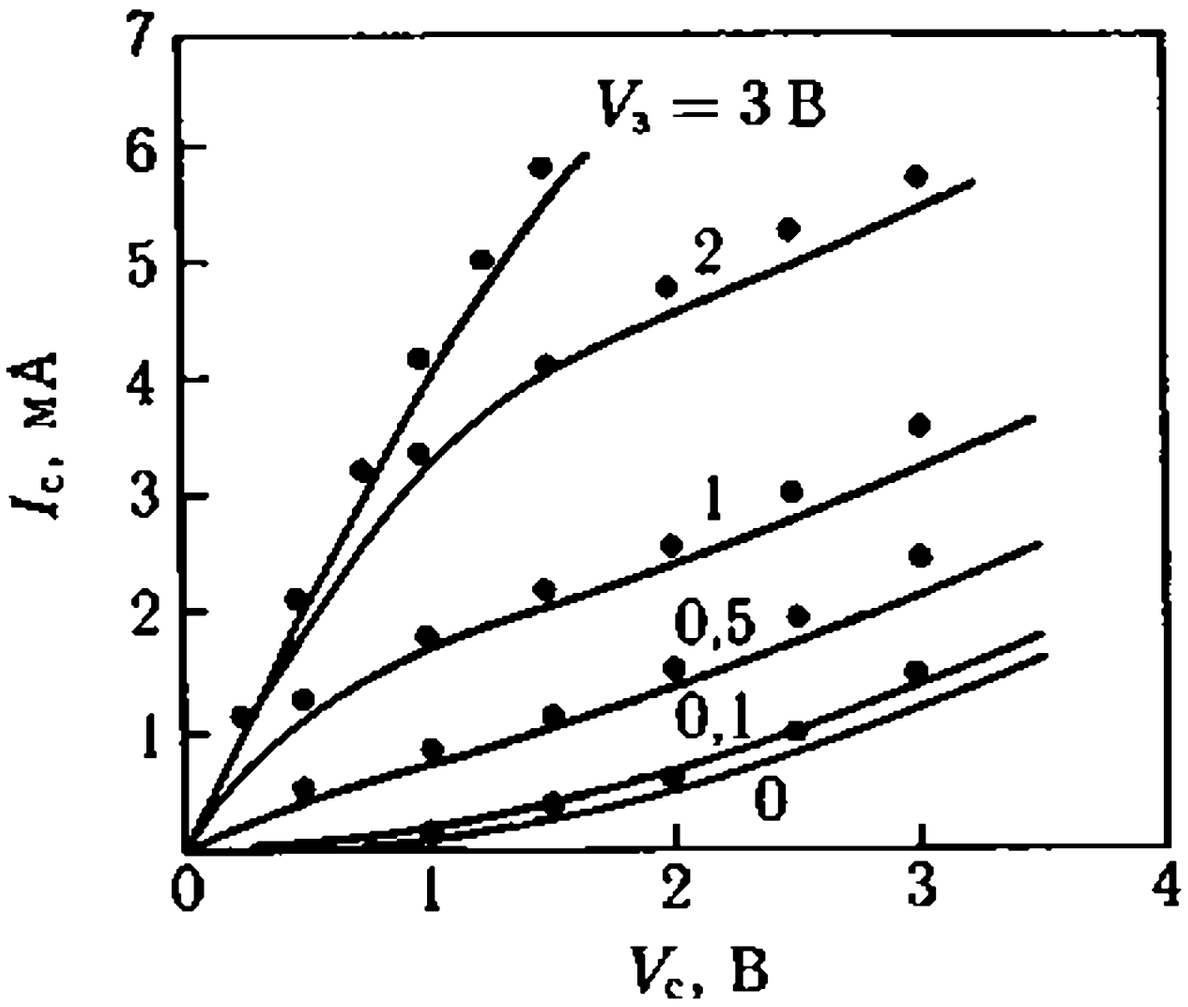


Рис. 4.10. Выходные (стоковые) характеристики МОП-транзистора с коротким каналом. Параметры структуры: $L = 0,23$ мкм, $W = 30$ мкм, $d_д = 258$ Å; подложка соединена со стоком [14]

Другим эффектом, проявляющимся в транзисторах с коротким каналом, является изменение вольт-амперных характеристик в области *подпороговых* токов, то есть в области напряжений $V_з < V_{пор}$. В этих условиях инверсия в канале является слабой, а зависимость энергии края зоны проводимости от координаты вдоль канала напоминает энергетическую диаграмму биполярного транзистора. При этом ток стока определяется высотой потенциального барьера, который носителям необходимо преодолеть, чтобы попасть в канал. В транзисторах с длинным каналом подпороговый ток практически не зависит от напряжения на стоке, а его зависимость от напряжения на затворе имеет следующий вид:

$$J_с \sim \exp\left(\frac{qV_з}{mkT}\right), \quad \text{где } m \approx 1 + \frac{C_{обедн}}{C_д} = 1 + \frac{\varepsilon d_д}{\varepsilon_д d_{обедн}}, \quad (4.19)$$

а $d_{\text{обедн}}$ — толщина обедненного слоя под затвором [14]. В транзисторах с коротким каналом ситуация качественно меняется. Из-за того, что распределение потенциала в таких транзисторах является двумерным, напряжение на стоке начинает заметно влиять на высоту указанного выше потенциального барьера и подпороговый ток транзистора начинает сильно зависеть от напряжения на стоке.

Важность области подпороговых токов для современных цифровых ИС связана с тем, что с уменьшением размеров транзисторов напряжение питания микросхем и разность напряжений, отвечающих уровням логического 0 и логической 1, также уменьшается. Это приводит к уменьшению отношения токов МОП-транзистора в открытом и закрытом состояниях. Поскольку ток транзистора в открытом состоянии должен оставаться большим, то с уменьшением размеров транзисторов ток закрытых транзисторов становится все больше и больше, что приводит к заметному росту потребления энергии микросхемами даже в неактивном состоянии. Именно подпороговый ток транзисторов становится тем фактором, который препятствует дальнейшему понижению рабочих напряжений цифровых ИС (современные процессоры работают при напряжении 1,3–1,5 В). По оценкам специалистов фирмы Intel, при рабочем напряжении микросхем ниже 1 В примерно половина рассеиваемой ими мощности может быть связана с токами утечки закрытых транзисторов [179]. В связи со сказанным следует добавить, что при уменьшении размеров МОП-транзисторов для сохранения высокого отношения токов транзисторов в открытом и закрытом состоянии значение параметра m в формуле (4.19) должно оставаться близким к 1, то есть соразмерно с уменьшением длины канала и увеличением уровня легирования подложки должна уменьшаться и толщина диэлектрика. При этом, как мы отмечали на с. 254, сильно возрастает туннельный ток утечки затвора.

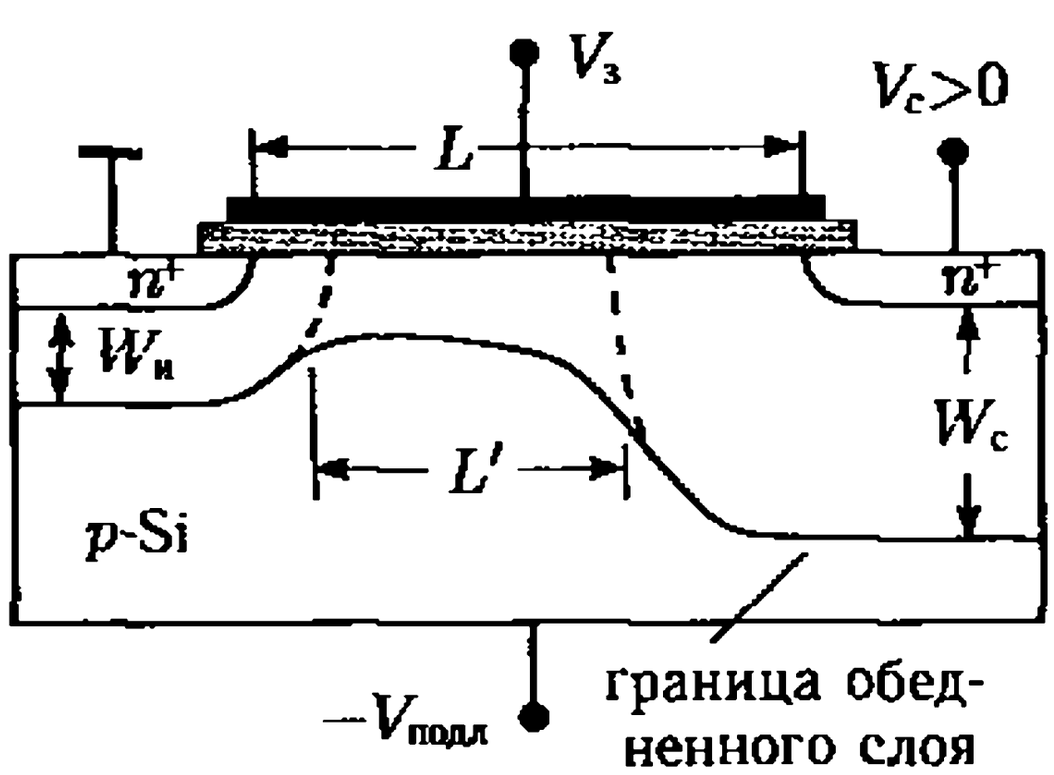


Рис. 4.11. Распределение заряда в обедненном слое полевого транзистора с коротким каналом

К сожалению, простое увеличение концентрации примеси в подложке, с целью ослабления эффектов короткого канала, приводит к трем нежелательным эффектам: увеличению емкостей,

появлению туннельного пробоя перехода сток–подложка и возрастанию напряженности электрического поля под затвором, которое понижает подвижность в канале. Математическое моделирование позволяет найти оптимальный профиль легирования области под затвором: оказывается, он должен быть неоднородным как по глубине, так и вдоль канала [181]. Именно так и легируются современные транзисторы. В качестве примера на рис. 4.12 показана конструкция МОП-транзистора, используемая в процессорах Pentium фирмы Intel. Эффективная длина канала транзисторов в процессорах Pentium 4 составляет всего 60 нм [180]. Для ослабления эффектов короткого канала толщина диэлектрика в этих транзисторах уменьшена до 15 Å, подложка имеет «ретроградный» профиль распределения примеси (концентрация примеси возрастает по мере удаления от поверхности вглубь), а в области под затвором с помощью ионной имплантации создан «ореол» (halo) — сильно легированная область с концентрацией акцепторов, достигающей нескольких единиц на 10^{18} см$^{-3}$.

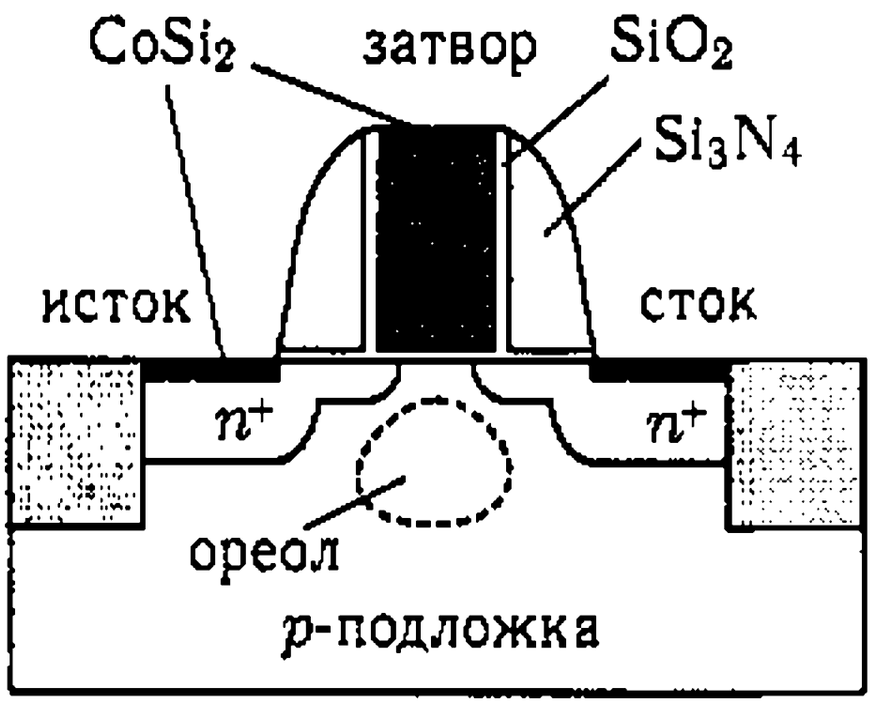


Рис. 4.12. Конструкция полевого транзистора с каналом n-типа, используемого в процессорах Pentium фирмы Intel [180]

По мере сокращения длины канала его сопротивление уменьшается и на характеристики МОП-транзисторов начинает все сильнее влиять конечное сопротивление областей истока и стока. Этому способствует и то, что глубину этих областей приходится делать все меньше и меньше. Внутренняя отрицательная обратная связь, создаваемая сопротивлениями истока и стока, заметно уменьшает крутизну вольт-амперной характеристики g_m, и поэтому в МОП-транзисторах с коротким каналом проблема снижения сопротивления этих областей и сопротивления контактов к ним становится особенно острой. В конструкции, показанной на рис. 4.12, для уменьшения сопротивления областей истока и стока рядом с n^+-областями истока и стока ($n = 5 \cdot 10^{19} - 10^{20}$ см$^{-3}$) создаются очень мелкие (~0,05 мкм) частично заходящие под затвор области n-типа с $n = (4–8) \cdot 10^{18}$ см$^{-3}$.

В МОП-транзисторах с коротким каналом напряженность электрического поля в канале обычно настолько высока, что в приборе возникают *горячие носители*. Появление таких носителей может приводить к нежелательным явлениям — захвату носителей на существующие в окисле ловушки, генерации в окисле новых дефектов. Эти явления вызывают дрейф порогового напряжения и уменьшение

крутизны вольт-амперной характеристики транзистора [156]. Высота барьера ΔE_c на границе раздела $Si–SiO_2$ составляет 3,2 эВ, а барьера ΔE_v — 3,7 эВ (см. табл. 1 в Приложении), и поэтому указанные явления появляются, когда носители приобретают в электрическом поле энергию, сравнимую с высотой соответствующего барьера. В интегральных схемах, работавших при напряжении питания 5 В, проблема горячих носителей становилась все более актуальной по мере уменьшения размеров транзисторов. Было установлено, что повысить стойкость МОП-структур к воздействию горячих электронов можно легируя SiO_2 атомами фтора и азота, а замена в форминг-газе (см. с. 248) водорода на дейтерий позволяет повысить эту стойкость в 10–40 раз. В настоящее время в результате заметного снижения рабочих напряжений микросхем проблема горячих носителей стала менее актуальной. Единственным случаем, когда горячие носители используются в работе МОП-транзисторов, являются специальные конструкции транзисторов, применяемые в энергонезависимых постоянных запоминающих устройствах (см. п. 4.2.3).

4.1.5. Быстродействие полевых транзисторов. Для оценки быстродействия МОП-транзистора рассмотрим его эквивалентную схему (рис. 4.13*а*). Из-за очень низкого тока утечки диэлектрика входную проводимость транзистора на постоянном токе можно считать равной нулю. На переменном токе входная проводимость транзистора определяется емкостями затвор–сток $C_{з–с}$, затвор–исток $C_{з–и}$ и затвор–подложка $C_{з–подл}$, каждая из которых рассчитывается как производная заряда затвора $Q_з$ по напряжению на соответствующем электроде при фиксированных напряжениях на остальных электродах (например, $C_{з–и} = -dQ_з/dV_и$). Важно отметить, что поскольку изменение напряжения на любом из выводов транзистора вызывает сложное перераспределение зарядов внутри структуры, то определенные указанным выше способом емкости зависят не только от геометрии электродов, но и от *режима работы*. Мы не будем приводить здесь расчет емкостей $C_{з–и}$, $C_{з–с}$ и $C_{з–подл}$ в зависимости от напряжений на затворе и стоке, который можно найти в [87]. Отметим, что в практически важном случае работы транзистора в режиме насыщения $C_{з–и} = C_{з–и.к} + 2C_д WL/3$, $C_{з–с} = C_{з–с.к}$, $C_{з–подл} = 0$, где $C_{з–и.к}$, $C_{з–с.к}$ — емкости перекрытия затвора с истоком и стоком, $C_д$ — удельная емкость диэлектрика, а L и W — длина и ширина затвора. Два генератора тока в эквивалентной схеме отражают тот уже упоминавшийся факт, что полевой транзистор является по сути *четырехэлектродным* прибором, который может управляться как напряжением на затворе, так и напряжением на подложке. Кроме того, в схеме учтены конечные сопротивления областей истока и стока

($R_и$, $R_с$), конечная выходная проводимость транзистора $g_{с-и}$, а также емкость и дифференциальная проводимость p–n-переходов исток–подложка и сток–подложка.

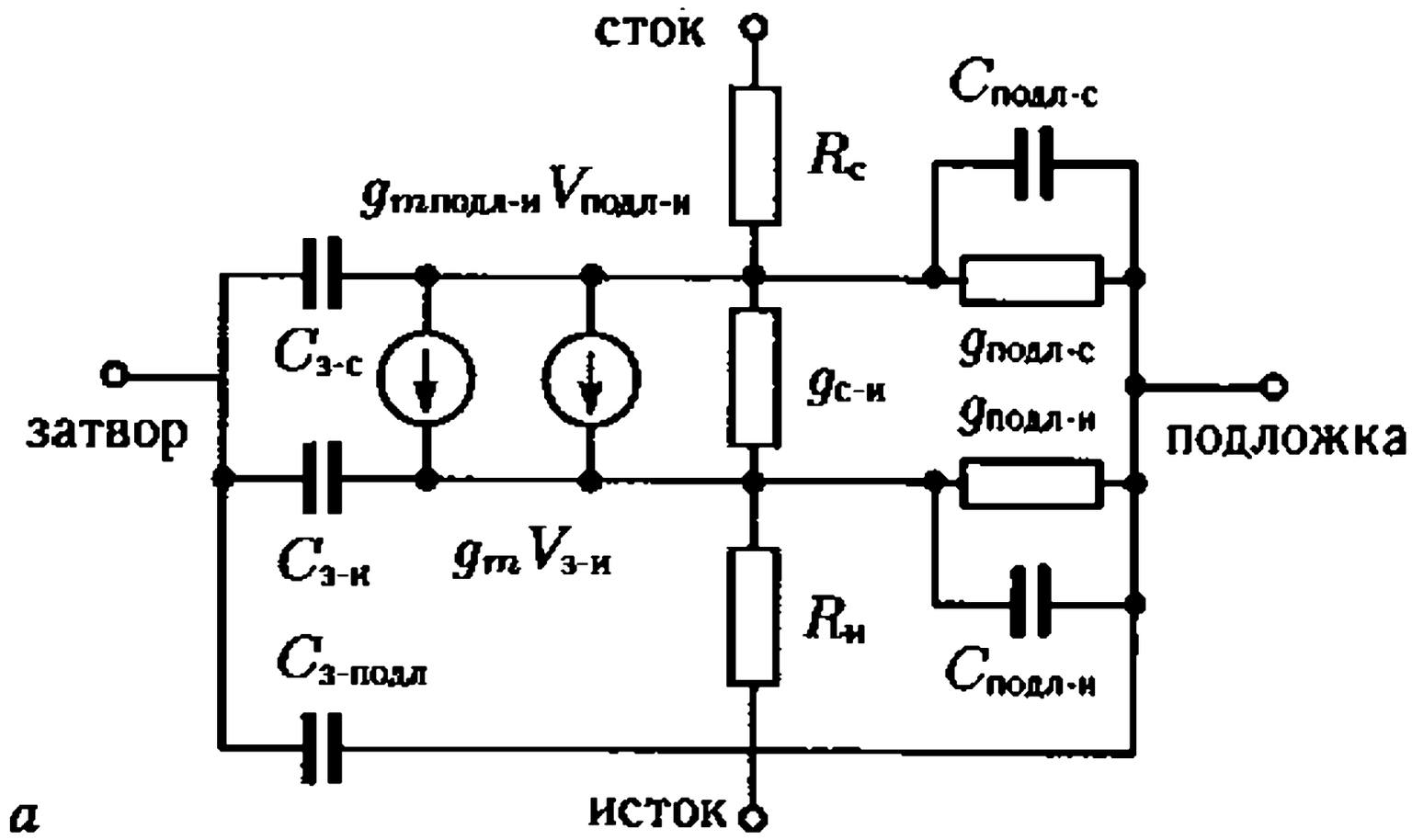


Рис. 4.13. Эквивалентная схема МОП-транзистора на переменном токе: *а* — полная эквивалентная схема, *б* — упрощенная эквивалентная схема

Полную эквивалентную схему можно упростить, если подложку и исток транзистора соединить вместе и пренебречь сопротивлениями истока и стока. Для упрощенной эквивалентной схемы (см. рис. 4.13*б*) нетрудно рассчитать коэффициент усиления по току A_i в режиме короткого замыкания на выходе:

$$A_i = \frac{-g_m}{i\omega(C_{з-и} + C_{з-с} + C_{з-подл})}. \qquad (4.20)$$

Частота, на которой модуль коэффициента усиления по току становится равным единице, называется *предельной частотой* f_T. Она равна

$$f_T = \frac{1}{2\pi}\frac{g_m}{(C_{з-и} + C_{з-с} + C_{з-подл})}. \qquad (4.21)$$

Из этой формулы становится ясным, почему разработчики МОП-транзисторов уделяют столь большое внимание получению

высокой крутизны вольт-амперной характеристики. Используя уравнение (4.13) для тока стока, перепишем выражение для g_m следующим образом:

$$g_m = \left.\frac{dI_c}{dV_з}\right|_{V_c=\text{const}} = \frac{d}{dV_з}(|qn_{ss}|v_{эф}W) =$$

$$= \frac{d}{dV_з}\left(\frac{Q'_з}{WL}v_{эф}W\right) \approx C_д\frac{v_{эф}}{L} = \frac{C_д}{t_{прол}}, \quad (4.22)$$

где $Q'_з$ — полный заряд подвижных носителей под затвором, $v_{эф}$ — эффективная скорость дрейфа носителей, а $t_{прол}$ — их время пролета через канал. Учитывая, что обычно $C_{з-и} + C_{з-с} + C_{з-подл} \approx C_д$, окончательно получаем, что

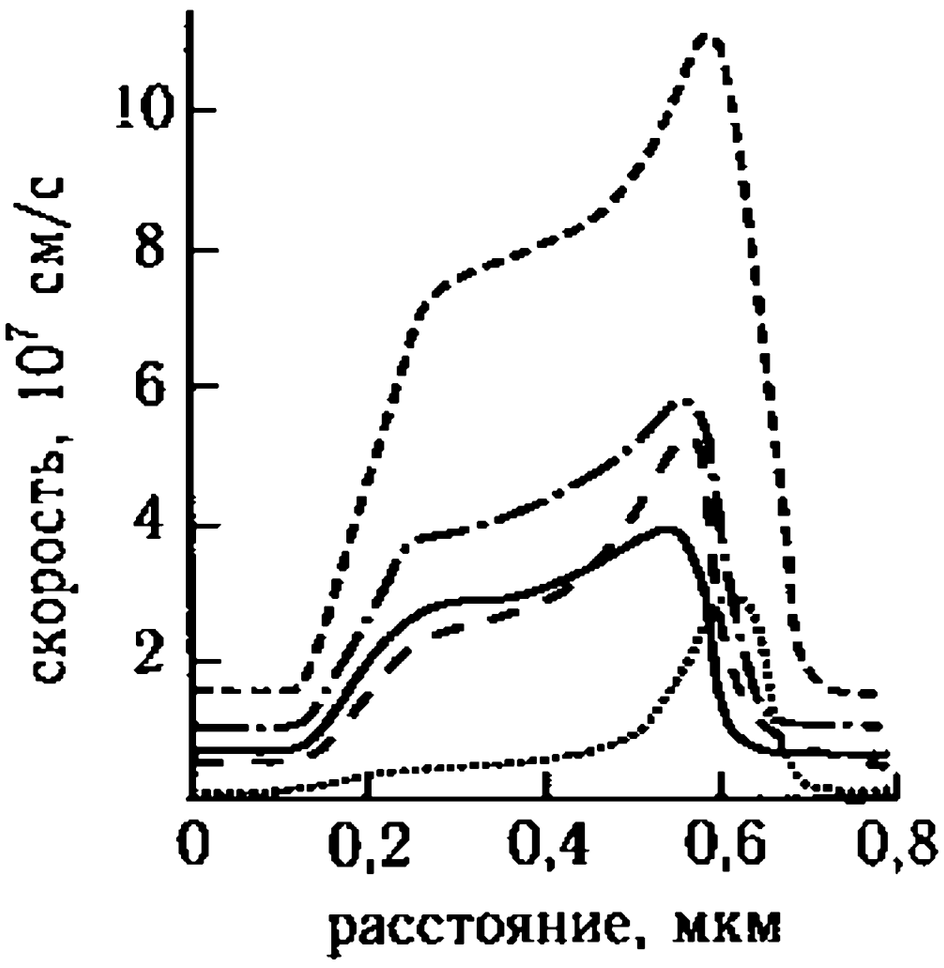
$$f_T \approx \frac{1}{2\pi t_{прол}}. \quad (4.23)$$

Из этого уравнения следует, что для создания быстродействующих полевых транзисторов необходимо стремиться максимально уменьшить время пролета носителей. Очевидный способ сделать это — сократить длину канала. Считая, что в кремниевых МОП-транзисторах носители движутся в канале со скоростью, близкой к скорости насыщения ($v_s \approx 10^7$ см/с), находим f_T (ГГц) $\approx$ $\approx 16/L$ (мкм). Длина канала в современных МОП-транзисторах достигает 400–600 Å.

Говоря о МОП-транзисторах, хотелось бы отметить их существенно более высокое быстродействие по сравнению с МОП-конденсатором. В то время как в МОП-конденсаторе носители, образующие канал, появляются в результате тепловой генерации неосновных носителей в подложке (что и определяет медленную реакцию индуцированного заряда на изменение напряжения на затворе), в МОП-транзисторах эти носители поступают в канал через области истока и стока, что и определяет их существенно более быструю реакцию. При этом, поскольку полевые транзисторы работают на *основных* носителях заряда (то есть в них отсутствуют те ограничения, которые накладывает конечная скорость рекомбинации неосновных носителей в приборах с инжекцией), это делает эти транзисторы потенциально более быстродействующими приборами.

4.1.6. Пути дальнейшего повышения быстродействия МОП-транзисторов. Возможность дальнейшего повышения быстродействия полевых транзисторов основана на использовании новых явлений: эффекта «всплеска скорости», явления

баллистического переноса и эффекта увеличения подвижности в напряженном кремнии.



материал		скорость насыщения, 10^7 см/с
Si	·····	1
GaAs	———	1,8
InP	– – – –	2,4
GaInAs	—·—·—	2,1
InAs	------	3,5

Рис. 4.14. Зависимость скорости электронов от расстояния в канале полевых транзисторов, изготовленных из различных полупроводников (эффект «всплеска скорости»). Для сравнения в таблице справа указаны максимальные скорости насыщения, полученные в структурах с длинным каналом [87]

Эффект «всплеска скорости» был обнаружен при моделировании динамики движения электронов в канале полевого транзистора методом Монте-Карло [182]. Оказалось, что когда электроны попадают в сильное электрическое поле, скорость их движения сначала быстро нарастает до значений, которые могут в несколько раз превышать скорость насыщения v_s, а затем эта скорость уменьшается до значения, равного v_s (см. рис. 4.14). Причиной такого поведения является существование в полупроводнике двух различных характерных времен: времени релаксации импульса τ_k и времени релаксации энергии τ_e. Поскольку $\tau_k \ll \tau_e$, то после попадания электронов в сильное электрическое поле за время порядка τ_k устанавливается их квазистационарное распределение по импульсам, и все носители начинают двигаться «синхронно» вдоль направления электрического поля. Однако через характерное время τ_e неупругие процессы рассеяния «растраивают» это движение и средняя скорость дрейфа уменьшается. Если носитель будет успевать пролететь через канал за время, меньшее τ_e, то его средняя скорость движения будет превосходить скорость насыщения и, в соответствии с формулой (4.23), частота отсечки может быть сделана выше. В транзисторах на основе GaAs эффект «всплеска скорости» проявляется в приборах с длиной канала менее 0,5 мкм, а в транзисторах на основе кремния — при длине канала менее 0,1 мкм [182].

Другой режим, называемый *режимом баллистического переноса*, может возникать в структурах со сверхкоротким ($\lesssim 500$ Å) каналом, когда носитель, движущийся в сильном электрическом поле, пролетает канал за время, меньшее τ_k, то есть с большой вероятностью вообще не испытав ни одного акта рассеяния.

Большой интерес к этому режиму связан с тем, что поскольку носитель движется максимально быстро (его скорость в любой точке канала определяется только значением электростатического потенциала в этой точке), этот режим позволяет реализовать минимально возможное для данной геометрии прибора время пролета и максимальную крутизну вольт-амперной характеристики.

МОП-структура, названная *баллистическим нанотранзистором*, была создана в Bell Laboratories в 1999 году. Длина канала в этой структуре составляла всего 250 Å. Оказалось, что для реализации баллистического режима переноса необходима чрезвычайно гладкая граница $Si–SiO_2$, и поэтому в этом транзисторе использовался сверхтонкий слой окисла. В структуре с толщиной окисла 16 Å при комнатной температуре был получен ток стока, составляющий 85% от максимально возможного значения, отвечающего чисто баллистическому режиму.

Исследования показали, что реализовать преимущества, связанные с эффектом «всплеска скорости» и баллистическим переносом, непросто. При уменьшении длины канала сильно возрастает рассеяние носителей на заряженных примесях, которые располагаются в сильно легированном затворе из поликристаллического кремния, в области ореола и в сильно легированных областях истока и стока. Кроме того, из-за увеличения напряженности поперечного электрического поля под затвором сильно возрастает рассеяние носителей на фононах и шероховатостях поверхности, а распределение потенциала в канале короткого МОП-транзистора таково, что при этом трудно обеспечить высокую напряженность электрического поля вблизи истока. Поэтому необходимо было пробовать другие подходы.

Эффект увеличения подвижности в слоях *напряженного кремния* известен сравнительно давно. Исследования упруго напряженных сверхрешеток $Si/Si_{1-x}Ge_x$ [183] обнаружили заметное увеличение подвижности электронов в слоях Si(100). Своим происхождением этот эффект обязан тому, что в результате растяжения кремния в плоскости слоя шестикратное вырождение его зоны проводимости снимается и наинизшими по энергии оказываются минимумы, эллипсоиды эффективной массы которых вытянуты в направлении, перпендикулярном

слою. Поскольку энергия этих минимумов оказывается заметно ниже края зоны проводимости в $Si_{1-x}Ge_x$, то электроны локализуются в слое Si. Из-за небольшой эффективной массы ($m_\perp \approx \approx 0,19m_0$) и меньшего темпа рассеяния этих электронов их подвижность в слое Si возрастает почти в 2 раза по сравнению с объемным материалом [184]. Хотя скорость насыщения при этом практически не изменяется, увеличение подвижности позволяет повысить крутизну вольт-амперной характеристики транзисторов на 60% [185] с соответствующим выигрышем в быстродействии. Полевые транзисторы из напряженного кремния можно получить наращиванием на подложку из Si варизонного слоя твердого раствора $Si_{1-x}Ge_x$ толщиной ~1,5 мкм, заканчивающегося слоем постоянного состава ($x \sim 0,2$) толщиной ~0,6 мкм; на поверхности этого «релаксированного» (ненапряженного) слоя затем создают тонкий (~100 Å) слой Si, который и будет служить каналом МОП-транзистора. Механические напряжения в структуре можно создать и другими способами. Так, в современных процессорах фирмы Intel напряжения в области канала n-МОП-транзисторов создают осаждением вокруг затвора слоев Si_3N_4 (см. рис. 4.12), а в транзисторах с каналом p-типа для создания напряжений используются области истока и стока, изготовленные из твердого раствора $Si_{1-x}Ge_x$ с помощью селективной эпитаксии. Напряженный кремний вот уже несколько лет используется фирмами IBM, Intel и AMD при производстве процессоров.

Наконец, очень перспективный путь повышения быстродействия — уменьшение напряженности поперечного электрического поля в канале. Наиболее эффективно это можно осуществить, создав у МОП-транзистора *два затвора* (симметрично сверху и снизу канала) или *окольцовывающий затвор*. Согласно теоретическим расчетам, в двухзатворных МОП-транзисторах подвижность носителей в канале может быть увеличена в 2–3 раза [186].

4.1.7. Мощные и высоковольтные МОП-транзисторы. Мощные полевые транзисторы с изолированным затвором находят применение в качестве усилителей мощности и ключевых элементов. Отсутствие явления накопления неосновных носителей обеспечивает лучшие динамические характеристики и более высокую скорость переключения этих транзисторов по сравнению с биполярными приборами. В настоящее время разработан целый ряд конструкций мощных МОП-транзисторов,

среди которых наибольшее значение имеют структуры с вертикальным направлением протекания тока [11, 163].

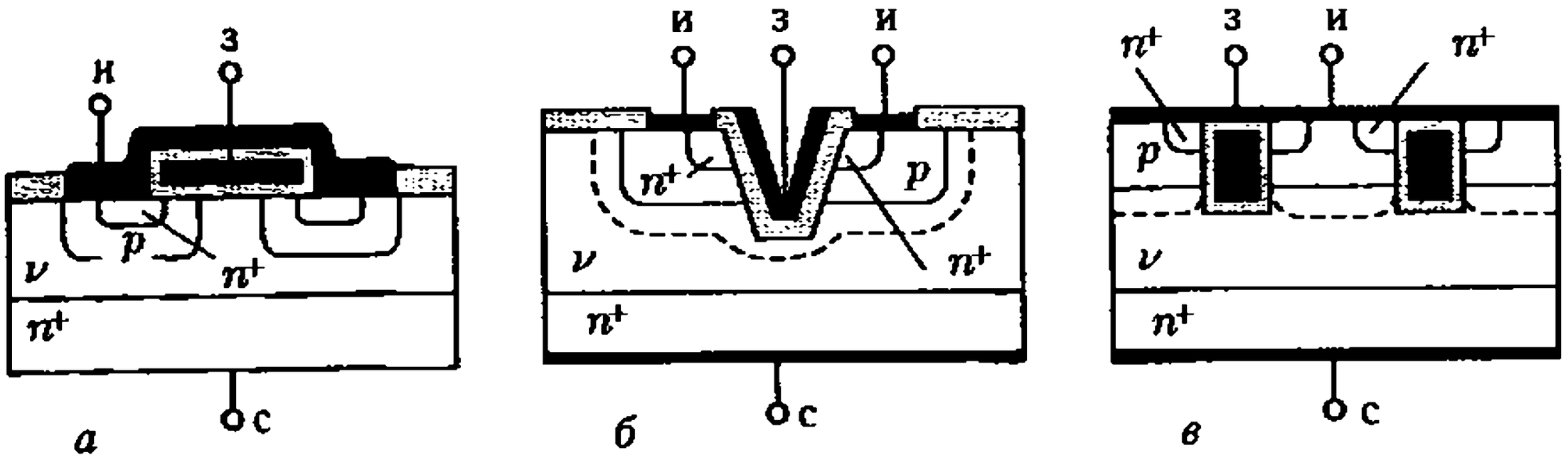


Рис. 4.15. Три конструкции мощных МОП-транзисторов с вертикальной структурой [163]

На рис. 4.15*а* показана конструкция мощного МОП-транзистора, созданного методом двойной диффузии (DMOS). Эти транзисторы получаются диффузией акцепторной и донорной примесей через окна, открытые в окисле на поверхности слабо легированного ($n \sim 10^{14}$ см$^{-3}$) эпитаксиального слоя ν-типа, выращенного на сильно легированной n^+-подложке. Слабое легирование эпитаксиального слоя позволяет сделать так, чтобы область пространственного заряда стока могла распространяться на всю толщину слоя и электроны, попадающие в ν-область из горизонтально расположенного канала, могли быстро дрейфовать в направлении к стоку. Большая протяженность области пространственного заряда позволяет заметно повысить напряжение пробоя структуры. Параллельное соединение большого числа таких транзисторов в одном приборе позволяет создать мощный полевой транзистор. Приборы такой конструкции характеризуются напряжением пробоя до 800 В, но при этом имеют сравнительно высокое (порядка нескольких Ом) сопротивление в открытом состоянии.

На рис. 4.15*б* показана конструкция мощного МОП-транзистора с V-образной канавкой (VMOS). На сильно легированной подложке n^+-Si (область стока) наращивается эпитаксиальный слой ν-типа, в котором путем диффузии создаются слой p-типа и поверхностные области n^+-типа, которые будут служить выводами истока. Затем путем селективного химического травления на поверхности кремния формируются V-образные канавки (см. с. 208). Поверхность канавок окисляется и поверх окисла наносится металлический затвор или затвор из легированного поликремния. Для исключения открывания паразитного n–p–n-транзистора, образованного областями n^+-типа и p-слоем, эмиттер и база этого транзистора при создании

контакта к истоку соединяются между собой. При подаче на затвор положительного смещения на границе окисла и p-слоя индуцируются проводящие каналы и ток в структуре течет почти вертикально, проходя часть пути в эпитаксиальном ν-слое. Концентрация примеси в этом слое намеренно выбирается невысокой ($10^{14} - 10^{15}$ см$^{-3}$), чтобы обеспечить высокое напряжение пробоя. При напряжении пробоя 30 В такой транзистор может обеспечить очень низкое сопротивление в открытом состоянии — до 0,05 Ом. Транзисторы с подобной структурой и длиной канала 0,5 мкм имеют частоту отсечки, превышающую 1 ГГц [11].

В современных МОП-транзисторах с вертикальным затвором (см. рис. 4.15*в*) на поверхности кремния методом анизотропного плазменного травления создаются глубокие канавки (trench), в которых формируются изолированные слоем окисла поликремниевые затворы. Практически вертикальные стенки канавок позволяют значительно повысить плотность упаковки и уменьшить сопротивление транзистора в открытом состоянии по сравнению с VMOS-структурами. Так, транзистор фирмы Infineon площадью 30 мм^2 с напряжением пробоя 50 В, состоящий более чем из 300 тысяч параллельно соединенных ячеек, имеет сопротивление открытого канала $3 \cdot 10^{-3}$ Ом [163].

Особенностью МОП-транзисторов по сравнению с биполярными транзисторами является то, что в области температур до ~150 °C сопротивление канала возрастает с увеличением температуры. Это способствует получению более однородного распределения плотности тока по площади структуры. Поэтому мощные МОП-транзисторы лучше переносят перегрузки по току и менее подвержены явлению вторичного пробоя.

Биполярные транзисторы с изолированным затвором (IGBT). Необходимость в поиске новых конструкций мощных полупроводниковых приборов вызвана рядом недостатков, присущих чисто полевым и чисто биполярным транзисторам. Так, полевые транзисторы при одинаковом с биполярными транзисторами максимальном токе должны иметь в несколько раз бóльшую площадь, в результате чего их стоимость оказывается в 2–3 раза выше. Недостатком мощных биполярных транзисторов является их невысокое быстродействие, связанное с накоплением в базе инжектированных носителей, что приводит к большим коммутационным потерям при их работе на переменном токе. Поэтому разработка гибридных структур, позволяющих объединить в себе достоинства каждого из указанных типов транзисторов, является весьма актуальной. В настоящее время разработаны

и выпускаются в промышленных масштабах несколько типов таких биполярно-полевых гибридных структур.

Первые монолитные биполярно-полевые структуры были изготовлены в конце 70-х годов и существенно усовершенствованы в 80-х годах [163]. Наибольшее распространение среди них нашли структуры, получившие название *биполярных транзисторов с изолированным затвором* (IGBT — insulated gate bipolar transistor), в которых объединены лучшие свойства полевых (быстродействие, малый ток управления) и биполярных (низкое напряжение насыщения) приборов. В настоящее время эти приборы занимают около 5% рынка выпускаемых мощных полупроводниковых приборов и используются в качестве мощных управляемых ключей, несколько потеснив тиристоры с занимаемых ими позиций.

Важным достоинством IGBT является их полная управляемость со стороны затвора и способность сравнительно быстро включаться и выключаться. По сравнению с запираемыми тиристорами (см. с. 224), эти приборы могут работать на более высоких частотах (до 50–100 кГц), однако для них характерно более высокое падение напряжения в открытом состоянии (1,5–3 В). Поэтому IGBT наиболее эффективно использовать в высоковольтной электронике. В настоящее время рабочее напряжение этих приборов достигает 6,5 кВ (изделие CM600HG-130H фирмы Powerex), а ток — 2400 А (изделие CM2400HC-34H той же фирмы). Еще одним важным достоинством IGBT является простота схем управления, сильно отличающая их от запираемых тиристоров.

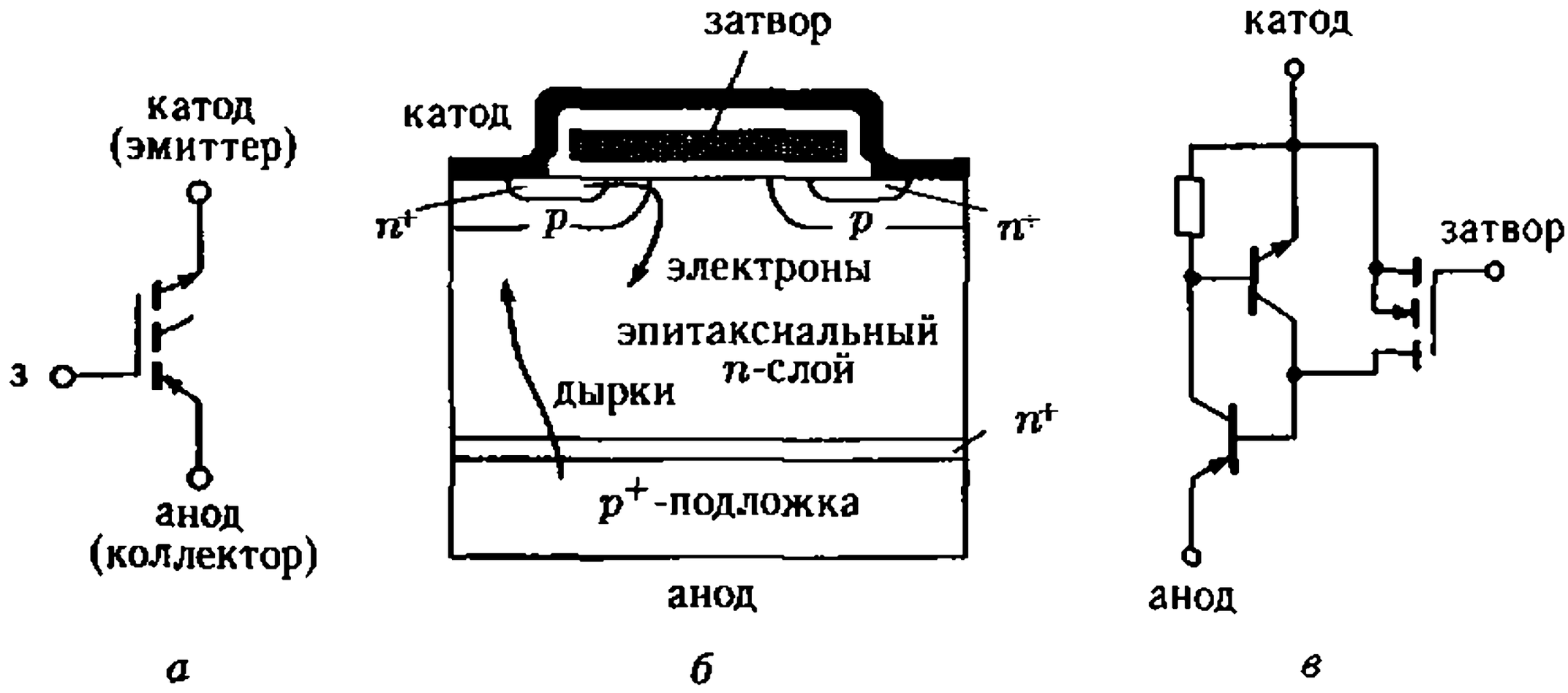


Рис. 4.16. Условное обозначение (*а*), поперечное сечение одной ячейки (*б*) и эквивалентная схема (*в*) биполярного транзистора с изолированным затвором (IGBT)

Устройство IGBT показано на рис. 4.16. По сути прибор представляет собой функционально-интегрированную структуру, состоящую из низковольтного полевого транзистора, истоком которого является n^+-область катода структуры, а стоком — высокоомный эпитаксиальный слой n-типа, и мощного p–n–p-транзистора, образованного p^+-подложкой (анодом структуры), эпитаксиальным слоем и p-областью катода структуры. Имеющийся в структуре n–p–n-транзистор, образованный n^+-областью катода, p-областью катода и эпитаксиальным слоем, является паразитным и в работе IGBT не используется: его открывание может привести к *эффекту защелкивания* (см. подробнее с. 281) и потере управляемости структурой, в результате чего прибор обычно выходит из строя. Чтобы этот транзистор никогда не открывался, n^+- и p-область катода IGBT изготавливают в виде закороченной структуры (подобно закороченному катоду тиристора, см. п. 3.1.1). Толщина эпитаксиального слоя составляет ~200 мкм, а выводы затвора и эмиттера располагаются на поверхности прибора в виде регулярной структуры с характерным расстоянием между ячейками, равным несколько мкм в современных приборах. Иногда в структуру между p^+-подложкой и эпитаксиальным слоем добавляют буферный n^+-слой, который позволяет уменьшить падение напряжение на приборе в открытом состоянии (этот вопрос уже обсуждался нами в п. 3.1 на с. 222) [1]. Дальнейшее улучшение характеристик IGBT достигается в структурах с вертикальным затвором [163].

Для перевода IGBT в открытое состояние на затвор подают положительное смещение, которое индуцирует проводящий канал между n^+-областью эмиттера и эпитаксиальным слоем (см. рис. 4.16). Протекающий по каналу ток электронов открывает p–n–p-транзистор и дырки начинают инжектироваться из p^+-области в эпитаксиальный слой. Эти дырки и электроны, подходящие из катода для обеспечения электронейтральности, сильно уменьшают сопротивление эпитаксиального слоя и тем самым обеспечивают низкое падение напряжения на приборе в открытом состоянии. Поскольку паразитный n–p–n-транзистор всегда закрыт, то, в отличие от тиристора, прибор *остается управляемым* и его можно легко перевести в закрытое состояние уменьшением напряжения на затворе. Время выключения IGBT определяется временем рассасывания инжектированных в базу

[1]) В зарубежной литературе структуры с таким слоем называют punch through (PT) IGBT.

p–n–p-транзистора дырок и может быть снижено до нескольких сотен наносекунд за счет создания радиационных дефектов при облучении прибора протонами.

4.2. Элементы интегральных схем на МОП-транзисторах

Путь к интегральной схеме на полевых транзисторах был нелегок. Несмотря на то, что первый МОП-транзистор был создан в 1960 г., использованию этих транзисторов в реальных приборах препятствовали высокая плотность поверхностных состояний на границе $Si–SiO_2$ и нестабильность характеристик. Балк, работавший в фирме IBM, в 1965 г. первым нашел способ уменьшения плотности поверхностных состояний путем термообработки окисла в атмосфере водорода, при которой происходит пассивация водородом ненасыщенных связей атомов Si и O (см. с. 247). Примерно в то же время сотрудники фирмы Fairchild Semiconductors установили, что нестабильность характеристик МОП-транзисторов связана с неконтролируемо присутствующими в окисле ионами натрия, и нашли способ связывания этих атомов. Только объединение этих двух идей позволило в 1965 г. выпустить первые интегральные схемы на МОП-транзисторах со стабильными характеристиками.

Первые интегральные схемы (ИС) на полевых транзисторах были построены на основе МОП-транзисторов с каналом *p*-типа. Технология получения этих транзисторов была очень проста и поэтому для создаваемых на их основе ИС был характерен намного более высокий выход годных по сравнению с ИС на биполярных транзисторах. Это предопределяло низкую стоимость МОП ИС и их быстрое развитие. Однако из-за большого встроенного положительного заряда в слое окисла (типичное пороговое напряжение *p*-МОП-транзистора равно 6 В) ИС на *p*-канальных транзисторах требовали высокого напряжения питания ($E_п = 27$ В) и характеризовались высокой потребляемой мощностью. При этом наряду с высокой *статической* рассеиваемой мощностью (связанной с протеканием постоянного тока через структуру), для них была характерна и большая *динамическая* мощность рассеяния (возникающая при переключении схем из состояния логической 1 в состояние логического 0 и обратно), которая равна $P_{дин} = fC_н U^2_{0-1}$, где f — частота переключения, $C_н$ — емкость шины, а U_{0-1} — разность напряжений, отвечающих уровням 0 и 1. Кроме того, было ясно, что из-за

примерно втрое более низкой подвижности дырок в Si по сравнению с электронами, в таких ИС нельзя полностью реализовать высокие потенциальные возможности полевых транзисторов.

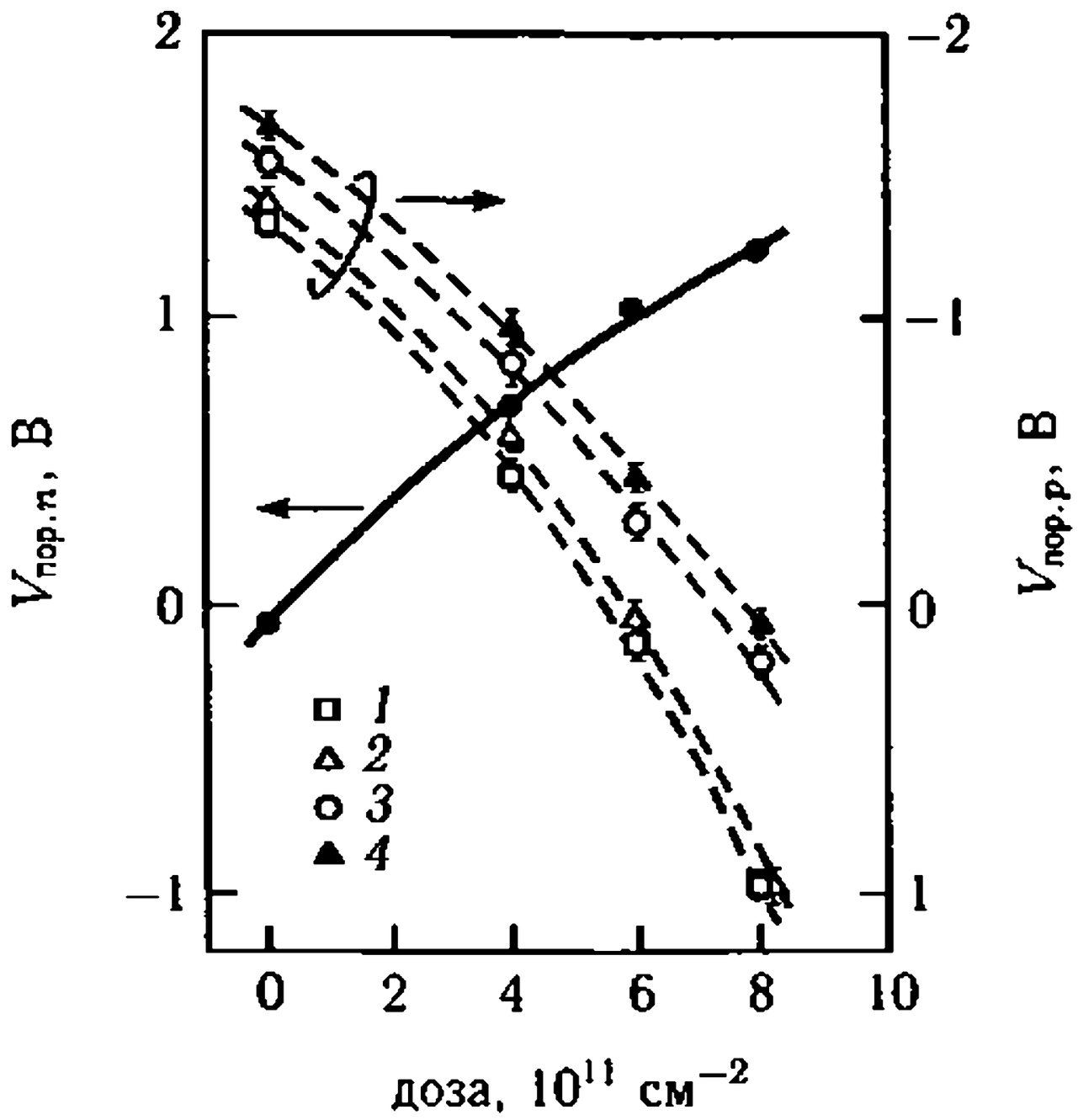


Рис. 4.17. Зависимость порогового напряжения МОП-транзисторов с каналами n- и p-типа от дозы имплантированных в приповерхностный слой ионов бора. Толщина подзатворного окисла 650 Å, поликремниевый затвор n^+-типа. Сплошная линия — транзистор с каналом n-типа, пунктирные линии — транзисторы с каналом p-типа. Последние транзисторы изготовлены в слое n-типа, созданном на поверхности подложки p-типа путем имплантации ионов фосфора с поверхностной концентрацией (в 10^{12} см$^{-2}$): *1* — 0,8; *2* — 1,0; *3* — 1,5; *4* — 2,0 [58]

При попытке создания ИС на МОП-транзисторах с каналом n-типа разработчики столкнулись с тем, что из-за положительного встроенного заряда в окисле (см. с. 247) на поверхности p-Si всегда образуется инверсионный слой, и необходимо было сначала научиться управлять пороговым напряжением n-МОП-транзистора. В первых микросхемах для коррекции $V_{пор}$ на подложку схемы подавалось большое отрицательное смещение; позже для коррекции $V_{пор}$ было предложено использовать ионную имплантацию. Действительно, если считать имплантированную примесь сосредоточенной в очень тонком слое вблизи границы раздела $Si–SiO_2$, то поверхностная плотность встроенного заряда в окисле просто изменяется на величину поверхностной плотности заряда внедренной примеси Q и сдвиг порогового напряжения равен $\Delta V_{пор} \approx Q/C_д$ [87]. Поэтому внедряя в приповерхностный слой полупроводника ионы акцепторных или

донорных примесей, можно направленно менять пороговое напряжение, при этом практически не изменяя емкость обедненного слоя. Имплантация доноров повышает пороговое напряжение в p-канальных МОП-транзисторах и понижает его в n-канальных приборах; имплантация акцепторов производит противоположное действие (см. рис. 4.17). Решение проблемы коррекции $V_{\text{пор}}$ позволило создать ИС на МОП-транзисторах с каналом n-типа, которые работали при напряжении $E_{\text{п}} = 5$ В и имели более высокое быстродействие.

4.2.1. Интегральные схемы на n-МОП-транзисторах. На рис. 4.18 показано устройство логического элемента «2И-НЕ», построенного на n-МОП-транзисторах. Элемент образован двумя последовательно включенными нормально закрытыми (работающими в режиме обогащения) транзисторами Т1 и Т2 и одним нормально открытым (работающим в режиме обеднения) транзистором Т3, который служит активной нагрузкой (см. рис. 4.18 *а*) [1]. Низкое выходное напряжение на выходе схемы (логический 0) устанавливается только тогда, когда оба транзистора Т1 и Т2 открыты, то есть на их входы подан высокий потенциал, отвечающий уровню логической 1.

Интегральные схемы на МОП-транзисторах создаются с помощью планарной технологии (см. п. 2.8.1). Основными этапами изготовления таких ИС являются [58]: создание в необходимых местах окисной изоляции; выращивание тонкого слоя подзатворного окисла; осаждение слоя поликремния, из которого затем с помощью фотолитографии изготавливаются затворы и нижний уровень контактных соединений; создание сильно легированных областей истока и стока с помощью ионной имплантации. После изготовления транзисторов поверхность структуры закрывается слоем фосфорно-силикатного стекла, которое является геттером подвижных ионов натрия и защищает прибор от их вредного воздействия. В этом стекле с помощью литографии вскрываются окна и с помощью верхнего слоя металлизации создаются необходимые соединения между элементами, а затем вся структура защищается слоем плазмохимически осажденного нитрида кремния (SiN_x), который герметизирует прибор и предохраняет его от загрязнения и механических повреждений.

---

[1]) Термины «нормально открытый» и «нормально закрытый» обозначают состояние транзистора при $V_{\text{зи}} = 0$, которое определяется знаком порогового напряжения $V_{\text{пор}}$. На электрических схемах каналы таких транзисторов изображаются, соответственно, сплошной и прерывистой линиями (см. рис. 4.18).

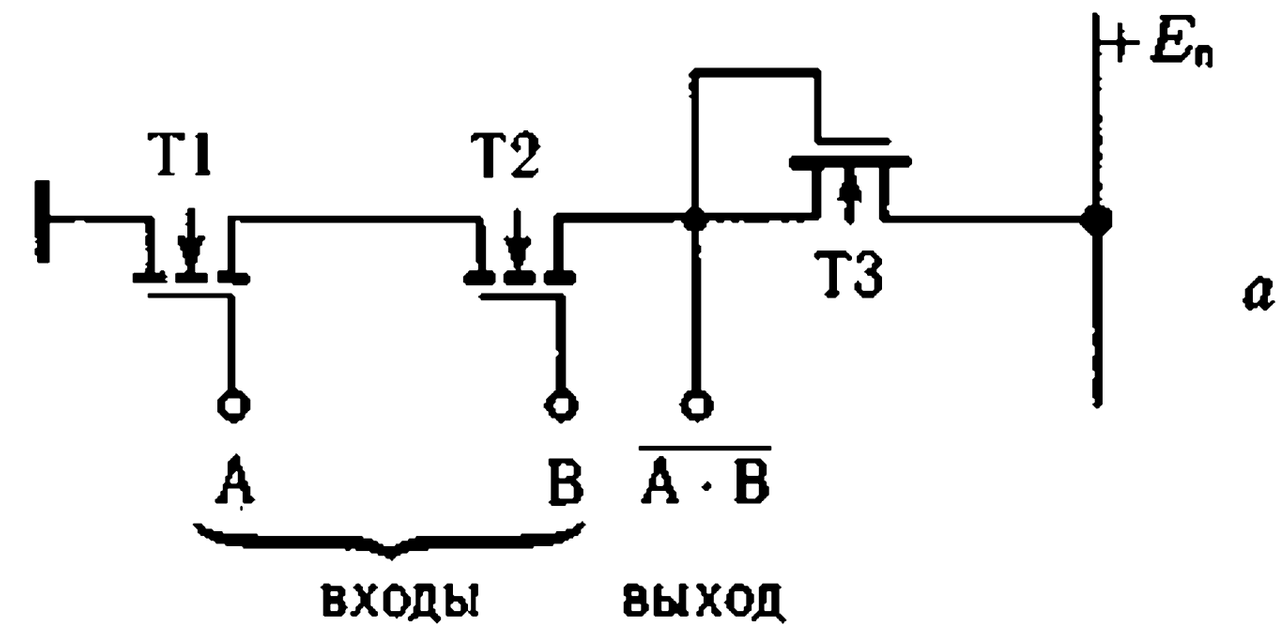


Рис. 4.18. Электрическая схема (*а*), топология (*б*) и поперечное сечение (*в*) логического элемента «2И-НЕ», построенного на n-МОП-транзисторах [58]

Особенностью процесса изготовления ИС на основе МОП-транзисторов с каналом n-типа является широкое использование ионной имплантации. Это связано с несколькими причинами. Во-первых, положительный встроенный заряд в окисле (см. с. 247) приводит к тому, что под окисной пленкой на поверхности p-Si всегда образуется инверсионный слой. Поэтому изоляция элементов в ИС требует не просто использования окисной изоляции, но и дополнительного легирования бором областей полупроводника, в которых будет создаваться эта изоляция, — создания *областей ограничения распространения канала* (см. рис. 4.18 *в*).

Это легирование проводится с помощью ионной имплантации. Во-вторых, использование в элементе одновременно нормально открытых и нормально закрытых транзисторов требует коррекции их пороговых напряжений, которое проводится путем имплантации примеси B для создания нормально закрытых транзисторов и примесей P или As для создания нормально открытых транзисторов. Наконец, для уменьшения паразитных емкостей перекрытия затвор–исток и затвор–сток транзисторов используется процедура *самосовмещения* областей истока, стока и затвора, при которой окончательное формирование границ истока и стока в транзисторе осуществляется путем имплантации донорных примесей *после* изготовления поликремниевого затвора. Энергия имплантируемых ионов подбирается так, чтобы они свободно проникали в подложку сквозь тонкий слой подзатворного окисла, но задерживались достаточно толстым затвором. При этом границы легированных n^+-областей истока и стока оказываются (независимо от точности совмещения фотошаблонов) точно совпадающими с краями затвора и паразитные емкости оказываются минимальными.

Статические и динамические запоминающие устройства. Большая доля выпускаемых в настоящее время полевых транзисторов входит в состав ИС запоминающих устройств. Микросхемы динамических запоминающих устройств и энергонезависимой памяти — это те направления разработок МОП СБИС, в которых используются самые передовые конструкции элементов, технологические приемы и методы проектирования. В этом разделе мы рассмотрим конструкцию и принцип действия статических и динамических запоминающих устройств, а работа энергонезависимой памяти будет рассмотрена в п. 4.2.3.

На рис. 4.19 показана электрическая схема и топология ячейки памяти статического оперативного запоминающего устройства (ОЗУ) [1]. Информация в ячейке хранится в триггере, образованном нормально закрытыми транзисторами T3 и T4, а нормально открытые транзисторы T1 и T2 выполняют роль активной нагрузки. При подаче уровня логической 1 на адресную шину АШ нормально закрытые транзисторы T5 и T6 открываются; при этом в режиме считывания хранящаяся в триггере информация считывается по разрядным шинам РШ расположенным в той же микросхеме усилителем считывания, а в режиме записи новая

[1]) В зарубежной литературе для обозначения этого типа устройств используется аббревиатура SRAM (static random access memory).

информация по сигналам, устанавливаемым на разрядных шинах схемой формирователя, записывается в триггер. Статическое ОЗУ может хранить записанную информацию сколь угодно долго, пока на микросхему подается напряжение питания.

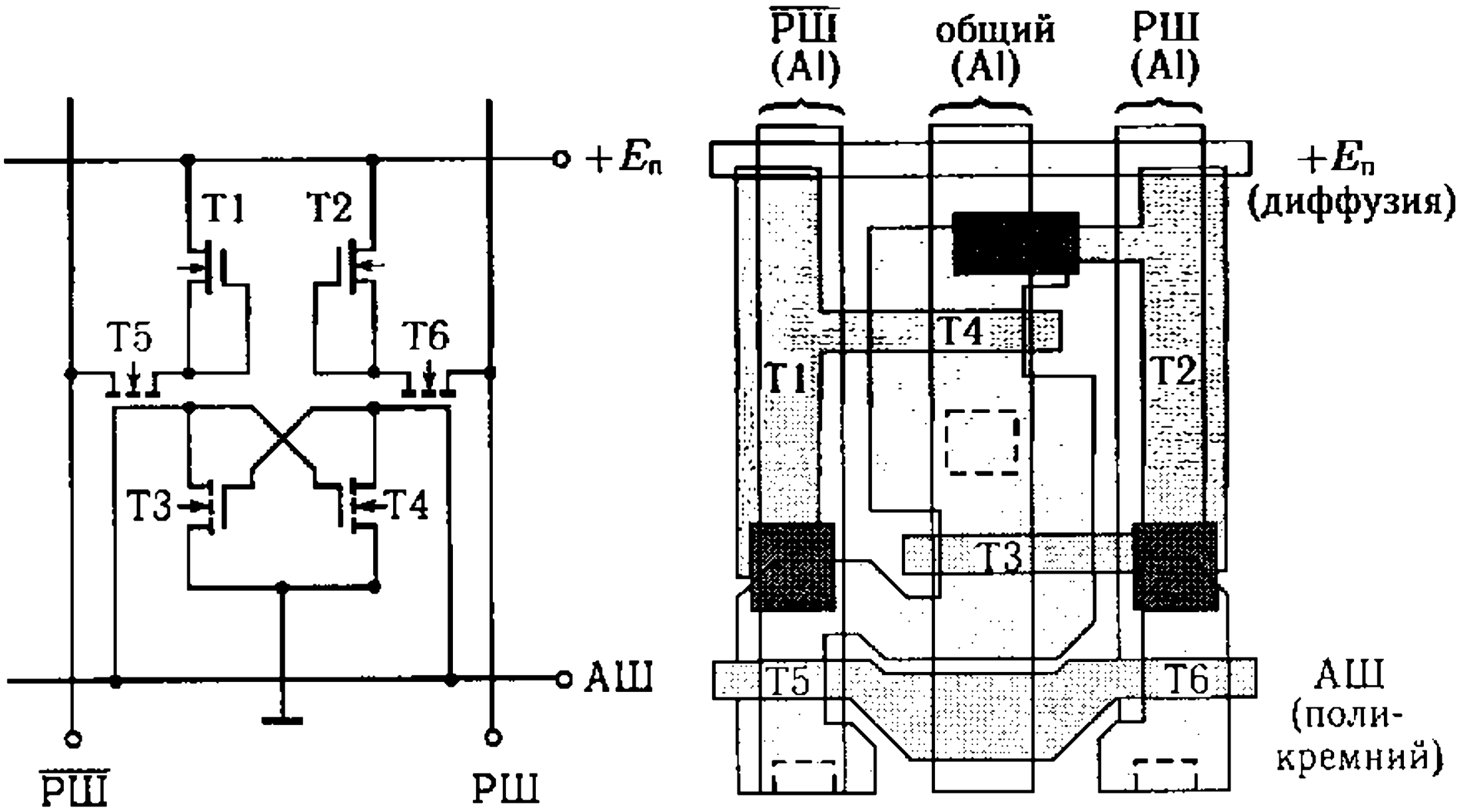


Рис. 4.19. Электрическая схема (*а*) и топология (*б*) ячейки памяти статического ОЗУ на *n*-МОП-транзисторах [58]

В поперечном сечении рассматриваемая ячейка памяти состоит из трех слоев, показанных различной штриховкой на рис. 4.19*б*:

1) диффузионного *n*-слоя, создаваемого на поверхности *p*-кремния в местах, где необходимо сформировать области истоков и стоков транзисторов и шину питания $+E_п$,

2) слоя поликремния, изолированного от подложки тонким слоем подзатворного окисла, из которого изготавливают затворы транзисторов и адресную шину, и

3) лежащего еще выше верхнего слоя металлизации из алюминия, отделенного от остальных элементов структуры толстым слоем окисла, из которого изготавливают разрядные шины и общую шину питания микросхемы.

Необходимые соединения между диффузионным слоем и поликремнием осуществляют в открытых в окисле окнах, показанных на рисунке наиболее темной штриховкой, а соединения между диффузионным слоем и алюминиевыми шинами — в окнах, обозначенных пунктирными квадратами.

Поскольку в ячейке памяти статического ОЗУ один из транзисторов Т3 или Т4 оказывается открытым, то через ячейку памяти всегда протекает некоторый ток. Величину этого тока можно менять, изменяя отношение длины к ширине канала нагрузочных транзисторов Т1 и Т2. Увеличение тока одновременно увеличивает быстродействие ячейки и потребляемую ею мощность. Для уменьшения потребляемой мощности в некоторых конструкциях статических запоминающих устройств нагрузочные транзисторы заменяются высокоомными резисторами из слабо легированного поликремния [58].

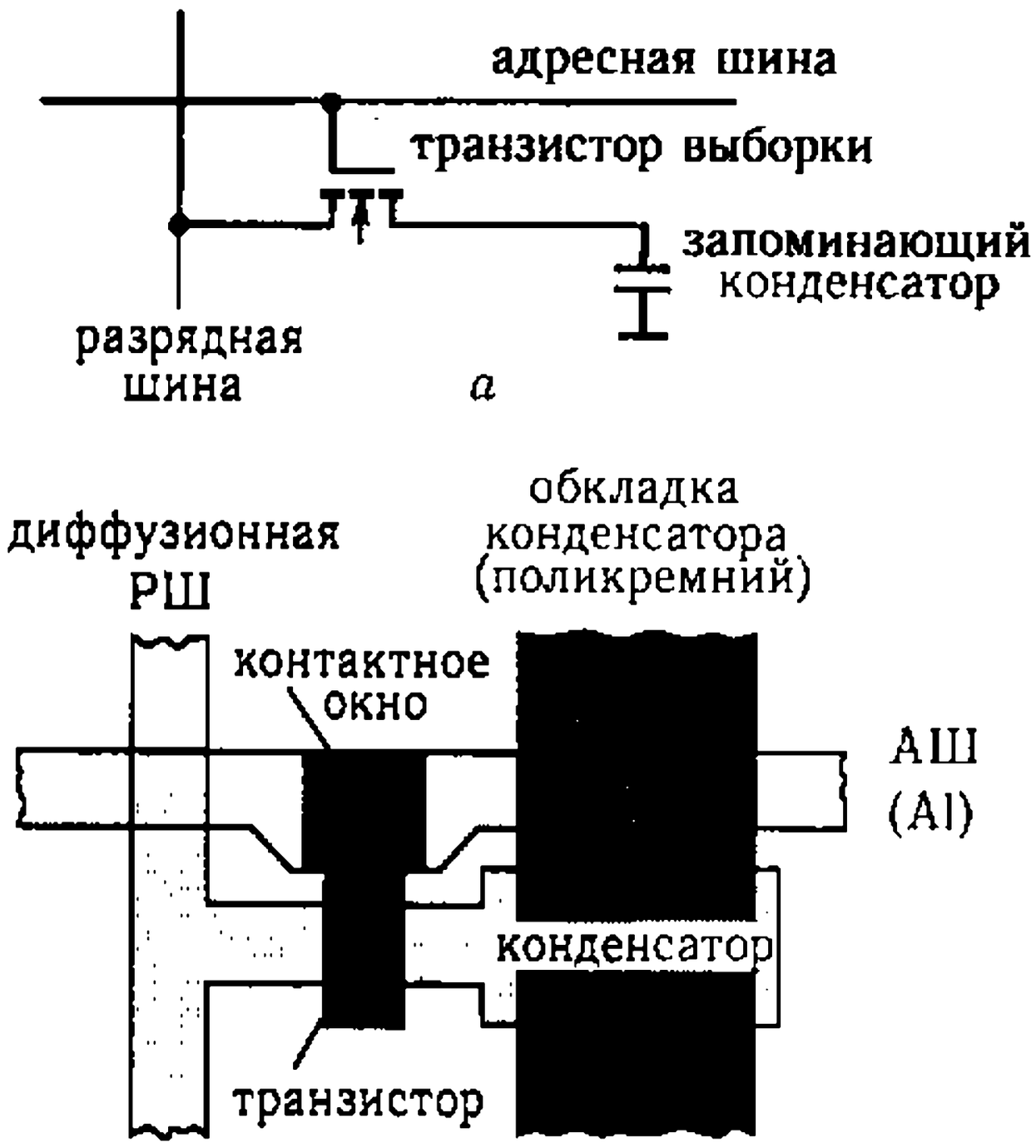
Из-за большой площади, занимаемой ячейкой статического запоминающего устройства, создание ОЗУ большой емкости на основе таких ячеек оказывается экономически неэффективным. Кроме того, статические запоминающие устройства, построенные на *n*-МОП-транзисторах, *обычно потребляют значительную* мощность. По этим причинам для создания ОЗУ большой емкости сотрудниками фирм IBM и Intel была разработана конструкция ячейки *динамической памяти*.

Работа динамических запоминающих устройств [1]) основана на хранении информации в виде зарядов на запоминающих конденсаторах. Первые микросхемы с ячейкой памяти, построенной на трех МОП-транзисторах (разработка фирмы Intel), были выпущены в 1970 г. (изделие Intel 1103), а микросхемы с ячейкой памяти, содержащей всего один МОП-транзистор (патент Деннарда [187] из IBM), — в 1972 г. (изделие Intel 2104). Последнее решение используется в настоящее время для производства практически всех микросхем динамических ОЗУ. Резкое уменьшение числа элементов в ячейке памяти позволяет разместить в микросхеме динамического ОЗУ намного больше ячеек, чем в микросхеме статического ОЗУ.

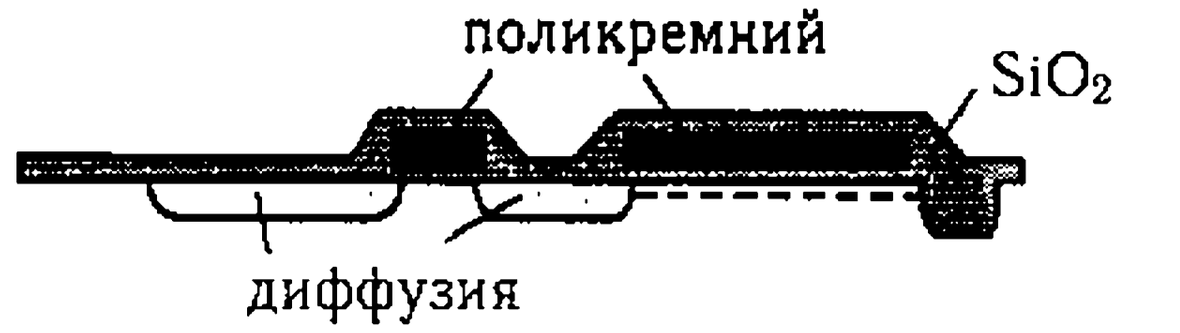
На рис. 4.20 показана электрическая схема и два варианта конструкции однотранзисторной ячейки динамического ОЗУ на полевом транзисторе с каналом *n*-типа. При подаче сигнала логической 1 на адресную шину транзистор открывается и конденсатор подключается к разрядной шине. При операции записи на разрядной шине устанавливается требуемый потенциал и конденсатор заряжается через открытый транзистор, а при операции чтения перетекающий на разрядную шину заряд с конденсатора считывается зарядочувствительным усилителем.

---

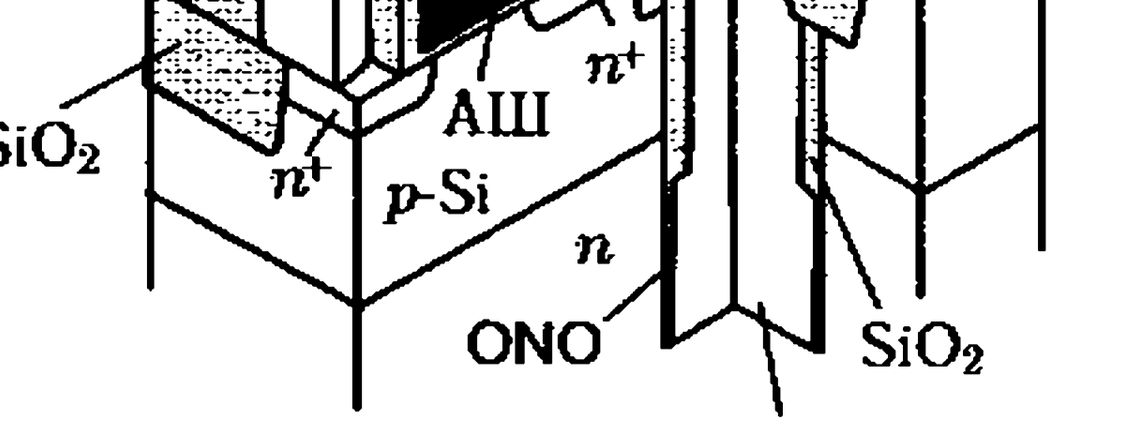
[1]) В зарубежной литературе для обозначения этого типа устройств используется аббревиатура DRAM (dynamic random access memory).



а



б



в

г

Рис. 4.20. Электрическая схема однотранзисторной ячейки памяти динамического ОЗУ на n-МОП-транзисторе (*а*); топология и поперечное сечение одной из первых конструкций ячейки (*б, в*) [58], конструкция ячейки, используемой в динамических ОЗУ емкостью 256 Мбит (*г*) [188]

Недостатком всех конструкций динамических ОЗУ является стекание заряда с конденсатора через обратно смещенный p–n-переход сток–подложка и по каналу закрытого транзистора [1]. Поэтому в запоминающих устройствах этого типа приходится проводить периодическую *регенерацию* зарядов на конденсаторах (с периодичностью от 1 до 100 мс), для чего в конструкции микросхем предусмотрена специальная схема регенерации. Поскольку при каждой операции чтения данных заряд с конденсатора перетекает на разрядную шину, имеющую существенно бóльшую емкость, и напряжение на конденсаторе уменьшается в ~10 раз, то после каждой операции чтения также необходимо проводить регенерацию (считывание информации является разрушающим). Из-за невысоких уровней сигнала и необходимости регенерации динамические ОЗУ имеют примерно в 10 раз более низкое быстродействие по сравнению со статическими ОЗУ. Достоинствами динамических ОЗУ являются низкая потребляемая ими мощность (основное потребление энергии происходит в момент зарядки конденсаторов) и заметно более низкая стоимость в расчете на единицу объема информации (по сравнению со статическими ОЗУ).

Одной из проблем, появившихся при разработке микросхем динамических ОЗУ большой емкости, является то, что с увеличением числа ячеек возникают все бóльшие трудности в размещении запоминающих конденсаторов. Разработчики считают, что независимо от размеров запоминающего элемента емкость конденсатора должна оставаться практически на одном уровне (~ ~0,04 пФ). Чтобы создать такой конденсатор с рабочим напряжением 1,5 В и диэлектриком из SiO_2 (толщина которого определяется максимально допустимым туннельным током утечки диэлектрика), необходима площадь обкладок ~5 мкм^2. Трудность в размещении конденсаторов возникла уже в микросхемах емкостью 4 Мбит, площадь ячейки памяти в которых составляла ~10 мкм^2; в современных микросхемах эта площадь уменьшилась почти в 100 раз. Поэтому в середине 80-х годов разработчики фирм Hitachi, Texas Instruments и IBM предложили формировать

---

[1]) Обратный ток p–n-перехода в заметной степени определяется током неосновных носителей, возбуждаемых в объеме подложки. Для уменьшения этого тока транзистор изготавливают в эпитаксиальном слое, выращенном на сильно легированной подложке, в которой концентрация неосновных носителей мала [58]. Утечка по каналу МОП-транзистора определяется подпороговым током транзистора, о котором мы говорили на с. 256. Этот канал утечки становится особенно заметным в настоящее время в связи с уменьшением напряжения питания микросхем памяти.

запоминающие конденсаторы в объеме кремниевой пластины в специальных глубоких «канавках», создаваемых методом реактивного ионного травления [188] (конструкция *trench capacitor*). В этой конструкции (см. рис. 4.20,*г*) обкладками конденсатора служат поликремний n^+-типа, заполняющий внутреннюю часть канала, и подложка n-типа проводимости. Изолятором в конденсаторе служит трехслойный композит оксид–нитрид–оксид (ONO) толщиной 40–60 Å; для уменьшения напряженности электрического поля в диэлектрике на подложку n-типа подается потенциал, равный полусумме уровней сигнала, отвечающих логическому 0 и 1. Глубина каналов составляет 4–8 мкм при диаметре 0,2–0,5 мкм. В 2001 г. площадь одной ячейки памяти в ОЗУ объемом 256 Мбит достигла 0,16 мкм^2 [189]. Микросхемы с конденсаторами такого типа выпускаются в настоящее время фирмами IBM, Toshiba, Infineon.

Другим решением проблемы размещения запоминающего конденсатора было решение располагать его *над* ячейкой памяти (конструкция *stacked capacitor*); микросхемы с такой конструкцией выпускаются в настоящее время фирмами Samsung, Hynix, Micron, NEC.

Пока не ясно, какая из двух описанных конструкций является более перспективной с точки зрения дальнейшего уменьшения размеров ячейки памяти: в конструкции stacked capacitor площадь и емкость конденсатора изменяются пропорционально площади ячейки, в то время как в конструкции trench capacitor при уменьшении площади остается возможность сохранить емкость конденсатора за счет углубления каналов и придания им бутылкообразной формы. С другой стороны, в конструкции stacked capacitor имеется возможность создавать многослойные конденсаторы или другие конструкции конденсаторов с развитой поверхностью.

Оба описанных подхода становятся малопригодными для микросхем емкостью $\geqslant 1$ Гбит из-за чрезвычайно малой площади, выделяемой под запоминающий элемент. Одним из путей решения проблемы может быть замена композитного диэлектрика ONO, используемого в запоминающих конденсаторах, на материал *с более высокой диэлектрической проницаемостью* [174, 175]. Считается, что это решение лучше совместимо с конструкцией stacked capacitor, хотя в последнее время разработаны эффективные методы внедрения различных диэлектриков и в узкие каналы в конструкции trench capacitor. Свойства некоторых предлагаемых для этой цели диэлектриков представлены в табл. 1 в Приложении. В конце 2004 г. фирма Samsung уже

приступила к массовому производству динамических ОЗУ с диэлектрическим слоем из Al_2O_3–HfO_2.

4.2.2. КМОП-структуры. По мере усложнения ИС и увеличения их уровня интеграции постепенно возрастала потребляемая микросхемами мощность. В то же время стало появляться все больше задач, которые требовали разработки ИС со сверхнизким потреблением энергии [1]. Решением проблемы создания высокоэкономичных ИС явилась реализация предложенной в 1963 г. идеи *комплементарной МОП-структуры (КМОП)*, в которой одновременно используются МОП-транзисторы с каналами *n*- и *p*-типа [190]. Простые логические интегральные схемы с КМОП-структурой (серия CD4000 фирмы RCA) появились в 1968 г., в 1971 г. были разработаны ИС для наручных часов (Epson/Seiko) и переносных калькуляторов (Toshiba), а в 1982 г. был выпущен первый процессор ЭВМ на КМОП БИС (80286 фирмы Intel). Именно с этого момента КМОП-структуры становятся основным направлением производства ИС на МОП-транзисторах: так, вплоть до самого последнего времени фирма Intel изготавливает свои процессоры по КМОП-технологии.

Рассмотрим устройство КМОП-структуры на примере наиболее простого логического элемента — инвертора. Электрическая схема инвертора показана на рис. 4.21. Основной идеей КМОП-структуры является последовательное включение МОП-транзисторов с каналами *p*- и *n*-типа. Концентрация и тип легирующей примеси, вводимой ионной имплантацией в приповерхностный слой кремния для изменения $V_{пор}$, выбираются так, чтобы при подаче на соединенные вместе затворы транзисторов любого напряжения транзисторы никогда бы не оказывались одновременно в открытом состоянии. Так, при подаче на вход инвертора напряжения, отвечающего уровню логического 0 ($U_0 \approx 0$) *n*-канальный транзистор закрыт, а *p*-канальный транзистор открыт, что формирует на

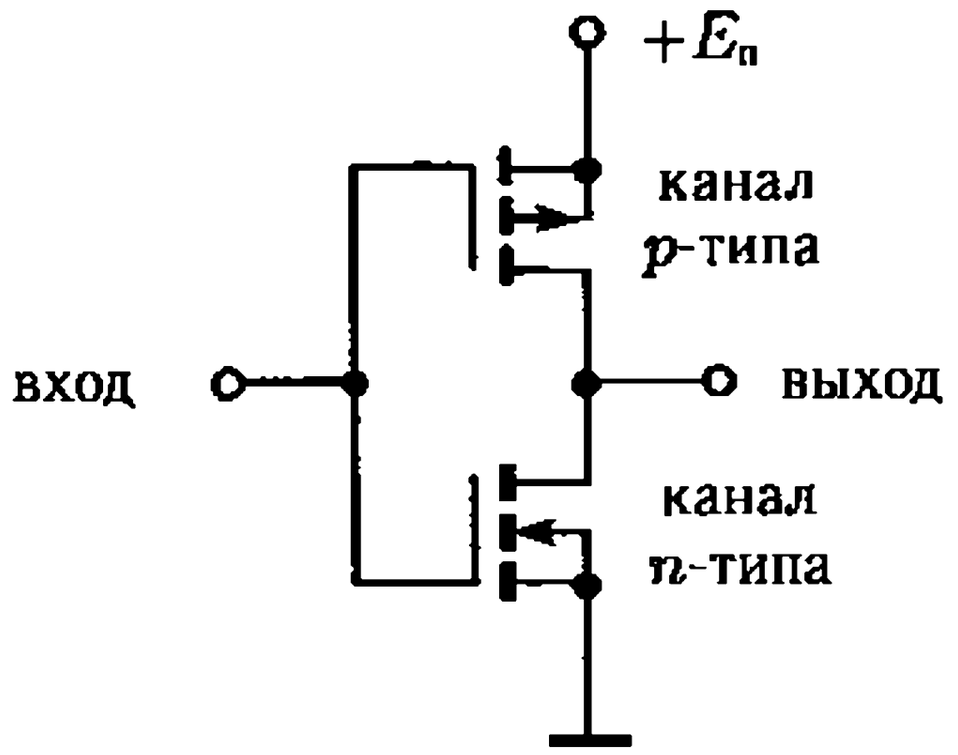


Рис. 4.21. Логический элемент (инвертор) на основе КМОП-структуры.

[1]) Примером таких задач может служить создание микросхем для переносных устройств с батарейным питанием, например, для наручных часов, элемент питания которых надо менять не чаще, чем раз в год.

выходе схемы напряжение, отвечающее логической 1 ($U_1 \approx E_п$). При подаче на вход схемы уровня логической 1 p-канальный транзистор закрыт, а n-канальный — открыт, что формирует на выходе схемы уровень логического 0. Поскольку один из транзисторов всегда закрыт, то через логический элемент в статическом режиме протекает лишь малый ток утечки p–n-переходов, образованных областями стока и истока с подложкой. Заметный ток через структуру протекает лишь в момент переключения элемента, так как при этом происходит перезаряд-ка емкости выходной цепи до напряжения, отвечающего новому логическому уровню. Соответствующее время переключения определяется произведением этой емкости на сопротивление канала открытого транзистора, которое обычно лежит в пределах 0,1–10 кОм. Недостатками КМОП-структур по сравнению со схемами на n-МОП-транзисторах являются более высокая входная емкость, связанная с необходимостью подключения каждого входа логического элемента одновременно к двум транзисторам, и примерно вдвое бóльшая площадь, занимаемая логическим элементом.

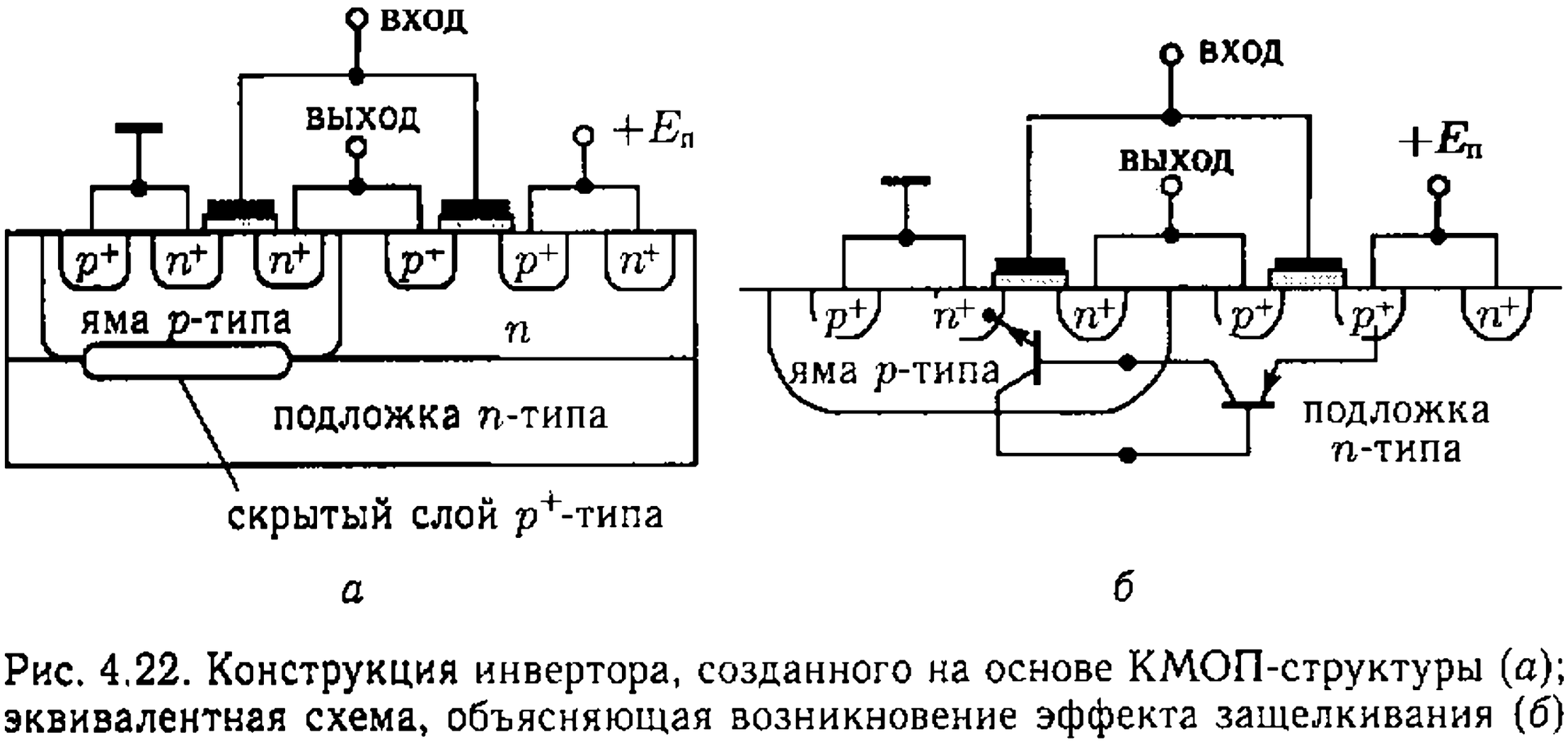


Рис. 4.22. Конструкция инвертора, созданного на основе КМОП-структуры (*а*); эквивалентная схема, объясняющая возникновение эффекта защелкивания (*б*)

Поперечное сечение КМОП-инвертора показано на рис. 4.22*а*. Поскольку подложка из Si и эпитаксиальный слой имеют n-тип проводимости, то на этой подложке можно создать только транзисторы с каналом p-типа. Для получения транзисторов с каналом n-типа в эпитаксиальном слое ионной имплантацией или диффузией формируются так называемые *карманы* (*ямы*) p-типа проводимости, в которых и создаются эти транзисторы. Подробнее о технологии изготовления КМОП-структур можно прочитать в книге [58].

Основной проблемой, присущей описанной конструкции КМОП-структуры, является возможность возникновения в ней так называемого *эффекта защелкивания* (тиристорного эффекта). Оказывается, что если подать (даже кратковременно) в выходную цепь этой структуры напряжение, выходящее за пределы $0-E_п$, то один из стоковых $p-n$-переходов может открыться и перевести структуру в состояние, при котором через нее протекает большой ток, грозящий вывести ее из строя [1]). Причину возникновения эффекта защелкивания объясняет эквивалентная схема, показанная на рис. 4.22*б*. В КМОП-структуре можно выделить два паразитных биполярных транзистора: вертикальный $n-p-n$-транзистор, образованный истоком n-канального транзистора, ямой p-типа и подложкой n-типа, и горизонтальный $p-n-p$-транзистор, образованный истоком p-канального транзистора, подложкой n-типа и ямой p-типа. В эквивалентной схеме эти транзисторы включены по схеме, аналогичной тиристору (ср. с рис. 3.4). Как мы знаем, эта схема характеризуется сильной положительной обратной связью и поэтому открывание любого из указанных транзисторов будет с неизбежностью вызывать открывание другого. Для уменьшения возможности появления эффекта защелкивания в создаваемые структуры добавляют дополнительные элементы (см. рис. 4.22*а*): скрытый слой p^+-типа, который создает дополнительный барьер для движения электронов в базе $n-p-n$-транзистора и уменьшает его коэффициент усиления, и дополнительные защитные кольца (p^+-область в кармане p-типа и n^+-область в эпитаксиальном слое), которые шунтируют выводы баз с эмиттерами обоих паразитных транзисторов и тем самым уменьшают возможность инжекции эмиттерами неосновных носителей.

Следует отметить, что найденное при разработке КМОП-структур технологическое решение — формирование МОП-транзисторов в кармане, имеющем другой тип проводимости, — в настоящее время используется и при создании других приборов: микросхем динамической памяти, приборов с зарядовой связью. Дело в том, что уменьшая объем полупроводника, расположенного под каналом транзистора, можно существенно уменьшить токи утечки $p-n$-переходов в этих приборах и таким способом добиться улучшения их характеристик.

---

[1]) Причиной выхода микросхемы из строя может оказаться даже электростатический разряд емкости проводника, подключаемого к выходу микросхемы.

Структуры кремний-на-изоляторе. Идеальным способом устранения эффекта защелкивания в КМОП-структурах является формирование транзисторов в тонких эпитаксиальных слоях кремния, выращенных на изолирующих подложках. Классическим примером такой эпитаксиальной структуры может служить *кремний-на-сапфире* (КНС). В структурах КНС путем травления или локального окисления удается полностью изолировать транзисторы с каналами p- и n-типа друг от друга, а низкая диэлектрическая проницаемость подложки позволяет существенно уменьшить паразитные емкости и при неизменном энергопотреблении почти вдвое повысить быстродействие микросхем [141]. Однако стоимость ИС, изготовленных по технологии КНС, заметно выше стоимости микросхем, изготовленных обычным способом, и поэтому их применение ограничивается специальными областями (например, высокая радиационная стойкость этих микросхем позволяет использовать их в условиях повышенной радиации в космосе).

Альтернативой дорогостоящей структуре КНС может служить структура, в которой тонкий поверхностный слой кремния отделен от кремниевой подложки слоем окисла («захороненным» слоем) [141]. Эту структуру не слишком правильно называют *кремний-на-изоляторе* (КНИ) [1].

Структуру кремний-на-изоляторе можно создать двумя способами. Первый способ основан на использовании предложенного в середине 60-х годов метода синтеза химических соединений с помощью ионной имплантации. В этом способе, называемом процессом SIMOX — *separation by implanted oxygen*, пластины кремния, нагретые до 600 °С, подвергаются имплантации ионами кислорода с энергией 180–200 кэВ и дозой $2 \cdot 10^{18}$ см$^{-2}$. После отжига при 1300 °С в имплантированной области возникает однородный слой SiO_2 толщиной около 0,4 мкм, на поверхности которого располагается слой монокристаллического кремния толщиной 0,2 мкм. На таких подложках в 1978 г. были созданы первые КМОП-структуры [191], а первые промышленные процессоры (PowerPC 750 фирмы IBM) появились в 1998 г. [192].

Другим способом создания структуры КНИ является метод сращивания–отслаивания (технология Smart Cut). В этом методе предварительно окисленная пластина кремния подвергается имплантации ионами H^+ с энергией 50 кэВ и дозой $(4–8) \cdot 10^{16}$ см$^{-2}$. После имплантации водород находится в решетке в состоянии пересыщенного твердого раствора, а также входит в состав вакансионно-водородных комплексов. При нагревании до 400–600 °С в результате распада

[1]) В зарубежной литературе для обозначения этого типа структур используется аббревиатура SOI (silicon on insulator).

твердого раствора и комплексов в объеме пластины образуется слой пористого кремния, заполненный газообразным водородом под высоким давлением, по которому верхний тонкий слой кремния может быть легко отслоен от объема пластины. Для переноса этого тонкого слоя на пластину-подложку поступают следующим образом: подложку и имплантированную пластину прижимают друг к другу и прогревают при 600 °С, после чего пластины расщепляют, причем тонкий слой остается на пластине-подложке. Затем полученная структура подвергается термообработке при 1100 °С, при которой происходит окончательное сращивание слоев кремния по слою и одновременно отжигаются радиационные дефекты в кремнии. Готовую структуру КНИ получают после удаления шероховатого слоя на отщепленной поверхности кремния.

Несмотря на то, что доза имплантации в технологии Smart Cut намного меньше, чем в процессе SIMOX, стоимость пластин, изготавливаемых по этой технологии, пока выше, чем по методу SIMOX.

Большое значение структуры КНИ связано с тем, что в настоящее время одним из существенных препятствий на пути создания сложных быстродействующих ИС (типа процессоров для ЭВМ) стала проблема отвода тепла от кристалла. Например, процессор Pentium-4 фирмы Intel с ядром Prescott, созданный по традиционной технологии, потребляет мощность около 100 Вт при рабочей частоте 3,2 ГГц. При этом близкий по производительности процессор PowerPC 970 фирмы IBM, изготовленный по технологии КНИ, потребляет всего 25 Вт. Возможность значительно уменьшить энергопотребление ИС при сохранении быстродействия (или еще больше повысить быстродействие) за счет уменьшения паразитных емкостей стимулирует все возрастающий интерес к структурам кремний-на-изоляторе [192, 193].

Уменьшение удельной емкости обедненного слоя в структурах КНИ позволяет добиться почти идеальной зависимости подпорогового тока в транзисторах ($m \approx 1$ в формуле (4.19)) и дополнительно понизить напряжение питания ИС, а возможность уменьшения толщины боковой окисной изоляции позволяет почти вдвое увеличить плотность упаковки. По этим причинам технология кремний-на-изоляторе считается весьма перспективной и предполагается, что к 2010 году около 10% кремниевых подложек будут изготавливаться по этой технологии. В настоящее время самые производительные современные 64-разрядные процессоры PowerPC 970 фирмы IBM и процессоры Opteron и Athlon 64 фирмы AMD изготавливаются по технологии КНИ.

Структуры BiCMOS. В конце 60-х годов у разработчиков появилась идея объединить в одной микросхеме все лучшее,

что позволяет достигнуть технология биполярных ИС и технология КМОП-структур. Первыми «гибридными» структурами, полученными в результате такого объединения и получившими название BiCMOS-структур (bipolar CMOS — биполярные КМОП-структуры), были операционные усилители, которые были разработаны фирмой RCA в 1973 г. В это время логические КМОП ИС работали при напряжении питания 5 В и имели разность напряжений логических уровней U_{0-1} в несколько раз выше, чем биполярные ИС. Поэтому для уменьшения динамической мощности рассеяния (см. с. 269) наиболее быстродействующие части цифровых ИС в это время было выгоднее создавать на основе биполярных приборов, а более медленные, но требующие экономии энергии схемы — на КМОП-структурах.

В дальнейшем, когда в начале 90-х годов напряжение питания КМОП-ядра микросхем начало быстро снижаться, различие между биполярными и КМОП-логическими элементами нивелировалось, и в это время биполярные транзисторы начинают все чаще применяться в периферийных узлах ИС для повышения их нагрузочной способности (выходного тока). В эти годы по технологии BiCMOS были созданы процессоры Pentium и PentiumPro фирмы Intel. После некоторого периода охлаждения в настоящее время интерес к BiCMOS-структурам вновь начал возрастать в связи с развитием таких приложений как телекоммуникация и мобильная связь, которые требуют размещать на одном кристалле и быстродействующие логические схемы (сейчас их выгоднее создавать на КМОП-структурах), и высокочастотные устройства и аналоговые схемы (для них лучше подходят биполярные транзисторы) [193].

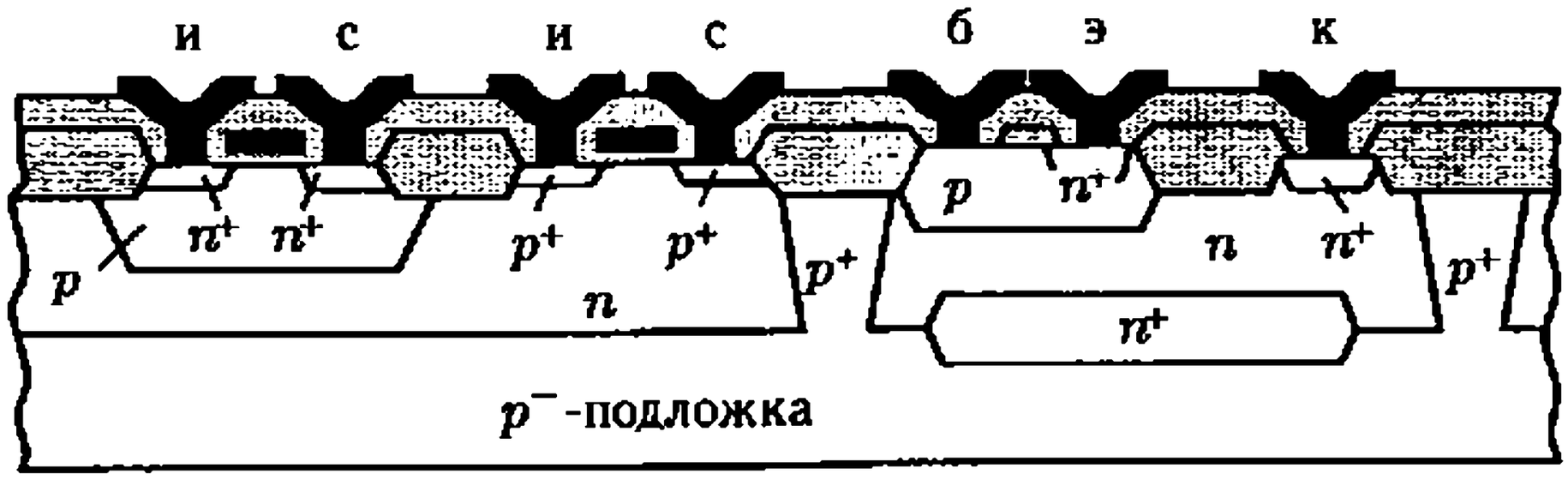


Рис. 4.23. Совмещение n-канальных, p-канальных МОП-транзисторов с биполярными транзисторами в структуре BiCMOS [141]

На рис. 4.23 показан фрагмент BiCMOS-структуры, на которой одновременно размещены p- и n-канальные МОП-

транзисторы и биполярный n–p–n-транзистор. Видно, что совмещение этих транзисторов в рамках одной структуры действительно не представляет большого труда с точки зрения технологии изготовления ИС, поскольку их создание основывается на одних и тех же приемах планарной технологии (см. п. 2.8.1).

4.2.3. Энергонезависимые постоянные запоминающие устройства на МОП-транзисторах.

Полевые транзисторы с плавающим затвором. В 1967 г. Канг и Зи [194] показали, что если в структуру МОП-транзистора между каналом и затвором добавить еще один элемент — изолированный от других частей структуры *плавающий затвор*, — то получится новый полупроводниковый прибор — энергонезависимый элемент памяти, который может сохранять записанную в него информацию (в виде заряда на плавающем затворе) при отключении источника питания. Эта идея лежит в основе большинства выпускаемых в настоящее время интегральных схем *репрограммируемых* (перепрограммируемых) постоянных запоминающих устройств (РПЗУ) [1]).

Первая конструкция МОП-транзистора с плавающим затвором показана на рис. 4.24*а*. Перемещение заряда между подложкой и плавающим затвором в ней осуществляется за счет туннелирования электронов сквозь тонкий слой SiO_2 в сильном (8–10 МВ/см) электрическом поле, создаваемом верхним (управляющим) затвором. В конструкции Канга и Зи слой диэлектрика между плавающим затвором и подложкой имел толщину $d_1 = 50$ Å и из-за недостаточно высокого качества окисла, который умели получать в то время, ток утечки затвора был слишком велик (заряд на затворе сохранялся всего несколько часов [194]).

Попытка увеличить толщину окисла приводила к неприемлемо высоким напряжениям, которые требовались для программирования ячейки. Поэтому в 1971 г. Фроман-Бенчковский [195] предложил другой подход: использовать структуру с толстым слоем окисла и программировать ее *горячими электронами*, образующимися при лавинном пробое перехода сток–подложка. Микросхемы на таких транзисторах были выпущены фирмой Intel в том же году и сразу же нашли широкое применение в качестве постоянных запоминающих устройств (ПЗУ) ЭВМ.

---

[1]) В зарубежной литературе для обозначения этого типа устройств используется аббревиатура EPROM (erasable programmable read only memory).

В отличие от выпускавшихся тогда однократно программируемых ПЗУ, эти запоминающие устройства допускали возможность их многократного перепрограммирования.

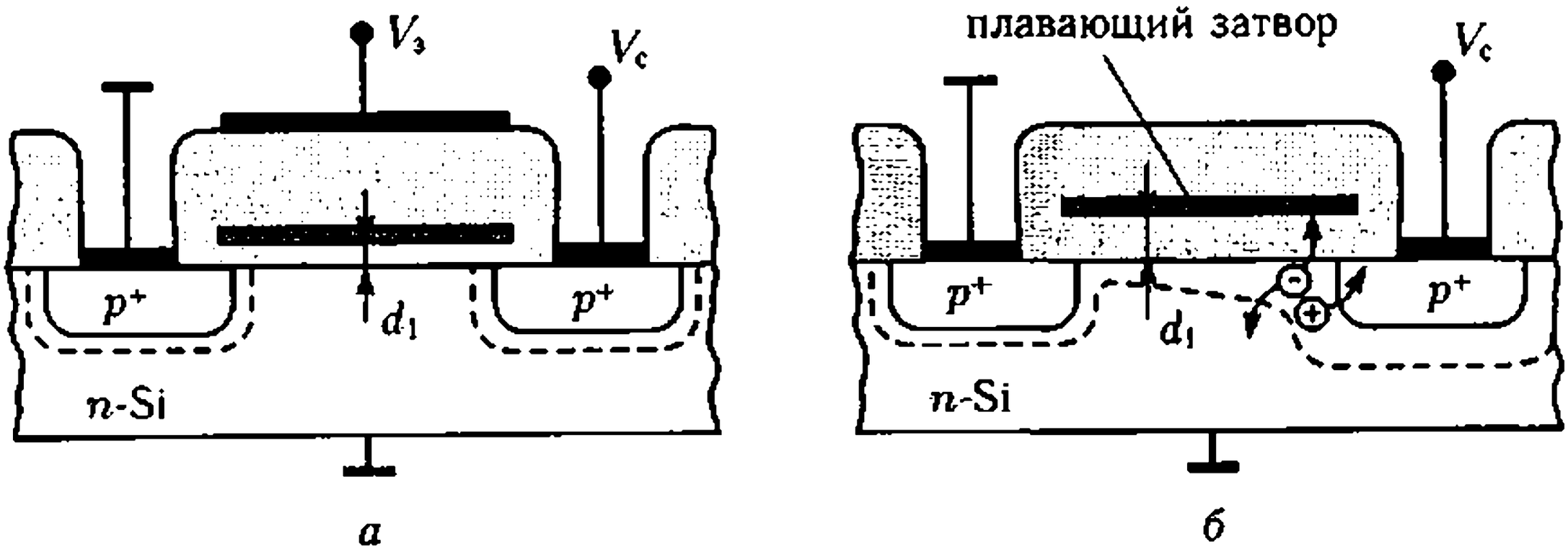


Рис. 4.24. Ячейка МОП-транзистора с двумя затворами, обеспечивающая возможность электрического перепрограммирования (*а*), и ячейка с плавающим затвором, программируемая лавинной инжекцией заряда (*б*)

Устройство ячейки памяти в конструкции Фромана-Бенчковского, названной ЛИЗМОП-структурой (МОП-транзистор с плавающим затвором и лавинной инжекцией заряда) [1], показано на рис. 4.24 *б*. Плавающий затвор из поликремния отделен от подложки из n-Si слоем окисла толщиной $d_1 \approx 1200$ Å и закрыт сверху окислом толщиной около 1 мкм. Второй (управляющий) затвор в структуре не требуется. Пороговое напряжение такого транзистора с каналом p-типа при нулевом заряде на плавающем затворе (состояние логического 0) составляет $V_{пор} \approx 8$ В. Для записи в ячейку памяти логической 1 используются горячие электроны, возникающие при лавинном пробое p–n-перехода между стоком и подложкой, энергия которых достаточна для преодоления потенциального барьера высотой 3,2 эВ между краями зоны проводимости в Si и SiO_2. Выбранная толщина окисла такова, что за счет термической активации и туннелирования за 10 лет хранения при температуре 125 °Сзаряд на затворе уменьшается всего на ~30% [195]. Попавший на затвор отрицательный заряд снижает $V_{пор}$ на 10–15 В и транзистор переходит в проводящее состояние. Характерное время записи информации в одну ячейку памяти составляет 5 мс. Для стирания информации (перепрограммирования) микросхему

[1]) В зарубежной литературе для обозначения этого типа структур используется аббревиатура FAMOS (floating-gate avalanche-injection MOS).

освещают ультрафиолетовым излучением от ртутной лампы в течение нескольких минут через расположенное на поверхности микросхемы кварцевое окно. При этом за счет фотоэмиссии электроны с затвора переходят в подложку и все ячейки, которые находились в состоянии логической 1, возвращаются в состояние логического 0. Недостатками ЛИЗМОП-структур являются: высокая продолжительность операций стирания и записи, невозможность изменения записанной информации без ее стирания во всей микросхеме, быстрая деградация слоя окисла под действием ультрафиолетового излучения и горячих электронов. Как показали исследования, последние могут не только захватываться на дефекты в слое окисла (изменяя $V_{пор}$), но также и создавать новые дефекты. Это ограничивает максимальное число циклов перепрограммирования ЛИЗМОП-структур величиной $\sim$100.

В настоящее время МОП-транзисторы с плавающим затвором вернулись к первоначальной конструкции ячейки (см. рис. 4.24 *а*). В ней используется высококачественный окисел толщиной $d_1 = 90$–100 Å, а изолирующий слой между двумя затворами имеет толщину около 200 Å и представляет собой композитную структуру ONO (оксид–нитрид–оксид) [1]. Эти изменения позволяют создать постоянное запоминающее устройство, стирание и запись информации в котором осуществляется *чисто электрическим способом* (ЭППЗУ — электрически перепрограммируемое ПЗУ) [2]. Для программирования ячейки на верхний (управляющий) затвор подают повышенное напряжение (12–20 В) и, в зависимости от его полярности, электроны туннельным способом переходят из подложки на плавающий затвор или в обратном направлении. Время записи логической 1 в такую ячейку памяти зависит от толщины окисла d_1 и изменяется в пределах 1 мкс–10 мс [3], а время записи логического 0

---

[1] Использование композитной структуры ONO позволяет понизить температуру изготовления второго слоя диэлектрика, поскольку высокая температура отрицательно влияет на качество созданного ранее подзатворного окисла. Кроме того, из-за высокой концентрации ловушек в Si_3N_4 эта структура обладает способностью «самозалечивания» дефектов (проколов в тонких слоях окисла).

[2] В зарубежной литературе для обозначения этого типа устройств используются аббревиатуры EEPROM или E^2PROM (electrically erasable programmable read only memory).

[3] Скорость записи информации в ячейку памяти можно увеличить, уменьшая толщину окисла. Однако при этом ухудшаются время сохранения информации и деградационная стойкость транзисторов.

(стирания информации) — 10–500 мс. Важным преимуществом микросхем ЭППЗУ по сравнению с ЛИЗМОП-структурами является возможность изменения информации в одной выбранной ячейке памяти без стирания информации во всей микросхеме, а также существенно более медленная деградация окисла (микросхемы ЭППЗУ выдерживают 10^4–10^5 циклов перепрограммирования).

Исследования показали, что деградация микросхем РПЗУ при многократном перепрограммировании связана с двумя явлениями: уменьшением разности пороговых напряжений, отвечающих уровням логического 0 и 1, и резким возрастанием тока утечки плавающих затворов в отдельных ячейках памяти [196]. Причинами первого из этих изменений являются захват электронов на глубокие ловушки в окисле и рождение под действием горячих носителей в объеме окисла и на его границах с полупроводником новых ловушек. Возрастание тока утечки в слабом электрическом поле после приложения к диэлектрику сильного электрического поля — хорошо известное явление. Причиной его является перемещение заряженных ионов в диэлектрике и такое перераспределение зарядов на ловушках, при котором в диэлектрике появляются локальные области с заметно более высокой напряженностью электрического поля. Более высокая вероятность туннелирования в этих областях и приводит к быстрому разряду плавающих затворов. Замечено, что последнее явление выражено тем сильнее, чем тоньше диэлектрик, и поэтому толщина подзатворного диэлектрика в микросхемах ЭППЗУ выбирается не из условия получения заданной величины туннельного тока утечки, а из условия получения разумного числа циклов перепрограммирования.

Следует заметить, что деградация ячеек памяти происходит при программировании их не только горячими электронами, но при использовании некоторых туннельных механизмов программирования. В частности, заметная деградация наблюдается при программировании ячеек с помощью туннелирования по механизму Фаулера–Нордгейма [28], при котором носители, туннелируя сквозь потенциальный барьер, сначала попадают в зону проводимости SiO_2, а затем уже на плавающий затвор. Напротив, при *прямом* туннелировании между полупроводниковой подложкой и поликремниевым плавающим затвором деградационные явления практически отсутствуют. К сожалению, прямое туннелирование проявляется только при толщине окисла менее 50 Å; при такой толщине окисла время сохранения информации на плавающем затворе получается слишком малым для практического использования в РПЗУ.

Примером отечественных микросхем ПЗУ с ЛИЗМОП-структурой и ультрафиолетовым стиранием могут служить микросхемы серий К573 и К1626, а ПЗУ с электрической перезаписью информации — микросхемы серии К1609.

Флэш-память. Транзисторы со структурой, показанной на рис. 4.24 *а*, являются основой построения флэш-памяти [1]), идея которой была предложена в работе [197], а первые микросхемы выпущены фирмой Toshiba в 1985 г. В настоящее время этот тип микросхем памяти практически вытеснил другие типы РПЗУ. По объему выпуска флэш-память занимает сейчас второе место после микросхем динамической памяти и широко используется в вычислительной технике, цифровой фотографии, мобильных телефонах, MP3-плеерах, электронных играх и т. д. [196].

Отличие флэш-памяти от рассмотренных выше традиционных типов РПЗУ состоит в изменении схем управления и упрощении соединений между элементами, которые позволили уменьшить размер ячеек, увеличить плотность упаковки, а также ускорить операцию стирания информации (наиболее медленную операцию в традиционных ЭППЗУ), которая теперь осуществляется одновременно в больших блоках памяти или во всей микросхеме.

Существуют два основных способа соединения ячеек в микросхемах флэш-памяти: по логике NOR и логике NAND. В микросхемах с логикой NOR ячейка памяти содержит один полевой транзистор с плавающим затвором, программирование которого обычно осуществляется инжекцией горячих электронов, создаваемых *в канале* транзистора при приложении между его истоком и стоком высокого напряжения, а стирание — туннельным способом (как в ЭППЗУ). При возбуждении одной из адресных шин АШ (см. рис. 4.25 *а*) информация о состоянии соответствующего транзистора передается на общую разрядную шину РШ, с помощью которой и реализуется логическая функция NOR («ИЛИ-НЕ»). Для увеличения плотности упаковки выводы истоков и стоков соседних ячеек памяти объединяются попарно. Характерное время записи в ячейку памяти составляет 10 мкс, время стирания — сотни мс.

Микросхемы с логикой NOR характеризуются невысокой плотностью упаковки, однако организация микросхемы допускает чтение и запись с произвольным доступом. Поэтому эти микросхемы оптимальны для хранения исполняемых компьютерных программ. Ячейки памяти выдерживают 10^4–10^5 циклов перепрограммирования.

---

[1]) Название этого типа микросхем (flash EEPROM) происходит от английского выражения *in a flash* — *в мгновение ока*, которое правильно отражает основную особенность этих ПЗУ — возможность быстрого стирания информации во всей микросхеме.

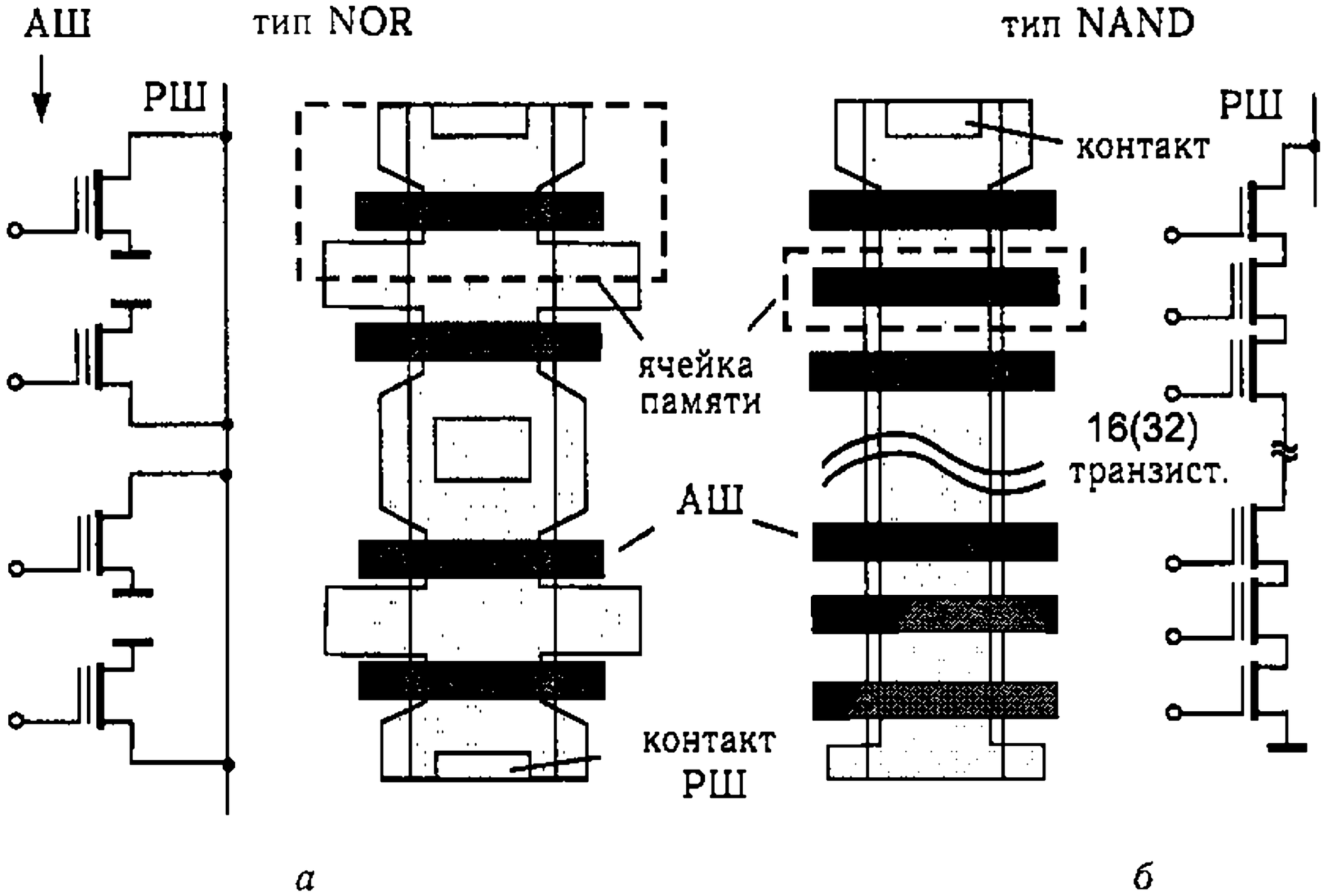


Рис. 4.25. Соединение ячеек памяти в микросхемах флэш-памяти с логикой NOR и NAND. При топологическом размере 0,13 мкм размер ячеек в микросхемах фирмы Samsung составляет 0,3×0,6 и 0,39×0,26 мкм, соответственно

В микросхемах с логикой NAND ячейки памяти объединены в цепочки, состоящие из 16–32 последовательно включенных полевых транзисторов с плавающим затвором, выводы истока и стока у которых создаются только на краях цепочек (см. рис. 4.25*б*). Отсутствие промежуточных выводов истока и стока у транзисторов внутри цепочек позволяет заметно увеличить плотность упаковки элементов в этих микросхемах. Поэтому типичная емкость микросхем с логикой NAND оказывается примерно на порядок выше, чем для микросхем с логикой NOR. Для адресации заданной ячейки памяти на все адресные шины кроме одной подается напряжение высокого уровня, соответствующие транзисторы открываются и по разрядной шине считывается информация о величине заряда на плавающем затворе «невыбранного» транзистора; таким образом в микросхеме реализуется функция NAND («И-НЕ»). Ячейки памяти в этих микросхемах стираются и программируются туннельным способом, поэтому время записи в эти ячейки заметно больше (~100 мкс), чем при программировании горячими электронами. Однако за счет параллельной записи в ячейки памяти целой страницы данных (4096+128 бит) скорость записи в микросхему больших объемов информации оказывается даже выше, чем в микросхемах

с логикой NOR. Поскольку работа микросхем с логикой NAND основана на вводе–выводе страниц данных, в них трудно реализовать произвольный доступ к ячейкам памяти и невозможна побайтная запись. Оптимальная область применения этих микросхем — хранение массивов данных, считываемых блоками по 512–2048 байт. Эти микросхемы выдерживают примерно на порядок большее число циклов перепрограммирования (10^5–10^6) по сравнению с микросхемами с логикой NOR.

В последнее время появились модификации микросхем флэш-памяти, сочетающие в себе лучшие достоинства микросхем с логикой NOR и NAND. Устройства DiskOnChip, DiskOnModule и карты памяти Compact Flash содержат в себе контроллер интерфейса ATA, позволяющий непосредственно использовать эти устройства в ЭВМ в качестве электронных аналогов жестких дисков емкостью до 32 Гбайт (по состоянию на 2006 г.). Эти устройства широко используются в переносной аппаратуре, бортовых системах сбора и регистрации информации, отказоустойчивых системах автоматики, которые работают в условиях повышенной вибрации, влажности и запыленности.

В связи с непрерывным уменьшением размеров ячеек запоминающих устройств большой интерес в настоящее время вызывает разработка новых типов таких устройств на основе *квантовых точек* [198]. По мере уменьшения размера плавающего затвора в РПЗУ уменьшается число хранимых на нем электронов и возрастает зависимость длительности хранения на нем информации от качества окисла. Использование массивов квантовых точек в качестве плавающих затворов позволяет практически избавиться от этой зависимости [1]. Действительно, в структурах с таким затвором утечка через дефектные места в окисле будет влиять только на часть квантовых точек, почти не отражаясь на работе транзистора в целом. Для уменьшения чувствительности транзисторов к дефектам в окисле необходимо создать как можно больше квантовых точек, изолированных друг от друга. Изоляция по отношению к туннельному переносу заряда между точками достигается при расстоянии между точками более 40 Å. При уменьшении размера квантовой точки становится заметной дискретность изменения ее электростатической энергии при добавлении на нее каждого дополнительного электрона. Этот эффект препятствует помещению большого заряда на квантовую точку малого размера (*эффект «кулоновской блокады»*) и, таким образом, накладывает ограничение на минимальный размер квантовой точки. По оценкам, для эффективной работы

[1]) Создание многоэлементного плавающего затвора как способ уменьшения влияния проколов в окисле на работу запоминающего элемента предлагалось еще в патенте Канга [194].

элементов памяти диаметр квантовых точек должен быть ~5 нм, а их плотность — 10^{12} см$^{-2}$.

Поскольку получение квантовых точек указанного размера методами современной литографии пока слишком сложно, можно попробовать использовать более технологичные решения и работать с массивами *нанокристаллов* приблизительно одинакового размера и формы (см. подробнее [198]). В частности, для создания таких массивов нанокристаллов для экспериментальных образцов запоминающих устройств использовались технология выращивания островков кремния на поверхности SiO_2 в процессе эпитаксии MOCVD и явление образования преципитатов из внедренных в подзатворный окисел атомов кремния в процессе термообработки после ионной имплантации [199, 200].

Для увеличения плотности записи информации в флэш-памяти фирма Intel в 1998 г. реализовала предложенную в патенте [201] идею записывать в каждую ячейку памяти *сразу несколько битов информации*, кодируя кодовые комбинации (00, 01, 10 и 11 в случае одновременной записи двух битов) разной величиной заряда на плавающем затворе и затем считывая эту информацию путем измерения тока насыщения открытого транзистора. Эта технология хранения получила название MLC (multi-level cell) и широко используется в настоящее время фирмами Intel, Toshiba, Samsung, Hynix, Micron в производстве флэш-памяти [196]. Примером таких микросхем может служить флэш-память StrataFlash фирмы Intel. Другим подходом, также позволяющим удвоить объем флэш-памяти, является создание двух ячеек памяти под одним затвором около истока и стока транзистора; этот подход развивается фирмами AMD (конструкция MirrorBit) и Infineon (конструкция TwinFlash).

Для микросхем с логикой NAND достаточно характерным видом ошибок является искажение отдельных битов данных. Характерная частота этих ошибок составляет 10^{-9} для ячеек с двумя уровнями логических сигналов и 10^{-6} для ячеек MLC с четырьмя уровнями. Для борьбы с этими ошибками при записи каждого блока данных в микросхему одновременно записывается несколько байтов ECC (кода, исправляющего ошибки); при чтении данных этот код используется для проверки правильности считанной информации и при необходимости искаженные биты исправляются. Если число ошибок в блоке становится слишком большим, то блок может быть помечен как дефектный и заменен резервным блоком. Микросхемы с ячейками памяти MLC выдерживают 10^4 циклов перепрограммирования, что примерно на порядок меньше, чем для микросхем с обычными ячейками.

В 2005 году максимальная емкость однокристалльных микросхем флэш-памяти, выпускаемых в промышленном масштабе, составляла: с логикой NOR — 512 Мбит (S29GL512N фирмы

Spansion), с логикой NAND — 4 Гбит (K9K4G08U0M фирмы Samsung). Минимальный топологический размер при производстве микросхем NAND в настоящее время составляет 55 нм.

МНОП-структуры. Полевые транзисторы с двухслойным диэлектриком, исследование которых было начато в 1967 г. [202], также используются для создания постоянных запоминающих устройств с электрической перезаписью информации. Наиболее распространенными среди них являются транзисторы со структурой МНОП (металл–нитрид–окисел–полупроводник) [1)].

В отличие от МОП-транзисторов с плавающим затвором, в МНОП-структурах заряд, кодирующий логическую информацию, хранится на глубоких ловушках в примыкающем к окислу слое нитрида кремния. Типичная толщина окисла в МНОП-структурах составляет 20–100 Å, толщина слоя Si_3N_4 — 100–1000 Å. Как и в рассмотренной выше структуре ЭППЗУ, для перемещения заряда на уровни ловушек достаточно приложить к затвору транзистора высокое напряжение, и за счет туннелирования ловушки начинают заполняться или опустошаться (в зависимости от полярности приложенного напряжения). Характерное время записи составляет ~10 мс.

К сожалению, в структурах с толщиной окисла менее 50 Å вероятность туннельной утечки электронов с ловушек в подложку достаточно высока, что ограничивает время хранения, а в структурах с толщиной окисла более 50 Å работа ячеек памяти осложняется тем, что при записи логической 1 электроны туннелируют по механизму Фаулера–Нордгейма и сначала попадают в зону проводимости SiO_2, а затем, пробретя высокую энергию за счет разрыва зон на гетерогранице SiO_2–Si_3N_4 и став горячими, уходят от границы раздела далеко в глубь нитрида. Поэтому последующее удаление этого заряда с ловушек при записи логического 0 происходит с большим трудом.

В работе [203] было предложено решение этой проблемы, состоявшее в преднамеренном легировании границы двух диэлектриков примесями, создающими уровни глубоких ловушек. В качестве примесей были успешно использованы W и Ir с поверхностной концентрацией 10^{14}–10^{15} см$^{-2}$. Легирование границы раздела позволило одновременно существенно уменьшить

[1)] Вместо нитрида кремния в подобных структурах может также использоваться Al_2O_3, Ta_2O_5 или TiO_2 [14]. Общим свойством этих диэлектриков является высокая концентрация присутствующих в них глубоких ловушек. Так, в Si_3N_4 их концентрация превышает 10^{19} см$^{-3}$.

и время программирования. Примером отечественных микросхем с МНОП-структурой могут служить микросхемы серии К558.

В последнее время большое распространение получила разновидность МНОП-структуры, называемая SONOS (silicon–oxide–nitride–oxide–silicon), в которой слой Si_3N_4 с глубокими ловушками отделен от кремниевой подложки и затвора из поликристаллического кремния тонкими (20–50 Å) слоями из SiO_2. Структура SONOS, в частности, используется в качестве запоминающей среды в некоторых типах флэш-памяти.

4.3. Полевые транзисторы с управляющим p–n-переходом и барьером Шоттки

Как мы уже говорили в начале главы, кроме полевых транзисторов с изолированным затвором существуют еще две конструкции полевых транзисторов: транзисторы с управляющим *p–n*-переходом и транзисторы с барьером Шоттки. Мы рассмотрим обе эти конструкции транзисторов вместе, поскольку они различаются только способом создания обедненного слоя, а теория работы этих приборов одна и та же.

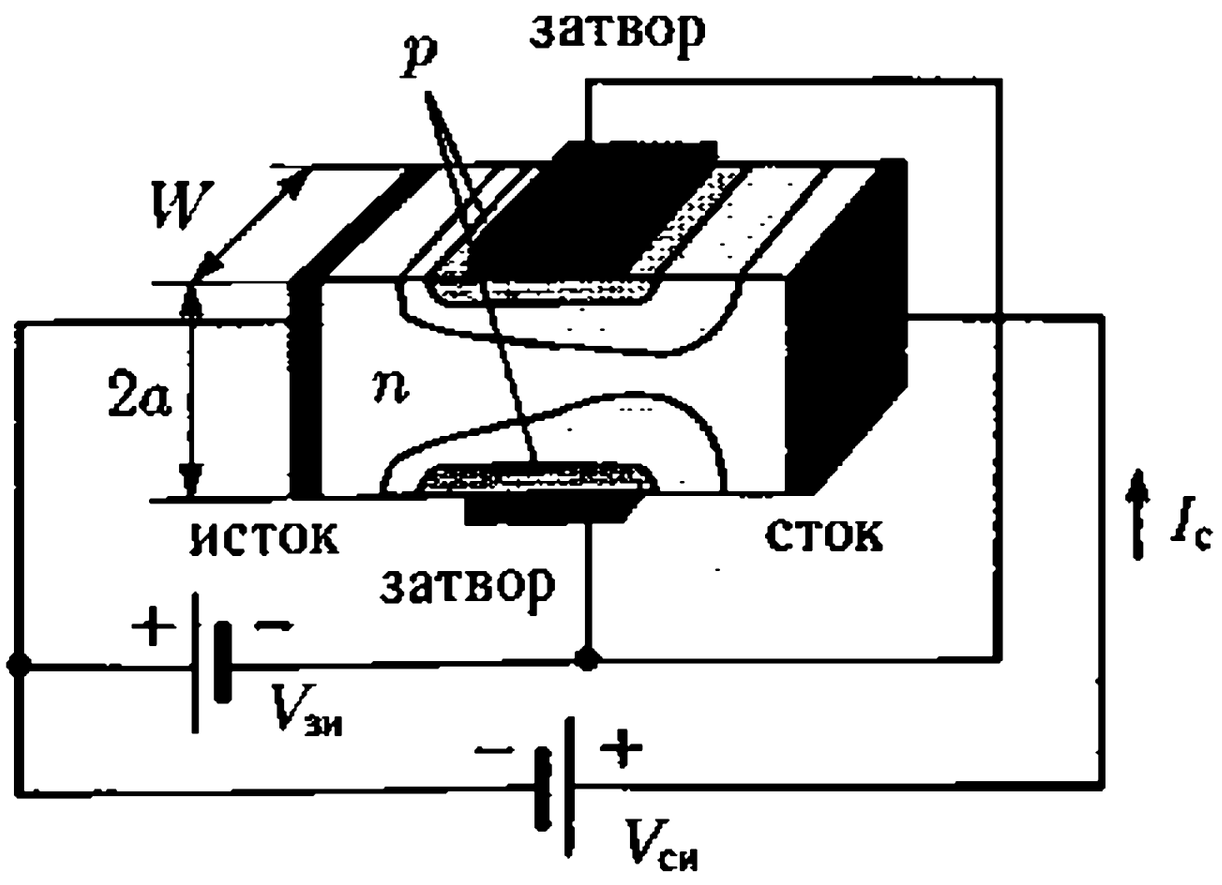


Рис. 4.26. Полевой транзистор с управляющим *p–n*-переходом [171]

Исторически первыми были разработаны *полевые транзисторы с управляющим p–n-переходом.* Теория такого транзистора была развита в 1952 г. Шокли [204], а первый транзистор был изготовлен Дейки и Россом [171] в 1953 году. Схематически устройство полевого транзистора с управляющим *p–n*-переходом показано на рис. 4.26. На боковых гранях бруска из полупроводника толщины $2a$ и ширины W диффузией создаются *p–n*-переходы. Обе диффузионные области соединяются между собой и образуют вывод затвора; истоком и стоком транзистора служат омические контакты, изготовленные на торцевых гранях образца. Работа этого прибора основана на изменении толщины проводящего канала в середине бруска за счет изменения толщины обедненных слоев *p–n*-переходов при подаче на затвор

напряжения смещения. С ростом приложенного к затвору обратного смещения толщина обедненных слоев возрастает, вызывая сужение проводящего канала и уменьшение протекающего через него тока. В определенном смысле работа этого транзистора напоминает работу электронной лампы, в которой управляющее напряжение на сетке модулирует ток электронов, движущихся от катода к аноду. Поскольку проводимость канала определяется главным образом основными носителями, этот тип полевых транзисторов также называют *униполярными* транзисторами (в отличие от биполярных транзисторов, рассмотренных нами в гл. 2).

Из-за довольно высокой емкости затвора и низкой крутизны вольт-амперной характеристики первые кремниевые полевые транзисторы с управляющим p–n-переходом имели невысокую предельную частоту f_T (см. формулу (4.21)) и применялись в основном для усиления сигналов в области низких и средних частот. В середине 60-х гг. стало понятным, что фактором, ограничивающим быстродействие полевых транзисторов, является не слишком высокая подвижность электронов в Si. Поэтому для увеличения быстродействия этих приборов было предложено использовать синтезированные в это время новые полупроводниковые материалы с более высокой подвижностью и скоростью насыщения электронов — GaAs и другие соединения $A^{III}B^{V}$. К сожалению, термодинамическая и химическая нестабильность собственного окисла, его недостаточно высокое сопротивление, а также высокая плотность поверхностных состояний на границе полупроводника с собственным окислом и другими диэлектриками (10^{13} см$^{-2}$ в случае GaAs и 10^{11} см$^{-2}$ в случае InP [141]) не позволяли получить на основе этих полупроводников хорошие МОП-транзисторы.

Чтобы сохранить основные преимущества полевых транзисторов (работа на основных носителях заряда) и использовать новые, более перспективные материалы, была предложена другая конструкция полевого транзистора — *полевой транзистор с барьером Шоттки*. Работоспособность такого транзистора из GaAs была продемонстрирована Мидом [172], а первый прибор из GaAs, в котором крутизна вольт-амперной характеристики была в 5 раз выше, чем в приборе из Si идентичной геометрии, был создан в 1967 году [205]. В настоящее время именно такую конструкцию имеют СВЧ полевые транзисторы из GaAs. В этих приборах с длиной канала 0,25 мкм получена предельная частота $f_T = 106$ ГГц; логические элементы на транзисторах с длиной канала 0,3 мкм имели задержку распространения

16 пс/вентиль [156]. Более высокое быстродействие имеют только полевые транзисторы с HEMT-структурой (см. с. 298). На основе полевых транзисторов с барьером Шоттки в настоящее время налажен выпуск мощных высокочастотных транзисторов из GaAs, а недавно появились и аналогичные транзисторы на основе карбида кремния SiC. Эти приборы позволяют получить выходную мощность 25–50 Вт на частотах до 2–4 ГГц.

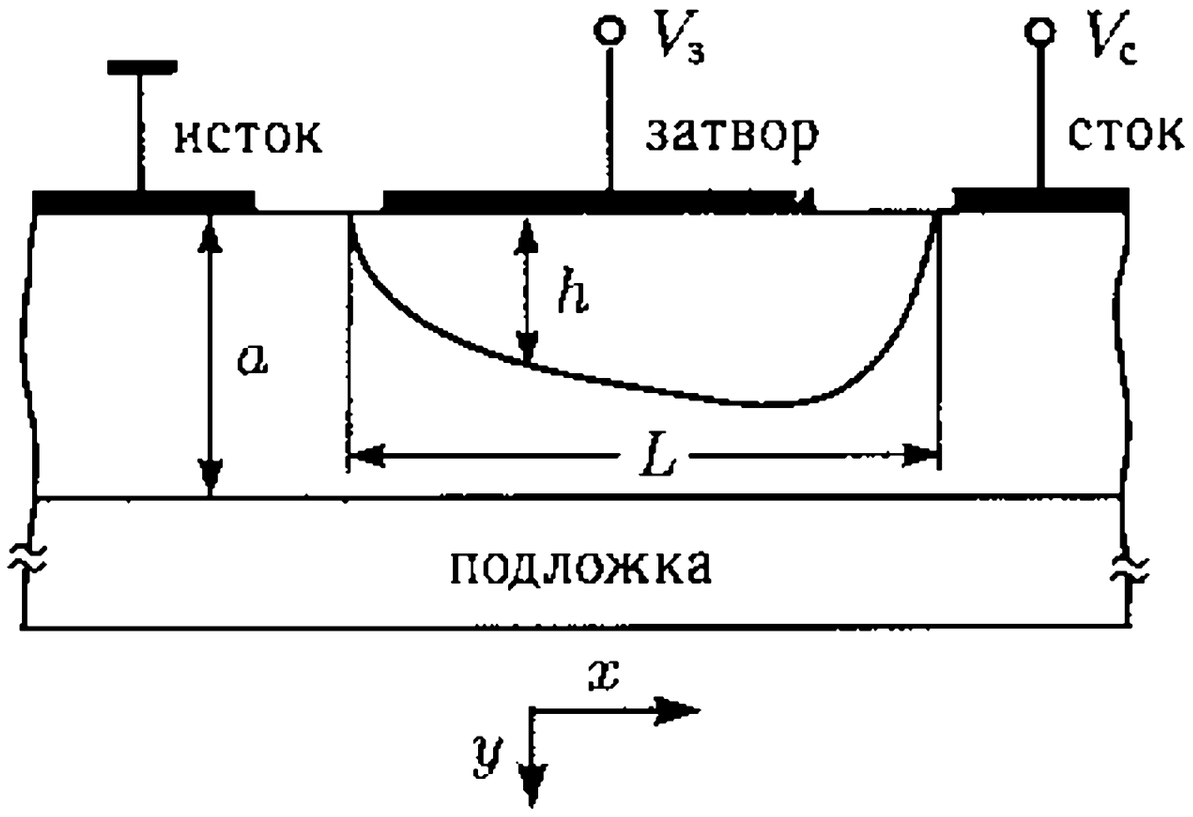


Рис. 4.27. Схематическое устройство полевого транзистора с барьером Шоттки. Характерные размеры прибора: $a = 0,3$ мкм, $L = 0,25$–$1,5$ мкм

Устройство полевого транзистора с барьером Шоттки (см. рис. 4.27) качественно похоже на устройство рассмотренного выше полевого транзистора с управляющим p–n-переходом. Отличием является то, что в этом транзисторе затвором является контакт металл–полупроводник (барьер Шоттки), а тонкий слой проводящего полупроводника с характерной концентрацией $n \sim 3 \cdot 10^{17}$ см$^{-3}$ получен эпитаксиальным наращиванием на полуизолирующую подложку, изготовленную из высокоомных GaAs(Cr) или InP(Fe).

Вольт-амперная характеристика полевого транзистора с барьером Шоттки. Рассчитаем вольт-амперную характеристику полевого транзистора с барьером Шоттки в случае однородно легированного канала n-типа проводимости [1]. Пусть, как показано на рис. 4.27, к стоку относительно истока приложено напряжение $V_с$, а к затвору — $V_з$. Протекающий в канале электрический ток вызывает в нем падение напряжения $V(x)$, которое увеличивается при перемещении от истока к стоку. Поскольку напряжение во всех точках затвора одинаково, то локальная толщина обедненного слоя $h(x)$, и, следовательно, толщина проводящего канала $[a - h(x)]$, также изменяются вдоль образца. Будем считать, что канал достаточно длинный ($L \gg a$), а продольная компонента напряженности электрического поля $\mathcal{E}_x$ намного меньше поперечной компоненты $\mathcal{E}_y$,

[1]) Решение задачи для произвольного профиля легирования можно найти в [14].

возникающей в обедненном слое. Эти условия позволяют пренебречь эффектами двумерности в распределении электрического поля в приборе и свести задачу к одномерной, то есть рассматривать ее в уже известном нам *приближении плавного канала*. Границу обедненной области будем считать резкой. Кроме того, будем считать, что напряженность продольного электрического поля невелика, так что зависимостью подвижности электронов μ_n от напряженности поля можно пренебречь.

В соответствии с формулой (1.102), локальная толщина обедненного слоя в барьере Шоттки в приближении плавного канала равна

$$h(x) = \sqrt{\frac{\varepsilon[V(x) + V_{bi} - V_з]}{2\pi q N_d}}\,, \tag{4.24}$$

где N_d — концентрация легирующей примеси в канале, а V_{bi} — величина встроенного потенциала.

Учитывая, что ток стока I_c одинаков во всех сечениях образца и вдоль образца меняется лишь толщина канала $[a - h(x)]$, нетрудно найти распределение электрического поля в канале:

$$\frac{dV}{dx} = \mathcal{E}_x = \frac{I_c}{q\mu_n N_d W[a - h(x)]}\,, \tag{4.25}$$

где W — ширина канала. Исключим из уравнения (4.25) переменную V. Для этого возведем в квадрат обе части уравнения (4.24) и, переходя к дифференциалам, получим $dV = (4\pi q N_d/\varepsilon)h\,dh$. Подставляя эту величину в уравнение (4.25) и перегруппировывая сомножители, приходим к следующему дифференциальному уравнению:

$$dx = \frac{4\pi q^2 N_d^2 \mu_n W}{\varepsilon I_c} h(a - h)\,dh. \tag{4.26}$$

Интегрируя это уравнение по x от истока ($x = 0$) до стока ($x = L$) с учетом вытекающих из (4.24) граничных условий,

$$h(0) = \sqrt{(\varepsilon/2\pi q N_d)[V_{bi} - V_з]}\,,$$

и

$$h(L) = \sqrt{(\varepsilon/2\pi q N_d)[V_c + V_{bi} - V_з]}\,,$$

в итоге получаем вольт-амперную характеристику полевого транзистора с барьером Шоттки:

$$I_c = q\mu_n N_d \frac{aW}{L}\left\{V_c - \frac{2[(V_c + V_{bi} - V_з)^{3/2} - (V_{bi} - V_з)^{3/2}]}{3V_{p0}^{1/2}}\right\}, \tag{4.27}$$

где $V_{p0} = 2\pi q N_d a^2/\varepsilon$ — *напряжение перекрытия* (*отсечки*) канала. Полученная формула справедлива при таких напряжениях на затворе и стоке, при которых канал не перекрывается даже в самой узкой его части вблизи стока. За этой границей (при $V > V_{\text{с.нас}} = V_{p0} - V_{bi} + V_{\text{з}}$) ток стока выходит на насыщение; как и в МОП-транзисторе, в этой области напряжений ток стока определяется инжекцией электронов из проводящего канала в обедненный слой.

Из формулы (4.27) нетрудно вычислить крутизну вольт-амперной характеристики транзистора, которая равна

$$g_m = \left.\frac{dI_{\text{с}}}{dV_{\text{з}}}\right|_{V_{\text{с}}=\text{const}} = q\mu_n N_d \frac{aW}{L}\left[\frac{(V_{\text{с}} + V_{bi} - V_{\text{з}})^{1/2} - (V_{bi} - V_{\text{з}})^{1/2}}{V_{p0}^{1/2}}\right] \qquad (4.28)$$

в линейной области и

$$g_{m.\text{нас}} = q\mu_n N_d \frac{aW}{L}\left(1 - \sqrt{\frac{V_{bi} - V_{\text{з}}}{V_{p0}}}\right) \qquad (4.29)$$

в области насыщения. Интересно, что величина $g_{m.\text{нас}}$ в точности равна проводимости канала транзистора при $V_{\text{с}} \to 0$.

Транзисторы с высокой подвижностью электронов (HEMT-транзисторы). Для увеличения быстродействия полевых транзисторов с барьером Шоттки необходимо увеличивать подвижность электронов в канале. В так называемых *транзисторах с высокой подвижностью электронов* (HEMT-транзисторах) [1] для этого используется обсуждавшаяся на с. 115 возможность увеличения подвижности носителей в структурах с гетеропереходом за счет пространственного разнесения области полупроводника, содержащей легирующие примеси, и канала, в котором движутся носители заряда [69, 156]

Конструкция полевого транзистора, канал которого располагается в потенциальной яме на границе двух полупроводников, была предложена в 1980 г. почти одновременно в пяти лабораториях мира, но первенство закреплено за Мимурой с соавторами [206], работавшими в фирме Fujitsu. Они-то и предложили называть эти приборы транзисторами с высокой подвижностью электронов (HEMT-транзисторами). Другим распространенным

[1] HEMT = high electron mobility transistor.

названием этой структуры является MODFET — *modulation-doped field-effect transistor*.

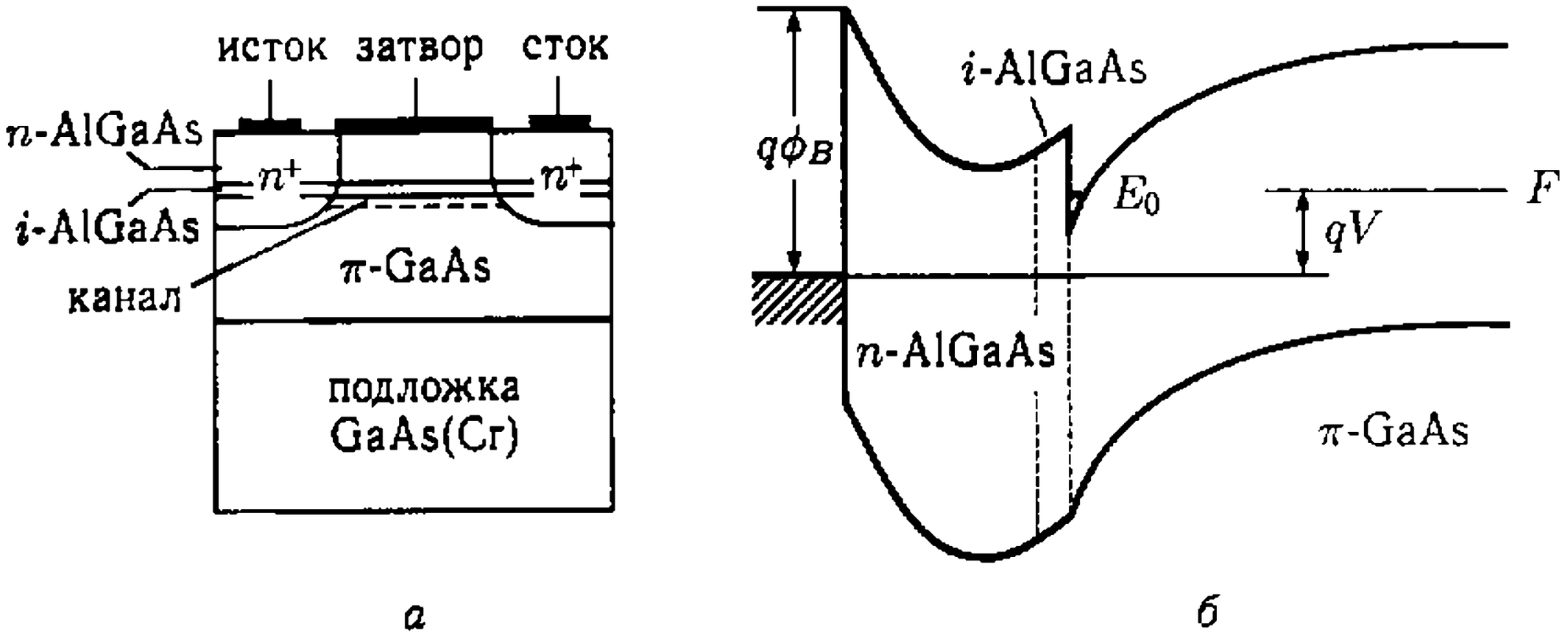


Рис. 4.28. Поперечное сечение (*а*) и энергетическая диаграмма (*б*) транзистора с высокой подвижностью электронов

Рассмотрим один из вариантов конструкции HEMT-транзистора, в котором используется гетеропереход GaAs–$Al_xGa_{1-x}As$ (см. рис. 4.28). На полуизолирующей подложке из GaAs(Cr) методами молекулярно-лучевой эпитаксии или MOCVD сначала выращивается тонкий (1–2 мкм) эпитаксиальный слой нелегированного π-GaAs с концентрацией дырок $p < 10^{14}$ см$^{-3}$. Далее на этом слое наращиваются сверхтонкий (30–60 Å) слой нелегированного твердого раствора i-$Al_{0,3}Ga_{0,7}As$ («спейсер»), а затем более толстый (500–700 Å) слой того же твердого раствора, легированного кремнием ($N_d \approx$ $\approx 2 \cdot 10^{18}$ см$^{-3}$). На поверхности структуры затем создается затвор (барьер Шоттки из Al). Далее, используя затвор в качестве маски, с помощью ионной имплантации формируются сильно легированные n^+-области стока и истока, и, наконец, к ним создаются омические контакты из AuGe/Ni.

При комнатной температуре донорные уровни в n-$Al_xGa_{1-x}As$ ионизуются, а электроны с них переходят в квантовую яму с треугольным профилем, образованную в GaAs вблизи гетерограницы. Пространственное разнесение с помощью слоя спейсера канала, в котором движутся электроны, и области, в которой остались ионизованные атомы доноров, резко ослабляет рассеяние электронов ионизованными примесями и увеличивает подвижность электронов в канале до значений, характерных для нелегированного материала (~9000 см^2/В · с при 300 К). Это значение почти вдвое выше характерной подвижности электронов в канале полевого транзистора из

GaAs с барьером Шоттки и в 10–20 раз выше, чем в канале кремниевого МОП-транзистора. Особенно разительное различие подвижностей наблюдается при понижении температуры: при 4,2 К подвижность электронов в канале может достигать $6,4 \cdot 10^6$ см^2/В $\cdot$ с [85].

Энергетическая диаграмма НЕМТ-транзистора показана на рис. 4.28 б. Ее отличительной особенностью является наличие двух взаимодействующих областей пространственного заряда — области, создаваемой барьером Шоттки к n-$Al_xGa_{1-x}As$, и области, создаваемой анизотипным гетеропереходом. Уровни легирования и толщины слоев в транзисторе, энергетическая диаграмма которого изображена на рисунке, могут быть подобраны так, чтобы транзистор работал как в режиме обогащения, так и в режиме обеднения. Для цифровых ИС, работающих от источника питания одной полярности, удобнее использовать транзисторы, работающие в режиме обогащения (то есть когда при $V_{зи} = 0$ транзистор закрыт), а для СВЧ ИС — транзисторы, работающие в режиме обеднения. Взаимодействие двух областей пространственного заряда обеспечивает то, что при подаче на затвор положительного смещения уровень размерного квантования в квантовой яме (обозначенный на рисунке E_0) опускается ниже уровня Ферми в π-GaAs и канал заполняется электронами. Из-за большой высоты барьера Шоттки к $Al_xGa_{1-x}As$ ток в цепи затвора даже при $V_{зи} > 0$ остается малым.

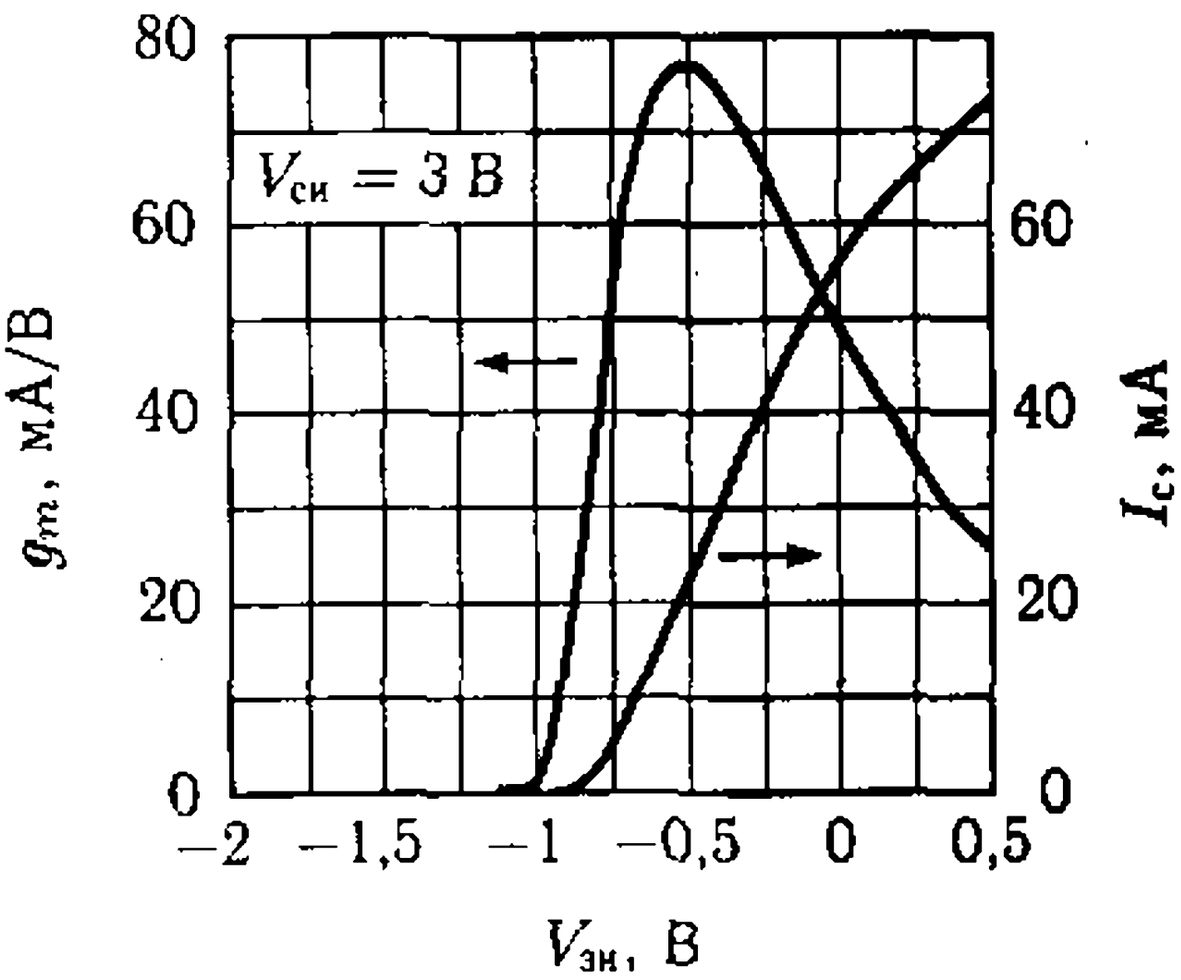


Рис. 4.29. Зависимость тока стока I_c и крутизны g_m вольт-амперной характеристики от напряжения на затворе в НЕМТ-транзисторе 2SK3001 фирмы Hitachi

Особенностью формирования энергетической диаграммы НЕМТ-транзистора является то, что при достаточно больших прямых смещениях на затворе край зоны проводимости в широкозонном полупроводнике может опуститься ниже уровня E_0 и тогда электроны начнут перетекать из квантовой ямы в яму, образующуюся в n-слое. Это приводит к уменьшению эффективной подвижности электронов (заметная часть электронов движется в области, в которой велико примесное рассеяние) и ослаблению управляемости

характеристиками транзистора со стороны затвора (электроны экранируют электрическое поле затвора). Оба эффекта приводят к ослаблению зависимости тока стока от напряжения на затворе и уменьшению крутизны вольт-амперной характеристики (см. рис. 4.29).

Расчеты показывают, что в HEMT-транзисторах с коротким каналом крутизна вольт-амперной характеристики в области насыщения определяется скоростью насыщения v_s и почти не зависит от подвижности электронов в слабом электрическом поле. Поэтому может показаться, что большого смысла бороться за увеличение подвижности нет. Однако это не так: увеличение подвижности немного увеличивает значение v_s и приводит к заметному уменьшению сопротивлений областей истока и стока (эти сопротивления, как мы отмечали на с. 258, оказывают негативное влияние на крутизну вольт-амперной характеристики). Поэтому стремление создавать все более высокочастотные приборы с неизбежностью приводит к поиску полупроводников с высокими значениями v_s и μ_n.

С другой стороны, для создания мощных транзисторов необходимо, чтобы удельный ток стока на единицу ширины канала в открытом состоянии транзистора был как можно выше. Чтобы добиться этого, необходима высокая поверхностная плотность электронов n_s в канале. К сожалению, описанный выше эффект перетекания электронов в широкозонную часть структуры не позволяет сколь угодно сильно увеличивать n_s; типичное значение n_s в HEMT-транзисторах на основе GaAs/AlGaAs составляет $\sim 10^{12}$ см$^{-2}$ в то время как в полевых транзисторах из GaAs с барьером Шоттки эта величина на порядок выше. Нетрудно показать, что максимальное значение n_s возрастает с увеличением энергии разрыва зон ΔE_c на гетерогранице. Однако увеличивать ΔE_c путем увеличения доли алюминия в твердом растворе n-$Al_xGa_{1-x}As$ выше $x = 0,25$ нельзя из-за возникновения в нем глубоких донорных уровней. Поэтому разработчики HEMT-транзисторов обратились к поиску других гетеропереходных пар с бóльшим значением ΔE_c. Так были получены HEMT-транзисторы на основе структуры InP–$Al_{0,48}In_{0,52}As$–$In_{0,53}Ga_{0,47}As$–$Al_{0,48}In_{0,52}As$, согласованной по параметру решетки. Однако решением проблемы стал полный отказ от согласованных по параметру решетки гетеропереходных пар и переход к использованию очень тонких *напряженных (псевдоморфных) слоев* в качестве узкозонной части структуры. Транзисторы с таким напряженным слоем называют *псевдоморфными* HEMT-транзисторами (pHEMT). В качестве узкозонного слоя

в них чаще всего используется твердый раствор $In_xGa_{1-x}As$ (толщина 100 Å, $x \sim 0.3$), в котором значения v_s и μ_n заметно выше, чем в GaAs, а более высокая величина ΔE_c в гетеропереходе AlGaAs/InGaAs обеспечивает получение высокого значения n_s (до $4 \cdot 10^{12}$ см$^{-2}$).

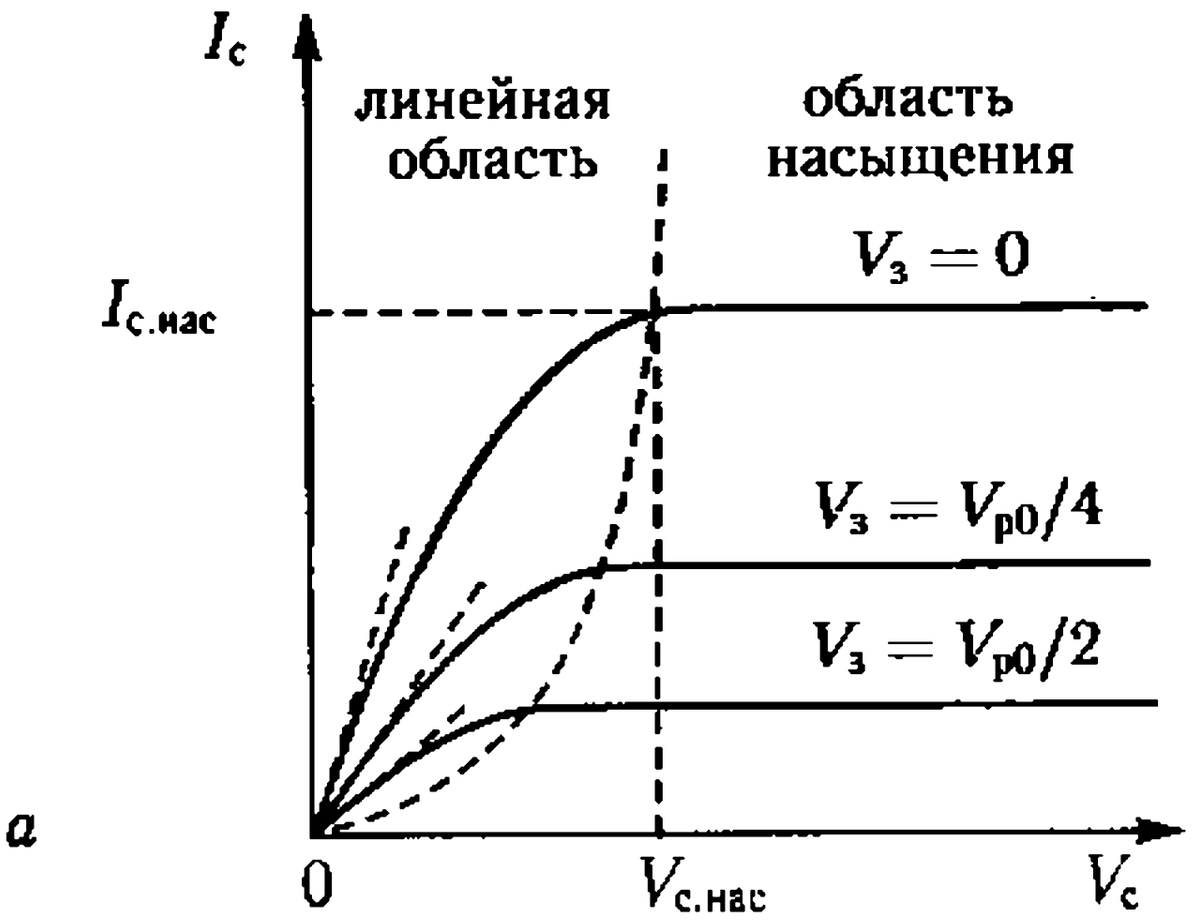
В последнее время большой интерес вызывают разработки HEMT-транзисторов на основе InAs — полупроводника, в котором v_s и μ_n примерно вдвое выше, чем в твердом растворе $In_{0,53}Ga_{0,47}As$, согласованном по параметру решетки с InP. Так, на основе гетеропереходов InAs/AlSb уже созданы опытные приборы, которые при сравнимом с приборами на основе $In_{0,53}Ga_{0,47}As$ быстродействии потребляют в 5–10 раз меньшую мощность.

Другим перспективным материалом для HEMT-транзисторов является структура на основе гетероперехода GaN/AlGaN. Исследование опытных образцов этих структур выявило ряд их интересных свойств, и прежде всего, возможность получения необычно высокой ($>10^{13}$ см$^{-2}$) поверхностной плотности заряда в канале, которая намного выше, чем в структуре GaAs/AlGaAs. Это оказывается возможным за счет поляризационных явлений (решетка GaN является полярной) и пьезоэлектрических полей, возникающих в механически напряженном слое вблизи гетерограницы. Вместе с высокой напряженностью поля пробоя (3,5 МВ/см), хорошей теплопроводностью и высокой скоростью насыщения электронов в GaN ($2,5 \cdot 10^7$ см/с) это предопределяет перспективную для этих структур область применения — создание мощных полевых транзисторов [207].

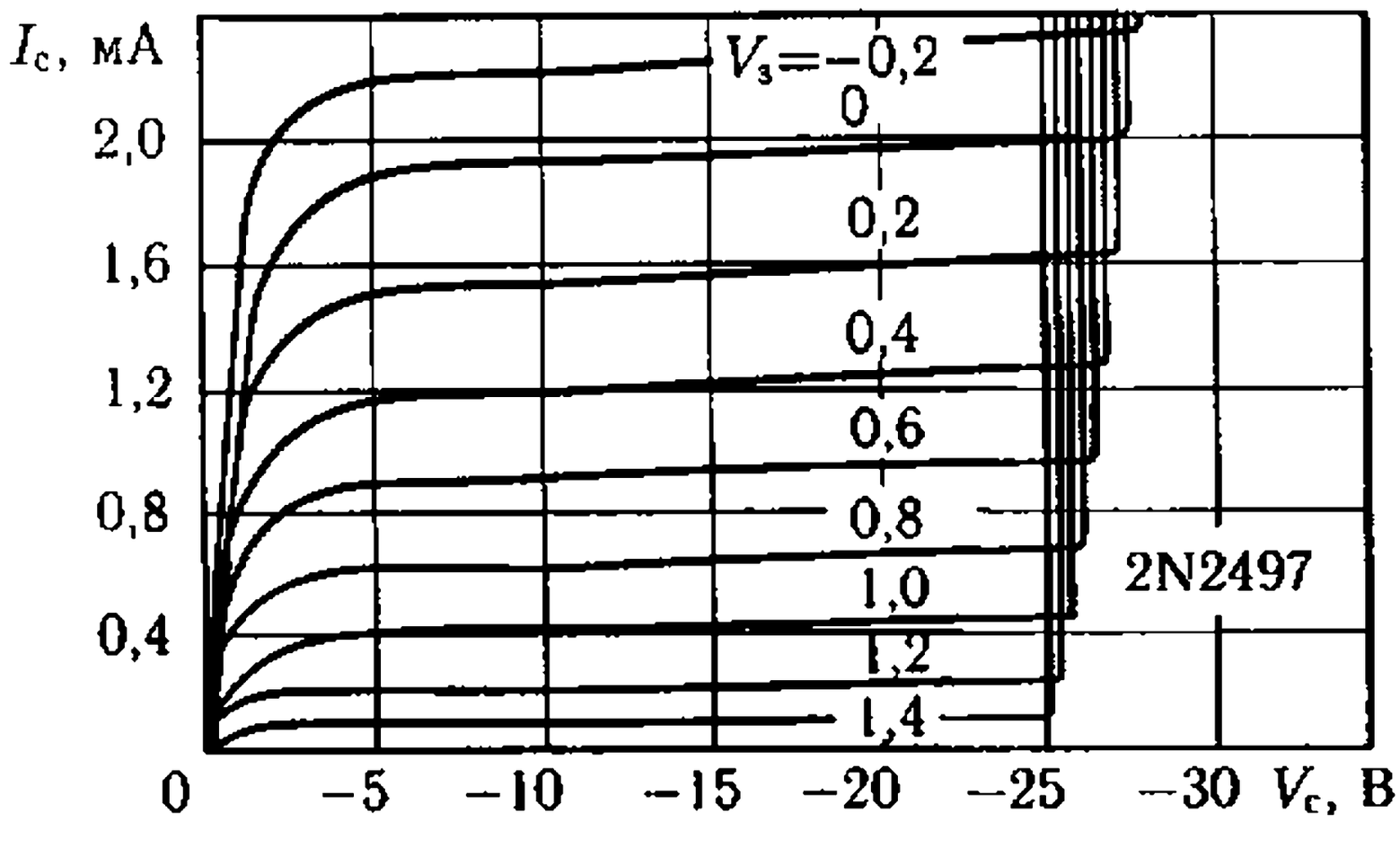
В настоящее время HEMT-транзисторы являются наиболее высокочастотными и быстродействующими среди всех типов полевых транзисторов, причем параметры транзисторов на основе InP заметно превосходят параметры транзисторов на основе GaAs. На основе гетеропереходов $Al_xIn_{1-x}As$–$In_xGa_{1-x}As$–InP были созданы транзисторы с предельной частотой $f_T = 562$ ГГц [208]. Использование HEMT-транзисторов в цифровых ИС позволяет уменьшить время задержки распространения до 3,2 пс [209]. В настоящее время налажен промышленный выпуск HEMT-транзисторов. Эти приборы характеризуются очень низким уровнем шума. Например, транзисторы MGF4951A и MGF4953A фирмы Mitsubishi работают на частотах до 26 ГГц и имеют шум-фактор 0,4 дБ на частоте 12 ГГц; еще более низкий коэффициент шума имеют HEMT-транзисторы на основе InP. По этой причине HEMT-транзисторы используются в системах спутниковой связи, радиоастрономии, в мобильных телефонах.

На основе HEMT-транзисторов из InP предполагают создавать цифровые ИС для будущих систем оптоволоконной связи со скоростью передачи 40–100 Гбит/с.

Полевые транзисторы с управляющим p–n-переходом. Сопоставляя рис. 4.26 и рис. 4.27, нетрудно заметить, что если разделить полевой транзистор с управляющим *p–n*-переходом на две части по плоскости симметрии, то каждая из полученных половинок окажется эквивалентной полевому транзистору с барьером Шоттки с единственным отличием, состоящим в том,



а



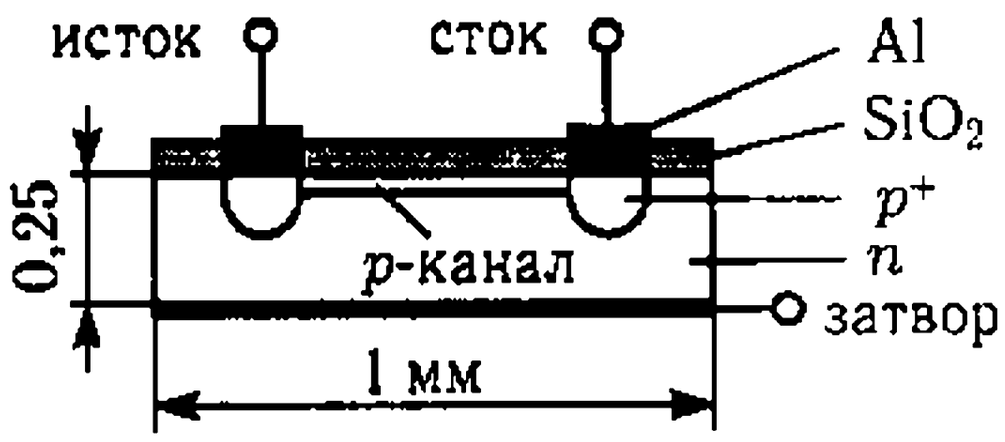
б

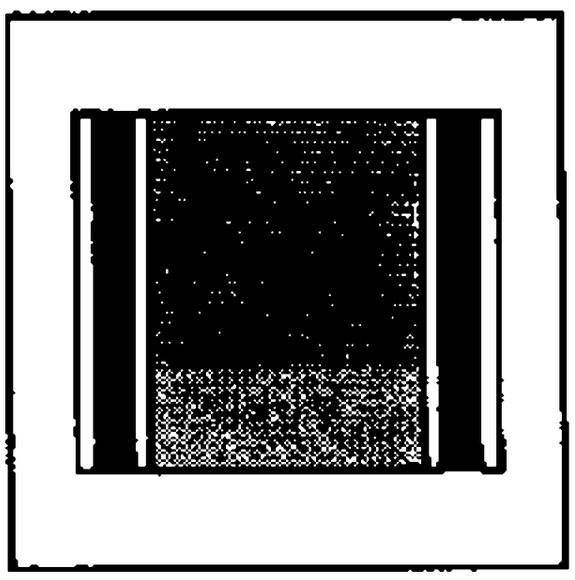
Рис. 4.30. Выходные вольт-амперные характеристики полевого транзистора с управляющим *p–n*-переходом: *а* — качественный вид; *б* — транзистор 2N2497

что в транзисторе с управляющим *p–n*-переходом вместо барьера Шоттки используется *p–n*-переход. Поэтому выражение для вольт-амперной характеристики полевого транзистора с управляющим *p–n*-переходом совершенно аналогично формуле (4.27), за исключением того, что ток стока следует увеличить в два

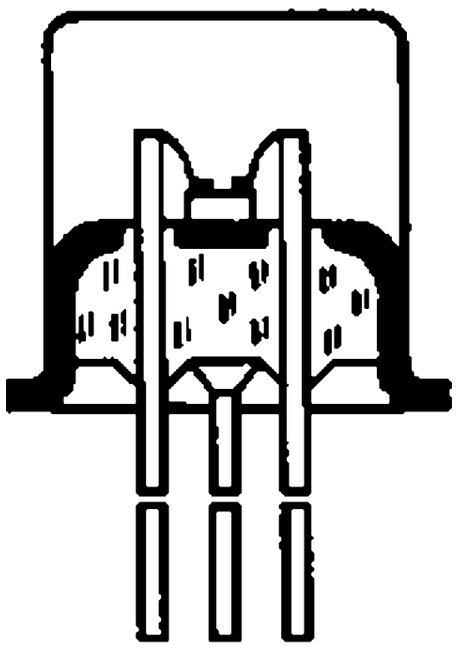
раза, а вместо встроенного потенциала V_{bi} использовать контактную разность потенциалов ϕ_k, рассчитываемую по формуле (1.4).

На рис. 4.30*а* показан качественный вид выходных вольтамперных характеристик полевого транзистора, рассчитанных по формуле (4.27), а на рис. 4.30*б* — характеристики промышленного полевого транзистора с управляющим p–n-переходом. Экспериментальные зависимости хорошо соответствуют расчетным за исключением области высоких напряжений на стоке, при которых ток стока резко возрастает (это связано с пробоем перехода затвор–сток).

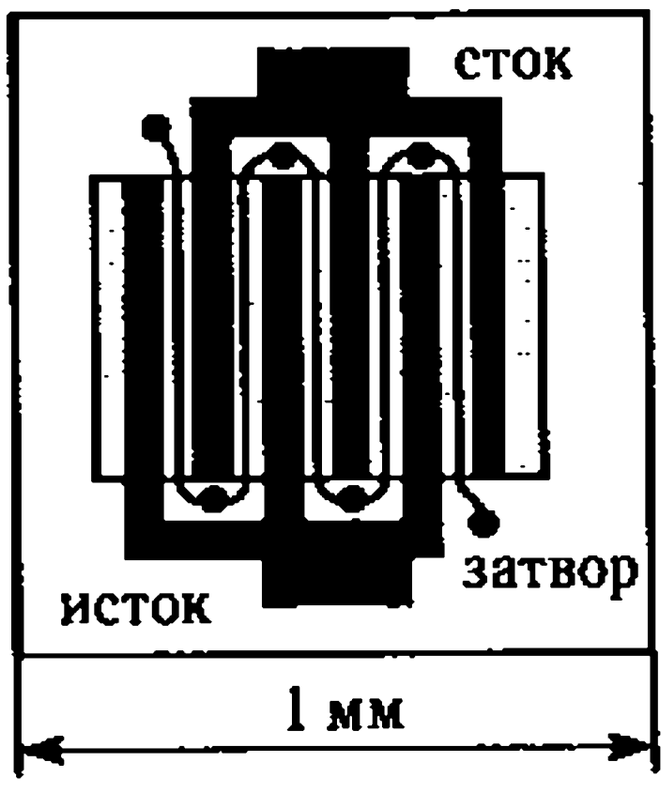




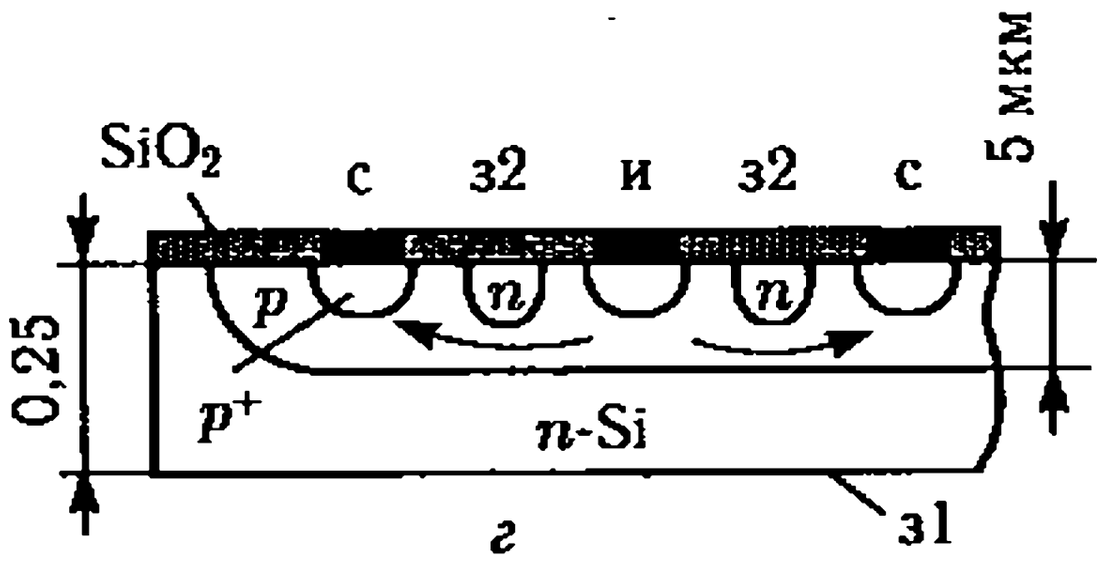
а



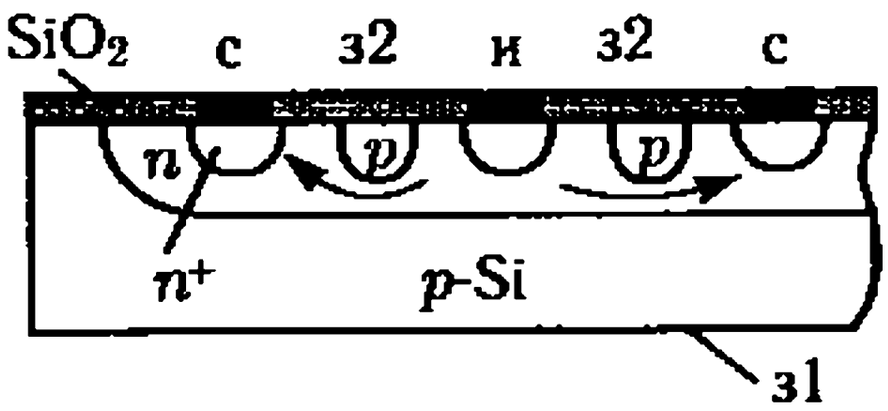
б



в



г



д

Рис. 4.31. Конструкции первых отечественных полевых транзисторов КП102 (*а*, *б*), КП103 (*в*, *г*) и КП302 (*д*)

При изготовлении первых полевых транзисторов с управляющим p–n-переходом основная трудность состояла в том, чтобы создать

достаточно тонкий канал и при этом обеспечить необходимую механическую прочность конструкции. Некоторые решения этой проблемы представлены на рис. 4.31. В конструкции транзистора КП102 с каналом p-типа проводимости на одной стороне пластины из Si с $n \sim$ $\sim 5 \cdot 10^{16}$ см$^{-3}$ с помощью диффузии акцепторной примеси создается слабо легированная область p-типа, которая будет служить каналом транзистора (см. рис. 4.31 *а*). Затем эту область закрывают защитным слоем SiO_2, в котором через окна формируют сильно легированные p^+-области истока и стока. Затвором в этой конструкции служит подложка.

В конструкции транзисторов КП103 с каналом p-типа и КП302 с каналом n-типа на одной стороне пластины сначала создаются неглубокие «карманы», тип проводимости которых противоположен типу проводимости подложки (см. рис. 4.31 *г,д*). После этого на поверхности пластины формируются диффузионные области истоков, стоков и затворов. Расположенные на поверхности полоски затворов з2 соединяются с подложкой, которая выполняет роль нижнего затвора з1. Для увеличения крутизны вольт-амперной характеристики каждый из описываемых приборов содержит пять параллельно соединенных транзисторов. Из-за достаточно большой длины канала предельная частота этих транзисторов не превышала 100 МГц.

Мощные полевые транзисторы с управляющим p–n-переходом строятся в основном по двум схемам [11]. В первой из них на поверхности эпитаксиального слоя n-типа, выращенного на подложке p^+-типа, вытравливаются V-образные канавки почти на всю глубину эпитаксиального слоя. Эти канавки отделяют друг от друга области истока и стока; затвором транзистора служит подложка. Благодаря V-образной канавке эффективная длина канала в транзисторе оказывается сравнительно короткой (порядка 1 мкм). Параллельное соединение большого числа таких транзисторов позволяет создать, например, транзистор с напряжением пробоя 37 В и током насыщения 270 А, предельная частота которого составляет 1,2 МГц [11]. Другой конструкцией мощного полевого транзистора является полевой транзистор со статической индукцией, рассматриваемый в следующем разделе.

Транзисторы со статической индукцией. Конструкция мощного полевого транзистора, называемого *полевым транзистором со статической индукцией* или аналоговым транзистором [11, 88, 163], показана на рис. 4.32 *а* [1]. Было предложено несколько похожих конструкций этого транзистора [210]. Транзистор имеет вертикальную структуру, затвор которой

[1] В зарубежной литературе для обозначения этих транзисторов используется аббревиатура SIT (static induction transistor).

выполнен в виде сетки из полупроводника p^+-типа, разделяющей области истока и стока n-типа проводимости. Носители, направляющиеся от истока к стоку, в области сетки движутся по каналам, ширина которых ограничена областями пространственного заряда p–n-переходов и зависит от приложенного к затвору напряжения. Благодаря очень короткому каналу транзистор характеризуется высоким током стока и высокой крутизной вольт-амперной характеристики, а из-за проявления эффектов короткого канала сами вольт-амперные характеристики сильно отличаются от характеристик обычных полевых транзисторов и очень напоминают характеристики вакуумного триода (см. рис. 4.32*б*). Транзисторы со статической индукцией могут иметь напряжение пробоя 800 В, максимальный ток стока 60 А, сопротивление в открытом состоянии 2 Ом и время включения 200 нс [11].

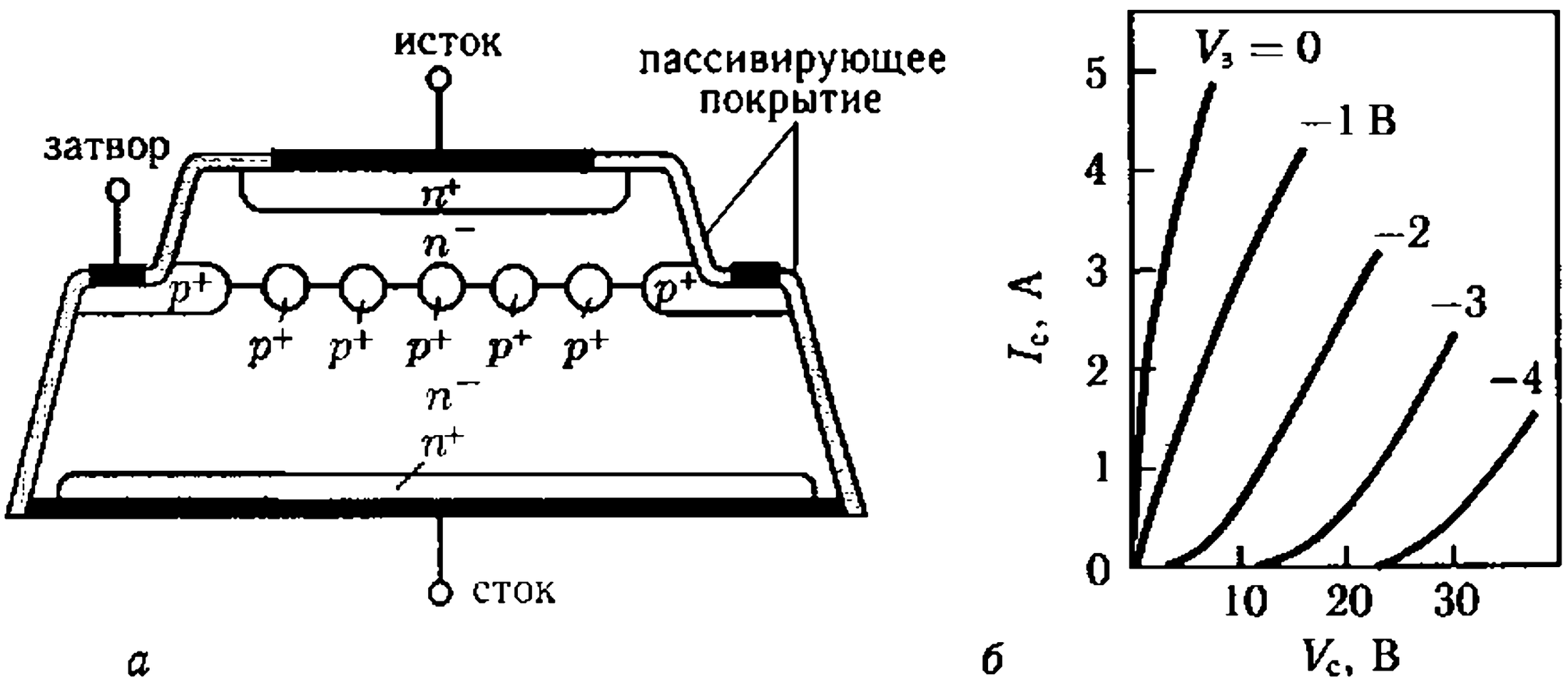


Рис. 4.32. Мощный полевой транзистор со статической индукцией: *а* — конструкция транзистора, *б* — его выходные вольт-амперные характеристики [11]

Изменение уровня легирования n-области структуры существенно влияет на вольт-амперные характеристики транзистора со статической индукцией. Выше мы описали случай, когда толщина области пространственного заряда вокруг сетки меньше половины расстояния между соседними проводниками сетки; в этом случае при нулевом напряжении на затворе транзистор открыт, а для его закрывания на затвор надо подать обратное смещение. В другом случае, когда области пространственного заряда перекрываются уже при нулевом напряжении на затворе, характеристики транзистора становятся похожими на характеристики биполярных транзисторов (см. гл. 2). Здесь перекрывающиеся области пространственного заряда образуют

«базу», которая разделяет верхнюю и нижнюю области n-типа. Протекание электронов через такую «базу» происходит там, где высота потенциального барьера для них минимальна — в местах, посередине между проводниками сетки. Такую конструкцию транзистора называют *биполярным транзистором со статической индукцией* [1]). Особенностью работы такого транзистора является то, что при подаче на затвор положительного смещения начинается инжекция дырок в слабо легированную n-область. Это вызывает приток электронов из контактов истока и стока для компенсации заряда инжектированных дырок, в результате чего сопротивление открытого канала резко падает. Из-за модуляции проводимости рабочая плотность тока в этих транзисторах может быть в несколько раз выше, чем в обычных биполярных транзисторах. Это предопределяет возможность использования биполярных транзисторов со статической индукцией в качестве мощных коммутационных устройств. Описанный выше режим работы называют *биполярным* режимом работы транзистора со статической индукцией в отличие от *полевого* режима, реализуемого при отрицательных смещениях на затворе.

Наконец, если концентрация легирующей примеси в n-области совсем низка (i-область), то возникает весьма необычная (триодная) вольт-амперная характеристика. Это связано с тем, что из-за слабого экранирования электрического поля в i-области высота потенциального барьера для инжекции электронов из истока оказывается линейной комбинацией потенциалов затвора и стока. Поэтому пока плотность тока инжекции невелика, ток стока

$$I_{\text{с}} \sim \exp\left(\frac{q(V_{\text{з}} + V_{\text{с}}/\mu)}{kT}\right),$$

где коэффициент μ зависит от геометрии прибора. При увеличении плотности тока концентрация электронов в канале становится выше концентрации легирующей примеси и ток стока начинает ограничиваться пространственным зарядом инжектированных носителей.

Существует еще один вариант транзистора со статической индукцией, в котором сетка затвора изготовлена из металла и погружена в достаточно сильно легированный полупроводник [69]. Это — так называемый *транзистор с проницаемой базой*. В этом транзисторе ширина каналов модулируется

---

[1]) В зарубежной литературе для обозначения этих транзисторов используется аббревиатура BSIT (bipolar static induction transistor).

с помощью барьеров Шоттки. Интерес к этому типу транзистора обусловлен тем, что из-за очень короткого канала (его длина определяется максимальной толщиной обедненного слоя барьера Шоттки) еще в 80-х годах на транзисторах из GaAs с вольфрамовой сеткой удалось получить время задержки распространения сигнала 4,3 пс [211].

Транзисторы со статической индукцией изготавливают не только из кремния и GaAs, но также и из SiC и GaN. Два последних полупроводника характеризуются широкой запрещенной зоной, высокой напряженностью поля пробоя и достаточно высокой подвижностью электронов, и поэтому хорошо подходят для создания мощных высокочастотных транзисторов (например, эти приборы позволяют получить выходную мощность 10 Вт на частоте 10 ГГц). Отечественной промышленностью разработан ряд мощных кремниевых транзисторов со статической индукцией, например, КП802 для работы в полевом режиме и КП934 для работы в биполярном режиме.

Шумы в полевых транзисторах. Основными источниками шума в полевых транзисторах являются тепловой шум канала и дробовой шум затвора. Подробный расчет этих шумов проведен в работе [49], здесь же мы обсудим только их результаты.

При нулевом напряжении сток–исток *тепловой шум* канала определяется его сопротивлением, которое, как мы показали выше, численно совпадает с величиной, обратной крутизне вольт-амперной характеристики транзистора $g_{m.\text{нас}}$ в режиме насыщения. Расчеты, проведенные при $V_{\text{си}} \neq 0$ (то есть когда канал становится неоднородным по толщине) показывают, что средний квадрат напряжения шума равен

$$\overline{e}^2_{\text{ш}} = \frac{4\gamma\pi kT}{g_{m.\text{нас}}}\,\Delta f, \tag{4.30}$$

где коэффициент γ определяется профилем изменения толщины канала вдоль образца и изменяется от 1 при $V_{\text{си}} = 0$ до $\gamma \approx \approx 2/3$ в режиме насыщения. То же справедливо и для МОП-транзисторов, созданных на высокоомных подложках.

Дробовой шум в полевых транзисторах достаточно мал, что обусловлено малым током утечки затвора. Однако на средних и высоких частотах в флуктуации тока затвора кроме дробового шума начинает давать вклад еще шум протекающего в цепи затвора емкостного тока, индуцированного тепловым шумом в канале. Если бы этот ток носил чисто емкостный характер, то он не давал бы вклад в шум (поскольку, в соответствии с принципом

Онзагера, шум связан с *диссипативными процессами*). Однако из-за конечного времени пролета носителей через канал сдвиг между фазой флуктуации напряжения и фазой емкостного тока немного отличается от 90°. В результате этого в токе затвора появляется действительная составляющая, которая дает дополнительный вклад в токовый шум. В итоге, полная величина среднего квадрата флуктуаций тока в цепи затвора равна

$$\bar{i}^2_{\text{ш}} = \left(2qI_{\text{з}} + \frac{4\gamma kT}{g_m}\omega^2 C_{\text{вх.экв}}\right)\Delta f, \qquad (4.31)$$

где $C_{\text{вх.экв}}$ — эффективная емкость, приблизительно равная 2/3 от емкости затвор–исток, а $\omega = 2\pi f$. Второе слагаемое в этой формуле обычно начинает превышать дробовой шум (первое слагаемое) на частотах выше ~50 кГц. Поэтому для высокочастотных применений следует использовать транзисторы с наименьшим отношением $C_{\text{вх.экв}}/g_m$.

В полевых транзисторах с МОП-структурой большую роль играет *фликкер-шум*, связанный с захватом и выбросом носителей поверхностными состояниями. Спектральная плотность этого шума пропорциональна плотности поверхностных состояний. Как мы уже отмечали, наименьшей плотностью поверхностных состояний на границе $Si–SiO_2$ характеризуется поверхность Si(100), и, следовательно, для ослабления фликкер-шума в МОП-транзисторах следует использовать кремний с такой ориентацией поверхности. Однако даже при этом избыточный шум в МОП-транзисторах оказывается столь сильным, что для усиления слабых сигналов эти транзисторы можно использовать только на частотах выше 1 МГц.

В полевых транзисторах с p–n-переходом источником дополнительного шума являются процессы генерации–рекомбинации с участием глубоких ловушек в области пространственного заряда. Захват и выброс носителей на эти ловушки модулирует толщину обедненного слоя и, следовательно, меняет толщину проводящего канала. В полевых транзисторах из GaAs с барьером Шоттки заметную роль играет также генерационно-рекомбинационный шум, связанный с глубокими ловушками в канале транзистора.

Полевые транзисторы с управляющим p–n-переходом и многоячеечной структурой (например, 2SK369 фирмы Toshiba) могут иметь очень низкое значение шумового напряжения (0,7 нВ/$\sqrt{\text{Гц}}$ на частоте 1 кГц) и при сопротивлении источника сигнала 10 кОм усилитель на таком транзисторе будет иметь температуру

шума менее 2 К в области звуковых частот. Из-за малого шума тока утечки затвора полевые транзисторы идеально подходят для создания низкошумящих усилителей с высоким входным сопротивлением.

Глава 5

ПРИБОРЫ С ЗАРЯДОВОЙ СВЯЗЬЮ

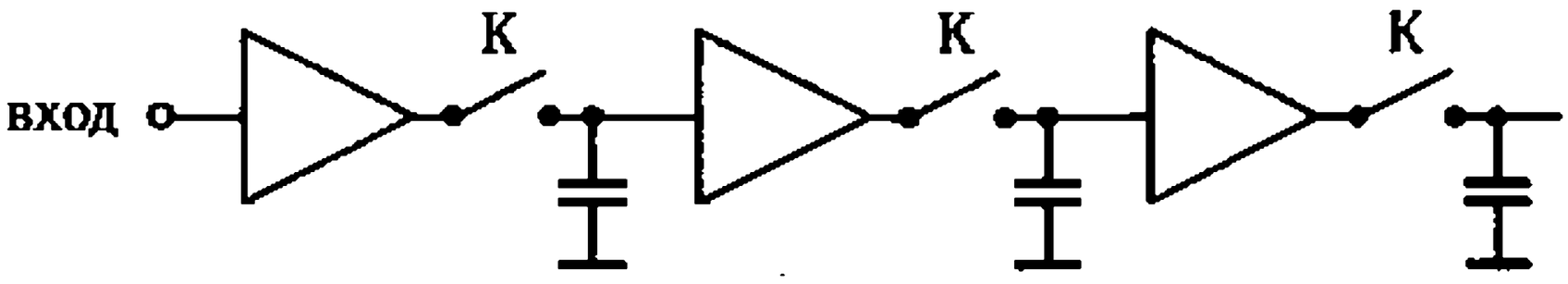
Структура, похожая на рассмотренный нами в п. 4.1.1 МОП-конденсатор, используется в *приборах с зарядовой связью* (ПЗС), иногда также называемых приборами с переносом заряда [1]). Приборами с зарядовой связью называют семейство полупроводниковых приборов, работа которых основана на управляемом перемещении *зарядовых пакетов* вдоль поверхности полупроводника путем подачи на систему поверхностных электродов определенной последовательности тактовых импульсов [212, 213]. Изменяя величину заряда в зарядовых пакетах, можно кодировать как цифровую, так и аналоговую информацию. Это предопределяет широкие возможности применения этих приборов.

Идея о возможности использования *не напряжения, а величины* заряда для представления информации и ее обработки возникла задолго до изобретения приборов с зарядовой связью. В. К. Зворыкин, изобретатель первой передающей телевизионной трубки (иконоскопа), еще в 1934 г. высказал идею создания запоминающих устройств, информация в которых хранилась бы в виде зарядов на множестве конденсаторов, созданных на поверхности изолирующей подложки. В 1952 г. были созданы первые электронные схемы аналоговых линий задержки, которые использовали принцип направленного перемещения заряда и получили название «пожарных цепочек» [2]).

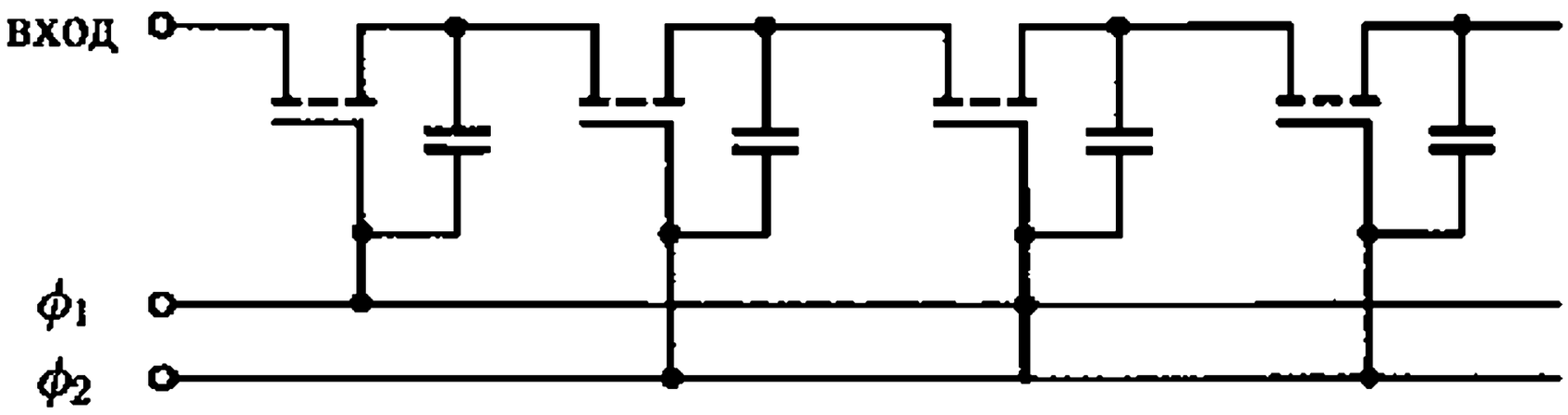
---

[1]) В зарубежной литературе эти приборы обозначаются CCD (charge coupled device).

[2]) Термин «пожарная цепочка» взят из обыденной жизни. Так называется известный принцип пожаротушения, когда люди, стоя на месте, передают ведра с водой друг другу по цепочке.



а



б

Рис. 5.1. Аналоговая линия задержки на «пожарных цепочках» (*а*) и ее реализация на дискретных полевых транзисторах (*б*) [212]

Пожарная цепочка представляет собой набор последовательно включенных запоминающих элементов (каждый из которых состоит из запоминающего конденсатора и электронной схемы повторителя), соединенных друг с другом с помощью ключей (см. рис. 5.1 *а*). При замыкании любого ключа К напряжение (заряд) с предыдущего конденсатора в цепочке копируется на следующий конденсатор. Если в такой схеме попеременно замыкать каждый второй ключ, то напряжение со входа цепочки при каждом переключении будет последовательно перемещаться на один шаг вдоль цепочки и на выходе цепочки сигнал появится с задержкой, равной произведению интервала времени между переключениями на число элементов в цепочке.

Реализация схемы пожарной цепочки на полевых транзисторах показана на рис. 5.1 *б*. Движущей силой, заставляющей в этой схеме заряд перемещаться с конденсатора на конденсатор, является напряжение между стоком и истоком, которое наводится конденсатором, подключенным к тактовой шине (ϕ_1 или ϕ_2), и появляется в момент открывания полевого транзистора.

Идея приборов с зарядовой связью была высказана в 1969 г. В. Бойлом и Дж. Смитом [214], работавшими в Bell Laboratories и искавшими пути создания электрического аналога запоминающего устройства на цилиндрических магнитных доменах [1]).

Простейшей конструкцией прибора с зарядовой связью является трехтактный регистр сдвига (см. рис. 5.2). Он состоит

---

[1]) Любопытно, что благодаря тому, что технологическая база для создания ПЗС была подготовлена предыдущими разработками интегральных схем на МОП-транзисторах, время, понадобившееся для практической реализации идеи ПЗС в виде прибора, составило всего несколько недель. По-видимому, ПЗС является единственным прибором, реализация которого была столь быстрой.

из цепочки МОП-конденсаторов и системы управляющих электродов, которые подсоединены к трем тактовым шинам ϕ_1, ϕ_2 и ϕ_3. Из-за более высокой подвижности электронов по сравнению с дырками ПЗС обычно создают на подложках p-типа. Принципиально важным моментом в работе этих приборов является то, что МОП-конденсаторы работают в условиях *неравновесного заполнения потенциальных ям электронами* [1]). В *режиме хранения* на одну из шин (ϕ_2 на рис. 5.2 *а*) подается положительный потенциал хранения $V_{\text{з.хран}} > V_{\text{пор}}$ и под подключенными к этой шине электродами возникают потенциальные ямы, в которых могут храниться зарядовые пакеты электронов. В режиме *передачи информации* на группу соседних электродов (ϕ_3) подается нарастающий потенциал, а потенциал на шине ϕ_2 через некоторое время снижается (см. рис. 5.2 *б*). При этом электроны из ям, в которой они хранились, перетекают в соседние ямы. После установления новых напряжений на тактовых шинах ПЗС снова переходит в режим хранения, однако зарядовые пакеты уже оказываются передвинутыми. При многократном повторении этих действий зарядовые пакеты перемещаются вдоль поверхности полупроводника в направлении, задаваемом расположением управляющих электродов. Для направленного перемещения зарядов при симметричной конструкции МОП-конденсаторов необходимо иметь не менее трех управляющих электродов, поскольку в режиме передачи по крайней мере один из них должен формировать барьер, предотвращающий перенос заряда в нежелательном направлении.

Основными параметрами приборов с зарядовой связью являются три величины: максимальный заряд в потенциальной яме, максимальное время хранения и эффективность переноса заряда.

Максимальный заряд в яме представляет собой заряд, который может удерживаться в яме без ее «переполнения». Поверхностная плотность этого заряда определяется напряжением на затворе МОП-конденсатора в режиме хранения $V_{\text{з.хран}}$ и составляет, как это следует из формул п. 4.1.2,

$$(qn_{ss})_{\max} = -C_{\text{д}}(V_{\text{з.хран}} - V_{\text{пор}}). \qquad (5.1)$$

---

[1]) Как мы уже отмечали на с. 246, после изменения напряжения на управляющем электроде МОП-конденсатора изменение заряда неосновных носителей в потенциальной яме происходит достаточно медленно. Это позволяет, работая в *динамическом режиме*, использовать эту особенность для временного хранения информации в виде зарядов в потенциальных ямах.

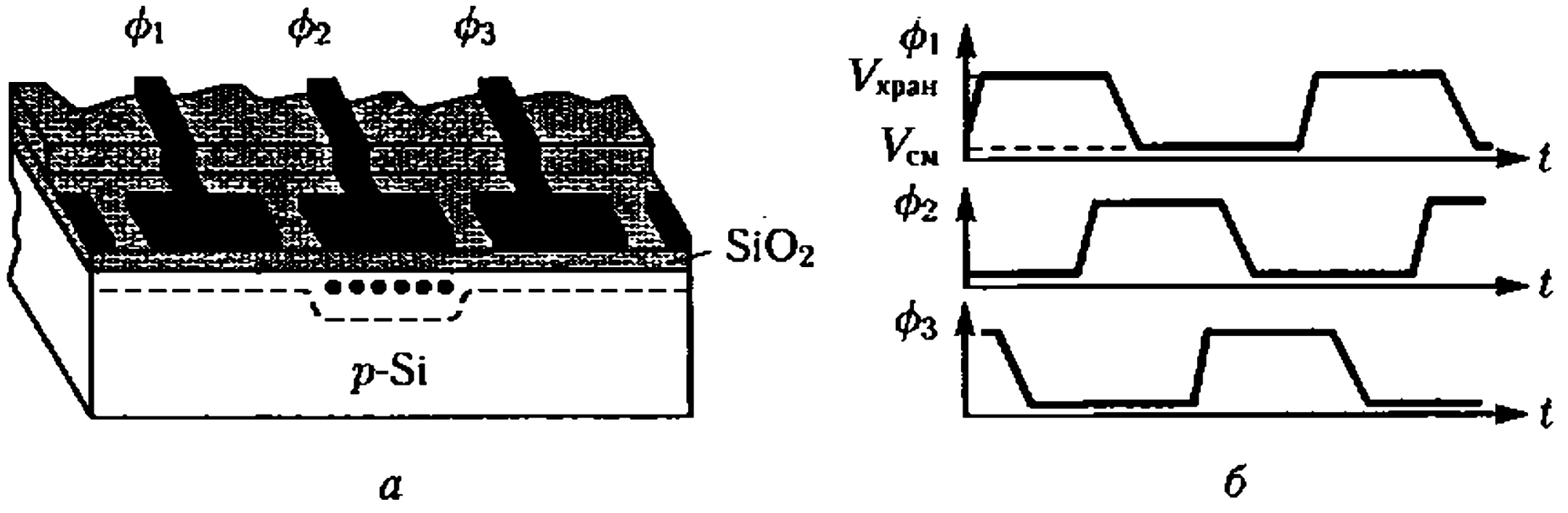


Рис. 5.2. Поперечное сечение регистра сдвига на ПЗС и потенциалы на его управляющих электродах в режиме хранения (*а*) и последовательность управляющих импульсов на тактовых шинах (*б*)

Эта величина обычно составляет $(0,3–1,6) \cdot 10^{-6}$ Кл/см^2 и ограничена сверху напряженностью электрического поля, при котором возникает пробой диэлектрика (характерная толщина окисла в ПЗС составляет 1000 Å).

Максимальное время хранения в ПЗС определяется как время, за которое неосновные носители, возникающие в результате тепловой генерации в полупроводнике, заполняют пустую потенциальную яму, находящуюся в режиме хранения. Значение соответствующего «темнового тока» определяется двумя величинами: плотностью поверхностных состояний на границе полупроводник–диэлектрик и концентрацией глубоких уровней в полупроводнике. В частности, в кремниевых ПЗС большую роль играет ток генерации, создаваемый глубокими уровнями примеси золота (остаточная концентрация которого может составлять 10^{11}–10^{12} см$^{-3}$), которые находятся в области изгиба зон. Для уменьшения темнового тока проводится геттерирование золота фосфором (см. с. 223). В кремнии максимальное время хранения составляет 1–10 с при 300 К.

Эффективность переноса в ПЗС. Очень важным параметром, характеризующим искажение сигнала при работе ПЗС, является *эффективность переноса* η. Эта величина определяется как доля заряда, переносимая из одной ямы в соседнюю за одну операцию переноса; типичные значения η лежат в пределах 0,99–0,999999. В качестве альтернативного параметра, описывающего эффективность переноса, часто используют величину *неэффективности* (*потерь*) переноса, которая определяется как $(1 - \eta)$. Поскольку цепочки ПЗС обычно состоят из тысяч последовательно включенных элементов, а полный коэффициент передачи цепочки равен η^N, где N — число элементов

в цепочке, то становится понятным стремление разработчиков сделать величину η как можно более близкой к единице. Если это будет не так, то при использовании ПЗС, например, в качестве запоминающего устройства ЭВМ «потеря контраста» между пакетами, отвечающими логическим 1, и пакетами, отвечающими логическим 0, может приводить к сбою в работе этих устройств.

Детальные исследования кремниевых ПЗС показали, что существует три механизма потерь переноса. Кроме возможной неполноты переноса заряда из одной ямы в другую (этот механизм потерь особенно характерен для схем пожарных цепочек в интегральном исполнении), существуют еще два механизма потерь, которые связаны с *захватом электронов из ям на поверхностные состояния* на границе раздела $Si–SiO_2$. Первый из этих механизмов связан с захватом электронов на медленные поверхностные состояния, которые характеризуются низким темпом выброса (см. с. 183); существование таких ловушек проявляется в особенно быстрой потере заряда в первом заполненном пакете, следующем за большой группой пустых пакетов. Второй механизм связан с захватом электронов на быстрые поверхностные состояния, которые характеризуются высоким темпом выброса; при этом часть захваченных носителей успевает вернуться в исходный зарядовый пакет за время переключения управляющего сигнала, а остальные носители попадают в следующий зарядовый пакет. Для механизмов, связанных с неполным переносом и захватом носителей на быстрые поверхностные состояния, характерно то, что суммарный заряд в пакетах не изменяется, а искажение сигнала проявляется в уменьшении «контраста» между соседними заполненными и пустыми пакетами по мере их продвижения в ПЗС.

Для уменьшения потерь переноса важно правильно выбрать значение минимального напряжения на управляющем электроде ($V_{см}$ на рис. 5.2*б*). Это значение должно быть таким, чтобы не подпускать дырки к границе раздела $Si–SiO_2$ и тем самым препятствовать их рекомбинации с захваченными на поверхностные состояния электронами. Обычно напряжение $V_{см}$ выбирается так, чтобы на поверхности полупроводника реализовывался режим обеднения или слабой инверсии.

К сожалению, величина потерь, связанных с захватом электронов на медленные поверхностные состояния в описанных выше ПЗС с *поверхностным каналом*, довольно велика и составляет 0,2–0,3%. Для борьбы с этими потерями в «нулевые» зарядовые пакеты часто вводят так называемый *фоновый заряд*

(порядка 10% от максимального заряда в яме). Этот заряд обеспечивает практически стационарное заполнение поверхностных состояний электронами и существенно ослабляет связанные с ними потери (до 0,01% [52]). Однако полностью избавиться от этих потерь невозможно, поскольку они начинают определяться краевыми эффектами, то есть тем, что площадь, занимаемая зарядовым пакетом, зависит от величины заряда. Краевые эффекты могут быть вызваны изменением размеров ямы как в поперечном (перпендикулярным направлению переноса), так и в продольном направлении. Из этого, в частности, следует, что для уменьшения потерь управляющие электроды должны следовать друг за другом без зазоров, а ширина канала в поперечном направлении должна быть достаточно большой.

В неэффективности переноса можно выделить статическую и динамическую компоненты. *Статическая компонента* в основном связана с обсуждавшимся выше захватом носителей из ямы на поверхностные состояния на границе полупроводник–диэлектрик. Эта компонента потерь существенно зависит от плотности поверхностных состояний.

Динамическая компонента связана с конечным временем перетекания носителей из одной ямы в другую. Если расстояние между электродами равно L, то постоянная времени процесса диффузионного переноса зарядового пакета составляет $4L^2/\pi^2 D_n$, где D_n — коэффициент диффузии электронов [212]. Для структуры с $L = 3$ мкм это время равно ≈ 1 нс. На самом деле перенос заряда между ямами протекает еще быстрее.

На начальном этапе переноса движение носителей происходит за счет так называемого *самоиндуцированного дрейфа*, связанного с электростатическим отталкиванием носителей, локализованных в потенциальной яме. После этого быстрого начального этапа уменьшение концентрации носителей происходит по экспоненциальному закону с указанной выше постоянной времени, а на заключительном этапе процесс вновь ускоряется за счет дрейфа носителей в *краевом поле*, возникающем в полупроводнике при подаче на два соседних электрода разных управляющих потенциалов. Напряженность этого поля в оптимально спроектированной структуре может достигать $2 \cdot 10^3$ В/см.

ПЗС с объемным каналом. Чтобы исключить высокие потери переноса, связанные с поверхностными состояниями, Бойл и Смит предложили конструкцию ПЗС с *объемным (скрытым) каналом* [215]. Особенностью этой конструкции ПЗС (см. рис. 5.3*а*) является то, что между слоем диэлектрика

и подложкой p-типа создается дополнительный слой n-типа проводимости толщиной несколько мкм [1]. Если на этот слой подать обратное смещение относительно подложки, то в структуре возникнет распределение потенциала, показанное на рис. 5.3*б* (при пустой потенциальной яме) или на рис. 5.3*в* (при заполненной электронами яме). Как и в ПЗС с поверхностным каналом, глубиной потенциальной ямы в этой конструкции можно управлять, меняя напряжение на управляющем электроде. Важно, что в этом ПЗС электроны движутся в канале, расположенном *в объеме* структуры, и поэтому они практически не взаимодействуют с поверхностными состояниями на границе $Si–SiO_2$. Это существенно уменьшает статическую компоненту потерь заряда, связанную с этими состояниями. Кроме того, для ПЗС с объемным каналом характерны более высокие краевые поля и более высокая подвижность носителей в канале (электроны не испытывают рассеяния на границе $Si–SiO_2$), что заметно увеличивает скорость переноса заряда из одной ямы в другую. Благодаря этому ПЗС с объемным каналом характеризуются более высоким быстродействием и более низкими потерями по сравнению с приборами с поверхностным каналом (в одном из ПЗС с объемным каналом частота переноса достигала 720 МГц при величине потерь 0,001 % [52]).

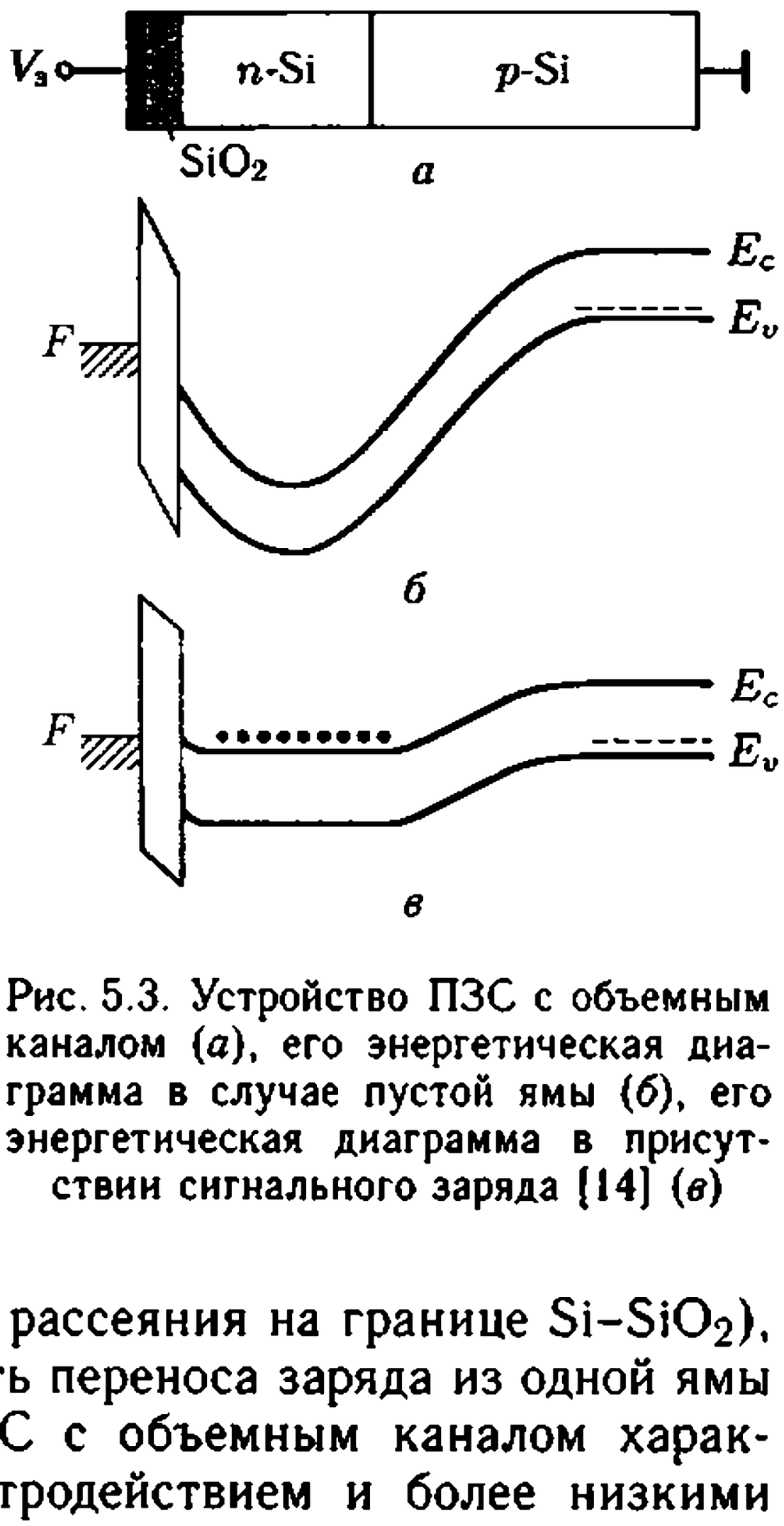


Рис. 5.3. Устройство ПЗС с объемным каналом (*а*), его энергетическая диаграмма в случае пустой ямы (*б*), его энергетическая диаграмма в присутствии сигнального заряда [14] (*в*)

Недостатком структур с объемным каналом является уменьшение величины максимальной плотности заряда $|qn_{ss}|_{max}$, которую можно удержать в яме при заданном напряжении на управляющем электроде. Расчеты показывают, что для однородно легированного дополнительного n-слоя толщиной d_n эта

---

[1]) Толщина и уровень легирования n-слоя выбираются так, чтобы в рабочем режиме этот слой был полностью обеднен электронами.

величина уменьшается в $1 + \varepsilon_д d_n/(2\varepsilon d_д)$ раз [212]. Правильно выбрав параметры структуры, в ПЗС с объемным каналом можно сохранить достаточно высокую величину $|qn_{ss}|_{max}$.

Рис. 5.4. Методы построения ПЗС с двухтактным управлением. Во всех трех конструкциях используется слой окисла переменной толщины (1000 и 4000 Å), а в конструкции *в* с помощью ионного легирования созданы дополнительные потенциальные барьеры [212]

Интересно, что конструкция ПЗС с объемным каналом позволяет в принципе обойтись и без слоя диэлектрика, если использовать в качестве управляющих электродов поверхностные p–n-переходы или барьеры Шоттки. Это не только упрощает технологию создания ПЗС-структур, но позволяет достигнуть бóльшей скорости переноса и лучшей стабильности, а также уменьшить амплитуду управляющих импульсов до ~2 В, что важно для уменьшения мощности потребления.

Возможность создания ПЗС без окисного слоя важна и для создания этих приборов из полупроводников, отличных от кремния, для которых задача получения высококачественного собственного окисла часто оказывается невыполнимой. Основная трудность при создании структур с управляющими электродами на основе барьеров Шоттки состоит в получении узких зазоров между электродами.

ПЗС с двухтактным управлением. Сложность электрической разводки тактовых шин в схемах с трехфазным управлением поставила задачу разработки таких конструкций ячеек, которые могли бы управляться с помощью *двух* тактовых шин. Понятно, что для направленного переноса зарядов в ПЗС энергетическая диаграмма таких ячеек должна быть асимметрична. Существует несколько конструкций ячеек, допускающих двухтактное управление (см. рис. 5.4). Для создания необходимого асимметричного потенциального рельефа в них используются слои окисла различной толщины, а для дополнительного

изменения потенциального рельефа в них может быть применено избирательное легирование приповерхностного слоя полупроводника с помощью ионной имплантации (см. области, помеченные крестиками на рис. 5.4*в*).

Альтернативный вариант конструкции приборов с зарядовой связью с двухтактным управлением основан на использовании описанной в начале главы идеи пожарных цепочек. При создании этих цепочек в интегральном исполнении (см. рис. 5.5) с помощью диффузии или ионной имплантации в полупроводнике

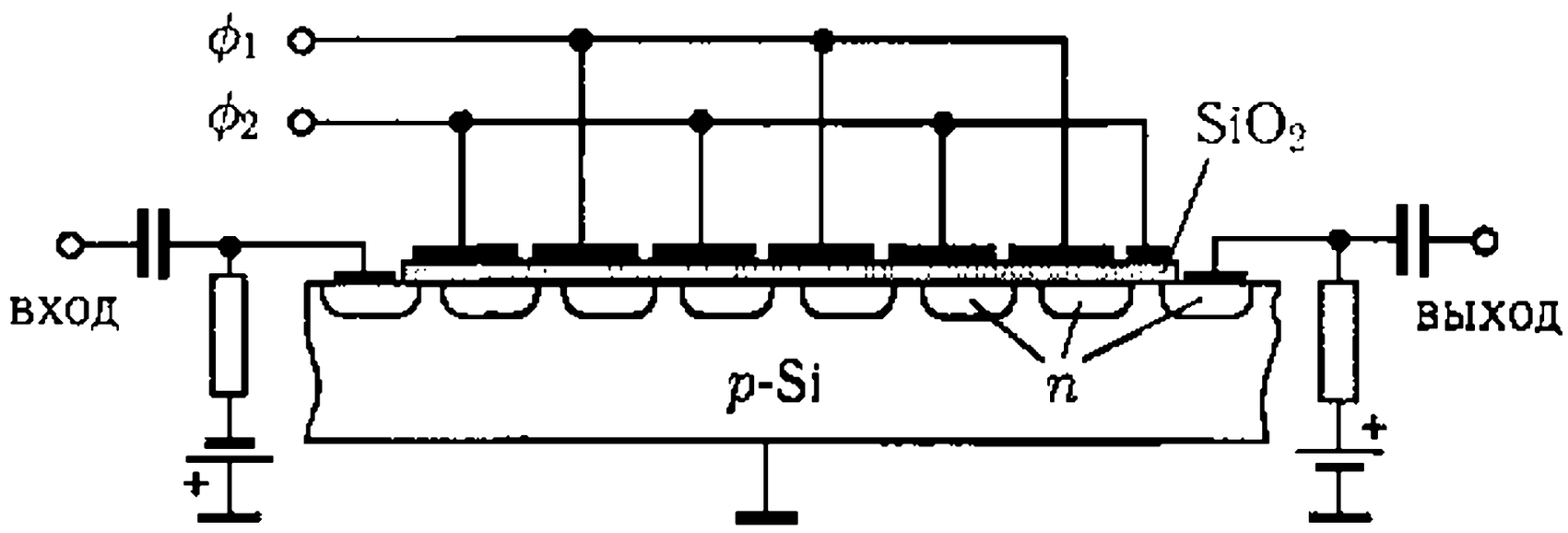


Рис. 5.5. ПЗС с двухтактным управлением, использующий идею «пожарной цепочки» [52]

создаются неглубокие p–n-переходы, над которыми на окисленной поверхности кремния асимметрично располагаются управляющие электроды. При подаче импульса напряжения на одну из тактовых шин одновременно происходит открывание каждого второго полевого транзистора и увеличение глубины потенциальных ям под протяженными участками управляющих электродов, на которые подано напряжение. Это и вызывает перемещение зарядовых пакетов из более мелких в соседние более глубокие потенциальные ямы.

Ввод и вывод информации в ПЗС. Ввод и вывод информации в ПЗС можно осуществить несколькими способами. Считывание информации на выходе ПЗС обычно производится электрическим способом, а ввод информации — либо электрическим, либо оптическим способом.

На рис. 5.6 показана схема задержки аналогового сигнала, построенная на основе ПЗС, в которой ввод и вывод информации осуществляется электрическим способом. Для этой цели в ней используются входной диод ID и выходной диод OD. Изменяя ток прямо смещенного входного диода, инжектирующего неосновные носители, можно управлять величиной заряда, который попадает в зарядовый пакет, формируемый под управляющим электродом ϕ_1. Далее эти пакеты перемещаются вдоль поверхности

полупроводника с помощью системы управляющих электродов (ϕ_1, ϕ_2, ϕ_3), а затем вновь преобразуются в аналоговый сигнал (импульсы тока) с помощью выходного диода OD, на который подано обратное смещение. Для корректного определения величины заряда в зарядовом пакете эти импульсы тока интегрируют с помощью специальной схемы зарядочувствительного усилителя на МОП-транзисторе, который создают при изготовлении ПЗС рядом с выходным диодом. Очень низкая емкость этой схемы обеспечивает существенное снижение уровня шума узла считывания [1]).

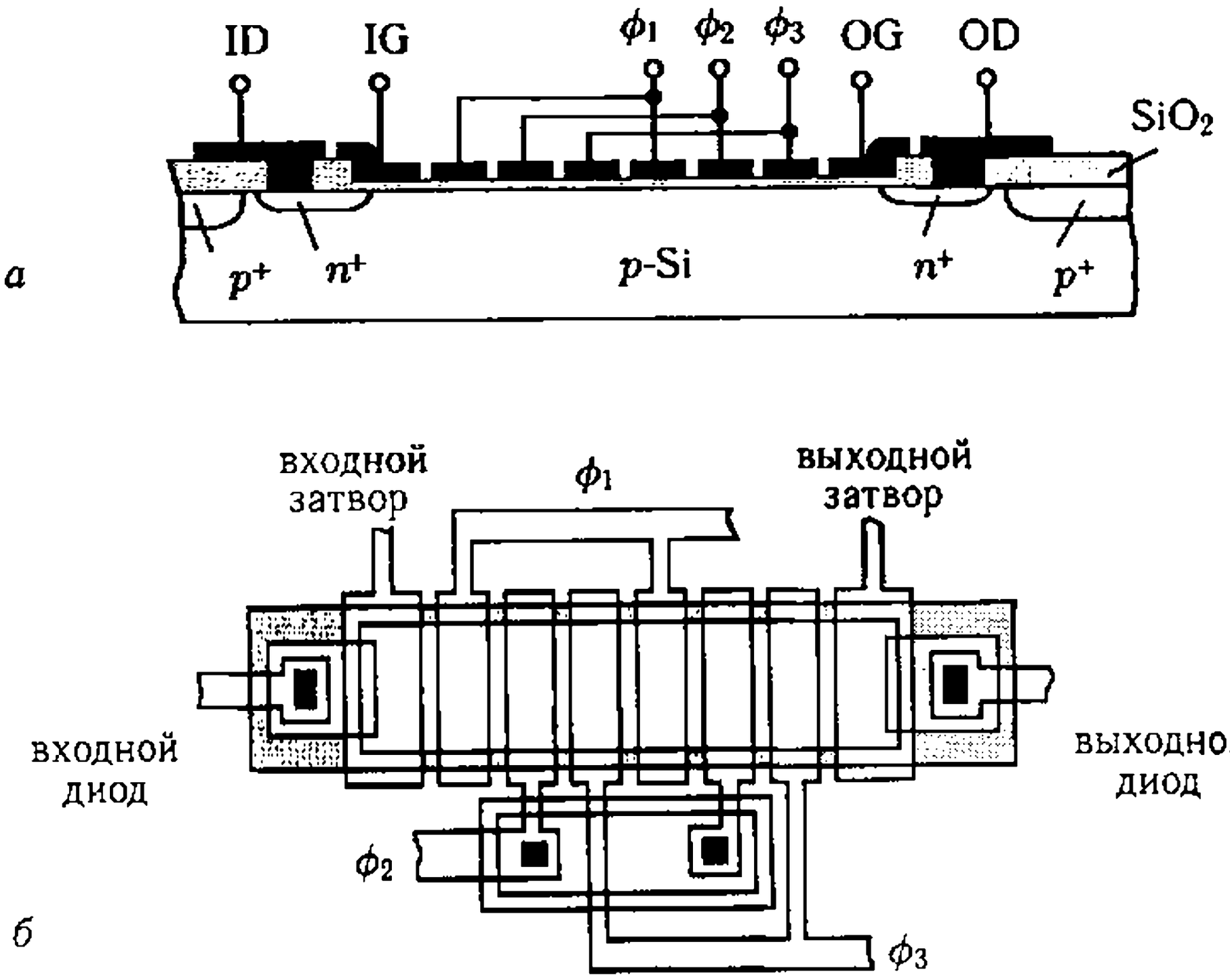


Рис. 5.6. Устройство (*а*) и топология (*б*) схемы задержки аналогового сигнала, созданной на основе ПЗС с каналом *n*-типа [14]

К сожалению, способ ввода информации с помощью модуляции тока входного диода характеризуется невысокой линейностью и достаточно высоким уровнем шума. Поэтому был предложен целый ряд модификаций схемы входного узла. Эти вопросы детально рассмотрены в книге [212] и здесь мы лишь укажем на другой режим работы описанного входного узла, в котором управление величиной заряда в зарядовых пакетах осуществляется не модуляцией тока диода ID, а изменением потенциала на

[1]) В ПЗС с объемным каналом шум узла считывания является основным источником шума приборов с зарядовой связью.

входном затворе IG (см. рис. 5.6) при фиксированной величине тока, протекающего через входной диод.

Описанный выше способ считывания информации с помощью обратно смещенного p–n-перехода обладает одним существенным недостатком: он является разрушающим, то есть после собирания заряда p–n-переходом зарядовый пакет исчезает. В то же время при построении сложных схем обработки сигналов желательно использовать такой способ измерения сигнального заряда, при котором сам зарядовый пакет оставался бы неизменным (*неразрушающий метод считывания*). Осуществить это можно введением в структуру ПЗС «плавающего затвора» (см. рис. 5.7). Проходящий под плавающим затвором зарядовый пакет наводит на нем потенциал, пропорциональный величине заряда. Плавающий затвор подключен к затвору расположенного рядом МОП-транзистора, на выходе которого и формируется импульс напряжения, пропорциональный величине сигнального заряда.

Приемники изображения на ПЗС. Ввод информации в приборах с зарядовой связью можно осуществлять не только электрическим способом, но и с помощью света. Последняя возможность оказывается особенно интересной и может быть использована для построения целого класса фоточувствительных приборов на ПЗС — *приемников изображения* [216, 217]. Работа этих приборов основана на том, что возбуждаемые при поглощении света электронно-дырочные пары разделяются электрическим полем области пространственного заряда МОП-конденсатора, причем неосновные носители затягиваются в потенциальные ямы, где за время экспозиции накапливается заряд, точно пропорциональный интенсивности света, падающего на данный участок поверхности полупроводника.

По способу ввода информации в приемники изображения их подразделяют на одномерные (линейные) и двумерные. *Одномерные приемники* более просты: они представляют собой линейку ПЗС, на которую проецируется изображение, механически сканируемое в направлении, перпендикулярном линейке. Такие приемники находят широкое применение в сканерах и аппаратах факсимильной связи.

Число фоточувствительных элементов в линейке зависит от области применения. Так, в современных сканерах их число находится в пределах 5–20 тысяч, хотя в некоторых сканерах число элементов может превышать 100 тысяч. Для обработки цветных изображений в сканерах используется блок из трех

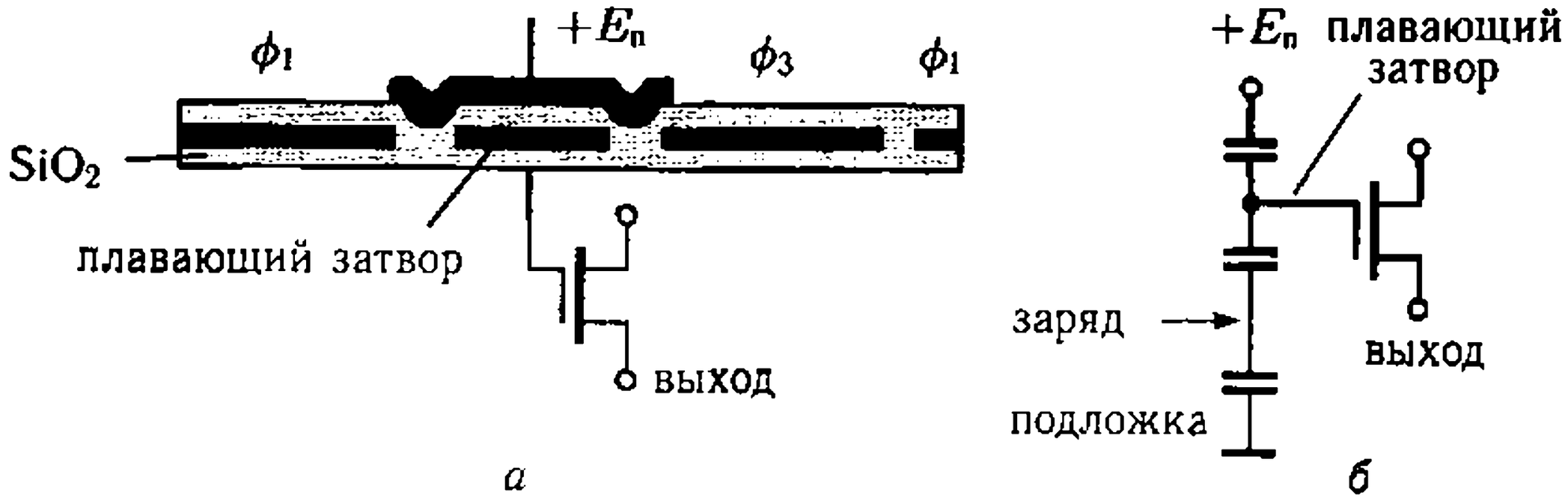


Рис. 5.7. Метод неразрушающего считывания информации с ПЗС с помощью плавающего затвора (*а*) и эквивалентная электрическая схема узла считывания (*б*)

близко расположенных одномерных линеек, закрытых красным, зеленым и синим светофильтрами.

Двумерные приемники изображения широко используются в телевизионных системах. При передаче телевизионного сигнала принято использовать чересстрочную развертку, при которой каждый передаваемый кадр изображения разбивается на два полукадра, состоящих из четных и нечетных строк. Удвоенная частота, с которой передаются полукадры, позволяет ослабить мерцание экрана и более качественно (без рывков) передавать изображение быстро движущихся объектов.

Для формирования чересстрочной развертки с помощью ПЗС было предложено два подхода. Первый из них проиллюстрирован на рис. 5.8. В этом подходе заряды, накапливаемые в двух группах фоточувствительных ячеек с полупрозрачными верхними электродами (*1* и *2*), которые отвечают четным и нечетным строкам, во время обратного хода кадровой развертки поочередно передаются в вертикальные регистры сдвига (ячейки хранения, защищенные от света) [1]. Каждый раз после того, как очередная строка передана (во время обратного хода строчной развертки) зарядовые пакеты в вертикальных регистрах сдвига перемещаются на строку вниз, а заряды из нижней строки передаются в горизонтальный регистр сдвига и во время прямого хода поэлементно поступают на усилитель считывания. По окончании передачи одного полукадра в ячейки хранения переносятся заряды, накопленные в фоточувствительных ячейках другого полукадра, и процедура поэлементного вывода изображения продолжается.

---

[1]) Защита ячеек от света необходима для того, чтобы на изображении не появлялись дефекты в виде полос, тянущихся в одном направлении от ярких точек на изображении.

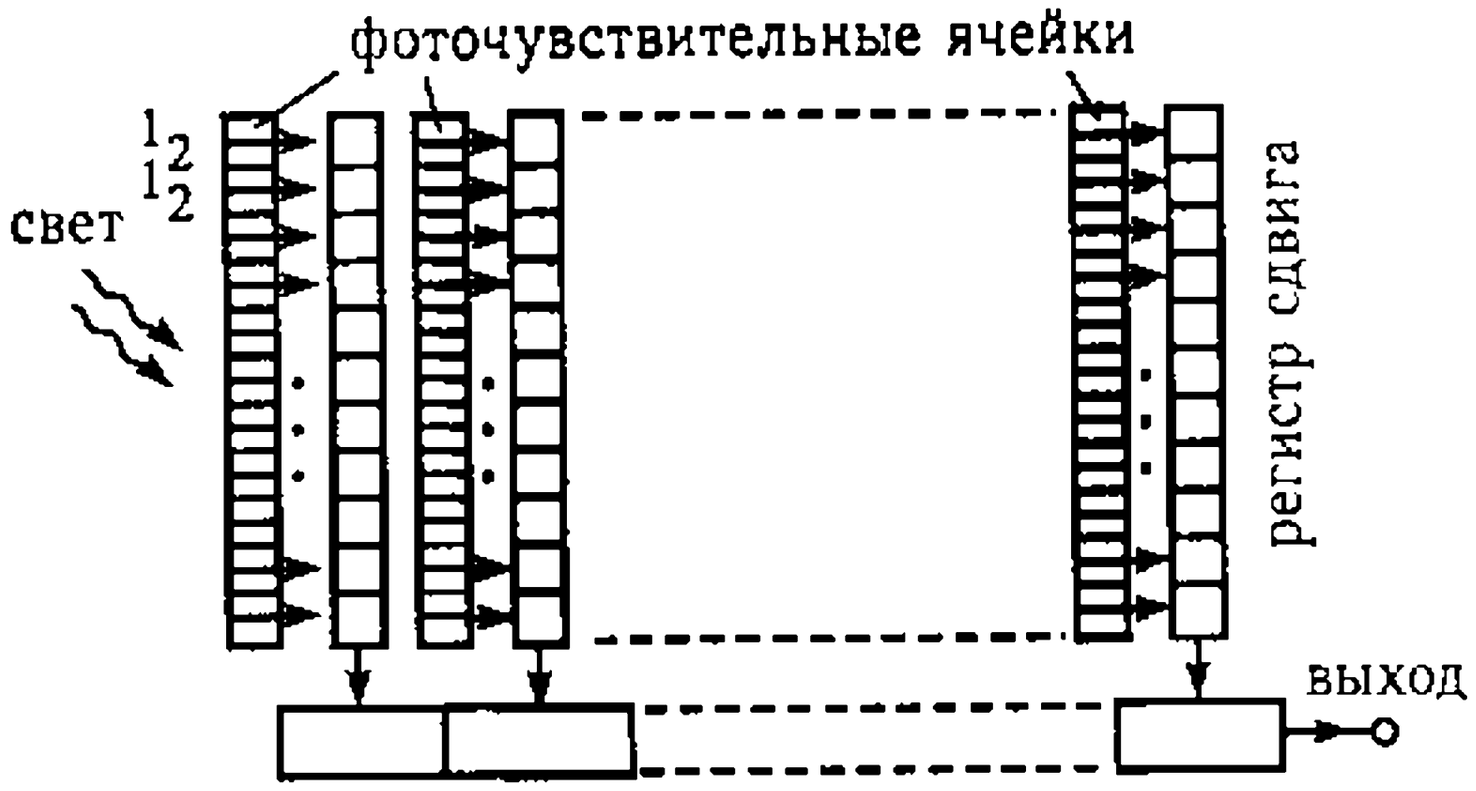


Рис. 5.8. Формирователь изображения на ПЗС [52]

Во втором подходе к формированию чересстрочной развертки фоточувствительные ячейки имеют вид вертикальных столбцов, простирающихся на всю высоту ПЗС, а изменение пространственного положения областей собирания носителей в конкретных местах этой ячейки осуществляется «электронным» способом за счет подачи определенных напряжений на проходящие над ячейками горизонтальные управляющие электроды [212]. Это позволяет перемещать центры потенциальных ям в фоточувствительных ячейках точно на половину расстояния между ячейками хранения и таким образом формировать четные и нечетные строки. Устройство вертикальных и горизонтальных регистров сдвига в этой конструкции аналогично рассмотренному выше.

Обе описанные конструкции приемников изображения на ПЗС позволяют одновременно производить накопление текущего кадра изображения в фоточувствительных ячейках и вывод предыдущего кадра из ячеек хранения. Управляющие электроды фоточувствительных ячеек обычно изготавливаются из поликристаллического кремния (слой поликремния толщиной 4000 Å достаточно прозрачен во всем диапазоне видимого света), а управляющие электроды ячеек хранения изготавливаются из какого-нибудь непрозрачного металла (например, алюминия).

Для создания цветных передающих камер для телевидения и видеокамер используются два варианта конструкций. В более дешевых камерах разделение цветов осуществляется путем нанесения на поверхность ПЗС-матрицы в строгом соответствии с положением фоточувствительных ячеек либо цветных полос, изготовленных из красителей, либо интерференционных фильтров. В более дорогих камерах одновременно используются три идентичных ПЗС, а цветоделение осуществляется с помощью

призменного устройства расщепления луча с дихроичными зеркалами.

Интересной особенностью ПЗС является то, что проецировать изображение на него можно как со стороны управляющих электродов, так и со стороны подложки. При освещении структуры со стороны контактов необходимо, чтобы электроды были полупрозрачными; обычно такие электроды изготавливают из поликремния. Хотя эффективность создания электронно-дырочной пары в полупроводнике при поглощении кванта света в области собственного поглощения близка к единице, следует иметь в виду, что при освещении структуры со стороны контактов из-за отражения света на нескольких границах кремния и поликремния с SiO_2 возникает интерференция, которая приводит к появлению дополнительных пиков и провалов на кривых спектральной чувствительности приемника, а из-за отражения света от структуры и поглощения света в электродах до фоточувствительных ячеек доходит лишь 35–50% падающих квантов. Поэтому характерное значение квантовой эффективности ПЗС-приемника при освещении со стороны контактов ограничивается указанными цифрами [1]).

При освещении ПЗС со стороны подложки квантовая эффективность и «гладкость» спектра фоточувствительности выше, однако пространственное разрешение приемника (то есть способность различать быстро чередующиеся светлые и темные линии на изображении) ниже, поскольку возбужденные у нижней границы подложки неосновные электроны в процессе диффузии к потенциальным ямам, расположенным на противоположной стороне пластины, могут заметно смещаться в поперечном направлении. Чтобы это диффузионное «размытие» было малым, подложка такого прибора должна быть очень тонкой. На интегральную чувствительность приемника также влияет и его конструкция; так, например, в конструкции, показанной на рис. 5.8, ячейки хранения закрыты системой непрозрачных металлических электродов, и поэтому интегральная чувствительность такой структуры оказывается вдвое ниже; цветные светофильтры также снижают чувствительность ПЗС-приемников. Для повышения интегральной чувствительности каждый элемент матрицы ПЗС может быть снабжен микролинзой, которая направляет падающий световой поток на фоточувствительную область ячейки.

---

[1]) Квантовую эффективность удается поднять до 70% если использовать контакты из прозрачных проводящих окислов металлов (In_2O_3–SnO_2).

Кроме более высокой квантовой эффективности по сравнению с современными вакуумными приемниками изображения (квантовая эффективность фотокатодов, используемых в супериконоскопах и суперортиконах, не превышает 20%), получению более высокой чувствительности ПЗС-приемников способствует и малая емкость усилителя считывания, который, в отличие от вакуумных приборов, можно создать на одном кристалле с матрицей ПЗС. Поэтому при приеме слабых сигналов уровень шума приемников изображения на ПЗС оказывается заметно ниже, а чувствительность — выше, чем у вакуумных приборов. По этой причине уже в 1989 году 97% всех приемников изображения, используемых в телевидении, создавались на основе ПЗС. В настоящее время приемники изображений на основе приборов с зарядовой связью широко используются в бытовой технике — видеокамерах и цифровых фотоаппаратах. Характеристики этих изделий (в частности, число элементов изображения) уже намного превзошли характеристики, принятые в телевидении. Так, число элементов разложения в цифровых фотоаппаратах уже перешагнуло границу в 12 миллионов, тогда как стандартный кадр телевизионного изображения в системах PAL и SECAM содержит примерно 250 тысяч элементов [1]).

Благодаря высокой чувствительности ПЗС-приемники находят широкое применение в системах регистрации слабых световых потоков, например, в астрономии; для уменьшения уровня шума при необходимости матрица ПЗС может дополнительно охлаждаться. Область спектральной чувствительности обычных кремниевых ПЗС простирается от 0,4–0,5 до 1,1 мкм. Чувствительность этих приборов в ближней инфракрасной области позволяет использовать их в охранных системах и приборах ночного видения (при инфракрасной подсветке объекта мощными светодиодами из GaAs). Более того, изготавливая ПЗС из полупроводников с различной шириной запрещенной зоны, можно

---

[1]) Следует отметить, что в середине 90-х годов у приемников изображения на основе ПЗС появились конкуренты — приемники изображения, построенные на основе КМОП-структур. Устройство этих приемников напоминает устройство запоминающих устройств (см. с. 273): каждая фоточувствительная ячейка имеет свой усилитель сигнала и может адресоваться с помощью адресной шины и передавать аналоговый сигнал по разрядной шине. К разрядным шинам подключены аналого-цифровые преобразователи, преобразующие сигнал в цифровую форму. Преимуществами этих приемников изображения являются существенно меньшая потребляемая мощность по сравнению с ПЗС и возможность использования всех достижений КМОП-технологии, позволяющая понизить их стоимость.

создать приемники изображения, работающие практически в любом спектральном диапазоне.

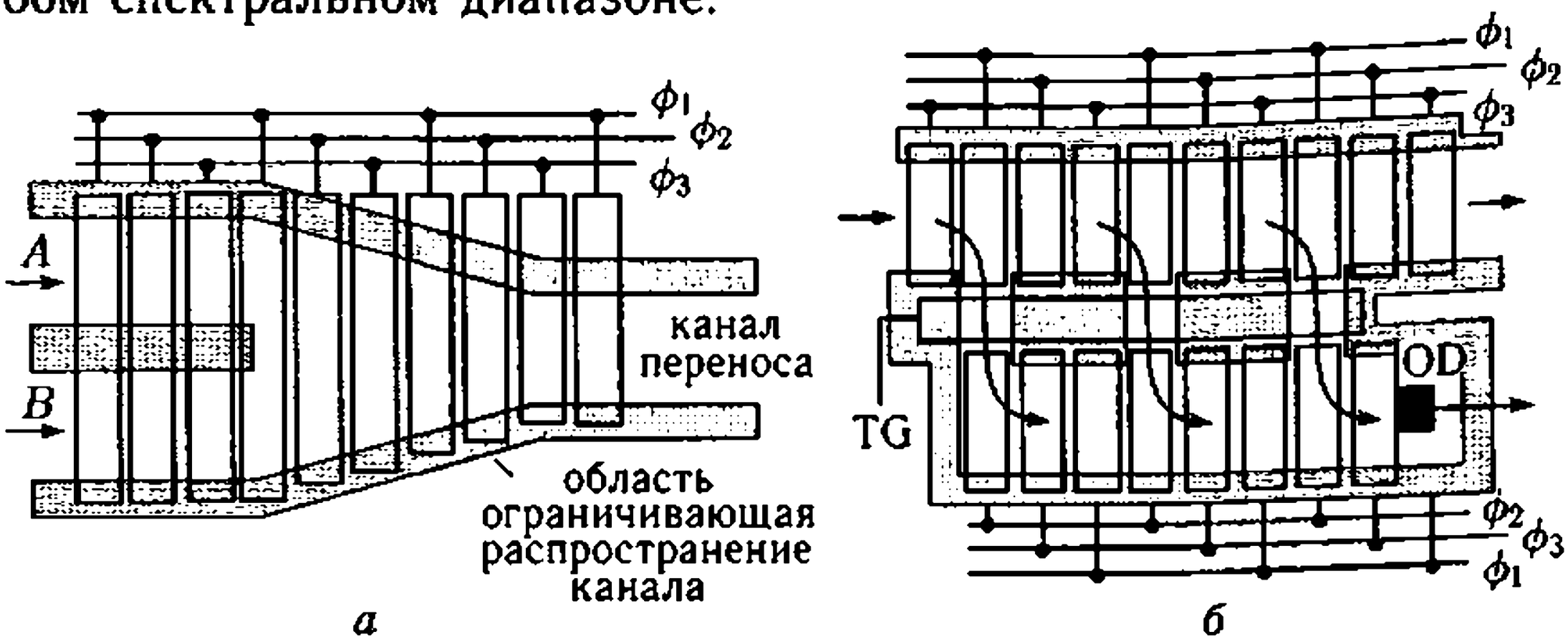


Рис. 5.9. Схемы объединения (*а*) и деления (*б*) зарядовых пакетов при аналоговой обработке сигналов [212]

Приемники изображения, работающие в инфракрасной области (ИК) спектра, необходимы во многих областях науки и техники. Среди возможных применений этих приемников можно указать решение задач разведки и наблюдения (в том числе и из космоса), создание приборов тепловидения для промышленных приложений и медицины. Наиболее актуальной является разработка приемников, работающих в окнах прозрачности атмосферы в диапазонах 2–2,5, 3,5–4,2 и 8–14 мкм, в которых ИК-излучение слабо поглощается присутствующими в воздухе углекислым газом и парами воды. Для тепловизионных приложений, регистрирующих собственное ИК-излучение объектов, необходимо достичь разрешения по температуре, равного 0,1 К. При одновременной засветке фотоприемника тепловым излучением фона с температурой 300 К для этого необходима высокая однородность полупроводника и идентичность характеристик фоточувствительных ячеек.

Существуют два типа приемников изображения на ПЗС, работающих в ИК-области: *монолитные* приборы, в которых фоточувствительная матрица и регистры сдвига изготавливаются из одного и того же полупроводника, и *гибридные* приборы, в которых схемы считывания и регистры сдвига ПЗС изготавливают из кремния, а фоточувствительные ячейки — из другого материала.

Монолитные ИК-приемники изображения из кремния используют два принципа: 1) фоточувствительность барьера Шоттки и 2) создание фоточувствительного слоя за счет легирования полупроводника. Первый из этих принципов основан на явлении фотоэмиссии электронов из металла при поглощении света в барьере Шоттки. Область спектральной чувствительности таких приемников определяется красной границей фотоэффекта, которая, например, составляет 5,5 мкм в барьере PtSi–*p*-Si. К сожалению, из-за низкого квантового выхода фотоэффекта (~1 %) эти приемники имеют невысокую чувствительность. Во втором способе в кремнии путем легирования его

различными примесями (P, Ga, In) создается фоточувствительный слой. При охлаждении прибора до температуры, отвечающей практически полному вымораживанию носителей на примесях, получается приемник, область спектральной чувствительности которого определяется энергией ионизации использованной примеси.

Монолитные приемники можно создавать и из других полупроводников; в этом плане были исследованы Ge, GaAs, InAs, InSb, $Hg_{0.8}Cd_{0.2}Te$, PbS, PbTe и $Pb_{0,76}Sn_{0,24}Te$, ширина запрещенной зоны в которых меньше, чем в Si. К сожалению, высокая плотность поверхностных состояний на границе этих полупроводников с диэлектриками мешает получить достаточно высокую эффективность переноса в изготовленных из них монолитных приемниках.

Гибридные приемники могут использовать как прямую инжекцию заряда из массива внешних детекторов (например, фотодиодов), так и «непрямую инжекцию». В последнем случае сигнал, снимаемый с фоточувствительного элемента (например, с фоторезистора), управляет величиной заряда, который вводится в зарядовые пакеты кремниевого ПЗС. Обе описанные конструкции гибридных приемников ИК-изображения требуют охлаждения до температур, оптимальных для работы используемых фоточувствительных элементов. Охлаждения приемника не требуется, если в качестве датчика использовать кристалл *пироэлектрика*. Нагрев пироэлектрика (например, триглицинсульфата) в результате поглощения ИК-излучения приводит к изменению величины спонтанной поляризации этого кристалла и возникновению напряжения на его гранях (порядка 200 мВ/град). Это напряжение можно использовать для управления величиной инжектируемого в ПЗС заряда. Недостатком пироэлектрических детекторов является их высокая инерционность и низкая чувствительность.

Аналоговые и цифровые функциональные устройства на ПЗС. Возможность перемещения зарядовых пакетов с помощью системы управляющих электродов позволяет создавать на основе ПЗС многофункциональные устройства обработки аналоговой информации [213]. Создавая определенную конфигурацию электродов, можно объединять и делить зарядовые пакеты (см. рис. 5.9), осуществлять задержку сигнала на заданное время. Все это позволяет легко реализовать такие практически важные устройства, как цифровые фильтры [1]. Эти устройства находят широкое применение при обработке речевых сигналов, в системах связи, радиолокации. Аналоговые линии задержки

---

[1] Цифровой фильтр — это электронное устройство, производящее взвешенное суммирование исходного сигнала с несколькими задержанными на различное время копиями того же сигнала. Выбирая весовые коэффициенты и времена задержки, можно реализовать практически любую частотную характеристику цифрового фильтра.

используются в схемах восстановления цветовых сигналов в цветном телевидении, а перестраиваемые линии задержки — в схемах коррекции временных искажений в устройствах видеозаписи.

Возможность одновременно осуществлять оптический и электрический ввод сигналов в ПЗС приводит к интересной возможности аналоговой обработки изображений непосредственно в этих приборах. Так, вводя в ячейки электрический сигнал, инвертированный по отношению к сигналу, полученному при закрытом оптическом затворе, удается эффективно компенсировать дефекты изображения, в частности, «геометрический шум», возникающий из-за неоднородности темнового тока на различных участках фоточувствительной матрицы. Другим применением аналоговой обработки изображений, основанным на вычитании из текущего кадра предыдущего кадра изображения, является выделение движущихся объектов, которое используется в радиолокации и охранных системах.

В связи с быстрым развитием МОП-транзисторов и интегральных схем на их основе в настоящее время ПЗС более не используются для создания цифровых ИС. Однако в связи с неуклонным развитием электроники в сторону наноразмерных систем, в которых логическая информация скорее всего будет представляться единичными зарядами, описываемые ниже решения, найденные для ПЗС, возможно, обретут вторую жизнь.

Использование ПЗС в цифровых устройствах в основном связано с созданием запоминающих устройств, хотя возможны и другие применения [213]. Основным элементом запоминающего устройства на ПЗС (см. рис. 5.10) является регистр сдвига, для компенсации потерь переноса в котором используется специальная *схема регенерации*. Создать схему регенерации довольно просто: для этого плавающую диффузионную область выходного диода одной цепочки (OD на рис. 5.6) достаточно подключить к управляющему электроду IG входной цепи следующей цепочки ПЗС. Эта схема по сути является инвертором (схемой «НЕ»). Действительно, если на выходе первой цепочки появляется заполненный электронами пакет (отвечающий состоянию логической 1), то потенциал электрода IG понижается и носители перестают инжектироваться во вторую цепочку, а если на выходе первой цепочки появляется пустой пакет (логический 0), то происходит инжекция во вторую цепочку. Правильно подобрав потенциалы на электродах, можно сделать схему инвертора нечувствительной к небольшой «деградации» величин зарядов, отвечающих логическому 0 и логической 1. Последовательное включение двух цепочек с двумя схемами регенерации по сути и представляет собой запоминающее устройство с последовательным доступом, которое является электронным аналогом магнитного диска.

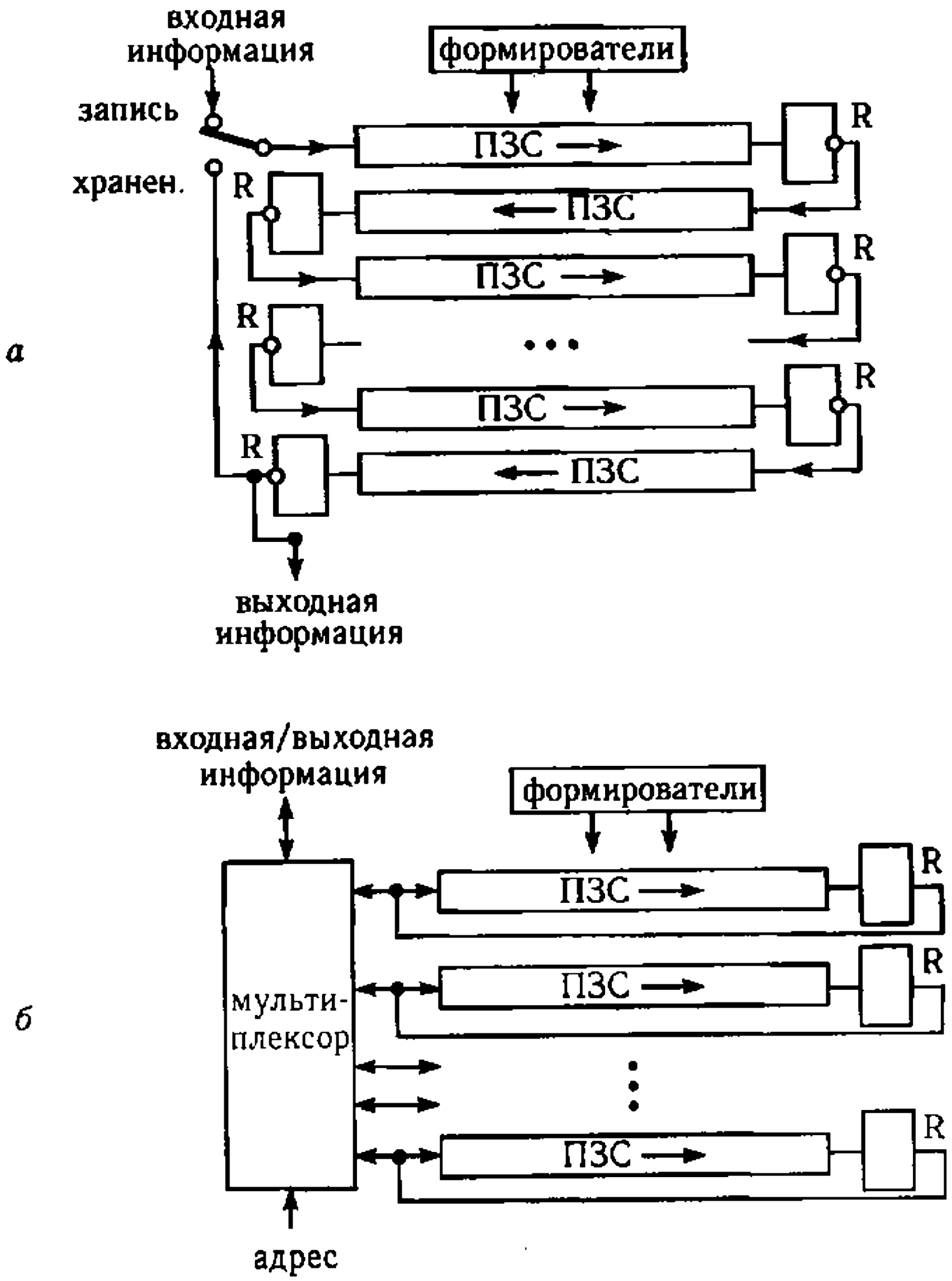


Рис. 5.10. Серпантинная (*а*) и петлевая (*б*) организация запоминающих устройств на ПЗС. R — схема регенерации сигнала [212]

Существуют два варианта организации запоминающих устройств на ПЗС (см. рис. 5.10). При серпантинной организации для доступа к необходимой информации приходится считывать все расположенные ранее данные (последовательный доступ к данным), но при этом схема управления устройством очень проста. При петлевой организации время доступа к данным существенно быстрее (память организована в виде параллельных «дорожек»), но при этом схема управления становится сложнее. Недостатком описываемых запоминающих устройств является их достаточно высокое энергопотребление, которое связано с необходимостью постоянного перемещения зарядов в цепочках ПЗС.

На основе приборов с зарядовой связью можно строить различные логические элементы, такие как схемы «И-НЕ» и «ИЛИ-НЕ», которые необходимы для создания периферийных схем запоминающих

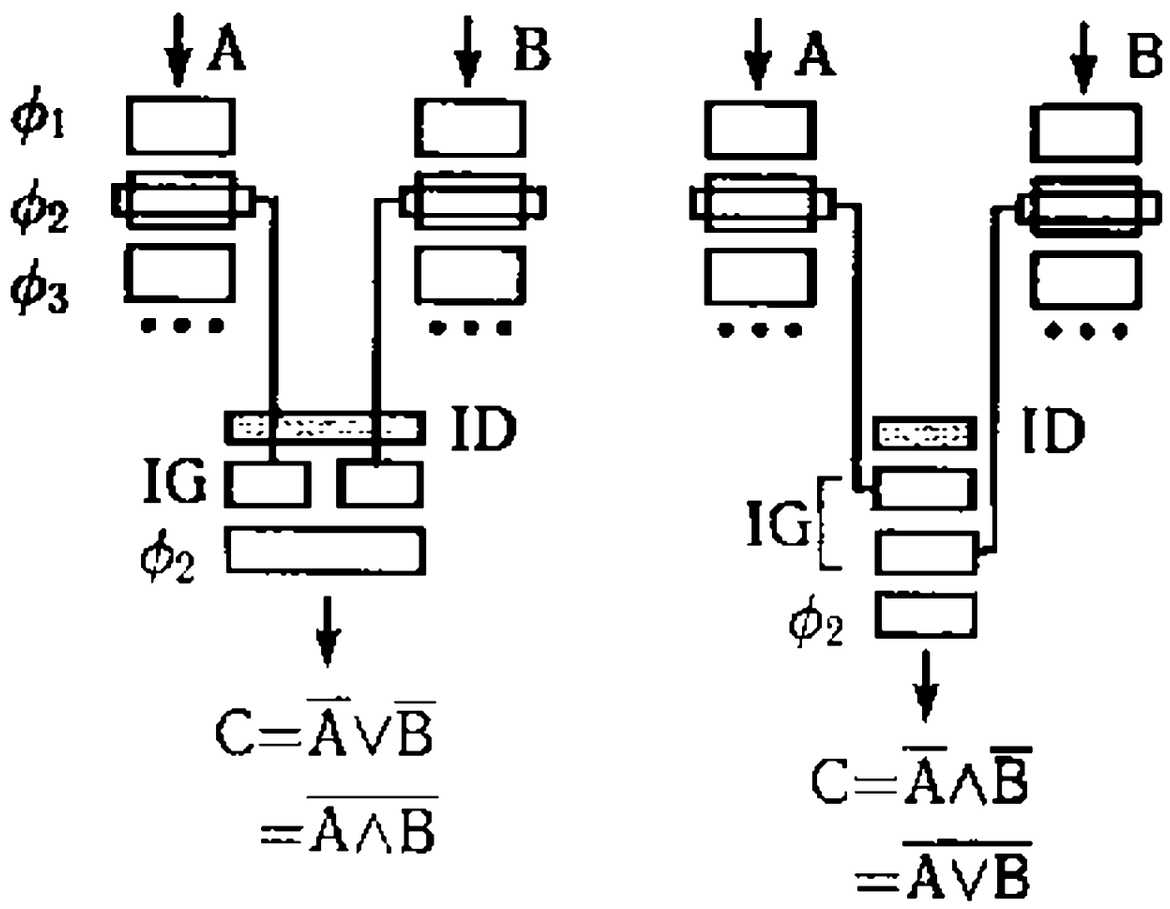


Рис. 5.11. Схема реализации логических функций «ИЛИ» и «И» путем создания соответствующей конфигурации управляющих электродов во входном узле ПЗС [213]

устройств. Функцию «ИЛИ» можно реализовать слиянием двух зарядовых пакетов (при переполнении ямы избыточный заряд из нее удаляется специальным устройством), а функцию «И» реализовать путем слияния зарядовых пакетов и проверкой потенциальной ямы на переполнение [213]. На рис. 5.11 показана более сложная схема, позволяющая организовать потоковую обработку данных. Используя идею неразрушающего считывания и подавая с плавающих затворов двух ПЗС-цепочек A и B электрические сигналы на включенные параллельно или последовательно управляющие затворы IG входного узла третьей ПЗС-цепочки C, можно реализовать указанные логические функции.

Шумы в приборах с зарядовой связью. Вопрос о шумах, присущих ПЗС, имеет первостепенное значение при использовании этих приборов в качестве приемников изображения и в схемах аналоговой обработки сигналов.

Поскольку зарядовые пакеты состоят из электронов, имеющих фиксированную величину заряда, большинство компонентов шума ПЗС носит дробовой характер [212, 213]. В структурах, для которых характерна неполнота переноса заряда (например, для пожарных цепочек), перемещение пакетов сопровождается появлением *шумов переноса*, которые связаны с флуктуациями величины заряда, теряемого при переносе пакета из одной ямы в другую. Однако даже если допустить, что заряд переносится полностью, его величина все равно флуктуирует из-за *шумов поверхностных состояний*, которые определяются флуктуацией числа электронов, захватываемых и выбрасываемых с поверхностных ловушек. Основной вклад в этот тип шума дают ловушки, темп выброса с которых близок к частоте переноса. Флуктуации заряда, связанные с поверхностными состояниями, равны

$$\Delta Q_s^2 = q^2 kT N_{ss} A \ln 2,$$

где N_{ss} — плотность поверхностных состояний, а A — площадь элемента. Спектральная плотность этого шума не зависит от частоты, поскольку энергетическое распределение поверхност-

ных состояний непрерывно. Глубокие уровни, присутствующие в обедненном слое, вызывают *шум объемных ловушек*. Этот тип шума также связан с процессами захвата и выброса электронов, но из-за дискретности энергетических уровней этот шум сильно зависит от температуры и частоты. Наконец, дополнительный шум в работе ПЗС создают флуктуации заряда при вводе и выводе информации. Шумы при вводе информации связаны с *шумами электрической инжекции* (при электрическом вводе) или *флуктуациями светового потока* (при оптическом вводе). Шумы при выводе информации в основном определяются тепловым шумом схемы считывания; при емкости узла считывания C их величина равна $\Delta Q^2 = kTC$. Полный шум современных ПЗС составляет 100–500 «шумовых» электронов/пакет, а в охлаждаемых устройствах, используемых в астрономии, — 2–6 электронов/пакет.

Глава 6

ПОЛУПРОВОДНИКОВЫЕ СВЧ ПРИБОРЫ

Область частот 3–300 ГГц, которая включает диапазоны сантиметровых, миллиметровых и субмиллиметровых (до 0,1 мм) длин волн, называется областью сверхвысоких частот (СВЧ). Выделение этого диапазона частот в отдельную область связано с тем, что техника СВЧ (волноводы, объемные резонаторы, полосковые линии и т. п.) существенно отличается от техники, используемой в области более длинных волн. Существуют два класса полупроводниковых приборов, используемых в технике СВЧ: детекторы высокочастотных колебаний и приборы для генерации и усиления таких колебаний.

Диоды, служащие для детектирования сигналов, в принципе мало чем отличаются от полупроводниковых диодов среднечастотного диапазона. Обычно в области СВЧ используются диоды, работающие на основных носителях заряда: обращенные диоды (см. п. 1.4.4) и диоды с барьером Шоттки (п. 1.5). Размеры этих диодов должны быть очень малы, что необходимо для получения малых емкости и индуктивности прибора.

Группа приборов, служащих для генерации и усиления колебаний, использует новые, не обсуждавшиеся нами ранее физические принципы. Их-то мы и рассмотрим в этой главе.

Субмиллиметровая (терагерцевая) часть диапазона СВЧ, в которой лежат частоты огромного числа электронных переходов в молекулах, в последнее время вызывает все больший интерес исследователей. Работа в этой области открывает огромные возможности для исследования атмосферы Земли, планет и межзвездного вещества методами молекулярной спектроскопии. Изучение этими методами химических реакций, идущих в верхних слоях атмосферы, важно для понимания глобальных экологических проблем (озоновые дыры, радиационный

баланс и потепление климата, контроль за загрязнением атмосферы). Около 40 тысяч линий тонкой структуры, обнаруженных в спектрах излучения межзвездного вещества, еще ждут расшифровки и, как ожидается, дадут важную информацию об эволюции Вселенной. Кроме того, субмиллиметровое излучение широко используется в диагностике плазмы и медицине. Для решения указанных задач требуются источники излучения, мощность которых составляла бы по крайней мере 0,5 мВт. Задачу создания таких источников и пытается решить современная полупроводниковая электроника СВЧ.

Как известно из радиотехники, колебания, возбуждаемые в электронных схемах, построенных из пассивных элементов, всегда затухают. Примером такой схемы может служить колебательный контур, в котором сопротивление материала, из которого сделана катушка индуктивности, и сопротивление утечки конденсатора приводят к диссипации энергии и затуханию возбужденных в контуре колебаний. Для получения незатухающих колебаний в схему необходимо включить активный элемент, который перераспределяет энергию внешнего источника так, что амплитуда колебаний начинает нарастать. Включение таких активных элементов на радиотехническом языке описывается подключением к колебательному контуру элемента, имеющего *отрицательное дифференциальное сопротивление* $dU/dI < 0$, то есть элемента, в котором протекающий переменный ток находится в противофазе с переменным напряжением.

В последующих разделах мы рассмотрим две ситуации, в которых в полупроводниковых приборах появляется отрицательное дифференциальное сопротивление:

1) случай, когда объемное отрицательное дифференциальное сопротивление возникает в полупроводнике, находящемся в сильном электрическом поле, и
2) случай, когда динамическое отрицательное дифференциальное сопротивление возникает на высокой частоте из обычного положительного сопротивления за счет специально создаваемой в приборе временно́й задержки.

6.1. Диоды Ганна

Отрицательное дифференциальное сопротивление в диодах Ганна возникает за счет эффекта *междолинного переноса*. Этот эффект проявляется в некоторых полупроводниках, помещенных в сильное электрическое поле, и приводит к N-образной зависимости плотности тока через образец от приложенного к нему электрического поля [218].

Развитие физики полупроводников в конце 50-х–начале 60-х годов характеризует большой интерес к изучению возможности возбуждения электрических колебаний в однородных полупроводниках. В 1958 г. Крёмер, основываясь на особенностях закона дисперсии тяжелых дырок в Ge, высказал идею СВЧ генератора на носителях с отрицательной эффективной массой, однако возможность ее практической реализации казалась маловероятной. В 1961–62 гг. Ридли и Уоткинс [219] и независимо Хилсум [220] теоретически исследовали возможность возникновения электрической неустойчивости в модели, в которой электрическое поле разогревает носители и вызывает их переход из зоны с более высокой подвижностью в зону с более низкой подвижностью. Авторы показали, что при определенных условиях это может приводить к появлению на вольт-амперных характеристиках образца участка с отрицательным дифференциальным сопротивлением. Расчеты Ридли и Уоткинса в основном ориентировались на p-Ge, подвергнутый одноосной деформации, а Хилсум в качестве модельных объектов выбрал GaSb и GaAs n-типа проводимости.

Таблица 6.1. Полупроводники, в которых наблюдается эффект междолинного переноса электронов [14, 218, 223]

Полупроводник	E_g, эВ	ΔE, эВ	$\mathcal{E}_T$, кВ/см
GaAs	1,42	0,28–0,36 (Γ–L)	3,2
InP	1,34	0,53–0,67 (Γ–L)	10,5
CdTe	1,53	0,51 (Γ–L)	13
ZnSe	2,70		38
InSb [1)]		(Γ–L)?	0,5
InAs [2)]		(Γ–L)	1,6
Ge [3)]			
$Ga_xIn_{1-x}Sb$ ($x = 0{,}4$–$0{,}91$)		(Γ–L)	0,43–0,79
$GaAs_xP_{1-x}$ ($x < 0,34$)		(Γ–L,X) [4)]	
$Al_xGa_{1-x}As$ ($x < 0,25$)		(Γ–L)	
$GaAs_{1-x}Sb_x$ ($x \leqslant 0,07$)			
$In_{0,16}Ga_{0,84}As$			~7
$InAs_{0,2}P_{0,8}$	1,10	0,95 (Γ–L)	5,7
GaN	3,39	2,1	80–150 (расч.)

[1)] При 77 К и давлении 12 кбар, когда ΔE становится меньше E_g.
[2)] При одноосном сжатии (14 кбар) вдоль <111>, когда ΔE становится меньше E_g.
[3)] Для германия известно свыше десяти видов электрической неустойчивости, часть которых по своей природе родственны эффекту Ганна (подробнее см. [218]).
[4)] При $x > 0,28$ зазор $\Delta E_{\Gamma-X}$ становится меньше $\Delta E_{\Gamma-L}$.

В 1963 г. Иан Ганн из фирмы IBM, занимавшийся явлениями, связанными с горячими носителями в полупроводниках, обнаружил, что в n-GaAs и n-InP в импульсных электрических полях возникают электрические колебания с частотой в несколько ГГц при превышении электрическим полем некоторой пороговой величины (2–4 кВ/см) [221]. Частота колебаний определялась длиной образца, не зависела от параметров внешней цепи, удельного сопротивления и кристаллографической ориентации исследованных образцов и была близка к обратному времени пролета электронов через образец. Эффект был интересен еще и потому, что эффективность преобразования постоянного тока в высокочастотные колебания уже в первых экспериментах оказалась довольно высокой (1–2%). Любопытно, что сам Ганн, зная о существовании работы Ридли и Уоткинса, склонялся к другому объяснению обнаруженного им эффекта, и только Крёмер [222] показал, что все основные черты обнаруженного Ганном явления соответствуют теоретическим моделям Ридли, Уоткинса и Хилсума.

Эффект междолинного переноса, названный по имени его первого исследователя эффектом Ганна, позже был обнаружен и в ряде других полупроводников (см. табл. 1).

Рассмотрим механизм формирования отрицательного дифференциального сопротивления на примере GaAs и InP n-типа проводимости — двух наиболее важных для создания диодов Ганна материалов. Зонная структура этих полупроводников (см. рис. 6.1) характеризуется наличием в зоне проводимости, кроме основного минимума в Г-точке зоны Бриллюэна, двух побочных

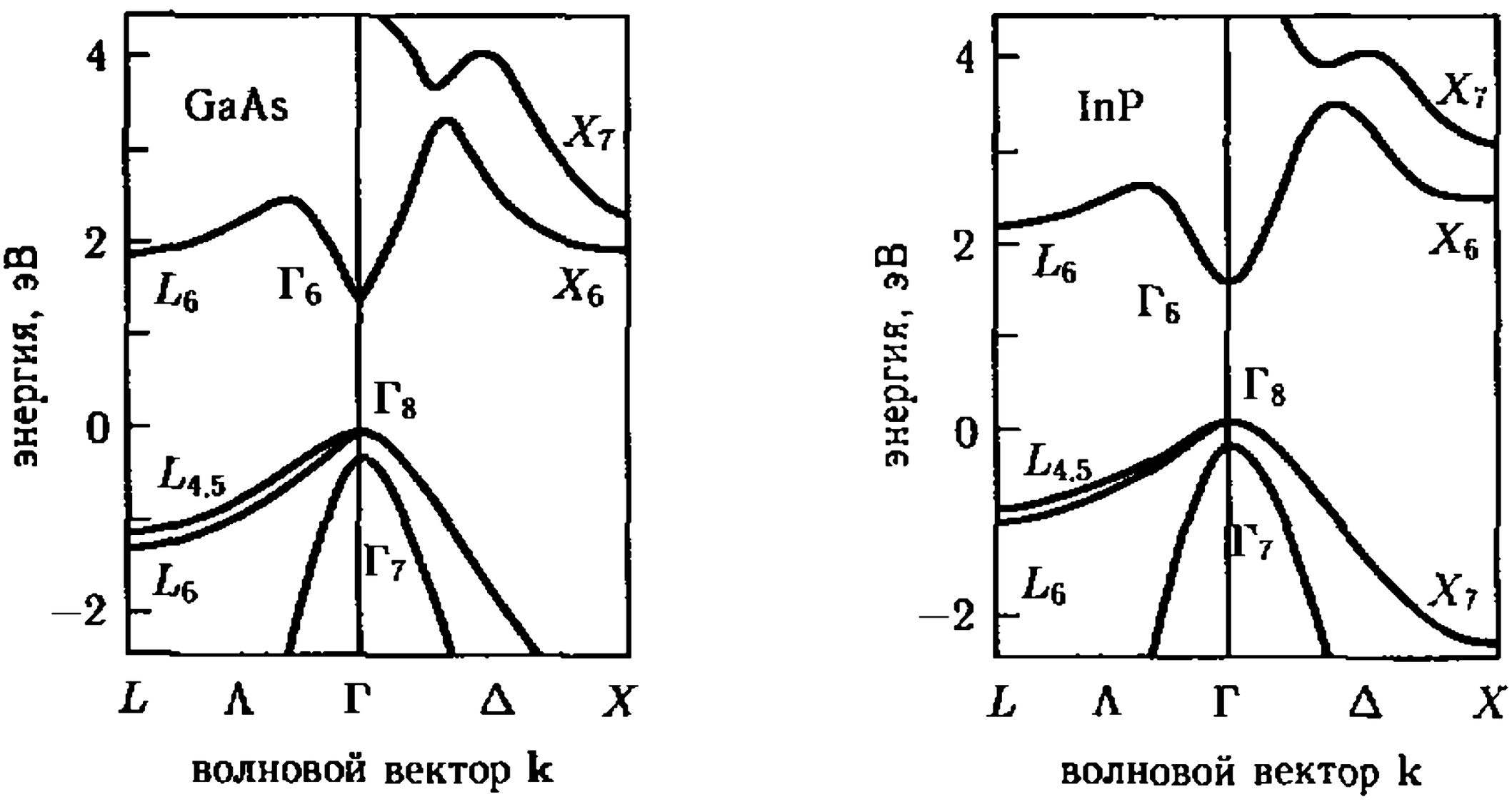


Рис. 6.1. Зонная структура GaAs и InP

минимумов в точках L и X. Энергетический зазор между основным минимумом и ближайшим к нему по энергии минимумом в точке L составляет $\Delta E \approx 0,3$ эВ в GaAs и $\approx 0,6$ эВ в InP.

В обоих материалах эффективная масса электронов в побочных минимумах намного больше, а подвижность — намного меньше, чем в основном минимуме.

Поскольку энергетический зазор между минимумами достаточно велик ($\Delta E \gg kT$), то в слабом электрическом поле подавляющее большинство электронов находится в основном минимуме, и плотность тока через образец определяется подвижностью электронов μ_1 в этом минимуме:

$$J \approx q\mu_1 n_0 \mathcal{E}. \tag{6.1}$$

Здесь $\mathcal{E}$ — напряженность электрического поля [1]), а n_0 — концентрация электронов.

С ростом напряженности электрического поля электронный газ разогревается и горячие электроны, энергия которых превышает ΔE, начинают переходить в побочный минимум. Если $\Delta E < E_g$, то процессами ударной ионизации (и, следовательно, возбуждением дырок) при этом можно пренебречь. В этом случае полная концентрация электронов не меняется, а электроны лишь перераспределяются между основным и побочными минимумами. Если обозначить концентрации электронов в этих минимумах через n_1 и n_2 (при этом $n_1 + n_2 = n_0 = \text{const}$) и предположить, что подвижности электронов в минимумах не зависят от $\mathcal{E}$, то плотность тока можно записать так:

$$J \approx q(\mu_1 n_1 + \mu_2 n_2)\mathcal{E}, \tag{6.2}$$

где μ_2 — подвижность электронов, находящихся в побочном минимуме.

Предполагая, что распределение электронов по энергиям в полупроводнике можно описать распределением Больцмана с некоторой *электронной температурой* T_e, отличной от температуры решетки T, величину T_e можно рассчитать, приравняв скорость набора энергии электроном в электрическом поле скорости ее диссипации:

$$q\mathcal{E}v(\mathcal{E}) = \frac{3}{2}\frac{k(T_e - T)}{\tau_e}, \tag{6.3}$$

где $v(\mathcal{E}) = (\mu_1 n_1 + \mu_2 n_2)\mathcal{E}/n_0$ — средняя скорость дрейфа, а τ_e — время релаксации энергии электронов (его характерное значение

---

[1]) Строго говоря, в этой и последующих формулах следовало бы выражать ток не через подвижность, а через скорость дрейфа электронов, поскольку в сильных электрических полях скорость дрейфа становится нелинейной функцией напряженности поля (мы уже говорили об этом на с. 167). Однако для качественной модели эффекта Ганна это не очень существенно.

составляет 10^{-12} с). Отношение заселенностей минимумов для электронов с температурой T_e равно

$$\frac{n_2}{n_1} = C \exp\left(-\frac{\Delta E}{kT_e}\right), \tag{6.4}$$

где $C = (m^*_{2c}/m^*_{1c})^{3/2}$ — отношение плотностей состояний, которое определяется значениями эффективных масс плотности состояний m_{ic} в соответствующих минимумах. Разрешая уравнение (6.3) в приближении $\mu_2 \ll \mu_1$ и подставляя в него отношение n_1/n_0, полученное из уравнения (6.4), мы приходим к следующему трансцендентному уравнению для зависимости $T_e(\mathcal{E})$:

$$T_e = T + \frac{2q\tau_e\mu_1}{3k}\mathcal{E}^2\left[1 + C\exp\left(-\frac{\Delta E}{kT_e}\right)\right]^{-1}. \tag{6.5}$$

При этом зависимость средней скорости дрейфа от напряженности электрического поля описывается формулой

$$v(\mathcal{E}) = \mu_1\mathcal{E}\left[1 + C\exp\left(-\frac{\Delta E}{kT_e}\right)\right]^{-1}. \tag{6.6}$$

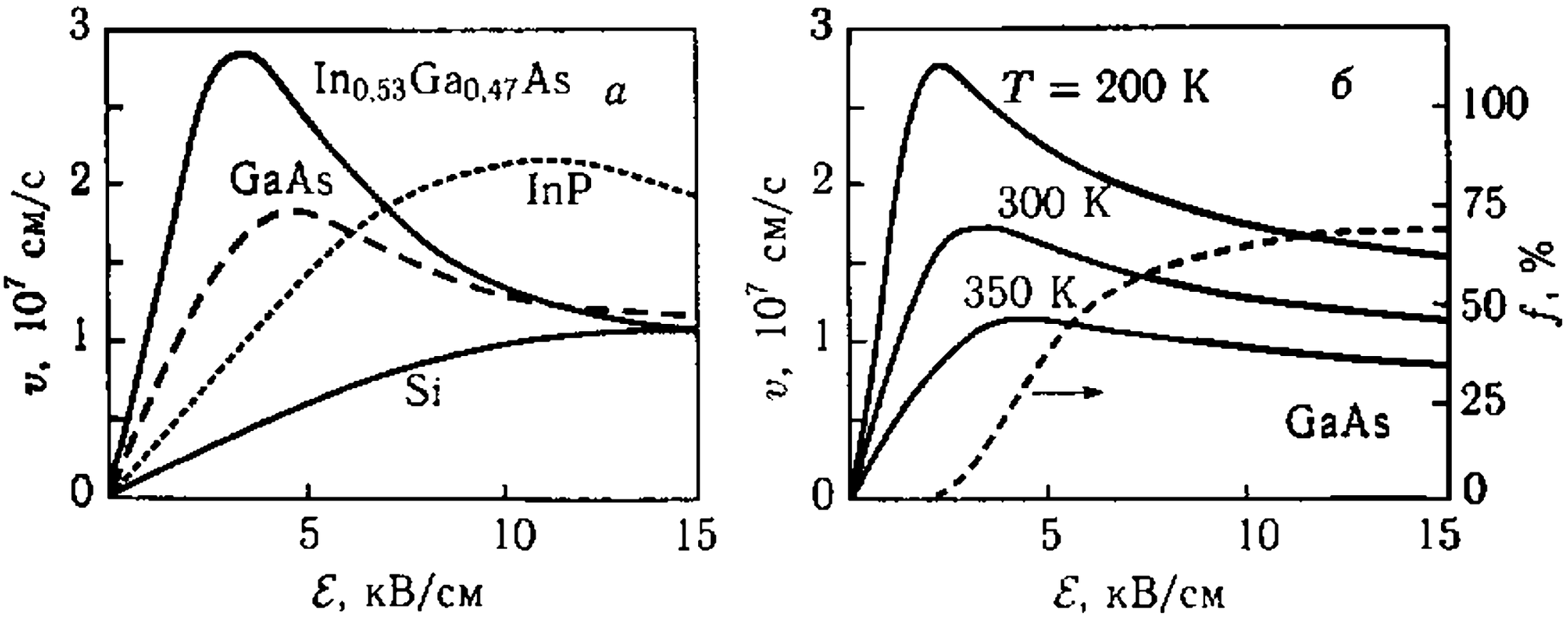


Рис. 6.2. Зависимость скорости дрейфа электронов v от напряженности электрического поля в некоторых полупроводниках [141] ($а$), расчет средней скорости дрейфа электронов v и заселенности побочного минимума зоны проводимости f от напряженности электрического поля в GaAs [14] ($б$)

Рассчитанные по этим формулам зависимости средней скорости дрейфа $v(\mathcal{E})$ и относительной заселенности побочного минимума $f = n_2/(n_1 + n_2)$ представлены пунктирной линией на рис. 6.2$б$. Поскольку f достаточно быстро возрастает с увеличением $\mathcal{E}$, то при некотором значении напряженности

электрического поля $\mathcal{E}_T$, называемом *пороговым полем*, ток в образце $J = qn_0v(\mathcal{E})$ достигает максимума и при более высоких значениях $\mathcal{E}$ полупроводник переходит в состояние с отрицательным дифференциальным сопротивлением ($dJ/d\mathcal{E} < 0$). Расчетная зависимость $v(\mathcal{E})$ для GaAs качественно согласуется с зависимостью, полученной из эксперимента (см. рис. 6.2 *а*). Пороговое поле равно 3,2 кВ/см в GaAs и 10,5 кВ/см в InP.

Анализ полученных выше уравнений показывает, что для появления отрицательного дифференциального сопротивления в полупроводнике необходимо, чтобы:

1) кроме основного минимума с высокой подвижностью носителей и небольшой плотностью состояний существовал побочный минимум с низкой подвижностью носителей и большой плотностью состояний, расположенный так, чтобы его заполнение носителями при комнатной температуре было несущественным, и

2) расстояние между минимумами ΔE было меньше E_g, чтобы междолинный перенос возникал раньше лавинного пробоя [1].

Представленная выше математическая модель эффекта Ганна носит чисто качественный характер. В ней не учитывается ряд существенных моментов, связанных с кинетикой внутридолинного и междолинного рассеяния электронов и наличием двух близко расположенных побочных минимумов (в точках L и X), из-за чего введение представления об электронной температуре становится невозможным [2]. Строгий подход к описанию явлений переноса в сильном электрическом поле требует решения кинетического уравнения Больцмана. Однако из-за трудности обоснования применимости различных аппроксимаций для функции распределения в этом подходе, наиболее широкое распространение в теории эффекта Ганна получил подход, основанный на компьютерном моделировании явлений переноса методом Монте-Карло [69, 218], который эквивалентен строгому решению уравнения Больцмана.

---

[1]) Теоретические расчеты показывают, что на самом деле для возникновения отрицательного дифференциального сопротивления наличие побочных минимумов не является обязательным; этот эффект может возникать в полупроводниках с *одной непараболической* зоной при преобладающем рассеянии на полярных оптических фононах [224]. Этот механизм может стать главным в полупроводниках с большим энергетическим зазором ΔE.

[2]) Для введения электронной температуры необходимо, чтобы темп электрон-электронных столкновений превышал темп рассеяния по другим механизмам. К сожалению, концентрация электронов в диодах Ганна недостаточна велика, чтобы это условие выполнялось.

Динамика развития флуктуаций в среде с отрицательным дифференциальным сопротивлением. Рассмотрим развитие флуктуаций в однородном полупроводнике n-типа проводимости, состояние которого характеризуется отрицательным дифференциальным сопротивлением. Распределение электрического поля в такой среде описывается уравнением Пуассона,

$$\nabla \cdot \mathcal{E} = \frac{4\pi q}{\varepsilon}(N_d - n) = -\frac{4\pi q}{\varepsilon}(n - n_0), \tag{6.7}$$

а уравнение для полной плотности тока должно включать помимо традиционных дрейфового и диффузионного токов (уравнение (1.2)) еще и слагаемое, описывающее *ток смещения*:

$$J = qnv(\mathcal{E}) + qD_n \nabla n + \frac{\varepsilon}{4\pi}\frac{d\mathcal{E}}{dt}. \tag{6.8}$$

В этом уравнении мы для простоты будем предполагать, что величина D_n не зависит от $\mathcal{E}$ [1]).

Для анализа динамики развития малых флуктуаций в пространстве и времени подставим в уравнения (6.7) и (6.8) решение в виде

$$\begin{gathered} n = n_0 + \delta n\, e^{i(\omega t - kx)}, \quad \mathcal{E} = \mathcal{E}_0 + \delta\mathcal{E}\, e^{i(\omega t - kx)}, \\ J = J_0 + \delta J\, e^{i(\omega t - kx)}. \end{gathered} \tag{6.9}$$

После несложных вычислений мы приходим к системе уравнений, связывающих δJ, $\delta\mathcal{E}$ и δn. Эти уравнения допускают существование нетривиального решения ($\delta\mathcal{E} \neq 0$, $\delta n \neq 0$ при $\delta J = 0$) только при условии, когда

$$\omega = -v(\mathcal{E}_0)k + i\left[\frac{4\pi q \mu_d n_0}{\varepsilon} + D_n k^2\right], \tag{6.10}$$

где $\mu_d = dv(\mathcal{E})/d\mathcal{E}$.

Уравнение (6.10) называется *дисперсионным уравнением*. Оно описывает эволюцию малой флуктуации с частотой ω и волновым вектором k. Действительная часть $\mathrm{Re}\,\omega$ описывает движение волны в пространстве со скоростью $\omega/k = -v(\mathcal{E}_0)$, а мнимая часть $\mathrm{Im}\,\omega$ — скорость либо затухания (при $\mathrm{Im}\,\omega > 0$), либо нарастания (при $\mathrm{Im}\,\omega < 0$) волны флуктуаций.

Рассмотрим мнимую часть $\mathrm{Im}\,\omega$ более подробно. Как следует из формулы (6.10), она состоит из двух слагаемых. Первое

---

[1]) Учет зависимости $D_n(\mathcal{E})$ делает бóльшую часть расчетов аналитически невыполнимыми (см. подробнее [218]).

слагаемое содержит величину μ_d, называемую *дифференциальной подвижностью*. В области отрицательного дифференциального сопротивления эта величина становится отрицательной; ее отрицательное максимальное значение достигает $\mu_d \approx$ ≈ -2400 см^2/В · с в GaAs и ≈ -2000 см^2/В · с в InP [14]. Величина, обратная первому слагаемому в квадратных скобках ($\tau_{Md} = \varepsilon/4\pi q\mu_d n_0$), носит название *дифференциального времени максвелловской (диэлектрической) релаксации*; она также становится отрицательной в области отрицательного дифференциального сопротивления. Второе слагаемое в формуле описывает затухание волны флуктуаций из-за диффузионного размытия и является всегда положительным [1]. При малых k (большие длины волн) это слагаемое оказывается малым, и эволюция флуктуации во времени определяется знаком первого слагаемого. Из этого следует, что среда с отрицательным дифференциальным сопротивлением характеризуется *флуктуационной неустойчивостью*: любые спонтанно возникшие в ней длинноволновые флуктуации плотности заряда и электрического поля всегда будут нарастать.

В образце конечных размеров величина k связана с длиной образца L уравнением $k \approx 2\pi m/L$, где $m \geqslant 1$ — номер возбуждаемой «гармоники». Чтобы флуктуации в таком образце нарастали, необходимо, чтобы дифференциальное время максвелловской релаксации $|\tau_{Md}|$ было короче характерного времени диффузии, то есть выражение в квадратных скобках уравнения (6.10) по крайней мере для $m = 1$ должно оставаться отрицательным. Это условие можно переписать следующим образом:

$$n_0 L^2 > \frac{\pi\varepsilon D_n}{q|\mu_d|}. \qquad (6.11)$$

Характерное значение выражения в правой части этого неравенства для GaAs равно $\sim 5 \cdot 10^7$ см$^{-1}$. Таким образом, мы видим, что в образце конечных размеров концентрация электронов, при которой в образце может наблюдаться усиление волн флуктуаций, ограничена снизу процессами диффузии.

Эволюция волны флуктуаций в образце конечных размеров зависит также и от того, успевает ли волна существенно

---

[1]) Следует заметить, что в условиях сильного отклонения от равновесия мы не можем больше пользоваться соотношением Эйнштейна для вычисления D_n. Эту величину можно найти моделируя явления переноса методом Монте-Карло [69, 218].

усилиться при прохождении через образец, или нет. Нарастание флуктуаций будет значительным, если время пролета электронов через образец L/v (здесь v — средняя скорость дрейфа) превышает характерное время нарастания флуктуаций $|\tau_{Md}|$. Это условие, называемое *критерием Крёмера*, можно переписать следующим образом:

$$n_0 L > \frac{\varepsilon v}{4\pi q |\mu_d|}. \tag{6.12}$$

Для GaAs характерное значение выражения в правой части (6.12) равно $\sim 10^{11}$ см$^{-2}$. При $n_0 L > 10^{12}$ см$^{-2}$ амплитуда флуктуации, зародившейся в образце, нарастает настолько быстро, что развитие флуктуаций перестает описываться рассмотренной выше линейной теорией, и в образце формируется движущийся *домен сильного поля*. Характеристики этого домена мы найдем в следующем разделе. О характере неустойчивости в таких образцах часто говорят как о *доменной неустойчивости* [1]).

Домены сильного поля. В образцах, удовлетворяющих критерию Крёмера (6.12), нарастание флуктуаций при средней напряженности электрического поля, превышающей пороговое поле $\mathcal{E}_T$, происходит настолько быстро, что в них за малую долю времени пролета успевает сформироваться *домен сильного поля*, который движется от катода к аноду с постоянной скоростью. Возникновение таких доменов было экспериментально установлено Ганном в 1964 г. с помощью зондовых измерений [221].

При образовании домена распределение электрического поля в образце становится сильно неоднородным (см. рис. 6.3 *а*). В образце сосуществуют области, в которых напряженность электрического поля $\mathcal{E} = \mathcal{E}_r < \mathcal{E}_T$, и область (домен), в которой напряженность поля $\mathcal{E}_m$ намного (иногда в десятки раз) превышает $\mathcal{E}_T$ [2]). Распределения электрического поля и плотности объемного

---

[1]) Явление доменной неустойчивости в однородных полупроводниках может быть связано не только с междолинным переносом электронов, но также с механизмами акустоэлектрической и рекомбинационной неустойчивости [10, 225]. Скорость движения акустоэлектрических и рекомбинационных доменов (10^{-3}–10^5 см/с) существенно меньше, чем ганновских доменов, и поэтому эти типы неустойчивости не представляют большого интереса для высокочастотной электроники.

[2]) При высокой напряженности электрического поля в домене (в GaAs величина $\mathcal{E}_m$ может достигать 200 кВ/см) в полупроводнике начинается ударная ионизация. Появляющиеся при этом дырки могут оказывать влияние на движение доменов. Одним из негативных последствий накопления дырок,

заряда в образце способствуют стабилизации домена. Действительно, более быстрое движение электронов на границах домена (там, где напряженность поля ниже $\mathcal{E}_T$, см. рис. 6.3 *а*) вызывает появление обогащенного слоя с одной стороны домена и обедненного слоя с другой его стороны, а такое перераспределение плотности заряда как раз и поддерживает сильное электрическое поле в домене.

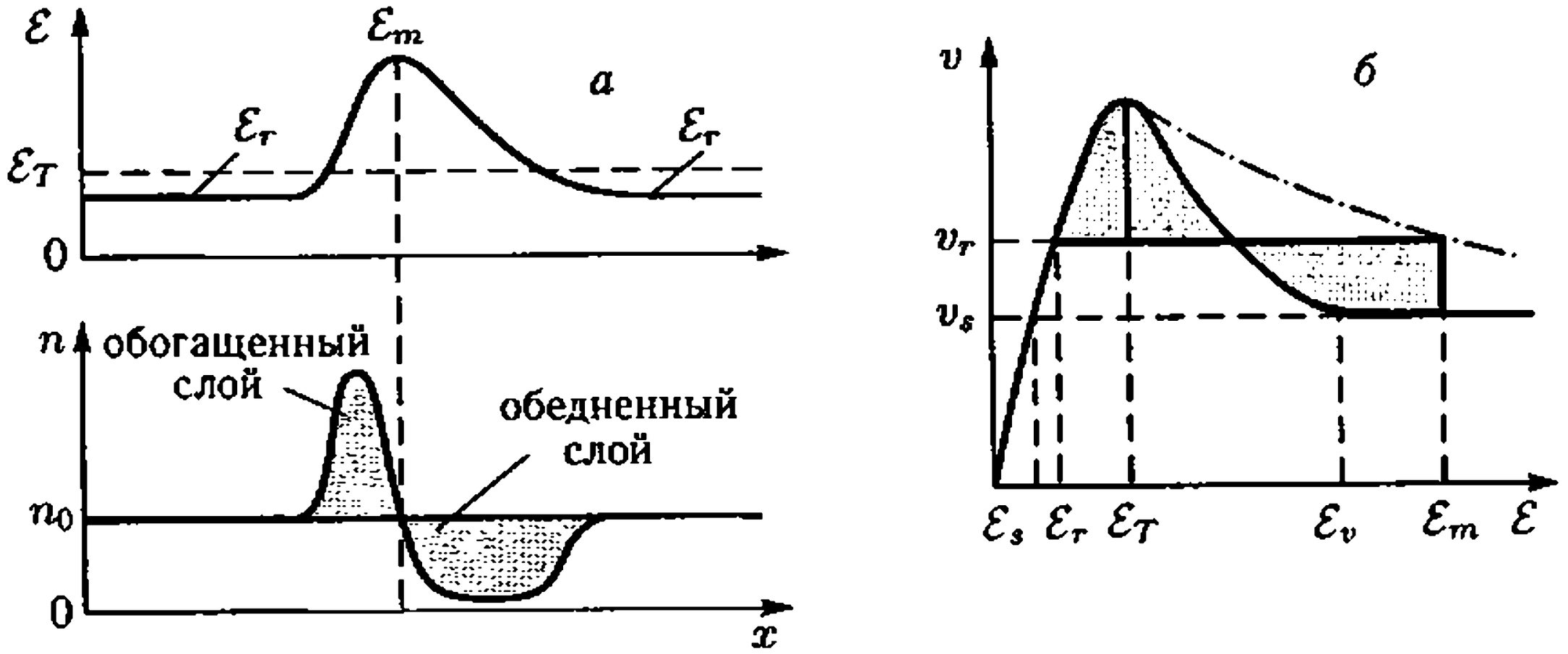


Рис. 6.3. Пространственное распределение электрического поля и концентрации электронов в движущемся домене сильного поля (*а*); использование «правила равных площадей» для нахождения связи $\mathcal{E}_r, \mathcal{E}_m$ и v_r (заштрихованные области имеют одинаковую площадь) (*б*). Штрих-пунктирная линия на рис. *б* — зависимость v_r от $\mathcal{E}_m$

Распределения поля и заряда в домене могут быть найдены путем совместного решения уравнения Пуассона и уравнения непрерывности, однако из-за сильной нелинейности задачи это обычно делается численными методами [218]. Ниже мы попытаемся ограничиться минимальным числом формул (без их вывода) и сосредоточимся больше на обсуждении самих явлений.

Одной из немногих задач, допускающих аналитическое решение, является задача о движении сформировавшегося (стабильного) домена. В предположении, что коэффициент диффузии D_n в уравнении (6.8) не зависит от $\mathcal{E}$, совместное решение уравнений (6.7) и (6.8) в движущейся вместе с доменом системе

которое связано с их невысокой подвижностью и возникает после прохождения нескольких доменов через образец, может быть нарушение когерентности колебаний в диодах Ганна. Поскольку напряженность поля в доменах быстро возрастает с увеличением n_0, то в образцах GaAs с $n_0 L > 10^{14}$ см$^{-2}$ интенсивное накопление дырок начинается уже на пороге возникновения эффекта Ганна.

координат позволяет придти к следующему уравнению, связывающему напряженность поля вне домена $\mathcal{E}_r$ с максимальным полем в домене $\mathcal{E}_m$ (так называемое «правило равных площадей»):

$$\int_{\mathcal{E}_r}^{\mathcal{E}_m} [v(\mathcal{E}) - v(\mathcal{E}_r)]\, d\mathcal{E} = 0. \qquad (6.13)$$

Для известной зависимости $v(\mathcal{E})$ и заданного значения $\mathcal{E}_m$ графическое решение этого уравнения (см. рис. 6.3*б*) позволяет рассчитать напряженность поля вне домена $\mathcal{E}_r$ и определить скорость v_r движения домена как единого целого. Считая, как показано на рис. 6.3*б*, что зависимость $v(\mathcal{E})$ выходит на насыщение при $\mathcal{E} > \mathcal{E}_v$, в практически важном случае большой амплитуды поля в домене [1]) из формулы (6.13) нетрудно получить приближенную формулу

$$\mathcal{E}_m(\mathcal{E}_r - \mathcal{E}_s) \approx \mathcal{E}_c^2,$$

где

$$\mathcal{E}_c^2 = (1/\mu_1) \int_{\mathcal{E}_s}^{\mathcal{E}_v} [v(\mathcal{E}) - v_s]\, d\mathcal{E}, \quad \text{а} \quad \mathcal{E}_s = v_s/\mu_1.$$

Используя характерные для GaAs значения $\mathcal{E}_c \approx 3$ кВ/см, $\mathcal{E}_s \approx 1{,}5$ кВ/см [218], для $\mathcal{E}_m = 60$ кВ/см получаем $\mathcal{E}_r \approx$ $\approx 1{,}65$ кВ/см, откуда следует, что домены в диодах Ганна движутся со скоростью, лишь немного превышающей скорость насыщения v_s.

Решение задачи о движении стабильного домена позволяет найти распределение плотности заряда в обогащенном и обедненном слоях и рассчитать вольт-амперную характеристику образца с доменом. Для этого уравнения (6.7), (6.8) и (6.13) необходимо дополнить условием на величину полного падения напряжения на образце:

$$U = \int_0^L \mathcal{E}\, dx = \mathcal{E}_r L + \frac{\varepsilon}{4\pi} \int_{\mathcal{E}_r}^{\mathcal{E}_m} \left(\frac{1}{|\rho_a|} + \frac{1}{\rho_d} \right) (\mathcal{E} - \mathcal{E}_r)\, d\mathcal{E}, \qquad (6.14)$$

где L — длина образца, а ρ_a и ρ_d — плотность объемного заряда в обогащенном и обедненном слоях, соответственно.

---

[1]) Характерные значения $\mathcal{E}_m$ в диодах Ганна из GaAs составляют 40–200 кВ/см и намного превышают $\mathcal{E}_v$.

При вычислении падения напряжения на домене U_d (эта величина равна интегралу в правой части уравнения (6.14)) мы перешли от интегрирования по координате к интегрированию по напряженности поля, использовав уравнение Пуассона $d\mathcal{E}/dx = 4\pi\rho/\varepsilon$. Можно показать [69], что при $\mathcal{E}_m \gg \mathcal{E}_т$ этот интеграл может быть представлен в виде

$$U_d \approx \varepsilon \mathcal{E}_m^2 F/(4\pi q n_0),$$

где множитель F, называемый параметром формы домена, является универсальной функцией безразмерной концентрации электронов $n_0/n_{кр}$. «Критическая концентрация» электронов,

$$n_{кр} = \varepsilon\mu_1 \mathcal{E}_c^2/(4\pi q D_n),$$

в GaAs составляет $\simeq 3 \cdot 10^{15}$ см$^{-3}$. При $n_0 \ll n_{кр}$ значения $\rho_d \approx \approx qn_0$, $|\rho_a| \gg qn_0$, $F \approx 1/2$, и распределение электрического поля в домене имеет приблизительно треугольную форму. При $n_0 \gg \gg n_{кр}$ плотности заряда ρ_d, $|\rho_a| \ll qn_0$,

$$F \approx \sqrt{32/3}\sqrt{n_0/n_{кр}},$$

и распределение поля в домене имеет форму, близкую к перевернутой параболе. Чем выше концентрация n_0 при фиксированном значении U_d, тем меньше ширина домена и тем выше напряженность электрического поля в нем.

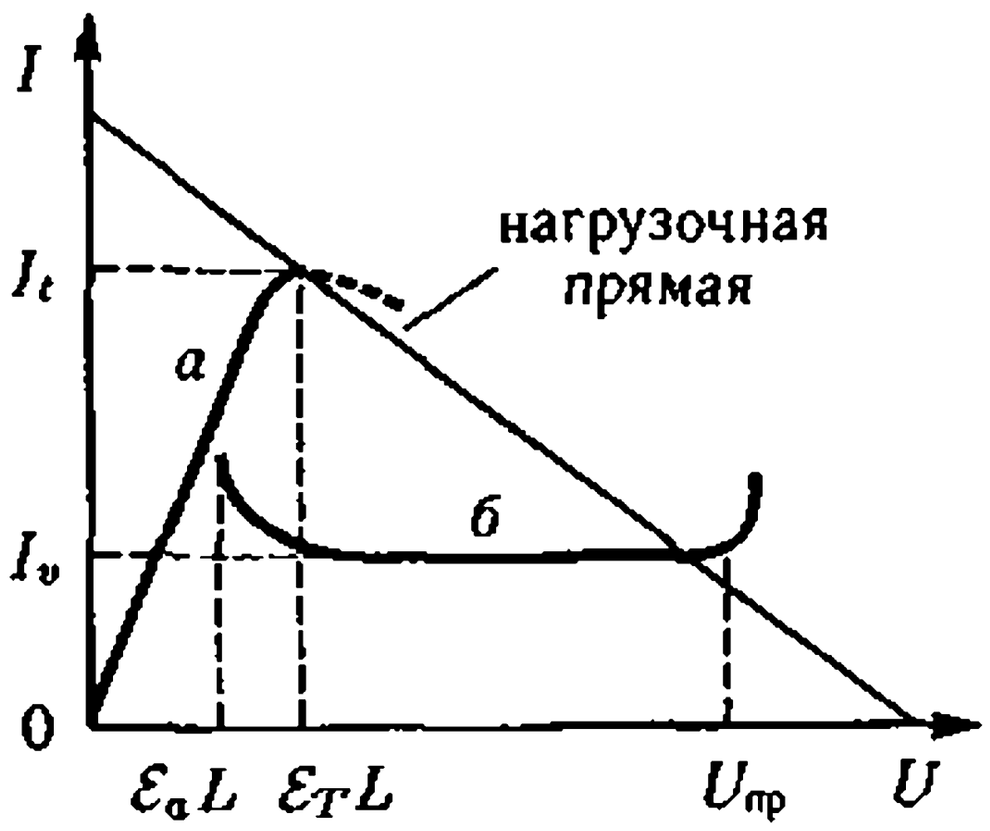


Рис. 6.4. Вольт-амперная характеристика диода Ганна без домена (*а*) и с доменом (*б*)

Уже первые исследования эффекта Ганна обнаружили, что неоднородности в образцах играют роль центров, на которых зарождаются домены. Контакты к образцу также могут выступать в качестве таких центров. В частности, контакты могут приводить к образованию *неподвижных доменов*: 1) так называемой области прикатодного падения потенциала и 2) стационарному анодному домену [69]. В то время как при увеличении приложенного к диоду напряжения от области прикатодного падения потенциала могут отделяться обычные движущиеся домены и ток в образце будет продолжать осциллировать, ситуация со стационарным анодным доменом сложнее: при появлении такого домена напряженность элек-

трического поля в объеме образца становится ниже $\mathcal{E}_T$ и движущиеся домены перестают возбуждаться (возбуждение колебаний в образцах с анодным доменом возможно лишь в высокодобротном резонаторе [69]). Расчеты показывают, что образование стационарного анодного домена возможно только при $n_0 \gtrsim n_{кр}$, и поэтому величину $n_{кр}$ обычно принимают за верхнюю границу концентрации электронов в образцах, из которых изготавливают диоды Ганна.

Совместное решение уравнений (6.13) и (6.14) позволяет рассчитать вольт-амперную характеристику образца с доменом. Вольт-амперные характеристики образца без домена и с доменом показаны на рис. 6.4. Видно, что возникновение домена приводит к появлению *гистерезиса* на вольт-амперной характеристике. После того, как домен сформировался, бóльшая часть приложенного к образцу напряжения падает на домене, а вне домена электрическое поле равно $\mathcal{E}_r < \mathcal{E}_T$. Поэтому если приложенное к образцу с доменом напряжение сделать немного ниже $\mathcal{E}_T L$, то домен в образце не исчезнет, просто его толщина немного уменьшится. Чтобы домен исчез, напряжение на образце необходимо понизить до $\mathcal{E}_a L$, где $\mathcal{E}_a$ — *пороговое поле исчезновения домена*. Расчеты (см. [218]) показывают, что $\mathcal{E}_a$ всегда меньше $\mathcal{E}_T$, причем при увеличении n_0 поле $\mathcal{E}_a \to \mathcal{E}_s = v_s/\mu_1$. При напряжении на образце $U > U_{пр}$ (см. рис. 6.4) из-за сильного возрастания напряженности поля в домене начинается процесс накопления дырок в образце (см. подстрочное замечание на с. 341) и на вольт-амперной характеристике появляются признаки «пробоя».

Пределы быстродействия диодов Ганна. До сих пор, рассматривая явление флуктуационной неустойчивости в среде с отрицательным дифференциальным сопротивлением, мы ничего не говорили о динамике физических процессов, лежащих в основе этого явления. Понимание этих процессов позволяет установить пределы быстродействия диодов Ганна.

Быстродействие диодов Ганна определяется тем, насколько быстро распределение электронов по энергиям и между минимумами (и, следовательно, средняя скорость дрейфа электронов) отслеживает изменение напряженности электрического поля. При изменении напряженности поля электроны в минимумах начинают разогреваться (или остывать) и путем междолинного рассеяния перераспределяться между минимумами. Поэтому быстродействие диодов Ганна определяется характеристиками этих процессов:

1) временами релаксации энергии в минимумах и

2) характерными временами междолинного рассеяния (из-за разной плотности состояний в минимумах времена прямого и обратного междолинного рассеяния могут заметно различаться).

Моделирование методом Монте-Карло показало, что среди указанных процессов наиболее медленным является процесс релаксации энергии в основном минимуме: его характерное время в GaAs и InP составляет 2–5 пс. Времена междолинного рассеяния $\Gamma \to L$, $\Gamma \to X$ составляют ~0,1 пс [218]. Это дает оценку максимальной частоты, на которой отрицательное дифференциальное сопротивление в диодах Ганна из GaAs и InP еще сохраняется, равную ~150 ГГц, что близко к максимально достигнутым частотам генерации 160 ГГц в GaAs и 170 ГГц в InP.

В настоящее время большое внимание исследователей эффекта Ганна обращено к GaN. Теоретические расчеты предсказывают для обеих кристаллических модификаций нитрида галлия (со структурой вюрцита и сфалерита) возможность появления отрицательного дифференциального сопротивления с наиболее высокой максимальной частотой генерации — ~700 ГГц [223]. Расчеты зонной структуры GaN не дают пока ясного ответа, связана ли доменная неустойчивость с существованием побочного минимума зоны проводимости, или же она связана с сильной непараболичностью основного минимума (см. подстрочное замечание на с. 338). Высокая максимальная частота генерации является следствием малого времени релаксации энергии в основном минимуме (0,15 пс) и небольшого времени междолинного рассеяния (1,2 пс). Если же окажется, что механизм неустойчивости связан с процессами, развивающимися только в основном минимуме, то быстродействие диодов может оказаться еще выше (до 4 ТГц). По сравнению с GaAs и InP нитрид галлия имеет более высокую напряженность поля лавинного пробоя и более высокую максимальную скорость дрейфа, что позволяет ожидать от диодов Ганна из GaN выходную мощность примерно на два порядка выше, чем от диодов из GaAs и InP.

Режимы генерации СВЧ колебаний в диодах Ганна.

Режим пролета обогащенного слоя. Диоды Ганна, для которых выполняется условие усиления флуктуаций (6.11), но не выполняется критерий Крёмера (6.12), принято называть *субкритически* легированными, а диоды, для которых критерий Крёмера выполняется — *суперкритически* легированными. Усиление и генерация СВЧ колебаний в субкритически

легированных образцах возможны в так называемом режиме пролета обогащенного слоя.

Пусть в какой-то момент времени к субкритически легированному образцу прикладывается напряжение, создающее в нем однородное электрическое поле с напряженностью $\mathcal{E} > \mathcal{E}_T$. В этот момент начинает течь ток и из катода в образец инжектируются избыточные электроны (так называемый обогащенный слой). Дальнейшая эволюция распределения электрического поля в образце происходит примерно так, как показано на рис. 6.5: флуктуация плотности объемного заряда в виде обогащенного слоя движется в направлении от катода к аноду, непрерывно возрастая по величине. Физической причиной увеличения амплитуды флуктуации является то, что в области с $\mathcal{E} > \mathcal{E}_T$ электроны движутся *медленнее*, чем в области с $\mathcal{E} < \mathcal{E}_T$, и заряд в обогащенном слое постоянно нарастает за счет электронов из прикатодной области, «догоняющих» флуктуацию заряда.

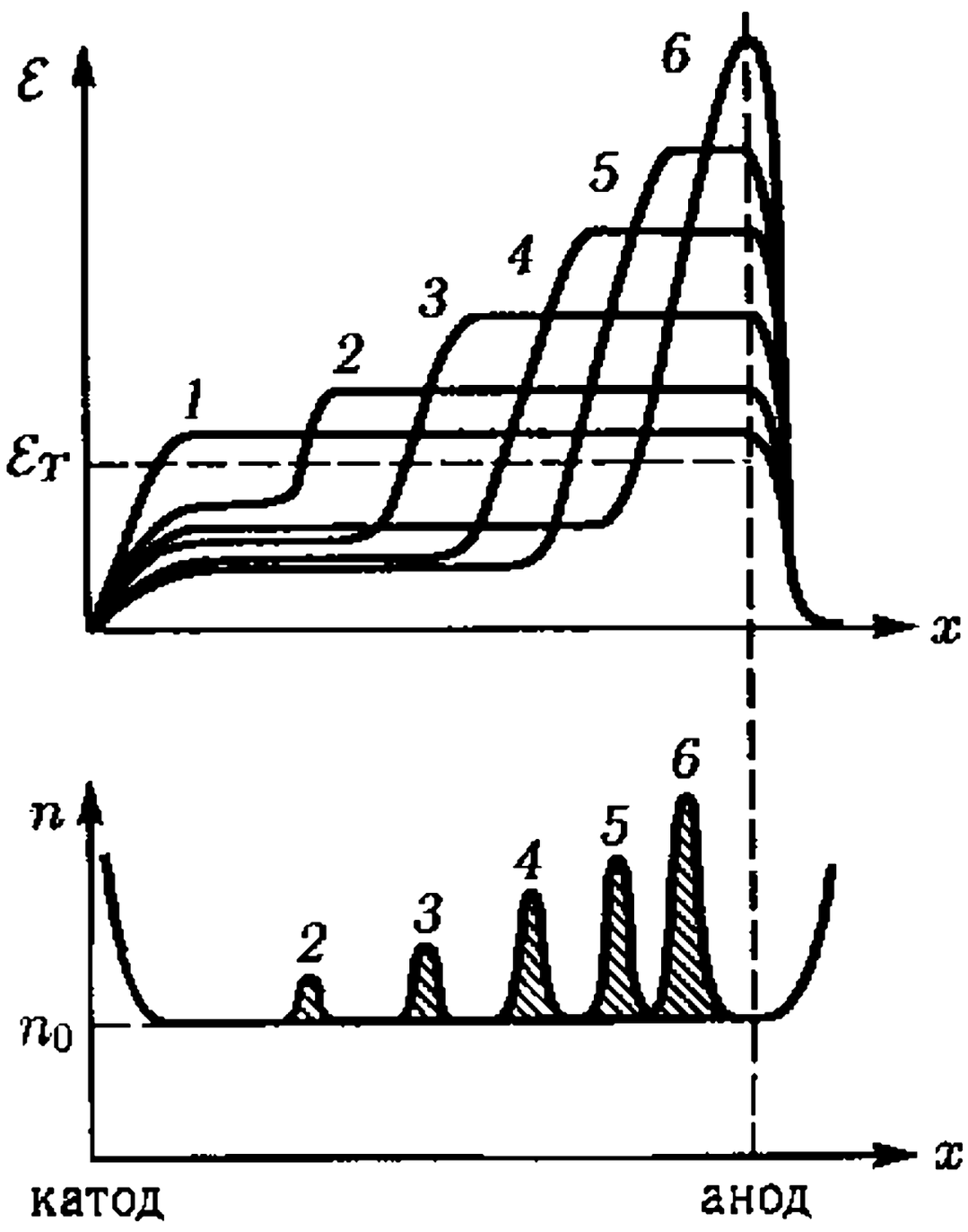


Рис. 6.5. Динамика распространения флуктуации заряда в диоде Ганна в режиме пролета обогащенного слоя. Цифры у кривых отвечают последовательным моментам времени [14]

Когда обогащенный слой уходит в анод, в образце, казалось бы, должны создаваться условия для инжекции следующего обогащенного слоя. На самом деле из-за диффузионного размытия обогащенного слоя к тому моменту, когда он достигает анода, заряд оказывается сильно «размазанным» и в образце устанавливается неоднородное распределение электрического поля и плотности заряда, а ток через образец при фиксированном напряжении становится стабильным. Эту стабильность однако легко нарушить, изменив напряжение на образце. То есть, если субкритически легированный образец поместить в высокодобротный резонатор, то в нем возможно возбуждение колебаний. Соответствующий режим генерации называют режимом с пролетом обогащенного слоя. К.п.д. генератора в этом режиме довольно мал (~5 %), а частота

генерации определяется длиной образца и скоростью пролета обогащенного слоя ($f \approx v_{об}/L$). Согласно расчетам, скорость пролета обогащенного слоя заметно превышает скорость движения доменов сильного поля и составляет $v_{об} = (1,5\text{–}4) \cdot 10^7$ см/с [69].

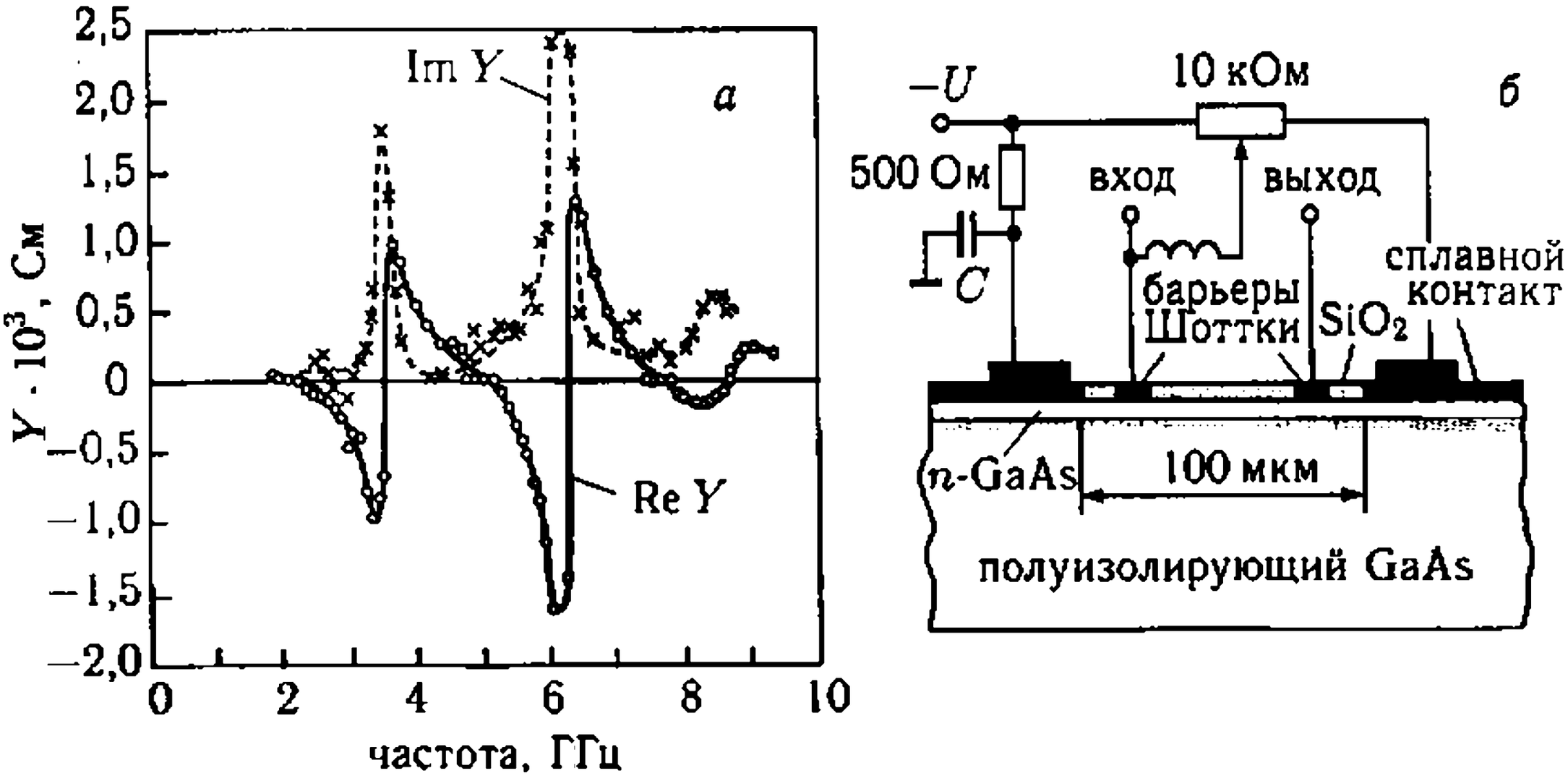


Рис. 6.6. Зависимость действительной и мнимой частей комплексной проводимости Y образца GaAs ($n_0 = 3 \cdot 10^{13}$ см$^{-3}$, $L = 70$ мкм) от частоты при $\mathcal{E} = 4,8$ кВ/см [14] (*а*); одна из конструкций усилителя типа бегущей волны на основе диода Ганна [69] (*б*)

Поскольку действительная часть полной проводимости диода Ганна в режиме пролета обогащенного слоя оказывается отрицательной на частотах, кратных гармоникам «пролетной частоты» (см. рис. 6.6 *а*), на этих частотах диод может быть использован в качестве усилителя. Усилители СВЧ колебаний могут быть построены либо по схеме усилителя бегущей волны, либо по схеме усилителя отражательного типа [69]. Пример конструкции усилителя первого типа показан на рис. 6.6 *б*. Возбуждение волны объемного заряда входным сигналом осуществляется с помощью барьера Шоттки. По мере распространения по образцу эта волна нарастает и поэтому сигнал, снимаемый со второго барьера Шоттки, оказывается усиленным. Коэффициент усиления по мощности усилителей подобного типа может достигать 10000 (40 дБ). В качестве усилителей могут быть использованы не только субкритически легированные образцы, но и суперкритически легированные диоды с анодным доменом или неподвижным доменом в области катода, а также диоды с движущимся доменом, в которых возможно параметрическое усиление сигнала [218].

К сожалению, усилители на основе эффекта междолинного переноса электронов характеризуются довольно высоким уровнем шума, связанным с флуктуациями скорости движения электронов (кроме того, в режимах, использующих формирование доменов, дополнительным источником шума становятся флуктуации времени зарождения домена). Наименьший достигнутый коэффициент шума усилителей

на диодах Ганна составляет 8–10 дБ на частоте 10 ГГц. Это существенно выше коэффициента шума современных полевых транзисторов из GaAs на той же частоте (1–2 дБ). Шумы диодов Ганна проявляются и при работе генераторов в виде паразитной частотной и (в меньшей степени) амплитудной модуляции выходного сигнала.

Режим пролета домена. Как мы уже говорили выше, в диодах Ганна, удовлетворяющих критерию Крёмера (6.12), нарастание флуктуаций происходит очень быстро. Поэтому уже в прикатодной области образца успевает сформироваться домен сильного поля, который далее движется от катода к аноду с постоянной скоростью, немного превышающей скорость насыщения. Генерация колебаний в таком образце происходит в режиме пролета домена.

Возбуждение колебаний тока в режиме пролета домена может быть понято на основе вольт-амперных характеристик, показанных на рис. 6.4. В начальный момент времени, когда на образец подается импульс напряжения и в образце создается однородное электрическое поле с напряженностью, немного превышающей $\mathcal{E}_T$, мы оказываемся на неустойчивом участке ветви a вольт-амперной характеристики. При этом через образец начинает течь ток, величина которого близка к I_t. Однако в течение короткого времени в области катода формируется домен сильного поля и мы переходим на ветвь $б$ этой характеристики, а ток понижается до I_v. Чтобы домен зарождался на катоде, а не в произвольной точке образца, необходимо, чтобы образец был достаточно однородным, а вблизи катода существовала область, в которой напряженность электрического поля была бы несколько выше, чем в остальной части образца. Сформировавшийся вблизи катода домен движется в образце, пока не достигнет анодного контакта, где он исчезает. В этот момент мы вновь возвращаемся на ветвь a вольт-амперной характеристики, и в электрической цепи появляется всплеск тока через образец (см. рис. 6.7). При этом средняя напряженность электрического поля в образце также возрастает и возникают условия для зарождения нового домена. Частота генерации в режиме пролета домена (*«пролетная частота»*) приблизительно обратно

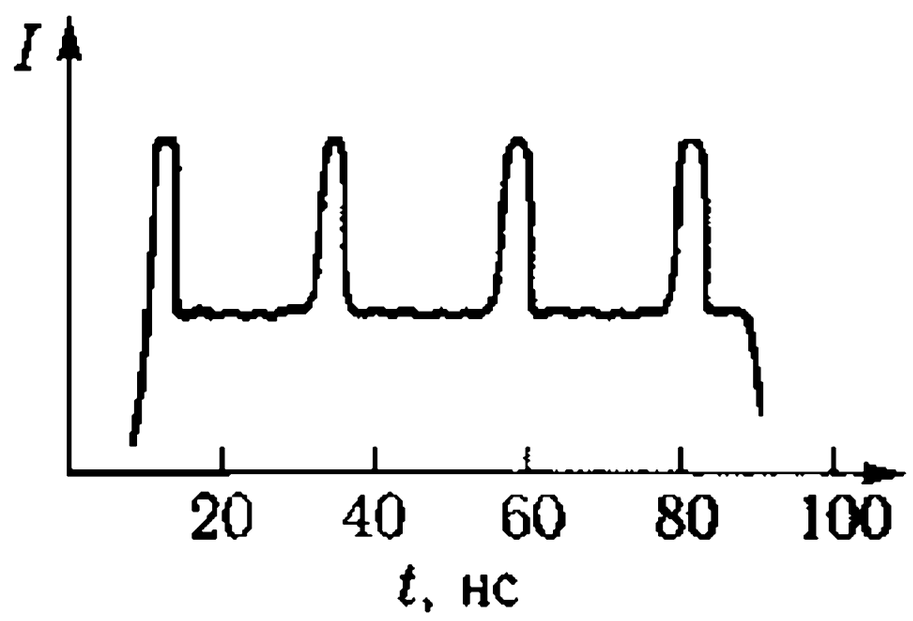


Рис. 6.7. Колебания тока в цепи, составленной из образца n-GaAs и омического сопротивления нагрузки [218]. Длина образца — 2 мм

пропорциональна длине образца ($f_{\text{прол}} \approx v_s/L$). Частоту генерации в небольших пределах можно перестраивать, изменяя напряжение на диоде [1]. К.п.д. генератора в режиме пролета домена при работе на омическую нагрузку мал, однако если поместить образец в СВЧ резонатор (см. рис. 6.8, *а*), то его к.п.д. возрастает и может достигать 29% (на частоте 2 ГГц), что близко к теоретическому пределу.

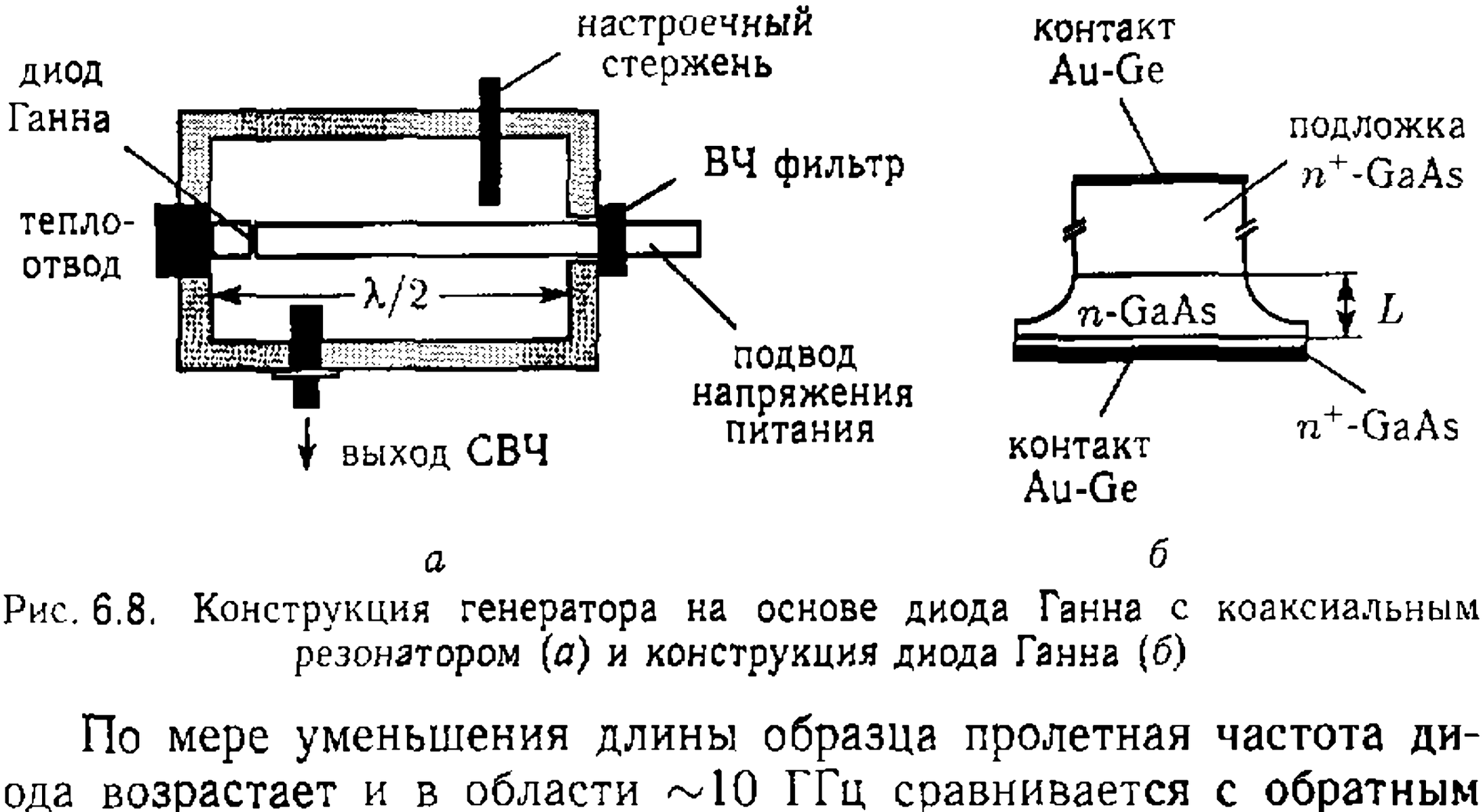


Рис. 6.8. Конструкция генератора на основе диода Ганна с коаксиальным резонатором (*а*) и конструкция диода Ганна (*б*)

По мере уменьшения длины образца пролетная частота диода возрастает и в области ~10 ГГц сравнивается с обратным временем формирования домена τ_f [2]. Режим генерации, когда домен за время пролета не успевает сформироваться полностью, называют *гибридным* режимом. К.п.д. генератора, работающего в гибридном режиме, несколько меньше, чем в режиме пролета домена.

Режимы с гашением и запаздыванием домена. Описанный выше режим пролета домена реализуется в том случае, когда амплитуда переменного напряжения в СВЧ резонаторе намного меньше постоянного напряжения, подаваемого на диод Ганна. Однако если настроить генератор так, чтобы эти напряжения были близки, то можно перевести диод Ганна в режимы,

---

[1]) Как мы говорили выше, с увеличением напряжения на диоде Ганна возрастает напряженность поля в домене $\mathcal{E}_m$. Поэтому, в соответствии с формулой (6.13), напряженность поля вне домена $\mathcal{E}_r$ и скорость движения домена v_r уменьшаются, и частота генерации также уменьшается. Перестраивать частоту генератора также можно изменяя частоту резонатора (механическим способом *или электронным способом с помощью варикапа*) [218].

[2]) Заметим, что время τ_f, вообще говоря, отличается от времени τ_{Md}, поскольку формирование домена не может быть описано просто как развитие малой флуктуации. Динамика формирования домена рассмотрена в [218].

характеризуемые бо́льшей гибкостью в перестройке частоты: режим с гашением (разрушением домена) и режим с запаздыванием формирования домена [1]). Работа диода Ганна в этих режимах использует явление гистерезиса на вольт-амперной характеристике (см. рис. 6.4).

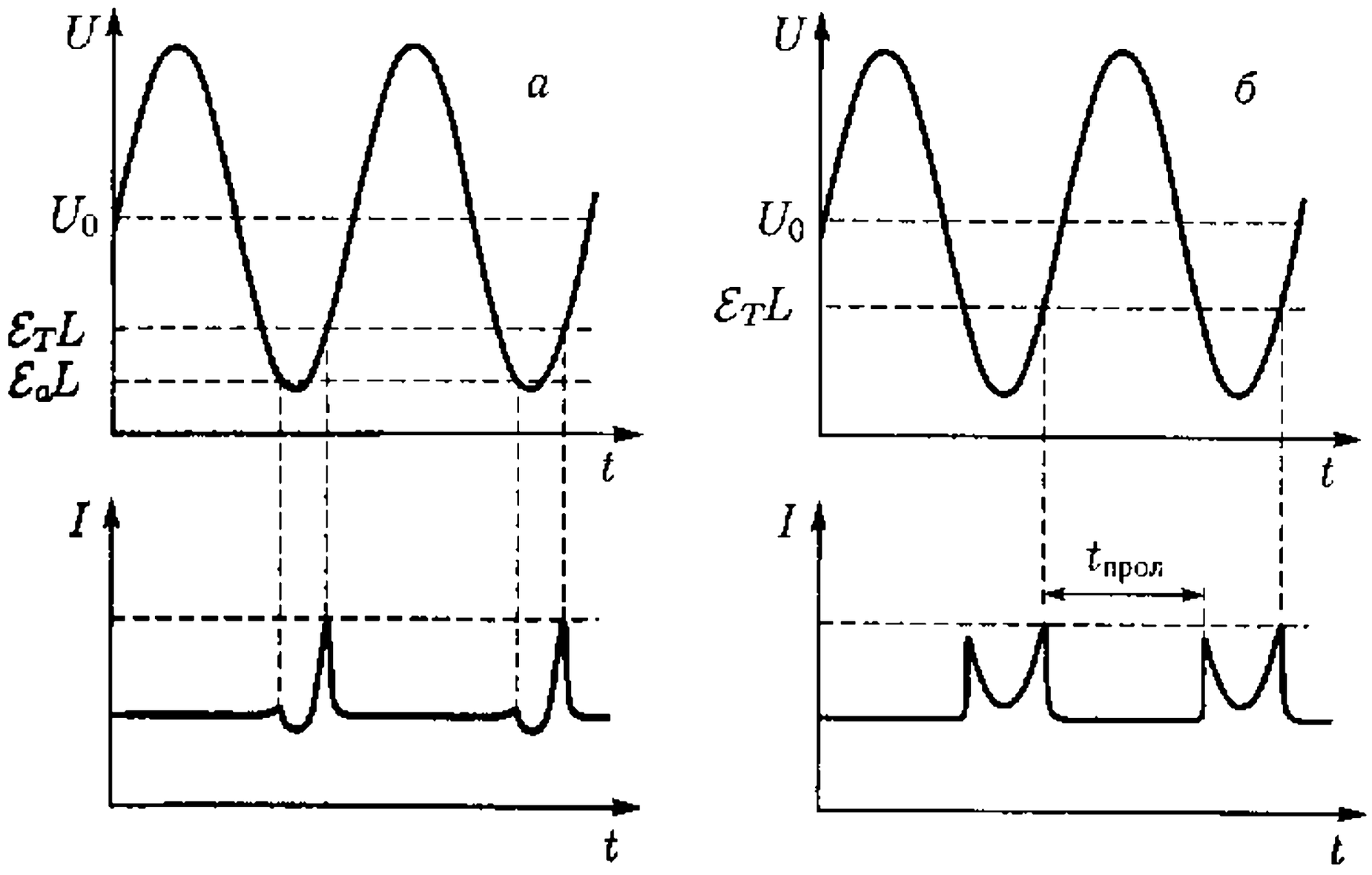


Рис. 6.9. Работа диода Ганна в режимах с гашением домена (*а*) и запаздыванием формирования домена (*б*)

Рассмотрим сначала режим работы с гашением (режим с разрушением домена). В этом режиме зародившийся вблизи катода домен движется в направлении к аноду, однако в момент времени, когда сумма приложенного к образцу постоянного и переменного напряжений оказывается ниже порога исчезновения домена ($\mathcal{E}_a L$ на рис. 6.9*а*), домен быстро рассасывается не достигнув анода. Всплеск тока, возникающий в этот момент и продолжающийся до появления следующего домена, находится в противофазе с переменным напряжением и, следовательно, передает энергию в резонатор. Режим с гашением позволяет существенно изменять частоту генерации и работать на частоте, задаваемой резонатором, которая может быть как немного ниже, так и существенно выше пролетной частоты диода Ганна $f_{\text{прол}}$.

В режиме с запаздыванием формирования домена (см. рис. 6.9*б*) после того, как домен достигнет анода (через $t_{\text{прол}}$ после его образования), требуется еще некоторое время, чтобы суммарное напряжение на диоде достигло $\mathcal{E}_T L$ и мог начать

---

[1]) Иногда, основываясь на близости физических принципов работы диодов Ганна в этих двух режимах и в пролетном режиме, все три режима генерации называют пролетными режимами.

формироваться следующий домен. В этом режиме диод Ганна также работает на частоте, задаваемой резонатором; очевидно, эта частота всегда ниже пролетной частоты диода. То, что в режиме с запаздыванием удается формировать достаточно продолжительные импульсы тока (ср. рис. 6.9*а* и *б*), позволяет в этом режиме добиться максимального к.п.д. преобразования (на частоте $f \approx f_{\text{прол}}/2$).

Описанные режимы работы показывают, что включая один и тот же диод Ганна в различные резонаторы и создавая для диода различные условия работы, можно изменять частоту генерации в широких пределах.

Режим с ограничением накопления объемного заряда. Несмотря на существование падающего участка на зависимости $U(\mathcal{E})$, на статических вольт-амперных характеристиках образца при напряжении $U > \mathcal{E}_T L$ этот участок никогда не наблюдается. Причиной этого является то, что в условиях отрицательного дифференциального сопротивления за короткое время в образце устанавливается неоднородное распределение электрического поля и плотности заряда, которое приводит к появлению области насыщения тока вместо падающего участка на вольт-амперной характеристике. Коуплэнд [226] предложил интересный режим возбуждения колебаний в диоде Ганна, при котором удается избежать возникновения объемного заряда, и *динамическая* вольт-амперная характеристика образца воспроизводит зависимость $U(\mathcal{E})$ с падающим участком. Этот режим генерации получил название режима с ограничением накопления объемного заряда (ОНОЗ) [1].

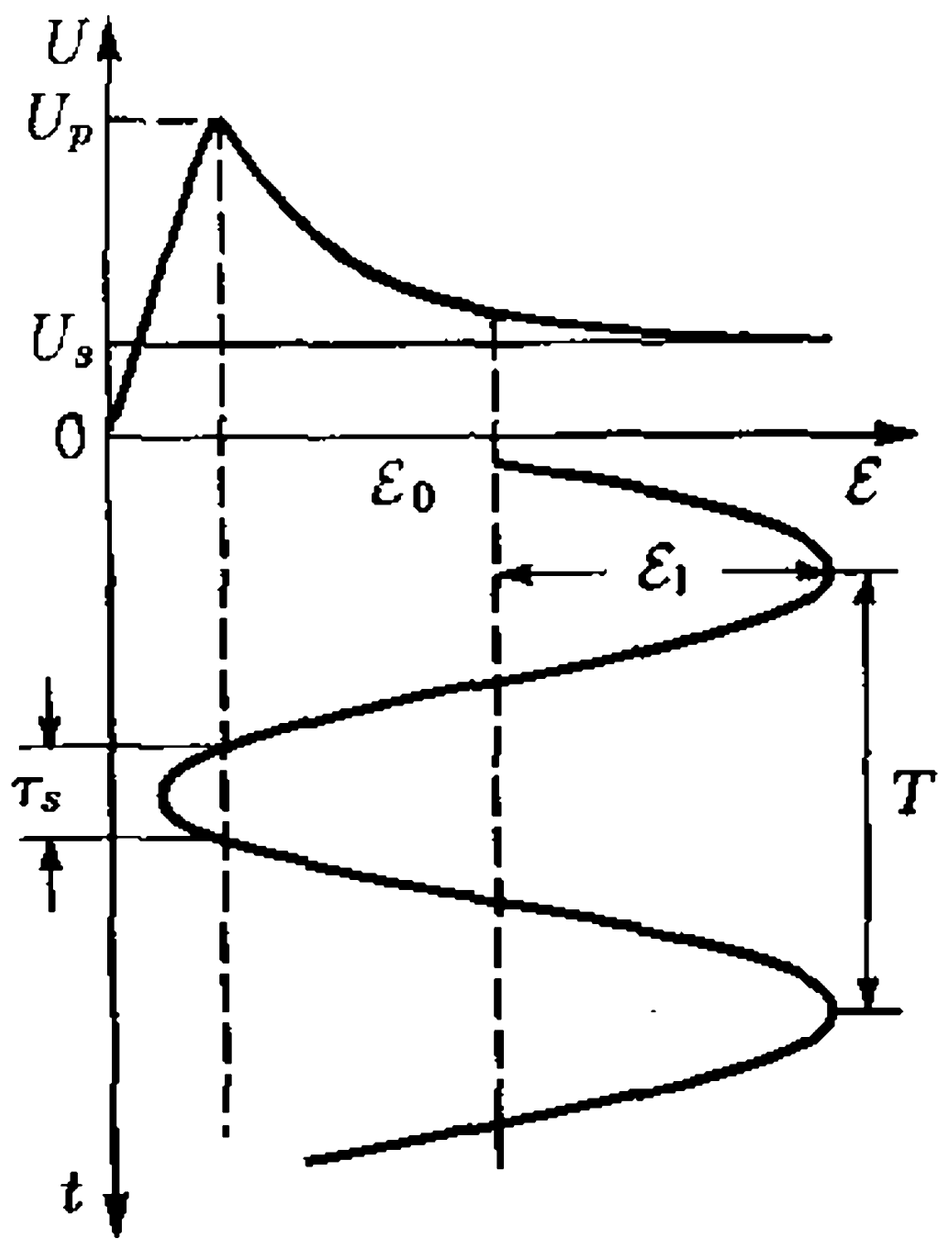


Рис. 6.10. Работа диода Ганна в режиме ОНОЗ

Суть режима ОНОЗ состоит в установлении такого режима генерации (см. рис. 6.10), при котором диод бóльшую часть периода колебаний находился бы в области с отрицательным

[1]) В зарубежной литературе для обозначения этого режима используется аббревиатура LSA (limited space-charge accumulation).

дифференциальным сопротивлением (когда в диоде распространяется нарастающий по амплитуде обогащенный слой), а в течение небольшой доли периода он попадал бы в область с положительным дифференциальным сопротивлением (когда накопленный в обогащенном слое объемный заряд рассасывается). Чтобы накопления объемного заряда не происходило, необходимо, чтобы частота колебаний превышала обратное время формирования домена ($\omega\tau_f \gtrsim 1$) и чтобы доля времени, в течение которой диод находится в области с положительным дифференциальным сопротивлением, удовлетворяло условию

$$\frac{\tau_s}{T} > \frac{|\overline{\mu}_d|}{|\overline{\mu}_d| + \mu_1},$$

где $|\overline{\mu}_d|$ — среднее по времени значение отрицательной дифференциальной подвижности (см. с. 340), τ_s — время нахождения в области с положительным дифференциальным сопротивлением, а T — период колебаний. Это условие следует из условия ненарастания объемного заряда за полный период колебаний:

$$\exp\left(\frac{T-\tau_s}{|\tau_{Md}|}\right)\exp\left(-\frac{\tau_s}{\tau_M}\right) < 1, \qquad (6.15)$$

где $\tau_M = \varepsilon/(4\pi q n_0 \mu_1)$ — максвелловское время релаксации в слабом поле.

Режим ОНОЗ позволяет диоду Ганна работать на частотах, многократно превышающих пролетную частоту. Максимальная частота генерации диодов Ганна в этом режиме составляет ~100 ГГц, а к.п.д. генераторов достигает 15% в импульсном режиме.

Чтобы получить с помощью диодов Ганна частоту выше ~ ~100 ГГц, используют *режим выделения гармоник*. Диод Ганна возбуждается на более низкой частоте и из спектра его колебаний выделяется вторая или третья гармоника [69] [1]). Мощность, которую можно получить таким образом, оказывается выше мощности, получаемой при возбуждении диода на той же частоте на основной гармонике. Таким способом в работе [227] в диоде из InP на второй гармонике была получена выходная мощность более 1 мВт в непрерывном режиме на частоте 315 ГГц.

---

[1]) Одной из причин, препятствующих возбуждению особо высокочастотных колебаний на основной гармонике, является существование вблизи катода «мертвого слоя» толщиной около 1 мкм. Он возникает из-за того, что для разогрева электронов и их перехода с побочные минимумы требуется конечное время.

Фосфит индия как материал для диодов Ганна характеризуется более высокой максимальной скоростью дрейфа и более высоким отношением скоростей в максимуме и минимуме на зависимости $v(\mathcal{E})$ по сравнению с GaAs. Поэтому он позволяет получить бóльшую мощность с более высоким к.п.д. Однако из-за примерно втрое более высокого порогового поля $\mathcal{E}_T$ в InP по сравнению с GaAs приборы из фосфида индия потребляют намного бóльшую мощность и возникает серьезная проблема отвода от них тепла. По этой причине InP используется только для создания диодов с короткой активной областью (пролетная частота 30–170 ГГц). При использовании алмазного теплоотвода (теплопроводность алмаза примерно в 5 раз выше теплопроводности меди) в диодах из InP на основной частоте 100 ГГц удается получить выходную мощность 200 мВт.

Для создания диодов Ганна на сильно легированной подложке n^+-типа выращивается эпитаксиальный слой n-типа, толщина которого L выбирается в соответствии с необходимой частотой генерации, а затем тонкий слой n^+-типа (см. рис. 6.8*б*). После этого подложка утончается путем шлифовки и на обеих сторонах пластины создаются омические контакты. Из этих структур затем путем травления выделяются диоды — меза-структуры диаметром 50–100 мкм и высотой ~20 мкм. Далее эти диоды монтируются эпитаксиальным слоем вниз (для улучшения отвода тепла) на держателях с хорошим теплоотводом.

Диоды Ганна являются сравнительно простыми и недорогими приборами, имеющими большой срок службы. Они используются в качестве радаров для предотвращения столкновений на транспорте и в авиации, как измерители дальности и скорости, бесконтактные вибродатчики, детекторы движения, сенсоры (например, в устройствах автоматического открывания дверей), в качестве задающих генераторов для умножителей частоты в радиоастрономии. Примерами отечественных диодов Ганна из GaAs могут служить приборы 3А703 (диапазон 8,2–12,5 ГГц) и 3А727 (37–54 ГГц). Мощность СВЧ излучения, полученная с помощью диодов Ганна на частоте 10 ГГц достигает 2 Вт в непрерывном режиме и 1 кВт в импульсном режиме. Максимальная мощность генерации изменяется с частотой примерно как f^{-2} [1]).

---

[1]) Появление такой зависимости легко понять, если учесть, что в пролетном режиме длина образца L обратно пропорциональна частоте, а мощность, отдаваемая в нагрузку, пропорциональна квадрату подаваемого на диод Ганна напряжения ($P \sim \mathcal{E}_T^2 L^2 \sim 1/f^2$).

6.2. Лавинно-пролетные диоды

Возможность получения динамического отрицательного дифференциального сопротивления за счет пролетного запаздывания была продемонстрирована Левелином и Боуеном [228] в 1939 г. на вакуумных диодах. В 1954 г. Шокли [229] предложил способ получения отрицательного дифференциального сопротивления за счет того же эффекта в полупроводниковых p–n–p- и n–p–n-структурах; позже эти идеи были использованы при разработке инжекционно-пролетных диодов (см. п. 6.3). Предсказываемая для рассмотренных Шокли структур величина отрицательного сопротивления, однако, казалась недостаточной для практических применений. Развивая работы Bell Laboratories в этом направлении, в 1958 г. Рид [230] предложил конструкцию лавинно-пролетного диода (ЛПД) [1], в которой наряду с пролетным запаздыванием используется возникающее при ударной ионизации лавинное запаздывание, которое существенно понижает величину динамического отрицательного сопротивления. [2] Из-за технологических трудностей первый работающий диод с предложенной Ридом конструкцией был создан только в 1964 г.

В то же время исследования СВЧ преобразователей частоты, проводившиеся в СССР А. С. Тагером с сотрудниками, обнаружили в 1959 г. возникновение СВЧ колебаний в условиях лавинного пробоя обычных диодов из Ge [231]; это явление было зарегистрировано как открытие [232]. Позже теоретические работы подтвердили, что отрицательное дифференциальное сопротивление, обусловленное лавинным и пролетным запаздыванием, может возникать не только в диодах Рида, но и в обычных p–n-переходах, p–i–n-диодах и

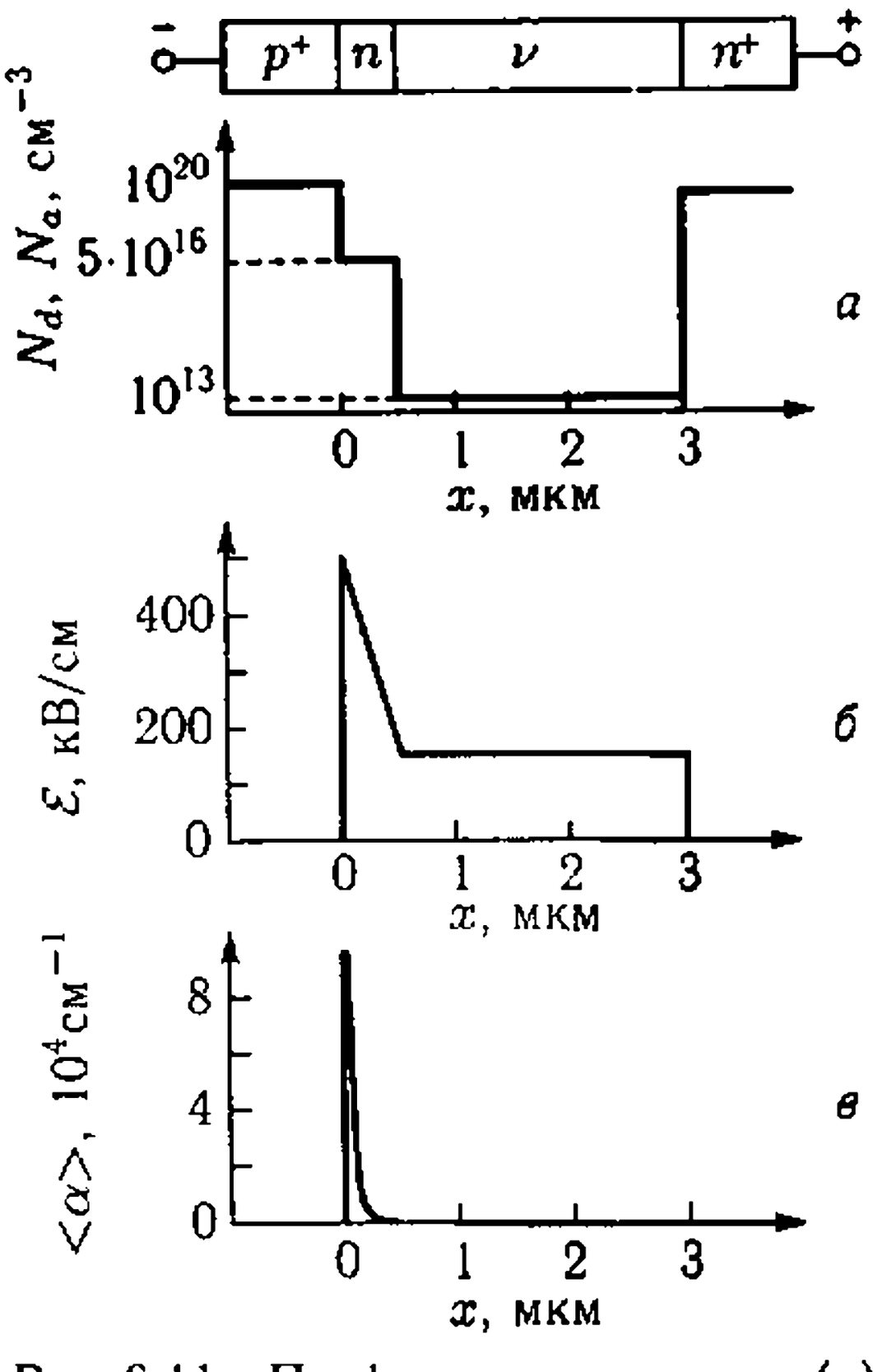


Рис. 6.11. Профиль легирования (*а*) и распределение электрического поля (*б*) и эффективного коэффициента ударной ионизации (*в*) в диоде Рида

[1]) В зарубежной литературе эти диоды называют IMPATT (Impact avalanche transit-time) диодами.

[2]) Для эффективной работы приборы с отрицательным дифференциальным сопротивлением должны иметь не высокое дифференциальное сопротивление, а высокую дифференциальную проводимость.

барьерах Шоттки с произвольным профилем легирования. В настоящее время лавинно-пролетные диоды являются самыми мощными твердотельными источниками СВЧ излучения (их мощность примерно на порядок выше мощности диодов Ганна).

Предложенная Ридом конструкция лавинно-пролетного диода называется *диодом Рида*. Этот диод (см. рис. 6.11) имеет структуру p^+–n–ν–n^+ и состоит из p^+–n-перехода, на который подается обратное смещение и в котором происходит лавинный пробой, и расположенной рядом с ним достаточно протяженной *области дрейфа* (ν-области), в которой инжектированные зарядовые пакеты дрейфуют в сильном и достаточно однородном электрическом поле. Область n^+ выполняет роль контакта.

Принцип действия. Малосигнальный импеданс. Чтобы понять принцип действия диода Рида, рассмотрим динамический отклик этой структуры сначала на качественном уровне. Пусть лавинно-пролетный диод помещен в некий резонатор, так что к нему одновременно приложены постоянное напряжение, достаточное для возникновения пробоя, и напряжение СВЧ колебаний. Чтобы диод поддерживал колебания в резонаторе, необходимо, чтобы он обладал отрицательным дифференциальным сопротивлением, то есть чтобы сдвиг фаз между колебаниями тока и напряжения на нем составлял 180°. Сдвиг фаз в 90° возникает при лавинном пробое p–n-перехода из-за того, что ток лавины запаздывает по отношению к приложенному к p–n-переходу напряжению: из-за конечного времени нарастания лавины наибольшая скорость прироста тока отвечает моментам, когда напряжение на p–n-переходе максимально. Дополнительный сдвиг фаз (еще на 90°) возникает динамически, за счет задержки при распространении зарядовых пакетов в области дрейфа.

Рассмотрим динамический отклик диода Рида на малом сигнале на частоте ω. Представим плотность тока и электрическое поле в образце в виде суммы постоянной и переменной составляющих:

$$J = J_0 + \delta J e^{i\omega t}, \quad \mathcal{E} = \mathcal{E}_0 + \delta\mathcal{E} e^{i\omega t},$$

где комплексные амплитуды δJ и $\delta\mathcal{E}$ могут зависеть от координаты. Лавинно-пролетный диод можно условно разделить на три области:

1) область лавинного умножения, которую Рид считал очень тонкой, чтобы не учитывать в ней явления запаздывания,

2) область дрейфа, в которой напряженность электрического поля велика и в которой носители движутся со скоростью насыщения v_s и

3) приконтактную область и контакты, которые определяют последовательное сопротивление диода.

При изложении этого материала мы будем следовать работе [233].

В любом поперечном сечении образца переменную часть плотности тока δJ можно представить в виде суммы переменной части плотности тока проводимости δJ_c и плотности тока смещения:

$$\delta J = \delta J_c(x) + \frac{\varepsilon}{4\pi}\frac{d\mathcal{E}(x)}{dt} = \delta J_c(x) + i\omega\frac{\varepsilon}{4\pi}\delta\mathcal{E}(x). \qquad (6.16)$$

Вкладом тока диффузии в δJ в первом приближении можно пренебречь. Ток проводимости связан с движением реальных носителей заряда, а ток смещения — с изменением напряженности электрического поля.

В области умножения плотность тока проводимости равна плотности тока лавинного пробоя J_{Ac}. Обозначим буквой γ отношение переменной части плотности этого тока δJ_{Ac} к переменной части полной плотности тока δJ; это — некоторая пока неизвестная величина, которую мы вычислим позже. Поскольку мы предположили, что область умножения тонка, то ток δJ_{Ac} достигает границы области дрейфа без запаздывания.

Расчет импеданса (комплексного сопротивления) диода Рида начнем с расчета импеданса области дрейфа. Плотность тока, связанную с движущейся волной объемного заряда, можно представить как волну, распространяющуюся со скоростью насыщения:

$$\delta J_c(x) = \delta J_{Ac}(0)e^{-i\omega x/v_s} = \gamma\cdot\delta J e^{-i\omega x/v_s}, \qquad (6.17)$$

где $\delta J_{Ac}(0)$ — амплитуда переменной части плотности тока проводимости на входе области дрейфа. Мы не учитываем затухания волны потому, что время пролета обычно на несколько порядков меньше времени жизни носителей. Подставляя уравнение (6.17) в (6.16), находим, что величина $\delta\mathcal{E}$ в области дрейфа зависит от x следующим образом:

$$\delta\mathcal{E}(x) = \delta J\frac{4\pi(1-\gamma e^{-i\omega x/v_s})}{i\omega\varepsilon}. \qquad (6.18)$$

Интегрирование $\delta\mathcal{E}(x)$ по всей области дрейфа позволяет рассчитать переменное напряжение, падающее на этой области. Отношение этого напряжения к амплитуде протекающего

переменного тока и есть импеданс Z_d области дрейфа:

$$Z_d = \frac{1}{S\delta J}\int_0^W \delta\mathcal{E}(x)\,dx = \frac{4\pi W}{i\omega\epsilon S}\left[1-\gamma\frac{1-e^{-i\vartheta}}{i\vartheta}\right]. \qquad (6.19)$$

Входящая в эту формулу величина $\vartheta = \omega W/v_s$ называется *пролетным углом*; W — толщина области дрейфа, S — площадь сечения диода.

Рассчитаем теперь импеданс области лавинного умножения и величину γ. Решение системы уравнений непрерывности с учетом генерации носителей, вызванной ударной ионизацией электронов и дырок, приводит к следующему уравнению для изменения плотности тока проводимости [14, 230]:

$$\frac{dJ_{Ac}}{dt} = \frac{2J_{Ac}}{\tau_A}\left[\int_0^{x_A}\alpha(x)\,dx - 1\right], \qquad (6.20)$$

справедливому при $J_0 \gg J_s$, где J_s — плотность тока насыщения диода. Здесь x_A — толщина области умножения, $\tau_A = x_A/v_s$ — время пролета носителями этой области, а α — коэффициент ударной ионизации. Если предположить, что электрическое поле в области умножения однородно, то выражение в квадратных скобках уравнения (6.20) можно заменить на $(\bar{\alpha}x_A - 1)$, где $\bar{\alpha}$ — среднее значение коэффициента ударной ионизации. Считая отклик коэффициента ударной ионизации на изменение электрического поля мгновенным, представим $\bar{\alpha}$ в виде суммы постоянной величины $\bar{\alpha}_0$ и переменной составляющей, пропорциональной напряженности переменного электрического поля в области умножения $\delta\mathcal{E}_A e^{i\omega t}$:

$$\bar{\alpha} = \bar{\alpha}_0 + \frac{d\bar{\alpha}}{d\mathcal{E}}\delta\mathcal{E}_A e^{i\omega t}, \quad \bar{\alpha}x_A = 1 + x_A\frac{d\bar{\alpha}}{d\mathcal{E}}\delta\mathcal{E}_A e^{i\omega t}. \qquad (6.21)$$

При записи второго уравнения мы учли условие $\bar{\alpha}_0 x_A = 1$, отвечающее стационарному решению уравнения (6.20). Подставляя разложение (6.21) в уравнение (6.20), находим амплитуду переменной компоненты плотности тока лавинного умножения:

$$\delta J_{Ac} \approx 2\frac{d\bar{\alpha}}{d\mathcal{E}}\frac{J_0 v_s}{i\omega}\delta\mathcal{E}_A, \qquad (6.22)$$

где J_0 — плотность протекающего в области лавинного умножения постоянного тока. Плотность тока смещения в области

умножения равна

$$\delta J_{Ad} = i\omega \frac{\varepsilon}{4\pi} \delta \mathcal{E}_A. \tag{6.23}$$

Выражения (6.22) и (6.23) позволяют рассчитать импеданс области умножения

$$Z_A = \frac{x_A \delta \mathcal{E}_A}{S(\delta J_{Ac} + \delta J_{Ad})} = \frac{4\pi x_A}{i\omega\epsilon S} \left(\frac{1}{1 - \omega_r^2/\omega^2} \right), \tag{6.24}$$

$$\omega_r^2 = 8\pi \frac{d\overline{\alpha}}{d\mathcal{E}} \frac{J_0 v_s}{\varepsilon},$$

и найти введенную нами ранее величину γ:

$$\gamma = \frac{\delta J_{Ac}}{\delta J} = \frac{\delta J_{Ac}}{\delta J_{Ac} + \delta J_{Ad}} = \frac{1}{1 - \omega^2/\omega_r^2}. \tag{6.25}$$

Физический смысл величины ω_r, входящей в эти формулы, может быть понят, если учесть, что ток лавинного умножения носит индуктивный, а ток смещения — емкостный характер (это следует из соотношения фаз напряжений и токов в уравнениях (6.22) и (6.23)). Таким образом, на переменном токе эквивалентная схема области лавинного умножения представляет собой параллельный колебательный контур, собственная частота которого («лавинная частота») и равна ω_r.

Подставляя найденное значение γ в уравнение (6.19) и суммируя импедансы трех отдельных областей (третья область имеет импеданс R_s), в итоге находим полный импеданс диода Рида:

$$Z = \frac{4\pi W}{i\omega\epsilon S} \left[1 - \frac{1}{1 - \omega^2/\omega_r^2} \left(\frac{1 - e^{-i\vartheta}}{i\vartheta} \right) \right] + \\ + \frac{4\pi x_A}{i\omega\epsilon S} \left(\frac{1}{1 - \omega_r^2/\omega^2} \right) + R_s. \tag{6.26}$$

Из формулы (6.26) следует, что для того, чтобы действительная часть импеданса была отрицательной, необходимо, чтобы $\omega > \omega_r$, а пролетный угол ϑ не был бы кратен 2π.

Мисава [234] провел расчеты импеданса другого предельного случая лавинно-пролетного диода — p–i–n-диода, в котором вся структура состоит только из одной однородной области умножения (эту структуру часто называют диодом Мисавы). Он показал, что при пробое p–i–n-диода в широкой области частот всегда возникает динамическое отрицательное дифференциальное сопротивление. В работе [235] Мисава развил технику расчета импеданса структур с произвольной последовательностью

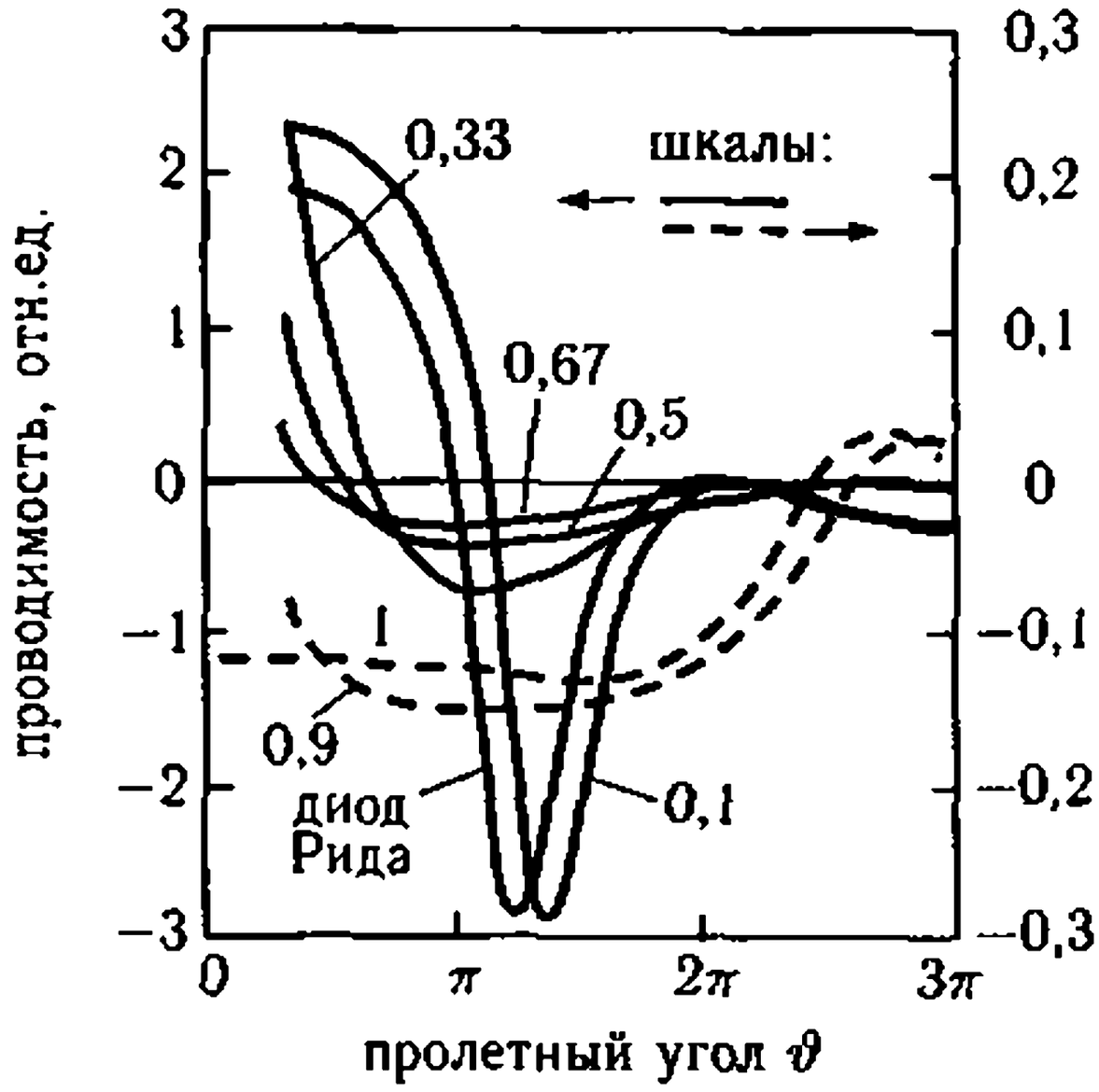


Рис. 6.12. Зависимость действительной части проводимости лавинно-пролетного диода от пролетного угла для диода Рида ($x_A/W = 0$) и ряда структур с конечным отношением x_A/W [235]. Значение частоты ω_r равно $\pi v_s/W$

областей умножения и дрейфа и подробно рассмотрел структуры, состоящие из области умножения конечной ширины x_A и области дрейфа ширины $(W - x_A)$. Расчетная зависимость действительной части комплексной проводимости от пролетного угла для диода Рида и ряда диодов с различным отношением x_A/W показана на рис. 6.12. Видно, что в определенном диапазоне частот (углов пролета) отрицательное дифференциальное сопротивление возникает в диодах с *любым* отношением x_A/W, причем в диодах с $x_A/W \gtrsim 0,3$ оно наблюдается и на частотах $\omega < \omega_r$ из-за большого вклада отрицательного дифференциального сопротивления области лавинного пробоя.

Анализ полученных Мисавой результатов позволяет заключить, что оптимальным режимом возбуждения колебаний в лавинно-пролетных диодах является режим, при котором пролетный угол близок к π. Поскольку частота ω_r зависит от J_0 (см. формулу (6.24)), при этом важно еще правильно выбрать величину тока диода. Оптимальные условия генерации достигаются, когда ω_r лежит на 20–30% ниже частоты генерации [235]. Из приведенного соотношения и формулы (6.24) следует, что с ростом частоты генерации плотность тока J_0 должна увеличиваться пропорционально квадрату этой частоты; на 100 ГГц эта плотность тока составляет ~30 кА/см^2 и продвижение в область более высоких частот связано с большими трудностями.

Шумы в лавинно-пролетных диодах. Отрицательное дифференциальное сопротивление лавинно-пролетных диодов позволяет использовать их для усиления СВЧ сигналов. Такие усилители обычно строятся по схеме отражательного типа [14]. К сожалению, ЛПД характеризуются довольно высоким уровнем шума, связанным с флуктуациями числа электронно-дырочных пар, рождаемых каждым участвующим в ударной ионизации носителем. Коэффициент шума диодов сильно зависит от соотношения коэффициентов ударной ионизации электронов и дырок в используемом материале (см. п. 1.3.1). Так, в диодах из GaAs, в котором $\alpha_n \approx \alpha_p$, наименьший достигнутый коэффициент шума в режиме малого сигнала ($F \approx 25$ дБ) был заметно ниже, чем в диодах из Si ($F \approx 40$ дБ), в котором $\alpha_n \gg \alpha_p$ [14]. Из-за высокого уровня шума лавинно-пролетные диоды практически не используются во входных цепях приемников СВЧ сигналов, однако они широко применяются в качестве усилителей мощности в выходных каскадах передатчиков.

Лавинно-пролетные диоды, построенные на основе гетероструктур, как правило, характеризуются более низким уровнем шума по сравнению с ЛПД на основе p–n-переходов. Наиболее низкий уровень шума, равный 7,7 дБ на частоте 190 ГГц, был получен в двухпролетном ЛПД на основе гетероперехода InP–$In_{0,53}Ga_{0,47}As$ с тщательно подобранной толщиной области лавинного пробоя [236].

Мощность и к.п.д. лавинно-пролетных диодов. Лавинно-пролетные диоды являются самыми мощными твердотельными источниками СВЧ колебаний. В отличие от диодов Ганна, лавинно-пролетные диоды могут быть изготовлены практически из любого полупроводника; на практике чаще всего используются Si и GaAs. Для оценки потенциальных возможностей лавинно-пролетных диодов рассмотрим их работу в режиме большого сигнала.

На рис. 6.13 показан характер изменения переменного напряжения и тока в диоде Рида в режиме большого сигнала. При изменении приложенного к образцу переменного напряжения с амплитудой V_a относительно его постоянного значения V_B (диод находится в СВЧ резонаторе) коэффициент умножения практически мгновенно отслеживает изменение напряженности электрического поля. Однако концентрация носителей в области лавинного умножения отстает от поля, поскольку она определяется числом уже имеющихся в этой области носителей. Концентрация носителей достигает максимума примерно в тот момент,

когда напряжение на образце уменьшается до значения V_B (см. рис. 6.13) и, таким образом, ток инжектируемых в область дрейфа электронов запаздывает по отношению к переменному напряжению приблизительно на 90°. Это — так называемое *лавинное запаздывание*.

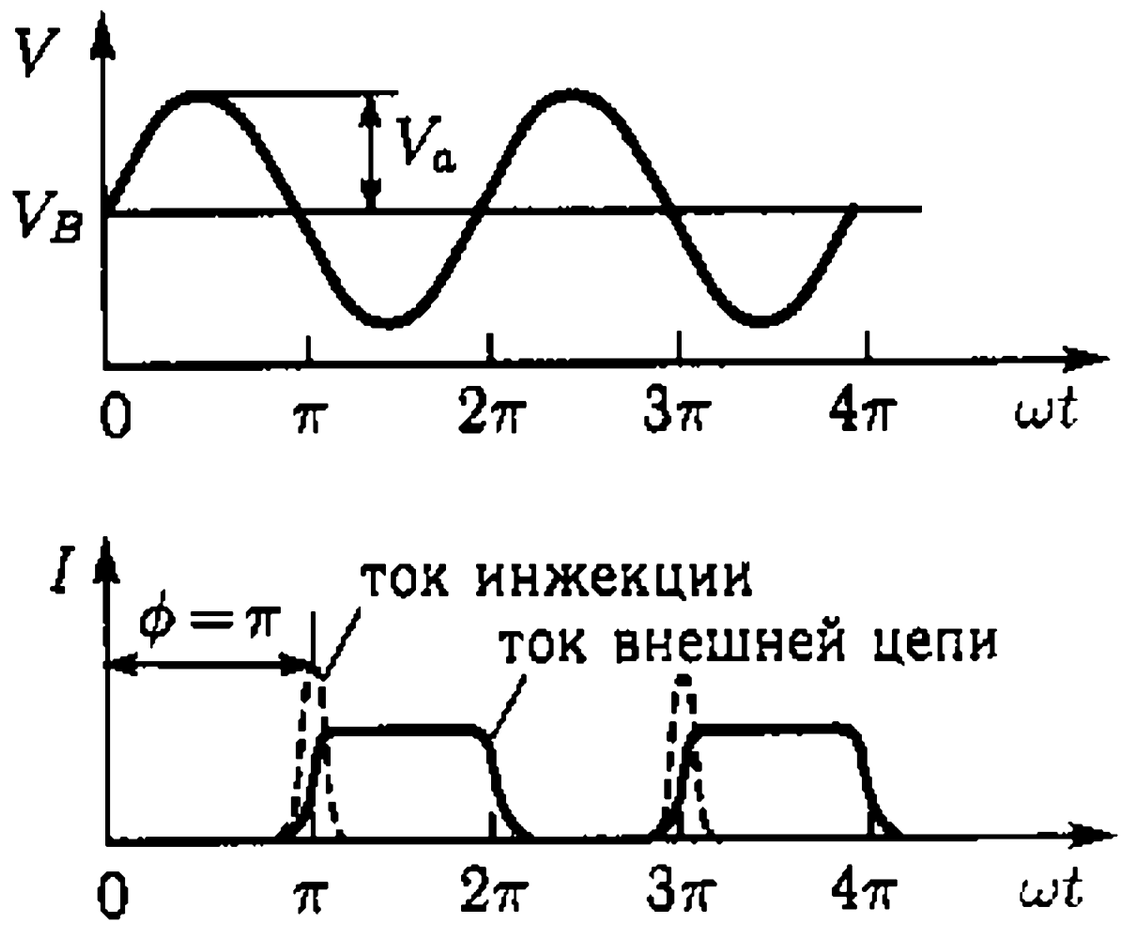


Рис. 6.13. Характер изменения переменного напряжения и тока в лавинно-пролетном диоде в режиме большого сигнала

Возбужденные в области лавинного умножения дырки уходят в p^+-область, а электроны попадают в область дрейфа и движутся там со скоростью насыщения. Ток в цепи продолжает течь пока электроны не достигнут анода (см. рис. 6.13), что вносит *пролетное запаздывание*. Нетрудно видеть, что на частоте, для которой время пролета равно половине периода колебаний, переменный ток в лавинно-пролетном диоде протекает в промежутки времени, когда переменное напряжение на диоде отрицательно, то есть знак дифференциального сопротивления прибора на этой частоте действительно отрицателен. Таким образом, оптимальные условия для возбуждения колебаний отвечают частоте

$$f \approx \frac{v_s}{2W}, \tag{6.27}$$

где W — толщина области дрейфа; при $W = 5$ мкм для кремния $f \approx 10$ ГГц.

Кривые, представленные на рис. 6.13, позволяют оценить максимальный к.п.д. генератора на ЛПД, который определяется как отношение мощности высокочастотных колебаний к потребляемой мощности. Считая форму тока во внешней цепи прямоугольной, а форму переменного напряжения синусоидальной, несложное интегрирование дает $\eta = 2V_a/(\pi V_B)$, что при $V_a = V_B$ составляет около 64 %. Эта оценка однако оказывается завышенной из-за ряда неучтенных эффектов. Главным из них является то, что сама инжекция зарядового пакета искажает распределение электрического поля в области дрейфа, увеличивая поле в области перед движущимся пакетом и уменьшая его в области

за пакетом (см. рис. 6.14) [1]. Поэтому необходимо, чтобы при движении зарядового пакета напряженность электрического поля в области дрейфа всегда оставалась бы достаточно высокой (чтобы носители двигались со скоростью насыщения), но при этом никогда бы не превышала напряженность поля лавинного пробоя $\mathcal{E}_{\text{проб}}$. Как показали Шарфеттер и Гуммель [237], напряжение, которое следует приложить к p–n-переходу для возникновения лавинного пробоя, сильно зависит от отношения коэффициентов ударной ионизации электронов и дырок α_n/α_p. Если учесть, что падение напряжения на области лавинного умножения практически не меняется за период колебаний, а в качестве среднего значения напряженности поля в области дрейфа взять величину $\mathcal{E}_{\text{проб}}/2$, то для максимального к.п.д. получаем более реалистичную величину:

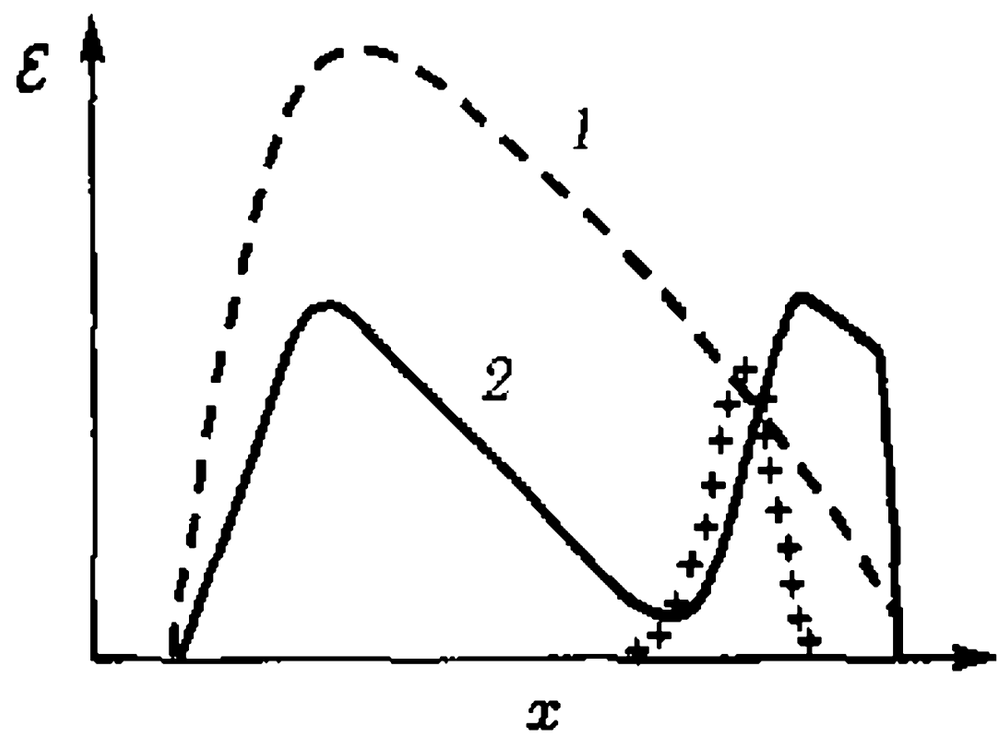


Рис. 6.14. Искажение распределения электрического поля в лавинно-пролетном диоде при движении большого зарядового пакета: *1* — распределение поля в отсутствие заряда, *2* — то же в присутствии заряда. *Крестиками показано распределение плотности движущегося заряда* [14]

$$\eta = \frac{2}{\pi}\,\frac{(V_d/2)}{V_d + V_A}\,, \tag{6.28}$$

где V_d и V_A — средние значения падений напряжения на области дрейфа и области лавинного умножения, соответственно. Оценка отношения V_A/V_d для Si и GaAs показала, что в полупроводниках с $\alpha_n \approx \alpha_p$ (в GaAs) величина $V_A/V_d \approx 1/3$ в то время как для кремния $V_A/V_d \approx 1,1$. Таким образом, теоретический предел к.п.д. ЛПД из Si составляет ~15%, а ЛПД из GaAs — ~23% [237].

Для создания оптимальных условий генерации, при которых средняя напряженность поля в области дрейфа должна быть примерно вдвое ниже, чем в области лавинного умножения, n-область ЛПД изготавливают в виде двух- (n^+—n) или трех-

[1]) Чтобы скачок напряженности электрического поля в области зарядового пакета не превышал $\Delta\mathcal{E}$, необходимо ограничить плотность тока через диод величиной $J_0 \approx \varepsilon f \Delta\mathcal{E}/4\pi$. На частоте 10 ГГц при $\Delta\mathcal{E} = 10^5$ В/см $J_0 \approx 10^3$ А/см^2.

слойных (n–n^+–n) структур с тщательно подобранным профилем легирования [14].

Одним из способов дальнейшего повышения к.п.д. лавинно-пролетных диодов является создание приборов не с одной областью дрейфа (как это было в конструкции, предложенной Ридом), а так называемых *двухпролетных* ЛПД, в которых формируются две области дрейфа, расположенные по разным сторонам области лавинного умножения: одна для электронов, другая для дырок. Выходная мощность ЛПД с такой структурой почти вдвое больше, а к.п.д. на ~40% выше, чем в приборах с одной областью дрейфа [14].

Другим способом повышения к.п.д. является создание лавинно-пролетных диодов на основе гетеропереходов [238]. Известно, что в оптимально сконструированных ЛПД на основе p–n-переходов большая доля приложенного напряжения падает на области лавинного умножения [14]. Поэтому если расположить эту область в узкозонной части гетероперехода, а область дрейфа сделать из широкозонного полупроводника, то из-за более низкой напряженности поля лавинного пробоя в узкозонном полупроводнике можно существенно снизить падение напряжения на области лавинного умножения (V_A в формуле (6.28)) и, следовательно, увеличить к.п.д. Так, в ЛПД на основе гетероперехода GaAs–Ge удалось получить к.п.д. 45% на частоте 8 ГГц в непрерывном режиме [239].

Максимальная мощность, которую можно получить от лавинно-пролетного диода, определяется, кроме к.п.д. прибора, еще двумя факторами: эффективностью отвода тепла от прибора (рабочая температура кристалла не должна превышать 200°С в случае кремния) и физическими свойствами используемого полупроводника. Для ЛПД, работающих на частотах до 30 ГГц в непрерывном режиме, первый фактор является основным. Лавинно-пролетный диод представляет собой меза-структуру, которая крепится к теплоотводу. Если считать, что отвод тепла ограничивается теплопроводностью полупроводника, то поскольку толщина структуры, которая определяется толщиной области дрейфа, изменяется обратно пропорциональна частоте генерации, то и максимальная рассеиваемая мощность, и мощность генерации изменяются как $1/f$.

При работе лавинно-пролетных диодов в импульсном режиме, а также для диодов, работающих в непрерывном режиме на частотах выше 30 ГГц, эффекты нагрева отходят на второй план, и основные ограничения выходной мощности связаны с электрон-

ными характеристиками полупроводника: напряженностью поля лавинного пробоя $\mathcal{E}_{\text{проб}}$ и скоростью насыщения v_s. Для оценки пригодности того или другого полупроводника для изготовления ЛПД часто используют величину $\mathcal{E}^2_{\text{проб}} v_s^2$, которая входит в формулу для оценки максимальной мощности генерации:

$$P_{\max} \approx \frac{\mathcal{E}^2_{\text{проб}} v_s^2}{8\pi f^2 X_c},$$

где $X_c \approx 2W/(\varepsilon f S)$ — импеданс диода [14]. В кремнии величина $\mathcal{E}^2_{\text{проб}} v_s^2$ примерно втрое больше, чем в GaAs.

При выборе материала для изготовления ЛПД следует иметь в виду следующее. На частотах $f \lesssim 10$ ГГц диоды из GaAs имеют более высокий к.п.д. [1]), однако из-за втрое более низкой теплопроводности арсенида галлия по сравнению с кремнием, отвод большого количества тепла от ЛПД, изготовленных из GaAs, затруднен. С другой стороны, благодаря более высокой теплопроводности Si, к кремниевым ЛПД можно подводить бóльшую электрическую мощность, и, несмотря на меньший к.п.д., получать от них сравнимую выходную мощность. Из-за более низкого быстродействия GaAs по сравнению с Si (см. ниже) к.п.д. этих диодов быстро уменьшается с ростом частоты, и в области высоких частот кремний, в соответствии с величиной $\mathcal{E}^2_{\text{проб}} v_s^2$, позволяет получить втрое более высокую мощность по сравнению с GaAs. Поэтому на частотах выше 35 ГГц большинство ЛПД изготавливают из кремния.

В настоящее время выходная мощность лавинно-пролетных диодов из Si и GaAs приблизилась к своему теоретическому пределу, и возникла необходимость в поиске новых полупроводников для создания более мощных ЛПД. По величине $\mathcal{E}^2_{\text{проб}} v_s^2$ наиболее перспективным материалом для лавинно-пролетных диодов считается карбид кремния, в котором эта величина примерно в 100 раз выше, чем для Si и GaAs. В лабораторных образцах ЛПД из SiC уже получена генерация на частотах 10 и 35 ГГц [240, 241]. По теоретическим оценкам, работа лавинно-пролетных диодов из карбида кремния возможна до частот 450–500 ГГц.

---

[1]) Одной из причин более высокого к.п.д. диодов из GaAs является формирование более четких границ зарядовых пакетов. Благодаря уменьшению скорости дрейфа с ростом напряженности электрического поля в области отрицательного дифференциального сопротивления, электроны в задней части пакета (той, где напряженность поля ниже, см. рис. 6.14), движутся *быстрее* и догоняют пакет.

Быстродействие лавинно-пролетных диодов. Генераторы на лавинно-пролетных диодах работают на частотах до ~390 ГГц и перекрывают весь диапазон сантиметровых и миллиметровых длин волн. Верхняя граница частоты генерации ЛПД определяется несколькими эффектами. Во-первых, это — конечное «время *лавинного отклика*», которое равно времени релаксации энергии в сильном электрическом поле. В Si, GaAs и InP значения этого времени примерно равны и составляют ~0,6 пс [239].

Другое ограничение связано с диффузионным расплыванием зарядового пакета при его движении в области дрейфа. Максимальную частоту, связанную с этим эффектом, можно оценить по формуле $f_{max} \approx v_s^2/6D_n$ [242]. Нетрудно видеть, что этот эффект должен быть особенно сильно выражен для электронов в GaAs, в котором подвижность и коэффициент диффузии электронов велики. Оценки показывают, что максимальная частота ЛПД из GaAs ограничена величиной ~80 ГГц; для кремния эта частота составляет ~460 ГГц [242]. При работе ЛПД с зарядовыми пакетами *дырок* максимальные частоты в Si и GaAs превышают 1000 ГГц, и поэтому не удивительно, что наивысшая частота генерации в ЛПД (394 ГГц [243]) была получена в приборе из Si, в котором использовался дрейф дырок. Наконец, существует ограничение, связанное с «насыщением» коэффициента ударной ионизации в сильных электрических полях и увеличением роли туннельного пробоя. Обсудим последнее ограничение более подробно.

Очевидно, что с ростом частоты генерации оптимальная толщина области лавинного умножения должна уменьшаться. Уменьшение размеров этой области влечет за собой резкое возрастание напряженности электрического поля, необходимого для возникновения пробоя. Однако, как мы знаем из материала п. 1.3, при этом заметно возрастает вероятность туннелирования электрона сквозь потенциальный барьер. Поэтому в наиболее высокочастотных ЛПД наряду с компонентой тока лавинной инжекции, запаздывающей по отношению к напряжению на 90°, появляется и компонента тока *туннельной инжекции*, для которой такого запаздывания нет. Уменьшение фазового угла инжекционного запаздывания приводит к ухудшению условий возбуждения колебаний, а при попытке дальнейшего увеличения частоты лавинно-пролетный диод переходит в режим чисто туннельной инжекции. Особенности работы диодов в этом режиме будут рассмотрены нами в п. 6.4.2.

Конструкция лавинно-пролетного диода похожа на конструкцию диода Ганна (см. рис. 6.8*б*). В наиболее простом варианте (структура p^+–n–n^+) необходимый профиль легирования создается путем диффузии акцепторной примеси в тонкий эпитаксиальный слой n-типа, выращенный на сильно легированной подложке [239]. Более сложные структуры (например, двухпролетные ЛПД) создают методом многослойного эпитаксиального наращивания. В высокочастотных приборах толщина области дрейфа настолько мала (менее 1 мкм), что для создания необходимого профиля легирования часто используется ионная имплантация. После создания структуры из пластины с помощью травления выделяются меза-диоды, которые монтируются на теплоотвод, причем для улучшения условий отвода тепла их часто монтируют в «перевернутом состоянии», чтобы укоротить путь от активной области к теплоотводу. Для улучшения отвода тепла в мощных ЛПД вместо теплоотвода из меди часто используется алмазный теплоотвод [239].

Выходная мощность лавинно-пролетных диодов из GaAs и Si на частоте 10 ГГц достигает ~50 Вт в импульсном режиме и ~5 Вт в непрерывном режиме; в диодах из Si в непрерывном режиме можно получить ~1 Вт на 100 ГГц и 200 мкВт на частоте 361 ГГц [14, 52]. Следует заметить, что в связи с быстрым развитием полевых транзисторов (см. гл. 4) многие характеристики ЛПД, работающих в области частот до ~100 ГГц, уже превзойдены этими транзисторами. Однако в коротковолновой части диапазона миллиметровых волн и в субмиллиметровом диапазоне ЛПД продолжают оставаться самыми мощными полупроводниковыми источниками СВЧ излучения.

Лавинно-пролетные диоды находят применение в системах космической связи [1]; в радиолокации; системах наведения (в том числе ракет); в связи (радиорелейные линии, беспроводные локальные сети); в медицине. Кроме того, как и диоды Ганна, ЛПД используются в охранных системах; для измерения расстояний, скорости и вибраций; в системах предотвращения столкновений (автотранспорт, авиация, судоходство); системах контроля скоростного режима на автострадах [239]. Отечественной промышленностью выпускается целый ряд ЛПД, например, кремниевые диоды 2А706 (диапазон 8,5–11,5 ГГц), 2А717 (31–

[1]) Диапазон частот 3,4–44 ГГц выделен под каналы передачи Земля–космос и космос–Земля, а область частот около 60 ГГц, и которой радиоволны сильно поглощаются молекулярным кислородом в атмосфере, используют для создания защищенных от перехвата систем межспутниковой связи.

62 ГГц), 2А756 (85–100 ГГц) и арсенид-галлиевые диоды 3А707 (8,3–16,7 ГГц), 3А759 (35–37 ГГц).

6.3. Инжекционно-пролетные диоды.

Основным недостатком лавинно-пролетных диодов является высокий уровень шума, связанный со статистическими флуктуациями числа электронно-дырочных пар, рождаемых носителями при ударной ионизации. В 1968 г. Рюэгг [244] и Райт [245] независимо предложили новый тип прибора с динамическим отрицательным дифференциальным сопротивлением, в котором для инжекции зарядовых пакетов используется потенциальный барьер (p–n-переход или барьер Шоттки) [1]) Генерация СВЧ колебаний в диодах этого типа, названных инжекционно-пролетными диодами (ИПД) [2]), впервые осуществлена в 1971 г. Колманом и Зи в Bell Laboratories [246].

Конструкция и принцип действия. Одна из возможных конструкций инжекционно-пролетного диода и ее энергетическая диаграмма показаны на рис. 6.15. В общем случае ИПД представляет собой структуру, в которой слой полупроводника расположен между двумя включенными навстречу друг другу выпрямляющими контактами: двумя барьерами Шоттки (структура металл–полупроводник n-типа–металл), двумя p–n-переходами (структура p^+–n–p^+) или барьером Шоттки и p–n-переходом. При подаче напряжения на такую структуру один из переходов оказывается смещенным в прямом направлении, а другой — в обратном. С ростом напряжения область пространственного заряда обратно смещенного перехода быстро расширяется, пока не произойдет смыкание областей пространственного заряда (так называемый *прокол*), после чего ток через структуру начинает очень быстро возрастать с увеличением напряжения [3]). Уровень легирования и толщина слоя полупроводника подбираются так, чтобы прокол

---

[1]) Впервые возможность получения отрицательного дифференциального сопротивления в приборах такой конструкции обсуждалась в работе Шокли [229].

[2]) В зарубежной литературе эти диоды называются BARITT (barrier-injection transit-time) диодами.

[3]) Подобные структуры, называемые *диодами с проколом*, используются в электронных схемах в качестве быстродействующих ограничителей напряжения, а также стабилитронов. Интересно, что созданные таким образом низковольтные стабилитроны (на напряжение, скажем, 1,5 В) по коэффициенту качества превосходят зенеровские диоды, которые мы рассматривали в п. 1.3.4.

структуры происходил раньше, чем начинал развиваться пробой (см. рис. 3.3).

Для понимания работы инжекционно-пролетного диода рассмотрим сначала статическую вольт-амперную характеристику структуры металл–полупроводник–металл [247]. Напряжение прокола можно найти, приравняв сумму толщин обедненных слоев W_1 и W_2 толщине слоя полупроводника W (см. рис. 6.15). Это напряжение равно

$$V_{\text{прок}} \approx \frac{2\pi q N_d}{\varepsilon} W^2 - W\left(\frac{8\pi q N_d V_{bi}}{\varepsilon}\right)^{1/2}. \quad (6.29)$$

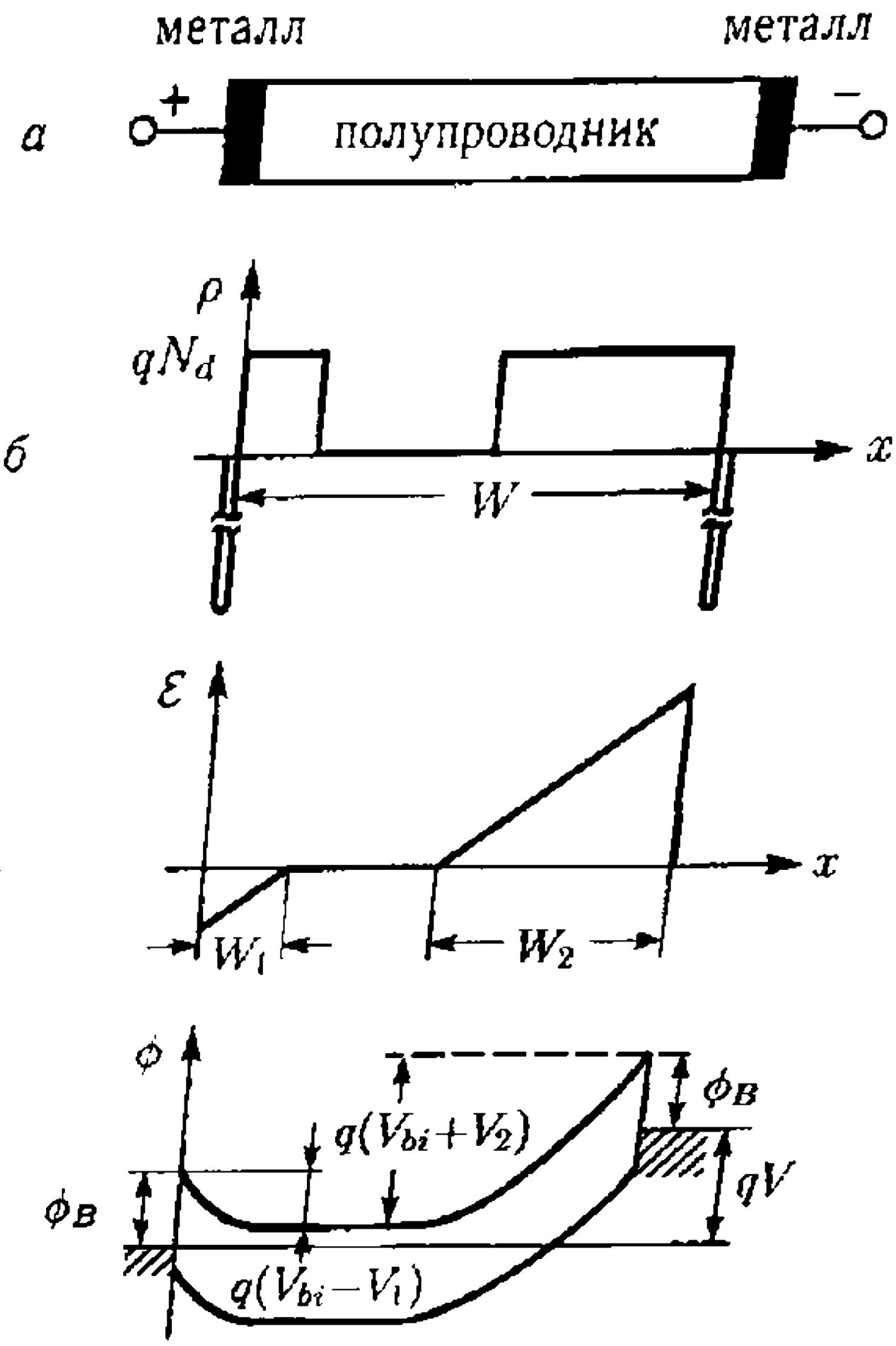


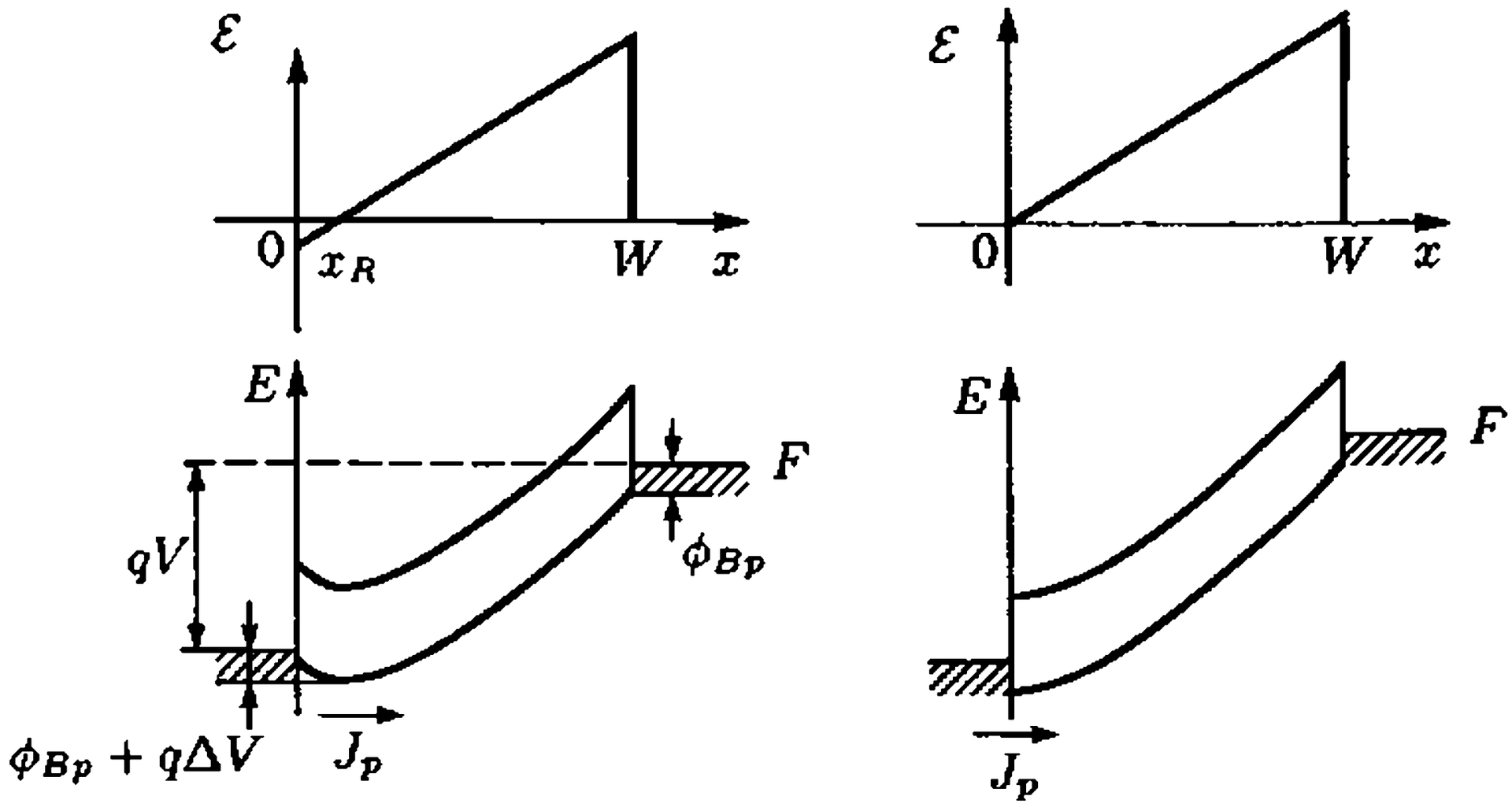
Рис. 6.15. Структура металл–полупроводник–металл (*а*) и распределение в ней плотности зарядов (*б*), напряженности электрического поля (*в*) и потенциала (*г*) при $V < V_{\text{прок}}$

Энергетическая диаграмма структуры при $V = V_{\text{прок}}$ показана на рис. 6.16*а*.

При напряжении $V < V_{\text{прок}}$ ток через структуру в основном определяется током электронов, втекающих из отрицательного контакта в полупроводник (током обратно смещенного барьера Шоттки), а ток дырок, инжектируемых положительным контактом, мал (см. рис. 6.17). При $V > V_{\text{прок}}$ вольт-амперная характеристика качественно изменяется. После смыкания областей пространственного заряда положение края валентной зоны $E_v(x)$ начинает напрямую зависеть от приложенного к структуре напряжения, причем с ростом V высота потенциального барьера для дырок быстро уменьшается. Поэтому при $V > V_{\text{прок}}$ дырочная компонента тока начинает быстро возрастать с увеличением напряжения. Так как электронная компонента тока при этом остается практически неизменной, ток дырок становится преобладающим. Этот рост тока дырок продолжается до напряжения

$$V_{\text{пз}} = \frac{2\pi q N_d}{\varepsilon} W^2, \quad (6.30)$$

при котором наклон краев зон вблизи положительного контакта становится нулевым (рис. 6.16*б*). Это напряжение называется *напряжением плоских зон*. При $V > V_{\text{пз}}$ величина дырочного тока изменяется



а *б*

Рис. 6.16. Распределение электрического поля и энергетическая диаграмма структуры металл–полупроводник–металл при $V = V_{\text{прок}}$ (*а*) и при $V = V_{\text{пз}}$ (*б*)

с напряжением значительно слабее, так как ток дырок теперь определяется только высотой барьера Шоттки к полупроводнику *p*-типа ϕ_{Bp}, которая уменьшается с ростом напряжения из-за проявления эффекта Шоттки (см. с. 93).

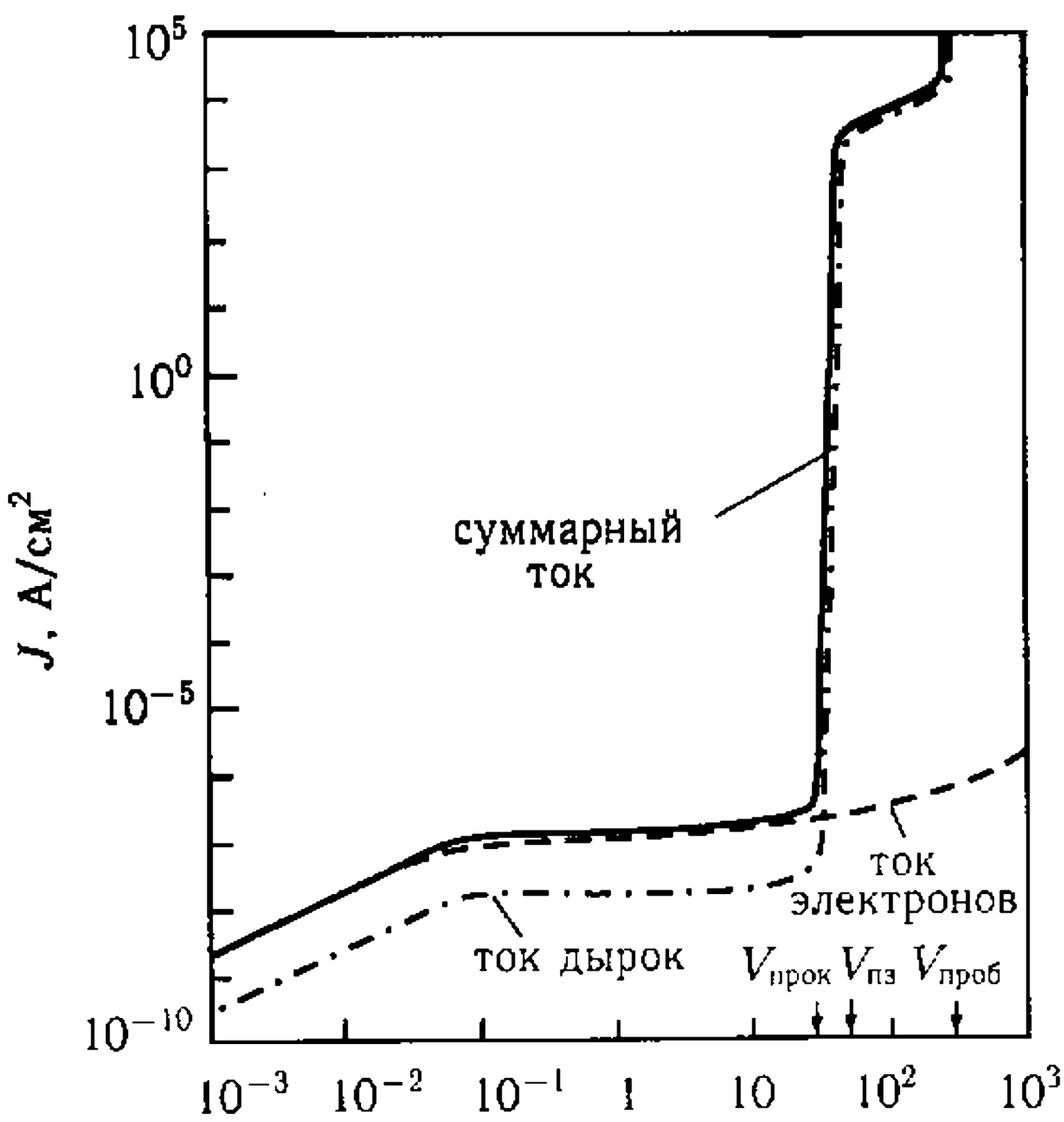


Рис. 6.17. Расчетная вольт-амперная характеристика структуры металл–полупроводник–металл в области прокола [247]

Рассчитаем теперь вклад дырок в вольт-амперную характеристику структуры металл–полупроводник–металл в практически важной области напряжений $V_{\text{прок}} < V < V_{\text{пз}}$ (ниже мы покажем, что для получения высокой частоты генерации необходима достаточно высокая крутизна вольт-амперной характеристики и поэтому ИПД работают именно в этой области напряжений). Поскольку при $V > V_{\text{прок}}$ слой полупроводника полностью обеднен электронами и плотность объемного заряда в нем равна $\rho = qN_d$, то зависимость $E_v(x)$ оказывается параболой, а величина прогиба зон равна

$$q\Delta V = q(V_{\text{пз}} - V)^2/4V_{\text{пз}}.$$

Таким образом, в интересующей нас области напряжений высота потенциального барьера, который должны преодолевать дырки, равна сумме $q\Delta V$ и высоты барьера Шоттки к полупроводнику p-типа ϕ_{Bp} (см. рис. 6.16 *а*). Поэтому в рассматриваемой области напряжений ток в структуре равен [247]

$$J_p \approx A_p^* T^2 e^{-\phi_{Bp}/kT} \exp\left[-\frac{q(V_{\text{пз}} - V)^2}{4kTV_{\text{пз}}}\right], \qquad (6.31)$$

где A_p^* — постоянная Ричардсона (см. с. 91) [1]).

Расчет вольт-амперных характеристик p^+–n–p^+-диодов с проколом показывает, что специфическая зависимость тока от напряжения, подобная (6.31), появляется и там, однако такая зависимость наблюдается, пока концентрация инжектированных дырок мала по сравнению с концентрацией доноров в полупроводнике. При высокой плотности тока объемный заряд инжектируемых дырок начинает заметно влиять на распределение электрического поля в структуре, возникают эффекты ограничения объемным зарядом и вольт-амперная характеристика существенно меняется [248]. В структурах металл–полупроводник–металл из-за того, что ток дырок ограничен высотой барьера ϕ_{Bp}, этого не происходит.

Чтобы понять принцип действия инжекционно-пролетного диода, рассмотрим его работу в режиме большого сигнала. Изменение переменного напряжения и тока в диоде показано на рис. 6.18. Инжекция носителей происходит в моменты времени, когда переменное напряжение на структуре близко к максимальному, после чего носители движутся в области дрейфа. Как следует из рисунка, ток во внешней цепи течет не только когда переменное напряжение на диоде отрицательно, но и часть

[1]) Строго говоря, использование авторами работы [247] формул термоэлектронной эмиссии для описания инжекции дырок в полупроводник не обосновано, так как условие Бете при этом не выполняется. Тем не менее, как показали авторы, в эксперименте наблюдается зависимость, близкая к предсказываемой формулой (6.31).

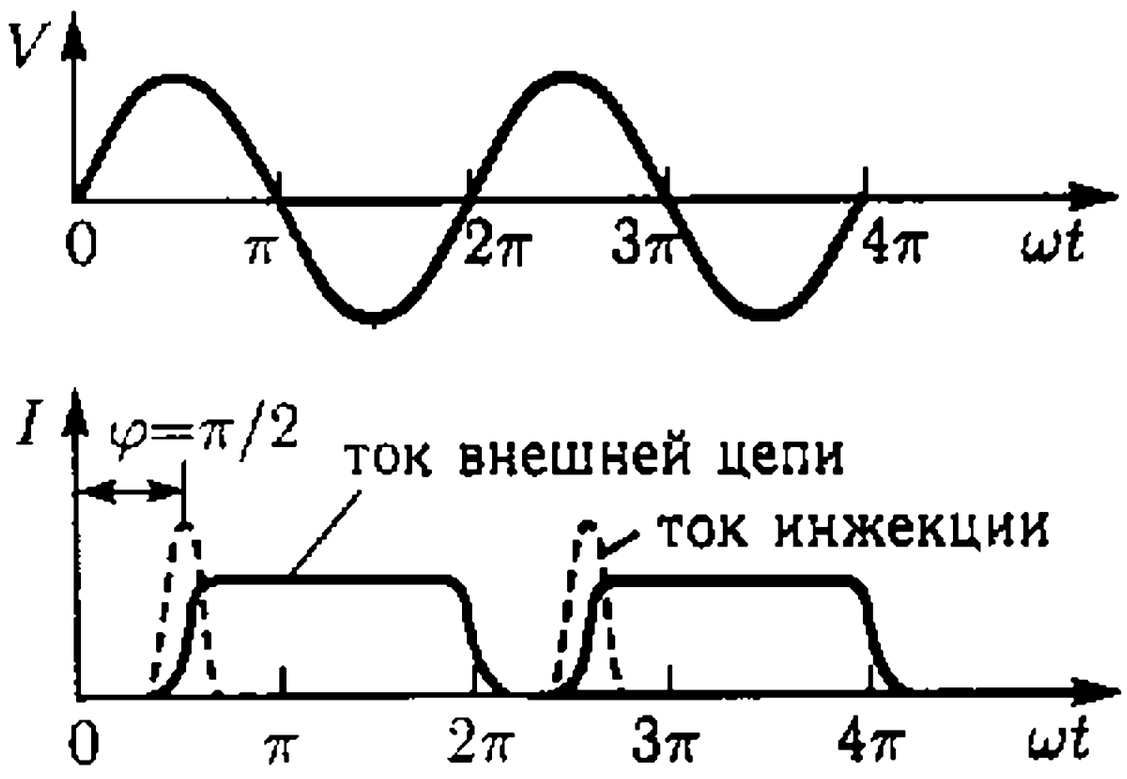


Рис. 6.18. Характер изменения напряжения и тока в инжекционно-пролетном диоде в режиме большого сигнала

периода, когда оно положительно, и поэтому для получения максимального к.п.д. необходимо, чтобы ток протекал в течение всего отрицательного полупериода.

Таким образом, длительность импульса тока τ_d должна составлять три четверти периода колебаний. Считая, что носители в области дрейфа движутся со средней скоростью, близкой к скорости насыщения v_s, можно связать частоту колебаний с толщиной слоя полупроводника W:

$$f \approx \frac{3}{4\tau_d} \approx \frac{3v_s}{4W}. \qquad (6.32)$$

Из-за того, что ток в диоде течет и во время положительного полупериода колебаний, ясно, что к.п.д. инжекционно-пролетного диода должен быть заметно ниже к.п.д. лавинно-пролетного диода.

Малосигнальный импеданс. Импеданс инжекционно-пролетного диода на малом сигнале рассчитывается аналогично импедансу рассмотренного в п. 6.2 лавинно-пролетного диода.

Первые расчеты импеданса ИПД со структурой металл–полупроводник–металл на малом сигнале провел Веллер [249], рассмотревший работу диода в условиях $V > V_{пз}$, то есть когда преодолевшая барьер дырка сразу же попадает в область сильного электрического поля. К сожалению, эти условия сильно отличаются от условий экспериментального наблюдения СВЧ генерации [246]. На самом деле генерация возникает в условиях, когда на энергетической диаграмме в объеме структуры присутствует потенциальный барьер, препятствующий инжекции дырок (см. рис. 6.16*а*), и поэтому ИПД следует рассматривать как структуру, состоящую из трех областей:

1) области перед барьером,

2) области слабого поля, в которой носители дрейфуют со скоростью, меньшей скорости насыщения v_s, и

3) области дрейфа, в которой носители дрейфуют со скоростью v_s.

Аналитически найти совместное решение уравнения непрерывности и уравнения Пуассона во всех трех областях не удается, и здесь можно пользоваться только результатами численного расчета.

Поскольку в окрестности максимума потенциального барьера напряженность электрического поля мала и носители движутся там в основном за счет диффузии, то понятно, что эта область может вносить заметный вклад в задержку распространения. На основании численных расчетов Сьёлунд [250] показал, что задержка распространения при движении в этой области действительно возрастает, но поскольку эффективная скорость движения в ней ненамного ниже скорости насыщения, а толщина области мала, то расчетах можно пренебречь дополнительным замедлением движения носителей в области слабого поля и рассматривать упрощенную модель ИПД, состоящую из области барьера и области дрейфа. Для максимального уменьшения влияния области слабого поля можно использовать конструкцию инжекционно-пролетного диода, предложенную Рюэггом [244], в которой область слабого поля сделана максимально короткой. На рис. 6.19 показано распределение электрического поля при $V > V_{\text{прок}}$ в этой конструкции.

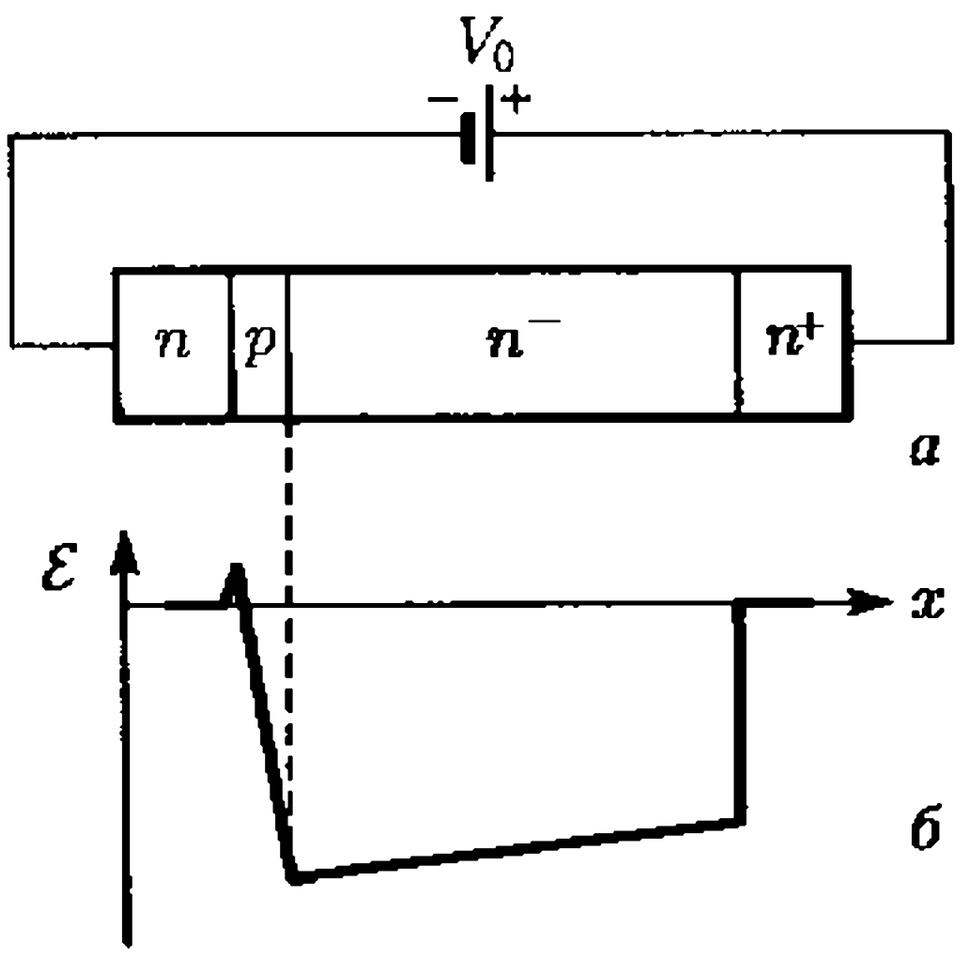


Рис. 6.19. Устройство (*а*) и распределение электрического поля (*б*) в конструкции инжекционно-пролетного диода, предложенной Рюэггом [244]

Сьёлунд [250] показал, что импеданс области барьера может быть аппроксимирован параллельно соединенными дифференциальным сопротивлением диода и барьерной емкостью этой области. Рассчитаем импеданс области барьера для ИПД со структурой металл–полупроводник–металл. Представим напряжение и плотность тока в диоде в виде суммы постоянной составляющей и изменяющейся по гармоническому закону малой добавки:

$$V(t) = V_0 + W\delta\mathcal{E}e^{i\omega t}, \quad J(t) = J_0 + \delta J e^{i\omega t}. \tag{6.33}$$

Здесь мы учли то обстоятельство, что при $V > V_{\text{прок}}$ амплитуда переменного электрического поля связана с амплитудой приложенного к диоду напряжения: $\delta V = W\delta\mathcal{E}$. Подставляя выражение для $V(t)$ в (6.31), найдем переменную составляющую плотности тока дырок δJ_c и рассчитаем отвечающую ей удельную инжекционную проводимость σ:

$$\sigma = \frac{\delta J_c}{\delta\mathcal{E}} = J_0 \frac{q(V_{\text{пз}} - V_0)}{2kTV_{\text{пз}}} \cdot W = J_0 \frac{\varepsilon(V_{\text{пз}} - V_0)}{4\pi kTN_d W}. \tag{6.34}$$

Из этой формулы следует, что σ немонотонно зависит от напряжения, достигая максимума при $V = V_{пз} - \sqrt{(2kT/q)V_{пз}}$. Поскольку полная плотность тока диода равна сумме плотностей тока проводимости δJ_c и тока смещения δJ_d, импеданс области барьера единичной площади равен

$$Z_b = \frac{\delta \mathcal{E}}{\delta J_c + \delta J_d} = \frac{1}{\sigma + i\omega\epsilon/4\pi}, \tag{6.35}$$

а отношение переменных составляющих инжекционного и полного токов $\gamma = \delta J_c/\delta J$ равно

$$\gamma = \frac{\delta J_c}{\delta J_c + \delta J_d} = \frac{\sigma}{\sigma + i\omega\epsilon/4\pi} = \frac{1}{1 + i(\omega/\omega_b)}, \tag{6.36}$$

где через ω_b обозначена характеристическая частота, равная $4\pi\sigma/\epsilon$.

Импеданс области дрейфа Z_d уже был рассчитан нами в п. 6.2 (формула (6.19)). Поскольку в конструкции ИПД толщина области барьера (x_R на рис. 6.16 *а*) намного меньше толщины области дрейфа $(W - x_R)$, то полный импеданс диода $Z = Z_b + Z_d$ можно приближенно считать равным импедансу области дрейфа. Подставляя в уравнение (6.19) условие (6.35), определяющее фазовый угол инжекции, в итоге получаем для диода единичной площади:

$$Z \approx \frac{4\pi(W - x_R)}{i\omega\epsilon}\left[1 - \frac{1}{1 + i(\omega/\omega_b)}\frac{1 - e^{-i\vartheta}}{i\vartheta}\right], \tag{6.37}$$

где $\vartheta = \omega(W - x_R)/v_s$ — угол пролета в области дрейфа. Анализ формулы (6.37) показывает, что действительная часть комплексного сопротивления становится отрицательной, если угол пролета одновременно удовлетворяет соотношениям

$$\pi < \vartheta < 2\pi \quad \text{и} \quad (\omega/\omega_b) > (1 - \cos\vartheta)/|\sin\vartheta|.$$

Последнее условие показывает, что запаздывание при инжекции носителей принципиально важно для работы ИПД, поскольку без него невозможно получить отрицательное дифференциальное сопротивление [1].

---

[1]) Дело в том, что единственной компонентой тока, способной к созданию отрицательного дифференциального сопротивления, является инжекционная

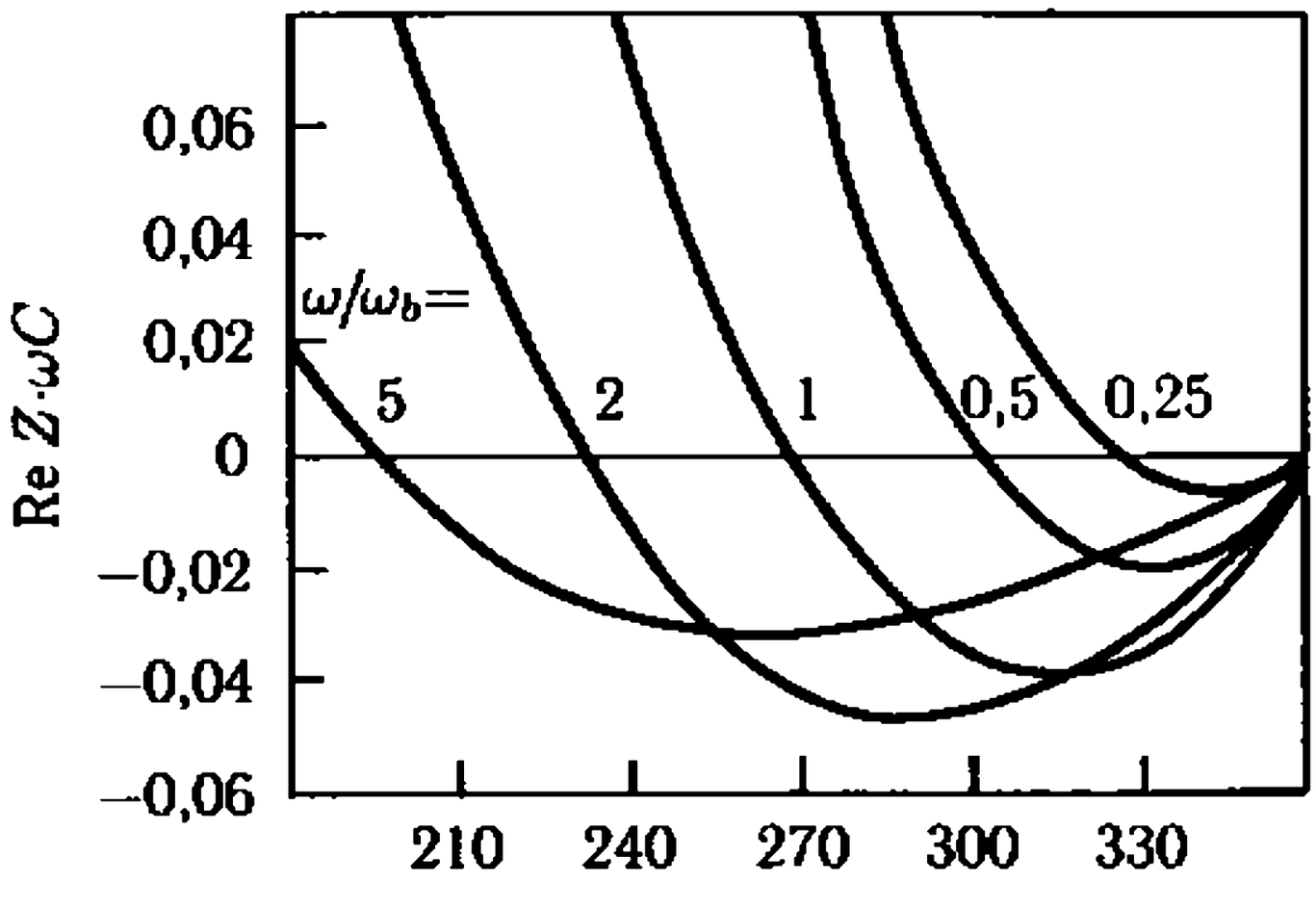


Рис. 6.20. Зависимость действительной части комплексного сопротивления структуры металл–полупроводник–металл от угла пролета ϑ для нескольких значений отношения ω/ω_b [249]

Расчетные зависимости величины действительной части Z как функции угла пролета для различных значений отношения ω/ω_b показаны на рис. 6.20. Наибольшее отрицательное значение $\mathrm{Re}\,Z$ достигается при $\omega/\omega_b \approx 1,88$ и $\vartheta \approx 292°$. Поскольку ω_b изменяется пропорционально удельной инжекционной проводимости σ, то для обеспечения возможности генерации на высоких частотах инжекционно-пролетные диоды должны работать на наиболее крутом участке вольт-амперной характеристики. Для получения высокого значения σ в структурах металл–полупроводник–металл контакты необходимо изготавливать из материалов, характеризуемых невысокими значениями ϕ_{Bp}, например PtSi ($\phi_{Bp} \approx 0,2$ эВ).

Как следует из малосигнального анализа, наряду с отрицательным дифференциальным сопротивлением импеданс ИПД характеризуется большой реактивной составляющей емкостного типа, которая при работе диода в качестве генератора колебаний должна быть скомпенсирована импедансом резонатора. Это усложняет согласование диода с резонатором и создает серьезные препятствия продвижению в область более высоких частот.

компонента. Волны объемного заряда, распространяющиеся в структуре, могут либо увеличивать, либо уменьшать мощность высокочастотных колебаний в зависимости от соотношения между фазой тока носителей и фазой локального переменного электрического поля. Оптимальным для возбуждения колебаний является фазовый сдвиг между током инжекции и напряженностью поля в 180°, который возникает в ЛПД при $\omega > \omega_r$ и обеспечивает этим диодам наибольшую эффективность преобразования.

Основными источниками шума в ИПД являются дробовой шум и диффузионный шум, связанный с флуктуациями скорости дрейфа носителей. По сути, природа шума здесь та же, что и диодах Ганна. Минимальный коэффициент шума, достигнутый в инжекционно-пролетных диодах из Si, составляет 10 дБ на частоте 7 ГГц, то есть спектральная плотность мощности шума в ИПД на три порядка ниже, чем в кремниевых ЛПД [14].

Инжекционно-пролетные диоды находят применение в качестве гетеродинов СВЧ приемников, в доплеровских радарных системах, охранных системах и для организации локальной связи. Максимальный теоретический к.п.д. ИПД составляет 21%, когда носители инжектируются в момент максимального напряжения на диоде [244], однако реально достигнутые к.п.д. не превышают 5%. Выходная мощность ИПД примерно на два порядка ниже, чем для ЛПД. Для увеличения к.п.д. и выходной мощности диодов можно попытаться увеличить угол инжекционного запаздывания [14]. Максимальная рабочая частота инжекционно-пролетных диодов составляет ~20 ГГц.

6.4. Другие способы генерации СВЧ колебаний

6.4.1. TRAPATT-режим. Режим генерации лавинно-пролетных диодов, названный режимом с захватом объемного заряда лавины [1]), был экспериментально обнаружен Прагером и др. из фирмы RCA [251]. Этот режим был назван «аномальным», поскольку он характеризуется:

1) частотой колебаний, которая в 2–10 раз ниже частоты нормальных колебаний ЛПД (6.27), и

2) необычно высоким к.п.д., по крайней мере вдвое превышающим теоретический к.п.д. лавинно-пролетного диода. Суть этого режима работы проще всего понять на примере p–i–n-диода [52].

Пусть на p–i–n-диод подано постоянное обратное смещение, близкое к напряжению пробоя. Если теперь на диод извне подать дополнительный импульс напряжения, достаточный для возникновения пробоя, то в объеме диода начнет развиваться процесс ударной ионизации, диод быстро перейдет в проводящее состояние и напряжение на диоде (при ограниченном токе в цепи) сильно уменьшится. Особенностью проводящего состояния явля-

[1]) В зарубежной литературе этот режим называют TRAPATT (trapped plasma avalanche triggered transit time).

ется то, что высокая концентрация электронов и дырок (плазма), возбужденная в процессе ударной ионизации, экранирует электрическое поле в объеме p–i–n-диода. Поэтому скорость дрейфа носителей в этой области невелика и они оказываются как бы «захваченными» в объеме структуры. Протекающий через диод ток вызывает экстракцию этих зарядов, которая продолжается до тех пор, пока все носители из объема диода не будут выведены, после чего диод возвращается в непроводящее состояние. Из-за низкой напряженности поля в объеме диода в состоянии с захваченной плазмой время восстановления может быть намного больше времени пролета.

Если такой p–i–n-диод подключить к короткозамкнутой длинной линии, то в схеме возникнут периодические колебания, поскольку отражающиеся в конце линии инвертированные импульсы напряжения через время задержки, определяемое длиной линии, будут возвращаться к диоду и вновь «запускать» его.

Аналогичный режим колебаний возникает и в ЛПД. Для возбуждения этого режима необходимо в момент «запуска» пропускать через диод большой ток с тем, чтобы фронт распространяющейся в образце волны ударной ионизации (пик напряженности электрического поля) не затухая проходил через область дрейфа и возбуждал во всем ее объеме высокую концентрацию электронов и дырок, необходимую для образования захваченной плазмы. Эти условия можно создать в специальном резонаторе, настроенном в резонанс на частоту, близкую к пролетной. Большая амплитуда колебаний на этой частоте обеспечивает быстрое нарастание напряжения и возможность получения большого тока в момент «запуска» при том, что основная мощность колебаний выделяется на одной из субгармоник пролетной частоты [252].

Интерес к TRAPATT-режиму связан прежде всего с тем, что в этом режиме диод работает почти как идеальный ключ, что позволяет ожидать от него очень высокого к.п.д. генерации. В этом режиме уже достигнут к.п.д. в 75% на частоте 600 МГц и 43% на частоте 8 ГГц [14]. При этом мощность колебаний, получаемая на одном диоде, достигает 400 Вт. Диоды, работающие в TRAPATT-режиме, используются для возбуждения фазированных антенных решеток. Этому режиму работы присущи два недостатка: очень высокий уровень шума, связанный с флуктуациями момента «запуска» диода, и то, что из-за высоких плотностей тока диод обычно работает только в импульсном режиме. Диапазон частот, перекрываемых диодами, работающи-

ми в TRAPATT-режиме, простирается от 600 МГц примерно до 10 ГГц.

6.4.2. Туннельно-пролетные диоды (TUNNETT). Как мы отмечали в п. 6.2, с ростом частоты генерации оптимальная толщина области лавинного умножения в ЛПД уменьшается, а напряженность электрического поля в p–n-переходе растет, и при попытке создать диод с очень высокой частотой генерации ток в p–n-переходе начинает определяться туннельным пробоем. При этом, поскольку туннелирование является очень быстрым процессом, разность фаз между током инжекции и напряжением на области пробоя (фазовый угол инжекции) становится близкой к нулю. Такие диоды принято называть туннельно-пролетными диодами. Первым изменение характеристик лавинно-пролетных диодов при учете туннелирования обсуждал Рид [230], а затем Аладинский [253] и Нишидзава [254].

Задача создания полупроводниковых СВЧ приборов с низким уровнем шума требует отказа от использования явления лавинного пробоя. Рассмотренные в п. 6.3 инжекционно-пролетные диоды не используют явление лавинного пробоя, однако их сравнительно невысокое быстродействие связано с тем, что в ИПД инжектированные носители часть своего пути в районе максимума потенциального барьера проходят с невысокой скоростью. В туннельно-пролетных диодах этого ограничения нет, поскольку протуннелировавшие носители сразу же попадают в область сильного электрического поля. Это позволяет создать низкошумящие генераторы СВЧ колебаний в миллиметровой и субмиллиметровой областях длин волн [254].

Как показали Нишидзава и Мотоя [255], все преимущества туннельно-пролетных диодов можно реализовать и в режиме, когда пробой p–n-перехода происходит с одновременным участием ударной ионизации и туннельного пробоя. Такие приборы уже имеют низкий уровень шума, а за счет еще достаточно большого фазового угла инжекции позволяют получать сравнительно высокую выходную мощность (30 мВт на частоте 150 ГГц в импульсном режиме).

Еще одной особенностью TUNNETT-диодов, обсуждаемой пока только теоретически (см., например, [256]), является возможность пролета протуннелировавших носителей в области дрейфа в *баллистическом режиме*, то есть со скоростями, примерно на порядок превышающими скорость насыщения. При сокращении толщины области дрейфа это позволяет ожидать существенного увеличения быстродействия туннельно-пролетных

диодов, которое, по оценкам, может достигать нескольких терагерц.

В качестве достижений в области создания TUNNETT-диодов укажем получение в диодах из GaAs в непрерывном режиме выходной мощности, равной 70 мВт (с к.п.д. около 5%) на частоте 105 ГГц [257], мощности 50 мкВт на частоте 322 ГГц [258] (на основной гармонике) и такой же мощности при выделении третьей гармоники на частоте 356 ГГц. В приборах из кремния еще в 1968 г. была получена генерация на частоте 256 ГГц с выходной мощностью 1 мВт в импульсном режиме [254].

6.4.3. QWITT-диоды. Исследования резонансно-туннельных диодов (РТД), принцип действия которых был рассмотрен нами в п. 1.6.2, показали, что наиболее высокие параметры этих диодов получаются при использовании широких спейсеров — областей, располагающихся с двух сторон двухбарьерной структуры. Это наводило на мысль, что эти слои играют важную роль в работе этих диодов, и поскольку слой со стороны анода обеднен носителями, то там могут проявляться пролетные эффекты. Последующие исследования показали, что именно так и происходит, и специально созданные структуры, состоящие из резонансно-туннельного диода и области дрейфа, получили название QWITT-диодов (QWITT = quantum well injection transit-time).

Особенностью QWITT-диодов является возможность их работы как на нарастающем, так и на падающем участке вольт-амперной характеристики резонансно-туннельного диода. Малосигнальный анализ, проведенный в работе [259], показал, что область дрейфа практически всегда оказывает определяющее влияние на величину отрицательного дифференциального сопротивления и максимальную частоту генерации этих диодов. Работа диода в случае, когда РТД работает на нарастающем участке вольт-амперной характеристики, по сути ничем не отличается от режима работы ИПД (см. п. 6.3).

Совершенно иная ситуация возникает, когда рабочая точка находится на падающем участке вольт-амперной характеристики резонансно-туннельного диода. В этом случае, хотя параметры QWITT-диода продолжают определяться областью дрейфа, оптимальная толщина области дрейфа сокращается примерно в 3 раза, что позволяет соответственно увеличить максимальную частоту генерации. Авторы [259] показали, что продвижению в область более высоких частот препятствуют конечное удельное сопротивление контактов к структуре и конечная величина

скорости насыщения. Учет эффекта всплеска скорости дрейфа и возможности баллистического переноса, по-видимому, позволит в дальнейшем расширить рабочий диапазон частот этих приборов. В связи с вышесказанным следует отметить, что по сути именно в QWITT-диодах была получена наиболее высокая частота генерации РТД — 712 ГГц [94].

Оптимизация размеров области дрейфа позволяет заметно увеличить отрицательное дифференциальное сопротивление резонансно-туннельного диода. Эксперименты показали, что при добавлении в структуру РТД области дрейфа толщиной 1000–2000 Å выходная мощность и к.п.д. действительно заметно увеличиваются. В практически реализованных приборах на основе гетероструктуры InGaAs/AlAs при работе в резонаторе, частота которого изменялась в пределах 1–10 ГГц, к.п.д. генератора на частоте 2,4 ГГц достигал 50% [260]. Выходная мощность при этом составляла 20 мВт.

Глава 7

ОПТОЭЛЕКТРОННЫЕ ПРИБОРЫ

Оптоэлектронные приборы — это приборы, работа которых основана на взаимодействии электромагнитного излучения с полупроводником. В этой главе мы рассмотрим два наиболее важных класса этих приборов — приемники и источники излучения.

7.1. Приемники излучения

7.1.1. Механизмы поглощения излучения в полупроводнике. Регистрация электромагнитного излучения в диапазоне от миллиметровых волн до жесткого рентгеновского излучения основана на поглощении этого излучения полупроводником, в результате которого в последнем происходит изменение концентрации или подвижности свободных носителей. Чтобы понять, к каким конкретным изменениям может приводить поглощение квантов той или иной энергии, рассмотрим основные механизмы поглощения электромагнитного излучения в полупроводнике.

На рис. 7.1 показаны спектры поглощения полупроводника в различных диапазонах энергий кванта как выше, так и ниже края собственного поглощения, который находится при энергии кванта $\hbar\omega \approx E_g$. Основным параметром, характеризующим поглощение электромагнитной волны с частотой ω в полупроводнике, является *коэффициент поглощения* $\alpha(\omega)$, который описывает уменьшение энергии электромагнитной волны $I(x)$ с увеличением глубины x ее проникновения в образец [261]:

$$I(x) = I(0)\, e^{-\alpha x}. \qquad (7.1)$$

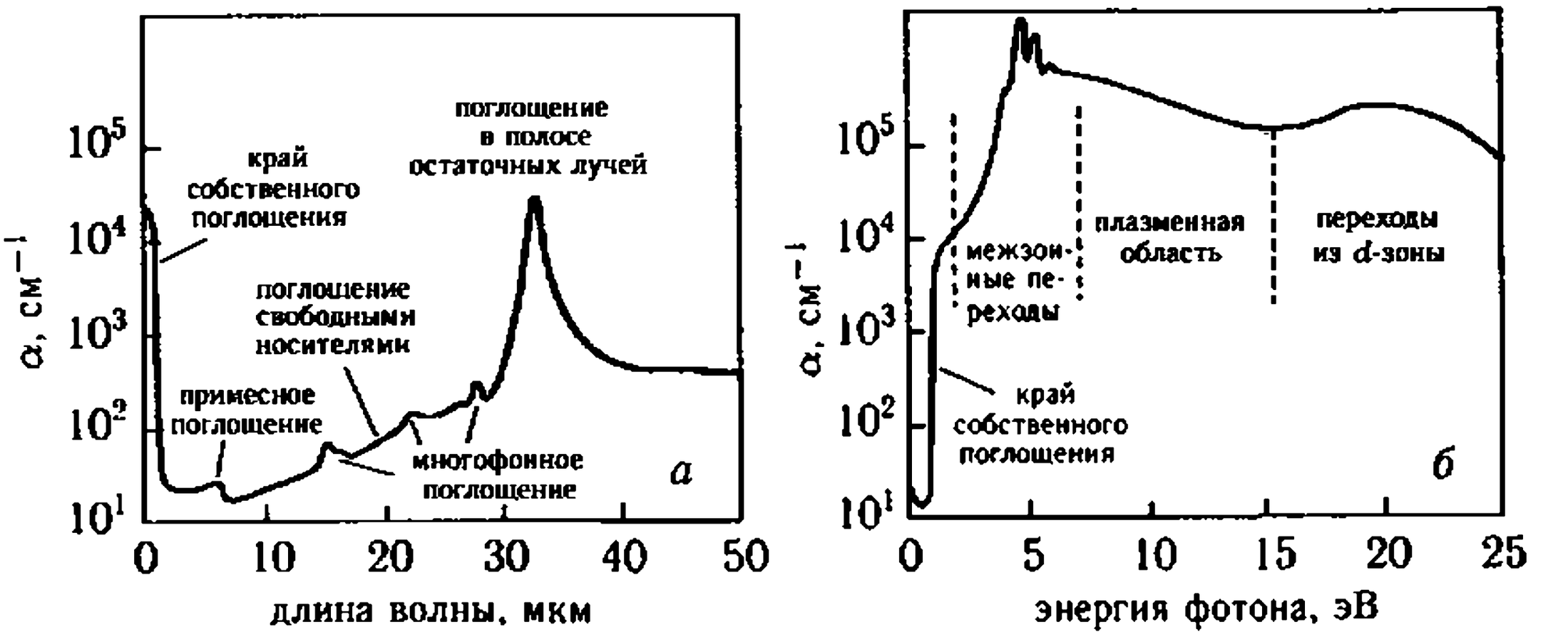


Рис. 7.1. Спектры поглощения полупроводника в инфракрасной (*а*) и ультрафиолетовой (*б*) областях спектра

Известно, что при поглощении излучения тела нагреваются. Если телом является полупроводник, сопротивление которого сильно зависит от температуры, то поглощение излучения будет приводить к изменению сопротивления образца, которое можно легко зарегистрировать. *Болометры* — тепловые приемники излучения, работа которых основана на измерении изменений сопротивления [262], — конструируются так, чтобы они практически полностью поглощали падающее на них излучение и поэтому они имеют практически постоянную чувствительность во всем диапазоне электромагнитного спектра. Недостатками болометров являются их низкая чувствительность и высокая инерционность (0,01–0,1 с).

Поглощение электромагнитного излучения в области *собственного поглощения* ($\hbar\omega \geqslant E_g$, см. рис. 7.1 *б*) сопровождается возбуждением электронно-дырочных пар и появлением неравновесных электронов и дырок. Наблюдаемое при поглощении света изменение проводимости полупроводника называют явлением *фотопроводимости* (фоторезистивным эффектом). Фотопроводимость, возникающую при поглощении света в области собственного поглощения, называют *собственной фотопроводимостью*. В отличие от болометрического эффекта, фотоотклик в эффекте фотопроводимости обычно является быстрым и определяется временем жизни неравновесных носителей в полупроводнике [1]).

---

[1]) Вопрос о кинетике фотопроводимости в полупроводниках с несколькими примесными уровнями, когда могут проявляться эффекты прилипания носителей, будет рассмотрен нами на с. 391.

При энергиях кванта $\hbar\omega \gg E_g$ в поглощении начинают проявляться электронные переходы из глубоко лежащих валентных зон или даже с остовных оболочек составляющих полупроводник атомов. Важность этих механизмов поглощения для первичных возбужденных носителей часто оказывается настолько высокой, что при своем движении в образце эти носители начинают сами ионизовать полупроводник, вызывая появление целого каскада вторичных электронно-дырочных пар. Поэтому *квантовый выход фотоионизации* η, то есть число электронно-дырочных пар, возбуждаемых одним квантом электромагнитного излучения, в случае квантов высокой энергии может заметно превышать единицу. В этом случае квантовый выход может быть оценен с помощью эмпирической формулы $\eta \approx \hbar\omega/3E_g$ [108]. Механизмы поглощения излучения в рентгеновской области спектра будут более подробно рассмотрены нами в п. 7.1.7.

В области спектра $\hbar\omega < E_g$ (см. рис. 7.1 *а*) энергия кванта меньше энергетического зазора между краем валентной зоны и зоной проводимости полупроводника и поэтому ее недостаточно, чтобы создать электронно-дырочную пару. Электронные переходы в этой области спектра возможны либо между энергетическим уровнем в запрещенной зоне и одной из разрешенных зон (*примесное поглощение*), либо внутри одной из зон (*поглощение на свободных носителях*). Примесное поглощение, при котором в одной из разрешенных зон появляются неравновесные носители, также приводит к увеличению проводимости образца (фотопроводимости); в данном случае, поскольку неравновесные носители возникают в результате примесного поглощения, фотопроводимость называют *примесной фотопроводимостью*.

Поглощение света на свободных носителях, при котором энергия квантов идет на разогрев носителей в разрешенных зонах, также может проявляться в изменении проводимости образца, поскольку подвижность носителей зависит от их энергии. Так как концентрация носителей заряда в полупроводнике при этом остается неизменной, а изменение проводимости целиком обусловлено изменением подвижности, этот механизм фотопроводимости часто называют *μ-фотопроводимостью* [263]. Хотя величина этого эффекта невелика, он может быть использован для регистрации СВЧ излучения и излучения в дальней инфракрасной области спектра, поскольку он является единственным нетепловым (а, следовательно, быстрым по времени отклика) способом регистрации излучения. Характерное время отклика

μ-фотопроводимости равно времени релаксации энергии в полупроводнике.

Поглощение в полосе остаточных лучей (на оптических фононах) и многофононное поглощение (см. рис. 7.1 *а*) происходит с прямой передачей энергии электромагнитной волны колебаниям решетки и не сопровождается ни возбуждением неравновесных носителей, ни разогревом носителей. Поэтому эти механизмы поглощения могут приводить только к болометрическому эффекту (см. выше).

7.1.2. Фотоприемники, основанные на явлении фотопроводимости. Фотоприемник, основанный на явлении фотопроводимости, — *фотосопротивление* (*фоторезистор*) — представляет собой пластинку или тонкую пленку полупроводника, на противоположных краях которой созданы омические контакты (см. рис. 7.2). При освещении пластинки светом в полупроводнике в результате примесного или собственного поглощения возбуждаются неравновесные носители заряда. В результате увеличения концентрации свободных носителей проводимость образца возрастает и это может быть зарегистрировано с помощью схемы, показанной на рис. 7.2.

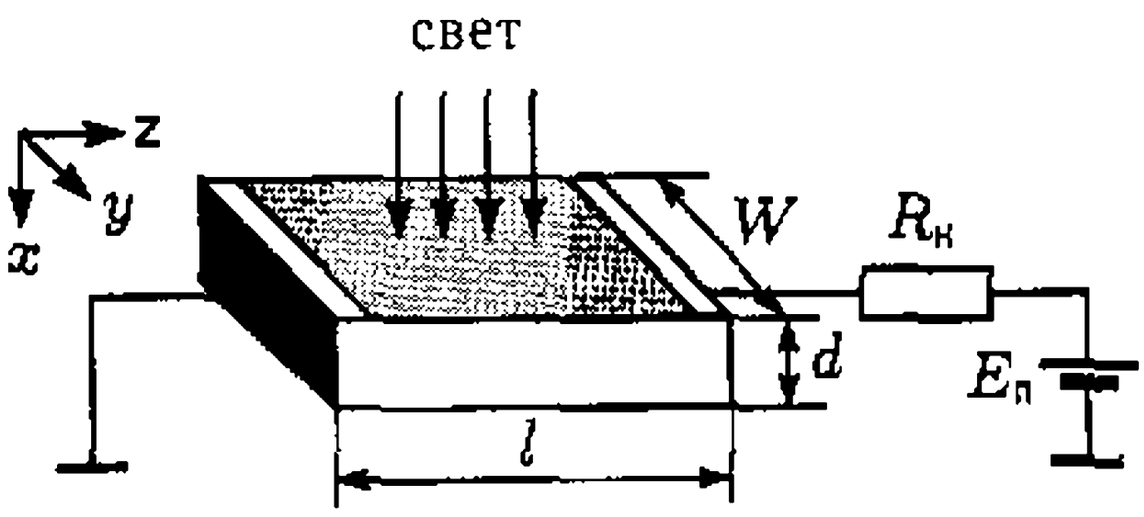


Рис. 7.2. Схема включения образца полупроводника в электрическую цепь для наблюдения фотопроводимости

Найдем условия получения максимального сигнала фотоотклика в схеме, приведенной на рис. 7.2. Падение напряжения на фоторезисторе, очевидно, равно

$$U = E_п \frac{R_{обр}}{R_{обр} + R_н} = E_п \frac{1}{1 + \sigma R_н}, \qquad (7.2)$$

где $R_{обр}$ — сопротивление образца, $\sigma = 1/R_{обр}$ — его проводимость, а $R_н$ — сопротивление нагрузки. Если при освещении образца его проводимость возрастает на величину $\Delta\sigma$, то отношение изменения напряжения на образце к вызвавшему его малому изменению проводимости составит

$$\frac{\Delta U}{\Delta \sigma} \approx -E_п \frac{R_н}{(1 + \sigma R_н)^2}. \qquad (7.3)$$

Условие оптимального согласования фотосопротивления с нагрузкой, при котором отношение $\Delta U/\Delta\sigma$ максимально, можно найти, приравняв нулю производную выражения (7.3) по $R_н$. Нетрудно убедиться, что

оптимальное согласование достигается при $R_н = R_{обр}$, при котором $\Delta U/\Delta\sigma = -E_п R_н/4$. Из последней формулы следует, что в условиях оптимального согласования напряжение сигнала на выходе схемы связано с изменением проводимости образца уравнением

$$|\Delta U| = \frac{E_п}{4}\frac{\Delta\sigma}{\sigma_0},$$

где через σ_0 обозначена темновая проводимость образца.

Задача создания максимально чувствительного фотоприемника состоит в нахождении такой геометрии прибора и выборе параметров полупроводника, при которых при заданном потоке излучения получается максимальное отношение $\Delta\sigma/\sigma_0$. Чтобы найти эти условия, рассчитаем распределение возбуждаемых светом неравновесных носителей в объеме полупроводника.

Если на интересующей нас длине волны коэффициент поглощения в полупроводнике равен α, то темп объемной генерации в пластинке полупроводника (см. рис. 7.2) на глубине x равен произведению квантового выхода фотоионизации η на отнесенное к единице длины число поглощенных квантов:

$$G(x) = \eta \cdot \left(-\frac{dI(x)}{dx}\right) = I_0(1-R)\eta\alpha e^{-\alpha x}. \qquad (7.4)$$

При выводе этой формулы из (7.1) мы учли, что падающий на поверхность полупроводника поток квантов плотностью I_0 (в квантах на квадратный сантиметр в секунду) частично отражается от границы раздела (R — коэффициент отражения). Поскольку в образце излучение поглощается неоднородно, для расчета пространственного распределения неравновесных носителей нам надо решить уравнение непрерывности (1.21).

Рассмотрим сначала случай стационарного (не зависящего от времени) монополярного возбуждения электронов (примесная фотопроводимость). В этом случае уравнение непрерывности сводится к следующему дифференциальному уравнению:

$$\frac{\partial(\delta n)}{\partial t} = G(x) - \frac{\delta n}{\tau_n} + D_n\frac{\partial^2(\delta n)}{\partial x^2} = 0. \qquad (7.5)$$

В простейшем (идеализированном) случае, когда поверхностной рекомбинацией носителей можно пренебречь, граничные условия на обеих поверхностях пластины ($x = 0$ и $x = d$) имеют вид

$$D_n\frac{\partial(\delta n)}{\partial x} = 0. \qquad (7.6)$$

Решение уравнения (7.5) будем искать в виде

$$\delta n(x) = Ae^{-\alpha x} + Be^{-x/L_n} + Ce^{x/L_n}, \qquad (7.7)$$

где $L_n = \sqrt{D_n \tau_n}$ — диффузионная длина электронов. Подставляя это решение в (7.5) и учитывая (7.4) и (7.6), находим:

$$A = \frac{I_0(1-R)\eta\alpha\tau_n}{1-\alpha^2 L_n^2}, \quad B = A\alpha L_n \frac{e^{-\alpha d} - e^{d/L_n}}{e^{d/L_n} - e^{-d/L_n}},$$
$$C = A\alpha L_n \frac{e^{-\alpha d} - e^{-d/L_n}}{e^{d/L_n} - e^{-d/L_n}}. \qquad (7.8)$$

В предельных случаях (*а*) очень сильного ($\alpha d, \alpha L_n \gg 1$) и (*б*) очень слабого ($\alpha d, \alpha L_n \ll 1$) поглощения распределение носителей по глубине имеет следующий вид:

$$\delta n(x) \approx \frac{I_0(1-R)\eta\tau_n}{L_n} \frac{\operatorname{ch}\left(\dfrac{d-x}{L}\right)}{\operatorname{sh}\left(\dfrac{d}{L_n}\right)}, \qquad (7.9)$$

$$\delta n(x) \approx I_0(1-R)\,\eta\alpha\tau_n\, e^{-\alpha x}.$$

Видно, что характерная длина, на которой концентрация неравновесных носителей изменяется в e раз, составляет L_n в случае (*а*) и $1/\alpha$ в случае (*б*).

При собственном поглощении, когда свет возбуждает носители обоих знаков, для расчета пространственного распределения неравновесных носителей необходимо совместно решить уравнения непрерывности для электронов и дырок, в которых следует учесть существование электрического поля (поля Дембера), возникающего из-за различия коэффициентов диффузии двух типов носителей. Эта задача по сути уже рассматривалась нами в п. 1.2.1 и мы знаем, что ее решение при $\tau_n = \tau_p$ формально сводится к решению уравнения (7.5), в котором вместо D_n стоит коэффициент амбиполярной диффузии D_a (1.25). Это позволяет использовать полученные выше решения (с заменой D_n на D_a) и в случае собственного поглощения при $\tau_n = \tau_p$.

Теперь, зная профиль распределения концентраций носителей по глубине образца, несложно рассчитать его полную

проводимость:

$$\sigma = q\frac{W}{l}\left[\mu_n \int_0^d n(x)\,dx + \mu_p \int_0^d p(x)\,dx\right] =$$

$$= q(\mu_n n_0 + \mu_p p_0)\frac{Wd}{l} + q\frac{W}{l}\left[\mu_n \int_0^d \delta n(x)\,dx + \mu_p \int_0^d \delta p(x)\,dx\right], \tag{7.10}$$

где W — ширина пластинки, l — ее длина, а d — ее толщина. Первое слагаемое представляет собой темновую проводимость образца σ_0, а второе слагаемое — изменение проводимости $\Delta\sigma$, вызванное освещением.

Интегрирование второго слагаемого в уравнении (7.10) для двух предельных случаев сильного собственного и слабого примесного поглощения приводит к следующим выражениям:

$$\Delta\sigma = q(\mu_n\tau_n + \mu_p\tau_p)\frac{W}{l} I_0(1-R)\,\eta, \tag{7.11}$$

$$\Delta\sigma = q\mu_n\tau_n\frac{W}{l} I_0(1-R)\,\eta(1-e^{-\alpha d}).$$

Смысл дополнительного сомножителя, который появляется в случае слабо поглощаемого излучения, вполне понятен: это — доля фотонов, поглощенных в образце.

Теперь, вычисляя $\Delta\sigma/\sigma_0$, для двух рассматриваемых случаев находим:

$$\frac{\Delta\sigma}{\sigma_0} = \frac{I_0(1-R)\eta}{d}\cdot\frac{\mu_n\tau_n + \mu_p\tau_p}{\mu_n n_0 + \mu_p p_0}, \tag{7.12}$$

$$\frac{\Delta\sigma}{\sigma_0} = \frac{I_0(1-R)\eta(1-e^{-\alpha d})}{n_0 d} \approx I_0(1-R)\frac{\eta\tau_n}{n_0}\alpha.$$

Из этих уравнений следует, что для достижения максимальной чувствительности фотосопротивлений, работающих в области собственного поглощения, необходимо уменьшать толщину пластинки до величины порядка $1/\alpha$ и изготавливать ее из материала с проводимостью, близкой к собственной (то есть максимально уменьшать σ_0). Фотосопротивления на основе примесной фотопроводимости также следует изготавливать из высокоомных полупроводников, но так как $\Delta\sigma/\sigma_0$ не зависит от d, то выбор толщины этих фотоприемников определяется другими факторами

(например, уровнем шума). В обоих случаях время жизни неравновесных носителей должно быть максимально высоким.

Рассмотрим теперь влияние поверхностной рекомбинации, которая всегда существует в реальных полупроводниках. Учесть поверхностную рекомбинацию можно, записав граничные условия для уравнения (7.5) на верхней и нижней поверхностях образца следующим образом:

$$D_n \frac{\partial(\delta n)}{\partial x} = s_1 \delta n(0), \quad -D_n \frac{\partial(\delta n)}{\partial x} = s_2 \delta n(d), \tag{7.13}$$

где s_1 и s_2 — скорости поверхностной рекомбинации на соответствующих поверхностях.

Мы не будем приводить здесь решение уравнения (7.5) с граничными условиями (7.13) из-за его громоздкости [1]) и ограничимся качественным обсуждением происходящего. Изменение профиля распределения электронов с увеличением скорости поверхностной рекомбинации в толстом кристалле ($d > L_n$) показано на рис. 7.3.

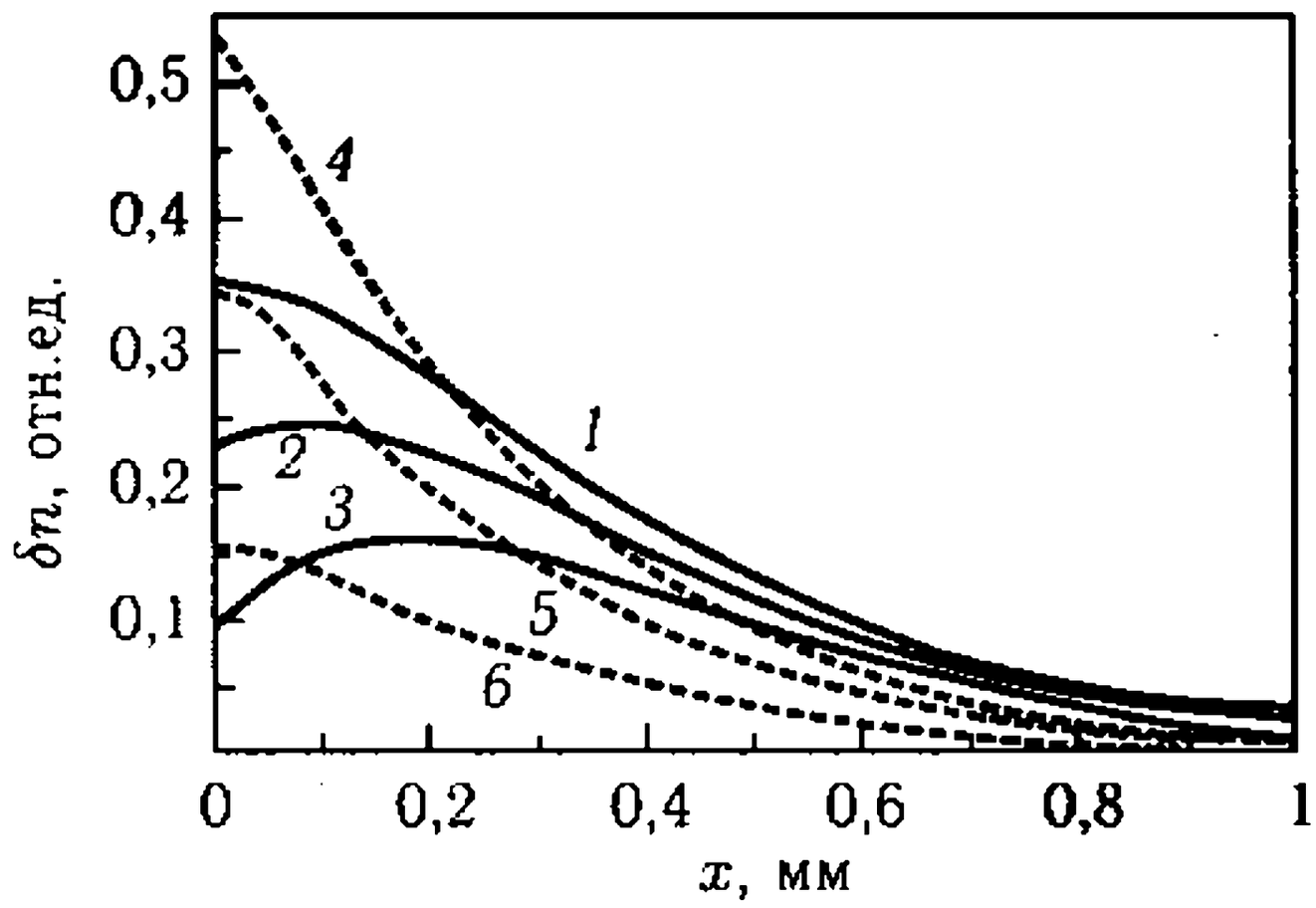


Рис. 7.3. Распределение неравновесных носителей по глубине пластины из p-Ge толщиной 1 мм при возбуждении светом с двумя разными длинами волн при трех скоростях поверхностной рекомбинации. s: *1*, *4* — 0; *2*, *5* — 2000; *3*, *6* — 10^4 см/с. α: *1*–*3* — 50 см$^{-1}$; *4*–*6* — 500 см$^{-1}$. $L_n = 400$ мкм

Видно, что поверхностная рекомбинация уменьшает концентрацию неравновесных носителей не только вблизи поверхностей, но и во всем объеме образца.

В тонком ($d \ll L_n$) образце влияние поверхностной рекомбинации на фотопроводимость в условиях объемного возбуждения (низкий коэффициент поглощения) может быть описано путем введения *эффективного времени рекомбинации*, величина

[1]) Решение этого уравнения можно найти в работах [261, 263].

которого при равных скоростях рекомбинации на верхней и нижней поверхностях образца ($s = s_1 = s_2$) вычисляется по формуле [1]

$$\frac{1}{\tau_{эф}} = \frac{1}{\tau_n} + \frac{2s}{d}. \qquad (7.14)$$

Этой же формулой можно воспользоваться и в случае сильного поглощения при условии, что $s \ll L_n/\tau_n$ [263]. Из этой формулы следует, что при учете поверхностной рекомбинации в тонких образцах эффективное время рекомбинации $\tau_{эф} \sim d$ и, следовательно, поверхностная рекомбинация сильно уменьшает рассчитываемые по формулам (7.12) значения $\Delta\sigma/\sigma_0$. В частности, в случае сильно поглощаемого излучения при учете поверхностной рекомбинации отношение $\Delta\sigma/\sigma_0$ перестает зависеть от толщины d. Поэтому при конструировании фотоприемников, работающих в области собственного поглощения, нет необходимости создавать очень тонкую пластинку и ее толщина обычно выбирается равной нескольким L_n. Для фотосопротивлений, работающих в области примесного поглощения, для уменьшения влияния поверхностной рекомбинации на $\Delta\sigma/\sigma_0$ толщину пластинки необходимо даже увеличивать; оптимальная толщина таких фотоприемников выбирается из условия достижения максимального отношения сигнал/шум (см. п. 7.1.3).

Наиболее важными параметрами фотосопротивлений являются их вольт-ваттовая чувствительность, быстродействие и спектр фоточувствительности. Рассмотрим эти параметры более подробно.

Вольт-ваттовая чувствительность. Вольт-ваттовая чувствительность фотоприемника определяется как отношение сигнала фотопроводимости ΔU к мощности падающего на образец излучения $P_{опт}$. Выражая $P_{опт}$ через плотность потока квантов I_0, их энергию $\hbar\omega$ и площадь чувствительной площадки образца Wl, в условиях оптимального согласования фотосопротивления с нагрузкой получаем

$$S = \frac{\Delta U}{P_{опт}} = \frac{\Delta U}{WlI_0\hbar\omega} = \frac{E_п}{4}\frac{1}{WlI_0\hbar\omega}\frac{\Delta\sigma}{\sigma_0}. \qquad (7.15)$$

Подставляя в эту формулу выражение для $\Delta\sigma/\sigma_0$ для случая собственного поглощения (формула (7.12)), находим

$$S = \frac{E_п}{4}\frac{(1-R)\eta}{V\hbar\omega}\frac{\mu_n\tau_n + \mu_p\tau_p}{\mu_n n_0 + \mu_p p_0}, \qquad (7.16)$$

где $V = Wdl$ — объем образца. Из этой формулы следует, что для получения высокой вольт-ваттовой чувствительности

фотоприемника надо стремиться уменьшать его объем (используя при этом фокусировку падающего излучения), а сам полупроводник должен характеризоваться большим временем жизни неравновесных носителей и проводимостью, близкой к собственной.

Наряду с вольт-ваттовой чувствительностью иногда вводится понятие *чувствительности по току*, которая определяется как отношение вольт-ваттовой чувствительности к сопротивлению образца. Путем несложных преобразований формулу (7.16) можно переписать в виде

$$S_I = \frac{S}{R_{\text{обр}}} = \frac{q(1-R)\eta}{2\hbar\omega}\left[\frac{\tau_n}{t_{\text{прол.}n}} + \frac{\tau_p}{t_{\text{прол.}p}}\right], \qquad (7.17)$$

где $t_{\text{прол.}i} = l^2/(\mu_i E_{\text{п}}/2)$ — время пролета i-того типа носителей через образец.

В связи с формулой (7.17) следует обратить внимание на одну особенность фотопроводимости полупроводников, содержащих уровни прилипания. Рассмотрим в качестве примера фотосопротивления из CdS. Дырки, возникающие при собственном поглощении света в этих кристаллах, быстро захватываются на акцепторные уровни прилипания (связанные с примесью Cu или с вакансиями кадмия) и могут удерживаться на них в течение достаточно долгого ($\sim 10^{-2}$ с) времени. В то же время возникающие при поглощении света электроны не захватываясь движутся к положительному контакту и через характерное время $t_{\text{прол}}$ уходят из образца. Поскольку после их ухода электронейтральность образца нарушается, то для ее поддержания с другого контакта в образец входят новые электроны, которые за то же время пролетают по образцу и уходят в контакт. Такое движение электронов продолжается до тех пор, пока дырки не будут выброшены с уровней прилипания и не рекомбинируют с электронами. Таким образом, среднее число носителей, дающее вклад в фототок при поглощении одного фотона, в этих кристаллах оказывается значительно больше единицы. Отношение $\tau/t_{\text{прол}}$, характеризующее величину этого эффекта, называется *коэффициентом усиления фототока* и в рассмотренном нами примере может достигать значений порядка 10^5 [263]. Коэффициент усиления фототока не следует путать с квантовым выходом фотоионизации η, который в рассмотренном случае остается близким к единице. Из формулы (7.17) следует, что усиление фототока дополнительно увеличивает вольт-ваттовую чувствительность фотоприемника.

Быстродействие фотоприемников. Быстродействие фотоприемников определяется тем, насколько быстро электрический сигнал на выходе приемника может следовать за изменением интенсивности падающего излучения. Его обычно характеризуют постоянной времени, с которой сигнал на выходе приемника спадает после выключения освещения. Поскольку изменение концентрации неравновесных носителей в полупроводнике при этом определяется процессами рекомбинации, то время релаксации сигнала фотопроводимости часто отождествляют с временем жизни неравновесных носителей заряда.

На самом деле вопрос о кинетике затухания сигнала фотопроводимости не так прост. Равенство времени релаксации фотопроводимости и времени жизни выполняется только тогда, когда генерация и рекомбинация неравновесных носителей идет по одному и тому же каналу. Примерами могут служить примесный полупроводник, в котором возбужденный светом электрон через некоторое время захватывается на тот же примесный центр, или $Hg_{1-x}Cd_xTe$, в котором неравновесные электроны и дырки, возбуждаемые при межзонных переходах, рекомбинируют друг с другом путем Оже-рекомбинации (с передачей энергии третьей частице).

В полупроводниках, содержащих глубокие уровни в запрещенной зоне, часто наблюдается явление прилипания носителей, о котором мы уже говорили в предыдущем разделе. В этом случае для описания фотопроводимости необходимо ввести два характерных времени — *время жизни фотопроводимости* $\tau_{фп}$, которое равно среднему времени, которое возбужденный носитель проводит в разрешенной зоне и в течение которого он участвует в проводимости, и *время релаксации фотопроводимости* $\tau_{рел}$, которое равно времени восстановления равновесного заполнения уровней в запрещенной зоне. Существенное различие этих времен связано с тем, что перед рекомбинацией возбужденный носитель много времени проводит на уровнях прилипания, выбрасываясь с которых и снова захватываясь на которые, он лишь в течение очень небольшой доли времени своего существования участвует в проводимости образца. Описанная особенность явления фотопроводимости в полупроводниках с уровнями прилипания позволяет заключить, что найденные из измерений фотопроводимости значения $\tau_{фп}$ лишь в редких случаях могут интерпретироваться как время жизни неравновесных носителей.

Спектры фоточувствительности. Спектры чувствительности фотосопротивлений в пренебрежении поверхностной рекомбинацией по сути дела уже были рассчитаны нами выше (см. формулу (7.12)). Как и следовало ожидать, спектральная чувствительность пропорциональна доле потока квантов, поглощенной в образце, которая, в свою очередь, определяется спектральной зависимостью коэффициента поглощения в полупроводнике $\alpha(\hbar\omega)$ и толщиной образца.

Нам остается лишь на качественном уровне обсудить влияние поверхностной рекомбинации на спектры фотопроводимости. Как мы отмечали выше, влияние поверхностной рекомбинации можно учесть, решив уравнение (7.5) с граничными условиями (7.13). Анализ этого решения показывает, что фоточувствительность образца уменьшается:

1) с увеличением скорости поверхностной рекомбинации s при фиксированном значении α и

2) с увеличением коэффициента поглощения α при фиксированном значении s (см. рис. 7.3).

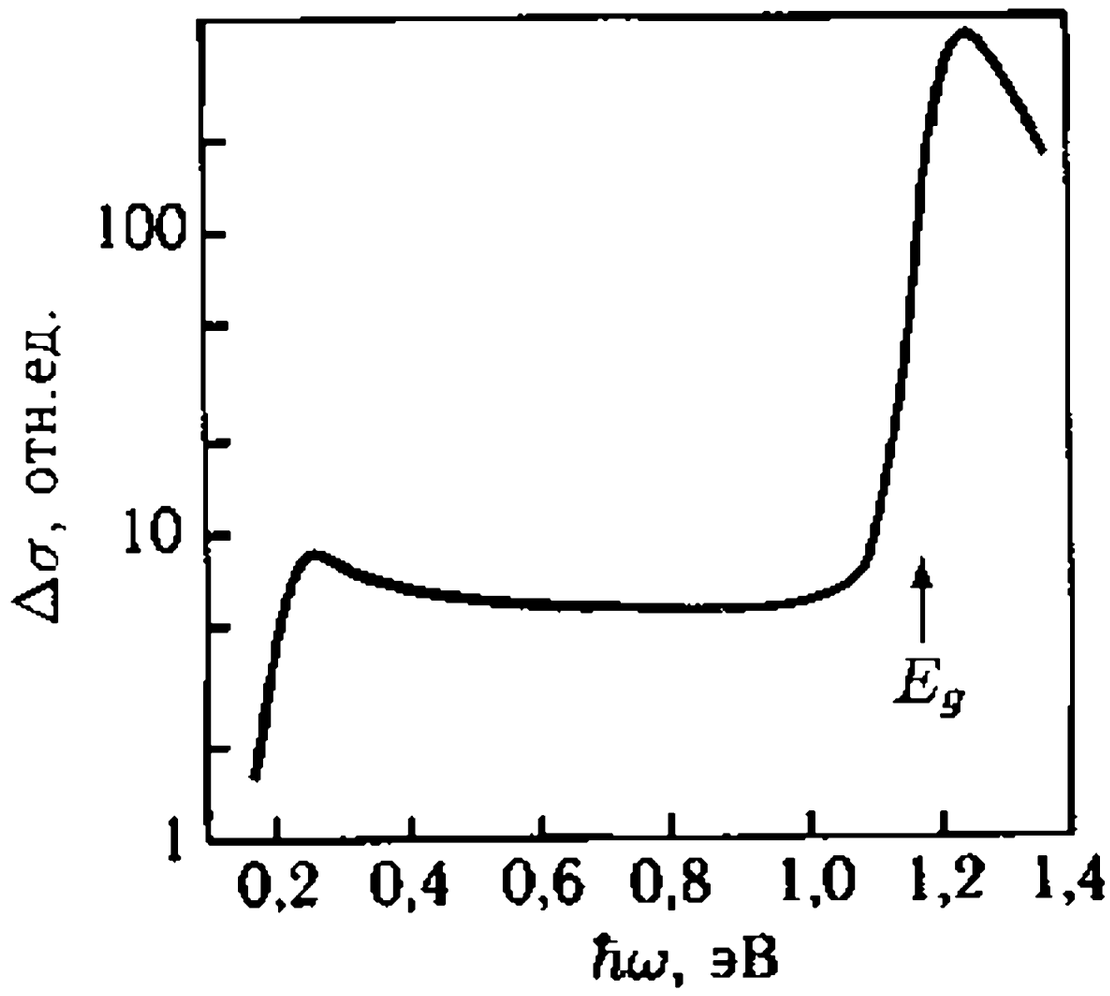


Рис. 7.4. Спектры фотопроводимости кремния, легированного индием, при $T = 93$ К. Спад сигнала выше края собственного поглощения связан с влиянием поверхностной рекомбинации [10]

Спектр фотопроводимости образца p-Si(In), на котором видны полосы примесной и собственной фотопроводимости, показан на рис. 7.4. Уменьшение фотопроводимости при $\hbar\omega > E_g$ обусловлено описанным выше влиянием поверхностной рекомбинации, а небольшое уменьшение примесной фотопроводимости в области энергий кванта 0,3–0,7 эВ — особенностью спектральной зависимости сечения фотоионизации для оптических переходов глубокая примесь–зона [10].

Инфракрасные приемники. Создание фотоприемников, работающих в инфракрасной (ИК) области спектра, представляет собой важную задачу. В этой области лежат полосы поглощения многих газов (прежде всего H_2O и CO_2), отделенные друг от друга так называемыми *окнами прозрачности атмосферы.* Фотоприемники, работающие в окнах прозрачности

атмосферы, используются для регистрации сигналов от удаленных источников (например, при наблюдении поверхности Земли из космоса, слежении за летящими целями, в астрономии). Они также применяются в системах ночного видения, тепловидении, секретной связи, для бесконтактного измерения температуры в технике и медицине. Фотоприемники, настроенные на полосы поглощения, используются в задачах контроля за состоянием окружающей среды (контроль загрязнения атмосферы CO, NO_2, NH_3, H_2S, определение содержания CH_4, озона и др.) [262].

Работа инфракрасных фотоприемников может основываться либо на явлении примесного поглощения, либо на собственном поглощении в полупроводнике. Несмотря на более высокие параметры, достигаемые в «собственных фотоприемниках», набор полупроводников, из которых они могут быть созданы, в общем, ограничен. Кроме Ge, Si, InSb, InAs, PbS, PbSe, которые характеризуются фиксированным значением ширины запрещенной зоны [262], приемники на основе собственной фотопроводимости изготавливают еще из ряда твердых растворов, наиболее важными среди которых являются $Hg_{1-x}Cd_xTe$ и $Pb_{1-x}Sn_xTe$ [261, 264]. В первом из этих твердых растворов ширина запрещенной зоны непрерывно возрастает от отрицательного значения [1]) до 1,53 эВ с ростом содержания кадмия, а во втором при увеличении x она сначала уменьшается до нуля, а затем снова возрастает. Таким образом, подбирая состав твердого раствора, в этих узкозонных полупроводниках можно изменять положение края собственного поглощения в достаточно широких пределах.

Работа «примесных фотоприемников» основана на явлении *примесного поглощения*, *длинноволновая граница* которого определяется энергией ионизации примеси. Эти фотоприемники обычно изготавливают из полупроводников, технология которых хорошо развита (Ge, Si). Наибольшее значение среди таких фотоприемников для средней ИК-области имеют Ge(Au), Ge(Hg), Ge(Cu), Ge(Zn) [262]. Для получения высокой чувствительности примесные фотоприемники требуют охлаждения до температуры, при которой kT составляет примерно 1/20 от энергии ионизации уровня. Спектры фоточувствительности некоторых из этих приемников можно увидеть ниже на *рис. 7.7*.

---

[1]) Теллурид ртути является полуметаллом — материалом, в котором дно зоны проводимости по энергии расположено ниже потолка валентной зоны.

При попытке продвинуться в еще более далекую ИК-область спектра разработчики фотоприемников сталкиваются с тем, что энергия ионизации примесей не может быть меньше энергии ионизации водородоподобных уровней. Поэтому понятно, что для каждого полупроводника существует некоторая максимальная длина волны, до которой в нем еще можно наблюдать примесную фотопроводимость. Для кремния наиболее длинноволновыми приемниками являются приемники из Si(B) ($\lambda_{гр} = 28$ мкм), для германия — из Ge(Sb) и Ge(Ga) ($\lambda_{гр} = 115$–130 мкм). Продвижение в более длинноволновую область в таких фотоприемниках возможно за счет эффекта уменьшения энергии ионизации акцепторных уровней в результате снятия вырождения зон легких и тяжелых дырок при одноосной деформации кристаллов вдоль оси [100]. Так, в фотоприемниках из напряженного Ge(Ga) удается сдвинуть длинноволновую границу фоточувствительности от 114 к 200 мкм [265].

Наиболее серьезным недостатком примесных фотоприемников является низкий коэффициент примесного поглощения. Хотя растворимость водородоподобных примесей в Ge и Si обычно высока (10^{19}–10^{21} см$^{-3}$ [10]), однако начиная с концентрации $\sim 5 \cdot 10^{14}$ см$^{-3}$ из-за перекрытия волновых функций отдельных атомов примеси в полупроводнике образуется *примесная зона* и появляется *прыжковая проводимость* по этой зоне. При этом темновой ток фотоприемников резко возрастает, а чувствительность — падает. По этой причине оптимальная концентрация акцепторов в фотоприемниках Ge(Ga) составляет всего $\sim 2 \times 10^{14}$ см$^{-3}$, а соответствующий ей коэффициент примесного поглощения равен 2 см$^{-1}$. Еще одной особенностью примесных фотоприемников является то, что их рабочее напряжение ограничено сверху напряжением возникновения примесного пробоя (см. подстрочное замечание на с. 45). Так, приемники из Ge(Ga) работают при напряжении смещения всего 30–50 мВ. Все это, вместе с необходимостью работать при температуре, близкой к температуре жидкого гелия (4,2 К), делает конструкцию этих примесных фотоприемников довольно сложной.

Фотоприемники с блокированной примесной зоной. В 1986 г. Петроф и Стейплброк [266] из Rockwell International получили патент на новую необычную конструкцию инфракрасного фотоприемника, которая позволяет одновременно заметно повысить чувствительность примесных фотоприемников и сдвинуть границу фотопроводимости в область длинных волн. Это — так называемый *фотоприемник с блокированной примесной*

зоной (BIB — blocked impurity band). В этих приемниках достаточно сильно легированный водородоподобными примесями слой полупроводника отделен от одного из контактов блокирующим слоем из нелегированного полупроводника толщиной в несколько мкм (см. рис. 7.5). Благодаря тому, что только возбужденные в зону проводимости электроны могут преодолеть блокирующий слой, а электроны, находящиеся в примесной зоне, этого сделать не могут, темновой ток фотоприемника при низкой температуре чрезвычайно мал (1–100 электронов в секунду). Эта особенность позволяет примерно в 100 раз повысить концентрацию легирующих примесей в поглощающем слое и за счет увеличения коэффициента поглощения увеличить чувствительность фотоприемника. При этом, поскольку за счет эффектов сильного легирования с ростом концентрации примесей их энергия ионизации уменьшается, это одновременно приводит и к сдвигу красной границы поглощения в длинноволновую область. Так, в фотоприемниках с блокированной примесной зоной из Ge(Ga) удалось сдвинуть границу чувствительности от $\lambda_{гр} = 115$ мкм к 220 мкм [267].

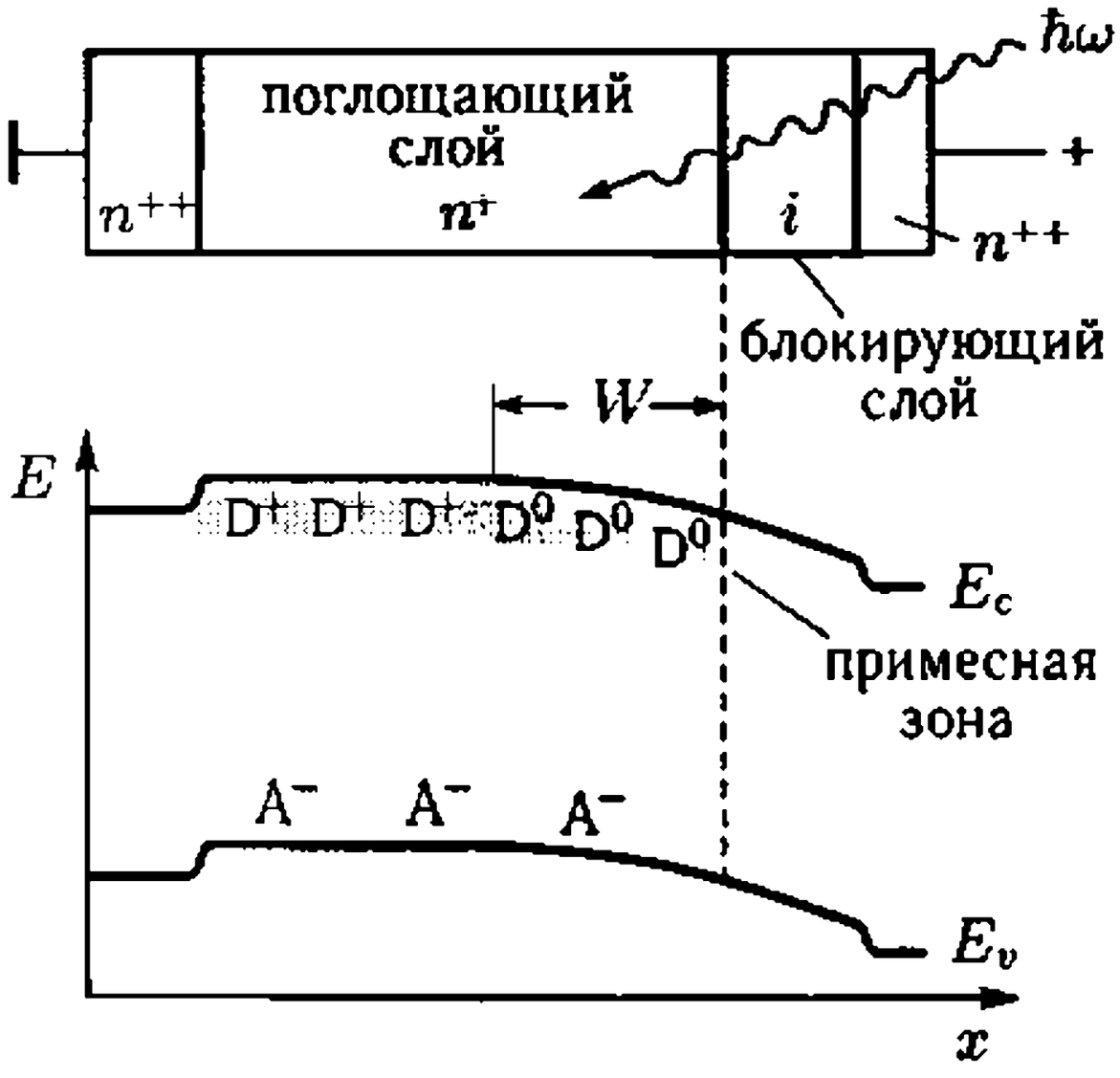


Рис. 7.5. Энергетическая диаграмма фотоприемника с блокированной примесной зоной

Следует особо подчеркнуть, что для изготовления BIB-фотоприемников необходимы полупроводники с очень низкой концентрацией остаточных примесей в i-области и очень малым коэффициентом компенсации в поглощающем слое, причем с уменьшением энергии ионизации уровней требования к чистоте полупроводника сильно возрастают. Первые BIB-приемники из Si с энергией ионизации примеси ~45 мэВ были изготовлены еще в конце 80-х годов. В разрабатываемых в настоящее время приемниках из Ge с энергией ионизации доноров ~10 мэВ концентрация остаточных примесей в i-области должна быть $<10^{13}$ см$^{-3}$, а коэффициент компенсации в поглощающем слое $<10^{-4}$. Последнее обусловлено тем, что квантовая эффективность фотоприемника определяется толщиной области пространственного

заряда W (см. рис. 7.5), а эта величина в случае, когда все доноры заполнены электронами [1]), определяется концентрацией *компенсирующей* примеси (отрицательно заряженными акцепторами). Из уравнения Пуассона следует, что толщина W, рабочее напряжение на фотоприемнике $V_{\text{раб}}$ и концентрация компенсирующей примеси N_a связаны соотношением

$$N_a \approx \varepsilon V_{\text{раб}}/(2\pi q W^2).$$

Для $V_{\text{раб}} = 50$ мВ и $W = 50$ мкм в случае германия величина N_a должна составлять всего $3 \cdot 10^{10}$ см$^{-3}$. Максимальная напряженность электрического поля при этом равна 20 В/см, что близко к напряженности поля примесного пробоя (см. подстрочное замечание на с. 45). Интересно, что из-за того, что в области изгиба зон все донорные уровни полностью заполнены, электроны, возбуждаемые ИК-излучением в зону проводимости, не могут рекомбинировать во время своего движения к положительному контакту, что обеспечивает полное собирание носителей, генерируемых в этой области.

Матрицы BIB-фотоприемников из Si(As) размером 256× ×256 элементов и из Si(Sb) размером 128×128 элементов в настоящее время используются для астрономических наблюдений из космоса на инфракрасном телескопе SIRTF; матрицы из Si(As) размером 1024×1024 элемента будут установлены на спутнике WISE, который планируют запустить в 2009 году. Имея примерно такую же чувствительность, как и традиционные примесные фотоприемники (токовая чувствительность ~30 А/Вт на длине волны 30 мкм), BIB-фотоприемники обладают существенно бóльшей помехоустойчивостью по отношению к радиационному космическому фону, что связано с меньшей толщиной поглощающего слоя.

В последние годы начаты работы по созданию BIB-приемников на основе GaAs(Te). При энергии ионизации доноров около 6 мэВ длинноволновая граница чувствительности в лабораторных образцах этих детекторов в настоящее время превышает 300 мкм [268]. Используя различные технологические ухищрения, разработчикам удалось получить GaAs с разностной концентрацией $|N_d - N_a| < 10^{12}$ см$^{-3}$. Для продвижения в область еще более длинных волн предполагается использовать InAs и InSb, энергия ионизации доноров в которых составляет

---

[1]) При подаче напряжения смещения на структуру необходимые для этого электроны подходят из объема поглощающего слоя.

соответственно 1,4 и 0,7 мэВ [269]. К сожалению, в настоящее время технология получения этих кристаллов недостаточно разработана, и концентрация остаточных примесей в них составляет $\sim 10^{14}$ см$^{-3}$.

Интересной особенностью BIB-фотоприемников является возможность их работы в *режиме счета фотонов* при регистрации квантов ИК-излучения. Создав в некоторой области поглощающего слоя, прилегающей к блокирующему слою, условия для ударной ионизации примесей и получив при этом коэффициент лавинного умножения $M = 10^4$–10^5, авторам работы [270] удалось создать на основе BIB-фотоприемника из Si(As) твердотельный аналог фотоэлектронного умножителя, который регистрировал отдельные кванты излучения в области длин волн 0,4–28 мкм с эффективностью 30% на длине волны 20 мкм.

Фотоприемники на переходах между уровнями размерного квантования. Новым типом ИК-фотоприемников для области длин волн 3–35 мкм являются фотоприемники, работающие на переходах между уровнями размерного квантования в сверхрешетках [271] [1]). Преимуществом этих приемников является бóльшая гибкость в формировании необходимого спектра фоточувствительности (длины волны в максимуме и ширины спектра). Из этих фотоприемников уже изготавливаются матрицы с числом элементов 1024×1024, однако темновые токи в них пока выше, чем в традиционных фотоприемниках.

Принципиальную возможность использования сверхрешеток в качестве приемников излучения обсуждали еще Есаки в 1977 г., но детальную теорию фотоприемников на оптических переходах между уровнями размерного квантования первыми развили Кун и Карунасири [272]. В одном из первых фотоприемников такого типа, созданном в Bell Laboratories [273], использовалась сверхрешетка, имевшая 50 периодов и состоявшая из квантовых ям (GaAs) шириной 76 Å, разделенных потенциальными барьерами из широкозонного полупроводника ($Al_{0,36}Ga_{0,64}As$) шириной 130 Å. Энергетический зазор между первым и вторым уровнями размерного квантования при этом составлял 123 мэВ, что соответствует длине волны $\lambda \approx 10$ мкм. Структура была легирована так, чтобы при низкой температуре в равновесии нижний уровень размерного квантования был полностью заполнен электронами. При поглощении кванта света

---

[1]) В зарубежной литературе для обозначения этого типа фотоприемников используется аббревиатура QWIP (quantum well infrared photodetector).

электроны возбуждались с уровня E_1 на уровень E_2 (см. рис. 7.6 *а*), откуда они путем туннелирования в приложенном к образцу электрическом поле ($8 \cdot 10^4$ В/см) переходили в область квазинепрерывного энергетического спектра и, таким образом, увеличивали проводимость образца. Спектры поглощения, отвечающие переходам между первым и вторым уровнями размерного квантования, имели лоренцеву форму с шириной на половине высоты, равной 11 мэВ. Достаточно узкая полоса фоточувствительности вообще характерна для всех фотоприемников, работающих на переходах между уровнями размерного квантования; ее ширина определяется конечным времени жизни возбужденного состояния (0,17 пс в рассматриваемом случае) и конечной вероятностью туннелирования. Токовая чувствительность описываемого приемника составляла 0,8 А/Вт для поляризованного ИК-излучения [1]).

Недостатком описанного выше фотоприемника является довольно большой темновой ток, связанный с туннелированием электронов между ямами в структуре с недостаточно толстыми барьерами. Флуктуации этого тока снижают чувствительность приемника. Этот недостаток удалось преодолеть в структуре, описанной в работе [274]. В ней сверхрешетка также имела 50 периодов и состояла из квантовых ям из GaAs шириной 40 Å, разделенных потенциальными барьерами из $Al_{0,31}Ga_{0,69}As$ шириной 300 Å. Из-за малой ширины ямы в ней может локализоваться *только один* уровень размерного квантования, а остальные уровни попадают в область квазинепрерывного спектра. При этом ширина спектра для межподзонных переходов увеличивается по сравнению со структурами, в которых яма содержит два уровня размерного квантования. Поскольку возбужденный электрон сразу попадает в область континуума и ему не нужно туннелировать сквозь барьер, толщина барьера может быть увеличена. Это позволяет сильно уменьшить темновой ток и повысить вольтваттовую чувствительность (см. рис. 7.6 *б*). Чтобы максимально ослабить темновой ток в структурах с толстыми барьерами, в которых он определяется тепловым выбросом электронов и термополевой эмиссией с уровня E_1, необходимо, чтобы этот уровень максимально далеко отстоял от границы квазинепрерывного

---

[1]) Согласно правилам отбора, оптические переходы между подзонами, образуемыми из зоны проводимости полупроводника, разрешены только для электромагнитных колебаний, в которых вектор электрического поля имеет компоненту, перпендикулярную слоям сверхрешетки (см. [77]).

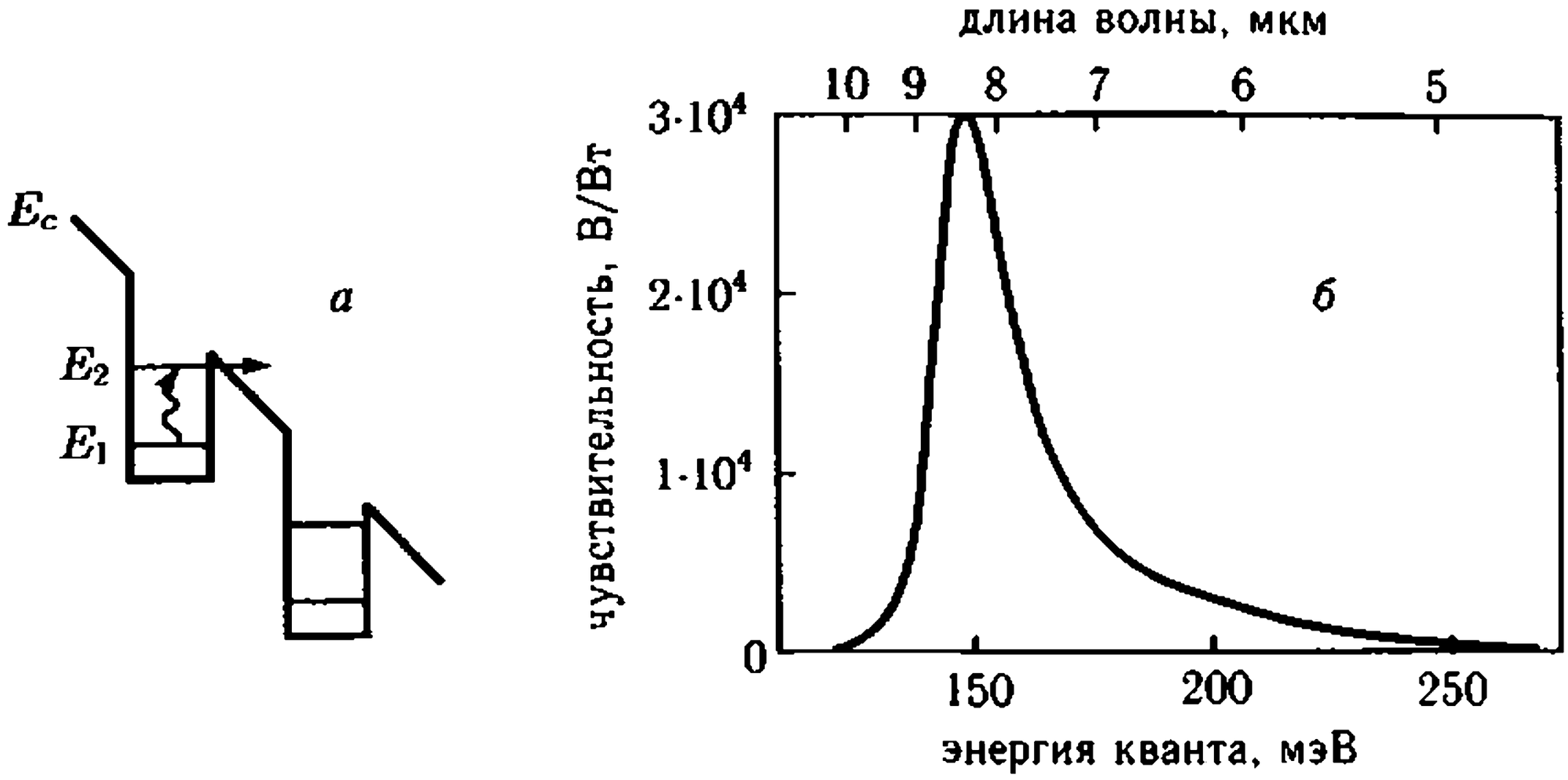


Рис. 7.6. Энергетическая диаграмма (*а*) и спектр фоточувствительности при 77 К (*б*) ИК-фотоприемника на основе сверхрешетки [274]

спектра. Для этого ширина ямы и высота барьера подбираются так, чтобы энергия уровня E_2 совпадала с положением края зоны проводимости в материале барьера.

В еще одном варианте конструкции QWIP-фотоприемника барьерные слои в сверхрешетке создаются настолько тонкими, чтобы перенос возбужденных носителей осуществлялся по *мини-зоне*, образованной из уровня E_2. Чтобы уменьшить темновой ток, связанный с проводимостью по первой мини-зоне, в структуру можно добавить блокирующий слой, аналогичный используемому в BIB-фотоприемниках (см. с. 394). Положение края зоны проводимости в этом слое выбирается так, чтобы он лежал между мини-зонами, образованными из уровней E_1 и E_2. Эта конструкция фотоприемника характеризуется значениями вольт-ваттовой чувствительности и обнаружительной способности, промежуточными между рассмотренными выше конструкциями. Улучшить эти параметры можно, расположив мини-зону в области квазинепрерывного спектра [275].

Коэффициент поглощения, связанный с оптическими переходами между уровнями размерного квантования, составляет 500–3000 $см^{-1}$. Из-за сильной поляризационной зависимости коэффициента поглощения для получения высокой квантовой эффективности в одиночных QWIP-приемниках излучение вводится в структуру под углом 45° к плоскости слоев (геометрия типа «скошенный край»), а для матриц фотоприемников более технологичным оказывается придание поверхности структуры «шероховатой» формы. При этом падающее на приемник излучение

преломляется под случайными углами и достаточно эффективно поглощается в объеме структуры.

Другими способами отклонения падающего излучения является создание двумерных дифракционных решеток на плоской поверхности структуры [276] и придание поверхности рифленой формы путем ее избирательного химического травления.

Интересная возможность ослабить описанную зависимость поглощения от поляризации излучения появляется при использовании переходов между уровнями размерного квантования в валентной зоне. Дело в том, что из-за сильного перемешивания состояний легких и тяжелых дырок запрет на поглощение излучения с ориентацией вектора электрического поля волны параллельно плоскости слоев снимается и в таких структурах исчезает необходимость в создании специальной геометрии ввода излучения в фотоприемник. К сожалению, из-за низкой подвижности дырок возбужденные носители рекомбинируют намного раньше, чем они достигают контактов, и чувствительность таких QWIP-приемников примерно на порядок ниже, чем для приемников, использующих переходы между состояниями, образуемыми из зоны проводимости. По-видимому, идеальным решением задачи создания «изотропного» приемника на основе переходов между уровнями размерного квантования является использование слоев самоорганизованных квантовых точек вместо квантовых ям [276–278]. Теория фотоприемников на квантовых точках (QDIP — quantum dot infrared photodetector) развита в работе [279].

Усложнение структуры сверхрешетки позволяет получить новые свойства в QWIP-приемниках. Так, для расширения спектральной области чувствительности фотоприемников сверхрешетки строят из повторяющихся блоков, состоящих из нескольких квантовых ям с несколько различной шириной и высотой барьера [280]. Создание асимметричных ям и барьеров позволяет создавать фотоприемники, область спектральной чувствительности которых перестраивается внешним электрическим полем [281].

Кроме рассмотренной выше гетеропереходной пары GaAs/AlGaAs, для создания QWIP-приемников используются пары GaAs/InGaP, GaAs/AlInP, InGaAs/InP, InGaAsP/InP, InGaAs/AlInAs, InGaAs/GaAs. Наиболее «длинноволновые» фотоприемники получены на основе сверхрешеток GaAs/ AlGaAs [282] и GaAs/InGaAs [283]. Быстродействие описываемых фотоприемников ограничивается временем жизни фотовозбужденных носителей, которое в структурах GaAs/AlGaAs состав-

ляет ~5 пс [284]. На основе QWIP-приемников в настоящее время выпускается широких набор переносных и стационарных тепловизионных систем с разрешением по температуре, достигающим 0,01 К, которые находят применение в пожаротушении, вулканологии, медицине, астрономии, а также в оборонной технике.

Фотоприемники, основанные на явлении внутренней фотоэмиссии. В последние годы группа фотоприемников, чувствительных в дальней инфракрасной области спектра (50–200 мкм), пополнилась новым типом фотоприемников, работа которых основана на явлении внутренней фотоэмиссии (HIWIP — homojunction interfacial workfunction internal photoemission) [285]. Приемники этого типа представляют собой многослойные структуры, построенные из очень тонких чередующихся слоев сильно легированного полупроводника, в котором происходит поглощение излучения на свободных носителях, и слоев почти собственного полупроводника.

Энергетический зазор между положением уровня Ферми в сильно легированном слое и краем зоны проводимости в нелегированном слое, который в значительной мере определяется эффектом сужения запрещенной зоны в сильно легированном полупроводнике (см. п. 2.2.1), определяет ту пороговую энергию, которую должен получить от кванта света электрон, чтобы преодолеть потенциальный барьер и начать участвовать в проводимости структуры.

Использование гетеропереходов (например, GaAs–$Al_{0.02}Ga_{0,98}$ As) для формирования барьеров позволяет не только сделать параметры этих фотоприемников более управляемыми, но также уменьшает темновой ток и повышает токовую чувствительность до значений, которые заметно превосходят соответствующие параметры QWIP-фотоприемников [286]. В отличие от барьеров Шоттки, в которых квантовый выход фотоэмиссии из металла не превышает 1 % (см. с. 326 в гл. 5), в HIWIP-фотоприемниках уже достигнута квантовая эффективность 24 % [286].

7.1.3. Шумы в фотоприемниках. Введенная на с. 389 вольт-ваттная чувствительность не может полностью охарактеризовать фотоприемник, поскольку в действительности регистрация сигнала проводится в условиях шума, и существует нижняя граница значения сигнала, который может быть обнаружен. Источниками шума при измерениях фотопроводимости являются: тепловой шум, дробовой шум, генерационно-рекомбинационный шум, фликкер-шум, а также *радиационный (фотонный) шум*.

Физические причины возникновения первых четырех видов шума были рассмотрены нами в п. 2.6. Фотонный шум представляет собой флуктуации проводимости образца, возникающие вследствие флуктуаций потока падающих на образец квантов излучения.

Поскольку напряжение генерационно-рекомбинационного шума в фотосопротивлениях увеличивается пропорционально току через образец, а напряжение теплового шума остается неизменным, то при высоких напряжениях на образце основным источником шума становится генерационно-рекомбинационный шум. Характеристики этого шума можно рассчитать, воспользовавшись формулой (2.30) для спектральной плотности флуктуаций тока. В компенсированном полупроводнике n-типа средний квадрат напряжения этого шума при оптимальном согласовании сопротивлений образца и нагрузки и $\omega\tau_n \ll 1$ равен

$$V_{\text{г-р}}^2 = \frac{E_{\text{п}}^2}{4}\frac{\tau_n}{n_0 V}\Delta f, \qquad (7.18)$$

где $E_{\text{п}}$ — напряжение питания схемы, V — объем образца, Δf — полоса частот, в которой производится измерение шума. Произведение $n_0 V$, стоящее в знаменателе этой формулы, представляет собой среднее число электронов в образце.

При увеличении напряжения на фотосопротивлении одновременно возрастают и вольт-ваттовая чувствительность (7.16), и напряжение генерационно-рекомбинационного шума (7.18). Поэтому при высоких напряжениях на образце, когда вклад других видов шума в полный шум становится несущественным, отношение сигнал/шум в фотоприемнике, работающем в области собственного поглощения, становится равным

$$\frac{\Delta U}{V_{\text{г-р}}} = \frac{S \cdot P_{\text{опт}}}{V_{\text{г-р}}} = \frac{(1-R)\eta}{2\hbar\omega}\sqrt{\frac{\tau_n}{n_0 V}}\left(1 + \frac{\mu_p}{\mu_n}\frac{\tau_p}{\tau_n}\right)\frac{P_{\text{опт}}}{\sqrt{\Delta f}}. \qquad (7.19)$$

Уравнение (7.19) позволяет нам рассчитать минимальную мощность оптического сигнала $P_{\text{пор}}$, которую можно зарегистрировать в условиях шума. Эта величина, называемая *эквивалентной мощностью шума* [1], определяется из условия равенства сигнала фотопроводимости среднеквадратичному напряжению

[1] В зарубежной литературе эта величина обозначается NEP (noise-equivalent power).

шума и равна

$$P_{\text{пор}} = \frac{V_{\text{г-р}}}{S} = \frac{2\hbar\omega}{(1-R)\eta}\sqrt{\frac{n_0 V}{\tau_n}}\left(1+\frac{\mu_p}{\mu_n}\frac{\tau_p}{\tau_n}\right)\sqrt{\Delta f}. \qquad (7.20)$$

Теперь можно ввести еще одну важную характеристику фотоприемника — его *обнаружительную способность*, которая определяется следующим образом:

$$\mathcal{D}^* = \frac{\sqrt{A\Delta f}}{P_{\text{пор}}}, \qquad (7.21)$$

где $A = Wl$ — площадь фоточувствительной площадки, а $P_{\text{пор}}$ — введенная выше эквивалентная мощность шума. Размерность этой величины — [см·Гц$^{1/2}$Вт$^{-1}$]. Нетрудно видеть, что величина $\mathcal{D}^*$ изменяется пропорционально $\sqrt{\tau_n/(n_0 d)}$ и, следовательно, чтобы создать наиболее чувствительный фотоприемник на основе собственной фотопроводимости, необходимо уменьшать толщину образца, увеличивать время жизни и использовать материал с минимальной концентрацией носителей заряда. Для фотоприемников, работающих на основе примесной фотопроводимости, $\mathcal{D}^* \sim \left(1-e^{-\alpha d}\right)/\sqrt{d}$, максимум $\mathcal{D}^*$ достигается при толщине образца, отвечающей условию $\alpha d \approx 1{,}25$.

Для фотоприемников, работающих в ИК-области спектра, характерна сильная фоновая засветка фотоприемника инфракрасным излучением, испускаемым окружающими предметами. Так, для охлаждаемого до 77 К фотосопротивления из PbSe с красной границей фоточувствительности 7 мкм поток фонового излучения с температурой 300 К, падающий на образец из телесного угла π стерадиан, составляет $Q_b = 1{,}1 \times \times 10^{17}$ квант/см^2·с [262]. Поэтому регистрация сигнала ИК-фотоприемниками производится в присутствии постоянной фоновой подсветки. В этом случае заметный вклад в шум фотоприемника начинает вносить фотонный шум, а проводимость образца σ_0 определяется уже не только носителями, возбуждаемыми за счет тепловой генерации примесных уровней, но и носителями, возбуждаемыми фоновой подсветкой. Если концентрация последних превышает концентрацию равновесных носителей, то говорят, что приемник работает *в режиме ограничения фона* (BLIP — background limited infrared photodetector). Очевидно, что в этом случае обнаружительная способность

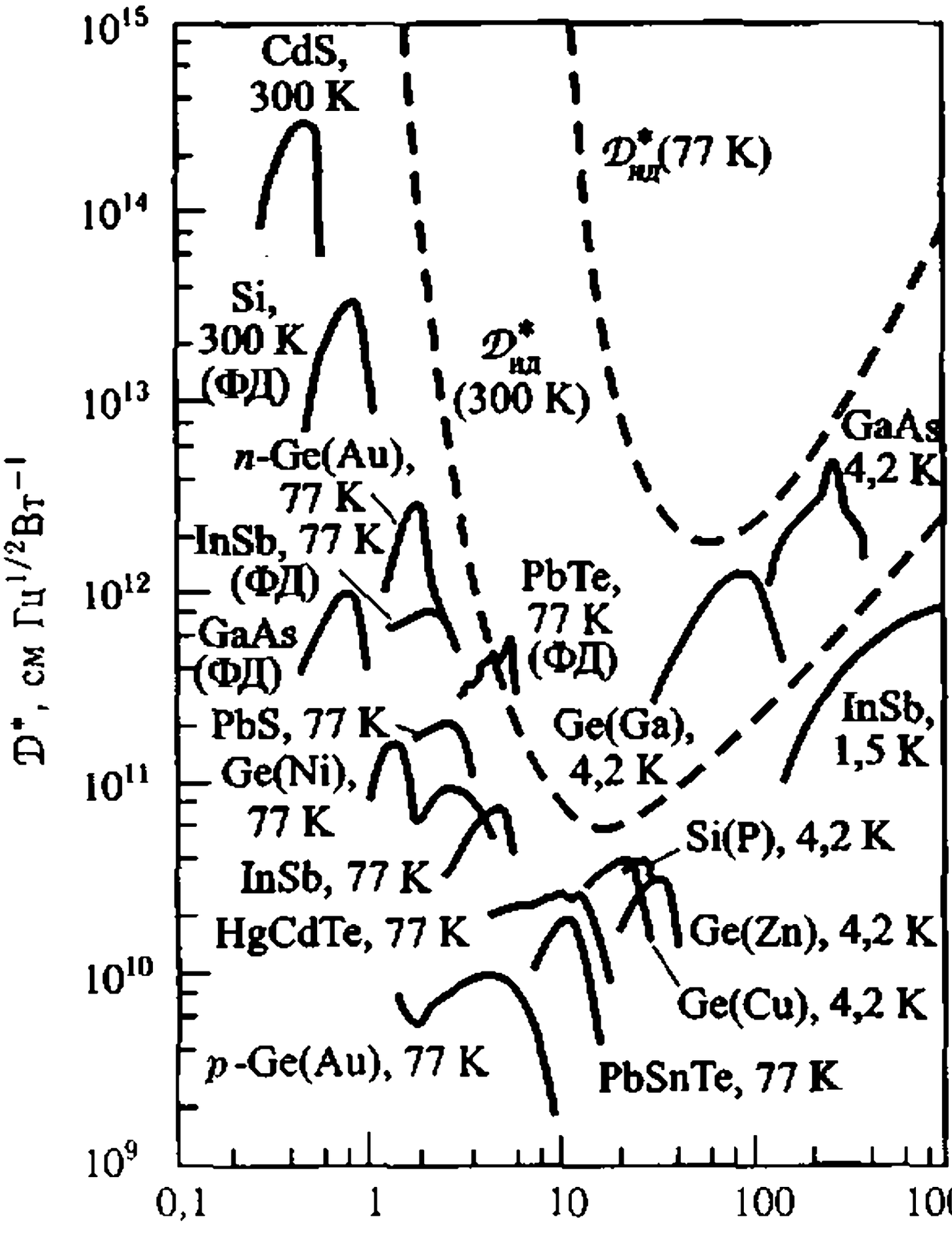


Рис. 7.7. Обнаружительная способность ряда инфракрасных фотоприемников. Штриховыми линиями показаны теоретические зависимости обнаружительной способности $\mathcal{D}^*_{ид}$, ограниченной флуктуациями фонового излучения, при температурах фона 77 и 300 К [14]

фотоприемника $\mathcal{D}^*$ не может превышать некую величину, определяемую *флуктуациями фона*. Для фотосопротивления эта величина равна

$$\mathcal{D}^*_{ид} = \frac{\lambda}{4\pi\hbar c}\left(\frac{\eta}{Q_b}\right)^{1/2},$$

где λ — длина волны, η — квантовая эффективность, а Q_b — фоновый поток фотонов в спектральной области чувствительности приемника [262]. Спектральные зависимости $\mathcal{D}^*_{ид}$ для идеального приемника ($\eta = 1$) при двух температурах фона показаны штриховыми линиями на рис. 7.7. На этом же рисунке с этими расчетными кривыми сопоставляются экспериментально полученные зависимости $\mathcal{D}^*(\lambda)$ для ряда инфракрасных фотоприемников. Очевидно, что для увеличения чувствительности

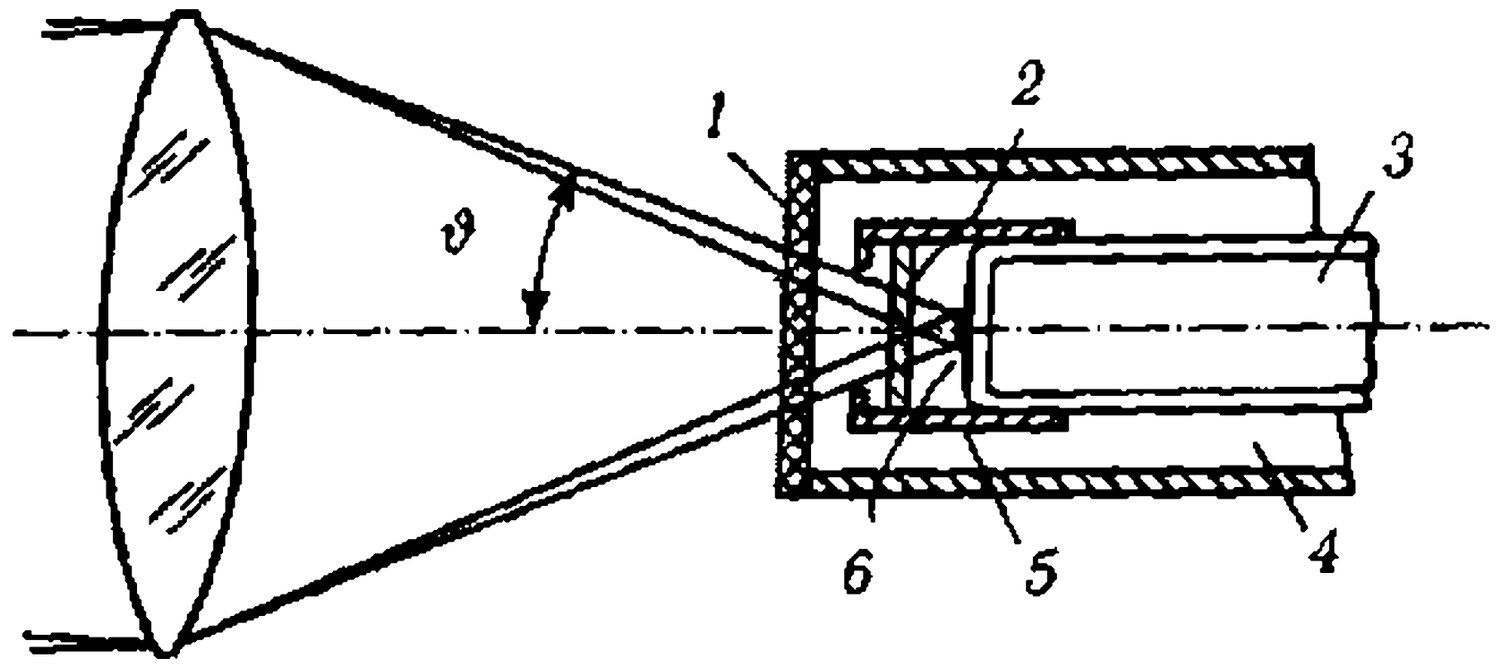


Рис. 7.8. Устройство фотоприемника с охлаждаемым радиационным экраном: *1* — входное окно, *2* — охлаждаемый фильтр, *3* — холодильная камера, *4* — вакуум, *5* — охлаждаемый экран, *6* — чувствительный элемент [262]

фотоприемников необходимо принимать меры по уменьшению интенсивности фоновой подсветки (см. рис. 7.8): уменьшать угол зрения ϑ (величина Q_b изменяется пропорционально $\sin^2\vartheta$), использовать охлаждаемые экраны.

7.1.4. Фотовольтаические приемники (фотодиоды). Фотовольтаическим эффектом называют эффект появления ЭДС в полупроводнике при его освещении. Как известно из курса физики полупроводников [1], для появления фото-ЭДС необходимо выполнение двух условий:

1) полупроводник должен быть неоднородным (в нем должны существовать локальные поля, способствующие пространственному разделению электронов и дырок) и
2) в полупроводнике должны одновременно возбуждаться электроны и дырки.

Примерами неоднородных полупроводниковых структур, в которых возможно появление фото-ЭДС, являются p–n-переходы и барьеры Шоттки [1]).

[1]) В принципе, появление фото-ЭДС возможно и в однородных полупроводниках. При *неоднородном возбуждении* носителей в приповерхностном слое полупроводника возникает *ЭДС Дембера*, а если образец при этом поместить в магнитное поле, то наблюдается ЭДС *фотоэлектромагнитного* (*ФЭМ*) *эффекта* [1]. Для этих двух эффектов важно, чтобы в полупроводнике возбуждались носители обоих знаков (собственное поглощение). При поглощении света свободными носителями в однородных полупроводниках возникает фототок, связанный с *увлечением электронов фотонами* [263]. Кроме того, в однородных кристаллах без центра инверсии в условиях однородного (в том числе, и монополярного) возбуждения может появляться *фотогальванический ток*, связанный с асимметрией процессов поглощения света, рассеяния носителей заряда и их рекомбинации [287]. Из указанных эффектов небольшое практическое значение имеет лишь ФЭМ-эффект в InSb.

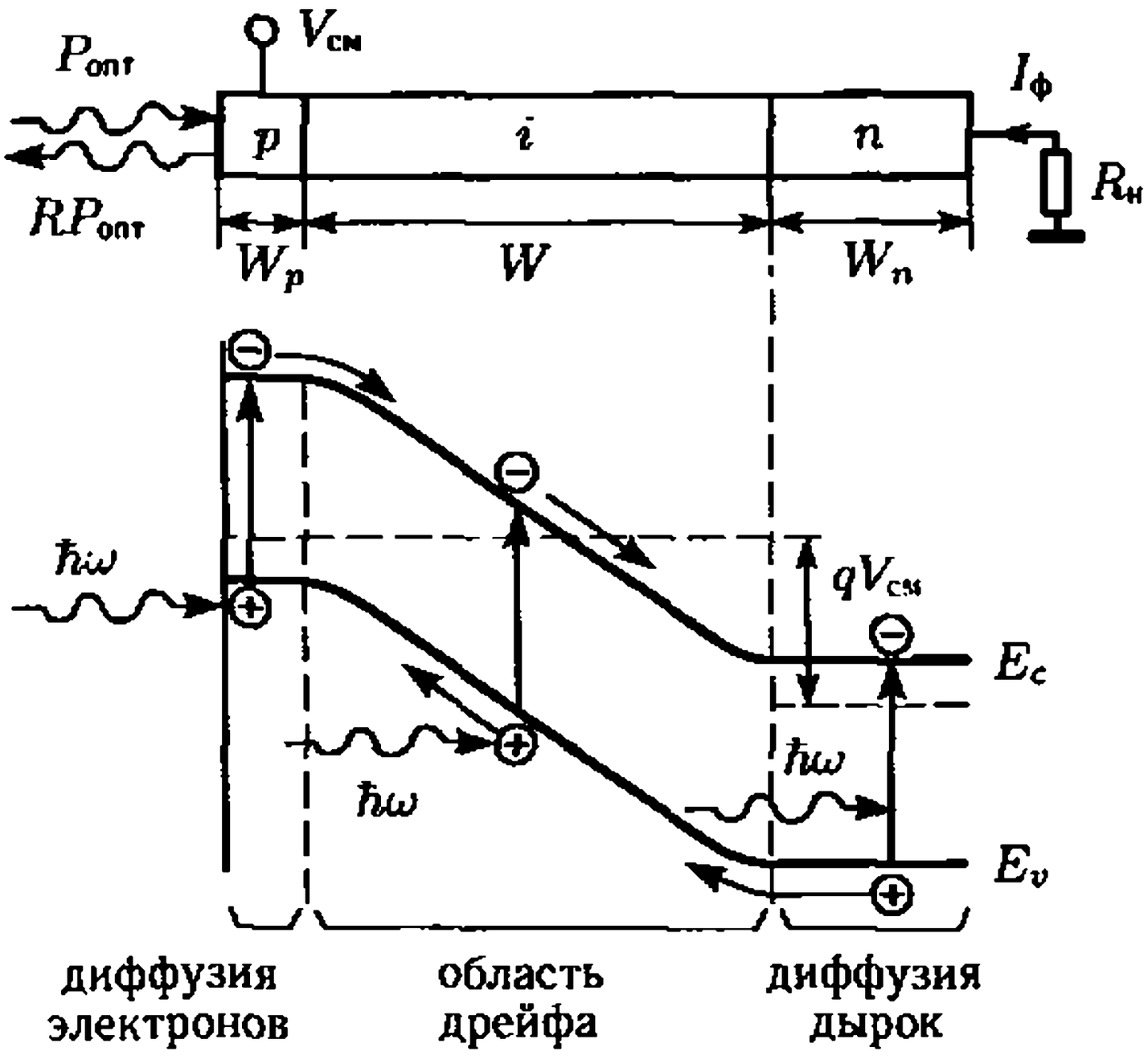


Рис. 7.9. Устройство и энергетическая диаграмма p–i–n-фотодиода при подаче на него обратного смещения $V_{см} < 0$

Рассчитаем квантовую эффективность p–i–n-фотодиода как функцию длины волны падающего излучения при освещении структуры сквозь тонкую p-область (см. рис. 7.9). Темп оптической генерации электронно-дырочных пар в полупроводнике описывается, как обычно, формулой (7.4). Будем считать толщину p-области W_p настолько малой, что все возбуждаемые в ней электроны затягиваются в область сильного электрического поля. В этом случае ток фотодиода будет складываться из тока электронов, возбуждаемых светом в p- и i-областях, и тока дырок, которые возбуждаются в объеме n-области и затем путем диффузии подходят к границе области сильного поля.

Ток фотодиода, связанный с возбуждаемыми в p- и i-области электронами, находится сравнительно легко [1]:

$$J_1 = -q \int_0^{W+W_p} G(x)dx = -qI_0(1-R)\eta(1-e^{-\alpha(W+W_p)}). \quad (7.22)$$

Отрицательный знак тока связан с тем, что положительные заряды (дырки), разделяемые электрическим полем перехода, движутся в направлении, противоположном направлению оси x.

[1] Будем предполагать, что рекомбинацией носителей за время пролета области сильного поля можно пренебречь.

Для вычисления компоненты тока, обусловленного дырками, возбуждаемыми в объеме n-области, необходимо решить уравнение непрерывности

$$D_p \frac{d^2 \delta p}{dx^2} - \frac{\delta p}{\tau_p} + G(x) = 0. \tag{7.23}$$

Считая, что толщина n-области $W_n \gg L_p$, решение этого уравнения будем искать в виде (7.7) с граничными условиями $\delta p = 0$ при $x = W + W_p$ (неравновесные дырки экстрагируются электрическим полем) и $\delta p = 0$ при $x \to \infty$. В результате несложных вычислений находим

$$J_2 = -qD_p \left.\frac{d\delta p}{dx}\right|_{x=W+W_p} = -qI_0(1-R)\eta \frac{\alpha L_p}{1+\alpha L_p} e^{-\alpha(W+W_p)}. \tag{7.24}$$

Таким образом, суммарный фототок по модулю равен

$$J = |J_1 + J_2| = qI_0(1-R)\eta \left(1 - \frac{e^{-\alpha(W+W_p)}}{1+\alpha L_p}\right), \tag{7.25}$$

откуда следует, что квантовая эффективность фотодиода равна

$$\eta_{\text{эф}} = \frac{J}{qI_0} = (1-R)\eta \left(1 - \frac{e^{-\alpha(W+W_p)}}{1+\alpha L_p}\right). \tag{7.26}$$

Нетрудно видеть, что для получения высокой квантовой эффективности необходимо, чтобы выполнялось соотношение $\alpha W \gg 1$ и были приняты все меры по уменьшению коэффициента отражения (например, использовано просветляющее покрытие) [1].

При выводе формулы (7.22) мы не учитывали влияния поверхностной рекомбинации на передней (лицевой) поверхности фотодиода. При очень высоких коэффициентах поглощения ($\alpha W_p \gtrsim 1$) поверхностная рекомбинация начинает играть заметную роль, понижая $\eta_{\text{эф}}$. Чтобы ослабить влияние поверхностной рекомбинации на квантовую эффективность кремниевых фотодиодов в синей и ультрафиолетовой области спектра, их конструкция рассчитывается так, чтобы создать вблизи лицевой поверхности фотодиода электрическое поле, затягивающее неосновные

[1] Из-за высокого показателя преломления, значения которого в полупроводниках находится в пределах $n = 2{,}5$–6, полупроводники характеризуются высоким коэффициентом отражения $R = (n-1)^2/(n+1)^2 = 18$–$51\,\%$.

области в глубь структуры. Это можно сделать, например, создав верхний контакт в виде барьера Шоттки [14].

Быстродействие фотодиодов ограничивается совместным действием нескольких факторов:

1) временем пролета носителей через область сильного поля,

2) постоянной времени заряда барьерной емкости структуры,

3) временем диффузии носителей, возбуждаемых в электрически нейтральной n-области структуры.

При разработке быстродействующих фотодиодов для исключения влияния последнего фактора можно изменить геометрию прибора и вводить излучение в фотодиод параллельно границам i-области, при этом одновременно достигая и высокой квантовой эффективности. Для увеличения быстродействия обычно стремятся максимально уменьшить толщину i-слоя (время пролета), сохранив при этом достаточно высокую квантовую эффективность и контролируя величину емкости структуры (которая при уменьшении толщины i-слоя возрастает).

Одним из наиболее быстродействующих фотодиодов в настоящее время является фотодиод DSC10W, выпускаемый фирмой Discovery Semiconductors. Его полоса пропускания достигает 75 ГГц при регистрации излучения с длиной волны 1,3–1,55 мкм. В конструкции этого фотодиода используется гетероструктура p^+-InP—i-$In_xGa_{1-x}As$—i-InP—n^+-InP с двумя областями i-типа. Структура спроектирована так, что возбуждаемые в освещаемом сквозь p-InP узкозонном слое $In_xGa_{1-x}As$ электроны и дырки, которые имеют *разные* скорости насыщения, пролетают структуру приблизительно за равное время.

Шумы фотодиодов, которые обычно работают при подаче обратного смещения («фотодиодный режим»), в основном определяются дробовым шумом протекающего через диод тока. Этот ток складывается из тока регистрируемого сигнала $I_{изм}$, тока $I_{фон}$, вызванного фоновой засветкой, и обратного темнового тока диода $I_{обр}$. Общий дробовой шум, в соответствии с формулой (2.28), равен

$$< i_s^2 >= 2q(I_{изм} + I_{фон} + I_{обр})\Delta f. \qquad (7.27)$$

Выход фотодиода подключается к усилителю, который для уменьшения постоянной времени заряда барьерной емкости обычно дополнительно шунтируется сопротивлением нагрузки. Если эквивалентное сопротивление параллельно включенных сопротивления нагрузки и входного сопротивления усилителя равно $R_{экв}$, то напряжение его теплового шума, пересчитанное

в ток, дает дополнительный вклад в величину шума:

$$< i_a^2 >= (4kT/R_{\text{экв}})\Delta f. \qquad (7.28)$$

Считая, что квантовая эффективность фотодиода (7.26) равна единице, из формул (7.27) и (7.28) легко рассчитать эквивалентную мощность шума *NEP*. Зависимость этой величины от режима работы фотодиода представлена на рис. 7.10.

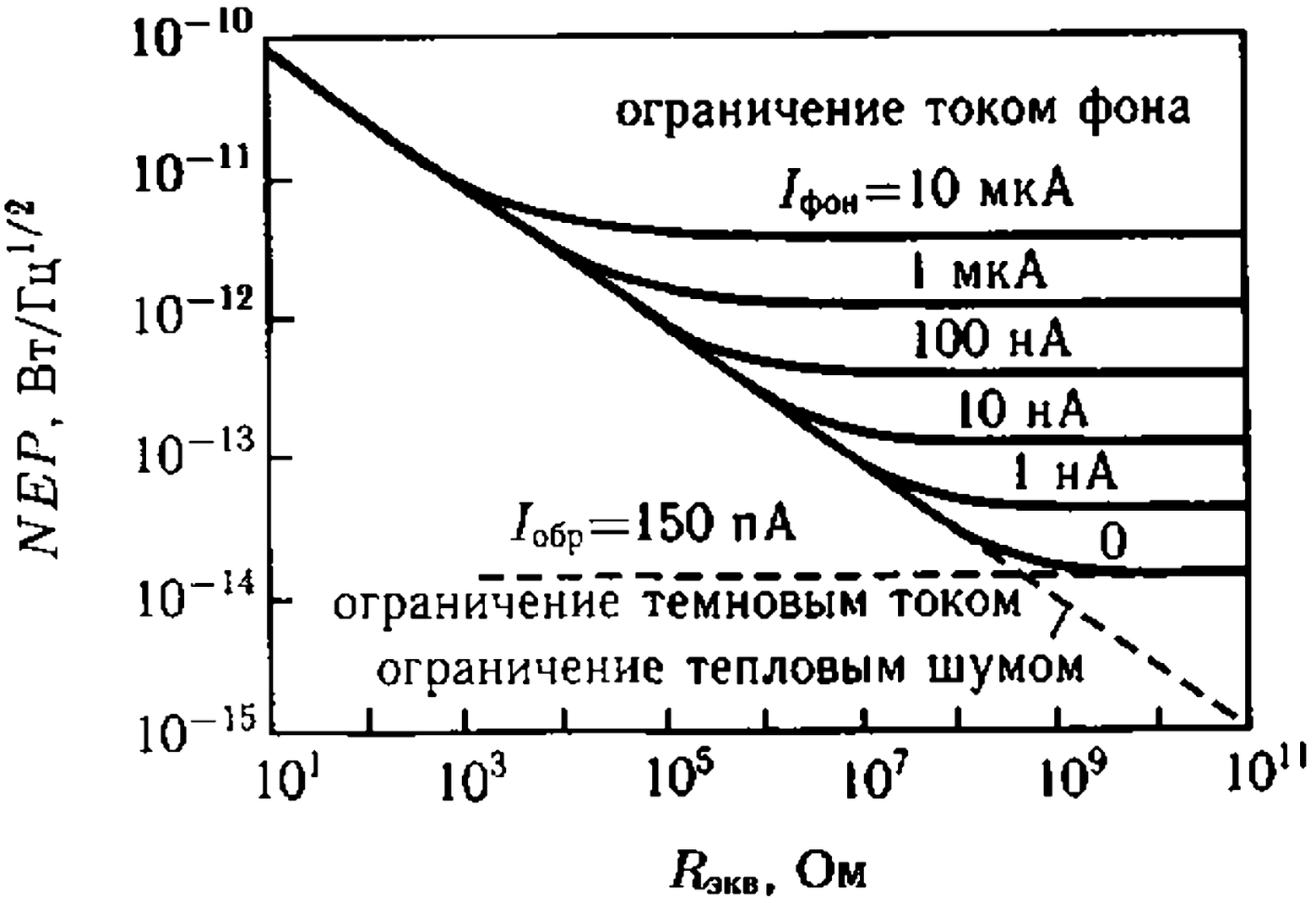


Рис. 7.10. Зависимость эквивалентной мощности шума *NEP* кремниевого фотодиода от сопротивления нагрузки при различных значениях фонового тока [14]

Фотодиоды изготавливают из многих полупроводников. В видимой области спектра широко используются Si, GaAs, GaP, в инфракрасной области — Ge и охлаждаемые приемники из InSb и $Hg_{1-x}Cd_xTe$, в ультрафиолетовой области ($\lambda < 380$ нм) — SiC, GaN, $Al_xGa_{1-x}N$. Большое значение также имеют твердые растворы $In_{1-x}Ga_xAs$ и $In_{1-x}Ga_xAs_{1-y}P_y$, на основе которых создаются быстродействующие фотодиоды для волоконно-оптических линий связи ($\lambda = 1{,}3$–$1{,}55$ мкм).

Фототранзисторы. Фототранзисторы являются функционально-интегрированными приборами, объединяющими в себе функции фотодиода и транзистора (см. рис. 7.11). Как следует из рисунка, фоточувствительной областью фототранзистора является коллекторный p–n-переход. Возникающие при его освещении электронно-дырочные пары разделяются электрическим полем p–n-перехода и фототок дырок I_Φ, направленный в базу транзистора, приводит к открыванию эмиттерного перехода. В результате ток коллектора оказывается

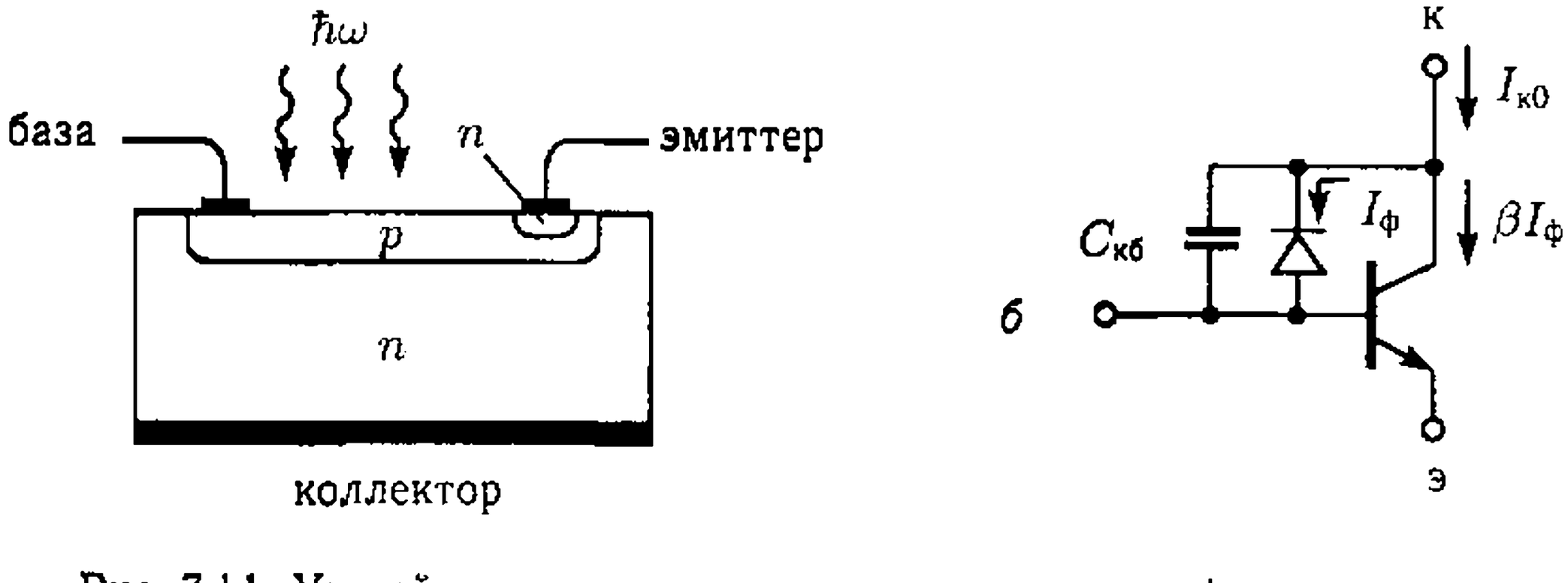


Рис. 7.11. Устройство и эквивалентная схема n–p–n-фототранзистора

в β раз больше вызвавшего его фототока, где β — коэффициент усиления транзистора по току. Из-за довольно большой емкости коллекторного перехода $C_{кб}$ быстродействие фототранзистора сравнительно невелико (время релаксации сигнала измеряется единицами мкс), однако присущее прибору усиление позволяет использовать его, например, в качестве приемного элемента оптрона [1]).

7.1.5. Лавинные фотодиоды. Лавинные фотодиоды представляют собой приборы, в которых за счет лавинного умножения возбужденных светом носителей осуществляется *внутреннее усиление* фототока. В определенном смысле лавинный фотодиод можно рассматривать как твердотельный аналог фотоэлектронного умножителя. Лавинные фотодиоды находят широкое применение в качестве быстродействующих приемников сигналов в системах волоконно-оптической связи.

При использовании лавинных фотодиодов необходимо иметь в виду, что лавинное умножение носителей может приводить к заметному снижению быстродействия фотодиодов. Когда коэффициенты ударной ионизации электронов и дырок близки ($\alpha_n \approx \alpha_p$), что характерно для многих полупроводников группы $A^{III}B^{V}$, вторичные электроны и дырки, рожденные путем ударной ионизации первичного электрона (дырки), еще долго остаются в области лавинного умножения после того, как вызвавший первичную ионизацию носитель покинет эту область. Это происходит потому, что электроны и дырки движутся в *разные* стороны, и характерное время затухания лавины во много раз превосходит

---

[1]) Оптрон — оптоэлектронный прибор, который состоит из полупроводниковых источника света и фотоприемника, связанных между собой световодом. Сигнал от светодиода к фотодиоду передается в оптроне с помощью потока света, что обеспечивает гальваническую развязку входа и выхода.

время пролета области сильного поля первичным носителем. Расчеты показывают, что при $\alpha_n \approx \alpha_p$ и большом коэффициенте умножения M произведение коэффициента умножения на ширину полосы пропускания фотодиода остается постоянным и равно

$$M \times \Delta f \approx 3/(2\pi t_{\text{ср}}), \tag{7.29}$$

где $t_{\text{ср}}$ — среднее время пролета электронами и дырками области умножения [14].

Чтобы сохранить высокое быстродействие при большом коэффициенте умножения M, необходимо использовать полупроводники, у которых $\alpha_n \ll \alpha_p$ или $\alpha_n \gg \alpha_p$. Действительно, если лавинное умножение осуществляется преимущественно электронами, а ударной ионизацией дырок можно пренебречь, то умножаться будут только электроны и задержка, связанная с удалением вторичных электронов из области сильного поля, будет минимальной. Это позволяет создавать лавинные фотодиоды из Si с полосой пропускания 30 ГГц и произведением $M \times \Delta f$, достигающим 315 ГГц.

Другой особенностью лавинных фотодиодов является специфический характер формирования шума при лавинном умножении носителей. Поскольку ударная ионизациякоторую мы рассматривали в п. 1.3.1, по своей природе является статистическим процессом и число электронно-дырочных пар, рождаемых каждым первичным электроном (дыркой), испытывает статистические флуктуации, то понятно, что в процессе лавинного умножения должен появляться избыточный шум. Этот шум принято характеризовать величиной шум-фактора $F = \langle M^2 \rangle / \langle M \rangle^2$. Расчеты показывают, что величина F зависит от коэффициента умножения M и соотношения коэффициентов ударной ионизации α_n и α_p (см. рис. 7.12). Если коэффициенты ударной ионизации

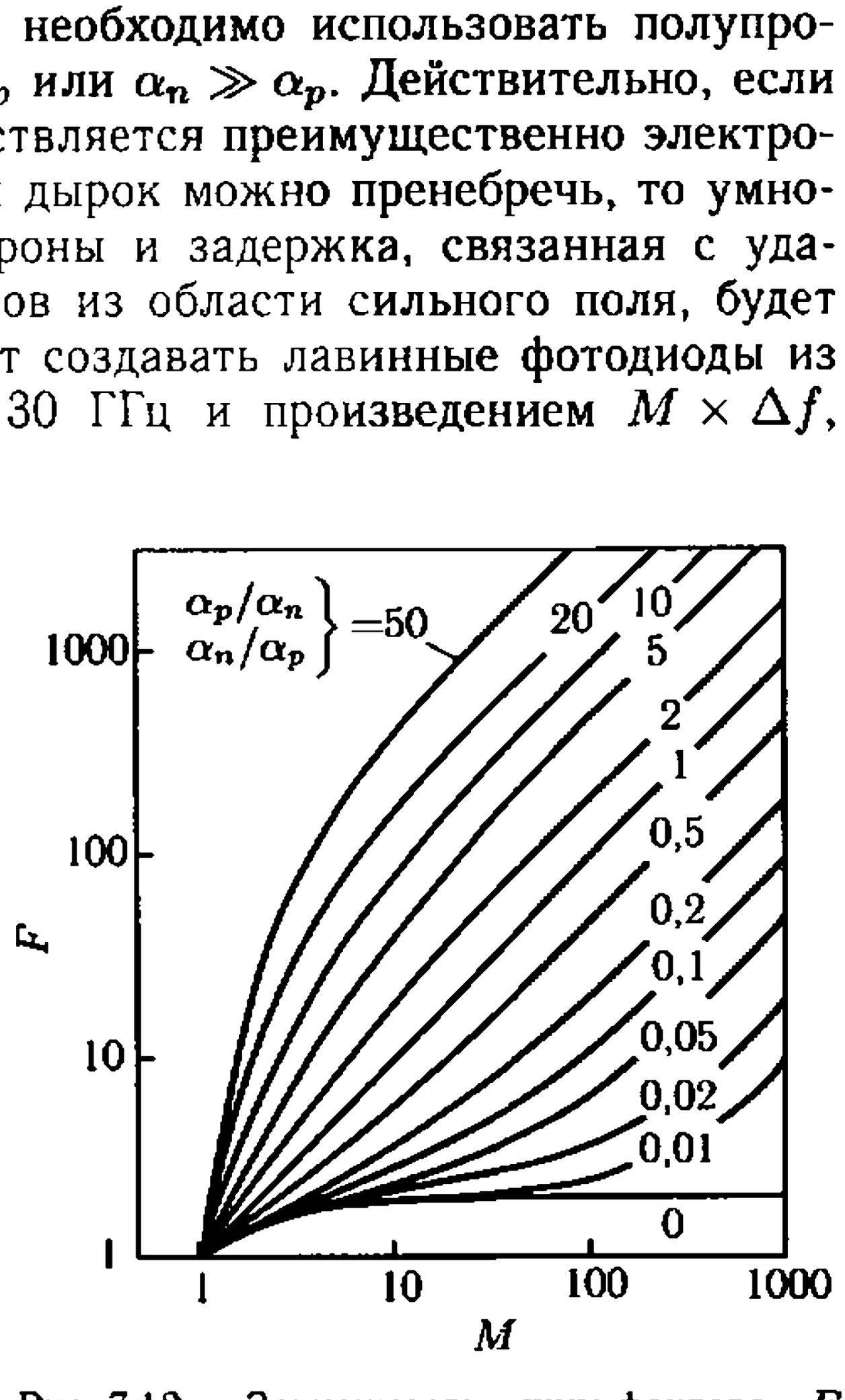


Рис. 7.12. Зависимость шум-фактора F от коэффициента умножения M для различных значений соотношений α_n/α_p и α_p/α_n [14]

электронов и дырок сильно различаются, в этом случае вклад в шум-фактор вносит лишь флуктуация величины M для одного типа носителей. Если же коэффициенты ударной ионизации электронов и дырок близки и оба типа носителей дают сравнимый вклад в умножение носителей, то флуктуация числа рождаемых электронами носителей умножается на флуктуацию числа рождаемых дырками носителей, что и вызывает большую относительную флуктуацию величины M (шум-фактор оказывается существенно выше). Поэтому для получения высокого отношения сигнал/шум в лавинных фотодиодах их стремятся изготовить из полупроводников с сильно различающимися α_n и α_p. На с. 120 мы уже обсуждали, как использование сверхрешеток может помочь изменить отношение α_n/α_p.

Для получения низкого значения F необходим не только соответствующий выбор полупроводника, но и правильная конструкция лавинного фотодиода. Расчеты показывают, что если в область сильного поля инжектируются только электроны, то шум-фактор равен [14]

$$F = \frac{\alpha_p}{\alpha_n} \cdot M + \left(2 - \frac{1}{M}\right)\left(1 - \frac{\alpha_p}{\alpha_n}\right). \qquad (7.30)$$

Из этой формулы следует, что для получения низкого шум-фактора необходимо, чтобы α_p/α_n было много меньше 1. Аналогичное (7.30) выражение с заменой отношения α_p/α_n на α_n/α_p может быть записано для случая, когда в область сильного поля инжектируются только дырки. Поэтому если для изготовления лавинного фотодиода выбран полупроводник с $\alpha_n \gg \alpha_p$, то для получения наименьшего значения F фоточувствительная область фотодиода должна располагаться так, чтобы в область умножения поступали электроны. В случае $\alpha_n \ll \alpha_p$ такими носителями должны быть дырки.

В последнее время при разработке сверхбыстродействующих лавинных фотодиодов разработчики столкнулись с трудностью получения высокой квантовой эффективности в таких диодах. Дело в том, что для создания приборов с полосой пропускания более 20 ГГц необходимо, чтобы толщина фоточувствительной области не превышала ~0,5 мкм. Даже при использовании прямозонных полупроводников, для которых характерен высокий коэффициент поглощения света, при одном проходе света через структуру доля поглощенных квантов оказывается небольшой. Чтобы преодолеть этот недостаток, была предложена конструкция лавинных фотодиодов с *оптическим резонатором*.

Для получения высокого коэффициента отражения ($\geqslant 95\%$) «зеркал» этого резонатора, их изготавливают в виде многослойной структуры из различных полупроводников (например, $Si_{0,7}Ge_{0,3}/Si$). Два таких зеркала заставляют падающее излучение многократно пройти через поглощающий слой и тем самым позволяют увеличить квантовую эффективность фотодиода. О различных подходах и достижениях в создании современных быстродействующих лавинных фотодиодов можно прочитать в обзорах [288, 289].

Рис. 7.13. Зависимость мощности сигнала и шума на выходе лавинного фотодиода от коэффициента умножения M и выбор оптимальных условий для получения наибольшего отношения сигнал/шум. На верхней шкале указано напряжение смещения на фотодиоде, необходимое для получения заданного коэффициента умножения [263]

Преимущества, которые дает использование лавинных фотодиодов, могут быть легко поняты из анализа экспериментальной кривой, показанной на рис. 7.13. Поскольку лавинные фотодиоды обычно используются для регистрации быстропротекающих процессов, то для получения высокого быстродействия их подключают к усилителям с низким входным сопротивлением (для получения низких значений постоянной времени заряда барьерной емкости p–n-перехода). В этой ситуации, в соответствии с рис. 7.10, основным источником шума фотоприемного устройства становится шум усилителя (шум сопротивления нагрузки). Поэтому, чтобы увеличить отношение сигнал/шум, необходимо использовать внутреннее усиление сигнала в лавинном фотодиоде. Пока уровень шума лавинного умножения мал по сравнению с шумами усилителя, отношение сигнал/шум быстро возрастает с увеличением M. Когда шумы лавинного умножения

становятся сравнимыми с шумами усилителя, отношение сигнал/шум проходит через максимум и при дальнейшем увеличении M начинает падать, поскольку шумы лавинного умножения возрастают с увеличением M обычно быстрее, чем сигнал. Оптимальным коэффициентом умножения $M_{\text{опт}}$ считают такое значение, при котором мощности шумов усилителя и лавинного фотодиода становятся равными (см. рис. 7.13). Лавинные фотодиоды изготавливаются не только из Si, но и из других полупроводников (Ge, GaAs, InAs, InSb, $Hg_{1-x}Cd_xTe$).

7.1.6. Преобразователи солнечной энергии (солнечные элементы). В связи с ростом потребления энергии и ограниченностью запасов традиционных сжигаемых видов топлива в настоящее время ведется поиск альтернативных источников энергии, среди которых важное место занимает ее практически неистощимый и экологически абсолютно безопасный источник — Солнце. Согласно существующим прогнозам, одним из основных преобразователей солнечной энергии со временем могут стать солнечные батареи, которые позволяют с высокой эффективностью преобразовывать свет непосредственно в электрический ток и при этом не требуют больших эксплуатационных расходов [290].

Принципиальная возможность прямого преобразования света в электрический ток была впервые продемонстрирована в 1839 г. Беккерелем, который наблюдал возникновение фото-ЭДС при освещении покрытой хлоридом серебра платиновой проволоки, помещенной в разбавленный раствор азотной кислоты. В 1876 г. Адамс и Дэй наблюдали появление фото-ЭДС при освещении контакта селена с платиной, а первый солнечный элемент на основе барьера Au–Se был создан Фриттсом в 1883 г. К 1914 г. на селеновых элементах с барьерным слоем из окиси меди была получена эффективность преобразования около 1%. В 1940 г. Рассел Ол, работавший в Bell Laboratories, создал первый солнечный элемент на p–n-переходе из кремния. Современная эра полупроводниковых солнечных элементов началась в 1954 г., когда Чапен, Фуллер и Пирсон из той же лаборатории на неглубоком диффузионном p–n-переходе, созданном на поверхности кремния, получили к.п.д. преобразования 6% [291]. В том же 1954 г. были созданы и первые солнечные элементы на основе барьеров Шоттки к CdS с эффективностью преобразования 6% [292]. В 1958 г. были запущены первые космические спутники «Спутник-3» (СССР) и «Авангард-1» (США), питание которых производилось от кремниевых солнечных батарей. К 1960 г. к.п.д. кремниевых элементов достиг 14%.

Распределение энергии солнечного излучения по спектру показано на рис. 7.14. Кривая, обозначенная AM0, соответствует спектру излучения Солнца за пределами земной атмосферы.

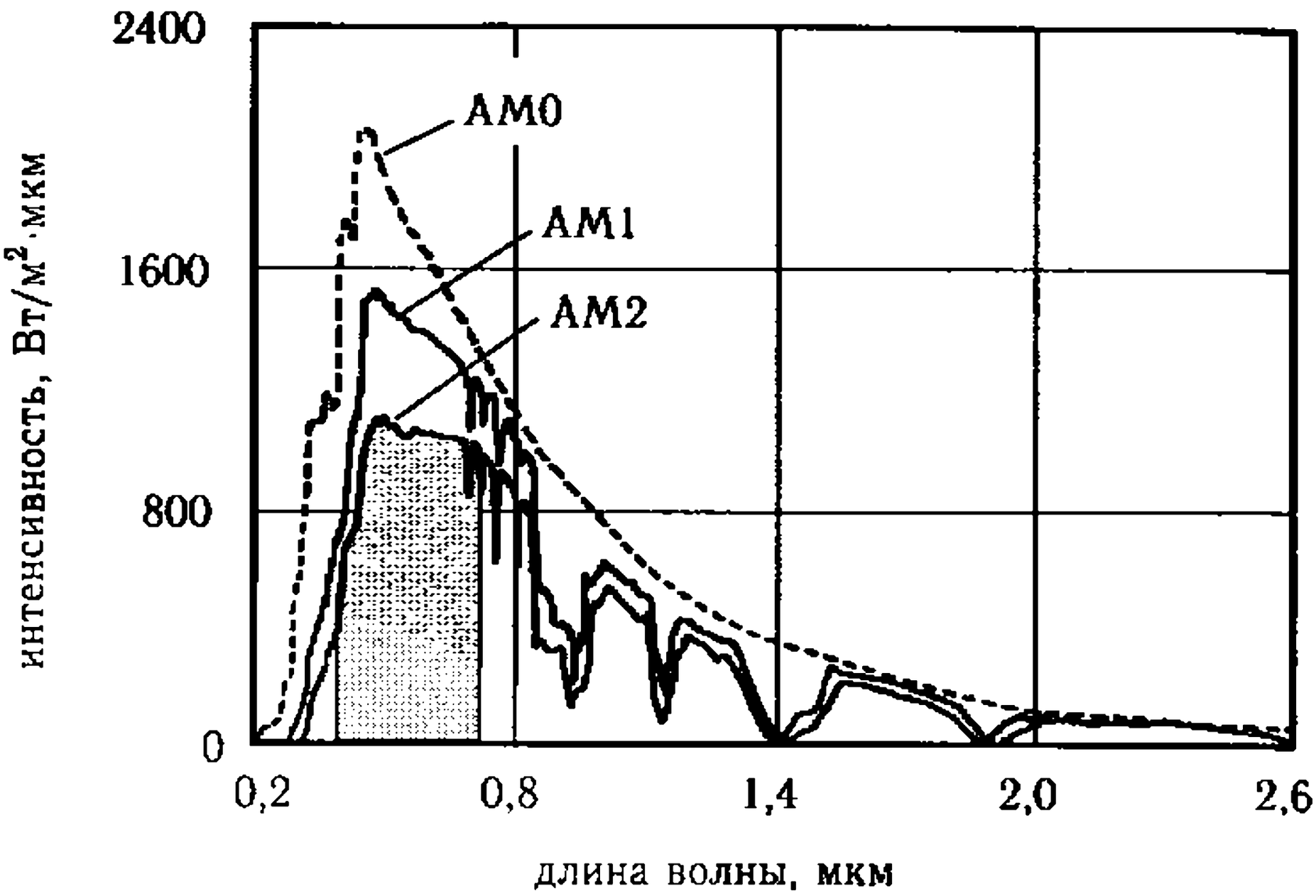


Рис. 7.14. Распределение солнечной энергии по спектру при различной толщине «воздушной массы». Заштрихованная область отвечает области видимого излучения [14]

При прохождении солнечного света через атмосферу его энергия немного ослабляется за счет поглощения излучения парами воды в инфракрасной области, поглощению ультрафиолетового излучения озоном и рассеянию света находящимися в атмосфере частицами пыли и аэрозолями. Изменение интенсивности солнечного излучения при прохождении атмосферы характеризуется «толщиной воздушной массой» (AM), которая численно равна секансу угла между положением Солнца и зенитом. Приведенные на рис. 7.14 спектры AM1 и AM2 отвечают спектральному распределению энергии Солнца, находящегося в зените и под углом 30° к горизонту. Полная мощность солнечного излучения при AM1 составляет 92,5 мВт/см^2, при AM2 — 69,1 мВт/см^2.

Фотовольтаические приемники, рассмотренные нами в п. 7.1.4, идеально подходят для создания фотоэлектрических преобразователей солнечной энергии в электрический ток. Однако при этом в конструкцию этих приборов требуется ввести изменения, необходимые для получения высокого к.п.д.

Поскольку свет, проникающий в полупроводник, создает там электронно-дырочные пары, которые разделяются электрическим полем p–n-перехода и создают фотоэлектрический ток, вольт-амперную характеристику фотоэлектрического преобразователя можно представить как вольт-амперную характеристику p–n-перехода, смещенную по оси ординат на величину

фотоэлектрического тока (см. рис. 7.15) [1]):

$$I = I_s\left[\exp\left(\frac{qV}{kT}\right) - 1\right] - I_L. \qquad (7.31)$$

Здесь I — ток на выходе преобразователя, I_L — фототок, получаемый от фотопреобразователя в режиме короткого замыкания и определяемый долей поглощенной полупроводником солнечной энергии, а I_s — ток насыщения p–n-перехода (см. (1.31)). Напряжение холостого хода на выходе фотопреобразователя может быть найдено из (7.31), если положить $I = 0$:

$$V_{xx} \approx \frac{kT}{q}\ln\left(\frac{I_L}{I_s}\right). \qquad (7.32)$$

Площадь заштрихованного прямоугольника на рис. 7.15 показывает, какую полезную мощность электрической энергии можно получить от такого преобразователя. Эта мощность может быть рассчитана из условия максимума произведения $V \cdot I$, в котором V и I связаны уравнением (7.31). Нетрудно показать, что напряжение и ток на выходе преобразователя, отвечающие максимальной выходной мощности, являются решениями следующего трансцендентного уравнения:

$$V_m = V_{xx} - \frac{kT}{q}\ln\left(1 + \frac{qV_m}{kT}\right), \text{ а } I_m = I_L\left(1 + \frac{kT}{qV_m}\right)^{-1}. \qquad (7.33)$$

Выходная мощность фотопреобразователя при оптимальной нагрузке может достигать 80 % от произведения $V_{xx}I_L$ (отношение $V_mI_m/V_{xx}I_L$ называется *фактором заполнения*).

Изменение температуры оказывает заметное влияние на к.п.д. солнечных элементов. С одной стороны, с ростом температуры диффузионные длины возрастают, что увеличивает эффективность собирания носителей и увеличивает I_L. С другой стороны, увеличение температуры вызывает экспоненциально быстрое возрастание тока насыщения (см. п. 1.2.1), в результате чего, в соответствии с уравнением (7.32), напряжение холостого хода уменьшается. Эксперимент показывает, что второй фактор преобладает, и в солнечных элементах из Si и GaAs к.п.д. преобразования уменьшается с ростом температуры со скоростью ≈0,4 %/град.

---

[1]) Мы используем эту формулу полагая, что при высоком прямом смещении вольт-амперная характеристика p–n-перехода уже определяется током инжекции.

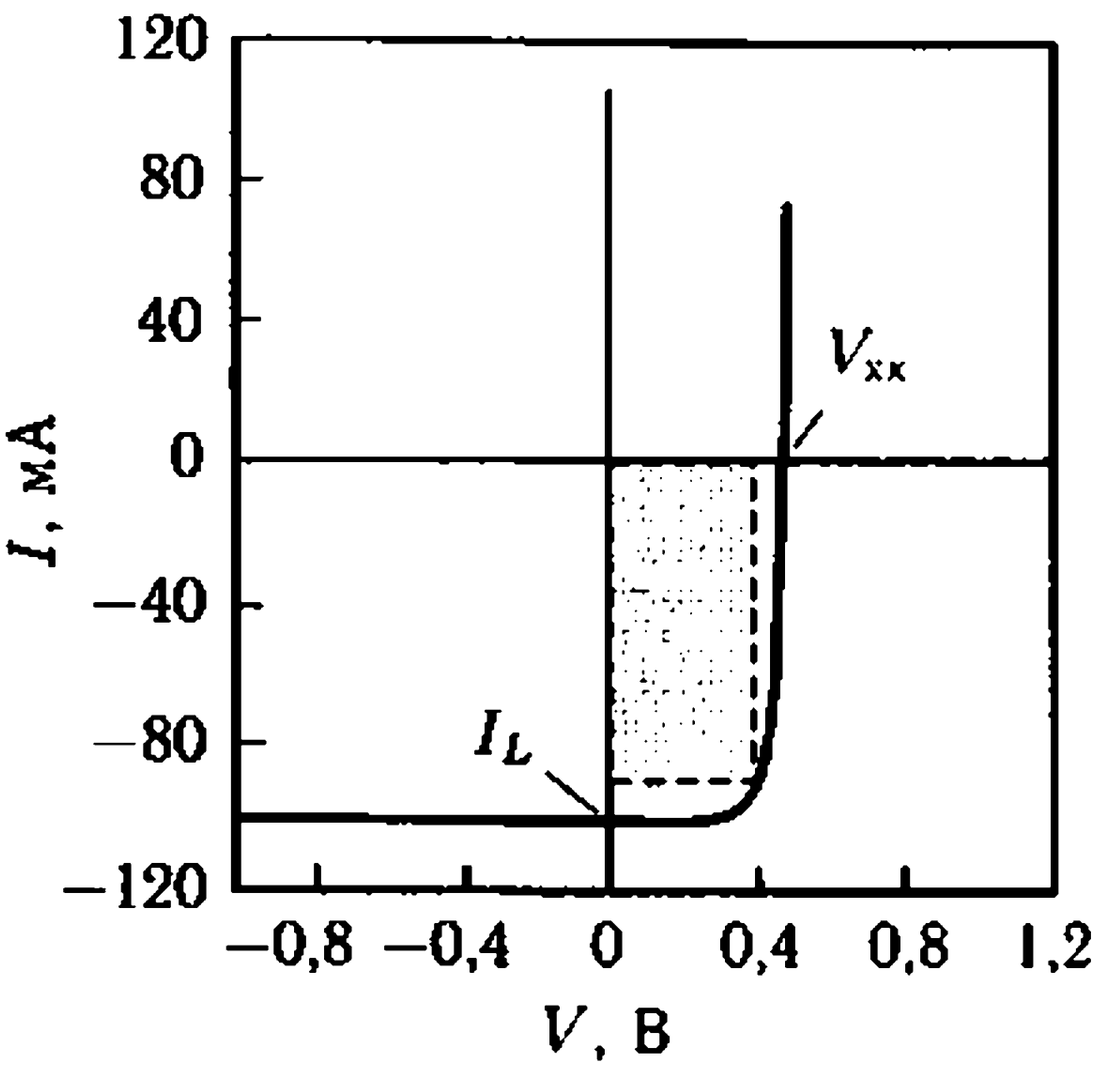


Рис. 7.15. Вольт-амперная характеристика освещенного солнечного элемента. Площадь заштрихованного прямоугольника равна максимальной выходной мощности, которую можно получить от данного преобразователя.

Чтобы определить, из какого полупроводника наиболее выгодно изготавливать преобразователи солнечной энергии, необходимо рассчитать теоретический к.п.д. преобразователя как функцию ширины запрещенной зоны полупроводника E_g. Ясно, что чем больше E_g, тем меньше будет плотность тока насыщения J_s и тем выше будет V_{xx}. С другой стороны, плотность фототока J_L уменьшается с ростом E_g, так как доля квантов в спектре солнечного излучения, которые могут создавать электронно-дырочные пары, уменьшается при сдвиге края собственного поглощения в коротковолновую область. Считая, что ток насыщения p–n-перехода описывается уравнением (1.31), для известного распределения энергии в спектре излучения Солнца можно рассчитать зависимость теоретического к.п.д. преобразователя солнечной энергии от ширины запрещенной зоны полупроводника. Эта зависимость, рассчитанная для элемента, работающего при 300 К при освещении солнечным излучением через воздушную массу толщиной AM1,5, показана на рис. 7.16 [14]. Видно, что для изготовления солнечных элементов наиболее подходят полупроводники с шириной запрещенной зоны в интервале 1–1,6 эВ. Максимальная теоретическая эффективность преобразования, равная 31 %, достигается при $E_g = 1,35$ эВ.

Концентрация солнечного излучения с помощью зеркальных или линзовых оптических систем позволяет за счет увеличения напряжения холостого хода увеличить к.п.д. преобразователя теоретически до 37 % (см. рис. 7.16).

Дальнейшее повышение к.п.д. возможно при использовании *каскадных фотопреобразователей*, изготовленных из освещаемых «насквозь» последовательно включенных p–n-переходов из нескольких полупроводников с разной шириной запрещенной зоны (см. рис. 7.17). Каждая из секций такого преобразователя осуществляет фотоэлектрическое преобразование определенного участка спектра излучения

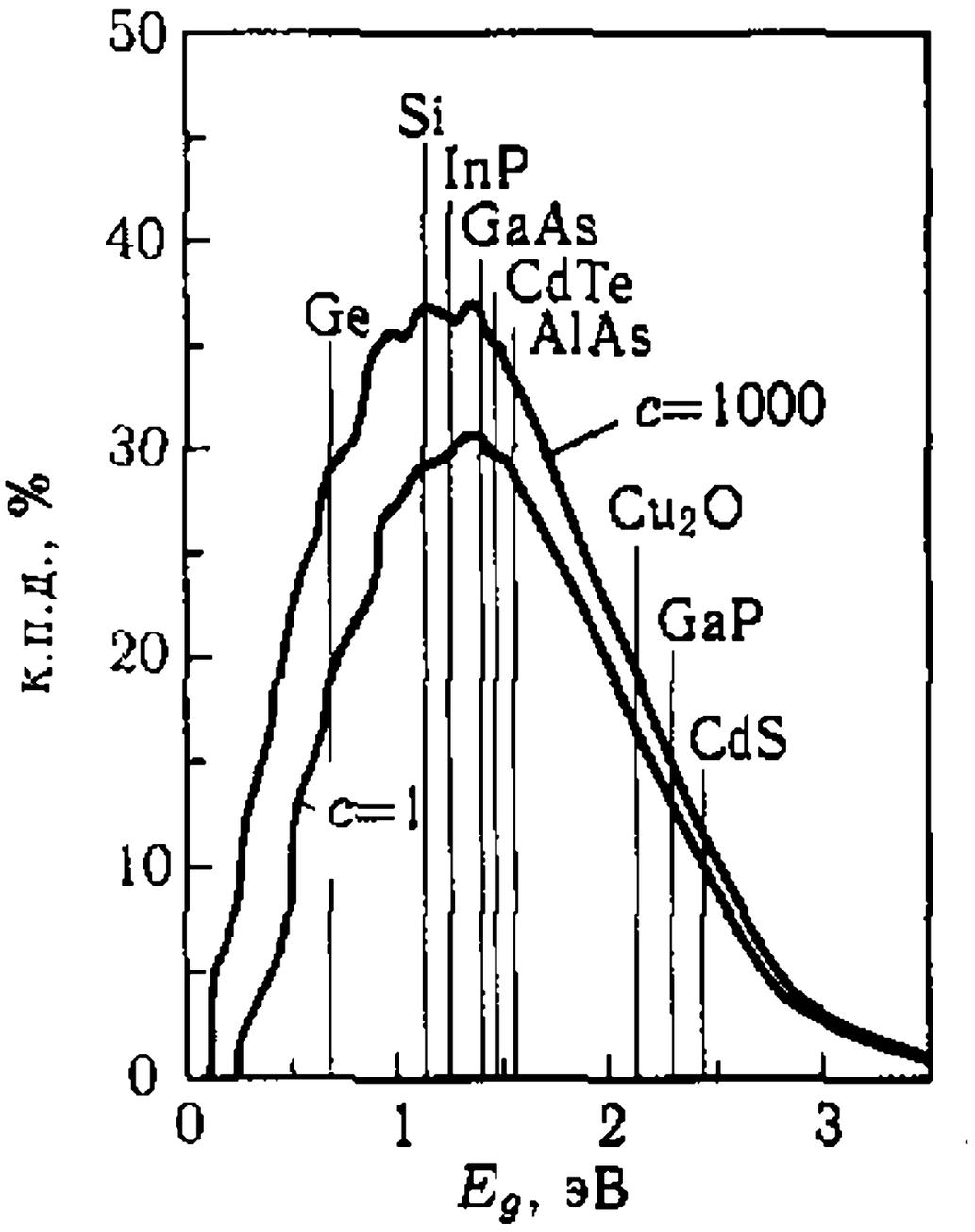


Рис. 7.16. Теоретический к.п.д. солнечных элементов, изготовленных из разных полупроводников, при естественном освещении (c=1) и 1000-кратной концентрации солнечного излучения (c = 1000). Толщина воздушной массы равна AM1,5, T = 300 К [14]

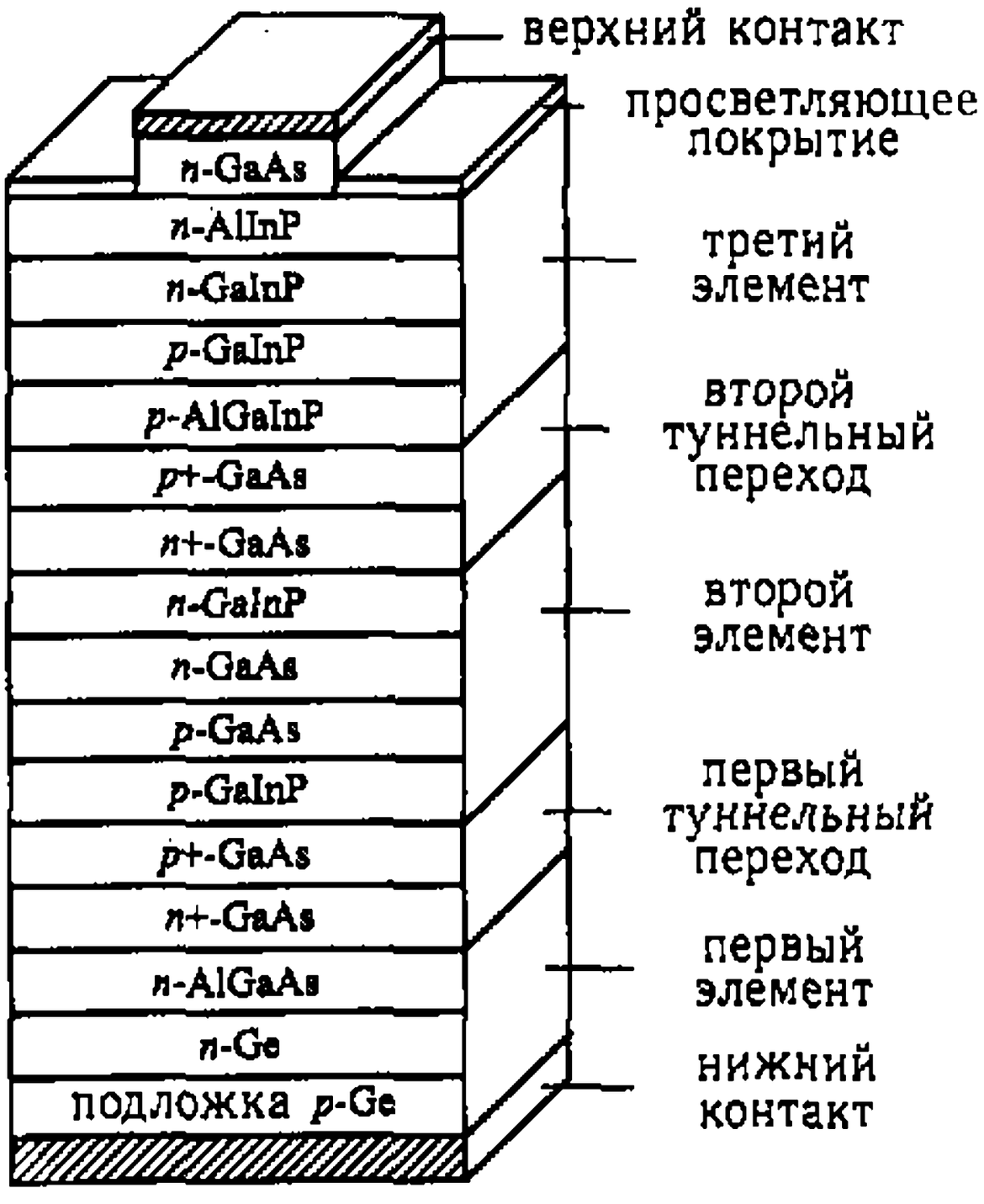


Рис. 7.17. Устройство трехкаскадного фотопреобразователя [293]

Солнца, при этом ЭДС последовательно включенных каскадов суммируется. Для электрического соединения отдельных каскадов используются туннельные p–n-переходы. Теоретический к.п.д. преобразования для двухкаскадного солнечного элемента составляет 50%, трехкаскадного — 56% [14]. В настоящее время на двух- и трехкаскадных монолитных солнечных элементах со структурами $In_{0,5}Ga_{0,5}P/GaAs$ и $In_{0,44}Ga_{0,56}P/In_{0,08}Ga_{0,92}As/Ge$ получены к.п.д. преобразования 30,3% и 32,0% при нормальном (неконцентрированном) освещении [293, 294]. В условиях концентрированного освещения на структуре InGaP/GaAs/Ge в июле 2003 года достигнут к.п.д. преобразования 36,9% [295], а в декабре 2006 г. эта цифра возросла до 40,7 %.

Максимальный к.п.д. преобразования, полученный на однокаскадных солнечных элементах, в условиях нормального освещения (AM1,5), в настоящее время составляет 24,7% для монокристаллического Si и 25,1% для GaAs [293, 294], что оказывается несколько ниже теоретического значения. Связано это с тем, что вольт-амперные характеристики реальных p–n-переходов отличаются от идеальных. Во-первых, как мы отмечали в п. 1.2.2, в полупроводниках с $E_g > 1$ эВ существенное влияние на вольт-амперные характеристики начинает оказывать рекомбинация в области пространственного заряда. Из-за этого напряжение холостого хода фотопреобразователя оказывается ниже, чем в идеальном случае, а сама вольт-амперная характеристика — более пологой, что дополнительно снижает величину напряжения V_m в режиме оптимальной нагрузки. Во-вторых, при большом токе нагрузки напряжение на выходе элемента дополнительно уменьшается из-за падения напряжения на сопротивлении контактов.

Конструкция простейшего солнечного элемента на основе p–n-перехода показана на рис. 7.18. Для получения высокого напряжения холостого хода (которое не может превышать контактную разность потенциалов ϕ_k, см. п. 1.1) и для обеспечения

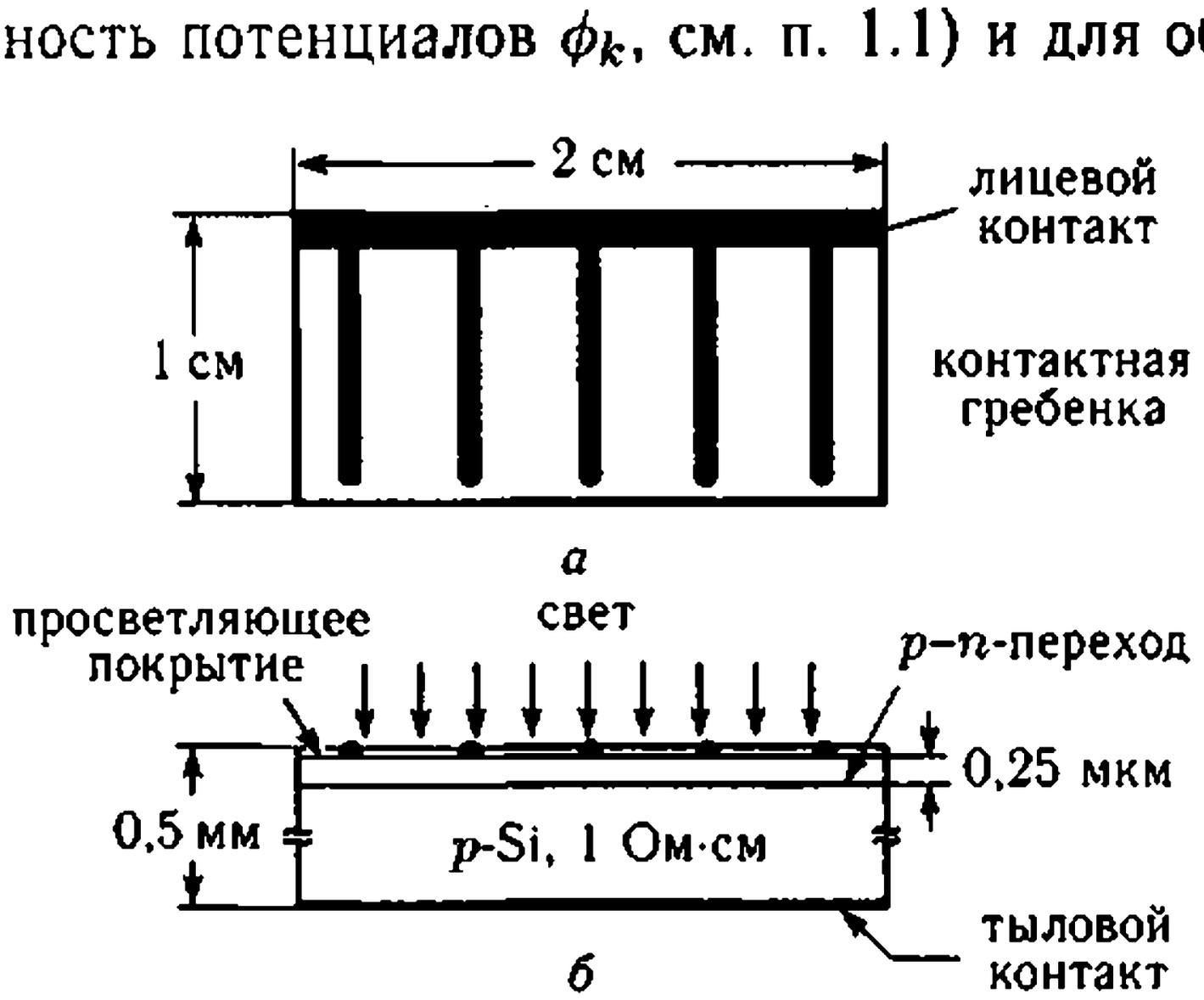


Рис. 7.18. Устройство простейшего кремниевого солнечного элемента с p–n-переходом и лицевым контактом в виде контактной гребенки к сильно легированному поверхностному слою

низкого сопротивления лицевого контакта поверхностный слой всегда легируется сильно. Контактная гребенка, создаваемая на поверхности солнечного элемента и обычно затеняющая 5–10% поверхности элемента, должна занимать как можно

меньшую площадь, обеспечивая при этом достаточно низкое сопротивление контакта. Поверхностный легированный слой намеренно делается очень тонким (0,1–0,5 мкм), чтобы элемент сохранял высокую эффективность преобразования и в коротковолновой части спектра излучения Солнца, которое поглощается в очень тонком слое полупроводника (сильно легированный слой характеризуется невысоким временем жизни и поэтому носители, рождаемые вблизи поверхности полупроводника, наиболее сильно подвержены поверхностной рекомбинации). Для уменьшения отражения света от границы раздела кремний–воздух на поверхности солнечных фотопреобразователей создается просветляющее покрытие (из SiO_2, Si_3N_4, Ta_2O_5) или поверхности элемента придается рифленая форма. В последнем случае геометрия бороздок на поверхности подбирается так, чтобы поток света, отраженный от поверхности элемента, вновь направлялся на солнечный элемент; это обеспечивает более полное поглощение света [14].

Отдельный кремниевый солнечный элемент площадью 2 $см^2$ развивает напряжение холостого хода $V_{xx} = 0{,}5$–$0{,}6$ В и имеет ток короткого замыкания 30–80 мА при освещении AM1,5; последовательно-параллельное соединение таких элементов в солнечную батарею позволяет получить от нее существенно бóльшие напряжение и ток.

При использовании солнечных элементов в условиях космоса особое значение приобретает вопрос об их радиационной стойкости. Дело в том, что облучение этих элементов частицами высоких энергий приводит к рождению в полупроводнике радиационных дефектов. Влияние этих дефектов на время жизни можно с хорошей точностью описать уравнением

$$\frac{1}{\tau} = \frac{1}{\tau_0} + K\Phi, \tag{7.34}$$

где τ и τ_0 — время жизни в облученном и необлученном материале, Φ — поглощенная доза радиации, а K — константа, зависящая от вида частиц. Уменьшение времени жизни приводит к уменьшению диффузионной длины (то есть уменьшению эффективности собирания электронно-дырочных пар) и увеличению тока насыщения и тока рекомбинации в области пространственного заряда. Оба эти фактора вызывают уменьшение выходной мощности солнечных элементов (их деградацию).

В работе [296] было показано, что повысить радиационную стойкость кремниевых p–n-переходов можно дополнительно легируя кристалл примесью лития, которая легко диффундирует

в Si и образует комплексы с радиационными дефектами. Эффективно нейтрализуя эти дефекты, литий препятствует деградации времени жизни. Для уменьшения повреждаемости солнечных батарей, работающих в условиях космоса, поверхность элементов дополнительно защищается с помощью специального покрытия.

Альтернативным вариантом солнечных элементов, пригодных для использования в условиях космоса, являются солнечные элементы из GaAs и гетероструктур AlGaAs/GaAs. Более высокая радиационная стойкость этих элементов по сравнению с кремнием, связанная с более высоким коэффициентом поглощения в прямозонных полупроводниках и возможности за счет этого уменьшить эффективную толщину активной области прибора, позволила, в частности, использовать такие солнечные элементы для питания орбитальной станции «Мир», проработавшей в космосе около 15 лет [293].

Для увеличения к.п.д. кремниевых солнечных элементов, чувствительность которых быстро уменьшается в синей области спектра, в 70-е годы была предложена конструкция так называемого *люминесцентного концентратора*. Идея этого устройства довольно проста: если на поверхность солнечного элемента нанести слой хорошо люминесцирующего органического красителя, который поглощал бы синее и ультрафиолетовое излучение и переизлучал бы свет в области, в которой чувствительность элемента заметно выше, то удалось бы более эффективно преобразовывать эту коротковолновую часть солнечного излучения в электрический ток. Однако реализации этой идеи мешала недостаточная светостойкость органических красителей. В 2000 г. авторы работы [297] предложили использовать в люминесцентных концентраторах полупроводниковые *квантовые точки*. Преимуществом квантовых точек по сравнению с органическими красителями является их стабильность, возможность перестройки энергии излучения подбором размера точек, бóльшая разность энергий поглощенного и излученного квантов. Квантовые точки CdSe/CdS вполне подходят для этой цели, так как их квантовый выход люминесценции превышает 80%. Первые работы по использованию люминесцентных концентраторов из квантовых точек позволили получить относительное увеличение к.п.д. солнечного элемента на 6–8%, а максимальное ожидаемое увеличение составляет 20–30% [298].

Широкому использованию фотоэлектрических преобразователей солнечной энергии препятствует их сравнительно высокая стоимость.

Поэтому поиску и изучению возможности использования в солнечной энергетике новых (в том числе и некристаллических) полупроводников уделяется большое внимание [14, 52, 57]. Среди этих материалов большое место занимает *гидрированный аморфный кремний* a-Si:H, к.п.д. солнечных элементов из которого достигает 10% в структурах большой площади. Более высокие значения коэффициента поглощения в аморфном кремнии по сравнению с кристаллическим кремнием позволяет эффективно поглощать солнечный свет уже в тонких пленках (порядка 1 мкм), что существенно снижает стоимость солнечных элементов. Аморфный кремний технологичен: p–i–n-структуры из него могут быть сравнительно легко получены методом плазменного разложения силана с осаждением тонкой пленки прямо на поверхность стальных листов большой площади или непрерывно движущейся стальной ленты. Хотя физика аморфного кремния сильно отличается от физики кристаллических полупроводников (подробнее см. [299]) и многие явления в a-Si:H (в частности, явления деградации под действием освещения) еще до конца не изучены, солнечные элементы из этого материала остаются альтернативой солнечным элементам из кристаллического кремния и уже сегодня находят практическое применение.

Кроме аморфного кремния, другими материалами, используемыми в настоящее время для промышленного выпуска тонкопленочных солнечных элементов, являются поликристаллический кремний, а также CdTe и твердые растворы на основе $CuInSe_2$ (последние два материала являются прямозонными полупроводниками). На элементах из твердого раствора $CuIn_{1-x}Ga_xSe_2$ ($x \approx 0,3, E_g \approx 1,2$ эВ) недавно получен к.п.д. преобразования 19,2%, который приближается к к.п.д. солнечных элементов из поликристаллического кремния (19,8%). В элементах из CdTe получен максимальный к.п.д. 16,5%. Лидером в производстве тонкопленочных солнечных элементов является Япония. Фотохимические преобразователи солнечной энергии на основе нанокристаллического TiO_2 и органических красителей имеют в настоящее время к.п.д. 11% [294]. Тенденции и перспективы дальнейшего развития гелиоэнергетики обсуждаются в обзорах [293, 295].

За последние тридцать лет стоимость солнечных элементов из кристаллического кремния уменьшилась в 10 раз и сейчас составляет менее 3 долл. за ватт пиковой мощности. Это означает, что если такой элемент будет работать в течение всего гарантийного срока (20 лет) при средней продолжительности солнечного освещения 2000 часов в год (типичное значение для Европы [14]), то без учета других затрат удельная стоимость производства электроэнергии составит около 0,08 долл. за киловатт-час, что приближается к стоимости производства электроэнергии другими способами (реальная себестоимость киловатт-часа электроэнергии, произведенной солнечными элементами сейчас составляет 0,2–0,5 долл.). Полная мощность производимой с помощью солнечных элементов электроэнергии в 2005 г. составила 1656 МВт; ведущими производителями этой энергии являются Германия и Япония.

7.1.7. Детекторы ядерных излучений. Специфической особенностью ядерных излучений, для регистрации которых используются полупроводниковые детекторы, является высокая проникающая способность этих излучений, и, следовательно, их низкий коэффициент поглощения в полупроводнике. Поэтому линейные размеры детекторов ядерных излучений могут достигать нескольких сантиметров, а объем — нескольких сотен $см^3$.

Работа полупроводниковых детекторов ядерных излучений основана на собирании электронно-дырочных пар, создаваемых при прохождении ионизирующих частиц высокой энергии через полупроводник. Энергия ионизирующей частицы обычно на 3–7 порядков превышает ширину запрещенной зоны полупроводника, и поэтому в результате ее взаимодействия с веществом образуется огромное число электронно-дырочных пар. Если регистрируемыми частицами являются заряженные частицы (электроны, протоны, альфа-частицы, высокоэнергичные ионы, заряженные π-мезоны), то ионизация полупроводника осуществляется самими частицами. Если регистрации подлежат рентгеновские или гамма-кванты, то в этом случае ситуация несколько меняется.

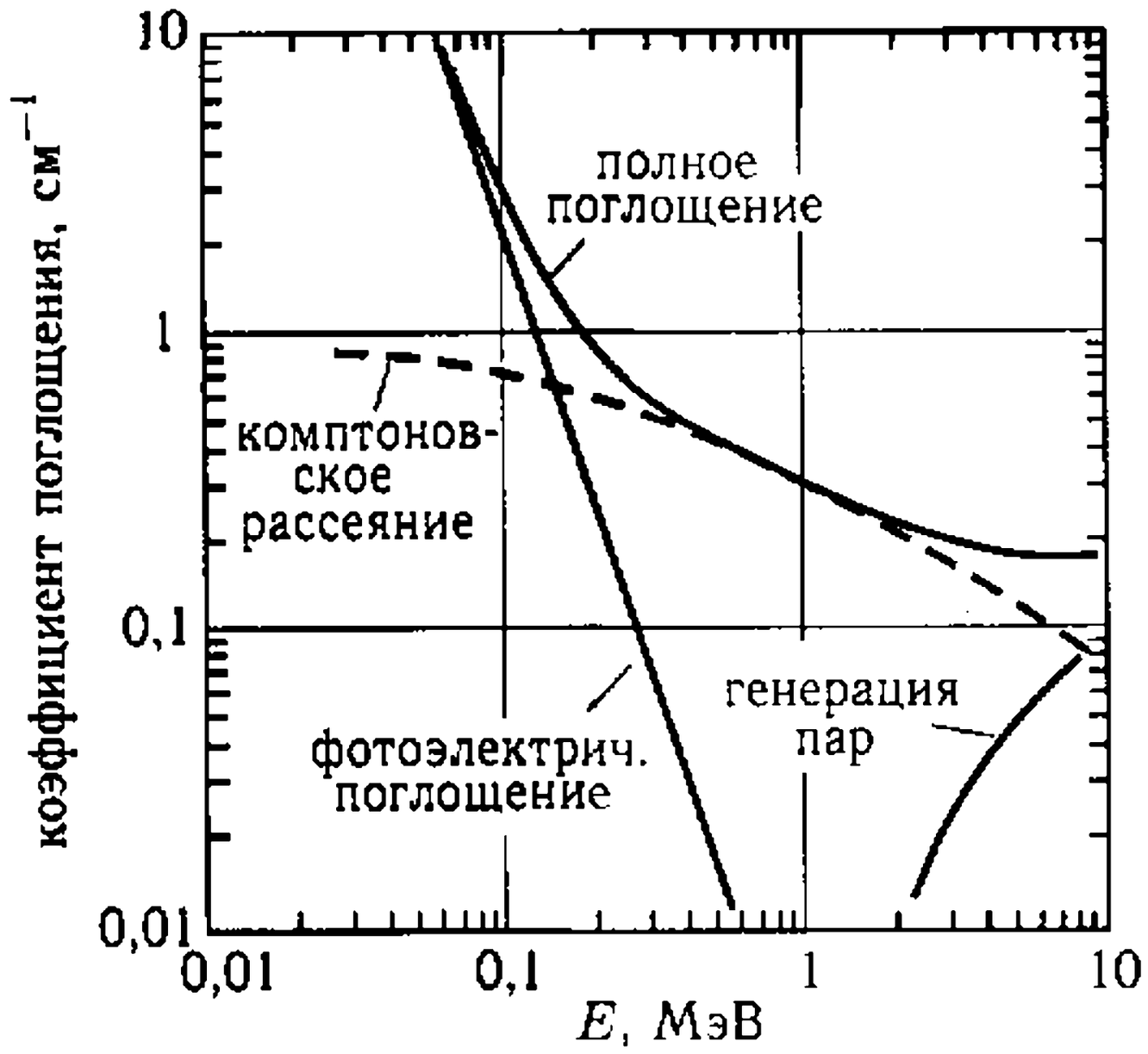


Рис. 7.19. Составляющие полного коэффициента поглощения высокоэнергичных фотонов в Ge [52]

Существуют три основных механизма взаимодействия высокоэнергичных фотонов с веществом (см. рис. 7.19):

1) фотоэлектрическое поглощение, при котором возбуждается фотоэлектрон, энергия которого равна разности энергии падающего кванта и энергии ионизации соответствующей (K, L, ...) оболочки поглощающего атома,

2) комптоновское рассеяние, при котором разность энергий падающего и рассеянного квантов «уносится» возбужденным электроном,

3) рождение электрон-позитронных пар. Заметим, что если вторичные частицы, рожденные при поглощении первичного кванта, не покидают детектор, то их энергия также используется для ионизации полупроводника. Например, при фотоэлектрическом поглощении возбужденный атом через $\sim 10^{-9}$ с испускает квант света или Оже-электрон, которые также рождают электронно-дырочные пары.

Поэтому число возбужденных в полупроводнике электронно-дырочных пар оказывается пропорциональным энергии регистрируемой частицы. Это дает основание использовать полупроводниковые детекторы не только для *регистрации* частиц, но и для *измерения их энергии*, то есть для решения задач ядерной спектрометрии. При создании детекторов рентгеновских и γ-квантов выгоднее использовать полупроводники, атомы которых имеют высокие атомные номера, поскольку коэффициент поглощения излучения быстро возрастает с ростом атомного номера.

Регистрация других видов нейтральных частиц (нейтронов и даже нейтрино) возможна только за счет вторичных частиц, возникающих в результате ядерных реакций первичных частиц с атомами полупроводника или атомами мишени, которая располагается в непосредственной близости от полупроводникового детектора.

Первыми детекторами ядерных излучений были *поверхностно-барьерные детекторы*, в которых возбужденные электронно-дырочные пары собирались электрическим полем барьера Шоттки (например, в барьере Au–n-Ge). Эти детекторы широко использовались для регистрации сильно ионизирующих полупроводник α-частиц, но из-за небольшой толщины обедненного слоя плохо подходили для регистрации слабо поглощаемых полупроводником излучений.

Для обеспечения эффективного собирания носителей с большого объема кристалла были предложены две конструкции детекторов ядерных излучений [52]. Первая из них представляет собой p–n-переход или p–i–n-диод с базой из высокоомного полупроводника, работающий при высоком обратном смещении. Основная проблема при создании этого типа детекторов состоит в нахождении способа получения очень низкой разностной концентрации доноров и акцепторов в базе диода, чтобы электрическое поле в структуре было достаточно однородным на расстояниях порядка нескольких сантиметров. Во второй

конструкции детекторов используется однородный высокоомный полупроводник, в котором с помощью внешнего источника напряжения создается сильное электрическое поле. Несмотря на простоту последней конструкции, она страдает определенными недостатками: в используемых в ней материалах (CdTe, HgI_2, полуизолирующий GaAs) обычно велика роль захвата носителей на ловушки, а контакты к полупроводнику недостаточно стабильны. Поэтому область применения этого типа детекторов ограничивается задачами, не требующими высокого энергетического разрешения.

Примером детектора первого типа является широко используемый дрейфовый германиевый детектор, создаваемый с помощью диффузии лития. Метод изготовления этого детектора был предложен Пеллом [300]. В пластинку германия p-типа проводимости (см. рис. 7.20) путем кратковременной диффузии лития (450 °C, 1,5 минуты) из поверхностного слоя «заго-

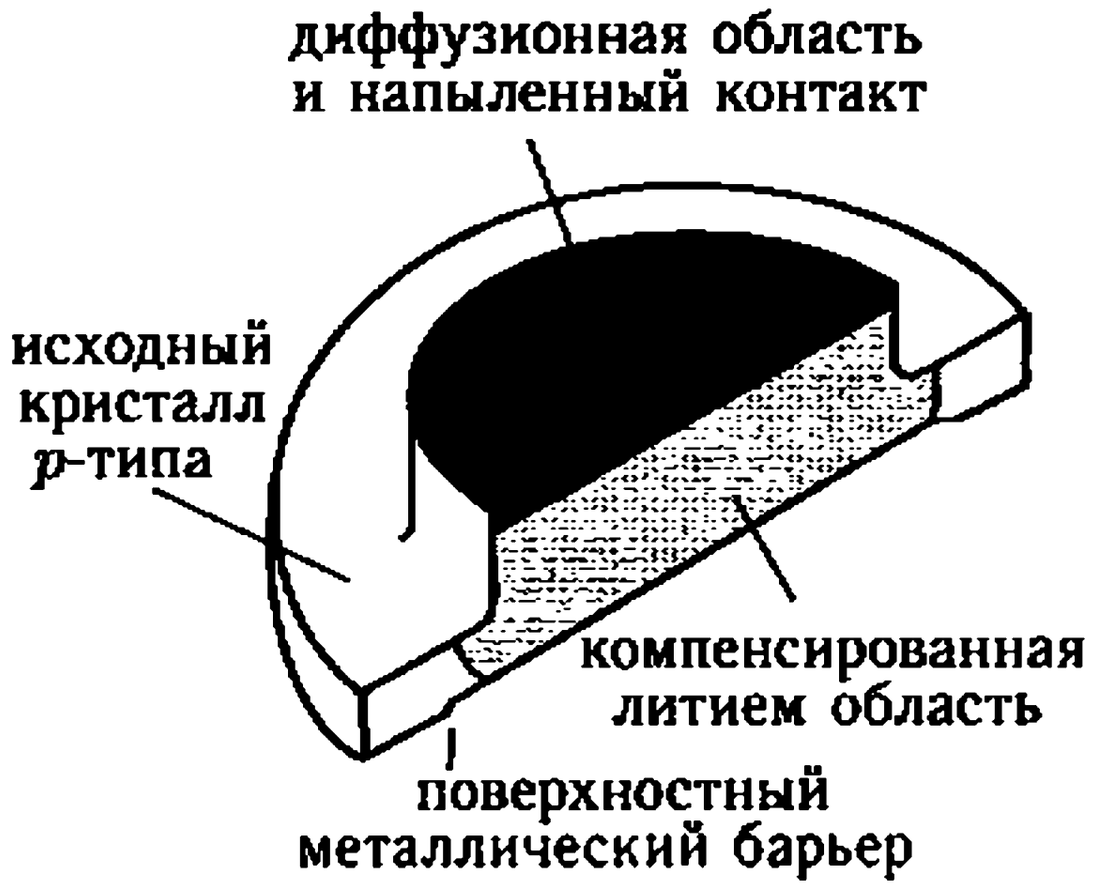


Рис. 7.20. Устройство дрейфового детектора Ge(Li) [52]

няется» примесь, которая создает область n-типа. Литий в Ge и Si является донорной примесью, которая располагается в междоузлиях и имеет высокий коэффициент диффузии. Поэтому при повышенной температуре (~ 125 °C), при приложении к p–n-переходу сильного обратного смещения, находящиеся в междоузлиях ионы Li^+ начинают дрейфовать в электрическом поле p–n-перехода в p-область структуры, компенсируя там акцепторную примесь. Подвижность ионов лития при 125 °C составляет $\sim 5 \cdot 10^{-11}$ см^2/В·с, что позволяет за несколько суток создать компенсированный слой толщиной в один сантиметр. Важно, что присутствие любой электрической неоднородности в кристалле вызывает такое перераспределение атомов Li, при котором эта неоднородность эффективно компенсируется. После того как литий продиффундирует почти на всю глубину пластины, обратное смещение снимают и детектор охлаждают до 77 К во избежание изменения установившегося в кристалле распределения атомов Li. Получаемая таким способом

разностная концентрация электрически активных примесей $(N_d - N_a)$ в компенсированной области оказывается порядка 10^8 см$^{-3}$, что достаточно для создания в структуре области сильного поля протяженностью в несколько сантиметров [1]. Аналогичным способом могут быть созданы и детекторы из Si(Li), однако из-за того, что коэффициент поглощения γ-квантов в кремнии в несколько раз меньше, чем в германии, область применения этих детекторов ограничивается работой в рентгеновской области спектра.

Выше мы уже отмечали, что полупроводниковые детекторы ядерных излучений позволяют по числу рожденных электронно-дырочных пар судить об энергии частицы [2]. Для этого необходимо измерить полный заряд носителей, рожденных частицей. Так как электронам и дыркам приходится дрейфовать через достаточно протяженную область сильного поля детектора, а эффективность собирания возбужденных электронно-дырочных пар должна достигать ~99,9%, становится понятным, почему на концентрацию уровней прилипания в полупроводнике и однородность кристалла накладываются жесткие ограничения. Далее, поскольку регистрируемые токи малы и их временна́я зависимость имеет сложную форму, для детектирования частиц приходится использовать специальные схемы зарядочувствительных усилителей, на выходе которых устанавливают формирователи импульсов стандартной формы [3]. Только в этом случае

---

[1]) В наиболее чистых кристаллах Ge и Si концентрация остаточных примесей составляет 10^{10}–10^{12} см$^{-3}$, что позволяет получить толщину области пространственного заряда при максимальном напряжении смещения, ограниченном пробоем, всего несколько мм.

[2]) При регистрации нестабильных элементарных частиц, таких как позитроны или заряженные π-мезоны, определение энергии первичной частицы требует учета того, что в результате аннигиляции позитрона или распада π-мезона (с образованием μ-мезона) в полупроводнике выделяется дополнительная энергия, связанная с ядерной реакцией [301].

Кроме того, при регистрации ионов *тяжелых* атомов проявляется так называемый амплитудный эффект, при котором амплитуда импульсов оказывается несколько ниже, чем при регистрации ионов легких атомов с той же энергией [301]. Это явление объясняется повышенным темпом рекомбинации неравновесных носителей в плотном треке, возникающем вдоль траектории частицы. Концентрация неравновесных носителей в области трека тяжелых ионов может достигать 10^{18} см$^{-3}$ и увеличение темпа рекомбинации в этой области может происходить за счет Оже-рекомбинации и возникновения радиационных дефектов.

[3]) Необходимость установки специального формирователя связана с тем, что из-за разного времени пролета электронов и дырок от места рождения до контактов токовый сигнал на выходе детектора имеет сложную форму.

амплитуда импульсов пропорциональна числу собранных электронно-дырочных пар (энергии частицы).

При использовании полупроводниковых детекторов ядерных излучений для измерения энергии частиц важной характеристикой детектора является его *энергетическое разрешение*, которое определяет минимальную разность энергий частиц, которую удается разрешить с помощью детектора. Поскольку при прохождении вторичных частиц через вещество они одновременно ионизуют вещество и рассеиваются на фононах, то вследствие статистического характера этих процессов при фиксированной энергии падающих частиц неизбежны флуктуации числа рождаемых ими электронно-дырочных пар. Среднее число рождаемых электронно-дырочных пар равно

$$N = E/E_0,$$

где E — энергия частицы, а E_0 — средняя энергия образования пары, которая приблизительно равна $3E_g$ (в случае Ge ее точное значение равно 2,96 эВ [52]). Флуктуация числа электронно-дырочных пар зависит от деталей развития ливня и в общем случае рассчитывается по формуле

$$\sigma = \sqrt{FN},$$

где F — так называемый *фактор Фано*. Экспериментально найденные значения этого фактора составляют $F = 0{,}06$–$0{,}12$ для Ge и 0,09–0,14 для Si. Это позволяет оценить энергетическое разрешение детектора

$$\Delta E/E = \sigma/N.$$

Например, при регистрации гамма-квантов с энергией 1 МэВ с помощью германиевого детектора ширина функции распределения импульсов будет составлять $2\sqrt{2\ln 2}\,\sigma \approx 1{,}15$ кэВ, то есть примерно 0,1 % от энергии частицы. Следует отметить, что поскольку средняя энергия образования пары E_0 в полупроводниках примерно на порядок меньше, чем в кристаллах сцинтилляторов и газах, энергетическое разрешение полупроводниковых детекторов ядерных излучений $\Delta E/E$ существенно выше, чем у сцинтилляционных детекторов и пропорциональных (газовых)

Из-за большого линейного размера детектора разброс времен пролета носителей может достигать 1 мкс. Формирователь интегрирует токовый сигнал, преобразуя его в импульс приблизительно гауссовой формы с фиксированной длительностью на половине высоты.

счетчиков, что и определяет широкое использование полупроводниковых детекторов.

Для экспериментов в области ядерной физики созданы специальные одномерные («полосковые») координатно-чувствительные детекторы на основе линеек кремниевых p–n-переходов, в которых достигнуто пространственное разрешение 4,5–7 мкм, а для регистрации двумерных «изображений» успешно используются охлаждаемые кремниевые ПЗС-структуры с пространственным разрешением 4–6 мкм [301].

7.2. Светодиоды и полупроводниковые лазеры

7.2.1. Механизмы излучательной рекомбинации. Излучение является процессом, обратным поглощению. При рекомбинации электрон, занимающий состояние с более высокой энергией, переходит в незаполненное состояние с более низкой энергией; при этом вся или бóльшая часть высвобождаемой энергии может выделиться в виде кванта электромагнитного излучения. Механизм рекомбинации электронов и дырок, в результате которого испускаются такие кванты, называют *излучательной рекомбинацией*. Само явление испускания излучения возбужденным каким-либо способом веществом называют *люминесценцией* [1]). Каналы рекомбинации, при которых кванты электромагнитного излучения не возбуждаются, называют *безызлучательными* каналами рекомбинации.

Оптические переходы в полупроводниках (как поглощение, так и люминесценция) существенно зависят от зонной структуры полупроводника и в первую очередь от того, является ли он прямозонным или непрямозонным [261, 302]. Причиной этого является то, что волновое число фотона $k = 2\pi/\lambda$ (здесь λ — длина волны излучения) на 3–4 порядка меньше волнового вектора, отвечающего краю зоны Бриллюэна. Поскольку при испускании и поглощении света должны выполняться законы сохранения энергии и импульса, то из этого следует, что оптические переходы без участия третьей частицы возможны только в прямозонных полупроводниках (например, в GaAs). В непрямозонных полупроводниках (Ge, Si, GaP) волновые векторы, отвечающие

---

[1]) Первые сведения о наблюдении люминесценции в полупроводниках относятся к началу XX века. Раунд (1907) и затем О. В. Лосев (1923) наблюдали свечение в области контакта к SiC при пропускании через образец электрического тока.

экстремумам зоны проводимости и валентной зоны, настолько различаются, что оптический переход без участия третьей частицы, которая «забирала» бы разность квазиимпульсов, просто невозможен. В качестве третьей частицы, как и в случае туннелирования в непрямозонном полупроводнике (см. п. 1.4), могут выступать фононы и атомы примесей.

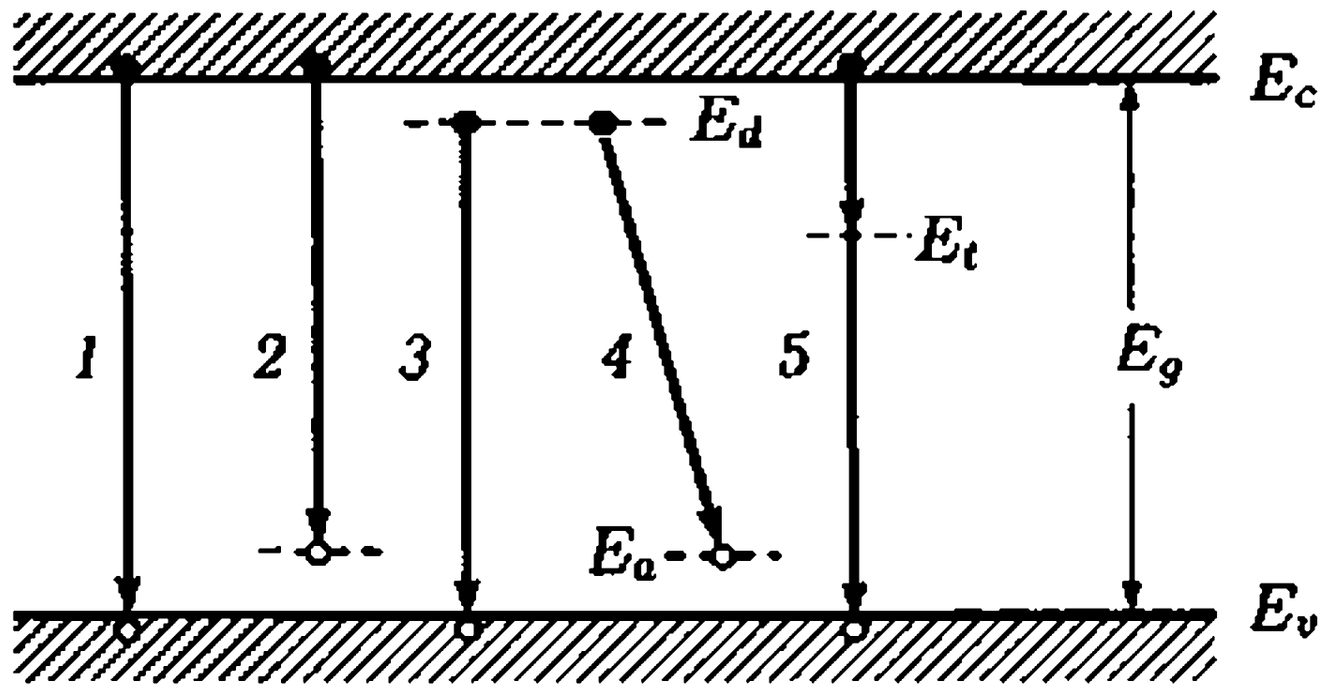


Рис. 7.21. Каналы излучательной (*1–4*) и безызлучательной (*5*) рекомбинации в полупроводнике: *1* — переходы зона–зона, *2* — переходы зона–акцептор, *3* — переходы донор–зона, *4* — переходы между донорами и акцепторами, *5* — рекомбинация через глубокий примесный уровень

Основные каналы излучательной рекомбинации в полупроводниках показаны на рис. 7.21. Люминесценция может возникать при рекомбинации свободных электрона и дырки (межзонные переходы, канал *1* на рис. 7.21), одного свободного и одного связанного на примеси носителя (переходы зона–примесь и примесь–зона, каналы 2,3), при переходах электронов между примесями (излучательная рекомбинация на донорно-акцепторных парах, канал *4*). Кроме того, возможна экситонная рекомбинация, при которой электрон и дырка, связанные кулоновским взаимодействием, могут двигаться в кристалле как единое целое (свободный экситон) или быть локализованными на каком-нибудь потенциале (связанный экситон). В реальных кристаллах с излучательной рекомбинацией всегда конкурирует безызлучательная рекомбинация (канал *5*), которая обычно идет через глубокие уровни в запрещенной зоне, хотя возможны и другие механизмы безызлучательной рекомбинации (например, Оже-рекомбинация).

Возможность расчета спектров излучательной рекомбинации на основании данных оптического поглощения была теоретически обоснована ван Русбреком и Шокли [303]. Они исходили из того, что в любом веществе в состоянии термодинамического равновесия темпы испускания и поглощения квантов света с заданной энергией в точности равны (*принцип детального*

равновесия). Основываясь на этом, ван Русбрек и Шокли показали, что темп излучательной рекомбинации в полупроводнике описывается уравнением

$$R_{\text{равн}}(\omega)d\omega = \frac{\omega^2 n_r(\omega)^2}{\pi^2 c^2[\exp(\hbar\omega/kT) - 1]}\,\alpha(\omega)d\omega, \qquad (7.35)$$

где $n_r(\omega)$ — показатель преломления света с частотой ω, c — скорость света в вакууме, а $\alpha(\omega)$ — коэффициент поглощения. Это уравнение позволяет из известного спектра поглощения рассчитать равновесный спектр излучения полупроводника.

Будем считать, что полупроводник, для которого определен спектр поглощения, невырожден. Тогда в неравновесных условиях, пока функции распределения электронов и дырок остаются больцмановскими, выражение для темпа излучательной рекомбинации можно записать следующим образом:

$$R = \frac{np}{n_i^2} R_{\text{равн}} = Bnp. \qquad (7.36)$$

Здесь под n и p мы будем подразумевать концентрацию не только свободных носителей, но и концентрацию электронов и дырок на всех разрешенных энергетических состояниях. Так, кроме концентрации свободных дырок, p может обозначать их концентрацию на любых состояниях ниже квазиуровня Ферми (например, дырок, захваченных на акцепторы) или дырок, связанных в экситоны. То же самое относится и к электронам, находящимся на состояниях выше квазиуровня Ферми.

Величину $B = R_{\text{равн}}/n_i^2$ в формуле (7.36) часто называют *вероятностью излучательной рекомбинации*. В полупроводниках с прямыми разрешенными оптическими переходами значения B лежат в пределах 10^{-11}–10^{-10} см^3/с [302]; в непрямозонных полупроводниках величина B на 3–4 порядка меньше. Выражение (7.36) позволяет ввести так называемое *излучательное время жизни*, которое описывает вклад излучательных переходов в общее время жизни. Например, в образце p-типа излучательное время жизни равно

$$\tau_{\text{изл}} = \frac{\delta n}{\delta R} = \frac{1}{Bp}. \qquad (7.37)$$

Для оценки, в p-GaAs с концентрацией дырок 10^{18} см$^{-3}$ излучательное время жизни составляет 3,5 нс [261]. Излучательное время жизни может служить оценкой быстродействия высокоэффективных светодиодов и лазеров, работающих на переходах зона–зона или зона–мелкая примесь.

Согласно уравнению (7.35), зная коэффициент поглощения, можно рассчитать равновесный спектр излучения полупроводника. Если отклонение от равновесия невелико, то этой же формулой можно воспользоваться и для расчета спектра люминесценции образца. Найдем в качестве примера форму спектра люминесценции, отвечающую межзонным оптическим переходам.

Для коэффициента поглощения для прямых разрешенных межзонных переходов в курсе физики полупроводников выводится следующее соотношение (см., например, [1]):

$$\alpha(\omega) = \frac{2(2m_r)^{3/2}q^2|\mathbf{p}_{cv}|^2}{3cm_0^2 n_r(\omega)\hbar^3\omega}(\hbar\omega - E_g)^{1/2}, \qquad (7.38)$$

где m_r — приведенная масса электрона и дырки ($m_r^{-1} = m_c^{-1} + m_v^{-1}$), $|\mathbf{p}_{cv}|^2$ — квадрат матричного элемента оператора импульса, $n_r(\omega)$ — показатель преломления, c — скорость света в вакууме. Подставляя выражение (7.38) в формулу (7.35) и учитывая, что $\hbar\omega \gg kT$, получаем следующее асимптотическое выражение для спектра межзонной (краевой) люминесценции:

$$R(\omega) \sim (\hbar\omega - E_g)^{1/2}\exp(-\hbar\omega/kT). \qquad (7.39)$$

На рис. 7.22 проведено сравнение спектра краевой люминесценции антимонида индия со спектром, рассчитанным по формуле (7.39). Как следует из рисунка, эта формула неплохо описывает форму спектра краевой люминесценции. На спектры краевой люминесценции сильное влияние могут оказывать *хвосты плотности состояний*, образующиеся вблизи краев разрешенных зон в легированных полупроводниках. Присутствие небольшого хвоста можно наблюдать на рис. 7.22. В сильно легированных полупроводниках эти хвосты сдвигают спектр краевой люминесценции в область более длинных волн. Излучательные переходы между состояниями в хвостах плотности состояний могут определять очень эффективную люминесценцию, как, например, в случае сильно компенсированного GaAs(Si). В полупроводниках с большой энергией связи экситона спектры краевой

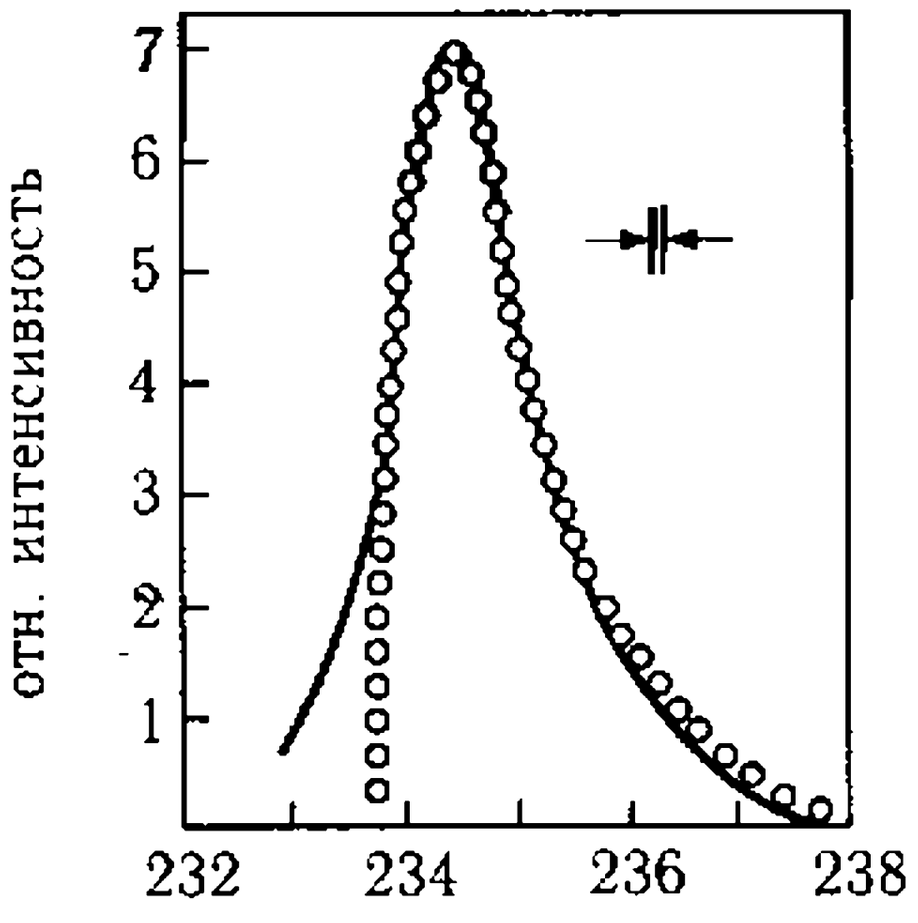


Рис. 7.22. Спектр краевой люминесценции нелегированного InSb при $T = 4,2$ К [304]

люминесценции могут усложняться из-за эффектов кулоновского взаимодействия электронов и дырок (экситонных эффектов).

Расчеты спектров люминесценции могут быть проведены и для излучательных переходов по каналу зона–акцептор (донор–зона). Соответствующие формулы можно найти в [261]. Однако говоря о люминесценции с участием примесей, следует отметить, что из-за захвата неравновесных носителей на примесные уровни концентрация этих носителей на уровнях всегда больше, чем в разрешенных зонах. Поэтому излучательная рекомбинация с участием примесных уровней часто оказывается более интенсивной, чем краевая люминесценция.

Обычно излучательная рекомбинация является не единственным механизмом рекомбинации в полупроводнике и часть электронно-дырочных пар рекомбинирует безызлучательно. Если в полном темпе рекомбинации выделить темп излучательной и безызлучательной рекомбинации и с помощью уравнений $\tau_{\text{изл}} = \Delta n / R_{\text{изл}}$ и $\tau_{\text{безызл}} = \Delta n / R_{\text{безызл}}$ определить излучательное и безызлучательное время жизни, то можно ввести понятие *квантового выхода* люминесценции, который равен доли излучательных переходов в полном темпе рекомбинации:

$$\eta_{\text{внутр}} = \frac{R_{\text{изл}}}{R_{\text{изл}} + R_{\text{безызл}}} = \frac{\tau_{\text{безызл}}}{\tau_{\text{безызл}} + \tau_{\text{изл}}}. \qquad (7.40)$$

7.2.2. Светодиоды. Светодиодами называют p–n-переходы, которые при пропускании через них электрического тока испускают электромагнитное излучение в видимой, инфракрасной или ультрафиолетовой областях спектра [304]. Это излучение возникает в результате излучательной рекомбинации инжектированных носителей; механизмы такой рекомбинации были рассмотрены нами в п. 7.2.1.

Среди светодиодов особое значение имеют светодиоды, которые излучают в видимой области спектра ($\lambda = 0,39$–$0,77$ мкм) — области чувствительности человеческого глаза. Эти светодиоды широко применяются для создания небольших информационных табло, в качестве сигнальных огней в автомобилях и светофорах, индикаторов в электронной аппаратуре. Их можно использовать в качестве источников света в оптронах (см. подстрочное замечание на с. 410) и оптотиристорах (см. с. 235). Они широко применяются для создания локальной подсветки (мобильные телефоны, велосипеды), а в последние годы начаты работы по использованию их для замены ламп накаливания в более широкомасштабных задачах освещения (см. с. 441).

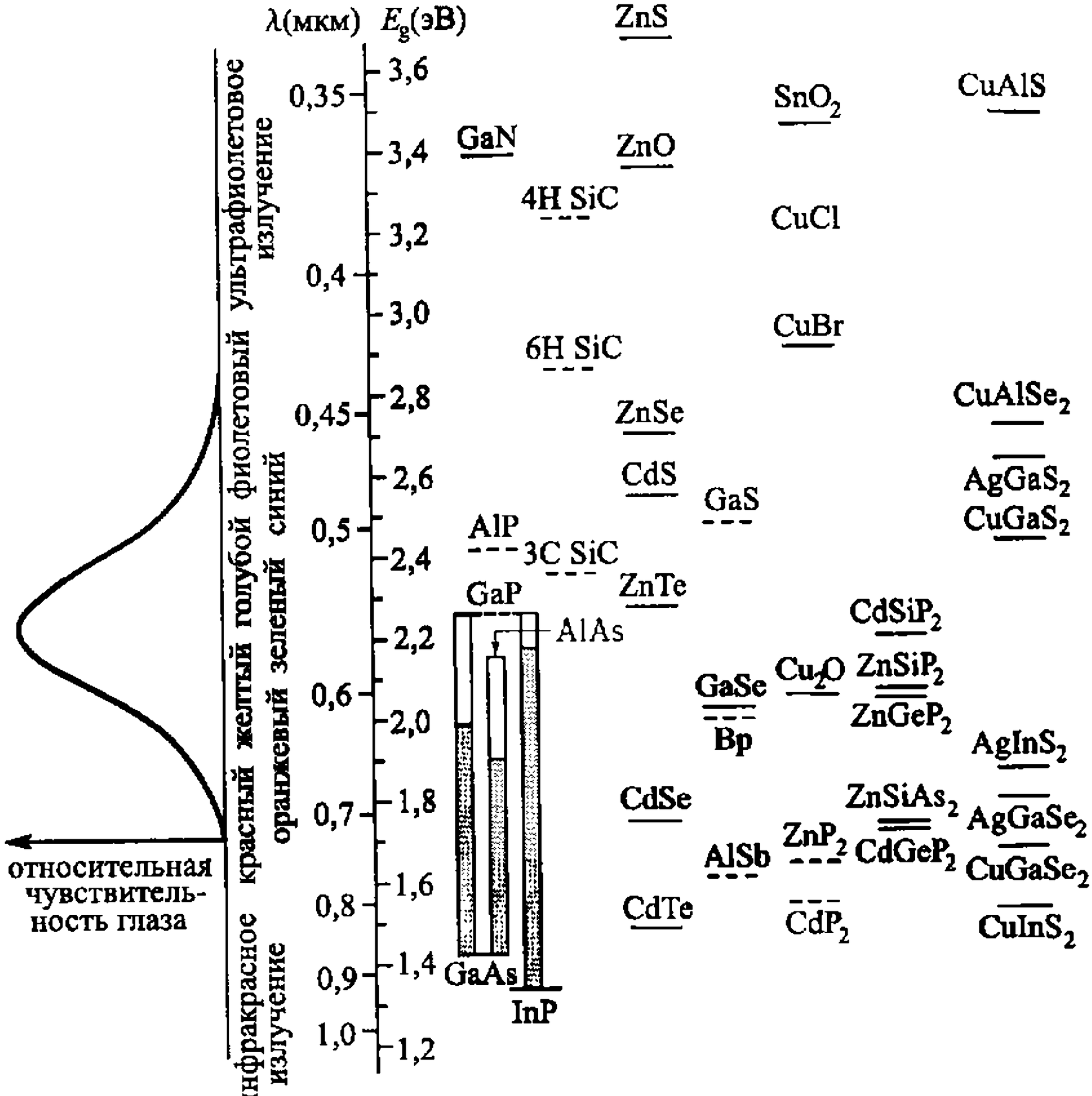


Рис. 7.23. Полупроводники, представляющие интерес для создания светодиодов, работающих в видимой области спектра [304]. Приводимые значения E_g уточнены по [305, 306]. Для тройных соединений $A^{II}B^{IV}C_2^V$, кристаллизующихся в структуре халькопирита, указаны значения энергии псевдопрямых переходов. Сплошные и пунктирные линии отвечают, соответственно, прямозонным и непрямозонным полупроводникам

Полупроводники, из которых могут быть созданы такие светодиоды, должны иметь ширину запрещенной зоны $E_g > 1,8$ эВ (см. рис. 7.23) [1]).

[1]) Речь идет о светодиодах, в которых фотон рождается в результате рекомбинации одной электронно-дырочной пары. В конце 60-х годов были разработаны светодиоды из GaAs с люминофорным покрытием, в которых видимое (голубое, зеленое или красное) излучение возбуждалось в результате поглощения люминофором двух или трех квантов инфракрасного излучения [304].

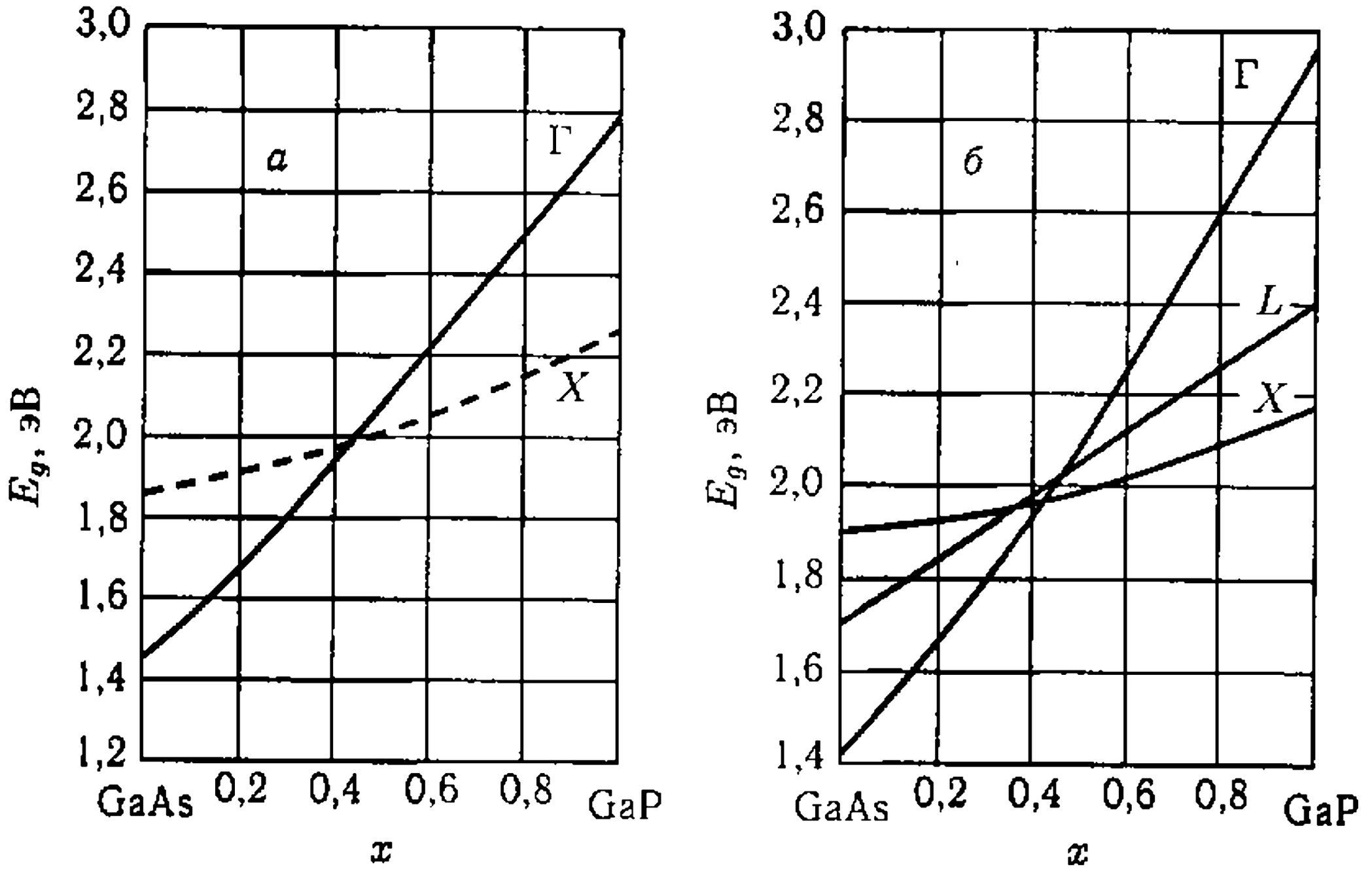


Рис. 7.24. Зависимость ширины запрещенной зоны для прямых (Г–Г) и непрямых (Г–X, Г–L) переходов в $GaAs_{1-x}P_x$ ($а$) и $Al_xGa_{1-x}As$ ($б$) от параметра состава x [304, 306]

Одним из первых практически важных материалов для светодиодов видимого диапазона стал твердый раствор $GaAs_{1-x}P_x$, относящийся к соединениям группы $A^{III}B^V$. На рис. 7.24 $а$ показана зависимость ширины запрещенной зоны этого твердого раствора от параметра состава x. В области $0 < x < 0,45$ наинизший по энергии минимум зоны проводимости лежит в Г-точке зоны Бриллюэна и оптические переходы — прямые. При $x > 0,45$ полупроводник становится непрямозонным. Поскольку, как мы уже говорили выше, в непрямозонном полупроводнике в излучательной рекомбинации должна принимать участие третья частица, то вероятность излучательной рекомбинации в таком полупроводнике обычно на несколько порядков меньше, чем в прямозонном. Поэтому эффективную краевую люминесценцию в $GaAs_{1-x}P_x$ удается получить только при $x < 0,45$ ($E_g < 1,99$ эВ), то есть таким способом можно создать светодиоды только с красным и оранжево-красным свечением.

Однако если в непрямозонном полупроводнике сделать эффективной передачу импульса, то и в таких материалах можно получить достаточно высокий внутренний квантовый выход люминесценции. Это достигается путем введения в кристаллы *центров эффективной излучательной рекомбинации.* В фосфиде галлия (GaP), являющемся типичным непрямозонным полупроводником, таким центром может служить изоэлектронная

ловушка N. Азот в GaP замещает атомы фосфора в узлах решетки и является изоэлектронной примесью. Азот и фосфор имеют одинаковую внешнюю электронную структуру (оба относятся к V группе элементов периодической системы), однако их атомные радиусы и электроотрицательности сильно различаются, и поэтому на месте примеси в кристалле формируется короткодействующий потенциал, способный связывать электроны. В исходном состоянии изоэлектронная ловушка нейтральна. В p-области диода инжектированный электрон сначала захватывается на центр. Отрицательно заряженный центр затем захватывает дырку из валентной зоны, формируя связанный экситон. Последующая аннигиляция этой связанной электронно-дырочной пары приводит к рождению фотона с энергией ~2,24 эВ, примерно равной разности между шириной запрещенной зоны и энергией связи экситона. Так как захваченный электрон сильно локализован на центре, то фурье-составляющая его волновой функции с волновым вектором, отвечающим волновому вектору дырки, достаточно велика и вероятность излучательной рекомбинации без участия третьей частицы оказывается достаточно большой. Внутренний квантовый выход люминесценции в случае GaP(N) при этом может достигать нескольких процентов. Те же явления наблюдаются и в твердом растворе $GaAs_{1-x}P_x$(N). Из GaP(N) и $GaAs_{1-x}P_x$(N) были изготовлены первые светодиоды с зеленым и желто-зеленым свечением ($\lambda_{max} \approx 565$ и 589 нм), внешний квантовый выход которых достигал ~0,7 % [304].

Другим примером эффективных центров излучательной рекомбинации в непрямозонных полупроводниках может служить примесный комплекс Zn–O в GaP. Первоначально считалось, что эффективная рекомбинация в p-области светодиодов из GaP(Zn,O) связана с рекомбинацией электронов, захваченных на глубокие донорные уровни кислорода, с дырками на акцепторных уровнях цинка (рекомбинация на донорно-акцепторных парах). Последующие исследования однако показали, что на самом деле основными каналами рекомбинации при 300 К являются рекомбинация связанных экситонов, локализованных на ближайших парах Zn–O, и рекомбинация связанных на комплексе Zn–O электронов со свободными дырками. Измерения сечения захвата электрона на комплекс Zn–O позволили оценить максимальный внутренний квантовый выход люминесценции в GaP(Zn,O), который составил ~35 %. Светодиоды из GaP(Zn,O) были первыми выпущенными промышленностью светодиодами с красным цветом свечения ($\lambda_{max} \approx 698$ нм). Рекордное значение внешнего

квантового выхода, полученное на этих светодиодах, составляет 12% [304].

В синих светодиодах из карбида кремния SiC ($\lambda_{max} \approx 470$ нм) излучательная рекомбинация идет через глубокий акцепторный уровень Al. Внешний квантовый выход излучения в светодиодах из этого непрямозонного полупроводника невысок: в начале 90-х годов его значение составляло 0,03%.

Хотя введение центров эффективной излучательной рекомбинации и позволяет создать на основе непрямозонных полупроводников светодиоды с достаточно высоким квантовым выходом, наибольший интерес все же представляют прямозонные полупроводники, на которых можно достигнуть более значимых результатов. Кроме упоминавшегося выше твердого раствора $GaAs_{1-x}P_x$ с $x < 0,45$, прямозонными полупроводниками являются твердый раствор $Al_xGa_{1-x}As$ при $x < 0,31$ ($E_g < 1,90$ эВ), твердый раствор $In_{1-x}Ga_xP$ при $x < 0,62$ ($E_g < 2,18$ эВ), твердый раствор $Al_xIn_{1-x}P$ при $x < 0,39$ ($E_g < 2,23$ эВ). Хотя ширина запрещенной зоны, отвечающая переходу к непрямозонному полупроводнику, в $Al_xGa_{1-x}As$ меньше, чем в $GaAs_{1-x}P_x$, из этого материала до сих пор изготавливают яркие красные светодиоды ($\lambda_{max} \approx 660$ нм). Среди других полупроводников с прямой структурой зон, имеющих в настоящее время наибольшее значение, следует отметить нитрид галлия GaN, твердые растворы на его основе $In_xGa_{1-x}N$ и $Al_xGa_{1-x}N$, а также четверной твердый раствор $Al_{1-x-y}In_yGa_xP$. Из последнего, в частности, изготавливают сверхъяркие светодиоды с желтым, оранжевым и красным свечением ($\lambda_{max} = 568$–635 нм). После тщательной проработки проблемы вывода излучения из образца [307] внешний квантовый выход в светодиодах из $Al_{1-x-y}In_yGa_xP$ был доведен до 55%, а светоотдача — до 100 лм/Вт [308]. Из твердых растворов $In_xGa_{1-x}N$ изготавливают яркие светодиоды с зеленым, сине-зеленым, синим и фиолетовым свечением с внешним квантовым выходом до 63% [?]. [1] Кроме того, разработаны зеленые и синие светодиоды из твердых растворов полупроводников группы $A^{II}B^{VI}$ — $ZnTe_{0,1}Se_{0,9}$ и $Zn_{0,9}Cd_{0,1}Se$ с внешним квантовым выходом до 5,3% [310] [2]. Наиболее коротковолновое излучение позволяют получить светодиоды из $Al_xGa_{1-x}N$,

---

[1] При этом внутренний квантовый выход излучения достигает 82% [309].

[2] В настоящее время основным недостатком светодиодов и инжекционных лазеров на основе гетероструктур в полупроводниках группы $A^{II}B^{VI}$ является очень быстрая деградация интенсивности люминесценции [310].

которые излучают в ультрафиолетовой области спектра (λ_{max} = = 250–370 нм) и имеют выходную мощность порядка 1 мВт.

Существенный прогресс, достигнутый за последние годы в создании высокоэффективных (сверхъярких) светодиодов, связан с использованием гетеропереходов и структур с квантовыми ямами. Использование гетеропереходов позволяет добиться направленной инжекции носителей в область диода, в которой происходит наиболее эффективная излучательная рекомбинация, а использование квантовых ям — свести к минимуму поглощение излучения в полупроводниковой структуре и тем самым обеспечить высокий внешний квантовый выход светодиода.

Инфракрасные светодиоды. Для изготовления светодиодов, работающих в инфракрасной (ИК) области спектра, используются полупроводники, ширина запрещенной зоны которых не превышает 1,5 эВ. Наиболее важное значение среди них имеют GaAs, GaSb, InAs, твердые растворы $In_xGa_{1-x}As_yP_{1-y}$, $In_xGa_{1-x}As$, $Ga_{1-x}Al_xAs_ySb_{1-y}$. Светодиоды из арсенида галлия характеризуются самым высоким внутренним (99,7%) и довольно высоким внешним (более 40% [311]) квантовым выходом люминесценции, что связано с тем, что среди всех прямозонных полупроводников он является наиболее технологически освоенным.

Одним из наиболее важных применений ИК-светодиодов является их использование в качестве источников излучения в оптронах и системах волоконно-оптической связи. В последнем случае, поскольку длина волны излучения светодиодов должна соответствовать одному из минимумов оптических потерь в кварцевом стекловолокне (1,3 или 1,55 мкм), их изготавливают из твердого раствора $In_xGa_{1-x}As_yP_{1-y}$.

Другим применением инфракрасных светодиодов является мониторинг окружающей среды (контроль за загрязнением атмосферы выхлопными газами). Максимумы в спектрах излучения таких светодиодов должны совпадать с полосами поглощения NO и CO_2 ($\lambda = 4$–5 мкм). Такие светодиоды могут быть созданы из твердых растворов полупроводников $A^{IV}B^{VI}$ ($PbS_{1-x}Se_x$).

Конструкции светодиодов. При разработке светодиодов важно создать такую конструкцию прибора, при которой возникающее в p–n-переходе излучение эффективно выводилось бы из образца. Одним из основных параметров светодиода является его *внешний квантовый выход*, который связан с внутренним квантовым выходом люминесценции (7.40)

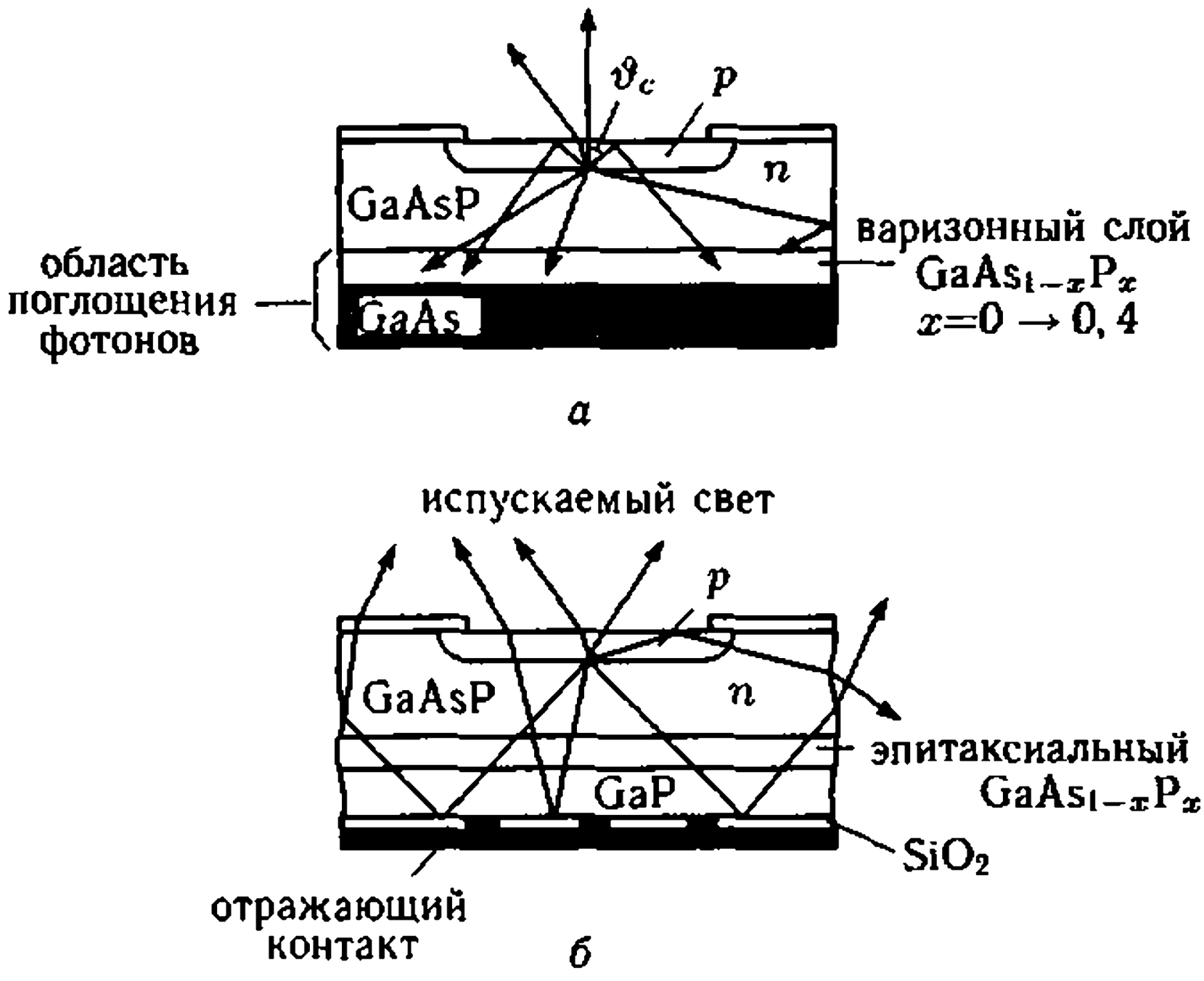


Рис. 7.25. Прохождение света, испускаемого p–n-переходом, в структурах с сильно и слабо поглощающей подложкой [14]

соотношением $\eta_{\text{внеш}} = \eta_o \eta_{\text{внутр}}$. Входящая в это соотношение величина η_o, называемая *коэффициентом вывода света*, как раз и характеризует эффективность вывода излучения из образца.

Среди светодиодных структур основной является структура с плоской геометрией (см. рис. 7.25). Обычно p–n-переход в светодиоде создают в эпитаксиальном слое, выращенном на подложке. Свет, генерируемый в области p–n-перехода, испускается во всех направлениях, однако из образца может выйти лишь небольшая его доля. Дело в том, что при переходе света из среды с высоким показателем преломления n_2 (в полупроводниках, используемых для изготовления светодиодов, n_2 находится в пределах 3,3–3,8) в среду с низким показателем преломления n_1 (например, в воздух) бóльшая его часть испытывает полное внутреннее отражение от границы раздела и возвращается обратно в образец. Это происходит для лучей, угол падения которых на границу раздела превышает критический угол,

$$\vartheta_c = \arcsin(n_1/n_2).$$

Кроме того, даже при $\vartheta < \vartheta_c$ часть фотонов отражается от границы раздела полупроводник–воздух и средний коэффициент пропускания составляет

$$\overline{T} \approx T_{\text{норм}} = 4n_1 n_2/(n_1 + n_2)^2 \approx 0,7.$$

Можно показать, что когда граница раздела плоская, излучение выводится только через верхнюю грань диода и многократным отражением света внутри образца можно пренебречь (то есть отраженный от границы раздела свет полностью поглощается в образце), коэффициент вывода света составляет всего

$$\eta_o = (1 - \cos\vartheta_c) \cdot \overline{T} \approx 1{,}3\text{–}1{,}6\,\%.$$

Коэффициент вывода света может дополнительно уменьшаться в результате поглощения света в объеме полупроводника даже на первом проходе, однако если придать поверхности образца специальную форму, использовать просветляющее покрытие и отражающие контакты, то можно заметно его повысить. Большую роль для получения высокого внешнего квантового выхода светодиода играет и выбор материала подложки: чем меньше поглощение в подложке, тем выше вероятность того, что после многократных отражений внутри структуры свет все-таки выйдет за пределы образца (ср. рис. 7.25*а* и рис. 7.25*б*).

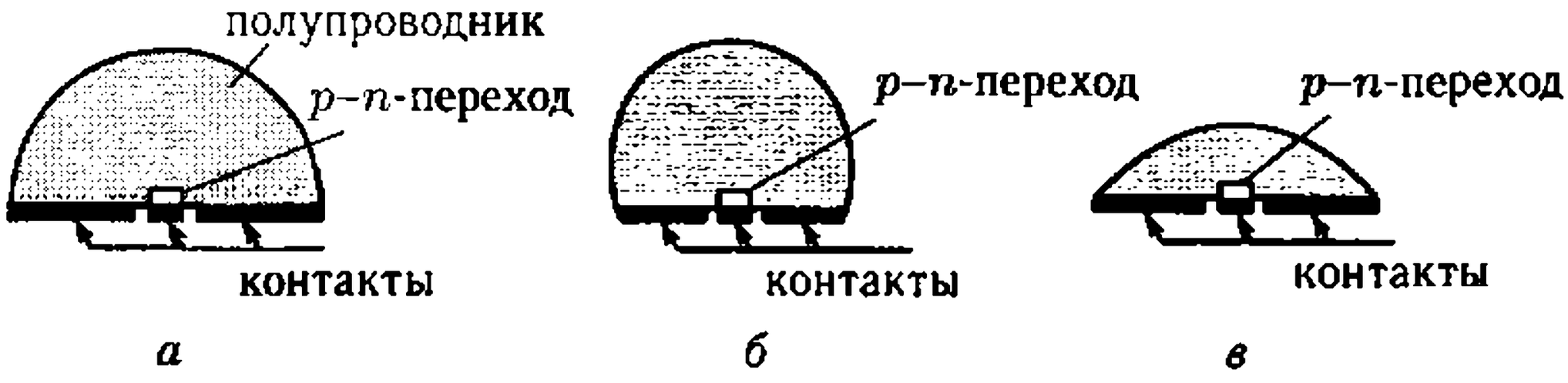


Рис. 7.26. Поперечное сечение трех конструкций светодиодов с улучшенной геометрией вывода излучения

На рис. 7.26 показаны поперечные сечения трех конструкций светодиодов с полусферической, усеченной сферической (сфера Вейерштрасса) и параболической геометрией. Основной особенностью этих конструкций является то, что все испускаемые p–n-переходом фотоны падают на границу раздела под углом меньше критического и, следовательно, не испытывают полного внутреннего отражения. При этом коэффициент вывода света может (теоретически) увеличиться в $\approx 2n^2$ раз. Различие рассматриваемых геометрий светодиодов будет проявляться только в угловом распределении интенсивности света (*диаграммах направленности*).

Использование просветляющего покрытия (например, помещение образца полупроводника в среду из пластмассы) увеличивает внешний квантовый выход излучения за счет увеличения критического угла ϑ_c. На практике показатель преломления используемых просветляющих покрытий лежит в пределах

1,4–1,8, что позволяет увеличить квантовый выход в 2–3 раза. Для инфракрасных светодиодов в качестве просветляющего покрытия можно также использовать легкоплавкие халькогенидные стекла, имеющие еще более высокий показатель преломления (2,4–2,9). Важно, что геометрическую форму, необходимую для формирования заданной диаграммы направленности, можно придавать просветляющему покрытию, а не полупроводниковому кристаллу, что существенно удешевляет стоимость производства светодиодов.

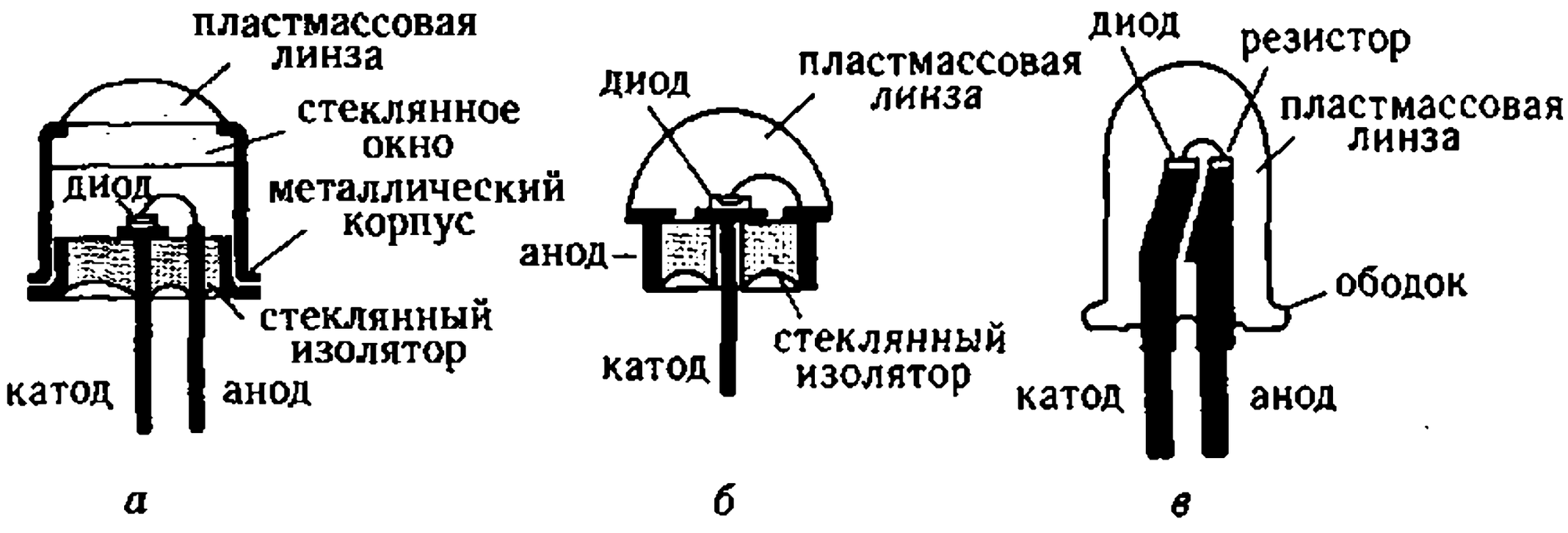


Рис. 7.27. Конструкции типичных светодиодов

Несколько типичных конструкций светодиодов показано на рис. 7.27. Их общими элементами являются полупроводниковый кристалл и пластмассовая линза, которая часто окрашивается с целью повышения контраста и коррекции спектра излучения. Конструкции, показанные на рис. 7.27 *а,б*, выполнены на базе обычных транзисторных и диодных корпусов, которые обеспечивают отвод выделяющегося при работе тепла. В конструкции, изображенной на рис. 7.27 *в*, полупроводниковый кристалл и токоограничительный резистор располагаются на толстых металлических выводах диода, по которым и отводится выделяющееся тепло.

При использовании инфракрасных светодиодов в качестве источников излучения для волоконно-оптических линий связи важной задачей является эффективный ввод излучения в световод, диаметр которого измеряется десятками мкм. Для этого разработаны специальные конструкции светодиодов [14], одна из которых представлена на рис. 7.28. В этой конструкции используется светодиодная структура с двумя гетеропереходами AlGaAs/GaAs, в которой тонкий активный слой из p-GaAs окружен слоями более широкозонного $Al_xGa_{1-x}As$. Это увеличивает эффективность излучательной рекомбинации за счет локализации инжектируемых электронов и дырок в узкозонной части

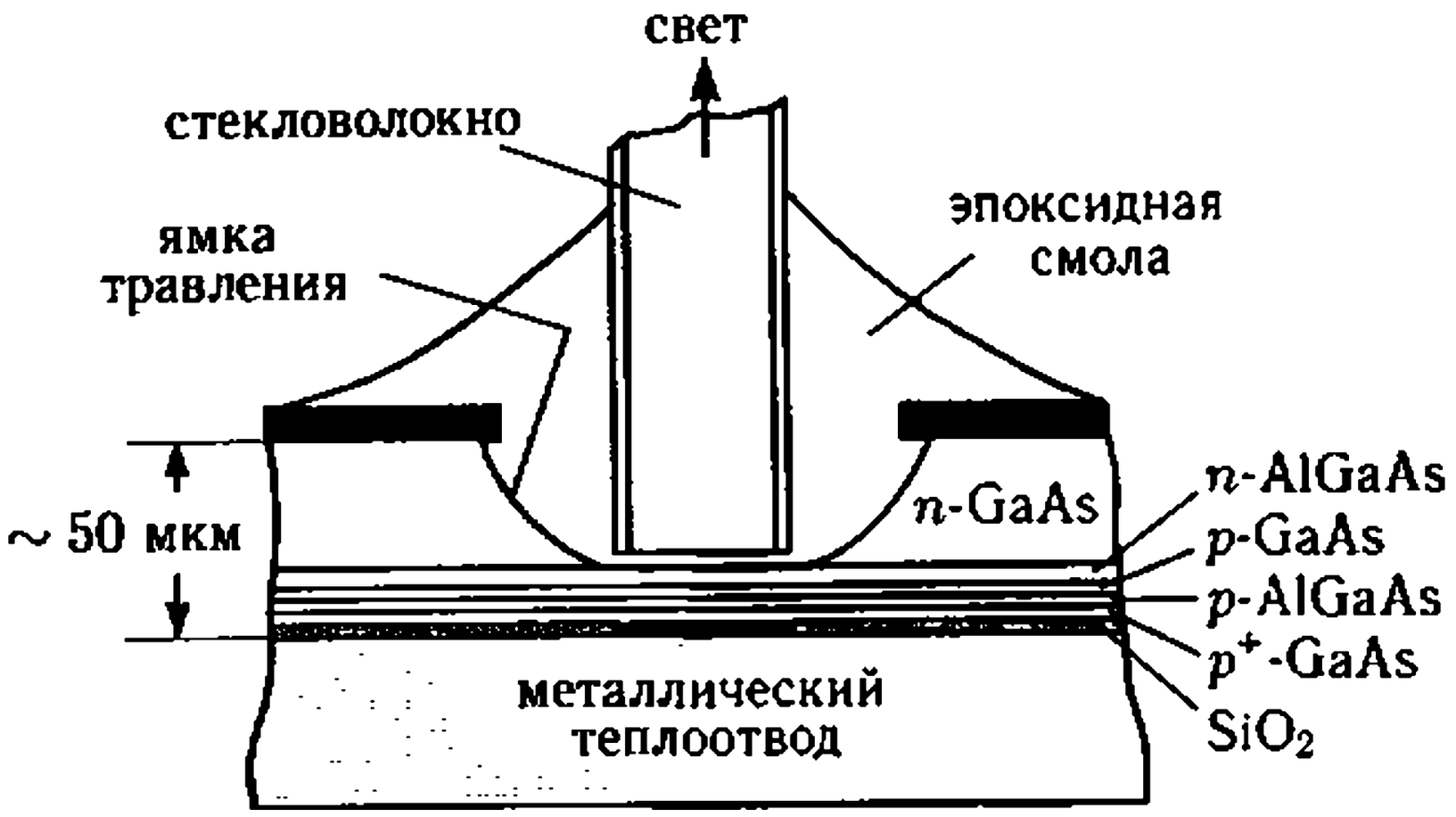


Рис. 7.28. Конструкция поверхностно-излучающего светодиода на основе двойной гетероструктуры AlGaAs/GaAs

структуры. Возникающее при этом излучение выводится через широкозонный слой практически не поглощаясь.

Локализация излучающей области в горизонтальной плоскости осуществляется путем ограничения размера окна в слое SiO_2, через которое создается контакт к нижнему p^+-слою структуры. Для уменьшения потерь на поглощение и обеспечения точного согласования положения торца световода и излучающей области в структуре с помощью фотолитографии вытравливается ямка, глубина которой делается такой, чтобы был полностью удален верхний контактный слой n-GaAs.

Белые светодиоды. Создание *«белых» светодиодов* является новым направлением в разработке полупроводниковых источников света. Благодаря успехам технологии и использованию новых полупроводниковых материалов (GaN и твердых растворов на его основе) стало возможным создавать синие, фиолетовые и ультрафиолетовые светодиоды, квантовый выход которых возрос настолько, что стало реально возможным их использование для решения задачи «твердотельного освещения». Важнейшим преимуществом светодиодов по отношению к лампам накаливания является высокая долговечность светодиодов, измеряемая десятками тысяч часов.

Полупроводниковые источники белого света можно создать несколькими способами. Прежде всего, можно смешать свет, испускаемый красным, зеленым и синим светодиодами. Хотя такие источники достаточно дороги, их преимуществом является возможность очень гибкой настройки оттенков белого света. Более дешевой является конструкция, в которой один светодиод

возбуждает люминофор (фосфо́р) или их смесь, дающую (иногда в комбинации с возбуждающим излучением) белое свечение [312].

Самой простой из таких конструкций является конструкция, в которой в качестве источника используется синий светодиод из $In_xGa_{1-x}N$ ($\lambda = 470$ нм), возбуждающий люминофор с желтым свечением ($Y_3Al_5O_{12}(Ce^{3+})$). Комбинация желтого и синего дает белый свет. Белые светодиоды с такой конструкцией были впервые созданы в начале 90-х годов сотрудниками фирмы Nichia. В настоящее время светоотдача таких белых светодиодов достигла 138 лм/Вт [309] [1]), то есть уже превысила светоотдачу ламп накаливания (15–20 лм/Вт) и люминесцентных ламп (70–100 лм/Вт) [2]).

Во второй конструкции синий светодиод возбуждает смесь люминофоров с красным ($SrS_2(Eu^{2+})$, $Y_2O_2S(Eu^{3+})$) и зеленым ($SrGa_2S_4(Eu^{2+})$, $ZnS(Cu,Al)$) свечением. Наконец, существует третья конструкция, в которой смесь синего, зеленого и красного люминофоров возбуждается ультрафиолетовым светодиодом. Светодиоды последних двух конструкций характеризуются белым свечением, более близким к естественному освещению, но требуют тщательного выбора люминофоров, спектры возбуждения люминесценции которых были бы согласованы со спектром возбуждающего излучения. Сотрудники фирмы Sandia в качестве люминофоров для белых светодиодов успешно используют квантовые точки на основе CdS, спектрами возбуждения и люминесценции которых можно управлять, изменяя размер точек.

К сожалению, по состоянию на 2006 г. стоимость получения единицы светового потока с помощью белых светодиодов составляла около 0,05 долл./лм, что примерно на два порядка выше, чем для ламп накаливания и люминесцентных ламп ($5 \cdot 10^{-4}$ долл/лм) [3]). Высокая стоимость полупроводниковых

---

[1]) В конце 2006 г. сотрудники фирмы Nichia сообщили о преодолении барьера в 150 лм/Вт в светоотдаче белых светодиодов.

[2]) При описании характеристик светодиодов видимого диапазона обычно пользуются не энергетическими, а *фотометрическими* единицами, которые учитывают спектральную чувствительность человеческого глаза. Напомним, что за единицу силы света в этой системе принимается *кандела*, которая определена как сила света, испускаемая с площади 1/60 $см^2$ абсолютно черным телом, температура которого равна температуре затвердевания платины (2042 К). Световой поток, излучаемый таким источником в телесный угол 1 стерадиан, равен 1 люмену (лм).

[3]) Для сравнения укажем, что стандартная 60-ваттная лампа накаливания производит световой поток около 850 лм.

источников света частично компенсируется их более высокой долговечностью. Поэтому применение светодиодов для целей освещения пока ограничивается небольшими объектами (светофоры, сигнальные фонари на автомобилях, карманные фонарики, подсветка мобильных телефонов, элементы рекламы). Согласно существующим прогнозам, в скором будущем светоотдача белых светодиодов достигнет 200 лм/Вт, стоимость их производства заметно снизится и они станут конкурентоспособными по отношению к люминесцентным лампам.

Светодиоды на основе органических полупроводников. Конкурентами источников белого света и светодиодов на основе полупроводников $A^{III}B^{V}$ и $A^{II}B^{VI}$ являются источники на основе *органических полупроводников*. То, что многие органические вещества обладают весьма интенсивной экситонной люминесценцией, было известно давно. Специалисты фирмы Eastman Kodak обнаружили явление электролюминесценции в органических полупроводниках еще в 70-х годах, однако чрезвычайно высокие напряжения (>1000 В), которые требовались для ее возбуждения, не позволяли практически использовать этот интересный эффект.

Существенный прорыв в создании источников света на основе органических полупроводников был сделан Тангом и ван Слайком [313], которые предложили использовать многослойную структуру (по сути *гетеропереход*) для осуществления эффективной инжекции электронов и дырок в органический материал. Это позволило им получить внешний квантовый выход 0,01 фотон/электрон, яркость >1000 кд/м^2 и светоотдачу 1,5 лм/Вт в зеленой области спектра при рабочих напряжениях менее 10 В. К настоящему времени найдены целые классы органических веществ с красным, зеленым, синим (и белым при их сложении) свечением, яркость люминесценции которых >1000 кд/м^2, а долговечность — 20–40 тысяч часов [314].

Основной трудностью в разработке светодиодов из органических полупроводников является создание условий для эффективной инжекции электронов и дырок в структуру и их доставки в область с высокой вероятностью излучательной рекомбинации. Энергетическая диаграмма одной из современных конструкций этих светодиодов показана на рис. 7.29 (указанные в ней значения энергий отсчитываются от уровня вакуума). Гетеропереходная структура состоит из двух очень тонких (несколько сотен Å)

слоев HMTPD и Alq_3, в которых по зонам HOMO и LUMO [1]) происходит перенос инжектируемых из контактов электронов и дырок, и тонкой области TAZ, в которую добавлены металлорганические комплексы $Ir(ppy)_3$ или $Ir(ppy)_2(acac)$ с очень эффективной фосфоресценцией. В этой области электрон и дырка встречаются, образуют экситон, возбуждение которого передается металлорганическому комплексу. Для эффективной инжекции дырок в широкозонный полупроводник контакт к HMTPD должен иметь высокую работу выхода (обычно им является ITO — оптически прозрачный проводящий окисел In_2O_3, легированный Sn), а для инжекции электронов в Alq_3 используют металл с низкой работой выхода (Ca или Mg с примесью Ag). Поскольку высота барьера для дырок на контакте тем не менее остается высокой (~0,9 эВ), инжекция в полупроводник ограничивается термоэлектронной эмиссией или же (при высоких приложенных напряжениях) происходит путем туннелирования через барьер в сильном электрическом поле (по механизму Фаулера–Нордгейма). Трудности инжекции и низкая подвижность носителей ($\sim 10^{-5}$ см^2/В · с) определяют довольно высокое падение напряжения в прямом направлении на светодиодах из органических полупроводников.

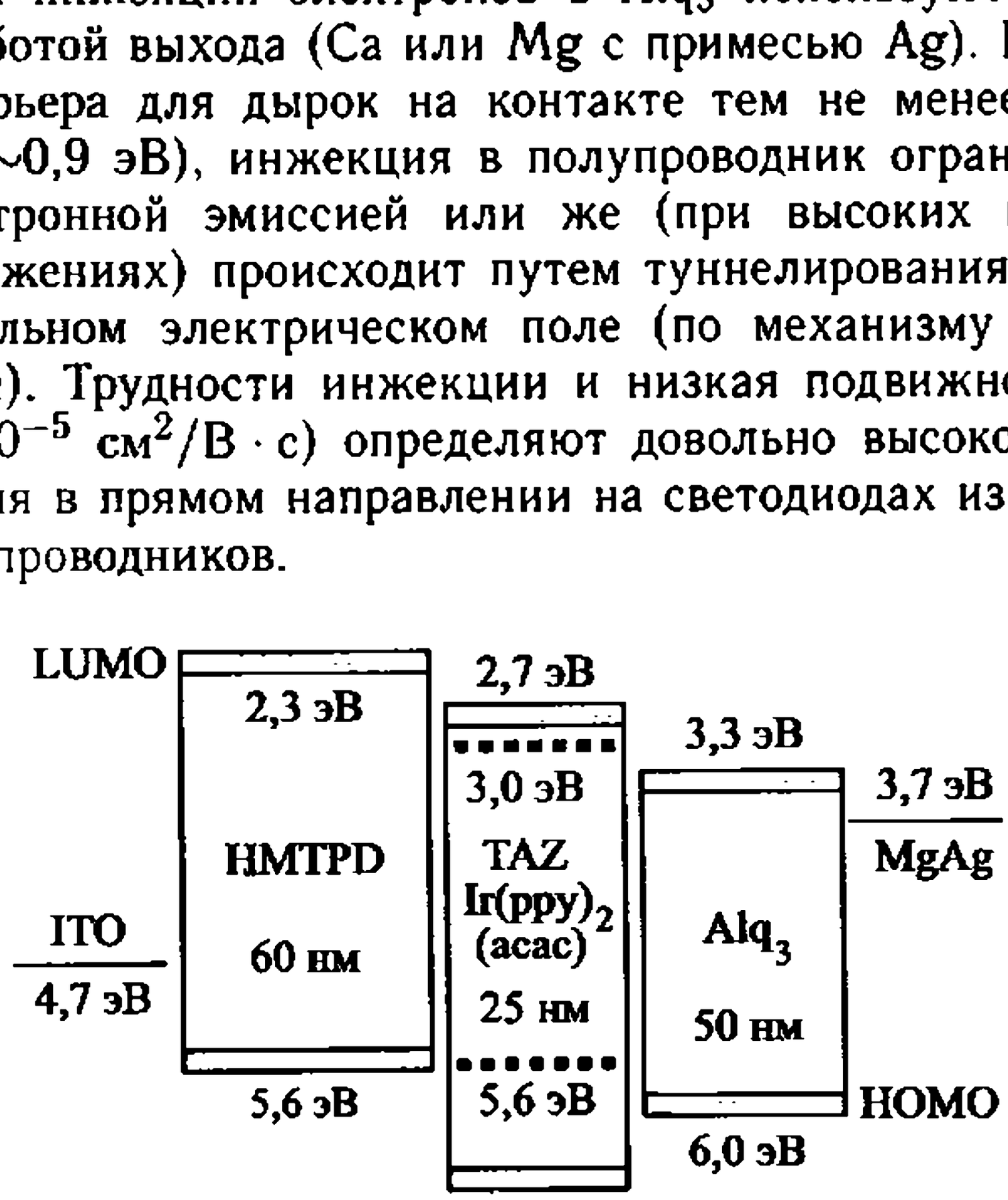


Рис. 7.29. Энергетическая диаграмма светодиода на основе органического полупроводника [315]. HMTPD, Alq_3, TAZ и $Ir(ppy)_2(acac)$ — аббревиатуры довольно длинных названий органических соединений. Квадратными точками показано энергетическое положение орбиталей HOMO и LUMO в красителе $Ir(ppy)_2(acac)$

[1]) Аббревиатуры HOMO (highest occupied molecular orbital) и LUMO (lowest unoccupied molecular orbital) обозначают самую верхнюю заполненную и самую нижнюю незаполненную молекулярные орбитали.

Внешний квантовый выход светодиодов с зеленым цветом свечения и активным слоем, содержащим $Ir(ppy)_3$, по данным работы [316] составляет 8%, а светоотдача — 31 лм/Вт; позже на похожей структуре был получен внешний квантовый выход 19% и светоотдача 60 лм/Вт при внутреннем квантовом выходе 87% [315]. Эти результаты близки к лучшим результатам, полученным на светодиодах из полупроводников $A^{III}B^{V}$. Фирма Eastman Kodak совместно с другими производителями бытовой техники уже освоила промышленный выпуск небольших дисплеев на основе органических полупроводников для цифровых фотоаппаратов и мобильных телефонов и проводит эксперименты по созданию мониторов для компьютеров и телевизоров, которые по своим параметрам (толщине, яркости свечения, экономичности) уже сейчас превосходят жидкокристаллические LCD-мониторы.

7.2.3. Инжекционные полупроводниковые лазеры. Анализируя формулу Планка для спектра излучения абсолютно черного тела, Эйнштейн [317] пришел к выводу, что наряду со спонтанными (самопроизвольными) излучательными переходами должны существовать и *вынужденные* переходы, вероятность которых пропорциональна плотности энергии электромагнитных колебаний в среде (см. рис. 7.30). При таких переходах второй

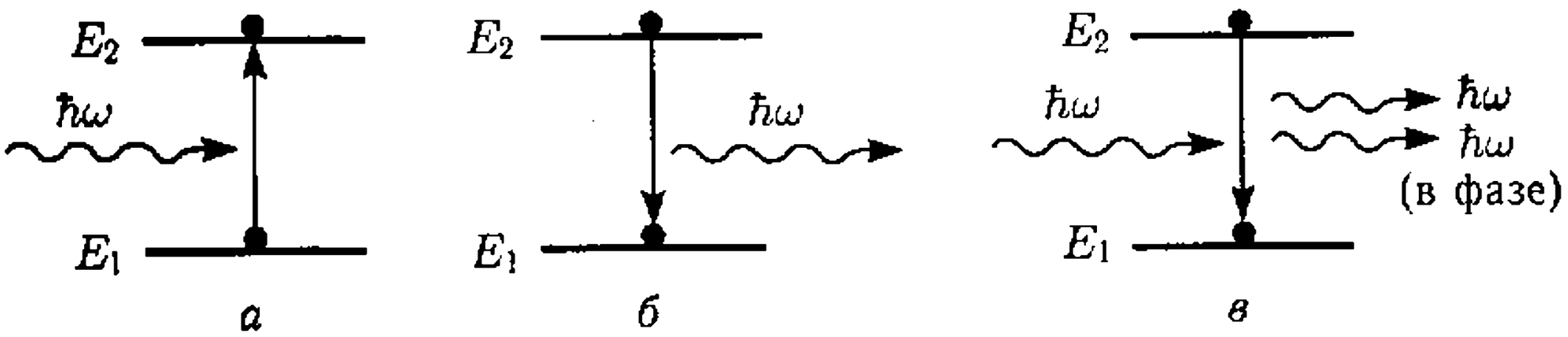


Рис. 7.30. Три типа оптических переходов между двумя энергетическими уровнями: *а* — поглощение, *б* — спонтанное излучение, *в* — вынужденное излучение

фотон, возникающий при прохождении первого фотона через вещество, обладает теми же частотой, фазой, поляризацией и направлением распространения, что и первый фотон. В состоянии термодинамического равновесия отношение темпа вынужденных переходов к темпу спонтанных переходов равно

$$1/[\exp(\hbar\omega/kT) - 1],$$

то есть исчезающе мало [302]. Однако если в среде создать высокую плотность электронного возбуждения, то ситуация может существенно измениться. Так, если вероятность вынужденного перехода окажется выше вероятности поглощения фотона,

то в среде возникнет *оптическое усиление* и в ней можно осуществить генерацию вынужденного излучения.

Впервые генерация вынужденного излучения была осуществлена в микроволновой (СВЧ) области в 1954 г., а в видимой области спектра — в 1960 году. Теоретические основы возможности возбуждения вынужденного излучения в полупроводниках были заложены в 1959–1961 гг. [318]. Генерацию вынужденного излучения в прямозонном полупроводнике GaAs впервые наблюдали в 1962 г. практически одновременно авторы работ [319–321]. Источники вынужденного микроволнового излучения получили название *мазеров*, а источники вынужденного излучения в видимой области и близких к ней инфракрасной и ультрафиолетовой областях спектра — *лазеров* [1]).

Полупроводниковые лазеры находят применение в устройствах чтения и записи компактных дисков (CD, DVD), в телекоммуникации, в системах прецизионного измерения перемещений, в спектроскопии высокого разрешения, для накачки твердотельных лазеров, в областях, связанных с обработкой материалов (резка, микросварка), в медицине (лазерная хирургия, дерматология) [322]. Мощность излучения одиночного лазера может достигать 5–10 Вт, а специально разработанных лазерных модулей — 100 Вт в квазинепрерывном режиме при к.п.д. 45–65 %. Долговечность современных лазеров измеряется миллионами часов.

Найдем условия возникновения оптического усиления в прямозонном полупроводнике. Рассмотрим оптические переходы с энергией кванта $\hbar\omega$ между состояниями зоны проводимости и валентной зоны. Если предположить, что матричный элемент перехода не зависит от энергии электрона E, то вероятность поглощения кванта при его взаимодействии с полупроводником равна

$$W_{\text{погл}}(\hbar\omega) = B' \int \rho_v(E - \hbar\omega)\rho_c(E) f_v(E - \hbar\omega)[1 - f_c(E)]dE, \tag{7.41}$$

а вероятность вынужденного излучения —

$$W_{\text{вын}}(\hbar\omega) = B' \int \rho_v(E - \hbar\omega)\rho_c(E) f_c(E)[1 - f_v(E - \hbar\omega)]dE. \tag{7.42}$$

---

[1]) Термины maser и laser являются аббревиатурами английских выражений microwave amplification by stimulated emission of radiation и light amplification by stimulated emission of radiation.

Разность этих двух интегралов, равная

$$W(\hbar\omega) = B' \int \rho_v(E - \hbar\omega)\rho_c(E)[f_c(E) - f_v(E - \hbar\omega)]dE, \quad (7.43)$$

и определяет, будет ли распространяющаяся в полупроводнике волна усиливаться или ослабляться. Усиление возможно только тогда, когда вероятность вынужденного испускания второго фотона первичным фотоном превышает вероятность поглощения первичного фотона. Если предположить, что распределение носителей в разрешенных зонах описывается квазиуровнями Ферми F_n^* и F_p^* (см. формулу (1.18)), то из формулы (7.43) следует, что для возникновения оптического усиления необходимо выполнение условия

$$F_n^* - F_p^* > \hbar\omega, \quad (7.44)$$

при котором для любой пары энергетических уровней, разделенных зазором $\Delta E = \hbar\omega$, выполняется соотношение $f_c > f_v$. Такое состояние называется *инверсной населенностью*. Если в полупроводнике нет примесных уровней и хвостов плотности состояний, то из уравнений (7.43) и (7.44) следует, что для возникновения в нем вынужденного излучения необходимо, чтобы электроны и дырки были вырождены хотя бы в одной из зон.

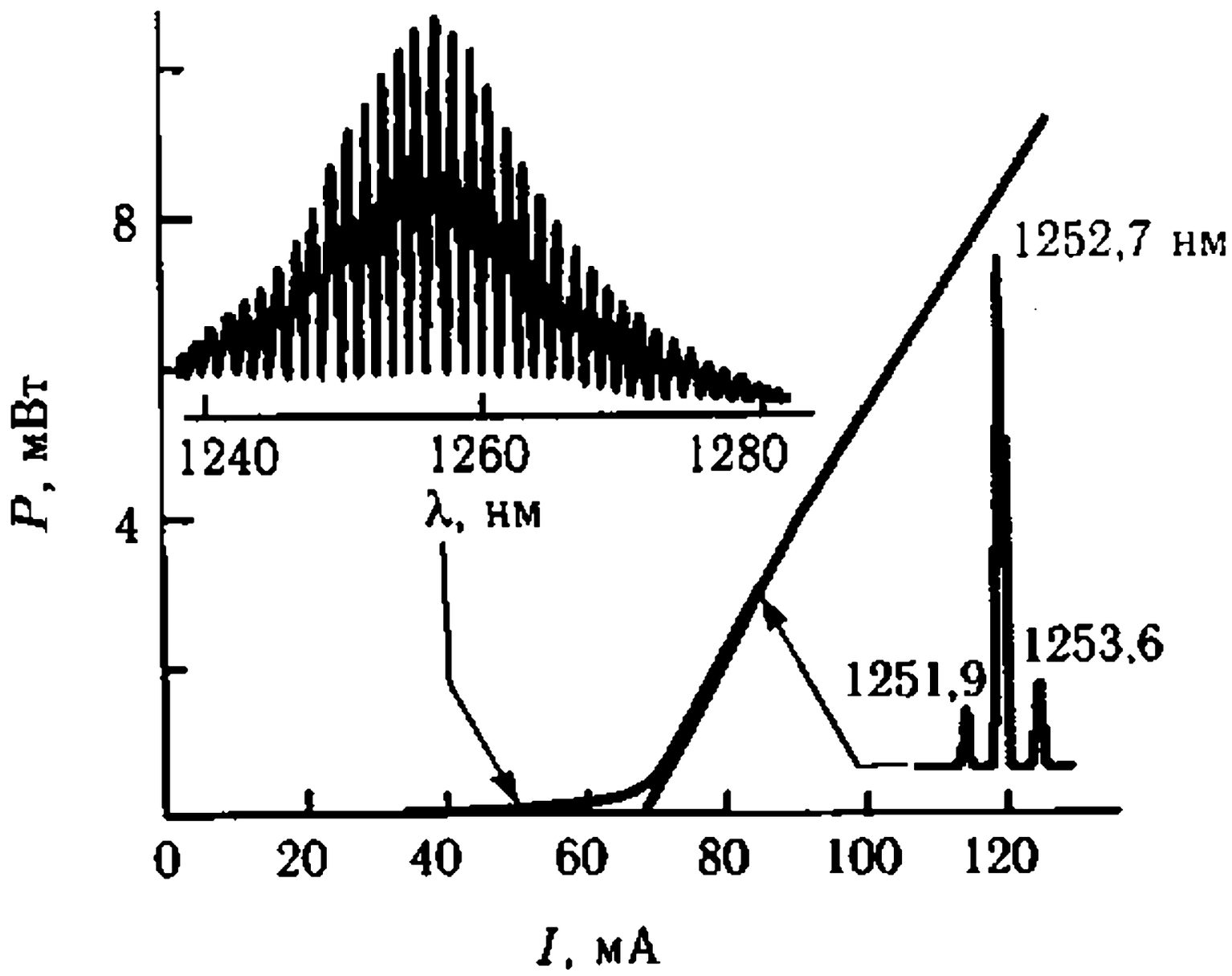


Рис. 7.31. Зависимость выходной мощности лазера на двойной гетероструктуре InP/InGaAsP/InP от тока диода и качественный вид спектров излучения ниже ($I = 50$ мА) и выше ($I = 85$ мА) порога генерации [323]. Интерференция в спектре спонтанного излучения обусловлена тем, что излучение выводится из резонатора Фабри–Перо

Однако одной только инверсной населенности и оптического усиления в среде недостаточно для создания лазера, поскольку возникающие спонтанным образом фотоны, усиленные в активной области, затем уходят в пассивные области кристалла, где происходит их поглощение. Для обеспечения *когерентности* колебаний необходимо, чтобы часть фотонов, прошедших через активную область, возвращалась в нее и поддерживала оптическую обратную связь. Иными словами, активная область лазера должна быть помещена в *оптический резонатор*. В первых лазерах из GaAs для создания резонатора использовалась замечательная способность этих кристаллов раскалываться по плоскостям спайности (110). Используя это свойство, в образце можно было создать две практически идеально параллельные грани, то есть поместить активную среду в резонатор Фабри–Перо. Когда оптическое усиление в активной области начинает превышать потери на вывод излучения из резонатора, мощность выходного излучения лазера резко (пороговым образом) возрастает, а в спектрах появляется модовая структура — узкие линии, отвечающие определенным модам собственных колебаний резонатора (см. рис. 7.31). Это и есть лазерная генерация.

Найдем условие возникновения лазерной генерации. Предполагая, что электронно-дырочные пары, создаваемые путем пропускания тока инжекции плотностью J, рекомбинируют с квантовым выходом η в активной области толщиной d и спектр излучения при этом имеет среднюю энергию ν и полуширину $\Delta\nu$, Лэшер и Стерн [324] рассчитали линейный коэффициент оптического усиления g в полупроводнике:

$$g = \beta_L J, \quad \beta_L = \frac{c^2\eta}{8\pi q n_r^2 \nu^2 \Delta\nu d}, \tag{7.45}$$

где n_r — показатель преломления среды, а c — скорость света в вакууме. В действительности коэффициент оптического усиления изменяется с током по закону

$$g = \beta_L(J - J_0), \tag{7.46}$$

где J_0 — плотность тока, необходимая для достижения инверсной населенности. Пусть коэффициент отражения зеркал резонатора лазера равен R, длина резонатора равна L, доля полной световой энергии в резонаторе, которая приходится на активную область структуры (*коэффициент оптического ограничения*), равен Γ, а коэффициент дополнительных потерь излучения за счет поглощения на свободных носителях в активной области

и рассеяния света на дефектах равен α. Тогда условием прохождения волны через резонатор без затухания будет

$$\Gamma g = \alpha + \frac{1}{L} \ln\left(\frac{1}{R}\right). \tag{7.47}$$

Из уравнений (7.46) и (7.47) легко найти пороговую плотность тока:

$$J_{\text{пор}} = J_0 + \frac{1}{\Gamma \beta_L}\left[\alpha + \frac{1}{L} \ln\left(\frac{1}{R}\right)\right]. \tag{7.48}$$

Из анализа этого уравнения и уравнения (7.45) следует, что для уменьшения пороговой плотности тока надо увеличивать длину резонатора L, уменьшать потери на поглощение α, уменьшать толщину активной области d (сохраняя при этом высокое значение Γ) и работать на переходах, имеющих наименьшую ширину спектра излучения ($\Delta\nu$).

Понятно, что для получения лазерной генерации необходим очень высокий уровень возбуждения электронной системы. Его можно создать с помощью оптической накачки, возбуждением пучком быстрых электронов или других заряженных частиц, ударной ионизацией, однако наиболее удобным и экономичным способом возбуждения полупроводника является *инжекция* неравновесных носителей с помощью p–n-перехода. В первых инжекционных лазерах на основе p–n-переходов в GaAs для достижения лазерной генерации при 300 К была необходима плотность тока $\sim 5 \cdot 10^4$ А/см^2.

Инжекционные лазеры на гетероструктурах. Чтобы понизить $J_{\text{пор}}$, необходимо создать условия для наиболее благоприятного взаимодействия возбужденной до состояния инверсной населенности электронной подсистемы с оптической подсистемой. Для этого, как следует из анализа формулы (7.48), надо локализовать электронное возбуждение в возможно более тонком слое полупроводника, в котором квантовый выход люминесценции наиболее высок. Алфёров и Казаринов [325] и, независимо, Крёмер [326] в 1963 г. предложили использовать для этой цели *гетеропереход*, который, как мы видели в п. 1.6.1, позволяет управлять направлением инжекции.

Лазеры с односторонней гетероструктурой действительно имели в несколько раз меньшую пороговую плотность тока по сравнению с p–n-переходами ($\sim$8 кА/см^2 при 300 К [322, 327]). Понижению $J_{\text{пор}}$ способствовал и дополнительный скачок показателя преломления на гетерогранице (см. рис. 7.32*б*), который ограничивал проникновение световой волны в пассивную

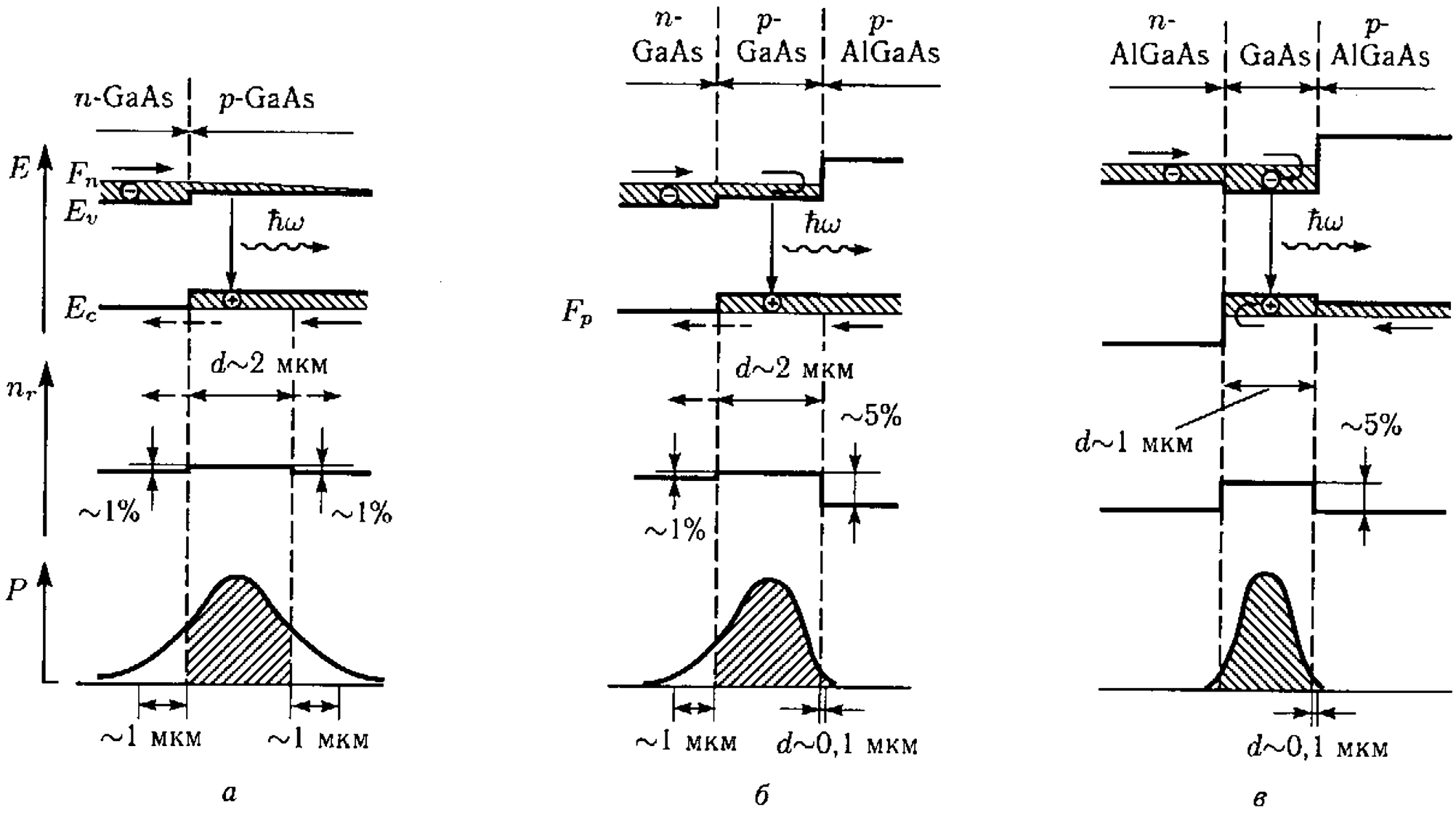


Рис. 7.32. Энергетическая диаграмма, профиль показателя преломления и распределение световой мощности в полупроводниковых лазерах с p–n-переходом (*а*), односторонней гетероструктурой (*б*) и с двойной гетероструктурой (*в*) [57]

область структуры и тем самым усиливал взаимодействие волн с электронной подсистемой и уменьшал ее поглощение в пассивной области. Еще более сильного снижения пороговой плотности тока удалось добиться в двойных гетероструктурах. В этих структурах (см. рис. 7.32 *в*) гетеропереходы одновременно локализуют и электронное, и оптическое возбуждение в тонкой активной области. При этом для достижения инверсной населенности не требуется сильно легировать активный слой, так как эффективная рекомбинация в нем осуществляется за счет двойной инжекции.

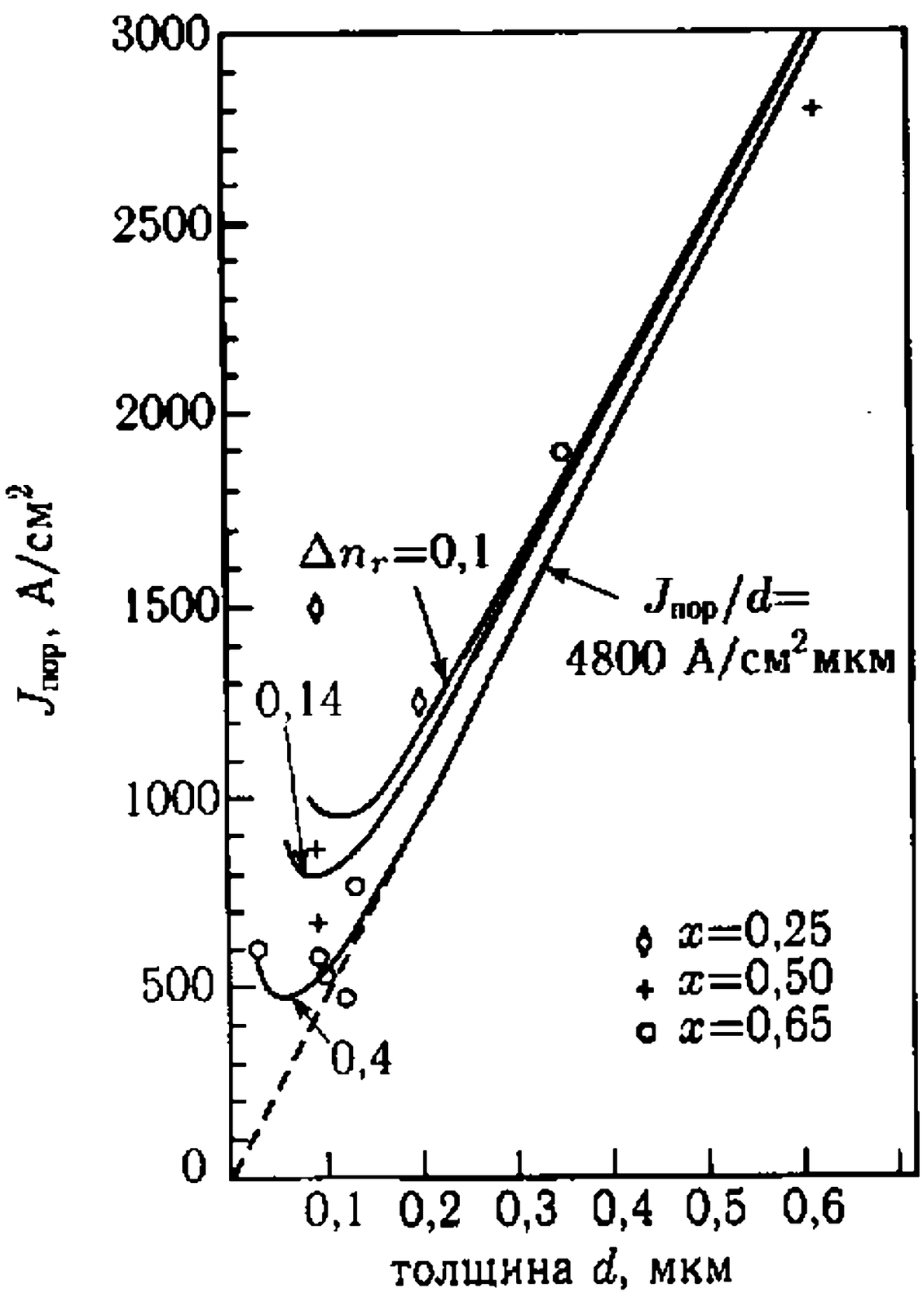


Рис. 7.33. Зависимость пороговой плотности тока от толщины активной области в двойной гетероструктуре $Al_xGa_{1-x}As/GaAs$ [52]

В соответствии с формулами (7.45) и (7.48), плотность тока, необходимая для получения лазерной генерации, уменьшается с уменьшением толщины активной области d; эксперимент (см. рис. 7.33) подтверждает этот вывод. Успехи в технологии выращивания структур с тонкими слоями позволили в 1970 г. осуществить режим непрерывной генерации при 300 К на лазерах с двойной гетероструктурой AlGaAs/GaAs/AlGaAs [328, 329]. Как следует из рис. 7.33, минимальная пороговая плотность тока (~500 А/см^2 для лазеров с активной областью из GaAs)

достигается при $d = 0,1$ мкм. При дальнейшем сужении активной области выход излучения за пределы оптического волновода уменьшает эффективность взаимодействия оптической волны сЁэлектронным возбуждением (коэффициент оптического ограничения Г) и пороговая плотность тока начинает возрастать. Для преодоления этих ограничений в современных лазерах создают более сложные структуры, которые мы обсудим ниже.

После разработки лазеров с двойной гетероструктурой, способных работать в непрерывном или квазинепрерывном режиме, возникла задача уменьшения площади активной области с целью понижения рабочих токов лазеров до уровней, используемых в традиционной электронике. На рис. 7.34 показаны две конструкции инжекционного лазера с полосковой геометрией.

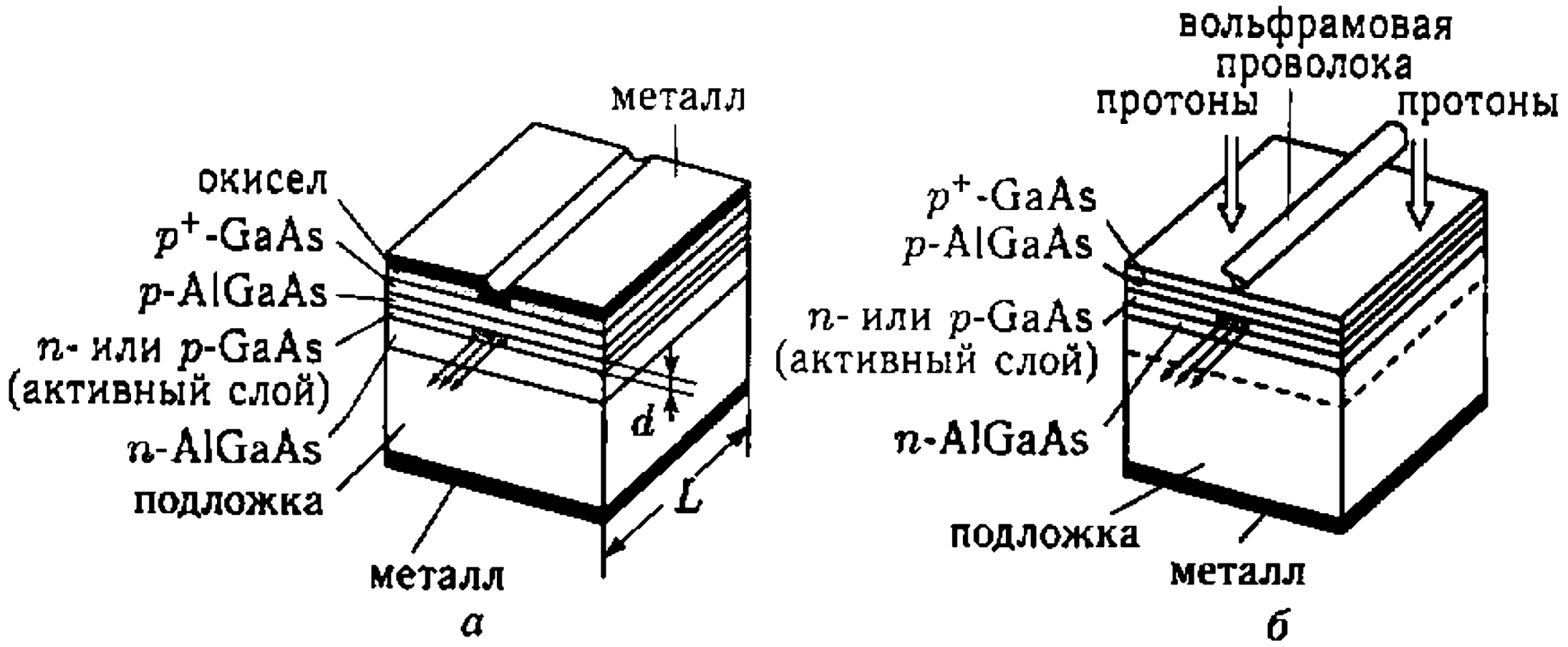


Рис. 7.34. Конструкции лазеров с двойной гетероструктурой и полосковой геометрией, созданной с использованием окисной изоляции (*а*) и путем бомбардировки незащищенной поверхности структуры протонами (*б*) [14]

Создать активную область в виде узкой полоски, перпендикулярной граням резонатора, можно либо ограничивая область электрического контакта к верхнему слою гетероструктуры с помощью окисной изоляции (см. рис. 7.34 *а*), либо сильно увеличивая удельное сопротивление верхнего слоя структуры за пределами активной области путем ее бомбардировки протонами (см. рис. 7.34 *б*). Кроме уменьшения порогового тока, полосковая геометрия необходима еще и для обеспечения одномодового характера генерации лазера, поскольку в лазерах с шириной активной области более ~20 мкм обычно одновременно возбуждаются несколько мод колебаний, что снижает степень когерентности излучения лазера и увеличивает ширину спектра излучения [1]).

[1]) Узкая линия излучения имеет первостепенное значение при использовании лазеров в волоконно-оптических линиях связи. Из-за дисперсии пока-

Одним из способов преодоления ограничений, связанных с выходом излучения за пределы активной области в двойных гетероструктурах с $d < 0,1$ мкм, является создание так называемой *структуры с раздельным ограничением* (SCH — separate confinement heterostructure или, что то же самое, LOC — large optical cavity) [330]. В этой структуре (см. рис. 7.35) электронное возбуждение локализовано в наиболее узкозонной части структуры — квантовой яме, ширина которой может достигать нескольких десятков ангстрем, а волноводом служит область шириной порядка 1 мкм, ограниченная с двух сторон двумя наиболее широкозонными областями. Первый инжекционный лазер с такой структурой и активной областью шириной 200 Å был создан в [331]. Структура с раздельным ограничением позволяет еще сильнее снизить пороговую плотность тока (до $\sim$ $\sim$100 А/см^2 в структуре с одной квантовой ямой [332]). Для уменьшения пороговой плотности тока в волноводной области структуры можно расположить не одну, а несколько квантовых ям. Дальнейшее улучшение характеристик лазеров достигается в структурах, в которых в области оптического ограничения создается градиент показателя преломления (GRINSCH — graded index separate confinement heterostructure). Структура с раздельным ограничением имеет еще одно преимущество над двойной гетероструктурой: она характеризуется более высокой стойкостью по отношению к катастрофическим отказам, связанным с разрушением зеркал оптического резонатора при высокой мощности выходного излучения [330].

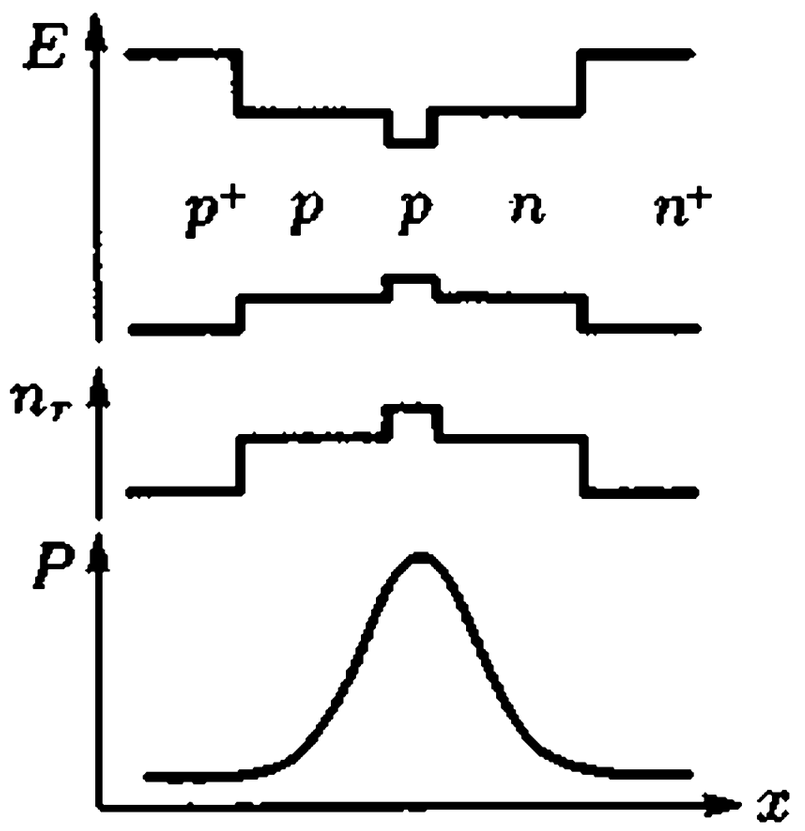


Рис. 7.35. Энергетическая диаграмма, профиль показателя преломления и распределение мощности световой волны в структуре с раздельным ограничением

Дальнейшее развитие инжекционных лазеров связано с расширением круга используемых в гетероструктурах материалов. На рис. 7.36 показана зависимость ширины запрещенной зоны (и соответствующая ей длина волны) для ряда тройных твердых растворов соединений $A^{III}B^{V}$ со структурой сфалерита как

зателя преломления излучение с разной длиной волны распространяется по волокну с разной скоростью. При большой длине канала передачи это вызывает уширение принимаемого оптического импульса и может сделать невозможной высокую скорость передачи данных.

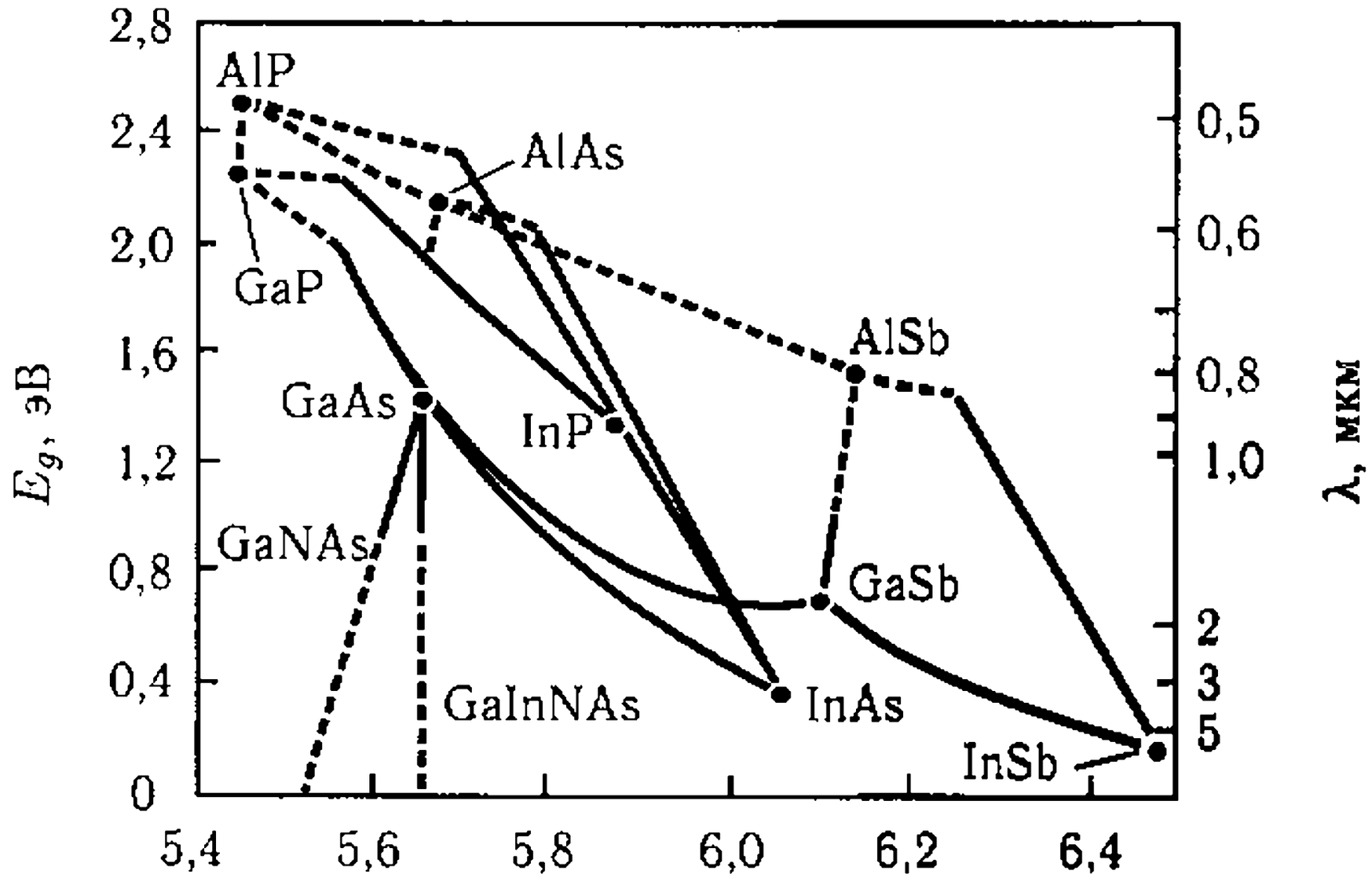


Рис. 7.36. Зависимость ширины запрещенной зоны и длины волны, отвечающей краю собственного поглощения, в тройных твердых растворах полупроводников группы $A^{III}B^{V}$ от параметра кристаллической решетки. Сплошные линии — прямозонные полупроводники, штриховые линии — непрямозонные полупроводники

функция параметра кристаллической решетки. Эта диаграмма позволяет найти E_g и параметры состава любой пары тройных твердых растворов, согласованных по параметру решетки, которые необходимы для создания лазера с требуемой длиной волны. Использование аналогичной диаграммы для четверных твердых растворов позволяет заметно расширить область поиска и рассчитать составы согласованной по параметру решетки гетеропереходной пары, которая может быть выращена на широко доступных подложках (например, на GaAs или InP). Так, взяв одним из материалов пары InP и подбирая параметры состава (x, y) четверного твердого раствора $In_xGa_{1-x}As_yP_{1-y}$, можно определить значения этих параметров, необходимые для создания лазеров с $\lambda = 1,3$ и 1,55 мкм (на этих длинах волн в кварцевом волокне находятся минимумы коэффициента поглощения).

Разработанные к настоящему времени инжекционные полупроводниковые лазеры, работающие на межзонных переходах, перекрывают широкий диапазон электромагнитного спектра от ультрафиолетовой области (GaN, $\lambda \approx 0,36$ мкм) до средней инфракрасной области спектра ($Pb_{1-x}Sn_xTe_{1-y}Se_y$, $\lambda \approx 46$ мкм). Продвинуться в более длинноволновую область (до 130 мкм) можно, используя *квантово-каскадные лазеры*, которые работают на переходах между уровнями размерного квантования в сверхрешетках [333].

Инжекционные лазеры с вертикальным резонатором. Сильное уменьшение объема активной области, сопровождаемое существенным понижением порогового тока лазеров, удалось реализовать в разработанных в 1979 г. структурах с *вертикальным резонатором* (VCSEL — vertical cavity surface emitting laser) [334] [1]. Пороговый ток в таких лазерах может составлять менее 10 мкА [335].

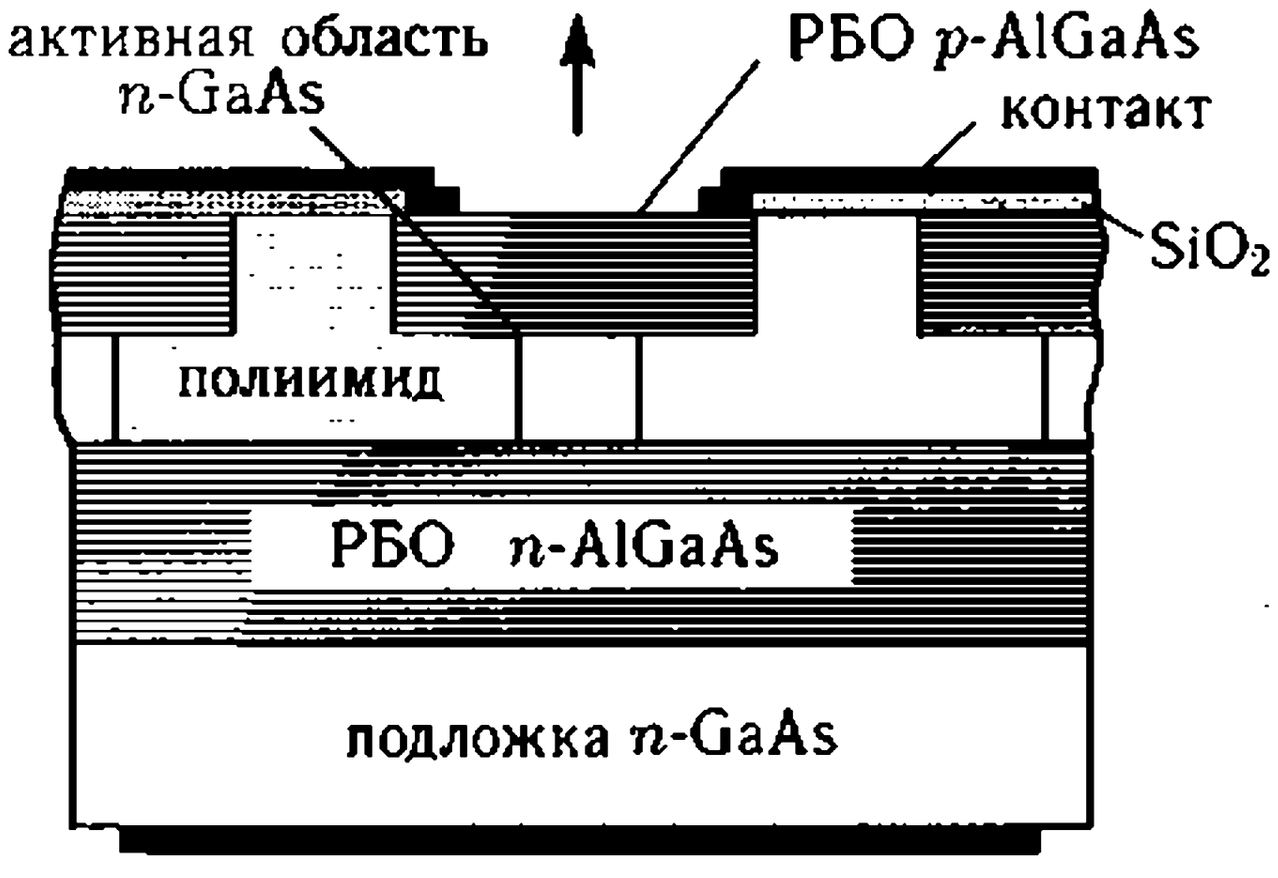


Рис. 7.37. Конструкция лазера с вертикальным резонатором [336]

Одна из конструкций VCSEL-лазера показана на рис. 7.37 [336]. В этой конструкции активная область из n-GaAs толщиной $L = 2$ мкм располагается между двумя горизонтально расположенными многослойными «зеркалами», которые изготавливаются из изотипных гетероструктур $Al_{0,05}Ga_{0,95}As/Al_{0,53}Ga_{0,47}As$ p- и n-типа и одновременно служат широкозонными эмиттерами. При этом вывод излучения происходит в вертикальном направлении через верхнюю поверхность структуры. Из-за очень малой толщины активной области для получения низкого порога генерации в соответствии с формулой (7.48) необходимо, чтобы зеркала имели высокий коэффициент отражения R. В качестве зеркал в лазерах с вертикальным резонатором обычно используют *распределенные брэгговские отражатели* (РБО), построенные из чередующихся слоев материалов с заметно различающимся показателем преломления (GaAs/AlGaAs, GaAs/Al_2O_3, ZnSe/MgF_2), каждый из которых имеет толщину $\lambda/4$. При числе слоев от 50 до 80 на таких зеркалах удается получить $R > 99,95\%$.

[1]) В отечественной литературе эти структуры также называют вертикально излучающими лазерами.

Большое внимание, уделяемое в настоящее время VCSEL-лазерам, связано с тем, что сопряжение таких лазеров с оптоволоконным кабелем оказывается намного проще по сравнению с лазерами, в которых излучение выводится через боковую поверхность. Лазеры с вертикальным резонатором имеют более высокую надежность и более низкую стоимость. Низкие пороговые токи и близкая к планарной технология изготовления лазерных структур позволяют создавать из таких лазеров двумерные массивы излучателей, что невозможно при других конструкциях инжекционных лазеров. Наконец, из-за бóльших геометрических размеров области вывода излучения диаграмма направленности этих лазеров оказывается ýже, чем для лазеров с боковым выводом излучения. Выходная мощность VCSEL-лазеров достигает 1,5 Вт [337].

Инжекционные лазеры на квантовых ямах, нитях и точках. В лазерных структурах с узкими квантовыми ямами дополнительному уменьшению пороговой плотности тока способствует и *качественное изменение энергетической зависимости плотности состояний* в двумерных структурах. Об изменении плотности состояний в структурах пониженной размерности мы уже говорили в п. 1.6.2.

Понять причину уменьшения $J_{пор}$ при понижении размерности можно из анализа рис. 7.38 [338]. Для того, чтобы в полупроводнике могла возникнуть лазерная генерация, значение оптического усиления в максимуме спектра должно стать больше величины, определяемой потерями в резонаторе.

Чтобы этого достигнуть, в полупроводнике необходимо создать сильное электронное возбуждение, однако при этом, как следует из рис. 7.38 *а*, в объемном полупроводнике распределение электронов по энергиям таково, что бóльшая часть возбужденных электронов никогда не будет давать вклад в оптическое усиление.

Иная ситуация складывается в двумерных структурах (квантовых ямах). Благодаря ступенчатой зависимости плотности состояний от энергии (см. рис. 7.38 *б*) в них значительная бóльшая часть возбужденных электронов с энергией вблизи края подзоны дает вклад в оптическое усиление. То есть из-за более эффективного согласования энергетической зависимости плотности состояний со спектрами усиления для возникновения лазерной генерации в двумерных структурах требуется более низкий уровень возбуждения.

Для одномерных (квантовые нити) и нульмерных структур (квантовые точки) энергетическая зависимость плотности

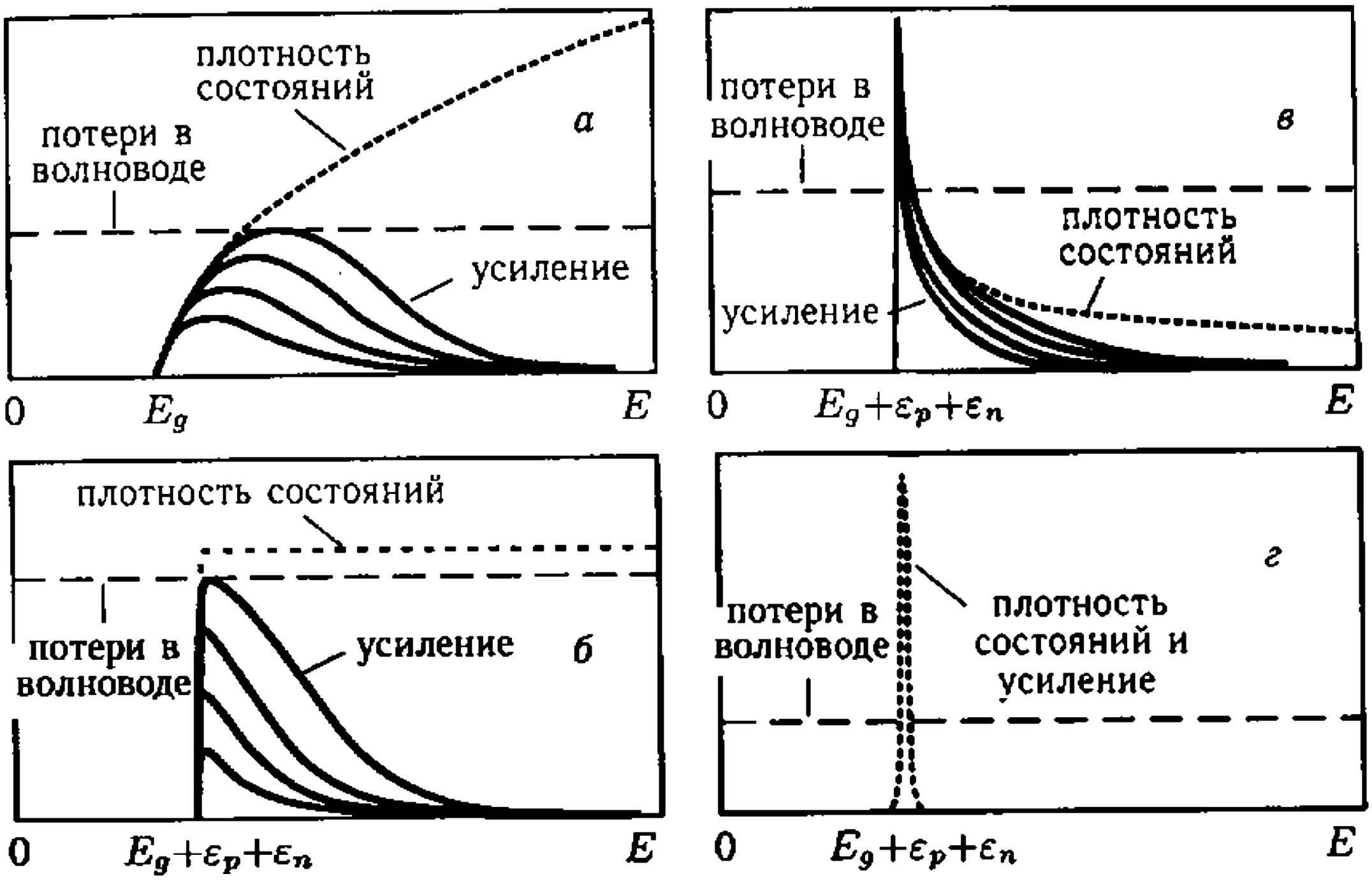


Рис. 7.38. Условия возбуждения лазерной генерации в полупроводниковых структурах различной размерности [338]: *а* — объемный полупроводник, *б* — квантовая яма, *в* — квантовая нить, *г* — квантовая точка

состояний все лучше и лучше согласуется со спектрами оптического усиления, что приводит к уменьшению числа электронных состояний, которые надо заполнить для достижения требуемого усиления. Это позволяет ожидать существенного снижения плотности порогового тока в таких структурах. Таким образом, из рис. 7.38 следует, что лазеры на квантовых точках являются наиболее перспективным типом полупроводниковых лазеров. К достоинствам лазеров на квантовых точках следует также добавить возможность гибкого управления длиной волны излучения путем изменения размера квантовых точек.

Лазеры на полупроводниковых структурах пониженной размерности имеют еще одно преимущество по сравнению с лазерами на основе объемных полупроводников — более слабую температурную зависимость порогового тока. Основной причиной быстрого увеличения $J_{\text{пор}}$ с ростом температуры является тепловое «размытие» функций распределения участвующих в излучательной рекомбинации носителей, из-за которого для достижения необходимого значения коэффициента оптического усиления требуется инжектировать все бóльшие концентрации носителей. Однако как показали Аракава и Сакаки [339], в структурах пониженной размерности из-за лучшего согласования энергетической зависимости плотности состояний

и спектров оптического усиления влияние теплового размытия должно сильно ослабевать. Этот эффект, действительно наблюдающийся в эксперименте, позволяет заметно расширить диапазон рабочих температур инжекционных лазеров.

Для реализации потенциальных возможностей инжекционных лазеров на структурах пониженной размерности в конце 80-х годов были начаты работы по созданию таких лазеров. Первые опыты по созданию квантовых нитей с использованием фотолитографии, травления и последующего заращивания нитей позволили получить лазерные структуры, которые работали только при низких температурах, характеризовались заметной неоднородностью (связанной с разбросом поперечных размеров отдельных квантовых нитей) и в которых отсутствовали особенности, свидетельствующие об одномерном характере движения носителей. Первый работающий при 300 К лазер на квантовых нитях был создан в 1989 году в системе AlGaAs/GaAs [340]. В нем квантовые нити из GaAs, окруженные слоями из AlGaAs, выращивались в V-образных канавках, которые шли в направлении $[01\bar{1}]$ и были получены путем избирательного травления подложек GaAs с ориентацией (100) (см. также с. 208). Невысокий коэффициент оптического ограничения Г, характерный для структуры с одиночной квантовой нитью, был впоследствии увеличен путем выращивания «стопки» из вертикально связанных квантовых нитей и созданием массивов квантовых нитей при выращивании короткопериодных сверхрешеток на вицинальных поверхностях GaAs или с использованием явлений самоорганизации при выращивании таких сверхрешеток на рассогласованных по параметру решетки подложках [341].

Дальнейший прогресс в создании лазеров на полупроводниковых структурах пониженной размерности был связан с развитием неожиданно открывшейся возможности формирования массивов практически одинаковых по размеру и форме квантовых точек.

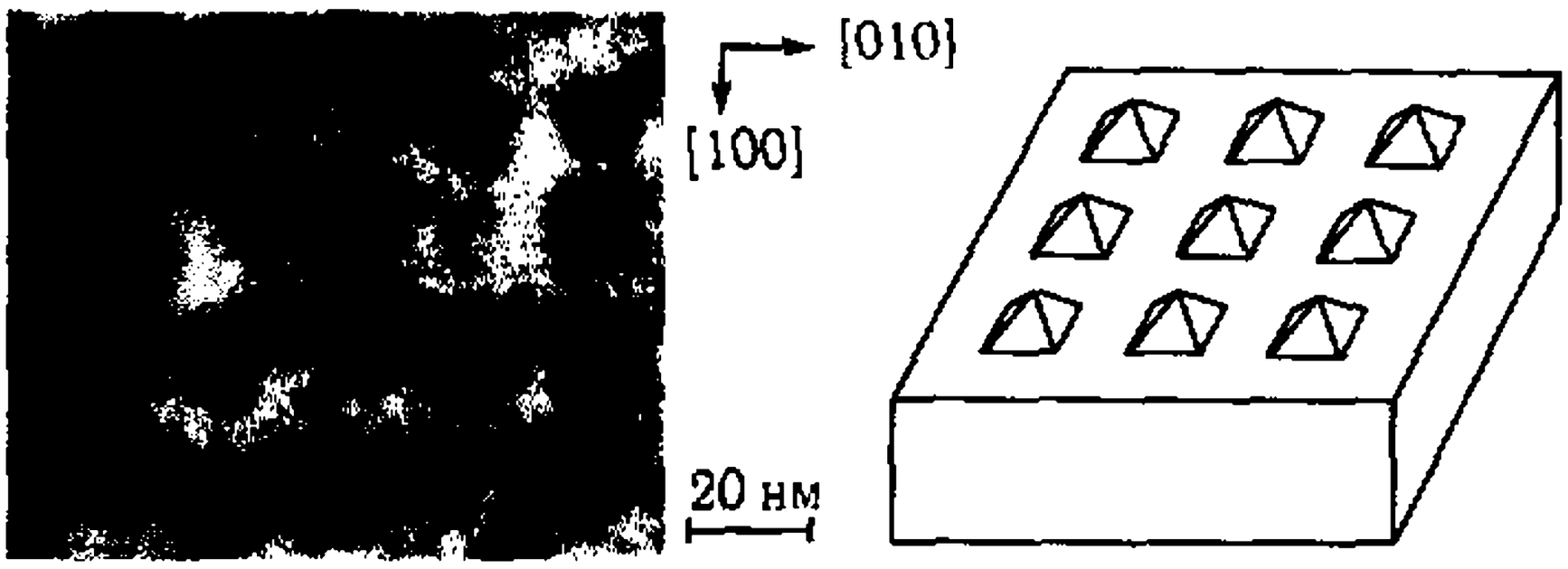
Возможность образования маленьких островков полупроводника при субмонослойном покрытии подложки (то есть когда количества нанесенного на подложку вещества не хватает для образования одного монослоя) была известна давно, однако слишком большая дисперсия размеров островков делала их малоперспективными для исследований. В начале 80-х годов были найдены технологические режимы, позволявшие при выращивании тонких пленок полупроводников, сильно отличавшихся по параметру решетки от материала подложки, реализовать как режим роста двумерной пленки, так и режим роста трех-

мерных кластеров. В 1985 г. было обнаружено *явление самоорганизации*, заключавшееся в формировании упорядоченных в направлении роста цепочек квантовых точек при выращивании многослойных структур InAs/GaAs [342]. Несколько позже были найдены технологические условия, при которых в одном слое образуются квантовые точки *приблизительно одинакового размера и формы* [343], а затем и условия образования этими точками практически *регулярной структуры* [344, 345]. Изображение одного слоя такой структуры в электронном микроскопе и качественный вид образующейся структуры показаны на рис. 7.39. Плотность островков в слое составляет $\sim 5 \cdot 10^{10}$ см$^{-2}$.

Получающиеся островки представляют собой пирамидки с характерным размером основания 100–150 Å и высотой 30–70 Å. При этом каждая пирамидка содержит от нескольких тысяч до нескольких десятков тысяч атомов. Вещество в пирамидке находится в когерентно напряженном состоянии, поскольку для островков небольшого размера такое состояние энергетически более выгодно, чем состояние, в котором упругие напряжения снимаются путем образования дислокаций несоответствия. Как было показано в работе [344], при определенных соотношениях между изменением поверхностной энергией при образовании пирамидки и энергией возникающих на ее ребрах упругих напряжений появляется некий оптимальный размер пирамидки, при котором ее термодинамическая энергия достигает минимума. Взаимодействие пирамидок посредством напряжений, создаваемых ими в подложке, может приводить к возникновению регулярной структуры. Поскольку размеры пирамидок достаточно малы и в них проявляется эффект ограничения движения носителей во всех трех измерениях, каждую из них можно рассматривать как квантовую точку.

Закрывая выращенный на поверхности подложки слой квантовых точек тонким слоем полупроводника, из которого изготовлена подложка, и выращивая следующие слои квантовых точек последовательно друг за другом на расстоянии $\sim$100 Å, можно создать трехмерную пространственную структуру из вертикально связанных квантовых точек.

То, что образующиеся при самоорганизованном росте пирамидки имеют приблизительно одинаковый размер и форму, приводит к тому, что электронный спектр всех квантовых точек очень близок. Это позволило сделать следующий шаг в технологии инжекционных лазеров и наблюдать сначала лазерную генерацию при 300 К при оптическом возбуждении [346], а затем создать и первый *инжекционный лазер на квантовых точках*



а *б*

Рис. 7.39. Электронно-микроскопическое изображение (*а*) и качественный вид (*б*) упорядоченного массива когерентно напряженных островков InAs на подложке GaAs [345]

в системе InAs/GaAs [347, 348] [1]). Уже сегодня в лазерах на квантовых точках достигнута рекордно низкая пороговая плотность тока 16 А/см^2 при комнатной температуре [350]; выходная мощность таких лазеров достигла 7 Вт в импульсном режиме на длине волны 1,55 мкм [351].

Кроме квантовых точек в обсуждавшейся выше системе InAs/GaAs, в настоящее время получены и достаточно подробно исследованы квантовые точки и в других полупроводниковых системах: Ge/Si, CdSe/ZnSe и др. Помимо создания полупроводниковых лазеров с низким порогом и высокой температурной стабильностью, квантовые точки начинают использовать в люминесцентных концентраторах (см. подробнее с. 421) для увеличения к.п.д. солнечных элементов [352], в фотоприемниках [353], светодиодах.

О свойствах, методах получения и последних достижениях в области создания гетероструктур с квантовыми точками можно прочитать в обзорах [345, 354].

---

[1]) Впервые лазерная генерация на квантовых точках наблюдалась в 1991 г. при оптическом возбуждении микрокристаллов CdSe, находящихся в стеклянной матрице [349].

ПРИЛОЖЕНИЕ

В табл. 1 приведены некоторые физические свойства диэлектриков, а в табл. 2 — физические свойства полупроводников, используемых при производстве полупроводниковых приборов. Сведения о других физических свойствах полупроводников можно найти в работе [306]. Сравнительно большой разброс данных для диэлектрической проницаемости в табл. 1 связан с тем, что ряд материалов может существовать в нескольких кристаллических модификациях, а низкосимметричные кристаллы к тому же могут обладать заметной анизотропией.

Таблица 1. Некоторые физические свойства диэлектриков, используемых при производстве полупроводниковых приборов ($T = 300$ К) [178, 355, 356]

Материал	E_g, эВ	ε	ρ, Ом · см	$\mathcal{E}_{\text{проб}}$, МВ/см	ΔE_c, эВ
SiO_2	9	3,9	10^{14}–10^{16}	10	3,2
Si_3N_4	5,3	7–7,9	10^{14}–10^{16}	10	2,4
Al_2O_3	8,8	9,5–12	$>10^{14}$	3–5	2,8
Y_2O_3	6	~15			2,3
La_2O_3	~6	~21			2,3
Ta_2O_5	4,4	23–28	$>3 \cdot 10^{12}$	4,5	0,36
TiO_2	3,0–3,2	60–170	10^{11}?	1	~0
ZrO_2	5,7–5,8	12–25			1,4–1,5
HfO_2	5,7–6	16–30		8,5	1,5
$ZrSiO_4$	~6	10–12			1,5
$HfSiO_4$	~6	10–25			1,5
BST*		~300			

* Твердый раствор $Ba_{0,65}Sr_{0,35}TiO_3$.

Таблица 2. Некоторые физические свойства полупроводников, используемых при производстве полупроводниковых приборов ($T = 300$ K) [306]

Материал	E_g, эВ*	μ, см^2/В · с		ε	Парам. решетки, Å
		μ_n	μ_p		
Полупроводники с решеткой типа алмаза					
C	5,47 (i)	2800	2100	5,7	3,567
Si	1,12 (i)	1450	500	11,7	5,431
Ge	0,661 (i)	3900	1900	16,2	5,658
Полупроводники с решеткой сфалерита					
GaAs	1,424 (d)	9200	400	13,2	5,653
InAs	0,354 (d)	33000	450	15,1	6,058
AlAs	2,15 (i)	280	200	10,1	5,660
GaP	2,27 (i)	160	140	11,1	5,451
InP	1,34 (d)	5000	150	12,5	5,869
AlP	2,45 (i)	—	—	9,8	5,464
AlSb	1,61 (i)	900	400	12,0	6,136
GaSb	0,726 (d)	2500	680	15,7	6,096
InSb	0,17 (d)	77000	850	16,8	6,479
ZnS	3,68 (d)	165	10–15	8,1	5,410
ZnSe	2,70 (d)	600	30	9,1	5,668
ZnTe	2,28 (d)	330	100	9,7–10,1	6,104
CdTe	1,49 (d)	1050	60	10,4	6,481
Полупроводники с решеткой вюрцита					
GaN	3,39 (d)	440	200	(9,5, 10,4)	(3,189, 5,178)
AlN	6,2 (d)	300	14	8,5–9,1	(3,112, 4,982)
SiC (4H)	3,23 (i)	850	120	—	(3,073, 10,053)
SiC (6H)	2,86 (i)	400	100	(9,7, 10,0)	(3,081, 15,117)
CdS	2,55 (d)	330	50	(8,7, 9,5)	(4,137, 6,716)
CdSe	1,75 (d)	900	50	(9,3, 10,2)	(4,300, 7,011)
ZnO	3,36 (d)	150	—	(7,8, 8,8)	(3,250, 5,207)
Полупроводники с решеткой каменной соли					
PbS	0,41 (d)	600	700	170	5,936
PbSe	0,29 (d)	1000	1000	250	6,124
PbTe	0,32 (d)	1700	840	400	6,460

*Символ в скобках означает: d — прямозонный, i — непрямозонный полупроводник.

Список литературы

1. *Бонч-Бруевич В.Л., Калашников С.Г.* Физика полупроводников. Изд. 2-е. — М.: Наука, 1990. — 688 с.
2. *Пикус Г.Е.* Основы теории полупроводниковых приборов. — М.: Наука, 1965. — 448 с.
3. *Shockley W.* // Bell Syst. Techn. J. 1949. V. 28. P. 435.
4. *Shockley W., Read W.T.* // Phys. Rev. 1952. V. 87, P. 835.
5. *Hall R.N.* // Phys. Rev. 1952. V. 87. P. 387.
6. *Sah C.T., Noyce R.N., Shockley W.* // Proc. IRE. 1957. V. 45. P. 1228.
7. *Гаман В.И.* Физика полупроводниковых приборов. — Томск: Изд-во НТЛ, 2000. — 426 с.
8. *Адирович Э.И., Карагеоргий-Алкалаев П.М., Лейдерман А.Ю.* Токи двойной инжекции в полупроводниках. — М.: Сов. радио, 1978. — 320 с.
9. *Стафеев В.И.* // ЖТФ. 1958. Т. 28. С. 1631.
10. *Милнс А.* Примеси с глубокими уровнями в полупроводниках. Пер. с англ. — М.: Мир, 1977. — 564 с.
11. *Блихер А.* Физика силовых биполярных и полевых транзисторов. Пер. с англ. — Л.: Энергоатомиздат, 1986. — 248 с.
12. *Davies L.W.* // Nature. 1962. V. 194, № 4830. P. 762.
13. *Hall R.N.* // Proc. IRE. 1952. V. 40. 1512.
14. *Зи С.* Физика полупроводниковых приборов. В 2-х томах. 2-е изд. Пер. с англ. — М.: Мир, 1984.
15. *Pearsall T.P., Capasso F., Nahory R.E. et al.* // Solid-State Electron. 1978. V. 21. P. 297.
16. *Lee C.A., Logan R.A., Batdorf R.L. et al.* // Phys. Rev. A 1964. V. 134. P. 761.
17. *Shockley W.* // Czech. J. Phys. 1961. V. B11. P. 81.
18. *Baraff G.A.* // Phys. Rev. 1962. V. 128. P. 2507.
19. *Miller S.L.* // Phys. Rev. 1955. V. 99. P. 1234.
20. *McKay K.G.* // Phys. Rev. 1954. V. 94. P. 877.
21. *Miller S.L.* // Phys. Rev. 1957. V. 105. P. 1246.
22. *Chynoweth A.G., McKay K.G.* // Phys. Rev. 1956. V. 102. P. 369.
23. *Ландау Л.Д., Лифшиц Е.М.* Квантовая механика. Нерелятивистская теория. — М.: Наука, 1989. — 768 с.
24. *Chynoweth A.G., McKay K.G.* // Phys. Rev. 1957. V. 106. P. 418.
25. *Zener C.* // Proc. Roy. Soc. (London), A 1934. V. 145. P. 523.

26. *Хоровиц П., Хилл У.* Искусство схемотехники. В 3-х тт. Пер. с англ. — М.: Мир, 1993.
27. *Esaki L.* // Phys. Rev. 1958. V. 109. P. 603.
28. *Fowler R. H., Nordheim L.* // Proc. Roy. Soc. (London). A 1928. V. 119. P. 173.
29. *Houston W. V.* // Phys. Rev. 1940. V. 57. P. 184.
30. *Келдыш Л. В.* // ЖЭТФ. 1957. Т. 33. С. 994.
31. *Price P. J., Radcliffe J. M.* // IBM J. Res. Develop. 1959. V. 3. P. 364.
32. *Kane E. O.* // J. Phys. Chem. Solids. 1960. V. 12. P. 181.
33. *Kane E. O.* // J. Appl. Phys. 1961. V. 32. P. 83.
34. *Иванчик И. И.* // ФТТ. 1961. Т. 3. С. 103.
35. *Бонч-Бруевич В. Л., Серебренников П. С.* // Радиотехн. и электрон. 1961. Т. 6. С. 2041.
36. *Серебренников П. С.* // Радиотехн. и электрон. 1962. V. 7. P. 536.
37. *Бонч-Бруевич В. Л., Серебренников П. С.* // Радиотехн. и электрон. 1963. Т. 8. С. 1002.
38. Туннельные диоды и их применение в схемах переключения и в устройствах СВЧ диапазона / Пер. с англ. (под ред. А. А. Визеля). — М.: Сов. радио, 1965. — 184 с.
39. *Ансельм А. И.* Введение в теорию полупроводников. — М.: Наука, 1978. — 616 с.
40. Туннельные явления в твердых телах / (Сб. под ред. Э. Бурштейна и С. Лундквиста). Пер. с англ. — М.: Мир, 1973. — 421 с.
41. *Brockhouse B. N., Iyengar P. K.* // Phys. Rev. 1958. V. 111. P. 747.
42. *Holonyak N., Lesk I. A., Hall R. N. et al.* // Phys. Rev. Lett. 1959. V. 3. P. 167.
43. *Moodera J. S., Kinder L. R., Wong T. M., Meservey R.* // Phys. Rev. Lett. 1995. V. 74. P. 3273.
44. *Pearton S. J., Overberg M. E., Thaler G. T. et al.* // Phys. Stat. Solidi (a). 2003. V. 195. P. 222.
45. *Jansen R.* // J. Phys. D. 2003. V. 36. R289.
46. *Yajima T., Esaki L.* // J. Phys. Soc. Jap. 1958. V. 13. P. 1281.
47. *Chynoweth A. G., Feldmann W. L., Logan R. A.* // Phys. Rev. 1961. V. 121. P. 684.
48. Туннельные диоды (физические основы работы). — М.: Наука, 1966. — 142 с.
49. *Ван дер Зил А.* Шум. Источники, описание, измерение. Пер. с англ. — М.: Сов. радио, 1973. — 229 с.
50. *Burrus C. A.* // J. Appl. Phys. 1961, V. 32. P. 1031.
51. *Karlovský J.* // Solid-State Electron. 1967. V. 10. P. 1109.
52. Handbook on Semiconductors. Vol. 4. / Device Physics (ed. by C. Hilsum). — North-Holland Publishing Co., 1981. — 970 p.
53. *Braun F.* // Ann. Phys. Chem. 1874. V. 153. P. 556.
54. *Schottky W.* // Naturwiss. 1938. V. 26. P. 843.
55. *Mott N. F.* // Proc. Cambr. Philos. Soc. 1938. V. 34. P. 568.

56. *Родерик Э.Х.* Контакты металл–полупроводник. Пер. с англ. — М.: Радио и связь, 1982. — 208 с.
57. *Милнс А., Фойхт Д.* Гетеропереходы и переходы металл–полупроводник. Пер. с англ. — М.: Мир, 1975. — 432 с.
58. Технология СБИС. / Под ред. С. Зи. Пер. с англ. В 2-х кн. — М.: Мир, 1986.
59. *Bardeen J.* // Phys. Rev. 1947. V. 71. P. 717.
60. *Тамм И.* // ЖЭТФ. 1933. Т. 3. P. 34.
61. *Benzer S.* // Phys. Rev. 1947. V. 71. P. 141.
62. *Де Грааф Х.* Использование поликристаллического кремния в технологии изготовления интегральных схем. В сб.: Поликристаллические полупроводники. Физические свойства и применения / Под ред. Г. Харбеке. Пер. с англ. — М.: Мир, 1989. — 244 с.
63. *Shockley W.* // Phys. Rev. 1939. V. 56. P. 317.
64. *Bethe H.A.* // MIT Radiat. Lab. Report. 1942. 43-12.
65. *Давыдов Б.* // ЖЭТФ. 1939. V. 9. P. 451.
66. *Heine V.* // Phys. Rev. 1965. V. 138. A1689.
67. *Crowell C.R., Sze S.M.* // Solid-State Electron. 1966. V. 9. P. 1035.
68. *Padovani F.A., Stratton R.* // Solid-State Electron. 1966. V. 9. P. 695.
69. *Шур М.С.* Современные приборы на основе арсенида галлия. Пер. с англ. — М.: Мир, 1991. — 632 с.
70. *Stall R., Wood C.E..C., Board K., Eastman L.F.* // Electron. Lett. 1979. V. 15. P. 800.
71. *Shockley W.* — US Patent 2,569,347 (1951), приоритет от 26 июня 1948 г.
72. *Губанов А.И.* // ЖТФ. 1950. Т. 20. С. 1287; ЖТФ. 1951. Т. 21. С. 304; ЖТФ. 1952. Т. 22. С. 729.
73. *Иоффе А.В.* // ЖТФ. 1948. Т. 18. № 1 С. 498.
74. *Крёмер Г.* Критический обзор теории гетеропереходов. В сб.: Молекулярно-лучевая эпитаксия и гетероструктуры / Под ред. Л. Ченга и К. Плога. Пер. с англ. — М.: Мир, 1989. — 274 с.
75. *Harrison W.A., Kraut E.A., Waldrop J.R., Grant R.W.* // Phys. Rev. B. 1978. V. 18. P. 4402.
76. *Kroemer H.* // Proc. IRE. 1957. V. 45. P. 1535.
77. *Воробьев Л.Е., Ивченко Е.Л., Фирсов Д.А., Шалыгин В.А.* Оптические свойства наноструктур. — СПб.: Наука, 2001. — 188 с.
78. *Андронов А.А., Нефедов И.М., Соснин А.В.* // ФТП. 2003. Т. 37. С. 378.
79. *Келдыш Л.В.* // ФТТ. 1962. Т. 4. С. 2265.
80. *Шик А.Я.* // ФТП. 1974. Т. 8. С. 1841.
81. *Силин А.П.* // УФН. 1985. Т. 147. С. 485.
82. *Esaki L., Tsu R.* // IBM J. Res. Develop. 1970. V. 14. P. 61.
83. *Херман М.* — Полупроводниковые сверхрешетки. Пер. с англ. — М.: Мир, 1989. — 240 с.

84. *Ю П., Кардона М.* Основы физики полупроводников. — М.: Физматлит, 2002. — 560 с.
85. *Pfeiffer L., West K. W., Stormer H. L., Baldwin K. W.* // Appl. Phys. Lett. 1989. V. 55. P. 1888.
86. *Dingle R., Störmer H. L., Gossard A. C., Wiegmann W.* // Appl. Phys. Lett. 1978. V. 33. P. 665.
87. *Шур М. С.* Физика полупроводниковых приборов. В 2-х томах. Пер. с англ. — М.: Мир, 1992.
88. *Пожела Ю.* Физика быстродействующих транзисторов. — Вильнюс: Мокслас, 1989. — 264 с.
89. *Davis R. H., Hosack H. H.* // J. Appl. Phys. 1963. V. 34. Pt. 1. P. 864.
90. *Йогансен Л. В.* // ЖЭТФ. 1963. Т. 45. С. 207; ЖЭТФ. 1964. Т. 47. С. 270.
91. *Бом Д.* Квантовая теория. Пер. с англ. — М.: Физматлит, 1961. — 728 с.
92. *Tsu R., Esaki L.* // Appl. Phys. Lett. 1973. V. 22. P. 562.
93. *Sollner T. C. L. G., Goodhue W. D., Tannenwald P. E. et al.* // Appl. Phys. Lett. 1983. V. 43. P. 588.
94. *Brown E. R., Söderström J. R., Parker C. D. et al.* // Appl. Phys. Lett. 1991. V. 58. P. 2291.
95. *Broekaert T. P. E., Lee W., Fonstad C. G.* // Appl. Phys. Lett. 1988. V. 53. P. 1545.
96. *Brown E. R., Sollner T. C. L. G., Parker C. D. et al.* // Appl. Phys. Lett. 1989. V. 55. P. 1777.
97. *Chin R., Holonyak N., Stillman G. E. et al.* // Electron. Lett. 1980. V. 16. P. 467.
98. *Берман Л. С.* Введение в физику варикапов. — Л.: Наука, 1968. — 180 с.
99. *Chattopadhyay G., Schlecht E., Ward J. S. et al.* // IEEE Trans. Microwave Theory Techn. 2004. V. 52. P. 1538.
100. *Kollberg E., Rydberg A.* // Electron. Lett. 1989. V. 25. P. 1696.
101. *Mélique X., Maestrini A., Lheurette E. et al.* // IEEE–MTT Int. Microwave Symposium (Anaheim, 1999).
102. *Носов Ю. Р.* Физические основы работы полупроводникового диода в импульсном режиме. — М.: Наука, 1968. — 263 с.
103. *Еремин С. А., Мокеев О. К., Носов Ю. Р.* Полупроводниковые диоды с накоплением заряда и их применение. — М.: Сов. радио, 1966. — 152 с.
104. *Moll J. L., Krakauer S., Shen R.* // Proc. IRE. 1962. V. 50. P. 43.
105. *Рожков А. В., Козлов В. А.* // ФТП. 2003. Т. 37. С. 1477.
106. *Williams R.* // J. Appl. Phys. 1966. V. 37. P. 3411.
107. *Lang D. V.* // J. Appl. Phys. 1974. V. 45. P. 3023.
108. *Вавилов В. С., Кекелидзе Н. П., Смирнов Л. С.* Действие излучений на полупроводники. — М.: Наука, 1988. — 192 с.

109. *Bardeen J., Brattain W.H.* // Phys. Rev. 1948. V. 74. P. 230; *Brattain W.H., Bardeen J.* // Phys. Rev. 1948. V. 74. P. 231; *Bardeen J., Brattain W.H.* — US Patent 2524035 (1950), приоритет от 17 июня 1948 г.
110. *Shockley W., Sparks M., Teal G.K.* // Phys. Rev. 1951. V. 83. P. 151.
111. *Lanyon H.P.D., Tuft R.A.* // IEEE Trans. Electron Devices. 1979. V. ED-26. P. 1014.
112. *Cazarre A., Tasselli J., Marty A. et al.* // Electron. Lett. 1985. V. 21. P. 1124.
113. *Early J.M.* // Proc. IRE. 1952. V. 40. P. 1401.
114. *Schafft H.A.* // Proc. IEEE. 1967. V. 55. P. 1272.
115. *Daw A.N., Mitra R.N., Choudhury N.K.D.* // Solid-State Electron. 1967. V. 10. P. 359.
116. *Knepper R.W., Gaur S.P., Chang F.-Y., Srinivasan G.R.* // IBM J. Res. Develop. 1985. V. 29. P. 218.
117. *Kirk C.T.* // IRE Trans. Electron Devices. 1962. V. ED-9. P. 164.
118. *Hafez W., Lai J.-W., Feng M.* // IEEE Electron Device Lett. 2003. V. 24. P. 436.
119. *Hafez W., Feng M.* // Appl. Phys. Lett. 2005. V. 86. P. 152101.
120. *Hafez W., Lai J.-W., Feng M.* // IEEE Electron Device Lett. 2003. V. 24. P. 292.
121. *Dvorak M.W., Bolognesi C.R., Pitts O.J., Watkins S.P.* // IEEE Electron Device Lett. 2001. V. 22. P. 361.
122. *Lee Q.* — Ph.D. Thesis, Univ. of California, Santa Barbara, June 1999.
123. *Chen Y.K., Levi A.F.J., Nottenburg R.N. et al.* // Appl. Phys. Lett. 1989. V. 55. P. 1789.
124. *Martin D.D., Stratton R.* // Solid-State Electron. 1966. V. 9. P. 237.
125. *Rieh J.S., Jagannathan B., Chen H., Schonenberg K.T. et al.* // IEDM Tech. Digest. 2002. P. 771.
126. *Dunn J.S., Ahlgren D.C., DuCoolbaugh D. et al.* // IBM J. Res. Develop. 2003. V. 47. P. 101.
127. *Борблик В.Л., Грибников З.С.* // ФТП. Т. 22. С. 1537.
128. *Heiblum M., Fischetti M.V.* // IBM J. Res. Develop. 1990. V. 34. P. 530.
129. *Mead C.A.* // Proc. IRE. 1960. V. 48. P. 359.
130. *Heiblum M., Anderson I.M., Knoedler C.M.* // Appl. Phys. Lett. 1986. V. 49. P. 207.
131. *Титце У., Шенк К.* Полупроводниковая схемотехника. Пер. с нем. — М.: Мир, 1983. — 512 с.
132. *Васильев Д.В., Витоль М.Р., Горшенков Ю.Г. и др.* Радиотехнические цепи и сигналы. — М.: Радио и связь, 1982. — 528 с.
133. *Букингем М.* Шумы в электронных приборах и системах. Пер. с англ. — М.: Мир, 1986. — 398 с.

134. *Van Vliet K.M.* // Proc. IRE. 1958. V. 46. P. 1004.
135. *Ржанов А.В.* Электронные процессы на поверхности полупроводников. — М.: Наука, 1971. — 480 с.
136. Аналоговые и цифровые интегральные схемы / Под ред. С. В. Якубовского. — М.: Сов. радио, 1979. — 336 с.
137. *Kilby J.S.* — US Patent 3,138,743 (1964), приоритет от 6 февраля 1959 г.
138. *Noyce R.N.* — US Patent 2,981,877 (1961), приоритет от 30 июля 1959 г.
139. *Моро У.* Микролитография. В 2-х ч. Пер с англ. — М.: Мир, 1990.
140. *Shaw J.M., Gelorme J.D., LaBianca N.C. et al.* // IBM J. Res. Develop. 1997. V. 41. P. 81.
141. *Сугано Т., Икома Т., Такэиси Ё.* Введение в микроэлектронику. Пер. с яп. — М.: Мир, 1988. 320 с.
142. *Березин А.С., Мочалкина О.Р.* Технология и конструирование интегральных микросхем. — М.: Радио и связь, 1983. — 232 с.
143. *Пичугин И.Г., Таиров Ю.М.* Технология полупроводниковых приборов. — М.: Высшая школа, 1984. — 288 с.
144. *Риссел Х., Руге И.* Ионная имплантация. Пер. с нем. М., Наука, 1983, 360 с.
145. *Ohl R.* — US Patent 2,750,541 (1956), приоритет от 31 января 1950 г.; *Ohl R.* // Bell Syst. Techn. J. 1952. V. 31. P. 104.
146. *Бредов М.М., Комарова Р.Ф., Регель А.Р.* // ДАН СССР. 1954. Т. 99. С 69.
147. *Cussins W.D.* // Proc. Phys. Soc. 1955. V. B68. P. 213.
148. *Shockley W.* — US Patent 2,787,564 (1957), приоритет от 28 октября 1954 г.
149. *Alväger T., Hansen N.J.* // Rev. Sci. Instrum. 1962. V. 33. P. 567.
150. *Гусев В.М., Титов В.В., Гусева М.И., Куринный В.И.* // ФТТ. 1965. Т. 7. С. 2077.
151. *King W.J., Burrill J.T., Harrison S. et al.* // Nucl. Instr. Methods. 1965. V. 38. P. 178.
152. *Двуреченский А.В., Качурин Г.А., Нидаев Е.В., Смирнов Л.С.* Импульсный отжиг полупроводниковых материалов. — М.: Наука, 1982. — 208 с.
153. *Conrad J.R., Radtke J.L., Dodd R.A. et al.* // J. Appl. Phys. 1987. V. 62. P. 4591.
154. *Chu P.K.* // Plasma Phys. Control. Fusion. 2003. V. 45. P. 555.
155. *Случинская И.А.* Основы материаловедения и технологии полупроводников. — М.: МИФИ, 2002. — 378 с.
156. *Ферри Д., Эйкерс Л., Гринич Э.* Электроника ультрабольших интегральных схем. Пер. с англ. — М.: Мир, 1991. — 328 с.
157. *Шагурин И.И., Петросянц К.О.* Проектирование цифровых микросхем на элементах инжекционной логики. — М.: Радио и связь, 1984. — 231 с.

158. *Washio K., Kondo M., Ohue E. et al.* // IEDM Tech. Digest. 1999. P. 557.
159. *Moll J.L., Tanenbaum M., Goldey J.M., Holonyak N.* // Proc. IRE. 1956. V. 44. P. 1174.
160. *Блихер А.* Физика тиристоров. Пер. с англ. — Л.: Энергоиздат, 1981. — 264 с.
161. *Евсеев Ю.А.* Полупроводниковые приборы для мощных высоковольтных преобразовательных устройств. М.: Энергия, 1978. — 192 с.
162. *Герлах В.* Тиристоры. Пер. с нем. М.: Энергоатомиздат, 1985. — 327 с.
163. *Воронин П.А.* Силовые полупроводниковые ключи. Семейства, характеристики, применение. — М.: Додэка-XXI, 2001. — 384 с.
164. *Aldrich R.W., Holonyak N.* // J. Appl. Phys. 1959. V. 30. P. 1819.
165. Легирование полупроводников методом ядерных реакций / Под ред. Л.С. Смирнова. — Новосибирск: Наука, 1981. — 182 с.
166. *Cleland J.W., Lark-Horovitz K., Pigg J.C.* // Phys. Rev. 1950. V. 78. P. 814.
167. *Gentry F.E., Scace R.I., Flowers J.K.* // Proc. IEEE. 1965. V. 53. P. 355.
168. *Lilienfeld J.E.* — US Patent 1,745,175 (1930), приоритет от 8 октября 1926 г.; *Lilienfeld J.E.* — US Patent 1,900,018 (1933), приоритет от 28 марта 1928 г.
169. *Shockley W., Pearson G.L.* // Phys. Rev. 1948. V. 74. P. 232.
170. *Kahng D., Atalla M.M.* // Solid-State Device Research Conf., Pittsburg, June 1960; *Kahng D.* — US Patent 3,102,230 (1963), приоритет от 31 мая 1960 г.; *Atalla M.M.* — US Patent 3,056,888 (1962), приоритет от 17 августа 1960 г.
171. *Dacey G.C., Ross I.M.* // Proc. IRE. 1953. V. 41. P. 970.
172. *Mead C.A.* // Proc. IEEE. 1966. V. 54. P. 307.
173. *Lee K., Choi J.-S., Sim S.-P., Kim C.-K.* // IEEE Trans. Electron Devices. 1991. V. ED-38. P. 1905.
174. *Thompson S., Packan P., Bohr M.* // Intel Technology J. 1998. № 3. P. 1.
175. *Buchanan D.A.* // IBM J. Res. Develop. 1999. V. 43. P. 245.
176. *Lucovsky G.* // IBM J. Res. Develop. 1999. V. 43. P. 301.
177. *Gusev E.P., Lu H.-C., ,Garfunkel E.L. et al.* // IBM J. Res. Develop. 1999. V. 43. P. 265.
178. *Kang L., Lee B.H., Qi W.-J.. et al.* // IEEE Electron Device Lett. 2000. V. 21. P. 181.
179. *Doyle B., Arghavani R., Barlage D. et al.* // Intel Technology J. 2002. V. 6, № 2. P. 42.
180. *Thompson S., Alavi M., Hussein M. et al.* // Intel Technology J. 2002. V. 6, № 2. P. 5.
181. *Taur Y.* // IBM J. Res. Develop. 2002. V. 46. P. 213.

182. *Ruch J.G.* // IEEE Trans. Electron Devices. 1972. V. ED-19. P. 652.
183. *Abstreiter G., Brugger H., Wolf T. et al.* // Phys. Rev. Lett. 1985. V. 54. P. 2441.
184. *Ismail K., Nelson S.F., Chu J.O., Meyerson B.S.* // Appl. Phys. Lett. 1993. V. 63. P. 660.
185. *Rim K., Hoyt J.L., Gibbons J.F.* // IEEE Trans. Electron Devices. 2000. V. ED-47. P. 1406.
186. *Lochtefeld A., Djomehri I.J., Samudra G., Antoniadis D.A.* // IBM J. Res. Develop. 2002. V. 46. P. 347.
187. *Dennard R.H.* — US Patent 3,387,286 (1968), приоритет от 14 июля 1967 г.
188. *Adler E., DeBrosse J.K., Geissler S.F. et al.* // IBM J. Res. Develop. 1995. V. 39. P. 167.
189. *Mandelman J.A., Dennard R.H., Bronner G.B. et al.* // IBM J. Res. Develop. 2002. V. 46. P. 187.
190. *Wanlass F.M.* — US Patent 3,356,858 (1967), приоритет от 18 июня 1963 г.
191. *Izumi K., Doken M., Ariyoshi H.* // Electron. Lett. 1978. V. 14. P. 593.
192. *Shahidi G.G.* // IBM J. Res. Develop. 2002. V. 46. P. 121.
193. *Ning T.H.* // IBM J. Res. Develop. 2002. V. 46. P. 181.
194. *Kahng D., Sze S.M.* // Bell Syst. Techn. J. 1967. V. 46. P. 1288; *Kahng D.* — US Patent 3,500,142 (1970), приоритет от 5 июня 1967 г.
195. *Frohman-Bentchkowsky D.* // Appl. Phys. Lett. 1971. V. 18. P. 332; Solid-State Electron. 1974. V. 17. P. 517.
196. *Bez R., Camerlenghi E., Modelli A., Visconti A.* // Proc. IEEE. 2003. V. 91. P. 489.
197. *Masuoka F., Asano M., Iwahashi H., Komuro T.* // IEDM Techn. Digest. 1984. P. 464.
198. *de Blauwe J.* // IEEE Trans. Nanotechnol. 2002. V. 1. P. 72.
199. *Tiwari S., Rana F., Hanafi H. et al.* // Appl. Phys. Lett. 1996. V. 68. P. 1377.
200. *Dimitrakis P., Kapetanakis E., Tsoukalas D. et al.* // Solid-State Electron. 2004. V. 48. P. 1511.
201. *Banks G.J.* // US Patent 5,218,569 (1993), приоритет от 8 февраля 1991 г.
202. *R. Wegener H.A., Lincoln A.J., Pao H.C. et al.* // IEDM-1967, Abstracts, p. 58; *R. Wegener H.A.* — US Patent 3,590,337 (1971), приоритет от 14 октября 1968 г.; *Frohman-Bentchkowsky D.* // Proc. IEEE. 1970. V. 58. P. 1207.
203. *Kahng D., Sundburg W.J., Boulin D.M., Ligenza J.R.* // Bell Syst. Techn. J. 1974. V. 53. P. 1723.
204. *Shockley W.* // Proc. IRE. 1952. V. 40. P. 1365.
205. *Hooper W.W., Lehrer W.I.* // Proc. IEEE. 1967. V. 55. P. 1237.

206. *Mimura T., Hiyamizu S., Fujii T., Nanbu K.* // Jap. J. Appl. Phys. 1980. V. 19. L225.
207. *Shealy J.R., Kaper V., Tilak V. et al.* // J. Phys.: Condens. Matter. 2002. V. 14. P. 3499.
208. *Yamashita Y., Endoh A., Shinohara K. et al.* // IEEE Electron Device Letters. 2002. V. 23. P. 573.
209. *Zeng A., Jackson M.K., van Hove M., De Raedt W.* // Appl. Phys. Lett. 1995. V. 67. P. 262.
210. *Zuleeg R., Hinkle V.O.* // Proc. IEEE. 1964. V. 52. 1245; *Teszner S., Gicquel R.* // Proc. IEEE. 1964. V. 52. 1502; *Nishizawa J., Terasaki T., Shibata J.* // IEEE Trans. Electron Devices. 1975. V. ED-22. P. 185.
211. *Bozler C.O., Alley G.D.* // Proc. IEEE. 1982. V. 70. P. 46.
212. *Секен К., Томпсет М.* Приборы с переносом заряда. Пер. с англ. — М.: Мир, 1978. — 327 с.
213. *Носов Ю.Р., Шилин В.А.* Полупроводниковые приборы с зарядовой связью. М., Сов. радио, 1976, 144 с; *Носов Ю.Р., Шилин В.А.* Основы физики работы приборов с зарядовой связью. — М.: Наука, 1986. — 318 с.
214. *Boyle W.S., Smith G.E.* // Bell Syst. Techn. J. 1970. V. 49. P. 587.
215. *Boyle W.S., Smith G.E.* — US Patent 3,792,322 (1974), приоритет от 19 апреля 1973 г.
216. *Барб Д.Ф., Кэмпана С.* — В сб.: Достижения в технике передачи и воспроизведения изображений. Т. 3. Под ред. Б. Кейзана. Пер. с англ. — М.: Мир, 1980. — С. 180–305.
217. *Пресс Ф.П.* Фоточувствительные приборы с зарядовой связью. — М.: Радио и связь, 1991. — 264 с.
218. *Левинштейн М.Е., Пожела Ю.К., Шур М.С.* Эффект Ганна. — М.: Сов. радио, 1975. — 288 с.
219. *Ridley B.K., Watkins T.B.* // Proc. Phys. Soc. 1961. V. 78. P. 293.
220. *Hilsum C.* // Proc. IRE. 1962. V. 50. P. 185.
221. *Gunn J.B.* // Solid State Commun. 1963. V. 1. P. 88; *Gunn J.B.* // IBM J. Res. Develop. 1964. V. 8. P. 141; *Gunn J.B.* — US Patent 3,365,583 (1968), приоритет от 12 июня 1964 г.
222. *Kroemer H.* // Proc. IEEE. 1964. V. 52. P. 1736.
223. *Alekseev E., Pavlidis D.* // Solid-State Electron. 2000. V. 44. P. 941.
224. *Hauser J.R., Glisson T.H., Littlejohn M.A.* // Solid-State Electron. 1979. V. 22. P. 487.
225. *Бонч-Бруевич В.Л., Звягин И.П., Миронов А.Г.* Доменная электрическая неустойчивость в полупроводниках. — М.: Наука, 1972. — 414 с.
226. *Copeland J.A.* // Proc. IEEE. 1966. V. 54. P. 1479; *Copeland J.A.* // J. Appl. Phys. 1967. V. 38. P. 3096.
227. *Eisele H.* // Electron. Lett. 1998. V. 34. P. 2412.
228. *Llewellyn F.B., Bowen A.E.* // Bell Syst. Techn. J. 1939. V. 18. P. 280.

229. *Shockley W.* // Bell Syst. Techn. J. 1954. V. 33. P. 799; *Shockley W.* — US Patent 2,794,917 (1957), приоритет от 27 января 1953 г.
230. *Read W. T.* — US Patent 2,899,652 (1959), приоритет от 24 июня 1954 г.; *Read W. T.* // Bell Syst. Techn. J. 1958. V. 37. P. 401.
231. *Тагер А. С., Вальд-Перлов В. М.* Лавинно-пролетные диоды и их применение в технике СВЧ. — М.: Сов. радио, 1968. — 480 с.
232. *Тагер А. С., Мельников А. И., Кобельков Г. П., Цебиев А. М.* — Диплом на открытие № 24, приоритет от 27 октября 1959 г.
233. *Gilden M., Hines M. E.* // IEEE Trans. Electron Devices. 1966. V. ED-13. P. 169.
234. *Misawa T.* // IEEE Trans. Electron Devices. 1966. V. ED-13. P. 137.
235. *Misawa T.* // IEEE Trans. Electron Devices. 1967. V. ED-14. P. 795.
236. *Mishra J. K., Panda A. K., Dash G. N.* // IEEE Trans. Electron Devices. 1997. V. ED-44. P. 2143.
237. *Scharfetter D. L., Gummel H. K.* // IEEE Trans. Electron Devices. 1969. V. ED-16. P. 64.
238. *Seidel T. E.* — US Patent 3,466,512 (1969), приоритет от 29 мая 1967 г.
239. *Шухостанов А. К.* Лавинно-пролетные диоды. Физика, технология, применение. — М.: Радио и связь, 1997. — 208 с.
240. *Vassileuski K. V., Zorenko A. V., Zekentes K.* // Electron. Lett. 2001. V. 37. P. 466.
241. *Yuan L., Cooper J. A., Melloch M. R., Webb K. J.* // IEEE Electron Device Lett. 2001. V. 22. P. 266.
242. *Culshaw B.* // Electron. Lett., V. 10. P. 143 (1974).
243. *Ino M., Ishibashi T., Ohmori M.* // Electron. Lett. 1976. V. 12. P. 148.
244. *Rüegg H. W.* // IEEE Trans. Electron Devices. 1968. V. ED-15. P. 577.
245. *Wright G. T.* // Electron. Lett. 1968. V. 4. P. 543.
246. *Coleman D. J., Sze S. M.* // Bell Syst. Techn. J. 1971. V. 50. P. 1695.
247. *Sze S. M., Coleman D. J., Loya A.* // Solid-State Electron. 1971. V. 14. P. 1209.
248. *Chu J. L., Persky G., Sze S. M.* // J. Appl. Phys. 1972. V. 43. P. 3510.
249. *Weller K. P.* // RCA Review. 1971. V. 32. P. 372.
250. *Sjölund A.* // Solid-State Electron. 1973. V. 16. P. 559.
251. *Prager H. J., Chang K. K. N., Weisbrod S.* // Proc. IEEE. 1967. V. 55. P. 586.
252. *Кэррол Дж.* СВЧ-генераторы на горячих электронах. — М.: Мир, 1972. — 382 с.
253. *Аладинский В. К.* // ФТП. 1968. Т. 2. С. 617.
254. *Nishizawa J., Okabe T., Okamoto K.* — US Patent 3,602,840 (1971), приоритет от 31 июля 1969 г.

255. *Nishizawa J., Motoya K.* — US Patent 4,745,374 (1988), приоритет от 17 июня 1986 г.

256. *Gribnikov Z.S., Vagidov N.Z., Mitin V.V., Haddad G.I.* // J. Appl. Phys. 2003. V. 93. P. 5435.

257., *HvEisele. Haddad G.I.* // IEEE Trans. Microwave Theory Techn. 1995. V. 43. P. 210.

258. *Plotka P., Nishizawa J., Kurabayashi T., Makabe H.* // IEEE Trans. Electron Devices. 2003. V. 50. P. 867.

259. *Kesan V.P., Neikirk D.P., Blakey P.A. et al.* // IEEE Trans. Electron Devices. 1988. V. 35. P. 405.

260. *Javalagi S., Reddy V., Gullapalli K., Neikirk D.* // Electron. Lett. 1992. V. 28. P. 1699.

261. *Мосс Т., Баррел Г., Эллис Б.* Полупроводниковая оптоэлектроника. Пер. с англ. — М.: Мир, 1976. — 432 с.

262. *Хадсон Р.* Инфракрасные системы. Пер. с англ. — М.: Мир, 1972. — 536 с.

263. *Аут И., Генцов Д., Герман К.* Фотоэлектрические явления. Пер. с нем. — М.: Мир, 1980. — 208 с.

264. Narrow-Gap Semiconductors. Springer Tracts in Modern Physics. Vol. 98. Contributiobs by R. Dornhaus, G. Nimtz, B. Schlicht. — Springer Verlag, Berlin–Heidelberg–New York–Tokyo, 1983. — 309 p.

265. *Kazanskii A.G., Richards P.L., Haller E.E.* // Appl. Phys. Lett. 1977. V. 31. P. 496.

266. *Petroff M.D., Stapelbroek M.G.* — US Patent 4,568,960 (1986), приоритет от 23 октября 1980 г.

267. *Watson D.M., Guptill M.T., Huffman J.E. et al.* // J. Appl. Phys. 1993. V. 74. P. 4199.

268. *Cardozo B.L., Haller E.E., Reichertz L.A., Beeman J.W.* // Appl. Phys. Lett. 2003. V. 83. P. 3990.

269. *Haller E.E., Beeman J.W.* // Proc. FIR, Submm, and mm Detector Technology Workshop (Monterey, Apr. 2002), NASA/CP-211408. 2003. paper 2-06.

270. *Petroff M.D., Stapelbroek M.G., Kleinhans W.A.* // Appl. Phys. Lett. 1987. V. 51. P. 406.

271. *Levine B.F.* // J. Appl. Phys. 1993 V. 74. № 8, R1.

272. *Coon D.D., Karunasiri R.P.G.* // Appl. Phys. Lett. 1984. V. 45. P. 649.

273. *Levine B.F., Bethea C.G., Choi K.K. at al.* // Appl. Phys. Lett. 1988. V. 53. P. 231.

274. *Levine B.F., Bethea C.G., Hasnain G. et al.* // Appl. Phys. Lett. 1988. V. 53. P. 296.

275. *Gunapala S.D., Levine B.F., Chand N.* // J. Appl. Phys. 1991. V. 70. P. 305.

276. *Воробьев Л.Е., Данилов С.Н., Зегря Г.Г. и др.* Фотоэлектрические явления в полупроводниках и размерно-квантовых структурах. – СПб.: Наука, 2001. — 248 с.
277. *Berryman K.W., Lyon S.A., Segev M.* // Appl. Phys. Lett. 1997. V. 70. P. 1861.
278. *Liu H.C.* // Opto-Electron. Rev. 2003. V. 11. № 1.
279. *Ryzhii V.* // Semicond. Sci. Technol. 1996. V. 11. P. 759.
280. *Bandara S.V., Gunapala S.D., Liu J.K. et al.* // Appl. Phys. Lett. 1998. V. 72. P. 2427.
281. *Mii Y.J., Karunasiri R.P.G., Wang K.L. et al.* // Appl. Phys. Lett. 1990. V. 56. P. 1986.
282. *Perera A.G.U., Shen W.Z., Matsik S.G. et al.* // Appl. Phys. Lett. 1998. V. 72. P. 1596.
283. *Perera A.G.U., Matsik S.G., Liu H.C. et al.* // Appl. Phys. Lett. 2000. V. 77. P. 741.
284. *Ehret S., Schneider H., Fleissner J., Koidl P.* // Appl. Phys. Lett. 1997. V. 71. P. 641.
285. *Perera A.G.U., Yuan H.X., Francombe M.H.* // J. Appl. Phys. 1995. V. 77. P. 915.
286. *Perera A.G.U., Matsik S.G., Yaldiz B. et al.* // Appl. Phys. Lett. 2001. V. 78. P. 2241.
287. *Лайнс М., Гласс А.* Сегнетоэлектрики и родственные им материалы. Пер. с англ. — М.: Мир, 1981. — 736 с.
288. *Kato K.* // IEEE Trans. Microwave Theory Techn. 1999. V. 47. P. 1265.
289. *Campbell J.C., Demiguel S., Ma F. et al.* // IEEE J. on Selected Topics in Quantum Electronics. 2004. V. 10. P. 777.
290. *Васильев А.М., Ландсман А.П.* Полупроводниковые фотопреобразователи. — М.: Сов. радио, 1971. — 246 с.
291. *Chapin D.M., Fuller C.S., Pearson G.L.* // J. Appl. Phys. 1954. V. 25. P. 676.
292. *Reynolds D.C., Leies G., Antes L.L., Marburger R.E.* // Phys. Rev. 1954. V. 96. P. 533.
293. *Алфёров Ж.И., Андреев В.М., Румянцев В.Д.* // ФТП. 2004. Т. 38. С. 937.
294. *Green M.A., Emery K., King D.L. et al.* // Progr. in Photovoltaics: Res. and Appl. 2003. V. 11. P. 347.
295. *Pietruszko S.M.* // Opto-Electron. Rev. 2004. V. 12, № 7.
296. *Вавилов В.С., Крюкова И.В., Чукичев М.В.* // ФТТ. 1964. Т. 6. С. 2634.
297. *Barnham K., Marques J.L., Hassard J., O'Brien P.* // Appl. Phys. Lett. 2000. V. 76. P. 1197.
298. *M. van Sark W.G.J.H., Meijerink A., Schropp R.E.I. et al.* // ФТП. 2004. Т. 38. С. 1000.
299. *Мотт Н., Дэвис Э.* Электронные процессы в некристаллических веществах. Изд. 2-е. В 2-х тт. Пер. с англ. — М.: Мир, 1982.

300. *Pell E.M.* // J. Appl. Phys. 1960. V. 31. P. 291.
301. *Акимов Ю.К., Игнатьев О.В., Калинин А.И., Кушнирук В.Ф.* Полупроводниковые детекторы в экспериментальной физике. — М.: Энергоатомиздат, 1989. — 344 с.
302. *Панков Ж.* Оптические процессы в полупроводниках. Пер. с англ. — М.: Мир, 1973. — 456 с.
303. *van Roosbroeck W., Shockley W.* // Phys. Rev. 1954. V. 94. P. 1558.
304. *Берг А., Дин П.* Светодиоды. Пер. с англ. — М.: Мир, 1979. — 688 с.
305. Data in Science and Technology. Semiconductors other than group IV elements and III-V compounds. — Springer Verlag, Berlin, 1992.
306. Landolt–Börnstein Numerical Data and Functional Relationships in Science and Technology. New Series. Group III. Vol. 22a. — Springer Verlag, Berlin, 1987.
307. *Streubel K., Linder N., Wirth R., Jaeger A.* // IEEE J. on Selected Topics in Quantum Electronics. 2002. V. 8. P. 321.
308. *Krames M.R., Ochiai-Holcomb M., Höfler G.E. et al.* // Appl. Phys. Lett. 1999. V. 75. P. 2365.
309. *Narukawa J., Narita J., Sakamoto T. et al.* // Jap. J. Appl. Phys. 2006. V. 45. L1084.
310. *Eason D.B., Yu Z., Hughes W.C. et al.* // Appl. Phys. Lett. 1995. V. 66. P. 115.
311. *Алфёров Ж.И., Андреев В.М., Гарбузов Д.З. и др.* // Письма в ЖТФ. 1977. Т. 3. С. 657.
312. *Mueller-Mach R., Mueller G.O., Krames M.R., Trottier T.* // IEEE J. on Selected Topics in Quantum Electronics. 2002. V. 8. P. 339.
313. *Tang C.W., VanSlyke S.A.* // Appl. Phys. Lett. 1987. V. 51. P. 913.
314. *Юрре Т.А., Рудая Л.И., Климова Н.В., Шаманин В.В.* // ФТП. 2003. Т. 37. С. 835.
315. *Adachi C., Baldo M.A., Thompson M.E., Forrest S.R.* // J. Appl. Phys. 2001. V. 90. P. 5048.
316. *Baldo M.A., Lamansky S., Burrows P.E. et al.* // Appl. Phys. Lett. 1999. V. 75. P. 4.
317. *Einstein A.* // Phys. Zs. 1917. V. 18. P. 121.
318. *Басов Н.Г., Вул Б.М., Попов Ю.М.* // ЖЭТФ. 1959. Т. 37. С. 587; *Басов Н.Г., Крохин О.Н., Попов Ю.М.* // УФН. 1960. Т. 72. С. 161; *Басов Н.Г., Крохин О.Н., Попов Ю.М.* // ЖЭТФ. 1961. Т. 40. С. 1879.
319. *Hall R.N., Fenner G.E., Kingsley J.D. et al.* // Phys. Rev. Lett. 1962. V. 9. P. 366.
320. *Nathan M.I., Dumke W.P., Burns G. et al.* // Appl. Phys. Lett. 1962. V. 1. P. 62.
321. *Quist T.M., Rediker R.H., Keyes R.J. et al.* // Appl. Phys. Lett. 1962. V. 1. P. 91.
322. *Елисеев П.Г.* // Квантов. электрон. 2002. Т. 32. С. 1085.

323. *Елисеев П.Г.* Введение в физику инжекционных лазеров. — М.: Наука, 1983. — 294 с.
324. *Lasher G., Stern F.* // Phys. Rev. 1964. V. A133. P. 553.
325. *Алфёров Ж.И., Казаринов Р.Ф.* — Авт. свид. № 181737 (1975), приоритет от 30 марта 1963 г.
326. *Kroemer H.* // Proc. IEEE. 1963. V. 51. P. 1782.
327. *Кейси Х., Паниш М.* Лазеры на гетероструктурах. В 2-х тт. Пер. с англ. — М.: Мир, 1981.
328. *Алфёров Ж.И., Андреев В.М., Гарбузов Д.З. и др.* // ФТП. 1970. Т. 4. С. 1826.
329. *Hayashi I., Panish M.B., Foy P.W., Sumski S.* // Appl. Phys. Lett. 1970. V. 17. P. 109.
330. *Lockwood H.F., Kressel H., Sommers H.S., Hawrylo F.Z.* // Appl. Phys. Lett. 1970. V. 17. P. 499; *Kressel H., Lockwood H.F., Hawrylo F.Z.* // Appl. Phys. Lett. 1971. V. 18. P. 43.
331. *Dupuis R.D., Dapkus P.D., Holonyak N. et al.* // Appl. Phys. Lett. 1978. V. 32. P. 295.
332. *Wang C.A., Walpole J.N., Missaggia L.J. et al.* // Appl. Phys. Lett. 1991. V. 58. P. 2208.
333. *Faist J., Capasso F., Sivco D.L. et al.* // Science. 1994. V. 264. P. 553.
334. *Soda H., Iga K., Kitahara C., Suematsu Y.* // Jap. J. Appl. Phys. 1979. V. 18. P. 2329.
335. *Yang G.M., MacDougal M.H., Dapkus P.D.* // Electron. Lett. 1995. V. 31. P. 886.
336. *Yang Y.J., Dziura T.G., Wang S.C.* // Appl. Phys. Lett. 1990. V. 56. P. 1839.
337. *Yan C., Ning Y., Qin L. et al.* // Electron. Lett. 2004. V. 40. P. 872.
338. *Асрян Л.В., Сурис Р.А.* // ФТП. 2004. Т. 38. № 3.
339. *Arakawa Y., Sakaki H.* // Appl. Phys. Lett. 1982. V. 40. P. 939.
340. *Kapon E., Hwang D.M., Bhat R.* // Phys. Rev. Lett. 1989. V. 63. P. 430.
341. *Chou S.T., Cheng K.Y., Chou L.J., Hsieh K.C.* // Appl. Phys. Lett. 1995. V. 66. P. 2220.
342. *Goldstein L., Glas F., Marzin J.Y. et al.* // Appl. Phys. Lett. 1985. V. 47. P. 1099.
343. *Leonard D., Krishnamurthy M., Reaves C.M. et al.* // Appl. Phys. Lett. 1993. V. 63. P. 3203.
344. *Shchukin V.A., Ledentsov N.N., Kop'ev P.S., Bimberg D.* // Phys. Rev. Lett. 1995. V. 75. P. 2968.
345. *Леденцов Н.Н., Устинов В.М., Щукин В.А. и др.* // ФТП. 1998. Т. 32. С. 385.
346. *Леденцов Н.Н., Устинов В.М., Егоров А.Ю. и др.* // ФТП. 1994. Т. 28. С. 1483.

347. *Kirstaedter N., Ledentsov N.N., Grundmann M. et al.* // Electron. Lett. 1994. V. 30. P. 1416.
348. *Алфёров Ж.И., Бимберг Д., Егоров А.Ю. и др.* // УФН. 1995. Т. 165. С. 224.
349. *Вандышев Ю.В., Днепровский В.С., Климов В.И., Окороков Д.К.* // Письма в ЖЭТФ. 1991. Т. 54. С. 441.
350. *Eliseev P.G., Li H., Stintz A. et al.* // Appl. Phys. Lett. 2000. V. 77. P. 262.
351. *Максимов М.В., Шерняков Ю.М., Крыжановская Н.В. и др.* // ФТП. 2004. Т. 38. С. 763.
352. *Chatten A.J., Barnham K.W.J., Buxton B.F. et al.* // ФТП. 2004. Т. 38. С. 949.
353. *Якимов А.И., Двуреченский А.В., Кириенко В.В. и др.* // ФТП. 2004. Т. 38. С. 1265.
354. *Устинов В.М.* // ФТП. 2004. Т. 38. С. 963.
355. *Wong H.-S.P.* // IBM J. Res. Develop. 2002. V. 46. P. 133.
356. *Osburn C.M., Kim I., Han S.K. et al.* // IBM J. Res. Develop. 2002. V. 46. P. 299.

Предметный указатель

μ-фотопроводимость, 383
h-параметры транзистора, 179–181
p–n-переход
— плавный, 17, 123
— резкий, 15, 122
— сверхрезкий, 125
BIB-фотоприемники, 394
DLTS-спектроскопия, 135–139
HIWIP-фотоприемники, 401
IGBT, 266
MODFET, 299
QWIP-фотоприемники, 397
Амбиполярная диффузия, 23
— коэффициент, 23, 36, 43
Амбиполярный дрейф, 23
База
— диода, 16, 33, 60
— транзистора, 141, 151, 154, 168, 176
Баллистический перенос, 174, 262, 378
Барьер Шоттки, 54, 83, 85, 86, 89, 95, 99, 123, 189, 240, 294, 296, 299, 300, 308, 318, 326, 368, 405, 408, 414, 424
— высота, 85–89, 93, 102, 300
— емкость, 123
— определение высоты, 92
Барьерная емкость, 106, 119, 121–124, 136, 167, 226, 408, 413
Биполярные транзисторы, 140–214
— с изолированным затвором (IGBT), 266
Болометры, 382
Быстродействие
— диодов, 132
— диодов Ганна, 345
— диодов Шоттки, 92, 97
— лавинно-пролетных диодов, 366
— лавинных фотодиодов, 410
— резонансно-туннельных диодов, 118
— светодиодов, 430
— тиристоров, 231
— транзисторов, 114, 145, 165, 171, 174, 188, 260, 261
— туннельно-пролетных диодов, 378
— туннельных диодов, 79, 118, 133
— фотодиодов, 408
— фоторезисторов, 391, 400
Варакторы, 125, 135
— HBV, 126
Варизонные структуры, 102, 107, 108, 170, 173, 264
Варикапы, 125, 350
Вольт-ваттовая чувствительность, 389, 398
Время жизни, 21, 22, 28, 39, 83, 129, 132, 135, 152, 183, 189, 222, 233, 382, 388, 390, 391, 398, 403, 420
— излучательное, 430
— определение, 128, 130
— фотопроводимости, 391
Время пролета, 38, 48, 167, 261, 390, 408
Время релаксации
— импульса, 21, 117, 118, 262
— энергии, 21, 262, 336, 345, 346, 366, 384

Встроенный заряд в окисле, 54, 228, 241, 246, 247, 269, 270, 272
Встроенный потенциал, 13, 89, 297
Вторичный пробой, 58, 163, 164
Вынужденное излучение, 446
Высота барьера Шоттки, 85–89, 93, 102, 300
Генерационно-рекомбинационный шум, 183, 402
Гетеропереходы, 83, 102–110, 126, 152, 153, 156, 170, 298, 300, 302, 361, 364, 401, 437, 440, 443, 449
Глубокие уровни, 28, 29, 36, 74, 132, 135, 189, 223, 314, 391
Горячие носители, 47, 120, 174, 176, 258, 285–288, 293, 336
Дебаевская длина экранирования, 16, 40, 152
Деградация
— ЛИЗМОП-структур, 287
— МОП-транзисторов, 248
— барьеров Шоттки, 88
— контактов, 103, 206
— светодиодов, 436
— солнечных элементов, 420, 422
— тиристоров, 235
— туннельных диодов, 78
— ячеек РПЗУ, 288
Детекторы ядерных излучений, 423–428
Динисторы, 216
Диоды
— QWITT, 379
— Ганна, см. Диоды Ганна
— Рида, 356
— инжекционно-пролетные, 368, 376
— лавинно-пролетные, 355, 368
— обращенные, 81–82
— опорные, 60
— резонансно-туннельные, 115–120, 379
— с барьером Шоттки, 83–99, 133, 189
— с накоплением заряда, 133–135
— с проколом, 368
— с тонкой базой, 25, 33, 99, 132
— светодиоды, 432–445
— со структурой p–i–n, 39–44, 133, 359, 376, 406, 422
— сплавные, 101
— точечные, 98
— туннельно-пролетные, 378
— туннельные, 62–81, 117, 133
— — получение, 76
— умножительные, 126
Диоды Ганна, 333–361
— гибридный режим, 350
— режим ОНОЗ, 352
— режим выделения гармоник, 353
— режим пролета домена, 349
— режим пролета обогащенного слоя, 346
— режим с гашением, 351
— режим с запаздыванием, 351
Дифференциальное сопротивление
— отрицательное, 64, 79, 112, 117, 118, 333, 355, 368
— стабилитрона, 61
Диффузионная длина, 19, 23, 25, 30, 43, 127, 142, 143, 150, 152, 218, 222, 386
— амбиполярная, 36, 41
Диффузионная емкость, 126
Диэлектрическая проницаемость, 14, 98, 105, 255, 278, 461
Длина свободного пробега, 47, 51, 57, 60, 90
Доменная неустойчивость, 341
Домены электрические, 341, 349
Дробовой шум, 182, 308, 330, 401
Емкостная спектроскопия, 123
Емкость

— барьера Шоттки, 123
— барьерная, 106, 119, 121–124, 136, 167, 226, 408, 413
— диффузионная, 126
Загонка примеси, 198, 204
Закон
— Мотта, 38
— сохранения импульса, 68, 428
Запаздывание
— лавинное, 355, 362
— пролетное, 355, 362
Затвор, 249
— плавающий, 285, 286, 289–292, 321, 330
Избыточные токи, 74–77, 117
Избыточный шум, 182, 309, 411
Излучательная рекомбинация, 428
— темп, 430
Изоляция
— окисная, 205, 207, 271, 272
— с помощью p–n-перехода, 207, 249
Изопланарный процесс, 208
Иммерсионная литография, 195
Инверсная населенность, 447
Инжекционно-пролетные диоды, 368, 376
Инжекционные лазеры, 445–460
— VCSEL, 455
— квантово-каскадные, 454
Инжекция
— в p–n-переходе, 20, 25, 449
— в гетеропереходе, 107, 153, 156, 170
— высокий уровень, 33, 39, 43, 222
Интегральные схемы
— n-МОП, 271–279
— И^{2}Л, 200, 210–214
— КМОП, 279–285
— ТТЛ, 189, 209, 214
— ТТЛШ, 190, 214
— ЭСЛ, 214
Ионная имплантация, 89, 101, 198, 205, 212, 248, 258, 270–272, 279, 280, 282, 292, 299, 319, 367
— метод PIII, 201
Ионная литография, 196
Исток, 249
КМОП-структуры, 279–285, 325
— эффект защелкивания, 281
Каналирование, 198
Квазиуровень Ферми, 21, 30, 139, 447
Квантовые нити, 112, 456–458
Квантовые точки, 112, 291, 400, 421, 442, 456, 457, 459, 460
Квантовые ямы, 110, 299, 300, 397, 437, 457
Квантовый выход
— люминесценции, 421, 432
— светодиода, 435–437, 439
— фотоионизации, 383, 385, 390
Коллектор, 141
Контакт
— металл–полупроводник, 84
— омический, 99–103
— точечный, 80, 98
Контактная разность потенциалов, 13, 18, 25, 36, 41, 104, 123, 162, 170, 304, 419
Коэффициент
— амбиполярной диффузии, 23, 36, 43, 386
— вывода света, 438, 439
— инжекции эмиттера, 150, 153, 155, 173, 221
— качества, 59, 368
— оптического ограничения, 448, 452, 458
— переноса через базу, 154, 221
— поглощения, 381, 385, 388, 392, 394, 395, 412, 430
— собирания носителей, 143, 147
— ударной ионизации, 47, 51, 120, 358, 361, 363, 410, 411
— умножения, 50, 220, 411

— усиления транзистора, 142, 144, 148, 151, 152, 154, 155, 158, 176, 177
— усиления фототока, 390
— шума, 186, 302, 348, 361, 376
Критерий Крёмера, 341, 346, 349
Крутизна вольт-амперной характеристики, 252, 254, 258
ЛИЗМОП-структуры, 248, 286
Лавинно-пролетные диоды, 355, 368
Лавинные фотодиоды, 53, 120, 410–414
Лавинный пробой, 45, 57, 160, 217, 230, 285, 355, 356, 364
Лазерный отжиг, 200
Лазеры, 446
Литография, 192
Люминесценция, 428
МНОП-структуры, 293
МОП-транзисторы, 240, 249–266, 308, 309, 320, 321
Максвелловское время релаксации, 38, 98, 133, 353
— дифференциальное, 340
Максимальная рабочая температура p–n-перехода, 25
Максимальная частота генерации, 118, 168, 171, 172
Междолинное рассеяние, 172, 175, 338, 345, 346
Междолинный перенос, 333, 341
Меза-структура, 148, 354, 367
Микроплазмы, 55, 62
Мини-зона, 113, 399
Модель
— Андерсона, 103
— термоэлектронной эмиссии, 90
— управления зарядом, 250
Напряжение
— Эрли, 160
— пороговое, 245, 248, 250, 256, 258, 269–271, 273, 286, 288
— пробоя, 44, 59, 160, 161, 217, 265
— прокола, 162, 217, 369
Напряженный кремний, 263
Непараболичность, 67, 68, 338
Обедненный слой, 11, 13, 29, 85
— толщина, 15, 17, 18, 55, 89, 121, 136, 162, 167, 244, 294, 295, 297, 369, 424
Область безопасной работы, 164
Область пространственного заряда, см. Обедненный слой
Обнаружительная способность, 403
Обращенные диоды, 81–82
Оже-рекомбинация, 39, 152, 222, 391, 426
Окисная изоляция, см. Изоляция окисная
Омические контакты, 99–103
— гетеропереходные, 101
Оптотиристоры, 235, 432
Оптроны, 410, 432
Охранные кольца, 53, 99, 230
ПЗС, 311–331, 428
— с объемным каналом, 316–318
Плазменное легирование, 201
Плазменное травление, 202–203, 205, 209, 266
Плазмоны, 175
Планарная технология, 147, 191–209, 271, 456
Поверхностная рекомбинация, см. Рекомбинация поверхностная
Поверхностные состояния, 86, 99, 183, 239, 248, 309, 314–316
Поверхностный пробой, 54, 228
Поглощение света
— в полосе остаточных лучей, 384
— коэффициент, см. Коэффициент поглощения
— на свободных носителях, 383, 401
— примесное, 383
— собственное, 382

Полевая эмиссия, 96, 100
Полевые транзисторы, 239–310
— мощные, 264–266, 296, 302, 305–308
— псевдоморфные (pHEMT), 301
— с p–n-переходом, 240, 294, 303–308
— с МНОП-структурой, 293
— с МОП-структурой, 240, 249–266
— с барьером Шоттки, 240, 294–298
— с высокой подвижностью электронов (HEMT), 298–303
— с изолированным затвором, 241, 249–266
— с плавающим затвором, 285
— со статической индукцией (SIT), 305
Полупроводник
— a-Si:H, 422
— $AgGaS_2$, 433
— $AgGaSe_2$, 433
— $AgInS_2$, 433
— AlAs, 88, 110, 118, 119, 380, 418, 433, 454, 462
— AlGaAs, 107, 108, 115, 120, 126, 153, 171, 299, 300, 334, 397, 398, 400, 401, 421, 433, 434, 436, 440, 451, 454, 455, 458
— AlGaN, 302, 409, 436
— AlInAs, 110, 171, 301, 302, 400, 454
— AlInGaP, 436
— AlInP, 400, 436, 454
— AlN, 462
— AlP, 433, 454, 462
— AlSb, 88, 116, 119, 171, 302, 433, 454, 462
— BP, 433
— C, 354, 462
— $CdGeP_2$, 433
— CdP_2, 433
— CdS, 87, 390, 414, 418, 421, 433, 442, 462
— CdSe, 83, 421, 433, 460, 462
— $CdSiP_2$, 433
— CdTe, 334, 418, 422, 425, 433, 462
— Cu_2O, 83, 418, 433
— $CuAlS_2$, 433
— $CuAlSe_2$, 433
— CuBr, 433
— CuCl, 433
— $CuGaS_2$, 433
— $CuGaSe_2$, 433
— $CuInS_2$, 433
— $CuInSe_2$, 422
— GaAlAsSb, 437
— GaAlSb, 454
— GaAs, 25, 32, 44–46, 48, 57, 69, 70, 77–80, 86–88, 96, 98, 100, 102, 104, 106–108, 110, 114, 115, 119, 120, 126, 139, 153, 171, 175, 192, 199, 214, 234, 235, 240, 262, 295, 299, 300, 302, 308, 309, 325, 327, 334, 335, 337, 338, 340, 341, 343, 346, 349, 361, 364–367, 379, 396–398, 400, 401, 409, 414, 416, 418, 419, 421, 425, 428, 431, 433, 437, 438, 440, 446, 448, 449, 451, 454, 455, 458, 460, 462
— GaAsP, 334, 433–435, 438, 454
— GaAsSb, 48, 110, 171, 334, 454
— GaInSb, 334, 454
— GaMnAs, 74
— GaMnN, 74
— GaMnP, 74
— GaN, 102, 107, 302, 308, 334, 346, 409, 433, 436, 441, 454, 462
— GaNAs, 454
— GaP, 45, 48, 86–88, 409, 418, 428, 433–435, 438, 454, 462
— GaS, 433

— GaSb, 77, 88, 110, 334, 437, 454, 462
— GaSe, 433
— Ge, 25, 27, 39, 44, 45, 48, 51, 58, 64, 70, 72, 76–81, 86, 88, 101, 102, 104, 106, 107, 110, 141, 143, 145, 146, 191, 197, 230, 327, 334, 355, 364, 393–395, 409, 414, 418, 423–428, 460, 462
— HgCdTe, 327, 391, 393, 409, 414
— HgI_2, 425
— InAs, 87, 110, 116, 118, 119, 171, 176, 262, 302, 327, 334, 393, 396, 414, 437, 454, 460, 462
— InAsP, 334, 454
— InGaAs, 48, 102, 110, 118, 153, 171, 172, 176, 262, 301, 302, 334, 361, 380, 400, 408, 409, 418, 437, 454
— InGaAsP, 110, 400, 409, 437, 447, 454
— InGaN, 436, 442
— InGaP, 110, 153, 171, 400, 418, 433, 436, 454
— InP, 48, 88, 102, 110, 118, 153, 171, 172, 192, 199, 262, 296, 301, 302, 334, 335, 338, 340, 346, 353, 361, 366, 400, 408, 418, 433, 447, 454, 462
— InSb, 69, 88, 327, 334, 393, 396, 405, 409, 414, 431, 454, 462
— PbS, 83, 327, 393, 462
— PbSe, 393, 403, 462
— PbSnTe, 110, 327, 393
— PbSnTeSe, 454
— PbSSe, 437
— PbTe, 327, 462
— Se, 83, 414
— Si, 25, 32, 40, 44, 45, 47, 48, 51, 54, 70, 75, 83, 85–89, 94, 98–102, 107, 110, 120, 126, 132, 138, 143, 146, 147, 152, 153, 155, 161, 167, 191, 197, 198, 203, 230, 231, 254, 262, 263, 269, 272, 282, 308, 309, 325, 326, 361, 365, 367, 376, 392–395, 409, 414, 416, 418, 419, 426–428, 460, 462
— SiC, 48, 88, 231, 296, 308, 365, 409, 428, 433, 436, 462
— SiGe, 120, 153, 173, 214, 263
— SiGeC, 110
— SnO_2, 324, 433
— ZnCdSe, 436
— $ZnGeP_2$, 433
— ZnO, 433, 462
— ZnP_2, 433
— ZnS, 433, 442, 462
— ZnSe, 110, 334, 433, 455, 460, 462
— $ZnSiAs_2$, 433
— $ZnSiP_2$, 433
— ZnTe, 433, 462
— ZnTeSe, 436
— вырожденный, 64, 96
— магнитный, 73
— непрямозонный, 70, 428, 430, 433, 434, 454, 462
— органический, 443
— прямозонный, 66, 69, 70, 412, 422, 428, 433, 436, 454, 462
Послеинжекционная ЭДС, 129
Потенциальный барьер, 11, 19, 55, 83, 84, 95, 104, 107, 112, 115
Предельная частота
— варикапа, 125
— полевого транзистора, 260
Приближение плавного канала, 250, 297
Приборы с зарядовой связью, см. ПЗС
Примесная зона, 71, 394
Пробой
— вторичный, 58, 163, 164
— лавинный, 45, 57, 160, 217, 230, 285, 355, 356, 364
— поверхностный, 54, 228

— примесный, 45, 394, 396
— тепловой, 57, 163
— туннельный, 55, 59, 65, 258, 366, 378
Прокол, 62, 160, 162, 217, 368
Пролетная частота, 348–351, 353, 377
Процесс SIMOX, 282
Псевдоморфные структуры, 109, 118, 173, 301
Работа выхода, 84, 85, 99, 104, 241
Радиационная стойкость, 139, 209, 282, 420, 421
Радиационные дефекты, 29, 74, 132, 139, 199, 269, 283, 420, 421, 426
Разгонка примеси, 198, 204
Режим
— TRAPATT, 376
— инверсии, 242
— насыщения, 188, 252
— обеднения, 242, 315
— обогащения, 242
— отсечки, 188
— плоских зон, 241
Резонансно-туннельные диоды, 115–120, 379
Резонансное туннелирование, 115
Рекомбинация
— Оже, 39, 152, 222, 391, 426
— безызлучательная, 428
— излучательная, 428
— поверхностная, 26, 184, 385, 388, 392, 407, 420
— темп, 22
— формула Шокли–Рида–Холла, 28
Рентгеновская литография, 195
СВЧ приборы, 332, 380
Самосовмещение, 207, 273
Сверхрешетки, 102, 110–121, 263, 397–399, 412, 454, 458
— *nipi*, 113
Светодиоды, 432–445
Селективное травление, 209, 265
Сечение захвата, 28, 137, 138, 435
Силициды, 85, 99, 153, 176
Симисторы, 236–238
Скорость
— насыщения, 167, 171, 174, 253, 261, 262, 295, 343, 349, 356, 365, 372, 378, 408
— поверхностной рекомбинации, 26, 101, 184, 388
Солнечные элементы, 414–422
— каскадные, 417
Соотношение Эйнштейна, 24, 43, 340
Сопротивление
— контактное, 100, 102
— растекания, 80, 98
Спектроскопия
— DLTS, 135–139
— емкостная, 123
— молекулярная, 332
— туннельная, 73
Спонтанное излучение, 445
Стабилитроны, 58–62, 368
Сток, 249
Структуры
— BiCMOS, 214, 284
— МНОП, 293
— кремний-на-изоляторе (SOI), 201
— кремний-на-изоляторе (КНИ), 282
— кремний-на-сапфире (КНС), 282
Схемы включения транзисторов, 177
Температура шума, 186, 309
Тепловой пробой, 57, 163
Тепловой шум, 182, 308, 401
Теплопроводность, 192, 231, 302, 364, 365
Термополевая эмиссия, 95, 97, 118, 398
Термоэлектронная эмиссия, 89, 175, 371, 444

Тиристоры, 39, 43, 215–238
— запираемые, 222, 224, 233, 267
— с закороченным катодом, 227
— симисторы, 236–238
— фототиристоры, 234
Ток
— генерационный, 29
— избыточный, 74–76, 117
— насыщения, 24, 32, 48, 92, 95, 152
— обратный коллектора, 160, 161, 178, 219, 220
— ограниченный пространственным зарядом, 38
— подпороговый, 256, 257, 277, 283
— рекомбинационный, 30, 31, 189
— смещения, 339, 357
— удержания, 221, 233
Транзисторы
— BSIT, 307
— DHBT, 171
— HBT, 170–177, 214
— HEMT, 119, 296, 298–303
— IGBT, 266
— RHET, 176
— SIT, 305
— баллистические, 174
— биполярные, 140–214
— гетеропереходные, 107, 108, 153, 170–177, 187, 214
— диффузионно-сплавные, 146
— дрейфовые, 145, 150, 154, 167
— многоколлекторные, 209
— многоэмиттерные, 209
— мощные, 147, 157, 158, 163
— на горячих электронах, 174
— плоскостные, 143, 145, 158
— полевые, см. Полевые транзисторы
— псевдоморфные (pHEMT), 301
— с меза-структурой, 148
— с металлической базой, 176
— с проницаемой базой, 307
— с туннельным эмиттером, 153
— со статической индукцией, 305
— сплавные, 101, 145
— точечные, 141, 198
— униполярные, 295
— фототранзисторы, 409
— эпитаксиальные, 147, 157, 170
Трансмутационное легирование, 230
Триаки, 236
Тринисторы, 216
Туннелирование, 17, 55, 63, 73, 81, 94, 95, 109, 184, 254, 285, 288, 366, 398, 444
— резонансное, 115
— с участием примесей, 72
— с участием фононов, 70, 117
Туннельная спектроскопия, 73
Туннельно-пролетные диоды, 378
Туннельное магнетосопротивление, 73
Туннельные диоды, 62–81, 117, 133
Туннельный пробой, 55, 59, 65, 258, 366, 378
Увлечение электронов фотонами, 405
Ударная ионизация, 45, 60, 62, 152, 220, 341, 355, 358, 361, 378, 397, 411, 449
— коэффициент, см. Коэффициент ударной ионизации
Уравнение
— Пуассона, 14, 18, 52, 53, 89, 105, 162, 229, 243, 339, 342, 396
— Шрёдингера, 111
— непрерывности, 22, 23, 37, 127, 134, 149, 342, 358, 385, 407
Уровни прилипания, 390, 391, 426
Уровни размерного квантования, 111, 120, 397

90, 94, 371
онейтральности, 22, 34,
0
27
ния, 416
ности, 31, 36, 94
4, 229
шум, 182, 248, 309, 401
мять, 195, 289–294
292
й спектр, 71
ческие, 71
ские, 60, 71, 167, 384
, 405–414
ые, см. Лавинные фо-
ы
рафия, 192–195, 271,
шум, 401, 403
мники
4–397
00
120, 397–401
димость, 382
т, 193, 196, 202
оры, 384–401
тивления, 384–401
торы, 234
исторы, 409
я, 92, 287, 326, 401
отности состояний,
сечки
ного транзистора,
, 177, 206
— генерационно-рекомбинаг
183, 402
— дробовой, 77, 182, 308, 3
401
— избыточный, 182, 309, 411
— лавинного умножения, 411
— переноса в ПЗС, 330
— поверхностный, 183, 330
— тепловой, 182, 308, 401
— фликкер-шум, 182, 248, 309,
401
— фотонный, 401, 403
Шум-фактор, 77, 186, 302, 411
Шумы
— в ПЗС, 320, 330
— в биполярных транзисторах,
182–187
— в гетеропереходных транзи-
сторах, 187
— в диодах Ганна, 348
— в инжекционно-пролетных ди-
одах, 375
— в лавинно-пролетных диодах,
361
— в полевых транзисторах, 308
— в стабилитронах, 62
— в туннельных диодах, 77
— в фотодиодах, 408
— в фотоприемниках, 401
ЭДС Дембера, 405
Эквивалентная мощность шума
(NEP), 402, 409
Эквивалентная схема
— МОП-конденсатора, 246
— полевого транзистора, 259
— тиристора, 218, 227
— транзистора, 180
— туннельного диода, 80
— шумов транзистора, 184

Электронное сродство, 84, 104, 106
Эмиссия
— полевая, 96, 100
— термополевая, 95, 97, 118, 398
— термоэлектронная, 89, 175, 371, 444
Эмиттер, 141
— коэффициент инжекции, 150, 153, 155, 173
Энергетическое разрешение, 425, 427
Эффект
— Ганна, 335
— Кирка, 157, 169
— Пула–Френкеля, 29
— Шоттки, 93, 370
— Эрли, 158, 181
— всплеска скорости, 262
— защелкивания, 268, 281, 282
— квантовомеханического отражения, 176
— кулоновской блокады, 291
— модуляции длины канала, 255
— модуляции проводимости, 33, 42, 43, 184
— насыщения скорости дрейфа, 167, 253
— оттеснения эмиттерного тока, 158
— поля, 239
— сужения запрещенной зоны, 151, 401
— туннельный, 62
— фотовольтаический, 405
— фотоэлектромагнитный, 405
Эффективность переноса в ПЗС, 314
Явление
— доменной неустойчивости, 341
— прилипания носителей, 390, 391
— самоорганизации, 458, 459
— сверхинжекции, 108

Учебное издание

ЛЕБЕДЕВ Александр Иванович

ФИЗИКА ПОЛУПРОВОДНИКОВЫХ ПРИБОРОВ

Редактор *М.Б. Козинцова*
Оригинал-макет: *Т.Н. Савицкая*
Оформление переплета: *Н.В. Гришина*

Подписано в печать 02.07.08. Формат 60×90/16. Бумага офсетная.
Печать офсетная. Усл. печ. л. 30,5. Уч.-изд. л. 33,5. Тираж 700 экз.
Заказ № 3607.

Издательская фирма «Физико-математическая литература»
МАИК «Наука/Интерпериодика»
117997, Москва, ул. Профсоюзная, 90
E-mail: fizmat@maik.ru, fmlsale@maik.ru;
http://www.fml.ru

Отпечатано с готовых диапозитивов
в ОАО «Ивановская областная типография»
153008, г. Иваново, ул. Типографская, 6
E-mail: 091-018@adminet.ivanovo.ru

ISBN 978-5-9221-0995-6

