

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение
высшего профессионального образования
«Томский государственный университет систем управления и
радиоэлектроники»

Кафедра электронных приборов

ТВЕРДОТЕЛЬНЫЕ ПРИБОРЫ И УСТРОЙСТВА

Учебное пособие
для студентов направления «Фотоника и оптоинформатика» и
«Электроника и наноэлектроника» (специальность «Электронные
приборы и устройства»)

2012

Шангин, Александр Сергеевич

Твердотельные приборы и устройства: учебное пособие для студентов направления «Фотоника и оптоинформатика» и «Электроника и нанoeлектроника» (специальность «Электронные приборы и устройства») / А.С. Шангин. Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования Томский государственный университет систем управления и радиоэлектроники, Кафедра электронных приборов. - Томск : ТУСУР, 2012. – 12 с.

В данном учебном пособии изложены основные сведения о твердотельных приборах и устройствах, а также принципы действия и свойства наиболее распространенных твердотельных и полупроводниковых приборов и элементов интегральных схем.

Пособие предназначено для студентов очной формы и заочной формы обучения, обучающихся по направлению «Фотоника и оптоинформатика» и «Электроника и нанoeлектроника» (специальность «Электронные приборы и устройства») по дисциплине «Твердотельные устройства», «Твердотельные приборы» и «Твердотельные приборы и устройства».

© Шангин Александр Сергеевич, 2012

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Томский государственный университет систем управления и
радиоэлектроники»

Кафедра электронных приборов

УТВЕРЖДАЮ
Зав.кафедрой ЭП
_____ С.М. Шандаров
«__» _____ 2012 г.

ТВЕРДОТЕЛЬНЫЕ ПРИБОРЫ И УСТРОЙСТВА

Учебное пособие для студентов направления «Фотоника и
оптоинформатика» и «Электроника и наноэлектроника» (специальность
«Электронные приборы и устройства»)

Разработчик

_____ А.С. Шангин
«__» _____ 2012 г.

2012

Содержание

| | |
|---|-----|
| 1. Введение | 5 |
| 2. Энергетические зоны полупроводников | 6 |
| 2.1. Простая модель энергетических зон | 6 |
| 2.2. Математическая модель энергетических зон | 10 |
| 3. Носители заряда в полупроводниках | 19 |
| 3.1.Равновесное состояние полупроводника | 19 |
| 3.2.Нерановесные процессы в полупроводниках | 38 |
| 3.3 Основные уравнения, описывающие процессы в полупроводниковых приборах | 49 |
| 4. Физические явления в р-п-переходе | 50 |
| 4.1. Равновесное состояние р-п перехода | 50 |
| 4.2. Работа р-п- перехода при внешнем напряжении | 56 |
| 4.3. Явления пробоя | 63 |
| 4.4. Малосигнальные модели диода | 66 |
| 4.5. Процессы переключения в диоде. Режим большого сигнала | 69 |
| 4.6. Переход металл-полупроводник | 71 |
| 4.7. Переход полупроводник-полупроводник (гетеропереход) | 73 |
| 4.8. Диоды для оптоэлектроники | 74 |
| 5. Физические принципы работы полевых транзисторов | 81 |
| 5.1. Введение в теорию идеального МОП- конденсатора | 82 |
| 5.2. Идеальный МОП-конденсатор | 91 |
| 5.3. Реальный МОП – конденсатор | 94 |
| 5.4. Идеальный МОП- транзистор | 97 |
| 5.5. Реальный МОП- транзистор | 102 |
| 5.6. Полевой транзистор с управляющим р-п- переходом | 103 |
| 5.7. Полевой транзистор с управляющим переходом металл - полупроводник | 110 |
| 6. Физические принципы работы транзистора и тиристора | 111 |
| 6.1 Структура биполярного транзистора и принцип его работы | 112 |
| 6.2. Параметры транзистора | 113 |
| 6.3. Частотные свойства транзистора | 117 |
| 6.4. Явления, наблюдаемые при высоком уровне инжекции | 118 |
| 6.5. Статистическая модель биполярного транзистора | 120 |
| 6.6. Модели транзистора в режиме малого сигнала | 126 |
| 6.7. Зарядовая модель биполярного транзистора | 132 |
| 6.8. Транзистор в режиме переключения | 134 |
| 6.9. Тиристоры | 135 |
| 7. Биполярные интегральные схемы | 137 |
| 7.1. Биполярные транзисторы ИС | 138 |
| 7.2. Диоды ИС | 143 |
| 7.3.Электрические модели биполярных ИС | 144 |
| 7.4 Применение электротермических моделей | 147 |
| 7.5. Семейства биполярных логических ИС | 147 |
| 7.6. Основные типы аналоговых биполярных ИС | 151 |

1. ВВЕДЕНИЕ

Предмет электроники и микроэлектроники

Электроникой называют науку, изучающую электронные свойства некоторых твердых тел, а также методы получения материалов с такими характеристиками, которые позволяют создавать устройства для передачи и накопления электронов. Второе определение характеризует физическую электронику с точки зрения процессов, происходящих на микроскопическом уровне.

Электроникой занимается:

- 1) исследованием физических явлений в приборах, действие которых основано на протекании электрического тока в твердом теле, вакууме или газе;
- 2) изучением электрических свойств, характеристик и параметров названных приборов;
- 3) разработкой и практическим применением этих приборов в различных устройствах и системах.

Так как в проводящих средах ток создается электронами, или преимущественно электронами, то при общем определении упомянутых выше приборов все они могут быть названы *электронными приборами*. Однако в связи с различными средами, в которых электроны перемещаются, различают следующие основные классы электронных приборов: *электровакуумные, газоразрядные (ионные), полупроводниковые*.

Основными областями современного развития технической электроники являются:

- 1) *радиоэлектроника*, связанная с разработкой и применением электронной аппаратуры для целей радиовещания, радиолокации, радионавигации, телевидения, радиоастрономии;
- 2) *промышленная электроника*, обеспечивающая внедрение электронных устройств во все отрасли производства, науки, техники, быта как для электропитания, так и с целью измерения, контроля, регулирования, защиты и управления различными промышленными объектами и автоматизации технологических и производственных процессов.

Современным этапом развития электроники является *микроэлектроника*, связанная с микроминиатюризацией электронной аппаратуры.

Микроэлектроника – это раздел электроники, охватывающий исследование и разработку интегральных микросхем и электронной аппаратуры на их основе.

Интегральная микросхема (интегральная схема, микросхема) – микроэлектронное функционально законченное изделие, имеющее высокую плотность упаковки электрически соединенных элементов (транзисторов, диодов, конденсаторов, резисторов и т.д.), изготовленное в едином технологическом цикле и рассматриваемое как единое целое с точки зрения требований к испытаниям, приемке, поставке и эксплуатации.

Характерной особенностью микроэлектроники является органическое единство всех трех её основных разделов: физики, технологии, микросхемотехники.

Микросхемотехника (интегральная схемотехника) представляет собой раздел микроэлектроники, охватывающий исследования и разработку схемотехнических решений (электрических и структурных схем), используемых в интегральных схемах и аппаратуре на их основе.

Современное состояние электроники

Современный этап развития электроники характеризуется созданием широкой номенклатуры и массовым выпуском интегральных микросхем, запоминающих устройств, микропроцессорных комплектов, однокристалльных ЭВМ, RISC-процессоров. Стремительно развивается рынок заказных больших интегральных схем на базе матричных кристаллов, программируемых логических матриц. Растет номенклатура специализированных БИС интерфейсов локальных вычислительных сетей.

Любые новые направления в электронике непременно будут связаны с технологией изготовления интегральных схем (ИС).

2. ЭНЕРГЕТИЧЕСКИЕ ЗОНЫ ПОЛУПРОВОДНИКОВ

Потенциальная энергия и импульс электрона в твердом теле определяются эффектами квантовых взаимодействий. Модель энергетических зон вводится для того, чтобы связать между собой энергию и импульс носителей заряда в кристалле.

Модель энергетических зон будет рассматриваться в двух аспектах. Прежде всего, данная модель позволяет качественно описать распределение энергии электронов в твердых телах и, кроме того, найти аналитическое выражение для энергии электронов в пространственно не ограниченном твердом теле путем решения уравнения Шрёдингера.

2.1. Простая модель энергетических зон

В некотором отношении зонная модель ведет свое начало от атомной модели Бора, которая позволяет объяснять электрические свойства жидкостей и газов. На смену модели Бора пришла квантово-механическая модель, согласно которой электрон, входящий в состав атома какого-либо вещества, рассматривается как некоторая заряженная частица, свойства которой можно определить, решив уравнение Шрёдингера

$$\nabla^2 \psi(x, y, z) + \frac{8\pi^2 m}{h^2} [E - V(x, y, z)] \psi(x, y, z) = 0 \quad (2.1)$$

где

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

$\psi(x, y, z)$ - волновая функция; E - полная энергия электрона; $V(x, y, z)$ - потенциальная энергия электрона; h - постоянная Планка.

Уравнение (2.1) имеет отличные от нуля решения только при некоторых дискретных значениях энергии E (это будет доказано позднее в рамках аналитической зонной модели). В этих случаях говорят, что имеет место некоторое *квантовое состояние*. Каждому квантовому состоянию отвечает некоторое определенное значение энергии электрона E , а также его импульса mv или hk . Для описания квантовых состояний используют набор квантовых чисел n , l , m и m_s , последнее из которых указывает значение спина электрона. Итак, электрон, входящий в состав свободного атома, характеризуется некоторым набором квантовых чисел и дискретных значений энергии.

Поскольку уравнение (2.1) имеет отличные от нуля решения только при некоторых дискретных значениях E , говорят, что электрон занимает некоторые разрешенные энергетические уровни (s, p, d, \dots). В соответствии с известным из квантовой механики принципом запрета Паули два электрона в любом атоме не могут находиться в одном и том же квантовом состоянии (т. е. описываться одной и той же волновой функцией и иметь один и тот же спин). По этой причине электроны в атоме не могут обладать одинаковыми наборами квантовых чисел (за исключением спинового числа); число электронов, находящихся на одном и том же энергетическом уровне, не может превышать числа квантовых состояний.

Часто используется также концепция *электронных оболочек* K, L, M, \dots , образованных одним или несколькими энергетическими уровнями с одним и тем же значением квантового числа.

Следует принять во внимание высокую концентрацию атомов в твердых кристаллических телах. При этом из-за близкого расположения атомов между ними возникает взаимодействие, и потенциальная энергия электронов становится периодической функцией пространственных координат.

Вследствие этого энергетические уровни любого атома становятся связанными с уровнями других атомов. Происходит расщепление каждого уровня на такое число новых уровней, которое совпадает с числом взаимодействующих атомов. Поскольку межатомные расстояния достаточно велики, отдельные уровни энергии оказываются очень близкими и образуют так называемые энергетические зоны. Это явление напоминает известное расщепление резонансной частоты, наблюдаемое в системе двух связанных колебательных контуров LC

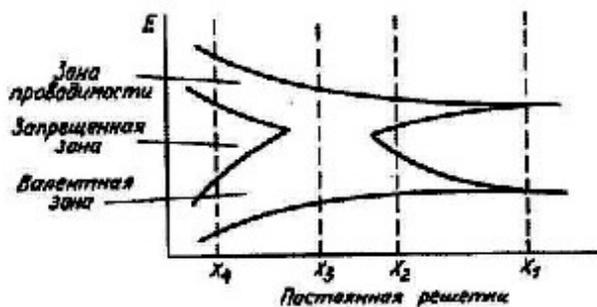


Рис. 2.1. Зонная модель решетки углерода при абсолютной температуре $T = 0 \text{ K}$, построенная в зависимости от межатомного расстояния

Чтобы пояснить процесс образования энергетических зон, рассмотрим вначале некоторый свободный атом углерода. В таком атоме имеются две электронные оболочки. Оболочка K содержит две подоболочки с двумя и шестью уровнями соответственно. Первая из них заполнена, а на втором из шести возможных уровней электронами занято только два. Поэтому четыре уровня остаются свободными. Такой структуре свободного атома углерода отвечает межатомное расстояние x_1 на рис. 2.1.

Из этого рисунка видно, что если атомы в веществе сближаются и уже не могут рассматриваться как свободные, то происходит трансформация подболочек. При расстоянии между атомами x_2 эти подболочки превращаются в две зоны, при x_3 - две зоны сливаются в одну, как у графита. Расстоянию x_4 соответствуют три зоны, что характерно для особой кристаллической формы углерода - алмаза: *валентная зона*, образованная четырьмя валентными электронами (два электрона на уровне $2s$ и два электрона на уровне $2p$), *зона проводимости* с четырьмя свободными вакансиями и, наконец, *запрещенная зона*, где электронов быть не может. *Наличие запрещенной зоны между двумя разрешенными зонами (зоной проводимости и валентной зоной) характерно для всех полупроводников. Этот факт имеет большое значение для физической электроники и для теории полупроводниковых устройств*

В соответствии с изложенным зонная модель кристаллической решетки углерода (алмаза) может быть представлена в упрощенной форме, изображенной на рис. 2.2. При этом считается, что каждый атом, входящий в состав кристалла, имеет четыре электрона в валентной зоне. Кроме того, каждому атому отвечает четыре вакантных уровня, находящихся в зоне проводимости. Ширину запрещенной зоны E_g выражают в электрон-вольтах (эВ).



Рис. 2.2. Простая модель энергетических зон

При нулевой абсолютной температуре зона проводимости пуста, а валентная зона заполнена целиком, так что электропроводность отсутствует. Единственная возможность создания

электропроводности заключается в том, чтобы сообщить электрону некоторую энергию, большую или равную E_g , за счет чего электрон попадает в зону проводимости. Этого можно достичь, например, путем нагрева или облучения. При комнатной температуре (300К) энергия E_g составляет 1,12 эВ для кремния, 0,72 эВ для германия и 7 эВ для углерода (алмаза).

Введя понятия разрешенных зон и запрещенной энергетической зоны, можно провести ряд последовательных рассуждений, которые подтверждают их существование.

- Электрон как материальная частица проявляет волновые свойства.
- Решив волновое уравнение, описывающие поведение электрона, получаем числовые значения параметров, называемых квантовыми числами.
- Упорядоченное расположение атомов в кристаллической решетке обуславливает то, что электрон может находиться на одном из разрешенных дискретных уровней; дополнительное ограничение налагает принцип Паули.

В зависимости от того, как расположены энергетические зоны, твердые тела принято делить на диэлектрики, полупроводники и металлы (рис 2.3).

Если ширина запрещенной зоны достаточно велика, то ни один из электронов, находящихся в валентной зоне, не может за счет теплового возбуждения получить порцию энергии, достаточную для перехода в зону проводимости.

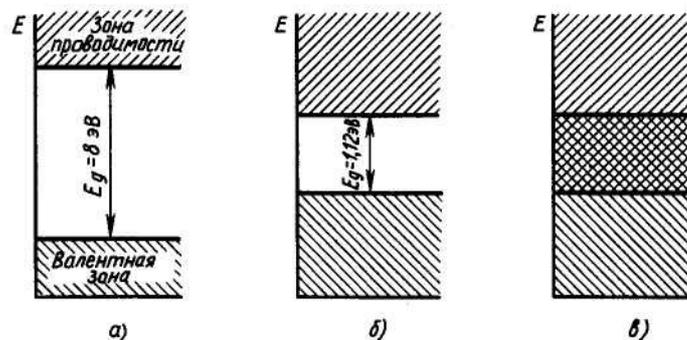


Рис. 2.3. простые модели энергетических зон различных материалов: а диэлектрики (SiO_2); б- полупроводник (Si); в- металл (Al).

Такие вещества имеют высокое электрическое сопротивление и называют *диэлектриками*.

Вещество называется *полупроводником*, если параметр E_g достаточно мал. При этом имеется конечная вероятность того, что электрон, ранее занимающий некоторый уровень в верхней части валентной зоны, приобретет дополнительную энергию, достаточную для того, чтобы скачком преодолеть интервал E_g и перейти в нижнюю часть зоны проводимости, которая до этого была пуста.

Реально механизм возникновения электропроводности в полупроводниках более сложен и, строго говоря, для его анализа требуется привлечь статистику Ферми-Дирака. Чтобы описать физические свойства полупроводников, нужно прежде всего определить структуру энергетических зон и вычислить плотность квантовых уровней в каждой зоне. Затем, используя статистику Ферми-Дирака, следует найти концентрацию носителей в каждой зоне.

В *металлах* из-за слабой связи валентных электронов со своими атомами валентная зона и зона проводимости перекрываются. Наложение внешнего электрического поля приводит к возникновению электропроводности даже при нулевой абсолютной температуре. Перемещаясь вдоль кристаллической решетки, электроны постоянно находятся в зоне проводимости.

Во многих случаях оказывается полезной упрощенная модель энергетических зон без указания числовых значений энергии. Однако следует иметь в виду, что эта модель является не более чем условным графическим изображением, которое облегчает изучение процесса движения носителей в полупроводниках. Энергетические зоны следует трактовать не как некие «каналы», в которых происходит движение электронов и дырок, а лишь как совокупность уровней энергии; при одних ее значениях электропроводность возможна, при других - невозможна.

Следует заметить, что на диаграмме энергетических зон ось отсчета направлена снизу вверх. Так как заряд электрона отрицателен, то значения, откладываемые по этой оси, численно равны электрическому потенциалу.

2.2. Математическая модель энергетических зон

Чтобы построить модель энергетических зон в общем случае, требуется количественно оценить уровни энергии, что связано с весьма сложными расчетами. Тем не менее удастся проанализировать диаграмму энергетических зон и получить важные сведения о квантово-механических свойствах электронов, обратившись к достаточно простой модели, известной под названием *модели Кронига-Пенни*.

Одна из задач зонной теории твердого тела состоит в том, чтобы математически описать поведение заряженной частицы (электрона) в кристалле. При этом микроскопический объект рассматривается как свободная частица с некоторой так называемой эффективной массой. Математическая модель должна учитывать большинство эффектов, обусловленных влиянием кристаллической решетки; ряд эффектов учитывается путем соответствующего выбора параметров, которые применяются для описания частицы.

Перемещаясь вблизи атома, электрон испытывает ускорения и замедления под действием электрического поля атома. Этот процесс является периодическим, так как периодическим является распределение потенциала в идеальном кристалле. Вычисления упрощаются, если предположить, что график функции, описывающей этот потенциал, имеет прямоугольную форму.

Модель Кронига-Пенни

Из квантовой механики известно, что дискретные уровни энергии электрона, находящегося в потенциальной яме шириной a с бесконечно высокими стенками, определяются равенством

$$E_n = \frac{1}{2m} \left(\frac{n\pi}{a} \right)^2, \quad n = 1, 2, 3, \dots \quad (2.2)$$

Рассмотрим аналитически задачу о поведении электрона, находящегося в периодической решетке (решетке Бравэ), предполагая, что кристалл является одномерным и неограниченно протяженным, а распределение потенциала в пространстве имеет прямоугольную форму (рис. 2.10). Уровни энергии при этом находятся в приближении одиночного электрона.

Волновая функция, описывающая поведение электрона, является решением уравнения Шрёдингера, которое не содержит время в явном виде:

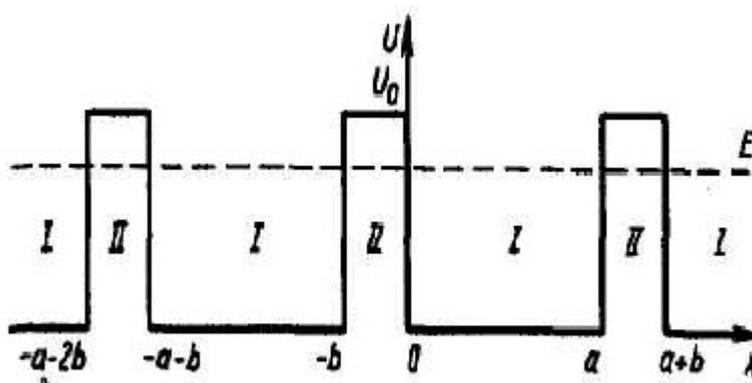


Рис. 2.4. Модель Кронига-Пенни с периодическим потенциалом прямоугольной формы: $a+b = L$ —период; a —ширина области I с нулевым потенциалом; b —ширина области II

$$\frac{\partial^2 \psi(x)}{\partial x^2} + \frac{2m}{\hbar^2} [E - V(x)] \psi(x) = 0 \quad (2.3)$$

Простейшее решение уравнения (2.3), имеющее четкий физический смысл, представляет плоскую волну с некоторым волновым числом k , промодулированную по амплитуде с помощью некоторой функции, период которой совпадает с периодом кристаллической решетки. Такие решения называют функциями Блоха; их существование вытекает из теоремы

Флоке, известной в теории дифференциальных уравнений с периодическими коэффициентами:

$$\psi(x) = U(x)e^{jkx} \quad (2.4)$$

где $U(x) = U(x + L) = U(x + NL)$; N – порядковый номер атома в кристалле. Очевидно, что $U(a) = U(-b)$.

$$\psi_1 = U_1 e^{jkx} \quad (2.5)$$

$$\frac{d\psi_1}{dx} = jkU_1 e^{jkx} + \frac{dU_1}{dx} e^{jkx} \quad (2.6)$$

$$\frac{d^2\psi_1}{dx^2} = -k^2 U_1 e^{jkx} + 2jk \frac{dU_1}{dx} + \frac{d^2 U_1}{dx^2} e^{jkx} \quad (2.7)$$

$$\alpha^2 = 2mE / \hbar^2 \quad (2.8)$$

$$\frac{d^2 U_1}{dx^2} + 2jk \frac{dU_1}{dx} + (\alpha^2 - k^2) U_1 = 0 \quad (2.9)$$

В области II потенциал $U(x) = V_0$. Обозначая:

$$\beta^2 = 2m(V_0 - E) / \hbar^2 \quad (2.10)$$

получаем соответствующее волновое уравнение:

$$\frac{d^2 U_2}{dx^2} + 2jk \frac{dU_2}{dx} + (\beta^2 + k^2) U_2 = 0 \quad (2.11)$$

Решения данных уравнений имеют следующий вид:

$$U_1 = A e^{j(\alpha-k)x} + B e^{-j(\alpha+k)x} \quad (2.12)$$

$$U_2 = C e^{(\beta-jk)x} + D e^{-(\beta+jk)x} \quad (2.13)$$

где A, B, C, D - некоторые постоянные коэффициенты.

Функции U и dU/dx должны быть непрерывными на концах интервала и, кроме того, должно выполняться равенство $U(a) = U(-b)$, так как функция U является периодической. Таким образом, граничные условия принимают вид:

$$\begin{aligned}
U_1(0) &= U_2(0) \\
\left. \frac{dU_1}{dx} \right|_{x=0} &= \left. \frac{dU_2}{dx} \right|_{x=0} \\
U_1(a) &= U_2(-b) \\
\left. \frac{dU_1}{dx} \right|_{x=a} &= \left. \frac{dU_2}{dx} \right|_{x=-b}
\end{aligned}$$

Отсюда получаем следующую систему линейных алгебраических уравнений относительно неизвестных коэффициентов:

$$A + B = C + D$$

$$j(\alpha - k)A - j(\alpha + k)B = (\beta - jk)C - (\beta + jk)D$$

$$Ae^{j(\alpha-k)a} + Be^{-j(\alpha+k)a} = Ce^{-(\beta-jk)b} + De^{(\beta+jk)a} \quad (2.14)$$

$$\begin{aligned}
&j(\alpha - k)Ae^{j(\alpha-k)a} - j(\alpha + k)Be^{-j(\alpha+k)a} = \\
&= (\beta - jk)Ce^{-(\beta-jk)b} - (\beta + jk)De^{(\beta+jk)a}
\end{aligned}$$

Перенеся все неизвестные величины в левые части выражений (2.14), приходим к системе однородных уравнений относительно A , B , C и D . Для разрешимости системы необходимо потребовать, чтобы ее определитель был равен нулю. Выполнив довольно утомительные преобразования, приходим к выводу, что для этого необходимо потребовать выполнения равенства:

$$\frac{\beta^2 - \alpha^2}{2\alpha\beta} \operatorname{sh}\beta b \cdot \sin \alpha a + \operatorname{ch}\beta b \cos \alpha a = \cos kL = \cos \frac{2n\pi}{N} \quad \text{при } E < V_0 \quad (2.15)$$

Здесь принято во внимание, что $L = a + b$ и что $k = 2n\pi / NL$.

Полученное тригонометрическое уравнение определяет допустимые уровни энергии, зависящие от величины α и β . Решить его можно как графическим, так и численным методом.

Правая часть уравнения (2.15) является периодической функцией переменной k . Значения переменной k изменяются дискретно с весьма малым шагом. Любому значению L соответствует столько дискретных шагов, каково число атомов N в кристалле. Так как это число очень велико, можно с полным основанием считать, что k является непрерывной переменной.

Следует заметить, что функция в правой части уравнения (2.15) может принимать значения, лежащие лишь в интервале от +1 до -1. Это означает, что при фиксированном значении V_0 существуют разрешенные уровни энергии E . Однако возможны и другие значения E , при которых

модуль левой части уравнения оказывается больше единицы. Очевидно, что энергия электрона не может принимать указанных значений, соответствующих *запрещенным энергетическим зонам*.

Чтобы выполнить графическое построение разрешенных и запрещенных зон, необходимо рассмотреть также случай $E > V_0$. При этом уравнение (2.15) сохраняет свою форму, однако параметр P становится мнимым. Поэтому, обозначив P символом $j\gamma$ и приняв во внимание, что $\text{sh}\beta b = \text{sh}(j\gamma b) = j \sin \gamma b$, $\text{ch}\beta b = \text{ch}(j\gamma b) = \cos \gamma b$, запишем (2.15) в виде:

$$-\frac{\alpha^2 + \gamma^2}{2\alpha\beta\gamma} \sin \gamma b \cdot \sin \alpha a + \cos \gamma b \cdot \sin \alpha a = \cos kL = \cos \frac{2n\pi}{N} \quad \text{при } E > V_0 \quad (2.16)$$

Здесь также можно выделить разрешенные и запрещенные энергетические зоны.

Уравнения (2.15) и (2.16) есть *математическая модель Кронига-Пенни*. Рассмотрим графическую интерпретацию этой модели. На рис. 2.5 независимо представлены левая и правая части данных уравнений. Граничные точки, отделяющие одну зону от другой, находятся из равенства $\cos kL = +1$ или $kL = \pm n\pi$, где $n=1, 2, 3, \dots$

Уравнения (2.15) и (2.16) можно упростить, не нарушая общности. Для этого ширину областей II на рис. 2.4 следует устремить к нулю и одновременно устремить к бесконечности значение потенциала V_0 . При этом площадь областей II должна оставаться неизменной. Тогда распределение потенциала будет описываться дельта-функцией, так как при $V_0 \rightarrow \infty$ и $b \rightarrow 0$ произведение $V_0 b = \text{const}$.

Введем параметр:

$$P = \lim_{\beta \rightarrow \infty, b \rightarrow 0} \frac{\beta^2 ab}{2}$$

Далее,

$$\beta^2 = \frac{2m(V_0 - E)}{2} \approx \frac{2mV_0}{2} \quad \text{при } V_0 \rightarrow \infty$$

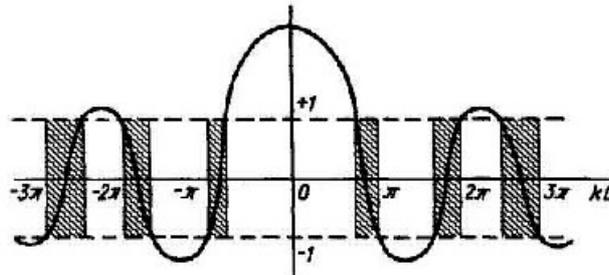


Рис. 2.5. Разрешенные (заштрихованные) и запрещенные (светлые) зоны для модели Кронига—Пенни

Если $b \rightarrow 0$, то $a \rightarrow L$. Кроме того, если $\beta^2 b = \text{const}$ при $\beta \rightarrow \infty$ и при $b \rightarrow 0$, то произведение $\beta b \rightarrow 0$. Если значения аргумента малы, то $\sin \beta b \rightarrow \beta b$; $\cos \beta b \rightarrow 1$. Таким образом, уравнение (2.15) упрощается:

$$-\frac{\alpha^2 - \beta^2}{2\alpha\beta} \beta b \sin \beta L + \cos \alpha L = \cos kL \quad (2.17)$$

Принимая во внимание, что $\beta^2 \gg \alpha^2$, получаем

$$\frac{\beta^2 b}{2\alpha} \sin \alpha L + \cos \alpha L = \cos kL$$

или

$$\frac{P}{2\alpha} \sin \alpha L + \cos \alpha L = \cos kL \quad (2.18)$$

Так как $a^2 = 2mEjh^2$, то данное уравнение действительно устанавливает связь между величинами E и k . На рис. 2.6,а показана структура энергетических зон в пространстве волнового вектора k . Точки с координатами $+n\pi/L$ представляют собой точки скачкообразного изменения энергии и служат границами n соответствующих зон Бриллюэна ($n=1,2,\dots$ -номер зоны). Запрещенным зонам отвечают такие решения уравнения (2.18), при которых значения левой части оказываются больше единицы, т.е. величина k принимает мнимые значения.

В реальном полупроводнике необходимо также указать направление волнового вектора k по отношению к осям кристалла. Поэтому кривые $E(k)$ имеют различный вид для разных направлений k .

Следует иметь в виду, что в реальном полупроводнике первая зона Бриллюэна представляет собой многогранник, ограниченный плоскостями, которые делят пополам отрезки, соединяющие узлы обратной решетки.

Модель Кронига-Пенни объясняет существование разрешенных и запрещенных зон. Этот результат, столь важный в зонной теории, не связан с конкретным выбором функции, которая описывает периодическое распределение потенциала.

Так как левые части уравнений (2.15) и (2.16) не изменяются при добавлении к аргументу kL величины $\pm 2n\pi$, то оказывается возможным получить изображение функции $E(k)$, приведенное к первой зоне Бриллюэна (рис. 2.6,б). Для этого графики, отвечающие зонам высших номеров, должны быть перемещены вдоль оси k . Так, отрезки кривых a' и b' , изображенные на рис. 2.6,б, получаются из отрезков a и b , которые представлены на рис. 2.6,а, путем сдвига на $-2\pi/L$ и $+2\pi/L$ соответственно.

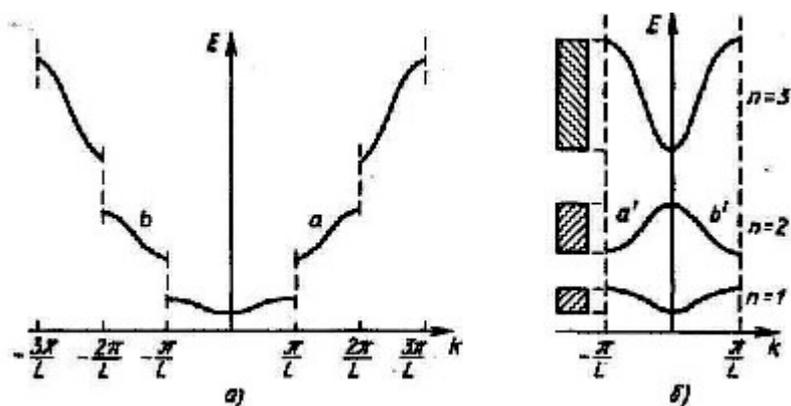


Рис. 2.6. Обычное представление функции $E(k)$ для простой одномерной кристаллической структуры в соответствии с моделью Кронига-Пенни (а), и приведение функции $E(k)$ к первой зоне Бриллюэна (б). Заштрихованы разрешенные зоны, не заштрихованы запрещенные зоны

Диаграмма $E(k)$, изображенная на рис. 2.7, дает представление о различиях между полупроводниками, диэлектриками и металлами.

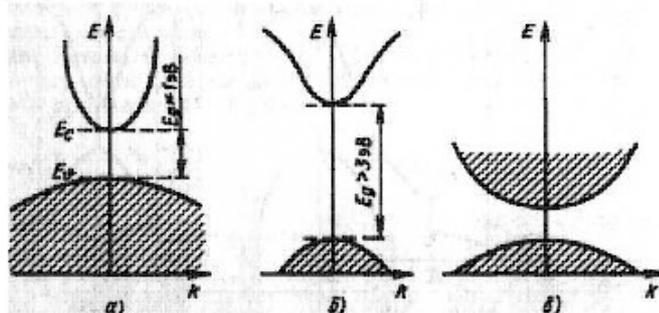


Рис. 2.7. Диаграммы энергетических зон в k -пространстве: а - для полупроводника (валентная зона заполнена электронами); б - для диэлектрика (валентная зона также заполнена); в - для металла (валентная зона заполнена полностью, а зона проводимости - частично). Можно заметить, что для диэлектрика значение E_g существенно больше, чем для полупроводника

Согласно квантово-механическим представлениям, если электроны полностью заполняют валентную зону, а зона проводимости пуста, ток проводимости возникнуть не может. Это связано с тем, что в заполненной зоне невозможны какие-либо изменения полной энергии электронов. Согласно упрощенной модели энергетических зон, в квантовой системе, которая либо получила порцию тепловой или лучистой энергии, либо оказалась под действием внешнего электрического поля, электроны из валентной зоны могут скачком переходить в зону проводимости и участвовать в создании электрического тока. Вероятность такого события тем меньше, чем больше энергия E . Поэтому с ростом E_g вещество по своим характеристикам приближается к диэлектрику.

Диаграмму $E(k)$ для свободной частицы можно описать с помощью формулы:

$$E = p^2 / (2m) \quad (2.19)$$

где p -импульс частицы массой m , пропорциональный ее волновому вектору. Для частицы, находящейся внутри кристалла $p = \hbar k$, следовательно:

$$\frac{1}{\hbar} \frac{dE}{dk} = \frac{k}{m} = v \quad (2.20)$$

Данное выражение показывает, что скорость частицы v в кристалле пропорциональна первой производной энергии по волновому вектору частицы или по ее импульсу $\hbar k$. (В точном квантово-механическом смысле следует говорить о скорости перемещения в пространстве волнового пакета, описывающего электрон. Такую скорость называют групповой скоростью частицы.— *Прим. пер.*) Тогда ускорение

$$a = \frac{dv}{dt} = \frac{dk}{m dt} = \frac{1}{\hbar} \frac{d}{dt} \frac{dE}{dk} \quad (2.21)$$

откуда

$$a = \frac{1}{\hbar} \frac{dk}{dt} \frac{d^2 E}{dk^2} \quad (2.22)$$

По определению

$$m = \frac{\hbar^2}{d^2 E / dk^2} = m^* \quad (2.23)$$

Здесь m^* - так называемая *эффективная масса* частицы, сугубо квантовая величина, которую не следует смешивать с массой свободной частицы в вакууме. Для частицы, находящейся внутри идеальной периодической решетки, эффективная масса может быть как положительной, так и отрицательной величиной.

В общем случае величина m^* зависит от k . Если кривая на диаграмме $E(k)$ имеет выпуклость вниз, как это соответствует окрестности точки E_c на рис. 2.8, то масса $m^* > 0$. Если же кривая имеет выпуклость вверх (окрестность точки E_v), то $m^* < 0$. В этом случае частица будет ускоряться в направлении, противоположном направлению ускорения электрона, т. е. будет вести себя как некоторая воображаемая частица с положительными массой и зарядом.

В рамках метода $E(k)$ диаграммы эту частицу следует называть дыркой. Очевидно, что такое определение дырки в достаточной мере отличается от определения, которое было дано в рамках модели ковалентной связи. Действительно, согласно модели ковалентной связи, движение дырки эквивалентно перемещению валентного

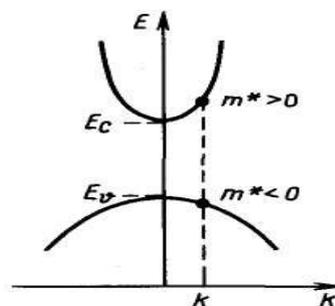


Рис. 2.8. Связь между эффективной массой m^* некоторой частицы и волновым вектором k при впадинах энергии E и E

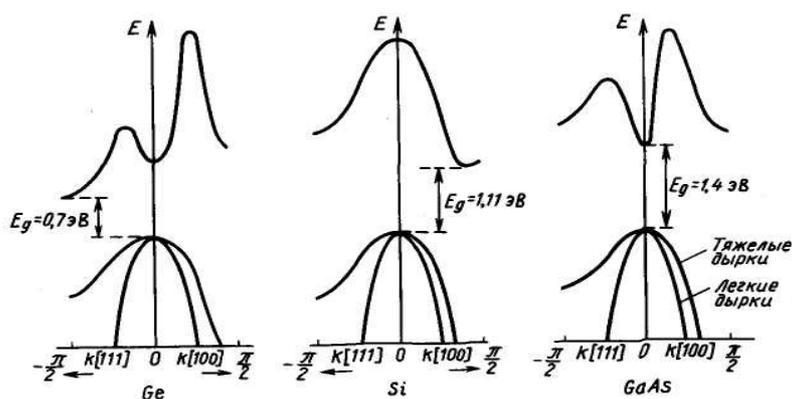


Рис. 2.9. Энергетические диаграммы полупроводников Ge, Si и GaAs в k -Пространстве

электрона в противоположном направлении, поэтому здесь нет необходимости говорить о существовании двух типов носителей заряда, хотя этот факт подтверждается при экспериментальном наблюдении явлений переноса, связанных с эффектом Холла.

Концепция эффективной массы является весьма важной, поскольку позволяет рассматривать дырки и электроны проводимости как классические заряженные частицы с эффективными массами m_p^* и m_n^* соответственно. В дальнейшем будем записывать эти символы, опуская звездочку; эффективную массу дырки будем обозначать символом m_p (индекс p указывает на то, что такой носитель имеет положительный заряд), а эффективную массу электрона - символом m_n (индекс n указывает на отрицательный заряд электрона).

Как уже отмечалось, в реальных полупроводниках диаграмма $E(k)$ гораздо более сложная и зависит от ориентации вектора k относительно осей решетки. Тем не менее общие закономерности, описываемые моделью Кронига-Пенни, остаются в силе. Любой полупроводник имеет набор энергетических зон, однако конфигурация запрещенной зоны может быть различной. На рис. 2.9 изображены основные валентные зоны и зоны проводимости для Ge, Si и GaAs. Такие графики получают как теоретическим путем, так и экспериментально.

Можно заметить, что представленные кривые несимметричны относительно оси E . Объясняется это тем, что для задания нескольких ориентации в кристалле и для определения нескольких минимумов в зоне проводимости требуется использовать различные направления k . Для кремния минимум зоны проводимости располагается вблизи зоны Бриллюэна. Полупроводник, у которого этот минимум не совпадает с точкой $k=0$, называют *полупроводником с непрямой запрещенной зоной*. Если же, как в случае GaAs, этот минимум наблюдается в точке $k=0$ то имеем *полупроводник с прямой запрещенной зоной*.

Ширина запрещенной зоны имеет большое практическое значение, потому что в ходе процессов генерации и рекомбинации носителей законы

сохранения энергии и момента количества движения не должны нарушаться. Это играет важную роль при изучении свойств фотодиодов, фотоэлементов и светоизлучающих диодов-приборов, поглощающих и излучающих свет.

Для создания полупроводниковых приборов, работающих при высоких температурах, а также светоизлучающих диодов видимого диапазона необходим полупроводник с большой шириной запрещенной зоны. Наоборот, чтобы создавать диоды, работающие при низких напряжениях или способные излучать кванты низких энергий, требуется полупроводник с узкой запрещенной зоной.

Как для кремния, так и для арсенида галлия максимум валентной зоны всегда наблюдается при $k=0$. Из характера кривых на рис. 2.15 видно, что для этих двух полупроводников валентные зоны являются вырожденными, т. е. одному и тому же значению энергии соответствуют различные состояния электрона. Для вырожденной зоны большая кривизна соответствует более тяжелым, а меньшая кривизна - более легким дыркам.

3. НОСИТЕЛИ ЗАРЯДА В ПОЛУПРОВОДНИКАХ

Знание плотности токов, концентрации носителей заряда (электронов и дырок) в полупроводниках, вместе с законом непрерывности тока, позволяют составить тот фундамент, на котором основано теоретическое изучение внутренних процессов в дискретных полупроводниковых приборах и интегральных схемах (ИС).

Проведенный здесь анализ относится как к состоянию термодинамического равновесия, так и к состоянию, когда это равновесие нарушено.

3.1. Равновесное состояние полупроводника

Система находится в состоянии термодинамического равновесия, если не существует никаких других взаимодействий с окружающей средой, кроме тех, которые необходимы для поддержания постоянной во времени и пространстве температуры. При этом концентрация носителей и полная энергия системы - в данном случае полупроводника - сохраняются неизменными. Понятие равновесия можно сформулировать и несколько по-иному: при заданной температуре некоторая система является термодинамически равновесной, если она находится в стационарном состоянии, и на нее не действуют никакие внешние возмущения.

Плотность заполнения уровней

Чтобы получить выражения, определяющие концентрацию электронов в зоне проводимости и концентрацию дырок в валентной зоне, нужно

прежде всего найти функцию $N(E)$, которая описывает распределение уровней в соответствующих зонах. После этого следует вычислить величину $N(E)dE$, представляющую собой число разрешенных энергетических уровней, приходящихся на единицу объема и лежащих в пределах от E до $E+dE$.

В соответствии с формулой (2.3) энергия:

$$E = \frac{2}{2m^* \cdot a} (k_x^2 + k_y^2 + k_z^2) \quad (3.1)$$

где m^* - эффективная масса частицы. Задав допустимые значения k_x , k_y и k_z , получаем, что некоторое квантовое состояние отвечает каждому элементу с объемом $(2\pi)^3/V$ в k -пространстве. Таким образом, $V/(8\pi^3) = N(k)$ - число частиц в «ячейке», где V - объем кристалла. В этом пространстве каждому элементу объема (кубу со стороной $a=1$) соответствует некоторое квантовое состояние; числовое значение всего объема определяет число таких состояний.

Если предположить, что поверхностью постоянной энергии является сфера Ферми радиуса $|k|$, то число квантовых уровней, заключенных внутри этой поверхности:

$$N = \frac{4}{3} \pi \cdot |k|^3 \quad (3.2)$$

Так как $a=1$, то, используя явное выражение вектора k , на основании (3.1) и (3.2) получаем:

$$N = \frac{4}{3} \pi \left(\frac{2m^*}{2} \right)^{3/2} E^{3/2} \quad (3.3)$$

Дифференцируя выражение (3.3) по E , находим число энергетических уровней, лежащих между E и $E+dE$:

$$dN(E) = 2\pi \left(\frac{2m^*}{2} \right)^{3/2} E^{1/2} dE \quad (3.4)$$

Фактически интерес представляет только восьмая часть объема полной сферы; в пределах этой части значений k_x , k_y и k_z положительны. Если к тому же учесть, что спин электрона имеет два возможных направления, то формула (3.4) принимает следующий вид:

$$dN(E) = \frac{4\pi}{3} (2m^*)^{3/2} E^{1/2} dE \quad (3.5)$$

Наконец, примем во внимание, что самый нижний уровень E_c энергии электрона в зоне проводимости есть потенциальная энергия «покоящегося» электрона. Если электрон, находящийся сначала на этом уровне, приобретает энергию E , которая превышает E_c , то разность $E - E_c$ является кинетической энергией электрона.

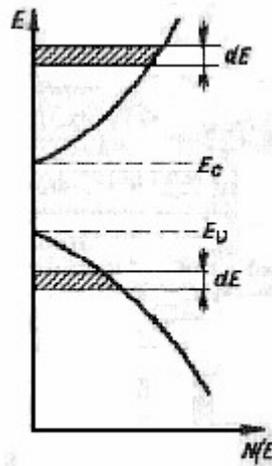


Рис. 3.1. Плотность квантовых уровней

Таким образом, приходим к выводу, что плотность квантовых уровней $N(E)$ в зоне проводимости

$$N(E) = \frac{4\pi}{3} (2m_n)^{3/2} (E - E_c)^{1/2} = N_c (E - E_c)^{1/2} \quad (3.6)$$

Аналогично в валентной зоне

$$N(E) = \frac{4\pi}{3} (2m_p)^{3/2} (E_v - E)^{1/2} = N_v (E_v - E)^{1/2} \quad (3.7)$$

где m_n и m_p — эффективная масса электрона и дырки соответственно; N_c и N_v — некоторые постоянные (рис. 3.1), где показана плотность квантовых уровней в зоне проводимости (электронная электропроводимость) и в валентной зоне (дырочная электропроводимость). Площадь заштрихованных областей пропорциональна числу уровней в интервале энергий dE при единичном объеме.

Как видно из формул (3.6) и (3.7), параметры m_n и m_p являются однородными величинами, характеризующими некоторую конкретную материальную среду. Чем больше числовые значения m_n и m_p , тем выше плотность квантовых уровней. Можно отметить также, что в окрестностях минимума зоны проводимости и максимума валентной зоны формулы (3.6) и (3.7) задают параболическую аппроксимацию. Применимость этих формул к исследованию свойств полупроводников определяется тем, что основная часть носителей заряда располагается или вблизи дна юны проводимости, или вблизи потолка валентной зоны.

Перейдем теперь к вопросу о том, как влияет абсолютная температура T на плотность распределения энергетических уровней электронов.

Функция распределения Ферми — Дирака

Одна из цепей любой статистической теории состоит в нахождении функции распределения. Так принято называть функцию, которая в условиях термодинамического равновесия при заданной температуре T

пропорциональна вероятности того, что некоторая частица занимает определенный энергетический уровень E .

Если рассматриваются классические (не квантовые) системы и не учитываются какие-либо специфические свойства частиц, то применима функция распределения Максвелла-Больцмана

$$f_{MB}(E) = Ce^{-E/(kT)} \quad (3.8)$$

где k - постоянная Больцмана; C - некоторый коэффициент пропорциональности.

При изучении систем неразличимых квантово-механических частиц возникают функции распределения двух видов. Одна из них, так называемое распределение Бозе - Эйнштейна, применима к частицам, которые описываются симметричными волновыми функциями:

$$f_{БЭ}(E) = [1 - \exp((E - E_B)/(kT))]^{-1} \quad (3.9)$$

где E_B - некоторая постоянная величина. При этом число частиц, которые могут занимать один и тот же энергетический уровень, в принципе не ограничено.

Другая функция описывает распределение Ферми-Дирака. Оно применимо к частицам, волновые функции которых антисимметричны и подчиняются запрету Паули (на одном и том же квантовом уровне не может оказаться более одного электрона). Данная функция имеет вид

$$f(E) = [1 + \exp((E - E_F)/(kT))]^{-1} \quad (3.10)$$

где E_F - так называемый уровень Ферми, который при заданной температуре занят электронами с вероятностью $1/2$.

Функция (3.10) представлена графически на рис. 3.2, причем в качестве параметра выбрана температура T . Можно видеть, что имеют место следующие соотношения:

$$\begin{aligned} E = E_F &\Rightarrow f(E) = 1/2, \\ E \gg 0 &\Rightarrow f(E) \rightarrow 0, \\ E = 0 &\Rightarrow f(E) \rightarrow 1, \\ E \gg E_F &\Rightarrow f(E) \approx e^{E_F/(kT)} e^{-E/(kT)}. \end{aligned} \quad (3.11)$$

Выражение (3.11) совпадает по форме с функцией распределения

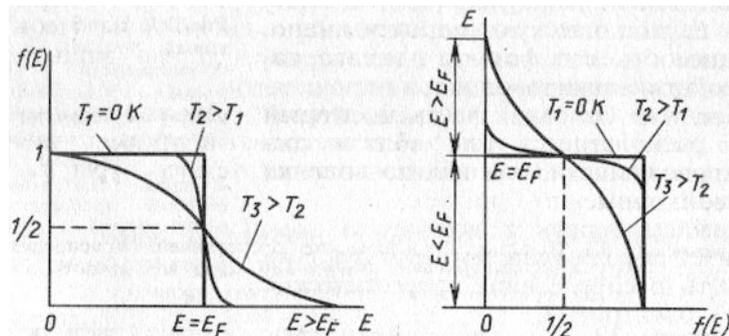


Рис. 3.2. Функция распределения Ферми—Дирака (графики эквивалентны)

Максвелла-Больцмана (3.8), которая описывает свойства некоторого газа, состоящего из классических частиц. Это позволяет упростить анализ ряда полупроводниковых устройств, например диодов и транзисторов, для которых можно положить, что $E - E_F > 3kT$ или $E \gg E_F$. Отметим, такое предположение несправедливо по отношению к туннельным диодам и инжекционным полупроводниковым лазерам.

Можно показать, что уровень Ферми представляет собой некоторую функцию температуры $E_F(T)$. Действительно, если $T = 0$ К, то

$$f(E) = \begin{cases} 1 & \text{при } E < E_F \\ 0 & \text{при } E > E_F \end{cases}$$

Это следует понимать так: при нулевой абсолютной температуре величина $E_F(0)$ представляет собой то максимально допустимое значение энергии, ниже которого все энергетические уровни заняты [так как $f(E) = 1$], а выше которого все уровни - пусты [так как $f(E) = 0$]. Короче говоря, при нулевой абсолютной температуре функция $f(E)$ имеет разрыв в точке координатой E_F . Интуитивно ясно, что при этой температуре уровень E_F располагается в центре запрещенной зоны.

На основании вида функции Ферми - Дирака можно утверждать, что вероятность иметь некоторый уровень энергии в занятом состоянии при повышении температуры растет, при ее понижении - падает.

Вероятность такого события, что некоторый энергетический уровень E занят не электроном, а дыркой, равна $1 - f(E)$. При этом, как можно видеть,

$$\begin{aligned} 1 - f(E) &= 1 - [1 + \exp((E - E_F)/(kT))]^{-1} = \\ &= \exp((E - E_F)/(kT)) / [1 + \exp((E - E_F)/(kT))] \end{aligned}$$

При изучении явлений, связанных с дырками, часто полагают что $E \ll E_F$ и $\exp((E - E_F)/(kT)) \ll 1$. Следовательно,

$$1 - f(E) \approx e^{E_F/(kT)} e^{-E/(kT)} \quad (3.12)$$

Данная формула напоминает выражение (3.11), однако здесь изменены знаки показателей экспонент. Это хорошо согласуется представлением о том, что дырки являются носителями положительного заряда. Если потенциал точки пространства, в которой находится дырка, становится более положительным, то ее энергия возрастает.

Концентрация носителей заряда

Как уже известно, распределения энергетических уровней, на краях зоны проводимости и валентной зоны не являются однородными, а описываются функциями параболического типа [см. формулы (3.6) и (3.7)]. В то же время функция распределения Ферми-Дирака вида (3.10) пропорциональна вероятности события, которое состоит в том, что

некоторый электрон занимает определенный энергетический уровень E при заданной температуре T .

Из изложенного ранее следует, что число частиц с энергиями в интервале от E до $E+dE$, приходящееся на единицу объема, составляет $N(E)f(E)dE$. Будем рассматривать электроны и полагать, что $E \gg E_F$.

В этом случае

$$dn = N(E)f(E)dE \quad (3.13)$$

где n - концентрация электронов в зоне проводимости.

Число электронов в зоне проводимости, приходящееся на единицу объема,

$$n = n(E) = \int_{E_c}^{E_{max}} N(E)f(E)dE \quad (3.14)$$

Интегрирование проводится в пределах от нижней границы зоны проводимости E_c до верхней границы этой зоны E_{max} . Сохраняя правильный смысл результата, можно существенно упростить выкладки, положив, что верхний предел интегрирования равен $+\infty$. Выполнив соответствующие подстановки, получим

$$\begin{aligned} n &= \frac{4\pi}{h^3} (2m_n)^{3/2} \int_{E_c}^{\infty} \frac{(E - E_c)^{1/2} dE}{1 + \exp[(E - E_F)/kT]} = \\ &= \frac{4\pi}{h^3} (2m_n)^{3/2} \exp\left(-\frac{E_c - E_F}{kT}\right) \cdot \int_0^{\infty} (E - E_c)^{1/2} \times \exp\left(-\frac{E - E_c}{kT}\right) d(E - E_c) \end{aligned} \quad (3.15)$$

Если обозначить $x = (E - E_c)/kT$, то

$$n = \frac{4\pi}{h^3} (2m_n kT)^{3/2} \exp\left(-\frac{E_c - E_F}{kT}\right) \int_0^{\infty} x^{1/2} e^{-x} dx.$$

Используя табличный интеграл $\int_0^{\infty} x^{1/2} e^{-x} dx = \frac{\sqrt{\pi}}{2}$, получаем

$$n = N_c \exp\left(-\frac{E_c - E_F}{kT}\right) \quad (3.16)$$

где

$$N_c = 2 \left(\frac{2\pi m_n kT}{h^2} \right)^{3/2} \quad (3.17)$$

Аналогично находим концентрацию дырок, приняв во внимание, что вероятность возникновения вакантного уровня в валентной зоне равна $1-f(E)$. Интегрирование следует проводить в пределах от $-\infty$ до E_v ,

$$p = \int_{-\infty}^{E_v} N(E)[1 - f(E)]dE \quad (3.18)$$

откуда

$$p = N_v e^{-\frac{E_v - E_F}{kT}} \quad (3.19)$$

где

$$N_v = 2 \left(\frac{2\pi m_p kT}{h^2} \right)^{3/2} \quad (3.20)$$

Рассмотрим физический смысл параметров N_c и N_v . Величина N_c - плотность уровней в зоне проводимости. Так как формула (3.16) определяет классический предел заполнения энергетических уровней вблизи дна зоны проводимости, т.е. в окрестности точки E_c , то число электронов n в единице объема окажется ровно таким, каким оно было бы, если бы единице объема соответствовало N_c уровней с одинаковой энергией E_c . Соответственно величина N_v является плотностью уровней в валентной зоне. Концентрация дырок p в данной зоне будет такой же, как и в том случае, когда имеется N_v уровней с одинаковой энергией E_v каждый. Можно заметить, что N_c и N_v являются величинами одного порядка, зависящими от температуры. Их числовые значения несколько различаются, так как не равны соответствующие эффективные массы. В кремнии при комнатной температуре $N_c = 2,8 \cdot 10^{19} \text{ см}^{-3}$ и $N_v = 10^{19} \text{ см}^{-3}$.

На рис. 3.3,а изображены графики функций, описывающих плотность уровней (число уровней, приходящихся на единичный интервал энергий и на единицу объема). Графики на рис. 3.3,б описывают функцию распределения - вероятность, с которой заданные уровни оказываются занятыми, — при $T=0 \text{ К}$ и при $T_1 > T$. Наконец, кривые рис. 3.3,в есть произведение этих двух функций (число электронов и дырок в единичном интервале энергий и в единичном объеме).

Если $T=0 \text{ К}$, то функция $f(E)$ равна нулю при $E > E_p$ и +1 при $E < E_F$. Тогда (рис. 3.3,в) произведение $N(E)f(E)$ начиная со значения E_c отображается прямой, которая совпадает с осью E . Значениям, меньшим E_v , отвечает кривая, изображенная штриховой линией и совпадающая с той, которая представлена на рис. 3.3,а. Иными словами, при $E \geq E_c$ имеем $N(E)f(E) = 0$ (электроны в зоне проводимости отсутствуют); при $E \leq E_v$ получаем $N(E)f(E) = N(E)$ (валентная зона заполнена полностью).

При температуре $T > 0 \text{ К}$ электроны, оказавшиеся в зоне проводимости, распределяются по закону $N(E)f(E)$, а дырки, возникшие в валентной зоне, - по закону $N(E)[1 - f(E)]$, которому отвечает штриховая кривая на рис. 3.3,в. Число электронов в зоне проводимости равно числу дырок в валентной зоне.

Если эффективные массы носителей заряда равны, то уровень Ферми E_F в чистом полупроводнике расположен посередине запрещенной зоны; кривые, описывающие распределение носителей, оказываются симметричными относительно точки E_F . При $m_p > m_n$ уровень Ферми смещается в верхнюю часть запрещенной зоны. Если же $m_p < m_n$, то это смещение происходит в противоположном направлении. Интегрируя в соответствующих пределах указанные функции, определяемые формулами (3.16) и (3.19), находим концентрацию электронов и дырок соответственно.

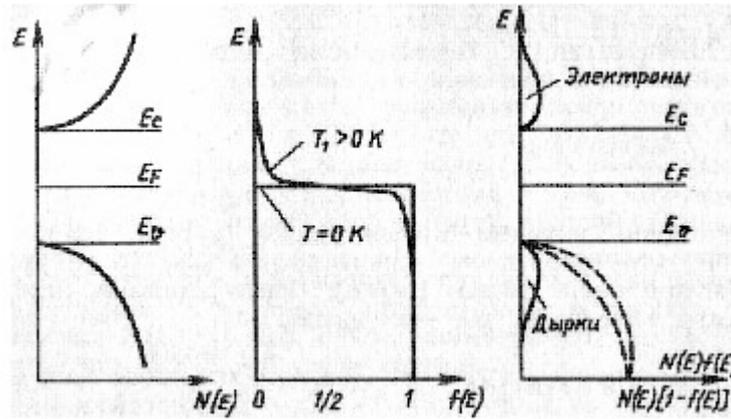


Рис. 3.3. Графический способ определения концентрации носителей (электронов и дырок) в чистом полупроводнике-площади, заключенной между кривыми $N(E)f(E)$, $N(E)[1-f(E)]$ и осью энергий E

Можно исключить уровень Ферми E_F , перемножив выражения (3.16) и (3.19)

$$np = N_c N_v \exp\left(-\frac{E_c - E_v}{kT}\right) = N_c N_v \exp\left(-\frac{E_g}{kT}\right) \quad (3.21)$$

где E_g - ширина запрещенной зоны. В действительности величина E_g связана с температурой довольно сложной зависимостью.

В интервале $150 \leq T \leq 300$ °К закон оказывается примерно линейным и выражается эмпирической формулой:

$$E_g(T) = E_{g0} - \alpha T \quad (3.22)$$

где $E_{g0} = E_g(0)$ - значение параметра E_g при нулевой абсолютной температуре; α - температурный коэффициент.

Для кремния $E_{g0} = 1,21$ эВ и $\alpha = 2,8 \cdot 10^{-4}$ эВ/К. Подставляя (3.22) в (3.21), получаем

$$np = N_c N_v e^{\alpha/k} e^{-E_{g0}/(kT)} = A_0 T^3 e^{-E_{g0}/(kT)} \quad (3.23)$$

где A_0 - некоторая постоянная. Правая часть этого уравнения зависит только от температуры и ширины запрещенной зоны полупроводника, но не зависит от концентрации носителей. Поэтому левая часть должна оставаться постоянной независимо от того, является ли полупроводник чистым или в нем присутствуют какие-нибудь примеси.

Собственные и примесные полупроводники

Удельное сопротивление полупроводников при комнатной температуре колеблется от 10^{-2} до 10^9 Ом·см. Если $T = 0$ К, то любые полупроводники ведут себя подобно диэлектрикам. Проводящие свойства этих материалов улучшаются с ростом температуры, а также при добавлении определенных примесей. В соответствии с этим принято говорить, что чистый материал, не содержащий примесей, влияющих на электропроводность, является собственным полупроводником. Если же в его

состав входят некоторые легирующие добавки, то такой материал называют примесным полупроводником.

Собственная концентрация, произведение pn и уровень Ферми

Как уже было показано, произведение концентраций носителей является некоторой постоянной величиной, которая зависит от температуры. Единственное условие, которое при этом налагается, состоит в том, что интервалы между уровнем E_F и границами разрешенных зон должны быть достаточно велики по сравнению с параметром kT . При комнатной температуре $kT=26$ мэВ значение E_g для кремния составляет 1,1 эВ (заметим, что в данном случае $E_{g0}=1,21$ эВ).

Если за счет термического возбуждения электрон покидает валентную зону, то на его месте в этой зоне образуется дырка. Таким образом, для собственного полупроводника

$$n = p = n_i = p_i \quad (3.24)$$

где n_i и p_i - собственные концентрации носителей. Генерация электронно-дырочных пар и их рекомбинация происходят непрерывно. При любой температуре устанавливается такое равновесное состояние, что произведение np постоянно. В этом можно убедиться, подставив (3.24) в (3.21):

$$n = p_i^2 = n_i^2 = const \quad (3.25)$$

при неизменных значениях T и E_g .

Равенство (3.25), играющее важную роль в теории полупроводников и полупроводниковых устройств, называют *уравнением полупроводника* или *законом действующих масс*. В случае термодинамического равновесия этот закон справедлив как для собственных, так и для примесных полупроводников. Если же равновесие нарушено и величины n и p в некотором конкретном приборе зависят от внешних факторов, то $np \neq n_i^2$.

Деля выражения (3.19) и (3.16), получаем

$$\frac{p}{n} = \frac{N_v}{N_c} \exp\left(\frac{E_c + E_v + 2E_F}{kT}\right)$$

или

$$E_F = \frac{1}{2}(E_c + E_v) - \frac{1}{2}kT \ln \frac{p}{n} - \frac{3}{4}kT \ln \frac{m_n}{m_p} \quad (3.26)$$

Уровень Ферми E_F собственного полупроводника, для которого $\ln(p/n) = 0$, обозначается символом E_i . Если $m_n \approx m_p$, что имеет место на практике, то уровень Ферми собственного полупроводника располагается посередине запрещенной зоны:

$$E_i = (E_c + E_v) / 2 \quad (3.27)$$

Если $n \neq p$, то уровень Ферми смещается в сторону той разрешенной зоны, которая имеет большую концентрацию носителей.

Концентрации носителей заряда можно представить в виде функций, зависящих от n и E . Действительно, положив $n=p=n_i$ и $E_F=E_i$ на основании формул (3.16) и (3.19) можно записать:

$$n_i = N_c \exp\left(-\frac{E_c - E_i}{kT}\right) = N_v \exp\left(-\frac{E_c - E_i}{kT}\right) \quad (3.28)$$

Поэтому указанные формулы принимают вид:

$$n = n_i \exp\left[\frac{E_F - E_i}{kT}\right] \quad (3.29)$$

$$p = n_i \exp\left[\frac{E_F - E_i}{kT}\right] \quad (3.30)$$

Формулы (3.29) и (3.30) в равной мере относятся как к собственным, так и к примесным полупроводникам.

Наконец, легко доказать, что параметр n_i^2 зависит от T и E_g . Действительно, $n_i^2 = np = N_c N_v e^{-E_g/(kT)} = AT^3 e^{-E_g/(kT)}$

С помощью этой формулы находим, что в кремнии при комнатной температуре $n_i^2 = 4,64 \cdot 10^{21} \text{ см}^{-6}$.

Примесные полупроводники

Чтобы превратить собственный полупроводник в примесный, необходимо ввести в его кристаллическую решетку некоторое количество специально подобранной химической добавки, т. е. осуществить легирование полупроводника. Примеси создают ряд энергетических уровней в запрещенной зоне. В результате вероятность образования электронно-дырочных пар при тепловом возбуждении оказывается гораздо более высокой, чем в случае собственного полупроводника.

Не следует думать, что все носители заряда возникают лишь под действием разрыва межатомных связей, как в модели ковалентной связи, или лишь под влиянием генерации электронно-дырочных пар, как в зонной модели. Дело в том, что большое число носителей обусловлено именно присутствием примесных атомов. На практике легирование осуществляется при помощи процесса диффузии. Этот процесс позволяет внедрять примеси в кристаллическую решетку полупроводника таким образом, что возникает именно тот профиль пространственного распределения примесей, который необходим для создания конкретного прибора. Легирование можно осуществить также при помощи других процессов, таких, как ионная имплантация и эпитаксиальное наращивание. Существуют примеси двух видов: *доноры* - пятивалентные элементы, такие, как P, As, Sb; концентрацию доноров будем обозначать символом N_d ; *акцепторы* - трехвалентные элементы, такие, как B, Al, Ga, их концентрацию будем обозначать символом N_a . На основании этого различают полупроводники n-типа и p-типа. В полупроводниках n-типа *типичными носителями* являются электроны ($n > p$), а *неосновными носителями* - дырки. Основные носители в данном случае создаются пятивалентными атомами элемента,

принадлежащего V группе Периодической системы элементов. Как упоминалось ранее, такой атом может «освободить» пятый электрон. В полупроводнике; p -типа основные носители - дырки ($p > n$), а неосновные - электроны. Здесь «свободную» дырку создает трехвалентный атом элемента, принадлежащего III группе периодической системы элементов. На рис. 3.4 изображены зонные диаграммы примесных полупроводников.

Рис. 3.4,а соответствует случаю, когда в собственный кремний введена легирующая добавка пятивалентного элемента. Пятый электрон примесного атома не может разместиться в валентной зоне кремния, которая полностью заполнена четырьмя электронами. Вследствие этого образуется новый разрешенный донорный уровень E_d , располагающийся на краю запрещенной зоны, которая граничит с зоной проводимости. Аналогично при легировании полупроводника трехвалентной примесью возникает разрешенный акцепторный уровень E_a (рис. 3.4,б).

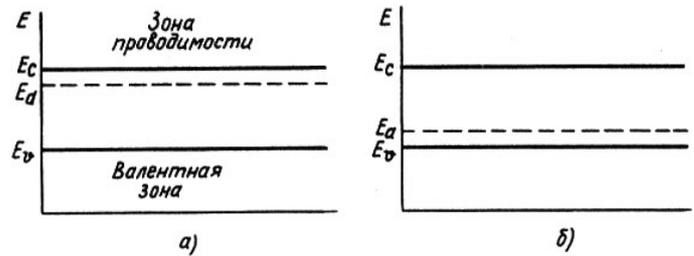


Рис. 3.4. Диаграммы энергетических зон примесных полупроводников при комнатной температуре: а — материал типа n , б — материал типа p

Уровни E_d и E_a оказываются весьма близкими к зоне проводимости и к валентной зоне соответственно. Понятно, что, получив небольшую порцию тепловой энергии, электрон, первоначально находившийся на уровне E_d , попадает в зону проводимости. Таким же образом электрон, находившийся вблизи потолка валентной зоны, переходит на уровень E_a . В результате таких процессов в зоне проводимости возникают электроны, а в валентной зоне - дырки. Если теперь к образцу приложить внешнее электрическое поле, то в нем начнет протекать электрический ток.

Установлено, что при комнатной температуре все примесные атомы оказываются ионизированными. Уровень Ферми E_f располагается либо между уровнем E_d и дном зоны проводимости (рис. 3.4, а), либо между уровнем E_a и потолком валентной зоны (рис. 3.4, б).

Если некоторый полупроводник одновременно легирован донорами и акцепторами, то можно получить материал любого типа в зависимости от того, какая из добавок имеет большую концентрацию. Этим часто пользуются при производстве полупроводниковых приборов, превращая, например, полупроводник n -типа в материал p -типа путем создания избыточной концентрации акцепторов.

Перейдем к расчету параметров n и p для примесного полупроводника. Здесь помимо уравнения (3.25) необходимо иметь еще одно уравнение, связывающее концентрации носителей заряда и концентрации примесных атомов и являющееся *условием электронейтральности*:

$$p + N_d = n + N_a \quad (3.31)$$

Справедливость данного равенства вытекает из следующих положений:

- полупроводник, на который не действует внешнее поле, является электрически нейтральным;
- введение донорных примесей с концентрацией N_d и акцепторных примесей с концентрацией N_a обуславливает появление добавочных электронов и дырок;
- все донорные и акцепторные атомы ионизированы.

Таким образом, зная параметры N_d и N_a , а также температуру, можно подсчитать концентрации n и p для любого примесного полупроводника, воспользовавшись формулами (3.25) и (3.31). Концентрация основных и неосновных носителей определяется соответственно формулами (3.32) и (3.33).

$$n_n = \frac{1}{2} N_d \left[1 + \left(1 + \frac{4n_i^2}{N_d^2} \right)^{1/2} \right] \quad (3.32)$$

$$p_n = \frac{1}{2} N_d \left[-1 + \left(1 + \frac{4n_i^2}{N_d^2} \right)^{1/2} \right] \quad (3.33)$$

Часто в практически важных случаях можно считать, что $n_i/N_d \ll 1$. Это позволяет упростить формулы (3.32) и (3.33), поскольку при $x \ll 1$ справедливо разложение $(1+x)^j = 1 + jx + \dots$. Тогда

$$n_n \approx N_d; p_n \approx n_i^2 / N_d \quad (3.34)$$

Аналогично убеждаемся, что для полупроводника p -типа

$$p_p \approx N_a; n_p \approx n_i^2 / N_a \quad (3.35)$$

Как можно видеть из формул (3.16) и (3.19), концентрация носителей заряда определяется положением уровня Ферми. В следующей главе при изучении свойств p - n -перехода будет показано, что значение E_F зависит от напряженности электрического поля в этом переходе. Эта напряженность, в свою очередь, определяется емкостью перехода. Емкость p - n -перехода является одним из важнейших параметров, влияющих на работу полупроводниковых приборов.

Получим простейшие выражения, определяющие уровень E_F для примесного полупроводника; в более общем случае нахождение его связано с графическими построениями и числовыми расчетами.

Для полупроводника n -типа, подставляя (3.34) в (3.16) имеем

$$N_d = N_c \exp[-(E_c - E_F) / kT] \quad (3.36)$$

откуда

$$E_F = E_c - kT \ln(N_c / N_d) \quad (3.37)$$

Если же рассматривать полупроводник p -типа, то, подставляя (3.35) в (3.19), имеем:

$$N_a = N_v \exp[-(E_F - E_v) / kT] \quad (3.38)$$

откуда

$$E_F = E_v - kT \ln(N_v / N_a) \quad (3.39)$$

Говорят, что полупроводник является *вырожденным*, если концентрация примесей настолько высока, что уровень E_F лежит на границах разрешенных зон или внутри них. За исключением некоторых особых случаев, все полупроводники, рассматриваемые в данной книге, являются невырожденными. В соответствии с формулами (3.37) и (3.39) также оказывается, что в невырожденных полупроводниках уровень E_F достаточно близок к границе зоны проводимости или валентной зоны.

В более строгом смысле вырожденный полупроводник характеризуется неравенством:

$$E_F - E_v < 3kT \quad (3.40)$$

Это условие не позволяет использовать приближенное выражение (3.11) и делает неприменимой формулу (3.37). Интересно отметить, что туннельный диод имеет две вырожденные области: *p*- и *n*-типа.

Если в определенных условиях, скажем при низких температурах, некоторые атомы донорной примеси оказываются неионизированными, то концентрация ионизированных атомов этой примеси, которую следует подставлять в выражение (3.37),

$$N_d^+ = N_d - N_{dn} \quad (3.41)$$

где N_d - общая концентрация атомов донорной примеси; N_{dn} - концентрация нейтральных атомов этой примеси, которая в соответствии с распределением Ферми - Дирака определяется формулой:

$$N_{dn} = \frac{N_d}{1 + 1/2 \exp[(E_d - E_F)/kT]} \quad (3.42)$$

Можно отметить, что статистика Ферми - Дирака предполагает наличие двух электронов в любом энергетическом состоянии. Однако на примесном уровне содержится лишь один электрон, что и учитывается множителем $1/2$ в формуле 3.42.

Аналогично, в случае легирования акцепторной примесью концентрация ионизированных примесных атомов в выражении (3.39)

$N_a^- = N_a - N_{an}$. Здесь концентрация нейтральных акцепторных атомов

$$N_{an} = \frac{N_a}{1 + 1/4 \exp[(E_a - E_F)/kT]} \quad (3.43)$$

Множитель $1/4$ учитывает наличие двух ориентации спина и существование двух вырожденных валентных зон, как это имеет место, например, в кремнии.

Явления переноса в условиях стационарной неравновесности

В тех случаях, когда на полупроводниках действуют некоторые внешние факторы, возникает ненулевая упорядоченная скорость носителей заряда. Как следствие, состояние термодинамического равновесия в полупроводнике нарушается. Связанные с этим явления называют

явлениями переноса. Внешнее возбуждение может иметь самый разнообразный характер - электрический, тепловой или оптический. Поэтому конкретные проявления эффектов переноса могут быть различными. В дальнейшем будем рассматривать два явления: 1) дрейфовую и диффузионную электропроводности и 2) эффект Холла.

Движение носителей заряда в образце полупроводника может возникать прежде всего под действием электрического поля или разности потенциалов. Образующийся электрический ток принято называть дрейфовым. Кроме того, движение носителей может обуславливаться пространственной неоднородностью их концентрации. При этом возникает так называемый диффузионный ток. Эффект Холла состоит в том, что между противоположными сторонами полупроводникового образца появляется некоторое электрическое поле или разность потенциалов. Это поле перпендикулярно плоскости, образованной двумя векторами: вектором напряженности электрического поля, вызывающего движение *неосновных* носителей заряда в образце, и перпендикулярным ему вектором напряженности магнитного поля. Эффект Холла будет подробно рассмотрен в дальнейшем. Очевидно, что этот эффект принадлежит к классу стационарно неравновесных явлений.

Дрейфовая и диффузионная электропроводности

В однородном полупроводнике свободные электроны, появившиеся за счет теплового возбуждения, совершают хаотические движения. Траектория отдельно взятого электрона прямолинейна до тех пор, пока не произойдет столкновения. Столкновения могут возникать по ряду причин, из-за нарушения периодичности потенциала под действием тепловых колебаний решетки (фотонов), из-за дефектов решетки, различных примесей, взаимодействий с другими носителями заряда и т. д. Средний ток в любом выбранном направлении равен нулю. Среднее расстояние между двумя последовательными столкновениями называют длиной свободного пробега частицы и составляет 10^{-6} - 10^{-4} см. Так как средняя скорость электрона порядка 10^7 см/с, то *время свободного пробега* $\tau_{сн}$ оказывается порядка 1 пс.

Явление дрейфа

Пусть к полупроводниковому образцу приложено некоторое достаточно слабое постоянное электрическое поле напряженностью ξ . В этом случае наряду с хаотическим движением свободных электронов наблюдается их упорядоченное движение в направлении, параллельном вектору напряженности поля, но противоположном по знаку. Иными словами, возникает некоторая отличная от нуля средняя скорость упорядоченного движения в направлении поля. Эту скорость принято называть средней скоростью дрейфа носителей.

Среднее ускорение электрона

$$a = q\xi / m_n \quad (3.44)$$

где q - заряд электрона. Можно считать, что в результате большого числа последовательных столкновений средняя скорость дрейфа электронов

$$v_{др} = a\xi / \tau_{cn} \quad (3.45)$$

Таким образом, если ввести понятие *подвижности* электронов

$$\mu_n = q\tau_{cn} / m_n \quad (3.46)$$

то

$$v_{др} = \mu_n \xi .$$

(3.47)

Таким же образом, рассматривая дрейфовое движение дырок, имеем

$$v_{др} = \mu_p \xi \quad (3.48)$$

Если определить плотность дрейфового тока электронов

$$Jn_{др} = nqv_{др} \quad (3.49)$$

то удельная объемная проводимость, вызванная движением электронов,

$$\sigma_n = \frac{Jn_{др}}{\xi} = nq \frac{v_{др}}{\xi} = nq\mu_n \quad (3.50)$$

Аналогичной формулой описывается дрейфовое движение дырок

$$\sigma_p = pq\mu_p \quad (3.51)$$

Электрическое поле напряженностью ξ действует как на электроны, так и на дырки; так как знаки зарядов и скорости этих частиц противоположны, то соответствующие токи складываются

Таким образом, приходим к формулировке закона Ома:

$$J_{др} = Jn_{др} + Jp_{др} = (n\mu_n + p\mu_p)q\xi = \sigma\xi$$

где

$$\sigma = \sigma_n + \sigma_p = q(n\mu_n + p\mu_p) \quad (3.52)$$

- удельная объемная проводимость материала. Эта величина зависит от температуры, с которой связаны подвижность и концентрация носителей.

Пусть образец полупроводникового материала имеет форму стержня длиной l с площадью поперечного сечения A . Сопротивление R этого образца и удельное сопротивление ρ материала определяются формулами

$$R = \rho \frac{l}{A} = \frac{U}{I}; \frac{1}{\rho} = \sigma = q(n\mu_n + p\mu_p).$$

Удельное сопротивление полупроводника является важным параметром при разработке полупроводниковых приборов.

Отметим попутно, что температурные зависимости сопротивления полупроводника и металла противоположны - во всем диапазоне температур, имеющих практический интерес, удельное сопротивление полупроводника с ростом температуры уменьшается, а удельное сопротивление металла растет.

Что касается числового значения подвижности, то для каждого вида носителей заряда в том или ином материале его можно считать постоянным. При этом, конечно, следует учитывать, что в общем случае подвижность связана с температурой и концентрацией легирующей примеси. Дело в том, что при достаточно высоких концентрациях примесей они влияют на длину свободного пробега между столкновениями. Длина свободного пробега зависит также от приложенного электрического поля, уменьшаясь с ростом его напряженности.

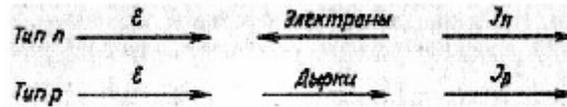


Рис. 3.5. Дрейфовый ток в полупроводниках n и p -типа. Следует обратить внимание на направление векторов напряженности поля и скорости носителей, а также на связь указанных векторов с вектором плотности дрейфового тока

Можно считать, что дрейфовый ток, обусловленный неосновными носителями, пренебрежимо мал по сравнению с током, который связан с основными носителями (рис. 3.5)

Полагают также, что распределение основных носителей практически остается равновесным, т.е. влияние на него внешнего поля можно не учитывать.

Диффузия

Если концентрация носителей заряда в полупроводнике оказывается пространственно неоднородной, то вместе с тепловым движением наблюдается перенос носителей из одной области кристалла в другую за счет диффузии (именно такое явление имеет место в p - n -переходе). Данный процесс аналогичен диффузии молекул газа в замкнутом сосуде при наличии градиента давления.

Пусть в момент времени $t=0$ имеется некоторое неоднородное распределение концентрации носителей. При этом возникает диффузионный ток носителей между областями с неодинаковыми концентрациями. По истечении достаточного времени во всем объеме кристалла устанавливается равновесное состояние.

Плотность диффузионного тока электронов $J_{n_{диф}}$ пропорциональна градиенту их концентрации:

$$J_{n_{диф}} = qD_n \frac{dn}{dx} \quad (3.53)$$

где D_n - коэффициент диффузии электронов. Аналогично, плотность диффузионного тока дырок

$$J_{p_{диф}} = -qD_p \frac{dp}{dx} \quad (3.54)$$

где D_p - коэффициент диффузии дырок. Отрицательный знак (3.54) возник потому, что вектор плотности тока дырок

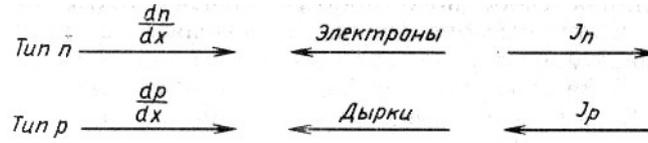


Рис. 3.6. Диффузионный ток в полупроводниках *n* и *p*-типа. Следует обратить внимание на связь между векторами градиентов концентраций, скоростями носителей и соответствующими плотностями диффузионного тока.

направлен в сторону, противоположную градиенту их концентрации (рис. 3.6).

Следует заметить, что плотности диффузионных токов неосновных носителей заряда сходным образом зависят от градиентов концентрации соответствующих носителей, т.е. от dn/dx и dp/dx . В дальнейшем будет показано, что диффузионный ток неосновных носителей заряда оказывает большое влияние на работу полупроводниковых приборов.

Явление дрейфа и диффузии могут наблюдаться одновременно. Поэтому можно записать следующие важнейшие соотношения для одномерной модели:

$$J_n = J_{n_{др}} + J_{n_{диф}} = q\mu_n n \xi + qD_n \frac{dn}{dx} \quad (3.55)$$

$$J_p = J_{p_{др}} + J_{p_{диф}} = q\mu_p p \xi + qD_p \frac{dp}{dx} \quad (3.56)$$

Для кремния при комнатной температуре и при незначительной степени легирования $D_n = 38 \text{ см}^2/\text{с}$ и $D_p = 13 \text{ см}^2/\text{с}$.

Следует ожидать, что имеется некоторая связь между величинами μ_n и μ_p , относящимися к процессу дрейфа, и параметрами D_n и D_p , которые характеризуют процесс диффузии. Дело в том, что оба указанных механизма определяются одной и той же причиной - столкновением носителей дефектами кристаллической решетки.

В условиях термодинамического равновесия $J_n = 0$, поэтому из (3.55) вытекает уравнение:

$$\frac{dn}{dx} + \frac{\mu_n}{D_n} \xi n = 0 \quad (3.57)$$

решение которого

$$n = A \exp \left[- \frac{\mu_n}{D_n} \int \xi dx \right] \quad (3.58)$$

Так как $-\int \xi dx = U = E/q$, где U - электростатический потенциал, то

$$n(E_2) = n(E_1) \exp \left[\frac{\mu_n (E_2 - E_1)}{D_n q} \right] \quad (3.59)$$

где E_1 и E_2 - некоторые пределы интегрирования. Сравнивая (3.29) и (3.59), получаем:

$$D_n = \frac{kT}{q} \mu_n \quad (3.60)$$

и аналогично имеем

$$D_p = \frac{kT}{q} \mu_p \quad (3.61)$$

Два последних равенства известны как соотношения Эйнштейна.

Рассуждая аналогично, убеждаемся, что всегда, в том числе и в равновесном состоянии, градиент концентрации связан с некоторым электрическим полем.

Объединяя формулы (3.60) и (3.61), получаем:

$$D_n = D_p / \mu_p = U_T \quad (3.62)$$

где $U_T = kT/q$ - так называемый температурный потенциал.

При $T = 300$ К имеем $U_T \approx 26$ мВ.

Выражения (3.60) и (3.61), относящиеся как к равновесному, так и к неравновесному состоянию, применимы при не слишком высоких концентрациях легирующих примесей, т. е. лишь для невырожденного полупроводника

Эффект Холла

Пусть через образец полупроводника протекает электрический ток

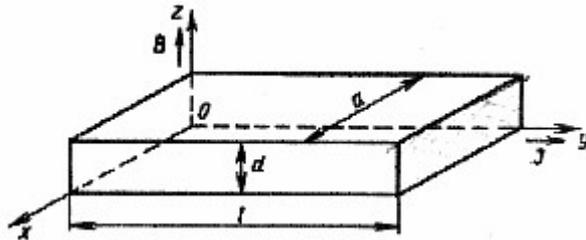


Рис. 3.7. Полупроводниковый образец, в котором наблюдается эффект Холла плотностью J . Образец помещен в слабое постоянное магнитное поле, индукция которого B перпендикулярна направлению тока. При этом в полупроводнике возникает электрическое поле, вектор напряженности которого нормален Π ноское 1 и, заданной векторами B и J (рис. 3.7). Описанное явление известно под названием эффекта Холла. Этот эффект служит важным средством, позволяющим экспериментально определить тип электропроводности, концентрацию основных носителей и их подвижность, а также измерить индукцию магнитного поля. Проводимый нами анализ не учитывает влияния соударений вдоль

оси x . Полученные результаты являются приближенными, однако они вполне применимы во многих практических случаях.

Скорости дрейфа электронов и дырок противоположно направлены, и знаки их зарядов также противоположны, поэтому на них действует одна и та же сила Лоренца, направленная вдоль положительной оси x :

$$F_x = qv_y B_z \quad (3.63)$$

где v_y – скорость дрейфа носителей.

Так как рассматриваемый образец изолирован и ток вдоль оси x не может быть замкнутым, то для уравнивания силы F_x на противоположных гранях образца должны накапливаться заряды. В результате возникает электрическое поле, перпендикулярное плоскости, в которой лежат векторы \mathbf{B} и \mathbf{J} .

Электростатическая сила $q\xi_x$ уравнивает силу F_x , т. е.

$$q\xi_x = qv_y B_z \quad (3.64)$$

Исходя из геометрии задачи, находим:

$$J_y = Mqv_y \quad (3.65)$$

где M - концентрация основных носителей. Следовательно,

$$q\xi_x = J_y B_z / M \quad (3.66)$$

или

$$R_n = \xi_x / (J_y B_z) = r_H / (qM). \quad (3.67)$$

Данную величину принято называть коэффициентом Холла. Для полупроводников n -типа и большинства металлов этот коэффициент отрицательный, а для полупроводников p -типа - положительный. Параметр r_n , входящий в последнюю формулу, называется множителем Холла, его значение колеблется от 1 до 1,5. Можно экспериментально найти значения трех величин, содержащихся в формуле (3.67), и получить, таким образом, концентрацию основных носителей M . Далее, измерив сопротивление образца, можно найти произведение μM , что позволяет определить подвижность основных носителей. Если основными носителями являются электроны, то ток:

$$I = adq\mu_n n \varepsilon_y \quad (3.68)$$

так как площадь сечения $A=ad$ и $J=I/A$. Приняв во внимание, что

$$F_x = qv_y B_z = q\mu_n \varepsilon_y B_z = +q\varepsilon_x, \quad (3.69)$$

можно, объединив формулы (3.68) и (3.69), получить

$$\varepsilon_x = \frac{\mu_n I B_z}{adq\mu_n n} \quad (3.70)$$

или

$$U_H = -a\varepsilon_x = -IB_z / (dq_n) \quad (3.71)$$

Последняя формула определяет так называемое *напряжение Холла*. Можно заметить, что знак U_H зависит от типа основных носителей, т.е. от типа полупроводника. Очевидно также, что величина U_H связана с концентрацией основных носителей заряда.

3.2. Неравновесные процессы в полупроводниках

Если в полупроводнике нарушено состояние термодинамического равновесия, то $np \neq n_i^2$. Неравновесное состояние возникает под действием какого-либо внешнего фактора, приводящего к изменению концентрации носителей.

Большинство полупроводниковых приборов работает в динамически неравновесных режимах. Поэтому важно изучить, каким образом устанавливается неравновесное состояние, каковы законы переноса носителей заряда при этих условиях, а также как система возвращается в равновесное состояние. Обсуждаемые здесь закономерности будут нами широко применяться при изучении физических процессов в полупроводниковых приборах. В каждом конкретном случае будут введены соответствующие начальные и граничные условия.

Процессы генерации и рекомбинации

При исследовании дрейфовой электропроводности указывалось, что хотя столкновения изменяют траектории носителей заряда, однако в равновесном состоянии общий эффект оказывается равным нулю. Изменение траектории не является единственным следствием столкновений. Действительно, может случиться так, что из-за столкновения энергия носителя заряда изменится, и он перейдет на другой разрешенный уровень, обусловленный наличием примесей. Может случиться, например, что электрон, первоначально находившийся в зоне проводимости, в результате столкновения с фононом (квантом колебаний кристаллической решетки) займет место дырки в валентной зоне. В данном случае говорят о явлении *рекомбинации* электрона. Рассуждая аналогично, можно допустить переход электрона из валентной зоны в зону проводимости, при этом говорят о *генерации* электронно-дырочной пары. При таких столкновениях, которые невозможны в условиях термодинамического равновесия, происходит перераспределение концентрации носителей, находящихся в различных энергетических зонах. Подобное явление называют процессом *генерации-рекомбинации* или просто процессом *рекомбинации* носителей.

Существуют различные механизмы генерации-рекомбинации. В дальнейшем будут изучены два из них. Прямые, или межзонные, процессы и непрямые процессы, обусловленные наличием промежуточных центров рекомбинации в объеме на поверхности кристалла. Оба эти механизма играют существенную роль. При изучении явления

рекомбинации предполагается, что полупроводник однороден, невырожден, электрически изолирован и находится при постоянной температуре. Считается также, что отклонение от равновесного состояния невелико.

Уровни инжекции

Неравновесная концентрация носителей заряда в полупроводнике не совпадает с равновесной. Если произведение $np > n_i^2$, то отклонение положительно и имеет место *инжекция* избыточных носителей. Если же $np < n_i^2$, то отклонение отрицательно; говорят, что при этом наблюдается *экстракция* носителей.

Если инжекция характеризуется неравенством

$$np - n_i^2 = M^2 \quad (3.72)$$

где M - избыточная концентрация основных носителей (n'_n, p'_n, n'_p или p'_p), то отклонение от равновесного состояния невелико. При этом говорят, что имеет место *низкий уровень инжекции*. Если же

$$np - n_i^2 \ll M^2$$

то имеет место *высокий уровень инжекции*.

При аналитическом исследовании полупроводниковых приборов всюду, где это возможно, уровень инжекции предполагается низким, так как то упрощает решение задачи.

В общем случае свойства полупроводникового материала зависят от числа инжектированных носителей заряда.

Межзонные процессы

В тех случаях, когда происходит инжекция носителей заряда, необходимо знать изменение скорости генерации-рекомбинации, а также механизмы возникновения этого явления.

Электрон, находящийся в зоне проводимости, может совершить переход в валентную зону и рекомбинировать с дыркой. Скорость рекомбинации, т.е. число носителей, рекомбинирующих в единице объема за 1 с, зависит от расположения вакантных уровней в валентной зоне (в данном случае это уровни, занятые дырками) и пропорциональна произведению np :

$$R = rnp \quad (3.74)$$

где r - коэффициент рекомбинации, имеющий статистическую природу и зависящий от температуры. Если параметр G_0 представляет скорость термической генерации электронно-дырочных пар при равновесии (рис.3.8), то

$$G_0 = R_0 = r \cdot n_0 \cdot p_0 \quad (3.75)$$

В том случае, когда происходит стационарное возбуждение полупроводника, например под действием электрического поля или света, возникает некоторая дополнительная скорость генерации электронно-дырочных пар G (рис. 3.8). Если при этом считать, что избыточные концентрации носителей не превышают соответствующих равновесных значений, как это имеет место в диодах и транзисторах.

$$R = rnp = r(n_0 + n')(p_0 + p') \quad (3.76)$$

$$G_i + G_0 + G \quad (3.77)$$

где G_i — общая скорость генерации; n' — $n - n_0$ и p' — $p - p_0$ — избыточные концентрации.

Так как возбуждение стационарно, то $R = G_i$ и истинная скорость процесса — общая скорость процесса

$$G = R - G_0 \quad (3.78)$$

$$G = r(n_0 + n')(p_0 + p') - rn_0 p_0 \quad (3.79)$$

Из условия электрической нейтральности следует, что $n' = p'$. Так как рассматриваемый процесс происходит при низком уровне инжекции, что $p' \ll n_0 + p_0$ и поэтому

$$G = r(n_0 + p_0)p' \quad (3.80)$$

Если материал является полупроводником n -типа, то $n_{n0} \gg p_{n0}$ так что

$$G_p = rn_{n0}(p_n - p_{n0}) \quad (3.81)$$

Если ввести величину $V_p = G_p$, представляющую скорость изменения концентрации дырок за счет рекомбинации, то получим уравнение

$$\frac{dp_n}{dt} = -V_p = \frac{p_n - p_{n0}}{\tau_p} \quad (3.82)$$

Скорость изменения концентрации электронов в полупроводнике p -типа за счет рекомбинации:

$$\frac{dn_p}{dt} = -V_n = \frac{n_p - n_{p0}}{\tau_n} \quad (3.83)$$

В двух последних формулах параметры $\tau_p = 1/(rn_{n0})$ и $\tau_n = 1/(rp_{p0})$ есть среднее время жизни дырок неосновных носителей в материале n -типа) и электронов (неосновных носителей в материале p -типа) соответственно.

Время жизни представляет оценку времени, в течение которого носитель (дырка) ведет себя как свободная частица, рекомбинируя затем с носителем заряда противоположного знака (электроном). Это время

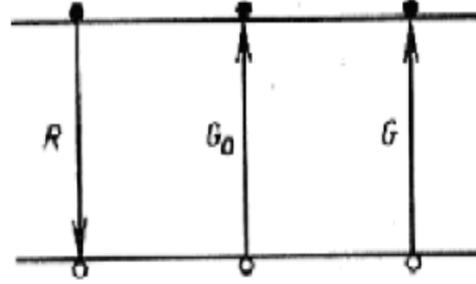


Рис. 3.8 Межзонные процессы: R — скорость электронно-дырочной рекомбинации в валентной зоне; G_0 — скорость термической генерации пар при равновесии; G — скорость генерации пар за счет внешнего возбуждения

колеблется от нескольких миллисекунд до нескольких микросекунд и существенно превышает время между двумя столкновениями (напомним, что неосновной носитель испытывает 10^6 - 10^8 столкновений, прежде чем рекомбинировать).

Процесс межзонной рекомбинации преобладает в полупроводниках с прямой запрещенной зоной, таких, как GaAs (см. рис. 2.9), поскольку здесь могут существовать переходы между потолком валентной зоны и дном зоны проводимости, когда волновой вектор k неизменен. Прямой механизм рекомбинации сопровождается излучением энергии в виде фотонов. По этой причине арсенид галлия используют для изготовления светоизлучающих диодов.

Результаты экспериментов показывают, что значения величин существенно меньше тех, которые входят в уравнения (3.82) и (3.83). Дело в том, что определенное влияние оказывают непрямые процессы генерации-рекомбинации, обусловленные наличием промежуточных уровней в запрещенной зоне.

Непрямые процессы

Непрямые процессы, основанные на использовании центров рекомбинации, имеют место во всех случаях, когда структуру кристаллической решетки нарушают примесные атомы. Чтобы упростить анализ, рассмотрим такие процессы, отдельно происходящие в объеме и на поверхности полупроводника.

Процессы в объеме полупроводника

Центры рекомбинации размещаются в запрещенной зоне и обуславливают обмен носителями между этой зоной и разрешенными зонами. Центр рекомбинации, расположенный вблизи середины запрещенной зоны (энергетический уровень ϵ_c), называют глубоким центром. Чтобы его создать, в качестве примесей используют такие элементы, как золото, платина, палладий, а также их сплавы. Центр рекомбинации, расположенный вблизи границы разрешенной зоны, называют неглубоким центром. Для его создания нужны примесные атомы элементов, принадлежащие III и V группам Периодической системы элементов. Процессы генерации-рекомбинации имеют место преимущественно вблизи уровня E_i потому что здесь примесные уровни почти в равной мере способны принимать носители обоих видов. Центры, удаленные от уровня E_i малоэффективны, так как здесь вероятность захвата носителя противоположного знака и снижается вероятность захвата носителя противоположного знака.

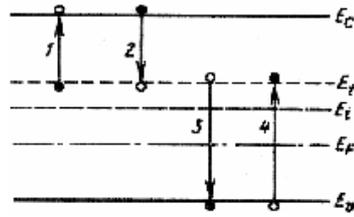


Рис. 3.9. Непрямые объемные процессы в простейшем случае одиночной примеси, создающей энергетический уровень E_i

Рассмотрим простой случай примеси, которая создает некоторый уровень E_i (рис. 3.9). Обозначим символом N_i концентрацию центров рекомбинации, а символом f_i - вероятностью того, что центр будет занят электроном. Прежде всего проанализируем процессы эмиссии и захвата при термодинамическом равновесии, когда имеет место распределение Ферми - Дирака. Число занятых центров $N_i f_i$, а число незанятых центров $N_i(1-f_i)$.

Отсюда, используя методику, развитую применительно к межзонным процессам, и обращаясь к рис. 3.9, получаем:

$$R_{1эм} = C_{эмн} N_i f_i \quad (3.84)$$

$$R_{2зах} = C_{зах} N_i (1 - f_i) \quad (3.85)$$

$$R_{эм} = C_{эмр} N_i (1 - f_i) \quad (3.86)$$

$$R_{4зах} = C_{захр} p N_i f_i \quad (3.87)$$

где R – скорости эмиссии и захвата носителей; C - коэффициенты эмиссии и захвата; n и p - соответствующие концентрации носителей.

В равновесном состоянии $R_{1эм} = R_{2зах}$. Тогда, принимая во внимание, что и

$n = n_i \exp\left(\frac{E_F - E_i}{kT}\right)$, получаем

$$C_{эмн} = C_{зах} n_i \exp\left(\frac{E_F - E_i}{kT}\right) N_i (1 - f_i) \quad (3.88)$$

Если воспользоваться функцией распределения Ферми-Дирака, то отсюда получаем

$$C_{эмн} = C_{зах} n_i \exp\left(\frac{E_i - E_i}{kT}\right) \quad (3.89)$$

Аналогично из равенства $R_{3эм} = R_{4зах}$ имеем

$$C_{эмр} = C_{захр} n_i \exp\left(\frac{E_i - E_i}{kT}\right) \quad (3.90)$$

Оба последних выражения справедливы и в состоянии термодинамического равновесия.

Если система возбуждена в результате некоторого стационарного внешнего воздействия, то истинные скорости процессов генерации-рекомбинации в зоне проводимости и в валентной зоне оказываются одинаковыми:

$$V \equiv G = R_{2\text{зах}} - R_{1\text{эм}} = R_{4\text{зах}} - R_{3\text{эм}} \quad (3.91)$$

Отсюда, полагая, что коэффициенты захвата носителей каждого сорта равны, т.е. $C_{ЭМ P} = C_{3АХ P} = C$ получаем:

$$V = \frac{C \cdot N_i (pn - n_i^2)}{n + 2n_i \cdot \text{ch}[(E_i - E_i)/kT]} \quad (3.92)$$

Если имеется полупроводник n -типа, у которого $E_i = E_i$ причем реализуется режим низкого уровня инжекции ($n_n = n_{n0}$), то

$$V_p \approx \frac{P_n - P_{n0}}{\tau_p} \quad (3.93)$$

где $\tau_p = 1/(C_{зах p} N_i)$ - среднее время жизни неосновных носителей, которое не зависит от коэффициента захвата основных носителей. Аналогично, рассматривая полупроводник p -типа, получаем среднее время жизни $\tau_n = 1/(C_{зах n} N_i)$.

Удобно воспользоваться понятием диффузионной длины L . Это среднее расстояние, которое проходит неосновной носитель в процессе диффузионного движения до момента рекомбинации:

$$L = (D\tau)^{1/2} \quad (3.94)$$

где D - коэффициент диффузии.

Интересно рассмотреть физическую интерпретацию понятия среднего времени жизни на примере динамики носителей, возбуждаемых под действием света. Предположим, что в начальный момент времени $t = 0$ некоторый образец внезапно освещают, в результате чего возникает избыток носителей. На основании (3.82) можно записать дифференциальное уравнение, которое описывает изменение во времени концентрации избыточных неосновных носителей в образце полупроводника n -типа при низком уровне инжекции:

$$\frac{d p'_n}{dt} = -\frac{p'_n(t)}{\tau_p}$$

Решение этого уравнения имеет вид:

$$p'_n(t) = p_n(t) - p_n(0) = p'_n(t) e^{-t/\tau_p} \quad (3.95)$$

где постоянная величина $p'_n(0)$ - избыточная концентрация носителей при $t=0$. Можно видеть, что функция $p'_n(t)$ уменьшается по экспоненциальному закону до тех пор, пока при $t \rightarrow \infty$ не будет достигнуто равновесие. Величина τ_p есть постоянная времени рассматриваемого процесса и может быть определена как время, в течение которого избыточная концентрация уменьшается в e раз по сравнению с начальным значением.

При изготовлении диодов и транзисторов, которые используются в качестве высокоскоростных переключателей, следует стремиться к тому, чтобы значение t было малым. При увеличении параметра N_i путем легирования кремния атомами золота длительность переключения может достичь 1 нс.

Процесс, основанный на участии глубоких центров рекомбинации, превалирует в полупроводниках с непрямой запрещенной зоной, например в

кремнии (см. рис. 2.9). Здесь при переходе носителей с нижних уровней зоны проводимости на верхние уровни валентной зоны изменяются волновой вектор k и энергия E .

Поверхностные процессы

Строго говоря, нельзя непосредственно связывать поверхностную рекомбинацию с существованием промежуточных центров. Однако можно утверждать, что само существование поверхности полупроводника нарушает периодичность кристаллической решетки и обуславливает возникновение состояний Тамма и Шокли. Как следствие, в запрещенной зоне появляются энергетические уровни, которые действуют подобно объемным центрам рекомбинации. Число таких уровней чрезвычайно велико, следствием чего является весьма малое значение параметра t . С такой ситуацией приходится сталкиваться, например, в тех случаях, когда поверхность кремния покрывают слоем оксида, чтобы получить систему $\text{SiO}_2 - \text{Si}$. Данная система служит основой при создании полевых транзисторов с изолированным затвором, известных под названием МОП-транзисторов (металл - оксид - полупроводник).

Будем прежде всего полагать, что все состояния Тамма и Шокли связаны с некоторым энергетическим уровнем, расположенным в запрещенной зоне. Кроме того, пусть известна плотность N_s , состояний на квадратный сантиметр, которая считается постоянной в пределах некоторого слоя, лежащего под поверхностью. Чтобы свести поверхностный случай к объемному, можно для приповерхностного квазинейтрального слоя записать выражение, аналогичное (3.93), справедливое при низком уровне инжекции:

$$V_s = S(n_s - n_{s0}) = S(p_s - p_{s0}) \quad (3.96)$$

Здесь величина $S = S_n = S_p$ заменяет величину $1/\tau = 1/\tau_n = 1/\tau_p$, имеет размерность скорости (см/с) и называется скоростью поверхностной рекомбинации. Заметим, что по физическому смыслу эта скорость отличается от скорости объемной рекомбинации V .

Внутреннее электрическое поле

До сих пор всегда считалось, что примеси распределены в объеме равномерно. Рассмотрим теперь случай неоднородного распределения примесей вдоль оси x , которое может иметь, вообще говоря, как детерминированный, так и случайный характер.

Неоднородное легирование ведет к появлению в образце внутреннего электрического поля. Подбирая закон распределения примесей, можно улучшить рабочие характеристики ряда полупроводниковых приборов.

Если к образцу не приложено какое-либо напряжение, внешняя цепь разомкнута и выполняется условие термодинамического равновесия, то в одномерном случае $J_n = J_p = 0$. Отсюда, используя формулы (3.55) и (3.56), находим:

$$\varepsilon_i = \frac{D_p}{\mu_p} \frac{1}{p} \frac{dp}{dx} = -\frac{D_n}{\mu_n} \frac{1}{n} \frac{dn}{dx}$$

Используя соотношение Эйнштейна (2.62), получаем

$$\varepsilon_i = U_T \frac{1}{p} \frac{dp}{dx} = -U_T \frac{1}{n} \frac{dn}{dx} \quad (3.97)$$

На практике следует считать, что $n \approx N_d(x)$ и $p \approx N_a(x)$, так как при малых значениях ε_i выполняются условия электрической квазинейтральности.

Физически возникновение поля напряженностью ε_i обусловлено наличием градиента концентрации носителей. Действительно, если имеется некоторое неоднородное распределение донорных примесей, то свободные электроны стремятся продиффундировать в область с низкой концентрацией (скажем, в положительном направлении оси x). При этом положительные ионы с концентрацией N_d^+ неподвижны. В результате разделения положительных и отрицательных зарядов возникает внутреннее электрическое поле, препятствующее диффузионному потоку и установлению равновесия

$$\varepsilon_i = \frac{-U_T \left(b \frac{dn}{dx} - \frac{dp}{dx} \right)}{nb + p} \quad (3.98)$$

где $b = \mu_n / \mu_p$

Квазиуровни и квазипотенциалы Ферми

Уравнения (3.29) и (3.30), относятся к собственным и примесным полупроводникам в состоянии термодинамического равновесия. В неравновесном состоянии применять данные уравнения нельзя, так как из-за инжекции концентрации носителей не совпадают с равновесными значениями. Как следствие, уравнение (3.25) не выполняется. В отсутствие термодинамического равновесия принято вводить две новые величины E_{Fn} и E_{Fp} , которые заменяют величину E_F в уравнениях (3.29) и (3.30). Таким образом, в невырожденных полупроводниках справедливы уравнения

$$n = n_i \exp\left(\frac{E_{Fn} - E_i}{kT}\right) = N_i \exp\left(-\frac{E_i - E_{Fn}}{kT}\right) \quad (3.99)$$

$$p = n_i \exp\left(\frac{E_i - E_{Fp}}{kT}\right) = N_i \exp\left(-\frac{E_{Fp} - E_i}{kT}\right) \quad (3.100)$$

Входящие сюда величины E_{Fn} и E_{Fp} называют квазиуровнями Ферми электронов и дырок соответственно. Понятие квазиуровней Ферми широко используют при изучении неравновесных физических процессов в диоде с p - n -переходом.

Выполнив подстановки $\psi_n = -E_{Fn} / q$, $\psi_p = -E_{Fp} / q$, $\psi_i = -E_i / q$ преобразуем формулы (3.99) и (3.100) к виду

$$n = n_i \exp\left(\frac{\psi - \psi_n}{U_T}\right) \quad (3.101)$$

$$p = n_i \exp\left(\frac{\psi_p - \psi}{U_T}\right) \quad (3.102)$$

где ψ_n и ψ_p - так называемые квазипотенциалы Ферми электронов и дырок соответственно; ψ - электростатический потенциал.

Уравнение полупроводника (3.25) для неравновесного состояния принимает вид:

$$np = n_i^2 \exp\left(\frac{E_{Fn} - E_i}{kT}\right) = n_i^2 \exp\left(\frac{\psi_p - \psi_n}{U_T}\right) \quad (3.103)$$

Свойства квазиуровней

Отметим некоторые интересные свойства квазиуровней и квазипотенциалов Ферми, тесно связанных между собой.

- При равновесии $\psi_n = \psi_p = \psi$ и $np = n_i^2$
- Если инжекция носителей увеличивается, то:
 - а) разность $E_{Fn} - E_i$ увеличивается вместе с ростом n [см. формулу (3.99)] и квазиуровень E_{Fn} стремится к E_c ;
 - б) разность $E_i - E_{Fp}$ уменьшается вместе с уменьшением p [см. формулу (3.100)] и квазиуровень E_{Fp} стремится к E_v .
- На основании равенства (3.101) имеем

$$\frac{dn}{dx} = \frac{n}{U_T} \left(\frac{d\psi}{dx} - \frac{d\psi_n}{dx} \right) \quad (3.104)$$

Так как в неравновесной системе выполняются соотношения Эйнштейна (3.62), то плотность токов, определяемых формулами (3.55) и (3.56), можно представить в виде

$$J_n = \frac{\sigma_n}{q} \frac{dE_{Fn}}{dx} = -\sigma_n \frac{d\psi_n}{dx} \quad (3.105)$$

$$J_p = \frac{\sigma_p}{q} \frac{dE_{Fp}}{dx} = -\sigma_p \frac{d\psi_p}{dx} \quad (3.105')$$

Данные формулы отображают обобщенный закон Ома, учитывая дрейф и диффузию носителей обоих видов. В неравновесном состоянии полупроводника квазиуровень заменяет произведение $q\mathcal{E}$, которое используется при описании явлений в проводниках и однородных полупроводниках. Соотношения (3.105) указывают на то, что ток существует лишь в том случае, когда имеется градиент квазиуровня Ферми.

Явление переноса в динамически неравновесном состоянии

Изучив явления дрейфа и диффузии, а также механизмы генерации-рекомбинации, можно приступить к выводу фундаментальных уравнений, которые описывают явления переноса в полупроводниках с избыточными носителями.

Уравнение непрерывности

Идея, лежащая в основе вывода этого уравнения, заключается в том, что разность между числом носителей, входящих в некоторый объем и выходящих из него, всегда равна изменению числа частиц в объеме. Такое динамическое равновесие носителей имеет место в полупроводниках, состояние которых лишь незначительно отличается от состояния термодинамического равновесия.

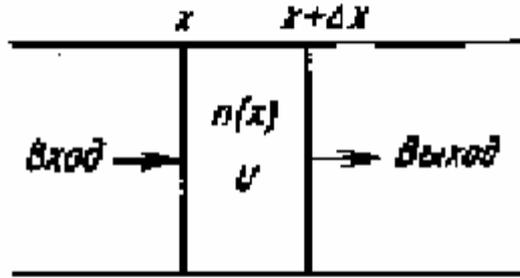


Рис. 3.10. Изменение концентрации электронов в полупроводнике.
«Вход» и «выход» представляют собой поперечные сечения
единичной площади

На рис. 3.10 изображен одномерный поток электронов в области длиной Δx . Предполагается, что полупроводник p -типа имеет единичную площадь поперечного сечения и наблюдается низкий уровень инжекции. Число электронов, входящих в указанную область за 1 с в направлении оси x вследствие процессов дрейфа и диффузии

$$-\mu_n(n_p \varepsilon)_x - D_n \left(\frac{dn_p}{dx} \right)_x$$

Число электронов, покидающих данную область

$$\mu_p(n_p \varepsilon)_{x+\Delta x} + D_n \left(\frac{dn_p}{dx} \right)_{x+\Delta x}$$

Так как скорости генерации и рекомбинации, отнесенные к единичному объему, равны G_n и V_n то уравнение непрерывности для электронов можно представить в виде:

$$\mu_p(n_p \varepsilon)_{x+\Delta x} - \mu_n(n_p \varepsilon)_x + D_n \left(\frac{dn_p}{dx} \right)_{x+\Delta x} - D_n \left(\frac{dn_p}{dx} \right)_x + (G_n - V_n)\Delta x = \frac{\partial}{\partial t}(n_p \Delta x)$$

где правая часть представляет изменение числа электронов в области Δx за время ∂t . Деля обе части данного равенства на Δx и переходя к пределу при $\partial t \rightarrow 0$, получаем уравнение непрерывности для электронов

$$\mu_n \frac{d}{dx}(n_p \varepsilon) + D_n \frac{d^2 n_p}{dx^2} + G_n - \frac{n_p - n_{p0}}{\tau_n} = \frac{\partial n_p}{\partial t} \quad (3.106)$$

Аналогично выводится уравнение непрерывности для дырок в полупроводнике p -типа:

$$-\mu_p \frac{d}{dx}(p_n \varepsilon) + D_p \frac{d^2 p_n}{dx^2} + G_p - \frac{p_n - p_{n0}}{\tau_p} = \frac{\partial p_n}{\partial t} \quad (3.107)$$

Здесь учитывается, что при низком уровне инжекции, когда концентрация инжектированных носителей значительно меньше концентрации основных носителей в равновесном состоянии, можно воспользоваться приближенными равенствами

$$V_n = \frac{n_p - n_{p0}}{\tau_n}, \quad V_p = \frac{p_n - p_{n0}}{\tau_p}$$

Уравнение Пуассона

В общем случае уравнения непрерывности являются нелинейными, так как напряженность электрического поля ε , а также времена жизни носителей τ_n и τ_p могут быть функциями координаты x (в трехмерных задачах координат x , y и z), а также времени t .

Напряженность электрического поля ε определяют с помощью теоремы Гаусса, согласно которой

$$d\varepsilon/dx = \rho/\varepsilon_{an} \quad (3.108a)$$

или на основании уравнения Пуассона

$$d^2 U/dx^2 = -\rho/\varepsilon_{an} \quad (3.108b)$$

где ρ - объемная плотность электрического заряда; ε_{an} - абсолютная диэлектрическая проницаемость полупроводника, равная произведению $\varepsilon_n \cdot \varepsilon_0$. Так как объемная плотность заряда

$$\rho = q[p + N_d - (n + N_a)] \quad (3.109)$$

$$\frac{d\varepsilon}{dx} = \frac{q}{\varepsilon_{an}} [p - n + N_d - N_a] \quad (3.110)$$

Уравнения (3.106), (3.107) и (3.110) вместе с соотношениями (3.55) и (3.56) описывают одномерное распределение плотности носителей и напряженности электрического поля. Эти уравнения можно решить численными методами при соответствующих начальных и граничных условиях. Возможен и аналитический подход, основанный на некоторых упрощающих предположениях.

3.3. Основные уравнения, описывающие процессы в полупроводниковых приборах

Ранее в одномерной постановке были рассмотрены статистические явления, процессы генерации-рекомбинации, а также явления переноса. Все они характеризуют макроскопические свойства полупроводников. Приведем в заключение сводку основных уравнений, объединив их в три группы.

а) Уравнения, описывающие плотность тока

Для одномерного случая (3.55) (3.56) запишем в виде

$$J_n = q\mu_n n\varepsilon + qD_n \frac{dn}{dx} \quad (3.111)$$

$$J_p = q\mu_p n\varepsilon + qD_p \frac{dp}{dx} \quad (3.112)$$

В трехмерном случае

$$J_n = q\mu_n n\varepsilon + qD_n \nabla n \quad (3.113)$$

$$J_p = q\mu_p n\varepsilon + qD_p \nabla p \quad (3.114)$$

$$J = J_n + J_p \quad (3.115)$$

где μ_n , μ_p , D_n и D_p - параметры невырожденного полупроводника, определяемые при некоторой заданной температуре T и известной концентрации примесей $N=N_d - N_a$.

б) Уравнения непрерывности

Уравнения (3.106) и (3.107) для одномерного случая:

$$\frac{\partial n_p}{\partial t} = \frac{n_p - n_{p0}}{\tau_n} + \mu_n \frac{d}{dx}(n_p \varepsilon) + D_n \frac{d^2 n_p}{dx^2} \quad (3.116)$$

$$\frac{\partial p_n}{\partial t} = \frac{p_n - p_{n0}}{\tau_p} + \mu_p \frac{d}{dx}(p_n \varepsilon) + D_p \frac{d^2 p_n}{dx^2} \quad (3.117)$$

В трехмерном случае:

$$\frac{\partial n}{\partial t} = G_n - V_n + \frac{1}{q} \nabla J_n \quad (3.118)$$

$$\frac{\partial p}{\partial t} = G_p - V_p + \frac{1}{q} \nabla J_p \quad (3.119)$$

Здесь G_n и G_p - скорости генерации носителей под действием внешних, например оптических или электрических, факторов; V_n и V_p - скорости рекомбинации электронов и дырок.

в) Уравнение Пуассона

Перепишем уравнение (3.110) для одномерного случая:

$$\frac{d\varepsilon}{dx} = \frac{q}{\varepsilon_{an}} [p - n + N_d - N_a] \quad (3.120)$$

В трехмерных задачах:

$$\nabla \varepsilon = \frac{q}{\varepsilon_{an}} [p - n + N_d - N_a]$$

Напряженность электрического поля \mathcal{E} представляет сумму напряженностей приложенного (внешнего) поля \mathcal{E}_n и внутреннего поля \mathcal{E}_i . Так как внешнее поле внутри полупроводника не имеет источников и стоков, то $\nabla \mathcal{E}_n = 0$ и поэтому, по определению,

$$\nabla \mathcal{E}_i = \frac{q}{\varepsilon_{an}} [p - n + N_d - N_a] \quad (3.121)$$

Приняв во внимание соответствующие начальные и граничные условия, на основании основных уравнений можно вычислить значения величин J_n , J_p , n , p и \mathcal{E} в каждой точке среды и в любой момент времени.

4. Физические явления в p-n-переходе.

Полупроводниковые диоды

Начиная с 1954 г., после того как Бардин, Браттейн и Шокли создали транзистор, в электронике постоянно наблюдаются быстрые изменения. В начале диоды, транзисторы и другие приборы изготавливали из германия. Через несколько лет основным материалом для производства полупроводниковых приборов стал кремний. В настоящее время кремний и арсенид галлия являются важнейшими материалами, используемыми для изготовления как дискретных приборов, так и изделий микроэлектроники.

Данная глава посвящена рассмотрению физических процессов в *p-n*-переходах, а также теории диодов, содержащих *p-n*-переход. Уделено внимание и другим структурам с *p-n*-переходами, в частности представляющим особый интерес в настоящее время диодам — с переходом металл—полупроводник, с гетеропереходом, светоизлучающим диодам и инжекционным лазерам.

Переход *p-n* является, прежде всего, полупроводниковым выпрямителем. Кроме того, этот переход служит важной составной частью многих дискретных и полупроводниковых приборов, интегральных схем (биполярные транзисторы, полевые транзисторы с управляющим *p-n*-переходом и т. д.).

4.1. Равновесное состояние p-n перехода

Пусть имеется образец монокристалла, изготовленного таким образом, что относительная доля примесных атомов составляет 10^{-11} . При этом можно

считать, что данный кристалл является собственным полупроводником. Из монокристалла вырезается тонкая пластина, в которую внедряются донорная или акцепторная примесь. Для этого можно использовать, например, управляемый процесс диффузии или вырастить на поверхности эпитаксиальный слой. В результате образуется $p-n$ -переход.

Примесный полупроводник, используемый для создания $p-n$ -перехода, должен содержать незначительную долю примесных атомов, например один примесный атом на сто миллионов атомов чистого вещества. Это нужно для того, чтобы носители, пересекающие $p-n$ -переход, не испытывали рекомбинации (Исходный кристалл должен быть высокоомным. При этом в нем возникает широкий $p-n$ -переход, имеющий высокое пробивное напряжение).

Переходы мощных дискретных приборов имеют квадратную форму со стороной примерно 3 мм; толщина переходов ИС не превышает 0,01 мм.

В полупроводнике с областями p - и n -типов, образующими переход, можно выделить следующие пространственные области: *металлургический переход* (контакт) (воображаемая плоскость, разделяющая p - и n -области), *область перехода*, или *область пространственного заряда*, или *обедненная область* (располагается по обе стороны металлургического перехода и имеет толщину от 10^{-6} до 10^{-4} см в зависимости от технологии производства), *нейтральные области* (p - и n -области), лежащие между областью пространственного заряда и границами полупроводников p - и n -типов, и, наконец, *омические контакты*, которыми оканчиваются нейтральные области.

В дальнейшем будет построена равновесная модель *резкого*, или *ступенчатого*, $p-n$ -перехода, в котором концентрация примесных атомов скачком изменяется от значения N_a в p -области до значения N_d в n -области. Резкий переход не является структурой, типичной для современных приборов. Тем не менее, такая упрощенная модель позволяет проанализировать наиболее важные характеристики, например вольтамперные. Другая причина, позволяющая воспользоваться столь простой моделью, заключается в том, что внутренние физические процессы и электрические свойства перехода лишь в малой степени зависят от способа его изготовления.

В случае резкого перехода предполагается, что концентрации акцепторов N_a в p -области и доноров N_d в n -области постоянны. Если $N_a = N_d$, то резкий переход называют *симметричным*. Если же $N_a \neq N_d$, то переход является *несимметричным*; при $N_a > N_d$ переход обозначают как p^+-n , при $N_a < N_d$ говорят о $p-n$ -переходе (этот случай на практике встречается чаще).

Будем рассматривать $p-n$ переход в состоянии термодинамического равновесия, т. е. при отсутствии внешних воздействий, таких, как внешнее напряжение. Математическая модель перехода строится на базе таких понятий, как *высота потенциального барьера* (контактная разность потенциалов) U_0 , *толщина области перехода* W , *максимальная напряженность внутреннего электрического поля* \mathcal{E}_{\max} и *плотность электрического заряда* Q .

Следствием предположения о термодинамическом равновесии является то, что:

- суммарные токи электронов и дырок в каждой точке объема полупроводника равны нулю ($I_n = I_p = 0$);
- квазиуровень Ферми $E_{F_n} = E_{F_p} = E_F$ постоянен для всего полупроводника; квазипотенциалы $\psi_n = \psi_p = \psi_F = const$. При этом энергетические зоны изогнуты и полностью заполнены во всех точках полупроводникового материала, в которых $np = n_i^2$.

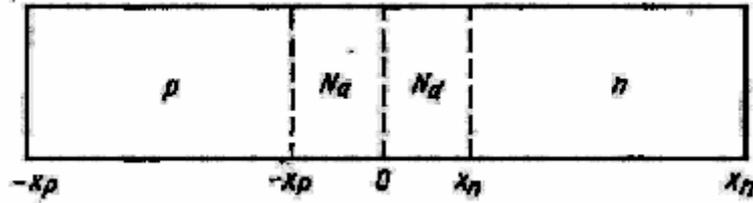


Рис. 4.1. Одномерная модель резкого p - n -перехода:
 x_p и x_n — границы области пространственного заряда, имеющей на практике толщину от 10^{-6} до 10^{-4} см ($W = x_n + x_p$; $W_n = X_n - x_n$; $W_p = X_p - x_p$)

Кроме того, принимается *гипотеза обеднения*, состоящая в том, что весь полупроводник с p - n -переходом мысленно разбивают на область перехода и на две нейтральные области. При этом в области перехода пренебрегают носителями заряда (рис. 4.1).

В равновесном состоянии $J_n = 0$ (это условие, относящееся как к дрейфовому току, так и к току диффузии, часто служит отправной точкой построения теории). Тогда на основании уравнения (3.111) можно записать

$$J_n = qn\mu_n \varepsilon + qD_n \frac{dn}{dx} = 0 \quad \text{и получить}$$

$$\varepsilon = -\frac{D_n}{n\mu_n} \frac{dn}{dx} \quad (4.1)$$

Отсюда, используя соотношение Эйнштейна (3.60) и равенство

$$\varepsilon = -\frac{d\psi(x)}{dx}, \quad \text{находим}$$

$$\frac{d\psi(x)}{dx} = \frac{U_T}{n} \frac{dn}{dx}, \quad (4.2)$$

где $U_T = kT/q$ - температурный потенциал.

Интегрируя по x в пределах от $-x_p$ до x_n , получаем

$$\psi(x_n) - \psi(-x_p) = U_T \ln \frac{n(x_n)}{n(-x_p)}. \quad (4.3)$$

Эта формула определяет высоту потенциального барьера p - n -перехода (рис. 4.2), равную

$$U_0 = \psi(x_n) - \psi(-x_p). \quad (4.4)$$

Так как примеси считаются ионизированными, то $n(x_n) = N_d$ и $n(-x_p) = n_i^2/p(-x_p)$, откуда находим

$$U_0 = U_T \ln \frac{N_a N_d}{n_i^2}. \quad (4.5)$$

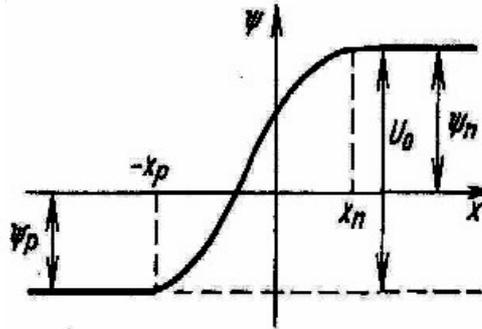


Рис. 4.2. Кривая распределения потенциала, знак которого противоположен знаку изменения энергии (см. рис. 4.3). График можно получить, проведя двукратное интегрирование уравнения Пуассона (4.6)

Это то напряжение, которое возникает в условиях термодинамического равновесия и ведет к прекращению диффузионного тока. Такое же выражение получается из формулы (3.112), в которой нужно положить $J_p = 0$.

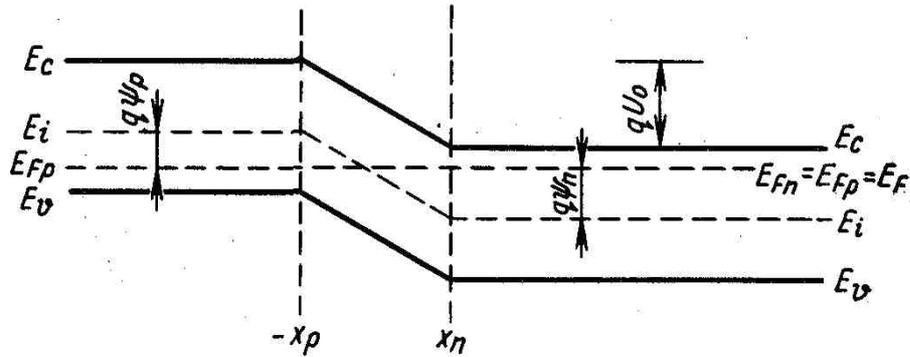


Рис. 4.3. Диаграмма энергетических уровней E_c и E_v для некоторого p - n -перехода в равновесном состоянии. Квазиуровни Ферми и при равновесии совпадают с уровнем Ферми E_F электростатический потенциал

Поскольку условия термодинамического равновесия считаются выполненными, единственным уравнением из трех групп, приведенных в конце гл. 3, остается уравнение Пуассона, в котором

$$N(x) = N_d^+ - N_a^- : \quad \frac{d\varepsilon}{dx} = \frac{d^2\psi}{dx^2} = \frac{q}{\varepsilon_{an}} [p - n - N(x)].$$

На основании гипотезы обеднения ($p \approx n \approx 0$) это уравнение принимает вид

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon_{an}} N(x). \quad (4.6)$$

Здесь ψ — электростатический потенциал, который, как это видно из формул (3.101) и (3.102), соответствует выражению $\psi = -E_i / q$.

Как можно видеть из рис. 4.1, в области перехода на отрезке $-x_p < x < 0$ $N(x) = N_a$. Поэтому уравнение (4.6) принимает вид

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon_{an}} N_a.$$

Интегрируя его, получаем

$$\psi(x) = -\frac{q}{\varepsilon_{an}} N_a \left(\frac{x^2}{2} + xx_p \right) + C_1. \quad (4.7)$$

Так как $\varepsilon = -d\psi / dx$, то напряженность электрического поля

$$\varepsilon = -\frac{q}{\varepsilon_{an}} N_a (x + x_p). \quad (4.8)$$

В области $0 \leq x \leq x_n$ концентрация $N(x) = N_d$; тогда из уравнения (4.6) следует, что

$$\frac{d^2\psi}{dx^2} = -\frac{q}{\varepsilon_{an}} N_d; \quad (4.9)$$

$$\psi(x) = -\frac{q}{\varepsilon_{an}} N_d \left(\frac{x^2}{2} + xx_n \right) + C_2. \quad (4.10)$$

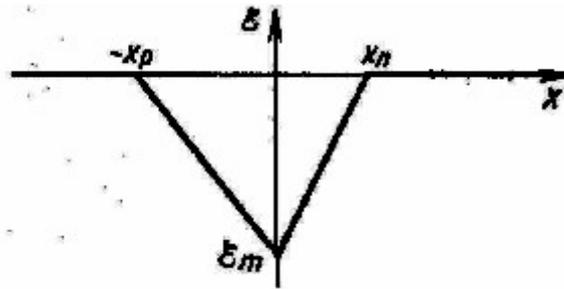


Рис. 4.4. Распределение напряженности электрического поля ε , полученное интегрированием уравнения Пуассона

Величина ε максимальна при $x=0$ (рис. 4.4); на основании равенств (4.8) и (4.10) имеем

$$\varepsilon_m = -\frac{q}{\varepsilon_{an}} N_a x_p = -\frac{q}{\varepsilon_{an}} N_d x_n, \quad (4.11)$$

откуда

$$N_a x_p = N_d x_n, \quad (4.12)$$

что является условием электрической нейтральности перехода. В соответствии с рис. 4.1 толщина обедненного слоя перехода

$$W = x_p + x_n. \quad (4.13)$$

Так как при $x = 0$ потенциал ψ должен быть непрерывным, то постоянные интегрирования в выражениях (4.8) и (4.10) должны быть одинаковыми:

$C_1 = C_2 = C$. Как следствие этого, используя формулы (4.13), (4.12), (4.4), (4.8) и (4.10), находим

$$W = \left[\frac{2\varepsilon_{an}}{q} \left(\frac{N_a + N_d}{N_a N_d} \right) U_0 \right]^{1/2} \quad (4.14)$$

В соответствии с равенством (4.12) плотность заряда, отнесенного к единичной площади, по обе стороны от сечения $x=0$ (рис. 4.5)

$$|Q| = qx_n N_d = qx_p N_a \quad (4.15)$$

или

$$|Q| = \left[2\varepsilon_{an} q \left(\frac{N_a + N_d}{N_a N_d} \right) U_0 \right]^{1/2} \quad (4.16)$$

Выражения (4.5), (4.11), (4.14) и (4.16) образуют математическую модель резкого p - n -перехода, находящегося в термодинамическом равновесии.

Как указывалось в начале параграфа, здесь рассматривается лишь случай резкого перехода. Однако определенный интерес представляет и плавный переход. Плавный переход - это структура, в которой концентрации легирующих примесей изменяются вдоль перехода линейно (вместо величин N_a и N_d указываются их градиенты). Часто такая структура служит хорошей приближенной моделью p - n -переходов кремниевых ИС, изготовленных по планарной технологии, так как в таких приборах примесные атомы диффундируют в область перехода со стороны поверхности.

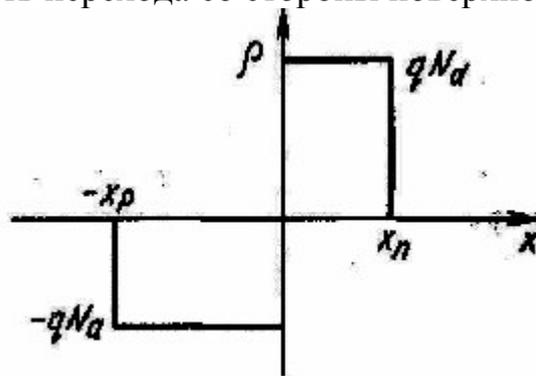


Рис. 4.5. Распределение плотности пространственного заряда, входящей в уравнение Пуассона (4.6)

4.2. Работа p-n- перехода при внешнем напряжении.

Полупроводниковый диод

Если к внешним выводам перехода приложить постоянное напряжение U , то это изменяет потенциалы и токи в переходе. При $U > 0$ говорят, что переход смещен в прямом направлении. Если же $U < 0$, то он смещен в обратном направлении. Различие свойств перехода при прямом и обратном напряжениях определяет возможность его работы в качестве выпрямляющего диода.

Качественное рассмотрение свойств смещенного перехода (при внешнем напряжении)

На рис. 4.6 изображены энергетические зоны для диода с p - n -переходом при прямом ($U > 0$) и при обратном ($U < 0$) напряжении. Можно заметить, что подача внешнего напряжения приводит к смещению квазиуровней Ферми относительно равновесного положения (см. рис. 4.2). Если $U > 0$, то эта величина вычитается из U_0 и ширина обедненной области уменьшается. Если же $U < 0$, то напряжения U и U_0 суммируются, что ведет к расширению обедненной области.

В соответствии с этим, приняв во внимание формулу (4.14), относящуюся к равновесному состоянию, можно записать следующее выражение для толщины

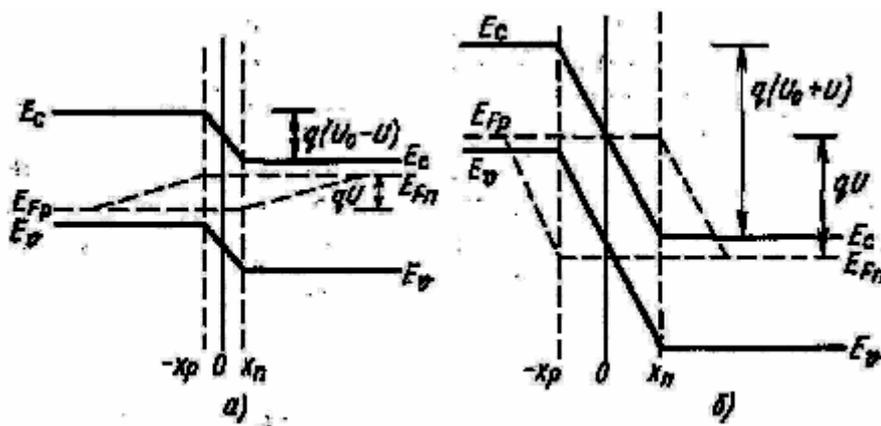


Рис. 4.6. Диаграмма энергетических зон и квазиуровней Ферми при прямом (а) и обратном (б) смещении перехода. Если $U > 0$ (прямое смещение), то оба квазиуровня Ферми изменяются на величину qU таким образом, что высота потенциального барьера снижается. Если же $U < 0$ (обратное смещение), то изменение квазиуровней ведет к увеличению высоты барьера

обедненной области в резком переходе, к которому приложено внешнее напряжение:

$$W = \left[\frac{2\varepsilon_{an}}{q} \left(\frac{N_a + N_d}{N_a N_d} \right) (U_0 - U) \right]^{1/2} \quad (4.17)$$

Формула справедлива как в случае $U > 0$, так и в случае $U < 0$.

Снижение высоты потенциального барьера при $U > 0$ позволяет основным носителям пересекать область перехода; при этом они становятся неосновными носителями, создавая заметный ток. При $U > 0$ эффекты диффузии более ощутимы, чем эффекты дрейфа.

Смещение квазиуровней Ферми относительно равновесного положения соответственно изменяет диаграмму энергетических зон. Интересно проследить за последовательностью построения такой диаграммы, определяющей свойства перехода.

Вначале изображают квазиуровень Ферми для любой из двух нейтральных областей диода и дополняют его соответствующими зоной проводимости и валентной зоной (если рассматривается область p -типа, то вблизи квазиуровня Ферми находится валентная зона). После этого с учетом направления смещения диода строят квазиуровень Ферми для другой нейтральной области, а затем соответствующие зону проводимости и валентную зону, причем ширина запрещенной зоны постоянна. И наконец, зоны проводимости, валентные зоны и квазиуровни Ферми соединяют через обедненную область (рис. 4.6). Если требуется больше информации, можно показать детали, определяемые свойствами свободных носителей в нейтральных областях.

Барьерная емкость обратносмещенного перехода

Так как толщина обедненной области зависит от смещающего напряжения, то эта область ведет себя как нелинейный конденсатор, емкость которого зависит от приложенного напряжения.

Барьерная емкость, отнесенная к единице площади,

$$C_{БАР} = \left| \frac{dQ}{dU} \right| \quad (4.18)$$

где Q - поверхностная плотность заряда [см. формулы (4.15) или (4.16)]. При $U < 0$ это емкость некоторого конденсатора, к которому приложено напряжение $U_0 - U$; диэлектриком конденсатора служит область перехода. Поэтому формулу (4.16) можно записать в виде

$$Q = \left(2\varepsilon_{an} q \frac{N_a N_d}{N_a + N_d} \right)^{1/2} (U_0 - U)^{1/2} \quad (4.19)$$

Отсюда

$$C_{БАР} = \left(\frac{\varepsilon_{an} q}{2(U_0 - U)} \frac{N_a N_d}{N_a + N_d} \right)^{1/2} \quad (4.20)$$

Если площадь перехода равна A , то

$$C_{\text{БАР}} = \varepsilon_{\text{ан}} A / W \quad (4.21)$$

где величина W определяется формулой (4.17).

Одна из принципиальных особенностей, отличающих емкость $C_{\text{БАР}}$ от емкости обычного плоского конденсатора, состоит в том, что в переходе направление вектора напряженности электрического поля не зависит от полярности приложенного напряжения.

При анализе работы транзистора будет показано, что величина $C_{\text{БАР}}$ является одним из факторов, которые ограничивают его применение. В равной мере емкость $C_{\text{БАР}}$ является нежелательной для большинства диодов. Тем не менее иногда требуются конденсаторы, емкость которых могла бы изменяться под действием электрических сигналов. Диоды, в которых используется это свойство барьерной емкости, называют варакторами (полупроводниковыми варикапами.— *Прим. пер.*). Варакторы широко применяют для настройки приемников частотно-модулированных сигналов.

Формула, определяющая значение $C_{\text{БАР}}$, была выведена в предположении $U < 0$. Применять ее для случая $U > 0$ следует с осторожностью, так как здесь модель обеднения менее адекватна, чем в случае $U < 0$. Если $U > 0$, то емкость $C_{\text{БАР}}$ велика, однако одновременно существует еще большая диффузионная емкость, связанная с наполнением заряда носителей в нейтральных областях.

Математическая модель диода. Уравнение Шокли

Поставим целью вывести математическую модель диода с p - n -переходом, способную объяснить существенные особенности его работы в неравновесном состоянии. Сформулируем следующие исходные предположения.

1. В области пространственного заряда выполняются условия обеднения: $|p - n| \ll N = N_d - N_a$.
2. Все внешнее напряжение приложено к обедненной области, толщина которой определяется формулой (4.17).
3. В нейтральных областях наблюдается либо избыток неосновных носителей (переход смещен в прямом направлении), либо их недостаток (переход смещен в обратном направлении). Отсчет ведется от равновесной концентрации при низком уровне инжекции.
4. В обедненной области носители заряда не рекомбинируют, поэтому здесь токи электронов и дырок неизменны.
5. Квазиуровни Ферми E_{F_n} и E_{F_p} по обе стороны от перехода различны и зависят от приложенного напряжения.

Последнее предположение позволяет вычислить концентрацию неосновных носителей по обе стороны перехода. Действительно, в данном случае

применимы формулы (3.101) и (3.102). На границе между областью перехода и нейтральной n -областью при низком уровне инжекции ($n \approx n_{n0}$) имеем:

- концентрация основных носителей

$$n_n = n_{n0} = n_i \exp\left(\frac{\psi - \psi_n}{U_T}\right) \quad (4.22)$$

- концентрация неосновных носителей

$$p_n = n_i \exp\left(\frac{\psi_p - \psi}{U_T}\right) \quad (4.23)$$

откуда

$$n_{n0} p_n = n_i^2 \exp\left(\frac{\psi_p - \psi_n}{U_T}\right) \quad (4.24)$$

В то же время из рис. 4.6 следует, что $E_{F_n} = -q\psi_n$ и $E_{F_p} = -q\psi_p$ откуда $|\psi_p - \psi_n| = q(E_{F_p} - E_{F_n}) = qU$, так что формула (4.24) принимает вид

$$n_{n0} p_n = n_i^2 \exp\left(\frac{U}{U_T}\right) \quad (4.25)$$

поскольку в равновесном состоянии $n_{n0} p_{n0} = n_i^2$, из равенства (4.25) следует, что

$$p_n(x) = p_{n0}(x_n) \exp\left(\frac{U}{U_T}\right) \quad (4.26)$$

Аналогично выводится формула, определяющая концентрацию неосновных носителей на границе перехода, примыкающей к p -области,

$$n_p(-x_p) = n_{p0}(-x_p) \exp\left(\frac{U}{U_T}\right)$$

так что имеет место равенство

$$\frac{p_n(x_n)}{p_{n0}(x_n)} = \frac{n_p(-x_p)}{n_{p0}(-x_p)} = \exp\left(\frac{U}{U_T}\right) \quad (4.27)$$

напоминающее соотношение Больцмана.

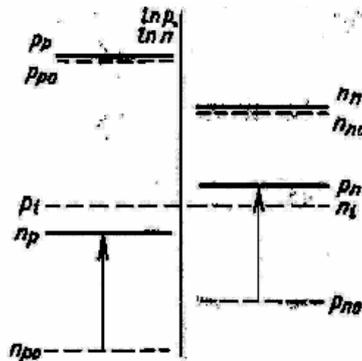


Рис. 4.7. Концентрации основных и неосновных носителей в p - n -переходе, находящемся под напряжением $U > 0$ при температуре 300 К. Использован логарифмический масштаб; значение n_i лежит между значениями концентраций основных и неосновных носителей

На рис. 4.7 в логарифмическом масштабе показаны концентрации основных и неосновных носителей в p - n -переходе при прямом напряжении. При таком выборе масштаба на графике в соответствии с условием равновесия уровни концентрации неосновных носителей по обеим сторонам находятся на одинаковом расстоянии, что указано стрелками. Заметим, что в p - n -переходе база - эмиттер биполярного транзистора основная часть носителей, определяющих работу транзистора, располагается в области с большей концентрацией примеси, а именно в p -области в случае, изображенном на рис. 4.7.

Уравнения (4.26) и (4.27) известны как *законы p - n -перехода*. Следует помнить, что данные соотношения выведены при малом отклонении от равновесного состояния и низком уровне инжекции.

Избыточные концентрации в указанных граничных точках определяются выражениями

$$p'_n(x_n) = p_n(x_n) - p_{n0}(x_n) = p_{n0}(x_n)(e^{U/U_T} - 1) \quad (4.28)$$

$$n'_p(-x_p) = n_p(-x_p) - n_{p0}(-x_p) = n_{p0}(-x_p)(e^{U/U_T} - 1) \quad (4.29)$$

Определим теперь распределение неосновных носителей в нейтральных областях, учитывая предположения 2 и 3, в соответствии с которыми напряженность электрического поля в нейтральных областях практически равна нулю. Таким образом, исходя из формул (3.112) и (3.117) получаем уравнение диффузии дырок (по предположению, дрейфом неосновных носителей в нейтральной n -области можно пренебречь по сравнению с диффузией):

$$D_p \frac{d^2 p_n(x)}{dx^2} - \frac{p_n(x) - p_{n0}(x)}{\tau_p} = 0 \quad (4.30)$$

Так как рассматриваемый переход является резким, то $p_{n0} = const$ во всей n -области. Поэтому уравнение (4.30) можно представить в виде

$$\frac{d^2 p'_n(x)}{dx^2} - \frac{p'_n(x)}{L_p^2} = 0 \quad (4.31)$$

где $L_p = (D_p \tau_p)^{1/2}$ - диффузионная длина дырок в n -области. Общее решение дифференциального уравнения (4.31) имеет вид

$$p'_n(x) = Ae^{x/L_p} + Be^{-x/L_p} \quad (4.32)$$

Используем граничное условие, согласно которому в точке с координатой X_n , расположенной в n -области, имеет место омический контакт и поэтому избыточная концентрация неосновных носителей здесь обращается в нуль. Кроме того, учтем, что на границе обедненной области выполняется равенство (4.28). Тогда

$$p'_n(X_n) = Ae^{X_n/L_p} + Be^{-X_n/L_p} = 0 \quad (4.33)$$

$$p'_n(x_n) = Ae^{x_n/L_p} + Be^{-x_n/L_p} \quad (4.34)$$

Подставив найденные значения постоянных A и B в формулу (4.32), находим закон распределения концентрации избыточных дырок в нейтральной n -области:

$$p'_n(x) = p_{n0} \frac{sh[(X_n - x)/L_p]}{sh[(X_n - x_n)/L_p]} (e^{U/U_T} - 1) \quad (4.35)$$

Аналогично находим распределение избыточных электронов в нейтральной p -области:

$$n'_p(x) = n_{p0} \frac{sh[(X_p - x)/L_n]}{sh[(X_p - x_p)/L_n]} (e^{U/U_T} - 1) \quad (4.36)$$

Выражение (4.35) упрощается, если в формуле (4.32) считать, что $x \approx X_n \rightarrow \infty$. Действительно, если $x \rightarrow \infty$, то в (4.32) величина $A = 0$, так что носители, инжектированные в p -область, не достигают точки с координатой X_n , которая слишком удалена от перехода. Положив $x = 0$, на основании (4.32) получаем, что $p'_n(0) = B$. При этом, используя формулу (4.28), находим

$$p'_n(0) = p_{n0} (e^{U/U_T} - 1)$$

$$p'_n(x) = p'_n(0) e^{-x/L_p} \quad (4.32')$$

$$p'_n(x) = p_{n0} (e^{U/U_T} - 1) e^{-x/L_p} \quad (4.35')$$

Можно показать, что

$$n'_p(x) = n_{p0} (e^{U/U_T} - 1) e^{-x/L_n} \quad (4.36')$$

Предположив, что ток неосновных носителей в нейтральной n -области создается за счет диффузии, можно записать

$$J_p(x) = -qD_p \frac{dp_p(x)}{dx} \quad (4.37)$$

откуда, принимая во внимание формулу (4.35), получаем

$$J_p(x) = \frac{qp_{n0}(x)D_p ch[(X_n - x)/L_p]}{L_p sh[(X_n - x_n)/L_p]} (e^{U/U_T} - 1) \quad (4.38)$$

Плотность тока неосновных носителей на границе между областью перехода и n -областью определяется выражением

$$J_p(x_n) = \frac{qp_{n0}(x)D_p}{L_p th[(X_n - x_n)/L_p]} (e^{U/U_T} - 1) \quad (4.39)$$

Аналогично получаем

$$J_p(-x_p) = \frac{qn_{p0}(-x_p)D_n}{L_n th[(X_p - x_p)/L_n]} (e^{U/U_T} - 1) \quad (4.40)$$

В соответствии с предположением 4 число носителей, входящих в область перехода через одну из ее границ в единицу времени, равно числу носителей, покидающих эту область через противоположную границу. Поэтому общая плотность тока, равная сумме плотностей токов электронов и дырок, в стационарном режиме должна быть постоянной:

$$\begin{aligned} J &= J_n(x_n) + J_p(x_n) = J_n(-x_p) + J_p(-x_p) = \\ &= \left\{ \frac{qn_{p0}(-x_p)D_n}{L_n th[(X_p - x_p)/L_n]} + \frac{qp_{n0}(x_n)D_p}{L_p th[(X_n - x_n)/L_p]} \right\} (e^{U/U_T} - 1) \end{aligned} \quad (4.41)$$

Используя то, что $n_{p0} = n_i^2 / N_a$ и $p_{n0} = n_i^2 / N_d$, а также вводя обозначения $X_p - x_p = W_p$ и $X_n - x_n = W_n$, получаем

$$J = qn_i^2 \left[\frac{D_n}{N_a L_n \operatorname{th}(W_p / L_n)} + \frac{D_p}{N_d L_p \operatorname{th}(W_n / L_p)} \right] (e^{U/U_T} - 1) \quad (4.42)$$

Так как $J=I/A$, где A —площадь поперечного сечения перехода, приходим к уравнению диода, называемому также уравнением Шокли,

$$I = I_{\text{нас}} (e^{U/U_T} - 1) \quad (4.43)$$

Здесь параметр $I_{\text{нас}}$ — обратный ток насыщения диода (часто обозначаемый как I_0):

$$I_{\text{нас}} = qn_i^2 A \left[\frac{D_n}{N_a L_n \operatorname{th}(W_p / L_n)} + \frac{D_p}{N_d L_p \operatorname{th}(W_n / L_p)} \right] \quad (4.44)$$

Формула (4.43) является математической моделью диода и достаточно точно описывает его внутренние свойства при условии, что или прямой ток, или обратное напряжение не слишком велики. Даже в тех случаях, когда имеют место те или иные особые эффекты и модель не может быть применена в точности, законы перехода, устанавливаемые формулами (4.26) и (4.27), всегда служат основой для описания физических процессов в диоде.

Прежде всего необходимо отметить, что сумма плотностей токов основных и неосновных носителей в каждой точке области перехода строго задается выражением (4.41). Кроме того, следует подчеркнуть, что ток диода (4.43) включает составляющие тока основных и неосновных носителей, существующих по обе стороны области перехода; их относительный вклад определяется параметрами N_a и N_d . Составляющие тока диода, смещенного в прямом направлении, показаны на рис. 4.8.

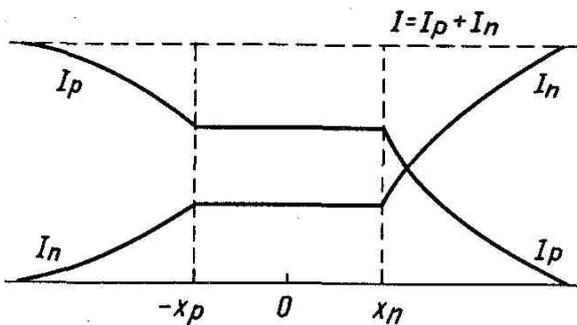


Рис. 4.8. Составляющие тока диода с p - n -переходом при прямом смещении

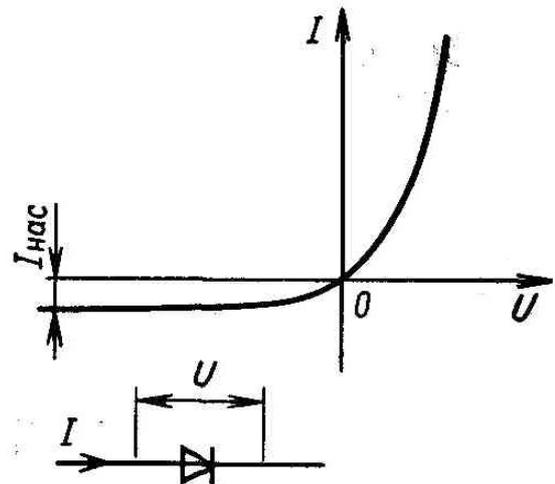


Рис. 4.9. Вольтамперная характеристика идеального диода, построенная в соответствии с уравнением (4.39). Здесь же приведено схемное изображение диода

Равенство (4.43) можно рассматривать как формулу идеального диода. Вольтамперная характеристика такого прибора изображена на рис. 4.9. Так как величина U_T весьма мала ($U_T = 26$ мВ при $T=300\text{К}$), то в случае $U > 0$ малые изменения приложенного напряжения U обуславливают значительные изменения тока I , который в данном случае описывается выражением $I = I_{\text{HAC}} \exp(U/U_T)$. Если же $U < 0$, причем $|U| \gg U_T$, то $I = -I_{\text{HAC}}$ и этот ток весьма мал, что существенно для многих применений диодов. Уменьшения I_{HAC} и соответственно n_i можно добиться, выбирая полупроводниковый материал с большой шириной запрещенной зоны, а также избегая слишком высоких температур.

Полупроводниковый диод выполняет роль выпрямителя, пропуская ток лишь в одном направлении. Ток диода может изменяться в пределах 10^{-4} - 10^2 А. При комнатной температуре ток I_{HAC} составляет несколько микроампер для германиевых диодов и несколько наноампер для кремниевых.

В заключение представляет интерес изучить температурную зависимость вольтамперной характеристики. Здесь существенными оказываются изменения параметра I_{HAC} , так как от температуры зависит величина n_i^2 . Действительно, известно, что

$$n_i^2 = N_c N_v \exp\left(-\frac{E_g}{kT}\right) = (4,83 \cdot 10^{15})^2 \left(\frac{m_p m_n}{m^2}\right)^{3/2} T^3 \exp\left(\frac{\alpha}{k}\right) \exp\left(-\frac{E_{g0}}{kT}\right)$$

или [см. формулу (3.23)]

$$n_i^2 = CT^2 \exp\left(-\frac{E_{g0}}{kT}\right) \quad (4.47)$$

Используя выражения (4.44) и (4.47), находим логарифмическую производную величины I_{HAC} по параметру T :

$$\frac{dI_{\text{HAC}}/I_{\text{HAC}}}{dT} \approx \frac{dn_i^2/n_i^2}{dT} = \frac{1}{n_i^2} \frac{dn_i^2}{dT} = \frac{3}{T} + \frac{E_{g0}}{kT^2} \quad (4.48)$$

Если на диод подано прямое напряжение, то $I \approx I_{\text{HAC}} \exp(U/U_T)$. Поэтому

$$\left(\frac{\partial I}{\partial T}\right)_{U=\text{const}} \approx \frac{I(E_{g0} - qU)}{kT^2} \quad (4.49)$$

$$\left(\frac{\partial U}{\partial T}\right)_{I=\text{const}} \approx -\frac{(E_{g0}/q) - U}{T} \quad (4.50)$$

4.3. Явления пробоя

Как уже отмечалось, ток диода при $U < 0$ в соответствии с уравнением (4.43) стремится к величине $-I_{\text{HAC}}$. Использовать формулу (4.43) при больших значениях обратного напряжения нецелесообразно, так как она перестает отражать явления, происходящие в реальных диодах. На практике ток I с увеличением отрицательного напряжения незначительно возрастает до тех пор,

пока напряжение не достигнет так называемого *напряжения пробоя* $U_{прб}$. После этого ток / возрастает скачкообразно (рис. 4.10).

Известны следующие механизмы пробоя: а) тепловая нестабильность; б) туннельный эффект, или явление Зенера; в) лавинный пробой. На практике встречаются и

смешанные случаи, так как любой из перечисленных механизмов не исключает других.

а) *Тепловая нестабильность* вызвана тем, что ток $I_{нас}$, связанный с величиной n_i^2 формулой (4.44), зависит от температуры. Выделение теплоты, обусловленное током I и напряжением U , оказывается достаточно большим и приводит к повышению температуры диода. С ростом температуры T ток увеличивается на величину ΔI и напряжение - на величину ΔU ; процесс носит лавинообразный характер и может привести к разрушению диода (см. рис. П-5.12 в приложении 5).

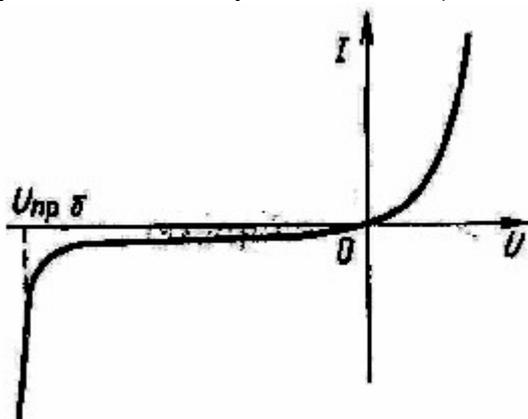


Рис. 4.10. Вольт-амперная характеристика реального диода с p - n -переходом

б) *Туннельный, или зеновский, эффект*. В силу квантово-механических свойств электрона вероятность события, состоящего в том, что электрон с энергией E преодолет потенциальный барьер шириной W и высотой $E_0 > E$, выражается формулой

$$P = \left[1 + \frac{E_0^2 \operatorname{sh}^2 \beta \cdot W}{4e(E_0 - E)} \right]^{-1} \quad (4.51)$$

где

$$\beta = \left[\frac{2m(E_0 - E)}{2} \right]^{1/2}$$

Если уровень Ферми располагается в зоне проводимости или в валентной зоне, а ширина барьера достаточно мала, существует конечная вероятность прохождения электрона сквозь потенциальный барьер. Такое явление называют туннельным эффектом.

При большой концентрации легирующей примеси ширина W значительно уменьшается, даже если обратное напряжение мало. Напряженность внутреннего электрического поля в обедненной области оказывается весьма

высокой, так что электроны способны преодолевать эту область за счет туннелирования. Возникающий при этом туннельный ток описывается выражением

$$I = \frac{A(2m_n)^{1/2} q^3 \varepsilon U}{4\pi^2 E_g^{1/2}} \exp\left(-\frac{4(2m_n)^{1/2} E_g^{1/2}}{3q \varepsilon}\right) \quad (4.52)$$

где ε - напряженность электрического поля в переходе; E - ширина запрещенной зоны; U - приложенное напряжение.

В кремнии и в арсениде галлия величина E_g уменьшается с ростом T . Поэтому наблюдается снижение пробивного напряжения, обусловленное туннельным эффектом, при увеличении температуры.

в) *Лавинный пробой* является наиболее важным, так как именно он обуславливает верхнюю границу напряжения U на диоде и предельно допустимое напряжение на коллекторе транзистора. Лавинный диод должен быть слаболегированным, так как здесь требуется, чтобы ширина обедненной области была существенно больше.

Если электрическое поле, возникающее в обедненной области при $U < 0$, достаточно велико, то электроны и дырки, пересекающие эту область, приобретают энергию, достаточную для того, чтобы выбивать электроны из атомов кристаллической решетки. При этом происходят разрыв ковалентных связей и генерация электронно-дырочных пар; возникающий ток складывается с током диода. В свою очередь, электронно-дырочные пары ускоряются электрическим полем и могут порождать новые пары на протяжении всего своего пути в данной области. Описанный механизм обладает кумулятивными свойствами, за счет чего происходит аномальный рост тока при небольшом приращении напряжения.

Коэффициент лавинного размножения M , который определяет обратный ток диода $I_{обр}$ в ходе описанного процесса, задается формулой

$$M = [1 - (|U|/U_{прб})^a]^{-1},$$

где $U_{прб}$ - напряжение пробоя; U - приложенное напряжение, не превышающее напряжение пробоя; a - величина, определяемая экспериментально и зависящая от уровней легирования p - и n -областей. При этом обратный ток диода

$$I_{обр} = I_{НАС} [1 - (|U|/U_{прб})^a]^{-1} \quad (4.53)$$

Как видно, ток $I_{обр}$ обращается в бесконечность при $U = U_{прб}$. Однако практически обратный ток ограничивается сопротивлением внешних цепей.

В диоде с резким переходом напряжения пробоя $U_{прб}$ связано с максимальной напряженностью поля

$$\varepsilon_{\max} = \left(\frac{2qN_a N_d}{\varepsilon_{ан} (N_a + N_d)} U_{прб} \right)^{1/2} \quad (4.54)$$

которая в кремнии имеет порядок $3 \cdot 10^7$ В/м. Отсюда

$$U_{прб} = \frac{\varepsilon_{an} |\varepsilon_{\max}|^2 (N_a + N_d)}{2qN_a N_d} \quad (4.55)$$

Предельное значение $U_{прб}$ составляет 500 В для германия и 1000 В для кремния.

В отличие от напряжения туннельного пробоя напряжение лавинного пробоя имеет положительный температурный коэффициент, т. е. увеличивается с ростом температуры, что было проверено экспериментально. Оба механизма пробоя могут вносить равный вклад. Тогда напряжение $U_{прб}$ почти не будет зависеть от температуры. Диоды с такими свойствами используются для создания опорных напряжений (полупроводниковые стабилитроны или диоды Зенера).

Процесс лавинного размножения носителей преобладает в тех диодах, у которых $|U_{прб}| > 8$ В. Если же величина $|U_{прб}|$ лежит в интервале 5 - 8 В, то пробой обусловлен одновременным существованием туннельного и лавинного пробоя (см. рис. П-5.13).

4.4. Малосигнальные модели диода

Перейдем к построению математической модели и эквивалентной схемы диода, соответствующих неравновесному состоянию в режиме малого сигнала. Этот случай весьма важен с точки зрения применения диодов в электрических цепях. Неравновесность системы обусловлена напряжением смещения. Предполагается, что в режиме малого сигнала колебания являются синусоидальными.

Математическая модель

Пусть к диоду приложено напряжение $U + u \exp(j\omega t)$. Здесь $U > 0$ - прямое напряжение; $u \exp(j\omega t)$ - малое гармоническое напряжение, колеблющееся около значения уровня U и называемое переменной составляющей [точнее следовало бы сказать, что $u \exp(j\omega t)$ - комплексная модель гармонического напряжения частотой ω . —Прим. пер.].

Если p -область диода легирована сильнее, чем n -область, то ток обусловлен диффузией дырок в n -область. В соответствии с формулами (3.112) и (3.117) уравнение диффузии дырок в нейтральную n -область имеет вид

$$\frac{\partial p_n}{\partial T} = -\frac{p_n'}{\tau_p} + D_p \frac{\partial^2 p_n}{\partial x^2} \quad (4.56)$$

В соответствии с выражением (4.28) концентрация избыточных носителей

$$p_n'(x_n, t) = p_{n0}(x_n) \left(e^{\frac{U + u \exp(j\omega t)}{U_T}} - 1 \right) \quad (4.57)$$

или

$$p'_n(x_n, t) = p_{n0}(x_n) \left(e^{\frac{U}{U_T}} e^{\frac{u \exp(j\omega t)}{U_T}} - 1 \right) \quad (4.58)$$

В режиме малого сигнала выполняются неравенства $u < U_T$ и $U > U_T$, поэтому единицей в формуле (4.58) можно пренебречь. Таким образом, из (4.58) следует, что

$$p'_n(x_n, t) \approx p_{n0}(x_n) e^{\frac{U}{U_T}} \left(1 + \frac{u}{U_T} e^{j\omega t} + \dots \right) \quad (4.59)$$

или

$$p'_n(x_n, t) \approx p_{n0}(x_n) e^{\frac{U}{U_T}} + \frac{u}{U_T} p_{n0}(x_n) e^{\frac{U}{U_T}} e^{j\omega t} \quad (4.60)$$

где

$$p'_{nU}(x_n) = p_{n0}(x_n) e^{\frac{U}{U_T}} \quad (4.61)$$

$$p'_{nu}(x_n) = \frac{u}{U_T} p'_{nU}(x_n) \quad (4.62)$$

Если диод легирован однородно, то выражения (4.61) и (4.62) можно подставить в (4.56). Отсюда приходим к выводу, что во всей n -области имеет место равенство

$$j\omega p'_{nu} e^{j\omega t} = -\frac{p'_{nU}}{\tau_p} - \frac{p'_{nu}}{\tau_p} e^{j\omega t} + D_p \frac{d^2 p'_{nU}}{dx^2} + D_p \frac{d^2 p'_{nu}}{dx^2} e^{j\omega t} \quad (4.63)$$

Сумма тех членов, которые содержат величину U , должна обратиться в нуль, так как при постоянном смещении $\partial p_n / \partial t = 0$ и поэтому

$$D_p \frac{d^2 p'_{nU}}{dx^2} - \frac{p'_{nU}}{\tau_p} = 0 \quad (4.64)$$

Оставшиеся члены, относящиеся к гармонической составляющей, приводят к уравнению

$$D_p \frac{d^2 p'_{nu}}{dx^2} - \frac{p'_{nu}}{\tau_p} = j\omega p'_{nu} \quad (4.65)$$

или

$$\frac{d^2 p'_{nu}}{dx^2} - \frac{p'_{nu}}{L_{pu}^2} = 0 \quad (4.66)$$

где диффузионная длина дырок в режиме малого гармонического сигнала

$$L_{pu} = \left(\frac{\tau_p D_p}{1 + j\omega \tau_p} \right)^{1/2} = \frac{L_p}{\sqrt{1 + j\omega \tau_p}} \quad (4.67)$$

Так как уравнения (4.65) и (4.64) совпадают по форме, ток в n -области можно искать в виде

$$i(x_n) = I_U + I_u e^{j\omega t} \quad (4.68)$$

или

$$i(x_n) = -qAD_p \left. \frac{dp'_{nU}}{dx} \right|_{x=x_n} - qAD_p \left. \frac{dp'_{nu}}{dx} \right|_{x=x_n} e^{j\omega t} \quad (4.69)$$

Таким же образом проводят вычисления применительно к избыточному распределению электронов в нейтральной p -области.

Выражения (4.70) - (4.72) образуют малосигнальную математическую модель диода.

Эквивалентная схема диода с p^+ - n -переходом

Для p^+ - n -диода можно ввести *проводимость на переменном токе*

$$Y = \frac{i}{u} = \frac{Aqn^2_i}{U_T} \left(\frac{D_n}{N_a L_{nu}} + \frac{D_p}{N_d L_{pu}} \right) e^{U/U_T}. \quad (4.73)$$

Чтобы упростить эквивалентную схему, ограничимся областью сравнительно низких частот, на которых выполняются неравенства $\omega\tau_p \ll 1$; $\omega\tau_n \ll 1$. При этом

$$\frac{1}{L_{pu}} = \frac{\sqrt{1+j\omega\tau_p}}{L_p} \approx \frac{1}{L_p} + \frac{j\omega\tau_p}{2L_p} \quad (4.74)$$

$$\frac{1}{L_{nu}} \approx \frac{1}{L_n} + \frac{j\omega\tau_n}{2L_n} \quad (4.75)$$

Подставляя равенства (4.74) и (4.75) в (4.73) и идентифицируя члены в соответствии с формулой

$$Y = G_{\text{диф}} + j\omega(C_{\text{диф}} + C_{\text{бар}}) \quad (4.76)$$

получаем величину

$$G_{\text{диф}} = \frac{Aqn^2_i}{U_T} \left(\frac{D_n}{N_a L_n} + \frac{D_p}{N_d L_p} \right) e^{U/U_T} \quad (4.77)$$

называемую *активной диффузионной проводимостью*, и величину

$$C_{\text{диф}} = \frac{Aqn^2_i}{2U_T} \left(\frac{D_n \tau_n}{N_a L_n} + \frac{D_p \tau_p}{N_d L_p} \right) e^{U/U_T} \quad (4.78)$$

называемую *диффузионной емкостью*. Здесь не учитывается барьерная емкость $C_{\text{бар}}$, значение которой мало по сравнению с $C_{\text{диф}}$ в режиме малого сигнала.

Легко убедиться, что формулу (4.77) можно записать в виде

$$G_{\text{диф}} = \frac{I_{\text{нас}}}{U_T} \exp\left(\frac{U}{U_T}\right) = \left. \frac{dI_U}{dU} \right|_{U=\text{const}} \quad (4.79)$$

откуда, так как $U > 0$ и $U > U_T$, получаем выражение для *дифференциального сопротивления* перехода

$$r = 1/G_{\text{диф}} \approx \tau_p I(U)/U_T = \tau_p / r \quad (4.80)$$

Данную формулу используют при описании входного сопротивления биполярного транзистора.

Далее, так как $U > U_T$, выражение (4.78) приводится к виду

$$C_{\text{диф}} \approx \tau_p I(U)/U_T = \tau_p / r \quad (4.81)$$

Величину $C_{\text{диф}}$ можно записать и по-иному:

$$C_{\text{диф}} = \frac{dQ}{dU} = \frac{dQ}{dI} \frac{dI}{dU} = \frac{W^2}{2D_p} G_{\text{диф}} \quad (4.81')$$

Диффузионная емкость пропорциональна постоянной составляющей тока. Происхождение термина связано с тем, что эта емкость обусловлена явлением диффузии.

Следует отметить, что емкость $C_{\text{диф}}$ существует лишь при $U > 0$, в то время как при $U < 0$ нужно учитывать лишь барьерную емкость $C_{\text{бар}}$.

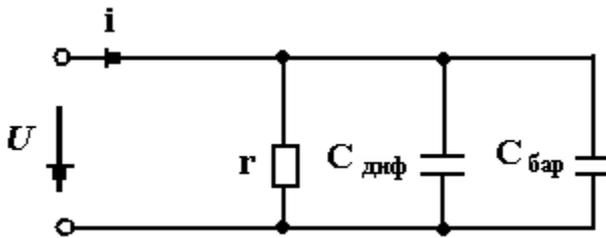


Рис. 4.11. Эквивалентная схема диода в режиме малого сигнала

Таким образом, если амплитуда приложенного гармонического напряжения удовлетворяет неравенству $u \ll U_T$, то диод можно описать эквивалентной схемой замещения в виде линейного двухполосника. Схема образована параллельным соединением дифференциального сопротивления r , диффузионной емкости $C_{\text{диф}}$ и барьерной емкости $C_{\text{бар}}$

(рис. 4.11). Если $U > 0$, то следует полагать, что $C_{\text{бар}} \ll C_{\text{диф}}$; если же $U < 0$, то выполняется обратное неравенство $C_{\text{бар}} \gg C_{\text{диф}}$.

4.5. Процессы переключения в диоде. Режим большого сигнала

Говорят, что в диоде происходит процесс переключения, если полярность приложенного к нему напряжения изменяется скачком. На практике к диоду прикладывают двуполярные импульсы напряжения. Такой режим диода принято называть *динамическим режимом* или *режимом большого сигнала*. Если приложенное напряжение изменяется от нуля до некоторого уровня $U > 0$, то происходит *включение* диода; при переходе с уровня $U > 0$ на уровень $U < 0$ наблюдается *выключение* диода.

Рассмотрим второй случай, так как он наиболее важен при работе прибора в режиме переключения.

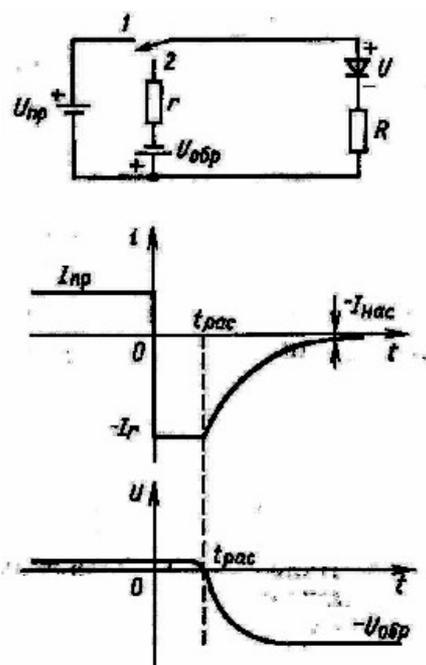


Рис. 4.12. Временные диаграммы тока и напряжения в процессе переключения диода

Явления при выключении диода

Чтобы изучить переходные процессы, наблюдаемые при выключении, предположим, что имеется диод с толстой базой, причем p -область диода легирована значительно сильнее, чем n -область (диод с p^+ - n -переходом). При этом необходимо учитывать электрический заряд лишь в n -области, который проявляется в форме избыточной концентрации дырок.

Используя уравнение непрерывности для дырок (3.117), путем интегрирования в пределах от x_n (граница раздела между обедненной областью и n -областью) до X_n (граница нейтральной n -области), получаем

$$J_p(x_n) - J_p(X_n) = q \int_{x_n}^{X_n} \left(\frac{p'_n}{\tau_p} + \frac{\partial p'_n}{\partial t} \right) dx \quad (4.82)$$

Так как диод имеет толстую базу, то $J_p(X_n) = 0$ и из (4.82) следует, что

$$i_p(x_n, t) = \frac{qA}{\tau_p} \int_{x_n}^{X_n} p'_n dx + qA \frac{\partial}{\partial t} \int_{x_n}^{X_n} p'_n dx \quad (4.83)$$

или

$$i_p(x_n, t) = \frac{Q_p(t)}{\tau_p} + \frac{dQ_p(t)}{dt} \quad (4.84)$$

где $Q_p = I\tau_p$ - избыточный заряд дырок в n -области при стационарных условиях.

Обратимся к рис. 4.12. Вначале переключатель находится в положении I в течение времени, необходимого для того, чтобы диод оказался смещенным в прямом направлении и в нем возник некоторый ток $I_{пр} = U_{пр}/R$. В

момент времени $t = 0$ переключатель переходит в положение 2 и ток скачкообразно принимает новое значение $-I_{обр} = -U_{обр}/r$

Используя в уравнении (4.84) преобразование Лапласа $Q_p(s) = L[Q_p(t)]$, получаем

$$-\frac{I_{обр}}{S} = \frac{Q_p(s)}{\tau_p} + sQ_p(s) - I_{np}\tau_p \quad (4.85)$$

где s - переменная, служащая аргументом преобразования Лапласа. Из (4.85) следует, что изображение

$$Q_p(s) = \frac{I_{np}\tau_p}{s + (1/\tau_p)} - \frac{I_{обр}}{s[s + (1/\tau_p)]} \quad (4.86)$$

Соответствующий оригинал

$$Q_p(t) = I_{np}\tau_p e^{-t/\tau_p} + I_{обр}\tau_p (e^{-t/\tau_p} - 1) \quad (4.87)$$

В рассматриваемом здесь диоде с толстой базой $p'_n(x) = p'_n(x_n) \exp(-x/L_p)$ поэтому

$$Q_p(t) \approx p'_n(x)L_p qA \quad (4.88)$$

Объединяя формулы (4.87) и (4.88), получаем

$$p'_n(x_n, t) = \frac{\tau_p}{qAL_p} [-I_{обр} + (I_{np} + I_{обр})e^{-t/\tau_p}] \quad (4.89)$$

Введем величину

$$t_{рас} = \tau_p \ln(1 + I_{np} / I_{обр}) \quad (4.90)$$

представляющую *время рассасывания* заряда. В течение этого времени избыточный заряд в точке x_n практически рассасывается. Параметр $t_{рас}$ служит мерой, ограничивающей быстродействие логических ИС, создаваемых по биполярной технологии. Для быстродействующих коммутационных диодов параметр τ_p стремятся уменьшать.

После того как накопленный заряд рассосется, диод переходит в стационарное состояние, характеризуемое напряжением $U_{обр}$ и током насыщения $I_{нас}$ (рис. 4.12). Переход совершается за время, называемое *временем восстановления*. Как время рассасывания, так и время восстановления зависят от среднего времени жизни избыточных носителей.

Аналогично можно рассмотреть диод с тонкой базой, причем методы анализа остаются прежними.

4.6. Переход металл-полупроводник

Если образцы металла и полупроводника привести в соприкосновение, то возникает некоторое движение зарядов, длящееся до тех пор, пока не установится равновесие. Контакт (переход) металл-полупроводник может быть как омическим, так и выпрямляющим. Омический контакт получается в предельном случае бесконечно большой скорости поверхностной рекомбинации.

Вольтамперная характеристика омического контакта линейна и симметрична. Контакты большинства полупроводниковых приборов являются омическими.

В выпрямляющем контакте, который называют диодом Шоттки, электропроводность обусловлена основными носителями. По этой причине диоду Шоттки свойственна высокая скорость переходных процессов, что позволяет использовать его для создания быстродействующих переключателей. К тому же здесь имеет место низкое напряжение отпираания при $U > 0$. Это обстоятельство благоприятно для работы выпрямителя большой мощности. Таким образом, диод Шоттки имеет по крайней мере два положительных отличия от диода на $p-n$ -переходе, у которого, во-первых, меньше скорость переходных процессов, так как ток создается движением неосновных носителей, и, во-вторых, выше напряжение отпираания (примерно 0,6 В для кремниевого диода с $p-n$ -переходом и 0,3 В для диода Шоттки).

При анализе перехода металл - полупроводник следует учитывать два фактора: разницу в работе выхода обоих материалов и наличие примесей и поверхностных дефектов полупроводника в области контакта. К счастью, поверхностные эффекты, которые осложняют изготовление высококачественных омических контактов, можно свести к нулю, выполнив

операцию пассивирования в процессе производства интегральных схем.

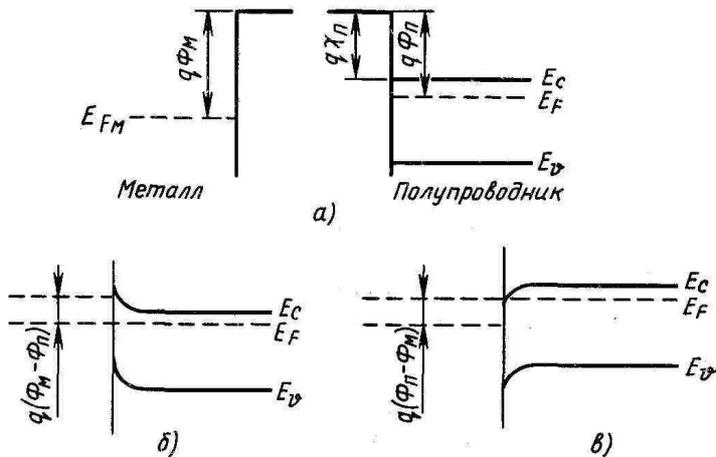


Рис. 4.13. Контакт металл-полупроводник n -типа:

a - при отсутствии смещения;

b - случай выпрямляющего контакта, ;

$в$ —случай омического контакта,

Характеристики перехода металл-полупроводник можно рассчитать пренебрегая влиянием поверхностных уровней. На рис. 4.13, a изображены диаграммы энергетических зон материалов до их контактного соединения. Диаграммы рис. 4.13, b , $в$ соответствуют случаям, когда материалы контактируют. На этих диаграммах верхний уровень, или уровень вакуума, является отсчетным и определяет энергию, которую

следует сообщить «покоящемуся» электрону для того, чтобы он покинул материал. Работа выхода металла $q\Phi_M$ и работа выхода полупроводника $q\Phi_P$ - это разности между энергией уровня вакуума и энергетическими уровнями материалов. Энергия электронного сродства $q\chi_P$ полупроводника представляет механическую энергию электрона, переходящего со дна зоны проводимости E_c на уровень вакуума. Для любого полупроводника величина $q\chi_P$ в отличие от величины $q\Phi_P$ не зависит от степени легирования.

Случай, изображенный на рис. 4.13,б, соответствует равновесному состоянию контакта металла с полупроводником n -типа при $q\Phi_M > q\Phi_{II}$. Можно отметить наличие барьера, препятствующего переходу электронов из полупроводника в металл и из металла в полупроводник. По этой причине данный переход можно рассматривать как выпрямляющий элемент. Случай, изображенный на рис. 4.13,в, соответствует неравенству $q\Phi_M < q\Phi_{II}$. Здесь электроны легко пересекают контакт, который ведет себя как омический. Такие же рассуждения справедливы и в случае полупроводника p -типа.

Диоды Шотки

Будем считать, что выполнены условия, соответствующие случаю, изображенному на рис. 4.13,б. Здесь потенциальный барьер возникает без внешнего напряжения. Если теперь к контакту приложить некоторое напряжение U , то высота барьера уменьшится до $U_0 - U$, в результате чего электронный ток возрастет подобно току в вакуумном диоде. Формула, описывающая вольтамперную характеристику, имеет вид

$$I = AR^*T^2 \exp\left(\frac{\Phi_{бар}}{U_T}\right) \left[\exp\left(\frac{U}{U_T}\right) - 1 \right] \quad (4.91)$$

где A - площадь поперечного сечения контакта; R^* - постоянная Ричардсона; $\Phi_{бар}$ - истинная высота потенциального барьера, которую вычисляют по экспериментально найденной формуле

$$\Phi_{бар} = 0,235\Phi_M - 0,352 \quad (4.92)$$

Внешне уравнение (4.91) весьма напоминает уравнение диода с p - n -переходом, смещенным в прямом направлении,

$$I = I_{нас} \left[\exp\left(\frac{U}{U_T}\right) - 1 \right]$$

откуда нетрудно определить ток $I_{нас}$.

Однако с точки зрения внутренних механизмов работы диод Шотки и диод с p - n -переходом существенно различны. В последнем используется явление диффузии неосновных носителей; напряжение смещения U определяет характер распределения этих носителей в обеих областях диода. Действие же диода Шотки основано на том, что напряжение U управляет потоком основных носителей, преодолевающих потенциальный барьер. Этим обусловлены достоинства диода Шотки.

4.7. Переход полупроводник-полупроводник (гетеропереход)

Можно получить некоторый переход между двумя полупроводниками, ширина запрещенных зон которых различна, т.е. $E_{g1} \neq E_{g2}$. На энергетической диаграмме гетероперехода (рис. 4.14) возникают разрывы. Поэтому исследовать его свой-

ства значительно труднее, чем характеристики перехода, выполненного из одного и того же материала.

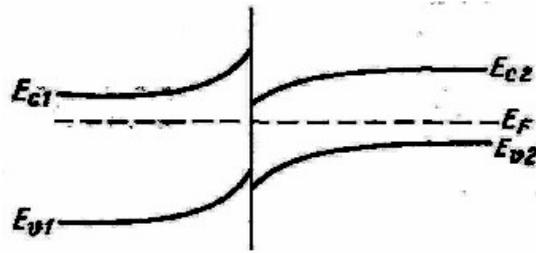


Рис. 4.14. Диаграмма энергетических уровней перехода полупроводник - полупроводник (гетеропереход). Между краями зон проводимости и валентных зон полупроводников 1 и 2 возникают разрывы, если $E_g^1 > E_g^2$

Один из способов обеспечить регулярность кристаллической структуры в области контакта двух полупроводников состоит в подборе таких веществ, межатомные расстояния которых одинаковы. Другой способ состоит в нанесении эпитаксиальной пленки одного вещества на поверхность кристалла другого. Примерами могут служить гетеропереходы Ge-GeAs и GaAs – GaAs_xP_{1-x}, применяемые в лазерах.

В лазерах и светоизлучающих диодах на основе гетеропереходов носители инжектируются в область с меньшим значением E_g . В результате те фотоны, которым удастся достигнуть области с большим значением E_g , излучаются в свободное пространство; энергия их примерно равна E_g .

Наиболее ценное свойство гетероперехода, к которому приложено некоторое напряжение $U > 0$, состоит в том, что здесь при соединении двух различных полупроводников для электронов и для дырок возникают потенциальные барьеры различной высоты.

4.8. Диоды для оптоэлектроники

Рассмотрим кратко некоторые диоды, играющие все более важную роль в современной технике,

- солнечные элементы, преобразующие энергию солнечного света в электрическую энергию, светоизлучающие диоды, на основе которых создают устройства отображения информации, и, наконец, полупроводниковые лазеры с *p-n*-переходом или с гетеропереходом, используемые для оптической связи на большие расстояния. Область применения таких лазеров весьма широка и включает, например, биоинженерию и голографию.

Солнечный элемент

Данный прибор представляет диод с *p-n*-переходом большой площади, способный преобразовывать энергию падающего солнечного излучения в энергию электрического тока с КПД, достигающим 20%. Получать энергию из угля, нефти или урана в настоящее время примерно в 50 раз дешевле, чем из

солнечного излучения при таком КПД. Однако следует ожидать, что в ближайшие десятилетия эта разница стоимостей будет сокращаться.

Чаще всего для изготовления солнечных элементов используют кремний, а также селен, арсенид галлия GaAs и сернистый индий InS. На рис. 4.15,а показана конструкция солнечного элемента, на рис. 4.15,б - соответствующая энергетическая диаграмма, на которой обозначена величина E_g .

Если солнечный свет с энергией кванта, равной или превышающей E_g , падает на n -слой p - n -перехода, то в обедненной области возникают электронно-дырочные пары. Существующее в этой области внутреннее электрическое поле заставляет электроны двигаться по направлению к n -области, а дырки—к p -области. Если имеется цепь нагрузки, то в ней под действием света возникает некоторый ток $I_{св}$.

Будем считать, что при отсутствии освещения солнечный элемент имеет вольтамперную характеристику идеального диода (рис. 4.15,в):

$$I = I_{НАС} [e^{U/U_T} - 1] \quad (4.93)$$

где $I_{НАС}$ - обратный ток насыщения, обусловленный тепловым возбуждением; для солнечного элемента

$$I_{НАС} = AqN_cN_v \left(\frac{\sqrt{D_n/\tau_n}}{N_a} + \frac{\sqrt{D_p/\tau_p}}{N_d} \right) e^{-E_g/kT} \quad (4.94)$$

где A - площадь элемента.

Если теперь подать освещение, то

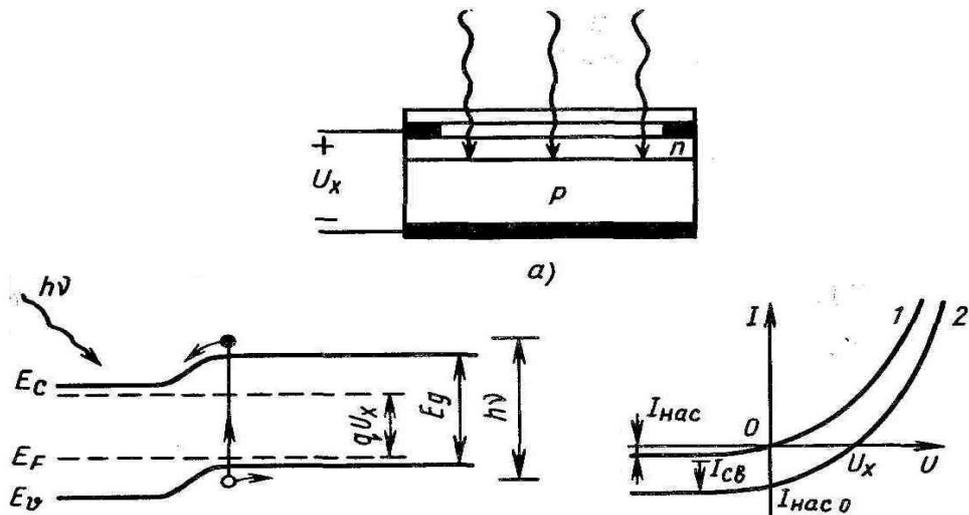


Рис. 4.15. Солнечный элемент: а - конструкция прибора (сверху вниз - стеклянная пластина, через которую проходит солнечный свет, собирающий электрод, весьма тонкая n -область, p -область, собирающий электрод), б - упрощенная диаграмма энергетических зон, в - вольтамперная характеристика (кривая 1 - темновой режим, кривая 2 - режим освещения)

$$I = I_{НАС} [e^{U/U_T} - 1] - I_{св} \quad (4.95)$$

причем знак тока $I_{cс}$ взят отрицательным, так как он создан неосновными носителями и направлен в противоположную сторону. На рис. 4.15,в показана вольтамперная характеристика солнечного элемента, полученная из характеристики в темновом режиме вычитанием величины $I_{cс}$.

Из формулы (4.95) можно найти напряжение, которое возникает в разомкнутой цепи (режим холостого хода) при $I=0$:

$$U_x = U_T \ln\left(\frac{I_{cс}}{I_{HAC}} + 1\right) \approx U_T \ln \frac{I_{cс}}{I_{HAC}} \quad (4.96)$$

Мощность на выходе

$$P = IU = I_{HAC} U [e^{U/U_T} - 1] - I_{cс} U \quad (4.97)$$

Эта мощность достигает максимального значения P_{max} в режиме, при котором имеет место равенство $dP/dU=0$. Отсюда следует, что соответствующее напряжение U_{max} должно удовлетворять уравнению

$$\frac{d}{dU} \{I_{HAC} U [e^{U/U_T} - 1] - I_{cс} U\} = 0$$

откуда

$$U_{max} = U_T \ln\left(\frac{I_{cс}/I_{HAC} + 1}{1 + U_{max}/U_T}\right) \approx U_x - U_T \ln(1 + U_{max}/U_T) \quad (4.98)$$

$$I_{max} = I_{HAC} \frac{U_{max}}{U_T} \exp\left(\frac{U_{max}}{U_T}\right) \approx I_{cс} \left(1 - \frac{U_T}{U_{max}}\right) \quad (4.99)$$

Следовательно,

$$P_{max} = I_{max} U_{max} = I_{HAC} \frac{U_{max}^2}{U_T} \exp\left(\frac{U_{max}}{U_T}\right) = \frac{I_{HAC} U_{max}^2}{U_T (1 + U_{max}/U_T)} \left(1 + \frac{I_{cс}}{I_{HAC}}\right) \quad (4.100)$$

или

$$P_{max} = I_{cс} \left[U_x - U_T \ln\left(1 + \frac{U_{max}}{U_T}\right) - U_T \right] \quad (4.101)$$

Данную формулу можно записать в виде

$$P_{max} = \frac{I_{cс}}{q} E_{max} \quad (4.102)$$

где

$$E_{max} = q \left[U_x - U_T \ln\left(1 + \frac{U_{max}}{U_T}\right) - U_T \right]$$

- энергия фотона, которая передается носителю заряда в режиме максимальной выходной мощности.

Идеальный КПД солнечного элемента (%)

$$\eta = \frac{P_{max}}{P_{пад}} \cdot 100 \quad (4.103)$$

или

$$\eta = \frac{I_{св} (E_{\max} / q)}{P_{\text{над}}} \cdot 100 \quad (4.104)$$

Так называемый коэффициент насыщения вольтамперной характеристики определяется формулой

$$K_{\text{НАС}} = I_{\max} U_{\max} / (I_{\text{НАС}} U_x) \quad (4.105)$$

Светоизлучающий диод

Светоизлучающий диод представляет прибор с *p-n*-переходом, который способен излучать видимый свет при подаче прямого напряжения 1,5 - 3 В. Существующие диоды излучают красный, зеленый, желтый, оранжевый, синий и белый свет. Наиболее широкое применение эти диоды находят, пожалуй, при создании устройств отображения цифр и знаков в микрокалькуляторах, а также для световых табло, управляемых с помощью ЭВМ.

Основным физическим процессом в светоизлучающем диоде является излучение фотонов при рекомбинации электронов или дырок, которые инжектируются при прямом смещении. Большинство актов рекомбинации носителей происходит на участке, длина которого равна удвоенной диффузионной длине перехода; именно здесь и происходит излучение света.

Одним из возможных механизмов служит прямой переход из зоны проводимости в валентную зону. Инжекция электронов в *p*-область приводит к излучению фотонов с энергией E_g . Могут иметь место и другие переходы на уровни, расположенные в запрещенной зоне. Используются полупроводники с прямыми или близкими прямым зонами, например GaAs, GaP и GaP_{1-x}As_x. В таких материалах существует большое число различных квантовых переходов, при которых излучается свет.

Составляющие диффузионного тока, зависящие от уровня легирования и от приложенного положительного напряжения, можно описать с помощью уравнения идеального диода:

- для электронов

$$I_n = \frac{qD_n n_i^2}{L_n N_a} \left[\exp\left(\frac{U}{U_T}\right) - 1 \right] \quad (4.106)$$

- для дырок

$$I_p = \frac{qD_p n_i^2}{L_p N_d} \left[\exp\left(\frac{U}{U_T}\right) - 1 \right] \quad (4.107)$$

КПД (%) процесса излучения фотонов в области *p*-типа определяется формулой

$$\eta = \frac{V_f}{V_f + V_{nf}} \cdot 100 \quad (4.108)$$

Здесь V_f - скорость рекомбинации электронов, приводящей к излучению фотонов (см. формулу (3.83)]; V_{nf} - скорость процесса рекомбинации, не сопровождающегося излучением. На основании выражения (3.83) получаем

$$\eta = \frac{1}{1 + (\tau_f / \tau_{nf})} \cdot 100 = \frac{\tau}{\tau_f} \cdot 100 \quad (4.109)$$

где τ - время жизни, вычисляемое по формуле

$$1/\tau = 1/\tau_f + 1/\tau_{nf} \quad (4.110)$$

Считая, что приближенно выполняется условие равновесия $n + p = N_a + N_a / n_i^2 \approx N_a$, применяя формулу (3.80) для вычисления мела излученных фотонов и находя величину τ_{nf} из формулы 3.84), получаем

$$\eta = \frac{1}{1 + GN_i / (rN_a)} \quad (4.111)$$

Лазеры с p - n -переходами и гетеропереходами

Инжекция электронов или дырок в область перехода приводит к спонтанной рекомбинации и вызывает излучение некогерентного света. В лазере на p - n -переходе за счет инжекции носителей возникает инверсия населенности уровней. Свет, излучаемый лазером, является когерентным (аббревиатура «лазер» происходит от - усиление света за счет вынужденного излучения).

Лазер с p - n -переходом отличается от других твердотельных (например, рубиновых), газовых (He-Ne, CO₂) и жидкостных лазеров. Он имеет весьма малые размеры, около 0,1×0,1×0,3 мм, и высокий электрооптический КПД, превышающий 50%. Выходной пучок света такого лазера легко модулировать, изменяя ток через переход. Полупроводниковый лазер имеет широкое применение, особенно важным является использование его в качестве источника света для волоконно-оптических линий связи.

Как известно, переход электрона с нижнего энергетического уровня на более высокий совершается за счет поглощения фотона. Обратный переход электрона с более высокого на более низкий уровень сопровождается либо спонтанным, либо вынужденным испусканием фотона. В рассматриваемом лазере имеют место оба процесса, однако второй является преобладающим. За счет инверсии населенности уровней число фотонов, участвующих в процессе вынужденного излучения, превышает число поглощенных фотонов. Кроме того, здесь используется явление удержания фотонов, в результате плотность потока фотонов вынужденного излучения превышает плотность потока спонтанного излучения (для удержания фотонов применяют оптический резонатор Фабри - Перо, образованный плоскопараллельными отполированными гранями кристалла).

Чтобы создать режим инверсии населенности, в полупроводниковом лазере используют p - и n -области с высоким уровнем легирования (вырожденный полупроводник). Кроме того, обеспечивают условие $qU > E_g$. Как следствие, число электронов вблизи дна зоны проводимости оказывается большим, чем вблизи потолка валентной зоны. Для этого требуется, чтобы рекомбинация электронно-дырочных пар происходила непосредственно.

Поэтому здесь нельзя использовать такие полупроводники, как кремний и германий.

Процесс инверсии населенности в рассматриваемой структуре можно адекватно описать с помощью квазиуровней Ферми. Концентрация электронов, инжектированных в слой с инверсной населенностью, а также различные диффузионные длины для p -области оказываются большими, чем при равновесии. Это же относится к дыркам, инжектированным в n -область. Однако уравнения, которые описывают концентрации носителей в стационарном состоянии, совпадают с уравнениями (3.99) и (3.100):

$$n = n_i \exp\left[\frac{E_{Fn} - E_i}{kT}\right] = N_c \exp\left[-\frac{E_c - E_{Fn}}{kT}\right] \quad (4.112)$$

$$p = n_i \exp\left[\frac{E_i - E_{Fp}}{kT}\right] = N_v \exp\left[-\frac{E_{Fp} - E_v}{kT}\right] \quad (4.113)$$

Отсюда можно получить энергетические диаграммы и вычислить соответствующие распределения концентрации носителей.

На рис. 4.16 представлена энергетическая диаграмма классического полупроводникового лазера на арсениде галлия GaAs, излучающего в инфракрасном диапазоне. Другой классический диодный лазер на GaAs создает излучение видимого диапазона.

Покажем, что в лазере на арсениде галлия напряжение $U > 0$ должно выбираться таким, чтобы выполнялось неравенство

$$qU > h\nu, \quad (4.114)$$

где ν - частота инфракрасного излучения.

Вероятность того, что электрон занимает энергетические уровни E_c и E_v , определяется функцией распределения Ферми - Дирака (3.10). Для случая, показанного на рис. 4.16,б,

$$f(E_c) = \frac{1}{1 + \exp[(E_c - E_{Fn})/(kT)]} \quad (4.115)$$

$$f(E_v) = \frac{1}{1 + \exp[(E_v - E_{Fp})/(kT)]} \quad (4.116)$$

причем

$$E_{Fn} - E_{Fp} = qU \quad (4.117)$$

$$E_c - E_v = h\nu \quad (4.118)$$

Если $N_{изл}$ - плотность потока инфракрасного излучения частотой ν , то

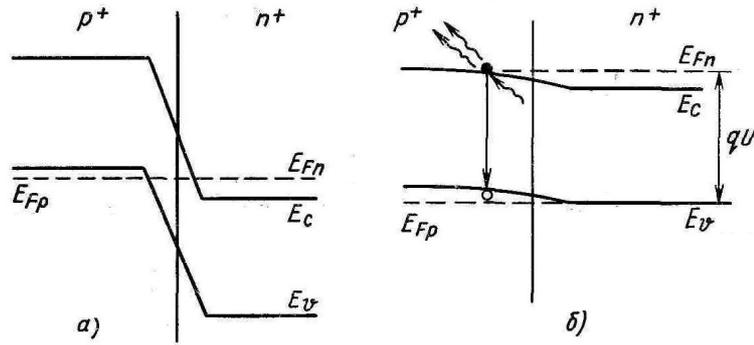


Рис. 4.16. Диаграмма энергетических уровней в полупроводниковом лазере:
 а - при отсутствии напряжения смещения;
 б - при прямом напряжении ($U > 0$), достаточном для создания инверсии населенности в p^+ -области вблизи перехода

число фотонов $N_{ног}$, поглощенных в единицу времени, удовлетворяет уравнению

$$\frac{dN_{ног}}{dt} = C_1 P_{cv} f(E_v) [1 - f(E_c)] N_{изл} \quad (4.119)$$

где $1 - f(E_c)$ - вероятность того, что энергетический уровень E_c остается незанятым; P_{cv} - вероятность перехода.

Аналогично, если $N_{вын}$ - плотность потока вынужденного инфракрасного излучения, то число таких фотонов, возникающих в единицу времени, удовлетворяет уравнению

$$\frac{dN_{вын}}{dt} = C_2 P_{cv} f(E_c) [1 - f(E_v)] N_{изл} \quad (4.120)$$

Так как числовые значения констант и вероятностей переходов равны, то для выполнения неравенства

$$f(E_c) [1 - f(E_v)] > f(E_v) [1 - f(E_c)] \quad (4.121)$$

или, что эквивалентно,

$$f(E_c) > f(E_v) \quad (4.122)$$

необходимо, чтобы

$$dE_{вын} / dt > dE_{ног} / dt \quad (4.123)$$

или

$$E_{Fn} - E_{Fp} > E_c - E_v \quad (4.124)$$

Отсюда, принимая во внимание формулы (4.117) и (4.118), приходим к неравенству (4.114).

В настоящее время полупроводниковый лазер на основе GaAs уступил место лазеру на гетеропереходе, образованном $Ga_{1-x}Al_xAs$ и GaAs. Этот лазер хорошо работает при комнатной температуре и удобен для применения в оптических линиях связи. В лазере на гетеропереходе инжектированные

носители сосредотачиваются в узкой области. По этой причине состояние инверсии населенности наступает при более низких значениях тока, чем в лазере на $p-n$ -переходе. Созданы лазеры на гетеропереходе с длиной волны $\lambda = 0,85$ мкм и пороговым током от 10 до 25 мА.

Существует несколько типов лазеров на гетеропереходе, среди которых следует отметить приборы с простым (SH) и двойным (DH) гетеропереходом. Первый имеет одиночный $p-n$ -переход, размещенный на расстоянии 2 мкм под гетеропереходом GaAs - $Al_xGa_{1-x}As$, и характеризуется сравнительно невысоким КПД. Лазер второго типа имеет более высокий КПД и более низкий пороговый ток.

Можно ожидать, что лазеры с гетеропереходами послужат основой интегральной оптики будущего, когда будут созданы передатчики светового диапазона, размещенные на едином кристалле.

5. ФИЗИЧЕСКИЕ ПРИНЦИПЫ РАБОТЫ ПОЛЕВЫХ ТРАНЗИСТОРОВ

Во многих современных электронных цепях используют транзисторы, ток носителей которых течет по так называемому каналу, образованному внутри кремниевого кристалла. Этим током можно управлять, прикладывая электрическое поле. Такие приборы называют полевыми транзисторами. В настоящее время эти транзисторы играют весьма важную роль, являясь элементами ИС которые содержат на одном кристалле от тысячи до сотен тысяч полупроводниковых приборов. В свою очередь, на базе таких ИС создают компьютеры, микропроцессорные системы, устройства обработки сигналов и т.д.

Наиболее распространенным видом полевого транзистора являются так называемые МОП-транзисторы. Их работа основана на использовании очень важной структуры, называемой МОП-конденсатором. Принцип действия полевого транзистора основан на дрейфовом движении основных носителей вдоль канала (в приборах с каналом n -типа перемещаются электроны, а в приборах с каналом p -типа - дырки). В отличие от полевого работа биполярного транзистора связана с диффузией неосновных носителей через базовую область.

Существуют три группы полевого транзисторов: типа МОП (металл-оксид-полупроводник), с управляющим $p-n$ -переходом, с управляющим переходом металл-полупроводник.

Изучение транзисторов начнем именно с полевого прибора, поскольку по структуре они проще биполярных транзисторов и легче объяснить принцип их работы, к тому же полевым транзисторам присущ ряд ценных качеств:

- производство этих приборов проще, они имеют меньшие габариты и можно добиться более высокой степени интеграции ИС;
- потребляемая ими мощность меньше чем, чем у биполярных транзисторов (мощность потребляемая МОП-транзисторами составляет единицы нановатт в то время как биполярные транзисторы потребляют единицы милливатт);

- применение полевых транзисторов улучшает экономические показатели изделий;
- входное сопротивление транзисторов весьма велико (свыше 10 МОм); на базе полевых транзисторов легко создавать запоминающие устройства;
- на базе полевых транзисторов легко создавать запоминающиеся устройства, работающие за счет накопления зарядов малыми внутренними емкостями; надежность полевых транзисторов выше надежности биполярных.

В данной главе качественно и количественно рассматриваются МОП-конденсатор, МОП-транзистор, а также полевой транзистор с управляющим $p-n$ -переходом и полевой транзистор с переходом металл-полупроводник.

5.1. Введение в теорию идеального МОП-конденсатора

Физической структурой, особенно важной при изучении поверхностных явлений в полупроводниках, является МОП-конденсатор. На базе анализа этой структуры удастся построить теорию, пригодную для описания дискретных полупровод-пиковых приборов и ИС, выполненных по МОП-технологии.

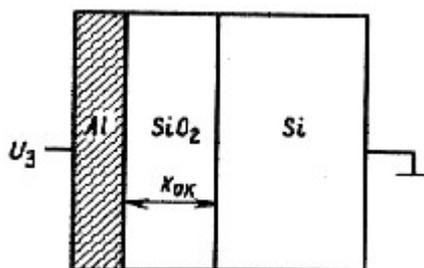


Рис 5.1. Поперечный разрез
МОП-конденсатора

Конструкция МОП-конденсатора изображена на рис. 5.1. Здесь имеется подложка из монокристаллического кремния, легированного атомами бора В или фосфора Р с концентрацией от 10^{14} до 10^{16} $см^{-3}$. На поверхности подложки методом высокотемпературного оксидирования создан слой оксида SiO_2 толщиной 40-100 нм. Поверхность оксида покрыта слоем металла толщиной 0,5 - 1 мкм, называемым *затвором*. Напряжение, приложенное между затвором и подложкой, определяет вид энергетических зон на границе раздела SiO_2 - Si и, следовательно, электрические свойства полупроводника в области существования поля.

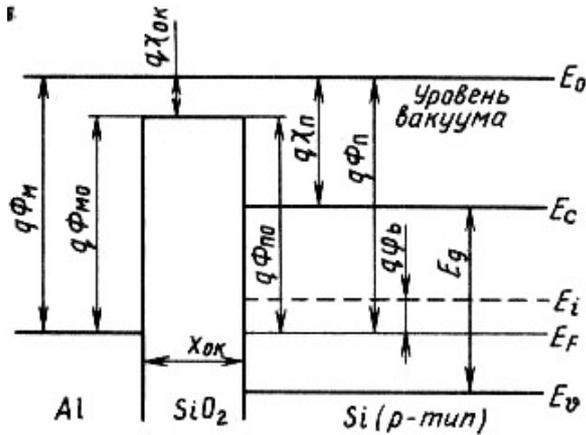


Рис. 5.2. Энергетическая диаграмма идеального МОП-конденсатора

- 3) оксидный слой однороден и не содержит каких-либо зарядов, т. е. представляет идеальный изолятор;
- 4) как алюминий, так и кремний имеют одинаковую работу выхода, и поэтому их уровни Ферми совпадают.

Так как напряжение $U_3 = 0$, т. е. контактная разность потенциалов между металлом и полупроводником $\Phi_{mn} = 0$, то энергетические зоны отображаются прямыми линиями. В соответствии с предположением 3) уровень Ферми постоянен и полупроводник при $U_3 = 0$ находится в состоянии термодинамического равновесия, т. е. $pn = n_i^2$ и ток между металлом и полупроводником отсутствует. На рис. 5.2 изображена энергетическая диаграмма, построенная для случая $U_3 = 0$ в соответствии с предположениями 2) - 4).

Можно видеть, что энергетические уровни - действительно прямые линии. Рассмотрим параметры, указанные на рис. 5.2:

$q\Phi_m$ - работа выхода из алюминия (энергия, необходимая для отрыва электрона от металла и перемещения в бесконечность);

$q\Phi_{mo} = 3,2\text{эВ}$ - определенная экспериментально разность работ выхода из алюминия и оксида кремния;

$q\Phi_{no} = 3,25\text{эВ}$ - определенная экспериментальная разность работ выхода из кремния и оксида кремния; $q\Phi_n$ - работа выхода из кремния;

$q\Phi_b = qU_F$ - разность между собственной энергией Ферми E_i и энергией Ферми E_F , относящейся к рассматриваемому объему полупроводника. Здесь U_F - высота потенциального барьера, или напряжение Ферми: $U_F = U_T \ln(Na/n_i)$, $U_T = kT/q$;

$x_{ок}$ - толщина слоя оксида кремния; qx_n - энергия электронного средства кремния (энергия, которую нужно сообщить электрону для того, чтобы он перешел в вакуум со дна зоны проводимости кремния); $qx_{ок}$ - энергия электронного средства оксида кремния.

Будем считать, что затвор выполнен из алюминия Al, а материалом подложки, неограниченно протяженной вглубь, служит кремний p-типа. Подложка равномерно легирована атомами бора, причем концентрация акцепторной примеси равна N_a . Для упрощения будем также считать, что:

- 1) структура является одномерной;
- 2) свойства материалов на поверхности и в объеме неизменны;

Наиболее важное следствие идеальности диэлектрических свойств SiO_2 состоит в том, что напряжение U_3 приложенное к затвору, не вызывает тока между металлом и полупроводником, а лишь приводит к появлению электрического поля и поверхностных зарядов на границах алюминия и кремния.

Режимы обогащения, обеднения и инверсии

В зависимости от знака напряжения на затворе МОП-конденсатора можно выделить три режима работы, отличающиеся характером явлений на поверхности полупроводника: а) обогащения; б) обеднения; в) инверсии.

Режим обогащения ($U_3 < 0$)

Если $U_3 < 0$, то возникает электрическое поле, направленное от полупроводника к затвору. Это поле смещает в кремнии основные носители (дырки) по направлению к границе раздела SiO_2 — Si. В результате на границе возникает обогащенный слой толщиной 10—100 нм с избыточной концентрацией дырок.

Это явление можно описать и по-другому. При $U_3 < 0$ потенциал на поверхности кремния уменьшается, в результате чего зона проводимости, валентная зона, а также уровень E_i изгибаются вверх.

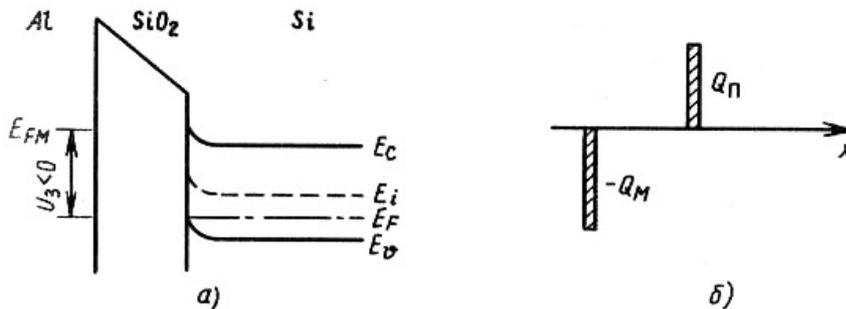


Рис. 5.3. Режим обогащения в подложке р-типа:
а - диаграмма энергетических уровней;
б - распределение заряда;

Как следствие этого, валентная зона приближается к уровню Ферми (рис. 5.3, а) и концентрация дырок в поверхностном слое становится большей, чем в объеме кремния. На границах раздела SiO_2 - Si и Al - Si возникают поверхностные заряды, одинаковые по модулю, с поверхностной плотностью $+Q_s$ на первой границе и $-Q_m$ на второй (рис. 5.3,б).

Электрическое поле в области между подложкой и затвором является однородным и описывается решением уравнения Пуассона

$$\frac{d^2U}{dx^2} = \frac{d\varepsilon}{dx} = \frac{-\rho}{\varepsilon_{\text{аок}}} \quad (5.1)$$

$$\varepsilon = |Q_m| / \varepsilon_{\text{аок}} \quad (5.2)$$

где $\varepsilon_{aок} = \varepsilon_{ок} \cdot \varepsilon_o$ - абсолютная диэлектрическая проницаемость оксида.

Интегрируя уравнение (5.1), получаем аналитическое выражение, описывающее распределение потенциала U в пренебрежении падением напряжения в кремнии:

$$dU = \varepsilon dx = -(U_3 / x_{ок}) dx \quad (5.3)$$

$$U = -\frac{U_3}{x_{ок}} \int_{x_{jr}}^x dx = -\frac{U_3}{x_{ок}} (x - x_{ок}) = U_3 \left(1 - \frac{x}{x_{ок}}\right)$$

Далее, обозначив символом $U_{ок}$ падение напряжения в оксидном слое, которое в режиме обогащения совпадает с напряжением U_3 , можно, используя формулу (5.2), записать:

$$U_{ок} = |\varepsilon| \cdot x_{ок} = Q_m \cdot x_{ок} / \varepsilon_{aок} \quad (5.4)$$

или

$$Q_m / U_3 = \varepsilon_{aок} / x_{ок} \quad (5.5)$$

Это выражение определяет емкость МОП-конденсатора с единичной площадью обкладок:

$$C_o = Q_m / U_3 \quad (5.6)$$

т.е.

$$C_o = \varepsilon_{aок} / x_{ок} \quad (5.7)$$

Режим обеднения ($U_3 > 0$)

Если напряжение $U_3 > 0$, то возникает электрическое поле, вектор напряженности которого направлен от затвора к кремниевой подложке. Это поле выталкивает дырки с границы раздела $\text{SiO}_2 - \text{Si}$ в глубь кристалла кремния. Как следствие этого, в непосредственной близости от границы возникает зона, обедненная дырками. С другой стороны, известно, что

$$p = n_i \exp\left(\frac{E_i - E_F}{kT}\right), \quad (5.8)$$

где параметр $n_i = 1,45 \cdot 10^{10} \text{ см}^{-3}$ при комнатной температуре. Поэтому если уменьшить значение p , то одновременно уменьшается разность $E_i - E_F$, так что энергетические уровни E_c , E_v и E_i в окрестности границы раздела $\text{SiO}_2 - \text{Si}$ изгибаются вниз (рис. 5.4,а)

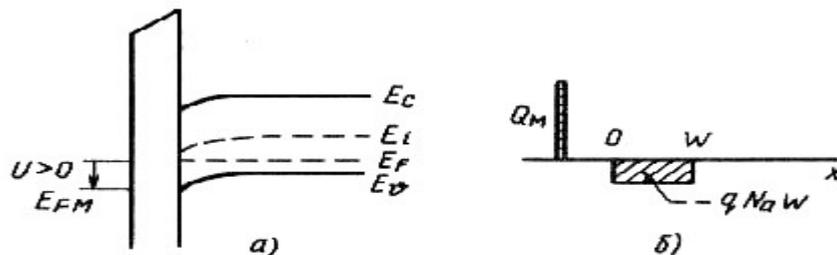


Рис. 5.4. Режим обеднения:

а - диаграмма энергетических уровней; б- распределение заряда

Если считать, что имеет место полное обеднение, то поверхностная плотность заряда выражается так же, как в p - n -переходе:

$$Q_S = Q_B = -qN_a \cdot W \quad (5.9)$$

Здесь N_a - концентрация акцепторной примеси, W - толщина обедненной области, Q_B - заряд, содержащийся в обедненной области и отнесенный к единице площади.

Напряженность электрического поля находят с помощью выражений (5.1) и (5.9). Максимальная напряженность поля

$$\varepsilon_{\max} = \varepsilon_S = q \cdot N_a \cdot W / \varepsilon_a \quad (5.10)$$

Поскольку значение W достаточно велико, приложенное напряжение (U_3 складывается из напряжения U_s (поверхностный потенциал границы раздела SiO_2 — Si) и напряжения $U_{ок}$ (напряжение на оксидном слое):

$$U_3 = U_{ок} + U_s. \quad (5.11)$$

Здесь величина $U_{ок}$ задается формулой (5.4), а напряжение, определяется из выражения (5.10), согласно которому:

$$U_s = \frac{\varepsilon_{\max} W}{2} = \frac{qN_a W^2}{2\varepsilon_a} \quad (5.12)$$

Решив уравнение Пуассона, дополненное граничными условиями $\varepsilon(W) = 0$ и $U(W) = 0$, определим пространственные распределения поля и потенциала:

$$\varepsilon = \frac{qN_a}{\varepsilon_a} \cdot \int_x^W dx = \frac{qN_a (w - x)}{\varepsilon_a} \quad (5.13)$$

$$U = \frac{qN_a}{\varepsilon_a} \cdot \int_x^W dx = \frac{qN_a}{\varepsilon_a} (W^2 + x^2 - 2Wx) = \frac{qN_a W^2}{2\varepsilon_a} \cdot \left(1 - \frac{x}{W}\right) = U_s \cdot \left(1 - \frac{x}{W}\right)^2 \quad (5.14)$$

Полученные выражения подобны тем, которые были найдены ранее применительно к резкому p - n -переходу с невысоким уровнем легирования p -области.

Режим инверсии ($U_3 \gg 0$).

Если к затвору приложить достаточно большое положительное напряжение, то это приведет к увеличению параметра W и в соответствии с формулой (5.12) к росту поверхностного электрического потенциала U_s . Данное явление является следствием того, что энергетические уровни сильно изгибаются вниз. Изгиб оказывается таким, что в малой окрестности Я границы раздела SiO_2 - Si уровень E_F в кремнии падает 1 ниже уровня E_F (рис. 5.5, а). При этом поверхностный потенциал U_s превышает потенциал U_F в кремнии. Изгиб энергетических уровней приводит также к тому, что уровни E_e и E_y сближаются.

В результате большое число электронов вытесняется из кремния на его поверхность, образуя заряженный слой толщиной от 3 до 30 нм. В пределах

этого слоя электроны становятся основными носителями, а кремний превращается в материал *n*-типа. Из-за инверсии электропроводности данный слой называют *инверсным слоем* или *каналом *n*-типа*.

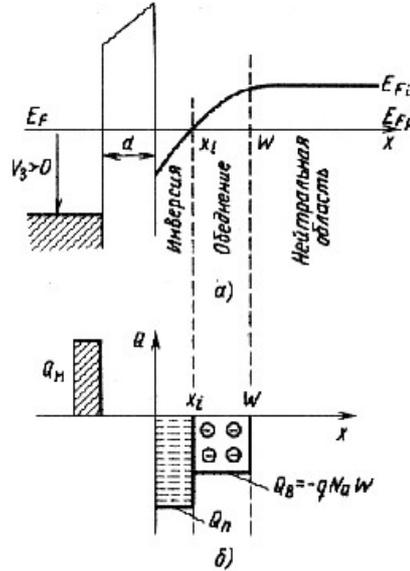


Рис. 5.5. Инверсный режим: а - диаграмма энергетических уровней; б - распределение заряда

Таким образом, в кремнии можно выделить три области. Одна из них, ближайшая к поверхности, представляет канал *n*-типа. Другая является нейтральной. Между этими областями существует обедненный слой толщиной W .

По сути, инверсный режим характеризуется преобладанием основных носителей. В этом режиме образуется канал с такими же свойствами, как у некоторого примесного полупроводника *n*-типа, в котором электроны являются основными носителями.

Говорят, что инверсия является слабой, если выполняли неравенство $U_F < U < 2U_F$

В таком режиме разница между концентрацией основных носителей (электронов) n и концентрацией неосновных носителей (дырок) N_a мала, так инверсный канал не образует. Если же $U > 2U_F$, то инверсия является сильной. Здесь концентрация основных носителей в слое велика ($n > N_a$) что ведет к возникновению инверсного канала.

При равенстве $U = 2U_F$ толщина обедненного слоя W_{max} достигает максимального значения (рис. 5.5,6). За счет изгиба энергетических зон малым изменениям ΔW соответствуют такие приращения ΔU , которые сопровождаются большими приращениями концентрации электронов в инверсном слое. Описанное явление служит причиной того, что малые изменения потенциала U на затворе МОП-структуры определяют тип носителей и их концентрацию в той окрестности границы раздела $\text{SiO}_2\text{-Si}$, которая располагается внутри объема кремния. Таким образом, МОП-структура проявляет емкостные свойства.

Поверхностная плотность индуцированного заряда в кремнии определяется соотношением

$$Q_S = Q_n - qN_a W_{\max} = Q_n + Q_B \quad (5.15)$$

где Q_n - поверхностная плотность заряда, индуцированного в тонком слое в инверсном режиме; $Q_B = -qN_a W_{\max}$ - плотность заряда в обедненной области при максимальной толщине W_{\max} .

Как напряженность электрического поля ε_{\max} , так и толщина обедненной области W_{\max} описываются теми же выражениями, которые ранее были приведены для режима обеднения. При этом символом W_{\max} обозначен параметр, который раньше фигурировал как W .

Используя рис. 5.5, можно записать следующие выражения, относящиеся к объему и поверхности соответственно:

$$p = N_v \exp\left(-\frac{E_F - E_v}{kT}\right) = n_i \cdot \exp\left(\frac{E_i - E_F}{kT}\right) = N_a \quad (5.16)$$

$$n_s = n_i \cdot \exp\left(\frac{E_F - E_i(x_{ок})}{kT}\right) \quad (5.17)$$

Режим сильной инверсии устанавливается в том случае, если концентрация электронов на поверхности и, оказывается равной концентрации дырок в объеме:

$$n_s = N_a \quad (5.18)$$

так что

$$E_F - E_i(x_{ок}) = E_i - E_F = qU_F \quad (5.19)$$

Как можно видеть из рис. 5.5, а, для реализации режима сильной инверсии требуется выполнение равенства

$$U_F(0) = U_T \ln(N_a / n_i) \quad (5.20)$$

откуда

$$U_S = 2U_F \quad (5.21)$$

что выражает связь между поверхностным потенциалом U_S и высотой потенциального барьера в объеме. Определив значение U_S можно, используя формулу (5.12), получить выражение для максимальной толщины:

$$W_{\max} = \left(\frac{2\varepsilon_a \cdot U_S}{q \cdot N_a}\right)^{1/2}$$

или

$$W_{\max} = 2 \cdot \left(\frac{\varepsilon_a \cdot U_F}{q \cdot N_a}\right)^{1/2} \quad (5.22)$$

Можно видеть, что выражение (5.22) аналогично формуле (4.14), которая относится к случаю резкого асимметричного перехода $p-n^+$ -типа,

характеризующегося неравенством $N_d \gg N_a$; внешнее смещение на переход не подается. Роль высоты потенциального барьера U_o здесь играет величина U_F . Однако первый из рассматриваемых переходов создан путем индуцирования n -слоя под действием электрического поля в подложке p -типа. В то же время второй переход обусловлен металлургическим процессом, в ходе которого кристалл легируется донорными примесями. Таким образом, удобно считать, что при инверсии в непосредственной близости от поверхности кремния возникает сильно асимметричный p - n^+ -переход, или, как говорят, канал n -типа. Именно свойства этого канала определяют характеристики МОП-транзистора.

Поверхностный заряд в режимах обогащения, обеднения и инверсии

Рассмотрим полубесконечный кристалл с равномерным распределением примесей, находящийся в состоянии термодинамического равновесия. Его границей служит плоскость $x=0$, а сам кристалл расположен в полупространстве $x>0$. Электрический потенциал в каждой точке зависит только от координаты x и подчиняется одномерному уравнению Пуассона.

$$\frac{d^2U(x)}{dx^2} = -\frac{\rho(x)}{\varepsilon_a} \quad (5.23)$$

В этом уравнении величина $\rho(x)$ - суммарная плотность статического заряда с учетом зарядов положительных и отрицательных ионов, а также зарядов свободных носителей, концентрации которых $n(x)$ и $p(x)$, ε_a - абсолютная диэлектрическая проницаемость полупроводника. Можно записать:

$$\rho(x) = q[p(x) - n(x) - N_a] \quad (5.24)$$

$$p(x) = p_0 \exp[-U(x)/U_T] \quad (5.25)$$

$$n(x) = n_0 \exp[-U(x)/U_T] \quad (5.26)$$

где p_0 и n_0 - концентрация дырок и электронов соответственно при $U(x) = 0$, т. е. на поверхности полупроводника. Очевидно, что $p_0 \approx N_a$ и $n_0 \approx n_i^2 / N_a$

На достаточном удалении от поверхности имеют место равенства $p(x) = 0$, $U(x) = 0$ и выполняется условие электрической нейтральности

$$N_a = p_0 - n_0 \quad (5.27)$$

Объединяя приведенные выражения, получаем

$$\frac{d^2U(x)}{dx^2} = -\frac{q}{\varepsilon_a} \cdot [p_0(e^{-U(x)/U_T} - 1) - n_0 \cdot (e^{U(x)/U_T} - 1)] \quad (5.28)$$

Напомним, что при малых возмущениях, когда $U(x)/U_T \leq 0,2$, справедливы приближенные равенства

$$p_0 e^{-U(x)/U_T} \approx p_0 \left(1 - \frac{U(x)}{U_T}\right) \quad (5.29)$$

$$n_0 e^{U(x)/U_T} \approx n_0 \left(1 + \frac{U(x)}{U_T} \right) \quad (5.30)$$

Эти формулы получены путем отбрасывания нелинейных слагаемых в разложении экспоненты $e^x = 1 + x - x^2/2 + \dots \approx 1 + x$. Таким образом, уравнение (5.28) принимает следующий вид:

$$\frac{d^2 U(x)}{dx^2} = \frac{q}{\varepsilon_a U_T} \cdot (p_0 - n_0) \cdot U(x) = \frac{U(x)}{L_D^2} \quad (5.31)$$

где L_D - так называемая *дебаевская длина* для дырок:

$$L_D = 2\varepsilon_a U_T / (qp_0) \quad (5.32)$$

Интегрируя уравнение (5.31), получаем

$$U(x) = U_s \exp(-x/L_D) \quad (5.33)$$

Таким образом, в частном случае малых возмущений распределение потенциала не зависит от значения U_F и связано с параметром L_D экспоненциальной зависимостью,

В общем случае решение уравнения (5.28) можно найти численными методами. Аналитическая форма записи напряженности электрического поля имеет вид

$$\varepsilon(x) = \frac{2U_T}{L_D} \cdot F \left\{ \frac{1}{U_T}, U(x), \frac{n_0}{p_0} \right\}. \quad (5.34)$$

Далее следует учесть, что поверхностный заряд Q_s связан с напряженностью электрического поля на границе раздела SiO_2 — Si законом Гаусса

$$Q_s = \varepsilon_{an} \varepsilon_s \quad (5.35)$$

Таким образом,

$$Q_s = \frac{2U_T}{\varepsilon_{an} L_D} \cdot F \left\{ \frac{1}{U_T}, U(x), \frac{n_0}{p_0} \right\} \quad (5.36)$$

или

$$Q_s = \tilde{p} - \tilde{n}. \quad (5.37)$$

Величина Q_s представляет суммарный избыточный заряд в равновесном состоянии, причем

$$\begin{aligned} \tilde{p} &= \int_0^\infty [p_x - p_0] dx = p_0 \int_0^\infty [e^{-U(x)/U_T} - 1] \cdot dx \\ \tilde{n} &= \int_0^\infty [n_x - n_0] dx = n_0 \int_0^\infty [e^{U(x)/U_T} - 1] \cdot dx \end{aligned}$$

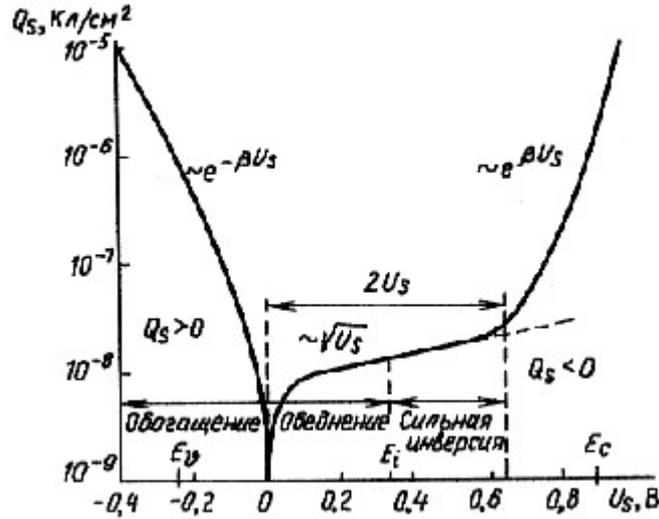


Рис. 5.6. Вольт-фарадная характеристика конденсатора единичной площади как функции поверхностного потенциала при комнатной температуре

На рис. 5.6 изображена зависимость величины Q_s от напряжения U_s для кремниевой подложки с концентрацией $N_a = 4 \cdot 10^{15} \text{ см}^{-3}$ при температуре 300 К. Если $U_s < 0$ и $Q_s > 0$, то возникает режим обогащения дырками. При $0 < U_s < U_F$ и $Q_s > 0$ система находится в режиме обеднения. Если выполняется неравенство $U_p > U_s > 2U_F$, то реализуется режим слабой инверсии. Здесь концентрация электронов еще мала и плотность зарядов Q_s определяется в основном акцепторными ионами в обедненном слое. Если же $U_s > 2U_F$, то возникает режим сильной инверсии. Здесь $Q_s < 0$, причем по модулю данная величина возрастает экспоненциально. При $U_s = 2U_F$ [см. формулу (5.21)] в соответствии с (5.20).

$$U_F = U_T \ln(N_a/n_i)$$

Из данного равенства видно, что можно управлять уровнем инверсии, т. е. минимальным значением напряжения U_s , которое нужно приложить к затвору, чтобы создать инверсный канал в чистом кремнии. Для этого следует изменять концентрации N_a . Отметим, что чаще всего МОП-приборы работают в режиме сильной инверсии.

5.2. Идеальный МОП-конденсатор

В таком конденсаторе напряжение на затворе U_3 складывается из напряжения на оксидном слое $U_{ок}$ и поверхностного напряжения [см. формулу (5.11)]:

$$U_3 = U_{ок} + U_s \quad (5.38)$$

Так как на границе раздела $\text{SiO}_2 - \text{Si}$ поверхностный заряд отсутствует, из теоремы Гаусса вытекает непрерывность нормальных составляющих вектора электрического смещения. Поэтому формула (5.35) применительно к данной поверхности имеет вид

$$\varepsilon_{aок} = \varepsilon_a \varepsilon_s = -Q_s = Q_M \quad (5.39)$$

где $\varepsilon_{aок} = 3,96\varepsilon_o$ - абсолютная диэлектрическая проницаемость оксида. Если даже положить, что в толще оксида существуют заряды, то напряженность электрического поля $\varepsilon_{ок}$ остается постоянной в пределах слоя толщиной $x_{ок}$ и обуславливает падение напряжения:

$$U_{ок} = \varepsilon_{ок} x_{ок} \quad (5.40)$$

Объединяя формулы (5.40) и (5.39), которые определяют удельную емкость оксидного слоя, и подставляя полученную величину в (5.38), находим, что

$$U_3 = -Q_s / C_o + U_s \quad (5.41)$$

Данная формула устанавливает связь между величинами U_3 и Q_s .

Вольт-фарадная характеристика в режиме малого сигнала

Наиболее простая модель МОП-структуры получается в режиме малого сигнала. В этом случае емкость МОП-конденсатора пропорциональна амплитуде переменной составляющей сигнала. Беря дифференциалы обеих частей формулы (5.41), получаем:

$$dU_3 = -dQ_s / C_o + dU_s \quad (5.42)$$

Так как удельная емкость МОП-конденсатора (емкость подложка - затвор) при малом сигнале определяется выражением:

$$C = -dQ_s / dU_3 \quad (5.43)$$

то

$$C = -\frac{dQ_s}{-dQ_s / C_o - dU_s} \quad (5.44)$$

или

$$C = \frac{1}{1/C_o - dU_s / dQ_s} = \frac{1}{1/C_o + 1/C_s} \quad (5.45)$$

где

$$C_s = -dQ_s / dU_s = \varepsilon_a / W \quad (5.46)$$

- удельная емкость, обусловленная существованием области пространственного заряда. В соответствии с формулой (5.45) эквивалентную схему МОП-структуры можно представить в виде последовательного соединения двух конденсаторов. Емкость первого из них $C_o = const$ обусловлена оксидным слоем. Второй конденсатор имеет емкость C_s , которая зависит от напряжения, приложенного к слою разделенного пространственного заряда толщиной W .

Если поверхность полупроводника находится в обедненном состоянии, а режим инверсии не реализуется, то применимы формулы (5.9) и (5.12). Используя эти выражения в (5.41), получаем

$$W = -\frac{\varepsilon_{an}}{C_0} + \left[\left(\frac{\varepsilon_{an}}{C_0} \right)^2 + \frac{\varepsilon_{an} \cdot U_3}{qN_a} \right]^{1/2} \quad (5.47)$$

Отсюда можно определить прежде всего величину $C = \varepsilon_{an} / W$, а затем отношение C/C_0 , которое описывает емкость МОП-структуры в режиме обеднения.

Формула (5.47) справедлива для значений U_3 лежащих интервале от нуля до так называемого *порогового напряжения* U_{nop} [см. выражение (5.52)], при котором поверхность кремния переходит в режим сильной инверсии.

Если известна функция, определяемая формулой (5.36), то ёмкость C_s можно получить в общем виде:

$$C_s = \frac{\varepsilon_{an}}{L_D} \cdot \frac{\left\{ 1 - \exp\left(-\frac{U_s}{U_T}\right) + \frac{n_0}{p_0} \cdot \left[\exp\left(\frac{U_s}{U_T}\right) - 1 \right] \right\}}{F \cdot \left\{ \frac{1}{U_T}, U(x), \frac{n_0}{p_0} \right\}} \quad (5.48)$$

Вольт-фарадная характеристика при низких частотах (5—100 Гц)

Если $U \ll 0$ (условие сильной инверсии), то применительно к формуле (5.45) имеет место неравенство $C_s \gg C_0$ и поэтому $C \approx C_0$. Если величина U_s возрастает и приближается к так называемому *напряжению плоских зон* U_{n3} то $C_s \approx C_0$ и $C < C_0$. Если же $C_s \gg C_0$ (в условиях сильной инверсии), то вновь

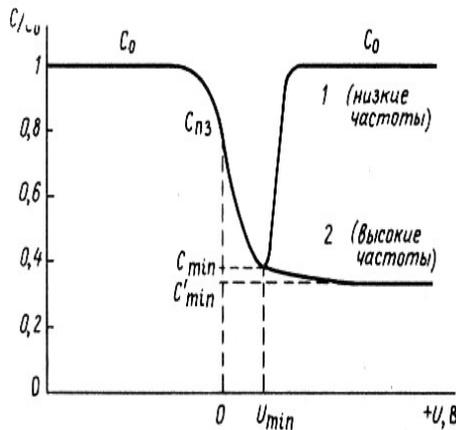


Рис. 5.7. Отношение C/C_0 как функция напряжения, приложенного к затвору в режиме малого сигнала на низких (1) и на высоких (2) частотах

$C \approx C_0$ (кривая 1 на рис. 5.7). На низких частотах изменение числа актов генерации-рекомбинации неосновных носителей (электронов в случае подложки р-типа) пропорционально амплитуде малого сигнала.

Вольт-фарадная характеристика на высоких частотах (около 1 МГц)

На высоких частотах условие обеднения и обогащения остаются теми же, что и на низких частотах. Прежними оказываются также условия инверсного режима, при котором дифференциальная емкость обусловлена генерацией электронов, создающих инверсную область. Однако переменный высокочастотный сигнал может

привести к появлению слоя толщиной W , в пределах которого сигнал изменяет свою полярность, так что

$$\Delta Q_s = -qN_a W \quad (5.49)$$

$$C_s = \frac{-\Delta Q_s}{\Delta U_s} = qN_a \frac{\Delta W}{\Delta U_s} \quad (5.50)$$

откуда на основании (5.12) получаем

$$C_s = \varepsilon_a / W \quad (5.51)$$

В случае $U_s > 2U_F$ величина $W = W_{max} = const$. Как следствие этого, отношение C/C_o в инверсной области постоянно (кривая 2 на рис. 5.7).

Итак, вольт-фарадные характеристики МОП-структуры в режиме малого сигнала зависят от частоты. Положение границы между областями высоких и низких частот зависит от тех факторов, которые определяют скорость процесса генерации-рекомбинации. К их числу относятся освещение кристалла и изменение его температуры.

Пороговое напряжение

Определим пороговое напряжение U_{nop} посредством равенства

$$U_{nop} = U_s - Q_B / C_o \quad (5.52)$$

полагая, что в инверсном слое отсутствуют подвижные заряды. Величина U_{nop} представляет собой то напряжение на затворе, при котором начинается режим сильной инверсии. Таким образом, значение отношения C/C_o , соответствующее пороговому напряжению, можно найти из (5.47), выполнив замену $U_3 = U_{nop}$:

$$\frac{C}{C_o} = \left(1 + \frac{2C_o^2 \cdot U_{nop}}{\varepsilon_a q \cdot N_a} \right)^{1/2}$$

Подставляя в формулу (5.15) величину Q_s из (5.41) и Q_B из (5.52), получаем

$$Q_n = -C_o(U_3 - U_{nop}) \quad (5.53)$$

Данное выражение устанавливает связь между величинами Q_n и $U_3 - U_{nop}$

5.3. Реальный МОП - конденсатор

Так как МОП-конденсатор создается в результате реакции высокотемпературного оксидирования кремния, приходится учитывать ряд эффектов, которые влияют на вид идеальных вольт-фарадных характеристик. К их числу относятся:

- возникновение на границе раздела $SiO_2 - Si$ зарядов, связанных с оксидом;
- различная работа выхода Al и Si;
- образование подвижных ионов, которые могут внедряться и кристаллическую решетку SiO_2 в ходе процесса оксидирования.

В дальнейшем будут рассмотрены первые два фактора, на основе которых удастся понять некоторые важные особенности МОП-технологий. Аналитическое исследование основано на понятии напряжения плоских зон U_{nz}

Факторы, определяющие идеальную вольт-фарадную характеристику

Заряды в оксидном слое

Предположим, что в толще оксида существует заряженный слой с зарядом Q , приходящимся на единицу площади. Данный слой размещается на расстоянии x от границы раздела Al - SiO₂. За счет этого возникает некоторое электрическое поле напряженностью ε , ориентированное в сторону границы раздела SiO₂ - Si. Как следствие, дырки в кремнии, находящиеся вблизи границы раздела, смещаются в глубь полупроводника, а электроны выходят на поверхность. Можно видеть, что при напряжении $U_3 = 0$ энергетические зоны искривляются, причем направление их изгиба в режиме обогащения или обеднения зависит от направления вектора ε . В соответствии со сказанным режим плоских зон будет возникать при $U_3 < 0$.

Напряжение плоских зон, найденное на основании решения уравнения Пуассона, имеет вид

$$U_{nz} = -x_{ok} \varepsilon = -\frac{x_{ok} Q_{ss}}{\varepsilon_{aok}} = -\frac{Q_{ss}}{C_0} \quad (5.54)$$

где Q_{ss} - плотность заряда на границе раздела.

Присутствие заряда плотностью Q_{ss} ведет к появлению поля ε , ориентированного в том же направлении, что и поле, созданное за счет напряжения U_3 . В результате теоретическая вольт-фарадная характеристика смещается параллельно самой себе на величину U_{B1} (как при низких, так и при высоких частотах). Если $Q_{ss} > 0$, то кривая смещается влево, а если $Q_{ss} < 0$, то вправо.

Разность работ выхода

Работа выхода из алюминия и кремния неодинаковы. В соответствии с рис. 5.2 разность между ними:

$$q\Phi_{mn} = q\Phi_{mo} - q\Phi_{no} \quad (5.55)$$

причем для алюминия

$$\begin{aligned} q\Phi_{mn} &= 3,2 - \left[3,25 - \frac{E_g}{2} + qU_F \right] = \\ &= 3,2 - (3,25 + 0,55 + qU_F) = -0,6 - qU_F, \text{ ЭВ} \end{aligned} \quad (5.56)$$

$$\Phi_{mn} = -0,6 - U_F, \quad (5.57)$$

где $U_F > 0$ в случае кремния p -типа и $U_F < 0$ в случае кремния n -типа.

Если Φ_{mn} не равно 0, то при коротком замыкании МОП-структуры возникает поток электронов в направлении Al - Si или в противоположном направлении. Как следствие, при $U_3 = 0$ режим плоских зон не реализуется и вольт-фарадная характеристика

смещается параллельно самой себе влево по отношению к идеальной кривой на величину:

$$U_{nz} = \Phi_{mn} \quad (5.58)$$

Итак, рабочее напряжение МОП-конденсатора в режиме плоских зон, обусловленное различием работ выхода алюминия и кремния, а также наличием поверхностного заряда, составляет:

$$U_{nz} = \Phi_{mn} - Q_{ss}/C_0. \quad (5.59)$$

Это выражение важно потому, что оно дает возможность вычислить поверхностную плотность заряда, так как значение Φ_{mn} находится экспериментально, а напряжение U_{nz} определяется путем сопоставления теоретической и экспериментальной кривых.

Пороговое напряжение и напряжение плоских зон

В заключение параграфа перечислим некоторые выражения, полезные при исследовании МОП-приборов.

Формулу (5.38) на основании (5.57) можно представить и виде

$$U_z = U_{ок} + U_s + \Phi_{mn} - Q_{ss}/C_0 \quad (5.60)$$

В идеальной МОП-структуре $\Phi_{mn} = 0$; $Q_{ss} = 0$ и

$$U_{ок} = Q_s/C_0 \quad (5.61)$$

где $Q_s = Q_B + Q_n$. Иными словами, плотность заряда в приповерхностной области кремния складывается из плотности заряда в обедненной области Q_B и плотности подвижного заряда Q_n , который выталкивается на поверхность за счет приложенного напряжения.

Вычисление порогового напряжения U_{nop} которое нужно приложить к затвору для того, чтобы обеспечить инверсный режим, можно упростить, если считать, что на границе режима инверсии выполняются равенства $Q_s = Q_B$ и $Q_n = 0$. При $U_z > U_{nop}$ любой заряд, появляющийся на поверхности кремния, обусловлен плотностью Q_n . Отсюда, используя результаты теории, развитой применительно к резкому p - n -переходу, получаем формулу для оценки толщины обедненного слоя в МОП-структуре с подложкой из кремния p -типа

$$W = \left[\frac{4\epsilon_{an} U_F}{qN_a} \right]^{1/2} \quad (5.62)$$

а также для оценки плотности заряда, связанного с обедненной областью,

$$Q_B = -qN_a W \quad (5.63)$$

Формула, определяющая пороговое напряжение [см. выражение (5.52)], принимает более полный вид:

$$U_{nap} = U_{nz} + U_s - Q_B/C_0 \quad (5.64)$$

причем

$$U_{nz} = \Phi_{mn} - \frac{Q_{ss}}{C_0} - \frac{1}{C_0} \cdot \int_0^{x_{ок}} \frac{x}{x_{ок}} \cdot \rho(x) dx \quad (5.65)$$

где последнее слагаемое правой части описывает заряд в оксиде кремния. При $U_3 > U_{nop}$, а именно в этом режиме работают приборы,

$$U_3 = U_{nop} + Q_n / C_0 \quad (5.66)$$

Следует еще раз отметить, что понятие порогового напряжения является одним из основных в теории МОП-приборов.

5.4. Идеальный МОП- транзистор

Как уже было указано в п. 5.1.1, в МОП-структуре создается тонкий слой носителей электрического заряда (электронов в кремнии *p*-типа), который располагается вблизи границы раздела SiO_2 - Si . Этот слой возникает при условии, что напряжения, приложенного к затвору, достаточно для инвертирования типа электропроводности подложки. Данное явление служит основой работы МОП-транзистора.

Полевой МОП-транзистор представляет собой прибор, в котором основные процессы происходят на поверхности кремния. Усилительный эффект этого транзистора основан на том, что здесь в обедненной области происходит изменение плотности заряда, за счет чего изменяется сопротивление канала, соединяющего исток и сток прибора. Управляющим фактором служит внешний потенциал, приложенный к затвору.

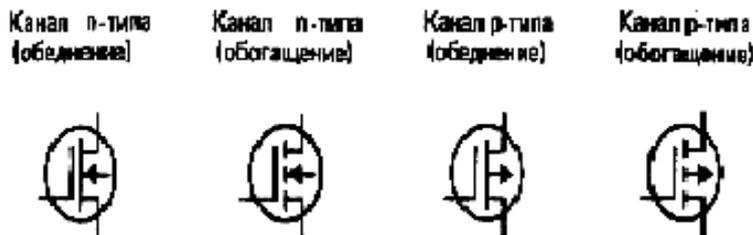


Рис. 5.8. Условные графические обозначения различных МОП - транзисторов

В зависимости от типа электропроводности канала различают два вида МОП-транзисторов: с каналом *p*- и *n*-типа. Каждый из этих двух типов приборов может работать в двух режимах: обеднения и обогащения (рис. 5,8).

Идеальная физическая структура

Чтобы предельно упростить изложение принципа работы данного прибора, ограничимся лишь случаем МОП-транзистора с каналом n^+ -типа, работающим в режиме обогащения. Такой прибор состоит из кремниевой подложки p -типа, на которой способом диффузии созданы два островка (области) n^+ -типа, называемые истоком (И) и стоком (С) (рис. 5.9). На участок поверхности полупроводника длиной L нанесен слой оксида SiO_2 толщиной d . Наружная поверхность этого слоя покрыта алюминий-евой пленкой, которая играет роль затвора (З). Четвертым выводом прибора служит вывод подложки (П).

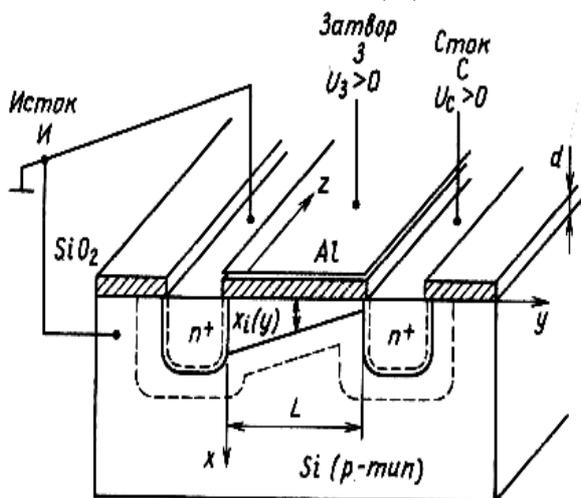


Рис. 5.9. Основная геометрическая конфигурация МОП-транзистора с каналом n -типа, работающего в режиме обогащения

Принцип работы

Говорят, что прибор работает в режиме обогащения, если при $U_3 = 0$ поверхностный канал отсутствует и проводимость участка цепи между истоком и стоком равна нулю. Будем считать, что $U_c = 0$ (напряжение U_c представляет собой потенциал стока относительно истока). При этом

p - n^+ - переход, образованный стоком и подложкой, смещен в обратном

направлении и ток в цепи оказывается весьма малым. Значение этого тока составляет около 1 нА при комнатной температуре. Чтобы ток между истоком и стоком оказался достаточно большим, нужно, чтобы в МОП-структуре в результате приложения напряжения затвор - подложка возникла сильная инверсия, обеспечивающая появление канала n -типа между областями истока и стока.

В соответствии с формулами (5.41), (5.9), (5.21) и (5.20) потенциал затвора относительно истока

$$U_3 = -\frac{Q_s}{C_0} + U_s = \frac{qN_a W_{\max}}{C_0} + 2U_T \ln \frac{N_a}{n_i} \quad (5.67)$$

На границе режима сильной инверсии эта величина есть пороговое напряжение, определяемое соотношением (5.52) или (5.64). В условиях сильной инверсии, т. е. при достаточно большом напряжении U_3 следует различать случаи больших и малых напряжений U_c

Если напряжение U_c мало, то электроны, индуцированные в канале, перемещаются от истока к стоку под действием сил электрического поля. Полевой транзистор ведет себя подобно резистору с сопротивлением

$$R = \frac{L}{Z} \cdot \frac{1}{g_c} = -\frac{L}{Z\mu_n Q_n} = \frac{1}{q\mu_n N_d W} \quad (5.68)$$

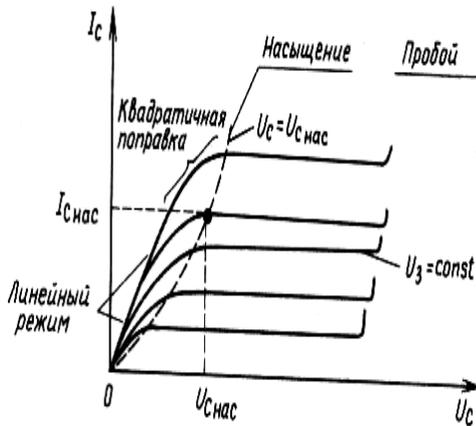


Рис. 5.10. Вольтамперные характеристики при различных значениях $U_3 = \text{const}$. Можно отметить существование омической или линейной области (с квадратичной поправкой), области насыщения и области пробоя.

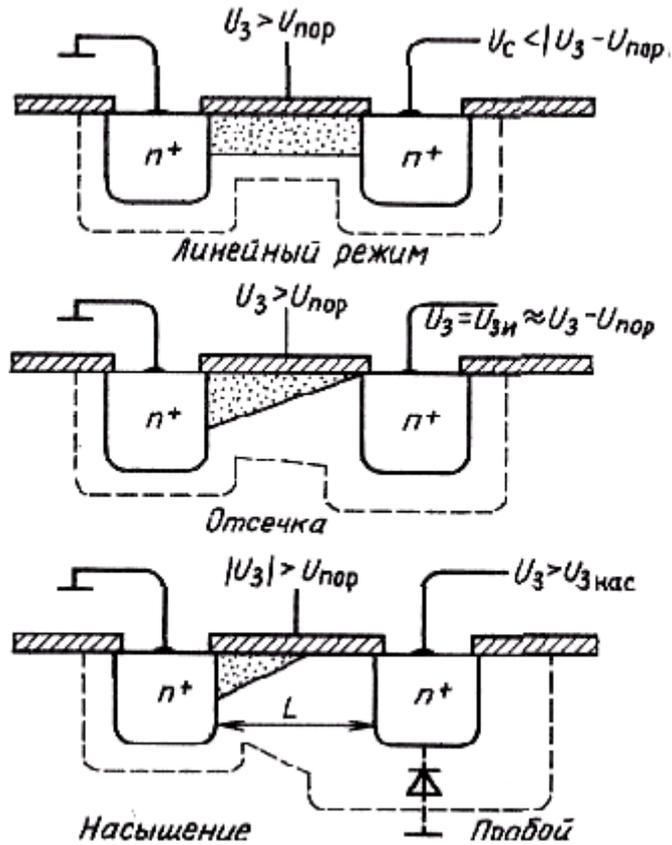
(рис. 5.10). Если продолжать увеличивать напряжение U_c , то напряжение, приложенное к слою SiO_2 в окрестности стока, уменьшается. Как следствие, в точке канала, ближайшей к стоку и называемой *точкой отсечки канала*, инверсии типа электропроводности не происходит. Такое явление наблюдается при некотором напряжении, называемом *напряжением насыщения* (рис. 5.10). Поверхность полупроводника в окрестности указанной точки находится не в режиме инверсии, а в режиме обеднения. Если $U_c > U_{c \text{ нас}}$, то точка отсечки канала перемещается, приближаясь к истоку. При этом ток между истоком и стоком не возрастает (режим насыщения тока). Другими словами, в таком режиме прибор ведет себя как источник тока, управляемый напряжением U_3

Вольт-амперные характеристики

До сих пор проводилось качественное рассмотрение вольтамперных характеристик $I_c - U_c$. Построим теперь математическую модель, которая описывает эти характеристики в ненасыщенном режиме. Покажем, что при определенных условиях эта же модель может описывать характеристики прибора, работающего в режиме насыщения.

Здесь L и Z -длина и ширина канала соответственно, g_c - удельная проводимость канала, μ_n - подвижность электронов в канале, Q_n - поверхностная плотность заряда электронов в канале. В условиях сильной инверсии характеристики прибора $I_c - U_c$ оказываются практически линейными (омическими) (рис. 5.10).

Если увеличивать напряжение U_c , оставляя неизменным напряжение $U_3 > U_{\text{пор}}$, то на участке канала, ближайшем к стоку, концентрация электронов уменьшается и поэтому сопротивление канала растет. В результате вольтамперные характеристики изгибаются вниз



Ненасыщенный режим-линейный режим

Рассмотрим канал длиной L (см. рис. 5.9). Падение напряжения элементе канала длиной dy

$$dU = I_c dR = -I_c \frac{dy}{Z\mu_n Q_n(y)} \quad (5.69)$$

Так как $I_3 = 0$, то ток I_c неизменен вдоль координаты y . Поэтому

$$I_c dy = -Z\mu_n Q_n(y) dU \quad (5.70)$$

Интегрируя равенство (5.70) по y от нуля до L и по U от нуля до U_c , получаем

$$I_c L = -Z\mu_n \int_0^{U_3} Q_n(y) dU. \quad (5.71)$$

Согласно формуле (5.15), в режиме инверсии плотность заряда инверсного слоя является функцией координаты точки наблюдения, измеряемой вдоль канала:

$$Q_n(y) = Q_s(y) - Q_B(y) \quad (5.72)$$

Используя формулу (5.41) и вводя напряжение плоских зон, получаем

$$Q_s(y) = -C_0 [U_3 - U_{n3} - U_s(y)] \quad (5.73)$$

Однако в соответствии с условием инверсии

$$U_s(y) = 2U_F - U(y) \quad (5.74)$$

где $U(y)$ - напряжение в точке канала с координатой y .

Следовательно,

$$Q_S(y) = -C_0[U_3 - U_{n3} - 2U_F - U(y)] \quad (5.75)$$

С другой стороны, используя выражения (5.64) и (5.63), находим

$$Q_B(y) = -qN_a W(y) = -2qN_a \varepsilon_{an} [2U_F - U(y)]^{1/2}. \quad (5.76)$$

Вычисления можно существенно упростить, если считать, плотность заряда Q_B в той области подложки, где сосредоточен пространственный заряд, не зависит от координаты y .

Поэтому в первом приближении будем считать, что

$$Q_B(y) \approx Q_{B\max} \quad (5.77)$$

где $Q_{B\max}$ - максимальное значение Q_B .

Подставляя выражения (5.75) и (5.77) в формулу (5.73), получаем

$$U(x) = U_s \exp(-x/L_D)$$

$$Q_n(y) = -C_0 \left[U_3 - U_{n3} - 2U_F - U_S(y) + \frac{Q_{B\max}}{C_0} \right] \quad (5.78)$$

Или с учетом формулу (5.65)

$$Q_n(y) = -C_0 [U_3 - U_{nop} - U(y)] \quad (5.79)$$

Если теперь подставить (5.79) в (5.71), то получим формулу, описывающую вольтамперную характеристику в линейной области:

$$I_C \approx \frac{Z\mu_n C_0}{L} \left[(U_3 - U_{nop})U_C - \frac{1}{2}U_C^2 \right] \quad (5.80)$$

Если учесть изменение величины $Q_n(y)$, описываемое выражением (5.76), то вместо формулы (5.80) получить более точное равенство.

$$I_C \approx \frac{Z\mu_n C_0}{L} \left\{ \left[(U_3 - U_{n3} - 2U_F - \frac{U_C}{2})U_C - \frac{2(2q\varepsilon_a N_a)}{3C_0} (U_C + 2U_F)^{3/2} - (2U_F)^{3/2} \right] \right\} \quad (5.81)$$

Режим насыщения

Равенство $I_C = I_{снас}$ означает появление точки отсечки канала (рис. 5.10 и 5.11). Это происходит при напряжении $U_C > U_{снас}$ таком, при котором $Q_n(L) = 0$. Используя в формуле (5.79) условия $Q_n(L) = 0$, получаем:

$$U_{снас} \ll (U_3 - U_{nop}) \quad (5.82)$$

Наконец, подставив выражение (5.82) в формулу (5.80), находим ток насыщения:

$$I_{снас} \approx \frac{Z\mu_n C_0}{L} \frac{[U_3 - U_{nop}]^2}{2} = \frac{Z\mu_n C_0}{2L} U_{снас} \quad (5.83)$$

Параметры эквивалентной схемы в динамическом режиме

Параметрами рассматриваемого прибора в динамическом режиме являются выходная проводимость g_c и крутизна стокзатворной характеристики S . В линейном режиме

$$Q = \left. \frac{\partial I}{\partial U} \right|_{U_3 = \text{const}} = \frac{Z\mu_n C_0}{L} (U_3 - U_{\text{пор}} - U_c) \quad (5.84)$$

При насыщении

$$g_c = 0 \quad (5.85)$$

Соответственно в линейном режиме

$$S = \left. \frac{\partial I}{\partial U} \right|_{U_3 = \text{const}} = \frac{Z\mu_n C_0}{L} U_c \quad (5.86)$$

В режиме насыщения

$$S = \frac{Z\mu_n C_0}{L} (U_3 - U_{\text{пор}}) \quad (5.87)$$

Если к напряжениям U_3 и U_c добавляются напряжения сигналов с малыми амплитудами, то можно считать, что связь между током I и напряжением U малого сигнала остается линейной и определяется малосигнальными параметрами. Работу МОП-транзистора на низких частотах можно описать эквивалентной схемой (рис. 5.12).

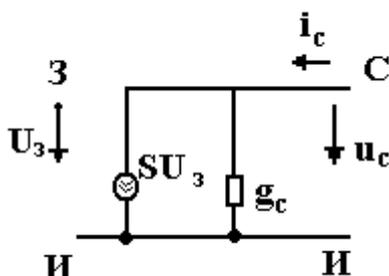


Рис. 5.12. Малосигнальная эквивалентная схема МОП-транзистора на низких частотах

5.5. Реальный МОП- транзистор

Рассмотрим принципиальные различия между идеальным реальным МОП-транзистором.

Выходное сопротивление в режиме насыщения

В соответствии с формулой (5.85) проводимость в данном режиме $g_c = 0$, т. е. выходное сопротивление прибора неограниченно велико. Это не совсем так, поскольку при $U_c > U_{\text{снас}}$ наличие точки отсечки канала обеспечивает равенство $U_{\text{снас}} = \text{const}$; в то же время эта точка смещается на малый отрезок

длиной l по направлению к истоку. Сокращение длины канала L приводит к росту тока стока, который выражается формулой:

$$I_c = I_{снас} L / (L - l) \quad (5.88)$$

Выходная вольтамперная характеристика прибора в области насыщения, т.е. при $U_c > U_{снас}$, имеет некоторый наклон из-за роста тока I_c при увеличении напряжения U_c .

Граничная частота передачи

Если напряжение на затворе получает приращение ΔU_3 , то возникает приращение тока стока

$$\Delta I_c = S \Delta U_3 \quad (5.89)$$

Однако можно указать отрезок времени такой длительности t_0 , при которой ток за это время остается постоянным; при этом заряд в канале увеличивается на величину ΔQ_n . Следовательно,

$$\Delta I_c t_0 = ZL \Delta Q_n = ZLC_0 \Delta U_3 \quad (5.90)$$

Для оценки рабочих частот прибора используют понятие граничной частоты, которую определяют выражением

$$f_{cp} = \frac{1}{2\pi t_0} = \frac{1}{2\pi ZLC_0 (\Delta U_3 / \Delta I_c)} = \frac{S}{2\pi C_0 ZL}$$

(5.91)

Отсюда с учетом формулы (5.86) находим максимальную, рабочую частоту

$$f_{max} = \mu_n U_c / (2\pi L^2) \quad (5.92)$$

имеющую порядок 10^8 Гц. Для повышения граничной частоты следует увеличивать параметр μ_n и уменьшать длину канала L . Так как $\mu_n > \mu_p$, то в современных ИС предпочитают использовать приборы с каналом n -типа, а не p -типа. Это в равной мере относится как к обогащенному, так и к обедненному режимам.

Ограничения, связанные с электрическим пробоем

Предельные режимы работы МОП-транзистора характеризуются тем, что напряженность внутреннего электрического поля в обедненных областях переходов исток-подложка и сток-подложка достигает 40—60 В/мкм. При такой напряженности поля происходит лавинное размножение носителей (см. рис. 5.10 и 5.11), в результате которого эти p - n -переходы могут разрушаться.

5.6. Полевой транзистор с управляющим p - n - переходом

Полевой транзистор с управляющим p - n -переходом имеет меньшее входное сопротивление по сравнению с МОП-транзистором. Однако оно все же больше, чем входное сопротивление биполярного транзистора. При правильном проектировании такие транзисторы должны иметь параметры,

недостижимые в рамках биполярной и МОП-технологий. Примером могут служить арсенид-галлиевые полевые транзисторы, предназначенные для работы на высоких частотах, хотя фактически приборы представляют особый класс транзисторов (см. § 5.8)

Существуют полевые транзисторы с управляющим p - n -переходом и каналами p - или n -типа. По физической структуре они аналогичны биполярным p - n - p - и n - p - n -транзисторам, принцип их работы совершенно иной. Общим является то, что как в полевом транзисторе с управляющим p - n -переходом, так и в биполярном транзисторе (см. гл. 6) используются явления, происходящие в объеме полупроводника.

Структура и принцип действия

На рис. 5.13 представлены структура и условное графическое изображение полевого транзистора с управляющим переходом и каналом p -типа, соединяющим области истока и стока. Здесь электрод истока

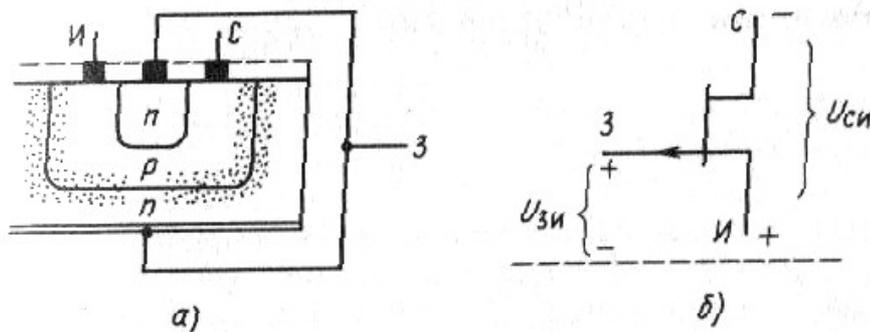


Рис. 5.13. Полевой транзистор с управляющим p - n -переходом и каналом p -типа: а – физическая структура; б – условное графическое обозначение.

заземлен; режим цепи стока выбран таким образом, что происходит дрейфовое движение основных носителей в p -канале от истока к стоку. К электроду затвора относительно истока приложено некоторое напряжение смещения $U_3 > 0$. При этом p - n -переходы смещены в обратном направлении и толщина обедненного слоя W возрастает, а толщина канала сокращается, что ведет к соответствующему изменению проводимости. Поэтому данный прибор можно рассматривать как переменный резистор, управляемый напряжением U_3 . Воспользуемся тем, что на основании формулы (4.17) для n^+ - p -перехода, у которого $N_d \gg N_a$, можно записать.

$$W \approx \left[\frac{2\varepsilon_{an}(U_0 + U_3)}{qN_a} \right]^{1/2} \quad (5.93)$$

В режиме, которому соответствует рис. 5.14, а когда $U_3 = U_c = 0$, ток отсутствует. Толщина обедненной области постоянна вдоль канала и определяется выражением

$$h \approx \left(\frac{2\varepsilon_{an} U_0}{qN_a} \right)^{1/2} \quad (5.94)$$

где $U_0 = U_T \ln(N_a N_d / n_i^2)$

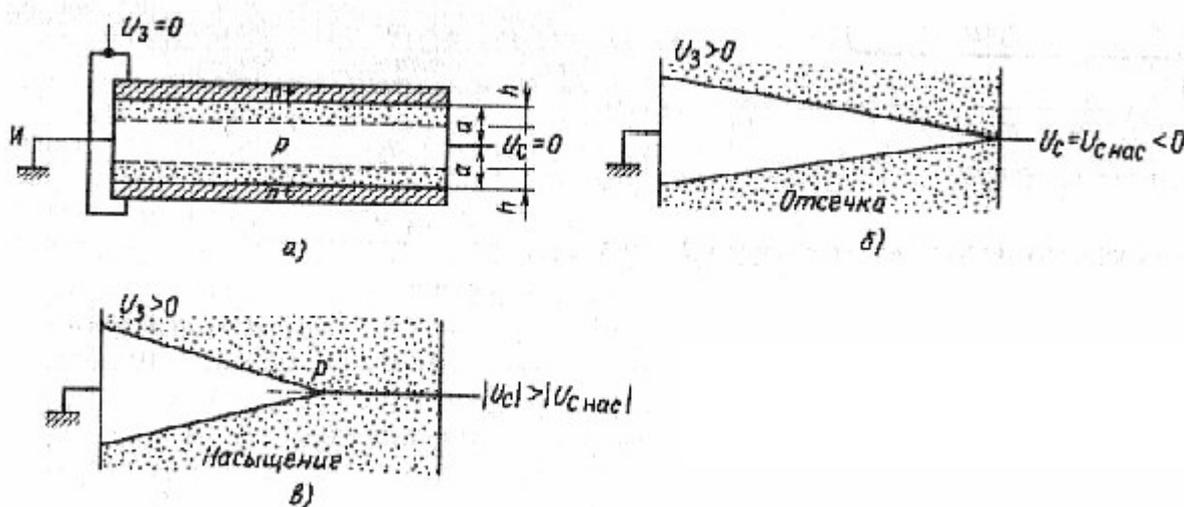


Рис 5.14. Поперечный разрез полевого транзистора с управляющим p - n -переходом и каналом p -типа при различных напряжениях

Если напряжение $U_3 > 0$, то возникает поток электронов и, как следствие, ток в цепи (рис. 5.14, б, в соответствуют случаи $U_c \geq U_{cнас}$).

Рабочие характеристики

Качественный анализ работы полевого транзистора с каналом p -типа

Кратко изучим выходные вольтамперные характеристик $I_c - U_c$ данного полевого транзистора (рис. 5.15). При фиксированном напряжении $U_3 = const$ ток I_c возрастает до тех пор, пока напряжение стока U_c не достигнет значения, выше которого ток I_c остается постоянным. Физически это означает, что канал переходит в режим отсечки.

Чем больше значение $|U_c|$, тем меньше сказывается влияние смещения, поданного на n^+ - p -перехода, и, как следствие, тем толще канал. Если $U_c > U_{cнас}$, то канал перекрывается в точке расположенной перед стоком. Ток стока остается постоянным, так как все носители, инжектированные в проводящую область, достигают стока, не испытывая рекомбинации.

Можно заметить, что чем более положителен потенциал затвора

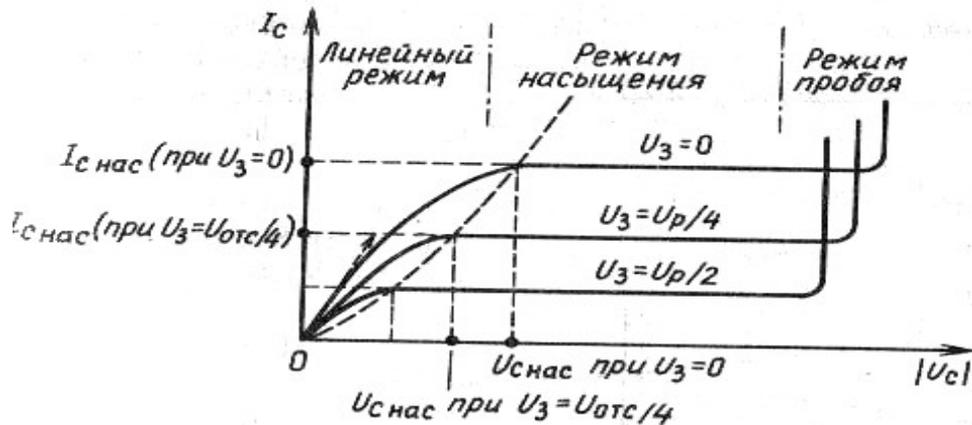


Рис. 5.15. Семейство выходных характеристик полевого транзистора с управляющим p-n-переходом

относительно истока, тем меньше ток насыщения.

Дело в том, что при этом сокращается начальная толщина канала, что приводит к возрастанию начального сопротивления.

Качественный анализ полевого транзистора с каналом n- типа

Будем считать, что выполняются следующие предположения: Размер области, обедненной зарядами, изменяется по линейному закону (плавный канал). Переход $p^+ - n$ является резким. Явлениями, происходящими вне канала толщиной W , можно пренебречь. В пределах канала n-типа концентрация легирующей примеси постоянна. Имея в виду физическую геометрическую симметрию прибора, можно провести аналитическое рассмотрение, обратившись к поперечному размеру (рис. 5.16).

Обозначим символами y_1 и y_2 толщину обедненной области у истока и стока соответственно.

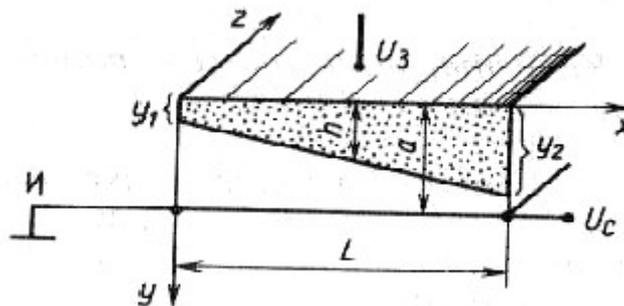


Рис. 5.16. Поперечный разрез верхней половины полевого транзистора с управляющим p-n-переходом и каналом n-типа (упрощенное изображение)

Если $N_a \gg N_d$, то толщина обедненного слоя на расстоянии x от истока

$$h = \left\{ \frac{2\varepsilon_{an}}{qN_d} [U_0 + U_3 + U(x)] \right\}^{1/2} \quad (5.95)$$

$U(x)$ - функция, описывающая распределение потенциала вдоль канала, возникающее под действием приложенного напряжения U_c . Напряжения $U(x)$ и U_3 складываются с напряжением U_0 и смещают p^+ - n -переход в обратном направлении. При $x = 0$ напряжение $U(x) = 0$ и поэтому

$$y_1 = \left[\frac{2\varepsilon_a}{qN_d} (U_0 + U_3) \right]^{1/2} \quad (5.96)$$

Если же $x=L$, то $U(x)=U_c$, так

$$y_2 = \left[\frac{2\varepsilon_a}{qN_d} (U_0 + U_3 + U_c) \right]^{1/2} \quad (5.97)$$

Канал перекрывается при условии, что $U_c = U_{снас}$, при этом $y_2 = a$. Следовательно,

$$a = \left[\frac{2\varepsilon_a}{qN_d} (U_0 + U_3 + U_{снас}) \right]^{1/2} \quad (5.98)$$

В это выражение входит напряжение отсечки канала

$$U_{омс} = U_0 + U_3 + U_{снас} \quad (5.99)$$

Поэтому

$$U_{омс} = qN_d a^2 / (2\varepsilon_a) \quad (5.100)$$

Откуда

$$U_{снас} = U_{омс} - U_3 - U_0 = \frac{qN_d W^2}{8\varepsilon_{an}} - U_3 - U_T \ln \frac{N_d N_a}{n_i^2} \quad (5.101)$$

Выведем выражение для тока I_c . В полевом транзисторе с управляющим p - n -переходом и каналом n -типа напряжение затвора относительно истока $U_3 < 0$, в то время как $U_c > 0$. Напряжение U_c создает электрическое поле с составляющей, вектора ξ_x , направленной в сторону, противоположную движению электронов. Это поле вызывает дрейф основных носителей в канале по направлению от истока к стоку.

Плотность тока в канале

$$J_x = \sigma(x) \cdot \varepsilon_x \quad (5.102)$$

С другой стороны,

$$J_x = \frac{I_c}{Z[a - h(x)]} \quad (5.103)$$

Объединяя формулы (5.102) и (5.103), получаем

$$I_c = \sigma \varepsilon_x Z[a - h(x)] \quad (5.104)$$

Так как

$$\varepsilon_x = -dU(x)/dx \quad (5.105)$$

то, беря абсолютное значение ξ_x , получаем

$$I_c = \sigma \frac{dU(x)}{dx} Z[a - h(x)] \quad (5.106)$$

Если напряжение $U_c > 0$ относительно истока, то электроны в канале перемещаются в сторону стока, так что ток $I_c > 0$ (в соответствии с общепринятой договоренностью положительное направление тока противоположно направлению движения электронов).

Имея в виду, что

$$\sigma = q\mu_n N_d \quad (5.107)$$

и что для всего прибора должен быть взят множитель $2a - 2p(x)$, преобразуем формулу (5.106) следующим образом:

$$I_c dx = 2q\mu_n N_d Z[a - h(x)] dU \quad (5.108)$$

Дифференцируя выражение (5.95), получаем

$$dU = \frac{qN_d}{\epsilon_a} h dh \quad (5.109)$$

Подставив дифференциал dU в формулу (5.108) и проведя интегрирование по x в пределах от 0 до L , находим

$$\begin{aligned} I_c &= \frac{Z\mu_n q^2 N_d^2}{\epsilon_a L} [a(y_2^2 - y_1^2) - \frac{2}{3}(y_2^3 - y_1^3)] = \\ &= I_{omc} \left[\frac{U_c}{U_{omc}} - \frac{2}{3} \left(\frac{U_c + U_3 + U_0}{U_{omc}} \right)^{3/2} + \frac{2}{3} \left(\frac{U_3 + U_0}{U_{omc}} \right)^{3/2} \right] \end{aligned}$$

где

$$I_{omc} = \frac{Zq^2 \mu_n N_d^2 a^2}{\epsilon_a L} \quad (5.110)$$

Напряжение отсечки канала U_{omc} определяется выражением (5.100).

Выражение (5.110) можно представить в ином виде:

$$I_c = G_0 \left[U_c - \frac{2}{3} U_{omc} \left(\frac{U_c + U_3 + U_0}{U_{omc}} \right)^{3/2} + \frac{2}{3} U_{omc} \left(\frac{U_3 + U_0}{U_{omc}} \right)^{3/2} \right] \quad (5.111)$$

где

$$G_0 = \frac{I_{omc}}{U_{omc}} = \frac{2aZq\mu_n N_d}{L} \quad (5.112)$$

- проводимость, мСм, канала n -типа при отсутствии обедненного слоя.

В линейной области, когда $U_c \ll U_3 + U_0$, выражение (5.110) или (5.111) можно разложить в степенные ряды, и тогда получим

$$I_c \approx G_0 U_c \left[1 - \left(\frac{U_3 + U_0}{U_{omc}} \right)^{1/2} \right] \quad (5.113)$$

При этом проводимость канала

$$g_c = \left. \frac{\partial I_c}{\partial U_c} \right|_{U_3=const} = G_0 \left[1 - \left(\frac{U_3 + U_0}{U_{omc}} \right)^{1/2} \right] \quad (5.114)$$

Крутизна стокзатворной характеристики

$$S = \left. \frac{\partial I_c}{\partial U_c} \right|_{U_c = \text{const}} = \frac{G_0 U_c}{2U_{omc}} \left(\frac{U_{omc}}{U_3 + U_0} \right)^{1/2} \quad (5.115)$$

В режиме насыщения

$$S_{HAC} = G_0 \left[1 - \left(\frac{U_3 + U_0}{U_{omc}} \right)^{1/2} \right] \quad (5.116)$$

Эквивалентная схема

Эквивалентная схема описывает работу полевого транзистора с управляющим p - n -переходом в режиме малого сигнала, под действием которого рабочая точка лишь незначительно перемещается. На рис. 5.17 изображена эквивалентная схема полевого транзистора, работающего на низких частотах. Следует заметить, что крутизна S здесь играет роль, аналогичную роли параметров α и β для биполярных транзисторов.

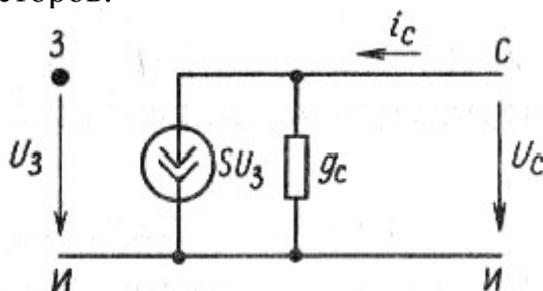


Рис. 5.17. Эквивалентная схема полевого транзистора с управляющим p - n -переходом для области низких частот в режиме малого сигнала.

Реальный полевой транзистор с управляющим p - n - переходом

Характеристики реального полевого транзистора отличаются от характеристик идеального прибора как в статическом, так и в динамическом режиме.

Статический режим

Ток стока реального транзистора слегка возрастает и после того как напряжение U_c достигает уровня $U_{cнас}$. В точке отсечки $U_c = U_{cнас}$; если же $U_c > U_{cнас}$, то разностное напряжение оказывается приложенным к обедненной области p - n -перехода стока. При этом эффективная длина канала уменьшается и ток стока возрастает.

Динамический режим

Диапазон рабочих частот реального полевого транзистора ограничен максимальной частотой

$$f_T = \frac{S}{2\pi C_{BX}} = \frac{2\mu_n q N_d a^2}{\pi \epsilon_a L} \quad (5.117)$$

где C_{BX} - входная емкость, входящая в малосигнальную эквивалентную схему полевого транзистора. Для работы на высоких частотах следует сокращать длину канала L и использовать каналы n -типа.

Реальному транзистору свойственны ограничения, связанные с лавинным пробоем за счет токов утечки затвора, тепловым пробоем из-за выделения теплоты в канале, а также с пробоем под действием процессов генерации-рекомбинации, обусловленных поверхностными явлениями.

5.7. Полевой транзистор с управляющим переходом металл - полупроводник

Такой транзистор по принципу действия напоминает полевой транзистор с управляющим p - n -переходом. Однако здесь затвором служит не p - n -переход, а выпрямляющий контакт металл-полупроводник. Полевые транзисторы создают на основе сложных полупроводников n -типа (группы III—V), подобных арсениду галлия GaAs. Здесь высокая подвижность электронов позволяет свести к минимуму последовательное сопротивление и увеличить граничную частоту.

В транзисторе с эпитаксиальным слоем, выращенном на полупроводниковой подложке, пороговое напряжение

$$U_{nop} = U_0 - U_{omc} \quad (5.118)$$

Здесь U_0 - контактная разность потенциалов в затворном переходе, достаточная для создания обедненного режима в канале; U_{omc} - напряжение отсечки канала. Можно показать, что для полевого транзистора с управляющим p - n -переходом справедлива следующая формула:

$$I_{Cна} = I_{omc} \left\{ \frac{1}{3} - \frac{U_3 + U_0}{U_{omc}} + \frac{2}{3} \left[1 - \left(\frac{U_3 - U_{nop}}{U_{omc}} \right)^{3/2} \right] \right\} \quad (5.119)$$

Аналогично, для полевого транзистора с управляющим переходом металл-полупроводник, работающего при насыщении, в малой окрестности точки $U_{нар}$ справедливо выражение

$$I_{CННА} = \frac{I_{omc}}{2} \left\{ \frac{1}{3} - \left(1 - \frac{U_3 + U_{nop}}{U_{omc}} \right) + \frac{2}{3} \left(1 - \left(\frac{U_3 + U_{nop}}{U_{omc}} \right)^{3/2} \right) \right\} \quad (5.120)$$

или

$$I_{Cнас} \approx \frac{z\mu_n \epsilon_a}{2aL} (U_3 + U_{nop})^2 \quad (5.121)$$

Здесь использовано разложение в ряд Тейлора, справедливое при условии $(U_3 - U_{ноп}) / U_{омс} \ll 1$.

Коэффициент $1/2$ в формуле (5.120) означает, что рассматривается лишь верхняя половина канала; отрицательный знак величины U_3 объясняется полярностью этого напряжения.

Арсенид - галлиевый полевой транзистор с управляющим переходом металл - полупроводник был одним из первых приборов, созданных на базе элементов III и V групп. Это единственный промышленно выпускаемый прибор такого типа, который может быть как дискретным элементом, так и составной частью ИС (см. п. 9.9.3). Время переключения подобных транзисторов менее 100 пс. Широкое внедрение их сдерживается трудностями, которые, несомненно, будут преодолены в ближайшем будущем.

6. ФИЗИЧЕСКИЕ ПРИНЦИПЫ РАБОТЫ БИПОЛЯРНОГО ТРАНЗИСТОРА И ТИРИСТОРА

Биполярный транзистор является наиболее распространенным активным полупроводниковым прибором. Такой транзистор состоит из двух взаимодействующих *p-n*-переходов, созданных в объеме монокристалла кремния или германия. Работа биполярного транзистора основана на явлениях, происходящих в объеме полупроводника. Два *p-n*-перехода разделяют три области, называемые *эмиттером*, *базой* и *коллектором*. В зависимости от характера примесей в этих областях принято различать транзисторы типа *n-p-n* и *p-n-p*. Ограничим наше рассмотрение приборами типа *n-p-n*, которые в настоящее время используются чаще, имеют лучшие характеристики в области высоких частот и большее усиление при одной и той же концентрации примесей и одинаковой геометрии. Это объясняется тем, что подвижность электронов в два-три раза выше подвижности дырок. Термин «биполярный транзистор» указывает на то, что работа данного прибора связана с движением как электронов, так и дырок.

В настоящей главе будут изучены модели, с помощью которых можно отразить внутренний механизм работы данного прибора. Известно несколько таких моделей. К их числу относятся: модель на постоянном токе, или модель для большого сигнала, известная под названием модели Эберса - Молла, модель на переменном токе, пригодная при малом уровне сигнала (линейная гибридная модель), а также динамическая (зарядовая) модель. Будут рассмотрены различные концепции, позволяющие описать физические явления в приборе, изучены частотные характеристики, а также явления, возникающие при высоком уровне инжекции. В конце главы кратко описаны особенности работы транзистора в режиме переключения, а также принципы действия тиристора.

6.1 Структура биполярного транзистора и принцип его работы

Для простоты будем изучать одномерную структуру, изображенную на рис. 6.1, а. Источники напряжений подключены таким образом, что усилительный прибор работает в активном режиме при нормальном включении: источник $U_{ЭБ}$ смещает переход эмиттер - база в прямом направлении, а источник $U_{КБ}$ смещает переход коллектор - база в обратном направлении (рис. 6.1,б). Возможны еще три режима, которые используются в переключательных устройствах: инверсный активный режим (аналогичный нормальному активному, но с взаимной переменой мест эмиттера и коллектора), режим отсечки (оба напряжения являются обратными) и, наконец, режим насыщения (оба перехода смещены в прямом направлении).

В активном нормальном режиме работа транзистора потенциал $U_{ЭБ}$ вызывает инжекцию электронов из эмиттера в область базы, которая располагается между границами $x_{БЭ}$ и $x_{БК}$ обедненных областей p - n -переходов (рис. 6.1). В активной области базы происходит диффузия электронов (Следует иметь в виду, что в базе так называемого дрейфового транзистора наряду с диффузией имеет место дрейф неосновных носителей под действием внутреннего поля). Некоторые электроны рекомбинируют с дырками, однако, большая их часть проходит область базы и достигает того участка,

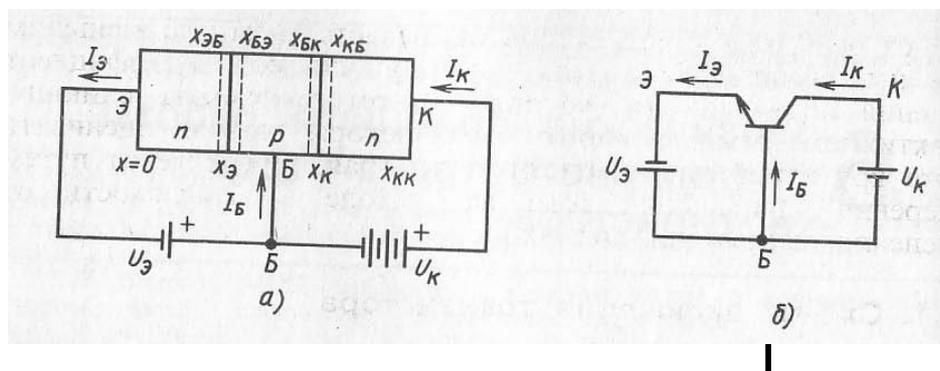


Рис. 6.1 Биполярный транзистор типа n - p - n . а – упрощенная модель транзистора, работающего в активной области (нормальный режим); область эмиттера легирована гораздо сильнее, чем область базы; б – условное графическое обозначение.

где источник напряжения $U_{КБ}$, включенный в обратном направлении, создает интенсивное электрическое поле, ускоряющее носители по направлению к коллектору. Чтобы этот процесс шел эффективно, активная область базы должна быть гораздо меньше диффузионной длины для электронов. Типичные значения таковы: толщина базы $W_B = x_{БК} - x_{БЭ} \approx 0,5$ мкм, диффузионная длина для электронов в базе $L_{нБ} \approx 10$ мкм.

Процесс усиления происходит следующим образом. Так как концентрация легирующих примесей в базе мала, инжекция дырок из базы в

эмиттер приводит к возникновению лишь небольшого дырочного тока, протекающего через вывод базы. С помощью этого малого тока можно управлять гораздо большим током коллектора (как уже отмечалось, работа транзистора основана на существовании носителей заряда обоих знаков). Поскольку напряжение $U_{КБ}$ является обратным, уровень импеданса, относящийся к этой части цепи, оказывается существенно выше того уровня, который связан с источником $U_{ЭБ}$. По этой причине транзистор является элементом цепи, создающим усиление по напряжению. Коэффициент передачи тока от эмиттера к коллектору оказывается немного меньшим единицы. Произведение этих двух величин есть коэффициент усиления по мощности, который может превышать единицу. В активном режиме работу транзистора можно оценивать также крутизной характеристики, которая определяется путем измерения приращения тока на выходе в зависимости от изменения напряжения на входе.

Схемы включения транзистора

В теории цепей различные устройства принято представлять в виде некоторых четырехполюсников с одним общим полюсом (выводом) (рис. 6.2). Если общим выводом является эмиттер, то принято говорить о схеме с общим эмиттером (ОЭ). Аналогично определяют схему с общей базой (ОБ) и общим коллектором (ОК).

Если эти схемы используются в цифровых устройствах, то выбирают две рабочие точки — соответствующие насыщению и отсечке. Режим транзистора соответствует либо одной, либо другой точке. Эти же схемы можно использовать для усиления слабых сигналов, обеспечив такое смещение, чтобы транзистор работал в нормальном (активном) режиме.

6.2. Параметры транзистора

В данном параграфе выводятся аналитические выражения для расчета параметров биполярного транзистора. Эти выражения важны для понимания внутренних процессов в приборе, которые рассмотрены в данной главе.

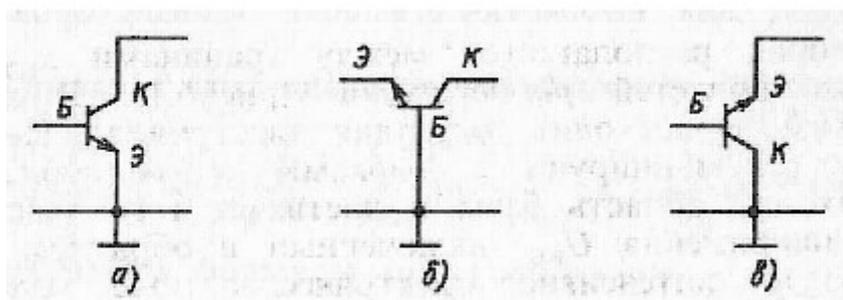


Рис. 6.2 Схемные изображения транзистора типа *n-p-n* в виде четырехполюсников:

а – схема ОЭ; б – схема ОБ; в – схема ОК

Параметры на постоянном токе

- Коэффициент инжекции, или эффективность эмиттера,

$$\gamma = \frac{I_{nЭ}}{I_{Э}} = \frac{I_{nЭ}}{I_{nЭ} + I_{pЭ}} \quad (6.1)$$

- Коэффициент переноса тока через базу

$$\beta_T = I_{nК} / I_{nЭ} \quad (6.2)$$

- Коэффициент умножения тока коллектора

$$M = I_K / I_{nК} \quad (6.3)$$

- Статический коэффициент передачи тока в схеме ОБ

$$\alpha_c = I_K / I_{Э}, \text{ при } U_{КБ} = 0 \quad (6.4)$$

Очевидно, что

$$\alpha_c = I_K / I_{Э} = \gamma \beta_T M \quad (6.5)$$

Поскольку $|I_{nК}| < |I_{nЭ}| < |I_{Э}|$, всегда $\alpha_c < 1$, хотя эта величина для высококачественных транзисторов близка единице.

- Статический коэффициент передачи тока в схеме ОЭ

$$B = h_{FE} = I_K / I_B, \text{ при } U_{КЭ} = 0 \quad (6.6)$$

Так как

$$I_{Э} = I_B + I_K \quad (6.7)$$

$$B = \alpha / (1 - \alpha) \quad (6.8)$$

Параметры на переменном токе (режим малого сигнала)

- Динамический коэффициент передачи тока в схеме ОБ

$$\alpha_D = \left. \frac{i_K}{i_{Э}} \right|_{U_{КБ}=0} = \left. \frac{i_K}{i_{Э}} \right|_{U_{КБ}=const} \quad (6.9)$$

- Динамический коэффициент передачи тока в схеме ОЭ

$$\beta = \left. \frac{i_K}{i_B} \right|_{U_{КЭ}=0} = \left. \frac{i_K}{i_{Э}} \right|_{U_{КЭ}=const} \quad (6.10)$$

Параметры α_D и β характеризуют качество транзистора. В типичном случае величина α_D лежит в пределах 0,990-0,998, чему соответствуют значения β от 100 до 500¹.

Аналитические выражения параметров транзистора на постоянном токе

¹ Заметим, что α и β - это коэффициенты передачи тока эмиттера и тока базы в одномерной теоретической модели транзистора. Эти же коэффициенты, а также элементы реального транзистора входят, в частности, и в физическую эквивалентную схему реального транзистора.

- Коэффициент инжекции, или эффективность эмиттера
В соответствии с формулой (6.1)

$$\gamma = \frac{I_{n\varepsilon}}{I_{\varepsilon}} = \frac{I_{n\varepsilon}}{I_{n\varepsilon} + I_{p\varepsilon}} = \left[1 + \frac{I_{p\varepsilon}}{I_{n\varepsilon}} \right]^{-1} \quad (6.11)$$

Приняв во внимание уравнение диода $I = I_{нас} (e^{U/U_T} - 1)$, имеем

$$I_{\varepsilon} = I_{\varepsilon нас} \left(\exp\left(\frac{U_{\varepsilon Б}}{U_T}\right) - 1 \right) \quad (6.12)$$

где

$$U_T = U_{kT}/q \quad (6.13)$$

$$I_{\varepsilon нас} = qn_{i0}^2 A \left[\frac{D_{n\delta}}{N_{a\delta} L_{n\delta} \text{th}(W_{\delta}/L_{n\delta})} + \frac{D_{p\varepsilon}}{N_{d\varepsilon} L_{p\varepsilon} \text{th}(W_{\varepsilon}/L_{p\varepsilon})} \right] \quad (6.14)$$

На основании формулы 6.11 получаем

$$\gamma = \left[1 + \frac{D_{p\varepsilon} N_{a\delta} L_{n\delta} \text{th}(W_{\delta}/L_{n\delta})}{D_{n\delta} N_{d\varepsilon} L_{p\varepsilon} \text{th}(W_{\varepsilon}/L_{p\varepsilon})} \right]^{-1}. \quad (6.15)$$

Как видно из равенства (6.15), наибольшее влияние на коэффициент инжекции оказывает концентрация примесей $N_{a\delta}$ и $N_{d\varepsilon}$. Если $N_{d\varepsilon} > N_{a\delta}$, то этот коэффициент близок единице. На практике для кремния максимальное значение $N_{d\varepsilon}$ имеет порядок 10^{21} см^{-3} , в то время как минимальное значение $N_{a\delta}$ имеет порядок 10^{12} см^{-3} . При этих крайних значениях концентраций эффективность эмиттера составляет 0,99999. В реальных приборах, у которых $N_{d\varepsilon} > 10^{18} \text{ см}^{-3}$, эффективность эмиттера несколько меньше, но все же она достаточно близка единице.

- Коэффициент переноса тока через базу

Принимая во внимание формулу $J_p(x) = -qD_p \frac{dp_p(x)}{dx}$, в которой отброшена составляющая с задержкой, для области базы имеем соотношение

$$I_n(x) = qAD_{n\delta} \frac{dn'(x)}{dx} \quad (6.16)$$

Согласно следующим равенствам,

$$n'(x) = n_{0B} (e^{U_{\varepsilon\delta}/U_T} - 1)$$

$$n'(x) = n_{0B} (e^{U_{\varepsilon\delta}/U_T} - 1)$$

Так как $|U_{KB}| \gg 0$ и $n'(x_{\delta\kappa}) \approx -n_{0B}$ то $n(x_{BK}) \approx 0$. Принимая это во внимание и используя формулу $n'_p(x) = n_{p0} (e^{U/U_T} - 1)e^{-x/L_n}$, получаем

$$n'(x) = \frac{n_{0\delta} \text{sh}[(x_{\delta\kappa} - x)/L_{n\delta}]}{\text{sh}(W_{\delta}/L_{n\delta})} (e^{U_{\varepsilon\delta}/U_T} - 1) \quad (6.17)$$

откуда

$$\frac{dn'(x)}{dx} = -\frac{n_{0\delta} \operatorname{ch}[(x_{\delta k} - x)/L_{n\delta}]}{\operatorname{ch}(W_{\delta}/L_{n\delta})} (e^{U_{БЭ}/U_E} - 1) \quad (6.18)$$

Подставляя эти равенства в формулу (6.16), с учетом (6.2) находим следующее выражение для коэффициента переноса тока через базу:

$$\beta_T = \frac{qD_{n\delta} \left. \frac{dn'}{dx} \right|_{x_{\delta k}}}{qD_{n\delta} \left. \frac{dn'}{dx} \right|_{x_{\delta\delta}}} = \frac{\operatorname{ch}(0)}{\operatorname{ch} \frac{W_{\delta}}{L_{n\delta}}} = \frac{1}{\operatorname{ch} \frac{W_{\delta}}{L_{n\delta}}} \quad (6.19)$$

Представим гиперболический косинус в виде разложения в ряд Тейлора:

$$\operatorname{ch} z = \frac{e^z + e^{-z}}{2} = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \dots \quad (6.20)$$

Так как $W_b \ll L_{n\delta}$, то имеет место приближенная формула, справедливая при условии, что примеси в базе распределены равномерно:

$$\beta_T \approx \frac{1}{1 + W_{\delta}^2 / (2L_{n\delta}^2)} \approx 1 - \frac{W_{\delta}^2}{2L_{n\delta}^2} \quad (6.21)$$

Здесь толщина базы W_{δ} зависит от напряжения смещения (эффект Эрли). В базе реального прибора примеси распределяются неоднородно. Можно показать, что при этом:

$$\beta_T \approx 1 - \frac{1}{L_{n\delta}^2} \int_0^{W_{\delta}} \left(\frac{1}{N_{a\delta}} \int_{x_{\delta\delta}}^{x_{\delta k}} N_{a\delta} dx \right) dx \quad (6.22)$$

- Коэффициент умножения тока коллектора

Этот параметр связан с пробоем р-n-перехода, т. е. лавинным процессом, и выражается соотношением:

$$M = \left[1 - \left(\frac{|U_{КБ}|}{U_{КБпр\delta}} \right)^a \right]^{-1} \quad (6.23)$$

Здесь показатель a может принимать значения в интервале 3-6; $U_{КБ}$ - напряжение смещения на переходе, не превышающее напряжение лавинного пробоя; $U_{КБпр\delta}$ - напряжение лавинного пробоя перехода база - коллектор. В формуле (6.3) следовало бы учесть также составляющую тока $I_{рК}$, однако, можно показать, что она пренебрежимо мала (имеются в виду кремниевые и германиевые транзисторы с сильно легированной областью коллектора).

- Статический коэффициент передачи тока в схеме ОБ

$$\alpha_c = \gamma\beta_T M = \left[1 + \frac{D_{p3} L_{n\bar{b}} N_{a\bar{b}} \operatorname{th}(W_{\bar{\sigma}} / L_{n\bar{b}})}{D_{n\bar{b}} L_{p3} N_{d3} \operatorname{th}(W_3 / L_{p3})} \right]^{-1} \times$$

$$\times \left(\frac{1}{\operatorname{ch} W_{\bar{\sigma}} / L_{n\bar{b}}} \right) \left[1 - \left(\frac{|U_{KB}|}{U_{KBпр\bar{b}}} \right)^a \right]^{-1} \quad (6.24)$$

В нормальном активном режиме работы высококачественного транзистора всегда $\gamma \leq 1$; $\beta_T \leq 1$ и $M \geq 1$. Если значение $|U_{KB}|$ приближается к $U_{KBпр\bar{b}}$, то коэффициент α_c становится больше единицы, что ведет к нестабильности работы прибора. Действительно, если в формуле (6.8) положить $a=1$, то величина β становится неограниченно большой.

Для многих современных транзисторов можно считать, что примеси распределены по объему равномерно, а полупроводник (кремний) в эмиттерной области является невырожденным.

При этом

$$B \approx \frac{D_{n\bar{b}} L_{p3} N_{d3} \operatorname{th}(W_3 / L_{p3})}{D_{p3} L_{n\bar{b}} N_{a\bar{b}} \operatorname{th}(W_{\bar{\sigma}} / L_{n\bar{b}})} \quad (6.25)$$

6.3. Частотные свойства транзистора

На низких частотах коэффициенты передачи тока α_d и β_d , определяемые в режиме малого сигнала и задаваемые формулами (6.9) и (6.10), не зависят от значения рабочей частоты. Данные величины начинают уменьшаться, лишь начиная с некоторой критической (предельной) частоты. Для функции $\alpha_d(j\omega)$ можно воспользоваться следующим приближенным выражением:

$$\alpha_d(j\omega) = \frac{\alpha}{1 + j\omega / \omega_\alpha} \quad (6.26)$$

где α - коэффициент передачи тока в области низких частот, ω_α - предельная частота (частота отсечки), определяемая в соответствии с формулой

$$\omega_\alpha = 2\pi f_\alpha \approx \pi^2 D_{n\bar{b}} / (4W_{\bar{\sigma}}^2) \quad (6.27)$$

Функция $\alpha_d(j\omega)$ зависит от ряда факторов. Обобщающим фактором служит параметр $\tau_{\bar{\sigma}}$ - время запаздывания сигнала, распространяющегося между эмиттером и коллектором.

Время пролета носителей через базу

На ранних этапах развития полупроводниковой электроники значение параметра $\tau_{\bar{\sigma}}$ в наибольшей степени лимитировало работу транзисторов. Применительно к современным ИС существует ряд подобных временных параметров, которые следует принимать во внимание.

Рассмотрим приближенный аналитический способ нахождения времени запаздывания сигнала в базе. Ток, обусловленный потоком электронов в базе,

перемещающихся от эмиттера к коллектору вследствие диффузии, определяется равенством

$$I_{n\bar{\sigma}}(x) = qAD_{n\bar{\sigma}} \frac{dn'(x)}{dx} \quad (6.28)$$

Если $W_{\bar{\sigma}} \ll L_{n\bar{\sigma}}$, то закон распределения концентрации избыточных носителей приблизительно таков:

$$n'(x) \approx n'(x_{\bar{\sigma}3}) \left(1 - \frac{x - x_{\bar{\sigma}3}}{x_{\bar{\sigma}k} - x_{\bar{\sigma}3}} \right) \quad (6.29)$$

Далее, используя формулу (6.28), получаем

$$I_{n\bar{\sigma}}(x) \approx -qAD_{n\bar{\sigma}} \frac{n'(x_{\bar{\sigma}3})}{W_{\bar{\sigma}}} \quad (6.30)$$

С другой стороны, известно, что

$$I_{n\bar{\sigma}}(x) = -qAn'(x)v(x) \quad (6.31)$$

где $v(x)$ - скорость носителей заряда в точке с координатой x . Приравнявая выражения (6.30) и (6.31), находим

$$v(x) = \frac{D_{n\bar{\sigma}}}{[W_{\bar{\sigma}} - (x - x_{\bar{\sigma}3})]}$$

Поскольку $dx = v(x)dt$, время, которое требуется электрону для того, чтобы пересечь область базы,

$$\tau_{\bar{\sigma}} = \int_0^{W_{\bar{\sigma}}} \frac{W_{\bar{\sigma}} - (x - x_{\bar{\sigma}3})}{D_{n\bar{\sigma}}} dx = \frac{W_{\bar{\sigma}}^2}{2D_{n\bar{\sigma}}} \quad (6.32)$$

Сравнивая выражения (6.32) и (6.27), находим

$$\varpi_{\alpha} = 2\pi f_{\alpha} = \pi^2 / (8\tau_{\bar{\sigma}}) \quad (6.33)$$

Отсюда видно, что при работе на высоких частотах следует уменьшать время пролета носителей через базу, уменьшая толщину базовой области.

6.4. Явления, наблюдаемые при высоком уровне инжекции

Вспомним, что уровень инжекции называют низким, если $n \ll p_0$; $p \approx p_0$; $p \gg n$. При высоком уровне инжекции $n' \approx p' \approx p$. Условия, характерные для высокого уровня инжекции, возникают при увеличении напряжений на транзисторе, работающем в активном режиме. При этом наблюдаются два принципиально важных явления: уменьшение эффективности эмиттера и рост коэффициента переноса. Первое явление влияет на характеристики транзистора существенно сильнее, чем второе. В результате коэффициент усиления по току падает в такой же мере, в какой возрастает уровень инжекции.

- Влияние эффективности эмиттера

Инжекция высокого уровня наблюдается, когда ток $I_{p(\text{ЭБ})}$ составляет существенную часть тока I_3 . Именно по этой причине эффективность эмиттера уменьшается.

- Влияние коэффициента диффузии

В результате инжекции носителей в область базы возникает некоторое электрическое поле напряженностью ε . При высоком уровне инжекции это поле влияет на движение неосновных носителей в базе. В данном случае плотность тока $J_{nБ} \approx 0$, и, согласно формуле

$$J_p = q\mu_p p\varepsilon - qD_p \frac{dp}{dx} :$$

$$\varepsilon = U_T \frac{1}{p} \frac{dp}{dx} \quad (6.34)$$

На основании $J_n = q\mu_n n\varepsilon + qD_n \frac{dn}{dx}$ имеем

$$J_{n\bar{0}} = q\mu_{n\bar{0}} n\varepsilon + qD_{n\bar{0}} \frac{dn}{dx} \quad (6.35)$$

Откуда

$$J_{n\bar{0}} = q\mu_{n\bar{0}} n U_T \frac{1}{p} \frac{dp}{dx} + qD_{n\bar{0}} \frac{dn}{dx} \quad (6.36)$$

Эту формулу удобно представить в виде

$$J_{n\bar{0}} = qD_{n\bar{0}} \left(1 + \frac{n}{p} \right) \frac{dn}{dp} \quad (6.37)$$

Здесь использовано соотношение Эйнштейна $D_{n\bar{0}} = U_T / \mu_{n\bar{0}}$, а также то, что в области базы $dp / dx \approx dn / dx$.

При низком уровне инжекции ($n \ll p$)

$$J_{n\bar{0}} = q(D_{n\bar{0}}) \frac{dn}{dx} \quad (6.38)$$

При высоком уровне инжекции ($n \approx p$)

Таким образом, повышение уровня инжекции приводит к росту коэффициента диффузии. В соответствии с формулой (6.19) при этом увеличивается коэффициент переноса зарядов через базу, а на основании равенства (6.32) сокращается время распространения.

6.5. Статистическая модель биполярного транзистора

В настоящем параграфе, а также в § 6.6 и 6.7 строятся простые, но весьма важные модели, "пригодные для описания свойств транзистора на постоянном токе, такие, как модель для большого сигнала и зарядовая модель. Обе они используются при анализе логических устройств. Изучается также модель на переменном токе (модель малого сигнала), применяемая в теории аналоговых цепей.

Модель для большого сигнала, количественно отражающая физические процессы в транзисторе, основывается на уравнении диода.

Упрощения и гипотезы

Продолжим описание, начатое ранее при исследовании свойств диода. Используем прежние ограничения и предположения, добавив ряд новых.

1. Площадь поперечного сечения A постоянна в любом сечении прибора.
2. Концентрация носителей зависит только от координаты x .
3. Концентрация легирующей примеси считается постоянной в каждой области, и поэтому р-п-переходы считаются резкими.
4. Концентрация примесей в эмиттере достаточно высока, однако вырождение не наблюдается ($n_{э0} > 100 p_{б0}$).
5. Уровень инжекции является низким.
6. Ни на внешней поверхности полупроводника, ни в обеих обедненных областях процессы рекомбинации не происходят.
7. Толщина активной области базы $W_б$ постоянна, т. е. эффект Эрли, связанный с модуляцией толщины базы под действием приложенных напряжений $U_{ЭБ}$ и $U_{КБ}$, не наблюдается.
8. Параметр $W_б$ мал по сравнению с диффузионной длиной $L_{нб}$ неосновных носителей в области базы ($W_{нб} < 0,3 L_{нб}$).
9. Пространственные размеры областей эмиттера и коллектора значительно превышают соответствующие диффузионные длины для неосновных носителей.
10. Возможность возникновения лавинного процесса (эффект Зенера) при подаче обратных смещений не принимается во внимание.

Распределение концентрации носителей заряда в области базы, эмиттера и коллектора

Будем считать, что точка начала отсчета ($x=0$) расположена в центре базовой области (см. рис. 6.1). Полагая, что толщина активной области базы $W_б = const$, обозначим символом $x_{тэ}$ толщину обедненной области перехода база - эмиттер, а символом $x_{тк}$ - толщину обедненной области перехода база - коллектор.

База. Неосновные носители — электроны

Уравнение непрерывности

$$\frac{d^2 n'_{p\delta}(x)}{dx^2} - \frac{n'_{p\delta}(x)}{L_{n\delta}^2} = 0 \quad (6.40)$$

где $L_{n\delta}^2 = D_{n\delta}\tau_n$.

Общее решение уравнения (6.40) имеет вид

$$n'_{p\delta}(x) = C_1 \operatorname{sh} \frac{W_\delta/2 - x}{L_{n\delta}} + C_2 \operatorname{sh} \frac{W_\delta/2 + x}{L_{n\delta}} \quad (6.41)$$

Граничные условия вытекают из выражения $\frac{p_n(x_n)}{p_{n0}(x_n)} = \frac{n_p(-x_p)}{n_{p0}(-x_p)} = \exp\left(\frac{U}{U_T}\right)$,

описывающего свойства перехода:

$$n_{\delta\alpha} = n_{p\delta}\left(x = -\frac{W_\delta}{2}\right) = n_{0\delta} \exp\left(\frac{U_{\text{ЭБ}}}{U_T}\right) \quad (6.42)$$

$$n_{\delta\kappa} = n_{p\delta}\left(x = +\frac{W_\delta}{2}\right) = n_{0\delta} \exp\left(\frac{U_{\text{КБ}}}{U_T}\right) \quad (6.43)$$

Применив эти равенства к формуле (6.41), получаем

$$n_{p\delta}\left(-\frac{W_\delta}{2}\right) - n_{0\delta} = n_{0\delta} \left[\exp\left(\frac{U_{\text{ЭБ}}}{U_T}\right) - 1 \right] = C_1 \operatorname{sh} \frac{W_\delta}{L_{n\delta}} \quad (6.44)$$

$$n_{p\delta}\left(+\frac{W_\delta}{2}\right) - n_{0\delta} = n_{0\delta} \left[\exp\left(\frac{U_{\text{КБ}}}{U_T}\right) - 1 \right] = C_2 \operatorname{sh} \frac{W_\delta}{L_{n\delta}} \quad (6.45)$$

Подставляя значения C_1 и C_2 , определяемые формулами (6.44) и (6.45), в выражение (6.41), находим

$$\begin{aligned} n'_{p\delta}(x) &= n_{p\delta} - n_{0\delta} \approx \\ &\approx \frac{n_{0\delta}}{W_\delta/L_{n\delta}} \left\{ \left[\exp\left(\frac{U_{\text{ЭБ}}}{U_T}\right) - 1 \right] \left(\frac{W_\delta/2 - x}{L_{n\delta}} \right) + \right. \\ &\left. + \left[\exp\left(\frac{U_{\text{КБ}}}{U_T}\right) - 1 \right] \left(\frac{W_\delta/2 + x}{L_{n\delta}} \right) \right\} \end{aligned} \quad (6.46)$$

что справедливо для любого значения x в пределах $-W_\delta/2 \leq x \leq W_\delta/2$

Эмиттер. Неосновные носители — дырки

Аналогично получаем уравнение непрерывности

$$\frac{d^2 p'_{n3}(x)}{dx^2} - \frac{p'_{n3}(x)}{L_{p3}^2} = 0 \quad (6.47)$$

Общее решение этого уравнения

$$p'_{n3}(x) = C_1 \exp(x/L_{p3}) + C_2 \exp(-x/L_{p3})$$

Используя граничные условия, имеем

$$p'_{n3}(-\infty) = 0 \Rightarrow C_2 = 0 \quad (6.48)$$

$$p'_{n3}\left(-\frac{W_6}{2} - x_{t3}\right) = p_{03} \left(e^{U_{ЭБ}/U_T} - 1 \right) \quad (6.49)$$

$$p_{03} \left[\exp\left(\frac{U_{ЭБ}}{U_T}\right) - 1 \right] = C_1 \exp\frac{-W_6/2 - x_{t3}}{L_{p3}} \quad (6.50)$$

$$p_{n3}(x) = p_{n3} - p_0 \left[\exp\left(\frac{U_{ЭБ}}{U_T}\right) - 1 \right] \times \\ \times \exp\frac{x + W_6/2 + x_{t3}}{L_{p3}} \quad (6.51)$$

Коллектор. Неосновные носители — дырки

Рассуждая аналогично, получаем

$$p'_{nk}(x) = p_{0k} \left[\exp\left(\frac{U_{КБ}}{U_T}\right) - 1 \right] \exp\left[-\frac{1}{L_{pk}} \left(x - \frac{W_6}{2} - x_{tk} \right) \right] \quad (6.52)$$

Модель Эберса — Мола

В соответствии с рис. 6.1 имеем

$$I_{Э} = I_{nЭ} + I_{pЭ} \quad (6.53)$$

$$-I_R = I_{nR} + I_{pR} \quad (6.54)$$

Ток эмиттера

$$I_{Э} = I_{nЭ} + I_{pЭ} = AJ_{dnЭ} \left(-\frac{W_6}{2} \right) + AJ_{dpp} \left(-\frac{W_6}{2} - x_{tЭ} \right) \quad (6.55)$$

На основании формулы $J_n = q\mu_n n \varepsilon + qD_n \frac{dn}{dx}$ можно записать

$$J_{dnЭ}(x) = \frac{qD_n n_{03}}{L_{n3}} \frac{1}{\text{sh}(W_6/L_{n3})} \left[-\left(\exp\left(\frac{U_{ЭБ}}{U_T}\right) - 1 \right) \text{ch}\frac{W_6/2 - x}{L_{n3}} + \left(\exp\left(\frac{U_{РБ}}{U_T}\right) - 1 \right) \text{ch}\frac{W_6/2 + x}{L_{n3}} \right] \quad (6.56)$$

Положив $x = -W_6/2$, $n_{03} = n_i^2/N_a$ и воспользовавшись разложением в ряд (при $z \ll 1$)

$$\text{sh}z = \frac{e^z - e^{-z}}{2} = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \dots$$

Получим

$$J_{dn\mathcal{E}}\left(-\frac{W_{\delta}}{2}\right) \approx \frac{qD_n n_i^2}{N_a W_{\delta}} \left[-\left(\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right) \operatorname{ch} \frac{W_{\delta}}{L_{n\mathcal{E}}} + \left(\exp\left(\frac{U_{RB}}{U_T}\right) - 1 \right) \operatorname{ch} 0 \right] \quad (6.57)$$

Выразив функцию $\operatorname{ch} z$ через экспоненциальные слагаемые, получим

$$J_{dn\mathcal{E}}\left(-\frac{W_{\delta}}{2}\right) \approx \frac{qD_n n_i^2}{N_a W_{\delta}} \left[-\left(\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right) + \left(\exp\left(\frac{U_{RB}}{U_T}\right) - 1 \right) \right] \quad (6.58)$$

На основании формулы $1 - f(E) \approx e^{E_F/kT} e^{E/(kT)}$

$$J_{dp\mathcal{E}}(x) = -\frac{qD_p p_{0\mathcal{E}}}{W_{\mathcal{E}}} \left[\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right] \exp\left(\frac{x + W_{\delta}/2 + x_{t\mathcal{E}}}{L_{p\mathcal{E}}}\right) \quad (6.59)$$

$$J_{dp\mathcal{E}}\left(-\frac{W_{\delta}}{2} - x_{t\mathcal{E}}\right) = -\frac{qD_p n_i^2}{N_{d\mathcal{E}} W_{\mathcal{E}}} \left[\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right] \quad (6.60)$$

Подставляя (6.58) и (6.60) в формулу (6.55), находим

$$I_{\mathcal{E}} = \frac{qAD_n n_i^2}{N_a W_{\delta}} \left[-\left(\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right) + \left(\exp\left(\frac{U_{RB}}{U_T}\right) - 1 \right) \right] - \frac{qAD_p n_i^2}{N_{d\mathcal{E}} L_{p\mathcal{E}}} \left[\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right] \quad (6.61)$$

$$I_{\mathcal{E}} = -qAn_i^2 \left(\frac{D_n}{N_a W_{\delta}} + \frac{D_p}{N_{d\mathcal{E}} W_{\mathcal{E}}} \right) \left(\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right) + \frac{qAD_n n_i^2}{N_a W_{\delta}} \left[\exp\left(\frac{U_{RB}}{U_T}\right) - 1 \right] \quad (6.62)$$

Ток коллектора

$$-I_K = I_{nK} + I_{pK} = AJ_{dnK} \left(-\frac{W_{\delta}}{2} \right) + AJ_{dpK} \left(\frac{W_{\delta}}{2} + x_{tk} \right) \quad (6.63)$$

Используя тот же метод, получаем

$$I_K = \frac{qAD_n n_i^2}{N_a W_{\delta}} \left[\exp\left(\frac{U_{\mathcal{E}B}}{U_T}\right) - 1 \right] - qAn_i^2 \left(\frac{D_n}{N_a W_{\delta}} + \frac{D_p}{N_{dK} L_{pK}} \right) \left(\exp\left(\frac{U_{KB}}{U_T}\right) - 1 \right) \quad (6.64)$$

Уравнения (6.62) и (6.64) описывают модель Эберса — Молла. Применительно к формуле (6.63) введем следующие параметры:

$$a_{11} = -qAn_i^2 \left(\frac{D_n}{N_a W_{\delta}} + \frac{D_p}{N_{d\mathcal{E}} W_{\mathcal{E}}} \right), \quad a_{12} = \frac{qAD_n n_i^2}{N_a W_{\delta}} \quad (6.65)$$

На основании (6.64) введем также параметры

$$a_{12} = \frac{qAD_n n_i^2}{N_a W_{\delta}}, \quad a_{22} = -qAn_i^2 \left(\frac{D_n}{N_a W_{\delta}} + \frac{D_p}{N_{dK} L_{pK}} \right) \quad (6.66)$$

Это позволяет представить модель Эберса - Молла в более компактном виде:

$$I_{\mathcal{E}} = a_{11}(e^{U_{\mathcal{EБ}}/U_T} - 1) + a_{12}(e^{U_{\mathcal{КБ}}/U_T} - 1) \quad (6.67)$$

$$I_{\mathcal{К}} = a_{21}(e^{U_{\mathcal{ЭБ}}/U_T} - 1) + a_{22}(e^{U_{\mathcal{КБ}}/U_T} - 1) \quad (6.68)$$

Можно заметить, что $a_{12} = a_{21}$ независимо от геометрии прибора. Данная модель справедлива при всех четырех возможных комбинациях знаков напряжений $U_{\mathcal{ЭБ}}$ и $U_{\mathcal{КБ}}$. Каждый из этих случаев соответствует одному из возможных режимов работы транзистора: активному (при нормальном и инверсном включении), режиму отсечки, а также режиму насыщения. Модель Эберса - Молла для $p-n-p$ -транзистора выражается следующими равенствами:

$$\begin{aligned} I_{\mathcal{Э}} &= a_{11}(e^{U_{\mathcal{ЭБ}}/U_T} - 1) - a_{12}(e^{U_{\mathcal{КБ}}/U_T} - 1) \\ I_{\mathcal{К}} &= -a_{21}(e^{U_{\mathcal{ЭБ}}/U_T} - 1) + a_{22}(e^{U_{\mathcal{КБ}}/U_T} - 1) \end{aligned}$$

Чем они отличаются от уравнений (6.67) и (6.68)?

Эквивалентная схема для модели Эберса - Молла

Исключив величину $U_{\mathcal{ЭБ}}$ из уравнений (6.67) и (6.68), получаем

$$I_{\mathcal{К}} = \frac{a_{21}}{a_{11}} I_{\mathcal{Э}} - \left(a_{22} - \frac{a_{21}a_{12}}{a_{11}} \right) (e^{U_{\mathcal{КБ}}/U_T} - 1) \quad (6.69)$$

На основании формул (6.4) и (6.12) можно записать:

$$I_{\mathcal{К}} = \alpha_F I_{\mathcal{Э}} + I_{\mathcal{Кнас}} (e^{U_{\mathcal{КБ}}/U_T} - 1) \quad (6.70)$$

где $I_{\mathcal{Кнас}}$ - ток насыщения перехода база - коллектор при обрыве цепи эмиттера ($I_{\mathcal{Э}}=0$); α_F - коэффициент передачи тока и схеме ОБ в нормальном активном режиме работы.

Если теперь из формул (6.67) и (6.68) исключить напряжение $U_{\mathcal{КБ}}$, то получим

$$I_{\mathcal{Э}} = \frac{a_{12}}{a_{22}} I_{\mathcal{К}} - \left(a_{11} - \frac{a_{12}a_{21}}{a_{22}} \right) (e^{U_{\mathcal{ЭБ}}/U_T} - 1) \quad (6.71)$$

По аналогии с (6.70) запишем эту формулу так:

$$I_{\mathcal{К}} = \alpha_R I_{\mathcal{Э}} + I_{\mathcal{Энас}} (e^{U_{\mathcal{ЭБ}}/U_T} - 1) \quad (6.72)$$

$I_{\mathcal{Энас}}$ - обратный ток насыщения перехода база - эмиттер; α_R - коэффициент передачи тока в схеме ОБ в активном инверсном режиме (электроны испускаются коллектором при оборванной цепи эмиттера).

Так как

$$\alpha_F = \left. \frac{I_{\mathcal{К}}}{I_{\mathcal{Э}}} \right|_{U_{\mathcal{КБ}}=const} = \frac{a_{21}}{a_{11}} \quad (6.73)$$

$$\alpha_R = \left. \frac{I_{\mathcal{К}}}{I_{\mathcal{Э}}} \right|_{U_{\mathcal{ЭБ}}=const} = \frac{a_{12}}{a_{22}} \quad (6.74)$$

можно записать

$$I_{\mathcal{Кнас}} = -a_{22} + a_{21}a_{12} / a_{11} \quad (6.75)$$

$$I_{ЭНАС} = a_{11} - a_{12} a_{21} / a_{22} \quad (6.76)$$

Легко доказать, что

$$\alpha_F I_{ЭНАС} \approx \alpha_R I_{КНАС} \quad (6.77)$$

Эта формула позволяет определить один неизвестный параметр транзистора, если три других заданы. Подчеркнем, что параметры транзистора на постоянном токе непосредственно связаны с коэффициентом диффузии, диффузионной длиной, концентрацией неосновных носителей, толщиной базы и площадью поперечного сечения.

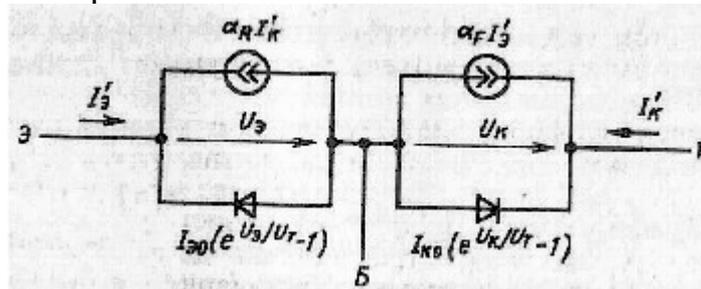


Рис. 6.3 Эквивалентная схема $n-p-n$ -транзистора, соответствующая модели Эберса-Молла

Представляя модель Эберса - Молла уравнениями (6.70) и (6.72), можно прийти к эквивалентной схеме, изображенной на рис. 6.3. Схема состоит из двух идеальных диодов, включенных встречно и подчиняющихся уравнениям Шотки с токами насыщения $I_{Энас}$ и $I_{Кнас}$, а также двух источников тока, наличие которых обусловлено переносом неосновных носителей через нейтральную область базы.

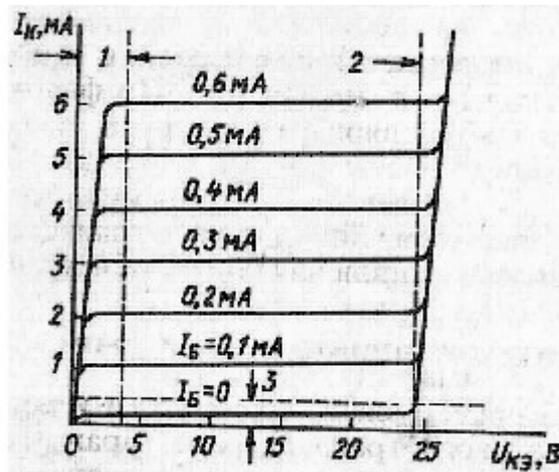


Рис. 6.4 Выходные характеристики $I_K=f(U_{КЭ})$ транзистора, включенного по схеме ОЭ: 1 – насыщения; 2 – реактивный режим; 3 – отсечный режим

Модель Эберса-Молла можно использовать для описания характеристик транзистора в режиме большого сигнала в схеме ОЭ (рис. 6.4). Эту модель целесообразно применять в том случае, если зависимости коэффициентов передачи α_F и α_R от частоты известны. Строго говоря, на очень высоких частотах модель Эберса -Молла не справедлива, поскольку она не может учесть изменения величины $n_0(x, t)$. Обычно пользуются формулами, которые вытекают из (6.26):

$$\alpha_F(\omega_\alpha) = \frac{\alpha_{F0}}{1 - \omega_\alpha / \omega_F} \quad (6.79)$$

$$\alpha_R(\omega_\alpha) = \frac{\alpha_{R0}}{1 - \omega_\alpha / \omega_R} \quad (6.80)$$

где параметр ω_α определяют согласно формуле (6.27); ω_F и ω_R - частоты отсечки для активного нормального и активного инверсного режимов соответственно.

6.6. Модели транзистора в режиме малого сигнала

При анализе работы биполярного транзистора в качестве усилительного прибора особый интерес представляет случай, когда напряжение база - эмиттер изменяется во времени периодически. Если амплитуда этого напряжения достаточно мала по сравнению с величиной U_T , то говорят, что транзистор работает в режиме малого сигнала (подчеркнем, что такой режим является динамическим). Для описания свойств транзистора в таком режиме принято использовать одну или несколько эквивалентных схем.

Любой биполярный транзистор, работающий в динамическом режиме, может быть представлен в виде некоторого четырехполюсника (рис. 6.5), для которого имеют место два уравнения, связывающие между собой четыре физические величины:

$$f_1(i_1, i_2, u_1, u_2) = 0, \quad f_2(i_1, i_2, u_1, u_2) = 0 \quad (6.81)$$

В данных уравнениях две переменные могут быть независимые, а две другие - выражаться через них.

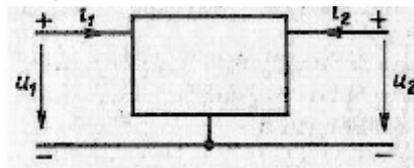


Рис. 6.5. Биполярный транзистор как четырехполюсник

При этом потребуется четыре коэффициента, играющих роль независимых параметров. Эти параметры определяют в соответствии с физическими принципами работы прибора. Необходимо ясно себе представлять, что эквивалентная схема описывает транзистор как некоторый «черный ящик» и поэтому не имеет непосредственного отношения к внутренним процессам в нем.

Число возможных эквивалентных схем транзистора соответствует числу параметров в режиме малого сигнала. Трех наиболее часто используемым видам соотношений между переменными i_1 , i_2 , u_1 , u_2 отвечают так называемые z -параметры, y -параметры и гибридные h -параметры. Система z -параметров задается уравнениями

$$\begin{aligned} u_1 &= z_{11}i_1 + z_{12}i_2 \\ u_2 &= z_{21}i_1 + z_{22}i_2 \end{aligned} \Rightarrow \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{pmatrix} \begin{pmatrix} i_1 \\ i_2 \end{pmatrix} \quad (6.82)$$

где

$z_{11} = \left. \frac{u_1}{i_1} \right|_{i_2=0}$ - входное сопротивление в режиме холостого хода на выходе по переменному току;

$z_{12} = \left. \frac{u_1}{i_2} \right|_{i_1=0}$ - сопротивление обратной связи в режиме холостого хода на входе по переменному току;

$z_{21} = \left. \frac{u_2}{i_1} \right|_{i_2=0}$ - сопротивление прямой передачи в режиме холостого хода на выходе;

$z_{22} = \left. \frac{u_2}{i_2} \right|_{i_1=0}$ - выходное сопротивление в режиме холостого хода на выходе по переменному току.

Аналогично, y -параметры транзистора определяются на основании системы уравнений:

$$\begin{aligned} i_1 &= y_{11}u_1 + y_{12}u_2 \\ i_2 &= y_{21}u_1 + y_{22}u_2 \end{aligned} \Rightarrow \begin{pmatrix} i_1 \\ i_2 \end{pmatrix} = \begin{pmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad (6.83)$$

в которой

$y_{11} = \left. \frac{i_1}{u_1} \right|_{u_2=0}$ - входная проводимость в режиме короткого замыкания на выходе по переменному току;

$y_{12} = \left. \frac{i_1}{u_2} \right|_{u_1=0}$ - проводимость обратной связи в режиме короткого замыкания на выходе по переменному току;

$y_{21} = \left. \frac{i_2}{u_1} \right|_{u_2=0}$ - проводимость прямой передачи в режиме короткого замыкания на входе по переменному току;

$y_{22} = \left. \frac{i_2}{u_2} \right|_{u_1=0}$ - выходная проводимость в режиме короткого замыкания на входе по переменному току

Таким же образом вводят систему гибридных h -параметров транзистора:

$$\begin{aligned} u_1 &= h_{11}i_1 + h_{12}u_2 \\ i_2 &= h_{21}i_1 + h_{22}u_2 \end{aligned} \Rightarrow \begin{pmatrix} u_1 \\ i_2 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} \begin{pmatrix} i_1 \\ u_2 \end{pmatrix} \quad (6.84)$$

где

$h_{11} = \left. \frac{u_1}{i_1} \right|_{u_2=0}$ - входное сопротивление в режиме короткого замыкания на входе по переменному току;

$h_{12} = \left. \frac{u_1}{u_2} \right|_{i_1=0}$ - коэффициент обратной связи в режиме холостого хода на входе по переменному току; показывает, какая доля выходного напряжения за счет обратной связи поступает на вход транзистора;

$h_{21} = \left. \frac{i_2}{i_1} \right|_{u_2=0}$ - коэффициент передачи тока в режиме короткого замыкания на выходе по переменному току (безразмерная величина);

$h_{22} = \left. \frac{i_2}{u_2} \right|_{i_1=0}$ - выходная проводимость в режиме холостого хода на входе по переменному току.

Для описания транзистора в режиме малого сигнала с помощью эквивалентной схемы чаще всего используют гибридные h -параметры, называемые так потому, что здесь физические размерности отдельных параметров оказываются неодинаковыми.

Поскольку перечисленные системы параметров не являются независимыми, достаточно измерить параметры лишь одной из групп, а затем пересчитать через нее все остальные параметры.

Если обратиться к эквивалентной схеме, соответствующей системе h -параметров (рис. 6.6), следует учесть, что первое уравнение системы (6.83) выражает входное напряжение в виде суммы двух составляющих, которые зависят от i_1 или от u_2 . Отсюда следует, что данному уравнению соответствует последовательная цепь; аналогично, второму уравнению соответствует параллельная цепь, так как ток i_2 есть сумма двух составляющих.

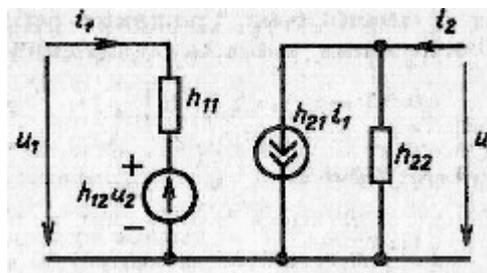


Рис 6.6. Т-образная эквивалентная схема биполярного транзистора в системе h – параметров в режиме малого сигнала

Любую последовательную цепь можно преобразовать в параллельную, воспользовавшись выражениями $1/h_{11}=y_{11}$; $-h_{12}/h_{11}=y_{12}$, которые доказываются аналитически на основании уравнений состояния четырехполюсника.

Гибридная П-образная эквивалентная схема транзистора на низких и высоких частотах

Наиболее распространенная эквивалентная схема транзистора в режиме малого сигнала называется гибридной П-образной схемой. Эта схема выгодно отличается тем, что входящие в нее параметры легко измерить и связать с физическими процессами в транзисторе. Предполагая, что транзистор включен по схеме ОЭ и работает в активном режиме, рассмотрим П-образные модели для области как низких, так и высоких частот.

Модель для области низких частот

В этом случае используются следующие выражения, которые непосредственно вытекают из формул Эберса - Молла:

$$I_{\text{Э}} = I_{\text{Б}} + I_{\text{К}} \quad (6.85)$$

$$I_{\text{Э}} = I_{\text{ЭНАС}} e^{U_{\text{БЭ}}/U_T}, \text{ так как } U_{\text{БЭ}}/U_T \gg 1 \quad (6.86)$$

$$I_{\text{К}} = \alpha I_{\text{ЭНАС}} e^{U_{\text{БЭ}}/U_T} = \alpha_F I_{\text{Э}}, \text{ так как } U_{\text{БК}}/U_T \ll 1 \quad (6.87)$$

$$I_{\text{Б}} = (1 - \alpha_F) I_{\text{Э}} \quad (6.88)$$

$$\beta_F = I_{\text{К}} / I_{\text{Б}} = \alpha_F / (1 - \alpha_F) \quad (6.89)$$

Малое приращение входного напряжения $u_{\text{бэ}}$ приводит к появлению переменной составляющей тока коллектора. Эти величины связаны между собой соотношением

$$i_k = S u_{\text{бэ}} \quad (6.90)$$

где

$$S = \frac{\partial I_{\text{К}}}{\partial U_{\text{БЭ}}} = \frac{I_{\text{К}}}{U_T} \quad (6.91)$$

где S - крутизна передаточной характеристики.

Из формул (6.85) и (6.87) следует, что

$$I_{\text{Б}} = (1 - \alpha_F) I_{\text{ЭНАС}} e^{U_{\text{БЭ}}/U_T} \quad (6.92)$$

Отсюда можно определить сопротивление базы r_{π} в режиме малого сигнала, воспользовавшись выражением

$$\frac{1}{r_{\pi}} = \frac{\partial I_{\text{Б}}}{\partial U_{\text{БЭ}}} = \frac{I_{\text{Б}}}{U_T} \quad (6.93)$$

откуда

$$r_{\pi} = \frac{U_T}{I_{\text{Б}}} = \frac{U_T}{I_{\text{К}} / \beta_F} = \frac{\beta_F}{S} \quad (6.94)$$

Чтобы найти сопротивление эмиттера $r_{\text{э}}$ в режиме малого сигнала, воспользуемся формулой

$$\frac{1}{r_{\text{э}}} = \frac{\partial I_{\text{Э}}}{\partial U_{\text{БЭ}}} = \frac{I_{\text{Э}}}{U_T} \quad (6.95)$$

из которой следует, что

$$r_{\text{э}} = U_T / I_{\text{Э}} \quad (6.96)$$

Итак

$$r_{\pi} = (1 + \beta_F) r_{\text{э}} \quad (6.97)$$

Модель для области высоких частот

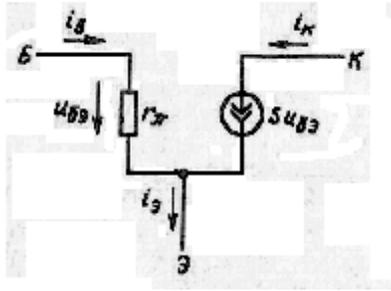


Рис. 6.7. Гибридная эквивалентная схема биполярного транзистора для области низких частот

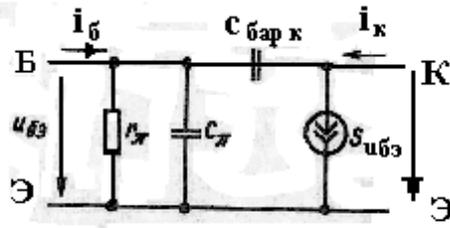


Рис.6.8. П-образная гибридная эквивалентная схема биполярного транзистора для области высоких частот

$$C_{пер} = C_{бар к}; C_{к} = C_{диф} + C_{бар э}$$

Гибридная П-образная модель, пригодная для описания высокочастотных свойств транзистора, должна учитывать влияние диффузионного сопротивления базовой области, модуляцию толщины базы, а также явления, связанные с накоплением заряда в области базы (рис. 6.8). Существенно, что такая модель учитывает связь между зарядом, накопленным в базе, и напряжением, приложенным к эмиттеру.

Введем крутизну передаточной характеристики транзистора

$$S = \partial I_{к} / \partial U_{БЭ} \tag{6.98}$$

Так как концентрация электронов на границе базовой области, прилегающей к переходу эмиттер-база, в транзисторе типа *n-p-n*

$$n_p(0) = n_{0б} e^{U_{БЭ}/U_T} \tag{6.99}$$

а абсолютное значение коллектора тока

$$|I_{к}| = \frac{qAD_n n_p(0)}{W_б} \text{ при } \frac{W_б}{L_n} \ll 1 \tag{6.100}$$

то

$$S = |I_{к}| / U_T \tag{6.101}$$

С другой стороны,

$$\frac{1}{r_{\pi}} = \frac{\partial I_{Б}}{\partial U_{БЭ}} \tag{6.102}$$

И так как ток базы

$$I_{Б} = qAn_i^2 \left(\frac{D_p}{W_{э}N_{дэ}} + \frac{D_n W_б}{2N_a L_n^2} \right) \left[\exp\left(\frac{U_{БЭ}}{U_T}\right) - 1 \right] \tag{6.103}$$

то

$$1/r_{\pi} = I_{Б} / U_T = I_{к} / (\beta U_T) = S / \beta \tag{6.104}$$

откуда

$$r_{\pi} = \beta / S \quad (6.105)$$

Диффузионная емкость определяется выражением

$$C_{\text{диф}} = \frac{dQ_B}{dU_{\text{БЭ}}} \quad (6.106)$$

Так как заряд, накопленный в области базы,

$$Q_B = \frac{q}{2} n_p(0) W_B A = \frac{q n_{p0} W_B A}{2 U_T} \exp\left(\frac{U_{\text{БЭ}}}{U_T}\right) \quad (6.107)$$

А ток эмиттера

$$I_{\text{Э}} \approx \frac{q n_{p0} D_{nB} A}{W_{\text{б}}} \exp\left(\frac{U_{\text{БЭ}}}{U_T}\right) \quad (6.108)$$

то диффузионная емкость

$$C_{\text{диф}} = W_B^2 I_{\text{Э}} / (2 D_{n0} U_T) \quad (6.109)$$

Если воспользоваться тем, что $I_{\text{Э}} \approx I_K$, $S = I_K / U_T$, то

$$C_{\text{диф}} = W_B^2 S / (2 D_{n0}) \quad (6.110)$$

Эквивалентная схема, представленная на рис.6.8, содержит также элементы: $C_{\text{барэ}}$ - барьерную емкость перехода эмиттер-база, $C_{\text{барк}}$ - барьерную емкость перехода коллектор-база.

Используя законы Кирхгофа, получаем выражение для граничной частоты

$$\omega_{\beta} = 2\pi f_{\beta} = [r_{\pi} (C_{\text{диф}} + C_{\text{барк}} + C_{\text{барэ}})]^{-1} \quad (6.111)$$

Анализируя П-образную эквивалентную схему, можно получить значение верхней граничной частоты усиления транзистора. Та частота, на которой модуль коэффициента передачи тока базы $|\beta|$ становится равным единице, называется граничной частотой коэффициента передачи тока базы и обозначается символом ω_T . В соответствии с формулой (6.26) эта частота удовлетворяет уравнению

$$\left| \frac{\beta}{1 + j\omega_T / \omega_{\beta}} \right| = 1 \quad (6.112)$$

Поскольку $\beta \gg 1$, равенство (6.112) выполняется в том случае, если

$$\omega_T = \beta \omega_{\beta} = 2 D_{n0} / W_{\text{б}}^2 \quad (6.113)$$

Можно заметить, что частота ω_T обратно пропорциональна времени распространения носителей в базе, определяемому по формуле (6.32), поскольку $C_{\text{диф}} \gg C_{\text{барэ}} + C_{\text{барк}} = C_{\pi}$ (рис. 6.8).

6.7. Зарядовая модель биполярного транзистора

Так принято называть одну из моделей, позволяющую в ряде случаев упростить анализ и расчеты. Модель основана на понятиях не тока или напряжения, а зарядов, которые накапливаются в различных областях прибора.

Зарядовая модель структуры с p - n -переходом

Если предположить, что ток, протекающий по переходу, является линейной функцией заряда Q и его производной dQ/dt , а также переменного во времени напряжения U , приложенного к переходу, то накопленный заряд можно представить в виде функции, которая зависит только от времени и удовлетворяет обобщенному уравнению сохранения заряда

$$\frac{d\rho}{dt} = -\frac{dJ}{dx} - \frac{q[p_n(x) - p_{n0}]}{\tau_p} + \frac{q[n_p(x) - n_{p0}]}{\tau_n} \quad (6.114)$$

Отсюда следует зарядовая модель перехода в виде

$$I = \frac{Q_n}{\tau_n'} + \frac{Q_p}{\tau_p'} + \frac{dQ_p}{dt} + \frac{dQ_{nep}}{dt} \quad (6.115)$$

где I - полный ток в переходе; Q_n - заряд, внесенный электронами в нейтральную p -область; Q_p - заряд, внесенный дырками в нейтральную n -область; τ_n' - среднее время жизни электрона с учетом процессов рекомбинации в объеме и на поверхности; τ_p' - среднее время жизни дырок; Q_{nep} - заряд в обедненной области; Q_n/τ_n' и Q_p/τ_p' - составляющие тока, связанные с поддержанием процессов рекомбинации в нейтральных областях; dQ_n/dt и dQ_p/dt - составляющие тока, обусловленные изменениями избыточных зарядов в нейтральных областях; dQ_{nep}/dt - составляющая тока, обусловленная изменением заряда в обедненной области.

При низком уровне инжекции данная модель справедлива также и в динамическом режиме. Она позволяет описывать свойства структуры с p - n -переходом на основании изучения зарядов, которые накапливаются в нейтральных областях и в обедненной области.

Зарядовая модель транзистора в активном режиме

Транзистор типа n - p - n , работающий как усилитель, управляется напряжением, которое прикладывается к переходу база - эмиттер. Изменение этого напряжения влияет на значение Q_B и другие составляющие заряда. Инжектированные заряды Q_B и Q_E определяют токи в стационарном режиме. Обозначим их суммой символом Q_F , подчеркнув тем самым, что эта величина относится к активному режиму. Будет показано, что данная модель содержит и другие составляющие, которые изменяются во времени и определяют режим работы транзистора.

Если заряд Q_F возрастает во времени, то появляется составляющая тока базы dQ_F/dt . Аналогично, изменение зарядов $Q_{БЭ}$ и $Q_{БК}$, которые накапливаются в переходах база-эмиттер и база-коллектор соответственно, обеспечивает дополнительный вклад в ток базы. В результате приходим к следующему выражению тока базы:

$$i_B = \frac{Q_F}{\tau_{BF}} + \frac{dQ_F}{dt} + \frac{dQ_{БЭ}}{dt} + \frac{dQ_{БК}}{dt} \quad (6.116)$$

Здесь τ_{BF} - время жизни неосновных носителей в базе при прямом смещении; три первых слагаемых правой части - это токи, протекающие через переход база-эмиттер, а четвертое слагаемое - ток через переход база-коллектор.

Примем во внимание прежде всего то, что ток коллектора можно выразить через заряд Q_F и характеристическое время τ_F :

$$I_K = Q_F / \tau_F \quad (6.117)$$

Далее учтем, что скорость, с которой заряд Q_B рекомбинирует в базе и инжектируется в виде дырок в область эмиттера, пропорциональна величине Q_F . Отсюда следует соотношение, определяющее входной ток базы:

$$I_B = Q_F / \tau_{BF} \quad (6.118)$$

Используя первый закон Кирхгофа, получаем зарядовую модель транзистора, работающего в активной области:

$$i_{\ominus} = -\frac{Q_F}{\tau_F} - \frac{Q_F}{\tau_{BF}} - \frac{dQ_F}{dt} - \frac{dQ_{БЭ}}{dt} \quad (6.119)$$

$$i_B = \frac{Q_F}{\tau_{BF}} + \frac{dQ_F}{dt} + \frac{dQ_{БЭ}}{dt} + \frac{dQ_{БК}}{dt} \quad (6.120)$$

$$i_K = \frac{Q_F}{\tau_F} - \frac{dQ_{БК}}{dt} \quad (6.121)$$

Дифференциальные уравнения, описывающие зарядовую модель транзистора, являются линейными, несмотря на то, что токи и напряжения в транзисторе связаны между собой нелинейными зависимостями. Эти уравнения весьма полезны для расчета устройств, в которых транзистор подключен к внешней нагрузке. Тем не менее данная модель не позволяет судить во всей полноте о внутренних процессах в транзисторе.

Чтобы лучше понять пределы применимости зарядовой модели, модели Эберса - Молла и гибридной П-образной модели, необходимо проводить сравнительное сопоставление параметров этих моделей.

Применение модели

Среди возможных применений зарядовой модели можно указать использование ее для нахождения тока коллектора i_K в транзисторе, который работает в активном режиме и имеет источник тока i_B в базовой цепи. Особенно часто эту модель применяют при исследовании работы

транзистора в режиме большого сигнала, а также при изучении нестационарного процесса, сопровождающего переход транзистора из режима отсечки в режим насыщения. Эта модель позволяет также определить ток стока в МОП-транзисторе на основании соотношения, которое связывает заряд в канале с временем перехода носителей через область канала.

6.8. Транзистор в режиме переключения

Биполярный транзистор может использоваться не только как усилительный элемент, но и как коммутатор, осуществляющий переключение напряжения с высокого уровня на низкий, и наоборот. При этом статическая рабочая точка скачкообразно переходит в новое положение, достаточно удаленное от первоначального. Очевидно, что методы линейного анализа в этом случае становятся непригодными.

Чтобы изменить положение рабочей точки, необходимо тем или иным образом изменить заряды, накопленные в нейтральных областях и в областях пространственного заряда, для чего требуется некоторое время. Поставим задачу оценить время накопления заряда, используя некоторые соображения, касающиеся процесса накопления заряда в нейтральных областях транзистора типа $n-p-n$.

В статическом режиме если $U_{KB} = 0$, то

$$I_B = I_{ЭНАС} (1 - \alpha_F) \left[\exp\left(\frac{U_{БЭ}}{U_T}\right) - 1 \right] \quad (6.122)$$

и если Q_F - избыточный заряд в нейтральных областях эмиттера и базы, то

$$I_B = Q_F / \tau_{BF} \quad (6.123)$$

где τ_{BF} - время жизни неосновных носителей в базе.

Аналогично, если $U_{ЭБ} = 0$, то

$$I_B = \frac{Q_R}{\tau_{BR}} = I_{КНАС} (1 - \alpha_R) \left[\exp\left(\frac{U_{KB}}{U_T}\right) - 1 \right] \quad (6.124)$$

где τ_{BR} - время жизни носителей при смещении в обратном направлении; данный параметр для переключательных транзисторов обычно приводится в справочниках.

Если $U_{KB} \neq 0$ и $U_{ЭБ} \neq 0$, то, просуммировав выражения (6.121) и (6.123), получим уравнения Эберса - Молла для тока базы.

В динамическом режиме для транзистора типа $n-p-n$ справедливы следующие уравнения зарядовой модели, как на постоянном, так и на переменном токе:

$$i_{\ominus} = \frac{Q_R}{\tau_{BR}} - \frac{Q_F}{\tau_{BF}} - \frac{dQ_F}{dt} - \frac{dQ_{ЭБ}}{dt} \quad (6.125)$$

$$i_K = \frac{Q_F}{\tau_F} - \frac{Q_R}{\tau_{BR}} - \frac{dQ_R}{dt} - \frac{dQ_{KB}}{dt} \quad (6.126)$$

$$i_B = \frac{Q_F}{\tau_{BF}} + \frac{Q_R}{\tau_{BR}} + \frac{dQ_{ЭБ}}{dt} + \frac{dQ_R}{dt} + \frac{dQ_{KB}}{dt} \quad (6.127)$$

Предположим теперь, что ток базы резко изменяется от значения I_{B1} до значения $-I_{B2}$. Избыточный заряд начинает уменьшаться, однако активный заряд Q_B неизменен в интервале времени от $t = 0$ до $t = t_{BR}$, где

$$dQ_F / dt = 0 \quad (6.128)$$

$$Q_F / \tau_{BF} = I_B \quad (6.129)$$

I_B - ток базы в режиме насыщения транзистора.

Аналогичные рассуждения приводят к тому, что

$$Q_R = \tau_{BR} (I_{B1} - I_B) \quad (6.130)$$

Используя это выражение в качестве начального условия для уравнения (6.126), получаем

$$Q_R = \tau_{BR} (I_{B1} - I_{B2}) e^{-t/\tau_{BR}} + \tau_{BR} (I_{B2} - I_B) \quad (6.131)$$

Если $t = t_{BR}$, то $Q_R = 0$. Отсюда, используя формулу (6.122), находим

$$t_{BR} = \tau_{BR} \ln[(I_{B1} - I_{B2}) / (I_B - I_{B2})] \quad (6.132)$$

Параметр t_{BR} имеет важное значение при оценке скорости процесса коммутации в транзисторе. Так или иначе, если требуется обеспечить высокую скорость переключения, следует не допускать работы транзистора в режиме насыщения.

6.9. Тиристоры

Тиристоры представляют собой четырехслойные полупроводниковые приборы, предназначенные для создания накопительных устройств, управляемых выпрямителей, регуляторов мощности и т. д. Эти приборы имеют два устойчивых состояния, в одном из которых они проводят ток («включено»), а в другом разрывают цепь («выключено»). Тиристоры могут работать с напряжениями до 1000 В и коммутировать токи до 500 А. Удастся достичь длительности переключения вплоть до десятков микросекунд.

Известны разновидности тиристоров: управляемые кремниевые вентили, тринисторы, динисторы, фототиристоры, программируемые однопереходные транзисторы и т. д.

Управляемый тиристор

Данный прибор используют как регулирующий элемент осветительной аппаратуры. Он применяется также в силовых устройствах

преобразования частоты, может служить мощным быстродействующим коммутатором и т. д.

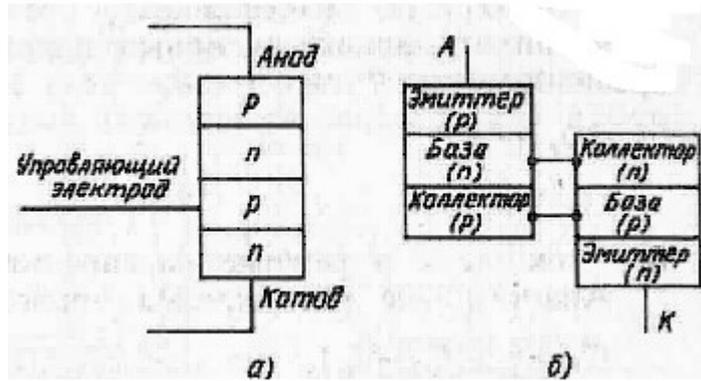


Рис. 6.9 Идеальная структура управляемого тиристора

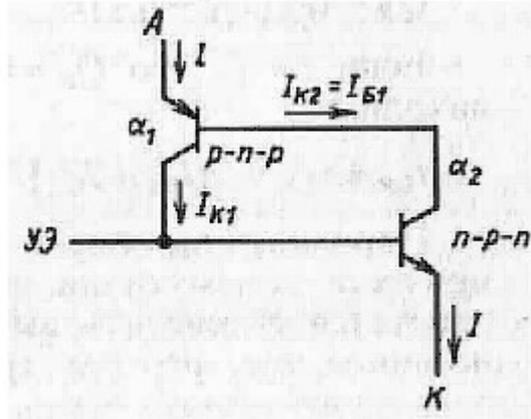


Рис. 6.10 Принципиальная схема кремниевого управляемого тиристора

Идеальная структура описываемого прибора (рис. 6.9, а) представляет объединение транзисторов типа $p-n-p$ и $n-p-n$ (рис. 6.9, б). Ток в цепи кремниевого управляющего электрода (УЭ) (рис. 6.10) усиливается $n-p-n$ -транзистором, поэтому в цепи базы $p-n-p$ -транзистора возникает ток $\beta_2 I_{\text{Э}}$. В свою очередь, этот ток усиливается транзистором типа $n-p-n$; ток $\beta_1 \beta_2 I_{\text{Э}}$ возрастает до тех пор, пока не становится равным току насыщения. В соответствии с рис. 6.10

$$I_{\text{Э}} = I_{\text{К}} + I_{\text{Б}} \quad (6.133)$$

$$I_{\text{К}} = \alpha I_{\text{Э}} + I_{\text{КБО}} \quad (6.134)$$

Объединяя эти два уравнения, получаем

$$I_{\text{Б1}} = (1 - \alpha_1) I_{\text{Э1}} + I_{\text{КБО1}} \quad (6.135)$$

Из рис. 6.10 следует также, что

$$I_{\text{К2}} = I_{\text{Б1}} \quad (6.136)$$

$$I_{\text{К2}} = \alpha_2 I_{\text{Э2}} + I_{\text{КБО2}} \quad (6.137)$$

Подставляя ток $I_{\text{Б1}}$ из (6.134) в (6.136), находим

$$I_{\text{К2}} = (1 - \alpha_1) I_{\text{Э1}} - I_{\text{КБО1}} \quad (6.138)$$

На основании формулы (6.135) запишем

$$\alpha_2 I_{\text{Э2}} + I_{\text{КБО2}} = (1 - \alpha_1) I_{\text{Э1}} - I_{\text{КБО1}} \quad (6.139)$$

Так как

$$I = I_{\text{Э1}} = I_{\text{Э2}} \quad (6.140)$$

то приходим к уравнению, которое описывает процесс в рассматриваемом приборе, наблюдаемый при переходе из области отсечки в проводящее состояние:

$$I = \frac{I_{КБО1} + I_{КБО2}}{1 - (\alpha_1 + \alpha_2)} \quad (6.141)$$

Если предположить, что коэффициенты умножения обоих транзисторов одинаковы, то явление лавинного пробоя будет наблюдаться при условии

$$(\alpha_1 + \alpha_2)M = 1 \quad (6.142)$$

На рис. 6.11 изображены статические характеристики прибора при токах управляющего электрода $I_{y3} > I_{y2} > I_{y1} > I_y = 0$. Можно заметить, что с ростом тока УЭ напряжение лавинного пробоя (напряжение включения) уменьшается.

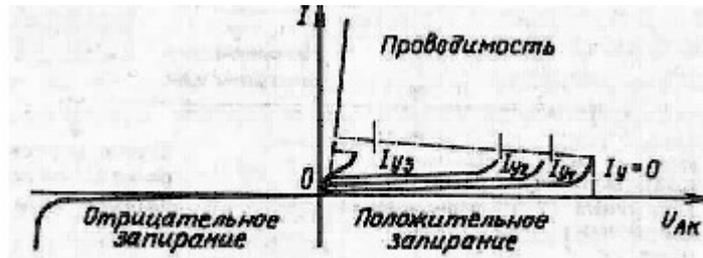


Рис. 6.11. Статические вольт-амперные характеристики кремниевого управляемого тиристора.

Показано, что прямое напряжение лавинного пробоя зависит от напряжения между управляющим электродом и катодом

На каждой отдельно взятой характеристике можно выделить четыре участка. На первом, линейном, участке в начале характеристики прибор имеет высокое сопротивление, так как здесь $\alpha_1 + \alpha_2 \ll 1$. Второй участок, определяемый равенством $dU/dI=0$, соответствует режиму лавинного пробоя, наступающему при $\alpha_1 + \alpha_2 = 1$. Третий участок характеристики изображен штриховой линией; здесь прибор имеет отрицательное динамическое сопротивление. Наконец, четвертому участку соответствует неравенство $\alpha_1 + \alpha_2 \gg 1$; здесь прибор находится в проводящем состоянии.

Как видно из рис. 6.11, при больших токах управляющего электрода лавинный пробой наблюдается в точках, близких к кривой, описывающей вольт-амперную характеристику выпрямляющего $p-n$ -перехода.

7. БИПОЛЯРНЫЕ ИНТЕГРАЛЬНЫЕ СХЕМЫ

В данной главе продолжено изучение биполярных транзисторов типа $n-p-n$ и $p-n-p$, а также диодов. Рассмотрены электротепловые модели этих приборов. Кроме того, приведены некоторые сведения, представляющие интерес как в данном контексте, так и в последующем, касающиеся БИС и тех семейств логических ИС, которые выполняются по биполярной технологии. В

заклучении кратко описаны два важнейших аналоговых устройства с биполярными транзисторами: дифференциальный и операционный усилители.

7.1. Биполярные транзисторы ИС

Транзисторы являются основными элементами ИС. Чтобы реализовать структуру микросхемы, в процессе ее изготовления необходимо выполнить большое число операций. Другие же элементы - диоды, резисторы и конденсаторы - можно создать на отдельных этапах общего технологического процесса.

Чаще используют транзисторы типа *n-p-n*, поскольку их параметры легче контролировать при изготовлении и лучше их частотные характеристики.

Транзистор типа *n-p-n* со скрытым слоем

Коллектор в транзисторе представляет *n*-область с повышенным сопротивлением. Эта область создается путем диффузии примеси в подложку *p*-типа, имеющую еще более высокое удельное сопротивление. При этом возникает электрическая изоляция между коллектором и подложкой. Данный способ называют изоляцией с помощью *p-n*-перехода.

Так как коллекторный контакт расположен в верхней части прибора, значительная часть тока I_k протекает в *n*-области. В результате возникает некоторое сопротивление, включенное последовательно с коллектором и приводящее к дополнительному падению напряжения. Это последовательное сопротивление эпитаксиального *n-p-n*-транзистора.

Один из способов, применяемых для оценки последовательного сопротивления $r_{кпос}$, известен под названием метода частичных областей. Согласно этому методу, коллекторную область делят на четыре подобласти (рис. 7.1).

Сопротивление каждого из них определяется формулой

$$r = \rho_k d / A_{эф} \quad (7.1)$$

где ρ_k - удельное сопротивление материального коллектора; d - расстояние между переходом эмиттер-база и коллекторно-эмиттерным переходом; $A_{эф}$ - площадь поперечного сечения на среднем уровне, т.е.

$$A_{эф} = (l_э + l_к) x_к / 2 \quad (7.2)$$

Здесь $l_э$ - длина стороны в верхнем основании перехода, $l_к$ - длина стороны в нижнем основании перехода, $x_к$ - толщина коллекторной пленки.

Окончательно сопротивление $r_{кпос}$ есть результат последовательного соединения сопротивлений областей, расположенных по вертикали между контактами коллектора и эмиттера:

$$r_{кпос} = \rho_k (x_к / A_э + x_к / A_{кк}) \quad (7.3)$$

где $A_э$ - площадь эмиттера; $A_{кк}$ - площадь коллекторного контакта.

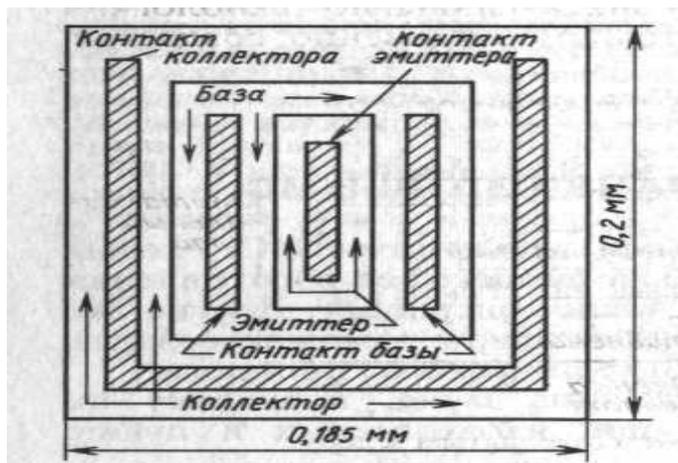
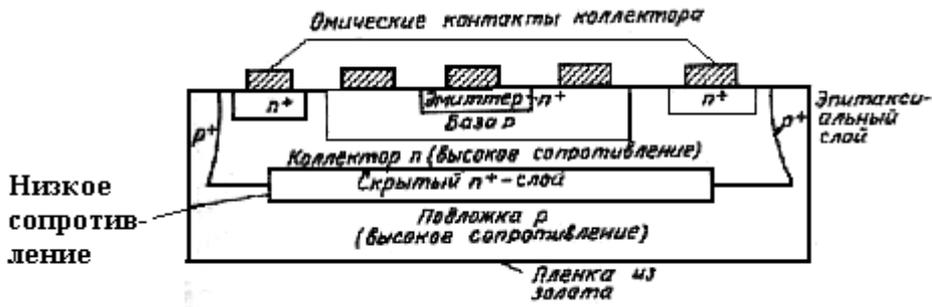


Рис. 7.1. Условное графическое обозначение транзистора.
Показана область коллектора, которую для оценки сопротивления коллектора $r_{к\text{ пос}}$ можно разделить на четыре трапециоида.

При изоляции с помощью p - n -переходов, смещенных в обратном направлении, возникает трудности, обусловленные появлением добавочных емкостей, токов утечки и влиянием паразитного транзистора типа p - n - p . Их можно избежать, используя диэлектрическую изоляцию. Однако это более дорогой способ. Существенно снизить последовательное сопротивление коллектора удастся перейдя к конструкции транзистора типа n - p - n со скрытым слоем. Сопротивление $r_{к\text{ пос}}$ такого транзистора становится пренебрежимо малым, благодаря чему эти транзисторы используются в составе биполярных ИС.

Структуру со скрытым слоем можно создать путем локальной диффузии донорной примеси, обеспечивающей низкое удельное сопротивление, в подложку p -типа, на которой выращивается эпитаксиальный слой коллектора. Можно также селективно выращивать слой n^+ -типа используя эпитаксиальную технологию с маскированием (рис. 7.2). При этом одновременно создается коллекторная область с повышенным сопротивлением. Как напряжение пробоя, так и емкость перехода коллектор-подложка определяются лишь удельным сопротивлением подложки.

Последовательность основных процессов, используемых для изготовления n - p - n -транзисторов со скрытым слоем, такая: на поверхность подложки p -типа методом селективной диффузии создается слой n^+ -типа; создается кремниевая пленка n -типа толщиной 3 мкм; проводится глубокая диффузия акцепторной примеси, обеспечивающая электрическую изоляцию этих элементов (данный процесс наиболее сложен); выполняется диффузия донорной примеси для создания сильно легированной области n^+ -типа под коллекторным электродом; диффузионным способом формируется база и

Рис. 7.2. Транзистор типа $n-p-n$ со скрытым слоем

эмиттер; создаются контактные окна; завершающими процессами является металлизация, проводимая для получения токоведущих дорожек, и пассивирование. Сюда входят классические процессы обработки кремния: фотолитография, диффузия и/или ионная имплантация, эпитаксия, высокотемпературное оксидирование, металлизация, очистка поверхности, травление и нанесение из газовой фазы защитной пленки (пассивирование).

При работе в аналоговых цепях биполярный транзистор обычно находится в режиме, далеком от насыщения, и должен иметь высокие значения коэффициента передачи тока и напряжения пробоя. Поэтому стремятся выбирать его параметры в следующих пределах: толщина эпитаксиального слоя 8-12 мкм, удельное сопротивление 1-5 Ом·см, толщина базы 0,5-1 мкм. В противоположность этому в логических цепях, где требуется высокая скорость переключения, необходимо применять транзисторы с малым напряжением насыщения и небольшим напряжением излома характеристики перехода коллектор-база. Эпитаксиальная пленка здесь выполняется более тонкой. Чтобы снизить паразитные емкости, вносимые изоляцией, диффузионные области базы и эмиттера располагают как можно ближе к поверхности.

При создании логических БИС и СБИС, для которых требуются одновременно высокие быстродействия и степень интеграции, необходимо по возможности уменьшать площадь эмиттера, чтобы предельно уменьшить емкость база-эмиттер (используется процесс изопланарной технологии), вывод базы следует размещать как можно ближе к эмиттеру. Наконец, желательно сокращать размеры изолирующей системы, оксидируя дополнительные участки (предусматривается добавочное окно, которое уменьшает емкости база-эмиттер и коллектор-подложка).

Транзистор типа $p-n-p$ с горизонтальной структурой

Как уже отмечалось, обычно ИС используют $n-p-n$ -транзисторы. В некоторых аналоговых и логических устройствах необходимо одновременно иметь транзисторы типов $n-p-n$ и $p-n-p$. Имеется ряд структур - боковая, или горизонтальная, вертикальная, или подложечная, комплементарная с изоляцией

$p-n$ -переходами и т.д., которые позволяют создавать транзисторы типа $p-n-p$. Наиболее простой и чаще всего применяемой в ИС является горизонтальная структура.

На рис. 7.3 схематически показан поперечный разрез горизонтального транзистора типа $p-n-p$. Здесь для создания областей эмиттера и коллектора используется диффузионная p -область базы $n-p-n$ - транзистора. Базу образуют эпитаксиальный слой, являющийся коллектором для транзистора типа $n-p-n$. Аналогично диффузионная n^+ - область, которая образует эмиттер $n-p-n$ -транзистора, служит контактом базы горизонтального транзистора типа $p-n-p$. Заглубленный слой обеспечивает низкое сопротивление базовой области и позволяет исключить влияние паразитного $p-n-p$ - транзистора, который может быть образован системой эмиттер-база-подложка.

Рассмотренная структура имеет ряд недостатков: большая толщина базы (6-12 мкм); меньший, чем у $n-p-n$ - транзистора, коэффициент передачи тока; худшие частотные свойства; меньшее напряжение пробоя перехода коллектор-база.

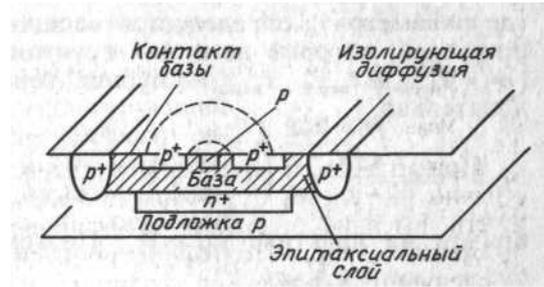


Рис. 7.3. Транзистор типа $p-n-p$ с горизонтальной структурой

Так как горизонтальный $p-n-p$ -транзистор имеет однородную базу, анализ его работы, основанный на учете эффектов первого порядка, оказывается несложным. Так, коэффициент инжекции определяется выражением

$$\gamma_I = \frac{1}{1 + W_I^2 (2L_p^2)} \approx 1 - W_I^2 (2L_p^2) \quad (7.4)$$

где W_I - эффективная толщина базы для носителей, инжектированных сбоку (приблизительно совпадает с толщиной базы, измеренной на границе).

Коэффициент передачи тока базы является частотным от деления величин I_K и I_B , причем

$$I_K = \gamma_I I_{\Sigma} A_{\Sigma} / (A_{\Sigma} + A_p) \quad (7.5)$$

где

$A_{\Sigma} \approx x_{пер\ к} (\ell + W)$ - эффективная площадь эмиттера для носителей, инжектированных сбоку ($x_{пер\ к}$ - глубина коллекторного перехода, ℓ и W - длина и ширина эмиттерного окна соответственно);

$$A_p \approx \ell (W + 2x_{пер\ к})$$

Ток базы, образуемый носителями, которые инжектируются вертикально и составляют основную часть тока рекомбинации, определяется выражением

$$I_B \approx (1 - \gamma_v) I_{\Sigma} A_p / (A_{\Sigma} + A_p) \quad (7.6)$$

где параметр γ_v определяется толщиной базы W_v для прочих носителей, которые не инжектируются в боковом направлении ($W_v = x_{пер\ n} - x_{пер\ к}$, $x_{пер\ n}$ – глубина перехода в подложке). Следовательно,

$$\beta = \frac{I_K}{I_B} = \frac{\gamma_1 A_{\text{э}} / A_p}{1 - \gamma_v} \approx \frac{A_{\text{э}}}{A_p} \left(1 - \frac{W_l^2}{2L_p^2} \right) \left(1 + \frac{W_v^2}{2L_p^2} \right) \frac{2L_p^2}{W_v^2} \quad (7.7)$$

причем на практике $L_p \gg W_l$. Поэтому

$$\beta \approx \frac{A_{\text{э}}}{A_p} \left(1 + \frac{2L_p^2}{W_v^2} \right) \quad (7.8)$$

Можно также получить приближенное выражение для отношения $A_{\text{э}}/A_p$:

$$\frac{A_{\text{э}}}{A_p} = \frac{1 + (W/l)}{2 + (W/x_{перк})}$$

Отсюда следует вывод о том, что, выбирая $l \approx 10W$, можно повысить коэффициент передачи β .

Транзистор типа *p-n-p* с вертикальной структурой

Структура этого транзистора такова, что здесь подложка выполняет роль коллектора, база *n-p-n*-транзистора служит эмиттер, в то время как базой является эпитаксиальный слой (рис. 7.4).

Данный транзистор имеет высокий коэффициент передачи тока и значительное пробивное напряжение перехода база – коллектор. Однако он имеет плохие частотные характеристики; кроме того, в такой цепи наиболее отрицательный потенциал имеет подложка. Последнее обстоятельство в схеме эмиттерного повторителя.

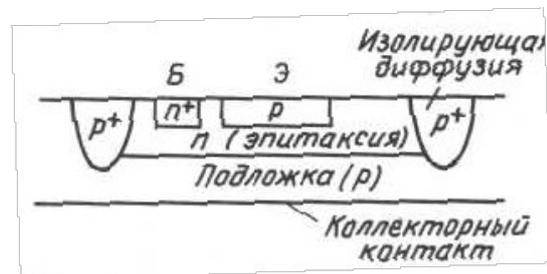


Рис. 7.4. Транзистор типа *p-n-p* с вертикальной структурой

Составной *p-n-p*-транзистор

Управлять толщиной базы и другими параметрами транзистора с горизонтальной структурой значительно сложнее, чем параметрами планарного транзистора, эмиттерный и коллекторный переходы которого параллельны. Так как значение β для транзистора с горизонтальной структурой близко единице, его часто используют в комбинации с планарным транзистором типа *p-n-p*, который позволяет повысить коэффициент передачи такого составного транзистора (рис. 7.5). Результирующий коэффициент передачи практически равен произведению коэффициента передачи отдельных транзисторов; относительно внешних зажимов такой составной прибор имеет характеристики *p-n-p* - транзистора.

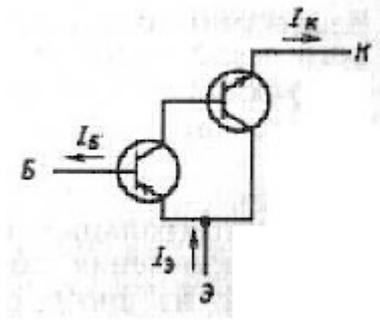


Рис 7.5. Схема составного транзистора

Коэффициент передачи составного транзистора

$$\beta_C = I_K / I_B \quad (7.9)$$

При этом

$$I_K / (1 + \beta_{npn}) = I_{Knp} \quad (7.10)$$

$$\beta_C = I_B / I_{Knp} = 1 / \beta_{np} \quad (7.11)$$

Объединяя формулы (7.10) и (7.11), получаем

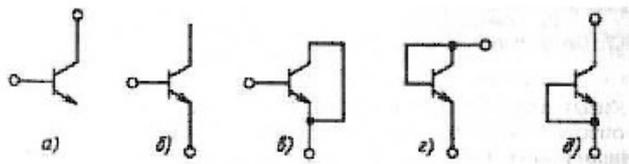
$$\beta_C = \frac{(1 + \beta_{npn}) I_{Knp}}{I_{Knp} / \beta_{np}} \quad (7.12)$$

откуда

$$\beta_C = (1 + \beta_{npn}) \beta_{np} \quad (7.13)$$

Хотя составной транзистор имеет более высокий коэффициент передачи тока, чем $p-n-p$ -транзистор, его площадь больше из-за необходимости размещать рядом $n-p-n$ -транзистор. Частотная характеристика типа $p-n-p$, а выходная проводимость выше, так как

$$g_{кэ} = \partial I_K / \partial U_{кэ} \approx (1 + \beta_{np}) g_{Knp} + g_{Knp} \quad (7.14)$$

Рис. 7.6. Пять типичных схем, позволяющих получать диоды из транзисторов типа $n-p-n$

и величина g_{Knp} может достигать больших значений из-за того, что становится невозможным управлять процессом модуляции толщины базы.

7.2. Диоды ИС

В интегральных схемах диоды с $p-n$ -переходом получают путем включения обычных $n-p-n$ -транзисторов в соответствии с одной из пяти схем, представленных на рис. 7.6. Такие диоды используются как нелинейные, а также как изолирующие элементы.

Легко показать, что, зная распределение неосновных носителей, соответствующие каждой из этих пяти моделей при смещении в прямом направлении, можно найти время накопления и рассасывания заряда

(длительность отклика). Выбор той или иной модели в каждой ситуации зависит от требований к длительности отклика, напряжению пробоя, обратному току (току утечки), емкости, а также проводимости при прямом смещении.

Диоды Шотки ИС

Принцип работы диода Шотки был объяснен ранее; там же были рассмотрены его принципиальные отличия от диода с $p-n$ -переходом. Существенным достоинством такого диода является прежде всего то, что для

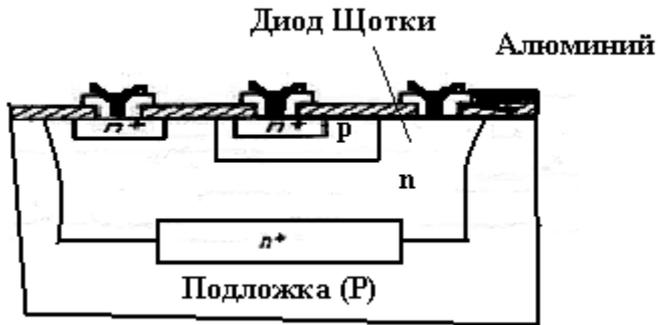


Рис. 7.7. Продольный разрез диода Шотки

его изготовления не требуется дополнительных этапов, не входящих в технологический цикл производства ИС.

Кроме того, с помощью диода Шотки удается повысить скорость нестационарных процессов в переключаемых транзисторах ИС. Такие диоды можно получить, нанося алюминиевые слои на кремний n -типа (рис. 7.7). Однако при этом трудно обеспечить достаточную воспроизводимость

параметров.

Поэтому лучше диода Шотки получают, применяя сплавы $P_t - N_i$.

В дальнейшем рассмотрим одно из применений диодов Шотки в логических ИС типа ТЛШ. Здесь за счет подключения этих диодов к $p-n$ -переходам удается избежать глубокого насыщения транзисторов.

7.3. Электрические модели биполярных ИС

При разработке дискретных полупроводниковых приборов приходится обращать серьезное внимание на тепловые эффекты. Еще в большей степени это относится к ИС, элементы которых находятся на расстоянии нескольких микрон друг от друга и выделение теплоты в одном элементе непосредственно влияет на характеристики соседних элементов. Следует иметь в виду, что резисторы ИС, как и транзисторы, имеют положительные температурные коэффициенты сопротивления – с ростом температуры их сопротивление увеличивается.

Интересно рассмотреть свойства кремния, который является основным материалом для производства ИС. Эти свойства определяются собственными параметрами кристалла (подвижностью, коэффициентом диффузии, концентрацией свободных носителей и т.д.). Основным электрическим параметром, через который можно выразить остальные параметры, является удельная электрическая проводимость σ , определяемая формулой

$$\sigma_p = q \mu_p p, \quad (7.15)$$

$$\sigma_n = q \mu_n n. \quad (7.16)$$

Подвижность основных носителей связана с числом столкновений, которые они испытывают при движении внутри кристалла. Подвижность связана с абсолютной температурой формулой

$$\mu = CT^\eta. \quad (7.17)$$

Здесь C – некоторая постоянная, η – параметр, определяемый экспериментально, для которого существуют графики. Следует отметить, что удельная проводимость зависит от температуры так же, как и подвижность.

Анализируя отдельные элементы, можно создать из электротермические модели, которые, в свою очередь, позволяют изучать влияние температуры на характеристики ИС.

Электротермическая модель резистора

Зная геометрию резистора и удельную проводимость материала, можно вычислить его сопротивление R . Например,

$$R = L/(\sigma A) = L/(q \mu_n n A) = LT^\eta/(qnAC) \quad (7.17)$$

где L – длина резистора; A – площадь его поперечного сечения в кремнии n -типа.

Можно убедиться в том, что если известно сопротивление R при некоторой температуре T_0 , то при повышении температуры dT значение сопротивления изменяется:

$$R(T) \approx R_{T_0}(1 + \delta_R dT) \quad (7.19)$$

где $T=T_0+dT$; δ_R – температурный коэффициент сопротивления (относительное изменение сопротивления при изменении температуры на 1К).

Пусть резистор имеет некоторое сопротивление R_{T_0} , находится при температуре T_0 и через него протекает некоторый постоянный ток I_{T_0} . При малом приращении тока i , считая температуру T_0 неизменной, получаем приращение напряжения

$$u = R_{T_0} i \quad (7.20)$$

Если температура изменяется на величину dT , то, используя выражение функции $R(T)$ и полагая справедливым закон Ома, получаем следующую электротермическую модель резистора:

$$u = R_{T_0} i + R_{T_0} I_{T_0} \delta_R dT.$$

Второе слагаемое в правой части эквивалентно некоторому источнику напряжения, зависящему от начального значения сопротивления и тока в цепи; последовательно с источником включен резистор сопротивлением R_{T_0} . Полярность источника совпадает с полярностью напряжения.

Электротермическая модель диода

Экспериментально найдено, что для диода, смещенного в прямом направлении, температурный коэффициент напряжения $\delta_U \approx -2$ мВ/К без учета

влияния объемного сопротивления. Соответствующая модель включает источник малого сигнала i , источник тока смещения I_{T0} и собственное сопротивление перехода, которое определяется из соотношения

$$u = ri + \delta_d dT \quad (7.22)$$

Рассматривая более реальную модель диода, смещенного в прямом направлении, и учитывая наличие двух областей, в которых заряды перемещаются по обе стороны от перехода и которые имеют эквивалентные сопротивления r_1 и r_2 , приходим к полной модели, описываемой формулой

$$u = (r_1 + r_2)i + \frac{kT_0}{qI_{T0}} + \delta_d dT + (r_1 + r_2)I_{T0} \delta_R dT \quad (7.23)$$

Электротермическая модель n-p-n-транзистора

Данная модель является более важной, чем модели диода, резистора и барьерной емкости (здесь не рассматриваемой). Объясняется это тем, что транзистор является основным элементом ИС. Кроме того, через транзистор протекает значительный ток, вызывающий его нагрев.

Будем рассматривать общий случай транзистора, который изолирован от подложки p - n -переходами, смещенными в обратном направлении. Токи транзистора определяются на основании модели Эберса - Молла:

$$\begin{aligned} I_{\text{Э}} &= I_{\text{ЭБО}} \left[\exp\left(\frac{U_{\text{ЭБ}}}{U_T}\right) - 1 \right] - \alpha_R I_{\text{КББ}} \left[\exp\left(\frac{U_{\text{КБ}}}{U_T}\right) - 1 \right] \\ I_{\text{К}} &= -\alpha_F I_{\text{ЭБО}} \left[\exp\left(\frac{U_{\text{ЭБ}}}{U_T}\right) - 1 \right] + I_{\text{КББ}} \left[\exp\left(\frac{U_{\text{КБ}}}{U_T}\right) - 1 \right] \end{aligned} \quad (7.24)$$

При этом считается, что ток, втекающий в транзистор, положителен, а ток, вытекающий из него, отрицателен.

Если транзистор работает в активном режиме при нормальном включении (переход база-эмиттер смещен в прямом направлении, а переход коллектор-база - в обратном), то экспоненциальным слагаемым, обусловленным обратным смещением перехода, можно пренебречь; тогда

$$\begin{aligned} I_{\text{Э}} &= I_{\text{ЭБО}} \left[\exp\left(\frac{U_{\text{ЭБ}}}{U_T}\right) - 1 \right] - \alpha_R I_{\text{КБ0}} \\ I_{\text{К}} &= -\alpha_F I_{\text{ЭБО}} \left[\exp\left(\frac{U_{\text{ЭБ}}}{U_T}\right) - 1 \right] - I_{\text{КБ0}} \end{aligned} \quad (7.25)$$

Это новая форма модели Эберса-Молла, определяющей токи в рабочей точке транзистора.

Предположим теперь, что имеется малый входной ток i_B . Если температура прибора неизменна, то ток коллектора

$$i_k = -\beta_{T0} i_B \quad (7.26)$$

Если же температура изменяется на dT , то изменяются коэффициент передачи тока β_{T0} и ток $I_{КБ0}$ конечном счете зависящие от коэффициента диффузии, концентрации и подвижности носителей. Обозначая через δ_β и $\delta_{нас}$ температурим коэффициенты параметров β и $I_{КБ0}$ соответственно, имеем

$$\delta_\beta = \frac{1}{\beta_{T0}} \frac{\partial \beta_{T0}}{\partial T} \quad (7.27)$$

$$\delta_{нас} = \frac{1}{I_{КБ0T0}} \frac{\partial I_{КБ0T0}}{\partial T} \quad (7.28)$$

Малые приращения токов коллектора и эмиттера при температуре $T=T_0 + dT$ образуют электротермическую моде транзистора:

$$i_K = \delta_\beta \beta_{T0} dT + \beta_{T0} i_B + \delta_{нас} I_{КБ0T0} dT \quad (7.29)$$

$$i_\varepsilon = qI_{\varepsilon T0} \delta_d dT / (kT_0)$$

Более точной является модель, учитывающая объемные сопротивления различных областей транзистора, которые располагаются между переходами и омическими контактам. Подобная модель должна включать соответствующие резисторы.

Применение электротермических моделей

Как уже указывалось, каждому элементу ИС соответствует электротермическая модель, отображающая зависимость его электрических параметров от изменения температуры. Используя такое представление, можно заменить исходную схему цепью, состоящей из резисторов, источников напряжения и тока. Чтобы исследовать такую систему, следует воспользоваться методами анализа пассивных цепей, а в случай достаточно сложных ИС - компьютерными методами численного анализа. Одним из примеров является изотермический расчет, применяемый в том случае, когда параметры двух или большего числа элементов должны оставаться строго пропорциональными, например при анализе дифференциальных усилителей и других уравновешенных цепей. Еще одним примером является задача термостабилизации подложки, выступающая как центральная проблема при разработке ИС| надежно работающих в условиях высоких температур.

7.5. Семейства биполярных логических ИС

Интегральные схемы играют исключительно важную роль в системах обработки данных, цифровых системах связи, и автоматического регулирования. Работа всех этих систем основана на комбинации и/или повторении небольшого чист базовых операций.

Обычно используют логические элементы(ключи) типа ИЛИ, И, НЕ, а также триггеры.

В зависимости от структуры различают семейства логических элементов. Элементы одного семейства работают единому принципу. Известно значительное число семейств логических элементов. Сравнить их между собой довольно затруднительно, поскольку приходится учитывать различные факторы, такие, как скорость переключения или вред задержки сигнала, мощность, рассеиваемую одним ключом число реализуемых функций, устойчивость к перегрузкам нагрузочные способности по входу и выходу, диапазон рабочих температур и другие. В данном параграфе весьма кратко будут рассмотрены семейства логических элементов, наиболее интересные с точки зрения построения СБИС. В рамках биполярной технологии они имеют малые времена задержки, высокую нагрузочную способность, малую потребляемую мощность и достигают высокой степени интеграции.

Интегральная инжекционная логика И²Л

Основной областью применения СБИС является цифровая техника, так как здесь логические элементы выполняют ограниченное число операций, многократно повторяющихся. Примером может служить элемент статической памяти с произвольным доступом. Число таких элементов в системах средней сложности, например в персональных компьютерах, составляет десятки тысяч и может быть снижено, если устройство памяти работает в динамическом режиме. Разработка биполярных ИС семейства И²Л позволила существенно сократить число транзисторов, приходящихся на один бит хранимой информации. В результате были созданы системы памяти емкостью до нескольких сотен килобит, размещаемые на одном кристалле размерами приблизительно 6х6 мм.

Базовая структура семейства И²Л была разработана в конце 60-х годов как ячейка статической памяти, потребляющая малую мощность (ток 1 нА/бит). Благодаря вертикальному расположению элементов площадь, занимаемая такой ячейкой на подложке, лишь незначительно превышает площадь, необходимую для размещения одиночного транзистора. Малая потребляемая мощность и миниатюрные размеры позволяют семейству И²Л успешно конкурировать с ИС типов n-МОП и КМОП. На основе таких ИС строят быстродействующие устройства памяти и персональные компьютеры.

На рис. 7.7 схематически *Вход* показаны разрез логического элемента И²Л и его принципиальная схема. Элемент содержит транзистор *p-n-p*, который инжектирует ток в базу многоколлекторного *n-p-n* инвертора.

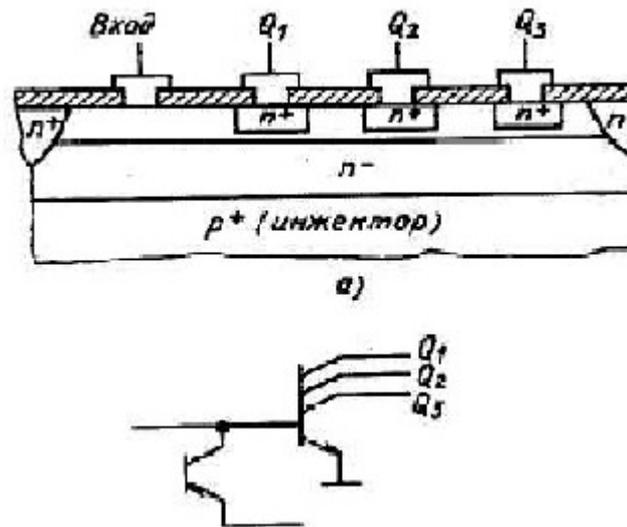


Рис. 7.8. Логика И²Л: а – поперечный разрез структуры; б – условное графическое обозначение

Используется подложка p -типа, на которую наносится эпитаксиальная пленка n -типа. Вслед за этим путем диффузии акцепторной примеси формируют базовые области вертикальных транзисторов и инжектора. Второй диффузией создают сложный коллектор. Всего требуется четыре фотошаблона, т. е. на один меньше, чем при изготовлении биполярного транзистора.

Логические элементы семейства И²Л имеют время задержки сигнала до 0,7 нс при шаге размещения 2,5 мкм. В динамическом режиме их параметры сравнимы с параметрами ИС КМОП при питающем напряжении около 1 В.

Логика с транзистором Шотки (ТЛШ)

Исследовались различные способы улучшения параметров И²Л-логики. Один из них основан на применении ионной имплантации. Другой реализуется в логике ТЛШ, которой свойственны достаточно высокое быстродействие и малое потребление мощности. Схема одного ключа логики ТЛШ показана на рис. 7.9. Диоды Шотки позволяют избегать перескоков напряжений логических уровней, а также препятствуют насыщению транзистора. Для изоляции ключа ТЛШ используются выпрямляющие диоды, которые образуются коллекторными n -областями. В результате схема содержит на один транзистор меньше, хотя для снижения токов утечки приходится использовать многоколлекторный транзистор.

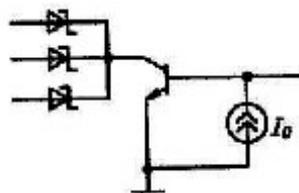


Рис. 7.9. Схема логического ключа с диодами Шотки

Инжекционная логика Шотки (И²ЛШ)

Данная структура представляет разновидность семейств, инжекционной логики. Ей присущи высокое быстродействие характерное для ИС с транзистором Шотки, и одновременно экономичность, характерная для И²Л ИС. Удастся достичь, скорости коммутации, в 5—10 раз более высокой по сравнению со скоростью коммутации И²Л ИС. Были созданы устройства И²Л ИС, которые при шаге размещения элементов 3 мкм, толщине эпитаксиальной пленки 1,2 мкм и с использованием технологии изоляции локальным оксидированием могли обеспечить время задержки 0,7 нс, потребляя ток 200 мкА.

Эмиттерно-связанная логика (ЭСЛ)

Интегральные схемы семейства ЭСЛ, называемые таю ИС с токовой связью, являются примером схем, работали в ненасыщенном режиме с малым перепадом напряжений логических уровней. В настоящее время ИС данного семейства являются наиболее быстродействующими.

Повидимому, такое положение сохранится и в будущем. Интегральные семейства схем ЭСЛ с уменьшенными геометрическими размерами могут иметь время задержки сигнала около 0,2 нс, потребляемую мощность в расчете на один инвертор 1 мВт и степень интеграции около 200 инверторов на 1 мм². На рис. 7.10 изображена принципиальная схема базового ключа типа ЭСЛ. На входе имеется два транзистора, один из которых находится в проводящем состоянии, другой — в режиме отсечки. Наименование данного семейства ИС обусловлено тем, что переход из одного логического состояния в другое происходит при появлении сигнала в общей эмиттерной цепи.

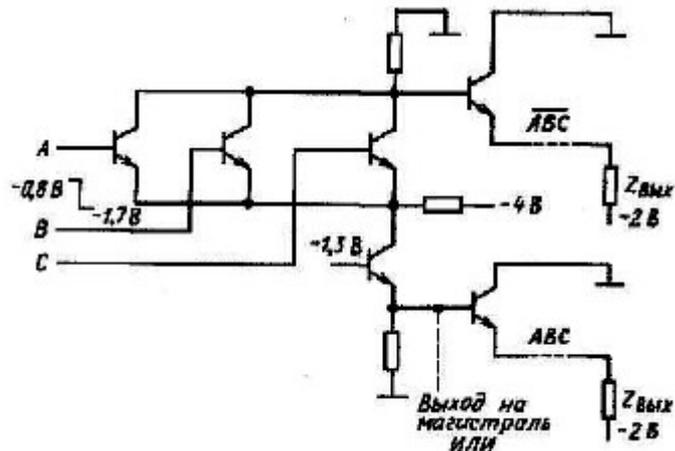


Рис. 7.10. Упрощенная схема ИС ЭСЛ

Диодно-транзисторная логика (ДТЛ)

До недавнего времени это семейство ИС применялось, редко. Интерес к нему возник вновь после того, как удалось создать логические устройства, работающие на подобном принципе, с временем задержки 1,6 нс, рассеиваемой мощностью 0,5 мВт/инвертор; степень интеграции 400 инверторов/мм². В новых ИС ДТЛ резисторы и диоды выполнены из поликристаллического кремния, что позволило увеличить площадь монокристалла, отводимую под активные элементы. На

рис. 7.11 изображена схема базового инвертора ИС ДТЛ. Устройство содержит T-образную диодную цепь, вслед за которой размещен транзисторный инвертор.

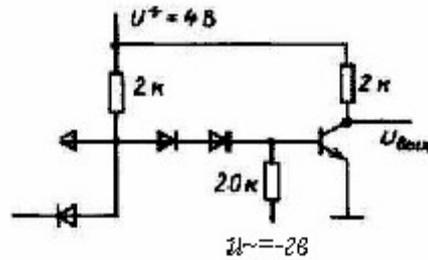


Рис. 7.11. Схема инвертора типа ДТЛ

Транзисторно-транзисторная логика (ТТЛ)

Данное семейство ИС, повидимому, не является перспективным для создания СБИС, так как свойственные ему качества можно реализовать с помощью других устройств, потребляющих меньшую мощность. Тем не менее данные ИС обладают наибольшим быстродействием среди ИС, работающих в насыщенном режиме. Быстродействие достигается за счет того, что диодная цепь, размещаемая на входе ИС ДТЛ, заменена многоэмиттерным транзистором (рис.7.12).

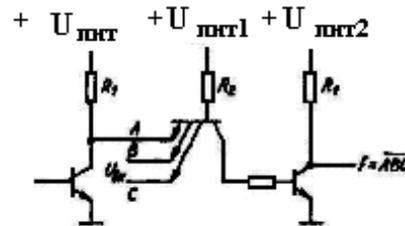


Рис. 7.12. Схема инвертора типа ТТЛ

7.6. Основные типы аналоговых биполярных ИС

В данном параграфе кратко описаны два типа биполярных ИС, которые используются в аналоговых (линейных) электронных цепях. Это дифференциальные и операционные усилители.

Как входные, так и выходные сигналы аналоговых ИС представляют напряжения, непрерывно изменяющиеся во времени. Входные сигналы амплитудой в несколько микровольт после усиления на выходе могут иметь амплитуду около 1 В.

Дифференциальные усилители

В линейных усилителях необходимо использовать непосредственные связи между отдельными звеньями. По этой причине изменения напряжений смещения по постоянному току усиливаются последующими ступенями, что приводит к появлению на выходе паразитного сигнала. Чтобы свести к минимуму этот эффект, используют дифференциальный усилитель.

Дифференциальный усилитель является основным элементом линейных ИС благодаря схемным особенностям, облегчающим процесс его изготовления. Большинство линейных ИС содержит одну или несколько усилительных ступеней такого вида. Примерами могут служить аналоговые вычислительные устройства, детекторы уровня, стабилизаторы напряжения, видеоусилители, аналоговые компараторы и операционные усилители. Дифференциальный усилитель может пропускать сигналы в широкой полосе частот, обладает высоким

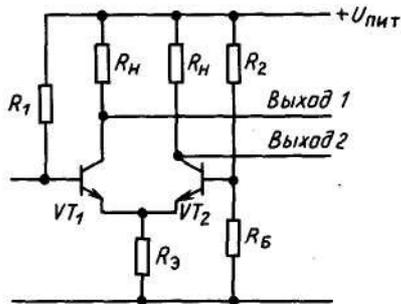


Рис. 7.13. Схема дифференциальной усилительной ступени

коэффициентом усиления и низким уровнем шума. На рис. 7.13 изображена принципиальная схема базового дифференциального усилителя, поясняющая принцип его работы. Так как рассматриваемая цепь симметрична относительно

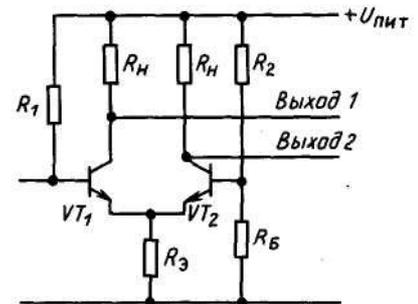


Рис. 7.13. Схема дифференциальной усилительной ступени

воображаемой прямой, проведенной между транзисторами VT_1 и VT_2 , а резисторы R_u коллекторных нагрузок одинаковы, то усилитель устойчив на всех частотах независимо от температуры. Любые изменения коэффициента передачи; или утечек в усилительном канале транзистора VT_1 компенсируются такими же изменениями в усилительном канале транзистора VT_2 .

Операционные усилители

Эти усилители состоят из нескольких ступеней, причем на входе включен дифференциальный усилитель. Операционные усилители имеют высокий коэффициент усиления, значительное входное и низкое выходное сопротивление. Происхождение термина связано с тем, что в свое время подобные усилители, выполненные на дискретных элементах, использовались в аналоговых вычислительных устройствах, которые проводили математические операции суммирования, умножения, интегрирования, дифференцирования и др. Гибкость применения, а также низкая стоимость обусловили использование данных ИС в качестве основных функциональных блоков линейных цепей, применяемых в измерительных устройствах, автогенераторах, активных фильтрах, линейных усилителях и т. д.

Технология изготовления линейных ИС в основном совпадает с планарной технологией, применяемой для производства биполярных транзисторов. Удастся реализовать все положительные качества: малые геометрические размеры, высокую надежность, низкую стоимость, слабую температурную зависимость и малый дрейф тока и напряжения.

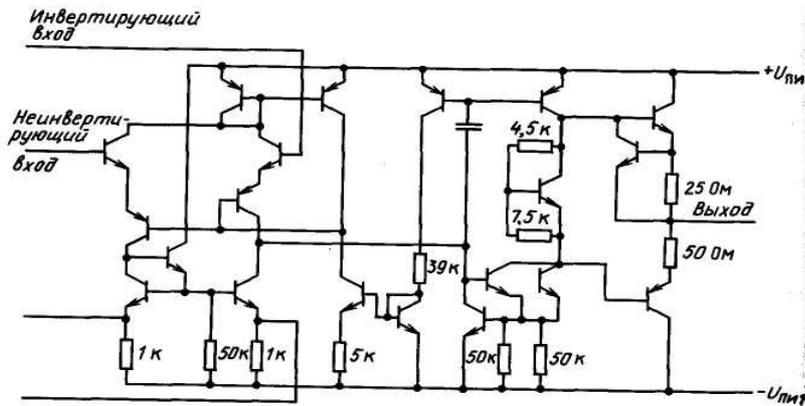


Рис. 7.14. Схема типичного операционного усилителя

На рис. 7.14 изображена принципиальная схема типичного операционного усилителя.

Первой ступенью служит дифференциальный усилитель, за ним обычно включены другие дифференциальные ступени соединенные каскадно. Это обеспечивает структурное единообразие и позволяет получать высокие значения коэффициента усиления.

Достоинства операционного усилителя и гибкость его применений проявляются при использовании внешней отрицательной обратной связи. Так как коэффициент усиления самого усилителя достаточно высок, то коэффициент передачи устройства, охваченного петлей обратной связи, зависит от характеристик внешних цепей, собранных из дискретных элементов. По этой причине рабочие характеристики системы не зависят от изменения параметров активных элементов, входящих в цепь прямого усиления. Введение отрицательной обратной связи улучшает также стабильность коэффициента усиления, повышает входное и снижает выходное сопротивление, расширяет полосу пропускания и снижает нелинейные искажения.

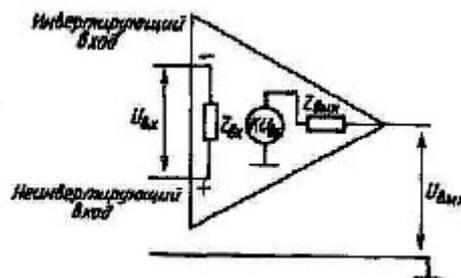


Рис. 7.15. Упрощенная схема идеального операционного усилителя

Условное графическое обозначение операционного усилителя представлено на рис. 7.15. Идеальному операционному усилителю присущи следующие свойства.

1. Коэффициент усиления напряжения со стороны инвертирующего входа положителен.
2. Коэффициент усиления напряжения со стороны инвертирующего входа отрицателен.
3. В частном случае один из входных зажимов усилителя может быть заземлен.
4. Входное сопротивление неограниченно велико.
5. Выходное сопротивление равно нулю.
6. Коэффициент усиления напряжения K_o неограниченно велик.
7. Ширина полосы пропускания бесконечно велика.
8. $u_{вых} = K_o u_{вх}$.
9. $u_{вых} = 0$, если $u_{вх неинв} = u_{вх инв}$
10. Характеристики прибора не зависят от температуры.
11. Ненасыщенное состояние восстанавливается мгновенно.
12. Выходная характеристика линейна и симметрична относительно нулевого уровня.

9. СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Гаман В.И. Физика полупроводниковых приборов.- Томск; Издательство научно – технической литературы; 2000.- 425 с.
2. Давыдов В.Н. Твердотельная электроника. Учебное пособие. ТУСУР. Томск. 2005.- 183 с.
3. Легостаев Н.С. Электроника и микроэлектроника. Учебное пособие. ТУСУР. Томск. 1999.- 176 с.
4. Шарапов А.В. Электронные цепи и микросхемотехника. Часть 1. Учебное пособие. ТУСУР. Томск. 1999.- 161 с.
5. Герасимов В.М., Скворцов В.А. Электронные цепи и микросхемотехника. Часть 2. Учебное пособие. ТУСУР. Томск. 1998.- 176 с.
6. Тугов Н.М., Глебов Б.А., Чарыков Н.А. Полупроводниковые приборы. М., Энергоатомиздат. 1990.- 576 с.
7. Викулин И.М., Стафеев В.И. Физика полупроводниковых приборов. М., «Радио и связь». 1990.- 262 с.
8. Росадо Л. Физическая электроника и микроэлектроника. М., Высшая школа. 1991.- 351 с.

Учебное пособие

Шангин А.С.

Твердотельные приборы и устройства

Усл. печ. л. Препринт
Томский государственный университет
систем управления и радиоэлектроники
634050, г.Томск, пр.Ленина, 40