

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Нижегородский государственный университет им. Н.И. Лобачевского
Национальный исследовательский университет

Е.В. Волкова
С.В. Оболенский

ПОЛУПРОВОДНИКОВЫЕ ДИОДЫ

Учебное пособие

Рекомендовано ученым советом радиофизического факультета
для студентов ННГУ, обучающихся по направлениям подготовки
011800 «Радиофизика», 010300 «Фундаментальная информатика
и информационные технологии»
и специальности 090302 «Информационная безопасность
телекоммуникационных систем»

Нижний Новгород
2013

УДК 53.082, 538.95

ББК 32.85

В67

В67 Волкова Е.В., Оболенский С.В. ПОЛУПРОВОДНИКОВЫЕ ДИОДЫ: Учебное пособие. – Нижний Новгород: Нижегородский госуниверситет, 2013. – 89 с.

Рецензенты: д.ф.-м.н. **В.И. Шашкин**,
к.т.н. **М.А. Китаев**

Данное пособие является продолжением цикла учебных пособий по полупроводниковой электронике и содержит информацию о физических основах работы полупроводниковых диодов.

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

Публикация учебного пособия поддержана грантом ФЦП "Научные и научно-педагогические кадры инновационной России".

Ответственные за выпуск:
председатель методической комиссии радиофизического факультета ННГУ,
к.ф.-м.н., доцент **Н.Д. Миловский**,
д.ф.-м.н., профессор **Е.З. Грибова**

УДК 53.082, 538.95

ББК 32.85

© Нижегородский государственный
университет им. Н.И. Лобачевского, 2013

СОДЕРЖАНИЕ

Предисловие	4
Часть 1. Физика явлений на границе раздела твёрдых тел. Выпрямляющие диоды	5
1.1. Классификация границ раздела между двумя полупроводниками	5
1.2. Изотипный (униполярный) гомопереход	6
1.3. Анизотипный (биполярный) гомопереход (p - n переход)	8
1.3.1. Электронно-дырочный переход в равновесном состоянии	8
1.3.2. Вольт-амперная характеристика p - n перехода	12
1.3.3. Емкость электронно-дырочного перехода	20
1.4. Анизотипный (биполярный) и изотипный (униполярный) гетеропереходы	23
1.5. Структура металл – диэлектрик – полупроводник (МДП)	31
1.5.1. Идеальная МДП-структура	31
1.5.2. Емкость МДП-структуры	34
1.6. Контакт металл – полупроводник	34
1.6.1. Зонная диаграмма	34
1.6.2. Теория процессов переноса зарядов	37
1.6.3. Омический контакт	42
1.7. Эквивалентные схемы диодов	43
Часть 2. Оптоэлектронные диоды	45
2.1. Фотодиоды	45
2.2. Солнечные батареи	52
2.3. Светодиоды и лазеры	55
Часть 3. СВЧ и КВЧ диоды	66
3.1. Диод Ганна	66
3.2. Лавинно-пролетные и инжекционно-пролетные диоды	74
3.2.1. Лавинно-пролетные диоды	74
3.2.2. Инжекционно-пролетные диоды	76
3.3. Диоды на основе эффекта туннелирования. Сверхрешетки	78
3.3.1. Туннельный диод	78
3.3.2. Туннельно-резонансный диод	83
3.3.3. Сверхрешётки	85
Список литературы	88

ПРЕДИСЛОВИЕ

Данная книга является второй из цикла учебных пособий, предназначенных для студентов 3 – 4-х курсов дневных и вечерних отделений высших технических учебных заведений, специализирующихся на изучении физики полупроводниковых приборов. В первой книге [1] авторами были изложены основные определения и понятия, касающиеся физики полупроводников, а также приведен расширенный объем сведений из области теории транспорта электронов в полупроводниковых материалах. Данное учебное пособие является логическим продолжением предыдущего и содержит сведения о физических процессах, протекающих вблизи контактов полупроводника с другими материалами, а также об основных физических принципах работы полупроводниковых диодов различных групп (выпрямительных, оптоэлектронных, генераторных). Для лучшего восприятия материала в пособии подробно рассмотрены принципы работы классических приборов, но, помимо этого, приведена информация и о современных гетеронаноструктурных диодах.

Пособие не заменяет лекций, лабораторных практикумов или учебников. Оно является дополнительным источником сведений, в котором объяснения простых базовых понятий сочетаются с детальными описаниями сложных физических эффектов. В связи с таким подходом к изложению материала пособие может использоваться в качестве как приложения к базовому курсу лекций, так и методической основы спецкурса по физике современных полупроводниковых приборов и наноэлектронике.

Используемая в пособии терминология сочетается с понятийным аппаратом как хорошо известных учебников, так и достаточно редких монографий по физике полупроводников и полупроводниковых приборов [2–8], а также оригинальных методических разработок кафедры электроники радиофизического факультета Нижегородского государственного университета имени Н. И. Лобачевского.

ЧАСТЬ 1

ФИЗИКА ЯВЛЕНИЙ НА ГРАНИЦЕ РАЗДЕЛА ТВЕРДЫХ ТЕЛ. ВЫПРЯМЛЯЮЩИЕ ДИОДЫ

В первой части представленного учебного пособия по полупроводниковым диодным структурам и приборам изложены элементарные основы теории контактных явлений. При этом речь пойдет как о контактах между различными материалами, например, полупроводником и металлом, так и о контактах между областями с разным типом проводимости внутри одного полупроводникового кристалла. Физические свойства подобных контактов, называемых также переходами, широко используются для выпрямления тока и лежат в основе работы базовых элементов целого ряда полупроводниковых сверхвысокочастотных устройств и быстродействующих интегральных схем, а также приборов полупроводниковой оптоэлектроники.

Параграфы 1.1. – 1.4. данной части посвящены изучению границ раздела между двумя полупроводниками (если быть более точным, то между областями различной проводимости внутри единого полупроводникового кристалла). В параграфах 1.5, 1.6. рассматривается физика контактных явлений на границах материалов с существенно отличающейся по величине проводимостью (границы раздела металл–полупроводник, металл–диэлектрик–полупроводник). В заключение раздела в параграфе 1.7. рассматриваются общие эквивалентные схемы выпрямительных диодов.

1.1. КЛАССИФИКАЦИЯ ГРАНИЦ РАЗДЕЛА МЕЖДУ ДВУМЯ ПОЛУПРОВОДНИКАМИ

Для обозначения уровня легирования полупроводниковых слоев, т.е. величин концентраций ионов доноров или акцепторов, в данном пособии будем использовать общепринятые сокращения:

- n^{++} и p^{++} – концентрация примеси в полупроводнике более 10^{19} см^{-3} ;
- n^{+} и p^{+} – $\sim 10^{18} \dots 10^{19} \text{ см}^{-3}$;
- n и p – $\sim 10^{16} \dots 10^{18} \text{ см}^{-3}$;
- n^{-} и p^{-} – $\sim 10^{15} \dots 10^{16} \text{ см}^{-3}$;
- n^{--} и p^{--} – менее 10^{15} см^{-3} .

Буквы « n » и « p » обозначают, соответственно, электронный и дырочный типы проводимости материала. Слои нелегированного (чистого) полупроводника обозначаются символом « i ». Подразумевается, что температура, при которой работает прибор, является комнатной (кроме случаев, где явно указывается иная ситуация), а концентрация основных носителей заряда равна концентрации введенной примеси.

Для обозначения границы раздела полупроводников, которую также часто называют переходом, используют дефис или косую черту, например, n^{+} - n Si переход, p -Si / n -Ge переход. При этом в первом случае предполагается, что

оба слоя полупроводника изготовлены из кремния, а во втором случае - первая область из кремния, а вторая из германия.

В данном пособии мы будем использовать следующую классификацию границ раздела между двумя полупроводниками (таблица 1.1).

Таблица 1.1. Классификация границ раздела

	Гомопереход	Гетеропереход
Изотипный (униполярный) переход	Один и тот же полупроводник и один и тот же тип проводимости по обе стороны от границы раздела (пример: n^+ - n и p^+ - p переходы)	Разные полупроводники, но один тип проводимости по обе стороны от границы раздела (пример: n -AlGaAs/ n -GaAs и p -AlGaAs/ p -GaAs переходы)
Анизотипный (биполярный) переход	Один и тот же полупроводник, но различные типы проводимости с обеих сторон границы раздела (пример: p - n переход)	Различные полупроводники и различные типы проводимости с обеих сторон границы раздела (пример: p -AlGaAs/ n -InP переход)

Примечание: GaAs читается как арсенид галлия, InP – фосфид индия, AlGaAs – тройное соединение алюминий-галлий-мышьяк.

1.2. ИЗОТИПНЫЙ (УНИПОЛЯРНЫЙ) ГОМОПЕРЕХОД

Рассмотрим униполярный гомопереход на примере n^+ - n перехода (рис. 1.1 а).

Проведем мысленный эксперимент. Возьмем два образца одного и того же полупроводника, отличающиеся друг от друга степенью легирования. Для каждого из них концентрация ионов доноров равна концентрации электронов, т.е. образцы электронейтральны¹. Соединим два полупроводниковых образца с различными концентрациями примесей так, чтобы кристаллические решетки образцов точно совпали, а на границе раздела не оказалось бы ни одного дефекта².

¹ В данной ситуации неосновные носители заряда можно не учитывать.

² Подобное механическое соединение на самом деле невозможно, так как на поверхности полупроводникового образца всегда присутствуют окислы, которые, будучи по своей природе диэлектриками, препятствуют протеканию тока. В процессе изготовления твердотельных приборов никакой «стыковки» кристаллических решеток не происходит, просто различные области единого монокристалла легируются разными примесями (подробнее о технологии роста кристаллов см. [8]).

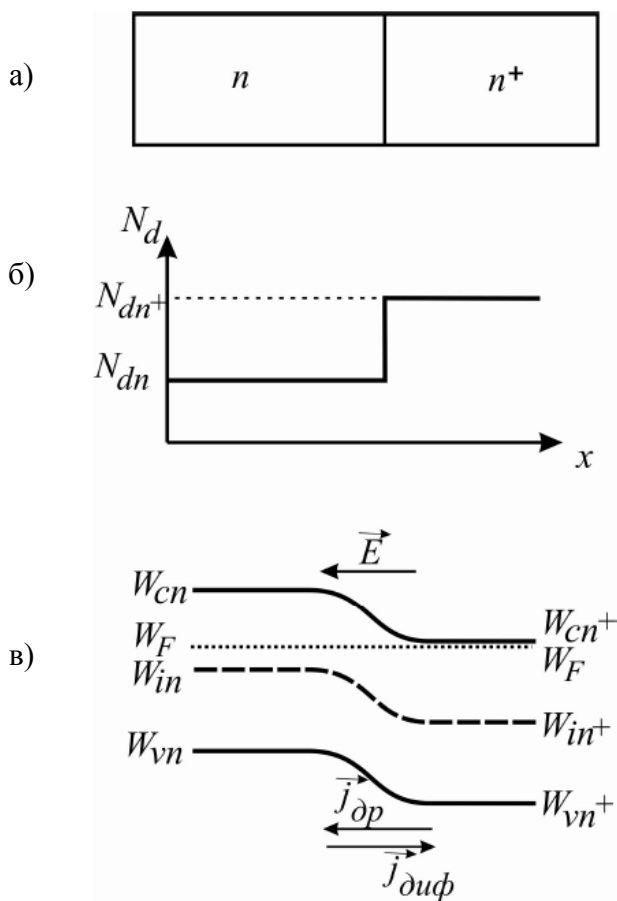


Рис. 1.1. Униполярный n^+ - n переход в состоянии равновесия: а) схематическое изображение конструкции; б) график зависимости концентрации доноров от координаты; в) энергетическая зонная диаграмма. W_c , W_v – соответственно, уровни дна зоны проводимости и потолка валентной зоны; W_F – уровень Ферми; W_i – уровень Ферми в собственном полупроводнике (для простоты будем считать, что он лежит в середине запрещенной зоны). E – напряженность встроенного электрического поля; \vec{J}_{dp} , \vec{J}_{dif} – соответственно, плотности дрейфового и диффузионно токов в переходе. На рисунке вторым индексом в обозначениях величин указан тип полупроводника, к которому данные величины относятся. В дальнейшем подобные уточнения мы будем опускать

Так как концентрация электронов в сильно легированной n^+ - области больше, чем в n – области (рис. 1.1. б), то возникнет диффузионный поток носителей заряда, который будет направлен справа налево (напомним, что диффузионный ток при этом направлен слева направо, поскольку направление тока выбирается по направлению движения положительных зарядов). Поскольку обе части полупроводника изначально нейтральны, то уход электронов из n^+ области означает, что там останется нескомпенсированный положительный заряд ионов доноров. Соответственно, в n -области будут копиться избыточные электроны. В итоге образуется *встроенное электрическое поле*, которое, в свою очередь, вызовет встречный дрейфовый ток носителей заряда. Сначала электроны будут преимущественно переходить из правой области в левую, затем внутреннее поле достигнет такой величины, при которой создаваемый им дрейфовый ток уравнивает ток диффузионный, и будет достигнуто *динамическое равновесие*. Зонная диаграмма перехода в равновесном состоянии показана на рис. 1.1 в.

Обратите внимание, что вектор напряженности электрического поля \vec{E} направлен в сторону роста энергии дна зоны проводимости (в данном случае – от положительных зарядов ионов доноров, обнажившихся в n^+ слое, к отрицательным зарядам электронов, скопившимся в n -слое). Поскольку на дне зоны проводимости кинетическая энергия носителей $W_k=0$, то вклад в полную энергию дает только потенциальная составляющая $U(r)=-e\phi(r)$, где e – абсолютная величина заряда электрона, $\phi(r)$ – потенциал. Поскольку $\vec{E}=-\nabla\phi$, то

вектор напряженности электрического поля будет сонаправлен с $\nabla U(r)$, а значит, направлен в сторону роста энергии дна зоны проводимости³.

На практике чаще всего используют изотипные переходы с уровнями легирования от 10^{14} до 10^{18} см⁻³. Высота потенциального барьера $n^+ - n^-$ - перехода зависит от разницы концентраций между областями перехода и имеет величину $1...10 kT$, а электрическое сопротивление, обусловленное таким барьером, меньше сопротивления остальных переходов, которые рассматриваются ниже.

1.3. Анизотипный (биполярный) ГОМОПЕРЕХОД (p - n ПЕРЕХОД)

Электронно-дырочным или p - n переходом называется приконтактная область между частями полупроводника с электронным (n) и дырочным (p) типами проводимости. Наиболее простой метод получения p - n переходов состоит во введении донорной и акцепторной примесей в процессе роста кристалла при эпитаксии, с помощью диффузии или ионного легирования.

В зависимости от характера распределения примесей различают резкий (ступенчатый) и плавный p - n переходы. Мы рассмотрим резкий p - n переход, в котором значения концентраций донорной N_d и акцепторной N_a примесей изменяются скачком на границе раздела.

1.3.1. Электронно-дырочный переход в равновесном состоянии

Снова проведем мысленный эксперимент: будем соединять в единое целое два полупроводниковых образца с различными типами проводимости так, что кристаллические решетки образцов точно совпадут, а на границе раздела не будет дефектов (рис.1.2).

Так как в p -области концентрация дырок (p_p) – основных носителей заряда – значительно больше, чем в n -области (p_n), то происходит

1) диффузия

дырок в n -область, где они окажутся неосновными носителями заряда. Таким образом, в некотором слое n -области, примыкающем к границе раздела, увеличивается концентрация неосновных носителей. Следовательно, там будет идти интенсивная

2) рекомбинация

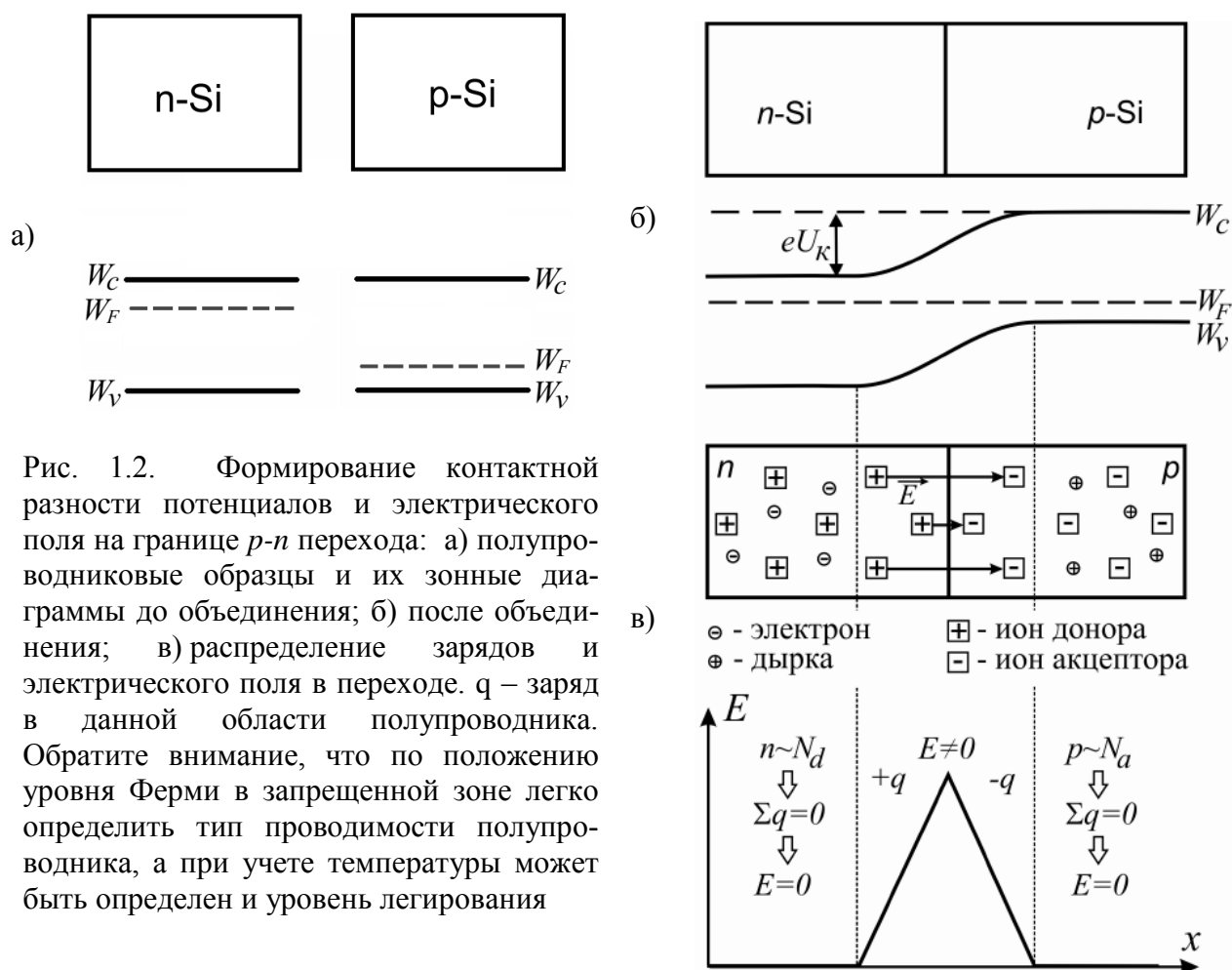
электронов и дырок. Поскольку изначально n -область полупроводника была электронейтральной (концентрация электронов равнялась концентрации ионов

³ Мы можем пользоваться соотношениями электростатики, т.к. скорость носителей заряда в полупроводниках на порядки меньше скорости света, т.е. можно считать, что при перемещении электронов и дырок поле мгновенно отслеживает изменение положения зарядов. Эта особенность характерна для полупроводниковых приборов, в отличие от приборов вакуумной электроники.

доноров), а электроны, оказавшиеся вблизи границы, рекомбинируют с дырками, пришедшими из p -слоя, то в приграничной области

3) *обнажится некомпенсированный объемный заряд*

положительного знака, обусловленный **ионами** донорной **примеси**. Аналогично, диффузия и рекомбинация электронов будут сопровождаться появлением в p -области отрицательного объемного заряда ионов акцепторной примеси (рис. 1.2 в). Таким образом, в приконтактной области появится встроенное электрическое поле, напряженность которого будет постепенно увеличиваться. В свою очередь, это поле вызовет встречный дрейфовый поток носителей заряда и будет препятствовать диффузионному движению. Очевидно, что когда диффузия будет полностью скомпенсирована дрейфом, система придет в состояние равновесия, а рост поля прекратится. При этом в полупроводнике установится определенная напряженность электрического поля и соответствующая ему *контактная разность потенциалов* U_K (рис. 1.2 в).



Подобное равновесие называют динамическим, поскольку движение носителей заряда прекращается лишь в среднем. Согласно распределению Больцмана, всегда найдутся высокоэнергетические носители заряда, которые смогут преодолеть имеющийся потенциальный барьер eU_K . Однако, точно такое же количество носителей заряда перейдет в противоположную сторону под действием встроенного электрического поля, поэтому суммарный ток будет

равен нулю ($\Sigma j = 0$). При установлении равновесия происходит выравнивание термодинамических характеристик, в частности, уровней Ферми, в областях с различными типами проводимости, т.е. слева и справа от перехода. Приконтактную область, где имеется электрическое поле, называют *p-n переходом, запирающим слоем* или *областью пространственного заряда (ОПЗ)*.

Обратите внимание, что ионы примесей присутствуют во всем полупроводнике (доноры – в *n*-области, акцепторы – в *p*-области), а не только в самом переходе. Однако, вне ОПЗ образец электронейтрален (рис. 1.2.в).

На рис. 1.3 представлены распределения концентрации примеси (а), плотности объемного заряда (б), напряженности электрического поля (в) и потенциала (г) в резком несимметричном *p-n* переходе.

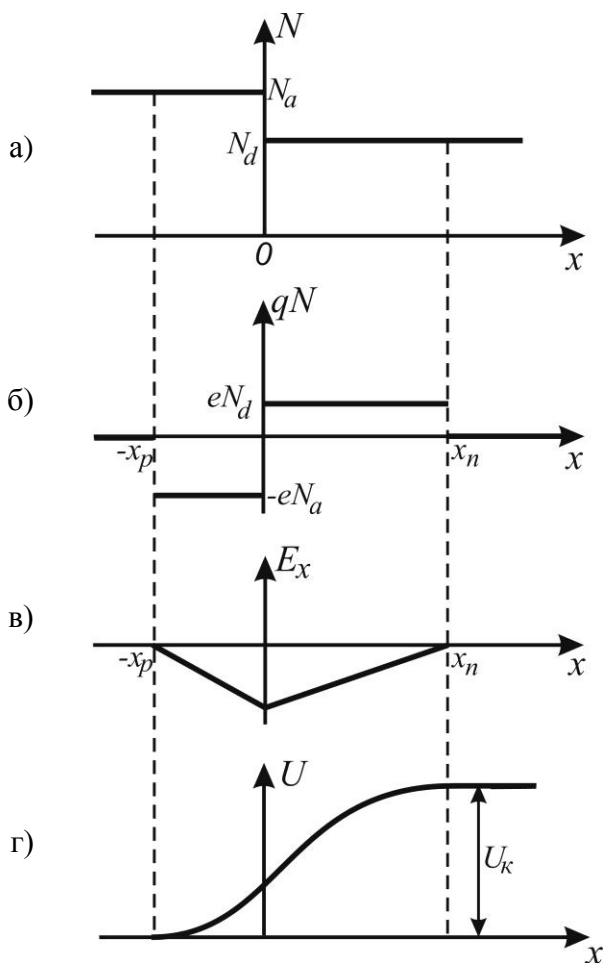


Рис. 1.3. Распределение концентрации примесей (а), плотности объемного заряда (б), поля (в) и потенциала (г) в *p-n* переходе. Обратите внимание, что на данном рисунке справа расположена область *n*-типа, а слева *p*-типа, так что электрическое поле, образованное ионами доноров и акцепторов, направлено справа налево ($E_x < 0$).

Так как по условию задачи величины концентраций доноров и акцепторов различны, а их суммарные заряды в обедненных электронами и дырками областях равны друг другу ($eN_d x_n = eN_a x_p$), то размеры самих областей различны. Они тем меньше, чем выше уровень легирования, что является следствием закона сохранения заряда

В равновесном состоянии величины концентраций электронов и дырок в невырожденном примесном полупроводнике следующим образом зависят от температуры:

$$n = N_c e^{-\frac{W_c - W_F}{kT}}, \quad p = N_v e^{-\frac{W_F - W_v}{kT}}. \quad (1.1)$$

Для собственного полупроводника эти концентрации равны:

$$n_i = p_i = N_c \cdot e^{-\frac{W_c - W_i}{kT}} = N_v \cdot e^{-\frac{W_i - W_v}{kT}}. \quad (1.2)$$

Здесь W_c , W_v , W_i – уровни энергии, соответствующие нижнему уровню зоны проводимости, верхнему уровню валентной зоны, середине запрещенной зоны; W_F – уровень Ферми; N_c , N_v – эффективные плотности квантовых состояний в зоне проводимости и валентной зоне, соответственно; n_i – собственная концентрация носителей заряда.

Энергию середины запрещенной зоны W_{ip} и W_{in} для p - и n -типа полупроводника и энергию Ферми W_F можно записать через соответствующие потенциалы:

$$W_{ip} = -e\varphi_{ip}, \quad W_{in} = -e\varphi_{in}, \quad W_F = -e\varphi_F \quad (1.3)$$

На основании соотношения (1.2) с учетом выражений (1.3) величины концентрации основных носителей вдали от p - n перехода могут быть выражены равенствами

$$p_p = n_i \cdot e^{\frac{\varphi_F - \varphi_{ip}}{\varphi_T}}, \quad n_n = n_i \cdot e^{\frac{\varphi_{in} - \varphi_F}{\varphi_T}}, \quad (1.4)$$

где $\varphi_T = \frac{kT}{e}$ – так называемый, тепловой потенциал. Равновесные концентрации основных (n_n , p_p) и неосновных (n_p , p_n) носителей заряда в невырожденных полупроводниках p - и n -типа связаны соотношением (так называемый, «закон действующих масс»):

$$n_p p_p = n_n p_n = n_i^2 = N_c \cdot N_v \cdot e^{-\frac{W_g}{kT}},$$

где W_g – ширина запрещенной зоны полупроводника.

Контактная разность потенциалов определяется соотношением:

$$U_k = \varphi_{in} - \varphi_{ip}. \quad (1.5)$$

Выразим U_k через равновесные концентрации электронов и дырок. Для этого из выражений (1.4) найдем φ_{ip} , φ_{in} и подставим в уравнение (1.5). После преобразований получим:

$$U_k = \varphi_T \ln \frac{n_n p_p}{n_i^2} = \varphi_T \ln \frac{N_d N_a}{n_i^2} = \varphi_T \ln \frac{n_n}{n_p} = \varphi_T \ln \frac{p_p}{p_n} \quad (1.6)$$

или

$$p_n = p_p \cdot e^{-U_k / \varphi_T}, \quad n_p = n_n \cdot e^{-U_k / \varphi_T}. \quad (1.7)$$

Таким образом, высота потенциального барьера p - n перехода определяется температурой, концентрацией легирующей примеси и собственной концен-

трацией носителей при заданной температуре (т.е. типом материала). При этом, как видно из зонной диаграммы, величина контактной разности потенциалов в невырожденном p - n переходе ограничена значением ширины запрещенной зоны полупроводника W_g . Если же одна область легирована сильно, а другая – слабо, то $U_k < W_g/2$.

1.3.2. Вольт-амперная характеристика p - n перехода

Пусть к электронно-дырочному переходу подключен источник ЭДС таким образом, чтобы потенциальный барьер уменьшился. Такое подключение называется *прямым* и оно соответствует подсоединению источника плюсом к p -области и минусом к n -области. В этом случае из источника ЭДС в область n -типа поставляются электроны, а p -типа – дырки. Эти носители компенсируют заряд части доноров и части акцепторов, так что оставшийся в ОПЗ объемный заряд ионов становится меньше, что приводит к уменьшению напряженности электрического поля. В свою очередь, это приводит к тому, что по обе стороны от ОПЗ число основных носителей, которые имеют возможность преодолевать барьер, экспоненциально возрастает. Точно так же возрастает и величина диффузионного тока $j_{диф}$. При этом дрейфовый ток неосновных носителей заряда, направленный в противоположную сторону, практически не меняется, поскольку в сильном поле перехода имеет место режим насыщения дрейфовой скорости носителей заряда. Таким образом, при положительном подключении в p - n переходе преобладает диффузионный ток, т.е. $j = j_{диф} - j_{др} > 0$. При этом основные носители переходят в область с противоположным типом проводимости. Такой процесс называется *инжекцией*. Носители проникают через энергетический барьер в области, где они оказываются неосновными и рекомбинируют. Эти избыточные неравновесные носители нарушают электронейтральность полупроводника вблизи перехода и вызывают в равном количестве приток основных носителей из глубины p - и n -областей. Скорость рекомбинации электронов и дырок конечна, поэтому неравновесные носители могут продвигаться вглубь полупроводника, причем глубина их проникновения значительно превысит толщину ОПЗ. При этом электронейтральность кристалла за пределами области объемного заряда не нарушается.

Таким образом, при приложении внешнего напряжения в прямом направлении в результате инжекции носителей через p - n переход будет протекать ток, величина которого будет нарастать с увеличением приложенного напряжения, т.е. с уменьшением высоты барьера перехода.

При *обратном* включении внешнего напряжения (минусом к p -области и плюсом к n -области) направления внешнего и встроенного полей совпадают, т.е. высота барьера увеличивается. При этом диффузионный поток носителей спадает практически до нуля, а суммарный ток определяется дрейфовым током неосновных носителей заряда и незначителен по величине (поскольку количество неосновных носителей на несколько порядков меньше, чем основных).

Зонные диаграммы p - n перехода при разной полярности приложенного напряжения изображены на рис. 1.4.

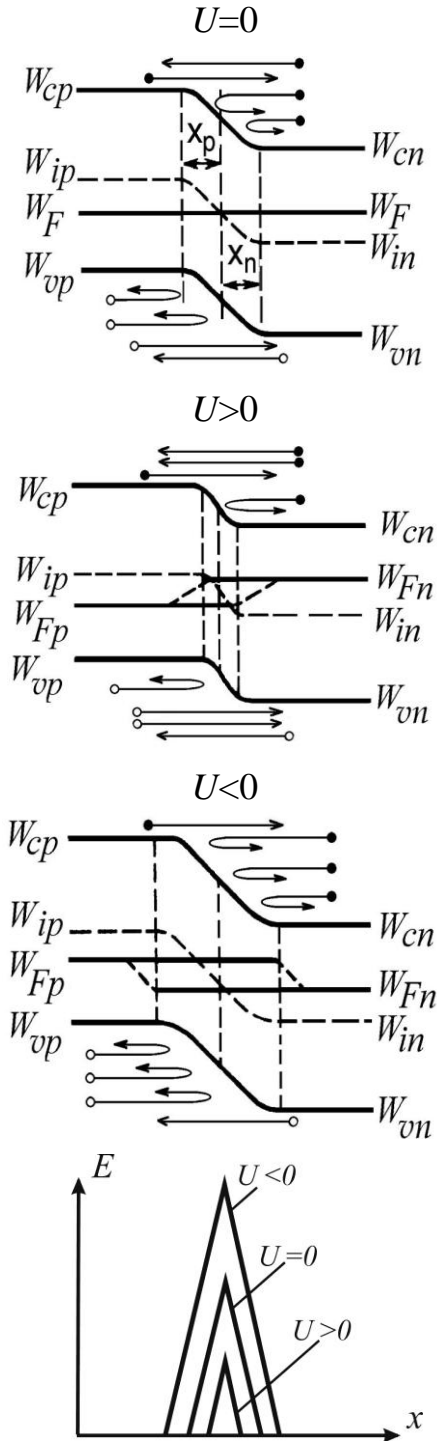


Рис. 1.4. Зонные диаграммы и распределение абсолютной величины напряженности электрического поля E в p - n переходе при различных внешних напряжениях U . Важно, что при изменении внешнего напряжения происходит расширение или сужение области перехода, так что в нее попадает большее или меньшее количество ионов доноров и акцепторов. Это определяет увеличение или уменьшение напряженности поля в переходе

При неравновесных условиях принято вводить два новых параметра распределения: W_{Fn} – для электронов и W_{Fp} – для дырок. Их значения выбирают таким образом, чтобы для концентраций электронов и дырок при наличии неравновесных носителей оставались справедливыми соотношения (1.1) и (1.4). Величины W_{Fn} и W_{Fp} называют квазиуровнями Ферми для электронов и дырок, соответственно. Таким образом, в невырожденных полупроводниках справедливы соотношения:

$$n = N_C e^{-\frac{W_C - W_{Fn}}{kT}}, \quad (1.7a)$$

$$p = N_V e^{-\frac{W_{Fp} - W_V}{kT}},$$

$$n = n_i \cdot e^{\frac{W_{Fn} - W_i}{kT}}, \quad (1.7б)$$

$$p = p_i \cdot e^{-\frac{W_{Fp} - W_i}{kT}}.$$

Ввиду сложности строгого анализа, обычно делают еще ряд допущений, упрощающих решение задачи:

1. Модель электронно-дырочного перехода одномерная, p - и n -области имеют бесконечную протяженность.
2. В p - и n -областях примеси распределены равномерно, а на границе раздела значения их концентраций изменяются скачком.

3. Уровень инжекции мал (полагают малым внешнее напряжение).
4. Концентрация неосновных носителей мала по сравнению с концентрацией основных носителей. В этом случае ток вдали от p - n -перехода будет определяться основными носителями.
5. Электроны и дырки исчезают только вследствие рекомбинации друг с другом. Пренебрегают наличием ловушек для носителей заряда.
6. Генерация и рекомбинация в ОПЗ отсутствуют.
7. Явления, связанные с пробоем перехода, отсутствуют.

При этих допущениях для токов неосновных носителей вне ОПЗ можно записать:

$$j_p = -eD_p \frac{dp}{dx}, \quad j_n = eD_n \frac{dn}{dx}, \quad (1.8)$$

где D_p, D_n – соответственно, коэффициенты диффузии дырок и электронов.

Одномерные уравнения непрерывности для дырок в n -области и электронов в p -области при отсутствии генерации можно записать в виде:

$$\frac{\partial p}{\partial t} = -\frac{1}{e} \cdot \frac{\partial j_p}{\partial x} - \frac{p - p_{n0}}{\tau_p}, \quad \frac{\partial n}{\partial t} = \frac{1}{e} \cdot \frac{\partial j_n}{\partial x} - \frac{n - n_{p0}}{\tau_n}, \quad (1.9)$$

где τ_p и τ_n – времена жизни носителей заряда, p_{n0} и n_{p0} – значения концентрации равновесных дырок в n -области и равновесных электронов в p -области, соответственно.

При нахождении статической вольт-амперной характеристики необходимо решить уравнение непрерывности для случая, когда концентрация неосновных носителей не меняется во времени:

$$\frac{\partial p}{\partial t} = \frac{\partial n}{\partial t} = 0. \quad (1.10)$$

Из уравнений (1.9) с учетом (1.8) и (1.10) получим:

$$\begin{aligned} \frac{d^2 p}{dx^2} - \frac{p - p_{n0}}{L_p^2} &= 0 \text{ при } x \geq x_n, \\ \frac{d^2 n}{dx^2} - \frac{n - n_{p0}}{L_n^2} &= 0 \text{ при } x \leq -x_p, \end{aligned} \quad (1.11)$$

где $L_p = \sqrt{D_p \tau_p}$ – диффузионная длина дырок в n -области, $L_n = \sqrt{D_n \tau_n}$ – диффузионная длина электронов в p -области. Граничными условиями являются:

$$\begin{aligned} p(x \rightarrow \infty) &\rightarrow p_{n0}; \quad n(x \rightarrow -\infty) \rightarrow n_{p0}; \\ p(x_n) &= p_{n1}; \quad n(-x_p) = n_{p1}. \end{aligned}$$

Решая уравнения (1.11) при этих условиях, получим:

$$p(x) = p_{n0} + (p_{n1} - p_{n0})e^{-(x-x_n)/L_p},$$

$$n(x) = n_{p0} + (n_{p1} - n_{p0})e^{(x+x_p)/L_n}.$$
(1.12)

При приложении внешнего напряжения U в прямом направлении высота потенциального барьера становится равной $e(U_k - U)$. При этом концентрации неосновных носителей на границе ОПЗ на основании соотношения (1.7) будут выражаться в следующем виде:

$$p_{n1} = p_p \cdot e^{-\frac{e(U_k - U)}{kT}} = p_{n0} e^{\frac{eU}{kT}},$$

$$n_{p1} = n_n \cdot e^{\frac{e(U_k - U)}{kT}} = n_{p0} e^{\frac{eU}{kT}}.$$
(1.13)

Подставляя (1.13) в (1.12), определим электронный и дырочный токи в точках $x = -x_p$ и $x = x_n$, соответственно:

$$j_n(-x_p) = eD_n \frac{dn}{dx} = \frac{eD_n n_{p0}}{L_n} \left(e^{\frac{eU}{kT}} - 1 \right),$$

$$j_p(x_n) = -eD_p \frac{dp}{dx} = \frac{eD_p p_{n0}}{L_p} \left(e^{\frac{eU}{kT}} - 1 \right).$$
(1.14)

В предположении отсутствия генерации и рекомбинации в ОПЗ плотности токов j_n и j_p в интервале $-x_p < x < x_n$ не зависят от координаты, т.е.:

$$j_n(x_n) = j_n(-x_p); \quad j_p(-x_p) = j_p(x_n);$$
(1.15)

Полный ток в стационарном режиме во всех сечениях одинаков. Проще всего вычислить ток на границах перехода в точках $x = -x_p$ или $x = x_n$. С учетом (1.15) плотность полного тока выражается соотношением:

$$j = j_n(-x_p) + j_p(x_n) = \left(\frac{eD_n n_{p0}}{L_n} + \frac{eD_p p_{n0}}{L_p} \right) \left(e^{\frac{eU}{kT}} - 1 \right).$$
(1.16)

На рис. 1.5 показаны распределения концентраций электронов и дырок в p - n переходе при прямом и обратном смещениях.

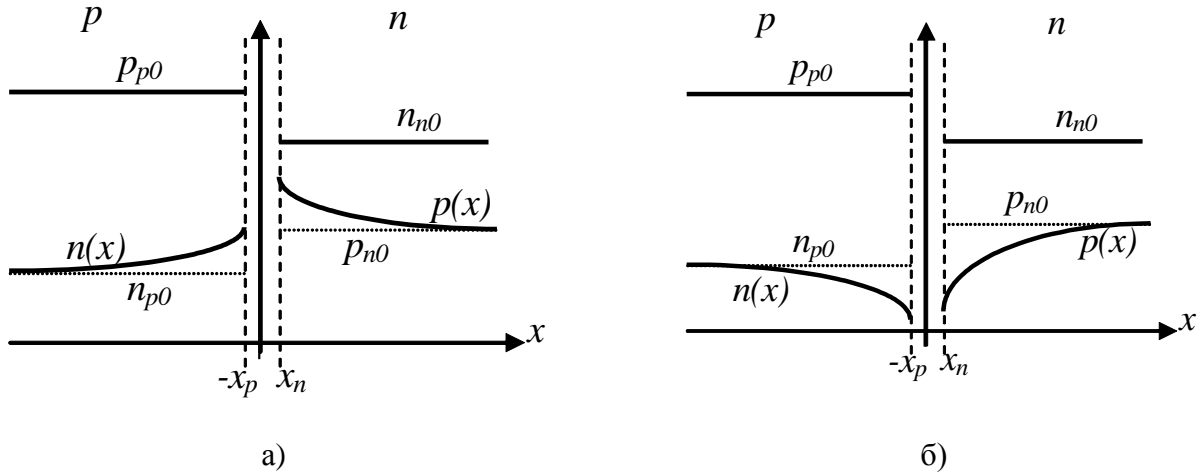


Рис. 1.5. Распределение величин концентраций электронов и дырок в p - n переходе: а) при прямом смещении; б) при обратном смещении

Если площадь перехода S , то полный ток через переход будет равен $J=jS$. Тогда вольт-амперную характеристику перехода можно записать в виде:

$$J = J_S \left(e^{\frac{U}{\phi_r}} - 1 \right). \quad (1.17)$$

Величина обратного тока перехода с увеличением обратного напряжения стремится к J_S , поэтому величину

$$J_S = \left(\frac{ep_{n0}L_p}{\tau_p} + \frac{en_{p0}L_n}{\tau_n} \right) S \quad (1.18)$$

называют *током насыщения* или *обратным током* p - n перехода. В (1.18) использованы соотношения: $L_n = \sqrt{D_n \phi_n}$, $L_p = \sqrt{D_p \phi_p}$.

Иногда, подчеркивая природу этого тока, его называют *током тепловой генерации* или просто *тепловым током*, т.к. обуславливающие его неосновные носители появляются в нейтральных p - и n -областях, прилегающих к переходу, за счет тепловой генерации. Эти носители диффундируют к границам перехода, захватываются его полем и переносятся в соседнюю область. Механизм образования теплового тока отражает формула (1.18), в которой p_{n0}/τ_p и n_{p0}/τ_n – скорости генерации неосновных носителей, а $SL_p p_{n0}/\tau_p$ и $SL_n n_{p0}/\tau_n$ – полное число неосновных носителей, генерируемых в слоях толщиной L_p и L_n за единицу времени. Именно эти носители без рекомбинации доходят до границы перехода и образуют обратный ток.

Можно рассматривать физический смысл выражения (1.18) и с другой стороны: отношения L_p/τ_p и L_n/τ_n имеют смысл скорости диффузии носителей заряда. Тогда формулу (1.18) для тока насыщения можно записать в классическом виде – как произведение заряда частиц, их концентрации, скорости и

площади сечения образца: $J_S = (ep_{n0} V_p + en_{p0} V_n)S$. Так как имеется ввиду скорость диффузионного движения носителей заряда в областях за пределами пространственного заряда перехода, то ток не зависит от величины обратного напряжения, подаваемого на переход, хотя напряженность поля перехода и ширина ОПЗ от внешнего напряжения зависеть будут. При попадании неосновных носителей в область перехода они будут подхвачены полем, и, независимо от величины последнего, переброшены на другую сторону ОПЗ. Таким образом, ток в данной ситуации будет ограничиваться сопротивлением областей сбора неосновных носителей и практически не будет зависеть от напряжения. Вольт-амперная характеристика диода, рассчитанная в рамках рассмотренной идеализированной модели, представлена на рисунке 1.10 сплошной линией.

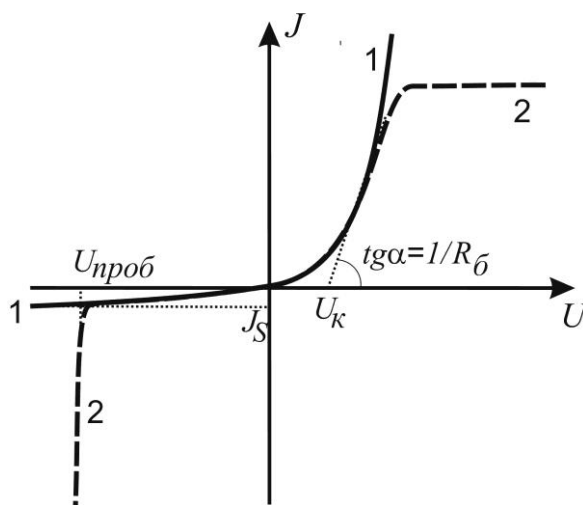


Рис. 1.6. ВАХ p - n перехода:
1) идеальная; 2) реальная.

Здесь R_b — сопротивление базы, U_k — контактная разность потенциалов, J_S — обратный ток p - n перехода, $U_{проб}$ — напряжение пробоя

Выражение (1.17) хорошо согласуется с экспериментальными данными. Отличия от теоретической зависимости наблюдаются при увеличении прямого тока и при достаточно больших обратных смещениях, когда имеет место резкое возрастание обратного тока перехода. Рассмотрим обе ситуации.

У реального диода последовательно с сопротивлением p - n перехода имеются сопротивления квазинейтральных областей полупроводника. Область с большей концентрацией примесей (а, следовательно, с большей концентрацией свободных носителей и меньшим сопротивлением) называют эмиттером, а с меньшей — базой; сопротивление эмиттера обычно мало и им пренебрегают. При выводе формулы (1.17) мы неявно предполагали, что все прикладываемое внешнее напряжение падает непосредственно на области перехода. Однако, это справедливо только при малых токах, т.е. при отрицательных и небольших положительных смещениях ($U \leq U_k$). В этих случаях переход существенно обеднен свободными носителями заряда, а сопротивление ОПЗ велико. Соответственно, и падение напряжения на этой области много больше, чем на других. При больших прямых токах падение напряжения на базе уже соизмеримо с падением напряжения на переходе. С учетом сопротивления базы аналитическое выражение зависимости тока диода от приложенного к нему напряжения может быть представлено в следующем виде:

$$J = J_s \left[e^{(U/\varphi_T - JR_\delta/\varphi_T)} - 1 \right], \quad (1.19)$$

где U - напряжение, приложенное к диоду, R_δ - сопротивление базы.

Проведя логарифмирование и дифференцирование выражения (1.19), определим дифференциальное сопротивление в произвольной точке вольт-амперной характеристики:

$$R_\delta = \frac{dU}{dJ} = \frac{1}{\frac{dJ}{dU}} = \frac{\varphi_T}{(J + J_s)} + R_\delta. \quad (1.20)$$

Видно, что при малых токах первое слагаемое в формуле (1.20) много больше второго, т.е. сопротивление диода в этом случае определяется сопротивлением самого p - n перехода (в этой области токов реальная и идеальная ВАХ совпадают). При больших токах дифференциальное сопротивление перехода мало и общее сопротивление определяется сопротивлением базы, т.е. зависимость тока от напряжения представляет собой прямую линию, тангенс угла наклона которой равен $1/R_\delta$. При дальнейшем увеличении прямого напряжения ток прибора выходит на насыщение. Причины этого явления зависят от конструкции прибора и определяются насыщением зависимости скорости носителей заряда в сильных электрических полях, малой концентрацией носителей заряда при слабом легировании полупроводника и разогревом полупроводника протекающим током.

Резкое возрастание обратного тока при увеличении напряжения выше критического, называемого напряжением пробоя $U_{проб}$, может быть вызвано следующими эффектами:

- туннелированием электронов сквозь узкий и высокий потенциальный барьер из валентной зоны в зону проводимости при большой напряженности электрического поля в ОПЗ (эффект Зинера или *туннельный пробой*);
- ударной ионизацией атомов полупроводника в сильном электрическом поле ОПЗ (*лавинный пробой*);
- перебросом электронов из валентной зоны в зону проводимости за счет саморазогрева (*тепловой пробой*).

Преобладание того или иного из упомянутых механизмов зависит от материала полупроводника, конструкции диода и температуры. *Туннельный пробой* наблюдается, как правило, в вырожденных полупроводниках. Толщина ОПЗ в них настолько мала (~ 10 нм), что при высоком значении напряженности поля в переходе (а значит, при сильном «наклоне» зон) становятся возможными туннельные переходы электронов с занятых состояний в валентной зоне на свободные состояния зоны проводимости. При этом величина тока экспоненциально возрастает.

Лавинный пробой развивается в случае, когда поле в полупроводнике (или в приборной структуре) настолько велико, что на длине свободного пробега

носители набирают энергию, достаточную для ударной ионизации атомов вещества. При этом налетающий горячий электрон, который приобрел энергию больше ширины запрещенной зоны, «выбивает» электрон из валентной зоны. В результате возникает пара противоположно заряженных частиц (электрон и дырка), одна или обе из которых также начинают участвовать в ударной ионизации вместе с исходным электроном. В такой ситуации происходит лавинообразное нарастание числа участвующих в процессе носителей, откуда и название данного типа пробоя.

Наконец, в случае протекания сильного тока полупроводниковая структура разогревается, а, значит, растет и количество термогенерированных носителей заряда, что, в свою очередь, приводит к дальнейшему росту тока. В результате развивается *тепловой пробой*, который приводит к разрушению кристаллической решетки полупроводника и выходу прибора из строя. Отметим, что туннельный и лавинный пробой сами по себе обратимы, однако, нельзя допускать их перехода в пробой тепловой. Величина напряжения пробоя у германиевых и кремниевых диодов с электронно-дырочными переходами может достигать сотен и даже тысяч вольт.

Получим выражение для связи напряжения на диоде с максимальной напряженностью поля в переходе, обуславливающей лавинный пробой. Учтем, что поле при равномерном распределении легирующей примеси зависит от координаты линейно, т.е. график имеет вид треугольника, а потенциал имеет смысл площади под графиком поля. Из геометрической формулы для площади треугольника имеем:

$$E_{\max} = \frac{2(U_{\kappa} + U)}{d}. \quad (1.20 \text{ а})$$

Здесь U – абсолютная величина обратного смещения, т.е. $U > 0$; d – ширина области пространственного заряда перехода

$$d = \sqrt{\frac{2(U_{\kappa} + U)\varepsilon\varepsilon_0}{eN_d}}, \quad (1.20 \text{ б})$$

где ε – диэлектрическая проницаемость полупроводника, ε_0 – диэлектрическая постоянная.

Тогда максимальное поле в переходе

$$E_{\max} = \sqrt{\frac{2eN_d(U_{\kappa} + U)}{\varepsilon\varepsilon_0}}. \quad (1.20 \text{ в})$$

Для того, чтобы в структуре начал развиваться лавинный пробой, необходимо, чтобы найденное максимальное значение электрического поля в переходе превысило некоторую определенную для заданного материала величину, т.е. стало бы больше напряженности *поля пробоя* $E_{\text{проб}}$. Для оценки этой величины вспомним, что при возникновении лавинного пробоя электрон на длине сво-

бодного пробега набирает энергию, необходимую для разрыва валентной связи, т.е. энергию, равную ширине запрещенной зоны W_g . Далее элементарная оценка на основе второго закона Ньютона с учетом эффективной массы электрона и характерного времени свободного пробега $\tau_{своб} \sim 10^{-13}$ с показывает, что электроны будут набирать указанную энергию при напряженности поля $E_{проб} \sim 10^5$ В/см, что для перехода с $N_d \approx N_a \approx 10^{17}$ см⁻³ соответствует напряжению ~ 10 В. Пробой будет развиваться там, где локализовано максимальное поле, т.е. на границе p - n -перехода. Область, в которой реализуется ударная ионизация, будет иметь толщину порядка длины свободного пробега носителей заряда. При увеличении внешнего напряжения толщина этой области увеличивается.

В заключение данного раздела отметим два момента, часто вызывающих вопросы у студентов при первом знакомстве с физикой работы полупроводниковых диодов.

1) Обратите внимание, что *«изогнуть» зонную диаграмму «в другую сторону» при подаче прямого смещения нельзя!* Это означает, что даже при подаче большого «открывающего» диод напряжения невозможно реализовать ситуацию, когда уровни энергии электронов в p -области стали бы ниже соответствующих уровней в n -области. Все подаваемое на диод смещение распределяется между запирающим слоем и базой. Чем большее напряжение мы подаем, тем выше ток, а значит, тем большая часть внешнего смещения падает на области базы и тем меньшая – на области самого перехода. В пределе можно предполагать «выглаживание» потенциального барьера перехода, но не его инверсию.

2) *При увеличении температуры прямой и обратный токи диода будут расти.* Физически это объясняется следующими причинами:

- а) увеличение температуры приводит к росту концентрации носителей заряда из-за тепловой генерации;
- б) при увеличении температуры уменьшается контактная разность потенциалов. Это легко понять из простых соображений. С ростом температуры уровни Ферми p и n областей будут стремиться к середине запрещенной зоны полупроводника, т.е. барьер между частями диода, имеющими разный тип проводимости, будет снижаться.

Таким образом, в формуле для ВАХ диода (1.17) основное влияние на зависимость результирующего тока от температуры оказывают $U_k(T)$ и $J_s(T)$, а не $\varphi_T = kT/e$, как может показаться на первый взгляд.

1.3.3. Емкость электронно-дырочного перехода

Всякий p - n переход, по существу, представляет собой систему двух полупроводниковых слоев, разделенных областью объёмного заряда. Такая система подобна плоскому конденсатору. Опыт показывает, что полупроводниковые диоды обладают значительной емкостью, которая ограничивает их применение

в высокочастотных приборах. Изучение емкости перехода во многих случаях помогает исследовать распределения поля и ионов доноров и акцепторов в ОПЗ, а также более детально объяснить механизм протекания тока.

Найдем ширину потенциального барьера и емкость p - n перехода. Помимо допущений, принятых при выводе вольт-амперной характеристики, будем считать, что вся примесь ионизирована и в области объемного заряда концентрация свободных носителей равна нулю. В этом случае плотность объемного заряда постоянна и определяется только значениями концентраций соответствующих примесей, т.е.

$$\rho_p = -eN_a, \quad \rho_n = eN_d. \quad (1.21)$$

Этот объемный заряд создает электрическое поле, которое проникает в p -область на глубину x_p , а в n -область - на x_n . Вне слоя пространственного заряда напряженность электрического поля равна нулю, т.е.

$$E_{p_x} = -\left. \frac{d\varphi}{dx} \right|_{x=-x_p} = 0; \quad E_{n_x} = -\left. \frac{d\varphi}{dx} \right|_{x=x_n} = 0. \quad (1.22)$$

Распределение электростатического потенциала φ в области объемного заряда можно найти, решив уравнение Пуассона:

$$\frac{d^2\varphi}{dx^2} = -\frac{\rho}{\varepsilon\varepsilon_0}, \quad (1.23)$$

где ε - диэлектрическая проницаемость полупроводника.

Интегрируя это уравнение с учетом граничных условий (1.22), получим выражение для напряженности электрического поля в p - и n -областях:

$$E_{p_x} = -\frac{d\varphi_p}{dx} = -\frac{e}{\varepsilon\varepsilon_0} N_a (x + x_p) \text{ при } -x_p < x < 0, \quad (1.24)$$

$$E_{n_x} = -\frac{d\varphi_n}{dx} = -\frac{e}{\varepsilon\varepsilon_0} N_d (x_n - x) \text{ при } 0 < x < x_n$$

В точке $x=0$ из условия непрерывности поля следует, что

$$E_p(0) = E_n(0). \quad (1.25)$$

Из последнего равенства с учетом (1.24) получим:

$$eN_a x_p = eN_d x_n. \quad (1.26)$$

Это соотношение выражает равенство положительного и отрицательного заряда в ОПЗ, т.е. условие электронейтральности образца.

При приложении к p - n переходу внешнего напряжения в соответствии с допущением об отсутствии электрического поля вне слоя объемного заряда граничные условия для потенциала можно представить следующим образом:

$$\varphi(-x_p)=0 \text{ и } \varphi(x_n)=U_k-U. \quad (1.27)$$

Величина напряжения U в формуле (1.27) положительна при прямом смещении перехода и отрицательна – при обратном.

Интегрируя уравнения (1.24) с учетом граничных условий (1.27), получим

$$\begin{aligned} \varphi_p &= \frac{eN_d}{2\varepsilon\varepsilon_0}(x_p+x)^2 \text{ при } -x_p < x < 0; \\ \varphi_n &= -\frac{eN_d}{2\varepsilon\varepsilon_0}(x_n-x)^2 + U_k - U \text{ при } 0 < x < x_n. \end{aligned} \quad (1.28)$$

Так как потенциал φ в пределах ОПЗ должен быть непрерывен, то $\varphi_p(0)=\varphi_n(0)$, т.е.

$$\frac{e}{2\varepsilon\varepsilon_0}(N_dx_n^2 + N_ax_p^2) = U_k - U. \quad (1.29)$$

Используя соотношение (1.29) и равенство (1.26), получим выражение для ширины области объёмного заряда:

$$d = x_n + x_p = \sqrt{\frac{2\varepsilon\varepsilon_0(N_a + N_d)(U_k - U)}{eN_aN_d}}. \quad (1.30)$$

Как видно из выражения (1.30), ширина ОПЗ уменьшается с увеличением прямого (положительного) напряжения и увеличивается при обратном напряжении.

Изменение ширины области объёмного заряда в связи с изменением напряжения приводит к изменению заряда в p - и n -областях. Поэтому p - n переход ведет себя подобно емкости. Эту емкость называют барьерной, т.к. она связана с образованием потенциального барьера между p - и n -областями.

$$C = \frac{dQ}{dU} = S \sqrt{\frac{e\varepsilon\varepsilon_0N_aN_d}{2(N_a + N_d)(U_k - U)}} = S \frac{\varepsilon\varepsilon_0}{d}, \quad (1.31)$$

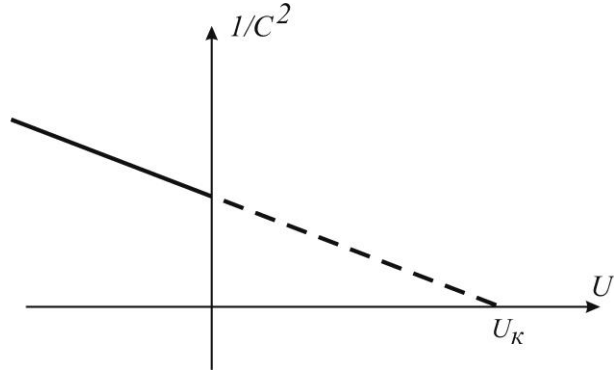
где S - площадь перехода. Обратите внимание, что формула (1.31) совпадает с выражением для емкости плоского конденсатора, хотя в отличие от последнего, в p - n переходе заряд является объемным.

В случае резко несимметричного p - n перехода (когда концентрация легирующей примеси в одной из частей много больше, чем в другой) переход заключен, в основном, в области со слабым легированием. Например, при $N_a \gg N_d$, переход простирается в n -область, а величина барьерной емкости не зависит от свойств p -области:

$$C = S \sqrt{\frac{e\varepsilon\varepsilon_0N_d}{2(U_k - U)}}. \quad (1.32)$$

Выражение (1.32) позволяет найти контактную разность потенциалов и концентрацию донорной примеси. График зависимости $\frac{1}{C^2} = f(U)$, изображенный на рис. 1.7, отсекает на оси абсцисс отрезок, равный по величине U_K . Если известна зависимость $C=f(U)$, то на основании равенства (1.31) можно построить зависимость ширины обеднённой области от приложенного напряжения.

Рис. 1.7. Вольт-фарадная характеристика p - n перехода. Продолжив зависимость в область прямых смещений, можно определить контактную разность потенциалов



Барьерная емкость, рассмотренная выше, вносит основной вклад в емкость перехода при обратном смещении. При прямом смещении в емкости преобладает диффузионная емкость, обусловленная изменением распределения концентрации неосновных носителей заряда. По сути, перераспределение концентрации электронов и дырок является перезарядкой обкладок эквивалентного конденсатора. Такая перезарядка будет происходить тем интенсивнее, чем большее прямое смещение приложено к переходу.

Исходя из зависимости (1.13) при прямом смещении может быть получено выражение для диффузионной емкости перехода:

$$C_{dif} = \frac{e^2}{2kT} (L_p p_{n0} + L_n n_{p0}) \exp\left(\frac{e(U + U_K)}{kT}\right). \quad (1.33)$$

При повышении частоты переменного сигнала емкость не успевает перезаряжаться полностью. Величина диффузионной емкости с увеличением частоты сигнала уменьшается по закону $\sim f^{-1/2}$. Вместе с тем диффузионная емкость быстро возрастает с ростом постоянного тока $\sim \exp\left(\frac{e(U + U_K)}{kT}\right)$. По этим причинам диффузионная емкость играет особенно большую роль на низких частотах и при прямом смещении сравнимом с U_K .

1.4. Анизотипный (биполярный) и изотипный (униполярный) ГЕТЕРОПЕРЕХОДЫ

Гетеропереходы образуются между различными по составу полупроводниками. На рис. 1.8 показана зонная диаграмма перехода между электронным и дырочным полупроводниками с разной шириной запрещенной зоны. Для

примера показана гетероструктура, в которой электронный полупроводник имеет большую ширину запрещенной зоны, чем дырочный – p - N -переход (в литературе более широкозонный полупроводник часто обозначается заглавной буквой).

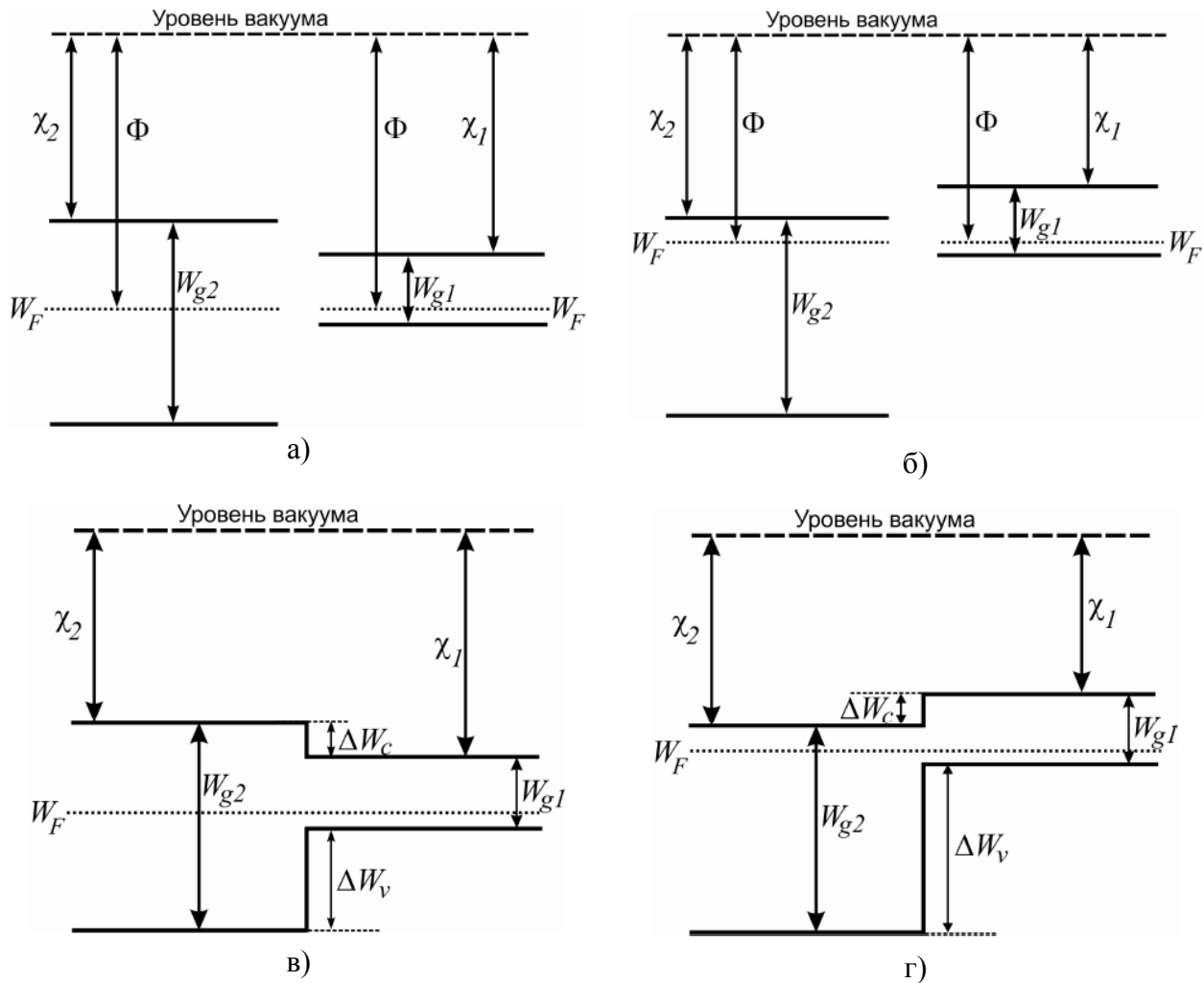


Рис. 1.8. Зонные диаграммы а), б) двух изолированных полупроводников N и p типов с одинаковыми работами выхода $\Phi_1 = \Phi_2 = \Phi$, в), г) гетеро- p - N переходов при условии равенства работ выхода материалов

В отличие от гомоструктурного p - n перехода, на зонной диаграмме гетероструктурного перехода имеются «разрывы», т.е. резкие перепады дна зоны проводимости и/или потолка валентной зоны, величины которых *определяются соотношениями электронного сродства материалов и их ширины запрещенной зоны*. Согласно модели Андерсона, величина «разрыва» дна зоны проводимости

$$\Delta W_c = \chi_1 - \chi_2, \quad (1.34)$$

где χ_1 и χ_2 - электронные сродства контактирующих полупроводников. Соответствующий «разрыв» потолка валентной зоны

$$\Delta W_v = \Delta W_g - \Delta W_c, \quad (1.35)$$

где ΔW_g - разность ширины запрещенных зон полупроводников. На рисунке 1.12 а, б для примера приведены варианты зонных структур, которые могут

наблюдаться при одинаковых значениях работы выхода, но различных соотношениях между шириной запрещенной зоны и электронным сродством материалов p и N -типа.

В реальных ситуациях полупроводниковые слои гетеропереходов отличаются между собой не только электронным сродством и шириной запрещенной зоны, но и работой выхода⁴. В этом случае, как и в гомоструктурном p - n переходе, будет происходить перераспределение зарядов и появление *встроенного электрического поля* вблизи границы раздела слоев, т.е. формирование ОПЗ. На зонной диаграмме это отразится появлением изгиба зон (рис. 1.9).

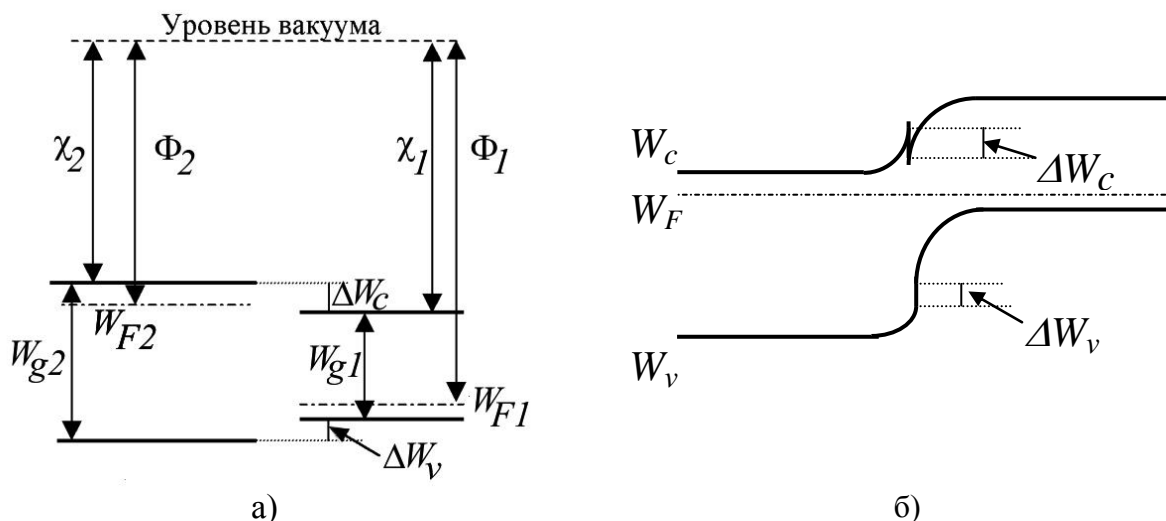


Рис. 1.9. Зонные диаграммы а) двух изолированных полупроводников N и p типов с разными работами выхода Φ_1 и Φ_2 (уровень вакуума принят за начало отсчета энергии); б) p - N перехода

При этом, поскольку непосредственно на границе слоев должны остаться разрывы дна зоны проводимости и потолка валентной зоны определенной величины, на зонной диаграмме может появиться характерный пикок (рис. 1.9). Отметим, что в зависимости от конкретной комбинации значений электронного сродства, уровней легирования и ширины запрещенной зоны материалов гетероструктуры подобный пикок может образоваться в зоне проводимости, в валентной зоне, в обеих зонах или ни в одной из них.

Сходство гетеро- и гомо- p - n переходов заключается в том, что электрическое поле в ОПЗ образовано зарядами ионов доноров и акцепторов (для изотипного гетероперехода электрическое поле будет образовано ионами доноров и электронами; см. n^+ - n переход). Их отличие в том, что на границе раздела гетероперехода химические связи между атомами кристаллической

⁴ Напомним, что под работой выхода понимается энергия, которую необходимо затратить для выхода электрона из твердого тела в вакуум в состояние с кинетической энергией, равной нулю. При этом предполагается, что электрон находился на уровне Ферми. Таким образом, работа выхода в полупроводниках определяется электронным сродством материала и степенью его легирования, поскольку последняя задает расстояние между дном зоны проводимости и положением уровня Ферми. Обратите внимание, что в случае невырожденного полупроводника уровень Ферми расположен в запрещенной зоне, где электронов нет.

решетки будут напряженными из-за того, что состав материала в этом месте меняется скачком. Поскольку связь представляет собой взаимодействие электронных облаков и ядер атомов, то напряжённая химическая связь обуславливает наличие сильного электрического поля на расстояниях порядка межатомных, составляющих единицы ангстрем. Это сильное поле обуславливает разрывы дна зоны проводимости и потолка валентной зоны.

По аналогии с гомоструктурным p - n переходом, ширина областей пространственного заряда гетероперехода при отсутствии внешнего напряжения определяется соотношениями:

$$x_n = \sqrt{\frac{2N_d\epsilon_1\epsilon_2(\varphi_2 - \varphi_1)}{eN_a(\epsilon_2N_d + \epsilon_1N_a)}}, \quad x_p = \sqrt{\frac{2N_a\epsilon_1\epsilon_2(\varphi_2 - \varphi_1)}{eN_d(\epsilon_2N_d + \epsilon_1N_a)}}, \quad (1.36)$$

а емкость гетероперехода - соотношением:

$$C = S \sqrt{\frac{e\epsilon_1\epsilon_2N_aN_d}{2(\epsilon_1N_a + \epsilon_2N_d)(\varphi_1 - \varphi_2 - U)}}. \quad (1.37)$$

Аналогично гомоструктурному p - n переходу, ВАХ гетероструктурного p - n перехода имеет вид:

$$J = J_S \left(e^{\frac{eU}{AkT}} - 1 \right), \quad (1.38)$$

где A – безразмерный коэффициент, характеризующий величину разрыва зон гетероструктуры. Для гомоструктурного p - n перехода $A = 1$.

На рис. 1.9 хорошо видно, что высота барьера для электронов и дырок различна. Резкий подъем потолка валентной зоны образует большой гетеробарьер для дырок, чем барьер для электронов в зоне проводимости. В этом случае, отношение электронного и дырочного токов будет иметь вид:

$$\frac{J_{S_n}}{J_{S_p}} = \frac{n_p V_n}{p_n V_p} \exp\left(\frac{\Delta W_V}{kT}\right). \quad (1.39)$$

Часто $\Delta W_V \approx 10...20kT$, т.е. высота барьера для дырок существенно больше их средней энергии $\frac{3}{2}kT$. Поскольку распределение частиц по энергии носит экспоненциальный характер, то в итоге в подобных переходах электронный ток на несколько порядков превышает дырочный. Такое явление называют *униполярной инжекцией*.

Рассмотрим теперь *варизонный гетеропереход* (рис. 1.10), т.е. такой гетеропереход в котором химический состав полупроводникового материала изменяется плавно. Например, в структуре $AlAs-Al_yGa_{1-y}As-GaAs$ химический состав среднего слоя плавно меняется так, что показатель соотношения Ga и Al

в тройном соединении зависит от координаты $y = y(x)$. Это позволяет плавно изменять ширину запрещенной зоны среднего слоя структуры.

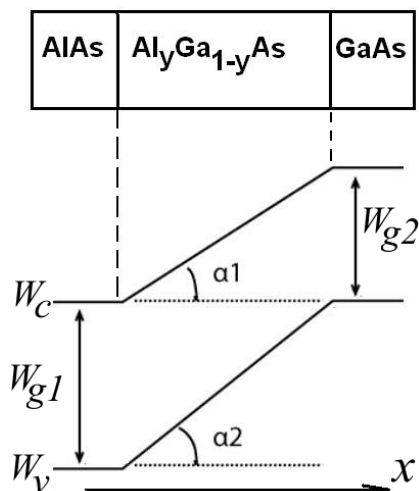


Рис. 1.10. Зонная диаграмма варизонного гетероперехода. В области, где ширина запрещенной зоны меняется, химический состав полупроводникового материала изменяется *плавно* в зависимости от координаты. Разные углы наклона дна зоны проводимости и потолка валентной зоны говорят о различных напряженностях электрических полей, действующих на электроны и дырки. Данный факт можно объяснить тем, что движение электронов в валентной зоне осуществляется «по химическим связям», где локально поля отличаются от тех, что действуют на свободные электроны

Важно, что углы наклона дна зоны проводимости и потолка валентной зоны различны, а значит, отличаются и величины напряженностей электрических полей, действующих на электроны и дырки. Этот эффект используется в некоторых приборах для увеличения скорости носителей заряда.

Анизотипный и изотипный варизонные переходы отличаются между собой типом зарядов, которые создают дополнительное электрическое поле в переходе. В первом случае поле создается ионами доноров с одной стороны и акцепторов — с другой, а во втором — ионами и свободными носителями заряда противоположного знака (либо донорами и электронами, либо акцепторами и дырками).

Вернемся теперь к резкому гетеропереходу, когда химический состав полупроводника сильно меняется на длинах порядка межатомного расстояния. Чтобы избежать дефектов кристаллической решетки, подберем химический состав слоев таким образом, чтобы постоянные кристаллических решеток совпадали, как, например, в гетеропереходе GaAs/Ga_{0,7}Al_{0,3}As.

На рис. 1.11 приведено распределение легирующей примеси и зонная диаграмма для *селективно легированного гетероперехода*. Такой гетеропереход специально изготавливается так, чтобы скачок в распределении концентрации доноров по координате не совпадал с границей гетероперехода, где реализуется разрыв зон.

В силу того, что потенциальная яма на границе подобного гетероперехода достаточно глубока и имеет размеры порядка длины волны электрона, происходит *квантование электронных уровней*. В такой ситуации энергия движения в поперечном слое ямы направлении (на рисунке 1.11 - по оси x) может измениться только на определенную величину, а значит, неупругое рассеяние по данной координате подавлено.

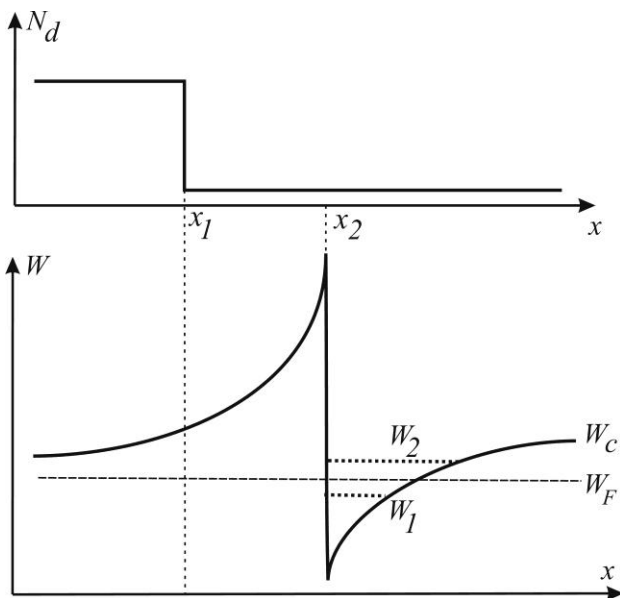


Рис. 1.11. Распределение легирующей примеси (вверху) и зонная диаграмма (внизу) для селективно легированного изотипного гетероперехода. Так как уровень Ферми расположен между двумя квантовыми уровнями W_1 и W_2 , то нижний уровень будет заселен электронами значительно больше, чем верхний [3]

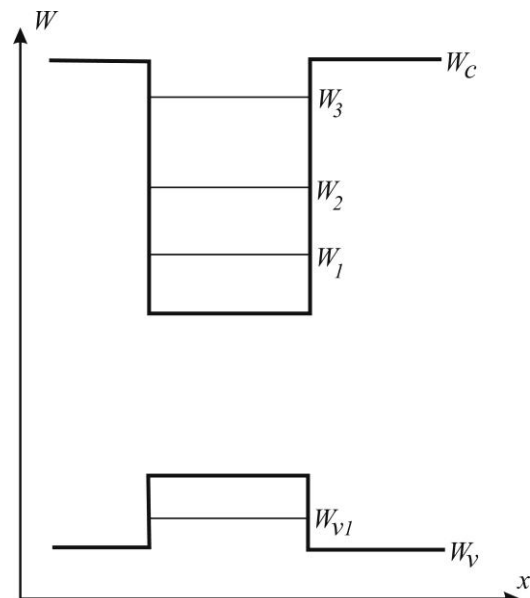


Рис. 1.12. Зонная диаграмма квантовой ямы, образованной двумя резкими гетеропереходами. Благодаря тому, что при изготовлении гетероперехода ширина среднего слоя может легко варьироваться, появляется возможность создания требуемой глубины квантовой ямы, т.е. регулировки количества и положения в ней энергетических уровней

Кроме того, поскольку в такой селективно легированной структуре область квантовой ямы изготовлена в материале с низким уровнем легирования, то рассеяние на ионах легирующей примеси практически отсутствует, а значит, подвижность электронов внутри квантовой ямы высока. Поэтому при движении в направлении, параллельном слоям ямы (по оси y или z) возможно достижение больших скоростей электронов, таких, как в нелегированном материале. Из-за того, что область сильно легированного материала изготавливается на расстояниях меньше диффузионной длины от квантовой ямы, концентрация электронов в яме будет высока. Подобный слой материала с высокой концентрацией носителей и одновременно с большой их подвижностью часто используется в качестве активной области полупроводниковых приборов, например, в полевых транзисторах с высокой подвижностью электронов (HEMT – high electron mobility transistor).

Другим вариантом образования квантовых ям является создание трёхслойной композиции из двух широкозонных и среднего узкозонного слоев полупроводника (рис. 1.12). Такая технология позволяет добиться реализации в одном полупроводниковом слое квантовой ямы как для электронов, так и для дырок. Важно, что для реализации среднего слоя необязательно подбирать материал с той же постоянной решетки, как у крайних слоев. Из-за того, что средний слой тонкий, при образовании гетероструктуры он растягивается, так что на границе раздела не возникают дефекты (псевдоморфный гетеропереход

[3]). За счет растяжения возможно варьирование параметров квантовой ямы. Подобные структуры используются для создания оптоэлектронных приборов.

Для треугольной ямы (рис.1.11) потенциал может быть аппроксимирован зависимостью $V(x) = eE_s x$ [3], тогда, решая уравнение Шредингера:

$$\frac{\hbar^2}{2m^*} \frac{d^2 \Psi_i}{dx^2} + (W_i - V(x)) \Psi_i = 0, \quad (1.40)$$

где \hbar – приведенная постоянная Планка, получаем распределение уровней в яме по энергии (уравнение Эйри):

$$W_i \approx \left(\frac{\hbar^2}{2m^*} \right)^{1/2} \left(\frac{3eE_s \pi(i + 3/4)}{2} \right)^{2/3}, \quad (1.41)$$

где для вычисления электрического поля E_s необходимо решить уравнение Пуассона. Для прямоугольной ямы в приближении ее бесконечной высоты положение уровней вычисляется по традиционной формуле:

$$W_n = \frac{n^2 \pi^2 \left(\frac{\hbar}{d} \right)^2}{2m^*}. \quad (1.42)$$

Важно понимать, что речь идет о квантовании только по одной координате (в нашем случае – по оси x). При этом по двум другим направлениям (в плоскости yz) электроны двигаются классически, поэтому такой электронный газ принято называть *двумерным*. Классическое движение частиц, как и в обычном, трехмерном, случае, описывается эффективной массой, подвижностью и скоростью насыщения, соответствующей объемному полупроводнику. Дисперсионное соотношение в двумерном электронном газе выглядит следующим образом⁵:

$$W = W_c + W_n + \frac{p_z^2 + p_y^2}{2m^*}, \quad (1.43)$$

где W_c – энергия дна зоны проводимости, W_n – положение уровней квантования по отношению к дну зоны проводимости.

Кроме потенциальных ям с помощью гетеропереходов могут быть созданы потенциальные барьеры для электронов и дырок (рис.1.13).

⁵ Предлагаем читателю подумать самостоятельно, в каких приближениях справедлива зависимость такого вида.

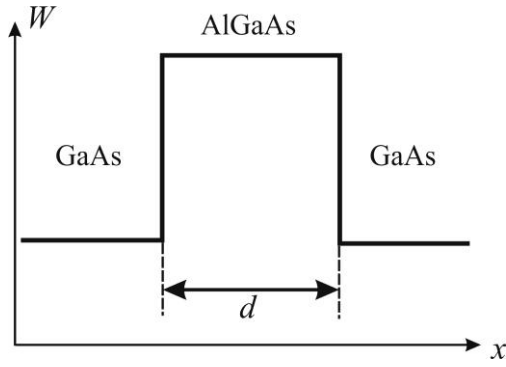


Рис. 1.13. Зонная диаграмма энергетического барьера, образованного двумя резкими гетеропереходами. Благодаря тому, что при изготовлении гетероперехода ширина среднего слоя легко варьируется, можно создавать барьеры с требуемым коэффициентом прохождения электронов за счет туннелирования

Коэффициент прохождения T через прямоугольный туннельно-прозрачный барьер высотой V_0 определяется толщиной барьера d , длиной волны туннелирующего электрона λ и его эффективной массой m^* :

$$T = \frac{1}{1 + \frac{(k_1^2 + k_2^2)^2}{4k_1^2 k_2^2} sh^2 k_2 d}, \quad (1.44)$$

где $k_1 = \frac{\sqrt{2m^*W}}{\hbar}$, $k_2 = \frac{\sqrt{2m^*(V_0 - W)}}{\hbar}$, \hbar – приведенная постоянная Планка.

Если $k_2 d \gg 1$, то $sh(k_2 d) \approx exp(k_2 d)/2$. Тогда выражение (1.44) примет вид

$$T \approx T_0 \exp\left(-2d \sqrt{\frac{2m^*(V_0 - W)}{\hbar^2}}\right), \quad (1.45)$$

где $T_0 = \left(\frac{k_1 k_2}{k_1^2 + k_2^2}\right)^2$.

1.5. СТРУКТУРА МЕТАЛЛ - ДИЭЛЕКТРИК - ПОЛУПРОВОДНИК (МДП)

1.5.1. Идеальная МДП-структура

Рассмотрим структуру металл-диэлектрик-полупроводник (МДП), которая схематически изображена на рис. 1.14.

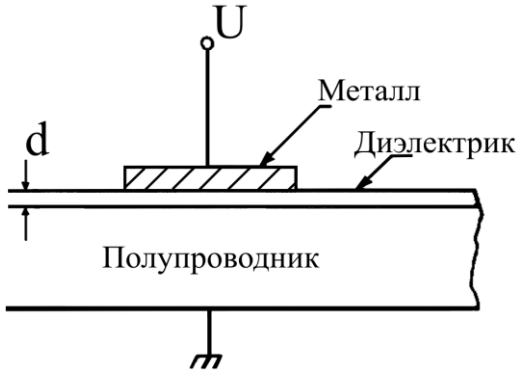


Рис. 1.14. Структура металл-диэлектрик-полупроводник (МДП-структура), где d — толщина слоя диэлектрика, а U — напряжение, приложенное к металлическому электроду

Зонные диаграммы идеальных МДП-структур при $U=0$ приведены на рис. 1.15. На рисунке введены следующие обозначения: $-e\phi_M$ — работа выхода из металла, χ — сродство к электрону полупроводника, W_g — ширина запрещенной зоны, $-e\psi_B$ — разность между уровнем Ферми W_F и серединой запрещенной зоны W_i , ϕ_b — потенциальный барьер между металлом и диэлектриком, χ_i — сродство к электрону для изолятора.

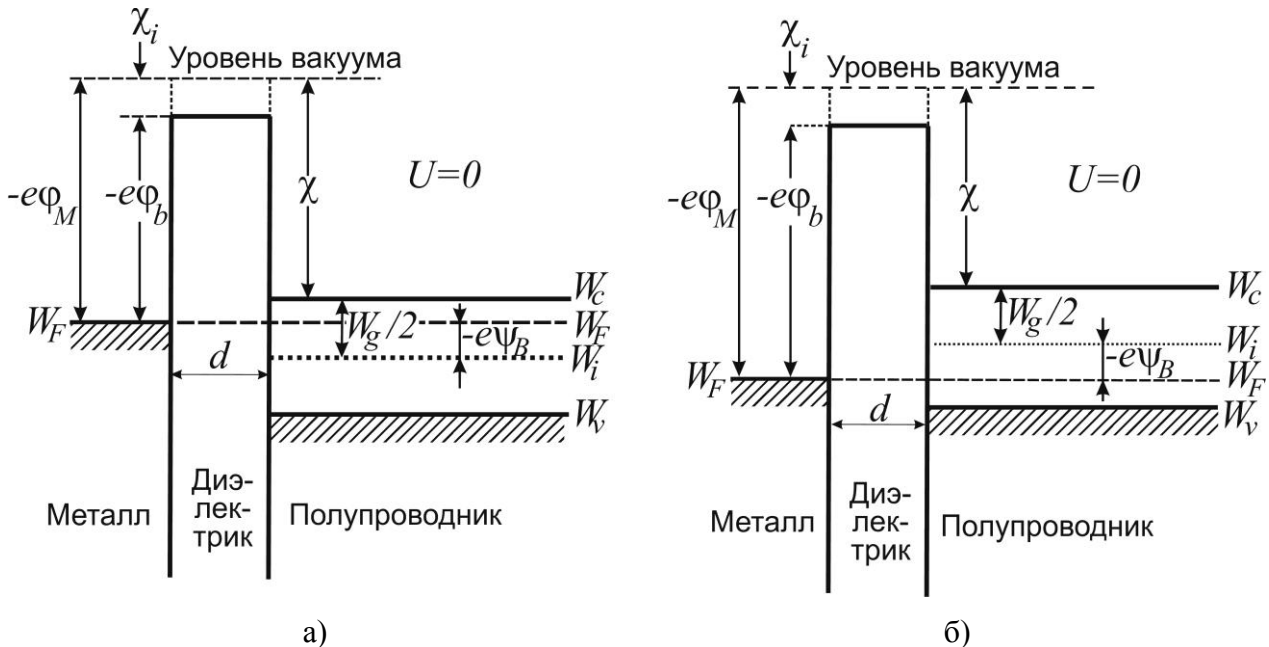


Рис. 1.15. Зонные диаграммы идеальных МДП-структур при $U = 0$: а) — полупроводник n -типа; б) — полупроводник p -типа

Понятие «идеальная МДП-структура» определим следующим образом:

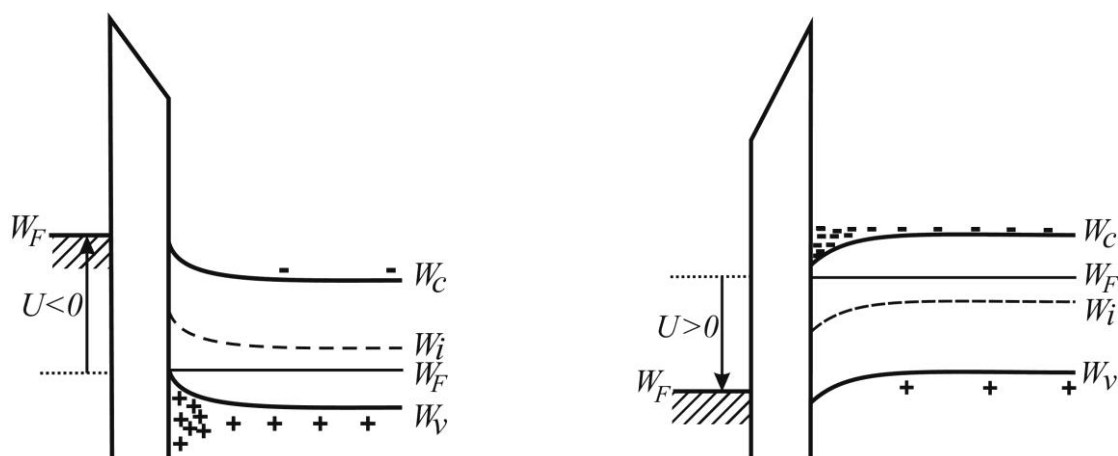
1. Работы выхода электронов из металла и полупроводника одинаковы, т.е. разность работ выхода из металла и полупроводника равна нулю. Это означает, что в отсутствие внешнего напряжения $U = 0$ энергетические зоны полупроводника не изогнуты (состояние плоских зон, см. рис. 1.15).
2. При любых напряжениях смещения в структуре могут существовать только заряд в ее полупроводниковой части и равный ему заряд противоположного знака на металлическом электроде, отделенном от полупроводника слоем диэлектрика.
3. При постоянном напряжении смещения отсутствует перенос носителей заряда через диэлектрик, т. е. сопротивление диэлектрика предполагается бесконечно большим.

Если к идеальной МДП-структуре приложить напряжение того или иного знака, то на полупроводниковой поверхности появится электрический заряд. При этом возможен один из трёх вариантов состояния равновесия. Рассмотрим их на примере МДП-структуры с полупроводником p -типа (рис. 1.16, левый столбец). Если к металлическому электроду структуры приложено отрицательное напряжение ($U < 0$), край валентной зоны у границы с диэлектриком изгибается вверх и приближается к уровню Ферми (рис. 1.16 *a*).

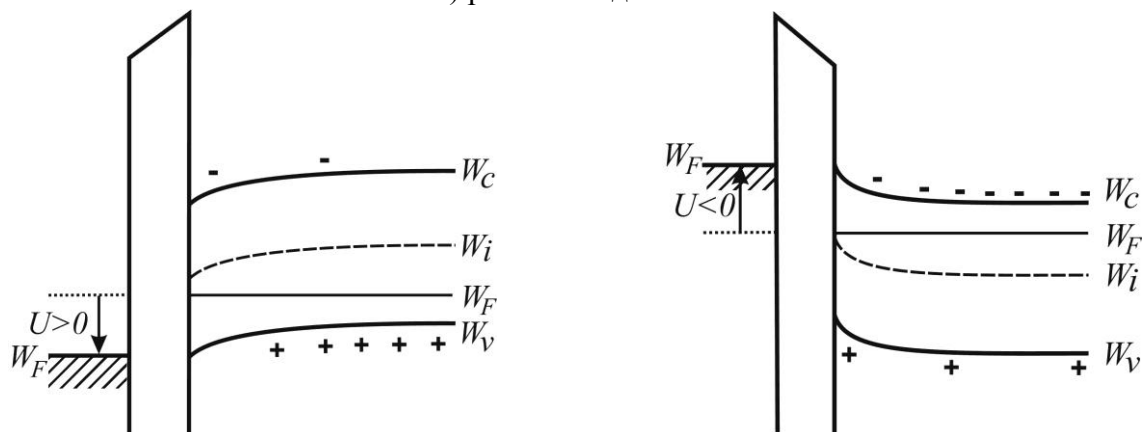
Поскольку в идеальной МДП-структуре сквозной ток равен нулю, уровень Ферми в полупроводнике остается постоянным. Так как концентрация дырок экспоненциально зависит от разности энергий ($W_F - W_V$), такой изгиб зон приводит к увеличению числа основных носителей (дырок) у поверхности полупроводника. Этот режим называется режимом обогащения (аккумуляции). Если к МДП-структуре приложено не слишком большое положительное напряжение ($U > 0$), зоны изгибаются в обратном направлении и приповерхностная область полупроводника обедняется основными носителями (рис. 1.16 *б*). Этот режим называют режимом обеднения или истощения поверхности. При больших положительных напряжениях зоны изгибаются вниз настолько сильно, что вблизи поверхности уровень Ферми пересекает собственный уровень W_i . В этом случае (рис. 1.16 *в*) концентрация неосновных носителей (электронов) у поверхности превосходит концентрацию основных носителей (дырок). Эта ситуация называется режимом инверсии.

Аналогичное рассмотрение можно провести и для МДП-структуры с полупроводником n -типа, при этом все указанные режимы будут осуществляться при напряжениях противоположной полярности (рис. 1.16, правый столбец).

а) режим обогащения



б) режим обеднения



в) режим инверсии

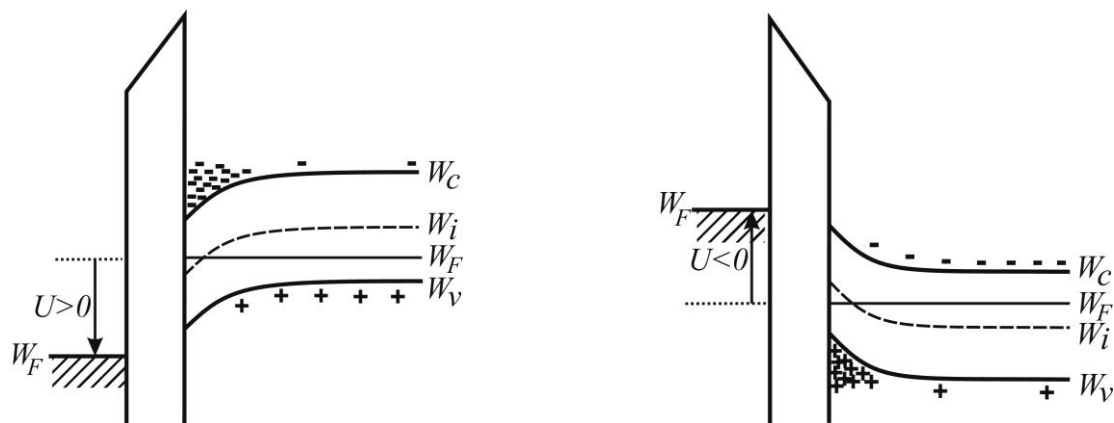


Рис. 1.16. Зонные диаграммы идеальных МДП – структур с полупроводниками р- (левый столбец) и n-типов (правый столбец) при подаче внешнего напряжения: а) режим аккумуляции; б) – режим обеднения; в) режим инверсии. Важно, что тангенс угла наклона дна зоны проводимости в слое диэлектрика и в прилегающем полупроводнике имеют одинаковый знак. Это объясняется тем, что нормальное к поверхностям раздела слоев электрическое поле создается зарядами, расположенными в полупроводнике и на поверхности металла (в идеальном диэлектрике зарядов нет). Поэтому вектор напряженности электрического поля имеет одно и то же направление в диэлектрике и прилегающем к нему слое полупроводника

1.5.2. Емкость МДП – структуры

Вольт-фарадная характеристика МДП-структуры приведена на рисунке 1.21:

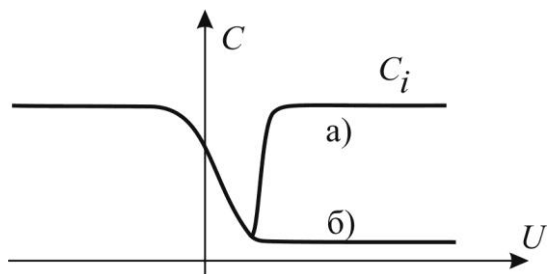


Рис. 1.17. Вольт-фарадная характеристика МДП-структуры на низкой частоте (а) и высокой частоте (б)

Поясним этот график на примере структуры металл – диэлектрик - полупроводник р-типа. При подаче значительного отрицательного потенциала на металл реализуется режим аккумуляции основных носителей заряда (дырок). Емкость в этом случае почти неизменна: $C_i \approx \epsilon \epsilon_0 S/d$, где d – ширина диэлектрика. При убывании отрицательного смещения вблизи поверхности образуется обедненная область, действующая как диэлектрик. Полная ёмкость при этом убывает. Далее ёмкость проходит через минимум и снова возрастает при образовании вблизи поверхности инверсионного слоя электронов при подаче положительного смещения. Отметим, что возрастание ёмкости в области положительных смещений зависит от способности электронов следовать за изменениями приложенного переменного сигнала. Это возможно лишь при низких частотах, когда скорость генерации - рекомбинации неосновных носителей достаточна для изменения заряда электронов в инверсионном слое в соответствии с изменением сигнала, на котором производятся измерения. В области же более высоких частот измерительного сигнала в правой части характеристики не наблюдается увеличения ёмкости (рис. 1.17 б).

1.6. КОНТАКТ МЕТАЛЛ - ПОЛУПРОВОДНИК

1.6.1. Зонная диаграмма

Чтобы показать, как формируется потенциальный барьер вблизи границы металла с полупроводником, имеющим другую работу выхода, предположим вначале, что материалы электрически нейтральны и изолированы друг от друга. На рис. 1.18 а представлена энергетическая зонная диаграмма для полупроводника n -типа, работа выхода из которого ($-e\varphi_n$) меньше, чем работа выхода из металла ($-e\varphi_m$). Именно этот случай приводит к возникновению потенциального

барьера в полупроводнике, называемого барьером Шоттки. Кроме того, будем считать, что заряженные поверхностные состояния⁶ отсутствуют.

Если металл и полупроводник теперь электрически соединить друг с другом, то часть электронов перейдет из полупроводника в металл. При этом уровни Ферми W_F в обоих материалах сравниваются - рис. 1.18 б. Вследствие такого перехода электронов в зазоре между металлом и полупроводником возникнет электрическое поле, и энергетические зоны у поверхности полупроводника изогнутся, как показано на рисунке 1.22 б. При дальнейшем уменьшении толщины вакуумного зазора энергетический барьер, образованный потенциалом этого зазора, становится туннельно-прозрачным для электронов, и им можно пренебречь (рис. 1.18 в).

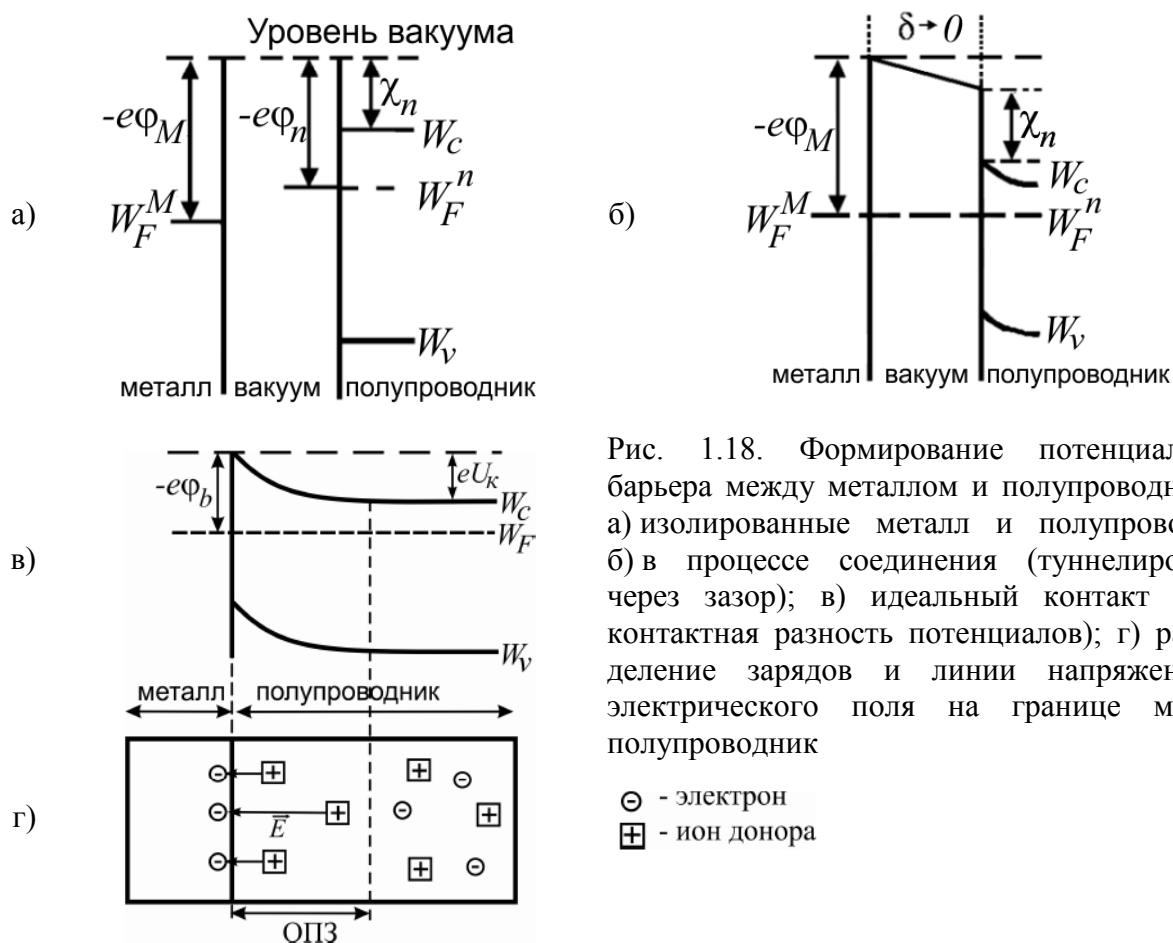


Рис. 1.18. Формирование потенциального барьера между металлом и полупроводником: а) изолированные металл и полупроводник; б) в процессе соединения (туннелирование через зазор); в) идеальный контакт (U_k – контактная разность потенциалов); г) распределение зарядов и линии напряженности электрического поля на границе металл-полупроводник

В действительности, идеализированная картина формирования барьера Шоттки (рис. 1.18 а-в) может нарушаться по целому ряду причин. Во-первых, из-за сохранения промежуточного слоя толщиной порядка нескольких межатомных расстояний (рис. 1.18 б). Во-вторых, из-за наличия поверхностных состояний в полупроводнике. Если плотность поверхностных состояний доста-

⁶ Очевидно, что на границе раздела полупроводника с какой-либо средой (в том числе, вакуумом) происходит нарушение периодичности потенциала кристаллической решетки. В связи с этим в данной области появляются электронные уровни в запрещенной зоне, которые принято называть *поверхностными состояниями*. Они могут захватывать электроны или дырки, приобретая, таким образом, отрицательный или положительный заряд.

точно велика, то заряд, связанный с ними, может эффективно экранировать полупроводник от электрического поля в промежуточном слое. Поэтому, величина заряда в обедненной области и высота барьера могут не зависеть от работы выхода металла. Третья причина, по которой в рис. 1.18 в. вносятся коррективы, связана с влиянием сил изображения (т.е. с тем, что электроны, приближающиеся к барьеру, заряжают поверхность раздела равным и противоположным по знаку зарядом, который притягивает электрон и, таким образом, изменяет профиль потенциального барьера, при этом снижая его высоту). Такое изменение формы барьера известно как эффект Шоттки.

Форма потенциального барьера зависит от распределения заряда в обедненной области. Потенциальная энергия W в области объёмного заряда удовлетворяет уравнению Пуассона:

$$\Delta W = \frac{e\rho}{\varepsilon\varepsilon_0}, \quad (1.46)$$

где ε - диэлектрическая постоянная, ρ - плотность электрического заряда. В одномерном случае, когда все величины зависят только от координаты x , отсчитываемой вглубь полупроводника, для идеального барьера получаем:

$$\frac{d^2W}{dx^2} = \frac{e^2}{\varepsilon\varepsilon_0} (N_d - n(x)), \quad (1.47)$$

где N_d и $n(x) = N_d \exp(-W(x)/kT)$ (распределение Больцмана) – соответственно, концентрации ионизированных доноров и свободных электронов. В объеме полупроводника при этом полный электрический заряд равен нулю в силу условия электронейтральности. Введя безразмерную потенциальную энергию

$w(x) = W(x)/kT$ и дебаевскую длину $l_D = \sqrt{\frac{\varepsilon\varepsilon_0 kT}{e^2 N_D}}$, преобразуем (1.47) к виду:

$$\frac{d^2w}{dx^2} = \frac{1}{l_D^2} (1 - e^{-w(x)}). \quad (1.48)$$

с граничными условиями $w(0) = \frac{-e(\varphi_m - \varphi_n)}{kT} = \frac{eU_k}{kT}$, $w(\infty) = 0$, $U_k > 0$ – контактная разность потенциалов. Для типичных параметров полупроводников при комнатной температуре дебаевская длина экранирования во много раз больше постоянной решетки.

При $w(0) \gg 1$, если пренебречь изгибом зон в переходной области, где концентрации электронов и доноров сравнимы (так называемое приближение полного обеднения), форма барьера может быть определена из уравнения (1.48) без экспоненциального слагаемого (учет его необходим лишь в области $w(x) \leq 1$):

$$w(x) \approx \frac{1}{2} \left(\frac{x}{l_D} - \sqrt{\frac{2eU_k}{kT}} \right)^2. \quad (1.49)$$

Получающийся барьер параболической формы известен как барьер Шоттки.

1.6.2. Теория процессов переноса зарядов

Перенос заряда через контакт металл-полупроводник осуществляется главным образом основными носителями в отличие от p - n переходов, где электрический ток обусловлен неосновными носителями. На рис. 1.19 показаны четыре основных транспортных процесса при прямом смещении: эмиссия электронов из полупроводника над барьером в металл, квантово-механическое туннелирование через барьер, рекомбинация в области пространственного заряда, рекомбинация в нейтральной области.

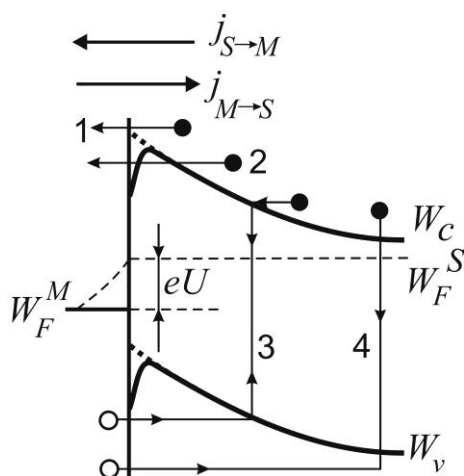


Рис. 1.19. Процессы токопереноса в контакте металл-полупроводник при прямом смещении. W_F^M , W_F^S – соответственно, квазиуровни Ферми в металле и полупроводнике; U – внешнее напряжение; $j_{S→M}$, $j_{M→S}$ – соответственно, плотности токов из полупроводника в металл и обратно. Стрелками обозначены следующие процессы:

1. эмиссия электронов из полупроводника над барьером в металл;
2. квантово-механическое туннелирование через барьер;
3. рекомбинация в области пространственного заряда;
4. рекомбинация в нейтральной области (инжекция дырок)

Кроме того, определенный вклад в полный ток может быть связан с токами утечки через ловушки на границе металл-полупроводник через периферийные области контакта, где в силу неоднородности возможны сильные краевые электрические поля. Хотя полностью исключить влияние указанных факторов, по-видимому, невозможно, тем не менее, современный уровень технологии позволяет изготовить диоды с барьером Шоттки, у которых механизм надбарьерного токопереноса является доминирующим, а их поведение вполне соответствует теоретическим представлениям.

Надбарьерное прохождение электронов из полупроводника в металл можно разделить на два этапа. Первый – дрейфово-диффузионный выход электронов из объема полупроводника к его поверхности. Второй – эмиссия электронов в металл. Оба процесса прохождения действуют последовательно, но, как правило, ток через контакт ограничивается одним из них. В соответствии с диффузионной теорией определяющим является первый процесс, согласно диодной (или термоэмиссионной) – второй [2, 4]. Ниже мы ограничимся рассмотрением основных положений теории термоэлектронной эмиссии, адекватно описывающей процессы токопереноса в полупроводниках с высокой подвижностью электронов, таких, например, как кремний или арсенид галлия, и

диффузионной теории. Познакомиться с более общей термоэмиссионной-диффузионной моделью токопереноса можно в [2, 7].

Транспортные свойства контакта металл/полупроводник определяются соотношением толщины области пространственного заряда и длины свободного пробега электронов (которая в свою очередь определяется процессами рассеяния на фононах, примесных центрах и других нарушениях кристаллической структуры). Если длина свободного пробега много больше толщины ОПЗ, можно воспользоваться так называемой диодной теорией выпрямления (Бете, 1942 г.) *в приближении термоэлектронной эмиссии.*

Предположим, что

- ✓ величина изгиба зон (т.е. высота барьера для электронов, движущихся из полупроводника в металл) $|eU_K| \gg kT$;
- ✓ область, определяющая термоэлектронную эмиссию, находится в термодинамическом равновесии (иными словами, положение квазиуровня Ферми в полупроводнике не изменяется вплоть до границы с металлом);
- ✓ протекание электрического тока не нарушает этого равновесия.

Ток через диод Шоттки представляет собой разность между током из металла в полупроводник и противоположным ему током, причем металл и полупроводник характеризуются каждый своим квазиуровнем Ферми. Величина тока в этом случае зависит только от высоты барьера и не зависит от его формы.

При приложении смещения U для электронов, покидающих полупроводник, высота барьера изменяется на eU , тогда как для электронов, движущихся в противоположном направлении, величина барьера меняется мало (только за счет эффекта Шоттки). Иными словами, концентрация электронов на границе раздела со стороны полупроводника растет как $e^{(eU/kT)}$ при увеличении напряжения U , а со стороны металла практически постоянна. Электроны на полупроводниковой стороне контакта находятся в термодинамическом равновесии с электронами в объеме полупроводника. Их концентрация выражается соотношением:

$$n(0) = n_0 \cdot e^{-\frac{e(U_K - U)}{kT}} = N_c \cdot e^{-\frac{W_c - W_F}{kT}} \cdot e^{-\frac{e(U_K - U)}{kT}} = N_c \cdot e^{-\frac{e(\phi_b - U)}{kT}}, \quad (1.50)$$

где n_0 – равновесная концентрация электронов в глубине полупроводника, ϕ_b – потенциал барьера для электронов, идущих из металла в полупроводник,

$N_c = 2 \left\{ \frac{2\pi m^* kT}{h^2} \right\}^{\frac{3}{2}}$ – эффективная плотность состояний в зоне проводимости.

Для полупроводников со сферическими изоэнергетическими поверхностями эти электроны имеют изотропное распределение по скоростям. Число электронов, падающих на единицу площади границы раздела в единицу времени, в

соответствии с элементарной кинетической теорией равно $n(0) \cdot \langle v \rangle / 4$, где $\langle v \rangle$ - средняя по абсолютной величине тепловая скорость электронов в полупроводнике. Тогда:

$$j_{M \rightarrow S} = \frac{eN_c \langle x \rangle}{4} e^{-\frac{e(\varphi_b - U)}{kT}}, \quad (1.51)$$

где $j_{M \rightarrow S}$ - плотность электрического тока из металла в полупроводник. Здесь индекс $M \rightarrow S$ соответствует направлению движения положительных зарядов (т.е. направлению вектора плотности тока). Поток электронов при этом направлен в противоположную сторону, т.е. из полупроводника в металл.

Плотность тока в обратном направлении не зависит от смещения (если пренебречь любой возможной полевой зависимостью φ_b). При нулевом смещении $j_{M \rightarrow S} = j_{S \rightarrow M}$, следовательно,

$$j_{S \rightarrow M} = \frac{eN_c \langle x \rangle}{4} e^{-\frac{e\varphi_b}{kT}}. \quad (1.52)$$

Учитывая, что полный ток $j = j_{M \rightarrow S} - j_{S \rightarrow M}$, получаем

$$j = \frac{eN_c \langle v \rangle}{4} e^{-\frac{e\varphi_b}{kT}} \left[e^{\frac{eU}{kT}} - 1 \right]. \quad (1.53)$$

При максвелловском распределении по скоростям:

$$\langle x \rangle = \sqrt{\frac{8kT}{\pi m^*}}. \quad (1.54)$$

Подставляя (1.54) и выражение для N_c в (1.53), получаем окончательное выражение для вольт-амперной характеристики диода Шоттки в модели термоэлектронной эмиссии:

$$j = j_s \left[e^{\frac{eU}{kT}} - 1 \right], \quad (1.55)$$

$$j_s = A^* T^2 e^{-\frac{e\varphi_b}{kT}}. \quad (1.56)$$

Здесь

$$A^* = 4\pi m^* e k^2 / h^3 \quad - \quad (1.58)$$

эффективная постоянная Ричардсона. Для напряжений смещения, больших $3kT/e$, единицей в квадратных скобках в (1.55) можно пренебречь, и тогда плотность тока будет пропорциональна $e^{(eU/kT)}$.

Реальная вольт-амперная характеристика имеет вид:

$$j = j_s \left[e^{\frac{eU}{nkT}} - 1 \right], \quad (1.59)$$

где $n > 1$ - почти постоянная величина. Обычно её называют фактором (или коэффициентом) неидеальности диода.

Отклонение ВАХ от идеальной при больших токах связано с падением напряжения смещения на последовательно включенном сопротивлении R (например, сопротивлении нейтральной области полупроводника R_s), которое всегда присутствует в реальных диодах. По этой причине фактическое падение напряжения на барьере Шоттки будет меньше напряжения на внешних выводах

диода. При этом плотность тока будет пропорциональна $(e^{\frac{e(U - JR)}{nkT}} - 1)$, где J - ток через диод.

Ёмкость диода Шоттки находится так же, как ёмкость резконеоднородного p-n перехода, и имеет вид:

$$C = S \sqrt{\frac{e\epsilon\epsilon_0 N_d}{2(U_k - U)}}. \quad (1.60)$$

Остановимся теперь на основах так называемой диффузионной теории выпрямления (Давыдов, Шоттки, Пекар, 1939 г.), применимой при длинах свободного пробега малых по сравнению с толщиной обеднённого слоя. В этом случае ток определяется выражением

$$j_{nx} = \sigma E_x + (-e) \cdot \left(-D \frac{dn(x)}{dx} \right), \quad (1.61)$$

где $\sigma = en\mu$ - проводимость, μ - подвижность носителей заряда, $j_{nx} = \text{const}$ - плотность электронного тока.

Кроме того, на поверхности $\varphi(0) = U_k - U$.

Из (1.61) с учетом соотношения Эйнштейна $\mu = eD/kT$ получим:

$$\frac{dn(x)}{dx} - \frac{e}{kT} \cdot \frac{d\varphi}{dx} \cdot n(x) - \frac{j_{nx}}{\mu kT} = 0. \quad (1.62)$$

Распределение электростатического потенциала $\varphi(x)$ связано, конечно, с распределением концентрации $n(x)$, однако, формально мы можем рассматривать (1.62) как линейное дифференциальное уравнение первого порядка для неизвестной функции $n(x)$, считая $\frac{d\varphi}{dx}$ заданной функцией x . Иначе говоря,

ограничимся линейной моделью. В этом случае общее решение уравнения (1.62) имеет вид

$$n(x) = n_0 e^{\frac{e\varphi(x)}{kT}} - \frac{j_{nx}}{\mu kT} \int_x^\infty e^{\frac{e}{kT}[\varphi(x)-\varphi(\xi)]} d\xi, \quad (1.63)$$

где $n_0 = n(x \rightarrow \infty) = \text{const.}$

На поверхности полупроводника (т.е. при $x=0$) концентрация электронов равна:

$$n(0) = n_0 e^{-\frac{eU_k}{kT}} \cdot e^{\frac{eU}{kT}} - \frac{j_{nx}}{\mu kT} \int_0^\infty e^{\frac{e}{kT}[\varphi(0)-\varphi(\xi)]} d\xi. \quad (1.64)$$

В рассматриваемом случае потоки электронов из металла в полупроводник и обратный (каждый в отдельности) много больше результирующего потока $(1/e)j_{nx}$. При этом на границе $x=0$ сохраняется практически равновесная концентрация электронов, т.е.

$$n(0) = n_0 e^{-\frac{eU_k}{kT}}. \quad (1.65)$$

Разрешая (1.64) относительно j_{nx} и используя (1.65), получим

$$j_{nx} = j_S \left[e^{\frac{eU}{kT}} - 1 \right], \quad (1.66)$$

где

$$j_S = \frac{n(0)\mu kT}{\int_0^\infty e^{\frac{e}{kT}[\varphi(0)-\varphi(\xi)]} d\xi}. \quad (1.67)$$

При вычислении интеграла в (1.67) ограничимся линейным членом в разложении $\varphi(\xi)$ в ряд по степеням ξ

$$\varphi(\xi) \approx \varphi(0) + \left(\frac{d\varphi}{d\xi} \right)_{\xi=0} \xi. \quad (1.68)$$

Тогда

$$\int_0^\infty e^{\frac{e}{kT}[\varphi(0)-\varphi(\xi)]} d\xi \approx \int_0^\infty e^{-\frac{e}{kT} \left(\frac{d\varphi}{d\xi} \right)_{\xi=0} \xi} d\xi = \frac{kT}{e} \frac{1}{\left(\frac{d\varphi}{dx} \right)_{x=0}}. \quad (1.69)$$

Подставляя (1.69) в (1.67), получим

$$j_S = en(0)\mu \left(\frac{d\varphi}{dx} \right)_{x=0}. \quad (1.70)$$

В (1.70) можно не учитывать слабую зависимость $\left(\frac{d\varphi}{dx} \right)_{x=0}$ от приложенного напряжения.

Таким образом, в случае диффузионной теории выпрямления ВАХ качественно имеет вид, изображенный на рис. 1.6, однако, при той же высоте потенциального барьера ток насыщения по величине значительно меньше, чем в модели термоэлектронной эмиссии.

1.6.3. Омический контакт

Омическим контактом называют контакт металл — полупроводник, сопротивление которого пренебрежимо мало по сравнению с объемным сопротивлением полупроводника. Один или несколько омических контактов присутствуют во всех без исключения полупроводниковых приборах. Важно, что хороший омический контакт не должен приводить к существенному изменению характеристик прибора, а падение напряжения на таком контакте при пропускании тока должно быть мало по сравнению с падением напряжения на активной области прибора.

Рассмотрим сначала удельное сопротивление контакта, определяемое как обратная величина от производной плотности тока по напряжению. Наиболее важной характеристикой контакта является сопротивление при нулевом смещении (величина, обратно пропорциональная тангенсу угла наклона ВАХ):

$$R_c = \left(\frac{dJ}{dU} \right)_{U=0}^{-1}. \quad (1.71)$$

Малые значения R_c достигаются при малой ширине барьера (высокой степени легирования) или его малой высоте. Именно из этих соображений исходят при изготовлении омических контактов (рис. 1.20 а, б).

Для широкозонных полупроводников трудно изготовить контакт с малой высотой барьера. Кроме того, используемые металлы не всегда имеют достаточно малую работу выхода. В таких случаях для изготовления омических контактов создают дополнительный высоколегированный подслой на поверхности полупроводника (рис. 1.20 б). Использование такого подслоя позволяет уменьшить ширину барьера до толщины, когда протекание тока в основном происходит за счет туннелирования «сквозь» барьер. Таким образом, если в некотором твердотельном приборе контакт между металлом и полупроводником должен быть невыпрямляющим (т.е. он не должен влиять на характеристики устройства), то под слоем металла обязательно содержится слой высоколегированного полупроводника, формирующий омический контакт. Обратите

внимание, что на зонной диаграмме омические контакты, как правило, не отображаются – там приводится лишь активная область прибора.

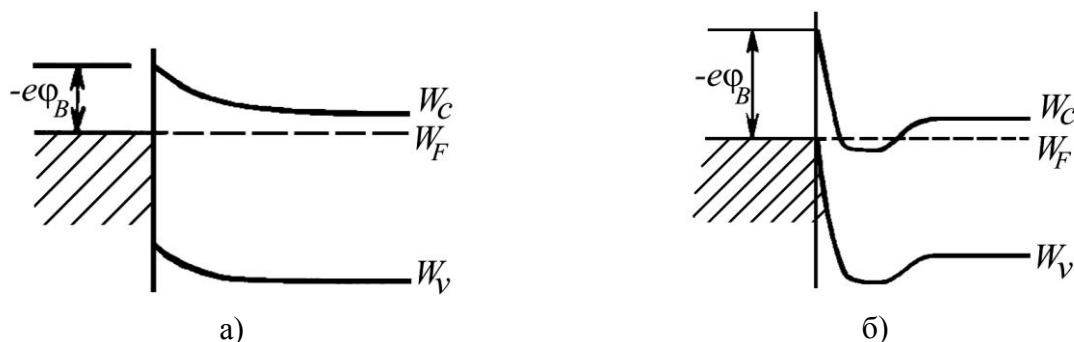


Рис. 1.20. Омические контакты с малой высотой барьера (а) и высокой степенью легирования (б)

1.7. ЭКВИВАЛЕНТНЫЕ СХЕМЫ ДИОДОВ

Для расчета параметров полупроводниковых приборов часто используют метод эквивалентной схемы, который заключается в замене единого устройства набором стандартных элементов. В случае пассивных приборов для такой замены используются конденсаторы, сопротивления и индуктивности. В усилительных устройствах, например, транзисторах, в эквивалентную схему необходимо включать генераторы тока и/или напряжения. Цель подобной замены – сведение задачи о расчете характеристик полупроводниковых элементов к известной задаче анализа токов в радиотехнической цепи с помощью уравнений Кирхгофа. Вычисление сопротивлений полупроводниковых слоев обычно производят на основе известной проводимости этих слоев, т.е. исходя из концентрации носителей заряда и их подвижности.

Выше мы уже отмечали, что в ОПЗ полупроводника есть электрическое поле и содержится мало свободных носителей заряда, поэтому такая область отчасти похожа на слой диэлектрика в конденсаторе. Отличие состоит в том, что через эту область течет ток, что учитывается в эквивалентной схеме с помощью сопротивления, включенного параллельно конденсатору.

Общий вид эквивалентных схем диодов на основе $n^+ - n$ перехода, $p - n$ перехода, барьера Шоттки и МДП структуры изображен на рис. 1.21.

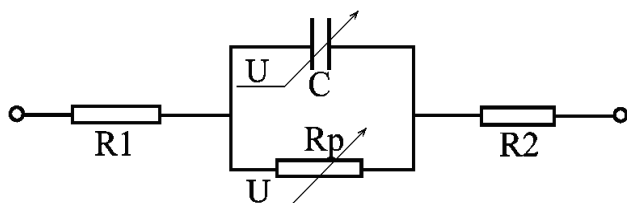


Рис. 1.21. Эквивалентная схема диодов на основе $p - n$ перехода, барьера Шоттки и МДП- структуры

Здесь C – управляемая напряжением емкость перехода; U – величина напряжения, приложенного непосредственно к переходу (т.е. без учета падения напряжения на сопротивлениях R_1 и R_2); R_1 , R_2 – последовательные сопротивления

областей полупроводникового материала по обе стороны от перехода, включая сопротивления металлических контактов; R_p – управляемое напряжением сопротивление перехода или барьера Шоттки (определяется из ВАХ диодов).

Обычно одно из последовательных сопротивлений R_1 , R_2 значительно меньше другого, например, в p^+n переходе или контакте металл-полупроводник. В такой ситуации меньшим сопротивлением пренебрегают. Большее сопротивление, по сути, является сопротивлением базы диода.

При обратном смещении сопротивление перехода R_p увеличивается, так что сопротивлениями R_1 и R_2 часто можно пренебречь. При небольшом прямом смещении, когда внешнее напряжение сравнимо с контактной разностью потенциалов, следует учитывать все сопротивления, а при больших прямых смещениях обычно существенным является только сопротивление базы.

Из-за того, что при больших прямых смещениях емкость диода шунтирована малым сопротивлением перехода R_p , ее можно не учитывать. При прямых напряжениях сравнимых с контактной разностью потенциалов обычно учитывают диффузионную емкость, а при обратных напряжениях – барьерную.

В случае МДП-структуры на эквивалентной схеме остаются лишь два элемента C и R_2 , включенные последовательно. Последнее означает, что постоянный ток через диэлектрик не протекает.

ЧАСТЬ 2

ОПТОЭЛЕКТРОННЫЕ ДИОДЫ

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

2.1. Фотодиоды

Объяснение принципов работы фотодиодов удобно начать с рассмотрения наиболее простого фотодетектора – *фоторезистора*. Он представляет собой пластину полупроводника с омическими контактами, присоединенными к противоположным концам пластины, как показано на рис. 2.1. Принцип работы фотоприемников (в том числе, и фоторезисторов) основывается на явлении *внутреннего фотоэффекта* или *фотопроводимости*: при падении излучения с энергией квантов больше ширины запрещенной зоны на поверхность полупроводника в нем генерируются электронно-дырочные пары, т.е. увеличивается число свободных носителей заряда (собственное возбуждение). Возможно также использование переходов с участием уровней, лежащих в запрещенной зоне (примесное возбуждение). В обоих случаях в соответствии с формулой $\sigma = e(\mu_n n + \mu_p p)$, где n, p – концентрации электронов и дырок, μ_n, μ_p – их подвижности, e – заряд электрона, облучение приводит к увеличению проводимости материала, т.е. появлению фототока, измерение которого позволяет детектировать падающее излучение.

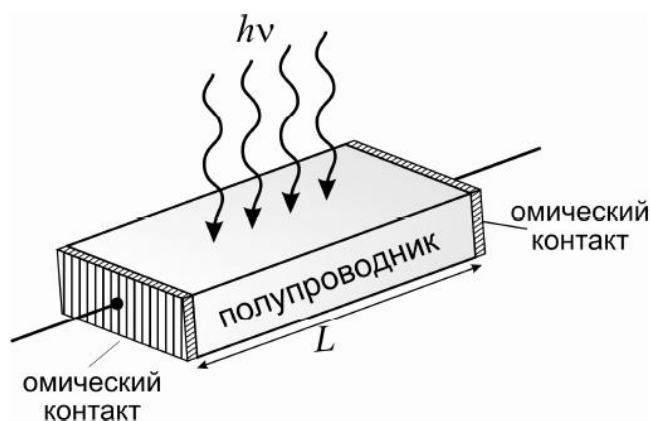


Рис. 2.1. Схематическое изображение фоторезистора

Существенным недостатком фотосопротивлений является значительная величина так называемого темнового тока, т.е. тока, который протекает в структуре в отсутствие излучения. Поэтому на практике в качестве фотоприем-

ников обычно используются фотодиоды, в которых данная проблема снимается. Семейство фотодиодов с обедненным слоем включает в себя p - n переходы, p - i - n диоды, диоды, образованные контактом металл - полупроводник (диоды с барьером Шоттки), гетеропереходы и точечно-контактные диоды.

В простейшем случае фотодиод представляет собой p - n переход, к которому приложено небольшое (по сравнению с напряжением пробоя) обратное смещение⁷. Как известно, ток обратносмещенного p - n перехода определяется неосновными носителями заряда, а, значит, в отсутствие излучения ток фотодиода мал. Поскольку толщина ОПЗ p - n перехода достаточно мала, то на практике обычно светят в торец прибора, как показано на рис. 2.2 а. Фотоны поглощаются в области пространственного заряда (ОПЗ) диода и в нейтральном объеме полупроводника. Электронно-дырочные пары, фотогенерированные в ОПЗ перехода, немедленно разделяются его полем и дают вклад в суммарный ток, что позволяет детектировать излучение (рис. 2.2 б). При этом носители, которые генерируются вне ОПЗ, попадают туда за счет диффузии. Область диффузии в этом случае должна быть достаточно тонкой (много меньше диффузионной длины), чтобы основная часть генерированных электронно-дырочных пар без рекомбинации достигала области перехода.

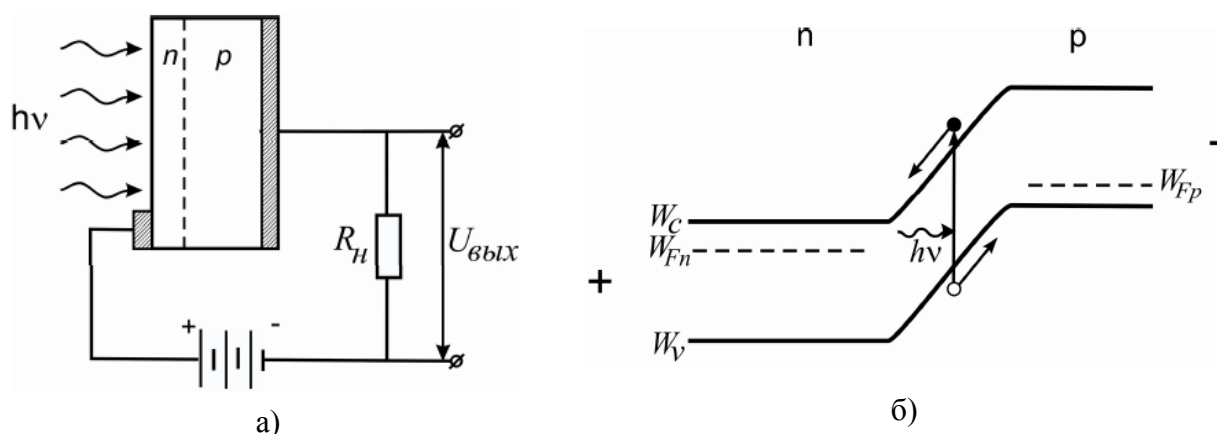


Рис. 2.2. Схематическое изображение фотодиода с подключением (а) и зонной диаграммы прибора при рабочем смещении (б). R_n – сопротивление нагрузки; $U_{вых}$ – напряжение выхода

Одной из основных характеристик любого фотодетектора, в том числе, и фотодиода, является *область спектральной чувствительности* – диапазон длин волн, в котором может быть генерирован заметный фототок. Одним из основных факторов, определяющих эту область, является коэффициент поглощения α . На рис. 2.3 представлены экспериментальные кривые зависимости коэффициента поглощения от длины волны в собственной области для германия и кремния [3]. Кроме того, указаны длины волн, соответствующие излучению некоторых распространенных лазеров.

⁷ Отметим, что в случае так называемых лавинных фотодиодов прикладывается высокое напряжение. Такие приборы будут рассматриваться ниже.

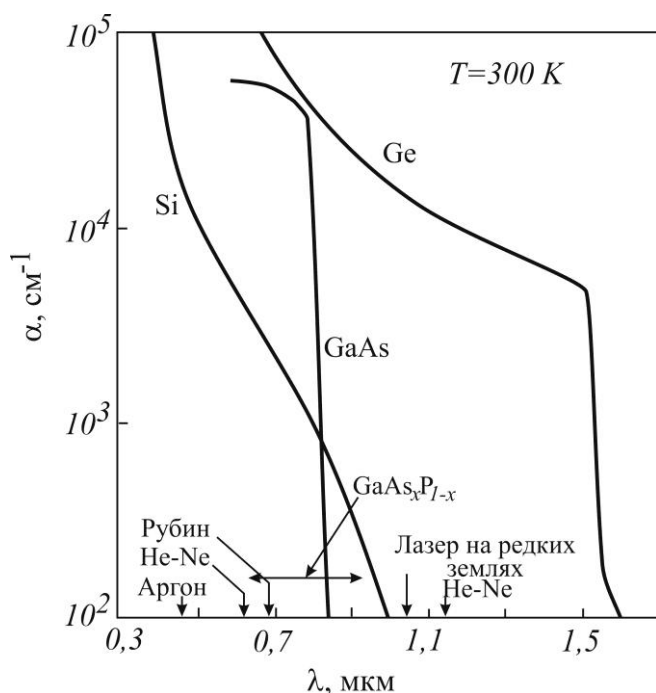


Рис. 2.3. Зависимость коэффициента поглощения от длины волны для Ge и Si при 300 K [3]. Указаны линии излучения некоторых лазеров

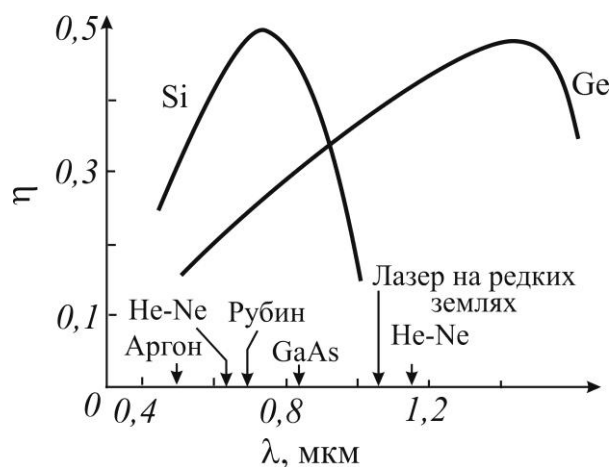


Рис. 2.4. Зависимость эффективного квантового выхода η (количества электронно-дырочных пар на один фотон) от длины волны λ для Ge и Si фоторезисторов [3]

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

$$\lambda_c = \frac{hc}{W_g} = \frac{1.24}{W_g[\text{эВ}]}, \text{ мкм.} \quad (2.1)$$

Здесь λ_c - длина волны, соответствующая ширине запрещенной зоны полупроводника W_g , c - скорость света, h - постоянная Планка. Излучение с длиной волны $\lambda < \lambda_c$ поглощается в полупроводнике с образованием пар электрон-дырка. При $\lambda > \lambda_c$ сравнительно небольшая по величине фотопроводимость может возникать в случае примесного поглощения фотона, т.е. при переходе электрона с уровня в запрещенной зоне в зону проводимости (или дырки в валентную зону). Коротковолновая граница спектральной чувствительности обусловлена тем, что в этой области длин волн коэффициент α очень велик (около 10^5 см^{-1}), т.е. излучение поглощается очень близко к поверхности, где времена рекомбинации малы из-за наличия дефектов. Вследствие этого фотоносители рекомбинируют раньше, чем уходят из поверхностного слоя к области пространственного заряда. На рис. 2.4 представлены типичные кривые

зависимости квантовой эффективности от длины волны для кремниевого и германиевого высокочастотных фотодиодов [3].

Еще одним важным параметром фотодетектора является его время отклика. Быстродействие фотодиодов определяется тремя факторами: временем диффузии носителей от места генерации к области пространственного заряда, временем дрейфа через обедненный слой и его емкостью. Чтобы свести к минимуму влияние диффузии, переход необходимо формировать очень близко к поверхности. Далее, чтобы основная часть излучения поглощалась непосредственно в ОПЗ, а свет не проникал бы вглубь за ее пределы, толщина области перехода должна быть порядка $1/\alpha$. При этом носители будут дрейфовать в ОПЗ со скоростями, ограниченными их рассеянием. Однако, область пространственного заряда не должна быть слишком широкой, так как времена переноса будут ограничивать предельную частоту работы прибора. С другой стороны, она не должна быть и слишком узкой, так как в этом случае становится большой емкость p - n перехода, а значит, и постоянная времени RC , где R - сопротивление базы диода. Оптимальным значением ширины области пространственного заряда считается такое, при котором время переноса примерно равно половине периода модуляции излучения. Например, для частоты модуляции, равной 10 ГГц , оптимальная ширина области пространственного заряда в кремнии (при скорости дрейфа порядка 10^7 см/сек) составляет около 5 мкм .

Типичным детектором с обедненным слоем является фотодиод со структурой p - i - n , где между p и n слоями заключена i -область, т.е. слой собственного полупроводника. Структура и зонная диаграмма такого фотодиода [3] показаны на рис. 2.5.

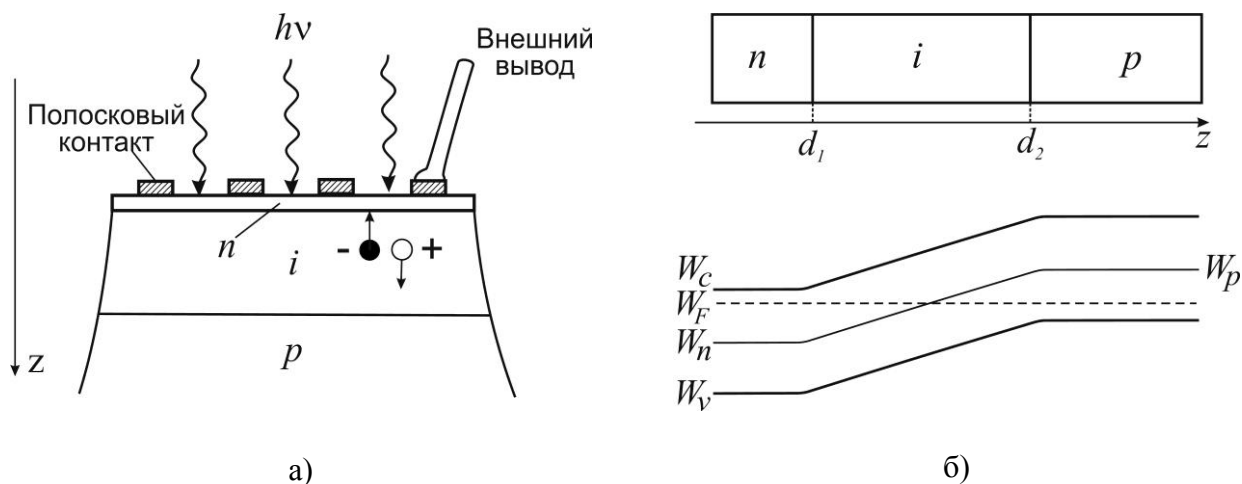


Рис. 2.5. Фотодиод с p - i - n структурой [3]: а) конструкция прибора; б) упрощенная структура и зонная диаграмма вдоль оси z в равновесном случае. Полосковые контакты объединены друг с другом – обычно это тонкая металлическая сетка на поверхности полупроводника

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

$$\begin{cases} \frac{d^2 W}{dz^2} = \frac{e^2}{\varepsilon \varepsilon_0} N_D, & d_1 - d_n < z < d_1; \\ \frac{d^2 W}{dz^2} = 0, & d_1 \leq z \leq d_2; \\ \frac{d^2 W}{dz^2} = -\frac{e^2}{\varepsilon \varepsilon_0} N_A, & d_2 < z < d_2 + d_p. \end{cases} \quad (2.2)$$

Здесь e – заряд электрона; N_D, N_A – соответственно, концентрации доноров в n и акцепторов в p областях; ε – диэлектрическая проницаемость материала; d_1, d_2 – координаты границ области собственной проводимости (рис. 2.5 б); d_n, d_p – толщины областей пространственного заряда в n и p областях, соответственно, причем, из закона сохранения заряда следует, что $N_D \cdot d_n = N_A \cdot d_p$. Если отсчитывать энергию от ее величины в n -области, то решение (2.2) будет иметь вид:

$$W(z) = \begin{cases} \frac{e^2 N_D}{2 \varepsilon \varepsilon_0} (z - d_1 + d_n)^2, & d_1 - d_n < z < d_1; \\ \frac{e^2 N_D}{2 \varepsilon \varepsilon_0} d_n^2 + \frac{e^2 N_D}{\varepsilon \varepsilon_0} d_n (z - d_1), & d_1 \leq z \leq d_2; \\ W_K - \frac{e^2 N_A}{2 \varepsilon \varepsilon_0} (d_2 + d_p - z)^2, & d_2 < z < d_2 + d_p. \end{cases} \quad (2.3)$$

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

$$W(z) = \begin{cases} 0, & z < d_1; \\ \frac{W_K}{d_2 - d_1} (z - d_1), & d_1 \leq z \leq d_2; \\ W_K, & d_2 < z. \end{cases} \quad (2.4)$$

Таким образом, мы получили линейный закон изменения потенциала между двумя легированными областями n и p -типа. То есть подобная система ведет себя, как простой конденсатор, при этом в области собственной проводимости постоянное электрическое поле определяется простым соотношением $E = eU_K / (d_2 - d_1)$.

P - i - n диоды характеризуются быстрым фотооткликом и достаточно высокой чувствительностью. Это объясняется тем, что емкость такой структуры значительно меньше емкости p - n перехода, а значит, меньше и постоянная времени RC . Кроме того, толщина i -слоя может быть сделана достаточно

большой для того, чтобы поглотить практически все излучение, т.е. поглощение в нейтральных областях сводится к минимуму.

Фотодиоды с барьером Шоттки также могут быть использованы в качестве высокоэффективных фотодетекторов [3]. Типичный пример приведен на рисунке 2.6.

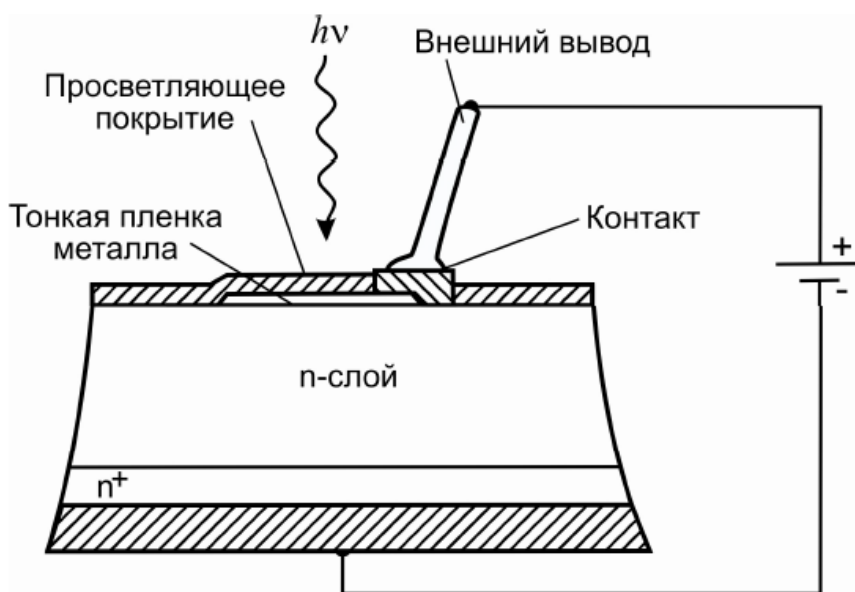


Рис. 2.6. Диод с барьером Шоттки и с просветляющим покрытием [3]

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

Фотодиоды с обедненным слоем изготавливаются на основе не только гомо-, но и гетероструктур, в которых переход образован между двумя полупроводниками с различной шириной запрещенной зоны. На рис. 2.7 схематически показан фотодиод, содержащий переход между n -GaAs и p -Ge, предназначенный для высокочастотного детектирования излучения с длиной волны $\lambda = 8450 \text{ \AA}$, испускаемого источником инфракрасного излучения, изготовленным из GaAs [3]. На этой длине волны коэффициент поглощения излучения равен 10 см^{-1} в GaAs и $3.2 \cdot 10^4 \text{ см}^{-1}$ в Ge. Таким образом, *основная доля излучения поглощается в германии в слое толщиной 1 мкм, т.е. там, где расположена обедненная область перехода*. В связи с тем, что носители создаются около перехода, а не близ поверхности, диоды имеют высокое быстродействие и нечувствительны к состоянию поверхности. Граничная частота диода слабо зависит от расстояния от поверхности до перехода, но более сильно зависит от того, какая часть излучения поглощается в той и другой частях обедненной области, образующей переход. Энергетический барьер на зонной диаграмме определяется разрывом дна зоны проводимости на границе раздела между GaAs и Ge. Преодолеть этот барьер носители могут либо за счет туннельного механизма, либо за счет надбарьерной эмиссии, имея достаточно большую энергию.

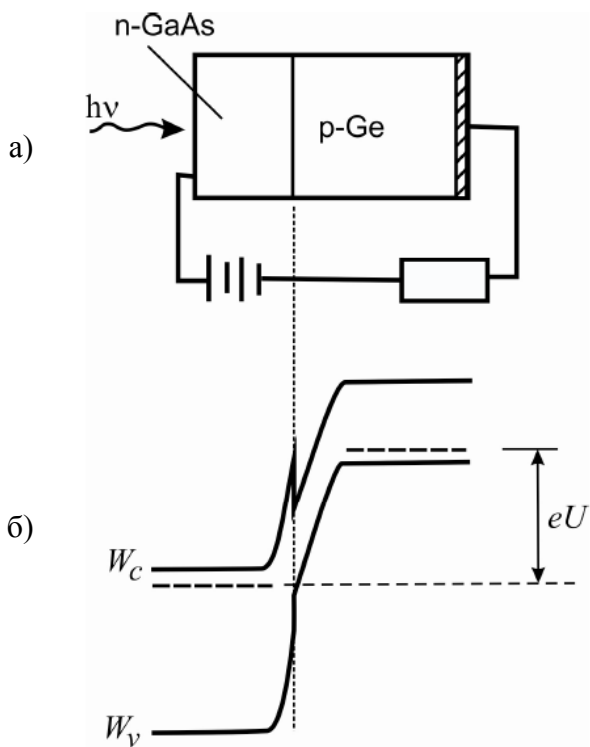


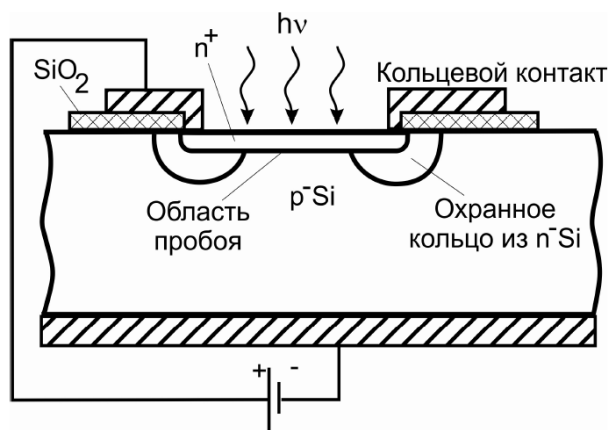
Рис. 2.7. Схематическое изображение фотодиода с гетеропереходом при приложении обратного смещения (а) и диаграмма энергетических зон диода (б)

Еще одним примером полупроводникового фотоприемника является *лавинный фотодиод*, т.е. диод, на который подаются большие (достаточные для реализации пробоя) обратные смещения. В такой ситуации фотовозбужденные электроны и дырки участвуют в формировании лавины свободных носителей заряда за счет процесса ударной ионизации полупроводника в сильном электрическом поле. Таким образом, с одного поглощенного фотона в лавинном фотодиоде образуется много свободных носителей за счет процессов вторичной генерации, т.е. прибор является фотодетектором с внутренним усилением тока. Минусом таких приборов является то, что лавинный процесс вносит дополнительные источники шума. Кроме того, появляются повышенные требования к термостабильности (лавинный пробой не должен переходить в термический).

Основным требованием для получения высокого усиления по току является исключение микроплазм или участков малой площади, пробивное напряжение которых ниже, чем пробивное напряжение всего перехода. Кроме того, для получения однородного по площади умножения необходимо исключить пробой у краев перехода. На рис. 2.8 показана типичная структура лавинного диода с охранным кольцом [3]. Градиент примеси в охранным кольце ниже, а радиус кривизны больше, поэтому центральная активная область пробивается раньше, чем периферийная (область охранным кольца).

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

Рис. 2.8. Поперечное сечение планарного фотодиода с охранным кольцом



2.2. СОЛНЕЧНЫЕ БАТАРЕИ

Солнечная батарея (фотоэлемент) представляет собой p - n переход большой площади, в котором энергия фотонов непосредственно преобразуется в электрическую энергию. Высокоэффективные солнечные батареи были впервые созданы в 1954 г. При этом были использованы диффузионные кремниевые p - n переходы. В настоящее время солнечные батареи изготавливаются из различных полупроводниковых материалов (Si, GaAs, CdS). Подобные приборы являются наиболее важными источниками питания на спутниках и космических кораблях.

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

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

Солнечная батарея, показанная на рис. 2.9, состоит из неглубокого p - n перехода, верхнего полоскового омического контакта и нижнего омического контакта, который полностью покрывает всю тыльную поверхность полупро-

водникового кристалла. Простейшая эквивалентная схема солнечной батареи при ее освещении показана на врезке рис. 2.10.

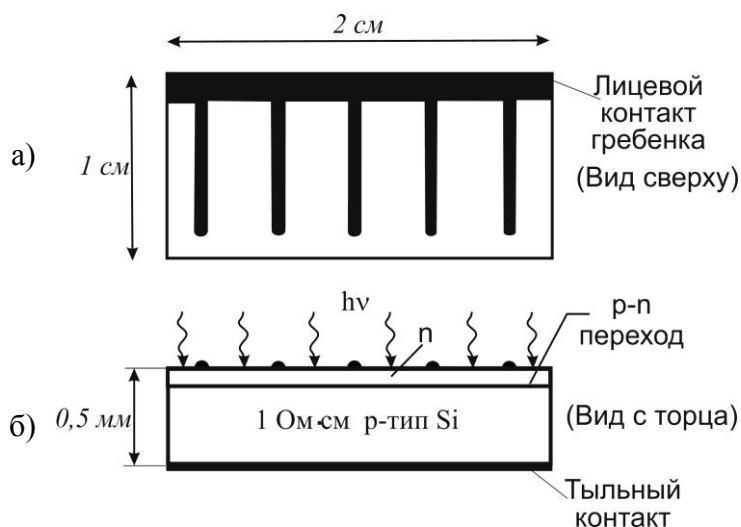


Рис. 2.9. Пример конструкции солнечной батареи: а) вид сверху; б) вид с торца

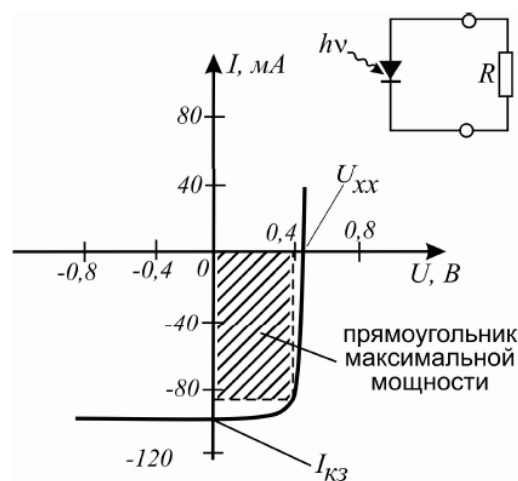


Рис. 2.10. ВАХ солнечной батареи с сопротивлением нагрузки R . На рисунке отмечены напряжение холостого хода U_{xx} и ток короткого замыкания $I_{кз}$

Вольт-амперная характеристика такого прибора может быть представлена в виде

$$I = I_s (e^{eU/kT} - 1) - I_\phi, \quad (2.5)$$

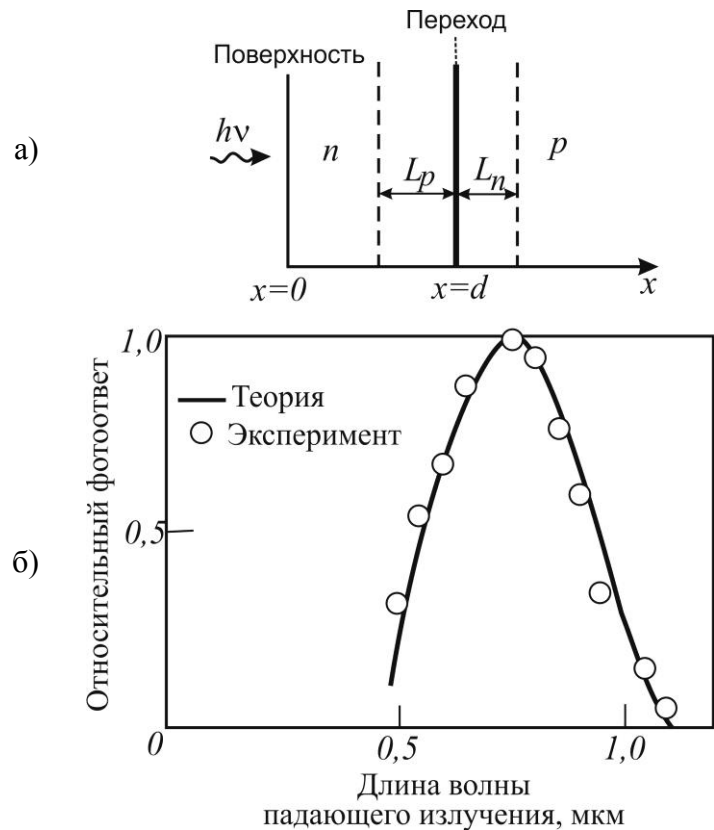
$$I_s = en_i^2 \left[\frac{1}{N_A} \left(\frac{D_n}{\tau_n} \right)^{1/2} + \frac{1}{N_D} \left(\frac{D_p}{\tau_p} \right)^{1/2} \right] \quad (2.6)$$

Здесь, помимо введенных выше стандартных обозначений, I_ϕ – ток, протекающий в структуре за счет генерации электронно-дырочных пар при освещении, т.е. ток в режиме короткого замыкания (сопротивление нагрузки $R=0$). На рис. 2.10 изображена ВАХ, соответствующая значениям $I_\phi=100$ мА, $I_s=10^{-9}$ А и $T=300$ К [3]. Кривая проходит через четвертый квадрант, следовательно, прибор может отдавать мощность во внешнюю цепь. ВАХ позволяет выбирать оптимальный режим работы фотоэлемента, т.е. оптимальное значение сопротивления нагрузки, которому соответствует наибольшая площадь прямоугольника с вершиной, расположенной на ВАХ, при данном значении светового потока Φ (фотонов / с·см²). КПД солнечных батарей, изготовленных из Si и GaAs, составляет 10-15%, а на основе тройных соединений с использованием гетеропереходов 20-50%. Основной вклад в уменьшение КПД дают потери на отражение, утечки и влияние сопротивления пассивных слоев полупроводника.

Для оценки величины тока короткого $I_{кз}$ замыкания и анализа спектрального распределения фоточувствительности солнечной батареи (т.е. зависимости

$I_{кз}(\lambda)$), удобно воспользоваться простой одномерной моделью, изображенной на рис. 2.11 а [3].

Рис. 2.11. Спектральная чувствительность тока солнечной батареи: а) - одномерное изображение солнечной батареи с глубиной залегания перехода d и диффузионной длиной неосновных носителей L_n и L_p для p - и n -областей, соответственно; б) сравнение типичных экспериментальных и расчетных кривых фотоответа для p - на n -солнечной батареи. Экспериментальные точки соответствуют данным, полученным при $d=2.0$ мкм. Сплошная кривая представляет собой расчетные результаты, полученные для $d=2.0$ мкм, $L_n=0.5$ мкм и $L_p=10.0$ мкм [3]



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

Если энергия квантов излучения превышает ширину запрещенной зоны полупроводника ($h\nu > W_g$), то за счет поглощения плотность потока фотонов Φ в материале изменяется в направлении x вглубь образца: $\Phi(x) = \Phi_0 \cdot \exp(-\alpha x)$, где Φ_0 – плотность падающего потока фотонов, α – коэффициент поглощения, зависящий от длины волны излучения (рис. 2.3). Тогда скорость поглощения фотонов в материале составит $\alpha \Phi_0 \cdot \exp(-\alpha x)$. Если при поглощении одного фотона рождается одна электронно-дырочная пара, то скорость их генерации (т.е. количество пар носителей, генерируемых в единице объема полупроводника за единицу времени) выражается формулой

$$G(x) = \Phi_0 \alpha e^{-\alpha x}. \quad (2.7)$$

Генерированные вне области перехода носители движутся к нему только за счет диффузии, при этом часть носителей рекомбинирует, не успевая добраться до области сильного поля. В n -области из общего числа неосновных носителей (дырок), генерированных на расстоянии x от поверхности, перехода достигает часть носителей, пропорциональная $\exp\left[-\frac{(d-x)}{L_p}\right]$. Тогда полное

число неосновных носителей, попадающих на единицу площади сечения перехода за единицу времени и появившихся за счет генерации электронно-дырочных пар в n -области, равно:

$$N_p \approx \int_0^d \Phi_0 \alpha \exp(-\alpha x) \exp\left[-\frac{|d-x|}{L_p}\right] dx = \frac{\Phi_0 \alpha}{\alpha - \frac{1}{L_p}} \left[\exp(-d/L_p) - \exp(-\alpha d) \right] \quad (2.8)$$

Аналогично, число неосновных носителей (электронов), достигающих перехода и обусловленных генерацией электронно-дырочных пар в p -области, равно интегралу (2.8), взятому в пределах от $x=d$ до $x=\infty$ при замене L_p на L_n :

$$N_n \approx \int_d^{+\infty} \Phi_0 \alpha \exp(-\alpha x) \exp\left[-\frac{|d-x|}{L_n}\right] dx = \frac{\Phi_0 \alpha \exp(-\alpha d)}{\alpha + \frac{1}{L_n}}. \quad (2.9)$$

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

$$N \approx \frac{\Phi_0 \alpha}{\alpha - \frac{1}{L_p}} \left[\exp(-d/L_p) - \exp(-\alpha \cdot d) \right] + \frac{\Phi_0 \alpha \cdot \exp(-\alpha \cdot d)}{\left(\alpha + \frac{1}{L_n} \right)}. \quad (2.10)$$

Таким образом, в стационарных условиях через структуру протекает постоянный ток, плотность которого $J_{кз} = eN$. Отметим, что выражение (2.10) получено для солнечных батарей, в которых освещаемая область имеет проводимость n -типа. Очевидно, что при освещении со стороны p -слоя необходимо лишь поменять местами L_p и L_n .

Полученное выражение позволяет также найти спектральное распределение фоточувствительности батареи, однако, при этом необходимо помнить, что, поскольку энергия одного фотона $h\nu$ пропорциональна частоте, то при одной и той же энергии излучения плотность потока фотонов Φ_0 пропорциональна длине волны λ . Рис. 2.11 б иллюстрирует хорошее совпадение расчетных и экспериментальных данных.

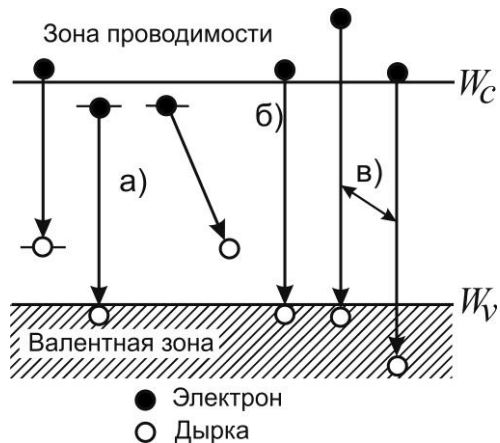
2.3. СВЕТОДИОДЫ И ЛАЗЕРЫ

Полупроводниковые светодиоды и лазеры представляют собой оптоэлектронные приборы, преобразующие электрическую энергию в энергию оптического излучения. В классических устройствах такого типа используются межзонные переходы электронов (рис. 2.12), т.е., в основном, генерация фотонов в материале происходит при излучательной рекомбинации электронов и дырок⁸.

⁸ Хотя в последнее время все большее распространение получают лазеры на внутризонных переходах.

В такой ситуации энергия излучаемых квантов, как видно из рисунка, определяется шириной запрещенной зоны кристалла: $h\nu > W_g$. В настоящее время разнообразие полупроводниковых материалов позволяет перекрыть широкий спектральный диапазон излучения: от ближнего ультрафиолета до среднего ИК диапазона.

Рис. 2.12. Основные виды переходов в полупроводнике [3]: а) переходы с участием примесей или дефектов (вакансии решетки и т. п.): между зоной проводимости и акцепторным уровнем, между донорным уровнем и валентной зоной, между донорным и акцепторным уровнями (межпримесное излучение); б) межзонные переходы: собственное излучение, энергетически очень близкое к ширине запрещенной зоны; в) излучение с более высокой энергией с участием «горячих» носителей, т.е. носителей, обладающих высокой энергией



Сразу отметим, что не все полупроводниковые материалы эффективно испускают свет при межзонной рекомбинации. Для объяснения данного факта необходимо вспомнить, что волновые числа фотонов в оптическом диапазоне составляют по порядку величины $k_{opt} = 2\pi/\lambda \sim 10^4 \dots 10^6 \text{ см}^{-1}$, а типичные значения волновых чисел носителей заряда в зоне Бриллюэна – $k_e = 2\pi/a \sim 10^8 \text{ см}^{-1}$, где a – постоянная решетки материала. Таким образом, волновой вектор излучения является пренебрежимо малым по сравнению с тем же параметром для электронов валентной зоны и зоны проводимости. То есть оптические межзонные переходы электронов происходят практически без изменения волнового числа (импульса). В этом смысле такие переходы называют *прямыми* или *вертикальными* (на зонной диаграмме начальное и конечное состояния электрона находятся на одной вертикали – рис. 2.13).

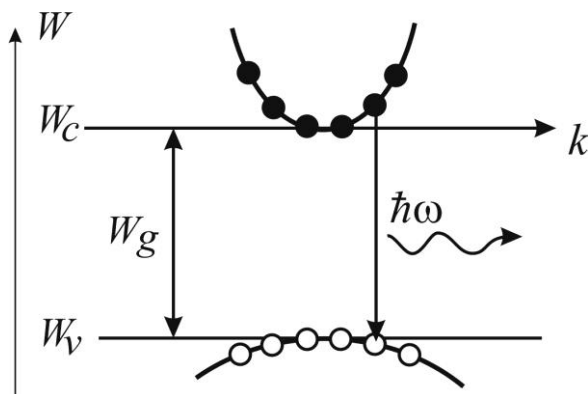


Рис. 2.13. Схематическое изображение вертикального перехода на зонной диаграмме полупроводника

Как известно [7], по структуре энергетических зон полупроводники разделяют на *прямозонные* и *непрямозонные*. К первым относятся кристаллы, у которых нижний минимум зоны проводимости и верхний максимум валентной зоны находятся в одной и той же точке зоны Бриллюэна (т.е. им соответствует

одно и то же значение волнового числа). В непрямозонных же полупроводниках экстремумам зон соответствуют различные волновые числа. Из прямозонных полупроводников наиболее широко используются арсенид галлия и фосфид индия. Самыми распространенными непрямозонными материалами являются кремний и германий. Важно, что в прямозонных полупроводниках возможны вертикальные переходы электронов, поэтому коэффициент поглощения и скорость излучательной рекомбинации в таких веществах высоки. В полупроводниках непрямозонных оптические переходы также возможны (рис. 2.14), но законы сохранения энергии и импульса в этом случае требуют, чтобы в таких процессах принимали участие не только фотоны и электроны, но и центры рассеяния (например, фононы), которые могли бы обеспечить существенное изменение импульса носителей. Поскольку в данном случае необходимо участие многих частиц, то, очевидно, вероятность подобных *непрямых переходов* будет ниже. Это отражается на зависимости коэффициента поглощения от длины волны для GaAs, Si и Ge, которую мы приводили на рис. 2.3. Видно, что у непрямозонных Si и Ge вблизи границы фотоэффекта коэффициент поглощения сравнительно мал, поскольку он определяется маловероятными непрямыми переходами. С увеличением энергии фотонов (с уменьшением длины волны излучения) становятся возможны переходы непосредственно в верхние долины, поэтому коэффициент поглощения растет. В то же время, у прямозонного GaAs наблюдается резкий рост коэффициента поглощения практически сразу при переходе через границу фотоэффекта. В итоге, вероятность межзонных излучательных переходов в непрямозонных полупроводниках значительно ниже, чем в материалах с прямой зоной. В связи с этим в настоящее время для изготовления высококачественных светодиодов и лазеров используют, в основном, прямозонные полупроводники.

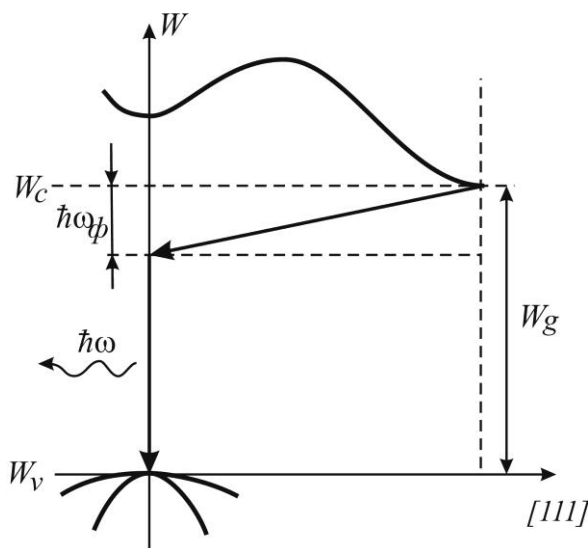


Рис. 2.14. Оптический межзонный переход с участием фонона в непрямозонном полупроводнике.

Приведенные выше рассуждения справедливы, однако, лишь для межзонных переходов в чистых полупроводниках. При межподзонных переходах в гетероструктурах или при межпримесных переходах в сильнолегированных полупроводниках ситуация меняется. Сейчас ведутся активные поиски возмож-

ностей организации лазерного излучения на непрямозонных материалах. Последнее необходимо, в первую очередь, для кремниевых цифровых СБИС, где важно заменить многоэлектродные шины данных, ограничивающие скорость передачи информации внутри микросхемы, на интегрированные оптические волноводы, имеющие существенно большую пропускную способность.

Перейдем далее непосредственно к рассмотрению принципов работы светоизлучающих диодов и полупроводниковых лазеров. Функционирование подобных приборов основано на явлении *люминесценции*. Напомним, что *люминесценцией* называют оптическое излучение, возникающее в результате возбуждения электронов в материале (к люминесценции, однако, не относится тепловое излучение). В зависимости от способа возбуждения различают следующие типы люминесценции: *фотолюминесценция*, при которой предполагается возбуждение системы оптическим излучением; *катодолюминесценция* – возбуждение пучком электронов; *радиолюминесценция* – возбуждение другими быстрыми частицами или излучением высокой энергии; *электролюминесценция* – возбуждение электрическим полем или током.

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

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

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

Как известно, необходимым условием для получения лазерного излучения является создание *инверсии населенности* в излучающей среде. Населенность считается инверсной, когда в некотором возбужденном состоянии с энергией W' частиц находится больше, чем на основном уровне с энергией

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

Рассмотрим далее, как эти общие принципы реализуются в полупроводниковых лазерах.

Для того чтобы рассмотреть условия получения инверсной населенности для полупроводниковых лазеров, обратимся к рисунку 2.15, на котором изображена зависимость энергии носителей от плотности состояний в прямозонном полупроводнике.

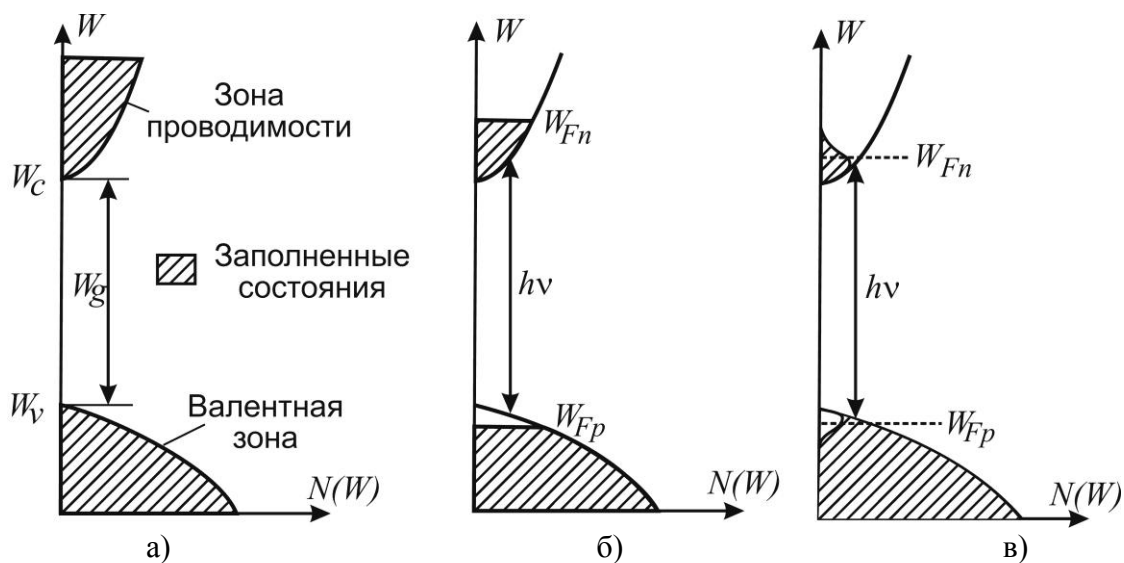


Рис. 2.15. Зависимость плотности состояний от энергии в полупроводнике [3]: а) состояние равновесия, $T=0$ К; б) инверсия, $T=0$ К; в) инверсия, $T>0$ К

На рис. 2.15, а представлены равновесные условия при $T=0$ К для собственного полупроводника. Заштрихованная площадь соответствует заполненным состояниям. Рис. 2.15, б соответствует случаю инверсной населенности при $T=0$ К. Валентная зона пуста до энергетического уровня W_{Fp} , а зона проводимости заполнена до уровня W_{Fn} . Фотоны, имеющие энергию $h\nu$, лежащую в

интервале $W_g < h\nu < (W_{Fn} - W_{Fp})$, вызывают переходы электронов на нижние уровни, т.е. возникает стимулированная эмиссия.

При отличных от абсолютного нуля температурах распределение носителей «размывается» по энергиям, как показано на рис. 2.15, в. Хотя общее тепловое равновесие всей системы отсутствует, можно считать, что носители в данной энергетической зоне находятся в тепловом равновесии друг с другом. В этом случае, как мы видели выше, пользуются распределением Ферми-Дирака, вводя квазиуровни Ферми для электронов и дырок. Тогда условие инверсии для какой-либо пары уровней, один из которых (W') лежит в зоне проводимости, а другой (W) – в валентной зоне, будет выглядеть следующим образом:

$$f(W') = \frac{1}{1 + \exp\left(\frac{W' - W_{Fn}}{kT}\right)} > f(W) = \frac{1}{1 + \exp\left(\frac{W - W_{Fp}}{kT}\right)}, \quad (2.12)$$

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

$$W_{Fn} - W_{Fp} > W' - W \quad (2.13)$$

или, что то же самое,

$$W_{Fn} - W_{Fp} > h\nu. \quad (2.14)$$

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

$$W_{Fn} - W_{Fp} > W_g. \quad (2.15)$$

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

Как упоминалось выше, наиболее распространенным методом возбуждения в полупроводниковых лазерах является инжекция электронов и дырок через прямосмещенный p - n переход. Энергия для генерации излучения в этом случае черпается из электрической батареи, подключенной к диоду. Для увеличения интенсивности свечения p - и n -области перехода легируются до вырождения электронно-дырочного газа. В такой ситуации в условиях равновесия уровень Ферми в p -области лежит в валентной зоне, а в n -области – в зоне проводимости (рис. 2.16 а).

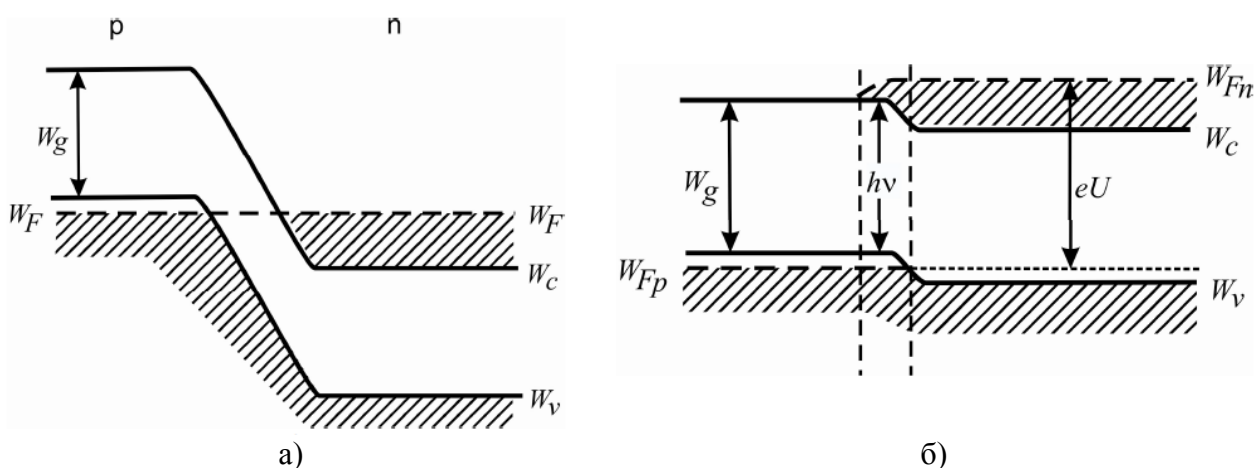


Рис. 2.16. Диаграмма энергетических зон лазера с p - n переходом (подобная диаграмме туннельного диода): а) без внешнего напряжения; б) при прямом смещении. Штриховкой изображены энергетические состояния, занятые электронами

Отметим, что зонная диаграмма, приведенная на рис. 2.16, подобна диаграмме туннельного диода (см. далее). Однако степень легирования областей лазера ниже, чем туннельного диода, поэтому обычно вольтамперная характеристика лазерного диода не имеет области отрицательного сопротивления или эта область пренебрежимо мала (сравните с ВАХ туннельного диода из главы 3).

При подаче определенного смещения, как видно из рисунка 2.16 б, вблизи границы перехода выполняется условие инверсии (2.15), необходимое для обеспечения однопроходного квантового усиления.

Для обеспечения генерации когерентного излучения, как мы помним, необходимо наличие резонатора. В отличие от твердотельных лазеров, в данном случае нет необходимости во внешнем оптическом резонаторе. Зеркалами служат сами грани кристалла, что автоматически обеспечивает их строгую параллельность при скалывании. Такая методика позволяет избежать проблем юстировки и снизить габариты прибора. Для формирования гладкой отражающей поверхности кристалл скалывается и полируется. Коэффициенты отражения получаемых зеркал сравнительно невысоки:

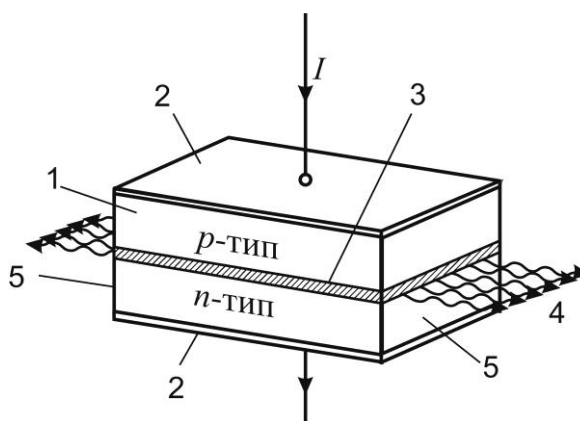
$$R = \frac{(n-1)^2}{(n+1)^2} \approx 20 \dots 40\% , \quad (2.16)$$

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

Общий вид инжекционного лазера представлен на рис. 2.17.

Рис. 2.17. Общий вид полупроводникового лазера с резонатором Фабри-Перо [3]:

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



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

Обратите внимание, что активная область лазера сосредоточена непосредственно вблизи границы p и n областей (рис. 2.16 б) и имеет поперечный размер порядка микрометра, что сравнимо с длиной волны излучения. Поэтому при выходе во внешнюю среду излучение сильно дифрагирует. Таким образом, полупроводниковые лазеры обладают высокой расходимостью ($20^\circ \dots 40^\circ$), которую компенсируют с помощью дополнительной фокусирующей оптики.

Ширина линии спонтанного излучения, характерного для светодиодов, определяется распределением носителей заряда по энергии и составляет несколько единиц kT (см. рис. 2.18). Это приводит к тому, что спектр излучения светодиодов будет иметь ширину⁹ примерно $4 kT$.

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

⁹ Данная оценка является достаточно грубой и отражает ширину спектра светодиода, измеренную по уровню 0.1 от максимального значения интенсивности.

нестабильной, что зависит от множества факторов — силы тока, температуры и т.д.

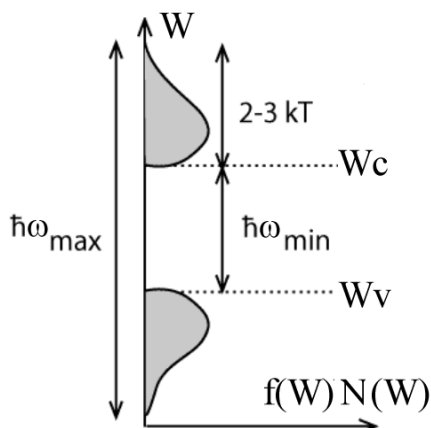


Рис. 2.18. Распределение электронов и дырок по энергии в зоне проводимости и валентной зоне, соответственно. Стрелками указаны минимальная и максимальная энергия кванта, генерируемого при электронно-дырочной рекомбинации

У лазерного диода есть два главных отличия от светодиода:

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

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

Важно, что в процессе формирования вынужденного излучения в полупроводниковом лазере участвуют одновременно электрон, дырка и первичный фотон, которые должны геометрически оказаться вблизи друг друга. В обычном p - n -переходе это затруднено, так как рекомбинирующие носители заряда «размазаны» в слое, размером порядка 1 мкм, т.е. на расстоянии диффузионной длины от границы перехода, а профиль светового поля существенно шире толщины активной области прибора. В связи с этим для обеспечения генерации при комнатной температуре необходима очень сильная инверсия населенности, что, в свою очередь, требует больших плотностей тока накачки (~ 1 кА/см²). Во избежание сильного разогрева и выхода прибора из строя подобные диоды работают в импульсном режиме либо требуют принудительного охлаждения. Поэтому все современные лазеры изготавливаются не на основе гомоструктур, а с использованием гетеропереходов.

Одной из самых распространенных лазерных конструкций является двойная гетероструктура (ДГС), изображенная на рисунке 2.19. Здесь слой материала с более узкой запрещенной зоной (GaAs) располагается между двумя слоями материала с более широкой запрещенной зоной (AlGaAs). Каждый из двух контактов GaAs–AlGaAs сам по себе является гетероструктурой, поэтому подобную трехслойную композицию и называют двойной (или двусторонней) гетероструктурой.

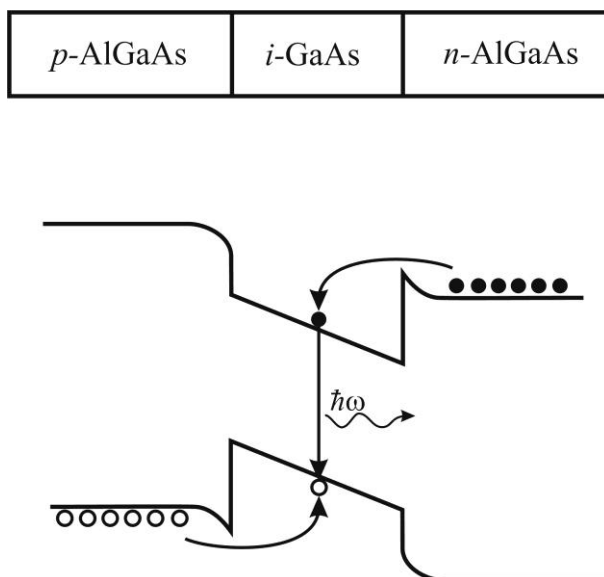


Рис. 2.19. Схематическое изображение конструкции и зонной диаграммы лазера с двумя гетеропереходами, ограничивающими диффузию инжектируемых электронов и дырок

ДГС лазеры имеют ряд преимуществ, основными из которых являются следующие.

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

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

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

В случае, когда толщина слоя узкозонного материала имеет размер порядка длины волны электрона, в активной области прибора реализуется квантовая яма (рис. 2.19), в которой электроны и дырки размещены на энергетических уровнях. Последнее приводит к резкому увеличению коэффициента усиления и позволяет уменьшить пороговый ток накачки лазерного диода еще на порядок.

Дальнейшее продвижение технологии изготовления лазерных диодов связано с использованием квантовых объектов еще меньшей (нулевой) размерности, чем квантовая яма — *квантовых точек*. По сути, квантовая точка — это объемное включение узкозонного материала внутри широкозонного, причем размеры этого объекта составляют доли длины волны электрона. Таким образом, по всем трем направлениям формируется квантовая яма для носителей

заряда. Используя термины электродинамики, можно говорить, что в случае квантовой точки мы имеем резонатор для одного электрона.

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

Первоначально на поверхность GaAs с помощью технологического процесса эпитаксии наносят тонкий слой InAs (рис. 2.20 а). Далее начинается рост трехмерных изолированных островков: для минимизации упругой энергии InAs собирается в «пирамидки» (рис. 2.20 б, в). После этого поверх сформированных «пирамидок» наносят следующий слой GaAs, «заращая» образованные островки. При нанесении второго слоя InAs новые островки будут формироваться в поле упругих напряжений, созданных островками первого слоя. Таким образом, новые квантовые точки будут образовываться непосредственно над старыми. Описанная процедура повторяется 20...50 раз, что, в итоге, позволяет вырастить трехмерный массив островков InAs в матрице GaAs.

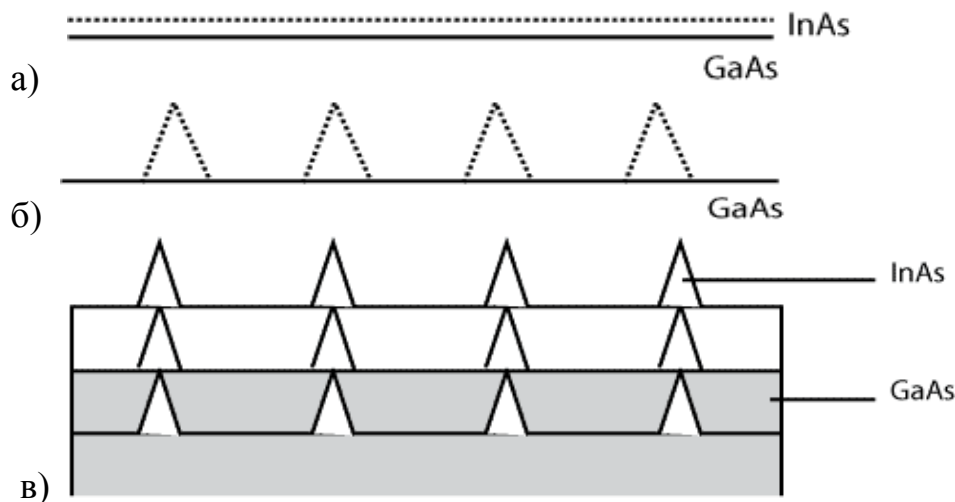


Рис. 2.20. Схема основных этапов технологии изготовления матрицы квантовых точек¹⁰

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

¹⁰ Изображение носит схематический характер. В действительности сформировать абсолютно упорядоченную структуру не удастся.

ЧАСТЬ 3 СВЧ И КВЧ ¹¹ ДИОДЫ

3.1. Диод ГАННА

Диодом Ганна называют тип полупроводниковых диодов, использующихся для генерации и преобразования колебаний в СВЧ диапазоне. Конструктивно диод Ганна представляет собой просто однородный полупроводник с нанесенными на него омическими контактами (рис. 3.1). Принцип действия диодов Ганна, в отличие от большинства полупроводниковых приборов, основан не на свойствах различных переходов, а на объёмных свойствах однородного полупроводника. В 1963 году Джон Ганн исследовал вольтамперные характеристики образцов GaAs и InP. Он обнаружил, что когда напряженность электрического поля, приложенного к образцу, которая определяется как

$$E = U/L \quad (3.1)$$

(где U – приложенное напряжение, L – длина образца), превосходит некоторую критическую величину E_t (~ 3 кВ/см для GaAs и ~ 6 кВ/см для InP), в цепи возникают спонтанные колебания тока (рис. 3.2). Частота этих колебаний примерно равнялась величине, обратной времени пролета носителей заряда через образец:

$$f = v_{dr}/L, \quad (3.2)$$

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



Рис. 3.1. Структура GaAs диода Ганна. Система слоев n^+ -Au реализует омический контакт

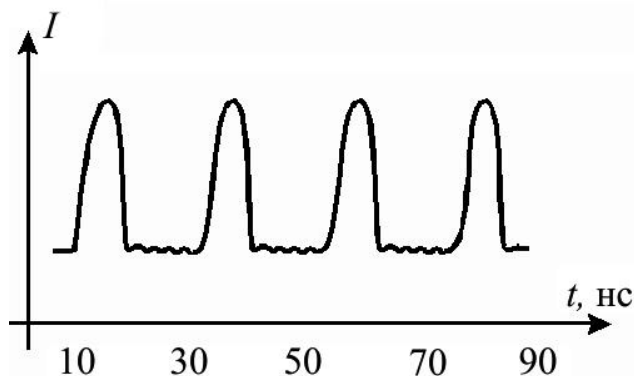


Рис. 3.2. Зависимость тока от времени для GaAs диода Ганна

¹¹ СВЧ – сверхвысокая частота, КВЧ – крайне высокая частота. Термины введены по аналогии с англоязычной литературой для обозначения частот выше 1 ГГц (СВЧ) и выше 100 ГГц (КВЧ). Следует отметить, что указанные границы частот весьма условны.

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

В 1963г. Ридли предсказал, что доменная неустойчивость должна появляться в полупроводниковом образце в случае, если на вольтамперной характеристике образца имеется участок *отрицательного дифференциального сопротивления (ОДС)* или, что то же самое, *отрицательной дифференциальной проводимости (ОДП)* N-типа (рис. 3.3). ВАХ будет иметь такой вид, если либо скорость носителей, либо их концентрация уменьшается с увеличением приложенного напряжения. Ридли, Уоткинс и Хилсум показали, что скорость носителей в GaAs и InP n-типа и в некоторых других полупроводниковых материалах должна уменьшаться с ростом электрического поля, когда напряженность поля превышает некоторое критическое значение. Данный эффект объясняется междолинными переходами электронов: в сильных полях энергия носителей заряда оказывается достаточной для их перехода из основной Γ -долины в более высоколежащие X и L-долины, где эффективная масса электронов увеличивается, что и приводит к снижению их дрейфовой скорости. График зависимости $v(E)$ для разных значений подвижности, определяемых уровнем легирования образца, приведен на рис. 3.4. На графике хорошо виден спадающий участок – область ОДП.

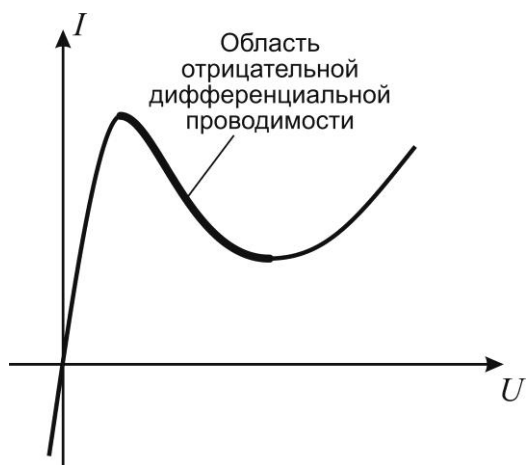


Рис. 3.3. Пример ВАХ N-типа. Жирной линией показан участок отрицательной дифференциальной проводимости, т.е. область, где $\frac{dI}{dU} < 0$

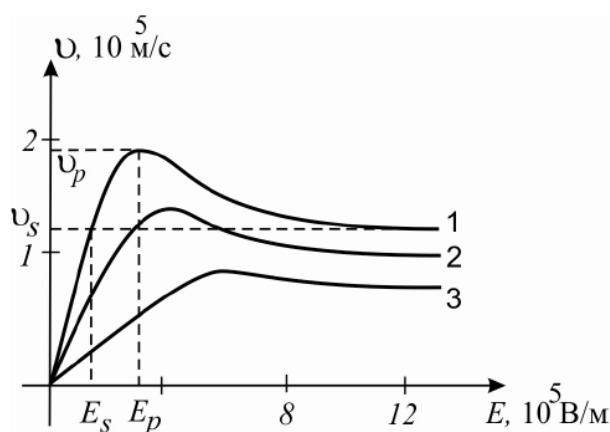


Рис. 3.4. Зависимости дрейфовой скорости электронов v от напряженности электрического поля E в GaAs для различных значений подвижности μ :

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

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

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

Рассмотрим подробнее процесс формирования области сильного поля (домена Ганна) в образце (рис. 3.5) и ее влияние на ток, протекающий через структуру.

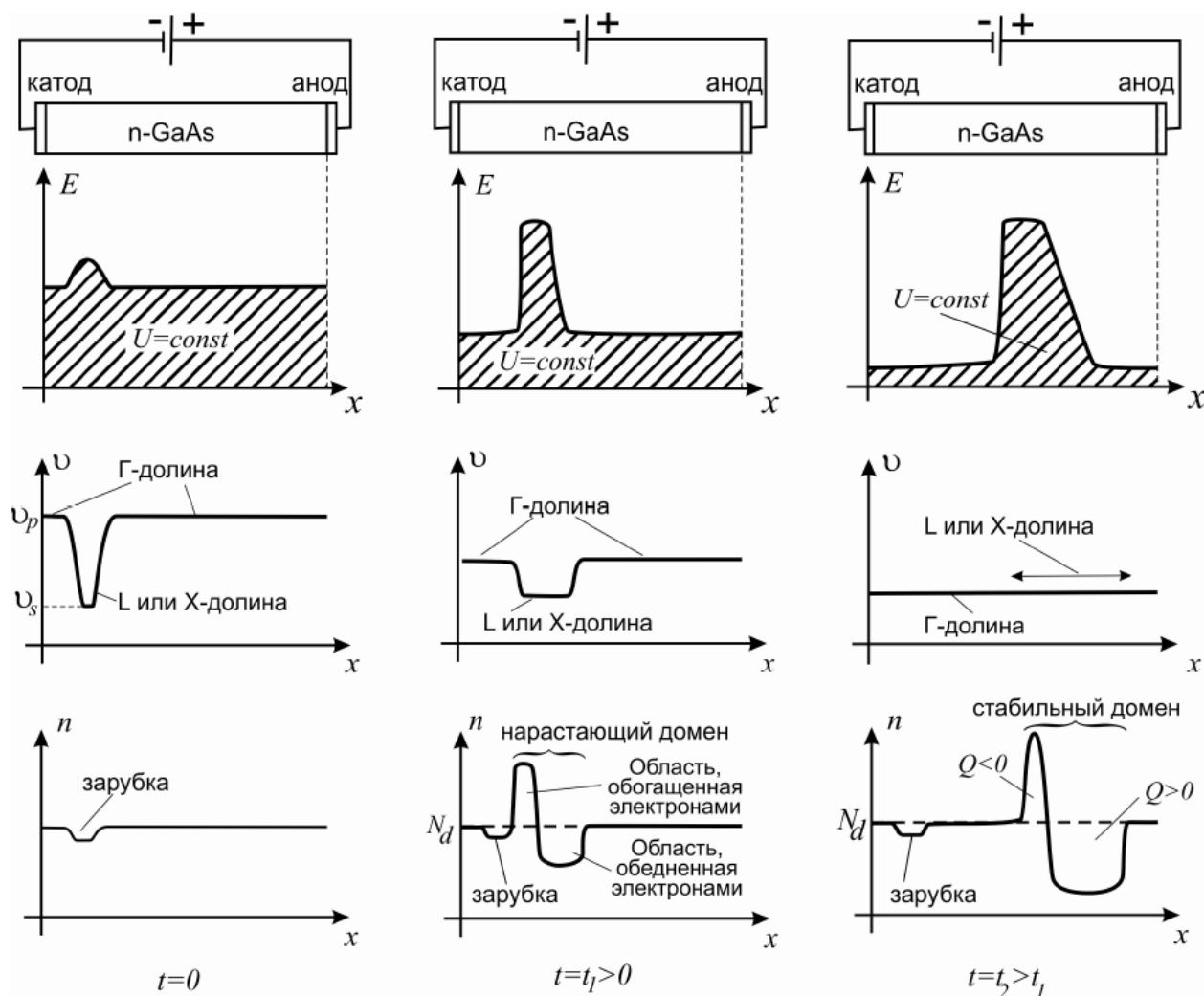


Рис. 3.5. Процесс формирования домена. Здесь Q – суммарный заряд электронов и доноров; E – напряженность электрического поля; v – скорость электронов; n – концентрация носителей заряда; x – координата вдоль образца. Пунктиром показана концентрация доноров N_d в объеме полупроводника. В области зарубки концентрация доноров (и электронов) имеет меньшее значение

Предположим, что в начальный момент времени $t=0$ напряженность электрического поля E в образце равна E_p всюду, за исключением небольшой области вблизи катода, где она немного выше вследствие неоднородности легирования. Так как напряженность поля в указанной области превышает E_p , скорость электронов здесь меньше, чем в остальных частях образца, поскольку при $E > E_p$ скорость v падает с ростом напряженности поля E (рис. 3.4). Следовательно, электроны перед областью сильного поля и за этой областью движутся быстрее, чем внутри нее, где они претерпевают междолинный переход и становятся более тяжелыми и медленными. Это ведет к обеднению электронами переднего и обогащению ими заднего фронтов движущейся области сильно-

го поля. Иными словами, быстрые электроны, находящиеся левее домена (рис. 3.5), догоняют его, попадают в область сильного поля и тормозятся. Быстрые электроны, находящиеся справа от домена, наоборот, убегают от него, обнажая положительный заряд ионизованных доноров. Таким образом, концентрация электронов в задней части домена превышает концентрацию ионов доноров, а в передней части – наоборот. То есть распределение зарядов в домене представляет собой дипольный слой, а направление электрического поля совпадает с его направлением в остальных частях образца. Тогда напряженность поля в области домена начинает расти и вызывает дальнейшее увеличение междолинных переходов и рост домена.

Если напряжение смещения U на образце поддерживается постоянным, то нарастание напряженности поля внутри домена происходит за счет уменьшения напряженности поля в остальных областях (площадь под графиком $E(x)$ остается постоянной). При этом как внутри, так и вне домена скорость носителей заряда уменьшается. Через некоторое время скорости электронов, находящихся вне и внутри области сильного поля, уравниваются, а домен перестает расти. В такой ситуации говорят, что он стабилизируется. Другой причиной стабилизации домена является диффузия электронов, вызванная градиентом их концентрации. После стабилизации домен движется к аноду, не меняя своей формы. Что касается тока диода, то он уменьшается в процессе формирования домена, а затем остается постоянным, пока домен не добежит до анода. Когда домен сильного поля достигает анода, он рассасывается, а ток при этом снова возрастает до исходного значения. Затем вблизи катода зарождается новый домен, и процесс повторяется. Обычно в образце формируется только один домен, поскольку напряженность поля везде, за исключением домена, меньше E_p , т.е. междолинные переходы в области вне домена подавлены. В итоге, мы получаем периодическую зависимость тока образца от времени (рис. 3.2)

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

$$j_s = qnv_s. \quad (3.4)$$

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

$$j_p = qnv_p. \quad (3.5)$$

Время движения домена (пролетное время) определяется выражением

$$T \approx L/v_s, \quad (3.6)$$

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

$$t_{gr} \approx 3\tau_m = \frac{3\varepsilon}{e|\mu_d|n}, \quad \mu_d = \frac{dv}{dE}. \quad (3.7)$$

Напомним, что по своему физическому смыслу максвелловское время релаксации может быть определено как время, за которое рассасывается флуктуация концентрации электронов Δn , возникшая из-за их случайного теплового движе-

ния: $\Delta n = \Delta n_0 e^{-\frac{t}{\tau_m}}$. В случае специальных условий, которые реализуются в диоде Ганна (т.е. при наличии отрицательной дифференциальной подвижности $\mu_d < 0$), флуктуация концентрации электронов будет не рассасываться, а, наоборот, нарастать до тех пор, пока домен не стабилизируется. Для формирования стабильного движущегося домена время t_{gr} должно быть существенно меньше, чем пролетное время

$$t_{gr} < \frac{L}{v_s}. \quad (3.8)$$

Из этого неравенства вытекает так называемый *критерий Кремера*:

$$nL > (nL)_1 = \frac{3\varepsilon v_s}{e|\mu_d|}. \quad (3.9)$$

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

Процесс формирования домена можно рассматривать как зарядку плоского конденсатора, где одной обкладкой является область, в которой электронов больше, чем доноров, а другой – соседняя область, где концентрация ионов доноров больше, чем электронов (рис. 3.5). Тогда время формирования домена $\tau_m = RC_{\text{домена}}$. Поскольку домен формируется во всем сечении диода Ганна, то при увеличении площади сечения емкость домена будет возрастать, а сопротивление соседних с доменом областей, через которые эта емкость заряжается – снижаться. В результате характерное время зарядки домена останется без изменений. Это говорит о том, что формирование ганновской неустойчивости – процесс, определяемый свойствами самого полупроводникового материала, поэтому характерное время формирования домена должно определяться константами материала ($\tau_m = \varepsilon/\sigma$), как мы и получили выше. Явление доменообразования¹² в диоде Ганна, по сути, является механизмом положительной обратной связи, которая необходима для возникновения колебаний в любой колебательной системе.

Важной особенностью ганновской генерации является тот факт, что сформировавшийся домен не исчезает, если во время его движения в диоде напряжение на образце уменьшается до величины, меньшей, чем пороговое

¹² Более точно было бы говорить о явлении нарастания волн пространственного заряда, поскольку усиление и генерация в диоде Ганна в некоторых случаях могут реализовываться и без возникновения ярко выраженного домена.

напряжение формирования домена. Происходит это потому, что поле внутри домена остается *большим*, чем E_p , даже если средняя напряженность поля по образцу становится ниже E_p . Для значений параметра $(nL) \geq 10^{16} \text{ м}^{-2}$ напряженность поля, необходимая для поддержания домена, близка к величине E_s .

Рассмотрим работу генератора на основе диода Ганна (рис. 3.6). Прибор представляет собой коаксиальный резонатор, который с одной стороны ограничен боковой стенкой коаксиального волновода с металлизированным торцом, к которому припаян диод, а с другой стороны – плунжером – подвижным поршнем с отверстием в центральной части. Перемещая плунжер, т.е. изменяя геометрические размеры резонатора, можно осуществлять его настройку на различные резонансные частоты. Таким образом, в генераторе имеется возможность управлять частотой колебаний выходного сигнала, хотя длина диода Ганна является фиксированной, так что время пролета домена строго задано. Кроме того, некоторые конструкции плунжера позволяют менять диаметр его внутреннего отверстия так, что возможна регулировка мощности выходного сигнала, а также мощности сигнала, который «остается» в резонаторе, детектируется центральным стержнем и прикладывается в виде переменного напряжения к диоду Ганна. Иными словами, диаметр внутреннего отверстия плунжера определяет глубину обратной связи в генераторе.

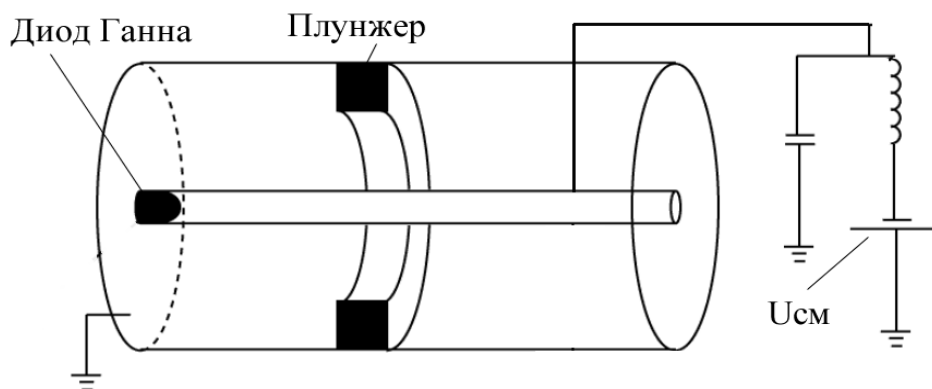


Рис. 3.6. Схема СВЧ генератора на основе коаксиального резонатора и диода Ганна. $U_{см}$ – напряжение смещения.

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

Сначала рассмотрим наиболее простой случай *пролетного* режима работы генератора. Пусть время пролета домена в диоде в точности равно периоду колебаний электромагнитного поля резонатора. Такая мода колебаний представлена на рис. 3.7.

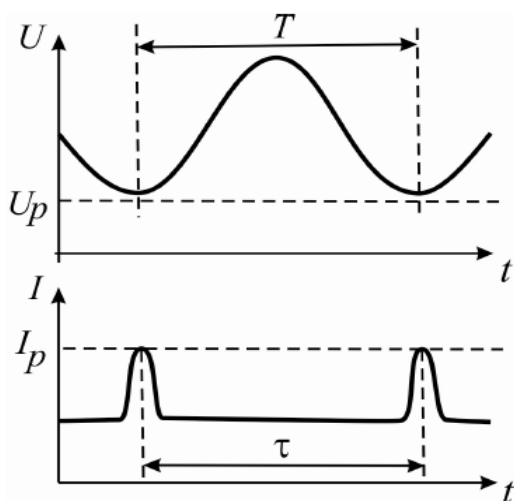


Рис. 3.7. Пролетный режим работы генератора (пролетное время домена в диоде τ равно периоду колебаний электромагнитного поля T)

Важно, что период переменной составляющей напряжения, прикладываемого к диоду, будет соответствовать периоду электромагнитного поля, так как это напряжение порождено самим полем. Характерная форма колебаний тока полностью аналогична рассмотренной выше форме колебаний для постоянного напряжения смещения. Следует отметить, что в данном режиме работы постоянную составляющую напряжения, подаваемого на диод от батареи, выбирают такой, что напряженность поля внутри диода всегда выше полей зарождения и удержания домена. Обозначим такое напряжение U_p . При исчезновении домена у анода и формировании вблизи катода нового домена поле в образце быстро достигает порогового значения, которое сохраняется, пока не сформирован новый домен. Поэтому генерируются короткие импульсы тока (рис. 3.7).

В режиме *запаздывания* пролетное время τ меньше периода колебаний резонатора T (рис. 3.8).

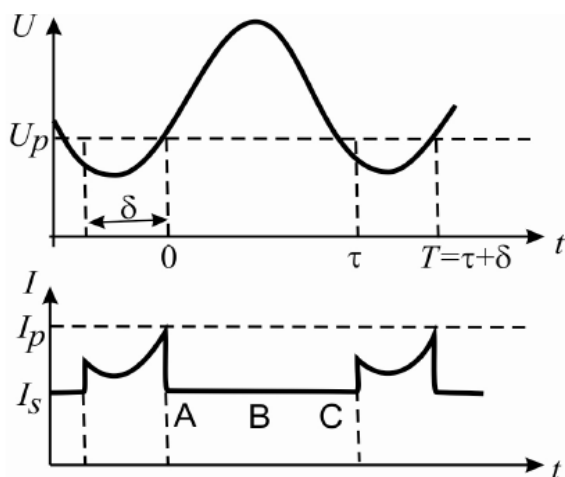


Рис. 3.8. Идеализированный режим запаздывания формирования домена, параллельный резонансный контур. В момент времени А домен формируется, В – распространяется вдоль образца, С – достигает анода

Пусть домен формируется в момент времени $t = 0$. Пусть также величина постоянного смещения и амплитуда СВЧ напряжения таковы, что перед тем, как домен достигнет анода в момент времени $t = \tau$ (τ - пролетное время), напряжение на диоде упадет ниже порогового значения U_p . Далее в некоторый момент времени $t = \tau + \delta$ напряжение снова превысит порог (рис. 3.8), и у катода

начнет образовываться новый домен. Таким образом, если предположить, что домен рассасывается на аноде мгновенно, то формирование нового домена запаздывает на время δ . Период колебаний при этом равен $T = \tau + \delta$. Максимальная выходная мощность в рассматриваемом режиме достигается, если домен достигает анода и рассасывается сразу же после того, как напряжение падает ниже порогового значения. Варьируя геометрические размеры резонатора, можно менять период колебаний T за счет изменения задержки δ , в то время как пролетное время τ остается неизменным и определяется длиной диода Ганна. Отметим, что в отсутствие домена форма зависимости тока от времени повторяет форму напряжения, т.е. диод представляет собой обычный резистор и подчиняется закону Ома. После зарождения нового домена ток снова резко падает (рис. 3.8).

Еще одним режимом работы генератора с диодом Ганна является режим с *гашением* домена (рис. 3.9). В этом случае в процессе перемещения по образцу домен разрушается (гасится), не успевая дойти до анода. Для такого разрушения дипольного слоя необходимо настолько сильно уменьшить поданное на диод напряжение, чтобы даже в максимуме напряженность электрического поля внутри домена стала меньше пороговой (рис. 3.5). Напряжение U_s , при котором сформированный домен разрушается, не доходя до анода, называют напряжением гашения, и, как упоминалось выше, оно меньше порогового напряжения U_p . Когда суммарное напряжение на диоде падает до величины, меньшей напряжения гашения, домен исчезает, а прибор возвращается в омическое состояние, в котором он остается до тех пор, пока напряжение вновь не превысит порог образования домена (рис. 3.9). Далее процесс повторяется.

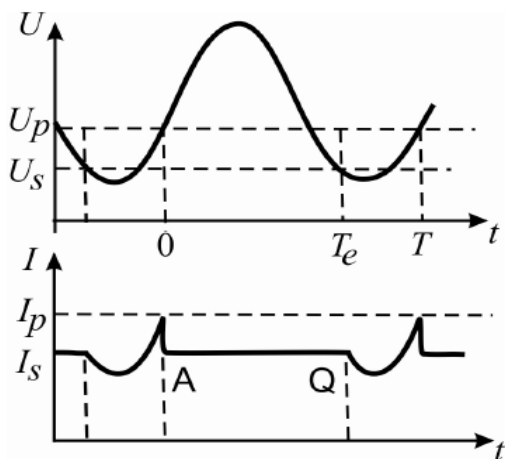


Рис. 3.9. Идеализированный режим гашения домена. В момент времени А домен формируется, а в момент Q – гасится, не дойдя до анода

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

3.2. ЛАВИННО-ПРОЛЕТНЫЕ И ИНЖЕКЦИОННО-ПРОЛЕТНЫЕ ДИОДЫ

3.2.1. Лавинно-пролетные диоды

В лавинно-пролетных диодах (ЛПД) для получения отрицательного сопротивления в СВЧ-диапазоне используются ударная ионизация в полупроводниковых структурах и пролетные эффекты сгустков электронов и дырок. В режиме ЛПД p - n переход может работать при двух условиях: а) на него подано обратное смещение, достаточное для развития лавинного пробоя; б) он помещен в микроволновый резонатор, например, такой как приведен на рис. 3.6 предыдущего раздела.

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

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

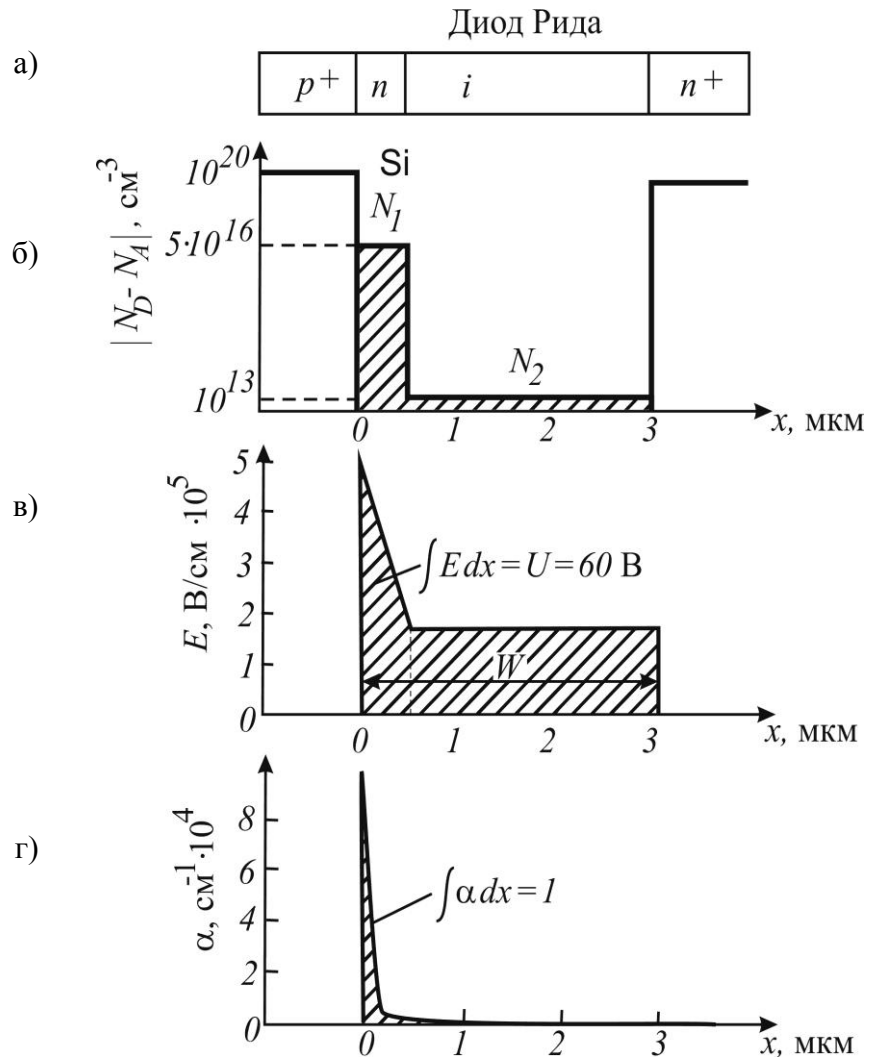
Рассмотрим принцип действия ЛПД на примере диода Рида. На рис. 3.10 показаны структура и распределения примесей, электрического поля и эффективной плотности ионизации в условиях пробоя для идеализированного диода Рида [3].

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

В диоде Рида зависимость напряженности поля от координаты является резко неоднородной (рис. 3.10 в), поэтому область «слоя умножения», т.е. слой, где происходит ударная ионизация, оказывается локализованным вблизи p - n перехода. Иными словами, большая часть актов ударной ионизации совершается в узком слое вблизи максимума напряженности электрического поля (рис. 3.10 г).

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

Рис. 3.10. Диод Рида: а) структура диода; б) распределение примесей в образце ($p^+ - n - i - n^+$); в) распределение электрического поля; г) распределение коэффициента ударной ионизации при лавинном пробое [2]



При помещении ЛПД в резонатор аналогично диоду Ганна, к прибору прикладывается переменное напряжение, так что суммарное напряжение в течение небольшой части периода колебаний становится больше напряжения пробоя $U_{проб}$ (рис. 3.11). В течение этого интервала времени вблизи $p-n$ перехода происходит лавинная генерация электронно-дырочных пар, но из-за того, что электроны достигают анода за конечное время (пролетный эффект), максимум тока запаздывает относительно максимума напряжения. Величина запаздывания определяется временем прохождения электрона через пролётную n -область:

$$\tau = \frac{L}{v_{нас}} \approx \frac{T}{2}, \quad (3.10)$$

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

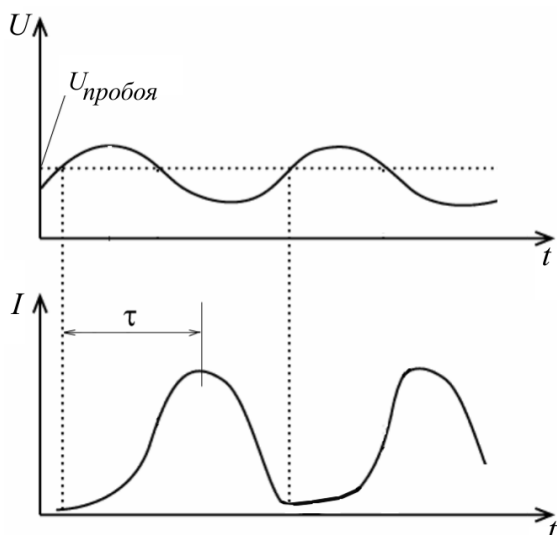


Рис. 3.11. Зависимость напряжения и тока от времени для ЛПД, помещенного в резонатор

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

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

$\frac{1}{4}T < \tau < \frac{3}{4}T$. На практике, однако, величина перестройки генератора обычно в 1.5...2 раза меньше.

3.2.2. Инжекционно-пролетные диоды

Принцип работы *инжекционно-пролетного диода* (ИПД), в целом, подобен ЛПД. Конструкция ИПД содержит два p - n -перехода и пролетную область (рис. 3.11), за счет которой осуществляется задержка между импульсами тока и напряжения на диоде. Разница заключается в режиме питания – ИПД работают при прямом смещении. На каждом периоде колебаний напряжения ИПД на короткое время отпирается и в пролетную n -область вбрасывается некоторое количество дырок, которые затем дрейфуют к аноду и создают импульс тока во внешней цепи. Схематические изображения зависимостей тока и напряжения от времени будут аналогичны приведенным на рис. 3.11. Поскольку правый p - n -переход находится при обратном смещении, большая часть пролетной области

ИПД в рабочем режиме обеднена электронами. При этом сгусток дырок, который инжектируется в n -область из левого, прямо смещенного, p - n -перехода, не рекомбинирует в пролетной области, а добегаёт до анода. Это позволяет получить эффективную задержку импульса тока диода относительно положительного полупериода напряжения, который открыл левый (инжектирующий) переход.

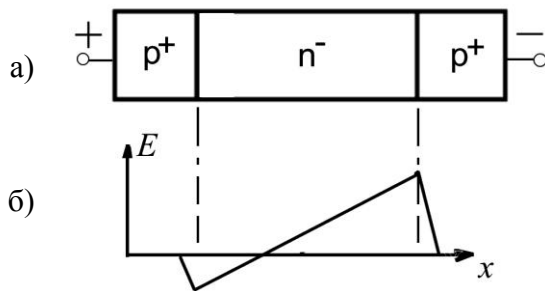


Рис. 3.12. Инжекционно-пролетный диод: а) схематическое изображение структуры диода; б) распределение напряженности электрического поля в образце.

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

Важной особенностью работы ИПД является возможность реализации такого напряжения питания, когда ОПЗ прямо смещенного перехода катода и обратосмещенного перехода анода смыкаются (рис. 3.12 б), так что во всей пролетной области (кроме одной точки, в которой и происходит смыкание) существует ненулевая напряженность электрического поля. Такое явление получило название «прокола». Данный режим работы диода позволяет существенно снизить эффект накопления инжектированных дырок внутри диода – поле высокой напряженности вблизи анода захватывает дырки и очень быстро переносит их в анод. В связи с этим на вольт-амперной характеристике прибора (рис. 3.13) участок резкого нарастания тока термоэмиссии сменяется участком ограничения тока пространственным зарядом инжектированных носителей только при очень больших плотностях тока. Это позволяет создавать не только инжекционно-пролетные диоды с относительно высоким КПД, но и использовать подобные приборы в качестве быстродействующих ограничителей напряжения.

В заключение, сравним основные характеристики ЛПД, ИПД и диода Ганна. Так же, как и ЛПД, ИПД отличается от диода Ганна тем, что отрицательная дифференциальная проводимость в приборе реализуется в узком спектральном диапазоне вблизи пролётной частоты. У диода Ганна ОДП реализуется в широком диапазоне частот. Энергия носителей заряда в ИПД меньше, чем в диоде Ганна (так как в ИПД нет междолинных переходов), а в диоде Ганна меньше, чем в ЛПД (так как в диоде Ганна нет пробоя), поэтому наименьшие шумы реализуются в ИПД и в диодах Ганна, а в ЛПД они значительно выше. Наоборот, амплитуды токов и выходные мощности соотносятся следующим образом: самые большие – у ЛПД, а у ИПД – наименьшие. Предельные частоты работы приборов составляют $f_{\text{ЛПД}} \approx 400 \dots 600$ ГГц, $f_{\text{на}} \approx 100 \dots 150$ ГГц, $f_{\text{ИПД}} \approx 30 \dots 50$ ГГц. Сравнение характеристик диодов приведено на рис. 3.14.

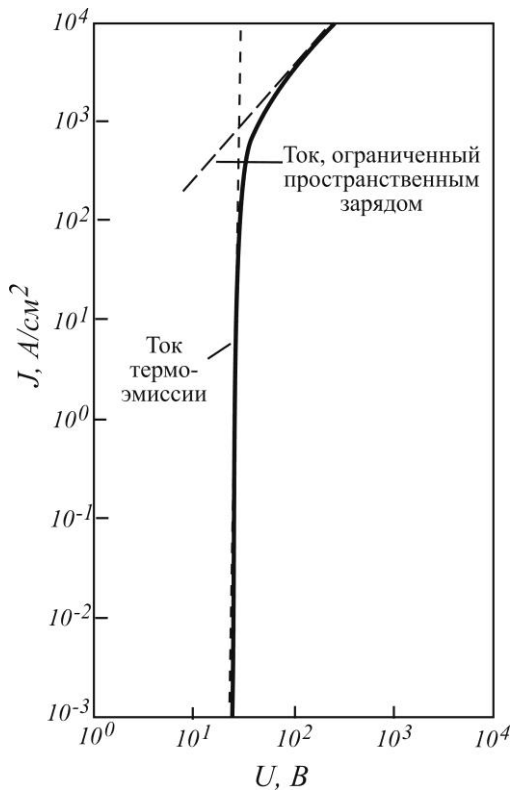


Рис. 3.13. Вольт-амперная характеристика p^+-n-p^+ диода в условиях работы с «проколом» [3]: пунктир - теоретический результат, сплошная линия - эксперимент

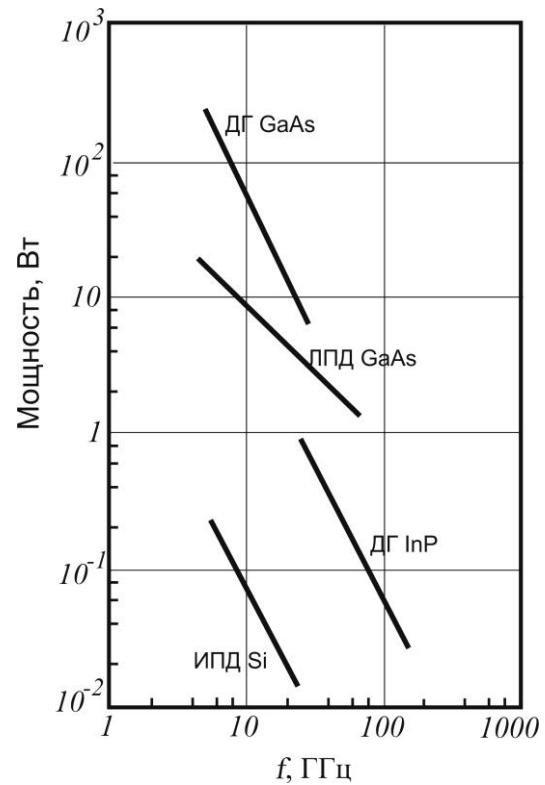


Рис. 3.14. Зависимость выходной мощности диодов Ганна (ДГ), ИПД и ЛПД от частоты генерируемого сигнала [3]

3.3. ДИОДЫ НА ОСНОВЕ ЭФФЕКТА ТУННЕЛИРОВАНИЯ. СВЕРХРЕШЕТКИ

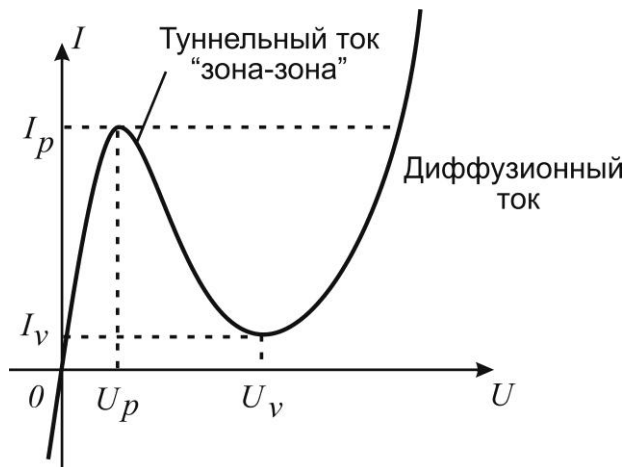
3.3.1. Туннельный диод

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

полученные вольт-амперные характеристики с помощью концепции квантового туннелирования носителей заряда и получил удовлетворительное соответствие между теорией и экспериментальными результатами.

Типичная статическая вольт-амперная характеристика туннельного диода показана на рисунке 3.15. На графике можно отметить следующие особенности. При приложении напряжения обратной полярности (отрицательный потенциал к p -области) ток монотонно возрастает. Таким образом, *туннельный диод, в отличие от классического, не обладает выпрямляющими свойствами*. При приложении прямого смещения ток сначала увеличивается до максимального значения I_p при напряжении U_p , а затем уменьшается до минимального значения I_v при напряжении U_v . При напряжениях, больших, чем U_v , ток ведет себя классическим образом, т.е. экспоненциально возрастает. Статическая характеристика складывается, в основном, из туннельного тока носителей между зонами и надбарьерного диффузионного тока (рис. 3.15).

Рис. 3.15. Вольт-амперная характеристика туннельного диода: I_p и U_p - пиковый ток и соответствующее напряжение; I_v и U_v - ток и напряжение в минимуме



Проведем качественное рассмотрение процесса туннелирования, используя упрощенную зонную структуру (рис. 3.16). Отметим, что, поскольку переход сильно легирован, электронно-дырочный газ в структуре является вырожденным, т.е. уровни Ферми p и n областей на зонной диаграмме находятся *внутри разрешенных зон*. В условиях теплового равновесия уровень Ферми постоянен по всей структуре (рисунок 3.16 б). Для простоты будем считать, что и с той, и с другой стороны p - n перехода выше уровня Ферми отсутствуют заполненные состояния, а ниже уровня Ферми отсутствуют свободные состояния (вообще говоря, это справедливо только при нулевой температуре, но качественной картины такое упрощение не изменит). Таким образом, в отсутствие внешнего напряжения область, содержащая свободные электроны и область, имеющая свободные уровни, не перекрываются по энергиям, т.е. туннельный ток в такой ситуации протекать не может.

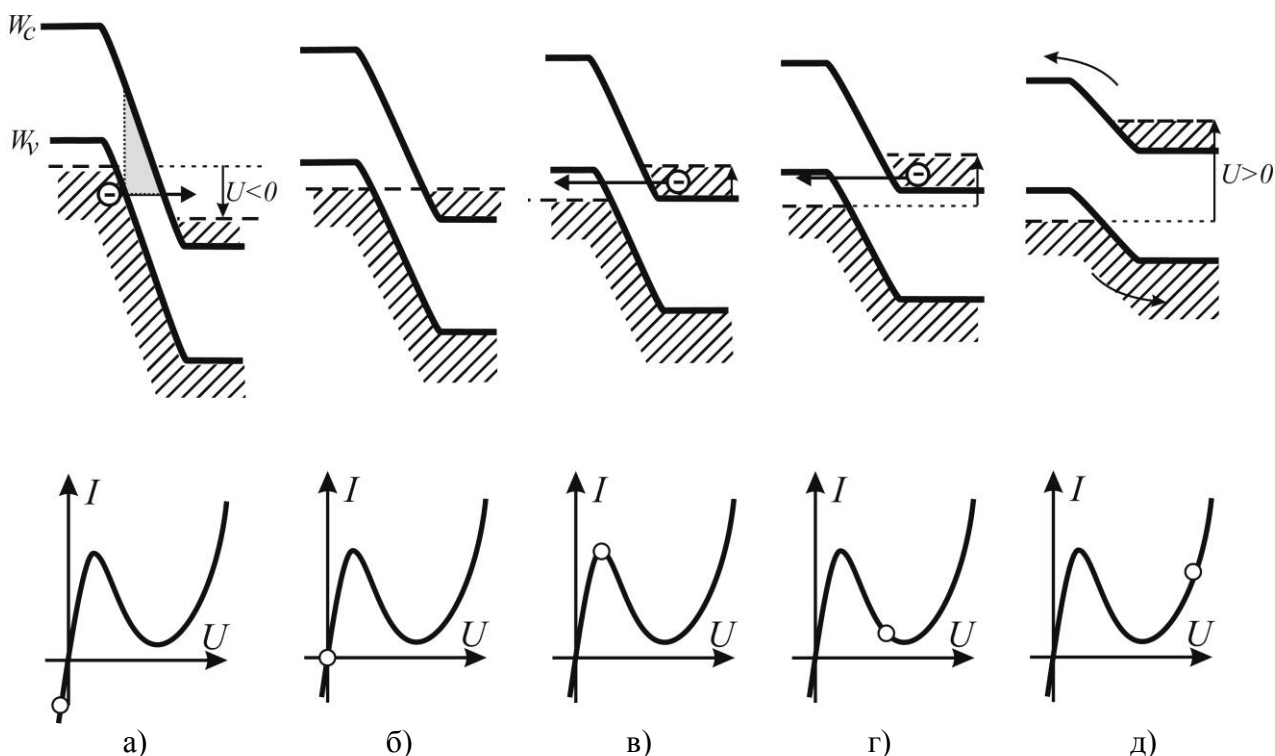


Рис. 3.16. Упрощенная зонная структура туннельного диода: а) - обратное смещение; б) термодинамическое равновесие; в) прямое смещение, при котором возникает пиковый ток; г) прямое смещение, при котором начинается «ток долины»; д) прямое смещение, при котором течет диффузионный ток [2]. Серым тоном выделен треугольник потенциального барьера, преодолеваемого электроном при туннелировании

При приложении внешнего смещения электроны могут туннелировать из валентной зоны в зону проводимости и наоборот (поэтому такой ток иногда называют током «зона – зона»). При этом носители преодолевают потенциальный барьер, который приближенно можно считать треугольным (рис. 3.16 а, треугольник серого тона). Следующие условия являются необходимыми для реализации процесса туннелирования:

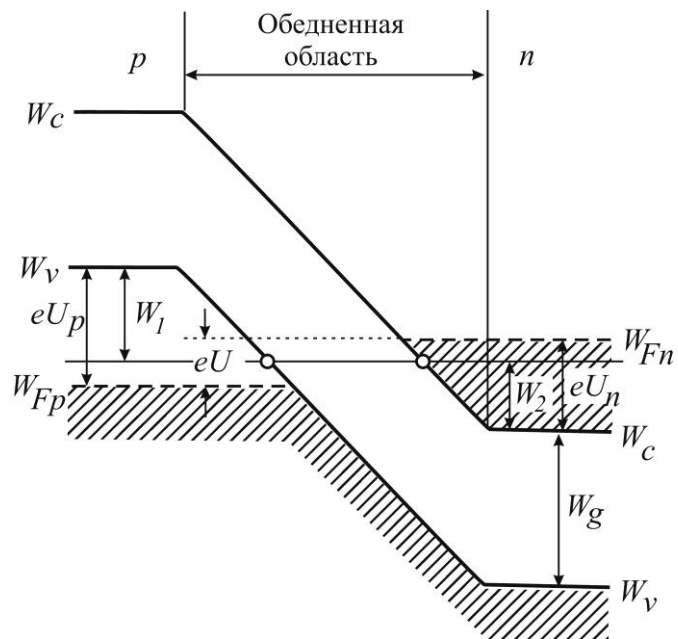
- 1) наличие заполненных состояний с той стороны перехода, откуда электрон туннелирует;
- 2) наличие свободных уровней той же энергии с другой стороны перехода;
- 3) высота и ширина потенциального барьера должны быть достаточно малыми, чтобы существовала конечная вероятность туннелирования электрона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

На рисунке 3.16 а показано туннелирование электрона из валентной зоны в зону проводимости при приложении обратного смещения. Соответствующие этому случаю ток и напряжение отмечены точкой на вольт-амперной характеристике. В случае небольшого прямого смещения существует область значений энергии, для которой имеются заполненные состояния с n -стороны и соответствующие свободные состояния с p -стороны. Причем, когда заполненная и свободная области максимально перекрываются по энергиям, в структуре протекает наибольший ток (рисунок 3.16 в). При дальнейшем увеличении прямого смещения количество свободных уровней с p -стороны, имеющих ту же

энергию, уменьшается (рисунок 3.16 г), а в запрещенной зоне состояний для электронов нет, т.е. переход туда невозможен. Если приложено прямое напряжение такой величины, что зоны «не перекрываются», т.е. дно зоны проводимости и вершина валентной зоны расположены на одном уровне, то в этом случае напротив заполненных состояний n -области лежат запрещенные состояния с p -области (наличием состояний в запрещенной зоне мы, для простоты, пренебрегаем). В этой точке туннелирование носителей прекращается. Однако, полный ток при этом нулю не равняется, поскольку присутствует еще и классическая компонента, обусловленная надбарьерными переходами электронов и дырок. В точке минимума этот ток мал, поскольку потенциальный барьер для носителей еще достаточно высок. При дальнейшем увеличении приложенного напряжения инжекционный ток будет экспоненциально расти (рисунок 3.16 д).

С ростом прямого напряжения туннельный ток увеличивается от нуля до максимума, а затем уменьшается до нуля, когда $U = U_n + U_p$, где U - приложенное напряжение, U_n - степень вырождения с n -стороны, $U_n \equiv (W_{Fn} - W_C)/e$, U_p - степень вырождения с p -стороны, $U_p \equiv (W_V - W_{Fp})/e$, как показано на рисунке 3.17. Обратите внимание (рис. 3.15), что область ВАХ, расположенная между экстремумами тока, является областью отрицательного дифференциального сопротивления. Наличие данной области позволяет использовать туннельный диод для усиления и генерации СВЧ сигналов.

Рис. 3.17. Упрощенная энергетическая диаграмма с постоянным электрическим полем в обедненном слое. Степени вырождения с n - и p -сторон равны eU_n и eU_p . U – напряжение смещения. W_1 и W_2 – энергии электрона, отсчитанные от краев зон в p - и n -областях



Отметим, что туннелирование в структуре может быть прямым и непрямым. Случай прямого туннелирования изображен на рисунке 3.18 а, на котором график зависимости энергии от волнового числа $W(k)$ совмещен с картиной туннельного перехода в W - x координатах. Электрон может туннелировать из окрестности минимума функции, описывающей зависимость энергии от волнового вектора для зоны проводимости, в точку с аналогичным значением квазиимпульса в окрестности максимума функции $W(k)$ валентной зоны. Для того, чтобы имело место прямое туннелирование, необходимо, чтобы минимум зоны

проводимости и максимум валентной зоны имели одинаковое значение импульса. Это условие, как уже упоминалось выше, выполняется в прямозонных полупроводниках.

Непрямое туннелирование происходит тогда, когда абсолютный минимум зоны проводимости и максимум валентной зоны лежат в разных точках зоны Бриллюэна (рисунок 3.18 б). При этом разность импульсов передается рассеивающим агентам, например, фононам. Законы сохранения энергии и импульса при туннелировании с участием фононов: $W_0 + w_{\text{фон}} = W_1$, $p_0 + \hbar k = p_1$, где W_0 – первоначальная энергии электрона, туннелирующего из n -области в p -область, $w_{\text{фон}}$ – энергия фонона, W_1 – конечная энергия электрона после туннелирования, p_0 – начальное значение импульса электрона, $\hbar k$ – импульс фонона, p_1 – импульс электрона после туннелирования. Отметим, что вероятность многофононного непрямого туннелирования много меньше вероятности туннелирования с использованием одного фонона. Кроме того, вероятность непрямого туннелирования существенно меньше, чем вероятность прямого.

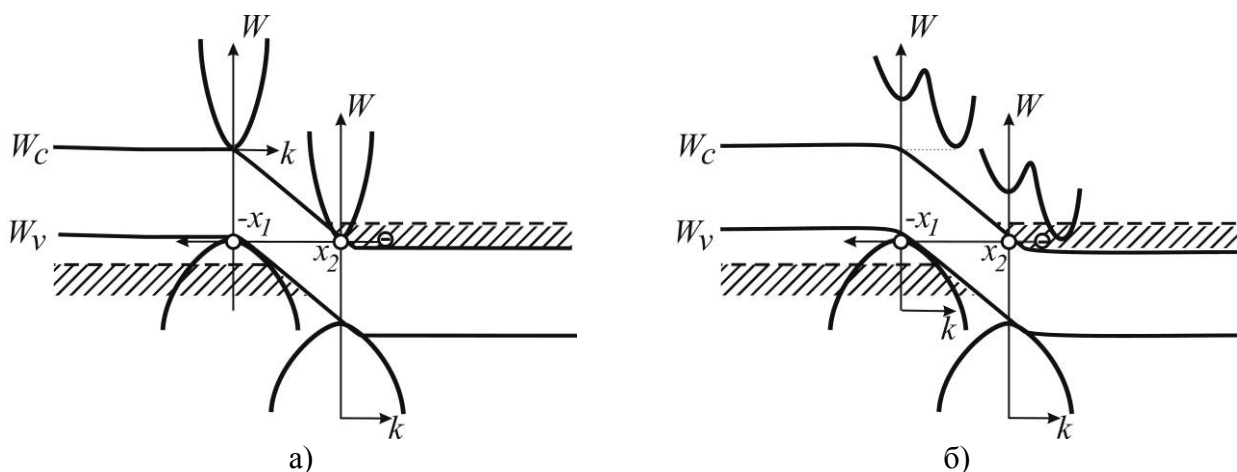


Рис. 3.18. Туннелирование: а) случай прямого туннелирования; б) случай непрямого туннелирования, когда $k_{\text{мин}} \neq k_{\text{макс}}$. Зависимость $W(k)$ наложена в характерных точках ($-x_1$ и x_2) на картину туннельного перехода в $W - x$ координатах

В заключение раздела приведем основную эквивалентную схему туннельного диода (рис. 3.19). Она состоит из четырех элементов: последовательной индуктивности L_S , последовательного сопротивления R_S , диодной емкости C и отрицательного сопротивления диода $-R$. Последовательное сопротивление R_S включает сопротивление вывода, омических контактов и сопротивление кристалла.

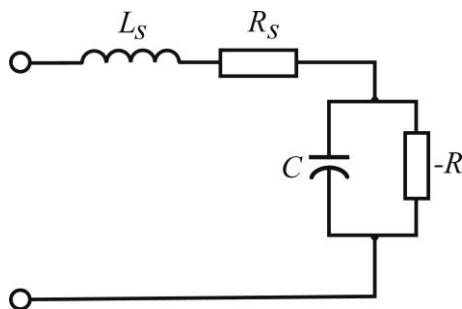


Рис. 3.19. Эквивалентная схема туннельного диода

3.3.2. Туннельно-резонансный диод

Структура туннельно-резонансного диода (ТРД) и его зонные диаграммы в равновесном случае и при смещении показаны на рис. 3.20.

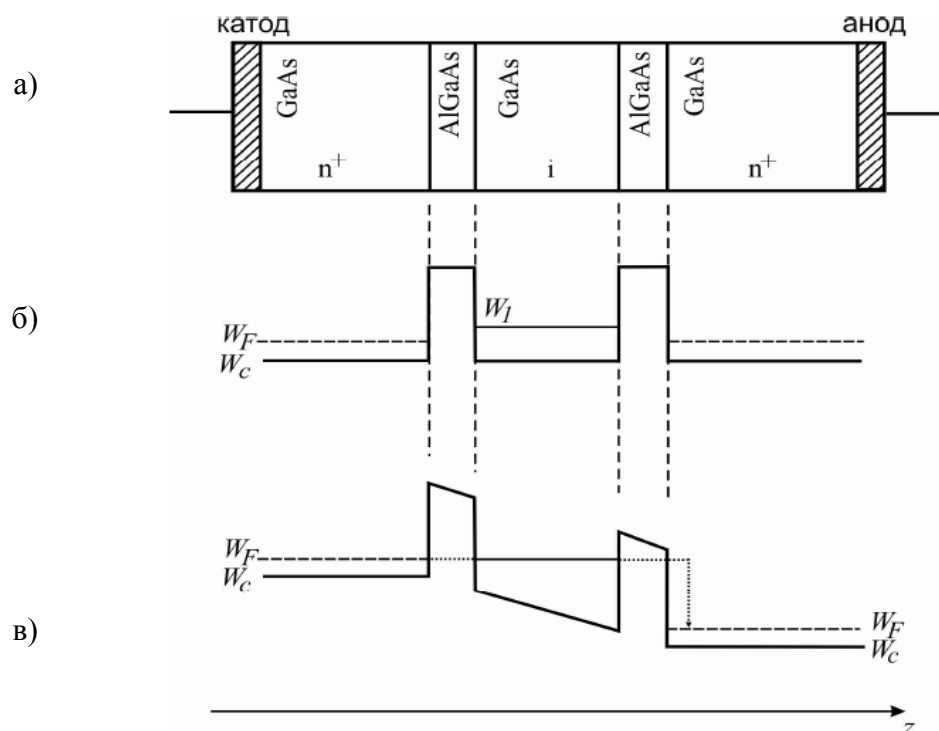


Рис. 3.20. Туннельно-резонансный диод на основе гетероструктуры GaAs/AlGaAs: а) структура диода; б) профиль дна зоны проводимости в отсутствие смещения; в) профиль дна зоны проводимости при подаче внешнего напряжения

ТРД состоит из трех слоев узкозонного полупроводника (на рисунке – GaAs), разделенных двумя слоями широкозонного материала (AlGaAs). Примыкающие к контактам сильно легированные области слева и справа от барьеров представляют собой «резервуары носителей заряда». Эти слои легируются настолько сильно, что электронный газ в таких областях становится вырожденным, а уровень Ферми лежит в зоне проводимости. Активная область диода состоит из двух потенциальных барьеров, между которыми расположена квантовая яма, содержащая один или несколько уровней для электронов (рис. 3.20 б). Прозрачность квантовой системы «барьер-яма-барьер» носит резонансный характер, откуда и берется название прибора. Если энергия налетающих носителей совпадает с энергией дискретного уровня в яме, то вероятность прохождения сквозь два барьера резко возрастает. Такая ситуация возникает при подаче напряжения на диод, когда квантовая яма опускается, как показано на рисунке 3.20 в и ток диода возрастает (см. рис.3.21 а). При напряжениях питания больше, чем высота барьера eU_0 происходит надбарьерная термоэмиссия электронов и ток увеличивается экспоненциально, как и в случае обычного рп-перехода.

Когда в яме существует несколько энергетических уровней, на вольт-амперной характеристике будет несколько резонансных пиков тока, каждый из которых соответствует своему уровню (см. рис.3.21 б). Если расстояние между соседними уровнями в яме меньше, чем расстояние от дна зоны проводимости до уровня Ферми в контактных слоях диода, то пики тока на ВАХ будут перекрываться, так как будет возможно одновременное туннелирование электронов через несколько уровней.

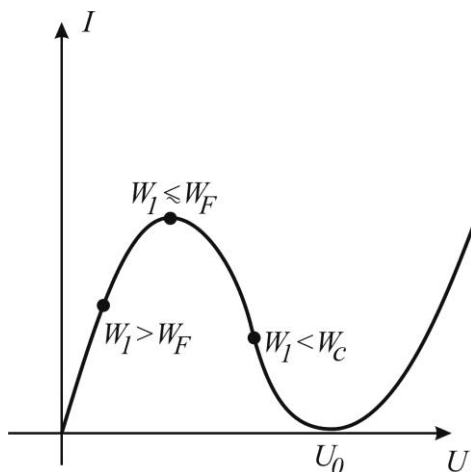


Рис. 3.21 а. ВАХ туннельно-резонансного диода с одним энергетическим уровнем в квантовой яме

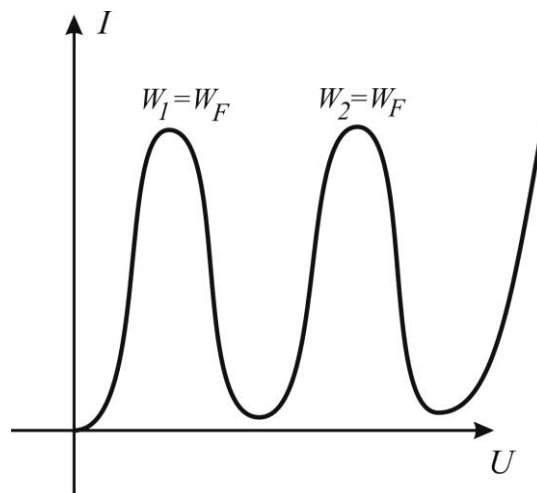


Рис. 3.21 б. ВАХ туннельно-резонансного диода с несколькими энергетическими уровнями в квантовой яме

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

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

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

3.3.3. Сверхрешётки

Сверхрешетками называют слоистые периодические структуры, образованные путем чередования тонких слоев полупроводников. По способу формирования сверхрешетки делятся на *композиционные* и *легированные*. Первые представляют собой чередующиеся слои полупроводников с различной шириной запрещенной зоны, вторые - чередующиеся слои *n*- и *p*-типов одного и того же полупроводника. В обоих случаях на электроны, кроме обычного потенциала кристаллической решётки, действует дополнительный одномерный периодический потенциал, называемый *потенциалом сверхрешетки*. Его период d меньше длины свободного пробега электронов, но значительно больше периода самой кристаллической решетки (1...10 нм). Таким образом, мы имеем периодическую систему квантовых ям, разделенных узкими, имеющими значительную туннельную прозрачность, потенциальными барьерами. В качестве примера на рисунке приведена композиционная сверхрешетка, изготовленная на основе широко используемой гетероструктуры GaAs-Al_xGa_{1-x}As.

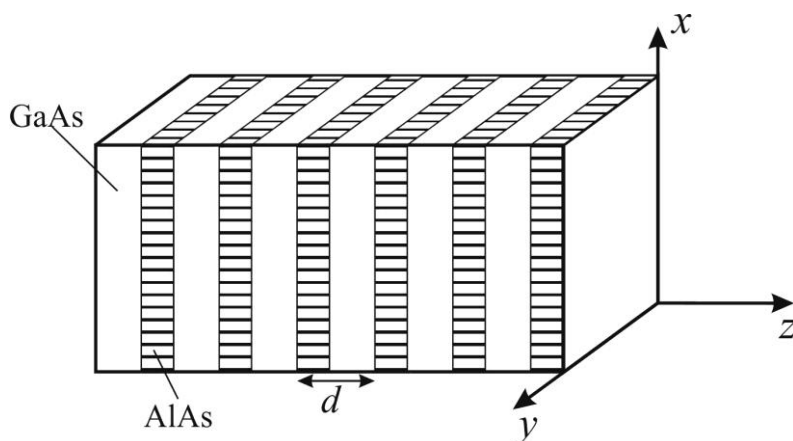


Рис. 3.22. Схематическая структура сверхрешетки GaAs/AlAs. z — продольная ось сверхрешетки, вдоль которой пропускают ток

Наличие дополнительного периодического потенциала, зависящего только от одной координаты, делает энергетический спектр электронов и дырок резко анизотропным. Поскольку потенциальные барьеры туннельно-прозрачны, энергетические уровни в квантовых ямах расщепляются и возникает одна или несколько разрешенных энергетических *минизон* (рис. 3.23). При этом в плоскости, перпендикулярной оси сверхрешетки, носители ведут себя как свободные частицы с соответствующей эффективной массой. Таким образом, сверхрешетка является искусственным одномерным кристаллом.

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

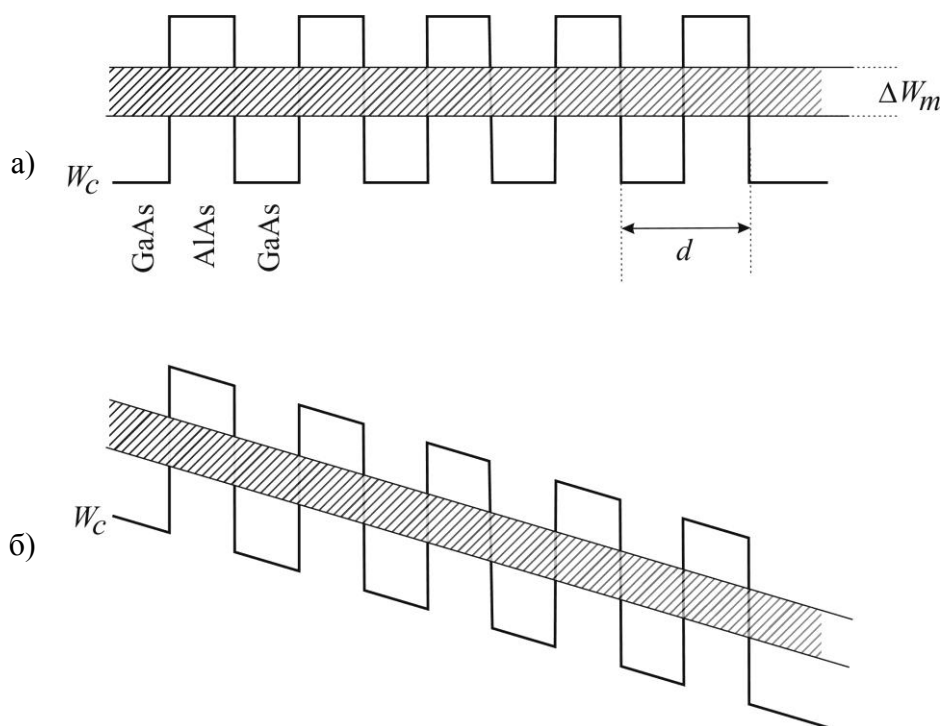


Рис. 3.23. Зонная диаграмма сверхрешетки: а) в отсутствие внешнего напряжения; б) при подаче внешнего напряжения. ΔW_m – ширина минизоны

При подаче на структуру внешнего напряжения минизона наклоняется (рис. 3.23 б) и в ней протекает ток электронов, как и в обычных разрешенных зонах полупроводника. При этом возможна реализация двух вариантов движения электронов.

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

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

Блоховские осцилляции возникают благодаря тому, что в нижней половине минизоны масса электронов положительна, а в верхней – отрицательна, также как и в обычных полупроводниковых кристаллах. В результате электрон отражается от потолка разрешённой зоны, хотя электрическое поле всегда направлено в одну сторону. В нижней половине минизоны электрон ускоряется электрическим полем, а в верхней – тормозится и разворачивается по полю. Период блоховских осцилляций должен быть меньше времени свободного пробега электронов, иначе осцилляций не будет, а будет только рассеяние, как это показано на рисунке 3.24 а.

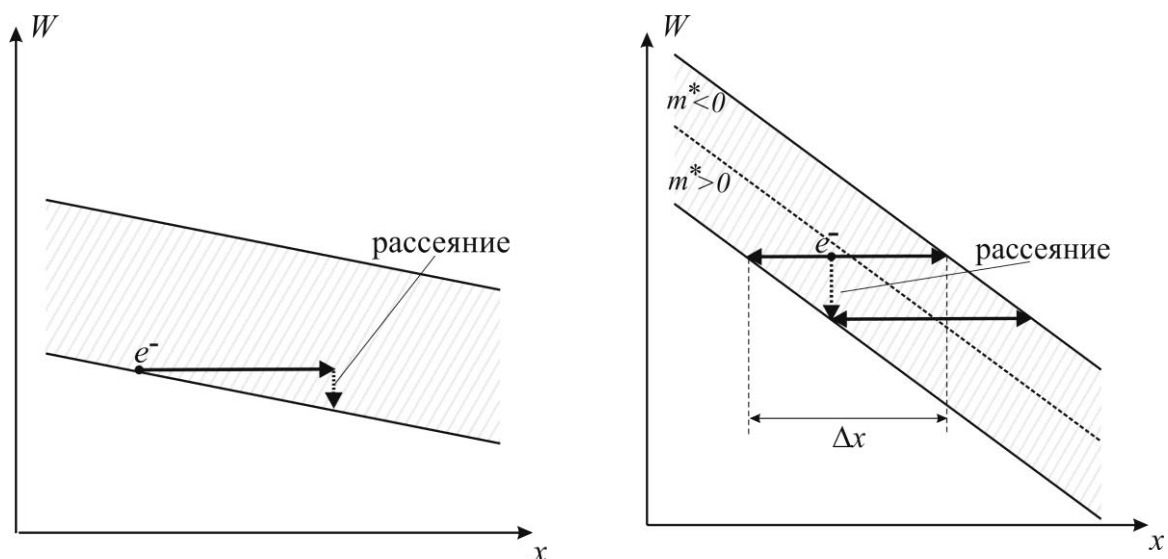


Рис. 3.24. Схема движения электронов в минизоне сверхрешетки: а) при малых напряжениях питания; б) при больших напряжениях питания. Пунктиром со стрелкой изображен процесс рассеяния, переводящий электрон на более низкий энергетический уровень

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

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

СПИСОК ЛИТЕРАТУРЫ

1. Оболенский С.В., Демарина Н.В., Волкова Е.В. Основы физики полупроводников. Транспорт носителей заряда в электрических полях: учебное пособие. – Нижний Новгород, Издательство Нижегородского госуниверситета, 2007. – 79 с.
2. Зи С.М. Физика полупроводниковых приборов. – М. Мир, 1981. – 567 с.
3. Пожела Ю. Физика быстродействующих транзисторов. – Вильнюс: Мокслас, 1989. – 264 с.
4. Степаненко И.П. Основы Микроэлектроники. – М.: Лаборатория базовых знаний, 2001. – 488с.
5. Шур М. Современные приборы на основе арсенида галлия. – М.: Мир, 1991. – 632 с.
6. Полевые транзисторы на арсениде галлия / ред. Ди Лоренцо Д.В., Канде-луола Д.Д. – М.: Радио и связь, 1988. – 496 с.
7. Бонч-Бруевич В.Л., Калашников С.Г. Физика полупроводников. – М.: Наука, 1990. – 559 с.
8. Технология СБИС / Пирс К., Адамс А., Кац Л., Цай Дж., Сейдел Т., Макгиллис Д.; под редакцией Зи С. – М: Мир, 1986. – 404 с.

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

ПОЛУПРОВОДНИКОВЫЕ ДИОДЫ

Учебное пособие

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

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

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