

**Physics of
Semiconductor
Devices**

Second edition

S. M. Sze

Bell Laboratories,
Incorporated
Murray Hill, New Jersey

A Wiley-Interscience
Publication
John Wiley & Sons
New York, Chichester, Brisbane,
Toronto, Singapore
1981

С. Зи

Физика

полупроводниковых

приборов

В 2-х книгах

1

Перевод с английского
канд. физ.-мат. наук В. А. Гергеля
и канд. техн. наук В. В. Ракитина
под редакцией
д-ра физ.-мат. наук Р. А. Суриса

Москва «Мир» 1984

ББК 32.852
З-59
УДК 621.382

Зи С.

З-59 Физика полупроводниковых приборов: В 2-х книгах.
Кн. 1. Пер. с англ. — 2-е перераб. и доп. изд. — М.: Мир,
1984. — 456 с., ил.

Монография написана известным американским специалистом в области полупроводниковой электроники. В переводе на русский язык выходят в двух книгах. Книга 1 посвящена физике биполярных приборов (диодов, транзисторов и тиристоров) и приборов на основных носителях (полевых транзисторов с $p-n$ -переходом и барьером Шоттки).

Для научных работников и инженеров, работающих в области электроники и вычислительной техники, а также для студентов старших курсов вузов.

Э $\frac{2403000000-430}{041(01)-84}$ 157-85, ч. 1

ББК 32.852
6Ф 0.32

Редакция литературы по новой технике

© 1981, by John Wiley and Sons, Inc.
All rights reserved. Authorized translation
from English language edition published
by John Wiley and Sons, Inc.

© Перевод на русский язык, «Мир», 1984

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА

Развитие полупроводниковой электроники и расширение областей ее применения происходят столь бурными темпами, что кажется почти невероятным появление книги по физическим основам функционирования полупроводниковых приборов, которая в течение длительного времени удовлетворяла бы исследователей и разработчиков полупроводниковых устройств. Тем не менее такая книга существует, и она перед вами. Ее автор — известный специалист в области полупроводниковой электроники, сотрудник фирмы Bell Laboratories, д-р С. Зи. Первое издание его монографии, ставшее настольным для специалистов, вышло в свет в 1969 г. О популярности этой книги свидетельствует огромное число ссылок, встречающихся до сих пор в оригинальных статьях. Ее перевод¹⁾ в нашей стране уже давно стал библиографической редкостью.

Второе издание столь существенно отличается от первого, что его следует рассматривать как новую книгу. И дело здесь не только и не столько в том, что переработана и дополнена большая часть текста, библиографии и иллюстраций. За время, прошедшее с момента выпуска первого издания, полупроводниковая электроника вышла на качественно новый уровень — создана и прочно вошла в жизнь техника интегральных схем. Теперь полупроводниковый прибор не рассматривается как нечто отдельное — он мыслится как составная часть ансамбля приборов, изготовленных на одном кристалле полупроводника, т. е. как элемент интегральной схемы. Это обстоятельство нашло четкое отражение в новом издании книги. Увеличенные степени интеграции и быстродействия интегральных схем с необходимостью приводит к уменьшению размеров приборов. Характерные размеры приборов становятся сравнимыми с размерами областей пространственного заряда и длинами свободного пробега электронов. В результате возникает ряд фундаментальных особенностей, которые не принимались во внимание в ранних моделях приборов. Эта тенденция нашла отражение в новом издании книги.

Последнее десятилетие ознаменовалось мощным развитием оптоэлектроники. Ее успехи связаны в значительной мере с внедрением полупроводниковых гетеропереходов. Автору удалось не только в весьма емкой форме изложить физику фотоэлектрических приборов, но и показать перспективу их развития. То же самое можно сказать и о главах, посвященных полупроводниковым СВЧ-приборам.

Нет сомнений в том, что перевод нового издания книги будет пользоваться такой же популярностью среди широкого круга специалистов в области физики полупроводниковых приборов и инженеров — разработчиков электронных устройств, как и первое, и послужит прекрасным пособием для студентов и аспирантов, специализирующихся в этих областях.

Перевод книги выполнен Гергелем В. А. (гл. 1, 5—8, приложения), Ракиным В. В. (гл. 2—4), Фуксом Б. И. (гл. 9, 14), Зыковым Н. В. (гл. 10, 11) и Хафизовым Р. З. (гл. 12, 13).

Р. А. Суриц

¹⁾ С. Зи. Физика полупроводниковых приборов: Пер. с англ. — М.: Энергия, 1973.

ПРЕДИСЛОВИЕ АВТОРА

С момента изобретения биполярного транзистора в 1947 г. началось бурное развитие полупроводниковой электроники. В соответствии с этим быстро увеличивалось число публикаций в области физики и технологии полупроводниковых приборов. Из-за огромного объема содержащейся в них информации возникла настоятельная необходимость в книге, которая послужила бы достаточно полным введением в физику полупроводниковых приборов и одновременно содержала бы справочный материал. С этой целью в 1969 г. было опубликовано первое издание «Физики полупроводниковых приборов», которое до сих пор остается одним из основных учебных пособий по этой теме для студентов и аспирантов, а приведенные в ней фактические данные широко используются исследователями и разработчиками полупроводниковых устройств.

За прошедшее десятилетие в области полупроводниковых приборов было опубликовано более 40 000 статей, содержащих большое число новых идей относительно функционирования полупроводниковых устройств и их изготовления. Все это потребовало существенной переработки содержания книги для второго издания (обновило более 80 % материала). Из 1000 содержащихся в книге ссылок на оригинальные статьи 70 % относятся к последнему десятилетию, а 65 % из более чем 600 иллюстраций заменены новыми.

В книге рассмотрено большинство наиболее важных современных полупроводниковых приборов, которые можно разделить на четыре основных класса: биполярные, униполярные, СВЧ- и оптоэлектронные приборы. Каждому из них посвящена отдельная часть книги. Каждая глава содержит небольшое историческое введение, а затем излагаются основные физические явления, определяющие характеристики приборов данного типа, и их математическое описание. Разделы глав расположены в логической последовательности, без излишней «привязки» к оригинальным работам первоисточникам. Главы в той или иной степени независимы друг от друга, что облегчает читателю использование книги в качестве справочника.

Естественно, в процессе написания книги мне помогали и оказывали поддержку многие мои коллеги. В первую очередь я хочу выразить признательность руководству фирмы Bell Laboratories, обеспечившему мне благоприятные условия для работы над книгой. Без такой поддержки эта книга, безусловно, не могла бы быть написана. Существенную помощь мне оказали советы моих рецензентов: д-ров из фирмы Bell Laboratories Эндриуса, Асплиса, Бидла, Брюса, Болла, Чена, Фичнера, Фукуи, Гуммели, Канга, Ли, Ленселтера, Николлиана, Нихауса, Панофиса, Паоли, Райдера, Шёдзи, Смита, Торнбера, Венпла; профессоров Кэзи (Университет Дьюка), Кроуэлла (Южнокалифорнийский университет), Фенга (Национальный Тайваньский университет), Ганди (Рэнселлерский политехнический институт), Крёмера (Калифорнийский университет), Ламмерта (Принстонский университет), Мельхиора (Швейцарский федеральный технологический институт), Редикера (Массачусетский технологический институт), Тима (Венский технический университет) и д-ров Чанга (фирма ИБМ), Гиббонса (из Plessey Research Limited) и Холла (фирма General Electric).

Я благодарен Мак-Грю, Чи, Фанк и Ланчу за техническое редактирование и Лэбэйт, Стивенс и Тетельбаум за их помощь при составлении литературы. Я признателен также Шаферу из Центра информации, статики и анализа числовых данных за предоставление новейших данных по свойствам полупроводников. Я хочу поблагодарить также Мак-Карти и Мэй за неоднократную перепечатку

разделов рукописи, Холмфелта и других сотрудников чертежного отдела фирмы Bell Laboratories за подготовку нескольких сотен иллюстраций, вошедших в книгу. Во всех случаях, когда в качестве иллюстраций были использованы рисунки из оригинальных работ, приведены соответствующие ссылки. Я признателен также работникам издательства Новотны, Альдцерис, Фаркас и Флетчер за помощь в публикации книги. И наконец, я хочу выразить особую благодарность моей жене Терезе Линг-Ю, сыну Раймонду и дочери Джулии, помогавшим мне на всем протяжении работы над книгой, от перепечатки первых набросков до окончательной обработки рукописи.

Мюррей Хилл, Нью-Джерси

С. М. Зи

Май 1981

ВВЕДЕНИЕ

Книга состоит из пяти частей. Часть I (гл. 1) представляет собой сводку данных об основных физических явлениях в полупроводниках и свойствах этих материалов, которые на протяжении всей книги будут использоваться для объяснения и расчета характеристик конкретных полупроводниковых приборов. Здесь кратко рассматриваются энергетические зонные схемы, распределения носителей и процессы переноса в трех наиболее важных в настоящее время полупроводниках: германии (Ge), кремнии (Si) и арсениде галлия (GaAs). Последние и наиболее точные значения параметров этих полупроводников приведены на рисунках в главе и в соответствующих таблицах, вынесенных для удобства в приложение.

Часть II (гл. 2—4) посвящена биполярным приборам, в которых в процессах переноса участвуют оба типа носителей тока — электроны и дырки. В гл. 2 рассматриваются характеристики $p-n$ -переходов и основные технологические способы их изготовления. Теория $p-n$ -перехода представляет собой основу физики полупроводниковых приборов, поскольку $p-n$ -переход — один из основных элементов большинства структур приборов. Глава 3 посвящена биполярному транзистору, работа которого основана на взаимодействии двух близко расположенных $p-n$ -переходов. Биполярный транзистор является одним из важнейших полупроводниковых приборов. Можно считать, что с его изобретением в 1947 г. началась эра современной электроники. Работа тиристора, представляющего собой три близко расположенных и взаимодействующих $p-n$ -перехода ($p-n-p-n$ -структура), рассматривается в гл. 4. Тиристоры обладают широким диапазоном уровня переключаемых мощностей. В зависимости от назначения они могут выдерживать токи от нескольких миллиампер до тысячи ампер и напряжения свыше 5000 В.

Часть III (гл. 5—8) посвящена униполярным приборам, работу которых определяют главным образом носители одного типа (основные). Описанный в гл. 5 контакт металл—полупроводник в значительной мере подобен резкому асимметричному $p-n$ -переходу, однако, как и все приборы на основных носителях, он обладает существенно более высоким быстродействием. Кроме того, контакты между металлом и сильнолегированным полупроводником широко используются в различных приборах в качестве

омического контакта. В гл. 6 рассмотрены родственные по принципу действия полевые транзисторы с $p-n$ -переходом в качестве затвора и полевые транзисторы, в которых роль затвора играет контакт металл—полупроводник (МП-транзисторы). В каждом из них электрическое поле затвора контролирует величину тока, протекающего вдоль перехода затвора. Вопросы физики поверхности и приборы со структурой металл—окисел—полупроводник (МОП-структуры) обсуждаются в гл. 7 и 8. Отметим, что свойства поверхностных состояний важны не только для рабочих характеристик приборов этого класса, но определяют надежность и стабильность работы остальных типов полупроводниковых приборов. В гл. 7 рассмотрен сравнительно новый тип полупроводниковых приборов — приборы с зарядовой связью (ПЗС). Они представляют собой совокупность близко расположенных МОП-конденсаторов и находят широкое применение для обработки сигналов и в качестве приемников изображения. В гл. 8 описаны полевые транзисторы металл—окисел—полупроводник (МОП-транзисторы), представляющие собой важнейший элемент современных сверхбольших интегральных схем (СБИС). МОП-транзисторы широко используются в современных микропроцессорах и полупроводниковых запоминающих устройствах, содержащих тысячи отдельных элементов в одном кристалле.

Часть IV (гл. 9—11) посвящена важнейшим полупроводниковым СВЧ-приборам. В гл. 9 изложены свойства туннельных диодов, представляющих собой сильнолегированные $p-n$ -переходы, в которых величина электрического поля достаточна для квантовомеханического межзонного туннелирования. В условиях лавинного пробоя $p-n$ перехода или контакта металл—полупроводник генерируется СВЧ-излучение. Рабочие характеристики таких лавинно-пролетных диодов (ЛПД) и некоторых родственных приборов приведены в гл. 10. СВЧ-излучение может генерироваться также при междолинных переходах электронов из нижней долины зоны проводимости с высокой подвижностью в расположенную при более высокой энергии долину с низкой подвижностью. Приборы этого типа описаны в гл. 11.

Часть V (гл. 12—14) посвящена оптоэлектронным приборам, которые детектируют, излучают и преобразуют энергию светового излучения в электрическую и наоборот. Светодиоды и полупроводниковые лазеры рассмотрены в гл. 12. Те и другие в настоящее время широко используются в качестве излучателей в современных волоконно-оптических линиях связи. Различные типы полупроводниковых фотодетекторов с высокой квантовой эффективностью и быстродействием описаны в гл. 13. Мировой энергетический кризис стимулировал поиск и разработку альтернативных источников энергии. Считается, что одним из главных направлений в этой области является развитие полупроводнико-

вых солнечных батарей, обеспечивающих непосредственное преобразование солнечной энергии в электрическую с достаточно высокой эффективностью. Полупроводниковые солнечные батареи различных типов и их рабочие характеристики рассмотрены в гл. 14.

Поговорим о принятых в книге обозначениях. Для простоты некоторые символы мы используем несколько раз, т. е. для обозначения различных понятий в различных приборах. Так, например, символом α мы обозначаем коэффициент усиления по току биполярного транзистора в схеме с общей базой, коэффициент оптического поглощения в фотодетекторах и коэффициент ударной ионизации в лавинно-пролетных диодах. В противном случае нам понадобилось бы для обозначений использовать более сложные (и менее употребительные) символы, что, как нам кажется, менее удобно. Тем не менее на протяжении отдельных глав каждый используемый символ имеет единственное значение, которое сразу же определяется. Большая часть использованных символов имеет одинаковое или близкое значение на протяжении всей книги. Для удобства они приведены в приложении А.

В настоящее время электроника вообще и полупроводниковые приборы в частности развиваются настолько быстро, что сегодняшние представления завтра уже могут оказаться устаревшими. Поэтому тем более важно ясно понимать фундаментальные физические процессы в полупроводниковых приборах. Это служит физической и математической основой, позволяющей специалисту самостоятельно усваивать новые идеи и методы, возникающие в нашей быстро развивающейся области. Следует отметить, что многие полупроводниковые приборы, в особенности униполярные и оптоэлектронные приборы, находятся в стадии интенсивного развития [1], и в настоящее время мы еще далеко не полностью снимаем их предельные возможности. Настоящая книга предназначена для того, чтобы дать читателю физические основы, необходимые для самостоятельной деятельности в области полупроводниковой электроники. Более подробные сведения по отдельным вопросам можно получить в оригинальных работах, список важнейших из которых приводится в конце каждой главы.

ЛИТЕРАТУРА

1. S. M. Sze Semiconductor Device Development in the 1970s and 1980s — A Perspective, IEEE Tech. Dig., Int. Electron Device Meet., Washington, D. C., Dec., 1980.

Часть I

Физика полупроводников

Глава I

ФИЗИКА И СВОЙСТВА ПОЛУПРОВОДНИКОВ

1.1. ВВЕДЕНИЕ

Физика полупроводниковых приборов, естественно, определяется физическими свойствами самих полупроводниковых материалов. В данной главе содержатся самые необходимые для понимания работы полупроводниковых приборов сведения. Для более детального изучения физики полупроводников следует воспользоваться стандартными учебниками¹⁾ и обзорными статьями Давилэпа [1], Маделунга [2], Молла [3], Мосса [4] и Смита [5].

Основное внимание в главе уделено трем важнейшим полупроводниковым материалам: германию (Ge), кремнию (Si) и арсениду галлия (GaAs). Германий и кремний — наиболее изученные в настоящее время полупроводники. В последние годы интенсивно изучается арсенид галлия, физические свойства которого отличаются от свойств кремния и германия. В частности, структура его энергетических зон, допускающая возможность прямых оптических переходов, интересна с точки зрения оптоэлектронных приложений, а междолинные переходы и высокая подвижность электронов — для генерации СВЧ-излучения.

1.2. КРИСТАЛЛИЧЕСКАЯ СТРУКТУРА

Кристаллическую решетку определяют три базисных вектора \mathbf{a} , \mathbf{b} и \mathbf{c} , таких, что любая трансляция на вектор

$$\mathbf{R} = m\mathbf{a} + n\mathbf{b} + p\mathbf{c}, \quad (1)$$

представляющий собой линейную суперпозицию базисных векторов (m , n и p — целые числа), переводит кристаллическую решетку саму в себя [6].

¹⁾ Для этой цели можно рекомендовать отечественные монографии, например: В. Л. Бонч-Бруевич, С. Г. Калашников, Физика полупроводников, Наука, Москва, 1977; А. И. Ансельм, Введение в теорию полупроводников, Наука, Москва, 1982. — *Прим. перев.*

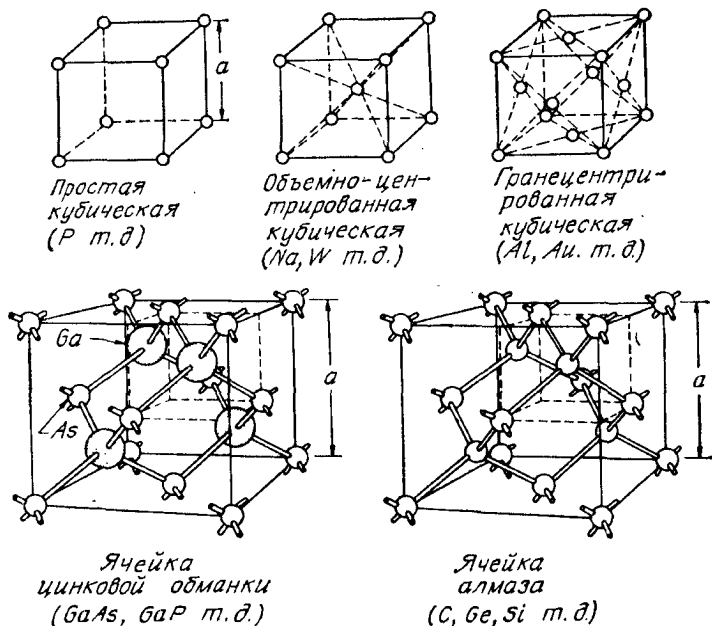


Рис. 1. Основные элементарные ячейки прямых решеток и кристаллическая структура ряда элементарных полупроводников и полупроводниковых соединений (a — постоянная решетки).

На рис. 1 показаны построенные на базисных векторах простейшие элементарные кристаллические ячейки. Большинство важнейших полупроводников имеет кристаллическую структуру типа алмаза или цинковой обманки, которые относятся к тетраэдрическим фазам, где каждый атом окружен четырьмя эквидистантными ближайшими соседями, расположенными в вершинах соответствующего тетраэдра. Связь между двумя ближайшими соседями обусловлена парой электронов с противоположными спинами. Решетки алмаза и цинковой обманки можно представить как две гранецентрированные кубические решетки, сдвинутые относительно друг друга на четверть объемной диагонали элементарной ячейки. В алмазоподобных полупроводниках (таких, как кремний) в узлах той и другой подрешетки находятся атомы кремния. В решетках типа цинковой обманки (например, в арсениде галлия) одну подрешетку составляют атомы галлия, а другую — атомы мышьяка. Арсенид галлия является соединением типа $A^{III}B^V$, поскольку образован элементами III и V групп периодической системы Менделеева. Большинство полупроводниковых соединений типа $A^{III}B^V$ кристаллизуется в решетку типа цинковой обманки [2, 7]. Многие полупроводники, однако, а среди

них и некоторые соединения $A^{III}B^V$, кристаллизуются в решетку типа вюрцита либо каменной соли. Решетка вюрцита показана на рис. 2, *а*. Ее также можно представить как две вставленные друг в друга плотноупакованные гексагональные подрешетки (например, кадмия и серы в случае CdS). Как и в решетках типа цинковой обманки, в структуре вюрцита отдельный атом также находится в тетраэдрическом окружении четырех ближайших соседей. На рис. 2, *б* показана решетка каменной соли, которую можно рассматривать как две гранецентрированные кубические решетки, вставленные друг в друга. В этой структуре каждый атом окружен шестью ближайшими соседями.

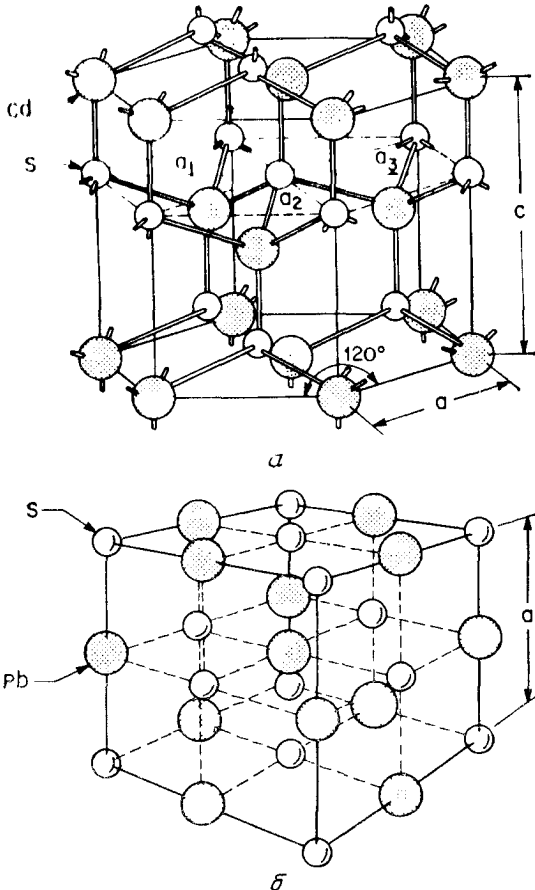


Рис. 2. Две элементарные ячейки решеток полупроводниковых соединений, *а* — ячейка решетки вюрцита (CdS , ZnS и т. д.); *б* — ячейка решетки каменной соли (PbS , $PbTe$ и т. д.).

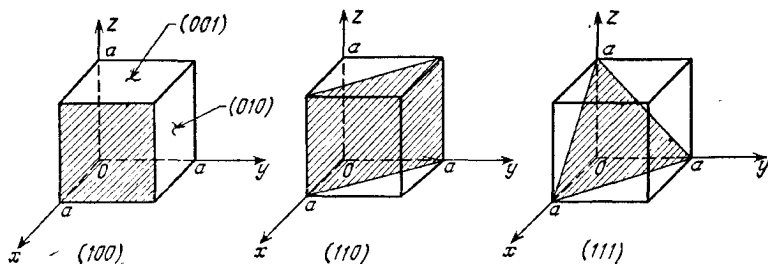


Рис. 3. Индексы Миллера основных плоскостей кубического кристалла.

Значения постоянных решетки важнейших полупроводниковых материалов и другие данные об их кристаллической структуре [8, 9] приведены в приложении Г. Отметим, что некоторые из полупроводниковых соединений (такие, как сульфид цинка или сульфид кадмия) могут в зависимости от условий кристаллизоваться и в структуру цинковой обманки, и в структуру вюрцита.

Для данного набора базисных векторов прямой решетки определим базисные векторы обратной решетки \mathbf{a}^* , \mathbf{b}^* , \mathbf{c}^* следующим образом:

$$\mathbf{a}^* \equiv 2\pi \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}}, \quad \mathbf{b}^* \equiv 2\pi \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}}, \quad (2)$$

$$\mathbf{c}^* \equiv 2\pi \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}}.$$

При этом $\mathbf{a} \cdot \mathbf{a}^* = 2\pi$; $\mathbf{a} \cdot \mathbf{b}^* = 0$ и т. д., а произвольный вектор обратной решетки имеет вид

$$\mathbf{G} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*, \quad (3)$$

где h , k , l — целые числа. Отсюда следует, что скалярное произведение $\mathbf{G}\mathbf{R} = 2\pi \cdot C$ (где C — целое число). Следовательно, любой вектор обратной решетки перпендикулярен соответствующим плоскостям прямой решетки, а объем элементарной ячейки обратной решетки обратно пропорционален объему элементарной ячейки прямой решетки, т. е. $V_c^* = (2\pi)^3/V_c$, где $V_c \equiv \mathbf{a} \cdot \mathbf{b} \times \mathbf{c}$.

Обычно положение атомных плоскостей в кристаллической решетке определяют индексами Миллера. Для этого нужно сначала найти точки, в которых рассматриваемая плоскость пересекает координатные оси, и записать их в единицах постоянных решетки, а затем взять обратные значения полученных целых чисел и привести их к наименьшему целому, кратному каждому из этих значений. Полученный результат заключают в круглые

скобки (hkl). Это и есть индексы Миллера отдельной плоскости или системы параллельных плоскостей. На рис. 3 показаны основные плоскости кубического кристалла и соответствующие индексы Миллера. Приведем также некоторые другие принятые обозначения:

- $(\bar{h}kl)$ — для плоскости, пересекающей ось x при отрицательных значениях этой координаты.
- $\{hkl\}$ — для плоскостей эквивалентной симметрии. Например, для кубического кристалла $\{100\}$ соответствует плоскостям (100) , (010) , (001) , $(\bar{1}00)$, $(0\bar{1}0)$ и $(00\bar{1})$.
- $[hkl]$ — для кристаллографического направления; например, $[100]$ — направление вдоль оси x .
- $\langle hkl \rangle$ — для совокупности всех эквивалентных направлений.
- $\{a_1, a_2, a_3\}$ для гексагональных решеток, где обычно используются четыре оси (рис. 2, a), причем ось c соответствует направлению $\{0001\}$.

Для двух элементарных полупроводников (германия и кремния) атомные плоскости $\{111\}$ являются плоскостями наиболее легкого разрушения или скола. В противоположность этому арсенид галлия с такой же кристаллической структурой и сравнительно небольшой ионной компонентой в атомных связях скалывается по плоскости $\{110\}$.

Элементарную ячейку обратной решетки кристалла можно представить в виде ячейки Вигнера—Зейтца. Для ее построения нужно провести перпендикулярные плоскости через середины отрезков, соединяющих выбранный центр с ближайшими эквивалентными узлами обратной решетки. На рис. 4, a показана построенная ячейка Вигнера—Зейтца для гранцентрированной кубической решетки $[10]$. Сначала проводятся отрезки из центральной точки куба (Γ) ко всем восьми его вершинам, а затем через середины этих отрезков проводятся соответствующие перпендикулярные плоскости; в конечном итоге получается усеченный октаэдр внутри куба — ячейка Вигнера—Зейтца. Можно показать $[11]$, что для гранцентрированной кубической прямой решетки с постоянной a обратной является объемно-центрированная кубическая решетка с постоянной $4\pi/a$. Следовательно, показанная на рис. 4, a ячейка Вигнера—Зейтца является элементарной ячейкой обратной решетки, соответствующей гранцентрированной кубической прямой решетке. Аналогичным образом можно построить ячейку Вигнера—Зейтца и для гексагональной структуры $[12]$ (рис. 4, b). Символы, использованные на рис. 4, заимствованы из теории групп. Некоторые из них использованы в разд. 1.3.

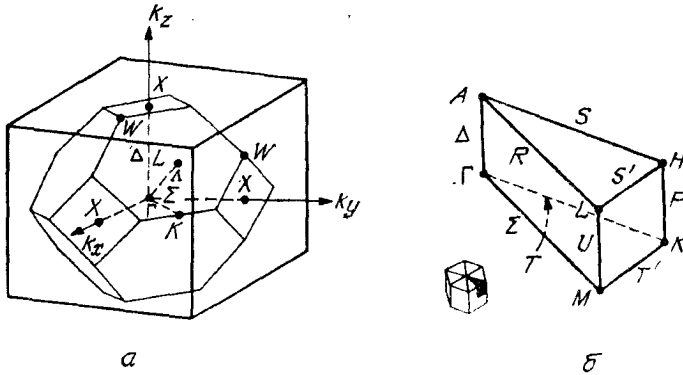


Рис. 4. Зоны Бриллюэна для решетки алмаза и цинковой обманки (а) и для решетки вюрцита (б) [10, 12].

Показаны главные точки и линии симметрии:

Γ : $\frac{2\pi}{a}(0, 0, 0)$ — центр зоны; L : $\frac{2\pi}{a}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$ — точки пересечения осей $\langle 111 \rangle$

(линия Δ) с краем зоны; X : $\frac{2\pi}{a}(0, 0, 1)$ — точки пересечения осей $\langle 100 \rangle$ (линия Δ)

с краем зоны; K : $\frac{2\pi}{a}\left(\frac{3}{4}, \frac{3}{4}, 0\right)$ — точки пересечения осей $\langle 110 \rangle$ (линия Σ) с краем зоны.

1.3. ЭНЕРГЕТИЧЕСКИЕ ЗОНЫ

Зонный характер энергетических спектров кристаллических твердых тел, т. е. зависимости энергии электрона от его импульса $E(k)$, следует из одночастичного уравнения Шредингера [11, 13]

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \right] \varphi_{\mathbf{k}}(\mathbf{r}) = E_{\mathbf{k}} \varphi_{\mathbf{k}}(\mathbf{r}). \quad (4)$$

Поскольку потенциал $V(\mathbf{r})$ периодически зависит от координат (с пространственными периодами, равными соответствующим постоянным кристаллической решетки), то, согласно известной теореме Блоха, решения уравнения (4) имеют вид

$$\varphi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{kr}} U_n(\mathbf{k}, \mathbf{r}), \quad (5)$$

где $U_n(\mathbf{k}, \mathbf{r})$ — периодические функции координат с периодом прямой решетки, а n — индекс соответствующей зоны. Выражения (5) называют блоховскими функциями. Из теоремы Блоха следует, что энергия $E_{\mathbf{k}}$ является периодической функцией волнового вектора \mathbf{k} в пространстве обратной решетки, т. е. $E_{\mathbf{k}} = E_{\mathbf{k}+\mathbf{G}}$, где \mathbf{G} — произвольный вектор обратной решетки, определяемый формулой (3). Таким образом, функцию $E_{\mathbf{k}}$ при данном значении зонного индекса n полностью определяют ее значения при волновых векторах, лежащих в пределах элементарной ячейки

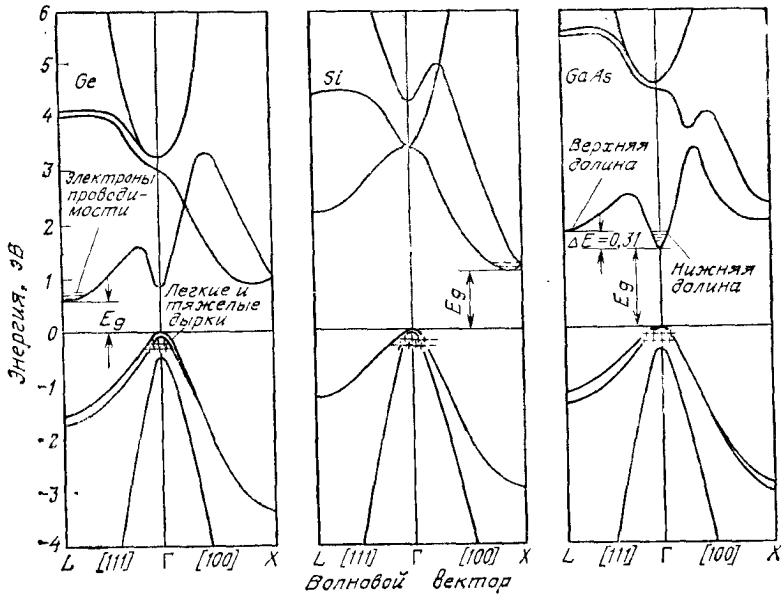


Рис. 5. Структура энергетических зон Ge, Si и GaAs [77].

обратной решетки кристалла. Обычно при этом используют ячейку Вигнера—Зейтца. Применительно к электронным энергетическим спектрам ее называют зоной Бриллюэна или первой зоной Бриллюэна [10]. Очевидно, что любой волновой вектор k можно привести к точке в первой зоне Бриллюэна, вычитая соответствующий вектор обратной решетки, и в дальнейшем рассматривать энергетические состояния только в этом редуцированном объеме k -пространства.

Зоны Бриллюэна для решетки алмаза и цинковой обманки такие же, как и зона, показанная на рис. 4, а. На рис. 4 также отмечены главные точки и линии симметрии.

Зонные энергетические спектры твердых тел рассчитываются теоретически с помощью различных приближенных методов. Для полупроводниковых кристаллов наиболее часто используются метод ортогонализированных плоских волн [14, 15], метод псевдопотенциала [16] и k - p -метод [5]. Результаты теоретических расчетов энергетических спектров Ge, Si и GaAs [17] приведены на рис. 5. Обратим внимание, что в спектрах каждого из этих полупроводников имеется область запрещенных энергий, в которой не существует электронных состояний. Эти состояния образуют разрешенные зоны с энергиями выше и ниже этой энергетической щели. Верхнюю разрешенную энергетическую зону называют зоной

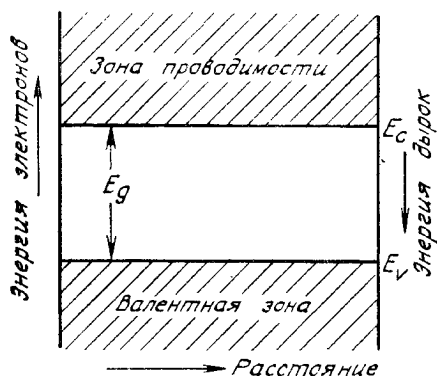


Рис. 6. Упрощенная зонная схема полупроводника.

проводимости, а нижнюю — валентной. Расстояние между нижним (дном зоны проводимости) и верхним (потолком валентной зоны) краями называется шириной запрещенной зоны E_g . Эта величина — один из важнейших параметров полупроводников. Прежде чем детально изучать структуру энергетических зон полупроводников, рассмотрим сначала упрощенную зонную схему, показанную на рис. 6. Здесь, как обычно, положение дна зоны проводимости обозначено символом E_C , а потолка валентной зоны — E_V . Энергия электрона считается увеличивающейся в направлении снизу вверх, в то время как для дырок энергия тем больше, чем ниже она отстоит от потолка валентной зоны. Значения ширины запрещенной зоны E_g для основных полупроводниковых материалов [9, 18] приведены в приложении Д.

Валентная зона в кристаллах со структурой цинковой обманки состоит из четырех подзон (если пренебречь спином в уравнении Шредингера), двукратно вырожденных по спину. Три из них вырождены в центре зоны при $k = 0$ (Γ -точка) и формируют верхний край валентной зоны, а четвертая подзона образует ее дно. Спин-орбитальное взаимодействие частично снимает вырождение при $k = 0$ и приводит к отщеплению одной подзоны. Как видно из рис. 5, две оставшиеся подзоны у потолка валентной зоны можно аппроксимировать параболическими зависимостями с различной кривизной. Зона, которой отвечает меньшая производная $\partial^2 E / \partial k^2$, называется зоной тяжелых дырок, а зона с большим значением $\partial^2 E / \partial k^2$ — зоной легких дырок. В общем случае эффективная масса является тензором, компоненты которого определяются соотношением

$$\frac{1}{m_{ij}^*} \equiv \frac{1}{\hbar^2} \frac{\partial^2 E(\mathbf{k})}{\partial k_i \partial k_j}. \quad (6)$$

Значения эффективных масс для важнейших полупроводников приведены в приложении Д.

Зона проводимости также состоит из нескольких подзон (рис. 5). Дно зоны проводимости может быть расположено либо на осях $\langle 111 \rangle$ (А), либо на осях $\langle 100 \rangle$ (Б), либо при $k = 0$ (Г). Более точных выводов о положении энергетических минимумов на этих осях из соображений симметрии сделать не удастся. Как показывает эксперимент, в Ge имеется восемь эквивалентных минимумов на осях $\langle 111 \rangle$, в Si — шесть на осях $\langle 100 \rangle$, а в GaAs дно зоны проводимости находится при $k = 0$. На рис. 7 показаны соответствующие изоэнергетические поверхности [19]. В Ge границы зоны Бриллюэна проходят точно по середине эллипсоидов, так что от каждого из них в первой зоне Бриллюэна остается половина (четыре полных эллипсоида в ячейке Вигнера—Зейтца), и, следовательно, эти поверхности постоянной энергии центрированы в L -точке. В Si имеется шесть эквивалентных эллипсоидов, центрированных на осях $\langle 100 \rangle$ на расстоянии от центра зоны, равном примерно трем четвертям полной длины соответствующей оси. В GaAs изоэнергетическими поверхностями являются сферы с центром в центре зоны Бриллюэна. Из экспериментальных данных для параболических зон можно получить эффективные массы электронов: одну для GaAs и две [продольную m_l^* (вдоль оси симметрии) и поперечную m_t^* (в направлении, перпендикулярном этой оси)] для Ge и Si. Значения электронных эффективных масс приведены в приложении Д.

При комнатной температуре и нормальном атмосферном давлении ширина запрещенной зоны германия составляет 0,66 эВ, кремния — 1,12 эВ, а арсенида галлия — 1,42 эВ. Отметим, что эти значения найдены для материалов с высокой степенью чистоты. В сильно легированных материалах ширина запрещенной зоны немного меньше. Как показывают экспериментальные результаты, ширина запрещенной зоны большинства полупроводников уменьшается с ростом температуры. Температурные зависимости для Ge, Si и GaAs приведены на рис. 8 [20]. При нулевой температуре (0 К) в этих полупроводниках ширина запрещенной зоны равна соответственно 0,743, 1,17 и 1,519 эВ. Температурную зависимость

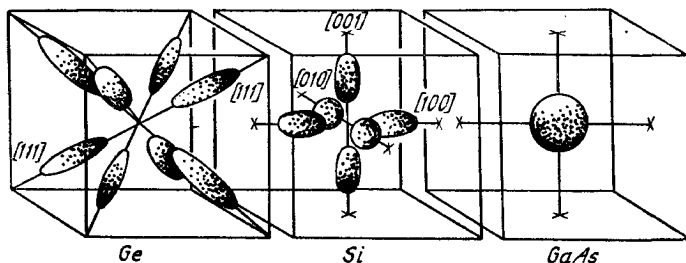


Рис. 7. Форма и расположение изоэнергетических поверхностей в Ge, Si и GaAs [19].

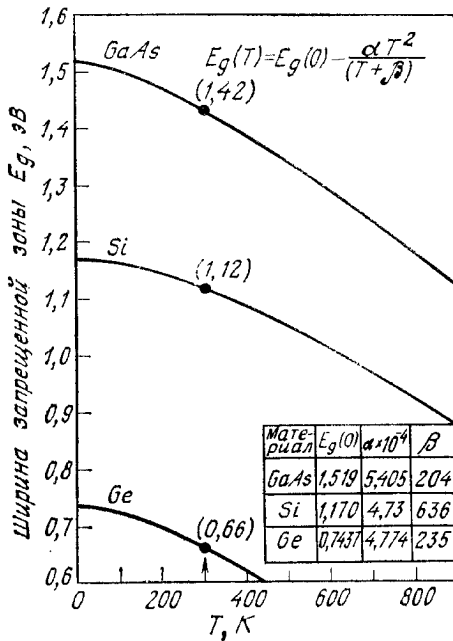


Рис. 8. Температурная зависимость ширины запрещенной зоны в Ge, Si и GaAs [20].

ширины запрещенной зоны в этих полупроводниках можно аппроксимировать универсальной функцией

$$E_g(T) = E_g(0) - \alpha T^2 / (T + \beta).$$

Числовые значения параметров $E_g(0)$, α и β приведены в таблице на рис. 8. Отметим, что для этих основных полупроводниковых материалов температурный коэффициент dE_g/dT отрицателен. В некоторых полупроводниках, однако, производная dE_g/dT положительна. Например, в PbS (приложение Д) ширина запрещенной зоны увеличивается от 0,286 эВ при $T = 0$ К до 0,41 эВ при $T = 300$ К. При комнатной температуре ширина запрещенной зоны в Ge и GaAs увеличивается с давлением: в Ge $dE_g/dP = 5 \cdot 10^{-6}$ эВ/(кг/см²), а в GaAs $dE_g/dP \approx 12,6 \cdot 10^{-6}$ эВ/(кг/см²) [21]. В кремнии ширина запрещенной зоны с ростом давления уменьшается ($dE_g/dP = -2,4 \cdot 10^{-6}$ эВ/(кг·см²)).

1.4. КОНЦЕНТРАЦИЯ НОСИТЕЛЕЙ ПРИ ТЕРМОДИНАМИЧЕСКОМ РАВНОВЕСИИ

На рис. 9 показаны условные картины электронных связей в кремниевых кристаллах. На рис. 9, а, иллюстрирующем свойства очень чистого кристалла (собственного кремния), все атомы, составляющие решетку, являются атомами кремния. Каждый из

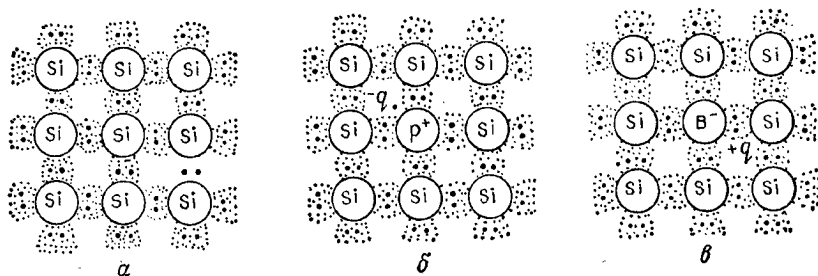


Рис. 9. Условные картины электронных связей в кристалле.

а — собственный Si; *б* — один из атомов Si замещен фосфором (донором); *в* — один из атомов Si замещен бором (акцептором).

них имеет четыре валентных электрона, участвующих в ковалентных связях с четырьмя соседними атомами (рис. 1). На рис. 9, б, иллюстрирующем полупроводник *n*-типа, один из кремниевых атомов замещен атомом фосфора, содержащим пять валентных электронов. Четыре электрона образуют связи с окружающими кремниевыми атомами, а пятый электрон «поставляется» в зону проводимости. Поэтому примесные атомы замещения типа фосфора называют «донорами». На рис. 9, в, иллюстрирующем полупроводник *p*-типа, один из атомов кремния замещен атомом бора, имеющим три валентных электрона. Оказывается, что бор может захватить недостающий для образования четырех атомных связей электрон у близлежащих кремниевых атомов. В результате в валентной зоне кремния появляется положительно заряженная «дырка». Такие примесные атомы называются «акцепторами».

1.4.1. Собственный полупроводник

Рассмотрим сначала случай собственного полупроводника. В условиях равновесия концентрация электронов в зоне проводимости равна

$$n = \int_{E_C}^{E_{top}} N(E) F(E) dE. \quad (7)$$

Здесь E_C и E_{top} — энергии дна и потолка зоны проводимости, $N(E)$ — плотность состояний на единичный интервал энергии, которую при достаточно малой концентрации и низких температурах можно считать примерно равной плотности состояний у дна зоны проводимости:

$$N(E) = M_C \frac{\sqrt{2}}{\pi^2} \frac{(E - E_C)^{1/2}}{\hbar^3} (m_{de})^{3/2}. \quad (8)$$

В этой формуле M_c — число эквивалентных минимумов в зоне проводимости, m_{de} — электронная эффективная масса плотности состояний [5]:

$$m_{de} = (m_1^* m_2^* m_3^*)^{1/3}, \quad (9)$$

где m_1^* , m_2^* , m_3^* — эффективные массы по главным осям эллипсоидов постоянной энергии. Например, в кремнии $m_{de} = (m_1^* m_2^*)^{1/3}$. Величина $F(E)$ в выражении (7) представляет собой функцию распределения Ферми—Дирака

$$F(E) = \frac{1}{1 + \exp\left(\frac{E - E_F}{kT}\right)}, \quad (10)$$

где k — постоянная Больцмана, T — абсолютная температура, E_F — энергия Ферми, величина которой определяется из условия электронейтральности (разд. 1.4.3).

Интеграл в выражении (7) можно представить в виде

$$n = N_c \frac{2}{\sqrt{\pi}} F_{1/2}\left(\frac{E_F - E_c}{kT}\right), \quad (11)$$

где N_c — эффективная плотность состояний в зоне проводимости:

$$N_c \equiv 2 \left(\frac{2\pi m_{dc} kT}{h^2}\right)^{3/2} M_c, \quad (12)$$

а $F_{1/2}(\eta_i)$ — интеграл Ферми—Дирака (рис. 10) [22]. В случае Больцмановской статистики (невырожденный полупроводник), когда уровень Ферми лежит ниже дна зоны проводимости, по крайней мере на несколько kT , интеграл Ферми приближенно равен $\sim \sqrt{\pi} e^{\eta_i/2}$. При этом из выражения (11) получаем

$$n = N_c \exp\left(-\frac{E_c - E_F}{kT}\right). \quad (13)$$

Аналогичным образом находим выражение для концентрации дырок в валентной зоне

$$p = N_v \frac{2}{\sqrt{\pi}} F_{1/2}\left(\frac{E_v - E_F}{kT}\right). \quad (14)$$

Здесь N_v — эффективная плотность состояний в валентной зоне:

$$N_v \equiv 2 \left(\frac{2\pi m_{dv} kT}{h^2}\right)^{3/2}, \quad (15)$$

m_{dv} — эффективная масса плотности состояний дырок [5]:

$$m_{dv} = (m_{iv}^{*3/2} + m_{hv}^{*3/2})^{2/3}, \quad (16)$$

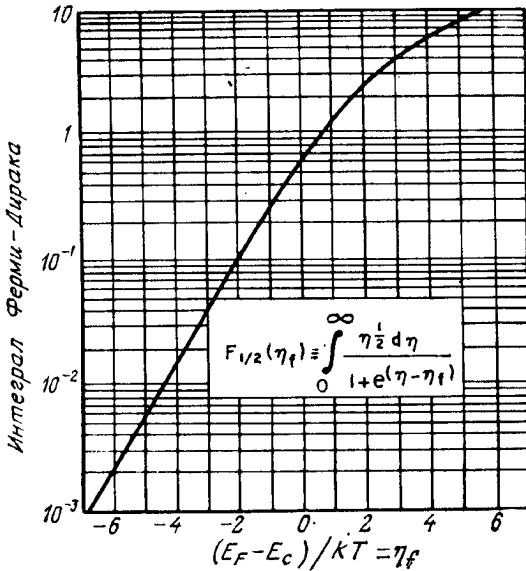


Рис. 10. Интеграл Ферми — Дирака $F_{1/2}$ как функция энергии Ферми [22].

где m_{lh}^* и m_{hh}^* — массы легких и тяжелых дырок. В невырожденном случае

$$p = N_V \exp\left(-\frac{E_F - E_V}{kT}\right). \quad (17)$$

При конечных температурах в полупроводнике непрерывно происходит процесс теплового возбуждения электронов из валентной зоны в зону проводимости. Этот процесс уравнивается рекомбинацией электронов из зоны проводимости и дырок из валентной зоны. В собственном полупроводнике число возбужденных электронов в зоне проводимости точно равно числу дырок, оставшихся в валентной зоне, т. е. $n = p = n_i$. Из этого условия с помощью формул (13) и (17) для уровня Ферми в собственном полупроводнике получим

$$\begin{aligned} E_F = E_i &= \frac{E_C + E_V}{2} + \frac{kT}{2} \ln\left(\frac{N_V}{N_C}\right) = \\ &= \frac{E_C + E_V}{2} + \frac{3kT}{4} \ln\left(\frac{m_{lh}}{m_{de} M_C^{2/3}}\right). \end{aligned} \quad (18)$$

Отсюда видно, что в собственном полупроводнике уровень Ферми лежит около середины запрещенной зоны.

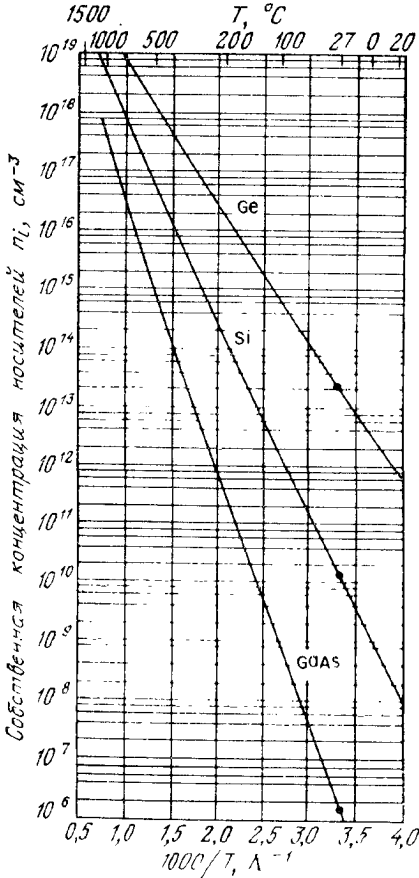


Рис. 11. Температурная зависимость собственной концентрации носителей в Ge, Si и GaAs [20].

Для собственной концентрации носителей n_i из выражений (13), (17) и (18) получим

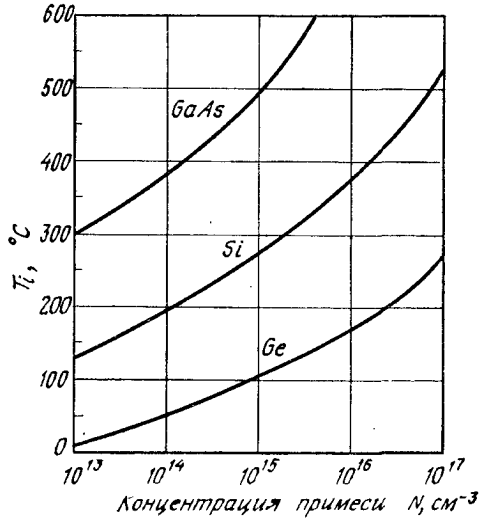
$$np = n_i^2 = N_C N_V \exp(-E_g/kT), \quad (19)$$

или

$$\begin{aligned} n_i &= \sqrt{N_C N_V} e^{-E_g/2kT} \\ &= 4,9 \cdot 10^{15} \left(\frac{m_{de} m_{dh}}{m_0^2} \right)^{3/4} M_C^{1/2} M_V^{1/2} e^{-E_g/2kT}, \end{aligned} \quad (19a)$$

где $E_g \equiv E_C - E_V$ — ширина запрещенной зоны, m_0 — масса свободного электрона. Температурные зависимости собственной концентрации n_i в Ge, Si и GaAs [20, 23] приведены на рис. 11. Как и следовало ожидать, чем шире запрещенная зона, тем меньше собственная концентрация n_i .

Рис. 12. Зависимость собственной температуры T_i от степени легирования.



При комнатных температурах собственная концентрация носителей тока n_i довольно мала по сравнению с уровнями легирования, характерными для полупроводниковой технологии. Однако с повышением температуры n_i быстро увеличивается. Так, например, в кремнии собственная концентрация удваивается при повышении температуры на каждые 11°C . Следовательно, при достаточно высоких температурах термогенерация становится доминирующим процессом, определяющим концентрацию носителей тока. На рис. 12 приведены зависимости собственной температуры T_i от степени легирования⁴⁾. Ниже значения T_i концентрация основных носителей в полупроводнике сравнительно слабо зависит от температуры, а выше T_i концентрация носителей тока увеличивается с ростом температуры экспоненциально. Собственная температура T_i является довольно важным параметром. Как показано в следующих главах, она имеет непосредственное отношение к явлению шнурования тока и вторичному пробоев в полупроводниковых приборах.

1.4.2. Доноры и акцепторы

При легировании полупроводника донорными или акцепторными примесями вводятся примесные уровни. Донорный уровень определяется как нейтральный при заполнении электроном, и

⁴⁾ Эта величина определяется как температура, при которой собственная концентрация носителей n_i сравнивается с концентрацией легирующей примеси. — Прим. перев.

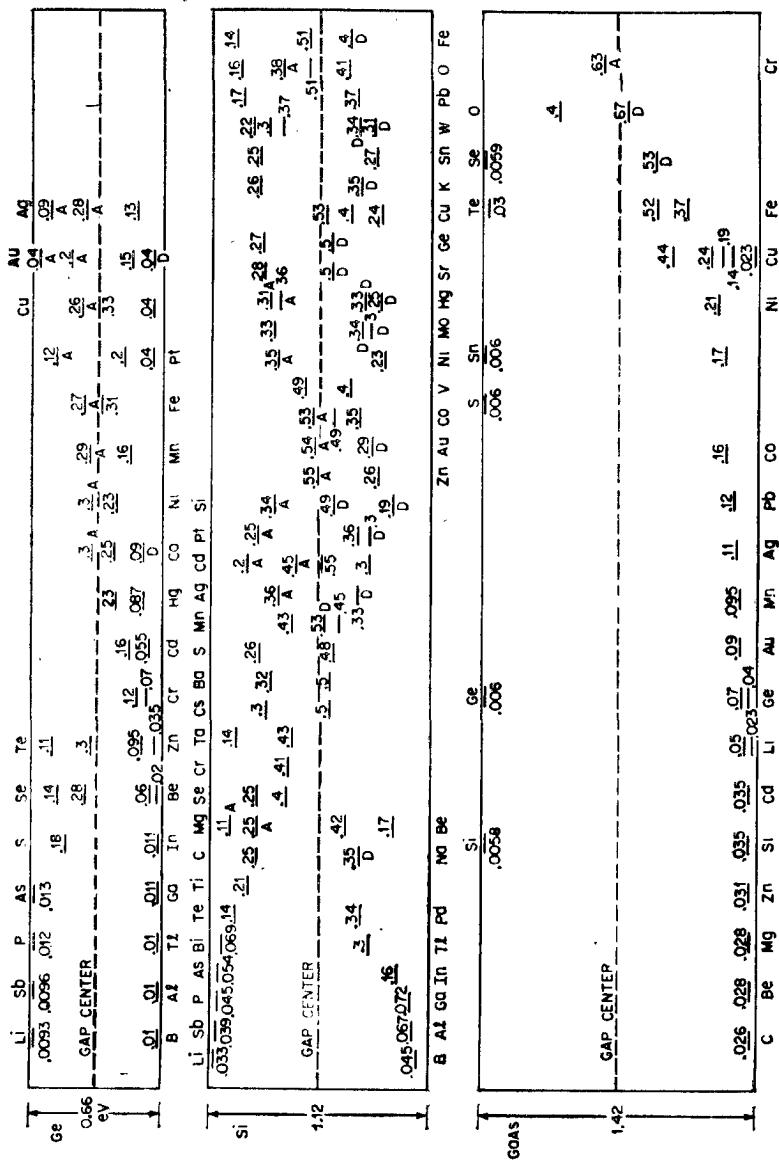


Рис. 13. Экспериментальные значения энергии ионизации различных примесных состояний в Ge, Si и GaAs. Энергия уровней, находящихся в нижней половине запрещенной зоны, отсчитана от края валентной зоны; все уровни акцепторные, за исключением донорных уровней, отмеченных буквой D. В верхней половине запрещенной зоны энергия уровней отсчитывается от дна зоны проводимости; все уровни донорные (за исключением акцепторных уровней, отмеченных буквой A [24, 27, 28]).

положительно заряженный в том случае, когда он пустой. Акцепторный уровень нейтрален в пустом состоянии и отрицательно заряжен при заполнении электроном.

Простейшей теоретической моделью примесных уровней в полупроводниках является водородоподобная модель. Энергия ионизации атома водорода равна

$$E_H = \frac{m_0 q^4}{32\pi^2 \epsilon_0^2 \hbar^2} = 13,6 \text{ эВ}, \quad (20)$$

где ϵ_0 — диэлектрическая постоянная вакуума. Энергия ионизации донора E_d получается отсюда при замене m_0 на эффективную массу электрона в зоне проводимости [5]:

$$m_{ce} = 3 \left(\frac{1}{m_1^*} + \frac{1}{m_2^*} + \frac{1}{m_3^*} \right)^{-1}, \quad (21)$$

а ϵ_0 на ϵ_s , где ϵ_s — диэлектрическая проницаемость полупроводника. Это даст

$$E_d = \left(\frac{\epsilon_0}{\epsilon_s} \right)^2 \left(\frac{m_{ce}}{m_0} \right) E_H. \quad (22)$$

Рассчитанная по формуле (22) энергия ионизации донорного состояния в Ge, Si и GaAs равна 0,006, 0,025 и 0,007 эВ соответственно. Аналогичным образом водородоподобная модель используется для расчета энергии ионизации акцепторного уровня. При этом валентная зона с отсутствующим в ней электроном рассматривается как суперпозиция полностью заполненной валентной зоны и изображающей «дырки», двигающейся в центральном поле отрицательно заряженного акцепторного примесного атома. Рассчитанная таким образом энергия связи акцепторного уровня (отсчитывается вверх от потолка валентной зоны) равна 0,015 эВ для Ge, 0,05 эВ для Si и примерно 0,05 эВ для GaAs.

Упрощенная водородоподобная модель, конечно, не позволяет точно рассчитать энергию связи примесного состояния. В особенности это относится к глубоким уровням в полупроводниках [24—26]. Однако для мелких уровней, таких, как В, Р и Si, выражение (22) дает правильные по порядку величины значения энергии ионизации. На рис. 13 приведены экспериментальные значения энергий ионизации различных примесных атомов в Ge, Si и GaAs [24—28]. Отметим, что некоторые примеси дают в запрещенной зоне полупроводника несколько различных уровней. Так, например, золото в Ge дает три акцепторных уровня и один донорный [29].

1.4.3. Определение положения уровня Ферми

Из выражения (18) видно, что уровень Ферми в собственном полупроводнике лежит вблизи середины запрещенной зоны. Этот случай показан на рис. 14, а, где слева направо приведены упро-

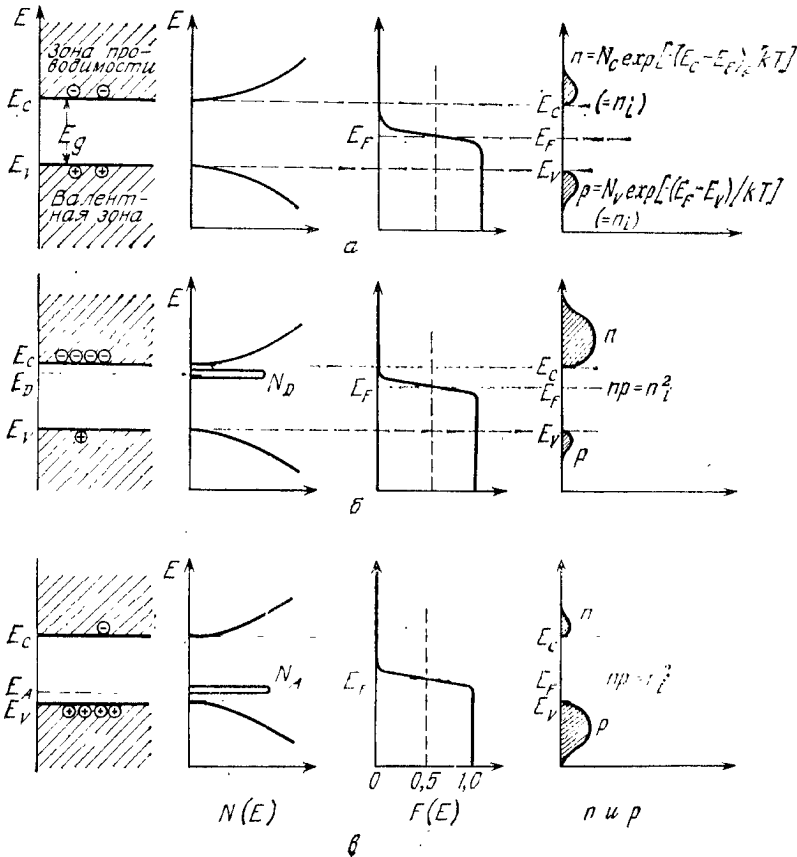


Рис. 14. Зонная диаграмма, плотность состояний, функция распределения Ферми — Дирака и концентрации носителей в собственном полупроводнике (а), в полупроводнике n -типа (б) и в полупроводнике p -типа (в).

щенная зонная схема, плотность состояний $N(E)$, функция распределения Ферми—Дирака $F(E)$ и концентрации электронов и дырок (заштрихованные области). Поскольку в собственном полупроводнике концентрации электронов и дырок равны друг другу ($n = p = n_i$), заштрихованные области на рис. 14, а в валентной зоне и зоне проводимости одинаковы.

В полупроводнике, содержащем примесные атомы, уровень Ферми смещается от середины запрещенной зоны настолько, насколько это необходимо, чтобы обеспечить выполнение условия электронейтральности образца (рис. 14, б и в). Рассмотрим сначала полупроводник, легированный донорной примесью с концентрацией N_D (см^{-3}) (рис. 14, б). Для сохранения электронейтраль-

ности концентрация всех отрицательных зарядов (электронов и ионизированных акцепторов) должна быть в точности равна концентрации положительных зарядов (дырок и ионизированных доноров). В рассматриваемом случае легирования донорной примесью ($N_A = 0$) условие электронейтральности имеет вид

$$n = N_D^+ + p. \quad (23)$$

Здесь n — концентрация электронов в зоне проводимости, p — концентрация дырок в валентной зоне, а N_D^+ — концентрация ионизированных доноров, равная [31]

$$N_D^+ = N_D \left[1 - \frac{1}{1 + \frac{1}{g} \exp\left(\frac{E_D - E_F}{kT}\right)} \right], \quad (24)$$

где g — фактор вырождения донорного примесного уровня. Он равен двум, поскольку электрон на этом уровне может иметь одно из двух значений спина. Для полупроводниковых кристаллов, легированных акцепторной примесью, условие электронейтральности записывается аналогичным образом. Концентрация ионизированных акцепторов в этом случае равна

$$N_A^- = \frac{N_A}{1 + g_A \exp\left(\frac{E_A - E_F}{kT}\right)}, \quad (25)$$

где g_A — фактор вырождения акцепторного состояния. В Ge, Si и GaAs $g_A = 4$ из-за двукратного вырождения валентной зоны при $k \rightarrow 0$.

Подставив в выражение (23) соответствующие выражения для концентраций, получим трансцендентное уравнение

$$N_C \exp\left(-\frac{E_C - E_F}{kT}\right) = N_D \frac{1}{1 + 2 \exp\left(\frac{E_F - E_D}{kT}\right)} + N_V \exp\left(\frac{E_V - E_F}{kT}\right), \quad (26)$$

позволяющее при заданных N_C , N_D , N_V , E_C , E_D , E_V и T однозначно определить положение уровня Ферми E_F . На рис. 15 показан изящный графический метод решения этого уравнения [31]. Для частного случая $N_D = 10^{18} \text{ см}^{-3}$ и $T = 300 \text{ К}$ уровень Ферми расположен у края зоны проводимости ниже энергии связи донорного состояния, так что почти все донорные центры при этом оказываются ионизированными. Для определения положения уровня Ферми при другой температуре нужно сначала вычислить соответствующие значения N_C и N_V (пропорциональные $T^{3/2}$), а затем, определив с помощью выражения (19) или по графикам рис. 11 величину $n_i(T)$, провести пересекающиеся прямые линии

$n(E_F)$ и $p(E_F)$. Пересечение первой из них с кривой $p + N_D^+$ дает положение уровня Ферми. При достаточно низких температурах уровень Ферми приближается и в конце концов пересекает донорный уровень (в полупроводнике n -типа). Для случая частичной компенсации, когда в полупроводнике кроме донорных содержится и некоторое число акцепторных примесей ($N_A < N_D$), при $N_A \gg \frac{1}{2} N_C \exp(-E_d/kT)$

$$n \approx \left(\frac{N_D - N_A}{2N_A} \right) N_C \exp(-E_d/kT), \quad (27)$$

где $E_d \equiv E_G - E_D$ — глубина донорного уровня по отношению к краю зоны проводимости, и

$$n \approx \frac{1}{\sqrt{2}} (N_D N_C)^{1/2} \exp(-E_d/2kT), \quad (28)$$

когда $N_D \gg \frac{1}{2} N_C \exp(-E_d/kT) \gg N_A$ [5]. Типичная температурная зависимость концентрации электронов в зоне проводимости приведена на рис. 16. При высоких температурах здесь расположена область собственной проводимости, где $n \approx p \gg N_D$.

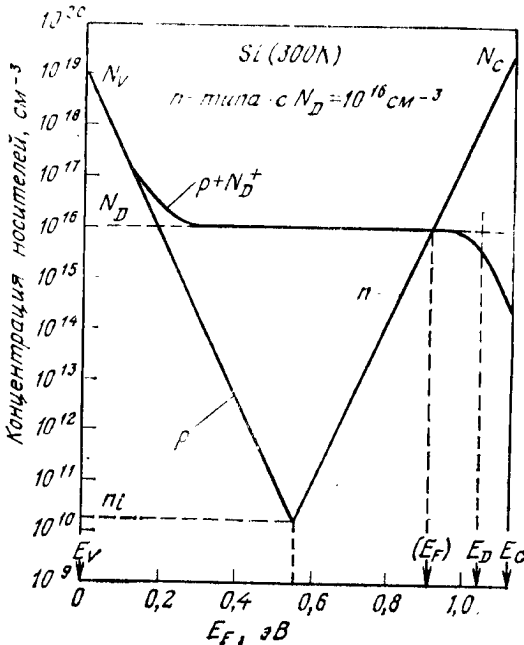


Рис. 15. Графический метод определения уровня Ферми [31].

Рис. 16. Температурная зависимость концентрации электронов в кремнии при концентрации доноров 10^{15} см^{-3} [5].

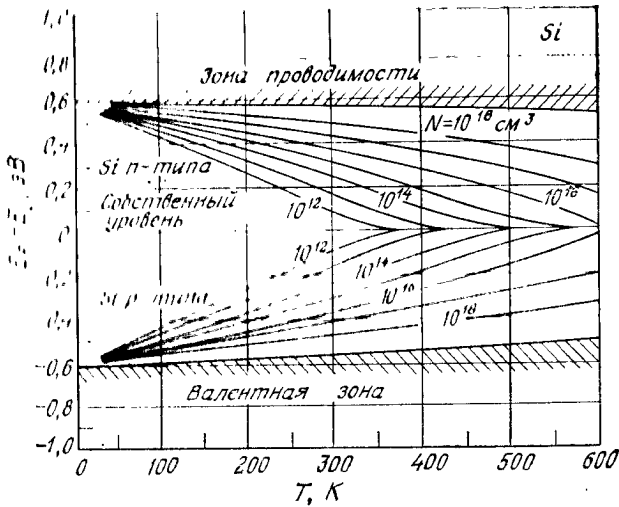
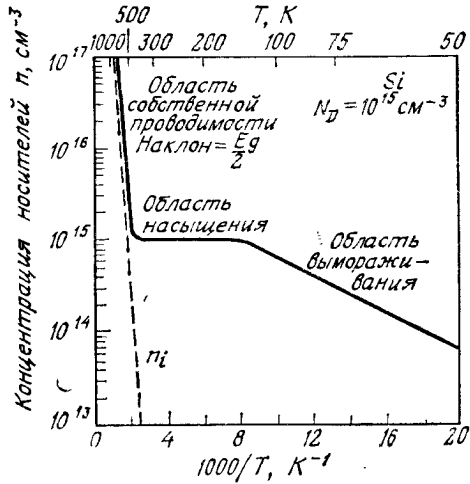


Рис. 17. Температурная зависимость положения уровня Ферми в кремневых образцах с различным уровнем легирования, рассчитанная с учетом температурной зависимости ширины запрещенной зоны [32].

При очень низких температурах происходит вымораживание электронов на примеси. В зависимости от степени компенсации наклон кривой $\ln n - \frac{1}{T}$ определяется выражением (27) или (28). В довольно широкой области промежуточных температур (на рис. 16 она простирается от ~ 100 до 500 К) концентрация

электронов в зоне проводимости слабо зависит от температуры и равна $\sim (N_D - N_A)$. На рис. 17 приведены зависимости положения уровня Ферми от температуры при различных концентрациях легирующей примеси [32], рассчитанные с учетом температурной зависимости ширины запрещенной зоны (рис. 8).

Отметим, что и в легированных полупроводниках независимо от концентрации примесей произведение np по-прежнему определяется формулой (19) (закон действующих масс). При относительно высоких температурах почти все доноры и акцепторы в полупроводнике ионизированы, и условие электронейтральности можно с достаточной точностью записать в виде

$$n + N_A = p + N_D. \quad (29)$$

Комбинируя выражения (29) и (19), для равновесных концентраций электронов и дырок в полупроводнике n -типа получим

$$n_{n0} = \frac{1}{2} [(N_D - N_A) + \sqrt{(N_D - N_A)^2 + 4n_i^2}] \approx N_D \quad (30)$$

и

$$p_{n0} = n_i^2/n_{n0} \approx n_i^2/N_D, \quad (31)$$

если

$$|N_D - N_A| \gg n_i \text{ и } N_D \gg N_A. \text{ При этом} \\ E_C - E_F = kT \ln \left(\frac{N_C}{N_D} \right). \quad (32)$$

Учитывая выражение (18), получим

$$E_F - E_i = kT \ln \left(\frac{n_{n0}}{n_i} \right). \quad (33)$$

Для концентрации дырок и электронов в полупроводнике p -типа выражения будут иметь вид

$$p_{p0} = \frac{1}{2} [(N_A - N_D) + \sqrt{(N_A - N_D)^2 + 4n_i^2}] \approx N_A \quad (34)$$

и

$$n_{p0} = n_i^2/p_{p0} \approx n_i^2/N_A \quad (35)$$

при

$$|N_A - N_D| \gg n_i \text{ и } N_A \gg N_D, \\ \text{а} \\ E_F - E_V = kT \ln \left(\frac{N_V}{N_A} \right), \quad (36)$$

или

$$E_i - E_F = kT \ln \left(\frac{p_{p0}}{n_i} \right). \quad (37)$$

В этих формулах индексы n и p обозначают тип проводимости полупроводника, а индекс 0 указывает, что речь идет о термодинамически равновесных концентрациях. В полупроводниках n -типа электроны называются основными носителями, поскольку их больше, чем дырок. Дырки называются здесь неосновными носителями. В дырочном полупроводнике основными носителями служат дырки, а неосновными — электроны.

1.5. ЯВЛЕНИЯ ПЕРЕНОСА

1.5.1. Подвижность

При малых электрических полях дрейфовая скорость носителей v_d пропорциональна напряженности электрического поля \mathcal{E} :

$$v_d = \mu \mathcal{E}. \quad (38)$$

Коэффициент пропорциональности есть подвижность μ [$\text{см}^2 \times \text{В}^{-1} \cdot \text{с}^{-1}$]. В неполярных полупроводниках, таких, как Ge и Si, основными механизмами, определяющими подвижность носителей, являются рассеяние на акустических фононах и рассеяние на ионизированных примесных атомах. Значения подвижности, определяемой рассеянием на акустических фононах, являются следующим выражением [33]:

$$\mu_i = \frac{\sqrt{8\pi} q \hbar^3 C_{11}}{m_i^* m^*{}^{5/2} (kT)^{3/2}} \sim (m^*)^{-5/2} T^{-3/2}, \quad (39)$$

где C_{11} — средний продольный модуль упругости полупроводника, V_{10} — смещение край зоны на единицу деформации кристаллической решетки, m^* — эффективная масса электрона проводимости. Согласно выражению (39), подвижность уменьшается с ростом температуры и при увеличении эффективной массы.

Подвижность, обусловленная рассеянием на ионизированных примесях, определяется выражением [34]

$$\begin{aligned} \mu_i &= \frac{64\sqrt{\pi} \epsilon_s^2 (2kT)^{3/2}}{N_i q^3 m^*{}^{1/2}} \left\{ \ln \left[1 + \left(\frac{12\pi \epsilon_s kT}{q^2 N_i^{1/3}} \right)^2 \right] \right\}^{-1} \sim \\ &\sim (m^*)^{-1/2} N_i^{-1} T^{3/2}, \end{aligned} \quad (40)$$

где N_i — концентрация ионизированных примесей, ϵ_s — диэлектрическая проницаемость. Видно, что в этом случае подвижность также уменьшается с ростом эффективной массы m^* , но с ростом температуры она увеличивается. Результирующая подвижность

$$\mu = \left(\frac{1}{\mu_i} + \frac{1}{\mu_0} \right)^{-1}. \quad (41)$$

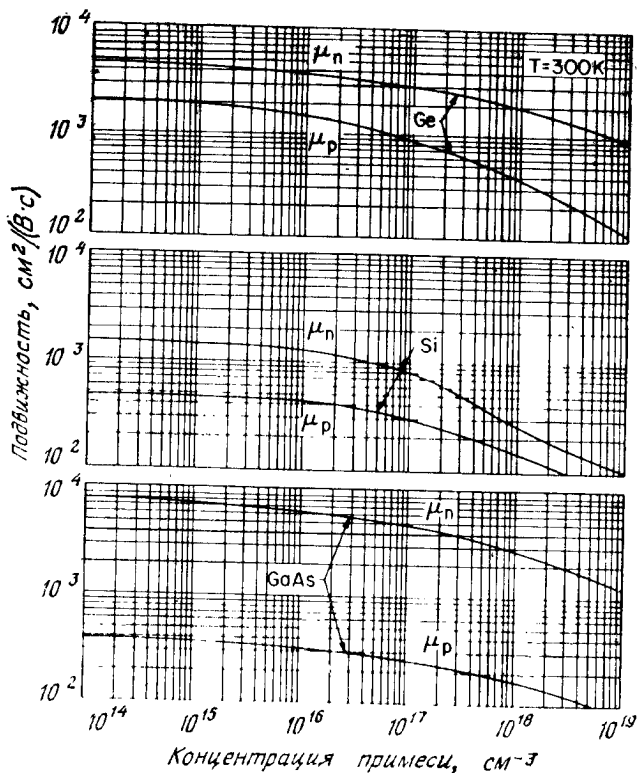


Рис. 18. Дрейфовая подвижность в Ge, Si и GaAs при $T = 300$ К в зависимости от концентрации легирующей примеси [9, 36, 38].

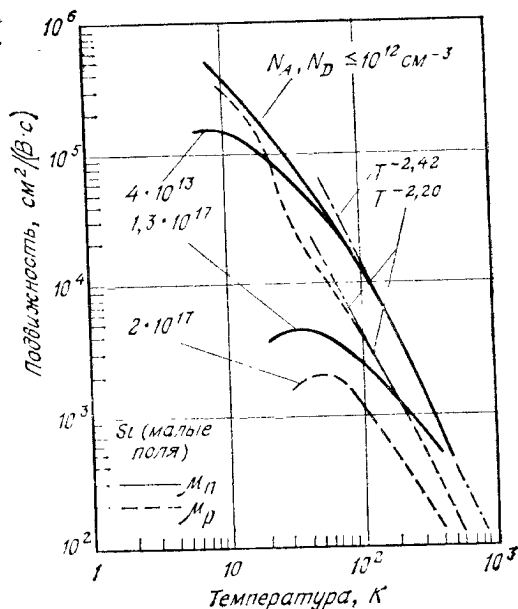
В полярных полупроводниках, таких, как GaAs, определяющую роль играет рассеяние на оптических фононах. В этом случае [35]

$$\mu \sim (m^*)^{-3/2} T^{1/2}. \quad (42)$$

Кроме названных выше механизмов, влияющих на подвижность носителей, следует отметить междолинное рассеяние, когда электрон при рассеянии переходит из одного энергетического минимума в другой (рис. 7), испуская или поглощая при этом соответствующий коротковолновый фотон.

На рис. 18 приведены экспериментальные зависимости подвижности в Ge, Si и GaAs при комнатной температуре от концентрации примеси [9, 36—38]. Видно, что с ростом концентрации примеси (при комнатной температуре большая часть примесных атомов ионизирована) подвижность уменьшается, что соответствует выражению (40). Поскольку при увеличении эффективной

Рис. 19. Температурная зависимость подвижности электронов и дырок в Si [37].



массы m^* подвижность уменьшается, в этих важнейших полупроводниковых материалах при той же концентрации примеси электронная подвижность больше дырочной (см. значения эффективных масс в приложении Д).

На рис. 19 приведены температурные зависимости подвижности в кремниевых образцах n - и p -типа с различной концентрацией примеси [37]. При малых концентрациях примесных атомов подвижность действительно уменьшается с ростом температуры, как это следует из выражения (39). Однако показатель степени экспериментальной зависимости отличается от теоретического значения $-\frac{3}{2}$, что, вероятно, обусловлено влиянием дополнительных механизмов рассеяния. В чистых материалах в области комнатных температур подвижность изменяется по закону $T^{-1,66}$ и $T^{-2,33}$ для n - и p -германия; $T^{-2,42}$ и $T^{-2,20}$ для n - и p -кремния; $T^{-1,0}$ и $T^{-2,1}$ для n - и p -GaAs соответственно.

Другим важным параметром, связанным с подвижностью, является коэффициент диффузии D_n (D_p) электронов и дырок. Его величина связана с величиной подвижности обобщенным соотношением Эйнштейна [5]

$$D_n = 2 \left(\frac{kT}{q} \mu_n \right) F_{1/2} \left(\frac{E_F - E_C}{kT} \right) / F_{-1/2} \left(\frac{E_F - E_C}{kT} \right), \quad (43)$$

где $F_{1/2}$ и $F_{-1/2}$ — интегралы Ферми—Дирака. Это выражение можно записать в виде ряда [74]

$$D_n = \frac{\mu_n kT}{q} \left[1 + 0,35355 \left(\frac{n}{N_G} \right) - 9,9 \cdot 10^{-3} \left(\frac{n}{N_G} \right)^2 + 4,45 \cdot 10^{-4} \left(\frac{n}{N_G} \right)^3 + \dots \right], \quad (43a)$$

где n — концентрация электронов, N_G — плотность состояний в зоне проводимости. В большинстве практических ситуаций достаточно учесть лишь первое и второе слагаемые этого выражения. Аналогичные соотношения связывают коэффициент диффузии и подвижность дырок. В невырожденных полупроводниках, где $n \ll N_G$ ($p \ll N_V$), выражение (43a) переходит в обычное соотношение Эйнштейна

$$D_n = \left(\frac{kT}{q} \right) \mu_n. \quad (44)$$

Коэффициенты диффузии при комнатной температуре легко получить из приведенных на рис. 18 значений подвижности. Для этого их надо умножить на $kT/q = 0,0259$ В ($T = 300$ К). Выше мы говорили исключительно о дрейфовой подвижности [27]. В следующем разделе мы рассмотрим холловскую подвижность, которая несколько отличается от дрейфовой.

1.5.2. Удельное сопротивление и эффект Холла

Удельное сопротивление ρ определяется как коэффициент пропорциональности между напряженностью электрического поля и плотностью тока:

$$\mathcal{E} = \rho J. \quad (45)$$

Обратная величина называется проводимостью $\sigma = 1/\rho$, т. е.

$$J = \sigma \mathcal{E}. \quad (46)$$

С учетом обоих типов носителей тока

$$\rho = \frac{1}{\sigma} = \frac{1}{q(\mu_n n + \mu_p p)}. \quad (47)$$

В полупроводнике n -типа, где $n \gg p$,

$$\rho \simeq \frac{1}{q\mu_n n} \quad (48)$$

или

$$\sigma \simeq q\mu_n n. \quad (48a)$$

Обычно для измерения удельного сопротивления используют четырехзондовый метод (рис. 20, вставка) [38, 39]. Через внешние

зонды пропускают малый измерительный ток от генератора тока, а внутренние зонды служат для измерения соответствующего падения напряжения. Для тонких образцов, толщина которых W много меньше их продольных размеров a и d , удельное сопротивление слоя определяется по формуле

$$R_s = \frac{V}{I} \cdot CF \quad [\text{Ом на квадрат}], \quad (49)$$

где CF — геометрический коэффициент (рис. 20). При этом $\rho = R_s W$ [Ом·см]. (49а)

Для образцов, диаметр которых d много больше расстояния между зондами s , геометрический коэффициент $CF = \pi/1\pi^2 = 4,54$.

На рис. 21 приведены экспериментальные зависимости удельного сопротивления от концентрации примеси (при $T = 300$ К) в кремниевых образцах, легированных бором (p -типа) и фосфором (n -типа) [38]. Аналогичные зависимости для Ge, GaAs и GaP [28, 38, 40] показаны на рис. 22. С помощью этих графиков можно рассчитать концентрацию легирующей примеси по измеренным значениям удельного сопротивления. Отметим, что концентрация примесей может существенно отличаться от концентрации носи-

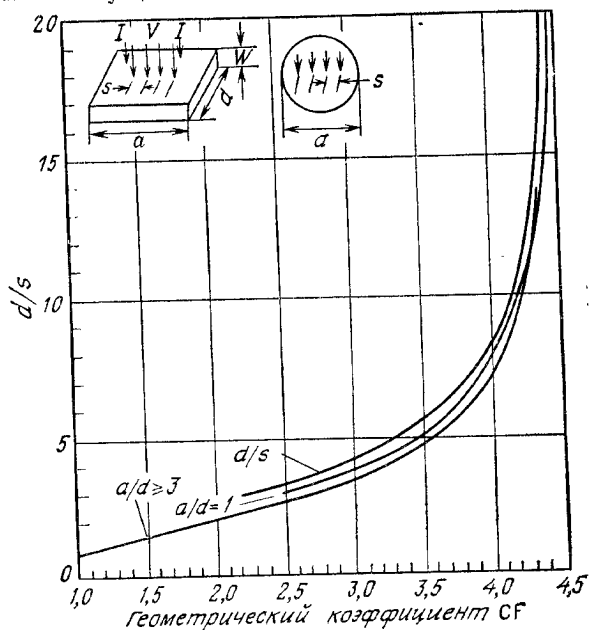


Рис. 20. Геометрический коэффициент для расчета удельного сопротивления по данным четырехзондового метода [38].

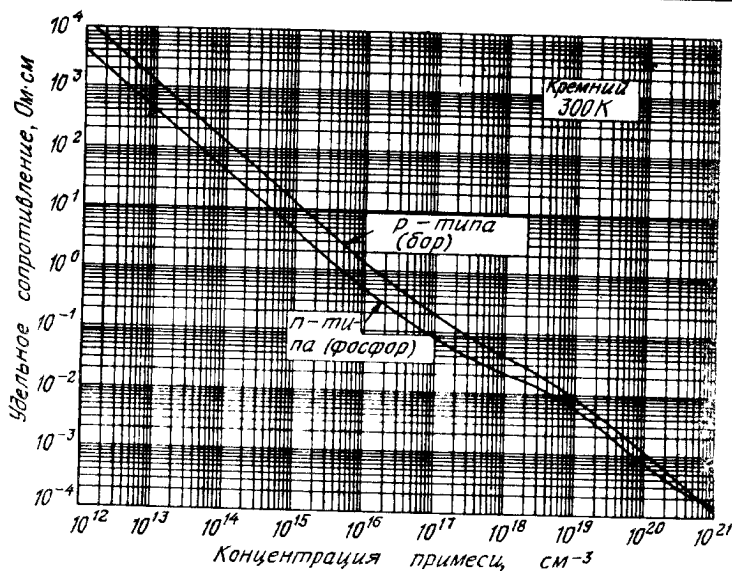


Рис. 21. Удельное сопротивление кремния при $T = 300 \text{ K}$ в зависимости от концентрации примесей [38].

телей тока. Так, например, в кремнии *p*-типа, легированного галнием с концентрацией 10^{17} см^{-3} , при комнатной температуре ионизировано всего лишь 77 % акцепторных атомов Ga, как это следует из формулы (25) и рис. 13 и 17. Другими словами, концентрация подвижных дырок при этом составляет $7,7 \cdot 10^{16} \text{ см}^{-3}$.

Для непосредственного определения концентрации носителей тока используют эффект Холла [41]. Принципиальная схема метода измерений иллюстрируется рис. 23 [42]. Вспомогательное поле \mathcal{E}_x приложено вдоль оси *x*. Перпендикулярно ему (вдоль оси *z*) направлено магнитное поле \mathcal{B}_z , а с верхнего и нижнего контактов снимается так называемое холловское напряжение V_H . Для определенности будем считать образец дырочным (*p*-типа). Сила Лоренца $qv \cdot \mathcal{B}_z$ отклоняет дырки к нижней поверхности образца, где они частично накапливаются, что приводит к возникновению вертикального электрического поля \mathcal{E}_y — холловского поля, которое компенсирует действие силы Лоренца на дырки и обеспечивает равенство нулю вертикального тока J_y . Холловское поле пропорционально плотности продольного тока J_x и напряженности магнитного поля \mathcal{B}_z . Его величину находят, измеряя холловское напряжение V_H :

$$\mathcal{E}_y = (V_H/W) = R_H J_x \mathcal{B}_z, \quad (50)$$

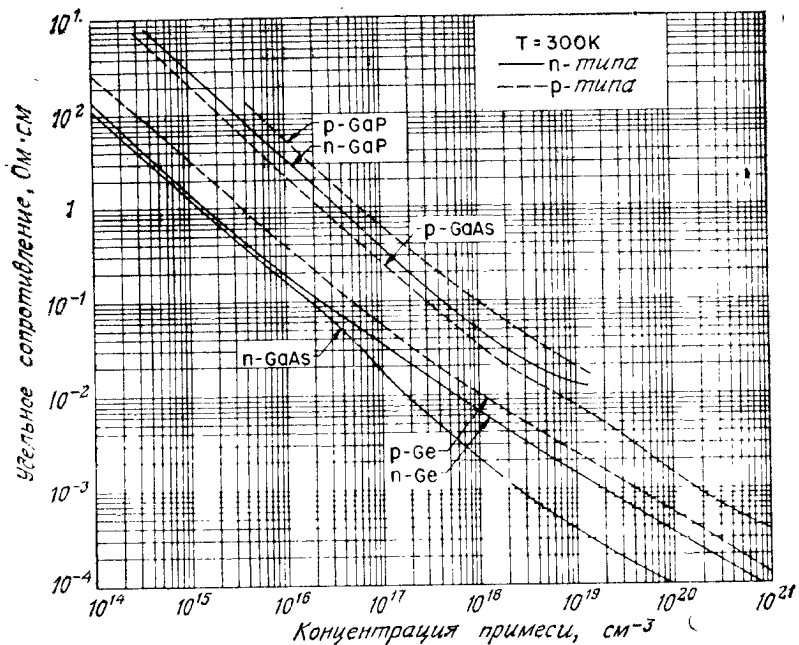


Рис. 22. Удельное сопротивление Ge, GaAs и GaP при $T = 300$ К в зависимости от концентрации примесей [28, 38].

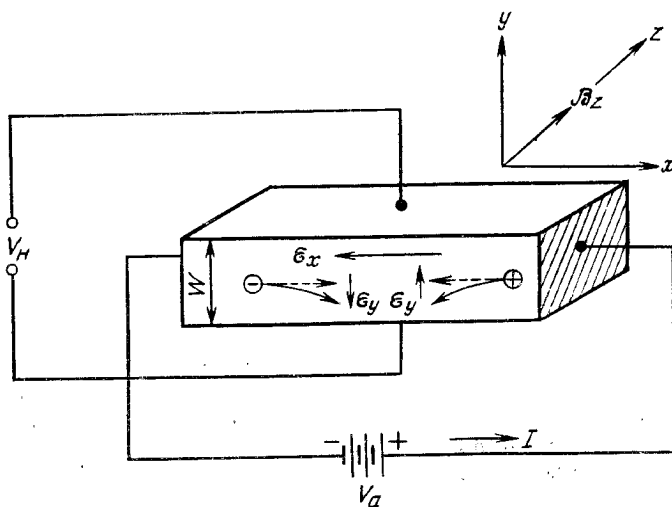


Рис. 23. Схема холловских измерений для определения концентрации.

где R_H — коэффициент Холла, определяемый выражениями [5]

$$R_H = r \frac{1}{q} \frac{p - b^2 n}{(p + bn)^2}, \quad b \equiv \mu_n / \mu_p, \quad (51)$$

$$r \equiv \langle \tau^2 \rangle / \langle \tau \rangle^2. \quad (52)$$

Параметр τ — среднее время свободного пробега носителей. Его величина зависит от энергии носителей E . В частности, в полупроводниках со сферическими изоэнергетическими поверхностями $\tau \sim E^{-1/2}$ при рассеянии на фононах и $\tau \sim E^{3/2}$ при рассеянии на ионизированных примесях. В общем случае можно считать, что $\tau = aE^{-s}$, где a и s — постоянные. При больцмановском распределении носителей по энергии (невырожденные полупроводники) среднее значение какой-либо степени τ равно

$$\langle \tau^n \rangle = \int_0^\infty \tau^n E^{3/2} \exp(-E/kT) dE \int_0^\infty E^{3/2} \exp(-E/kT) dE. \quad (53)$$

Отсюда для $\tau \sim E^{-s}$ легко получить

$$\langle \tau^2 \rangle = a^2 (kT)^{-2s} \Gamma\left(\frac{5}{2} - 2s\right) / \Gamma\left(\frac{5}{2}\right) \quad (54a)$$

и

$$\langle \tau \rangle^2 = \left[a (kT)^{-s} \Gamma\left(\frac{5}{2} - s\right) / \Gamma\left(\frac{5}{2}\right) \right]^2, \quad (54b)$$

где $\Gamma(n)$ — соответствующие гамма-функции:

$$\Gamma(n) \equiv \int_0^\infty x^{n-1} e^{-x} dx, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

Для рассмотренных выше механизмов рассеяния ($s = \frac{1}{2}$ и $s = -\frac{3}{2}$) коэффициент r оказывается равным $3\pi/8 \approx 1,18$ при рассеянии на фононах и $315\pi/512 \approx 1,93$ при рассеянии на ионизированных примесях.

Холловская подвижность μ_H определяется как произведение коэффициента Холла на проводимость:

$$\mu_H = |R_H \sigma|. \quad (55)$$

Ее следует отличать от дрейфовой подвижности μ_n (или μ_p), которая определяется с помощью выражения (48a). Для полупроводников с ярко выраженным типом проводимости ($n \gg p$ или $p \gg n$) из выражения (51) получим соответственно

$$R_H = r \left(\frac{-1}{qn} \right) \text{ и } R_H = r \left(\frac{+1}{qp} \right). \quad (56a), (56b)$$

Следовательно, в этих случаях из холловских измерений можно непосредственно определить и тип проводимости (электроны или дырки), и концентрацию носителей.

До сих пор мы фактически предполагали, что приложенное магнитное поле мало и не оказывает заметного влияния на удельное сопротивление образца. Однако в достаточно сильных магнитных полях наблюдается существенное увеличение удельного сопротивления — так называемый магниторезистивный эффект. Для полупроводников со сферическими изоэнергетическими поверхностями относительное изменение удельного сопротивления в магнитном поле равно

$$\frac{\Delta\rho}{\rho_0} = \left\{ \left[\frac{\Gamma^2\left(\frac{5}{2}\right)\Gamma\left(\frac{5}{2}-3s\right)}{\Gamma^3\left(\frac{5}{2}-s\right)} \right] \left(\frac{\mu_n^3 n + \mu_p^3 p}{\mu_n n + \mu_p p} \right) - \right. \\ \left. - \left[\frac{\Gamma\left(\frac{5}{2}\right)\Gamma\left(\frac{5}{2}-2s\right)}{\Gamma^2\left(\frac{5}{2}-s\right)} \right]^2 \left(\frac{\mu_n n - \mu_p p}{\mu_n n + \mu_p p} \right)^2 \right\} \mathcal{B}_z^2, \quad (57)$$

где ρ_0 — удельное сопротивление в отсутствие магнитного поля. Это отношение пропорционально квадрату компоненты магнитного поля, направленной перпендикулярно току. При $n \gg p$ имеем $(\Delta\rho/\rho_0) \sim \mu_n^3 \mathcal{B}_z^2$. Аналогичное соотношение получается и для полупроводника p -типа.

1.5.3. Рекомбинационные процессы

Каждый раз, когда физическая система выходит из термодинамического равновесия (т. е. когда $pn \neq n_i^2$), начинают проявляться кинетические процессы, посредством которых система стремится к восстановлению равновесия (т. е. $pn = n_i^2$). На рис. 24 схематически показаны основные генерационно-рекомбинационные процессы, ответственные за восстановление равновесных концентраций носителей тока. Рис. 24, а иллюстрирует рекомбинацию зона—зона, когда электрон непосредственно рекомбинирует с дыркой. Закон сохранения энергии при этом обеспечивается либо за счет испускания соответствующего светового кванта (излучательная рекомбинация), либо избыточная энергия отдается другому свободному электрону или дырке (оже-рекомбинация). Последний процесс является обратным по отношению к процессу ударной ионизации, а первый — к процессу оптического возбуждения электронно-дырочных пар, особенно существенному в прямозонных полупроводниках типа $A^{III}B^V$.

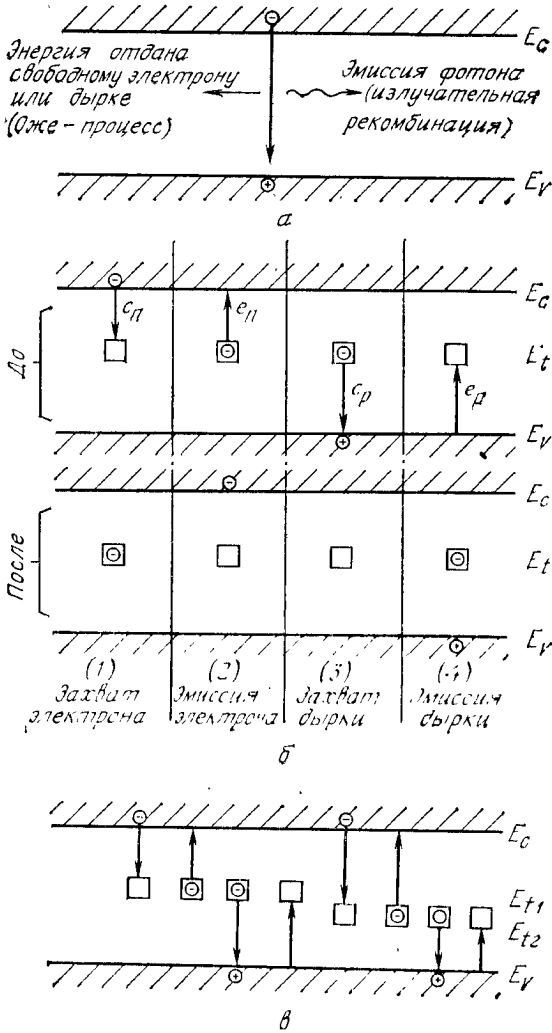


Рис. 24. Основные рекомбинационные процессы. a — рекомбинация зона — зона (излучательная или зона Оже); b — рекомбинация через моноэнергетический уровень; v — рекомбинация с участием двух уровней [43].

На рис. 24, b показана рекомбинация через дискретный энергетический уровень, расположенный в запрещенной зоне полупроводника, а на рис. 24, v — рекомбинация с участием двух таких дискретных уровней. При рекомбинации через дискретный уровень на это состояние захватывается электрон, а затем дырка. Кроме того, при расчете результирующего темпа такой рекомби-

нации U ($\text{см}^{-3} \cdot \text{с}^{-1}$) необходимо принять во внимание и обратные процессы эмиссии электрона и дырки с уровня в зоны. Согласно теории Шокли—Рида—Холла [43—45],

$$U = \frac{\sigma_p \sigma_n v_{th} (pn - n_i^2) N_t}{\sigma_n \left[n + n_i \exp\left(\frac{E_t - E_i}{kT}\right) \right] + \sigma_p \left[p + n_i \exp\left(-\frac{E_t - E_i}{kT}\right) \right]}, \quad (58)$$

где σ_p и σ_n — сечения захвата электрона и дырки; v_{th} — тепловая скорость носителей, равная $\sqrt{3kT/m^*}$; N_t — концентрация ловушек; E_t — уровень ловушек; n_i — собственная концентрация носителей; E_i — собственный уровень Ферми. Очевидно, что при тепловом равновесии $pn = n_i^2$ и $U = 0$. Полагая для простоты $\sigma_n = \sigma_p = \sigma$, вместо выражения (58) будем иметь

$$U = \sigma v_{th} N_t \frac{pn - n_i^2}{n + p + 2n_i \operatorname{ch}\left(\frac{E_t - E_i}{kT}\right)}. \quad (59)$$

Из этого выражения видно, что темп рекомбинации максимален в том случае, когда рекомбинационный уровень расположен вблизи середины запрещенной зоны ($E_t \approx E_i$). Именно такие ловушки будут наиболее эффективными рекомбинационными центрами.

При малых уровнях инжекции, т. е. когда концентрация инжектированных носителей ($\Delta n = \Delta p$) много меньше концентрации основных носителей, рекомбинационный процесс можно описывать приближенным выражением

$$U = \frac{p_n - p_{n0}}{\tau_p}, \quad (60)$$

где p_{n0} — равновесная концентрация неосновных носителей, $p_n = p_{n0} + \Delta p$, τ_p — время жизни неосновных носителей. В случае полупроводника n -типа, где $n \approx n_{n0} \gg n_i$ и p_n , разлагая выражение (58) по $\Delta p/n$, получим

$$U = \sigma_p v_{th} N_t (p_n - p_{n0}). \quad (61)$$

Следовательно, время жизни неосновных носителей (дырок) в полупроводнике n -типа

$$\theta_p = \frac{1}{\sigma_p v_{th} N_t}. \quad (62)$$

Аналогично в полупроводнике p -типа время жизни электронов

$$\tau_n = \frac{1}{\sigma_n v_{th} N_t}. \quad (63)$$

На многоуровневых ловушках рекомбинационный процесс в общем случае происходит аналогично. Конечно, имеются определенные количественные различия, особенно при высоких уровнях инжекции (когда величина $\Delta n \approx \Delta p$ — порядка концентрации основных носителей), когда результирующее время жизни асимптотически стремится к среднему по всем (положительно заряженным, отрицательно заряженным и нейтральным) уровням ловушки.

Справедливость соотношений (62) и (63) проверена экспериментально при диффузионном введении в материал дополнительных примесей с энергией уровня вблизи середины запрещенной зоны (рис. 13). Согласно изложению выше, такие примеси являются эффективными рекомбинационными центрами. Типичным примером служит золото в кремнии [29], где время жизни неосновных носителей уменьшалось обратно пропорционально концентрации введенного золота от $2 \cdot 10^{-7}$ до $2 \cdot 10^{-10}$ с при увеличении концентрации от 10^{14} до 10^{17} см⁻³. Этот эффект используется при изготовлении некоторых быстродействующих переключателей, где необходимы малые времена жизни. В качестве другого способа уменьшения времени жизни неосновных носителей используется облучение полупроводников частицами высоких энергий, которое вызывает смещение атомов из их равновесных положений в решетке и создаст другие дефекты кристаллической структуры полупроводника. При этом образуются глубокие уровни в запрещенной зоне полупроводника. Так, например, при электронном облучении кремния возникают акцепторные уровни с энергией 0,4 эВ выше потолка валентной зоны и донорные уровни, лежащие на 0,36 эВ ниже края зоны проводимости [18]. При облучении нейтронами в Si возникает акцепторный уровень 0,56 эВ, а облучение дейтерием дает промежуточное состояние с энергией 0,25 эВ над краем валентной зоны. Аналогичные результаты получаются при радиационных воздействиях на Ge, GaAs и другие полупроводники. Отметим, что в отличие от рекомбинационных центров, введенных в полупроводник диффузией, радиационные рекомбинационные центры отжигаются при сравнительно низких температурах.

Для экспериментального определения времени жизни неосновных носителей обычно используют измерения фотопроводимости (ФП) [46], или фотоэлектромагнитный эффект (ФЭМ) [47]. Плотность добавочного фототока $J_{\text{ФП}}$, возникающего при освещении полупроводника, равна

$$J_{\text{ФП}} = q(\mu_n + \mu_p)\Delta n \mathcal{E}, \quad (64)$$

где \mathcal{E} — продольное электрическое поле, приложенное к образцу; Δn — концентрация фотовозбужденных носителей, равная произведению темпа оптической генерации электронно-дырочных

пар G на время жизни τ ($\Delta n = G\tau$). Таким образом, время жизни носителей можно определить с помощью соотношения

$$\tau = \frac{\Delta n}{G} = \frac{J_{\text{ФП}}}{G\mathcal{E}q(\mu_n + \mu_p)} \sim J_{\text{ФП}}. \quad (65)$$

Экспериментальная схема, используемая при определении τ по фотопроводимости, рассмотрена в разд. 1.7. При фотоэлектромагнитном эффекте измеряют ток короткого замыкания, который появляется в освещенном образце, помещенном в магнитное поле, направленное перпендикулярно направлению падения возбуждающего света:

$$J_{\text{ФЭМ}} = q(\mu_n + \mu_p) \mathcal{E}_z \frac{D}{L} (\tau G), \quad (66)$$

где $L \equiv \sqrt{D\tau}$ — диффузионная длина. Отсюда для времени жизни получим

$$\tau = \left[\frac{J_{\text{ФЭМ}}}{\mathcal{E}_z \sqrt{DG}q(\mu_n + \mu_p)} \right]^2 \sim (J_{\text{ФЭМ}}/\mathcal{E}_z)^2. \quad (67)$$

1.6. ФОНОННЫЕ СПЕКТРЫ. ОПТИЧЕСКИЕ И ТЕПЛОВЫЕ СВОЙСТВА ПОЛУПРОВОДНИКОВ. ПОВЕДЕНИЕ ПОЛУПРОВОДНИКОВ ПРИ СИЛЬНЫХ ЭЛЕКТРИЧЕСКИХ ПОЛЯХ

В предыдущих разделах мы рассматривали процессы переноса электронов и дырок в слабых и умеренно сильных электрических полях. В этом разделе мы кратко обсудим некоторые другие эффекты и свойства полупроводников, важные для работы полупроводниковых приборов.

1.6.1. Фононные спектры

Хорошо известно, что в одномерной цепочке, состоящей из перемежающихся атомов двух разных типов (с отличающимися массами m_1 и m_2), существуют две ветви колебаний, частоты которых (при учете лишь взаимодействия между ближайшими соседями) равны [3]

$$\nu_{\pm} = \left[\alpha_f \left(\frac{1}{m_1} + \frac{1}{m_2} \right) \pm \alpha_f \sqrt{\left(\frac{1}{m_1} + \frac{1}{m_2} \right)^2 - \frac{4 \sin^2(qa)}{m_1 m_2}} \right]^{1/2}, \quad (68)$$

где α_f — силовая константа, a — расстояние между ближайшими атомами в цепочке, q — волновое число. При $q \rightarrow 0$ частота ν_{\pm} пропорциональна q . Это акустическая ветвь спектра, поскольку

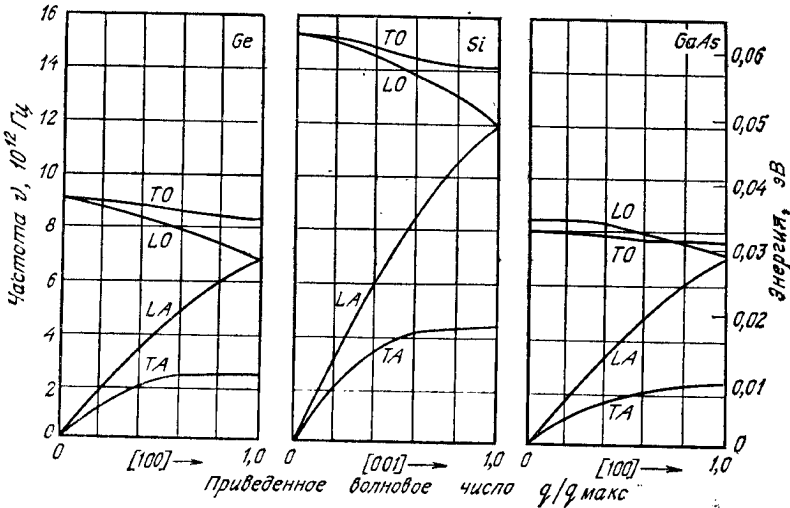


Рис. 25. Экспериментальные фоновые спектры Ge, Si и GaAs [48—50].

дисперсия ее длинноволнового участка совпадает с законом дисперсии звука, распространяющегося в среде. В свою очередь частота

$$v_+ \rightarrow \left[2\alpha_1 \left(\frac{1}{m_1} + \frac{1}{m_2} \right) \right]^{1/2}$$

при $q \rightarrow 0$. Эту ветвь колебаний, существенно отличающуюся от акустической, называют оптической, поскольку соответствующие частоты обычно лежат в оптическом диапазоне. Отметим, что в акустической моде атомы обеих подрешеток с отличающимися массами двигаются при колебаниях в одном направлении, в то время как в оптической моде подрешетки смещаются в противоположных направлениях.

В трехмерных решетках с одним атомом на элементарную ячейку, таких, как простая кубическая, объемно-центрированная и гранцентрированная кубические решетки, существуют лишь три акустические моды колебаний. В трехмерных решетках с двумя атомами на элементарную ячейку таких, как решетки Ge, Si и GaAs, имеются три акустические и три оптические моды. В продольно поляризованных модах атомы смещаются в направлении, параллельном волновому вектору, а в поперечно-поляризованных модах — перпендикулярно ему. Имеются одна продольная акустическая (LA) и одна продольная оптическая (LO) моды и соответственно две поперечные акустические (TA) и две поперечные оптические (TO) моды.

На рис. 25 приведены экспериментальные фоновые спектры Ge, Si и GaAs [48—50]. Отметим, что при малых значениях q

частоты LA- и TA-колебаний пропорциональны волновому числу q . Фононная энергия, соответствующая пику первого порядка рамановского рассеяния света, представляет собой энергию продольного оптического фонона при $q = 0$. Она равна 0,037 эВ в Ge, 0,063 эВ в Si, 0,035 эВ в GaAs. Эти значения вместе с другими важнейшими параметрами Ge, Si и GaAs приведены в приложении Е.

1.6.2. Оптические свойства полупроводников

Оптические измерения являются одним из основных методов определения зонной структуры полупроводников. Световые кванты вызывают электронные переходы между различными зонами. Это явление называется межзонным (фундаментальным) поглощением света, и оно используется для определения ширины запрещенной зоны материала. Кроме того, фотоны могут поглощаться, возбуждая электроны в состоянии с более высокой энергией в пределах той же энергетической зоны. Это явление называется поглощением на свободных носителях. Оптические измерения используются также для определения фононных спектров кристалла.

Обычно измеряют коэффициент пропускания T и коэффициент отражения света R , которые для случая нормального падения равны

$$T = \frac{(1 - R^2) \exp(-4\pi x k \cdot \lambda)}{1 - R^2 \exp(-8\pi x k \cdot \lambda)}, \quad (69)$$

$$R = \frac{(1 - \bar{n})^2 + k^2}{(1 + \bar{n})^2 + k^2}, \quad (70)$$

где λ — длина волны света, n — показатель преломления, k — коэффициент экстинкции, x — толщина образца. Коэффициент поглощения света на единице длины

$$\alpha \equiv \frac{4\pi k}{\lambda}. \quad (71)$$

Анализируя зависимости коэффициентов T и R от длины волны при нормальном падении света или любого из них при нескольких различных углах падения, можно определить и показатель преломления \bar{n} , и коэффициент поглощения α , что позволяет затем рассчитать характерные энергии электронных оптических переходов.

Вблизи края фундаментального поглощения [5]

$$\alpha \sim (h\nu - E_g)^\gamma, \quad (72)$$

где $h\nu$ — энергия фотона, E_g — ширина запрещенной зоны, γ — показатель степени. Теоретически (в одноэлектронном приближении) $\gamma = \frac{1}{2}$ для разрешенных прямых переходов и $\gamma = \frac{3}{2}$

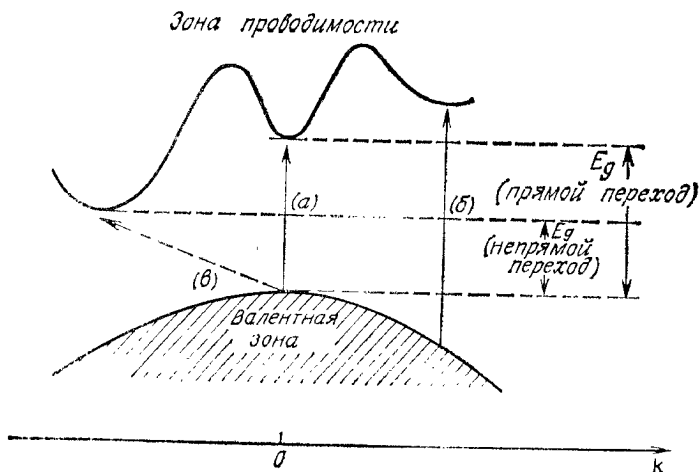


Рис. 26. Оптические переходы.

для запрещенных прямых переходов (случай (а) и (б) на рис. 26, когда экстремумы соответствующих зон расположены при одном и том же значении волнового вектора $k_{\text{мин}} = k_{\text{мин}}$). Для непрямых переходов (случай (в) на рис. 26), которые происходят с участием фононов, $\gamma = 2$. Кроме того, в случае непрямых переходов в экситонное состояние $\gamma = \frac{1}{2}$. Экситон — это связанное состояние электрона и дырки с энергией, чуть меньшей ширины запрещенной зоны. Такая связанная электронно-дырочная пара может перемещаться по кристаллу как одно целое.

Вблизи края поглощения, когда значения разности $h\nu - E_g$ сравнимы с энергией связи экситона, при расчете α нужно учитывать кулоновское взаимодействие между генерированными электроном и дыркой. В этой области частот ($h\nu \approx E_g$) край поглощения с образованием свободных электрона и дырки непрерывно сливается с размытыми линиями поглощения, которые соответствуют рождению экситонов в высоковозбужденных состояниях. При высоких частотах ($h\nu \gg E_g$) в процессах оптического поглощения участвуют более высокие энергетические зоны полупроводника с более сложной зонной структурой, что отражается на форме зависимости $\alpha(h\nu)$.

На рис. 27 приведены экспериментальные зависимости коэффициента поглощения α от энергии фотонов в Ge, Si и GaAs вблизи и выше края фундаментального поглощения [51—54]. Наблюдаемое смещение кривых поглощения в сторону высоких энергий при понижении температуры обычно связывают с температурной зависимостью ширины запрещенной зоны (рис. 8).

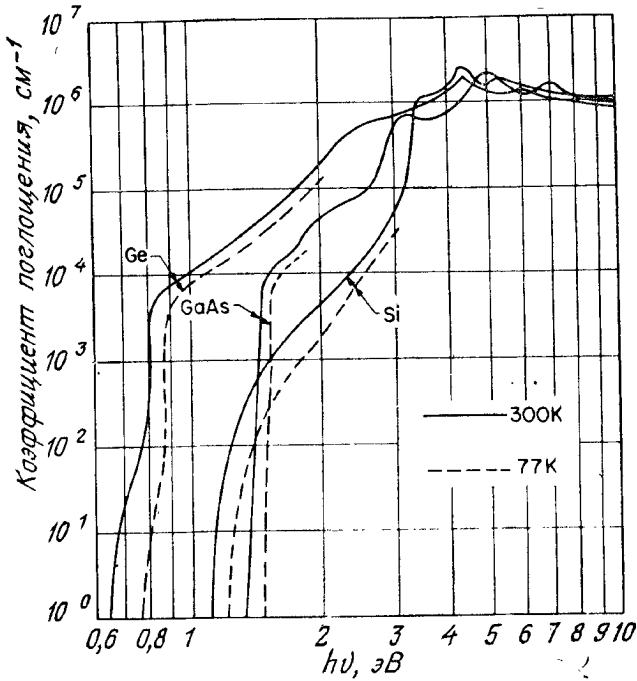


Рис. 27. Экспериментальные спектры оптического поглощения в чистых кристаллах Ge, Si и GaAs [51—54].

1.6.3. Тепловые свойства полупроводников

Если в полупроводнике наряду с электрическим полем имеется градиент температуры, то полная плотность тока (в одномерном случае) определяется выражением [5]

$$J = \sigma \left(\frac{1}{q} \frac{\partial E_F}{\partial x} - \mathcal{P} \frac{\partial T}{\partial x} \right), \quad (73)$$

где σ — проводимость, E_F — энергия Ферми (электрохимический потенциал), \mathcal{P} — дифференциальная термо-э. д. с. Для невырожденных полупроводников при степенной зависимости среднего времени свободного пробега от энергии $\tau \sim E^{-s}$ термо-э. д. с. равна

$$\mathcal{P} = - \frac{k}{q} \left\{ \frac{\left[\frac{5}{2} - s + \ln(N_C/n) \right] n \mu_n - \left[\frac{5}{2} - s - \ln(N_V/p) \right] p \mu_p}{n \mu_n + p \mu_p} \right\}, \quad (74)$$

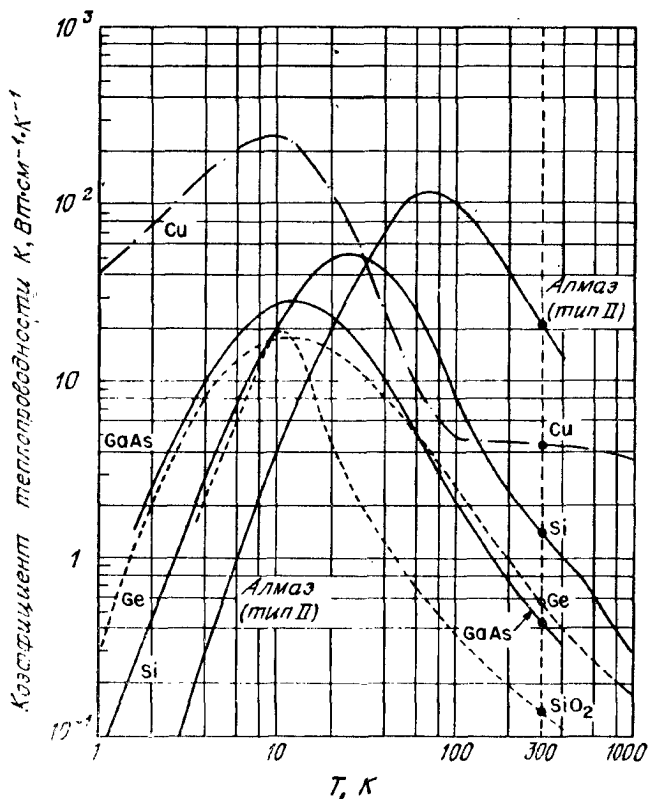


Рис. 28. Экспериментальные температурные зависимости коэффициента теплопроводности в чистых Ge, Si, GaAs, Cu, SiO₂ и алмазе [тип II] [55–57].

где k — постоянная Больцмана, N_C и N_V — эффективные плотности состояний в зоне проводимости и в валентной зоне соответственно. Отсюда следует, что в полупроводниках n -типа термо-э. д. с. отрицательна, а в полупроводниках p -типа термо-э. д. с. положительна. Этот результат часто используют для определения типа проводимости полупроводниковых образцов. Измерения термо-э. д. с. также можно использовать для определения положения уровня Ферми относительно краев разрешенных зон. При комнатных температурах в кремнии p -типа термо-э. д. с. увеличивается с ростом удельного сопротивления (от 1 мВ·К⁻¹ в образце с $\rho = 0.1$ Ом·см до 1,7 мВ·К⁻¹ в образце с $\rho = 100$ Ом·см). Аналогичные результаты (с точностью до знака \mathcal{P}) получаются в кремниевых образцах с проводимостью n -типа.

Другим важным параметром, характеризующим тепловые свойства полупроводников, является коэффициент теплопроводности κ

$$\kappa = \kappa_L + \frac{\left(\frac{5}{2} - s\right) k^2 \sigma T}{q^2} + \frac{k^2 \sigma T}{q^2} \frac{(\sigma - 2s + E_g/kT)^2 n p \mu_n \mu_p}{(n \mu_n + p \mu_p)^2} \quad (75)$$

в случае, когда для электронов и дырок $\tau \sim E^{-s}$. В этом выражении первое слагаемое соответствует решеточной теплопроводности, второе — электронной теплопроводности, а третье — комбинированным электронно-решеточным процессам. Вклад второго слагаемого в полную теплопроводность обычно мал. Третий член может быть достаточно велик, когда $E_g \gg kT$. Установлено, что коэффициент теплопроводности сначала увеличивается с ростом T (при низких температурах), а затем при высоких температурах уменьшается.

На рис. 28 приведены экспериментальные температурные зависимости коэффициента теплопроводности в Ge, Si и GaAs [55, 56], а в приложении Е даны значения этих коэффициентов при комнатной температуре. На рис. 28 приведены также соответствующие зависимости для меди, алмаза (тип II) [55] и SiO₂ [57]. Медь — наиболее часто используемый материал для теплоотводов в приборах с p — n -переходами. Алмаз (тип II) имеет наибольшую среди всех известных материалов теплопроводность при комнатной температуре. Его применение в качестве теплоотвода в лазерах с p — n -переходом и в ЛПД-генераторах рассмотрено в соответствующих разделах книги.

1.6.4. Поведение носителей заряда при сильных электрических полях

Как уже говорилось в разд. 1.5.1, при малых электрических полях дрейфовая скорость носителей в полупроводниках пропорциональна напряженности электрического поля, а коэффициент этой пропорциональности, который называется подвижностью, не зависит от электрического поля. Однако при достаточно сильных электрических полях полевая зависимость дрейфовой скорости носителей становится нелинейной, и в некоторых случаях происходит насыщение дрейфовой скорости. При более сильных полях начинается ударная ионизация. Сначала мы рассмотрим нелинейность дрейфовой скорости (подвижности).

При термодинамическом равновесии носители заряда в процессах рассеяния испускают столько же фононов, сколько и поглощают. Распределение носителей по энергии при этом описывается максвелловской функцией с температурой, равной температуре кристаллической решетки. Когда к полупроводнику при-

ложено внешнее электрическое поле, носители приобретают в нем дополнительную энергию и отдают ее в решетку в результате превышения эмиссии фононов над их поглощением при рассеянии. При этом средняя энергия носителей оказывается больше, чем в равновесном случае, а их распределение по энергии характеризуется эффективной электронной температурой T_e , которая больше температуры решетки T . Уравнение, определяющее электронную температуру T_e , получают из условия равенства скорости поступления энергии в электронную подсистему от электрического поля и скорости, с которой эта энергия переходит из электронной подсистемы в решетку за счет испускания фононов [3]:

$$\frac{T_e}{T} = \frac{1}{2} \left\{ 1 + \left[1 + \frac{3\pi}{8} \left(\frac{\mu_0 \mathcal{E}}{C_s} \right)^2 \right]^{1/2} \right\}. \quad (76)$$

Здесь μ_0 — подвижность при малых полях, C_s — скорость звука, \mathcal{E} — электрическое поле. При этом

$$v_d = \mu_0 \mathcal{E} \sqrt{\frac{T}{T_e}}. \quad (77)$$

При малых электрических полях, когда $\mu_0 \mathcal{E} < C_s$, соответствующим образом разлагая выражения (76) и (77), получим

$$T_e \simeq T \left[1 + \frac{3\pi}{32} \left(\frac{\mu_0 \mathcal{E}}{C_s} \right)^2 \right] \quad (78)$$

и

$$v_d \simeq \mu_0 \mathcal{E} \left[1 - \frac{3\pi}{64} \left(\frac{\mu_0 \mathcal{E}}{C_s} \right)^2 \right]. \quad (79)$$

При напряженности электрического поля $\mu_0 \mathcal{E} \simeq 8C_s/3$ электронная температура T_e становится в 2 раза больше решеточной, а подвижность носителей уменьшается на 30 %. При еще больших полях дрейфовая скорость в Si и Ge перестает зависеть от поля и становится равной скорости насыщения

$$v_s = \sqrt{\frac{8E_p}{3\pi m_0}} \simeq 10^7 \text{ см/с}, \quad (80)$$

где E_p — энергия оптического фонона (приложение Е).

В GaAs полевая зависимость дрейфовой скорости более сложная, чем в Ge и Si, что обусловлено особенностями энергетического спектра зоны проводимости этого материала (рис. 5). Основным минимумом зоны проводимости (долина) здесь расположен в центре зоны Бриллюэна и характеризуется высокой подвижностью ($\mu \approx 4000\text{--}8000 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$), а на осях (111) [58] расположены долины с малой ($\sim 100 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$) подвижностью и энергией на 0,3 эВ выше основного минимума. В нижней долине эффективная масса $m^* = 0,068 m_0$, а в верхних долинах $m^* = 1,2 m_0$.

Следовательно, плотность состояний в верхней долине примерно в 70 раз больше, чем в нижней. В достаточно сильных электрических полях электроны в основной долине разогреваются настолько, что становятся возможными их переходы в верхние долины с малой подвижностью. Это приводит к возникновению отрицательного дифференциального сопротивления. Более подробно механизм междолинных переходов и полевая зависимость дрейфовой скорости в GaAs рассмотрены в гл. 11.

На рис. 29, а приведены зависимости дрейфовой скорости от электрического поля, измеренные при комнатной температуре в довольно чистых кристаллах Ge, Si и GaAs [37, 59, 60]. Сильнолегированные кристаллы характеризуются меньшей подвижностью носителей при малых полях. Однако, как оказалось [75], скорость насыщения слабо зависит от концентрации примесей и примерно одинакова в чистых и сильнолегированных образцах. В Ge скорость насыщения для электронов и дырок примерно одинакова и равна $6 \cdot 10^6$ см/с, а в Si $v_s = 1 \cdot 10^7$ см/с. В GaAs имеется довольно протяженный участок с отрицательной дифференциальной подвижностью (при полях, превышающих $3 \cdot 10^3$ В/см). При значительно больших полях дрейфовая скорость электронов в GaAs стремится к значению $6 \cdot 10^6$ В/см. На рис. 29, б приведена температурная зависимость скорости насыщения электронов в Si и GaAs [37, 61, 62]. Как можно видеть, с ростом температуры скорость насыщения в Si и GaAs уменьшается.

Рассмотрим теперь ударную ионизацию. В достаточно сильном электрическом поле носители тока «разгоняются» до таких больших энергий, что могут возбуждать новые электронно-дырочные пары за счет ударной ионизации. Темп генерации электронно-дырочных пар в единице объема при ударной ионизации можно записать в виде

$$G = \alpha_n n v_n + \alpha_p p v_p, \quad (81)$$

где α_n и α_p — коэффициенты ударной ионизации, определяемые как число электронно-дырочных пар, генерируемых носителем тока на единице длины траектории. Коэффициенты α_n и α_p являются весьма резкими функциями электрического поля. Для этих величин в работе [76] получено следующее соотношение:

$$\alpha(\mathcal{E}) = (q\mathcal{E}/E_i) \exp \{ -\mathcal{E}_i/\mathcal{E} (1 + \mathcal{E}/\mathcal{E}_p) + \mathcal{E}_{KT} \}, \quad (82)$$

где E_i — пороговая энергия ударной ионизации, \mathcal{E}_{KT} , \mathcal{E}_p и \mathcal{E}_i — пороговые электрические поля, при превышении которых начинают проявляться эффекты замедления носителей за счет испускания тепловых и оптических фононов, а также ионизационных потерь соответственно. Как было установлено, в Si для электронов $E_i = 3,6$ эВ и для дырок $E_i = 5,0$ эВ. В ограниченном

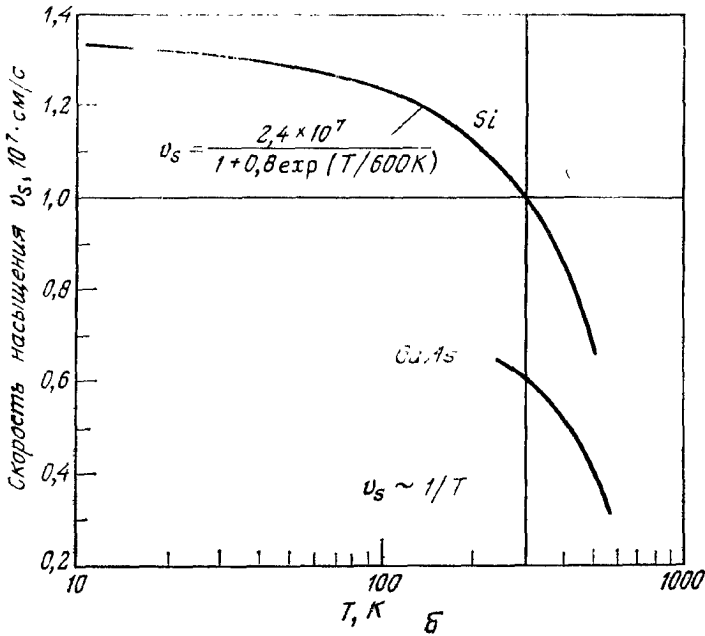
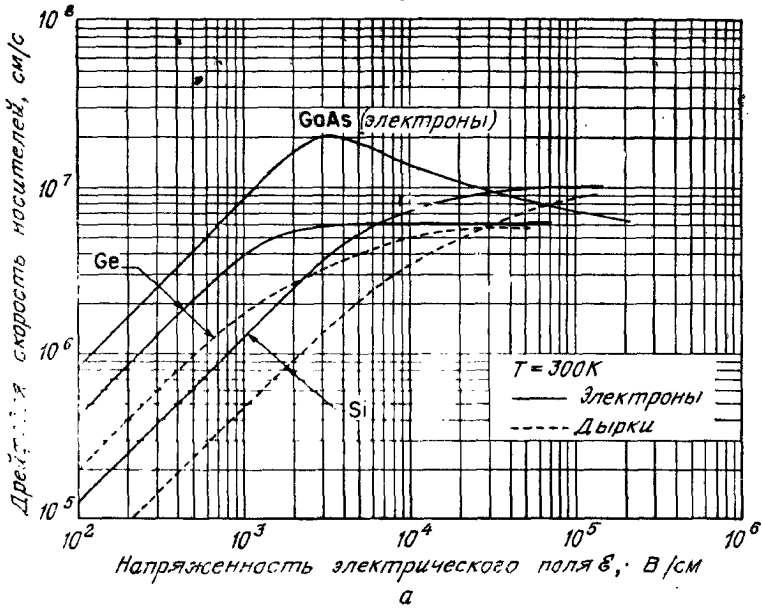


Рис. 29. Экспериментальные зависимости дрейфовой скорости (а) от электрического поля в чистых кристаллах Ge, Si и GaAs [37, 59, 60] и температурная зависимость скорости насыщения электронов (б) в Si и GaAs [37, 61, 62].

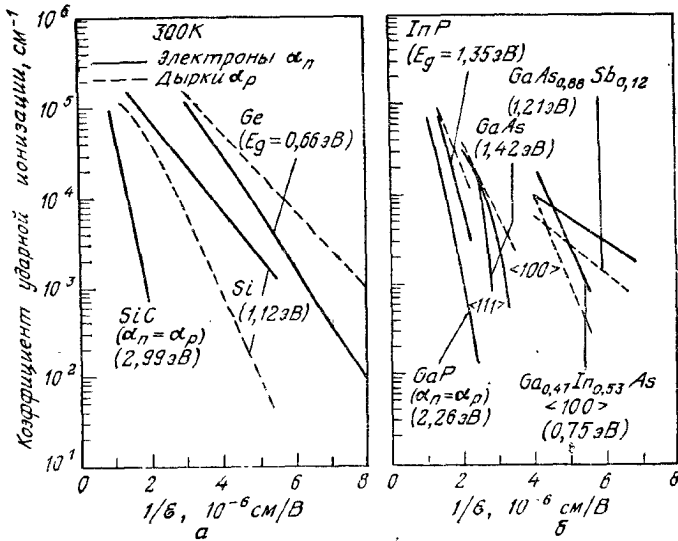


Рис. 30. Полевая зависимость коэффициентов ударной ионизации при $T = 300 \text{ K}$ в Ge, Si, GaAs и некоторых других полупроводниковых соединениях [63—69].

диапазоне электрических полей вместо выражения (82) можно записать следующие упрощенные соотношения:

$$\alpha(\mathcal{E}) = (q\mathcal{E}/E_I) \exp(-\mathcal{E}_I/\mathcal{E}) \text{ для } \mathcal{E}_p > \mathcal{E} > \mathcal{E}_{нт}, \quad (83a)$$

$$\alpha(\mathcal{E}) = (q\mathcal{E}/E_I) \exp(-\mathcal{E}_I \mathcal{E}_p / \mathcal{E}^2) \text{ для } \mathcal{E} > \mathcal{E}_p, \sqrt{\mathcal{E}_p \mathcal{E}_{нт}}. \quad (83б)$$

На рис. 30, а показаны результаты экспериментального определения коэффициентов ударной ионизации [63—65] в Ge, Si, SiC, а на рис. 30, б — коэффициенты ионизации в GaAs и некоторых других бинарных и тройных соединениях типа $A^{III}B^V$ [66—70]. Эти результаты получены при измерении фотоумножения в p — n -переходах. Отметим, что в некоторых полупроводниках, например в GaAs, коэффициенты ионизации зависят от ориентации кристалла. Как правило, характерное электрическое поле, соответствующее определенному (например, 10^4 см^{-1}) значению коэффициента ионизации, уменьшается с уменьшением ширины запрещенной зоны. Отметим также, что приведенные на рис. 30 экспериментальные зависимости для большинства полупроводников хорошо аппроксимируются выражением (83а), за исключением GaAs и GaP, для которых лучшее соответствие обеспечивает формула (83б).

Температурную зависимость коэффициентов ударной ионизации можно выразить в терминах модифицированной трехпараме-

трической теории Бараффа [71, 72]. Этими параметрами являются [69] пороговая энергия ударной ионизации E_I , средняя длина свободного пробега носителя до испускания оптического фонона λ и средняя потеря энергии при фононном рассеянии $\langle E_p \rangle$. Они определяются формулами [69]

$$\lambda = \lambda_0 \operatorname{th} \left(\frac{E_p}{2kT} \right), \quad (84)$$

$$\langle E_p \rangle = E_p \operatorname{th} \left(\frac{E_p}{2kT} \right), \quad (85)$$

$$\frac{\lambda}{\lambda_0} = \frac{\langle E_p \rangle}{E_p}, \quad (86)$$

где E_p — энергия оптического фонона (приложение E), λ_0 — асимптотическое низкотемпературное значение средней длины свободного пробега носителя с высокой энергией до испускания оптического фонона.

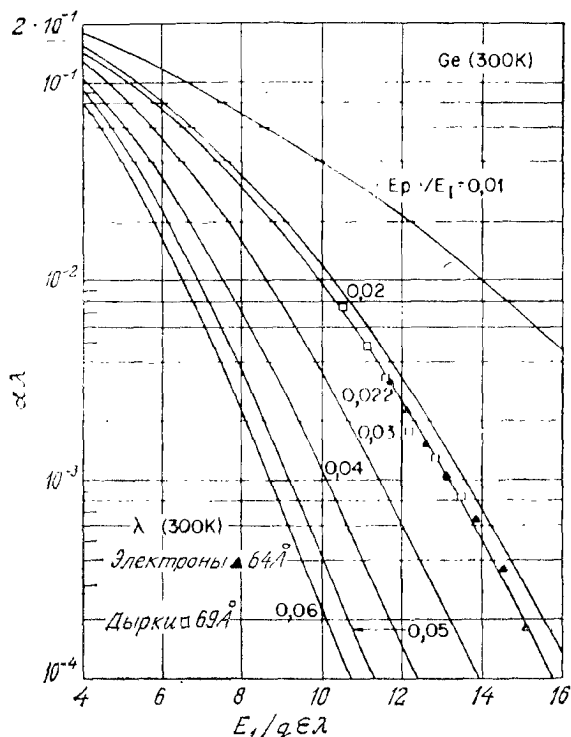


Рис. 31. Графики трехпараметрической теории Бараффа для расчета зависимости α (\mathcal{E}) при различных температурах [71].

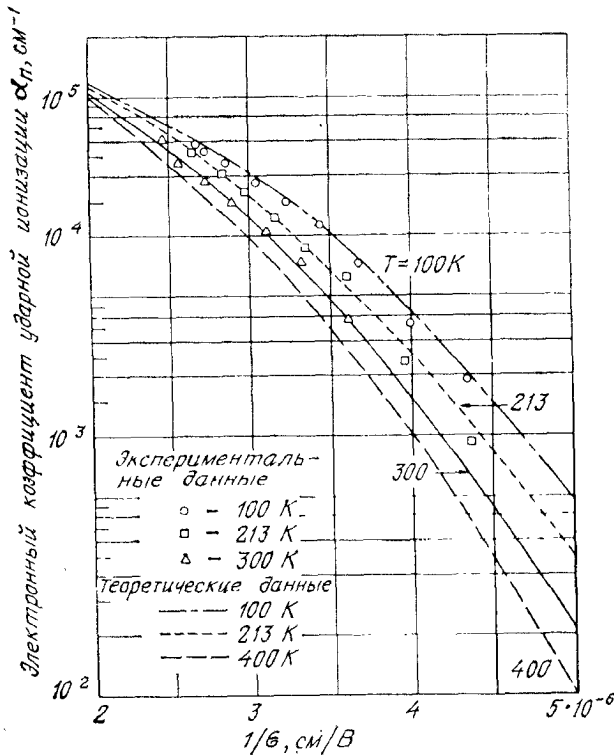


Рис. 32. Зависимость коэффициента ударной ионизации от обратного электрического поля в кремнии при четырех различных температурах [72].

Результаты теории Бараффа показаны на графиках рис. 31, где произведение $\alpha\lambda$ представлено в зависимости от величины $E_1/q\mathcal{E}\lambda$. Параметром кривых является величина $\langle E_p \rangle / E_1$ — отношение средней энергии оптического фонона к пороговой энергии ионизации. Поскольку в конкретной серии ионизационных измерений значения E_1 ($\approx 1,0$ эВ для Ge), коэффициента α и его зависимости от поля фиксированы, сравнивая экспериментальные результаты с графиками Бараффа, можно определить λ — характерную среднюю длину пробега носителя до испускания оптического фонона. Типичный пример такого сравнения для Ge при 300 К показан на рис. 31. Так как параметр $\langle E_p \rangle / E_1$ в данном случае равен 0,022, получается, что при комнатной температуре $\lambda = 64 \text{ \AA}$ для электронов и $\lambda = 69 \text{ \AA}$ для дырок. Подобные результаты были получены также в кристаллах Si, GaAs и GaP. Определить таким способом величину λ при комнатной температуре, по формуле (84) находим значение λ_0 (для основных полупровод-

никовых материалов они приведены в приложении Е). Следовательно, теперь можно рассчитать величину λ при любой интересующей нас температуре T_1 . Учитывая температурную зависимость $\langle E_p \rangle$ (выражение (85)), можно найти соответствующий T_1 график Барарфа, который даст в конечном итоге искомую зависимость $\alpha(\mathcal{E})$ при $T = T_1 \neq 300$ К. На рис. 32 приведены полученные описанным выше способом теоретические зависимости коэффициента ударной ионизации от электрического поля в кремнии при нескольких различных температурах и экспериментальные результаты. Отметим в заключение, что при одинаковых электрических полях коэффициент ударной ионизации тем меньше, чем выше температура.

1.7. основные уравнения для анализа работы полупроводниковых приборов

1.7.1. Основные уравнения [31]

Основные уравнения для анализа работы полупроводниковых приборов определяют статику и динамику поведения носителей тока в полупроводниках, которые находятся под действием внешних полей, обуславливающих соответствующие отклонения от равновесных условий. Их можно разделить на три группы: уравнения Максвелла, уравнения для плотности токов и уравнения непрерывности.

Уравнения Максвелла для однородных и изотропных материалов:

$$\nabla \times \mathcal{E} = -\frac{\partial \mathcal{H}}{\partial t}, \quad (87)$$

$$\nabla \times \mathcal{H} = \frac{\partial \mathcal{D}}{\partial t} + \mathbf{J}_{\text{cond}} = \mathbf{J}_{\text{tot}} \quad (88)$$

$$\nabla \cdot \mathcal{D} = \rho(x, y, z), \quad (89)$$

$$\nabla \cdot \mathcal{B} = 0, \quad (90)$$

$$\mathcal{B} = \mu_0 \mathcal{H}, \quad (91)$$

$$\mathcal{D}(r, t) = \int_{-\infty}^t \epsilon_s(t-t') \mathcal{E}(r, t') dt'. \quad (92)$$

Здесь \mathcal{E} и \mathcal{D} — электрическое поле и вектор электрической индукции, \mathcal{H} и \mathcal{B} — магнитные поле и индукция, ϵ_s и μ_0 — диэлектрическая и магнитная проницаемости, $\rho(x, y, z)$ — объемная плотность полного электрического заряда, \mathbf{J}_{cond} — плотность тока проводимости, \mathbf{J}_{tot} — плотность полного тока (состоящего из тока проводимости и тока смещения) ($\nabla \cdot \mathbf{J}_{\text{tot}} = 0$). Среди этих

шести уравнений наиболее важным в практическом отношении является уравнение Пуассона (уравнение (89)), определяющее свойства обедненных слоев в p - n -переходах (гл. 2).

Уравнения для плотности токов:

$$\mathbf{J}_n = q\mu_n n \mathcal{E} + qD_n \nabla n, \quad (93)$$

$$\mathbf{J}_p = q\mu_p p \mathcal{E} - qD_p \nabla p, \quad (94)$$

$$\mathbf{J}_{\text{cond}} = \mathbf{J}_n + \mathbf{J}_p. \quad (95)$$

Здесь \mathbf{J}_n и \mathbf{J}_p — плотности электронного и дырочного тока, состоящие из полевой и диффузионной (обусловленной градиентом концентрации) компонент. Вопрос о подвижностях электронов μ_n и дырок μ_p мы уже обсуждали в разд. 1.5.1. В невырожденных полупроводниках коэффициенты диффузии D_n и D_p связаны со значениями подвижностей μ_n и μ_p соотношением Эйнштейна $D_n = (kT/q) \mu_n$.

В одномерном случае уравнения (93) и (94) имеют вид

$$J_n = q\mu_n n \mathcal{E} + qD_n \frac{dn}{dx} = q\mu_n \left(n \mathcal{E} + \frac{kT}{q} \frac{dn}{dx} \right), \quad (93a)$$

$$J_p = q\mu_p p \mathcal{E} - qD_p \frac{dp}{dx} = q\mu_p \left(p \mathcal{E} - \frac{kT}{q} \frac{dp}{dx} \right), \quad (94a)$$

которые справедливы при не очень сильных электрических полях. При сильных полях в этих уравнениях следует заменить произведения $\mu_p \mathcal{E}$ и $\mu_n \mathcal{E}$ на соответствующие скорости насыщения v_s . Для полупроводников, находящихся во внешних магнитных полях, к правой части уравнений (93) и (94) нужно добавить слагаемые $J_{n\perp} \operatorname{tg} \theta_n$ и $J_{p\perp} \operatorname{tg} \theta_p$ соответственно, где $J_{n\perp}$ и $J_{p\perp}$ — компоненты плотности электронного и дырочного тока, перпендикулярные магнитному полю, а $\operatorname{tg} \theta_n = q\mu_n n R_H / \mathcal{H}$ ($\operatorname{tg} \theta_p = q\mu_p p R_H / \mathcal{H}$). Напомним, что электронный холловский коэффициент R_H отрицателен, а дырочный R_H положителен.

Уравнения непрерывности:

$$\frac{\partial n}{\partial t} = G_n - U_n + \frac{1}{q} \nabla \cdot \mathbf{J}_n, \quad (96)$$

$$\frac{\partial p}{\partial t} = G_p - U_p - \frac{1}{q} \nabla \cdot \mathbf{J}_p, \quad (97)$$

где G_n и G_p — темп генерации электронов и дырок в единице объема ($\text{см}^{-3}/\text{с}$), вызываемой внешними воздействиями, такими, как оптическое возбуждение или ударная ионизация при сильных полях. Скорость рекомбинации электронов в полупроводнике p -типа здесь обозначена символом U_n . При малых уровнях инжекции, когда концентрация инжектированных носителей много меньше равновесной концентрации основных носителей тока,

$U_n \approx (n_p - n_{p0})/\tau_n$, где n_p — концентрация неосновных носителей тока, n_{p0} — термодинамически равновесное значение этой концентрации, τ — время жизни неосновных носителей (электронов). Аналогичным образом в полупроводнике n -типа скорость рекомбинации дырок определяется через дырочное время жизни τ_p . Если электроны и дырки генерируются и рекомбинируют парами без прилипания или других аналогичных эффектов, то $\tau_n = \tau_p$. В одномерном случае в условиях малой инжекции уравнения (96) и (97) имеют вид

$$\frac{\partial n_p}{\partial t} = G_n - \frac{n_p - n_{p0}}{\tau_n} + n_p \mu_n \frac{\partial \mathcal{E}}{\partial x} + \mu_n \mathcal{E} \frac{\partial n_p}{\partial x} - D_n \frac{\partial^2 n_p}{\partial x^2}, \quad (96a)$$

$$\frac{\partial p_n}{\partial t} = G_p - \frac{p_n - p_{n0}}{\tau_p} + p_n \mu_p \frac{\partial \mathcal{E}}{\partial x} + \mu_p \mathcal{E} \frac{\partial p_n}{\partial x} - D_p \frac{\partial^2 p_n}{\partial x^2}. \quad (97a)$$

1.7.2. Примеры применения основных уравнений

Релаксация фотовозбужденных носителей. Рассмотрим образец n -типа, освещаемый так, что свет генерирует электронно-дырочные пары равномерно по его объему (рис. 33, а). При $\mathcal{E} = 0$ и $\partial p_n / \partial x = 0$ уравнение (97a) имеет вид

$$\frac{\partial p_n}{\partial t} = G - \frac{p_n - p_{n0}}{\tau_p}. \quad (98)$$

В стационарных условиях $\partial p_n / \partial t = 0$ и

$$p_n = p_{n0} + \tau_p G = \text{const}. \quad (99)$$

Пусть в некоторый момент времени, например при $t = 0$, освещение выключается. В последующие моменты времени ($t > 0$) концентрация определяется уравнением

$$\frac{\partial p_n}{\partial t} = - \frac{p_n - p_{n0}}{\tau_p}, \quad (100)$$

которое нужно решить с начальным условием $p_n(0) = p_{n0} + \tau_p G$ (выражение (99)). Это решение

$$p_n(t) = p_{n0} + \tau_p G e^{-t/\tau_p} \quad (101)$$

показано на рис. 33, б.

Рассмотренный пример лежит в основе метода измерения времени жизни неосновных носителей [46]. Соответствующая экспериментальная установка схематически изображена на рис. 33, в. Избыточные носители здесь генерируются однородно в образце короткими световыми импульсами. Если по образцу пропускается постоянный ток, появление фотоносителей вызывает соответствующее уменьшение падения напряжения на образце. Наблюдая релаксацию фотопроводимости в промежутке между возбуждающими импульсами на осциллографе, определяют время жизни τ

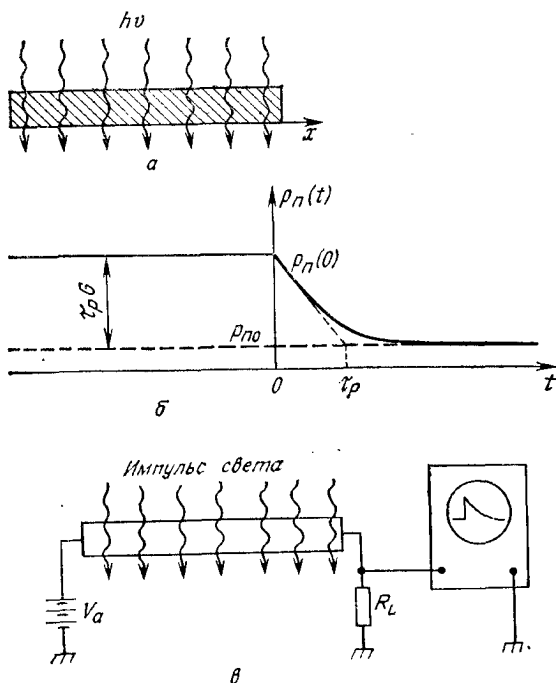


Рис. 33. Релаксация фотовозбужденных носителей [46].

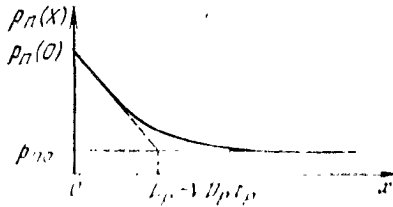
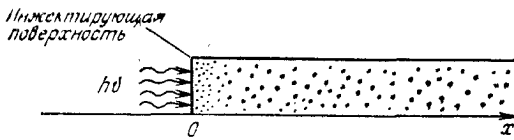
a — образец *n*-типа при постоянном освещении; *б* — зависимость концентрации неосновных носителей (дырок) от времени; *в* — схема эксперимента для определения времени жизни неосновных носителей.

(длительность световых импульсов при этом должна быть много меньше τ).

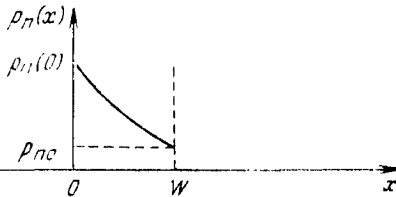
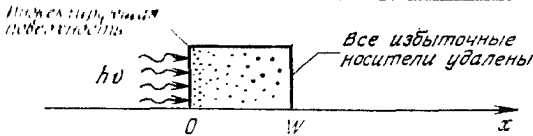
Стационарная инжекция с одной стороны образца. На рис. 34, *a* рассмотрен пример, когда избыточные носители инжектируются с одной стороны образца, например при освещении коротковолновым светом, который генерирует электронно-дырочные пары в тонком приповерхностном слое. Из графиков, приведенных на рис. 27, для фотонов с $h\nu = 3,5 \text{ эВ}$ $\alpha = 10^6 \text{ см}^{-1}$, т. е. интенсивность такого излучения ослабляется в *e* раз в приповерхностном слое толщиной 100 Å, где и генерируются в основном избыточные носители тока.

В стационарных условиях ($\frac{\partial \rho_n}{\partial t} = 0$) поверхностная генерация приводит к возникновению градиента концентрации неосновных носителей в приповерхностной области образца. При этом уравнение (97а) принимает вид

$$-\frac{\rho_n - \rho_{n0}}{\tau_p} + D_p \frac{\partial^2 \rho_n}{\partial x^2} = 0. \quad (102)$$



а



б

Рис. 34. Стационарная инжекция с одной стороны образца.

а — полубесконечный образец; б — образец конечных размеров.

Его решение, удовлетворяющее граничным условиям $p_n(x=0) = p_n(0)$ и $p_n(x \rightarrow \infty) = p_{n0}$, имеет вид (рис. 34, а)

$$p_n(x) = p_{n0} + [p_n(0) - p_{n0}] e^{-x/L_p}, \quad (103)$$

где $L_p \equiv \sqrt{D_n \tau_p}$ — диффузионная длина неосновных носителей. В чистых образцах Ge и Si L_p и $L_n \equiv \sqrt{D_n \tau_n}$ достигают значений порядка 1 см, а в GaAs $L_{\text{max}} \approx 10^{-2}$ см.

Изменим теперь второе граничное условие, полагая, что все избыточные носители удаляются из образца при $x=W$, т. е. $p_n(W) = p_{n0}$. В этом случае решением уравнения (102) является функция (рис. 34, б)

$$p_n(x) = p_{n0} + (p_n(0) - p_{n0}) \frac{\text{sh} \frac{W-x}{L_p}}{\text{sh} \frac{W}{L_p}}. \quad (104)$$

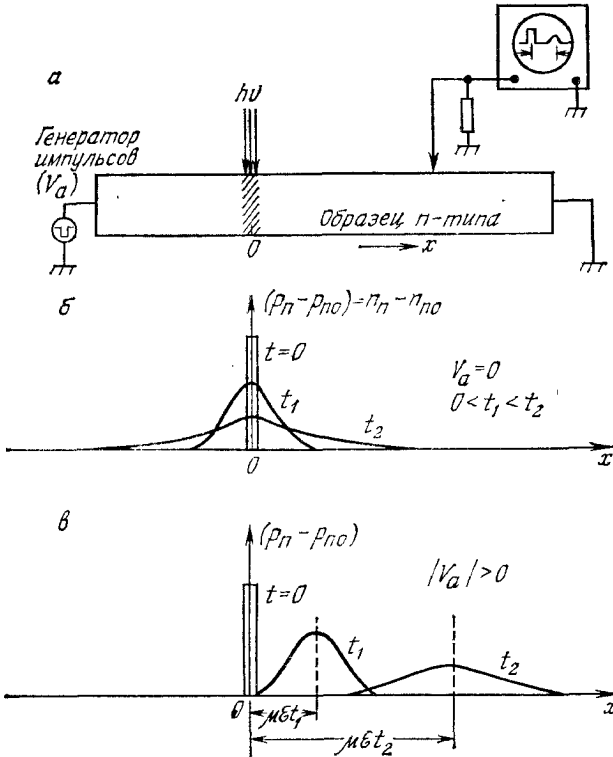


Рис. 35. Нестационарный процесс диффузионного расплывания локального «пакета» основных носителей [73].

a — схема эксперимента; *б* — поле отсутствует; *в* — поле наложено.

С помощью выражения (94а) можно рассчитать плотность дырочного тока при $x = W$:

$$J_p = -qD_p \left. \frac{\partial p}{\partial x} \right|_{x=W} = q(p_n(0) - p_{n0}) \frac{D_p}{L_p} \frac{1}{\text{sh} \frac{W}{L_p}}. \quad (105)$$

Как показано в гл. 3, это выражение фактически определяет коэффициент усиления тока в биполярном транзисторе.

Стационарная и нестационарная диффузия. Рассмотрим теперь случай, когда избыточные носители генерируются узким импульсом света в локальном участке образца (рис. 35, *a*). Распределение концентрации неравновесных дырок в образце по окончании возбуждающего светового импульса ($G = 0$) определяется уравнением

$$\frac{\partial p_n}{\partial t} = -\frac{p_n - p_{n0}}{\tau_p} - \mu_p \mathcal{E} \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2}, \quad (106)$$

которое следует из уравнения (97а) в предположении, что электрическое поле в образце однородно ($\frac{\partial \mathcal{E}}{\partial x} = 0$). В отсутствие внешнего поля ($\mathcal{E} = 0$) решение уравнения (106), удовлетворяющее этим «импульсным» начальным условиям, имеет вид

$$p_n(x, t) = \frac{N}{\sqrt{4\pi D_p t}} \exp\left(-\frac{x^2}{4D_p t} - \frac{t}{\tau_p}\right) + p_{n0}, \quad (107)$$

где N — полное число избыточных неосновных носителей, генерированных световым импульсом, на единицу площади поперечного сечения образца. Из рис. 35, б видно, что неосновные носители диффундируют от точки инжекции и при этом рекомбинируют.

Для образца во внешнем электрическом поле в выражении (107) x следует заменить на $x - \mu_p \mathcal{E} t$. Характер такого решения показан на рис. 35, в. В этом случае первоначально узкий «пакет» неосновных носителей диффузионно расплывается и рекомбинирует, одновременно смещаясь во внешнем электрическом поле к отрицательному контакту образца с дрейфовой скоростью $\mu_p \mathcal{E}$. Рассмотренный пример представляет собой известный эксперимент Хайнса—Шокли по измерению дрейфовой подвижности в полупроводниках [73]. Зная длину образца и приложенное электрическое поле и измерив время задержки между инжектирующим импульсом и детектированным импульсом, можно легко определить подвижность $\mu = x/\mathcal{E} t$.

Поверхностная рекомбинация. При учете поверхностной рекомбинации граничное условие на поверхности образца ($x = 0$) имеет вид

$$qD_n \left. \frac{\partial p_n}{\partial x} \right|_{x=0} = qS_p (p_n(0) - p_{n0}). \quad (108)$$

Оно означает, что число рекомбинирующих на единице площади поверхности кристалла неосновных носителей пропорционально отклонению их концентрации на этой границе от равновесного значения p_{n0} . Коэффициент пропорциональности S_p имеет размерность скорости и называется скоростью поверхностной рекомбинации. Распределение концентрации неосновных носителей в однородно освещенном образце определяется уравнением непрерывности

$$G - \frac{p_n - p_{n0}}{\tau_p} + D_p \frac{\partial^2 p_n}{\partial x^2} = 0 \quad (109)$$

(стационарное освещение). Пренебрегая взаимодействием границ, в качестве второго граничного условия при $x \rightarrow \infty$ используем выражение (99). Решение уравнения (109), удовлетворяющее указанным граничным условиям, имеет вид

$$p_n(x) = p_{n0} + \tau_p G \left[1 - \frac{\tau_p S_p \exp(-x/L_p)}{L_p + \tau_p S_p} \right]. \quad (110)$$

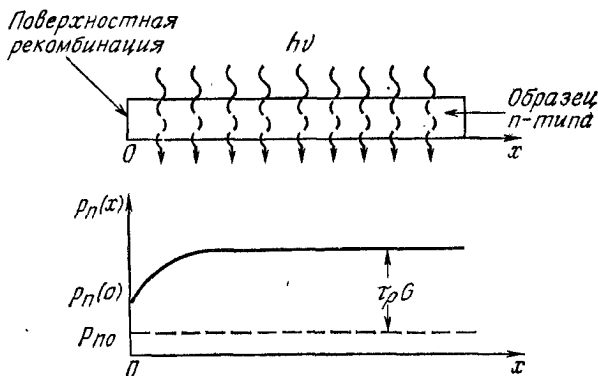


Рис. 36. Распределение неосновных носителей вблизи границы образца при учете поверхностной рекомбинации.

Его характер иллюстрирует график на рис. 36. При $S_p \rightarrow 0$ концентрация дырок однородна по длине образца ($p_n(x) \rightarrow p_{n0} + \tau_p G$); при $S_p \rightarrow \infty$ $p_n(x) \rightarrow p_{n0} + \tau_p G [1 - \exp(-x/L_p)]$, т. е. концентрация неосновных носителей на границе образца вне зависимости от интенсивности освещения остается равной термодинамически равновесному значению p_{n0} . По аналогии с рекомбинационным процессом в объеме полупроводника, где при малых уровнях инжекции обратное время жизни неосновных носителей τ^{-1} равно $\sigma_p v_{th} N_{st}$, скорость поверхностной рекомбинации определяется выражением

$$S_p = \sigma_p v_{th} N_{st}, \quad (111)$$

где N_{st} — число поверхностных рекомбинационных центров на единицу площади границы кристалла.

ЛИТЕРАТУРА

1. Duniap W. C. An Introduction to Semiconductors, Wiley, N. Y., 1957.
2. Madelung O. Physics of III—V Compounds, Wiley, N. Y., 1964.
3. Moll J. L. Physics of Semiconductors, McGraw-Hill, N. Y., 1964.
4. Moss T. S., Ed., Handbook on Semiconductors. Vols. 1—4, North-Holland, Amsterdam, 1980.
5. Smith R. A. Semiconductors, 2nd ed., Cambridge University Press, London, 1979.
6. See, for example, Kittel C. Introduction to Solid State Physics, Wiley, N. Y., 1976.
7. Willardson R. K., Beer A. C., Eds. Semiconductors and Semimetals, Vol. 2, Physics of III—V Compounds, Academic, N. Y., 1966.
8. Pearson W. B. Handbook of Lattice Spacings and Structure of Metals and Alloys, Pergamon, N. Y., 1967.
9. Casey H. C., Jr., Panish M. B. Heterostructure Lasers, Academic, N. Y., 1978.
10. Brillouin L. Wave Propagation in Periodic Structures, 2nd ed., Dover, N. Y., 1963.

11. Ziman J. M. Principles of the Theory of Solids, Cambridge University Press, London, 1964.
12. Cohen M. L. Pseudopotential Calculations for II—VI Compounds, Thomas D. G., Ed., II—VI Semiconducting Compounds, Benjamin W. A., N. Y., 1967, p. 462.
13. Kittel C. Quantum Theory of Solids, Wiley, N. Y., 1963.
14. Allen L. C. Interpolation Scheme for Energy Bands in Solids, *Phys. Rev.*, **98**, 993 (1955).
15. Herman F. The Electronic Energy Band Structure of Silicon and Germanium. *Proc. IRE*, **43**, 1703 (1955).
16. Phillips J. C. Energy-Band Interpolation Scheme Based on a Pseudopotential, *Phys. Rev.*, **112**, 685 (1958).
17. Chelikowsky J. R., Cohen M. L. Nonlocal Pseudopotential Calculations for the Electronic Structure of Eleven Diamond and Zinc-Blende Semiconductors, *Phys. Rev.*, **B14**, 556 (1976).
18. Neuberger M. Germanium Data Sheets, DS-143 (Feb. 1965, Oct. 1960); Silicon Data Sheets, DS-137 (May 1964, July 1968); Gallium Arsenide Data Sheets, DS-144 (Apr. 1965, Sept. 1967), Compiled from Data Sheets of Electronic Properties Information Center (EPIC), Hughes Aircraft Co., Culver City, Calif.
 - (a) Dalven R. A. Review of the Semiconductor Properties of PbTe, PbSe, PbS and PbO, *Infrared Phys.*, **9**, 141 (1969).
 - (b) Strzalkowski I., Joshi S., Crowel C. R. Dielectric Constant and Its Temperature Dependence for GaAs, CdTe and ZnSe, *Appl. Phys. Lett.*, **28**, 350 (1976).
 - (c) Jensen G. H. Temperature Dependence of Bandgap in ZnO from Reflection Data, *Phys. Status Solidi*, **64**, K51 (1974).
 - (d) Kesamany F. P. GaN: Band Structure, Properties and Potential Applications, *Sov. Phys. Semicond.*, **8**, 147 (1974).
19. Ziman J. M. Electrons and Phonons, Clarendon, Oxford, 1960.
20. Thurmond C. D. The Standard Thermodynamic Function of the Formation of Electrons and Holes in Ge, Si, GaAs and GaP, *J. Electrochem. Soc.*, **122**, 1133 (1975).
21. Paul W., Warschauer D. M., Eds., Solids under Pressure, McGraw-Hill, N. Y., 1963.
22. Blackmore J. S. Carrier Concentrations and Fermi Levels in Semiconductors, *Electron. Commun.*, **29**, 131 (1952).
23. Hall R. N., Racette J. H. Diffusion and Solubility of Copper in Extrinsic and Intrinsic Germanium, Silicon, and Gallium Arsenide, *J. Appl. Phys.*, **35**, 379 (1964).
24. Milnes A. G. Deep Impurities in Semiconductors, Wiley, N. Y., 1973.
25. Hermannson J., Phillips J. C. Pseudopotential Theory of Excitation and Impurity States, *Phys. Rev.*, **150**, 652 (1966).
26. Callaway J., Hughes A. J. Localized Defects in Semiconductors, *Phys. Rev.*, **156**, 860 (1967).
27. Conwell E. M. Properties of Silicon and Germanium, Part II, *Proc. IRE*, **46**, 1281 (1958).
28. Sze S. M., Irvin J. C. Resistivity, Mobility, and Impurity Levels in GaAs, Ge, and Si at 300 K, *Solid State Electron.*, **11**, 599 (1968).
29. Bullis W. M. Properties of Gold in Silicon, *Solid State Electron.*, **9**, 143 (1966).
30. Wolfstirn K. B. Holes and Electron Mobilities in Doped Silicon from Radio Chemical and Conductivity Measurements, *J. Phys. Chem. Solids*, **16**, 279 (1960).
31. Shockley W. Electrons and Holes in Semiconductors, D. Van Nostrand, Princeton, N. J., 1950.
32. Grove A. S. Physics and Technology of Semiconductor Devices, Wiley, N. Y., 1967.

33. Bardeen J., Shockley W. Deformation Potentials and Mobilities in Nonpolar Crystals, *Phys. Rev.*, **80**, 72 (1950).
34. Conwell E., Weisskopf V. F. Theory of Impurity Scattering in Semiconductors, *Phys. Rev.*, **77**, 388 (1950).
35. Ehrenreich H. Band Structure and Electron Transport in GaAs, *Phys. Rev.*, **120**, 1951 (1960).
36. Prince M. B. Drift Mobility in Semiconductors I, Germanium, *Phys. Rev.*, **92**, 681 (1953).
37. Jacoboni C., Canali C., Ottaviani G., Quaranta A. A. A Review of Some Charge Transport Properties of Silicon, *Solid State Electron.*, **20**, 77 (1977).
38. Beadle W. F., Plummer R. D., Tsai J. C. C. Quick Reference Manual for Semiconductor Engineers (в печати).
39. Smits F. M. Measurement of Sheet Resistivities with the Four-Point Probe, *Bell Syst. Tech. J.*, **37**, 711 (1958).
40. Irvin J. C. Resistivity of Bulk Silicon and of Diffused Layers in Silicon, *Bell Syst. Tech. J.*, **41**, 387 (1962).
41. Hall E. H. On a New Action of the Magnet on Electric Currents, *Am. J. Math.*, **2**, 287 (1979).
42. Van der Pauw L. J. A Method of Measuring Specific Resistivity and Hall Effect of Disc or Arbitrary Shape, *Phillips Res. Rep.*, **13**, 1 (Feb. 1958).
43. Sah C. T., Noyce R. N., Shockley W. Carrier Generation and Recombination in $p-n$ Junction and $p-n$ Junction Characteristics, *Proc. IRE*, **45**, 1228 (1975).
44. Hall R. N. Electron-Hole Recombination in Germanium, *Phys. Rev.*, **87**, 387 (1952).
45. Shockley W., Read W. T. Statistics of the Recombination of Holes and Electrons, *Phys. Rev.*, **87**, 835 (1952).
46. Stevenson D. T., Keyes R. J. Measurement of Carrier Lifetime in Germanium and Silicon, *J. Appl. Phys.*, **26**, 190 (1955).
47. Gartner W. W. Spectral Distribution of the Photomagnetic Electric Effect, *Phys. Rev.*, **105**, 823 (1957).
48. Brockhouse B. N., Iyengar P. K. Normal Modes of Germanium by Neutron Spectrometry, *Phys. Rev.*, **111**, 747 (1958).
49. Brockhouse B. N. Lattice Vibrations in Silicon and Germanium, *Phys. Rev. Lett.*, **2**, 256 (1959).
50. Waugh J. L. T., Dolling G. Crystal Dynamics of Gallium Arsenide, *Phys. Rev.*, **132**, 2410 (1963).
51. Dash W. C., Newman R. Intrinsic Optical Absorption in Single-Crystal Germanium and Silicon at 77° K and 300° K, *Phys. Rev.*, **99**, 1151 (1955).
52. Philipp H. R., Taft E. A. Optical Constants of Germanium in the Region 1 to 10 eV, *Phys. Rev.* **113**, 1002 (1959); Optical Constants of Silicon in the Region 1 to 10 eV, *Phys. Rev. Lett.*, **8**, 13 (1962).
53. Hill D. E. Infrared Transmission and Fluorescence of Doped Gallium Arsenide, *Phys. Rev.*, **133**, A866 (1964).
54. Casey H. C., Jr., Sell D. D., Wecht K. W. Concentration Dependence of the Absorption Coefficient for n - and p -type GaAs between 1,3 and 1,6 eV, *J. Appl. Phys.*, **46**, 250 (1975).
55. Ho C. Y., Powell R. W., Liley P. E. Thermal Conductivity of the Elements — A Comprehensive Review, American Chemical Society and American Institute of Physics, N. Y., 1975.
56. Holland M. G., Phonon Scattering in Semiconductors from Thermal Conductivity Studies, *Phys. Rev.*, **134**, A471 (1964).
57. Armstrong B. H. Thermal Conductivity in SiO₂, in Pantelides S. T., Ed., The Physic of SiO₂ and Its Interface, Pergamon, N. Y., 1978.
58. Aspnès D. E. GaAs Lower Conduction-Band Minima: Ordering and Properties, *Phys. Rev.*, **B14**, 5331 (1976).

59. Smith P., Inoue M., Frey J. Electron Velocity in Si and GaAs at Very High Electric Fields, *Appl. Phys. Lett.*, **37**, 797 (1980).
60. Ruch J. G., Kino G. S. Measurement of the Velocity — Field Characteristics of Gallium Arsenide, *Appl. Phys. Lett.*, **10**, 40 (1967).
61. Okamoto H., Ikeda M. Measurement of the Electron Drift Velocity in Avalanche GaAs Diodes, *IEEE Trans. Electron. Devices*, **ED-23**, 372 (1976).
62. Kramer B., Mircea A. Determination of Saturated Electron Velocity in GaAs, *Appl. Phys. Lett.*, **26**, 623 (1975).
63. Logan R. A., Sze S. M. Avalanche Multiplication in Ge and GaAs $p - n$ Junctions, Proc. Int. Conf. Phys. Semicond., Kyoto, and *J. Phys. Soc. Jpn. Suppl.*, **21**, 434 (1966).
64. Grant W. N. Electron and Hole Ionization Rates in Epitaxial Silicon at High Electric Fields, *Solid State Electron.*, **16**, 1189 (1973).
65. Glover G. H. Charge Multiplication in Au — SiC (6H) Schottky Junction, *J. Appl. Phys.*, **46**, 4842 (1975).
66. Pearsall T. P., Capasso F., Nahory R. E., Pollack M. A., Chelikowsky J. R. The Band Structure Dependence of Impact Ionization by Hot Carriers in Semiconductors GaAs, *Solid State Electron.*, **21**, 297 (1978).
67. Umeha I., Choudhury A. N. M. M., Robson P. N., Ionization Coefficients Measured in Abrupt InP Junction, *Appl. Phys. Lett.*, **36**, 302 (1980).
68. Logan R. A., White H. G. Charge Multiplication in GaP $p - n$ Junctions, *J. Appl. Phys.*, **36**, 3945 (1965).
69. Pearsall T. P. Impact Ionization Rates for Electrons and Holes in $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$, *Appl. Phys. Lett.*, **36**, 218 (1980).
70. Pearsall T. P., Nahory R. E., Pollack M. A. Impact Ionization Rates for Electrons and Holes in $\text{GaAs}_{1-x}\text{Sb}_x$ Alloys, *Appl. Phys. Lett.*, **28**, 403 (1976).
71. Baraff G. A. Distribution Junctions and Ionization Rates for Hot Electrons in Semiconductors, *Phys. Rev.*, **128**, 2507 (1962).
72. Crowell C. R., Sze S. M. Temperature Dependence of Avalanche Multiplication in Semiconductors, *Appl. Phys. Lett.*, **9**, 242 (1966).
73. Haynes J. R., Shockley W. The Mobility and Life of Injected Holes and Electrons in Germanium, *Phys. Rev.*, **81**, 835 (1951).
74. Kroemer H. The Einstein Relation for Degenerate Carrier Concentration, *IEEE Trans. Elec. Dev.* **ED-25**, 850 (1978).
75. Thornber K. K. Relation of Drift Velocity to Low-Field Mobility and High Field Saturation Velocity, *J. Appl. Phys.*, **51**, 2127 (1980).
76. Thornber K. K. Application of Scaling to Problems in High-Field Electronic Transport, *J. Appl. Phys.* **52**, 279 (1981).

Часть II

Биполярные приборы

Глава 2

ПЛОСКОСТНЫЕ ДИОДЫ

2.1. ВВЕДЕНИЕ

Плоскостные $p-n$ -переходы играют важную роль в современной электронике как самостоятельные изделия (диоды) и для понимания работы других полупроводниковых приборов. Теория плоскостных $p-n$ -переходов лежит в основе объяснения работы полупроводниковых приборов. Основы теории вольт-амперных характеристик плоскостных $p-n$ -переходов были заложены Шокли [1] и развиты Са, Нойсом и Шокли [2], а также Моллом [3].

В начале главы кратко описаны основы технологии полупроводниковых приборов, используемой не только для получения $p-n$ -переходов, но и для изготовления большинства других полупроводниковых приборов. Затем с помощью уравнений, рассмотренных в гл. 1, получены статические и динамические характеристики идеальных $p-n$ -переходов. Далее обсуждаются отклонения характеристик реального прибора от характеристик идеального прибора, обусловленные генерацией и рекомбинацией носителей в обедненном слое $p-n$ -перехода, высоким уровнем инжекции носителей и наличием последовательно включенного с ним сопротивления. Подробно рассмотрен пробой $p-n$ -перехода, и особое внимание уделено пробую, вызванному лавинным размножением носителей. Один раздел посвящен переходным процессам и шумам в $p-n$ -переходе. Плоскостной $p-n$ -переход является двухполюсником. В зависимости от профиля легирующей примеси, геометрии прибора и режима работы прибор может выполнять различные схемные функции (разд. 2.7). В конце главы рассмотрены такие приборы, как гетеропереходы, — переходы, сформированные между различными полупроводниками (например, между Ge n -типа и $GaAs$ p -типа), играющие большую роль в полупроводниковой электронике.

2.2. ОСНОВЫ ТЕХНОЛОГИИ

В этой главе речь преимущественно пойдет о кремниевой технологии [4], поскольку ее развитие намного опережает технологию на других полупроводниковых материалах. Несколько важ-

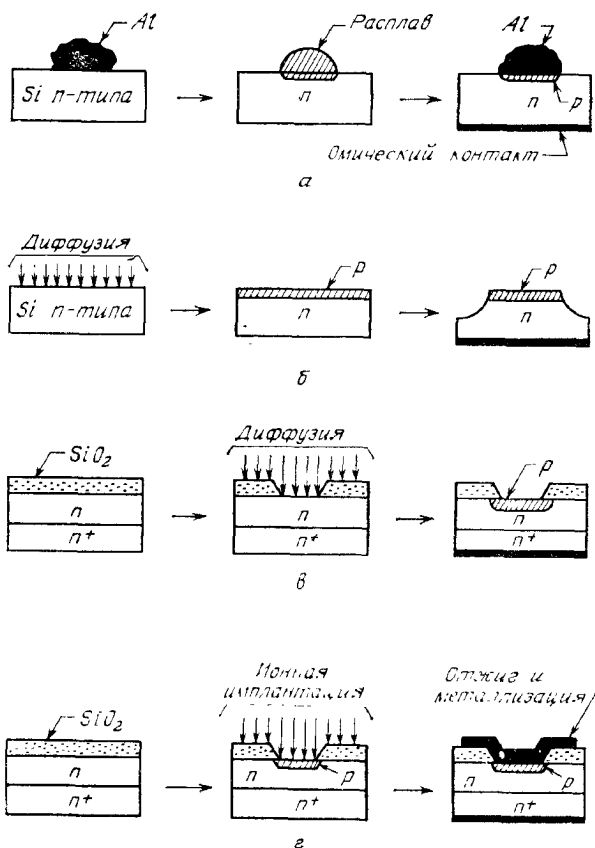


Рис. 1. Методы изготовления полупроводниковых приборов.

a — сплавной переход; *б* — диффузионный мезопереход; *в* — диффузионный планарный переход на эпитаксиальной подложке; *г* — ионно-имплантированный переход.

нейших методов изготовления полупроводниковых приборов схематически показано на рис. 1. При методе сплавления [5] (рис. 1, *a*) небольшую таблетку алюминия помещают на поверхность кремниевой пластины с проводимостью *n*-типа, имеющей ориентацию (111). Затем пластину с таблеткой нагревают до температуры, немного превышающей температуру эвтектики (~580 °С для системы Al—Si). При расплавлении таблетки образуется небольшая капля смеси Al—Si, которая с последующим понижением температуры начинает затвердевать. В результате образуется рекристаллизованная область, насыщенная акцепторной примесью и имеющая ту же кристаллографическую ориентацию, что и исходная пластина. Таким образом, в подложке проводимостью *n*-типа сформи-

рована сильнолегированная область p -типа (p^+). Алюминиевый выступ на поверхности пластины можно использовать как контакт к области p -типа. Чтобы получить омический контакт к пластине n -типа, на ее нижнюю поверхность напыляют сплав Au—Sb, содержащий $\sim 0,1\%$ Sb, и вплавляют его при температуре 400°C для формирования сильнолегированного слоя n -типа (n^+). При использовании пластины p -типа роли алюминия и сплава Au—Sb меняются. Первый служит для создания омического контакта p^+ — p -типа, а второй — для получения n^+ — p -перехода. При сплавном методе нельзя контролировать положение p — n -перехода, поскольку оно сильно зависит от температуры и длительности цикла сплавления.

Диффузионный метод (или метод диффузии в твердой фазе), разработанный в 1956 г., позволяет более точно управлять распределением примеси [6]. Рисунок 1, б иллюстрирует диффузионный метод получения p — n -перехода с мезоструктурой. В подложку n -типа проводят диффузию примеси p -типа, например бора из соединения BВг_3 . После диффузии определенные участки поверхности защищают от действия травителей, например, слоем воска или металлическим покрытием. Незащищенные участки поверхности подложки удаляют с помощью травления. В результате образуются мезоструктуры.

Более точный контроль геометрических размеров диффузионного перехода был достигнут за счет использования изолирующей пленки, которая препятствует диффузии большинства донорных и акцепторных примесей в глубь подложки [7]. Рисунок 1, в иллюстрирует типичный пример такого процесса. На поверхности кремневой подложки при высокой температуре выращивают тонкий слой двуокиси кремния толщиной ~ 1 мкм. Литографическими методами (например, фотолитографией, рентгенолитографией, электронной литографией) удаляют определенные участки окисла, образуя в нем окна или более сложные рисунки. Примеси диффундируют в открытые участки поверхности кремния, и p — n -переходы образуются только в местах окон в окисле. Этот процесс, получивший название планарного [8], является основой технологии изготовления полупроводниковых приборов и интегральных схем¹⁾.

Эпитаксиальные структуры типа показанного на рис. 1, в [9] обычно используются в планарной технологии для уменьше-

¹⁾ В планарном процессе используются такие давно известные технологические методы, как окисная защита при диффузии. Принципиальная особенность планарного процесса заключается в сочетании методов, позволяющем получить большую точность контроля размеров, формы электродов и диффузионных областей приборов. Однако в этом случае требуется, чтобы поверхность пластины была гладкой без рельефа, в противном случае слой фоторезиста не будет ровным и появятся дефекты.

ния последовательного сопротивления. Слово «эпитаксия» имеет греческое происхождение: «эпи» означает «на» и «таксис» — «расположено в порядке». Эпитаксией называют метод выращивания путем химической реакции на поверхности кристалла тонких слоев полупроводниковых материалов с сохранением кристаллической структуры исходного кристалла. Таким методом на поверхности сильнолегированной низкоомной подложки выращивают высокоомные эпитаксиальные слои, добиваясь желаемых электрических свойств и механической прочности.

На рис. 1, *г* показан *p-n*-переход, полученный с помощью ионной имплантации [10]. В настоящее время этот метод позволяет наиболее точно контролировать распределение примесей. Ионную имплантацию можно производить при комнатной температуре, а возникающие в процессе имплантации дефекты кристаллической решетки устраняют путем последующего отжига при температуре 700 °С и ниже. Следовательно, ионная имплантация является относительно низкотемпературным процессом по сравнению с диффузией, которая обычно проводится при температуре 1000 °С и выше.

Рассмотрим кратко четыре основных процесса планарной технологии: эпитаксиальное выращивание, окисление, диффузию примесей и ионную имплантацию.

Эпитаксиальные слои можно получать методом выращивания из газовой фазы [4]. Основная реакция, в результате которой на подложке наращиваются кремниевые слои, состоит в восстановлении тетрахлорсилана в водороде: $\text{SiCl}_4 + 2\text{H}_2 \rightarrow \text{Si (твердый)} + 4\text{HCl (газ)}$. Обычно кремниевые слои выращивают со скоростью ~ 1 мкм/мин при температуре 1200 °С и выше. Молярная концентрация SiCl_4 , т. е. отношение числа молекул SiCl_4 к общему числу молекул газовой смеси, составляет 0,01 %. Эпитаксиальные слои также получают выращиванием из жидкой фазы [11], что широко используется при работе со сложными полупроводниками. При молекулярно-лучевой эпитаксии удается контролировать с большой точностью состав полупроводниковой пленки в слоях толщиной порядка атомных размеров [12, 13].

Для создания пленок двуокиси кремния наиболее часто используют метод термического окисления [14] кремния, сопровождающийся химической реакцией $\text{Si (твердый)} + \text{O}_2$ (сухой кислород) $\rightarrow \text{SiO}_2$ (твердый) или $\text{Si (твердый)} + 2\text{H}_2\text{O (пар)} \rightarrow \text{SiO}_2$ (твердый) $+ 2\text{H}_2$. Можно показать, что в начале реакции толщина слоя окисла растет линейно со временем, а при длительном окислении — как квадратный корень из длительности [15]. При образовании слоя двуокиси кремния толщиной d расходуется слой кремния толщиной $0,45d$. Экспериментальные зависимости толщины окисла от времени и температуры реакции при выращивании в сухом кислороде и в парах воды приведены на рис. 2.

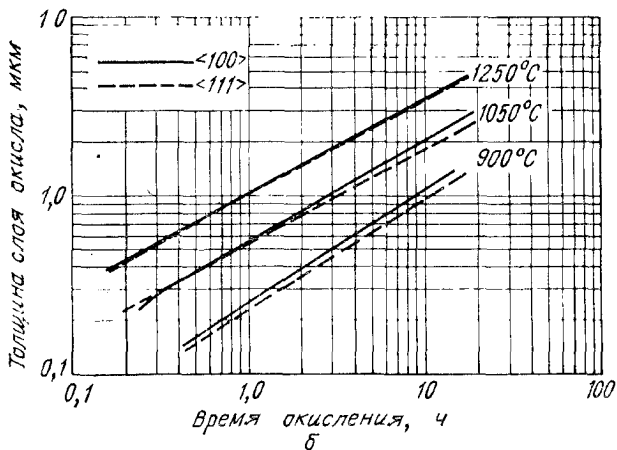
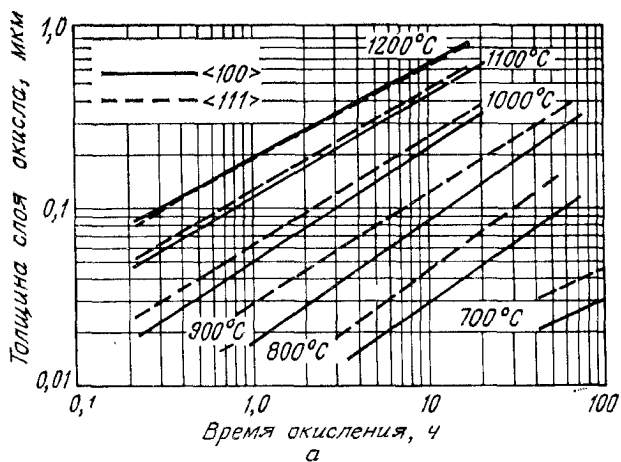


Рис. 2. Экспериментальные зависимости толщины окисла от времени окисления и температуры для двух ориентаций подложки [16]:

а — окисление в сухом кислороде; *б* — окисление в парах воды.

Для заданной температуры скорость окисления в парах воды в 5—10 раз выше, чем в сухом кислороде. Кроме того, при низких температурах скорость окисления зависит от ориентации кристалла [16].

Простой одномерный диффузионный процесс описывается законом Фика [17]

$$\frac{\partial C(x, t)}{\partial t} = D \frac{\partial^2 C(x, t)}{\partial x^2}, \quad (1)$$

где C — концентрация примеси, D — коэффициент диффузии. Это уравнение аналогично уравнению (96а) (гл. 1), но в нем от-

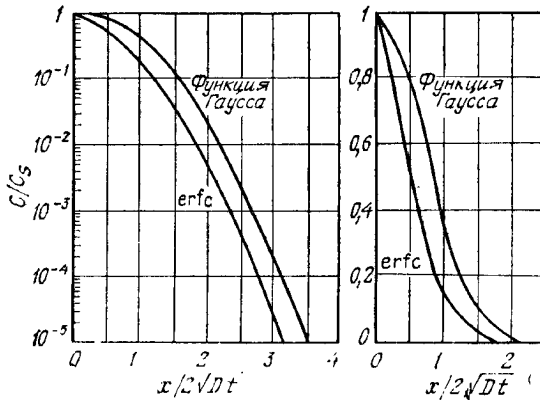


Рис. 3. Нормированные распределения концентрации для функции Гаусса и дополнительной функции ошибок (erfc) в полулогарифмическом и линейном масштабах [17].

сутствуют члены, связанные с генерацией, рекомбинацией и электрическим полем. При выполнении условий диффузии из «ограниченного источника», содержащего общее количество примеси S , решением уравнения (1) является функция Гаусса

$$C(x, t) = \frac{S}{\sqrt{\pi l t}} \exp\left(-\frac{x^2}{4Dt}\right). \quad (2a)$$

При выполнении условий «постоянной поверхностной концентрации», равной C_s , решением уравнения (1) является дополнительная функция ошибок

$$C(x, t) = C_s \operatorname{erfc}\left(\frac{x}{2\sqrt{Dt}}\right). \quad (26)$$

Зависимости нормированной концентрации от нормированного расстояния для обоих решений приведены на рис. 3. Диффузионные профили многих примесей действительно хорошо аппроксимируются полученными выражениями. Вместе с тем распределение многих примесей, например As в Si, имеют более сложные профили вследствие сильной зависимости процесса диффузии от концентрации примеси [18].

Коэффициент диффузии D в общем случае зависит от температуры и концентрации примеси. При низкой концентрации величина D не зависит от концентрации. (Концентрация примеси считается низкой, если она меньше, чем концентрация собственных носителей при температуре диффузии, например при 1100°C , $n_i \approx 10^{19}\text{см}^{-3}$, как это следует из рис. 11 гл. 1.) В ограниченном температурном диапазоне и при низкой концентрации коэффициент диффузии может быть представлен в виде

$$D(T) = D_0 \exp(-\Delta E/kT), \quad (3)$$

где D_0 — коэффициент диффузии при высокой температуре, ΔE — энергия активации процесса диффузии. Значения $D(T)$ для различных примесей в Ge, Si и GaAs при низкой их концентрации приведены на рис. 4 [19, 20]. При увеличении концентрации примеси коэффициент $D(T)$ начинает сильно зависеть от концентрации.

Коэффициент диффузии примеси связан с предельной растворимостью примеси в твердой фазе, равной максимальной концентрации примеси, которая может быть введена в твердую фазу при любой заданной температуре. Зависимости предельных растворимостей, наиболее важных примесей для кремния от температуры представлены на рис. 5 [21]. Из него следует, что в качестве примесей, обеспечивающих наиболее сильное легирование кремния n -типа, следует использовать мышьяк и фосфор, а для получения сильнолегированного кремния p -типа следует использовать бор.

Ионная имплантация представляет собой введение в подложку заряженных атомных частиц, обладающих определенной энергией, с целью изменения электрических, металлургических и химических свойств подложки. Обычно используют ионы с энергией в диапазоне 10—400 кэВ, а типичные дозы ионов варьируют от 10^{11} до 10^{16} ион/см². Основные достоинства ионной имплантации заключаются в следующем: 1) общая доза примеси, глубина профиля и поверхностная однородность строго контролируются; 2) процесс

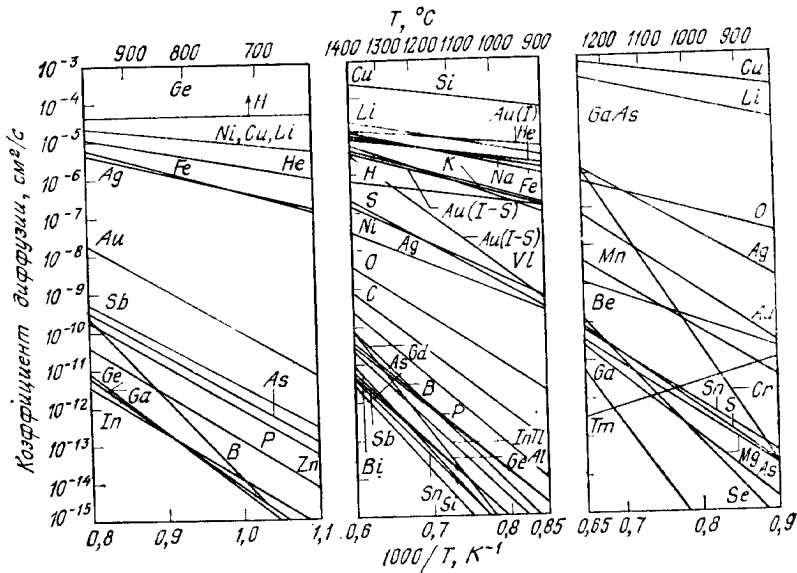


Рис. 4. Температурная зависимость коэффициентов диффузии примесей в Ge, Si и GaAs [19, 20].

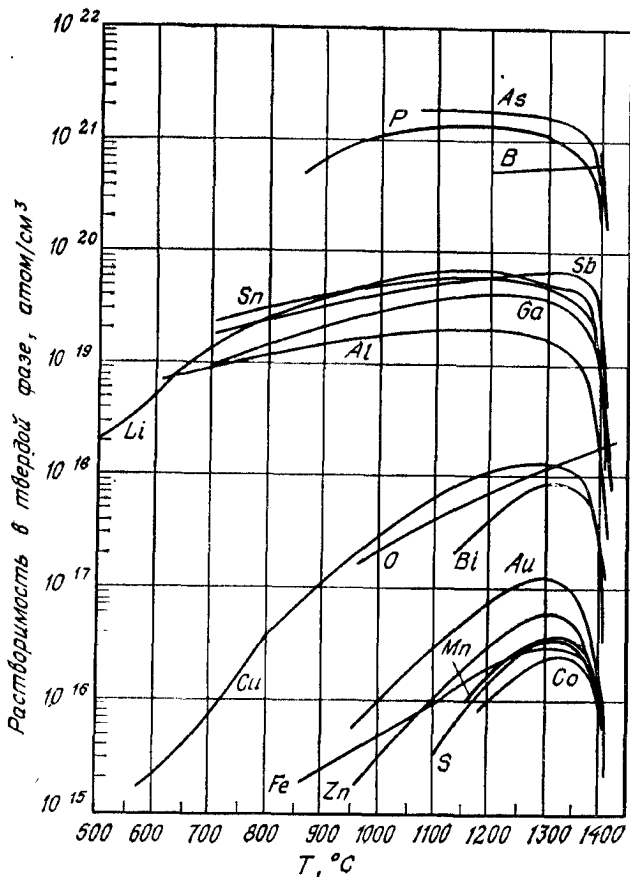


Рис. 5. Температурная зависимость растворимости различных элементов в Si [21].

протекает при низкой температуре; 3) имплантированная область точно совмещается с краем маски.

При имплантации ионным лучом бесконечно малого диаметра распределение ионов в подложке (вставка на рис. 6) описывается формулой ¹⁾ [10]

$$n(x, y) = \frac{s}{(2\pi)^{3/2} \Delta R_p \Delta R_T^2} \exp \left[- \left(\frac{x - R_p}{\sqrt{2} \Delta R_p} \right)^2 \right] \exp \left[- \left(\frac{y}{\sqrt{2} \Delta R_T} \right)^2 \right], \quad (4)$$

¹⁾ Точнее, правую часть выражения надо еще умножить на время воздействия пучка. — Прим. ред.

где s — интенсивность облучения, измеряемая числом ионов в 1 с; R_p — нормальная длина пробега по направлению падения ионов; ΔR_p — нормальная дисперсия; ΔR_T — поперечная дисперсия. Зависимость R_p от энергии имплантируемых ионов бора, фосфора и мышьяка в Si и SiO₂ приведена на рис. 6. Нормальная длина пробега приблизительно линейно возрастает с энергией. Ионы бора имеют почти одинаковые нормальные длины пробега в Si и SiO₂; вместе с тем для фосфора и мышьяка нормальная длина пробега в SiO₂ примерно на 20 % меньше, чем в Si. Нормальная и поперечная дисперсия для этих ионов показана на рис. 7. Дисперсия также возрастает с увеличением энергии ионов. Отношение ΔR_p к R_p во всем диапазоне энергий лежит в интервале 0,2—0,5.

При равномерном сканировании имплантирующим лучом бесконечно малого диаметра по подложке концентрация легирующей примеси перестает зависеть от координат при удалении от края

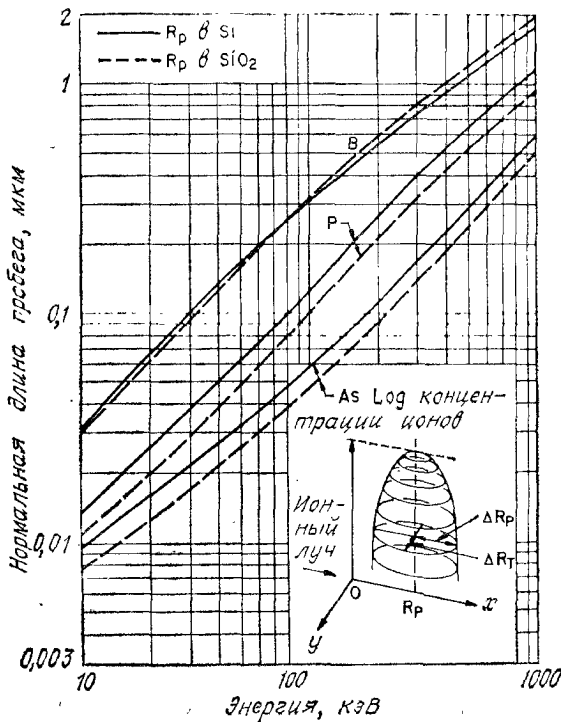


Рис. 6. Зависимость нормальной длины пробега ионов бора, фосфора и мышьяка в Si (сплошные линии) и SiO₂ (штриховые линии) от энергии имплантации. На вставке показано распределение имплантированных ионов [10].

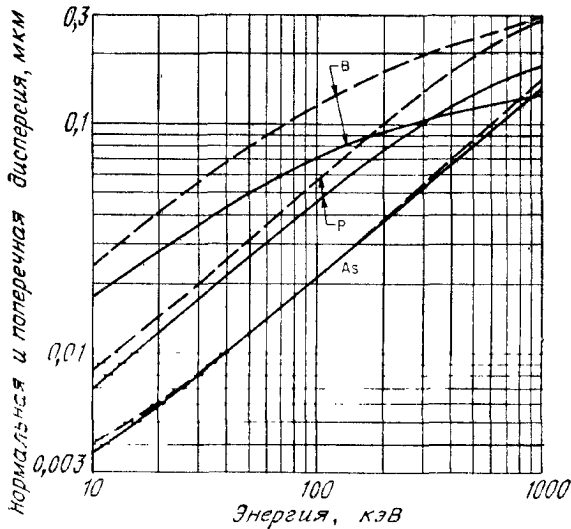


Рис. 7. Зависимость нормальной (сплошные линии) и поперечной (штриховые линии) дисперсии длины пробега ионов бора, фосфора и мышьяка в кремний от энергии имплантации [10].

области сканирования на расстояние, равное нескольким ΔR_T . В этом случае распределение примеси принимает вид

$$n(x) = \frac{\varphi}{\sqrt{2\pi}\Delta R_p} \exp\left[-\left(\frac{x - R_p}{\sqrt{2}\Delta R_p}\right)^2\right], \quad (4a)$$

что соответствует распределению Гаусса с общим числом ионов на единицу площади φ и максимальной концентрации $\varphi/(\sqrt{2\pi}\Delta R_p)$ на глубине $x = R_p$.

Начиная с 1974 г. интенсивно изучается лазерная обработка полупроводников [22, 23]. Лазерное излучение высокой интенсивности (импульсная генерация рубинового лазера или непрерывная генерация аргонового лазера) способно устранить нарушения кристаллической решетки, вызванные ионной имплантацией, а также рекристаллизовать аморфные полупроводниковые пленки. Потенциальные преимущества лазерного процесса обработки состоят в следующем: 1) в управлении глубиной расположения отжигаемого слоя и профиля примеси за счет избирательного поглощения лазерного излучения и времени действия лазерного импульса или сканирования луча. Особенность лазерного излучения заключается в возможности активировать имплантированную примесь без ее перераспределения; 2) в высокой степени локализации процесса обработки вплоть до микронных масштабов, поскольку луч лазера допускает фокусировку до таких размеров; 3) в ре-

кристаллизации материала из аморфного слоя на кристаллической подложке или образовании поликристаллических пленок с большим размером зерна из пленок, нанесенных на изоляторы.

Практически большинство примесных профилей в $p-n$ -переходах можно отнести к одному из двух предельных случаев: резкому переходу со ступенчатым распределением примеси и плавному переходу с линейным распределением примеси (рис. 8). Ступенчатая аппроксимация обеспечивает адекватное приближение для сплавных переходов, мелких диффузионных переходов и ионно-имплантированных переходов. Линейная аппроксимация пригодна для глубоких диффузионных переходов.

В результате планарного процесса возникает еще одно важное явление. При образовании $p-n$ -перехода путем диффузии через

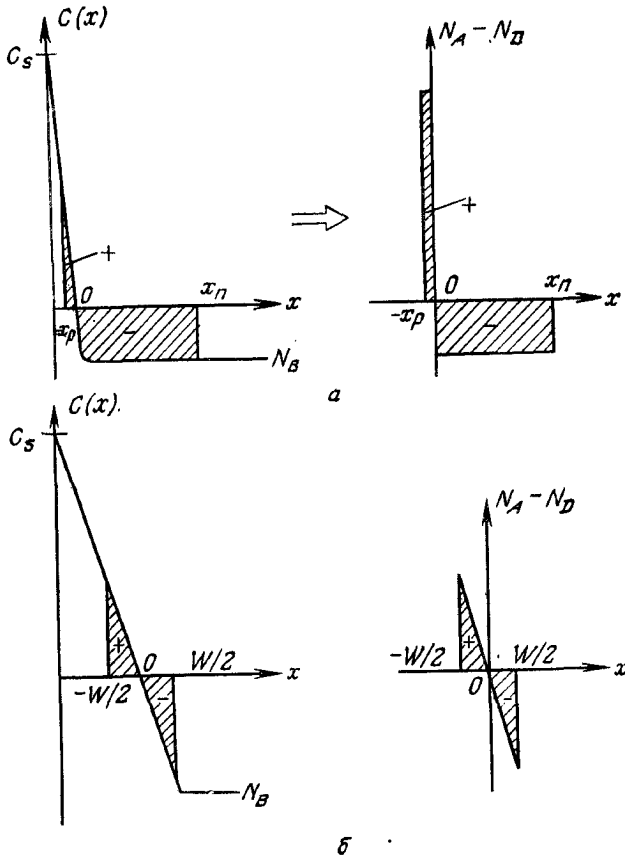


Рис. 8. Аппроксимированное распределение примеси.

a — резкий переход со ступенчатым распределением; b — плавный переход с линейным распределением.

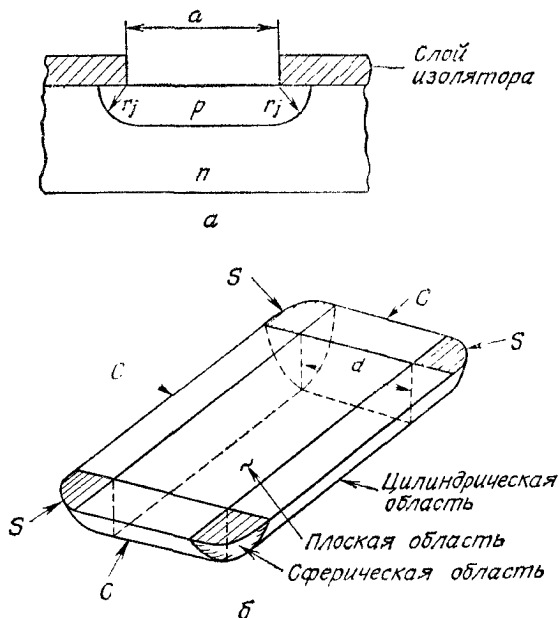


Рис. 9. Форма $p-n$ -перехода, полученного путем планарного диффузионного процесса (а) (r_j — радиус кривизны) и образование областей приблизительно цилиндрической и сферической формы при диффузии через прямоугольное окно в маске (б) [24].

окно в изолирующем слое примеси диффундируют не только в глубь полупроводника, но и в боковых направлениях. Поэтому боковые границы плоскостного $p-n$ -перехода имеют приблизительно цилиндрическую форму (рис. 9, а) [24]. Кроме того, если в защитной маске имелись острые углы, то форма $p-n$ -перехода вблизи них будет близка к сферической (рис. 9, б). Сферические и цилиндрические области оказывают существенное влияние на работу $p-n$ -перехода, особенно в режиме лавинного умножения [25], рассматриваемого в разд. 2.5.

2.3. ОБЕДНЕННЫЙ СЛОЙ И БАРЬЕРНАЯ ЕМКОСТЬ

2.3.1. Резкий переход

Диффузионный потенциал и ширина обедненного слоя. Для резкого (ступенчатого) перехода характерно резкое изменение концентрации примесей в полупроводнике (от концентрации доноров N_D до концентрации акцепторов N_A) (рис. 10, а). В частности, если $N_A \gg N_D$, то резкий p^+-n -переход оказывается несимме-

тричным. Вначале рассмотрим такой переход в состоянии теплового равновесия, когда отсутствует приложенное напряжение и не протекает ток. Из уравнений (33) и (93а) гл. 1 следует

$$J_n = 0 = q\mu_n \left(n\mathcal{E} + \frac{kT}{q} \frac{dn}{dx} \right) = \mu_n n \frac{dE_F}{dx}, \quad (5)$$

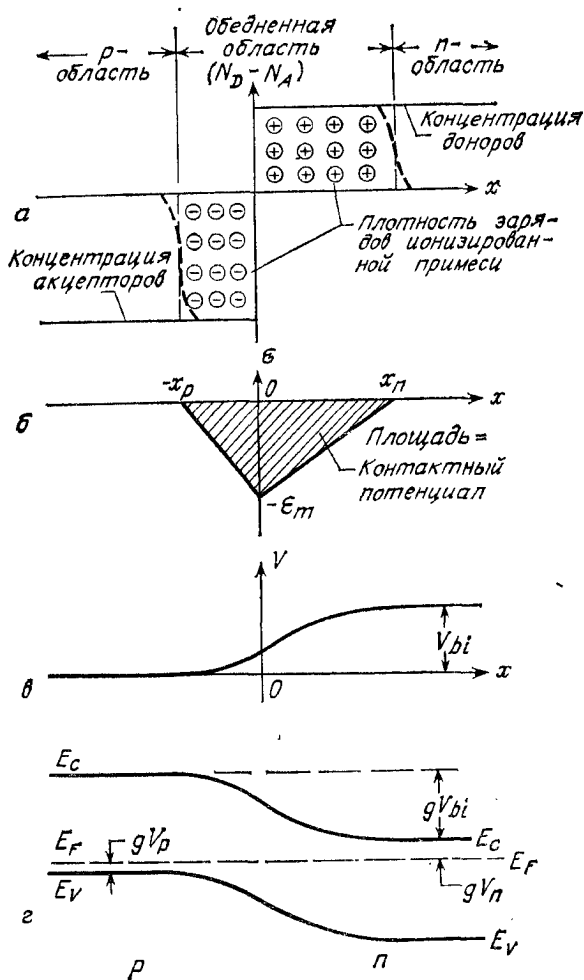


Рис. 10. Резкий переход при тепловом равновесии.

a — распределение пространственного заряда. Штриховыми линиями обозначены «хвосты» распределения основных носителей; *б* — распределение электрического поля; *в* — применение потенциала с расстоянием (V_{bi} — контактная разность потенциалов); *г* — энергетическая диаграмма.

или

$$\frac{\partial E_F}{\partial x} = 0. \quad (5a)$$

Аналогично

$$J_p = 0 = \mu_p p \frac{\partial E_F}{\partial x}. \quad (6)$$

Таким образом, равенство нулю электронного и дырочного токов ведет к постоянству уровня Ферми во всем объеме образца. Контактная разность потенциалов V_{bi} (рис. 10, б, з) определяется выражением

$$qV_{bi} = E_g - (qV_n + qV_p) = kT \ln \left(\frac{N_C N_V}{n_i^2} \right) - \left[kT \ln \left(\frac{N_C}{n_{n0}} \right) + kT \ln \left(\frac{N_V}{p_{p0}} \right) \right] = kT \ln \left(\frac{n_{n0} p_{p0}}{n_i^2} \right) \simeq kT \ln \left(\frac{N_A N_D}{n_i^2} \right). \quad (7)$$

Поскольку в равновесии $n_{n0} p_{p0} = n_p p_{p0} = n_i^2$,

$$V_{bi} = \frac{kT}{q} \ln \left(\frac{p_{p0}}{p_{n0}} \right) = \frac{kT}{q} \ln \left(\frac{n_{n0}}{p_{p0}} \right). \quad (7a)$$

Выражение (7a) связывает концентрации дырок и электронов по обе стороны перехода:

$$p_{n0} = p_{p0} \exp \left(- \frac{qV_{bi}}{kT} \right), \quad (8a)$$

$$n_{p0} = n_{n0} \exp \left(- \frac{qV_{bi}}{kT} \right). \quad (8б)$$

Приближенные значения V_{bi} для несимметричных резких $p-n$ -переходов в Ge, Si и GaAs приведены на рис. 11.

При тепловом равновесии электрическое поле в нейтральных областях полупроводника (достаточно далеко удаленных от перехода) равно нулю. Поэтому общий отрицательный заряд на единицу площади в p -области перехода в точности равен общему положительному заряду на единицу площади в n -области перехода:

$$N_A x_p = N_D x_n. \quad (9)$$

Из уравнения Пуассона получим (для резкого перехода)

$$- \frac{\partial^2 V}{\partial x^2} \equiv \frac{\partial \mathcal{E}}{\partial x} = \frac{\rho(x)}{\epsilon_s} = \frac{q}{\epsilon_s} [p(x) - n(x) + N_D^+(x) - N_A^-(x)], \quad (10)$$

или

$$- \frac{\partial^2 V}{\partial x^2} \approx \frac{q}{\epsilon_s} N_D \quad \text{для } 0 < x \leq x_n, \quad (10a)$$

$$\frac{\partial^2 V}{\partial x^2} \approx \frac{-q}{\epsilon_s} N_A \quad \text{для } -x_p \leq x < 0. \quad (10б)$$

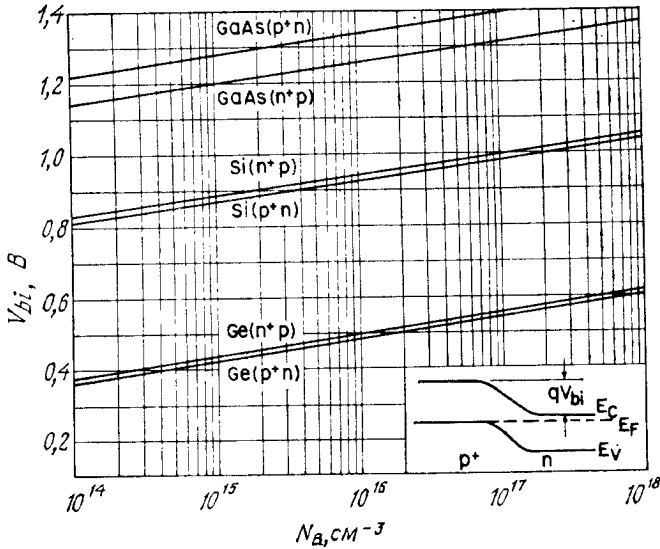


Рис. 11. Контактная разность потенциалов для несимметричных резких переходов в Ge, Si и GaAs как функция концентрации примеси в слаболегированной области перехода.

Интегрируя уравнения (10а) и (10б), определим величину электрического поля (рис. 10, б):

$$\mathcal{E}(x) = -\frac{qN_A(x+x_p)}{\epsilon_s} \text{ для } -x_p \leq x < 0, \quad (11a)$$

$$\mathcal{E}(x) = -\mathcal{E}_m + \frac{qN_D x}{\epsilon_s} = \frac{qN_D}{\epsilon_s}(x-x_n) \text{ для } 0 < x \leq x_n, \quad (11б)$$

где \mathcal{E}_m — максимальное значение электрического поля, которое достигается при $x = 0$ и равно

$$|\mathcal{E}_m| = \frac{qN_D x_n}{\epsilon_s} = \frac{qN_A x_p}{\epsilon_s}. \quad (12)$$

Еще раз интегрируя уравнение (10), получим распределение потенциала и диффузионный потенциал $V(x)$ (рис. 10, в):

$$V(x) = \mathcal{E}_m \left(x - \frac{x^2}{2W} \right), \quad (13)$$

$$V_{bi} = \frac{1}{2} \mathcal{E}_m W \equiv \frac{1}{2} \mathcal{E}_m (x_n + x_p), \quad (14)$$

где W — полная ширина обедненной области. Исключив \mathcal{E}_m из выражений (12) и (14), получим для резкого симметричного перехода

$$W = \sqrt{\frac{2\epsilon_s}{q} \left(\frac{N_A + N_D}{N_A N_D} \right) V_{bi}}. \quad (15)$$

Для несимметричного резкого перехода выражение (15) упрощается:

$$W = \sqrt{\frac{2\epsilon_s V_{bi}}{q N_B}}, \quad (15a)$$

где $N_B = N_D$ или $N_B = N_A$ в зависимости от того, выполняется условие $N_A \gg N_D$ или не выполняется.

Более точное выражение для ширины обедненной области получается из уравнения (10), если кроме концентрации ионизированной примеси дополнительно учесть вклад основных носителей, т. е. положить $p \approx -q [N_A - p(x)]$ в p -области перехода и $p \approx q [N_D - n(x)]$ в n -области перехода. Ширина обедненной области в таком приближении вычисляется по той же формуле (15), если V_{bi} заменить на $V_{bi} - 2kT/q$. Поправочный член $2kT/q$ появляется из-за наличия двух «хвостов» распределения основных носителей [26] (электронов в n -области перехода и дырок в p -области, как показано штриховой линией на рис. 10, а). Каждый из них вносит поправку kT/q . Сама поправка представляет собой просто дипольный момент несточности в распределении, равной разности между истинным распределением и распределением в резком переходе. Учитывая эти замечания, ширину обедненного слоя несимметричного резкого перехода при тепловом равновесии можно записать в виде

$$W = \sqrt{\frac{2\epsilon_s}{q N_B} (V_{bi} - 2kT/q) = L_D \sqrt{2(\beta V_{bi} - 2)}}, \quad (16)$$

где $\beta = q/kT$ и L_D — дебаевская длина. Дебаевская длина является характеристическим параметром для полупроводников и определяется выражением

$$L_D \equiv \sqrt{\frac{\epsilon_s kT}{q^2 N_B}} = \sqrt{\frac{\epsilon_s}{q N_B \beta}}. \quad (17)$$

При тепловом равновесии ширина обедненной области резкого перехода составляет $\sim 6L_D$ для Ge, $\sim 8L_D$ для Si и $\sim 10L_D$ для GaAs. Зависимость дебаевской длины от концентрации легирующей примеси для кремния при комнатной температуре приведена на рис. 12. При концентрации примеси 10^{16} см $^{-3}$ дебаевская длина составляет 400 Å и изменяется по закону $1/\sqrt{N_B}$, т. е. с возрастанием концентрации на порядок она уменьшается в 3,16 раза.

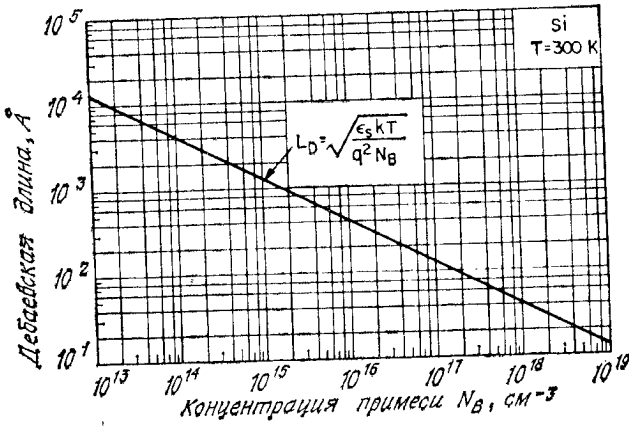


Рис. 12. Зависимость дебаевской длины в Si от концентрации примеси.

Зависимость величины W от концентрации примеси для резкого несимметричного перехода в кремнии приведена на рис. 13 (штриховая линия соответствует нулевому смещению на переходе). При подаче на переход напряжения V полное изменение электростатического потенциала на нем составит $V_{bi} + V$ для обратного смещения (положительное напряжение на n -области по отношению к p -области) и $V_{bi} - V$ для прямого смещения. Подставив эти значения напряжения в выражение (16), получим зависимость ширины обедненной области от приложенного напряжения.

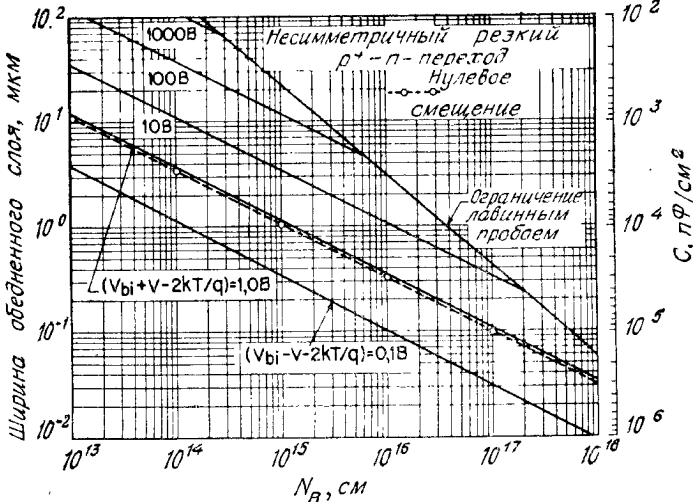


Рис. 13. Зависимость ширины обедненного слоя и удельной емкости от концентрации примеси для несимметричного резкого перехода в Si.

На рис. 13 приведены значения W для несимметричных резких переходов в кремнии. Значения, лежащие ниже линии нулевого смещения (штриховая линия), относятся к прямому смещению, а значения выше этой линии относятся к обратному смещению.

Приведенные результаты справедливы также и для GaAs, так как Si и GaAs имеют примерно одинаковые значения низкочастотной диэлектрической постоянной. Чтобы получить значение ширины обедненной области для германия, достаточно умножить величину, полученную для кремния, на коэффициент $\sqrt{\epsilon_s(\text{Ge})/\epsilon_s(\text{Si})} = 1,16$.

Рассмотренная выше простая модель дает достаточно точные оценки для большинства резких $p-n$ -переходов. Однако в случае сильнонесимметричных переходов или приборов со сверхмелким залеганием переходов для получения точных результатов необходимо выполнять численные расчеты [27]. Электрическое поле вблизи перехода не может быть ограничено областью, заштрихованной на рис. 8, а в левой части, поскольку любой градиент примеси дает свой вклад в поле (в соответствии с выражением (5)). Большой градиент легирующей примеси, очевидно, приведет к распространению электрического поля за пределы отмеченной области. На рис. 14, в в качестве примера показан диффузионный $p-n$ -переход глубиной $0,25 \text{ мкм}$ с $N_s = 2 \cdot 10^{20} \text{ см}^{-3}$, с профилем типа erfc и $N_A = 5 \cdot 10^{15} \text{ см}^{-3}$. Зонная диаграмма в состоянии равновесия приведена на рис. 14, а. Имеется несколько существенных различий в профиле электрического поля, полученном с помощью простой модели и численных расчетов. Во-первых, протяженность области с электрическим полем в диффузионной части перехода в пять раз больше, чем следует из простой модели. Во-вторых, электрическое поле не снижается ниже 10^4 В/см вплоть до границы области, поэтому процессы переноса носителей будут сильно ускорены. Распределение пространственного заряда (без учета поверхностных эффектов) приведено на рис. 14, в. Очевидно, что пространственный заряд в диффузионной области распространяется на значительно большее расстояние, чем ширина обедненной области — x_p , полученная в рамках простой модели.

Барьерная емкость. Удельная барьерная емкость $p-n$ -перехода определяется выражением $C \equiv dQ_c/dV$, где dQ_c — дифференциальное приращение плотности заряда, вызванное достаточно малым изменением приложенного напряжения.

Для несимметричных резких переходов удельная емкость

$$C \equiv \frac{dQ_c}{dV} = \frac{d(qN_B W)}{d[(qN_B/2\epsilon_s) W^2]} = \frac{\epsilon_s}{W} = \sqrt{\frac{q\epsilon_s N_B}{2}} \times \\ \times (V_{bi} \pm V - 2kT/q)^{-1/2} = \frac{\epsilon_s}{\sqrt{2}L_D} (\beta V_{bi} \pm \beta V - 2)^{-1/2} [\Phi/\text{см}^2], \quad (18)$$

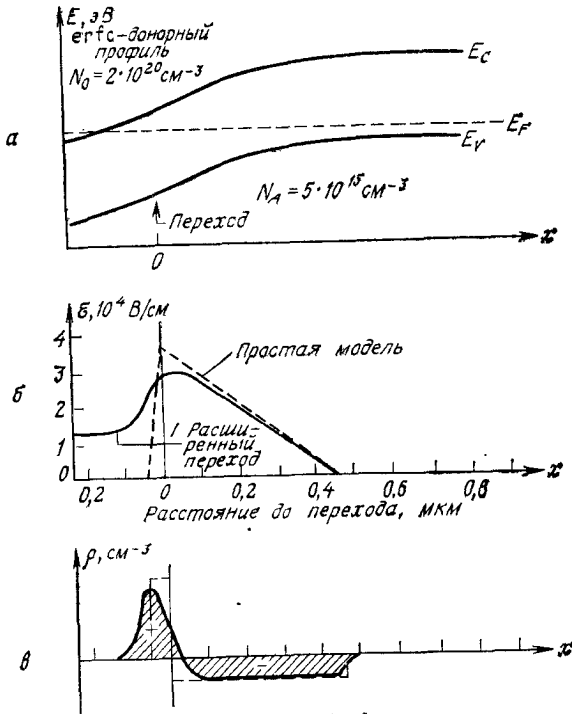


Рис. 14. Сильнонесимметричный переход [27].

a — зонная диаграмма; *b* — распределение электрического поля; *v* — распределение пространственного заряда.

или

$$\frac{1}{C^2} = \frac{2L_D^2}{e_s^2} (\beta V_{bi} \pm \beta V - 2), \tag{18a}$$

$$\frac{d(1/C^2)}{dV} \cong \frac{2L_D^2 \beta}{e_s^2} = \frac{2}{q \epsilon_s N_B}, \tag{18б}$$

где знаки плюс и минус соответствуют обратному и прямому смещению. Из выражения (18а) следует, что для несимметричного резкого перехода зависимость $1/C^2$ от V представляет собой прямую линию. Ее наклон определяет концентрацию примеси в подложке (N_B), а точка пересечения с осью абсцисс (при $1/C^2 = 0$) дает величину $V_{bi} - 2kT/q$. Результаты вычисления емкостей приведены на рис. 13. Отметим, что при прямом смещении к барьерной емкости добавляется диффузионная емкость (разд. 3.4).

Формула (18б) остается справедливой и для переходов с более сложными распределениями примеси, чем для резкого $p^+ - n$ -перехода. В общем случае имеем

$$\frac{d(1/C^2)}{dV} = \frac{2}{q\varepsilon_s N(W)} \text{ и } W = \frac{\varepsilon_s}{C(V)}, \quad (18\text{в}), (18\text{г})$$

где $N(W)$ — плотность примеси при $x = W$.

Следует отметить, что зависимость емкости от напряжения нечувствительна к изменениям профиля примеси в высоколегированной области, если они имеют место на расстояниях, меньших дебаевской длины. Поэтому при определении распределения примесей $C-V$ -методом обеспечивается пространственное разрешение порядка дебаевской длины [28].

2.3.2. Плавный линейный переход

Вначале рассмотрим состояние теплового равновесия. На рис. 15, а показан случай линейного распределения примеси в плавном переходе. Для такой ситуации уравнение Пуассона имеет вид

$$-\frac{\partial^2 V}{\partial x^2} = \frac{\partial \mathcal{E}}{\partial x} = \frac{\rho(x)}{\varepsilon_s} = \frac{q}{\varepsilon_s} (p - n) \approx \frac{q}{\varepsilon_s} ax, \quad -\frac{W}{2} \leq x \leq \frac{W}{2}, \quad (19)$$

где a — градиент концентрации примеси размерностью см^{-4} . Интегрируя уравнение (19), получим распределение напряженности электрического поля (рис. 15, б)

$$\mathcal{E}(x) = -\frac{qa}{\varepsilon_s} \frac{(W/2)^2 - x^2}{2}. \quad (20)$$

В точке $x = 0$ поле принимает максимальное значение \mathcal{E}_m :

$$|\mathcal{E}_m| = \frac{qaW^2}{8\varepsilon_s}. \quad (20\text{а})$$

Повторное интегрирование уравнения (20) дает величину контактной разности потенциалов (рис. 15, в)

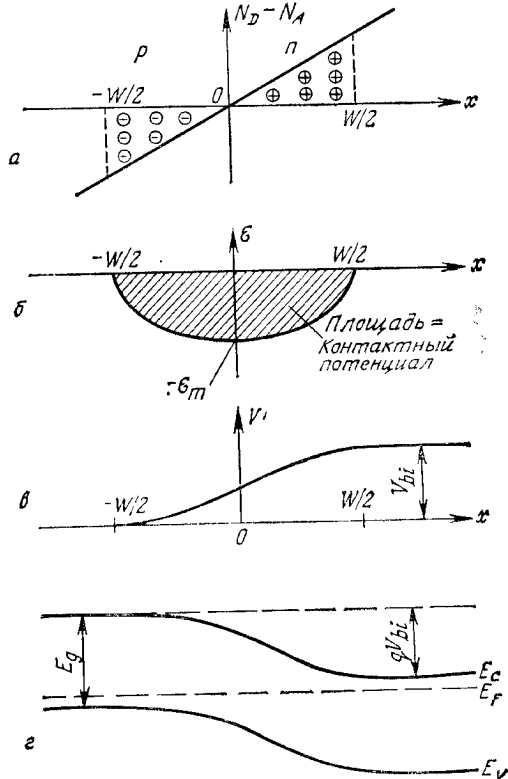
$$V_{bi} = \frac{qaW^3}{12\varepsilon_s} \quad (21)$$

или

$$W = \left(\frac{12\varepsilon_s V_{bi}}{qa} \right)^{1/3}. \quad (21\text{а})$$

Так как значения концентрации примесей на границах обедненной области ($-W/2$ и $W/2$) равны $aW/2$, то для контактной раз-

Рис. 15. Плавный линейный переход в тепловом равновесии
 а — распределение пространственного заряда; б — распределение электрического поля; в — изменение потенциала с расстоянием;
 г — зонная диаграмма.



ности потенциалов на линейном переходе справедливо приближенное выражение, аналогичное выражению (7):

$$V_{bi} \approx \frac{kT}{q} \ln \left[\frac{(aW/2)(aW/2)}{n_i^2} \right] = \frac{kT}{q} \ln \left(\frac{aW}{2n_i} \right)^2. \quad (22)$$

Барьерная емкость линейного перехода вычисляется по формуле

$$C \equiv \frac{dQ_c}{dV} = \frac{d(qaW^2/8)}{d(qaW^3/12e_s)} = \frac{e_s}{W} = \left[\frac{qa^2e_s^2}{12(V_{bi} \pm V)} \right]^{1/3} [\Phi/\text{см}^2], \quad (23)$$

где знаки плюс и минус соответствуют обратному и прямому смещению. Более точным расчетом [29] с использованием численных методов получают выражение, аналогичное выражению (23), в котором вместо величины V_{bi} нужно подставить «градиентное напряжение» V_g :

$$V_g = \frac{2}{3} \frac{kT}{q} \ln \left[\frac{a^2e_s kT/q}{8qn_i^2} \right]. \quad (24)$$

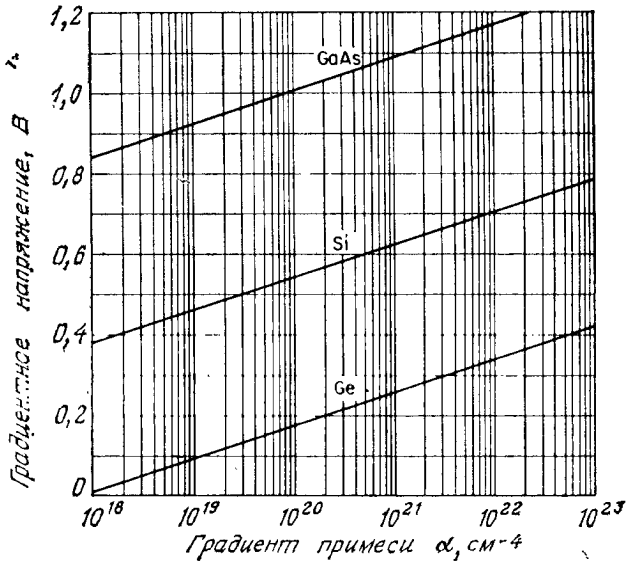


Рис. 16. Градиентное напряжение для линейного перехода в Ge, Si и GaAs.

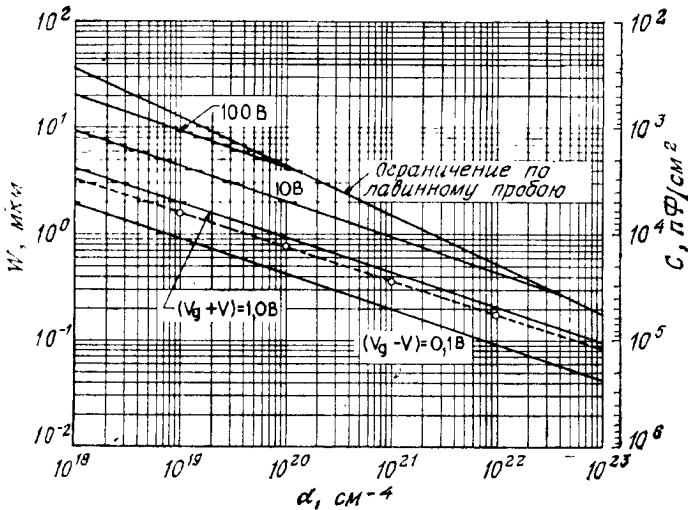


Рис. 17. Зависимости ширины обедненного слоя и удельной барьерной емкости от градиента концентрации примеси для линейных переходов в Si.

Зависимость градиентного напряжения от градиента концентрации примеси в Ge, Si и GaAs приведена на рис. 16. Это напряжение меньше вычисленного по формуле (22) более чем на 100 мВ. Зависимости ширины обедненной области и соответствующей барьерной емкости от градиента концентрации примеси для кремния приведены на рис. 17 (штриховая линия соответствует нулевому смещению на переходе).

2.4. ВОЛЬТ-АМПЕРНЫЕ ХАРАКТЕРИСТИКИ

2.4.1. Идеальный случай. Формула Шокли [1]

Идеальные вольт-амперные характеристики вычисляются на основе следующих четырех допущений: 1) приближения обедненного слоя с резкими границами, т. е. контактная разность потенциалов и приложенное напряжение уравновешены двойным заряженным слоем с резкими границами, вне которых полупроводник считается нейтральным; 2) приближения Больцмана, т. е. в обедненной области справедливы распределения Больцмана, аналогичные выражениям (33) и (37) из гл. 1; 3) приближения низкого уровня инжекции, т. е. плотность инжектированных неосновных носителей мала по сравнению с концентрацией основных носителей; 4) отсутствия в обедненном слое токов генерации и постоянства протекающих через него электронного и дырочного токов.

Вначале рассмотрим соотношения Больцмана для состояния теплового равновесия

$$n = n_i \exp\left(\frac{E_F - E_i}{kT}\right) \equiv n_i \exp\left[\frac{q(\psi - \varphi)}{kT}\right], \quad (25a)$$

$$p = n_i \exp\left(\frac{E_i - E_F}{kT}\right) \equiv n_i \exp\left[\frac{q(\varphi - \psi)}{kT}\right], \quad (25b)$$

где ψ и φ — потенциалы, соответствующие середине запрещенной зоны и уровню Ферми ($\psi \equiv -E_i/q$, $\varphi \equiv -E_F/q$). Очевидно, что поскольку в состоянии теплового равновесия справедливы выражения (25a) и (25b), то произведение pn равно n_i^2 . При подаче на переход напряжения по обеим его сторонам происходит изменение концентрации неосновных носителей и произведение pn уже не равно n_i^2 . Определим теперь квазиуровни Ферми с помощью следующих выражений:

$$n \equiv n_i \exp\left[\frac{q(\psi - \varphi_n)}{kT}\right], \quad (26a)$$

$$p \equiv n_i \exp\left[\frac{q(\varphi_p - \psi)}{kT}\right], \quad (26b)$$

где φ_n и φ_p — квазиуровни Ферми для электронов и дырок соответственно. Из выражений (26а) и (26б) имеем

$$\varphi_n \equiv \psi - \frac{kT}{q} \ln \left(\frac{n}{n_i} \right), \quad (27a)$$

$$\varphi_p \equiv \psi + \frac{kT}{q} \ln \left(\frac{p}{n_i} \right). \quad (27б)$$

Тогда

$$pn = n_i^2 \exp \left[\frac{q(\varphi_p - \varphi_n)}{kT} \right]. \quad (28)$$

При прямом смещении $(\varphi_p - \varphi_n) \gg 0$ и $pn > n_i^2$, а при обратном смещении $(\varphi_p - \varphi_n) < 0$ и $pn < n_i^2$.

Из выражения (93) гл. 1 и выражения (26а) с учетом того, что $\mathcal{E} \equiv -\nabla\psi$, получим

$$J_n = q\mu_n \left(n\mathcal{E} + \frac{kT}{q} \nabla n \right) = q\mu_n n (-\nabla\psi) + \\ + q\mu_n \frac{kT}{q} \left[\frac{qn}{kT} (\nabla\psi - \nabla\varphi_n) \right] = -q\mu_n n \nabla\varphi_n. \quad (29)$$

Аналогично для дырочного тока имеем

$$J_p = -q\mu_p p \nabla\varphi_p. \quad (30)$$

Таким образом, плотности электронного тока и дырочного тока пропорциональны градиентам квазиуровней Ферми для электронов и дырок соответственно. Если $\varphi_n = \varphi_p = \varphi = \text{const}$ в состоянии теплового равновесия, то $J_n = J_p = 0$.

Идеализированные функции распределения потенциала и концентрации носителей в p - n -переходе при прямом и обратном смещении приведены на рис. 18. Изменения φ_n и φ_p с расстоянием определяются распределением концентрации носителей в соответствии с выражениями (27). В n - и p -областях перехода разность концентраций электронов n составляет несколько порядков, а плотность электронного тока J_n почти не меняется. Поэтому в пределах обедненного слоя потенциал φ_n также должен быть почти постоянен. Разность электростатических потенциалов на p - n -переходе определяется величиной

$$V = \varphi_p - \varphi_n. \quad (31)$$

Из выражений (28) и (31) получаем концентрацию электронов на границе обедненного слоя в p -области перехода ($x = -x_p$):

$$n_p = \frac{n_i^2}{p_p} \exp \left(\frac{qV}{kT} \right) = n_{p0} \exp \left(\frac{qV}{kT} \right), \quad (32)$$

где n_{p0} — равновесная концентрация электронов в p -области. Аналогично

$$p_n = p_{n0} \exp \left(\frac{qV}{kT} \right), \quad (33)$$

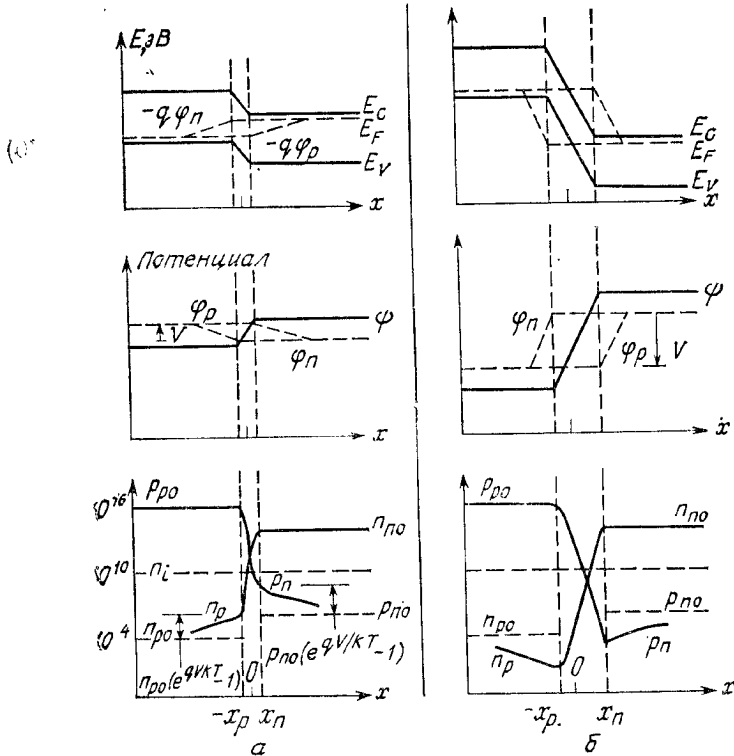


Рис. 18. Зонная диаграмма, положение собственного уровня Ферми ψ и квазиуровней Ферми для электронов ψ_n и дырок ψ_p и распределение концентрации носителей [1].

а — при прямом смещении; б — при обратном смещении.

где p_n — концентрация дырок на границе обедненного слоя в n -области перехода при $x = x_n$, а p_{n0} — равновесная концентрация дырок в n -области. Полученные выражения являются основными граничными условиями при вычислении идеальной вольт-амперной характеристики.

В стационарном состоянии уравнения непрерывности записываются в виде

$$-U + \mu_n \mathcal{E} \frac{\partial n_n}{\partial x} + \mu_n n_n \frac{\partial \mathcal{E}}{\partial x} + D_n \frac{\partial^2 n_n}{\partial x^2} = 0, \quad (34a)$$

$$-U - \mu_p \mathcal{E} \frac{\partial p_n}{\partial x} - \mu_p p_n \frac{\partial \mathcal{E}}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} = 0. \quad (34б)$$

В этих уравнениях U обозначает результирующую скорость рекомбинации. Поскольку в первом приближении соблюдается за-

рядовая нейтральность, то $(n_n - n_{n0}) \approx (p_n - p_{n0})$. Умножая уравнение (34а) на $\mu_p p_n$ и уравнение (34б) на $\mu_n n_n$ и учитывая соотношение Эйнштейна $D = (kT/q) \mu$, получим

$$-\frac{p_n - p_{n0}}{\tau_a} + D_a \frac{\partial^2 p_n}{\partial x^2} - \frac{n_n - p_n}{n_n/\mu_p + p_n/\mu_n} \frac{\mathcal{E} \partial p_n}{\partial x} = 0, \quad (35)$$

где $D_a = \frac{n_n + p_n}{n_n/D_p + p_n/D_n}$ — коэффициент абиполярной диффузии, (36)

$$\tau_a = \frac{p_n - p_{n0}}{U} = \frac{n_n - n_{n0}}{U} \text{ — амбиполярное время жизни.} \quad (37)$$

В случае малого уровня инжекции (т. е. при $p_n \ll n_n \approx n_{n0}$ в полупроводнике n -типа) уравнение (35) упрощается:

$$-\frac{p_n - p_{n0}}{\tau_p} - \mu_p \mathcal{E} \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} = 0. \quad (38)$$

В уравнении (38) (в отличие от уравнения (34б)) отсутствует член $\mu_p p_n \partial \mathcal{E} / \partial x$. При низком уровне инжекции этот член того же порядка малости, что и члены, которыми мы пренебрегли.

В нейтральной области, где отсутствует электрическое поле, уравнение (38) существенно упрощается и принимает вид

$$\frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{D_p \tau_p} = 0. \quad (39)$$

Решение уравнения (39) с граничными условиями, задаваемыми выражением (33), и при условии $p_n(x = \infty) = p_{n0}$ имеет вид

$$p_n - p_{n0} = p_{n0} (e^{qV/kT} - 1) e^{-(x-x_n)/L_p}, \quad (40)$$

$$L_p \equiv \sqrt{D_p \tau_p}. \quad (41)$$

В результате при $x = x_n$ плотность дырочного тока равна

$$J_p = -qD_p \left. \frac{\partial p_n}{\partial x} \right|_{x_n} = \frac{qD_p p_{n0}}{L_p} (e^{qV/kT} - 1). \quad (42)$$

Аналогично, рассматривая p -область, получим плотность электронного тока

$$J_n = qD_n \left. \frac{\partial n_p}{\partial x} \right|_{-x_p} = \frac{qD_n n_{p0}}{L_n} (e^{qV/kT} - 1). \quad (43)$$

Распределение концентрации неосновных носителей и плотности тока для прямого и обратного смещения на p — n -переходе приведено на рис. 19.

Общий ток через переход равен сумме токов выражений (42) и (43):

$$J = J_p + J_n = J_s (e^{qV/kT} - 1), \quad (44)$$

$$J_s = \frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n}. \quad (45)$$

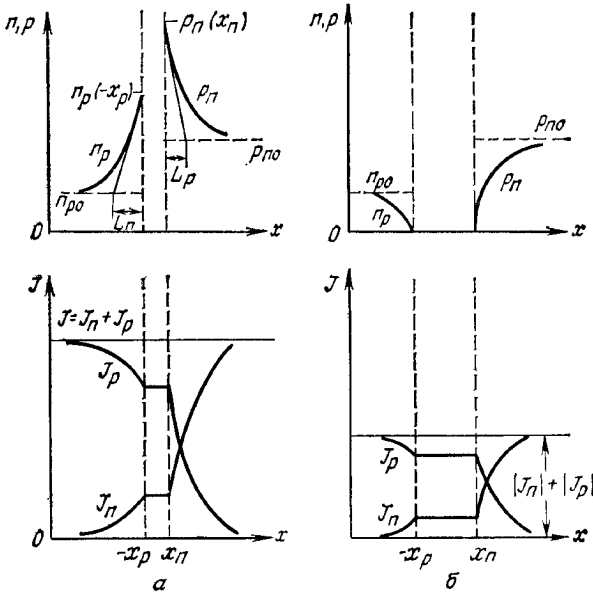


Рис. 19. Распределение концентрации носителей и плотности тока (линейный масштаб) [1].
 а — при прямом смещении; б — при обратном смещении.

Выражение (45) представляет собой известную формулу Шокли [1], описывающую вольт-амперную характеристику идеального диода (рис. 20). При прямом смещении (при подаче на p -область положительного напряжения) и при $V > 3kT/q$ наклоном характеристики постояен (рис. 20, б); при температуре 300 °К для изменения тока на порядок требуется изменить напряжение на 59,5 мВ ($=2,3 kT/q$). При обратном смещении плотность тока насыщается и ограничена величиной $-J_s$.

Обсудим теперь кратко влияние температуры на плотность тока насыщения J_s . Ограничимся рассмотрением роли первого слагаемого в формуле (45), так как роль второго слагаемого аналогична. В случае несимметричного резкого p^+-n -перехода (с концентрацией доноров в n -области N_D) $p_{n0} \gg n_{p0}$ и вторым членом можно [вообще пренебречь. Все величины D_p , p_{n0} и L_p ($\equiv \sqrt{D_p \tau_p}$) зависят от температуры. Если отношение D_p/τ_p пропорционально T^γ , где γ — постоянная, то

$$\begin{aligned}
 J_s &\simeq \frac{qD_p p_{n0}}{L_p} \simeq q \sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} \sim \left[T^3 \exp\left(-\frac{E_g}{kT}\right) \right] T^{\gamma/2} = \\
 &= T^{(3+\gamma/2)} \exp\left(-\frac{E_g}{kT}\right). \quad (46)
 \end{aligned}$$

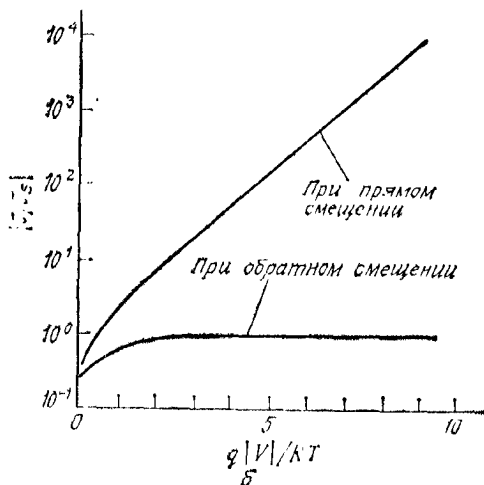
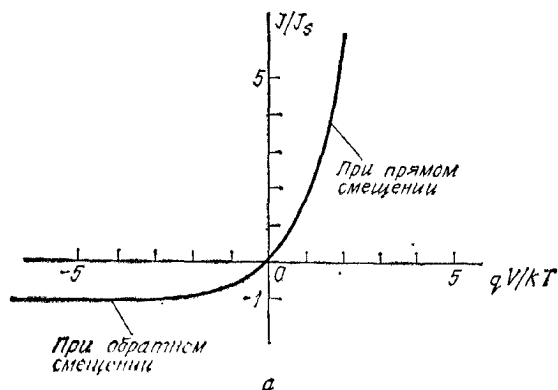


Рис. 20. Вольт-амперные характеристики идеального $p-n$ -перехода.
 а — линейный масштаб; б — полулогарифмический масштаб.

Температурная зависимость степенного множителя $T^{(3+v/2)}$ гораздо слабее, чем экспоненциального. Наклон кривой, описывающей зависимость J_s от $1/T$, определяется шириной запрещенной зоны E_g . Можно полагать, что при обратном смещении (когда $|J_R| \sim J_s$) ток будет расти с температурой приблизительно по экспоненте $e^{-E_g/kT}$, а при прямом смещении (когда $J_F \sim J_s e^{qV/kT}$) ток будет расти приблизительно по закону $\exp [-(E_g - qV)/kT]$.

2.4.2. Процессы генерации-рекомбинации носителей [2]

Формула Шокли удовлетворительно описывает вольт-амперные характеристики германиевых $p-n$ -переходов при низких плотностях токов. Однако для $p-n$ -переходов в Si и GaAs эта формула дает лишь качественное согласие с реальными характеристиками. Основными причинами отклонения характеристики от идеальной являются: 1) влияние поверхности; 2) генерация и рекомбинация носителей в обедненном слое; 3) туннелирование носителей между состояниями в запрещенной зоне; 4) высокий уровень инжекции, наблюдаемый даже при относительно небольшом прямом смещении; 5) влияние последовательного сопротивления. Кроме того, под действием достаточно большого электрического поля при обратном смещении возникает пробой перехода (например, в результате лавинного умножения) (разд. 2.5).

Поверхность оказывает влияние на $p-n$ -переход в основном за счет ионных зарядов на ней или вблизи нее, которые индуцируют заряды в полупроводнике. Это приводит к образованию так называемых поверхностных каналов или поверхностных обедненных слоев. Наличие канала влияет на обедненную область $p-n$ -перехода и вызывает возрастание поверхностных токов утечки. Более подробно поверхностные эффекты рассмотрены в гл. 7 и 8. В кремниевых планарных $p-n$ -переходах поверхностные токи утечки обычно намного меньше генерационного тока в обедненной области.

Сначала рассмотрим генерационный ток при обратном смещении. В этом случае из-за уменьшения концентрации носителей при обратном смещении ($pn \ll n_i^2$), согласно рассмотренным в гл. 1 явлениям генерации—рекомбинации, преобладающим будет процесс эмиссии. Скорость генерации электроно-дырочных пар, согласно уравнению (58) гл. 1, при условиях $p \ll n_i$ и $n \ll n_i$ составляет

$$U = - \left[\frac{\sigma_p \sigma_n v_{th} N_t}{\sigma_n \exp\left(\frac{E_t - E_i}{kT}\right) + \sigma_p \exp\left(\frac{E_i - E_t}{kT}\right)} \right] n_i \equiv \frac{n_i}{\tau_e}, \quad (47)$$

где τ_e — эффективное время жизни носителей, равное обратной величине выражения в квадратных скобках. Плотность тока, обусловленного генерацией в обедненной области, принимается равной

$$J_{\text{gen}} = \int_0^W q|U| dx \simeq q|U|W = \frac{qn_i W}{\tau_e}. \quad (48)$$

где W — ширина обедненного слоя. Если эффективное время жизни слабо изменяется с температурой, то генерационный ток будет иметь ту же температурную зависимость, что и n_i . При за-

данной температуре $J_{\text{ген}}$ пропорционален ширине обедненного слоя, которая в свою очередь зависит от приложенного обратного смещения. Таким образом, следует ожидать, что для резкого перехода

$$J_{\text{ген}} \sim (V_{bi} + V)^{1/2}, \quad (49a)$$

а для плавного перехода

$$J_{\text{ген}} \sim (V_{bi} + V)^{1/3}. \quad (49b)$$

Полный обратный ток (при $p_{n0} \gg n_{p0}$ и $|V| > 3kT/q$) можно приближенно представить суммой диффузионного тока в нейтральной области и генерационного тока в обедненной области:

$$J_R = q \sqrt{\frac{D_p}{\tau_p} \frac{n_i^2}{N_D} + \frac{qn_i W}{\tau_e}}. \quad (50)$$

В полупроводниках с большим значением n_i (таких, как Ge) при комнатной температуре преобладает диффузионный ток и обратный ток подчиняется закону Шокли. Если же n_i мало (что имеет место, например, в Si), то может преобладать генерационный ток. Типичные результаты для кремния [3] приведены на рис. 21 (кривая δ).

При прямом смещении, когда в генерационно-рекомбинационных явлениях в обедненном слое определяющими становятся процессы захвата носителей, к диффузионному току добавляется рекомбинационный ток J_{rec} . Подставляя выражение (28) в формулу (58) из гл. 1, получим

$$U = \frac{\sigma_p \sigma_n v_{th} N_i n_i^2 (e^{qV/kT} - 1)}{\sigma_n \left[n + n_i \exp\left(\frac{E_t - E_i}{kT}\right) \right] + \sigma_p \left[p + n_i \exp\left(\frac{E_i - E_t}{kT}\right) \right]}. \quad (51)$$

При выполнении условий $E_t = E_i$ и $\sigma_n = \sigma_p = \sigma$ выражение (51) упрощается и принимает вид

$$U = \frac{\sigma v_{th} N_i n_i^2 (e^{qV/kT} - 1)}{n + p + 2n_i} = \frac{\sigma v_{th} N_i n_i^2 (e^{qV/kT} - 1)}{n_i \left\{ \exp\left[\frac{q(\Psi - \Phi_n)}{kT}\right] + \exp\left[\frac{q(\Phi_p - \Psi)}{kT}\right] + 2 \right\}}. \quad (52)$$

Величина U достигает максимального значения в той точке обедненной области, где $\Psi = (\Phi_n + \Phi_p)/2$, при этом знаменатель в формуле (52) принимает вид $2n_i [\exp(qV/2kT) + 1]$. В результате при $V \geq kT/q$ имеем

$$U \simeq \frac{1}{2} \sigma v_{th} N_i n_i \exp\left(\frac{qV}{2kT}\right) \quad (53)$$

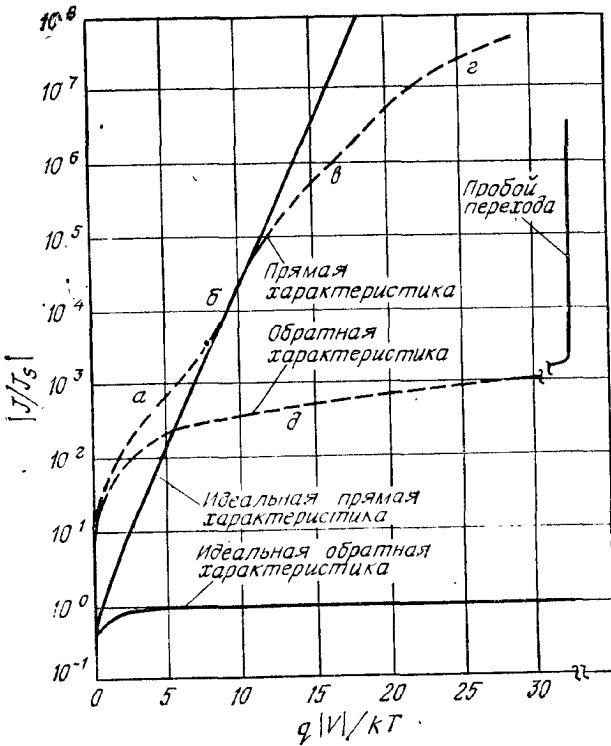


Рис. 21. Вольт-амперные характеристики кремниевого диода [3].

а — преобладание генерационно-рекомбинационного тока; б — преобладание диффузионного тока; в — высокий уровень инжекции; г — влияние последовательного сопротивления; д — обратный ток утечки.

и

$$J_{\text{ген}} = \int_0^W qU dx \approx \frac{qW}{2} \sigma v_{th} N_t n_i \exp\left(\frac{qV}{2kT}\right) \sim n_i N_t. \quad (54)$$

Рекомбинационный ток при прямом смещении, подобно генерационному току при обратном смещении, пропорционален n_i . Полный прямой ток приближенно равен сумме токов, определяемых формулами (44) и (54), и при $p_{n0} \gg n_{p0}$ и $V \gg kT/q^1$

$$J_F = q \sqrt{\frac{D_p}{\tau_p}} \frac{n_i^2}{N_D} \exp\left(\frac{qV}{kT}\right) + \frac{qW}{2} \sigma v_{th} N_t n_i \exp\left(\frac{qV}{2kT}\right). \quad (55)$$

¹ Строго говоря, в выражениях (54) и (55) используется не ширина обедненного слоя W , а меньшая величина $W \frac{kT}{e(V_{bi} - V)}$, представляющая собой толщину слоя, в котором эффективно происходят процессы рекомбинации электронов и дырок. — Прим. ред.

Экспериментальные результаты в общем случае можно описать следующим выражением:

$$J_F \sim \exp\left(\frac{qV}{nkT}\right), \quad (56)$$

где коэффициент $n = 2$, если преобладает рекомбинационный ток (рис. 21, кривая *a*), и $n = 1$, если преобладает диффузионный ток (рис. 21, кривая *b*). Если оба тока сравнимы по величине, то n лежит между 1 и 2.

2.4.3. Высокий уровень инжекции

При прямом смещении в условиях высокой плотности токов, когда плотность инжектированных неосновных носителей сравнима с концентрацией основных носителей, необходимо учитывать дрейфовую и диффузионную составляющие тока. Для плотностей токов проводимости справедливы выражения

$$J_p = -q\mu_p p \nabla \varphi_p, \quad J_n = -q\mu_n n \nabla \varphi_n.$$

Поскольку величины J_p , q , μ_p и p положительны, то квазиуровень Ферми для дырок монотонно уменьшается с координатой (т. е. слева направо на рис. 18, *a*). Аналогично ведет себя квазиуровень Ферми для электронов, монотонно возрастая справа налево. Следовательно, всюду в переходе расстояние между квазиуровнями Ферми должно быть меньше или равно приложенному напряжению, и даже в случае высокого уровня инжекции справедливо соотношение

$$pn \leq n_i^2 \exp\left(\frac{qV}{kT}\right). \quad (57)$$

Следует отметить, что приведенные выше рассуждения не зависят от наличия рекомбинации в обедненной области. Пока где-либо присутствует рекомбинация, продолжается протекание тока.

На рис. 22 приведены расчетные зависимости уровня Ферми в собственном полупроводнике ψ , квазиуровней Ферми φ_n и φ_p и концентрации носителей в кремниевом резком $p-n$ -переходе со следующими параметрами: $N_A = 10^{18} \text{ см}^{-3}$, $N_D = 10^{16} \text{ см}^{-3}$, $\tau_n = 3 \cdot 10^{-10} \text{ с}$, $\tau_p = 8,4 \cdot 10^{-10} \text{ с}$. При плотности тока 10 А/см^2 диод работает в режиме низкого уровня инжекции. Почти все приложенное напряжение падает на $p-n$ -переходе. Концентрация дырок в n -области мала по сравнению с концентрацией электронов. При плотности тока 10^5 А/см^2 концентрация электронов в n -области вблизи перехода существенно превышает концентрацию доноров и возникает падение напряжения на сопротивлении n -области. Плотность тока, равная 10^4 А/см^2 , приводит к очень высокому уровню инжекции; падение напряжения на переходе становится незначительным по сравнению с омическим падением

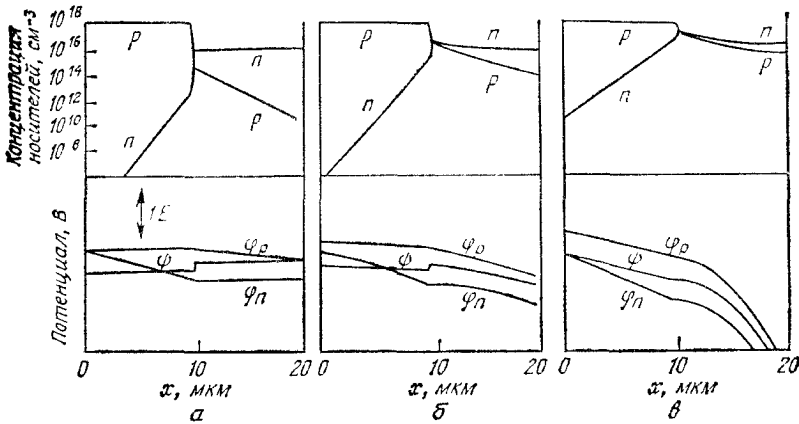


Рис. 22. Распределение концентрации носителей, положение собственного уровня Ферми ψ и квазиуровней Ферми для $p-n$ -перехода в Si при различных плотностях тока [30].

$a - 10 \text{ A/cm}^2$; $б - 10^3 \text{ A/cm}^2$; $в - 10^4 \text{ A/cm}^2$.

напряжения в обеих областях диода. Хотя на рис. 22 изображена только центральная часть диода, очевидно, что при любом прямом смещении расстояние между квазиуровнями Ферми перехода меньше или равно разности между квазиуровнями Ферми для электронов справа от перехода.

Из рис. 22, б и в следует, что концентрации носителей на границе перехода в n -области приблизительно равны ($n \approx p$). Подставляя это условие в формулу (57), получим $p_n(x = x_n) \approx n_i \exp(qV/2kT)$. Следовательно, зависимость тока от приложенного напряжения пропорциональна $\exp(qV/2kT)$, как показано на рис. 21 (кривая в).

При высоком уровне инжекции необходимо учитывать также эффект, обусловленный конечной величиной сопротивления квазинейтральных областей вблизи перехода. На этом сопротивлении падает большая часть напряжения, приложенного к диоду (рис. 21, кривая г). Влияние последовательного сопротивления существенно уменьшается при использовании материалов с эпитаксиальными слоями.

В.4.4. Диффузионная емкость

Барьерная емкость, рассмотренная выше, при обратном смещении вносит основной вклад в общую емкость перехода. При прямом смещении в емкости перехода преобладает диффузионная емкость, обусловленная изменением распределения концентрации неосновных носителей. Если на переход, смещенный в прямом

направлении, подать небольшой сигнал переменного тока, то общее напряжение и плотность тока можно описать выражениями

$$V(t) = V_0 + V_1 e^{j\omega t}, \quad J(t) = J_0 + J_1 e^{j\omega t}, \quad (58)$$

где V_0 и J_0 — постоянные напряжение и плотность тока, а V_1 и J_1 — амплитуды напряжения и плотности тока входного сигнала. Концентрации электронов и дырок на границах обедненной области можно получить из выражений (32) и (33), если вместо V подставить $(V_0 + V_1 e^{j\omega t})$. Переменная составляющая концентрации дырок, соответствующая малому входному сигналу, определяется соотношением

$$\tilde{p}_n(x, t) = p_{n1}(x) e^{j\omega t}. \quad (59)$$

Ограничиваясь случаем $V_1 \ll V_0$, получим

$$\begin{aligned} p_n &= p_{n0} \exp \left[\frac{q(V_0 + V_1 e^{j\omega t})}{kT} \right] \simeq \\ &\simeq p_{n0} \exp \left(\frac{qV_0}{kT} \right) + \frac{p_{n0} q V_1}{kT} \exp \left(\frac{qV_0}{kT} \right) e^{j\omega t}. \end{aligned} \quad (60)$$

Аналогичное выражение получается и для концентрации электронов. Первое слагаемое в выражении (60) описывает постоянную составляющую, а второе слагаемое — переменную составляющую на границе обедненного слоя $|p_{n1}(x_n) e^{j\omega t}|$, обусловленную входным сигналом. Подставляя p_n в уравнение непрерывности (уравнение (97) из гл. 1 при $G_p = 0$), имеем

$$j\omega \tilde{p}_n = -\frac{\tilde{p}_n}{\tau_p} + D_p \frac{\partial^2 \tilde{p}_n}{\partial x^2}$$

или

$$\frac{\partial^2 \tilde{p}_n}{\partial x^2} - \frac{\tilde{p}_n}{D_p \tau_p (1 + j\omega \tau_p)} = 0. \quad (61)$$

Уравнение (61) тождественно уравнению (39), если время жизни носителей положить равным

$$\tau_p^* = \frac{\tau_p}{1 + j\omega \tau_p}. \quad (62)$$

Сделав соответствующие подстановки в выражение для полного тока (44), получим плотность переменного тока

$$J_1 = \frac{qV_1}{kT} \left[\frac{qD_p p_{n0}}{L_p / \sqrt{1 + j\omega \tau_p}} + \frac{qD_n n_{p0}}{L_n / \sqrt{1 + j\omega \tau_n}} \right] \exp \left(\frac{qV_0}{kT} \right). \quad (63)$$

Из выражения (63) непосредственно вычисляем полную проводимость перехода по переменному току

$$Y \equiv \frac{J_1}{V_1} = G_d + j\omega C_d. \quad (64)$$

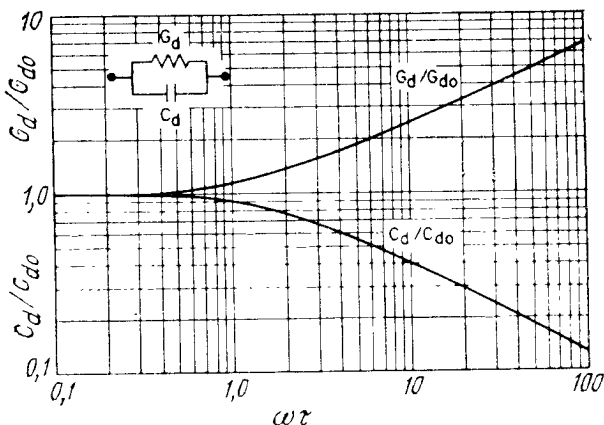


Рис. 23. Зависимость диффузионной проводимости и диффузионной емкости от $\omega\tau$. На вставке приведена эквивалентная схема $p-n$ -перехода при прямом смещении.

Для относительно низких частот ($\omega\tau_p, \omega\tau_n \ll 1$) диффузионная проводимость G_{d0} определяется выражением

$$G_{d0} = \frac{q}{kT} \left(\frac{qD_p p_{n0}}{L_p} + \frac{qD_n n_{p0}}{L_n} \right) e^{qV_0/kT} \text{ [МО/см}^2\text{]}, \quad (65)$$

которое совпадает с производной от полного тока, описываемого формулой (44). Низкочастотная диффузионная емкость записывается в виде

$$C_{d0} = \frac{q}{kT} \left(\frac{qL_p p_{n0}}{2} + \frac{qL_n n_{p0}}{2} \right) e^{qV_0/kT} \text{ [Ф/см}^2\text{]}. \quad (66)$$

Зависимость проводимости и емкости от нормализованной частоты $\omega\tau$ приведена на рис. 23, причем в расчет принимался только один член выражения (63) (например, член, содержащий p_{n0} , если $p_{n0} \gg n_{p0}$). На вставке показана эквивалентная схема полной проводимости перехода для переменного сигнала. Из рис. 23 следует, что диффузионная емкость снижается с повышением частоты сигнала. Для высоких частот она уменьшается приблизительно по закону $(\omega)^{-1/2}$. Вместе с тем диффузионная емкость быстро возрастает с уровнем постоянного тока ($\sim e^{qV_0/kT}$). По этим причинам диффузионная емкость играет особенно большую роль на низких частотах и при прямом смещении.

2.5. ПРОБОЙ $p-n$ -ПЕРЕХОДА [31]

При большом обратном смещении на $p-n$ -переходе, которое создает в нем большое электрическое поле, переход «пробивается» и через него протекает очень большой ток. Существуют три основ-

ных механизма пробоя: тепловая неустойчивость, туннельный эффект и лавинное умножение. Первые два механизма мы рассмотрим кратко, а лавинное умножение — более подробно.

2.5.1. Тепловая неустойчивость

Пробой, обусловленный тепловой неустойчивостью, определяет электрическую прочность большинства изоляторов при комнатной температуре, а также является основным фактором в полупроводниках с относительно узкой шириной запрещенной зоны, например в германии. В $p-n$ -переходе при обратном напряжении протекает обратный ток и, следовательно, выделяется тепло. Увеличение обратного напряжения вызывает повышение температуры перехода. В свою очередь это приводит к возрастанию обратного тока. Влияние температуры на обратную ветвь вольт-амперной характеристики перехода показано на рис. 24 [32], где семейство горизонтальных линий представляет собой различные значения обратного тока. Каждая линия соответствует определенному току при постоянной температуре перехода, а сам ток изменяется с температурой по закону $T^{3+v/2} \exp(-E_g/kT)$. Количество выделяемого переходом тепла, определяющее его температуру, пропорционально произведению $I \cdot V$. Поэтому гиперболические кривые постоянной температуры в двойном логарифмическом масштабе представляют собой семейство прямых. Если соединить точки пересечения кривых, соответствующих одинаковой температуре, получим обратную ветвь вольт-амперной

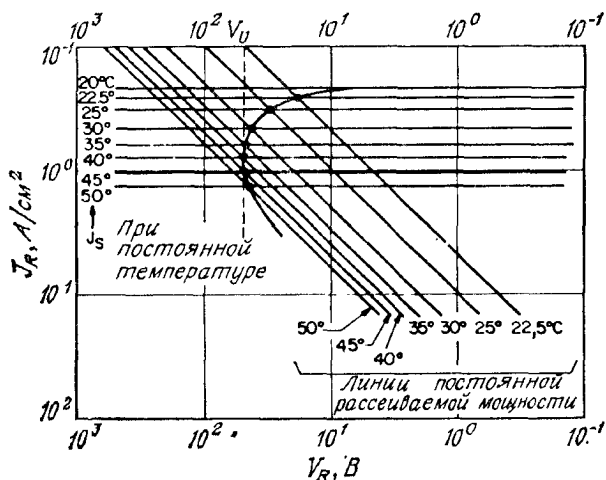


Рис. 24. Обратная ветвь вольт-амперной характеристики при тепловом пробое (V_U — напряжение тепловой неустойчивости) [32].

характеристики перехода. Вследствие теплового нагрева при больших обратных напряжениях на характеристике возникает участок с отрицательным дифференциальным сопротивлением. На этом участке диод выходит из строя, если не принять специальных мер для ограничения тока (например, включить последовательное сопротивление). Этот эффект называют тепловой неустойчивостью, а напряжение V_U (рис. 24) — напряжением тепловой неустойчивости. В $p-n$ -переходах со сравнительно большими токами насыщения, например в германиевых переходах, тепловая неустойчивость проявляется уже при комнатных температурах. Однако при очень низких температурах тепловая неустойчивость становится несущественной по сравнению с другими механизмами пробоя.

2.5.2. Туннельный эффект

Хорошо известно, что квантовомеханическая вероятность прохождения сквозь одномерный прямоугольный потенциальный барьер с высотой E_0 и шириной W определяется выражением [33]

$$T_t = \left[1 + \frac{E_0^2 \operatorname{sh}^2 \kappa W}{4E(E_0 - E)} \right]^{-1}, \quad (67)$$

где $\kappa \equiv \sqrt{\frac{2m(E_0 - E)}{\hbar^2}}$,

а E — энергия частицы. С уменьшением энергии E вероятность туннелирования монотонно убывает. При $\kappa W \gg 1$ она приближенно равна

$$T_t \approx \frac{16E(E_0 - E)}{E_0^2} \exp(-2\kappa W). \quad (67a)$$

Аналогичное выражение получится и для $p-n$ -переходов. Подробное математическое обоснование приведено в гл. 9. Плотность туннельного тока определяется выражением [31]

$$J_t = \frac{\sqrt{2m^*} q^3 \mathcal{E} V}{4\pi^2 \hbar^2 E_g^{1/2}} \exp\left(-\frac{4\sqrt{2m^*} E_g^{3/2}}{3q\mathcal{E}\hbar}\right), \quad (68)$$

где \mathcal{E} — электрическое поле в переходе, E_g — ширина запрещенной зоны, V — приложенное напряжение, m^* — эффективная масса носителя.

Когда электрическое поле в Ge или Si достигает 10^6 В/см, начинают протекать большие токи, обусловленные туннельными переходами между зонами. Чтобы получить такое сильное поле, концентрации примеси в p - и n -областях должны быть достаточно высокими. Доказано, что механизм пробоя в кремниевых и германиевых переходах обязан туннельному эффекту при напряжениях пробоя, меньших $4E_g/q$. В переходах с напряжением пробоя,

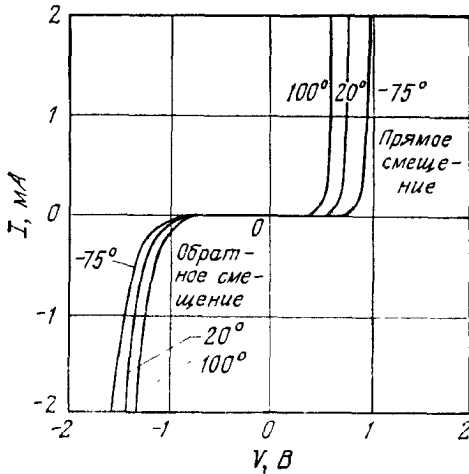


Рис. 25. Вольт-амперная характеристика перехода с туннельным пробоем [32].

превышающим $6E_g/q$, механизм пробоя обусловлен лавинным умножением. При напряжении пробоя, лежащем в интервале $4-6E_g/q$, в пробое участвуют оба механизма (лавинный и туннельный). Поскольку ширина запрещенной зоны E_g в Ge, Si и GaAs уменьшается с повышением температуры (гл. 1), то напряжение пробоя в этих полупроводниках, связанное с туннельным эффектом, имеет отрицательный температурный коэффициент, т. е. падает с повышением температуры. Это объясняется тем, что при более высоких температурах заданная величина тока при пробое J_t достигается при меньших обратных напряжениях (или полях). Типичный пример семейства характеристик диода с туннельным пробоем при различных температурах показан на рис. 25. Температурная зависимость обычно используется для распознавания механизмов пробоя, поскольку лавинный пробой имеет положительный температурный коэффициент, т. е. напряжение пробоя возрастает с повышением температуры.

2.5.3. Лавинное умножение

Лавинное умножение, или ударная ионизация, является наиболее важным механизмом пробоя $p-n$ -переходов. Напряжение лавинного пробоя определяет верхний предел обратного напряжения большинства диодов, коллекторного напряжения биполярных транзисторов (гл. 3), напряжения стока МЭП-транзисторов (гл. 6) и МОП-транзисторов (гл. 8). Кроме того, механизм ударной ионизации используется в мощных генераторах СВЧ-колебаний, таких, как лавинно-пролетные диоды (гл. 10), и в приемниках оптических сигналов, таких, как лавинные фотодиоды (гл. 13).

Найдем сначала основное выражение для условия пробоя. Пусть с левой стороны обедненной области шириной W втекает ток I_{p0} . Если электрическое поле в обедненной области настолько велико, что вследствие ударной ионизации могут генерироваться электронно-дырочные пары, то дырочный ток I_p будет нарастать с координатой и на правой границе обедненной области ($x = W$) достигнет величины $M_p I_{p0}$. Аналогично электронный ток I_n будет возрастать в направлении от $x = W$ к $x = 0$. Полный ток $I (= I_p + I_n)$ в стационарных условиях постоянен. Приращение дырочного тока в точке с координатой x равно числу электронно-дырочных пар, генерируемых за 1 с на расстоянии dx :

$$d(I_p/q) = (I_p/q)(\alpha_p dx) + (I_n/q)(\alpha_n dx) \quad (69)$$

или

$$dI_p/dx - (\alpha_p - \alpha_n) I_p = \alpha_n I. \quad (70)$$

Коэффициенты ионизации электронов и дырок (α_n и α_p) рассмотрены в гл. 1.

Решение ¹⁾ уравнения (70) с граничным условием $I = I_p(W) = M_p I_{p0}$ записывается в виде

$$I_p(x) = I \left\{ \frac{1}{M_p} + \int_0^x \alpha_n \exp \left[- \int_0^x (\alpha_p - \alpha_n) dx' \right] dx \right\} / \exp \times \\ \times \left[- \int_0^x (\alpha_p - \alpha_n) dx' \right], \quad (71)$$

где M_p — коэффициент умножения дырок, равный

$$M_p \equiv \frac{I_p(W)}{I_p(0)}. \quad (72)$$

Уравнение (71) можно записать в виде

$$1 - \frac{1}{M_p} = \int_0^W \alpha_p \exp \left[- \int_0^x (\alpha_p - \alpha_n) dx' \right] dx. \quad (73)$$

¹⁾ Уравнение (70) имеет вид $y' + Py = Q$, где $y \equiv I_p$. Его решением является

$$y = \left[\int_0^x Q e^{\int_0^x P dx'} dx + C \right] / e^{\int_0^x P dx'}$$

где C — постоянная интегрирования.

Напряжением лавинного пробоя называется напряжение, при котором M_p стремится к бесконечности. Следовательно, условие пробоя задается интегралом ионизации

$$\int_0^W \alpha_p \exp \left[- \int_0^x (\alpha_p - \alpha_n) dx' \right] dx = 1. \quad (74)$$

Если лавинный процесс инициируется не дырками, а электронами, то интеграл ионизации имеет вид

$$\int_0^W \alpha_n \exp \left[- \int_x^W (\alpha_n - \alpha_p) dx' \right] dx = 1. \quad (75)$$

Уравнения (74) и (75) эквивалентны [34], т. е. условия пробоя зависят только от процессов внутри обедненной области, а не от носителей или первоначальных токов, вызвавших лавинный процесс. Ситуация не меняется, если пробой вызван смешанными первичными токами: лишь бы выполнялось одно из условий — (74) или (75).

Для полупроводников с равными коэффициентами ионизации ($\alpha_n = \alpha_p = \alpha$), например для GaP, интегралы (74) и (75) принимают вид

$$\int_0^W \alpha dx = 1. \quad (76)$$

Используя полученные условия пробоя и располагая зависимостью коэффициентов ионизации от поля, можно определить напряжение пробоя, максимальное электрическое поле и ширину обедненной области. Электрическое поле и потенциал обедненного слоя находятся из уравнения Пуассона. Положение границ слоя, удовлетворяющее уравнению (74), можно определить численно методом последовательных приближений. Если ширина обедненной области известна, то напряжение пробоя V_B несимметричного резкого перехода

$$V_B = \frac{\mathcal{E}_m W}{2} = \frac{e_s \mathcal{E}_m^2}{2q} (N_B)^{-1}, \quad (77a)$$

а для линейного перехода

$$V_B = \frac{2\mathcal{E}W}{3} = \frac{4\mathcal{E}_m^3}{3} \left(\frac{2e_s}{q} \right)^{1/2} (a)^{-1/2}, \quad (77b)$$

где N_B — концентрация ионизированной примеси в слаболегированной области, e_s — диэлектрическая проницаемость полупроводника, a — градиент концентрации примеси, \mathcal{E}_m — максимальное электрическое поле.

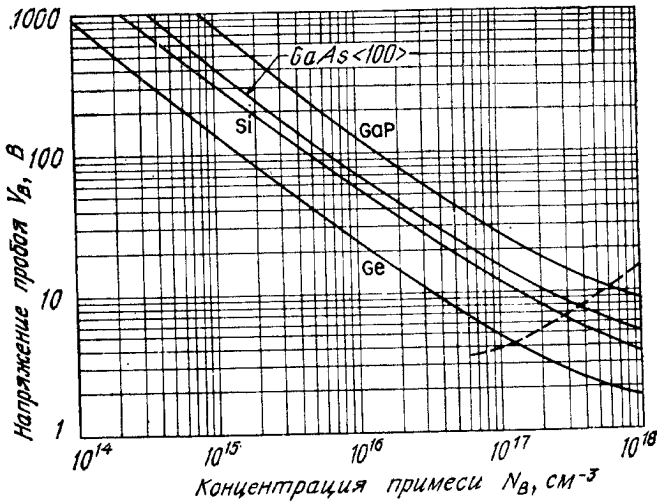


Рис. 26. Зависимость напряжения лавинного пробоя от концентрации примеси для несимметричного ступенчатого перехода в Ge, Si, GaAs с ориентацией $\langle 100 \rangle$ и в GaP [35].

Расчетные зависимости напряжения пробоя от N_B для резких переходов в Ge, Si, GaAs с ориентацией $\langle 100 \rangle$ и в GaP [35] приведены на рис. 26. Экспериментальные результаты обычно хорошо согласуются с расчетными значениями [36]. Штриховой линией отмечен верхний предел N_B , при котором справедливы выражения

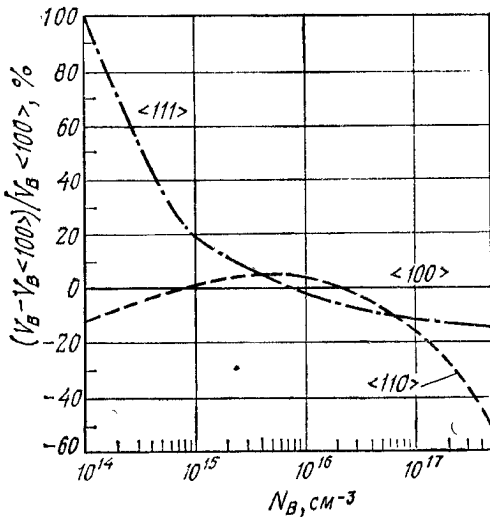


Рис. 27. Зависимость напряжения пробоя в несимметричных резких переходах в GaAs для различных ориентаций [37].

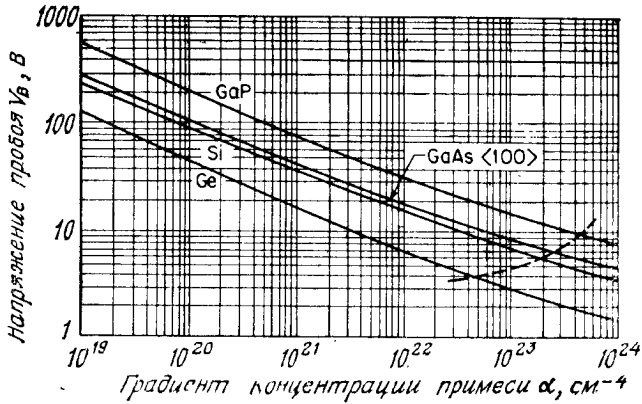


Рис. 28. Зависимость напряжения лавинного пробоя от градиента концентрации примеси для плавных линейных переходов в Ge, Si, GaAs с ориентацией $\langle 100 \rangle$ и GaP [35].

для расчета лавинного пробоя. Этот предел выбирается на основе критерия $6E_g/q$. При более высоких концентрациях существенный вклад в пробой начинает вносить туннельный механизм, и в конце концов он станет преобладающим.

В арсениде галлия коэффициент ударной ионизации зависит от ориентации кристалла (гл. 1). На рис. 27 приведены зависимости напряжения пробоя V_B при ориентации $\langle 111 \rangle$ и $\langle 110 \rangle$ и ориентации $\langle 100 \rangle$. Видна слабая зависимость напряжения пробоя от ориентации при концентрации примеси $\sim 10^{16} \text{ см}^{-3}$. При более слабом легировании V_B максимально в случае ориентации $\langle 111 \rangle$, а при более сильном легировании V_B максимально в случае ориентации $\langle 100 \rangle$ [37].

На рис. 28 приведена зависимость напряжения пробоя от градиента концентрации примеси для плавных линейных переходов в перечисленных выше полупроводниках. Штриховой линией отмечен верхний предел α , при котором еще верен расчет лавинного пробоя.

Расчетные зависимости максимального поля \mathcal{E}_m и ширины обедненного слоя при пробое резких переходов в четырех рассмотренных полупроводниках даны на рис. 29 [35]. На рис. 30 аналогичные зависимости приведены для линейных переходов. В $p-n$ -переходе на кремнии максимальное поле определяется выражением [38]

$$\mathcal{E}_m = \frac{4 \cdot 10^3}{1 - \frac{1}{3} \lg(N_B/10^{16})} \left[\frac{\text{В}}{\text{см}} \right], \quad (78)$$

где N_B измеряется в см^{-3} .

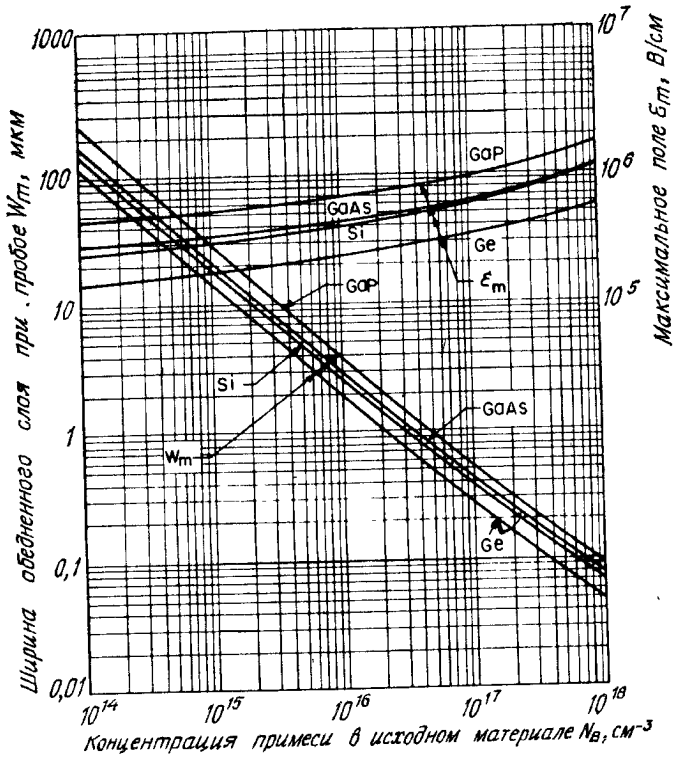


Рис. 29. Экспериментальные зависимости дрейфовой скорости (а) от электрического поля в чистых кристаллах Ge, Si и GaAs [37, 59, 60] и температурная зависимость скорости насыщения электронов (б) в Si и GaAs [37, 61, 62].

Вследствие сильной зависимости коэффициентов ионизации от поля величина максимального поля очень слабо зависит от N_B и α . Поэтому в первом приближении для данного полупроводника можно считать \mathcal{E}_m постоянным. Тогда из выражения (77) следует, что для резкого перехода $V_B \sim N_B^{-1,0}$, а для линейного $V_B \sim \alpha^{-0,5}$. Подтверждением этого служат графики на рис. 26 и 28. Кроме того, как и следовало ожидать, при фиксированных N_B или α напряжение пробоя возрастает с увеличением ширины запрещенной зоны, поскольку при лавинном умножении происходит переход носителей из одной зоны в другую.

На основе рассмотренных выше результатов можно получить приближенную универсальную формулу напряжения пробоя резкого перехода, справедливую для всех изученных полупроводников:

$$V_B \cong 60 (E_R/1,1)^{3/2} (N_B/10^{16})^{-3/4} [\text{В}], \quad (79a)$$

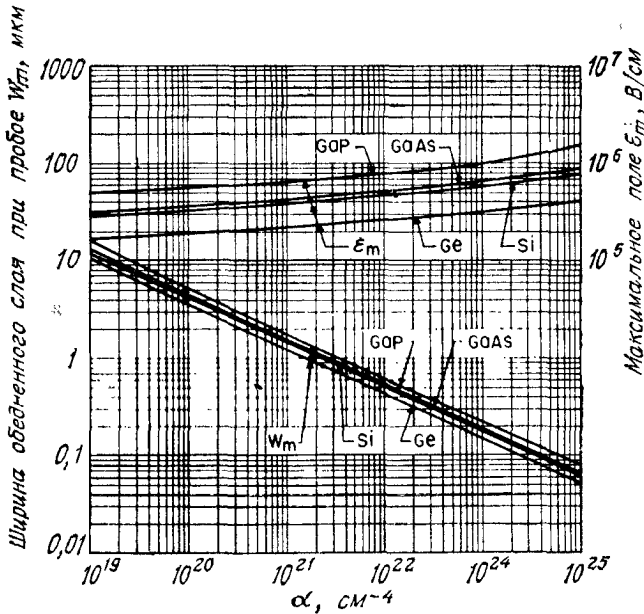


Рис. 30. Ширина обедненного слоя и максимальное поле для плавных линейных переходов в Ge, Si, GaAs с ориентацией (100) и GaP [35].

где E_g — ширина запрещенной зоны при комнатной температуре, эВ; N_B — концентрация примеси в слаболегированной области, см⁻³. Аналогичное выражение для линейного перехода имеет вид

$$V_B \cong 60 (E_g/1,1)^{6/5} (\alpha/3 \cdot 10^{20})^{-2/5} [B], \quad (796)$$

где α — градиент концентрации примеси, см⁻⁴.

Для $p-n$ -перехода, полученного диффузионным методом, с линейным распределением примеси на одной стороне и с постоянной концентрацией примеси на другой стороне перехода (вставка на рис. 31) напряжение пробоя принимает промежуточное значение между напряжениями пробоя ступенчатого и линейного переходов (рис. 26 и 28) [39]. Для больших значений α и низких значений N_B напряжение пробоя диффузионного перехода (рис. 31) приближается к напряжениям для резкого перехода (нижняя линия); с другой стороны, для малых α и высоких N_B напряжение близко к напряжениям для линейного перехода (параллельные линии).

Результаты, представленные на рис. 26—30, получены в предположении достаточно большой толщины полупроводникового слоя, чтобы при пробое мог существовать обедненный слой шириной W_m (рис. 29). Однако если ширина полупроводникового слоя

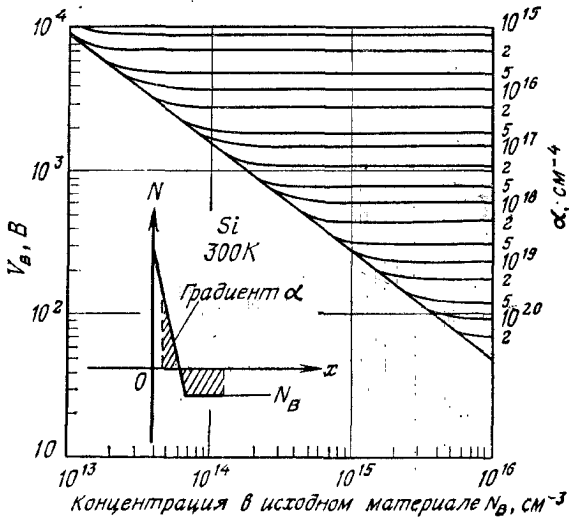


Рис. 31. Напряжение пробоя для диффузионных переходов. На вставке показано распределение пространственного заряда [39].

W меньше W_m (рис. 32, вставка), то еще до лавинного пробоя произойдет прокол прибора (т. е. обедненный слой достигнет границы раздела $n - n^+$). При дальнейшем увеличении обратного смещения неизбежно наступит пробой прибора. Максимальное электрическое поле \mathcal{E}_m примерно такое же, как и в приборе без

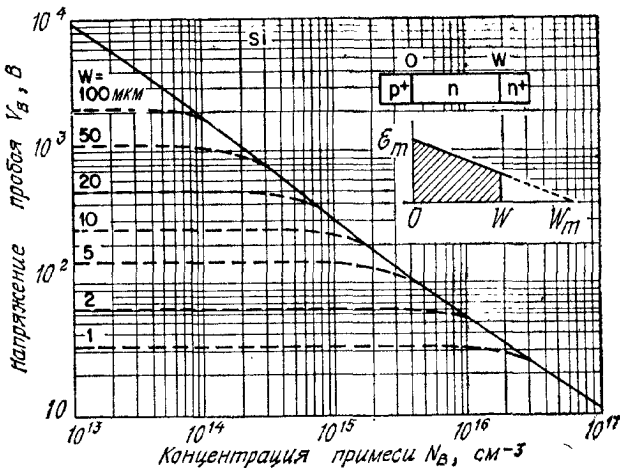


Рис. 32. Напряжение пробоя $p^+ - n - n^+$ и $p^+ - v - n^+$ -переходов.

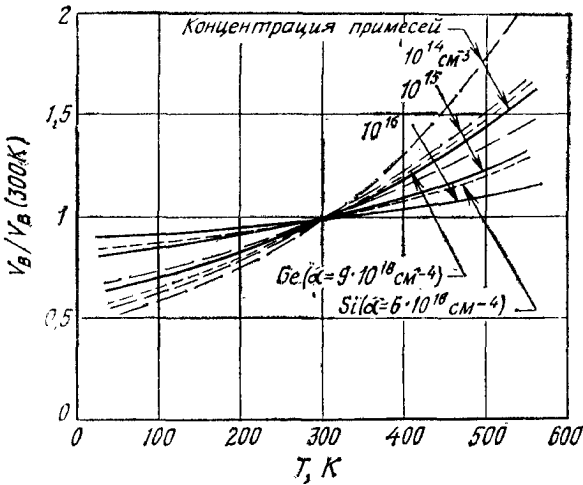


Рис. 33. Температурная зависимость нормализованного напряжения лавинного пробоя [40].

— резкий переход в Si; — — — резкий переход в Ge; - - - линейные переходы

прокола. Следовательно, для напряжения пробоя V_{BT} диода с проколом можно написать

$$\begin{aligned} \frac{V_{BT}}{V_B} &= \frac{\text{Заштрихованная область (вставка на рис. 32)}}{(\mathcal{E}_m W_m)/2} = \\ &= \left(\frac{W}{W_m} \right) \left(2 - \frac{W'}{W_m} \right). \end{aligned} \quad (80)$$

Прокол обычно возникает при достаточно низкой концентрации легирующей примеси N_B , что имеет место в $p^+ - n - n^+$ и $p^+ - n - n^+$ -диодах (n означает слаболегированный полупроводник p -типа, а n — слаболегированный полупроводник n -типа). Рассчитанная по формуле (80) зависимость напряжения пробоя таких диодов от концентрации примеси в слаболегированной области несимметричного резкого перехода в кремнии приведена на рис. 32 (сплошная линия). Предполагается, что переход сформирован на эпитаксиальной подложке (т. е. на n^+ -подложке выращен эпитаксиальный слой n -типа толщиной W). Для заданной толщины (штриховая линия) при уменьшении концентрации примеси напряжение пробоя стремится к постоянному значению, соответствующему проколу эпитаксиального слоя.

Результаты, приведенные на рис. 26—32, относятся к лавинному пробоя при комнатной температуре. С повышением температуры напряжение пробоя возрастает. Это можно объяснить тем, что горячие носители, проходя через обедненный слой под действием сильного поля, теряют часть своей энергии при столк-

повении с оптическими фононами. Средняя длина свободного пробега электронов между электрон-фононными столкновениями λ падает с повышением температуры (выражение (84) гл. 1). Поэтому при постоянном поле носители, проходя заданное расстояние, отдают кристаллической решетке большую часть энергии. Следовательно, для того чтобы носители набрали энергию, достаточную для генерации электронно-дырочных пар, требуется большая разность потенциалов. Результаты расчетов на основе модифицированной теории Бараффа [40] (гл. 1) приведены на рис. 33, где напряжения пробоя V_B для Ge и Si нормализованы относительно значений при комнатной температуре. Из вычислений следует, что при одинаковых профилях легирования относительные изменения V_B перехода с температурой в GaAs примерно совпадают с изменениями в Ge, а в GaP — с изменениями в Si. Необходимо обратить внимание на существенное увеличение напряжения пробоя с повышением температуры при относительно низких концентрациях примеси (или малых градиентах концентрации) [41]. На рис. 34 приведены экспериментальные результаты [42], которые достаточно хорошо согласуются с теоретическими.

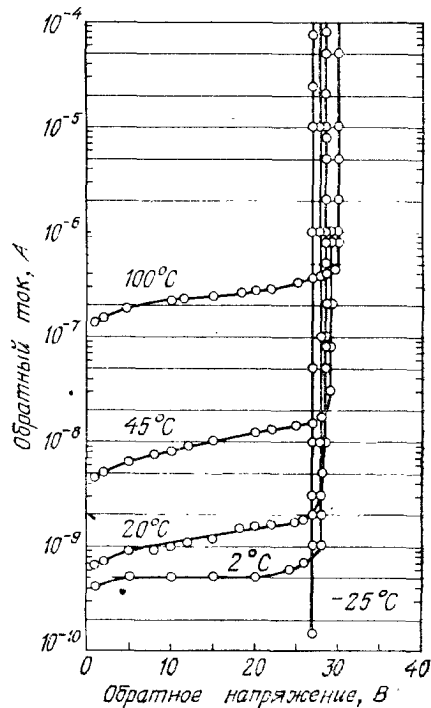


Рис. 34. Обратная ветвь вольт-амперной характеристики кремниевого $n^+ - p$ -диода без микроплазм и с охраным кольцом ($N_B = 2,5 \cdot 10^{16} \text{ см}^{-3}$) при различных температурах. Температурный коэффициент равен $0,024 \text{ В/}^\circ\text{C}$ [42].

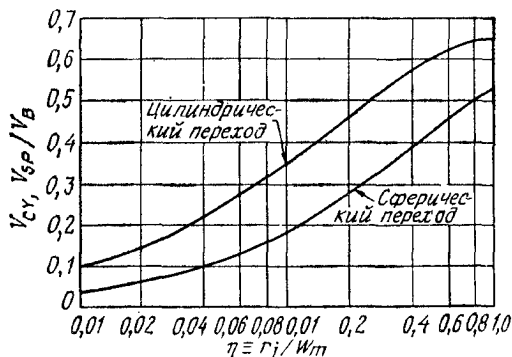


Рис. 35. Нормализованные зависимости напряжения пробоя цилиндрического и сферического переходов от кривизны перехода [39].

В планарных $p-n$ -переходах необходимо учитывать очень сильное влияние кривизны перехода. Схема планарного перехода приведена на рис. 9, б. Поскольку напряженность электрического поля на цилиндрических и/или сферических областях перехода выше, то напряжение пробоя определяется именно этими областями. Потенциал $V(r)$ и электрическое поле $\mathcal{E}(r)$ в цилиндрическом или сферическом переходе можно найти, решая уравнение Пуассона

$$\frac{1}{r^n} \frac{d}{dr} |r^n \mathcal{E}(r)| = -\frac{\rho(r)}{\epsilon_s}, \quad (81)$$

где $n = 1$ для цилиндрических переходов и $n = 2$ для сферических. Решая уравнение (81), получим

$$\mathcal{E}(r) = \frac{1}{\epsilon_s r^n} \int_{r_j}^r r^n \rho(r) dr + \frac{\text{const}}{r^n}, \quad (82)$$

где r_j — радиус кривизны металлургического перехода, а константа должна быть выбрана такой, чтобы удовлетворялись условия пробоя (74) или (75).

Из полученных численными методами результатов следует, что при 300 К для несимметричных резких цилиндрических переходов в кремнии справедливо следующее простое аналитическое выражение [39]:

$$\frac{V_{CY}}{V_B} = \left[\frac{1}{2} (\eta^2 + 2\eta^{6/7}) \ln(1 + 2\eta^{-8/7}) - \eta^{6/7} \right], \quad (83)$$

а для сферических переходов

$$\frac{V_{SP}}{V_B} = [\eta^2 + 2,14\eta^{6/7} - (\eta^3 + 3\eta^{13/7})^{2/3}], \quad (84)$$

где V_{CY} и V_{SP} — напряжение пробоя цилиндрического и сферического перехода соответственно, V_B — напряжение пробоя пло-

скостного перехода, имеющего ту же концентрацию примесей, и $\eta \equiv r_j/W_m$. Зависимость напряжения пробоя от η для цилиндрического и сферического резких переходов приведена на рис. 35. Что касается цилиндрического и сферического переходов с линейным распределением примесей, то из результатов вычислений следует сравнительно слабая зависимость напряжения пробоя от радиуса кривизны [25].

2.6. ПЕРЕХОДНЫЕ ПРОЦЕССЫ И ШУМЫ

2.6.1. Переходные процессы

При использовании диодов в качестве переключателей требуется, чтобы переход от прямого смещения к обратному был как можно резче, а время переключения — как можно меньше. На рис. 36, а приведена простейшая схема переключения диода. Через переход протекает прямой ток I_F . В момент времени $t = 0$ ключ S мгновенно переводится в правое положение, и некоторое время через переход протекает обратный ток $I_R \approx V/R$. Время переключения определяется как время, за которое ток снижается до 10 % от начального тока I_R . Как показано на рис. 36, б, оно

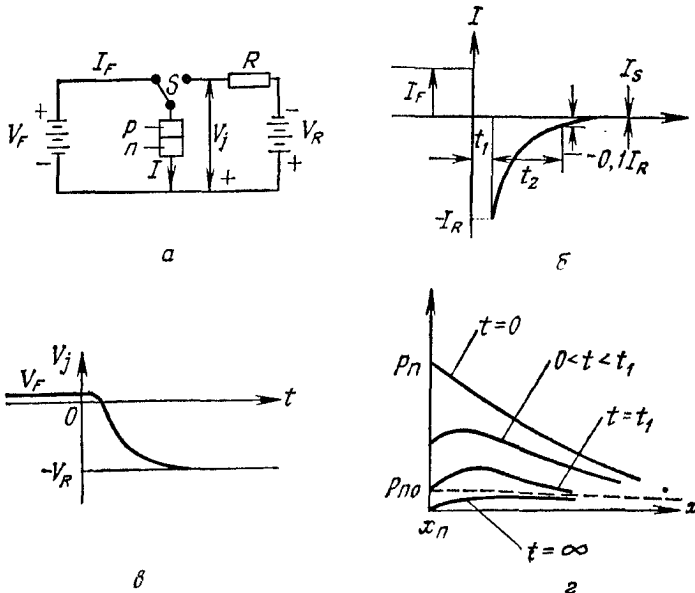


Рис. 36. Переходные процессы в $p-n$ -переходе [43].

а — простейшая схема переключения; б — изменение тока при переключении; в — изменение напряжения; г — распределение неосновных носителей в различные моменты времени.

равно сумме t_1 и t_2 , где t_1 — интервал времени фазы постоянного тока и t_2 — интервал времени фазы спада тока. Сначала рассмотрим фазу постоянного тока (называемую также фазой хранения). Уравнение непрерывности, приведенное в гл. 1, можно записать для области n -типа (при условии $p_{p0} \ll n_{n0}$) в виде

$$\frac{\partial p_n(x, t)}{\partial t} = D_p \frac{\partial^2 p_n(x, t)}{\partial x^2} - \frac{p_n(x, t) - p_{n0}}{\tau_p}, \quad (85)$$

где τ_p — время жизни неосновных носителей. В качестве граничных условий выступают начальное распределение дырок в момент $t = 0$, вытекающее из стационарного решения уравнения диффузии, и напряжение на переходе, определяемое выражением (33):

$$V_i = \frac{kT}{q} \ln \left[\frac{p_n(0, t)}{p_{n0}} \right]. \quad (86)$$

Зависимость распределения плотности неосновных носителей p_n от времени [43] показана на рис. 36, *г*. Из формулы (86) следует, что до тех пор, пока $p_n(0, t)$ больше p_{n0} (на интервале времени $0 < t < t_1$), напряжение $V_i \approx kT/q$ (рис. 36, *в*), а ток I_R приблизительно равен $\frac{V}{R} = \text{const}$. Следовательно, на этом этапе обратный ток не меняется, и мы имеем дело с фазой постоянного тока. Однако по мере приближения к моменту t_1 концентрация дырок стремится к нулю, обратное напряжение на переходе быстро возрастает и начинает выполняться новое граничное условие. Наступает фаза спада тока, для которой справедливо граничное условие $p(0, t) = p_{n0} = \text{const}$. Времена t_1 и t_2 в соответствии с решением, полученным Кингстоном [43], находятся из трансцендентных уравнений

$$\text{erf} \sqrt{\frac{t_1}{\tau_p}} = \frac{1}{1 + I_R/I_F}, \quad (87)$$

$$\text{erf} \sqrt{\frac{t_2}{\tau_p}} = \frac{\exp(-t_2/\tau_p)}{\sqrt{\pi} (t_2/\tau_p)} - 1 + 0,1 \left(\frac{I_R}{I_F} \right). \quad (88)$$

На рис. 37 приведены результаты расчета для плоскостного перехода с шириной базы n -типа W , много большей диффузионной длины дырок L_p (сплошные линии), и для перехода с узкой базой ($W \ll L_p$) (штриховые линии). Для больших отношений I_R/I_F в случае $W \gg L_p$ время переключения равно

$$t_1 + t_2 \approx \frac{\tau_p}{2} \left(\frac{I_R}{I_F} \right)^{-1}, \quad (89a)$$

а в случае $W \ll L_p$

$$t_1 + t_2 \approx \frac{W^2}{2D_p} \left(\frac{I_R}{I_F} \right)^{-1}. \quad (89b)$$

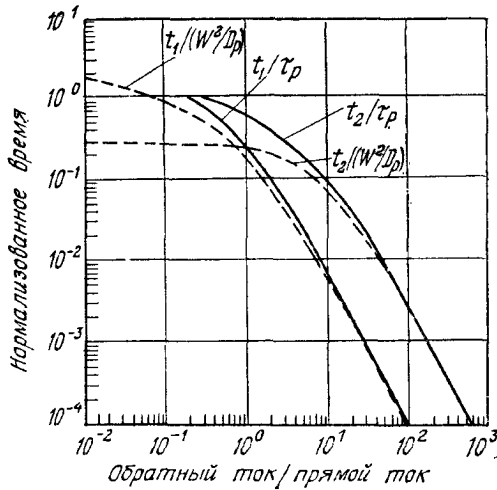


Рис. 37. Зависимость нормализованного времени от отношения обратного тока к прямому току [43].

При переключении перехода с широкой базой ($W \gg L_p$) от прямого тока 10 мА к обратному току 10 мА длительность фазы постоянного тока составляет $0,3\tau_p$, а длительность фазы спада $0,6\tau_p$. Полное время переключения в данном случае равно $0,9\tau_p$. Таким образом, для быстрого переключения требуется уменьшить время жизни неосновных носителей τ_p . Для его снижения вводят примеси, создающие глубокие уровни в запрещенной зоне, такие, как золото в кремнии.

2.6.2. Шумы

Термин «шумы» относится к самопроизвольным флуктуациям тока, протекающего через полупроводниковые материалы и приборы, и к флуктуациям напряжения на них. Поскольку приборы в основном используются для измерения малых физических величин или для усиления слабых сигналов, то спонтанные флуктуации тока или напряжения ограничивают снизу предел измеряемых величин или усиливаемых сигналов. Важно знать факторы, определяющие эти пределы, чтобы целенаправленно оптимизировать условия работы приборов и искать новые способы и новые технологические методы снижения уровня шумов.

Наблюдаемые шумы обычно подразделяют на тепловой шум, фликер-шум и дробовой шум. Тепловой шум имеет место в любом проводнике или полупроводнике и обусловлен хаотическим движением носителей тока. Средний квадрат напряжения тепловых

шумов в разомкнутой цепи $\langle V_n^2 \rangle$ определяется выражением [44, 45]

$$\langle V_n^2 \rangle = 4kTBR, \quad (90)$$

где k — постоянная Больцмана; T — абсолютная температура, К; B — ширина полосы пропускания, Гц; R — действительная часть импеданса, подключенного к выходным клеммам. При комнатной температуре в полупроводниковом материале с сопротивлением 1 кОм среднеквадратичное напряжение шума, измеренное в полосе частот 1 Гц, составляет ~ 4 нВ ($1 \text{ нВ} = 10^{-9} \text{ В}$).

Фликер-шум отличается специфическим спектральным распределением, пропорциональным $1/f^\alpha$, где величина α обычно близка к 1 (так называемый шум $1/f$) и играет важную роль на низких частотах. Для большинства полупроводниковых приборов происхождение фликер-шума связано с поверхностными эффектами. Установлена качественная и количественная корреляция спектральной плотности шумов $1/f$ с той частью потерь в импедансе структуры металл — диэлектрик — полупроводник (МДП), которая обусловлена комбинацией носителей на поверхностных ловушках у границы раздела.

Дробовой шум является преобладающим для большинства полупроводниковых приборов. При низких и средних частотах он не зависит от частоты, т. е. имеет «белый» спектр. Средний квадрат тока дробового шума p — n -перехода определяется выражением

$$\langle i_n^2 \rangle = 2qBI, \quad (91)$$

где I — абсолютная величина тока через переход. При низком уровне инжекции средний квадрат полного тока (без учета шума $1/f$) составляет

$$\langle i_n^2 \rangle = 4kTBG - 2qBI. \quad (92)$$

Из формулы Шокли получим

$$G = \frac{\partial I}{\partial V} = \frac{\partial}{\partial V} [I_s (e^{qV/kT} - 1)] = \frac{qI_s}{kT} e^{qV/kT}. \quad (93)$$

Подставляя выражение (93) в выражение (92), найдем, что в случае прямого смещения

$$\langle i_n^2 \rangle = 2qI_s B e^{qV/kT} + 2qBI_s. \quad (94)$$

Экспериментальные данные действительно подтверждают, что средний квадрат шумового тока пропорционален току насыщения I_s , который можно изменять с помощью освещения перехода.

2.7. СХЕМНЫЕ ФУНКЦИИ

Полупроводниковый $p-n$ -переход является двухполюсником, который способен выполнять в схемах различные функции в зависимости от приложенного напряжения, а также от профиля примесей и геометрии перехода. В этом разделе кратко рассмотрен ряд интересных применений диодов, основанных на свойствах вольт-амперных и вольт-фарадных характеристик, а также на характеристиках пробоя, рассмотренных в предыдущих разделах. Некоторые типы двухполюсников описаны в последующих главах (например, туннельные диоды в гл. 9 и 10).

2.7.1. Выпрямители

Выпрямитель представляет собой плоскостной диод, спроектированный специально для выпрямления переменного тока. Он должен обладать очень низким сопротивлением для тока, протекающего в одном направлении, и очень высоким сопротивлением — в другом направлении. Прямое и обратное сопротивления выпрямителя легко находятся из формулы вольт-амперной характеристики реального диода

$$I = I_s (e^{qV/nkT} - 1), \quad (95)$$

где I_s — ток насыщения, а множитель n обычно принимает значения между 1 (для диффузионного тока) и 2 (для рекомбинационного тока). Из формулы (95) можно определить прямое сопротивление на постоянном токе (статическое сопротивление) R_F и сопротивление для малого сигнала (динамическое сопротивление) r_F :

$$\begin{aligned} R_F &\equiv \frac{V_F}{I_F} \left(\simeq \frac{V_F}{I_s} e^{-qV_F/nkT} \text{ для } V \geq 3kT/q \right), \\ z_F &\equiv \frac{dV_F}{dI_F} = \frac{nkT}{qI_F}. \end{aligned} \quad (96)$$

Обратные сопротивления на постоянном токе R_R и на малом сигнале r_R определяются выражениями

$$\begin{aligned} R_R &\equiv \frac{V_R}{I_R} \simeq \left(\frac{V_R}{I_s} \text{ для } |V_R| \geq 3kT/q \right), \\ r_R &\equiv \frac{\partial V_R}{\partial I_R} = \frac{nkT}{qI_s} e^{q|V_R|/kT}. \end{aligned} \quad (97)$$

Из сравнения выражений (96) и (97) следует, что коэффициент выпрямления на постоянном токе R_R/R_F изменяется по закону $\exp(qV_F/nkT)$, а коэффициент выпрямления на переменном токе r_R/r_F — по закону $I_R/I_s \exp(-q|V_R|/kT)$.

Выпрямители обычно обладают низкой скоростью переключения; другими словами, переключение из открытого состояния с высокой проводимостью в закрытое состояние с высоким импедансом сопровождается большой задержкой во времени. Такая задержка, пропорциональная времени жизни неосновных носителей (рис. 37), не имеет значения при выпрямлении переменного тока с частотой 60 Гц. Для сохранения эффективного выпрямления на высоких частотах следует существенно уменьшить время жизни. Большинство выпрямителей имеет мощность рассеяния 0,1—10 Вт, напряжение обратного пробоя 50—2500 В (в высоковольтных выпрямителях включают последовательно два и более $p-n$ -перехода) и время переключения от 50 нс для маломощных диодов до ~ 500 нс для мощных диодов.

2.7.2. Стабилитроны

Стабилитрон представляет собой плоскостной диод, работающий при обратном смещении в режиме пробоя. До наступления пробоя диод имеет очень высокое сопротивление, а после пробоя — очень малое динамическое сопротивление. В результате происходит ограничение, или стабилизация напряжения на уровне напряжения пробоя. Стабилитроны в основном изготавливаются из кремния, так как кремний имеет низкий ток насыщения и технология кремния лучше разработана. Как уже говорилось в разд. 3.5, при напряжении пробоя V_B , превышающем $6E_g/q$ (~ 8 В для кремния), основным механизмом пробоя является лавинное умножение (температурный коэффициент положителен). При $V_B < 4E_g/q$ (~ 5 В для кремния) механизм пробоя обусловлен межзонным туннелированием (температурный коэффициент отрицателен). При $4E_g/q < V_B < 6E_g/q$ пробой зависит от обоих механизмов. Соединив диод с отрицательным температурным коэффициентом последовательно с диодом с положительным температурным коэффициентом, можно получить стабилизатор напряжения с низким температурным коэффициентом порядка 0,002 %/°C, который пригоден в качестве источника опорного напряжения.

2.7.3. Варисторы

Варистором или регулируемым сопротивлением называется двухполюсник с нелинейной вольт-амперной характеристикой [46]. Нелинейность характеристики плоскостного диода следует из выражений (96) и (97). Аналогичными нелинейными характеристиками обладают диоды со структурой металл — проводник, рассматриваемые в гл. 5. Интересным применением варисторов является использование их в качестве симметричных ограничи-

телей напряжения на уровне 0,5 В. Для этого они соединяются параллельно разноименными полюсами. Такой двухдиодный прибор обладает прямой диодной вольт-амперной характеристикой независимо от полярности прикладываемого напряжения.

2.7.4. Варакторы

Варактором называется прибор, реактивностью которого можно управлять с помощью напряжения смещения. Варакторные диоды широко используются в параметрических усилителях, генераторах гармоник, смесителях, детекторах и в системах с электронной настройкой.

Основные вольт-фарадные характеристики перехода были выведены в разд. 2.3. Целесообразно представить полученные выше соотношения для резкого и линейного переходов в более общем виде. Одномерное уравнение Пуассона записывается следующим образом:

$$\frac{d^2V}{dx^2} = -\frac{qN}{\epsilon_s}, \quad (98)$$

где N — произвольное распределение примеси. В общем случае при условии, что одна из областей полупроводника сильно легирована, можно (рис. 38, а) считать

$$N = Bx^m \quad \text{для } x \geq 0. \quad (99)$$

При $m = 0$ имеем $B = N_N$, что соответствует однородному распределению примеси в слаболегированной области или несимметричному резкому переходу. При $m = 1$ распределение примеси соответствует несимметричному линейному переходу. Переход

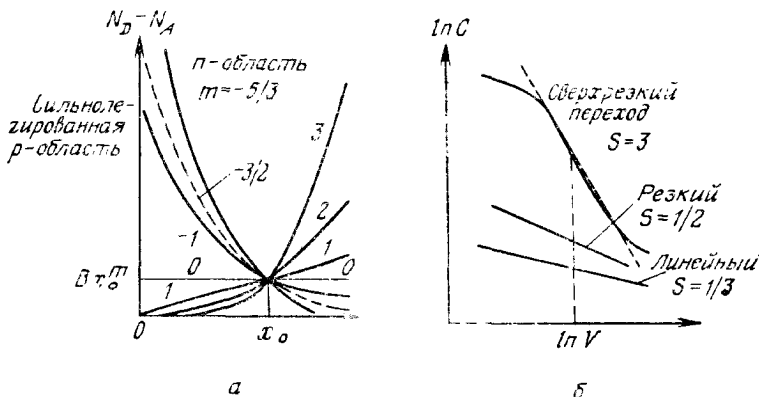


Рис. 38. Различные распределения примеси (а) в варакторах и зависимость барьерной емкости от обратного смещения (б) (двойной логарифмический масштаб) [48, 47].

с $m < 0$ называют сверхрезким. Сверхрезкий профиль примеси можно получить в эпитаксиальном процессе или с помощью ионного легирования. Граничные условия для уравнения (98) имеют вид $V(x=0) = 0$ и $V(x=W) = V + V_{bi}$, где V — приложенное напряжение, а V_{bi} — контактная разность потенциалов.

Интегрируя уравнение Пуассона с учетом указанных граничных условий, получим выражения для ширины обедненного слоя и удельной дифференциальной емкости перехода [47]:

$$W = \left[\frac{\epsilon_s (m+2) (V + V_{bi})}{qB} \right]^{1/(m+2)}, \quad (100)$$

$$C \equiv \frac{\partial Q_c}{\partial V} = \left[\frac{qB \epsilon_s^{m+1}}{(m+2) (V + V_{bi})} \right]^{1/(m+2)} \sim (V + V_{bi})^{-s}, \quad (101)$$

$$s \equiv \frac{1}{m+2},$$

где Q_c — заряд на единицу площади, равный произведению величины ϵ_s и максимального электрического поля (при $x=0$). Одним из основных характеристических параметров варактора является чувствительность $s(V)$ [48]:

$$s \equiv - \frac{dC}{C} \frac{V}{dV} = - \frac{d(\log C)}{d(\log V)} = \frac{1}{m+2}. \quad (102)$$

Чем больше s , тем больше изменение емкости под действием приложенного напряжения. Для линейных переходов $m=1$ и $s=1/3$, для резких $m=0$ и $s=1/2$ и для сверхрезких $m=-1$, $m=-3/2$, $m=-5/3$ и $s=1$, $s=2$, $s=3$ соответственно. Вольт-фарадные характеристики диодов с такими переходами приведены на рис. 38, б. Как и следовало ожидать, сверхрезкий переход имеет наивысшую чувствительность и обеспечивает наибольшие изменения емкости.

Упрощенная эквивалентная схема варактора [47] приведена на рис. 39 (вставка), где C_J — емкость перехода, R_S — последовательное сопротивление, R_P — параллельное эквивалентное сопротивление, обусловленное генерационно-рекомбинационным током, диффузионным током и током поверхностной утечки. С ростом обратного смещения C_J и R_S уменьшаются, а R_P обычно возрастает. Эффективность варактора определяется его добротностью Q , которая равна отношению запасенной в нем энергии к рассеиваемой:

$$Q \approx \frac{\omega C_J R_P}{1 + \omega^2 C_J^2 R_P R_S}. \quad (103)$$

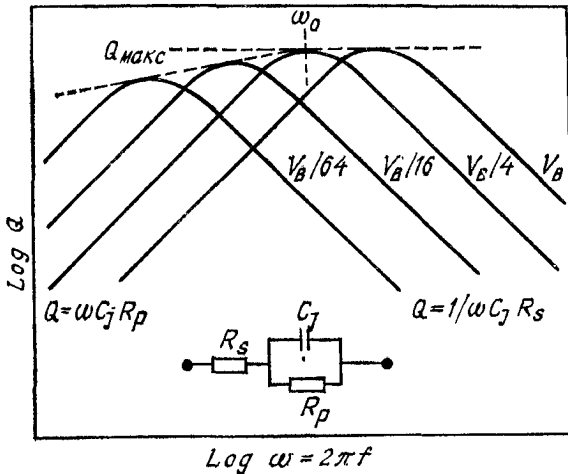


Рис. 39. Зависимость добротности варактора Q от частоты при различных смещениях. На вставке приведена эквивалентная схема варактора [47].

Дифференцируя это выражение, получим угловую частоту ω_0 , соответствующую максимальной добротности, и саму величину $Q_{\text{макс}}$:

$$\omega_0 \approx \frac{1}{C_J (R_P R_S)^{1/2}}, \quad (104)$$

$$Q_{\text{макс}} \approx \left(\frac{R_P}{4R_S} \right)^{1/2}. \quad (105)$$

На рис. 39 приведены кривые, поясняющие соотношение между Q , частотой и напряжением смещения. При фиксированном смещении добротность Q изменяется по закону $\omega C_J R_P$ на низких частотах и по закону $1/\omega C_J R_S$ на высоких частотах. Максимальное напряжение смещения ограничено напряжением пробоя V_B .

2.7.5. Диоды с быстрым восстановлением

Диоды с быстрым восстановлением используются для получения сверхвысоких скоростей переключения. Их можно разделить на два типа: диоды с диффузионным переходом и диоды с переходом металл — полупроводник. Эквивалентные схемы аналогичны схемам варактора (вставка на рис. 39), а их поведение при переключении в общих чертах отражено на рис. 36, б.

Полное время восстановления $t_1 + t_2$ диода с $p-n$ -переходом может быть существенно уменьшено путем введения центров рекомбинации, таких, какие золото создает в кремнии. Хотя время восстановления прямо пропорционально времени жизни τ

(рис. 37), к сожалению, отсутствует возможность уменьшить его до нуля путем введения максимально большого числа центров рекомбинации N_t . Это связано с возрастанием обратного тока $p - n$ -перехода, генерационная компонента которого также пропорциональна N_t (выражения (47) и (48)). В прямозонных полупроводниках, где возможны прямые переходы между зонами (таких, как GaAs), время жизни неосновных носителей обычно намного меньше, чем в Si. Поэтому диоды с $p - n$ -переходами в GaAs обладают сверхвысокой скоростью переключения и имеют время восстановления $\sim 0,1$ нс и менее. Кремниевые диоды практически позволяют получить время переключения от 1 до 5 нс.

Диоды типа металл — полупроводник (диоды Шоттки) также позволяют достичь сверхвысокой скорости переключения. Эти диоды обычно работают на основных носителях и в них эффекты накопления незначительны. Переходы металл — полупроводник рассмотрены в гл. 5.

2.7.6. Диоды с накоплением заряда

В противоположность диодам с быстрым восстановлением диоды с накоплением заряда конструируются таким образом, чтобы они накапливали заряд во время протекания прямого тока, а после переключения пропускали ток в обратном направлении за короткое время. Особый интерес среди них представляют диоды со ступенчатым восстановлением (называемые также диодами с мгновенным восстановлением). Они в течение короткого времени проводят ток в обратном направлении, а затем, как только накопленный заряд рассеется, в них происходит резкая отсечка тока. Время отсечки лежит в пикосекундном диапазоне, и поэтому фронт выключения содержит большое число гармоник. Эти диоды используются в качестве генераторов гармоник и формирователей импульсов.

Диоды с накоплением заряда в основном изготавливаются из Si с относительно большим временем жизни неосновных носителей (0,5—5 мкс), которое приблизительно в 1000 раз больше, чем у диодов с быстрым восстановлением.

2.7.7. $p - i - n$ -Диоды

В $p - i - n$ -диодах $p - n$ -переход обладает таким распределением примеси, что между его p - и n -слоями оказывается заключенным слой с проводимостью, близкой к собственной проводимости полупроводника (i -слой) (рис. 40, а). На практике, однако, идеальный i -слой заменяется либо высокоомным слоем p -типа (π -слой), либо высокоомным слоем n -типа (ν -слой). Распределения примеси, плотности объемного заряда и электрического

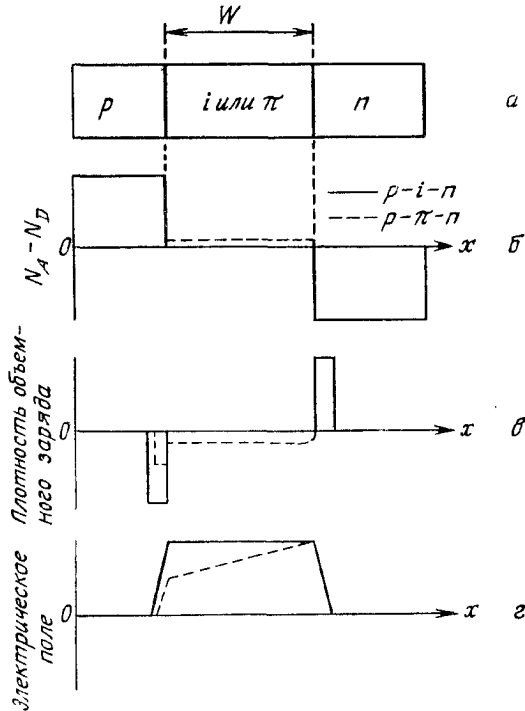


Рис. 40. Распределение примеси, плотности объемного заряда и электрического поля в $p-i-n$ - и $p-\pi-n$ -диодах [49].

поля в $p-i-n$ - и $p-\pi-n$ -диодах [49] приведены на рис. 40, б—г. Вследствие низкой концентрации примеси в i -слое на нем падает большая часть приложенного напряжения. В реальных $p-i-n$ -диодах распределение примеси в p - и n -слоях имеет более плавный характер, чем представлено на рис. 40. Такие диоды можно изготовить следующим образом: 1) эпитаксиальным методом, 2) методом диффузии p - и n -примесей с обеих сторон высокоомной полупроводниковой подложки и 3) методом ионного дрейфа (например, лития), создающего сильнокомпенсированную область с собственной проводимостью [50].

Широкое распространение $p-i-n$ -диоды получили в СВЧ-электронике. Их можно использовать в качестве СВЧ-переключателя с практически постоянной барьерной емкостью и высокой нагрузочной способностью. Время переключения составляет $\sim W/2v_s$ [51], где v_s — предельная скорость движения носителей в i -слое. Кроме того, $p-i-n$ -диод можно использовать в качестве управляемого аттенюатора, сопротивление которого почти линейно зависит от прямого тока. Он пригоден также для моду-

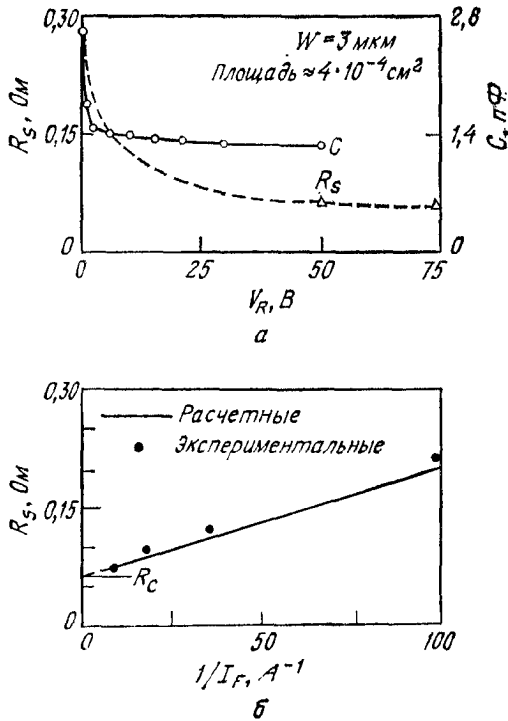


Рис. 41. Зависимость барьерной емкости и последовательного сопротивления от обратного напряжения (а) и зависимость последовательного сопротивления от прямого тока (б) [52].

ляции на частотах вплоть до гигагерцового диапазона. Следует отметить, что прямые характеристики тиристора (гл. 4) в включенном состоянии очень похожи на характеристики $p-i-n$ -диода.

На рис. 32 приведены значения напряжения пробоя $p-i-n$ -диода при обратных смещениях. Поскольку максимальное поле \mathcal{E}_m в Si с низкой концентрацией примеси составляет $2,5 \cdot 10^5$ В/см, то напряжение пробоя

$$V_B \approx \mathcal{E}_m W \approx 25W \text{ [В]}, \quad (106)$$

где W — ширина обедненного слоя, мкм. Результаты измерений емкости и последовательного сопротивления обратносмещенного $p-i-n$ -диода, предназначенного для переключения малых мощностей, представлены на рис. 41 [52]. Отметим, что емкость достигает величины, равной $\epsilon_s A/W$ (A — площадь перехода) уже при смещении 5 В, которое намного ниже напряжения пробоя, составляющего 75 В. При обратном смещении выше 10 В

емкость дополнительно уменьшается только на 3 %. Последовательное сопротивление R_s содержит две составляющие:

$$R_s = R_i + R_c, \quad (107)$$

где R_i — сопротивление i -области и R_c — сопротивление контакта. С увеличением обратного смещения R_i приближается к нулевому значению и последовательное сопротивление быстро уменьшается до значения, соответствующего контактному сопротивлению.

В случае прямого смещения p^+ — i -контакт инжектирует дырки, а i — n^+ -контакт инжектирует электроны. Вначале рассмотрим протекание тока, вызванного электронно-дырочной рекомбинацией в i -области. Плотность тока определяется соотношением [38, 53]

$$J = \int_0^W qU dx, \quad (108)$$

где U — скорость рекомбинации, равная $n(x)/\tau$. Пусть концентрация электронов, инжектированных в i -область, равна n' , тогда

$$J = \frac{qn'W}{\tau_a}, \quad (109)$$

где τ_a — амбиполярное время жизни. Если предположить, что концентрация носителей в i -области примерно постоянна, то диффузионным током можно пренебречь. Поскольку концентрация инжектированных носителей намного выше концентрации примеси в i -области, то p — i — n -диод обычно работает в условиях высокого уровня инжекции, т. е. $n' \approx p' \gg n_i$, где p' — средняя концентрация дырок, инжектированных в i -область. В таком случае полный дрейфовый ток записывается в виде

$$\begin{aligned} J &= q\mu_n n' \mathcal{E}' + q\mu_p p' \mathcal{E}' = q(\mu_n + \mu_p) n' \mathcal{E}' = \\ &= \frac{q}{kT} \left(1 + \frac{1}{b}\right) qD_n n' \mathcal{E}', \end{aligned} \quad (110)$$

где $b \equiv \mu_n/\mu_p$ и \mathcal{E}' — среднее электрическое поле в i -области. Коэффициент амбиполярной диффузии в соответствии с выражением (36) равен

$$D_a = \frac{2D_n}{1+b}. \quad (111)$$

Подставляя выражение (111) в выражение (110), имеем

$$J = \frac{p}{kT} \frac{(b+1)^2}{2b} qD_a n' \mathcal{E}'. \quad (112)$$

Напряжение V_i на i -области определяется выражением

$$V_i = \mathcal{E}' W. \quad (113)$$

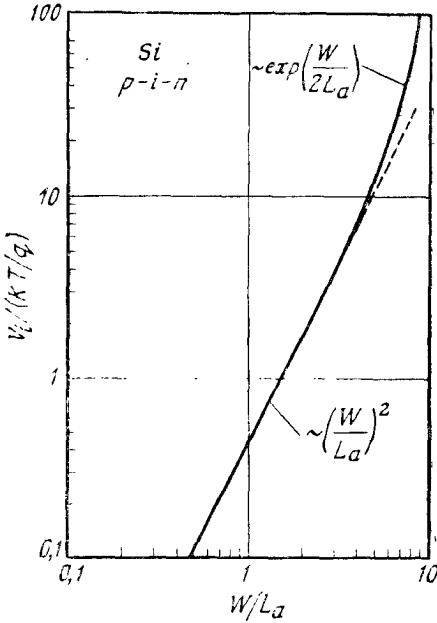


Рис. 42. Зависимость падения напряжения на внутренней области $p-i-n$ диода от W/L_a [39].

Из выражения (109) с учетом соотношения (113) и того, что $L_a = \sqrt{D_a \tau_a}$ — длина амбиполярной диффузии, имеем

$$V_i = \frac{kT}{q} \frac{2b}{(1+b)^2} \left(\frac{W}{L_a} \right)^2. \quad (114)$$

Для кремния $b \approx 3$ и

$$V_i \approx \frac{3kT}{8q} \left(\frac{W}{L_a} \right)^2. \quad (115)$$

Сопротивление R_i вычисляется по формуле

$$R_i \equiv \frac{V_i}{I_F} = \left(\frac{3kTW^2}{8qD_a\tau_a} \right) \frac{1}{I_F}. \quad (116)$$

На рис. 41, б в качестве примера показано поведение последовательного сопротивления при прямом смещении. Отметим, что оно изменяется обратно пропорционально I_F и когда $1/I_F \rightarrow 0$, то R_i асимптотически стремится к величине контактного сопротивления R_c .

Более точно величину V_i можно получить, решая уравнение (35) с соответствующими граничными условиями (рис. 42). Следует отметить, что при $W/L_a \leq 2$ напряжение V_i достаточно точно определяется выражением (115). Однако при $W/L_a \geq 2$ вели-

чина V_i начинает расти быстрее и в первом приближении удовлетворяет соотношению

$$V_i \simeq \frac{3kT}{8q} \exp\left(\frac{W}{2L_a}\right). \quad (117)$$

Таким образом, условие $W = 2L_a$ определяет границу между «коротким» и «длинным» $p-i-n$ -диодами. Для коротких структур ($W \leq 2L_a$) градиент концентрации носителей по i -области слабо влияет на величину напряжения (при комнатной температуре отклонение не превышает 0,05 В) и им можно пренебречь. Однако для протяженных структур падение напряжения на i -области оказывается гораздо выше. При проектировании прибора выбор толщины i -области определяется допустимым напряжением пробоя. Для сохранения работоспособности коротких структур величина L_a должна выбираться большой. В условиях высокого уровня инжекции коэффициент амбиполярной диффузии D_a уменьшается с ростом концентрации инжектированных носителей за счет эффектов взаимного рассеяния носителей. Время жизни неосновных носителей будет также снижаться из-за процессов оже-рекомбинации при высокой концентрации носителей. Следовательно, при повышении плотности тока L_a будет уменьшаться, что приведет к увеличению отношения W/L_a .

На рис. 43 представлены результаты вычислений [54] характеристик высоковольтных приборов с $W = 600$ мкм и $\tau_a = 30$ мкс. Кривая a учитывает только рекомбинацию в i -области

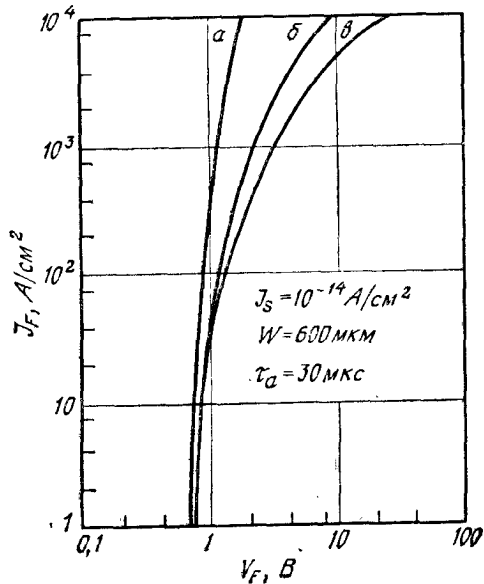


Рис. 43. Прямые вольт-амперные характеристики [54].

кривая *b* — влияние взаимного рассеяния носителей и кривая *в* — оже-рекомбинацию. При учете дополнительных факторов проявляется неуклонное ухудшение характеристик прибора. Очевидно, что оже-рекомбинация и взаимное рассеяние носителей устанавливают предел работоспособности прибора.

2.8. ГЕТЕРОПЕРЕХОДЫ

Гетеропереходом называется переход, образованный между двумя различными полупроводниками. Если полупроводники имеют одинаковый тип проводимости, то они образуют изотипный гетеропереход. Если тип их проводимости различен, то получается анизотипный гетеропереход. В 1951 г. Шокли предложил использовать резкий гетеропереход в качестве эффективного эмиттера в биполярном транзисторе [55]. В том же году были опубликованы теоретические работы Губанова по гетеропереходам [56]. Позже Кремер проанализировал аналогичный плавный гетеропереход в качестве эмиттера с широкой запрещенной зоной [57]. В дальнейшем продолжалось интенсивное изучение гетеропереходов, которые нашли широкое применение в качестве инжекционных лазеров, работающих при комнатной температуре, светодиодов, фотодетекторов и элементов солнечных батарей. Кроме того, на основе многослойной периодической структуры из гетеропереходов с толщиной слоев ~ 100 Å созданы так называемые сверхрешетки. Гетеропереходы рассмотрены в работах [58, 59, 11].

2.8.1. Принцип действия

Модель зонной структуры идеального резкого гетероперехода без ловушек на границе раздела была предложена Андерсоном [60], который использовал результаты работы Шокли. Рассмотрим эту модель, так как она позволяет хорошо объяснить большинство процессов переноса носителей в гетеропереходах, а при небольших уточнениях пригодна и для описания неидеальных случаев. На рис. 44 приведена зонная диаграмма двух изолированных полупроводников. Предполагается, что последние различаются шириной запрещенной зоны, диэлектрическими проницаемостями ϵ , работой выхода φ_m и сродством к электрону χ . Работа выхода и сродство к электрону определяются как энергия, необходимая для удаления электрона с уровня Ферми E_F и со дна зоны проводимости E_C в вакуум соответственно. Различие в положении дна зоны проводимости полупроводников обозначено ΔE_C , а различие в положении потолка валентной зоны ΔE_V . На рис. 44, *a* показан случай, когда $\Delta E_C = (\chi_1 - \chi_2)$. Соотношение $\Delta E_C = \Delta \chi$ может оказаться невыполненным. Однако если

рассматривать ΔE_C как эмпирическую величину, то модель Андерсона не потребует изменений [60а].

Зонная диаграмма анизотипного $p-n$ -гетероперехода в равновесии, образованного такими полупроводниками, приведена на рис. 44, б. Положение уровня Ферми в равновесном состоянии должно быть одинаково по обе стороны перехода, а уровень энергии, соответствующий вакууму, должен быть параллелен краям зон и непрерывен. Поэтому разрыв в положении краев зоны проводимости (ΔE_C) и краев валентной зоны (ΔE_V) не связан с уровнем легирования, если, конечно, сами величины E_g и χ не зависят от концентрации примеси (т. е. в случае невырожденных полупроводников). Полный контактный потенциал V_{bi} равен сумме потенциалов $V_{b1} + V_{b2}$, где V_{b1} и V_{b2} — электростатические потенциалы равновесного состояния первого и второго полупроводников соответственно.

Ширину обедненного слоя в каждом полупроводнике и барьерную емкость можно найти, решив уравнение Пуассона для резкого перехода с каждой стороны границы раздела. Одним из граничных условий является непрерывность электрической индукции на границе раздела, т. е. $\epsilon_1 E_1 = \epsilon_2 E_2$. В результате имеем

$$x_1 = \left[\frac{2N_{A2}\epsilon_1\epsilon_2(V_{bi} - V)}{qN_{D1}(\epsilon_1N_{D1} + \epsilon_2N_{A2})} \right]^{1/2}, \quad (118)$$

$$x_2 = \left[\frac{2N_{D1}\epsilon_1\epsilon_2(V_{bi} - V)}{qN_{A2}(\epsilon_1N_{D1} + \epsilon_2N_{A2})} \right]^{1/2}, \quad (119)$$

$$C = \left[\frac{qN_{D1}N_{A2}\epsilon_1\epsilon_2}{2(\epsilon_1N_{D1} + \epsilon_2N_{A2})(V_{bi} - V)} \right]^{1/2}. \quad (120)$$

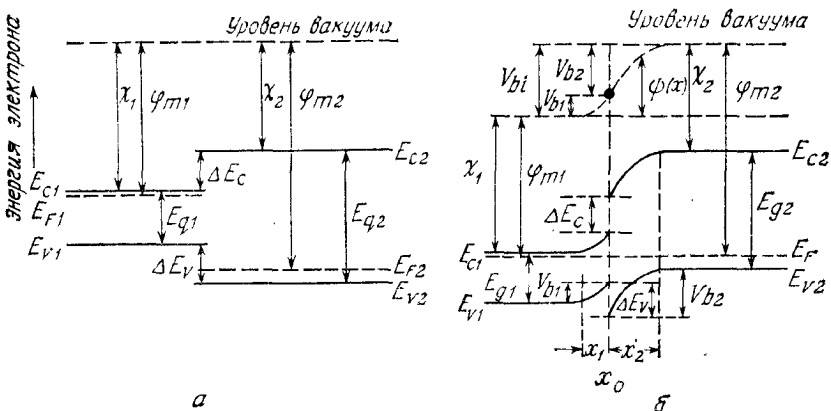


Рис. 44. Зонные диаграммы двух изолированных полупроводников при условии электронейтральности (а) и идеального анизотипного $p-n$ -гетероперехода при тепловом равновесии (б) [60].

Отношение напряжений на каждом полупроводнике составляет

$$\frac{V_{b1} - V_1}{V_{b2} - V_2} = \frac{N_{A2} \epsilon_2}{N_{D1} \epsilon_1}, \quad (121)$$

где $V = V_1 + V_2$. Очевидно, что полученные выражения преобразуются к выражению для обычного (гомогенного) $p - n$ -перехода, рассмотренного в разд. 2.3, если материалы по обе стороны гетероперехода одинаковы.

Для изотипного $n - n$ -гетероперехода из тех же двух полупроводников ситуация несколько другая. Поскольку работа выхода полупроводника с широкой запрещенной зоной меньше, то энергетические зоны изогнутся в противоположную сторону по сравнению со случаем $p - n$ -перехода¹⁾ (рис. 45, а) [61]. Отношение величин $V_{b1} - V_1$ и $V_{b2} - V_2$ можно найти из условия непрерывности электрической индукции на границе раздела. Если область 1 обогащена носителями, подчиняющимися статистике Больцмана (подробный анализ проведен в разд. 7.2), то для электрической индукции \mathcal{D}_1 в точке x_0 справедлива формула

$$\mathcal{D}_1 = \epsilon_1 \mathcal{E}_1(x_0) = \left\{ 2\epsilon_1 q N_{D1} \left[\frac{kT}{q} \left(\exp \frac{q(V_{b1} - V_1)}{kT} - 1 \right) - (V_{b1} - V_1) \right] \right\}^{1/2}, \quad (122)$$

Электрическая индукция на границе раздела для обедненной области 2 определяется выражением

$$\mathcal{D}_2 = \epsilon_2 \mathcal{E}_2(x_0) = [2\epsilon_2 q N_{D2} (V_{b2} - V_2)]^{1/2}. \quad (123)$$

Из равенства правых частей выражений (122) и (123) находим отношение величин $(V_{b1} - V_1)$ и $(V_{b2} - V_2)$, имеющее весьма сложный вид. Однако если отношение $\epsilon_1 N_{D1} / \epsilon_2 N_{D2} \approx 1$ и $V_{bi} (\equiv V_{b1} + V_{b2}) \gg kT/q$, то имеем [61]

$$\exp \left[\frac{q(V_{b1} - V_1)}{kT} \right] \approx \frac{q}{kT} (V_{b1} - V), \quad (124)$$

где V — полное приложенное напряжение, равное $(V_1 + V_2)$. На рис. 45 приведены также идеализированные зонные диаграммы для $p - n$ - (материал p -типа с узкой запрещенной зоной и материал n -типа с широкой запрещенной зоной) и $p - p$ -гетеропереходов в состоянии равновесия.

Переходя к вольт-амперным характеристикам гетеропереходов, рассмотрим интересный случай, показанный на рис. 45, а. Про-

¹⁾ Условимся обозначать материал с меньшей шириной запрещенной зоны индексом 1.

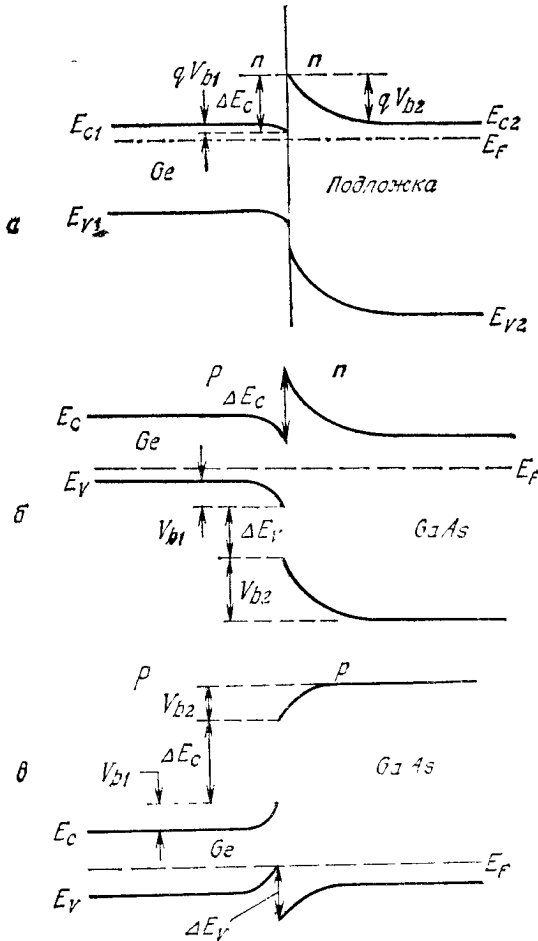


Рис. 45. Зонные диаграммы идеального изотипного $n-n$ -гетероперехода [61] (а), а также идеальных $p-n$ - (б) и $p-p$ -гетеропереходов (в) [60].

текание тока в данном случае связано с термоионной эмиссией (подробнее см. гл. 5), и плотность тока равна [61]

$$J = A^* T^2 \exp\left(-\frac{qV_2}{kT}\right) \left[\exp\left(\frac{qV_2}{kT}\right) - \exp\left(\frac{qV_1}{kT}\right) \right], \quad (125)$$

где A^* — эффективная постоянная Ричардсона. Подставляя соотношение (124) в формулу (125), получим вольт-амперную характеристику в виде

$$J = J_0 \left(1 - \frac{V}{V_{b1}}\right) \left[\exp\left(\frac{qV}{kT}\right) - 1 \right], \quad (126)$$

где

$$J_0 \equiv \frac{qA^*TV_{bi}}{k} \exp\left(-\frac{qV_{bi}}{kT}\right).$$

Приведенное выражение отличается от вольт-амперной характеристики контакта металл — полупроводник множителем J_0 , а также характером зависимости от температуры. Обратный ток не имеет насыщения, а при больших значениях V линейно возрастает с напряжением. В прямом направлении зависимость J от qV/kT допускает аппроксимацию экспоненциальной функцией, т. е. $J \sim \exp(qV/kT)$.

2.8.2. Приборы на гетеропереходах

Успешное применение гетеропереходов в различных приборах в первую очередь обязано эпитаксиальной технологии выращивания пленок, согласованных с решеткой подложки. В получаемых изотипных и анизотипных гетеропереходах практически отсутствуют состояния на границе раздела [62]. Гетеропереходы используются также в биполярных приборах в качестве эмиттеров с широкой запрещенной зоной (гл. 3) и в униполярных приборах с МОП-структурой (гл. 6). Наиболее важным является применение гетеропереходов в оптоэлектронных приборах (полупроводниковых лазерах, фотодетекторах и элементах солнечных батарей). Их характеристики подробно описаны в гл. 12—14.

В этом разделе кратко рассмотрено несколько новых конфигураций гетеропереходов, которые также могут найти применение. На рис. 46, а приведена униполярная структура выпрямителя [63], состоящего из тройного соединения $Al_xGa_{1-x}As$, заключенного между двумя слоями GaAs n -типа. За счет изменения параметра состава x в пределах 0—0,4 ширина запрещенной зоны $Al_xGa_{1-x}As$ линейно увеличивается от 1,42 до 1,92 эВ, что приводит к повышению дна зоны проводимости в такой (вариантной) структуре (рис. 46, б). При прямом смещении падение напряжения на слое с изменяющимся составом уменьшает наклон потенциального барьера и вызывает возрастание термической эмиссии над барьером (рис. 46, в). При обратном смещении электроны не могут преодолеть резкую ступеньку потенциального барьера в том месте, где пилообразная зависимость обрывается. В таких приборах выпрямляющие характеристики наблюдались в диапазоне температур 77—300 К.

Сверхрешетками называют многослойные структуры из гетеропереходов с типичной толщиной слоя 80—100 Å, а также структуры с периодическим изменением уровня легирования полупроводника, образующего последовательность гомопереходов. Методом молекулярно-лучевой эпитаксии можно получать ато-

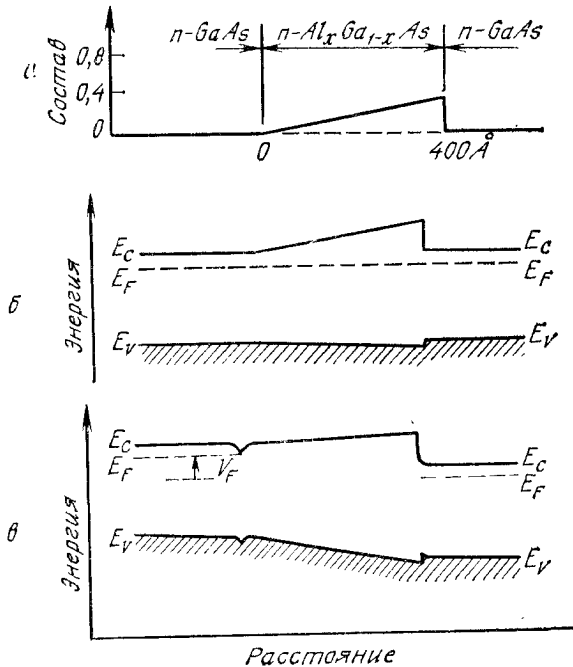


Рис. 46. Варизонная гетероструктура [63].

a — изменение состава вдоль структуры; *b* — равновесная зонная диаграмма; *в* — зонная диаграмма при прямом смещении.

марно ровные слои и управлять их толщиной. Упрощенная зонная диаграмма сверхрешетки со структурой $\text{GaAs} - \text{Al}_x\text{Ga}_{1-x}\text{As}$ [64] приведена на рис. 47, *a*. Полупроводниковые слои не легированы. Следовательно, уровень Ферми лежит вблизи середины запрещенной зоны. Для состава с $x = 0,3$ различие в ширине запрещенных зон составит ~ 300 мэВ. Степень легирования можно модулировать путем одновременного введения Al и Si (последний является донором $\text{Al}_x\text{Ga}_{1-x}\text{As}$) так, что кремний будет введен только в слои $\text{Al}_x\text{Ga}_{1-x}\text{As}$. Зонная диаграмма переменного-легированной сверхрешетки показана на рис. 47, *b*. Здесь уровень Ферми переместился ближе ко дну зоны проводимости. Так как край зоны проводимости в GaAs лежит ниже, чем донорные уровни в $\text{As}_x\text{Ga}_{1-x}\text{As}$, электроны с донорных атомов перейдут в область GaAs. В результате все подвижные носители окажутся в слоях GaAs, а ионизированные атомы донорной примеси (в слоях $\text{As}_x\text{Ga}_{1-x}\text{As}$) будут пространственно разделены. Плотность электронов в каналах GaAs намного превысит плотности нейтральных и ионизированных примесных центров рассеяния в тех же кана-

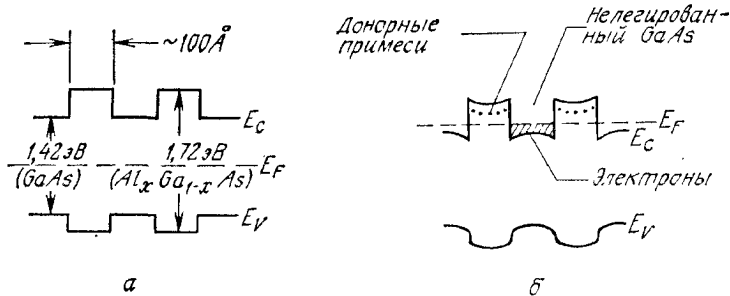


Рис. 47. Зонные диаграммы для сверхрешетки из чередующихся слоев GaAs и $Al_{0,3}Ga_{0,7}As$ (а) и для переменнo-легированной сверхрешетки (б) [64].

лах. Это приведет к существенному изменению поведения подвижности носителей при изменении температуры и плотности в том случае, когда рассеяние на примесях играет важную роль. Экспериментальные зависимости подвижности носителей в многослойной структуре от температуры приведены на рис. 48. Отметим, что в переменнo-легированной сверхрешетке подвижность существенно выше, чем в объемном материале. Если напряженно приложено перпендикулярно многослойной структуре, может наступить резонансное туннелирование, приводящее к туннельному току при напряжениях, соответствующих квазистационарным состояниям

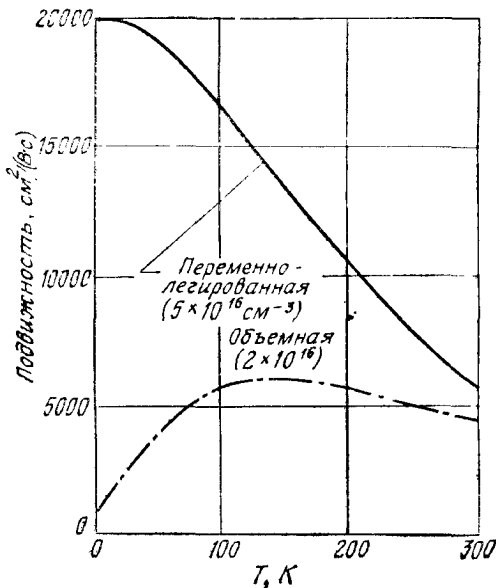


Рис. 48. Температурная зависимость подвижности в GaAs и в переменнo-легированной сверхрешетке [64].

в потенциальных ямах сверхрешетки. Эти свойства открывают широкие возможности для создания новых полезных приборов [64—66].

ЛИТЕРАТУРА

1. Shockley W. The Theory of $p - n$ Junctions in Semiconductors and $p - n$ Junction Transistors, *Bell Syst. Tech. J.*, **28**, 435 (1949); Electrons and Holes in Semiconductors, D. Van Nostrand, Princeton, N. J., 1950.
2. Sah C. T., Noyce R. N., Shockley W. Carrier Generation and Recombination in $p - n$ Junction and $p - n$ Junction Characteristics, *Proc. IRE*, **45**, 1228 (1957).
3. Moll J. L. The Evolution of the Theory of the Current — Voltage Characteristics of $p - n$ Junctions, *Proc. IRE*, **46**, 1076 (1958).
4. For example, see A. G. Grove, Physics and Technology of Semiconductor Devices, Wiley, N. Y., 1967.
5. Hall R. N., Dunlap W. C. $p - n$ Junctions Prepared by Impurity Diffusion, *Phys. Rev.*, **80**, 467 (1950).
6. Tanenbaum M., Thomas D. E. Diffused Emitter and Base Silicon Transistors, *Bell Syst. Tech. J.*, **35**, 1 (1956).
7. Frosch C. J., Derrick L. Surface Protection and Selective Masking during Diffusion in Silicon, *J. Electrochem. Soc.*, **104**, 547 (1957).
8. Hoerni J. A. Planar Silicon Transistor and Diodes, IRE Electron Devices Meet., Washington, D. C., 1960.
9. Theurer H. C., Kleimack J. J., Loar H. H., Christenson H. Epitaxial Diffused Transistors, *Proc. IRE*, **48**, 1642 (1960).
10. For a review, see, for example, Pickar K. A. Ion Implantation in Silicon-Physics, Processing and Microelectronic Devices, in R. Wolfe, ed., Applied Solid State Science, Vol. 5, Academic, N. Y., 1975.
11. Casey H. C., Jr., Panish M. B. Heterostructure Lasers, Academic, N. Y., 1978.
12. Cho A. Y. Recent Developments in Molecular Beam Epitaxy, *J. Vac. Sci. Technol.*, **16**, 275 (1979).
13. Bean J. C. Growth of Doped Silicon Layers by Molecular Beam Epitaxy, in: F. F. Y. Wang, Ed., Impurity Doping Processes in Silicon, North-Holland, Amsterdam, 1981.
14. Atalla M. M. Semiconductor Surfaces and Films; the Silicon — Silicon Dioxide System, in Gatos H., Ed. Properties of Elemental and Compound Semiconductors, Vol. 5, Interscience, N. Y., 1960, pp. 163—181.
15. Deal B. E., Grove A. S. General Relationship for the Thermal Oxidation of Silicon, *J. Appl. Phys.*, **36**, 3770 (1965).
16. Meindl J. P., Dutton R. W., Saraswat K. C., Plummer J. D., Kamins T. I., Deal B. E. Silicon Epitaxy and Oxidation, in F. Van de Wiele, W. L. Engl, P. O. Jespers, Eds., Process and Device Modeling for Integrated Circuit Design, Noordhoff, Leyden, 1977.
17. For a general reference, see H. S. Carslaw, J. C. Jaeger, Conduction Heat in Solids, 2nd ed., Oxford University Press, London, 1959.
18. Fair R. B. Concentration Profiles of Diffused Dopants in Silicon, in: F. F. Y. Wang, Ed. Impurity Doping Processes in Silicon, North-Holland, Amsterdam, 1981.
19. Burger R. M., Donovan R. P., Eds., Fundamentals of Silicon Integrated Device Technology, Vol. 1, Prentice-Hall, Englewood Cliffs, N. J., 1967.
20. Kendall D. L., DeVries D. B. Diffusion in Silicon, in Haberecht R. R., Kern E. L., Eds., Semiconductor Silicon, Electrochemical Society, N. Y., 1969, p. 358.

21. Trunbore F. A. Solid Solubilities of Impurity Elements in Germanium and Silicon, *Bell Syst. Tech. J.*, **39**, 205 (1960).
22. Khaibullin I. B. et al., VINITI dep. N2661 (1974).
23. Ferris S. D., Leamy H. J., Poate J. M., Eds. Laser — Solid Interaction and Laser Processing, American Institute of Physics, N. Y., 1979.
24. Lee T. P., Sze S. M. Depletion Layer Capacitance of Cylindrical and Spherical $p - n$ Junctions, *Solid State Electron.*, **10**, 1105 (1967).
25. Sze S. M., Gibbons G. Effect of Junction Curvature on Breakdown Voltages in Semiconductors, *Solid State Electron.*, **9**, 831 (1966).
26. Garrett G. G. B., Brattain W. H. Physical Theory of Semiconductor Surfaces, *Phys. Rev.*, **99**, 376 (1955); Kittel C., Kroemer H., Thermal Physics, 2nd ed., Freeman W. H. and Co., San Francisco, 1980.
27. Redfield D. Revised Model of Asymmetric $p - n$ Junctions, *Appl. Phys. Lett.*, **35**, 182 (1979).
28. Johnson W. C., Panosis P. T. The Influence of Debye Length on the $C - V$ Measurement of Doping Profiles, *IEEE Trans. Electron. Devices*, **ED-18**, 965 (1971).
29. Chawla B. R., Gummel H. K. Transition Region Capacitance of Diffused $p - n$ Junctions, *IEEE Trans. Electron Devices*, **ED-18**, 178 (1971).
30. Gummel H. K. Hole — Electron Product of $p - n$ Junctions, *Solid State Electron*, **10**, 209 (1967).
31. For a general discussion, see Moll J. L. Physics of Semiconductors, McGraw-Hill, N. Y., 1964.
32. Strutt M. J. O. Semiconductor Devices, Vol. 1, Semiconductor and Semiconductor Diodes, Academic, N. Y., 1966, Chap. 2.
33. Schiff L. J. Quantum Mechanics, 2nd ed., McGraw-Hill, N. Y. 1955.
34. Lundberg P. J., private communication.
35. Sze S. M., Gibbons G. Avalanche Breakdown Voltages of Abrupt and Linearly Graded $p - n$ Junctions in Ge, Si, GaAs, and GaP, *Appl. Phys. Lett.*, **8**, 111 (1966).
36. Warner R. M., Jr., Avalanche Breakdown in Silicon Diffused Junctions, *Solid State Electron.*, **15**, 1303 (1972).
37. Lee M. H., Sze S. M. Orientation Dependence of Breakdown Voltage in GaAs, *Solid State Electron.*, **23**, 1007 (1980).
38. Waldhauser F., private communication.
39. Ghandhi S. K. Semiconductor Power Devices, Wiley, N. Y., 1977.
40. Crowell C. R., Sze S. M. Temperature Dependence of Avalanche Multiplication in Semiconductors, *Appl. Phys. Lett.*, **9**, 242 (1966).
41. Chang C. Y., Chiu S. S., Hsu L. P. Temperature Dependence of Breakdown Voltage in Silicon Abrupt $p - n$ Junctions, *IEEE Trans. Electron Devices*, **ED-18**, 301 (1971).
42. Goetzberger A., McDonal B., Haiz R. H., Scarlet R. M. Avalanche Effects in Silicon $p - n$ Junction. II. Structurally Perfect Junctions, *J. Appl. Phys.*, **34**, 1591 (1963).
43. Kingston R. H. Switching Time in Junction Diodes and Junction Transistors, *Proc. IRE*, **42**, 829 (1954).
44. Van der Ziel A. Noise in Measurements, Wiley, N. Y., 1976.
45. Van der Ziel A., Chenette C. H. Noise in Solid State Devices, in Advances in Electronics and Electron Physics, Vol. 46, Academic, N. Y., 1978.
46. Levin J. P. Theory of Varistor Electronic Properties, *Crit. Rev. Solid State Sci.*, **5**, 597 (1975).
47. For a review, see Norwood M. H., Shatz E. Voltage Variable Capacitor Tuning — A Review, *Proc. IEEE*, **56**, 788 (1968).
48. Moline R. A., Foxhall G. F. Ion-Implanted Hyperabrupt Junction Voltage Variable Capacitors, *IEEE Trans. Electron Devices*, **ED-19**, 267 (1972).
49. Veloric H. S., Prince M. B. High Voltage Conductivity-Modulated Silicon Rectifier, *Bell Syst. Tech. J.*, **36**, 975 (1957).

50. Pell E. M. Ion Drift in an $n - p$ Junction, *J. Appl. Phys.*, **31**, 291 (1960); also Mager J. W. Characteristics of $p - i - n$ Junction Produced by Ion-Drift Techniques in Silicon, *J. Appl. Phys.*, **33**, 2894 (1962).
51. Lucovsky G., Schwartz R. F., Emmons R. B. Transit-Time Considerations in $p - i - n$ Diodes, *J. Appl. Phys.*, **35**, 622 (1964).
52. Chiang Y. S., Denlinger E. J. Low-Resistance All-Epitaxial pin Diode for Ultra-High-Frequency Applications, *RCA Rev.*, **38**, 390 (1977).
53. Hall R. N. Power Rectifiers and Transistors, *Proc. IRE*, **40**, 1512 (1952).
54. Burtcher J., Dannhauser F., Krausse J. Recombination in Thyristor and Rectifier in Silicon, *Solid State Electron.*, **18**, 35 (1975).
55. Shockley W. U. S. Patent 2569347 (1951).
56. Gubanov A. I., *Zh. Tekh. Fiz.*, **21**, 304 (1951); *Zh. Eksp. Teor. Fiz.*, **21**, 721 (1951).
57. Kroemer H. Theory of a Wide-Gap Emitter for Transistors, *Proc. IRE*, **45**, 1535 (1957).
58. Milnes A. G., Feucht D. L. Heterojunctions and Metal - Semiconductor Junctions, Academic, N. Y., 1972.
59. Sharma B. L., Purohit R. K. Semiconductor Heterojunctions, Pergamon, London, 1974.
60. Anderson R. L. Experiments on Ge - GaAs Heterojunctions, *Solid State Electron.*, **5**, 341 (1962).
- 60a. Frensley W. R., Kroemer H. Theory of the Energy-Band Lineup at an Abrupt Semiconductor Heterojunction, *Phys. Rev. B.*, **16**, 2642 (1977).
61. Chang L. L. The Conduction Properties of Ge - GaAs_{1-x}P_x $n - n$ Heterojunctions, *Solid State Electron.*, **8**, 721 (1965).
62. Lang D. V., Logan R. A. A Search for Interface States in an LPE GaAs/Al_xGa_{1-x}As Heterojunction, *Appl. Phys. Lett.*, **31**, 683 (1977).
63. Allyn C. L., Gossard A. C., Wiegmann W. New Rectifying Semiconductor Structure by Molecular Beam Epitaxy, *Appl. Phys. Lett.*, **36**, 373 (1980).
64. Dingle R., Stormer H. L., Gossard A. C., Wiegmann W. Electron Mobilities in Modulation-Doped Semiconductor Heterojunction Superlattices, *Appl. Phys. Lett.*, **33**, 665 (1978).
65. Chang L. L., Esaki L., Tsu R. Resonant Tunneling in Semiconductor Double Barriers, *Appl. Phys. Lett.*, **24**, 593 (1974).
66. Hess K., Morkoc H., Shichijo H., Streetman. Negative Differential Resistance through Real-Space Electron Transfer, *Appl. Phys. Lett.*, **35**, 469 (1979).

Глава 3

БИПОЛЯРНЫЕ ТРАНЗИСТОРЫ

3.1. ВВЕДЕНИЕ

Биполярный транзистор ¹⁾ — один из важнейших полупроводниковых приборов — был изобретен группой исследователей фирмы Bell Laboratories в 1947 г. Это был беспрецедентный случай в электронике и в полупроводниковой электронике в частности. До 1974 г. полупроводники применялись лишь как термисторы, фотодиоды и выпрямители. В 1948 г. Бардин и Брэттен создали точечно-контактный транзистор [1]. В 1949 г. Шокли опубликовал классическую работу по плоскостным диодам и транзисторам [2]. С этого момента теория транзисторов продолжала интенсивно развиваться и включала такие вопросы, как поведение транзисторов на высоких частотах, при больших мощностях и в режиме переключения. Больших успехов достигла технология транзисторов, особенно в разработке сплавных транзисторов [3], транзисторов с переходами, полученных вытягиванием из расплава [4] и зонной плавкой [5], транзисторов диффузионных [6], эпитаксиальных [7], планарных [8], с балочными выводами [9], изготовленных с помощью ионной имплантации [10], методами литографии и сухого травления [11]. Благодаря достижениям технологии была увеличена мощность, улучшены частотные свойства и повышена надежность транзисторов. Кроме того, исследования в области физики полупроводников, теории и технологии транзисторов расширили наши знания и способствовали развитию других полупроводниковых приборов.

Биполярные транзисторы в настоящее время являются важнейшими компонентами в быстродействующих вычислительных машинах, в космических кораблях и спутниках, во всех современных средствах связи и в силовых установках. По физике, проектированию и применению биполярных транзисторов опубликовано много монографий: среди них учебники Филлиса [12] и Гартнера [13], серия книг, изданных комитетом по образованию в области полупроводниковой электроники [14], и монографии Притчарда [15], Ганди [16], Мюллера и Каминса [17].

3.2. СТАТИЧЕСКИЕ ХАРАКТЕРИСТИКИ

3.2.1. Вольт-амперные характеристики

Ниже рассмотрены основные статические характеристики биполярных $p-n-p$ - и $n-p-n$ -транзисторов. На рис. 1 приведены условные обозначения и названия элементов

¹ Transistor — transfer resistor — преобразователь (трансформатор) резисторов.

$p - n - p$ - и $n - p - n$ -транзисторов. Стрелкой указано направление тока при нормальных условиях работы, т. е. при прямом смещении эмиттерного $p - n$ -перехода и обратном смещении коллекторного перехода. На рис. 2 показано включение $p - n - p$ -транзистора по схеме с общей базой, общим эмиттером и общим коллектором. Направления токов и знаки напряжений соответствуют нормальным условиям работы. Для $n - p - n$ -транзистора все направления токов и полярности напряжений необходимо изменить на противоположные. Ниже рассмотрены $p - n - p$ -транзисторы; результаты сохраняются и для $n - p - n$ -транзисторов при соответствующем изменении полярностей.

На рис. 3, а показан условный разрез $p - n - p$ -транзистора, включенного по схеме усилителя с общей базой, на рис. 3, б — профиль легирования транзистора, имеющего однородную концентрацию примеси в каждой из областей, а на рис. 3, в приведена зонная диаграмма при нормальном режиме работы транзистора.

Статические характеристики транзистора можно непосредственно получить из теории $p - n$ -перехода (гл. 2). Говоря об

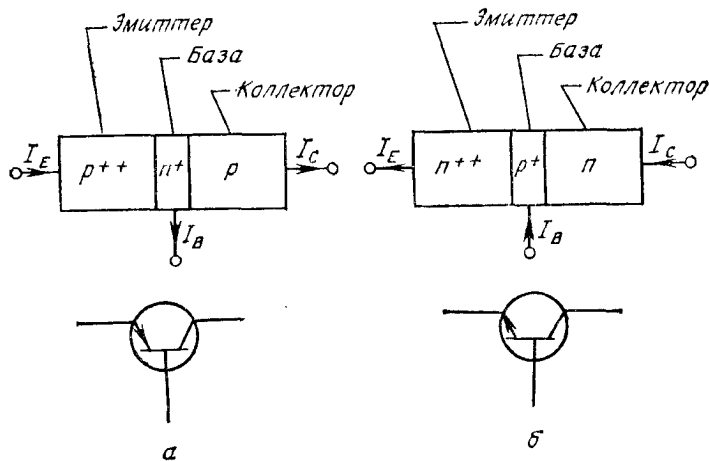


Рис. 1. Условные обозначения и названия элементов биполярного транзистора. а) $p - n - p$ -транзистор; б) $n - p - n$ -транзистор.

основных свойствах транзистора, будем считать, что вольт-амперные характеристики эмиттерного и коллекторного переходов подчиняются уравнениям идеального диода [2], т. е. можно пренебречь эффектами, обусловленными поверхностной рекомбинацией — генерацией, последовательным сопротивлением и высоким уровнем инжекции. Эти эффекты будут учтены позже.

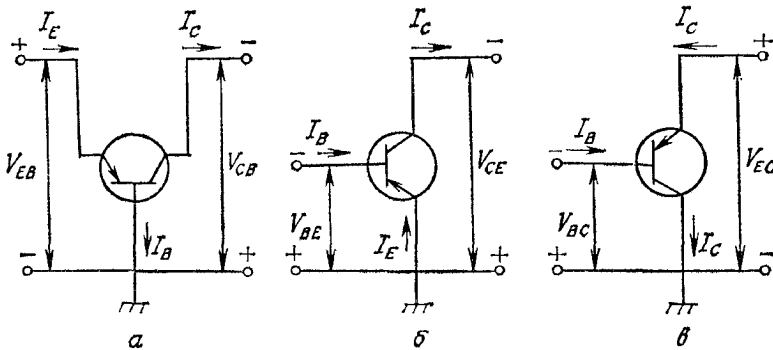


Рис. 2. Три схемы включения $p-n-p$ -транзистора.

а — схема с общей базой; б — схема в общий эмиттер; в — схема с общим коллектором.

Если весь потенциал падает на обедненной области $p-n$ -перехода (рис. 3, б), из уравнения непрерывности и уравнения для плотности токов определяются равновесные характеристики. Для нейтральной области базы уравнения имеют вид

$$0 = -\frac{p-p_B}{\tau_B} - D_B \frac{\partial^2 p}{\partial x^2}, \quad (1)$$

$$J_p = -qD_B \frac{\partial p}{\partial x}, \quad (2) \quad J_n = J_{\text{tot}} - qD_B \frac{\partial p}{\partial x}, \quad (3)$$

где p_B — равновесная плотность неосновных носителей в базе, J_{tot} — полная плотность токов проводимости, τ_B — время жизни неосновных носителей, D_B — коэффициент диффузии. Условия для концентрации избыточных носителей на границе обедненной области эмиттера можно записать следующим образом:

$$\begin{aligned} p'(0) &\equiv p(0) - p_B = p_B [\exp(qV_{EB}/kT) - 1], \\ n'(-x_E) &\equiv n(-x_E) - n_E = n_E [\exp(qV_{EB}/kT) - 1], \end{aligned} \quad (3)$$

где n_E — равновесная плотность неосновных носителей (электронов) в эмиттере. Аналогичные соотношения можно записать для коллекторного перехода:

$$\begin{aligned} p'(W) &= p(W) - p_B = p_B [\exp(qV_{CB}/kT) - 1], \\ n'(x_C) &= n(x_C) - n_C = n_C [\exp(qV_{CB}/kT) - 1]. \end{aligned} \quad (4)$$

Решения уравнения (1), описывающие распределение неосновных носителей в приборе, т. е. дырок в базе и электронов в эмиттере и коллекторе, имеют вид

$$p(x) = p_B + \left[\frac{p'(W) - p'(0)e^{-W/L_B}}{2 \operatorname{sh}(W/L_B)} \right] e^{x/L_B} - \left[\frac{p'(W) - p'(0)e^{W/L_B}}{2 \operatorname{sh}(W/L_B)} \right] \times e^{-x/L_B}, \quad (5)$$

$$n(x) = n_E + n'(-x_E) \exp\{(x + x_E)/L_E\}, \quad x < -x_E, \quad (6)$$

$$n(x) = n_C + n'(x_C) \exp[-(x - x_C)/L_C], \quad x > x_C, \quad (7)$$

где $L_B = \sqrt{\tau_B D_B}$ — диффузионная длина дырок в базе, L_E и L_C — диффузионные длины электронов в эмиттере и коллекторе. Особенное значение имеет выражение (5), так как оно связывает ширину базы W с распределением неосновных носителей. Если $W \rightarrow \infty$ или $W/L_B \gg 1$, выражение (5) сводится к

$$p(x) = p_B + p(0) e^{-x/L_B}, \quad (8)$$

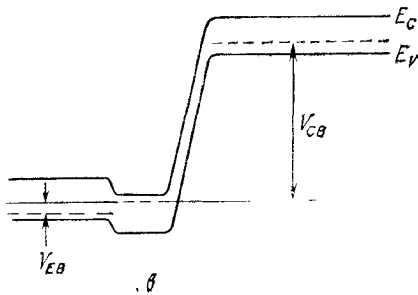
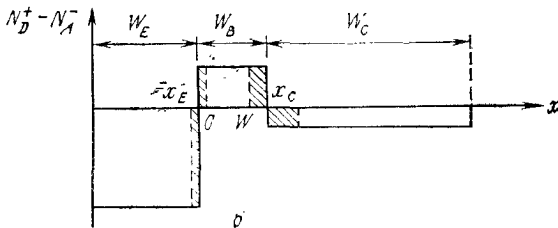
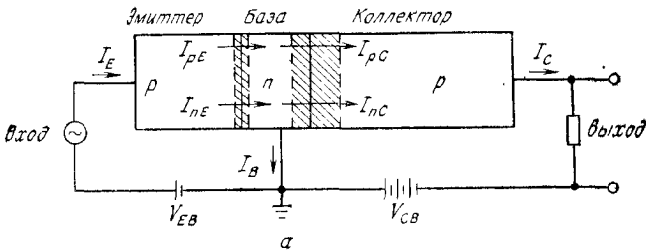


Рис. 3. Биполярный транзистор $p-n-p$ -типа, включенный по схеме с общей базой (а), профиль легирования транзистора со ступенчатым распределением примесей (б) и зонная диаграмма при нормальной работе (в).

т. е. распределение носителей в базе аналогично распределению в $p-n$ -переходе. В этом случае отсутствует связь между эмиттерным и коллекторным токами, которые определяются градиентами плотности носителей в точках $x = 0$ и $x = W$ соответственно. Таким образом, здесь отсутствует «транзисторный» эффект. Из формул (2) и (3) можно получить зависимость полного эмиттерного тока от приложенного напряжения

$$I_E = AJ_p(x=0) + AJ_n(x=-x_E) = A \left(-qD_B \frac{\partial p}{\partial x} \Big|_{x=0} \right) + \\ + A \left(-qD_E \frac{\partial n}{\partial x} \Big|_{x=-x_E} \right) = Aq \frac{D_B p_B}{L_B} \operatorname{cth} \left(\frac{W}{L_B} \right) \left[(e^{qV_{EB}/kT} - 1) - \right. \\ \left. - \frac{1}{\operatorname{sh}(W/L_B)} (e^{qV_{CB}/kT} - 1) \right] + Aq \frac{D_E n_E}{L_E} (e^{qV_{EB}/kT} - 1) \quad (9)$$

и зависимость полного коллекторного тока

$$I_C = AJ_p(x=W) + AJ_n(x=x_C) = A \left(-qD_B \frac{\partial p}{\partial x} \Big|_{x=W} \right) + \\ + A \left(-qD_C \frac{\partial n}{\partial x} \Big|_{x=x_C} \right) = Aq \frac{D_B p_B}{L_B} \frac{1}{\operatorname{sh}(W/L_B)} \left[(e^{qV_{EB}/kT} - 1) - \right. \\ \left. - \operatorname{cth} \left(\frac{W}{L_B} \right) (e^{qV_{CB}/kT} - 1) \right] - Aq \frac{D_C n_C}{L_C} (e^{qV_{CB}/kT} - 1), \quad (10)$$

где A — площадь поперечного сечения транзистора. Разность этих двух токов невелика и равна току базы

$$I_B = I_E - I_C. \quad (11)$$

Изменим теперь распределение примеси в базе (рис. 3, б) и рассмотрим более общее распределение примеси [19], показанное на рис. 4. Транзистор с подобным распределением примеси называют дрейфовым транзистором, так как в его базу встроено электрическое поле, ускоряющее дрейф дырок. Концентрация доноров N и плотность электронов в базе при $N \gg n_i$ определяются выражением

$$n \approx N = n_i \exp [q(\psi - \varphi)/kT], \quad (12)$$

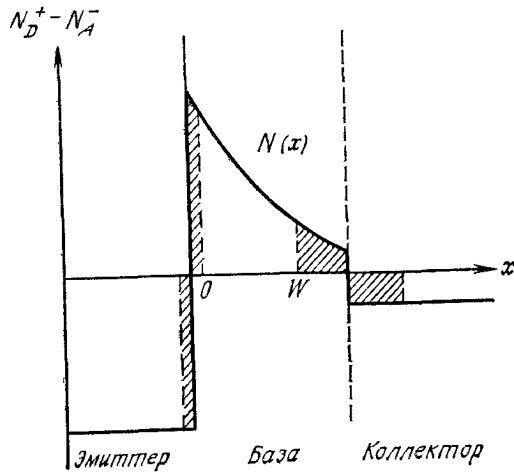
где n_i — концентрация носителей в собственном полупроводнике, φ — потенциал Ферми и ψ — потенциал Ферми в собственном полупроводнике. Из выражения (12) получим величину встроенного электрического поля:

$$\mathcal{E} \equiv -\frac{\partial \psi}{\partial x} = -\frac{kT}{q} \frac{1}{N} \frac{dN}{dx}. \quad (13)$$

Плотность дырочного тока определяется выражением

$$J_p = q\mu_{BP}\mathcal{E} - qD_B \frac{dp}{dx}. \quad (14)$$

Рис. 4. Профиль легирования транзистора с градиентом концентрации примеси в базе [19].



Подставляя выражение (13) в выражение (14), получим

$$J_p = -qD_B \left(\frac{p}{N} \frac{dN}{dx} + \frac{dp}{dx} \right). \quad (15)$$

Решение уравнения (15) для равновесного состояния с граничными условиями $p = 0$ при $x = W$ имеет вид

$$p = \frac{J_p}{qD_B} \frac{1}{N(x)} \int_x^W N(x) dx. \quad (16)$$

Концентрация дырок при $x = 0$ записывается в виде

$$p(x=0) = \frac{J_p}{qD_B} \frac{1}{n_{B0}} \int_0^W N(x) dx \simeq p_{B0} \exp\left(\frac{qV_{EB}}{kT}\right), \quad (17)$$

где n_{B0} определяется из концентрации доноров при $x = 0$, а p_{B0} — равновесная концентрация дырок при $x = 0$ (поэтому $n_{B0}p_{B0} = n_i^2$). Ток $I_p = AJ_p$ (где A — площадь) равен

$$I_p = \frac{qAD_B n_i^2}{W} \exp\left(\frac{qV_{EB}}{kT}\right) = I_1 \exp\left(\frac{qV_{EB}}{kT}\right). \quad (18)$$

Полный ток коллектора

$$I_C = I_1 \exp\left(\frac{qV_{EB}}{kT}\right) + I_2, \quad (19)$$

где I_2 — ток насыщения. Типичные экспериментальные результаты приведены на рис. 5 [20]. Отметим, что экспоненциальный

закон (выражение (19)) хорошо выполняется почти во всем диапазоне токов, за исключением очень высоких токов, при которых плотность инжектированных носителей сравнима или превосходит концентрацию примесей в коллекторе. Постоянная I_1 находится путем экстраполяции тока при $V_{EB} = 0$. Количество примесей на единицу площади базы (так называемое число Гуммеля [21]) можно получить из выражения

$$Q_b \equiv \int_0^w N(x) dx = \frac{q}{I_1} ADV_{EB}^2. \quad (20)$$

Для кремниевых биполярных транзисторов число Гуммеля лежит в диапазоне $10^{12} - 10^{13} \text{ см}^{-2}$.

На рис. 5 приведена также типичная характеристика базового тока, где можно выделить четыре участка: 1) область малых токов,

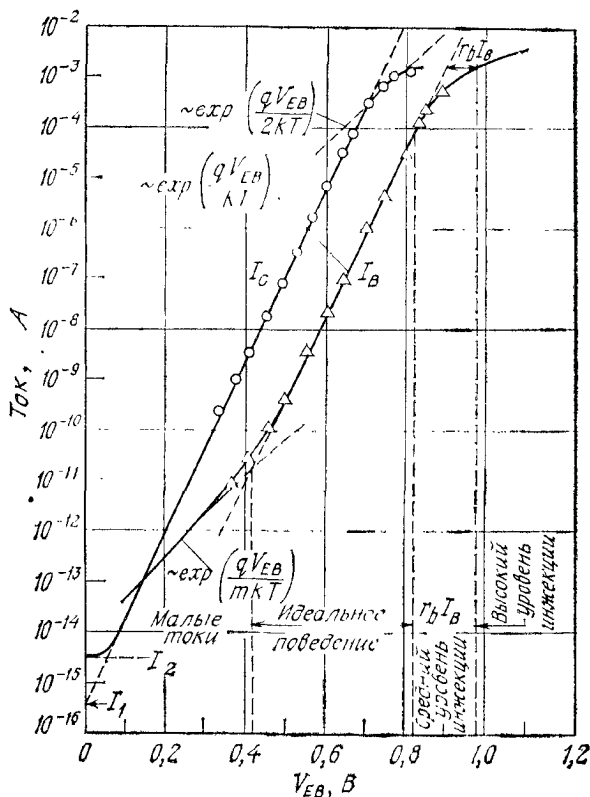


Рис. 5. Зависимость коллекторного и базового тока от напряжения эмиттер — база [20].

где базовый ток изменяется по закону $\exp(qV_{EB}/mkT)$ с $m \sim 2$; 2) область идеального поведения; 3) область среднего уровня инжекции, отличающаяся значительным падением напряжения на сопротивлении базы; 4) область высокого уровня инжекции. Для улучшения характеристик в области малых токов необходимо уменьшить плотность ловушек в обедненной области и на поверхности полупроводника. Для уменьшения сопротивления базы и ослабления эффектов, обусловленных высоким уровнем инжекции, необходимо изменить профиль легирования базы и конструкцию самого транзистора.

3.2.2. Коэффициент усиления тока

Когда $p-n-p$ -транзистор смещен в активную область, как показано на рис. 3, а, эмиттерный ток состоит из двух компонент дырочного тока $I_{pE} = AJ_p(x=0)$, инжектируемого в базу, и электронного тока $I_{nE} = AJ_n(x=x_E)$, инжектируемого из базы в область эмиттера. Коллекторный ток также содержит две компоненты: дырочный ток $I_{pC} = AJ_p(x=W)$ и электронный ток $I_{nC} = AJ_n(x=x_C)$. Величины указанных токов определяются выражениями (9) и (10).

Коэффициент усиления по току в схеме с общей базой α_0 , обозначаемый в гибридной системе параметров четырехполюсника как h_{FB} (где индексы F и B есть начальные буквы слов forward — прямо и base — база соответственно), определяется следующим образом:

$$\alpha_0 \equiv h_{FB} = \frac{\partial I_C}{\partial I_E} = \frac{\partial I_{pE}}{\partial I_E} \frac{\partial I_{pC}}{\partial I_{pE}} \frac{\partial I_C}{\partial I_{pC}}. \quad (21)$$

Первый из сомножителей $\partial I_{pE}/\partial I_E$ называют эффективностью эмиттера γ , сомножитель $\partial I_{pC}/\partial I_{pE}$ — коэффициентом переноса в базе α_T , а сомножитель $\partial I_C/\partial I_{pC}$ — коэффициентом умножения коллектора M . Так как при нормальной работе транзистора смещение на переходе коллектор — база гораздо ниже напряжения пробоя, то статический коэффициент усиления по току в схеме с общей базой имеет вид

$$\alpha_0 = \gamma \alpha_T M \simeq \gamma \alpha_T. \quad (22)$$

Статический коэффициент усиления по току в схеме с общим эмиттером β_0 , обозначаемый также h_{FE} , равен

$$\beta_0 \equiv h_{FE} = \frac{\partial I_C}{\partial I_B}. \quad (23)$$

Из соотношения (11) видно, что α_0 и β_0 взаимосвязаны:

$$\beta_0 = \frac{\alpha_0}{1 - \alpha_0}. \quad (24)$$

Так как величина α_0 в биполярных транзисторах близка к 1, β_0 обычно много больше 1. Например, если $\alpha_0 = 0,99$, то $\beta_0 = 99$, а если $\alpha_0 = 0,998$, то $\beta_0 = 499$.

При нормальной работе $p-n-p$ -транзистора $V_{EB} > 0$ и $V_{CB} \ll 0$, поэтому в выражениях (9) и (10) можно пренебречь членом, содержащим V_{CB} . В этом случае справедливы следующие соотношения:

$$\begin{aligned} \gamma &= \frac{\text{Приращение дырочного тока из эмиттера}}{\text{Приращение общего эмиттерного тока}} = \\ &= \frac{\partial A J_p(x=0)}{\partial I_E} = \left[1 + \frac{n_E}{p_B} \frac{D_E}{D_B} \frac{L_B}{L_E} \operatorname{th} \left(\frac{W}{L_B} \right) \right]^{-1}, \end{aligned} \quad (25)$$

$$\begin{aligned} \alpha_T &= \frac{\text{Приращение дырочного тока, достигшего коллектора}}{\text{Приращение дырочного тока из эмиттера}} = \\ &= \frac{J_p(x=W)}{J_p(x=0)} = \frac{1}{\operatorname{ch}(W/L_B)} \approx 1 - \frac{W^2}{2L_B^2}, \end{aligned} \quad (26)$$

где γ — эффективность эмиттера и α_T — коэффициент переноса в базе. Отметим, что $\gamma < 1$ и $\alpha_T < 1$, а величины, дополняющие их до 1, пропорциональны электронному току, вытекающему из базового контакта. В биполярном транзисторе с шириной базы, меньшей $0,1L_B$, $\alpha_T > 0,995$, и коэффициент усиления по току почти полностью определяется эффективностью эмиттера. При условии, что $\alpha_T \sim 1$,

$$h_{FE} = \frac{\gamma}{1-\gamma} = \frac{p_B D_B^2 L_E}{L_E D_E L_B} \operatorname{cth} \left(\frac{W}{L_B} \right) \sim \frac{p_B}{n_E} \left(\frac{1}{W} \right) \sim \frac{N_E}{N_B W} \sim \frac{N_E}{Q_b}, \quad (27)$$

где N_B и N_E — концентрации примеси в базе и эмиттере соответственно, Q_b — число Гуммеля, определенное по формуле (20). Следовательно, для данной концентрации N_E статический коэффициент усиления по току обратно пропорционален Q_b . Эта зависимость для пассивированных транзисторов с односторонне легированными эмиттерами приведена на рис. 6 [22]. Доза ионов в базе прямо пропорциональна Q_b , и видно, что с уменьшением дозы h_{FE} возрастает.

Коэффициент усиления по току в общем случае зависит от тока коллектора. На рис. 7 приведена типичная зависимость, полученная из графика на рис. 5 по формуле (23). При очень малых токах коллектора вклад рекомбинационно-генерационного тока (так называемого тока Са — Нойса — Шокли [23]) в обедненной области эмиттера и поверхностных токов утечки может превышать полезный диффузионный ток неосновных носителей в базе. Следовательно, эффективность эмиттера оказывается низкой. Коэф-

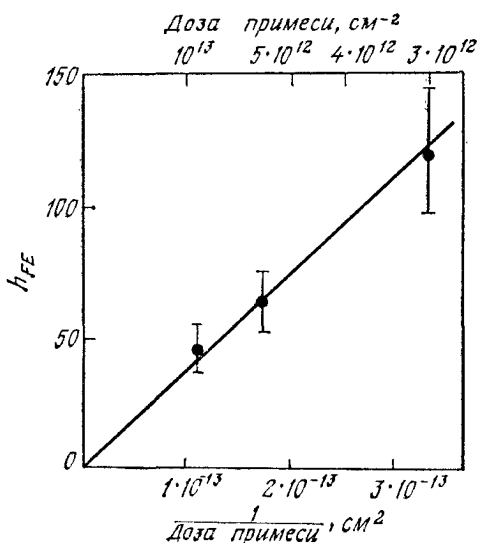


Рис. 6. Зависимость коэффициента усиления транзистора (на частоте 5 ГГц) от дозы примеси, имплантированной в базу [22].

Коэффициент усиления по току h_{FE} возрастает с током коллектора по следующему закону:

$$h_{FE} = \frac{\partial I_C}{\partial I_B} \sim \frac{e^{qV_{EB}/kT}}{e^{qV_{EB}/m kT}} = \exp \left[\frac{qV_{EB}}{kT} \left(1 - \frac{1}{m} \right) \right] \sim (I_C)^{1-1/m}. \quad (28)$$

Снижение числа объемных и поверхностных ловушек приводит к возрастанию h_{FE} при низких уровнях тока [24]. Когда величина базового тока попадает в интервал, соответствующий идеальному поведению, h_{FE} достигает области максимальных значений. При дальнейшем увеличении коллекторного тока плотность не-

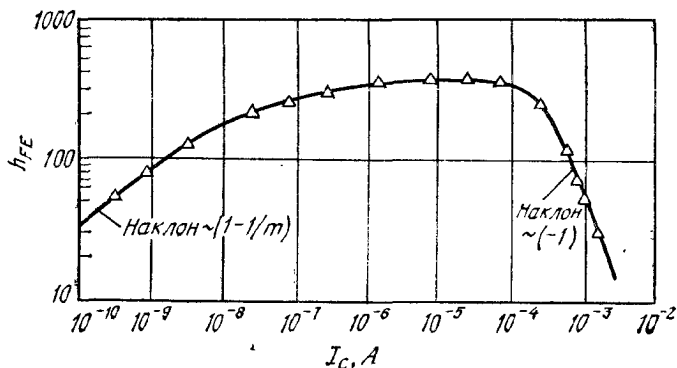


Рис. 7. Зависимость коэффициента усиления транзистора от тока коллектора.

основных носителей, инжектированных в базу, приближается к исходной плотности основных носителей (условие высокого уровня инжекции). Инжектированные носители вызывают повышение плотности основных носителей в базе, что в свою очередь приводит к снижению эффективности эмиттера. Анализ протекающих на базе процессов требует совместного решения уравнений непрерывности и уравнений для плотности токов с учетом как дрейфовой, так и диффузионной компонент. Уменьшение коэффициента усиления при возрастании I_C известно под названием эффекта Вебстера [25]. Как видно из рис. 7, при высоком уровне инжекции h_{FE} изменяется обратно пропорционально I_C :

$$h_{FE} = \frac{\partial I_C}{\partial I_E} \sim \frac{e^{qV_{EB}/kT}}{e^{qV_{EB}/kT}} \sim e^{-qV_{EB}/kT} \sim (I_C)^{-1}. \quad (29)$$

В формуле (27) кроме числа Гуммеля фигурирует еще один важный фактор — концентрация примеси в эмиттере N_E . Чтобы иметь высокий коэффициент усиления h_{FE} , степень легирования эмиттера должна быть во много раз выше, чем степень легирования базы, т. е. $N_E/N_B \gg 1$. Однако при очень высокой концентрации примеси в эмиттере начинают сказываться эффект сужения ширины запрещенной зоны и эффект оже-рекомбинации: оба вызывают уменьшение h_{FE} .

Сужение ширины запрещенной зоны в сильнолегированном кремнии, как показали исследования, связано с повышением энергии электростатического взаимодействия основных и неосновных носителей. Уменьшение ширины запрещенной зоны E_g [26] определяется выражением

$$\Delta E_g = \frac{3q^2}{16\pi\epsilon_s} \left(\frac{q^2 N_E}{\epsilon_s kT} \right)^{1/2}. \quad (30)$$

При комнатной температуре сужение зоны описывается формулой

$$\Delta E_g = 22,5 (N_E / 10^{18})^{1/2} \text{ [мэВ]}, \quad (31)$$

где N_E — концентрация примеси эмиттера, см⁻³. Экспериментальные данные, приведенные на рис. 8, хорошо согласуются с формулой (31).

При таких условиях плотность собственных носителей в эмиттере равна

$$n_{iE}^2 = N_C N_V \exp [-(E_g - \Delta E_g)/kT] = n_i^2 \exp (\Delta E_g/kT), \quad (32)$$

где N_C и N_V — плотности состояний в зоне проводимости и в валентной зоне соответственно, n_i — плотность собственных носителей в отсутствие эффекта сужения зоны. Концентрации не-

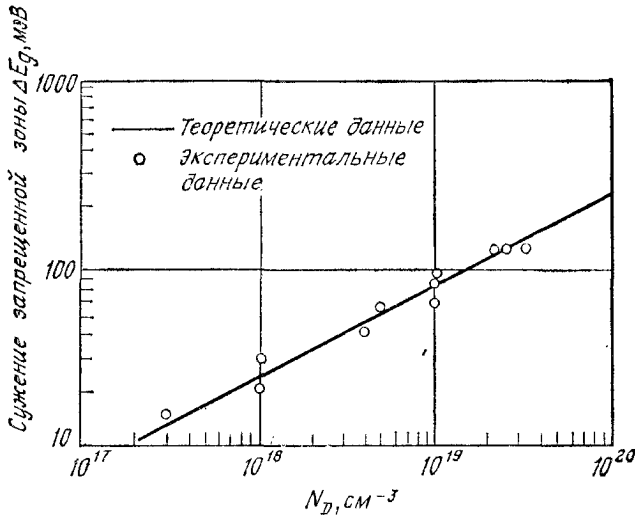


Рис. 8. Сужение ширины запрещенной зоны в кремнии [26].

основных носителей, фигурирующие в формуле (27), допускают следующую замену:

$$p_B = \frac{n_i^2}{N_B} \tag{33a}$$

и

$$n_E = \frac{n_{iE}^2}{N_E} = \frac{n_i^2}{N_E} \exp(\Delta E_g/kT). \tag{33b}$$

Следовательно,

$$h_{FE} \sim \frac{p_B}{n_E} \sim \exp(-\Delta E_g/kT). \tag{34}$$

Поскольку E_g возрастает, усиление по току падает.

Оже-рекомбинация заключается во взаимном уничтожении электрона и дырки, сопровождающейся передачей энергии другой свободной дырке [16]. Такой процесс, протекающий при участии двух дырок и одного электрона, возможен при инжекции электронов в высоколегированную p^+ -область; такой областью является эмиттер $p^+ - n - p$ -транзистора. Оже-рекомбинация — процесс, противоположный лавинному умножению. Время жизни при оже-рекомбинации $\tau = 1/G_p p^2$, где p — концентрация основных носителей, а G_p — скорость рекомбинации (равная $1-2 \times 10^{-31}$ см⁶/с для Si при комнатной температуре). Аналогично протекает рекомбинация в высоколегированной n^+ -области при участии двух электронов и одной дырки с характерным временем

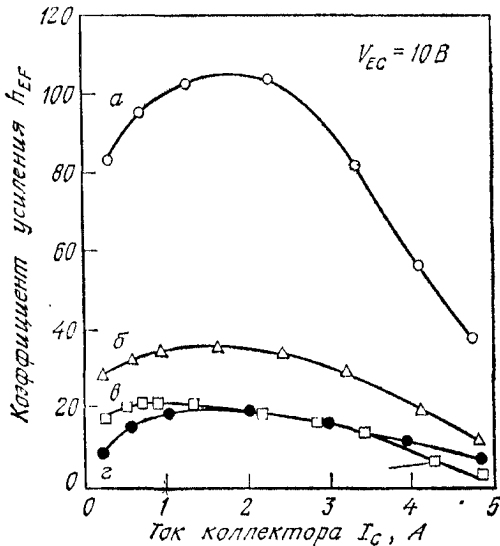


Рис. 9. Зависимость коэффициента усиления транзистора от тока коллектора [27].

жизни $\tau_A = 1/G_n n^2$. Время жизни электронов (неосновных носителей) τ в p -эмиттере определяется формулой

$$\frac{1}{\tau} = \frac{1}{\tau_p} + \frac{1}{\tau_A}, \quad (35)$$

где τ_p — время жизни, связанное с рекомбинацией типа Са — Нойса — Шокли. При возрастании концентрации носителей оже-рекомбинация становится преобладающей, вызывая уменьшение времени жизни неосновных носителей в эмиттере. В свою очередь это приводит к сокращению диффузионной длины L_E , что в соответствии с формулой (25) снижает эффективность эмиттера.

На рис. 9 приведены результаты вычисления зависимости коэффициента усиления от коллекторного тока, полученные из двумерной модели, а также результаты экспериментальных измерений [27]. Кривая *а* относится к прибору, на характеристики которого влияет только генерация Шокли — Холла — Рида (ШХР), кривая *б* отражает совместное действие ШХР и сужение запрещенной зоны, кривая *в* учитывает ШХР, сужение запрещенной зоны и оже-рекомбинацию, а кривая *г* представляет результаты измерений. Из рис. 9 следует, что для точного вычисления коэффициента усиления по току необходимо принимать во внимание оба фактора: сужение запрещенной зоны и оже-рекомбинацию. Относительный вклад каждого из трех рассмотренных выше эффектов зависит от глубины эмиттера и уровня инжекции.

В современных биполярных транзисторах со слаболегированным эпитаксиальным коллектором на коэффициент усиления

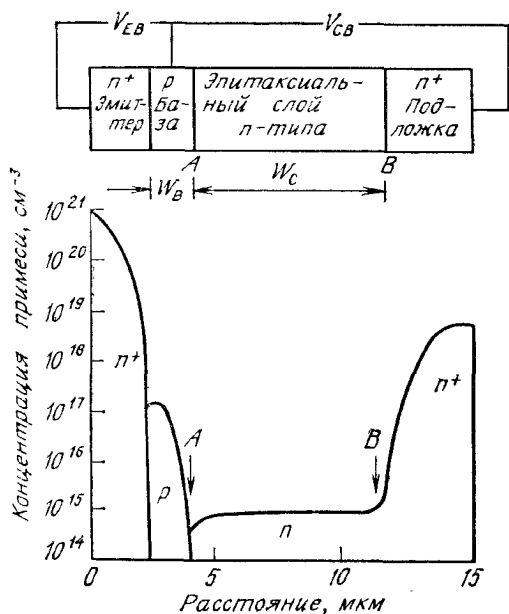


Рис. 10. Профиль легирования $n - p - n$ -транзистора с эпитаксиальным коллектором [28].

влияет перемещение под действием больших токов области с высоким электрическим полем из точки A в точку B (рис. 10) [28]. В результате эффективная ширина базы возрастает от W_B до $(W_B + W_C)$. Это явление, получившее название эффекта Кирка [29], приводит к увеличению числа Гуммеля в базе Q_b и к снижению h_{FE} . Важно отметить, что в условиях высокого уровня инжекции токи настолько велики, что в области коллектора возникают большие поля. В связи с этим классическая модель, предполагающая наличие четких границ области переноса между переходом эмиттер — база и переходом коллектор — база, непригодна. Необходимо решать численными методами полную систему уравнений в частных производных (для плотности тока, непрерывности и уравнения Пуассона) с граничными условиями, определенными только на внешних контактах. На рис. 11 приведены результаты вычисления распределения электрического поля при $|V_{CB}| = 2$ В и при различных плотностях коллекторного тока для транзистора с профилем легирования, изображенным на рис. 10. Из рисунка видно, что при повышении тока максимум электрического поля сдвигается от точки A к точке B .

Результаты, приведенные на рис. 11, свидетельствуют, что фактическая ширина базы W_{CB} при протекании тока зависит от

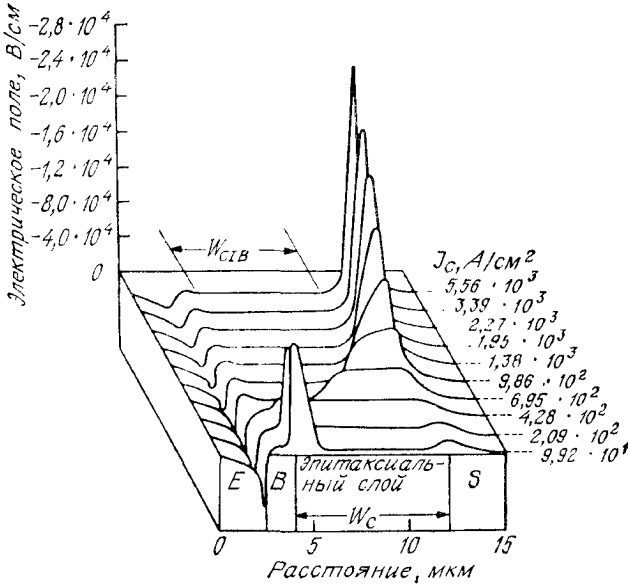


Рис. 11. Распределение электрического поля при различных плотностях коллекторного тока. Профиль легирования показан на рис. 10 [28].

концентрации примеси в коллекторе и от плотности коллекторного тока и определяется следующим выражением [16]:

$$W_{c1b} = W_c \left[1 - \left(\frac{J_1 - qv_s N_G}{J_G - qv_s N_G} \right)^{1/2} \right] \quad (36)$$

при

$$J_1 \equiv qv_s \left(N_C + \frac{2\varepsilon_s V_{CB}}{qW_c^2} \right),$$

где v_s — предельная скорость носителей (равная 10^7 см/с в кремнии при 300 К), N_C — концентрация примесей в эпитаксиальной пленке, V_{CB} — напряжение, приложенное между коллектором и базой. Как только J_C превысит J_1 , ток W_{c1b} начнет расти, а когда J_C станет много больше J_1 , величина W_{c1b} достигнет значения W_C .

3.2.3. Выходные характеристики

В разд. 3.2.2 показано, что токи на выводах транзистора связаны с распределением неосновных носителей в области базы. В случае транзистора с высокой эффективностью эмиттера в формулах для тока эмиттера и коллектора (выражения (9) и (10)) остаются только члены, пропорциональные градиенту неосновных

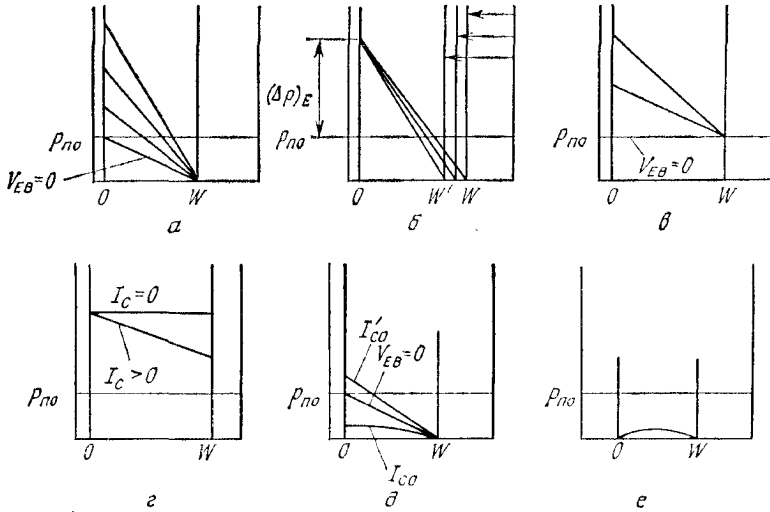


Рис. 12. Распределение плотности дырок в базе $p-n-p$ -транзистора при различных приложенных напряжениях [30].

a — нормальное включение: $V_{CB} = \text{const}$, V_{EB} — переменное; b — нормальное включение: $V_{EB} = \text{const}$, V_{CB} — переменное; c — V_{EB} — положительное, $V_{CB} = 0$; e — оба перехода смещены в прямом направлении; d — учет токов I_{C0} и I'_{C0} ; e — оба перехода смещены в обратном направлении.

носителей ($\partial p/\partial x$) при $x = 0$ и $x = W$ соответственно. Поэтому основные соотношения в транзисторе можно сформулировать следующим образом:

1. Приложенные напряжения задают плотности токов основных носителей на границах областей с помощью фактора $\exp(qV/kT)$.
2. Эмиттерный и коллекторный токи пропорциональны градиентам плотности неосновных носителей (дырок) на границах переходов, т. е. при $x = 0$ и $x = W$.
3. Базовый ток равен разности между эмиттерным и коллекторным токами.

На рис. 12 показано распределение дырок в базе $p-n-p$ -транзистора для различных величин приложенного напряжения [30]. С помощью этих графиков можно объяснить статические вольт-амперные характеристики транзисторов. Для данного транзистора эмиттерный ток I_E и коллекторный ток I_C являются функциями приложенных напряжений V_{EB} и V_{CB} , т. е. из формул (9) и (10) следует, что $I_E = f_1(V_{EB}, V_{CB})$ и $I_C = f_2(V_{EB}, V_{CB})$. Семейство выходных характеристик транзистора, включенного по схеме с общей базой и с общим эмиттером, представлено на рис. 13. В схеме с общей базой (рис. 13, a) кол-

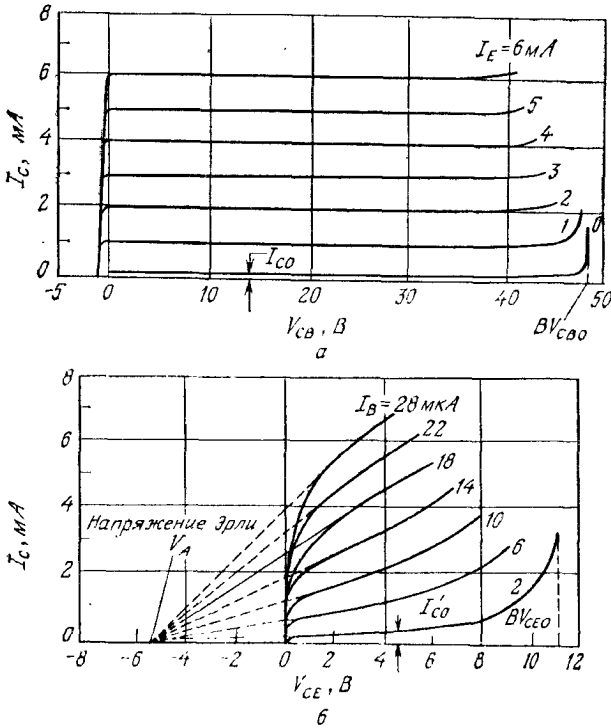


Рис. 13. Выходные характеристики $p-n-p$ -транзистора [30, 35].
 а — в схеме с общей базой; б — в схеме с общим эмиттером.

латорный ток практически равен эмиттерному току ($\alpha_0 = 1$) и фактически не зависит от V_{CE} . Коллекторный ток практически остается неизменным даже при нулевом напряжении, так как избыточные дырки продолжают извлекаться коллектором, о чем свидетельствует вид профиля распределения дырок на рис. 12, в. Для уменьшения коллекторного тока до нуля необходимо приложить к переходу коллектор — база небольшое прямое смещение ($\sim 1 \text{ В}$ для Si), которое создаст плотность дырок в базе у коллектора ($x = W$) такую же, как у эмиттера при $x = 0$ (рис. 12, г).

Обратный ток утечки коллектора I_{C0} (обозначаемый также I_{CBO}) измеряется в схеме с отключенным эмиттером. Он существенно меньше, чем обычный обратный ток $p-n$ -перехода, так как наличие эмиттерного перехода, у границы которого при $x = 0$ градиент дырок равен 0 (поскольку эмиттерный ток отсутствует), уменьшает градиент дырок при $x = W$ (рис. 12, д).

По этой же причине ток I_{CO} меньше, чем в случае, когда эмиттер коротко соединен с базой ($V_{EB} = 0$).

Если увеличивать V_{CB} вплоть до BV_{CBO} , то коллекторный ток начинает быстро возрастать. Обычно такое поведение связано с пробоем перехода коллектор — база, а величина пробивного напряжения примерно такая же, какая получена в гл. 2 для отдельного $p - n$ -перехода. При очень узкой базе или при относительно слабом ее легировании пробой может быть вызван проколом базы, т. е. с увеличением V_{CB} ширина нейтральной области базы сузится до нуля и обедненная область коллектора сомкнется с обедненной областью эмиттера. В момент смыкания коллектор оказывается коротко соединен с эмиттером, что приведет к протеканию большого тока.

Рассмотрим выходные характеристики транзистора в схеме с общим эмиттером. На рис. 13, б приведены выходные характеристики (зависимость I_C от V_{CE}) обычного $p - n - p$ -транзистора. Отметим большой коэффициент усиления по току ($h_{FE} = \partial I_C / \partial I_B$) и возрастание тока I_C с повышением напряжения V_{CE} . Обратный ток при таком включении I'_{CO} есть коллекторный ток при отсутствии тока базы (разомкнутая база), и он намного больше I_{CO} . В самом деле, из выражения (11) находим

$$I_B = I_E - I_C = I_E - (I_{CO} + \alpha_0 I_E). \quad (37)$$

Следовательно,

$$I'_E (I_B = 0) = \frac{I_{CO}}{1 - \alpha_0}. \quad (38)$$

Так как в данных условиях (рис. 12, в) эмиттерный и коллекторный токи равны ($I'_{CO} = I'_E$), то

$$I_{CEO} = I'_{CO} = \frac{I_{CO}}{1 - \alpha_0} \approx \beta_0 I_{CO} = \beta_0 I_{CBO}. \quad (39)$$

С увеличением V_{CE} ширина базы W уменьшается и наблюдается возрастание β_0 (рис. 12, б). Отсутствие насыщения выходных характеристик транзистора в схеме с общим эмиттером обусловлено значительным увеличением β_0 с ростом V_{CE} . Этот факт получил название эффекта Эрли [31]. Напряжение V_A , при котором пересекаются экстраполированные выходные характеристики, называется напряжением Эрли. В транзисторе с шириной базы W_B , много большей размера обедненной области в базе, напряжение Эрли равно

$$V_A \approx \frac{qN_B W_B^2}{\epsilon_s}. \quad (40)$$

При малых напряжениях между коллектором и эмиттером коллекторный ток быстро падает до нуля. Напряжение V_{CE} делится между двумя переходами: в результате создаются малое прямое

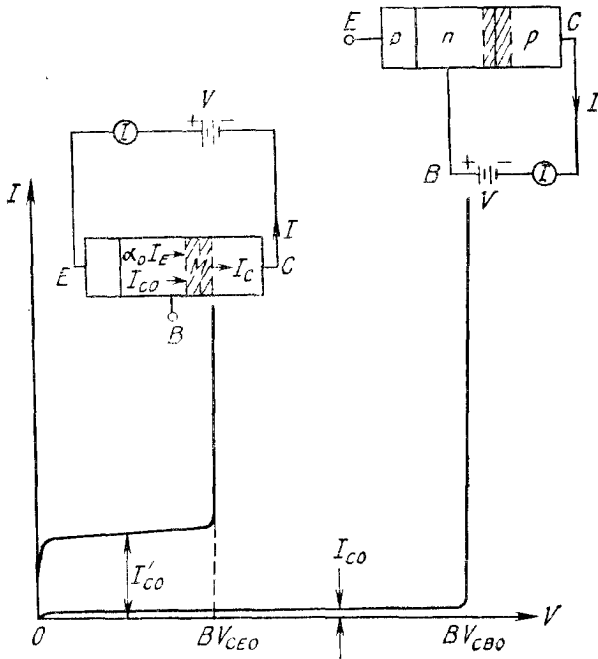


Рис. 14. Напряжение пробоя BV_{CBO} и ток насыщения I_{CO} для схемы с общей базой и соответствующие величины BV_{CEO} и I'_{CO} для схемы с общим эмиттером [13].

смещение на эмиттере и большое обратное смещение на коллекторе. Если поддерживать базовый ток постоянным, то должен оставаться постоянным и потенциал на эмиттерном переходе. Поэтому с уменьшением V_{CE} до определенной величины (~ 1 В для кремниевого транзистора) смещение на коллекторном переходе принимает нулевое значение (рис. 12, в). При дальнейшем снижении V_{CE} коллектор смещается в прямом направлении (рис. 12, г), и из-за быстрого уменьшения градиента дырок при $x = W$ происходит быстрое падение коллекторного тока.

Напряжение пробоя при разомкнутой базе можно найти следующим образом. Пусть коэффициент умножения на коллекторном переходе M определяется выражением

$$M = \frac{1}{1 - (V/BV_{CBO})^n}, \quad (41)$$

где BV_{CBO} — напряжение пробоя в схеме с общей базой, n — постоянный коэффициент. Когда базовая цепь разомкнута, $I_E =$

$= I_C = I$. Протекая через коллекторный переход, токи I_{CO} и $\alpha_0 I_E$ умножаются в M раз (рис. 14). В результате имеем

$$M(I_{CO} + \alpha_0 I) = I \quad (42)$$

или

$$I = \frac{MI_{CO}}{1 - \alpha_0 M}. \quad (43)$$

Если $\alpha_0 M = 1$, ток I ограничивается только внешним сопротивлением. Из выражения (41) и условия $\alpha_0 M = 1$ вытекает формула для определения напряжения пробоя BV_{CBO} в схеме с общим эмиттером:

$$BV_{CEO} = BV_{CBO}(1 - \alpha_0)^{1/n}. \quad (44)$$

При $\alpha_0 \approx 1$ значение BV_{CEO} намного меньше BV_{CBO} .

3.2.4. Моделирование прибора

Модель Эберса—Молла [32]. Задача моделирования заключается в установлении связи между физическими параметрами и электрическими характеристиками приборов. Моделирование особенно важно при разработке интегральных схем, когда по простым и точным моделям приборов необходимо определить поведение сложной схемы. Естественно, что чем точнее модель, тем она более сложная. Следовательно, необходимо находить компромисс между точностью и сложностью [33, 34].

Основной моделью биполярного транзистора считается модель Эберса—Молла: два диода и два источника тока включены навстречу друг другу (рис. 15, а). Источники тока управляются токами диодов, сами диоды считаются идеальными, а их характеристики имеют вид

$$I_F = I_{FO} (e^{qV_{EB}/kT} - 1), \quad (45a)$$

$$I_R = I_{RO} (e^{qV_{CB}/kT} - 1), \quad (45b)$$

где I_{FO} и I_{RO} — соответственно токи насыщения диодов, смещенных нормально в прямом и обратном направлениях. Внешние токи транзистора запишем следующим образом:

$$I_E = I_F - \alpha_I I_R, \quad (46a)$$

$$I_C = I_R - \alpha_N I_F, \quad (46b)$$

$$I_B = -(1 - \alpha_N) I_F - (1 - \alpha_I) I_R, \quad (46в)$$

где α_N и α_I — соответственно прямой и инверсный коэффициенты усиления транзистора в схеме с общей базой. Приведенные выше формулы устанавливают соотношения между внешними токами I_E

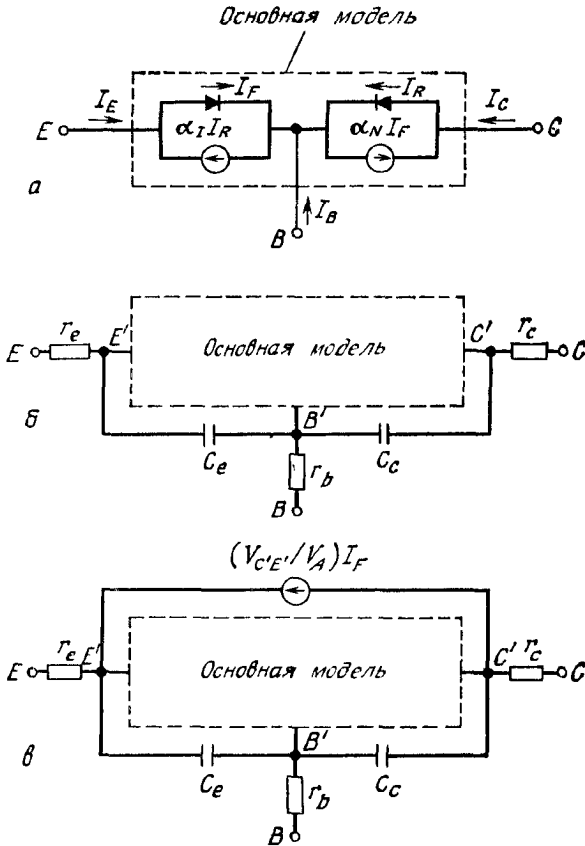


Рис 15. Эквивалентные схемы модели Эберса — Молла [32].

а — простейшая модель; б — модель с последовательными сопротивлениями и барьерными емкостями; в — модель с дополнительным источником тока для учета эффекта Эрнст

и I_C и приложенными к переходам напряжениями V_{EB} и V_{CB} . Простейшая модель содержит четыре параметра: I_{RO} , I_{FO} , α_N и α_I .

С помощью формул (9) и (10) для эмиттерного и коллекторного токов можно записать такие общие выражения:

$$I_E = a_{11}(e^{qV_{LB}/kT} - 1) + a_{12}(e^{qV_{C}/kT} - 1) \quad (47a)$$

$$I_C = a_{21}(e^{qV_{EB}/kT} - 1) + a_{22}(e^{qV_{CB}/kT} - 1). \quad (47б)$$

Из сравнения выражений (47) и (46) вытекает

$$\begin{aligned} a_{11} &= I_{FO}, & a_{12} &= -\alpha_I I_{RO}, \\ a_{21} &= -\alpha_N I_{FO}, & a_{22} &= I_{RO}. \end{aligned} \quad (48)$$

Из взаимности характеристик четырехполюсника ($a_{12} = a_{21}$) вытекает, что $a_1 I_{FO} = \alpha_N I_{FO}$. Следовательно, для основной модели требуются только три параметра.

Чтобы улучшить точность модели, в нее добавляют последовательные сопротивления и барьерные емкости [33] (рис. 15, б). Отметим, что в этом случае диоды уже управляются внутренними напряжениями на переходах $V_{E'B'}$ и $V_{C'B'}$, а не внешними приложенными напряжениями. Для учета эффекта Эрли в модель необходимо добавить источник тока между внутренним эмиттером и коллектором (рис. 15, в, где V_A — напряжение Эрли). Теперь число параметров модели выросло с 3 до 9. В основную модель могут быть дополнительно введены параметры, учитывающие зависимость α_N и α_I от плотности тока и рабочей частоты; можно дополнительно подсоединить к базе диоды, учитывающие двумерный эффект концентрации тока вдоль перехода база — эмиттер (этот эффект рассмотрен в разд. 3.4). Очевидно, что уточнение модели неизбежно приводит к увеличению числа параметров и модель транзистора становится более сложной.

Модель Гуммеля—Пуна [35]. Основана на интегральных соотношениях для зарядов и связывает внешние электрические характеристики с зарядом в базе транзистора. Это очень точная модель, объясняющая многие физические эффекты, но для ее описания требуется много параметров: так, для работы в широком диапазоне необходимо 25 параметров. Последовательное упрощение модели Гуммеля — Пуна в конце концов приводит к простейшей модели Эберса — Молла

Чтобы получить интегральные зарядовые соотношения, вначале рассмотрим уравнения для токов, выведенные в гл. 2:

$$J_n = q\mu_n n \frac{\partial \varphi_n}{\partial x}, \quad (49a)$$

$$J_p = -q\mu_p p \frac{\partial \varphi_p}{\partial x}, \quad (49б)$$

где

$$n = n_i \exp [q(\psi - \varphi_n)/kT], \quad (50a)$$

$$p = n_i \exp [q(\varphi_p - \psi)/kT]. \quad (50б)$$

Производную по координате от произведения pn можно представить в виде

$$\frac{d}{dx} (pn) = \frac{q(pn)}{kT} \left(\frac{\partial \varphi_p}{\partial x} - \frac{\partial \varphi_n}{\partial x} \right). \quad (51)$$

Выполняя интегрирование уравнения (51) по x от $x = 0$ до $x = W$ (рис. 3, б) с использованием выражений (49) и пренебрегая рекомбинацией, получим

$$(pn)_{x=0} - (pn)_{x=W} = \frac{J_{CC}}{kT} \int_0^W \frac{n(x)}{\mu_p} dx, \quad (52)$$

где J_{CC} — ток, протекающий от эмиттера к коллектору в транзисторе с единичным коэффициентом усиления. Подстановка выражений (50) в уравнение (52) дает

$$\begin{aligned} \exp [q(\varphi_p - \varphi_n)/kT] |_{x=0} - \exp [q(\varphi_p - \varphi_n)/kT] |_{x=W} = \\ = \frac{J_{CC}}{n_i^2 kT} \int_0^W \frac{n(x)}{\mu_p} dx. \end{aligned} \quad (53)$$

Предположим, что квазиуровень Ферми для электронов в базе постоянен. Тогда

$$V_{EB} = \varphi_p(0) - \varphi_n(0), \quad V_{CB} = \varphi_p(W) - \varphi_n(W). \quad (54)$$

Эти напряжения отличаются от напряжений на выводах транзистора омическими потерями. Уравнение (53) запишем в виде

$$I_{CC} = AJ_{CC} = (qn_i A)^2 D_B \frac{e^{qV_{EB}/kT} - e^{qV_{CB}/kT}}{qA \int_0^W n(x) dx}, \quad (55)$$

где A — активная площадь транзистора. Модель Гуммеля — Пуна основана на уравнении (53), связывающем напряжения на переходах, коллекторный ток и заряд в базе. Задача моделирования транзистора оказалась сведенной к моделированию заряда в базе

$$Q_B = qA \int_0^W n(x) dx, \quad (56)$$

который состоит из пяти компонент:

$$Q_B = Q_{BO} + Q_{jE} + Q_{jC} + Q_{dE} + Q_{dC}, \quad (56a)$$

где Q_{BO} — заряд в базе при нулевом смещении; Q_{jE} и Q_{jC} — заряды, связанные с барьерными емкостями эмиттера и коллектора; Q_{dE} и Q_{dC} — заряды неосновных носителей, обусловленные диффузионными емкостями эмиттера и коллектора. При увеличении уровня инжекции возрастают и диффузионные емкости, в связи с чем происходит снижение коэффициента усиления.

Выражение для тока (55) можно представить в виде

$$J_{CC} = I_F - I_R, \quad (57)$$

где

$$I_F = I_S Q_{BO} \frac{e^{qV_{EB}/kT} - 1}{Q_B}, \quad (58a)$$

$$I_R = I_S Q_{BO} \frac{e^{qV_{CB}/kT} - 1}{Q_B}. \quad (58b)$$

Отметим, что выражения (58а) и (58б) аналогичны выражениям (45а) и (45б) для модели Эберса — Молла. Заряд Q_{dE} в выражении (56) можно представить в виде $B\tau_F I_F$, где τ_F — время жизни неосновных носителей при протекании прямого тока, B — коэффициент, обычно равный 1, но из-за наличия эффекта Кирка может превышать 1. Заряд Q_{dC} можно записать в виде $\tau_R I_R$, где τ_R — время жизни неосновных носителей при протекании обратного тока.

Подставляя выражения (58а) и (58б) в уравнение (56а), получим квадратное уравнение относительно Q_B , и его решение имеет вид

$$Q_B = \frac{Q_{BO} + Q_{JE} + Q_{JC}}{2} + \left\{ \left(\frac{Q_{BO} + Q_{JE} + Q_{JC}}{2} \right)^2 + I_S Q_{BO} [B\tau_F (e^{qV_{EB}/kT} - 1) + \tau_R (e^{qV_{CB}/kT} - 1)] \right\}^{1/2}. \quad (59)$$

Ток базы записывается в виде

$$I_B = dQ_B/dt + I_{rec}, \quad (60)$$

где рекомбинационный базовый ток можно разделить на два слагаемых:

$$I_{rec} = I_{EB} + I_{CB}, \quad (61)$$

причем

$$I_{EB} = I_1 (e^{qV_{EB}/kT} - 1) + I_2 (e^{qV_{EB}/m_e kT} - 1), \quad (61a)$$

$$I_{CB} = I_3 (e^{qV_{EB}/m_c kT} - 1). \quad (61б)$$

В этих выражениях m_e и m_c — коэффициенты неидеальности характеристик эмиттера и коллектора. Для идеальных токов $m_e = m_c = 1$, и для рекомбинационно-генерационных токов в обедненной области $m_e = m_c = 2$. Полные токи эмиттера и коллектора теперь определяются выражениями

$$I_E = I_{CC} + I_{BB} + \tau_F (dI_F/dt) + C_{JE} (dV_{EB}/dt), \quad (62a)$$

$$I_C = I_{CC} - I_{CB} - \tau_R (dI_R/dt) + C_{JC} (dV_{EC}/dt). \quad (62б)$$

Эквивалентная схема модели Гуммеля — Пуна, дополненная последовательными сопротивлениями, показана на рис. 16. Поскольку заряд Q_B зависит от напряжения, модель учитывает эффект высокого уровня инжекции в базе ($\tau_F I_F$ становится больше, чем Q_{BO}). Расширение базы под действием тока (эффект Кирка) учитывается множителем B , который зависит от I_C и V_{CB} . Вклад эмиттера в базовый ток моделируется двумя параллельными диодами, один из которых идеальный, а другой имеет коэффициент неидеальности $m_e > 1$. Это приводит к зависимости коэффициента усиления от смещения при низких уровнях тока. Эффект

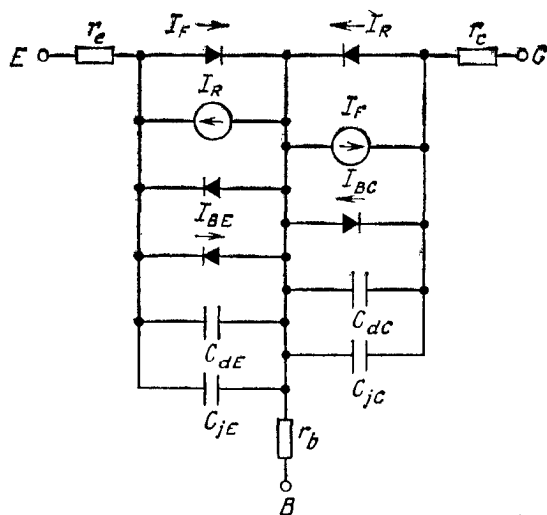


Рис. 16. Эквивалентная схема модели Гуммеля — Пуна [35].

Эрли моделируется зависимостью заряда Q_{jC} от напряжения ($Q_{jC} = C_{jC}V_{CB}$). Модель Гуммеля — Пуна обладает высокой точностью и позволяет учитывать многие физические эффекты с помощью зависимости Q_B от смещения. Однако для ее описания требуется 25 параметров. Поэтому при анализе конкретных схем приходится искать разумный компромисс между точностью решения и сложностью модели.

3.3. СВЧ-ТРАНЗИСТОРЫ

Ниже рассмотрены биполярные транзисторы, предназначенные для работы на высоких частотах (выше 1 ГГц). На рис. 17, а показан типичный кремниевый СВЧ-транзистор. Поскольку подвижность электронов в кремнии выше подвижности дырок, то все кремниевые СВЧ-транзисторы — приборы $n-p-n$ -типа. Для уменьшения последовательного коллекторного сопротивления в качестве подложки используют эпитаксиальную $n-n^+$ -структуру. На ее поверхности формируют изолирующий слой (например, термическим окислением получают слой SiO_2). Базовый и эмиттерный слой создаются диффузией или ионной имплантацией. Обычно для ослабления эффекта смещения границы эмиттера и улучшения его эффективности используют мышьяк в качестве эмиттерной примеси [36]. За счет топологии транзистора достигают необходимых токовых характеристик. Для этого варьируют число полосковых областей эмиттера и контактов к базе.

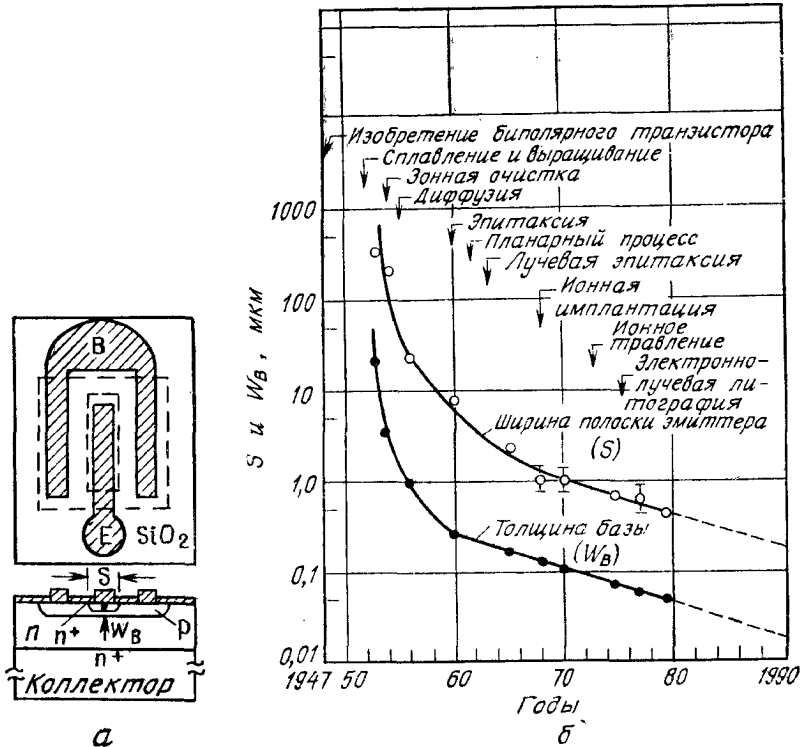


Рис. 17. Уменьшение двух критических размеров дискретных транзисторов (ширины эмиттерной полоски и толщины слоя базы), достигнутое с 1952 г. Указаны также основные этапы развития технологии. На рис. 17, а показан транзистор с полосковой геометрией электродов. В интегральных транзисторах критические размеры примерно в 10 раз больше для предотвращения закороток между эмиттером и коллектором [37, 11].

С помощью изменения профиля легирования добиваются требуемых частотных свойств и приемлемых пробивных напряжений.

Низкочастотные транзисторы отличаются от СВЧ-транзисторов размерами активных областей, значениями паразитных параметров полупроводниковой структуры и корпуса. Для улучшения высокочастотных свойств транзисторов должны быть существенно уменьшены размеры активных областей и значения паразитных параметров. В высокочастотных транзисторах определяющими критическими размерами являются ширина эмиттерной полоски S и толщина базы W_b . На рис. 17, б приведены данные, иллюстрирующие тенденцию уменьшения этих размеров со временем (начиная с 1952 г.), а также отмечены основные события, связанные с развитием технологии транзисторов [11, 37]. Уменьше-

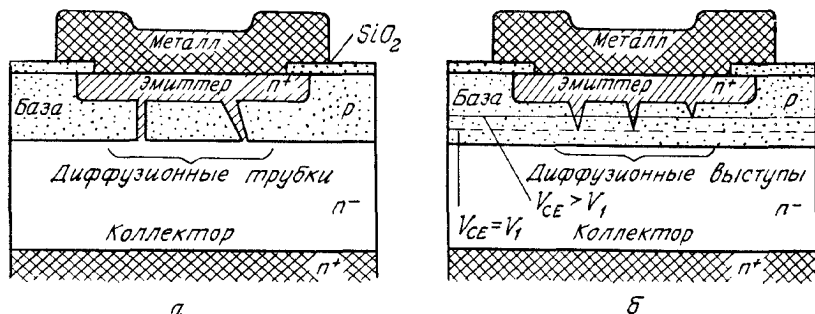


Рис. 18. Диффузионные трубки (а) и диффузионные выступы (б) в базе вдоль дислокаций [38].

ние вертикальных размеров в основном обязано развитию диффузионных процессов и ионной имплантации, в то время как уменьшение горизонтальных размеров связано с успехами литографии. В настоящее время ширина эмиттерной полоски составляет меньше 1 мкм, а толщина базы может составлять всего несколько сотен ангстрем. С уменьшением ширины базы первостепенную важность приобретает устранение возможных закороток между эмиттером и коллектором, вызванных диффузионными трубками и диффузионными выступами в базе вдоль дислокаций (рис. 18) [38]. Поэтому должны использоваться технологические процессы, исключающие образование дефектов упаковки вследствие окисления, появление дислокаций скольжения (обусловленных эпитаксиальным выращиванием) и другие нарушения, сопутствующие технологическим операциям.

3.3.1. Частота отсечки

Частота отсечки f_T является наиболее важным показателем качества СВЧ-транзисторов. Она определяется [40] как частота, на которой коэффициент усиления по току в режиме короткого замыкания схемы с общим эмиттером h_{Tc} ($\equiv \partial I_C / \partial I_B$) равен 1. Частота отсечки связана с физической структурой транзистора через время задержки носителей, пролетающих от эмиттера к коллектору, τ_{ec} :

$$f_T = \frac{1}{2\pi\tau_{ec}}. \quad (63)$$

Время задержки τ_{ec} представляет собой сумму четырех времен задержки, характеризующих последовательные фазы движения носителей от эмиттера к коллектору:

$$\tau_{ec} = \tau_E + \tau_B + \tau_C + \tau'_C. \quad (64)$$

Здесь τ_E — время зарядки обедненного слоя эмиттера, определяемое выражением

$$\tau_E = r_e(C_e + C_c + C_p) \approx \frac{kT}{qI_E}(C_e + C_c + C_p), \quad (65)$$

где r_e — сопротивление эмиттера; C_e — емкость эмиттера; C_c — емкость коллектора; C_p — другие паразитные емкости, связанные с базовым выводом; I_E — эмиттерный ток, примерно равный коллекторному току I_C . Величина r_e является производной по напряжению эмиттерного тока, определяемого соотношением (9).

Вторая составляющая времени задержки в формуле (64) представляет собой время пролета через слой базы и равна

$$\tau_B = \frac{W^2}{\eta D_B}, \quad (66)$$

где $\eta = 2$ для случая равномерного легирования базы. Формулу (66) можно получить, подставив $L_B = \sqrt{D_B \tau_B} / (1 + j\omega \tau_B)$ в выражения (25) и (26), что дает малосигнальный коэффициент усиления в схеме с общей базой [41]:

$$\alpha \approx \frac{1}{\text{ch}(W \sqrt{(1 + j\omega \tau_B) / D_B \tau_B})} \approx \frac{1}{1 + jW^2 \omega / 2D_B}. \quad (67)$$

Время задержки τ_B определяется из формулы $1/2\pi f_\alpha$, где f_α — так называемая «альфа»-частота отсечки — равна частоте, на которой усиление падает в $1/\sqrt{2}$ раз по сравнению со значением на низкой частоте. Вклад эффективности эмиттера γ во время задержки мал и в формуле (67) не учитывается. В случае неравномерного распределения примесей в базе, например, как в дрейфовом транзисторе (рис. 4), коэффициент η в формуле (66) должен быть увеличен. Если встроенное поле \mathcal{E}_{bi} постоянно, то коэффициент η принимает значение [42]

$$\eta \approx 2 \left[1 + \left(\frac{\mathcal{E}_{bi}}{\mathcal{E}_0} \right)^{3/2} \right], \quad (68)$$

где $\mathcal{E}_0 = 2D_B/\mu_B W$. При $\mathcal{E}_{bi}/\mathcal{E}_0 = 10$ $\eta \approx 60$, т. е. за счет большого встроенного поля достигается значительное снижение τ_B . Встроенное поле создается автоматически в реальных транзисторах, в которых база формируется диффузией. На рис. 10 показан типичный пример высокочастотного эпитаксиального $n-p-n$ -транзистора, полученного двойной диффузией.

Третья составляющая времени задержки связана с пролетом носителей через обедненный слой коллектора (рис. 3):

$$\tau_c = \frac{x_c - W}{2v_s}, \quad (69)$$

где v_s — предельная скорость носителей в коллекторе.

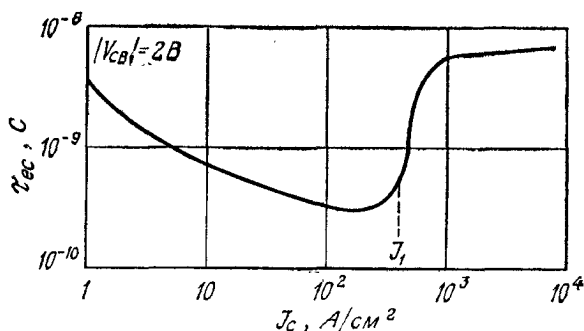


Рис. 19. Время движения носителей от эмиттера к коллектору в зависимости от плотности коллекторного тока для транзистора, изображенного на рис. 10 [28].

Четвертая компонента задержки обусловлена временем, в течение которого заряжается объединенная емкость коллектора:

$$\tau'_c = r_c C_c, \quad (70)$$

где r_c — последовательное сопротивление коллектора, C_c — емкость коллектора. В эпитаксиальном транзисторе r_c может быть существенно уменьшено и время задержки τ'_c пренебрежимо мало по сравнению с другими временами задержки.

Таким образом, выражение для частоты отсечки f_T имеет вид

$$f_T = \frac{1}{2\pi\tau_{ec}} = \left\{ 2\pi \left[\frac{kT(C_e + C_c + C_D)}{qI_C} + \frac{W^2}{\eta D_B} + \frac{x_C - W}{2v_s} \right] \right\}^{-1}. \quad (71)$$

Из выражения (71) видно, что для повышения частоты отсечки необходимо уменьшать толщину базы транзистора (один из критических размеров, показанных на рис. 17), толщину коллектора и работать при высоких плотностях тока. Однако при уменьшении толщины коллектора происходит соответствующее снижение пробивного напряжения. Следовательно, необходимо искать компромисс между высокочастотными свойствами транзистора и его способностью выдерживать высокие напряжения.

С увеличением рабочего тока частота отсечки повышается, так как время заряда эмиттера τ_E обратно пропорционально току. Вместе с тем, когда ток становится достаточно большим и плотность инжектированных неосновных носителей сравнивается или превышает концентрацию примеси в базе, эффективная толщина базы возрастает от W_B до $(W_B + W_C)$ (разд. 3.2). На рис. 19 приведены теоретические значения времени задержки носителей при движении от эмиттера к коллектору τ_{ec} для транзистора, показанного на рис. 10 [28]. При низких плотностях токов τ_{ec}

падает с ростом J_C , как следует из формулы (71), и коллекторный ток переносится в основном за счет дрейфа, т. е.

$$J_C \approx q\mu_C N_C \mathcal{E}_C, \quad (72)$$

где μ_C , N_C и \mathcal{E}_C — подвижность, концентрация примеси и электрическое поле в коллекторном эпитаксиальном слое соответственно. При дальнейшем нарастании тока время задержки принимает минимальное значение, а затем начинает возрастать, особенно быстро при токе J_1 . Этому току соответствует максимальное однородное электрическое поле $\mathcal{E}_C = (V_{CO} + |V_{CB}|)/W_C$, где V_{CO} — контактный потенциал коллектора, V_{CB} — напряжение, приложенное между базой и коллектором. Токи, превышающие J_1 , не могут уже переноситься через эпитаксиальную область коллектора только за счет дрейфовой компоненты. Величина J_1 определяется выражением (72) и равна

$$J_1 = q\mu_C N_C (V_{CO} + |V_{CB}|)/W_C. \quad (73)$$

В результате эффекта Кирка этот ток является оптимальным с точки зрения максимальной частоты отсечки. Следует отметить, что с увеличением V_{CB} одновременно возрастает и величина J_1 .

3.3.2. Высокочастотные характеристики

Для описания высокочастотных характеристик транзисторов широко используют матрицу рассеяния (*s*-параметры), так как на высоких частотах их измерять проще, чем другие системы параметров [43, 44]. На рис. 20 показан обобщенный четырехполюсник с падающими (a_1 , a_2) и отраженными (b_1 , b_2) волновыми сигналами, используемыми для определения *s*-параметров. Четырехполюсник описывается линейными соотношениями

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad (74)$$

где параметры s_{11} , s_{22} , s_{12} и s_{21} имеют следующее значение:

$s_{11} = \left. \frac{b_1}{a_1} \right|_{a_2=0}$ — коэффициент отражения на входе при согласованной нагрузке на выходе ($Z_L = Z_0$ означает, что $a_2 = 0$ (Z_0 — внутренний импеданс));

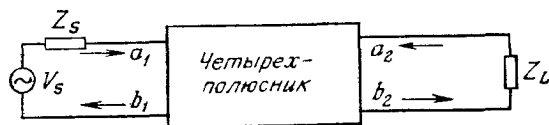


Рис. 20. Четырехполюсник с падающими (a_1 , a_2) и отраженными (b_1 , b_2) волновыми сигналами, используемыми для определения *s*-параметров.

$s_{22} = \frac{b_2}{a_2} \Big|_{a_1=0}$ — коэффициент отражения на выходе при согласованной нагрузке на входе ($Z_S = Z_0$ означает, что $a_1 = 0$);

$s_{21} = \frac{b_2}{a_1} \Big|_{a_2=0}$ — коэффициент прямого усиления при согласованной нагрузке на выходе;

$s_{12} = \frac{b_1}{a_2} \Big|_{a_1=0}$ — коэффициент обратного усиления при согласованной нагрузке на входе.

Определим ряд показателей качества СВЧ-транзисторов на основе s -параметров. Коэффициент усиления по мощности G_p определим как отношение мощности, передаваемой в нагрузку, к мощности на входе схемы [44]:

$$G_p = \frac{|s_{21}|^2 (1 - \Gamma_L^2)}{(1 - |s_{11}|^2) + \Gamma_L^2 (|s_{22}|^2 - D^2) - 2\text{Re}(\Gamma_L N)}, \quad (75)$$

где

$$\begin{aligned} \Gamma_L &= (Z_L - Z_0)/(Z_L + Z_0), \\ D &= s_{11}s_{22} - s_{12}s_{21}, \quad N = s_{22} - Ds_{11}^*. \end{aligned}$$

В этой формуле Re означает действительную часть комплексной величины, а звездочка $*$ — комплексно сопряженную величину.

По коэффициенту устойчивости K можно судить о том, будет ли транзистор генерировать при любом сочетании пассивной нагрузки и импеданса источника в отсутствие внешней обратной связи. Коэффициент записывается в виде

$$K = \frac{1 + |D|^2 - |s_{11}|^2 - |s_{22}|^2}{2|s_{12}s_{21}|}. \quad (76)$$

Если $K \gg 1$, прибор абсолютно устойчив, т. е. в отсутствие внешней обратной связи пассивная нагрузка и любой импеданс источника не приведут к колебаниям. Если $K < 1$, прибор потенциально неустойчив, т. е. при определенной комбинации пассивной нагрузки и импеданса источника может возникнуть генерация.

Коэффициент максимального усиления $G_{\text{макс}}$ представляет собой максимально возможный коэффициент усиления по мощности транзистора без внешней обратной связи. Он характеризует прямое усиление транзистора по мощности, когда вход и выход одновременно и сопряженно согласованны, и определен только для абсолютно устойчивого транзистора ($K > 1$):

$$G_{\text{макс}} = \left| \frac{s_{21}}{s_{12}} (K + \sqrt{K^2 - 1}) \right|. \quad (77)$$

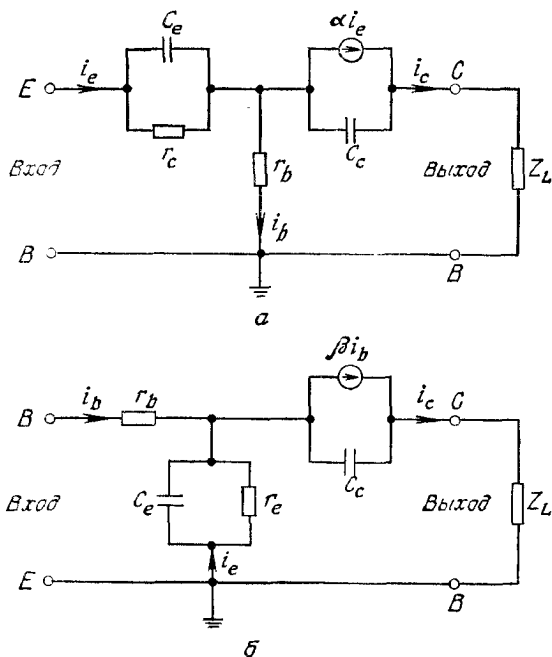


Рис. 21. Упрощенная эквивалентная схема транзистора на высоких частотах. а — схема с общей базой; б — схема с общим эмиттером.

Из выражения (77) видно, что при $K < 1$ выражение в круглых скобках становится комплексным, а $G_{н.м.н.с}$ — неопределенным.

Коэффициент одностороннего усиления U характеризует прямое усиление транзистора по мощности в усилителе с обратной связью, в котором обратное усиление по мощности сведено к нулю за счет настройки идеальной цепи взаимной обратной связи. Этот коэффициент не зависит от реактивности высшей цепи и схемы включения транзистора. Он определяется выражением

$$U = \frac{|s_{11}s_{22}s_{12}s_{21}|}{(1 - |s_{11}|^2)(1 - |s_{22}|^2)}. \quad (78)$$

Найдем теперь связь полученных выше характеристик четырехполюсника с внутренними параметрами транзистора. Упрощенная эквивалентная схема высокочастотного биполярного транзистора приведена на рис. 21. Его параметры были определены выше. Малосигнальным коэффициентом усиления по току в схеме с общей базой является величина α . Рассмотрим транзистор с полосковой базой и полосковым эмиттером (рис. 17). Пусть ширина эмиттера S , длина L , а расстояние между краями эмиттера и базового

контакта S . В такой конструкции емкость коллектора приблизительно равна $C_c = C_0SL$, где C_0 — удельная емкость коллектора. Сопротивление базы при такой геометрии составляет $r_b = r_0S/L$, причем $r_0 \approx \rho_B/W$, где ρ_B — усредненное удельное сопротивление слоя базы.

Малосигнальный коэффициент усиления α в схеме с общей базой определяется выражением

$$\alpha \equiv h_{fb} = \frac{dI_C}{dI_E}. \quad (79)$$

Аналогично определяется малосигнальный коэффициент усиления в схеме с общим эмиттером:

$$\beta \equiv h_{fe} = \frac{dI_C}{dI_B}. \quad (80)$$

Из выражений (21), (22), (79), (80) имеем

$$\alpha = \alpha_0 + I_E \frac{\partial \alpha_0}{\partial I_E}, \quad \beta = \beta_0 + I_B \frac{\partial \beta_0}{\partial I_B} \quad \text{и} \quad \beta = \frac{\alpha}{1 - \alpha}. \quad (81)$$

При низких плотностях тока величины α_0 и β_0 возрастают с током (рис. 7), причем α и β превышают свои статические значения. При больших плотностях тока справедливо обратное утверждение.

Для эквивалентной схемы, приведенной на рис. 21, а, коэффициент U определяется выражением [68]

$$U \equiv \frac{|\alpha(f)|^2}{8\pi^2 r_b C_c \left\{ -\operatorname{Im} [\alpha(f)] + \frac{2\pi f r_e C_c}{1 + 4\pi^2 f^2 r_e^2 C_c^2} \right\}}, \quad (82)$$

где $\operatorname{Im} [\alpha(f)]$ — мнимая часть коэффициента усиления по току α . Если $\alpha(f)$ можно представить в виде $\alpha_0/(1 + jf/f_T)$ и если $f < f_T$, то $\operatorname{Im} [\alpha(f)] \approx -\alpha_0/f_T = -\alpha_0 \omega_{TC}$. Тогда справедливо выражение

$$U \approx \frac{\alpha_0}{16\pi^2 r_b C_c f^2 (\tau_{ec} + r_e C_c \alpha_0)} = \frac{\alpha_0 f^2}{16\pi^2 S^2 r_0 C_c \tau_{ec}^*}, \quad (83)$$

где использованы соотношения $r_b = r_0S/L$ и $C_c = C_0SL$, а $\tau_{ec}^* = \tau_{ec} + r_e C_c / \alpha_0$. При $\alpha_0 \approx 1$ и $\tau_{ec} > r_e C_c$ выражение (83) принимает более простой вид

$$U = \frac{f_T}{8\pi^2 r_b C_c} = \frac{f_T f^2}{8\pi^2 S^2 r_0 C_0}. \quad (84)$$

Другим важным показателем качества является максимальная частота генерации f_{\max} — частота, на которой коэффициент

однаправленного усиления равен 1. Из выражений (83) и (84) экстремализованная величина $f_{\text{макс}}$ равна

$$f_{\text{макс}} \approx \frac{1}{4\pi S} \left(\frac{\alpha_0}{r_0 C_0 \tau_{ec}^*} \right)^{1/2}, \quad (85a)$$

или

$$f_{\text{макс}} \approx \frac{1}{2S} \left(\frac{f_T}{2\pi r_0 C_0} \right)^{1/2}. \quad (85b)$$

Обратим внимание, что и коэффициент однонаправленного усиления, и максимальная частота колебаний возрастают с уменьшением S , вследствие чего ширина эмиттерной полоски является определяющим размером СВЧ-транзисторов.

Другим важным показателем качества выступает шум-фактор — отношение полного среднего квадрата напряжения шума на выходе транзистора к среднему квадрату напряжения шума на источнике теплового шума с сопротивлением R_S . На низких частотах доминирующий источник шума в транзисторе связан с поверхностными эффектами, которые обуславливают возрастание шума по закону $1/f$. На средних и высоких частотах шум-фактор описывается выражением [45]

$$NF = 1 + \frac{r_b}{R_S} + \frac{r_c}{2R_S} + \frac{(1 - \alpha_0) [1 + (1 - \alpha_0)^{-1} (f/f_\alpha)^2] (R_S + r_b + r_c)^2}{2\alpha_0 r_e R_S}. \quad (86)$$

Из выражения (86) видно, что на средних частотах при $f \ll f_\alpha$ шум-фактор примерно постоянен и определяется r_b , r_c , $(1 - \alpha_0)$ и R_S . Существует оптимальное значение R_S , которое находится из условия $d(NF)/dR_S = 0$. Минимальный шум-фактор обозначают $NF_{\text{мин}}$. Для создания малошумящих приборов очень важно снизить величину $(1 - \alpha_0)$, или, что то же самое, повысить α_0 . На высоких частотах, превышающих критическую частоту $f = \sqrt{1 - \alpha_0} f_\alpha$, шум-фактор возрастает примерно как квадрат частоты, т. е. как f^2 .

3.3.3. Геометрия и характеристики приборов

Простейший биполярный транзистор с полосковыми базовыми контактами показан на рис. 17. В настоящее время все биполярные СВЧ-транзисторы имеют планарную конструкцию, а большинство из них представляют собой кремниевые $n-p-n$ -транзисторы. По геометрии их можно подразделить на три основные конфигурации (рис. 22) [46]: встречноштыревые, многоэмиттерные и ячеистые. Как уже говорилось выше, из-за падения напряжения вдоль перехода база—эмиттер эмиттерный ток в основном про-

текает по периферии эмиттера. Следовательно, и способность биполярного транзистора пропускать большой ток тем выше, чем больше длина эмиттера. Все конфигурации, представленные на рис. 22, имеют большое отношение периметра эмиттера к его площади.

Во всех конструкциях транзисторов для образования омических контактов к эмиттеру, базе и коллектору использован процесс финишной металлизации. Особо важную роль процесс металлизации играет в технологии транзисторов с балочными выводами [9, 47]. На рис. 23 показан кремниевый высокочастотный транзистор с балочными выводами и полосковой геометрией базовых контактов, как в транзисторе, показанном на рис. 17. Металлические выводы толщиной ~ 10 мкм выполняют роль держателя кремниевого кристалла и одновременно служат электрическими контактами. Технология приборов с балочными выводами позволяет получить исключительно высокую надежность и хорошие электрические характеристики.

Для уменьшения расстояния между электродами базы и эмиттера были созданы транзисторы со ступенчатыми выводами

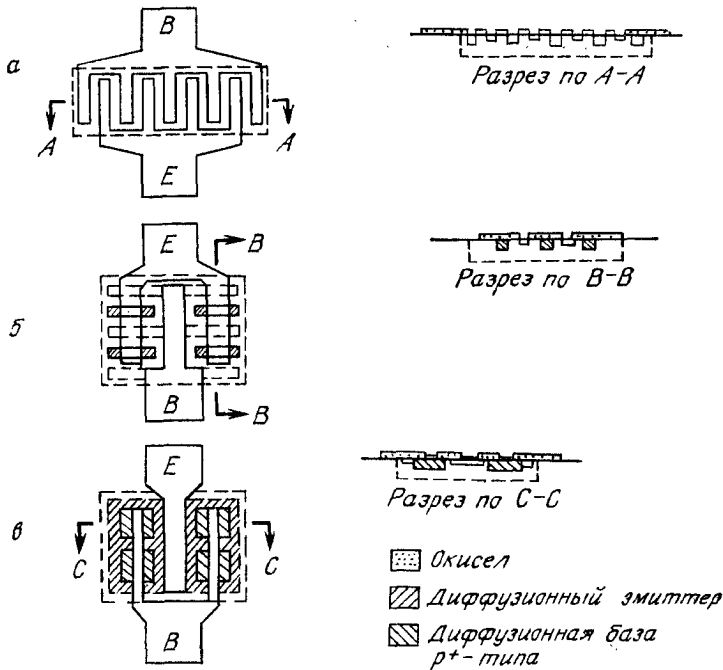


Рис. 22. Три конфигурации СВЧ-транзисторов [46].
а — встречноштыревая; б — многоэмиттерная; в — ячеистая.

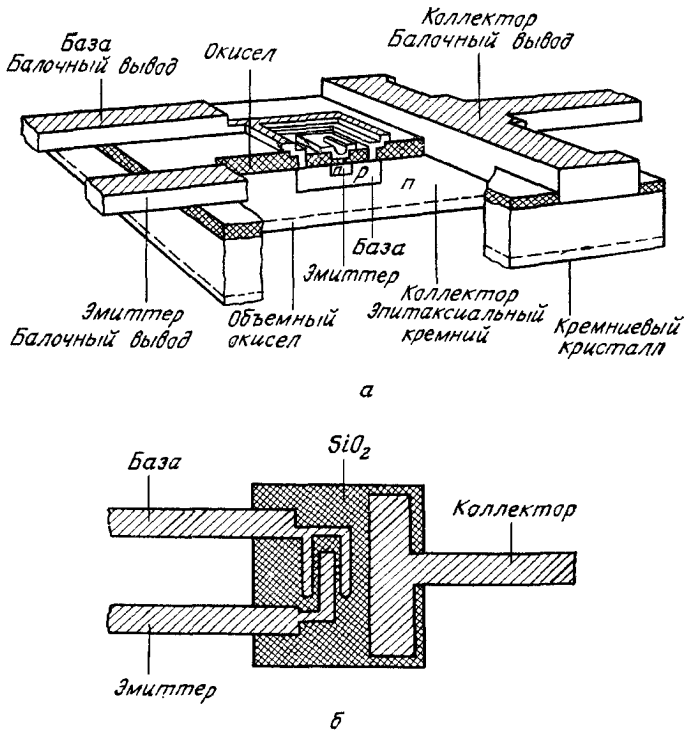
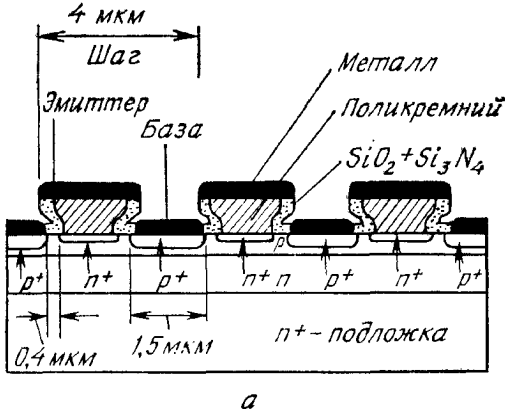


Рис. 23. Транзистор с балочными выводами [9].

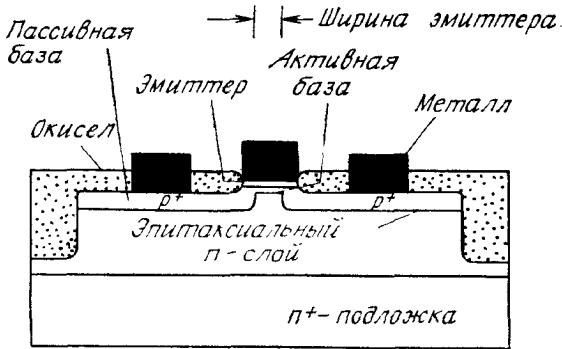
а — в разрезе; *б* — вид сверху.

(рис. 24, *а*) [48]. При помощи селективного травления поликристаллического кремния получают электроды в форме обратной трапеции. Боковые ребра обратных трапеций образуют изолирующие слои между напыленными выводами базы и эмиттера. Расстояние между диффузионным слоем эмиттера и контактом к базе может быть уменьшено до 0,4 мкм и менее; частота отсечки в этом случае составляла 8,4 ГГц.

Другая усовершенствованная конструкция СВЧ-транзистора, в котором активная и пассивная области базы сформированы разными допускающими независимую оптимизацию процессами, приведена на рис. 24, *б* [49]. За счет уменьшения эффективной ширины эмиттера и сопротивления базы достигается высокая частота отсечки и низкий шум-фактор. На рис. 25 приведены коэффициент усиления по мощности и другие показатели качества такого транзистора. Результаты получены из измерений *s*-параметров. Коэффициент однонаправленного усиления *U* в соответствии с формулой (84) изменяется по закону f^{-2} . Экстраполируя



а



б

Рис. 24. Транзистор со ступенчатыми электродами (а) и с имплантированной базой (б) [49].

зависимость до $U = 1$ (0 дБ), получим, что у исследуемого СВЧ-транзистора $f_{\text{макс}}$ составляет 25 ГГц. Коэффициент максимального усиления $G_{0 \text{ макс}}$ и коэффициент усиления по току в схеме с общим эмиттером h_{fe} зависят от частоты по закону f^{-2} . Частота отсечки f_T , определяемая из условия $h_{fe} = 1$ (0 дБ), оказалась выше 5 ГГц. Минимальный шум-фактор увеличивается с 1 дБ при 1 ГГц до 4 дБ при 8 ГГц.

На рис. 26 приведена зависимость выходной СВЧ-мощности от частоты при современном состоянии технологии биполярных транзисторов [50]. Выходная мощность обратно пропорциональна квадрату частоты, что является результатом ограничения, налагаемого полем лавинного пробоя и предельной скоростью носителей [18]. (гл. 10). В импульсном режиме работы на частоте 1 ГГц

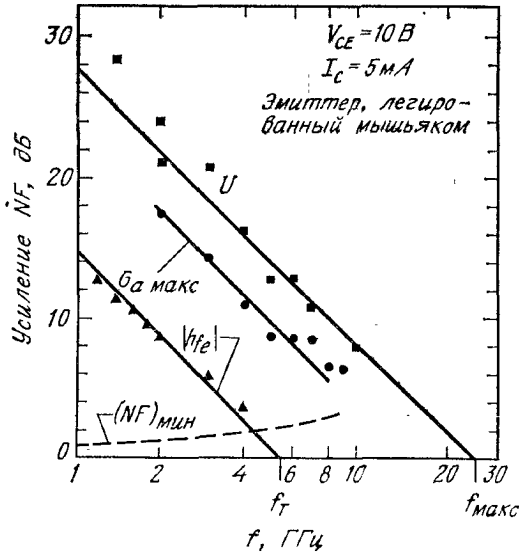


Рис. 25. Коэффициент однонаправленного усиления, коэффициент максимального усиления, коэффициент усиления в схеме с общим эмиттером и фактор шума биполярного транзистора, полученного имплантацией [49].

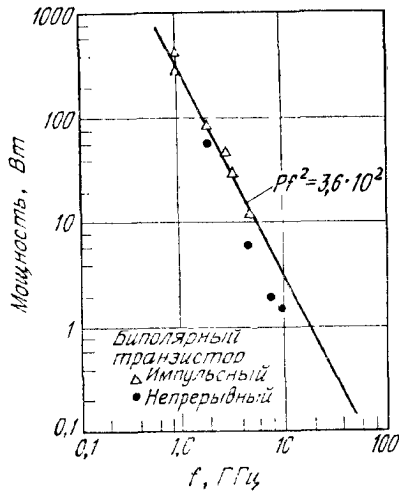


Рис. 26. Зависимость мощности от частоты для современных биполярных СВЧ-транзисторов [50].

может быть достигнута мощность ~ 500 Вт. В непрерывном режиме получены следующие значения: 60 Вт на 2 ГГц, 6 Вт на 5 ГГц и 1,5 Вт на 10 ГГц. С развитием новых способов и технологических процессов следует ожидать трехкратного повышения предельных параметров по сравнению с полученными в настоящее время.

3.4. МОЩНЫЕ ТРАНЗИСТОРЫ

3.4.1. Распределение температуры и балластное сопротивление эмиттера

Мощные транзисторы проектируются для использования в усилителях мощности и регулирования высоких напряжений и больших токов. При разработке мощных транзисторов основное внимание уделяется абсолютным величинам рассеиваемой мощности и ограничениям работоспособности из-за вторичного пробоя. При описании СВЧ-транзисторов в разд. 3.3 особое внимание уделялось частоте отсечки и коэффициенту усиления по мощности. Однако между мощными транзисторами и СВЧ-транзисторами нет четкой очерченной границы, так как производство мощности на частоту в основном ограничено параметрами материала [18].

В мощных транзисторах с увеличением рассеиваемой мощности повышается температура переходов T_j . Максимум T_j ограничен температурой, при которой в базе достигается собственная проводимость полупроводника. При температуре, превышающей T_j , прибор перестает работать как транзистор из-за короткого замыкания между коллектором и эмиттером. Для улучшения характеристик транзистора последний необходимо помещать в корпус с хорошим теплоотводом, обеспечивающим эффективное рассеяние выделяемого при работе тепла.

Чтобы получать большие уровни мощности, необходимо правильно выбрать ширину полоски эмиттера S и толщину базы транзистора W_b . В мощных транзисторах также используют встречно-штыревые и многоэмиттерные структуры (рис. 22). Это позволяет пропускать через них большой ток и распределять его более равномерно.

На рис. 27, *а* и *б* приведены распределения тока и температуры вдоль половины перехода эмиттер—база транзистора (ширина полоски эмиттера составляет 250 мкм) для четырех различных условий смещения [51]. Полная входная мощность во всех случаях одинакова (60 Вт). Отметим, что при высоких напряжениях и малых токах (кривые *A* и *B*) ток концентрируется в центре полоски, где наблюдается также очень сильное повышение температуры. При низких напряжениях и больших токах (кривая *Г*) ток концентрируется вблизи края эмиттера (эффект вытеснения

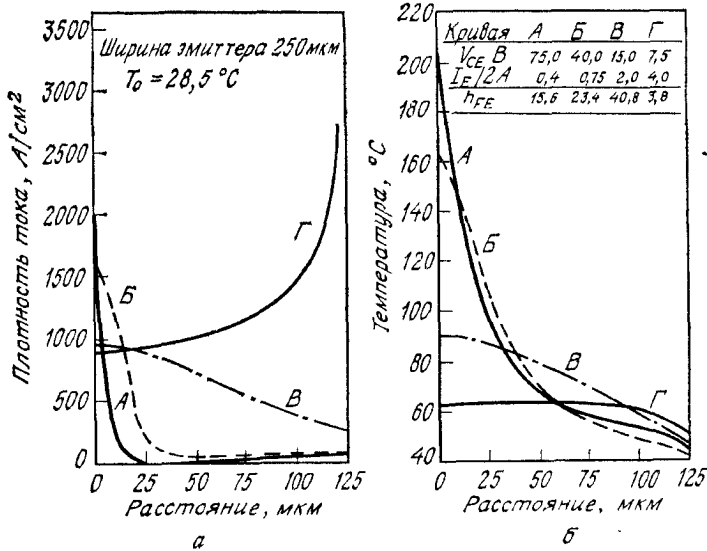


Рис. 27. Изменение плотности тока (а) и температуры (б) с расстоянием от центра перехода эмиттер — база [51].

эмиттера), а повышение температуры в центре полоски намного слабее. Следовательно, работа при высоких напряжениях и малых токах сопряжена с более сложными тепловыми условиями.

Для улучшения равномерности распределения тока во встречно-штыревом и многоэмиттерном транзисторах добавляют распределенное эмиттерное сопротивление R_E , которое ограничивает любое нежелательное увеличение тока через отдельный эмиттер. Такие последовательно включенные резисторы называют стабилизирующими резисторами или балластными эмиттерными резисторами. На рис. 28 (вставка) приведена схема включения балластного резистора R_E . Коллекторный ток в этой схеме [52] определяется выражением

$$I_C = I_{SO} \exp \left\{ \frac{qV_{in} - qI_C [(R_E + r_e)/\alpha + r_b (1/\alpha - 1) - E_g(T_j)]}{k(T_0 + R_{th}I_C/V_C)} \right\}, \quad (87)$$

где I_{SO} — постоянная, не зависящая от температуры; E_g — ширина запрещенной зоны при температуре перехода T_j ; T_0 — температура окружающей среды; R_{th} — тепловое сопротивление. На рис. 28 приведены экспериментальные результаты (светлые кружки), которые хорошо согласуются с расчетными кривыми, полученными по формуле (87) (сплошные линии). При малых бал-

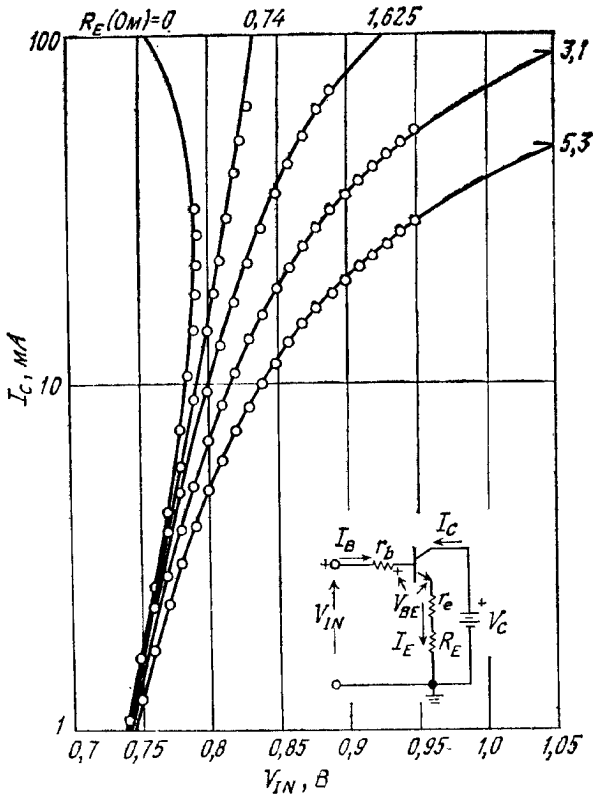


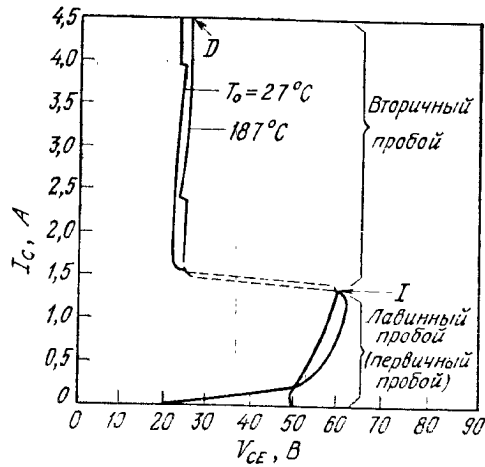
Рис. 28. Зависимость коллекторного тока от напряжения при различных балластных эмиттерных резисторах R_E . На вставке приведена эквивалентная схема с R_E [52].

ластных резисторах (менее 0,74 Ом) и при высоких плотностях тока дифференциальное сопротивление dV_{IN}/dI становится отрицательным, а ток — неуправляемым. Если R_E достаточно велико, транзистор абсолютно устойчив, так как дифференциальное сопротивление положительно.

3.4.2. Вторичный пробой

Использование мощных транзисторов и других полупроводниковых приборов часто ограничивается явлением, называемым вторичным пробоем, признаком которого служит резкое снижение напряжения на приборе с одновременным внутренним шнурованием тока. Об этом явлении впервые сообщили Торнтон и Симмонс [53], и с тех пор оно широко изучается при исследовании мощных

Рис. 29. Вольт-амперные характеристики транзистора в режиме вторичного пробоя для двух значений температуры [56].



полупроводниковых приборов [54, 55]. Транзисторы большой мощности должны работать в условиях безопасного режима, при которых мала вероятность возникновения вторичного пробоя.

Особенности вольт-амперной характеристики транзистора в режиме вторичного пробоя показаны на рис. 29 [56]. Когда напряжение, приложенное между эмиттером и коллектором, превышает величину, определенную формулой (44), происходит лавинный пробой (первичный пробой). При дальнейшем повышении напряжения наступает вторичный пробой. Экспериментальные результаты позволяют в общем случае выделить четыре стадии вторичного пробоя: на первой стадии наблюдается нестабильность тока при напряжениях, близких к пробивному; на второй стадии происходит переключение из области высокого в область низкого напряжения; третья стадия характеризуется возрастанием тока при низком напряжении; четвертая стадия соответствует разрушению прибора (буква *D* на рис. 29). Возникновение нестабильности вызвано в первую очередь температурными эффектами. Когда на транзистор поступает достаточно мощный импульс $P = I_C \cdot BV_{CEO}$, вызывающий вторичный пробой, последний наступает с некоторой задержкой во времени. Это время называют временем включения вторичного пробоя. На рис. 30 приведена типичная зависимость времени включения вторичного пробоя от величины мощности приложенного импульса при различной температуре окружающей среды [57]. Для определенного времени включения t существует температура включения вторичного пробоя T_{tr} , под которой следует понимать температуру «горячих» точек в момент, предшествующий лавинному пробую. Вели-

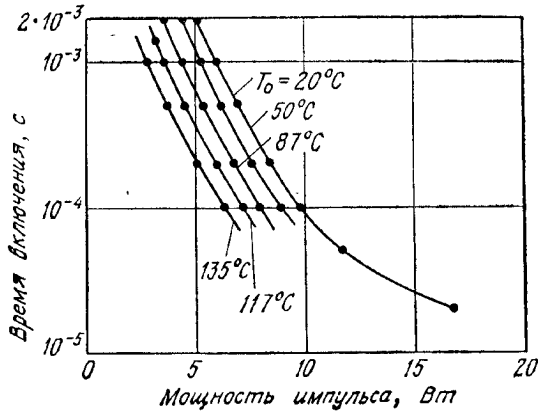


Рис. 30. Зависимость времени включения вторичного пробоя от мощности приложенного импульса при различной температуре окружающей среды [57].

чина T_{tr} связана с мощностью импульса P и с температурой окружающей среды T_0 следующим соотношением:

$$T_{tr} - T_0 = C_1 P, \quad (88)$$

где C_1 — постоянная. Из рис. 30 видно, что при фиксированной температуре соотношение между мощностью импульса и временем включения приблизительно равно

$$\tau \sim \exp(-C_2 P), \quad (89)$$

где C_2 — постоянная.

Подставляя выражение (88) в формулу (89), получим

$$\tau \sim \exp\left[-\frac{C_2}{C_1}(T_{tr} - T_0)\right]. \quad (90)$$

Температура вторичного пробоя T_{tr} зависит от различных параметров и геометрии прибора. Для большинства кремниевых диодов и транзисторов температура T_{tr} соответствует той температуре, при которой концентрация носителей собственного материала n_i равна концентрации примеси в коллекторе (рис. 12, гл. 1). «Горячие» точки обычно располагаются вблизи центра прибора. Величина T_{tr} изменяется в зависимости от концентрации примеси в коллекторе, а отношение постоянных C_2/C_1 определяется геометрией прибора. Поэтому время включения вторичного пробоя, зависящее экспоненциально от этих параметров (формула (90)), сильно колеблется для разных приборов.

За пределами области нестабильных токов напряжение на приборе резко падает. В течение этой второй стадии вторичного пробоя сопротивление точек локального пробоя резко уменьшается. На третьей стадии (стадии низкого напряжения) полу-

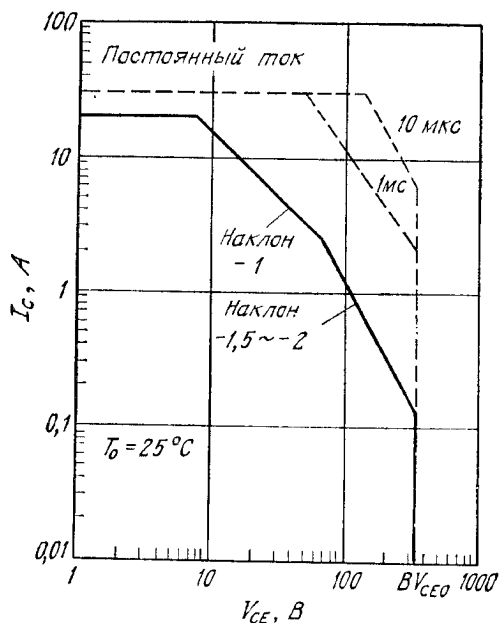


Рис. 31. Область безопасной работы транзистора [58].

проводник находится при высокой температуре, а вблизи точек пробоя становится собственным полупроводником. При дальнейшем возрастании тока точки пробоя начинают плавиться и наступает четвертая стадия — разрушение прибора.

Для защиты транзистора от постепенного разрушения необходимо установить область его безопасной работы. Типичный пример такой области для кремниевого мощного транзистора, включенного по схеме с общим эмиттером, приведен на рис. 31. Нагрузочная кривая коллекторной цепи для определенной схемы должна лежать ниже указанного допустимого предела. Ограничение вытекает из следующего требования: максимальная температура перехода не должна превышать 150 °С. На постоянном токе тепловой предел области безопасной работы определяется тепловым сопротивлением прибора R_{th} [58]:

$$R_{th} = \frac{T_j - T_0}{P}, \quad (91)$$

где P — рассеиваемая мощность. Следовательно, тепловой предел ограничивает максимально допустимую температуру перехода:

$$R_{th \text{ макс}} = \frac{T_{j \text{ макс}} - T_0}{(I_C \times V_{CE})_{\text{макс}}}. \quad (92)$$

Если $T_{j \text{ макс}} = 150^\circ\text{C}$ и $R_{th \text{ макс}}$ считать постоянным, то

$$(I_C \times V_{CE})_{\text{макс}} = \frac{T_{j \text{ макс}} - T_0}{R_{th \text{ макс}}} = \text{const.} \quad (93)$$

Следовательно, между $\ln I_C$ и $\ln V_{CE}$ существует линейная зависимость, и наклон их графика (рис. 31) равен -1 . Как следует из рис. 27, при высоких напряжениях и малых токах в центре эмиттерной полоски может существенно повыситься температура, что ведет к возникновению вторичного пробоя, и наклон прямой принимает значение, лежащее между $-1,5$ и -2 . Другое ограничение области безопасной работы связано с напряжением первичного пробоя BV_{CEO} (вертикальная линия на рис. 31). В импульсном режиме область безопасной работы расширяется в сторону больших токов. При повышенной температуре окружающей среды тепловые ограничения снижают мощность, которую способен пропускать прибор (формула (93)), и область безопасной работы сужается.

3.5. ПЕРЕКЛЮЧАЮЩИЕ ТРАНЗИСТОРЫ

Переключающие транзисторы используются в качестве ключей, которые могут за очень короткое время перейти из состояния с высоким напряжением и малым током в состояние с низким напряжением и большим током. Условия работы переключающих транзисторов отличаются от условий работы СВЧ-транзисторов. Дело в том, что процесс переключения представляет собой переходный процесс при больших сигналах, в то время как СВЧ-транзисторы обычно усиливают малые сигналы. Тем не менее геометрия переключающего транзистора аналогична геометрии СВЧ-транзистора (рис. 17).

Наиболее важными параметрами переключающего транзистора являются коэффициент усиления по току и время переключения. Чтобы улучшить коэффициент усиления по току, достаточно снизить степень легирования базовой области. Для уменьшения времени переключения транзистор может быть легирован золотом, которое образует центры рекомбинации вблизи середины запрещенной зоны.

Переключающий транзистор допускает различные режимы работы. Два основных режима и соответствующие нагрузочные линии приведены на рис. 32, а. По степени использования определенных областей выходных характеристик транзисторов они получили названия режима насыщения и токового режима. Выходные характеристики принято делить на три области: Область I: область отсечки, коллекторный ток отсутствует, эмиттерный и коллекторный переходы обратно смещены.

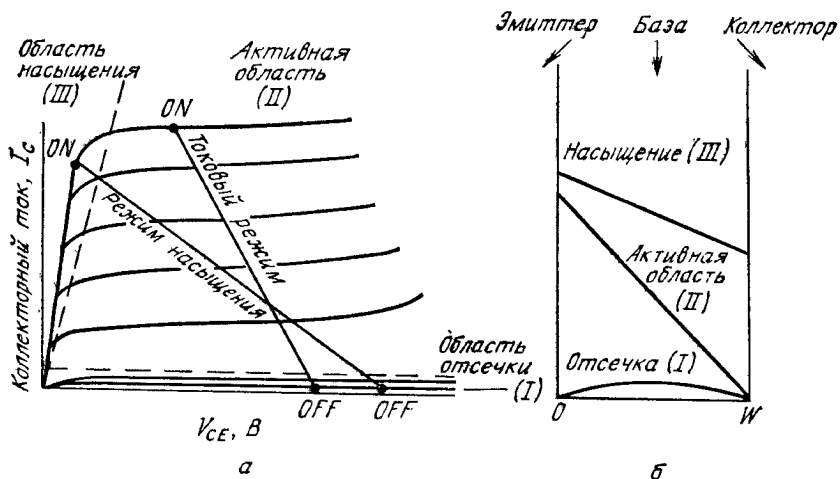


Рис. 32. Режимы работы переключающего транзистора [59].

a — способы переключения; *б* — распределение неосновных носителей в базе при отсечке, в активной области и при насыщении.

Область II: активная область, эмиттер смещен в прямом направлении, а коллектор в обратном.

Область III: область насыщения, эмиттер и коллектор смещены в прямом направлении.

Распределения неосновных носителей в базе транзистора, соответствующие областям отсечки, активной области и области насыщения, приведены на рис. 32, б [59].

Для обоих режимов выключенное состояние (OFF) характеризуется сдвигом рабочей точки по нагрузочной линии в область отсечки транзистора. Итак, способ переключения в первую очередь определяется уровнем прямого тока и положением рабочей точки во включенном состоянии (ON). Чаще всего применяют режим насыщения как наиболее близкий к режиму работы идеального ключа. Фактически в транзисторе в выключенном состоянии разомкнута цепь между эмиттером и коллектором, а во включенном состоянии она короткозамкнута. Токовый режим используют для высокоскоростного переключения, так как при этом исключается время рассасывания, связанное с пребыванием транзистора в области насыщения.

Рассмотрим теперь поведение транзистора при переключении на основе модели Эберса—Молла. Введенные выше в выражениях (47а) и (47б) коэффициенты a с соответствующими индексами можно определить из совокупности четырех величин, допускающих непосредственное измерение:

I_{EO} : обратный ток насыщения эмиттерного перехода при разомкнутом коллекторе,

$$e^{qV_{EB}/kT} \ll 1, \quad I_C = 0.$$

I_{CO} : обратный ток насыщения коллекторного перехода при разомкнутом эмиттере,

$$e^{qV_{CB}/kT} \ll 1, \quad I_E = 0.$$

α_N : нормальный коэффициент усиления по току при нормальных условиях включения, когда эмиттер смещен в прямом направлении, а коллектор в обратном. Тогда коллекторный ток равен $I_C = -\alpha_N I_E + I_{CO}$.

α_I : инверсный коэффициент усиления по току при обратном включении, т. е. когда эмиттер смещен в обратном направлении, а коллектор в прямом. Тогда эмиттерный ток равен $I_E = -\alpha_I I_C + I_{EO}$. В большинстве транзисторов площадь эмиттера меньше площади коллектора и последний немного эффективнее собирает носители, инжектированные эмиттером. Поэтому обычно $\alpha_N > \alpha_I$.

Приведенные выше величины позволяют представить коэффициенты входящие в выражения (47а) и (47б), в следующей форме:

$$a_{11} = -\frac{I_{EO}}{1 - \alpha_N \alpha_I}, \quad a_{12} = \frac{\alpha_I I_{CO}}{1 - \alpha_N \alpha_I}, \quad (94)$$

$$a_{21} = \frac{\alpha_N I_{EO}}{1 - \alpha_N \alpha_I}, \quad a_{22} = -\frac{I_{CO}}{1 - \alpha_N \alpha_I}.$$

В областях I и II коллекторный переход смещен в обратном направлении. Поэтому формулы (47а) и (47б) можно записать в более простом виде

$$I_E = -\frac{I_{EO}}{1 - \alpha_N \alpha_I} e^{qV_{EB}/kT} + \frac{(1 - \alpha_N) I_{EO}}{1 - \alpha_N \alpha_I}, \quad (95a)$$

$$I_C = \frac{\alpha_N I_{EO}}{1 - \alpha_N \alpha_I} e^{qV_{EB}/kT} + \frac{(1 - \alpha_I) I_{CO}}{1 - \alpha_N \alpha_I}. \quad (95b)$$

Эквивалентная схема, соответствующая выражениям (95), приведена на рис. 33, а. Для учета конечной проводимости полупроводникового материала в схему добавлены сопротивления базы r_b , эмиттера r_e и коллектора r_c . В области III за независимые переменные удобно принять токи. Из приведенных выше выражений следует

$$V_{EB} = \frac{kT}{q} \ln \left(-\frac{I_E + \alpha_I I_C}{I_{EO}} + 1 \right), \quad (96a)$$

$$V_{CB} = \frac{kT}{q} \ln \left(-\frac{I_C + \alpha_N I_E}{I_{CO}} + 1 \right). \quad (96b)$$

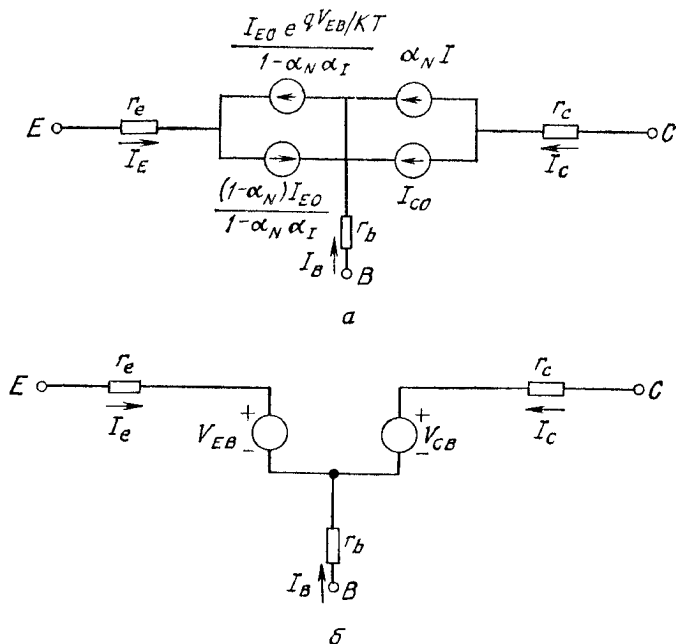


Рис. 33. Эквивалентные схемы переключающего транзистора [32].
 а — для областей I и II; б — для области III.

На рис. 33, б приведена эквивалентная схема транзистора для области III. Формулы (95) и (96) позволяют анализировать логические задачи переключения транзисторов в режиме большого сигнала.

Для описания характеристик переключающего транзистора необходимо определить следующие пять параметров: предельный ток, максимально допустимое напряжение, импедансы во включенном и выключенном состояниях и время переключения. Предельный ток определяется допустимой мощностью рассеяния и так же, как в мощных транзисторах, задается тепловым ограничением. Максимально допустимое напряжение определяется напряжением пробоя или прокола, рассмотренным выше. Импедансы во включенном и выключенном состояниях могут быть получены из формул (95) и (96) при соответствующих граничных условиях. Например, для схемы с общей базой эти импедансы имеют вид

$$V_C/I_C \text{ (выкл., область I)} = \frac{V_C(1 - \alpha_N \alpha_I)}{I_{C0} - \alpha_N I_{E0}}, \quad (97)$$

$$V_C/I_C \text{ (вкл., область III)} = \frac{kT}{qI_C} \ln \left(-\frac{I_C + \alpha_N I_E}{I_{C0}} \right). \quad (98)$$

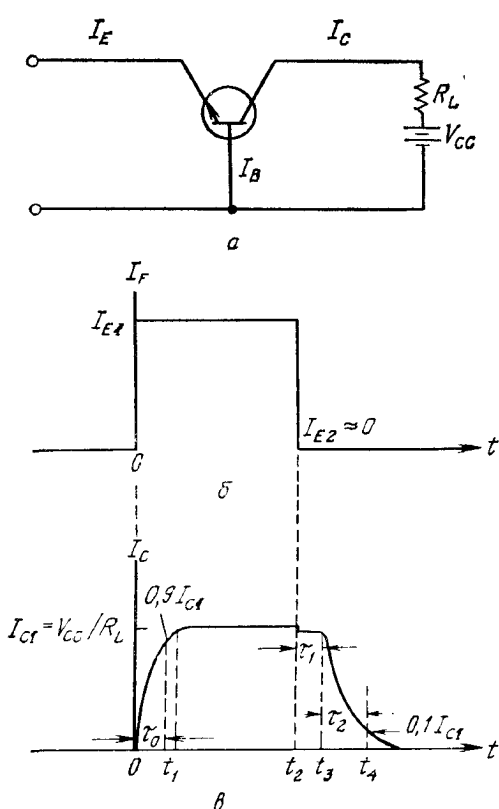


Рис. 34. Переключение $n-p-n$ -транзистора в схеме с общей базой [59].

a — схема включения; b — входной импульс эмиттерного тока; v — соответствующий отклик коллекторного тока.

Из выражения (97) видно, что импеданс в выключенном состоянии выше при малых обратных токах насыщения переходов I_{CO} и I_{EO} . Импеданс во включенном состоянии, определяемый формулой (98), в первом приближении обратно пропорционален коллекторному току I_C и очень мал при большой величине I_C . Омические сопротивления (рис. 33, б) вносят существенный вклад в полный импеданс транзистора и должны приниматься в расчет.

Рассмотрим теперь время переключения, т. е. время, необходимое для перевода транзистора из выключенного состояния во включенное или обратно. В общем случае время включения отличается от времени выключения [59]. На рис. 34, a представлен процесс переключения транзистора, включенного по схеме с общей базой. После подачи импульса на эмиттерный вывод (рис. 34, б) происходит включение транзистора на интервале времени от $t = 0$ до t_1 (рис. 34, в), а переходный процесс включения определяется параметрами активного режима (область II). В момент времени t_1 рабочая точка попадает в область насыщения

тока (область III). Время, необходимое для нарастания тока до уровня 90 % тока насыщения ($=V_{CC}/R_L$), называют временем включения τ_0 . В момент времени t_2 ток эмиттера снижается до нуля и начинается процесс выключения. В интервале от t_2 до t_3 плотность неосновных носителей в базовом слое остается большой, что соответствует режиму области III (рис. 32; б), но постепенно она снижается до нуля. В течение отрезка времени τ_1 импеданс транзистора сохраняется низким и коллекторный ток определяется внешней цепью. В момент времени t_3 плотность носителей около коллекторного перехода становится близкой 0. Одновременно быстро возрастает импеданс коллекторного перехода и транзистор начинает работать в активной области II. Интервал времени τ_1 называют временем рассасывания носителей. Начиная с момента t_3 переходный процесс рассчитывается, исходя из параметров активной области II. К моменту времени t_4 коллекторный ток снижается до 10 % максимального значения. Интервал времени τ_2 , заключенный между t_3 и t_4 , называют временем спада.

Время включения τ_0 можно получить из анализа переходного процесса в активном режиме. Для ступенчатого входного сигнала I_{E1} преобразование Лапласа имеет вид I_{E1}/s . Если коэффициент усиления в схеме с общей базой представить в виде $\alpha_N/(1 + j\omega/\omega_N)$, где ω_N — частота отсечки коэффициента α , при которой $\alpha/\alpha_N = 1/\sqrt{2}$, то преобразование Лапласа для коэффициента усиления имеет вид $\alpha_N/(1 + s/\omega_N)$. Тогда для коллекторного тока преобразование Лапласа будет иметь следующий вид:

$$I_C(s) = \frac{\alpha_N}{1 + s/\omega_N} \frac{I_{E1}}{s}. \quad (99)$$

Обратное преобразование Лапласа этого выражения записывается в виде

$$I_C = I_{E1}\alpha_N(1 - e^{-\omega_N t}). \quad (100)$$

Обозначив через $I_{C1} \approx V_{CC}/R_L$ величину коллекторного тока в режиме насыщения и подставив $I_C = 0,9I_{C1}$ в формулу (100), получим значение τ_0 :

$$\tau_0 = \frac{1}{\omega_N} \ln \left(\frac{I_{E1}}{I_{E1} - 0,9I_{C1}/\alpha_N} \right). \quad (101)$$

На основе подхода, аналогичного изложенному выше, можно получить формулу для времени рассасывания и времени спада переходного процесса выключения транзистора в схеме с общей базой [59]:

$$\tau_1 = \frac{\omega_N + \omega_I}{\omega_N \omega_I (1 - \alpha_N \alpha_I)} \ln \left[\frac{I_{E1} - I_{E2}}{(I_{C1}/\alpha_N) - I_{E2}} \right], \quad (102)$$

$$\tau_2 = \frac{1}{\omega_N} \ln \left(\frac{I_{C1} - \alpha_N I_{E2}}{0,1I_{C1} - \alpha_N I_{E2}} \right), \quad (103)$$

где ω_1 — частота отсечки инверсного коэффициента α_1 (величины I_{E1} и I_{E2} указаны на рис. 34). Из формулы (102) следует, что время рассасывания становится равным 0, если транзистор не попадает в область насыщения III (как при токовом режиме работы), поскольку в этом случае $I_{C1} = \alpha_N I_{E1}$. Приведенные выше выражения можно использовать для схемы с общим эмиттером, если выполнить следующую замену величин: при определении τ_0 и τ_2 частота ω_N заменяется на частоту отсечки β , равную $\omega_N/(1 - \alpha_N)$, I_{E1} и I_{E2} заменяются на I_{B1} и I_{B2} , а α_N заменяется на $\alpha_N/(1 - \alpha_N)$; при определении τ_1 производятся две последние замены: I_{E1} и I_{E2} заменяются на I_{B1} и I_{B2} , а α_N заменяется на $\alpha_N/(1 - \alpha_N)$.

Из приведенных выше соотношений вытекает, что время переключения, т. е. время включения τ_0 и время выключения ($\tau_1 + \tau_2$), обратно пропорционально частоте отсечки. Для повышения скорости переключения необходимо увеличить частоту отсечки. Важно отметить, что быстродействие большинства переключающих транзисторов ограничено накоплением в коллекторе, которое может быть снижено при увеличении частоты отсечки.

3.6. РАЗНОВИДНОСТИ БИПОЛЯРНЫХ ТРАНЗИСТОРОВ

3.6.1. Масштабирование транзисторов

В интегральных схемах (ИС), и особенно в сверхбольших интегральных схемах (СБИС), для повышения быстродействия и плотности размещения элементов необходимо уменьшать размеры биполярных транзисторов. Прогресс в уменьшении размеров биполярных транзисторов за последнее десятилетие проиллюстрирован на рис. 35 [60]. Отметим, что основное отличие транзисторов интегральных схем (интегральных транзисторов) от дискретных транзисторов заключается в планарном расположении всех выводов на верхней поверхности подложки, а также в необходимости электрической изоляции транзисторов для предотвращения паразитного взаимодействия между ними. До 1970 г. использовалась изоляция $p-n$ -переходом (рис. 35, а) и коллекторные области n -типа отделялись друг от друга изолирующими p -областями, смещенными в обратном направлении. С 1971 г. начали применять изоляцию приборов термически выращенным окислом, что привело к существенному уменьшению размеров (рис. 35, б). В 1975 г. научились ограничивать окисной стенкой эмиттер, и площадь дополнительно уменьшилась на 50 % (рис. 35, в). В настоящее время уменьшение всех размеров осуществляется на основе принципа масштабирования: если коэффициент масштабирования равен 2, то в результате площадь уменьшается в четыре раза, а ширина базы — в два раза (рис. 35, г).

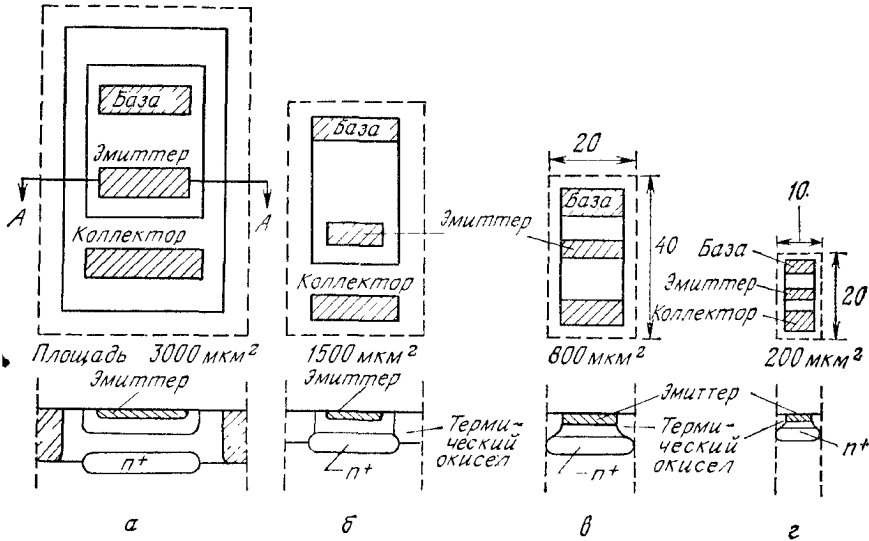


Рис. 35. Уменьшение горизонтальных и вертикальных размеров биполярных транзисторов за последнее десятилетие [60].

а — изоляция p-n-переходом; б — окисная изоляция; в, г — масштабирование транзисторов с окисной изоляцией.

Трехмерное изображение транзистора с окисной изоляцией приведено на рис. 36. Изолирующие области можно выполнять непосредственно внутри транзистора, что не только уменьшает общую площадь, но и значительно улучшает характеристики прибора из-за уменьшения паразитных емкостей. С развитием ионной имплантации, литографии и методов сухого травления можно

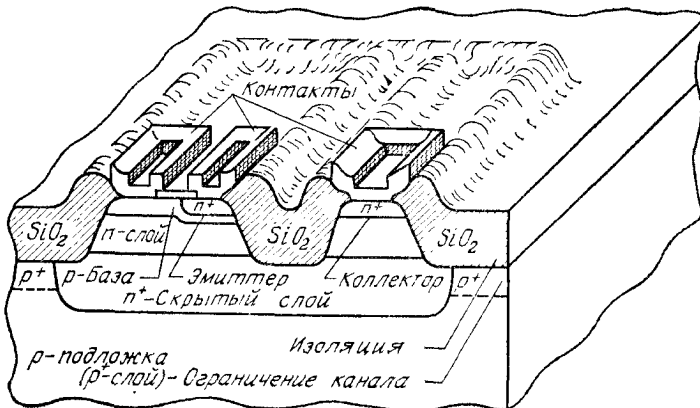


Рис. 36. Трехмерное изображение транзистора с окисной изоляцией [11].

в недалеком будущем ожидать появления биполярных транзисторов с субмикронными размерами, обладающих произведением мощности на время задержки менее 1 пДж.

3.6.2. Интегральная инжекционная логика [61, 62]

Интегральная инжекционная логика (И²Л) с 1972 г. получила широкое распространение в логических и запоминающих устройствах. Ее привлекательными особенностями являются совместимость с технологией биполярных транзисторов, простота топологии и высокая плотность упаковки. Эквивалентная электрическая схема и разрез структуры интегральной инжекционной логики представлены на рис. 37. Базовый элемент И²Л содержит продольный $p-n-p$ -транзистор Q_1 и инверсный вертикальный многоколлекторный $n-p-n$ -транзистор Q_2 . Поскольку для И²Л не требуются изолирующие области и резисторы, плотность упа-

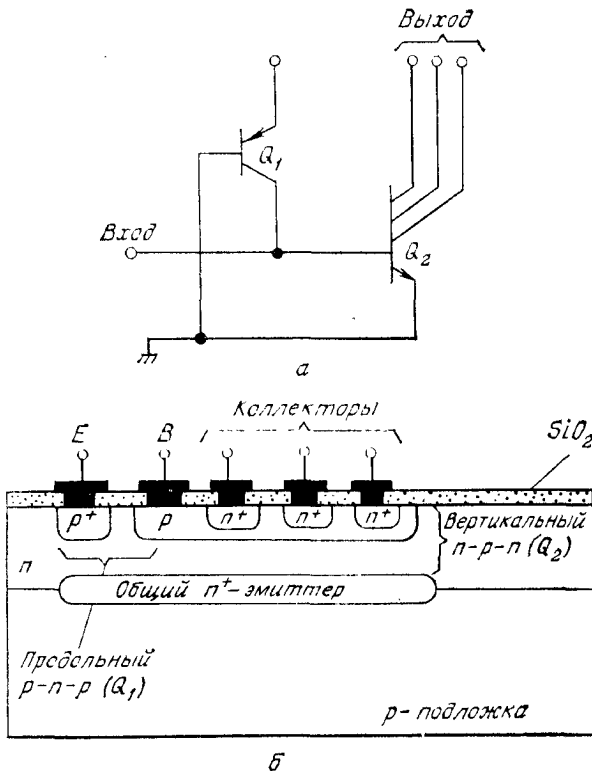


Рис. 37. Интегральная инжекционная логика.
а — эквивалентная схема элемента; б — разрез элемента.

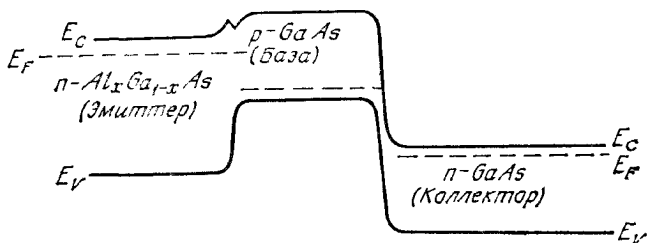


Рис. 38. Зонная диаграмма гетеропереходного $n - p - n$ -транзистора.

ковки схем может быть очень высокой. Для повышения быстродействия в качестве коллекторов целесообразно использовать диоды Шоттки (гл. 5) [63].

3.6.3. Транзисторы с гетеропереходами [64]

Энергетическая зонная диаграмма гетеропереходного транзистора с широкозонным эмиттером приведена на рис. 38. Прибор имеет эмиттер n -типа из $\text{Al}_x\text{Ga}_{1-x}\text{As}$, базу p -типа из GaAs и коллектор n -типа из GaAs . К преимуществам гетеротранзистора относится следующее: 1) высокая эффективность эмиттера, так как потоку дырок (неосновных носителей в эмиттере) из базы в эмиттер препятствует высокий барьер в валентной зоне; 2) уменьшенное сопротивление базы, поскольку база может быть сильно легирована без снижения эффективности эмиттера; 3) меньшее вытеснение тока в эмиттере вследствие малого падения напряжения вдоль перехода эмиттер—база; 4) улучшенная переходная характеристика из-за высокого коэффициента усиления по току и низкого сопротивления базы; 5) расширенный температурный диапазон — гетеротранзистор может работать при более высоких температурах ($\sim 350^\circ\text{C}$), так как ширина его запрещенной зоны больше, и при низких температурах (вплоть до гелиевых (4 К)), так как содержит мелкие примесные уровни. В настоящее время получены гетеротранзисторы с коэффициентом усиления по току $\beta_0 \approx 350$ [65]. Однако из-за технологических ограничений частота отсечки не превышает 1 ГГц.

3.6.4. Транзисторы на горячих электронах

Горячими считаются те электроны, энергия которых превышает энергию Ферми на несколько kT , где k — постоянная Больцмана, а T — температура решетки. Было предложено много трехэлектродных структур, подобных биполярным транзисторам, с переносом горячих электронов от эмиттера к коллектору. Первый такой прибор со структурой металл—диэлектрик—металл—

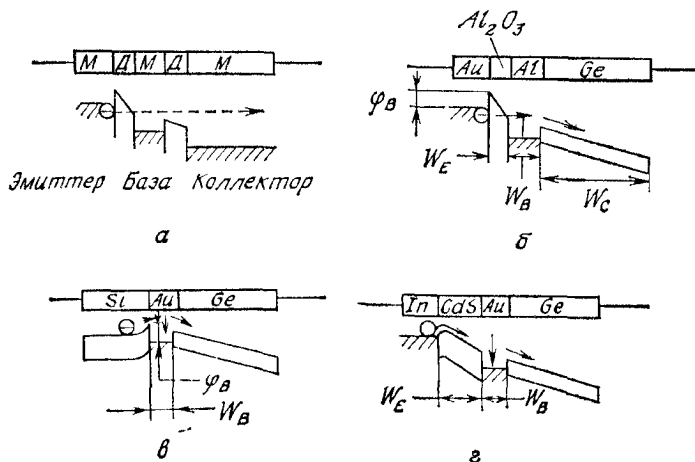


Рис. 39. Транзисторы на горячих электронах [67].

a — МДМДМ-структура; *б* — туннельный транзистор; *в* — транзистор с металлической базой; *г* — транзистор с ограничением тока пространственным зарядом.

диэлектрик—металл (МДМДМ), в котором протекание тока в слое диэлектрика происходило за счет туннелирования (рис. 39, *a*), известен с 1960 г. [66]. Лучшие характеристики имеет прибор с коллектором в виде диода Шоттки (рис. 39, *б*). Дальнейшее улучшение прибора связано с использованием эмиттера в виде барьера Шоттки (рис. 39, *в*). Предлагалось также использовать эмиттер в режиме ограничения тока пространственным зарядом (рис. 39, *г*).

Основное отличие этих транзисторов от других приборов заключается в способе инжекции электронов в базу [67]. Только транзистор с металлической базой обладает потенциально лучшими СВЧ-характеристиками, чем биполярный транзистор. Однако экспериментально полученные значения α_0 при комнатной температуре малы ($\sim 0,3$). Эти результаты получены для транзистора со структурой Si—Au—Ge при толщине золотой пленки 90 Å [68]. Использование перспективных технологических методов (молекулярно-лучевой эпитаксии, выращивания монокристаллических пленок металла на полупроводниках [69] и переход на работу при низких температурах) позволяет надеяться на получение хороших высокочастотных характеристик и высоких коэффициентов усиления по току.

3.6.5. Транзисторы с проникаемой базой [70]

Трехмерное изображение транзистора с проникаемой базой приведено на рис. 40. Прибор имеет четырехслойную структуру: подложку из GaAs n^+ -типа, слой эмиттера из GaAs n -типа, фигур-

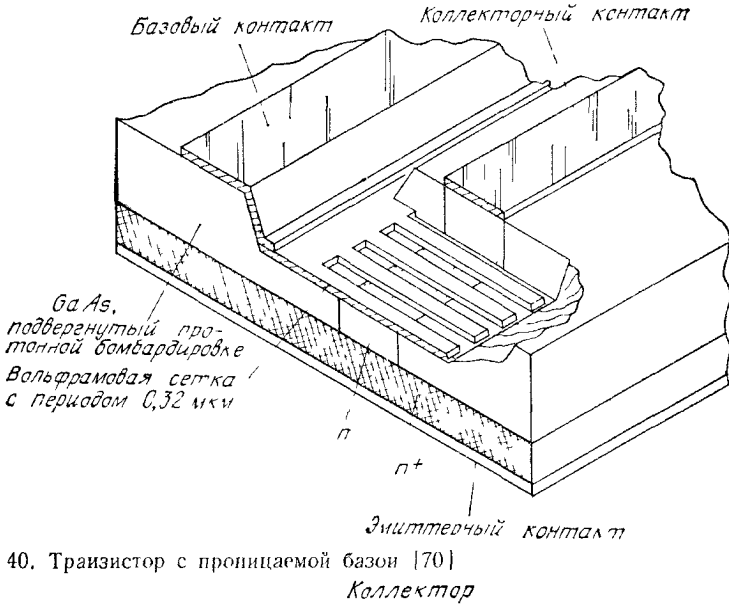


Рис. 40. Транзистор с пористой базой [70]

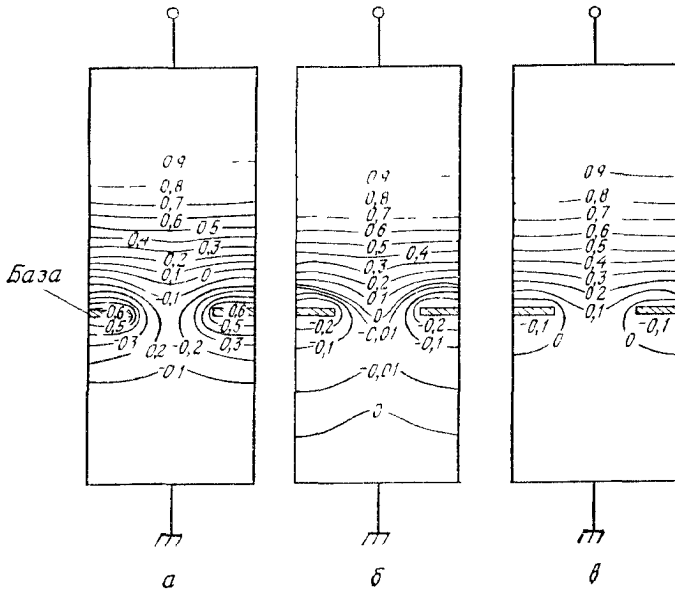


Рис. 41. Эквипотенциальные линии отдельной ячейки транзистора с пористой базой при различных условиях смещения [70].

а — при $V_{BE} = 0$ В; б — при $V_{BE} = 0,3$ В; в — при $V_{BE} = 0,5$ В.

ную металлическую сетку (из вольфрама толщиной 200 Å), образующую барьер Шоттки высотой 0,8 В, и слой коллектора *n*-типа. Ширина металлических полосок и зазоры между ними составляют 1600 Å. При изготовлении транзистора с проникаемой базой были использованы рентгенолитография и эпитаксиальное наращивание полупроводникового слоя над металлическим слоем. На рис. 41 приведены теоретически полученные эквипотенциальные линии в поперечном сечении прибора для трех различных условий смещения. На рис. 41, *a* к коллектору приложено напряжение 1 В, а база и эмиттер заземлены. При движении от эмиттера к коллектору электроны должны пройти через область с отрицательным потенциалом, т. е. над потенциальным барьером. В центре зазора между металлическими полосками барьер наиболее низкий (0,25 В), а около границы металл—полупроводник барьер наиболее высокий (0,6 В). Наличие высокого барьера обуславливает малую плотность тока в цепи коллектора (~ 5 А/см²). Однако, если сместить базу в прямом направлении на 0,3 В (рис. 41, *б*), барьер понизится и плотность коллекторного тока возрастет до 900 А/см². При дальнейшем повышении смещения на базе накопление носителей приведет к ограничению тока пространственным зарядом (рис. 41, *в*). Коэффициент усиления по току в схеме с общей базой при смещении на базе ниже 0,4 В остается выше 0,99 благодаря потенциалу барьера Шоттки. Когда напряжение на базе достигает 0,5 В, величина α_0 уменьшается до 0,96, а при дальнейшем повышении напряжения на базе α_0 резко снижается. При базовой сетке с достаточно малыми межэлектродными зазорами и при соответствующих концентрациях носителей за счет барьерного ограничения тока можно получить высокие плотности управляемых токов и, как следствие, большую крутизну и высокую предельную частоту $f_{\text{макс}}$ прибора. Название «транзистор с проникаемой базой» происходит из аналогии с биполярным транзистором (протекание тока, управляемого барьером) и из аналогии (по структуре) с транзистором с металлической базой [70]. На первых образцах приборов [70] были получены $f_{\text{макс}} \approx 17$ ГГц, коэффициент усиления 13 дБ на 4 ГГц, коэффициент шума 3,5 дБ на 4 ГГц. Численное моделирование указывает на возможность значительного увеличения $f_{\text{макс}}$ при соответствующем масштабировании размеров.

ЛИТЕРАТУРА

1. Bardeen J., Brattain W. H. The Transistor, A Semiconductor Triode, *Phys. Rev.*, 74, 230 (1948).
2. Shockley W. The Theory of *p* — *n* Junctions in Semiconductors and *p* — *n* Junction Transistors, *Bell Syst. Tech. J.*, 28, 435 (1949).
3. Hall R. N., Dunlap W. C. *p* — *n* Junctions Prepared by Impurity Diffusion, *Phys. Rev.*, 80, 467 (1950).

4. Teal G. K., Sparks M., Buehler E. Growth of Germanium Single Crystals Containing $p - n$ Junctions, *Phys. Rev.*, **81**, 637 (1951).
5. Piann W. H. Principles of Zone-Refining, *Trans. AIME*, **194**, 747 (1952).
6. Tanenbaum M., Thomas D. E. Diffused Emitter and Base Silicon Transistor, *Bell Syst. Tech. J.*, **35**, 1 (1956).
7. Theuerer H. C., Kleimack J. J., Loar H. H., Christenson H. Epitaxial Diffused Transistors, *Proc. IRE*, **48**, 1642 (1960).
8. Hoerni J. A. Planar Silicon Transistor and Diodes, IRE Electron Devices Meet., Washington, D. C., 1960.
9. Lespelter M. P. Beam-Lead Technology, *Bell Syst. Tech. J.*, **45**, 233 (1966).
10. Shockley W., U. S. Patent 2,787,564 (1954). For a view on ion implantation, see Gibbons J. F. Ion Implantation in Semiconductors — Part I, Range Distribution Theory and Experiments, *Proc. IEEE*, **56**, 295 (1968).
11. Labuda E. F., Clemens J. T. Integrated Circuit Technology, in Kirk R. E., Othmer D. F., Eds., *Encyclopedia of Chemical Technology*, Wiley, N. Y., 1980.
12. Phillips A. B. Transistor Engineering, McGraw-Hill, N. Y., 1962.
13. Gartner W. W. Transistors, Principle, Design and Application, D. Van Nostrand, Princeton, N. Y., 1960.
14. SEEC (Semiconductor Electronics Education Committee), 4 vols.
 - (1) Adler R. B., Smith A. C., Longini R. L. Introduction to Semiconductor Physics, SEEC, Vol. 1, Wiley, N. Y., 1966.
 - (2) Gray P. E., DeWitt D., Boothroyd A. R., Gibbons J. F. Physical Electronics and Circuit Models of Transistors, SEEC, Vol. 2, Wiley, N. Y., 1966.
 - (3) Searle C. L., Boothroyd A. R., Angelo E. J., Gray P. E., Pederson D. O. Elementary Circuit Properties of Transistors, SEEC, Vol. 3, Wiley, N. Y., 1966.
 - (4) Thornton R. D., DeWitt D., Chenette E. R., Gray P. E. Characteristics and Limitations of Transistors, SEEC, Vol. 4, Wiley, N. Y., 1966.
15. Pritchard R. L. Electrical Characteristics of Transistors, McGraw-Hill, N. Y., 1967.
16. Ghandhi S. K. Semiconductor Power Devices, Wiley, N. Y., 1977.
17. Muller R. S., Kamins T. I. Device Electronics for Integrated Circuits, Wiley, N. Y., 1977.
18. Johnson E. O. Physical Limitations on Frequency and Power Parameters of Transistors, IEEE Int. Conv. Rec., Pt. 5, p. 27 (1965).
19. Moll J. L., Ross I. M. The Dependence of Transistor Parameters on the Distribution of Base Layer Resistivity, *Proc. IRE*, **44**, 72 (1956).
20. Jespers P. G. A. Measurements for Bipolar Devices, in Van de Wiele F., Engl W. L., Jespers P. G., Eds., *Process and Device Modeling for Integrated Circuit Design*, Noordhoff, Leyden, 1977.
21. Gummel H. K. Measurement of the Number of Impurities in the Base Layer of a Transistor, *Proc. IRE*, **49**, 834 (1961).
22. Payne R. S., Scavuzzo R. J., Olson K. H., Nacci J. M., Moline R. A. Fully Ion-Implanted Bipolar Transistors, *IEEE Trans. Electron Devices*, **ED-21**, 273 (1974).
23. Sah C. T., Noyce R. N., Shockley W. Carrier Generation and Recombination in $p - n$ Junction and $p - n$ Junction Characteristics, *Proc. IRE*, **45**, 1228 (1957).
24. Werner W. M. The Influence of Fixed Interface Charges on Current Gain Fal-lout of Planar $n - p - n$ Transistors, *J. Electrochem. Soc.*, **123**, 540 (1976).
25. Webster W. M. On the Variation of Junction-Transistor Current Amplification Factor with Emitter Current, *Proc. IRE*, **42**, 914 (1954).
26. Lanyon H. P. D., Tuft R. A. Bandgap Narrowing in Heavily Doped Silicon, IEEE Tech. Dig., Int. Electron Device Meet., 1978, p. 316.

27. McGrath E. J., Navon D. H. Factors Limiting Current Gain in Power Transistors, *IEEE Trans. Electron Devices*, **ED-24**, 1255 (1977).
28. Poon H. C., Gummel H. K., Scharfetter D. L. High Injection in Epitaxial Transistors, *IEEE Trans. Electron Devices*, **ED-16**, 455 (1969).
29. Kirk C. T. A Theory of Transistor Cutoff Frequency (f_T) Fall-Off at High Current Density, *IEEE Trans. Electron Devices*, **ED-9**, 164 (1962).
30. Morant M. J. Introduction to Semiconductor Devices, Addison-Wesley, Reading, Mass., 1964.
31. Early J. M. Effects of Space-Charge Layer Widening in Junction Transistors, *Proc. IRE*, **40**, 1401 (1952).
32. Ebers J. J., Moll J., L. Large-Signal Behavior of Junction Transistors, *Proc. IRE*, **42**, 1761 (1954).
33. deGraaff H. C. Review of Models for Bipolar Transistors, in Ref. 20.
34. Getreu I. E. Modeling the Bipolar Transistor, Elsevier, N. Y., 1978.
35. Gummel H. K., Poon H. C. An Integral Charge Control Model of Bipolar Transistors, *Bell Syst. Tech. J.*, **49**, 827 (1970).
36. Tsukamoto K., Akasaka Y., Watari Y., Kusano Y., Hirose Y., Nakamura G. Arsenic-Implanted Emitter and Its Application to UHF Power Transistors, *Jpn. J. Appl. Phys.*, **17**, *Suppl.*, 17-1, 187 (1977).
37. Edwards R. Fabrication Control Is Key to Microwave Performance, *Electronics*, **41**, 109 (Feb. 1968).
38. Wang A. C. M., Kakihana S. Leakage and h_{FE} Degradation in Microwave Bipolar Transistors, *IEEE Trans. Electron Devices*, **ED-21**, 667 (1974).
39. Parrillo L. C., Payne R. S., Seidel T. F., Robinson McD., Reutlinger G. W., Post D. E., Field R. L. The Reduction of Emitter-Collector Shorts in a High-Speed, All Implanted, Bipolar Technology, *IEEE Tech. Dig., Int. Electron Device Meet.*, 1979, p. 348.
40. Pritchard R. L., Angell J. B., Adler R. B., Early J. M., Webster W. M. Transistor Internal Parameters for Small-Signal Representation *Proc. IRE*, **49**, 725 (1961).
41. Moll J. L. Physics of Semiconductors, McGraw-Hill, N. Y., 1964.
42. Daw A. N., Mitra R. N., Choudhury N. K. D. Cutoff Frequency of a Drift Transistor, *Solid State Electron.*, **10**, 359 (1967).
43. Kurokawa K. Power Waves and the Scattering Matrix, *IEEE Trans. Microwave Theory Tech.*, **MTT-13**, 194 (1965).
44. S-Parameter Techniques for Faster, More Accurate Network Design, Application Note 95-1, *Hewlett-Packard J.*, **18**, No. 6 (Feb. 1967).
45. Nielson E. G. Behavior of Noise Figure in Junction Transistors, *Proc. IRE*, **45**, 957 (1957).
46. Cooke H. F. Microwave Transistors: Theory and Design, *Proc. IEEE*, **59**, 1163 (1971).
47. Yau L. D., Tsai T. N. Fabrication of a Low-Noise Beam-Leaded Microwave Bipolar Transistor by Electron and Photolithography, *IEEE Trans. Electron Devices*, **ED-25**, 413 (1978).
48. Sakai T., Sunohara Y., Sakakibara Y., Murota J. Stepped Electrode Transistor, *Jpn. J. Appl. Phys.*, **16**, *Suppl.*, 16-1, 43 (1977).
49. Archer J. A. Low-Noise Implanted-Base Microwave Transistors, *Solid State Electron.*, **17**, 387 (1974).
50. Allison R. Silicon Bipolar Microwave Power Transistors, *IEEE Trans. Microwave Theory Tech.*, **MTT-27**, 415 (1979).
51. Gaur S. P., Navon D. H., Teerlinck R. W. Transistor Design and Thermal Stability, *IEEE Trans. Electron Devices*, **ED-20**, 527 (1973).
52. Arnold R. P., Zoroglu D. S. A Quantitative Study of Emitter Ballasting, *IEEE Trans. Electron Devices*, **ED-21**, 385 (1974).
53. Thornton C. G., Simmons C. D. A New High Current Mode of Transistor Operation, *IRE Trans. Electron Devices*, **ED-5**, 6 (1958).

54. Schafft H. A. Second-Breakdown — A Comprehensive Review, *Proc. IEEE*, **55**, 1272 (1967).
55. Klein N. Electrical Breakdown in Solids, in Marton L., Eds., *Advances in Electronics and Electron Physics*. Academic, N. Y., 1968.
56. Dunn L., Nuttall K. I. An Investigation of the Voltage Sustained by Epitaxial Transistors in Current Mode Second Breakdown, *Int. J. Electron.*, **45**, 353 (1978).
57. Melchior H., Strutt M. J. O. Secondary Breakdown in Transistors, *Proc. IEEE*, **52**, 439 (1964).
58. Oettinger F. F., Blackburn D. L., Rubin S. Thermal Characterization of Power Transistors, *IEEE Trans. Electron Devices*, **ED-23**, 831 (1976).
59. Moll J. L. Large-Signal Transient Response of Junction Transistors, *Proc. IRE*, **42**, 1773 (1954).
60. Rice D. Isoplanar-S Scales Down for New Heights in Performance, *Electronics*, **52**, 137 (1979).
61. Hart K., Slob A. Integrated Injection Logic — A New Approach to LSI, *ISSCC Dig. Tech. Pap.*, p. 92 (1972); *IEEE J. Solid State Circuits*, **SC-7**, 346 (1972).
62. Berger H. H., Widemann S. K. Merged Transistor Logic — A Low-Cost Bipolar Logic Concept, *ISSC Dig. Tech. Pap.*, p. 90 (1972); also *IEEE J. Solid State Circuits*. **SC-7**, 340 (1972).
63. Berger H. H., Wiedmann S. K. Advanced Merged Transistor Logic by Using Schottky Junctions, *Microelectronics*, **7**, 35 (1976).
64. Shockley W. U. S. Patent 2569347 (1951).
65. Konagai M., Takahashi K. GaAlAs — GaAs Heterojunction Transistors with High Injection Efficiency, *J. Appl. Phys.*, **46**, 2120 (1975).
66. Mead C. A. Tunnel-Emission Amplifiers, *Proc. IRE*, **48**, 359 (1960).
67. Moll J. L. Comparison of Hot Electrons and Related Amplifiers, *IEEE Trans. Electron Devices*, **ED-10**, 299 (1963).
68. Sze S. M., Gummel H. K. Appraisal of Semiconductor — Metal — Semiconductor Transistors, *Solid State Electron.*, **9**, 751 (1966).
69. Cho A. Y., Dernier P. D. Single-Crystal-Aluminium Schottky-Barrier Diodes Prepared by Molecular-Beam Epitaxy (MBE) on GaAs, *J. Appl. Phys.*, **49**, 3328 (1978).
70. Bozler C. O., Alley G. D., Murphy R. A., Flanders D. C., Lindley W. T. Fabrication and Microwave Performance of the Permeable Base Transistor, *IEEE Tech. Dig.*, *Int. Electron Device Meet.*, 1979, p. 384.

Глава 4

ТИРИСТОРЫ

4.1. ВВЕДЕНИЕ

Тиристорами называют большое семейство полупроводниковых приборов, которые обладают бистабильными характеристиками и способны переключаться из одного состояния в другое. В одном состоянии тиристор имеет высокое сопротивление и малый ток (закрытое, или выключенное состояние), в другом — низкое сопротивление и большой ток (открытое, или включенное состояние). Принцип действия тиристора тесно связан с принципом действия биполярного транзистора, в котором и электроны, и дырки участвуют в механизме проводимости. Название «тиристор» произошло от слова «тиратрон», поскольку электрические характеристики обоих приборов во многом аналогичны.

Используя введенное в 1950 г. Шокли [1] понятие транзистора с ловушкой в коллекторе, Эберс предложил двухтранзисторную модель для объяснения характеристик простейшего тиристора — четырехслойного $p-n-p-n$ прибора [2]. Подробно принцип действия и первые работающие $p-n-p-n$ -приборы описаны в работе [3], которая послужила основой для всех последующих работ по исследованию тиристорov. Благодаря наличию двух устойчивых состояний и низкой мощности рассеяния в этих состояниях тиристоры обладают уникальными полезными свойствами, позволяющими использовать их для решения широкого диапазона задач (от регулирования мощности в домашних бытовых электроприборах до переключения и преобразования энергии в высоковольтных линиях электропередачи). В настоящее время созданы тиристоры, работающие при токах от нескольких миллиампер до 5000 А и выше и при напряжениях, превышающих 10 000 В [4]. Принцип действия и способы применения тиристорov описаны в работах [5, 6]. Подробное описание функционирования и технологии изготовления тиристорov можно найти в монографии Гапди [7].

4.2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

Основная схема тиристорной структуры показана на рис. 1. Она представляет собой четырехслойный $p-n-p-n$ -прибор, содержащий три последовательно соединенных $p-n$ -перехода J1, J2 и J3. Контакт к внешнему p -слою назван анодом, а контакт к внешнему n -слою — катодом. В общем случае $p-n-p-n$ -

прибор может иметь два управляющих электрода (называемых также базами), подсоединенных к внутренним p - и n -слоям. Прибор без управляющих электродов (рис. 1, б) работает как двухполюсник со структурой $p-n-p-n$ и называется диодным тристором (а также динистором и диодом Шокли). Прибор с одним управляющим электродом (рис. 1, в) является трехполюсником и называется триодным тристором или просто тристором (также используют термин «полупроводниковый управляемый вентиль» (ПУВ)).

Типичный профиль легирующей примеси в диффузионно-сплавном приборе показан на рис. 2, а. В качестве исходного материала выбрана подложка n -типа. Диффузией с обеих сторон подложки одновременно создают слои p_1 и p_2 . На заключительной стадии путем сплавления (или диффузии) с одной стороны подложки создают слой n_2 . Поперечный разрез тристора, предназначенного для работы с умеренными токами, приведен на рис. 2, б [5]. Медное основание в виде головки болта служит для отвода тепла.

Вольт-амперная характеристика тристора (с управляющими электродами или без них), приведенная на рис. 3, имеет несколько различных участков. Участок характеристики между точками 0 и 1 соответствует закрытому состоянию (прямому запираанию) с очень высоким сопротивлением. Прямое переключение (или включение) тристора происходит при выполнении условия $dV/dI = 0$ (ему соответствуют напряжение включения V_{BF}

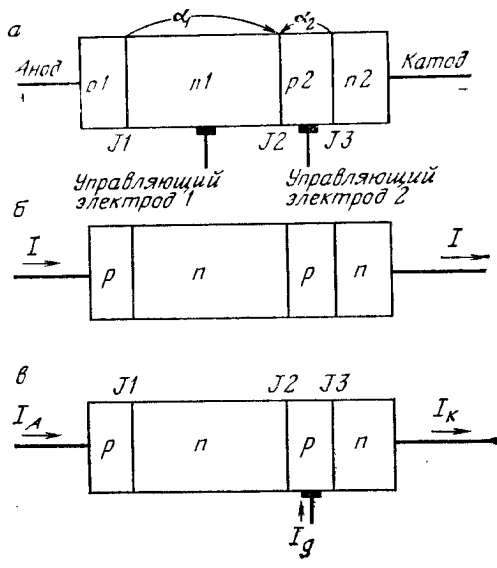


Рис. 1. Схемы тристора.

а — основная четырехслойная $p-n-p-n$ -структура; α_1 — коэффициент усиления по току $p-n-p$ -транзистора, α_2 — коэффициент усиления по току $n-p-n$ -транзистора. При нормальном смещении центральный переход обратно смещен и служит коллектором для $p-n-p$ - и $n-p-n$ -транзисторов; б — диодный тристор; в — триодный тристор.

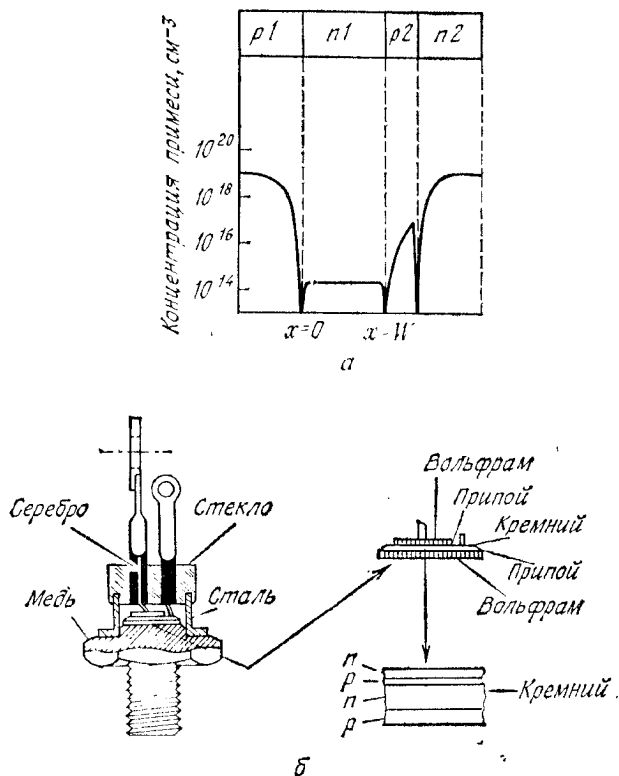


Рис. 2. Типичный профиль легирования (а). Наиболее важными параметрами являются концентрация примеси и ширина базы n_1 . Разрез тиристора, предназначенного для переключения умеренных токов (б) [5].

и ток включения I_n (рис. 3)). Между точками 1 и 2 находится участок характеристики с отрицательным сопротивлением; участок между точками 2 и 3 соответствует открытому состоянию (прямой проводимости). В точке 2, где снова $dV/dI = 0$, через прибор протекает минимальный удерживающий ток I_n , а на приборе показано минимальное удерживающее напряжение V_n . Участок между точками 0 и 4 описывает режим обратного запирания прибора, а участок между точками 4 и 5 — режим обратного пробоя.

Таким образом, тиристор в области прямых смещений является бистабильным элементом, способным переключаться из закрытого состояния с высоким сопротивлением и малым током в открытое состояние с низким сопротивлением и большим током и наоборот. Ниже рассмотрены основные режимы работы тиристора.

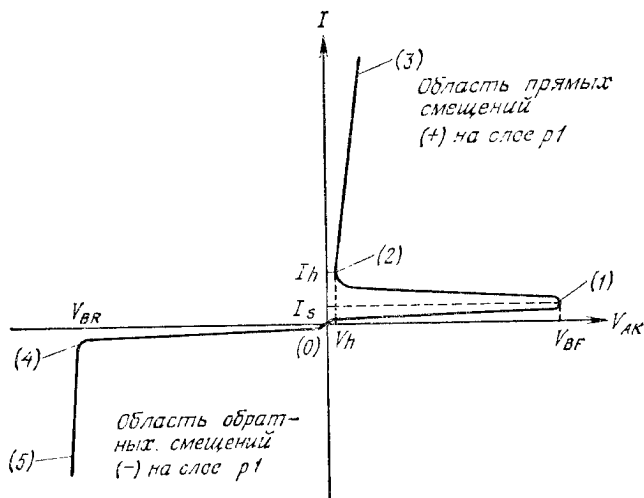


Рис. 3. Вольт-амперная характеристика тиристора.

4.2.1. Режим обратного запирания

Напряжение пробоя. Два основных фактора ограничивают напряжение обратного и прямого пробоя: лавинный пробой и прокол обедненной области.

В режиме обратного запирания к аноду прибора приложено напряжение, отрицательное по отношению к катоду; переходы J1 и J3 смещены в обратном направлении, а переход J2 смещен в прямом (рис. 4, а). Если профиль примеси такой, как на рис. 2, а, то большая часть приложенного напряжения падает на переходе J1. В зависимости от толщины W_{n1} слоя $n1$ пробой вызывается лавинным умножением (толщина обедненной области при пробое меньше W_{n1}) либо проколом (обедненный слой распространяется на всю область $n1$, и происходит смыкание переходов J1 и J2).

Для несимметричного резкого $p^+ - n$ -перехода на кремнии с сильнолегированной областью $p1$ напряжение лавинного пробоя при комнатной температуре описывается выражением, приведенным в гл. 2 [7, 8]:

$$V_B = 5,34 \cdot 10^{13} (N_{n1})^{-0,75} [\text{В}], \quad (1)$$

где N_{n1} — концентрация примеси в области $n1$. Для плавного перехода напряжение лавинного пробоя равно [7, 8]

$$V_B = 9,17 \cdot 10^9 a^{-0,4} [\text{В}], \quad (2)$$

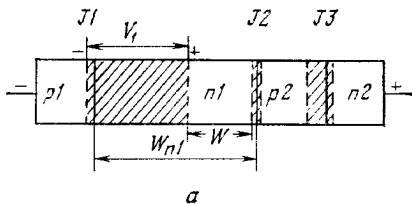
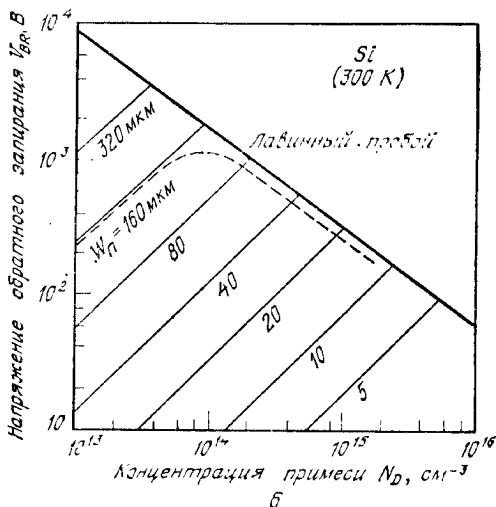


Рис. 4. Режим обратного запирающего тиристора. Линия лавинного пробоя указывает максимально допустимое напряжение на слое $n1$ в зависимости от концентрации примеси. Параллельные линии соответствуют проколу слоя $n1$ в зависимости от его ширины [9].



где a — градиент концентрации примеси, см^{-4} . Напряжение пробоя несимметричного резкого перехода

$$V_{PT} \approx \frac{qN_{n1}W_{n1}^2}{2\epsilon_s}. \quad (3)$$

На рис. 4, б приведены основные ограничения режима обратного запирающего кремниевых тиристоров [9]. Например, для $W_{n1} = 160$ мкм максимальное напряжение пробоя не превышает 2000 В и соответствует $N_{n1} = 8,5 \cdot 10^{13} \text{ см}^{-3}$; для меньших концентраций напряжение пробоя снижается из-за пробоя, а при больших концентрациях падает из-за лавинного умножения.

В действительности напряжение обратного запирающего меньше указанных пределов вследствие того, что переход $J1$ связан с соседним переходом $J2$, образуя $p-n-p$ транзистор с разомкнутой базой, включенный по схеме с общим эмиттером; работа этого транзистора снижает напряжение пробоя. Условия обратного пробоя для конфигурации с общим эмиттером соответствуют

коэффициенту лавинного умножения $M = 1/\alpha_1$, и напряжение пробоя описывается формулой (гл. 3)

$$V_{BR} = V_B (1 - \alpha_1)^{1/n}, \quad (4)$$

где α_1 — коэффициент усиления транзистора по току в схеме с общей базой, V_B — напряжение лавинного пробоя $p1-n1$ -перехода, n — постоянная, равная ~ 6 для p^+-n -диодов и 4 для $n-p$ -диодов. Поскольку выражение $(1 - \alpha_1)^{1/n} < 1$, напряжение обратного пробоя тиристора меньше V_B . В большинстве практических случаев коэффициент инжекции транзистора $\gamma \approx 1$, так как область $p2$ сильно легирована. Следовательно, коэффициент усиления по току равен коэффициенту переноса по базе α_T :

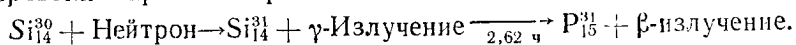
$$\alpha_1 = \gamma \alpha_T \approx \alpha_T = \operatorname{sech}(W/L_{n1}), \quad (5)$$

где L_{n1} — диффузионная длина дырок в области $n1$ и

$$W = W_{n1} [1 - (V/V_{PT})^{1/2}]. \quad (6)$$

Для заданных значений W_{n1} и L_{n1} отношение W/L_{n1} уменьшается с ростом напряжения обратного смещения. Следовательно, вклад коэффициента переноса по базе в общее усиление становится более весомым по мере приближения напряжения к пределу, обусловленному проколом. На рис. 4, б показан (штриховой линией) пример снижения напряжения обратного запираения для $W_{n1} = 160$ мкм и $L_{n1} = 150$ мкм. Отметим, что напряжение V_{BR} приближается к V_{PT} при меньших концентрациях примеси в области $n1$. При возрастании концентрации напряжение V_{BR} всегда несколько ниже V_B из-за конечной величины W/L_{n1} .

Нейтронное легирование [10]. Мощные высоковольтные тиристоры занимают большую площадь; часто целиком вся пластина (диаметром 100 мм и более) служит подложкой одного прибора. Такие размеры предъявляют очень жесткие требования к однородности исходного материала. Для снижения разброса удельного сопротивления и повышения однородности распределения легирующих примесей применяют способ нейтронного облучения. Обычно используют кремниевые пластины со средним удельным сопротивлением, значительно большим по величине, чем требуется. Затем пластина облучается тепловыми нейтронами. В результате происходят превращение части кремния в фосфор и легирование кремния примесью n -типа:



Период полураспада составляет 2,62 ч. Поскольку глубина проникновения нейтронов в кремний равна ~ 100 см, примесь распределяется по пластине очень равномерно. На рис. 5 приведены распределения продольного макроскопического удельного сопротивления кремния, легированного обычным способом и нейтрон-

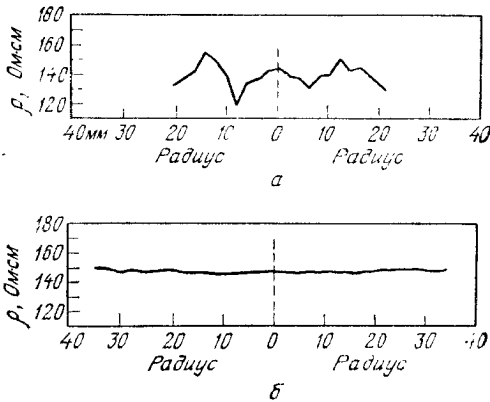


Рис. 5. Типичное продольное распределение удельного сопротивления в кремнии, легированном обычным способом (а) и нейтронным облучением (б) [11].

ным облучением [11]. Для кремния, легированного обычным способом, неоднородность удельного сопротивления составляла $\sim \pm 15\%$, а при нейтронном легировании $\sim \pm 1\%$.

Структуры с фаской. Чтобы получить тиристор с максимальным напряжением пробоя, обычно стараются изготовить сплавлением или диффузией плоские $p-n$ -переходы, так как напряжения пробоя цилиндрических и сферических переходов ниже (гл. 2). Но даже и в плоскостных переходах наблюдается преждевременный пробой по поверхности, который приводит к увеличению тока на краю прибора и снижению динамических токовых характеристик тиристора. В структурах с фаской поверхностные электрические поля существенно ниже по сравнению с полями в объеме материала. В результате пробой однородно распределен по объему.

Схемы структуры с фаской приведены на рис. 6. Положительная фаска означает уменьшение площади поперечного сечения в направлении от сильнолегированной области к слаболегированной области (рис. 6, а). Отрицательная фаска, наоборот, означает увеличение площади в том же направлении (рис. 6, б). Две тиристорные структуры с фаской показаны на рис. 6, в и г. Прибор на рис. 6, в имеет отрицательные фаски для переходов J2 и J3 и положительную фаску для перехода J1. Структура на рис. 6, г имеет положительные фаски для всех трех переходов [12].

В переходах с положительной фаской поле на поверхности снижено в $\sin\theta$ раз (в первом приближении). На рис. 7 приведены значения электрического поля на поверхности, полученные путем решения двумерного уравнения Пуассона для p^+-n -перехода при обратном смещении 600 В [13], и значения внутреннего электрического поля в объеме. Отметим, что максимум поля на поверхности всегда меньше, чем в объеме; с уменьшением угла

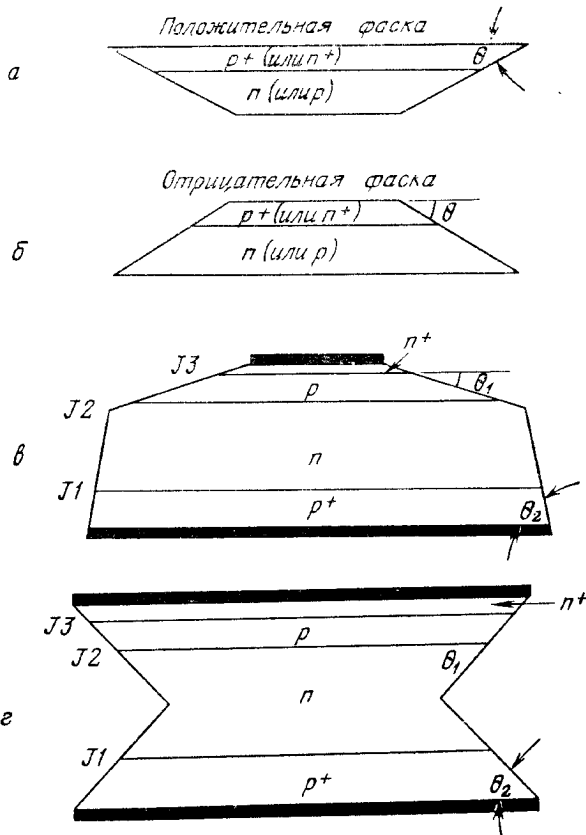


Рис. 6. Схемы тиристоров с фаской.

а — положительная фаска; б — отрицательная фаска; в — тиристор с отрицательной и положительной фасками; г — тиристор с двумя положительными фасками.

фаски снижается пик электрического поля и его положение сдвигается в глубь слаболегированной области. Напряжение пробоя для перехода с положительной фаской такое же, как для плоскостного перехода.

Чтобы в переходе с отрицательной фаской пробой происходил внутри объема, сам переход должен быть сильноасимметричен, а угол фаски — мал по величине. Численные расчеты показывают, что для отрицательных фасок с малыми углами максимум поверхностного поля меньше, чем в объеме, и расположить в p -области (концентрация в слое $p1$ выше, чем в слое $n1$). Внутренний пробой наступает при напряжениях, более низких, чем в плоском переходе. На рис. 8 приведена зависимость нормализованного на-

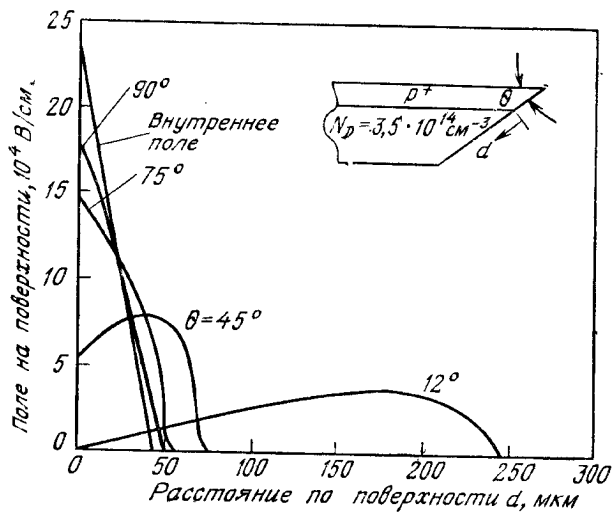


Рис. 7. Зависимость поверхностного и внутреннего полей в приборе с положительной фаской (показанного на вставке) от угла фаски [13].

пряжения объемного пробоя (отнесенного к напряжению пробоя плоского перехода) от эффективного угла фаски

$$0_{\text{eff}} = (0,04) \theta \left(\frac{W}{d} \right)^2, \quad (7)$$

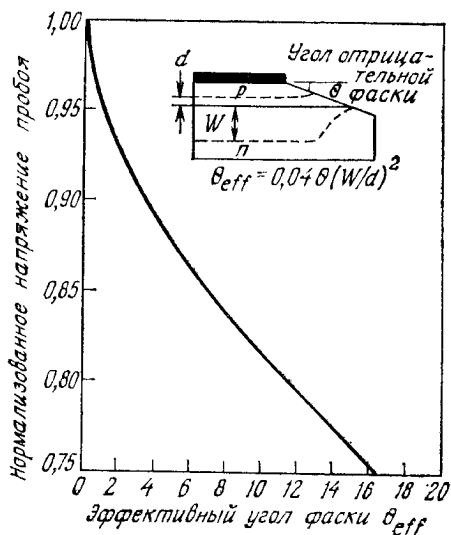


Рис. 8. Нормализованное напряжение пробоя в зависимости от угла отрицательной фаски структуры, показанной на вставке [14].

где W и d — соответствующие значения глубин обедненных слоев перехода, показанного на вставке [14]. Чтобы приблизиться к напряжению пробоя плоского перехода, необходимо использовать фаски с малыми углами.

4.2.2. Режим прямого запираания

Влияние коэффициентов. При прямом запираании напряжение на аноде положительно по отношению к катоду и обратно смещен только переход J2. Переходы J1 и J3 смещены в прямом направлении. Большая часть приложенного напряжения падает на переходе J2. Для объяснения характеристик прибора в режиме прямого запираания используем двухтранзисторную модель [2]. Из рис. 1 следует, что тиристор можно рассматривать как соединение $p-n-p$ -транзистора с $n-p-n$ -транзистором, причем коллектор каждого из них соединен с базой другого, как показано на рис. 9, а и б для триодного тиристора. Центральный переход действует как коллектор дырок, инжектируемых переходом J1, и электронов, инжектируемых переходом J3. Взаимосвязь между токами эмиттера I_E , коллектора I_C и базы I_B и статическим коэффициентом усиления по току α_1 $p-n-p$ -транзистора при-

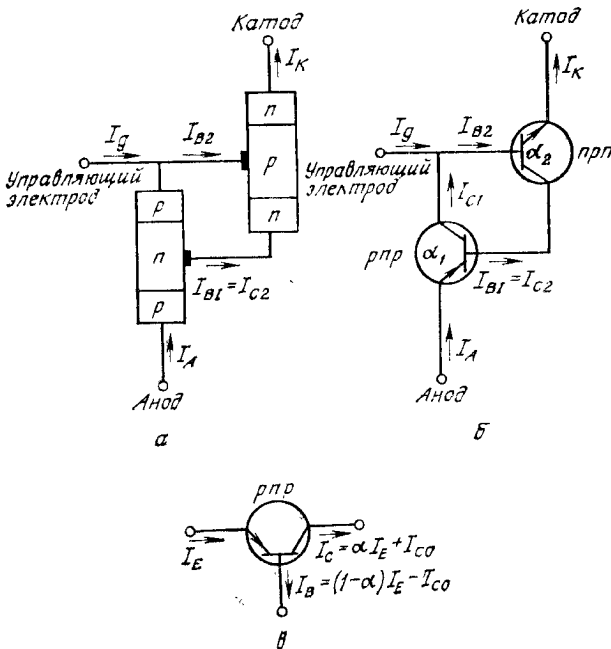


Рис. 9. Двухтранзисторная модель триодного тиристора (а), соединение транзисторов (б) и соотношение токов в $p-n-p$ -транзисторе (в) [2].

ведена на рис. 9, в, где I_{CO} — обратный ток насыщения перехода коллектор—база. Аналогичные соотношения можно получить для $n-p-n$ -транзистора при изменении направления токов на противоположное. Из рис. 9, б следует, что коллекторный ток $n-p-n$ -транзистора является одновременно базовым током $p-n-p$ -транзистора. Аналогично коллекторный ток $p-n-p$ -транзистора и управляющий ток I_g втекают в базу $n-p-n$ -транзистора. В результате, когда общий коэффициент усиления в замкнутой петле превысит 1, оказывается возможным регенеративный процесс.

Ток базы $p-n-p$ -транзистора равен

$$I_{B1} = (1 - \alpha_1) I_A - I_{CO1}. \quad (8)$$

Этот ток протекает также через коллектор $n-p-n$ -транзистора. Ток коллектора $n-p-n$ -транзистора с коэффициентом усиления α_2 равен

$$I_{C2} = \alpha_2 I_K + I_{CO2}. \quad (9)$$

Приравняв I_{B1} и I_{C2} , получим

$$(1 - \alpha_1) I_A - I_{CO1} = \alpha_2 I_K + I_{CO2}.$$

Так как $I_K = I_A + I_g$, то из формулы (9) следует

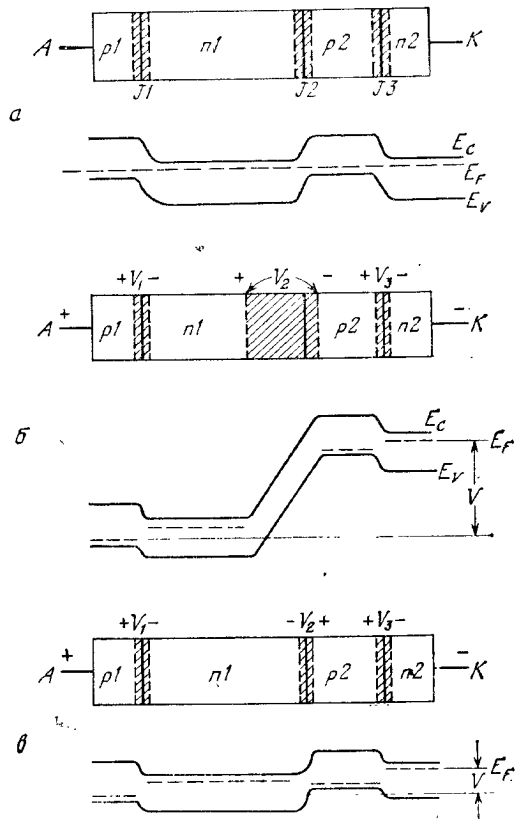
$$I_A = \frac{\alpha_2 I_g + I_{CO1} + I_{CO2}}{1 - (\alpha_1 + \alpha_2)}. \quad (10)$$

Ниже показано, что α_1 и α_2 зависят от I_A и обычно растут с увеличением тока. Уравнение (10) описывает статическую характеристику прибора в диапазоне напряжений вплоть до пробоя. После пробоя прибор работает как $p-i-n$ -диод. Отметим, что все слагаемые в числителе правой части уравнения (10) малы, следовательно, пока член $(\alpha_1 + \alpha_2) < 1$, ток I_A мал. Если $\alpha_1 + \alpha_2 = 1$, знаменатель дроби обращается в нуль и происходит прямой пробой (или включение тиристора). Следует отметить, что если полярность напряжения между анодом и катодом сменить на обратную, то переходы J1 и J3 будут смещены в обратном направлении, а J2 — в прямом. При таких условиях пробой не происходит, так как в качестве эмиттера работает только центральный переход и регенеративный процесс становится невозможным.

Ширина обедненных слоев и энергетические зонные диаграммы в равновесии, в режимах прямого запираания и прямой проводимости показаны на рис. 10. В равновесии обедненная область каждого перехода и контактный потенциал определяются профилем распределения примесей. Когда к аноду приложено положительное напряжение, переход J2 стремится сместиться в обратном направлении, а переходы J1 и J3 — в прямом. Падение

Рис. 10. Энергетическая зонная диаграмма в режиме прямого смещения.

a — состояние равновесия; *б* — закрытое состояние, когда большая часть приложенного напряжения падает на центральном переходе; *в* — открытое состояние, когда все три перехода смещены прямо.



напряжения между анодом и катодом равно алгебраической сумме падений напряжения на переходах:

$$V_{AK} = V_1 + V_2 + V_3. \quad (11)$$

По мере повышения напряжения возрастает ток через прибор и, следовательно, увеличиваются α_1 и α_2 . Благодаря регенеративному характеру этих процессов прибор в конце концов перейдет в открытое состояние. После включения тиристора протекающий через него ток должен быть ограничен внешним сопротивлением нагрузки; в противном случае при достаточно высоком напряжении тиристор выйдет из строя. Во включенном состоянии переход J2 смещен в прямом направлении (рис. 10, *в*), и падение напряжения $V_{AK} = (V_1 - |V_2| + V_3)$ приблизительно равно сумме напряжения на одном прямосмещенном переходе и напряжения на насыщенном транзисторе.

Переключение тиристора происходит при $dV_{AK}/dI_A = 0$. Обычно это условие достигается раньше, чем $(\alpha_1 + \alpha_2) = 1$. Можно показать, что переключение начинается, когда сумма малосигнальных коэффициентов усиления достигает 1 [15]. Рассмотрим ситуацию, возникающую при приращении управляющего тока I_g на величину ΔI_g . Вследствие такого приращения анодный ток получит приращение ΔI_A , а приращение катодного тока составит

$$\Delta I_K = \Delta I_A + \Delta I_g. \quad (12)$$

По определению малосигнальные коэффициенты усиления равны

$$\tilde{a}_1 \equiv \frac{dI_C}{dI_A} = \lim_{\Delta I_A \rightarrow 0} \frac{\Delta I_C}{\Delta I_A}, \quad (13a)$$

$$\tilde{a}_2 \equiv \frac{dI_C}{dI_K} = \lim_{\Delta I_K \rightarrow 0} \frac{\Delta I_C}{\Delta I_K}. \quad (13b)$$

Дырочный ток через переход J2 равен $\tilde{a}_1 \Delta I_A$, а электронный ток равен $\tilde{a}_2 \Delta I_K$. Приравняв изменение анодного тока изменению тока через J2, получим

$$\Delta I_A = \tilde{a}_1 \Delta I_A + \tilde{a}_2 \Delta I_K. \quad (14)$$

Подставляя выражение (14) в формулу (12), имеем

$$\frac{\Delta I_A}{\Delta I_g} = \frac{\tilde{a}_2}{1 - (\tilde{a}_1 + \tilde{a}_2)}. \quad (15)$$

Когда $\tilde{a}_1 + \tilde{a}_2 = 1$, сколько угодно малое увеличение I_g приведет к потере устойчивости прибора, так как, согласно выражению (15), любое возрастание I_g вызывает неограниченное нарастание I_A . Хотя при анализе в качестве возмущения был использован управляющий ток, к тому же результату приведет небольшое повышение температуры или напряжения.

Статический коэффициент усиления транзистора по току равен

$$\alpha_1 = \alpha_T \gamma, \quad (16)$$

где α_T — коэффициент переноса, по определению равный отношению тока, дошедшего до коллектора, к инжектированному току, а γ — коэффициент инжекции, равный отношению инжектированного тока к полному току эмиттера. Из рис. 9, в следует, что

$$I_G = \alpha_1 I_E + I_{CO}. \quad (17)$$

Дифференцируя выражение (17) по эмиттерному току, получим соотношение для малосигнального коэффициента усиления

$$\tilde{a}_1 \equiv \frac{dI_C}{dI_E} = \alpha_1 + I_E \frac{\partial \alpha_1}{\partial I_E}. \quad (18)$$

Подставляя выражение (16) в соотношение (18), получим

$$\tilde{\alpha}_1 = \gamma \left(\alpha_T + I_E \frac{\partial \alpha_T}{\partial I_E} \right) + \alpha_1 I_E \frac{\partial \gamma}{\partial I_E}. \quad (19)$$

Простейшие оценки для α_T и γ имеют вид

$$\alpha_T = \frac{1}{\text{ch}(W/\sqrt{D\tau})} \simeq 1 - \frac{W^2}{2D\tau}, \quad (20a)$$

$$\gamma \simeq \frac{1}{1 + N_B W / N_E L_E}, \quad (20б)$$

где W — ширина базы, D и τ — коэффициент диффузии и время жизни неосновных носителей в базе, N_B и N_E — концентрации примесей в базе эмиттера, L_E — диффузионная длина в эмиттере. Чтобы получить большие значения коэффициента усиления, необходимо использовать приборы с малыми отношениями $W/\sqrt{D\tau}$ и N_B/N_E .

Для исследования зависимости статического и малосигнального коэффициентов усиления от тока необходимы более детальные вычисления, учитывающие как диффузионную, так и дрейфовую компоненты тока. Дырочные токи через переходы J1 и J2 можно рассчитать по формуле

$$I_p(x) = qA_s \left(p_n \mu_p \mathcal{E} - D_p \frac{\partial p_n}{\partial x} \right), \quad (21)$$

где A_s — площадь перехода. Уравнение непрерывности для области n1 (рис. 2, а) имеет вид

$$\frac{\partial p_n}{\partial t} = - \frac{p_n - p_{n0}}{\tau_p} - \mu_p \mathcal{E} \frac{\partial p_n}{\partial x} + D_p \frac{\partial^2 p_n}{\partial x^2} \quad (22)$$

с граничными условиями $p_n(x=0) = p_{n0} \exp(\beta V)$, где $\beta \equiv \equiv q/kT$, и $p_n(x=W) = 0$. Стационарное решение уравнения (22) при указанных граничных условиях записывается в виде

$$\begin{aligned} p_n(x) = & p_{n0} \exp(\beta V) \exp[(C_1 + C_2)x] - p_{n0} [\exp(\beta V) \exp(C_2 W) + \\ & + \exp(-C_1 W)] \exp(C_1 x) \text{csch}(C_2 W) \text{sh}(C_2 x) + \\ & + p_{n0} \exp(C_1 x) \text{csch}(C_2 W) \text{sh}(x - W) C_2 + p_{n0}, \end{aligned} \quad (23)$$

где

$$C_1 \pm C_2 = \frac{\mu_p \mathcal{E}}{2D_p} \pm \left[\left(\frac{\mu_p \mathcal{E}}{2D_p} \right)^2 + \frac{1}{D_p \tau_p} \right]^{1/2}.$$

Из выражений (21), (22) и (23) можно определить коэффициент переноса

$$\alpha_T = \frac{C_2 \exp(C_1 W)}{C_1 \text{sh}(C_2 W) + C_2 \text{ch}(C_2 W)}. \quad (24)$$

Эффективность эмиттера описывается следующим выражением:

$$\begin{aligned} \gamma &\equiv \frac{I_p}{I_p + I_n + I_r} \approx \frac{I_p}{I_p + I_r} = \\ &= \frac{I_{p0} \exp(\beta V)}{I_{p0} \exp(\beta V) + I_R \exp(\beta V/n)}. \end{aligned} \quad (25)$$

Здесь I_p и I_n — токи, инжектированные в базу и эмиттер соответственно, I_r — ток рекомбинации в области пространственного заряда, равный $I_R \exp(\beta V/n)$, где I_R и n — постоянные (обычно $1 < n < 2$), $I_{p0} = qD_p A_s p_n |C_1 + C_2 \operatorname{cth}(C_2 W)|$. Для распределения примеси, приведенного на рис. 2, а, $p_{p0} (p_1) \gg n_{n0} (n_1)$, и током I_n в формуле (25) можно пренебречь.

Исходя из выражений (24) и (25), можно вычислить зависимость α_1 от тока эмиттера и ширины базы (W). Кроме того, используя выражения (19), (24) и (25), можно получить величину малосигнального коэффициента $\tilde{\alpha}$. На рис. 11 приведены результаты вычислений для распределения примеси, показанного на рис. 2, а, с использованием параметров, типичных для кремния [16]. Отметим, что во всем диапазоне токов малосигнальный коэффициент усиления всегда превышает статический коэффициент усиления для большого сигнала. Отношение ширины базы к диффузионной длине W/L является важным параметром прибора, определяющим зависимость коэффициента усиления от тока.

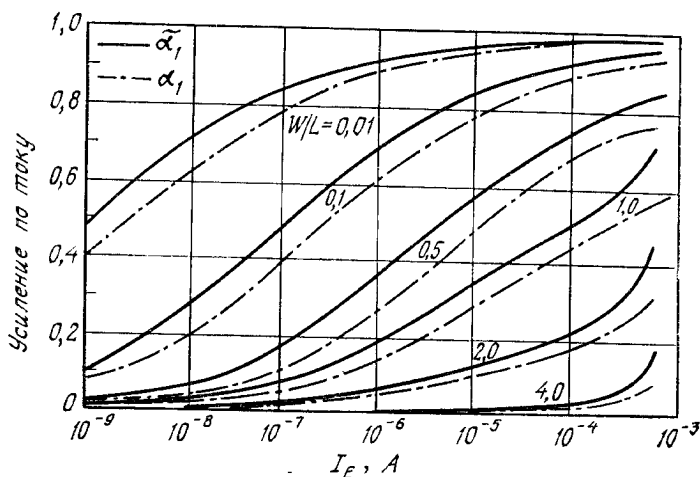


Рис. 11. Зависимость малосигнального и статического коэффициентов переноса от тока и ширины базы для транзистора со следующими параметрами: $n_{n0} = 3 \cdot 10^{14} \text{ см}^{-3}$, $p_{n0} = 7,5 \cdot 10^9 \text{ см}^{-3}$, $A_s = 0,16 \text{ мм}^2$, $\mu_n = 1400 \text{ см}^2/(\text{В} \cdot \text{с})$, $\mu_p = 500 \text{ см}^2/(\text{В} \cdot \text{с})$, $D_p = 13 \text{ см}^2/\text{с}$, $\tau_p = 0,5 \text{ мкс}$, $L_p = 25,5 \text{ мкм}$, $I_R = 2,5 \times 10^{-10} \text{ А}$ и $n = 1,5$ [16].

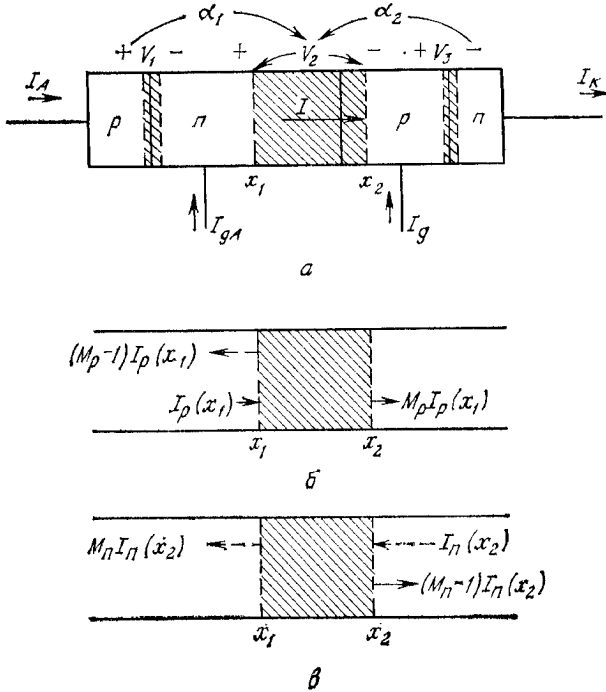


Рис. 12. Обобщенная схема тиристора.

I — ток, протекающий через центральный переход; i_n и i_p — первичные электронный и дырочный токи, генерирующие токи $M_n i_n$ и $M_p i_p$ в условиях лавинного умножения.

При малых значениях W/L коэффициент переноса не зависит от тока и коэффициент усиления изменяется с током только благодаря коэффициенту инжекции. Это условие выполняется в узкой базе прибора (в базе $n-p-n$ -транзистора). При большом значении W/L (в базе $p-n-p$ -транзистора) коэффициент переноса и коэффициент инжекции зависят от тока. Таким образом, величина коэффициента усиления в принципе может лежать в требуемом диапазоне при правильном подборе диффузионной длины и степени легирования.

Напряжение включения. Для определения напряжения включения V_{BF} рассмотрим обобщенную конфигурацию тиристора с выводными контактами ко всем четырем слоям, показанную на рис. 12, а. Пусть центральный переход смещен в обратном направлении и напряжение V_2 на переходе J2 достаточно для лавинного умножения носителей, пересекающих его обедненную область. Обозначим коэффициент умножения электронов через M_n , а коэффициент умножения дырок через M_p и будем считать, что они

оба зависят от V_2 . В результате умножения стационарный ток дырок $I_p(x_1)$, втекающий в обедненный слой в сечении x_1 (рис. 12, б), принимает значение $M_p I_p(x_1)$ в сечении $x = x_2$. Аналогичный результат имеет место для электронного тока $I_n(x_2)$, втекающего в обедненный слой в сечении x_2 (рис. 12, в). Общий ток I равен

$$I = M_p I_p(x_1) + M_n I_n(x_2). \quad (26)$$

Так как, с другой стороны, $I_p(x_1)$ является коллекторным током $p-n-p$ -транзистора, то с учетом схемы на рис. 9, в его можно представить в виде

$$I_p(x) = \alpha_1(I_A) I_A + I_{CO1}. \quad (27)$$

Аналогично можно представить первичный электронный ток $I_n(x_2)$:

$$I_n(x) = \alpha_2(I_K) I_K + I_{CO2}. \quad (28)$$

Подставляя выражения (27) в (28) в формулу (26), получим

$$I = M_p [\alpha_1(I_A) I_A + I_{CO1}] + M_n [\alpha_2(I_K) I_K + I_{CO2}]. \quad (29)$$

Если допустить, что $M_p = M_n = M$, то выражение (29) принимает более простой вид

$$\frac{1}{M(V_2)} = \frac{\alpha_1(I_A) I_A}{I} + \frac{\alpha_2(I_K) I_K}{I} + \frac{I_0}{I}, \quad (30)$$

где $I_0 = I_{CO1} + I_{CO2}$.

Если $I_g = I_{gA} = 0$ (здесь I_{gA} — ток управляющего электрода, расположенного ближе к аноду) и $I = I_A = I_K$, то выражение (30) принимает вид

$$\frac{1}{M(V_2)} = \alpha_1(I) + \alpha_2(I) + I_0/I. \quad (31)$$

Коэффициент умножения M можно представить в виде

$$M(V_2) = \frac{1}{1 - (V_2/V_B)^n}, \quad (32)$$

где V_B — напряжение пробоя (разд. 4.2.1) и n — постоянная величина. Напряжение включения теперь может быть получено из формул (31) и (32) при условии, что $I \gg I_0$. В результате имеем

$$M(V_2) = \frac{1}{\alpha_1 + \alpha_2} = \frac{1}{1 - (V_{BF}/V_B)^n} \quad (33)$$

или

$$V_{BF} = V_B (1 - \alpha_1 - \alpha_2)^{1/n}. \quad (34)$$

Из сравнения с напряжением обратного пробоя $V_{BR} = V_B (1 - \alpha_1)^{1/n}$ видно, что V_{BF} всегда меньше V_{BR} . При малых значе-

ниях суммы $\alpha_1 + \alpha_2$ напряжение V_{BF} примерно равно напряжению обратного пробоя, приведенному на рис. 4. Если же $\alpha_1 + \alpha_2 \approx 1$, напряжение включения существенно меньше V_{BR} .

4.2.3. Режим прямой проводимости

Когда тиристор находится во включенном состоянии, все три перехода смещены в прямом направлении. Дырки инжектируются из области $p1$, а электроны — из области $n2$, и структура $p1-p2-n2$ ведет себя аналогично насыщенному транзистору с удаленным диодным контактом к области $n1$. Следовательно, прибор в целом аналогичен $p1-i-n2$ (p^+-i-n^+)-диоду.

В p^+-i-n^+ -диоде с i -областью шириной W плотность прямого тока определяется скоростью рекомбинации дырок и электронов внутри i -области. Плотность тока определяется выражением

$$J = \int_0^W qR dx. \quad (35)$$

Здесь R — скорость рекомбинации, описываемая соотношением [17]

$$R = G(n^2p + p^2n) + \frac{np - n_i^2}{\tau_{p0}(n + n_i) + \tau_{n0}(p + n_i)}, \quad (36)$$

где первое слагаемое характеризует процесс оже-рекомбинации с коэффициентом G , равным для кремния $(1 \div 2) \cdot 10^{-31}$ см⁶/с; второе слагаемое описывает рекомбинацию через центры с энергией вблизи середины запрещенной зоны, а τ_{p0} и τ_{n0} — времена жизни дырок и электронов соответственно. В предельном случае при $n = p \gg n_i$ выражение (36) принимает более простой вид

$$\tau_{\text{eff}} = \frac{n}{R} = \left(2Gn^2 \cdot \left[\frac{1}{\tau_{p0} + \tau_{n0}} \right] \right)^{-1}. \quad (37)$$

Если во всей i -области концентрация носителей примерно постоянна, то из выражений (35) и (36) вытекает следующая величина плотности тока:

$$J = qnW/\tau_{\text{eff}}. \quad (38)$$

Плотность тока можно представить также в виде

$$J = q(\mu_n + \mu_p)n\bar{\mathcal{E}} = \frac{q}{kT} \frac{(b+1)^2}{2b} qD_a n \bar{\mathcal{E}}, \quad (39)$$

где $\bar{\mathcal{E}}$ — усредненное значение электрического поля, b — отношение μ_n/μ_p и D_a — коэффициент амбиполярной диффузии.

Падение напряжения в i -области V_m равно

$$V_m = 2d\bar{\mathcal{E}}. \quad (40)$$

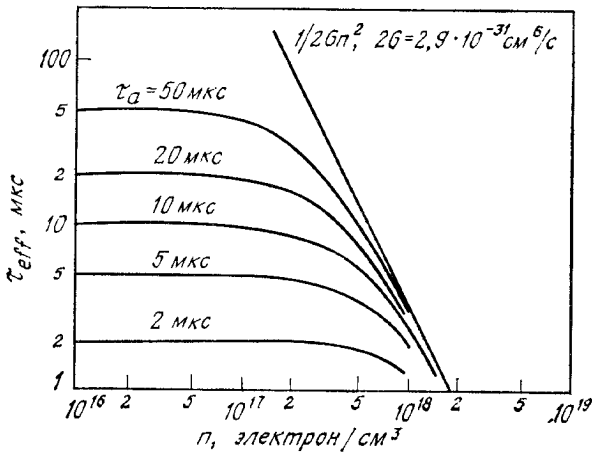


Рис. 13. Эффективное время жизни при высоком уровне инжекции (τ_a — амбиполярное время жизни, G — коэффициент оже-рекомбинации) [18].

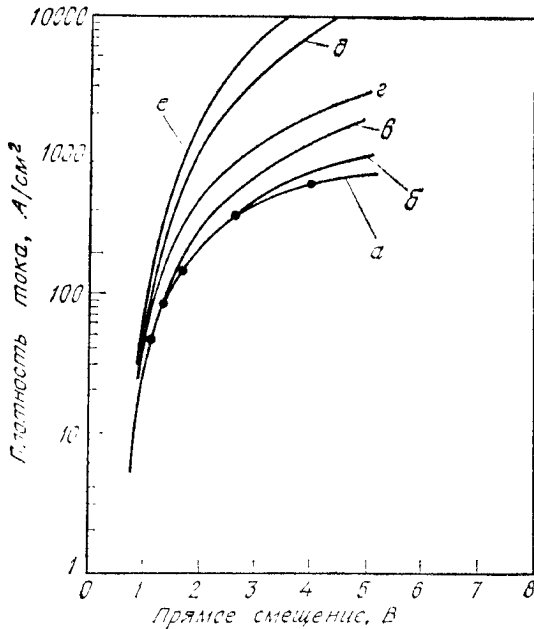


Рис. 14. Теоретические (линии) и экспериментальные (точки) вольт-амперные характеристики тиристора, полученные при учете различных физических механизмов [17].

a — при учете всех факторов и при теплопроводности $50 \text{ Вт}/(\text{см}^2 \cdot \text{К})$; b — при постоянной температуре; g — без учета оже-рекомбинации; δ — без учета оже-рекомбинации и взаимного рассеяния носителей; ϵ — без учета сужения зоны, оже-рекомбинации и взаимного рассеяния носителей.

Объединяя выражения (38) и (39), получим

$$V_m = \frac{kT}{q} \frac{8b}{(1+b)^2} \left(\frac{d^2}{D_a \tau_{\text{eff}}} \right) \quad (41)$$

или

$$V_m \sim (\tau_{\text{eff}})^{-1}. \quad (42)$$

Поскольку V_m обратно пропорционально эффективному времени жизни, для уменьшения V_m необходимо увеличить τ_{eff} . Результаты численных расчетов τ_{eff} [18] для ряда значений амбиполярного времени жизни $\tau_a = \tau_{p0} + \tau_{n0}$ приведены на рис. 13. При низких концентрациях носителей эффективное время жизни равно амбиполярному времени жизни; однако при концентрации носителей, превышающей 10^{17} см^{-3} , эффективное время жизни благодаря оже-рекомбинации быстро уменьшается по закону n^{-2} .

Был проведен численный анализ режима прямой проводимости с учетом различных физических механизмов. Серия расчетных вольт-амперных характеристик тиристора с напряжением включения 2500 В при температуре теплоотвода 400 К приведена на рис. 14 [17]. Плотность тока 1000 А/см^2 соответствует максимально допустимому уровню тока в импульсном режиме, а плотность тока 100 А/см^2 — максимально допустимому уровню тока в статическом режиме. Как видно из рис. 14, при плотностях 100 А/см^2 и выше основными ограничивающими механизмами являются взаимное рассеяние носителей и оже-рекомбинация. Сужение запрещенной зоны фактически не сказывается до тех пор, пока плотность тока не превысит 1000 А/см^2 . Рекомбинация через центры в середине запрещенной зоны выступает в качестве ограничения уже при плотностях, меньших 100 А/см^2 , и остается важным фактором при больших плотностях тока. Температура перехода влияет уже при плотностях тока, превышающих 500 А/см^2 . Кривая a относится к случаю, когда при расчете учитывались все описанные выше механизмы. Приведенные на рис. 14 экспериментальные результаты хорошо согласуются с кривой a .

4.3. ДИОДНЫЙ И ТРИОДНЫЙ ТИРИСТОРЫ

4.3.1. Статические вольт-амперные характеристики

Как уже говорилось выше, диодный тиристор представляет собой двухполюсный $p-n-p-n$ -прибор [19а, б]. Исходя из общего соотношения (30), опишем графический метод построения его вольт-амперной характеристики. Так как в диодном тиристоре $I_g = I_{gA} = 0$ и $I_A = I_K = I$, то соотношение (30) принимает более простой вид

$$\frac{1}{M(V_2)} = \alpha_1(I) + \alpha_2(I) + I_0/I = f(I), \quad (43)$$

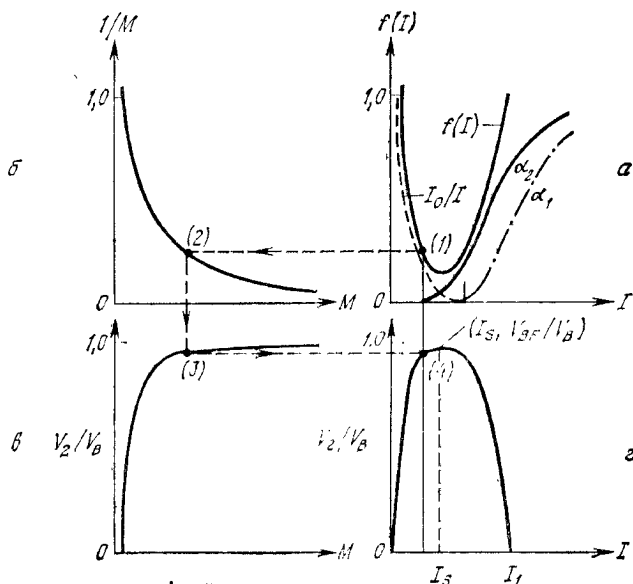


Рис. 15. Графическое построение вольт-амперной характеристики диодного тиристора [20].

где коэффициент $M (V_2)$ определяется выражением (32). Пусть I_0 — некоторая известная постоянная, а зависимости α_1 и α_2 от тока имеют вид, показанный на рис. 11. Графическое решение уравнения (43) продемонстрировано на рис. 15 [20]. Сложив три кривые, получим функцию $f(I)$ (рис. 15, а). На рис. 15, б приведена зависимость $1/M$ от M , причем масштаб по вертикали идентичен масштабу на рис. 15, а, а масштаб по горизонтали идентичен масштабу на рис. 15, в, на котором приведена зависимость (32). Выберем теперь на оси абсцисс (рис. 15, в) исходную точку — значение тока I , для которого требуется найти падение напряжения на приборе. Спроецируем эту точку вертикально вверх на рис. 15, а и найдем значение функции $f(I)$ — точка 1. Затем спроецируем точку 1 по горизонтали налево до пересечения с графиком $1/M$ и получим точку 2, т. е. определим M . По известной величине M , проецируя точку 2 вертикально вниз, найдем необходимую величину (V_2/V_B) на рис. 15, в (точка 3). Выполним последнюю горизонтальную проекцию точки 3 до пересечения с первой вертикальной проекцией в точке 4. Эта точка дает величину нормализованного падения напряжения на приборе, необходимого для поддержания заданного тока I . Полная вольт-амперная характеристика прибора (рис. 15, в) получается повторением этого геометрического построения для разных значений тока.

Отметим, что точка переключения (I_s , V_{BF}) расположена в диапазоне токов, где функция $f(I)$ достигает минимума. Точка удержания определяется как точка с низким напряжением и большим током, в которой $dV/dI = 0$. Приведенный анализ не позволяет ее найти. Однако в первом приближении ее координатами можно считать $(I_1, 0)$, где $f(I_1) = 1$. При $f(I_1) = 1$ $M(V_2) = 1$, а это означает, что $V_2 = 0$. Если напряжение $V_2 = 0$, то ток насыщения центрального перехода также стремится к 0. Тогда из уравнения (43) следует $\alpha_1(I) + \alpha_2(I) = 1$. Отсюда при известных α_1 и α_2 можно найти ток, при котором смещение на центральном переходе становится нулевым. В результате полное падение напряжения на приборе равно просто сумме падений на

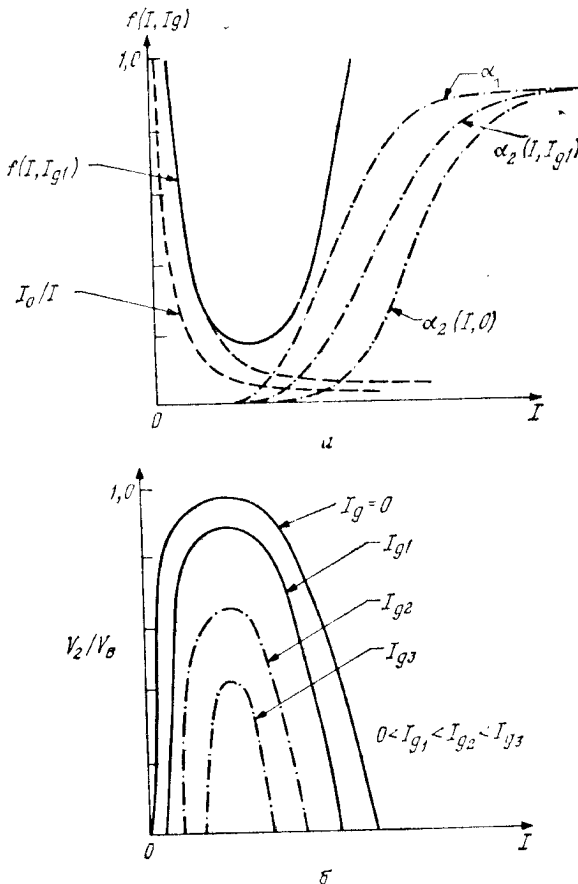


Рис. 16. Графическое построение вольт-амперной характеристики триодного тиристора [20].

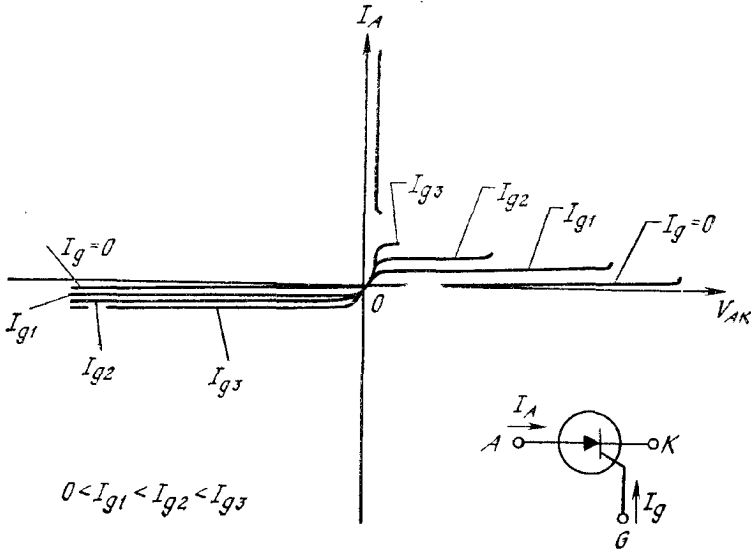


Рис. 17. Вольт-амперная характеристика тиристора при различных управляющих токах [5].

пряжений на двух прямосмещенных внешних переходах ($\sim 1,2-1,4$ В для кремниевых приборов).

При токе, превышающем I_1 , все переходы смещены в прямом направлении. (Проведенный выше анализ неприменим, так как верен в предположении, что переход J2 остается смещенным в обратном направлении.) Если продолжать увеличивать ток, полное падение напряжения еще более уменьшается и прибор остается на участке с дифференциальным отрицательным сопротивлением, пока общий ток не превысит ток удержания I_h . При больших токах падение напряжения на центральном переходе становится сравнимым с напряжением на эмиттерном переходе и динамическое сопротивление прибора снова оказывается положительным [21]. После прохождения точки (I_h, V_h) на характеристике прибор находится в области прямой проводимости (разд. 4.2.3).

Для тиристора с одним управляющим электродом уравнение (30) можно преобразовать к виду

$$\frac{1}{M(V_2)} = \alpha_1(I) + \alpha_2(I + I_g) + \frac{\alpha_2(I + I_g)}{I} I_g + \frac{I_0}{I} = f(I, I_g). \quad (44)$$

В этом уравнении ток I_K заменен на $I + I_g$ и $I_{gA} = 0$. Уравнение (44) равносильно уравнению (43), если $I_g = 0$. На рис. 16 приве-

дены функция $f(I, I_g)$ и вольт-амперная характеристика тиристора при $I_g = 0$. Семейство вольт-амперных характеристик для различных значений I_g получается путем построения $\alpha_2(I + I_g)$ для каждого значения I_g и включения в функцию $f(I, I_g)$ члена $\alpha_2(I + I_g)/I$. Таким образом строят семейство кривых $f(I, I_g)$. Необходимо отметить, что с увеличением I_g напряжение включения падает. Это позволяет управлять включением структуры.

Полное семейство вольт-амперных характеристик триодного тиристора при различных управляющих токах приведено на рис. 17. В режиме прямого запираания характеристики аналогичны кривым, приведенным на рис. 16, б, с учетом расположения координатных осей.

4.3.2. Время включения и выключения

В процессе переключения тиристора из закрытого состояния в открытое ток через него должен вырасти до достаточно высокого уровня, чтобы удовлетворялось условие $\tilde{\alpha}_1 + \tilde{\alpha}_2 = 1$. Существует несколько способов переключения тиристорov. Единственным способом включения диодного тиристора является переключение анодным напряжением. Переключение напряжением можно осуществить двумя путями: медленным повышением анодного напряжения вплоть до напряжения включения и импульсом анодного напряжения. Последний вариант, называемый dV/dt -переключением, рассмотрен в разд. 4.3.3.

Наиболее важным способом включения триодного тиристора является запуск управляющим током. При подаче запускающего тока (управляющего тока в базу) анодный ток тиристора реагирует на него с запаздыванием. Процесс изменения анодного тока можно характеризовать временем включения $t_{он}$ (рис. 18, а). Поскольку процесс переключения тиристора имеет регенеративную природу, время включения приблизительно равно среднему геометрическому времени диффузии в $n1$ - и $p2$ -областях, или

$$t_{он} = \sqrt{t_1 t_2}, \quad (45)$$

где $t_1 \equiv W_{n1}^2/2D_p$, $t_2 \equiv W_{p2}^2/2D_n$, W_{n1} и W_{p2} — толщины слоев $n1$ и $p2$ соответственно, D_n и D_p — коэффициенты диффузии элементов и дырок соответственно.

Приведенный выше результат вытекает из рис. 9, если воспользоваться методом заряда. Положим, что в $p-n-p$ - и $n-p-n$ -транзисторах накоплены заряды Q_1 и Q_2 . Коллекторные токи транзисторов $I_{c2} \approx Q_1/t_1$ и $I_{c1} \approx Q_2/t_2$. В идеальных условиях при $dQ_1/dt = I_{c2}$ и $dQ_2/dt = I_g + I_{c1}$ справедливо следующее уравнение:

$$\frac{d^2 Q_1}{dt^2} - \frac{Q_1}{t_1 t_2} = \frac{I_g}{t_2}. \quad (46)$$

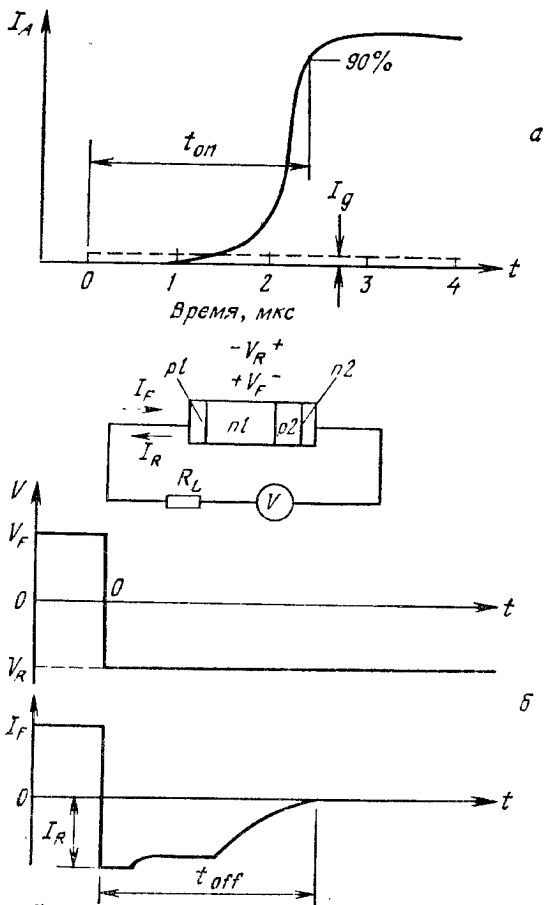


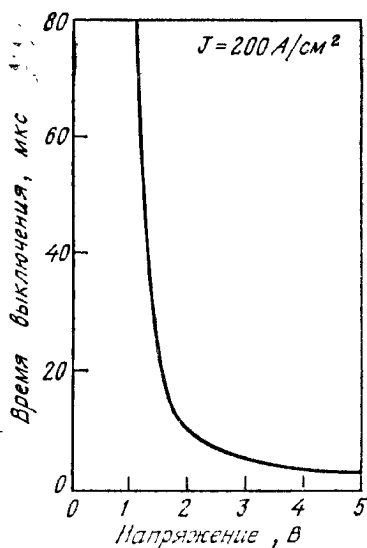
Рис. 18. Переходные процессы в триодном тиристоре [5].

a — включение при подаче запускающего тока; *б* — выключение при быстром изменении полярности напряжения.

Его решением является функция $\exp(-t/t_{\text{он}})$ с постоянной времени $t_{\text{он}}$, определяемой формулой (45). Для уменьшения времени включения необходимо создавать приборы с узкими слоями $n1$ и $p2$. Однако такое требование противоречит условию высокого пробивного напряжения. Поэтому мощные высоковольтные тиристоры обладают большим временем включения.

Когда тиристор находится во включенном состоянии, все три его перехода смещены в прямом направлении. Следовательно, в приборе накапливаются избыточные неосновные и основные носители заряда, а их избыток n пропорционален прямому току. Для перевода прибора в закрытое состояние избыточные носители

Рис. 19. Типичная взаимосвязь между прямым падением напряжения и временем выключения мощного тиристора [27].



либо должны быть удалены электрически полем, либо должны рекомбинировать [22, 23]. Типичная форма тока при выключении прибора показана на рис. 18, б. Основной вклад во время задержки дает время рекомбинации в слое $n1$.

Поскольку дырочный ток через структуру пропорционален избыточному заряду в слое $n1$, то можно написать

$$I = I_F \exp\left(-\frac{t}{\tau_p}\right), \quad (47)$$

где $I = I_F$ в момент $t = 0$, τ_p — время жизни неосновных носителей. Для перевода прибора в закрытое состояние ток через него должен быть меньше тока удержания I_h . Отсюда время выключения равно

$$t_{\text{off}} = \tau_p \ln\left(\frac{I_F}{I_h}\right). \quad (48)$$

Чтобы добиться малого времени выключения, необходимо снизить время жизни неосновных носителей в слое $n1$. Этого достигают введением центров рекомбинации во время диффузионного процесса (в кремнии такие центры дают золото и платина) или путем электронного и гамма-облучения [24—26]. Золото образует в кремнии акцепторный уровень вблизи середины запрещенной зоны и служит эффективным генерационно-рекомбинационным центром. Однако с повышением концентрации золота возрастают токи утечки. В результате при легировании золотом снижается напряжение включения. Легирование платиной или облучение электронами не вызывает снижения напряжения. Уменьшение времени жизни

одновременно увеличивает прямое падение напряжения во включенном состоянии (выражение (42)). Типичная взаимосвязь прямого падения напряжения и времени выключения мощного тиристора приведена на рис. 19 [27]. Приемлемое время выключения составляет 10 мкс, и прямое падение напряжения равно 2 В.

Для сокращения времени выключения обычно во время фазы запирания подают обратное смещение между управляющим электродом и катодом. Такой способ называется управляемым выключением [28, 29]. Улучшение запирания связано с тем, что обратносмещенный управляющий электрод поглощает большую часть тока рассасывания, которая в противном случае протекала бы через катод, мешая восстановлению прямого напряжения на аноде.

4.3.3. Закороченный катод и эффект dV/dt

В современных диодных и триодных тиристорах для улучшения характеристик приборов используют закороченный катод [6, 7]. Упрощенная конструкция тиристора с закороченным катодом показана на рис. 20, а. Эквивалентная двухтранзисторная схема приведена на рис. 20, б, где полный ток катода, обозначенный через $I_{K'}$, равен сумме тока I_K и шунтирующего тока $I_{шунт}$. Если сопротивление $R_{шунт}$ настолько мало, что большая часть анодного тока течет через закороченную часть катода, то $I_K \ll I_A$. Из уравнения (30) при $I_{KA} = 0$ и $I = I_A$ получим

$$I_A = \frac{MI_0}{1 - M\alpha_1 - M\alpha_2 (I_K/I_A)}. \quad (49)$$

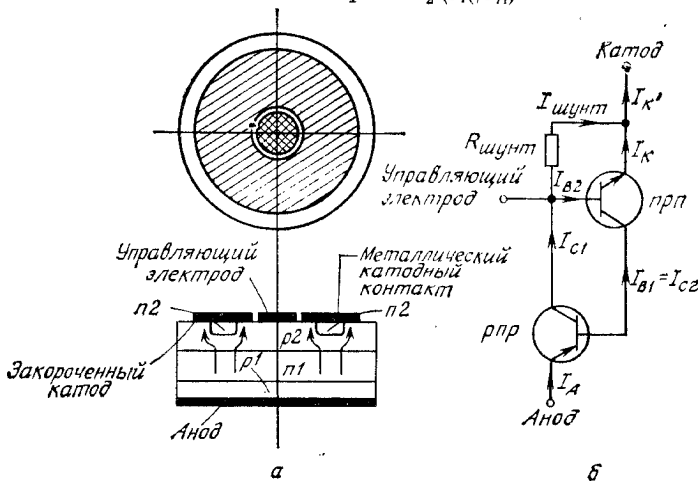


Рис. 20. Тиристор с закороченным катодом.

а — часть тока протекает через катодную аакоротку; б — двухтранзисторная модель тиристора с закороченным катодом.

Если катод не закорочен, то $I_K = I_A$, поэтому

$$I_A = \frac{MI_0}{1 - M\alpha_1 - M\alpha_2}. \quad (50)$$

Из выражения (33) следует, что напряжение включения V_{BF} определяется условием $M = 1/(\alpha_1 + \alpha_2)$. При закороченном катоде, когда $I_K \ll I_A$, уравнение (49) принимает вид

$$I_A = \frac{MI_0}{1 - M\alpha_1}. \quad (51)$$

В этих условиях напряжение включения становится равным напряжению обратного пробоя, определяемому выражением (4).

В динамическом режиме закрытый тиристор может быть включен при напряжениях, намного более низких, чем напряжение включения. Уменьшение напряжения включения при этом зависит от амплитуды импульса анодного напряжения и скорости его нарастания. Это явление называется эффектом dV/dt . Его можно использовать для включения тиристора (dV/dt -переключение). Эффект dV/dt обусловлен тем, что быстро изменяющееся анодное напряжение приводит к увеличению тока смещения $d(CV)/dt$, где C — емкость перехода J_2 . Емкостный ток в свою очередь может вызвать возрастание суммарного коэффициента усиления ($\bar{\alpha}_1 + \bar{\alpha}_2$) до единицы, и в результате произойдет включение тиристора. В мощных тиристорах, которые должны обладать высокими значениями V_{BF} , необходимо ослабить влияние эффекта dV/dt . С этой целью можно сместить цепь управляющий электрод — катод в обратном направлении, чтобы ток смещения протекал через управляющий вывод и не вызывал увеличения коэффициента усиления. Можно также снизить время жизни неосновных носителей в областях n_1 и p_2 , что приведет к уменьшению коэффициентов усиления, но в режиме прямой проводимости α остается высоким.

Эффективным способом ослабления эффекта dV/dt является использование закороченного катода [30]. Как показано на рис. 20, а, ток смещения протекает через шунт и не влияет на коэффициент усиления α_2 $n-p-n$ -транзистора. Закороченный катод может обеспечить существенно большую устойчивость к эффекту dV/dt . Обычно тиристоры с незакороченным катодом допускают нарастание анодного напряжения 20 В/мкс. В приборах с закороченным катодом устойчивость к эффекту dV/dt может увеличиваться в 10—100 раз и более.

4.3.4. Ограничение dI/dt [7]

В начале процесса включения тиристора начинает проводить только небольшая область катода вблизи управляющего контакта. Эта область высокой проводимости поставяет запускающий ток,

необходимый для включения других близко расположенных областей, пока процесс не распространится на всю площадь катода. Этот процесс характеризуется скоростью распространения v_{sp} . Рассмотрим концентрическую структуру (рис. 20, а) с управляющим электродом радиусом r_0 , расположенным в центре, и предположим, что анодный ток и напряжение между анодом и катодом линейно изменяются со временем. Таким образом, на этапе включения

$$i_A = \frac{dI_A}{dt} t, \quad v_{AK} = V_{AK} \left(1 - \frac{t}{t_0}\right), \quad (52), (53)$$

$$r = r_0 + v_{sp} t, \quad (54)$$

где V_{AK} — стационарное напряжение между анодом и катодом, t_0 — время включения, i_A , v_{AK} , r — изменяющиеся во времени ток, напряжение и радиус проводящей области соответственно.

Мгновенная рассеиваемая мощность и проводящая площадь составляют

$$i_A v_{AK} = V_{AK} \frac{dI_A}{dt} \left(1 - \frac{t}{t_0}\right) t, \quad (55)$$

$$\text{Площадь} = \pi [(r_0 + v_{sp} t)^2 - r_0^2]. \quad (56)$$

Отсюда находим плотность рассеиваемой мощности

$$P = \frac{i_A v_{AK}}{\text{Площадь}} = \frac{V_{AK} (dI_A/dt) (1 - t/t_0) t}{\pi [(r_0 + v_{sp} t)^2 - r_0^2]} \quad (57)$$

и повышение температуры в наиболее горячей точке

$$\Delta T = \frac{1}{\rho C} \int_0^{\infty} P dt \sim (dI_A/dt), \quad (58)$$

где ρ и C — плотность и удельная теплоемкость кремния соответственно. Из формулы (58) следует, что при постоянной величине V_{AK} повышение температуры пропорционально dI_A/dt . Для предотвращения перегрева и постепенного разрушения прибора важное значение приобретает такой параметр, как допустимая скорость нарастания тока. Для ее увеличения можно расширить площадь области начальной проводимости или уменьшить отношение W/L в области $n1$, повышая скорость распространения области проводящего состояния.

Предложено несколько гребенчатых конструкций, в которых расстояние от любой части катода до управляющего электрода не превышает максимально допустимой величины. Изящный пример использования с этой целью эвольвенты показан на рис. 21. Уравнение эвольвенты ABC в полярных координатах имеет вид

$$r = r_0 (1 + 0^2)^{1/2}, \quad (59)$$

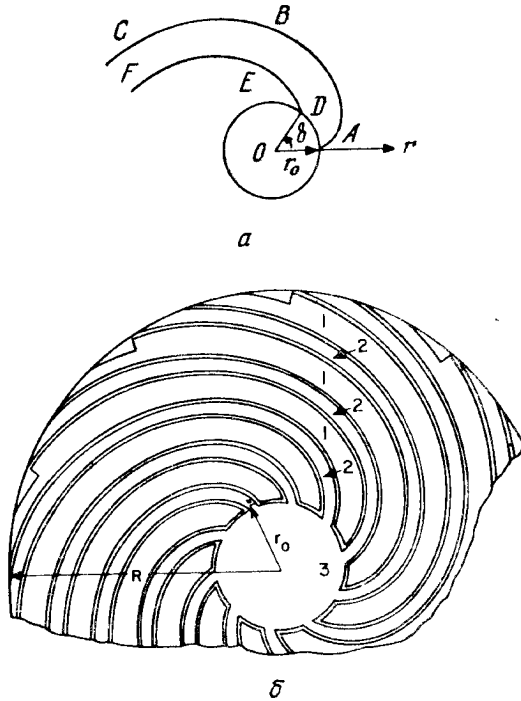


Рис. 21. Структура с электродами в форме эвольвенты [31].

а эвольвенты DEF

$$r = r_0 [1 - |(0 - \delta)^2|^{1/2}]. \quad (60)$$

Эти две эвольвенты везде эквидистантны и отстоят друг от друга на расстоянии $r_0\delta$. На рис. 21, б показана геометрия катодов и управляющих электродов прибора. Катоды, обозначенные цифрой 1, имеют мезоструктуру. Металлический слой на управляющих электродах (цифра 2) присоединен к управляющему контакту 3.

Другой способ расширения области начального включения состоит в использовании дополнительного усилительного управляющего электрода [32] (рис. 22, а). Эквивалентная схема тиристора с таким усилительным электродом приведена на рис. 22, б. Когда на центральный управляющий электрод подан небольшой запускаяющий ток, структура с усилительным электродом, играющая роль запускаяющего прибора, включается гораздо быстрее, так как ее продольные размеры невелики. Запускающий ток намного больше исходного переключающего тока и обеспечивает

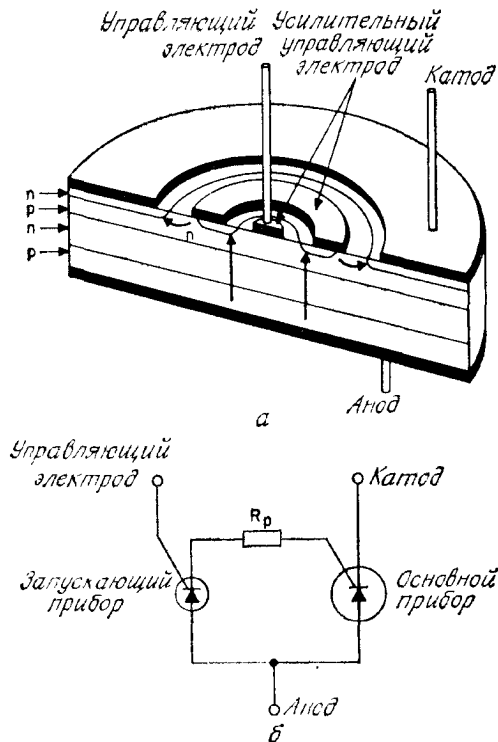


Рис. 22. Тиристор с усилительным управляющим электродом [32].
 а — конструкция; б — эквивалентная схема.

основному прибору более сильный возбуждающий ток. Увеличенный возбуждающий ток расширяет область начального включения основного тиристора.

4.3.5. Максимальная рабочая частота

При низких скоростях переключения тиристор в общем случае является более эффективным ключом, чем биполярный транзистор. Поэтому тиристоры получили широкое применение в области промышленного регулирования мощности, где рабочие частоты обычно составляют 50—60 Гц. В последнее время расширяется их применение в схемах с более высокими скоростями переключения. Рассмотрим теперь максимально достижимые рабочие частоты тиристоров.

Напряжение на выводах тиристора и протекающий через него во время переключения ток приведены на рис. 23 [33]. Основным фактором, влияющим на время включения и выключения тири-

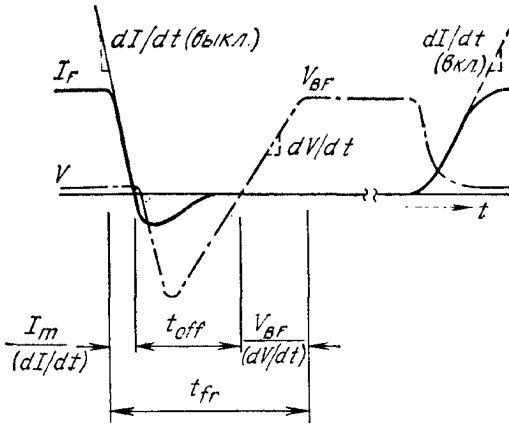


Рис. 23. Изменение токов и напряжений на выводах тиристора во время переключения [33].

стора, является скорость изменения тока dI/dt в процессе включения и выключения. В закрытом состоянии dI/dt определяется главным образом внешней цепью. Оптимизируя параметры этой цепи, необходимо ограничить dI/dt при включении (разд. 4.3.4). Время выключения рассмотрено в разд. 4.3.2. Скорость нарастания прямого падения напряжения на тиристоре dV/dt после периода восстановления высокой проводимости ограничивается емкостным током смещения.

Последний может вызвать увеличение коэффициента усиления $n-p-n$ -транзистора, достаточное для включения тиристора, раньше, чем к тиристору будет приложено полное прямое напряжение, или раньше, чем поступит запускаящий сигнал на управляющий электрод. Такой эффект существенно ослабляется при закороченном катode. Время прямого восстановления равно сумме трех указанных выше составляющих:

$$t_{fr} = I_F/(dI/dt) + t_{off} + V_{BF}/(dV/dt), \quad (61)$$

где I_F — максимальный прямой ток, V_{BF} — напряжение включения. Максимальная рабочая частота определяется выражением

$$f_m = \frac{1}{2t_{fr}}. \quad (62)$$

На рис. 24 приведено соотношение между временем выключения t_{off} , диффузионной длиной L и шириной W базы $n1$ в зависимости от времени жизни неосновных носителей в области $n1$ для тиристорov с рабочими напряжениями 800 и 1100 В. Из рисунка видно, что в соответствии с выражением (48) время t_{off} прямо пропорционально τ_p . Напряжение включения обратно пропорцио-

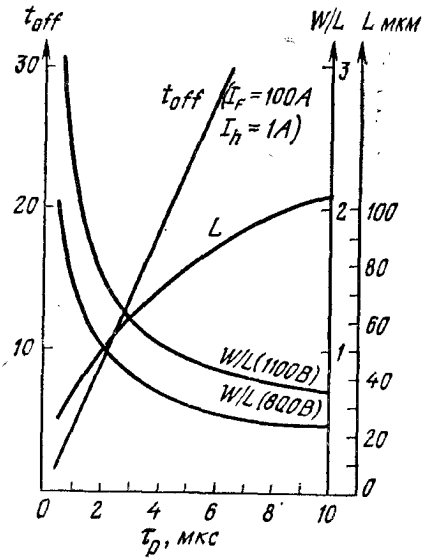


Рис. 24. Соотношение между временем выключения t_{off} , диффузионной длиной L и шириной W базы n_1 в зависимости от времени жизни неосновных носителей в базе [33].

нально частоте для любого отношения W/L . При заданном напряжении включения ширина W обычно фиксирована и отношение W/L изменяется обратно пропорционально $\sqrt{\tau_p}$.

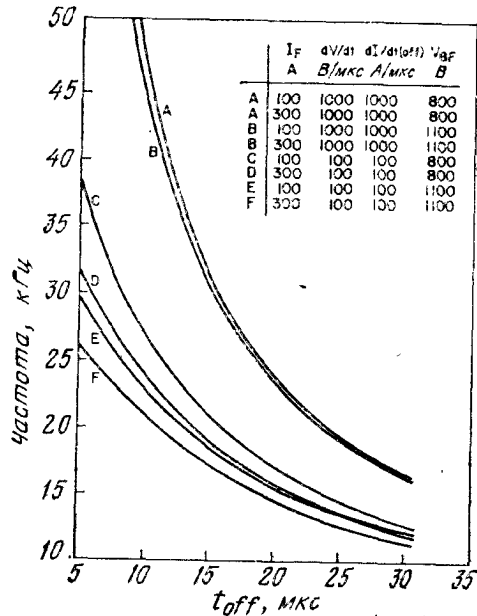


Рис. 25. Максимальные рабочие частоты тиристоров [33].

Максимальная рабочая частота тиристорov, изготовленных обычным образом, с рабочими напряжениями 800 и 1100 В приведена на рис. 25 для двух значений каждого из параметров: рабочих токов dI/dt и dV/dt . Например, тиристор с рабочим напряжением 800 В и током 100 А при $W/L = 0,75$ ($t_{off} = 18$ мкс из рис. 24) и при $dV/dt = 100$ В/мкс и $dI/dt = 100$ А/мкс имеет $f_m \approx 20$ кГц. Для повышения f_m необходимо увеличивать dV/dt или dI/dt , или усложнять структуру прибора, или делать и то и другое. Таким образом, сохраняя рабочее напряжение и величину прямого тока, можно за счет снижения времени жизни уменьшить время выключения тиристора.

4.4. МОЩНЫЕ ТИРИСТОРЫ

4.4.1. Тиристоры с обратной проводимостью

Тиристор с обратной проводимостью (reverse conducting thyristor) представляет собой многослойный тиристор с тремя выводами, характеристики которого аналогичны характеристикам обычного тиристора в режиме прямой проводимости, но который проводит большой ток в обратном направлении. Он используется в качестве схемы возбуждения электролюминесцентных ламп и двунаправленной схемы переключателя переменного тока. Особенность этого тиристора состоит в том, что он имеет закороченные катод и анод. Поперечный разрез тиристора с обратной проводимостью приведен на рис. 26, а [34]. Отметим, что для перехода J_2 пригодна только положительная фаска. Между p - и n -областями вблизи анодного контакта введены n^+ -области. В результате при работе в обратном направлении, когда анод отрицательно смещен по отношению к катоду, переход J_1 , обычно находящийся под обратным смещением, теперь оказывается закороченным на анод. Следовательно, в обратном направлении протекают большие токи. Прямая вольт-амперная характеристика (рис. 26, б) такая же, как для обычного тиристора.

Мощные тиристоры с обратной проводимостью могут работать при температурах переходов, превышающих 150 °С, в то время как обычные тиристоры сохраняют работоспособность лишь до 125 °С. Поэтому первые способны пропускать большие токи в открытом состоянии.

На рис. 27 приведены зависимости напряжения включения от температуры тиристора с обратной проводимостью и тиристора с закороченным катодом. Как уже говорилось выше, V_{BF} для тиристора с закороченным катодом равно напряжению обратного запираения, которое выше, чем для тиристора с незакороченным катодом. В закрытом состоянии работа тиристора с обратной проводимостью не зависит от коэффициентов усиления составляющих

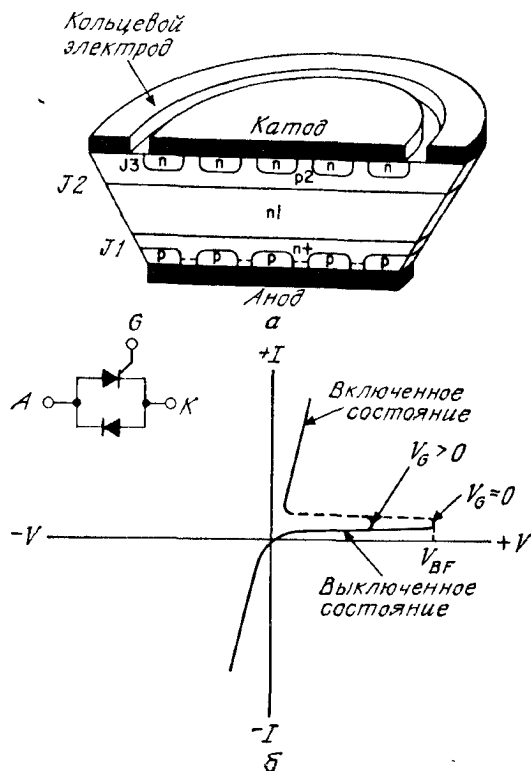


Рис. 26. Высоковольтный высокотемпературный тиристор с обратной проводимостью [34].

а — конструкция; б — вольт-амперная характеристика и условное обозначение тиристора.

транзисторов и V_{BF} равно напряжению лавинного пробоя перехода J2. Поскольку лавинный пробой имеет положительный температурный коэффициент, V_{BF} для тиристора с обратной проводимостью увеличивается с температурой (рис. 27). Видно, что при температурах выше 125°C V_{BF} тиристора с закороченным катодом быстро падает, в то время как V_{BF} тиристора с обратной проводимостью продолжает возрастать и при температурах, превышающих 150°C .

Тиристоры с обратной проводимостью обладают быстрым восстановлением, так как в закрытом состоянии их характеристики не зависят от работы транзисторов. Следовательно, эти тиристоры допускают более высокий уровень легирования золотом или платиной, чем обычные тиристоры.

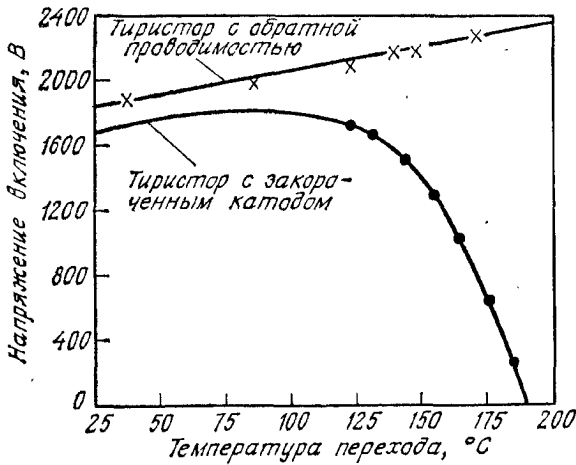


Рис. 27. Температурные зависимости напряжения включения тиристора с обратной проводимостью и тиристора с закороченным катодом [34].

4.4.2. Фототиристор

Фототиристор (или переключатель, управляемый излучением) представляет собой четырехслойный (несимметричный) тиристор с тремя выводами, который может включаться под действием светового излучения, превышающего некоторый порог. Фототиристор способен обеспечивать совершенную электрическую изоляцию между мощной схемой и схемой возбуждения за счет передачи энергии запускающего импульса по волоконно-оптической линии. Фототиристоры применяются в системах фотоэлектрического контроля и управления, оптической связи и оптоэлектронных переключательных схемах.

Упрощенная структура фототиристора и его условное обозначение приведены на рис. 28 [35]. Для ослабления эффекта dV/dt и повышения температурной стабильности напряжения включения прибор имеет закороченный катод. Катодная площадка равномерно облучается источником света через оптическое волокно радиусом r_1 , в результате чего в пределах облучаемой области возникают однородно распределенные электронно-дырочные пары. Внутренняя область катода закорочена по окружности радиусом $r = r_2$.

Наиболее существенный вклад во внешний ток дают электронно-дырочные пары, генерируемые внутри области пространственного заряда обратносмещенного перехода $J2$. Они разделяются электрическим полем порядка 1 нс, т. е. практически мгновенно по сравнению с временем включения обычного тиристора (~ 1 мкс). Дырки переносятся в область $p2$, а электроны —

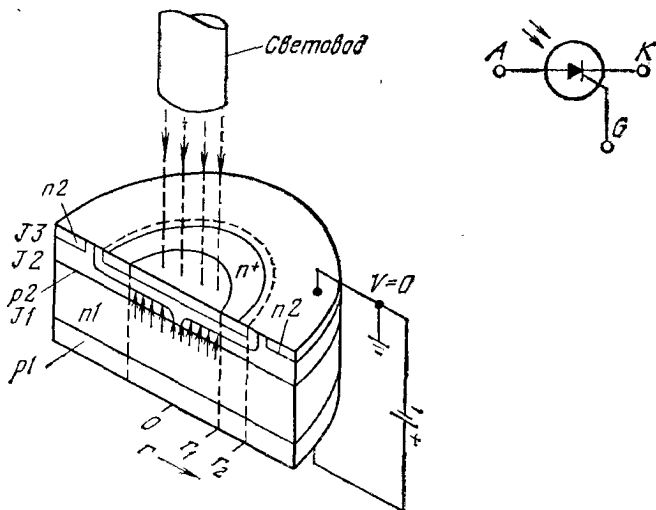


Рис. 28. Конструкция и условное обозначение фототиристора с закороченным катодом [35].

в область $n1$. Обе области получают одинаковые количества основных носителей без задержки на их генерацию. Поэтому в момент включения под действием света анодный ток резко возрастает (рис. 29).

Фототок затем усиливается двухтранзисторной $p-n-p-n$ -структурой, работающей в регенеративном режиме. Анодный ток после некоторой задержки, связанной с временем движения инжектированных неосновных носителей в базе, продолжает возрастать. Если сумма коэффициентов усиления остается меньше 1 (I_{ph1} достигает I_{A1}), то ток асимптотически стремится к стадио-

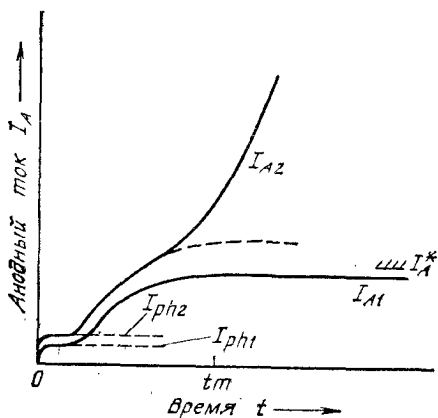


Рис. 29. Характеристики включения фототиристора при различных фототоках [35].

нарной величине и достигает ее за время t_m , равное среднему времени жизни носителей в областях $n1$ и $p2$. Пусть I_A^* — стационарный ток, при котором сумма коэффициентов усиления равна 1, и пусть I_{ph2} — фототок, при котором соответствующий анодный ток I_{A2} стремится к стационарной величине, превышающей I_A^* . Тогда вскоре после превышения анодным током значения I_A^* вступит в действие регенеративная положительная обратная связь по току, что приведет к быстрому нарастанию анодного тока. В результате тиристор переключится в открытое состояние (рис. 29). При увеличении фототока процесс включения сдвигается в сторону уменьшения времени задержки; следовательно, время включения уменьшается с увеличением интенсивности света.

Экспериментальные характеристики включения высоковольтного тиристора приведены на рис. 30. Штриховой линией представлено включение с помощью управляющего электрода. Кривая *a* характеризует включение при низкой световой энергии а кривая *б* — при более сильном световом возбуждении. При повышении возбуждения время задержки сильно уменьшается.

Величина фототока, включающего тиристор, зависит от длины волны излучения λ . Для кремния максимум спектральной чувствительности лежит в диапазоне $\lambda = 0,85—1,0$ мкм. Эффективными источниками света в этом диапазоне являются лазеры с двойной гетероструктурой GaAs—GaAlAs, лазеры на GaAs и светодиоды из GaAs (гл. 12).

Мощные тиристоры могут включаться под действием очень малой интенсивности света ($\sim 0,2$ мВт для тиристора с рабочим напряжением 3 кВ), так как излучение допускает фокусировку на очень небольшой площади. Например, при диаметре стекловолокна 100 мкм площадь поверхности начального включения не превышает 10^{-2} мм². Следовательно, плотность мощности на площадке начального включения будет очень высокой. Для фототи-

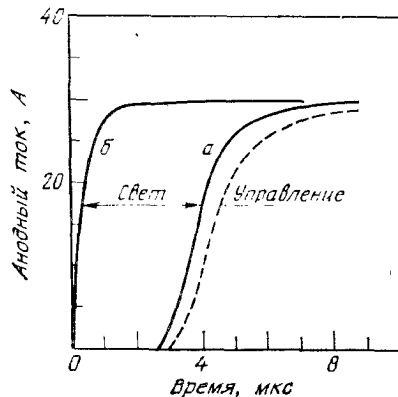


Рис. 30. Экспериментальные характеристики включения мощного фототиристора излучением неодимового лазера. Энергия лазерного импульса $2,5 \cdot 10^{-6}$ Вт·с (а) и $5 \cdot 10^{-4}$ Вт·с (б) [35].

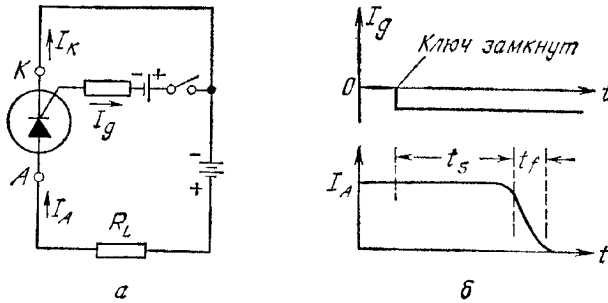


Рис. 31. Выключаемый тиристор [36].

а — схема включения; *б* — переходный процесс выключения.

ристора с закороченным катодом (рис. 28) минимальная световая мощность, необходимая для запуска, изменяется приблизительно обратно пропорционально r_2/r_1 . Поэтому при больших значениях r_2/r_1 требуется небольшая мощность света. Однако даже при $r_2/r_1 = 5$ для запуска необходима световая мощность около 5 мВт, что на порядок выше, чем для фототиристора с незакороченным катодом. Следовательно, необходимо искать компромисс между световой чувствительностью и устойчивостью к эффекту dV/dt .

Как только световое излучение возбудит область начального включения, регенеративный процесс приводит к расширению включенной площади, и в конце концов включается весь катод прибора. После переключения в проводящее состояние, в котором анодный ток превышает фототок, световое излучение можно отключить, не изменяя анодный ток.

4.4.3. Выключаемый тиристор

Выключаемый тиристор (или тиристор с выключением по управляющему электроду) представляет собой четырехслойную структуру, которая допускает включение положительным управляющим напряжением и выключение отрицательным управляющим напряжением. Обычно тиристор выключают путем снижения анодного тока ниже тока удержания или путем включения анодного тока противоположного направления. Выключаемые тиристоры используются в качестве инверторов, импульсных генераторов, прерывателей и в схемах переключателей постоянного тока и часто предпочитают применять их вместо транзисторов в высокоскоростных мощных устройствах, поскольку они способны выдерживать высокое напряжение в закрытом состоянии.

Принципиальная схема включения тиристора [36] приведена на рис. 31, *а*. При одномерном представлении процесса выключения следует рассмотреть поведение этого тиристора под действием

отрицательного управляющего тока с амплитудой I_g . Если пренебречь всеми токами утечки, то для поддержания $n-p-n$ -транзистора во включенном состоянии необходим базовый ток, равный $(1 - \alpha_2) I_K$. В действительности базовый ток составляет $(\alpha_1 I_A - I_g)$. Следовательно, условие включения записывается следующим образом:

$$\alpha_1 I_A - I_g < (1 - \alpha_2) I_K. \quad (63)$$

Поскольку $I_A = I_K + I_g$, то для выполнения неравенства (63) необходимо, чтобы

$$I_g > \left(\frac{\alpha_1 + \alpha_2 - 1}{\alpha_2} \right) I_A. \quad (64)$$

Введем коэффициент усиления выключения β_{off} в виде отношения I_A/I_g :

$$\beta_{off} \equiv \frac{I_A}{I_g} < \frac{\alpha_2}{\alpha_1 + \alpha_2 - 1}. \quad (65)$$

Для увеличения β_{off} необходимо, чтобы α_2 для $n-p-n$ -транзистора был как можно ближе к 1 и в то же время чтобы α_1 для $p-n-p$ -транзистора был малым.

В реальных транзисторах процесс выключения носит двумерный характер. До подачи отрицательного тока I_g оба транзистора включены и находятся в состоянии сильного насыщения. Важным этапом процесса выключения является удаление избыточного заряда накопленных носителей. Из-за падения напряжения в области $p2$ за счет протекания продольного тока величина смещения перехода $J3$ в положительном направлении падает от центра к управляющему контакту (рис. 32). Очевидно, что часть перехода

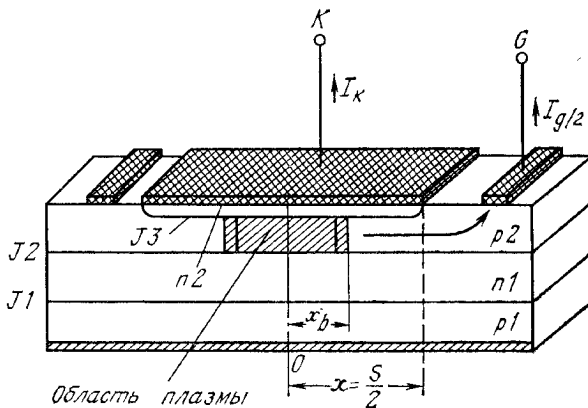


Рис. 32. Концентрация плазмы в базе выключаемого тиристора [36].

$J3$, наиболее близкая к контакту, под действием выключающего тока окажется смещенной в обратном направлении. С этого момента прямой ток будет отклоняться к той части перехода $J3$, которая еще остается прямосмещенной и будет последовательно вытесняться во все уменьшающуюся область, пока последняя не сократится до некоторого критического размера. С этого момента остаток избыточного заряда начнет быстро удаляться из $p2$ -области, и фаза рассасывания закончится. Время рассасывания определяется выражением

$$t_s = t_{p2} (\beta_{\text{off}} - 1) \ln \left(\frac{Sx_0/W_{p2}^2 + 2x_0^2/W_{p2}^2 - \beta_{\text{off}} + 1}{4x_0^2/W_{p2}^2 - \beta_{\text{off}} + 1} \right), \quad (66)$$

где t_{p2} — время пролета по области $p2$; S — длина катода; x_0 — постоянная затухания концентрации электронов для $x > x_b$, приблизительно равная диффузионной длине электронов; W_{p2} — ширина области $p2$. Время рассасывания возрастает с увеличением коэффициента усиления β_{off} . Существует компромисс между временем рассасывания и коэффициентом усиления выключением. Для сокращения времени рассасывания необходимо использовать низкие значения β_{off} (что приводит к увеличению управляющего тока). Время спада на рис. 31, б соответствует времени, необходимому для распространения обедненного слоя по переходу $J2$ и удалению заряда из этой области. Полный заряд на единицу площади в области $n1$ равен

$$Q \simeq qp^*W(V_A) \simeq J_A t_f, \quad (67)$$

где p^* — средняя концентрация дырок в области $n1$, W — ширина обедненного слоя при данном анодном напряжении V_A и J_A — плотность анодного тока. Из выражения (67) следует формула для времени спада

$$t_f \simeq \frac{qp^*W(V_A)}{J_A}. \quad (68)$$

Время спада уменьшается с увеличением плотности анодного тока и возрастает по закону $\sqrt{V_A}$, так как W изменяется приблизительно по закону $\sqrt{V_A}$.

Для надежного функционирования выключаемого тиристора необходимо, чтобы площадь сжатой плазмы в конечный момент была еще достаточно большой для того, чтобы плотность тока была бы не слишком велика. Из этого требования вытекает целесообразность встречноштыревой конструкции (рис. 21). Использование управления с усилением желательно также и для получения быстрого включения. Основное отличие выключаемого тиристора от описанного выше с дополнительным усилительным электродом заключается в том, что первый может быть выключен

подачей отрицательного смещения на управляющий электрод при сохранении положительного потенциала между анодом и катодом. Второй же для выключения требует коммутации приложенного напряжения, а обратное смещение на электроде используется для уменьшения времени выключения.

4.5. ДИАК И ТРИАК

Диак (диодный переключатель переменного тока) и триак (триодный переключатель переменного тока) являются двунаправленными тиристорами [37, 38]. Они могут быть включены и выключены и при положительном, и при отрицательном анодном напряжении и, следовательно, предназначены для работы с переменным током.

Структура с двумя диодами образует переключательный диод переменного тока и двунаправленный $p-n-p-n$ -переключатель. Первый является просто трехслойным прибором, аналогичным по конструкции биполярному транзистору, за исключением того, что концентрация примесей в обоих переходах примерно одинакова и отсутствует контакт к базе. Одинаковые уровни легирования обуславливают симметричную двунаправленную характеристику (рис. 33). При любой полярности напряжения на диаке один его переход смещен в прямом направлении, а другой — в обратном. Ток через него ограничен током утечки обратносмещенного перехода. Если приложенное напряжение достаточно велико, наступает пробой. Напряжение пробоя равно $BV_{CBO}(1-\alpha)^{1/n}$, где BV_{CBO} — напряжение лавинного пробоя $p-n$ -перехода, α — коэффициент усиления по току, n — постоянная величина. Выражение для напряжения пробоя то же, что и для $n-p-n$ -транзистора с разомкнутой базой (гл. 3). Если после наступления пробоя ток продолжает расти, то увеличение коэффициента α приведет к уменьшению падения напряжения на выводах прибора. Такое уменьшение означает возникновение области с отрицательным сопротивлением.

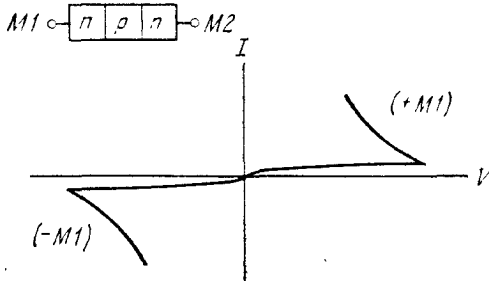


Рис. 33. Типичная характеристика переключательного диода переменного тока.

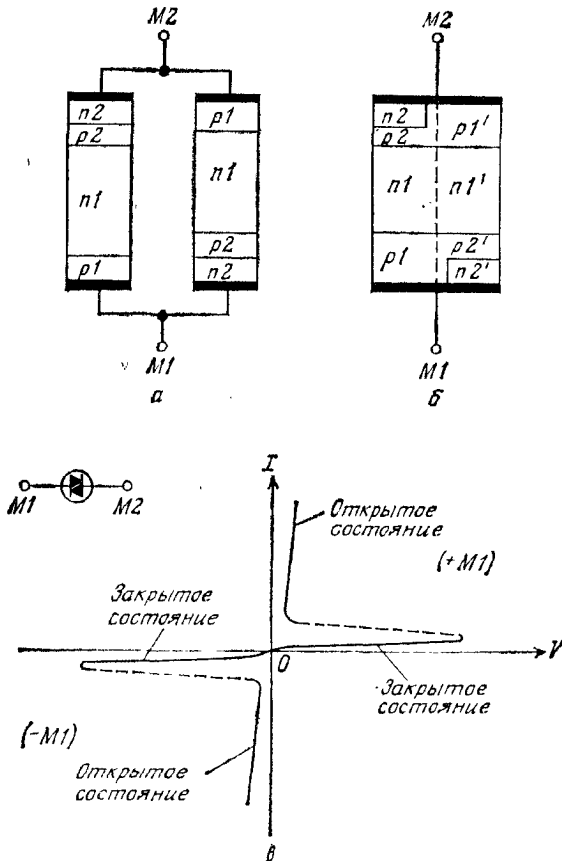


Рис. 34. Два диодных тиристора, соединенных навстречу друг другу (а), и их объединение в единый двухполюсник (б), а также вольт-амперная характеристика и условное обозначение диода (в).

Двунаправленный диодный $p-n-p-n$ -переключатель функционирует как два обычных диодных тиристора, соединенных навстречу друг другу, поэтому можно переключать напряжение обеих полярностей (рис. 34, а, где M1 означает магистральную клемму 1, а M2 — магистральную клемму 2). Используя принцип закорачивания катода, можно преобразовать такое соединение в один двухвходовый диак (рис. 34, б). Из симметрии структуры следует, что прибор может работать при любой полярности приложенного напряжения. Симметричная вольт-амперная характеристика прибора и его условное обозначение приведены на рис. 34, в. Подобно диодному тиристору, диак может включаться

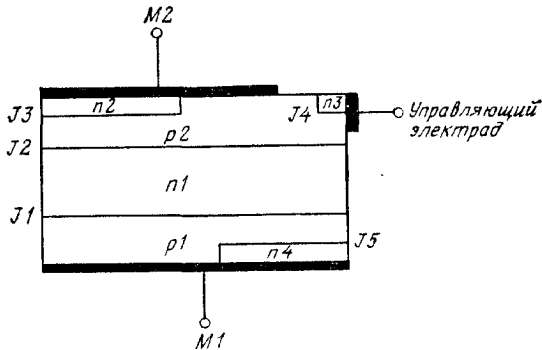


Рис. 35. Поперечное сечение триака (шестислойной структуры с пятью $p-n$ -переходами и тремя закороченными электродами).

в проводящее состояние при напряжении, превышающем напряжение включения, а также за счет эффекта dV/dt . Благодаря регенеративному характеру процесса двунаправленный диодный $p-n-p-n$ -переключатель имеет более высокое отрицательное сопротивление и меньшее прямое падение напряжения, чем диодный переключатель переменного тока.

Триак может переключать ток любого направления при подаче низковольтного импульса слабого тока, приложенного между управляющим электродом и одной из двух магистральных клемм $M1$ или $M2$ (рис. 35). Триак целесообразно использовать в регуляторах напряжения освещения, в схемах управления скоростью электродвигателей и в аналогичных приложениях. Конструкция триака значительно сложнее, чем конструкция обычных тиристоров. Кроме четырех основных слоев $p1-n1-p2-n2$ триак имеет область управляющего перехода $n3$ и область $n4$ вблизи контакта $M1$. Отметим также, что область $p1$ закорочена на $n4$, область $p2$ — на $n2$ и область $n3$ — на $p2$ с помощью клемм $M1$ и $M2$ и управляющего электрода.

Функционирование прибора при различных условиях смещения проиллюстрировано на рис. 36. Когда магистральная клемма $M1$ положительна по отношению к $M2$ и на управляющем электроде G напряжение положительно (также по отношению к $M2$), прибор аналогичен обычному тиристоры (рис. 36, а). Переход $J4$ смещен в обратном направлении, и управляющий ток протекает по шунту цепи управления вблизи области $n3$. Так как переход $J5$ также смещен в обратном направлении, то магистральный ток протекает по левой части $p1-n1-p2-n2$ -тиристора.

На рис. 36, б клемма $M1$ положительна по отношению к $M2$, но на управляющий электрод подано отрицательное напряжение. Переход $J4$ между слоями $n3$ и $p2$ теперь смещен в прямом направ-

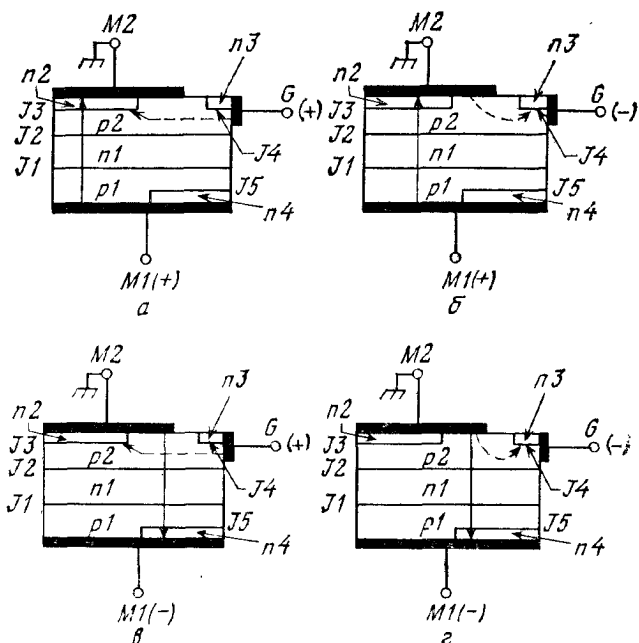


Рис. 36. Распределение токов при четырех способах включения триака [37].

лении, и электроны инжектируются из слоя $n3$ в слой $p2$. Вспомогательный тиристор $p1-n1-p2-n3$ включится продольным током по базе $p2$ к электроду $n3$ при возрастании усиления по току транзистора $n3-p2-n1$. Включение вспомогательного тиристора приведет к появлению тока по направлению к области $n2$, который служит запускающим током и включает левую часть тиристора $p1-n1-p2-n2$. Когда клемма $M1$ отрицательна по отношению к $M2$, а напряжение V_G положительно, переход $J3$ оказывается прямосмещенным между $M2$ и закороченным управляющим электродом (рис. 36, в). Электроны инжектируются из $n2$ в $p2$ и диффундируют к слою $n1$, вызывая в результате увеличение прямого смещения на переходе $J2$. За счет регенеративного процесса полный ток в конце концов потечет через шунт к $M2$. Управляющий переход $J4$ смещен в обратном направлении, поэтому полный ток через прибор протекает по правой части $p2-n1-p1-n4$ -тиристора.

На рис. 36, г клемма $M1$ отрицательна по отношению к $M2$ и напряжение V_G также отрицательно. При таких условиях переход $J4$ смещен в прямом направлении и включение вначале происходит за счет инжекции электронов из области $n3$ в область $n1$.

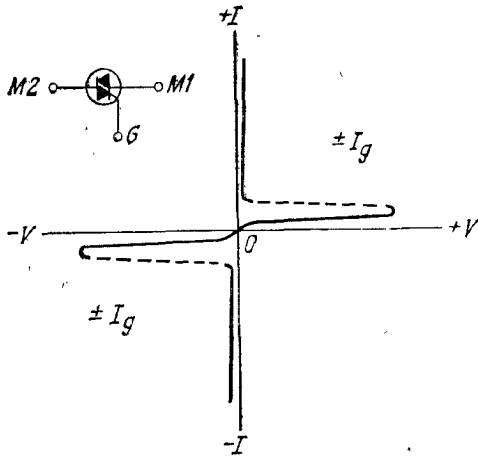


Рис. 37. Вольт-амперная характеристика и условное обозначение триака.

При этом потенциал слоя $n1$ снижается и из области $p2$ в область $n1$ начинают инжектировать дырки. Они создают базовый ток в транзисторе $p2-n1-p1$, и в конце концов правая часть $p2-n1-p1-n4$ -тиристора включается. Поскольку переход J3 смещен в обратном направлении, магистральный ток протекает из шунта к M2 через область $n4$.

Вольт-амперная характеристика и условное обозначение триака приведены на рис. 37. Триак является симметричным триодным переключателем, способным управлять нагрузкой, питаемой переменным током. Интеграция двух тиристоров на одном кристалле означает, что в любой момент времени работает только половина структуры (рис. 36). Следовательно, триак использует только половину площади двух произвольно соединенных тиристоров. Основное преимущество прибора заключается в хорошем согласовании выходных характеристик, экономии одного корпуса и двух внешних выводов. Однако его входные характеристики сильно рассогласованы. Для компенсации входного рассогласования используется диак. В настоящее время триаки перекрывают широкий диапазон рабочих напряжений (вплоть до 1600 В) и токов (выше 300 А).

4.6. ОДНОПЕРЕХОДНЫЕ ТРАНЗИСТОРЫ И ПЕРЕКЛЮЧАЮЩИЕ ТИРИСТОРЫ

В этом разделе рассмотрено несколько тиристоров малой мощности. Они пригодны для работы с низкими напряжениями и малыми токами. Одно из наиболее важных их применений — использование в качестве запускающих схем для мощных тиристоров.

4.6.1. Однопереходные транзисторы

Однопереходный транзистор принадлежит к семейству тиристоров. Подобно тиристорам, он обладает высоким импедансом в закрытом состоянии и низким импедансом в открытом состоянии. Однако переключение из закрытого в открытое состояние происходит не за счет регенеративного процесса, а за счет модуляции проводимости. Однопереходный транзистор является трехполюсником с одним эмиттерным переходом и двумя базовыми контактами. Конструкция прибора относится к сплавным структурам на брусках германия, впервые описанным Шоуки, Пирсоном и Хайнсом [39]. В то время такая структура называлась нитевидным транзистором. В процессе развития прибор имел объемную структуру, затем диффузионно-планарную и, наконец, эпитаксиально-планарную. Изменялось и его название от «диода с двойной базой» до последнего — «однопереходного транзистора» [40—44]. Его упрощенная конструкция приведена на рис. 38, а. Два омических базовых контакта называются базой 1 (В1) и базой 2 (В2). Между контактами В1 и В2 расположен $p-n$ -переход, называемый эмиттерным.

Эквивалентная схема однопереходного транзистора приведена на рис. 38, б. При нормальных условиях работы вывод В1 заземлен, а на базу В2 подано положительное напряжение смещения V_{BB} (рис. 38, а). Сопротивление между В1 и В2 обозначим через R_{BB} , между В2 и А — через R_{B2} и между А и В1 — через R_{B1} ($R_{BB} = R_{B2} + R_{B1}$). Величины G_p и G_n — избыточная дырочная и электронная проводимости между эмиттером и В1. Внешнее напряжение V_{BB} задает ток и электрическое поле вдоль полупроводникового стержня и приводит к появлению напряжения на n -области эмиттерного перехода, которое составляет η -ю часть приложенного напряжения V_{BB} . Коэффициент η называют вну-

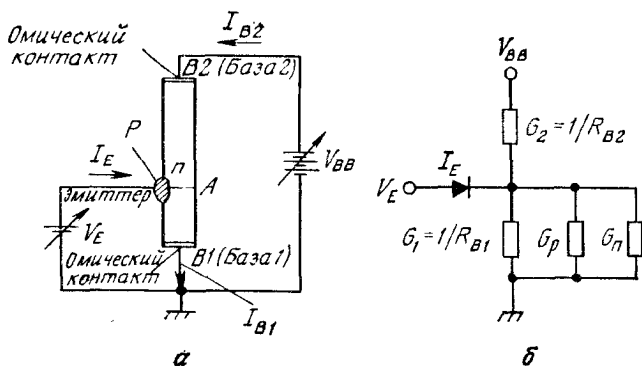


Рис. 38. Однопереходный транзистор.

а — упрощенная конструкция и схема включения; б — эквивалентная схема.

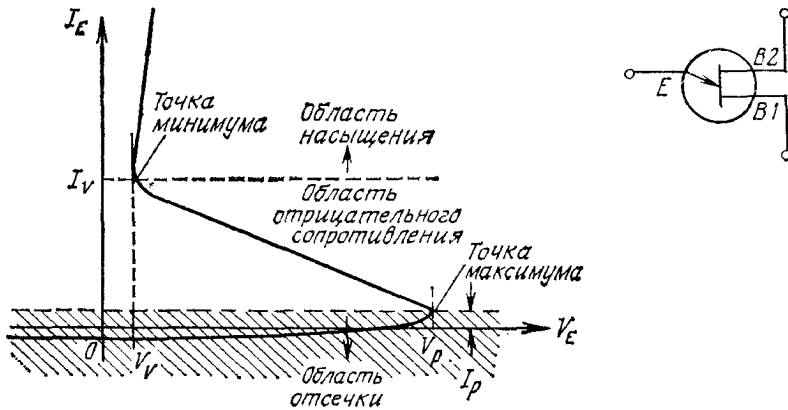


Рис. 39. Вольт-амперная характеристика и условное обозначение однопереходного транзистора.

тренним коэффициентом деления, и для него можно написать выражение

$$\eta \equiv \frac{R_{B1}}{R_{B1} + R_{B2}} = \frac{R_{B1}}{R_{BB}}. \quad (69)$$

Когда эмиттерное напряжение V_E меньше ηV_{BB} , эмиттерный переход смещен в обратном направлении и в цепи эмиттера протекает лишь небольшой обратный ток насыщения. Если напряжение V_E превышает ηV_{BB} на величину, равную прямому падению на эмиттерном переходе, в стержень инжектируются дырки. За счет электрического поля внутри полупроводникового стержня дырки будут двигаться по направлению к базе 1, увеличивая проводимость стержня в области между эмиттером и базой 1. При повышении тока I_E напряжение на эмиттере падает из-за увеличения проводимости, а прибор оказывается на участке характеристики с отрицательным сопротивлением.

Вольт-амперная характеристика и условное обозначение однопереходного транзистора приведены на рис. 39. Отметим две важные точки на кривой — максимальное и минимальное напряжения. В этих точках наклон $dV_E/dI_E = 0$. Область с током, меньшим I_P , называют областью отсечки (закрытое состояние). Область между максимальным и минимальным напряжениями называют областью отрицательного сопротивления (здесь важна модуляция проводимости). Область с током, превышающим I_V , называют областью насыщения (открытое состояние). Время переключения из точки с максимальным напряжением к точке с минимальным напряжением зависит от геометрии прибора и условий смещения. Оно пропорционально расстоянию от эмиттера до контакта к базе 1 [45].

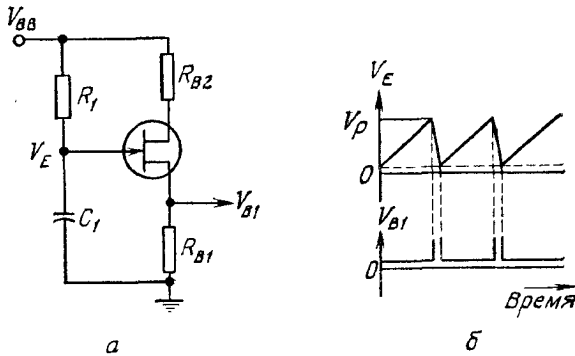


Рис. 40. Мультивибратор на однопереходном транзисторе [46].
 а — простейшая схема мультивибратора; б — форма релаксационных колебаний.

Однопереходный транзистор можно использовать в качестве мультивибратора (рис. 40, а) [46]. Конденсатор C_1 через сопротивление R_1 заряжается до тех пор, пока напряжение на нем не достигнет V_p , после чего транзистор переключится (открытое состояние) и конденсатор C_1 начинает разряжаться через сопротивление R_{B1} . Когда напряжение на эмиттере снижается до ~ 2 В, транзистор выключается и цикл повторяется. Формы напряжения на конденсаторе V_E и выходного напряжения приведены на рис. 40, б. Период колебаний равен

$$T \simeq R_1 C_1 \ln \left(\frac{1}{1 - \eta} \right). \quad (70)$$

Выходное напряжение может быть подано на тиристор для его переключения.

4.6.2. Программируемый однопереходный транзистор

Программируемый однопереходный транзистор представляет собой $p-n-p-n$ -прибор с анодным управляющим электродом (контакт управляющего электрода подсоединен к n -базе вместо p -базы, как в обычном тиристоре). Структура прибора, его условное обозначение и типичная схема включения приведены на рис. 41. Форма колебаний, наблюдаемых в такой схеме, похожа на форму колебаний в схеме с однопереходным транзистором (рис. 40, б). Период колебаний также вычисляется по формуле (70), если положить $\eta = R_2 / (R_1 + R_2)$, где R_1 и R_2 — сопротивления делителя напряжения. При заданном произведении RC частотой колебаний можно управлять, изменяя элементы R_1 и (или) R_2 . Следовательно, программируемый однопереходный транзистор является более гибким прибором для использования в качестве генератора, чем однопереходный транзистор.

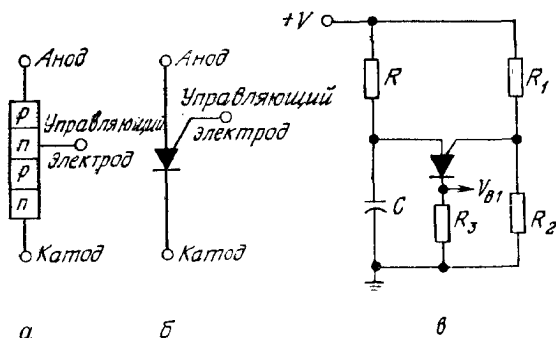


Рис. 41. Программируемый однопереходный транзистор [6].

a — структура; *б* — условное обозначение; *в* — типичная схема.

4.6.3. Кремниевый одноплавленный переключатель

Кремниевым одноплавленным переключателем называют $p-n-p$ -прибор с анодным управляющим электродом, аналогичный программируемому однопереходному транзистору. Это тиристор небольшого размера с лавинным диодом, встроенным между анодным управляющим электродом и катодом. Когда напряжение на анодном управляющем электроде превышает напряжение лавинного пробоя диода, переключатель включается. Его вольт-амперная характеристика такая же, как характеристика тиристора. Эти приборы могут быть использованы в качестве релаксационного генератора (рис. 40, *a*). Поскольку напряжение включения ограничено пробивным напряжением диода и ток включения достаточно велик, то диапазон возможных времен задержки генератора на кремниевом одноплавленном переключателе ограничен сильнее, чем для генератора на однопереходном транзисторе. Так как напряжение пробоя встроенного лавинного диода слабо зависит от температуры, то V_{BF} также обладает слабой температурной зависимостью. Следовательно, эти переключатели пригодны для создания стабильных генераторов низкого напряжения.

4.6.4. Кремниевый двунаправленный переключатель

Кремниевый двунаправленный переключатель состоит из двух одноплавленных переключателей, соединенных противоположными выводами и выполненных на одной и той же кремниевой подложке. Его вольт-амперная характеристика подобна характеристике триака. Этот переключатель работает при положительных и отрицательных полярностях приложенных напряжений. Он особенно полезен для запуска триака при разнополярных импульсах напряжения смещения V_G .

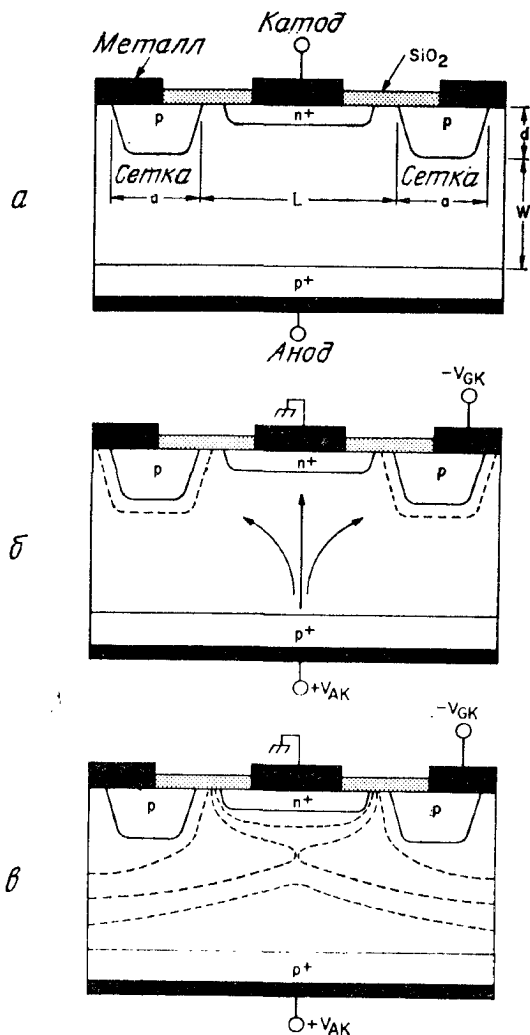


Рис. 42. Планарный полевой тиристор [47].

a — разрез структуры; b — использование обратносмещенной сетки для отклонения катодного тока; c — эквипотенциальные линии в обедненной области.

4.7. ПОЛЕВЫЕ ТИРИСТОРЫ

Полевыми тиристорами называют переключатели мощности, состоящие из p - v - n -диода с управляющим электродом в виде сетки [47] (рис. 42, a). Конструкция сетки аналогична той, что используется в мощных полевых транзисторах [48].

Если анодный и катодный переходы смещены в прямом направлении и сеточная цепь открыта, электроны и дырки инжектируются в область n -базы, снижают ее сопротивление, и в результате напряжение на приборе оказывается низким. Это состояние называется открытым. Если на сетку по отношению к катоду подано обратное напряжение, ток, текущий от анода к катоду, будет отводиться к сетке, которая в данном случае выступает как эффективный коллектор дырок (рис. 42, б). При достаточно больших напряжениях на сетке ее обедненная область достигает катодного контакта и возникает потенциальный барьер, показанный на рис. 42, в, где изображены эквипотенциальные линии в обедненной области. Потенциал вблизи анода положителен. По мере приближения к катоду он принимает нулевое значение, становится отрицательным, а затем вновь равен нулю у заземленного катода. Таким образом, существует потенциальная яма, представляющая барьер для электронов и препятствующая их инжекции из катода. В отсутствие источника электронов дырки также не могут инжектироваться анодом, и, следовательно, прибор оказывается в закрытом состоянии.

Вольт-амперная характеристика полевого тиристора приведена на рис. 43 [47, 49]. В открытом состоянии типичное прямое напряжение на приборе составляет ~ 1 В. В закрытом состоянии максимальное напряжение между анодом и катодом (напряжение включения V_{AK}) возрастает с увеличением отрицательного смещения на сетке V_{GK} . Коэффициент усиления запираания по определению равен

$$\mu \equiv - \frac{\Delta V_{AK}}{\Delta V_{GK}}. \quad (71)$$

На рис. 44 приведена зависимость напряжения включения от смещения на сетке и от глубины залегания сетки тиристора, изготовленного на кремневой подложке, легированной фосфором с сопротивлением $100 \text{ Ом} \cdot \text{см}$. При заданном смещении на сетке напряжение включения возрастает с увеличением глубины сетки. Коэффициент усиления запираания (рис. 44, б) экспоненциально растет с глубиной сетки. Обнаружено, что напряжение выключения уменьшается в 2—4 раза при увеличении глубины сетки в пределах 16—36 мкм.

Полевой тиристор может быть использован в схемах для быстрого включения и выключения обычных тиристоров, так как сам включается при снижении сеточного смещения и выключается при удалении неосновных носителей из базы без нарушения регенеративного процесса. Поскольку регенеративный процесс отсутствует, приборы устойчивы по отношению к эффекту dV/dt и могут функционировать при более высоких температурах, чем обычные тиристоры.

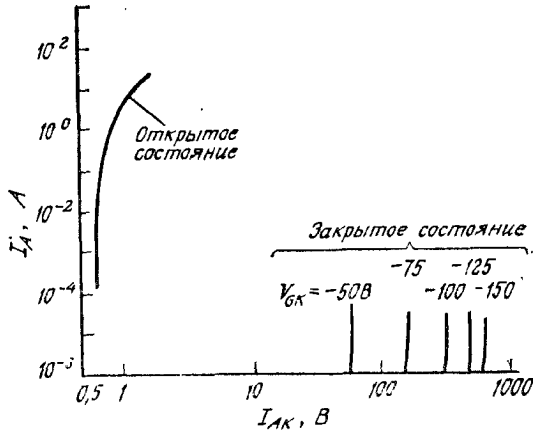


Рис. 43. Характеристики полевого тиристора в открытом и закрытом состояниях [47].

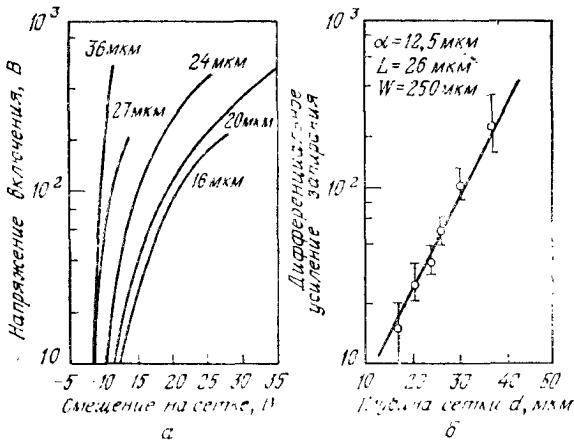


Рис. 44. Характеристики полевого тиристора в режиме прямого запирания [50]. а — зависимость напряжения включения от смещения; б — зависимость усиления запирания от глубины сетки.

ЛИТЕРАТУРА

1. Shockley W. Electronics and Holes in Semiconductors, D. Van Nostrand, Princeton, N. J., 1950, p. 112.
2. Ebers J. J. Four-Terminal $p-n-p-n$ Transistors, *Proc. IRE*, 40, 1361 (1952).
3. Moli J. L., Tanenbaum M., Goldey J. M., Holonyak N. $p-n-p-n$ Transistor Switches, *Proc. IRE*, 44, 1174 (1956).
4. Thyristor DATA Book, DATA Inc., Pine Brook, N. J., 1979.
5. Gentry F. E., Gutzwieler F. W., Holonyak N. H., Von Zastrow E. E. Semiconductor Controlled Rectifiers, Prentice-Hall, Englewood Cliffs, N. J., 1964.

6. Blicher A. Thyristor Physics, Springer, N. Y., 1976.
7. Ghandhi S. K. Semiconductor Power Devices, Wiley, N. Y., 1977.
8. Sze S. M., Gibbons G. Avalanche Breakdown Voltages of Abrupt and Linearly Graded $p - n$ Junctions in Ge, Si, GaAs, and GaP, *Appl. Phys. Lett.*, **8**, 111 (1966).
9. Herlet A. The Maximum Blocking Capability of Silicon Thyristors, *Solid State Electron.*, **8**, 655 (1965).
10. Special Issue on High-Power Semiconductor Devices, *IEEE Trans. Electron Devices*, **ED-23** (1976).
11. Haas E. W., Schnoller M. S. Phosphorus Doping of Silicon by Means of Neutron Irradiation, *IEEE Trans. Electron Devices*, **ED-23**, 803 (1976).
12. Cornu J., Schweitzer S., Kuhn O. Double Positive Beveling: A Better Edge Contour for High Voltage Devices, *IEEE Trans. Electron Devices*, **ED-21**, 181 (1974).
13. Davis R. L., Gentry F. E. Control of Electric Field at the Surface of $p - n$ Junctions, *IEEE Trans. Electron Devices*, **ED-11**, 313 (1964).
14. Adler M. S., Temple V. A. K. A General Method for Predicting the Avalanche Breakdown Voltage of Negative Bevelled Devices, *IEEE Trans. Electron Devices*, **ED-23**, 956 (1976).
15. Gentry F. E. Turn-on Criterion for $p - n - p - n$ Devices, *IEEE Trans. Electron Devices*, **ED-11**, 74 (1964).
16. Yang E. S., Voulgaris N. C. On the Variation of Small-Signal Alphas of a $p - n - p - n$ Device with Current, *Solid State Electron*, **10**, 641 (1967).
17. Adler M. S. Accurate Calculation of the Forward Drop and Power Dissipation in Thyristors, *IEEE Trans. Electron Devices*, **ED-25**, 16 (1978).
18. Krause J. Auger Recombination in Forward Biased Silicon Rectifiers and Thyristors, *Solid State Electron.*, **17**, 427 (1974).
- 19a. Schroen W. H. Characteristics of a High-Current, High-Voltage Shockley Diode, *IEEE Trans. Electron Devices*, **ED-17**, 694 (1970).
- 19b. Chu C. K., Johnson J. E., Brewster J. B. 1200V and 5000 A Peak Reverse Blocking Diode Thyristor, *Jpn. J. Appl. Phys.*, **16**, *Suppl.*, **16-1**, 537 (1977).
20. Gibbons J. F. Graphical Analysis of the I-V Characteristics of Generalized $p - n - p - n$ Devices, *Proc. IEEE*, **55**, 1366 (1967).
21. Gibbons J. F. A Critique of the Theory of $p - n - p - n$ Devices, *IEEE Trans. Electron Devices*, **ED-11**, 406 (1964).
22. Yang E. S. Turn-off Characteristics of $p - n - p - n$ Devices, *Solid State Electron*, **10**, 927 (1967).
23. Sundresh T. S. Reverse Transient in $p - n - p - n$ Triodes, *IEEE Trans. Electron Devices*, **ED-14**, 400 (1967).
24. Baliga B. J., Sun E. Comparison of Gold, Platinum, and Electron Irradiation for Controlling Lifetime in Power Rectifiers, *IEEE Trans. Electron Devices*, **ED-24**, 685 (1977).
25. Dubeck I., Kassing R. Gold as an Optimum Recombination Center for Power Rectifiers and Thyristors, *Solid State Electron.*, **20**, 1033 (1977).
26. Baliga B. J., Krishna S. Optimization of Recombination Levels and Their Capture Cross Section in Power Rectifiers and Thyristors, *Solid State Electron.*, **20**, 225 (1977).
27. Schlegel E. S. A Technique for Optimizing the Design of Power Semiconductor Devices, *IEEE Trans. Electron Devices*, **ED-23**, 925 (1976).
28. Shimizu J., Oka H., Funakawa S., Gamo H., Iida T., Kawakami A. High-Voltage High-Power Gate-Assisted Turn-Off Thyristor for High-Frequency Use, *IEEE Trans. Electron Devices*, **ED-23**, 883 (1976).
29. Schlegel E. Gate Assisted Turn-Off Thyristors, *IEEE Trans. Electron Devices*, **ED-23**, 888 (1976).
30. Munoz-Yague A., Leturcq P. Optimum Design of Thyristor Gate-Emitter Geometry, *IEEE Trans. Electron Devices*, **ED-23**, 917 (1976).

31. Storm H. F., Clair J. G. St. An Involute Gate-Emitter Configuration for Thyristors, *IEEE Trans. Electron Devices*, **ED-21**, 520 (1974).
32. Gentry F. E., Moynson J. The Amplifying Gate Thyristor, Paper No. 19.1, IEEE Meet. Prof. Group Electron Devices, Washington, D. C., 1968.
33. Roberts F. M., Wilkinson E. L. G. The Relative Merits of Thyristors and Power Transistors for Fast Power-Switching Application, *Int. J. Electron.*, **33**, 319 (1972).
34. Kokosa R. A., Tuit B. R. A High-Voltage High-Temperature Reverse Conducting Thyristor, *IEEE Trans. Electron Devices*, **ED-17**, 667 (1970).
35. Gerlach W. Light Activated Power Thyristors, *Inst. Phys. Conf. Ser.*, **32**, 111 (1977).
36. Wolley E. D. Gate Turn-Off in $p - n - p - n$ Devices, *IEEE Trans. Electron Devices*, **ED-13**, 590 (1966).
37. Gentry F. E., Scace R. I., Flowers J. K. Bidirectional Triode $p - n - p - n$ Switches, *Proc. IEEE*, **53**, 355 (1965).
38. Esson J. F. Bidirectional Triode Thyristor Applied Voltage Rate Effect Following Conduction, *Proc. IEEE*, **55**, 1312 (1967).
39. Shockley W., Pearson G. L., Haynes J. R. Hole Injection in Germanium — Quantitative Studies and Filamentary Transistors, *Bell Syst. Tech. J.*, **28**, 344 (1949).
40. Bluhm V. A., Sylvan T. P. A High Performance Unijunction Transistor Using Conductivity Modulation of Spreading Resistance, *Solid State Des.*, **5**, 26 (1964).
41. Senhouse L. S. A Unique Filamentary-Transistor Structure, Paper No. 23.6, IEEE Electron Device Meet., Washington, D. C., Oct. 1967.
42. Lesk I. A., Mathis V. P. The Double-Base Diode — A New Semiconductor Device, *IRE Conv. Rec.*, Pt. 6, p. 2 (1963).
43. Trofimenkoff F. N., Huff G. J. DC Theory of the Unijunction Transistor, *Int. J. Electron.*, **20**, 217 (1966).
44. Clark L. E. Now, New Unijunction Geometries, *Electronics*, **38**, 93 (1965).
45. Scharfetter D. L., Jordan A. G. Reactive Effects in Semiconductor Filaments Due to Conductivity Modulation and an Extension of the Theory of the Double-Base Diode, *IRE Trans. Electron Devices*, **ED-9**, 461 (1962).
46. SCR Manual, 5th ed., General Electric, Syracuse, N. Y., 1972.
47. Houston D. E., Krishna S., Piccone D. E., Einke R. J., Sun Y. S. A Field Terminated Diode, *IEEE Trans. Electron Devices*, **ED-23**, 905 (1976).
48. Zuleeg R. Multi-Channel Field-Effect Transistor Theory and Experiment, *Solid State Electron.*, **10**, 559 (1967).
49. Barandon R., Laurenceau P. Power Bipolar Gridistor, *Electron. Lett.*, **12**, 486 (1976).
50. Baliga B. J. Grid Depth Dependence of the Characteristics of Vertical Channel Field Controlled Thyristors, *Solid State Electron.*, **22**, 237 (1979).

Часть III

Униполярные приборы

Глава 5

КОНТАКТЫ МЕТАЛЛ — ПОЛУПРОВОДНИК

5.1. ВВЕДЕНИЕ

Первые исследования выпрямляющих свойств контактов металл — полупроводник обычно приписывают Брауну, обнаружившему в 1874 г. зависимость полного сопротивления контактов от полярности приложенного напряжения и особенностей их изготовления [1]. Практическое использование выпрямляющих свойств точечного контакта началось с 1904 г. [2]. В 1931 г. Вильсон построил теорию переноса заряда в полупроводниках, основанную на зонной теории твердых тел [3]. Впоследствии эта теория была применена к контактам металл—полупроводник. В 1938 г. Шоттки высказал предположение, что потенциальный барьер создается неподвижным пространственным зарядом в полупроводнике, а не за счет возникновения между металлом и полупроводником промежуточного химического слоя [4]. Такой барьер называют барьером Шоттки. В 1938 г. Мотт также предложил соответствующую теоретическую модель для контактов металла к тонким слоям полупроводника [5]. Образующийся при этом барьер получил название барьера Мотта. Основополагающие теории и исторический обзор экспериментальных исследований в этой области изложены в монографии [6].

Контакты металл—полупроводник широко используются для выпрямления тока, в качестве одного из основных элементов ряда полупроводниковых СВЧ-устройств, а также для исследования фундаментальных физических параметров полупроводниковых материалов и поэтому изучены достаточно подробно. Теория процессов переноса и технология изготовления приборов рассмотрены в обзорах [7, 8]. В последующих главах также рассмотрены применения контактов металл—полупроводник. Такие контакты, в частности, используются как затворы в полевых транзисторах (гл. 6), в качестве стока и истока в МОП-транзисторах (гл. 8), электродов в мощных ЛПД-генераторах (гл. 10), третьего электрода в приборах, основанных на междолинных переходах (гл. 11), как фотодетекторы и солнечные элементы (гл. 13 и 14).

5.2. Зонные диаграммы

Ниже рассмотрены основные зонные энергетические диаграммы барьера, образующегося при контакте металла с полупроводником. Показано, что обедненный слой контакта металл—полупроводник аналогичен обедненному слою в резком асимметричном (например, $p^+ - n$) переходе.

5.2.1. Идеализированная модель и поверхностные состояния

При непосредственном контакте металла с полупроводником уровни Ферми этих материалов при термодинамическом равновесии должны совпадать. Рассмотрим сначала два предельных случая (рис. 1) [6]. Более общая ситуация рассмотрена в разд. 5.5. На рис. 1, а показаны энергетические диаграммы для идеального контакта металла с полупроводником n -типа при отсутствии поверхностных состояний. На первом слева рисунке металл и полупроводник не приведены в соприкосновение друг с другом и система не находится в термодинамическом равновесии. Если затем их электрически соединить, то из полупроводника в металл перетечет некоторый заряд и установится термодинамическое равновесие. При этом уровни Ферми в обоих материалах сравняются, т. е. уровень Ферми в полупроводнике понизится относительно уровня Ферми металла на величину, равную разности соответствующих работ выхода.

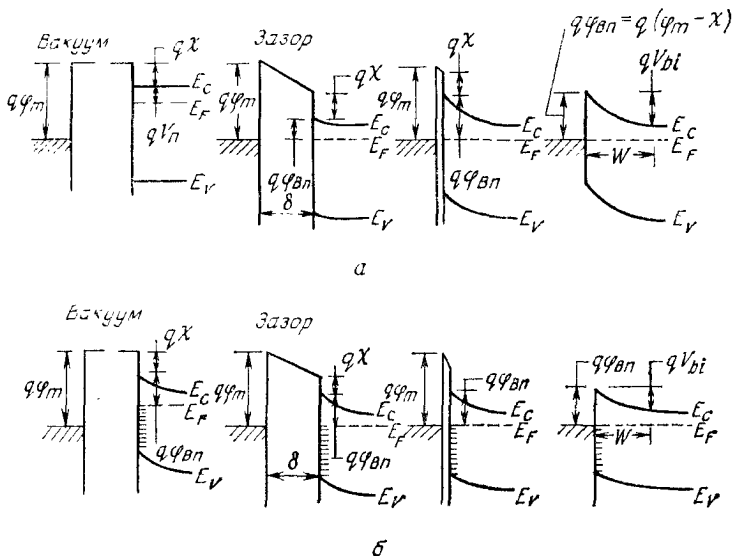


Рис. 1. Зонные энергетические диаграммы контактов металл — полупроводник [6].

Работой выхода называется разность энергий между уровнем вакуума и уровнем Ферми. Для металла эта величина составляет $q\varphi_m$ (φ_m измеряется в вольтах), а в полупроводнике она равна $q(\chi + V_n)$, где $q\chi$ — электронное сродство, т. е. разность между энергией дна зоны проводимости E_c и уровнем вакуума, а qV_n — положение уровня Ферми в запрещенной зоне полупроводника. Разность $q\varphi_m - q(\chi + V_n)$ называется контактной разностью потенциалов. По мере уменьшения расстояния δ отрицательный заряд на поверхности металла увеличивается. При этом в полупроводнике образуется равный ему по величине положительный заряд. Вследствие относительно низкой концентрации носителей этот положительный заряд распределен в некоторой области вблизи поверхности полупроводника. Когда расстояние δ становится сравнимым с межатомными расстояниями и зазор становится пропущиваемым для электронов, имеет место первый предельный случай (первый справа на рис. 1, а). Очевидно, что высота барьера в этом предельном случае равна разности между работой выхода металла и электронным сродством полупроводника:

$$q\varphi_{Vn} = q(\varphi_m - \chi). \quad (1)$$

При идеальном контакте между металлом и полупроводником p -типа высота барьера $q\varphi_{Vp}$ определяется аналогичным выражением

$$q\varphi_{Vp} = E_g - q(\varphi_m - \chi). \quad (2)$$

Для данного полупроводника и любого металла сумма высот барьеров на образцах n - и p -типа должна, таким образом, быть равной ширине запрещенной зоны, т. е.

$$q(\varphi_{Vn} + \varphi_{Vp}) = E_g. \quad (3)$$

Второй предельный случай, когда на поверхности полупроводника имеется большая плотность поверхностных состояний, показан на рис. 1, б. На первом слева рисунке показана ситуация, соответствующая равновесию между поверхностными состояниями и объемными состояниями полупроводника при отсутствии термодинамического равновесия между металлом и полупроводником. Поверхностные состояния в этом случае заполнены вплоть до уровня Ферми E_F . Когда система металл—полупроводник приходит в равновесие, уровень Ферми полупроводника понижается относительно уровня Ферми металла на величину, равную контактной разности потенциалов, в результате чего в зазоре δ возникает электрическое поле. Если плотность поверхностных состояний достаточно велика и «принимает» на себя весь дополнительный положительный заряд, возникающий по мере уменьшения δ , без заметного сдвига уровня заполнения E_F , то величина пространственного заряда в полупроводнике остается прежней.

То есть в этом случае высота барьера определяется свойствами поверхности полупроводника и не зависит от работы выхода металла.

5.2.2. Обедненный слой

Из сказанного выше ясно, что при контакте металла с полупроводником валентная зона проводимости полупроводника занимает определенное энергетическое положение по отношению к уровню Ферми металла. Если это положение известно, то оно служит граничным условием при решении уравнения Пуассона в полупроводнике, которое записывается в том же виде, что и для случая p - n -переходов. Зонные энергетические диаграммы контактов металла с полупроводниками n - и p -типов при разных смещениях показаны на рис. 2.

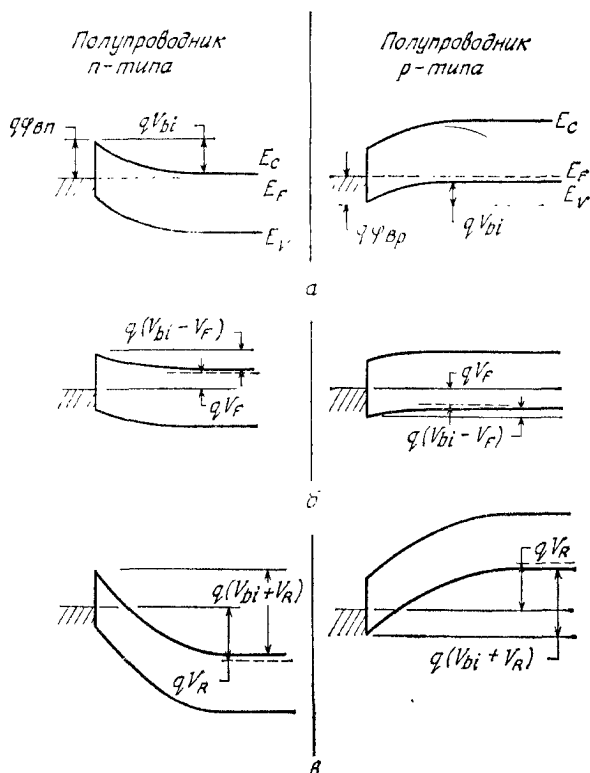


Рис. 2. Зонные энергетические диаграммы контактов металла с полупроводниками n - и p -типов при разных смещениях, a — при термодинамическом равновесии; b — при прямом смещении; $в$ — при обратном смещении.

В приближении резкой границы обедненного слоя (т. е. $\rho \approx qN_D$ при $x < W$ и $\rho \approx 0$, $dV/dx \approx 0$ при $x > W$, где W — ширина обедненного слоя) для барьера металл — полупроводник, как и для резкого несимметричного p — n -перехода, получаем

$$W = \sqrt{\frac{2\epsilon_s}{qN_D} \left(V_{bi} - V - \frac{kT}{q} \right)}, \quad (4)$$

$$|\mathcal{E}(x)| = \frac{qN_D}{\epsilon_s} (W - x) = \mathcal{E}_m - \frac{qN_D}{\epsilon_s} x, \quad (5)$$

$$V(x) = \frac{qN_D}{\epsilon_s} \left(Wx - \frac{1}{2} x^2 \right) - \Phi_{Bn}, \quad (6)$$

где слагаемое kT/q возникает за счет вклада в электрическое поле хвоста функции распределения основных носителей, а \mathcal{E}_m — максимальная величина напряженности электрического поля, которая достигается при $x = 0$:

$$\begin{aligned} \mathcal{E}_m = \mathcal{E}(x=0) &= \sqrt{\frac{2qN_D}{\epsilon_s} \left(V_{bi} - V - \frac{kT}{q} \right)} = \\ &= \frac{2(V_{bi} - V - kT/q)}{W}. \end{aligned} \quad (7)$$

Величина пространственного заряда на единицу площади поверхности полупроводника Q_{sc} и соответствующая удельная емкость обедненного слоя C определяются формулами

$$Q_{sc} = qN_D W = \sqrt{2q^2 N_D \left(V_{bi} - V - \frac{kT}{q} \right)} \quad [\text{Кл/см}^2], \quad (8)$$

$$C \equiv \left| \frac{\partial Q_{sc}}{\partial V} \right| = \sqrt{\frac{q^2 N_D}{2(V_{bi} - V - kT/q)}} = \frac{q}{W} \quad [\text{Ф/см}^2]. \quad (9)$$

Уравнение (9) можно переписать в виде

$$\frac{1}{C^2} = \frac{2(V_{bi} - V - kT/q)}{q\epsilon_s N_D}, \quad (10a)$$

или

$$-\frac{d(1/C^2)}{dV} = \frac{2}{q\epsilon_s N_D}, \quad (10б)$$

$$N_D = \frac{2}{q\epsilon_s} \left[-\frac{1}{d(1/C^2)/dV} \right]. \quad (10в)$$

Если концентрация N_D постоянна во всей области обедненного слоя, то на графике зависимости $1/C^2$ от V мы получим прямую

линию. Если концентрация N_D не постоянна, то, измеряя дифференциальную емкость, можно с помощью уравнения (10в) определить профиль легирования.

5.3. ЭФФЕКТ ШОТТКИ

Эмиссии электронов из металла препятствует потенциальный барьер, образующийся за счет электрических сил изображения. Понижение этого барьера по мере увеличения приложенного внешнего электрического поля называется эффектом Шоттки. Рассмотрим сначала систему металл—вакуум. Минимальная энергия, необходимая для перехода электрона с уровня Ферми в вакуум, называется работой выхода $q\varphi_m$ (φ_m измеряется в вольтах) (рис. 3). Для металлов величина $q\varphi_m$ обычно колеблется в пределах 2—6 эВ и очень чувствительна к загрязнениям поверхности. Наиболее надежные значения работы выхода для чистых поверхностей [9] приведены на рис. 4.

Электрон, находящийся в вакууме на некотором расстоянии x от металла, индуцирует на его поверхности положительный заряд. Сила притяжения между электроном и этим индуцированным поверхностным зарядом равна по величине силе притяжения к эффективному положительному заряду $+q$, называемому зарядом изображения. Эта сила, называемая силой изображения, равна

$$F = \frac{-q^2}{4\pi(2x)^2\epsilon_0} = \frac{-q^2}{16\pi\epsilon_0 x^2}, \quad (11)$$

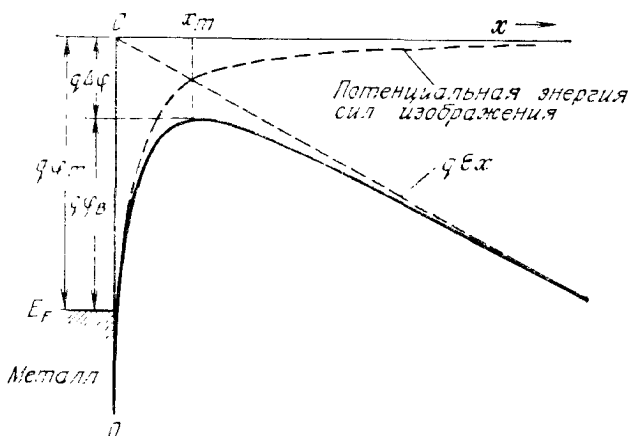


Рис. 3. Энергетическая диаграмма системы металл — вакуум.

Эффективная работа выхода (т. е. высота энергетического барьера) при приложении внешнего электрического поля уменьшается. Это уменьшение является следствием суперпозиции внешнего поля и силы изображения.

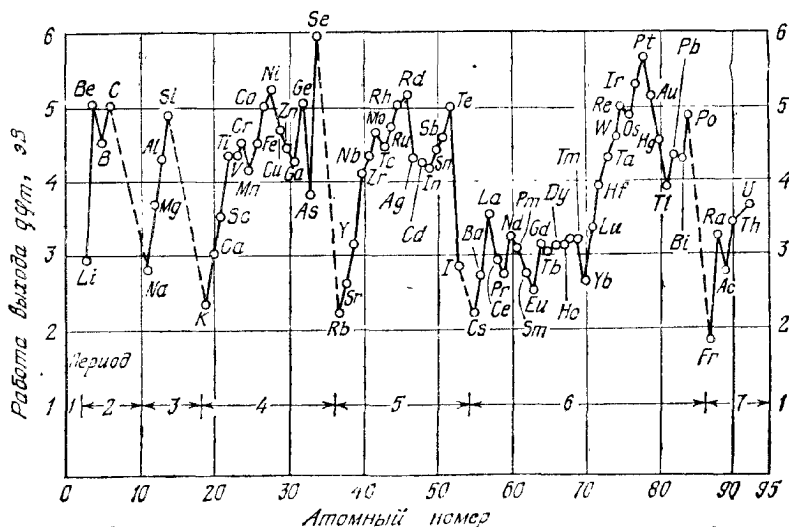


Рис. 4. Значения работ выхода для чистых поверхностей металлов в зависимости от их атомного номера. Отметим периодический характер этой зависимости внутри каждой группы [9].

где ϵ_0 — диэлектрическая постоянная вакуума. Работа, которую необходимо совершить, чтобы переместить электрон из бесконечности в точку x , равна

$$E(x) = \int_{\infty}^x F dx = \frac{q^2}{16\pi\epsilon_0 x}. \quad (12)$$

Эта работа есть потенциальная энергия электрона на расстоянии x от поверхности. Зависимость $E(x)$ приведена на рис. 3 штриховой линией.

Если к системе приложено внешнее электрическое поле \mathcal{E} , то потенциальная энергия электрона PE равна сумме

$$PE(x) = \frac{q^2}{16\pi\epsilon_0 x} + q\mathcal{E}x \text{ [эВ]}. \quad (13)$$

Понижение барьера Шоттки $\Delta\phi$ и расстояние x_m , на котором величина потенциала достигает максимума (рис. 3), определяются из условия $d[PE(x)]/dx = 0$. Отсюда

$$x_m = \sqrt{\frac{q}{16\pi\epsilon_0 \mathcal{E}}} \text{ [см]}, \quad (14)$$

$$\Delta\phi = \sqrt{\frac{q\mathcal{E}}{4\pi\epsilon_0}} = 2\mathcal{E}x_m \text{ [В]}. \quad (15)$$

Из уравнений (14) и (15) получаем $\Delta\varphi = 0,12$ В, $x_m \approx 60$ Å при $\mathcal{E} = 10^5$ В·см⁻¹ и $\Delta\varphi = 1,2$ В, $x_m \approx 10$ Å при $\mathcal{E} = 10^7$ В·см⁻¹. Таким образом, сильные электрические поля вызывают значительное понижение барьера Шоттки. В результате эффективная работа выхода металла для термоэлектронной эмиссии ($q\varphi_B$) уменьшается.

Полученные выше результаты могут быть перенесены на системы металл — полупроводник. В этом случае электрическое поле \mathcal{E} заменяется полем в полупроводнике вблизи границы раздела (где оно достигает своего максимального значения), а диэлектрическая постоянная вакуума ϵ_0 заменяется диэлектрической проницаемостью, характеризующей полупроводник (ϵ_s), т. е.

$$\Delta\varphi = \sqrt{\frac{q\mathcal{E}'}{4\pi\epsilon_s}}. \quad (15a)$$

Значение ϵ_s может отличаться от статической диэлектрической проницаемости полупроводника. Последнее объясняется тем, что, если время пролета электрона от поверхности раздела металл — полупроводник до точки x_m (x_m — точка, в которой потенциальная энергия электрона достигает своего максимального значения) меньше времени диэлектрической релаксации полупроводника, то последний не успевает поляризоваться. Поэтому наблюдаемое значение диэлектрической проницаемости может оказаться меньше статической (низкочастотной) диэлектрической проницаемости. В кремнии, однако, эти величины практически совпадают друг с другом.

Эффективная диэлектрическая проницаемость ϵ_s/ϵ_0 для контакта золото — кремний определена по результатам фотоэлектрических измерений, которые более подробно мы рассмотрим в разд. 5.5. На рис. 5 приведено измеренное при этом уменьшение высоты потенциального барьера как функция от корня квадратного из электрического поля [10]. Из полученной зависимости с помощью уравнения (15a) находим, что эффективная диэлектрическая проницаемость сил изображения равна $12 \pm 0,5$. При $\epsilon_s/\epsilon_0 = 12$ расстояние x_m изменяется от 10 до 50 Å в диапазоне изменения электрического поля, показанном на рис. 5. Если предположить, что скорость носителей $\sim 10^7$ см·с⁻¹, то время пролета этих расстояний составляет 10^{-14} — $5 \cdot 10^{-14}$ с. Оказывается, что диэлектрическая проницаемость, полученная из формулы для силы изображения, близка к значениям диэлектрической проницаемости (~ 12) для электромагнитного излучения соответствующих частот (с длиной волны 3—15 мкм) [11]. Так как диэлектрическая проницаемость кремния (11,7) практически постоянна в диапазоне частот от нуля до частоты, соответствующей $\lambda = 1$ мкм, за время пролета электрона через обедненный слой решетка успевает поляризоваться. Поэтому значения диэлектрической проницаемости,

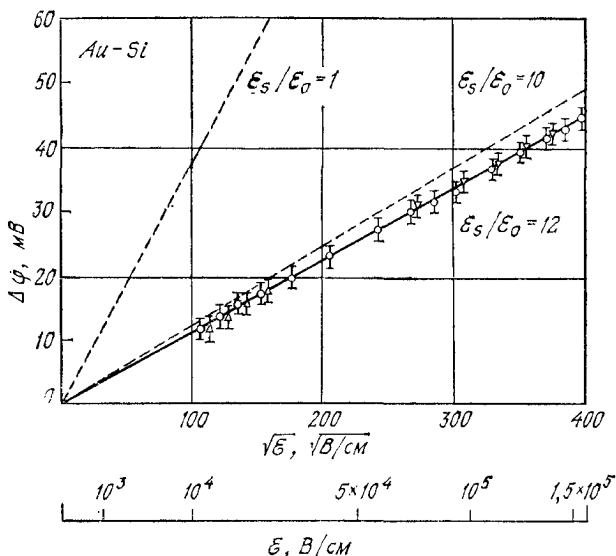


Рис. 5. Понижение энергетического барьера как функция электрического поля в диодах Au — Si.

полученные из фотоэлектрических и оптических измерений, близки друг к другу. Германий и арсенид галлия имеют аналогичную частотную зависимость диэлектрической проницаемости. Поэтому можно ожидать, что для этих полупроводников значение

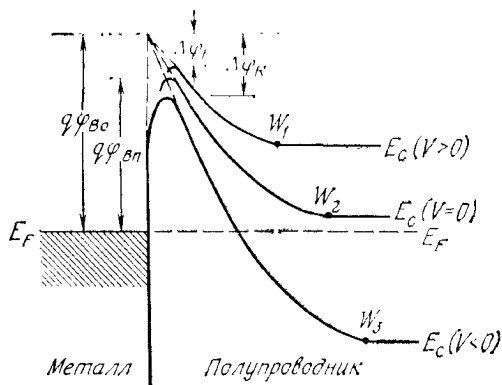


Рис. 6. Энергетические диаграммы барьера Шоттки между металлом и полупроводником *n*-типа при различных напряжениях смещения [8].

$q\varphi_{B0}$ — высота барьера в отсутствие поля; $q\varphi_{Bn}$ — высота барьера при термодинамическом равновесии; $\Delta\varphi_F$ — повышение барьера при прямом смещении; $\Delta\varphi_R$ — понижение барьера при обратном смещении.

диэлектрической проницаемости, определяющей силу изображения, в указанном выше интервале полей приблизительно совпадает со статическим значением.

На рис. 6 приведены энергетические диаграммы барьера Шоттки между металлом и полупроводником n -типа при различных напряжениях смещения. Отметим, что увеличение высоты барьера $q\Phi_{\text{во}} - q\Delta\Phi_{\text{Ф}}$ при прямом смещении ($V > 0$) и ее уменьшение при обратном ($V < 0$) являются очень малыми по сравнению с изменением высоты барьера в системе металл—вакуум. Это объясняется большой величиной ϵ_s . Например, при $\epsilon_s = 12\epsilon_0$ и $\mathcal{E} = 10^5 \text{ В} \cdot \text{см}^{-1}$ из уравнения (15а) получаем, что $\Delta\phi$ составляет всего 0,035 В, а при меньших полях это значение еще меньше. Хотя понижение барьера является незначительным, оно оказывает существенное влияние на процессы переноса в системах металл—полупроводник. Этот вопрос мы обсудим в разд. 5.4.

5.4. ТЕОРИИ ПРОЦЕССОВ ПЕРЕНОСА ЗАРЯДА

Перенос заряда через контакт металл—полупроводник осуществляется главным образом основными носителями в отличие от p — n -переходов, где электрический ток обусловлен неосновными носителями. На рис. 7 показаны четыре основных транспортных процесса при прямом смещении контакта Шоттки (при отрицательном смещении имеют место обратные процессы) [7]:

1) надбарьерный перенос, преобладающий в диодах Шоттки на умеренно легированных полупроводниках ($N_D \leq 10^{17} \text{ см}^{-3}$ для Si) при не слишком низких температурах ($\sim 300 \text{ К}$); 2) квантовомеханическое туннелирование электронов через барьер (важное при более высоких уровнях легирования, например, в большинстве омических контактов); 3) рекомбинация в области пространственного заряда (процесс, аналогичный рекомбинации в p — n -переходе) (гл. 2); 4) инжекция дырок из металла в полу-

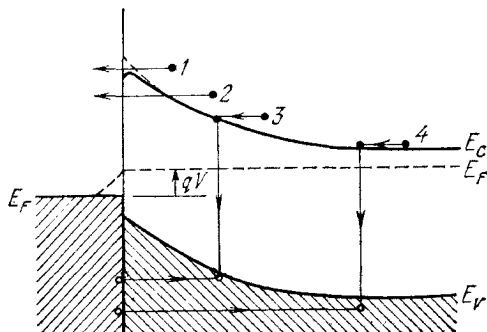


Рис. 7. Четыре основных процесса переноса при прямом смещении [7].

проводник (процесс, аналогичный рекомбинации в нейтральном объеме). Кроме того, вклад в полный ток через диод Шоттки могут давать ловушки вблизи границы раздела металл — полупроводник и токи утечки в периферийных областях контакта, обусловленные высокими краевыми электрическими полями. Для уменьшения тока через ловушки используют различные методы улучшения качества поверхности. Значительно снизить либо вообще устранить краевые токи утечки позволяет целый ряд предложенных в настоящее время конструктивных решений (разд. 5.6).

Рассмотрим сначала надбарьерный перенос электронов. В полупроводниках с высокой подвижностью (например, в Si) процесс переноса может быть полностью описан в рамках теории термоэлектронной эмиссии. Затем мы рассмотрим диффузионную теорию, описывающую контакты к полупроводникам с низкой подвижностью, и полную эмиссионно-диффузионную теорию, являющуюся обобщением первых двух.

5.4.1. Теория термоэлектронной эмиссии

В теории термоэлектронной эмиссии, развитой Бете [12], предполагается, что 1) высота барьера $q\phi_{Bn}$ много больше kT ; 2) область, определяющая термоэлектронную эмиссию, находится в термодинамическом равновесии; 3) протекание полного тока не нарушает этого равновесия. Данные предположения позволяют считать, что полный ток представляет собой разность между током из металла в полупроводник и противоположным ему током из полупроводника в металл, причем металл и полупроводник характеризуются каждый своим квазиуровнем Ферми. Ясно, что в этом случае ток не зависит от формы барьера, а зависит лишь от его высоты.

Плотность тока из полупроводника в металл $J_{s \rightarrow m}$ определяется числом электронов,двигающихся к металлу (в направлении x) с энергией, достаточной для преодоления потенциального барьера:

$$J_{s \rightarrow m} = \int_{E_F + q\phi_B}^{\infty} qv_x dn, \quad (16)$$

где $E_F + q\phi_B$ — минимальная энергия, необходимая для термоэлектронной эмиссии в металл; v_x — скорость носителей в направлении переноса. Концентрация электронов с энергией в интервале от E до $E + dE$ равна

$$dn = N(E) F(E) dE = \frac{4\pi (2m^*)^{3/2}}{h^3} \sqrt{E - E_c} \exp[-(E - E_c + qV_n)/kT] dE, \quad (17)$$

где $N(E)$ — плотность состояний, $F(E)$ — функция распределения носителей по энергии, m^* — эффективная масса в полупроводнике, $qV_n \equiv E_C - E_F$.

Если предположить, что полная энергия электрона в зоне проводимости представляет собой только кинетическую энергию, то

$$E - E_C = \frac{1}{2} m^* v^2, \quad (18a)$$

$$dE = m^* v dv, \quad (18б)$$

$$\sqrt{E - E_C} = v \sqrt{m^*/2}. \quad (18в)$$

Подставляя формулы (18a) — (18в) в формулу (17), получим

$$dn = 2 \left(\frac{m^*}{h} \right)^3 \exp \left(- \frac{qV_n}{kT} \right) \exp \left(- \frac{m^* v^2}{2kT} \right) (4\pi v^2 dv). \quad (19)$$

Уравнение (19) определяет число электронов в единице объема со скоростями в интервале от v до $v + dv$, двигающихся в любых направлениях. Разлагая скорость электронов на компоненты вдоль осей и выбирая ось x параллельно направлению переноса, имеем

$$v^2 = v_x^2 + v_y^2 + v_z^2. \quad (20)$$

После преобразования $4\pi v^2 dv = dv_x dv_y dv_z$ из уравнений (16), (19) и (20) получаем

$$\begin{aligned} J_{s \rightarrow m} &= 2q \left(\frac{m^*}{h} \right)^3 \exp \left(- qV_n/kT \right) \times \\ &\times \int_{v_{0x}}^{\infty} v_x \exp \left(- m^* v_x^2 / 2kT \right) dv_x \int_{-\infty}^{\infty} \exp \left(- m^* v_y^2 / 2kT \right) dv_y \times \\ &\times \int_{-\infty}^{\infty} \exp \left(- m^* v_z^2 / 2kT \right) dv_z = \\ &= \left(\frac{4\pi q m^* k^2}{h^3} \right) T^2 \exp \left(- qV_n/kT \right) \exp \left(- \frac{m^* v_{0x}^2}{2kT} \right). \end{aligned} \quad (21)$$

Минимальная скорость (v_{0x}) в направлении x , необходимая для преодоления барьера, определяется соотношением

$$\frac{1}{2} m^* v_{0x}^2 = q(V_{bi} - V), \quad (22)$$

где V_{bi} — высота потенциального барьера при нулевом смещении (рис. 1, а). Подставляя выражение (22) в выражение (21), находим

$$\begin{aligned} J_{s \rightarrow m} &= \left(\frac{4\pi q m^* k^2}{h^3} \right) T^2 \exp \left[- \frac{q(V_n + V_{bi})}{kT} \right] \exp \left(\frac{qV}{kT} \right) = \\ &= A^* T^2 \exp \left(- \frac{q\Phi_B}{kT} \right) \exp \left(\frac{qV}{kT} \right). \end{aligned} \quad (23)$$

Здесь $\varphi_B \equiv V_n + V_{bi}$ и

$$A^* = \frac{4\pi q m^* k^2}{h^3} \quad (24)$$

— эффективная постоянная Ричардсона для термоэлектронной эмиссии в пренебрежении рассеянием на оптических фонах и квантовомеханическим отражением (разд. 5.4.3). Для свободных электронов постоянная Ричардсона A равна $120 \text{ A} \cdot \text{см}^{-2} \text{ K}^{-2}$. Если учесть понижение потенциального барьера за счет силы изображения, то в уравнении (23) высоту барьера φ_B нужно уменьшить на $\Delta\varphi$.

Для полупроводников с изотропной эффективной массой в нижнем минимуме зоны проводимости, таких, как GaAs n -типа, $A^*/A = m^*/m_0$, где m^* и m_0 — соответственно эффективная масса и масса свободного электрона. Для многодолинных полупроводников соответствующая постоянная Ричардсона A^* для каждого энергетического минимума задается выражением [13]

$$\frac{A_1^*}{A} = \frac{1}{m_0} (l_1^2 m_y^* m_z^* + l_2^2 m_z^* m_x^* + l_3^2 m_x^* m_y^*)^{1/2}, \quad (25)$$

где l_1, l_2 и l_3 — косинусы углов между нормалью к плоскости контакта и главными осями эллипсоидов, m_x^*, m_y^* и m_z^* — компоненты тензора эффективной массы. Минимумы зоны проводимости германия расположены на краю зоны Бриллюэна в направлении $\langle 111 \rangle$. Эти минимумы эквивалентны четырем эллипсоидам с продольной массой $m_l^* = 1,64m_0$ и поперечной $m_t^* = 0,082 m_0$. Сумма всех A_1^* минимальна в направлении $\langle 111 \rangle$:

$$\left(\frac{A^*}{A} \right)_{n\text{-Ge}\langle 111 \rangle} = m_t^*/m_0 \cdot [(m_l^*)^2 + 8m_t^* m_l^*]^{1/2} / m_0 = 1,11. \quad (26)$$

Максимальным оказывается значение A^* в направлении $\langle 100 \rangle$:

$$\left(\frac{A^*}{A} \right)_{n\text{-Ge}\langle 100 \rangle} = \frac{4}{m_0} \left[\frac{(m_t^*)^2 + 2m_l^* m_t^*}{3} \right]^{1/2} = 1,19. \quad (27)$$

Минимумы зоны проводимости кремния расположены в направлении $\langle 100 \rangle$. При этом $m_t^* = 0,98m_0$, а $m_l^* = 0,19m_0$. Все минимумы дают одинаковый вклад в ток в направлении $\langle 111 \rangle$, в котором достигается максимальное значение A^* :

$$\left(\frac{A^*}{A} \right)_{n\text{-Si}\langle 111 \rangle} = \frac{6}{m_0} \left[\frac{(m_t^*)^2 + 2m_l^* m_t^*}{3} \right]^{1/2} = 2,2. \quad (28)$$

Минимальное значение A^* достигается здесь в направлении $\langle 100 \rangle$:

$$\left(\frac{A^*}{A} \right)_{n\text{-Si}\langle 100 \rangle} = 2m_t^*/m_0 + 4(m_l^* m_t^*)^{1/2}/m_0 = 2,1. \quad (29)$$

Валентные зоны Ge, Si и GaAs имеют два энергетических минимума при $k = 0$. Поэтому ток, создаваемый легкими и тяжелыми дырками, практически не зависит от направления. Суммируя вклад этих двух типов носителей, получаем

$$\left(\frac{A^*}{A}\right)_{p\text{-типа}} = (m_{ih}^* + m_{hh}^*)/m_0. \quad (30)$$

Значения величин A^*/A для этих полупроводников приведены в табл. 1 [13].

Поскольку высота барьера для электронов,двигающихся из металла в полупроводник, остается неизменной, величина соответствующего тока не зависит от приложенного напряжения. В термодинамическом равновесии (т. е. при $V = 0$) этот ток равен току, текущему из полупроводника в металл. Соответствующая плотность тока получается из уравнения (23) подстановкой $V = 0$:

$$J_{m \rightarrow s} = -A^*T^2 \exp\left(-\frac{q\Phi_{Bn}}{kT}\right). \quad (31)$$

Сумма выражений (23) и (31) представляет собой плотность полного тока:

$$\begin{aligned} J_n &= \left[A^*T^2 \exp\left(-\frac{q\Phi_{Bn}}{kT}\right) \right] \left[\exp\left(\frac{qV}{kT}\right) - 1 \right] = \\ &= J_{ST} \left[\exp\left(\frac{qV}{kT}\right) - 1 \right], \end{aligned} \quad (32)$$

где

$$J_{ST} \equiv A^*T^2 \exp\left(-\frac{q\Phi_{Bn}}{kT}\right). \quad (33)$$

Уравнение (32) имеет тот же вид, что и уравнение для плотности тока в p - n -переходе, однако выражения для плотности тока насыщения различаются.

Таблица 1. Значения A^*/A

Полупроводник	Ge	Si	GaAs
p -типа	0,34	0,66	0,62
n -типа $\langle 111 \rangle$	1,11	2,2	0,068 (слабые поля), 1,2 (сильные поля)
n -типа $\langle 100 \rangle$	1,19	2,1	0,068 (слабые поля), 1,2 (сильные поля)

5.4.2. Диффузионная теория

Диффузионная теория Шоттки [4] основана на следующих предположениях: 1) высота барьера много больше kT ; 2) рассеяние электронов при их движении в обедненном слое играет существенную роль; 3) концентрация носителей при $x = 0$ и $x = W$ не зависит от тока (т. е. она совпадает со своим равновесным значением); 4) концентрация примесей в полупроводнике достаточно мала, и вырождение отсутствует.

В данном случае приходится учитывать обе компоненты электрического тока (диффузионную и полевою):

$$J_x = J_n = q \left[n(x) \mu \mathcal{E} + D_n \frac{\partial n}{\partial x} \right] = q D_n \left[- \frac{q n(x)}{kT} \frac{\partial V(x)}{\partial x} + \frac{\partial n}{\partial x} \right]. \quad (34)$$

В стационарном случае плотность тока не зависит от x и уравнение (34) можно проинтегрировать, предварительно умножив его на $\exp [-qV(x)/kT]$. В результате получим

$$J_n \int_0^W \exp \left[- \frac{qV(x)}{kT} \right] dx = q D_n \left\{ n(x) \exp \left[- \frac{qV(x)}{kT} \right] \right\}_0^W \quad (35)$$

с граничными условиями

$$\begin{aligned} qV(0) &= -q(V_n + V_{bi}) = -q\Phi_{Bn}, \\ qV(W) &= -qV_n - qV, \\ n(0) &= N_c \exp \left[- \frac{E_c(0) - E_F}{kT} \right] = N_c \exp \left(- \frac{q\Phi_{Bn}}{kT} \right), \\ n(W) &= n = N_c \exp \left(- \frac{qV_n}{kT} \right). \end{aligned} \quad (36)$$

Подставляя уравнение (36) в выражение (35), находим

$$J_n = q N_c D_n \left[\exp \left(\frac{qV}{kT} \right) - 1 \right] / \int_0^W \exp \left[- \frac{qV(x)}{kT} \right] dx. \quad (37)$$

Распределение потенциала в барьере Шоттки в пренебрежении силами изображения определяется уравнением (6):

$$qV(x) = \frac{q^2 N_D}{\epsilon_s} \left(Wx - \frac{x^2}{2} \right) - q\Phi_{Bn}.$$

Подставив его в уравнение (37) и выразив W через $V_{bi} + V$, получим окончательно

$$J_n \cong \left\{ \frac{q^2 D_n N_D}{kT} \left[\frac{q(V_{bi} - V) 2N_D}{\epsilon_s} \right]^{1/2} \times \right. \\ \left. \times \exp \left(-\frac{q\Phi_{Bn}}{kT} \right) \right\} \left[\exp \left(\frac{qV}{kT} \right) - 1 \right] = J_{SD} \left[\exp \left(\frac{qV}{kT} \right) - 1 \right]. \quad (38)$$

Выражения для плотности тока в теории термоэлектронной эмиссии и в диффузионной теории (уравнения (32) и (38)) в основном очень похожи, однако в диффузионной теории «плотность тока насыщения» J_{SD} сильнее зависит от напряжения и менее чувствительна к температуре, чем J_{SI} в теории термоэлектронной эмиссии.

Б.4.3. Термоэмиссионно-диффузионная теория

Обобщение двух описанных выше теорий было сделано Кроуэлом и Зи [14]. В этой теории в качестве граничного условия используется скорость термоэлектронной рекомбинации v_R на поверхности раздела металл — полупроводник.

Поскольку диффузия носителей существенно зависит от распределения потенциала, мы будем сразу учитывать понижение потенциального барьера за счет эффекта Шоттки (рис. 8). Рассмотрим случай, когда высота барьера настолько велика, что плотность заряда в области между поверхностью металла и $x = W$ полностью определяется концентрацией заряженных доноров

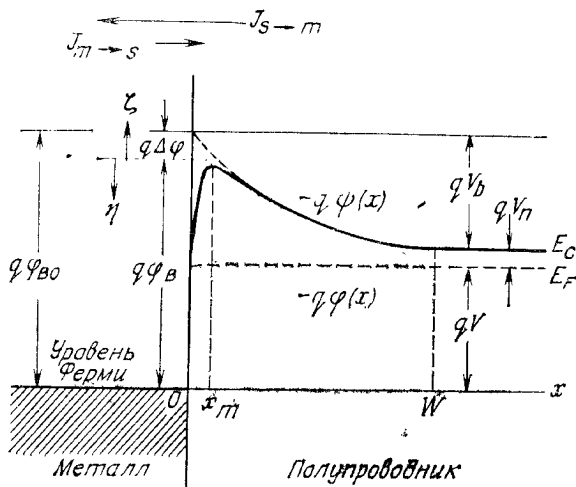


Рис. 8. Энергетическая диаграмма контакта с учетом эффекта Шоттки. $q\phi(x)$ — потенциальная энергия электрона, $q\phi(x)$ — положение квазиуровня Ферми.

(т. е. W является границей обедненного слоя). Как уже говорилось выше, максимум $q\psi$ вблизи границы раздела металл—полупроводник обусловлен наложением двух электрических полей: поля ионизированных доноров (штриховая линия) и поля силы изображения, притягивающего электрон к металлу. Как показано на рис. 8, приложенное напряжение V создает поток электронов из полупроводника в металл. На этом рисунке показан также изменяющийся в области барьера квазиуровень Ферми, соответствующий полному току J :

$$J = -q\mu n \frac{d\varphi_n}{dx}. \quad (39)$$

Здесь n — плотность электронов в точке x :

$$n = N_C e^{-q(\varphi_n - \psi)/kT}, \quad (40)$$

где N_C — эффективная плотность состояний в зоне проводимости, T — температура электронов. Будем предполагать, что область между x_m и W изотермична и, кроме того, температура электронов равна температуре решетки.

Поскольку область барьера между x_m и поверхностью раздела ($x = 0$) действует как сток для электронов, ток носителей можно описать выражением, воспользовавшись скоростью эффективной поверхностной рекомбинации v_R в максимуме потенциальной энергии:

$$J = q(n_m - n_0)v_R, \quad (41)$$

где n_m — плотность электронов в точке x_m при протекании тока, n_0 — концентрация электронов, которая была бы в равновесии ($V = 0$) при той же высоте и том же положении максимума потенциала $\psi(x)$. Как φ , так и ψ удобно измерять относительно уровня Ферми в металле. При этом

$$\varphi(W) = -V, \quad n_0 = N_C e^{-q\varphi_{Bn}/kT},$$

$$\text{и } n_m = N_C \exp\left[\frac{-q\varphi(x_m) - q\varphi_{Bn}}{kT}\right], \quad (42)$$

где $q\varphi_{Bn}$ — высота барьера, $q\varphi(x_m)$ — расстояние до квазиуровня Ферми в точке x_m .

Подставив выражение (40) в выражение (39) и интегрируя от x_m до W , получим

$$\exp\left[\frac{q\varphi(x_m)}{kT}\right] - \exp\left(\frac{qV}{kT}\right) = -\frac{J}{\mu N_C kT} \int_{x_m}^W \exp\left(\frac{-q\psi}{kT}\right) dx. \quad (43)$$

Используя затем выражения (41) — (43), найдем

$$J = \frac{qN_c v_R}{1 + v_R/v_D} \exp\left(-\frac{q\varphi_{Bn}}{kT}\right) \left[\exp\left(\frac{-qV}{kT}\right) - 1 \right], \quad (44)$$

где

$$v_D \equiv \left[\int_{x_m}^W \frac{q}{\mu kT} \exp\left[-\frac{q}{kT}(\varphi_{Bn} + \psi)\right] dx \right]^{-1} \quad (45)$$

— эффективная скорость диффузии, описывающая перенос электронов от границы обедненного слоя (точки W) к точке максимума потенциальной энергии. Если предположить, что функция распределения электронов при $x \geq x_m$ максвелловская, и если никакие электроны, кроме тех, которые связаны с плотностью тока $qn_0 v_R$, не возвращаются из металла, то полупроводник ведет себя как термоэлектронный эмиттер. При этом v_R оказывается равной средней тепловой скорости:

$$v_R = \int_0^{\infty} v_x \exp(-m^* v_x^2 / 2kT) dv_x \left/ \int_{-\infty}^{\infty} \exp(-m^* v_x^2 / 2kT) dv_x \right. = \\ = (kT/2m^* \pi)^{1/2} = A^* T^2 / qN_c, \quad (46)$$

где A^* — эффективная постоянная Ричардсона, значения которой приведены в табл. 1. При 300 К величина v_R равна $7,0 \cdot 10^6$, $5,2 \cdot 10^6$ и $1,0 \cdot 10^7$ см/с для ориентированного в направлении $\langle 111 \rangle$ Ge n -типа, $\langle 111 \rangle$ Si n -типа и GaAs n -типа соответственно. Если $v_D \gg v_R$, то в экспоненциальном члене уравнения (44) остается только v_R и справедлива теория термоэлектронной эмиссии. И наоборот, при $v_D \ll v_R$ преобладающим является процесс диффузии. Если пренебречь влиянием сил изображения и не учитывать зависимость подвижности электронов от электрического поля, скорость v_D будет равна $\mu \mathcal{E}$, где \mathcal{E} — электрическое поле в полупроводнике вблизи поверхности. При этом получается результат, который следует также из обычной диффузионной теории (уравнение (37)):

$$J \simeq qN_c \mu \mathcal{E} \exp\left(-\frac{q\varphi_{Bn}}{kT}\right) \left[\exp\left(\frac{qV}{kT}\right) - 1 \right]. \quad (47)$$

Для учета влияния сил изображения при вычислении v_D нужно в выражение (45) подставить соответствующее выражение для ψ :

$$\psi = \varphi_{Bn} + \Delta\varphi - \mathcal{E}x - \frac{q}{16\pi\epsilon_0 x}, \quad (48)$$

где $\Delta\varphi$ — понижение потенциального барьера, определяемое из формулы (15а) (в предположении, что электрическое поле посто-

но при $x < x_m$). Подставляя выражение (48) в (45), получаем, что $v_D \approx \mu \mathcal{E}$ при $\Delta\varphi < kT/q$, а по мере увеличения $\Delta\varphi$ до $20kT/q$ происходит уменьшение v_D до $0,3\mu\mathcal{E}$.

Результат, выражаемый уравнением (44), является объединением диффузионной теории Шоттки и теории термоэлектронной эмиссии Бете. Из этого уравнения видно, что ток хорошо описывается в рамках теории термоэлектронной эмиссии при $\mu\mathcal{E}(x_m) > v_R$. Последний критерий является более точным, чем условие Бете $\mathcal{E}(x_m) > kT/q\lambda$, где λ — длина свободного пробега.

Выше скорость поверхностной рекомбинации v_R , входящая в выражение для тока термоэлектронной эмиссии, была введена как граничное условие, учитывающее только коллекторные свойства металла в барьере Шоттки. Однако во многих случаях существует достаточно большая вероятность того, что электрон, проходящий над барьером, отразится обратно вследствие рассеяния на оптическом фоне [15, 16]. Вероятность преодоления электроном барьера в первом приближении можно представить в виде $f_p = \exp(-x_m/\lambda)$. Кроме того, функция распределения электронов по энергии может сильно отличаться от максвелловской как вследствие квантовомеханического отражения электронов от барьера Шоттки, так и вследствие туннелирования через барьер [17, 18]. Отношение f_Q полного тока к току, который имел бы место в пренебрежении квантовомеханическим туннелированием и отражением, сильно зависит от электрического поля и энергии электрона относительно максимума потенциала.

Окончательное выражение для вольт-амперной характеристики, учитывающее f_p и f_Q , имеет вид

$$J = J_S(e^{qV/kT} - 1), \quad (49)$$

$$J_S = A^{**}T^2 \exp\left(-\frac{q\varphi_B n}{kT}\right), \quad (50)$$

где

$$A^{**} = \frac{f_p f_Q A^*}{1 + f_p f_Q v_R / v_D}. \quad (50a)$$

На рис. 9 представлены результаты расчета эффективной постоянной Ричардсона A^{**} для системы металл — кремний при комнатной температуре и концентрации примеси 10^{16} см^{-3} [19]. Видно, что величина A^{**} для электронов в кремнии n -типа в интервале полей $10^4 - 2 \cdot 10^5 \text{ В/см}$ практически постоянна и равна приблизительно $110 \text{ А} \cdot \text{см}^{-2} \cdot \text{К}^{-2}$. Для дырок в кремнии p -типа A^{**} тоже постоянна в этом интервале полей, но ее значение существенно меньше ($\sim 30 \text{ А} \cdot \text{см}^{-2} \cdot \text{К}^{-2}$).

На основании изложенного выше можно сделать вывод, что при комнатной температуре в интервале полей $10^4 - 10^5 \text{ В/см}$ перенос заряда в большинстве диодов Шоттки на Ge, Si и GaAs

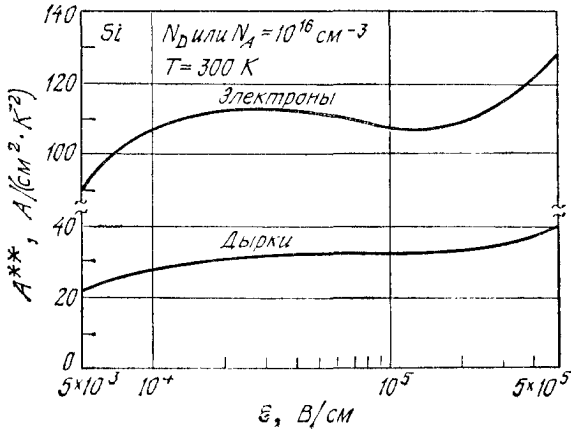


Рис. 9. Расчетные значения эффективной постоянной Ричардсона как функции электрического поля в барьере металл — кремний [19].

осуществляется путем термоэлектронной эмиссии основных носителей.

Пространственное изменение квазиуровня Ферми в области барьера можно оценить, подставив выражения (6) и (40) в выражение (39). Оказывается [20], что это изменение весьма незначительно (рис. 8). В диодах Au-Si с $N_D = 1,2 \cdot 10^{15} \text{ см}^{-3}$ при температуре 300 К и прямом смещении 0,2 В разность $\phi(W) - \phi(0)$ составляет всего 8 мВ, а при более высоком уровне легирования она еще меньше. Эти результаты еще раз подтверждают, что в полупроводниках с высокой подвижностью и умеренной концентрацией примеси применима теория термоэлектронной эмиссии.

5.4.4. Туннельный ток

В приборах на сильнолегированных полупроводниках, а также при низких температурах преобладающим процессом переноса заряда оказывается туннелирование. В этом случае к $J_{s \rightarrow m}$ (выражение (16)), описывающему термоэлектронную эмиссию, нужно добавить туннельную компоненту тока, которая пропорциональна квантовому коэффициенту прохождения, умноженному на вероятность, что соответствующее состояние в полупроводнике заполнено, а в металле свободно, т. е.

$$J_{s \rightarrow m} = \frac{A^* T}{k} \int_0^{\infty} T(\xi) \exp \left[\frac{-q(V_b + V_n + \xi - \Delta\phi)}{kT} \right] d\xi + \frac{A^* T}{k} \int_0^{q(V_b - \Delta\phi)} F_s(V) T(\eta) (1 - F_m) d\eta, \quad (51)$$

где энергии $q\xi$ и $q\eta$ отсчитываются соответственно вверх и вниз от максимума потенциального барьера (рис. 8). Первое слагаемое в выражении (51) соответствует термоэлектронной компоненте тока и переходит в выражение (23) при $T(\xi) = 1$. Второе слагаемое соответствует туннельной компоненте. Величины F_s и F_m — функции распределения Ферми—Дирака соответственно в полупроводнике и в металле, а $T(\xi)$ и $T(\eta)$ — коэффициенты прозрачности барьера соответственно выше и ниже максимума потенциала.

Аналогичное выражение можно записать для тока $J_{m \rightarrow s}$, текущего из металла в полупроводник:

$$J_{m \rightarrow s} = -\frac{A^*T}{k} \exp\left(\frac{-q\Phi_{Bn}}{kT}\right) \int_0^{\infty} T(\xi) \exp\left(-\frac{\xi}{kT}\right) d\xi - \frac{A^*T}{k} \int_0^{q(V_b - \Delta\Phi)} F_m T(\eta) (1 - F_s) d\eta. \quad (52)$$

Полная плотность тока описывается алгебраической суммой выражений (51) и (52).

Теоретические и экспериментальные вольт-амперные характеристики барьеров Au—Si приведены на рис. 10 [18]. Отметим, что плотность тока J можно записать в виде

$$J = J_s [\exp(qV/nkT) - 1], \quad (53)$$

или

$$J \simeq J_s \exp(qV/nkT) \text{ при } V \gg kT/q, \quad (53a)$$

где J_s — плотность тока насыщения, получаемая экстраполяцией линейного в логарифмическом масштабе участка зависимости $J(V)$ к точке $kV = 0$, а n — фактор неидеальности, определяемый выражением

$$n \equiv \frac{q}{kT} \frac{\partial V}{\partial (\ln J)}. \quad (53b)$$

Зависимости плотности тока насыщения и фактора n от концентрации легирующей примеси в диодах Au—Si при различных температурах приведены на рис. 11 [18]. Интересно отметить, что J_s почти постоянна при низких уровнях легирования, но при $N_D \gg 10^{17} \text{ см}^{-3}$ быстро возрастает. Фактор неидеальности n при низком уровне легирования и сравнительно высоких температурах близок к 1. Однако по мере повышения уровня легирования либо понижения температуры отличие n от единицы становится существенным.

Плотность туннельной компоненты тока, преобладающей при высоком уровне легирования и низких температурах, определяется простым выражением

$$J_t \sim \exp(-q\Phi_{Bn}/E_{00}), \quad (54)$$

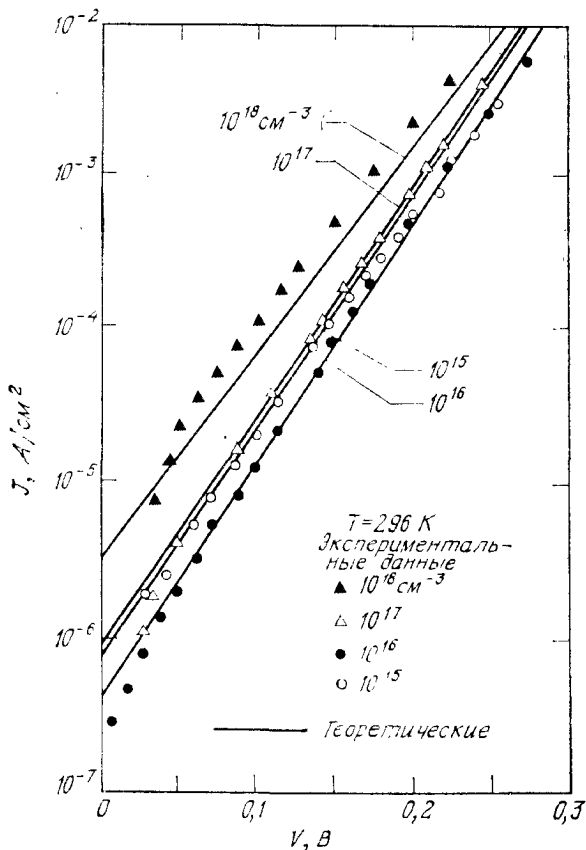


Рис. 10. Теоретические и экспериментальные вольт-амперные характеристики диодов Au — Si [18].

поскольку коэффициент прозрачности барьера пропорционален:

$$T(\eta) \sim \exp(-q\varphi_{Bn}/E_{00}), \quad (54a)$$

где

$$E_{00} \equiv \frac{q\hbar}{2} \sqrt{\frac{N_D}{\epsilon_8 m^*}}.$$

Из выражения (54) видно, что туннельный ток экспоненциально зависит от $\sqrt{N_D}$.

На рис. 12 показано отношение туннельного тока к току термоэлектронной эмиссии в диодах Au—Si [18]. Видно, что при $N_D \leq 10^{17} cm^{-3}$ и $T \gtrsim 300 K$ это отношение много меньше 1 и туннель-

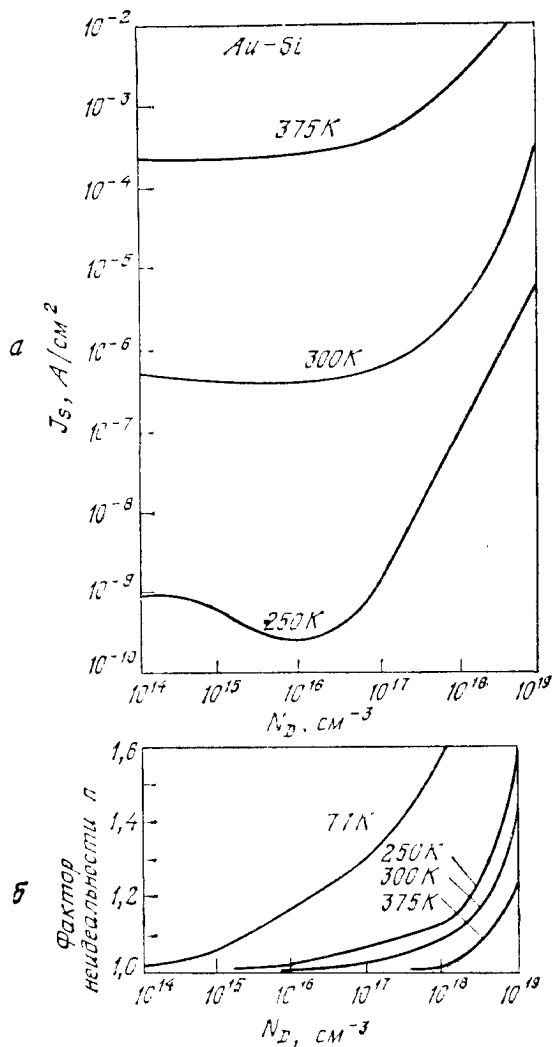


Рис. 11. Зависимости плотности тока насыщения (а) и фактора неидеальности n (б) от концентрации легирующей примеси в диоде Au — Si при разных температурах [18].

ной компонентой можно пренебречь. Однако при высоких уровнях легирования и при низких температурах оно становится много больше 1. Это указывает на то, что в последнем случае преобладающим является туннельный ток.

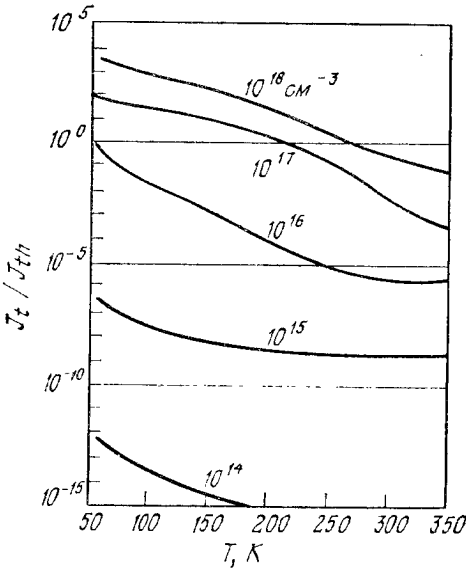


Рис. 12. Отношение туннельного тока к току термоэлектронной эмиссии в диодах Au — Si [18].

5.4.5. Инжекция неосновных носителей

При низком уровне инжекции перенос заряда в диодах Шоттки осуществляется главным образом основными носителями. Однако при достаточно больших прямых смещениях коэффициент инжекции неосновных носителей γ (т. е. отношение тока неосновных носителей к полному току) увеличивается с ростом тока. Последнее объясняется увеличением дрейфовой компоненты тока, которая становится много больше диффузионной.

В стационарном состоянии одномерные уравнения непрерывности и плотности тока для неосновных носителей записываются в виде

$$0 = -\frac{p_n - p_{n0}}{\tau_p} - \frac{1}{q} \frac{\partial J_p}{\partial x}, \quad (55)$$

$$J_p = q\mu_p p_n \mathcal{E} - qD_p \frac{\partial p_n}{\partial x}. \quad (56)$$

Рассмотрим энергетическую диаграмму, показанную на рис. 13. Здесь x_1 — граница обедненной области и x_2 — граница раздела между эпитаксиальным слоем n -типа и подложкой n^+ -типа. Из теории выпрямления (гл. 2) следует, что плотность неосновных носителей в точке x_1 равна

$$p_n(x_1) = p_{n0} \left(\exp\left(\frac{qV}{kT}\right) - 1 \right) \approx \frac{n_i^2}{N_D} \left[\exp\left(\frac{qV}{kT}\right) - 1 \right], \quad (57)$$

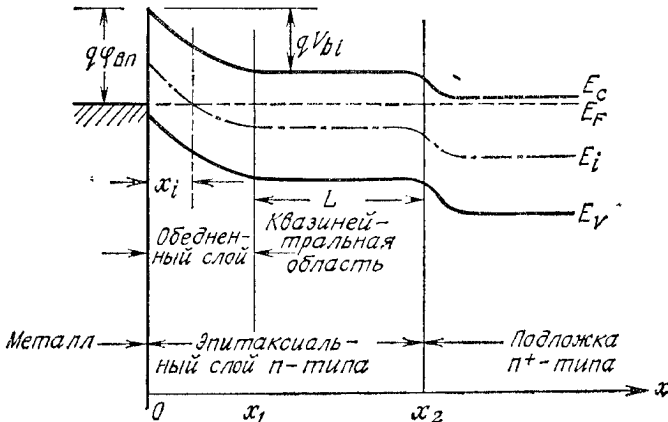


Рис. 13. Энергетическая диаграмма эпитаксиального барьера Шоттки [23].

где N_D — концентрация доноров в p -слое. Используя уравнения (49) и (57), можно выразить $p_n(x_1)$ через плотность тока

$$p_n(x_1) = \frac{n_i^2}{N_D} \frac{J}{J_s} \quad (58)$$

Граничное условие для $p_n(x)$ в точке $x = x_2$ можно получить, если учесть, что скорость переноса неосновных носителей v_T равна D_p/L_p . При этом

$$J_p(x_2) = qv_T p_n = q \left(\frac{D_p}{L_p} \right) p_{n0} \left[\exp \left(\frac{qV}{kT} \right) - 1 \right] \quad (59)$$

в случае $L \ll L_p$,

где D_p — коэффициент диффузии неосновных носителей, L_p — их диффузионная длина, L — длина квазинейтральной области.

При низком уровне инжекции в уравнении (56) можно пренебречь дрейфовым членом по сравнению с диффузионным. В результате для коэффициента γ получаем

$$\gamma \equiv \frac{J_p}{J_p + J_n} \simeq \frac{J_p}{J_n} = \frac{q n_i^2 D_p}{N_D L_p A^{**} T^2 \exp(-q\phi_{Bn}/kT)} \quad (60)$$

Экспериментально было обнаружено [22], что для диодов Au—Si с $N_D = 10^{16} \text{ см}^{-3}$ коэффициент γ очень мал ($\sim 5 \cdot 10^{-5}$). Последнее находится в хорошем соответствии с результатом, получаемым из выражения (60).

По мере увеличения положительно смещающего напряжения в квазинейтральной области растет поле \mathcal{E} . Это приводит к тому,

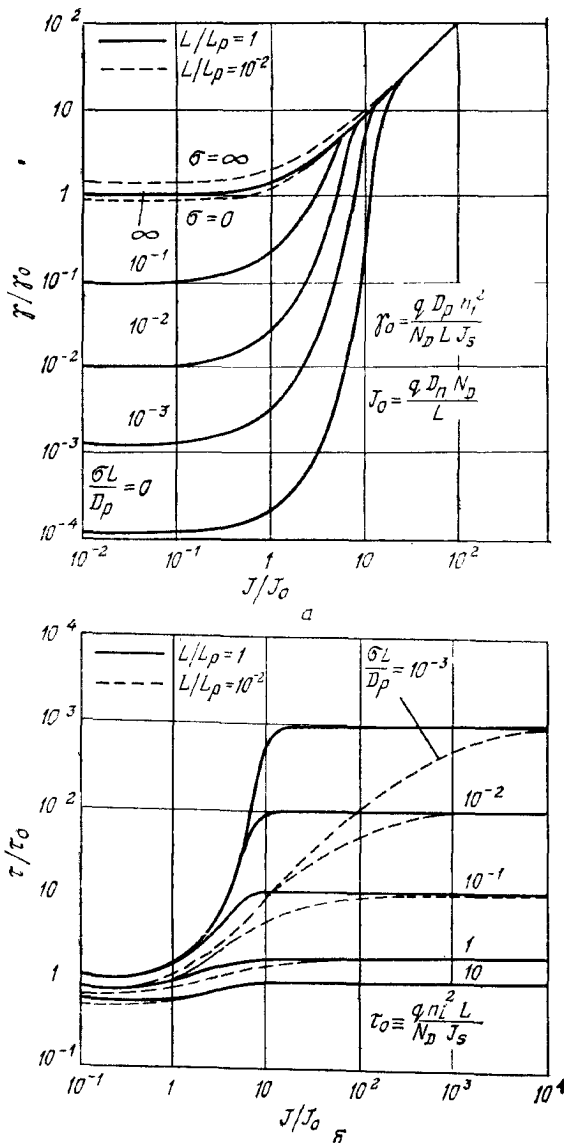


Рис. 14. Зависимость нормированного отношения плотности тока неосновных носителей к плотности полного тока (а) и нормированного времени накопления неосновных носителей (б) от нормированной плотности полного тока [23].

что в токе неосновных носителей преобладающей становится дрейфовая компонента. Тогда из уравнений (56) и (58) получаем

$$\gamma \sim q\mu_p p_n \mathcal{E} \sim \frac{J}{J_s}, \quad (61)$$

т. е. γ растет пропорционально плотности тока. Например, для диода Au—Si с $N_D = 10^{15} \text{ см}^{-3}$ и $J_S = 5 \cdot 10^{-7} \text{ А/см}^2$ при плотности тока инжекции 350 А/см^2 коэффициент γ составляет $\sim 5\%$. Промежуточный случай рассмотрен в работе [23]. Результаты расчетов представлены на рис. 14, а, где γ и J нормированы на

$$J_0 \equiv \frac{qD_n N_D}{L}, \quad (62a)$$

$$\gamma_0 \equiv \frac{qD_p n_i^2}{N_D L J_S}. \quad (62b)$$

Из рис. 14, а видно, что для уменьшения коэффициента инжекции неосновных носителей следует брать образцы с большой концентрацией N_D (соответствующей низкому сопротивлению материала), большим J_S (соответствующим малой высоте барьера) и малой концентрацией n_i (соответствующей большой ширине запрещенной зоны).

Другой величиной, связанной с коэффициентом γ , является время накопления неосновных носителей τ_s . Оно определяется как отношение заряда неосновных носителей в квазинейтральной области к плотности тока:

$$\tau_s \equiv \frac{\int_{x_1}^{x_2} qp(x) dx}{J}. \quad (63)$$

В пределе больших токов

$$\tau_s \simeq \frac{qn_i^2 L_p}{N_D J_S}. \quad (64)$$

Зависимость τ_s от плотности тока показана на рис. 14, б, где использована та же нормировка, что и на рис. 14, а. Например, для диода Au—Si с $N_D = 1,5 \cdot 10^{14} \text{ см}^{-3}$, $L = 7 \text{ мкм}$ и $D_p/L_p = 2000 \text{ см/с}$ время накопления при $J = 10 \text{ А/см}^2$ составляет $\sim 1 \text{ нс}$. При увеличении N_D до $1,5 \cdot 10^{16} \text{ см}^{-3}$ это время уменьшается до $0,01 \text{ нс}$.

5.5. ВЫСОТА БАРЬЕРА

5.5.1. Общее выражение для высоты барьера

Основные зонные диаграммы контактов металл — полупроводник мы рассмотрели в разд. 5.2. В общем случае высота барьера на таком контакте зависит от работы выхода металла и плотности

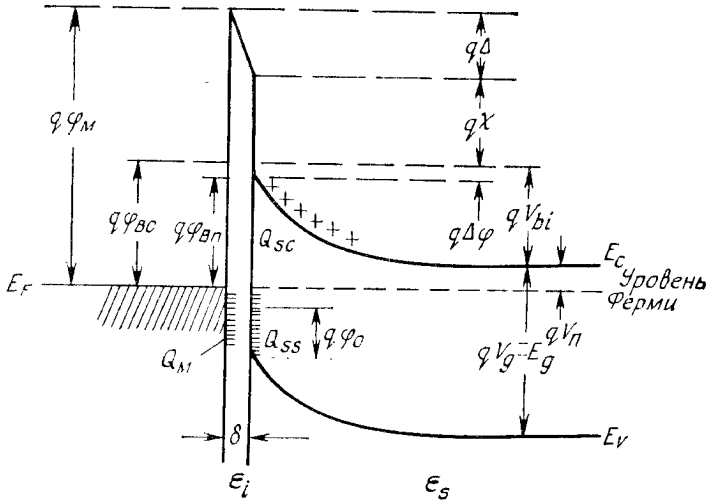


Рис. 15. Подробная энергетическая диаграмма контакта металл — полупроводник n -типа при наличии промежуточного слоя толщиной порядка межатомных расстояний [24].

φ_M — работа выхода металла; φ_{Bn} — высота энергетического барьера; φ_{B0} — асимптотическое значение φ_{Bn} при нулевом электрическом поле; φ_n — энергетический уровень на поверхности; $\Delta\varphi$ — понижение барьера за счет сил изображения; Δ — падение потенциала на промежуточном слое; χ — электронное сродство полупроводника; V_{bi} — встроенный потенциал; ϵ_s — диэлектрическая проницаемость полупроводника; ϵ_i — диэлектрическая проницаемость промежуточного слоя; δ — толщина промежуточного слоя; Q_{sc} — плотность объемного заряда в полупроводнике; Q_{ss} — плотность заряда поверхностных состояний полупроводника; Q_M — плотность поверхностного заряда в металле.

поверхностных состояний. Общее выражение для высоты потенциального барьера [24] может быть получено на основе двух предположений: 1) толщина промежуточного слоя, возникающего между поверхностями полупроводника и металла, либо равна нулю, либо порядка межатомных размеров и поэтому является туннельно прозрачной для электронов, а влияние этого слоя сводится лишь к падению потенциала на нем; 2) энергетическая плотность поверхностных состояний не зависит от типа металла и определяется только свойствами поверхности полупроводника.

Более подробная энергетическая диаграмма контакта металл — полупроводник n -типа — обозначения различных физических величин, которые будут использованы в дальнейшем, показаны на рис. 15. Энергетический уровень $q\varphi_0$ обозначает положение, которое занимал бы уровень Ферми в силу условия электронейтральности на свободной поверхности полупроводника (в отсутствие металлического контакта) [25]. Другой важной величиной является высота барьера $q\varphi_{Bn}$, который преодолевает электрон,

двигающийся из металла в полупроводник. При этом предполагается, что промежуточный слой имеет толщину всего несколько ангстрем и поэтому туннельно прозрачен для электронов.

Будем считать поверхностные состояния полупроводника акцепторными, а их энергетическую плотность D_s [см⁻²·эВ⁻¹] в интервале энергий от $q\varphi_0$ до уровня Ферми постоянной. Тогда плотность заряда на этих поверхностных состояниях описывается выражением

$$Q_{ss} = -qD_s(E_g - q\varphi_0 - q\varphi_{Bn} - q\Delta\varphi) \quad [\text{Кл/см}^2], \quad (65)$$

где $q\Delta\varphi$ — понижение барьера за счет эффекта Шоттки. В этой формуле величина в круглых скобках представляет собой разность энергий между уровнем Ферми на поверхности и уровнем $q\varphi_0$, а величина Q_{ss} — заряд заполненных поверхностных состояний, энергия которых превышает $q\varphi_0$.

Поверхностная плотность заряда обедненного слоя полупроводника в термодинамическом равновесии определяется выражением (8). Запишем ее в виде

$$Q_{sc} = \sqrt{2qe_s N_D \left(\varphi_{Bn} - V_n + \Delta\varphi - \frac{kT}{q} \right)}. \quad (66)$$

Полная плотность заряда на единицу площади полупроводника равна сумме выражений (65) и (66). Поскольку пространственный заряд в тонком промежуточном слое обычно пренебрежимо мал, можно считать, что равный по величине и противоположный ему по знаку заряд Q_M локализован на поверхности металла:

$$Q_M = - (Q_{ss} + Q_{sc}). \quad (67)$$

В соответствии с законом Гаусса для разности потенциалов Δ между поверхностями металла и полупроводника получим

$$\Delta = -\delta \frac{Q_M}{\epsilon_i}, \quad (68)$$

где ϵ_i — диэлектрическая проницаемость промежуточного слоя, δ — его толщина. Кроме того, из зонной диаграммы на рис. 15 видно, что вследствие постоянства уровня Ферми в системе при термодинамическом равновесии для Δ можно записать соотношение

$$\Delta = \varphi_m - (\chi + \varphi_{Bn} + \Delta\varphi). \quad (69)$$

Исключив Δ из выражений (68) и (69), получим

$$\begin{aligned} (\varphi_m - \chi) - (\varphi_{Bn} + \Delta\varphi) &= \sqrt{\frac{2qe_s N_D \delta^2}{\epsilon_i^2} \left(\varphi_{Bn} + \Delta\varphi - V_n - \frac{kT}{q} \right)} - \\ &\quad - \frac{qD_s \delta}{\epsilon_i} (E_g - q\varphi_0 - q\varphi_{Bn} - q\Delta\varphi). \quad (70) \end{aligned}$$

Уравнение (70) можно теперь решить относительно φ_{Bn} . Вводя параметры

$$c_1 \equiv \frac{2q\epsilon_s N_D \delta^2}{\epsilon_i^2}, \quad c_2 \equiv \frac{\epsilon_i}{\epsilon_i + q^2 \delta D_s}, \quad (71a), (71b)$$

запишем решение уравнения (70) в виде

$$\begin{aligned} \varphi_{Bn} = & \left[c_2 (\varphi_m - \chi) + (1 - c_2) \left(\frac{E_g}{q} - \varphi_0 \right) - \Delta\varphi \right] + \\ & + \left\{ \frac{c_1^3 c_2}{2} - c_2^{3/2} \left[c_1 (\varphi_m - \chi) + (1 - c_2) \left(\frac{E_g}{q} - \varphi_0 \right) \frac{c_1}{c_2} - \right. \right. \\ & \left. \left. - \frac{c_1}{c_2} \left(V_n + \frac{kT}{q} \right) + \frac{c_2 c_1^2}{4} \right]^{1/2} \right\}. \quad (72) \end{aligned}$$

Чтобы вычислить c_1 по формуле (71a), надо оценить значения параметров δ и ϵ_i . Для сколотых в вакууме либо хорошо очищенных полупроводниковых поверхностей промежуточный слой имеет атомные размеры, т. е. δ составляет 4—5 Å. В первом приближении диэлектрическую проницаемость такого тонкого слоя можно считать равной диэлектрической проницаемости вакуума. Поскольку последняя является нижним пределом возможных значений ϵ_i , это приближение дает завышенное значение c_1 . При $\epsilon_s \approx 10\epsilon_0$, $\epsilon_i = \epsilon_0$ и $N_D < 10^{18}$ см⁻³ величина c_1 мала ($\sim 0,01$ В) и выражение в фигурных скобках в формуле (72) меньше 0,04 В. Пренебрегая этим слагаемым, приведем выражение (72) к виду

$$\varphi_{Bn} = c_2 (\varphi_m - \chi) + (1 - c_2) \left(\frac{E_g}{q} - \varphi_0 \right) - \Delta\varphi \equiv c_2 \varphi_m + c_3. \quad (73)$$

Если c_2 и c_3 можно определить экспериментально, а значение χ известно, то

$$\varphi_0 = \frac{E_g}{q} - \frac{c_2 \chi + c_3 + \Delta\varphi}{1 - c_2} \quad (74)$$

и из выражения (71b)

$$D_s = \frac{(1 - c_2) \epsilon_i}{c_2 \delta q^2}. \quad (75)$$

Подставив в выражение (75) те же значения $\delta \approx 4\text{—}5$ Å и $\epsilon_i = 1$, получим

$$D_s \approx 1,1 \times 10^{13} (1 - c_2) / c_2 \text{ [см}^{-2} \text{ эВ}^{-1} \text{]}. \quad (75a)$$

Из уравнения (73) можно непосредственно получить два рассмотренных выше предельных случая:

1. Если $D_s \rightarrow \infty$, то $c_2 \rightarrow 0$ и

$$q\varphi_{Bn} = (E_g - q\varphi_0) - q\Delta\varphi. \quad (76a)$$

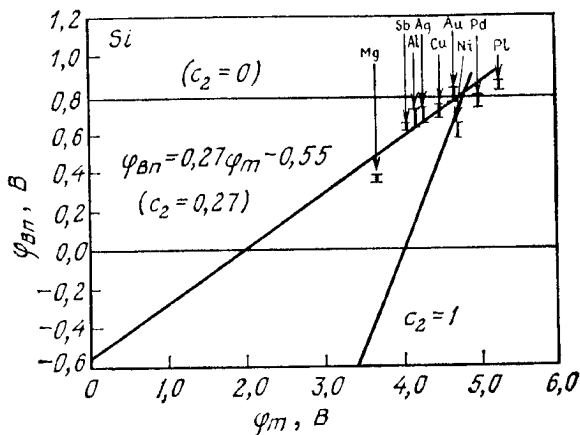


Рис. 16. Экспериментальные значения высоты энергетического барьера в системах металл — кремний n -типа [24].

В этом случае уровень Ферми на поверхности фиксируется поверхностными состояниями на энергии, превышающей край валентной зоны на величину $q\varphi_0$. При этом высота барьера не зависит от работы выхода металла и полностью определяется степенью легирования и поверхностными свойствами полупроводника.

2. Если $D_s \rightarrow 0$, то $c_2 \rightarrow 1$ и

$$q\varphi_{Вп} = q(\varphi_m - \chi) - q\Delta\varphi. \quad (76б)$$

Это выражение определяет высоту энергетического барьера идеального диода Шоттки (в отсутствие поверхностных состояний). Оно аналогично выражению (1) (за исключением слагаемого, описывающего понижение барьера за счет эффекта Шоттки).

Экспериментальные результаты, полученные для систем металл—кремний n -типа, приведены на рис. 16. Линейная аппроксимация этих данных с помощью метода наименьших квадратов приводит к результату $\varphi_{Вп} = 0,27\varphi_m - 0,55$. Сравнивая это выражение с выражением (73) и используя формулы (74) и (75а), получим $c_2 = 0,27$, $q\varphi_0 = 0,33$ эВ и $D_s = 4 \cdot 10^{13}$ см $^{-2}$ ·эВ $^{-1}$. Аналогичные результаты получены для GaAs, GaP и CdS. Они показаны на рис. 17 и приведены в табл. 2.

Отметим, что значения $q\varphi_0$ для Si, GaAs и GaP довольно близки к $1/3$ запрещенной зоны. Аналогичный результат получается также для других полупроводников [26]. Этот факт указывает на то, что большинство поверхностей ковалентных полупроводников имеет большой пик плотности поверхностных состояний с энергией на $1/3$ запрещенной зоны выше края валентной зоны.

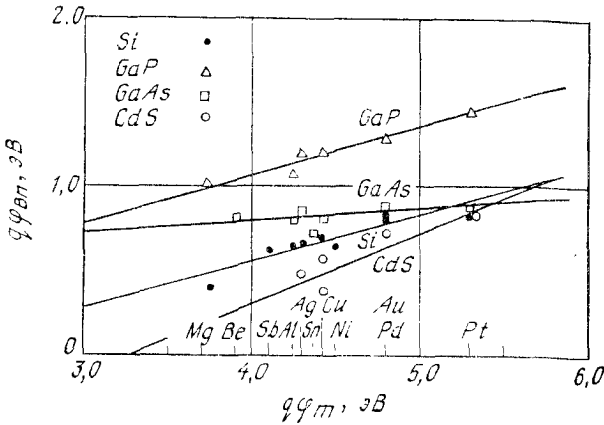


Рис. 17. Экспериментальные значения высоты энергетического барьера в системах металл — полупроводник [24].

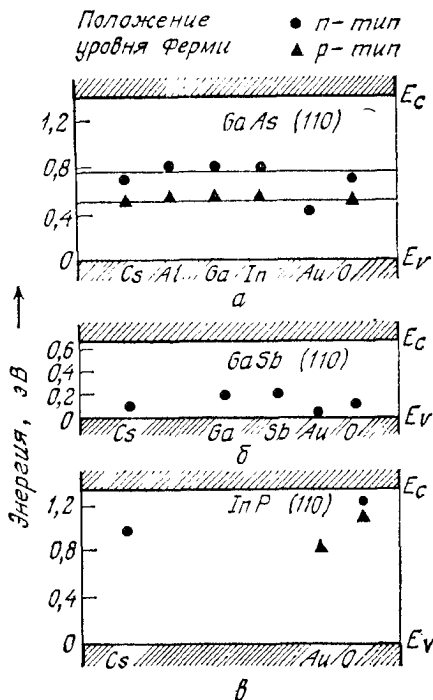
По расчетам, выполненным Пейджем [27] для поверхности алмаза с ориентацией $\langle 111 \rangle$, действительно получается узкая зона поверхностных состояний чуть ниже середины запрещенной зоны. Подобная ситуация, по-видимому, имеет место и для других полупроводников.

Экспериментальные исследования с использованием фотоэмиссионной спектроскопии показывают, что в полупроводниках типа $A^{III}B^V$ на формирование барьеров Шоттки определяющее влияние оказывают приповерхностные дефекты кристаллической структуры, образующиеся при нанесении металла [72]. На рис. 18 [73] показано положение уровня Ферми на поверхности для контактов некоторых металлов с полупроводниками GaAs, GaSb и InP. Видно, что положение уровня Ферми на поверхности этих полупроводников не зависит от типа металла. (Интересно также отметить, что на окисленных поверхностях уровень Ферми на границе с окислами локализован примерно при тех же энергиях, что и на границе с металлами.) Такая фиксация уровня

Таблица 2. Значения величин, используемых при вычислении высоты барьера в Si, GaP, GaAs и CdS

Полупроводник	c_2	c_3 , В	χ , В	$D_s \cdot 10^{-18}$, эВ $^{-1}$ /см 2	$q\phi_0$, эВ	$q\phi_0/E_g$
Si	$0,27 \pm 0,05$	$-0,55 \pm 0,22$	4,05	$2,7 \pm 0,7$	$0,30 \pm 0,36$	0,27
GaP	$0,27 \pm 0,03$	$-0,01 \pm 0,13$	4,0	$2,7 \pm 0,4$	$0,66 \pm 0,2$	0,294
GaAs	$0,07 \pm 0,05$	$+0,49 \pm 0,24$	4,07	$12,5 \pm 10,0$	$0,53 \pm 0,33$	0,38
CdS	$0,38 \pm 0,16$	$-1,20 \pm 0,77$	4,8	$1,6 \pm 1,1$	$1,5 \pm 1,5$	0,6

Рис. 18. Положение уровня Ферми на поверхности раздела полупроводников GaAs, GaSb и InP с некоторыми металлами и окислами. Отметим, что это положение слабо зависит от химической природы металлов и окисла [73].



Ферми на поверхности объясняет тот факт, что для большинства полупроводников типа $A^{III}B^V$ высота барьера слабо зависит от работы выхода металла.

В ионных полупроводниках, таких, как CdS и ZnS, высота барьера обычно сильно зависит от природы нанесенного металла, причем обнаружена связь между характеристиками поверхности раздела и электроотрицательностью. Электроотрицательность X_M определяется как энергия, необходимая для удаления электрона из атома в молекуле. На рис. 19 приведена шкала электроотрицательности, составленная Полингом [28]. Отметим, что эта шкала имеет ту же периодичность, что и работа выхода (рис. 4).

На рис. 20, а [29] показан график зависимости высоты барьера от электроотрицательности металлов, нанесенных на Si, GaSe и SiO_2 . Тангенс угла наклона графика назовем степенью влияния химической природы металла на свойства поверхности полупроводника S :

$$S \equiv \frac{d\Phi_{Bn}}{dX_M}. \quad (77)$$

На рис. 20, б показана зависимость величины S от разности электроотрицательностей компонент полупроводника. Отметим

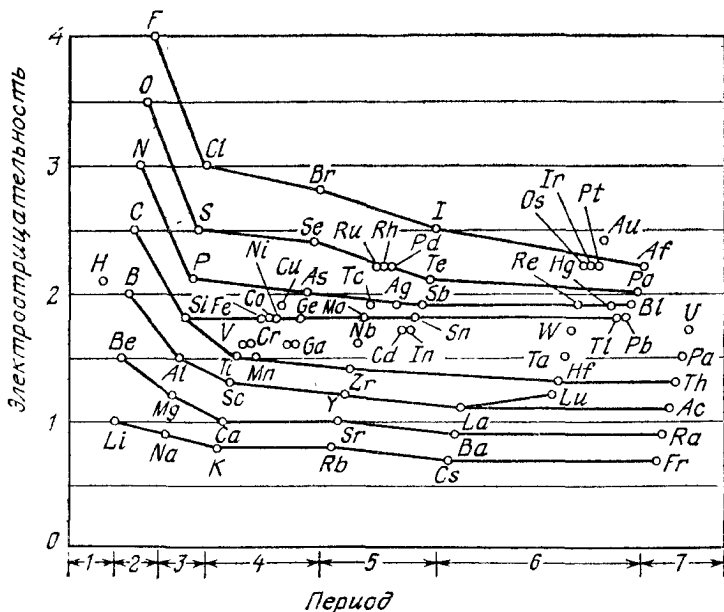


Рис. 19. Шкала электроотрицательности [28]. Отметим тенденцию к увеличению электроотрицательности внутри каждой группы.

резкий скачок при переходе от ковалентных полупроводников (таких, как GaAs с $\Delta X = 0,4$) к ионным (таким, как AlN с $\Delta X = 1,5$). В полупроводниках с $\Delta X < 1$ величина S мала, что указывает на слабую зависимость высоты барьера от электроотрицательности металла (или его работы выхода). При $\Delta X > 1$ величина S приближается к единице и высота барьера начинает зависеть от электроотрицательности металла (или его работы выхода).

В кремниевой микроэлектронной технологии широко применяются барьеры Шоттки, образующиеся в результате химической реакции металла с кремниевой подложкой. Условия протекания этих реакций хорошо известны и легко контролируемы, что позволяет получить на границе раздела силицид металла — кремний требуемые и хорошо воспроизводимые барьеры Шоттки. На рис. 21 представлены эмпирические значения высоты барьера между силицидами 12 различных переходных металлов и кремнием n -типа в зависимости от теплоты образования силицидов [71]. За исключением PtSi, эти данные можно представить в виде

$$\Phi_{\text{Вн}} = 0,81 - 0,17 (\Delta H), \quad (78)$$

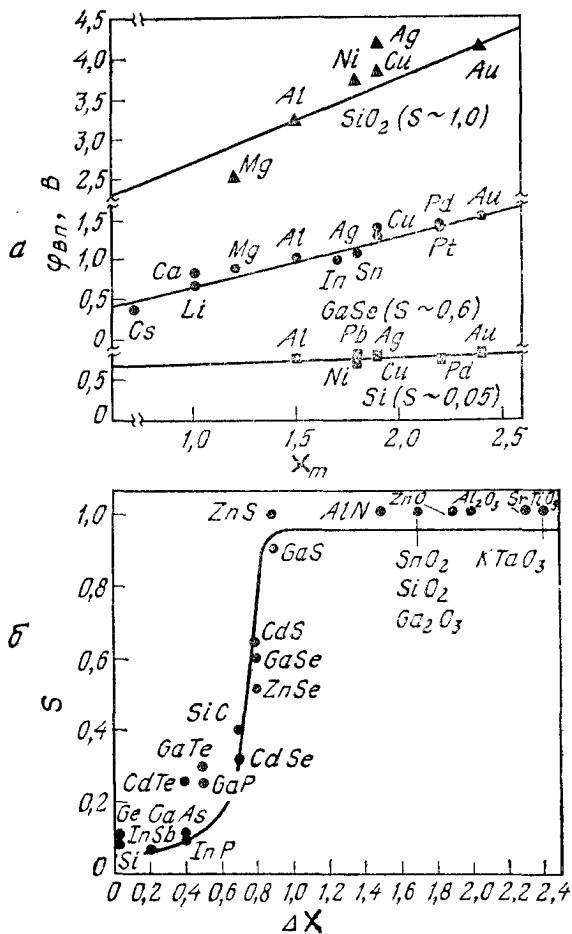


Рис. 20. Зависимость высоты барьера (а) от электроотрицательности металла, нанесенного на Si, GaSe и SiO₂, и зависимость величины S от разности электроотрицательностей компонент полупроводников (б) [29].

где ΔH — теплота образования (в электронвольтах). Значение 0,81 эВ, полученное при нулевой теплоте образования, соответствует высоте барьера на свободной поверхности. Наклон прямой, аппроксимирующей эти данные, следует из того, что высота барьера между двумя образцами кремния *n*-типа равна нулю. Последнее означает, что прямая линия на рис. 21 должна пересечь ось абсцисс в точке, соответствующей энергии связи кремния (4,67 эВ). Эта величина довольно близка к значению ΔH , получаемому из формулы (78) при $\phi_{Bn} = 0$.

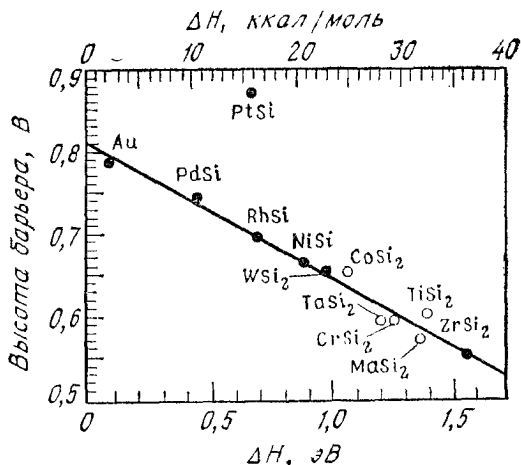


Рис. 21. Зависимость высоты барьера в структурах силицид металла — кремний от теплоты образования силицидов [71].

● — экспериментальные значения; ○ — предполагаемые значения.

5.5.2. Измерения высоты барьера

Высоту потенциального барьера на контакте металл — полупроводник измеряют в основном четырьмя методами, описанными ниже [30].

Метод вольт-амперной характеристики. Для умеренно легированных полупроводников вольт-амперная характеристика в области прямых смещений с $V \gg 3kT/q$ в соответствии с выражением (49) имеет вид

$$J = A^{**} T^2 \exp\left(-\frac{q\Phi_{B0}}{kT}\right) \exp\left[\frac{q(\Delta\Phi + V)}{kT}\right], \quad (79)$$

где Φ_{B0} — асимптотическое значение высоты барьера при нулевом поле (рис. 15), A^{**} — эффективная постоянная Ричардсона, $\Delta\Phi$ — понижение барьера за счет эффекта Шоттки. Поскольку A^{**} и $\Delta\Phi$ являются функциями приложенного напряжения, вольт-амперную характеристику при прямом смещении (и при $V \gg 3kT/q$) можно представить в виде $J \sim \exp(qV/nkT)$, где фактор неидеальности

$$n \equiv \frac{q}{kT} \frac{\partial V}{\partial (\ln J)} = \left[1 + \frac{\partial \Delta\Phi}{\partial V} + \frac{kT}{q} \frac{\partial (\ln A^{**})}{\partial V}\right]^{-1}. \quad (80)$$

Типичные примеры вольт-амперных характеристик показаны на рис. 22. Фактор неидеальности характеристики диода W—Si равен $n = 1,02$, а для диода W—GaAs $n = 1,04$ [31]. В резуль-

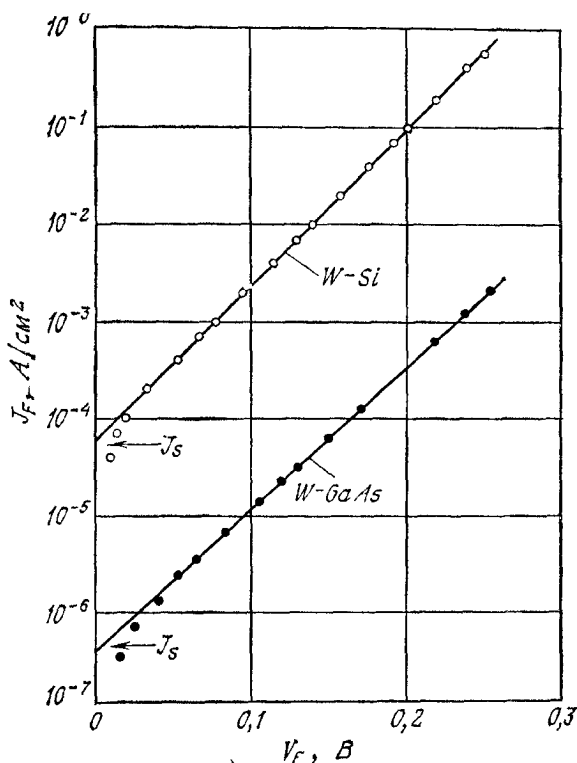


Рис. 22. Зависимости плотности тока в диодах W — Si и W — GaAs от приложенного в прямом направлении напряжения [31].

тате линейной экстраполяции этих характеристик к $V = 0$ найдем ток насыщения J_S . Высоту барьера получим из формулы

$$\varphi_{Bn} = \frac{kT}{q} \ln \left(\frac{A^{**} T^2}{J_S} \right). \quad (81)$$

Значение φ_{Bn} не очень чувствительно к выбору A^{**} ; так, например, при комнатной температуре увеличение A^{**} на 100 % приводит к увеличению φ_{Bn} только на 0,018 В. Теоретическое соотношение между J_S и φ_{Bn} (или φ_{Bp}) при комнатной температуре и $A^{**} = 120 \text{ А} \cdot \text{см}^{-2} \cdot \text{К}^{-2}$ представлено на рис. 23. Соответствующее соотношение при других значениях A^{**} можно получить параллельным смещением этих линий. При обратных смещениях изменение тока обусловлено главным

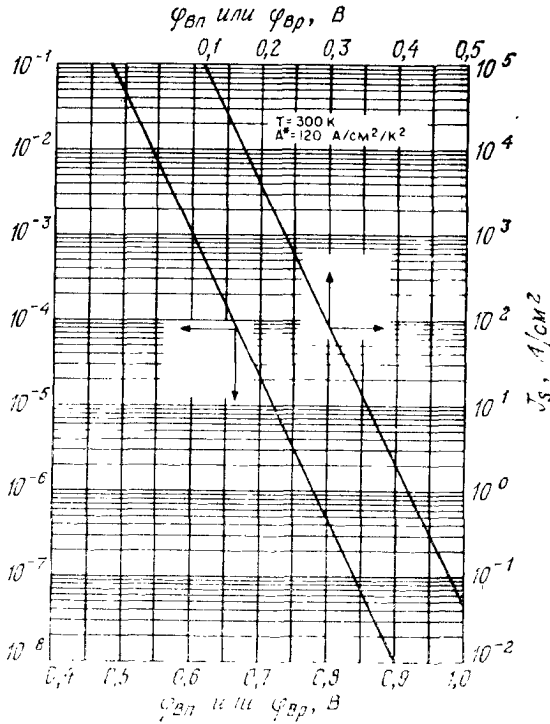


Рис. 23. Теоретическая зависимость плотности тока насыщения от высоты барьера при $T = 300$ К и $A^{**} = 120$ А·см⁻²·К⁻².

образом понижением высоты барьера за счет эффекта Шоттки:

$$J_R \approx J_S \quad (\text{при } V_R > 3kT/q) = A^{**}T^2 \exp\left(-\frac{q\Phi_{B0}}{kT}\right) \times \exp\left(+\frac{q\sqrt{q\mathcal{E}/4\pi\epsilon_s}}{kT}\right), \quad (82)$$

где

$$\mathcal{E} = \sqrt{\frac{2qN_D}{\epsilon_s} \left(V + V_{bi} - \frac{kT}{q}\right)}.$$

Когда высота барьера $q\Phi_{Bn}$ значительно меньше ширины запрещенной зоны, ток, обусловленный генерационно-рекомбинационными процессами, мал по сравнению с током эмиссии через барьер. В этом случае обратный ток будет возрастать с ростом напряжения в соответствии с выражением (82).

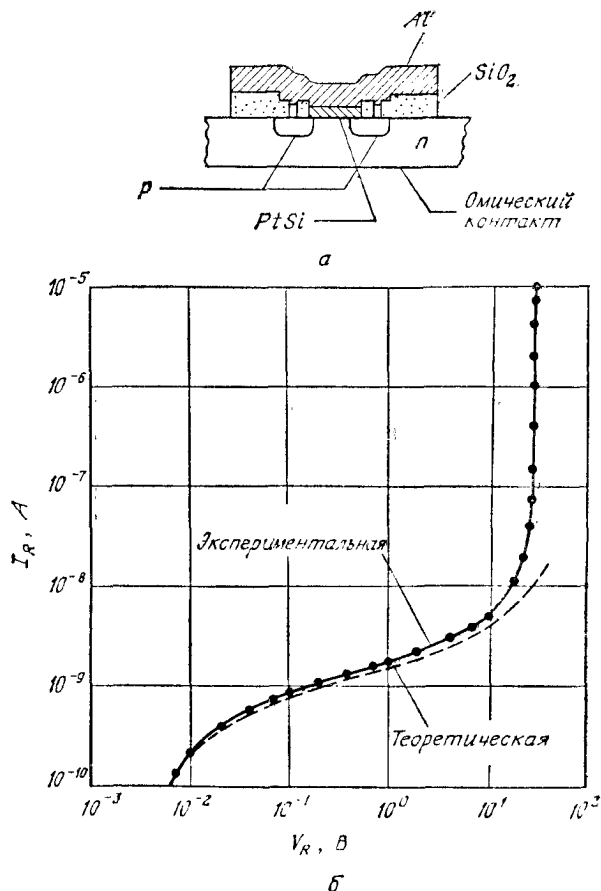


Рис. 24. Диод PtSi — Si с диффузионным охранным кольцом (а) и его вольт-амперные характеристики (экспериментальная и рассчитанная по формуле (82)) [32].

В большинстве применяемых на практике диодах Шоттки основной вклад в обратный ток вносят краевые токи утечки. Они обусловлены концентрацией электрического поля на краю металлического электрода и аналогичны токам в $p-n$ -переходах с малым радиусом кривизны ($r_j \rightarrow 0$) (гл. 2). Для уменьшения краевых токов используют диффузионные охранные кольца [32] (рис. 24, а). Охранное кольцо представляет собой диффузионную область p -типа, профиль легирования которой подбирается так, чтобы напряжение пробоя перехода было больше напряжения пробоя контакта металл—полупроводник. При устранении краевых эффектов прямые и обратные вольт-амперные характеристики

диодов Шоттки получают близкими к идеальным. На рис. 24, б приведены экспериментальная и теоретическая характеристики диода PtSi—Si с охранным кольцом. Соответствие между ними очень хорошее. Резкое увеличение тока при напряжении ~ 30 В обусловлено лавинным пробоем, как этого и следовало ожидать при концентрации доноров $\sim 2,5 \cdot 10^{16}$ см $^{-3}$.

Эффективность использования охранного кольца с целью предупреждения преждевременного пробоя и утечки через поверхность может быть установлена из зависимости тока утечки от диаметра контакта при постоянном напряжении. Для этого на полупроводнике нужно изготовить несколько диодов Шоттки разных размеров. На рис. 25 приведена зависимость тока утечки при обратном смещении от диаметра диода [33]. Прямая, проведенная через экспериментальные точки, имеет тангенс угла наклона, равный 2. Последнее означает, что ток утечки пропорционален площади прибора. Если бы основной вклад в ток давали краевые токи утечки, то наклон этой прямой был бы равен 1.

В некоторых диодах Шоттки возникает дополнительная зависимость обратного тока от напряжения, являющаяся следствием того, что в случае отсутствия на поверхности раздела металл—

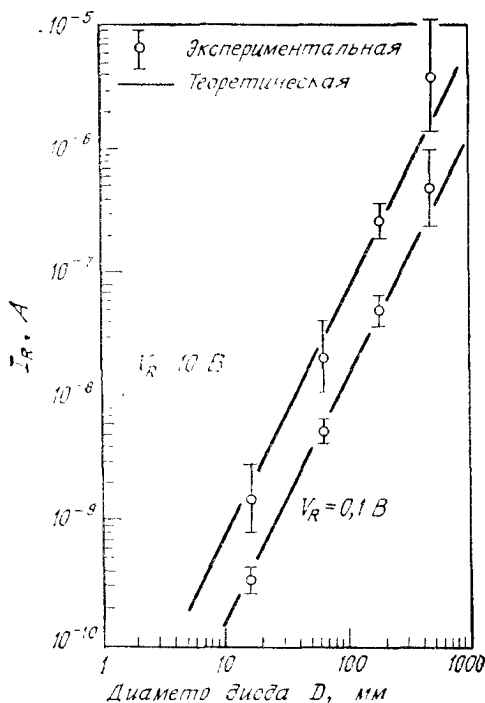


Рис. 25. Зависимость обратного тока утечки от диаметра NiSi — Si-диода на кремнии n -типа с $N_D = 6 \cdot 10^{15}$ см $^{-3}$ [33].

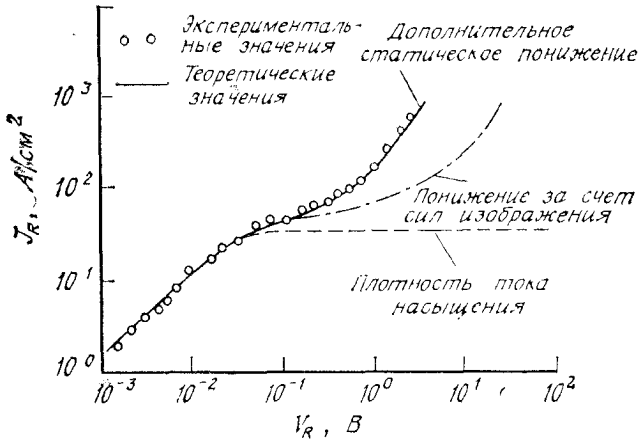


Рис. 26. Теоретические и экспериментальные вольт-амперные характеристики RhSi — Si-диода при обратном смещении [19].

полупроводник промежуточного слоя волновая функция электрона в металле проникает в запрещенную зону полупроводника. Этот квантовомеханический эффект приводит к образованию на поверхности раздела статического дипольного слоя. Вызванное им изменение высоты барьера зависит от приложенного напряжения (т. е. от $d\varphi_{BO}/d\mathcal{E}_m \neq 0$) и в первом приближении может быть представлено в виде

$$(\Delta\varphi)_{\text{статич}} \simeq \alpha \mathcal{E}_m, \quad (83)$$

где $\alpha \equiv d\varphi_{BO}/d\mathcal{E}_m$. Из рис. 26 видно, что при $\alpha \approx 17 \text{ \AA}$ достигается хорошее соответствие между теоретическими и экспериментальными значениями обратного тока.

Метод энергии активации. Принципиальное преимущество этого метода измерения высоты барьера Шоттки заключается в том, что не требуется о площади дополнительных предположений электрически активной области. Метод энергии активации часто используется при исследовании таких поверхностей раздела, где неизвестно точное значение площади контакта. В случае если поверхность плохо очищена или на ней не полностью произошла реакция, электрически активная площадь может составлять только часть геометрической площади. Кроме того, если металлургическая граница получена на неровной поверхности полупроводника, то электрически активная площадь может оказаться больше внешней геометрической.

Умножая уравнение (49) на A_e (площадь электрически активной области), получим

$$\ln(I_F/T^2) = \ln(A_e A^{**}) - q(\varphi_{Bn} - V_F)/kT, \quad (84)$$

где $q (\phi_{Bn} - V_F)$ — энергия активации. В ограниченном интервале температур (например, $273 \text{ K} < T < 373 \text{ K}$) можно считать, что величины A^{**} и ϕ_{Bn} не зависят от температуры. Если при этом напряженье смещения в прямом направлении поддерживается постоянным, то из тангенса угла наклона зависимости $\ln(I_F/T^2)$ от $1/T$ найдем высоту барьера ϕ_{Bn} , а из линейной экстраполяции до $1/T = 0$ получим произведение электрически активной площади A_e на эффективную постоянную Ричардсона A^{**} .

Метод энергии активации занимает важное место при исследовании металлургических реакций на поверхности полупроводника. На рис. 27 приведены зависимости тока через контакты Al—Si *n*-типа от температуры, полученные после отжига при разных температурах. Изменение наклона этих прямых указывает на то, что эффективная высота барьера Шоттки почти линейно увеличивается от 0,71 до 0,81 В при изменении температуры отжига от 450 до 650 °С. Измерения высоты барьера методами вольт-амперных и вольт-фарадных характеристик приводят к аналогичному результату.

Очевидно, что когда достигается эвтектическая температура системы Al—Si ($\sim 580^\circ\text{C}$), структура контакта должна существенно измениться. После отжига при температуре выше эвтек-

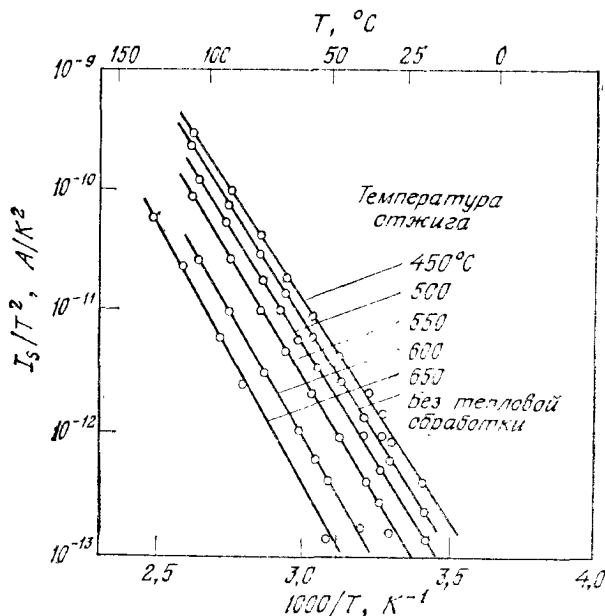


Рис. 27. Зависимость тока от температуры в координатах, используемых для определения высоты барьера [47].

тической точка пересечения прямых на рис. 27 с осью ординат смещается, что соответствует двукратному увеличению электрически активной области диода Al—Si.

Метод вольт-фарадной характеристики. Если к постоянному напряжению смещения добавить слабое переменное, то на поверхности металла и в полупроводнике будут индуцироваться дополнительные переменные заряды противоположных знаков. Соотношение между C и V представляется формулой (9). На рис. 28 приведены типичные зависимости $1/C^2$ от приложенного напряжения. Найдя точку пересечения экстраполирующей прямой с осью напряжений, можно определить высоту барьера [31, 34]:

$$\Phi_{Вп} = V_i + V_n + \frac{kT}{q} - \Delta\phi, \quad (85)$$

где V_i — точка пересечения с осью напряжений, а qV_n — разность энергий между уровнем Ферми и дном зоны проводимости в полупроводнике, которую можно вычислить, если известна концентрация легирующей примеси. Последнюю в свою очередь можно найти из тангенса угла наклона рассматриваемой зависимости (формула (10в)). (Этот метод можно использовать также для определения профиля легирования эпитаксиальных слоев.)

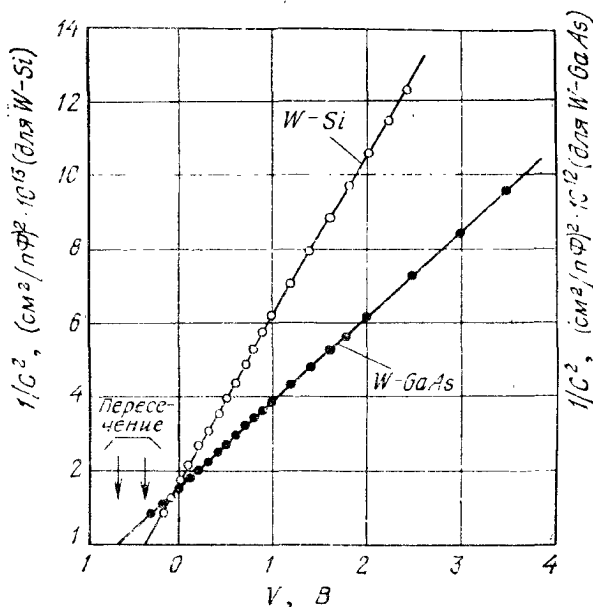


Рис. 28. Зависимость $1/C^2$ от приложенного напряжения для диодов W — Si и W — GaAs [31].

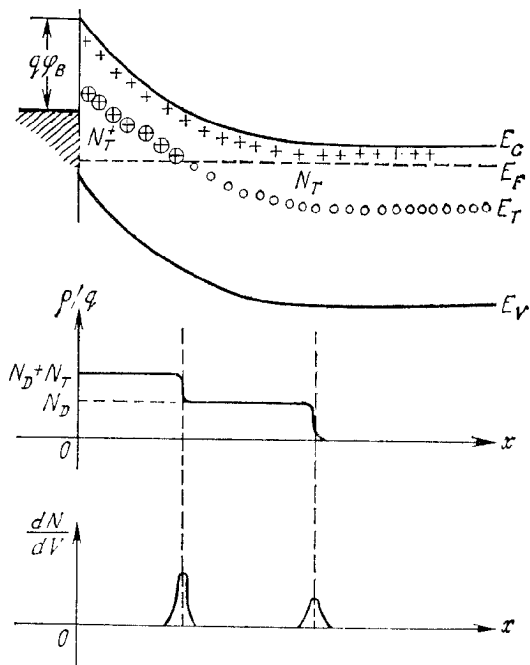


Рис. 29. Полупроводник с одним мелким и одним глубоким донорными уровнями. N_D и N_T — концентрации соответственно мелких и глубоких доноров [35].

Метод вольт-фарaday характеристики может быть использован для изучения глубоких примесных уровней. На рис. 29 показан полупроводник с одним мелким и одним глубоким донорными уровнями [35]. При соответствующем изгибе зон все глубокие примеси с энергией выше уровня Ферми ионизированы, что приводит к более высокой плотности пространственного заряда вблизи поверхности. Если на постоянное напряжение смещения наложено слабое переменное и если заполнение глубоких примесных центров удается следовать за этим сигналом, то примесные центры дадут дополнительный вклад dN/dV в емкость. На рис. 30 приведена зависимость $1/C^2$ от V при разных частотах. Низкочастотные зависимости можно использовать для изучения свойств глубоких примесей. Для определения высоты барьера в полупроводнике с одним мелким и одним глубоким примесными уровнями (рис. 29) надо получить вольт-фарaday характеристику при двух различных температурах [39].

Фотоэлектрический метод. Этот метод является прямым и самым точным методом определения высоты барьера [36]. Принцип

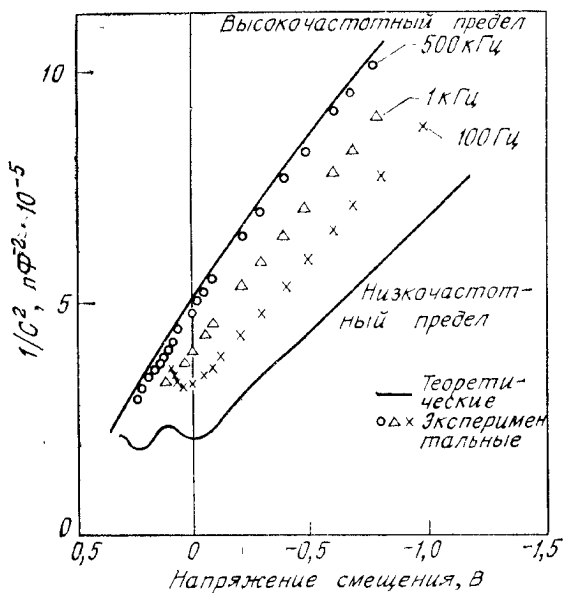


Рис. 30. Зависимость $1/C^2$ от напряжения при разных частотах [35].

ципиальная схема эксперимента показана на рис. 31. Когда образец освещается со стороны металла монохроматическим светом с энергией $h\nu > q\phi_{Вн}$, в металле возбуждаются электроны с энергией, достаточной для преодоления барьера — процесс (1). Если при этом $h\nu > E_g$, а пленка металла достаточно тонкая, то свет, частично проходящий через нее, генерирует электронно-дырочные пары и в полупроводнике — процесс (2). При освещении образца со стороны подложки также возможно возбуждение электронов в металле с энергией, достаточной для преодоления потенциального барьера, т. е. с $h\nu > q\phi_{Вн}$ — процесс (1). Однако, когда $h\nu > E_g$, свет будет сильно поглощаться вблизи тыловой поверхности полупроводника и вероятность того, что генерированные здесь электронно-дырочные пары достигнут границы раздела металл — полупроводник, очень мала.

В теории Фаулера [37] зависимость квантового выхода R от энергии фотона $h\nu$ выражается формулой

$$R \sim \frac{T^2}{\sqrt{E_s - h\nu}} \left[\frac{x^2}{2} + \frac{\pi^2}{6} - \left(e^{-x} - \frac{e^{-2x}}{4} + \frac{e^{-3x}}{9} - \dots \right) \right]$$

при $x \geq 0$, (86)

где $h\nu_0 \equiv q\phi_{Вн}$ — высота барьера, E_s — сумма $h\nu_0$ и энергии Ферми, отсчитанной от дна зоны проводимости металла, $x \equiv h(\nu -$

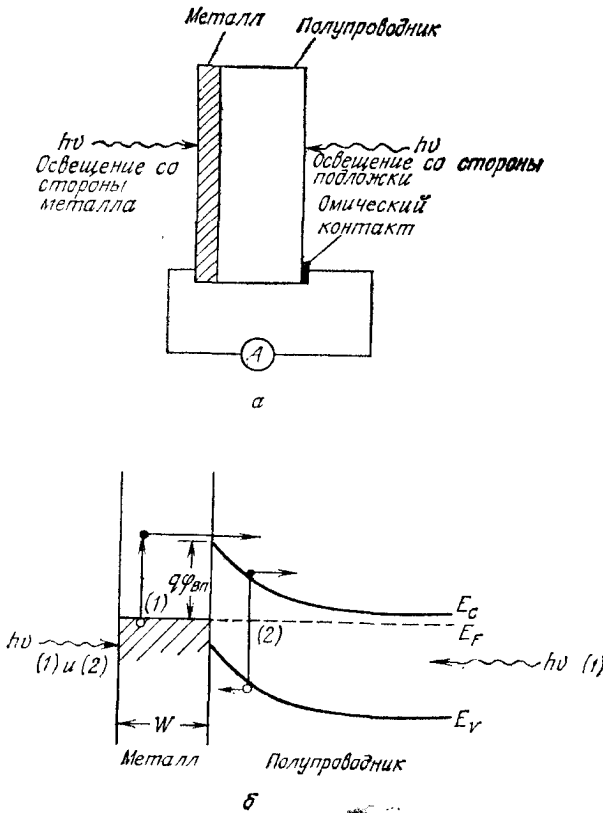


Рис. 31. Принципиальная схема установки для фотоэлектрических измерений (а) и энергетическая диаграмма процессов фотовозбуждения (б).

$\nu_0)/kT$. При условии, что $E_s \gg h\nu$ и $x > 3$, вместо выражения (86) можно использовать упрощенные выражения

$$R \sim (h\nu - h\nu_0)^2 \text{ при } h(\nu - \nu_0) > 3kT, \quad (86a)$$

или

$$\sqrt{R} \sim h(\nu - \nu_0). \quad (86б)$$

Построив зависимость корня квадратного из фотоотклика от энергии фотона, получим прямую линию. Экстраполируя эту прямую на ось энергий, сразу получим высоту энергетического барьера. На рис. 32 показана зависимость фотоотклика диодов $W-Si$ и $W-GaAs$ от энергии фотона. Видно, что высота барьера для этих приборов равна соответственно 0,65 и 0,80 эВ.

Фотоэлектрические измерения можно использовать также для определения других параметров прибора и свойств входящих в него материалов. Эти измерения были использованы, в частности, для определения высокочастотной диэлектрической проницаемости кремния [10]. Находя граничную частоту фотоотклика при различных запирающих напряжениях, можно определить соответствующее понижение барьера $\Delta\phi$. Построив, как показано на рис. 5, зависимость $\Delta\phi$ от $\sqrt{\mathcal{E}}$, найдем эффективную диэлектрическую проницаемость ϵ_s/ϵ_0 . Кроме того, фотоэлектрические измерения были использованы для определения зависимости высоты барьера от температуры [38]. С этой целью измерялась граничная частота фотоотклика при разных температурах диода Au—Si. Оказалось, что сдвиг граничной частоты достаточно хорошо «отслеживает» температурную зависимость ширины запрещенной зоны. Это означает, что уровень Ферми на поверхности раздела Au—Si занимает фиксированное положение относительно края валентной зоны. Последнее вполне соответствует выводам разд. 5.5.1.

Результаты измерения высоты барьера. Для измерения высоты потенциального барьера в системе металл—полупроводник использовались все описанные выше методы. В случае совершенного контакта с чистой поверхностью полупроводника эти методы приводят к одному и тому же значению с точностью $\pm 0,02$ эВ. Причиной существенных расхождений между результатами, полученными различными методами, могут быть загрязнения поверхности,

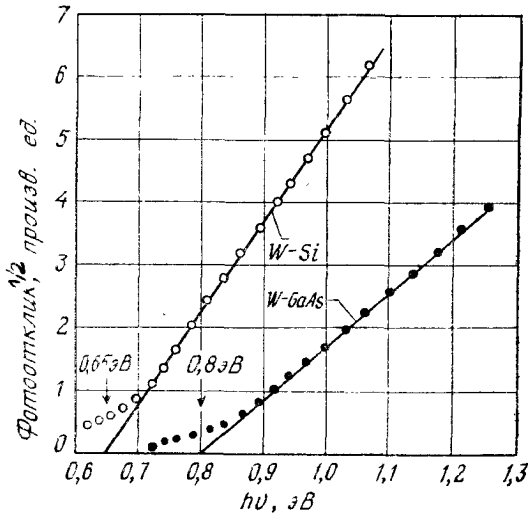


Рис. 32. Зависимость корня квадратного из фотоотклика, пересчитанного на один фотон, от энергии фотона для диодов W — Si и W — GaAs [31].

Таблица 3. Экспериментальные значения высоты барьера Шоттки при 300 К (в вольтах)

Полупроводник	Тип	E_g' , эВ	Ag	Al	Au	Cr	Cu	Hf	In	Mg	Mo	Ni	Pb	Pd	Pt	Ta	Ti	W
Алмаз	p	5,47			1,71		0,52		0,64			0,49	0,38					0,48
Ge	n	0,66	0,54	0,48	0,59				0,55								0,50	0,67
Ge	p		0,50	0,72	0,30	0,61	0,58				0,68	0,61		0,81	0,90		0,50	
Si	n	1,12	0,78	0,58	0,34	0,50	0,46	0,58			0,42	0,51	0,55				0,61	0,45
Si	p		0,54	2,00	1,95									1,00				
SiC	n	3,00			1,20													
AlAs	n	2,16			0,55													
AlSb	p	1,63			3,10													
BN	p	7,50			0,87													
BP	p	6,00			0,60													
GaSb	n	0,67			0,90		0,82	0,72								0,85		0,80
GaAs	n	1,42	0,88	0,80	0,42			0,68										
GaAs	p		0,63		1,30	1,06	1,20	1,84		1,04	1,13	1,27					1,12	
GaP	n	2,24	1,20	1,07	0,72													
GaP	p				0,17													
InSb	n	0,16	0,18		0,47													
InAs	p	0,33			0,52													
InP	n	1,29	0,54		0,76													
InP	p				0,78		0,50					0,45	0,59	0,62	1,10		0,84	
CdS	n	2,43	0,56	Оми- че- ский														
CdSe	n	1,70	0,43		0,49		0,33								0,37			
CdTe	n		0,81	0,76	0,71									0,76	0,76			
ZnO	n	3,20		0,68	0,65		0,45		0,30					0,68	0,75			
ZnS	n	3,60	1,65	0,80	2,00		1,75		1,50	0,82				1,87	1,84	0,30		
ZnSe	n		1,21	0,76	1,36		1,10		0,91				1,16	1,40	1,10			
PbO	n		0,95						0,93			0,96	0,95					

промежуточный изолирующий слой, краевые токи утечки или глубокие примесные уровни.

В табл. 3 приведены результаты измерений [39—41] высоты барьера Шоттки на ряде полупроводников. Они получены на барьерах, изготовленных напылением сверхчистых металлов в системах с высоким вакуумом на сколотую или химически очищенную поверхность полупроводника. Как и следовало ожидать, наиболее изученными являются контакты к кремнию и GaAs. Из металлов чаще всего используют золото, алюминий и платину. Результаты измерений высоты барьеров в системах силицид металла—кремний *n*-типа приведены в табл. 4 [42—43а].

Высота энергетического барьера обычно очень чувствительна к термообработке образца до и после напыления [44]. На рис. 33 показана зависимость высоты барьера в диодах на основе Si и GaAs *n*-типа, измеренной при комнатной температуре, от температуры отжига. Высота барьера в диоде Pt—Si составляет 0,9 В. После отжига при 300 °С и выше на поверхности раздела образуется силицид PtSi [45], что приводит к уменьшению Φ_{bn} до 0,85 В. В диодах Pt—GaAs на поверхности раздела после отжига образуется PtAs₂ и высота барьера увеличивается с 0,84 до 0,87 В [46]. При отжиге диода Al—Si выше 450 °С высота

Таблица 4. Характеристики барьеров силицид металла—кремний *n*-типа

Силицид металла	Φ_{bn} , В	Структура	Температура образования, °С	Температура плавления, °С
CoSi	0,68	Кубическая	400	1460
CoSi ₂	0,64	»	450	1326
CrSi ₂	0,57	Гексагональная	450	1475
HfSi	0,53	Орторомбическая	550	2200
IrSi	0,93	—	300	—
MnSi	0,76	Кубическая	400	1275
Mn ₁₁ Si ₁₉	0,72	Тетрагональная	800 ¹	1145
MoSi ₂	0,55	»	1000 ¹	1980
Ni ₂ Si	0,7—0,75	Орторомбическая	200	1318
NiSi	0,66—0,75	»	400	992
NiSi ₂	0,7	Кубическая	800 ¹	993
Pd ₂ Si	0,72—0,75	Гексагональная	200	1330
PtSi	0,84	Орторомбическая	300	1229
RhSi	0,69	Кубическая	300	—
TaSi ₂	0,59	Гексагональная	750 ¹	2200
TiSi ₂	0,60	Орторомбическая	650	1540
WSi ₂	0,65	Тетрагональная	650	2150
ZrSi ₂	0,55	Орторомбическая	600	1520

¹ На чистой поверхности может быть $\leq 700^\circ \text{C}$.

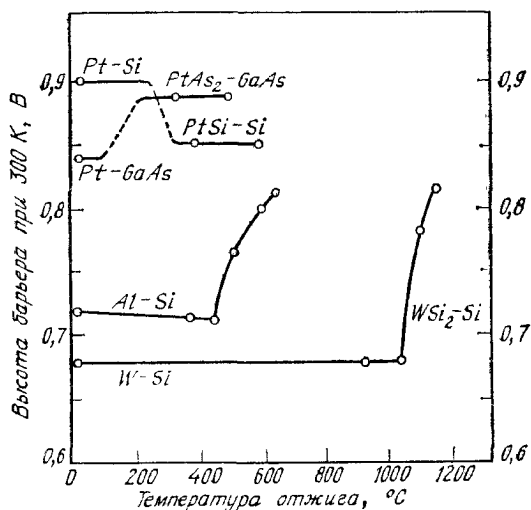


Рис. 33. Высота барьера в диодах Шоттки на основе Si и GaAs *n*-типа при 300 К после отжига при различных температурах.

барьера начинает увеличиваться [47], что происходит, вероятнее всего, вследствие диффузии Si в Al (рис. 27). В диодах W—Si высота барьера остается постоянной вплоть до температур отжига, превышающих 1000 °С, после чего образуется WSi_2 [48].

5.5.3. Регулирование высоты барьера

Высота обычного барьера Шоттки в основном определяется свойствами металла и поверхности раздела металл—полупроводник и почти не зависит от степени легирования. Поэтому на заданном полупроводнике (обычно Si *n*- и *p*-типа) можно получить лишь конечный набор высот барьеров Шоттки (табл. 3 и 4).

Варьировать высоту потенциального барьера при контакте заданных материалов можно путем создания на поверхности полупроводника тонкого ($\sim 100 \text{ \AA}$ и меньше) слоя с другой концентрацией легирующей примеси (такие слои можно создавать, например, ионной имплантацией) [49—51]. Этот способ используется, в частности, в том случае, когда выбор металла обусловлен его металлургическими свойствами.

На рис. 34, в показаны идеализированные зонные диаграммы контактов к полупроводнику *n*-типа с тонким *n*⁺- или *p*⁺-слоем, служащим соответственно для уменьшения и увеличения высоты барьера. Рассмотрим сначала случай уменьшения барьера. Рас-

пределение поля (рис. 34, б) при этом определяется выражением

$$\begin{aligned} \mathcal{E} &= -\mathcal{E}_m + qn_1x/\epsilon_s \quad \text{при } 0 < x < a, \\ \mathcal{E} &= -\frac{qn_2}{\epsilon_s}(W-x) \quad \text{при } a < x < W, \end{aligned} \quad (87)$$

где \mathcal{E}_m — максимальная величина электрического поля, которая достигается на границе раздела металл—полупроводник и равна

$$\mathcal{E}_m = \frac{q}{\epsilon_s} [n_1a + n_2(W-a)]. \quad (88)$$

Вызванное этим полем понижение высоты потенциального барьера из-за силы изображения есть

$$\Delta\varphi = \sqrt{q\mathcal{E}_m/4\pi\epsilon_s}. \quad (89)$$

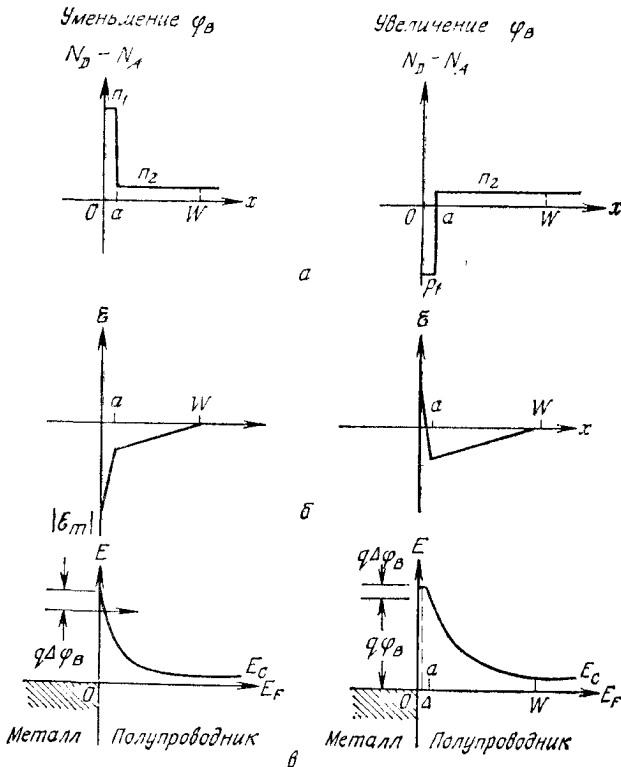


Рис. 34. Идеализированные зонные диаграммы контактов к полупроводнику n-типа с тонкими n⁺- или p⁺-слоями, служащими соответственно для уменьшения и увеличения высоты барьера.

Для барьера Шоттки на Si и GaAs с $n_2 \approx 10^{16}/\text{см}^3$ и меньше величина $n_2(W - a)$ при нулевом смещении составляет $\sim 10^{11}/\text{см}^2$. Поэтому если $n_1 a$ существенно больше 10^{11} см^{-2} , то из выражений (88) и (89) имеем

$$\mathcal{E}_m \approx qn_1 a / \epsilon_s, \quad \Delta \varphi \approx \sqrt{\frac{n_1 a}{4\pi}}. \quad (90), (91)$$

При $n_1 a = 10^{12}$ и $n_1 a = 10^{13} \text{ см}^{-2}$ барьер понижается соответственно на 0,045 и 0,14 В.

Хотя использование n^+ -слоя приводит к понижению высоты барьера, наиболее важными в таких структурах обычно становятся эффекты туннелирования. Из выражения (90) следует, что при $n_1 a = 10^{13} \text{ см}^{-2}$ поле \mathcal{E}_m равно $1,6 \cdot 10^6 \text{ В/см}$. Такое поле возникает в диоде Au—Si с концентрацией примесей 10^{19} см^{-3} при нулевом смещении. Из рис. 11, а видно, что плотность тока насыщения этого диода составляет $\sim 10^{-3} \text{ А/см}^2$. Такая плотность тока соответствует высоте барьера 0,6 В (рис. 23), т. е. на 0,2 В меньше высоты барьера в низколегированном диоде Au—Si, где она равна 0,8 В. Расчетные зависимости эффективной высоты барьеров на Si и GaAs от \mathcal{E}_m приведены на рис. 35 [52]. Увеличивая поле \mathcal{E}_m от 10^5 до 10^6 В/см , можно уменьшить эффективную высоту барьера на 0,2 В в кремнии и более чем на 0,3 В в GaAs.

Выбирая параметры n_1 и a , необходимо следить за тем, чтобы при прямом смещении сильное понижение барьера и дополнительный туннельный ток существенно не увеличили фактор неидеальности n . При обратном смещении в обычно используемом интервале напряжений эти эффекты не приводят к большим токам утечки.

Если на поверхности полупроводника сформирован тонкий слой с противоположным типом легирования, то эффективная высота барьера может увеличиться. На рис. 34, а показан профиль легирования полупроводника n -типа с поверхностным слоем p^+ -типа. Можно показать, что при $x = 0$ потенциальная энергия электрона в полупроводнике равна $q\varphi_B$ и достигает максимума при $x = \Delta$, где

$$\Delta = \frac{1}{p_1} [ap_1 - (W - a)n_2]. \quad (92)$$

Эффективная высота такого барьера равна

$$\varphi'_B = \varphi_B + \mathcal{E}_m \Delta - \frac{qp_1 \Delta^2}{2\epsilon_s}. \quad (93)$$

Если $p_1 \gg n_2$ и $ap_1 \gg n_2 W$, то выражение (93) стремится к $\varphi_B + qp_1 a^2 / \epsilon_s$. При этом с ростом $p_1 a$ будет увеличиваться эффективная высота барьера.

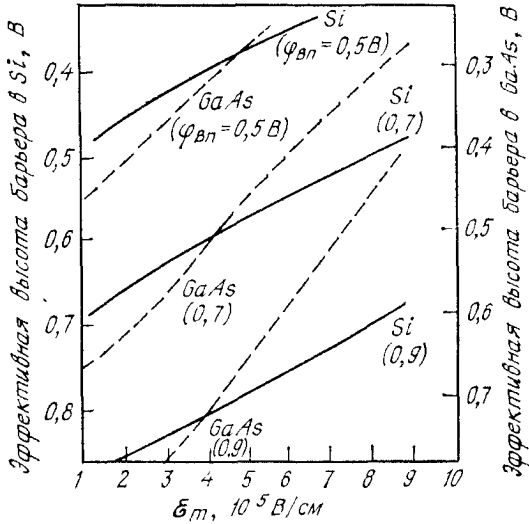


Рис. 35. Расчетные значения эффективной высоты барьеров на Si и GaAs в зависимости от поля E_m [52].

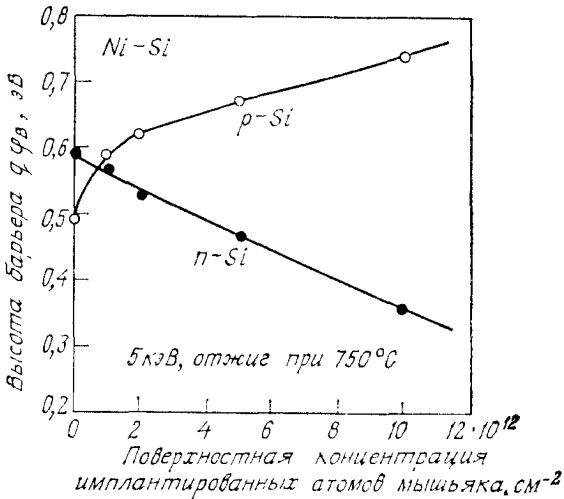


Рис. 36. Зависимость эффективной высоты энергетического барьера для дырок в подложке p-типа и для электронов в подложке n-типа от имплантированной дозы мышьяка [52].

На рис. 36 приведены результаты измерения высоты барьера в диоде Ni—Si с тонким поверхностным слоем, полученным имплантацией мышьяка [52]. Видно, что по мере увеличения дозы имплантации эффективная высота барьера в диоде с подложкой *p*-типа увеличивается, а в диоде с подложкой *n*-типа уменьшается.

5.6. СТРУКТУРЫ ПРИБОРОВ

Как уже говорилось выше, свойства диода Шоттки во многом сходны со свойствами резкого несимметричного *p—n*-перехода. Однако, поскольку перенос заряда здесь осуществляется основными носителями, быстрдействие диодов Шоттки значительно выше. Отсюда следует, что диоды Шоттки в принципе могут выполнять почти все функции диодов с *p—n*-переходами. Исключения составляют лишь *p—n*-диоды с накоплением заряда, поскольку в приборах, работающих на основных носителях, время накопления чрезвычайно мало. В этом разделе мы рассмотрим различные структуры с барьерами Шоттки и области их применения.

Первыми диодными структурами были точечно-контактные выпрямители, которые изготавливались путем контакта к поверхности полупроводника заостренного металлического проводника. Контакт может быть создан либо механически, либо с помощью электрического разряда, в результате которого возникает небольшой сплавной *p—n*-переход.

Точечно-контактный выпрямитель имеет обычно плохие прямые и обратные вольт-амперные характеристики по сравнению с планарным диодом Шоттки. Кроме того, эти характеристики непредсказуемы теоретически, поскольку трудно учесть и проконтролировать такие параметры, как давление проводника, площадь контакта, структура кристалла, состояние поверхности, химический состав проводника, тепловой режим образования контакта. Преимуществом точечно-контактного выпрямителя является маленькая площадь и, следовательно, очень маленькая емкость, что весьма желательно при использовании в СВЧ-технике. К недостаткам такого выпрямителя относятся: большое сопротивление растекания ($R_s = \rho/2\pi r_0$, где r_0 — радиус полусферического точечного контакта); большой ток утечки, возникающий в основном вследствие влияния поверхности и приводящий к ухудшению коэффициента выпрямления; плавные предпробойные характеристики, обусловленные концентрацией силовых линий поля под контактом.

В настоящее время большинство диодов Шоттки изготавливается методами планарной технологии, такими, как термическое испарение, химическое разложение, электронно-лучевая бомбардировка, распыление, нанесение металлов анодированием. При этом

поверхность полупроводника предварительно обрабатывается такими методами, как химическое травление, полировка, скол в вакууме, тепловая обработка, ионная бомбардировка. Большинство контактов металл—полупроводник создается в вакуумных системах [53]. Одним из наиболее важных параметров, характеризующих вакуумное напыление, является давление насыщенных паров, определяемое как давление, при котором твердое тело или жидкость находится в равновесии с собственным паром. Зависимость давления насыщенных паров от температуры [54] для наиболее часто используемых элементов приведена на рис. 37.

Диод с малой площадью контакта (рис. 38, а) изготавливают методами планарной технологии на эпитаксиальном слое n -типа, образованном на подложке n^+ -типа, и обычно используют как СВЧ-детектор [55]. Для получения диода с хорошими характеристиками нужно минимизировать его последовательное сопротивление и емкость.

Сопротивление контакта R_j можно получить, используя выражение (53):

$$R_j \equiv \frac{\partial V}{\partial I} = \frac{nkT}{qIA_j}, \quad (94)$$

где A_j — площадь контакта. Типичная экспериментальная зависимость R_j от I для диодов Au—Si и Au—GaAs и соответствующая зависимость для кремниевого точечно-контактного диода, рассмотренного выше, приведены на рис. 39. Отметим, что при больших прямых смещениях сопротивление прибора не уменьшается до нуля, как это следует из выражения (94), а принимает постоянное значение, которое является последовательным сопротивлением R_s , определяемым по формуле (рис. 13)

$$R_s = \frac{1}{A_j} \int_{x_1}^{x_2} \rho(x) dx + \frac{\rho_B}{4r} + R_c. \quad (95)$$

Здесь первое слагаемое является сопротивлением квазинейтральной области, расположенной между x_1 (границей обедненного слоя) и x_2 (границей между эпитаксиальным слоем и подложкой). Второе слагаемое есть сопротивление растекания от круглого контакта радиусом r ($A_j = \pi r^2$) в подложку с удельным сопротивлением ρ_B . Последнее слагаемое описывает сопротивление омических контактов к подложке.

При использовании диодов Шоттки в СВЧ-схемах важной характеристикой является частота отсечки при прямом смещении f_{co} , определяемая выражением

$$f_{co} \equiv \frac{1}{2\pi R_j C_F}, \quad (96)$$

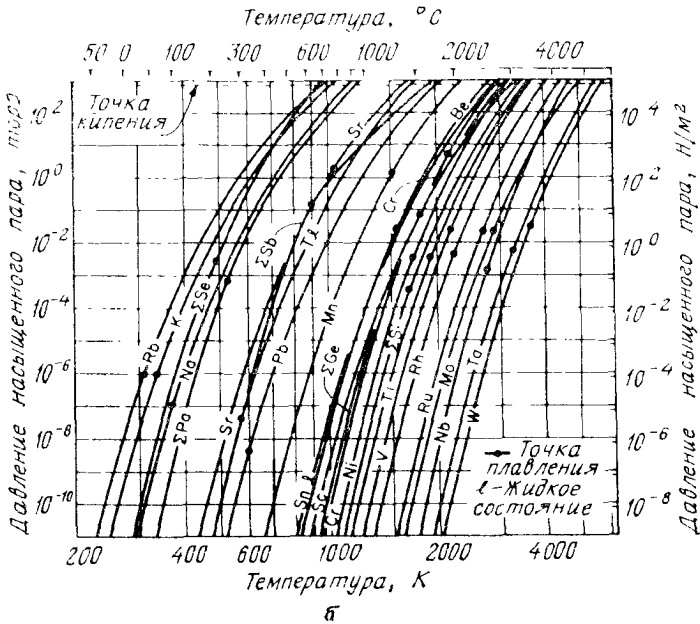
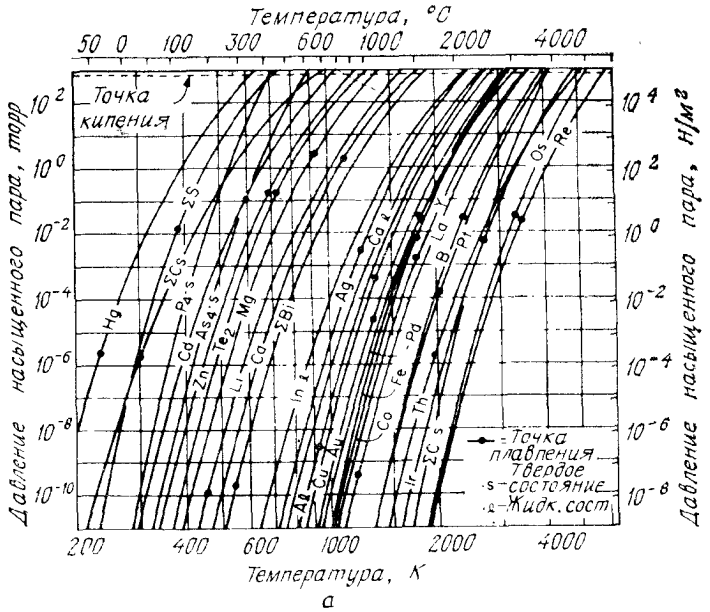


Рис. 37. Зависимость давления насыщенных паров от температуры для жидких и твердых элементов [54].

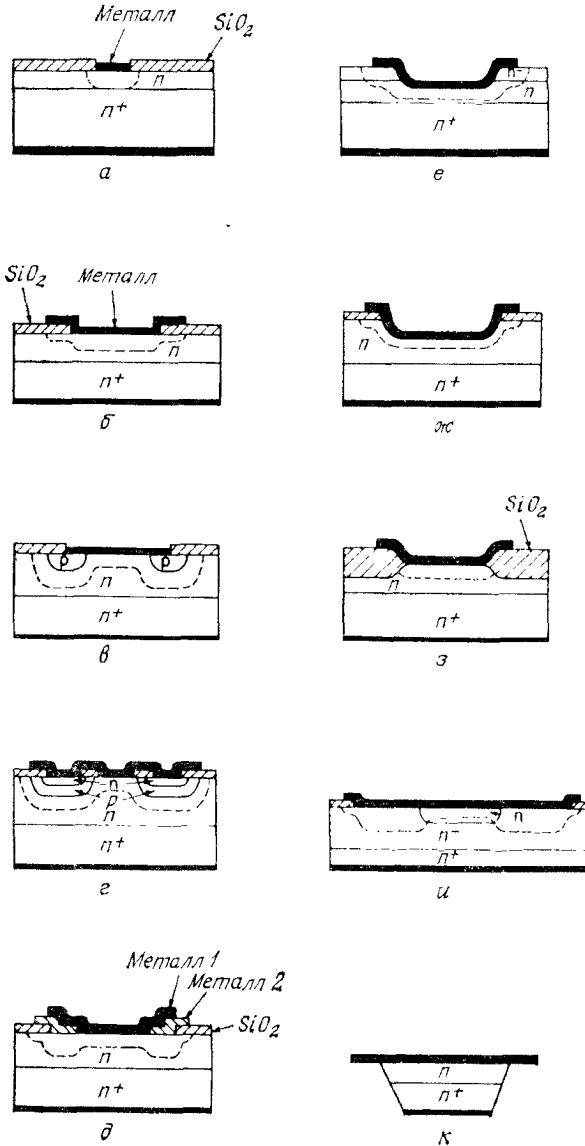


Рис. 38. Структуры с контактами металл — полупроводник.

a — диод с малой площадью контакта; *б* — диод с перекрытием металла; *в* — диод с диффузионным охранным кольцом; *г* — диод с двойным диффузионным охранным кольцом; *д* — диод с металлическим охранным кольцом; *е* — диод с высокоомным охранным слоем; *ж* — диодная структура с углублением в полупроводнике; *з* — диод с изоляцией окислом; *и* — диод с дополнительной диффузией и перекрытием металла; *к* — коническая структура диода.

где R_F — сопротивление, C_F — емкость диода при прямом смещении 0,1 В [55а]. Частота f_{co} значительно меньше частоты отсечки при нулевом смещении, и ее можно считать нижним пределом для частот отсечки. Типичные значения этой величины [56] приведены на рис. 40. Отметим, что при указанной степени легирования и диаметре контакта (~ 10 мкм) большую частоту отсечки имеет диод Шоттки на n -GaAs. В основном это обусловлено более высокой подвижностью электронов в GaAs.

Для повышения надежности и стабильности приборов нужны диоды с достаточно большой площадью контакта, однако желательно, чтобы их емкость при этом оставалась малой. Было показано, что данным требованиям может удовлетворить барьер Мотта. Барьером Мотта называется контакт металл—полупроводник, в котором эпитаксиальный слой полупроводника очень слабо легирован, а его толщина значительно меньше толщины, требуемой для формирования обедненного слоя. В таком приборе электроны в эпитаксиальном слое практически отсутствуют даже при прямом смещении. На рис. 41 показана зонная диаграмма барьера Мотта. Удельная емкость его обедненной области не зависит от напряжения и равна ϵ_s/W . Поскольку толщину обедненной об-

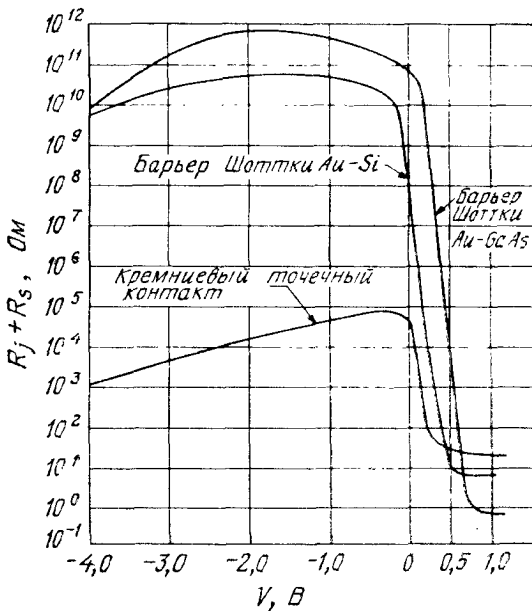


Рис. 39. Зависимость полного сопротивления (сопротивления перехода и последовательного сопротивления образца) от приложенного напряжения для диодов Au — Si, Au — GaAs и кремниевого точечно-контактного диода [56].

ласти барьера Мотта можно сделать значительно большей, чем в случае обычного диода Шоттки, то при заданной частоте отсечки можно увеличить площадь прибора [57]. Перенос заряда в барьерах Мотта определяется в основном диффузионными процессами, которые описываются уравнением (38).

Диод с перекрытием металла [58] (рис. 38, б) имеет почти идеальные вольт-амперные характеристики при прямом смещении и низкие токи утечки при не слишком больших обратных смещениях. Однако с увеличением обратного смещения влияние краев металлического электрода приводит к росту обратного тока. Такие структуры широко используются в интегральных схемах, поскольку их легко получить при металлизации. Важным примером их применения является транзистор с диодом Шоттки между базой и коллектором [59, 60] (рис. 42), позволяющим существенно уменьшить время насыщения. В стандартной технологии изготовления транзисторов со скрытым коллектором такой диод получают путем нанесения более широкого базового контакта, перекры-

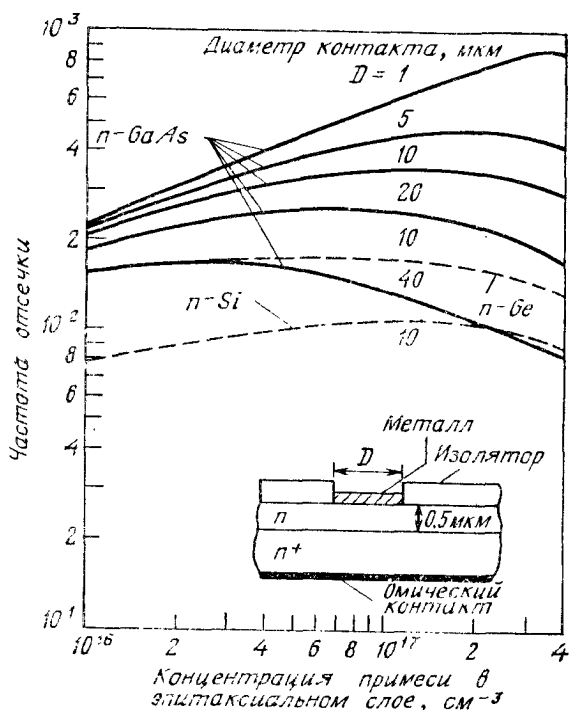


Рис. 40. Зависимость частоты отсечки (в гигагерцах) при прямом смещении от степени легирования эпитаксиального слоя для приборов с контактами разных диаметров и толщиной эпитаксиального слоя 0,5 мкм.

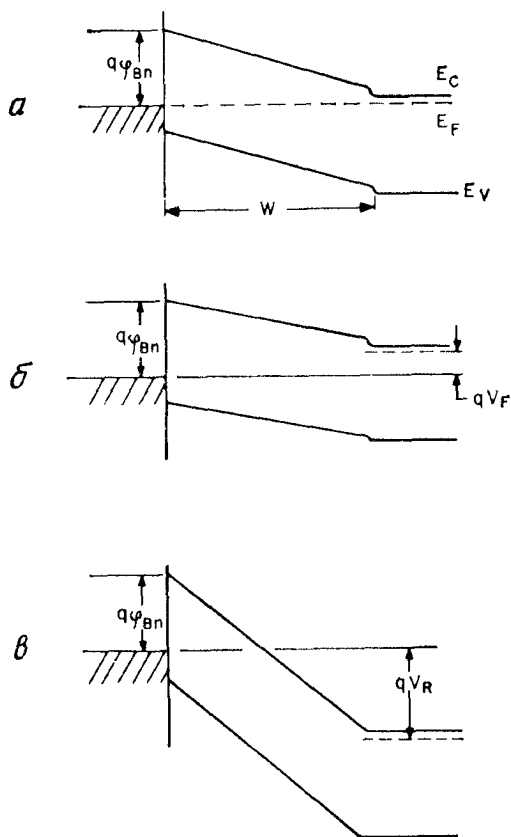
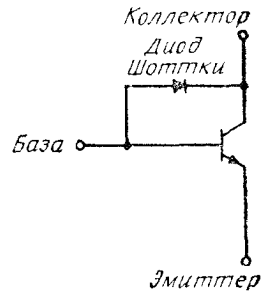


Рис. 41. Зонная диаграмма барьера Мотта при разных смещениях [5].

вающегося с внешним, не подвергнутым диффузии эпитаксиальным слоем [41]. В области насыщения коллекторный переход слегка смещен в прямом направлении, а не в обратном. Если падение напряжения на диоде Шоттки при прямом смещении меньше, чем напряжение между базой и коллектором исходного транзистора в режиме насыщения, то большая часть базового тока потечет через диод, в котором неосновные носители не накапливаются. Поэтому время насыщения транзистора с диодом Шоттки существенно меньше времени насыщения одного транзистора (без диода Шоттки).

Для устранения краевых эффектов было предложено много структур. Структура с диффузионным охранным кольцом (рис. 38, в) [32] обладает почти идеальными прямыми и обратными вольт-амперными характеристиками (рис. 24). К недостаткам такого диода относятся большое время восстановления и большая паразитная емкость, обусловленные наличием $p-n$ -перехода.

Рис. 42. Биполярный транзистор с диодом Шоттки между базой и коллектором [60].



Поэтому его применяют обычно для изучения статических характеристик. Структура с двойным диффузионным охранным кольцом (рис. 38, *а*) [61] позволяет уменьшить время восстановления, но технология ее изготовления относительно сложна. В структуре, показанной на рис. 38, *б*, для создания охрannого кольца используют два типа металла с разной высотой барьера на контакте. Однако на ковалентных полупроводниках большую разность высот барьеров получить трудно. Еще одна структура с охранным кольцом, образующимся в результате нанесения на эпитаксиальный слой тонкого высокоомного слоя [62], приведена на рис. 38, *в*. Поскольку диэлектрическая проницаемость высокоомного слоя больше диэлектрической проницаемости изолятора, паразитная емкость такого диода обычно больше, чем емкость диода, показанного на рис. 38, *б*. В структуре с углублением в полупроводнике (рис. 38, *ж*) возникает проблема надежности, связанная с ухудшением качества поверхности раздела в углублении [63]. Другим подходом к проблеме уменьшения краевых полей является изоляция окислом [64] (рис. 38, *з*). При этом для создания локальной ступеньки окисла требуется специальная планарная операция.

Структура с дополнительной диффузией и перекрытием металла (рис. 38, *и*) [65] представляет собой сдвоенный (параллельно) диод Шоттки, причем в этой структуре отсутствует $p-n$ -переход. Она имеет почти идеальные прямые и обратные вольт-амперные характеристики и очень малое время восстановления. Однако процесс изготовления такого диода включает дополнительное окисление и диффузию, а внешнее кольцо может увеличить его емкость.

В некоторых мощных СВЧ-генераторах (например, в лавинно-пролетных диодах) используют коническую структуру (рис. 38, *к*) [66]. Угол между металлическим покрытием и конусом полупроводника должен превышать 90° . При этом напряженность электрического поля на периферии контакта обычно меньше, чем в центре, что обеспечивает однородность лавинного пробоя по всей его площади.

5.7. ОМИЧЕСКИЙ КОНТАКТ

Омическим контактом называют контакт металл—полупроводник, сопротивление которого пренебрежимо мало по сравнению с объемным сопротивлением полупроводника и сопротивлением растекания. Хороший омический контакт не должен приводить к существенному изменению характеристик прибора, а падение напряжения на таком контакте при пропускании через него требуемого тока должно быть достаточно мало по сравнению с падением напряжения на активной области прибора.

Рассмотрим сначала удельное сопротивление контакта, определяемое как обратная величина от производной плотности тока по напряжению. Наиболее важной характеристикой контакта является удельное сопротивление при нулевом смещении [67]:

$$R_c \equiv \left(\frac{\partial J}{\partial V} \right)_{V=0}^{-1}. \quad (97)$$

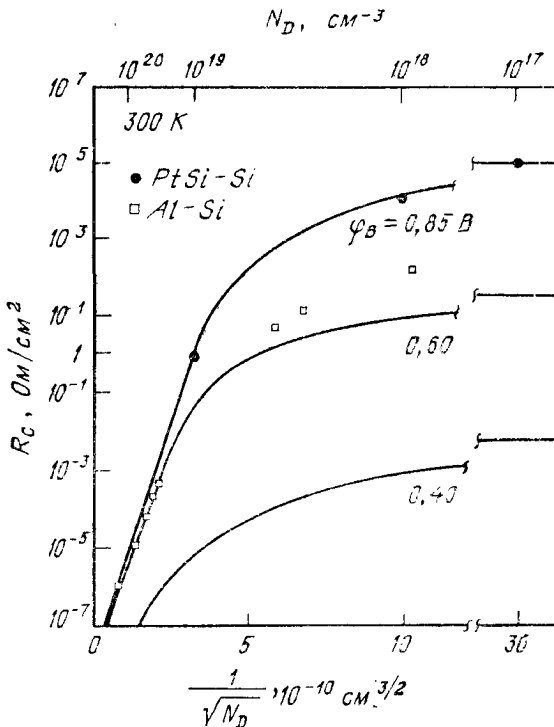
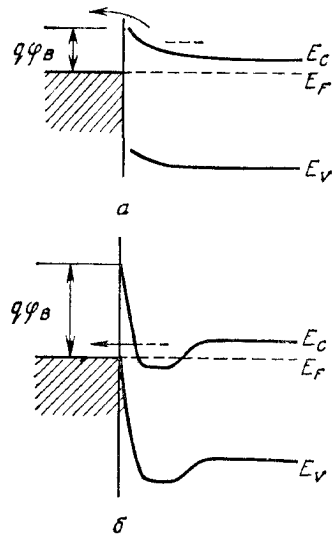


Рис. 43. Теоретические и экспериментальные зависимости удельного сопротивления контактов от $1/\sqrt{N_D}$ [67, 68].

Рис. 44. Омические контакты с малой высотой барьера (а) и высокой степенью легирования (б).



В контактах металл—полупроводник с низким уровнем легирования преобладает термоэлектронная компонента тока, определяемая формулой (49). В этом случае

$$R_c = \frac{k}{qA^*T} \exp\left(\frac{q\varphi_{Bn}}{kT}\right). \quad (98)$$

При выводе этого выражения мы пренебрегли слабой зависимостью высоты барьера от напряжения. Из выражения (98) видно, что для получения малых R_c нужно изготавливать контакты с малой высотой барьера.

В контактах металл—полупроводник с более высоким уровнем легирования преобладает туннельная компонента тока, определяемая формулой (54а). При этом

$$R_c \sim \exp\left(\frac{q\varphi_{Bn}}{E_{00}}\right) = \exp\left[\frac{2\sqrt{\epsilon_s m^*}}{\hbar} \left(\frac{\varphi_{Bn}}{\sqrt{N_D}}\right)\right]. \quad (99)$$

Отсюда видно, что в туннельной области удельное сопротивление контакта экспоненциально зависит от $\varphi_{Bn}/\sqrt{N_D}$.

На рис. 43 [67, 68] приведены результаты расчета зависимости R_c от $1/\sqrt{N_D}$. При $N_D \geq 10^{19} \text{ см}^{-3}$ R_c определяется в основном туннельными процессами и быстро уменьшается по мере повышения степени легирования. При $N_D \leq 10^{17} \text{ см}^{-3}$ ток обусловлен термоэмиссией и R_c почти не зависит от легирования. На рис. 43 приведены также экспериментальные результаты для диодов PtSi—Si и Al—Si. Видно, что они находятся в хорошем соответствии с расчетными.

Таблица 5. Технология изготовления омических контактов к полупроводникам типа $A^{III}B^V$

Полупроводник типа $A^{III}B^V$	E_g , эВ	Тип	Материал контакта	Технология изготовления	Температура сплавления, °C
AlN	5,9	Полу-изолятор	Si	Формовка	
		То же	Al, Al—In	»	1500—1800
AlP	2,45	»	Mo, W	Распыление	1000
AlAs	2,16	<i>n</i>	Ga—Ag	Формовка	500—1000
		<i>n, p</i>	In—Te	»	150
		<i>n, p</i>	Au	»	160
		<i>n, p</i>	Au—Ge	»	700
		<i>n</i>	Au—Sn	»	
GaN	3,36	Полу-изолятор	Al—In	»	
GaP	2,26	<i>p</i>	Au—Zn (99 : 1)	»	700
		<i>p</i>	Au—Ge	Напыление	700
		<i>n</i>	Au—Si	Формовка	»
		<i>n</i>	(62 : 38)	»	360
		<i>n</i>	Au—Si (98 : 2)	Напыление	700
GaAs	1,42	<i>p</i>	Au—Zn (99 : 1)	Электролиз, напыление	
		<i>p</i>	In—Au (80 : 20)	Формовка	
		<i>n</i>	Au—Ge (88 : 12)	Напыление	
		<i>n</i>	In—Au (90 : 10)	Напыление	350—450
		<i>n</i>	Au—Si (94 : 6)	»	550
		<i>n</i>	Au—Sn (90 : 10)	»	300
		<i>n</i>	Au—Te (98 : 2)	»	350—700
GaSb	0,72	<i>p</i>	In	Формовка	500
		<i>n</i>	In	»	
InP	1,35	<i>p</i>	In	»	
		<i>n</i>	In, In—Te	»	350—600
		<i>n</i>	Ag—Sn	»	350—600
InAs	0,36	<i>n</i>	In	Напыление	600
			Sn—Te (99 : 1)	Формовка	
				»	

Продолжение табл. 5

Полупроводник типа A ^{III} B ^V	E_g , эВ	Тип	Материал контакта	Технология изготовления	Температура сплавления, °С	
InSb	0,17	<i>n</i>	In	Формовка		
		<i>n</i>	Sn—Te (99 : 1)	»		
	1,42—2,31	<i>p</i>	Au—Zn	Напыление		500
		<i>p</i>	Al	»		500
		<i>n</i>	Au—Ge—Ni	»		450
Al _x Ga _{1-x} As	1,42—2,16	<i>n</i>	Au—Sn	»	450	
		<i>p</i>	Au—In	Анодирование	400—450	
		<i>p</i>	Au—Zn	Напыление		
		<i>p</i>	Al	»	500	
		<i>n</i>	Au—Ge—Ni	»	500	
		<i>n</i>	Au—Sn	»	450—485	
					Электролиз	450
Ga _{1-x} In _x Sb	0,70—0,17	<i>n</i>	Au—Si	Напыление		
Al _x Ga _{1-x} P	2,31—2,45	<i>n</i>	Sn—Te	Напыление		
Ga _{1-x} In _x As	1,47—2,35	<i>n</i>	Sn	Формовка		
InAs _x Sb _{1-x}	0,17—0,35	<i>n</i>	In—Te	»		

Из анализа рис. 43 следует, что для получения малых R_c нужна либо высокая степень легирования, либо малая высота барьера (либо то и другое вместе). Именно из этих соображений исходят при изготовлении омических контактов (рис. 44).

На широкозонных полупроводниках обычно очень трудно изготовить контакт с малой высотой барьера. Кроме того, используемые металлы не всегда имеют достаточно малую работу выхода. В таких случаях для изготовления омических контактов создают дополнительный высоколегированный слой на поверхности полупроводника. Создать такой слой можно различными методами: мелкой диффузией, перекристаллизацией, диффузией одной из компонент материала контакта, двойной эпитаксией, ионной имплантацией. Для изготовления омических контактов к Ge и Si *n*-типа сначала напыляют сплав Au—Sb (с 0,1 % Sb). Затем при соответствующей эвтектической температуре сплавление этот контакт в полупроводник в атмосфере инертного газа (такого, как аргон или азот) [69]. Для изготовления омических контактов к GaAs и другим полупроводникам типа A^{III}B^V разработаны различные методы [70]. Они представлены в табл. 5.

ЛИТЕРАТУРА

1. Braun F. Über die Stromleitung durch Schwefelmetalle, *Ann. Phys. Chem.*, **153**, 556 (1874).
2. Bosc J. C. U. S. Patent 775,840 (1904).

3. Wilson A. H. The Theory of Electronic Semiconductors, *Proc. R. Soc. Lond. Ser. A*, **133**, 458 (1931).
4. Schottky W. Halbleitertheorie der Sperrschicht, *Naturwissenschaften*, **26**, 843 (1938).
5. Mott N. F. Note on the Contact between a Metal and an Insulator or Semiconductor, *Proc. Camb. Philos. Soc.*, **34**, 568 (1938).
6. Henisch H. K. Rectifying Semiconductor Contacts, Clarendon, Oxford, 1957.
7. Rhoderick E. H. Metal-Semiconductor Contacts, Clarendon, Oxford, 1978;
8. Transport Processes in Schottky Diodes, in Pepper K. M., Ed., *Inst. Phys. Conf. Ser.*, No. 22, Institute of Physics, Manchester, England, 1974, p. 3.
9. Rideout V. L. A Review of the Theory, Technology and Applications of Metal-Semiconductor Rectifiers, *Thin Solid Films*, **48**, 261 (1978).
10. Michaelson H. B. Relation between an Atomic Electronegativity Scale and the Work Function, *IBM J. Res. Dev.*, **22**, 72 (1978).
11. Sze S. M., Crowell C. R., Kahng D. Photoelectric Determination of the Image Force Dielectric Constant for Hot Electrons in Schottky Barriers, *J. Appl. Phys.*, **35**, 2534 (1964).
12. Salzberg C. D., Villa G. G. Infrared Refractive Indexes of Silicon Germanium and Modified Selenium Glass, *J. Opt. Soc. Am.*, **47**, 244 (1975).
13. Bethe H. A. Theory of the Boundary Layer of Crystal Rectifiers, *MIT Radiat. Lab. Rep.* 43—12 (1942).
14. Crowell C. R. The Richardson Constant for Thermionic Emission in Schottky Barrier Diodes, *Solid State Electron.*, **8**, 395 (1965).
15. Crowell C. R., Sze S. M. Current Transport in Metal — Semiconductor Barriers, *Solid State Electron.*, **9**, 1035 (1966).
16. Crowell C. R., Sze S. M. Electron — Optical — Phonon Scattering in the Emitter and Collector Barriers of Semiconductor — Metal — Semiconductor Structures, *Solid State Electron.*, **8**, 979 (1965).
17. Kao C. W., Anderson L., Crowell C. R. Photoelectron Injection at Metal — Semiconductor Interface, *Surface Sci.*, **95**, 321 (1980).
18. Crowell C. R., Sze S. M. Quantum-Mechanical Reflection of Electrons at Metal — Semiconductor Barriers: Electron Transport in Semiconductor — Metal — Semiconductor Structures, *J. Appl. Phys.*, **37**, 2685 (1966).
19. Chang C. Y., Sze S. M. Carrier Transport across Metal — Semiconductor Barriers, *Solid State Electron.*, **13**, 727 (1970).
20. Andrews J. M., Lepseter M. P. Reverse Current — Voltage Characteristics of Metal-Silicide Schottky Diodes, *Solid State Electron.*, **13**, 1011 (1970).
21. Crowell C. R., Beguwalla M. Recombination Velocity Effects on Current Diffusion and Inefficiency in Schottky Barriers, *Solid State Electron.*, **14**, 1149 (1971).
22. Padovani F. A., Stratton R. Field and Thermionic-Field Emission in Schottky Barriers, *Solid State Electron.*, **9**, 695 (1966).
23. Yu A. Y. C., Snow E. H. Minority Carrier Injection of Metal — Silicon Contacts, *Solid State Electron.*, **12**, 155 (1969).
24. Scharfetter D. L. Minority Carrier Injection and Charge Storage in Epitaxial Schottky Barrier Diodes, *Solid State Electron.*, **8**, 299 (1965).
25. Cowley A. M., Sze S. M. Surface States and Barrier Height of Metal — Semiconductor Systems, *J. Appl. Phys.*, **36**, 3212 (1965).
26. Bardeen J. Surface States and Rectification at a Metal Semiconductor Contact, *Phys. Rev.*, **71**, 717 (1974).
27. Mead C. A., Spitzer W. G. Fermi-Level Position at Metal-Semiconductor Interfaces, *Phys. Rev.*, **134**, A713 (1964).
28. Pugh D. Surface States on the {111} Surface of Diamond, *Phys. Rev. Lett.*, **12**, 390 (1964).
29. Pauling L. The Nature of the Chemical Bond., 3rd ed., Cornell University Press, Ithaca, N. Y., 1960.
30. Kurtin S., McGill T. C., Mead C. A. Fundamental Transition in Electronic Nature of Solids, *Phys. Rev. Lett.*, **22**, 1433 (1969).

30. Mead C. A. Metal-Semiconductor Surface Barriers, *Solid State Electron.*, **9**, 1023 (1966).
31. Crowell C. R., Sarace J. C., Sze S. M. Tungsten-Semiconductor Schottky-Barrier Diodes, *Trans. Met. Soc. AIME*, **233**, 478 (1965).
32. Lepselter M. P., Sze S. M. Silicon Schottky Barrier Diode with Near-Ideal I—V Characteristics, *Bell. Syst. Tech. J.*, **47**, 195 (1968).
33. Andrews J. M., Korch F. B. Formation of NiSi and Current Transport across the NiSi — Si Interface, *Solid State Electron.*, **14**, 901 (1971).
34. Goodman A. M. Metal — Semiconductor Barrier Height Measurement by the Differential Capacitance Method — One Carrier System, *J. Appl. Phys.*, **34**, 329 (1963).
35. Roberts G. I., Crowell C. R. Capacitive Effects of Au and Cu Impurity Levels in Pt *n*-type Si Schottky Barriers, *Solid State Electron.*, **16**, 29 (1973).
36. Crowell C. R., Spitzer W. G., Howarth L. E., Labate E. Attenuation Length Measurements of Hot Electrons in Metal Films, *Phys. Rev.*, **127**, 2006 (1962).
37. Fowler R. H. The Analysis of Photoelectric Sensitivity Curves for Clean Metals at Various Temperatures, *Phys. Rev.*, **38**, 45 (1931).
38. Crowell C. R., Sze S. M., Spitzer W. G. Equality of the Temperature Dependence of the Gold — Silicon Surface Barrier and the Silicon Energy Gap in Au *n*-type Si Diodes, *Appl. Phys. Lett.*, **4**, 91 (1964).
39. Beguwala M., Crowell C. R. Characterization of Multiple Deep Level Systems in Semiconductor Junctions by Admittance Measurements, *Solid State Electron.*, **17**, 203 (1974).
40. McCaldin J. O., McGill T. C., Mead C. A. Schottky Barriers on Compound Semiconductors: The Role of the Anion, *J. Vac. Sci. Technol.*, **13**, 802 (1976).
41. Andrews J. M. The Role of the Metal — Semiconductor Interface in Silicon Integrated Circuit Technology, *J. Vac. Sci. Technol.*, **11**, 972 (1974).
42. Andrews J. M., Phillips J. C. Chemical Bonding and Structure of Metal — Semiconductor Interfaces, *Phys. Rev. Lett.*, **35**, 56 (1975).
43. van Gorp G. J. The Growth of Metal Silicide Layers on Silicon, in Huff H. R., Sirtl E., Eds., *Semiconductor Silicon, 1977*, Electrochemical Society, Princeton, N. J., 1977, p. 342.
- 43a. Ohdomari I., Tu K. N., d'Heurle F. M., Kuan T. S., Petersson S. Schottky-Barrier Height of Iridium Silicide, *Appl. Phys. Lett.*, **33**, 1028 (1978).
44. Saltich J. L., Terry L. E. Effects of Pre- and Post-Annealing Treatments on Silicon Schottky Barrier Diodes, *Proc. IEEE*, **58**, 492 (1970).
45. Sinha A. K. Electrical Characteristics and Thermal Stability of Platinum Silicide-to-Silicon Ohmic Contacts Metalized with Tungsten, *J. Electrochem. Soc.*, **120**, 1767 (1973).
46. Sinha A. K., Smith T. E., Read M. H., Poate J. M. *n*-GaAs Schottky Diodes Metalized with Ti and Pt/Ti, *Solid State Electron.*, **19**, 489 (1976).
47. Chino K. Behavior of Al — Si Schottky Barrier Diodes under Heat Treatment, *Solid State Electron.*, **16**, 119 (1973).
48. Itoh Y., Hashimoto N. Reaction-Process Dependence of Barrier Height between Tungsten Silicide and *n*-Type Silicon, *J. Appl. Phys.*, **40**, 425 (1969).
49. Shannon J. M. Reducing the Effective Height of a Schottky Barrier Using Low-Energy Ion Implantation, *Appl. Phys. Lett.*, **24**, 369 (1974).
50. Shannon J. M. Increasing the Effective Height of a Schottky Barrier Using Low-Energy Ion Implantation, *Appl. Phys. Lett.*, **25**, 75 (1974).
51. Andrews J. M., Ryder R. M., Sze S. M. Schottky Barrier Diode Contacts, U. S. Patent 3964084 (1976).
52. Shannon J. M. Control of Schottky Barrier Height Using Highly Doped Surface Layers, *Solid State Electron.*, **19**, 537 (1976).
53. For general references on vacuum deposition, see Holland L. *Vacuum Deposition of Thin Films*, Chapman & Hall, London, 1966; Roth A. *Vacuum Technology*, North-Holland, Amsterdam, 1976.

54. Honig R. E. Vapor Pressure Data for the Solid and Liquid Elements, *RCA Rev.*, **23**, 567 (1962).
55. Young D. T., Irvin J. C. Millimeter Frequency Conversion Using Au-n-type GaAs Schottky Barrier Epitaxy Diode with a Novel Contacting Technique, *Proc. IEEE*, **53**, 2130 (1965); Kahng D., Ryder R. M. Small Area Semiconductor Devices, U. S. Patent 3360851 (1968).
- 55a. Vanderwal N. C. A Microwave Schottky-Barrier Varistor Using GaAs for Low Series Resistance, IEEE Int. Electron Device Meet., Washington, D. C., Oct. 18–20, 1967.
56. Irvin J. C., Vanderwal N. C. Schottky-Barrier Devices, in Watson H. A., Ed., *Microwave Semiconductor Devices and Their Circuit Applications*, McGraw-Hill, N. Y., 1968.
57. McColl M., Millea M. F. Advantages of Mott Barrier Mixer Diodes, *Proc. IEEE*, **61**, 499 (1973).
58. Yu A. Y. C., Mead C. A. Characteristics of Al — Si Schottky Barrier Diode, *Solid State Electron.*, **13**, 97 (1970).
59. Baker R. H. Maximum Efficiency Switching Circuit, MIT Lincoln Lab. Rep. TR-110, Lexington, Mass., 1956.
60. Tada K., Laraya J. L. R. Reduction of the Storage Time of a Transistor Using a Schottky-Barrier Diode, *Proc. IEEE*, **55**, 2064 (1967).
61. Saltich J. L., Clark L. E. Use of a Double Diffused Guard Ring to Obtain Near Ideal I—V Characteristics in Schottky-Barrier Diodes, *Solid State Electron.*, **13**, 857 (1970).
62. Linden K. J. GaAs Schottky Mixer Diode with Integral Guard Layer Structure, *IEEE Trans. Electron Devices*, **ED-23**, 363 (1976).
63. Rhee C., Saltich J. K., Zwernemann R. Moat-Etched Schottky Barrier Diode Displaying Near Ideal I—V Characteristics, *Solid State Electron.*, **15**, 1181 (1972).
64. Anantha N. G., Ashar K. G., *IBM. J. Res. Dev.*, **15**, 442 (1971).
65. Rusu A., Bulucea C., Postolache C. The Metal-Ovrlap-Laterally-Diffused (MOLD) Schottky Diode, *Solid State Electron.*, **20**, 499 (1977).
66. Colemann D. J., Jr., Irvin J. C., Sze S. M. GaAs Schottky Diodes with Near-Ideal Characteristics, *Proc. IEEE*, **59**, 1121 (1971).
67. Chang C. Y., Fang Y. K., Sze S. M. Specific Contact Resistance of Metal — Semiconductor Barriers, *Solid State Electron.*, **14**, 541 (1971).
68. Yu A. Y. C. Electron Tunneling and Contact Resistance of Metal — Silicon Contact Barriers, *Solid State Electron.*, **13**, 239 (1970).
69. Lepselter M. P., Andrews J. M. Ohmic Contacts to Silicon, in Schwartz B., Ed., *Ohmic Contacts to Semiconductors*, The Electrochemical Society Symposium Series, N. Y., 1969, p. 159.
70. Rideout V. L. A Review of the Theory and Technology for Ohmic Contacts to Group III—V Compound Semiconductors, *Solid State Electron.*, **18**, 541 (1975).
71. Andrews J. M. Extended Abstracts, Electrochem. Soc. Spring Meet. Abstr. 191, 1975, p. 452.
72. Spicer W. E., Chye P. W., Garner C. M., Lindau I., Pianetta P. The Surface Electronic Structure of III—V Compounds and the Mechanism of Fermi Level Pinning by Oxygen (Passivation) and Metals (Schottky Barriers), *Surface Sci.*, **86**, 763 (1979).
73. Spicer W. E., Lindau I., Skeath P., Su C. Y., Chye P. Unified Mechanism for Schottky-Barrier Formation and III—V Oxide Interface States, *Phys. Rev. Lett.*, in press (1981).

Глава 6

ПОЛЕВЫЕ ТРАНЗИСТОРЫ С $p-n$ -ПЕРЕХОДОМ В КАЧЕСТВЕ ЗАТВОРА И ПОЛЕВЫЕ ТРАНЗИСТОРЫ ТИПА МЕТАЛЛ — ПОЛУПРОВОДНИК¹

6.1. ВВЕДЕНИЕ

Полевой транзистор с $p-n$ -переходом в качестве затвора впервые предложил и проанализировал Шокли в 1952 г. [1]. Этот прибор выполняет функцию резистора, управляемого напряжением. Поскольку проводимость полевого транзистора с $p-n$ -переходом в качестве затвора определяется главным образом основными носителями, его также называют «униполярным» (в отличие от биполярных транзисторов). Первый униполярный транзистор, который работал по принципу, предложенному Шокли, продемонстрировали Дейки и Росс [2]. Впоследствии они учли влияние полевой зависимости подвижности на характеристики полевого транзистора с $p-n$ -переходом [3].

Полевой транзистор типа металл—полупроводник (МП-транзистор) был предложен в 1966 г. [4], а затем был изготовлен на GaAs, эпитаксиально выращенном на полуизолирующей GaAs-подложке [5]. Принцип действия МП-транзистора идентичен принципу работы полевого транзистора с $p-n$ -переходом в качестве затвора. Различие состоит лишь в том, что в МП-транзисторе в качестве затвора использован выпрямляющий контакт металл—полупроводник (вместо $p-n$ -перехода). МП-транзисторы обладают технологическими и некоторыми другими достоинствами. Так, например, при изготовлении МП-транзисторов используются сравнительно низкотемпературные технологические процессы, в то время как для создания полевого транзистора с $p-n$ -переходом требуется высокотемпературная диффузия. Отметим также малое сопротивление и малое падение напряжения вдоль канала МП-транзистора. Кроме того, в МП-транзисторах оказываются допустимыми более высокие уровни рассеиваемой мощности (металлический затвор является хорошим теплоотводом). Однако полевые транзисторы с $p-n$ -переходом допускают применение некоторых специальных затворных структур (таких, как гетеропереходы, затворы с буферным слоем и т. д.) для улучшения высокочастотных характеристик.

На рис. 1 показано «генеалогическое» дерево полевых транзисторов. Кроме рассматриваемых в настоящей главе полевых

¹ В англоязычной литературе эти приборы называются — junction field-effect transistor (JFET) и metal — semiconductor field-effect transistor (MESFET).— *Прим. перев.*

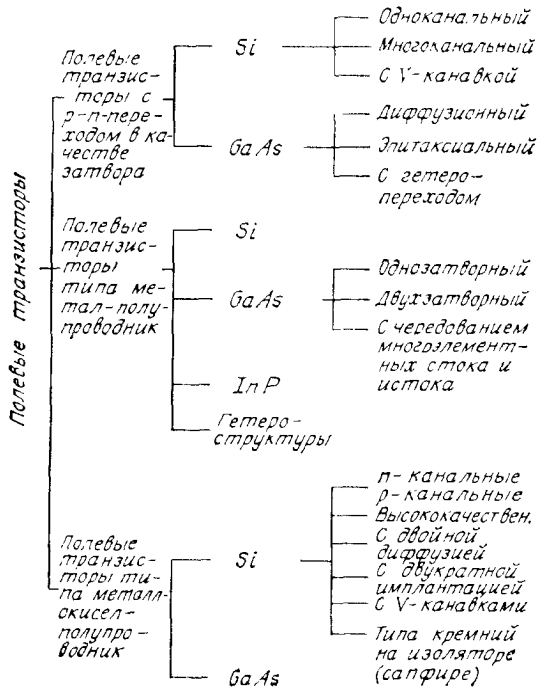


Рис. 1. «Генеалогическое» дерево полевых транзисторов.

транзисторов с $p-n$ -переходом и МП-транзисторов к этому классу приборов относятся и МОП-транзисторы (металл—окисел—полупроводник), свойства которых рассмотрены в гл. 8. В настоящее время относящиеся к этим трем классам полевых транзисторов разнообразные приборы на всевозможных полупроводниковых материалах довольно хорошо изучены. Для более подробного ознакомления с рабочими характеристиками, шумовыми свойствами и СВЧ-модификациями полевых транзисторов можно рекомендовать работы [6—8].

Полевые транзисторы обладают достоинствами при их использовании в аналоговых переключателях, усилителях с высокоомным входом, СВЧ-усилителях и интегральных схемах. Они имеют существенно более высокое входное сопротивление, чем биполярные транзисторы, что облегчает их сопряжение со стандартными СВЧ-устройствами. В области высоких токов полевые транзисторы имеют отрицательный температурный коэффициент, т. е. ток в данных приборах уменьшается с повышением температуры. Благодаря этому возникает более однородное распределение температуры по площади прибора и снижается вероятность развития

теплового или вторичного пробоя, характерного для биполярных транзисторов. Высокая термостабильность наблюдается даже в полевых транзисторах с большой активной площадью канала, а также в схемах, в которых большое число этих приборов включено параллельно. Поскольку полевые транзисторы являются униполярными приборами, они не чувствительны к эффектам накопления неосновных носителей, и поэтому имеют более высокие граничные частоты и скорости переключения. Кроме того, поскольку характеристики полевых транзисторов линейны либо квадратичны (а не экспоненциальны), они оказываются значительно менее чувствительными к перекрестным наводкам, чем биполярные транзисторы.

6.2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ПРИБОРОВ

Полевой транзистор с p - n -переходом в качестве затвора показан на рис. 2. Он представляет собой проводящий канал с двумя омическими контактами — стоком и истоком. Когда на сток подано положительное напряжение V_D относительно истока, электроны в канале перемещаются от истока к стоку. Третий электрод структуры — затвор — образует выпрямляющий p - n -переход с каналом. Очевидно, что сопротивление канала будет изменяться с изменением толщины обедненных слоев, распространяющихся в канал, поэтому рассматриваемый полевой транзистор представ-

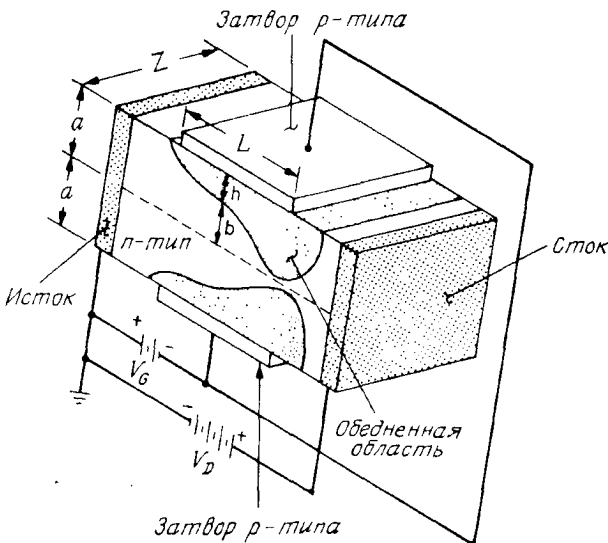


Рис. 2. Полевой транзистор с p - n -переходом в качестве затвора (модель Шокли) [2].

ляет собой резистор, управляемый напряжением на затворе. На рис. 2 указаны также основные размеры, характеризующие эту структуру: длина канала (или длина затвора) L , его ширина Z и глубина a . Кроме того, показаны локальная ширина обедненного слоя h и соответствующая локальная глубина проводящего канала b . Полярность приложенных напряжений соответствует рассматриваемому n -канальному полевому транзистору. Для p -канального прибора полярность питающих напряжений должна быть противоположной. Обычно исток полевого транзистора заземляют и напряжения затвора и стока отсчитывают по отношению к заземленному истоку.

В отсутствие напряжений смещения ($V_G = V_D = 0$) прибор находится в термодинамическом равновесии, и все токи равны нулю. При фиксированном напряжении смещения на затворе V_G (нулевом или отрицательном) ток канала увеличивается с ростом напряжения на стоке V_D до тех пор, пока при некотором сравнительно большом напряжении $V_D = V_{D\text{ sat}}$ не происходит насыщения тока $I_D = I_{D\text{ sat}}$. Типичные вольт-амперные характеристики полевого транзистора с p - n -переходом приведены на рис. 3. На этих характеристиках следует различать три области: линейную (при малых напряжениях на стоке), где ток стока I_D пропорционален V_D ; область насыщения, где ток стока $I_D = I_{D\text{ sat}}$ не зависит от напряжения стока; область пробоя, где ток стока стремительно увеличивается для сравнительно небольших приращений напряжения ΔV_D . При увеличении отрицательного на-

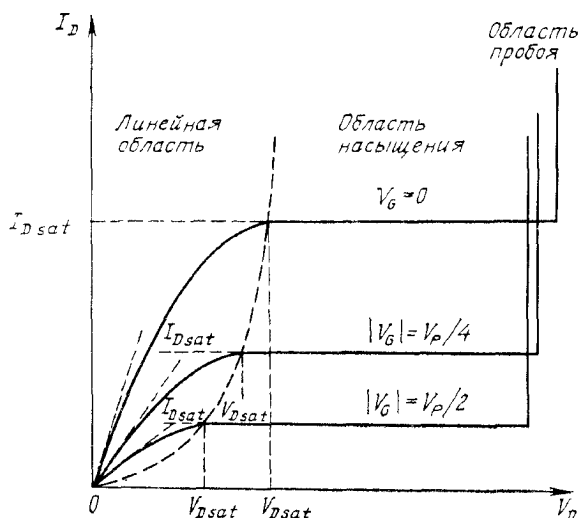


Рис. 3. Вольт-амперные характеристики полевого транзистора с p - n -переходом.

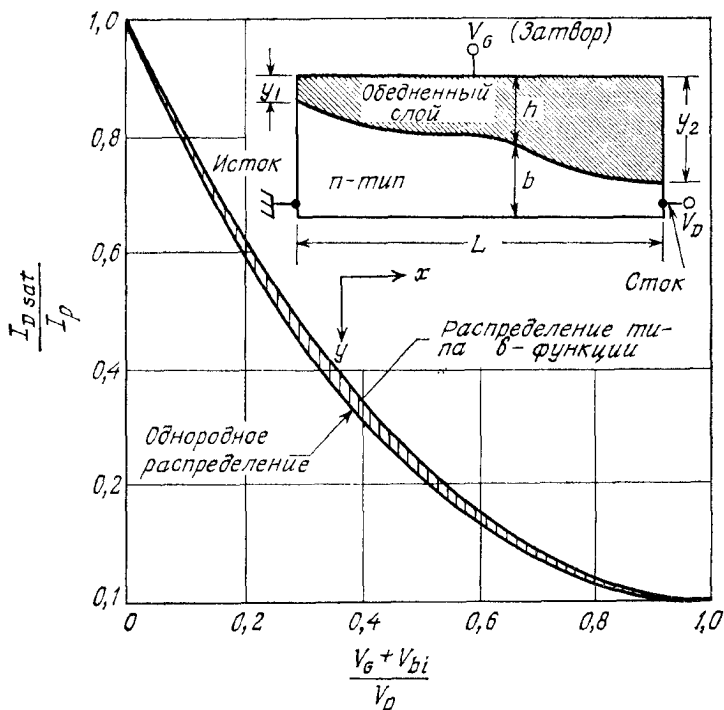


Рис. 4. Передаточные характеристики длинноканального полевого транзистора для двух предельных распределений легирующей примеси в канале. На вставке показано поперечное сечение верхней половины прибора (y_1 и y_2 — ширина обедненного слоя у истока и стока, n — локальная толщина обедненного слоя) [9, 10].

пряжения смещения на затворе V_G ток насыщения $I_{D\text{ sat}}$ и напряжение, соответствующее началу насыщения $V_{D\text{ sat}}$, уменьшаются. Это обусловлено понижением начальной глубины проводящего канала, что в свою очередь приводит к большему начальному (при $V_D = 0$) сопротивлению канала транзистора.

Проведём теперь анализ вольт-амперных характеристик длинноканального ($L \gg a$) полевого транзистора с p - n -переходом, воспользовавшись при этом следующими предположениями:

- 1) приближением плавного канала; 2) приближением резкого края обедненного слоя; 3) независимостью подвижности носителей тока от электрического поля. Более общий анализ, включающий короткоканальные и двумерные эффекты, проведен в разд. 6.3. Учитывая симметрию прибора (рис. 2), рассмотрим только верхнюю половину полевого транзистора (рис. 4, вставка).

6.2.1. Однородно легированный канал

Приближенное плавного канала состоит в том, что для распределения потенциала в обедненном слое можно записать одномерное уравнение Пуассона

$$-\frac{d^2V}{dy^2} = \frac{d\mathcal{E}_y}{dy} = \frac{\rho(y)}{\epsilon_s}, \quad (1)$$

которое для однородно легированного канала в приближении резкой границы обедненного слоя имеет вид

$$-\frac{d^2V}{dy^2} = \frac{qN_D}{\epsilon_s}. \quad (1a)$$

Здесь \mathcal{E}_y — поперечное электрическое поле (в направлении y). Отсюда для локальной ширины обедненного слоя h находим

$$h = \{2\epsilon_s [V(x) + V_G + V_{bi}/qN_D]\}^{1/2}, \quad (2)$$

где V_{bi} — встроенный потенциал p^+ - n -перехода (контактная разность потенциалов), равный $(kT/q) \ln(N_D/n_i)$, а $V(x)$ — локальное значение потенциала в сечении проводящего канала на расстоянии x от истока. Отметим, что в n -канальных приборах напряжение на затворе отрицательно по отношению к истоку, так что V_G в выражении (2) и последующих выражениях обозначает абсолютное значение напряжения на затворе. В соответствии с формулой (2) ширина обедненного слоя на границах канала (у стока и истока) определяется выражением

$$\begin{aligned} y_1 &= [2\epsilon_s(V_G + V_{bi})/qN_D]^{1/2} \text{ при } x = 0, \\ y_2 &= [2\epsilon_s(V_D + V_G + V_{bi})/qN_D]^{1/2} \text{ при } x = L. \end{aligned} \quad (3)$$

Когда величина y_2 становится равной a , происходит смыкание обедненных областей от верхнего и нижнего затвора у стока (отсечка канала). Из этого условия для напряжения отсечки канала (начало области насыщения) получим

$$V_P = V(y_2 = a) \equiv qN_D a^2 / 2\epsilon_s. \quad (4)$$

Воспользовавшись дифференциальным законом Ома, для плотности тока J_x в проводящем канале запишем

$$J_x = \sigma(x) \mathcal{E}_x. \quad (5)$$

Для случая однородного легирования имеем

$$J_x = qN_D \mu \mathcal{E}_x. \quad (5a)$$

Здесь $\sigma(x)$ — проводимость, μ — подвижность электронов, которая предполагается не зависящей от электрического поля, а $\mathcal{E}_x = -dV/dx$ — продольное электрическое поле в канале. Полный

ток в рассматриваемой верхней половине канала определяется выражением

$$I_D = qN_D\mu \left(\frac{dV}{dx} \right) (a - h) Z \quad (6)$$

или
$$I_D dx = Z\mu qN_D (a - h) dV. \quad (6a)$$

Выразив с помощью соотношения (2) дифференциал напряжения dV :

$$dV = \frac{qN_D}{\epsilon_s} h dh, \quad (7)$$

после интегрирования выражения (6a) от $x = 0$ до $x = L$ получим

$$\begin{aligned} I_D &= \frac{1}{L} \int_{y_1}^{y_2} Z\mu qN_D (a - h) \frac{qN_D}{\epsilon_s} h dh = \\ &= \frac{Z\mu q^2 N_D^2 a^3}{6\epsilon_s L} \left[\frac{3}{a^2} (y_2^2 - y_1^2) - \frac{2}{a^3} (y_2^3 - y_1^3) \right]. \end{aligned} \quad (8)$$

Обозначим множитель перед квадратной скобкой символом

$$I_P \equiv Z\mu q^2 N_D^2 a^3 / 6\epsilon_s L \quad (9)$$

и назовем его током отсечки канала. Введя безразмерные длины

$$\begin{aligned} u &\equiv h/a = [(V + V_G + V_{bi})/V_P]^{1/2}, \\ u_1 &\equiv y_1/a = [(V_G + V_{bi})/V_P]^{1/2}, \\ u_2 &\equiv y_2/a = [(V_D + V_G + V_{bi})/V_P]^{1/2}, \end{aligned} \quad (10)$$

выражение (8) запишем в виде

$$I_D = I_P [3(u_2^2 - u_1^2) - 2(u_2^3 - u_1^3)], \quad (11)$$

или

$$\begin{aligned} I_D &= I_P \{3V_D/V_P - 2[(V_D + V_G + V_{bi})^{3/2} - \\ &\quad - (V_G + V_{bi})^{3/2}]/V_P^{3/2}\}. \end{aligned} \quad (11a)$$

При фиксированном напряжении на затворе V_G максимальное значение тока (ток насыщения $I_{D \text{ sat}}$) соответствует отсечке канала. Подставив в выражение (11) $u_2 = 1$ ($y_2 = a$), получим

$$\begin{aligned} I_{D \text{ sat}} &= I_P (1 - 3u_1^2 + 2u_1^3) = \\ &= I_P \left[1 - 3 \left(\frac{V_G + V_{bi}}{V_P} \right) + 2 \left(\frac{V_G + V_{bi}}{V_P} \right)^{3/2} \right]. \end{aligned} \quad (12)$$

Вольт-амперные характеристики полевого транзистора с $p-n$ -переходом, рассчитанные по формуле (11a), приведены на рис. 3.

Величина тока насыщения $I_{D \text{ sat}}$ на этих графиках определяется выражением (12), а напряжение, соответствующее началу насыщения:

$$V_{D \text{ sat}} = V_P - V_G - V_{bi} = \frac{qN_D a^2}{2\epsilon_s} - V_G - \frac{kT}{q} \ln\left(\frac{N_D}{n_i}\right). \quad (13)$$

При дальнейшем увеличении напряжения на стоке транзистора ток стока изменяется слабо и остается примерно равным току насыщения до тех пор, пока не начнется лавинный пробой $p^+ - n$ -диода затвор—канал, после чего ток стока резко возрастает при увеличении напряжения. Отметим, что лавинный пробой происходит вблизи стока, где обратное смещение диода затвор—канал максимально:

$$V_B = V_D + V_G, \quad (14)$$

или

$$V_D = V_B - V_G.$$

Как уже говорилось выше, V_G обозначает абсолютную величину отрицательного напряжения на затворе. Если $V_G = 0$, лавинный пробой происходит при напряжении на стоке, равном V_B . Если $V_G > 0$, пробой диода затвор—канал наступает при том же напряжении пробоя V_B , но требуемое теперь для этого напряжение на стоке оказывается равным $V_B - V_G$.

Из уравнения (11) можно получить два важных параметра: крутизну

$$g_m \equiv \frac{\partial I_D}{\partial V_G} = \frac{2Z\mu q N_D}{L} (y_2 - y_1) \quad (15a)$$

и проводимость канала, которую иногда называют проводимостью стока:

$$g_D \equiv \frac{\partial I_D}{\partial V_D} = \frac{2Z\mu q N_D}{L} (a - y_2). \quad (15b)$$

В линейной области характеристики ($V_D \rightarrow 0$) проводимость канала определяется выражением

$$g_{D0}(V_D \rightarrow 0) = g_{\text{макс}} \left(1 - \sqrt{\frac{V_G + V_{bi}}{V_P}}\right), \quad (16)$$

где $g_{\text{макс}} \equiv qN_D a \mu Z/L$. В области насыщения крутизна g_m равна

$$g_m = \frac{\partial I_{D \text{ sat}}}{\partial V_G} = g_{\text{макс}} (1 - u_1) = g_{\text{макс}} \left(1 - \sqrt{\frac{V_G + V_{bi}}{V_P}}\right). \quad (17)$$

Отметим полную идентичность выражений (16) и (17).

6.2.2. Канал с произвольным профилем легирования [9]

Для полевых транзисторов с произвольным профилем легирования канала определим поверхностную плотность заряда выражением

$$Q(Y) \equiv \int_0^Y \rho(y) dy \quad [\text{Кл/см}^2] \quad (18)$$

или

$$Q(h) \equiv \int_0^h \rho(y) dy \quad [\text{Кл/см}^2], \quad (18a)$$

где $\rho(y)$ — объемная плотность заряда, Кл/см³. Зависимость локального напряжения обратного смещения диода затвор—канал $V(h)$ от h и $\rho(y)$ можно определить из уравнения Пуассона (1). Интегрируя последнее от $y = 0$ до $y = h$, получим

$$\mathcal{E}_y \equiv -\frac{\partial V}{\partial y} = \frac{1}{\epsilon_s} \int_0^y \rho(y) dy + \text{const.} \quad (19)$$

Постоянная интегрирования в выражении (19) определяется граничным условием на краю обедненного слоя: $\mathcal{E}_y = 0$ при $y = h$. Отсюда получим

$$\frac{\partial V}{\partial y} = \frac{1}{\epsilon_s} \left[\int_0^h \rho(y) dy - \int_0^y \rho(y) dy \right] = \frac{1}{\epsilon_s} [Q(h) - Q(y)]. \quad (20)$$

Интегрируя снова от $y = 0$ до $y = h$, найдем

$$V(h) = \frac{1}{\epsilon_s} \left[Q(h) \int_0^h dy - \int_0^h Q(y) dy \right] = \frac{1}{\epsilon_s} \left[hQ(h) - \int_0^h Q(y) dy \right] \quad (21)$$

или после интегрирования по частям

$$V(h) = \frac{1}{\epsilon_s} \int_0^h y\rho(y) dy. \quad (22)$$

Напомним, что напряжение $V(h)$ наряду с приложенными напряжениями смещения содержит контактную разность V_{bi} . Положив в выражении (22) величину h равной a , получим напряжение отсечки канала

$$V_P = V(h = a) = \frac{1}{\epsilon_s} \int_0^a y\rho(y) dy. \quad (23)$$

Дифференцируя выражение (22), получим соотношение

$$dV/dh = h\rho(h)/\epsilon_s, \quad (24)$$

из которого следует, что приращение напряжения, необходимое для определенного уширения обедненного слоя, пропорционально ширине этого слоя и объемной плотности заряда на его границе. Для локальной удельной емкости перехода (на единицу площади) получим

$$C \equiv dQ(h)/dV = \left(\frac{dQ}{dh}\right) \left(\frac{dh}{dV}\right) \quad [\Phi \cdot \text{см}^{-2}]. \quad (25)$$

Следовательно, обедненный слой можно рассматривать как плоский конденсатор с межэлектродным зазором, равным h , поскольку его дифференциальная емкость не зависит от профиля распределения заряда внутри обедненной области.

Найдем теперь выражение для вольт-амперной характеристики полевого транзистора и его крутизны. Из дифференциального закона Ома (5) для полного тока вместо выражения (6) следует написать

$$I_D = Z\mu \frac{dV}{dx} \int_h^a \rho(y) dy \quad (26)$$

или

$$I_D dx = Z\mu (dV/dh) dh \int_h^a \rho(y) dy. \quad (26a)$$

Подставив сюда выражение (24) и интегрируя по x с граничными условиями $h = y_1$ при $x = 0$ и $h = y_2$ при $x = L$, получим

$$\int_0^L I_D dx = I_D L = \frac{2Z\mu}{\epsilon_s} \int_0^L h\rho(h) dh \int_h^a \rho(y) dy \quad (27)$$

или

$$I_D = \frac{2Z\mu}{\epsilon_s L} \int_{y_1}^{y_2} [Q(a) - Q(h)] h\rho(h) dh. \quad (28)$$

Это есть основное уравнение, определяющее вольт-амперные характеристики длинноканального полевого транзистора с произвольным профилем легирования в канале.

Дифференцируя выражение (28), получим уравнение для крутизны

$$g_m \equiv \frac{\partial I_D}{\partial V_G} = \frac{\partial I_D}{\partial y_1} \frac{\partial y_1}{\partial V_G} + \frac{\partial I_D}{\partial y_2} \frac{\partial y_2}{\partial V_G}. \quad (29)$$

Вычислив соответствующие частные производные из выражений (24) и (28), найдем

$$g_m = \frac{2Z\mu}{L} [Q(y_2) - Q(y_1)]. \quad (29a)$$

Отметим, что, согласно выражению (29a), крутизна транзистора равна проводимости прямоугольной части канала, ограниченной в поперечном направлении плоскостями $y = y_1$ и $y = y_2$. Аналогичным образом из выражений (24) и (28) для проводимости канала получим

$$g_D = \frac{\partial I_D}{\partial V_D} = \frac{2Z\mu}{L} [Q(a) - Q(y_2)]. \quad (30)$$

Как и следовало ожидать, дифференциальная проводимость канала g_D становится равной нулю при $y_2 \approx a$, т. е. когда $V_D + V_G = V_P - V_{bi}$ (отсечка канала). Сравним между собой выражения (29a) и (30) для крутизны и дифференциальной проводимости полевого транзистора. С одной стороны, согласно выражению (30), в линейной области ($V_D \rightarrow 0$, $y_2 \rightarrow a$) дифференциальная проводимость g_D пропорциональна $[Q(a) - Q(y)]$. С другой стороны, в области насыщения ($V_D + V_G \geq V_P$, $y_2 \rightarrow a$) крутизна $g_m \sim [Q(a) - Q(y_1)]$, т. е. пропорциональна той же разности зарядов. Тем самым мы получили полезное соотношение

$$\begin{aligned} g_{D0}(V_D \rightarrow 0) &= g_m(V_D \gg V_P) = \frac{2Z\mu}{L} [Q(a) - Q(y_1)] \equiv \\ &\equiv g_{\max} \left[1 - \frac{Q(y_1)}{Q(a)} \right], \end{aligned} \quad (31)$$

где $g_{\max} \equiv \frac{2Z\mu}{L} Q(a)$, справедливое при любом распределении легирующей примеси по глубине канала.

Рассмотренные ранее характеристики полевого транзистора с однородно легированным каналом представляют собой частный случай выведенных выше общих соотношений (выражения (28)—(30)). В табл. 1 приведены результаты, полученные с помощью выражения (28) для трех предельных профилей распределения примеси в канале [9]. В столбцах *A* и *C* приведены результаты расчета для двух предельных ситуаций, когда весь заряд сосредоточен в виде δ -функции при $y = 0$ или $y = a$. Столбец *B* соответствует рассмотренному выше случаю однородного легирования. Отметим, что безразмерный параметр $g_{\max} V_P / I_P$ зависит только от распределения примеси в канале, однако изменяется в довольно ограниченных пределах (от 2 до 4). Приведем здесь полученное

Таблица 1. Соотношения для предельных распределений концентрации легирующей примеси в полевом транзисторе с прямоугольным каналом

Параметр	Общий множитель	Коэффициент, обусловленный характером распределения заряда в канале		
		A (весь заряд при $y = a$)	B (однородное распределение)	C (весь заряд при $y = 0$)
$g_{\text{макс}}$	$\frac{2Z\mu r a}{L}$	1	1	1
V_P	$\frac{4\rho a^2}{\epsilon_s}$	$\frac{1}{4}$	$\frac{1}{8}$	0
I_P	$\frac{8Z\mu\rho^2 a^3}{\epsilon_s L}$	$\frac{1}{8}$	$\frac{1}{24}$	0
$\frac{g_{\text{макс}} V_P}{I_P}$	1	2	3	4

в случае A выражение для передаточной характеристики транзистора в области насыщения

$$I_{D \text{ sat}} = I_P \left[1 - \left(\frac{V_G + V_{bi}}{V_P} \right) \right]^2. \quad (32)$$

Соответствующая зависимость показана на рис. 4 вместе с графиком формулы (12) для однородно легированного канала. Отметим, что эти характеристики оказались неожиданно близки друг к другу. Таким образом, различным распределениям примеси по глубине канала, промежуточным между этими предельными ситуациями, соответствует весьма узкая область (заштрихованная на рис. 4). Расположенные в ней характеристики хорошо аппроксимируются выражением

$$\left[1 - \left(\frac{V_G + V_{bi}}{V_P} \right) \right]^n$$

с показателем $n = 2-2,25$ [10].

6.2.3. Нормально закрытый полевой транзистор

Электрические символы, обозначающие полевые транзисторы с p - n -переходом и МП-транзисторы, показаны на рис. 5. Выше мы рассматривали нормально открытый n -канальный полевой транзистор (обедненного типа), т. е. прибор, в котором проводящий канал существует при нулевом напряжении на затворе ($V_G = 0$).

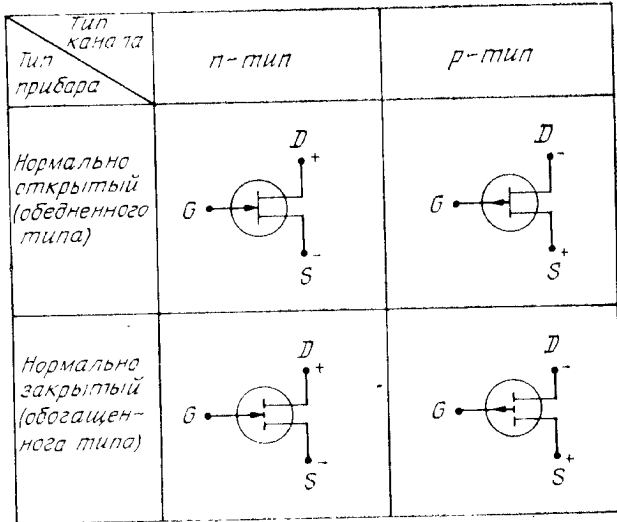


Рис. 5. Электрические символы для обозначения нормально открытых и нормально закрытых полевых транзисторов с *p* — *n*-переходом и МП-транзисторов.

С точки зрения применений в маломощных высокочастотных схемах представляют интерес нормально закрытые *p*-канальные транзисторы (обогащенного типа), в которых канал настолько узок, что при $V_G = 0$ он уже перекрыт встроенным потенциалом *p*⁺ — *n*-перехода V_{bi} . В таком транзисторе ток начинает протекать, когда положительное смещение на затворе превышает определенное пороговое значение V_T , величина которого

$$V_T \approx V_{bi} - V_P \quad (33)$$

или

$$V_{bi} \approx V_T + V_P. \quad (33a)$$

где V_P — напряжение отсечки канала, определяемое формулой (4) или (23). В окрестности порогового напряжения выражение для тока стока транзистора в режиме насыщения можно получить, подставив формулу (33a) для V_{bi} в соотношение (12) и затем разлагая его в ряд Тейлора по разности $V_G - V_T$ [11]:

$$I_D = \frac{Z \mu \epsilon_s}{2aL} (V_G - V_T)^2. \quad (34)$$

Выражение (34) подобно аналогичному соотношению для МОП-транзисторов обогащенного типа (гл. 8). Различие состоит лишь в том, что для МОП-транзисторов вместо глубины канала a в соответствующей формуле фигурирует толщина диэлектрика.

Выходные вольт-амперные характеристики нормально открытых и нормально закрытых приборов аналогичны (рис. 6). Переходные характеристики этих приборов (рис. 6) отличаются друг

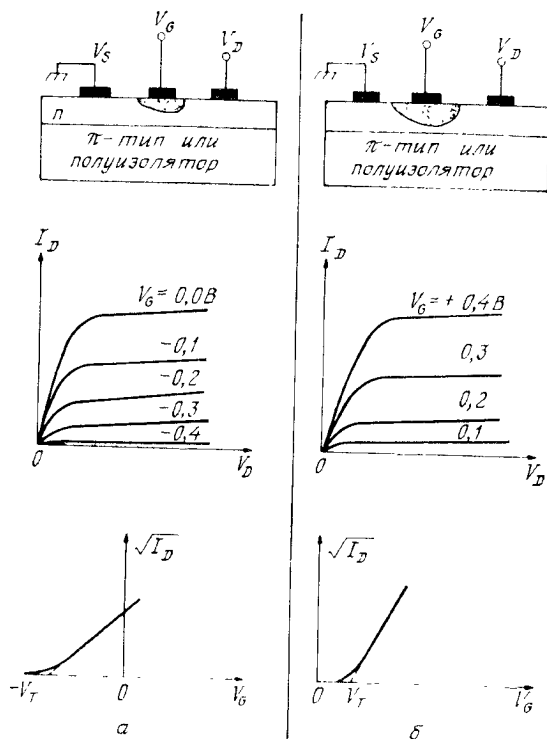


Рис. 6. Вольт-амперные характеристики нормально открытого (а) и нормально закрытого (б) МП-транзисторов.

от друга положением порогового напряжения V_T на оси напряжений. В нормально закрытом транзисторе ток при $V_G = 0$ не протекает, а начинает увеличиваться лишь при $V_G > V_T$ (в соответствии с формулой (34)). Отметим, что, поскольку величина встроенного потенциала $p^+ - n$ -затвора не превышает 1 В, диапазон прямых смещений на затворе транзистора ограничен величиной, равной $\sim 0,5$ В, чтобы избежать больших токов в цепи затвора из-за инжекции. Ниже мы рассмотрим работу нормально открытых приборов. Однако все выводы будут справедливы и для нормально закрытых приборов (с точностью до соответствующего сдвига напряжений на затворе).

6.3. ОСОБЕННОСТИ ХАРАКТЕРИСТИК РЕАЛЬНЫХ ПРИБОРОВ

6.3.1. Полевая зависимость подвижности

Для длинноканальных полевых транзисторов ($L \gg a$) предположения, использованные в разд. 6.2 при выводе вольт-амперных характеристик, в основном справедливы, и характеристики

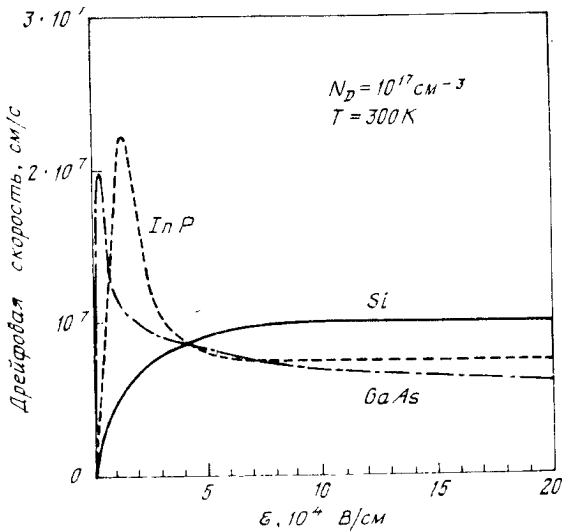


Рис. 7. Зависимость дрейфовой скорости электронов от электрического поля в Si, GaAs и InP [12, 13].

таких приборов довольно хорошо описываются соотношениями (12) или (28). Однако экспериментальные характеристики полевых транзисторов с короткими каналами (у которых отношение длины канала к его глубине не слишком велико) существенно отличаются от результатов приближенной теории. Одной из главных причин такого отличия является зависимость подвижности носителей тока от величины электрического поля, которая приводит к насыщению дрейфовой скорости при больших значениях поля. Зависимость дрейфовой скорости от электрического поля в Si, GaAs и InP приведена на рис. 7 [12, 13]. При малых полях дрейфовая скорость увеличивается пропорционально электрическому полю, а коэффициент этой пропорциональности и есть подвижность ($\mu \equiv dv/d\mathcal{E}$). В кремнии дрейфовая скорость монотонно увеличивается с электрическим полем и достигает скорости насыщения $v \approx 10^7$ см/с при полях, превышающих $5 \cdot 10^4$ В/см. В GaAs и InP зависимость $v(\mathcal{E})$ немонотонна: дрейфовая скорость сначала достигает максимума, а затем уменьшается, стремясь к значениям, равным $\sim (6-8) \cdot 10^6$ см/с.

Зависимость дрейфовой скорости от электрического поля в кремнии можно аппроксимировать простым аналитическим выражением (рис. 8)

$$v = \frac{\mu \mathcal{E}_x}{1 + \mu \mathcal{E} / v_{xs}} \quad (35)$$

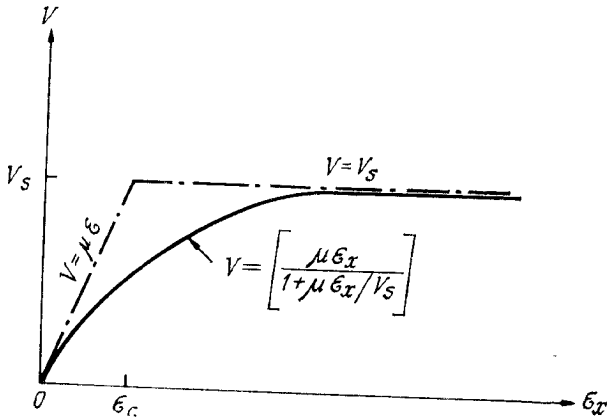


Рис. 8. Две аппроксимации полевой зависимости дрейфовой скорости.

При этом полный ток канала определяется выражением (вместо выражения (6))

$$I_D = qN_D \frac{\mu \mathcal{E}_x}{1 + \mu \mathcal{E}_x / v_s} (a - h) Z. \quad (36)$$

Используя введенные выше безразмерные глубины обедненного слоя (выражение (10)), преобразуем это выражение к виду [14]

$$I_D = 6I_P (1 - u) u' / (1 + 2uu'z), \quad (37)$$

где

$$u' = \partial u / \partial (x/L) = (L/V_P) \mathcal{E}_x (2u)^{-1},$$

а параметр

$$z \equiv \mu V_P / v_s L \quad (38)$$

представляет собой отношение характерной скорости $\mu V_P / L$ к скорости насыщения. Отметим, что в приближении постоянной подвижности ($\mu = \text{const}$) $z = 0$. Интегрируя выражение (37) вдоль канала от истока ($x = 0$) до произвольной точки x , получим

$$x/L = \frac{I_P}{I_D} \left[3 \left(1 - \frac{z}{3} \frac{I_D}{I_P} \right) (u^2 - u_1^2) - 2(u^3 - u_1^3) \right], \quad (39)$$

где, согласно граничному условию при $x = 0$ ($V_D = 0$),

$$u^2|_{x=0} \equiv u_1^2 = \frac{V_G + V_{bi}}{V_P}.$$

Подставив в выражение (39) граничное условие на стоке $x = L$, $u = u_2$, для полного тока транзистора получим следующее выражение [14]:

$$I_D = \frac{I_P [3(u_2^2 - u_1^2) - 2(u_2^3 - u_1^3)]}{1 + \mu V_D / v_s L}. \quad (40)$$

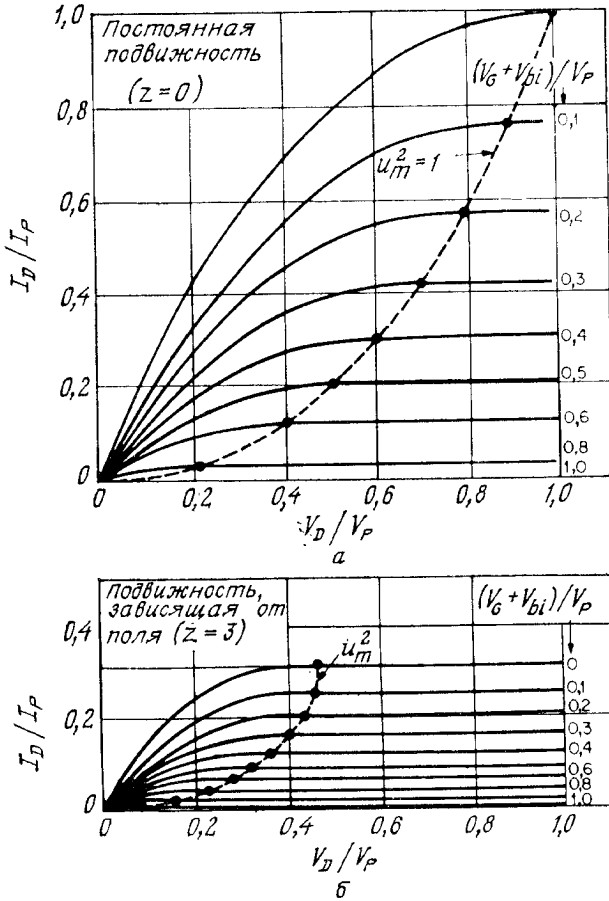


Рис. 9. Вольт-амперные характеристики, рассчитанные в приближении постоянной подвижности ($z = 0$) (а) и с учетом полевой зависимости подвижности ($z = 3$) (б) [14].

Из сравнения выражений (40) и (11) следует, что при одинаковых I_P , u_1 и u_2 полевая зависимость подвижности уменьшает ток стока в $(1 + \mu V_D/v_s L)$ раз. Зависимости тока I_D от напряжения стока (в относительных единицах), рассчитанные по уравнению (40) в приближении постоянной подвижности ($z = 0$) и при $z = 3$, приведены на рис. 9. Видно, что полевая зависимость подвижности приводит к значительному уменьшению тока транзистора. Дифференцируя, находим, что выражение (40) имеет максимум при значении $u_2 = u_m$, которое определяется следующим уравнением:

$$u_m^3 + 3u_m \left(\frac{1}{z} - u_1^2 \right) + 2u_1^3 - \frac{3}{z} = 0. \quad (41)$$

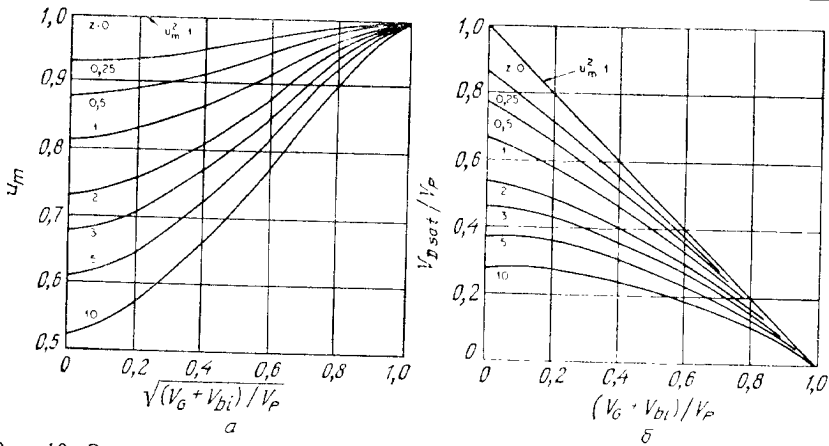


Рис. 10. Зависимость ширины обедненного слоя у стока в начале насыщения u_m от напряжения на затворе V_G и зависимость напряжения начала насыщения $V_{D \text{ sat}}$ от V_G . (б) [14].

Выразив отсюда u_1^3 через u_m и подставив его в выражение (40), получим для тока насыщения

$$I_{D \text{ sat}} = \frac{3I_P(1 - u_m)}{z}. \quad (42)$$

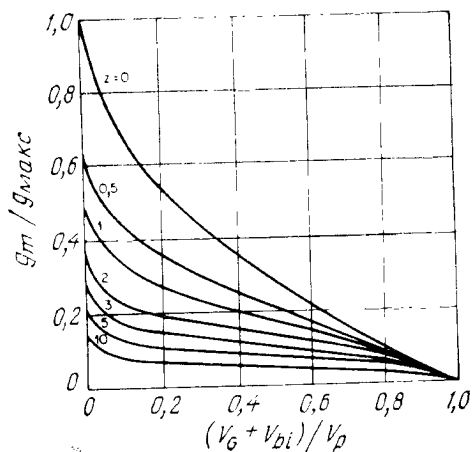
Численные решения уравнения (41) при различных z приведены на рис. 10, а. Здесь же показано и решение $u_m = 1$ для случая постоянной подвижности ($z = 0$), соответствующее граничному условию $(V_D + V_G + V_{bi}) = V_P$. На рис. 10, б приведены графики зависимости $V_{D \text{ sat}}/V_P$ от $(V_G + V_{bi})/V_P$, т. е. зависимости напряжения начала участка насыщения от напряжения на затворе для различных значений параметра z . Отметим, что с ростом z уменьшаются ток насыщения и напряжение $V_{D \text{ sat}}$, соответствующее началу области насыщения тока.

Дифференцируя выражение (42), получим выражение для крутизны транзистора

$$g_m = \frac{\partial I_{D \text{ sat}}}{\partial V_G} = \frac{g_{\text{макс}}(u_m - u_1)}{1 + z(u_m^2 - u_1^2)}. \quad (43)$$

Крутизна также оказывается меньше соответствующей крутизны в приближении постоянной подвижности (выражение (17)). Последняя получается из выражения (43) при $z = 0$ и $u_m = 1$. Зависимости $g_m/g_{\text{макс}}$ от напряжения на затворе при различных z приведены на рис. 11. Отметим, что при больших значениях z крутизна транзистора становится почти не зависящей от напряжения на затворе.

Рис. 11. Зависимость безразмерной крутизны от напряжения на затворе [14].



6.3.2. Модель двух областей

Изложенная в предыдущем разделе модель учета полевой зависимости подвижности неплохо согласуется с экспериментальными характеристиками кремниевых короткоканальных полевых и МП-транзисторов. В GaAs полевая зависимость подвижности более сложная, чем в кремнии, и насыщение дрейфовой скорости в GaAs происходит при значительно меньших электрических полях. Поэтому для GaAs было предложено [7] использовать кусочно-линейную аппроксимацию зависимости $v(\mathcal{E}_x)$ (рис. 8). При малых электрических полях подвижность считается постоянной, и $v = \mu \mathcal{E}_x$, а при $\mathcal{E}_x > \mathcal{E}_c$ скорость носителей считается не зависящей от электрического поля ($v = v_s$). При использовании этого приближения канал полевого транзистора, работающего в режиме насыщения (отсечки канала), как бы разбивается на две области (рис. 12). В области I (вблизи истока) подвижность считается постоянной и используется приближение плавного канала, описанное в разд. 6.2. В области II (вблизи стока) скорость носителей считается равной скорости насыщения ($v = v_s$), а глубина проводящего канала подбирается такой, чтобы обеспечить непрерывность полного тока канала транзистора на границе этих областей. Отметим, что плоскость $y = y_c$, разграничивающая эти две области, теперь уже не фиксируется на стоке ($y_c = L$), как в предыдущем случае (разд. 6.3.1). Напротив, положение этой области (L_1) будет изменяться в зависимости от величины напряжений на стоке и затворе транзистора. Оно определяется условием $\mathcal{E}_x(L_1) = \mathcal{E}_c$. Следовательно, сформулированная выше модель двух областей должна описывать также и линейный участок вольт-амперных характеристик полевого транзистора (при этом $L_1 = L$, т. е. область I распространяется на весь канал).

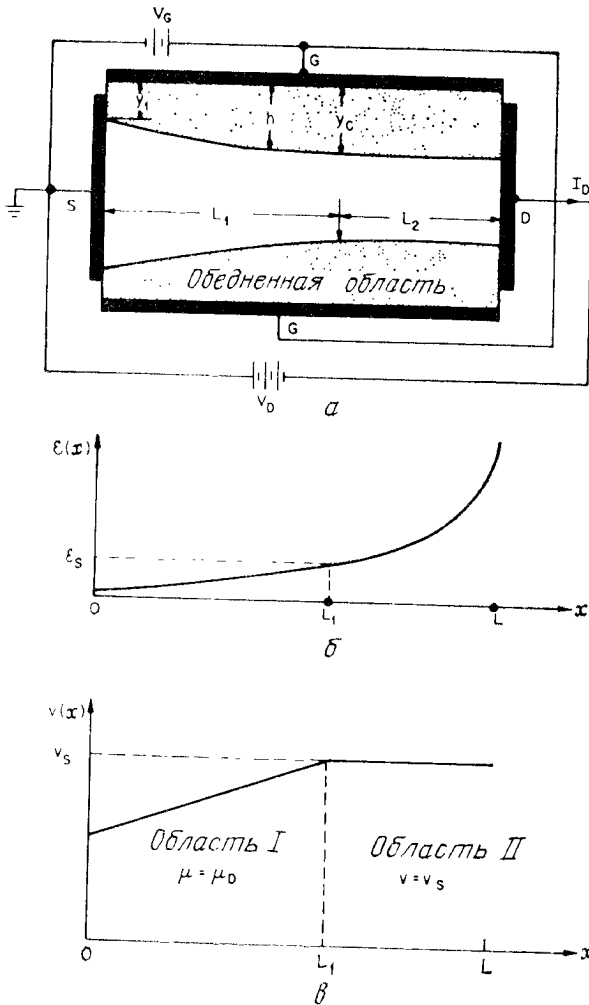


Рис. 12. Модель двух областей (область I — область с постоянной подвижностью, область II — область, где скорость носителей равна скорости насыщения) [7]

Обозначим ширину обедненного слоя в точке $x = L_1$, разграничивающей области I и II, символом y_c (рис. 12, а) и введем безразмерную величину $u_c = y_c/a$. Рассмотрим сначала область I. Интегрируя в ней уравнение (6) от $x = 0$ до $x = L_1$ (а не до $x = L$, как раньше), получим выражение

$$I_D = I_1 [3(u_c^2 - u_1^2) - 2(u_c^3 - u_1^3)], \quad (44)$$

где

$$I_1 \equiv Z \mu q^2 N_D^2 a^3 / 6 \epsilon_s L_1,$$

которое отличается от выражения (11) тем, что в нем вместо u_2 стоит u_c , а вместо I_P — величина I_1 . Длина L_1 определяется из условия непрерывности тока на границе областей I и II. Поскольку в области II носители движутся со скоростью насыщения, то

$$I_D = q N_D v_s (a - y_c) Z = I_S (1 - u_c), \quad (45)$$

где $I_S = q N_D v_s a Z$ — ток насыщения полностью открытого канала. Приравнявая выражения (44) и (45), получим соотношение, определяющее L_1 :

$$L_1 = zL \frac{(u_c^2 - u_1^2) - \frac{2}{3}(u_c^3 - u_1^3)}{1 - u_c}, \quad (46)$$

где $z = \mu V_P / v_s L$. Чем больше z , тем сильнее сказывается эффект насыщения дрейфовой скорости на характеристиках полевых транзисторов. В полевых СВЧ-транзисторах $z = 2-20$. Еще большие значения параметра z характеризуют приборы, предназначенные для работы в частотном диапазоне ~ 10 ГГц.

Если безразмерные толщины обедненного слоя u_c и u_1 известны, то с помощью выражения (46) можно определить L_1 , а затем ток стока I_D . И наоборот, при заданных u_1 и значении тока стока I_D по формуле (45) можно рассчитать u_c , а из формулы (46) определить L_1 .

Чтобы получить величину падения напряжения в канале транзистора от истока до стока, нужно проинтегрировать продольное электрическое поле \mathcal{E}_x от $x = 0$ до $x = L$. Часть этого напряжения в области I равна

$$V_I = V_P - V_S = V_P (u_c^2 - u_1^2). \quad (47)$$

Электрическое поле в области II определяется свободными зарядами, локализованными на электроде стока. Принимая во внимание наименьшую гармонику соответствующего решения уравнения Лапласа, для падения напряжения в области II можно написать

$$V_{II} \simeq \frac{2a}{\pi} \mathcal{E}_c \cos \frac{\pi y}{2a} \operatorname{sh} \left[\frac{\pi(L - L_1)}{2a} \right]. \quad (48)$$

Полное падение напряжения в канале равно сумме выражений (47) и (48);

$$V_D = V_P \left\{ (u_c^2 - u_1^2) + \frac{2}{\pi} \frac{a}{zL} \operatorname{sh} \left[\frac{\pi(L - L_1)}{2a} \right] \right\}. \quad (49)$$

Уравнения (46) и (49) являются той парой уравнений, которая позволяет исключить параметр L_1 и выразить безразмерную тол-

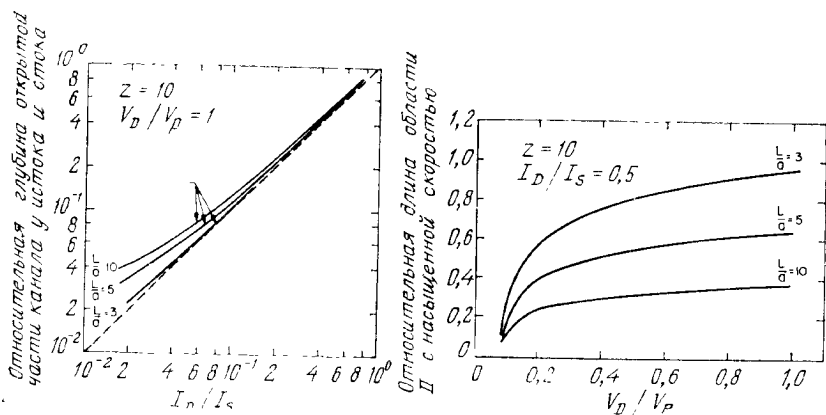


Рис. 13. Относительная глубина (а) открытой части канала в зависимости от тока стока при $z = 10$ и $V_D/V_P = 1$ и относительная длина (б) области с насыщенной скоростью в зависимости от напряжения на стоке при различных L/a [7].

щину u_c через напряжения на затворе V_G и стоке V_D транзистора, чтобы затем с помощью выражения (45) получить соответствующую зависимость тока стока.

На рис. 13, а показана [7] относительная глубина проводящего канала у истока ($1 - u_1$) и стока ($1 - u_c$) в зависимости от безразмерного тока I_D/I_S при $z = 10$ и некоторых различных значений отношения длины канала к глубине L/a . Следует указать, что в обычных режимах работы транзисторов отношение I_D/I_S не превышает 0,5. Отметим, что, согласно этим графикам, изменение ширины проводящего канала от истока до стока не превышает 10—20 %, за исключением области очень малых токов. Иными словами, граница между обедненной областью и проводящим каналом почти параллельна плоскости электрода затвора. Графики, приведенные на рис. 13, б, свидетельствуют о сильном проникновении области II, где дрейфовая скорость равна скорости насыщения, в канал под затвор транзистора. Так, например, при $L/a = 3$ и $V_D = V_P$ область II занимает примерно 95 % всей длины канала.

6.3.3. Модель с полностью насыщенной скоростью

Эта модель предназначена для описания характеристик приборов с очень короткими каналами, в которых, как можно предполагать, область насыщения дрейфовой скорости охватывает всю длину канала. При этом ток насыщения транзистора будет изменяться прямо пропорционально изменению ширины проводящего канала $a - h$, где h — ширина обедненного слоя. В простейшем

случае однородно легированного канала эта модель дает следующее выражение для тока транзистора:

$$I = qv_s Z (a - h) N_D. \quad (50)$$

Выражение (50) довольно хорошо описывает экспериментальные характеристики короткоканальных ($L \lesssim 2$ мкм) полевых транзисторов с p — n -переходом на GaAs. Соответствующее значение скорости насыщения при этом оказывается равным $1,2 \cdot 10^7$ см/с (при $T = 300$ К).

Для произвольного профиля легирования канала эта модель дает

$$I = v_s Z \int_h^a \rho(y) dy. \quad (51)$$

Характер распределения электрического поля и потенциала в неоднородно легированных обедненных слоях определяется выражениями (19) и (22). Дифференцируя выражения (51) и (22), получим

$$dI/dh = v_s Z \rho(h) \quad (52a)$$

и

$$dV/dh = h\rho(h)/\epsilon_s. \quad (52b)$$

Следовательно, крутизна

$$g_m = \frac{dI}{dV} = v_s Z \epsilon_s / h (V_G). \quad (53)$$

Из этого выражения следует, что для обеспечения линейности передаточных характеристик ($g_m = \text{const}$) необходимо выбирать профиль легирования так, чтобы глубина обеднения h оказалась слабо меняющейся функцией напряжения на затворе V_G . На рис. 14 приведены передаточные характеристики, рассчитанные для некоторых типичных профилей легирования [15]. Здесь напряжение $V = V_G + V_{bi}$, т. е. содержит также встроенный потенциал p^+ — n -затвора. Что же касается напряжения V_D (между стоком и истоком), то в рассматриваемой модели оно вообще не фигурирует. Отметим, что для всех приведенных на рис. 14 профилей легирования передаточные характеристики обладают существенной нелинейностью (линейность возникает лишь при предельных значениях параметров, характеризующих эти распределения). Последнее обстоятельство существенно отличает рассматриваемую модель с полностью насыщенной скоростью от рассмотренной выше модели с постоянной подвижностью, где изменения профиля легирования оказывают пренебрежимо малое влияние на передаточные характеристики. Согласно выражению (53), крутизна транзистора g_m уменьшается с ростом напряжения на затворе V_G , однако другой важный параметр g_m/C_{GS} , характеризующий быстроедействие поле-

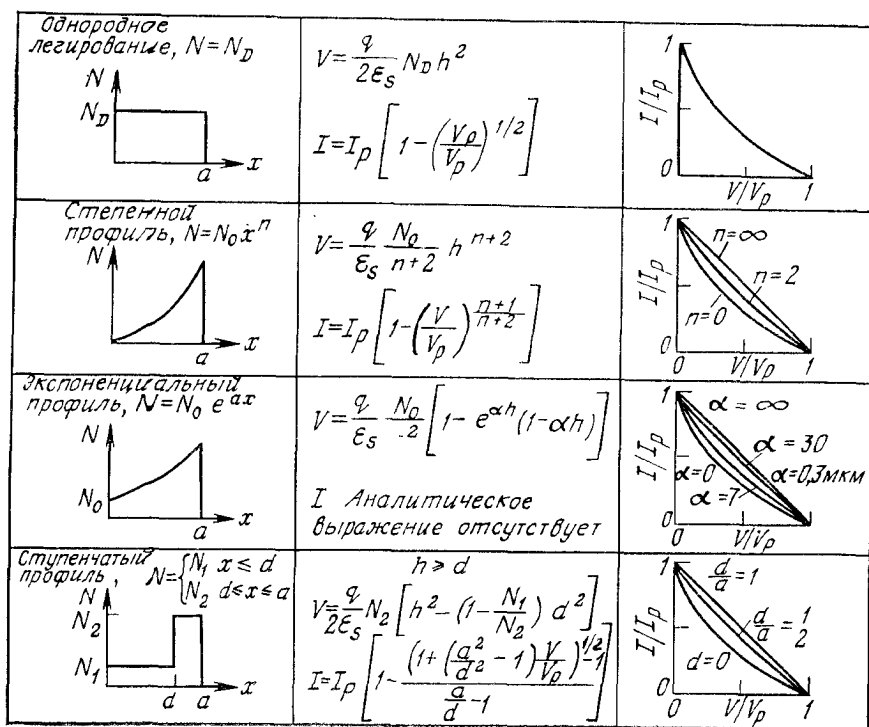


Рис. 14. Передаточные характеристики для некоторых типичных профилей легирования [15].

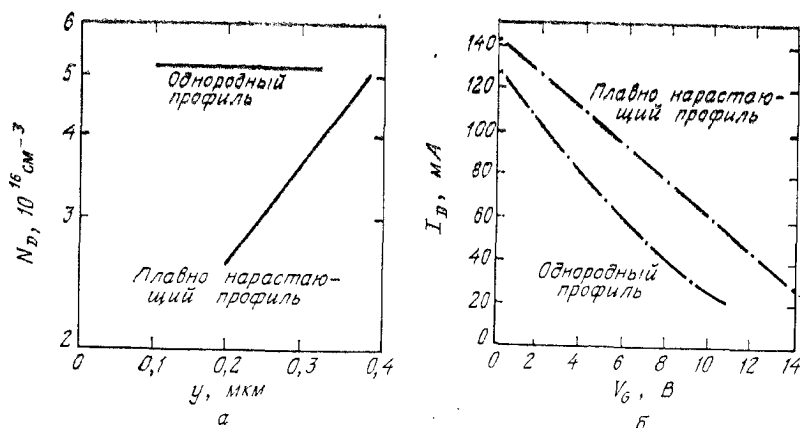


Рис. 15. Экспериментально измеренные профили легирования (а) и передаточные характеристики полевых транзисторов с плавно нарастающим и однородным профилем легирования канала (б) [16].

вого транзистора, остается при этом неизменным, поскольку емкость затвор—исток $C_{GS} = \epsilon_s/h$ обратно пропорциональна глубине обедненного слоя, т. е.

$$g_m/C_{GS} = v_s Z = \text{const.} \quad (54)$$

Из экспериментальных результатов следует, что полевые транзисторы с плавно [16] и ступенчато легированным [17] каналом имеют передаточные характеристики с более высокой степенью линейности. Типичный пример показан на рис. 15. В транзисторе с плавно нарастающим к центру канала профилем легирования крутизна $g_m = 9,5 \text{ МОм}^{-1}$ остается постоянной в диапазоне напряжений на затворе $0 < V_G < 12 \text{ В}$.

6.3.4. Двумерный анализ

В предыдущих разделах мы вывели аналитические соотношения для вольт-амперных характеристик полевых транзисторов с p — n -переходом и МП-транзисторов, используя различные предположения. Однако в приборах с коротким каналом, в особенности при больших напряжениях на стоке, начинают доминировать двумерные эффекты, которые трудно учесть аналитически [18]. Соответствующий анализ мы проведем отдельно для кремниевых полевых транзисторов с p — n -переходом и аналогичных приборов на GaAs, поскольку зависимости дрейфовой скорости от поля в этих материалах существенно различаются (рис. 7).

Рассмотрим сначала вольт-амперные характеристики тонкого слоя кремния n -типа, нанесенного на изолирующую подложку [8] (рис. 16, *a*). Как уже говорилось выше, при малых напряжениях на стоке этот кремниевый слой подобен резистору. При больших напряжениях темп увеличения дрейфовой скорости электронов с ростом электрического поля ослабевает, и вольт-амперная характеристика пленки отклоняется вниз от начальной прямой. При еще больших напряжениях V_D дрейфовая скорость достигает скорости насыщения, что приводит к соответствующему насыщению тока. На рис. 16, *б* показана модифицированная структура с затвором между стоком и истоком в виде p — n -перехода или выпрямляющего контакта металл—полупроводник. Здесь обедненная область под затвором действует как слой изолятора, что сокращает область канала, «открытую» для протекания электронного тока. Ширина обедненного слоя определяется величиной приложенных напряжений. При малом напряжении на стоке и затворе, закороченном на исток (рис. 16, *б*), обедненный слой под затвором уменьшает поперечное сечение проводящего канала $b < a$. Следовательно, сопротивление между стоком и затвором увеличивается.

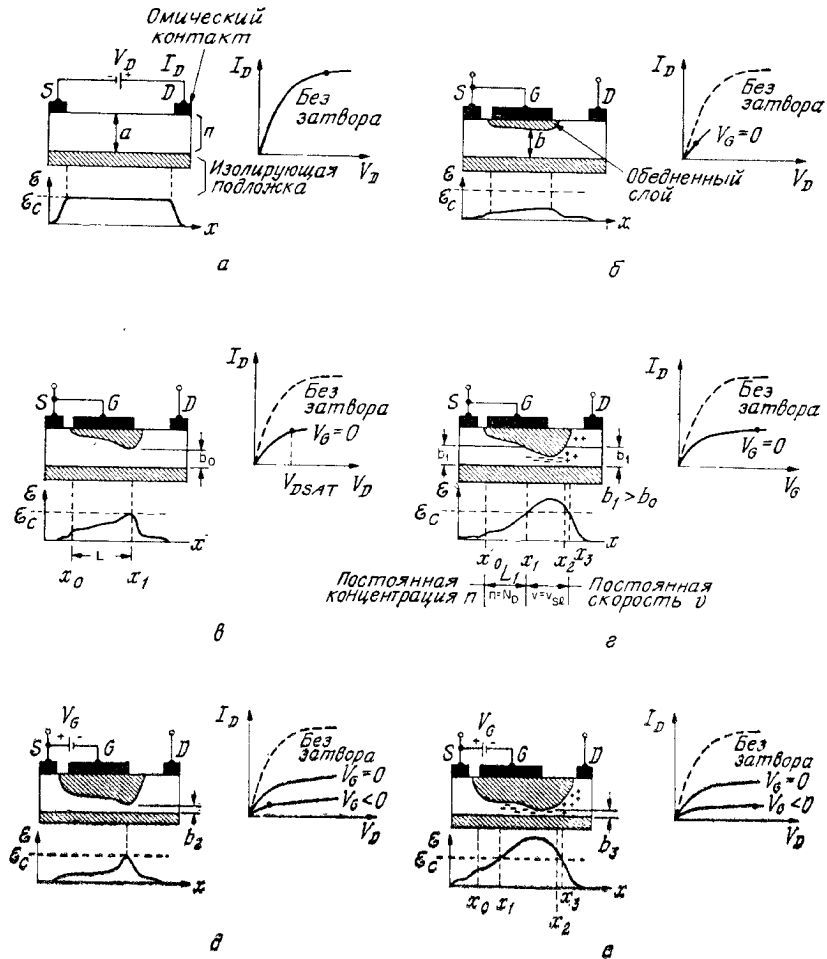


Рис. 16. Распределение электрического поля и вольт-амперные характеристики кремниевых МП-транзисторов при различных соотношениях напряжений на стоке и затворе [8].

Используя закон сохранения полного тока, для любого поперечного сечения канала можно написать

$$I_D = Zqn(x)v(x)b(x). \quad (55)$$

Концентрация электронов $n(x)$ равна концентрации легирующей примеси N_D , пока электрическое поле не превышает критического значения \mathcal{E}_c . Потенциал вдоль канала увеличивается от нуля на истоке до V_D на стоке. Следовательно, обратное смещение перехода затвор—канал и локальная ширина обедненного слоя также уве-

ливаются по мере приближения к стоку. Результирующее уменьшение ширины проводящего канала должно быть скомпенсировано соответствующим увеличением продольного электрического поля и дрейфовой скорости электронов, чтобы величина полного тока оставалась неизменной по длине канала. С увеличением напряжения стока (при некотором $V_D = V_{D\text{ sat}}$) электрическое поле в канале у стока достигает критического значения \mathcal{E}_c , а скорость электронов — скорости насыщения (рис. 16, в)). При этом наименьшая ширина канала у стока достигает минимального значения b_0 , а ток транзистора начинает насыщаться.

При дальнейшем увеличении напряжения стока ($V_D > V_{D\text{ sat}}$) обедненная область расширяется к стоку (рис. 16, з)). Однако точка x_1 , где электроны впервые достигают скорости насыщения, смещается в противоположном направлении (к истоку), падение напряжения между истоком и точкой x_1 уменьшается, и, следовательно, ширина канала в точке b_1 увеличивается ($b_1 > b_0$). Поэтому ток инжекции электронов из электронейтральной части канала ($x < x_1$) в область насыщения дрейфовой скорости ($x_1 < x < x_2$) увеличивается, благодаря чему вольт-амперная характеристика транзистора на участке насыщения (конечное выходное сопротивление) имеет положительный наклон [19].

По мере продвижения от точки x_1 к стоку потенциал в канале растет, ширина обедненной области увеличивается, а проводящий канал сужается. Но, поскольку скорость электронов в этой области уже не зависит от электрического поля и равна скорости насыщения, для компенсации этого сужения канала и обеспечения сохранения полного тока концентрация электронов здесь увеличивается и становится больше концентрации доноров. Поэтому в соответствии с уравнением (55) в части канала, где $b < b_1$, т. е. при $x_1 < x < x_2$, аккумулируются электроны, и эта часть канала оказывается отрицательно заряженной. При $x = x_2$, где глубина проводящего канала снова становится равной b_1 , концентрация электронов равна концентрации доноров. Этот отрицательный заряд области канала $x_1 < x < x_2$ компенсируется положительно заряженным слоем $x_2 < x < x_3$ с некоторым дефицитом электронов. Следовательно, часть напряжения стока, избыточная над $V_{D\text{ sat}}$, падает на дипольном слое, который расширяется в стоковой части канала при дальнейшем росте V_D .

Отрицательное напряжение на затворе (рис. 16, д)) увеличивает ширину обедненной области и, следовательно, сужает канал, что увеличивает сопротивление на линейном (при малых V_D) участке характеристики транзистора. При этом уменьшается напряжение $V_{D\text{ sat}}$, при котором в наиболее узкой части канала (глубиной $b_2 < b_1$) достигается критическое поле \mathcal{E}_c . Поэтому при $V_G > 0$ участок насыщения характеристики транзистора начинается при меньших напряжениях и токах, чем при $V_G = 0$. При

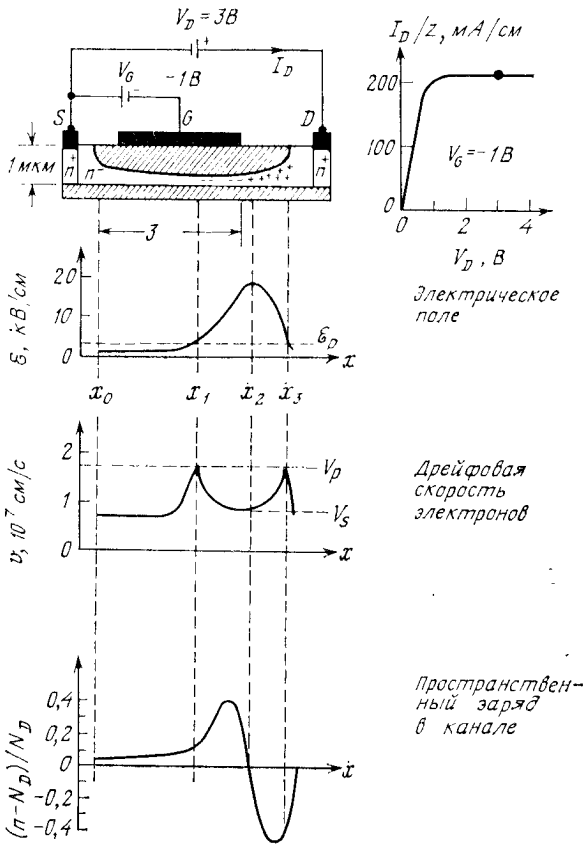


Рис. 17. Поперечное сечение канала и распределения поля, дрейфовой скорости и концентрации электронов по длине канала в МП-транзисторе из GaAs в режиме насыщения [8].

дальнейшем увеличении напряжения стока $V_D > V_{D \text{ sat}}$ у стокового края затвора также образуется дипольный слой (рис. 16, e), обеспечивающий непрерывность полного тока в канале.

Особенности поведения полевых транзисторов из GaAs, где зависимость дрейфовой скорости от электрического поля более сложная, чем в кремнии, иллюстрируют графики, приведенные на рис. 17 [8, 20]. Здесь также наиболее узкая часть канала расположена у стокового края затвора. Дрейфовая скорость электронов достигает максимального значения в точках x_1 и x_3 и имеет некоторый провал в промежутке между ними. Поэтому в полевых транзисторах из GaAs эффект аккумуляции электронов в узкой части канала более ярко выражен, чем в кремнии. Для компенсации отрицательного заряда в области $x_1 < x < x_3$ за этой областью,

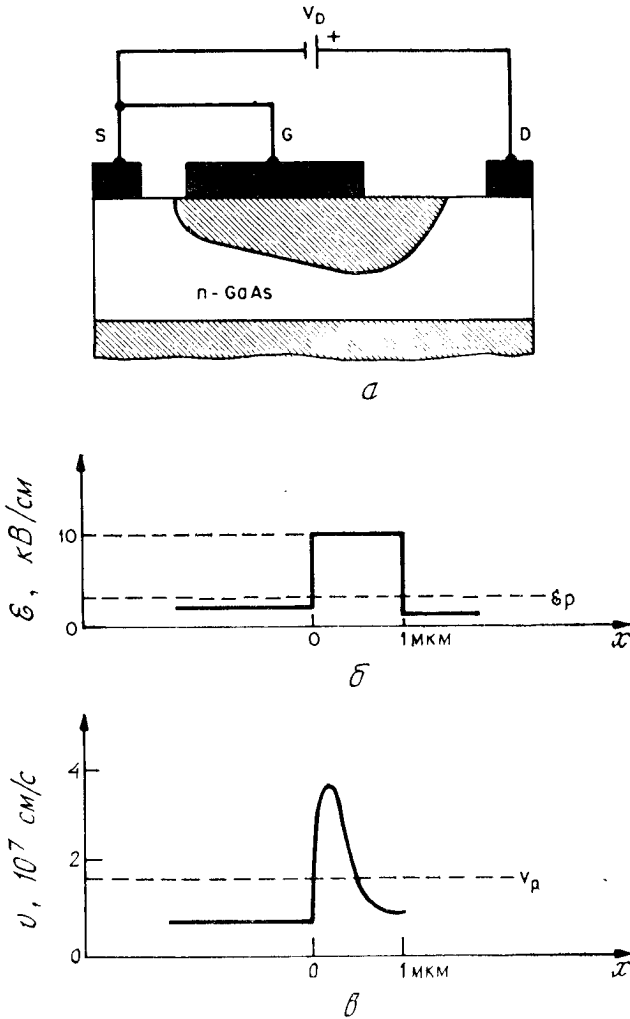


Рис. 18. Повышение стационарной дрейфовой скорости в области высоких полей ($\mathcal{E} > \mathcal{E}_p$) под затвором [8].

ближе к стоку ($x_2 < x < x_3$), образуется положительно заряженный слой, обедненный электронами. Его заряд примерно равен отрицательному заряду в области аккумуляции, и на этом стационарном дипольном слое обычно падает большая часть стокового напряжения.

В приборах с очень короткими каналами возможна такая ситуация, когда электроны настолько быстро пролетают область

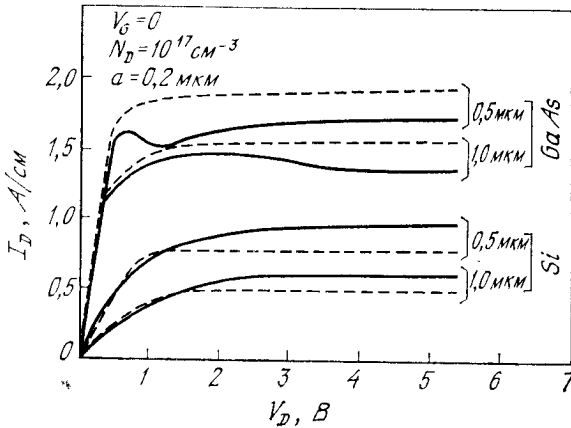


Рис. 19. Вольт-амперные характеристики кремниевых МП-транзисторов и из GaAs с длиной канала 1 и 0,5 мкм [22].

высокого поля в канале у стока, что при этом не успевают термализоваться [21]. Такую неравновесную ситуацию иллюстрирует рис. 18. Можно считать, что в области малых полей ($\mathcal{E} < \mathcal{E}_p$, где \mathcal{E}_p — электрическое поле, соответствующее пиковому значению дрейфовой скорости) электроны не разогреваются и остаются равновесными. В области же высоких полей ($\mathcal{E} > \mathcal{E}_p$) электроны ускоряются до более высокого значения скорости, чем стационарная скорость насыщения v_s . На рис. 18, в показано, что ожидаемое максимальное значение электронной скорости более чем в два раза превышает пиковую скорость; затем на длине ~ 1 мкм она релаксирует и достигает стационарной скорости насыщения. Такое превышение скорости уменьшает время пролета электронов через область высокого поля, что улучшает характеристики прибора, а также приводит к сдвигу области аккумуляции в зазор между затвором и стоком.

На рис. 19 приведены вольт-амперные характеристики кремниевых полевых МП-транзисторов и из GaAs, полученные двумерным численным расчетом (сплошные линии) [22]. Здесь же приведены характеристики, рассчитанные с помощью приближенных теоретических моделей (штриховые линии). Для кремниевых приборов использовалась модель, учитывающая полевую зависимость подвижности, а для приборов из GaAs — модель двух областей. Следует отметить очень хорошее соответствие характеристик в линейной области. В области насыщения это соответствие хуже (однако и здесь разность не превышает 15 %). Отметим, кроме того, что предсказываемые двумерным расчетом провалы на характеристиках GaAs-приборов никогда не наблюдались экспериментально.

6.4. ВЫСОКОЧАСТОТНЫЕ ХАРАКТЕРИСТИКИ

6.4.1. Малосигнальная эквивалентная схема

Канальные полевые транзисторы (в особенности на GaAs) широко применяются в высокочастотных схемах в качестве малошумящих усилителей, высокоэффективных генераторов и быстродействующих логических элементов. Мы сначала рассмотрим малосигнальную эквивалентную схему полевого транзистора с $p-n$ -переходом или МП-транзистора (рис. 20, а) (режим насыщения, схема

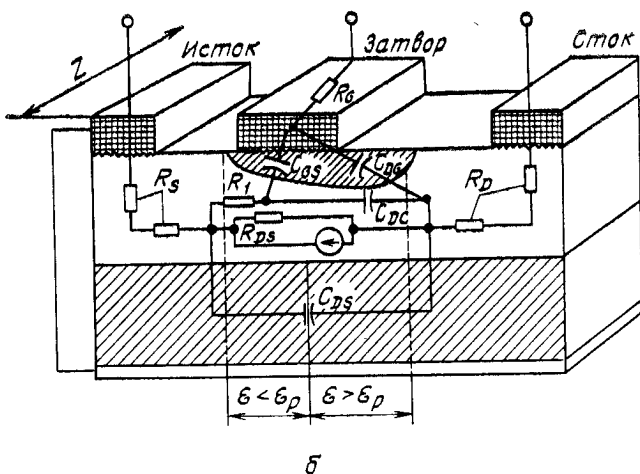
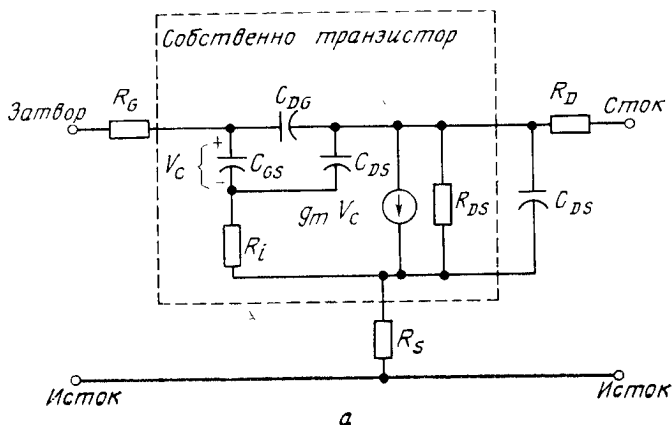


Рис. 20. Эквивалентная электрическая схема МП-транзистора (а) и физические элементы структуры, соответствующие элементам эквивалентной электрической схемы (б) [8].

включения с общим истоком) [8]. На рис. 20, б показано, каким физическим элементам полевого транзистора соответствуют элементы этой эквивалентной электрической схемы. Среди них следует выделить собственно транзисторные элементы (такие, как $(C_{DG} + C_{GS})$ — полная емкость перехода затвор—канал, входное сопротивление R_i , выходное сопротивление канала R_{DS}) и внешние (паразитные) элементы (такие, как омические сопротивления стока и истока R_S и R_D и емкость подложки).

Ток обратносмещенного перехода затвор—канал определяется выражением

$$I_G = I_S [\exp(qV_G/nkT) - 1], \quad (56)$$

где n — фактор неидеальности ($1 < n < 2$), а I_S — ток насыщения. Отсюда для входного сопротивления получим

$$R_i \equiv (\partial I_G / \partial V_G)^{-1} = nkT/q(I_G + I_S). \quad (57)$$

Согласно этой формуле, при $I_G \rightarrow 0$ и $I_S = 10^{-10}$ А входное сопротивление при комнатной температуре составляет ~ 250 МОм.

Последовательные сопротивления стока и истока не модулируются напряжением затвора и поэтому обуславливают дополнительное омическое падение напряжения между затвором и контактами стока и истока. Это приводит к соответствующему уменьшению выходной проводимости и крутизны прибора. С учетом последовательных сопротивлений R_S и R_D напряжения V_D и V_G в формуле (11а) следует заменить на $[V_D - I_D(R_S + R_D)]$ и $(V_G - I_D R_S)$. В линейном режиме транзистора сопротивления R_S , R_D и $1/g_{D0}$ включены последовательно, поэтому выходная проводимость прибора равна $g_{D0}/[1 + (R_S + R_D)g_{D0}]$. Такой же результат получается и для крутизны транзистора в линейной области. В режиме насыщения на крутизну влияет только последовательное сопротивление истока $g_m/(1 + R_S g_m)$, а последовательное сопротивление стока приводит лишь к некоторому увеличению напряжения начала насыщения $V_{D \text{ sat}}$.

Высокочастотные характеристики канальных транзисторов зависят от двух основных факторов: времени пролета и характерного времени зарядки затвора RC . Временем пролета называют конечный временной интервал τ , который требуется носителям для преодоления расстояния от истока до стока. В приближении постоянной подвижности

$$\tau = \frac{L}{\mu \mathcal{E}_x} \approx \frac{L^2}{\mu V_D}, \quad (58)$$

а при высоких полях в приближении скорости насыщения

$$\tau = L/v_s. \quad (59)$$

Так, например, в полевом транзисторе из GaAs с длиной затвора 1 мкм время пролета составляет ~ 10 пс (10^{-11} с). Обычно время

пролета невелико по сравнению с временной постоянной RC , определяемой входной емкостью затвора и крутизной транзистора.

В соответствии с эквивалентной схемой (рис. 20, а) граничная частота f_T определяется как частота, при которой ток через C_{GS} равен току генератора $g_m V_C$:

$$f_T = \frac{g_m}{2\pi C_{GS}} \left(= \frac{1}{2\pi\tau} = \frac{v_s}{2\pi L} \right). \quad (60)$$

Максимальная частота колебаний определяется выражением

$$f_{\max} \simeq \frac{f_T}{2\sqrt{r_1 + f_T \tau_s}}, \quad (61)$$

где

$$r_1 = (R_G + R_t + R_S)/R_{DS}$$

— отношение входного и выходного сопротивлений, а постоянная времени

$$\tau_s \equiv 2\pi R_G C_{DG}.$$

Коэффициент однонаправленного усиления

$$U \simeq (f_{\max}/f)^2. \quad (62)$$

С ростом частоты он уменьшается со скоростью 6 дБ/октава. При $f = f_{\max}$ коэффициент усиления мощности становится равным 1. Для повышения f_{\max} следует увеличивать f_T и оптимизировать отношение собственно транзисторных сопротивлений R_t/R_{DS} , а также уменьшать внешние сопротивления R_G , R_S и емкость обратной связи C_{DG} .

6.4.2. Ограничения по мощности и частоте

Граничная частота полевых канальных транзисторов определяется геометрией прибора и параметрами материала. В Si и GaAs электроны имеют более высокую подвижность, чем дырки. Поэтому в СВЧ-схемах используются только n -канальные полевые транзисторы. Кроме того, поскольку в GaAs электронная подвижность при малых полях примерно в пять раз выше, чем в кремнии, граничная частота f_T в GaAs-транзисторах должна быть больше, чем в аналогичных кремниевых приборах.

Важнейшим геометрическим параметром полевого транзистора является длина затвора L . При уменьшении L уменьшается емкость затвора C_{GS} и увеличивается крутизна, а следовательно, увеличивается и граничная частота f_T . На рис. 21 приведены теоретические зависимости f_T от длины канала L для транзисторов из Si, GaAs и InP [23, 24]. Отметим, что в InP-приборах ожидается более высокая граничная частота f_T , чем в GaAs-транзисторах, что

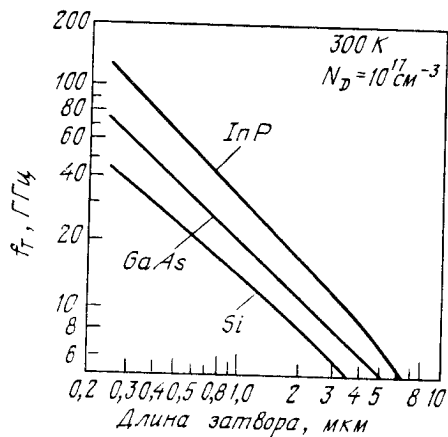


Рис. 21. Теоретические зависимости частоты отсечки от длины затвора для транзисторов из Si, GaAs и InP [23, 24].

обусловлено более высоким значением пиковой дрейфовой скорости в этом материале [25]. Для МП-транзисторов из Si и из GaAs с длиной затвора менее 0,5 мкм граничная частота f_T соответствует миллиметровому диапазону спектра электромагнитного излучения (≥ 30 ГГц).

Для того чтобы затвор транзистора мог эффективно контролировать транспорт электронов в канале, его длина должна превышать глубину канала ($L/a > 1$) [24]. Поэтому, сокращая L , нужно одновременно уменьшать и глубину канала a , обеспечивая выполнение условия $L/a > 1$. Для этого в полевых транзисторах из Si и GaAs обычно используют более высокий уровень легирования канала, не превышающий, однако, $\sim 5 \cdot 10^{17} \text{ см}^{-3}$ (во избежание пробоя). При таком уровне легирования минимальная длина канала L ограничена величиной около 0,1 мкм, что соответствует граничной частоте $f_T \approx 100$ ГГц.

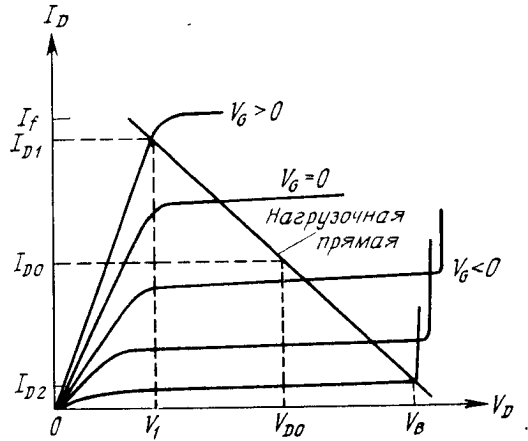
На рис. 22 приведены вольт-амперные характеристики мощного МП-транзистора. Его максимальная выходная мощность на синусоидальном сигнале зависит от максимальных значений тока и напряжения (рис. 22) следующим образом:

$$P_{\text{out}} = \frac{(I_{D1} - I_{D2})(V_B - V_1)}{8}. \quad (63)$$

Здесь ток I_{D1} близок к значению I_f (максимальному току канала при предельно допустимом прямом смещении затвора), а V_B — напряжение лавинного пробоя. Для повышения выходной мощности следует максимизировать I_f и V_B . Величину максимального тока можно оценить [26] с помощью уравнений (44) и (45), подставив в них $u_1 = 0$:

$$u_2^3 - 1,5u_2^2 - 1,5(u_2 - 1)/z = 0, \quad (64)$$

Рис. 22. Вольт-амперные характеристики мощного (нормально открытого) *n*-канального МП-транзистора [48].



Решение этого уравнения определяет минимальную глубину обедненного слоя в канале у стока $u_{\text{мин}}$. Максимальный ток канала определяется из выражения (45):

$$I_m = I_S(1 - u_{\text{мин}}). \tag{65}$$

На рис. 23 приведены теоретические зависимости [26] глубины канала от концентрации N_D для некоторых значений длины канала L и отношения I_m/Z (максимальный ток на единицу ширины

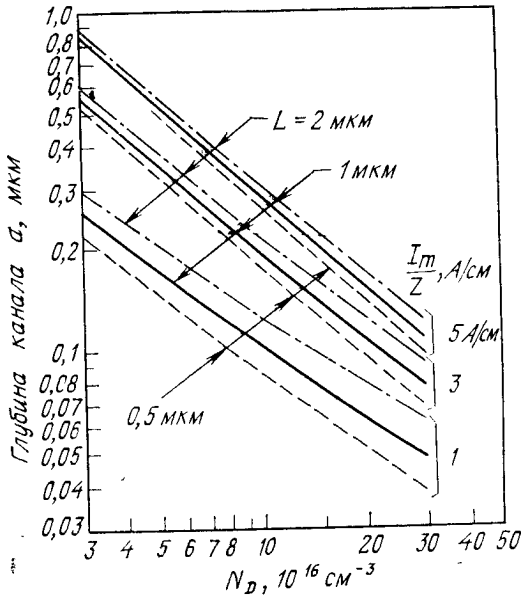


Рис. 23. Расчетные зависимости глубины канала от уровня легирования для GaAs МП-транзисторов с различной длиной затвора и плотностью максимального тока на единицу ширины канала [26].

рины канала). Из них следует, что при заданных L и I_m/Z глубина канала a изменяется примерно по закону N_D^{-1} . Так, при $L = 1$ мкм, $N_D = 10^{17}$ см $^{-3}$ и $a = 0,2$ мкм максимальный ток на единицу ширины канала составляет 3 А/см.

Для увеличения напряжения пробоя V_B нужно снижать уровень легирования канала N_D , что, однако, при заданных значениях L и a приводит к уменьшению максимального тока I_m . Поэтому при оптимизации выходной мощности приходится искать компромиссные решения. При этом следует учитывать, что для мелких каналов, в которых полный заряд на единицу площади $Q_c = N_D a$ не превышает $2 \cdot 10^{12}$ см $^{-2}$, из-за двумерного распределения электрического поля напряжение пробоя значительно превышает соответствующее пробивное напряжение плоского (одномерного) $p^+ - n$ -перехода. С учетом двумерных эффектов для напряжения пробоя в работе [47] было получено следующее выражение:

$$V_B = 5 \cdot 10^{13} / Q_C \text{ (см}^{-2}\text{)} \text{ (В)}. \quad (66)$$

При работе с высокими уровнями мощности приборы разогреваются. С ростом температуры уменьшаются подвижность электронов ($\sim [T \text{ (К)}]^{-2}$) и скорость насыщения ($v_s \sim [T \text{ (К)}]^{-1}$) [27]. Поэтому каналные полевые транзисторы имеют отрицательный температурный коэффициент, что обуславливает их сравнительно высокую термостабильность при работе с высокими уровнями мощности. Соотношение между предельной мощностью и граничной частотой МП-транзисторов из GaAs [28] иллюстрируют данные,

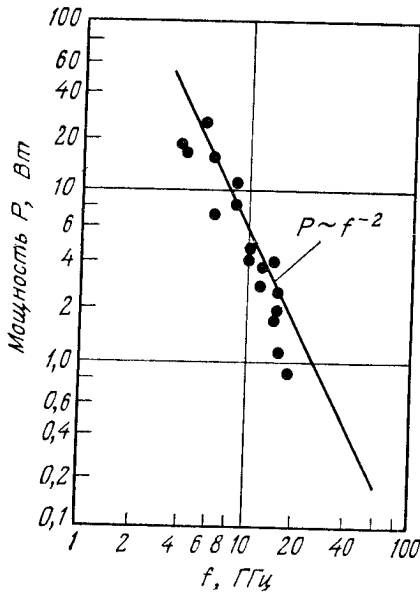


Рис. 24. Соотношение между предельной мощностью и граничной частотой для современных МП-транзисторов из GaAs [28].

приведенные на рис. 24. Предельная мощность изменяется пропорционально f^{-2} , что характерно для электронных ограничений, которые более подробно рассмотрены в гл. 10.

Очевидно, что миниатюризация, освоение субмикронных размеров, уменьшение паразитных сопротивлений и емкостей и т. д. приведут к дальнейшему улучшению частотных и мощностных характеристик канальных полевых транзисторов.

6.4.3. Шумовые характеристики

Полевые транзисторы с $p-n$ -переходом и МП-транзисторы в принципе являются сравнительно малошумящими приборами, поскольку их функционирование обуславливают только основные носители тока. Поэтому обычно главными источниками шума в реальных приборах оказываются внешние паразитные сопротивления, которые, к сожалению, не удается исключить полностью.

На рис. 25 показана эквивалентная шумовая схема прибора [7]. Источники шума i_{ng} , i_{nd} , e_{ng} и e_{ns} соответствуют индуцированному шуму затвора, шуму в цепи стока, тепловому шуму последовательного паразитного сопротивления истока. На этой схеме e_s — источник сигнала, а Z_S — импеданс входной цепи. В прямоугольнике, очерченном штриховой линией, сосредоточены собственно транзисторные элементы. Оптимальный шум-фактор, полученный с помощью этой эквивалентной схемы, для МП-транзистора из GaAs [29] определяется выражением

$$F_0 \simeq 1 + fL \sqrt{g_m(R_S + R_G)/4}. \quad (67)$$

Очевидно, что для улучшения шумовых характеристик следует уменьшать длину канала и снижать паразитные сопротивления

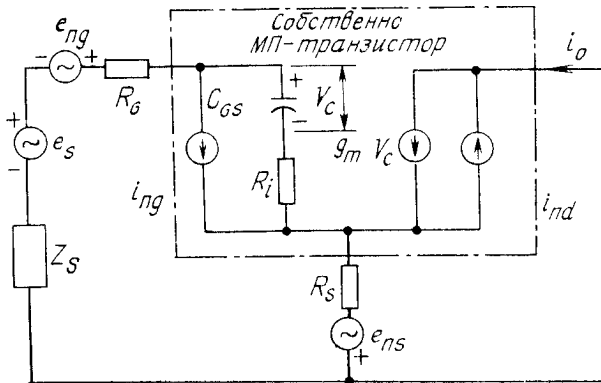


Рис. 25. Эквивалентная шумовая схема полевого транзистора с $p-n$ -переходом [7].

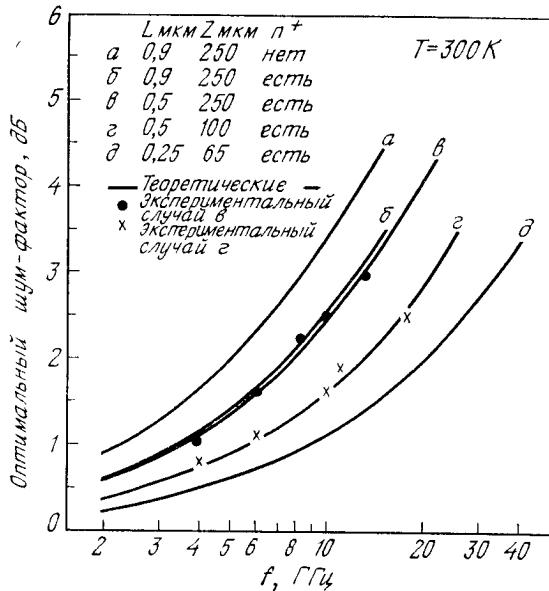


Рис. 26. Теоретические зависимости оптимального шум-фактора от частоты для МП-транзисторов с различной длиной и шириной канала и результаты соответствующих измерений при $T = 300 \text{ K}$ [29, 30].

затвора и истока. На рис. 26 приведены экспериментальные [29, 30] и теоретические значения шум-фактора. При фиксированной частоте шум уменьшается с уменьшением длины затвора. Отметим, что шум уменьшается и при сокращении ширины канала, поскольку при этом уменьшается сопротивление металлического электрода затвора.

Приборы с плавно легированным каналом [15] (рис. 15) оказались менее шумящими по сравнению с приборами тех же размеров и однородно легированным каналом (соответствующее уменьшение шума от 1 до 3 Дб). Это различие в шумах связано с крутизной g_m , которая входит в выражение (67) для шум-фактора. Поэтому некоторое уменьшение g_m (но не g_m/C_{GS}) в случае плавно легированного канала дает улучшение шумовых характеристик.

6.4.4. Конструкции приборов

Несколько типичных конструкций мощных МП-транзисторов для СВЧ-диапазона [48] схематически показаны на рис. 27. Все приборы изготовлены на полужолирующих подложках и имеют промежуточный (буферный) эпитаксиально выращенный слой, предназначенный для уменьшения влияния сильно дефектной

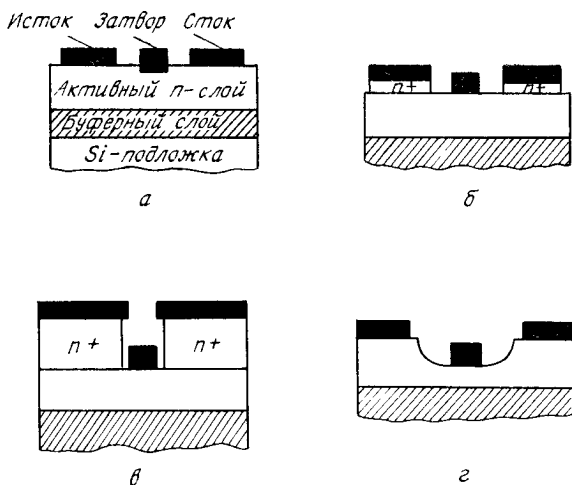


Рис. 27. Различные способы формирования контактов стока и истока в МП-транзисторах из GaAs [48].

полуизолирующей подложки на проводимость канала. На рис. 27, а показан простейший планарный прибор с вплавленными контактами. Поскольку омические контакты стока и истока оказывают существенное влияние на характеристики и надежность приборов [31], на практике используются различные способы формирования стока и истока, направленные на снижение последовательных сопротивлений этих контактов и повышение пробивных напряжений стока [49]. В конструкции, приведенной на рис. 27, б, для этого использован эпитаксиальный n^+ -слой, который затем частично вытравливается при формировании затвора. Аналогичная, но со значительно более толстым n^+ -слоем структура показана на рис. 27, в. В процессе ее изготовления металлизация стока и истока используется в качестве маски при формировании затвора (самосовмещение). На рис. 27, г показана структура с заглубленным затвором. Она характеризуется повышенными пробивными напряжениями за счет уменьшения электрических полей в окрестности стока. На рис. 28 показана мощная МП-структура из GaAs, общий затвор которой одновременно выполняет функцию теплоотвода [32]. Верхний слой металлизации представляет собой секционированные электроды стока и истока, «вставленные» друг в друга. Вне активной области прибора GaAs превращен с помощью протонной бомбардировки в полуизолирующий материал.

Для увеличения рабочих частот и снижения шумов предложено большее число способов формирования затворов МП-транзисторов. На рис. 29, а показана структура, в которой под затвором ионной бомбардировкой (Ar^+) сформирована промежуточная полуизоли-

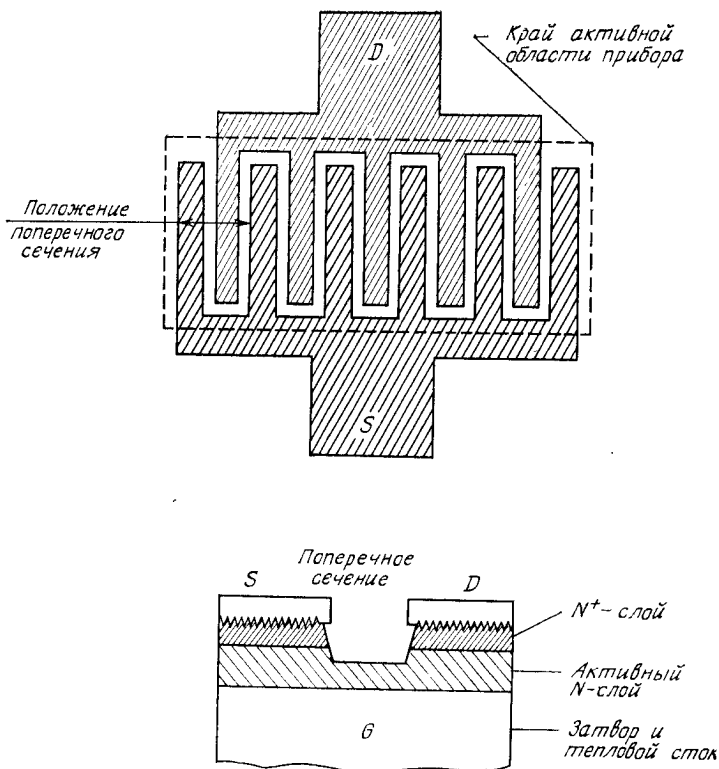


Рис. 28. Мощный МП-транзистор с подложкой, одновременно выполняющей роль затвора и теплового стока, и секционированными стоком и истоком [32].

рующая область [33]. При этом уменьшаются емкость затвора и его токи утечки, а также увеличивается напряжение пробоя. На рис. 29, б приведена аналогичная структура с буферным слоем между затвором и каналом [34].

На рис. 29, в показан полевой транзистор с гетеропереходом [35]. Здесь роль подзатворной буферной области играет слой более широкозонного, чем GaAs, полупроводника AlGaAs, что значительно уменьшает ток утечки затвора по сравнению с обычным барьером Шоттки. Для создания приборов с субмикронной длиной затвора в настоящее время широко используются различные способы самосовмещения. На рис. 29, г показан МП-транзистор с двумя затворами [36]. Этот прибор, как и гетероструктура на рис. 29, в, перспективен для использования в малощумящих схемах.

В работе [50] исследован полевой МП-транзистор с двойной гетероструктурой, поперечное сечение которого показано на

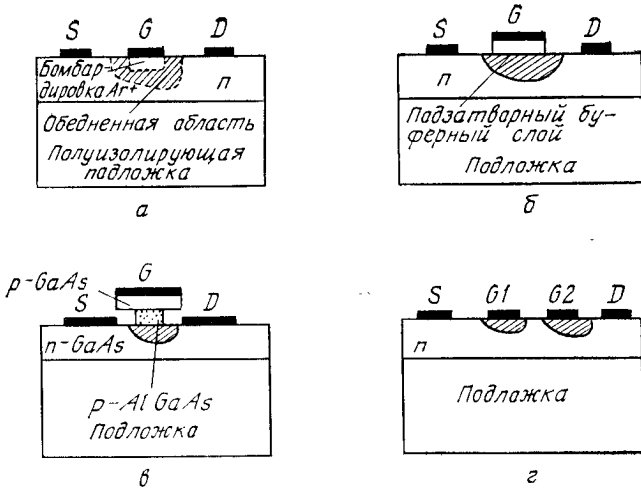


Рис. 29. Различные конфигурации затворов, используемые в высококачественных полевых транзисторах.

рис. 30, а. Активный слой прибора представляет собой тройное соединение $Ga_{0,47}In_{0,53}As$. Все полупроводниковые и металлические слои структуры последовательно нанесены на полуизолирующую InP-подложку с ориентацией $\langle 100 \rangle$ методом молекулярно-лучевой эпитаксии [50]. Эти слои хорошо согласованы с точки зрения постоянной решетки между собой и с InP-подложкой, что обуславливает низкую плотность поверхностных ловушек на соответствующих границах раздела. На рис. 30, б показана зонная схема рассматриваемой двойной гетероструктуры в условиях равновесия.

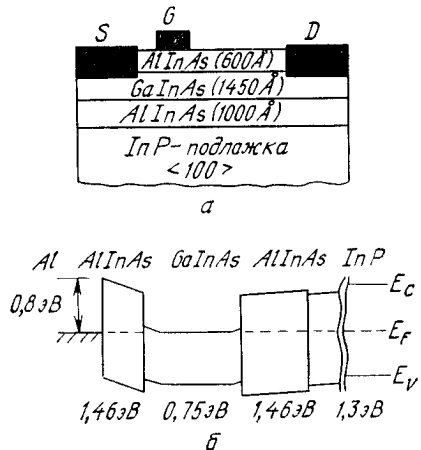


Рис. 30. Условное поперечное сечение МП-транзистора с двойной гетероструктурой (а) и зонная диаграмма прибора (б) [50].

Верхний полупроводниковый слой $Al_{0,48}In_{0,52}As$ образует барьер Шоттки с внешним алюминиевым затвором ($\Phi_{Вн} = 0,8$ В). Подвижные электроны здесь сосредоточены в активном узкозонном слое $Ga_{0,17}In_{0,53}As$. Поскольку подвижность при слабых полях и пиковая дрейфовая скорость в этом материале выше, чем в GaAs, в рассматриваемой двойной гетероструктуре можно получить большую крутизну. Такие МП-транзисторы весьма перспективны для применения в быстродействующих схемах.

6.5. ДРУГИЕ ПОЛЕВЫЕ ПРИБОРЫ

6.5.1. Ограничители (стабилизаторы) тока

Ниже рассмотрены два типа токорегулирующих диодов, представляющих собой двухполюсные полевые приборы: полевой диод и диод с насыщением скорости.

Полевой диод представляет собой обычный полевой транзистор с затвором, закороченным на исток [37, 38]. Его вольт-амперная характеристика (рис. 31) подобна характеристике полевого транзистора при $V_G = 0$ (рис. 3). Полевой диод как стабилизатор тока характеризуется четырьмя основными рабочими параметрами: стабилизируемым током I_l , напряжением насыщения V_{sat} , наклоном характеристики на участке стабилизации g_l и напряжением пробоя V_B . Каждый из этих параметров был рассмотрен применительно к работе полевого транзистора. Отметим здесь только, что для уменьшения V_{sat} следует использовать приборы с малой глубиной канала a и малым уровнем легирования канала N_D (выражение (4)). При этом уменьшается и стабилизируемый ток I_l . Кроме того, для снижения I_l можно уменьшать отношение Z/L (ширины к длине канала). Для уменьшения g_l следует увеличивать длину канала L , а для увеличения напряжения пробоя снижать уровень легирования канала.

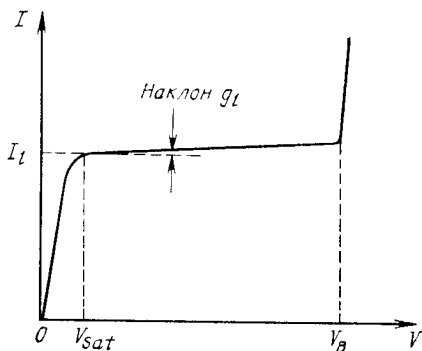
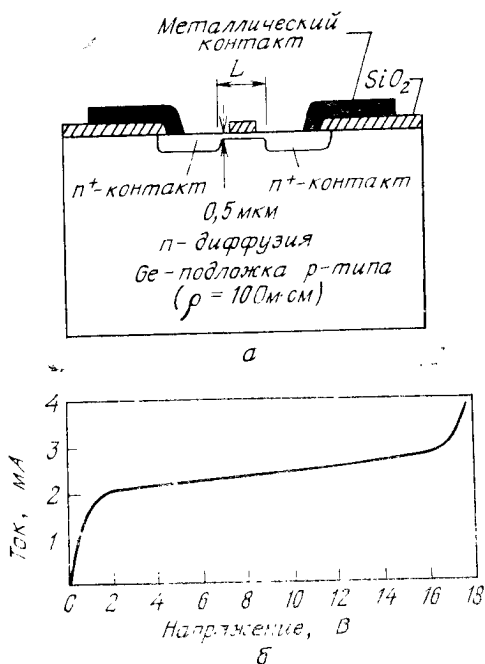


Рис. 31. Вольт-амперная характеристика стабилизатора тока (полевого диода).

Рис. 32. Конструкция диода с насыщением скорости (а) и экспериментальная вольт-амперная характеристика (б) [39].



В диодах с насыщением дрейфовой скорости участок стабилизации тока на вольт-амперной характеристике возникает в результате эффекта насыщения дрейфовой скорости носителей тока в сильных электрических полях [39]. Структура такого прибора показана на рис. 32, а. Областью с высоким электрическим полем здесь служит мелкий ($\sim 0,5$ мкм) диффузионный n -слой длиной 3 мкм. В качестве материала структуры выбран германий, так как в нем насыщение дрейфовой скорости начинается при сравнительно малых (~ 4 КВ/см) электрических полях. В кремнии, например, дрейфовая скорость начинает насыщаться только при 30 КВ/см.

Обратимся снова к показанным на рис. 31 четырем главным параметрам, характеризующим стабилизатор тока, и рассмотрим их применительно к диодам с насыщением скорости. Стабилизируемый ток в этом случае равен

$$I_t = qN_D v_s A + I_S, \quad (68)$$

где v_s — скорость насыщения, A — площадь канала, I_S — ток насыщения обратносмещенного p — n -перехода. С ростом температуры ток насыщения увеличивается, а скорость насыщения падает. Следовательно, когда эти два конкурирующих механизма полностью компенсируют друг друга, величина стабилизированного тока практически не зависит от температуры.

Напряжение насыщения V_{sat} определяется соотношением

$$V_{\text{sat}} = \mathcal{E}_s L + I_1 R_c, \quad (69)$$

где \mathcal{E}_s — электрическое поле, соответствующее началу насыщения дрейфовой скорости; R_c — последовательное паразитное сопротивление контактов. В идеальном стабилизаторе тока $V_{\text{sat}} = 0$, поэтому на практике стремятся по возможности уменьшать V_{sat} .

Наклон вольт-амперной характеристики диода на участке стабилизации тока g_I обусловлен двумя эффектами. Во-первых, дрейфовая скорость носителей тока на участке насыщения зависит (хотя и слабо) от величины электрического поля. Второй причиной, обуславливающей положительное g_I , является инжекция носителей из контакта в канал. Она приводит к появлению тока, ограниченного пространственным зарядом в канале. Как показано в гл. 10, этот механизм характеризуется сопротивлением $L^2/2\epsilon_s \cdot v_s \cdot A$.

Пробой диода обусловлен ударной ионизацией в канале. Напряжение пробоя определяется выражением

$$V_B \approx \mathcal{E}_m L/2, \quad (70)$$

поскольку напряженность электрического поля линейно увеличивается по длине канала и достигает характерного поля ударной ионизации \mathcal{E}_m на положительном контакте. Для германия $\mathcal{E}_m \approx \approx 1,5 \cdot 10^6$ В/см, поэтому для прибора с длиной канала 3 мкм $V_B \approx 20$ В.

Экспериментальная вольт-амперная характеристика стабилизирующего диода с насыщением скорости приведена на рис. 30, б. Следует отметить, что диоды этого типа обладают большим быстродействием, чем аналогичные полевые диоды. Дело в том, что при одинаковом токе стабилизации I_1 ширина обедненного слоя в диоде с насыщением скорости может быть существенно больше, чем в соответствующем полевом диоде. При этом шунтирующая емкость диода с насыщением скорости будет много меньше входной емкости аналогичного полевого диода.

6.5.2. Полевые транзисторы с V-канавками

Полевой транзистор с V-канавками не является планарной структурой, причем именно эта «непланарность» используется для оптимизации характеристик приборов [40, 41]. При этом достигаются большие значения крутизны и меньшие сопротивления в открытом состоянии, чем в аналогичных планарных транзисторных структурах. Поперечное сечение одного из таких приборов показано на рис. 33, а. Здесь использовано три типа V-канавок: сравнительно мелкие (для стока и истока), средняя (для затвора) и глубокие (для изоляции отдельных приборов друг от друга). Эта структура сформирована на эпитаксиальном слое

n-кремния с ориентацией $\langle 100 \rangle$, выращенном на *p*-подложке. V-канавки создаются в процессе анизотропного травления. Стенка V-канавки составляет с горизонталью $54,7^\circ$. Другая модификация полевого транзистора с V-канавками показана на рис. 33, б. Здесь V-канавка используется для сужения канала. Затвором служит высоколегированная p^+ -подложка, а контакты стока и истока содержат промежуточный эпитаксиальный n^+ -слой. При нулевом напряжении на стоке V_D (рис. 34, а) проводимость канала равна

$$g_D = \frac{qN_D\mu a_1 Z}{L_1} \left[1 + \left(\frac{V_G}{V_P} \right)^{1/2} \right], \quad (71)$$

где a_1 — расстояние от острия затворной V-канавки до *p*-подложки, N_D и N_A — концентрации легирующей примеси в *n*-канале и *p*-подложке, L_1 — эффективная длина канала. Поскольку

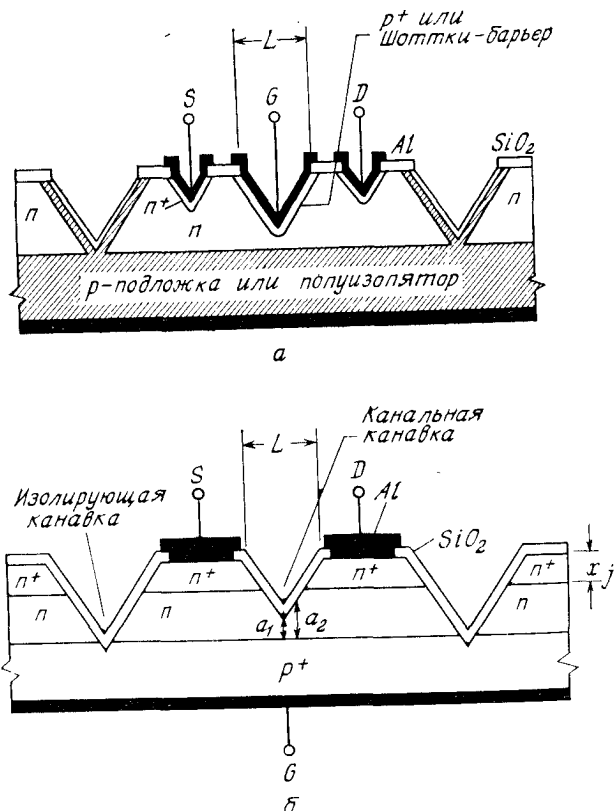


Рис. 33. Полевой транзистор с V-канавкой в затворе (а) и полевой транзистор с V-канавкой в канале (б) [41].

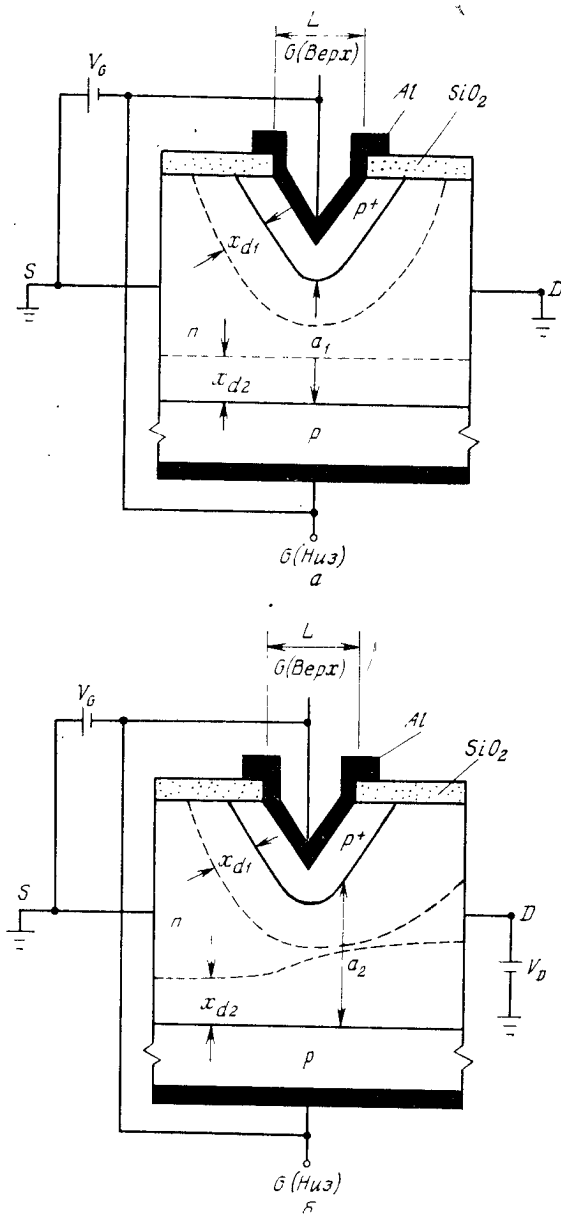


Рис. 34. Работа полевого транзистора с V-канавкой [41].
 а — при $V_D = 0$; б — в режиме насыщения [41].

длина L_1 меньше ширины затворной V-канавки L , g_D в такой структуре больше, чем в обычном планарном полевом транзисторе с той же геометрией. Отметим, что в приборах с V-канавками отсечка канала в режиме насыщения происходит вблизи центра канала (рис. 34, б), а не у стока, как в обычном полевом транзисторе. Эффективная длина канала в режиме насыщения L_3 также меньше ширины канавки L . Это обуславливает повышенную крутизну прибора ($\sim 1/L_2$).

6.5.3. Многоканальные полевые транзисторы

Для повышения уровня переключаемых мощностей предложены и изучены так называемые многоканальные полевые транзисторы [42—44]. Эти приборы фактически представляют собой несколько

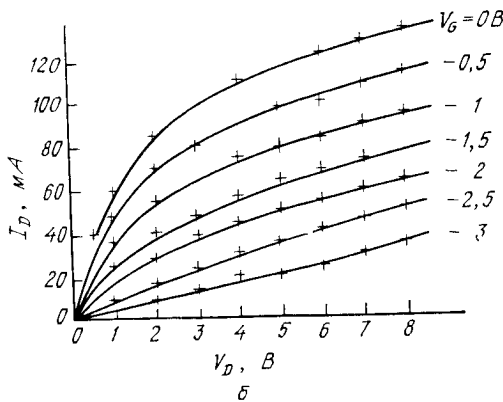
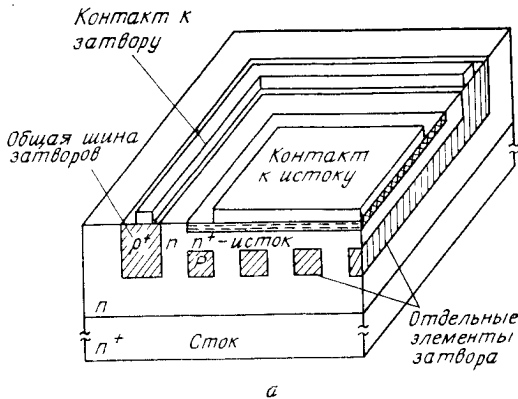


Рис. 35. Изометрическое изображение многоканального полевого транзистора (а) и экспериментальные вольт-амперные характеристики (б) прибора с $N_D = 10^{16} \text{ см}^{-3}$ [45].

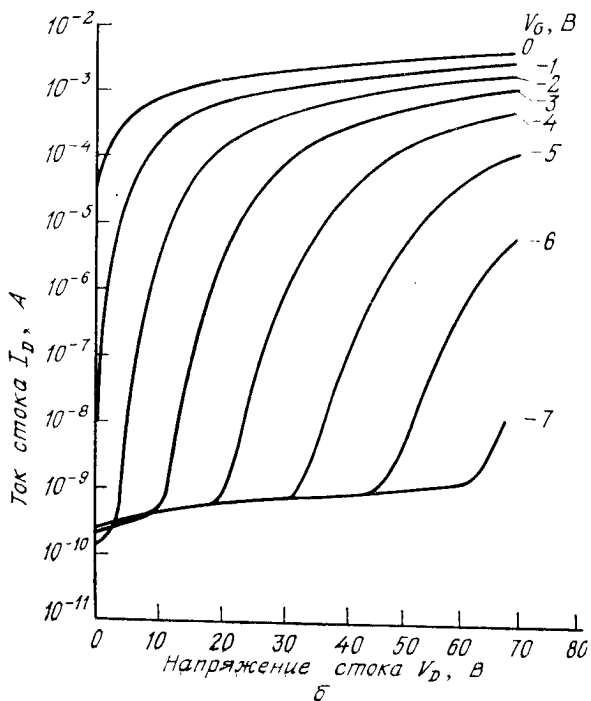
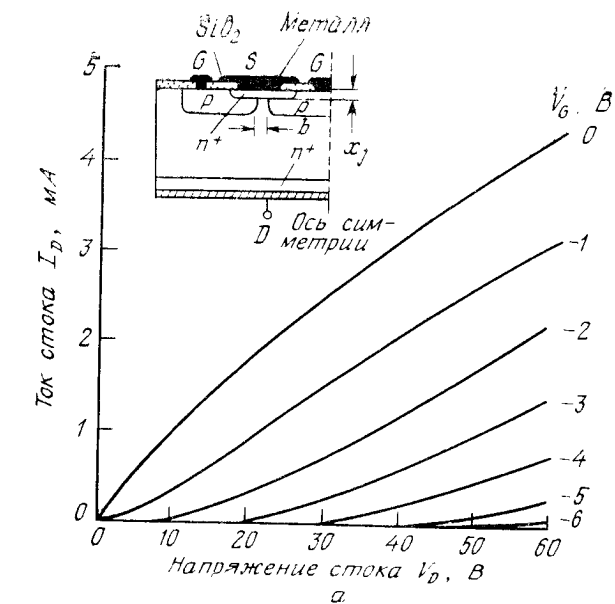
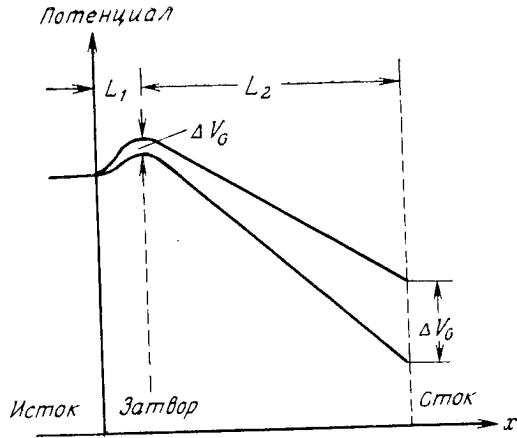


Рис. 36. Экспериментальные вольт-амперные характеристики многоканального полевого транзистора с $N_D = 10^{14} \text{ см}^{-3}$ в линейном (а) и полулогарифмическом масштабе (б). На вставке показано поперечное сечение отдельной ячейки прибора [46].

Рис. 37. Распределение потенциала между стоком и истоком вертикального полевого транзистора. При $V_G = 0$ канал перекрыт встроенным потенциалом.



отдельных полевых транзисторов, соединенных параллельно (рис. 35, а), поэтому, как и следовало ожидать, их вольт-амперные характеристики подобны характеристикам обычного одноканального полевого транзистора (рис. 35, б). Характерный уровень токов в этом случае значительно выше, чем для отдельного прибора [45].

Однако в тех случаях, когда уровень легирования активной области достаточно низкий и канал уже при нулевом напряжении на затворе перекрыт встроенным потенциалом V_{bi} , вольт-амперные характеристики отличаются от обычных [44]. В качестве примера семейства таких триодноподобных характеристик [46] приведено на рис. 36, а. На вставке показана структура прибора (вертикальный полевой транзистор) с уровнем легирования 10^{14} см^{-3} и расстоянием между затворами 4 мкм. На рис. 36, б те же характеристики изображены в полулогарифмическом масштабе. Из графиков, построенных на рис. 36, а в линейном масштабе, следует, что при больших уровнях тока вольт-амперные характеристики прибора триодного типа, т. е. ток стока увеличивается линейно с напряжением стока V_D . При малых уровнях тока стока увеличивается с ростом напряжения V_D экспоненциально (рис. 36, б). Причина такой экспоненциальной зависимости состоит в том, что в данном случае напряжение на стоке понижает энергетический барьер в канале, образовавшийся вследствие смыкания обедненных областей от обоих затворов (рис. 37). Очевидно, что такое понижение барьера $\Delta V_G \approx (L_1/L_2)/\Delta V_D$, где ΔV_D — соответствующее изменение напряжения стока V_D . Поэтому ток прибора, определяемый надбарьерной термоэлектронной инжекцией, пропорционален

$$I \sim \exp\left(\frac{q\Delta V_G}{kT}\right) \approx \exp\left(\frac{qI_1 \Delta V_D}{kTL_2}\right), \quad (72)$$

что и объясняет наблюдаемый экспоненциальный характер вольт-амперных характеристик прибора в области малых токов (рис. 36, б). При больших токах плотность электронов в канале между затворами становится больше концентрации фиксированного заряда доноров, и в этом случае основную роль начинают играть ограничения тока пространственным зарядом подвижных носителей, что обуславливает примерно линейный характер вольт-амперных характеристик прибора в этой области.

ЛИТЕРАТУРА

1. Shockley W. A Unipolar Field-Effect Transistor, *Proc. IRE*, **40**, 1365 (1952).
2. Dacey G. C., Ross I. M. Unipolar Field-Effect Transistor, *Proc. IRE*, **41**, 970 (1953).
3. Dacey G. C., Ross I. M. The Field-Effect Transistor, *Bell Syst. Tech. J.*, **34**, 1149 (1955).
4. Mead C. A. Schottky Barrier Gate Field-Effect Transistor, *Proc. IEEE*, **54**, 307 (1966).
5. Hooper W. W., Lehrer W. I. An Epitaxial GaAs Field-Effect Transistor, *Proc. IEEE*, **55**, 1237 (1967).
6. Hauser J. R. Junction Field Effect Transistors, in Burger R. M., Donovan R. P., Eds., *Fundamental of Silicon Integrated Device Technology*, Vol. 2, Prentice-Hall, Englewood Cliffs, N. J., 1968, Chap. 3.
7. Pucel R. A., Haus H. A., Stalz H. Signal and Noise Properties of GaAs Microwave Field-Effect Transistors, in Martin L., Ed., *Advances in Electronics and Electron Physics*, Vol. 38, Academic, N. Y., 1975, p. 195.
8. Liechti C. A. Microwave Field-Effect Transistors — 1976, *IEEE Trans. Microwave Theory Tech.*, **MTT-24**, 279 (1976).
9. Brokemuehl R. R. Analysis of Field-Effect Transistors with Arbitrary Charge Distribution, *IEEE Trans. Electron Devices*, **ED-10**, 31 (1963).
10. Middlebrook R. D., Richer I. Limits on the Power-Law Exponent for Field-Effect Transistor Transfer Characteristics, *Solid State Electron.*, **6**, 542 (1963).
11. Zuleeg R., Notthoff J. K., Lehovec K. Femtojoule High-Speed Planar GaAs E-JFET Logic, *IEEE Trans. Electron Devices*, **ED-25**, 628 (1978).
12. Jacoboni C., Canali C., Ottaviani G., Quaranta A. A. A Review of Some Charge Transport Properties of Silicon, *Solid State Electron.*, **20**, 77 (1977).
13. Smith P., Inoue M., Frey J. Electron Velocity in Si and GaAs at Very High Electric Fields, *Appl. Phys. Lett.*, **37**, 797 (1980).
14. Lehovec K., Zuleeg R. Voltage — Current Characteristics of GaAs JFETs in the Hot Electron Range, *Solid State Electron.*, **13**, 1415 (1970).
15. Williams R. E., Shaw D. W. Graded Channel FET's Improved Linearity and Noise Figure, *IEEE Trans. Electron Devices*, **ED-25**, 600 (1978).
16. Williams R. E., Shaw D. W. GaAs FET's with Graded Channel Doping Profiles, *Electron. Lett.*, **13**, 408 (1977).
17. Pucel R. A. Profile Design for Distortion Reduction in Microwave Field-Effect Transistors, *Electron. Lett.*, **14**, 204 (1978).
18. Kennedy D. P., O'Brien R. R. Computer Aided Two-Dimensional Analysis of the Junction Field-Effect Transistor, *IBM J. Res. Dev.*, **14**, 95 (1970).
19. Lehovec K., Miller R. Field Distribution in Junction Field Effect Transistors at Large Drain Voltages, *IEEE Trans. Electron Devices*, **ED-22**, 273 (1975).

20. Himsforth B. A Two-Dimensional Analysis of GaAs Junction Field-Effect Transistors with Long and Short Channels, *Solid State Electron.*, **15**, 1353 (1972).
21. Ruch J. Electron Dynamics in Short Channel Field-Effect Transistors, *IEEE Trans. Electron Devices*, **ED-19**, 652 (1972).
22. Wada T., Frey J. Physical Basis of Short-Channel MESFET Operation, *IEEE Electron Devices*, **ED-26**, 476 (1979).
23. Reiser M., Wolf P. Computer Study of Submicrometer FET's, *Electron. Lett.*, **8**, 254 (1972).
24. Maloney T. J., Frey J. Frequency Limits of GaAs and InP Field-Effect Transistors at 300 K and 77 K with Typical Active Layer Doping, *IEEE Trans. Electron Devices*, **ED-23**, 519 (1976).
25. Barrera J. S., Archer R. J. InP Schottky-Gate Field-Effect Transistors, *IEEE Trans. Electron Devices*, **ED-22**, 1023 (1975).
26. Fukui H. Drain Current Limitations and Temperature Effects in GaAs MESFET, *IEEE Tech. Dig. Int. Electron Device Meet.* 1978, pp. 140—143.
27. Sevin L. J. Field Effect Transistors, McGraw-Hill, N. Y., 1965.
28. Ikoma T. Status of Microwave and High Speed Devices, Proc. 7th Bienn. Cornell Electr. Eng. Conf. Active Microwave Semicond. Devices and Circuits, Cornell University, Ithaca, N. Y., p. 7 (1979).
29. Fukui H. Optimal Noise Figure of Microwave GaAs MESFETs, *IEEE Trans. Electron Devices*, **ED-26**, 1032 (1979).
30. Hewitt B. S., Cox H. M., Fukui H., DiLorenzo J. V., Schlosser W. O., Iglesias D. E. Low Noise GaAs MESFETs, *Electron. Lett.*, **12**, 309 (1976).
31. Mizuishi K., Kurono H., Sato H., Kodera H. Degradation Mechanics of GaAs MESFETs, *IEEE Trans. Electron Devices*, **ED-26**, 1008 (1979).
32. Blocker T., Macksey H., Adams R. X-Band RF Power Performance of GaAs FET's, *IEEE Tech. Dig. Int. Electron Device Meet.*, 1974, pp. 288—291.
33. Macksey H. M., Shaw D. W., Wisseman W. R. GaAs Power FETs with Semi-Insulating Gates, *Electron. Lett.*, **12**, 192 (1976).
34. Nagashima A., Umehachi S., Kano G. Calculation of Microwave Performance of Buffer Layer Gate GaAs MESFETs, *IEEE Trans. Electron Devices*, **ED-25**, 537 (1978).
35. Umehachi S., Ashahi K., Inoue M., Kano G. A New Heterojunction Gate GaAs FET, *IEEE Trans. Electron Devices*, **ED-12**, 613 (1975).
36. Turner J. A., Waller A. J., Kelly E., Parker D. Dual-Gate GaAs Microwave Field-Effect Transistor, *Electron. Lett.*, **7**, 661 (1971).
37. Warner R. M., Jackson W. H., Doucette E. I., Sone H. A. A Semiconductor Current Limiter, *Proc. IRE*, **47**, 45 (1959).
38. Lawrence H. A Diffused Field-Effect Current Limiter, *IRE Trans. Electron Devices*, **ED-9**, 82 (1962).
39. Boll H. J., Iwersen J. E., Perry E. W. High-Speed Current Limits, *IEEE Trans. Electron Devices*, **ED-13**, 904 (1966).
40. Salama C. A., Oakes J. G. Nonplanar Power Field-Effect Transistors, *IEEE Trans. Electron Devices*, **ED-25**, 1222 (1978).
41. Mok T. K., Salama C. A. T. The Characteristics and Application of a V-Shaped Notched Channel Field-Effect Transistor (VFET), *Solid State Electron.*, **19**, 159 (1976).
42. Teszner S., Gicquel R. Gridistor — A New Field Effect Device, *Proc. IEEE*, **52**, 1502 (1964).
43. Zulleg R. Multichannel Field-Effect Transistor Theory and Experiment, *Solid State Electron.*, **10**, 559 (1967).
44. Nishizawa J. I., Terasaki T., Shibata J. Field Effect Transistor Versus Analog Transistor (Static Induction Transistor), *IEEE Trans. Electron Devices*, **ED 22**, 185 (1975).

45. Lecrosnier D. R., Pelous G. P. Ion Implanted FET for Power Applications, *IEEE Trans. Electron Devices*, **ED-21**, 113 (1974).
46. Morenza J. L., Esteve D. Entirely Diffused Vertical Channel JFET: Theory and Experiment, *Solid State Electron*, **21**, 739 (1978).
47. Wemple S. H., Niehaus W. C., Cox H. M., DiLorenzo J. V., Schlosser W. O. Control of Gate — Drain Avalanche in GaAs MESFETs, *IEEE Trans. Electron Devices*, **ED-27**, 1013 (1980).
48. DiLorenzo J. H., Wisseman W. R. GaAs Power MESFET: Design, Fabrication and Performance, *IEEE Trans. Microwave Theory Tech.*, **MIT-27**, 367 (1979).
49. Fukuta M., Syama K., Suzuki H., Nakayama K., Ishidawa H. Power GaAs MESFET with High Drain — Source Breakdown Voltage, *IEEE Trans. Microwave Theory Tech.*, **MTT-24**, 312 (1976).
50. Narnard J., Ohno H., Wood C. E. C., Eastman L. F. Double Heterostructure $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ MESFETs with Submicron Gates, *IEEE Electron Devices, Lett.*, **EDL-1**, 174 (1980).

Глава 7

МДП-СТРУКТУРЫ. ПРИБОРЫ С ЗАРЯДОВОЙ СВЯЗЬЮ

7.1. ВВЕДЕНИЕ

Структуры металл—диэлектрик—полупроводник (МДП-структуры) имеют большое значение при изучении свойств поверхности полупроводников. Поскольку различные поверхностные эффекты непосредственно влияют на надежность и стабильность работы практически всех типов полупроводниковых приборов, изучение физики поверхности с помощью МДП-структур играет большую роль во всей полупроводниковой технологии. В этой главе рассматриваются преимущественно структуры металл—окисел—полупроводник (кремний) (МОП-структуры), которые являются основным элементом конструкции большинства современных планарных приборов и интегральных схем и наиболее хорошо изучены в настоящее время.

МДП-структуры были предложены в качестве управляемой напряжением емкости впервые в 1959 г. Моллом [1], Пфанном и Гарретом [2]. Затем характеристики МДП-структур анализировали Франкл [3] и Линднер [4]. Для изучения термически окисленной поверхности кремния МОП-структуры были впервые использованы Терманом [5], Леховеком и Слободским [6]. Обстоятельное и глубокое изложение принципов физики Si—SiO₂-структур можно найти в книге Никколиана и Брюса [7].

Принцип зарядовой связи был сформулирован Бойлом и Смитом [8] в 1970 г. Первый прибор с зарядовой связью (ПЗС) сконструировали Амелио, Томпсет и Смит [9]. В наиболее простом виде прибор с зарядовой связью можно рассматривать как матрицу близко расположенных МДП-конденсаторов. При соответствующей последовательности тактовых импульсов напряжения на электродах МДП-конденсаторов в ПЗС могут контролируемым образом перемещаться вдоль поверхности полупроводника так называемые зарядовые пакеты. Используя этот принцип, с помощью ПЗС можно осуществить целый ряд функций: прием изображений, хранение данных, обработку сигналов, логические операции. Обстоятельное изложение физических принципов и механизмов работы ПЗС содержится в монографиях Секена и Томпсета [10] и Кима [11]. Важнейшие оригинальные статьи по ПЗС можно найти в журнале ТИИЭР [12].

7.2. ИДЕАЛЬНАЯ МДП-СТРУКТУРА

Схематически МДП-структура приведена на рис. 1, где d — толщина слоя диэлектрика, а V — напряжение, приложенное к металлическому полевому электроду. Ниже мы будем считать напряжение положительным, если металлический электрод структуры смещен положительно относительно омического контакта к подложке, и отрицательным в противоположном случае.

Зонные диаграммы идеальных МДП-структур при $V = 0$ приведены на рис. 2. Понятие «идеальная МДП-структура» определим следующим образом:

1. Работы выхода электронов из металла и полупроводника одинаковы, или разность работ выхода металла и полупроводника Φ_{ms} равна нулю:

$$\Phi_{ms} \equiv \Phi_m - \left(\chi + \frac{E_g}{2q} - \psi_B \right) = 0 \text{ для } n\text{-типа,} \quad (1a)$$

$$\Phi_{ms} \equiv \Phi_m - \left(\chi + \frac{E_g}{2q} + \psi_B \right) = 0 \text{ для } p\text{-типа.} \quad (1б)$$

Здесь Φ_m — работа выхода металла, χ — сродство к электрону полупроводника, E_g — ширина запрещенной зоны, ψ_B — разность между уровнем Ферми E_F и положением уровня Ферми в собственном полупроводнике E_i . Условия (1) означают, что в отсутствие внешнего напряжения $V = 0$ энергетические зоны полупроводника не изогнуты (состояние плоских зон).

2. При любых смещениях в структуре могут существовать только заряд в ее полупроводниковой части и равный ему заряд противоположного знака на металлическом электроде, отделенном от полупроводника слоем диэлектрика.

3. При постоянном напряжении смещения отсутствует перенос носителей тока сквозь диэлектрик, т. е. сопротивление диэлектрика предполагается бесконечным.

Теория идеальной МДП-структуры, рассматриваемая в этом разделе, служит основой для понимания и исследования свойств реальных МДП-структур и поверхности полупроводников.

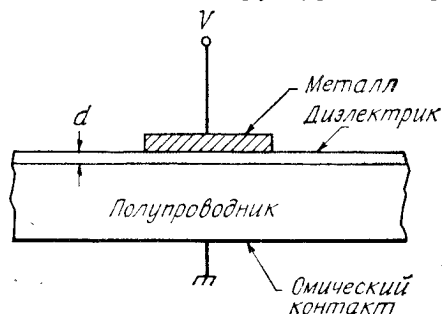
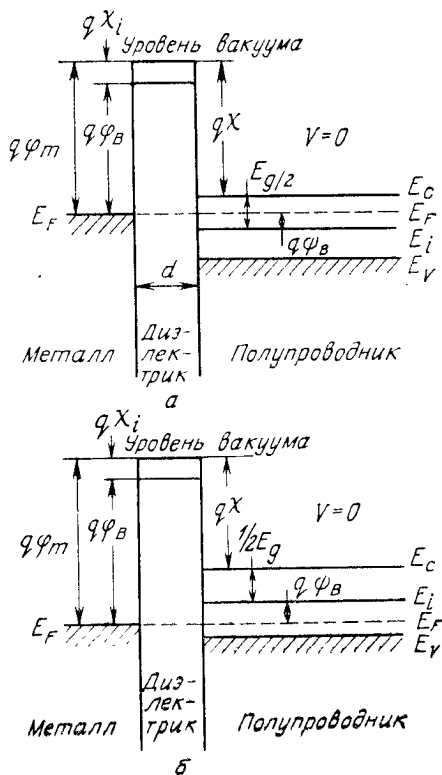


Рис. 1. Структура металл — диэлектрик — полупроводник (МДП-структура).

Рис. 2. Зонные диаграммы идеальных МДП-структур при $V = 0$.

a — полупроводник n -типа; b — полупроводник p -типа.



Когда к идеальной МДП-структуре приложено напряжение того или другого знака, на полупроводниковой поверхности могут возникнуть три основные ситуации (рис. 3). Рассмотрим их сначала для МДП-структуры с полупроводником p -типа. Если к металлическому электроду структуры приложено отрицательное напряжение ($V < 0$), край валентной зоны у границы с диэлектриком изгибается вверх и приближается к уровню Ферми (рис. 3, a). Поскольку в идеальной МДП-структуре сквозной ток равен нулю, уровень Ферми в полупроводнике остается постоянным. Так как концентрация дырок экспоненциально зависит от разности энергий ($E_F - E_V$), такой изгиб зон приводит к увеличению числа основных носителей (дырок) у поверхности полупроводника. Этот режим называется режимом обогащения (аккумуляции). Если к МДП-структуре приложено не слишком большое положительное напряжение ($V > 0$), зоны изгибаются в обратном направлении и при поверхностная область полупроводника обедняется основными носителями (рис. 3, b). Этот режим называют режимом обеднения

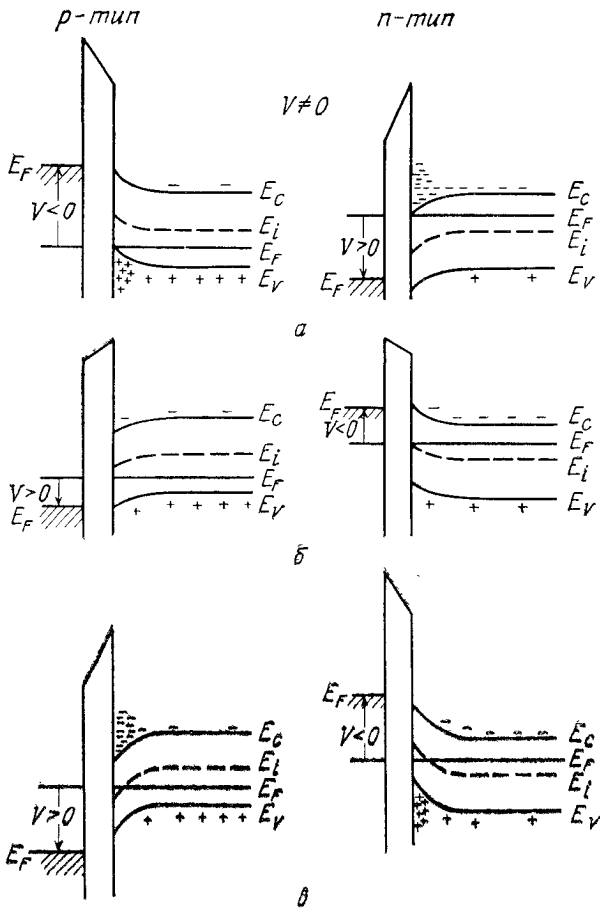


Рис. 3. Зонные диаграммы идеальной МДП-структуры при $V \neq 0$.
 а — режим аккумуляции; б — режим обеднения; в — режим инверсии.

или истощения поверхности. При ббльших положительных напряжениях зоны изгибаются вниз настолько сильно, что вблизи поверхности происходит пересечение уровня Ферми E_F с собственным уровнем E_i . В этом случае (рис. 3, в) концентрация неосновных носителей (электронов) у поверхности превосходит концентрацию основных носителей (дырок). Эта ситуация называется режимом инверсии. Аналогичное рассмотрение можно провести и для МДП-структуры с полупроводником n -типа. Указанные режимы осуществляются при напряжении противоположной полярности.

7.2.1. Приповерхностная область пространственного заряда

Ниже получены соотношения, связывающие поверхностный потенциал, пространственный заряд (отнесенный к единице площади поверхности) и электрическое поле. Эти соотношения использованы в разд. 7.2.2 для расчета вольт-фарадных характеристик идеальных МДП-структур.

На рис. 4 в более крупном плане приведена зонная диаграмма приповерхностной области полупроводника p -типа идеальной МДП-структуры. В этой области электростатический потенциал ψ изменяется от значения на поверхности ψ_s до потенциала электро-нейтральной области полупроводника $\psi = 0$, выбранного за точку отсчета. Зависимость концентраций электронов и дырок от потенциала ψ определяется соотношениями

$$n_p = n_{p0} \exp(q\psi/kT) = n_{p0} \exp(\beta\psi), \quad (2)$$

$$p_p = p_{p0} \exp(-q\psi/kT) = p_{p0} \exp(-\beta\psi), \quad (3)$$

где n_{p0} и p_{p0} — равновесные плотности электронов и дырок в объеме полупроводника $\beta \equiv q/kT$. Потенциал ψ положителен, если зоны изогнуты вниз (рис. 4). Соответствующие поверхностные концентрации определяются соотношениями

$$n_s = n_{p0} \exp(\beta\psi_s),$$

$$p_s = p_{p0} \exp(-\beta\psi_s). \quad (4)$$

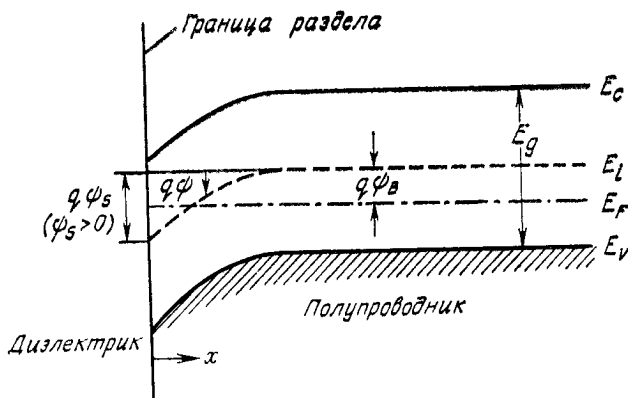


Рис. 4. Зонная диаграмма приповерхностной области полупроводника p -типа. Потенциал ψ определен по отношению к объему подложки ($\psi = 0$ в ее электро-нейтральной части) и отсчитывается от собственного уровня E_i . Изгибу зон на рисунке соответствует положительный знак поверхностного потенциала ψ_s . В режиме аккумуляции $\psi_s \leq 0$, в режиме обеднения $\psi_s > 0$ и при инверсии $\psi_s > \psi_B$.

В соответствии с тем, что было сказано выше, характерные интервалы изменения поверхностного потенциала ψ_s можно определить следующим образом:

- $\psi_s < 0$ — аккумуляция дырок (зоны изогнуты вверх);
 $\psi_s = 0$ — состояние плоских зон;
 $\psi_B > \psi_s > 0$ — режим обеднения (зоны изогнуты вниз);
 $\psi_s = \psi_B - n_s = p_s = n_i$ (n_i — собственная концентрация);
 $\psi_s > \psi_B$ — режим инверсии (накопление у поверхности неосновных носителей (электронов), зоны изогнуты вниз).

Зависимость потенциала ψ от расстояния до границы раздела x можно получить с помощью одномерного уравнения Пуассона

$$\frac{d^2\psi}{dx^2} = -\frac{\rho(x)}{\epsilon_s}. \quad (5)$$

Здесь ϵ_s — диэлектрическая проницаемость полупроводника, а $\rho(x)$ — плотность полного объемного заряда:

$$\rho(x) = q(N_D^+ - N_A^- + p_p - n_p), \quad (6)$$

где N_D^+ и N_A^- — концентрации ионизированных доноров и акцепторов соответственно. Отметим, что в объеме полупроводника вдали от поверхности выполняется условие электронейтральности, т. е. $\rho(x) = 0$ при $\psi = 0$, а

$$N_D^+ - N_A^- = n_{p0} - p_{p0}. \quad (7)$$

В общем случае, согласно выражениям (2) и (3), для всех значений ψ имеем

$$p_p - n_p = p_{p0} \exp(-\beta\psi) - n_{p0} \exp(\beta\psi). \quad (8)$$

В результате вместо уравнения (5) получим

$$\frac{\partial^2\psi}{\partial x^2} = -\frac{q}{\epsilon_s} [p_{p0}(e^{-\beta\psi} - 1) - n_{p0}(e^{\beta\psi} - 1)]. \quad (9)$$

Интегрирование уравнения (9) [13]:

$$\int_0^{\partial\psi/\partial x} \left(\frac{\partial\psi}{\partial x}\right) d\left(\frac{\partial\psi}{\partial x}\right) = -\frac{q}{\epsilon_s} \int_0^{\psi} [p_{p0}(e^{-\beta\psi} - 1) - n_{p0}(e^{\beta\psi} - 1)] d\psi \quad (10)$$

дает соотношение, связывающее электрическое поле ($\mathcal{E} \equiv -d\psi/dx$) и потенциал ψ :

$$\mathcal{E}^2 = \left(\frac{2kT}{q}\right)^2 \left(\frac{qp_{p0}\beta}{2\epsilon_s}\right) \left[(e^{-\beta\psi} + \beta\psi - 1) + \frac{n_{p0}}{p_{p0}} (e^{\beta\psi} - \beta\psi - 1) \right]. \quad (11)$$

Для сокращения записи последующих формул обозначим

$$F\left(\beta\psi, \frac{n_{p0}}{p_{p0}}\right) \equiv \left[(e^{-\beta\psi} + \beta\psi - 1) + \frac{n_{p0}}{p_{p0}} (e^{\beta\psi} - \beta\psi - 1) \right]^{1/2} \geq 0 \quad (12)$$

и введем так называемую дебаевскую длину дырок

$$L_D \equiv \sqrt{\frac{kT\varepsilon_s}{p_{p0}q^2}} \equiv \sqrt{\frac{\varepsilon_s}{qp_{p0}\beta}}. \quad (13)$$

Тогда электрическое поле

$$\mathcal{E} = -\frac{\partial\psi}{\partial x} = \pm \frac{\sqrt{2}kT}{qL_D} F\left(\beta\psi, \frac{n_{p0}}{p_{p0}}\right). \quad (14)$$

В этом выражении знак $+$ нужно использовать при $\psi > 0$, а знак $-$ при $\psi < 0$. Величину поверхностного электрического поля получим, подставив в выражение (14) $\psi = \psi_s$:

$$\mathcal{E}_s = \pm \frac{\sqrt{2}kT}{qL_D} F\left(\beta\psi_s, \frac{n_{p0}}{p_{p0}}\right). \quad (15)$$

По закону Гаусса объемный заряд, отнесенный к единице площади границы раздела, индуцировавший это поле, составляет

$$Q_s = -\varepsilon_s \mathcal{E}_s = \mp \frac{\sqrt{2}\varepsilon_s kT}{qL_D} F\left(\beta\psi_s, \frac{n_{p0}}{p_{p0}}\right). \quad (16)$$

Чтобы определить избыточные поверхностные плотности электронов Δn и дырок Δp при данном значении поверхностного потенциала ψ_s , необходимо вычислить следующие интегралы [14]:

$$\Delta p = p_{p0} \int_0^{\infty} (e^{-\beta\psi} - 1) dx = \frac{qp_{p0}L_D}{\sqrt{2}kT} \int_{\psi_s}^0 \frac{e^{-\beta\psi} - 1}{F(\beta\psi, n_{p0}/p_{p0})} d\psi \text{ [см}^{-2}\text{]}, \quad (17)$$

$$\Delta n = n_{p0} \int_0^{\infty} (e^{\beta\psi} - 1) dx = \frac{qn_{p0}L_D}{\sqrt{2}kT} \int_{\psi_s}^0 \frac{e^{\beta\psi} - 1}{F(\beta\psi, n_{p0}/p_{p0})} d\psi \text{ [см}^{-2}\text{]}. \quad (18)$$

Типичная зависимость полного заряда Q_s от поверхностного потенциала ψ_s показана на рис. 5. Она рассчитана для МДП-структуры на кремниевой подложке p -типа с $N_A = 4 \cdot 10^{15} \text{ см}^{-3}$ при комнатной температуре. При отрицательных ψ_s заряд Q_s положителен, что отвечает аккумуляции дырок на поверхности. В этом случае в выражении (12) доминирует первое слагаемое, так что $Q_s \sim \exp(q|\psi_s|/2kT)$. В состоянии плоских зон $\psi_s = 0$ и $Q_s = 0$. В режиме обеднения $\psi_B > \psi_s > 0$, а заряд Q_s отрицателен. При этом в выражении (12) доминирует второе слагаемое,

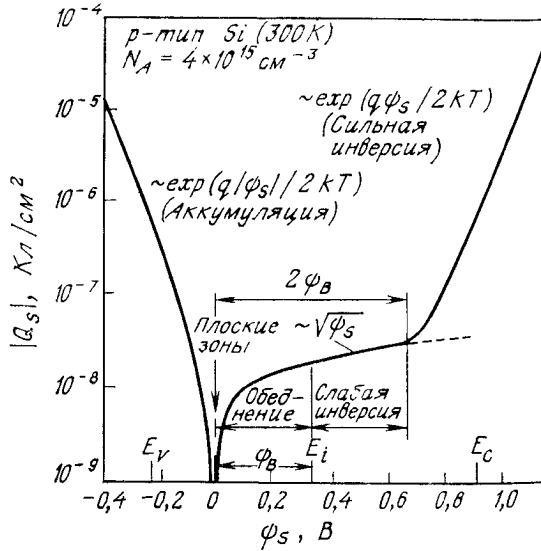


Рис. 5. Зависимость плотности объемного заряда в полупроводнике (на единицу площади границы раздела) от поверхностного потенциала ψ_s для кремния p -типа с $N_A = 4 \cdot 10^{15} \text{ см}^{-3}$ при комнатной температуре; потенциал ψ_B соответствует разности уровня Ферми и собственного уровня E_i в объеме полупроводника [13].

так что $Q_s \sim \sqrt{\psi_s}$. При сильной инверсии $\psi_s \gg \psi_B$ главным в выражении (12) является четвертое слагаемое, и в этом случае $Q_s \sim -\exp(q\psi_s/2kT)$. Сильная инверсия наступает при поверхностном потенциале

$$\psi_s(\text{inv}) \simeq 2\psi_B = \frac{2kT}{q} \ln\left(\frac{N_A}{n_i}\right), \quad (19)$$

когда поверхностная концентрация неосновных носителей (электронов) становится равной исходной концентрации основных носителей p_{p0} .

По определению полная дифференциальная емкость полупроводника

$$C_D \equiv \frac{\partial Q_s}{\partial \psi_s} = \frac{\epsilon_s}{\sqrt{2}L_D} \frac{[1 - e^{-\beta\psi_s} + (n_{p0}/p_{p0})(e^{\beta\psi_s} - 1)]}{F(\beta\psi_s, n_{p0}/p_{p0})} [\Phi/\text{см}^2]. \quad (20)$$

В состоянии плоских зон, т. е. при $\psi_s = 0$, C_D можно легко определить, разложив в ряд соответствующие экспоненты, что дает

$$C_D(\psi_s = 0) = \epsilon_s/L_D [\Phi/\text{см}^2]. \quad (21)$$

7.2.2. Характеристики идеальной МДП-структуры

На рис. 6, а приведена зонная диаграмма идеальной МДП-структуры с тем же, что и на рис. 4, характером изгиба зон. На рис. 6, б приведено распределение заряда в структуре. Ясно, что для обеспечения электронейтральности структуры заряд на ее

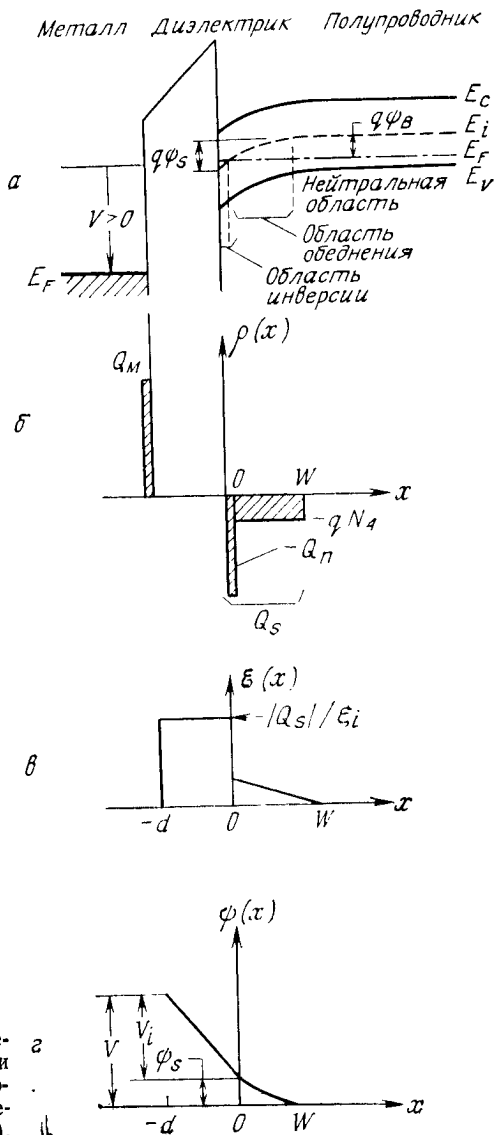


Рис. 6. Зонная диаграмма идеальной МДП-структуры (а) и распределения зарядов (в условиях инверсии) (б), электрического поля (в) и потенциала (г).

металлическом электроде Q_M должен быть равен сумме электронного заряда в инверсионном слое Q_n и заряда ионизированных акцепторов в области обеднения полупроводника:

$$Q_M = Q_n + qN_A W = Q_s. \quad (22)$$

Все заряды здесь отнесены к единице площади границы раздела, W — толщина обедненного слоя, Q_s — полная поверхностная плотность заряда в полупроводнике. На рис. 6, *в* и *г* показаны распределения электрического поля и потенциала в структуре, которые соответствуют первому и второму интегралам уравнения Пуассона.

Ясно, что в отсутствие разности работ выхода приложенное напряжение делится между полупроводником и изолятором, т. е.

$$V = V_i + \psi_s, \quad (23)$$

где V_i — падение напряжения на слое диэлектрика, равное (см. рис. 6, *в*)

$$V_i = \mathcal{E}_i d = \frac{|Q_s| d}{\epsilon_i} \left(\equiv \frac{|Q_s|}{C_i} \right). \quad (24)$$

Полная емкость структуры

$$C = \frac{C_i C_D}{C_i + C_D} \quad [\text{Ф/см}^2], \quad (25)$$

что соответствует последовательному соединению емкости полупроводника C_D и емкости слоя диэлектрика $C_i = \epsilon_i/d$. Последняя определяется толщиной диэлектрика и представляет собой максимально возможную емкость всей структуры. Емкость полупроводника изменяется при изменении напряжения на структуре (уравнение (20)). Используя формулы (20), (23)—(25), можно определить зависимость полной емкости идеальной МДП-структуры (рис. 7, кривая (а)) от приложенного напряжения. Особый интерес представляет значение полной емкости структуры в состоянии плоских зон, т. е. при $\psi_s = 0$. Для этой величины из выражений (21) и (25) получим

$$C_{FB}(\psi_s = 0) = \frac{\epsilon_i}{d + (\epsilon_i/\epsilon_s) L_D} = \frac{\epsilon_i}{d + (\epsilon_i/\epsilon_s) \sqrt{kT\epsilon_s/\rho p_0 q^2}}, \quad (26)$$

где ϵ_i и ϵ_s — диэлектрические проницаемости изолятора и полупроводника соответственно, а L_D — дебаевская длина дырок (см. выражение (13)).

Зависимость емкости идеальной МДП-структуры от напряжения при отрицательных значениях последнего отвечает аккумуляции дырок у границы раздела. В этом режиме дифференциальная емкость полупроводника (выражение (20)) существенно больше емкости диэлектрика, поэтому полная емкость структуры близка

к величине C_i . Когда напряжение, приложенное к МДП-структуре, становится больше 0, в приповерхностном слое полупроводника образуется обедненная область, которая действует как добавочный слой диэлектрика. Это приводит к уменьшению полной емкости МДП-структуры. Затем, проходя через минимум, обозначенный на рис. 7 символами C_{\min} и V_{\min} , полная дифференциальная емкость структуры резко возрастает, снова приближаясь к величине C_i . Последнее обусловлено тем, что в данной области напряжений у границы раздела с диэлектриком образуется электронный инверсионный слой, дифференциальная емкость которого также значительно превышает емкость диэлектрика. Отметим, что нарастание емкости в области положительных смещений зависит от того, успевает ли концентрация инверсионных электронов следовать за изменениями приложенного к структуре переменного напряжения, с помощью которого осуществляется измерение емкости. Данный режим осуществляется лишь при сравнительно малых частотах, когда скорость генерационно-рекомбинационных процессов, ответственных за изменение концентрации неосновных носителей (в нашем случае электронов), достаточна, чтобы электронная плотность изменялась в фазе с напряжением измерительного сигнала. Экспериментально установлено, что для структур металл — SiO_2 — Si соответствующая частотная граница лежит в диапазоне 5—100 Гц [15, 16]. При более высоких частотах увеличения дифференциальной емкости структуры при положительных

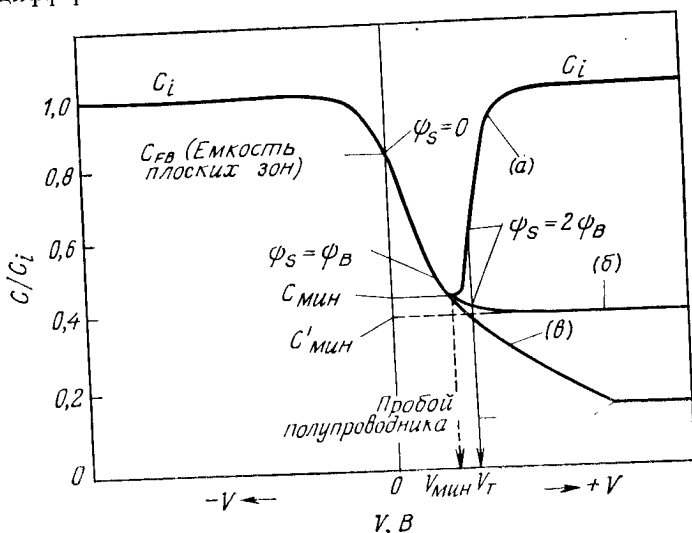


Рис. 7. Вольт-фарадные кривые идеальной МДП-структуры [16].

(а) — в области низких частот; (б) — в области высоких частот; (в) — в режиме глубокого обеднения.

напряжениях не наблюдается (кривая (б) на рис. 7). Кривая (в) на этом рисунке соответствует вольт-фарадной характеристике идеальной МДП-структуры в условиях глубокого обеднения (импульсное напряжение смещения). Эта зависимость непосредственно связана с режимами работы приборов с зарядовой связью (ПЗС), которые рассмотрены в разд. 7.4. В области сравнительно высоких напряжений кривая (в) становится пологой в результате лавинного пробоя приповерхностной области полупроводника МДП-структуры. Более подробно влияние ударной ионизации на характеристики МДП-структур рассмотрено ниже.

На кривых, приведенных на рис. 7, указаны также характерные значения поверхностного потенциала. Как уже говорилось, в идеальной МДП-структуре состояние плоских зон $\psi_s = 0$ осуществляется при нулевом напряжении ($V = 0$). Режиму обеднения соответствуют поверхностные потенциалы от $\psi_s = 0$ до $\psi_s = \psi_B$. Величина ψ_B оказывается несколько меньше значения $V_{\text{мин}}$. При поверхностном потенциале $\psi_s = 2\psi_B$ начинается сильная инверсия. Режиму слабой инверсии соответствуют поверхностные потенциалы $\psi_B < \psi < 2\psi_B$.

Форму высокочастотной $C-V$ -кривой идеальной МДП-структуры можно рассчитать так, как и вольт-фарадную зависимость резкого асимметричного $p-n$ -перехода [17, 70]. В режиме обеднения заряд, приходящийся на единицу площади границы раздела, равен $-qN_A W$, где N_A — концентрация ионизированных акцепторов, а W — глубина обедненного слоя. Интегрирование соответствующего уравнения Пуассона дает следующее распределение потенциала в обедненном слое:

$$\psi = \psi_s \left(1 - \frac{x}{W}\right)^2, \quad (27)$$

причем поверхностный потенциал

$$\psi_s = \frac{qN_A W^2}{2\epsilon_s}. \quad (27a)$$

С увеличением приложенного напряжения ψ_s и W растут, пока не наступает сильная инверсия ($\psi_s (\text{inv}) \simeq 2\psi_B$) (рис. 5), во время которой рост ψ_s и W резко замедляется. В режиме сильной инверсии небольшое увеличение поверхностного потенциала $\Delta\psi_s$ вызывает существенное увеличение плотности заряда инверсионного слоя, который экранирует объем полупроводника от проникновения электрического поля с электрода структуры. Можно считать, что в режиме сильной инверсии достигается максимальная толщина обедненного слоя, равная, согласно формулам (19) и (27a),

$$W_m \simeq \frac{2\epsilon_s \psi_s (\text{inv})}{qN_A} = \sqrt{\frac{4\epsilon_s kT \ln(N_A/n_i)}{q^2 N_A}}. \quad (28)$$

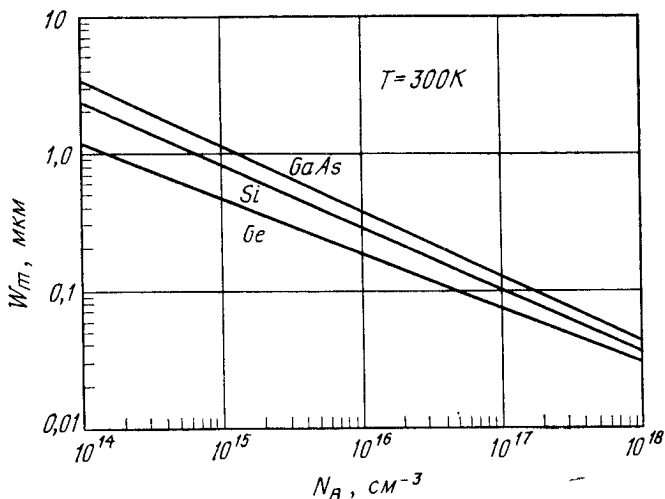


Рис. 8. Зависимость максимальной ширины обедненного слоя от концентрации примеси в Ge, Si и GaAs при сильной инверсии.

Зависимости W_m от концентрации легирующей примеси, рассчитанные для Ge, Si и GaAs при комнатной температуре, приведены на рис. 8. Здесь N_B равно либо N_A , либо N_D в зависимости от типа проводимости полупроводника. Важной величиной является так называемое напряжение включения (пороговое напряжение) V_T , при котором начинается сильная инверсия. Используя формулы (19) и (23), находим

$$V_T = \frac{Q_s}{C_i} + 2\psi_B, \quad (29)$$

где $Q_s = qN_A W_m$. Отсюда с помощью выражения (22) получим окончательно:

$$V_T \approx \sqrt{\frac{2\varepsilon_s q N_A (2\psi_B)}{C_i}} + 2\psi_B. \quad (29a)$$

Соответствующее V_T значение дифференциальной емкости идеальной МДП-структуры

$$C'_{\text{мин}} \approx \frac{\varepsilon_i}{d + (\varepsilon_i/\varepsilon_s) W_m}. \quad (30)$$

Высокочастотная вольт-фарадная характеристика идеальной МДП-структуры [16] приведена на рис. 9. При $0 < V < V_T$ ее можно аппроксимировать зависимостью, полученной в приближении обедненного слоя (штриховая линия), а при $V < 0$ и $V >$

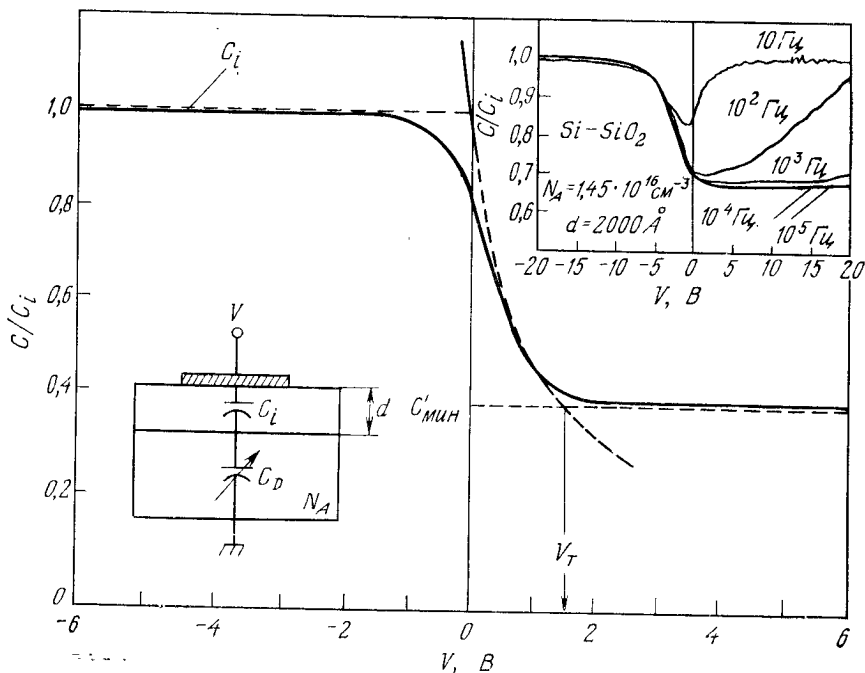


Рис. 9. Высокочастотная вольт-фардадная зависимость МДП-структуры и ее аппроксимация (штриховые линии). На вставке приведены $C-V$ -кривые при разных частотах тестирующего напряжения [16].

$> V_T$ — постоянными значениями C_i и C'_{\min} соответственно. На вставке к рис. 9 приведены экспериментальные $C-V$ -кривые кремниевых МОП-конденсаторов, снятые при разных частотах.

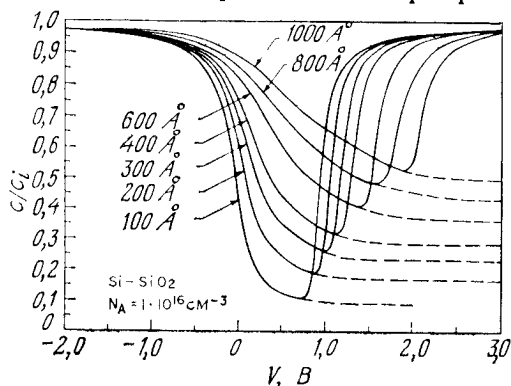


Рис. 10. $C-V$ -кривые идеальных МДП-структур. Сплошными линиями показаны низкочастотные характеристики, а штриховыми — высокочастотные [18].

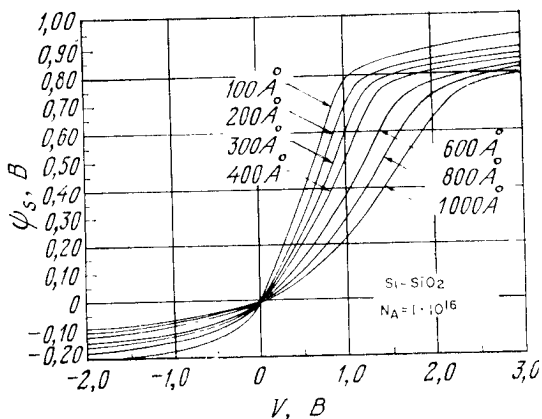


Рис. 11. Зависимость поверхностного потенциала от приложенного напряжения в идеальных МДП-структурах [18].

Отметим, что низкочастотный характер $C-V$ -зависимости, как уже говорилось выше, начинает проявляться при $f \approx 100$ Гц.

Серия вольт-фарадных характеристик идеальных МДП-структур (металл — SiO_2 —Si) с различными значениями толщины слоя окисла и степени легирования полупроводника рассчитана в ра-

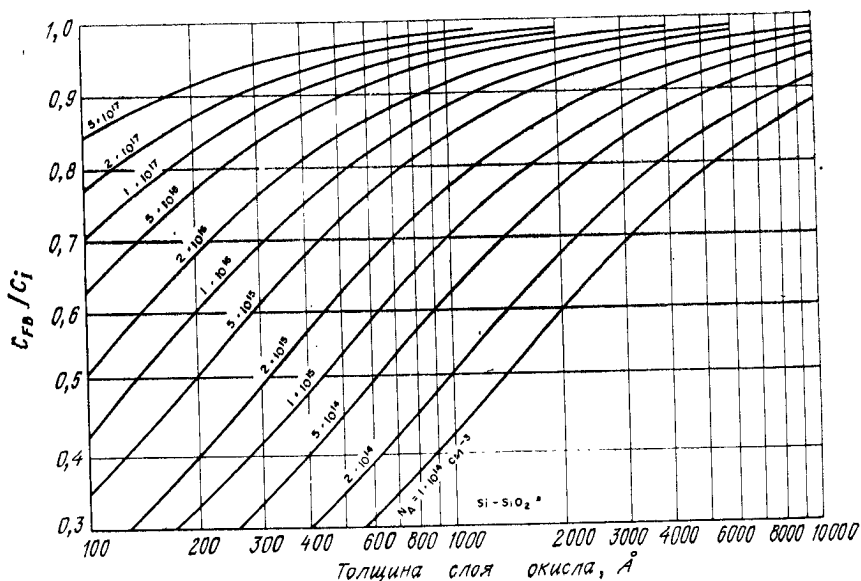


Рис. 12. Зависимость емкости плоских зон от толщины слоя окисла для идеальных МДП-структур. Параметр кривых — уровень легирования кремния [18],

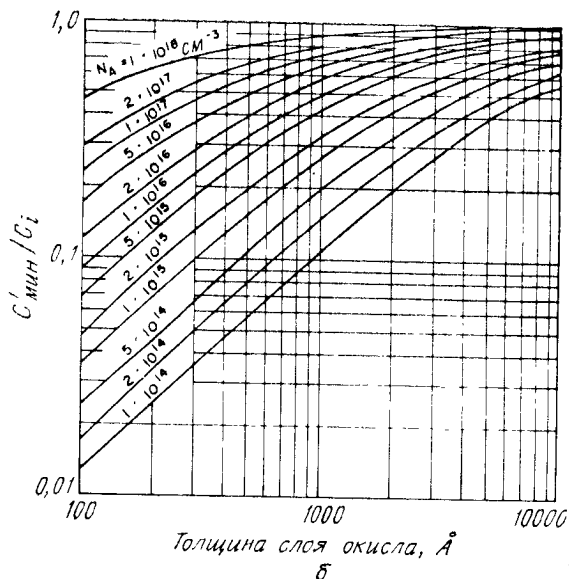
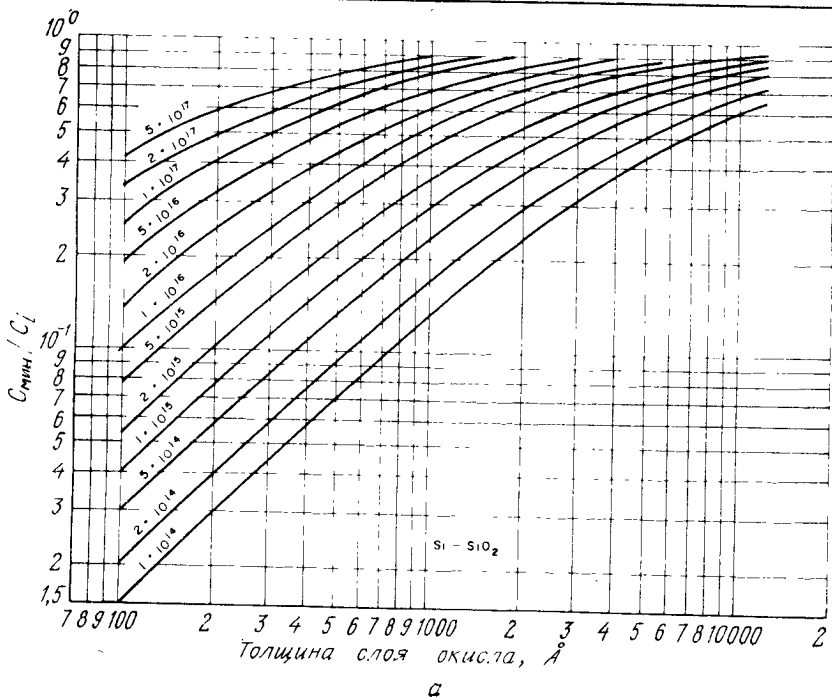


Рис. 13. Зависимость нормированной минимальной низкочастотной (а) и высокочастотной (б) емкости от толщины слоя окисла для идеальных МДП-структур. Параметр кривых — уровень легирования кремния [18].

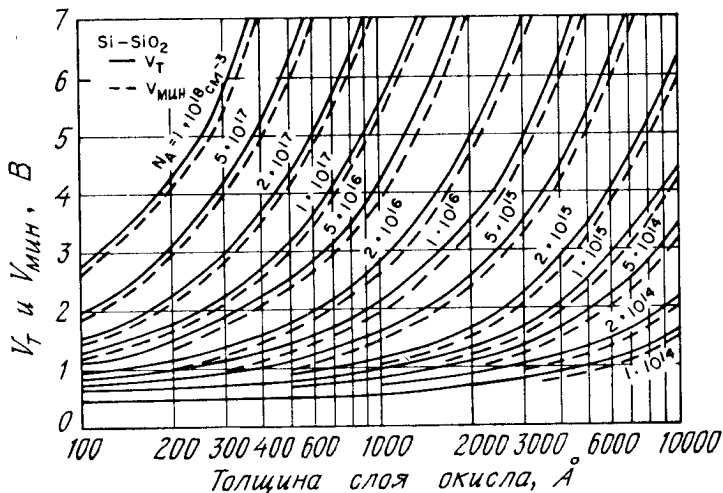


Рис. 14. Зависимости порогового напряжения V_T и напряжения V_{\min} , соответствующего минимальной низкочастотной емкости C_{\min} , от толщины слоя окисла для идеальных МДП-структур. Параметр кривых — уровень легирования кремния [18].

боте [18]. Некоторые из них (для типичных значений толщины SiO_2) приведены на рис. 10. Отметим, что в структурах с более тонким слоем SiO_2 наблюдается более сильная зависимость емкости от напряжения. На рис. 11 приведены зависимости поверхностного потенциала от напряжения в тех же МДП-структурах. Графики на рис. 12—14 показывают зависимость характерных значений C — V -характеристик (C_{\min} , C'_{\min} , C_{FB}), нормированных на величину емкости диэлектрика, а также V_{\min} и V_T от параметров МДП-структуры. Кривые, приведенные на рис. 10 и 11, рассчитаны для МДП-структур с полупроводником p -типа. Их можно использовать и для анализа характеристик МДП-структур на n -подложке. Для этого нужно поменять знак на оси напряжений на рис. 10 и 11. Данные кривые можно использовать и для анализа характеристик МДП-структур с другими диэлектриками (а не только с диэлектриком SiO_2), при этом следует воспользоваться выражением

$$d_c = d_i \frac{\epsilon_i(\text{SiO}_2)}{\epsilon_i(\text{диэлектрика})}, \quad (31)$$

где d_i — действительная толщина диэлектрического слоя, а d_c — эффективная толщина, служащая параметром для графиков рис. 10 и 11 (ϵ_i — диэлектрическая проницаемость этого изолятора, а $\epsilon(\text{SiO}_2) = 3,5 \cdot 10^{-13} \text{ Ф} \cdot \text{см}^{-1}$). С помощью формул (26—29) анало-

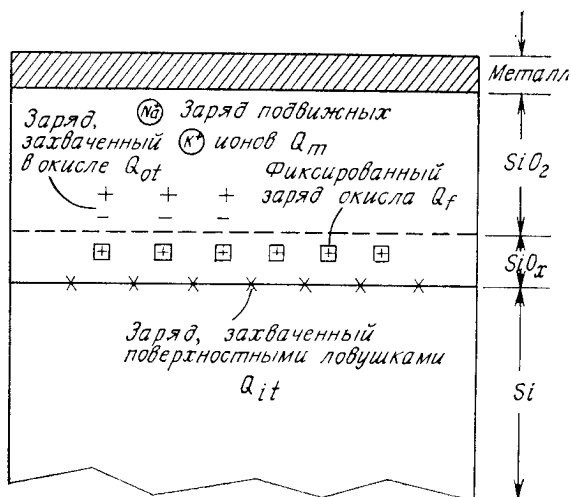


Рис. 15. Классификация зарядов, присутствующих в термически окисленном кремнии [19].

гичные кривые можно рассчитать и для МДП-структур на других полупроводниковых материалах.

Приведенные на рис. 10—14 зависимости идеальных МДП-структур в последующих разделах использованы для сравнения с экспериментальными результатами и для анализа характеристик реальных МДП-конденсаторов.

7.3. Si—SiO₂—МОП-СТРУКТУРЫ

Из всех МДП-структур наиболее важными являются структуры металл—SiO₂—Si (МОП). Однако природа различных электрофизических процессов, происходящих на границе раздела Si—SiO₂, в настоящее время еще не получила достаточно полного объяснения. Многие исследователи [7] считают, что границу раздела Si—SiO₂, возникающую при термическом окислении кремния, следует рассматривать как некоторый переходный слой с переменным химическим составом. Предполагается, что на границе монокристаллического кремния находится моноатомный слой нестехиометрического SiO_x ($1 < x < 2$; стехиометрической двуокиси кремния соответствует $x = 2$), представляющий собой не полностью окисленный кремний. Затем следует промежуточный слой SiO₂ с большими внутренними механическими напряжениями толщиной 10—40 Å, который переходит в обычный ненапряженный стехиометрический аморфный SiO₂. Отличие характеристик реальных МОП-структур от соответствующих зависимостей идеальных

МДП-конденсаторов обусловлено существованием поверхностных ловушек и зарядов в окисле.

В настоящее время принята следующая классификация этих зарядов и ловушек [19] (рис. 15):

1. Заряд, захваченный поверхностными ловушками Q_{it} , представляющий собой заряд электронных состояний, которые локализованы на границе раздела Si—SiO₂ и энергия которых лежит в глубине запрещенной зоны полупроводника. Эти поверхностные состояния, называемые также быстрыми (а иногда пограничными), могут достаточно быстро перезаряжаться, обмениваясь электронами (дырками) с кремнием. Поверхностные состояния, вероятно, обусловлены избыточными атомами кремния (трехвалентным кремнием), избыточным кислородом или примесными атомами.

2. Фиксированный заряд окисла Q_f , расположенный на границе раздела или в непосредственной близости от нее. Величина этого заряда остается практически постоянной во всей области электрических полей, характерных для рабочего диапазона напряжений на МОП-структурах.

3. Заряд, захваченный в окисле Q_{oi} . Этот заряд возникает, например, при рентгеновском облучении структур или инжекции горячих электронов в диэлектрик. Соответствующие ловушки более или менее равномерно распределены по толщине слоя окисла.

4. Заряд подвижных ионов Q_m (например, ионов натрия), который может перемещаться в слое окисла при стрессовых термополевых нагрузках МДП-структур.

Величину всех этих зарядов обычно относят к единице площади границы раздела, т. е. измеряют в единицах Кл·см⁻². Вместо Q_i часто используют соответствующие поверхностные плотности, которые обозначают символом N_i с теми же индексами ($N_i = Q_i/q$ — число зарядов на 1 см²). Поскольку энергетические уровни состояний, захватывающих поверхностный заряд Q_{it} , непрерывно распределены в запрещенной зоне полупроводника, полезной характеристикой является энергетическая плотность поверхностных состояний

$$D_{it} = \frac{1}{q} \frac{dQ_{it}}{dE} \text{ [число зарядов/см}^2 \cdot \text{эВ]}. \quad (32)$$

7.3.1. Заряд, захваченный поверхностными ловушками

Поверхностные состояния изучали Тамм [20], Шокли [21] и многие другие авторы [7]. Основная причина возникновения этих состояний в запрещенной зоне полупроводника заключается в том, что сама граница раздела является нарушением пространственной периодичности кристаллической решетки. Первое экспериментальное доказательство существования поверхностного

захваченного заряда Q_{it} получили Шокли и Пирсон [22] при измерении поверхностной проводимости. Измерения на чистых поверхностях, полученных сколом в сверхвысоком вакууме [23], показывают, что в этом случае плотность Q_{it} чрезвычайно велика: порядка числа атомов на свободной поверхности кристалла ($\sim 10^{15}$ атом/см²). В современных МОП-структурах, получаемых термическим окислением кремния, подавляющая часть поверхностного заряда Q_{it} нейтрализуется в процессе низкотемпературного (450 °С) отжига в атмосфере водорода. В лучших образцах величина N_{it} не превышает 10^{10} см⁻², что соответствует одному захваченному заряду на каждые 10^5 атомов границы раздела.

Поверхностные состояния считаются донорными, если, отдавая электрон, они становятся нейтральными или положительно заряженными. Акцепторными называют поверхностные состояния, которые становятся нейтральными или отрицательно заряженными, захватывая электрон. Заполнение поверхностных состояний, так же как и объемных (гл. 1), определяется распределениями Ферми

$$F_{SD}(E_t) = \left[1 - \frac{1}{1 + \frac{1}{g} \exp\left(\frac{E_t - E_F}{kT}\right)} \right] = \frac{1}{1 + g \exp\left(\frac{E_F - E_t}{kT}\right)} \quad (33a)$$

для донорных поверхностных ловушек и

$$F_{SA}(E_t) = \frac{1}{1 + \frac{1}{g} \exp\left(\frac{E_t - E_F}{kT}\right)} \quad (33б)$$

для акцепторных поверхностных ловушек. Здесь E_t — энергия поверхностного состояния, а g — фактор вырождения, равный 2 для донорных и 4 для акцепторных поверхностных состояний.

При изменениях приложенного к МДП-структуре напряжения положение энергетических уровней поверхностных ловушек изменяется, следуя за смещением краев разрешенных зон полупроводника на границе раздела, в то время как положение уровня Ферми остается неизменным. В результате происходит изменение зарядового состояния этих ловушек (в первом приближении можно считать, что оно происходит, когда энергетический уровень ловушки пересекает уровень Ферми). Изменение поверхностного связанного заряда дает дополнительный вклад в дифференциальную емкость МОП-структуры, в результате чего вольт-фарадная характеристика последней отличается от $C-V$ -зависимости идеальной МДП-структуры. Эквивалентная электрическая схема МОП-структуры [24], учитывающая влияние поверхностных состояний, показана на рис. 16, а. На этом рисунке C_i и C_D — емкости изолятора и обедненного слоя полупроводника соответственно, идентичные соответствующим емкостям, показанным на вставке внизу

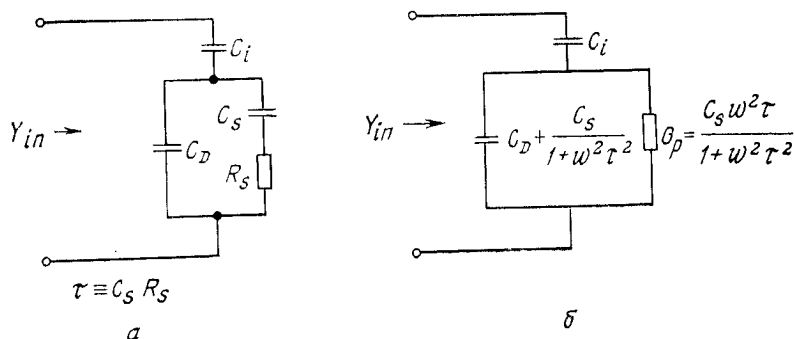


Рис. 16. Эквивалентные электрические схемы, учитывающие влияние поверхностных состояний, C_s и R_s связаны с плотностью поверхностных состояний D_{it} [24].

слева на рис. 9. Эквивалентные емкость C_s и сопротивление R_s зависят от поверхностного потенциала и описывают перезарядку поверхностных состояний. Произведение $C_s R_s$ определяет время релаксации заполнения поверхностных ловушек и частотную зависимость процесса перезарядки поверхностных состояний. Параллельные ветви электрической схемы на рис. 16, а можно представить (рис. 16, б) в виде зависящей от частоты эквивалентной емкости

$$C_p = C_D + \frac{C_s}{1 + \omega^2 \tau^2}, \quad (34)$$

включенной параллельно с эквивалентной проводимостью

$$\frac{G_p}{\omega} = \frac{C_s \omega \tau}{1 + \omega^2 \tau^2}, \quad (35)$$

где $\tau \equiv C_s R_s$. Активная и реактивная составляющие полной комплексной проводимости эквивалентных схем, приведенных на рис. 16:

$$Y_{in} \equiv G_{in} + j\omega C_{in}, \quad (36)$$

определяются выражениями

$$G_{in} = \frac{\omega^2 C_s \tau C_i^2}{(C_i + C_D + C_s)^2 + \omega^2 \tau^2 (C_i + C_D)^2}, \quad (36a)$$

$$C_{in} = \frac{C_i}{C_i + C_D + C_s} \left[C_D + C_s \frac{(C_i + C_D + C_s)^2 + \omega^2 \tau^2 C_D (C_i + C_D)}{(C_i + C_D + C_s)^2 + \omega^2 \tau^2 (C_i + C_D)^2} \right]. \quad (36б)$$

Емкостные методы. Для определения величины заряда, захваченного на поверхностные ловушки, можно использовать измере-

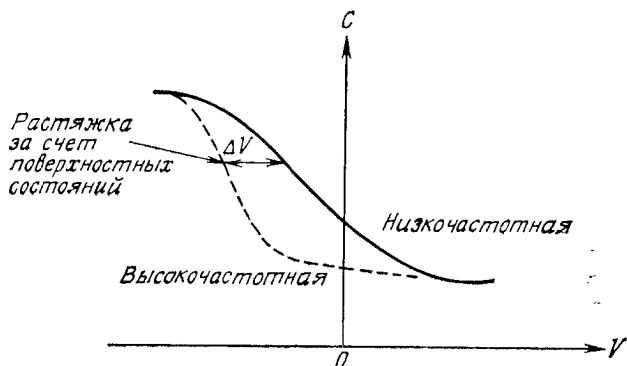


Рис. 17. Вытягивание $C - V$ -кривых за счет поверхностных состояний.

ния входной емкости МДП-структуры и измерения ее входной проводимости, поскольку формулы (36а) и (36б) в принципе содержат одинаковую информацию о поверхностных состояниях. Ниже показано, что метод проводимости обеспечивает несколько более высокую точность определения плотности поверхностных состояний. Это оказывается особенно важным при исследовании МОП-структур с относительно малой ($\sim 10^{10}$ см $^{-2}$ эВ $^{-1}$) плотностью поверхностных состояний. Преимущество емкостных методов состоит в том, что с их помощью можно сравнительно легко определять сдвиг напряжения плоских зон и величину полного захваченного заряда Q_{it} (интегральную плотность поверхностных состояний).

На рис. 17 проиллюстрировано вытягивание (уширение) $C - V$ -характеристики МОП-структуры за счет поверхностного захваченного заряда. При высоких частотах ($\omega\tau \gg 1$) заряд на поверхностных состояниях не успевает следовать за изменениями переменного тестирующего напряжения. В этом случае выражение для емкости (36б) принимает вид выражения (25). Высокочастотная $C - V$ -зависимость МОП-структуры, в которую не входит емкость поверхностных состояний C_s , изображена на рис. 17 штриховой линией. Однако и в этом случае поверхностные состояния влияют на форму вольт-фарадной характеристики, сдвигая и растягивая ее вдоль оси напряжений. Дело в том, что при наличии связанного поверхностного заряда требуется соответствующее увеличение (по сравнению с ситуацией в идеальной МДП-структуре) заряда на металлическом электроде, чтобы достичь того же значения поверхностного потенциала. По форме высокочастотной вольт-фарадной зависимости плотность поверхностных состояний определяется с помощью следующего выражения:

$$D_{it} = \frac{C_t}{q} [(d\psi_s/dV)^{-1} - 1] - \frac{C_D}{q} [\text{см}^{-2} \text{ эВ}^{-1}]. \quad (37)$$

При этом сначала по значению полной емкости C с помощью выражения (25) определяют дифференциальную емкость полупроводника C_D , а затем с помощью формулы (20) находят поверхностный потенциал ψ_s (для этого необходимо знать концентрацию легирующей примеси). После этого производную $d\psi_s/dV$ определяют графическим дифференцированием.

Интегральный метод [25] основан на измерении низкочастотной зависимости емкости структуры от напряжения. В условиях, когда можно пренебречь изменением заряда, захваченного в окисле, и заряда подвижных ионов, из выражений (23) и (25) (если принять во внимание, что $dQ = C_i dV_i = CdV$) следует

$$\left(\frac{\partial\psi_s}{\partial V}\right) = 1 - \frac{C}{C_i}, \quad (38)$$

$$\frac{d\psi_s}{dV_i} = \frac{C_i}{C} - 1. \quad (39)$$

Интегрируя выражения (38), найдем

$$\psi_s(V_1) - \psi_s(V_2) = \int_{V_2}^{V_1} \left[1 - \frac{C}{C_i}\right] dV. \quad (40)$$

Следовательно, зависимость поверхностного потенциала от приложенного напряжения можно найти интегрированием зависимости $(1 - C/C_i)$, где $C \equiv C(V)$ — экспериментальная C — V -кривая. Отметим, что уравнение (40) справедливо только в том случае, когда в процессе измерений успевает устанавливаться равновесное заполнение поверхностных состояний. Для этого должны быть достаточно малы частота напряжения измерительного сигнала и скорость изменения напряжения смещения на структуре. Воспользуемся теперь условием электронейтральности МОП-структуры. В данном случае к пространственному заряду полупроводника Q_s (рис. 6, б) нужно добавить заряд, связанный на донорных и акцепторных поверхностных состояниях, т. е. $Q_{it}^d(\psi_s)$ и $Q_{it}^a(\psi_s)$. Из условия электронейтральности находим

$$\frac{\varepsilon_i V_i}{d} = q \int_{E_V}^{E_C} [D_{it}^d F_{SD}(E_t) - D_{it}^a F_{SA}(E_t)] dE_t + Q_s, \quad (41)$$

где D_{it}^d и D_{it}^a — плотности донорных и акцепторных поверхностных состояний, а $F_{SD}(E_t)$ и $F_{SA}(E_t)$ — функции Ферми (выражение (33)). Дифференцируя выражение (41) по ψ_s , получим окончательно

$$\frac{\partial\psi_s}{\partial V_i} = \frac{\varepsilon_i/d}{(dQ_s/d\psi_s) + qD_{it}(q\psi_s)}. \quad (42)$$

Отсюда определяется плотность поверхностных состояний D_{it} . Для этого нужно знать концентрацию легирующей примеси и температуру. Отметим, что требуемая зависимость $\partial\psi_s/\partial V_i$ от ψ_s получается непосредственно из измеренной низкочастотной $C-V$ -характеристики МОП-структуры с помощью формул (39) и (40).

Метод проводимости. Как уже говорилось выше, активная и реактивная составляющие комплексной проводимости МДП-структуры содержат одинаковую информацию о поверхностных состояниях. Однако при использовании емкостных методов возникают определенные трудности с выделением емкости поверхностных состояний из измеренной полной емкости структуры, поскольку последняя содержит емкость диэлектрика и емкость обедненного слоя полупроводника. Эти трудности особенно велики при исследовании МДП-структур с малой плотностью состояний. В предложенном в работе [26] методе проводимости, основанном на измерении активной составляющей отклика МДП-структуры, эти трудности устранены, поскольку активная составляющая комплексной проводимости МДП-структуры обусловлена исключительно запаздыванием процесса перезарядки поверхностных состояний. Поэтому метод проводимости дает более точные и надежные результаты, что особенно важно для систем с относительно малой плотностью поверхностных состояний, таких, как, например, система с термически окисленным кремнием.

Приведенные соображения иллюстрирует рис. 18, где показаны зависимости емкости и проводимости МДП-структуры от напряжения, измеренные на частотах 5 и 100 кГц. Отметим, что наибольшее отличие емкости здесь составляет 14 %, в то время как амплитуды пиков проводимости отличаются более чем на порядок.

Для иллюстрации метода проводимости снова воспользуемся упрощенной эквивалентной схемой МДП-структуры (рис. 16). Комплексная проводимость всей структуры определяется с помощью измерительного моста. Емкость диэлектрика C_i находят по величине емкости структуры в режиме аккумуляции. Затем рассчитывается комплексная проводимость составляющей эквивалентной схемы (рис. 16), представляющей собой параллельное соединение емкости полупроводника и последовательной цепочки $R_s C_s$, соответствующей поверхностным состояниям. Для этого комплексная входная проводимость пересчитывается во входной импеданс, из которого вычитается реактивная составляющая, обусловленная емкостью диэлектрика. Полученный при этом импеданс затем снова преобразуется в комплексную проводимость, составляющие которой теперь вычисляются по формулам (34) и (35). Активная составляющая этой проводимости определяется только параметрами C_s и R_s , которые описывают поверхностные состояния, и не зависит от емкости полупроводника C_D . Таким путем определяют

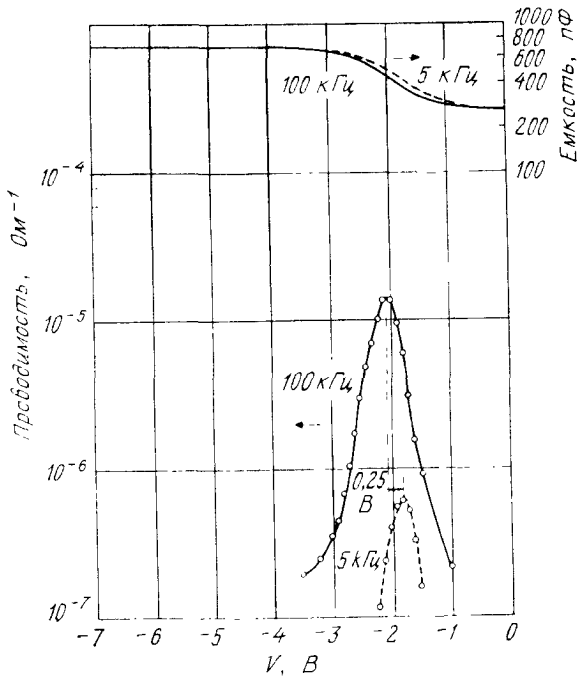


Рис. 18. Результаты измерений емкости и проводимости МДП-структуры при двух частотах [26].

частотную зависимость G_p/ω при фиксированных значениях напряжения смещения на структуре. График $G_p(\omega)/\omega$ проходит через максимум при $\omega = \tau^{-1}$, что позволяет непосредственно определить характерное время перезарядки поверхностных ловушек. Значение G_p/ω в максимуме равно $C_s/2$. Таким образом, метод проводимости позволяет определить величину τ и емкость поверхностных состояний C_s . Плотность поверхностных состояний $D_{it} = C_s/qA$, где A — площадь металлического полевого электрода структуры.

Типичные зависимости D_{it} для Si—SiO₂-структур [27] приведены на рис. 19. Вблизи середины запрещенной зоны D_{it} изменяется относительно слабо, но затем резко нарастает по мере приближения к краям разрешенных зон. Отметим существенную зависимость плотности поверхностных состояний от ориентации полупроводниковой подложки МОП-структуры. В структурах с ориентацией подложки $\langle 100 \rangle$ плотность поверхностных состояний оказывается примерно на порядок меньше, чем в структурах с ориентацией $\langle 111 \rangle$. Этот факт коррелирует с числом свободных связей на соответствующих атомных плоскостях кремния [28, 29]. В табл. 1 приведены параметры атомных плоскостей $\langle 111 \rangle$, $\langle 110 \rangle$ и $\langle 100 \rangle$

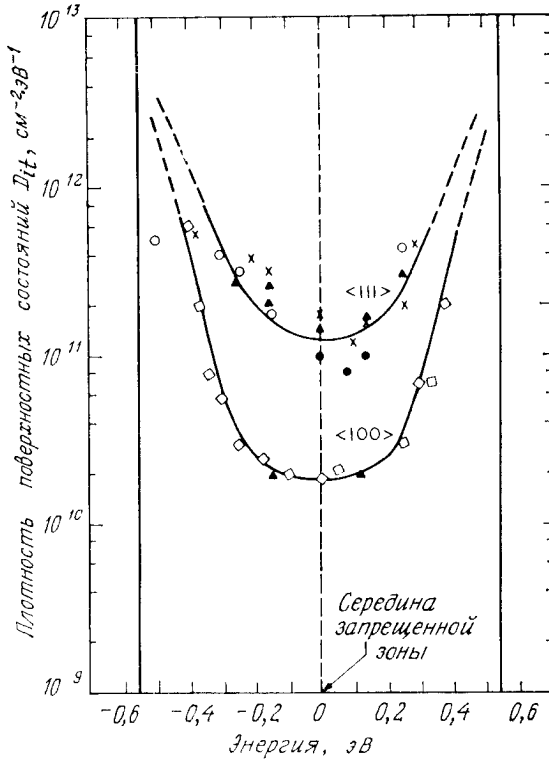


Рис. 19. Плотность поверхностных состояний в термически окисленном кремнии [27].

монокристаллического кремния. Отметим, что поверхность $\langle 111 \rangle$ характеризуется наибольшим числом свободных связей на квадратный сантиметр, а поверхность $\langle 100 \rangle$ — наименьшим. Можно ожидать, что скорость окисления поверхности $\langle 100 \rangle$ также будет наименьшей. Если предположить, что за поверхностные состояния

Таблица 1. Свойства кристаллографических плоскостей кремния

Ориентация	Площадь элементарной ячейки	Число атомов на площади элементарной ячейки	Число свободных связей на площади элементарной ячейки	Поверхностная плотность атомов, 10^{14} см^{-2}	Поверхностная плотность свободных связей, 10^{14} см^{-2}
$\langle 111 \rangle$	$\sqrt{3}a^2/2$	2	3	7,85	11,8
$\langle 110 \rangle$	$\sqrt{2}a^2$	4	4	9,6	9,6
$\langle 100 \rangle$	a^2	2	2	6,8	6,8

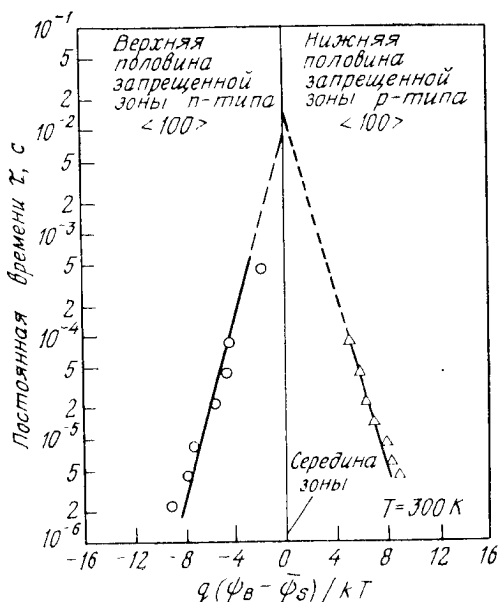


Рис. 20. Зависимость постоянной времени перезарядки от поверхностного потенциала [26].

ответственны избыточные атомы кремния в окисле, то становится понятной ориентационная зависимость плотности состояний (рис. 19), поскольку при меньшей скорости окисления в SiO_2 возникает меньше избыточных атомов кремния. Так или иначе, но все современные кремниевые МОП-транзисторы изготавливаются только на подложках с ориентацией $\langle 100 \rangle$ (гл. 8).

На рис. 20 приведены результаты определения постоянной времени перезарядки τ при некоторых значениях поверхностного потенциала в МОП-структурах, полученных влажным окислением кремниевых подложек с ориентацией $\langle 100 \rangle$. Здесь ψ_B — потенциал, соответствующий равенству собственного уровня и энергии Ферми, а $\bar{\psi}_s$ — средний поверхностный потенциал (смысл этого термина пояснен ниже). Эти результаты можно аппроксимировать экспоненциальными зависимостями, подобными тем, которые описывают генерационно-рекомбинационные процессы на объемных ловушках (гл. 1):

$$\tau = \frac{1}{\bar{v}\sigma_p n_i} \exp \left[-\frac{q(\psi_B - \bar{\psi}_s)}{kT} \right] \text{ для } p\text{-типа,} \quad (43)$$

$$\tau = \frac{1}{\bar{v}\sigma_n n_i} \exp \left[\frac{q(\psi_B - \bar{\psi}_s)}{kT} \right] \text{ для } n\text{-типа.}$$

Здесь σ_p и σ_n — сечения захвата дырок и электронов соответственно, а \bar{v} — средняя тепловая скорость. Из приведенных на рис. 20 экспериментальных результатов следует, что сечения захвата носителей тока на поверхностные состояния не зависят от энергии. В работе [26] получены абсолютные значения сечений захвата $\sigma_p = 4,3 \cdot 10^{-16} \text{ см}^2$ и $\sigma_n = 8,1 \cdot 10^{-16} \text{ см}^2$. При этом были использованы следующие значения параметров: $\bar{v} = 10^7 \text{ см} \cdot \text{с}^{-1}$ и $n_i = 1,6 \cdot 10^{10} \text{ см}^{-3}$. МОП-структуры на кремнии с ориентацией $\langle 111 \rangle$ характеризуются такой же экспоненциальной зависимостью времени перезарядки поверхностных состояний от поверхностного потенциала. Сечения захвата в этих структурах составляют $\sigma_p = 2,2 \cdot 10^{-16} \text{ см}^2$ и $\sigma_n = 5,9 \cdot 10^{-16} \text{ см}^2$.

Из результатов, приведенных на рис. 19, следует, что спектр поверхностных состояний в системе Si — SiO₂ состоит из большого числа энергетических уровней, расположенных так близко друг к другу, что спектр можно считать непрерывным во всей запрещенной зоне. Следовательно, необходимо модифицировать эквивалентную электрическую схему, приведенную на рис. 16, соответствующую моноэнергетическому поверхностному состоянию. Кроме того, необходимо учесть статистические флуктуации поверхностного потенциала, обусловленные пространственной неоднородностью распределения поверхностных зарядов (фиксированного заряда в диэлектрике Q_f и связанного поверхностного заряда Q_{it}). Это оказывается важным, поскольку в соответствии с выражениями (43) небольшие флуктуации ψ_s приводят к весьма значительным флуктуациям τ . Вставка на рис. 21 иллюстрирует, как хаотичность расположения заряженных центров на границе раздела приводит к соответствующим флуктуациям электрического поля и, следовательно, к флуктуациям поверхностного потенциала. На рис. 21 кривыми показаны частотные зависимости $G_p(\omega)/\omega$, рассчитанные для системы Si — SiO₂ с учетом непрерывности энергетического спектра поверхностных состояний, и статистического (пуассоновского) распределения плотности поверхностного заряда $Q_{it} + Q_f$ в плоскости границы раздела. Совпадение экспериментальных точек и теоретических зависимостей несомненно свидетельствует в пользу статистической модели перезарядки поверхностных состояний [26].

Приведенные выше экспериментальные результаты можно интерпретировать с помощью модифицированных эквивалентных электрических схем (рис. 22). Эквивалентная схема на рис. 22, а учитывает дисперсию характерного времени перезарядки поверхностных состояний, обусловленную статистическими флуктуациями поверхностного потенциала. Каждая из составляющих ее последовательных цепочек $R'_i C'_i$ соответствует непрерывному спектру поверхностных состояний на характеристической площадке A_c , величина которой пропорциональна квадрату глу-

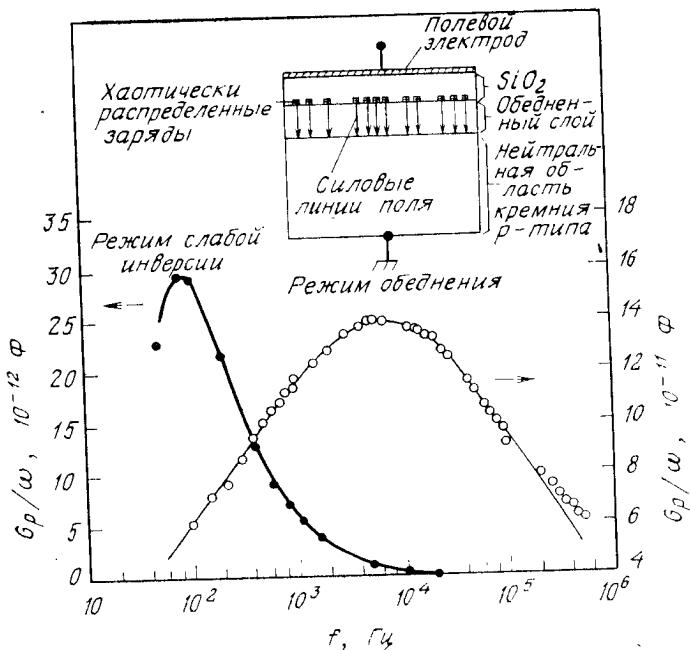


Рис. 21. Зависимость G_p/ω от частоты для Si — SiO₂ — МДП-структуры в режиме обеднения и в режиме слабой инверсии [26].

●, ○ — экспериментальные результаты; — — — — расчетные результаты.

бины обедненного слоя¹⁾. Эта эквивалентная схема описывает поведение МДП-структуры в режиме обеднения. Схема на рис. 22, б отвечает ситуации, когда уровень Ферми на границе раздела проходит вблизи середины запрещенной зоны, т. е. когда $q\psi_s = q\psi_B \pm$ несколько kT . R_{ns}^1 и R_{ps}^1 — эквивалентные сопротивления захвата электронов и дырок соответственно. Наличие двух сопротивлений в каждой элементарной цепочке этой эквивалентной схемы отражает тот факт, что при $\psi \approx \psi_B$ плотности электронов и дырок на границе с окислом являются величинами одного порядка, и, следовательно, необходимо учитывать вклад носителей обоих типов в перезарядку поверхностных состояний. В режиме слабой инверсии ($2\psi_B > \psi_s > \psi_B$) (рис. 22, в) эквива-

¹⁾ В пределах каждой из этих площадок поверхностный потенциал считается постоянным, а статистическое распределение значений ψ_s по различным площадкам описывается гауссовой функцией с дисперсией, пропорциональной заряду $Q_f + Q_{it}$. Для более глубокого ознакомления с физикой перезарядки поверхностных состояний кроме оригинальной статьи [26] можно рекомендовать довольно полный обзор Гейтцбергера, Клаусмана и Шульда в журнале CRC, Critical Reviews Solid State Sciences, 1976, January, pp. 1—43. — Прим. перев.

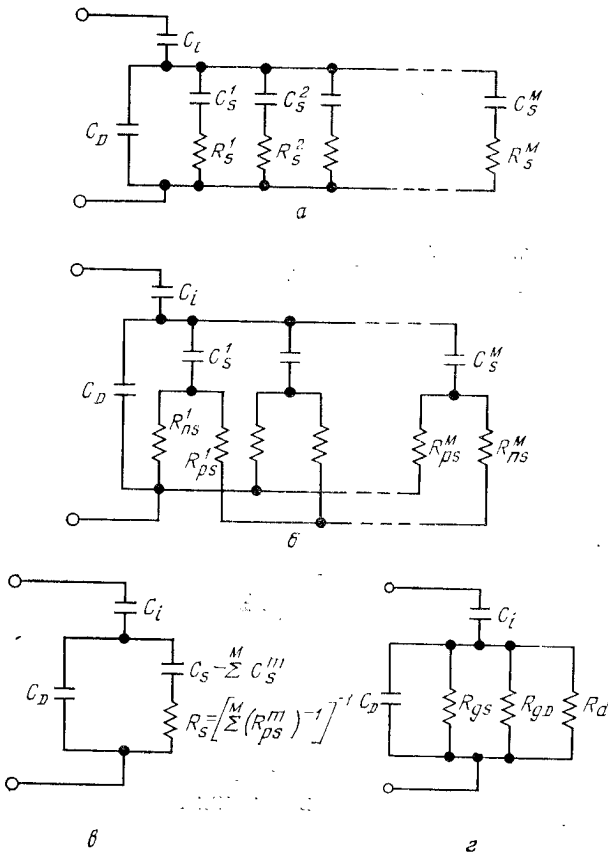


Рис. 22. Модифицированные эквивалентные электрические схемы МДП-структур [26, 15].

a — режим объединения; *b* — промежуточный режим между объединением и слабой инверсией; *в* — слабая инверсия; *г* — сильная инверсия.

лентные сопротивления захвата неосновных носителей R_{ns} много меньше эквивалентных сопротивлений R_{ps} , и поэтому эквивалентные емкости C_s^i можно считать соединенными параллельно. При этом эквивалентная схема на рис. 22, *б* становится схемой на рис. 22, *в*, частотные свойства которой характеризуются единственной постоянной времени $C_s R_s$. Это согласуется с экспериментальными частотными зависимостями активной проводимости МДП-структур в режиме слабой инверсии. Эквивалентная схема МДП-структуры в режиме сильной инверсии [15] ($\psi_s > 2\psi_B$) приведена на рис. 22, *г*, где C_i и C_D — емкости изолятора и полупроводника соответственно. Сопротивление R_d соответствует компоненте тока перезарядки инверсионного слоя, обусло-

вленной диффузией неосновных носителей из электронейтрального объема подложки к краю области обеднения, а R_{gs} — генерации неосновных носителей через поверхностные состояния. Сопротивление R_{gD} связано с генерацией и рекомбинацией носителей тока в области пространственного заряда. Экспериментально установлено, что в кремниевых МОП-структурах именно генерация и рекомбинация в области пространственного заряда являются главными процессами, определяющими частотную зависимость отклика МОП-структуры в режиме сильной инверсии. Отметим, что в этом режиме емкость полупроводника, включающая емкость инверсионного слоя, много больше емкости поверхностных состояний, поэтому последней можно пренебречь.

7.3.2. Заряды в окисле

В соответствии с классификацией, предложенной в работе [19] и сформулированной в начале разд. 7.3, заряд в окисле состоит из фиксированного заряда окисла Q_f , заряда, захваченного в окисле Q_{ot} , и заряда подвижных ионов Q_m (рис. 15).

Фиксированный заряд окисла Q_f имеет следующие свойства: 1) величина остается неизменной в широком диапазоне изменений поверхностного потенциала ψ_s ; 2) локализован в слое толщиной порядка 30 \AA вблизи границы раздела $\text{Si} - \text{SiO}_2$ [7]; 3) величина слабо зависит от толщины слоя окисла типа проводимости и концентрации легирующей примеси в подложке; 4) в большинстве случаев положителен, а величина зависит от ориентации подложки и технологических режимов окисления и отжига структур. Предполагается, что фиксированный заряд Q_f обусловлен либо избыточным (трехвалентным) кремнием, либо избыточным (несвязанным, потерявшим один электрон) кислородом в приповерхностном слое SiO_2 . При анализе электрических характеристик МОП-структур фиксированный заряд Q_f можно рассматривать как заряженный слой, локализованный на границе раздела $\text{Si} - \text{SiO}_2$.

На рис. 23 приведены высокочастотные вольт-фарадные зависимости, сдвинутые вдоль оси напряжений в результате присутствия либо положительного, либо отрицательного фиксированного заряда Q_f на границе раздела [7]. Положение этих кривых характеризует так называемый сдвиг напряжения плоских зон, определенный по отношению к $C - V$ характеристике идеальной МДП-структуры с $Q_f = 0$. Независимо от типа проводимости подложки положительный заряд Q_f сдвигает $C - V$ -характеристику в сторону отрицательных напряжений смещения, а отрицательный заряд Q_f — в сторону положительных смещений.

Характер влияния фиксированного заряда Q_f на $C - V$ -характеристики можно легко пояснить с помощью рис. 24, где условно показано «поперечное сечение» МОП-структуры с положитель-

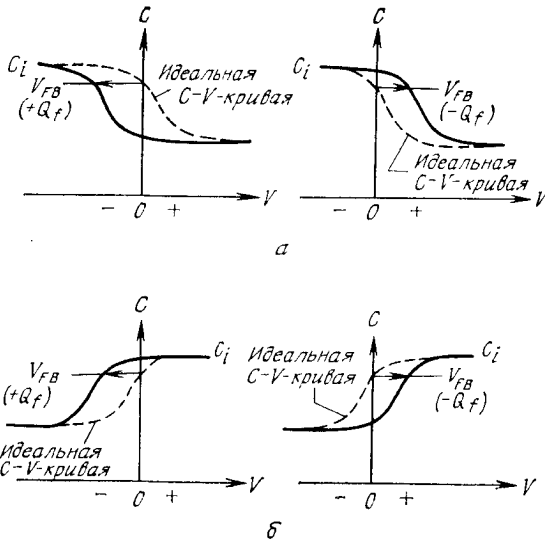


Рис. 23. Сдвиг $C - V$ -кривых вдоль оси напряжений, обусловленный положительным или отрицательным фиксированным зарядом окисла [7].
 а — для полупроводника p -типа; б — для полупроводника n -типа.

ным Q_f при отрицательном напряжении смещения. Для полной электронейтральности структуры необходимо, чтобы каждый отрицательный заряд на ее металлическом электроде компенсировался равным по величине и противоположным по знаку зарядом в диэлектрике или в полупроводнике. В идеальной МДП-структуре $Q_f = 0$, и эта компенсация осуществляется только за счет заряда ионизированных доноров в обедненном слое полупроводника. В реальной МОП-структуре с положительным Q_f часть зарядов на металлическом электроде компенсируется фиксированным за-

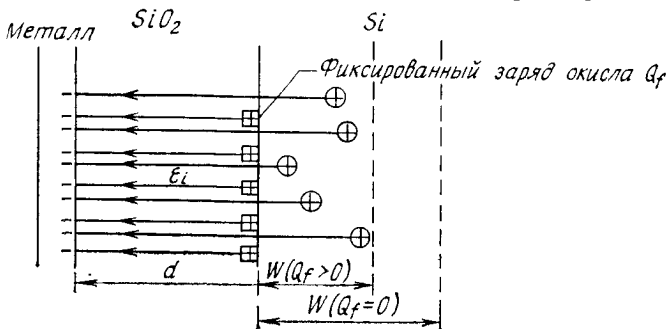


Рис. 24. Влияние фиксированного заряда окисла на свойства МДП-структур [24].

рядом окисла, что естественно приводит к соответствующему уменьшению глубины области обеднения по сравнению с идеальной МДП-структурой при том же напряжении смещения. Этот результат проиллюстрирован на рис. 24, где показано, что часть силовых линий электрического поля замыкается на фиксированном заряде Q_f . Так как глубина области обеднения уменьшается, емкость МОП-структуры увеличивается, и, следовательно, вся $C - V$ -кривая сдвигается по отношению к идеальной в сторону отрицательных напряжений. При отрицательных Q_f $C - V$ -кривая сдвигается в противоположном направлении. Абсолютная величина этого сдвига

$$\Delta V_f = \frac{Q_f}{C_i}. \quad (44)$$

Как впервые установлено в работе [30], присутствие ионов щелочных металлов, таких, как натрий, в термически выращенном SiO_2 является основной причиной нестабильности характеристик полупроводниковых приборов, пассивированных с помощью окисления. Ионы щелочных металлов влияют также и на надежность работы полупроводниковых приборов при повышенных температурах и напряжениях, поскольку в этих условиях ионы могут передвигаться в слое диэлектрика, что приводит к нежелательным изменениям характеристик приборов (например, сдвигу пороговых напряжений).

На рис. 25 приведено распределение ионов натрия в слое SiO_2 (штриховая кривая), установившееся в результате термополевого дрейфа [31]. В этой структуре исходная толщина слоя окисла составляла 5400 \AA , а концентрация акцепторов в кремнии $5 \times 10^{14} \text{ см}^{-3}$. Отметим, что концентрация ионов натрия на границе раздела довольно велика ($3 \cdot 10^{17} \text{ см}^{-3}$). Это приводит к большому сдвигу напряжения плоских зон и нестабильности характеристик. По закону Гаусса сдвиг напряжения плоских зон, обусловленный рядом подвижных ионов, составляет

$$\Delta V_m = \frac{Q_m}{C_i} = \frac{1}{C_i} \left[\frac{1}{d} \int_0^d x \rho_m(x) dx \right], \quad (45)$$

где Q_m — эффективный заряд подвижных ионов, приходящийся на единицу площади границы раздела, а $\rho_m(x)$ [$\text{Кл} \cdot \text{см}^{-3}$] — объемная плотность заряда подвижных ионов. На рис. 25 приведены профили концентрации, полученные в процессе диффузии натрия в пленках нитрида кремния [32] с различными размерами кристаллитов. Отметим, что наиболее устойчивым по отношению к проникновению ионов натрия оказывается аморфный Si_3N_4 .

Чтобы предотвратить проникновение подвижного ионного заряда в окисел в процессе рабочего цикла приборов, можно исполь-

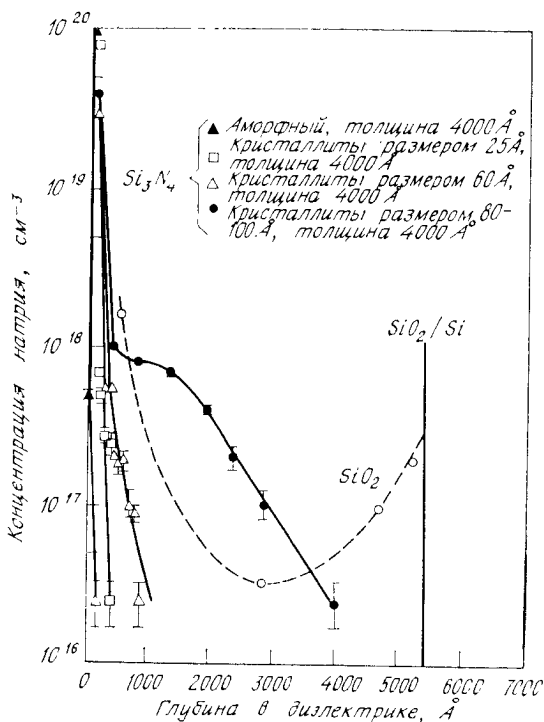
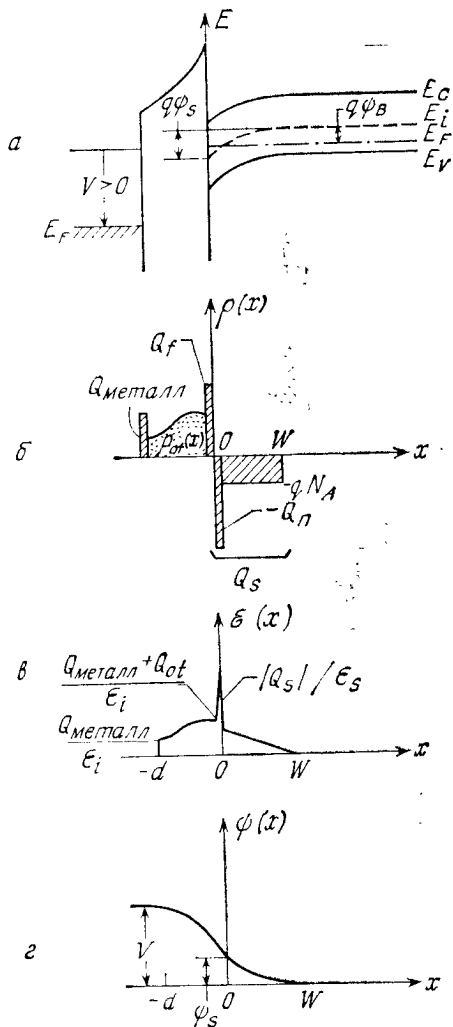


Рис. 25. Распределение концентрации ионов натрия по толщине пленок двуокиси кремния и нитрида кремния при различных размерах кристаллитов [31, 32].

звать непроницаемые для подвижных ионов защитные пленки, такие, как, например, аморфный нитрид кремния. Окисел Al_2O_3 и фосфоросиликатные стекла также могут выполнять роль барьерных слоев для ионов натрия. Однако МДП-структуры с составным изолятором металл — $(Al_2O_3 - SiO_2) - Si$ характеризуются довольно большим сдвигом $C - V$ -кривых в область отрицательных напряжений. Это означает, что в структурах на p -подложке поверхность уже инвертирована при нулевом и даже при не слишком больших отрицательных напряжениях смещения. Фосфоросиликатные стекла [34], представляющие собой окисел SiO_2 , обогащенный P_2O_5 , образуются вблизи внешней поверхности слоя окисла в процессе диффузии фосфора и существенно уменьшают нестабильность рабочих характеристик МОП-структур за счет захвата ионов натрия, поскольку растворимость натрия в этих стеклах намного выше, чем в SiO_2 . Структуры с составным изолятором металл — $(Si_3N_4 - SiO_2) - Si$ используются в настоящее время главным образом как энергонезависимые элементы памяти

Рис. 26. МДП-структура с фиксированным и захваченным в окисле зарядами.

a — зонная диаграмма; *б* — распределение варяда; *в* — электрическое поле; *г* — потенциал.



(гл. 8). Эти структуры характеризуются малыми значениями плотности поверхностных состояний, что обусловлено высоким качеством границы раздела $Si - SiO_2$. Кроме того, внешний слой Si_3N_4 обеспечивает их устойчивость по отношению к ионному дрейфу. В видимой области спектра электромагнитного излучения пленки Si_3N_4 имеют несколько большую диэлектрическую проницаемость, чем пленки SiO_2 . В табл. 2 проведено сравнение цветов пленок SiO_2 и Si_3N_4 соответствующей толщины [36].

Заряд, захваченный в слое окисла $Q_{от}$, также приводит к сдвигу $C - V$ -характеристик МОП-структур. Этот заряд обусловлен

Таблица 2. Сравнение цветов пленок SiO_2 и Si_3N_4

Порядок	Цвет	Диапазон толщины ¹ пленки SiO_2 , мкм	Диапазон толщины пленки Si_3N_4 , мкм
Первый	Песочный	0—0,027	0—0,020
	Коричневый	0,027—0,053	0,020—0,040
	Золотисто-коричневый	0,053—0,073	0,040—0,055
	Красный	0,073—0,097	0,055—0,073
	Темно-голубой	0,097—0,10	0,073—0,077
	Голубой	0,10—0,12	0,077—0,093
	Светло-голубой	0,12—0,13	0,093—0,10
	Бледно-голубой	0,13—0,15	0,10—0,11
	Песочный	0,15—0,16	0,11—0,12
	Светло-желтый	0,16—0,17	0,12—0,13
Второй	Желтый	0,17—0,20	0,13—0,15
	Красно-оранжевый	0,20—0,24	0,15—0,18
	Красный	0,24—0,25	0,18—0,19
	Темно-красный	0,25—0,28	0,19—0,21
	Голубой	0,28—0,31	0,21—0,23
	Зелено-голубой	0,31—0,33	0,23—0,25
	Светло-зеленый	0,33—0,37	0,25—0,28
	Желто-оранжевый	0,37—0,40	0,28—0,30
	Красный	0,40—0,44	0,30—0,33

¹ Отношение показателей преломления равно $\frac{\bar{n}(\text{Si}_3\text{N}_4)}{\bar{n}(\text{SiO}_2)} = \frac{1,97}{1,48} = 1,33 =$
 $= \frac{\text{толщина SiO}_2}{\text{толщина Si}_3\text{N}_4}$.

структурными дефектами в слое окисла. Ловушки в окисле обычно нейтральны, но могут заряжаться, захватывая электроны и дырки. На рис. 26 приведены зонная диаграмма и распределения заряда, электрического поля и потенциала в МОП-структуре, содержащей как фиксированный заряд, так и заряд, захваченный в окисле. Из сравнения рис. 26 и 6 следует, что в первом случае те же значения поверхностного потенциала ψ_s достигаются при меньших напряжениях смещения ($C - V$ -кривая сдвигается в область отрицательных напряжений). Сдвиг напряжения, обусловленный зарядом, захваченным в окисле, записывается в виде

$$\Delta V_{0i} = \frac{Q_{0i}}{C_i} = \frac{1}{C_i} \left[\frac{1}{d} \int_0^d x \rho_{0i}(x) dx \right], \quad (46)$$

где Q_{0i} — эффективная поверхностная плотность этого заряда, приведенная к единице площади границы раздела $\text{Si} - \text{SiO}_2$, а ρ_{0i} — истинная объемная плотность заряда, захваченного

в окисле. Результирующий сдвиг напряжения плоских зон ΔV_{FB} , обусловленный всеми компонентами заряда в окисле, есть

$$\Delta V_{FB} = \Delta V_f + \Delta V_m + \Delta V_{ot} = \frac{Q_0}{C_i}, \quad (47)$$

где $Q_0 \equiv Q_f + Q_m + Q_{ot}$ — сумма соответствующих эффективных зарядов на единицу площади границы раздела Si — SiO₂.

7.3.3. Влияние разности работ выхода и других внешних факторов

Разность работ выхода. Согласно нашему определению, в идеальной МДП-структуре разность работ выхода электрона из металла и полупроводника равна нулю (рис. 2):

$$\varphi_{ms} \equiv \varphi_m - \left(\chi + \frac{E_g}{2q} - \psi_B \right). \quad (48)$$

Если же эта разность отлична от нуля, а кроме того, в диэлектрике МДП-структуры присутствует заряд Q_0 (выражение (47)), вольт-фарадные характеристики реальной МДП-структуры будут сдвинуты вдоль оси напряжений относительно идеальной $C - V$ кривой на величину

$$V_{FB} = \varphi_{ms} - \frac{Q_0}{C_i} = \varphi_{ms} - \frac{Q_f + Q_m + Q_{ot}}{C_i}, \quad (49)$$

которая называется сдвигом напряжения плоских зон. (Для простоты мы не учитывали заряд, захваченный на поверхностных состояниях.) Если по каким-либо причинам можно пренебречь величиной заряда подвижных ионов и зарядом, захваченным на объемных ловушках диэлектрика, выражение (49) упрощается и принимает вид

$$V_{FB} = \varphi_{ms} - \frac{Q_f}{C_i}. \quad (49a)$$

Энергетическая зонная диаграмма границы раздела Si — SiO₂ была определена по результатам измерений фотоэмиссии электронов [37]. Установлено, что ширина запрещенной зоны SiO₂ примерно равна 9 эВ, а сродство к электрону $q\chi_i = 0,9$ эВ. Работа выхода из металла в МОП-структурах обычно определяется по результатам измерений фотоотклика или вольт-фарадных характеристик. На рис. 27 приведены экспериментальные зависимости фотоотклика МОП-структур с электродами из различных металлов от энергии возбуждающих фотонов [38]. По оси ординат здесь отложен корень кубический из фотоотклика в произвольных единицах. Пересечение этих прямых с осью абсцисс определяет величину энергетического барьера $q\varphi_B$ между металлом и SiO₂.

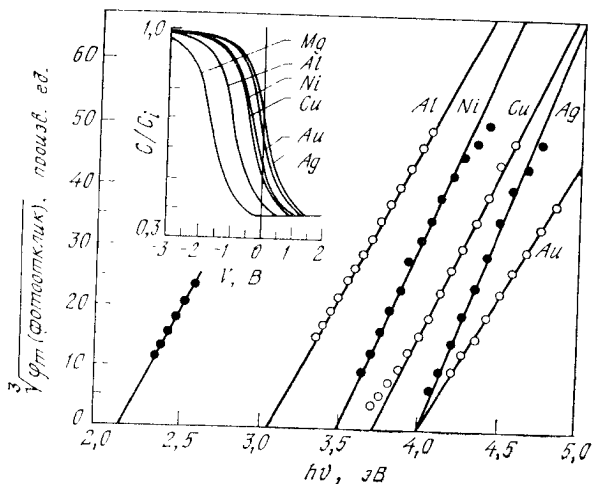


Рис. 27. Зависимость фотоотклика МОП-структур с различными металлическими электродами от энергии фотонов. На вставке приведены соответствующие $C - V$ -кривые [38].

Следовательно, работа выхода из металла Φ_m равна сумме $\Phi_B + \chi_i$, где χ_i — сродство к электрону в окисле (рис. 2). Подобные результаты получаются также из анализа вольт-фарадных характеристик серии экспериментальных МОП-структур с одинаковыми параметрами окисного слоя и подложки, различающихся лишь материалом полевого электрода (вставка на рис. 27). Как следует из уравнения (49), $C - V$ характеристики таких МОП-структур должны быть сдвинуты относительно друг друга на величину, равную разности работ выхода из соответствующих металлов: $\Phi_{m1} - \Phi_{m2} = \Phi_{B1} - \Phi_{B2}$. Поэтому, если работа выхода электрона для одного из металлов известна, по сдвигу $C - V$ -кривых можно определить абсолютные значения Φ_m для всех остальных металлов.

Экспериментальные значения Φ_m для ряда металлов, определенные указанными методами, приведены в табл. 3, где указаны также работы выхода из соответствующих металлов в вакуум. Отметим совпадение значений Φ_m , определенных вольт-фарадным и фотоэмиссионным методами, и заметное отличие этих значений от работы выхода электрона в вакуум. Последнее не является неожиданным, если учесть поликристалличность металлических пленок, нанесенных в качестве полевого электрода на слой окисла металла — окисел довольно сильно отличаются от условий на границе монокристаллический металл — вакуум при измерениях «вакуумной» работы выхода электрона. С помощью рассмотренных выше экспериментальных методов было также установлено, что

Таблица 3. Работа выхода металлов

Металл	Φ_m (по $C - V$ -критериям), В	Φ_m (по фотоотклику), В	Φ_m (работа выхода в вакууме), В
Mg	3,35	3,15	3,7
Al	4,1 ¹	4,1	4,25
Ni	4,55	4,6	4,5
Cu	4,7	4,7	4,25
Au	5,0	5,0	4,8
Ag	5,1	5,05	4,3

¹ Это значение Φ_m для Al является суммой высоты барьера (3,2 В) и сродства к электрону в SiO_2 (0,9 В).

высота энергетического барьера на границе кремний — двуокись кремния практически не зависит от кристаллической ориентации подложки (в пределах погрешности 0,1 эВ).

Из приведенных выше результатов следует, что разность работ выхода Φ_{ms} может составлять заметную долю наблюдаемого сдвига напряжения плоских зон МОП-структуры, и, таким образом, ее необходимо учитывать (см. уравнение (49)) при оценках величины фиксированного заряда окисла по сдвигу $C - V$ -характеристик. В качестве примера на рис. 28, а приведена зонная диаграмма МОП-структуры с алюминиевым полевым электродом и окислом толщиной 500 Å, термически выращенным на кремниевой подложке n -типа ($N_D \approx 10^{18} \text{ см}^{-3}$). В этом случае $\Phi_m = 4,1$ В, работа выхода из полупроводника $\chi + E_g/2q = \Phi_B = 4,35$ В и, следовательно, разность работ выхода $\Phi_{ms} = 4,1 - 4,35 = -0,25$ В.

При изготовлении современных интегральных схем в качестве затворного электрода широко применяются пленки сильнолегированного поликремния. Для поликремниевых затворов n^+ -типа, где уровень Ферми практически совпадает с положением дна зоны проводимости, эффективная работа выхода Φ_m равна величине сродства к электрону в кремнии ($\chi_{\text{Si}} = 4,15$ В). В поликремниевых затворах p^+ -типа, где уровень Ферми лежит у потолка валентной зоны, эффективная работа выхода $\Phi_m = \chi_{\text{Si}} + E_g/q = 5,25$ В.

На рис. 28, б приведена зависимость разности работ выхода Φ_{ms} от уровня легирования кремниевой подложки для МОП-структур с затворными электродами из Al, Au и поликремния n^+ - и p^+ -типа [38, 39]. Из этих графиков следует, что в зависимости от материала затвора при нулевом напряжении смещения приповерхностная область полупроводника МОП-структуры может оказаться практически в любом состоянии (от аккумуляции до инверсии).

Влияние внешних факторов. Внешние факторы, такие, как температура, облучение структуры и ионизирующая радиация,

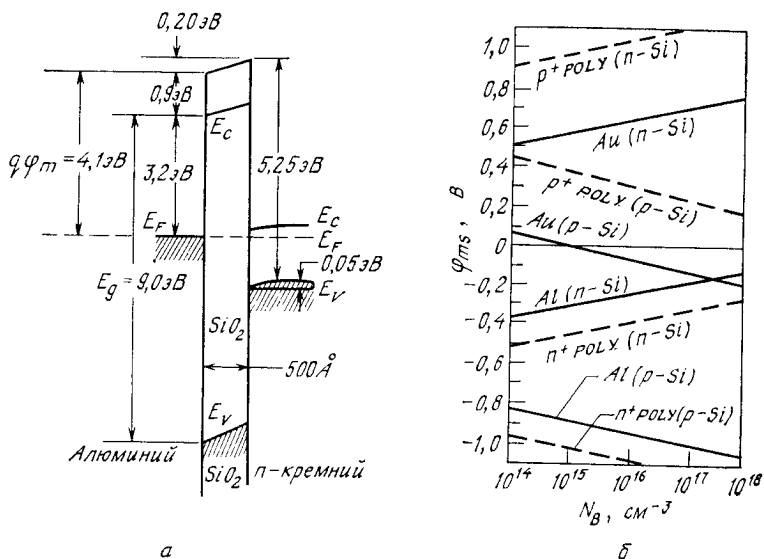


Рис. 28. Зонная диаграмма (а) структуры Al — SiO₂ — Si с толщиной окисла 500 Å и $N_D = 10^{16}$ см⁻³ и зависимость (б) разности работ выхода от степени легирования в МОП-структурах с электродами из вырожденного поликремния, Al и Au [38, 39].

могут оказывать существенное влияние на характеристики МОП-структур. Рассмотрим сначала влияние температуры. Как известно, заряд инверсионного слоя в МОП-структурах связан с объемом полупроводниковой подложки только через генерационно-рекомбинационные процессы, т. е. последние обеспечивают установление стационарного значения плотности заряда инверсионного слоя, соответствующего заданному напряжению на структуре. Характерное время установления такого равновесия определяется интенсивностью генерационно-рекомбинационных процессов, которая сильно (экспоненциально) зависит от температуры. Обычно соответствующая граничная частота для Si — SiO₂-систем при комнатной температуре не превышает 100 Гц, а в отдельных случаях она ниже 1 Гц. При понижении температуры скорость накопления и рассасывания заряда инверсионного слоя настолько уменьшается, что, например, при рассасывании инверсионного слоя область пространственного заряда структуры оказывается под прямым смещением до 0,25 В, чтобы обеспечить достаточную инжекцию инверсионного заряда в подложку [40]. При этом происходит значительная деформация $C - V$ -характеристик структуры в области инверсии (гистерезис), и для их правильного определения необходимо в процессе измерений довольно долго выдерживать МОП-структуру при данном напря-

жении (в каждой экспериментальной «точке»). В области повышенных температур скорость генерации существенно увеличивается, что значительно облегчает изучение влияния температуры на свойства МДП-структур. В первую очередь это относится к изучению механизмов генерационных явлений.

Эквивалентная электрическая схема МОП-структуры в режиме сильной инверсии приведена на рис. 22, а на рис. 29 приведены температурные зависимости полной проводимости $G = R_{gs}^{-1} + R_{gd}^{-1} + R_d^{-1}$ экспериментальной МОП-структуры на n -подложке. Напомним, что скорость рекомбинационного процесса в области пространственного заряда пропорциональна величине n_i , т. е. характеризуется энергией активации $E_g/2$, в то время как скорость диффузионного процесса накопления пропорциональна n_i^2 с энергией активации, равной E_g . Данные, приведенные на рис. 29, показывают, что генерационно-рекомбинационный процесс в области пространственного заряда $R_{gd}^{-1} \sim n_i$ является доминирующим механизмом установления термодинамического равновесия при температурах ниже 140°C . В этом температурном диапазоне значение энергии активации $0,56\text{ эВ}$, соответствующее экспериментальной зависимости на рис. 29, практически совпадает с ожидаемой величиной $E_g/2$. При температурах выше 140°C

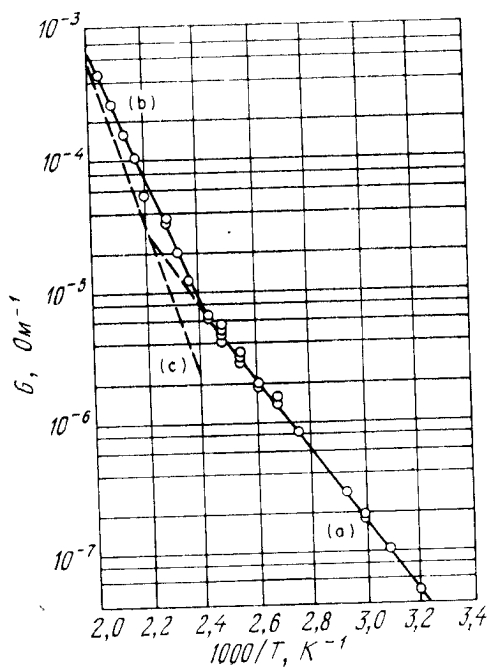


Рис. 29. Температурная зависимость проводимости [41].

наклон экспериментальной кривой увеличивается. Это свидетельствует о том, что в данной области преобладает другой генерационно-рекомбинационный механизм. Оказалось, что характерный наклон линии (с), которая получена вычитанием из полной проводимости G (кривая (b)) проводимости области пространственного заряда R_{gb}^{-1} (кривая (a)), равен $1,17 \text{ эВ} \approx E_g$, что соответствует ожидаемому результату для диффузионного механизма $R_d^{-1} \sim n_i^2$. Отметим, что приведенные выше экспериментальные результаты свидетельствуют также о том, что эквивалентная схема на рис. 22, г хорошо описывает электрические свойства реальных МОП-структур в режиме сильной инверсии.

При облучении МОП-структур светом увеличивается высокочастотная емкость структуры на участке C — V -кривой, соответствующей режиму сильной инверсии. Причем с ростом интенсивности облучения высокочастотная емкость на этом участке все более приближается к своему низкочастотному значению C_i . Такое поведение обусловлено, во-первых, уменьшением под действием облучения характерного времени генерации неосновных носителей τ_{inv} [16]. Во-вторых, световая генерация электронно-дырочных пар в приповерхностной области полупроводника МОП-структуры приводит к уменьшению ¹⁾ поверхностного потенциала ψ_s при постоянном полном напряжении, приложенном к структуре (рис. 30, а). Уменьшение ψ_s приводит к соответствующему уменьшению толщины области пространственного заряда, а следовательно, к увеличению высокочастотной емкости [42]. Последний механизм доминирует при достаточно высоких частотах измерительного сигнала.

Основной процесс, обуславливающий изменение характеристик МОП-структур под действием ионизирующих излучений, таких, как рентгеновское [43] и γ -излучения [44], иллюстрирует зонная схема на рис. 30, б. Проникающее в слой окисла ионизирующее излучение, разрывая связи Si — O, генерирует электронно-дырочные пары, которые затем разделяются электрическим полем в окисле, если в процессе экспозиции МОП-структура

¹⁾ При оптическом облучении инвертированной МДП-структуры область пространственного заряда последней оказывается как бы под прямым смещением, равным $\psi_l = \frac{kT}{q} \ln \frac{I_\phi}{I_s}$ и необходимым для того, чтобы током инжекции инверсионных носителей в подложку полностью скомпенсировать фототок I_ϕ , который устремляется на границу раздела в результате аккумуляции областью пространственного заряда электронно-дырочных пар, генерированных фотонами в приповерхностном слое полупроводниковой подложки. При этом $\psi_s = 2\varphi_B - \psi_l$. Указанное уменьшение ψ_s под действием облучения приводит также к соответствующему увеличению эффективной скорости генерации, пропорциональному $\exp \left\{ \frac{q\psi_l}{2kT} \right\}$. — Прим. перев.

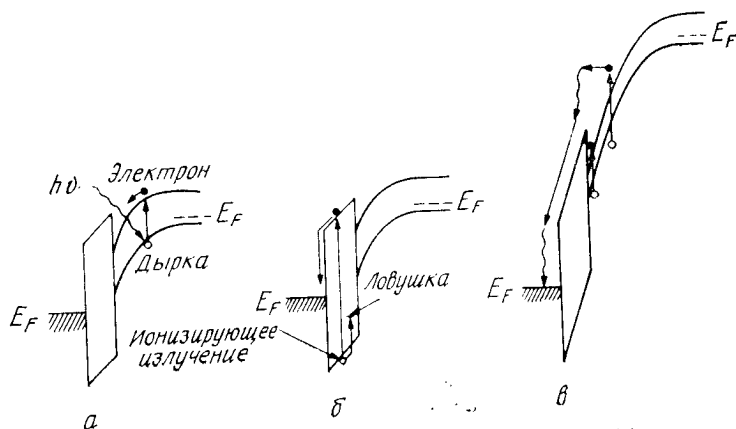


Рис. 30. Зонные диаграммы МДП-структуры при освещении (а), радиационных воздействиях (б) и лавинной инжекции (в) [7].

находится под напряжением. Более подвижные электроны при этом достаточно быстро покидают окисел и уходят в затвор или в подложку, а более медленные дырки частично захватываются ловушками в окисле. При положительном напряжении на затворе (рис. 30, б) эти дырки захватываются вблизи границы с полупроводником и, следовательно, оказывают сильное влияние на характеристики МОП-структуры. При этом наблюдается значительный радиационный сдвиг напряжения плоских зон в сторону отрицательных напряжений (образование положительного фиксированного заряда). Возможно, что эти захваченные дырки ответственны также и за увеличение плотности поверхностных состояний, которое обычно наблюдается после радиационных воздействий на МОП-структуру [7]. При противоположной полярности напряжения (минус на затворе) дырки будут захватываться вблизи металлического электрода, который сильно экранирует их влияние на полупроводник. В этом случае величина эффективного положительного заряда, индуцированного радиацией, оказывается значительно меньшей, и вольт-фарадные характеристики претерпевают относительно малый сдвиг вдоль оси напряжений. Отметим, что экспериментальные результаты хорошо соответствуют изложенной модели дырочного захвата в окисле.

Рассмотрим также влияние лавинной инжекции в окисел [45], которая может происходить в МОП-структурах при их работе в условиях глубокого обеднения (рис. 7, кривая в)). При быстром изменении напряжения на электроде структуры почти все напряжение оказывается приложенным к нестационарному обедненному слою в полупроводнике, где при достаточно большой величине

этого напряжения развивается лавинный процесс ударной ионизации. В результате лавинного процесса у границы с окислом возникает довольно большое число электронов и дырок с энергиями, достаточными для преодоления соответствующих энергетических барьеров. Так как величина электронного энергетического барьера на границе Si — SiO₂, равная $q\chi_{Si} - q\chi_i = 4,1 - 0,9 = 3,2$ эВ, меньше, чем высота энергетического барьера для дырок $E_g(\text{SiO}_2) - E_g(\text{Si}) - q\chi_{Si} - q\chi_i = 4,7$ эВ, электронная инжекция в окисел существенно превышает дырочную.

Условия возникновения лавинного пробоя в МОП-структурах исследованы в работе [46] на основе двумерной модели распределения электрического поля (вставка на рис. 31). Электрическое поле у границы раздела полупроводник — диэлектрик неодинаково и достигает максимального значения \mathcal{E}_m вблизи края затворного электрода. Напряжение пробоя структуры определяется как напряжение на затворе, при котором ионизационный интеграл вдоль оптимальной силовой линии электрического поля достигает значения, равного 1. Оптимальная силовая линия начинается в той точке границы раздела Si — SiO₂ вблизи края электрода, где $\mathcal{E} = \mathcal{E}_m$, и кончается на краю обедненной области, где $\mathcal{E} = 0$. Результаты расчета пробивного напряжения для МОП-структур, различающихся толщиной окисла и уровнем легирования подложки, представлены на рис. 31. Как видно из графиков, при каждом определенном значении толщины окисла существует концентрация легирующей примеси, при которой напряжение лавинного пробоя оказывается минимальным. Было установлено, что слева от этих минимумов лавинный пробой действительно происходит у края электрода ($\mathcal{E}_m > \mathcal{E}_1$), а правой части графиков соответствует пробой, однородный по площади структуры ($\mathcal{E}_m = \mathcal{E}_1$). В условиях однородного пробоя величина электрического поля на границе раздела \mathcal{E}_1 тем больше, чем больше концентрация легирующей примеси (гл. 2). При этом соответственно увеличиваются электрическое поле в окисле и падение напряжения на нем, что приводит к росту пробивного напряжения с увеличением концентрации в правой части графиков (рис. 31). Отметим в заключение, что лавинный пробой происходит однородно по площади электрода, когда отношение $d/W_{\text{макс}}$ (где $W_{\text{макс}}$ — максимальная толщина обедненного слоя при напряжении пробоя) превышает 0,3.

На рис. 32 приведены зависимости емкости и поперечной проводимости кремниевой МОП-структуры с окислом, выращенным во влажной атмосфере, измеренные до и после лавинной инжекции электронов. Видно, что инжекция горячих электронов в окисел приводит к сдвигу напряжения плоских зон в сторону более положительных напряжений (это свидетельствует об увеличении отрицательного заряда в окисле). При этом увеличивается амплитуда

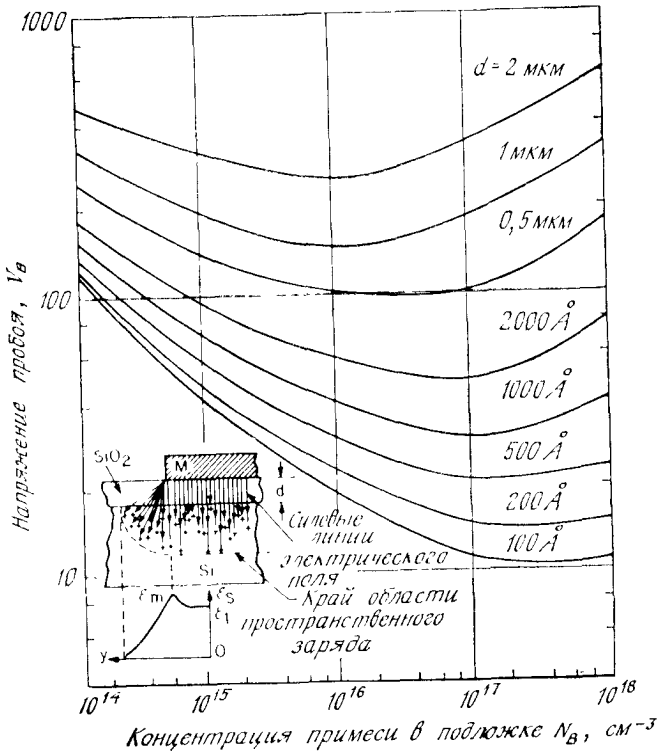


Рис. 31. Зависимость напряжения пробоя МДП-структуры в условиях глубокого обеднения от концентрации примеси в подложке. Параметром кривых является толщина окисла. На вставке показано распределение электрического поля вдоль границы с окислом [46].

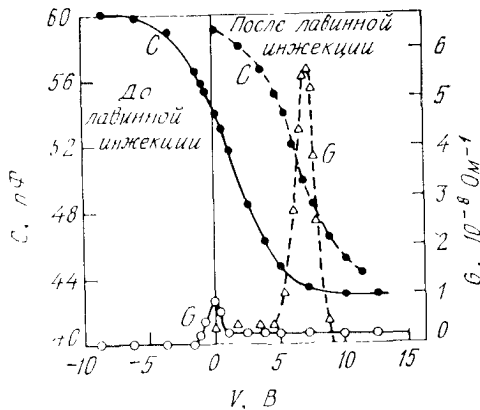


Рис. 32. Зависимости емкости и проводимости МДП-структуры от напряжения до и после лавинной инжекции электронов в окисел [7].

пика поперечной проводимости, что говорит о росте плотности поверхностных состояний (от $1,2 \cdot 10^{11}$ до $7,9 \cdot 10^{11}$ $\text{см}^{-2} \cdot \text{эВ}^{-1}$). Следует отметить, что лавинная инжекция или инжекция горячих носителей имеет самое непосредственное отношение к выбору рабочих режимов полупроводниковых приборов со структурой металл — диэлектрик — полупроводник. Так, например, в МОП-транзисторе носители в канале могут ускоряться в продольном электрическом поле от истока к стоку до энергий, достаточных для преодоления энергетического барьера на границе Si — SiO₂. Захват этих носителей в окисле приводит к возникновению дополнительного фиксированного заряда и к увеличению плотности поверхностных состояний. Все это нежелательно, поскольку приводит к изменению рабочих параметров прибора (дрейфу характеристик). С другой стороны, с помощью лавинной инжекции при пробое стокового перехода МОП-транзистора можно, например, заряжать сквозь окисел поликремниевый затвор. Этот метод используется в полупроводниковых элементах памяти некоторых типов.

7.3.4. Явления переноса в пленках изоляторов

Проводимость слоя изолятора в идеальной МДП-структуре предполагается равной нулю. Реальные диэлектрики обладают проводимостью, особенно заметной при повышенных температурах и электрических полях. По закону сохранения электрической индукции поле в диэлектрике \mathcal{E}_i и электрическое поле \mathcal{E}_s в полупроводнике на границе с диэлектриком связаны соотношением

$$\mathcal{E}_i = \mathcal{E}_s \left(\frac{\epsilon_s}{\epsilon_i} \right), \quad (50)$$

где ϵ_s и ϵ_i — соответствующие диэлектрические проницаемости. Так, например, в системе Si — SiO₂, когда поле в кремнии достигает характерной для лавинного пробоя величины [47] $3 \times 10^5 \text{ В} \cdot \text{см}^{-1}$, соответствующее поле в окисле примерно в три раза больше ($\sim 10^6 \text{ В} \cdot \text{см}^{-1}$), поскольку $\epsilon_{\text{Si}}/\epsilon_{\text{SiO}_2} = 11,7/3,9$. При таких значениях электрического поля электронная и дырочная проводимость SiO₂ все еще пренебрежимо мала даже при повышенных температурах. Однако уже при этих полях в окисле может происходить заметный дрейф подвижных ионов, таких, как ионы натрия, что приводит к нестабильности характеристик приборов и гистерезисным эффектам. В ультратонких слоях SiO₂, а также при очень больших электрических полях проводимость определяют "эффекты" туннелирования. Отметим, что проводимость других изолирующих слоев, таких, как Si₃N₄ и Al₂O₃, обычно намного превышает проводимость двуокиси кремния.

Таблица 4. Основные механизмы проводимости диэлектриков

Механизм	Выражение	Зависимость от температуры и напряжения
Эмиссия Шоттки	$J = A^* T^2 \exp \times \left[\frac{-q(\varphi_B - \sqrt{q\mathcal{E}_i/4\pi\epsilon_i})}{kT} \right]$	$\sim T^2 \exp(+a\sqrt{V}/T - q\varphi_B/kT)$
Эмиссия Пула—Френкеля	$J \sim \mathcal{E} \exp \times \left[\frac{-q(\varphi_B - \sqrt{q\mathcal{E}_i/\pi\epsilon_i})}{kT} \right]$	$\sim V \exp(+2a\sqrt{V}/T - q\varphi_B/kT)$
Туннельная или полевая эмиссия	$J \sim \mathcal{E}^2 \exp \times \left[\frac{4\sqrt{2m^*}(q\varphi_B)^{3/2}}{3q\hbar\mathcal{E}} \right]$	$\sim V^2 \exp(-b/V)$
Ограничение пространственным зарядом	$J = \frac{8\epsilon_i \mu V^2}{9d^3}$	$\sim V^2$
Омическая проводимость	$J \sim \mathcal{E} \exp(-\Delta E_{ae}/kT)$	$\sim V \exp(-c/T)$
Ионная проводимость	$J \sim \frac{\mathcal{E}}{T} \exp(-\Delta E_{ai}/kT)$	$\sim \frac{V}{T} \exp(-d'/T)$

A^* — постоянная Ричардсона; φ_B — высота барьера; \mathcal{E} — электрическое поле; ϵ_i — динамическая диэлектрическая проницаемость изолятора; m^* — эффективная масса; d — толщина изолятора; ΔE_{ae} — энергия активации электронов; ΔE_{ai} — энергия активации ионов; $a = \sqrt{q/4\pi\epsilon_i}d$, $V = \mathcal{E}d$; b , c и d — положительные константы, не зависящие от напряжения и температуры.

Основные физические процессы, определяющие проводимость изолирующих пленок, представлены в табл. 4. Мы не будем подробно останавливаться на эмиссии Шоттки, поскольку этот процесс уже рассматривался в гл. 5 при обсуждении термоионной эмиссии носителей тока на границе раздела металл — изолятор или изолятор — полупроводник. Напомним только, что при реализации этого механизма график зависимости $\ln(J/T^2)$ от величины $1/T$ представляет собой прямую линию, наклон которой зависит от величины диэлектрической проницаемости изолятора ϵ_i .

Эмиссия Пула — Френкеля [48], механизм которой проиллюстрирован на рис. 33 (вставка), представляет собой ускоренный электрическим полем процесс термовозбуждения электронов с ловушек в зону проводимости диэлектрика. В случае кулоновских ловушек соответствующее выражение для тока фактически совпадает с выражением для эмиссии Шоттки. При этом, однако, φ_B означает глубину ловушки, а не высоту энергетического барьера. Кроме того, для механизма Пула — Френкеля слагаемое

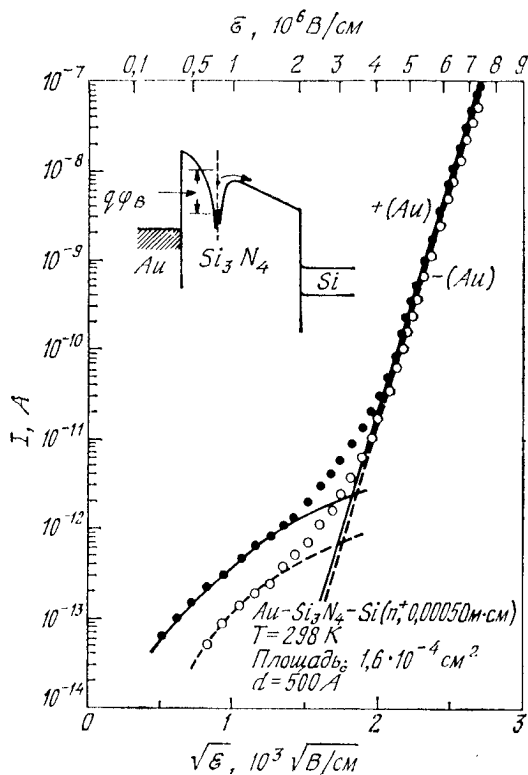


Рис. 33. Вольт-амперные характеристики структуры Au — Si₃N₄ — Si при комнатной температуре. Вставка иллюстрирует механизм эмиссии Пула — Френкеля [50].

$\sqrt{q/\pi\epsilon_i}$, соответствующее полевому уменьшению энергетического барьера, в два раза больше, чем для эмиссии Шоттки. Туннельному механизму проводимости соответствуют либо чисто полевая ионизация захваченных электронов с ловушек диэлектрика в его зону проводимости, либо непосредственные туннельные переходы электронов из металла в зону проводимости диэлектрика. Характерными свойствами туннельного механизма являются слабая температурная зависимость тока и очень сильная зависимость его величины от приложенного напряжения. Токи, ограниченные пространственным зарядом, наблюдаются при сильной монополярной инжекции носителей тока в диэлектрик и при отсутствии в нем подвижных компенсирующих зарядов противоположного знака. В этом случае, если пренебречь захватом электронов на ловушки, ток пропорционален квадрату приложенного напря-

жения. При низких напряжениях и повышенных температурах ток в диэлектрике может определять обычное термовозбуждение электронов с ловушек в зону проводимости. При этом вольт-амперная характеристика линейна, а соответствующая проводимость экспоненциально зависит от $1/T$. Ионная проводимость диэлектриков подобна диффузионным процессам. Однако постоянная компонента ионного тока существенно уменьшается с течением времени после приложения напряжения, поскольку инжекция ионов в диэлектрик и их экстракция из него сильно затруднены. Поэтому сравнительно большие начальные ионные токи приводят лишь к перераспределению ионного заряда от одного края диэлектрика к другому. После снятия внешнего напряжения в диэлектрике остается большое внутреннее поле, расталкивающее ионы. Однако за конечные времена равновесное распределение ионного заряда по толщине диэлектрика не успевает установиться, что приводит к соответствующим гистерезисным эффектам.

Применительно к конкретным изолирующим слоям каждый из рассмотренных механизмов проводимости обычно доминирует в определенном диапазоне напряжений и температур. Однако при анализе экспериментальных результатов следует принимать во внимание и то, что эти основные механизмы не в полной мере независимы друг от друга. Так, например, туннельные характеристики с учетом перераспределения электрического поля при больших плотностях пространственного заряда приближаются к форме характеристик, соответствующих эмиссии Шоттки [49].

На рис. 33 в качестве примера приведены вольт-амперные характеристики пленок нитрида кремния [50], построенные в так называемых координатах Пула — Френкеля ($\ln I - \sqrt{\mathcal{E}}$). Пленки Si_3N_4 осаждались на подложку из вырожденного кремния n -типа ($\rho = 0,0005 \text{ Ом}\cdot\text{см}$) в процессе реакции $\text{SiCl}_4 + \text{NH}_3$ при 1000°C . В качестве электрода напылялась пленка золота Au. Отметим, что при высоких полях зависимости, соответствующие положительным и отрицательным напряжениям, практически идентичны. Небольшое различие этих характеристик при малых полях, по-видимому, связано с отличием высоты энергетических барьеров на границах нитрид — золото и нитрид — кремний. Отметим также, что при высоких полях ток сквозь пленку нарастает экспоненциально (показатель экспоненты пропорционален корню квадратному из электрического поля), а при малых полях вольт-амперная характеристика является линейной. Было установлено, что при комнатных температурах форма приведенных характеристик практически не зависит от толщины пленки, площади и материала верхнего электрода и полярности напряжения. Все это указывает на то, что проводимость пленок определяют объемные процессы, а не свойства приэлектродной области, как в диодах с барьером Шоттки. Отметим, кроме того, что при низких

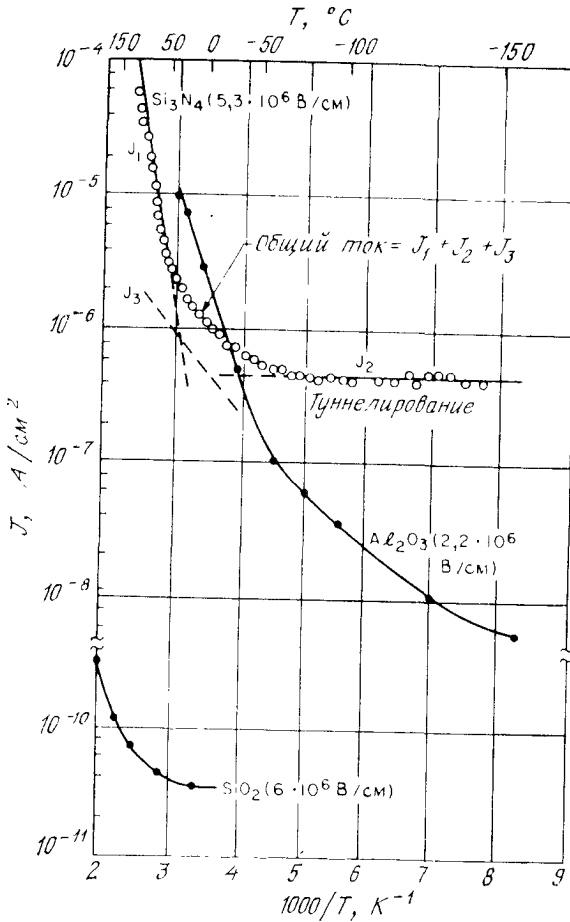


Рис. 34. Температурная зависимость плотности тока в пленках Si_3N_4 [50], Al_2O_3 [51] и SiO_2 [52].

температурах проводимость пленок нитрида кремния практически не зависит от температуры. На рис. 34 приведены температурные зависимости плотности тока в трех разных изоляторах. На кривой, полученной для пленки Si_3N_4 при электрическом поле $5,3 \times 10^6$ В·см⁻¹, можно выделить три компоненты: J_1 , J_2 и J_3 . Ток J_1 обусловлен эмиссией Пула — Френкеля. Он преобладает при высоких электрических полях и сравнительно высоких температурах. Отметим, что величина высокочастотной (динамической) диэлектрической проницаемости, которую можно определить по наклону этого участка характеристики ($\epsilon_i = 5,5$), близка к соответствующим значениям, полученным из оптических изме-

рений [36]. Ток J_3 обусловлен туннельной эмиссией электронов с ловушек в зону проводимости. Он преобладает при низких температурах и высоких электрических полях. Ток J_3 представляет собой омическую компоненту проводимости, существенную при малых полях и умеренных температурах. Аналогичные свойства проявляют и пленки Al_2O_3 [51]. В рассматриваемой области значений электрических полей пленки двуокиси кремния имеют гораздо меньшую электропроводность [52]. Например, при $\mathcal{E} = 6 \cdot 10^6 \text{ В} \cdot \text{см}^{-1}$ и комнатной температуре плотность тока в SiO_2 составляет $\sim 4 \cdot 10^{-11} \text{ А} \cdot \text{см}^{-2}$, что на много порядков меньше плотности соответствующих токов в пленках Si_3N_4 и Al_2O_3 .

Другой важной характеристикой диэлектрических слоев является максимальная диэлектрическая прочность. Так называют электрическое поле, при котором происходит пробой изолятора. Основными механизмами, определяющими диэлектрическую прочность изоляторов, являются электрический и тепловой пробои [53]. Они конкурируют друг с другом, и в конкретных ситуациях доминирующим оказывается тот из них, который происходит при меньшей напряженности электрического поля. При низких температурах причиной пробоя диэлектрика обычно являются электрические процессы, и диэлектрическая прочность при этом практически не зависит от температуры. При высоких температурах происходит тепловой пробой, и диэлектрическая прочность уменьшается с ростом температуры. В качестве примера температурной зависимости диэлектрической прочности на рис. 35 приведены результаты измерений в пленках нитрида кремния. Величину пробивного поля \mathcal{E}_m для теплового пробоя можно оценить с помощью выражения

$$\mathcal{E}_m \simeq \left(\frac{\pi \varepsilon_i}{q} \right) (\varphi_B - CT)^2, \quad (51)$$

полученного в работе [50] из условия равенства тепловых потерь и джоулева тепловыделения в структуре. Здесь φ_B — высота энергетического барьера, а C — слабоменяющаяся функция температуры, зависящая от длительности импульса пробивного напряжения. Отметим, что при высоких температурах (рис. 35) значения \mathcal{E}_m уменьшаются с температурой в соответствии с выражением (51), а низкотемпературное значение $\mathcal{E}_m \simeq 10^7 \text{ В} \cdot \text{см}^{-1}$.

Аналогичная температурная зависимость \mathcal{E}_m наблюдается и при пробое термически выращенных пленок SiO_2 . Кроме того, в них наблюдается зависимость пробивного поля от толщины пленки. Из рис. 36 видно, что \mathcal{E}_m увеличивается при уменьшении толщины пленки [54, 55]. Для пленок SiO_2 , толщина которых превышает 1000 \AA , \mathcal{E}_m примерно постоянно и составляет $9 \times 10^6 \text{ В} \cdot \text{см}^{-1}$ (рис. 36, а). Для очень тонких пленок SiO_2 напря-

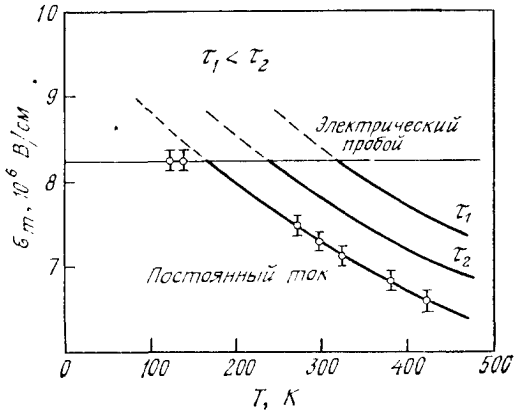


Рис. 35. Температурная зависимость пробивного поля в Si_3N_4 [50], [53].

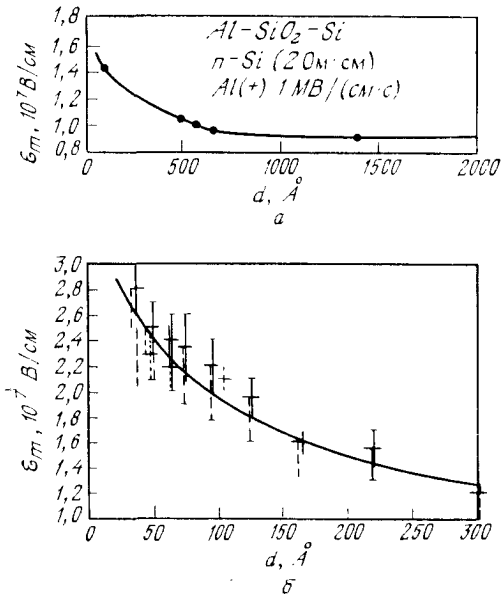


Рис. 36. Зависимость пробивного поля от толщины пленок SiO_2 [54, 55].

а) $d = 100-2000 \text{ \AA}$; б) $d = 40-300 \text{ \AA}$.

женность \mathcal{E}_m достигает величины $3 \cdot 10^7 \text{ В} \cdot \text{см}^{-1}$, которая близка к предельным значениям напряженности электрического поля, при которых происходит разрыв связей $\text{Si} - \text{O}$ (рис. 36, б).

7.4. ПРИБОРЫ С ЗАРЯДОВОЙ СВЯЗЬЮ

Приборы с зарядовой связью (ПЗС) представляют собой матрицы близко расположенных друг к другу МОП-конденсаторов. Соответствующие последовательности тактовых импульсов на затворах такой матрицы смещают ее отдельные МОП-конденсаторы в режим глубокого обеднения, так что зарядовые пакеты могут храниться под электродами матрицы и контролируемым образом перемещаться вдоль поверхности кристалла, перетекая из-под одних электродов матрицы к соседним электродам.

Основными типами приборов с зарядовой связью являются ПЗС с поверхностным каналом и ПЗС со скрытым каналом. В ПЗС с поверхностным каналом заряды хранятся и переносятся у границы раздела полупроводник — диэлектрик. В ПЗС со скрытым каналом благодаря специальному легированию подложки эти процессы происходят в толще полупроводника на некотором удалении от границы с диэлектриком. Отметим также, что при конструировании конкретных микрорелектронных устройств на ПЗС (в зависимости от их назначения) применяются различные схемы организации тактового питания и взаимного расположения затворов.

7.4.1. Хранение заряда

Основным элементом ПЗС (с поверхностным каналом) является МОП-конденсатор, работающий в режиме глубокого обеднения. Его зонная диаграмма для случая, когда сигнальный заряд равен нулю ($Q_{\text{sig}} = 0$), приведена на рис. 37, б. Здесь ψ_{s0} — поверхностный потенциал, $(V_G - V_{FB})$ — эффективное напряжение на затворе (напряжение, приложенное к металлическому электроду, будем называть затворным напряжением V_G), V_{FB} — напряжение плоских зон [56]. Применительно к работе ПЗС минимум электронной потенциальной энергии на границе раздела с окислом называют потенциальной ямой. При $Q_{\text{sig}} = 0$ эта яма пустая. Когда у границы раздела хранится сигнальный зарядовый пакет, поверхностный потенциал уменьшается, что соответствует заполнению потенциальной ямы (рис. 37, б).

Глубину пустой потенциальной ямы легко определить с помощью уравнений (23), (24) и (27а), устанавливающих связь поверхностного потенциала ψ_{s0} и напряжения на затворе:

$$V_G - V_{FB} = V_i + \psi_s = \frac{qN_A W}{C_i} + \psi_s, \quad (52)$$

$$\psi_s = qN_A W^2 / 2\epsilon_s. \quad (53)$$

Отметим, что в рассматриваемой нестационарной ситуации глубина обеднения W может существенно превышать W_m — макси-

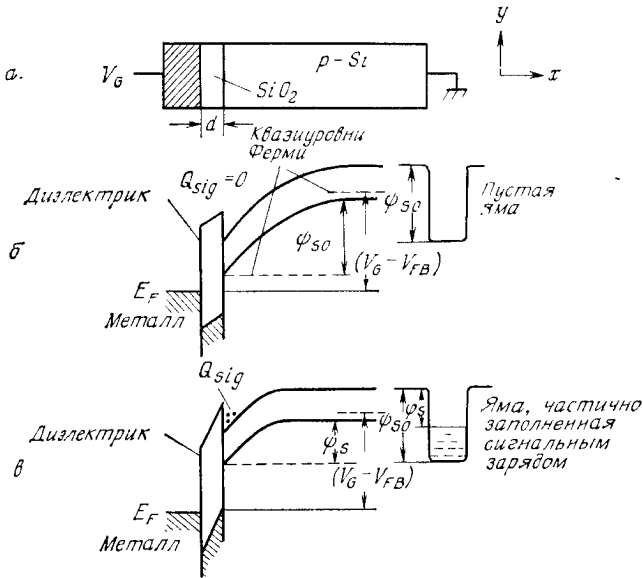


Рис. 37. Зонные диаграммы МОП-структуры с поверхностным каналом [56].
а — изгиб зон в состоянии глубокого обеднения (пустая яма); *б* — изгиб зон на границе раздела Si — SiO₂; *в* — частично заполненная яма.

мальную глубину стационарного обедненного слоя. Исключив из уравнений (52) и (53) W , получим

$$V_G - V_{FB} = \psi_s + \frac{1}{C_t} \sqrt{2\epsilon_s q N_A \psi_s}. \quad (54)$$

Это нелинейное соотношение между поверхностным потенциалом и напряжением на затворе иллюстрируется рис. 38. Отметим, что при одном и том же напряжении на затворе, варьируя d или N_A , можно получать различные значения поверхностного потенциала. Так, например, при $(V_G - V_{FB}) = 10$ В и $N_A = 10^{15}$ см⁻³, поверхностный потенциал уменьшается с 8,5 до 4,2 В, когда толщина окисла возрастает от 0,1 до 0,5 мкм. Этот факт используется при конструировании так называемых двухфазных ПЗС и для предотвращения бокового растекания зарядовых пакетов.

При наличии сигнального зарядового пакета в потенциальной яме ($Q_{sig} \neq 0$) поверхностное электрическое поле в полупроводнике и поле в окисле описываются соответственно выражениями

$$\mathcal{E}_s = (Q_{sig} + qN_A W)/\epsilon_s, \quad (55a)$$

$$\mathcal{E}_t = (Q_{sig} + qN_A W)/\epsilon_t, \quad (55б)$$

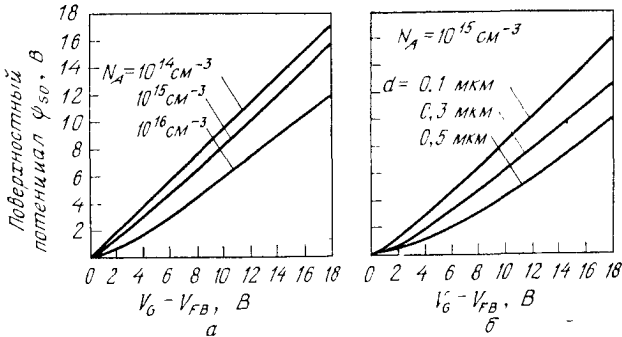


Рис. 38. Зависимость поверхностного потенциала от напряжения на затворе для пленки SiO_2 толщиной 0,1 мкм (а) и для концентрации акцепторов (б) 10^{15} см^{-3} [11].

где Q_{sig} — поверхностная плотность хранящегося сигнального заряда. В этом случае для напряжения на затворе V_G вместо выражения (52) будем иметь

$$V_G - V_{FB} = \frac{Q_{\text{sig}}}{C_i} + \frac{qN_A W}{C_i} + \psi_s. \quad (56)$$

Снова воспользовавшись выражением (53), найдем

$$V_G - V_{FB} = \frac{Q_{\text{sig}}}{C_i} + \frac{\sqrt{2\varepsilon_s q N_A \psi_s}}{C_i} + \psi_s. \quad (57)$$

Решив это уравнение относительно ψ_s , получим окончательно

$$\psi_s = V'_G + V_0 - (2V'_G V_0 + V_0^2)^{1/2}, \quad (58)$$

где

$$V'_G \equiv V_G - V_{FB} - \frac{Q_{\text{sig}}}{C_i}, \quad (58a)$$

$$V_0 \equiv qN_A \varepsilon_s / C_i^2. \quad (58b)$$

Зависимость поверхностного потенциала от величины хранящегося заряда приведена на рис. 39 для трех значений напряжения на затворе [57]. Отметим, что при данных напряжениях на затворе величина ψ_s практически линейно уменьшается с ростом величины сигнального заряда Q_{sig} . Часто эту зависимость интерпретируют с помощью качественных «гидравлических» представлений о потенциальной яме емкостью C_i , «заливаемой» сигнальным зарядом. С увеличением Q_{sig} верхний уровень этой «жидкости», поверхностный потенциал, практически линейно приближается к нулю — верхнему краю ямы (рис. 37, б).

Найдем теперь зависимость объемной концентрации электронов на границе раздела $n_s = n_{p0} \exp(q\psi_s/kT)$ (электрон·см⁻³)

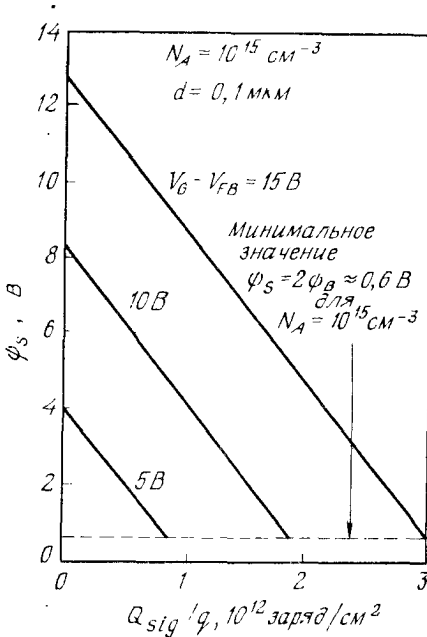


Рис. 39. Зависимость поверхностного потенциала от величины сигнала и напряжения на затворе [57].

от поверхностной плотности сигнала Q_{sig} . Для этого сначала запишем с помощью выражений (53) и (56) выражение для глубины обедненного слоя в присутствии сигнала:

$$W = \frac{\epsilon_s}{C_i} \left(\sqrt{1 + \frac{V_G'}{2V_0} - 1} \right), \quad (59)$$

где V_G' и V_0 — определяются формулами (58а, б). Обратившись теперь к выражению (11), в правой части которого в режиме глубокого обеднения главную роль играют второе и четвертое слагаемые, можно записать

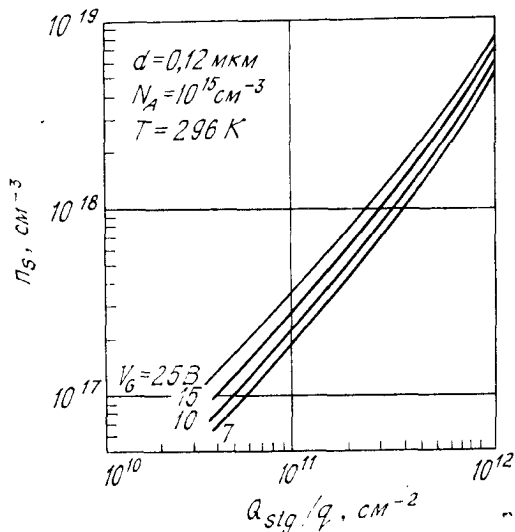
$$\mathcal{E}_s^2 \approx 2 \left(\frac{kT}{qL_D} \right)^2 \left(\frac{q\psi_s}{kT} + \frac{n_{p0}}{N_A} e^{q\psi_s/kT} \right) = \left(\frac{qN_A W}{\epsilon_s} \right)^2 + 2 \left(\frac{kT}{qL_D} \right)^2 \frac{n_s}{N_A}. \quad (60)$$

Комбинируя выражения (55а) и (60), получим окончательно

$$n_s = \frac{W Q_{sig}}{qL_D^2} + \frac{1}{2N_A} \left(\frac{Q_{sig}}{qL_D} \right)^2, \quad (61)$$

где W — определяется соотношением (59). Графики этой зависимости n_s от Q_{sig} , рассчитанные в работе [58], приведены на рис. 40. Видно, что в интервале значений $Q_{sig} = 4 \cdot 10^{10} - 10^{13} \text{ см}^{-2}$ объемная плотность электронов на границе раздела увеличивается пропорционально $(Q_{sig})^{3/2}$.

Рис. 40. Зависимость объемной концентрации электронов на границе с окислом от поверхностной плотности сигнала заряда и напряжения на затворе [58].



7.4.2. Основные ПЗС-структуры

Поперечное сечение типичного трехфазного ПЗС показано на рис. 41 [59]. Это устройство представляет собой полупроводниковую подложку, покрытую однородным слоем изолятора (окисла), на котором достаточно близко друг к другу располо-

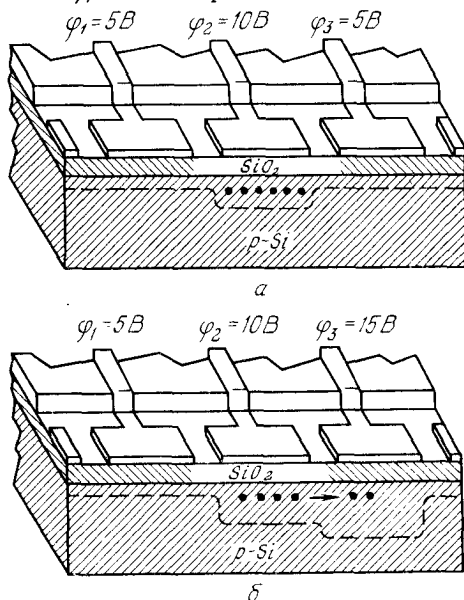


Рис. 41. Поперечное сечение трехфазного ПЗС [59].

a — высокое напряжение на электроде φ_2 ; *б* — высокое напряжение на электроде φ_3 (происходит перенос заряда).

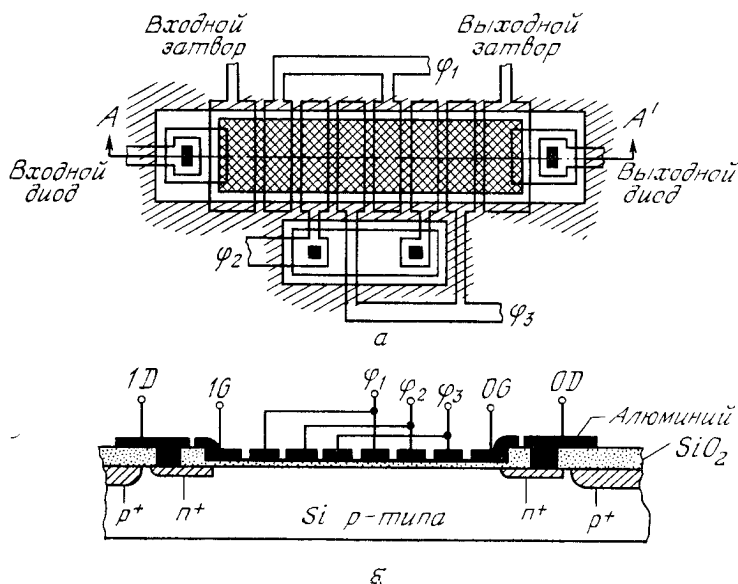


Рис. 42. *n*-Канальный прибор с зарядовой связью [11].

a — вид сверху; *б* — поперечный разрез по *AA'*.

жены затворы — электроды переноса. На рис. 41, *a* представлено состояние хранения информации, когда сигнальный зарядовый пакет находится под средним электродом, напряжение на котором выше, чем на соседних. Процесс переноса сигнала заряда начинается в тот момент, когда на правый затвор подается импульс более высокого напряжения (рис. 41, *б*).

Трехфазный *n*-канальный ПЗС [11] в более подробном виде (вместе с входным и выходным устройствами) показан на рис. 42. Собственно ПЗС, или ПЗС-регистр, здесь составляют три пары электродов переноса (затворов), подсоединенные к шинам тактового питания φ_1 , φ_2 и φ_3 . Входное устройство, состоящее из входного диода и входного затвора, обеспечивает ввод сигнальных зарядовых пакетов под первый электрод переноса регистра. Экстракция и детектирование зарядовых пакетов обеспечиваются выходными затвором и диодом. Временные диаграммы напряжений на шинах тактового питания, а также входной и выходной сигналы приведены на рис. 43, *a* [11]. Расположение потенциальных ям и распределение сигнального заряда в регистре представлены на рис. 43, *б*.

В начальный момент времени $t = t_1$ на тактовую шину φ_1 подано высокое напряжение, а на шины φ_2 и φ_3 — низкое. Следовательно, потенциальные ямы под затворами фазы φ_1 глубже

потенциальных ям под остальными электродами переноса. Входной (ID) и выходной (OD) диоды заперты большим положительным напряжением, что исключает инжекцию электронов в канал под входным (IG) и выходным (OG) затворами, а следовательно, и под электродами переноса ПЗС-регистра. Иными словами, все потенциальные ямы в регистре при $t = t_1$ пустые. В момент t_2 напряжение на входном диоде уменьшается, последний приоткрывается и инжектирует электроны через входной затвор в потенциальную яму под первым затвором фазы φ_1 .

В конце инжекционного процесса поверхностные потенциалы под первым электродом фазы φ_1 и входным затвором равны потенциалу на входном диоде. При $t = t_3$ входной диод снова запирается высоким напряжением, а избыточный заряд из-под первого электрода фазы φ_1 через входной затвор стекает обратно в диод. В результате под первым затвором фазы φ_1 остается вполне определенное количество электронов — сигнальный заряд, величина которого определяется разностью потенциалов фазы φ_1 и входного затвора. В момент t_4 напряжение на фазе φ_2 становится высоким, а на фазе φ_1 оно начинает уменьшаться. При этом сигнальный заряд из-под затвора φ_1 перетекает под первый затвор фазы φ_2 , где поверхностный потенциал превышает потенциал фазы φ_1 . Этот процесс называется переносом. Отметим, что с учетом конечности времени, которое требуется для перетекания заряда из-под одного затвора переноса под другой, задний фронт тактовых импульсов специально делают достаточно пологим. В конце процесса переноса в момент времени t_5 весь сигнальный заряд хра-

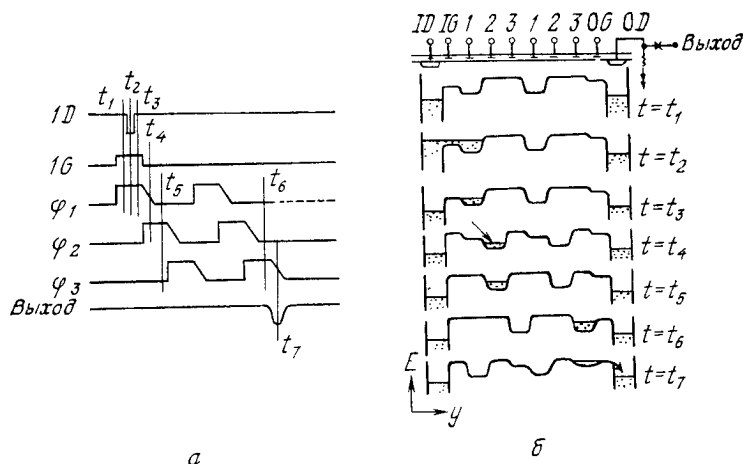


Рис. 43. Диаграммы (а) тактовых импульсов и выходной сигнал в ПЗС и распределения (б) заряда и потенциала вдоль ПЗС в различные моменты времени, отмеченные на тактовых диаграммах [11].

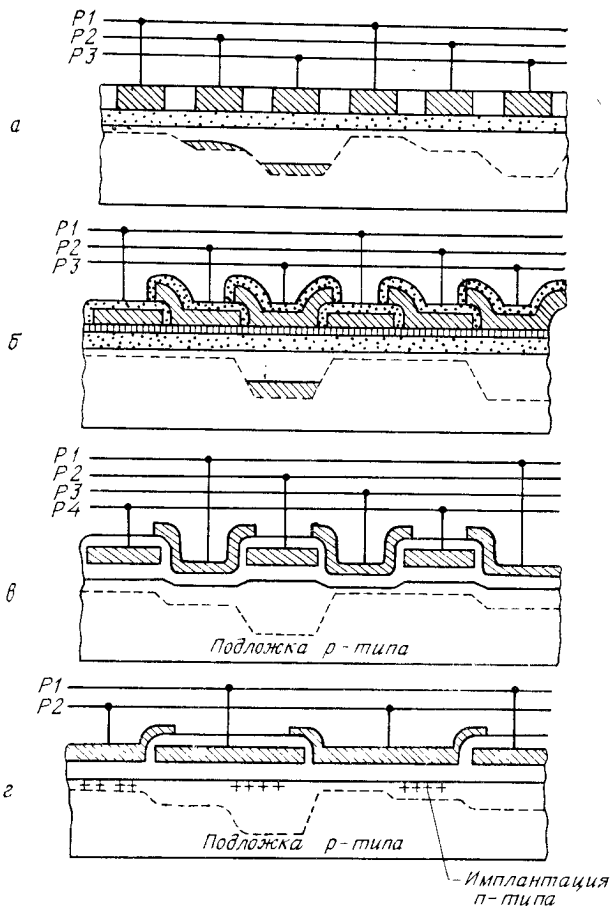


Рис. 44. Электродные структуры и схемы тактового питания ПЗС [60].

а — трехфазная электродная структура, полученная селективным легированием пленки собственного поликремния; *б* — трехфазная электродная структура, в которой использованы три перекрывающихся уровня окисленного кремния; *в* — базовая четырехтактная двухуровневая поликремниевая электродная структура; *г* — компланарная двухтактная двухуровневая поликремниевая электродная структура.

няется уже под первым электродом фазы φ_2 . Через соответствующее число таких циклов переноса (в момент t_6) сигнальный заряд оказывается под последним затвором фазы φ_3 , и, после того как в момент t_7 напряжение на этой фазе начинает уменьшаться, сигнальный заряд через выходной затвор «выталкивается» в выходной диод. При этом выходное устройство вырабатывает ток-овый или потенциальный выходной сигнал, пропорциональный величине зарядового пакета (рис. 43, *а*).

Описанный выше способ ввода сигнального заряда используется в аналоговых и запоминающих устройствах на ПЗС. В системах регистрации оптического изображения зарядовые пакеты формируются в результате генерации электронно-дырочных пар светом, проникающим в полупроводниковую подложку. При этом выходные сигналы оказываются пропорциональными локальной освещенности.

В современной практике конструирования и изготовления ПЗС используются различные электродные структуры и схемы тактового питания. Некоторые из них показаны на рис. 44. Трехфазные электродные структуры (рис. 44, а, б) обладают тем преимуществом [60], что при их изготовлении проблема обеспечения малых (1—2 мкм) межэлектродных зазоров решается значительно легче, чем в случае моноуровневых электродных систем (рис. 42). Одно из достоинств поликремниевых электродных структур (рис. 44, б и в) состоит в том, что эти структуры относительно мало чувствительны к межуровневым закороткам.

Для предотвращения растекания сигнального заряда через боковые границы затворов в направлении, перпендикулярном переносу, на практике используются различные способы осуществления так называемого бокового ограничения канала переноса [10] (рис. 45), слабое подлегирование подложки ионной имплантацией под толстым слоем окисла (рис. 45, б); краевые

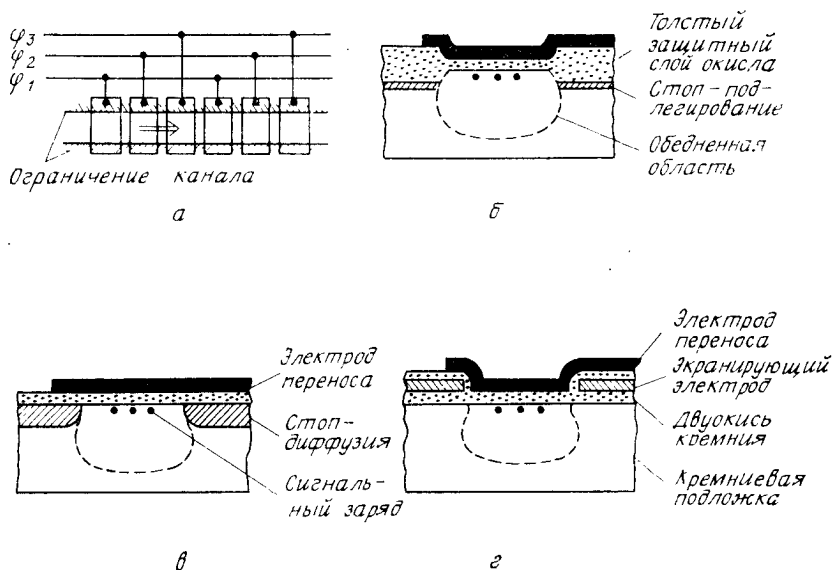


Рис. 45. Методы бокового ограничения канала, используемые в ПЗС [10].

а — ограничение электродов переноса; б — толстый защитный слой окисла; в — высоколегированная диффузионная область; г — поликремниевый экранирующий электрод.

высоколегированные диффузионные области, так называемая стоп-диффузия (рис. 45, *б*); использование дополнительных экранирующих электродов (рис. 45, *г*), смещающих приграничную область полупроводника в режим аккумуляции.

7.4.3. Перенос заряда и частотные свойства

Процесс переноса сигнального заряда в ПЗС определяется тремя основными физическими механизмами: термической диффузией, самоиндуцированным дрейфом и дрейфом в краевых электрических полях. Для относительно малых зарядовых пакетов доминирующим механизмом переноса является термодиффузия. В этом случае полный заряд, оставшийся под электродом, из-под которого осуществляется перенос, к моменту времени t от начала переноса экспоненциально уменьшается со временем. Характерная постоянная времени, определяющая скорость такого диффузионного стекания [11],

$$\tau_{th} = \frac{4L^2}{\pi^2 D_n}, \quad (62)$$

где L — длина электрода переноса, а D_n — коэффициент диффузии неосновных носителей.

Стекание относительно больших (сравнимых с полной глубиной потенциальной ямы) зарядовых пакетов происходит за счет самоиндуцированного дрейфа, обусловленного простым электростатическим расталкиванием инверсионных электронов. Возникающее при этом продольное (направленное параллельно границе с окислом) электрическое поле пропорционально градиенту поверхностной плотности сигнального заряда:

$$\mathcal{E}_{ys} = \frac{q}{C_i} \frac{\partial Q_s(y, t)}{\partial y}. \quad (63)$$

В результате сигнальный заряд уменьшается во времени по гиперболическому закону [10]

$$\frac{Q(t)}{Q} = \frac{t_0}{t + t_0}, \quad (64)$$

где $t_0^* \equiv \pi L^3 W_e C_i / 2 \mu_n Q$, Q — начальная величина полного сигнального заряда, W_e — ширина электрода, а μ_n — подвижность носителей.

Оказывается, что и в отсутствие сигнального заряда под электродом переноса на границе с окислом существует определенное продольное электрическое поле. Это так называемое краевое поле, для вычисления которого необходимо учитывать двумерный характер распределения потенциала в обедненном слое ПЗС-структуры, обусловлено влиянием потенциалов на соседних электродах

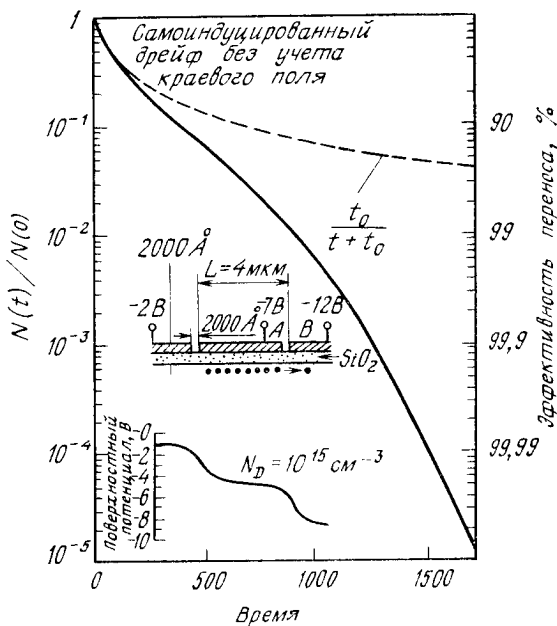


Рис. 46. Теоретическая зависимость нормированного оставшегося заряда от времени с начала переноса в ПЗС с длиной затвора 4 мкм и уровнем легирования 10^{15} см^{-3} . На вставке приведено распределение поверхностного потенциала вдоль направления переноса [61].

(вставка на рис. 46) [61]. Величина краевых полей зависит от толщины слоя окисла, длины электродов, уровня легирования и амплитуды тактовых напряжений. При типичных значениях параметров напряженность продольного краевого поля под центром электрода переноса составляет $\sim 2 \cdot 10^3 \text{ В} \cdot \text{см}^{-1}$. Ясно, что на последних стадиях процесса переноса, когда оставшаяся под электродом часть сигнального заряда становится настолько малой, что самоиндуцированное поле оказывается меньше краевого, именно последнее определяет динамику стекания оставшихся малых порций сигнального заряда. Результаты соответствующих расчетов [61] приведены на рис. 46, где в логарифмическом масштабе построена зависимость заряда, оставшегося под затвором переноса к моменту времени t от начала дрейфа, от времени. Сплошная кривая получена с учетом действия краевого поля, а штриховая соответствует «чистому» самоиндуцированному дрейфу.

Определим теперь эффективность переноса η как отношение полного заряда, успевшего за время, равное длительности тактового импульса T , протечь под затвор B , к величине полного

заряда, находившегося под затвором A в начале процесса переноса (вставка на рис. 46):

$$\eta = 1 - \frac{Q(t=T)}{Q(t=0)}. \quad (65)$$

Соответственно неэффективностью переноса будем называть величину

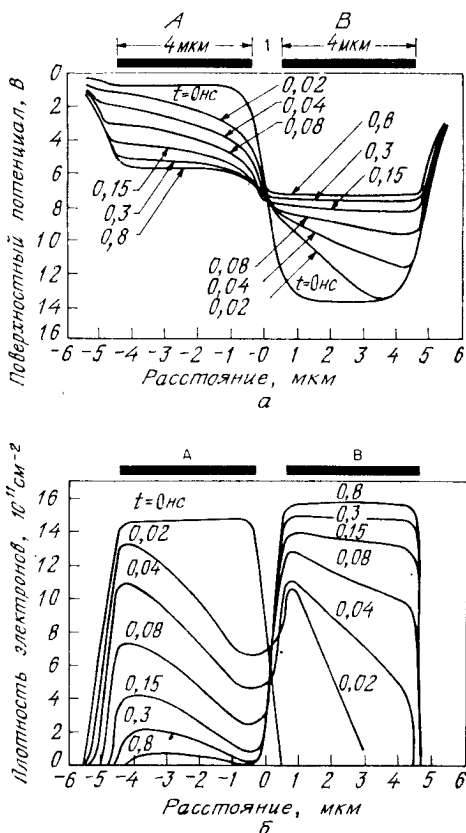
$$\varepsilon \equiv 1 - \eta = Q(t=T)/Q(t=0). \quad (66)$$

Как следует из рис. 46, для того чтобы эффективность переноса превышала 99,99 %, или, что то же самое, чтобы неэффективность ε была меньше 10^{-4} , тактовые частоты не должны превышать нескольких десятков мегабит в 1 с (10^7 Гц). В тех случаях, когда требуются более высокие тактовые частоты, следует использовать структуры с более короткими электродами. При этом увеличиваются краевые поля, а следовательно, и скорость перетекания заряда.

Динамика процесса переноса сигнального заряда в ПЗС исследовалась также численным анализом системы уравнений Пуассона, непрерывности и транспортных соотношений с учетом двумерности распределения потенциала в активной области прибора [62]. Полученные при этом распределения плотности сигнального заряда и поверхностного потенциала в различные моменты времени с начала переноса приведены на рис. 47. Анализируя графики распределения поверхностного потенциала (рис. 47, а), легко видеть, что скорость стекания заряда весьма велика в начальные моменты времени, что обусловлено высокой скоростью дрейфа электронов в большом самоиндуцированном электрическом поле. Отметим, что уже через 0,8 нс после начала переноса поверхностный потенциал под затвором A практически перестает изменяться со временем. Это означает, что большая часть сигнального заряда успела к этому времени вытечь из-под левого электрода. При этом разность поверхностных потенциалов под двумя соседними затворами близка к стационарному значению 1,5 В, что также приводит к существенному замедлению скорости переноса. Из распределений плотности подвижного сигнального заряда (рис. 47, б) видно, что распределение сигнального заряда под электродом A более плавное, чем под электродом B . Это обусловлено тем, что в более глубокой потенциальной яме большими оказываются и краевые поля, «поджимающие» сигнальный заряд к центру ямы. Отметим также, что в соответствии с графиками, приведенными на рис. 47, б, через 0,8 нс после начала переноса 99 % всех сигнальных электронов находятся уже под электродом B .

Рассматривая работу ПЗС, мы до сих пор считали сигнальные электроны свободными и анализировали их движение в соответ-

Рис. 47. Распределение потенциала под затворами в различные моменты времени (а) и соответствующие распределения плотности информационного заряда под затворами в процессе переноса (б) [62].



стующим образом искривленной зоне проводимости, пренебрегая при этом взаимодействием между зоной проводимости и связанными состояниями в запрещенной зоне кристалла, такими, например, как поверхностные ловушки. Рассмотренные выше механизмы переноса называются моделью переноса свободных зарядов. Эта модель справедлива при сравнительно высоких тактовых частотах и дает разумную оценку предельных тактовых частот ПЗС, превышение которых приводит к резкому снижению эффективности переноса. Так, например, для ПЗС с длиной затворов ~ 10 мкм и при типичных значениях остальных параметров для обеспечения эффективности переноса не ниже 99,99 % тактовая частота не должна превышать нескольких десятков мегагерц.

В области промежуточных тактовых частот эффективность переноса сигнального заряда определяется главным образом процессом захвата сигнальных электронов на поверхностные ловушки. Процессы захвата и освобождения электронов с поверх-

ностных состояний, сопровождающие движение сигнального пакета в трехфазном ПЗС, показаны на рис. 48 [63]. Когда сигнальный заряд «входит» в контакт с пустыми поверхностными ловушками, все они практически мгновенно заполняются электронами. После того как этот сигнальный пакет уходит под соседний затвор, захваченные электроны начинают генерироваться с ловушек в зону проводимости. Поскольку энергетический спектр поверхностных ловушек квазинепрерывен, процесс освобождения электронов характеризуется довольно широким спектром сравнительно больших постоянных времени (рис. 20). Поэтому электроны, захваченные на сравнительно мелкие ловушки, освобождаются достаточно быстро и успевают догнать «свой» сигнальный пакет, в то время как электроны с более глубоких ловушек поступают уже в «чужой» пакет, следующий за первым (рис. 48). В результате этого процесса, если на вход ПЗС подать конечную последовательность одинаковых зарядовых пакетов, первые из них подойдут к выходному устройству значительно ослабленными, в то время как за последним еще некоторое число тактов будет тянуться «хвост» отставшего заряда. Неэффективность переноса, обусловленная рассматриваемым взаимодействием сигнального заряда с поверхностными состояниями, описывается выражением

$$\varepsilon \approx \frac{qkTD_{it}}{C_i V_s} \ln(p + 1), \quad (67)$$

где $C_i V_s$ — поверхностная плотность сигнального заряда, D_{it} — плотность поверхностных состояний, а p — число фаз (на рис. 48 $p = 3$). Для уменьшения неэффективности ε следует уменьшать плотность поверхностных состояний.

Влияние поверхностных ловушек на эффективность переноса можно существенно уменьшить, постоянно пропуская через ПЗС-

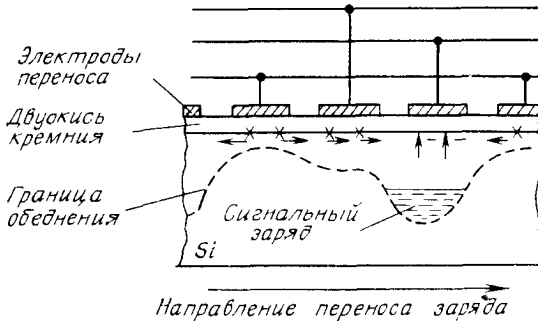


Рис. 48. Захват и освобождение электронов с поверхностных ловушек под различными электродами в процессе переноса сигнального заряда в трехфазном ПЗС [63].

регистр некоторое количество фонового (опорного) заряда — так называемый «непустой нуль». При этом поверхностные ловушки оказываются постоянно заполненными электронами из «непестого нуля», что значительно снижает степень их взаимодействия с сигнальным зарядом. Характерная величина заряда такого «непестого нуля» обычно на практике составляет 10—25 % полной емкости потенциальной ямы. Главный недостаток такого способа увеличения эффективности переноса заключается в соответствующем уменьшении динамического диапазона прибора.

При малых тактовых частотах главным фактором, определяющим работоспособность ПЗС, является темновой ток прибора. Плотность темнового тока можно записать в следующем виде [11]:

$$J_{\tau} = \frac{qn_i W}{2\tau} + \frac{qD_n}{L_n} \frac{n_i^2}{N_A} + \frac{qS_0 n_i}{2}. \quad (68)$$

Здесь первое слагаемое — ток, обусловленный объемной генерацией в обедненном слое; второе слагаемое — диффузионный ток генерации неосновных носителей в электронейтральной части подложки, поступающий через область обеднения в инверсионный слой; последнее слагаемое — ток поверхностной генерации; τ — время жизни неосновных носителей; D_n — коэффициент диффузии; L_n — диффузионная длина; S_0 — скорость поверхностной рекомбинации.

Процесс релаксации состояния глубокого обеднения МОП-структуры, обусловленный темновым током [64], показан на рис. 49. В начальный момент времени передний фронт импульса напряжения на затворе V_G быстро переводит МОП-структуру

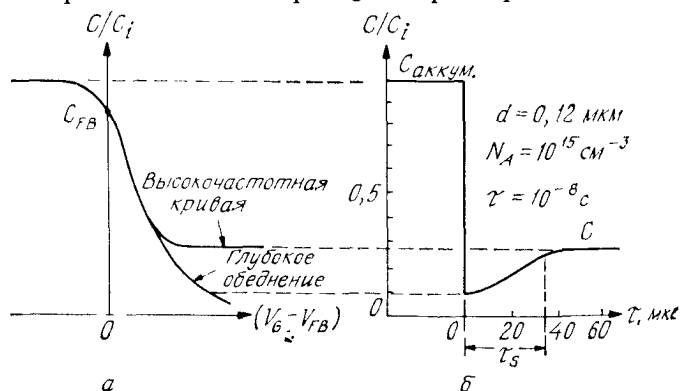


Рис. 49. Зависимость емкости МДП-структуры от напряжения и времени при импульсном смещении в режим глубокого обеднения [64].

а — нормированная емкость в зависимости от напряжения на затворе; *б* — нормированная емкость в зависимости от времени.

туру из режима аккумуляции в режим глубокого обеднения (рис. 49, а). С течением времени ток термогенерации постепенно заполняет потенциальную яму неосновными носителями, поверхностный потенциал уменьшается, а емкость структуры растет, асимптотически приближаясь к величине, соответствующей стационарному режиму инверсии (рис. 49, б). Характерная длительность этого процесса τ_s зависит от скорости термогенерации неосновных носителей, т. е. от величины темнового тока. Для Si — SiO₂-систем типичные значения τ_s составляют 10⁻³—10⁻² с. Ясно, что для работоспособности ПЗС тактовая частота f должна быть достаточно велика, чтобы дополнительный заряд, поступающий в сигнальный пакет при его прохождении через ПЗС-регистр за счет темнового тока

$$Q_T = AJ_T N / pf, \quad (69)$$

не превышал соответствующей доли максимального сигнального заряда

$$Q_{\text{макс}} = AC_i \Delta\psi_s. \quad (70)$$

В этих выражениях A — площадь затвора, N — полное число электродов ПЗС-регистра, p — число фаз (в примере на рис. 42 $N = 6$, $p = 3$), $\Delta\psi_s$ — максимально допустимый перепад поверхностного потенциала в канале. Следовательно,

$$Q_T / Q_{\text{макс}} = J_T N / (pf C_i \Delta\psi_s). \quad (71)$$

Если это отношение равно 10⁻³, то для 128-затворного трехфазного ПЗС ($N = 128$) при толщине слоя затворного окисла 0,1 мкм $J_T = 10^{-9}$ А·см⁻² и $\Delta\psi_s = 5$ В. Низкочастотный предел тактовой частоты, обусловленный темновым током, составляет ~1 кГц.

В цифровых устройствах на ПЗС сигнальный заряд определенной величины соответствует 1, а пустая яма 0. На рис. 50 приведена частотная зависимость нормированной допустимой (наи-

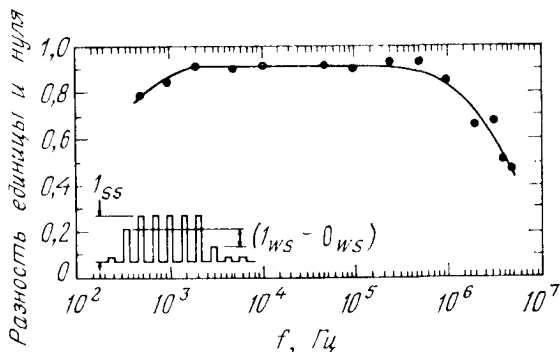


Рис. 50. Частотная характеристика ПЗС с поверхностным каналом [65].

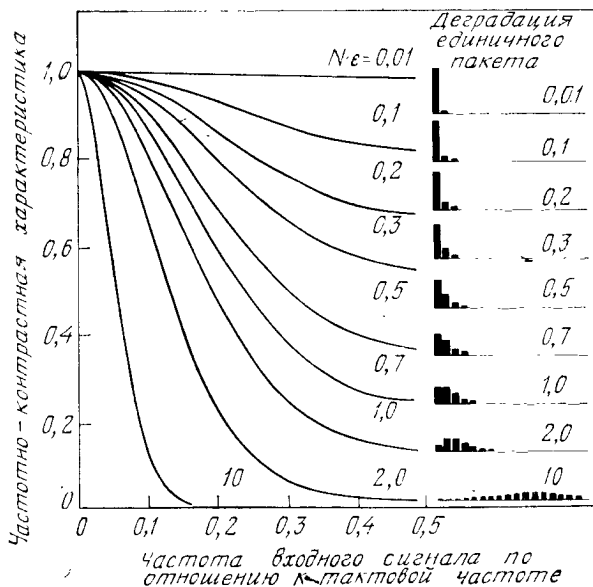


Рис. 51. Частотно-контрастные характеристики и пространственное разрешение в ПЗС для различных значений неэффективности $N\epsilon$ [60].

худшей) разности между 1 и 0, которая равна $(I_{wc} - 0_{wc})/I_{ss}$. Поскольку эта разность непосредственно связана с эффективностью переноса заряда, приведенные данные означают, что величина эффективности переноса характеризует работоспособность ПЗС в широком диапазоне тактовых частот.

Низкочастотный завал частотной характеристики обусловлен встраиванием дополнительного (за счет темнового тока) заряда в сигнальные пакеты, что, естественно, искажает амплитуду передаваемых сигналов. Ее высокочастотный спад обусловлен резким уменьшением эффективности переноса, когда длительность тактового импульса оказывается недостаточной для полного перетекания сигнального заряда из-под одного затвора под другой.

Для улучшения низкочастотных свойств ПЗС следует уменьшать все компоненты темнового тока, увеличивая время жизни, диффузионную длину и снижая скорость поверхностной рекомбинации. Для расширения рабочего диапазона ПЗС в сторону высоких частот можно уменьшать длину затвора L , использовать n -канальные структуры (поскольку подвижность электронов выше, чем у дырок) и уменьшать межэлектродный зазор. Перспективным для создания сверхвысокоскоростных ПЗС является использование GaAs из-за высокой подвижности электронов в этом материале. Уже имеются ПЗС со скрытым каналом на GaAs, работающие с тактовыми частотами до 500 МГц [71].

Неэффективность переноса может также приводить к большим фазовым задержкам. Частотно-контрастные характеристики и деградация единичного зарядового пакета [60] в зависимости от произведения $N\epsilon$ приведены на рис. 51, где N — полное число переносов, которое требуется, чтобы сигнальный пакет достиг выходного узла¹⁾. Здесь ясно видны характер расплывания индивидуального зарядового пакета и образование соответствующих «хвостов» при различных значениях $N\epsilon$. Крайнее левое положение сигнального столбика в каждом «кадре» (рис. 51) соответствует временному положению индивидуального зарядового пакета, которое он занимал бы на выходе идеального ПЗС ($\epsilon = 0$). Легко видеть, что при $N\epsilon \geq 1$ значительная часть сигнального заряда оказывается сдвинутой в хвостовые пакеты. Это препятствует адекватной обработке информации в таком приборе.

7.4.4. ПЗС со скрытым каналом

В ПЗС с поверхностным каналом неосновные носители сигнальных пакетов перемещаются под действием тактовых импульсов напряжения на затворах непосредственно у границы с окислом. При этом они сильно взаимодействуют с поверхностными ловушками, что является одним из главных факторов, ограничивающих эффективность переноса сигнального заряда. Чтобы исключить захват информационного заряда на эти поверхностные ловушки и тем самым повысить эффективность переноса, была предложена конструкция ПЗС со скрытым каналом [66]. В этом приборе движение зарядовых пакетов ограничено в пределах объемного канала, расположенного под границей с окислом. Поперечное сечение такого ПЗС показано на рис. 52 [67]. Этот прибор состоит из кремниевой подложки p -типа с приповерхностным слоем проводимости n -типа и контактов n^+ -типа на обоих концах n -канала.

¹⁾ Величина N , следовательно, равна полному числу электродов переноса. — Прим. перев.

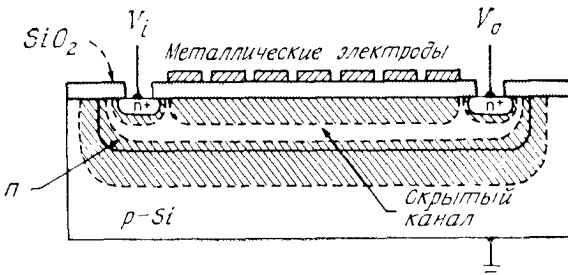


Рис. 52. Поперечное сечение ПЗС со скрытым каналом [67].

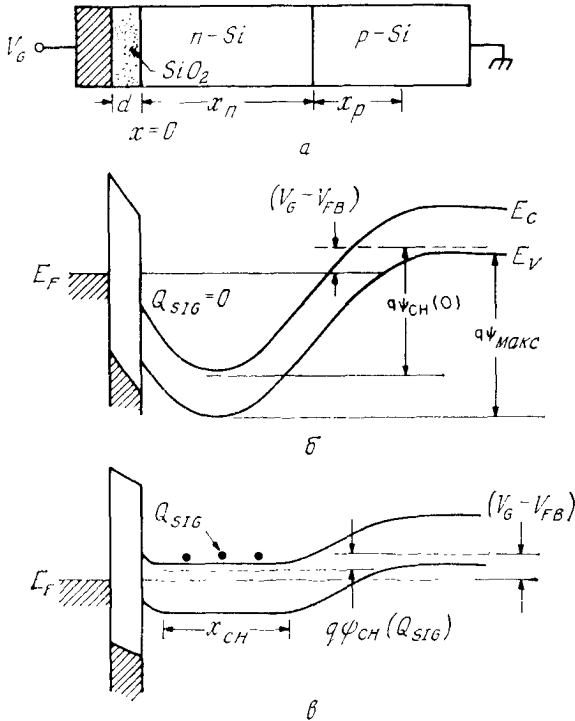


Рис. 53. Схема ПЗС со скрытым каналом (а), его зонная диаграмма для пустой ямы (б) и зонная диаграмма в присутствии сигнального заряда (в) [68].

Зонные диаграммы ПЗС со скрытым каналом показаны на рис. 53 [68]. В отсутствие сигнального заряда сравнительно узкий n -слой прибора полностью обеднен тактовыми импульсами положительного напряжения на затворах структуры (рис. 53, б). Когда сигнальный заряд вводится в скрытый канал, он будет храниться в его центральной части (рис. 53, в).

Распределение потенциала для случая, показанного на рис. 53, б, можно рассчитать аналитически, когда концентрации примеси в n - и p -областях постоянны, воспользовавшись приближением полного обеднения. В этом случае его определяют с помощью уравнений Пуассона [11]:

$$\frac{d^2\psi}{dx^2} = 0, \quad -d < x < 0, \quad (72a)$$

$$\frac{d^2\psi}{dx^2} = -qN_D/\epsilon_s, \quad 0 < x < x_n, \quad (72б)$$

$$\frac{d^2\psi}{dx^2} = qN_A/\epsilon_s, \quad x_n < x < x_n + x_p \quad (72в)$$

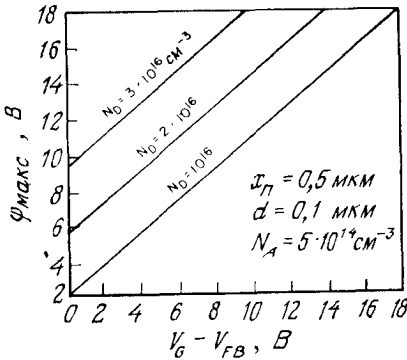


Рис. 54. Зависимости максимального потенциала в канале $\psi_{\text{макс}}$ от напряжения на затворе в ПЗС со скрытым каналом [11].

при следующих граничных условиях: 1) $\psi = (V_G - V_{FB})$ при $x = -d$; 2) $\psi = 0$ при $x = x_n + x_p$; 3) непрерывность потенциала и электрической индукции при $x = 0$ и $x = x_n$. Решение системы уравнений (72) с указанными граничными условиями дает распределение потенциала по толщине прибора с максимумом (рис. 53, б), величина которого

$$\psi_{\text{макс}} = \psi_J \left(1 + \frac{N_A}{N_D} \right), \quad (73)$$

где

$$V_G - V_{FB} + V_1 = \psi_J + \sqrt{V_{0x} \psi_J}, \quad (74)$$

$$\psi_J \equiv \frac{qN_A x_p^2}{2\epsilon_s}; \quad V_{0x} \equiv \frac{2qN_A x_p^2}{\epsilon_s} \left(1 + \frac{\epsilon_s}{\epsilon_i} \frac{d}{x_n} \right)^2,$$

$$V_1 \equiv \frac{qN_D x_n^2}{2\epsilon_s} \left(1 + \frac{2\epsilon_s}{\epsilon_i} \frac{d}{x_n} \right).$$

Соответствующие этим формулам зависимости $\psi_{\text{макс}}$ от напряжения на затворе V_G , рассчитанные при типичных для ПЗС со скрытым каналом значениях параметров, приведены на рис. 54.

Чтобы получить распределение потенциала в скрытом канале при наличии сигнального заряда (рис. 53, в), нужно решить более сложные уравнения, отличающиеся от системы уравнений (72) тем, что вместо N_D надо брать $N_D - n(x)$, а вместо N_A надо брать $N_A + n(x) - p(x)$, где $n(x)$ и $p(x)$ — концентрации свободных электронов и дырок соответственно. Результаты численных расчетов [69] для сигнальных пакетов различной величины приведены на рис. 55, где плотность сигнального заряда (на единицу площади границы раздела) нормирована на величину $N_A L_D$ (L_D — дебаевская длина). Например, при $N_A = 10^{14} \text{ см}^{-3}$ дебаевская длина равна $0,415 \text{ мкм}$, а $N_A L_D = 4,15 \cdot 10^9 \text{ заряд/см}^2$. Графики, приведенные на рис. 55, б, показывают, что с увеличением сигналь-

Рис. 55. Распределение потенциала (а) в скрытом канале ПЗС с однородно легированными n -слоем и p -областью ($V_G = 4$ В) и соответствующее распределение объемной плотности сигнала (б) [69].

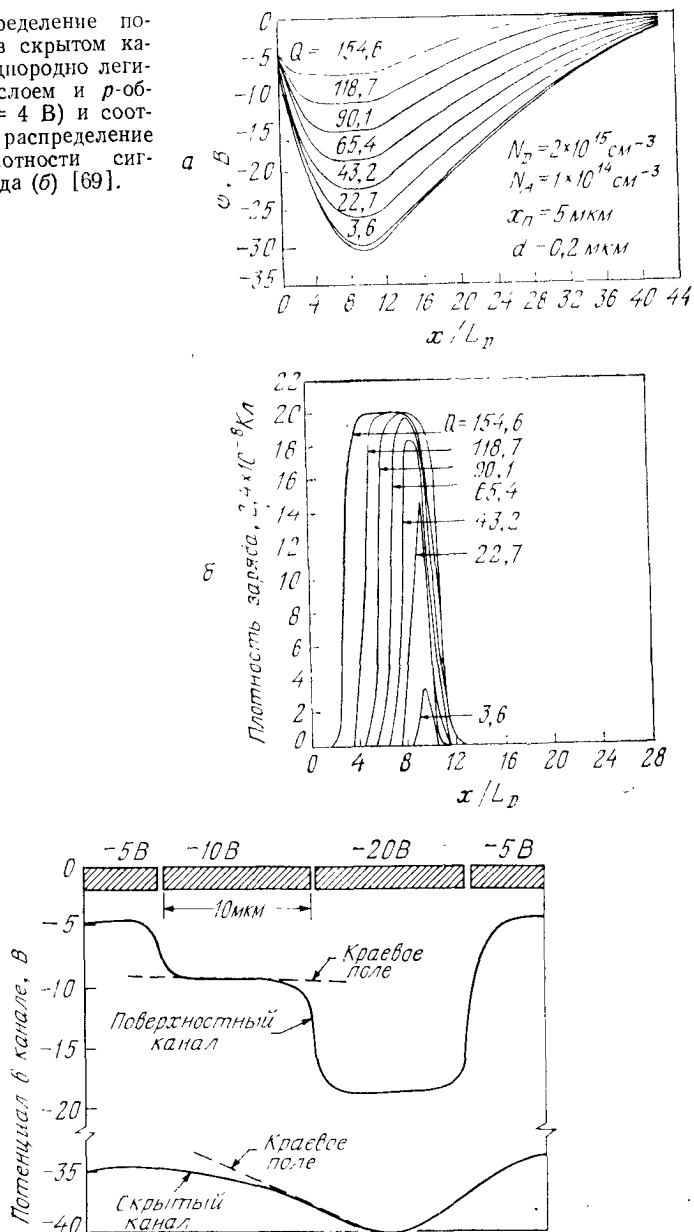


Рис. 56. Результаты двумерного расчета продольного распределения потенциала в центре скрытого канала ПЗС. Для сравнения приведен соответствующий профиль потенциала в аналогичном ПЗС с поверхностным каналом [67].

ного заряда его пространственное распределение становится более широким и сдвигается в сторону границы раздела с окислом.

Результаты двумерного численного расчета изменения потенциала вдоль скрытого канала приведены на рис. 5б. Для сравнения здесь также показана соответствующая зависимость в аналогичном приборе с поверхностным каналом. Ясно, что больший градиент потенциала в ПЗС со скрытым каналом под центром затвора (краевое поле) будет способствовать более высокой скорости переноса заряда. Действительно, как оказалось, значения неэффективности переноса $\epsilon \sim 10^{-4}$ — 10^{-5} сравнительно легко достигаются в ПЗС со скрытым каналом, что на порядок меньше, чем в типичных конструкциях ПЗС с поверхностным каналом тех же геометрических размеров.

ЛИТЕРАТУРА

1. Moll J. L. Variable Capacitance with Large Capacity Change, *Wescon Conv. Rec.*, Pt. 3, p. 32 (1959).
2. Pfann W. G., Garrett C. G. Semiconductor Varactor Using Space-Charge Layers, *Proc. IRE*, 47, 2011 (1959).
3. Frankl D. R. Some Effects of Material Parameters on the Design of Surface Space-Charge Varactor, *Solid State Electron.*, 2, 71 (1961).
4. Lindner R. Semiconductor Surface Varactor, *Bell Syst. Tech. J.*, 41, 803 (1962).
5. Termann L. M. An Investigation of Surface States at a Silicon/Silicon Dioxide Interface Employing Metal — Oxide — Silicon Diodes, *Solid State Electron.*, 5, 285 (1962).
6. Lehovec K., Slobodsky A. Field-Effect Capacitance Analysis of Surface States on Silicon, *Phys. Status Solidi*, 3, 447 (1963).
7. Nicollian E. H., Brews J. R. *MOS Physics and Technology*, Wiley, N. Y., 1982.
8. Boyle W. S., Smith G. E. Charge Coupled Semiconductor Devices, *Bell Syst. Tech. J.*, 49, 587 (1970).
9. Amelio G. F., Tompsett M. F., Smith G. E. Experimental Verification of the Charge Coupled Diode Concept, *Bell Syst. Tech. J.*, 49, 593 (1970).
10. Sequin C. H., Tompsett M. F. *Charge Transfer Devices*, Academic, N. Y., 1975.
11. Kim C. K. The Physics of Charge-Coupled Devices, in Howes M. J., Morgan D. V., Eds., *Charge-Coupled Devices and Systems*, Wiley, N. Y., 1979, p. 1.
12. Melen R., Buss D., Eds. *Charge-Coupled Devices: Technology and Applications*, IEEE Press, N. Y., 1977.
13. Garrett C. G. B., Brattain W. H. Physical Theory of Semiconductor Surfaces, *Phys. Rev.*, 99, 376 (1955).
14. Kingston R. H., Neustadter S. F. Calculation of the Space Charge, Electric Field, and Free Carrier Concentration at the Surface of a Semiconductor, *J. Appl. Phys.*, 26, 718 (1955).
15. Hofstein S. R., Warfield G. Physical Limitation on the Frequency Response of a Semiconductor Surface Inversion Layer, *Solid State Electron.*, 8, 321 (1965).
16. Grove A. S., Deal B. E., Snow E. H., Sah C. T. Investigation of Thermally Oxidized Silicon Surfaces Using Metal — Oxide — Semiconductor Structures, *Solid State Electron.*, 8, 145 (1965).

17. Grove A. S., Snow E. H., Deal B. E., Sah C. T. Simple Physical Model for the Space-Charge Capacitance of Metal — Oxide — Semiconductor Structures, *J. Appl. Phys.*, **33**, 2458 (1964).
18. Goetzberger A. Ideal MOS Curves for Silicon, *Bell Syst. Tech. J.*, **45**, 1097 (1966).
19. Deal B. E. Standardized Terminology for Oxide Charges Associated with Thermally Oxidized Silicon, *IEEE Trans. Electron Devices*, **ED-27**, 606 (1980).
20. Tamm I. Über eine mögliche Art der Elektronenbindung an Kristalloberflächen, *Phys. Z. Sowjetunion*, **1**, 733 (1933).
21. Shockley W. On the Surface States Associated with a Periodic Potential, *Phys. Rev.*, **56**, 317 (1939).
22. Shockley W., Pearson G. L. Modulation of Conductance of Thin Films of Semiconductors by Surface Charges, *Phys. Rev.*, **74**, 232 (1948).
23. Allen F. G., Gobeli G. W. Work Function, Photoelectric Threshold and Surface States of Atomically Clean Silicon, *Phys. Rev.*, **127**, 150 (1962).
24. Nicollian E. H., Goetzberger A. MOS Conductance Technique for Measuring Surface State Parameters, *Appl. Phys. Lett.*, **7**, 216 (1965).
25. Berglund C. N. Surface States at Steam-Grown Silicon — Silicon Dioxide Interface, *IEEE Trans. Electron Devices*, **ED-13**, 701 (1966).
26. Nicollian E. H., Goetzberger A. The Si — SiO₂ Interface — Electrical Properties as Determined by the MIS Conductance Technique, *Bell Syst. Tech. J.*, **46**, 1055 (1967).
27. White W. H., Cricchi J. R. Characterization of Thin-Oxide MNOS Memory Transistors, *IEEE Trans. Electron Devices*, **ED-19**, 1280 (1972).
28. Deal B. E., Sklar M., Grove A. S., Snow E. H. Characteristics of the Surface-State Charge (Q_{ss}) of Thermally Oxidized Silicon, *J. Electrochem. Soc.*, **114**, 266 (1967).
29. Ligenza J. R. Effect of Crystal Orientation of Oxidation Rates of Silicon in High Pressure Steam, *J. Phys. Chem.*, **65**, 2011 (1961).
30. Snow E. H., Grove A. S., Deal B. E., Sah C. T. Ion Transport Phenomena in Insulating Films, *J. Appl. Phys.*, **36**, 1664 (1965).
31. Yon E., Ko W. H., Kuper A. B. Sodium Distribution in Thermal Oxide on Silicon by Radiochemical and MOS Analysis, *IEEE Trans. Electron Devices*, **ED-13**, 276 (1966).
32. Dalton J. V., Drobek J. Structure and Sodium Migration in Silicon Nitride Films, *J. Electrochem. Soc.*, **115**, 865 (1968).
33. Cheney G. T., Jacobs R. M., Korb H. W., Nigh H. E., Stack J. Al₂O₃ — SiO₂ IGFET Integrated Circuits, Paper No. 2.2, IEEE Device Meet., Washington, D. C., Oct. 18—21, 1967.
34. Snow E. H., Deal B. E. Polarization Phenomena and Other Properties of Phosphosilicate Glass Films on Silicon, *J. Electrochem. Soc.*, **113**, 2631 (1966).
35. Chu T. L., Szedon J. R., Lee C. H. The Preparation and $C - V$ Characteristics of Si — Si₃N₄ and Si — SiO₂ — Si₃N₄ Structure, *Solid State Electron.*, **10**, 897 (1967).
36. Reizman F., Van Gelder W. Optical Thickness Measurement of SiO₂ — Si₃N₄ Films on Silicon, *Solid State Electron.*, **10**, 625 (1967).
37. Williams R. Photoemission of Electrons from Silicon into Silicon Dioxide, *Phys. Rev.*, **140**, A569 (1965).
38. Deal B. E., Snow E. H., Mead C. A. Barrier Energies in Metal — Silicon Dioxide — Silicon Structures, *J. Phys. Chem. Solids*, **27**, 1873 (1966).
39. Werner W. M. The Work Function Difference of the MOS-System with Aluminium Field Plates and Polycrystalline Silicon Field Plates, *Solid State Electron.*, **17**, 769 (1974).
40. Goetzberger A. Behavior of MOS Inversion Layers at Low Temperature, *IEEE Trans. Electron Devices*, **ED-14**, 787 (1967).
41. Goetzberger A., Nicollian E. H. Temperature Dependence of Inversion Layer Frequency Response in Silicon, *Bell Syst. Tech. J.*, **46**, 513 (1967).

42. Grosvalet J., Jund C. Influence of Illumination on MIS Capacitance in the Strong Inversion Region, *IEEE Trans. Electron Devices*, **ED-14**, 777 (1967).
43. Collins D. R., Sah C. T. Effects of X-Ray Irradiation on the Characteristics of MOS Structures, *Appl. Phys. Lett.*, **8**, 124 (1966).
44. Snow E. H., Grove A. S., Fitzgeralds D. J. Effect of Ionization Radiation on Oxidized Silicon Surfaces and Planar Devices, *Proc. IEEE*, **55**, 1168 (1967).
45. Nicollian E. H., Goetzberger A., Berglund C. N. Avalanche Injection Currents and Charging Phenomena in Thermal SiO₂, *Appl. Phys. Lett.*, **15**, 174 (1969).
46. Rusu A., Bulucea C. Deep-Depletion Breakdown Voltage of SiO₂/Si MOS Capacitors, *IEEE Trans. Electron Devices*, **ED-26**, 201 (1979).
47. Sze S. M., Gibbons G. Effects of Junction Curvature on Breakdown Voltage in Semiconductors, *Solid State Electron.*, **9**, 831 (1966).
48. Frenkel J. On the Theory of Electric Breakdown of Dielectrics and Electronic Semiconductors, *Tech. Phys. USSR*, **5**, 685 (1938); On Pre-Breakdown Phenomena in Insulators and Electronic Semiconductors, *Phys. Rev.*, **54**, 647 (1938).
49. O'Dwyer J. J. The Theory of Electrical Conduction and Breakdown in Solid Dielectrics, Clarendon, Oxford, 1973.
50. Sze S. M. Current Transport and Maximum Dielectric Strength of Silicon Nitride Films, *J. Appl. Phys.*, **38**, 2951 (1967).
51. Johnson W. C. Study of Electronic Transport and Breakdown in Thin Insulating Films, Tech. Rep. No. 7, Princeton University, 1979.
52. Av-Ron M., Shatzkes M., DiStefano T. H., Cadoff I. B. The Nature of Electron Tunneling in SiO₂, in Pantelider S. T., Ed. The Physics of SiO₂ and Its Interfaces, Pergamon, N. Y., 1978, p. 46.
53. Klein N. Electrical Breakdown in Solids, *Advances in Electronics and Electron Physics*, Vol. 26, Academic, N. Y., 1969.
54. Osburn C. M., Ormond D. W. Dielectric Breakdown in Silicon Dioxide Films on Silicon, *J. Electrochem. Soc., Solid State Sci. Technol.*, **119**, 591 (1972).
55. Harari E. Dielectric Breakdown in Electrically Stressed Thin Films of Thermal SiO₂, *J. Appl. Phys.*, **49**, 2478 (1978).
56. Barbe D. F. Imaging Devices Using the Charge-Coupled Concept, *Proc. IEEE*, **63**, 38 (1975).
57. Beynon J. D. E. The Basic Principles of Charge-Coupled Devices, *Microelectronics*, **7**, 7 (1975).
58. Ong D. G., Pierret R. F. Approximate Formula for Surface Carrier Concentration in Charge-Coupled Devices, *Electron. Lett.*, **10**, 6 (1974).
59. Boyle W. S., Smith G. E. Charge-Coupled Devices — A New Approach to MIS Device Structures, *IEEE Spectrum*, **8**, 18 (1971).
60. Tompsett M. F. Video-Signal Generation, in McLean T. P., Schagen P., Eds., *Electronic Imaging*, Academic, N. Y., 1979, p. 55.
61. Carnes J. E., Kosonocky W. F., Ramberg E. G. Free Charge Transfer in Charge-Coupled Devices, *IEEE Trans. Electron Devices*, **ED-19**, 789 (1972).
62. Elsaid M. H., Chamberlain S. G., Watt L. A. K. Computer Model and Charge Transport Studies in Short Gate Charge-Coupled Devices, *Solid State Electron.*, **20**, 61 (1977).
63. Tompsett M. F. The Quantitative Effect of Interface States on the Performance of Charge-Coupled Devices, *IEEE Trans. Electron. Devices*, **ED-20**, 45 (1973).
64. Tao T. F., Ellis J. R., Kost L., Doshier A. Feasibility Study of PbTe and PbSnTe Infrared Charge Coupled Imager, Proc. Appl. Conf. CCD, Naval Electron. Lab. Center, San Diego, p. 259 (1973).
65. Augusta B., Harroun T. Conceptual Desing of an Eight Megabyte High Performance Charge-Coupled Storage Device, Proc. Appl. Conf. CCD, Naval Electron. Lab. Center, San Diego, p. 55 (1973).

66. Boyle W. S., Smith G. E. U. S. Patent 3792322 (1974).
67. Walden R. H., Krambeck R. H., Strain R. J., McKenna J., Schryer N. L., Smith G. E. The Buried Channel Charge Coupled Device, *Bell Syst. Tech. J.*, **51**, 1635 (1972).
68. Burt D. J. Basic Operation of the Charge Coupled Device, Int. Conf. Technol. Appl. CCD, University of Edinburgh, p. 1 (1974).
69. Kent W. H. Charge Distribution in Buried-Channel Charge-Coupled Devices, *Bell Syst. Tech. J.*, **52**, 1009 (1973).
70. Brews J. R. A Simplified High-Frequency MOS Capacitance Formula, *Solid State Electron.*, **20**, 607 (1977).
71. Deyhimi I., Eden R. C., Anderson R. J., Harris I. S., Jr. A 500-MHz GaAs Charge-Coupled Device, *Appl. Phys. Lett.*, **36**, 151 (1980).

Оглавление

ПРЕДИСЛОВИЕ РЕДАКТОРА ПЕРЕВОДА	5
ПРЕДИСЛОВИЕ АВТОРА	6
ВВЕДЕНИЕ	8
ЧАСТЬ I. ФИЗИКА ПОЛУПРОВОДНИКОВ	11
Глава 1. Физика и свойства полупроводников	11
1.1. Введение	11
1.2. Кристаллическая структура	11
1.3. Энергетические зоны	16
1.4. Концентрация носителей при термодинамическом равновесии	20
1.5. Явления переноса	33
1.6. Фононные спектры. Оптические и тепловые свойства полупроводников. Поведение полупроводников при сильных электрических полях	45
1.7. Основные уравнения для анализа работы полупроводниковых приборов	58
Литература	65
ЧАСТЬ II. БИПОЛЯРНЫЕ ПРИБОРЫ	69
Глава 2. Плоскостные диоды	69
2.1. Введение	69
2.2. Основы технологии	69
2.3. обедненный слой и барьерная емкость	80
2.4. Вольт-амперные характеристики	91
2.5. Пробой $p-n$ -перехода	103
2.6. Переходные процессы и шумы	117
2.7. Схемные функции	121
2.8. Гетеропереходы	132
Литература	139
Глава 3. Биполярные транзисторы	142
3.1. Введение	142
3.2. Статические характеристики	143
3.3. СВЧ-транзисторы	166
3.4. Мощные транзисторы	180
3.5. Переключающие транзисторы	186
3.6. Разновидности биполярных транзисторов	192
Литература	198

Глава 4. Тиристоры	202
4.1. Введение	202
4.2. Основные характеристики	202
4.3. Диодный и триодный тиристоры	221
4.4. Мощные тиристоры	235
4.5. Диак и триак	243
4.6. Однопереходные транзисторы и переключающие тиристоры	247
4.7. Полевые тиристоры	252
Литература	254
ЧАСТЬ III. УНИПОЛЯРНЫЕ ПРИБОРЫ	257
Глава 5. Контакты металл — полупроводник	257
5.1. Введение	257
5.2. Зонные диаграммы	258
5.3. Эффект Шоттки	262
5.4. Теория процессов переноса заряда	266
5.5. Высота барьера	283
5.6. Структуры приборов	310
5.7. Омический контакт	318
Литература	321
Глава 6. Полевые транзисторы с $p-n$-переходом в качестве затвора и полевые транзисторы типа металл — полупроводник	325
6.1. Введение	325
6.2. Основные характеристики приборов	327
6.3. Особенности характеристик реальных приборов	338
6.4. Высокочастотные характеристики	355
6.5. Другие полевые приборы	366
Литература	374
Глава 7. МДП-структуры. Приборы с зарядовой связью	377
7.1. Введение	377
7.2. Идеальная МДП-структура	378
7.3. $Si-SiO_2$ — МОП-структуры	394
7.4. Приборы с зарядовой связью	429
Литература	450

УВАЖАЕМЫЙ ЧИТАТЕЛЬ!

Ваши замечания о содержании книги, ее оформлении, качестве перевода и другие просим присылать по адресу: 129820, Москва, И-110, ГСП, 1-й Рижский пер., д. 2, издательство «Мир».

С. Зи

ФИЗИКА ПОЛУПРОВОДНИКОВЫХ ПРИБОРОВ

В 2-х томах

Книга 1

Старший научный редактор И. М. Андреева
Младший научный редактор Л. С. Сысоева
Художник Д. А. Анникеев
Художественный редактор В. Б. Прищепа
Технический редактор В. П. Сизова
Корректор В. С. Соколов

ИБ № 3809

Сдано в набор 10.04.84.
Подписано к печати 13.09.84.
Формат 60×90¹/₁₆.
Бумага типографская № 1.
Гарнитура литературная. Печать высокая.
Объем 14,25 бум. л. Усл. печ. л. 28,50. Усл. кр.-отт. 28,50.
Уч.-изд. л. 26,13. Изд. № 20/2918
Тираж 16 000 экз. Зак. 106. Цена 2 р. 20 к.

ИЗДАТЕЛЬСТВО «МИР»

129820, Москва, И-110, ГСП, 1-й Рижский пер., 2.

Ленинградская типография № 6 ордена Трудового Красного Знамени
Ленинградского объединения «Техническая книга» им. Евгении Соколовой
Союзполиграфпрома при Государственном комитете СССР
по делам издательств, полиграфии и книжной торговли.
193144, г. Ленинград, ул. Монсеенко, 10.

**Physics of
Semiconductor
Devices**

Second Edition

S. M. Sze

Bell Laboratories,
Incorporated
Murray Hill, New Jersey

A Wiley-Interscience
publication
John Wiley & Sons
New York · Chichester · Brisbane ·
Toronto · Singapore
1981

С. Зи

Физика

полупроводниковых

приборов

В 2-х книгах

2

Перевод с английского
канд. физ.-мат. наук В. А. Гергеля,
канд. физ.-мат. наук Н. В. Зыкова,
д-ра физ.-мат. наук Б. И. Фукса
и канд. физ.-мат. наук Р. З. Хафизова
под редакцией
д-ра физ.-мат. наук Р. А. Суриса

Москва «Мир» 1984

ББК 32.852
З-5С
УДК 621.382

Зи С.
З-59 Физика полупроводниковых приборов: В 2-х книгах.
Кн. 2. Пер. с англ. -- 2-е перераб. и доп. изд. -- М.: Мир,
1984. -- 456 с., ил.

Монография написана известным американским специалистом в области полупроводниковой электроники. В переводе на русский язык выходит в двух книгах. Книга 2 посвящена физике приборов на туннельном эффекте и онтоэлектронных устройств (светодиодов, лазеров, фотодетекторов и солнечных батарей).

Для научных работников и инженеров, работающих в области электроники и вычислительной техники, а также студентов старших курсов вузов.

З $\frac{2403000000-431}{041(01)-84}$ 157-85, ч. 1

ББК 32.852
6Ф 0.32

Редакция литературы по новой технике

- © 1981, by John Wiley and Sons, Inc.
All rights reserved. Authorized translation
from English language edition published
by John Wiley and Sons, Inc.
© Перевод на русский язык, «Мир», 1984

МОП-ТРАНЗИСТОРЫ

8.1. ВВЕДЕНИЕ

Полевой транзистор со структурой металл—окисел—полупроводник (МОП-транзистор)¹ в настоящее время является основным элементом сверхбольших интегральных схем (СБИС), таких, как микропроцессоры и полупроводниковые запоминающие устройства. В последнее время они находят широкое применение и в мощных переключающих схемах. Принцип работы поверхностного полевого транзистора впервые предложен Лилленфельдом [1] и Хейлом [2] в начале 30-х годов. Позднее, в конце 40-х годов, теорию работы этих приборов развивали Шокли и Пирсон [3]. В 1960 г. Канг и Аталла [4] изготовили первый кремниевой МОП-транзистор, используя термическое окисление. Основные характеристики МОП-транзисторов затем исследовали Нанотти и Мотт [5, 6], Са [7], Хоффштейн и Хейман [8]. Для ознакомления с современным состоянием технологии, примененной в физике МОП-транзисторов можно рекомендовать обзорные статьи Уолмарка и Джонсона [9], Ричмана [10] и Брюса [11].

Поскольку ток в МОП-транзисторах переносится в основном носителями только одного типа (например, электронами в *n*-канальных приборах), МОП-транзисторы обычно относят к классу униполярных приборов. Работа других приборов этого класса — полевых транзисторов с *p-n*-переходом в качестве затвора (JFET) и полевых транзисторов с затвором Шоттки (MESFET) — была рассмотрена в гл. 6. Хотя в настоящее время известны МОП-транзисторы, изготовленные на различных полупроводниковых материалах, таких, как Ge [12], Si и GaAs [13], с использованием различных диэлектрических слоев (SiO_2 , Si_3N_4 и Al_2O_3), наибольшее распространение получили МОП-транзисторы на Si-SiO_2 .

Сначала мы рассмотрим основные характеристики обычных, так называемых длинноканальных МОП-транзисторов, в которых

¹ Приборы этого типа в зарубежной литературе носят разные названия, в том числе MOSFET (metal — oxide — semiconductor field-effect transistor, т. е. полевой транзистор металл — окисел — полупроводник), IGfet (insulation gate field-effect transistor, т. е. полевой транзистор с изолированным затвором), MISFET (metal — insulator — semiconductor field-effect transistor, т. е. полевой транзистор металл — изолятор — полупроводник), MOST (metal — oxide — semiconductor transistor, т. е. транзистор металл — окисел — полупроводник). В отечественной литературе наиболее часто используются термины «МОП-транзистор» и «МДП-транзистор». — *Прим. перев.*

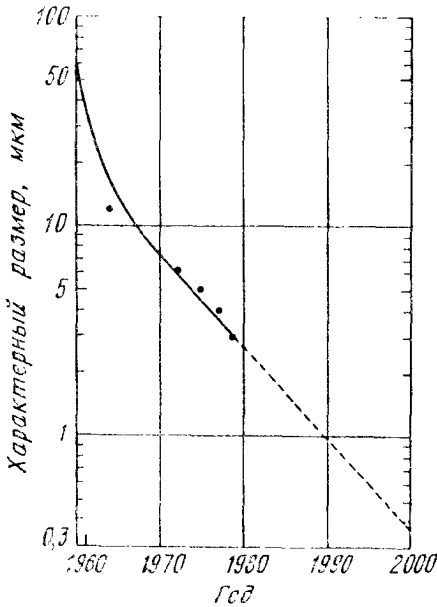


Рис. 1. Зависимость характерного размера отдельного прибора в интегральных схемах коммерческого назначения от времени [14].

длина канала L значительно превышает сумму толщин обедненных слоев стока и истока $(W_S + W_D)^1$. Это поможет также понять специфику работы так называемых короткоканальных приборов, в которых $L \leq (W_S + W_D)$. Данные, представленные на рис. 1, показывают, что с 1959 г. — «начала эры

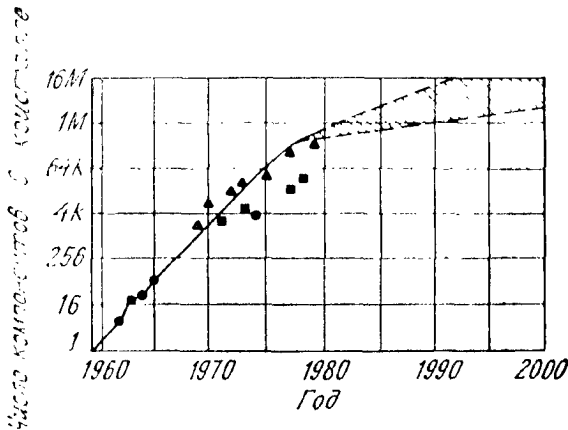


Рис. 2. Зависимость сложности интегральных схем от времени [15].

▲ — МОП-память; ■ — МОП-логика; ● — биполярная логика.

¹ Эти термины определены в разд. 8.2.

интегральных схем» — и до настоящего времени характерный размер отдельного элемента уменьшался, а число компонентов в одном кристалле соответственно увеличивалось экспоненциально во времени [14]. Ожидается, что «скорость» этих процессов в ближайшее время должна существенно снизиться из-за возрастающих трудностей конструирования и производства интегральных схем такого большого объема. Однако есть основания полагать, что к 1990 г. сложность интегральных схем будет составлять 10^6 и более компонентов на кристалл при характерном размере отдельного элемента ≤ 1 мкм (рис. 2) [15]. На рабочие характеристики МОП-транзисторов таких малых размеров могут существенно влиять так называемые короткоканальные эффекты: двумерный характер распределения потенциала под затвором, насыщение дрейфовой скорости носителей в сильных электрических полях, зарядка окисла и т. д.

В настоящее время для улучшения тех или иных рабочих характеристик предложено довольно большое число различных структурных модификаций МОП-транзисторов. Особенности работы некоторых, наиболее представительных из них описаны ниже. Мы рассмотрим также работу специальных МОП-транзисторных структур с многослойным затвором, которые используются в настоящее время в качестве основного элемента энергонезависимых полупроводниковых запоминающих устройств.

8.2. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ

Основная структура полевого транзистора металл—окисел—полупроводник (МОП-транзистора) показана на рис. 3. Этот четырехполюсный прибор состоит из полупроводниковой подложки

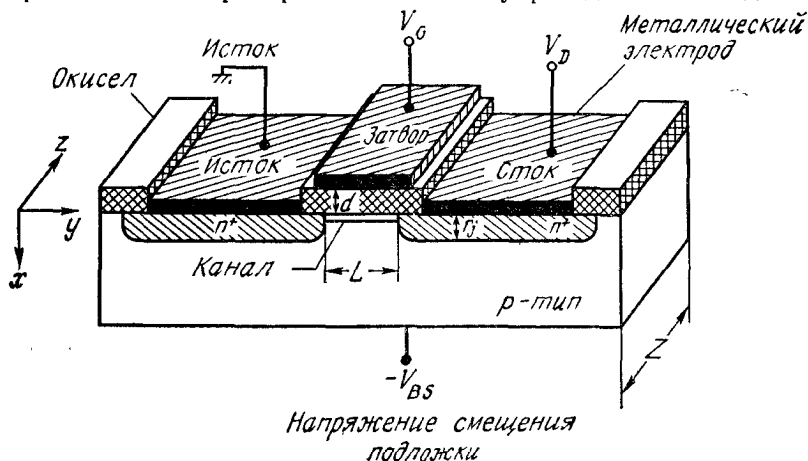


Рис. 3. Схема структуры МОП-транзистора [4].

p -типа, в которой сформированы (например, с помощью ионной имплантации) две высоколегированные n^+ -области — сток и исток¹. Металлический электрод, отделенный от подложки слоем окисла, называется затвором. В последнее время в качестве затвора используются высоколегированный поликремний, а также некоторые комбинации поликремния и силицидов. Основными параметрами структуры являются длина канала L — расстояние между металлургическими границами n^+ — p -переходов стока и истока, ширина канала Z , толщина слоя изолятора d , глубина переходов r_j и уровень легирования подложки N_A . В кремниевых интегральных схемах отдельный МОП-транзистор окружен в целях изоляции областью с толстым слоем окисла, который называется пассивирующим или полевым (в отличие от тонкого слоя подзатворного окисла).

Рассматривая работу прибора, мы будем все напряжения отсчитывать от потенциала истока, т. е. считать его заземленным. Когда напряжение на затворе отсутствует, электрическая цепь исток — сток представляет собой два n^+ — p -перехода, включенных навстречу друг другу. При этом в ней может протекать очень малый ток, равный току утечки обратносмещенного перехода². Если же к затвору приложено достаточно большое положительное напряжение, у границы с диэлектриком образуется инверсионный слой или канал, соединяющий n^+ -области стока и истока. Проводимость этого инверсионного канала модулируется при изменении напряжения на затворе. Тыловой контакт к подложке может находиться либо под тем же опорным потенциалом, что и исток, либо под напряжением, соответствующим обратному смещению n^+ — p -перехода исток—подложка. Напряжение обратного смещения подложки также влияет на проводимость инверсионного канала.

8.2.1. Неравновесные условия

Когда между стоком и истоком транзистора приложено напряжение, МОП-структура оказывается в неравновесных условиях. При этом квазиуровень Ферми неосновных носителей (в данном случае электронов) оказывается ниже равновесного уровня Ферми. На рис. 4 приведены двумерные зонные диаграммы, соответствующие различным режимам работы МОП-транзистора, который для удобства показан развернутым на 90° (рис. 4, а) [16]. Зонная

¹ Это так называемый n -канальный прибор. Для анализа работы p -канального МОП-транзистора следует просто поменять местами индексы n и p , а также изменить полярность приложенных напряжений.

² Здесь рассматривается работа нормально закрытого n -канального МОП-транзистора (обогащенного типа). Работа МОП-транзисторов других типов описана в последующих разделах.

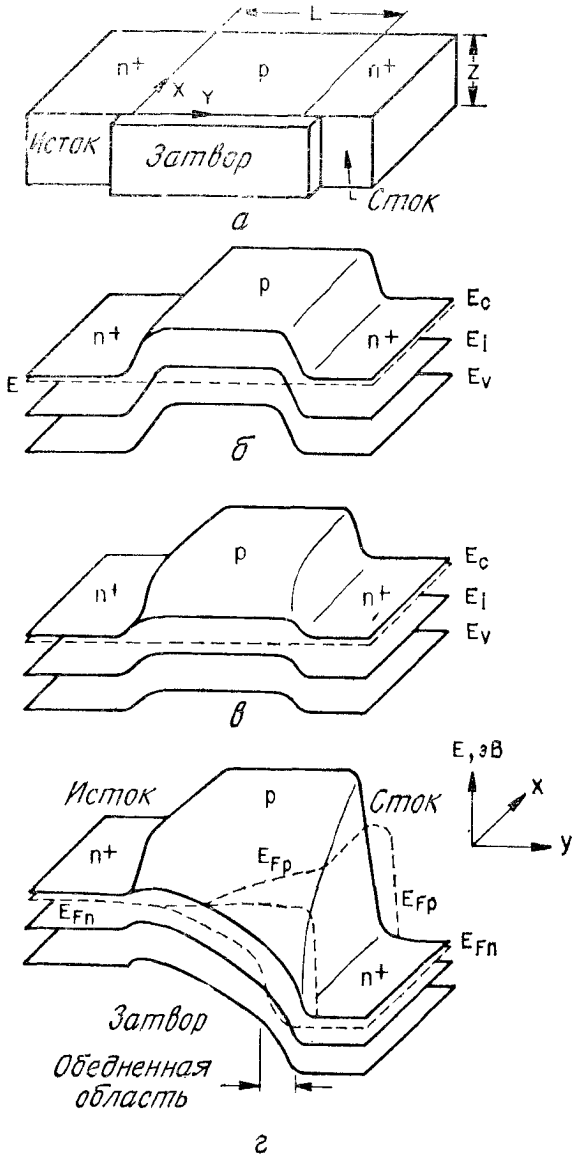


Рис. 4. Двумерная зонная диаграмма *n*-канального МОП-транзистора [16].
 а — конфигурация прибора; б — равновесная зонная диаграмма (плоские зоны, нулевое смещение подложки); в — равновесная зонная диаграмма при инвертирующем смещении на затворе; г — неравновесная ситуация (напряжения поданы на затвор и сток структуры).

схема на рис. 4, б соответствует равновесным условиям ($V_D = 0$), отсутствию обратного смещения на подложке ($V_{DS} = 0$) и нулевому напряжению на затворе ($V_G = 0$, плоские зоны). На рис. 4, в показана также равновесная ситуация, но уже с инвертированной границей раздела ($V_G > 0$). Неравновесная ситуация, когда соответствующие положительные напряжения приложены и к затвору, и к стоку, показана на рис. 4, г. Штриховыми линиями показаны дырочный и электронный квазиуровни Ферми. В то время как дырочный квазиуровень Ферми E_{Fp} не зависит от координат (остается равным равновесному уровню Ферми в объеме подложки), электронный квазиуровень Ферми E_{Fn} уменьшается по мере приближения к стоку. Из рис. 4, г видно также, что в неравновесных условиях, для того чтобы инвертировать поверхность вблизи стока, требуется большее напряжение на затворе, чем в равновесном случае, когда $\psi_s(\text{inv}) \simeq 2\psi_B$. Это обусловлено тем, что напряжение на стоке понижает электронный квазиуровень Ферми, а инверсионный слой на поверхности может образоваться только в том случае, когда поверхностный потенциал, отсчитанный от собственного уровня энергии E_i , пересекает квазиуровень Ферми неосновных носителей.

На рис. 5 приведены для сравнения пространственные распределения зарядов и зонные диаграммы в примыкающей к стоку части p -области в равновесных и неравновесных условиях [17]. В первом случае, который подробно рассмотрен в гл. 7, при инверсии толщина обедненной области достигает своего максимального равновесного значения W_m . В неравновесном случае толщина обедненного слоя зависит от напряжения на стоке V_D , а поверхностный потенциал, соответствующий началу сильной инверсии, с достаточной степенью точности можно считать равным

$$\psi_s(\text{inv}) \simeq V_D + 2\psi_B. \quad (1)$$

Соотношения, устанавливающие связь между зарядом инверсионного слоя, зарядом в обедненной области и напряжением на затвор для неравновесного случая выводятся аналогично тому, как это сделано в гл. 7 для равновесного случая. При этом, однако, используются еще два дополнительных предположения: 1) квазиуровни Ферми основных и неосновных носителей не зависят от расстояния до границы раздела; 2) разность между этими квазиуровнями равна напряжению на стоковом переходе, т. е. $E_{Fp} = -E_{Fn} + qV_D$ (для p -подложки). Первое предположение в условиях инверсии приводит к несущественным ошибкам, поскольку в этом случае заряд основных носителей составляет пренебрежимо малую часть полного пространственного заряда в приповерхностном слое. Второе предположение справедливо только в условиях сильной инверсии, когда заряд инверсионного слоя составляет

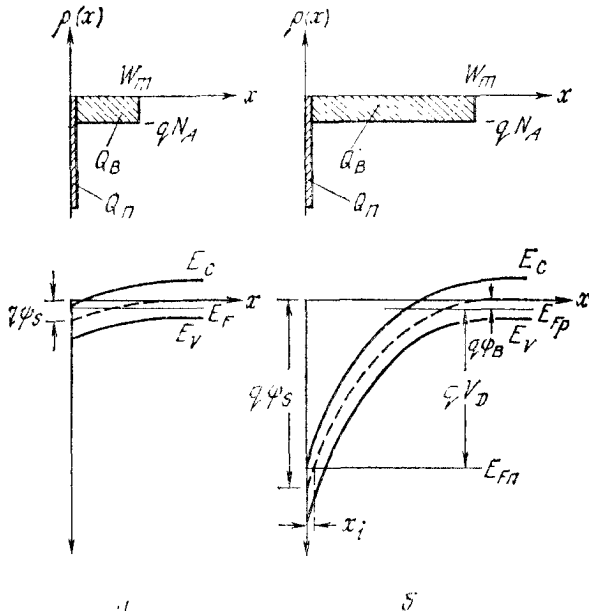


рис. 6. Распределение зарядов и энергетические диаграммы в инвертированной части p -подложки [17].

a — при равновесных условиях; *b* — неравновесные условия вблизи стока

существенную часть полного заряда приповерхностной области полупроводниковой подложки.

Основываясь на этих предположениях, запишем одномерное уравнение Пуассона для области пространственного заряда в прилегающей к стоку части p -подложки в виде

$$\frac{\partial^2 \psi}{\partial x^2} = -\frac{q}{\varepsilon_s} (N_D^+ - N_A^- + p - n), \quad (2)$$

где, как и в гл. 7,

$$\begin{aligned} N_D^+ - N_A^- &= n_{p0} - p_{p0}, \quad p_{p0} \approx N_A, \quad p = p_{p0} e^{-\beta \psi}, \\ \text{но } n &= n_{p0} e^{\beta \psi - \beta V_D}, \quad \beta \equiv q/kT. \end{aligned} \quad (3)$$

Тем же способом, что и в гл. 7, получим отсюда

$$\mathcal{E} = -\frac{\partial \psi}{\partial x} = \pm \frac{\sqrt{2} kT}{q L_D} F(\beta \psi, V_D, \frac{n_{p0}}{p_{p0}}) \quad (4)$$

и

$$Q_s = -\varepsilon_s \mathcal{E}_s = \mp \frac{\sqrt{2} \varepsilon_s kT}{q L_D} F(\beta \psi_s, V_D, \frac{n_{p0}}{p_{p0}}), \quad (5)$$

где

$$F\left(\beta\psi, V_D, \frac{n_{p0}}{\rho_{p0}}\right) \equiv \left[e^{-\beta\psi} + \beta\psi - 1 + \frac{n_{p0}}{\rho_{p0}} e^{-\beta V_D} (e^{\beta\psi} - \beta\psi e^{\beta V_D} - 1) \right]^{1/2}, \quad (6)$$

а

$$L_D \equiv \left(\frac{kT\epsilon_s}{\rho_{p0}q^2} \right)^{1/2}. \quad (7)$$

В режиме сильной инверсии полный поверхностный заряд (5) можно представить в виде

$$Q_s = Q_n + Q_B, \quad (8)$$

где

$$Q_B = -qN_A W_m = -\sqrt{2qN_A \epsilon_s (V_D + 2\psi_B)} \quad (9)$$

и Q_n заряда, обусловленный неосновными носителями в инверсионном слое:

$$|Q_n| \equiv q \int_0^{x_i} n(x) dx = q \int_{\psi_s}^{\psi_B} \frac{n(\psi) d\psi}{d\psi/dx}, \quad (10)$$

$$|Q_n| \approx q \int_{\psi_s}^{\psi_B} \frac{n_{p0} e^{(\beta\psi - \beta V_D)} d\psi}{(V_D^2 kT/qL_D) F(\beta\psi, V_D, n_{p0}/\rho_{p0})}. \quad (11)$$

Здесь x_i — точка, в которой собственный уровень Ферми пересекает квазиуровень Ферми неосновных носителей — электронов. Для используемых на практике уровней легирования кремния величина x_i довольно мала (30—300 Å). Выражение (11) является одним из основных соотношений теории МОП-транзистора с длинным каналом, и в конкретных ситуациях соответствующая зависимость $Q_n(\psi_s)$ рассчитывается численными методами. Более простое и удобное выражение для Q_n , справедливое при достаточно сильной инверсии, дает так называемая модель заряженных слоев [18]:

$$|Q_n| = \sqrt{2}qN_A L_D \left\{ \left[\beta\psi_s + \left(\frac{n_{p0}}{\rho_{p0}} \right) e^{(\beta\psi_s - \beta V_D)} \right]^{1/2} - (\beta\psi_s)^{1/2} \right\}. \quad (12)$$

Выражения (11) и (12) для Q_n выведены для случая, когда обратное смещение на подложке отсутствует ($V_{BS} = 0$). При наличии такого смещения при расчете Q_n следует подставить $\beta(V_D + V_{BS})$ в формулу (12) вместо βV_D .

8.2.2. Линейная область характеристик. Область насыщения

Рассмотрим сначала качественно работу МОП-транзистора. Будем считать, что напряжение на затворе структуры G достаточно для сильной инверсии на границе с окислом (рис. 6, a). Тогда, если напряжение на стоке V_D не слишком велико, инвер-

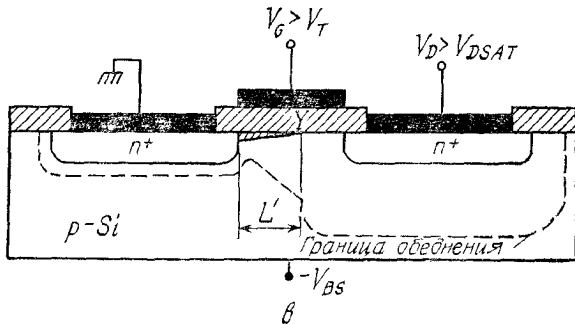
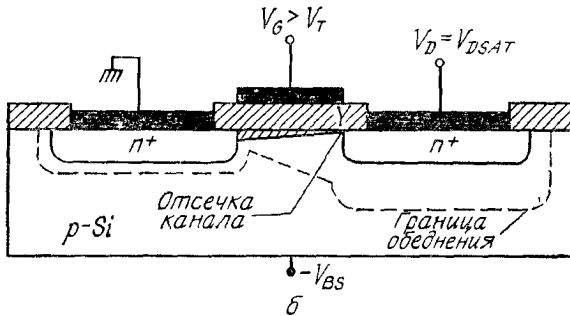
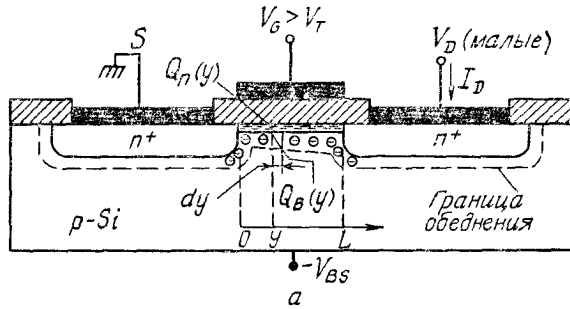


Рис. 6. МОП-транзистор в линейном режиме (малые напряжения на стоке) (a), в начале насыщения (отсечка канала на границе со стоком $Y = L$) (b) и в режиме насыщения (показано уменьшение эффективной длины канала) ($в$).

сионный слой действует как обычное сопротивление, и ток через проводящий инверсионный канал I_D будет увеличиваться пропорционально напряжению стока V_D . Эта область называется линейной областью работы прибора. С увеличением напряжения на стоке в конце концов достигается момент, когда ширина канала x_i , а следовательно, и заряд инверсионного слоя Q_n в точке $y = L$ (прямо на границе стока) становится равным нулю. Эти условия соответствуют началу режима отсечки канала (рис. 6, б). Напряжение стока в данном режиме обозначим $V_{D\text{sat}}$. При больших напряжениях ($V_D > V_{D\text{sat}}$) точка отсечки Y сдвигается к истоку, а ток стока при этом увеличивается лишь незначительно (рис. 6, в), поскольку напряжение между истоком и точкой отсечки канала Y по-прежнему равно $V_{D\text{sat}}$. Инжекция неосновных носителей из точки Y в обедненную область стока подобна процессу инжекции носителей переходом эмиттер—база биполярного транзистора в обедненную область его коллекторного перехода.

Теперь выведем аналитические соотношения, определяющие основные характеристики МОП-транзисторов, используя при этом следующие допущения: 1) подзатворная область прибора представляет собой идеальную МОП-структуру, что в соответствии с определением этого понятия, данным в гл. 7, означает отсутствие поверхностных состояний, фиксированного заряда в окисле, разности работ выхода между полупроводником и металлом и т. д.; 2) учитывается только дрейфовая компонента тока; 3) подвижность носителей в инверсионном слое постоянна (не зависит от электрического поля); 4) концентрация легирующей примеси в канале однородна; 5) обратные токи утечки p — n -переходов пренебрежимо малы; 6) поперечное электрическое поле \mathcal{E}_x , направленное перпендикулярно границе раздела (по оси x), значительно превышает продольное электрическое поле \mathcal{E}_y , направленное вдоль поверхности (по оси y). Последнее допущение называется приближением плавного канала.

В соответствии с принятыми допущениями поверхностная плотность полного заряда в полупроводнике на расстоянии y от истока и поверхностный потенциал $\psi_s(y)$ в той же точке связаны следующим соотношением:

$$Q_s(y) = [-V_G + \psi_s(y)] C_i, \quad (13)$$

где $C_i \equiv \epsilon_i/d$ — удельная емкость окисла. Используя выражение (8) для заряда инверсионного слоя, получим

$$Q_n(y) = Q_s(y) - Q_B(y) = -[V_G - \psi_s(y)] C_i - Q_B(y). \quad (14)$$

В режиме сильной инверсии поверхностный потенциал ψ_s равен $2\psi_B + V(y)$, где $V(y)$ — напряжение обратного смещения

в точке y по отношению к потенциалу истока (который мы считаем заземленным). Как уже говорилось выше, заряд обедненного слоя

$$Q_B(y) = -qN_A W_m = -\sqrt{2\epsilon_s q N_A [V(y) + 2\psi_B]}. \quad (15)$$

Подставив это выражение в выражение (14), получим

$$Q_n(y) = -[V_G - V(y) - 2\psi_B] C_i + \sqrt{2\epsilon_s q N_A [V(y) + 2\psi_B]}. \quad (16)$$

Проводимость канала представляет собой интеграл

$$g = \frac{Z}{L} \int_0^{x_i} \sigma(x) dx \quad (17)$$

от объемной удельной проводимости

$$\sigma(x) = qn(x) \mu_n(x). \quad (18)$$

Полагая, что подвижность μ не зависит от x , получим

$$g = \frac{qZ\mu_n}{L} \int_0^{x_i} n(x) dx = qZ\mu_n |Q_n|/L. \quad (19)$$

Сопротивление элементарного участка канала длиной dy есть

$$dR = \frac{dy}{gL} = \frac{dy}{Z\mu_n |Q_n(y)|} \quad (20)$$

и, следовательно, падение напряжения на этом участке

$$dV = I_D dR = \frac{I_D dy}{Z\mu_n |Q_n(y)|}, \quad (21)$$

где ток стока I_D — константа, не зависящая от y . Умножив выражение (16) на выражение (21) и проинтегрировав полученное выражение от истока ($y = 0$, $V = 0$) до стока ($y = L$, $V = V_D$), для рассматриваемого идеализированного случая найдем

$$I_D = \frac{Z}{L} \mu_n C_i \left\{ \left(V_G - 2\psi_B - \frac{V_D}{2} \right) V_D - \frac{2}{3} \frac{\sqrt{2\epsilon_s q N_A}}{C_i} \left[(V_D + 2\psi_B)^{3/2} - (2\psi_B)^{3/2} \right] \right\}. \quad (22)$$

Из выражения (22) следует, что при фиксированном напряжении на затворе V_G ток стока I_D сначала линейно увеличивается с напряжением стока V_D (линейная область), затем крутизна характеристики плавно уменьшается до нуля, после чего начинается область насыщения. Семейство выходных ($I_D - V_D$) характеристик идеального МОП-транзистора приведено на рис. 7.

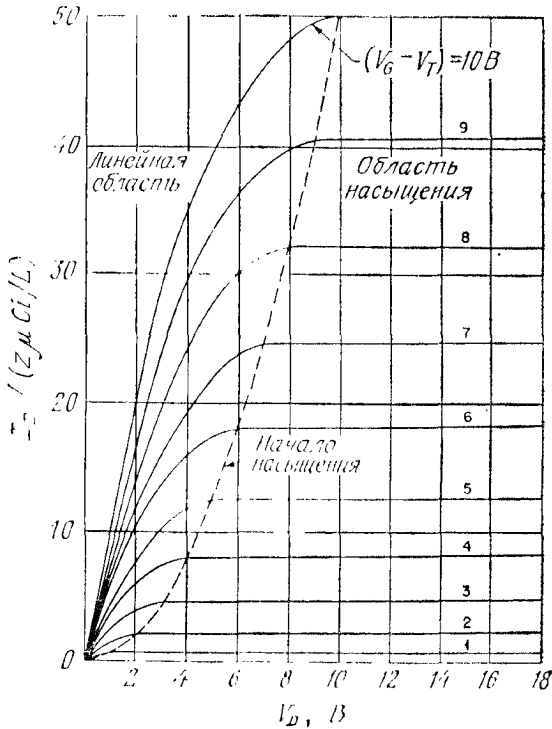


Рис. 7. Выходные (стоковые) характеристики идеального МОП-транзистора (I_D в зависимости от V_D). Штриховая линия соединяет точки начала насыщения на характеристиках, соответствующих различным емещениям на затворе транзистора ($J_D \text{ sat. } V_D \text{ sat.}$). При $V_D > V_D \text{ sat}$ ток стока остается постоянным.

Рассмотрим названные выше две области более подробно. Для малых напряжений стока, разлагая выражение (22), получим

$$I_D \approx \frac{z}{L} \mu_n C_i \left[(V_G - V_T) V_D - \left(\frac{1}{2} + \frac{\sqrt{\varepsilon_s q N_A (2\psi_B)}}{4C_i} \right) V_D^2 \right], \quad (23)$$

или еще проще (при $V_D \ll (V_G - V_T)$)

$$I_D \approx \frac{z}{L} \mu_n C_i (V_G - V_T) V_D, \quad (23a)$$

где

$$V_T = 2\psi_B + \frac{\sqrt{2\varepsilon_s q N_A (2\psi_B)}}{C_i} \quad (24)$$

— пороговое напряжение. Теоретические зависимости порогового напряжения от концентрации примеси и толщины изолятора для Si—SiO₂-систем были приведены в гл. 7 (рис. 14). Эксперимен-

тально пороговое напряжение определяют, линейно продолжая измеренную при малых V_D сток-затворную характеристику ($I_D - V_G$) до пересечения с осью напряжений.

Дифференцируя выражение (23а), получим для проводимости g_D и крутизны g_m идеального МОП-транзистора в линейной области

$$g_D \equiv \left. \frac{\partial I_D}{\partial V_D} \right|_{V_G = \text{const}} = \frac{Z}{L} \mu_n C_i (V_G - V_T), \quad (25)$$

$$g_m \equiv \left. \frac{\partial I_D}{\partial V_G} \right|_{V_D = \text{const}} = \frac{Z}{L} \mu_n C_i V_D. \quad (26)$$

По мере увеличения напряжения стока V_D заряд инверсионного слоя Q_n в точке $y = L$ (у края стока) уменьшается и, наконец, становится равным нулю при $V_D = V_{D \text{ sat}}$. Это явление по аналогии с подобным эффектом в полевых транзисторах с $p-n$ -переходом в качестве затвора называется отсечкой канала. При $V_D = V_{D \text{ sat}}$ ток стока достигает максимального значения $I_{D \text{ sat}}$. При больших напряжениях ($V_D > V_{D \text{ sat}}$) наступает насыщение ($I_D = I_{D \text{ sat}}$). Величину напряжения отсечки $V_{D \text{ sat}}$ получим, приравняв нулю выражение (16) с $V(L) = V_{D \text{ sat}}$:

$$V_{D \text{ sat}} = V_G - 2\psi_n - K^2 (1 - \sqrt{1 + 2V_G/K^2}), \quad (27)$$

где $K = \sqrt{qN_A/C_i}$. Ток насыщения $I_{D \text{ sat}}$ можно найти, подставив выражение (27) в выражения (22) и (23):

$$I_{D \text{ sat}} \approx \frac{mZ}{L} \mu_n C_i (V_G - V_T)^2, \quad (28)$$

где коэффициент m в общем случае зависит от уровня легирования подложки. При сравнительно малых концентрациях примеси $m = 1/2$ [11]. Пороговое напряжение V_T в этом выражении в режиме насыщения для МОП-транзисторов с низколегированными подложками определяется также уравнением (24). В сильнолегированных структурах V_T оказывается зависящим от V_G . Крутизна МОП-транзистора в области насыщения в соответствии с выражением (28) равна

$$g_m = \left. \frac{\partial I_D}{\partial V_G} \right|_{V_D = \text{const}} = \frac{2mZ}{L} \mu_n C_i (V_G - V_T). \quad (29)$$

При выводе соотношений, определяющих основные характеристики МОП-транзисторов, мы использовали ряд упрощающих предположений. Попробуем теперь отказаться от первых двух предположений, т. е. проанализируем, какое влияние на характеристики МОП-транзистора оказывают неидеальность МОП-структуры и диффузионная компонента тока в канале. Как уже говорилось в гл. 7, фиксированный заряд окисла и разность

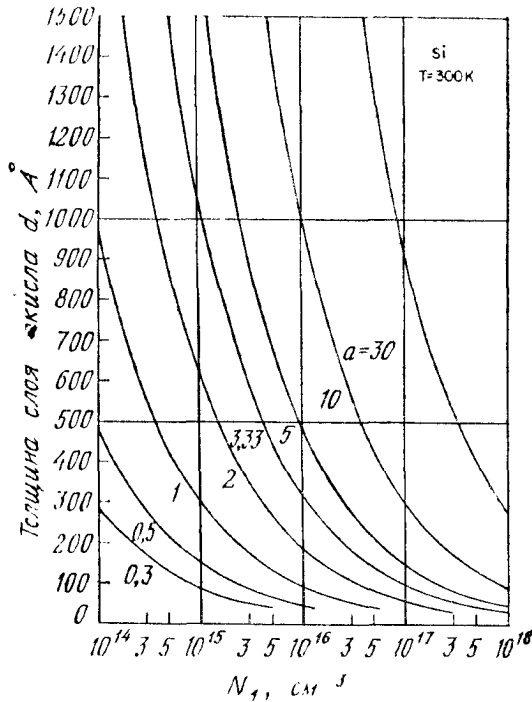


Рис. 8. Номограмма для определения значений безразмерного параметра a [19].

работ выхода электрона из металла и полупроводника обуславливают сдвиг напряжения плоских зон МОП-структуры V_{FB} . Ясно, что на эту же величину должно сдвинуться и пороговое напряжение МОП-транзистора:

$$\begin{aligned}
 V_T &= V_{FB} - 2\psi_B - \frac{\sqrt{2\epsilon_s q N_A (2\psi_B)}}{C_i} = \\
 &= \left(\varphi_{ms} - \frac{Q_f}{C_i} \right) + 2\psi_B + \frac{\sqrt{4\epsilon_s q N_A \psi_B}}{C_i}. \quad (30)
 \end{aligned}$$

На пороговое напряжение МОП-транзистора влияет также напряжение обратного смещения подложки V_{BS} :

$$V_T = V_{FB} + 2\psi_B + \sqrt{2\epsilon_s q N_A (2\psi_B + V_{BS})} / C_i, \quad (31)$$

или

$$\begin{aligned}
 \Delta V_T &= V_T(V_{BS}) - V_T(V_{BS} = 0) = \\
 &= \frac{\sqrt{2\epsilon_s q N_A}}{C_i} (\sqrt{2\psi_B + V_{BS}} - \sqrt{2\psi_B}) =
 \end{aligned}$$

$$= \frac{a}{\beta} (\sqrt{2\beta\psi_B + \beta V_{BS}} - \sqrt{2\beta\psi_B}), \quad (32)$$

где

$$a \equiv \sqrt{2} (\epsilon_s/L_D)/C_i = \sqrt{2} (\epsilon_s/\epsilon_i) (d/L_D). \quad (33)$$

При выполнении конкретных расчетов для нахождения a удобно использовать номограмму, приведенную на рис. 8 [19]. Семейство зависимостей сдвига порогового напряжения ΔV_T от смещения подложки V_{BS} для некоторых типичных значений параметра a приведено на рис. 9. Легко видеть, что чем больше a , тем больше ΔV_T .

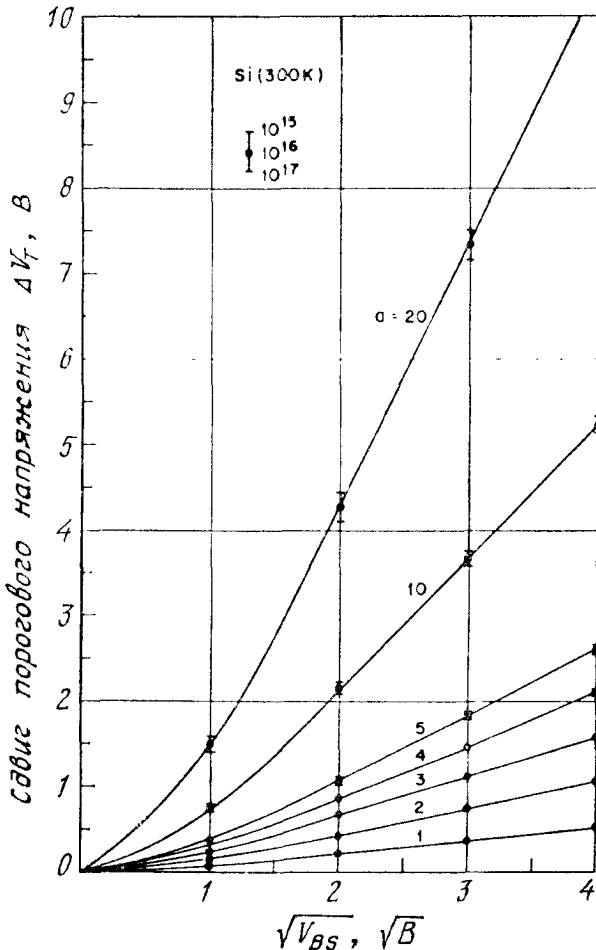


Рис. 9. Сдвиг порогового напряжения за счет обратного смещения подложки при различных значениях параметра a .

Для исследования влияния диффузионной компоненты тока канала рассмотрим двумерную зонную схему МОП-транзистора в неравновесных условиях (рис. 4, *з*), где, в частности, показано изменение положения электронного квазиуровня Ферми по длине канала. С помощью этого квазиуровня ψ_{Fn} плотность полного тока канала, включая обе его компоненты (диффузионную и дрейфовую), можно записать в виде

$$J_D(x, y) = q\mu_n n \mathcal{E}_y + qD_n \nabla n = -qD_n n(x, y) \nabla \psi_{Fn}. \quad (34)$$

Снова используя приближение плавного канала, из выражения (34) получим для полного тока

$$\begin{aligned} I_D &= \int_0^{x_i} J_D(x, y) Z dx = \\ &= \frac{1}{L} \int_0^L D_n q Z \left(\frac{\partial \psi_{Fn}}{\partial y} \right) \int_0^{x_i} n(x, y) dx dy = \\ &= \frac{Z}{L} \frac{\varepsilon_s \mu_n}{L_D} \int_0^{V_D} \int_{\psi_B}^{\psi_s} \frac{e^{\beta\psi - \beta V}}{F(\beta\psi, V, n_{p0}/p_{p0})} d\psi dV. \end{aligned} \quad (35)$$

Напряжение на затворе связано с поверхностным потенциалом ψ_s соотношением

$$\begin{aligned} V_G &= V_G - V_{FB} = -\frac{Q_s}{C_i} + \psi_s = \\ &= \frac{2\varepsilon_s kT}{C_i q L_D} F\left(\beta\psi_s, V, \frac{n_{p0}}{p_{p0}}\right) + \psi_s. \end{aligned} \quad (36)$$

Когда напряжение на затворе V_G существенно превышает пороговое напряжение V_T , выражение (35) переходит в приближенную формулу (22). Однако в окрестности порогового напряжения, а также вблизи отсечки канала точность формулы (22) недостаточна, и здесь следует использовать точное выражение (35). Для конкретных структур с определенными значениями физических параметров (размеров, концентрации примеси, подвижности) уравнения (35) и (36) решаются численными методами. Получаемые при этом зависимости справедливы при любых напряжениях на стоке (и в линейной области, и в области насыщения). В качестве примера на рис. 10 приведены результаты одного из таких расчетов [16], которые отчетливо демонстрируют эффект насыщения тока — типичное свойство выходных характеристик МОП-транзисторов с длинным каналом.

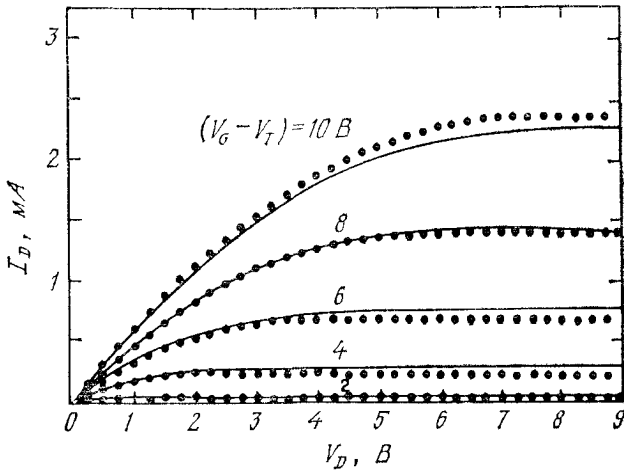


Рис. 10. Теоретические (●) и экспериментальные (—) стоковые характеристики p -канального МОП-транзистора с параметрами $d = 2000 \text{ \AA}$, $N_D = 4,6 \cdot 10^{14} \text{ см}^{-3}$ и $\mu_p = 256 \text{ см}^2 \cdot \text{В}^{-1} \cdot \text{с}^{-1}$ [16].

8.2.3. Подпороговая область

Если напряжение на затворе МОП-транзистора ниже порогового и граница с окислом находится в условиях слабой инверсии, соответствующий ток стока называют подпороговым током [20, 21]. Подпороговая область характеристик особенно важна для МОП-транзисторов, предназначенных для работы в низковольтных полупроводниковых устройствах с малым энергопотреблением (цифровые логические схемы, устройства памяти), поскольку именно этот режим соответствует закрытому состоянию МОП-транзистора, а также описывает процесс переключения из закрытого состояния в открытое.

В условиях слабой инверсии основной компонентой тока канала является диффузионная составляющая. Формула для тока стока выводится так же, как и выражение для коллекторного тока в биполярном транзисторе с однородно легированной базой. Рассматривая МОП-транзистор как $n-p-n$ -исток—подложка—сток)-биполярный транзистор, можно написать

$$I_D = -qAD_n \frac{dn}{dy} = qAD_n \frac{n(0) - n(L)}{L}, \quad (37)$$

где A — эффективная поперечная площадь канала, а $n(0)$ и $n(L)$ — объемные концентрации электронов у истока и стока соответственно (рис. 6, а):

$$n(0) = n_{p0} e^{\beta \psi_s}, \quad (38a)$$

$$n(L) = n_{p0} e^{\beta \psi_s - \beta V_D}. \quad (38b)$$

Здесь ψ_s — поверхностный потенциал у истока. Эффективная поперечная площадь канала представляет собой произведение ширины канала Z на эффективную толщину инверсионного слоя в направлении, перпендикулярном поверхности. Поскольку концентрация электронов экспоненциально зависит от потенциала ψ , эффективная толщина канала равна расстоянию от поверхности, на котором потенциал уменьшается на величину порядка kT/q . Следовательно, эффективная толщина канала равна $kT/q\mathcal{E}_s$, где \mathcal{E}_s — поверхностное электрическое поле, которое в режиме слабой инверсии определяется выражением

$$\mathcal{E}_s = -Q_B/\epsilon_s = \sqrt{2qN_A\psi_s/\epsilon_s}. \quad (39)$$

Подставив выражения (38) и (39) в выражение (37), получим [18, 22]

$$I_D = \mu_n \left(\frac{Z}{L}\right) \frac{aC_i}{2\beta^2} \left(\frac{n_i}{N_A}\right)^2 (1 - e^{-\beta V_D}) e^{\beta\psi_s} (\beta\psi_s)^{-1/2}. \quad (40)$$

В последнем выражении были использованы соотношение Эйнштейна $D_n = \mu_n kT/q$ и безразмерный параметр a , характеризующий МОП-структуру. Поверхностный потенциал ψ_s у истока в этих обозначениях записывается через напряжения на затворе V_G следующим образом [18, 19]:

$$\psi_s = (V_G - V_{FB}) - \frac{a^2}{2\beta} \left\{ \left[1 + \frac{4}{a^2} (\beta V_G - \beta V_{FB} - 1) \right]^{1/2} - 1 \right\}. \quad (41)$$

Из выражения (40) следует, что в подпороговой области ток стока экспоненциально увеличивается с ростом напряжения на затворе V_G и перестает зависеть от напряжения стока при $V_D > 3kT/q$. С помощью формул (40) и (41) можно определить характерное изменение напряжения на затворе S , требуемое для изменения тока стока на порядок. По определению

$$\begin{aligned} S &= \ln 10 dV_G/d(\ln I_D) = \\ &= (kT/q) \ln 10 d(\beta V_G)/d(\ln I_D) = \\ &= (kT/q) \ln 10 \left[1 + C_D(\psi_s)/C_i \right] \left\{ 1 - \left(\frac{2}{a^2} \right) [C_D(\psi_s)/C_i]^2 \right\}. \end{aligned} \quad (42)$$

В обычных ситуациях, когда $a \gg C_D/C_i$,

$$S \approx \frac{kT}{q} \ln 10 (1 + C_D/C_i). \quad (43)$$

Слагаемое в скобках здесь представляет собой отношение емкостного делителя $(C_i + C_D)/C_i$. В структурах с высокой плотностью поверхностных состояний D_{it} эффективную емкость поверхностных состояний $C_{it} = qD_{it}$ нужно «включить» параллельно емкости обед-

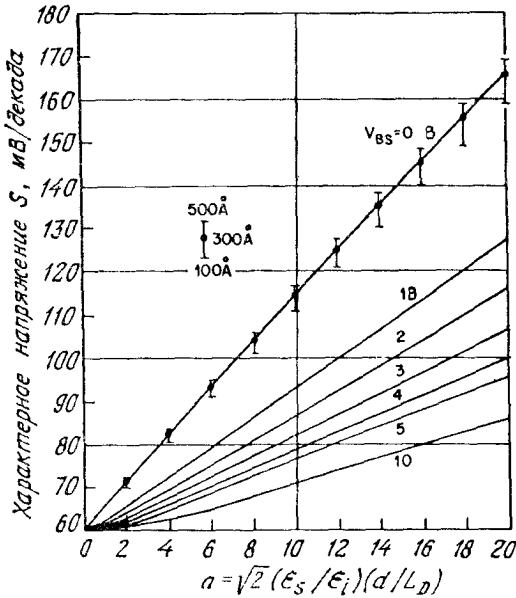


Рис. 11. Характерный сдвиг затворного напряжения S , обеспечивающий в подпороговой области изменение тока на порядок, при различных обратных смещениях подложки [19].

нишного слоя C_D , т. е. заменить в выражении (43) C_D на $C_D + C_{it}$. При этом

$$S_{Dit} \approx \frac{kT}{q} \ln 10 \left(1 + \frac{C_D + C_{it}}{C_i} \right). \quad (44)$$

Когда к подложке МОП-транзистора приложено обратное смещение, поверхностный потенциал ψ_s увеличивается и соответственно уменьшается емкость обедненного слоя C_D , а следовательно, и характерное изменение затворного напряжения S . На рис. 11 приведены теоретические зависимости S от величины параметра a при различных обратных смещениях на подложке МОП-транзистора [19]. Следует отметить, что существенное уменьшение S вызывает уже «первый вольт» обратного смещения подложки.

Экспериментальные подпороговые характеристики МОП-транзистора с длинным каналом (15,5 мкм) для трех значений напряжения обратного смещения V_{BS} приведены на рис. 12 [23]. Как и следовало ожидать, ниже порогового напряжения V_T , отмеченного на каждой из кривых стрелочкой, ток стока МОП-транзистора экспоненциально зависит от напряжения на затворе.

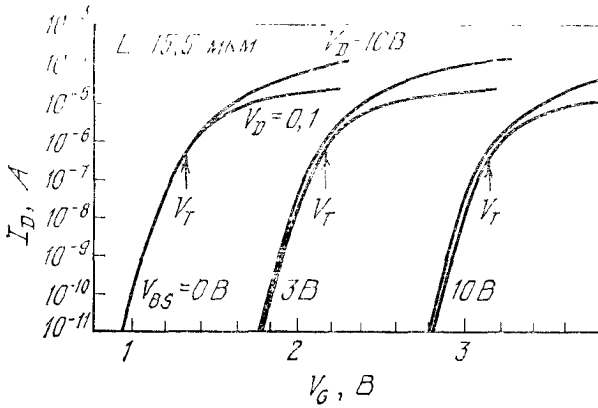


Рис. 12. Экспериментальные подпороговые характеристики МОП-транзистора с длинным каналом ($L = 15,5$ мкм) [23].

Отметим также, что подпороговый участок экспериментальных кривых практически не изменяется при изменении напряжения стока от 0,1 до 10 В. Независимость I_D от V_D на подпороговом участке характерна для длинноканального поведения МОП-транзистора. Толщина слоя окисла и уровень легирования подложки прибора, характеристики которого приведены на рис. 12, составляли 570 \AA и $5,6 \cdot 10^{16} \text{ см}^{-3}$. (Параметр a равен 4.) При этом теоретические значения $S = 83$ мВ на декаду для $V_{BS} = 0$, $S = 67$ мВ на декаду для $V_{BS} = 3$ В и $S = 63$ мВ на декаду при $V_{BS} = 10$ В (рис. 11), а $\Delta V_T = 0,75$ В для $V_{BS} = 3$ В и $\Delta V_T = 1,7$ В для $V_{BS} = 10$ В (рис. 9). Эти результаты близки к экспериментальным значениям соответствующих величин, найденным по характеристикам прибора (рис. 12).

8.2.4. Подвижность носителей

При малых продольных электрических полях \mathcal{E}_y дрейфовая скорость носителей инверсионного слоя пропорциональна приложенному полю $v_d = \mu_n \mathcal{E}_y$ (\mathcal{E}_y — параллельно направлению тока). Из результатов измерений следует, что дрейфовая поверхностная подвижность μ_n зависит от поперечного электрического поля \mathcal{E}_x (\mathcal{E}_x перпендикулярно направлению тока). На рис. 13 приведены зависимости $\mu_n(\mathcal{E}_x)$, измеренные [24] при трех различных температурах в кремниевых МОП-структурах на p -подложке с ориентацией $\langle 100 \rangle$. Отметим, что, согласно результатам работы [24], подвижность μ_n не зависит от обработки поверхности и уровня легирования подложки при $N_A < 10^{17} \text{ см}^{-3}$. При постоянной

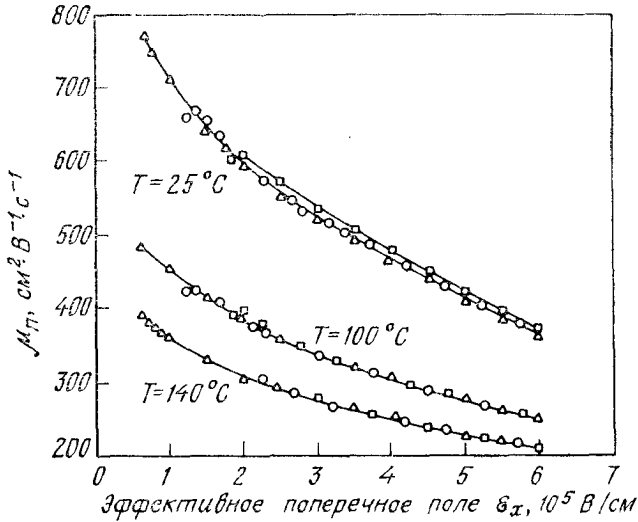


Рис. 13. Зависимость подвижности инверсионного слоя μ_n от эффективного поперечного поля ϵ_y при трех температурах ($\epsilon_x < 10^3 \text{ В/см}$) [24].

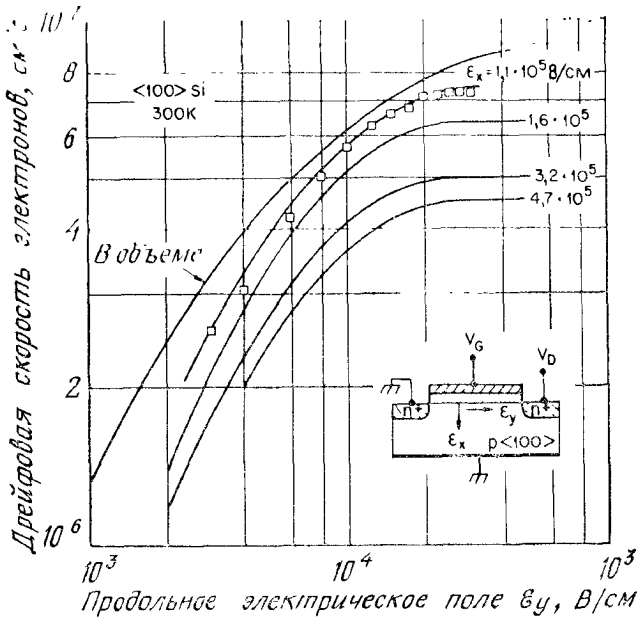


Рис. 14. Зависимость дрейфовой скорости электронов от продольного поля при некоторых значениях поперечного поля [25, 26].

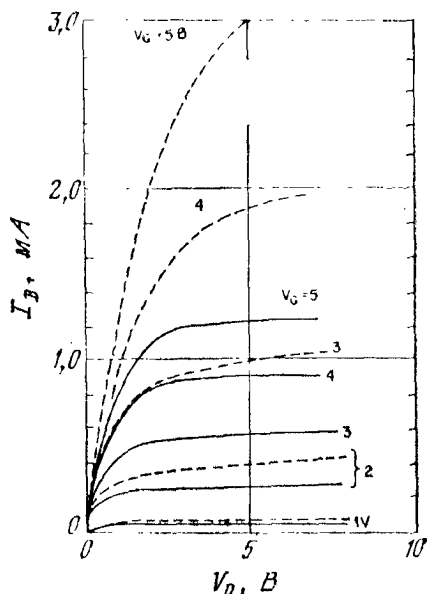


Рис. 15. Стоковые характеристики, рассчитанные в предположении постоянной подвижности (---) и с учетом полевой зависимости подвижности (—) [28].

температуре поверхностная подвижность уменьшается с увеличением эффективного поперечного электрического поля

$$(\mathcal{E}_x)_{\text{eff}} = \frac{1}{v_s} \left(Q_n + \frac{1}{2} Q_n \right), \quad (45)$$

где $(\mathcal{E}_x)_{\text{eff}}$ — поперечное электрическое поле, усредненное по распределению электронов в инверсионном слое.

При увеличении продольного электрического поля \mathcal{E}_y наступает насыщение дрейфовой скорости (аналогично тому, что наблюдается в объеме кремния). Экспериментальные зависимости дрейфовой скорости инверсионных электронов от величины продольного электрического поля приведены на рис. 14 [25, 26]. При фиксированном значении поперечного электрического поля \mathcal{E}_x дрейфовая скорость при малых значениях \mathcal{E}_y пропорциональна продольному полю \mathcal{E}_y , причем коэффициент пропорциональности равен подвижности μ_n . Зависимость μ_n от \mathcal{E}_x приведена на рис. 13. При увеличении поля \mathcal{E}_y дрейфовая скорость стремится к определенному пределу. Ее зависимость от поля \mathcal{E}_y можно аппроксимировать довольно общим выражением [27]

$$v_d = v_0 \left[1 + \left(\frac{v_0}{v_c} \right)^2 \left(\frac{v_0}{v_c} + G \right)^{-1} + \left(\frac{v_0}{v_s} \right) \right]^{-1/2}, \quad (46)$$

где v_c , v_s и G — подгоночные параметры, а

$$v_0 \cong \mu_n(\mathcal{E}_x) \mathcal{E}_y. \quad (47)$$

Отметим, что подвижность $\mu_n(\mathcal{E}_x)$ — однозначная функция поперечного поля \mathcal{E}_x . Если $\mathcal{E}_y \rightarrow 0$, зависимость $v_d(\mathcal{E}_y)$ стремится к $v_d = v_0 = \mu_n(\mathcal{E}_x)\mathcal{E}_y$. Если же $\mu_n\mathcal{E}_y > v_s$, дрейфовая скорость примерно равна величине v_s , которая также зависит от поперечного электрического поля \mathcal{E}_x .

В гл. 6 рассматривалось влияние насыщения дрейфовой скорости на характеристики полевых транзисторов с p — n -переходом в качестве затвора. Аналогичные результаты для МОП-транзистора приведены на рис. 15, где сравниваются расчетные характеристики, полученные в предположении постоянной подвижности, и экспериментальные зависимости в том же приборе с учетом насыщения дрейфовой скорости [28]. Эффект насыщения дрейфовой скорости приводит, во-первых, к существенному уменьшению тока насыщения МОП-транзистора и, во-вторых, к тому, что ток насыщения примерно пропорционален напряжению на затворе, а не квадратичен, как предсказывает теория идеального МОП-транзистора [28]. В условиях насыщения дрейфовой скорости ток насыщения МОП-транзистора

$$I_{D \text{ sat}} = ZC_i(V_G - V_T)v_s, \quad (48)$$

и, следовательно, его крутизна в этом режиме постоянна:

$$g_m = (dI_{D \text{ sat}}/dV_G) = ZC_iv_s. \quad (49)$$

8.2.5. Температурная зависимость

Температура влияет на многие параметры и характеристики МОП-транзисторов, в особенности на подвижность, пороговое напряжение и форму подпорогового участка характеристик. При напряжениях на затворе, обеспечивающих сильную инверсию, подвижность инверсионных носителей при $T > 300$ К уменьшается с температурой по закону T^{-2} [24].

Рассмотрим температурную зависимость порогового напряжения (выражение (30)) в линейной области работы МОП-транзистора

$$V_T = \varphi_{ms} - \frac{Q_f}{C_i} + 2\psi_B + \frac{\sqrt{4\epsilon_s q N_A \psi_B}}{C_i}. \quad (50)$$

Поскольку разность работ выхода φ_{ms} и фиксированный заряд окисла не зависят от температуры, дифференцируя выражение (50) по температуре, получим [29]

$$\frac{dV_T}{dT} = \frac{d\psi_B}{dT} \left(2 + \frac{1}{C_i} \sqrt{\frac{\epsilon_s q N_A}{\psi_B}} \right), \quad (51)$$

где

$$\frac{d\psi_B}{dT} \approx \pm \frac{1}{T} \left[\frac{E_g(T=0)}{2q} - |\psi_B(T)| \right]. \quad (51a)$$

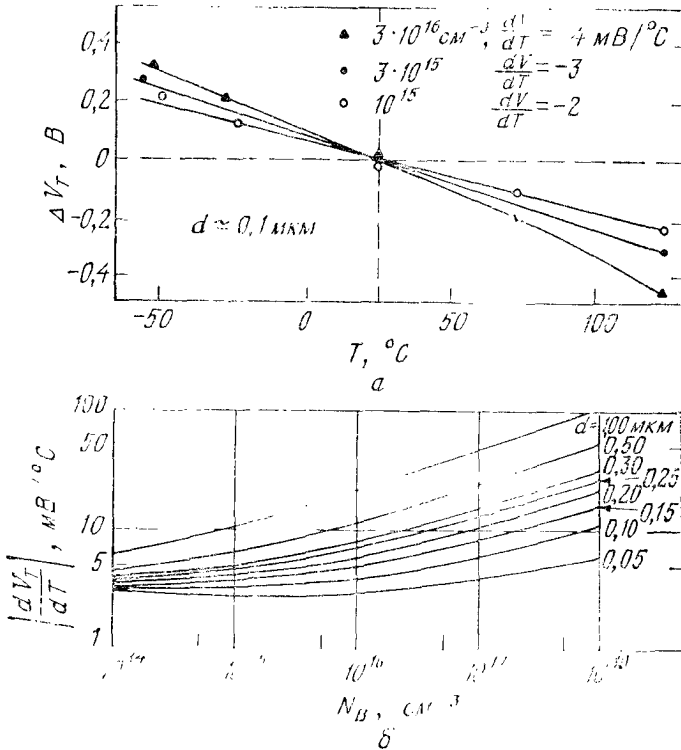


Рис. 16. Зависимость порогового напряжения от температуры (а) и зависимость dV_T/dT от уровня легирования в Si — SiO₂-структурах с различной толщиной окисного слоя (б) [29, 30].

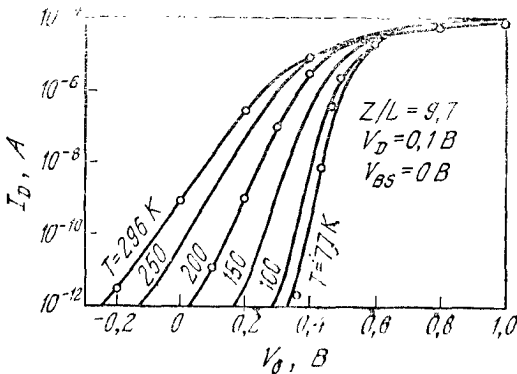


Рис. 17. Передаточные характеристики длинного прибора ($L = 9 \text{ мкм}$) при различных температурах [31].
 — теоретические; \circ — экспериментальные.

Типичные экспериментальные значения порогового напряжения в Si—SiO₂-системах в окрестности комнатной температуры приведены на рис. 16, а [29, 30]. Видно, что экспериментальные точки в этом температурном диапазоне лежат на почти прямых линиях. Значения соответствующих производных можно рассчитать, воспользовавшись формулой (51) при комнатной температуре. Результаты такого расчета для МОП-структур с различной толщиной слоя окисла в зависимости от уровня легирования подложки приведены на рис. 16, б. Отметим, что при заданной толщине слоя окисла величина производной dV_T/dT в основном увеличивается с ростом уровня легирования.

С понижением температуры характеристики МОП-транзисторов улучшаются (особенно в подпороговой области). На рис. 17 приведены передаточные (сток-затворные) характеристики длинноканального МОП-транзистора ($L = 9$ мкм), параметром которых является температура [31]. Отметим прежде всего, что при уменьшении температуры от 296 до 77 К пороговое напряжение возрастает от $V_T = 0,25$ до 0,5 В. Такое увеличение V_T подобно показанному на рис. 16, а. Однако наиболее важным улучшением является значительное уменьшение характерного перепада затворного напряжения S , требуемого для изменения подпорогового тока на порядок (от 80 мВ на декаду при 296 К до 22 мВ на декаду при 77 К). Такое (примерно четырехкратное) уменьшение S в основном обусловлено температурным множителем kT/q в формуле (42). Другими преимуществами работы МОП-транзисторов при 77 К являются повышенная подвижность, более высокие крутизна и проводимость в окрестности порога, пониженное рассеяние мощности, меньшие токи утечки $p-n$ -переходов и меньшие сопротивления металлических шин. Основным недостатком при этом является необходимость погружения приборов в специальную инертную охлаждающую среду (например, в жидкий азот) и связанные с этим дополнительные, чисто технические трудности.

8.2.6. Типы МОП-транзисторов

Идеальный МОП-транзистор представляет собой усилитель с бесконечным входным сопротивлением и генератором тока на выходе. Электрические свойства реальных приборов представляются более сложной эквивалентной схемой. На рис. 18 приведена одна из таких схем (для схемы включения с общим истоком) [32]. О дифференциальной крутизне g_m мы уже говорили выше. Входная проводимость G_{in} определяется током утечки через тонкий слой подзатворного окисла. Для термически выращенных слоев двуокиси кремния ток утечки между затвором и каналом очень мал (его плотность составляет 10^{-10} А·см⁻²), и

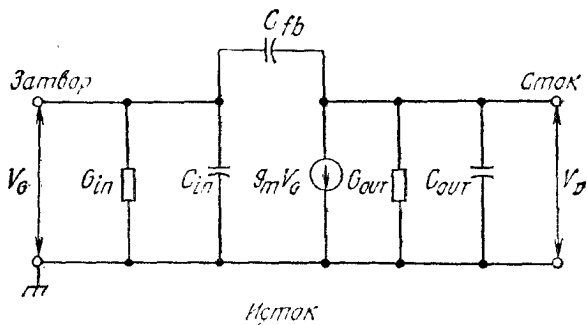


Рис. 18. Эквивалентная электрическая схема МОП-транзистора для включения с общим истоком [6].

обычно входной проводимостью пренебрегают. Входная емкость C_{in} равна $\partial Q_M / \partial V_G$, где Q_M — полный заряд на затворе [16]. В реальных приборах тонкий слой окисла и затвор частично перекрывают области стока и истока. Этот краевой эффект дает основной вклад в проходную емкость C_{fb} . Выходная проводимость G_{out} равна проводимости стокового $p-n$ -перехода. Выходная емкость C_{out} представляет главным образом емкости $p-n$ -переходов стока и истока, последовательно соединенные через объем полупроводниковой подложки. В линейном режиме МОП-транзистора максимальная рабочая частота прибора

$$f_m = \frac{\omega_m}{2\pi} = \frac{g_m}{2\pi C_{in}} \approx \frac{\mu_n V_D}{2\pi L^2}. \quad (52)$$

Этот результат следует из выражения (26) и того факта, что $C_{in} \approx ZLC_i$. В области насыщения (см. выражение (49))

$$f_m \approx \frac{v_s}{2\pi L}, \quad (53)$$

что соответствует времени пролета канала со скоростью насыщения

$$\tau = \frac{L}{v_s}. \quad (54)$$

Для $L = 1$ мкм и $v_s = 10^7$ см·с⁻¹ время пролета τ составляет всего лишь 10 пс. Однако в типичных кольцевых осцилляторах на МОП-транзисторах с длиной канала 1 мкм измеряемые времена задержки обычно на порядок больше. Следовательно, такие задержки обусловлены главным образом различными паразитными емкостями и сопротивлениями в окружении прибора.

В зависимости от типа инверсионного канала различают четыре основных типа МОП-транзисторов. Если при нулевом

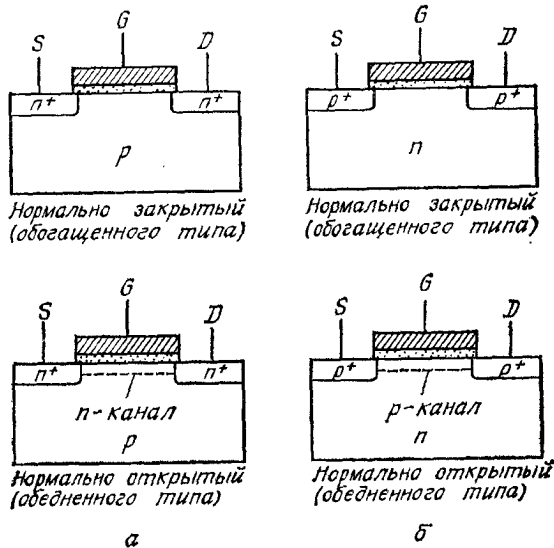


Рис. 19. Основные типы МОП-транзисторов.
 а — n-канальные; б — p-канальные.

Тип	Электрический символ	Выходные характеристики	Передачные характеристики
n-канал Нормально закрытый (обогащенного типа)			
n-канал Нормально открытый (обедненного типа)			
p-канал Нормально закрытый (обогащенного типа)			
p-канал Нормально открытый (обедненного типа)			

Рис. 20. Электрические символы, передачные и выходные характеристики четырех типов МОП-транзисторов [33].

напряжении на затворе проводимость канала очень мала и для образования инверсионного n -канала к затвору нужно приложить положительное напряжение, такой прибор называется нормально закрытым (обогащенного типа) n -канальным МОП-транзистором. Если же n -канал существует уже при нулевом смещении на затворе и, чтобы уменьшить его проводимость, к затвору следует приложить отрицательное обедняющее напряжение, такой прибор называют нормально открытым (обедненного типа) n -канальным МОП-транзистором. Оба этих типа n -канальных приборов показаны на рис. 19, а. Аналогичным образом классифицируются p -канальные МОП-транзисторы: нормально закрытые (обогащенного типа) и нормально открытые (обедненного типа) (рис. 19, б).

Электрические символы, передаточные и выходные характеристики этих четырех типов МОП-транзисторов приведены на рис. 20 [33]. Отметим, что для нормально закрытого n -канального прибора, чтобы получить заметный ток стока, нужно приложить к затвору достаточно большое положительное смещение, превышающее пороговое напряжение V_T . В нормально открытых приборах значительные токи могут протекать уже при $V_G = 0$, а смещение затвора увеличивает или уменьшает их величину. Сказанное выше можно легко распространить и на p -канальные приборы, соответствующим образом изменив полярность напряжений.

8.3. НЕОДНОРОДНО ЛЕГИРОВАННЫЕ И ПРИБОРЫ СО СКРЫТЫМ КАНАЛОМ

В разд. 8.2 концентрация легирующей примеси в подложке предпологалась постоянной. В реальных же приборах распределение примеси практически всегда неоднородно. Определенная неоднородность возникает даже в приборах, изготавливаемых на начально однородно легированных подложках, за счет перераспределения примесей в процессе термического окисления. Более того, в современной МОП-технологии специальное неоднородное легирование подложки, главным образом с помощью ионной имплантации, широко используется для улучшения рабочих параметров приборов. Так, например, в настоящее время ионная имплантация используется: во-первых, для самосмещения стока и истока с затвором, чтобы уменьшить смещение перекрытия; во-вторых, для подлегирования границы раздела $\text{Si}-\text{SiO}_2$ для корректировки пороговых напряжений; в-третьих, для создания имплантированных каналов на слаболегированных подложках, что предотвращает прокол (смыкание) стока с истоком; в четвертых, для создания приборов со скрытым каналом путем введения в приповерхностную область примесей противоположного типа.

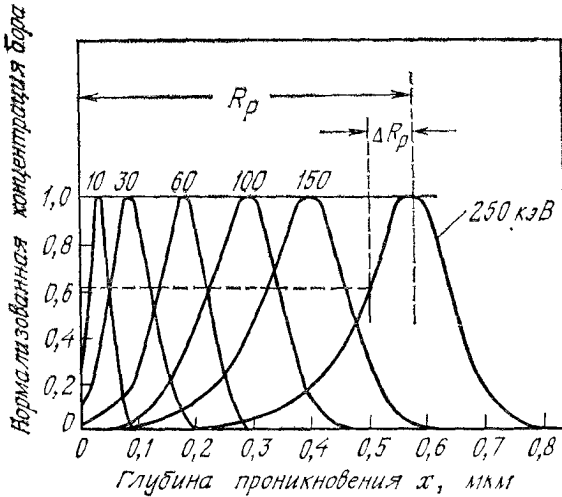


Рис. 21. Нормализованные распределения концентрации имплантированного бора по глубине кремниевой подложки при различных энергиях имплантации [34].

Профили концентрации примеси при ионной имплантации близки к форме гауссова распределения с максимумом на средней длине пробега R_p и стандартным отклонением ΔR_p :

$$N(x) = \frac{D}{\sqrt{2\pi}\Delta R_p} \exp \left[-\frac{(x - R_p)^2}{2(\Delta R_p)^2} \right]. \quad (55)$$

Здесь D , — доза ионов на единицу площади (рис. 21) [34]. Средняя длина пробега и стандартное отклонение увеличиваются с ростом энергии имплантируемых ионов [34]. Значения R_p и ΔR_p для бора, мышьяка и фосфора в кремнии приведены в гл. 2.

Ниже мы рассмотрим влияние неоднородного легирования канала на характеристики, в особенности на пороговые напряжения и характерный наклон подпорогового участка характеристик. Мы также рассмотрим МОП-транзисторы со скрытым каналом, которые можно использовать в качестве нагрузки (нормально открытый) в схеме инвертора и как нормально закрытые быстродействующие приборы.

8.3.1. Сдвиг порога

Чтобы рассчитать сдвиг порогового напряжения, обусловленный ионной имплантацией, рассмотрим сначала идеализированный ступенчатый профиль концентрации (рис. 22) [35]. Отметим, что первоначальный профиль концентрации имплантированных ионов заметно изменяется в процессе термического отжига (рис. 22).

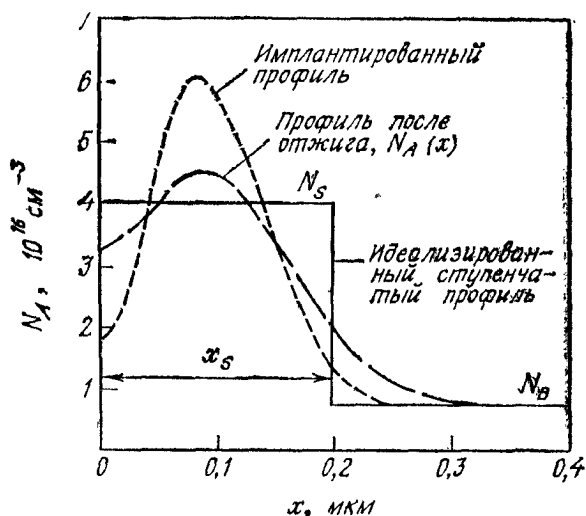


Рис. 22. Профиль легирования в области имплантации под затворным окислом [35].

Профиль после отжига можно аппроксимировать ступенчатой функцией с глубиной x_s , равной сумме средней длины пробега и стандартного отклонения первоначального профиля. Высота ступеньки $N_S - N_B$ определяется соотношением

$$(N_S - N_B) x_s = \int_0^{\infty} |N_A(x) - N_B| dx = D_I. \quad (56)$$

При $x_s \rightarrow 0$ (распределение в виде δ -функции на границе раздела) влияние имплантированных акцепторов (например, отрицательно заряженных ионизированных атомов бора) сводится к уменьшению положительного фиксированного заряда окисла на величину qD_I . Следовательно, в этом случае

$$V_T = V_{FB} + 2\psi_B + \frac{\sqrt{2\epsilon_s q N_B (2\psi_B + V_{BS})}}{C_i} + \frac{qD_I}{C_i}, \quad (57)$$

а соответствующий сдвиг порога

$$\Delta V_T = V_T(D_I) - V_T(D_I = 0) = \frac{qD_I}{C_i}. \quad (58)$$

Для широких распределений, таких, что $x_s \gg W_m$ (W_m — максимальная толщина обедненного слоя при сильной инверсии), приповерхностный слой полупроводника можно рассматривать как однородно легированную область с концентрацией N_S . В этом случае пороговое напряжение определяется выражением (31),

в котором N_A следует заменить на N_S и соответствующим образом скорректировать величину ψ_B .

В случае $x_s \ll W_m$ для определения толщины обедненного слоя и падения напряжения в окисле необходимо решать уравнение Пуассона для ступенчатого профиля концентрации примеси:

$$W_m = \sqrt{\frac{2\epsilon_s}{qN_B} \left[\psi_s + V_{BS} - \frac{qx_s^2}{2\epsilon_s} (N_S - N_B) \right]^{1/2}} \quad (59)$$

$$V_i = \frac{q}{C_i} \int_0^{W_m} N_A(x) dx = \\ = \frac{q}{C_i} [N_B W_m + (N_S - N_B) x_s] = \frac{q}{C_i} (N_B W_m + D_i), \quad (60)$$

где ψ_s — поверхностный изгиб зон при сильной инверсии, а V_i — падение напряжения в слое окисла. Отсюда пороговое напряжение

$$V_T = V_{FB} + \psi_s + \frac{\sqrt{2q\epsilon_s N_B}}{C_i} \left(\psi_s + V_{BS} - \frac{qx_s}{2\epsilon_s} D_i \right)^{1/2} + \frac{qD_i}{C_i}. \quad (61)$$

Легко видеть, что, когда $V_{BS} + \psi_s$ становится много больше $qx_s D_i / 2\epsilon_s$, выражение (61) стремится к приближенной формуле (57), соответствующей импульсному (типа δ -функции) профилю примеси на границе с окислом.

Вычисленные с помощью формул (57)—(61) зависимости порогового напряжения от смещения подложки V_{BS} приведены на рис. 23. Кривые *a* и *б* соответствуют случаю однородного легирования с $N_B = 7,5 \cdot 10^{15}$ и $4 \cdot 10^{16}$ см⁻³ соответственно. Эти зависимости аналогичны зависимостям, приведенным на рис. 9. Кривая *в* на рис. 23 получена в приближении импульсного профиля имплантированной примеси на границе раздела. В соответствии с формулой (57) она просто сдвинута относительно кривой *a* на величину $qD_i/C_i = 1,1$ В. Кривая *г* представляет точное решение уравнения Пуассона для ступенчатого профиля. Легко видеть, что при малых V_{BS} , когда $W_m < x_s = 0,2$ мкм, она близка к кривой *б*, соответствующей однородному распределению с высокой концентрацией примеси ($4 \cdot 10^{16}$ см⁻³). При больших V_{BS} , когда $W_m \gg x_s$, кривая *a* приближается к кривой *в*, соответствующей импульсной аппроксимации профиля концентрации имплантированных ионов.

Со схемотехнической точки зрения желательно иметь как можно меньшую чувствительность порогового напряжения МОП-транзистора к смещению подложки. Из рис. 23 видно, что для корректировки V_T следует использовать мелкую (с малыми x_s) имплантацию. Это позволяет при изменениях пороговых на-

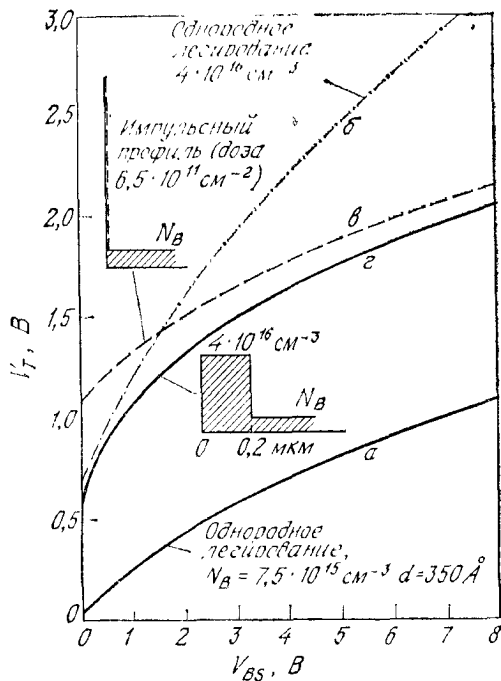


Рис. 23. Теоретическая зависимость чувствительности порога к смещению на подложке для различных профилей легирования [35].

пряжений сохранить исходную малую чувствительность V_T к смещению подложки.

Описанное выше приближение ступенчатого профиля представляет собой оценку первого порядка для сдвига порогового напряжения при ионной имплантации. Чтобы получить точный результат, необходимо решить уравнение Пуассона с точным профилем концентрации имплантированной примеси. Оказывается, что для сравнительно глубоких и узких распределений концентрации имплантированных ионов (рис. 24, а) лучшей, чем ступенчатая, является аппроксимация реального распределения импульсным профилем (типа δ -функции), центрированным на

$$x_c = \int_0^{x_I} [N_A(x) - N_B] x dx / D_I \quad (62)$$

и с дозой

$$D_I = \int_0^{x_I} [N_A(x) - N_B] dx. \quad (63)$$

Соответствующее этому импульсному профилю распределение электрического поля показано на рис. 24, б [36]. Площадь под

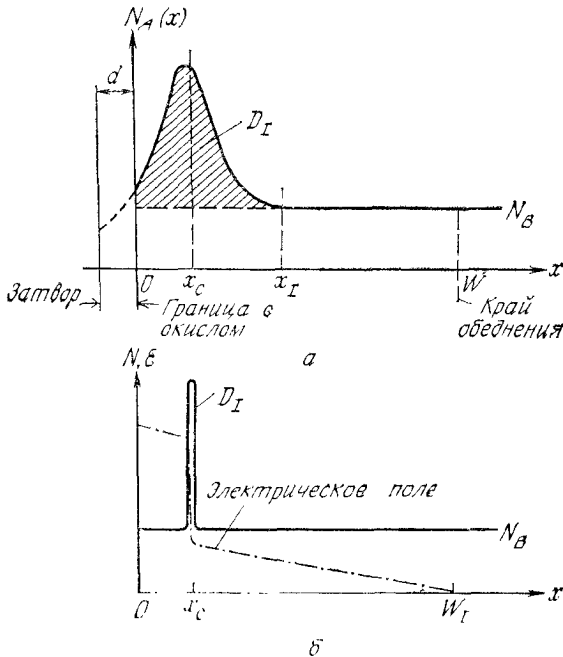


Рис. 24. Реальный профиль концентрации имплантированной примеси (а) и его импульсная аппроксимация (б) [36].

кривой равна поверхностному потенциалу ψ_s . Из уравнения Пуассона можно найти глубину обедненного слоя:

$$W_I = \sqrt{2L_D} [\beta\psi_s - D_I x_c / (N_B L_D^2) - 1]^{1/2}. \quad (64)$$

Напомним, что в однородно легированной подложке глубина обедненного слоя

$$W_U = \sqrt{2L_D} [\beta\psi_s - 1]^{1/2}. \quad (65)$$

Из сравнения выражений (64) и (65) следует, что ионная имплантация уменьшает глубину обедненного слоя.

Теперь с помощью выражения (64) найдем обусловленный имплантацией сдвиг порогового напряжения ΔV_T , определив его [11, 36] как разность между напряжениями на затворе, требуемыми для создания одного и того же инверсионного заряда Q_n в МОП-структуре с имплантированной примесью, и в аналогичной структуре на однородно легированной подложке. В однородно легированной МОП-структуре с напряжением на затворе V_{GU} и поверхностным потенциалом ψ_{SU} плотность инверсионного заряда равна

$$Q_n = Q_s - Q_B = C_i (V_{GU} - \psi_{SU}) - qN_B W_U. \quad (66)$$

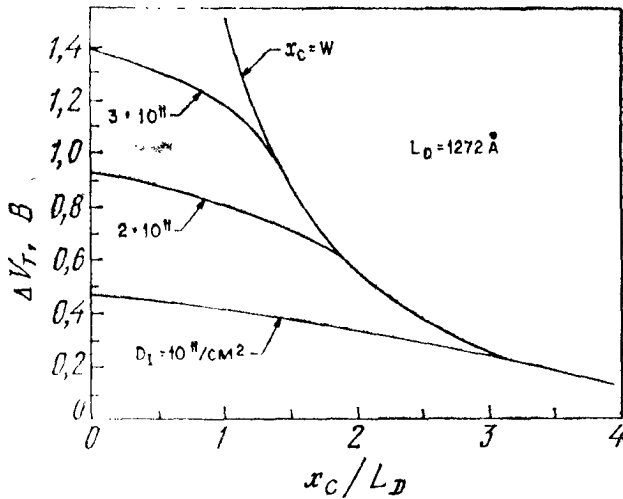


Рис. 25. Зависимость сдвига порогового напряжения от центрады распределения имплантированной примеси при различных дозах ионов [36].

В случае имплантации при напряжении на затворе V_{GI} и изгибе зон ψ_{Si}

$$Q_n = C_i (V_{GI} - \psi_{Si}) - qN_B W_l - qD_l. \quad (67)$$

Приравняв выражения (66) и (67), для сдвига порогового напряжения ΔV_T получим

$$\Delta V_T = V_{GI} - V_{GU} = \psi_{Si} - \psi_{SU} - qN_B (W_U - W_l) / C_i + qD_l / C_i. \quad (68)$$

На практике при достаточно большом Q_n (в режиме сильной инверсии) поверхностные потенциалы ψ_{Si} и ψ_{SU} различаются на величину, которая меньше kT/q , так что в выражении (68) можно пренебречь этой разностью. Подставив в выражение (68) выражения (64) и (65), найдем

$$\Delta V_T = \frac{qD_l}{C_i} - \frac{V \sqrt{2} q L_D N_B}{C_i} \left[(\beta \psi_{SU} - 1)^{1/2} - \left(\beta \psi_{SU} - \frac{D_l x_c}{N_B L_D^2} - 1 \right)^{1/2} \right]. \quad (69)$$

Как уже говорилось выше, изгиб зон ψ_{SU} при сильной инверсии примерно равен $2\psi_B + V_{BS}$, и для расчета зависимости сдвига порога от дозы D_l и центрады x_c имплантированной примеси остается лишь подставить $\psi_{SU} = 2\psi_B + V_{BS}$ в выражение (69).

Полученные таким образом зависимости ΔV_T от центрады x_c при различных дозах [36] приведены на рис. 25. При $x_c = 0$, т. е. когда δ -функция имплантации расположена на границе

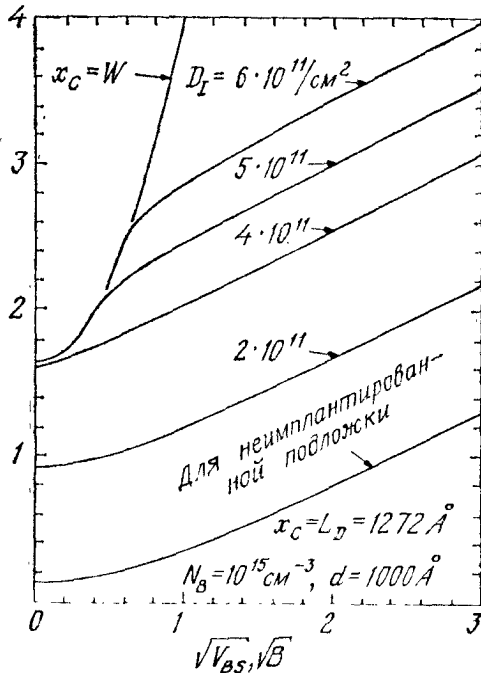


Рис. 26. Зависимость порогового напряжения от смещения на подложке при различных дозах имплантированных ионов [36].

раздела Si—SiO₂, $\Delta V_T = qD_I/C_i$, поскольку слагаемое в квадратных скобках в формуле (69) равно нулю. С ростом x_c величина ΔV_T уменьшается. Наконец, x_c становится равным толщине слоя обеднения W , и край обедненной области совпадает с имплантированным слоем $W = x_c$. Значение дозы, при которой $W = x_c$, можно получить из уравнения (64):

$$D_I(x_c = W) = N_B (W_U^2 - x_c^2) / 2x_c. \quad (70)$$

На рис. 26 приведена зависимость порогового напряжения от смещения на подложке для импульсного профиля имплантации с $x_c = L_D = 1272 \text{ \AA}$. При $D_I < 4 \cdot 10^{11} \text{ см}^{-2}$ имплантация приводит к параллельному сдвигу зависимости $V_T(V_{BS})$ относительно исходной кривой, соответствующей однородно легированной подложке (неимплантированной). При таких малых D_I главную роль в уравнении (69) играет первое слагаемое qD_I/C_i . При $D > 4 \cdot 10^{11} \text{ см}^{-2}$ край обедненного слоя фиксируется в плоскости имплантации. В этом случае с ростом смещения V_{BS} изменения порогового напряжения сначала описываются кривой, соответ-

ствующей условно $x_c = W$. Однако при больших смещениях на подложке, таких, что $(2\psi_B + V_{BS}) > D_I x_c / \beta N_B L_B^2$, край обедненного слоя «отрывается» от плоскости имплантации x_c , первое слагаемое в выражении (69) опять становится доминирующим, а зависимость $V_T(V_{BS})$ становится параллельной исходной («неимплантированной») кривой.

8.3.2. Наклон подпороговых характеристик

Как уже говорилось выше, в подпороговой области прибора с однородно легированной подложкой логарифм тока стока $\ln I_D$ увеличивается почти линейно с ростом напряжения на затворе. Обратный наклон этой прямой определяет, на сколько следует уменьшить напряжение на затворе МОП-транзистора, чтобы уменьшить ток прибора до требуемого уровня. Поэтому подпороговую область, соответствующую закрытому состоянию МОП-транзистора, обычно характеризуют напряжением S , необходимым для уменьшения тока стока на порядок величины. В случае однородного легирования [19]

$$S \approx \frac{kT}{q} \ln 10 (1 + C_D/C_I) \left/ \left[1 - \left(\frac{2}{a^2} \right) \left(\frac{C_D}{C_I} \right)^2 \right] \right., \quad (71)$$

где C_D — удельная дифференциальная емкость обедненного слоя:

$$C_D = \epsilon_s/W_I. \quad (72)$$

Влияние имплантации на подпороговое напряжение S определяется главным образом изменением (в соответствии с выражением [64]) дифференциальной емкости обедненного слоя. На рис. 27 приведены зависимости S от центриды имплантированного профиля примеси для различных доз D_I . При $x_c = 0$ исходное значение S , как и в однородно легированной подложке, определяется только концентрацией примеси и толщиной слоя окисла. При фиксированной дозе D_I характерное напряжение S сначала увеличивается с ростом x_c , поскольку при этом край обедненного слоя приближается к границе раздела $\text{Si}-\text{SiO}_2$ и дифференциальная емкость C_D растет. Затем край слоя обеднения попадает в плоскость имплантации, и теперь при росте x_c он уже удаляется от границы с окислом вместе с плоскостью имплантации. При этом, естественно, C_D и S уменьшаются до исходных значений, соответствующих неимплантированной подложке. Излом кривых на рис. 27 соответствует ситуации, когда $x_c = W$. В реальных условиях острота пика, естественно, сглаживается за счет ненулевой толщины слоя имплантированной примеси. Из данных, приведенных на рис. 27, следует, что с точки зрения поддержания высокой крутизны участка выключения МОП-транзистора (малые значения S) предпочтительна мелкая (с малыми x_c) имплантация.

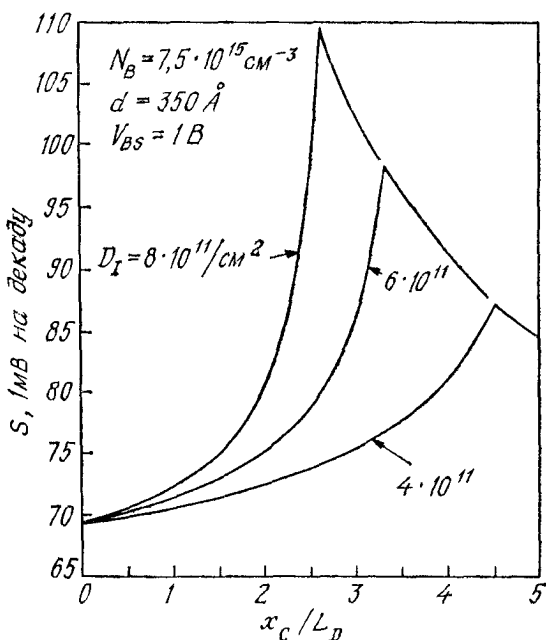


Рис. 27. Зависимость характерного подпорогового напряжения S от центрады распределения имплантированных ионов при различных дозах [19].

8.3.3. Приборы со скрытым каналом

Когда тип имплантированной примеси противоположен типу примесных атомов в исходной подложке (например, при имплантации мышьяка в p -кремний), в приповерхностной части подложки может сформироваться так называемый скрытый канал. На рис. 28, а показан условный поперечный разрез такого n -канального МОП-транзистора со скрытым каналом, где в отличие от обычного МОП-транзистора проводящий канал заглублен в объем полупроводника от границы раздела $\text{Si}-\text{SiO}_2$. Истинный профиль легирования приповерхностной области подложки такого прибора и его ступенчатая аппроксимация приведены на рис. 28, б. Ясно, что в окрестности металлургического $p-n$ -перехода ($x = x_j$) образуется обедненная область, локальная толщина которой определяется величиной напряжений, приложенных к структуре. Напряжение на затворе, кроме того, модулирует локальную толщину приповерхностной подзатворной области обеднения.

МОП-транзисторы со скрытым каналом могут быть нормально открытыми и нормально закрытыми приборами в зависимости от глубины и уровня легирования канала. Рассмотрим сначала

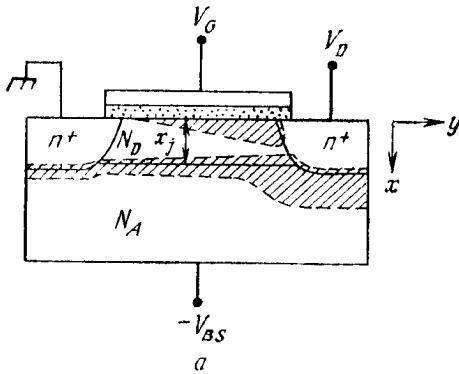
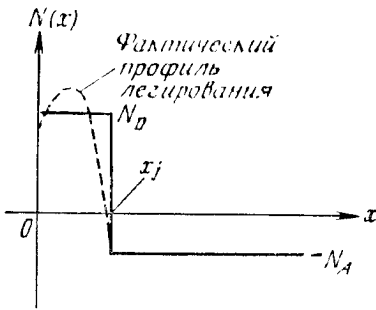


Рис. 28. МОП-транзистор со скрытым каналом (а) и профиль легирования канала (б).



б

работу нормально открытого прибора. Его условное поперечное сечение в области вблизи истока [37] показано на рис. 29 (вставка). Когда напряжение на затворе увеличивается (в рассматриваемом случае оно отрицательно), ширина скрытого канала уменьшается. Наконец, при $V_G = V_P$ величина x_b становится равной нулю и происходит отсечка канала. Напряжение отсечки канала V_P рассчитывается с помощью уравнения Пуассона и закона сохранения электрической индукции (Гаусса) на границе Si—SiO₂ в качестве граничного условия. Для ступенчатого профиля (рис. 28, б) в предположении $N_D \gg N_A$ получаем

$$V_P - V_{KB} = -\frac{qN_D x_j}{C_i} \left(1 + \frac{C_i}{2C_D}\right) + \frac{\sqrt{2\epsilon_s q N_A}}{C_i} \left(1 + \frac{C_i}{C_D}\right) (V_{bi} + V_{BS})^{1/2}, \quad (73)$$

где $C_D = \epsilon_s/x_j$, а V_{bi} — контактная разность (встроенный потенциал) p — n -перехода. На рис. 29 приведены рассчитанные по этой формуле зависимости напряжения отсечки канала V_P от

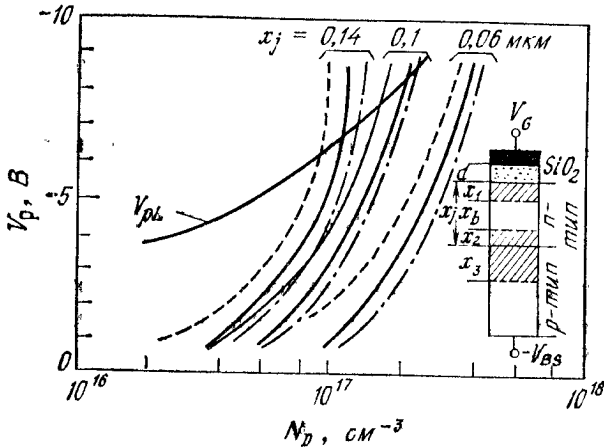


Рис. 29. Зависимость напряжения отсечки от концентрации доноров при различных значениях глубины канала и уровня легирования подложки. На вставке показана толщина обедненных слоев в области у истока [37].

--- $N_A = 10^{16} \text{ см}^{-3}$, — $N_A = 3 \cdot 10^{16} \text{ см}^{-3}$, — — $N_A = 5 \cdot 10^{16} \text{ см}^{-3}$.

концентрации доноров в приповерхностном слое при нескольких значениях глубины перехода x_1 для случая $V_{DS} = 0$. Видно, что при фиксированных x_1 напряжение отсечки канала увеличивается с ростом концентрации. Однако при слишком больших N_D (либо x_1) полностью перекрыть проводящий канал напряжением на затворе невозможно, поскольку в конечном счете увеличение x_1 с ростом V_G будет стабилизировано (на уровне W_m) образованием дырочного инверсионного слоя на границе раздела Si—SiO₂. Соответствующее предельное значение напряжения отсечки канала равно

$$V_{PL} = V_{FB} + 2\psi_B + \sqrt{2\varepsilon_s q N_D (2\psi_B) / C_i}. \quad (74)$$

Предельное значение глубины перехода x_{iL} , при превышении которого канал уже не перекрывается полностью ($x_b \neq 0$), легко вычисляется по формуле

$$x_{iL} = \frac{l}{N_D} \sqrt{\frac{2\varepsilon_s N_A}{q}} (\sqrt{2V_{bi} N_D / N_A} + \sqrt{V_{bi} + V_{BS}}). \quad (75)$$

Выходные (стоковые) характеристики МОП-транзистора со скрытым каналом аналогичны характеристикам полевого транзистора с p — n -переходом в качестве затвора (гл. 6). Они также линейны при малых V_D и насыщаются при больших стоковых напряжениях. Согласно работе [38], ток стока транзистора со скрытым каналом

$$I_D = Z (\mu D_i - \mu Q_s - \mu Q_B) \frac{dV}{dy}, \quad (76)$$

где

$$D_I = \int_0^{x_1} [N(x) - N_A] dx \approx (N_D - N_A) x_1 \quad (77)$$

— плотность имплантированных центров на единицу площади, а

$$Q_s = -\bar{C} [V_G - V_{FB} - V(y)] \quad (78)$$

— плотность заряда (приповерхностного слоя обеднения либо аккумуляции) на единицу площади границы раздела, \bar{C} — усредненная емкость затвора,

$$Q_B = \sqrt{2\epsilon_s q N_A' [V_{bi} + V_{BS} + V(y)]^{1/2}} \quad (79)$$

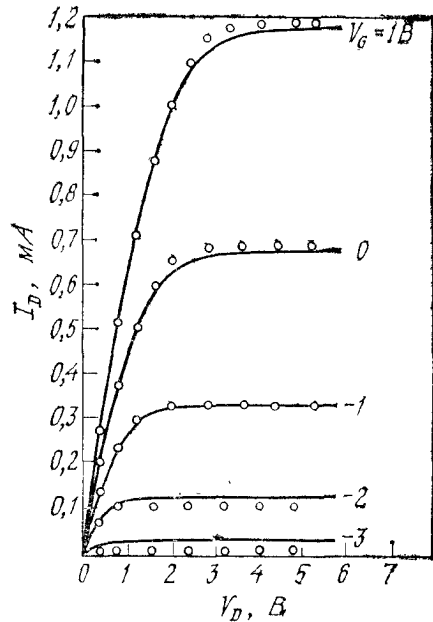
— заряд в обедненном слое p - n -перехода, а $N_A' = N_D N_A / (N_D + N_A)$. В нормально обедненном режиме [$(V_G - V_{FB}) \leq 0$] ток стока протекает только в скрытом объемном канале, поэтому в выражение (76) следует подставлять значение объемной подвижности μ_B . Подставив выражения (77)—(79) в выражение (76) и проинтегрировав от 0 до L , получим

$$I_D = \frac{\mu_B^2}{L} \left\{ D_I V_D + \bar{C} \left[(V_G - V_{FB}) V_D - \frac{1}{2} V_D^2 \right] - \right. \\ \left. - \frac{2}{3} \sqrt{2\epsilon_s q N_A'} [(V_{bi} + V_{BS} + V_D)^{3/2} - (V_{bi} + V_{BS})^{3/2}] \right\}. \quad (80)$$

Рассчитанные по этой формуле выходные характеристики и экспериментальные зависимости I_D (V_D) хорошо соответствуют друг другу (рис. 30). Отметим, что, поскольку рассматриваемый МОП-транзистор является нормально открытым прибором, в нем уже при $V_G = 0$ протекают довольно большие токи. При изменении напряжения на затворе ток стока также изменяется.

На рис. 31, а показаны профиль легирования и толщины обедненных слоев нормально открытого n -канального МОП-транзистора со скрытым каналом в режиме, предшествующем отсечке канала. На рис. 31, б и в приведены распределения зарядов и подвижных электронов в момент отсечки ($x_1 + x_2 = x_1$, $V_G = V_P$). Закрытое состояние прибора при $V_G > V_P$ полностью соответствует подпороговому режиму обычного МОП-транзистора обогащенного типа. Главной компонентой тока при этом является диффузионная компонента, а сам подпороговый ток зависит экспоненциально от напряжения затвора. Характеризующий наклон подпорогового (экспоненциального) участка параметр S

Рис. 30. Экспериментальные (—) и рассчитанные (○) стоковые характеристики нормально открытого МОП-транзистора [38].
 $V_{FB} = -1,25$ В; $V_{BS} = 0$ В.



определяется соотношением емкостей, показанных на рис. 31, а [39]:

$$S = \frac{kT}{q} \ln 10 [(C_1 + C_2)/C_1] =$$

$$= \frac{kT}{q} \ln 10 \left[1 + \frac{\epsilon_i x_1 + \epsilon_s d}{\epsilon_i (x_2 + x_3)} \right]. \quad (81)$$

Очевидно, что рассматриваемый МОП-транзистор со скрытым каналом будет нормально закрытым прибором, когда обедненные слои (приповерхностный и $p-n$ -перехода) будут сомкнуты уже при нулевом напряжении на затворе $V_G = 0$. Для этого необходимо, чтобы напряжение отсечки V_P , которое теперь можно по аналогии назвать пороговым напряжением V_T :

$$V_{FB} - V_T = \frac{qN_D(x_j - x_1)}{C_i} + \frac{qN_D(x_j - x_1)^2}{2\epsilon_s}, \quad (82)$$

было бы больше нуля, что возможно в структурах с большими напряжениями плоских зон V_{FB} . На практике для этого применяются затворы из сильнолегированного бором поликремния или из металлов с большой работой выхода.

По сравнению с поверхностными МОП-транзисторами в приборах со скрытым каналом можно ожидать более высоких значений подвижности, поскольку движущиеся в объеме носители не

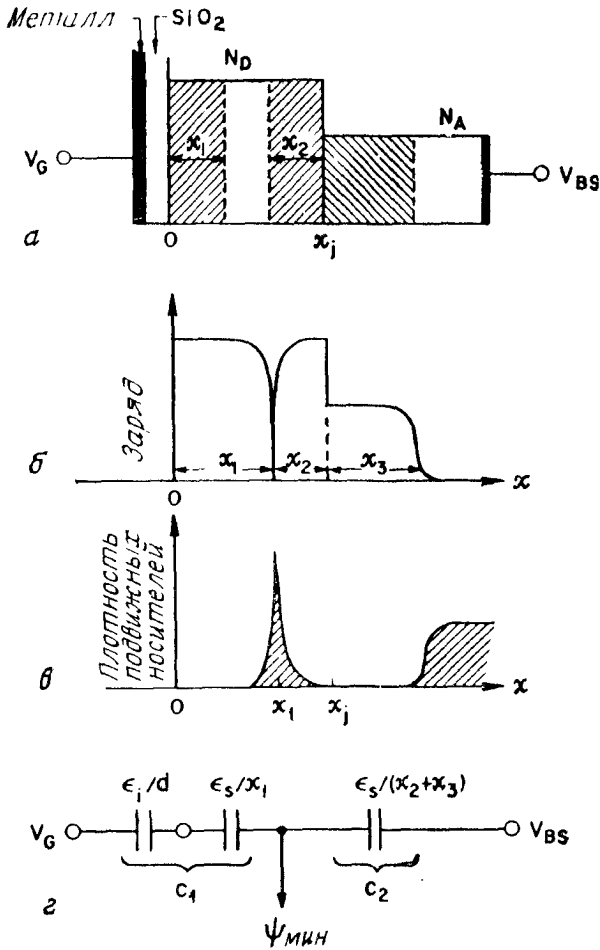


Рис. 31. МОП-транзистор со скрытым каналом в режиме, предшествующем отсечке канала [39]

а — профиль легирования; б — плотность заряда; в — плотность подвижных носителей; г — эквивалентные емкости.

рассеиваются на неоднородностях границы раздела с диэлектриком. Эти соображения подтверждаются результатами соответствующих измерений (рис. 32, а), из которых видно, что в приборах со скрытым каналом подвижность примерно на 50 % выше, чем в обычных МОП-транзисторах. На рис. 32, б для сравнения приведены зависимости порогового напряжения соответствующих приборов от длины канала L . В приборах с длинным каналом напряжение V_T не зависит от длины L . При малых L , однако,

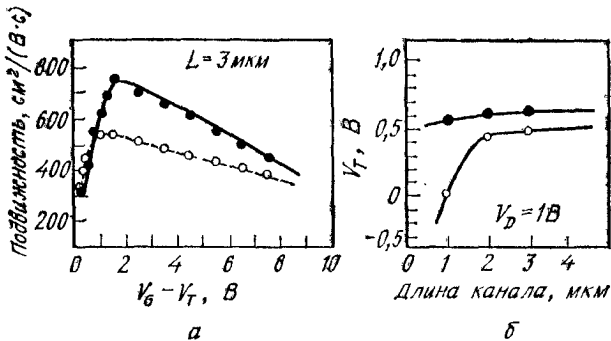


Рис. 32. Характеристики МОП-транзисторов с поверхностным (○) и скрытым (●) каналами [40].

а — зависимость подвижности от напряжения затвора; б — зависимость порогового напряжения от длины канала.

происходит нежелательное уменьшение V_T с длиной канала, обусловленное так называемыми короткоканальными эффектами, которые мы рассмотрим в разд. 8.4. Однако уже теперь можно отметить большую «устойчивость» МОП-транзисторов со скрытым каналом к влиянию короткоканальных эффектов по сравнению с обычными приборами. Поэтому нормально закрытые МОП-транзисторы со скрытым каналом весьма полезны для высокочастотных применений.

8.4. КОРОТКОКАНАЛЬНЫЕ ЭФФЕКТЫ

С 1959 г., который можно считать «началом эры интегральных схем», до настоящего времени характерный размер отдельного полупроводникового прибора уменьшился по крайней мере на два порядка. Можно полагать, что такое сокращение будет продолжаться и в обозримом будущем (рис. 1). Однако по мере уменьшения длины канала МОП-транзистора свойства последнего начинают резко отличаться от свойств обычных длинноканальных приборов, рассмотренных в разд. 8.2. Эти отклонения — так называемые короткоканальные эффекты — обусловлены существенно двумерным характером распределения электрических полей в активной области и сравнительно высокими абсолютными значениями напряженности полей.

Если при неизменной концентрации легирующей примеси в подложке сокращать длину канала прибора, в конце концов она станет величиной порядка толщины обедненных слоев $p-n$ -переходов стока и истока. При этом распределение потенциала в канале будет равным образом определяться поперечным полем \mathcal{E}_x , обусловленным напряжениями на затворе и подложке, и продольным полем \mathcal{E}_y , инициированным напряжением смещения стока

транзистора. Иными словами, распределение потенциала в таком короткоканальном приборе имеет двумерный характер, и для его описания уже нельзя использовать приближение плавного канала, которое предполагает, что $\mathcal{E}_x \gg \mathcal{E}_y$. Двумерный характер распределения потенциала существенно изменяет подпороговый участок характеристики прибора, обуславливает нежелательную зависимость порогового напряжения от длины канала и напряжений смещения на электродах, уменьшает выходное сопротивление, препятствуя отсечке канала.

При повышенных значениях электрических полей, характерных для короткоканальных приборов, становится важной полевая зависимость подвижности, которая в конечном итоге приводит к насыщению дрейфовой скорости (разд. 8.2). При еще больших полях в окрестности стокового перехода начинается ударная ионизация, становится существенной дополнительная проводимость по подложке и происходит так называемое включение паразитного биполярного транзистора. Высокие электрические поля приводят также к разогреву носителей и соответствующей инжекции горячих носителей в окисел. Такая зарядка окисла обуславливает сдвиг порога, дрейф характеристик и ухудшение крутизны прибора.

Все перечисленные короткоканальные эффекты усложняют работу прибора и ухудшают его рабочие характеристики. Следовательно, в практике конструирования следует стремиться к устранению или минимизации этих эффектов, чтобы «физически» короткоканальный прибор был «электрически» подобен длинноканальному.

8.4.1. Подпороговые токи

На рис. 33 приведены экспериментальные передаточные (т. е. $\ln I_D$ от V_G) характеристики нескольких МОП-транзисторов, изготовленных по стандартной n -канальной технологии на пластинах p -кремния с ориентацией (100). После выращивания подзатворного окисла заданной толщины с помощью рентгеновской литографии на пластинах формировались поликремниевые затворы длиной от 1 до 10 мкм и одинаковой шириной 70 мкм. Переходы стока и истока были получены имплантацией ионов мышьяка с последующим отжигом. Глубина переходов варьировалась от 0,25 до 1,5 мкм в зависимости от энергии имплантации и режима отжига. Для контактной металлизации был использован алюминий.

В МОП-транзисторах, характеристики которых приведены на рис. 33, *а*, толщина окисла составляла 130 Å, концентрация легирующей примеси в подложке 10^{15} см⁻³, а глубина переходов 0,33 мкм. Прибор с каналом длиной 7 мкм демонстрирует типичное

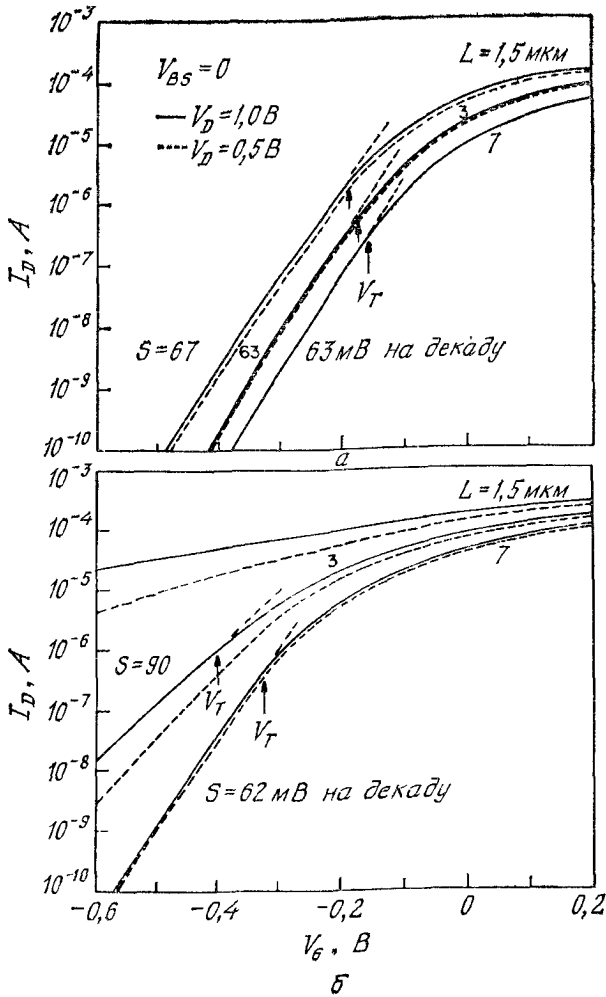


Рис. 33. Подпороговые характеристики МОП-транзисторов с различной длиной канала.

а — $N_B = 10^{15} \text{ см}^{-3}$; б — $N_B = 10^{14} \text{ см}^{-3}$.

длинноканальное поведение. Его подпороговая характеристика не зависит от напряжения стока, как и должно быть при $V_D > > 3kT/q$ согласно формуле (40). В приборе с $L = 3 \text{ мкм}$ изменение V_D немного сдвигает подпороговый участок. Несколько больший сдвиг при изменении напряжения стока от $V_D = 0,5$ до $1,0 \text{ В}$ наблюдается в транзисторе с самым коротким ($L = 1,5 \text{ мкм}$) каналом. Соответственно изменяются и пороговые напряжения V_T ,

определенные здесь как точка (показанная на рис. 33 стрелками), в которой I — V -зависимость начинает отклоняться от прямой линии. При увеличении V_D уменьшается также крутизна подпорогового участка (S увеличивается).

Иные характеристики имеют приборы с идентичными (за исключением концентрации примеси в подложке $N_B = 10^{14} \text{ см}^{-3}$) параметрами (рис. 33, б). Здесь небольшой сдвиг характеристик при изменении V_D наблюдается уже в приборе с $L = 7$ мкм. В МОП-транзисторе с каналом длиной 3 мкм подпороговые токи и соответствующие S существенно увеличены. И наконец, характеристики прибора с самым коротким каналом ($L = 1,5$ мкм) кардинально отличаются от характеристик обычных длинноканальных приборов (прибор «не выключается»). Критерий, разграничивающий длинно- и короткоканальные приборы, можно попытаться определить, используя одно из двух характерных свойств обычного длинноканального МОП-транзистора: 1) обратно пропорциональную зависимость тока стока от длины канала $I_D \sim 1/L$; 2) независимость подпорогового тока обычного прибора от напряжения стока при $V_D > 3kT/q$. На рис. 34 приведены зависимости I_D и $\Delta I_D/I_D$ от $1/L$, где I_D — ток стока транзисторов при напряжении на затворе V_G , равном пороговому напряже-

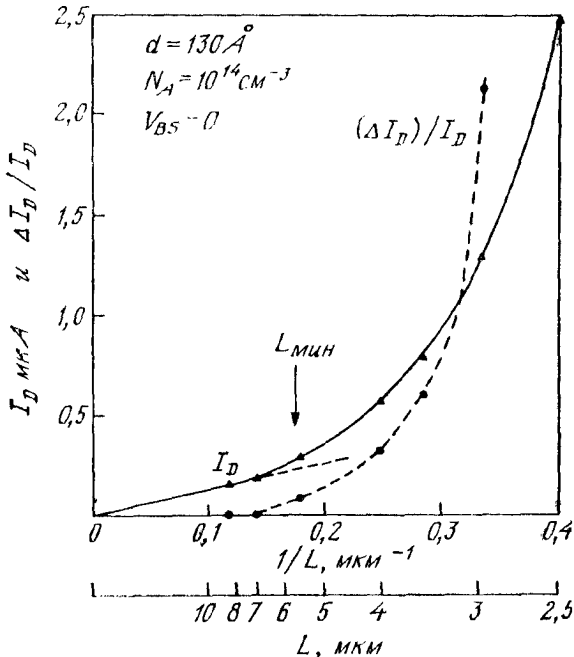


Рис. 34. Зависимость тока стока и отношения $\Delta I_D/I_D$ от величины $1/L$.

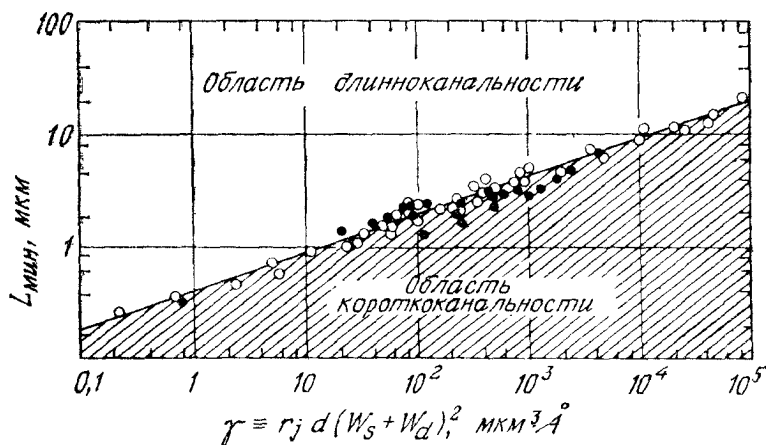


Рис. 35. Зависимость минимальной длинноканальной длины L_{\min} от параметра $\gamma \equiv r_j d (W_s + W_d)^2$ [41].

нию V_T , а $\Delta I_D/I_D$ — относительная разность токов при том же $V_G = V_T$ и двух различных напряжениях стока V_D . Началом короткоканального поведения будем считать момент, когда отклонение тока стока от длинноканальной зависимости $I_D \sim 1/L$ составляет 10 % или когда относительная разность $\Delta I_D/I_D$ равна 0,1.

Результаты большого количества измерений, выполненных в МОП-транзисторах, параметры которых варьировались в широких пределах (толщина подзатворного окисла 100—1000 Å, концентрация примеси в подложке 10^{14} — 10^{17} см $^{-3}$, глубина переходов 0,18—1,5 мкм при напряжениях на стоке до 5 В), обобщаются довольно простым эмпирическим соотношением [41]

$$L_{\min} = 0,4 [r_j d (W_s + W_d)^2]^{1/3} \equiv 0,4 (\gamma)^{1/3}. \quad (83)$$

Здесь L_{\min} — минимальная длина канала, при которой подпороговый участок еще сохраняет длинноканальный характер, мкм; r_j — глубина переходов, мкм; d — толщина слоя окисла, Å; $W_s + W_d$ — сумма толщин обедненных слоев стока и истока, вычисленная в приближении одномерного резкого p — n -перехода:

$$W_D = \sqrt{\frac{2\epsilon_s}{qN_A} (V_D + V_{bt} + V_{BS})} \quad [\text{мкм}], \quad (84)$$

где V_{bt} — контактная разность перехода, а V_{BS} — смещение на подложке. При нулевом смещении стока ($V_D = 0$) толщина W_D равна W_s .

На рис. 35 приведены для сравнения результаты соответствующих экспериментальных измерений (—●—) и вычислений по

эмпирическому соотношению (83). Здесь также приведены результаты двумерных машинных расчетов зависимости $L_{\text{мин}}$ от γ ($-\infty$ -), полученные на основе сформулированных выше критериев короткоканального поведения. Отметим, что в самых худших случаях экспериментальные и расчетные результаты отличаются от эмпирического соотношения (83) не более чем на 20 %, а для основного числа точек наблюдается значительно лучшее соответствие. Следовательно, эмпирическую формулу (83) можно использовать в качестве главного ориентира в практике миниатюризации МОП-транзисторов. Все МОП-транзисторы с параметрами, которым соответствует точка в заштрихованной области рис. 35, по своим электрическим свойствам будут короткоканальными приборами. МОП-транзисторы с параметрами, попадающими в незаштрихованную область рис. 35, в электрическом смысле являются длинноканальными. Так, например, свойства прибора с $L = 10$ мкм, у которого $\gamma = 10^5$ мкм³ Å, будут короткоканальными, в то время как МОП-транзистор с $L = 0,5$ мкм, но с $\gamma = 1$ мкм³ Å будет вести себя как длинноканальный прибор.

В первом приближении отклонения от длинноканального поведения можно проанализировать, воспользовавшись принципом электронейтральности (рис. 36, а) [42]

$$Q'_M + Q'_O + Q'_n + Q'_B = 0, \quad (85)$$

где Q'_M — полный заряд на затворе, Q'_O — полный эффективный (отнесенный к границе раздела Si—SiO₂) заряд окисла, Q'_n — полный заряд инверсионного слоя, а Q'_B — полный эффективный заряд ионизированных примесей в области обеднения под затвором. Поделив на $C_i A$, это соотношение можно записать в виде

$$V_G = V_{FB} + \psi_s + Q'_B/C_i A, \quad (86)$$

где V_{FB} — напряжение плоских зон, ψ_s — поверхностный потенциал и A — площадь. Пороговое напряжение V_T получим, подставив в выражение (86) $\psi_s = 2\phi_n$:

$$V_T = V_{FB} - 2\phi_n - Q'_B/C_i A. \quad (87)$$

Для длинноканального прибора $Q'_B = qN_A W A$, где W — толщина обедненного слоя. В приборах с коротким каналом влияние Q'_B на пороговое напряжение уменьшается, поскольку в крайних областях канала вблизи стока и истока часть силовых линий с заряженных примесных атомов «уходит в сторону», закорачиваясь на высоколегированных n^+ -областях, а не на затворе, как это показано на рис. 36, а для $V_D = 0$. Естественно, что при $V_D > 0$ обедненная область вблизи стока шире, чем вблизи истока (рис. 36, б). Отметим, что с учетом влияния поперечных электрических полей на распределение потенциала в при-

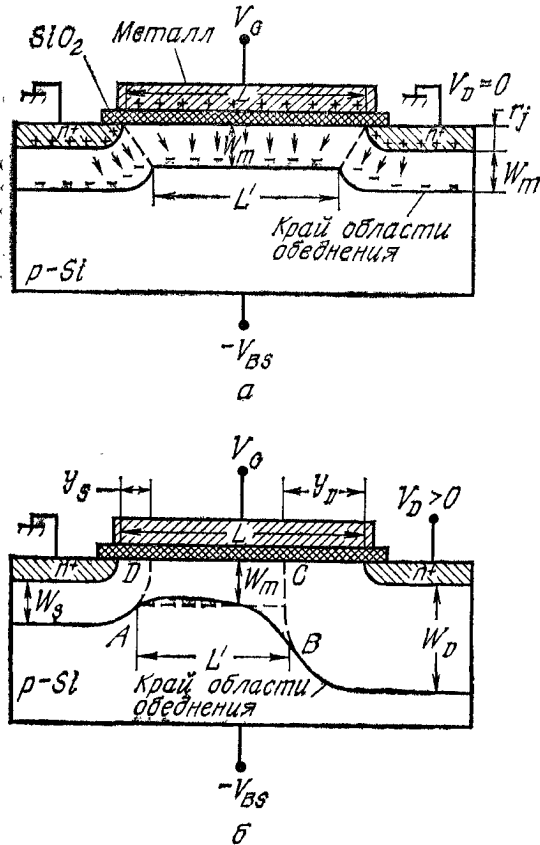


Рис. 36. Модель принципа электронейтральности при рассмотрении короткоканальных эффектов [42].

а) $V_D = 0$; б) $V_D > 0$.

поверхностном слое «горизонтальные» толщины обеднения y_S и y_D оказываются несколько меньше «вертикальных» глубин обеднения W_S и W_D соответственно.

В результате уменьшения эффективного заряда обедненного слоя Q_B поверхностный потенциал ψ_s для данного значения напряжения на затворе увеличивается, что приводит к соответствующему увеличению подпороговых токов (выражения (37) и (38)). Поверхностный потенциал можно записать следующим образом [43]:

$$V_G - V_{FB} = \psi_s + \frac{1}{C_t} \sqrt{q\epsilon_s N_A (\psi_s + V_{BS})/2} \left(1 + \frac{L - W_D - W_S}{L - y_D - y_S} \right), \quad (88)$$

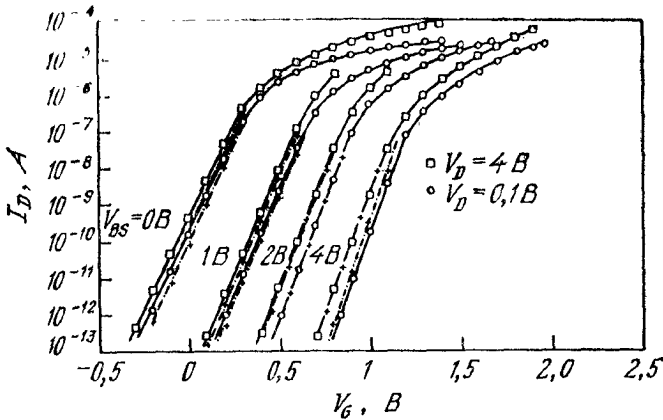


Рис. 37. Подпороговые токи в короткоканальном МОП-транзисторе с $d = 530 \text{ \AA}$, $N_A = 1,2 \cdot 10^{16} \text{ см}^{-3}$, $L = 2,56 \text{ мкм}$ и $Z = 21,5 \text{ мкм}$ [22].
 — □ — по формуле (90); — ○ — двумерное моделирование.

где W_D и W_S определяются выражением (84), а

$$y_S \approx \sqrt{\frac{2\varepsilon_s}{qN_A} (V_{bi} - \psi_s)} \quad (89a)$$

и

$$y_D \approx \sqrt{\frac{2\varepsilon_s}{qN_A} (V_{bi} - \psi_s - V_D)}. \quad (89b)$$

Для подпорогового тока будем иметь

$$I_D = \mu_n \left(\frac{Z}{L - y_S - y_D} \right) \frac{aC_i}{2\beta^2} \left(\frac{n_i}{N_A} \right)^2 (1 - e^{-\beta V_D}) e^{\beta \psi_s} (\beta \psi_s)^{1/2}. \quad (90)$$

Это выражение отличается от выражения (40) только тем, что в нем фигурирует меньшая длина канала:

$$L_{\text{eff}} = L - y_S - y_D. \quad (90a)$$

Для сравнения на рис. 37 приведены экспериментальные подпороговые кривые, характеристики, рассчитанные по формуле (90), и соответствующие результаты, полученные двумерным численным моделированием. Все они находятся в достаточно близком соответствии друг с другом.

8.4.2. Пороговые напряжения

В первом приближении величину порогового напряжения короткоканального прибора можно оценить, используя простые геометрические соображения [42]. Будем считать, что эффектив-

ный полный заряд Q'_B равен заряду ионизированных примесей в трапецидальной области под затвором (рис. 36, а):

$$\frac{Q'_B}{Z} = qN_A W_m \left(\frac{L+L'}{2} \right). \quad (91)$$

Из обычных тригонометрических соотношений получим для $L+L'$ следующее выражение:

$$\frac{L+L'}{2L} = 1 - \frac{r_j}{L} \left(\sqrt{1 + \frac{2W_m}{r_j}} - 1 \right). \quad (92)$$

При этом сдвиг порогового напряжения

$$\begin{aligned} \Delta V_T &= \frac{1}{C_i} \left(\frac{Q'_B}{ZL} - qN_A W_m \right) = - \frac{qN_A W_m}{C_i} \left(1 - \frac{L+L'}{2L} \right) = \\ &= - \frac{qN_A W_m r_j}{C_i L} \left(\sqrt{1 + \frac{2W_m}{r_j}} - 1 \right). \end{aligned} \quad (93)$$

Для учета влияния напряжения стока и смещения на подложке выражение (93) запишем в модифицированной форме [22]:

$$\Delta V_T = - \frac{qN_A W_m r_j}{2C_i L} \left[\left(\sqrt{1 + \frac{2y_S}{r_j}} - 1 \right) + \left(\sqrt{1 + \frac{2y_D}{r_j}} - 1 \right) \right], \quad (94)$$

где y_S и y_D определяются выражениями (89), а

$$W_m = \sqrt{2\epsilon_s (2\psi_B + V_{BS}) / qN_A}. \quad (95)$$

Как следует из рис. 38, выражение (94) достаточно хорошо описывает результаты соответствующих измерений.

Определенное влияние на пороговое напряжение оказывает также краевой эффект на боковой границе затвора. Соответствующий сдвиг порога при этом обусловлен тем фактом, что область обеднения выходит (в направлении z , перпендикулярно току I_D) за боковой край затвора (вставка на рис. 39). Предполагая эту краевую часть области пространственного заряда цилиндрической [44], для полного заряда в области обеднения будем иметь

$$Q_{BT} = qN_A ZLW \left(1 + \frac{\pi}{2} \frac{W}{Z} \right). \quad (96)$$

Из этого выражения видно, что боковое расширение области пространственного заряда описывается соответствующим нагрузочным коэффициентом $1 + \pi W / 2Z$. В результате наблюдается соответствующее увеличение порогового напряжения:

$$V_T = V_{FB} + 2\psi_B + \frac{\sqrt{2\epsilon_s qN_A (2\psi_B + V_{BS})}}{C_i} \left(1 + \frac{\pi}{2} \frac{W}{Z} \right). \quad (97)$$

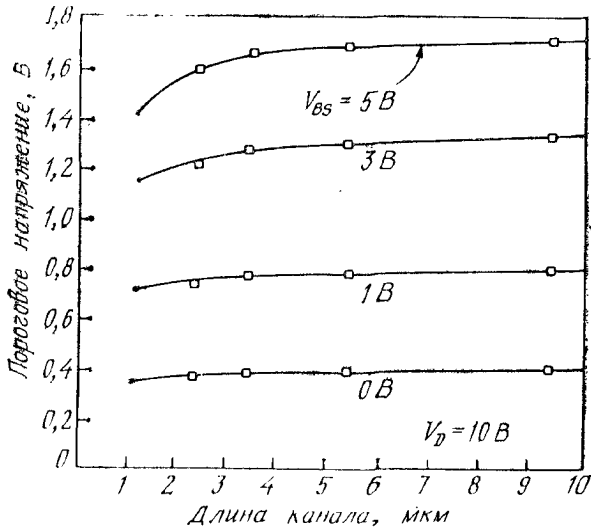


Рис. 38. Зависимость порогового напряжения от длины канала для МОП-транзисторов с $d = 360 \text{ \AA}$, $N_A = 2,5 \cdot 10^{16} \text{ см}^{-3}$ и $L = 2,4-9,4 \text{ мкм}$ [22].

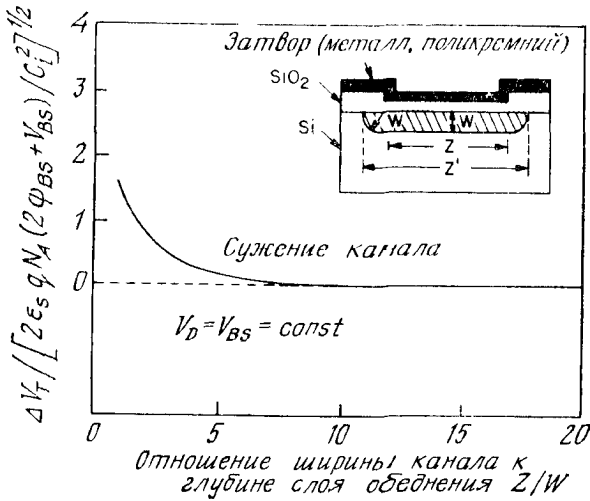


Рис. 39. Влияние бокового сужения канала на пороговое напряжение [44].

Влияние бокового сужения канала проиллюстрировано на рис. 39. Ясно, что заметного эффекта в этом случае можно ожидать, когда ширина канала Z становится сравнимой с характерной глубиной слоя обеднения W . Так, например, для прибора с шириной затвора $L = 1$ мкм соответствующее увеличение V_T становится существенным при уровнях легирования подложки, меньших 10^{16} см $^{-3}$.

8.4.3. Линейная область и режим насыщения

На рис. 40 приведены выходные вольт-амперные характеристики МОП-транзистора с длиной канала 0,73 мкм и толщиной слоя окисла 258 Å [45]. В область канала прибора сквозь окисел была имплантирована доза ионов бора $2 \cdot 10^{12}$ см $^{-2}$ с энергией 150 кэВ. Результирующий профиль концентрации можно было считать ступенчатым с глубиной $x_s = 0,3$ мкм и поверхностной концентрацией $7 \cdot 10^{16}$ см $^{-3}$. При напряжениях стока 0—5 В параметр γ (выражение (83)) изменялся в пределах 3—10 мкм $^2 \cdot \text{Å}$, что соответствует $L_{\text{мин}} = 0,6$ —0,9 мкм. Следовательно, рассматриваемый прибор (в соответствии с классификацией, данной на рис. 35) занимает промежуточное положение у края области, соответствующей длинноканальным приборам. В первом приближении токи насыщения транзистора можно вычислить, используя выражение (28) и подставив в него вместо L редуцированную эффективную длину канала (выражение (90а)). С ростом V_D

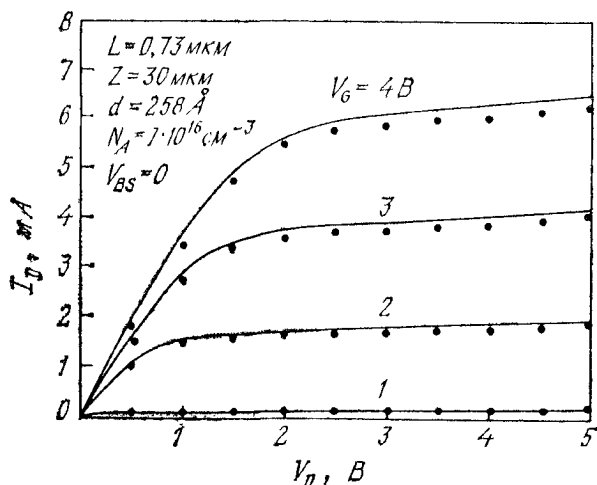


Рис. 40. Стоковые характеристики МОП-транзистора с длиной канала 0,73 мкм [45].

— экспериментальные; ●●● теоретические.

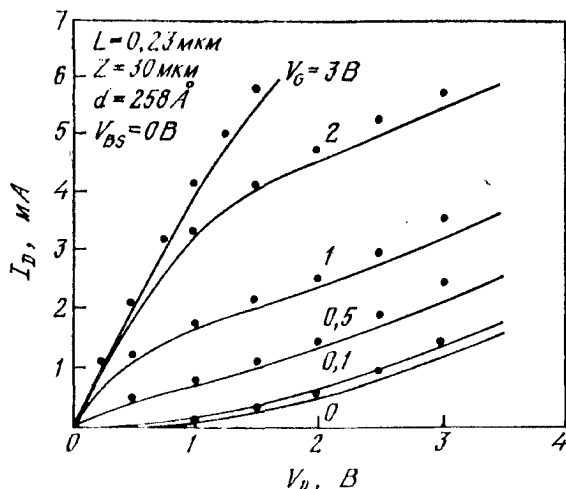


Рис. 41. Стоковые характеристики МОП-транзистора с длиной канала 0,23 мкм [45].

— экспериментальные; ●●● теоретические.

длина L_{eff} уменьшается, что обуславливает соответствующее увеличение тока стока I_D .

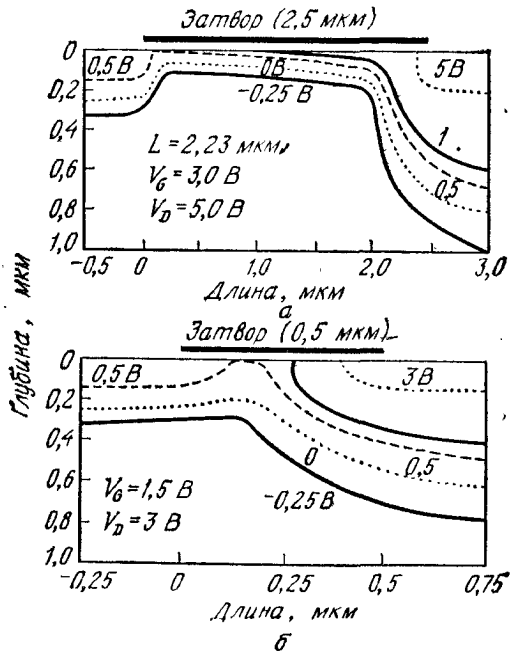
На рис. 41 приведены $I-V$ -характеристики прибора, все параметры которого идентичны параметрам МОП-транзистора с характеристиками, приведенными на рис. 40, за исключением длины канала, которая в данном случае составляла 0,23 мкм. Характеристики этого прибора демонстрируют ярко выраженное короткоканальное поведение. В нем уже при $V_D = 0$ происходит смыкание обедненных областей стока и истока, поскольку соответствующая сумма $y_S + y_D = 0,26$ мкм больше длины канала L . Следовательно, и во всем диапазоне стоковых напряжений (рис. 41) рассматриваемый прибор работает в режиме смыкания (прокола). В этом случае электроны из истока будут инжектироваться непосредственно в объем обедненной области канала, где они подхватываются электрическим полем и сразу же коллектируются стоком. Напряжение на стоке, требуемое для такого смыкания [10], равно

$$V_{pi} \approx \frac{qN_A(L - y_S)^2}{2\epsilon_s} - V_{bi}. \tag{98}$$

Если V_D существенно превышает эту величину, рассматриваемый ток инжекции ограничивает накопление заряда подвижных носителей в области пролета. В этом так называемом режиме ограничения тока пространственным зарядом

$$I_D \approx 9\epsilon_s \mu_n A V_D^2 / 8L^3, \tag{99}$$

Рис. 42. Распределение потенциала в структурах, характеристики которых приведены на рис. 40 и 41 [45].
 а — длинноканальная структура с $L = 2,23$ мкм; б — короткоканальная структура с $L = 0,23$ мкм.



где параметр A по порядку величины равен площади $n^+ - p$ -перехода истока. Ток инжекции I_D , увеличивающийся пропорционально V_G^2 , течет в объеме полупроводника параллельно току инверсионного канала, пропорциональному напряжению на затворе V_G .

Теоретические точки на рис. 40 и 41 получены в результате сложного двумерного численного моделирования при учете эффекта смыкания и полевой зависимости подвижности без использования каких-либо подгоночных параметров. Поэтому следует особо отметить хорошее соответствие этих результатов экспериментальным данным. На рис. 42, а и б приведены двумерные распределения потенциала в структурах длинноканального МОП-транзистора и рассмотренного выше короткоканального прибора с $L = 0,23$ мкм. Все параметры обоих приборов идентичны, за исключением длины канала, которая в длинноканальном приборе ($L = 2,3$ мкм) была в 10 раз больше, чем в короткоканальном. Отметим, что длина канала в обоих приборах примерно на 0,27 мкм короче длины соответствующих затворов в результате боковой диффузии имплантированных ионов под край затвора. В длинноканальном приборе (рис 42, а) эквипотенциали в канале параллельны друг другу и сосредоточены

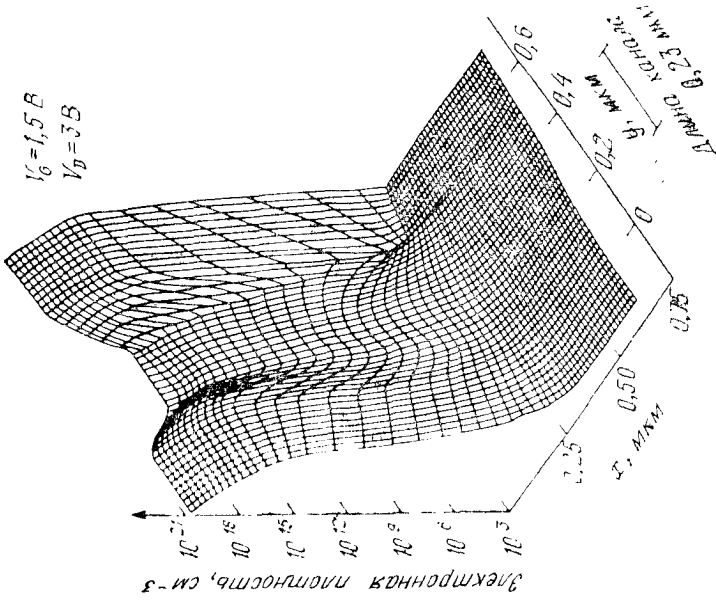


Рис. 44. Двумерное распределение электронной плотности в МОП-транзисторе с $L = 0,23 \text{ мкм}$ [45].

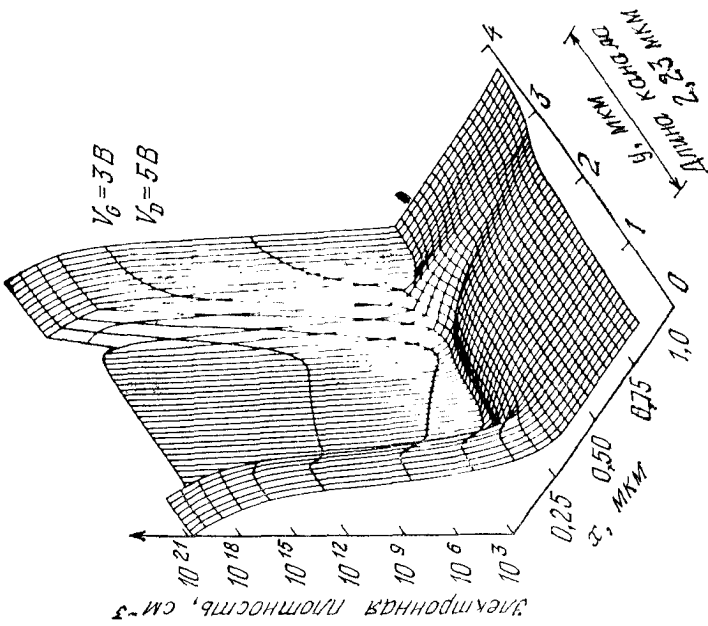


Рис. 43. Двумерное распределение электронной плотности в МОП-транзисторе с $L = 2,23 \text{ мкм}$ [45].

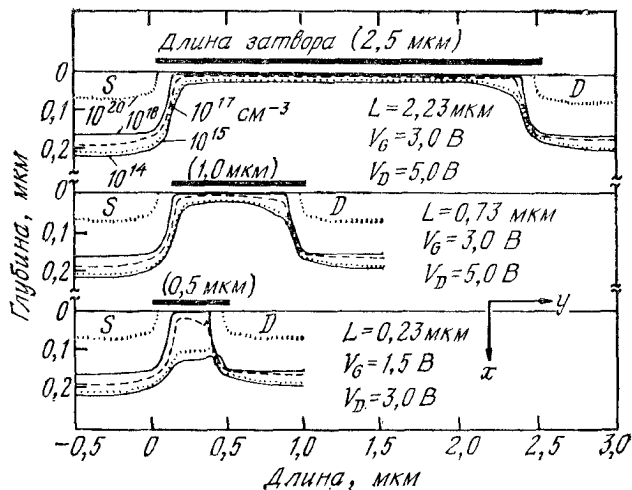


Рис. 45. Линии постоянной электронной плотности в МОП-транзисторах с длиной канала 2,23; 0,73 и 0,23 мкм [45].

в непосредственной близости от границы раздела с окислом. Напротив, в короткоканальном приборе (рис. 42, б) эквипотенциали уходят глубоко в объем подложки, что указывает на эффект смыкания.

На рис. 43 и 44 приведено распределение электронной плотности в рассматриваемых приборах. Из рис. 43, соответствующего режиму насыщения длинноканального прибора, ясно видно характерное уменьшение электронной плотности у края стока (отсечка канала). В длинноканальном приборе (рис. 43) электроны локализованы у границы с окислом, в то время как в короткоканальном МОП-транзисторе (рис. 44) электронная плотность довольно глубоко простирается в толщу подложки.

На рис. 45 показаны линии одинаковой электронной плотности в плоскости x — y , рассчитанные для рассматриваемых МОП-транзисторов с длиной канала 2,23; 0,73 и 0,23 мкм. Видно, что, как уже говорилось выше, в длинноканальном МОП-транзисторе с длиной затвора 2,5 мкм электронный ток сосредоточен в слое толщиной ~ 200 Å у границы раздела. Отметим также, что у края стока, где происходит отсечка канала, часть линий тока уходит от поверхности в объем подложки. В приборе с длиной затвора 1 мкм ток все еще остается сосредоточенным у границы с окислом, однако область отсечки уже захватывает значительно большую часть длины канала. При уменьшении длины канала до 0,23 мкм (длина затвора 0,5 мкм) линии тока уходят глубоко в подложку, а свойства прибора теряют длинноканальный характер.

В заключение следует констатировать, что из-за сложной природы двумерного распределения потенциала выразить в простой аналитической форме вольт-амперные характеристики короткоканальных приборов невозможно. Подходящие приближенные выражения можно получить, определенным образом модифицируя основное выражение (22), тем или иным способом «выделяя» действующие заряды (например так, как это показано на рис. 36). Обычно в таких упрощенных аналитических моделях удается путем подбора подходящих значений подгоночных параметров обеспечить неплохое соответствие с экспериментальными данными [46—48].

8.4.4. Умножение носителей и зарядка окисла

В длинноканальных приборах при достаточно высоком напряжении на стоке в области отсечки у границы канала со стоком происходит слабый лавинный пробой. Генерированные при этом электроны сразу же уходят в сток, а генерированные дырки коллектируются подложкой, создавая ток подложки I_{BS} . Зависимость этого тока от напряжения на затворе в длинноканальном приборе с $L = 10$ мкм приведена на рис. 46 [49]. Здесь показаны также полный ток стока I_D транзистора и генерационно-рекомбинационный ток в области обеднения. Диапазон изменения тока на рис. 46 перекрывает все характерные режимы

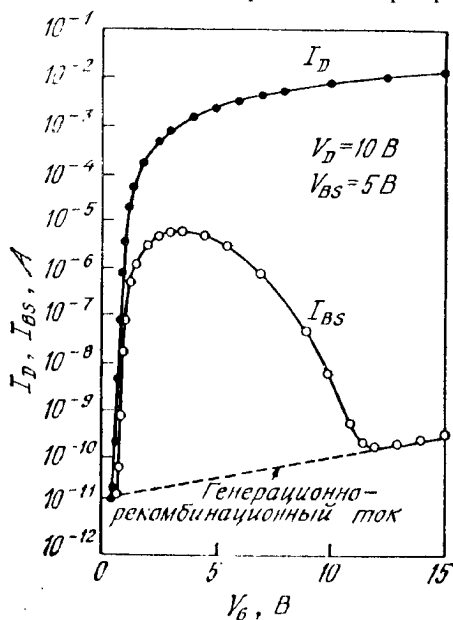


Рис. 46. Зависимость тока стока и тока подложки в длинноканальном МОП-транзисторе от напряжения на затворе [49].

работы прибора (подпороговый участок, линейную область и насыщение). При увеличении напряжения на затворе ток подложки I_{BS} сначала увеличивается, а затем, проходя через максимум, снова уменьшается. Такое поведение тока I_{BS} можно объяснить следующим образом. Предположим для простоты, что ударная ионизация происходит однородно по длине области отсечки канала ΔL . Тогда для тока подложки I_{BS} можно написать

$$I_{BS} = I_D \alpha \Delta L, \quad (100)$$

где α — коэффициент ударной ионизации, т. е. число электронно-дырочных пар, генерируемых электроном на единице длины. При заданном напряжении на стоке и при увеличении V_G одновременно увеличиваются ток I_D и напряжение $V_{D\text{ sat}}$. Однако с ростом $V_{D\text{ sat}}$ уменьшается поверхностное поле в области отсечки $(V_D - V_{D\text{ sat}})/L$, что, естественно, приводит к уменьшению α . Следовательно, имеются два противоположных фактора: 1) рост I_D обуславливает увеличение I_{BS} при малых V_G и 2) уменьшение α приводит к снижению I_{BS} при больших напряжениях на затворе. Максимуму I_{BS} соответствует напряжение на затворе, при котором действие обоих факторов сбалансировано.

В короткоканальных приборах возникает дополнительный эффект, обусловленный током дырок, генерированных в области пробоя. Дело в том, что при малом расстоянии между стоком и истоком и при достаточном напряжении стока V_D значительная часть этих дырок увлекается электрическим полем в исток, а не в подложку (рис. 47) [50]. Встречая на пути энергетический барьер $n^+ - p$ -перехода истока, дырки накапливаются у его внешней границы и «открывают» переход. В результате исток начинает инжектировать в объем подложки электроны, которые коллектируются стоком. Все это представляет собой механизм «включения» паразитного биполярного $n - p - n$ (исток—подложка—сток)-транзистора в короткоканальных приборах. Ясно, что развитие этого процесса приводит к электрическому пробояю структуры, начало которого определяется условием

$$\alpha_{npn} M = 1. \quad (101)$$

Здесь α_{npn} — коэффициент усиления тока биполярного транзистора в схеме с общей базой:

$$\alpha_{npn} = \text{sech}(L_G/L_{diff}) \simeq 1 - \frac{L_G^2}{2L_{diff}^2}, \quad (102)$$

где L_G — эффективная толщина базы, равная по порядку величины длине затвора, а L_{diff} — диффузионная длина в подложке. Коэффициент умножения M можно записать в виде

$$M = \left[1 - \left(\frac{BV_{CEO}}{V_{D\text{ sub}}} \right)^n \right]^{-1}, \quad (103)$$

где BV_{CEO} — напряжение пробоя биполярного транзистора в схеме с общим эмиттером и разомкнутой базой, а $V_{D\text{sub}}$ — напряжение пробоя перехода сток—подложка. Из выражений (101)–(103) для результирующего напряжения пробоя короткоканального МОП-транзистора получим

$$V_B = BV_{CEO} \approx \frac{V_{D\text{sub}}}{2^{1/n}} \left(\frac{L_G}{L_{diff}} \right)^{2/n}. \quad (104)$$

На рис. 48 приведены результаты измерений напряжения пробоя в структурах с различной длиной канала L . Удовлетворительное согласие этих данных с результатами расчета по формуле (104) получается при $n = 5,43$. Различие пробивных напряжений структур с различной глубиной стокового перехода можно объяснить зависимостью $V_{D\text{sub}}$ от кривизны перехода (гл. 2).

Другим эффектом, существенным для работы короткоканальных приборов, который возникает при больших полях, является зарядка окисла [51]. Оказывается, что при высоких продольных электрических полях часть электронов, двигаясь в канале, разогревается до энергии, достаточной для преодоления энергетического барьера на границе Si-SiO_2 (3,1 эВ), и инжектируется в окисел (рис. 49 (вставка)). Горячие электроны могут инжектироваться в окисел также из электронно-дырочной плазмы лавинного пробоя в области отсечки канала у границы со стоком. Кроме того, в SiO_2 могут инжектироваться электроны, термически гене-

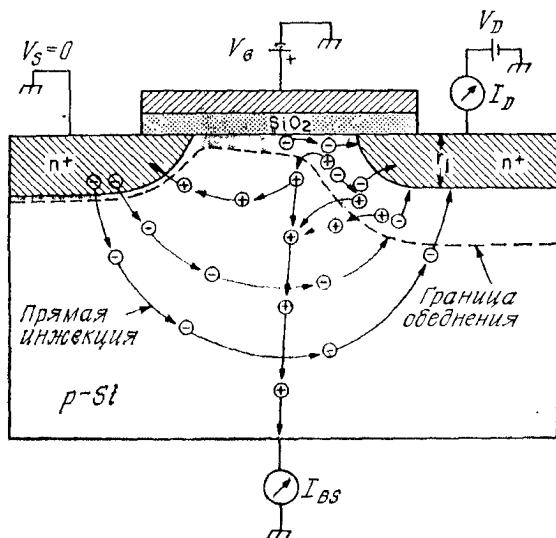


Рис. 47. Эффект паразитного биполярного транзистора [50].

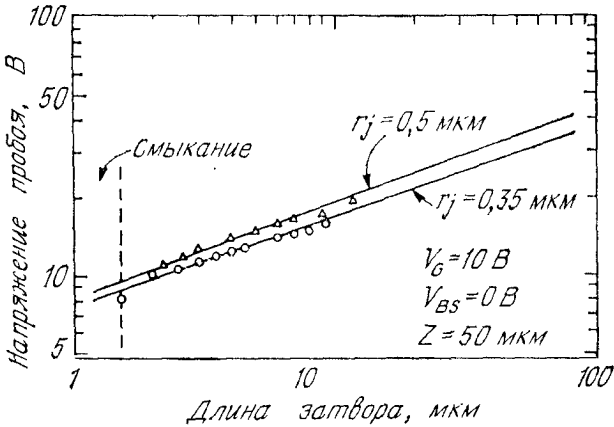


Рис. 48. Зависимость напряжения пробоя от длины затвора, обусловленная эффектом паразитного биполярного транзистора [50].

рированные в объеме подложки и разогретые большим поперечным электрическим полем на пути к границе раздела. Влияние инжекции горячих электронов в окисел на свойства МОП-транзистора показано на рис. 49. Отметим, что, во-первых, пороговое напряжение при зарядке окисла становится более положительным и, во-вторых, снижается крутизна МОП-транзистора (меньший наклон I — V -зависимости) вследствие уменьшения подвижно-

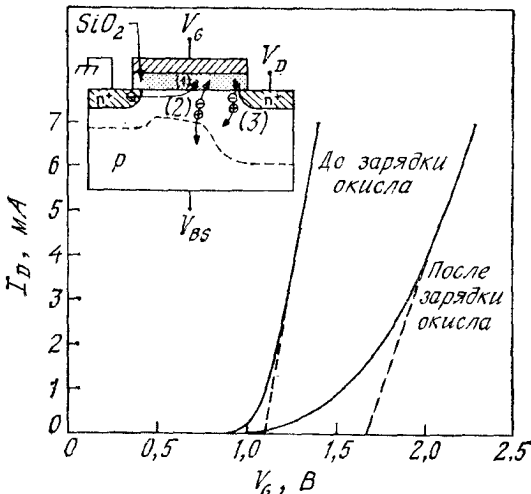


Рис. 49. Процессы, обуславливающие зарядку окисла и результирующее изменение характеристики МОП-транзистора [51].

(1) — ток в канале; (2) — горячие электроны; (3) — электронно-дырочная плазма лавинного пробоя.

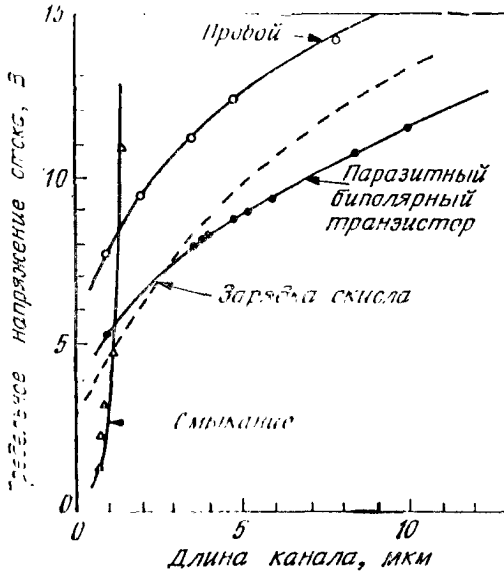


Рис. 50. Зависимость предельных напряжений стока от длины канала при различных эффектах [52].

сти в канале. Кроме того, возрастают подпороговые токи прибора вследствие увеличения плотности поверхностных состояний.

Зарядка окисла оказывает существенное влияние на стабильность работы МОП-транзистора и в конечном счете определяет срок его службы, поскольку этот процесс приводит к непрерывной деградации рабочих характеристик прибора. Для замедления деградации и тем самым увеличения срока службы прибора приходится ограничивать максимальную амплитуду рабочих напряжений МОП-транзистора. Из графиков, приведенных на рис. 50, можно выявить влияние рассмотренных выше механизмов на предельные напряжения стока [52]. Эти зависимости получены для некоторых определенных значений параметров прибора (d , r_f , N_A). При ином выборе этих параметров относительная роль рассмотренных механизмов может, естественно, изменяться.

Для устранения эффекта паразитного транзистора следует уменьшать сопротивление подложки $R_{суб}$, так чтобы произведение тока подложки на это сопротивление не превышало напряжения 0,6 В, требуемого для открывания истока, когда напряжение на стоке равно или превышает соответствующее напряжение BV_{CEO} . При этом напряжение пробоя короткоканального МОП-транзистора уже не ограничивается величиной BV_{CEO} и прибор может работать при более высоких напряжениях, что обеспечивает большую надежность [50]. Чтобы уменьшить зарядку окисла, следует уменьшать плотность электронных ловушек в SiO_2 [53].

Для увеличения напряжения смыкания применяют одно- или даже двукратно ионно-имплантированные структуры с повышенным легированием приповерхностной области подложки. Некоторые из них рассмотрены в разд. 8.5.

8.5. МОП-ТРАНЗИСТОРНЫЕ СТРУКТУРЫ

Для улучшения рабочих характеристик МОП-транзисторов (повышения быстродействия, уменьшения энергопотребления, повышения надежности и уровня переключаемых мощностей) в настоящее время предложено большое число разнообразных структурных решений. Ниже рассмотрены некоторые наиболее типичные и употребительные из этих структур.

8.5.1. Масштабное уменьшение размеров

Как уже говорилось выше, короткоканальные эффекты ухудшают рабочие характеристики МОП-транзисторов и поэтому нежелательны. Одним из способов устранения этих эффектов является пропорциональное уменьшение всех характерных размеров прибора. При этом следует во столько же раз уменьшить и характерные значения рабочих напряжений, с тем чтобы внешние электрические поля в приборе оставались на прежнем уровне. Такое масштабное уменьшение размеров представляет собой простейший подход к проблеме миниатюризации МОП-транзисторов.

На рис. 51, а показан обычный прибор, а на рис. 51, б — прибор, все характерные размеры которого (толщина окисла, длина и ширина канала, размеры n^+ — p -переходов) [54] уменьшены в κ раз (κ — масштабный фактор). Уровень легирования подложки во втором приборе в κ раз больше, чем в обычном приборе, и, поскольку питающие напряжения также уменьшены в κ раз, толщина соответствующих обедненных областей оказывается в κ раз меньше. На рис. 51, в приведены сток-затворные характеристики обоих приборов. Отметим, что пороговое напряжение масштабного уменьшенной структуры также уменьшилось в κ раз. Такое пропорциональное уменьшение характерных размеров отдельного прибора увеличивает плотность компоновки приборов на кристалле в κ^2 раз, уменьшает характерные времена задержки (увеличивает быстродействие) в κ раз и уменьшает мощность, рассеиваемую в отдельном приборе, в κ^2 раз.

Обратим внимание на то, что область подпороговых токов на характеристиках обоих приборов (рис. 51, в) практически одинакова. Это неудивительно, поскольку при таком масштабном преобразовании характерное напряжение $S \sim (1 + C_D/C_i)$ остается неизменным (емкости C_D и C_i увеличились в κ раз). Отметим также, что контактная разность переходов V_{ib} и характерный

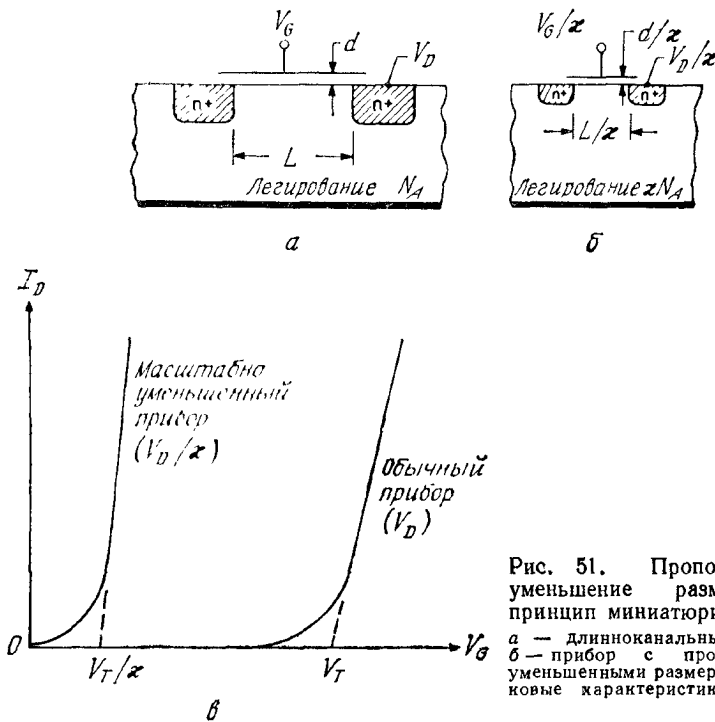


Рис. 51. Пропорциональное уменьшение размеров как принцип миниатюризации [54]. а — длинноканальный прибор; б — прибор с пропорционально уменьшенными размерами; в — статические характеристики.

поверхностный потенциал ψ_D не уменьшаются при таком «масштабировании», а даже увеличиваются (на $\sim 10\%$ при увеличении уровня легирования в 10 раз). Поэтому не изменяется и характерная разность затворных напряжений ($\sim 0,5$ В), соответствующих началу сильной инверсии и режиму обеднения. Паразитные емкости при таком масштабном уменьшении размеров могут и не уменьшиться. Сопротивление шин разводки обычно увеличивается с уменьшением размеров.

Большую «гибкость» в решении вопросов миниатюризации допускает подход, основанный на использовании эмпирического соотношения (83) [41]. В этом случае при обеспечении заданного минимального размера $L_{\text{мин}}$ можно в определенных пределах варьировать различные параметры структуры, поддерживая при этом неизменной требуемую величину синтетического параметра γ . Следовательно, при таком подходе не обязательно уменьшать все характерные размеры структуры в одинаковое число раз, а в определенных пределах можно даже варьировать геометрию структуры. Последнее можно использовать в целях упрощения технологии либо для оптимизации других аспектов функционирования прибора.

8.5.2. Высококачественные МОП-структуры (HMOS)

Так называют МОП-транзисторы, изготовленные с использованием ионной имплантации. Структура, показанная на рис. 52, а, содержит один имплантированный слой [55]. Он контролирует пороговое напряжение МОП-транзистора и в определенной мере препятствует смыканию обедненных областей стока и истока. Отметим, что имплантированный слой здесь довольно узкий, и обедненная область стока выходит за его пределы в низколегированную подложку. Это обеспечивает достаточно малую выходную емкость МОП-транзистора (емкость перехода сток — подложка). На рис. 52, б показана такая структура с двойной ионной имплантацией [56]. Параметры первого имплантированного слоя p_1 выбираются для обеспечения требуемого порогового напряжения МОП-транзистора, а характеристики второго слоя p_2 — для предотвращения смыкания. Использование двойной ионной имплантации позволяет существенно минимизировать

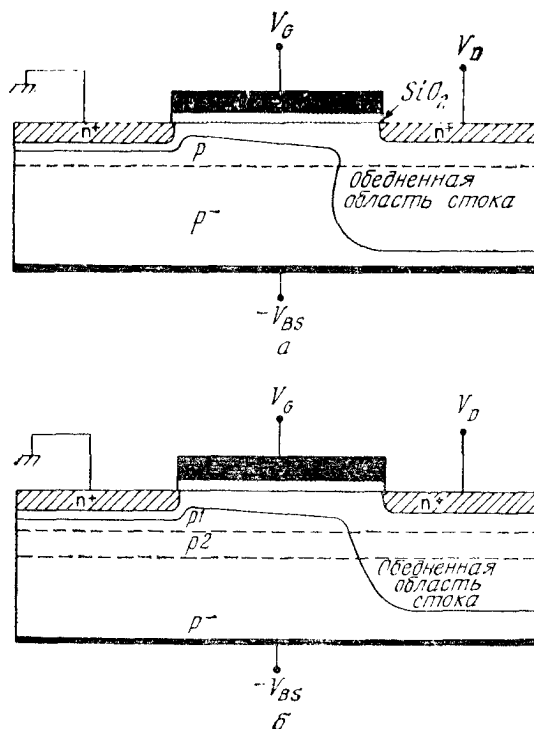


Рис. 52. Высококачественная МОП-структура (HMOS) [55, 56].

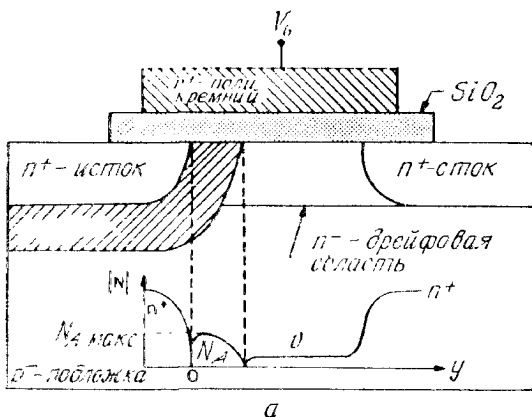
а — с одним имплантированным слоем; б — двукратно имплантированная структура.

короткоканальные эффекты в таких МОП-структурах в физически малыми длинами каналов.

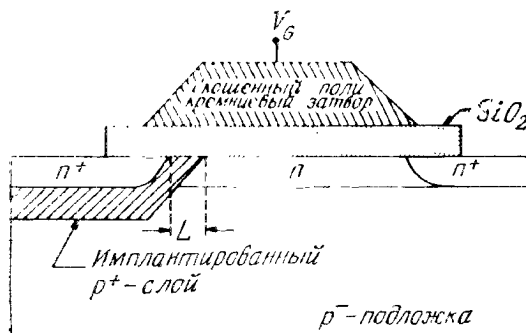
Имплантация, однако, ухудшает подпороговый участок характеристик [19] (увеличенные значения характерного напряжения S) и повышает чувствительность порога к смещению подложки V_{BS} . Эти нежелательные эффекты необходимо учитывать при оптимизации параметров структуры.

8.5.3. МОП-структуры с двойной диффузией (DMOS)

МОП-структура, схема которой приведена на рис. 53, а, изготовлена с использованием так называемой двойной диффузии [57]. Этот процесс, широко применяемый в биполярной технологии, основан на том, что скорость диффузии p -примеси (например, бора) в кремнии выше, чем для n -примеси (например, фосфора).



а



б

Рис. 53 МОП-структуры с двойной диффузией (а) и двукратной ионной имплантацией (б) [57, 58].

В рассматриваемом случае он позволяет получать структуры со сравнительно короткими каналами. На рис. 53, а также показан профиль легирования в такой структуре вдоль границы с окислом. Отметим, что сразу за каналом следует низколегированная дрейфовая область. В аналогичной структуре, изготовленной с использованием двукратной ионной имплантации [58] (рис. 53, б), в качестве маски используется поликремниевый затвор транзистора, скошенные края которого обеспечивают выход имплантированной p -области на границу с окислом.

Такие структуры могут иметь очень короткие каналы, и при их изготовлении используется обычный литографический процесс. Обе структуры (рис. 53) весьма устойчивы по отношению к смыканию из-за высокой концентрации примеси в p -слое. Сравнительно протяженная низколегированная дрейфовая область служит как бы нагрузкой, стабилизируя электрическое поле в канале на уровне $\approx 10^4$ В·см⁻¹, достаточном для достижения носителями скорости насыщения [59]. Такое ограничение максимальных электрических полей в структуре препятствует развитию нежелательных процессов лавинного умножения носителей и зарядки окисла. В этом отношении структуры с двойной диффузией имеют значительные преимущества перед МОП-транзисторами, изготовленными по стандартной и ПМOS-технологии [11].

Однако в этих структурах труднее контролировать пороговое напряжение МОП-транзистора, которое (рис. 53, а) определяется максимальным значением концентрации $N_{A \text{ макс}}$ на границе с окислом [60]. Кроме того, сравнительно высокая концентрация примеси в узком p -слое, необходимая для предотвращения смыкания стока с истоком, ухудшает подпороговый участок характеристик, соответствующий закрытому состоянию МОП-транзисторов.

8.5.4. МОП-транзисторы с заглубленным каналом

Конструкция этого прибора [61], показанная на рис. 54 (вставка), служит примером использования концепции миниатюризации, основанной на эмпирическом соотношении (83): минимальная длина канала $L_{\text{мин}}$, при которой еще сохраняется «длинноканальность» электрических характеристик МОП-транзистора и уменьшается с уменьшением глубины переходов по закону $r_j^{1/3}$. Представленные на рис. 54 зависимости порогового напряжения от длины канала демонстрируют такую минимизацию короткоканальных эффектов с уменьшением r_j . Легко видеть, что рассматриваемая конструкция с заглубленным каналом позволяет снизить r_j до нуля и даже до отрицательных значений.

Очевидным недостатком структур с заглубленным каналом, особенно существенным для приборов с субмикронными разме-

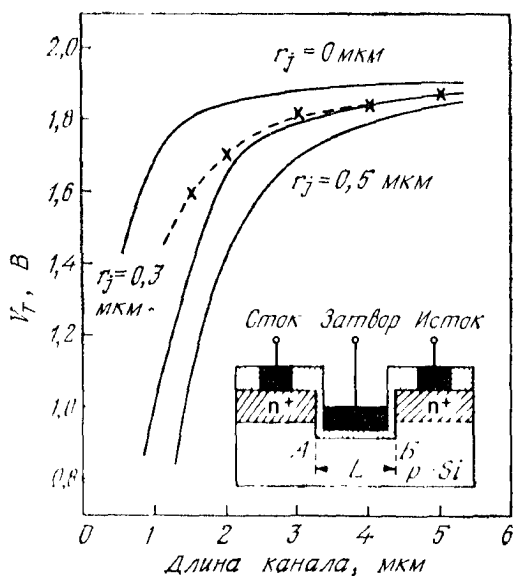


Рис. 54. Расчетные (—) и экспериментальные (---x---) зависимости порогового напряжения от длины канала при различных значениях глубины перехода. На вставке показан МОП-транзистор с заглубленным каналом [61].

рами, являются технологические трудности, возникающие при формировании требуемой «канавки». Кроме того, в такой геометрии плохо контролируется толщина окисла в «углах» А и Б, которая определяет пороговое напряжение транзистора. Отметим также, что в структурах с заглубленным каналом можно ожидать повышенную скорость зарядки окисла в результате более интенсивной инжекции горячих электронов.

8.5.5. Барьеры Шоттки в качестве стока и истока

Использование барьеров Шоттки в качестве стока и истока МОП-транзисторов имеет ряд достоинств с точки зрения рабочих характеристик приборов и технологии их изготовления. Схематически такая МОП-структура показана на рис. 55, а [62]. Барьеры Шоттки имеют сравнительно малую толщину перехода, что весьма благоприятно для минимизации короткоканальных эффектов. Высокая проводимость контактов Шоттки также обеспечивает достаточно малую величину эффективного последовательного сопротивления истока. С точки зрения технологии исключение ряда высокотемпературных операций способствует повышению качества окисного слоя и облегчает воспроизведение геометрии структуры. Кроме того, МОП-транзисторы с барьерами Шоттки могут быть изготовлены на полупроводниках (таких, как, например, CdS), в которых не удается получить достаточно качественных $p-n$ -переходов.

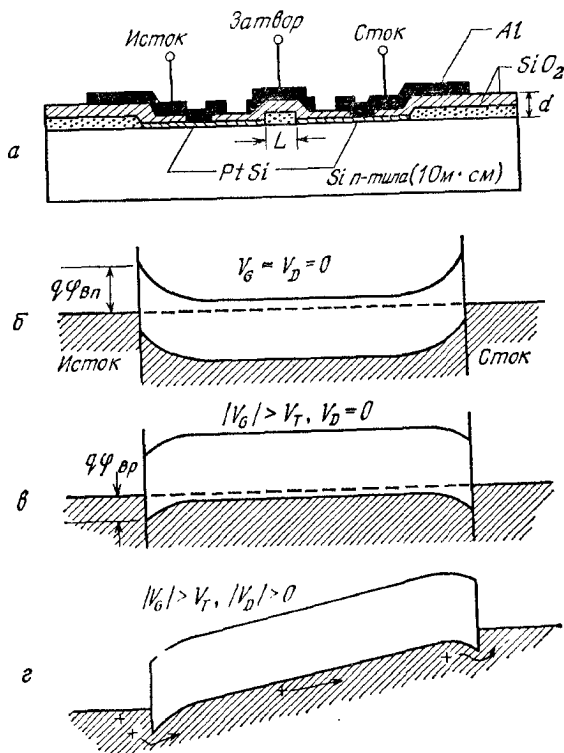


Рис. 55. МОП-транзистор с барьерами Шоттки в качестве стока и истока [62]. а — схема поперечного разреза прибора; б, в и г — зонные диаграммы при различных комбинациях напряжений на электродах.

На рис. 55, б приведена зонная диаграмма структуры в условиях термодинамического равновесия ($V_G = V_D = 0$). Здесь $q\varphi_{Bn}$ — высота энергетического барьера между металлом и n -подложкой (например, для контакта PtSi—Si $q\varphi_{Bn} = 0,85$ эВ). Когда напряжение на затворе достаточно велико, чтобы инвертировать приповерхностную область полупроводника от n -типа к p -типу (рис. 55, в), энергетический барьер между истоком и p -инверсионным слоем равен уже $q\varphi_{Bp} = 0,25$ эВ (для того же контакта PtSi—Si). Термоионная эмиссия через такой малый барьер поставляет достаточное число дырок в инверсионный слой структуры, где они увлекаются электрическим полем к контакту Шоттки стока (рис. 55, г). Отметим, что в этом рабочем режиме прибора барьер Шоттки истока оказывается обратномещенным. При комнатных температурах плотность термоионного тока через барьер высотой 0,25 эВ составляет $\sim 10^3$ А·см⁻². При необходи-

мости ее увеличения следует использовать контакты Шоттки с большей высотой энергетического барьера по отношению к n -кремнию. Изготовление аналогичных МОП-транзисторов на p -подложке (с n -каналом) в настоящее время представляет большие трудности, поскольку еще не найдены контакты Шоттки с достаточно большой высотой энергетического барьера по отношению к p -кремнию.

8.5.6. Тонкопленочный транзистор

Тонкопленочный транзистор (ТПТ) схематически изображен на рис. 56 (вставка). Его структуру составляют последовательно нанесенные на изолирующую подложку слои полупроводника (например, CdS), диэлектрика и металла [63]. Вольт-амперные характеристики ТПТ обогащенного типа (рис. 56) подобны соответствующим характеристикам обычных МОП-транзисторов. Поскольку полупроводниковая пленка в ТПТ обычно наносится простым распылением, в ней содержится значительно большее число дефектов и нарушений кристаллической решетки, чем в соответствующих монокристаллических образцах. Поэтому транспортные процессы в таких пленках более сложны, чем в моно-

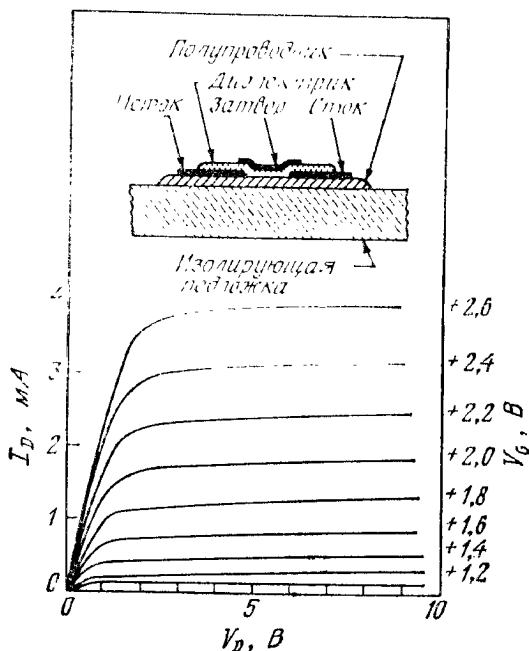


Рис. 56. Тонкопленочный транзистор и его стоковые характеристики [63].

кристаллах. Для улучшения приборных характеристик ТПТ, надежности и воспроизводимости следует снижать плотность объемных и поверхностных ловушек в полупроводниковой пленке.

8.5.7. Структуры типа «кремний на изоляторе» (КНИ)

К настоящему времени предложено довольно большое число структур типа «кремний на изоляторе» (КНИ): кремний на сапфире (КНС), кремний на шпинели, кремний на нитриде и кремний на окисле. КНИ-приборы формируются по стандартной МОП-технологии на монокристаллической кремниевой пленке, эпитаксиально выращенной на изолирующей подложке (например, на Al_2O_3 в случае КНС) [64]. Схема такой структуры приведена на рис. 57, а. Диэлектрическая подложка обеспечивает электрическую изоляцию соседних приборов. КНИ-структуры имеют низкие паразитные емкости, что обеспечивает высокое быстродействие этих приборов. Типичные характеристики n -канального КНС-прибора показаны на рис. 57, б. Перегибы в области насыщения характеристик можно объяснить следующим образом. Когда напряжение стока увеличивается, в области вблизи стокового контакта начинается слабый процесс лавинного умножения. Генерированные при этом электроны сразу же поступают в сток, а генерированные дырки двигаются к истоку, составляя ток подложки. Однако, поскольку КНС-прибор не имеет контакта к подложке, эти дырки накапливаются у границы истока, смещая его в прямом направлении. При этом прибор фактически как бы находится в режиме с положительно смещенной подложкой $V_{BS} > 0$, что эквивалентно уменьшению порогового напряжения. Этим объясняются наблюдаемые скачки тока. Амплитуда скачков порогового напряжения сильно зависит от степени легирования под-

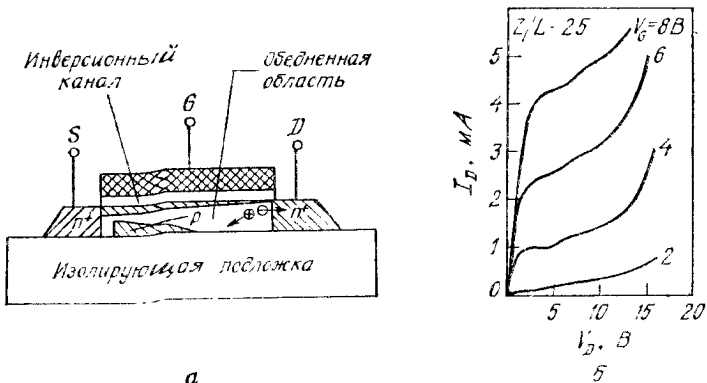


Рис. 57. МОП-транзистор типа «кремний на изоляторе» (а) и стоковые характеристики КНС-прибора (б) [64].

ложки: чем больше концентрация, тем больше скачок V_T и тем заметнее перегибы вольт-амперных характеристик.

В работах [65, 66] для изготовления КНИ-приборов предложено использовать рекристаллизацию лазерным отжигом пленок аморфного кремния на изолирующих подложках, таких, как Si_3N_4 и SiO_2 . В настоящее время рабочие характеристики КНИ-структур значительно уступают соответствующим характеристикам приборов на монокристаллических подложках. Это обусловлено высокой плотностью дефектов в объеме таких эпитаксиальных кремниевых пленок и большой плотностью поверхностных состояний на границах раздела кремний—изолятор. При улучшении качества пленок КНИ-приборы нашли бы широкое применение в сверхбыстродействующих схемах. Кроме того, на основе КНИ-структур можно было бы попытаться освоить трехмерный уровень интеграции микросхем, «накладывая» друг на друга соответствующие плоские базовые КНИ-структуры [80].

8.5.8. МОП-транзистор с V-канавкой

Структура МОП-транзистора с V-канавкой [67] схематически показана на рис. 58, а, а структура МОП-транзистора с U-канавкой — на рис. 58, б [68]. Эти приборы изготавливаются на кремниевых подложках с ориентацией $\langle 100 \rangle$ на основе эффекта анизотропного травления. Вытравливаемые V-канавки имеют характерный угол наклона на $54,7^\circ$ по отношению к горизонтальной плоскости. Как видно из рис. 58, а, распределение легирующих примесей здесь аналогично распределению примесей в МОП-транзисторах с двойной диффузией. В этих приборах два параллельных друг другу канала длиной L расположены на обеих сторонах вытравленной канавки. Общим стоком служит подложка. В такой конфигурации легко соединить параллельно требуемое число отдельных приборов, что позволяет создавать таким способом мощные переключатели, выдерживающие большие токи [69]. В тех ситуациях, когда требуется планарная организация схемы, т. е. нужно, чтобы все контакты выходили на внешнюю поверхность кристалла, можно исключить из структуры n^+ -область и увеличить степень легирования p -слоя (заменить p на p^+). Левая p^+ -область тогда будет стоком, правая — истоком (или наоборот), а собственно канал будет расположен по границе ABC .

8.5.9. Гексагональный МОП-транзистор

На рис. 59, а показана полупроводниковая структура, представляющая собой плотно упакованный массив так называемых гексагональных МОП-транзисторов [70]. Отдельная ячейка структуры представляет собой МОП-транзистор с двойной диффузией

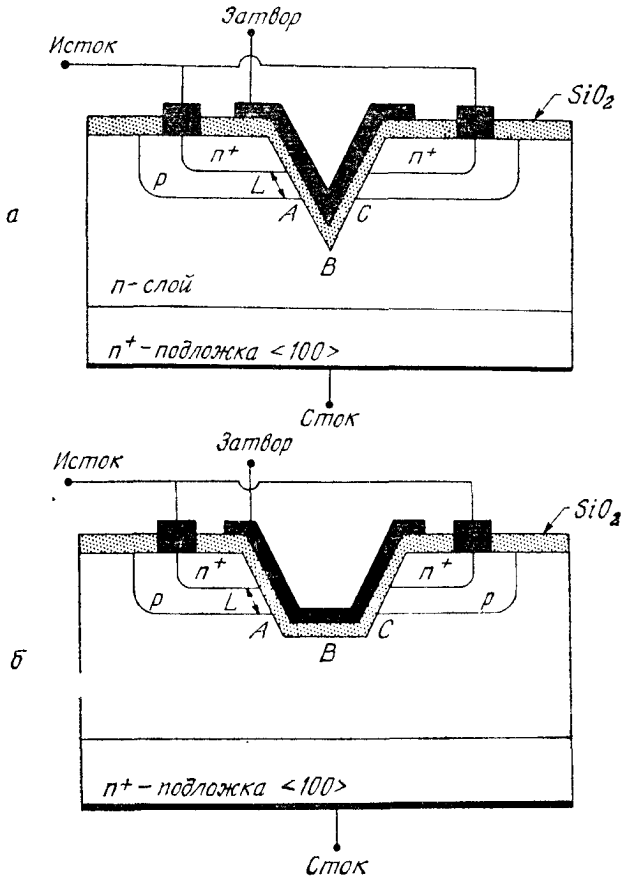


Рис. 58. МОП-транзистор с V-образной канавкой (а) и U-образной канавкой (б) [67, 68].

и шестиугольным истоком. Узкий диффузионный канал, расположенный по его периферии, перекрыт поликремниевым затвором. Общим стоком служит подложка структуры. Гексагональная структура имеет высокую плотность размещения элементов на кристалле ($\sim 10^5$ см $^{-2}$). Большие значения отношения ширины к длине канала ($6ZN/L$) обеспечивают высокую крутизну и малое выходное сопротивление во включенном состоянии.

На рис. 59, б приведены выходные характеристики мощной гексагональной структуры, рассчитанной на переключение токов до 100 А. Напряжения, соответствующие участку насыщения характеристик, имеют те же значения, что и в аналогичных биполярных приборах таких же размеров. Выходное сопротивление

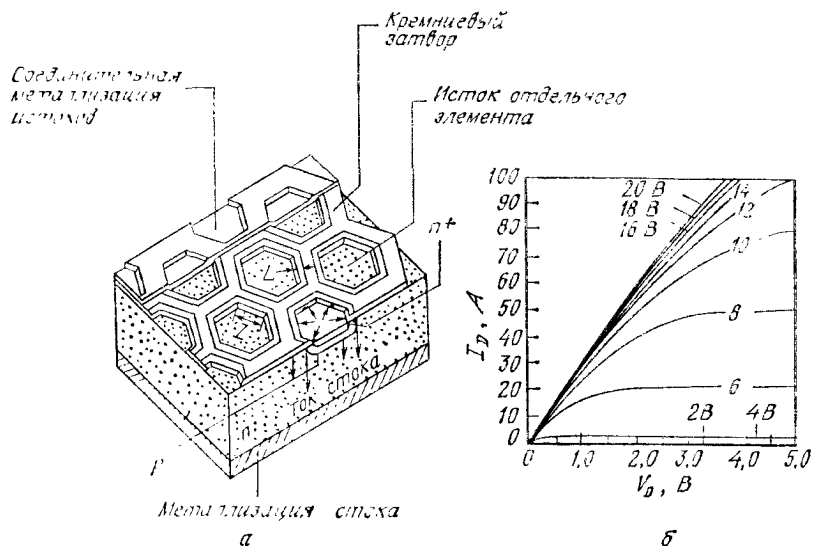


Рис. 59. Гексагональный МОП-транзистор (а) и его выходные характеристики (б) [70].

во включенном состоянии составляет всего лишь 0,05 Ом. При необходимости можно получить меньшие выходные сопротивления, увеличив отношение $6NZ/L$.

8.6. ЭНЕРГОНЕЗАВИСИМЫЕ ЭЛЕМЕНТЫ ПАМЯТИ

Если модифицировать затвор обычного МОП-транзистора таким образом, чтобы в подзатворном диэлектрике мог храниться электрический заряд, то мы получим новый прибор — энергонезависимый элемент памяти. С тех пор как в 1967 г. был предложен первый такой прибор [71], энергонезависимые элементы памяти получили широкое развитие и распространение. На их основе в настоящее время разрабатывается и изготавливается широкий класс полупроводниковых интегральных запоминающих устройств (ЗУ), таких, как электрически репрограммируемые постоянные ЗУ (ПЗУ с электрической перезаписью), программируемые ПЗУ со стиранием, энергонезависимые ЗУ с произвольной выборкой и т. п.

Энергонезависимые элементы памяти можно разделить на два класса: приборы с плавающим затвором и структуры с двухслойным диэлектриком — МДОП (металл—диэлектрик—окисел—полупроводник). «Запись» заряда в них осуществляется инжекцией носителей из кремния через слой окисла. Заряд хранится

либо в плавающем затворе, либо на границе окисла со вторым диэлектриком (в МДОП-структурах). Этот заряд, изменяя пороговое напряжение МОП-транзистора, «сдвигает» прибор в состояние с высоким порогом. В хорошо сконструированной ячейке такое состояние сохраняется более 100 лет. «Стирание» хранящегося заряда (возврат прибора в исходное состояние с низким порогом) осуществляется либо электрически (с помощью соответствующих импульсов напряжения на затворе), либо некоторыми другими способами, например ультрафиолетовым облучением структуры.

8.6.1. Приборы с плавающим затвором

В структуре с плавающим затвором (рис. 60) на первом, тонком слое окисла I (I) расположен металлический электрод M (I) — плавающий затвор, отделенный вторым, толстым слоем диэлектрика I (2) от внешнего металлического затвора M (2). Зонные диаграммы структуры с плавающим затвором, соответствующие режимам записи, хранения и стирания, приведены на рис. 61.

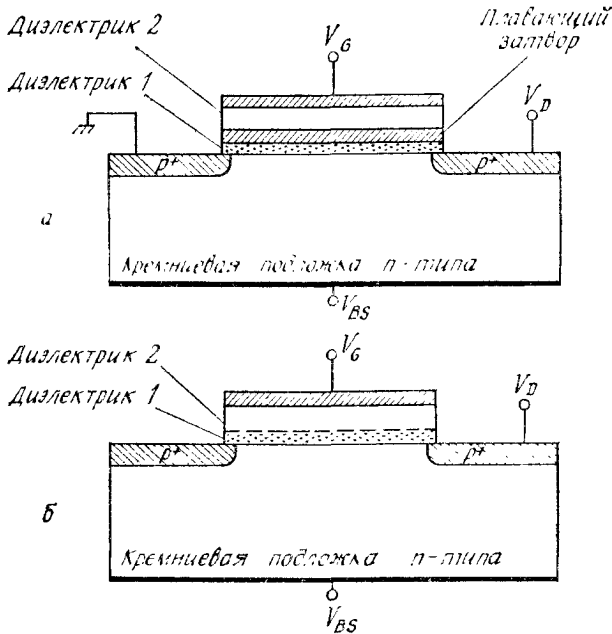


Рис. 60. Энергонезависимый элемент памяти с плавающим затвором (а) и энергонезависимый элемент памяти типа МДОП (б).

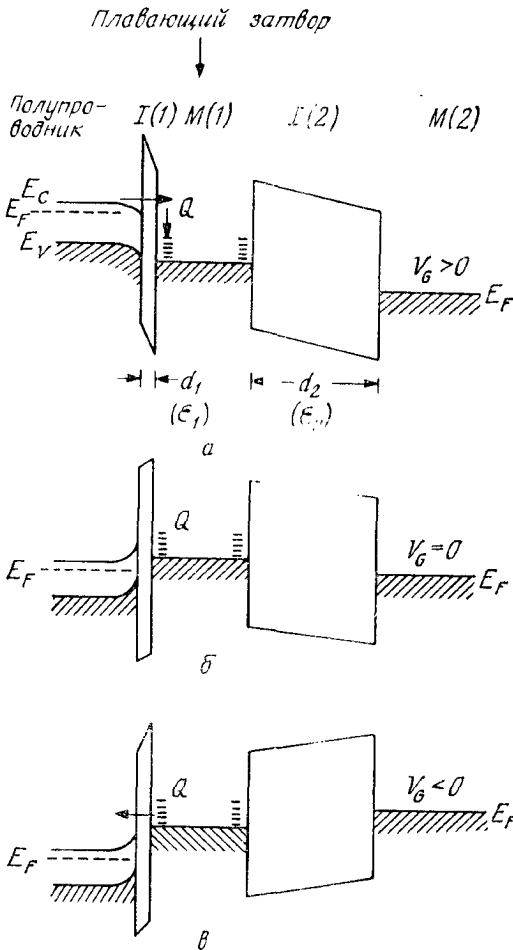


Рис. 61. Зонные диаграммы элемента памяти с плавающим затвором [71].

a — зарядка (операция записи);
 $б$ — хранение заряда; $в$ — разрядка плавающего затвора (операция стирания).

Напряжение V_G , приложенное к внешнему затвору структуры $M(2)$, делится между обоими диэлектрическими слоями (рис. 61, а):

$$V_G = V_1 + V_2 = d_1 \mathcal{E}_1 + d_2 \mathcal{E}_2, \quad (105)$$

где \mathcal{E}_1 и \mathcal{E}_2 — напряженность электрического поля в этих слоях. Поскольку, согласно закону Гаусса,

$$\epsilon_1 \mathcal{E}_1 = \epsilon_2 \mathcal{E}_2 + Q, \quad (106)$$

электрическое поле в каждом из слоев (например, в первом) можно записать в виде

$$\mathcal{E}_1 = \frac{V_G}{d_1 + d_2 (\epsilon_1 / \epsilon_2)} + \frac{Q}{\epsilon_1 + \epsilon_2 (d_1 / d_2)}, \quad (107)$$

где Q — заряд в плавающем затворе, а ϵ_1 и ϵ_2 — диэлектрические проницаемости первого и второго диэлектрических слоев.

Если токи в обоих диэлектрических слоях не равны друг другу, заряд плавающего затвора Q будет изменяться во времени:

$$Q(t) = \int_0^t [J_1(\mathcal{E}_1) - J_2(\mathcal{E}_2)] dt \quad [\text{Кл}\cdot\text{см}^{-2}], \quad (108)$$

где $J_1(\mathcal{E}_1)$ и $J_2(\mathcal{E}_2)$ — плотности тока в диэлектриках 1 и 2.

Обычно токи в диэлектриках сильно зависят от напряженности электрического поля. Так, например, при туннелировании по механизму Фаулера—Нордгейма

$$J = C_1 \mathcal{E}^2 \exp(-\mathcal{E}_0/\mathcal{E}), \quad (109)$$

где \mathcal{E} — электрическое поле, а C_1 и \mathcal{E}_0 — константы, зависящие от эффективной массы носителей и высоты барьера. Как уже говорилось в гл. 7, данный механизм проводимости играет основную роль в тонких слоях SiO_2 и Al_2O_3 . Когда транспорт заряда в диэлектрике осуществляется по механизму Пула—Френкеля, как, например, в Si_3N_4 , тогда

$$J = C_2 \mathcal{E} \exp[-q(\varphi_B - \sqrt{q\mathcal{E}/\pi\epsilon_1})/kT], \quad (110)$$

где C_2 — константа, пропорциональная плотности ловушек в диэлектрике; φ_B — глубина ловушки, а ϵ_1 — динамическая диэлектрическая проницаемость.

В результате разбаланса токов в диэлектриках 1 и 2 за время действия большого положительного напряжения V_G , приложенного к внешнему затвору (импульса записи), в плавающем затворе накапливается заряд Q (рис. 61, б), который сдвигает пороговое напряжение структуры на величину

$$\Delta V_T = -\frac{d_2}{\epsilon_2} Q. \quad (111)$$

Стирание хранящегося заряда осуществляется импульсом противоположной полярности $V_G < 0$ (рис. 61, в).

На рис. 62 приведены результаты теоретического расчета динамики записи заряда по формулам (105)—(109) для следующих значений параметров: $d_1 = 50 \text{ \AA}$, $\epsilon_1 = 3,85\epsilon_0$ (для SiO_2); $d_2 = 1000 \text{ \AA}$, $\epsilon_2 = 30\epsilon_0$ (для ZrO_2), при напряжении $V_G = 50 \text{ В}$ и в предположении $J_2 = 0$. Как видно из рис. 62, в начальной стадии записи заряд Q линейно увеличивается со временем, но затем эта зависимость насыщается. В свою очередь ток инжекции сначала слабо меняется, а затем начинает резко уменьшаться. Такое поведение можно объяснить следующим образом. В начале записи при $t = 0$ заряд Q также равен нулю, а электрическое поле в слое 1 максимально ($\mathcal{E}_1 = V_G/[d_1 + (\epsilon_1/\epsilon_2)d_2]$). Пока Q

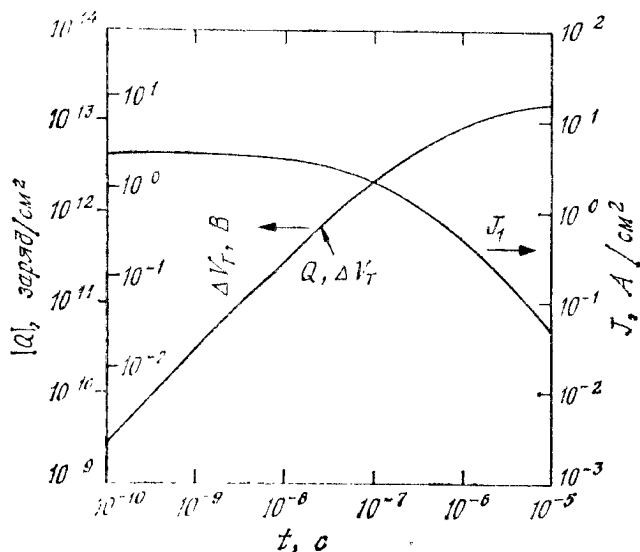


Рис. 62. Расчетные зависимости тока зарядки и накопленного заряда от времени [71].

мало и соответствующее уменьшение \mathcal{E}_1 незначительно, ток J_1 уменьшается слабо. Поэтому Q в начале процесса увеличивается линейно во времени. Однако в дальнейшем увеличение Q становится существенным и приводит к уменьшению поля \mathcal{E}_1 , в результате чего ток J_1 начинает заметно снижаться, а его интеграл (заряд Q) насыщаться.

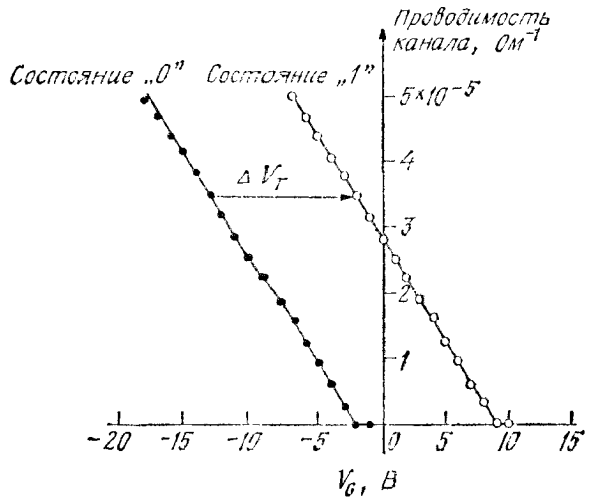
На рис. 62 приведен также соответствующий заряду Q сдвиг порогового напряжения ΔV_T (выражение (111)). В реальных приборах время зарядки (время записи), необходимое для сдвига порогового напряжения на 1 В, не превышает 10^{-7} с. Экспериментально сдвиг порогового напряжения можно определить, измеряя проводимость канала g_D . При малых напряжениях на стоке (для p -канального транзистора)

$$g_D = -\frac{Z}{L} \mu_p C_i (V_G - V_T), \quad V_G > V_T. \quad (112)$$

После изменений заряда на плавающем затворе Q (Q — электронный заряд и поэтому отрицателен) график зависимости $g_D(V_G)$ сдвигается вправо на величину ΔV_T (рис. 63) [73].

Мы подробно проанализировали работу предельно упрощенной запоминающей ячейки (рис. 60, а), чтобы проиллюстрировать основные принципы функционирования реальных энергонезависимых элементов памяти, о которых речь пойдет ниже: среди них

Рис. 63. Проводимость канала (состояние «0») и после (состояние «1») выполнения операции записи [73].



принцип туннельной инжекции, используемый в энергонезависимых запоминающих ячейках типа МДЮП, и принцип хранения заряда в плавающем затворе, на котором основана работа реальных ЗУ с плавающим поликремниевым затвором.

В первых программируемых ПЗУ со стиранием в качестве материала плавающего затвора использовался сильнолегированный поликремний (рис. 64, а). В этих приборах, получивших название МОП-ЗУ с плавающим затвором и лавинной инжекцией [74], поликремниевый затвор со всех сторон окружен слоем окисла. Довольно большая толщина слоя окисла ($d_1 \approx 1000 \text{ \AA}$) выбрана для того, чтобы исключить в окисле слабые места и коротки, через которые поликремниевый плавающий затвор мог бы быть электрически связан с подложкой. Инжекция заряда осуществляется горячими электронами, генерированными в области лавинного пробоя у стока (рис. 64, а), для чего сток смещается до напряжений, требуемых для пробоя. Горячие электроны при этом втягиваются в плавающий затвор электрическим полем, возникающим вследствие его емкостной связи с электродами стока и истока. Для стирания информационного заряда в такой структуре используется ультрафиолетовое или рентгеновское облучение. Электрическое стирание здесь невозможно, поскольку в структуре отсутствует внешний затвор.

При необходимости электрического стирания используется аналогичная ячейка памяти с двумя поликремниевыми затворами (рис. 64, б) [75]. Внешний затвор, кроме того, улучшает эффективность записи. Иногда в приборах этого типа изоляция между поликремниевыми затворами осуществляется слоем SiO_2 , обога-

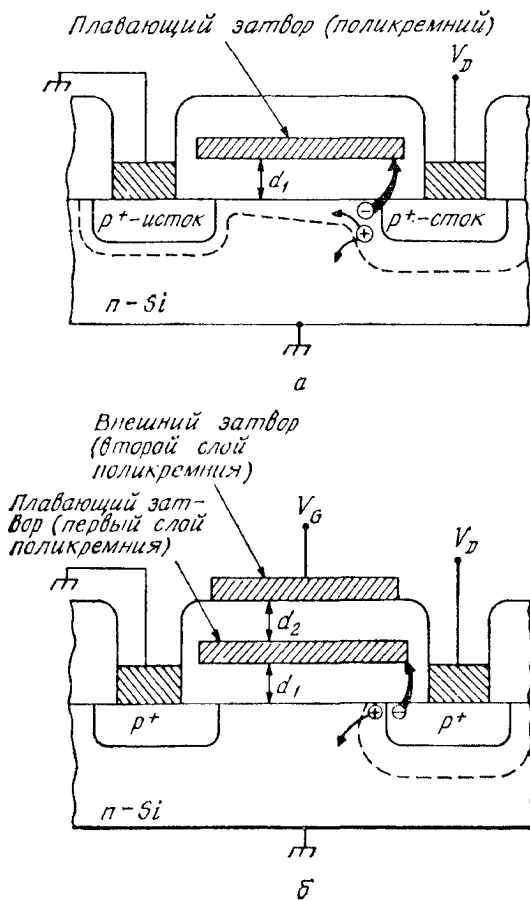
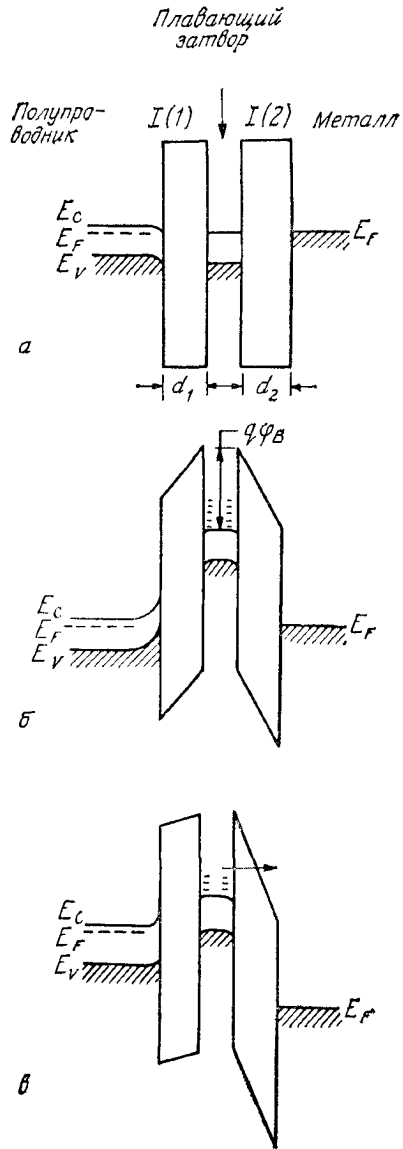


Рис. 64. Ячейка МОП-ЗУ с плавающим затвором и лавинной инжекцией (а) [74] и аналогичная ячейка с двумя затворами, обеспечивающая возможность электрического стирания (б) [75].

ценным кремнием [81]. На рис. 65 приведены зонные диаграммы, поясняющие работу такой ячейки с двумя затворами. Равновесное состояние структуры с $Q = 0$ (рис. 65, а) соответствует хранению информационного нуля «0». Состояние с накопленным в процессе лавинной инжекции зарядом на плавающем затворе (рис. 65, б) соответствует хранению информационной единицы «1». При больших положительных напряжениях на внешнем затворе накопленный заряд удаляется путем его вытягивания во внешний затвор (рис. 65, в), и элемент памяти возвращается в исходное состояние «0».

Рис. 65. Зонные диаграммы двухзатворного элемента памяти в состоянии «0» (а), состоянии «1» (б) и в режиме разрядки (стирания) (в).



На рис. 66 приведены теоретические кривые и соответствующие экспериментальные результаты, показывающие влияние напряжения на внешнем затворе на эффективность записи в такой структуре [75]. При фиксированном напряжении на стоке V_D напряжение на внешнем затворе V_G сначала способствует записи

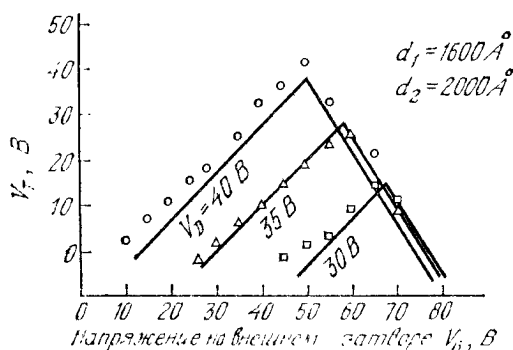


Рис. 66. Зависимость сдвига порогового напряжения при записи в двухзатворном элементе памяти от напряжения на внешнем затворе [75].

— теоретические данные;
 ○, △, □ — экспериментальные данные.

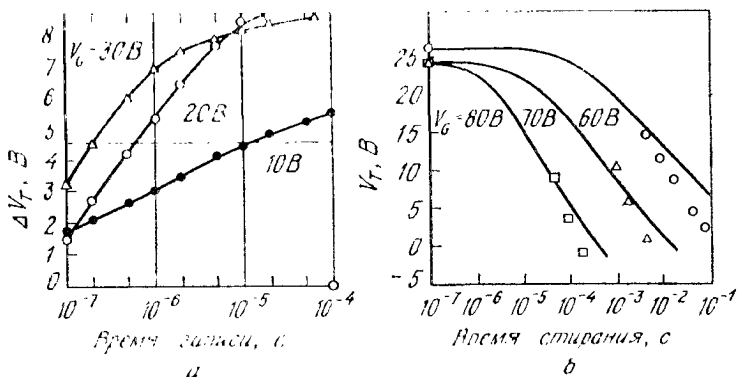


Рис. 67. Зависимость сдвига порогового напряжения от длительности импульсов записи и стирания в двухзатворном элементе памяти [75].

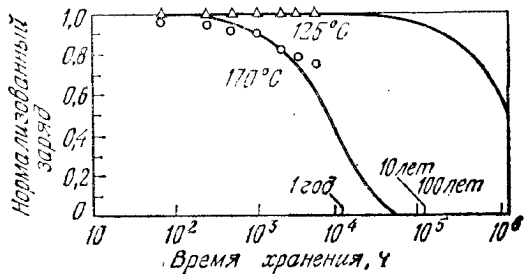
— теоретические данные; ○, △, □ экспериментальные данные.

(V_T увеличивается), усиливая дрейф горячих электронов через первый слой SiO_2 на плавающий затвор. Однако при достаточно больших V_G начинает преобладать конкурирующий процесс полевого вытягивания заряда из плавающего затвора во внешний. Поэтому с определенного значения V_G эффективность записи падает (V_T уменьшается). На рис. 67 приведены зависимости изменения порогового напряжения от времени записи и стирания для той же, что и на рис. 66, структуры. Отметим, что при $V_G = 20 \text{ В}$ для сдвига порога на 5 В требуется время $\sim 10^{-6} \text{ с}$, в то время как в аналогичных структурах без внешнего затвора (рис. 64, а) для этого необходимы времена, на два порядка большие. Кроме того, скорость операции стирания, естественно, увеличивается с ростом напряжения затвора V_G .

Важной характеристикой энергонезависимых элементов памяти является время хранения информации. Оно определяется

Рис. 68. Кривые стекания хранящегося заряда для двух температур [72].

— теоретические данные
 Δ , \circ экспериментальные данные.



как время, за которое записанный заряд уменьшается в результате стекания на 50 %:

$$t_R = \ln 2 / \{ \nu \exp(-q\phi_B/kT) \}. \quad (113)$$

Здесь ν — частота диэлектрической релаксации, а $q\phi_B$ — высота энергетического барьера (рис. 65, б). Кривые стекания заряда, рассчитанные при $q\phi_B = 1,7$ эВ для температуры 125 и 170 °С, приведены на рис. 68 вместе с соответствующими экспериментальными данными. Как следует из рис. 68, время хранения информации при $T = 125$ и 170 °С составляет примерно 100 лет и 8000 ч соответственно.

8.6.2. МДОП-структуры

Среди различных запоминающих устройств со структурой металл—диэлектрик— SiO_2 — Si (МДОП) наиболее популярны МНОП-структуры (металл— Si_3N_4 — SiO_2 — Si). Кроме того, известны МДОП-структуры, в которых вместо нитрида кремния используются другие диэлектрические пленки, такие, как окись алюминия, окись тантала и окись титана. К этому типу приборов относится также структура, содержащая только один слой двуокиси кремния, во внешнюю часть которого имплантированы ионы золота (или других металлов) [76]. Такая имплантация изменяет проводимость в соответствующей части окисла, а также приводит к возникновению пограничных центров, способных захватывать и хранить заряд.

На рис. 69 приведены зонные диаграммы МНОП-структуры, соответствующие операциям записи и стирания информационного заряда [77]. Ток в окисле J_0 обусловлен туннелированием Фаулера—Нордгейма, в то время как ток через нитрид J_N соответствует эмиссии Пула—Френкеля. Уравнения, описывающие процесс накопления заряда в МНОП-структурах, близки к системе уравнений (105)—(108). Различие состоит в том, что электронные ловушки, захватывающие заряд в нитриде у границы с окислом, вероятно, непрерывно распределены по энергии связи, а также

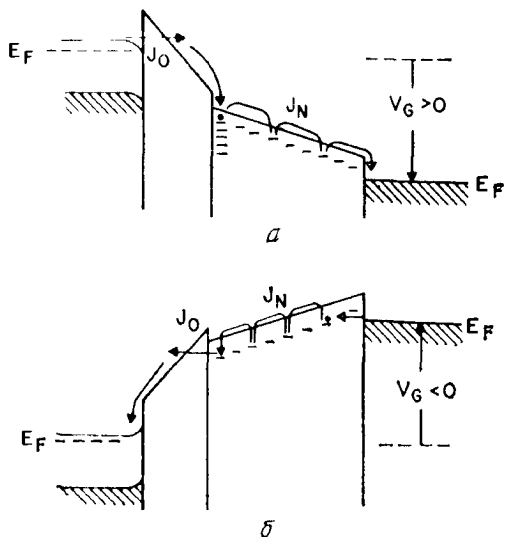


Рис. 69. Операции записи (а) и стирания (б) в МНОП-структуре [77].

по толщине приграничной области нитрида кремния. Этот факт тем или иным способом учитывается в различных теоретических моделях, предложенных в настоящее время для объяснения экспериментальных характеристик записи в МНОП-структурах. Данные, приведенные на рис. 70, представляют собой пример такого сопоставления, демонстрирующий неплохое соответствие теоретических зависимостей и результатов эксперимента [78].

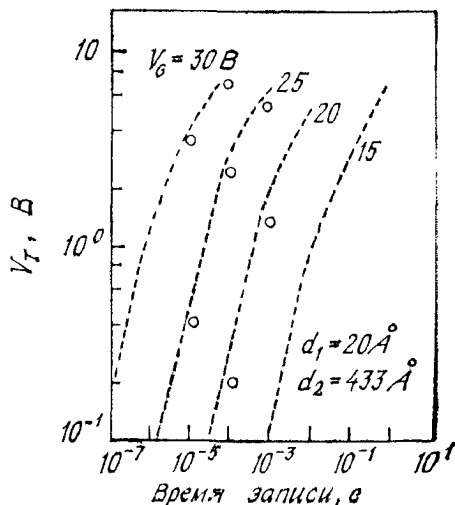


Рис. 70. Рассчитанные (— — —) и измеренные (°) сдвиги порога в МНОП-структурах [78].

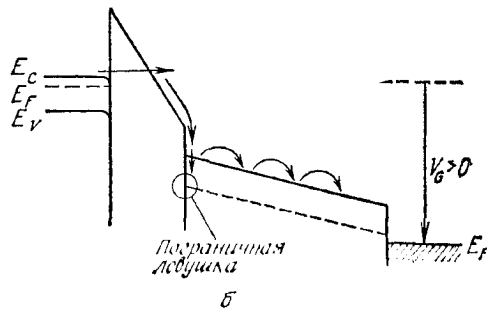
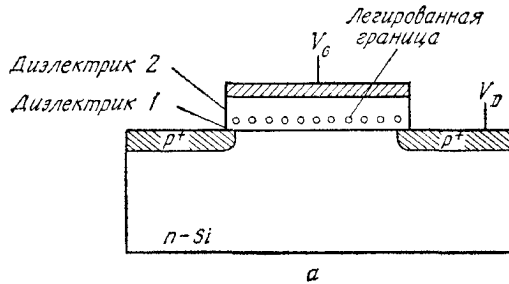


Рис. 71. Элемент памяти с легированной границей раздела диэлектриков [79].
 а — поперечное сечение прибора; б — зонная диаграмма при наличии напряжения на затворе.

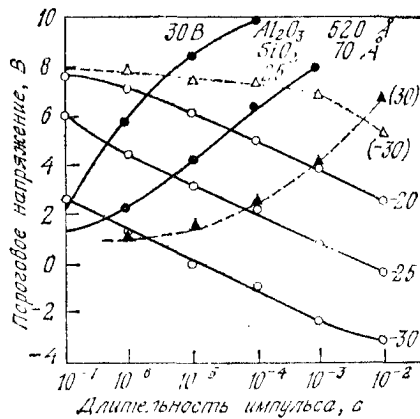


Рис. 72. Зависимости сдвига порогового напряжения от длительности импульсов записи и стирания в структуре с двойным диэлектриком [79].
 ▲ запись и Δ стирание в нелегированной структуре; ● запись и ○ стирание в структуре с легированной границей раздела ($1,5 \cdot 10^{18}$ см⁻²).

Как видно из рис. 70, время записи, необходимое для обеспечения заданного сдвига порога, быстро уменьшается с ростом V_G . В свою очередь при фиксированном V_G сдвиг порога ΔV_T при больших временах записи стремится к насыщению (так же, как на рис. 62).

В качестве одного из способов улучшения рабочих характеристик МДОП-элементов памяти было предложено легировать границу раздела между диэлектриками (например, вольфрамом) для создания дополнительных центров захвата заряда (рис. 71, а) [79]. Оказывается, что соответствующее значительное увеличение плотности поверхностных ловушек позволяет существенно снизить времена записи и стирания (рис. 71, б). Из сравнения экспериментальных результатов, приведенных на рис. 72, следует, что времена записи ($V_G = 30$ В) и времена стирания ($V_G = -30$ В) в структурах с легированной границей раздела диэлектриков примерно на три порядка меньше соответствующих времен в обычных МНОП-структурах.

ЛИТЕРАТУРА

1. Lilienfeld J. E. U. S. Patent 1745175 (1930).
2. Heil O. British Patent 439457 (1935).
3. Shockley W., Pearson G. L. Modulation of Conductance of Thin Films of Semiconductors by Surface Charges, *Phys. Rev.*, **74**, 232 (1948).
4. Kahng D., Atalla M. M. Silicon — Silicon Dioxide Field Induced Surface Devices, IRE Solid-State Device Res. Conf., Carnegie Institute of Technology, Pittsburgh, Pa., 1960; Kahng D. A Historical Perspective on the Development of MOS Transistors and Related Devices, *IEEE Trans. Electron Devices*, **ED-23**, 655 (1976).
5. Ihantola H. K. Design Theory of a Surface Field-Effect Transistor, Stanford Electron. Lab. Tech. Rep. No. 1661-1 (1961).
6. Ihantola H. K. J., Moll J. L. Design Theory of a Surface Field-Effect Transistor, *Solid State Electron.*, **7**, 423 (1964).
7. Sah C. T. Characteristics of the Metal — Oxide — Semiconductor Transistors, *IEEE Trans. Electron Devices*, **ED-11**, 324 (1964).
8. Holstein S. R., Heiman F. P. The Silicon Insulated-Gate Field-Effect Transistor, *Proc. IEEE*, **51**, 1190 (1963).
9. Walmark J. T., Johnson H. Field-Effect Transistors, Physics, Technology, and Applications, Prentice-Hall, Englewood Cliffs, N. J., 1966.
10. Richman P. MOSFET's and Integrated Circuits, Wiley, N. Y., 1973.
11. Brews J. R. Physics of the MOS Transistor, in Kahng D., Ed., *Applied Solid State Science*, Suppl. 2A, Academic, N. Y., 1981.
12. Chang L. L., Yu H. N. The Germanium Insulated-Gate Field-Effect Transistor (FET), *Proc. IEEE*, **53**, 316 (1965).
13. Wilmsen C. W., Szpak S. MOS Processing for III—V Compound Semiconductors: Overview and Bibliography, *Thin Solid Films*, **46**, 17 (1977).
14. Looking Ahead to the Year 2000, Technology, *Electronics*, **53**, (9) 530 (1980).
15. Moore G. VLSI: Some Fundamental Challenges, *IEEE Spectrum*, **16** (4) 30 (1979).
16. Pao H. C., Sah C. T. Effects of Diffusion Current on Characteristics of Metal — Oxide (Insulator) — Semiconductor Transistors (MOST), *IEEE Trans. Electron Devices*, **ED-12**, 139 (1965).

17. Grove A. S., Fitzgerald D. J. Surface Effects on $p - n$ Junctions: Characteristics of Surface Space-Charge Regions under Nonequilibrium Conditions, *Solid State Electron.*, **9**, 783 (1966).
18. Brews J. R. A Charge-Sheet Model of the MOSFET, *Solid State Electron.*, **21**, 345 (1978).
19. Brews J. R. Subthreshold Behavior of Uniformly and Nonuniformly Doped Long-Channel MOSFET, *IEEE Trans. Electron Devices*, **ED-26**, 1282 (1979).
20. Barron M. B. Low Level Currents in Insulated Gate Field Effect Transistors, *Solid State Electron.*, **15**, 293 (1972).
21. Gosney W. M. Subthreshold Drain Leakage Current in MOS Field-Effect Transistors, *IEEE Trans. Electron Devices*, **ED-19**, 213 (1972).
22. Fichtner W., Potzl H. W. MOS Modeling by Analytical Approximations. I. Subthreshold Current and Threshold Voltage, *Int. J. Electron.*, **46**, 33 (1979).
23. Troutman R. R. Subthreshold Design Considerations for IGFET's, *IEEE J. Solid State Circuits*, **SC-9**, 55 (1974).
24. Sabnis A. G., Clemens J. T. Characterization of the Electron Mobility in the Inverted (100) Si Surface, *IEEE Tech. Dig. Int. Electron Device Meet.*, 1979, p. 18.
25. Coen R. W., Muller R. S. Velocity of Surface Carriers in Inversion Layers on Silicon, *Solid State Electron.*, **23**, 35 (1980).
26. Cooper J. A., Nelson D. F. Observation of High Field Transport in Si Inversion Layer by a Time-of-Flight Technique, *IEEE Device Res. Conf.*, June 23, 1980.
27. Thornter K. K. Relation of Drift Velocity to Low-Field Mobility and High Field Saturation Velocity, *J. Appl. Phys.*, **51**, 2127 (1980).
28. Yamaguchi K. Field-Dependent Mobility Model for Two-Dimensional Numerical Analysis of MOSFETs, *IEEE Trans. Electron Devices*, **ED-26**, 1068 (1979).
29. Vadasz L., Grove A. S. Temperature Dependence of MOS Transistor Characteristics below Saturation, *IEEE Trans. Electron Devices*, **ED-13**, 863 (1966).
30. Wang R., Dunkley J., DeMassa T. A., Jelsma L. F. Threshold Voltage Variations with Temperature in MOS Transistors, *IEEE Trans. Electron Devices*, **ED-18**, 386 (1971).
31. Gaensslen F., Rideout V. L., Walker E. J., Walker J. J. Very Small MOSFET's for Low-Temperature Operation, *IEEE Trans. Electron Devices*, **ED-24**, 218 (1977).
32. Fischer W. Equivalent Circuit and Gain of MOS Field-Effect Transistors, *Solid State Electron.*, **9**, 71 (1966).
33. Gallagher R. C., Corak W. S. A Metal — Oxide — Semiconductor (MOS) Hall Element, *Solid State Electron.*, **9**, 571 (1966).
34. Wittmack K., Maul J., Schulz F. Ion Implantation in Semiconductor and Other Materials, Plenum, N. Y., 1973.
35. Rideout V. L., Gaensslen F. H., LeBlanc A. Device Design Consideration for Ion Implanted n -channel MOSFETs, *IBM J. Res. Dev.*, p. 50 (Jan. 1975).
36. Brews J. R. Threshold Shifts Due to Nonuniform Doping Profiles in Surface Channel MOSFETs, *IEEE Trans. Electron Devices*, **ED-26**, 1696 (1979).
37. Merckel G. Ion Implanted MOS Transistors — Depletion Mode Devices, in Van de Wiele F., Engle W. L., Jespers P. G., Eds., *Process and Device Modeling for IC Design*, Noordhoff, Leyden, 1977.
38. Huang J. S. T., Taylor G. W. Modeling of an Ion-Implanted Silicon-Gate Depletion-Mode IGFET, *IEEE Trans. Electron Devices*, **ED-22**, 995 (1975).
39. Hendrikson T. E. A Simplified Model for Subpinchoff Condition in Depletion Mode IGFET's, *IEEE Trans. Electron Devices*, **ED-25**, 435 (1978).
40. Nishituchi K., Oka H., Nakamura T., Ishikawa H., Shinoda M. A Normally-Off Type Buried Channel MOSFET for VLSI Circuits, *IEEE Tech. Dig., Int. Electron Device Meet.*, 1978, p. 26.

41. Brews J. R., Fichtner W., Nicollian E. H., Sze S. M. Generalized Guide for MOSFET Miniaturization, *IEEE Electron Devices Lett.*, EDL-1, 2 (1980).
42. Yau L. D., A Simple Theory to Predict the Threshold Voltage of Short-Channel IGFET's, *Solid State Electron.*, 17, 1059 (1974).
43. Taylor G. W. Subthreshold Conduction in MOSFET's, *IEEE Trans. Electron Devices*, ED-25, 337 (1978).
44. Merckel G. Short Channels — Scaled Down MOSFET's, in Ref. 37, p. 705.
45. Fichtner W. Scaling Calculation for MOSFET's, IEEE Solid State Circuits and Technology Workshop on Scaling and Microlithography, N. Y., Apr. 22, 1980.
46. Taylor G. W. The Effects of Two-Dimensional Charge Sharing on the Above Threshold Characteristics of Short-Channel IGFET's, *Solid State Electron.*, 22, 701 (1979).
47. Yau L. D. Simple I—V Model for Short-Channel IGFET's in the Triode Region, *Electron. Lett.*, 11, 44 (1975).
48. Fukuma M., Matsuura M. A Simple Model for Short Channel MOSFET's, *Proc. IEEE*, 65, 1212 (1977).
49. Kamata T., Tanabashi K., Kobayashi K. Substrate Current Due to Impact Ionization in MOSFET, *Jpn. Appl. Phys.*, 15, 1127 (1976).
50. Sun E., Moll J., Berger J., Alders B. Breakdown Mechanism in Short-Channel MOS Transistors, IEEE Tech. Dig. Int. Electron Device Meet., Washington D. C., 1978, p. 478.
51. Ning T. H., Osburn C. M., Yu H. N. Effect of Electron Trapping on IGFET Characteristics, *J. Electron. Mater.*, 6, 65 (1977).
52. Matsunaga J., Konaka M., Kohyama S., Iizuku H. 11th Int. Conf. Solid State Devices, Aug. 27, 1979, p. 45.
53. Nicollian E. H., Berglund C. N. Avalanche Injection of Electrons into Insulating SiO₂ Using MOS Structures, *J. Appl. Phys.*, 41, 3052 (1970).
54. Dennard R. H., Gaensslen F. H., Yu H., Rideout V. L., Bassens E., LeBlanc A. R. Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions, *IEEE J. Solid State Circuits*, SC-9, 256 (1974).
55. Shannon J. M., Stephen J., Freeman J. H. MOS Frequency Soars with Ion-Implanted Layers, *Electronics*, p. 96 (Feb. 3, 1969).
56. Nihira H., Konaka M., Iwai H., Nishi Y. Anomalous Drain Current in n-MOSFET's and Its Suppression by Deep Ion Implantation, IEEE Tech. Dig., Int. Electron Device Meet., 1978, p. 487.
57. Tarui Y., Hayashi Y., Sekigawa T. Diffusion Self-Aligned Enhance-Depletion MOS-IC, Proc. 2nd Conf. Solid State Devices, *Suppl. J. Jpn. Soc. Appl. Phys.*, 40, 193 (1971).
58. Tihanyi J., Widmann D. DIMOS — A Novel IC Technology with Submicron Effective Channel MOSFET's, IEEE Tech. Dig., Int. Electron Device Meet., 1977, p. 399.
59. Masuhara T., Muller R. S. Analytical Technique for the Design of DMOS Transistors, *Jpn. J. Appl. Phys.*, 16, 173 (1976).
60. Pocha M. D., Gonzalez A. G., Dutton R. W. Threshold Voltage Controllability in Double-Diffused MOS Transistors, *IEEE Trans. Electron. Devices*, ED-21, 778 (1974).
61. Nishimatsu S., Kawamoto Y., Masuda H., Hori R., Minato O. Grooved Gate MOSFET, *Jpn. J. Appl. Phys.*, 16; *Suppl. 16-1*, 179 (1977).
62. Lepsciter M. P., Sze S. M. SB-IGFET: An Insulated-Gate Field-Effect Transistor Using Schottky Barrier Contacts as Source and Drain, *Proc. IEEE*, 56, 1088 (1968).
63. Weimer P. K. The TFT — A New Thin-Film Transistor, *Proc. IRE*, 50, 1462 (1962).
64. Tihanyi J., Schlotterer H. Influence of the Floating Substrate Potential on the Characteristics of ESF1 MOS Transistors, *Solid State Electron.*, 18, 309 (1975).

65. Lam H. W., Tasch A. F., Holloway T. C., Lee K. F., Gibbons J. F. Ring Oscillators Fabricated in Laser-Annealed Silicon-on-Insulator, *IEEE Electron Devices Lett.*, **EDL-1**, 99 (1980).
66. Geis M. W., Flanders D. C., Antoniadis D. A., Smith H. I. Crystalline Silicon on Insulators by Graphoepitaxy, IEEE Tech. Dig., Int. Electron Device Meet., 1979, p. 210.
67. Holmes F. E., Salama C. A. T. VMOS — A New MOS Integrated Circuit Technology, *Solid State Electron.*, **17**, 791 (1974).
68. Salama C. A. T. A New Short Channel MOSFET Structure (UMOST), *Solid State Electron.*, **20**, 1003 (1977).
69. Lidow A., Herman T., Collins H. W. Power MOSFET Technology, IEEE Tech. Dig., Int. Electron Device Meet., 1979, p. 79.
70. Collins H. W., Pelly B. HEXFET, A New Power Technology, Cutts On-Resistance, Boosts Ratings, *Electron. Des.*, **17** (12), 36 (1979).
71. Kahng D., Sze S. M. A Floating Gate and Its Application to Memory Devices, *Bell Syst. Tech. J.*, **46**, 1283 (1967).
72. Nishi Y., Iizuka H. Nonvolatile Memories, in Kahng D., Ed., Applied Solid State Science, Suppl. 2A, Academic, N. Y., 1981.
73. Card H. C., Worrall A. G. Reversible Floating-Gate Memory, *J. Appl. Phys.*, **44**, 2326 (1973).
74. Frohman-Bentchkowsky D. FAMOS — A New Semiconductor Charge Storage Device, *Solid State Electron.*, **17**, 517 (1974).
75. Iizuka H., Masuoka F., Sato T., Ishikawa M. Electrically Alterable Avalanche-Injection-Type MOS Read-Only Memory with Stacked-Gate Structures, *IEEE Trans. Electron Devices*, **ED-23**, 379 (1976).
76. Chen L. I., Pickar K. A., Sze S. M. Carrier Transport and Storage Effects in Au Ion Implanted SiO₂ Structures, *Solid State Electron.*, **15**, 979 (1972).
77. Frohman-Bentchkowsky D. The Metal — Nitride — Oxide — Silicon (MNOS) Transistor Characteristics and Applications, *Proc. IEEE*, **58**, 1207 (1970).
78. Card H. C., Elmasry M. I. Functional Modeling of Nonvolatile MOS Memory Devices, *Solid State Electron.*, **19**, 863 (1976).
79. Kahng D., Sundburg W. J., Boulin D. M., Ligenza J. R. Interfacial Dopants for Dual-Dielectric Charge-Storage Cells, *Bell Syst. Tech. J.*, **53**, 1723 (1974).
80. Gibbons J. J., Lee K. F. One-Gate-Wide CMOS Inverter on Laser-Recrystallized Polysilicon, *IEEE Electron Devices Lett.*, **EDL-1**, 117 (1980).
81. DiMaria D. J., DeMeyer K. M., Dong D. W. Electrically-Alterable Memory Using a Dual Electron Injector Structure, *IEEE Electron Devices Lett.*, **EDL-1**, 179 (1980).

Часть IV

Полупроводниковые приборы СВЧ-диапазона

Глава 9

ТУННЕЛЬНЫЕ ПРИБОРЫ

9.1. ВВЕДЕНИЕ

В этой главе, а также в гл. 10 и 11 рассмотрены некоторые полупроводниковые приборы СВЧ-диапазона, перекрывающего область частот $\sim 0,1$ ГГц (10^8 Гц) — 1000 ГГц (300 см — 0,3 мм). Этот диапазон в свою очередь подразделяется на отдельные полосы частот. Названия полос и их границы приведены в табл. 1. Полоса частот 30—300 ГГц называется также диапазоном миллиметровых волн, поскольку длины волн, соответствующие этим частотам, лежат между 10 и 1 мм. Полоса более высоких частот называется полосой субмиллиметровых волн. В 1970 г. были приняты новые обозначения полос частот СВЧ-диапазона (табл. 2). Во избежание недоразумений рекомендуется одновременно указывать как название полосы частот, так и ее границы. В СВЧ-диапазоне могут работать многие полупроводниковые приборы, рассмотренные в предыдущих главах. В табл. 3 указаны основные приборы СВЧ-диапазона и принципы их работы.

В данной главе описаны приборы, работа которых основана на явлении квантовомеханического туннелирования. Первая работа, подтверждающая реальность создания туннельных приборов, была посвящена туннельному диоду, называемому также диодом Эсаки, и опубликована Л. Эсаки в 1958 г. [1]. Эсаки в процессе изучения внутренней полевой эмиссии в вырожденном германиевом p - n -переходе обнаружил «аномальную» вольт-амперную характеристику: дифференциальное сопротивление на одном из участков характеристики было отрицательным. Этот эффект он объяснил с помощью концепции квантовомеханического туннелирования и при этом получил приемлемое согласие между теоретическими и экспериментальными результатами.

В явлении туннелирования главную роль играют основные носители. Время туннелирования носителей через потенциальный барьер не описывается на привычном языке времени пролета ($\tau = W/v$, где W — ширина барьера, v — скорость носителей); оно определяется с помощью вероятности квантовомеханического

Таблица 1. Названия диапазонов
длин волн и полос частот [63]

Название диапазона (полосы)	Размер волновода, см	Полоса частот, ГГц
L	—	1,0—2,6
S	7,6×3,8	2,60—3,95
G	5×2,5	3,95—5,85
C	4,4×2,3	4,90—7,05
J	3,8×1,9	5,30—8,20
H	3,2×1,3	7,05—10,00
X	2,5×1,25	8,20—12,40
M	2,1×1,2	10,00—15,00
P	1,8×1,0	12,40—18,00
N	1,5×0,85	15,00—22,00
Ku	—	15,30—18,00
K	1,2×0,65	18,00—26,50
R	0,9×0,56	26,50—40,00
Миллиметровый		>30—300
Субмиллиметровый		>300

Таблица 2. Новые названия диапазонов
и полос частот [64]

Название диапазона	Полоса частот, ГГц
A	0,100—0,250
B	0,250—0,500
C	0,500—1,000
D	1,000—2,000
E	2,000—3,000
F	3,000—4,000
G	4,000—6,000
H	6,000—8,000
I	8,000—10,000
J	10,000—20,000
K	20,000—40,000
L	40,000—60,000
M	60,000—100,000
Миллиметровый	>30—300
Субмиллиметровый	>300

перехода в единицу времени. Эта вероятность пропорциональна $\exp[-2k(0)W]$, где $k(0)$ — среднее значение волнового вектора в процессе туннелирования, приходящееся на один носитель с нулевым поперечным импульсом и энергией, равной энергии Ферми [2]. Отсюда следует, что время туннелирования пропор-

Таблица 3. Полупроводниковые приборы СВЧ-диапазона

Прибор	Принцип работы
Варактор (гл. 2) $p-i-n$ -диод (гл. 2)	Зависимость емкости от напряжения смещения Емкость практически постоянна, высокое напряжение пробоя
Биполярный транзистор (гл. 3) Диод с точечным контактом (гл. 5) Диод Шоттки (гл. 5)	Совместное участие электронов и дырок в транспортных процессах Малая площадь, малая емкость Ток основных носителей, гермоэлектронная эмиссия
Полевой транзистор с $p-n$ -переходом (гл. 6) Полевой транзистор с барьером Шоттки (гл. 6) МОП-транзистор (гл. 8)	Основные носители; ток, модулируемый напряжением на затворе Основные носители; ток, модулируемый смещением на барьере Шоттки Неосновные носители движутся в инверсионном канале у поверхности
Туннельный диод (гл. 9)	Туннелирование в прямосмещенном p^+-n^+ -переходе, отрицательное дифференциальное сопротивление
Обращенный диод (гл. 9)	Туннелирование в p^+-n^+ -переходе при обратном или почти нулевом смещении
ЛПД (гл. 10)	Лавинное умножение и пролетные эффекты для генерации мощного излучения
ИПД (гл. 10) Пролетный диод с захваченным объемным зарядом лавины (гл. 10) Прибор на эффекте междолинного перехода электронов (гл. 11)	Эмиссия через барьер и пролетные эффекты Захват плазмы в лавинно-пролетном диоде Электроны переходят из низкоэнергетических состояний с высокой подвижностью в высокоэнергетические состояния с низкой подвижностью

ционально $\exp [2k(0)W]$. Оно очень мало, и поэтому туннельные приборы можно использовать в диапазоне миллиметровых волн.

9.2. ТУННЕЛЬНЫЙ ДИОД

Благодаря высокой надежности и совершенству технологии изготовления туннельные диоды используются в специальных СВЧ-приборах с низким уровнем мощности, таких, как гетеродин и схемы для синхронизации частоты. В данном разделе описаны характеристики туннельных диодов на постоянном токе и в СВЧ-диапазоне.

Туннельный диод представляет собой простой $p-n$ -переход, обе стороны которого вырождены (т. е. очень сильно легированы примесями). На рис. 1 приведена энергетическая диаграмма туннельного диода, находящегося в состоянии термического

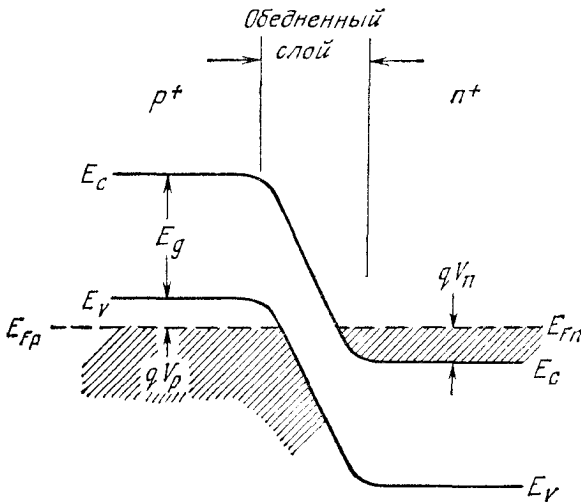


Рис. 1. Энергетическая диаграмма туннельного диода в состоянии термического равновесия. V_p и V_n — степени вырождения p -области и n -области соответственно.

равновесия. В результате сильного легирования уровень Ферми проходит внутри разрешенных зон. Степени вырождения V_p и V_n обычно составляют несколько kT/q , а ширина обедненного слоя $\sim 100 \text{ \AA}$ и меньше, т. е. намного меньше, чем в обычном p — n -переходе.

На рис. 2, а приведена типичная статическая вольт-амперная характеристика туннельного диода, из которой видно, что ток в обратном направлении (потенциал p -области отрицателен по отношению к потенциалу n -области) монотонно увеличивается. В прямом направлении ток сначала возрастает до максимального значения (пикового значения I_p) при напряжении V_p , а затем уменьшается до минимальной величины I_V при напряжении V_V . При напряжениях, превышающих V_V , ток возрастает с ростом напряжения по экспоненциальному закону. Полный статический ток диода представляет собой сумму тока туннелирования из зоны в зону, избыточного и диффузионного тока (рис. 2, б).

Сначала проведем качественный анализ туннельных процессов при температуре абсолютного нуля для упрощенной зонной структуры (рис. 3) [3]. Отметим, что уровни Ферми проходят внутри разрешенных зон полупроводника, и в состоянии термического равновесия (рис. 3, б) уровень Ферми постоянен по всему полупроводнику. Выше уровня Ферми все состояния по обеим сторонам перехода оказываются пустыми, а ниже уровня Ферми все разрешенные состояния по обеим сторонам перехода запол-

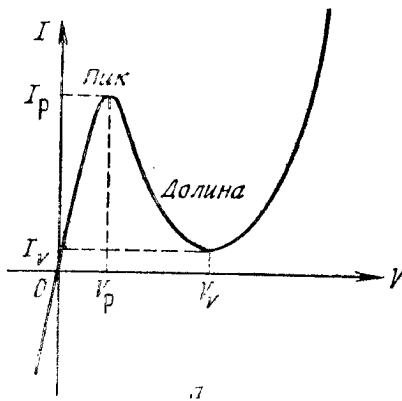
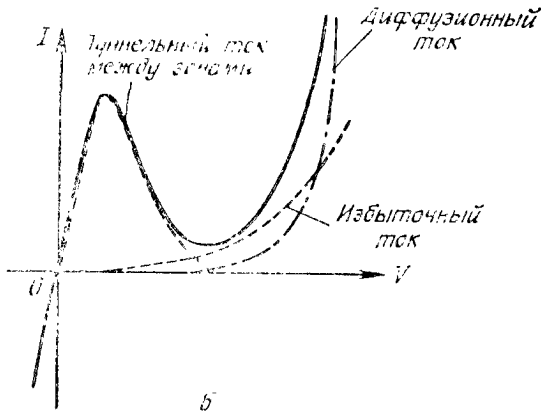


Рис. 2. Типичная статическая вольт-амперная характеристика туннельного диода (а) и три компоненты полного тока в туннельных диодах (б).



лены электронами. Поэтому в отсутствие приложенного напряжения туннельный ток не протекает.

При подаче напряжения на переход электроны могут туннелировать из валентной зоны в зону проводимости или наоборот. Для протекания туннельного тока необходимо выполнение следующих условий: 1) энергетические состояния на той стороне перехода, откуда туннелируют электроны, должны быть заполнены; 2) на другой стороне перехода энергетические состояния с той же энергией должны быть пустыми; 3) высота и ширина потенциального барьера должны быть достаточно малыми, чтобы существовала конечная вероятность туннелирования; 4) должен сохраняться квазинимпульс.

На рис. 3, а показано, как туннелируют электроны из валентной зоны в зону проводимости при обратном напряжении на диоде. Соответствующая величина тока отмечена точкой на вольт-ам-

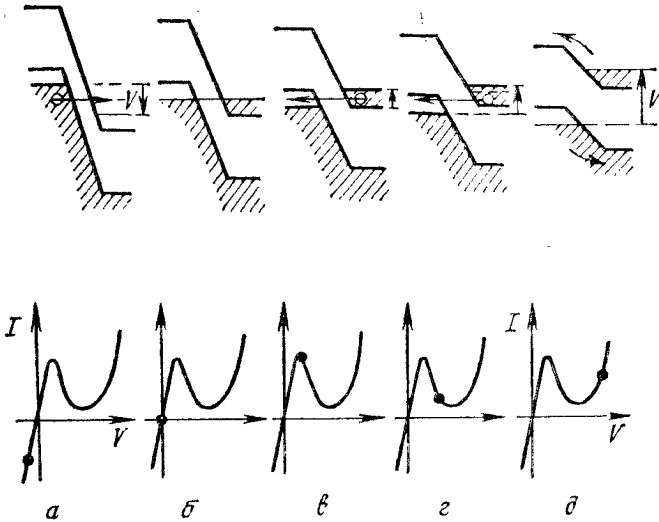


Рис. 3. Упрощенные энергетические диаграммы туннельного диода [3].

a — при обратном смещении; *б* — в тепловом равновесии, при нулевом смещении; *в* — при прямом смещении, обеспечивающем максимальное значение тока; *г* — при прямом смещении, соответствующем протеканию тока, близкого к доливному; *д* — при прямом смещении, соответствующем диффузионному току.

перной характеристике. При прямом напряжении (рис. 3, *в*) существует диапазон энергий, при которых состояния в *n*-области заполнены, а разрешенные состояния в *p*-области пусты. Естественно, что при этом электроны могут туннелировать из *n*-области в *p*-область. При увеличении прямого напряжения число разрешенных пустых состояний в *p*-области, в которые могут туннелировать электроны из *n*-области, уменьшается (рис. 3, *г*). Если же прямое напряжение имеет такое значение, что зоны «не перекрываются», т. е. энергия дна зоны проводимости точно совпадает с энергией потолка валентной зоны, то неразрешенные пустые состояния, соответствующие заполненным состояниям, отсутствуют. Следовательно, в этой точке туннельный ток должен исчезать. При дальнейшем увеличении напряжения будет протекать обычный диффузионный ток (рис. 3, *д*), который экспоненциально возрастает с ростом напряжения. Таким образом, следует ожидать, что при увеличении прямого напряжения туннельный ток сначала возрастает от нуля до максимального значения I_p , а затем уменьшается до нуля, когда приложенное прямое напряжение $V = V_n + V_p$ (рис. 1), где V_n — степень вырождения *n*-области ($V_n \equiv (E_{Fn} - E_C)/q$), а V_p — степень вырождения *p*-области ($V_p \equiv (E_V - E_{Fp})/q$). Падающий участок

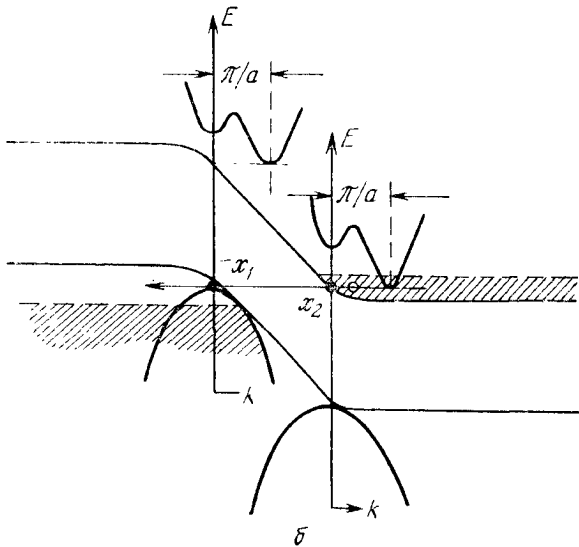
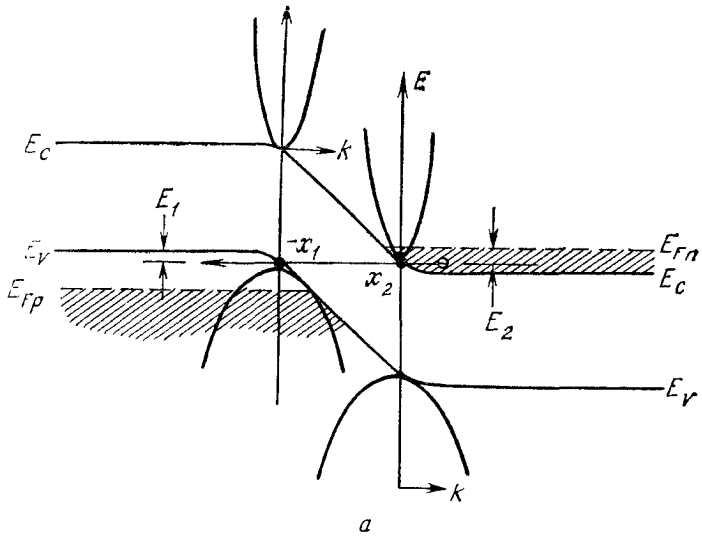


Рис. 4. Суперпозиция в классических точках поворота ($-x_1$ и x_2) структуры зон в импульсном пространстве $E - k$ и энергетической диаграммы туннельного перехода в координатном пространстве $E - x$.
а — случай прямого туннелирования ($k_{\text{мин}} = k_{\text{макс}}$); **б** — случай непрямого туннелирования ($k_{\text{мин}} \neq k_{\text{макс}}$).

вольт-амперной характеристики соответствует области отрицательного дифференциального сопротивления.

Процесс туннелирования может быть прямым и непрямым. Случай прямого туннелирования показан на рис. 4, а, где структура зон в импульсном пространстве $E - k$ в классических точках поворота наложена на энергетическую диаграмму туннельного перехода в координатном пространстве $E - x$. При такой структуре зон электроны могут туннелировать из окрестности минимума зоны проводимости в окрестность максимума валентной зоны, сохраняя значение импульса. Таким образом, для того чтобы происходило прямое туннелирование, положения дна зоны проводимости и потолка валентной зоны в пространстве импульсов должны совпадать. Это условие выполняется в полупроводниках с прямой запрещенной зоной (в таких, как GaAs и GaSb). Оно может выполняться также в полупроводниках с непрямой запрещенной зоной (например, в Ge) при достаточно больших приложенных напряжениях, таких, что максимум валентной зоны (точка Г) [4] находится на одном уровне с непрямым минимумом зоны проводимости (точка Г).

Непрямое туннелирование происходит в том случае, когда минимум зоны проводимости и максимум валентной зоны расположены при различных значениях импульса (рис. 4, б). Для выполнения закона сохранения импульса разность между импульсами, соответствующими минимуму зоны проводимости и потолку валентной зоны, должна передаваться рассеивателям, таким, как фононы или примесные центры. Законы сохранения энергии и импульса для туннелирования с участием фононов формулируются следующим образом: сумма энергии фонона и начальной энергии электрона, туннелирующего из n -области в p -область, равна конечной энергии электрона, протуннелировавшего в p -область; сумма начального импульса электрона и импульса фонона ($\hbar k_p$) равна конечному импульсу протуннелировавшего электрона. В общем случае вероятность непрямого туннелирования гораздо меньше, чем вероятность прямого. Кроме того, вероятность непрямого туннелирования с участием нескольких фононов гораздо меньше вероятности аналогичного процесса с участием лишь одного фонона.

9.2.1. Вероятность туннелирования и туннельный ток

Если электрическое поле в полупроводнике достаточно велико ($\sim 10^6$ В/см), вероятность квантовомеханического туннелирования, т. е. прямого возбуждения электронов из валентной зоны в зону проводимости, становится существенной. Использование приближения ВКБ (метода Вентцеля—Крамерса—Брил-

люэна) приводит к следующему выражению для вероятности туннелирования T_t [5]:

$$T_t \simeq \exp \left[-2 \int_{-x_1}^{x_2} |k(x)| dx \right], \quad (1)$$

где $|k(x)|$ — модуль волнового вектора частицы в области барьера, а $-x_1$ и x_2 — классические точки поворота (рис. 4).

Туннелирование электрона через запрещенную зону формально идентично туннелированию частицы через барьер. Проанализируем ситуацию в случае потенциальных барьеров двух видов: треугольного и параболического (рис. 5, а и б). Внутри треугольного барьера модуль волнового вектора равен

$$k(x) = \sqrt{\frac{2m^*}{\hbar^2} (PE - E)} = \sqrt{\frac{2m^*}{\hbar^2} \left(\frac{E_g}{2} - q\mathcal{E}x \right)}, \quad (2)$$

где PE — потенциальная энергия, E — энергия электрона, E_g — ширина запрещенной зоны полупроводника, \mathcal{E} — напряженность электрического поля. Подставляя выражение (2) в формулу (1), получаем

$$\begin{aligned} T_t &\simeq \exp \left[-2 \int_{-x_1}^{x_2} \sqrt{\frac{2m^*}{\hbar^2} \left(\frac{E_g}{2} - q\mathcal{E}x \right)} dx \right] = \\ &= \exp \left[+ \frac{4}{3} \frac{\sqrt{2m^*}}{\mathcal{E}\hbar} \left(\frac{E_g}{2} - q\mathcal{E}x \right)^{3/2} \right]_{-x_1}^{x_2}. \end{aligned} \quad (3)$$

Поскольку

$$\frac{E_g}{2} - q\mathcal{E}x = 0 \quad \text{при } x = x_2,$$

$$\frac{E_g}{2} - q\mathcal{E}x = E_g \quad \text{при } x = -x_1,$$

выражение (3) принимает вид

$$T_t \simeq \exp \left(- \frac{4\sqrt{2m^*} E_g^{3/2}}{3\mathcal{E}\hbar} \right). \quad (4)$$

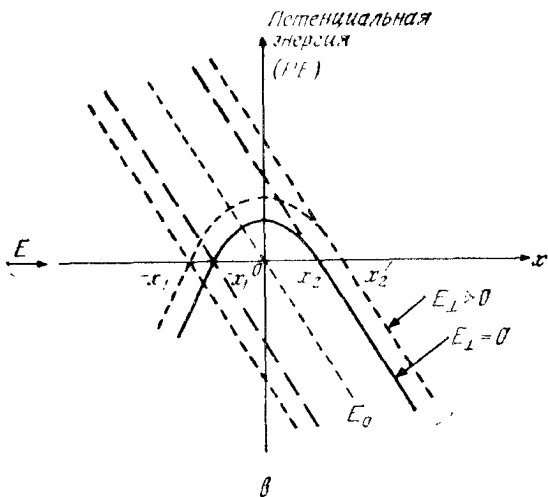
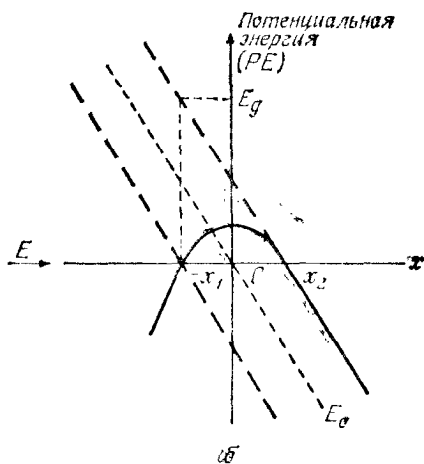
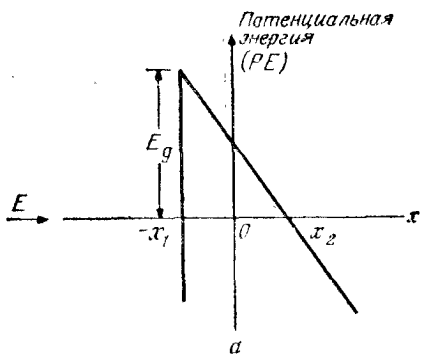
Для параболического барьера

$$PE - E = \frac{(E_g/2)^2 - E_0^2}{E_g} = \frac{(E_g/2)^2 - (q\mathcal{E}x)^2}{E_g}, \quad (5)$$

где E_0 — энергия электрона в электрическом поле. Это выражение представляет собой простейшую алгебраическую функцию,

Рис. 5. Потенциальные барьеры [10].

а — треугольный; б — параболический; в — параболический при $E_L > 0$.



правильно описывающую туннелирование вблизи краев запрещенной зоны [6]. При этом вероятность туннелирования равна

$$\begin{aligned}
 T_t &\approx \exp \left[-2 \int_{-x_1}^{x_2} \sqrt{\frac{2m^*}{\hbar^2} \left(\frac{E_g^2/4 - q^2 \mathcal{E}^2 x^2}{E_g} \right)} dx \right] = \\
 &= \exp \left[-\frac{m^* 1/2 E_g^{3/2}}{2 \sqrt{2} q \hbar \mathcal{E}} \int_{-1}^1 (1 - y^2)^{1/2} dy \right] \Big|_{y=2q \mathcal{E} x / E_g} = \\
 &= \exp \left(-\frac{\pi m^* 1/2 E_g^{3/2}}{2 \sqrt{2} q \hbar \mathcal{E}} \right). \quad (6)
 \end{aligned}$$

Выражение (6) совпадает с выражением (4) с точностью до числового множителя.

Поскольку при туннелировании полный импульс должен сохраняться, при вычислении вероятности туннелирования необходимо учитывать роль поперечной компоненты импульса. В связи с этим полную энергию будем представлять в виде двух компонент E_x и E_{\perp} , где E_{\perp} — энергия, связанная с движением в плоскости, перпендикулярной направлению туннелирования (или с поперечной компонентой импульса), а E_x — энергия движения в направлении туннелирования. Тогда при $E_{\perp} > 0$ (рис. 5, в)

$$pE - E_x = \frac{E_g^2/4 - E_0^2}{E_g} + E_{\perp}, \quad (7)$$

и координаты классических точек поворота равны

$$-x'_1, x'_2 = \mp \frac{1}{q\mathcal{E}} \sqrt{E_g^2/4 + E_g/E_{\perp}}. \quad (8)$$

Из соотношений (1), (7) и (8) получаем вероятность туннелирования

$$T_t \approx \exp \left(-\frac{\pi m^* 1/2 E_g^{3/2}}{2 \sqrt{2} q \hbar \mathcal{E}} \right) \exp \left(-\frac{2E_{\perp}}{E} \right), \quad (9)$$

где первая экспонента та же, что и в выражении (6), и соответствует вероятности туннелирования с нулевым поперечным импульсом. Характерная энергия при туннелировании с отличным от нуля поперечным импульсом \bar{E} равна

$$\bar{E} \equiv \frac{4 \sqrt{2} q \hbar \mathcal{E}}{3 \pi m^* 1/2 E_g^{1/2}}. \quad (10)$$

Если эта величина очень мала, могут туннелировать лишь электроны с малым поперечным импульсом. Другими словами, нали-

чие поперечного импульса у частицы приводит к дополнительному уменьшению вероятности туннелирования в $\text{exp}(-2E_{\perp}/\bar{E})$ раз. При прямом туннелировании поперечный импульс должен сохраняться. Из приведенных выше результатов следует, что вероятность туннелирования тем больше, чем меньше эффективная масса и ширина запрещенной зоны и чем больше электрическое поле.

Теперь найдем величину туннельного тока. Сначала сделаем это в первом приближении, используя плотности состояний в зоне проводимости и валентной зоне [1], а затем обсудим более строгую теорию прямого туннелирования [6].

В условиях теплового равновесия туннельный ток $I_{V \rightarrow C}$ с заполненных состояний валентной зоны на пустые состояния в зоне проводимости равен току $I_{C \rightarrow V}$ с заполненных состояний зоны проводимости на пустые состояния валентной зоны в соответствии с принципом детального равновесия. Выражения для $I_{C \rightarrow V}$ и $I_{V \rightarrow C}$ имеют следующий вид:

$$I_{C \rightarrow V} = A \int_{E_C}^{E_V} F_C(E) n_C(E) T_t [1 - F_V(E)] n_V(E) dE, \quad (11a)$$

$$I_{V \rightarrow C} = A \int_E^{E_V} F_V(E) n_V(E) T_t [1 - F_C(E)] n_C(E) dE, \quad (11b)$$

где A — константа; T_t — вероятность туннелирования, которая считается равной для обоих направлений; $F_C(E)$ и $F_V(E)$ — функции распределения Ферми—Дирака; $n_C(E)$ и $n_V(E)$ — плотности состояний в зоне проводимости и валентной зоне соответственно. При смещении перехода наблюдаемый ток I_t равен

$$I_t = I_{C \rightarrow V} - I_{V \rightarrow C} = A \int_{E_C}^{E_V} [F_C(E) - F_V(E)] T_t n_C(E) n_V(E) dE. \quad (12)$$

В результате последующих вычислений формула (12) приводится к виду [7]

$$I_t = I_P (V/V_P) \exp(1 - V/V_P), \quad (13)$$

где I_P и V_P — ток и напряжение, определенные в соответствии с рис. 2. Напряжение V_P можно оценить из условия максимума произведения числа заполненных состояний в n -области на число пустых состояний с той же энергией в p -области. Результат можно записать в виде [7]

$$V_P \approx (V_n + V_p)/3. \quad (14)$$

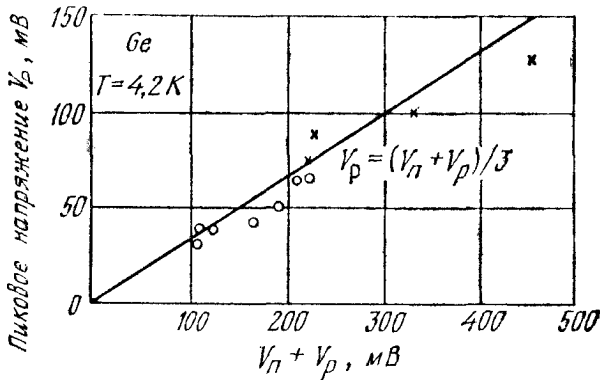


Рис. 6. Зависимость напряжения V_p для германиевых туннельных диодов от суммы V_n и V_p [7, 9].

× при изменении V_p ; ○ при изменении V_n .

Степень вырождения n -области V_n можно оценить из интеграла Ферми—Дирака [8]:

$$V_n \approx \frac{kT}{q} \left[\ln \left(\frac{N_D}{N_g} \right) + 0,35 \left(\frac{N_D}{N_g} \right) \right], \quad (15)$$

где N_D — концентрация доноров, N_C — эффективная плотность состояний в зоне проводимости. Аналогичное выражение можно написать для степени вырождения p -области V_p , если только в соотношении (15) заменить N_D на N_A , а N_C на N_V .

На рис. 6 приведена зависимость напряжения V_p от суммы степеней вырождения V_n и V_p , измеренная на германиевых туннельных диодах [9]. Видно, что сдвиг напряжения V_p к большим значениям при увеличении легирования и экспериментальные значения V_p достаточно хорошо согласуются с формулой (14).

При вычислении тока I_p мы сначала рассмотрим случай прямого туннелирования, т. е. будем считать импульс сохраняющимся. Предельная плотность тока в интервале энергий $dE_x dE_{\perp}$ равна [6, 10]

$$dJ_x = \frac{q(m_y^* m_z^*)^{1/2}}{2\pi^2 \hbar^3} dE_x dE_{\perp}, \quad (16)$$

где

$$E = E_x + E_{\perp}, \quad E_x = \hbar^2 k_x^2 / 2m_x^*. \quad (17)$$

Полагая m^* изотропной и равной для электронов и дырок и используя соотношения (9) и (17), получим выражение для плотности туннельного тока J_t :

$$J_t = \frac{qm^*}{2\pi^2\hbar^3} \exp\left(-\frac{\pi m^*{}^{1/2} E_g^{3/2}}{2\sqrt{2} q\hbar \mathcal{E}}\right) \int [F_C(E) - F_V(E)] \times \\ \times \exp(-2E_\perp/\bar{E}) dE dE_\perp. \quad (18)$$

Для того чтобы в качестве переменных интегрирования фигурировали E и E_\perp , мы использовали выражение (16). Пределы интегрирования по E_\perp определяются условиями $0 \leq E_\perp \leq E_1$ и $0 \leq E_\perp \leq E_2$, где E_1 и E_2 — энергии электрона, отсчитанные от краев зоны проводимости и валентной зоны соответственно. Пределы интегрирования по E определяются краями зон.

После интегрирования по E_\perp получаем

$$J_t = \frac{qm^*}{2\pi^2\hbar^3} \exp\left(-\frac{\pi m^*{}^{1/2} E_g^{3/2}}{2\sqrt{2} q\hbar \mathcal{E}}\right) \left(\frac{\bar{E}}{2}\right) D, \quad (19)$$

$$D \equiv \int [F_C(E) - F_V(E)] [1 - \exp(-2E_S/\bar{E})] dE, \quad (20)$$

$$\mathcal{E} \equiv (qV_{bi} N^*/2\epsilon_s)^{1/2}, \quad (21)$$

где в уравнение (20) введена величина E_S , обозначающая меньшую из величин E_1 и E_2 , а величина \bar{E} определяется выражением (10). Величина \mathcal{E} — средняя напряженность электрического поля резкого перехода; qV_{bi} — разность потенциальных энергий на краях перехода, примерно равная ширине запрещенной зоны; N^* — эффективная концентрация легирования:

$$N^* \equiv N_A N_D / (N_A + N_D). \quad (22)$$

Для германового туннельного диода в качестве эффективной массы в соотношении (19) следует [11] подставлять величину

$$m^* = 2 \left(\frac{1}{m_e^*} + \frac{1}{m_{lh}^*} \right)^{-1},$$

если речь идет о туннелировании из зоны легких дырок в точки $\langle 000 \rangle$ зоны проводимости Ge. Здесь $m_{lh}^* = 0,04m_0$ — масса легких дырок, а $m_e^* = 0,036m_0$ — масса электрона в зоне проводимости в точке $\langle 000 \rangle$. При туннелировании в направлении $\langle 100 \rangle$ в минимумы $\langle 111 \rangle$ эффективная масса равна

$$m^* = 2 \left[\left(\frac{1}{3m_l^*} + \frac{2}{3m_i^*} \right) + \frac{1}{m_{lh}^*} \right]^{-1},$$

где $m_l^* = 1,6m_0$ и $m_i^* = 0,082m_0$ — соответственно продольная и поперечная массы минимумов $\langle 111 \rangle$. Однако показатели экспоненты в уравнении (19) отличаются для этих двух случаев лишь

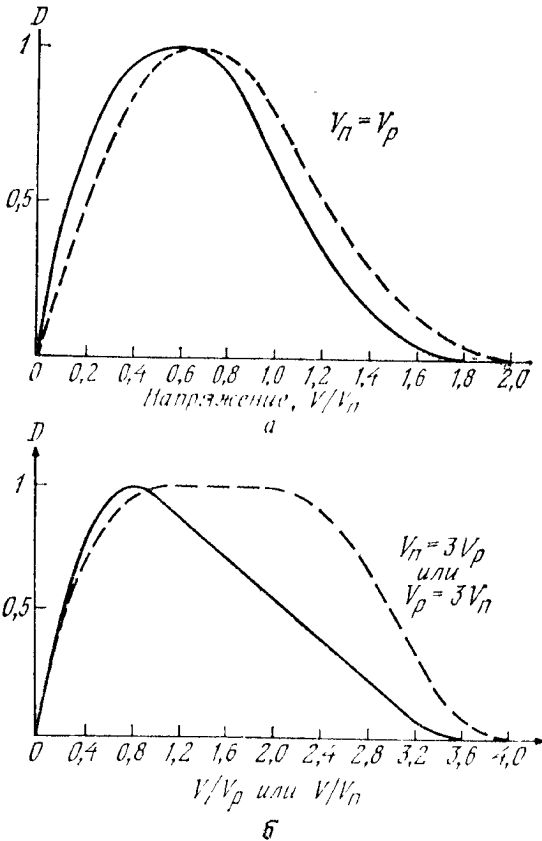


Рис. 7. Зависимость эффективной плотности состояний D от прямого смещения перехода для случаев прямого (штриховые кривые) и непрямого (сплошные кривые) туннелирования (величина \bar{E} полагается очень большой) [6].

на 5 %. Величина D , определяемая уравнением (20), представляет собой интеграл перекрытия, от которого зависит форма вольт-амперной характеристики. Она имеет размерность энергии (эВ) и зависит от температуры и степеней вырождения qV_n и qV_p . При $T = 0$ К F_C и F_V имеют вид ступенек. На рис. 7 приведены зависимости величины D от прямого смещения (штриховые кривые) для случаев $V_n = V_p$ и $V_n = 3V_p$. Видно, что максимум D достигается при $V = (V_n + V_p)/3$, т. е. в полном соответствии с формулой (14).

Аналогичные зависимости приведены на рис. 7 и для случая непрямого туннелирования (сплошные кривые). Для непрямого туннелирования с участием фононов вероятность туннелирования равна [6, 12]

$$T_t \approx \exp \left[\frac{-4\sqrt{2} m_{rx}^* 1/2 (E_g - E_p)^{3/2}}{3q\hbar\mathcal{E}} \right], \quad (23)$$

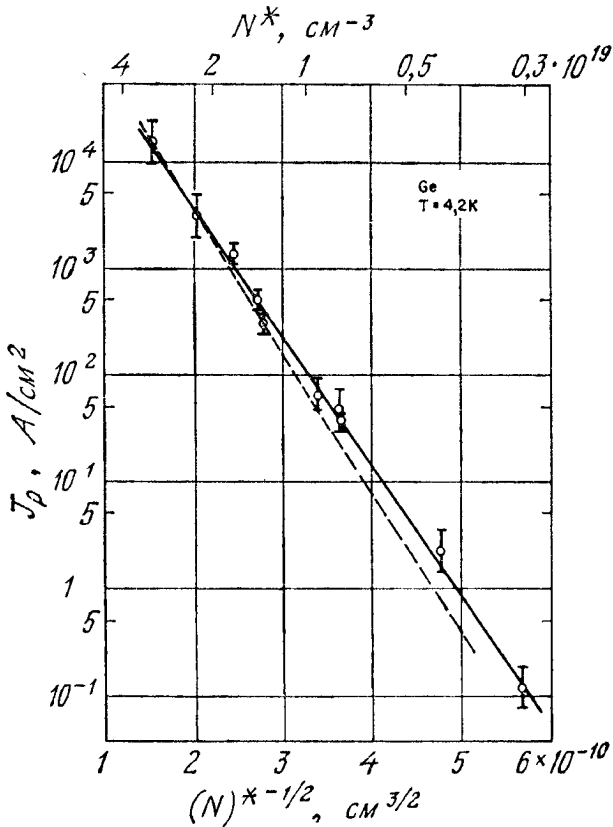


Рис. 8. Зависимость плотности пикового тока от эффективной концентрации легирования для германиевых диодов [7, 9].

где m_{rx}^* — приведенная эффективная масса в направлении туннелирования, а E_p — энергия фонона. Выражение для туннельного тока в этом случае аналогично соотношению (19). Максимум тока на прямой ветви вольт-амперной характеристики достигается при $V = V_n + V_p - (V_n^2 + V_p^2)^{1/2}$. На рис. 8 приведены экспериментальная (сплошная линия) и теоретическая, вычисленная по формуле (19), (штриховая линия) зависимости пикового тока для нескольких германиевых диодов [7].

9.2.2. Вольт-амперные характеристики

Как следует из рис. 2, б, вид статической вольт-амперной характеристики определяется тремя компонентами тока: туннельным, избыточным и диффузионным токами. В идеальном туннель-

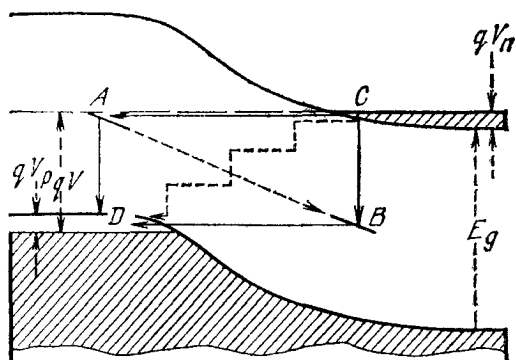


Рис. 9. Зонная диаграмма, иллюстрирующая различные механизмы туннелирования через состояния в запрещенной зоне [13].

ном диоде туннельный ток должен уменьшаться до нуля при напряжениях $V \geq V_n + V_p$; при больших напряжениях должен течь только обычный диодный ток, вызванный инжекцией неосновных носителей. Однако на практике ток туннельного диода при таких напряжениях существенно превышает обычный диодный ток, в связи с чем он получил название «избыточный ток». Основная причина избыточного тока — туннелирование носителей через энергетические состояния, расположенные в запрещенной зоне.

Для определения величины избыточного тока воспользуемся рис. 9 [13], на котором приведены несколько возможных путей туннелирования. На первом пути электрон, выходящий из точки C в зоне проводимости, может протуннелировать на подходящий локализованный уровень в точку A , из которой, отдав энергию, может попасть затем в точку D , расположенную в валентной зоне. По второму пути электрон «падает» из точки C на пустой уровень в точку B , из которой затем туннелирует в точку D . По третьему пути (маршрут $CABD$) электрон отдает свою избыточную энергию в процессе, который можно назвать проводимостью по примесной зоне между точками A и B . И наконец, четвертый путь — лестница из точки C в точку D , которая состоит из целого ряда туннельных переходов между различными локальными уровнями; сопровождаемых вертикальными скачками — потерями энергии при переходах электрона с одного уровня на другой; подобный процесс возможен при достаточно большой концентрации промежуточных уровней. Основным можно считать маршрут CBD ; остальные пути фактически являются лишь более сложными его модификациями. Рассмотрим электрон, туннелирующий из точки B в точку D при напряжении смещения на переходе V . Высота барьера, под которым он должен туннелировать, равна

$$E_x \approx E_g - qV + q(V_n + V_p) \approx q(V_{bi} - V), \quad (24)$$

где V_{bi} — контактная разность потенциалов (здесь полагается, что электрон появляется в валентной зоне вблизи ее потолка). Вероятность туннелирования T_i для электрона, движущегося на уровне точки B , равна

$$T_i \approx \exp\left(\frac{-4\sqrt{2}m_x^*{}^{1/2}E_x^{3/2}}{3q\hbar\mathcal{E}}\right) = \exp(-\alpha_x E_x^{3/2}/\mathcal{E}). \quad (25)$$

Это выражение аналогично формуле (4), в которой величина E_g заменена величиной E_x . Максимальная напряженность электрического поля в резком переходе равна

$$\mathcal{E} = 2(V_{bi} - V)/W, \quad (26)$$

где W — ширина обедненного слоя, описываемая выражением

$$W = \left[\frac{2e_s}{q} \left(\frac{N_A + N_D}{N_A N_D} \right) (V_{bi} - V) \right]^{1/2}. \quad (27)$$

Если плотность заполненных состояний, расположенных выше потолка валентной зоны на величину E_x , равна D_x , то плотность избыточного тока определяется выражением

$$J_x \approx A_1 D_x T_i, \quad (28)$$

где A_1 — константа, а зависимость избыточного тока от различных параметров определяется в основном их влиянием на вероятность туннелирования T_i , а не на плотность состояний D_x . Тогда, подставляя выражения (24)—(27) в формулу (28), получим следующее выражение для избыточного тока [13]:

$$J_x \approx A_1 D_x \exp\{-\alpha'_x [E_g - qV + 0,6q(V_n + V_p)]\}, \quad (29)$$

где α'_x — константа. Из выражения (29) следует, что избыточный ток возрастает при увеличении плотности локализованных состояний в запрещенной зоне (благодаря множителю D_x), а также экспоненциально зависит от приложенного напряжения V (при условии, что $E_g \gg qV$). Выражение (29) можно также записать в виде [14]

$$\begin{aligned} J_x &= J_V \exp\left[\frac{4}{3} \left(\frac{m_x^* e_s}{N^*}\right)^{1/2} (V - V_V)\right] = \\ &= J_V \exp[A_2 (V - V_V)], \end{aligned} \quad (30)$$

где J_V — плотность тока в минимуме (в долине) при напряжении V_V , а A_2 — множитель в экспоненте. Экспериментальные зависимости $\ln J_x$ от V для туннельных диодов из кремния и арсенида галлия в соответствии с формулой (30) оказываются линейными.

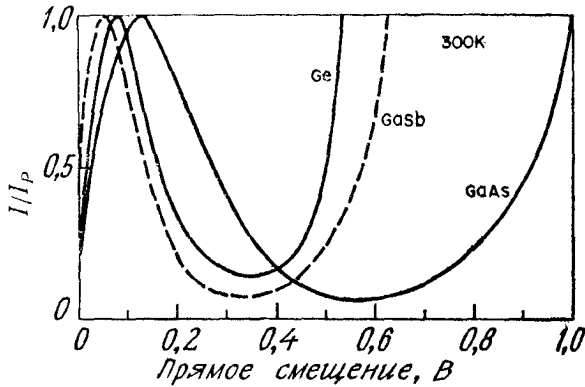


Рис. 10. Типичные вольт-амперные характеристики туннельных диодов из Ge, GaSb и GaAs при 300 К.

Диффузионный ток представляет собой ток инжекции неосновных носителей в p - n -переходах

$$J_{th} = J_0 (e^{qV/kT} - 1), \quad (31)$$

где J_0 — плотность тока насыщения (гл. 2).

Полная статическая вольт-амперная характеристика определяется суммой этих трех компонент тока:

$$\begin{aligned} J &= J_t + J_x + J_{th} = \\ &= J_p \left(\frac{V}{V_p} \right) \exp \left(1 - \frac{V}{V_p} \right) + J_v \exp [A_2 (V - V_v)] + J_0 \exp \left(\frac{qV}{kT} \right). \end{aligned} \quad (32)$$

Вклад туннельного тока в полный ток является определяющим при $V < V_v$, избыточный ток наиболее существен при $V \approx V_v$, а диффузионный ток преобладает при $V > V_v$.

На рис. 10 для сравнения одновременно представлены типичные вольт-амперные характеристики туннельных диодов из Ge, GaSb [15] и GaAs [16, 17] при температуре 300 К. Величина отношения I_p/I_v равна 8 для Ge и 12 для GaSb и GaAs. Туннельные диоды были изготовлены из многих полупроводников. В диодах из Si отношение I_p/I_v равнялось примерно 4 [18], из InAs оно составляло 2 при комнатной температуре и возросло до 10 при 4 К [19] (вследствие малой ширины запрещенной зоны), из InP отношение I_p/I_v составляло 5 [20], а в диодах на тройном соединении $Ga_{0,7}Al_{0,3}As$ достигало 12 [21]. В общем случае эта величина для данного полупроводника возрастает с увеличением уровня легирования как n -, так и p -областей. Предельное значение этого соотношения зависит, во-первых, от пикового тока,

который определяется эффективной массой туннелирования и шириной запрещенной зоны, и, во-вторых, от минимального тока, который определяется распределением и концентрацией локализованных состояний в запрещенной зоне.

Рассмотрим кратко влияние на вольт-амперную характеристику температуры, облучения электронами и давления. Температурная зависимость пикового значения тока объясняется зависимостью от температуры величин D и E_g , входящих в формулу (19). При высоких концентрациях температурная зависимость слабая и изменения вероятности туннелирования определяются в первую очередь отрицательной величиной dE_g/dT , поэтому пиковый ток растет с увеличением температуры. В слаболегированных туннельных диодах важнейшим фактором является уменьшение D с ростом температуры. Естественно, что при этом пиковый ток уменьшается с ростом температуры. Для типичных германиевых туннельных диодов изменения пикового тока в диапазоне температур -50 — $+100$ °C составляют примерно ± 10 % [22]. Ток в минимуме в общем случае растет с увеличением температуры, так как ширина запрещенной зоны уменьшается с ростом температуры (см. выражение (29)).

Основное влияние электронного облучения состоит в возрастании избыточного тока, что связано с увеличением концентрации энергетических уровней в запрещенной зоне [23]. Возросший избыточный ток можно постепенно спизить с помощью отжига. Подобные результаты получают и для других видов облучения, например γ -лучей. Влияние давления на вольт-амперную характеристику сводится к увеличению избыточного тока германиевых и кремниевых туннельных диодов [24], что обусловлено глубокими уровнями, связанными с возникающими под давлением дефектами решетки в обедненном слое. Исследования показали обратимость этих изменений. Однако для GaSb I_P и I_V уменьшаются с увеличением гидростатического давления [25], что может быть объяснено расширением запрещенной зоны и уменьшением степеней вырождения V_n и V_p при возрастании давления.

9.2.3. Характеристики прибора

Большинство туннельных диодов изготавливаются одним из следующих способов. 1) *Вплавление таблетки*. Маленькая таблетка металлического сплава, в котором в высокой концентрации растворена легирующая примесь одного типа, вплавляется в поверхность полупроводниковой подложки с сильнолегированной примесью противоположного типа при точном соблюдении режима вплавления в атмосфере инертного газа или водорода (например, мышьяк из легированной мышьяком тонкой оловянной

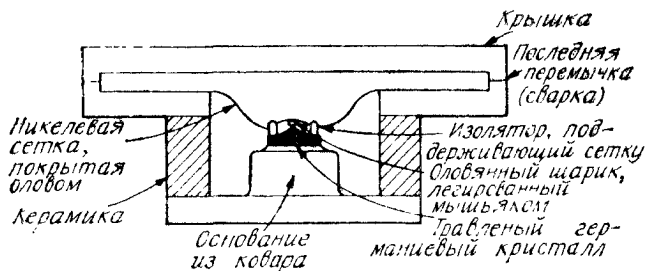


Рис. 11. Корпус, используемый в СВЧ-технике, вместе с помещенным в него германиевым туннельным диодом [27].

таблетки создает области n^+ -типа на поверхности германиевой подложки (рис. 11, p^+ -типа). Необходимая величина пикового тока достигается после введения процесса травления. 2) *Импульсная сварка*. Контакт и переход формируются одновременно при приложении импульсного напряжения к полупроводниковой подложке и помещенному на нее сплаву, содержащему компенсирующую примесь. 3) *Планарная технология* [26]. Планарные туннельные диоды создаются с использованием таких процессов, как эпитаксия, диффузия и контролируемое плавление. На рис. 11 показан корпус, используемый в СВЧ-технике, с помещенным в него германиевым туннельным диодом [27]. Верхним контактом к оловянной таблетке служит луженая проволочная сетка. Диаметр перехода очень мал: он составляет $\sim 2,5$ мкм для прибора, работающего на частоте 6 ГГц, и меньше для более высокочастотных диодов.

На рис. 12, *а* показано символическое обозначение туннельного диода [28], а на рис. 12, *б* — его основная эквивалентная схема, в которую входят четыре элемента: последовательно соединенные индуктивность L_S и сопротивление R_S и параллельно соединенные емкость диода C и отрицательное сопротивление $-R$.

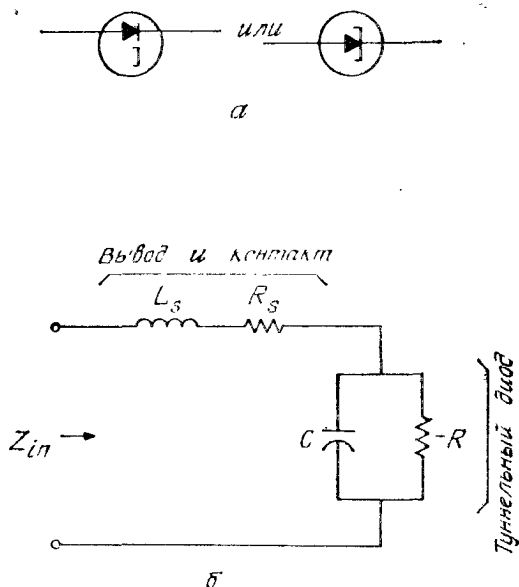
Последовательное сопротивление R состоит из сопротивлений вывода, омических контактов и сопротивления растекания подложки, равного $\rho/2d$, где ρ — удельное сопротивление полупроводника, а d — диаметр площади диода. Последовательная индуктивность L_S для коаксиальной линии равна [29]

$$L_S = \frac{2,303\mu_0 l}{2\pi} \ln\left(\frac{r_2}{r_1}\right), \quad (33)$$

где μ_0 — магнитная проницаемость среды, l — длина, а r_1 и r_2 — соответственно внутренний и внешний радиусы коаксиальной линии. Ниже показано, что эти паразитные элементы существенно ограничивают характеристики туннельного диода.

Для определения величины емкости и отрицательного сопротивления диода воспользуемся его типичной вольт-амперной ха-

Рис. 12. Символическое обозначение туннельного диода (а) и его эквивалентная схема (б) [28].



рактической (рис. 13, а). Получаемая отсюда зависимость дифференциального сопротивления, определяемого как $(dI/dV)^{-1}$, приведена на рис. 13, б. Величина отрицательного сопротивления в точке перегиба вольт-амперной характеристики, где отрицательное сопротивление достигает минимального значения, обозначена через $R_{\text{мин}}$. Эту величину можно аппроксимировать выражением

$$R_{\text{мин}} \approx 2V_p/I_p, \quad (34)$$

где I_p и V_p — пиковый ток и соответствующее напряжение. На рис. 13, в приведена зависимость проводимости dI/dV от напряжения. При напряжениях, соответствующих пиковому току и току долины, проводимость оказывается равной нулю; емкость диода обычно измеряется в минимуме вольт-амперной характеристики и обозначается C_j .

Импеданс эквивалентной цепи $Z_{\text{ин}}$, изображенной на рис. 12, равен

$$Z_{\text{ин}} = \left[R_s + \frac{-R}{1 + (\omega RC)^2} \right] + j \left[\omega L_s + \frac{-\omega CR^2}{1 + (\omega RC)^2} \right]. \quad (35)$$

Из этого соотношения видно, что активная (действительная) составляющая импеданса при некоторой частоте может оказаться равной нулю, в то время как для реактивной (мнимой) составляющей импеданса подобное соотношение может выполняться при другом значении частоты. Мы будем называть эти частоты актив-

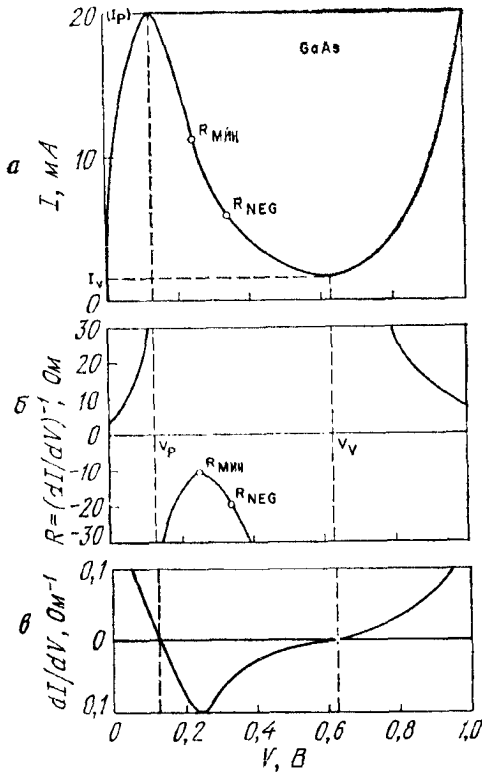


Рис. 13 Вольт-амперные характеристики туннельного диода из GaAs при 300 К (а), зависимость дифференциального сопротивления $(dI/dV)^{-1}$ от напряжения (б) ($R_{\text{мин}}$ — минимальное отрицательное сопротивление, R_{NEG} — сопротивление в области минимального шум-фактора) и зависимость дифференциальной проводимости $G \equiv dI/dV$ от напряжения (в) (при пиковом и долинном токе $G = 0$).

ной частотой отсечки f_r и реактивной частотой отсечки f_x соответственно. Они равны

$$f_r \equiv \frac{1}{2\pi RC} \sqrt{\frac{R}{R_s} - 1}, \quad (36)$$

$$f_x \equiv \frac{1}{2\pi} \sqrt{\frac{1}{L_s C} - \frac{1}{(RC)^2}}. \quad (37)$$

Для частот отсечки, определенных при минимальном сопротивлении и емкости, соответствующей точке минимума вольт-амперной характеристики, имеем

$$f_{r0} \equiv \frac{1}{2R_{\text{мин}}C_j} \sqrt{\frac{R_{\text{мин}}}{R_s} - 1} \geq f_r, \quad (38)$$

$$f_{x0} \equiv \frac{1}{2\pi} \sqrt{\frac{1}{L_s C_j} - \frac{1}{(R_{\text{мин}}C_j)^2}} \leq f_x, \quad (39)$$

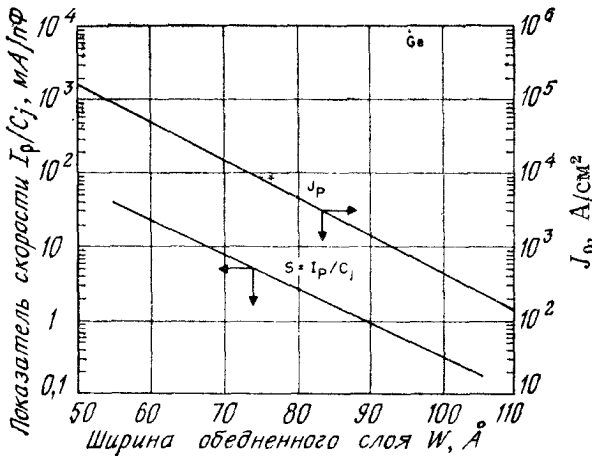


Рис. 14. Зависимости средней величины плотности пикового тока и показателя скорости переключения I_P/C_j от ширины обедненного слоя для германиевых туннельных диодов при 300 К [26].

где f_{r0} — максимальная активная частота отсечки, выше которой сопротивление диода всегда положительно; $f_{\lambda 0}$ — минимальная реактивная частота отсечки либо резонансная частота диода, при которой реактивность диода равна нулю (диод служит генератором колебаний, если $f_{r0} > f_{\lambda 0}$). Для большинства приложений, в которых используется работа диода в области отрицательного сопротивления, требуется выполнение соотношений $f_{\lambda 0} > f_{r0}$ и $f_{r0} \gg f_0$, где f_0 — рабочая частота. Из формул (38) и (39) следует, что для выполнения требования $f_{\lambda 0} > f_{r0}$ следует уменьшать величину последовательной индуктивности L_S . Характеристикой качества туннельных диодов служит показатель скорости переключения, который определяется как отношение пикового тока к величине емкости I_P/C_j при напряжении V_P . Скорость переключения туннельного диода определяется током, необходимым для зарядки емкости перехода, и поэтому зависит от величины тока источника питания и средней величины произведения RC . Поскольку отрицательное сопротивление R обратно пропорционально пиковому току, для быстрого переключения требуется большой показатель скорости переключения (или малое произведение RC). На рис. 14 приведены зависимости пикового тока и показателя скорости переключения [26] от ширины обедненного слоя в германиевых туннельных диодах при 300 К. Мы видим, что для получения большого показателя скорости переключения необходима малая ширина слоя, т. е. более сильное легирование.

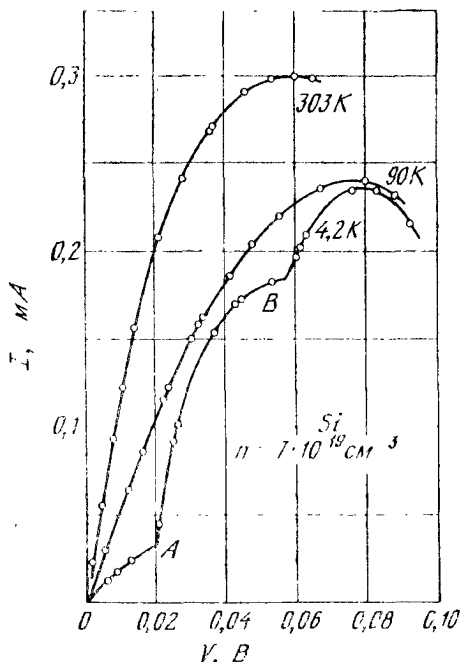


Рис. 15. Вольт-амперные характеристики кремниевого туннельного диода при трех температурах [30].

С эквивалентной цепью связана другая важная величина — шум-фактор, определяемый выражением

$$NF \approx 1 + \frac{q}{2kT} |RI|_{\min}, \quad (40)$$

где $|RI|_{\min}$ — минимальное для данной вольт-амперной характеристики значение произведения тока на отрицательное сопротивление. Соответствующая величина R (обозначенная как R_{NFG}) приведена на рис. 13. Произведение $q|RI|_{\min}/2kT$ называется шумовой постоянной K ; эта величина является константой для данного материала. Типичные значения K при комнатной температуре равны 1,2 для Ge, 2,4 для GaAs, 0,9 для GaSb. Шум-фактор для германиевых туннельных диодов равен ~ 5 дБ при 6 ГГц и ~ 6 дБ при 14 ГГц.

Помимо использования в СВЧ- и вычислительной технике туннельный диод полезен для изучения фундаментальных физических параметров. Такой диод можно использовать для туннельной спектроскопии — метода определения энергетического спектра полупроводника, в котором зондирование спектра осуществляется при помощи туннелирования электронов с известным распределением по энергиям вместо применяемого в оптической спектроскопии поглощения фотонов фиксированной частоты. Туннельная спек-

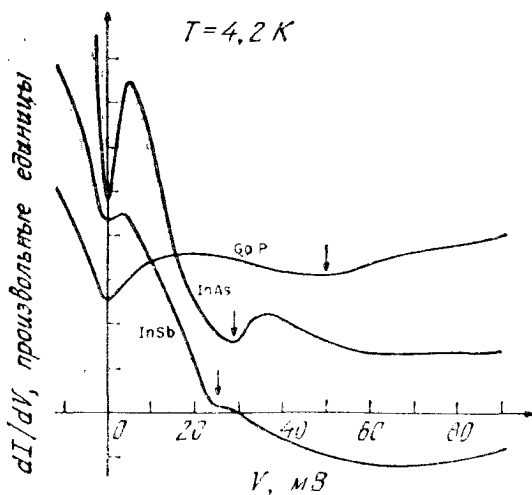


Рис. 16. Зависимость проводимости dI/dV от напряжения для туннельных диодов, созданных на полупроводниках группы $A^{III}B^V$ [31].

троскопия использовалась для изучения электронных энергетических состояний в твердых телах и для наблюдения мод возбуждения в таких приборах с $p-n$ -переходами. На рис. 15 приведены вольт-амперные характеристики кремниевое туннельного диода [30]. При уменьшении температуры до 4,2 К на кривой появляются две точки перегиба A и B . Эти точки перегиба соответствуют началу туннелирования с испусканием фононов, а энергии (или напряжения) в точках A и B равны энергиям акустических и оптических фононов соответственно. Подобные наблюдения проведены на туннельных переходах в сложных полупроводниках типа $A^{III}B^V$. На рис. 16 приведены зависимости проводимости dI/dV от напряжения V для GaP, InAs и InSb при 4,2 К [31]. Стрелками указаны энергии, соответствующие энергиям оптических фононов в этих полупроводниках. Более детальное изучение непрямого туннелирования может быть проведено при анализе второй производной d^2I/dV^2 [32].

9.3. ОБРАЩЕННЫЙ ДИОД

Если концентрации легирующих примесей таковы, что p - и n -области $p-n$ -перехода близки к вырождению либо слабо вырождены, то при малых смещениях ток в «обратном» направлении (рис. 17) превышает ток в «прямом» направлении. Отсюда происходит название такого туннельного диода — «обращенный диод». В состоянии теплового равновесия уровень Ферми в обра-

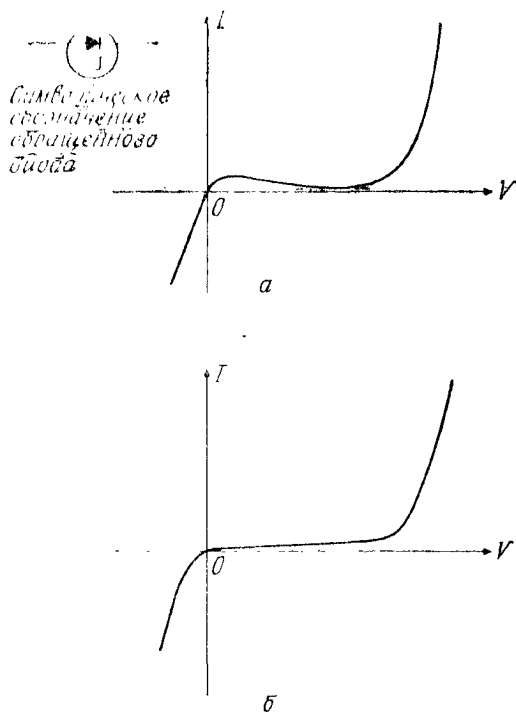


Рис. 17. Символическое обозначение обращенного диода и его вольт-амперные характеристики при наличии отрицательного сопротивления (а) и без отрицательного сопротивления (б).

щенном диоде очень близок к краям зон. При малых обратных смещениях (при отрицательном потенциале p -области по отношению к потенциалу n -области) диаграмма энергетических зон обращенного диода подобна диаграмме, приведенной на рис. 3, а, с одним отличием — обе стороны перехода теперь не вырождены. При обратном смещении электроны могут легко туннелировать из валентной зоны в зону проводимости, что по формуле (19), которую можно записать в виде

$$J \approx B_1 \exp(+|V|/B_2), \quad (41)$$

будет приводить к возрастанию туннельного тока с ростом приложенного напряжения V . Здесь B_1 и B_2 — положительные величины, слабо зависящие от V . Из формулы (41) следует, что обратный ток зависит от напряжения примерно по экспоненциальному закону.

Обращенный диод можно использовать как выпрямитель малых сигналов, детектор СВЧ-излучения и смеситель. Подобно туннельному диоду, обращенный диод имеет хорошие частотные характеристики, поскольку его работа не связана с накоплением неосновных носителей. Кроме того, вольт-амперная характери-

стика обращенного диода нечувствительна к влиянию температуры и облучения, и в таких диодах очень мал $1/f$ -шум [33—36].

При работе в нелинейном режиме, например в качестве высокоскоростного переключателя, характеристикой совершенства прибора служит параметр γ , определяемый как отношение второй и первой производных вольт-амперной характеристики. Этот параметр также называют коэффициентом кривизны [37]:

$$\gamma \equiv \frac{d^2 I / dV^2}{dI / dV}. \quad (42)$$

Величина γ служит мерой степени нелинейности, нормированной на величину рабочей полной проводимости. При прямом смещении p — n -перехода или барьера Шоттки (гл. 5) величина γ равна q/nkT , т. е. обратно пропорциональна T . При комнатной температуре для идеального p — n -перехода ($n = 1$) $\gamma \simeq 40 \text{ В}^{-1}$ независимо от смещения. Однако при обратном смещении p — n -перехода величина γ очень мала при малых смещениях и растет пропорционально фактору лавинного умножения вблизи напряжения пробоя [38]. Хотя идеальная обратная вольт-амперная характеристика дает вблизи напряжения пробоя величину γ , большую чем 40 В^{-1} , однако из-за статистического распределения примесей и ограничения тока объемным зарядом реальные значения γ могут оказаться гораздо меньше.

Величину γ для обращенного диода можно получить на основании формул (4), (13), (14), (26). Она равна [39]

$$\gamma (\text{при } V = 0) = \frac{4}{V_n + V_p} + \frac{2}{\hbar} \sqrt{\frac{2\varepsilon_s m^*}{N^*}}, \quad (43)$$

где m^* — средняя эффективная масса носителей:

$$m^* \simeq m_e^* m_h^* / (m_e^* + m_h^*),$$

а N^* — эффективная концентрация легирования, определяемая формулой (22). Ясно, что коэффициент кривизны γ зависит от концентраций примесей в обеих областях перехода и от эффективных масс. В отличие от барьеров Шоттки в обращенном диоде величина γ относительно слабо зависит от изменений температуры, поскольку параметры, фигурирующие в выражении (43), являются медленно меняющимися функциями температуры.

На рис. 18, а и б приведены теоретические и экспериментальные значения γ для германиевых обращенных диодов. Сплошными кривыми представлены результаты вычислений по формуле (43) с использованием следующих значений эффективных масс: $m_e^* = 0,22m_0$ и $m_h^* = 0,39m_0$. Совпадение результатов во всем рассмотренном диапазоне концентраций легирующих примесей в об-

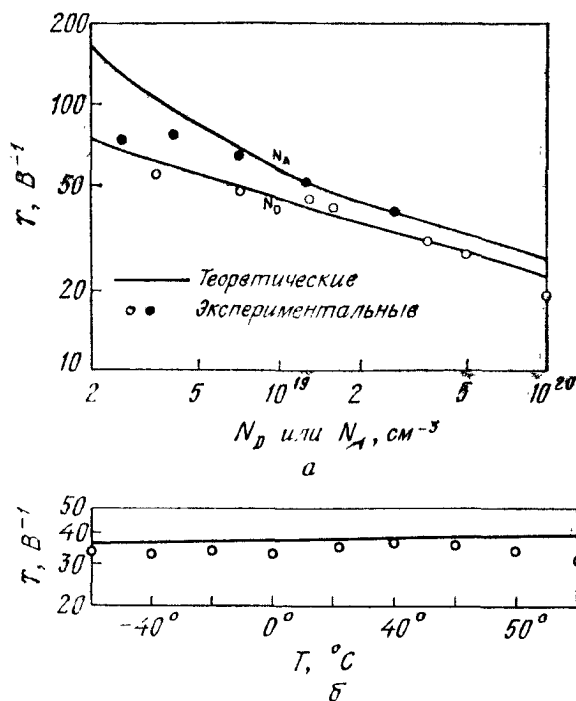


Рис. 18. Зависимость γ при 300 К и $V \approx 0$ в германиевых диодах от концентрации акцепторов (при фиксированном значении $N_D = 2 \cdot 10^{19} \text{ см}^{-3}$) и доноров (а) (при фиксированном значении $N_A = 10^{19} \text{ см}^{-3}$ [39]) и от температуры (б) [39].

щем довольно хорошее. Отметим также две интересные особенности поведения γ для обращенных диодов: 1) γ может быть больше 40 V^{-1} , 2) нечувствительность γ к изменениям температуры.

9.4. туннельный МДП-диод

Для диода со структурой металл—диэлектрик—полупроводник (МДП) вольт-амперные характеристики критическим образом зависят от толщины диэлектрика. Если диэлектрический слой достаточно толстый (например, для системы $\text{Si}-\text{SiO}_2$ больше 50 \AA), то переходом носителей сквозь этот слой можно пренебречь, и МДП-диод представляет собой обычную МДП-емкость (гл. 7). Если же, наоборот, диэлектрический слой очень тонкий (меньше 10 \AA), то он практически не мешает движению носителей между металлом и полупроводником, и тогда структура представляет собой барьер Шоттки (гл. 5). Если же диэлектрический

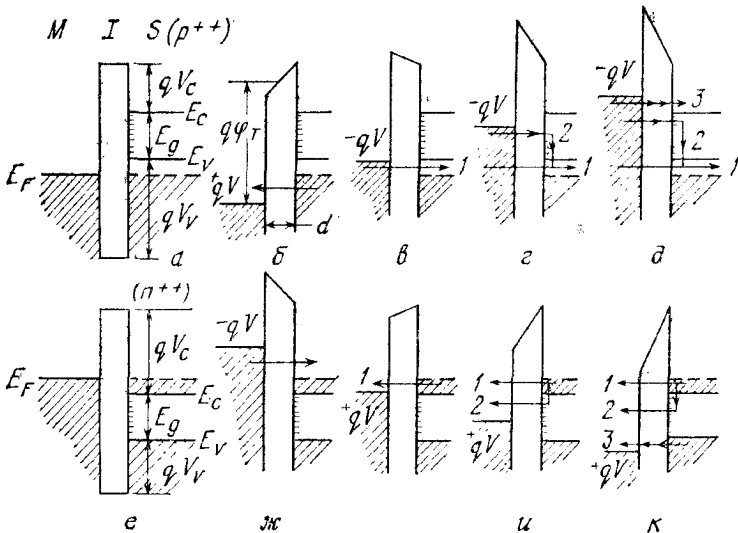


Рис. 19. Упрощенные энергетические диаграммы (учитывающие влияние поверхностных состояний) туннельных МДП-диодов на вырожденных подложках [40].

слой имеет промежуточную толщину ($10 \text{ \AA} < d < 50 \text{ \AA}$), то МДП-диод представляет собой прибор третьего класса — туннельный МДП-диод. В этом разделе рассмотрены сначала свойства туннельных МДП-диодов на вырожденных полупроводниковых подложках. Такие диоды могут обладать отрицательным сопротивлением такой же природы, что и простой туннельный диод. Затем рассмотрены свойства туннельных МДП-диодов на невырожденных подложках.

9.4.1. Вырожденный полупроводник

На рис. 19 приведены упрощенные энергетические диаграммы, учитывающие роль поверхностных состояний, для туннельных МДП-диодов на p^{++} - и n^{++} -полупроводниковых подложках [40]. Для простоты здесь не учитываются изгиб зон под действием силы изображения и падение потенциала на слое оксида в условиях равновесия. Сначала рассмотрим случай подложки p^{++} -типа. Приложение положительного напряжения к металлу (рис. 19, б) вызывает туннелирование электронов из валентной зоны в металл. Величина туннельного тока определяется выражением, аналогичным формуле (12). Используя приближение ВКБ и законы сохранения энергии и поперечного импульса k_{\perp} , выражение для плотности туннельного тока, протекающего в направлениях x

между двумя проводящими областями через запрещенную область, можно записать в виде [41]

$$J = \frac{q}{4\pi^2\hbar} \iint T_t [F_1(E) - F_2(E)] dk_{\perp}^2 dE, \quad (44)$$

где F_1 и F_2 — распределения Ферми в обеих проводящих областях, а T_t — вероятность туннелирования. В рассматриваемом нами здесь случае МДП-диода следует иметь в виду, что поверхности Ферми для электронов в полупроводнике значительно меньше, чем в металле. Вследствие этого можно полагать, что туннелирование электронов из полупроводника в металл всегда разрешено. Если же кроме этого предположить, что энергетические зоны имеют параболическую форму с изотропной электронной массой m^* , то формулу (44) можно записать в более простом виде

$$J = \frac{m^*q}{2\pi^2\hbar^3} \iint T_t dE_{\perp} dE, \quad (45)$$

где E_{\perp} и E — поперечная и полная кинетические энергии электронов в полупроводнике. Пределами интегрирования по E_{\perp} являются нуль и E , а пределами интегрирования по E служат два уровня Ферми. Вероятность туннелирования через прямоугольный барьер с эффективной высотой $q\varphi_T$ и шириной d (рис. 19, б) может быть получена из формулы (1) [42]:

$$T_t \simeq \exp[-2(2m^*/\hbar^2)^{1/2}(q\varphi_T)^{1/2}d] \simeq \quad (46)$$

$$\simeq \exp(-\alpha_T \varphi_T^{1/2} d), \quad (46a)$$

где постоянная $\alpha_T \simeq 1$, если эффективная масса в диэлектрике равна массе свободного электрона и если значения φ_T и d измеряются соответственно в вольтах и ангстремах.

Этот туннельный ток (рис. 19, б) монотонно возрастает с увеличением разности энергий между положениями уровней Ферми; кроме того, он увеличивается с уменьшением высоты барьера. Приложение к металлу малого отрицательного напряжения (рис. 19, в) приводит к туннелированию электронов из металла на свободные состояния в валентной зоне полупроводника. Как видно из рис. 19, г, увеличение напряжения $-V$ приводит к увеличению эффективной высоты барьера, через который электроны туннелируют из металла на пустые состояния в валентной зоне, а следовательно, к появлению падающего участка на вольт-амперной характеристике (при условии, что $qV_G < qV$ (рис. 19, а)). Однако вклад в ток может вносить и компонента другой природы, обусловленная туннелированием электронов с большими энергиями из металла на пустые поверхностные состояния, где эти электроны будут рекомбинировать с дырками из валентной зоны. Поскольку высота барьера для такого туннелирования умень-

шается с ростом смещения, вольт-амперная характеристика этой компоненты тока будет монотонно возрастающей. Наконец, при дальнейшем увеличении смещения появляется третья, очень быстро нарастающая компонента туннельного тока, обусловленная переходом носителей из металла в зону проводимости полупроводника (рис. 19, д).

Теперь рассмотрим туннелирование в структуре с полупроводником n^{++} -типа. Эффективная высота диэлектрического барьера в случае полупроводника n^{++} -типа (рис. 19, е) должна быть меньше, чем в случае полупроводника p^{++} -типа, а следовательно, должна быть больше и величина туннельного тока. При отрицательном напряжении на металле электроны туннелируют из металла на пустые состояния в зоне проводимости полупроводника, что приводит к большому, быстро нарастающему току (рис. 19, ж). Малое положительное напряжение на металле приводит к усилению туннелирования электронов из зоны проводимости полупроводника в металл (рис. 19, з). Если поверхностные состояния заполнены электронами, захваченными из зоны проводимости, дальнейшее увеличение напряжения приведет к росту второй компоненты тока, обусловленной туннелированием электронов с поверхностных состояний в металл. Эта компонента тока растет с увеличением напряжения вследствие уменьшения эффективного диэлектрического барьера (рис. 19, и). При больших напряжениях, кроме того, оказывается возможным туннелирование из валентной зоны в металл. Однако вклад этой компоненты тока в полную вольт-амперную характеристику невелик из-за относительно большой высоты барьера (рис. 19, к). Таким образом, влияние зонной структуры полупроводника на туннельные характеристики МДП-структур с подложками n^{++} -типа оказывается гораздо более слабым, чем в случае структур p^{++} -типа.

На рис. 20 приведены вольт-амперные характеристики трех образцов, изготовленных на кремниевых p^{++} -подложках со слоями окисла толщиной 20 Å (0,0006 Ом·см) и подвергнутых различным обработкам. Окислы выращивались в атмосфере сухого кислорода (кривая а), во влажной атмосфере (кривая б) и во влажной атмосфере с последующим 30-минутным отжигом при 350 °С в атмосфере H_2 (кривая в). Зонная структура полупроводника лучше проявляется на вольт-амперных характеристиках, снятых при отрицательных напряжениях. Видно, что ток относительно медленно нарастает при увеличении напряжения до значения $V \approx -1$ В, после чего скорость возрастания тока увеличивается. Это напряжение соответствует ширине запрещенной зоны кремния при сильном легировании, приводящем к образованию хвостов плотности состояний вблизи краев зон.

Очевидно, что отрицательное сопротивление, предсказываемое при малых отрицательных напряжениях, как это говорилось

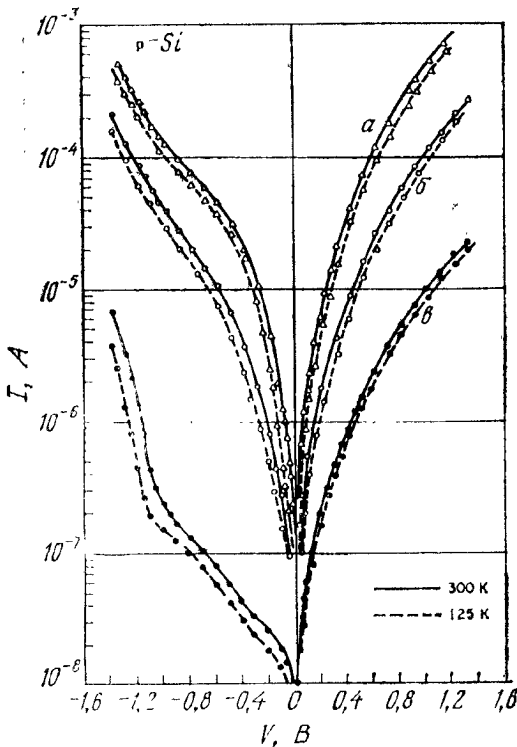
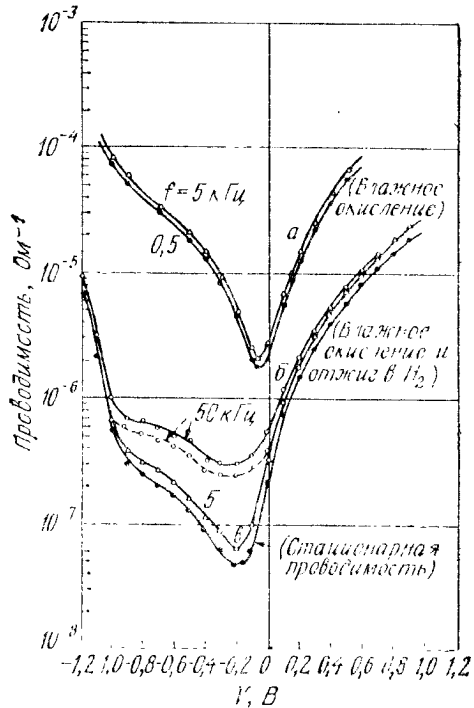


Рис. 20. Вольт-амперные характеристики трех образцов на кремниевой p^{++} -подложке со слоем окисла толщиной 20 Å, созданным различными способами. Характеристики, измеренные при комнатной температуре (300 К), указаны сплошными линиями, а измеренные при температуре жидкого азота (125 К) — штриховыми линиями [40].

выше, маскируется туннелированием электронов из металла на поверхностные состояния. Кривые *a* и *б* принципиально не отличаются от кривой *в*, однако для них характерны существенно высокие значения тока при напряжениях, соответствующих запрещенной зоне ($-1,1 \text{ В} < V < 0$). Если предположить, что величина тока в этом диапазоне напряжений пропорциональна плотности поверхностных состояний, то отсюда следует вывод о возрастании плотности поверхностных состояний на один-два порядка величины при отказе от водородного отжига в случае влажного окисления или при выращивании окисла в атмосфере сухого кислорода. Этот вывод качественно согласуется с экспериментальными данными, полученными при изучении плотности поверхностных состояний, возникающих при аналогичных способах выращивания окислов большой толщины ($d \approx 1000 \text{ Å}$) [43]. Влияние зонной структуры полупроводника и плотности поверхностных состояний на характеристики туннелирования в стационарных условиях оказывается еще более ярко выраженным на зависимости проводимости от напряжения (рис. 21, кри-

Рис. 21. Зависимости проводимости от напряжения, измерения при различных частотах. Кривая для стационарной проводимости получена с помощью дифференцирования кривых, приведенных на рис. 20 [40].
 — 300 К; --- 125 К.



вая в), которая получена дифференцированием кривых, приведенных на рис. 20. Левые ветви кривых при $V < -1,1$ В соответствуют туннелированию электронов из металла в зону проводимости; правые ветви при $V > 0$ соответствуют туннелированию из валентной зоны в металл. Минимум проводимости при малых отрицательных напряжениях есть результат суперпозиции двух компонент тока (рис. 19, з). Очевидно, что ожидаемая отрицательная проводимость, связанная с первой компонентой тока, компенсируется большей положительной проводимостью, связанной с протеканием тока через поверхностные состояния.

На рис. 21 и 22 приведены соответственно зависимости проводимости, измеренной на переменном токе, от приложенного напряжения и частоты. Как и следовало ожидать, кривые для емкости неотожженных образцов расположены выше, чем кривые для емкости образцов, подвергнутых отжигу, поскольку в первом случае плотность поверхностных состояний выше. Частотная зависимость проводимости оказывается более сильной. При частотах, меньших 5 кГц, зависимость проводимости от напряжения практически совпадает со статической. На рис. 22 (вставка) показана основная эквивалентная схема, с помощью которой учиты-

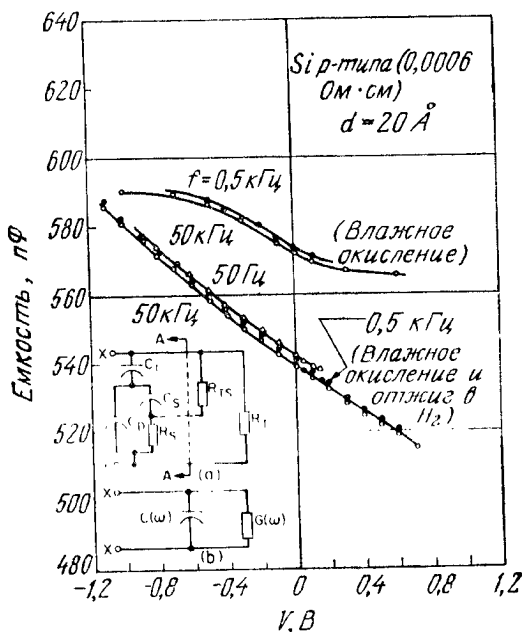


Рис. 22. Зависимости емкости от напряжения, измеренные при различных частотах. На вставке показана эквивалентная схема туннельного МДП-диода [40].

вается влияние туннелирования. RC -цепь, расположенная левее линии AA , идентична RC -цепи для более толстого слоя окисла. R_T — эквивалентное сопротивление, соответствующее туннелированию электронов в валентную зону и (или) в зону проводимости p^{++} -полупроводника (ему соответствуют первая и третья компоненты тока на рис. 19, в и д). R_{TS} — эквивалентное сопротивление, соответствующее туннелированию электронов на поверхностные состояния и последующей их рекомбинации с дырками из валентной зоны (чему соответствует вторая компонента тока на рис. 19, г или д). Величины R_T и R_{TS} являются функциями приложенного напряжения. Полную цепь можно упростить, заменив ее параллельно соединенными емкостью $C(\omega)$ и проводимостью $G(\omega)$, зависящими от частоты.

Можно показать, что $d|C(\omega)|/d\omega \leq 0$ и $d|G(\omega)|/d\omega \geq 0$. Для тонких окисных слоев и (или) сильнолегированных полупроводниковых подложек емкость увеличивается по сравнению с высокочастотным значением [40]

$$C(\infty) = \frac{C_i C_D}{C_i + C_D} \quad (47a)$$

вплоть до величины

$$C(0) \approx C(\infty) + AC_s, \quad (47b)$$

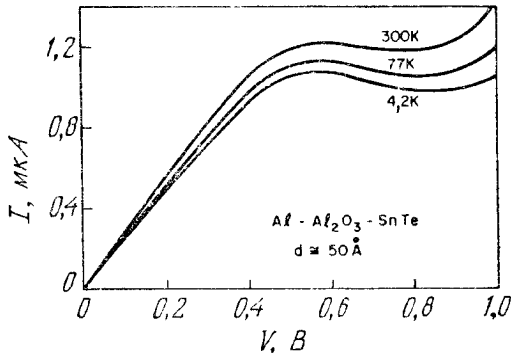


Рис. 23. Вольт-амперные характеристики туннельного МДП-диода (Al — Al₂O₃ — SnTe) при трех температурах [44].

если $\omega \rightarrow 0$. Проводимость увеличивается от низкочастотного значения

$$G(0) = \frac{1}{R_T} + \frac{1}{R_{TS} + R_s} \quad (48a)$$

вплоть до величины

$$G(\infty) = G(0) + \frac{A(R_{TS} + R_s)}{R_{TS}R_s} \quad (48b)$$

при $\omega \rightarrow \infty$. Величина

$$A \equiv \frac{(C_i R_{TS} - C_D R_s)^2}{(C_i + C_D)^2 (R_{TS} + R_s)^2} \quad (49)$$

Экспериментальные результаты находятся в хорошем согласии с приведенными выше качественными соображениями (рис. 21 и 22).

Отрицательное сопротивление наблюдалось в туннельных МДП-диодах из Al—Al₂O₃—SnTe [44]. Материал SnTe представлял собой сильнолегированный полупроводник *p*-типа с концентрацией носителей $8 \cdot 10^{20} \text{ см}^{-3}$; толщина слоя Al₂O₃ составляла $\sim 50 \text{ \AA}$. На рис. 23 приведены измеренные вольт-амперные характеристики такой структуры при трех различных температурах. Участок отрицательного сопротивления существует в диапазоне 0,6—0,8 В. Эти результаты хорошо согласуются с теоретическими результатами [41], полученными по формуле (45).

9.4.2. Невырожденный полупроводник

Туннельные МДП-диоды с очень тонким слоем окисла, выращенным на невырожденной полупроводниковой подложке, не обладают отрицательным сопротивлением. Этот диод используется

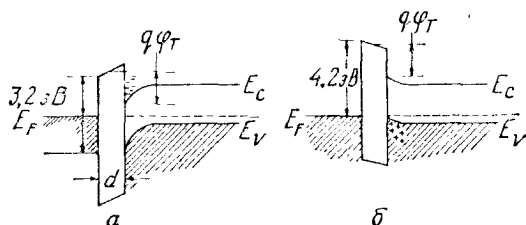


Рис. 24 Энергетические диаграммы туннельных МДП-диодов на невырожденных подложках [45].

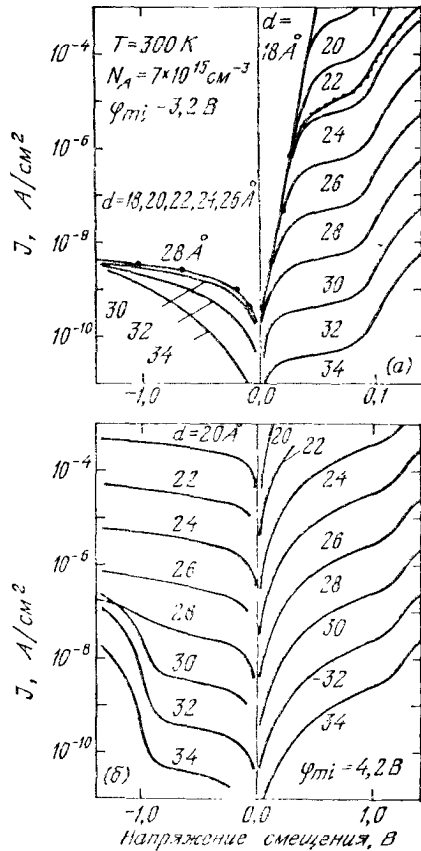
a — низкий барьер в системе металл — диэлектрик; *б* — высокий барьер в системе металл — диэлектрик.

в элементах памяти (например, МНОП-приборы), в преобразователях энергии (например, солнечные батареи) и в СВЧ-приборах в качестве перехода эмиттер—база туннельного транзистора.

Одним из важнейших параметров такого диода является высота барьера на границе металл—диэлектрик, которая существенно влияет на вольт-амперные характеристики [45, 46]. На рис. 24 приведены диаграммы энергетических зон туннельных МДП-диодов, находящихся в состоянии теплового равновесия. Диаграммы соответствуют подложке *p*-типа и двум разным высотам барьера в системе металл—диэлектрик. При малой высоте барьера ($\varphi_{mi} = 3,2$ В, что соответствует системе Al—SiO₂) поверхность дырочного кремния инвертирована. При большой высоте барьера ($\varphi_{mi} = 4,2$ В, что соответствует системе Au—SiO₂) на поверхности кремния накапливаются дырки. Туннельный ток состоит из двух основных компонент: тока из зоны проводимости в металл (J_{ct}) и тока из валентной зоны в металл (J_{vt}). Обе эти компоненты определяются выражениями, подобными формуле (44). На рис. 25 приведены теоретические вольт-амперные характеристики диодов с малой и большой высотой барьера. Если высота барьера невелика (рис. 25, *a*), то при малых прямых и обратных смещениях главным является ток неосновных носителей (электронов) J_{ct} , поскольку уровень Ферми проходит ближе к краю зоны проводимости. При увеличении прямого смещения (положительного напряжения на полупроводнике) ток монотонно возрастает. При фиксированном смещении ток резко возрастает с уменьшением толщины диэлектрика в результате того, что ограничивается вероятностью туннелирования (выражение (46)), которая экспоненциально зависит от толщины диэлектрика. При обратном смещении ток практически не зависит от толщины диэлектрика до тех пор, пока $d < < 30$ Å, поскольку в этом случае ток ограничивается скоростью генерации неосновных носителей (электронов) в объеме полупроводника, т. е. он аналогичен току насыщения обратносмещенного *p*—*n*-перехода. На рис. 25, *a* приведены также экспериментальные

Рис. 25. Вольт-амперные характеристики туннельных МДП-диодов [45].

а — с низким барьером; б — с высоким барьером.



результаты (они представлены черными точками), полученные при $d = 23,5 \text{ \AA}$. Отметим хорошее соответствие теоретических данных экспериментальным, а также то, что полученные вольт-амперные характеристики подобны вольт-амперным характеристикам $p-n$ -перехода.

Если высота барьера большая (рис. 25, б), то при прямых смещениях главным является ток основных носителей (дырок), туннелирующих из валентной зоны в металл, вследствие чего величина тока экспоненциально возрастает с уменьшением толщины диэлектрика. Однако в этом случае и при обратных смещениях не должно быть такой независимости тока от толщины диэлектрика, которая проиллюстрирована рис. 25, а. Вместо этого ток резко возрастает с уменьшением толщины диэлектрика, поскольку ток основных носителей в обоих направлениях ограни-

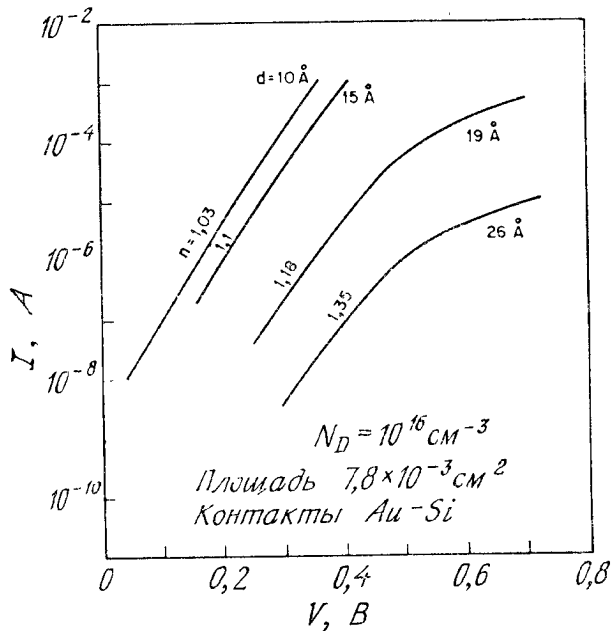


Рис. 26. Экспериментальные вольт-амперные характеристики туннельных МДМ-диодов с различной толщиной слоя оксида [47].

чивается вероятностью туннелирования, а не скоростью генерации носителей.

Подставив формулу (46а) в выражение (45) и оценив интеграл по энергиям, получим соотношение [42, 47]

$$J = A^* T^2 e^{(-\alpha_T \varphi_T^{1/2} d)} e^{-q\varphi_B/kT} (e^{qV/kT} - 1), \quad (50)$$

где $A^* = 4\pi m^* q^2 k^2 h^{-3}$ — эффективная постоянная Ричардсона, φ_B — высота барьера. Выражение (50) идентично обычному выражению для термоэлектронной эмиссии в случае барьера Шоттки, за исключением множителя $\exp(-\alpha_T \varphi_T^{1/2} d)$, который представляет собой вероятность туннелирования. Отсюда следует, что при $\varphi_T \approx 1$ В и $d > 50$ Å вероятность туннелирования составляет $\sim e^{-50} \approx 10^{-22}$ и туннельным током действительно можно пренебречь. При уменьшении d и (или) φ_T ток быстро возрастает вплоть до величины тока термоэлектронной эмиссии. На рис. 26 приведены прямые ветви вольт-амперных характеристик четырех туннельных диодов типа Au—SiO₂—Si. При $d = 10$ Å зависимость тока от напряжения совпадает с обычной зависимостью для диода

Шоттки с фактором идеальности $n \approx 1$. При увеличении толщины диэлектрика ток резко уменьшается, а фактор идеальности начинает отличаться от 1.

9.5. ПЕРЕКЛЮЧАЮЩИЙ МДП-ДИОД

Переключающий МДП-диод представляет собой четырехслойную структуру (рис. 27, а). Этот диод, так же как и диод Шокли (гл. 4), был предложен для реализации S-образной вольт-амперной характеристики (рис. 27, б [48]). Когда к прибору прикладывается отрицательное напряжение (p^+ -подложка при этом заземлена), реализуется вольт-амперная характеристика с большим значением импеданса (прибор находится в состоянии «выключено»). При достаточно большом напряжении (напряжении переключения V_S) прибор резко переключается в состояние с большим током и низким напряжением (в состояние «включено»). Переключение

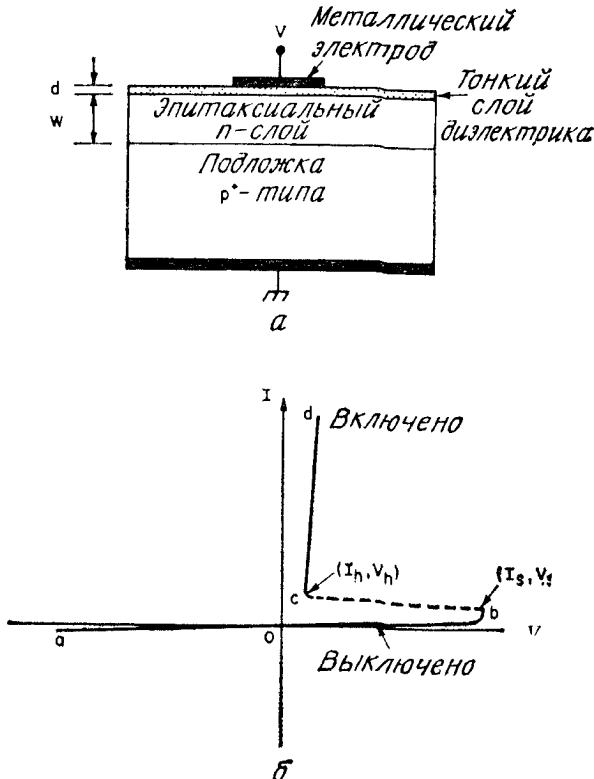


Рис. 27. Переключающий МДП-диод (а) и его S-образная вольт-амперная характеристика (б).

происходит либо благодаря расширению приповерхностного обедненного слоя вплоть до границы $p^+ - n$ -перехода (прокол n -слоя), либо благодаря лавинному пробою вблизи поверхности n -слоя [49]. К достоинствам переключающего МДП-диода относится совпадение уровня токов и напряжений с уровнем, который используется в цепях цифровой техники, высокая скорость переключения (1 нс и меньше) и высокая чувствительность напряжения переключения V_S к оптической или электрической инжекции. Впервые такой диод был изготовлен из кремния с использованием SiO_2 в качестве туннельного диэлектрика. Эффект переключения наблюдался также при использовании других диэлектриков (например, Si_3N_4) и толстых слоев поликристаллического кремния. Было показано, что, так же как в тиристоре (гл. 4), инжекция носителей в n -слой через третий контакт контролирует величину напряжения переключения.

Для того чтобы понять эффект переключения, рассмотрим энергетические диаграммы, приведенные на рис. 28 [50]. На рис. 28, а показана диаграмма идеального диода в состоянии теплового равновесия, а на рис. 28, б — аналогичная диаграмма для случая, когда на металлический электрод подано положительное напряжение. Если проводимость туннельного диэлектрика гораздо выше, чем проводимость обратносмещенного $p^+ - n$ -перехода, то большая часть приложенного напряжения падает на $p^+ - n$ -переходе. При этом вольт-амперная характеристика прибора практически та же, что и для обратносмещенного $p^+ - n$ -перехода, и определяется генерацией носителей в обедненном слое:

$$J_s = \frac{qn_i W_j}{2\tau_g} \approx \frac{n_i}{\tau_g} \left(\frac{q^2}{2N_D} \right)^{1/2} (V + V_{bi})^{1/2}, \quad (51)$$

где n_i — собственная концентрация носителей, τ_g — время жизни неосновных носителей, W_j — ширина обедненного слоя, N_D — концентрация доноров в n -слое, V_{bi} — контактная разность потенциалов, V — приложенное напряжение. Поведение этого тока иллюстрируется участком кривой $0a$ вольт-амперной характеристики, приведенной на рис. 27, б.

При увеличении отрицательного напряжения на электроде (рис. 28, в) свободные электроны вытягиваются из эпитаксиального слоя через $p^+ - n$ -переход и обедненный слой под электродом расширяется. При генерации электронно-дырочных пар в этом слое электроны вытягиваются из прибора через $p^+ - n$ -переход, а дырки подтягиваются к границе раздела $\text{SiO}_2 - \text{Si}$. В случае толстого слоя окисла ($d \geq 50 \text{ \AA}$) дырки накапливаются у границы раздела и инвертируют поверхность. Если же слой окисла достаточно тонкий, дырки туннелируют через него. Тогда при увеличении напряжения происходит обеднение поверхности n -слоя

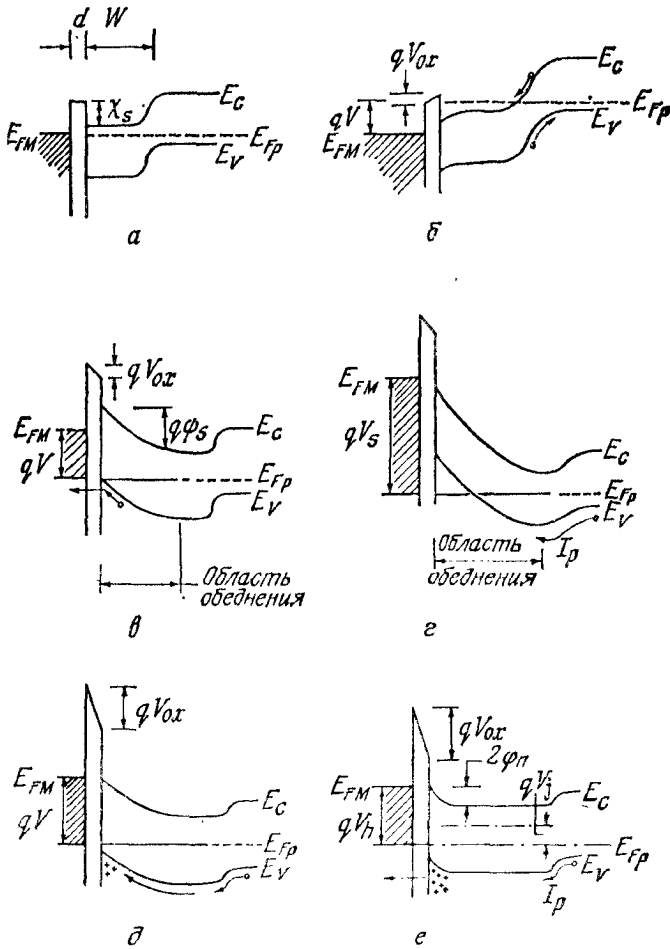


Рис. 28. Энергетические диаграммы переключающего МДП-диода при различных напряжениях [50].

вместо инверсии. При этом генерационный ток определяется выражением, аналогичным выражению (51), но в последнем множитель $(V + V_{dt})^{1/2}$ заменяется на $(\psi_s)^{1/2}$, где ψ_s — величина поверхностного потенциала. Этот ток также слабо зависит от напряжения (участок кривой *Ob* на рис. 27, б).

Если приложено достаточно большое напряжение, например $V = V_s$ (напряжение переключения), обедненный слой под окис-

лом расширяется вплоть до $p^+ - n$ -перехода (рис. 28, *г*), и происходит прокол перехода. Напряжение прокола равно

$$V_s \approx \frac{qN_D(W - W_{j_0})^2}{2\epsilon_s}, \quad (52)$$

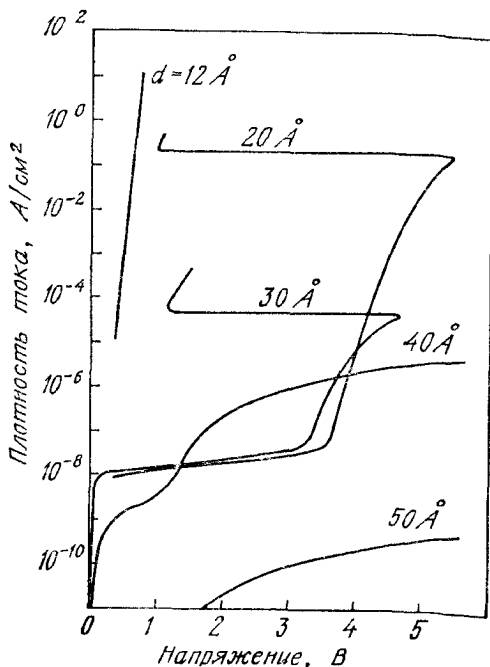
где W — толщина эпитаксиального слоя n -типа, W_{j_0} — толщина обедненного слоя $p^+ - n$ -перехода при нулевом смещении. При проколе уровень инжекции дырок из перехода возрастает экспоненциально с увеличением смещения перехода. В первый момент после прокола величина поля в окисле оказывается недостаточной для того, чтобы большой поток инжектированных дырок мог туннелировать через диэлектрик, поскольку при этом ток ограничен туннелированием. Следовательно, большое увеличение напряжения приведет к накоплению дырок на границе раздела $\text{Si} - \text{SiO}_2$. Это накопление переведет поверхность из глубокого обеднения в область инверсии (рис. 28, *д*); при этом ψ_s и напряжение на приборе уменьшаются. Падение напряжения на окисле увеличится, вследствие чего возрастет ток электронов, туннелирующих из металла. Этот ток увеличит смещение на $p^+ - n$ -переходе. Данный механизм обеспечивает положительную обратную связь и приводит к появлению на вольт-амперной характеристике участка с отрицательным дифференциальным сопротивлением (участок кривой *bc* на рис. 27, *б*). Из энергетической диаграммы прибора при остаточном напряжении (рис. 28, *е*) видно, что остаточное напряжение равно

$$V_h = V_{ox} + \psi_s + V_j, \quad (53)$$

где V_{ox} — падение напряжения на окисле, ψ_s — поверхностный потенциал, V_j — напряжение на $p^+ - n$ -переходе. Приведенный анализ касается случая переключения в условиях прокола n -слоя. Тот же механизм положительной обратной связи ответственный за переключение в условиях лавинного пробоя. При этом напряжение переключения, определяемое выражением (52), следует заменить на напряжение лавинного пробоя n -слоя вблизи границы раздела $\text{Si} - \text{SiO}_2$.

В соответствии с проведенным анализом основным параметром для эффекта переключения служит толщина слоя окисла. Этот вывод подтверждают вольт-амперные характеристики, приведенные на рис. 29 [49]. Для слоев окисла большой толщины ($d \geq 50 \text{ \AA}$) сопротивление туннелирования слишком высоко для переключения диода при низких напряжениях ($V \leq 5 \text{ В}$). В случае очень тонкого слоя окисла ($d < 15 \text{ \AA}$) переход может полностью сместиться в прямом направлении еще до того, как произойдет глубокое обеднение поверхности; следовательно, вольт-амперная

Рис. 29. Рассчитанные вольтамперные характеристики переключающего МДП-диода при различной толщине слоя окисла. Параметры прибора: $W = 10$ мкм, $N_D = 10^{18}$ см $^{-3}$ и $\tau_g = 3,5 \cdot 10^{-5}$ с [49].



характеристика прибора будет совпадать с характеристикой $p-n$ -перехода. Эффект переключения наблюдался только на слоях окисла промежуточной толщины ($15 \text{ \AA} < d < 40 \text{ \AA}$).

9.6. туннельный мдм-диод

Туннельный диод на структуре металл—диэлектрик—металл (МДМ) представляет собой тонкопленочный прибор, в котором электроны из первого металла могут туннелировать через пленку диэлектрика и коллектироваться вторым металлом. На рис. 30 (вставка) приведены основные энергетические диаграммы симметричного МДМ-диода. Поскольку все приложенное напряжение падает на диэлектрике, плотность туннельного тока через него в соответствии с выражением (44) равна

$$J = \frac{4\pi q m^*}{\hbar^3} \iint T_t [F(E) - F(E + qV)] dE_{\perp} dE. \quad (54)$$

При абсолютном нуле температуры выражение (54) принимает более простой вид [51]

$$J = J_0 [\bar{\varphi} \exp(-A\sqrt{\bar{\varphi}}) - (\bar{\varphi} + V) \exp(-A\sqrt{\bar{\varphi} + V})], \quad (55)$$

где

$$J_0 \equiv q^2 / [2\pi\hbar (\Delta d)^2], \quad A \equiv 4\pi (\Delta d) \sqrt{2mq/h},$$

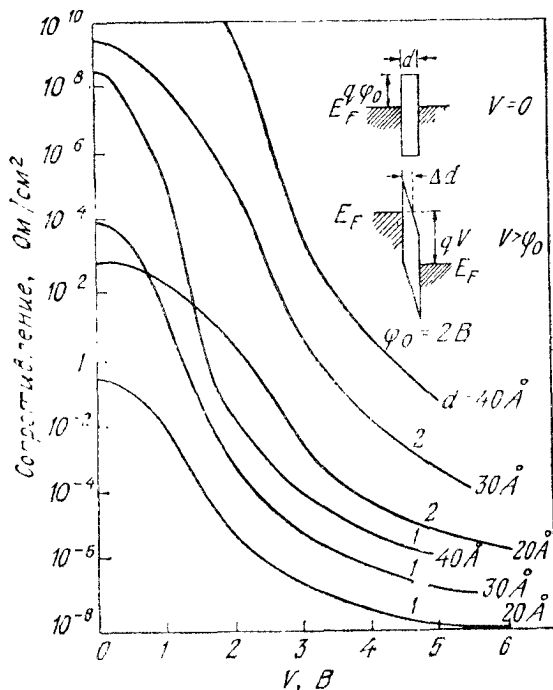


Рис. 30. Зависимости туннельного сопротивления симметричной МДМ-структуры от напряжения. На вставке показаны зонные диаграммы при $V = 0$ и при $V > \varphi_0$ [51].

а $\bar{\varphi}$ — средняя высота барьера над уровнем Ферми. Выражение для плотности тока (55) можно интерпретировать как результат суперпозиции тока с плотностью $J_0 \bar{\varphi} \exp(-A \sqrt{\bar{\varphi}})$, текущего от электрода 1 к электроду 2, и тока с плотностью $J_0 (\bar{\varphi} + V) \exp(-A \sqrt{\bar{\varphi} + V})$, текущего от электрода 2 к электроду 1.

Ниже мы используем выражение (55) для описания свойств идеальной МДМ-структуры (рис. 30). Под идеальностью здесь понимается пренебрежение влиянием температуры, сил изображения и проникновения поля в глубь металлических электродов. При $0 \leq V \leq \varphi_0$, $\Delta d = d$ и $\bar{\varphi} = \varphi_0 - V/2$ плотность тока равна

$$J = J_0 [(\varphi_0 - V/2) \exp(-A \sqrt{\varphi_0 - V/2}) - (\varphi_0 + V/2) \exp(-A \sqrt{\varphi_0 + V/2})]. \quad (56)$$

При большем напряжении ($V > \varphi_0$) имеем $\Delta d = d\varphi_0/V$ и $\bar{\varphi} = \varphi_0/2$. Тогда плотность тока равна

$$J = \frac{q^2 \mathcal{E}^2}{4\pi \hbar \varphi_0} [\exp(-\mathcal{E}_0/\mathcal{E}) - (1 + 2V/\varphi_0) \exp(-\mathcal{E}_0 \sqrt{1 + 2V/\varphi_0}/\mathcal{E})], \quad (57)$$

где $\mathcal{E} = V/d$ — поле в диэлектрике, а $\mathcal{E}_0 \equiv \frac{8}{3} \sqrt{\pi q} (\varphi_0)^{3/2}$. При очень больших напряжениях, таких, что $V > (\varphi_0 + E_F/q)$, вторым слагаемым в квадратных скобках в выражении (57) можно пренебречь, и тогда мы получим хорошо известную формулу Фаулера—Нордгейма. На рис. 30 приведены результаты расчета туннельного сопротивления V/J при разных высотах барьера и толщинах слоя диэлектрика. Отметим, что туннельное сопротивление быстро уменьшается с увеличением приложенного напряжения.

Для идеальной асимметричной МДМ-структуры (рис. 31, вставка) в диапазоне низких напряжений $0 < V < \varphi_1$ величины Δd и $\bar{\varphi}$ равны d и $(\varphi_1 + \varphi_2 - V)/2$ и не зависят от полярности напряжения. При этом не зависит от полярности и вольт-амперная характеристика. При больших напряжениях ($V > \varphi_2$) средняя высота барьера $\bar{\varphi}$ и эффективная длина туннелирования Δd начинают зависеть от полярности. Поэтому значения тока при различных полярностях оказываются разными. На рис. 31 приведены зависимости туннельного сопротивления от напряжения при $d = 20, 30, 40 \text{ \AA}$, $\varphi_1 = 1 \text{ В}$ и $\varphi_2 = 2 \text{ В}$. Экспериментальные результаты, полученные на структурах $\text{Al}-\text{Al}_2\text{O}_3-\text{Al}$ при $d = 30 \text{ \AA}$, $\varphi_1 = 1,6 \text{ В}$ и $\varphi_2 = 2,5 \text{ В}$, хорошо согласуются с предсказаниями теории, основанными на выражениях (56) и (57). Однако эффективная площадь, использованная для сравнения с экспериментальными данными, была принята равной 1 % площади электрода [52]. Совсем иная ситуация наблюдается для туннельных МДМ-диодов на структурах $\text{Al}-\text{SiO}_2-\text{SiO}_2$, где эффективная площадь обычно равна площади алюминиевого электрода. Относительно малые значения площади туннелирования в МДМ-диоде могут быть объяснены статистической природой образования диэлектрических пленок на металлической подложке [53]. Туннельный ток определяется только самыми тонкими участками диэлектрической пленки. Из-за статистических флуктуаций толщины пленки емкость МДМ-структуры всегда оказывается больше той, которая вычисляется исходя из средней толщины слоя диэлектрика [54]. Кроме того, на величину емкости может влиять распределение потенциала на металлических электродах. В результате проникновения электрического поля в глубь металла

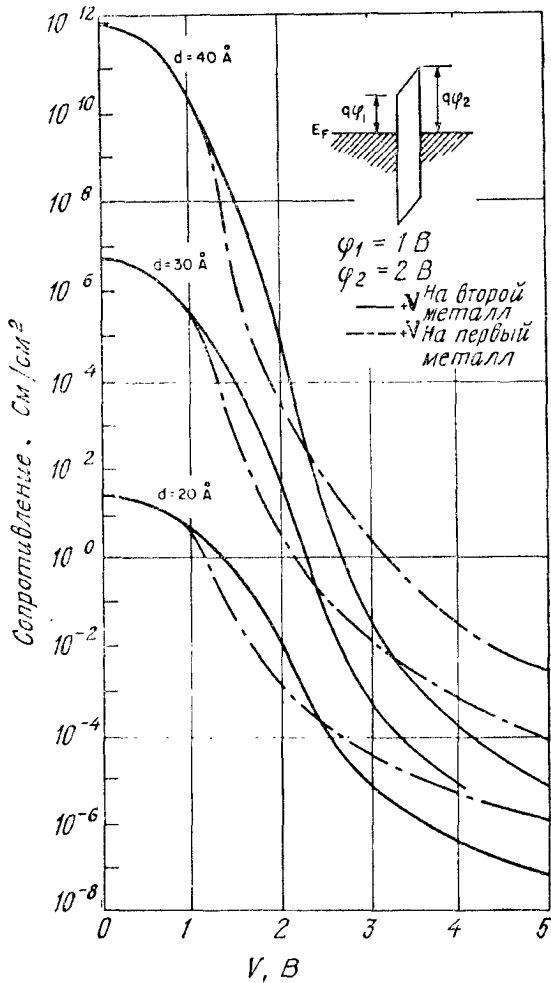


Рис. 31. Зависимости туннельного сопротивления асимметричной МДМ-структуры от напряжения. На вставке показана зонная диаграмма при $V = 0$ [51].

полная емкость фактически равняется двум последовательно соединенным емкостям [55]:

$$1/C = d/\epsilon_i + 2,3\lambda_m/\epsilon_m, \quad (58)$$

где ϵ_i и ϵ_m — диэлектрические проницаемости диэлектрика и металла соответственно, а λ_m — характерная глубина проникновения электрического поля в металл ($\sim 0,5 \text{ \AA}$ для благородных металлов). Из выражения (58) видно, что емкость, обусловленная

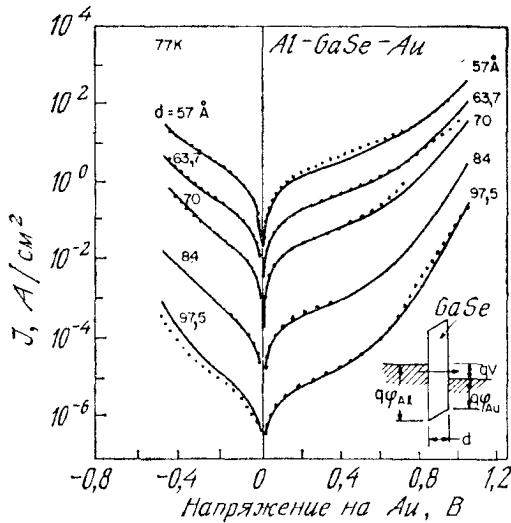


Рис. 32. Зависимость плотности тока от приложенного напряжения для нескольких структур Al — GaSe — Au. Точками представлены экспериментальные данные, кривыми — результаты вычислений. На вставке показана зонная диаграмма при приложенном напряжении [56].

вторым слагаемым, составляет 10^6 пФ/см². Полная величина емкости единицы площади при $d \rightarrow 5$ Å окажется гораздо меньше ϵ_s/d .

Туннельные МДМ-диоды использовались для изучения связи между энергией и импульсом в запрещенной зоне монокристаллических диэлектриков, например GaSe ($E_g = 2,0$ эВ, $\epsilon_s/\epsilon_0 = 8$ и плотность носителей $p \approx 3 \cdot 10^{14}$ см⁻³ при 300 K). GaSe имеет слоистую кристаллическую структуру, что упрощает изготовление монокристаллических чешуек требуемой толщины (< 100 Å). Туннельную МДМ-структуру создают путем размещения монокристаллического образца между двумя металлическими электродами, например Al—GaSe—Au (барьер на границе Al—GaSe равен 1,08 В, а на границе Au—GaSe 0,52 В). Зонная диаграмма такой структуры приведена на рис. 32 (вставка). Здесь приведены также результаты измерений на структурах Al—GaSe—Au различной толщины. Используя одну из $J(V)$ -кривых, можно с помощью формул (44) и (54) вычислить зависимость импульса от энергии $k(E)$ для GaSe. Определение зависимости $k(E)$ позволяет без использования подгоночных параметров вычислить значения туннельного тока при любой толщине слоя диэлектрика. Таким способом были рассчитаны теоретические вольт-амперные характеристики, приведенные на рис. 32, из которого видно идеальное

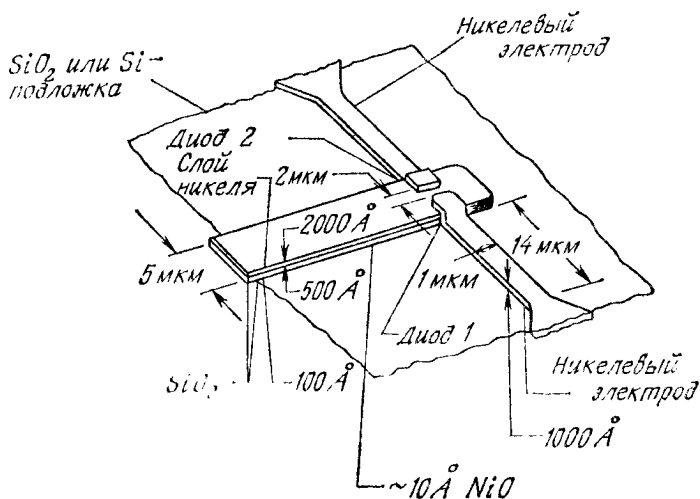


Рис. 33. Схема краевого МОМ-диода [57].

совпадение между теоретическими и экспериментальными данными для всех значений напряжения и толщины.

Стабильным и воспроизводимым МДМ-диодом является крайевой МОМ-диод (металл—окисел—металл), который отличается строго фиксированной площадью туннелирования [57]. На рис. 33 приведено схематическое изображение такого прибора. Перекрытие между металлической (Ni) полоской шириной 1 мкм и оксидированным краем металлической (Ni) полоски, толщина которой близка к 100 Å, приводит к площади туннелирования $\sim 10^{-10}$ см². Толщина слоя окисла (NiO) составляет 10 Å. Коэффициент кризисы γ (уравнение (42)) для краевого МОМ-диода со структурой Ni—NiO—Ni лежит в диапазоне 1—10. Этот прибор можно использовать в качестве широкополосного детектора и смесителя, поскольку нелинейность его вольт-амперной характеристики постоянна в диапазоне от звуковых частот и вплоть до частот ближнего инфракрасного излучения.

9.7. ТУННЕЛЬНЫЙ ТРАНЗИСТОР

За последние годы было предпринято много попыток изобрести или сконструировать новые твердотельные приборы с лучшими рабочими характеристиками по сравнению с биполярными или полевыми транзисторами. К числу наиболее интересных предложений относятся туннельные транзисторы. Первый туннельный транзистор был создан Мидом в 1960 г. [58]. На рис. 34, а по-

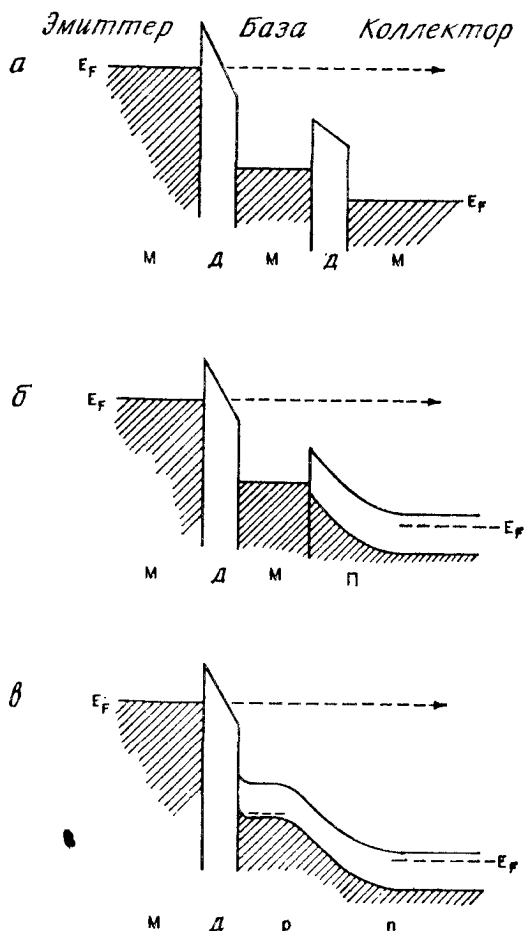


Рис. 34. Туннельный МДМДМ-транзистор (а), туннельный МДМП-транзистор (б) и туннельный МД (р — п)-транзистор (в) [58, 60, 61].

казана энергетическая диаграмма прибора со структурой металл—диэлектрик—металл—диэлектрик—металл (МДМДМ), в котором ток течет через первый диэлектрический слой благодаря туннелированию. После туннелирования из эмиттера в базу энергия электронов не менее чем на несколько kT превышает энергию Ферми в базе; эти электроны называются горячими, поскольку они не находятся в тепловом равновесии с решеткой. Такой транзистор называют также транзистором на горячих электронах [59]. Спратт, Шварц и Кейн установили, что коэффициент усиления по току такой структуры можно существенно

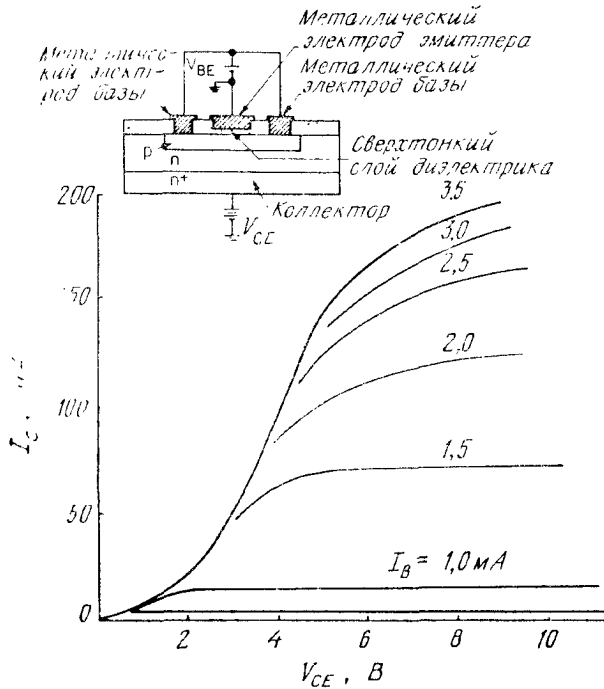


Рис. 35. Вольт-амперные характеристики коллектора туннельного МД ($p-n$)-транзистора. На вставке показана диаграмма поперечного сечения прибора [61].

улучшить, заменив коллектирующий диэлектрик полупроводником, образующим барьер Шоттки со вторым металлом (рис. 34, б) [60]. Максимальная рабочая частота для МДМП-структуры меньше, чем для биполярного транзистора в основном из-за того, что в такой структуре время зарядки эмиттера больше (что обусловлено большой емкостью эмиттера), а коэффициент усиления по току в схеме с общей базой меньше (что обусловлено интенсивным рассеянием горячих электронов в области базы).

На рис. 34, в приведена энергетическая диаграмма туннельного транзистора [61], представляющего собой МДП-структуру, в которой электроны могут туннелировать из эмиттирующего металлического электрода в базу (слой p -типа, полученный диффузией). На рис. 35 (вставка) показана диаграмма поперечного сечения такой структуры. В качестве металлического электрода используется Al, нанесенный на тонкий слой SiO_2 толщиной около 20 Å, термически выращенный в атмосфере сухого кислорода (550 °C в течение 30 мин). Толщина базы p -типа составляет ~1500 Å, а концентрация акцепторов в ней ~ 10^{18} см $^{-3}$. Концентрация до-

норов в эпитаксиальном слое n -типа $8 \cdot 10^{17} \text{ м}^{-3}$. На рис. 35 приведены также вольт-амперные характеристики коллектора. Коэффициент усиления по току в схеме с общим эмиттером при низких уровнях тока в базе может превышать 100.

В работе [62] предложен транзистор с туннельно-тонкой базой, состоящий из двух гетеропереходов. На рис. 36, *a* приведена зонная диаграмма такого транзистора в состоянии теплового равновесия. Для нормального функционирования этого транзистора необходимо выполнение трех условий: 1) разрывы зон ΔE_C и ΔE_V должны происходить в таком направлении и иметь такую величину, чтобы образовались яма для электронов и барьер для дырок (рис. 36, *a*); 2) n -область при всех условиях должна оставаться проводящей; 3) толщина базы и конфигурация энергетических зон должны быть такими, чтобы основным током

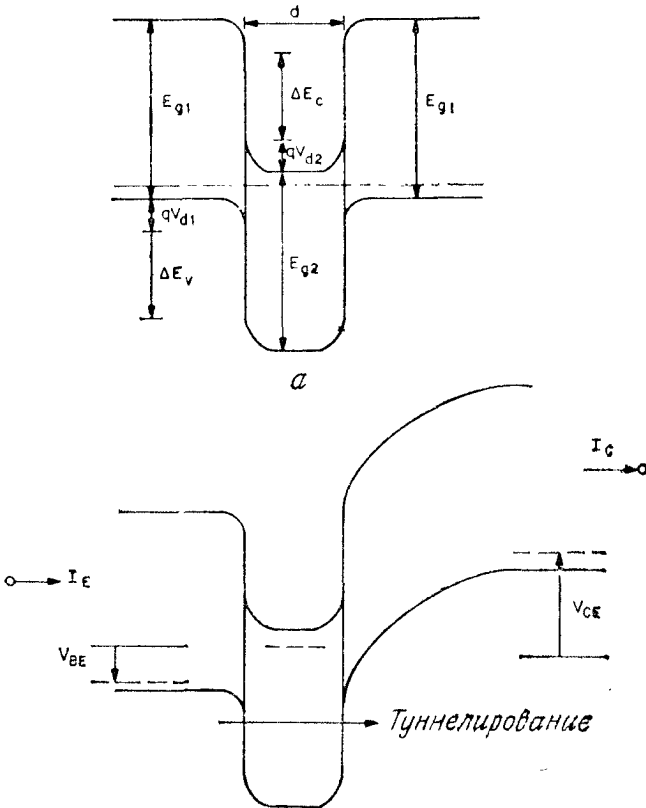


Рис. 36. Энергетические диаграммы транзистора с туннельно-тонкой базой в состоянии теплового равновесия (*a*) и при напряжении (*b*) [62].

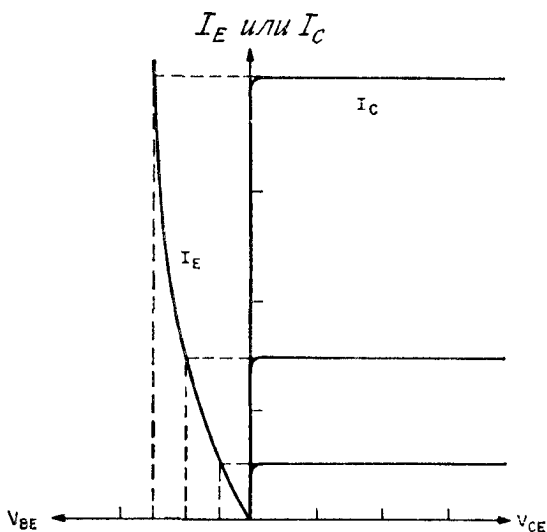


Рис. 37. Зависимость тока эмиттера и тока коллектора от напряжения на базе и коллекторе [62].

был туннельный. Этим условиям удовлетворяют, например, $\text{GaAs}_{0,5}\text{Sb}_{0,5}$ p -типа с $N_A \approx 10^{18} \text{ см}^{-3}$ в качестве материала эмиттера и коллектора и $\text{Ga}_{0,5}\text{In}_{0,5}\text{As}$ n -типа с $N_D \approx 10^{19} \text{ см}^{-3}$ при $d \approx 50 \text{ \AA}$ в качестве материала базы. Для этой пары $E_{g1} \approx E_{g2} = 0,8 \text{ эВ}$ и $\Delta E_C = \Delta E_V = 0,5 \text{ эВ}$. Выращивание этих слоев можно осуществить методом молекулярно-лучевой эпитаксии.

На рис. 36, б приведена зонная диаграмма такого туннельного транзистора при подаче напряжения V_{CE} . Основным током здесь будет туннельный ток через базу, для которого можно записать $J_t \sim \exp(qV_{BE}/kT)$. На рис. 37 приведены зависимости тока эмиттера I_E и тока коллектора I_C от напряжения. Близость величин I_E и I_C означает, что эффективность эмиттера и коэффициент переноса по базе равны 1, а следовательно, транзистор имеет большой коэффициент усиления по току. Этот туннельный транзистор обладает тремя преимуществами перед обычным биполярным транзистором: 1) пренебрежимо малым временем пролета через базу, поскольку времена туннелирования всегда малы; 2) малой емкостью эмиттера, что обусловлено его слабым легированием; 3) низким сопротивлением базы, обусловленным ее сильным легированием. Можно ожидать, что благодаря этим достоинствам транзистор с туннельно-тонкой базой будет иметь лучшие характеристики в СВЧ-диапазоне.

ЛИТЕРАТУРА

1. Esaki L. New Phenomenon in Narrow Germanium $p-n$ Junctions, *Phys. Rev.* **109**, 603 (1958); Long Journey into Tunneling, *Proc. IEEE*, **62**, 825 (1974); Discovery of the Tunnel Diode, *IEEE Trans. Electron Devices*, **ED-23**, 644 (1976).
2. Thornber K. K., McGill Thomas C., Mead C. A. The Tunneling Time of an Electron, *J. Appl. Phys.*, **38**, 2384 (1967).
3. Hall R. N. Tunnel Diodes, *IRE Trans. Electron Devices*, **ED-7**, 1 (1960).
4. Morgan J. V., Kane E. O. Observation of Direct Tunneling in Germanium, *Phys. Rev. Lett.*, **3**, 466 (1959).
5. Ландау Л. Д., Лифшиц Е. М. Квантовая механика, ГИФМЛ, Москва, 1963, стр. 211.
6. Kane E. O. Theory of Tunneling. *J. Appl. Phys.*, **32**, 83 (1961); Tunneling in InSb, *Phys. Chem. Solids*, **2**, 181 (1960).
7. Demassa T. A., Knott D. P. The Prediction of Tunnel Diode Voltage — Current Characteristics, *Solid State Electron.*, **13**, 131 (1970).
8. Kroemer H. The Einstein Relation for Degenerate Carrier Concentration, *IEEE Trans. Electron Devices*, **ED-25**, 850 (1978).
9. Meyerhofer D., Brown G. A., Sommers H. S., Jr. Degenerate Germanium I, Tunnel, Excess, and Thermal Current in Tunnel Diodes, *Phys. Rev.*, **126**, 1329 (1962).
10. Moll J. L. Physics of Semiconductors, McGraw-Hill, N. Y., 1964, p. 252.
11. Butcher P. N., Hulme K. F., Morgan J. R. Dependence of Peak Current Density on Acceptor Concentration in Germanium Tunnel Diodes, *Solid State Electron.*, **3**, 358 (1962).
12. Келдыш Л. В. Поведение неметаллических кристаллов в сильных электрических полях, *ЖЭТФ*, **6**, 763 (1958).
13. Chynoweth A. G., Feldmann W. L., Logan R. A. Excess Tunnel Current in Silicon Esaki Junctions, *Phys. Rev.*, **121**, 684 (1961).
14. Roy D. K. On the Prediction of Tunnel Diode I—V Characteristics, *Solid State Electron.*, **14**, 520 (1971).
15. Carr W. N. Reversible Degradation Effects in GaSb Tunnel Diodes, *Solid State Electron.*, **5**, 261 (1962).
16. Nanavati R. P., Morato De Andrade C. A. Excess Current in Gallium Arsenide Tunnel Diodes, *Proc. IEEE*, **52**, 869 (1964).
17. Holonyak N., Jr. Evidences of States in the Forbidden Gap of Degenerate GaAs and InP — Secondary Tunnel Current and Negative Resistance, *J. Appl. Phys.*, **31**, 130 (1960).
18. Franks V. M., Hulme K. F., Morgan J. R. An Alloy Process for Making High Current Density Silicon Tunnel Junction, *Solid State Electron.*, **8**, 343 (1965).
19. Kleinknecht H. P. Indium Arsenide Tunnel Diodes, *Solid State Electron.*, **2**, 133 (1961).
20. Burrus C. A. Indium Phosphide Esaki Diodes, *Solid State Electron.*, **3**, 357 (1962).
21. Yoshihito A., Konagai M., Sakai Y. Mixed Crystal Tunnel Diode, *Jpn. J. Appl. Phys.*, **12**, 480 (1973).
22. Minton R. M., Glicksman R. Theoretical and Experimental Analysis of Germanium Tunnel Diode Characteristics, *Solid State Electron.*, **7**, 491 (1964).
23. Logan R. A., Augustyniak W. M., Gilber J. F. Electron Bombardment Damage in Silicon Esaki Diodes, *J. Appl. Phys.*, **32**, 1201 (1961).
24. Bernard W., Rindner W., Roth H. Anisotropic Stress Effect on the Excess Current in Tunnel Diodes., *J. Appl. Phys.*, **35**, 1860 (1964).
25. Голованов В. В., Панахов А. З. Влияние гидростатического давления на туннельный ток в GaAs-диодах. ФТП, **6**, (1972).

26. Davis R. E., Gibbons G. Design Principles and Construction of Planar Ge Esaki Diodes, *Solid State Electron.*, **10**, 461 (1967).
27. Hindin H. J. Tunnel Diode Flying High, *Electronics*, **52**, 81 (1979).
28. Standards on Definitions, Symbols, and Methods of Test for Semiconductor Tunnel (Esaki) Diodes and Backward Diodes, *IEEE Trans. Electron Devices*, **12**, 374 (1965).
29. Hauer W. B. Definition and Determination of the Series Inductance of Tunnel Diodes, *IRE Trans. Electron Devices*, **8**, 470 (1961).
30. Esaki L., Miyahara Y. A New Device Using the Tunneling Process in Narrow $p-n$ Junctions, *Solid State Electron.*, **1**, 13 (1960).
31. Hall R. N., Racette J. H., Ehrenreich H. Direct Observation of Polarons and Phonons During Tunneling in Group 3—5 Semiconductor Junctions, *Phys. Rev. Lett.*, **4**, 456 (1960).
32. Chynoweth A. G., Logan R. A., Thomas D. E. Thion-Assisted Tunneling in Silicon and Germanium Esaki Junctions, *Phys. Rev.*, **125**, 877 (1962).
33. Shurmer H. V. Backward Diodes as Microwave Detectors, *Proc. Inst. Electr. Eng., Lond.*, **111**, 15111 (1964).
34. Eng. S. T. Low Noise Properties of Microwave Backward Diodes, *IRE Trans. Microwave Theory Tech.*, **MTT 8**, 419 (1961).
35. Hopkins J. B. Microwave Backward Diodes In InAs, *Solid State Electron.*, **13**, 697 (1970).
36. Bhattacharyya A. B., Sami S. L. Switching Time Analysis of Backward Diodes, *Proc. IEEE* **58**, 513 (1970).
37. Torrey H. C., Whitmer C. A., Crystal Rectifiers, McGraw-Hill, N. Y., 1948, Ch. 8.
38. Sze S. M., Ryder R. M. The Nonlinearity of the Reverse Current — Voltage Characteristics of a $p-n$ Junction near Avalanche Breakdown, *Bell Syst. Tech. J.*, **46**, 1135 (1967).
39. Karlovsky J. The Curvature Coefficient of Germanium Tunnel and Backward Diodes, *Solid State Electron.*, **10**, 1109 (1967).
40. Dahlke W. E., Sze S. M. Tunneling in Metal — Oxide — Silicon Structures, *Solid State Electron.*, **10**, 865 (1967).
41. Chang L. L., Stiles P. J., Esaki L. Electron Tunneling between a Metal and a Semiconductor: Characteristics of Al — Al₂O₃ — SnTe and GeTe Junctions, *J. Appl. Phys.*, **38**, 4440 (1967).
42. Kumar V., Dahlke W. E. Characteristics of Cr — SiO₂ — nSi Tunnel Diodes, *Solid State Electron.*, **20**, 143 (1977).
43. Nicollian E. H., Goetzberger A. The Si — SiO₂ Interface Electrical Properties as Determined by the MIS Conductance Technique, *Bell Syst. Tech. J.*, **46**, 1055 (1967).
44. Esaki L., Stiles P. J. New Type of Negative Resistance In Barrier Tunneling, *Phys. Rev. Lett.*, **16**, 1108 (1966).
45. Green M. A., King F. D., Schewchun J. Minority Carrier MIS Tunnel Diodes and Their Application to Electron and Photovoltaic Energy Conversion: I. Theory, *Solid State Electron.*, **17**, 551 (1974). II Experiment, *Solid State Electron.*, **17**, 563 (1974).
46. Temple V. A. K., Green M. A., Schewchun J. Equilibrium-to-Nonequilibrium Transition in MOS Tunnel Diodes, *J. Appl. Phys.*, **45**, 4934 (1974).
47. Card H. C., Rhoderick E. H. Studies of Tunnel MOS Diodes I. Interface Effects in Silicon Schottky Diodes, *J. Phys. D Appl. Phys.*, **4**, 1589 (1971).
48. Yamamoto T., Morimoto M. Thin-MIS-Structure Si Negative Resistance Diode, *Appl. Phys. Lett.*, **20**, 269 (1972).
49. Habib S. E.-D., Simmons J. G. Theory of Switching In $p-n$ Insulator (Tunnel) — Metal Devices, *Solid State Electron.*, **22**, 181 (1979).
50. Simmons J. G., El-Badry A. Theory of Switching Phenomena In Metal/Semi-Insulator/ $n-p^+$ Silicon Devices, *Solid State Electron.*, **20**, 955 (1977); A. El-

- Badry and Simmons J. G., Experimental Studies of Switching in Metal Semi-Insulating np^+ Silicon Devices, *Solid State Electron.*, **20**, 963 (1977).
51. Simmons J. G. Generalized Formula for the Electric Effect between Similar Electrodes Separated by a Thin Insulating Film. *J. Appl. Phys.*, **34**, 1793 (1963).
 52. Pollack S. R., Morris C. E. Tunneling through Gaseous Oxidized Films of Al_2O_3 , *Trans. AIME*, **233**, 497 (1965).
 53. Hirych Z. Influence of Nonuniform Thickness of Dielectric Layers on Capacitance and Tunnel Current, *Solid State Electron.*, **9**, 967 (1966).
 54. Pochobradsky J. On the Capacitance of Metal — Insulator — Metal Structures with Nonuniform Thickness, *Solid State Electron.*, **10**, 973 (1967).
 55. Ku H. Y., Ullman F. G. Capacitance of Thin Dielectric Structures, *J. Appl. Phys.* **35**, 265 (1964).
 56. Kurtin S., McGill T. C., Mead C. A. Tunneling Currents and $E - k$ Relation, *Phys. Rev. Lett.*, **25**, 756 (1970); Direct Interelectrode Tunneling in GaAs., *Phys. Rev.* **B3**, 3368 (1971).
 57. Heiblum M., Wang S. Y., Whinnery J. R., Gustafson T. K. Characteristics of Integrated MOM Junction at DC and at Optical Frequencies, *IEEE J. Quantum Electron.*, **QE-14**, 159 (1978).
 58. Mead C. A. Tunnel-Emission Amplifiers, *Proc. IRE*, **48**, 359 (1960).
 59. Crowell C. R., Sze S. M. Hot Electron Transport and Electron Tunneling in Thin-Film Structures, in Haas G., Thun R. E., Eds. Physics of Thin Films, Vol. 4, Academic, N. Y., 1967.
 60. Spratt J. P., Schwartz R. F., Kane W. M. Hot Electrons in Metal Films: Injection and Collection, *Phys. Rev., Lett.*, **6**, 341 (1961).
 61. Kasaki H. Tunnel Transistor, *Proc IEEE*, **61**, 1053 (1973).
 62. Chang L. L., Esaki L. Tunnel Triode — A Tunneling Base Transistor, *Appl. Phys. Lett.*, **31**, 687 (1977).
 63. Hewlett-Packard Electronic Test Instruments, Hewlett-Packard, Palo Alto, Calif., 1961.
 64. Liao S. Y. Microwave Devices and Circuits, Prentice-Hall, Englewood Cliffs, N. J., 1980.

Глава 10

ЛАВИННО-ПРОЛЕТНЫЕ ДИОДЫ И ДРУГИЕ ПРИБОРЫ НА ПРОЛЕТНЫХ ЭФФЕКТАХ

10.1. ВВЕДЕНИЕ

Принцип работы лавинно-пролетных диодов¹ (ЛПД) основан на возникновении отрицательного сопротивления в диапазоне сверхвысоких частот, которое обусловлено процессами лавинного умножения носителей и их пролета через полупроводниковую структуру. Появление отрицательного сопротивления связано с временным запаздыванием этих двух процессов, приводящим к фазовому сдвигу между током и напряжением. «Лавинное запаздывание» появляется за счет конечного времени нарастания лавинного тока, а «пролетное запаздывание» — за счет конечного времени прохождения носителями области дрейфа. Сопротивление диода отрицательно на некоторой частоте, если сумма этих времен равна полупериоду колебания. Впервые экспериментальное наблюдение осцилляций было осуществлено² в 1965 г. Джоном, Делушем и Коэнном в смещенном в область обратного лавинного пробоя кремниевом диоде, который был помещен в СВЧ-резонатор [1, 2].

Возникающее за счет пролетных эффектов в полупроводниковых диодах отрицательное сопротивление впервые изучил в 1954 г. Шокли, который считал, что двухконтактные приборы из-за структурной простоты обладают потенциальными преимуществами по сравнению с трехконтактными транзисторами [3]. В 1958 г. Рид предложил конструкцию высокочастотного полупроводникового диода, в котором область лавинного умножения расположена на одном конце относительно высокоомного слоя, служащего дрейфовым пролетным пространством для генерируемых носителей заряда (т. е. $p^+ - n - i - n^+$ или $n^+ - p - i - p^+$) [4]. Осцилляции в диоде Рида впервые наблюдали Ли и др. [5]. Результаты малосигнальной теории, развитой Мисавой [6], Джилдеином и Хайнсом [7], подтвердили, что отрицательное сопротивление, обусловленное лавинно-пролетным эффектом, может возникать

¹ В английской литературе эти приборы сокращенно называются IMPATT (impact ionization avalanche transit time) — *Прим. перев.*

² Лавинно-пролетный диод был впервые создан в СССР на основе обнаруженного в 1959 г. эффекта генерации когерентных СВЧ-колебаний при лавинном пробое германиевых дифференциальных диодов (Тагер А. С., Мельников А. И., Кобельков Г. П., Цибнев А. М. Генерация и усиление радиоволн сантиметрового и миллиметрового диапазонов с помощью полупроводникового диода в области положительного наклона его статической вольт-амперной характеристики. Диплом на открытие № 24, приоритет 27.X.1959 г.). — *Прим. ред.*

в обычном диоде с $p-n$ -переходом или контакте металл—полупроводник с произвольным профилем распределения примеси.

В настоящее время лавинно-пролетный диод является одним из самых мощных твердотельных источников СВЧ-излучения [8, 9]. ЛПД могут генерировать в непрерывном режиме наибольшую мощность в диапазоне частот, соответствующих миллиметровым длинам волн (т. е. свыше 30 ГГц) [10, 10а]. Однако следует сказать о трудностях, с которыми сталкиваются при работе ЛПД во внешней цепи: 1) высокий уровень шума; 2) необходимость тщательного расчета цепей (чтобы избежать расстройки или даже перегорания диода, поскольку реактивность большая и сильно зависит от амплитуды осцилляций) [11].

Другим прибором на пролетном эффекте является инжекционно-пролетный диод¹ (ИПД), название которого указывает на механизмы, лежащие в основе его работы. Колмен и Зи в 1971 г. сообщили о работе структуры металл—полупроводник—металл в режиме с проколом в качестве ИПД [12]. Аналогичные структуры были предложены в 1968 г. Ругом [13] (им был проведен анализ работы в режиме большого сигнала) и Райтом [14], который рассмотрел механизм переноса, ограниченного объемным зарядом. Поскольку в ИПД нет запаздывания, связанного с конечным временем нарастания лавины, предполагается, что эти приборы будут генерировать меньшие мощности при более низких к. п. д. по сравнению с лавинно-пролетными диодами. С другой стороны, шум, связанный с инжекцией носителей заряда через барьер, меньше лавинного шума ЛПД. Низкий уровень шума и хорошая стабильность позволяют применять ИПД в качестве гетеродина. Инжекционно-пролетные диоды обладают большей мощностью и меньшей величиной минимального обнаружимого сигнала, чем ЛПД или приборы на эффекте междолиního перехода электронов при их использовании в качестве доплеровских детекторов [15, 16]. Родственным инжекционно-пролетному диоду является двухскоростной пролетный диод² (ДСПД), в котором скорость носителей у одного контакта значительно меньше, чем у другого. До сих пор не было сообщений об экспериментальном наблюдении СВЧ-генерации в ДСПД. Тем не менее, согласно теоретическим оценкам, эти приборы будут иметь более высокие к. п. д. по сравнению с инжекционно-пролетными диодами, поскольку в ДСПД есть дополнительное запаздывание на время пролета области, в которой скорость носителей относительно низка.

¹ В английской литературе принято сокращение BARITT (barrier injection transit time). — *Прим. перев.*

² В английской литературе используется сокращение DOVETT (double velocity transit time). — *Прим. перев.*

Родственным ЛПД прибором является также пролетный диод с захваченным объемным зарядом лавины¹. Впервые такой режим работы был обнаружен Прагером, Чангом и Вайсбродом в 1967 г. в кремниевых лавинных диодах [17]. Рабочая частота этого прибора намного меньше пролетной, а к. п. д. значительно выше, чем у ЛПД. В результате теоретических исследований было установлено, что при работе в режиме большого сигнала лавинный процесс начинается в области высокого поля, а затем быстро распространяется на весь образец, в результате чего последний заполняется высокопроводящей электронно-дырочной плазмой, пространственный заряд которой снижает напряжение на диоде до очень малых величин [18—20]. Так как плазма не может быть быстро выведена из прибора, этот режим работы назван режимом с захваченным объемным зарядом лавины. Пролетные диоды с захваченным объемным зарядом лавины применяются в импульсных передатчиках и в радарях с фазными решетками [15, 21].

10.2. СТАТИЧЕСКИЕ ХАРАКТЕРИСТИКИ

Основными представителями семейства лавинно-пролетных диодов являются диод Рида, асимметричный резкий $p-n$ -переход, симметричный $p-n$ -переход (диод с двумя дрейфовыми областями), диод с двухслойной базой, диод с трехслойной базой (модифицированный диод Рида) и $p-i-n$ -диод. Ниже рассмотрены статические характеристики этих диодов (распределение электрического поля, напряжение пробоя и эффекты пространственного заряда).

На рис. 1 приведены распределения примесей, электрического поля и эффективного коэффициента ударной ионизации в идеализированном диоде Рида ($p^+-n-i-n^+$ или $n^+-p-i-p^+$ -диоде) при пробое. Эффективный коэффициент ударной ионизации равен

$$\langle \alpha \rangle \equiv \alpha_n \exp \left[- \int_x^W (\alpha_n - \alpha_p) dx' \right], \quad (1)$$

где α_n и α_p — коэффициенты ударной ионизации электронов и дырок соответственно и W — ширина обедненной области.

Как показано в гл. 2, условие лавинного пробоя можно записать в виде

$$\int_0^W \langle \alpha \rangle dx = 1. \quad (2)$$

¹ В английской литературе этот прибор сокращенно называется TRAPATT (trapped plasma avalanche triggered transit). — *Прим. перев.*

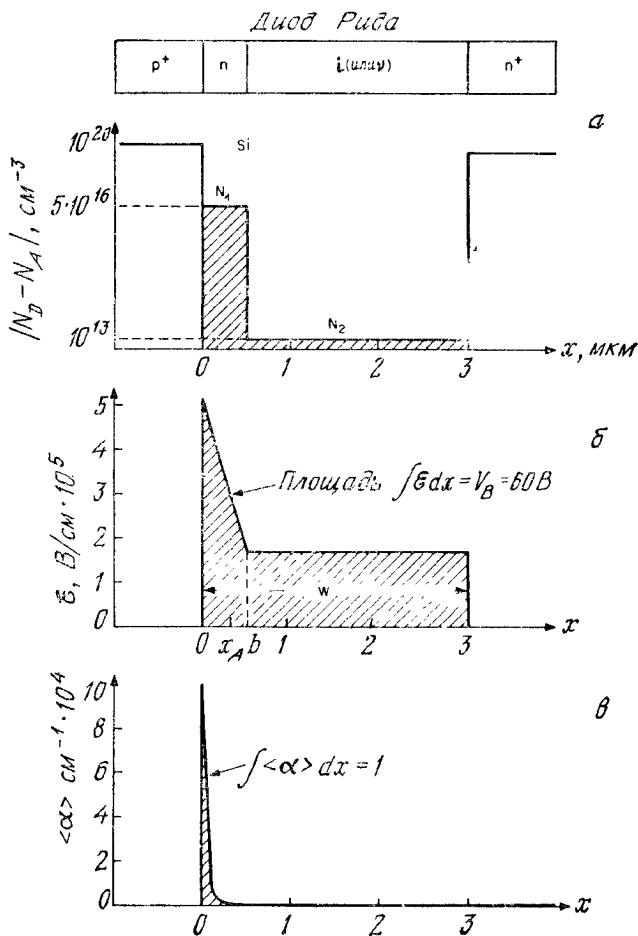


Рис. 1. Распределения примесей (а), напряженности электрического поля (б) и эффективного коэффициента ионизации (в) при пробое в $p^+ - n - i - n^+$ -диоде Рида [4].

Из-за сильной зависимости коэффициентов ударной ионизации от напряженности электрического поля область лавинного умножения сильно локализована, т. е. практически весь процесс умножения носителей происходит в узком слое высокого поля от 0 до x_A , где x_A — ширина области умножения. Падение напряжения на области умножения обозначим V_A . Ниже мы покажем, что оптимальная плотность тока и максимальный к. п. д. лавинно-пролетных диодов сильно зависят как от x_A , так и от V_A . Слой вне области умножения ($x_A \leq x \leq W$) называется областью дрейфа.

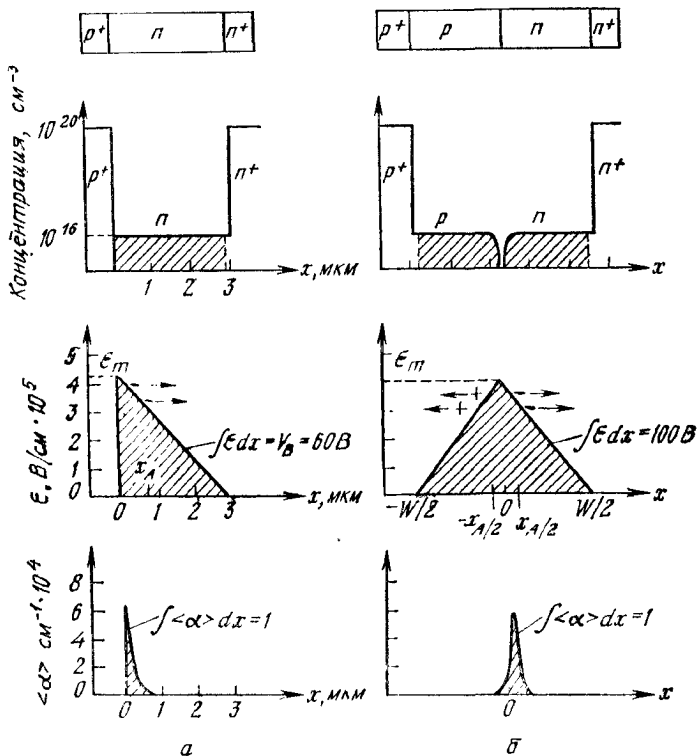


Рис. 2. Распределения примесей, напряженности электрического поля и эффективного коэффициента ионизации в диоде с асимметричным резким $p^+ - n$ -переходом (диод с одной дрейфовой областью) (а) и в $p^+ - p - n - n^+$ -диоде с симметричным резким $p - n$ -переходом (диод с двумя дрейфовыми областями) (б).

Существуют два предельных случая распределения примеси в диодах Рида. Если отсутствует область с концентрацией примеси N_2 , то мы имеем резкий $p^+ - n$ -переход. Если же отсутствует область с концентрацией N_1 , то диод Рида вырождается в $p - i - n$ -диод, который иногда называют диодом Мисавы [6]. На рис. 2, а приведены статистические характеристики диода с асимметричным резким $p - n$ -переходом. Область лавинного умножения сильно локализована. На рис. 2, б приведены характеристики кремниевого диода с симметричным резким $p - n$ -переходом. Область умножения расположена вблизи центра обедненной области. Небольшая асимметрия эффективного коэффициента ударной ионизации по отношению к точке, в которой напряженность электрического поля максимальна, связана с тем, что в кремнии α_n и α_p

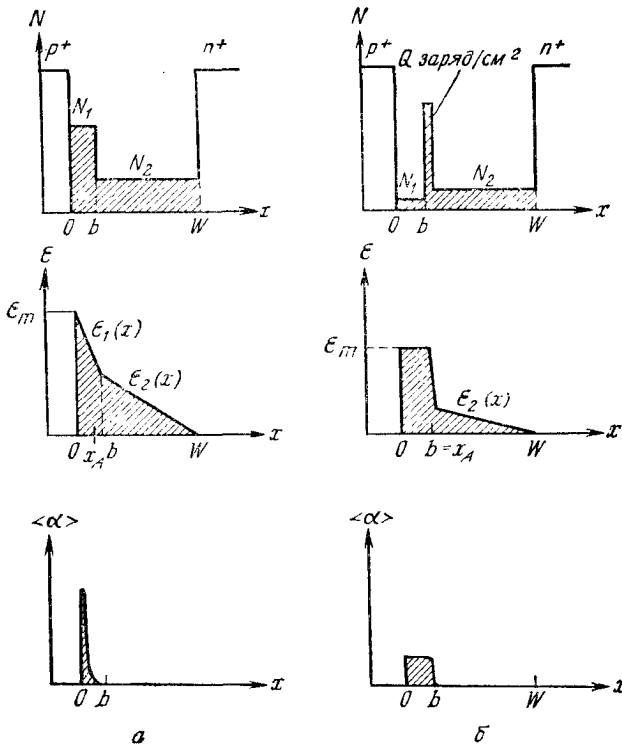


Рис. 3. Распределения примесей, напряженности электрического поля и эффективного коэффициента ионизации в модифицированных диодах Рида: с двухслойной базой (а) и с трехслойной базой (б).

сильно различаются. Если $\alpha_n \simeq \alpha_p$, как это имеет место в GaP, эффективный коэффициент ионизации

$$\langle \alpha \rangle = \alpha_n = \alpha_p, \quad (3)$$

и область лавинного умножения симметрична по отношению к точке $x = 0$.

На рис. 3, а приведены статические характеристики модифицированной структуры Рида — диода с двухслойной базой, в котором концентрация N_2 существенно превышает соответствующую величину для диода Рида [22]. Статические характеристики другой модификации — диода с трехслойной базой приведены на рис. 3, б. В этом диоде «сгусток» заряда Q расположен в точке $x = b$. Поскольку достаточно высокая напряженность электрического поля остается постоянной между точками $x = 0$ и $x = b$, то максимальное поле может быть меньше, чем в диоде

в двухслойной базе. В $p-i-n$ -диоде электрическое поле однородно во всей i -области при малых токах. В этом случае область лавинного умножения соответствует всей ширине слоя собственной концентрации.

10.2.1. Напряжение пробоя

Напряжение пробоя для асимметричных резких $p-n$ -переходов было рассмотрено в гл. 2. Мы можем воспользоваться развитым в этой главе методом для того, чтобы найти напряжения пробоя для других типов диодов. Для симметричных резких переходов (рис. 2, б) напряжение пробоя определяется равенством

$$V_B = \frac{1}{2} \mathcal{E}_m W = \frac{\epsilon_s \mathcal{E}_m^2}{qN_B}, \quad (4)$$

где \mathcal{E}_m — величина максимальной напряженности электрического поля при $x = 0$. На рис. 4 приведены максимальные напряженности электрического поля при пробое в симметричных и асимметричных $p-n$ -переходах из кремния и арсенида галлия с ориентацией $\langle 100 \rangle$ [23, 24]. Если известна концентрация примеси, то напряжение пробоя можно найти из равенства (4), пользуясь графиками, приведенными на рис. 4. Обратное напряжение при пробое равно $(V_B - V_{bi})$, где V_{bi} — контактная разность потенциалов, причем $V_{bi} = (2kT/q) \ln(N_B/n_i)$ в случае симметричных резких переходов (в работающих ЛПД величиной V_{bi} можно пренебрегать).

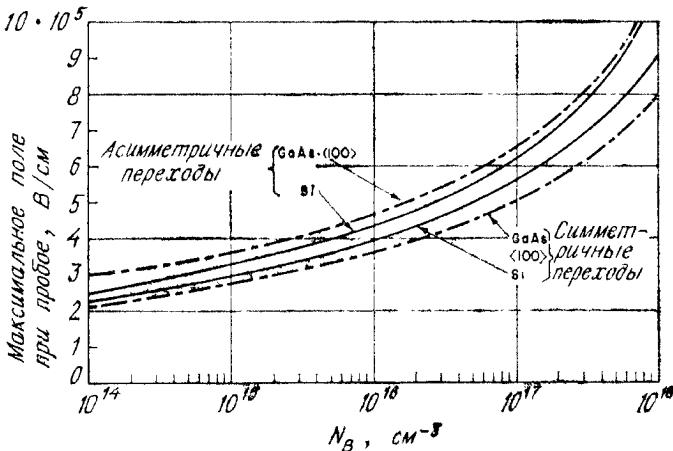


Рис. 4. Максимальная напряженность электрического поля при пробое как функция концентрации для асимметричных и симметричных резких переходов из Si и GaAs [23, 24].

Напряжение пробоя и ширина обедненной области в диоде Рида и диоде с двухслойной базой определяются следующими равенствами:

$$V_B = \left(\mathcal{E}_m - \frac{qN_1 b}{2\epsilon_s} \right) b - \frac{1}{2} \left(\mathcal{E}_m - \frac{qN_1 b}{\epsilon_s} \right) (W - b), \quad (5)$$

$$W = \frac{\epsilon_s \mathcal{E}_m}{qN_2} - b \left(\frac{N_1}{N_2} - 1 \right). \quad (6)$$

Ширина обедненной области в диоде Рида ограничена толщиной эпитаксиального слоя, которая меньше величины, вычисленной с помощью равенства (6). Максимальная напряженность электрического поля при пробое в диоде Рида или диоде с двухслойной базой с точностью до 1 % совпадает с соответствующей величиной для асимметричного резкого $p^+ - n$ -перехода с такой же концентрацией примеси N_1 при условии, что ширина области лавинного умножения $x_A < b$ [25]. Следовательно, напряжение пробоя можно найти из равенства (5) и (6), используя величину максимальной напряженности электрического поля из графиков, приведенных на рис. 4.

В диоде с трехслойной базой при очень узком слое объемного заряда с высокой концентрацией примеси напряжение пробоя и полная ширина обедненной области определяются выражениями

$$V_B = \mathcal{E}_m b + \frac{1}{2} \left(\mathcal{E}_m - \frac{qQ}{\epsilon_s} \right) (W - b), \quad (7)$$

$$W = \frac{\epsilon_s}{qN_2} \left(\mathcal{E}_m - \frac{qQ}{\epsilon_s} \right) + b, \quad (8)$$

где Q — число атомов примеси на 1 см^2 в слое объемного заряда. Поскольку напряженность электрического поля приблизительно постоянна при $0 \leq x \leq b$, эффективный коэффициент ударной ионизации $\langle \alpha \rangle = 1/b$. Максимальную напряженность электрического поля можно найти, если известна зависимость коэффициента ионизации от поля.

10.2.2. Область лавинного умножения и область дрейфа

В идеальном $p-i-n$ -диоде область лавинного умножения занимает всю ширину слоя собственной концентрации. Однако в диодах Рида и $p-n$ -переходах область умножения носителей заряда очень узкая и расположена вблизи металлургической границы перехода. Вклад в интеграл в выражении (2) быстро уменьшается по мере удаления x от металлургической границы. Поэтому имеет смысл определить ширину области умножения x_A таким

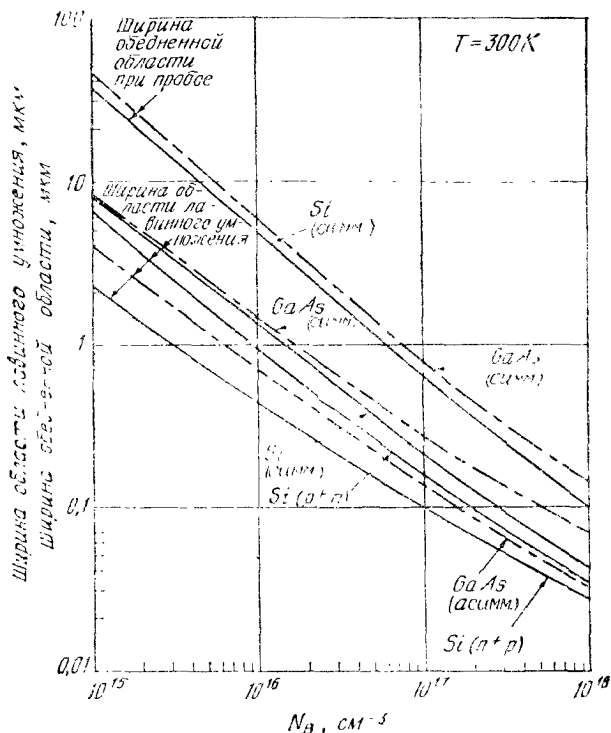


Рис. 5. Ширина области лавинного умножения в переходах из кремния и арсенида галлия и ширина области обеднения в зависимости от концентрации примеси в симметричных переходах из Si и GaAs [24].

образом, чтобы при интегрировании эффективного коэффициента ионизации от 0 до x_A получить 95 % полной величины:

$$\int_0^{x_A} \langle \alpha \rangle dx = 0,95, \quad (9)$$

а для диода Рида, диода с двухслойной базой, асимметричного резкого p^+n -перехода и симметричных $p-n$ -переходов при пробое

$$\int_{-x_A/2}^{x_A/2} \langle \alpha \rangle dx = 0,95. \quad (10)$$

На рис. 5 приведены зависимости ширины области лавинного умножения от концентрации примеси в диодах из кремния и арсенида галлия [24], а также зависимости ширины обедненной области

в симметричных переходах. Из-за того что в кремнии коэффициенты ионизации различны ($\alpha_n > \alpha_p$), при одинаковой концентрации примеси ширина области умножения в $n^+ - p$ -переходе меньше, чем в $p^+ - n$ -переходе. В диоде Рида или в диоде с двухслойной базой ширина области умножения такая же, как и в асимметричном резком $p^+ - n$ -переходе, если концентрация примеси N_1 одинакова. В диоде с трехслойной базой область умножения занимает пространство от металлургической границы до слоя заряда высокой концентрации при $x_A = b$.

Область дрейфа совпадает с обедненной областью, за исключением слоя лавинного умножения носителей, или $x_A < x < W$. Важнейшей характеристикой области дрейфа является скорость, с которой носители заряда пролетают через эту область. Время пролета области дрейфа минимально при достаточно большой напряженности электрического поля, когда носители движутся со скоростью насыщения v_s . В кремнии напряженность электрического поля должна превышать 10^4 В/см. В арсениде галлия поле может быть значительно меньше ($\sim 10^3$ В/см).

В $p - i - n$ -диодах это требование выполняется автоматически, так как электрическое поле при пробое приблизительно одинаково на всей длине области собственной концентрации и существенно превышает поле, необходимое для насыщения скорости. В диоде Рида минимальная напряженность электрического поля в области дрейфа равна $\mathcal{E}_{\text{мин}} = \mathcal{E}_m - q [N_1 b + N_2 (W - b)] / \epsilon_s$.

Из сказанного выше следует, что диод Рида нужно рассчитать так, чтобы $\mathcal{E}_{\text{мин}}$ было достаточно велико. В резких $p - n$ -переходах всегда существуют области, в которых напряженность электрического поля меньше необходимой для начала пробоя. Однако эти области занимают лишь небольшую часть всего обедненного слоя. Например, в кремниевом $p^+ - n$ -переходе с концентрацией примеси 10^{16} см $^{-3}$ максимальное поле при пробое равно 4×10^5 В/см. Отношение ширины области, в которой напряженность электрического поля относительно низка (меньше 10^4 В/см), к полной ширине обедненной области составляет $10^4 / 4 \cdot 10^5 = 2,5$ %. В $p^+ - n$ -переходе из арсенида галлия при такой же концентрации примеси ширина области слабого поля составляет менее 0,2 % ширины обедненной области. Поэтому существование области слабого поля практически не сказывается на времени пролета носителей через обедненную область.

10.2.3. Влияние температуры и объемного заряда носителей

Напряжение пробоя и максимальные напряженности электрического поля, о которых речь шла выше, вычисляются при комнатных температурах без учета эффектов объемного заряда носителей, разогрева и в отсутствие генерации. Однако в рабочих

условиях ЛПД смещен в область лавинного пробоя, и обычно плотность тока очень велика. Это приводит к заметному увеличению температуры в переходе и появлению эффектов, связанных с объемным зарядом носителей.

Коэффициенты ионизации электронов и дырок уменьшаются при увеличении температуры [26]. Поэтому при заданной концентрации примеси напряжение пробоя ЛПД будет возрастать по мере того, как температура повышается. С ростом потребляемой на постоянном токе мощности (произведение напряжения на ток при обратном смещении) возрастают как температура в переходе, так и напряжение пробоя. В конце концов диод выходит из строя в основном из-за сильного разогрева, вызывающего локальные разрушения. Поэтому увеличение температуры в $p-n$ -переходе накладывает сильные ограничения на режим работы диода. Для того чтобы предотвратить увеличение температуры в приборе, необходим хороший теплоотвод. Этот вопрос мы обсудим в разд. 10.4.

Объемный заряд генерируемых носителей приводит к изменению напряженности электрического поля в обедненной области. Этот эффект вызывает появление дополнительного статического сопротивления, которое положительно для резких переходов и отрицательно для $p-i-n$ -диодов [27].

Сначала рассмотрим p^+-n-n^+ -диод с асимметричным резким переходом (рис. 6, а). Когда приложенное напряжение V равно напряжению пробоя V_B , абсолютная величина напряженности электрического поля $|\mathcal{E}_x|$ максимальна в точке $x = 0$ и равна \mathcal{E}_m . Если предположить, что электроны движутся в обедненной области со скоростью насыщения v_s , то ток I равен

$$I = v_s \rho A, \quad (11)$$

где ρ — плотность заряда носителей и A — площадь. Изменение напряженности электрического поля $\Delta \mathcal{E}(x)$, связанное с объемным зарядом носителей, можно найти из равенства (11) и уравнения Пуассона

$$\Delta \mathcal{E}(x) \approx \frac{Ix}{A \epsilon_s v_s}. \quad (12)$$

Если предположить, что все носители генерируются в области умножения шириной x_A , то изменение напряжения, связанное с объемным зарядом в области дрейфа $W - x_A$, можно найти, проинтегрировав $\Delta \mathcal{E}(x)$:

$$\Delta V_B \approx \int_0^{W-x_A} \frac{Ix}{A \epsilon_s v_s} dx = I \frac{(W-x_A)^2}{2A \epsilon_s v_s}. \quad (13)$$

Полная величина приложенного напряжения равна

$$V = V_B + \Delta V_B = V_B + IR_{SC}, \quad (14)$$

где R_{SG} — сопротивление области пространственного заряда [28], которое определяется из выражений (13) и (14):

$$R_{SG} \cong \frac{\Delta V_B}{I} \cong \frac{(W - x_A)^2}{2A\epsilon_s v_s}. \quad (15)$$

Для значений параметров, приведенных на рис. 6, а, сопротивление области пространственного заряда приблизительно равно 20 Ом при $A = 10^{-4} \text{ см}^2$.

В $p-i-n$ - или $p-v-n$ -диодах ситуация отличается от случая p^+-n -перехода. Когда приложенное напряжение достаточно для начала лавинного пробоя, обратный ток мал. Влиянием объемного

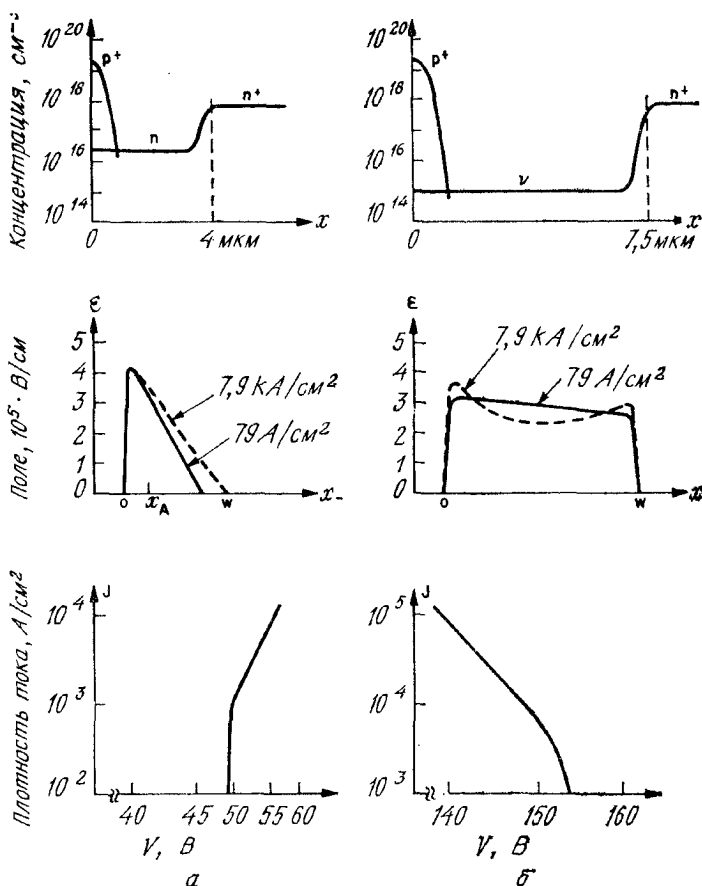


Рис. 6. Распределения примесей электрического поля и вольт-амперные характеристики p^+-n-n^+ -диода (а) и p^+-v-n^+ -диода (б) [27].

заряда можно пренебречь, а поле считать практически однородным в обедненной области. По мере увеличения тока все большее количество электронов генерируется за счет ударной ионизации вблизи $p-i$ -перехода, а дырок — вблизи $n-i$ перехода (так как поле имеет два максимума, см. рис. 6, б). Возникающий пространственный заряд снижает электрическое поле в центре v -области.

Поэтому уменьшается и напряжение, которое равно $\int_0^w \mathcal{E} dx$. Это приводит к уменьшению сопротивления $p-v-n$ -диода по постоянному току (рис. 6, б).

10.3. ДИНАМИЧЕСКИЕ ХАРАКТЕРИСТИКИ

10.3.1. Запаздывание, связанное с инжекцией.

Пролетный эффект

Сначала рассмотрим запаздывание фазы, связанное с инжекцией, и пролетный эффект в идеализированном приборе [29]. Предположим, что импульс тока проводимости инжектируется при $x = 0$ (рис. 7) с фазовым углом φ по отношению к полному току, а приложенное к диоду постоянное смещение таково, что инжектированные носители заряда пролетают область дрейфа со скоростью насыщения v_s . Плотность переменного тока проводимости \tilde{J}_c при $x = 0$ равна полной плотности тока \tilde{J} со сдвигом по фазе:

$$\tilde{J}_c(x=0) = \tilde{J} \exp(-j\varphi). \quad (16)$$

В произвольной точке области дрейфа полный переменный ток равен сумме тока проводимости и тока смещения:

$$\tilde{J}(x) = \tilde{J}_c(x) + \tilde{J}_d(x) = \tilde{J}_c(x=0) e^{-j\omega x/v_s} + j\omega \varepsilon_s \tilde{\mathcal{E}}(x), \quad (17)$$

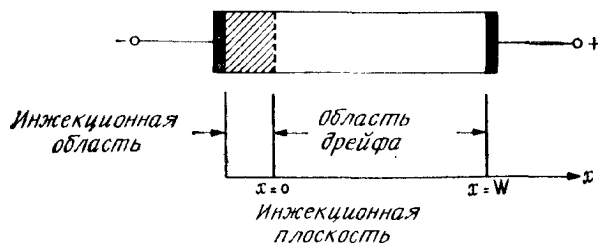


Рис. 7. Идеализированный диод, в котором носители инжектируются при $x = 0$ и движутся в области дрейфа со скоростью насыщения.

где $\tilde{\mathcal{E}}(x)$ — переменная компонента напряженности электрического поля. Из выражений (16) и (17) получаем

$$\tilde{\mathcal{E}}(x) = \frac{\tilde{J} (1 - e^{-j\omega x/v_s - j\varphi})}{j\omega \varepsilon_s}. \quad (18)$$

Проинтегрировав выражение (18), найдем импеданс диода

$$Z \equiv \frac{\int_0^W \tilde{\mathcal{E}}(x) dx}{\tilde{J}} = \frac{1}{j\omega C} \left[1 - \frac{e^{-j\varphi} (1 - e^{-j\theta})}{j\theta} \right], \quad (19)$$

где $C = \varepsilon_s/W$ — емкость на единицу площади и $\theta = \omega W/v_s$ — пролетный угол. Выделив действительную и мнимую части выражения (19), получим

$$R = \frac{\cos \varphi - \cos(\varphi + \theta)}{\omega C \theta}, \quad (20)$$

$$X = -\frac{1}{\omega C} + \frac{\sin(\theta + \varphi) - \sin \varphi}{\omega C \theta}. \quad (21)$$

Ниже мы изучим влияние инжекционного фазового угла φ на сопротивление по переменному току R . Когда угол φ равен нулю (нет фазового запаздывания), сопротивление пропорционально величине $(1 - \cos \theta)/\theta$, которая всегда больше или равна нулю (рис. 8, а), и поэтому сопротивление положительно. Следовательно, только пролетный эффект не может приводить к возникновению отрицательного сопротивления. Однако если $\varphi \neq 0$, то сопротивление отрицательно при некоторых значениях пролетных углов. Например, при $\varphi = \pi/2$ наибольшее отрицательное сопротивление соответствует углу пролета $\theta \approx 3\pi/2$ (рис. 8, б). На этом принципе основана работа инжекционно-пролетного диода: инжекция несущих носителей через барьер вносит фазовое запаздывание 90° , а пролетный угол, равный 270° , оптимизирует характеристику прибора. Детальный анализ работы ИПД проведен в разд. 10.7. Если $\varphi = \pi$, сопротивление максимально при $\theta = \pi$ (рис. 8, в). Эта ситуация соответствует лавинно-пролетному режиму, в котором фазовый сдвиг 180° достигается за счет конечного времени нарастания лавинного тока, а пролетный эффект приводит к добавочному запаздыванию на 180° .

Изложенное выше подтверждает важность инжекционного запаздывания. Проблема создания прибора, использующего при работе пролетные эффекты, сводится, таким образом, к поиску способа, с помощью которого можно вызвать запаздывание инжекции носителей в область дрейфа. Из рис. 8 видно, что сумма инжекционной фазы и оптимального угла пролета $\varphi + \theta_{\text{opt}}$ приблизительно равна 2π , а отрицательное сопротивление рас-

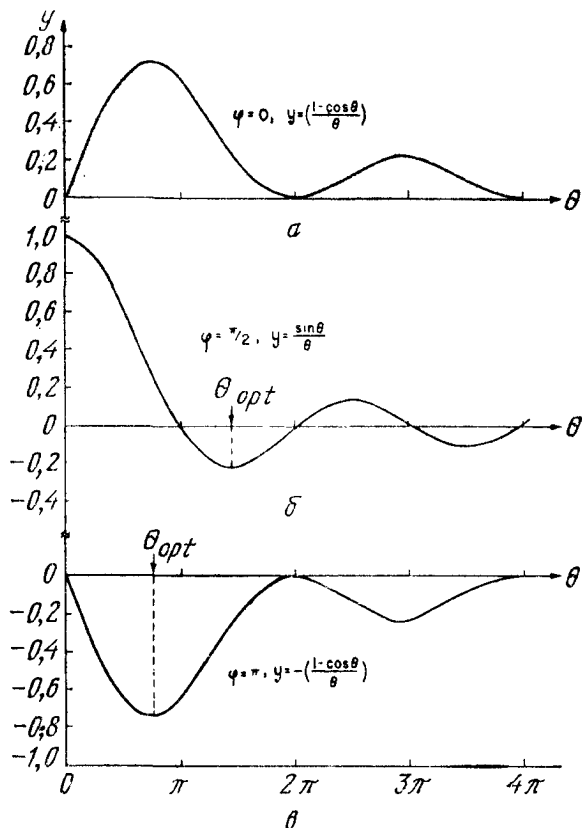


Рис. 8. Зависимость сопротивления от пролетного угла при трех значениях инжекционного запаздывания фазы.

тет по мере того, как φ становится все более положительным. Из-за более высокого сопротивления лавинно-пролетные диоды имеют лучшие к. п. д. и генерируют большие мощности, чем инжекционно-пролетные диоды.

10.3.2. Малосигнальный анализ

Анализ малосигнальных процессов впервые проведен Ридом [4] и затем развит Джилденом и Хайнсом [7]. Для простоты предположим, что коэффициенты ионизации $\alpha_n = \alpha_p = \alpha$ и скорости насыщения электронов и дырок равны. На рис. 9, а приведена схема диода Рида. Воспользовавшись соображениями, изложенными в разд. 10.2, разобьем диод на три области: 1) область

лавинного умножения, которая предполагается настолько тонкой, что в ней можно не учитывать объемный заряд и запаздывание сигнала; 2) область дрейфа, в которой отсутствует генерация носителей, а имеющиеся носители движутся со скоростями насыщения; 3) пассивную область, которая вносит нежелательное паразитное сопротивление.

Две активные области взаимодействуют друг с другом, так как переменное электрическое поле непрерывно на их границе. Будем использовать нижний индекс «0» для обозначения постоянных величин, а тильду (\sim) для малосигнальных переменных. Величины, содержащие постоянную и переменную составляющие, обозначим буквами без индексов. Плотность лавинного тока в области умножения \tilde{J}_A равна плотности переменного тока проводимости в этой области, а полная плотность переменного тока равна \tilde{J} . Поскольку мы предположили, что область умножения узка, то ток \tilde{J}_A достигает границы области дрейфа без запаздывания. Если носители движутся со скоростью насыщения v_s , плотность переменного тока проводимости $\tilde{J}_c(x)$ в области дрейфа представляет собой незатухающую волну (в которой изменяется лишь фаза), распространяющуюся со скоростью насыщения:

$$\tilde{J}_c(x) = \tilde{J}_A e^{-j\omega x/v_s} \equiv \gamma \tilde{J} e^{-j\omega x/v_s}, \quad (22)$$

где $\gamma \equiv \tilde{J}_A/\tilde{J}$ — комплексная величина, равная отношению плотности лавинного тока к плотности полного тока.

В любом поперечном сечении плотность полного тока \tilde{J} равна сумме плотностей тока проводимости \tilde{J}_c и тока смещения \tilde{J}_d . Эта сумма не зависит от x :

$$\tilde{J} = \tilde{J}_c(x) + \tilde{J}_d(x) \neq f(x). \quad (23)$$

Плотность тока смещения связана с напряженностью переменного электрического поля $\tilde{\mathcal{E}}(x)$ следующим соотношением:

$$\tilde{J}_d = j\omega \epsilon_s \tilde{\mathcal{E}}(x). \quad (24)$$

Из выражений (22)—(24) можно получить выражение для переменной компоненты напряженности электрического поля в области дрейфа как функции x и \tilde{J} :

$$\tilde{\mathcal{E}}(x) = \tilde{J} \frac{1 - \gamma e^{-j\omega x/v_s}}{j\omega \epsilon_s}. \quad (25)$$

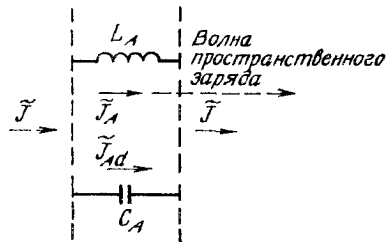
В результате интегрирования $\tilde{\mathcal{E}}(x)$ получим зависимость падения напряжения на этой области. Выражение для коэффициента γ мы получим ниже.

Область лавинного умножения. Сначала рассмотрим область умножения. Плотность постоянного тока $J_0 = J_{p0} + J_{n0}$ связана с плотностью обратного тока насыщения, обусловленного термической генерацией $J_s = J_{ns} + J_{ps}$, соотношением

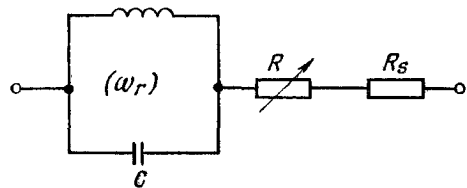
$$\frac{J_s}{J_0} = 1 - \int_0^W (\alpha) dx. \quad (26)$$



а



б



в

Рис. 9. Модель диода Рида с областью лавинного умножения, областью дрейфа (б), эквивалентная схема диода Рида при малых пролетных углах (а) и за

При пробое J_0 стремится к бесконечности, а $\int_0^W \langle \alpha \rangle dx = 1$.

В стационарном случае интеграл от эффективного коэффициента ионизации не превышает 1. Это не обязательно имеет место в случае быстро изменяющегося поля. Ниже мы получим дифференциальное уравнение для тока, зависящего от времени. Если предположить, что коэффициенты ионизации и скорости насыщения электронов и дырок равны и дрейфовые токи существенно превышают диффузионные, то основные уравнения, описывающие работу прибора, в одномерном приближении имеют вид

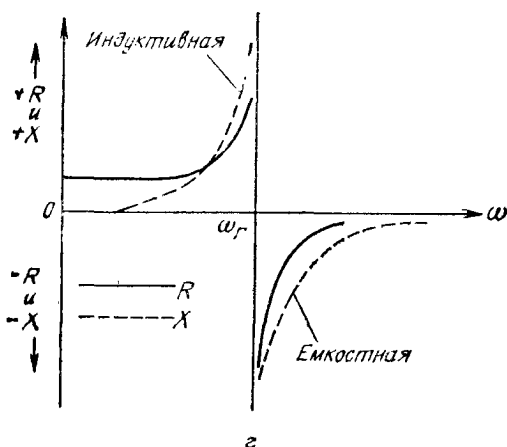
$$\frac{\partial \mathcal{E}}{\partial x} = \frac{q}{\varepsilon_s} (N_D^+ - N_A^- + p - n) \quad (\text{уравнение Пуассона}), \quad (27)$$

$$\left. \begin{aligned} J_n &= qv_s n, \\ J_p &= qv_s p, \\ J &= J_n + J_p \end{aligned} \right\} \quad (\text{равенства для плотности тока}), \quad (28)$$

$$\frac{\partial n}{\partial t} = \frac{1}{q} \frac{\partial J_n}{\partial x} + \alpha v_s (n + p), \quad (29a)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q} \frac{\partial J_p}{\partial x} + \alpha v_s (n + p) \quad (\text{уравнения непрерывности}). \quad (29b)$$

Вторые слагаемые в правых частях выражений (29a) и (29b) равны скорости генерации электронно-дырочных пар за счет лавинного умножения. Эта величина значительно превышает скорость термической генерации, которой можно пренебречь. Про-



■ пассивной областью (а), эквивалентная схема для области лавинного умножения активных и реактивных составляющих импеданса от частоты (б) [7].

суммировав выражения (29а) и (29б) с учетом выражения (28) и проинтегрировав от $x = 0$ до $x = x_A$, получим

$$\tau_A \frac{dJ}{dt} = -(J_p - J_n)_{0^+}^{x_A} + 2J \int_0^{x_A} \alpha dx, \quad (30)$$

где $\tau_A = x_A/v_s$ — время пролета носителями области умножения.

Граничным условием служит равенство тока электронов при $x = 0$ обратному току насыщения J_{ns} . Поэтому в плоскости $x = 0$ $J_p - J_n = -2J_n + J = -2J_{ns} + J$. В плоскости $x = x_A$ дырочный ток равен обратному току насыщения J_{ps} генерируемым в области пространственного заряда дырок. Следовательно, $J_p - J_n = 2J_p - J = 2J_{ps} - J$. С учетом этих граничных условий уравнение (30) принимает следующий вид:

$$\frac{dJ}{dt} = \frac{2J}{\tau_A} \left[\int_0^{x_A} \alpha dx - 1 \right] + \frac{2J_s}{\tau_A}. \quad (31)$$

В стационарных условиях J совпадает с постоянным током J_0 , и уравнение (31) переходит в равенство (26).

Упростим уравнение (31), подставив вместо α средний коэффициент ионизации $\bar{\alpha}$, полученный интегрированием по области умножения. Если пренебречь слагаемым с J_s , получим

$$\frac{dJ}{dt} = \frac{2J}{\tau_A} (\bar{\alpha} x_A - 1). \quad (32)$$

Сделаем предположение о малости переменного сигнала:

$$\begin{aligned} \bar{\alpha} &= \bar{\alpha}_0 + \bar{\alpha}' e^{j\omega t} \simeq \bar{\alpha}_0 + \bar{\alpha}' \tilde{\mathcal{E}}_A e^{j\omega t}, \\ \bar{\alpha} x_A &= 1 + x_A \alpha' \tilde{\mathcal{E}}_A e^{j\omega t}, \\ J &= J_0 + \tilde{J}_A e^{j\omega t}, \quad \mathcal{E} = \mathcal{E}_0 + \tilde{\mathcal{E}}_A e^{j\omega t}, \end{aligned} \quad (33)$$

где $\alpha' \equiv \partial\alpha/\partial\mathcal{E}$ и $\bar{\alpha}' = \alpha' \mathcal{E}_A$. Подставляя приведенные выше соотношения в уравнение (32) и пренебрегая членами более высокого порядка малости, получим выражение для переменной компоненты плотности лавинного тока проводимости

$$\tilde{J}_A = \frac{2\alpha' x_A J_0 \tilde{\mathcal{E}}_A}{j\omega\tau_A}. \quad (34)$$

Ток смещения в области умножения равен

$$\tilde{J}_{Ad} = j\omega\epsilon_0 \tilde{\mathcal{E}}_A. \quad (35)$$

Это две компоненты полного тока, протекающего в цепи. Для заданной величины напряженности электрического поля лавинный

ток \tilde{J}_A имеет реактивный характер и изменяется обратно пропорционально ω , как в индуктивности. Другая составляющая тока \tilde{J}_{Ad} также имеет реактивный характер и изменяется прямо пропорционально частоте, как в емкости. Таким образом, область умножения эквивалентна цепи, в которой индуктивность и емкость соединены параллельно (рис. 9, б), причем их величины определяются следующими равенствами (A — площадь поперечного сечения диода):

$$\begin{aligned} L_A &= \tau_A / 2J_0 \alpha' A, \\ C_A &= \epsilon_s A / x_A. \end{aligned} \quad (36)$$

Резонансная частота такой цепи равна

$$f_r \equiv \frac{\omega_r}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{2\alpha' v_s J_0}{\epsilon_s}}. \quad (37)$$

Импеданс области умножения имеет простой вид

$$z_A = \frac{x_A}{j\omega\epsilon_s A} \left(\frac{1}{1 - \omega_r^2/\omega^2} \right) = \frac{1}{j\omega C_A} \left(\frac{1}{1 - \omega_r^2/\omega^2} \right). \quad (38)$$

Множитель γ равен

$$\gamma \equiv \frac{\tilde{J}_A}{\tilde{J}} = \frac{1}{1 - \omega_r^2/\omega^2}. \quad (39)$$

Таким образом, тонкая область умножения ведет себя как параллельная резонансная цепочка с собственной частотой, пропорциональной квадратному корню из плотности постоянного тока (выражение (37))

Область дрейфа и полный импеданс. Подставляя выражение (39) и выражение (25) и интегрируя поле по длине дрейфовой области ($W - x_A$), получим выражение для падения переменного напряжения

$$\tilde{V}_d = \frac{(W - x_A) \tilde{J}}{j\omega\epsilon_s} \left[1 - \frac{1}{1 - \omega_r^2/\omega^2} \left(\frac{1 - e^{-j\theta_d}}{j\theta_d} \right) \right], \quad (40)$$

где θ_d — угол пролета в области дрейфа:

$$\theta_d \equiv \frac{\omega(W - x_A)}{v_s} \equiv \omega\tau_d \quad (41)$$

и

$$\tau_d = \frac{W - x_A}{v_s}. \quad (41a)$$

Определим $C_d = Ae_s/(W - x_A)$ как емкость области дрейфа. Тогда из выражения (40) мы получим импеданс этой области

$$Z_d = \frac{\tilde{V}_d}{\tilde{I}_A} = \frac{1}{\omega C_d} \left[\frac{1}{1 - \omega^2/\omega_r^2} \left(\frac{1 - \cos \theta_d}{\theta_d} \right) \right] + \\ + \frac{j}{\omega C_d} \left[-1 + \frac{1}{1 - \omega^2/\omega_r^2} \left(\frac{\sin \theta_d}{\theta_d} \right) \right] = R + jX, \quad (42)$$

где R и X — активная и реактивная компоненты импеданса соответственно. Очевидно, что реальная часть импеданса (активное сопротивление) будет отрицательной для всех частот, больших ω_r (рис. 9, ε), за исключением нулей при $\theta_d = 2\pi n$, где n — целое. Активное сопротивление положительно при $\omega < \omega_r$, а при уменьшении частоты стремится к

$$R(\omega \rightarrow 0) = \frac{\tau_d}{2C_d} = \frac{(W - x_A)^2}{2Ae_s v_s}.$$

Низкочастотное малосигнальное сопротивление соответствует конечной величине объемного заряда в области дрейфа, а приведенное выше выражение совпадает с равенством (15).

Полный импеданс диода равен сумме импедансов области лавинного умножения, области дрейфа и сопротивления пассивной области R_s :

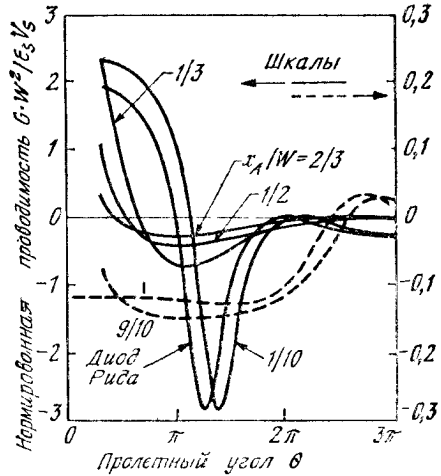
$$Z = \frac{(W - x_A)^2}{2Ae_s v_s} \left(\frac{1}{1 - \omega^2/\omega_r^2} \right) \frac{1 - \cos \theta_d}{\theta_d^2/2} + R_s + \\ + \frac{j}{\omega C_d} \left[\left(\frac{\sin \theta_d}{\theta_d} - 1 \right) - \left(\frac{\sin \theta_d}{\theta_d} + \frac{x_A}{W - x_A} \right) \frac{1}{1 - \omega_r^2/\omega^2} \right]. \quad (43)$$

Равенство (43) записано в форме, удобной для упрощения в случае малых углов пролета θ_d . При $\theta_d \ll \pi/4$ выражение (43) имеет вид

$$Z = \frac{(W - x_A)^2}{2Av_s e_s (1 - \omega^2/\omega_r^2)} + R_s + \frac{j}{\omega C} \left[\frac{1}{(\omega_r^2/\omega^2) - 1} \right], \quad (44)$$

где $C \equiv \varepsilon_s A/W$ — емкость всей области обеднения. Отметим, что первое слагаемое в выражении (44) представляет собой активное сопротивление, которое становится отрицательным при $\omega > \omega_r$. Третье слагаемое — реактивная компонента сопротивления параллельной резонансной цепочки, в которой емкость диода зашунтирована индуктивностью. Реактивное сопротивление носит индуктивный характер при $\omega < \omega_r$ и емкостной при $\omega > \omega_r$. Другими словами, сопротивление становится отрицательным на частотах, соответствующих смене знака реактивной компоненты. Эквивалентная схема и частотные зависимости активной и ре-

Рис. 10. Зависимости нормированной величины проводимости от пролетного угла для шести лавинно-пролетных диодов [30].



активной составляющих импеданса приведены на рис. 9, в и г соответственно.

Импеданс диода Z связан с адмиттансом Y следующим соотношением:

$$Z \equiv \frac{\tilde{V}}{\tilde{I}_A} = R + jX = \frac{1}{G + jB} = \frac{1}{Y}, \quad (45)$$

где G и B — активная и реактивная компоненты адмиттанса соответственно. Теоретические зависимости малосигнальной активной проводимости от пролетного угла для шести ЛПД приведены на рис. 10 [30]. Эти диоды имеют одинаковые области обеднения, но различные по ширине области лавинного умножения ($x_A/W = 1/10, 1/3, 1/2, 2/3, 9/10, 1$). Отметим, что при увеличении толщины слоя лавинного умножения расширяется диапазон частот, в котором сопротивление диода отрицательно. Диод Риды обладает большой отрицательной проводимостью в узком диапазоне частот. Напротив, отрицательная проводимость диода Мисавы ($p-i-n$ -диода с $x_A/W = 1$) мала, но диапазон частот шире и содержит низкие частоты.

10.4. мощность и коэффициент полезного действия

10.4.1. Анализ работы в режиме большого сигнала

На рис. 11 приведены характеристики диода Риды при работе в режиме большого сигнала. Область лавинного умножения с высоким электрическим полем, которое генерирует электронно-дырочные пары, расположена вблизи $p^+ - n$ -перехода (рис. 11, б),

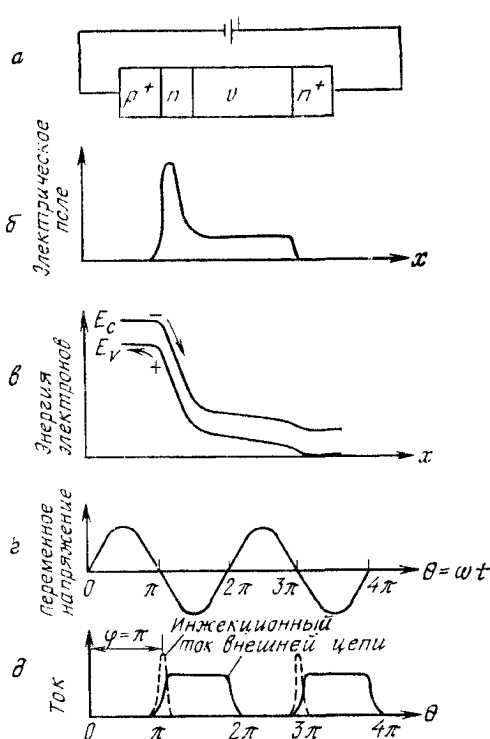


Рис. 11. Характеристики диода при работе в режиме большого сигнала [4].

a — $p^+ - n - v - p^+$ -структура;
 $б$ — распределение поля при про-
 бое; $в$ — энергетическая зонная
 диаграмма; $г$ — переменное напря-
 жение; $д$ — инжекционный ток
 и ток внешней цепи для ди-
 ода Рида.

а слаболегированный v -слой образует область дрейфа с практически однородным полем. Генерируемые дырки быстро попадают в p^+ -область, а генерируемые электроны инжектируются в область дрейфа (рис. 11, б), где совершают работу, равную выделяемой СВЧ-мощности. При периодических колебаниях электрического поля относительно среднего значения (рис. 11, г) коэффициент ионизации, отнесенный к числу носителей заряда, практически мгновенно отслеживает колебания поля. Однако концентрация носителей не повторяет изменения поля, так как генерация зависит от числа уже имеющихся носителей. Даже после того, как поле достигает максимального значения, концентрация носителей продолжает возрастать, поскольку скорость генерации все еще превышает среднее значение. Концентрация носителей максимальна приблизительно в тот момент, когда поле уменьшилось до среднего значения. Таким образом, изменение концентрации инжектированных носителей запаздывает по отношению к переменному напряжению приблизительно на 90° (это так называемое лавинное запаздывание) даже тогда, когда коэффициент ионизации изменяется в фазе с полем. Описанная выше ситуация проиллю-

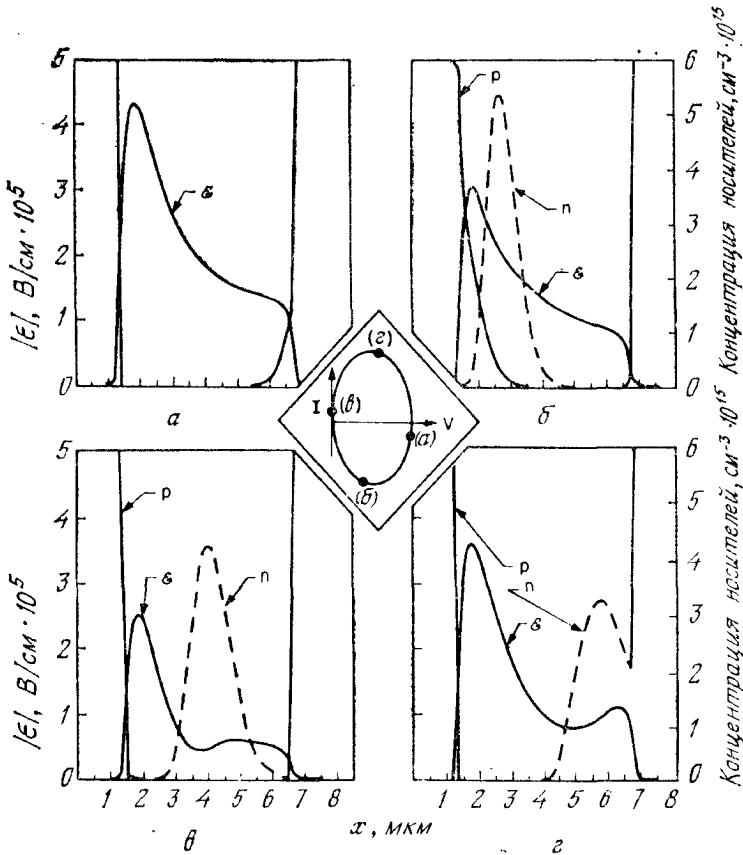


Рис. 12. Результаты расчета на ЭВМ основной моды колебаний ЛПД [31] для четырех случаев, разделенных интервалом в четверть периода. На вставке показана фазовая траектория в координатах ток — напряжение на зажимах.

стрирована на рис. 11, δ изменением инжекционного тока. Максимум переменного электрического поля (или напряжения) соответствует фазе $\pi/2$, а наибольшая концентрация инжектированных носителей — фазе π (т. е. $\varphi = \pi$, как показано на рис. 11). Инжектированные носители затем попадают в область дрейфа, которую они пролетают со скоростью насыщения, внося «пролетное запаздывание». Форма электрического тока, возбуждаемого во внешней цепи, также приведена на рис. 11, δ . Из сравнения временных зависимостей тока и напряжения следует, что сопротивление диода отрицательно.

Детализированные характеристики диода при работе в режиме большого сигнала можно получить с помощью соотношений (27)—

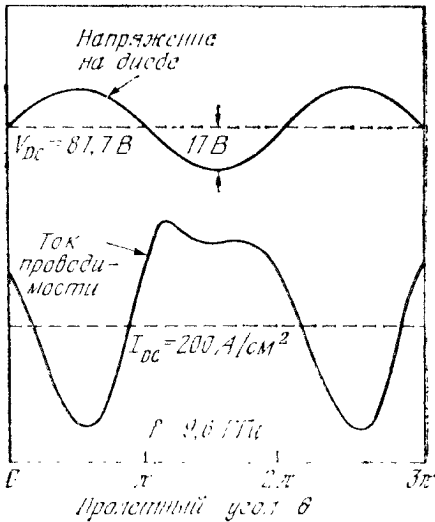


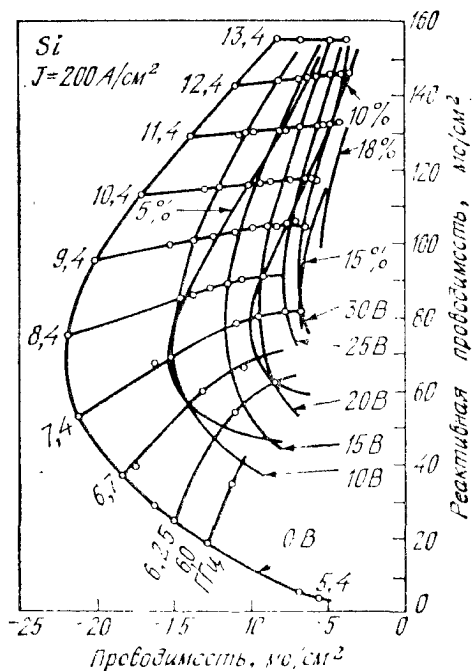
Рис. 13. Зависимости напряжения на зажимах и тока проводимости от пролетного угла для диода с характеристиками, приведенными на рис. 12, при 9,6 ГГц [31].

(29) и приближенных граничных условий. Рассчитанные на ЭВМ зависимости электрического поля и концентраций электронов и дырок от расстояния x в обедненной области кремниевого $p^+ - n - v - n^+$ -диода (при $N_1 = 10^{16} \text{ см}^{-3}$, $b = 1 \text{ мкм}$, $N_2 = 10^{18} \text{ см}^{-3}$ и $W = 6 \text{ мкм}$) в моменты времени, разделенные промежутками в четверть периода, приведены на рис. 12 [31]. Отметим, что 1) генерация электронов и дырок начинается при максимальном напряжении, а через четверть периода импульсы заряда уже полностью сформированы и движутся через соответствующие области дрейфа; 2) дырки быстро покидают активную область, в то время как электроны дрейфуют приблизительно в течение полупериода, индуцируя положительный ток проводимости при отрицательном напряжении; 3) за последующую четверть периода оставшиеся электроны выносятся из активной области, так как напряжение снова приближается к максимальному значению; 4) ток смещения достаточно большой и положительный (соответствует положительному смещению p^+ -контакта), в то время как напряжение на зажимах диода отрицательно; 5) амплитуда модуляции напряжения и тока велика, а эффективность высокая, если область лавинного умножения достаточно широкая.

Зависимости напряжения на зажимах диода и тока проводимости (рис. 12) от пролетного угла приведены на рис. 13. Соотношения между фазами колебаний напряжения на зажимах и тока проводимости близки к идеальным, показанным на рис. 11.

Реактивная проводимость и отрицательная активная проводимость диода в режиме большого сигнала при различных значениях

Рис. 14. Результаты расчетов реактивной проводимости и отрицательной активной проводимости [31].



переменного напряжения приведены на рис. 14. Там же указаны значения к. п. д. Отметим, что реактивная проводимость обычно возрастает с увеличением амплитуды напряжения, а отрицательная активная проводимость убывает. Результаты расчетов показывают, что к. п. д. может достигать 18 %. Эта величина близка к максимальной экспериментально полученной величине к. п. д. кремниевого лавинного-пролетного диода. Тем не менее в некоторых модифицированных диодах Рида из арсенида галлия к. п. д. достигал 40 %. Ниже мы рассмотрим ограничения на величину генерируемой мощности и коэффициент полезного действия.

10.4.2. Температурные ограничения

При низких частотах характеристики ЛДП в непрерывном режиме генерации ограничены в первую очередь температурным фактором, т. е. мощностью, которая может быть рассеяна в виде тепла в полупроводниковом кристалле. Типичный способ монтажа прибора показан на рис. 15, а. Упрощенная структура диода и теплоотвода изображена на рис. 15, б. Полное тепловое сопротив-

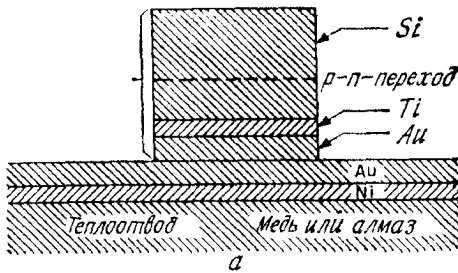
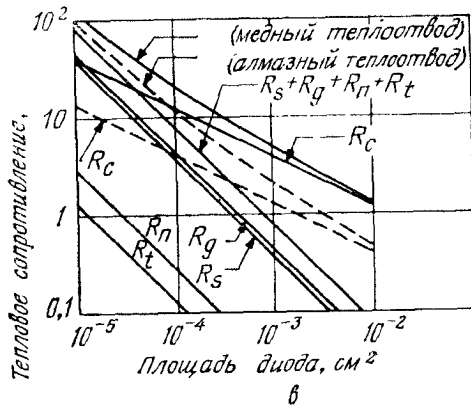
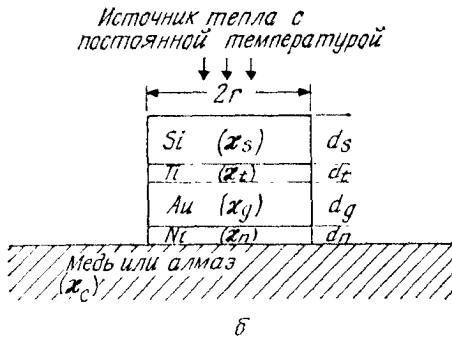


Рис. 15. Диод, смонтированный на медном теплоотводе, в металлических слоях, используемых для термокомпрессированного соединения золото — золото (а), упрощенная структура диода и теплоотвода (б) в зависимости полного теплового сопротивления R_T ($^{\circ}\text{C}/\text{Вт}$) и его компонент от площади диода (в) [32].



ление для кругового источника тепла радиусом r на глубине d_s в кремнии равно [32]

$$R_T = R_s + R_t + R_g + R_n + R_o = \frac{1}{A} \left(\frac{d_s}{x_s} + \frac{d_t}{x_t} + \frac{d_g}{x_g} + \frac{d_n}{x_n} \right) + \frac{1}{4\pi k_c} \quad (46)$$

Символы поясняются рис. 15, б. Последнее слагаемое равно тепловому сопротивлению теплоотвода, занимающего бесконечное полу-

Таблица 1. Теплопроводность и типичные значения толщин слоев различных материалов для диода, работающего в диапазоне частот 15—18 ГГц при 300 К

Материал	Теплопроводность κ , Вт/(см·К)	Толщина d , мкм	d/μ 10 ⁻⁴ см ² ·К/Вт
Кремний [‡]	0,80	3,00	3,80
Титан	0,16	0,02	0,13
Золото	3,00	12,50	4,20
Никель	0,71	0,20	0,28
Медь	3,90	—	—
Алмаз	20,00	—	—

[‡] Теплопроводность для кремния дана при 500 К (по предположению при максимальной рабочей температуре).

пространство. Зависимости различных компонент теплового сопротивления от площади диода приведены на рис. 15, в. Штриховыми линиями показаны составляющая теплового сопротивления алмазного (тип II) теплоотвода и соответствующее полное тепловое сопротивление R_T . Предполагается, что теплопроводность алмаза при 300 К в 3 раза превышает соответствующую величину для меди, а значение теплопроводности кремния соответствует максимальной рабочей температуре 500 К (табл. 1). Из рисунка видно, что алмазный теплоотвод уменьшает тепловое сопротивление приблизительно в 2 раза, а R_T убывает с ростом площади диода.

Мощность P , которая может быть рассеяна в виде тепла в диоде, должна быть равна мощности, которая может быть передана в теплоотвод. Поэтому P равно $\Delta T/R_T$, где ΔT — разность температур p - n -перехода и теплоотвода. Если реактивная проводимость $2\pi fC$ (где $C = A\epsilon_s/W$) остается постоянной, а основной вклад в тепловое сопротивление вносит полупроводник (в предположении, что $d_s \approx W$), то для заданной величины ΔT

$$Pf = \left(\frac{\Delta T}{R_T} \right) f \sim \left[\frac{\Delta T}{(1/A)(W/\kappa_s)} \right] \left(\frac{W}{A\epsilon_s} \right) = \frac{\kappa_s \Delta T}{\epsilon_s} = \text{const.} \quad (47)$$

В этих условиях величина генерируемой в непрерывном режиме мощности уменьшается пропорционально $1/f$.

10.4.3. Ограничения, связанные со свойствами носителей

Максимальная выходная мощность диода на заданной частоте зависит от предельных характеристик полупроводниковых материалов, а также от достижимого в СВЧ-цепях уровня импеданса.

Ограничения, вытекающие из свойств полупроводниковых материалов, связаны со следующими параметрами: 1) величиной напряженности электрического поля \mathcal{E}_m , при которой происходит лавинный пробой; 2) максимально достижимой скоростью дрейфа носителей в полупроводнике v_s — скоростью насыщения.

Максимальное приложенное к полупроводниковому образцу напряжение ограничено напряжением пробоя, которое в случае однородной лавины равно $V_m = \mathcal{E}_m W$, где W — ширина обедненной области. Поскольку ток увеличивает напряженность электрического поля в области пространственного заряда (это следует из уравнения Пуассона), максимальная величина протекающего в образце тока также ограничена лавинными процессами. Если максимальное электрическое поле равно \mathcal{E}_m , максимальный ток определяется выражением $I_m = \mathcal{E}_m \epsilon_s v_s A / W$. Следовательно, верхний предел потребляемой от источника питания мощности

$$P_m = V_m I_m = \mathcal{E}_m^2 \epsilon_s v_s A. \quad (48)$$

Частота, соответствующая углу пролета $\theta = \pi$, равна

$$f = \frac{v_s}{2(W - x_A)} \quad (49)$$

$$\text{и} \quad f \simeq \frac{v_s}{2W} \quad \text{для } W \gg x_A. \quad (49a)$$

Равенство (48) можно записать в виде

$$P_m f^2 \simeq \mathcal{E}_m^2 v_s^2 / 8\pi X_c, \quad (50)$$

где $X_c = (2\pi f C)^{-1}$ — реактивная составляющая импеданса.

Если предположить, что величина импеданса цепи не может быть меньше некоторого значения, то из выражения (50) следует, что максимальная мощность, которая может быть передана подвижным носителям, уменьшается с частотой по закону $1/f^2$. Можно ожидать, что это электронное ограничение будет основным на частотах выше 30 ГГц в кремнии и арсениде галлия.

Следовательно, в режиме непрерывной генерации на относительно низких частотах существенны температурные ограничения ($P \sim 1/f$), а на более высоких частотах — ограничения, связанные со свойствами носителей ($P \sim 1/f^2$). Граничная частота, разделяющая эти интервалы, для данного полупроводника зависит от максимально допустимого увеличения температуры, минимально достижимого уровня импеданса цепи и произведения \mathcal{E}_m на v_s .

Для рабочих температур в диапазоне 150—200 °С максимальная напряженность электрического поля \mathcal{E}_m в кремнии почти на 10 % меньше, чем в арсениде галлия, а скорость насыщения v_s в кремнии почти в 2 раза больше. Поэтому на частотах, больших 30 ГГц, выходная мощность кремниевого ЛПД должна по рас-

четам приблизительно в 3 раза превышать выходную мощность ЛПД из арсенида галлия, работающего на тех же частотах [33]. Можно полагать, что в области субмиллиметровых длин волны выгоднее использовать диоды Мисавы, в которых электрическое поле однородно, так как они обладают широким диапазоном частот, в котором сопротивление отрицательно, а пролетные эффекты не играют такой важной роли в появлении отрицательного сопротивления, как в диодах Рида [34].

При работе в импульсном режиме, когда температурные эффекты можно не учитывать (т. е. при очень коротких импульсах), предельно достижимая мощность на любых частотах $P \sim 1/f^2$.

10.4.4. Ограничения на коэффициент полезного действия

Для эффективной работы ЛПД необходима генерация в области лавинного умножения максимально допустимой величины заряда Q_m при условии, что электрическое поле в области дрейфа не упадет настолько, что скорость носителей станет меньше скорости насыщения. Движение заряда Q_m через область дрейфа приводит к появлению переменной составляющей напряжения на диоде mV_D , где m — коэффициент модуляции ($m \leq 1$) и V_D — среднее падение напряжения на области дрейфа. На оптимальной частоте $\sim v_s/2W$ движение заряда Q_m также приводит к возникновению переменного тока проводимости, который отстает по фазе на угол ϕ от переменного напряжения на диоде. Средняя величина тока проводимости равна постоянному току J_0 . Следовательно, ток проводимости может изменяться лишь в пределах $0-2J_0$. В случае прямоугольного импульса тока проводимости и синусоидального изменения напряжения на диоде по амплитуде и фазе к. п. д. генерации СВЧ-мощности η равен [31, 35]

$$\eta = \frac{\text{Мощность генерируемого сигнала}}{\text{Мощность, потребляемая от источника питания}} = \frac{(2J_0/\pi)(mV_D)}{J_0(V_A + V_D)} |\cos \phi| = \frac{2m}{\pi} \frac{1}{1 + V_A/V_D} |\cos \phi|, \quad (51)$$

где V_A — падение напряжения на области лавинного умножения, а сумма V_A и V_D равна полной величине приложенного постоянного напряжения. Угол ϕ равен запаздыванию фазы тока проводимости. При идеальных условиях $\phi = 180^\circ$, а $|\cos \phi| = 1$. Для диодов с двумя дрейфовыми областями напряжение V_D в выражении (51) следует заменить на $2V_D$. Здесь мы не учли вклад области умножения, поскольку напряжение на этом слое носит индуктивный характер и сдвинуто по фазе относительно тока проводимости на $+\pi/2$. Ток смещения сдвинут по фазе относительно напряжения на $+\pi/2$ и поэтому тоже не дает вклада в среднюю генерируемую мощность.

Из выражения (51) следует, что для повышения к. п. д. диода необходимо увеличивать коэффициент модуляции, оптимизировать угол запаздывания фазы так, чтобы он был близок к 180° , и уменьшать отношение V_A/V_D . Тем не менее для того, чтобы лавинный процесс начинался быстро, напряжение V_A должно быть достаточно велико. При уменьшении отношения V_A/V_D ниже некоторого оптимального значения происходит резкое снижение к. п. д. диода практически до нуля [31].

Если зависимость скорости от поля такая, что насыщение дрейфовой скорости происходит при очень малых значениях напряженности электрического поля, то коэффициент модуляции $m \approx 1$ и скорость носителей все время равна скорости насыщения. В арсениде галлия *n*-типа эффективное насыщение дрейфовой скорости происходит при полях $\sim 10^8$ В/см, а для кремния *n*-типа — при $\sim 2 \cdot 10^4$ В/см. Следовательно, в диодах из арсенида галлия *n*-типа можно получить значительно большие амплитуды переменного напряжения и более высокие к. п. д. [36—38].

Для того чтобы оценить оптимальное значение отношения V_A/V_D , предположим, что пролетная частота, определяемая равенством (49), приблизительно на 20 % превышает резонансную частоту (выражение (37)):

$$f = \frac{v_s}{2(W - x_A)} = 1,2f_r = \frac{1,2}{2\pi} \sqrt{\frac{2\alpha' v_s J_0}{\epsilon_s}}. \quad (52)$$

Напряжение V_D равно

$$V_D = \langle \mathcal{E}_D \rangle (W - x_A) = \langle \mathcal{E}_D \rangle v_s / 2f, \quad (53)$$

где $\langle \mathcal{E}_D \rangle$ — средняя напряженность электрического поля в области дрейфа. При 100 %-ной модуляции тока ($J_0 = J_{dc} = J_{ac}$) максимальная величина заряда $Q_m = m\epsilon_s \langle \mathcal{E}_D \rangle$ определяет плотность тока:

$$J_0 = Q_m f = m\epsilon_s \langle \mathcal{E}_D \rangle f. \quad (54)$$

Если коэффициент ионизации $\alpha \sim \mathcal{E}^\zeta$, величина α' может быть получена следующим образом:

$$\alpha' \equiv \frac{d\alpha}{d\mathcal{E}} = \frac{\zeta\alpha}{\mathcal{E}} \simeq \frac{\zeta(W - x_A)\alpha}{V_D}. \quad (55)$$

Комбинируя выражения (52)—(55), получим соотношение [35]

$$(V_A/V_D)_{\text{opt}} \simeq 4m \left(\frac{1,2}{2\pi} \right)^2 \zeta \alpha x_A. \quad (56)$$

При частоте $f \simeq 10$ ГГц для арсенида галлия оптимальное значение $V_A/V_D = 0,65$ и $m \simeq 1$, в то время как для кремния оно равно $\sim 1,1$ при $m \simeq 1/2$.

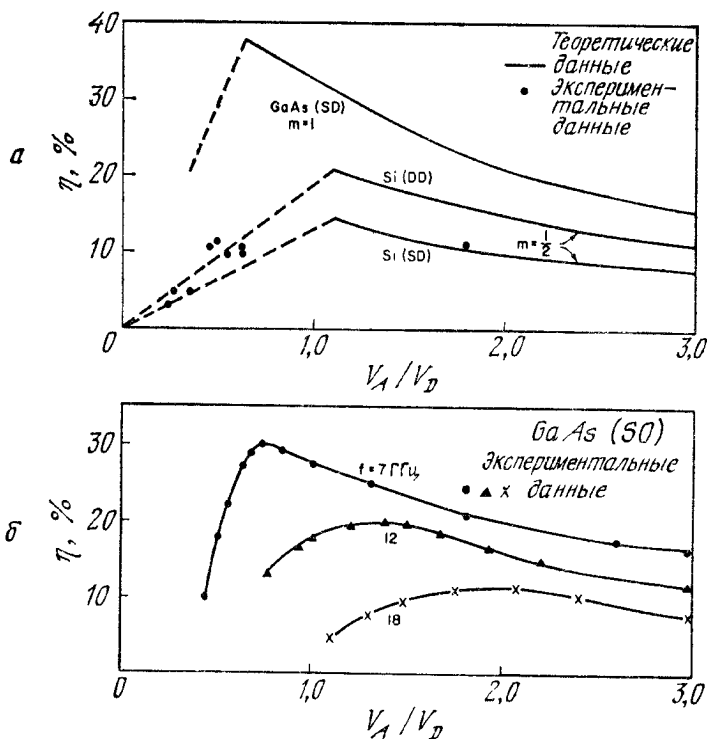


Рис. 16. Зависимость к. п. д. от V_A/V_D для диодов из кремния и арсенида галлия (а) [35] и экспериментально полученные значения к. п. д. в зависимости от V_A/V_D в диодах с двухслойной базой из арсенида галлия (б) [39]. SD — одна область дрейфа; DD — две области дрейфа.

На рис. 16, а приведены зависимости к. п. д. диодов из кремния и арсенида галлия от V_A/V_D . При вычислении максимальных к. п. д. использованы оптимальные значения величин, которые мы получили выше. Ожидаемое значение максимального к. п. д. для кремниевого диода с одной дрейфовой областью составляет около 15 %, с двумя дрейфовыми областями 21 %, а для диодов из арсенида галлия с одной дрейфовой областью — около 38 %. Предыдущие оценки совпадают с экспериментально полученными результатами. На высоких частотах происходит увеличение оптимального значения V_A/V_D , что приводит к уменьшению максимального к. п. д. Приведенные на рис. 16, б экспериментальные результаты для диода из арсенида галлия с одной областью дрейфа вполне согласуются с полученными нами выше оценками [39].

На практике множество различных факторов служит причиной уменьшения к. п. д. лавинно-пролетных диодов. К ним можно отнести влияние пространственного заряда носителей, обратного

тока насыщения, последовательного сопротивления, скин-эффект, насыщение коэффициента ионизации, туннельный эффект, конечность «времени лавинного отклика» и эффекты, связанные с накоплением неосновных носителей.

Влияние пространственного заряда [40] иллюстрируется рис. 17. Генерируемые дырки уменьшают напряженность электрического поля (рис. 17, а), что может привести к преждевременному прекращению процесса лавинного умножения и уменьшению угла запаздывания фазы, который в оптимальных условиях равен 180° . По мере того как дырки дрейфуют вправо (рис. 17, б), пространственный заряд может снизить электрическое поле на левом краю зарядового пакета до значений, меньших, чем это необходимо для насыщения скорости. Такое уменьшение в свою очередь приведет к изменению формы импульса тока и снижению уровня мощности, генерируемой на пролетных частотах.

Влияние обратного тока насыщения [41] иллюстрируется рис. 18. Большой ток насыщения вызывает быстрое нарастание

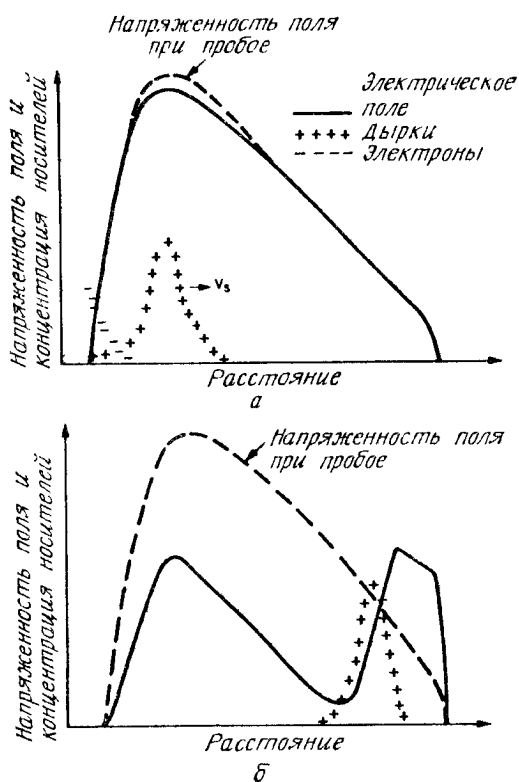


Рис. 17. Мгновенные распределения электрического поля и носителей заряда в диоде Рида [40]

а — процесс лавинного умножения завершен, заряд начинает двигаться через диод; б — пролет заряда почти закончен. Видно, что пространственный заряд носителей сильно снижает электрическое поле.

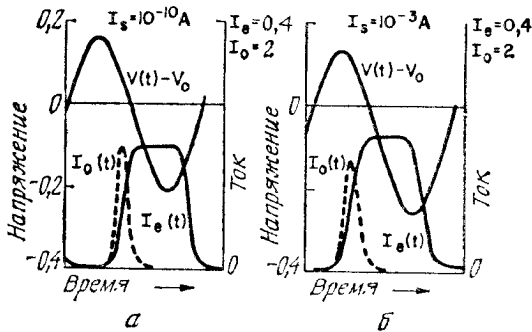


Рис. 18. Фазовые сдвиги для инжекционного тока $I_0(t)$ и тока внешней цепи $I_e(t)$ по отношению к переменному напряжению на диоде. Результаты расчета на ЭВМ для тока насыщения 10^{-10} А и 10^{-3} А [41].

лавны, приводя к уменьшению соответствующего запаздывания фазы. Для $I_s = 10^{-10}$ А $|\cos \phi|$ близок к 1 (рис. 18, а), но для $I_s = 10^{-3}$ А $|\cos \phi|$ равен лишь 0,43 (рис. 18, б), а к. п. д. меньше примерно в два раза. Инжекция неосновных носителей из «плотных» омических контактов также приводит к увеличению обратного тока насыщения и уменьшению к. п. д.

Эффект неполного «перекрывтия» эпитаксиального слоя подложки обедненной областью [42, 43] иллюстрируется рис. 19. Наличие области несмыкания приводит к дополнительному последовательному сопротивлению и снижает отрицательное сопротивление диода. Отметим, что этот эффект в арсениде галлия вы-

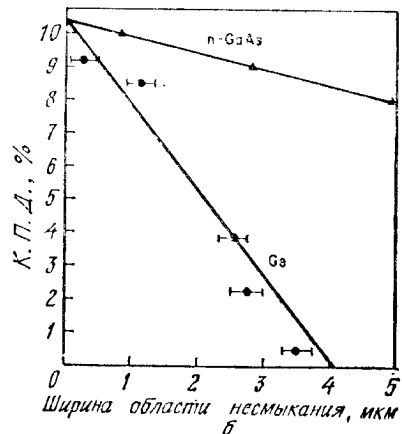
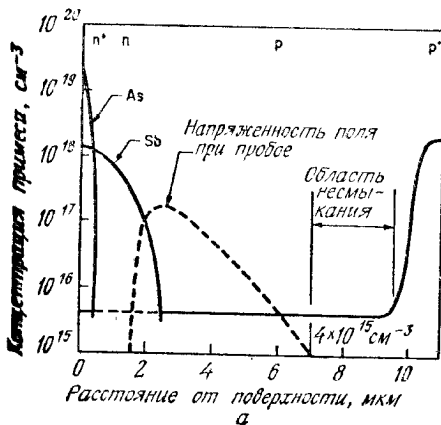


Рис. 19. Распределение электрического поля при пробое в германиевом ЛПД (а) и к. п. д. в зависимости от ширины области несмыкания в эпитаксиальном слое для n -GaAs и Ge-диодов (б) [42, 43].

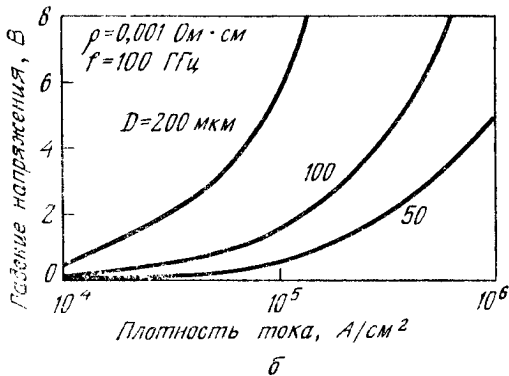
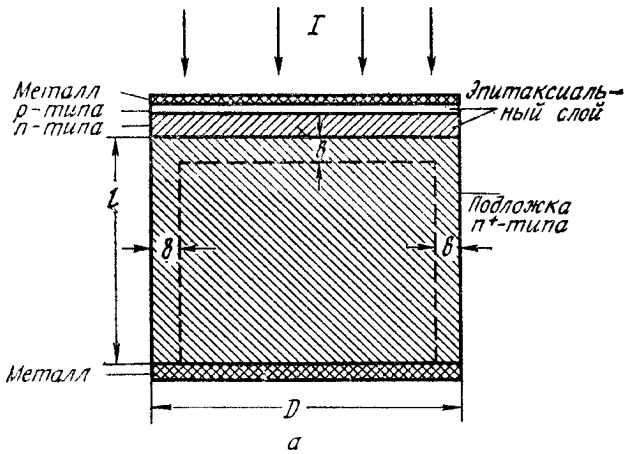


Рис. 20. Скин-эффект [44].

а — протекание тока ограничено тонким поверхностным слоем толщиной δ , что приводит к неоднородности и резистивным потерям; *б* — рассчитанные зависимости падения напряжения на подложке при 100 ГГц для диодов с различными диаметрами.

ражен слабее, поскольку при малых электрических полях в GaAs подвижность носителей существенно выше.

Рассмотрим теперь влияние скин-эффекта [44] (рис. 20). Если рабочая частота ЛПД лежит в диапазоне миллиметровых длин волн, область протекания тока ограничена скин-слоем глубиной δ на поверхности подложки. Поэтому эффективное сопротивление подложки возрастает и возникает падение напряжения в плоскости, перпендикулярной направлению тока (рис. 20, б). Это в свою очередь приводит к неоднородному распределению плотности тока в данной плоскости и дополнительному последовательному сопротивлению, что также снижает к. п. д.

При достаточно высоких рабочих частотах ширина обедненной области очень мала, а необходимое для начала ударной ионизации электрическое поле велико. При сильных полях имеют место два основных эффекта. Во-первых, коэффициент ионизации медленно изменяется, что приводит к уширению импульса инжекционного тока [45], изменению формы тока на выходе и уменьшению к. п. д. Во-вторых, туннельный ток может стать доминирующим. Поскольку этот ток находится в фазе с приложенным полем, запаздывание в 180° , характерное для лавинного умножения, в этом случае не может быть достигнуто. Туннельные механизмы изучались в связи с туннельно-пролетным режимом работы [4, 46, 47]. Предполагается, что шум туннельно-пролетного диода будет меньше, чем шум ЛПД, однако его выходная мощность и к. п. д. также будут значительно меньше.

Другим ограничивающим фактором при работе в субмиллиметровом диапазоне длин волны является конечное запаздывание отклика коэффициента ионизации на изменение электрического поля. Для кремния «время лавинного отклика» $\tau_i < 10^{-18}$ с. Предполагается, что кремниевые ЛПД будут эффективно работать на частотах 300 ГГц и выше, так как τ_i мало по сравнению с временем пролета в диапазоне субмиллиметровых волн. Однако в арсениде галлия τ_i на порядок больше, чем в кремнии [48]. Такие большие значения τ_i могут ограничить рабочий диапазон GaAs-ЛПД частотами, меньшими 100 ГГц.

Эффекты накопления неосновных носителей в $p^+ - n$ - или $n^+ - p$ -диодах, возникающие из-за диффузии генерируемых в активном слое электронов или дырок в нейтральную p^+ - или n^+ -область, также приводят к снижению к. п. д. В то время как основные носители пролетают область дрейфа, неосновные носители накапливаются в нейтральной области, а в следующую фазу цикла диффундируют обратно в активную область, вызывая преждевременную лавину, которая искажает соотношение фаз между током и напряжением.

10.4.5. Необратимый термический пробой диода вследствие образования шнура тока

Причиной перегорания диода может быть не только перегрев, но и неоднородное распределение тока по площади при образовании шнуров с высокой плотностью тока. Такая ситуация возможна в диоде в режиме с отрицательным сопротивлением по постоянному току, когда локальные области максимальной плотности тока характеризуются минимальным напряжением пробоя. Это причина выхода из строя $p-i-n$ -диодов. Объемный заряд движущихся в области дрейфа носителей препятствует возникновению режима с низкочастотным отрицательным сопротивлением

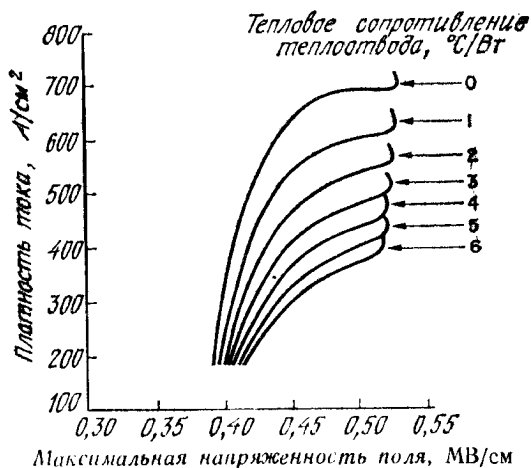


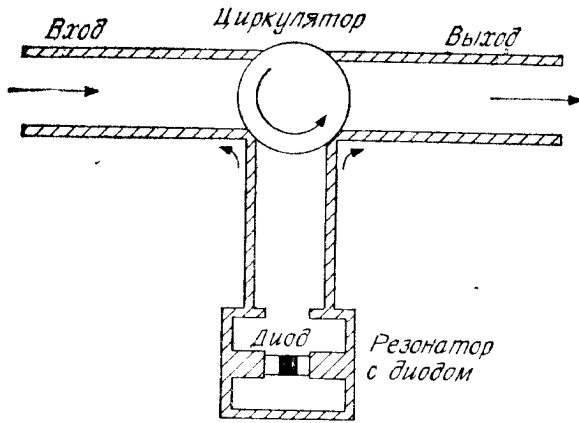
Рис. 21. Зависимость плотности тока от максимальной напряженности электрического поля в кремниевых барьерах Шоттки с $N_D = 6,2 \cdot 10^{15} \text{ см}^{-3}$; $\Phi_{ВР} = 0,79 \text{ В}$ при температуре 22°С [49].

и, следовательно, перегоранию из-за образования шнура тока. Однако диоды, обладающие положительным статическим сопротивлением при малых токах, могут иметь отрицательное динамическое сопротивление и перегорать при больших токах.

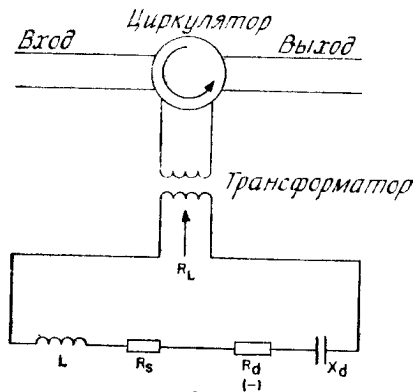
На рис. 21 приведена зависимость плотности тока от максимальной напряженности электрического поля в GaAs-диоде с барьером Шоттки [49] при концентрации примеси $6,2 \cdot 10^{15} \text{ см}^{-3}$ и высоте барьера $0,79 \text{ В}$. Если в режиме непрерывной генерации увеличивается ток прибора, то возрастает и температура в переходе. В конце концов после достижения поворотной точки зависимости тока от поля, за которой начинается область отрицательного дифференциального сопротивления, диод выходит из строя из-за образования шнура тока с высокой плотностью. Отметим, что диод может работать при больших плотностях тока, предшествующих точке поворота, если использовать теплоотвод с малым тепловым сопротивлением.

10.5. ШУМЫ ЛПД

Основной причиной шумов ЛПД является статистическая природа генерации электронно-дырочных пар в области лавинного умножения. Поскольку шумы определяют минимальный уровень СВЧ-сигнала, который может быть усилен, теория шумов в ЛПД очень важна.



а



б

Рис. 22. ЛПД, помещенный в резонатор (а), и эквивалентная схема (б) [50].

Для усиления сигнала лавинно-пролетный диод можно поместить в резонатор, связанный с передающей линией [50], которая в свою очередь связана с входным и выходным трактами с помощью циркулятора (рис. 22, а). На рис. 22, б приведена эквивалентная схема, используемая при малосигнальном анализе. Ниже мы введем две характеристики: коэффициент шума и шумовое отношение. Коэффициент шума определим следующим образом:

$$NF = 1 + \frac{\text{Выходная мощность шума, возникающего в усилителе}}{\text{Коэффициент усиления мощности} \times (kT_0 B_1)} =$$

$$= 1 + \frac{\bar{I}_n^2 R_L}{P_G k T_0 B_1}, \quad (57)$$

где P_G — коэффициент усиления мощности, R_L — сопротивление нагрузки, k — постоянная Больцмана, $T_0 = 290$ К, B_1 — шумовая полоса, $\overline{I_n^2}$ — средний квадрат шумового тока диода, индуцируемого в замкнутой цепи (рис. 22, б). Шумовое отношение определим как

$$M \equiv \frac{\overline{I_n^2}}{4kT_0GB_1} \quad (58)$$

или

$$M \equiv \frac{\overline{V_n^2}}{4kT_0(-Z_{\text{real}})B_1}, \quad (59)$$

где G — отрицательная действительная компонента адмиттанса, $-Z_{\text{real}}$ — действительная часть импеданса диода и $\overline{V_n^2}$ — средний квадрат шумового напряжения. Отметим, что коэффициент шума и шумовое отношение зависят от среднеквадратичного шумового тока (или среднеквадратичного шумового напряжения). Ниже мы покажем, что при частотах, больших резонансной f_r , шум диода уменьшается, однако при этом уменьшается также отрицательное сопротивление. Поэтому приемлемой величиной, характеризующей работу диода в качестве усилителя, является шумовое отношение, и мы заинтересованы в том, чтобы оно было минимальным.

Коэффициент шума в усилителе с большим коэффициентом усиления равен [50]

$$NF = 1 + \frac{qV_A/kT_0}{4\zeta^2 V_A (\omega^2 - \omega_r^2)}, \quad (60)$$

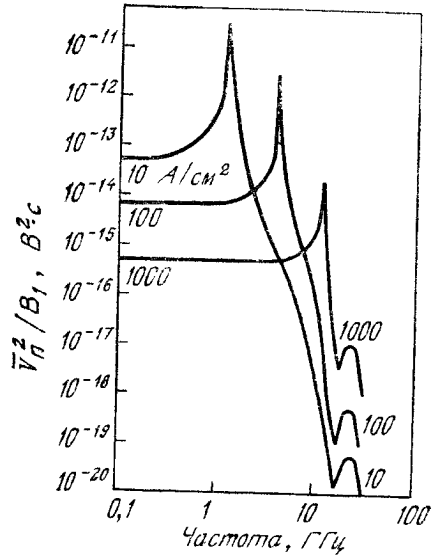
где ζ — показатель степени в зависимости $\alpha \sim \mathcal{E}^\zeta$, τ_A и V_A — время пролета области умножения и падение напряжения на ней соответственно, f_r — резонансная частота (разд. 10.3). Соотношение (60) получено в приближении тонкой области умножения и равенства коэффициентов ионизации электронов и дырок. При $\zeta = 6$ (для кремния), $\omega = 2\omega_r$ и $V_A = 3$ В коэффициент шума на частоте $f = 10$ ГГц по расчетам равен 11 000, или 40,5 дБ.

При реальных значениях коэффициентов ионизации ($\alpha_n \neq \alpha_p$ для кремния) и произвольном профиле распределения примеси выражение для среднего квадрата шумового напряжения на низких частотах имеет вид [51]

$$\overline{V_n^2} = \frac{2qB_1}{J_0 A} \left(\frac{1 + W/x_A}{\alpha'} \right)^2 \sim \frac{1}{J_0}, \quad (61)$$

где $\alpha' = \partial\alpha/\partial\mathcal{E}$. На рис. 23 приведена зависимость $\overline{V_n^2}/B_1$ от частоты для кремниевоего ЛПД с $A = 10^{-4}$ см², $W = 5$ мкм и $x_A = 1$ мкм. Отметим, что, как следует из выражения (61), при низких частотах средний квадрат шумового напряжения $\overline{V_n^2}$

Рис. 23. Средний квадрат шумового напряжения, отнесенный к ширине полосы, в зависимости от частоты для кремниевого ЛПД [51].



обратно пропорционален плотности постоянного тока. Вблизи резонансной частоты (которая пропорциональна $\sqrt{J_0}$) $\overline{V_n^2}$ достигает максимального значения, а затем быстро убывает приблизительно как четвертая степень частоты. Таким образом, уровень шума может быть снижен, если работать при малых токах и на частотах, превышающих лавинную. Эти условия противоположны режиму большой мощности и высокого к. п. д., и поэтому необходимо выбирать промежуточный режим, оптимальный для конкретных применений.

На рис. 24 приведены типичные теоретические и экспериментально полученные [52] зависимости шумового отношения для ЛПД из арсенида галлия. На пролетной частоте 6 ГГц шумовое отношение приблизительно равно 32 дБ. Минимальное шумовое отношение, равное 22 дБ, соответствует частотам, превышающим приблизительно в 2 раза пролетную частоту. Шумовое отношение для ЛПД из арсенида галлия значительно ниже, чем для кремневых диодов. В табл. 2 приведены шумовые отношения для диодов из германия, кремния и арсенида галлия. Шумы в режимах усиления и генерации даны для идеальной цепи без потерь при частотах, соответствующих максимальной эффективности генерации без гармонической настройки.

Основная причина низкого уровня шумов в GaAs-ЛПД состоит в том, что при заданной напряженности электрического поля коэффициенты ионизации электронов и дырок в арсениде галлия одинаковы, в то время как в кремнии они существенно различа-

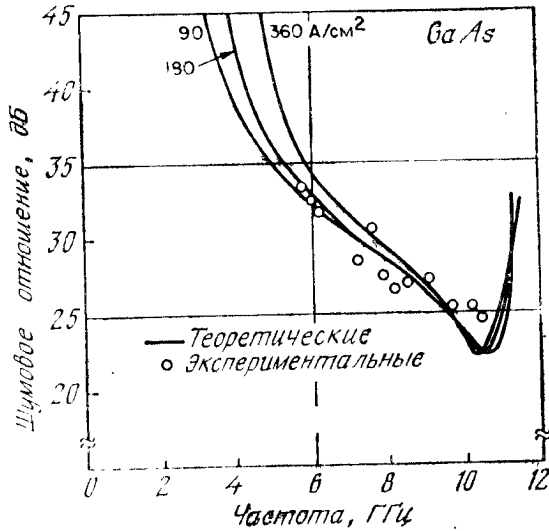


Рис. 24. Шумовое отношение для GaAs-ЛПД с пролетной частотой 6 ГГц [52].

ются. Используя интеграл лавинного умножения, можно показать, что для получения больших коэффициентов умножения M средняя длина ионизации $1/\langle\alpha\rangle$ должна приблизительно равняться ширине слоя умножения x_A , если $\alpha_n = \alpha_p$, и $x_A/\ln M$, если $\alpha_n \gg \alpha_p$. Так, при заданной ширине x_A в кремнии будет происходить значительно больше актов ионизации, и, следовательно, шумы ЛПД будут больше.

На рис. 25 приведены зависимости выходной мощности от шумового отношения для кремниевых и GaAs-ЛПД на частоте 6 ГГц (в случае частотно-модулированного шума) [53]. Мощность нормирована на 1 мВт, т. е. равна $10 \log (P \cdot 10^3)$ дБ, где мощность P выражена в ваттах. Диоды помещались в настроенный на определенную частоту коаксиальный резонатор, причем сопротивле-

Таблица 2. Шумовые отношения для лавинно-пролетных диодов

Полупроводник	Ge	Si	Ga As
Шумовое отношение в режиме малого сигнала, дБ	30	40	25
Шумовое отношение в режиме большого сигнала, дБ	40	55	35

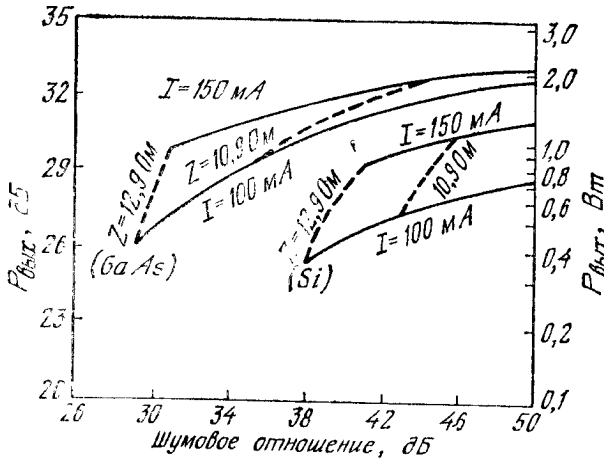


Рис. 25. Зависимости выходной мощности от шумового отношения для генератора с синхронизированной фазой. Величина синхронизирующей мощности поддерживалась постоянной на уровне 4 дБ. Показаны линии, соответствующие постоянной нагрузке Z и постоянному току диода I [53].

ние нагрузки изменялось постепенно при помощи импедансного трансформатора Z . При максимальной выходной мощности шумовое отношение относительно мало. Шумовое отношение можно уменьшить за счет небольшого понижения уровня выходной мощности. Еще раз отметим, что при заданном уровне мощности (например, 1 Вт или 30 дБ) шум ЛПД из арсенида галлия приблизительно на 10 дБ меньше шума кремниевого ЛПД.

10.6. КОНСТРУКЦИЯ ПРИБОРА И ХАРАКТЕРИСТИКИ

10.6.1. Изготовление диода

Основываясь на соображениях, изложенных в разд. 10.4, можно сделать вывод, что лавинно-пролетные диоды обычно конструируются так, чтобы выходная мощность и к. п. д. были максимальны. На рис. 26 показаны некоторые типичные структуры мощных ЛПД.

Диодная структура, изображенная на рис. 26, *a*, получена с помощью двойной эпитаксии или диффузии в эпитаксиальный слой [54]. Для снижения последовательного сопротивления используют n^+ -подложку. Толщину эпитаксиального слоя следует контролировать, чтобы при пробое отсутствовала область несмыкания. При работе на высоких частотах даже подложка n^+ -типа должна быть достаточно тонкой (порядка нескольких микромет-

ров), чтобы снизить потери и влияние неоднородностей из-за скин-эффекта.

На рис. 26, б показан ЛПД с барьером Шоттки, который представляет собой выпрямляющий контакт металл—полупроводник (гл. 5) [55—57]. Несмотря на то что распределения электрического поля в структурах, показанных на рис. 26, а и б, идентичны, диод с барьером Шоттки обладает рядом преимуществ. Во-первых, напряженность электрического поля максимальна на металлургической границе раздела, и поэтому выделяющееся тепло легко отводится через металлический контакт. Диод может быть изготовлен в форме усеченного конуса (рис. 26, б) с целью снижения влияния краевых эффектов, связанных с большой напряженностью электрического поля, и получения одиородной области лавинного пробоя. Во-вторых, диод с барьером Шоттки можно создать при относительно низких температурах, при которых не происходит нарушения высококачественной структуры эпитаксиального слоя. Однако такие диоды обладают существенным недостатком. Дело в том, что в присутствии электронов и дырок высоких энергий атомы полупроводника могут химически воздействовать на металл, что приводит к ухудшению характеристик контакта.

Барьер Шоттки можно также использовать в модифицированном диоде Рида, если вместо p^+ -слоя нанести металлический контакт (рис. 26, в). Поскольку в барьерах Шоттки главную роль играют основные носители, эффект накопления неосновных носителей [58], который имеет место в обычных структурах Рида, здесь несуществен, а к. п. д. может быть даже выше.

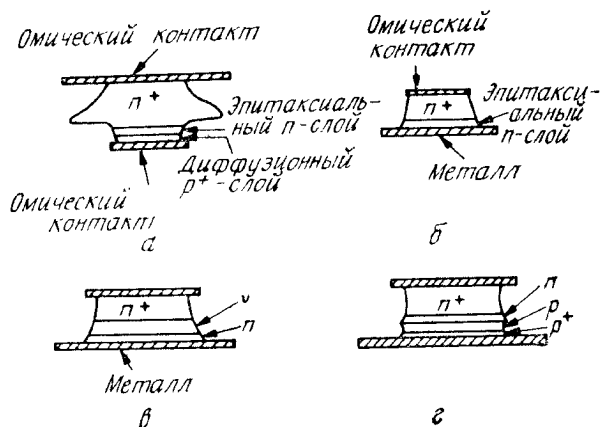


Рис. 26. Структуры некоторых ЛПД.

а — диод, созданный с помощью диффузии или двойной эпитаксии; б — барьер Шоттки; в — диод с двухслойной базой; г — диод с двумя областями дрейфа, созданный с помощью ионной имплантации.

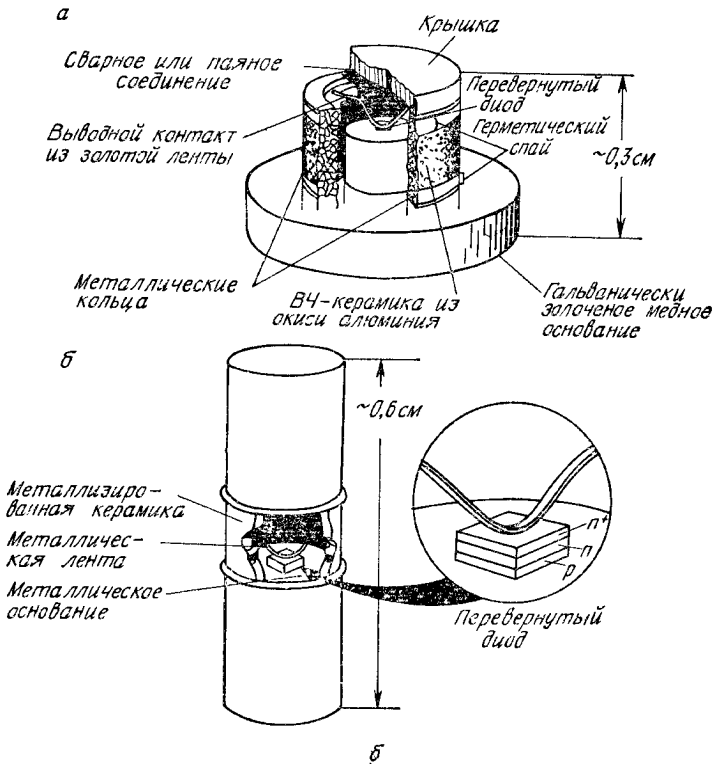


Рис. 27. Два СВЧ-корпуса с помещенными в них ЛПД.

Используя модифицированные структуры Рида, можно получить значительно большие к. п. д. по сравнению с обычными диодами с постоянным профилем примеси. Однако при изготовлении модифицированных диодов Рида требуется более строгий контроль профиля примеси, чтобы получить прибор с определенной частотной характеристикой. Метод самоограниченного анодного травления используется для уменьшения толщины высоколегированного слоя (в диоде с двухслойной базой) и низколегированного слоя (в диоде с трехслойной базой) с целью получения определенного напряжения пробоя (т. е. рабочей частоты) [59]. В большинстве барьеров Шоттки на GaAs высота барьера большая, а обратный ток насыщения мал. Однако платина реагирует с арсенидом галлия при рабочих температурах, образуя $PtAs_2$ и вызывая смещение поверхности барьера. Это «введение» платины в арсенид галлия изменяет напряжение пробоя и приводит к деградации прибора. Эффект можно контролировать, если нанести на подложку слой платины толщиной 200—500 Å с последующим

нанесением слоя ртути или тантала для ограничения реакции платины с арсенидом галлия [60, 61].

На рис. 26, з показана структура диода, созданного с помощью метода ионной имплантации [62]. В случае бора или фосфора, которые наиболее часто используются при легировании полупроводников, глубина проникновения ионов составляет 0,5 мкм/100 кэВ. Поэтому слой шириной 1 мкм можно легко получить на установке для ионной имплантации с энергией ионов, равной нескольким сотням килоэлектронвольт. Такие источники используются при изготовлении ЛПД, работающих в диапазоне миллиметровых длин волн. К достоинствам этого метода следует отнести возможность изготовления структур с двумя дрейфовыми областями (рис. 26, з), в которых выходная мощность и импеданс, отнесенные к единичной площади, приблизительно в 2 раза больше. Поэтому ожидается, что такие структуры будут генерировать большие выходные мощности с более высокими к. п. д. [62].

ЛПД могут быть изготовлены методом молекулярно-лучевой эпитаксии [63]. Поскольку при этом можно контролировать уровень легирования и толщину вплоть до атомных размеров, предполагается, что данный метод окажется весьма полезным при создании ЛПД, работающих в миллиметровом и субмиллиметровом диапазонах длин волн.

Изготовленный диод обычно монтируют в СВЧ-корпус. Два типичных корпуса показаны на рис. 27. В обоих случаях диод крепится диффузионной областью или металлическим электродом на медный или алмазный теплоотвод для обеспечения эффективного охлаждения перехода во время работы.

10.6.2. Связь параметров диода с рабочей частотой и СВЧ-характеристики

Из малосигнальной теории можно получить приближенные соотношения между различными параметрами диода и рабочей частотой. Выражение (43) для сопротивления можно записать в следующем виде:

$$-R \sim \frac{W^2}{2A\varepsilon_s v_s} \left(\frac{1}{\omega^2/\omega_r^2 - 1} \right) \left(\frac{1 - \cos \theta_d}{\theta_d^2/2} \right), \quad (62)$$

где θ_d — угол пролета, равный $\omega\tau$, и $\tau \approx W/v_s$. Для того чтобы сопротивление $-R$ не зависело от ω/ω_r , необходимо выполнить следующие равенства (выражение (62)):

$$W^2/2A\varepsilon_s v_s = \text{const}, \quad \omega_r \tau = \text{const}. \quad (63a), (63b)$$

Поскольку ширина обедненной области W обратно пропорциональна рабочей частоте (выражение (49)), площадь прибора, пропорциональная W^2 , зависит по закону ω^{-2} . Из условия лавинного

Таблица 3. Приближенные соотношения между параметрами ЛПД и рабочей частотой

Параметр	Зависимость от частоты
Площадь поперечного сечения перехода A	f^{-2}
Плотность постоянного тока J	f
Ширина обедненной области W	f^{-1}
Напряжение пробоя V_B	f^{-1}
Выходная мощность $P_{\text{вых}}$	
температурные ограничения	f^{-1}
ограничения, связанные со свойствами полупроводников	f^{-2}
К. п. д. η	Не зависит

пробоя (равенство (2)) можно показать, что коэффициент ионизации α и его производная по напряженности электрического поля α' обратно пропорциональны ширине обедненной области. Комбинируя выражения (37) и (636) с соотношением $\alpha' \sim 1/W$, получим выражение для плотности постоянного тока

$$J_0 \sim \frac{\omega_r^2}{\alpha'} \sim \frac{\omega^2}{1/W} \sim \omega. \quad (64)$$

Соотношения между параметрами диода и рабочей частотой приведены в табл. 3. Они используются при экстраполяции характеристик и проектировании приборов, работающих в других частотных диапазонах.

Ограничения на выходную мощность рассмотрены в разд. 10.4. Предполагается, что при низких частотах коэффициент полезного действия будет очень слабо зависеть от частоты. Однако в диапазоне миллиметровых длин волн плотности рабочих токов велики ($\sim f$), а площадь мала ($\sim f^{-2}$), поэтому рабочая температура диода высокая. Повышение температуры в свою очередь приводит к увеличению обратного тока насыщения и снижению к. п. д. Кроме того, скин-эффект, туннельный и другие эффекты, которые связаны с высокими частотами и сильными электрическими полями, также ухудшают к. п. д. Следовательно, по мере увеличения частоты следует ожидать уменьшения к. п. д.

На рис. 28 показана СВЧ-цепь для измерений на частотах в диапазонах Н (6—8 ГГц) и I (8—10 ГГц) [64]. Наиболее важными компонентами схемы служат: 1) корпус диода, который обеспечивает необходимый теплоотвод, а также снабжен подвижными втулками, используемыми при настройке на резонанс в цепи; 2) измеритель мощности, который служит источником СВЧ-мощности; 3) генератор развертки, с помощью которого детекти-

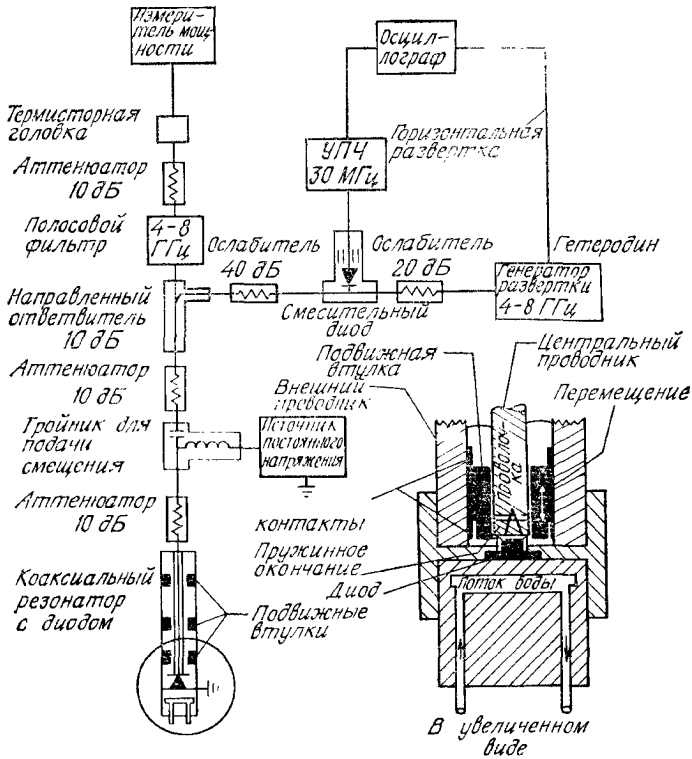


Рис. 28. Блок-схема широкодиапазонной измерительной установки для ЛПД [65].

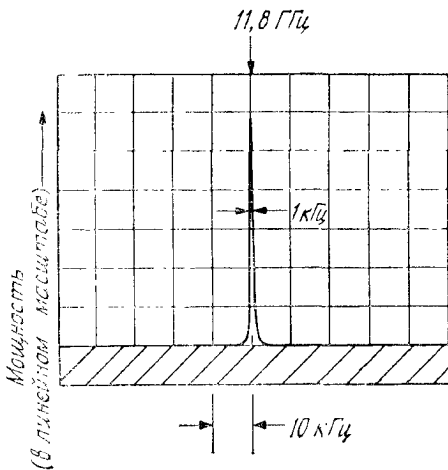


Рис. 29. Спектр генерируемых $p-n$ -диодом колебаний ($V_B = 54$ В) [66].

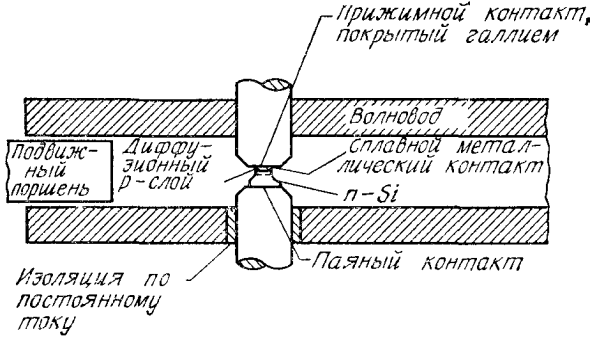


Рис. 30. СВЧ-резонатор для ЛПД миллиметрового диапазона длин волн [67].

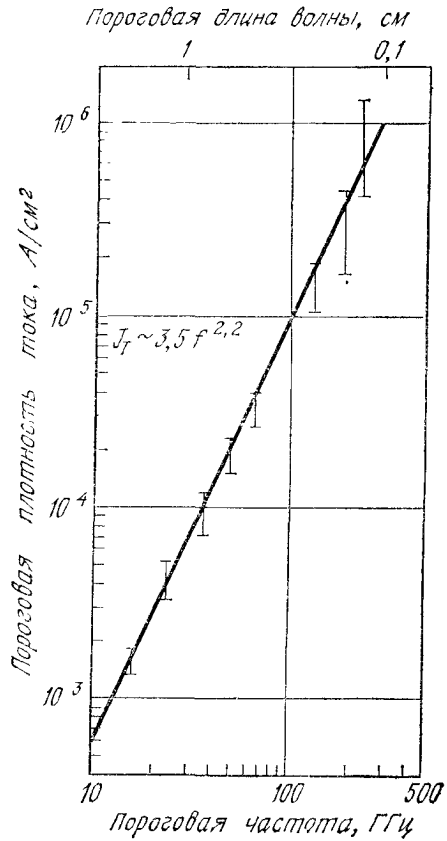


Рис. 31. Зависимость пороговой частоты от плотности тока [67].

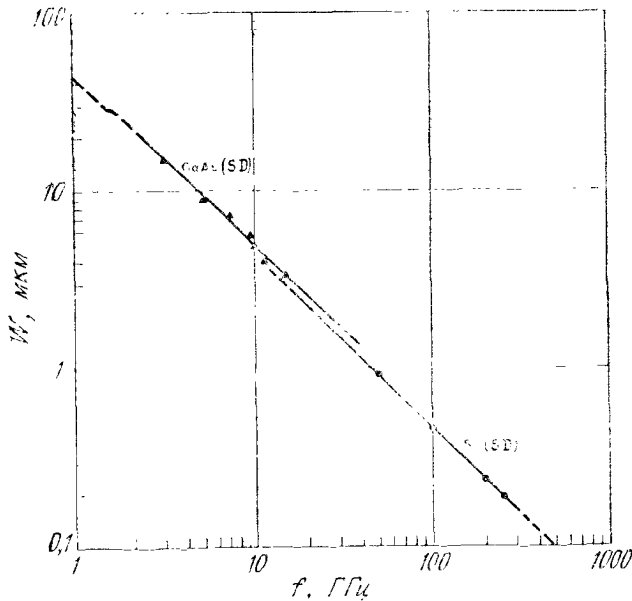


Рис. 32. Зависимость ширины обедненной области от частоты в Si- и GaAs-ЛПД [68, 69] (SD — одна область дрейфа).

руются СВЧ-сигналы; 4) осциллограф для наблюдения выходного сигнала [65]. Типичный спектр колебаний (генерируемый кремниевым $p-v-n$ -диодом), полученный с экрана осциллографа, показан на рис. 29 [66]. Рабочая частота равна 11,8 ГГц, а ширина полосы меньше 1 кГц. Для диапазона миллиметровых длин волн следует использовать другую измерительную схему и другой способ крепления диода; типичный СВЧ-резонатор с диодом показан на рис. 30 [67].

На рис. 31 приведена зависимость пороговой плотности тока, т. е. минимальной плотности тока, при которой начинается генерация, от частоты [67]. Отметим, что пороговая плотность тока изменяется приблизительно как квадрат частоты в соответствии с общими выводами, сделанными нами выше. Зависимость оптимальной ширины обедненной области от частоты для ЛПД из кремния и арсенида галлия [68, 69] приведена на рис. 32. Ширина обедненной области изменяется, как и ожидалось, обратно пропорционально частоте. Интересно, что при частотах выше 100 ГГц ширина обедненной области менее 0,5 мкм. Такая малая ширина слоя может дать представление о трудностях, с которыми приходится сталкиваться при изготовлении модифицированных

диодов Рида и диодов с двумя дрейфовыми областями, работающих в этом диапазоне частот.

В настоящее время наибольшая величина произведения мощности на квадрат частоты получена для диодов с двумя дрейфовыми областями. На рис. 33 приведены характеристики [70] диодов с одной и двумя дрейфовыми областями, работающих на частоте 50 ГГц. Созданный методами ионной имплантации кремниевый ЛПД с двумя дрейфовыми областями в непрерывном режиме генерирует мощность выше 1 Вт на частоте 50 ГГц при максимальном к. п. д. 14 %. Эти данные можно сравнить с характеристиками аналогичного диода с одной дрейфовой областью, который генерирует мощность 0,5 Вт при к. п. д. 10 %. Преимущество диодов с двумя дрейфовыми областями обусловлено тем, что в этих диодах электроны и дырки, генерируемые в лавине, пролетая соответствующие дрейфовые области, совершают работу против сил высокочастотного электрического поля. В диодах с одной дрейфовой областью используется лишь один тип носителей.

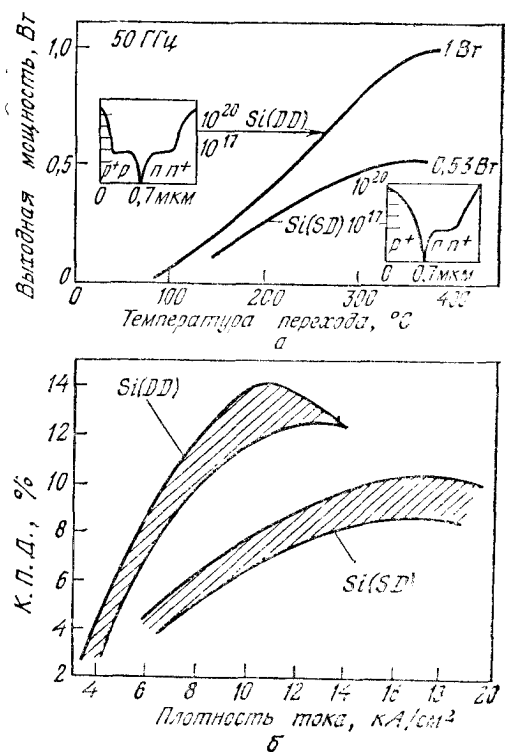


Рис. 33. Выходная мощность (а) и к. п. д. (б) для Si-ЛПД с одной (SD) и двумя (DD) областями дрейфа при 50 ГГц. Значения к. п. д. приведены для четырех диодов каждого типа [70].

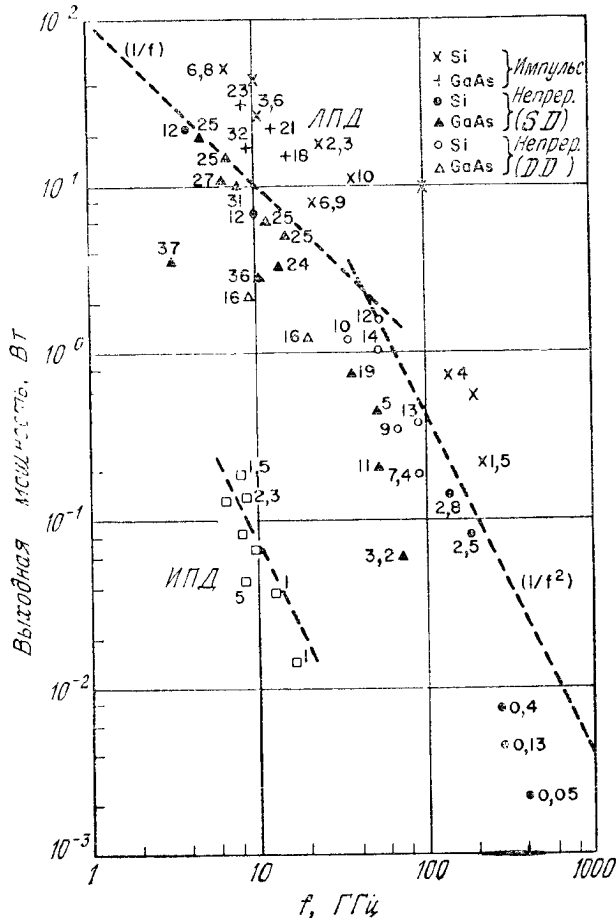


Рис. 34. Характеристики современных ЛПД и ИПД. Рядом с экспериментальными точками указаны значения к. п. д. в процентах [10, 85, 97]. SD — одна области дрейфа; DD — две области дрейфа.

Характеристики современных ЛПД [10, 71, 72] приведены на рис. 34 вместе с характеристиками инжекционно-пролетных диодов (ИПД), которые изучены в разд. 10.7. При низких частотах выходная мощность ограничена тепловыми эффектами и зависит от частоты как $1/f$, а при более высоких частотах (>50 ГГц) мощность ограничена характеристиками носителей и пропорциональна $1/f^2$. На частотах, близких к 10 ГГц, для модифицированных диодов с трехслойной базой и контактом Pt—GaAs получены к. п. д., близкие к 40 % [73]. Из рис. 34 видно, что ЛПД являются наиболее мощными твердотельными источниками СВЧ-

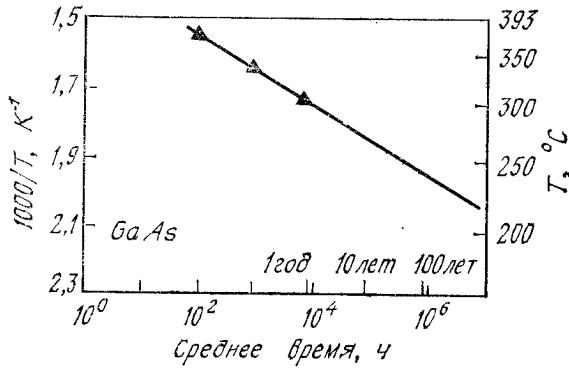


Рис. 35. Зависимость среднего времени работы GaAs-ЛПД до выхода из строя от температуры [74].

колебаний. Из всех твердотельных приборов ЛПД могут генерировать максимальную мощность в непрерывном режиме в миллиметровом диапазоне длин волн.

ЛПД очень надежны при работе в высокотемпературных условиях, если при их изготовлении применяется улучшенная технология. На рис. 35 приведена зависимость среднего времени работы прибора до выхода из строя от температуры для ЛПД из арсенида галлия [74]. Величина энергии активации, стоящая в показателе экспоненты этой зависимости, равна 1,9 эВ. Из экстраполяции этой зависимости следует, что при рабочих температурах в переходе $\sim 200^\circ\text{C}$ срок службы такого диода составит $\sim 10^7$ ч. Аналогичные результаты получены для кремниевых ЛПД [75].

10.7. ИНЖЕКЦИОННО-ПРОЛЕТНЫЕ ДИОДЫ

И ДВУХСКОРОСТНЫЕ ПРОЛЕТНЫЕ ДИОДЫ

Инжекционно-пролетный диод принадлежит к семейству пролетных СВЧ-диодов [12]. В основе его работы лежат два механизма: термоэмиссия (инжекция) и диффузия неосновных носителей через прямосмещенный барьер и пролет носителей через область дрейфа (что приводит к запаздыванию фазы на угол пролета $3\pi/2$). Обычно ИПД маломощны и имеют низкий к. п. д., однако отношение сигнал/шум лучше, чем у других диодов. ИПД используются чаще всего в качестве гетеродинов в СВЧ-приемниках.

Инжекционно-пролетный диод представляет собой два выпрямляющих контакта с общей базой, которая в рабочих условиях полностью обеднена носителями. Сначала мы рассмотрим протекание тока в таком диоде. Затем проведем анализ работы в режимах

малого и большого сигналов. В разд. 10.7.4 мы рассмотрим работу родственного прибора — двухскоростного пролетного диода (ДСПД).

10.7.1. Протекание тока

Сначала рассмотрим протекание электрического тока в симметричной структуре металл—полупроводник—металл (МПМ) с однородно легированным слоем полупроводника n -типа (рис. 36, а) [76]. На рис. 36, б—г приведены распределения объемного заряда, электрического поля и энергетическая диаграмма

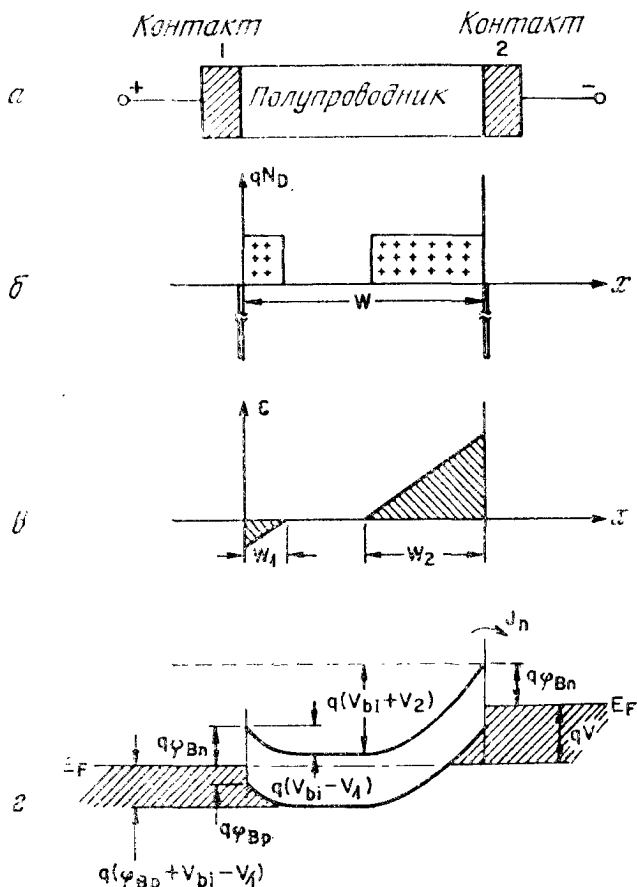


Рис. 36. Структура металл — полупроводник — металл (МПМ-структура) [76]. а — МПМ-структура с однородно легированным полупроводником n -типа; б — распределение пространственного заряда при малых смещениях диода; в — распределение поля; г — энергетическая диаграмма.

смещенного диода соответственно (меньшее положительное напряжение приложено к контакту 1, причем контакт 1 смещен, таким образом, в прямом направлении, а контакт 2 — в обратном). Ширина обедненных областей определяется следующими равенствами:

$$W_1 = \sqrt{\frac{2\varepsilon_s}{qN_D} (V_{bi} - V_1)} \quad (65)$$

и

$$W_2 = \sqrt{\frac{2\varepsilon_s}{qN_D} (V_{bi} + V_2)}, \quad (66)$$

где W_1 и W_2 — ширина обедненных областей прямо- и обратнo-смещенных барьеров соответственно, N_D — концентрация ионизованной примеси и V_{bi} — контактная разность потенциалов. В этих условиях полный ток равен сумме обратного тока насыщения диода Шоттки с высотой барьера Φ_{Bn} , генерационно-рекомбинационного тока и тока поверхностных утечек.

По мере увеличения напряжения в конце концов произойдет смыкание обедненной области контакта, смещенного в обратном направлении, с обедненной областью прямосмещенного контакта (рис. 37, а). Соответствующее напряжение называется напряжением смыкания (или напряжением прокола) V_{RT} . Это напряжение можно найти, приравняв $W_1 + W_2 = W$ (где W — ширина n -области):

$$\begin{aligned} V_{RT} &= \frac{qN_D}{2\varepsilon_s} W^2 - W \left[\frac{2qN_D}{\varepsilon_s} (V_{bi} - V_1) \right]^{1/2} \simeq \\ &\simeq \frac{qN_D}{2\varepsilon_s} W^2 - W \left(\frac{2qN_D V_{bi}}{\varepsilon_s} \right)^{1/2}. \end{aligned} \quad (67)$$

При дальнейшем увеличении напряжения энергетические зоны на контакте 1 становятся плоскими. В этом случае электрическое поле при $x = 0$ равно нулю, а падение напряжения на первом контакте $V_1 = V_{bi}$ (рис. 37, б). Соответствующее напряжение называется напряжением плоских зон V_{FB} :

$$V_{FB} \equiv \frac{qN_D W^2}{2\varepsilon_s}. \quad (68)$$

Зависимость напряжения плоских зон от концентрации примеси в кремниевых диодах с различной шириной базы приведена на рис. 38. Для заданной ширины базы максимальная величина V_{FB} ограничена напряжением лавинного пробоя.

Величина постоянного смещения ИПД в условиях генерации СВЧ-мощности обычно лежит между V_{RT} и V_{FB} . В этом случае ($V_{RT} < V < V_{FB}$) приложенное напряжение V связано с высотой барьера, смещенного в прямом направлении, следующим соотношением:

$$V_{bi} - V_1 = (V_{FB} - V)^2 / 4V_{FB}. \quad (69)$$

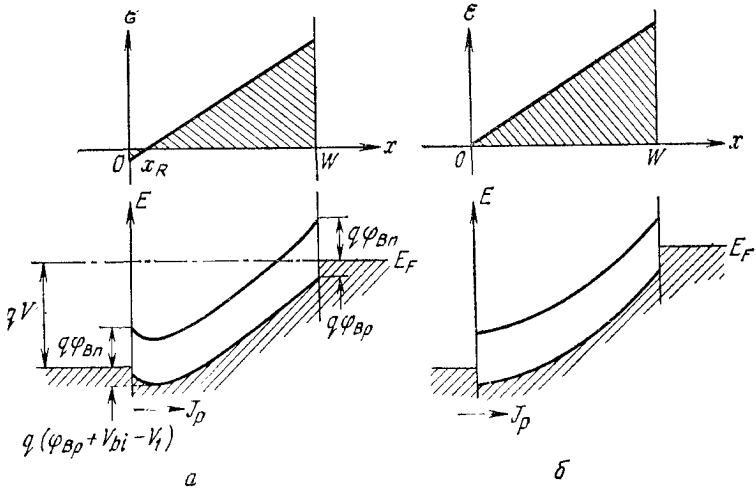


Рис. 37. Распределение электрического поля и энергетическая диаграмма МПМ-структуры при проколе (а) и в условиях плоских зон (б) [76].

Положение точки x_R , в которой происходит смыкание обедненных областей, определяется равенством

$$x_R/W = (V_{FB} - V)/2V_{FB}. \quad (70)$$

После смыкания обедненных областей ток термически эмиттированных через барьер φ_{BP} дырок становится доминирующим:

$$J_p = A_p^* T^2 e^{-q(\varphi_{BP} + V_{bi})/kT} (e^{qV/kT} - 1), \quad (71)$$

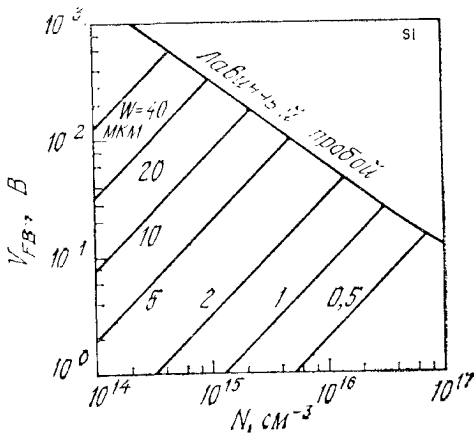


Рис. 38. Зависимость напряжения плоских зон от концентрации примеси в кремниевых диодах с различной шириной базы. Для заданной ширины базы максимальная величина напряжения плоских зон ограничена напряжением лавинного пробоя [76].

где A_p^* — эффективная постоянная Ричардсона (гл. 5). Из равенства (69) для $V \geq V_{RT}$ получаем

$$J_p = A_p^* T^2 e^{-q\phi_{BP}} \exp \left[-\frac{q(V_{FB} - V)^2}{4kTV_{FB}} \right]. \quad (72)$$

Таким образом, после смыкания обедненных носителями областей зависимость тока от напряжения будет экспоненциальной.

Если ток настолько велик, что концентрация инжектированных носителей сравнивается с концентрацией ионизованной примеси, объемный заряд подвижных носителей влияет на распределение электрического поля в области дрейфа. Это явление называется эффектом ограничения объемным зарядом. Если дырки пересекают n -область со скоростью насыщения v_s , а ток $J \gg qv_s N_D$, то уравнение Пуассона принимает вид

$$\frac{d\mathcal{E}}{dx} = \frac{\rho}{\epsilon_s} = \frac{q}{\epsilon_s} \left(N_D + \frac{J}{qv_s} \right) \simeq \frac{J}{\epsilon_s v_s}. \quad (73)$$

Дважды проинтегрировав это уравнение с граничными условиями $\mathcal{E} = 0$, $V = 0$ при $x = 0$, получим [77]

$$J = \left(\frac{2\epsilon_s v_s}{W^2} \right) V = qv_s N_D \left(\frac{V}{V_{FB}} \right). \quad (74)$$

Изложенные выше соображения можно применить для других структур, таких, как $p^+ - n - p^+$ и $p^+ - i - n - \pi - p^+$ -диоды (рис. 39). Выражения для напряжений смыкания и плоских зон в $p^+ - n - p^+$ -диоде имеют такой же вид, как и для МПМ-структуры. Для усложненной $p^+ - i - n - \pi - p^+$ -структуры из аналогичных рассуждений можно получить величины V_{RT} , V_{FB} , x_R и соотношения между приложенным напряжением и падением напряжения на смещенном в прямом направлении переходе [78].

Протекание тока через $p^+ - n - p^+$ -диод, в котором произошел прокол, определяется теми же механизмами, что и в МПМ-структуре. Единственным отличием является отсутствие в равенствах (71) и (72) множителя $\exp(-q\phi_{BP}/kT)$ в случае инжекции носителей через смещенный в прямом направлении $p^+ - n$ -переход [77], т. е.

$$J = A^* T^2 \exp[-q(V_{FB} - V)/4kTV_{FB}] = J_{FB} \exp \left[-\frac{q(V_{FB} - V)^2}{4kTV_{FB}} \right]. \quad (75)$$

Высота барьера в структуре PtSi—Si равна 0,2 эВ. Следовательно, при температуре 300 К и заданном напряжении, превышающем напряжение смыкания, ток $p^+ - n - p^+$ -диода будет приблизительно в 3000 раз больше тока МПМ-структуры. При комнатной температуре $J_{FB} (\equiv A^* T^2)$ составляет $\sim 10^7$ А/см². Поэтому в обычных рабочих условиях эффект ограничения объемным зарядом становится существенным при токах, значительно меньших J_{FB} .

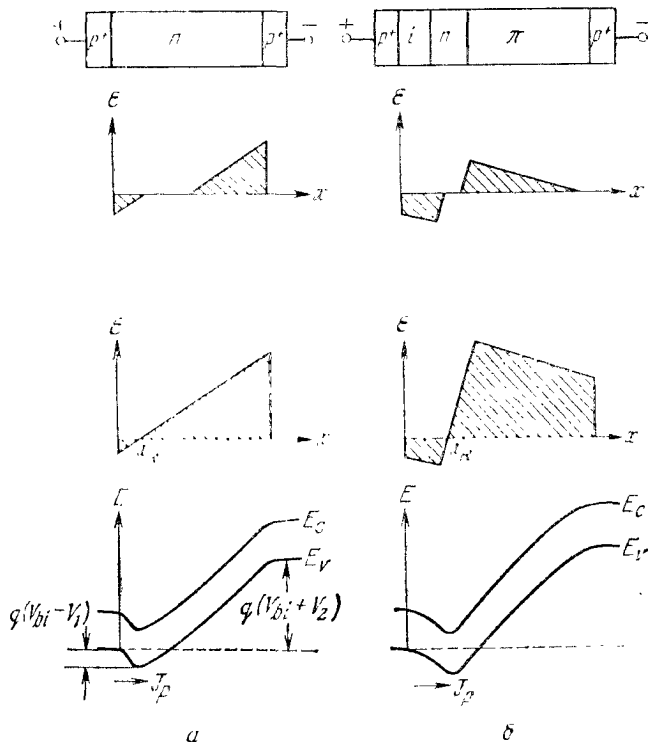


Рис. 39. Распределение электрического поля при малых смещениях и проколе и энергетические диаграммы при проколе для $p^+ - n$ -структуры (а) и $p^+ - i - n - p^+$ -структуры (б).

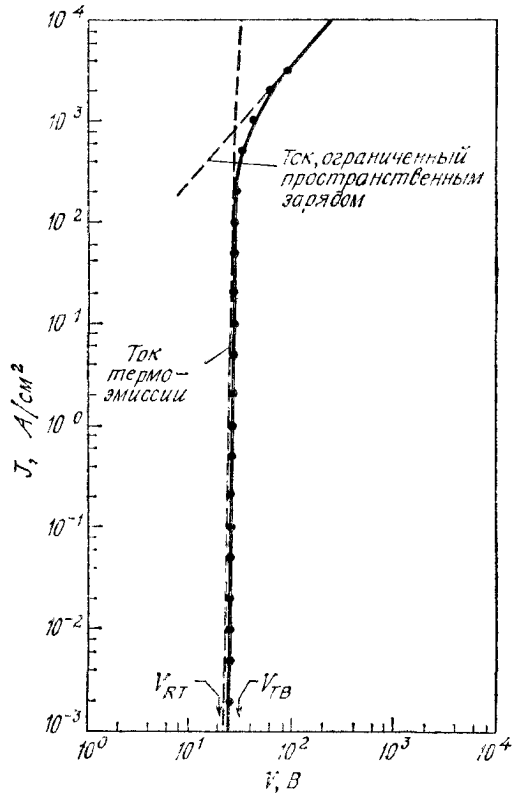
Типичная вольт-амперная характеристика кремниевого $p^+ - n - p^+$ -диода с концентрацией примеси $5 \cdot 10^{14} \text{ см}^{-3}$ и шириной базы 8,5 мкм приведена на рис. 40. Напряжение плоских зон равно 29 В, а напряжение смыкания составляет ~ 21 В. Отметим, что сначала ток возрастает экспоненциально, а затем линейно с напряжением. Экспериментальные результаты находятся в хорошем соответствии с рассчитанными с помощью равенств (74) и (75) зависимостями.

Для эффективной работы ИПД необходимо контролировать величину инжектированного заряда. Поэтому ток должен резко возрастать с увеличением напряжения. Линейная вольт-амперная зависимость, связанная с эффектом ограничения объемным зарядом, будет ухудшать характеристики прибора. Обычно оптимальная плотность тока значительно меньше $J = qv_s N_D$.

Описанный выше диод со смыканием обедненных областей (который также называют диодом с проколом) используется в каче-

Рис. 40. Вольт-амперная характеристика кремниевого $p^+ - n - p^+$ -диода в условиях работы с проколом [77].

— — — теоретическая; —●— экспериментальная.



стве быстродействующего ограничителя напряжения, поскольку ток в таком диоде резко возрастает, как только напряжение превысит напряжение смыкания. Эффект накопления заряда практически отсутствует; кроме того, диод обладает хорошей температурной стабильностью. Были созданы диоды со смыканием с рабочим напряжением $\sim 1,5$ В, причем характеристики были сравнимы и даже лучше, чем характеристики зинеровского диода, в котором используются явления лавинного пробоя или туннелирования (при больших напряжениях — лавинный пробой, при малых — туннелирование) [79].

10.7.2. Малосигнальные характеристики

Ниже показано, что малосигнальное сопротивление инжекционно-пролетного диода отрицательно и, следовательно, в нем самопроизвольно возбуждаются колебания. Рассмотрим $p^+ - n - p^+$ -

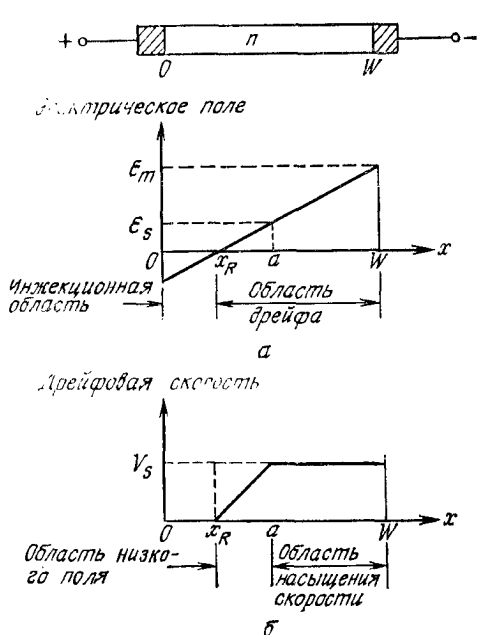


Рис. 41. Распределения электрического поля (а) и скорости (б) в области дрейфа в ИПД [81].

структуру. Распределение электрического поля при напряжении смещения, большем напряжения смыкания, приведено на рис. 41, а. Точка x_R соответствует максимальной потенциальной энергии инжектированных дырок; ее положение определяется выражением (70). Точка a отделяет область относительно слабых электрических полей от области, в которой дрейфовая скорость носителей равна скорости насыщения, т. е. $\mathcal{E} \geq \mathcal{E}_s$, а $v = v_s$ (рис. 41, б). Если концентрация инжектированных носителей невелика, то

$$a \approx \frac{v_s \mathcal{E}_s}{qN_D} + x_R. \quad (76)$$

Время пролета носителями области дрейфа ($x_R < x < W$) определяется следующим выражением [80]:

$$\tau_d = \int_{x_R}^a \frac{dx}{\mu_n \mathcal{E}(x)} + \int_a^W \frac{dx}{v_s} = \int_{x_R}^a \frac{dx}{\mu_n q N_D x / \varepsilon_s} + \frac{W - a}{v_s} \approx \frac{3,75 \varepsilon_s}{q \mu_n N_D} + \frac{W - a}{v_s}. \quad (77)$$

Для того чтобы получить выражение для малосигнального импеданса диода, используем подход, аналогичный подходу, примененному в разд. 10.3.2, и представим зависящие от времени вели-

чны в виде сумм, не зависящих от времени и малосигнальных переменных компонент:

$$\begin{aligned} J(t) &= J_0 + \tilde{J}e^{j\omega t}, \\ V(t) &= V_0 + W\tilde{\mathcal{E}}e^{j\omega t}. \end{aligned} \quad (78)$$

Подставив эти соотношения в выражение (75), получим выражение для линеаризованной переменной составляющей инжекционного тока дырок

$$\tilde{J} = \sigma\tilde{\mathcal{E}}, \quad (79)$$

где σ — инжекционная проводимость на единичную площадь, равная

$$\sigma = J_0 \frac{\epsilon_s (V_{FB} - V_0)}{N_D W kT}, \quad (80)$$

причем J_0 — плотность тока, определяемая выражением (75), в котором V следует заменить на V_0 . Инжекционная проводимость возрастает при увеличении напряжения, достигает максимума, а затем быстро убывает, когда V_0 приближается к V_{FB} . Из выражений (75) и (80) можно найти смещение диода при максимальной проводимости σ :

$$V_{0\sigma \text{ макс}} = V_{FB} - \sqrt{2kT V_{FB}/q}. \quad (81)$$

Поскольку переменное электрическое поле непрерывно на границе инжекционной и дрейфовой областей, эти области взаимодействуют друг с другом. Определим \tilde{J} как плотность полного переменного тока, а \tilde{J}_I — плотность инжекционного тока. Предположим, что инжекционная область настолько тонкая, что носители заряда без запаздывания поступают в область дрейфа. Переменная компонента плотности тока проводимости определяется следующим равенством:

$$\tilde{J}_c(x) = \tilde{J}_I e^{-j\omega\tau(x)} \equiv \gamma \tilde{J} e^{-j\omega\tau(x)} \quad (82)$$

и представляет собой незатухающую волну, распространяющуюся по направлению к точке $x = W$ с пролетным запаздыванием по фазе, равным $\omega\tau(x)$. Комплексная величина $\gamma \equiv \tilde{J}_I/\tilde{J}$ равна отношению плотностей переменных инжекционного и полного токов.

В заданной точке области дрейфа плотность полного переменного тока \tilde{J} равна сумме тока проводимости \tilde{J}_c и тока смещения \tilde{J}_d :

$$\tilde{J} = \tilde{J}_c(x) + \tilde{J}_d(x) \neq f(x). \quad (83)$$

Ток смещения связан с напряженностью переменного электрического поля $\tilde{\mathcal{E}}(x)$ следующим соотношением:

$$\tilde{J}_d(x) = j\omega\epsilon_s\tilde{\mathcal{E}}(x). \quad (84)$$

Комбинируя выражения (80), (82) и (84), получим переменную составляющую напряженности электрического поля в области дрейфа как функцию x и \tilde{J} :

$$\mathcal{E}(x) = \frac{\tilde{J}}{j\omega\varepsilon_s} [1 - \gamma e^{-j\omega\tau(x)}]. \quad (85)$$

Проинтегрировав $\mathcal{E}(x)$, получим зависимость переменной компоненты падения напряжения на области дрейфа от плотности тока \tilde{J} . Коэффициент γ определяется выражением

$$\gamma = \frac{\tilde{J}_1}{\tilde{J}_1 + \tilde{J}_d} = \frac{\sigma}{\sigma + j\omega\varepsilon_s}. \quad (86)$$

Подставив γ в выражение (85) и проинтегрировав по длине области дрейфа ($W - x_R$) с учетом граничных условий $\tau = 0$ при $x = x_R$ и $\tau = \tau_d$ при $x = W$, получим переменную составляющую падения напряжения на области дрейфа

$$V_d = \frac{\tilde{J}(W - x_R)}{j\omega\varepsilon_s} \left[1 - \frac{\sigma}{\sigma + j\omega\varepsilon_s} \frac{1 - e^{j\theta_d}}{j\theta_1} \right], \quad (87)$$

где θ_d — угол пролета в области дрейфа:

$$\theta_d = \frac{\omega [(W - x_R) + (a - x_R)]}{v_s} = \omega\tau_d, \quad (88)$$

а θ_1 — постоянная, равная

$$\theta_1 \equiv \omega \left(\frac{W - x_R}{v_s} \right). \quad (89)$$

Определим $C_d = \varepsilon_s/(W - x_R)$ как емкость области дрейфа. Из выражения (87) получим малосигнальный импеданс структуры

$$Z \equiv \frac{\tilde{V}_d}{\tilde{J}} = R_d - jX_d, \quad (90)$$

где R_d и X_d — малосигнальные действительная и мнимая составляющие импеданса соответственно:

$$R_d = \frac{1}{\omega C_d} \left(\frac{\sigma}{\sigma^2 + \omega^2\varepsilon_s^2} \right) \left[\frac{\sigma(1 - \cos\theta_d) + \omega\varepsilon_s \sin\theta_d}{\theta_1} \right],$$

$$X_d = \frac{1}{\omega C_d} - \frac{1}{\omega C_d} \left(\frac{\sigma}{\sigma^2 + \omega^2\varepsilon_s^2} \right) \left[\frac{\sigma \sin\theta_d - \omega\varepsilon_s(1 - \cos\theta_d)}{\theta_1} \right].$$

Отметим, что действительная составляющая (сопротивление) отрицательна при условии, что угол пролета θ_d больше π , но меньше 2π и если $|(1 - \cos\theta_d)/\sin\theta_d|$ меньше $\omega\varepsilon_s/\sigma$.

На рис. 42 приведены рассчитанные значения (штриховые линии) и экспериментально полученные значения (сплошные

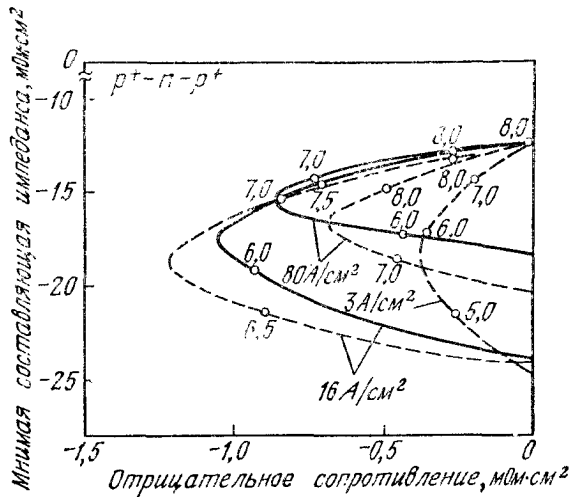


Рис. 42. Малосигнальный импеданс $p^+ - n - p^+$ -диода в области отрицательного сопротивления при трех различных плотностях тока [81]:

--- теоретические значения; — экспериментальные значения.

линии) действительной и мнимой составляющих малосигнального импеданса $p^+ - n - p^+$ -диода с $L = 8,5$ мкм и $N = 5 \cdot 10^{14}$ см $^{-3}$ [81]. Для плотности тока 3 А/см 2 отрицательное сопротивление $-R_d$ максимально на частоте $f = 5,7$ ГГц. При увеличении плотности тока до 16 А/см 2 растут амплитуда максимума $-R_d$ и соответствующая частота. При увеличении плотности тока до 80 А/см 2 максимальное отрицательное сопротивление $-R_d$ начинает убывать в результате уменьшения инжекционной проводимости (выражение (80)).

Теоретические и экспериментальные результаты совпадают в общих чертах. Малосигнальная теория также применима к инжекционно-пролетным диодам с МПМ-структурой, для которых получено хорошее соответствие с экспериментальными данными.

Из полученных выше результатов следует, что 1) инжекционно-пролетные диоды имеют отрицательное малосигнальное сопротивление и, следовательно, могут самовозбуждаться; 2) инжекция через смещенный в прямом направлении $p^+ - n$ -переход или барьер Шоттки служит источником носителей заряда; 3) время пролета носителями области дрейфа является важной величиной для частотных характеристик.

Инжекционно-пролетный диод представляет собой прибор с низким уровнем шума, основными источниками которого являются дробовой шум инжектированных носителей (инжекцион-

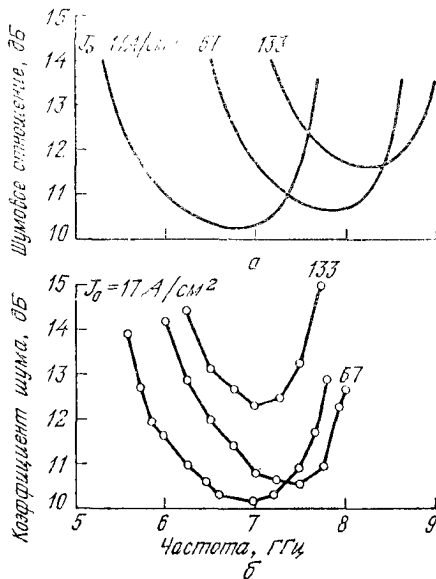


Рис. 43. Рассчитанные шумовые отношения (а) и экспериментально полученные коэффициенты шума (б) Si-ИПД с $W = 7,9$ мкм и $N_D = = 1,2 \cdot 10^{15}$ см $^{-3}$ [82].

ный шум) и случайные флуктуации скорости в области дрейфа (диффузный шум).

Рассчитанные и экспериментально полученные результаты [82] находятся в хорошем соответствии при малых плотностях токов (рис. 43). Для $p^+ - n - p^+$ -диода с $N_D = 1,2 \cdot 10^{15}$ см $^{-3}$ и $W = 7,9$ мкм теоретическая величина шумового отношения составляет ~ 10 дБ при плотности тока 17 А/см 2 . При увеличении плотности тока кривые сдвигаются в сторону более высоких частот. Из сравнения этих результатов с шумами ЛПД видно, что уровень шума в инжекционно-пролетных диодах существенно ниже.

10.7.3. Характеристики при работе в режиме большого сигнала [83]

Основные характеристики ИПД при работе в режиме большого сигнала приведены на рис. 44. Носители инжектируются в виде δ -импульса, когда величина переменного напряжения максимальна ($\varphi = \pi/2$). В течение трех четвертей периода эти носители движутся в направлении отрицательно смещенного контакта:

$$\theta_d = \omega \tau_d = 3\pi/2 \quad (91)$$

или

$$i = \frac{3}{4\tau_d}, \quad (91a)$$

где θ_d — угол пролета и τ_d — время пролета. В первом приближении τ_d равно W/v_s . Следовательно, частота осцилляций

$$f \approx \frac{3v_s}{4W}. \quad (92)$$

Более точное выражение для оптимальной частоты можно получить, подставив выражение (77) в выражение (92).

По оценкам максимальный к. п. д. ИПД равен $\sim 10\%$ при условии, что носители инжектируются при $\varphi = \pi/2$ (рис. 44). Однако можно достигнуть больших к. п. д., если инжектировать носители в более поздний момент цикла, т. е. при $\pi/2 < \varphi \leq \pi$. С этой целью был создан многослойный $n^+ - i - p - v - n^+$ -ИПД [84] (на рис. 39, б показана комплементарная $p^+ - i - n - \pi - p^+$ -структура). Тормозящее поле $n^+ - i - p$ -области увеличивает время инжекционного запаздывания. Измеренные значения выходной мощности и к. п. д. приведены на рис. 45. Максимальное значение к. п. д. составляет более 5% при токе $J \approx 50$ А/см². Соответствующее значение выходной мощности равно 40 мВт

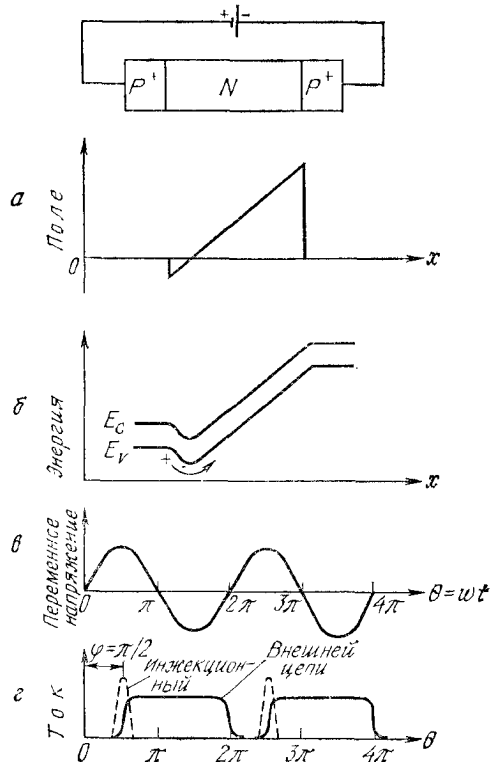


Рис. 44. Работа ИПД в режиме большого сигнала:

а — $p^+ - n - p^+$ -структура; б — распределение поля после пролета; в — энергетическая диаграмма; г — переменное напряжение; д — инжекционный ток и ток внешней цепи.

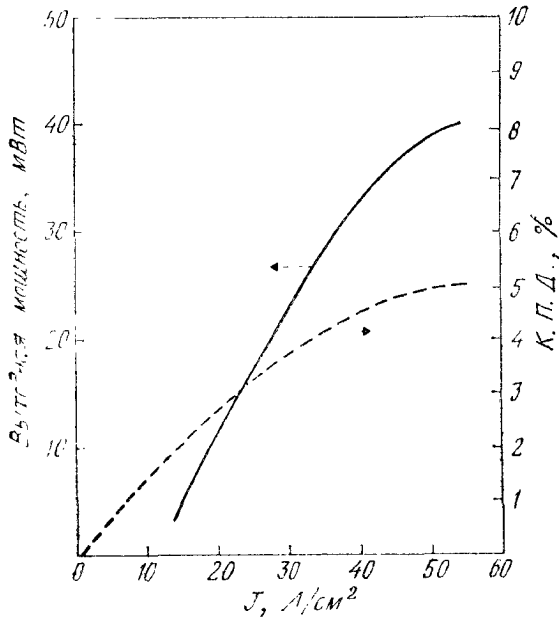


Рис. 45. Зависимости выходной мощности и к. п. д. $n^+ - i - p - v - n^+$ -ИПД от плотности тока [84].

на частоте 7,6 ГГц. После нанесения платинового контакта Шоттки на эпитаксиальный $n-p^+$ -слой с концентрацией $N_D = 2,5 \times 10^{15} \text{ см}^{-3}$ и $W = 6,5 \text{ мкм}$ была получена выходная мощность 152 мВт на частоте 8,6 ГГц при максимальном к. п. д. 2,3 % [85]. Характеристики современных ИПД приведены на рис. 34 [84, 85, 97]. Хотя выходная мощность на частоте $\sim 10 \text{ ГГц}$ приблизительно на два порядка меньше, чем в ЛПД, шумовое отношение ИПД меньше приблизительно во столько же раз. Предполагается, что путем оптимизации времени инжекционного запаздывания можно реализовать потенциальные преимущества ИПД как малошумящего прибора с приемлемыми выходной мощностью и к. п. д.

10.7.4. Двухскоростной пролетный диод

Диод с двухскоростным режимом пролета, или двухскоростной пролетный диод (ДСПД), подобен ИПД за одним исключением: скорость носителей вблизи инжектирующего контакта в нем значительно ниже, чем вблизи другого контакта [86, 87]. Для $n^+ - p - n - n^+$ -ДСПД предложена гетероструктура с $n^+ - p$ -переходом из $Al_xGa_{1-x}As$ и $n - n$ -переходом из арсенида галлия

(вставка на рис. 46). На этой вставке также приведено распределение дрейфовых скоростей носителей в диоде, причем скорость насыщения v_1 в области I меньше скорости насыщения v_2 в области 2. Постоянное смещение «проколота» электрическим полем, а инжектируемый через смещенный в прямом направлении $n^+ - p$ -переход ток очень большой.

Из-за дополнительного запаздывания фазы при пролете носителями области с малой скоростью насыщения ожидается, что отрицательное сопротивление ДСПД будет больше, чем сопротивление ИПД (который является предельным случаем ДСПД с равными скоростями насыщения $v_1 = v_2$). Максимальное малосигнальное отрицательное сопротивление определяется выражением [87]

$$R_{\text{макс}} = \frac{[v_1 (1 + \cos \varphi) - 2v_2] \cos \varphi}{A\omega^2 v_s}, \quad (93)$$

где φ — инжекционное запаздывание фазы, A — площадь прибора. На рис. 46 приведены теоретические значения $R_{\text{макс}}$ в зависимости от инжекционного запаздывания фазы φ при различных соотношениях между скоростями насыщения. Отметим, что

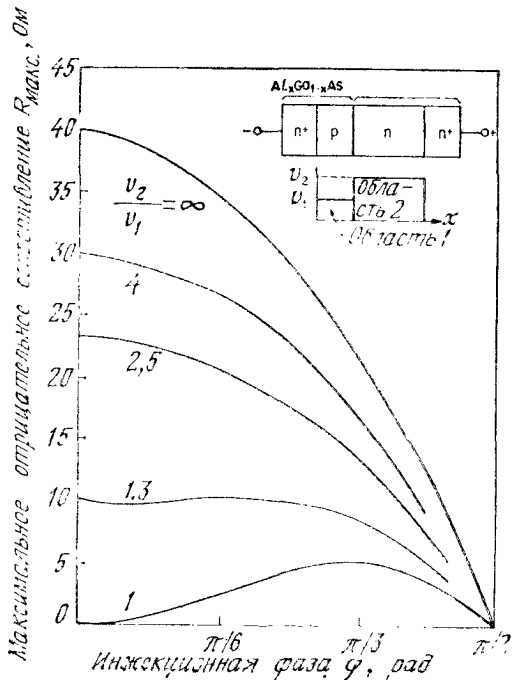


Рис. 46. Зависимость максимального отрицательного сопротивления от инжекционного запаздывания фазы в ДСПД при 10 ГГц и $v_s = 8 \cdot 10^8 \text{ см/с}$ и $A = 10^{-4} \text{ см}^2$. На вставке показаны структура прибора и распределение дрейфовых скоростей [87].

в ИПД ($v_2/v_1 = 1$) для получения максимального отрицательного сопротивления необходимо инжекционное запаздывание фазы, равное $\sim \pi/3$ рад. Из рис. 46 ясно видны преимущества ДСПД с большим отношением скоростей насыщения. Отрицательное сопротивление сравнительно велико, а прибор может работать даже при условии отсутствия инжекционного запаздывания ($\varphi = 0$). Поэтому токи в ДСПД могут быть больше, чем в ИПД, в котором с ростом плотности тока происходит уменьшение инжекционного запаздывания до тех пор, пока не исчезнет отрицательное сопротивление. Для ДСПД теоретическое значение к. п. д. превышает 25 %.

10.8. ПРОЛЕТНЫЙ ДИОД С ЗАХВАЧЕННЫМ ОБЪЕМНЫМ ЗАРЯДОМ ЛАВИНЫ

Этот диод является мощным прибором с высоким коэффициентом полезного действия. В настоящее время максимальная экспериментально полученная выходная мощность диода в импульсном режиме генерации равна 1,2 кВт на частоте 1,1 ГГц (пять последовательно соединенных диодов [88]), а максимальный к. п. д. равен 75 % на частоте 0,6 ГГц [89]. Однако способ генерации сигнала в таком диоде весьма сложен, и, кроме того, работа в этом режиме требует точного контроля параметров цепи и самого диода. Шумовое отношение этого диода существенно больше, чем шумовое отношение ЛПД, а максимально достижимая рабочая частота, по-видимому, меньше частот диапазона миллиметровых длин волн.

Для того чтобы понять, как возникает пролетный режим с захваченным объемным зарядом лавины, рассмотрим $n^+ - p - p^+$ -диод, в котором ток резко возрастает в момент $t = 0$ (рис. 47, а). Если при $t = 0$ максимальная напряженность поля в диоде меньше критической напряженности \mathcal{E}_m , которую необходимо достичь для того, чтобы начался лавинный пробой, то со временем будет происходить увеличение напряженности электрического поля (рис. 47, б). Причина этого в том, что ток связан с полем следующим соотношением [20, 90]:

$$J = \varepsilon_s \frac{d\mathcal{E}}{dt}. \quad (94)$$

Следовательно, напряженность электрического поля равна

$$\mathcal{E}(x, t) = \mathcal{E}_m - \frac{qN_A}{\varepsilon_s} x + \frac{Jt}{\varepsilon_s}, \quad (95)$$

где N_A — концентрация примеси в p -области. Поэтому промежуток времени, в течение которого электрическое поле достигает критического значения \mathcal{E}_m на заданном расстоянии x от контакта

в обедненной области, можно найти, приравняв $\mathcal{E}(x, t) = \mathcal{E}_m$, т. е.

$$\frac{qN_A x}{\epsilon_s} = Jt. \quad (96)$$

Продифференцировав соотношение (96), найдем скорость распространения границы области лавинного умножения (или скорость лавинной зоны) v_z (рис. 47, в):

$$v_z \equiv \frac{dx}{dt} = \frac{J}{qN_A}. \quad (97)$$

Так, например, для $N_A = 10^{15} \text{ см}^{-3}$ и $J = 10^4 \text{ А/см}^2$ скорость v_z значительно превышает скорость насыщения и равна $6 \cdot 10^7 \text{ см/с}$. То есть лавинная зона [19] (или ударный фронт лавины) быстро перескакает практически всю базовую область, которая заполняется высокопроводящей электроно-дырочной плазмой, причем пространственный заряд плазмы сильно снижает напряжение на диоде [18, 91].

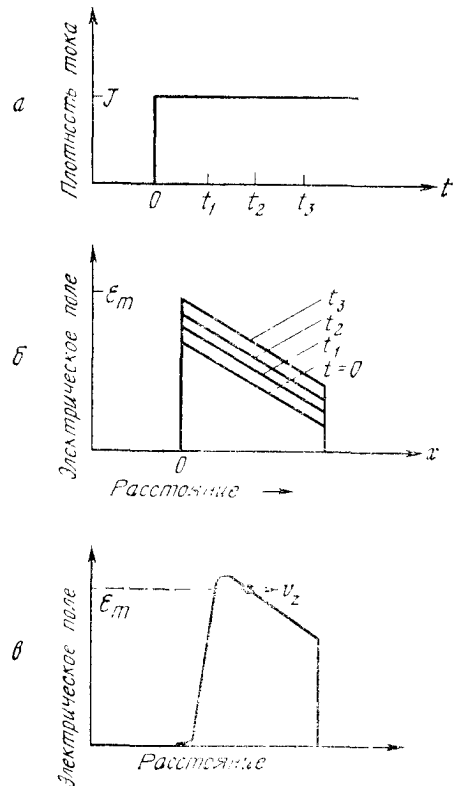


Рис. 47. Характеристики пролетного диода с захваченным объемом зарядом лавины [20, 94]:

a — плотность тока в зависимости от времени; *б* — идеальное распределение электрического поля, возрастающего со временем; *в* — ударный фронт лавины, распространяющийся в диоде со скоростью v_z , большей скорости насыщения v_s .

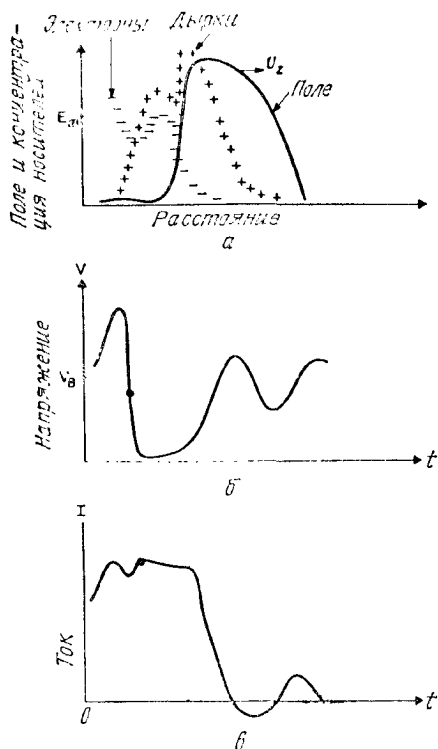
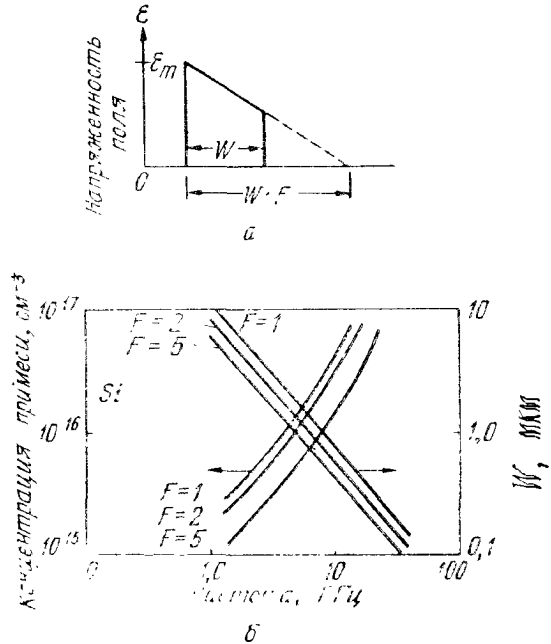


Рис. 48. Характеристики работы диода в режиме с захваченным объемным зарядом лавины [92]: a — результаты расчета на ЭВМ для мгновенного распределения поля и пространственного заряда; b и c — зависимости напряжения и тока от времени соответственно. Точкой на зависимостях на рис. b и c отмечен момент времени, для которого распределения приведены на рис. a .

Рассчитанные на ЭВМ распределения плотности заряда и электрического поля в диоде [92] в некоторый момент времени приведены на рис. 48, a . Зависимости напряжения и тока диода от времени приведены на рис. 48, b и c соответственно. Падение напряжения на диоде в начале периода может существенно превышать стационарное напряжение пробоя. Оно уменьшается после того, как лавинная зона пройдет через базу прибора, однако вскоре опять увеличится, поскольку электрическое поле поляризует объемный заряд подвижных носителей. Если скорость дрейфа носителей не слишком велика, то она пропорциональна напряженности электрического поля (т. е. $v = \mu \mathcal{E}$), а время пролета электронов и дырок через область дрейфа определяется подвижностями и будет значительно больше W/v_s .

Таким образом, в этом диоде возможна генерация колебаний значительно меньших частот, поскольку время рассасывания объемного заряда плазмы, равное отношению заряда Q к току I , может сильно превышать обычное пролетное время при больших напряженностях электрического поля. Этот диод является пролетным прибором в том смысле, что время запаздывания носите-

Рис. 49. Параметры прибора, используемые при расчете. Коэффициент перекрытия F определяется, как показано на рис. 49, а [40].



лей, возникающее из-за конечной скорости пересечения областй дрейфа (т. е. время между инжекцией на одном контакте и сбором носителей заряда на другом контакте), используется здесь для получения соответствующего сдвига по фазе тока относительно о приложенного напряжения.

Для работы диода необходима внешняя цепь, которая хорошо пропускает гармоники основной частоты при больших амплитудах напряжения. Кроме того, гармоник должно быть много, чтобы обеспечить требуемую задержку тока по фазе при таких низких частотах. Диод не должен выходить из строя при максимальном напряжении основной частоты. Необходимо, чтобы диод выдерживал смещение приблизительно в течение четверти периода до тех пор, пока пик гармонического напряжения не «запустит» процесс лавинного умножения. Так как напряжение на зажимах такого диода может быть малым при больших токах, и наоборот, к. п. д. довольно высок. Для некоторых кремниевых диодов были предсказаны к. п. д. [93, 94], превышающие 50 %.

Конструкция и характеристики таких диодов более сложные, поскольку сильная взаимосвязь прибора и внешней цепи предъявляет особые требования к большинству параметров. Для того чтобы уменьшить напряжение, необходимое для начала движения лавинной зоны, диод обычно изготавливают с большим коэффициентом

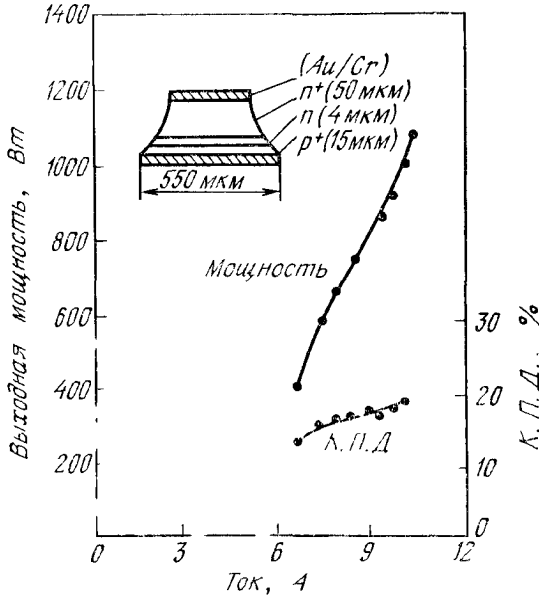


Рис. 50. Зависимость выходной мощности и к. п. д. от тока для пяти соединенных последовательно диодов при 1,1 ГГц [88].

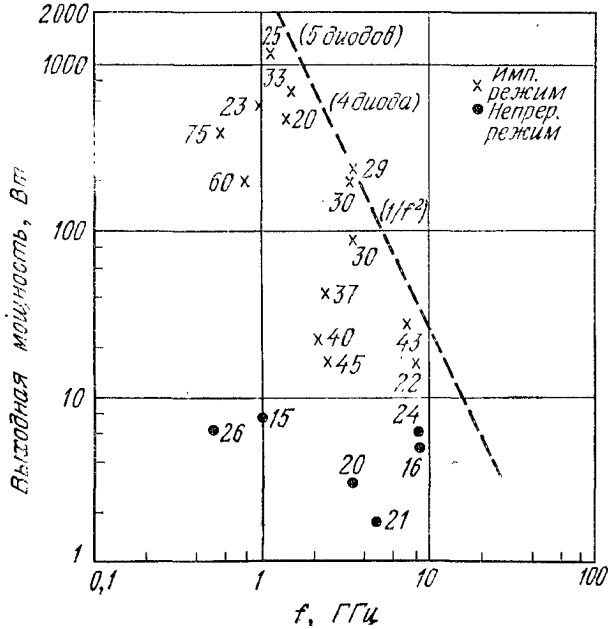


Рис. 51. Характеристики современных диодов с захваченным объемным зарядом лавины.

том перекрытия F (рис. 49, а). Чем больше коэффициент перекрытия F , тем больше прибор напоминает $p-i-n$ -диод. Зависимости ширины обедненной области и концентрации примеси от рабочей частоты для трех значений коэффициента перекрытия приведены на рис. 49, б.

Характеристики наиболее мощных современных диодов с зашваченным объемным зарядом приведены на рис. 50 и 51 [88, 8, 95, 96]. Выходная мощность пяти соединенных последовательно диодов в импульсном режиме генерации достигает 1,2 кВт при к. п. д. 25 %. Отметим, что верхний предел рабочих частот близок к 10 ГГц, а наибольшие выходные мощности получены в диодах, работающих в импульсном режиме. Такие высокие значения мощности, генерируемой в импульсном режиме, существенно превышают выходные мощности других полупроводниковых СВЧ-приборов. Однако большая чувствительность к малым изменениям параметров цепи, рабочих условий или температуры затрудняет создание приборов со стабильными характеристиками при работе в изменяющихся внешних условиях.

ЛИТЕРАТУРА

1. Johnson R. L., DeLoach B. C., Jr., Cohen B. G. A Silicon Diode Oscillator, *Bell Syst. Tech. J.*, **44**, 369 (1965).
2. DeLoach B. C., Jr. The IMPATT Story, *IEEE Trans. Electron Devices*, **ED-23**, 57 (1976).
3. Shockley W. Negative Resistance Arising from Transit Time in Semiconductor Diodes, *Bell Syst. Tech. J.*, **33**, 799 (1954).
4. Read W. T. A Proposed High-Frequency Negative Resistance Diode, *Bell Syst. Tech. J.*, **37**, 401 (1958).
5. Lee C. A., Batdorf R. L., Wiegman W., Kaminsky G. The Read Diode and Avalanche, Transit-Time, Negative-Resistance Oscillator, *Appl. Phys. Lett.*, **6**, 89 (1965).
6. Misawa T. Negative Resistance on $p-n$ Junction under Avalanche Breakdown Conditions, Parts I and II, *IEEE Trans. Electron Devices*, **ED-13**, 137 (1966).
7. Gildep M., Hines M. F. Electronic Tuning Effects in the Read Microwave Avalanche Diode, *IEEE Trans. Electron Devices*, **ED-13**, 169 (1966).
8. Sze S. M., Ryder R. M. Microwave Avalanche Diodes, *Proc. IEEE*, **59**, 1140 (1971).
9. Schroeder W. E., Haddad G. I. Nonlinear Properties of IMPATT Devices, *Proc. IEEE*, **61**, 153 (1973).
10. Special Issue on Solid-State Microwave Millimeter-Wave Power Generation, Amplification, and Control, *IEEE Trans. Microwave Theory Tech.*, **MTT-27** (May 1979).
- 10a. Eastman L. F. Microwave Semiconductor Devices: State-of-the-Art and Limiting Effects, *IEEE Tech. Dig., Int. Electron Devices Meet.*, 1978, p. 364.
11. Brackett C. A. The Elimination of Tuning Induced Burnout and Bias Circuit Oscillation in IMPATT Oscillators, *Bell Syst. Tech. J.*, **52**, 271 (1973).
12. Coleman D. J., Jr., Sze S. M. The Baritt Diode — A New Low Noise Microwave Oscillator, *IEEE Device Res. Conf.*, Ann Arbor, Mich., June 28, 1971; A Low-Noise Metal — Semiconductor — Metal (MSM) Microwave Oscillator, *Bell. Syst. Tech. J.*, **50**, 1695 (1971).

13. Ruegg H. W. A Proposed Punch-Through Microwave Negative Resistance Diode, *IEEE Trans. Electron Devices*, **ED-15**, 577 (1968).
14. Wright G. T. Punch-Through Transit-Time Oscillator, *Electron. Lett.*, **4**, 543 (1968).
15. Sobol H., Sterzer F. Microwave Power Sources, *IEEE Spectrum*, **9**, 20 (1972).
16. East J. R., Nguyen-Ba H., Haddad G. I. Microwave and mm Wave Baritt Doppler Detector *Microwave J.*, **19**, 51 (1976).
17. Prager H. J., Chang K. K. N., Weisbrod S. High Power, High Efficiency Silicon Avalanche Diodes at Ultrahigh Frequencies, *Proc. IEEE*, **55**, 586 (1967). [Имеется перевод: Прагер, Чанг, Вайсброд. Мощные кремниевые лавинные диоды с высоким к п. д. для ультравысоких частот. — ТИИЭР, 1967, 55, № 4, 119—121.]
18. Scharfetter D. L., Bartelink D. J., Gummel H. K., Johnston R. L. Computer Simulation of Low-Frequency High-Efficiency Oscillation in Germanium IMPATT Diodes *IEEE Trans. Electron Devices*, **ED-15**, 691 (1968).
19. Bartelink D. J., Scharfetter D. L. Avalanche Shock Fronts in $p - n$ Junctions, *Appl. Phys. Lett.*, **14**, 420 (1969).
20. Cloutiere A. S., Koike R. J., Napoli L. S. A Theory for the High-Efficiency Mode of Oscillation in Avalanche Diodes, *RCA Rev.*, **30**, 397 (1969).
21. Chang K. K. N., Kawamoto H., Prager H. J., Reynolds J. F., Rosen A., Mikenas V. A. TRAPATT Amplifiers for Phased-Array Radar Systems, *Microwave J.*, **16**, 27 (1973).
22. Salmer G., Pribelich H., Farrayre A., Kramer B. Theoretical and Experimental Study of GaAs Impact Oscillator Efficiency, *J. Appl. Phys.*, **44**, 314 (1973).
23. Sze S. M., Gibbons G. Avalanche Breakdown Voltages of Abrupt and Linearly Graded $p - n$ Junctions in Ge, Si, GaAs, and GaP, *Appl. Phys. Lett.*, **8**, 111 (1966).
24. Schroeder W. E., Haddad G. I. Avalanche Region Width in Various Structures of IMPATT Diodes, *Proc. IEEE*, **59**, 1245 (1971). [Имеется перевод: Шредер, Хаддад, Ширина слоя умножения в IMPATT-диодах различных структур. — ТИИЭР, 1971, 59, № 8, 129—132.]
25. Gibbons G., Sze S. M. Avalanche Breakdown in Read and $p - i - n$ Diodes, *Solid State Electron.*, **11**, 225 (1968).
26. Crowell C. R., Sze S. M. Temperature Dependence of Avalanche Multiplication in Semiconductors, *Appl. Phys. Lett.*, **9**, 242 (1966).
27. Bowers H. C. Space-Charge-Limited Negative Resistance in Avalanche Diodes, *IEEE Trans. Electron Devices*, **ED-15**, 343 (1968).
28. Sze S. M., Shockley W. Unit-Cube Expression for Space-Charge Resistance, *Bell Syst. Tech. J.*, **46**, 837 (1967).
29. Weissglas P. Avalanche and Barrier Injection Devices, in Howes M. J., Morgan D. V., Eds., *Microwave Devices — Device Circuit Interactions*, Wiley, N. Y., 1976, Chap. 3.
30. Misawa T. Multiple Uniform Layer Approximation in Analysis of Negative Resistance in $p - n$ Junctions in Breakdown, *IEEE Trans. Electron Devices*, **ED-14**, 795 (1967).
31. Scharfetter D. L., Gummel H. K. Large-Signal Analysis of a Silicon Read Diode Oscillator, *IEEE Trans. Electron Devices*, **ED-16**, 64 (1969).
32. Swan C. B., Misawa T., Marinaccio L. Composite Avalanche Diode Structures for Increased Power Capability, *IEEE Trans. Electron Devices*, **ED-14**, 684 (1967).
33. Scharfetter D. L. Power-Impedance-Frequency Limitation of IMPATT Oscillators Calculated from a Scaling Approximation, *IEEE Trans. Electron Devices*, **ED-18**, 536 (1971).
34. Thim H. W., Poetze H. W. Search for Higher Frequencies in Microwave Semiconductor Devices, 6th Eur. Solid State Device Res. Conf., *Inst. Phys. Conf. Ser.*, **32**, 73 (1977).

35. Seidel J. E., Niehaus W. C., Iglesias D. E. Double-Drift Silicon IMPATTs at X Band, *IEEE Trans. Electron Devices*, **ED-21**, 523 (1974).
36. Kuvas R. L. Carrier Transport in the Drift Region of Read-Type Diodes, *IEEE Trans. Electron Devices*, **ED-25**, 660 (1978).
37. Hirachi Y., Kobayashi K., Ogasawara K., Toyama Y. A New Concept for High Efficiency Operation of Hi — Lo-Type GaAs IMPATT Diodes, *IEEE Trans. Electron Devices*, **ED-25**, 666 (1978).
38. Blakey P. A., Culshaw B., Giblin R. A. Comprehensive Models for the Analysis of High Efficiency GaAs IMPATTs, *IEEE Trans. Electron Devices*, **ED-25**, 674 (1978).
39. Nishitani K., Sawano H., Ishihara O., Ishii T., Mitsui S. Optimum Design for High-Power and High Efficiency GaAs Hi — Lo IMPATT Diodes, *IEEE Trans. Electron Devices*, **ED-26**, 210 (1979).
40. Evans W. J. Avalanche Diode Oscillators, in Hershberger W. D., Ed. *Solid State and Quantum Electronics*, Wiley, N. Y., 1971.
41. Misawa T. Saturation Current and Large Signal Operation of a Read Diode, *Solid State Electron.*, **13**, 1363 (1970).
42. Kovel S. R., Gibbons G. The Effect of Unswep Epitaxial Material on the Microwave Efficiency of IMPATT Diodes, *Proc. IEEE Lett.*, **55**, 2066 (1967).
43. Aono Y., Okuto Y. Effect of Undepleted High Resistivity Region on Microwave Efficiency of GaAs IMPATT Diodes, *Proc. IEEE*, **63**, 724 (1975). [Имеется перевод: Аоно, Окуто. Влияние нейтрального участка высокоомной области на к. п. д. IMPATT-диодов из арсенида галлия. — ТИИЭР, 1975, 63, № 3, 56—57.]
44. DeLoach B. C., Jr. Thin Skin IMPATTs, *IEEE Trans. Microwave Theory Tech.*, **MTT-18**, 72 (1970).
45. Misawa T. High Frequency Fall-Off of IMPATT Diode Efficiency, *Solid State Electron.*, **15**, 457 (1972).
46. Elta M. E., Haddad G. I. Mixed Tunneling and Avalanche Mechanisms in $p - n$ Junctions and Their Effects on Microwave Transit-Time Devices, *IEEE Trans. Electron Devices*, **ED-25**, 694 (1978).
47. Nishizawa J. I., Motoya K., Okuno Y. 200 GHz TUNNETT Diodes, *Jpn. J. Appl. Phys.*, **17**, Suppl. 17—1, 167 (1977).
48. Berenz J. J., Kinoshita J., Hierl T. L., Lee C. A. Orientation Dependence of n -type GaAs Intrinsic Avalanche Response Time, *Electron. Lett.*, **15**, 150 (1979).
49. Olson H. M. A Mechanism for Catastrophic Failure of Avalanche Diodes, *IEEE Trans. Electron Devices*, **ED-22**, 842 (1975).
50. Hines M. F. Noise Theory of Read Type Avalanche Diode, *IEEE Trans. Electron Devices*, **ED-13**, 158 (1966).
51. Gummel H. K., Blue J. L. A Small-Signal Theory of Avalanche Noise of IMPATT Diodes, *IEEE Trans. Electron Devices*, **ED-14**, 569 (1967).
52. Blue J. L. Preliminary Theoretical Results on Low Noise GaAs IMPATT Diodes, *IEEE Device Res. Conf.*, Seattle, Wash., June 1970.
53. Irvin J. C., Coleman D. J., Johnson W. A., Tatsuguchi I., Decker D. R., Dunn C. N. Fabrication and Noise Performance of High-Power GaAs IMPATTs, *Proc. IEEE*, **59**, 1212 (1971). [Имеется перевод: Ирвин, Коулмен, Джонсон и др. Изготовление и шумовые свойства мощных IMPATT-диодов из арсенида галлия. — ТИИЭР, 1971, 59, № 8, 95—99.]
54. Misawa T. Microwave Si Avalanche Diode With Nearly-Abrupt-Type Junction, *IEEE Trans. Electron Devices*, **ED-14**, 580 (1967).
55. Sze S. M., Lepselter M. P., MacDonald R. W. Metal — Semiconductor IMPATT Diode, *Solid State Electron.*, **12**, 107 (1969).
56. De Nobel D., Kock H. G. A Silicon Schottky Barrier Avalanche Transit Time Diode, *Proc. IEEE*, **57**, 2088 (1969). [Имеется перевод: де Нобел, Кок. Кремниевый лавинно-пролетный диод с барьером Шоттки. — ТИИЭР, 1969, 57, № 10, 200.]

57. Kim C. K., Armstrong L. D. GaAs Schottky-Barrier Avalanche Diode. *Solid State Electron.*, **13**, 53 (1970).
58. Misawa T. Minority Carrier Storage and Oscillation Efficiency in Read Diodes, *Solid State Electron.*, **13**, 1369 (1970)
59. Niehaus W. C., Schwartz B. A Self-Limiting Anodic Etch-to-Voltage Technique for Fabrication of Modified Read IMPATTs, *Solid State Electron.*, **19**, 175 (1976).
60. Mahoney G. E. Retardation of IMPATT Diode Aging by Use of Tungsten in the Electrodes, *Appl. Phys. Lett.*, **27**, 613 (1975).
61. Heaton J. L., Walline R. E., Carroil J. F. Low — High — Low Profile GaAs IMPATT Reliability, *IEEE Trans. Electron Devices* ED-26, 96 (1979).
62. Scharfetter D. L., Evans W. J., Johnston R. L. Double-Drift-Region (p^+pnn^+) Avalanche Diode Oscillators, *Proc. IEEE*, **58**, 1131 (1970). [Имеется перевод: Шарфеттер, Эванс, Джонстон ГЛПД на диодах с двумя пролетными пространствами (структуры $p^+ - p - n - n^+$). — ТИИЭР 1970, 58 № 8, 113—114.]
63. Cho A. Y., Dunn C. N., Kuvus R. L., Schroeder W. E. GaAs IMPATT Diodes, Prepared by Molecular Beam Epitaxy, *Appl. Phys. Lett.*, **25**, 224 (1974).
64. Lilo S. Y. Microwave Devices and Circuits, Prentice-Hall, Englewood Cliffs, N. J., 1980.
65. Iglesias D. E. Circuit for Testing High Efficiency IMPATT Diodes, *Proc. IEEE*, **55**, 2065 (1967). [Имеется перевод: Игlesiас. Схема для проверки лавинно-пролетных диодов с большим к. п. д. — ТИИЭР, 1967, 55, № 11, 319—320.]
66. Misawa T., Marinaccio L. P. A 1/4 Watt Si $p - v - n$ X-Band IMPATT Diode, *Int. Electron Device Meet.*, Washington, D. C., Oct. 1966.
67. Bowman L. S., Burrus C. A., Jr., Pulse-Driven Silicon $p - n$ Junction Avalanche Oscillators for the 0,9 to 20 mm Band *IEEE Trans. Electron Devices*, ED-14, 411 (1967).
68. Ino M., Ishibashi T., Ohmori M. Submillimeter Wave Si p^+pn^+ IMPATT Diodes, *Jpn. J. Appl. Phys.*, **16** Suppl. 16-1, 89 (1977).
69. Pribetich J., Chive M., Constant E., Farrayre A. Design and Performance, of Maximum-Efficiency Single and Double-Drift-Region GaAs IMPATT Diodes in the 3—18 HCz Frequency Range, *J. Appl. Phys.*, **49**, 5584 (1978).
70. Seidel T. E., Davls R. E., Iglesias D. E. Double-Drift-Region Ion-Implanted Millimeter-Wave IMPATT Diodes, *Proc. IEEE*, **59**, 1222 (1971). [Имеется перевод: Зейдел, Дэвис, Игlesiас. Двухпролетные IMPATT-диоды миллиметрового диапазона, изготовленные методом ионного внедрения. — ТИИЭР, 1971, 59, № 8, 107—114.]
71. Cohen E. D. Trapatts and Impatts — State of the Art and Application, *Micro-wave J.*, **20**, 22 (1977).
72. Grey K. W. Recent Advances in Microwave Devices, *Jpn. J. Appl. Phys.*, **16**, Suppl. 16-1, 81 (1974).
73. Bozler C. O., Donnelly J. P., Murphy R. A., Laton R. W., Subduny R. W. High-Efficiency Ion-Implanted Lo — Hi — Lo GaAs IMPATT Diodes, *Appl. Phys. Lett.*, **29**, 125 (1976).
74. Hierl T. L., Berenz J. J., Kinoshita J., Zubeck I. U. High Efficiency Pulsed GaAs Read IMPATT Diodes, *Electron. Lett.*, **14**, 155 (1978).
75. Potteliger D. C., private communication.
76. Sze S. M., Coleman D. J., Loya A. Current Transport in Metal — Semiconductor — Metal (MSM) Structures, *Solid State Electron.*, **14**, 1209 (1971).
77. Chu J. L., Persky G., Sze S. M. Thermionic Injection and Space-Charge-Limited Current in Reach-Through p^+np^+ Structures, *J. Appl. Phys.*, **43**, 3510 (1972).
78. Eknoyan O., Yang E. S., Sze S. M. Multilayered Ion-Implanted Baritt Diodes with Improved Efficiency, *Solid State Electron*, **20**, 291 (1977).

79. De Cogan D. The Punch-Through Diode, *Microelectronics*, **8**, 20 (1976).
80. Nguyen-Ba H., Haddad G. I. Effects of Doping Profile on the Performance of Baritt Devices, *IEEE Trans. Electron Devices*, ED-24, 1154 (1977).
81. Chu J. L., Sze S. M. Microwave Oscillation in *pnp* Reach-Through Baritt Diodes, *Solid State Electron.*, **16**, 85 (1973).
82. Sjolund A. Small-Signal Analysis of Punch-Through Injection Microwave Devices, *Solid State Electron.*, **16**, 559 (1973).
83. Kwok S. P., Haddad G. I. Power Limitation in Baritt Devices, *Solid State Electron.*, **19**, 795 (1976).
84. Eknayan O., Sze S. M., Yang E. S. Microwave Baritt Diode with Retarding Field — An Investigation, *Solid State Electron.*, **20**, 285 (1977).
85. Ahmad S., Freyer J. High-Power Pt Schottky Baritt Diodes, *Electron. Lett.*, **12**, 238 (1976).
86. Sitch J. E., Majerfeld A., Robson P. N., Hasegawa F. Transit — Time — Induced Microwave Negative Resistance in GaAlAs — GaAs Heterostructure Diodes, *Electron. Lett.*, **11**, 457 (1975).
87. Sitch J. E., Robson P. N. Efficiency of BARITT and DOVETT Oscillators, *Solid State Electron Devices*, **1**, 31 (1976).
88. Liu S. G., Risko J. J. Fabrication and Performance of Kilowatt L-Band Avalanche Diodes, *RCA Rev.*, **31**, 3 (1970).
89. Kostichack D. F. UHF Avalanche Diode Oscillator Providing 400 Watts Peak Power and 75 Percent Efficiency, *Proc. IEEE Let.*, **58**, 1282 (1970).
90. DeLoach B. C., Scharfetter Jr. and D. L. Device Physics of TRAPATT Oscillators, *IEEE Trans. Electron Devices*, ED-17, 9 (1970).
91. Johnston R. L., Scharfetter D. L., Bartelink D. J. High-Efficiency Oscillations in Germanium Avalanche Diodes Below the Transit-Time Frequency, *Proc. IEEE*, **56**, 1611 (1968). [Имеется перевод: Джонстон, Шарфеттер, Бартелинк. Генерация колебаний германиевыми лавинно-пролетными диодами с высоким к. п. д. на частотах ниже номинальной. — ТИИЭР, 1968, № 9, 222—225.]
92. Evans W. J. Circuits for High-Efficiency Avalanche-Diode Oscillators, *IEEE Trans. Microwave Theory Tech.*, MTT-17, 1060 (1969).
93. Scharfetter D. L. Power-Frequency Characteristics of the PRAPATT Diode Mode of High Efficiency Power Generation in Germanium and Silicon Avalanche Diodes, *Bell Syst. Tech. J.*, **49**, 799 (1970).
94. Trew R. J., Haddad G. I., Masnari N. A. A Simplified Model of a TRAPATT Diode, *IEEE Trans. Electron Devices*, ED-23 28 (1976).
95. Obah C. O. G., Benko E., Midford T. A., Bowers H. C., Chao P. Y. Single-Diode 0.5 kW TRAPATT Oscillators *Electron. Lett.*, **10**, 430 (1974).
96. Ying R. S., Foug T. T. C-Band Complementary TRAPATT Diodes, *Proc. IEEE*, **62**, 287 (1974). [Имеется перевод: Ин, Фын. TRAPATT-диоды С-диапазона с дополнительной структурой. — ТИИЭР, 1974, № 3, 147—148.]
97. Ahmad S., Freyer J. Design and Development of High-Power Microwave Silicon BARITT Diodes, *IEEE Trans. Electron Devices*, ED-26, 1370 (1979).

Глава 11

ПРИБОРЫ НА ЭФФЕКТЕ МЕЖДОЛИННОГО ПЕРЕХОДА ЭЛЕКТРОНОВ

11.1. ВВЕДЕНИЕ

Приборы на эффекте междолинного перехода электронов¹⁾ (одни из важнейших СВЧ-приборов) широко используются в качестве гетеродина и усилителя мощности в диапазоне частот 1—100 ГГц. Твердотельные генераторы на этих приборах применяются в радарх, системах раннего оповещения и контрольно-измерительной аппаратуре.

В 1963 г. Ганн [1] обнаружил генерацию когерентных СВЧ-колебаний при наложении на произвольно ориентированный короткий образец из арсенида галлия или фосфида индия n -типа постоянного электрического поля, превышающего пороговое значение, равное нескольким киловольтам на 1 см. Частота колебаний приблизительно равнялась величине, обратной времени пролета носителей через образец. Кремер [2] указал на то, что все наблюдаемые свойства этого процесса согласуются с результатами теории отрицательного дифференциального удельного сопротивления, предложенной независимо Ридли и Уоткинсом [3, 3а], а также Хилсумом [4, 4а]. Причиной возникновения отрицательного дифференциального удельного сопротивления служит стимулированный электрическим полем переход электронов зоны проводимости из низкоэнергетической долины, в которой их подвижность относительно велика, в более высокоэнергетические долины с меньшими подвижностями. Хатсон и др. [5] обнаружили влияние давления на характеристики прибора из арсенида галлия, а Аллен и др. [6], исследуя генерацию в тройных соединениях $\text{GaAs}_{1-x}\text{P}_x$, показали, что величина порогового электрического поля уменьшается при уменьшении энергетического зазора между основным и побочным минимумами зоны проводимости. Все это убедительно доказывало, что причиной ганновской генерации является эффект междолинного перехода электронов.

Этот эффект также называют механизмом Ридли—Уоткинса—Хилсума или эффектом Ганна. Приборы на эффекте междолинного перехода электронов подробно описаны в работах [7—9]. В работе [10] дан обзор твердотельных СВЧ-генераторов.

¹⁾ В зарубежной литературе эти приборы называются TED (Transferred-Electron Devices). — *Прим. перев.*

11.2. ЭФФЕКТ МЕЖДОЛИННОГО ПЕРЕХОДА ЭЛЕКТРОНОВ

11.2.1. Объемное отрицательное дифференциальное удельное сопротивление

Физические механизмы возникновения объемного отрицательного дифференциального удельного сопротивления весьма различны. Одним из важнейших служит эффект междолинного перехода электронов. Поскольку случайная флуктуация концентрации свободных носителей в любой точке образца с объемным отрицательным удельным сопротивлением приводит к моментальному возникновению пространственного заряда, величина которого возрастает со временем по экспоненциальному закону, такой полупроводник является по своей природе нестабильным. Запишем уравнение непрерывности в одномерном случае ¹⁾

$$\frac{\partial n}{\partial t} + \frac{1}{q} \frac{\partial J}{\partial x} = 0. \quad (1)$$

Малое локальное отклонение концентрации основных носителей от однородной равновесной n_0 приводит к появлению пространственного заряда плотностью $n - n_0$. Уравнение Пуассона и выражение для плотности тока имеют вид

$$\frac{\partial \mathcal{E}}{\partial x} = \frac{q(n - n_0)}{\epsilon_s}, \quad (2)$$

$$J = \mathcal{E}/\rho - qD \frac{\partial n}{\partial x}, \quad (3)$$

где ϵ_s — диэлектрическая проницаемость, ρ — удельное сопротивление и D — коэффициент диффузии. Продифференцировав выражение (3) по x и учитывая уравнение Пуассона, получим

$$\frac{1}{q} \frac{\partial J}{\partial x} = \frac{n - n_0}{\rho \epsilon_s} - D \frac{\partial^2 n}{\partial x^2}. \quad (4)$$

Подстановка этого выражения в уравнение (1) приводит к уравнению

$$\frac{\partial n}{\partial t} + \frac{n - n_0}{\rho \epsilon_s} - D \frac{\partial^2 n}{\partial x^2} = 0. \quad (5)$$

Уравнение (5) можно решить методом разделения переменных. Если концентрация зависит только от координаты, то решение уравнения (5) имеет вид

$$n - n_0 = (n - n_0)_{x=0} \exp(-x/L_D), \quad (6)$$

¹⁾ Во избежание появления многочисленных отрицательных знаков мы будем считать в этой главе заряд электронов положительным.

где L_D — дебаевская длина, равная

$$L_D \equiv \sqrt{\frac{kT\epsilon_s}{q^2 n_0}} \quad (7)$$

и определяющая расстояние, на котором затухают малые флуктуации заряда. Если концентрация зависит только от времени, то решение уравнения (5) имеет вид

$$n - n_0 = (n - n_0)_{t=0} \exp(-t/\tau_R), \quad (8)$$

где τ_R — время диэлектрической релаксации:

$$\tau_R \equiv \rho\epsilon_s = \epsilon_s/q\mu n \approx \epsilon_s/q\mu n_0, \quad (9)$$

за которое происходит рассасывание пространственного заряда, если дифференциальное удельное сопротивление и дифференциальная подвижность μ положительны. Однако в полупроводнике с отрицательным дифференциальным сопротивлением (ОДС) любая флуктуация концентрации будет нарастать с постоянной времени, равной $|\tau_R|$, а не затухать.

Приборы с объемным ОДС можно разделить на две группы: приборы, управляемые напряжением (с N-образными характеристиками), и приборы, управляемые током (с S-образными характеристиками). Типичные зависимости плотности тока от напряженности электрического поля, характерные для этих двух типов приборов, приведены на рис. 1, а и б [11]. Также показано удельное дифференциальное сопротивление ($\partial\mathcal{E}/\partial J$) при изменении напряжения смещения от нуля до порогового, при котором начинается участок отрицательного сопротивления (рис. 1, в и г). В приборах с N-образными характеристиками зависимость напряженности электрического поля от плотности тока (рис. 1, а) неоднозначна, а в приборах с S-образными J — \mathcal{E} -характеристиками наоборот (рис. 1, б).

Характеристики на рис. 1 подобны приведенным в главах, посвященных контактным явлениям. Так, например, туннельный диод является прибором с N-образной J — \mathcal{E} -характеристикой, а тиристор — прибором с S-образной характеристикой. Объемное ОДС обусловлено микроскопическими свойствами полупроводника: стимулированным электрическим полем захватом носителей на глубокие уровни [12], ударной ионизацией мелких примесных уровней в компенсированных полупроводниках [13] и переходом электронов из основного минимума зоны проводимости в побочные (эффект междолинного перехода электронов). Мы рассмотрим основные характеристики приборов с ОДС независимо от конкретных физических механизмов возникновения отрицательного сопротивления.

В однородном полупроводнике с ОДС распределение электрического поля, соответствующее устойчивому состоянию, существ-

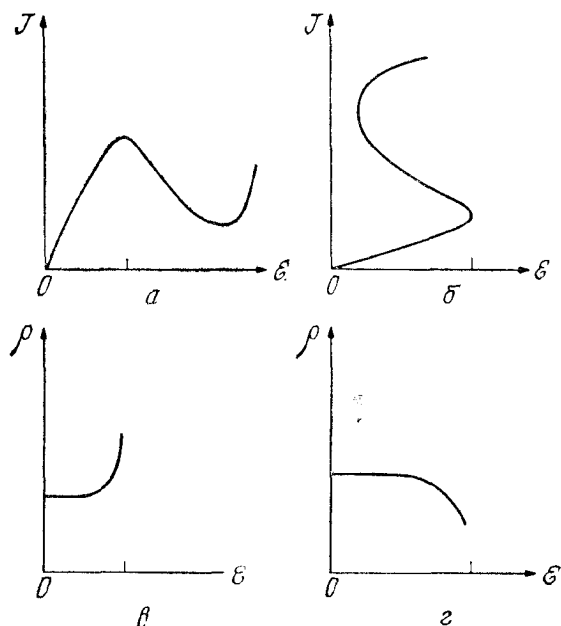


Рис. 1. Зависимости плотности тока и дифференциального удельного сопротивления от напряженности электрического поля. *a* и *v* — для приборов с N-образными $J - \mathcal{E}$ -характеристиками; *b* и *z* — для приборов с S-образными $J - \mathcal{E}$ -характеристиками.

венно неоднородно. Ниже мы приведем простые соображения, из которых следует, что в приборах с N-образными $J - \mathcal{E}$ -характеристиками будут образовываться домены высокого электрического поля (или обогащенные слои), а в приборах с S-образными характеристиками — шнуры тока с высокой плотностью [11]. При описании различных режимов работы приборов в последующих разделах мы лишь видоизменим нашу простую модель.

Как следует из рис. 1, *v*, положительное дифференциальное удельное сопротивление увеличивается с ростом электрического поля, т. е. $dp/d\mathcal{E} > 0$. Если напряженность электрического поля в некоторой области полупроводника немного больше средней (рис. 2, *a*), то ее удельное сопротивление также больше. Поэтому плотность тока будет меньше. Это приведет к увеличению размеров области и образованию домена (области высокого электрического поля), вне которого поле относительно мало. Поверхности, ограничивающие домен высокого поля, эквипотенциальны и поэтому перпендикулярны направлению электрического тока.

Для прибора с S-образной характеристикой первоначальная величина положительного дифференциального удельного сопро-

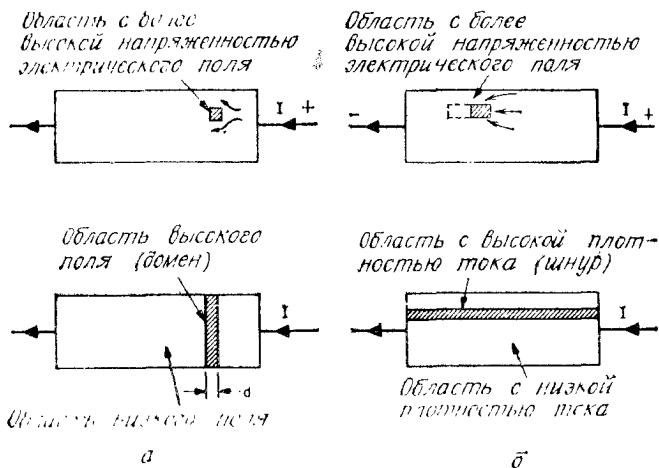


Рис. 2. Формирование домена высокого электрического поля (а) в образце с отрицательным дифференциальным удельным сопротивлением, управляемым напряжением (N-образной характеристикой), а также формирование шнура тока с высокой плотностью (б) в образце с отрицательным дифференциальным удельным сопротивлением, управляемым током (S-образной характеристикой).

тивления уменьшается при увеличении напряженности электрического поля, т. е. $d\rho/d\mathcal{E} < 0$. Если поле в некоторой области полупроводника немного больше среднего, то удельное сопротивление этой области меньше. Поэтому ток будет втекать в эту область, что приведет к ее удлинению в направлении электрического поля и образованию шнура тока с высокой плотностью.

Для более подробного изучения нестабильностей объемного заряда [14] в приборах с N-образными J — \mathcal{E} -характеристиками обратимся к рис. 3. На рис. 3, а приведена типичная мгновенная зависимость J от \mathcal{E} , а на рис. 3, б — схема прибора. Предположим, что в точке А образовался избыток отрицательного заряда (см. рис. 3, в). Причиной этого могут быть случайные шумы или неоднородное легирование (рис. 3, в). Проинтегрировав один раз уравнение Пуассона, мы получим распределение электрического поля в образце, приведенное на рис. 3, г, причем поле слева от точки А меньше чем справа. Если постоянное напряжение на приборе соответствует точке \mathcal{E}_A на J — \mathcal{E} -кривой, то втекающий в точку А ток будет больше вытекающего, что приведет к накоплению отрицательного пространственного заряда в точке А. При этом электрическое поле слева от точки А становится еще меньше, что способствует дальнейшему накоплению заряда. Этот процесс продолжается до тех пор, пока

напряженности электрического поля в области вне точки A не будут соответствовать точкам 1 и 2 J — \mathcal{E} -кривой на рис. 3, a , в которых токи равны, а дифференциальное удельное сопротивление положительно. В результате образуется движущийся слой, обогащенный носителями. Этот процесс, безусловно, зависит от числа электронов в кристалле, которое должно быть достаточ-

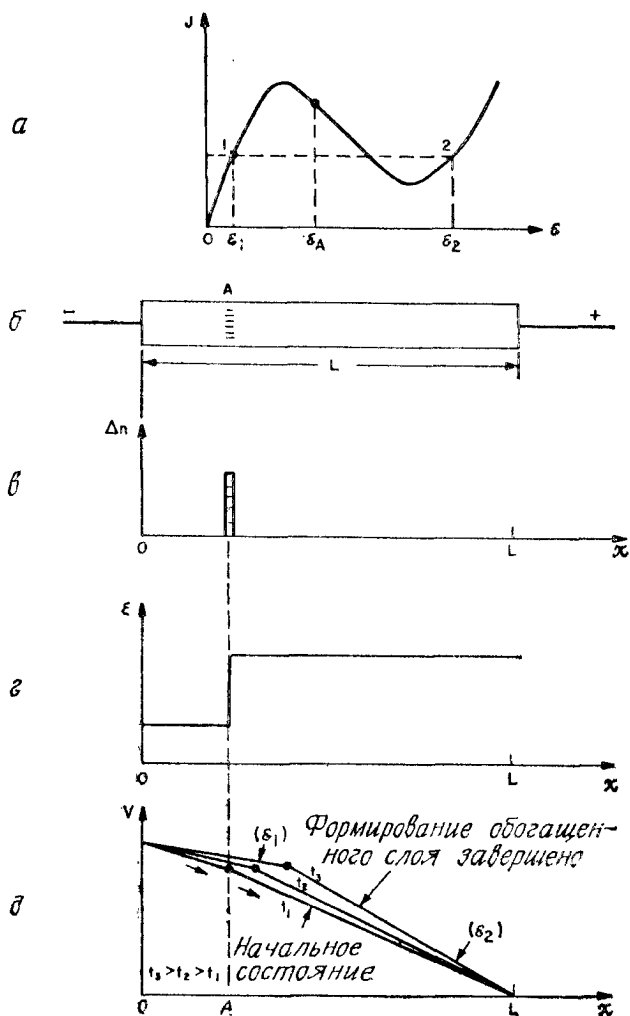


Рис. 3. Формирование обогащенного электронами слоя в возмущенной среде с ОДС [14].

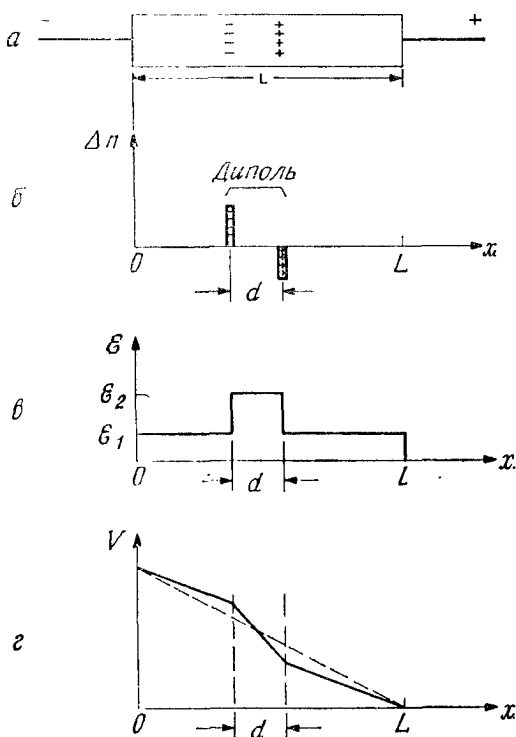


Рис. 4. Формирование электронного дипольного слоя в возмущенной среде с ОДС [14].

ным для формирования слоя пространственного заряда за время его пролета через образец [14].

Идеальный обогащенный слой, о котором до сих пор говорилось, представляет собой простейший пример неустойчивости объемного заряда. Если положительные и отрицательные заряды разделены небольшим пространственным промежутком (рис. 4), то мы имеем дело с образованием дипольного слоя (домена). Поле внутри домена больше, чем вне его (рис. 4, б). Из-за ОДС ток в области низкого электрического поля будет больше, чем внутри домена. Значения напряженности электрического поля вне и внутри домена будут стремиться к стационарным, соответствующим лежащим вне участка ОДС точкам зависимости J от \mathcal{E} , в которых электрические токи равны (рис. 3, а). (При этом мы не учитывали толщину доменных стенок.) Таким образом домен достигает стабильной конфигурации. Дипольный слой движется через кристалл и исчезает на аноде. В этот момент электрическое поле начинает возрастать, оставаясь однородным в полупроводнике, до тех пор, пока не превысит пороговое $\mathcal{E} \geq \mathcal{E}_T$ (рис. 5, а), что приведет к образованию нового домена и повторению процесса.

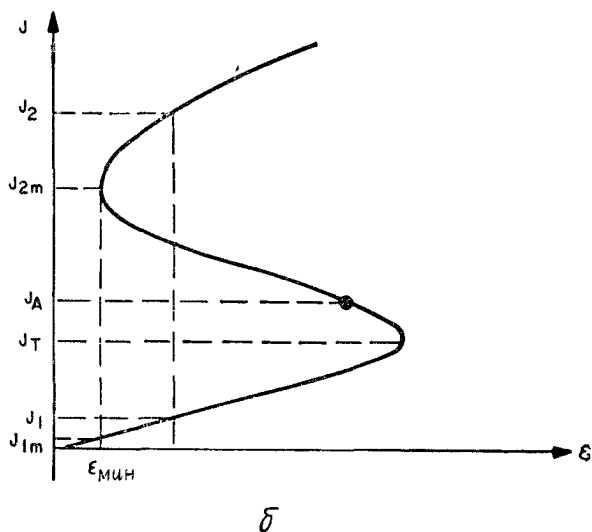
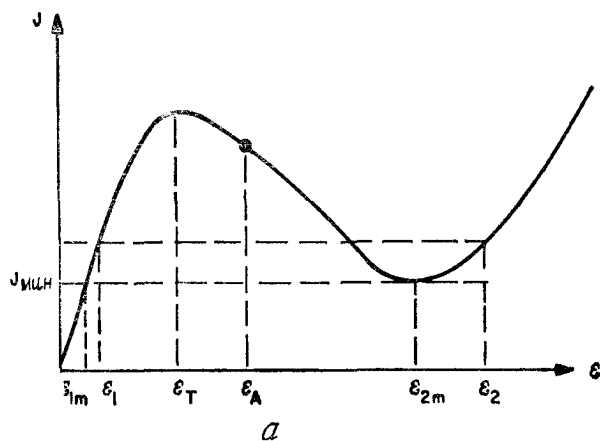


Рис. 5. Минимальная плотность тока и соответствующая напряженность электрического поля для приборов с N-образными $J - \mathcal{E}$ -характеристиками, управляемых напряжением (а), и для приборов с S-образными $J - \mathcal{E}$ -характеристиками, управляемых током (б) [11].

Для того чтобы оценить толщину дипольного слоя d , используем следующие соотношения:

$$V = \mathcal{E}_A L = \mathcal{E}_2 d + (L - d) \mathcal{E}_1 \quad (10)$$

и

$$d = L \left(\frac{\mathcal{E}_A - \mathcal{E}_1}{\mathcal{E}_2 - \mathcal{E}_1} \right), \quad (11)$$

где L — длина образца. Предположим, что наиболее стабильное состояние соответствует минимальной мощности, потребляемой прибором от источника питания. Это означает, что при постоянном напряжении ток должен быть минимальным. Как следует из рис. 5, a , толщина дипольного слоя d определяется выражением (11), если заменить в нем \mathcal{E}_1 и \mathcal{E}_2 на \mathcal{E}_{1m} и \mathcal{E}_{2m} соответственно.

Аналогичное рассмотрение можно провести и для прибора с S-образной J — \mathcal{E} -характеристикой. Вместо домена мы рассмотрим шнур с площадью поперечного сечения a . В стационарных условиях для заданной величины тока I (рис. 5, б) получим

$$I = J_A A = J_2 a + (A - a) J_1 \quad (12)$$

и

$$a = A \left(\frac{J_A - J_1}{J_2 - J_1} \right), \quad (13)$$

где A — площадь поперечного сечения прибора. Если потребляемая мощность минимальна, напряженне также должно быть минимальным, а площадь поперечного сечения шнура можно найти из выражения (13), заменив в нем J_2 и J_1 на J_{2m} и J_{1m} соответственно.

11.2.2. Междолинный переход электронов

Переход электронов из основного минимума зоны проводимости с относительно большой подвижностью в побочные более высокоэнергетические минимумы с меньшими подвижностями называется эффектом междолинного перехода электронов. Для того чтобы понять, как этот эффект приводит к появлению отрицательного дифференциального удельного сопротивления, рассмотрим зависимости энергии электрона от квазиимпульса в арсениде галлия и фосфиде индия (рис. 6) — двух наиболее важных для таких приборов полупроводниках [15, 16]. Видно, что зонные структуры этих полупроводников почти одинаковы. Зона проводимости состоит из нескольких подзон. Дно зоны проводимости соответствует $k = 0$ (точка Γ). Первая более высокоэнергетическая подзона расположена в направлении $\langle 111 \rangle$ (L), а следующая — вдоль оси $\langle 100 \rangle$ (X). Таким образом, подзоны в этих полупроводниках имеют следующую последовательность: Γ — L — X . Обычно предполагалось, что первая подзона в арсениде галлия отделена при комнатной температуре энергетическим зазором $\sim 0,36$ эВ от основного минимума (X -подзона). Однако в 1976 г. Аспнес [15] измерил электроотражение синхротронного излучения на барьере Шоттки и установил правильную последовательность подзон в арсениде галлия Γ — L — X (которая идентична фосфиду индия (рис. 6)).

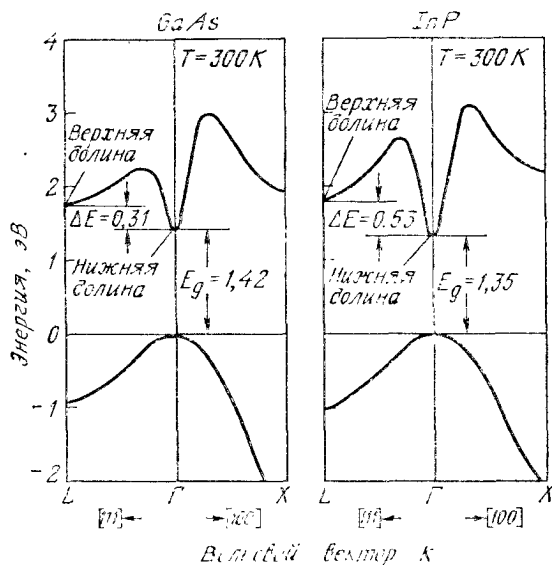


Рис. 6. Структура энергетических зон арсенида галлия и фосфида индия [15, 16].

Теперь мы получим приближенное соотношение между дрейфовой скоростью и напряженностью электрического поля, опираясь на предположение о равенстве электронных температур T_e в нижней (Γ) и верхней (L) долинах [4, 17]. Величина энергетического зазора между минимумами зоны проводимости $\Delta E \approx 0,31$ эВ для арсенида галлия и $\sim 0,53$ эВ для фосфида индия. Введем обозначения: m_1^* и m_2^* — эффективные массы, μ_1 и μ_2 — подвижности, n_1 и n_2 — концентрации электронов в нижней и верхней долинах соответственно, причем полная концентрация носителей заряда равна $n = n_1 + n_2$. Плотность стационарного тока в полупроводнике можно представить следующим образом:

$$J = q (\mu_1 n_1 + \mu_2 n_2) \mathcal{E} = q n v, \quad (14)$$

где v — средняя дрейфовая скорость и

$$v = \left(\frac{\mu_1 n_1 + \mu_2 n_2}{n_1 + n_2} \right) \mathcal{E} \approx \frac{\mu_1 \mathcal{E}}{1 + (n_2/n_1)}, \quad (15)$$

так как $\mu_1 \gg \mu_2$. Отношение заселенностей верхней и нижней долин, разделенных энергетическим зазором ΔE , равно

$$\frac{n_2}{n_1} = R \exp \left(-\frac{\Delta E}{kT} \right), \quad (16)$$

где R — отношение плотностей состояний и

$$R = \frac{\text{Число состояний в верхней долине}}{\text{Число состояний в нижней долине}} = \frac{M_2}{M_1} \left(\frac{m_2^*}{m_1^*} \right)^{3/2}, \quad (17)$$

а M_1 и M_2 — число верхних и нижних долин соответственно. Для арсенида галлия $M_1 = 1$, а число верхних долин вдоль оси L равно 8, но они расположены у края первой зоны Бриллюэна, и поэтому $M_2 = 4$. Используя значения эффективных масс электронов в арсениде галлия $m_1^* = 0,067m_0$ и $m_2^* = 0,55m_0$, получим $R = 94$.

Поскольку электрическое поле ускоряет электроны и увеличивает их кинетическую энергию, электронная температура T_e превышает температуру решетки T . Электронная температура определяется с помощью времени релаксации энергии

$$q\mathcal{E}v = \frac{3}{2} k (T_e - T)/\tau_e, \quad (18)$$

где время релаксации энергии τ_e предполагается равным $\sim 10^{-12}$ с. Подставив v из выражения (15) и n_2/n_1 из выражения (16) в выражение (18), получим

$$T_e = T + \frac{2q\tau_e\mu_1}{3k} \mathcal{E}^2 \left[1 + R \exp \left(-\frac{\Delta E}{kT_e} \right) \right]^{-1}. \quad (19)$$

Используя это равенство, можно рассчитать зависимость T_e от напряженности электрического поля при заданной величине T . Из выражений (15) и (16) следует соотношение между дрейфовой скоростью и полем:

$$v = \mu_1 \mathcal{E} \left[1 + R \exp \left(-\frac{\Delta E}{kT_e} \right) \right]^{-1}. \quad (20)$$

Рассчитанные с помощью выражений (19) и (20) типичные зависимости v от \mathcal{E} для арсенида галлия при трех температурах решетки приведены на рис. 7 (сплошные кривые). Показана также заселенность верхней долины как функция напряженности электрического поля (штриховая кривая). Из рассмотренной выше простой модели можно сделать следующие выводы: 1) существует определенное пороговое значение напряженности электрического поля \mathcal{E}_T , при котором возникает участок ОДС (или отрицательной дифференциальной подвижности); 2) пороговое значение напряженности электрического поля увеличивается с ростом температуры решетки; 3) участок отрицательной дифференциальной подвижности может отсутствовать, если температура решетки достаточно высока или энергетический зазор ΔE между минимумами зоны проводимости мал. Поэтому для того, чтобы междолинный переход электронов приводил к возникновению объемного ОДС, необходимо выполнение следующих условий: 1) температура решетки должна быть настолько малой, чтобы в отсут-

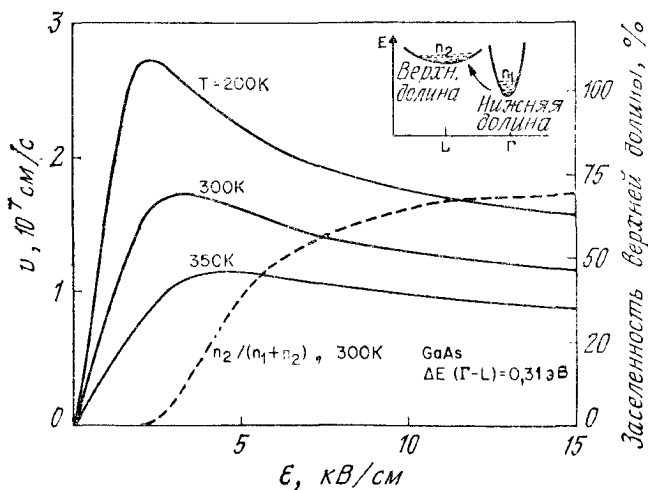


Рис. 7. Зависимости дрейфовой скорости от напряженности электрического поля в GaAs при трех температурах решетки (двухдолинная модель в предположении равенства электронных температур в обоих минимумах).

ствие поля большинство электронов находилось в основном минимуме зоны проводимости (т. е. $kT < \Delta E$); 2) в основном минимуме зоны проводимости электроны должны иметь высокую подвижность, малую эффективную массу и малую плотность состояний, в то время как в побочных минимумах электроны должны иметь низкую подвижность, большую эффективную массу, а плотность состояний должна быть высокой; 3) энергетический зазор между минимумами должен быть меньше ширины запрещенной зоны, чтобы лавинный пробой не произошел раньше междолинных переходов электронов.

Из всех полупроводников, удовлетворяющих этим условиям, арсенид галлия и фосфид индия n -типа наиболее исследованы и широко используются. Тем не менее эффект междолинного перехода электронов наблюдался во многих других полупроводниках, в том числе в германии, бинарных, тройных и четверных соединениях (табл. 1) [9, 18, 19]. Эффект междолинного перехода электронов наблюдался также в InAs и InSb при воздействии гидростатического давления, которое прикладывается для уменьшения энергетического зазора ΔE , при обычных условиях превышающего ширину запрещенной зоны. Особый интерес для возможного использования в маломощных, но быстродействующих приборах представляют тройные соединения элементов III—V групп GaInSb, так как в них пороговые поля малы, а дрейфовые скорости велики. В полупроводниках с большим энергетическим за-

Таблица 1. Полупроводниковые материалы, в которых наблюдается эффект междолинного перехода электронов при 300 К

Полупроводник	E_g , эВ	Энергетический зазор		\mathcal{E}_T , кВ/см	v_p , 10^7 см/с
		между минимумами	ΔE , эВ		
GaAs	1,42	$\Gamma-L$	0,31	3,2	2,2
InP	1,35	—	0,53	10,5	2,5
Ge ^{а)}	0,74	$L-\Gamma$	0,18	2,3	1,4
CdTe	1,50	$\Gamma-L$	0,51	11,0	1,5
InAs ^{б)}	0,36	$\Gamma-L$	1,28	1,6	3,6
InSb ^{в)}	0,28	$\Gamma-L$	0,41	0,6	5,0
ZnSe	2,60	$\Gamma-L$	—	38,0	1,5
Ga _{0,5} In _{0,5} Sb	0,36	$\Gamma-L$	0,36	0,6	2,5
Ga _{0,3} In _{0,7} Sb	0,24	$\Gamma-L$	—	0,6	2,9
InAs _{0,2} P _{0,8}	1,10	$\Gamma-L$	0,95	5,7	2,7
Ga _{0,13} In _{0,87} As _{0,37} P _{0,63}	1,05	—	—	5,5—8,6	1,2

а) При 77 К и ориентации $\langle 100 \rangle$ или $\langle 110 \rangle$.

б) Давление 14 кбар.

в) При 77 К и давлении 8 кбар.

зором между долинами (например, в $Al_{0,25}In_{0,75}As$ $\Delta E = 1,12$ эВ, а в $Ga_{0,6}In_{0,4}As$ $\Delta E = 0,72$ эВ) причиной возникновения ОДС могут быть электроны, находящиеся в основном (Γ) минимуме зоны проводимости [59]. Численные расчеты на ЭВМ по методу Монте-Карло показали, что в таких полупроводниках наличие побочных минимумов не является необходимым условием существования ОДС. Причиной появления максимума в зависимости дрейфовой скорости от поля и возникновения ОДС может служить эффект рассеяния электронов на полярных оптических фононах в непараболическом основном минимуме.

Экспериментальные зависимости дрейфовой скорости от напряженности электрического поля в арсениде галлия и фосфиде индия при комнатной температуре приведены на рис. 8 [16, 20]. Результаты теоретического анализа процессов переноса электронов в сильных электрических полях находятся в хорошем соответствии с экспериментальными данными [21, 22]. Пороговая величина напряженности электрического поля \mathcal{E}_T , определяющая начало участка ОДС, равна $\sim 3,2$ кВ/см для GaAs и 10,5 кВ/см для InP. Максимальная величина дрейфовой скорости v_p составляет $\sim 2,2 \cdot 10^7$ см/с для сверхчистых кристаллов арсенида галлия и $\sim 2,5 \cdot 10^7$ см/с для фосфида индия. Максимальная величина отрицательной дифференциальной подвижности приблизительно

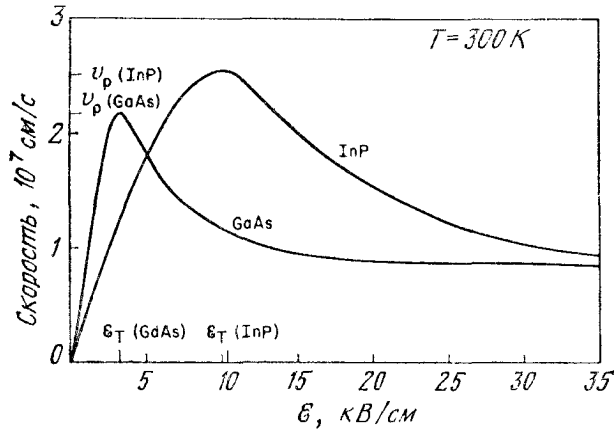


Рис. 8. Экспериментальные зависимости дрейфовой скорости от напряженности электрического поля в GaAs и InP [20, 16].

равна $\sim 2400 \text{ см}^2/\text{В} \cdot \text{с}$ для GaAs и $\sim 2000 \text{ см}^2/\text{В} \cdot \text{с}$ для InP. Экспериментальные зависимости нормированных величин напряженности порогового поля $\mathcal{E}_T/\mathcal{E}_T(300\text{K})$ и максимальной дрейфовой скорости $v_p/v_p(300\text{K})$ от температуры решетки арсенида галлия показаны на рис. 9 [23]. Полученные с помощью простой теоретической модели результаты (рис. 7) качественно согласуются с экспериментальными. Если известна зависимость дрейфовой скорости v от напряженности электрического поля \mathcal{E} , то равенство для плотности тока (3) можно записать в следующем виде:

$$J = qnv(\mathcal{E}) - qD \frac{dn}{dx}, \quad (21)$$

или

$$J = q(n_1\mu_1 + n_2\mu_2)\mathcal{E} - q\left(D_1 \frac{dn_1}{dx} + D_2 \frac{dn_2}{dx}\right), \quad (21a)$$

где первое слагаемое в правой части обусловлено дрейфом, а второе — диффузией носителей заряда, причем предполагается, что коэффициент диффузии не зависит от поля.

В стационарных условиях при малых градиентах концентрации носителей выражение (21) принимает вид $J = qnv(\mathcal{E})$. Подставив этот результат в уравнение Пуассона (уравнение (2)), получим

$$\frac{\partial \mathcal{E}}{\partial x} = -\frac{qn_0}{\epsilon_s} \left[1 - \frac{J/qn_0}{v(\mathcal{E})} \right]. \quad (22)$$

Граничное условие для этого нелинейного дифференциального уравнения первого порядка формулируется на основании утверждения о непрерывности напряженности электрического поля

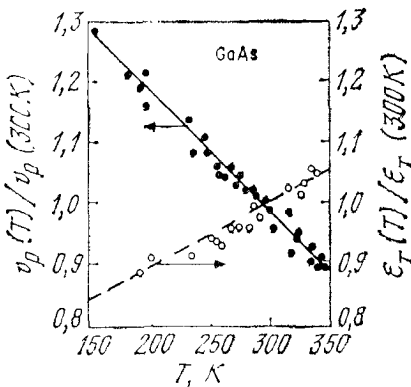


Рис. 9. Экспериментальные зависимости нормированных величин максимальной дрейфовой скорости и пороговой напряженности электрического поля от температуры в GaAs [23].

$\mathcal{E}(x)$ в полупроводнике. Результаты численного решения уравнения (22) для образца из арсенида галлия приведены на рис. 10, где $J_T = qn_0v_p$ [17]. Из рисунка видно, что в любой точке x напряженность электрического поля является монотонно возрастающей функцией плотности тока. Поэтому, если правильно сформулировать граничные условия, стационарное решение не дает отрицательного сопротивления. Это не удивительно, поскольку еще Шокли указал на то, что приборы с объемным ОДС должны

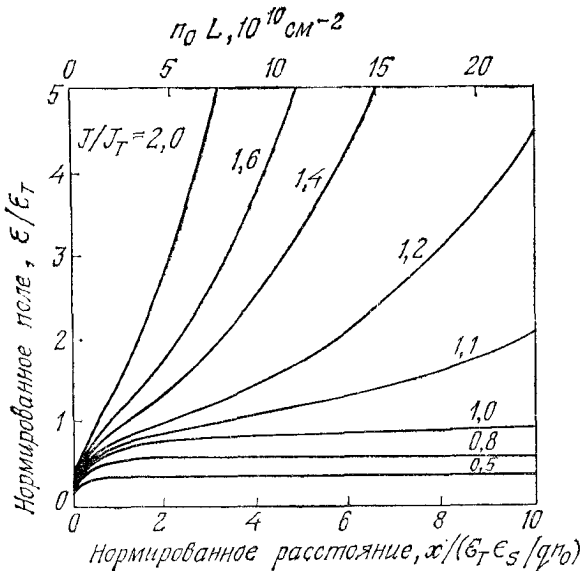


Рис. 10. Зависимости напряженности электрического поля от расстояния x ($\mathcal{E}(0) = 0$ при $x = 0$) [17].

быть стабильными на постоянном токе [24]. Действительно, в полупроводнике имеется внутренний пространственный заряд (избыточные электроны, инжектированные катодом), который возрастает при увеличении напряжения настолько, что ток также увеличивается, несмотря на уменьшение дрейфовой скорости электронов из-за отрицательной дифференциальной подвижности. Однако эти стационарные решения не обязательно являются устойчивыми по отношению к малым флюктуациям.

11.3. РЕЖИМЫ РАБОТЫ

С тех пор как в 1963 г. Ганн впервые наблюдал СВЧ-колебания в приборах на эффекте междолинного перехода электронов из арсенида галлия и фосфида индия, были изучены различные режимы работы этих приборов. Пять основных факторов определяют тот или иной режим ¹⁾: 1) величина концентрации примеси и однородность ее распределения в образце; 2) ширина активной области; 3) свойства катодного контакта; 4) тип используемой внешней цепи; 5) величина напряжения смещения на приборе.

Как уже говорилось выше, в нестационарном состоянии в приборах на эффекте междолинного перехода электронов формируются обогащенные слои, или домены. Процесс образования домена высокого поля зависит от числа носителей заряда в полупроводнике и длины прибора, которые должны быть достаточно большими, чтобы за пролетное время происходило формирование пространственного заряда требуемой величины. Исходя из этого, устанавливается критерий для того или иного режима работы прибора. Как следует из выражения (8), увеличение со временем пространственного заряда в приборе с ОДС описывается на начальной стадии следующей зависимостью:

$$(n - n_0) = (n - n_0)_{t=0} \exp(t / \tau_R),$$

где $|\tau_R| = e/qn_0 |\mu_-|$ и μ_- — отрицательная дифференциальная подвижность. Если это соотношение остается справедливым на протяжении времени пролета слоя пространственного заряда через образец, то максимальный фактор роста будет равен $\exp(L/v|\tau_R|)$, где v — средняя скорость дрейфа слоя. Для существенного увеличения заряда этот фактор должен превышать 1, и, следовательно, $L/v|\tau_R| > 1$ или

$$n_0 L > \epsilon_s v / q |\mu_-|. \quad (23)$$

¹⁾ В отдельных работах возникает ряд вопросов в связи с определением граничных условий на катоде. Поэтому представляется необходимым дальнейшее изучение этого вопроса с целью установления справедливости некоторых утверждений, которые включены в эту главу.

Для арсенида галлия и фосфида индия n -типа правая часть неравенства составляет $\sim 10^{12}$ см $^{-2}$. Если произведение $n_0L < 10^{12}$ см $^{-2}$, то распределение поля в таком приборе устойчиво (рис. 10). Поэтому важной границей, разделяющей режимы работы, является величина произведения концентрации носителей и длины прибора, равная $n_0L = 10^{12}$ см $^{-2}$. Ниже мы рассмотрим некоторые важные режимы работы, причем начнем с идеального режима однородного электрического поля.

11.3.1. Идеальный режим однородного электрического поля

В идеализированных условиях, когда в исходном образце отсутствует пространственный заряд и электрическое поле однородно, вольт-амперная характеристика прибора может быть получена путем пересчета зависимости скорости от поля. Простейшей формой импульсов напряжения является прямоугольная (рис. 11) [25]. Введем два безразмерных параметра $\alpha \equiv I_V/I_T$ и $\beta \equiv V_0/V_T$. Из предполагаемой формы импульсов следует, что средняя величина тока I_0 равна

$$I_0 = (1 + \alpha) I_T/2. \quad (24)$$

Мощность, потребляемая прибором от источника питания,

$$P_0 = V_0 I_0 = \frac{1}{2} \beta (1 + \alpha) V_T I_T. \quad (25)$$

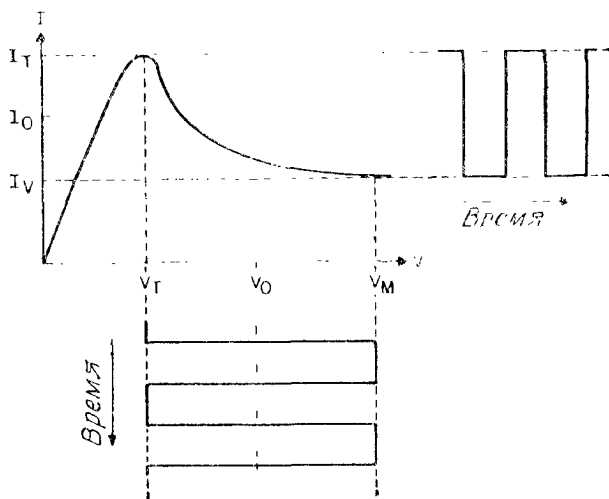


Рис. 11. Идеализированные прямоугольные формы импульсов тока и напряжения в режиме однородного электрического поля [25].

Полная величина СВЧ-мощности, выделяемой на нагрузке, равна

$$P_{rf} = \frac{V_M - V_T}{2} \frac{I_T - I_V}{2} = \frac{(\beta - 1)(1 - \alpha)}{2} V_1 I_T. \quad (26)$$

Поэтому к. п. д. прибора

$$\eta = \frac{(1 - \alpha)(\beta - 1)}{(1 + \alpha)\beta}, \quad (27)$$

а для основной частоты

$$\eta_1 = \frac{8}{\pi^2} \eta. \quad (27a)$$

Из выражения (27) следует, что наибольший к. п. д. получается при максимально возможных напряжении постоянного смещения ($\beta \rightarrow \infty$) и отношении $1/\alpha$. Теоретически достижимая величина к. п. д. прибора из арсенида галлия равна 30 % ($1/\alpha = 2,2$) и 45 % для прибора из фосфида индия ($1/\alpha = 3,5$). Эти значения не должны зависеть от частоты до тех пор, пока последние меньше величины, обратной времени релаксации энергии, и величины, обратной времени междолинного рассеяния.

В экспериментальных условиях такие высокие значения к. п. д. не были получены, а рабочая частота обычно равнялась пролетной. Это объясняется следующими причинами: 1) величина постоянного смещения ограничена напряжением лавинного пробоя; 2) формирование слоя пространственного заряда в обычных условиях приводит к неоднородному распределению поля в образце; 3) в резонансной цепи трудно получить идеальную форму импульсов тока и напряжения.

11.3.2. Режим с обогащенным слоем

Распределение электрического поля в коротком или слаболегированном образце ($n_0 L < 10^{12} \text{ см}^{-2}$) устойчиво, а его сопротивление по постоянному току положительно. На рис. 10 приведены стационарные распределения напряженности электрического поля в таком приборе при различных плотностях тока. Если к образцу внезапно прикладывается однородное электрическое поле, то динамику обогащенного слоя можно понять из упрощенных зависимостей, приведенных на рис. 12 [26]. В момент времени 1 обогащенный слой (т. е. избыточные электроны) инжектируется из катода, что приводит к «расщеплению» распределения электрического поля на два участка (момент времени 2). При этом скорости носителей вне обогащенного слоя изменяются, как показано на рис. 12, а. Поскольку напряжение на приборе предполагается постоянным, площади под кривыми на рис. 12, б должны быть равны. Из зависимости дрейфовой скорости от напряженности электрического поля следует, что по мере продвижения обога-

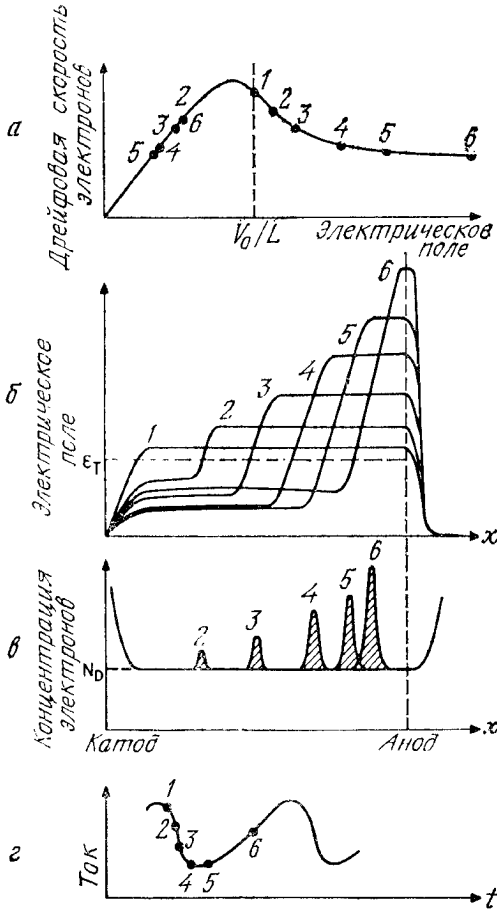


Рис. 12. Характеристики прибора в режиме работы с пролетом обогащенного слоя при постоянном напряжении на приборе [26].

щенного слоя к аноду это равенство может иметь место только при уменьшении скорости носителей вне слоя (моменты времени 3—5). В конце концов в момент времени 6 обогащенный слой достигает анода и исчезает. Поле вблизи катода возрастает до порогового, инжектируется другой обогащенный слой, и процесс повторяется. Как видно из рис. 12, г, зависимость тока от времени достаточно гладкая.

Если n_0L меньше критического значения ($n_0L < 10^{12}$ см⁻²), то сопротивление прибора может быть отрицательным на частотах, близких к пролетной частоте или ее гармоникам. Такой прибор может работать как усилитель [27]. На рис. 13 приведены экспериментальные частотные зависимости действительной и мнимой составляющих проводимости образца из арсенида галлия n -типа

с $n_0 = 3 \cdot 10^{13} \text{ см}^{-3}$ и $L = 70 \text{ мкм}$ ($n_0 L = 2,1 \cdot 10^{11} \text{ см}^{-2}$) [28]. Максимумы отрицательной действительной составляющей проводимости приблизительно соответствуют гармоникам пролетной частоты. Экспериментально полученная зависимость коэффициента усиления от частоты приведена на рис. 14 для прибора с такой же длиной, но меньшей концентрацией ($n_0 = 1,5 \cdot 10^{13} \text{ см}^{-3}$). При увеличении напряженности электрического поля пик коэффициента усиления сдвигается в область более высоких частот, поскольку растет максимум отрицательной действительной составляющей проводимости и сдвигается в ту же область.

Прибор будет генерировать сигнал в режиме с пролетом обогащенного слоя, если величина $n_0 L$ меньше критического значения, а прибор включен в параллельную резонансную цепь с сопротивлением нагрузки $\sim 10R_0$, где R_0 — сопротивление прибора при малых полях. На рис. 15 приведены распределения напряженности электрического поля в образце в четыре различных момента одного СВЧ-периода [29]. Показан также профиль легирования ($n_0 = 2 \cdot 10^{14} \text{ см}^{-3}$) и зависимость напряжения и тока от времени. Напряжение всегда превышает пороговое ($V > V_T = \mathcal{E}_T L$). Форма генерируемых колебаний далека от идеальной, а к. п. д. составляет лишь 5%. При последовательном соединении прибора с нагрузочным сопротивлением и катушкой индуктивности можно получить более приемлемую форму колебаний и к. п. д. $\sim 10\%$.

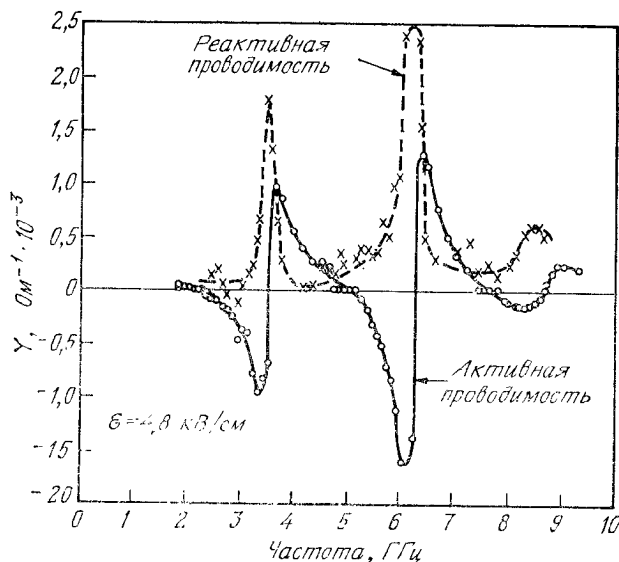


Рис. 13. Экспериментальная зависимость проводимости образца из арсенида галлия ($n_0 = 3 \cdot 10^{13} \text{ см}^{-3}$ и $L = 70 \text{ мкм}$) от частоты [28].

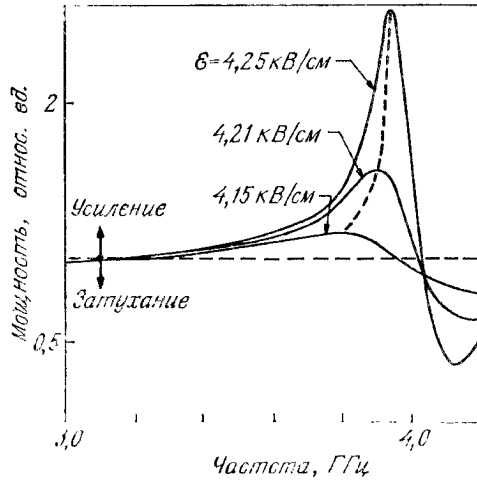


Рис. 14. Экспериментально полученный коэффициент усиления как функция частоты при различных напряжениях смещения на образце из арсенида галлия с $n_0 = 1,5 \cdot 10^{13} \text{ см}^{-3}$ и $L = 70 \text{ мкм}$ [28].

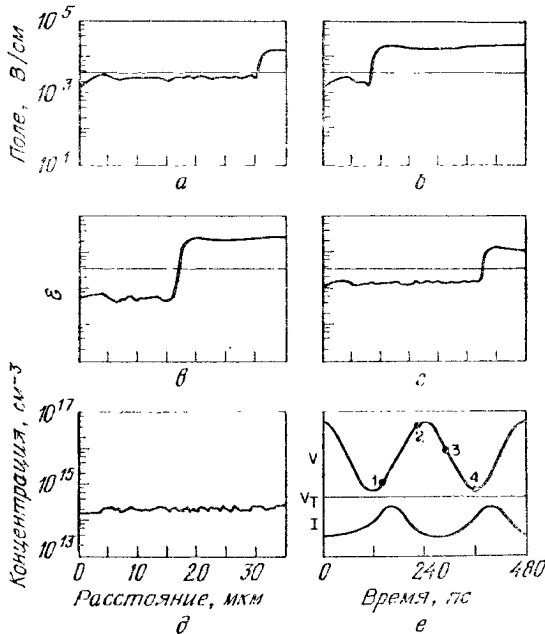


Рис. 15. Распределения электрического поля в четыре различных момента одного СВЧ-периода (а — г), профиль легирования примесью (б) и зависимости тока и напряжения от времени (е) прибора из GaAs при работе в резонансной цепи в режиме с обогащенным слоем для $fL = 1,4 \cdot 10^7 \text{ см/с}$ и $n_0/f = 5 \cdot 10^4 \text{ с/см}^3$ [29].

11.3.3. Режим пролета домена

Если произведение n_0L превышает 10^{12} см⁻², то флуктуации пространственного заряда в полупроводнике возрастают экспоненциально со временем и расстоянием и формируются домены, которые движутся к аноду. Обычно домен образуется вблизи катодного контакта, поскольку флуктуации концентрации примеси и объемного заряда здесь максимальны. Причиной экспериментально наблюдаемой ганновской генерации служат периодическое образование и последующее исчезновение на аноде доменов, которые успевают полностью сформироваться. В разд. 11.2 мы качественно рассмотрели неустойчивости пространственного заряда и формирование доменов. В этом разделе мы изучим диполи более подробно. Дипольные слои являются стабильными в том смысле, что движутся с определенной скоростью и не изменяются со временем. Мы будем считать, что скорость дрейфа электронов подчиняется статической зависимости, приведенной на рис. 16 (сплошная кривая). Система уравнений, описывающих поведение электронов, состоит из уравнения Пуассона (уравнение (2)) и уравнения для плотности полного тока

$$J = qnv(\mathcal{E}) - q \frac{dD(\mathcal{E})n}{dx} + \epsilon_s \frac{d\mathcal{E}}{dt}. \quad (28)$$

За исключением третьего слагаемого, которое равно току смещения, это уравнение идентично выражению (21).

Решение должно иметь вид домена высокого электрического поля, который без изменения формы движется со скоростью $v_{дом}$. Вне домена концентрация носителей и электрическое поле постоянны и равны $n = n_0$ и $\mathcal{E} = \mathcal{E}_0$, соответственно. Для решения такого вида напряженность электрического поля \mathcal{E} и концентра-

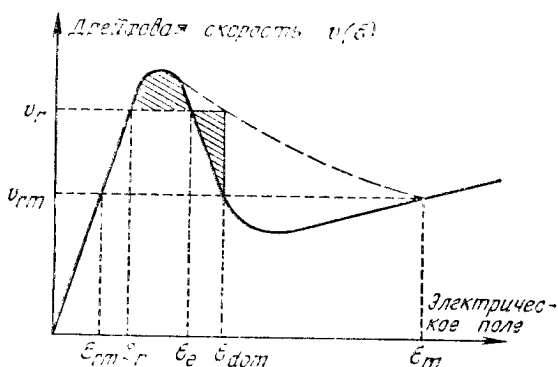


Рис. 16. Зависимости скорости от напряженности электрического поля (сплошная кривая) и максимальной напряженности электрического поля в домене от дрейфовой скорости вне его (штриховая кривая) [30].

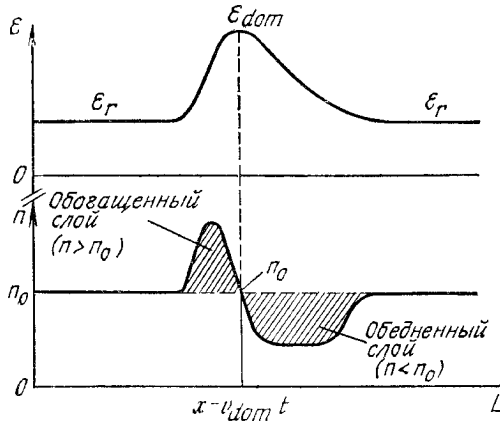


Рис. 17. Распределения электрического поля и концентрация электронов для стабильного домена высокого поля, движущегося со скоростью v_r [31].

ция n зависят от одной переменной $x' = x - v_{\text{дом}}t$, а их распределения приведены на рис. 17. Отметим, что концентрация является двучленной функцией электрического поля. Домен состоит из обогащенного носителями слоя, в котором $n > n_0$, и обедненного слоя ($n < n_0$). Концентрация носителей n равна n_0 при двух значениях напряженности электрического поля: при $\mathcal{E} = \mathcal{E}_r$ вне домена и при $\mathcal{E} = \mathcal{E}_{\text{дом}}$ в максимуме поля в домене.

Предположим, что величина поля вне домена \mathcal{E}_r известна. (Ниже мы покажем, что величину \mathcal{E}_r можно легко найти.) Ток вне домена состоит только из тока проводимости и определяется выражением $J = qn_0v_r$, в котором $v_r = v(\mathcal{E}_r)$. Так как

$$\frac{\partial \mathcal{E}}{\partial x} = \frac{\partial \mathcal{E}}{\partial x'} \quad \text{и} \quad \frac{\partial \mathcal{E}}{\partial t} = -v_{\text{дом}} \frac{\partial \mathcal{E}}{\partial x'},$$

выражения (2) и (28) можно записать следующим образом:

$$\frac{\partial \mathcal{E}}{\partial x'} = \frac{q}{\epsilon_s} (n - n_0) \quad (29)$$

и

$$\frac{d}{dx'} [D(\mathcal{E})n] = n[v(\mathcal{E}) - v_{\text{дом}}] - n_0(v_r - v_{\text{дом}}). \quad (30)$$

Разделив выражение (30) на выражение (29), мы можем исключить переменную x' и получить дифференциальное уравнение для $[D(\mathcal{E})n]$ как функцию электрического поля:

$$\frac{q}{\epsilon_s} \frac{d}{d\mathcal{E}} [D(\mathcal{E})n] = \{n[v(\mathcal{E}) - v_{\text{дом}}] - n_0(v_r - v_{\text{дом}})\} / (n - n_0). \quad (31)$$

В общем виде уравнение (31) может быть решено только численными методами [30–32]. Однако задача существенно упрощается,

если предположить, что коэффициент диффузии не зависит от электрического поля, т. е. $D(\mathcal{E}) = D$. С учетом этого приближения решение уравнения (31) будет иметь вид

$$\frac{n}{n_0} - \ln\left(\frac{n}{n_0}\right) - 1 = \frac{e_s}{qn_0D} \int_{\mathcal{E}_r}^{\mathcal{E}} \left\{ [v(\mathcal{E}') - v_{\text{dom}}] - \frac{n_0}{n} (v_r - v_{\text{dom}}) \right\} d\mathcal{E}', \quad (32)$$

в чем можно убедиться, продифференцировав его по \mathcal{E} .

Отметим, что если $\mathcal{E} = \mathcal{E}_r$ или $\mathcal{E} = \mathcal{E}_{\text{dom}}$, то концентрация $n = n_0$ (рис. 17) и левая часть выражения (32) равна 0. Поэтому интеграл в правой части выражения (32) должен также равняться 0 при $\mathcal{E} = \mathcal{E}_{\text{dom}}$. Однако проинтегрировать от \mathcal{E}_r до \mathcal{E}_{dom} можно либо по обедненному слою, в котором $n < n_0$, либо по обогащенному слою ($n > n_0$). Поскольку первое слагаемое под интегралом в выражении (32) не зависит от n , а результаты интегрирования второго слагаемого различаются в двух указанных выше случаях, для равенства интеграла нулю при интегрировании как по обедненному слою, так и по обогащенному необходимо, чтобы $v_r = v_{\text{dom}}$. Тогда при $\mathcal{E} = \mathcal{E}_{\text{dom}}$ уравнение (32) принимает вид

$$\int_{\mathcal{E}_r}^{\mathcal{E}_{\text{dom}}} [v(\mathcal{E}') - v_r] d\mathcal{E}' = 0. \quad (33)$$

Это условие выполняется, если равны площади заштрихованных областей на рис. 16. Воспользовавшись «правилом равных площадей» [30], можно определить максимальную величину поля в домене \mathcal{E}_{dom} , если известна величина \mathcal{E}_r . Штриховая кривая на рис. 16 представляет собой зависимость \mathcal{E}_{dom} от v_r , найденную с помощью этого правила. Кривая начинается в максимуме зависимости скорости от поля, в котором напряженность равна пороговой, и заканчивается в точке (\mathcal{E}_m, v_{rm}) . Если напряжение смещения прибора такое, что скорость $v(\mathcal{E}_r)$ меньше v_{rm} , то правило равных площадей не может быть выполнено и стабильный домен не существует [31].

Для решения уравнения (31) с учетом полевой зависимости коэффициента диффузии применяются численные методы. Из полученных решений следует, что при заданной величине напряженности электрического поля вне домена \mathcal{E}_r существует не более одного значения превышения скорости доменом ($v_{\text{dom}} - v_r$), при котором это решение существует. Другими словами, для любой \mathcal{E}_r существует единственная устойчивая доменная конфигурация. На рис. 18 приведены зависимости превышения скорости доменом от дрейфовой скорости носителей вне его v_r при двух значениях концентрации n_0 [32]. Отметим, что для $n_0 = 10^{15} \text{ см}^{-3}$ выпол-

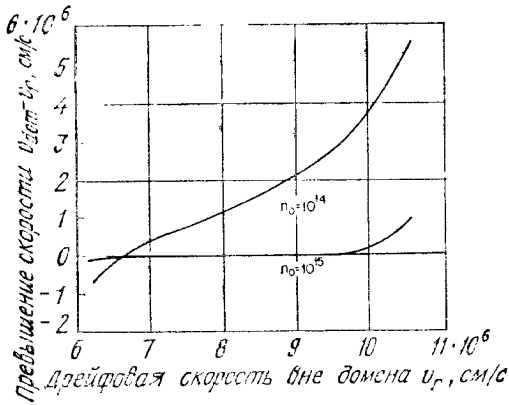


Рис. 18. Зависимость превышения скорости доменом высокого поля от дрейфовой скорости вне его [32].

няется равенство $v_{\text{дом}} = v_r$ в широком диапазоне v_r , в котором применимо правило равных площадей.

Теперь рассмотрим некоторые характеристики домена высокого поля. Если диполь не находится в контакте с каким-нибудь электродом, ток прибора зависит от напряженности электрического поля вне домена \mathcal{E}_r :

$$J = qn_0v(\mathcal{E}_r). \quad (34)$$

Поэтому при заданной концентрации n_0 поле вне домена определяет плотность тока J . Избыточное напряжение на диполе можно найти следующим образом:

$$V_{ex} = \int_{-\infty}^{\infty} [\mathcal{E}(x) - \mathcal{E}_r] dx. \quad (35)$$

Результаты вычисления интеграла (35) на ЭВМ для различных значений концентрации носителей и напряженности поля вне домена приведены на рис. 19 [32]. Эти кривые можно использовать для определения поля вне домена \mathcal{E}_r в образце длиной L с концентрацией примеси n_0 при напряжении смещения V , если учесть, что одновременно с уравнением (35) должно выполняться равенство

$$V_{ex} = V - L\mathcal{E}_r. \quad (36)$$

Прямая линия, определяемая этим уравнением, называется приборной прямой. Для частного случая $L = 25$ мкм и $V = 10$ В она нанесена штрихами на рис. 19. Если $V/L > \mathcal{E}_T$, то пересечение приборной прямой с зависимостью (35) однозначно определяет \mathcal{E}_r , и, следовательно, ток прибора. Наклон приборной прямой задается величиной L , а точка пересечения, соответствующая \mathcal{E}_r , может смещаться при изменении постоянного напряжения V .

Когда домен достигает анода, ток во внешней цепи возрастает, а электрическое поле перераспределяется в полупроводнике так, что образуется новый домен. Частота колебаний тока зависит от скорости движения домена в образце $v_{\text{дом}}$. Если возрастает $v_{\text{дом}}$, то увеличивается частота, и наоборот. Зависимость $v_{\text{дом}}$ от напряжения смещения может быть легко найдена.

На рис. 20 приведена зависимость ширины домена от избыточного напряжения [32]. Отметим, что при заданной величине V_{ex} ширина диполя уменьшается с увеличением концентрации. При бесконечно малом коэффициенте диффузии домен будет иметь треугольную форму, так как при $\mathcal{E}_r < \mathcal{E} < \mathcal{E}_{\text{дом}}$ правая часть выражения (32) стремится к бесконечности, если $D \rightarrow 0$. Поэтому левая часть выражения (32) также должна неограниченно возрастать. Это означает, что $n \rightarrow 0$ в обедненном слое и $n \rightarrow \infty$ в обогащенном слое. Напряженность электрического поля линейно уменьшается от $\mathcal{E}_{\text{дом}}$ до \mathcal{E}_r с расстоянием, а ширина домена равна

$$d = \frac{\epsilon_s}{qn_0} (\mathcal{E}_{\text{дом}} - \mathcal{E}_r). \quad (37)$$

Избыточное напряжение на домене

$$V_{\text{ex}} = (\mathcal{E}_{\text{дом}} - \mathcal{E}_r) d/2 = \frac{\epsilon_s (\mathcal{E}_{\text{дом}} - \mathcal{E}_r)^2}{2qn_0}. \quad (38)$$

На практике в приборах из арсенида галлия и фосфида индия генерировались только домены треугольной формы.

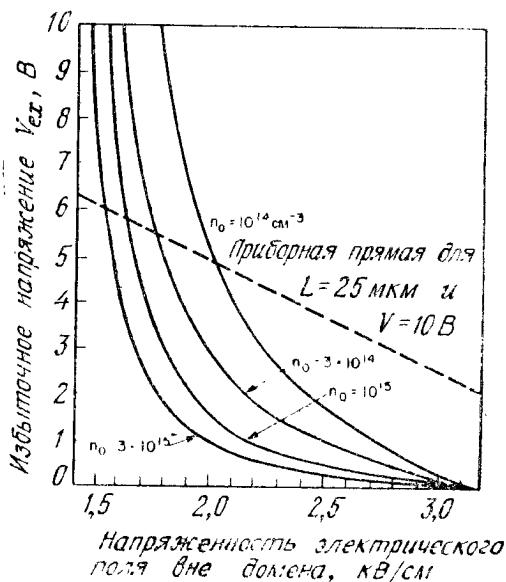


Рис. 19. Зависимость избыточного напряжения на домене от напряженности электрического поля вне его при различных концентрациях примеси [32].

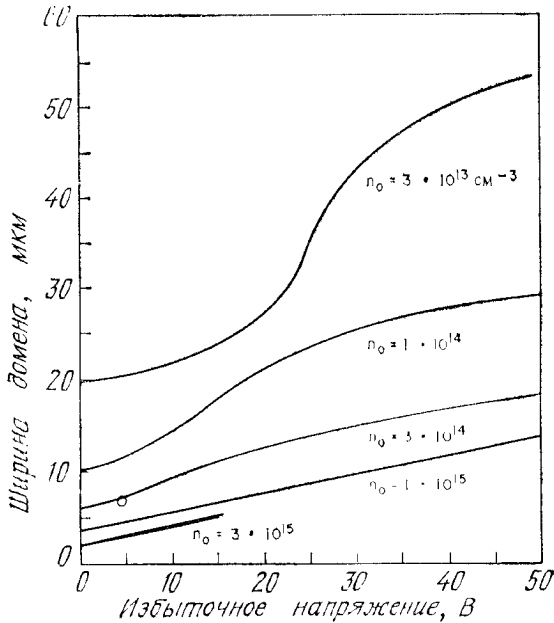
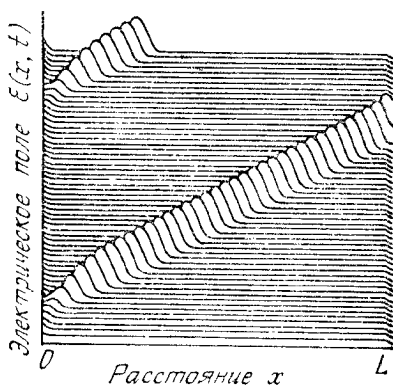


Рис. 20. Зависимость ширины домена от избыточного напряжения при различных концентрациях примеси [32].

Режим пролета домена может быть реализован, если прибор с величиной произведения n_0L , превышающей критическое значение, включить в параллельную резонансную цепь (например, в СВЧ-резонатор с высокой добротностью). В этом режиме домен высокого поля формируется у катода и движется через образец к аноду. Когда домен достигает анода, напряженность электрического поля в образце увеличивается до тех пор, пока не превысит пороговую, и новый диполь образуется у катода. На рис. 21 показаны результаты численного моделирования процесса распространения домена в приборах из арсенида галлия длиной 100 мкм при концентрации примеси $5 \cdot 10^{14} \text{ см}^{-3}$ ($n_0L = 5 \cdot 10^{12} \text{ см}^{-2}$) [33]. Две пространственные зависимости $\mathcal{E}(x, t)$, изображенные соседними кривыми, разделены промежутком $16\tau_R$, где τ_R — время диэлектрической релаксации (выражение (9)), равное для данного прибора 1,5 пс. В момент исчезновения домена на аноде происходит увеличение тока во внешней цепи; поэтому зависимость тока от времени в образце, длина которого значительно превышает ширину домена, будет иметь резкие «всплески», а не плавную синусоидальную форму. На рис. 22 показана экспериментальная зависимость тока от времени в образце длиной 100

Рис. 21. Численное моделирование процесса распространения возникающего у катода дипольного слоя в режиме пролета домена. Распределения поля, изображенные соседними кривыми, соответствуют моментам времени, разделенным интервалом 24 пс [33].



мкм и с $n_0L = 3 \cdot 10^{13} \text{ см}^{-2}$ [34]. Очевидно, что для получения формы колебаний, близкой к синусоидальной, следует уменьшать длину образца (что приведет к увеличению частоты) или увеличивать ширину домена. Из рис. 20 видно, что ширина домена растет при уменьшении концентрации примеси n_0 . В общем случае, уменьшая величину произведения n_0L , можно получить форму колебаний, более близкую к синусоидальной. На рис. 23 приведены распределения напряженности электрического поля в образце длиной 35 мкм в четыре различных момента в течение одного СВЧ-периода, а также профиль легирования и зависимости напряжения и тока от времени [29]. Для этого прибора форма тока больше похожа на синусоидальную. Теоретический анализ показывает, что к. п. д. пролетного режима максимален, когда произведение $n_0L \approx 10^{12} \text{ см}^{-2}$, а домен занимает приблизительно половину длины образца. В этом случае форма колебаний тока близка к синусоидальной. Максимально достижимая величина к. п. д. для этого режима составляет 10 %. Если форма импульсов тока близка к прямоугольной, то к. п. д. будет выше. Такую фор-

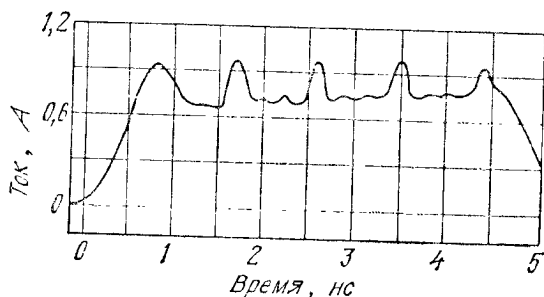


Рис. 22. Экспериментальная зависимость тока от времени в образце из арсенида галлия длиной 100 мкм при концентрации примеси $3 \cdot 10^{15} \text{ см}^{-3}$.

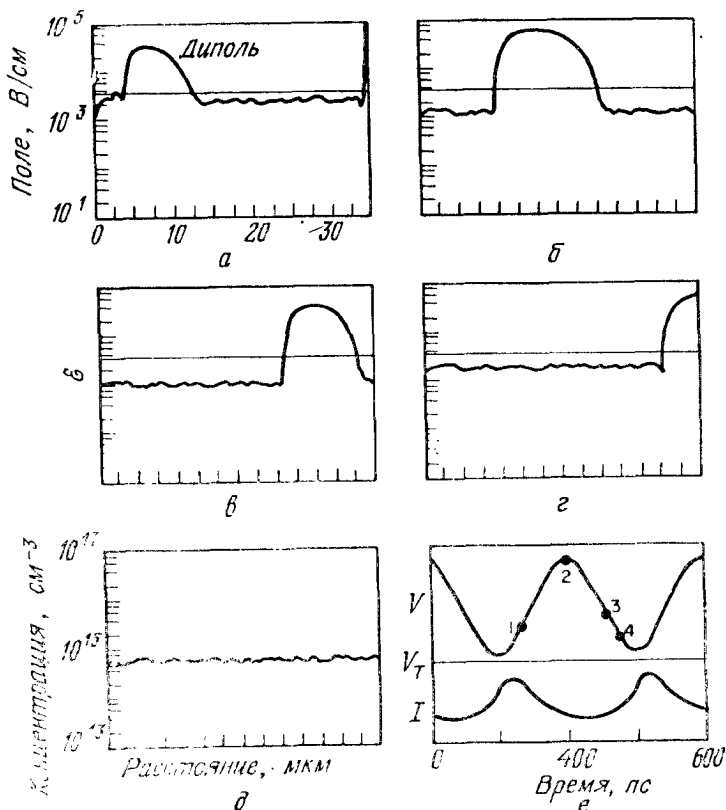


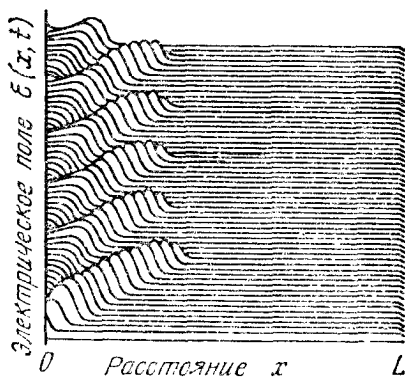
Рис. 23. Распределение поля в четыре различных момента одного СВЧ-периода для режима с пролетом домена (а—г), а также профиль легирования примесью (д) и зависимости тока и напряжения от времени (е) для образца из GaAs при $n_0L = 2,1 \cdot 10^{12} \text{ см}^{-2}$ и $fL = 0,9 \cdot 10^7 \text{ см/с}$ [29].

му генерируемых импульсов можно получить, если в момент исчезновения домена на аноде напряжение будет меньше порогового. Домен сформируется только после того, как напряжение превысит пороговое. Однако при реализации режима с запаздывающим формированием домена возникают большие трудности, связанные с настройкой.

11.3.4. Режим с разрушением домена

Включенный в резонансную цепь прибор на эффекте междлинного перехода электронов может работать на частотах, больших пролетной, при условии, что домен высокого поля рассеивается до того, как достигнет анода. В режиме пролета диполь-

Рис. 24. Численное моделирование режима с разрушением домена [33].



ного слоя большая часть падения напряжения на приборе приходится на домен. Поэтому ширина диполя уменьшается при понижении напряжения на диоде за счет изменения тока в резонансной цепи (рис. 20) до тех пор, пока в некоторой точке обедненный и обогащенный слои не нейтрализуют друг друга. Напряжение в этот момент равно V_s . Рассасывание домена происходит при уменьшении напряжения на приборе ниже V_s . Когда напряжение вновь превысит пороговое, возникнет новый домен, и процесс повторится. Следовательно, генерация будет происходить на частоте, определяемой параметрами резонансной цепи, а не на пролетной частоте.

На рис. 24 показан режим с разрушением домена [33]. Параметры прибора такие же, как и для кривых на рис. 21. Домен рассасывается на расстоянии $\sim L/3$ от катода, а рабочая частота примерно в три раза превышает пролетную (см. рис. 21).

Экспериментально [35] и теоретически [29] показано, что в режиме с разрушением домена в образцах, для которых резонансная частота цепи в несколько раз превышает пролетную (т. е. $fL > 2 \cdot 10^7$ см/с), а рабочая частота равна величине, обратной времени диэлектрической релаксации (т. е. $n_0/f \approx \epsilon_s/q|\mu_-|$), обычно образуются несколько диполей, поскольку электрическое поле не успевает перераспределиться в один домен. На рис. 25 показано формирование нескольких дипольных слоев в режиме работы с разрушением домена. Для этого образца $n_0L = 4,2 \times 10^{12}$ см $^{-2}$, $fL = 4,2 \cdot 10^7$ см/с и $n_0/f = 10^5$ с/см 3 .

Верхний предел рабочих частот в данном режиме определяется скоростью рассасывания домена, которая зависит от двух постоянных времени: времени диэлектрической релаксации и времени RC -цепи, где R равно сопротивлению той части полупроводника, в которой нет диполей, а C — емкость следующих друг за другом доменов. Первая постоянная времени определяет минимальную величину отношения $n_0/f \approx 10^4$ с/см 3 для арсенида гал-

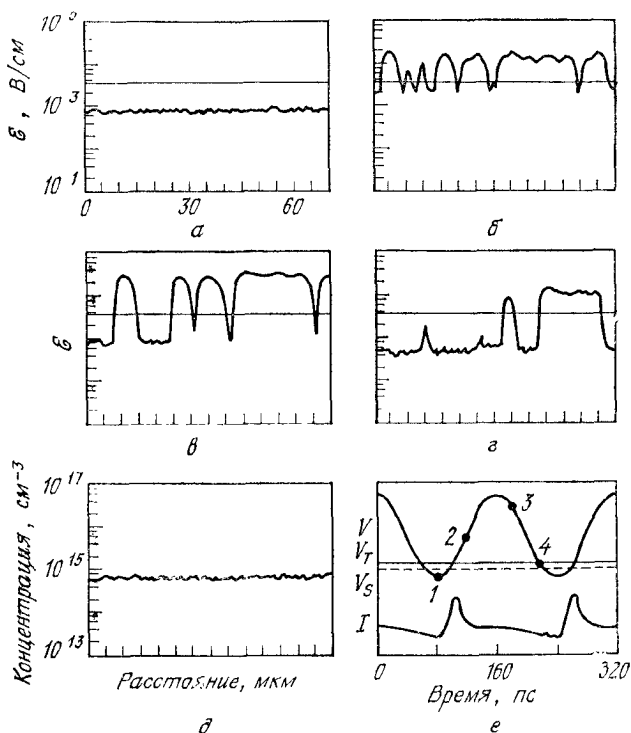


Рис. 25. Образование нескольких диполей в режиме с разрушением домена [29].

лия и фосфида индия n -типа [37, 38]. Вторая постоянная времени зависит от числа доменов и длины образца. Теоретическое максимальное значение к. п. д. генератора, работающего в режиме с разрушением домена, равно 13 % [36].

11.3.5. Режим ограниченного накопления объемного заряда [37]

В модели этого режима предполагается, что напряженность электрического поля увеличивается до пороговой и затем уменьшается так быстро, что распределение пространственного заряда, приводящее к образованию домена, не успевает сформироваться. У катода образуется только первичный обогащенный слой, а распределение электрического поля в остальной части прибора почти однородно, если только флуктуации концентрации примесей малы и их величина недостаточна для возникновения доменов. В этих условиях напряженность электрического поля в большей части прибора одинакова, что приводит к эффективной генерации мощности на частоте, определяемой параметрами внешней цепи. Чем

выше частота, тем меньше длина пути, проходимого слоем пространственного заряда, причем в большей части полупроводника дифференциальная подвижность остается отрицательной. Для работы прибора в этом режиме необходимо, чтобы в течение одного СВЧ-периода объемный заряд не успевал заметно возрасти, а обогащенный слой, наоборот, успел полностью разрушиться. Поэтому τ_R (выражение (9)), соответствующее отрицательной дифференциальной подвижности, должно быть больше по абсолютной величине, чем СВЧ-период, а τ_R , соответствующее положительной дифференциальной подвижности, должно быть меньше. Эти требования можно записать в виде неравенств [38]

$$\frac{e_s}{q\mu_+} \ll \frac{n_0}{f} < \frac{e_s}{q|\mu_-|}, \quad (39)$$

где μ_+ — положительная дифференциальная подвижность при малых напряженностях электрического поля, а μ_- — средняя отрицательная дифференциальная подвижность при полях, больших порогового. Для GaAs и InP получаем

$$10^4 < n_0/f < 10^5 \text{ с/см}^3. \quad (40)$$

Интересно отметить, что режим с разрушением нескольких доменов можно реализовать и в образце с флуктуациями примесной концентрации в некотором диапазоне значений n_0/f . Работающие в этом режиме приборы пригодны для генерации коротких импульсов большой мощности, поскольку можно использовать диоды большой длины (непролетная мода), для которых затруднен теплоотвод. Однако максимальная рабочая частота приборов, работающих в данном режиме, существенно ниже, чем приборов, работающих в пролетном режиме. Это связано с тем, что релаксация энергии электронов в основном минимуме зоны проводимости происходит достаточно медленно, что приводит к большим временам рассасывания домена. Из результатов численного моделирования на ЭВМ следует, что при генерации в непрерывном режиме максимальное время, в течение которого прибор на основе GaAs должен находиться при допороговом смещении, составляет ~ 20 пс; это ограничивает верхний предел рабочих частот величиной ~ 20 ГГц [39, 40]. Ожидается, что приборы из фосфида индия будут иметь более высокий верхний предел.

11.4. ХАРАКТЕРИСТИКИ ПРИБОРОВ НА ЭФФЕКТЕ МЕЖДОЛИННОГО ПЕРЕХОДА ЭЛЕКТРОНОВ

11.4.1. Катодные контакты

Для изготовления приборов на эффекте междолинного перехода электронов необходимы чрезвычайно чистые и однородные полупроводниковые материалы с низкой концентрацией глубоких

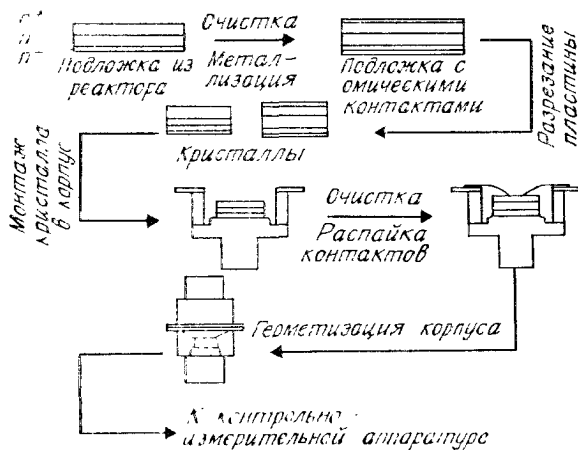


Рис. 26. Основные стадии процесса изготовления приборов на эффекте междолинного перехода электронов [41].

донорных уровней и ловушек, особенно для приборов, работающих в режиме с рассасыванием объемного заряда. Первые приборы изготавливались из арсенида галлия и фосфида индия со сплавными омическими контактами. В настоящее время для этих целей используются методами эпитаксии из газовой и жидкой фаз или методом молекулярно-лучевой эпитаксии. Типичный диапазон донорных концентраций составляет 10^{14} — 10^{16} см⁻³, а длины прибора — от нескольких микрометров до нескольких сотен микрометров. Основные стадии процесса изготовления такого прибора на эпитаксиальном арсениде галлия n^+ — n — n^+ -типа показаны на рис. 26 [41]. Кристаллы с приборами на эффекте междолинного перехода электронов монтируются в корпуса, которые, как и теплоотводы, аналогичны корпусам для ЛПД, описанным в гл. 10. Некоторые высокоомощные приборы изготавливаются при помощи селективной металлизации или травления мезоструктур.

Для улучшения характеристик прибора вместо омических n^+ -контактов применяют контакты с ограниченной инжекцией [42—44]. При использовании таких контактов можно реализовать ситуации, когда пороговое поле для тока приблизительно равно пороговой напряженности \mathcal{E}_T , определяющей начало участка ОДС. Поэтому поле в полупроводнике может быть однородным. В приборе с омическими контактами формирование обогащенного слоя или домена происходит на некотором расстоянии от катода из-за конечного времени разогрева электронов в основном минимуме зоны проводимости. Ширина «мертвой зоны» может достигать 1 мкм, что накладывает ограничения на минимальную

длину диода и, следовательно, максимальную рабочую частоту. В контактах с ограниченной инжекцией горячие электроны эмитируются из катода, что приводит к уменьшению ширины «мертвой зоны». Поскольку влияние пролетных эффектов можно свести к минимуму, отрицательное сопротивление прибора не будет зависеть от частоты и будет шунтироваться геометрической емкостью диода. Если к прибору подсоединить индуктивность и достаточно малое сопротивление, то прибор может генерировать колебания резонансной частоты в режиме однородного электрического поля. Теоретическая величина к. п. д. для этого случая найдена в разд. 11.3.1.

Были изучены два класса контактов с ограниченной инжекцией: контакт Шоттки с малой высотой барьера и двухслойная катодная структура. На рис. 27 приведены характеристики трех катодных контактов. В приборе с омическими контактами (рис. 27, а) всегда существует область низкой напряженности электрического поля у катода, а поле в образце неоднородно. В обратнорасположенном барьере Шоттки [45] можно получить достаточно однородное распределение электрического поля (рис. 27, б). Обратный ток равен (гл. 5)

$$J_R = A^{**} T^2 \exp(-q\phi_B/kT), \quad (41)$$

где A^{**} — эффективная константа Ричардсона, а $q\phi_B$ — высота барьера. Плотностям тока в диапазоне $10^2 - 10^4$ А/см² соответствуют

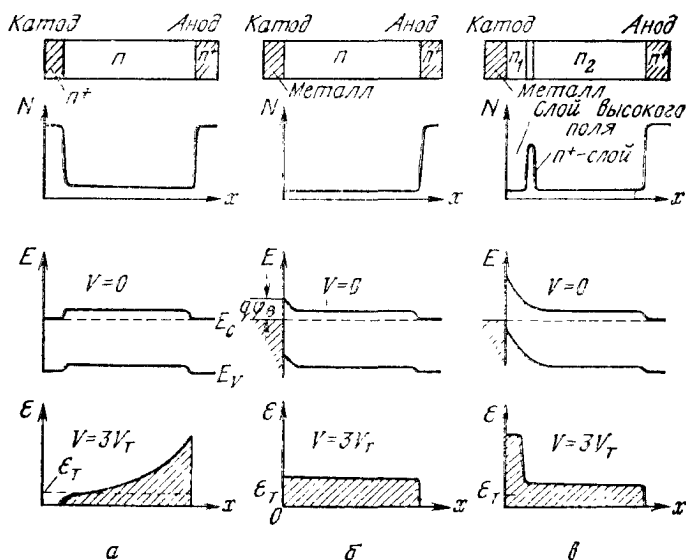


Рис. 27. Характеристики трех катодных контактов.

а — омический; б — барьер Шоттки; в — двухслойный контакт с барьером Шоттки.

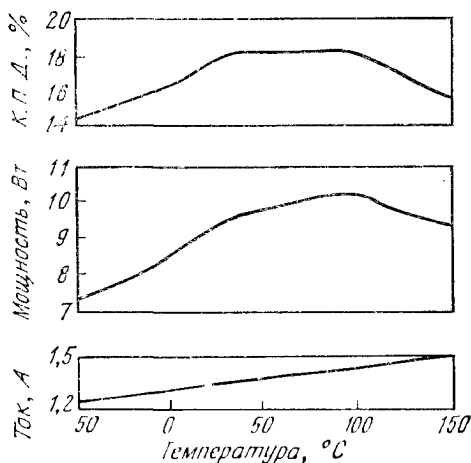


Рис. 28. К. п. д., генерируемая СВЧ-мощность и ток в приборе из фосфида индия с двухслойным катодным контактом, работающим в импульсном режиме при $n_0 = 2 \cdot 10^{15} \text{ см}^{-3}$ и $l = 8 \text{ мкм}$ и в области температур -50 — $150 \text{ }^\circ\text{C}$ [46].

значения высот барьера в пределах $0,15$ — $0,3 \text{ эВ}$. Однако нелегко получить барьер Шоттки небольшой высоты в полупроводниковых соединениях типа $\text{A}^{\text{III}}\text{B}^{\text{V}}$; кроме того, область рабочих температур ограничена, поскольку ток изменяется экспоненциально с температурой (выражение (41)).

Двухслойный катодный контакт состоит из слоя высокой напряженности электрического поля и n^+ -слоя (рис. 27, в) [46]. Его конфигурация аналогична конфигурации ЛПД с трехслойной базой (гл. 10). Электроны разогреваются в слое высокого электрического поля, а затем инжектируются в активную область, где поле однородно. Такие структуры успешно используются в широком диапазоне температур. Коэффициент полезного действия, генерируемая в импульсном режиме мощность и ток прибора из фосфида индия в области температур -50 — $150 \text{ }^\circ\text{C}$ приведены на рис. 28. Отметим, что в этом температурном интервале к. п. д. и мощность изменяются приблизительно всего лишь на $\pm 10 \%$. Наибольший к. п. д., полученный у приборов из фосфида индия с двухслойным катодным контактом, равен 24% . Из-за захвата уровня Ферми на поверхностные состояния катодные контакты с ограниченной инжекцией в приборах из GaAs до сих пор не созданы.

11.4.2. Зависимость мощности от частоты и шум

Процесс передачи энергии электрического поля электронам и переход последних из основного минимума зоны проводимости в побочные занимают конечный промежуток времени. Поэтому рабочие частоты приборов ограничены сверху величинами, обратными времени релаксации энергии и времени междолинного

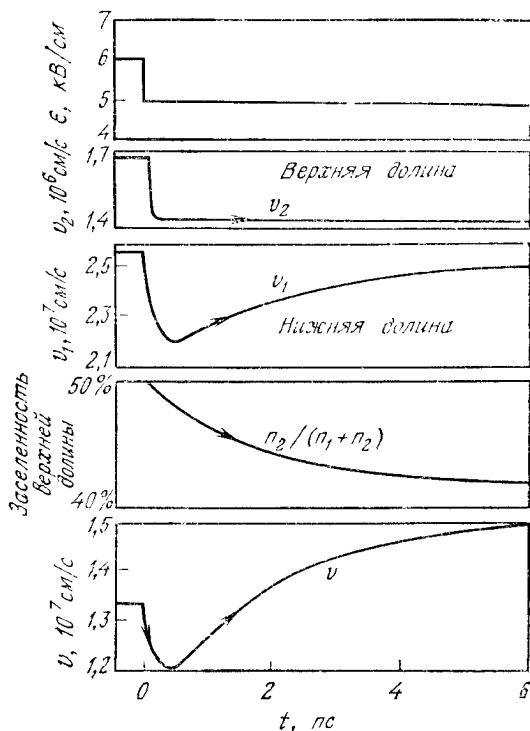


Рис. 29. Временные зависимости при внезапном изменении напряженности электрического поля от 6 до 5 кВ/см в момент $t = 0$ [47].

рассеяния. На рис. 29 приведены временные зависимости дрейфовой скорости носителей в верхней и нижней долинах, средней скорости и заселенности верхней долины при внезапном уменьшении напряженности электрического поля от 6 до 5 кВ/см [47]. Отметим, что скорость электронов в верхней долине v_2 практически мгновенно реагирует на изменение поля. Однако скорость в нижней долине v_1 изменяется медленно (характерное время отклика составляет ~ 5 пс), поскольку горячие электроны в основном минимуме слабо рассеиваются. Кроме того, причиной медленного изменения концентрации n_2 является большая величина времени междолинного рассеяния электронов. Поэтому отклик средней скорости v связан частично с изменением скорости v_1 и частично с междолинным рассеянием. Оценка верхнего предела рабочей частоты прибора из-за конечного времени отклика составляет ~ 150 ГГц.

В пролетном режиме рабочая частота обратно пропорциональна длине диода, т. е. $f = v/L$. Соотношение между генерируемой мощностью и частотой имеет вид

$$P_{rf} = V_{rf}^2/R = \mathcal{E}'_1 L^2/R = \frac{\mathcal{E}'_1 v^2}{Rj^2} \sim \frac{1}{j^2}, \quad (42)$$

где V_{rf} и \mathcal{E}'_1 — СВЧ-напряжение и напряженность электрического поля соответственно, а R — импеданс. Поэтому ожидаемое изменение мощности с частотой пропорционально $1/f^2$. Зависимость генерируемой в непрерывном и импульсном режимах СВЧ-мощности от частоты для современных приборов на эффекте междолинного перехода электронов из арсенида галлия и фосфида индия показана на рис. 30 [48—53, 57, 58]. В скобках рядом с экспериментальными точками приведены значения к. п. д. в процентах. Как следует из формулы (42), изменение мощности пропорционально $1/j^2$. В импульсном режиме в приборах из GaAs с достаточно длинной базой получена СВЧ-мощность до 6 кВт на частоте, близкой к 2 ГГц. Генерируемая в непрерывном режиме мощность приблизительно равна 2 Вт на 10 ГГц, что примерно в пять раз меньше, чем для лавинно-пролетного диода (ЛПД). Однако при фиксированной частоте напряжение смещения в приборе на эффекте междолинного перехода электронов меньше (в 2—5 раз), чем в лавинно-пролетном диоде.

Характеристики приборов из фосфида индия в непрерывном режиме сравнимы с характеристиками приборов из арсенида галлия. Однако в импульсном режиме они хуже, несмотря на то что по теоретическим оценкам к. п. д. приборов из фосфида индия выше. Причиной этого является более развитая технология арсенида галлия.

В приборах на эффекте междолинного переноса обнаруживаются шумы двух видов: амплитудно-модулированный (АМ) и частотно-модулированный (ЧМ), причем эти шумы возникают из-за флюктуаций тепловой скорости электронов. Обычно АМ-шум мал, поскольку амплитуда относительно стабильна из-за сильной нелинейности зависимости скорости от напряженности электрического поля. Величина среднего отклонения частоты в случае ЧМ-шума равна [54]

$$f_{rms} = \frac{f_0}{Q_{ex}} \sqrt{\frac{kT_{eq}(f_m)B}{P}}, \quad (43)$$

где f_0 — пролетная частота, Q_{ex} — добротность внешней цепи, P_0 — генерируемая мощность, B — ширина полосы. Эквивалентная ЧМ-шумовая температура T_{eq} равна

$$T_{eq}(f_m) = qD/k|\mu_-|, \quad (44)$$

где величина средней отрицательной дифференциальной подвижности μ_- зависит от амплитуды напряжения. Поскольку отноше-

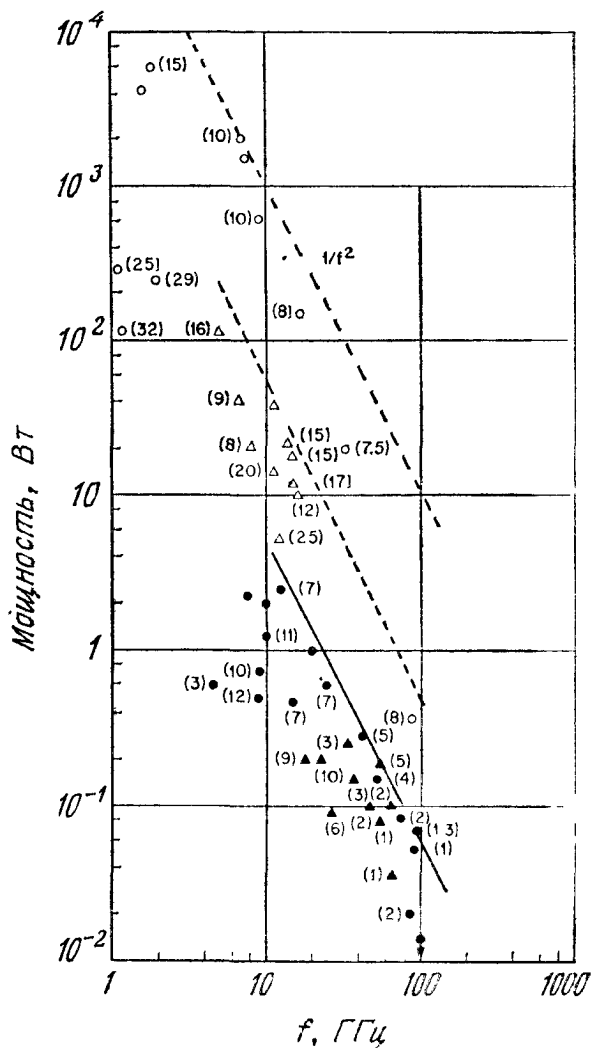


Рис. 30. Зависимость генерируемой СВЧ-мощности от частоты для приборов на эффекте междолинного перехода электронов.

○ импульсный режим, прибор из GaAs, Δ непрерывный режим, прибор из GaAs, ● импульсный режим, прибор из InP, ▲ непрерывный режим, прибор из InP.

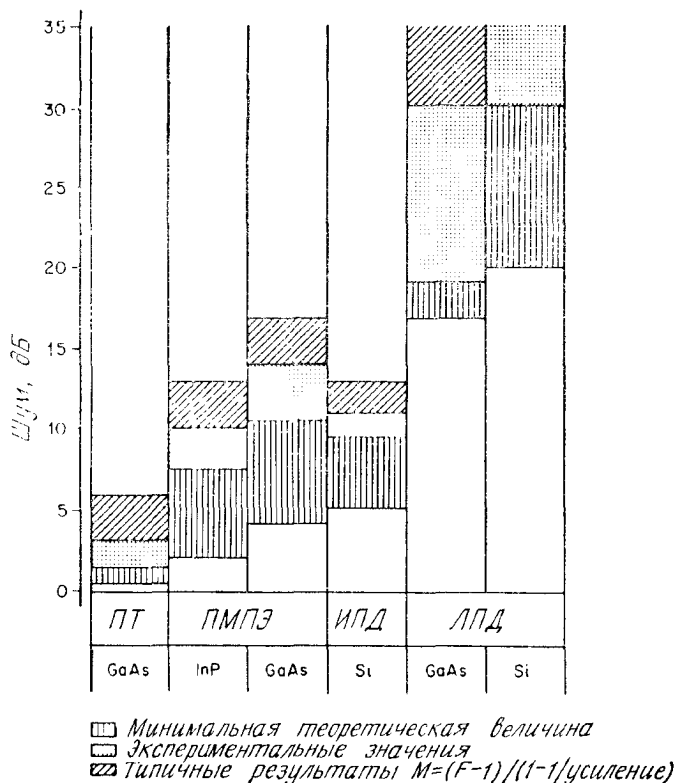


Рис. 31. Малосигнальный шум в полупроводниковых СВЧ-приборах в диапазоне частот 8—12 ГГц [55].

ПТ — полевой транзистор с затвором Шоттки; ПМПЭ — приборы на междолинном переходе электронов.

ние $D/|\mu_-|$ в фосфиде индия меньше, чем в арсениде галлия, ожидается, что уровень шума в приборах из InP будет также ниже.

Результаты измерения малосигнального шума в СВЧ-приборах в диапазоне частот 8—12 ГГц приведены на рис. 31 [55]. Наименьший шум наблюдается у полевого транзистора с затвором Шоттки из арсенида галлия (гл. 6). Шумы кремниевого ИПД сравнимы с шумами приборов на эффекте междолинного перехода электронов. Причиной того, что шумы в ЛПД больше, чем в других приборах, являются лавинные процессы.

11.4.3. Другие приборы

Мы рассмотрели эффект междолинного перехода электронов и его использование в СВЧ-генераторах и усилителях. Приборы на этом эффекте применяются в быстродействующих цифровых и ана-

логовых устройствах. Ниже мы рассмотрим приборы с изменяющейся вдоль направления тока площадью поперечного сечения и приборы с тремя контактами.

При анализе процессов в генераторе с переменной площадью поперечного сечения можно воспользоваться результатами одномерной теории доменов высокого поля, если предположить, что толщина диполя мала, и рассматривать явления в практически однородной области полупроводника вблизи домена. Эти допущения справедливы, если $n_0L \gg 10^{12} \text{ см}^{-2}$, а площадь поперечного сечения диода изменяется плавно при удалении от контакта. Воспользовавшись теоретическими результатами, полученными в предыдущем разделе, можно показать, что существует такое избыточное напряжение на домене $V_{ex} = V_{gm}$, при превышении которого электрическое поле вне домена \mathcal{E}_r не изменяется. Соответствующая величина напряженности электрического поля вне домена равна $\mathcal{E}_{r,m}$ (рис. 16). Когда избыточное напряжение на домене $V_{ex} = V_{gm}$, плотность тока равна

$$J_{gm} = qn_0v_{gm}. \quad (45)$$

Такие «насыщенные» домены движутся в генераторе с постоянной скоростью. Пусть толщина объемного генератора равна s , а изме-

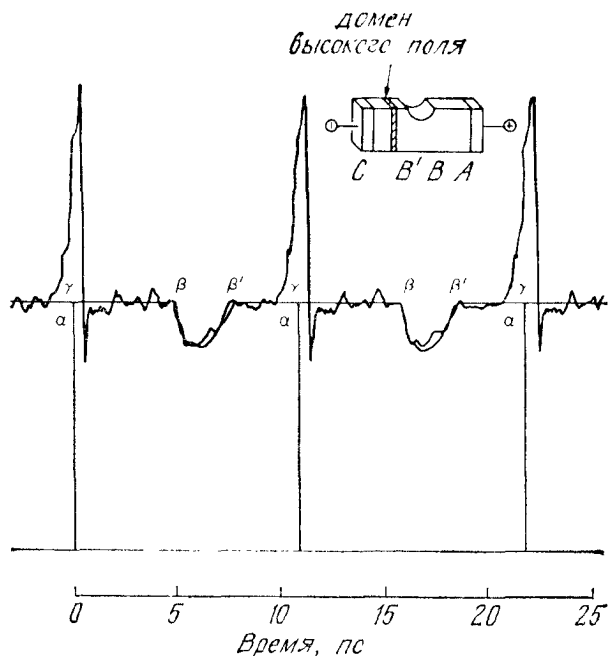


Рис. 32. Форма тока прибора, показанного в верхней части рисунка [56].

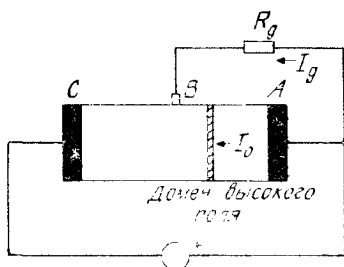
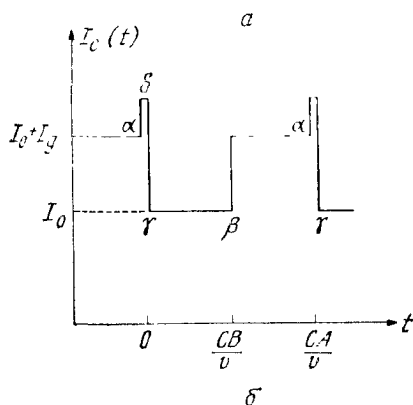


Рис. 33. Генератор управляемых прямоугольных импульсов тока (а) и их форма (б).



нящаяся с расстоянием от катода x ширина — $b(x)$. Если домен высокого поля зарождается у катода при $t = 0$, то в момент времени t диполь находится на расстоянии $x(t) = v_{rm}t$. В любом поперечном сечении прибора ток должен быть одинаков, а вблизи домена равен

$$I(t) = J_{rm}sb(v_{rm}t) + I_g(t), \quad (46)$$

где $I_g(t)$ — связанный с исчезновением домена ток, который отличен от нуля только в конце периода. Поэтому из равенства (46) следует, что ток пропорционален $b(v_{rm}t)$.

На рис. 32 приведена форма импульсов тока прибора, показанного в верхней части рисунка [56]. Тонкими сплошными линиями показана ожидаемая для такой конфигурации прибора форма тока. Обозначения α , β , β' и γ соответствуют моментам времени, в которые домен находится в точках A , B , B' и C .

До сих пор мы рассматривали только приборы с двумя контактами. Формой генерируемых импульсов можно управлять с помощью одного или нескольких электродов, расположенных вдоль длины прибора. На рис. 33, а показана конфигурация такого прибора с электродом в точке B [56]. Ожидаемая форма тока приве-

дена на рис. 33, б. Здесь снова используется теория насыщенного домена, описанная выше. После отрыва домена от катода при $t = 0$ ток катодного контакта равен току насыщенного домена (Aqn_0v_{rm}) до тех пор, пока диполь не достигнет электрода, расположенного в точке B . В этот момент ток становится равным сумме тока насыщенного домена и протекающего через сопротивление R_g тока I_g . Ток I_g равен напряжению между точками A и B при наличии между ними домена, поделенному на R_g . После этого ток катода остается равным

$$I_c(t) = Aqn_0v_{rm} + I_g \quad (47)$$

до тех пор, пока домен не достигнет анода, вызвав короткий «всплеск» тока.

ЛИТЕРАТУРА

1. Gunn J. B. Microwave Oscillation of Current in III—V Semiconductors, *Solid State Commun.*, **1**, 88 (1963); Instabilities of Current in III—V Semiconductors, *IBM J. Res. Dev.*, **8**, 141 (1964).
2. Kroemer H. Theory of the Gunn Effect, *Proc. IEEE*, **52**, 1736 (1964).
3. Ridley B. K., Watkins T. B. The Possibility of Negative Resistance Effects in Semiconductors, *Proc. Phys. Soc. Lond.*, **78**, 293 (1961).
- 3a. Ridley B. K. Anatomy of the Transferred-Electron Effect in III—V Semiconductors, *J. Appl. Phys.*, **48**, 754 (1977).
4. Hilsum C. Transferred Electron Amplifiers and Oscillators, *Proc. IRE*, **50**, 185 (1962).
- 4a. Hilsum C. Historical Background of Hot Electron Physics, *Solid State Electron.*, **21**, 5 (1978).
5. Hutson A. R., Jayaraman A., Chynoweth A. G., Coriel A. S., Feldmann W. L. Mechanism of the Gunn Effect from a Pressure Experiment, *Phys. Rev. Lett.*, **14**, 639 (1965).
6. Allen J. W., Shyam M., Chen Y. S., Pearson G. L. Microwave Oscillations in $\text{GaAs}_{1-x}\text{Px}$ Alloys, *Appl. Phys. Lett.*, **7**, 78 (1965).
7. Carroll J. E. Hot Electron Microwave Generator, Edward Arnold, London, 1970.
8. Bulman P. J., Holson G. S., Taylor B. S. Transferred Electron Devices, Academic, N. Y., 1972.
9. Bosch B. G., Engelmann R. W. H. Gunn-Effect Electronics, Wiley, N. Y., 1975.
10. Thim H. W. Solid State Microwave Sources, in Hilsum C., Ed., Handbook on Semiconductors, Vol. 4, Device Physics, North-Holland, Amsterdam, 1980.
11. Ridley B. K. Specific Negative Resistance in Solids, *Proc. Phys. Soc. Lond.*, **82**, 954 (1963).
12. Ridley B. K., Pratt R. G. A Bulk Differential Negative Resistance Due to Electron Tunneling through an Impurity Potential Barrier, *Phys. Lett.*, **4**, 300 (1963).
13. McWhorter A. L., Rediker R. H. The Cryosar — A New Low-Temperature Computer Component, *Proc. IRE*, **47**, 1207 (1959).
14. Kroemer H. Negative Conductance in Semiconductors, *IEEE Spectrum*, **5**, 47 (1968).
15. Aspnes D. E. GaAs Lower Conduction Band Minimum: Ordering and Properties, *Phys. Rev.*, **14**, 5331 (1976).
16. Rees H. D., Gray K. W. Indium Phosphide: A Semiconductor for Microwave Devices, *Solid State Electron Devices*, **1**, 1 (1976).

17. McCumber D. E., Chynoweth A. G. Theory of Negative Conductance Application and Gunn Instabilities in 'Two-Valley' Semiconductors, *IEEE Trans. Electron Devices*, **ED-13**, 4 (1966).
18. Sakai K., Ikoma T., Adachi Y. Velocity-Field Characteristics of $\text{Ga}_x\text{In}_{1-x}\text{Sb}$ Calculated by the Monte Carlo Method, *Electron Lett.*, **10**, 402 (1974).
19. Hayes R. E., Raymond R. M. Observation of the Transferred-Electron Effect in GaInAsP , *Appl. Phys. Lett.*, **31**, 300 (1977).
20. Ruch J. G., Kino G. S. Measurement of the Velocity-Field Characteristics of Gallium Arsenide, *Appl. Phys. Lett.*, **10**, 40 (1967).
21. Butcher P. N., Fawcett W. Calculation of the Velocity-Field Characteristics for Gallium Arsenide, *Phys. Lett.*, **21**, 489 (1966).
22. Littlejohn M. A., Hauser J. R., Glisson T. H. Velocity — Field Characteristics of GaAs with F — L — X Conduction-Band Ordering, *J. Appl. Phys.*, **48**, 4587 (1977).
23. Mojzes I., Podor B., Balogh I. On the Temperature Dependence of Peak Electron Velocity and Threshold Field Measured on GaAs Gunn Diodes, *Phys. Status Solidi*, **39**, K123 (1977).
24. Shockley W. Negative Resistance Arising from Transit Time in Semiconductor Diodes, *Bell Syst. Tech. J.*, **33**, 799 (1954).
25. Kino G. S., Kuru I. High-Efficiency Operation of a Gunn Oscillator in the Domain Mode, *IEEE Trans. Electron Devices*, **ED-16**, 735 (1969).
26. Hobson G. S. The Gunn Effect, Clarendon, Oxford, 1974.
27. Thim H. W., Haydl W. Microwave Amplifier Circuit Consideration, in Howes M. J., Morgan D. V., Eds., *Microwave Devices*, Wiley, N. Y., 1976, Chap. 6.
28. Hakki B. W. Amplification in Two-Valley Semiconductors, *J. Appl. Phys.*, **38**, 808 (1967).
29. Thim H. W. Computer Study of Bulk GaAs Devices with Random One-Dimensional Doping Fluctuations, *J. Appl. Phys.*, **39**, 3897 (1968).
30. Butcher P. N. Theory of Stable Domain Propagation in the Gunn Effect, *Phys. Lett.*, **19**, 546 (1965).
31. Butcher P. N., Fawcett W., Hilsun C. A Simple Analysis of Stable Domain Propagation in the Gunn Effect, *Br. J. Appl. Phys.*, **17**, 841 (1966).
32. Copeland J. A. Electrostatic Domains in Two-Valley Semiconductors, *IEEE Trans. Electron Devices* **ED-13**, 187 (1966).
33. Shaw M., Grubin H. L., Solomon P. R. The Gunn — Hilsun Effect, Academic N. Y. 1979.
34. Fukui H. New Method of Observing Current Waveforms in Bulk GaAs *Proc. IEEE*, **54**, 792 (1966).
35. Thim H. W., Barber M. R. Observation of Multiple High-Field Domains in *n*-GaAs, *Proc. IEEE*, **56**, 110 (1968).
36. Barber M. R. High Power Quenched Gunn Oscillators, *Proc. IEEE*, **56**, 752 (1968).
37. Copeland J. A. A New Mode of Operation for Bulk Negative Resistance Oscillators, *Proc. IEEE*, **54**, 1479 (1966).
38. Copeland J. A. LSA Oscillator Diode Theory, *J. Appl. Phys.*, **38**, 3096 (1967).
39. Jones D., Rees H. D. Electron-Relaxation Effects in Transferred-Electron Devices Revealed by New Simulation Method, *Electron. Lett.*, **8**, 363 (1972).
40. Kroemer H. Hot Electron Relaxation Effects in Devices, *Solid State Electron.*, **21**, 61 (1978).
41. Sterzer F. Transferred Electron Amplifiers and Oscillators for Microwave Application, *Proc. IEEE*, **59**, 1155 (1971).
42. Kroemer H. The Gunn Effect under Imperfect Cathode Boundary Condition, *IEEE Trans. Electron Devices*, **ED-15**, 819 (1968).
43. Atalla M. M., Moll J. L. Emitter Controlled Negative Resistance in GaAs, *Solid State Electron*, **12**, 619 (1969).

44. Yu S. P., Tantraporn W., Young J. D. Transit-Time Negative Conductance in GaAs Bulk — Effect Diodes, *IEEE Trans. Electron. Devices*, **ED-18**, 88 (1971).
45. Colliver D. J., Irving L. D., Pattison J. E., Rees H. D. High-Efficiency InP Transferred-Electron Oscillators, *Electron. Lett.*, **10**, 221 (1974).
46. Gray K. W., Pattison J. E., Rees J. E., Prew B. A., Clarke R. C., Irving L. D. InP Microwave Oscillator with 2-Zone Cathode, *Electron. Lett.*, **11**, 402 (1975).
47. Rees H. D. Time Response of the High-Field Electron Distribution Function in GaAs, *IBM J. Res. Dev.*, **13**, 537 (1969).
48. Hobson G. S. Recent Development in Transferred Devices, *J. Phys. E*, **7**, 229 (1974).
49. Irving L. D., Pattison J. E., Braddock P. W., Gray K. W. Improved Mean Power and Long Pulse-Width Operation of InP TEDs in J Band, *Electron. Lett.*, **14**, 116 (1978).
50. Hamilton R. J., Fairman R. D., Long S. I., Omori M., Fank F. B. InP Gunn-Effect Devices for Millimeter-Wave Amplifiers and Oscillators, *IEEE Trans. Microwave Theory Tech.*, **MTT-24**, 775 (1976).
51. Day W. R. Gunn Oscillators, *Microwave Syst. News*, **8**, 56 (1978).
52. Suzuki T., Ito M., Ishii T., Mitsui S. Design and Fabrication for High Efficiency and High Output Power Gunn Diodes, *Trans. Inst. Electron. Commun. Eng., Jpn.*, **E61**, 932 (1978).
53. Mun J. High-Efficiency and High-Peak-Power InP Transferred-Electron Oscillators, *Electron. Lett.*, **13**, 275 (1977).
54. Ataman A., Harth W. Intrinsic FM Noise of Gunn Oscillators, *IEEE Trans. Electron Devices*, **ED-20**, 12 (1973).
55. Constant E. Noise in Microwave, Injection, Transit Time and Transferred-Electron Devices, *Physica*, **83B**, 24 (1976).
56. Shoji M. Functional Bulk Semiconductor Oscillators, *IEEE Trans. Electron Devices*, **DED-14**, 535 (1967).
57. Ondria J. Wide-Band Mechanically Turnable W-Band (75—110GHz) CW GaAs Diode Oscillators, Proc. 7th Bienn. Conf. Active Microwave Semicond. Devices Circuits, Cornell University, Ithaca, N. Y., 1979, p. 309.
58. Crowley J. D., Fank F. B., Hyder S. B., Sowers J. J., Tringali D. Millimeter Wave InP Transferred Electron Devices, Proc. 7th Bienn. Conf. Active Microwave Semicond. Devices Circuits, Cornell University, Ithaca, N. Y., 1979, p. 331.
59. Hauser J. R., Glisson T. H., Littlejohn M. A. Negative Resistance and Peak Velocity in the Central (000) Valley of III—V Semiconductors, *Solid State Electron.*, **22**, 487 (1979).

Часть V

Оптоэлектронные приборы

Глава 12

СВЕТОДИОДЫ И ПОЛУПРОВОДНИКОВЫЕ ЛАЗЕРЫ

12.1. ВВЕДЕНИЕ

К оптоэлектронным приборам относятся приборы, в которых основные процессы протекают с участием квантов света — фотонов. Оптоэлектронные приборы можно разделить на три группы: 1) приборы, преобразующие электрическую энергию в оптическое излучение (светодиоды, полупроводниковые лазеры); 2) приборы, детектирующие оптические сигналы за счет протекающих под действием света электронных процессов (фотодетекторы); 3) приборы, осуществляющие преобразование оптического излучения в электрическую энергию (фотовольтаические приборы, солнечные батареи).

В данной главе рассматривается первая группа приборов, а гл. 13 и 14 посвящены соответственно фотодетекторам и солнечным батареям.

Явление электролюминесценции было открыто в 1907 г. [1]. Оно состоит в генерации света в материале, сквозь который протекает электрический ток, вызванный электрическим полем. Излучение электролюминесценции в отличие от теплового характеризуется сравнительно узким спектром (для светодиодов спектральная ширина линии лежит обычно в диапазоне длин волн 100—500 Å). Излучение может быть практически монохроматическим; например, для лазерного диода спектральная ширина излучения лежит в пределах 0,1—1 Å. Подробное изучение светодиодов проведено в работе [2], а вопросам практического использования светодиодов в качестве источников света, дисплеев и оптронов посвящена работа [3].

После открытия в 1954 г. лазера и последующей разработки оптических и рубиновых лазеров [5, 6] было предложено использовать полупроводники в качестве материала для лазеров [7—10]. В 1961 г. на основе теоретических расчетов [11] были найдены условия, необходимые для осуществления режима генерации лазерного излучения с использованием концепции квазиуровней Ферми. В 1962 г. в работе [12] было показано, что генерация

может быть практически реализована в прямозонных полупроводниках, и были найдены важнейшие критерии для такого режима. В том же 1962 г. три группы авторов почти одновременно сообщили об осуществлении режима лазерной генерации в полупроводниках [13—15]. Было получено импульсное излучение с длиной волны 0,84 мкм в $p-n$ -переходах на основе GaAs, смещенных в прямом направлении и охлажденных до температуры жидкого азота. Вскоре после этого появилось сообщение о получении генерации излучения на длине волны 0,71 мкм в $p-n$ -переходе на основе тройного соединения GaAs_{1-x}P_x [16]. В 1970 г. авторы работы [17] реализовали режим непрерывной генерации при комнатной температуре в лазерах на основе двойных гетеропереходов. Первые структуры такого типа были предложены в работах [18, 19].

Со времени этих первых открытий было найдено множество новых материалов для лазеров. Диапазон длин волн когерентного излучения включает близкую ультрафиолетовую, видимую и дальнюю инфракрасную области спектра (от ~0,3 до ~30 мкм). Полупроводниковые лазеры рассматриваются как основные источники света для волоконно-оптических систем связи. Они могут найти также широкое применение во многих областях фундаментальных исследований, таких, как газовая спектроскопия высокого разрешения и контроль атмосферных загрязнений. Свойства лазеров на основе гетероструктур подробно рассмотрены в работах [20, 21].

12.2. ИЗЛУЧАТЕЛЬНЫЕ ПЕРЕХОДЫ

Светодиоды и полупроводниковые лазеры принадлежат к классу люминесцентных приборов. Люминесценцией называется оптическое излучение (в ультрафиолетовой, видимой или инфракрасной областях спектра), возникающее в результате электронного возбуждения материала, в отличие от теплового излучения, обусловленного только тепловым нагревом материала. На рис. 1 приведена диаграмма электромагнитного спектра. Несмотря на то что возбуждение излучения на разных длинах волн осуществляется разными методами, природа излучения во всех случаях одинакова. Излучение, видимое человеческим глазом, лежит в узкой области длин волн примерно от 0,4 до 0,7 мкм. На рис. 1 показаны основные цветовые зоны видимого излучения от фиолетовой до красной. Инфракрасная область простирается от 0,7 до ~1000 мкм, а ультрафиолетовая область включает длины волн от 0,4 до ~0,01 мкм (т. е. 100 Å, или 10 нм). В этой и последующих главах нас в основном будут интересовать области длин волн от близкой ультрафиолетовой (~0,3 мкм) до близкой инфракрасной (~1,5 мкм).

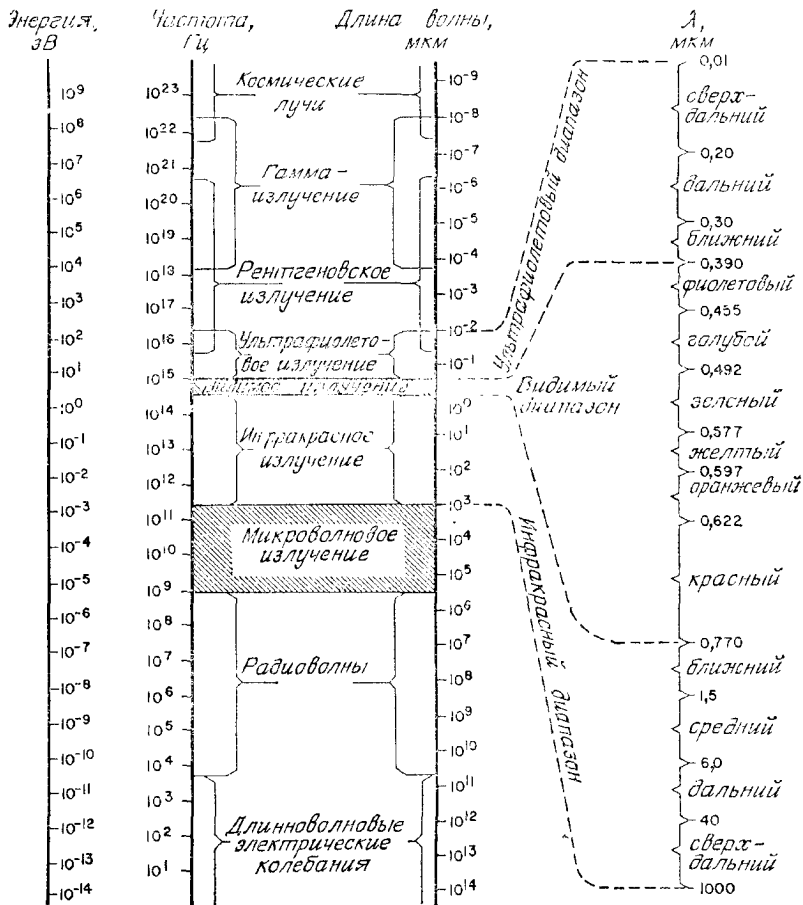


Рис. 1. Диаграмма электромагнитного спектра.

В зависимости от источника энергии возбуждения люминесценцию можно разделить [22] на: 1) фотолюминесценцию (возбуждение оптическим излучением); катодлюминесценцию (возбуждение под действием пучка электронов, или катодных лучей); 3) радиационную люминесценцию (возбуждение другими быстрыми частицами или излучением высокой энергии) и 4) электролюминесценцию (возбуждение электрическим полем или током). Мы будем рассматривать в основном электролюминесценцию, и прежде всего инжекционную электролюминесценцию, которая представляет собой оптическое излучение, возникающее при инжекции неосновных носителей в область полупроводникового $p-n$ -перехода, где имеются излучательные переходы.

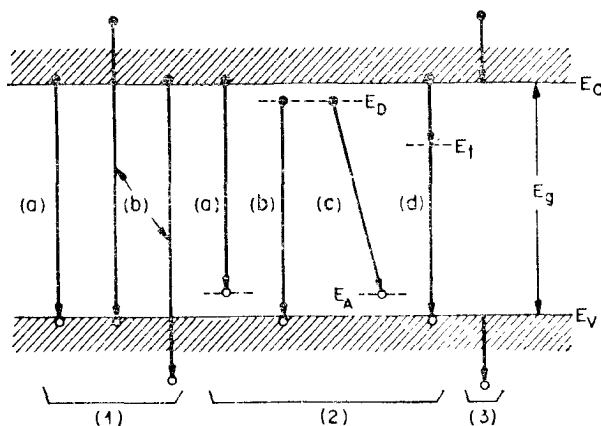


Рис. 2. Основные переходы в полупроводнике [23].

● электрон; ○ дырка.

На рис. 2 схематически представлены основные переходы в полупроводнике [23]. Эти переходы можно классифицировать следующим образом. Во-первых, это междузонные переходы, вызывающие: а) собственное излучение с энергией, очень близкой к ширине запрещенной зоны, которое может сопровождаться возбуждением фононов или экситонов; б) излучение с более высокой энергией с участием энергичных или «горячих» носителей, которое иногда может быть связано с лавинным пробоем. Во-вторых, это переходы с участием химических примесей или физических дефектов: а) между зоной проводимости и акцепторным уровнем; б) между донорным уровнем и валентной зоной; в) между донорным и акцепторным уровнями (межпримесное излучение); г) через глубокие уровни. В-третьих, это внутрizonные переходы, которые вызывают излучение, называемое иногда тормозным, и которые протекают с участием «горячих» носителей.

Не все переходы могут возникать в одном и том же материале или при одних и тех же условиях и не все из них являются излучательными. Эффективным с точки зрения люминесценции является такой материал, в котором излучательные переходы преобладают над безызлучательными (к последним относятся, например, переходы при оже-рекомбинации) [2].

12.2.1. Спектры излучения

Существуют три типа взаимодействия между фотонами и электронами в твердом теле. Фотон может поглотиться в результате перехода электрона из заполненного состояния валентной зоны в свободное состояние зоны проводимости. Кроме того, он может

стимулировать излучение подобного себе фотона, вызывая переход электрона из заполненного состояния в зоне проводимости в свободное состояние в валентной зоне. Могут также возникать спонтанные обратные переходы электронов из зоны проводимости на свободные состояния в валентной зоне, вызывающие испускание фотонов.

Интенсивность спонтанного излучения зависит от плотности заполненных состояний в зоне проводимости и плотности свободных состояний в валентной зоне [20]:

$$I(h\nu) \sim \nu \langle M \rangle^2 N_c N_v F_c(E) F_v(E), \quad (1)$$

где $\langle M \rangle$ — матричный элемент перехода; N_c — плотность состояний в зоне проводимости; N_v — плотность состояний в валентной зоне; $F_c(E)$ и $F_v(E)$ — функции распределения Ферми — Дирака для электронов и дырок соответственно.

Спектр спонтанного излучения обычно имеет вид [24]

$$I(h\nu) \sim \nu^2 (h\nu - E_g)^{1/2} \exp[-(h\nu - E_g)/kT], \quad (2)$$

где E_g — ширина запрещенной зоны.

Следует отметить, что при характерных для светодиодов и полупроводниковых лазеров концентрациях примеси представление локализованных примесных уровней, отстоящих от краев запрещенной зоны, уже неприменимо. В этом случае случайное распределение заряженных примесей в кристалле приводит к флуктуациям потенциала, которые создают на краях запрещенной зоны хвосты плотности состояний валентной зоны и зоны проводимости (разд. 12.4.3).

По мере увеличения концентрации примеси энергия ионизации донорных и акцепторных уровней уменьшается до нуля, уровни ионизируются (предполагается, что они не образуют комплексов), и концентрация свободных носителей уже не зависит от температуры. Поэтому в примесном полупроводнике с низкой степенью компенсации концентрация свободных электронов n_0 или дырок p_0 будет равна концентрации доноров или акцепторов соответственно. Значение концентрации, при которой энергия ионизации приближается к нулю, находится из условия

$$r/a^* \approx 3,0. \quad (3)$$

В уравнении (3) величина r составляет среднее расстояние между примесными атомами, равное

$$r \equiv \left(\frac{3}{4\pi n} \right)^{1/3}. \quad (4)$$

Здесь n — концентрация примеси и a^* — радиус водородоподобного примесного центра, равный

$$a^* = \frac{1}{2} \left(\frac{\epsilon_s}{\epsilon_0} \right) \left(\frac{m_0}{m^*} \right), \quad (5)$$

где m^* — эффективная масса электрона или дырки соответственно для донорного или акцепторного полупроводника.

В основе стандартной теории оптических междузонных переходов лежит так называемое правило k -отбора. Волновой вектор k_1 , соответствующий волновой функции валентной зоны, и волновой вектор k_2 , соответствующий волновой функции зоны проводимости, должны различаться на волновой вектор фотона, т. е. матричный элемент $\langle M \rangle$ равен нулю. Поскольку волновой вектор электрона существенно превосходит волновой вектор фотона, правило k -отбора обычно записывается в виде равенства

$$k_1 = k_2. \quad (6)$$

Разрешенными являются переходы, при которых начальное и конечное состояния характеризуются одинаковыми волновыми векторами; такие переходы называются «прямыми» или «вертикальными». Если минимум зоны проводимости и максимум валентной зоны не соответствуют одному и тому же значению вектора k , то для сохранения квазиимпульса при переходах необходимо участие фонона; эти переходы называются «непрямыми». Введение в полупроводник примесей приводит к возмущению волновых функций и матричных элементов перехода, вследствие чего правило k -отбора строго не выполняется. Поэтому в обычно рас-

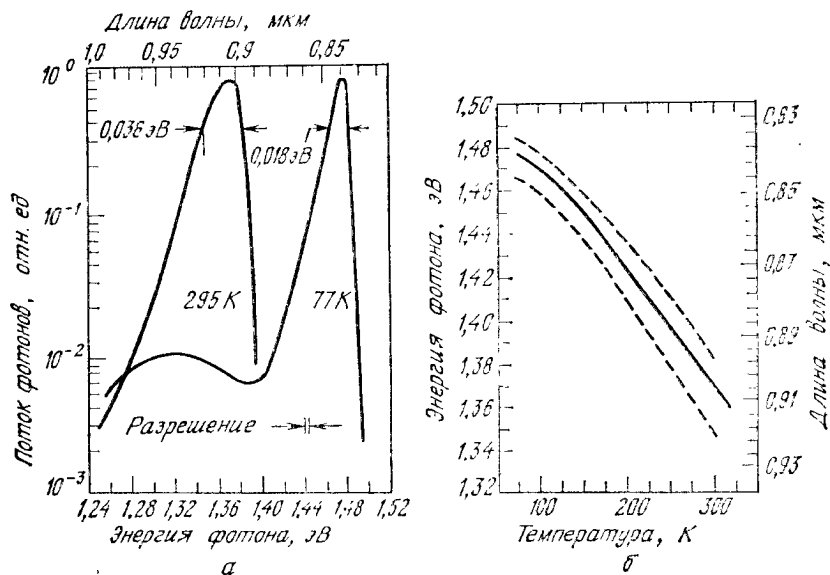


Рис. 3. Спектры излучения диода на GaAs (а) при 295 и 77 К и температурная зависимость энергии максимума (сплошная кривая) и уровней полуширины спектра излучения (штриховые кривые) (б) [25].

смаатриваемых случаях наблюдаются «запрещенные» переходы, и матричный элемент перехода зависит от энергии фотона.

На рис. 3, а представлены спектры излучения $p-n$ -перехода на основе GaAs, наблюдаемые при температурах 77 и 295 К [25]. Уменьшение энергии фотона, соответствующей максимуму спектра, при возрастании температуры обусловлено в основном сужением запрещенной зоны. На рис. 3, б приведена температурная зависимость энергии максимума и уровней полуширины спектра излучения диода. Полуширина спектра слабо увеличивается с ростом температуры, как это и следует ожидать из уравнения (2).

12.2.2. Эффективность люминесценции

При заданной энергии возбуждения наряду с излучательной рекомбинацией протекают конкурирующие безызлучательные процессы (рис. 4) [26]. Квантовая эффективность люминесценции определяется как отношение числа возбужденных носителей, дающих вклад в излучение, к полному числу носителей, участвующих в рекомбинации, и может быть выражена через времена жизни следующим образом:

$$\eta_q = R_r/R = \tau_{nr}/(\tau_{nr} + \tau_r), \quad (7)$$

где τ_r и τ_{nr} — времена жизни излучательной и безызлучательной рекомбинации соответственно, а R_r и R — скорости излучательной и полной рекомбинации. Для полупроводниковых слоев p -типа проводимости скорость рекомбинации и время жизни связаны соотношением

$$R = (n - n_0)/\tau. \quad (8)$$

Аналогично для слоев n -типа

$$R = (p - p_0)/\tau, \quad (9)$$

где n_0 и p_0 — концентрации электронов и дырок в состоянии теплового равновесия, а n и p — электронная и дырочная кон-

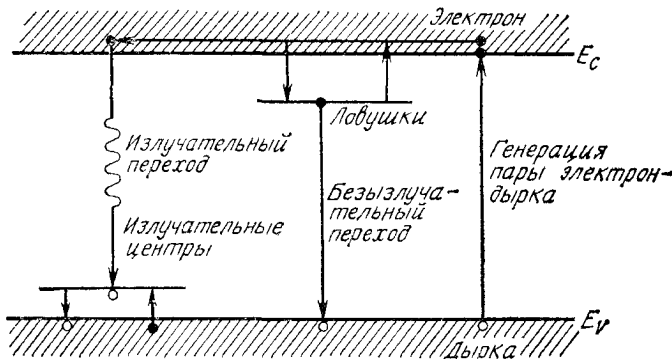


Рис. 4. Процессы излучательной и безызлучательной рекомбинации [26].

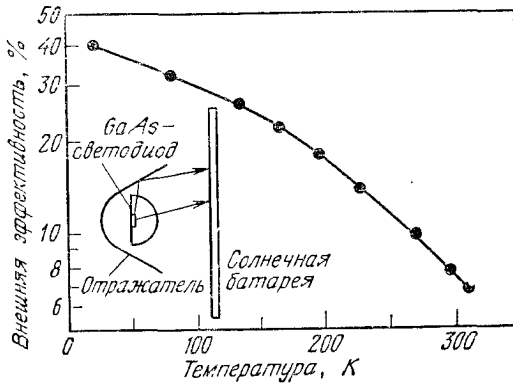


Рис. 5. Зависимость квантовой эффективности GaAs-светодиода от температуры. На вставке показано поперечное сечение устройства, используемого для измерения эффективности [25].

центрации при оптическом возбуждении. Время жизни τ неосновных носителей дается выражением

$$\tau = \tau_r \tau_{nr} / (\tau_{nr} + \tau_r). \quad (10)$$

Из уравнения (7) видно, что для обеспечения высокой квантовой эффективности время жизни τ_r должно быть мало.

Вставка на рис. 5 иллюстрирует схему измерения эффективности. Измерение излучения производится непосредственно по току короткого замыкания солнечной батареи. Внешняя эффективность, измеренная как функция от температуры, представлена на рис. 5 [25]. С ростом температуры эффективность уменьшается и имеет значения около 40, 32 и 7 % при температурах 20, 77 и 300 К соответственно. Уменьшение эффективности обусловлено уменьшением отношения τ/τ_r с ростом температуры.

12.2.3. Методы возбуждения

К способам, которыми можно возбуждать электролюминесценцию, относятся междузонное (собственное) возбуждение, лавинное возбуждение и возбуждение при туннелировании и инжекции. Междузонное возбуждение обеспечивается при воздействии переменного электрического поля на порошкообразный полупроводник (например, ZnS), впрессованный в диэлектрик (пластмассу или стекло). Электролюминесценция обычно имеет место при частотах, лежащих в звуковом диапазоне. Ее эффективность обычно низка ($\leq 1\%$). Основными механизмами возбуждения при этом являются ударная ионизация ускоренных электронов или полевая эмиссия электронов с ловушек [23, 27].

Для того чтобы создать лавинное возбуждение, $p - n$ -переход или контакт металл — полупроводник смещают в обратном направлении в область лавинного пробоя. Электронно-дырочные пары, возникающие при ударной ионизации, могут вызвать излучение за счет междузонных (лавинное излучение) или внутризонных (тормозное излучение) переходов. Электролюминесценция может быть также результатом туннелирования в область прямо-смещенного или обратнoсмещенного перехода. При достаточно большом обратном смещении на контакте металл — полупроводник (вырожденный, p -типа) дырки с уровня Ферми в металле могут туннелировать в валентную зону и вызывать излучательную рекомбинацию с электронами, туннелирующими из валентной зоны в зону проводимости [28].

Наиболее важным методом возбуждения электролюминесценции является инжекционный [29]. При прямом смещении на $p - n$ -переходе инжекция неосновных носителей через переход может привести к весьма эффективной излучательной рекомбинации, так как в этом случае электрическая энергия непосредственно преобразуется в фотоны. В последующих разделах мы в первую очередь будем рассматривать инжекционные электролюминесцентные приборы.

12.3. СВЕТОДИОДЫ

Светодиоды — это $p - n$ -переходы, которые при прямом смещении могут испускать спонтанное излучение в ультрафиолетовой, видимой и инфракрасной областях электромагнитного спектра. Мы рассмотрим сначала светодиоды видимого участка спектра, которые широко применяются в информационных каналах, обеспечивающих связь электронной аппаратуры с ее пользователями, а затем инфракрасные светодиоды, которые эффективно используются в оптронах и представляют собой перспективные источники света для волоконно-оптических линий связи.

12.3.1. Светодиоды видимого диапазона

Эффективность воздействия света на глаз человека определяется функцией относительной видности глаза $V(\lambda)$, зависящей от длины волны. На рис. 6 приведена функция относительной видности глаза при угле зрения 2° , определенная Международной комиссией по освещению (МКО) [2]. Чувствительность глаза максимальна на длине волны $0,555 \text{ мкм}$ ($V(0,555) = 1,0$) и спадает до нуля при длинах волн $0,39$ и $0,77 \text{ мкм}$, соответствующих краям видимого диапазона спектра. Эффективность использования лучистой энергии при зрительном восприятии характеризуется яркостью лучистой энергии. Для стандартного фотометрического

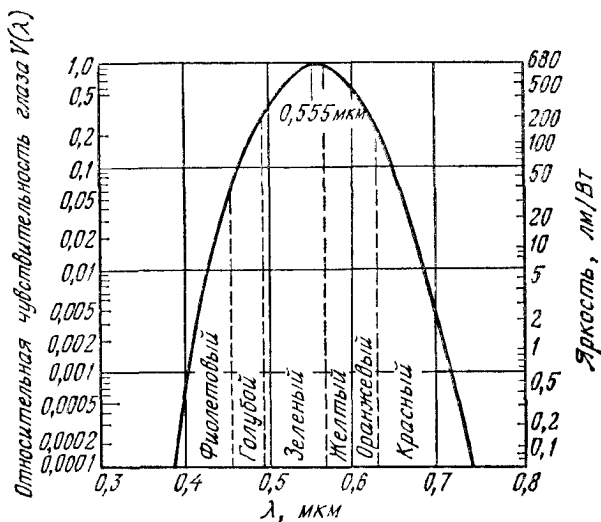


Рис. 6. Функция относительной видности глаза, определенная Международной комиссией по освещению (МКО) для стандартного фотометрического наблюдателя. Обозначены основные цветовые зоны видимого излучения.

наблюдателя 1 Вт лучистой энергии в максимуме чувствительности глаза (0,555 мкм) соответствует 680 лм. На рис. 6 показаны шесть основных цветовых участков видимого диапазона от фиолетового до красного.

Материалы для светодиодов видимого диапазона. Поскольку глаз чувствителен только к свету с энергией $h\nu \geq 1,8$ эВ ($\sim 0,7$ мкм), то полупроводники, которые могут быть использованы для создания светодиодов видимого диапазона, должны иметь ширину запрещенной зоны больше этого значения. На рис. 7 приведен перечень полупроводников, представляющих интерес с этой точки зрения. Заштрихованные прямозонные полупроводники особенно важны для электролюминесцентных приборов, так как излучательная рекомбинация в них является процессом первого порядка (протекает без участия фононов) и ожидаемая квантовая эффективность должна быть значительно более высокой, чем для полупроводников с непрямой зоной, в которых при переходах возбуждаются фононы. Среди всех приведенных на рис. 7 полупроводников наибольший интерес представляет $\text{GaAs}_{1-x}\text{P}_x$, относящийся к соединениям типа $A^{III} - BV$.

На рис. 8, а приведена зависимость ширины запрещенной зоны $\text{GaAs}_{1-x}\text{P}_x$ от молярного содержания компонент x [20]. В диапазоне $0 < x < 0,45$ запрещенная зона прямая, а ее ширина увеличивается от $E_g = 1,424$ эВ при $x = 0$ до $E_g = 1,977$ эВ при

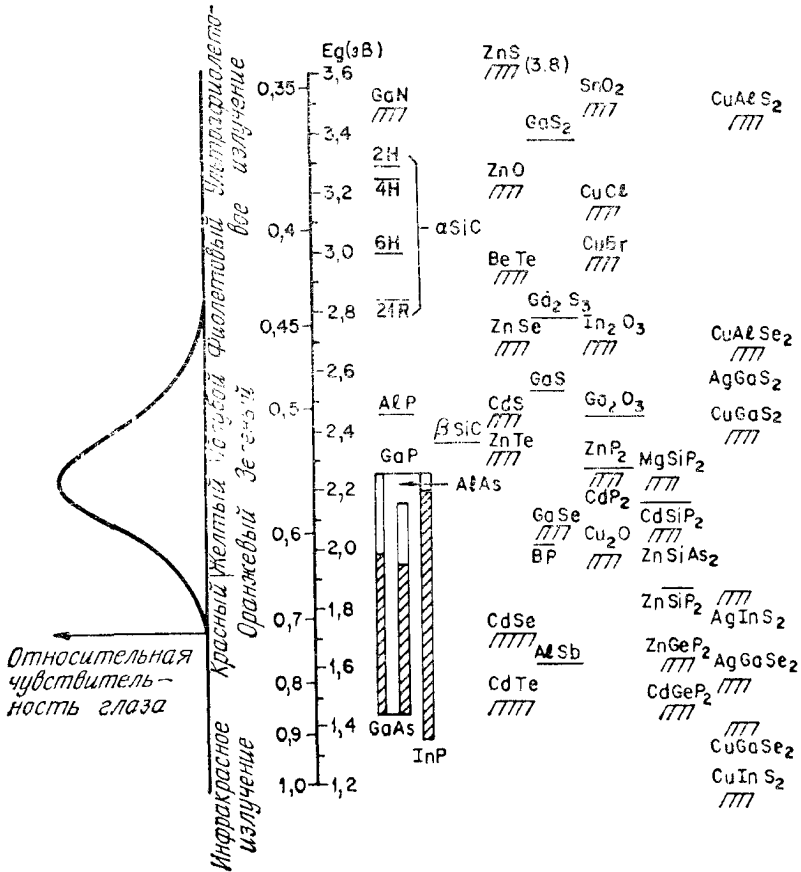


Рис. 7. Полупроводники, на основе которых могут быть созданы светодиоды со спектром излучения, включающим диапазон видимого света [2].

$x = 0,45$. При $x > 0,45$ полупроводник становится непрямозонным. На рис. 8, б приведена зависимость энергии от волнового вектора для нескольких значений x [30], из которой следует, что зона проводимости имеет два минимума. Те минимумы, которые расположены в точке Г, являются прямыми, тогда как другие, расположенные на оси x , не прямые. Электроны в прямом минимуме зоны проводимости и дырки в максимуме валентной зоны обладают одинаковыми квазиимпульсами; электроны в непрямом минимуме имеют другое значение квазиимпульса. Для таких прямозонных полупроводников, как GaAs и GaAs_{1-x}P_x ($x \leq 0,45$), квазиимпульс при междозонных переходах сохраняется, поэтому эти переходы характеризуются высокой степенью вероятности.

При этом энергия фотона приблизительно равна ширине запрещенной зоны полупроводника. В прямозонных материалах процесс излучательной рекомбинации является доминирующим. В то же время для $\text{GaAs}_{1-x}\text{P}_x$ при $x > 0,45$ и GaP , у которых запрещенная зона не прямая, вероятность междозонных переходов чрезвычайно мала, поскольку в этом случае для преобразования квазиимпульса при переходе требуется участие фононов или других факторов рассеяния. Поэтому для усиления излучательных процессов в непрямозонных полупроводниках, таких, например, как GaP , специально создаются рекомбинационные центры.

Эффективные центры излучательной рекомбинации в $\text{GaAs}_{1-x}\text{P}_x$ могут быть созданы путем внедрения специальных примесей, например азота [31]. Азот, введенный в полупроводник, замещает атомы фосфора в узлах решетки. Азот и фосфор имеют одинаковую внешнюю электронную структуру (оба относятся к V группе элементов периодической системы), а структуры их внутренних оболочек сильно различаются. Это приводит к возникновению вблизи зоны проводимости электронного уровня захвата. Полученный таким образом рекомбинационный центр называется изоэлектронным центром. В нормальном состоянии изоэлектронные центры нейтральны. В материале p -типа инжектированный электрон сначала захватывается на центр. Заряженный отрица-

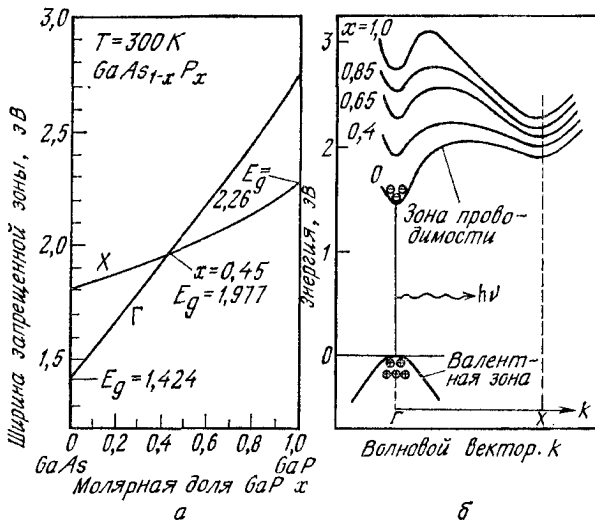


Рис. 8. Зависимость ширины энергетических зазоров (а) для прямого (Γ) и непрямого (X) минимумов зоны проводимости от состава $\text{GaAs}_{1-x}\text{P}_x$ [20] и энергетическая зонная структура $\text{GaAs}_{1-x}\text{P}_x$ (б). Значения состава соответствуют красному ($x = 0,4$), оранжевому ($0,65$), желтому ($0,85$) и зеленому свету ($1,0$) [30].

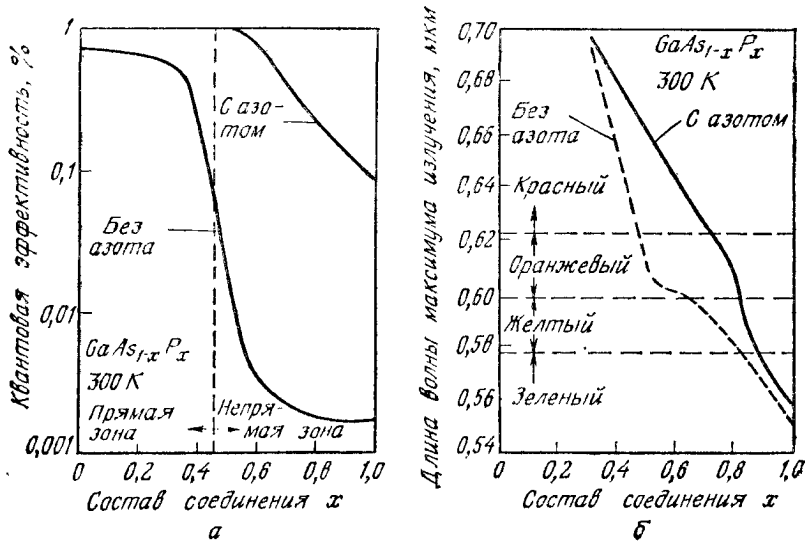


Рис. 9. Зависимость квантовой эффективности (а) и длины волны, соответствующей максимуму излучения (б), от состава соединения при наличии и в отсутствие изоэлектронной примеси азота [31].

тельно центр затем захватывает дырку из валентной зоны, формируя связанный экситон. Последующая аннигиляция этой электронно-дырочной пары приводит к рождению фотона с энергией, примерно равной разности между шириной запрещенной зоны и энергией связи центра. Так как захваченный электрон сильно локализован на центре, его импульс рассеивается. Таким образом обеспечивается преобразование квазиимпульса, вследствие чего вероятность прямого перехода существенно возрастает. В непрямозонных материалах, таких, как GaP, описанный механизм излучательной рекомбинации является преобладающим.

На рис. 9, а приведена зависимость квантовой эффективности от состава $GaAs_{1-x}P_x$ при наличии и в отсутствие изоэлектронной примеси азота [31]. Резкое уменьшение эффективности излучения $GaAs_{1-x}P_x$ без примеси азота в диапазоне $0,4 < x < 0,5$ обусловлено изменением механизма рекомбинации: при $x = 0,45$ преобладающими становятся непрямые переходы. Наличие азота обеспечивает более высокие значения эффективности при $x > 0,45$, однако последняя все же постепенно уменьшается с ростом x вследствие увеличения разности глубин прямой и непрямой запрещенных зон (рис. 8, б). Кроме того, длина волны максимума излучения для соединений, легированных азотом, смещается под влиянием энергии связи изоэлектронного центра (рис. 9, б).

Конструкции и характеристики светодиодов. Среди светодиодных структур основной является структура с плоской геометрией (рис. 10). Обычно прямозонные светодиоды (красное излучение) формируются на подложках GaAs (рис 10, а), тогда как непрямозонные (оранжевое, желтое и зеленое излучения) — на подложках GaP (рис 10, б). При использовании подложки GaAs на нее эпитаксиально наращивается переходный слой $\text{GaAs}_{1-x}\text{P}_x$ переменного состава с x , изменяющимся в пределах 0—0,4, а затем слой $\text{GaAs}_{1-x}\text{P}_x$ с постоянным составом. Переходная область ограничивает образование безызлучательных центров, обусловленных различием решеток. Фотоны, генерируемые в области перехода, испускаются во всех направлениях, однако наблюдателя достигает лишь та их часть, которая проходит через поверхность.

Уменьшение количества излучаемых светодиодом фотонов обусловлено поглощением в материале светодиода, потерями за счет отражения и потерями за счет полного внутреннего отражения. Потери, связанные с поглощением, весьма существенны в светодиодах на подложках GaAs (рис. 10, а), так как в этом случае

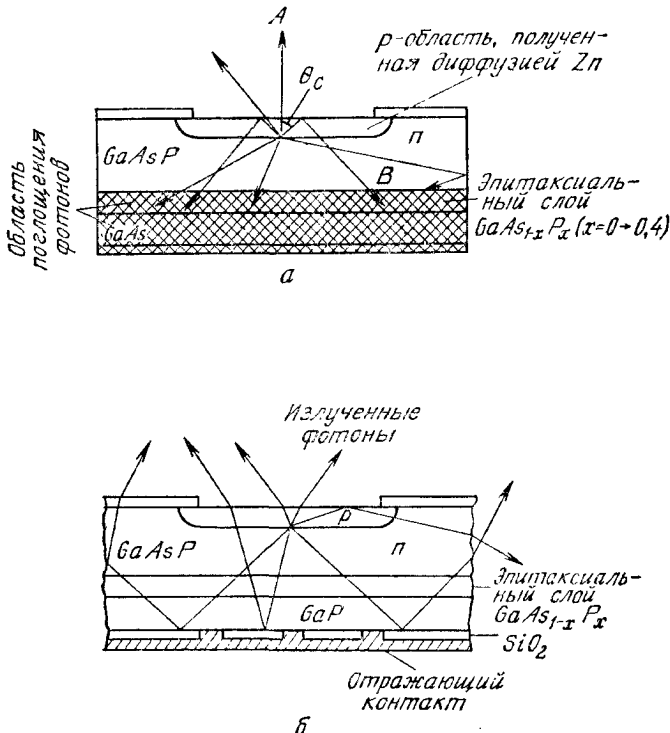


Рис. 10. Прохождение света, излучаемого переходом в структурах с непрозрачной (а) и прозрачной подложками (б) [3].

подложка непрозрачна для света и поглощает примерно 85 % фотонов, излучаемых переходом. В светодиодах на подложках GaP (рис. 10, б) фотоны, испускаемые в направлении тылового контакта, могут отражаться от него, при этом поглощение составляет $\sim 25\%$, и эффективность излучения может быть существенно увеличена. При переходе фотонов из среды с показателем преломления \bar{n}_2 (например, из GaAs с $\bar{n}_2 = 3,66$) в среду с показателем преломления \bar{n}_1 (например, в воздух с $\bar{n}_1 = 0,1$) часть света претерпевает отражение от границы раздела сред в обратном направлении. Эти потери называются потерями отражения. При нормальном падении света коэффициент отражения равен

$$R = \left(\frac{\bar{n}_2 - \bar{n}_1}{\bar{n}_2 + \bar{n}_1} \right)^2. \quad (11)$$

Третий фактор обусловлен полным внутренним отражением света, падающего на границу раздела под углом, больше критического угла θ_c , определяемого выражением

$$\theta_c = \sin^{-1}(\bar{n}_1/\bar{n}_2). \quad (12)$$

Для GaAs критический угол составляет $\sim 16^\circ$, а для GaP с $\bar{n}_2 = 3,45$ он равен $\sim 17^\circ$.

При плоской геометрии (рис. 10, а) луч *A* ослабляется вследствие поглощения, а также частично из-за потерь на отражение. Лучи *B*, которые приходят к границе раздела полупроводник — воздух под углами $\theta \geq \theta_c$, претерпевают полное внутреннее отражение. Полная эффективность преобразования электрического сигнала в оптический дается выражением [32]

$$\eta_F = \frac{q}{P} (1 - R) (1 - \cos \theta_c) \frac{\int \Phi(\lambda) (1 + R_1 e^{-2\alpha_1(\lambda) x_1}) e^{-\alpha_2(\lambda) x_2} d\lambda}{\int \Phi(\lambda) d\lambda}, \quad (13)$$

$$\eta_F \approx \frac{4\bar{n}_2\bar{n}_1}{(\bar{n}_2 + \bar{n}_1)^2} (1 - \cos \theta_c), \quad (14)$$

где P — мощность на входе; $4\bar{n}_2\bar{n}_1/(\bar{n}_2 + \bar{n}_1)^2$ — коэффициент передачи, равный $4\bar{n}_2/(\bar{n}_2 + 1)^2$ для границы раздела полупроводник — воздух; $(1 - \cos \theta_c)$ — телесный угол; $\Phi(\lambda)$ — скорость генерации фотонов в единицах фотон/(с·см²); R_1 — коэффициент отражения от тылового контакта; $\alpha(\lambda)$ и x — соответственно коэффициенты поглощения и толщина *p*- и *n*-областей прибора.

На рис. 11 показаны поперечные разрезы других светодиодов, которые имеют полусферическую, усеченно сферическую и параболическую геометрию [33]. Аналогичные выражения для эффективности могут быть получены также и для этих конфигураций. Основное отличие этих трех структур от структуры с плоской

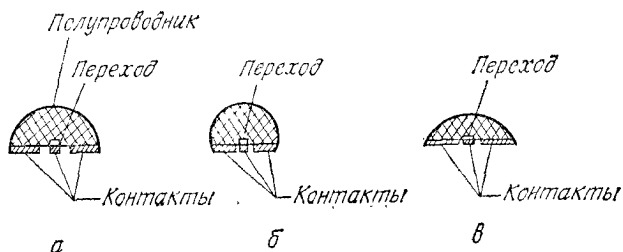


Рис. 11. Разрез трех светодиодов [33].
 а — полусфера; б — усеченная сфера; в — параболоид.

геометрией (рис. 10, а) состоит в том, что телесный угол для них равен 1. Таким образом, отношение эффективностей равно

$$\frac{\eta}{\eta_F} \approx \frac{1}{1 - \cos \theta_c} = \frac{1}{1 - \sqrt{1 - 1/\bar{n}_2^2}}, \quad (15)$$

$$\eta/\eta_F = 2\bar{n}_2^2 \text{ при } \bar{n}_2 \gg 1. \quad (16)$$

Это означает, что для структур на GaP с $\bar{n}_2 = 3,45$ при данной геометрии можно ожидать увеличения эффективности на порядок величины (рис. 11). На рис. 12 приведены типичные диаграммы направленности для плоского, полусферического и параболического диодов [34]. Очевидно, что требуемая диаграмма направленности может быть получена за счет изменения геометрии прибора.

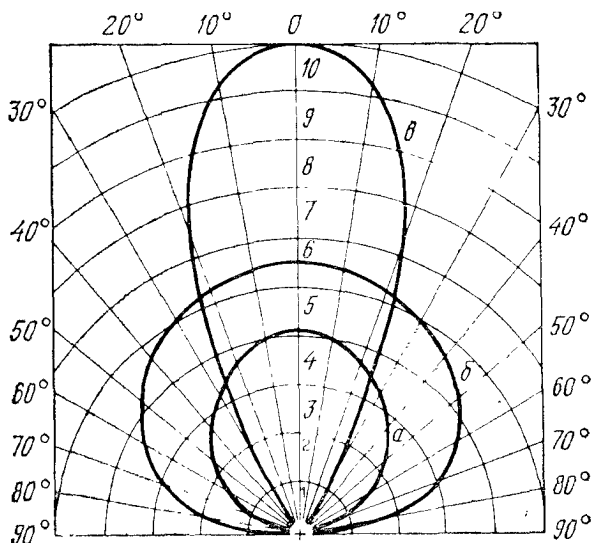


Рис. 12. Диаграмма направленности излучения светодиодов плоской (а), полусферической (б) и параболической (в) геометрии [34].

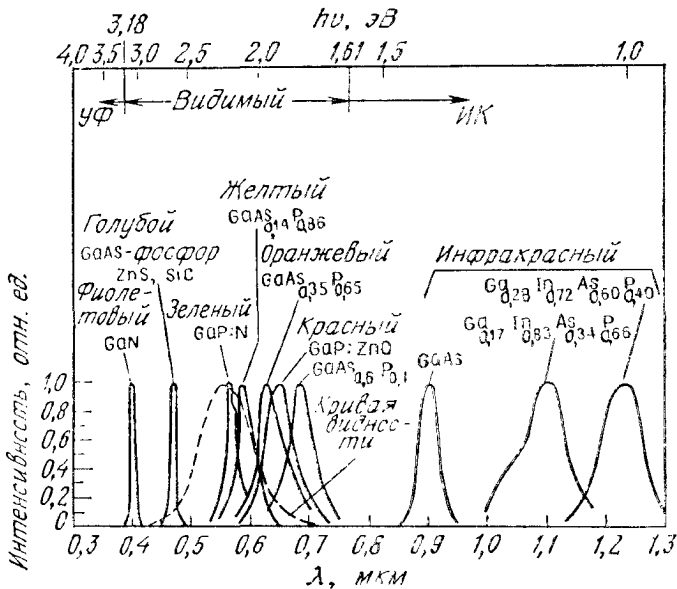


Рис. 13. Относительные спектры излучения различных светодиодов видимого и инфракрасного диапазонов.

На рис. 13 представлены относительные спектры излучения некоторых типичных светодиодов при комнатной температуре. Инфракрасные светодиоды из GaAs и соединений типа $A^{III}B^V$ будут рассмотрены в разделе 12.3.2. Красные светодиоды могут быть изготовлены из прямозонного соединения $GaAs_{0,6}P_{0,4}$, а также на основе GaP с примесью ZnO, которая, обладая таким же числом общих валентных электронов, как и пары соседних атомов галлий — фосфор, замещает их в решетке полупроводника [2]. ZnO является одной из примесей, формирующих изоэлектронные центры. Уровень рекомбинации, обусловленный этой примесью, расположен примерно на 0,3 эВ ниже края зоны проводимости GaP, в результате чего испускаются кванты с энергией 1,95 эВ. Светодиоды, излучающие оранжевый, желтый и зеленый свет [35], создаются на основе непрямозонных соединений $GaAs_{1-x}P_x$ или GaP, легированных изоэлектронной примесью азота. Светодиоды, излучающие голубой свет, могут быть изготовлены из соединений ZnS, SiC [36], а также в виде преобразователей инфракрасного излучения в видимое [37]. Рис. 14, а иллюстрирует основную структуру преобразователя, в котором инфракрасное излучение светодиода GaAs поглощается слоем фосфора с добавками ионов редкоземельных элементов, например иттербия (Yb^{3+}) или эрбия (Er^{3+}). Характеристики преобразования зависят от

процесса последовательного поглощения двух фотонов в инфракрасной области спектра, сопровождающегося испусканием одного фотона в видимой области спектра (рис. 14, б). На рис. 13 в фиолетовой части спектра расположен спектр излучения светодиода с $p-n$ -переходом на основе GaN. Отметим, что с увеличением длины волны λ_0 , соответствующей максимуму спектра излучения, увеличивается также и полуширина спектра. Это обусловлено тем, что ширина спектра спонтанного излучения пропорциональна λ_0^2 , как и следует из уравнения (2).

На рис. 13 приведена также характеристика чувствительности глаза (штриховая кривая) с максимумом на длине волны 0,555 мкм (зеленый свет). По мере смещения длины волны в красную область видимого спектра чувствительность глаза быстро падает. Спектральная характеристика чувствительности глаза должна учитываться при оценках эффективности визуального восприятия излучения от различных светодиодов с известными значениями энергии излучения. Для светодиодов с высокой эффективностью и широким спектром яркостный эквивалент излучения дается выражением [29]:

$$\text{Яркость} = L_0 \int V(\lambda) P(\lambda) d\lambda \quad [\text{лм/Вт}], \quad (17)$$

где L_0 — максимальное значение яркости, равное 680 лм/Вт; $V(\lambda)$ — функция относительной видности глаза; $P(\lambda)$ — спектр излучения. Например, как показано на рис. 15, максимум красного излучения диода на основе GaP — ZnO лежит на длине волны 0,69 мкм. Максимум кривой, соответствующей суперпозиции спектра излучения с кривой видности, смещается к значению $\lambda = 0,65$ мкм и соответствует яркостному эквиваленту, равному 15 лм/Вт.

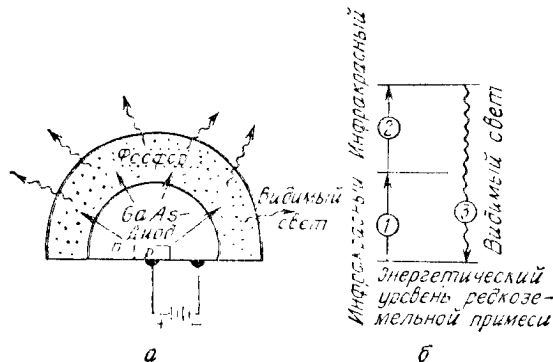


Рис. 14. Преобразователь, в котором излучение GaAs-диода возбуждает испускание света фосфором (а) и энергетические уровни преобразования (б) [37].

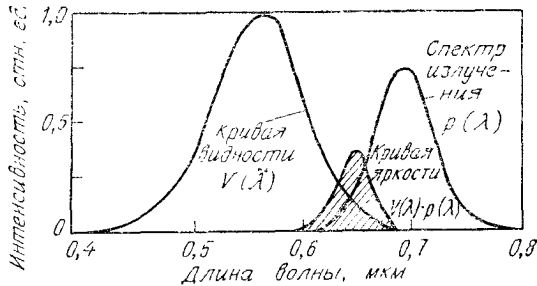


Рис. 15. Оценка яркостного эквивалента излучения [29].

Светодиоды, излучающие в видимой области спектра, могут использоваться в качестве индикаторов и дисплеев, а также в оптронах [3]. На рис. 16 приведены схемы конструкций различных светодиодных источников [2]. Источник содержит кристалл со светодиодом и пластмассовую линзу, которая обычно окрашивается с целью усиления контраста и фильтрации излучения. Источники, показанные на рис. 16, а и б, выполнены на базе обычных транзисторных и диодных корпусов. В источнике, изображенном на рис. 16, в, используется базовый держатель в виде металлических проводников, на которых размещены светодиод и последовательное сопротивление. Пластмассовая линза задает диаграмму направленности и угол зрения прибора. Конструкция корпуса на рис. 16, г удовлетворяет требованиям для размещения прозрачных светодиодных кристаллов, например, таких, как GaP, которые излучают через все пять своих граней.

На рис. 17 приведены основные форматы светодиодных дисплеев. Для отображения цифр от 0 до 9 обычно применяются 7-сегментные дисплеи или светодиодные матрицы с числом элементов 3×5 . В алфавитно-цифровых дисплеях (0 — 9 и A — Z) используются форматы, содержащие 14 сегментов или 5×7 элементов. Дисплеи можно изготавливать на базе интегральной технологии так же, как кремниевые монолитные схемы, или используя гибридную сборку, в ходе которой на отражателе формируются сегменты из отдельных светодиодных чипов.

Светодиоды могут использоваться в оптронах, которые обеспечивают электрическую развязку между входным или управляющим сигналом и выходом [2, 3]. На рис. 18, а показана схема оптрона, содержащего светодиод в качестве источника света и фотодиод в качестве детектора. Излучение, генерируемое светодиодом при подаче на него входного электрического сигнала, детектируется фотодиодом. Затем излучение вновь преобразуется в электрический сигнал в виде тока, который протекает через сопротивление нагрузки. Типичное значение коэффициента пре-

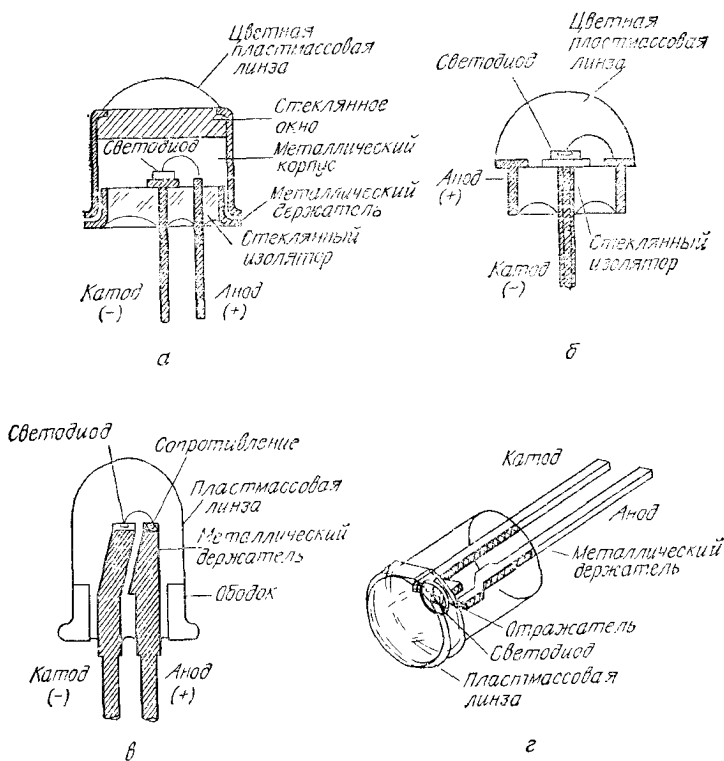


Рис. 16. Схемы конструкций некоторых светодиодных излучателей [2].

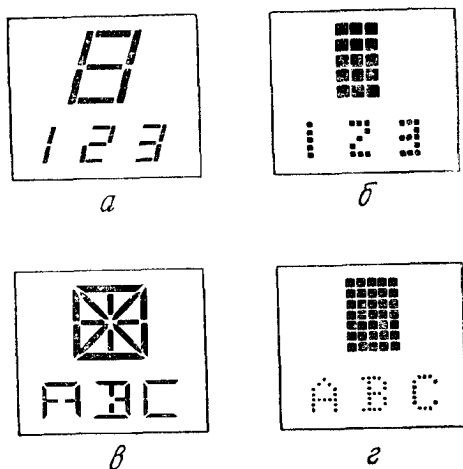


Рис. 17. Форматы цифровых и буквенных светодиодных дисплеев.

а — 7-сегментный (цифровой); б — матричный 3 × 5 (цифровой); в — 14-сегментный (алфавитно-цифровой); г — матричный (алфавитно-цифровой).

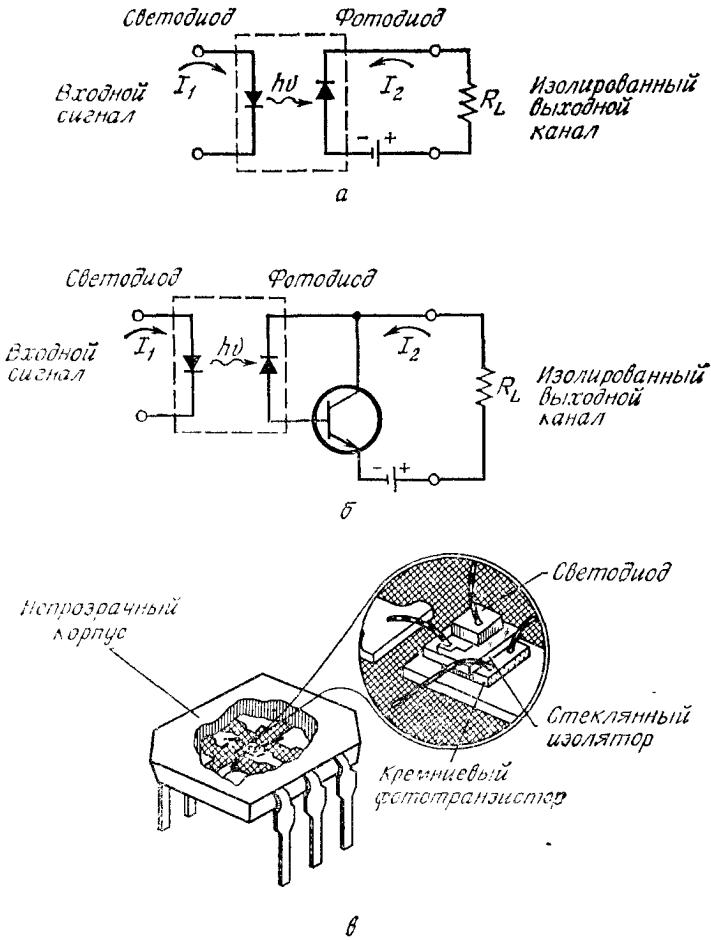


Рис. 18. Оптоны [2].

а — схема; **б** — оптрон с высоким коэффициентом преобразования; **в** — оптрон, расположенный на одной контактной площадке.

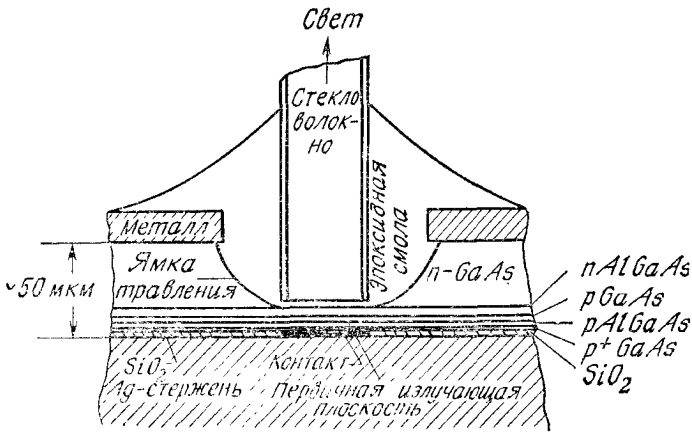
образования по току I_2/I_1 составляет $\sim 10^{-3}$. Для усиления коэффициента преобразования в оптронах в качестве детекторов используются фототранзисторы, при этом отношение токов может быть повышено до $0,1 : 1$. На рис. 18, **в** показана оптронная структура, размещенная в одном корпусе. В таких приборах осуществляется оптическая передача сигналов со скоростью света и обеспечивается электрическая изоляция, так как отсутствует обратная связь с выхода на вход.

12.3.2. Инфракрасные светодиоды

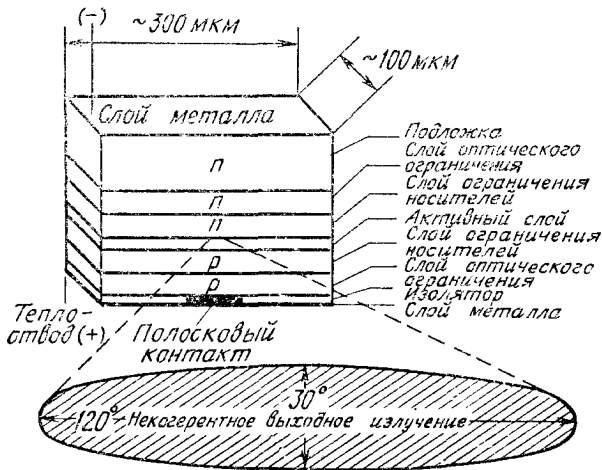
Инфракрасные светодиоды, как и светодиоды видимого излучения, имеют различную конфигурацию (рис. 10 и 11). На рис. 13 среди других спектров приведен также спектр излучения светодиода на основе GaAs. Этот инфракрасный источник в настоящее время обладает наибольшей эффективностью электролюминесценции в основном благодаря тому, что среди всех прямозонных полупроводников GaAs является технологически наиболее освоенным. Одним из важных применений диодов на основе GaAs являются источники излучения в оптронах. Для изготовления инфракрасных светодиодов используются многие другие полупроводники, имеющие запрещенную зону шириной менее 1,5 эВ. К ним относятся твердые растворы, в состав которых входят три или четыре элемента III и V групп периодической системы. На рис. 13 приведены спектры излучения инфракрасных светодиодов из четырехэлементных полупроводников $Ga_{1-x}In_xAs_{1-y}P_y$ [38]. Структура запрещенной зоны и постоянные решетки этих полупроводников рассмотрены в разд. 12.4.

Инфракрасные светодиоды являются перспективными источниками для волоконно-оптических линий связи. Светодиоды по сравнению с лазерами имеют как недостатки, так и преимущества. К последним относятся высокая рабочая температура, меньшее влияние температуры на мощность излучения, а также простота конструкции и схемы питания. К недостаткам относится меньшая яркость, более низкие частоты модуляции и большая спектральная ширина линии излучения, типичное значение которой составляет 100—500 Å, тогда как лазеры характеризуются шириной линии 0,1—1 Å.

Различают два основных типа светодиодов, обеспечивающих ввод излучения в оптические волокна малого диаметра: светодиоды с излучающей поверхностью и светодиоды с излучающей гранью [39, 40]. В поверхностных излучателях (рис. 19, а) излучающая область перехода ограничена слоем окисла, а диаметр контакта обычно составляет 15—100 мкм. Для уменьшения потерь поглощения и обеспечения плотного контакта излучающей поверхности с торцом волокна слой полупроводника, через который проходит излучение, должны быть очень тонкими (10—15 мкм). В гетеропереходах (например, GaAs — $Al_xGa_{1-x}As$) мощность излучения на выходе может быть повышена за счет эффекта ограничения носителей, который возникает при окружении слоя излучательной рекомбинации (например, GaAs) областью полупроводника с более широкой запрещенной зоной (например, $Al_xGa_{1-x}As$). Подробно этот эффект рассмотрен в последующих разделах. Гетеропереход может также служить в качестве окна,



а



б

Рис. 19. Поверхностно-излучающий светодиод (а) на основе двойной гетероструктуры AlGaAs, имеющий малую площадь и высокую излучающую способность и совмещенный со стекловолокном [39], и светодиод с излучающей гранью на основе двойной гетероструктуры (б) [40].

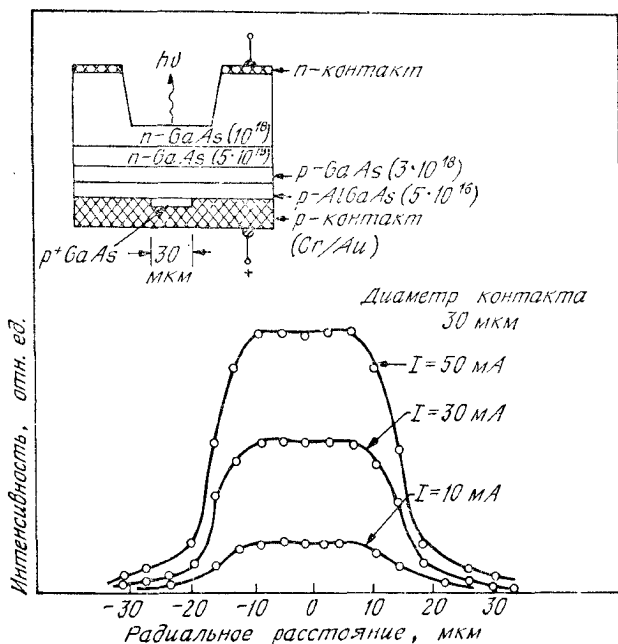


Рис. 20. Распределение интенсивности света вблизи излучающей поверхности светодиода. На вставке показана структура светодиода [41].

которое не поглощает излучение, возникающее в полупроводнике с меньшей запрещенной зоной.

При выводе через грань (рис. 19, б) излучение может быть сосредоточено в относительно направленном луче, что повышает эффективность связи светодиода с волокнами, имеющими малый приемный угол. Пространственное распределение излучения при этом аналогично распределению для лазеров на гетеропереходах, которые рассматриваются в разд. 12.4.

Вставка на рис. 20 иллюстрирует структуру поверхностного излучателя, в котором область излучения ограничена областью с высоким контактным сопротивлением [41]. Диаметр омического контакта к слою p^+ -GaAs составляет 30 мкм. Слой металлизации Cr/Au вокруг области омического контакта имеет высокое контактное сопротивление со слоем p -AlGaAs (концентрация легирования $5 \cdot 10^{16} \text{ см}^{-3}$). Поэтому основная часть тока протекает через область омического контакта. На рис. 20 приведено распределение интенсивности света вблизи излучающей поверхности. Видно, что диаметр области излучения ограничен размером 30 мкм, задаваемым конфигурацией контакта.

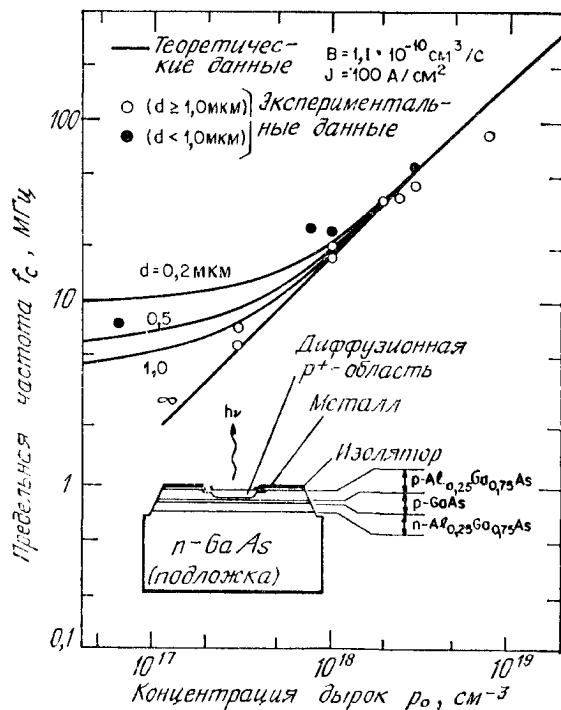


Рис. 21. Зависимость предельной частоты светодиода от концентрации дырок в активном слое [42].

Другим важным параметром, который должен учитываться при конструировании светодиодов для оптических систем связи, является диапазон рабочих частот. Как следует из уравнений (7) и (10), квантовая эффективность люминесценции и полное время жизни связаны с временем жизни излучательной рекомбинации. При внешнем возбуждении скорость полной излучательной рекомбинации определяется выражением [41a]

$$R_r = Gnp/n_0p_0, \quad (18)$$

где G — скорость полной термической генерации. Обычно это соотношение записывается следующим образом:

$$R_r = Bnp, \quad (19)$$

где B — константа излучательной рекомбинации, равная G/n_0p_0 . При достаточно низких уровнях возбуждения, таких, что в материале p -типа $p \approx p_0$, время жизни излучательной рекомбинации становится равным

$$\tau_r = (n - n_0)/R_r \approx 1/Bp_0, \quad (20)$$

а для материала n -типа, когда $n \approx n_0$,

$$\tau_r = (p - p_0)/R_r \approx 1/Vn_0. \quad (21)$$

Предельная частота светодиода

$$f_c = \frac{1}{2\pi\tau}. \quad (22)$$

Если $\tau_r < \tau_{nr}$, то τ в уравнении (22) становится близким к τ_r . Таким образом, как следует из уравнений (20) и (21), при увеличении концентрации τ_r уменьшается, а f_c растет. Результаты теоретического расчета и экспериментальные данные измерения зависимости предельной частоты от концентрации дырок для светодиода на основе гетероструктуры с выводом излучения через поверхность приведены на рис. 21 [42]. Видно, что экспериментально измеренная предельная частота возрастает с концентрацией p_0 , что находится в соответствии с уравнениями (10) и (20). Для достижения высоких значений f_c необходимо уменьшать толщину рекомбинационной области и увеличивать концентрацию носителей.

12.4. ФИЗИКА ПОЛУПРОВОДНИКОВЫХ ЛАЗЕРОВ

12.4.1. Общие замечания

Полупроводниковые лазеры, подобно другим лазерам (таким, как рубиновый лазер или же лазер на смеси He — Ne), испускают излучение, когерентное в пространстве и во времени. Это означает, что излучение лазера высокомонохроматично (имеет узкую полосу спектра) и создает строго направленный луч света. Вместе с тем по ряду важных характеристик полупроводниковые лазеры существенно отличаются от лазеров других типов:

1. В обычных лазерах квантовые переходы происходят между дискретными энергетическими уровнями, тогда как в полупроводниковых лазерах переходы обусловлены зонной структурой материала.

2. Полупроводниковые лазеры имеют очень малые размеры ($\sim 0,1$ мм в длину), и, так как активная область у них очень узкая (~ 1 мкм и меньше), расхождение лазерного луча значительно больше, чем у обычного лазера.

3. Пространственные и спектральные характеристики излучения полупроводникового лазера сильно зависят от свойств материала, из которого сделан переход (таких свойств, как структура запрещенной зоны и коэффициент преломления).

4. В лазере с $p - n$ -переходом лазерное излучение возникает непосредственно под действием тока, протекающего через прямо-смещенный диод. В результате система в целом очень эффективна, поскольку позволяет легко осуществлять модуляцию излучения

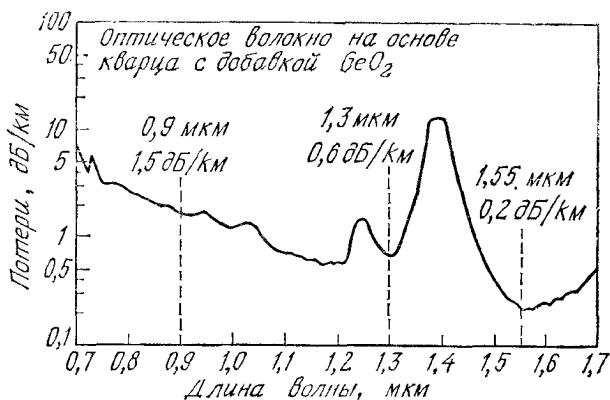


Рис. 22. Зависимость потерь от длины волны излучения в кварцевом волокне. Обозначены три значения длины волны, представляющие наибольший интерес [43].

за счет модуляции тока. Так как полупроводниковые лазеры характеризуются очень малыми временами стимулированного излучения, модуляция может проводиться на высоких частотах.

Благодаря малым размерам и возможности высокочастотной модуляции полупроводниковый лазер представляет собой один из перспективных источников излучения для волоконно-оптических систем связи. На рис. 22 приведена зависимость потерь от длины волны для экспериментальных оптических волокон [43]. На графике выделены три значения длины волны, которые представляют практический интерес. В интервале длин волн вблизи 0,9 мкм в качестве источников излучения используются гетеролазеры на основе $\text{GaAs} - \text{Al}_x\text{Ga}_{1-x}\text{As}$, а в качестве фотодетекторов — Si-фотодиоды или Si-фотодиоды с лавинным умножением. Вблизи длины волны 1,3 мкм волокно имеет низкие потери (0,6 дБ/км) и слабую дисперсию, а в окрестности длины волны 1,55 мкм потери достигают минимального значения 0,2 дБ/км. Для этих двух длин волн в качестве источников излучения могут использоваться лазеры на основе четверных соединений типа $A^{\text{III}}B^{\text{V}}$, таких, как $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y} - \text{InP}$, а в качестве фотодетекторов — германиевые лавинные фотодиоды, а также фотодиоды на тройных и четверных соединениях [44]. Волоконно-оптические линии связи для больших значений длин волн рассмотрены в работе [45]. Для создания таких систем необходимы волокна с исключительно низкими потерями, а также эффективные источники излучения (например, светодиоды $\text{PbS}_{0,1}\text{Se}_{0,9}$ [77]) и чувствительные фотодетекторы.

Полупроводниковые материалы. Перечень полупроводниковых материалов, проявляющих лазерные свойства, постоянно растет.

В настоящее время фактически все лазерные полупроводники обладают прямой запрещенной зоной. Это обусловлено тем, что излучательные переходы в прямозонных полупроводниках представляют собой процесс первого порядка (импульс сохраняется автоматически) и вероятность переходов высока. В полупроводниках с непрямой зоной излучательная рекомбинация выступит как процесс второго порядка (т. е. сохранение импульса и энергии при переходах обеспечивается участием фононов или другими факторами рассеяния), так что вероятность излучательных переходов существенно ниже. Кроме того, в непрямозонных полупроводниках при увеличении степени возбуждения потери, связанные с поглощением излучения на нижективируемых свободных носителях, возрастают быстрее, чем усиление [12].

На рис. 23 представлен диапазон длин волн лазерного излучения, включающий спектральные интервалы излучения различных полупроводников и охватывающий области спектра от ультрафиолетовой до инфракрасной [20]. В связи с тем, что GaAs стал первым лазерным материалом, наиболее интенсивное исследование и развитие получили родственные ему твердые растворы элементов III и V групп. Такие соединения типа $A^{IV}B^V$, как PbS, PbTe, PbSe и родственные им твердые растворы, также обладают

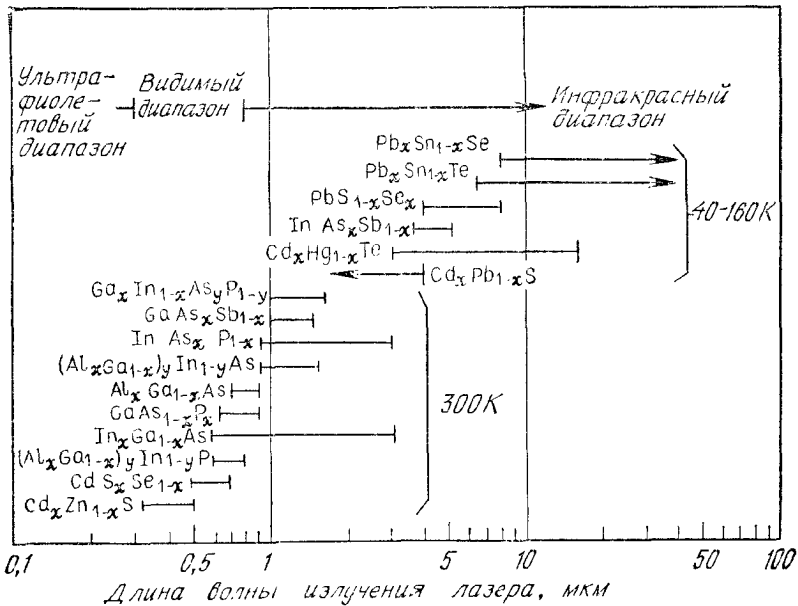


Рис. 23. Интервалы длин волн, для которых имеются или могут быть созданы гетеролазеры на соединениях $A^{III}B^{IV}$ и $A^{IV}B^{VI}$ [20].

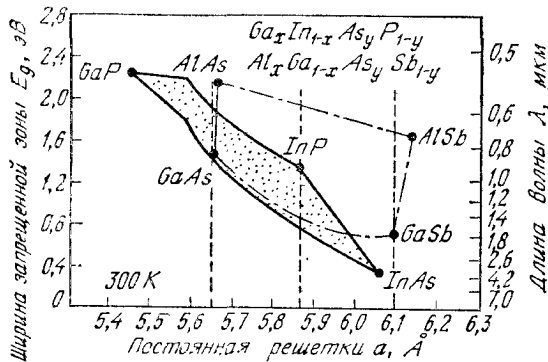


Рис. 24. Ширина запрещенной зоны и постоянная решетки для твердых растворов типа $A^{III}B^{IV}$ [20].

лазерными свойствами. В этих соединениях, относящихся к прямозонным материалам, энергетические экстремумы в зоне Бриллюэна локализованы вдоль направления $\langle 111 \rangle$ [46] в отличие от GaAs, у которого экстремумы расположены в центре зоны.

В этом разделе мы рассмотрим основы физики лазеров, причем в качестве иллюстраций будем пользоваться в основном примерами, относящимися к приборам на основе полупроводниковых соединений типа $A^{III}B^{V}$. Если узлы решетки, соответствующие элементам III группы, случайным образом заняты различными атомами элементов этой группы или подобная ситуация реализуется в узлах решетки элемента V группы, то такие соединения относятся к кристаллическим твердым растворам. При их описании часто используются обозначения типа $A_x B_{1-x} C$ и $A_x B_{1-x} C_y D_{1-y}$ соответственно для тройных и четверных соединений, где A и B — элементы III группы, а C и D — элементы V группы. В пределах каждой группы элементы могут быть расположены в алфавитном порядке или в соответствии с их атомным номером. В этой главе мы будем пользоваться алфавитным порядком расположения элементов внутри группы.

К наиболее важным из сложных соединений относятся твердые растворы $Al_x Ga_{1-x} As_y Sb_{1-y}$ и $Ga_x In_{1-x} As_y P_{1-y}$ [20]. На рис. 24 приведена диаграмма значений ширины запрещенной зоны и постоянной решетки для бинарных полупроводников типа $A^{III}B^{V}$ и их промежуточных тройных и четверных соединений. Для создания гетероструктур с пренебрежимо малой плотностью состояний на границе раздела необходимы пары полупроводников с близкими постоянными решетки. При использовании GaAs ($a = 5,6533 \text{ \AA}$) в качестве подложки для гетероструктуры с тройным соединением $Al_x Ga_{1-x} As$ можно достичь

рассогласования решеток менее 0,1 %. Аналогично подложка из InP ($a = 5,8686 \text{ \AA}$) может иметь почти идеальное согласование постоянных решетки с четверным соединением $\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$, что обозначено вертикальными линиями на рис. 24.

Стимулированное излучение. Как говорилось выше, работа лазера связана с тремя основными процессами, обусловленными переходом носителей: поглощением, спонтанной эмиссией и стимулированным излучением. Для описания этих процессов мы будем использовать простую модель [47]. Рассмотрим два энергетических уровня E_1 и E_2 , один из которых E_1 характеризует основное, а другой E_2 — возбужденное состояние атома (рис. 25). Любой переход между этими состояниями сопровождается испусканием или поглощением фотона с частотой ν_{12} , определяемой из соотношения $h\nu_{12} = E_2 - E_1$, где h — постоянная Планка. При обычных температурах большинство атомов находится в основном состоянии. Эта ситуация нарушается в результате воздействия на систему фотона с энергией, равной $h\nu_{12}$. Атом в состоянии E_1 поглощает фотон и переходит в возбужденное состояние E_2 . Это и составляет процесс поглощения излучения (рис. 25, а). Возбужденное состояние является нестабильным и через короткий промежуток времени без какого-либо внешнего воздействия атом переходит в основное состояние, испуская фотон с энергией $h\nu_{12}$. Этот процесс называется спонтанной эмиссией (рис. 26, б). Время жизни, связанное со спонтанной эмиссией (т. е. среднее время возбужденного состояния), может изменяться в широком диапазоне, обычно в пределах 10^{-9} — 10^{-3} с, в зависимости от параметров полупроводника, таких, как структура зон (прямая или непрямая) и плотность рекомбинационных центров. Столкновение фотона, обладающего энергией $h\nu_{12}$, с атомом, находящимся в возбужденном состоянии, стимулирует мгновенный переход атома в основное состояние с испусканием фотона с энергией $h\nu_{12}$ и фазой, соответствующей фазе падающего излучения. Этот процесс называется стимулированным излучением (рис. 25, в).

Лазеры на основе двойных гетероструктур. На рис. 26 показана базовая структура лазера с p — n -переходом. Две боковые грани структуры скальваются или полируются перпендикулярно плоскости перехода. Две другие грани делаются шероховатыми для того, чтобы исключить излучение в направлениях, не совпадающих с главным. Такая структура называется резонатором Фабри — Перо. Смещение лазерного диода в прямом направлении вызывает протекание тока. Вначале, при низких значениях тока, возникает спонтанное излучение, распространяющееся во всех направлениях. При увеличении смещения ток достигает порогового значения, при котором создаются условия для стимулированного излучения, и p — n -переход испускает монохроматичный и строго направленный луч света (рис. 26).

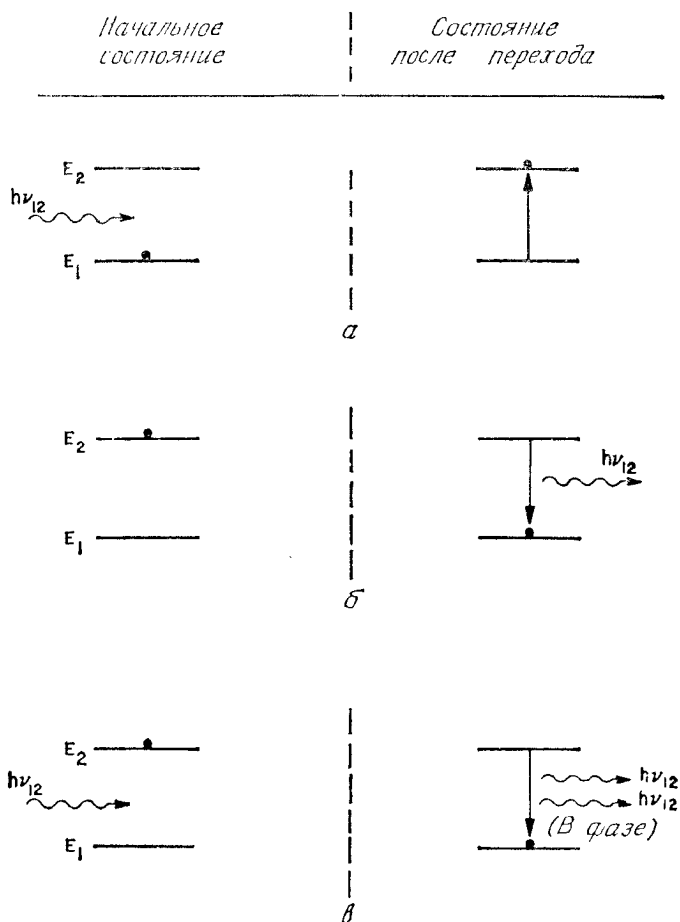


Рис. 25. Три основных типа переходов между двумя энергетическими уровнями E_1 и E_2 . Черными кружками обозначены состояния атома [47].
a — поглощение; *б* — спонтанное излучение; *в* — стимулированное излучение.

Для гомоструктур (например, $p-n$ -переходов на основе GaAs) пороговая плотность тока быстро увеличивается с ростом температуры. При комнатной температуре типичное значение J_{th} (полученное при измерении в импульсном режиме) составляет $\sim 5,0 \cdot 10^4$ А/см². Такая большая плотность тока создает серьезные трудности для реализации режима непрерывной генерации при 300 К.

С целью уменьшения пороговой плотности тока были предложены, а затем с помощью эпитаксиальной технологии реализованы

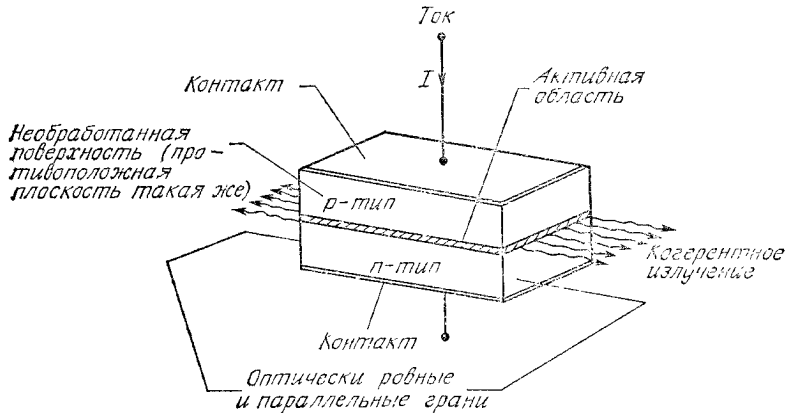


Рис. 26. Основная структура лазера с $p-n$ -переходом в виде резонатора Фабри — Перо.

лазеры на гетероструктурах. На рис. 27 схематически изображены гомоструктура, структуры с одним и двумя гетеропереходами, их зонные диаграммы в условиях прямого смещения [48], а также профили распределения показателя преломления и интенсивности генерируемого в $p-n$ -переходе света. Как видно из рис. 27, в структуре с двумя гетеропереходами носители сосредоточены внутри активной области d , ограниченной с обеих сторон потенциальными барьерами; излучение также ограничено этой

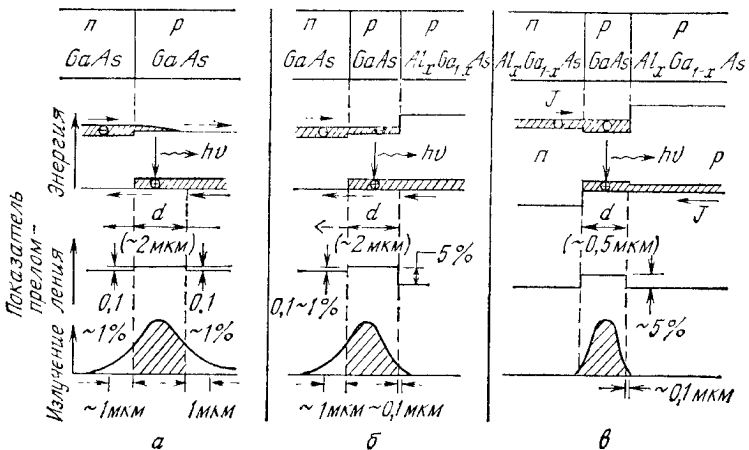


Рис. 27. Сравнительные характеристики некоторых лазерных структур [48]. а — гомоструктура; б — структура с одним гетеропереходом; в — структура с двумя гетеропереходами.

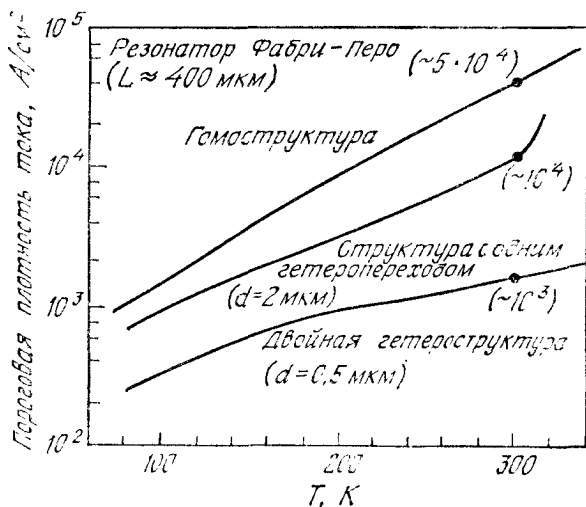


Рис. 28. Температурная зависимость пороговой плотности тока для трех лазерных структур [48].

областью вследствие скачкообразного уменьшения показателя преломления за ее пределами. Эти ограничения способствуют усилению стимулированного излучения и соответственно уменьшению пороговой плотности тока.

На рис. 28 для сравнения приведены зависимости J_{th} от рабочей температуры для трех структур, изображенных на рис. 27. Следует отметить, что самая слабая температурная зависимость наблюдается для лазеров на двойных гетероструктурах (ДГ-лазеры). Поскольку J_{th} в ДГ-лазерах при 300 К может достигать значений $\sim 10^3$ А/см² и менее, оказывается возможным реализовать режим непрерывной генерации при комнатной температуре. Благодаря этому лазеры нашли широкое применение в науке и промышленности, в частности в волоконно-оптических системах связи. Ниже рассмотрены в основном ДГ-лазеры.

12.4.2. Волноводный эффект

Коэффициент оптического ограничения. В ДГ-лазере свет удерживается и направляется диэлектрическим волноводом. На рис. 29, а показан диэлектрический волновод на основе трехслойной структуры, в которой активный слой расположен между двумя неактивными слоями. Если показатели преломления \bar{n}_1 , \bar{n}_2 и \bar{n}_3 удовлетворяют условию

$$\bar{n}_2 > \bar{n}_1 \geq \bar{n}_3, \quad (23)$$

то угол падения луча θ_{12} на границу раздела между слоями 1 и 2 (рис. 29, б) будет больше критического угла, определяемого уравнением (12). Аналогичная ситуация наблюдается на границе раздела слоев 2 и 3. Таким образом, если показатель преломления активной области больше показателей преломления окружающих ее слоев (неравенство (23)), то электромагнитное излучение распространяется в направлении, параллельном границам раздела слоев.

В лазерах с гомоструктурой разность между показателем преломления центрального слоя волновода и показателями преломления соседних слоев составляет лишь 0,1—1%. В гетеролазерах можно задать более существенное изменение показателей преломления на каждом гетеропереходе (до $\sim 10\%$) и тем самым обеспечить четко выраженное ограничение излучения.

Для более подробного анализа свойств волноводов обратимся к рис. 30, на котором схематически изображен лазер в декартовой системе координат. Плоскости $z = 0$ и $z = -L$ совпадают с двумя отражающими гранями или «зеркалами» лазера, которые представляют собой либо сколотые, либо отполированные поверх-

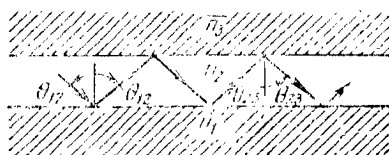
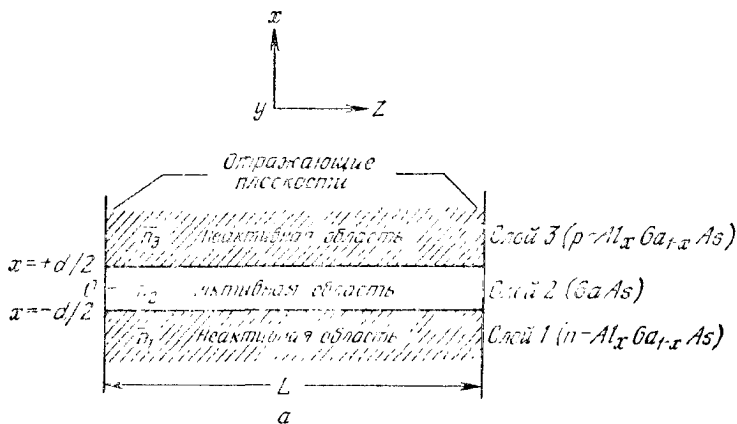


Рис. 29. Трехслойный диэлектрический волновод (а) и ход лучей в нем (б).

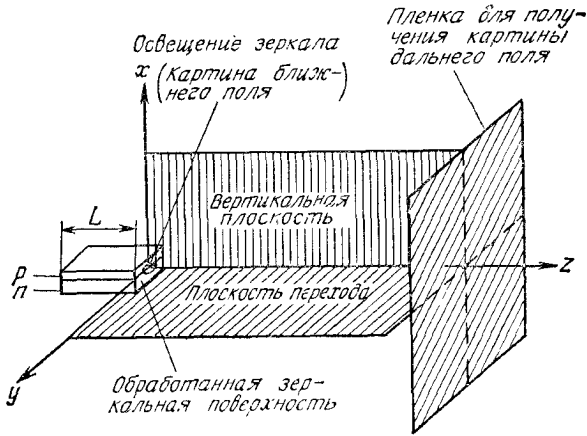


Рис. 30. Лазер с $p-n$ -переходом в декартовой системе координат.

ности. Переднее зеркало может излучать в полупространство $z > 0$. Пересекающиеся координатные оси x и y направлены соответственно перпендикулярно и параллельно плоскости перехода.

Рассмотрим симметричный трехслойный диэлектрический волновод с показателями преломления $\bar{n}_2 > \bar{n}_1 = \bar{n}_3$ (рис. 29). В электромагнитной волне с поперечной поляризацией электрического поля (ТЕ-поляризацией) компонента \mathcal{E}_z равна 0. Используя предположение о неограниченности волновода в направлении y , получим (вследствие $\partial/\partial y = 0$) упрощенное волновое уравнение

$$\partial^2 \mathcal{E}_y / \partial x^2 + \partial^2 \mathcal{E}_y / \partial z^2 = \mu_0 \epsilon \partial^2 \mathcal{E}_y / \partial t^2, \quad (24)$$

где μ_0 — магнитная проницаемость, а ϵ — диэлектрическая постоянная. Решая (методом разделения переменных) уравнение (24), получим для волны с ТЕ-поляризацией в пределах $-d/2 \leq x \leq d/2$ активного слоя

$$\mathcal{E}_y(x, z, t) = A_e \cos(\kappa x) \exp [j(\omega t - \beta z)]. \quad (25)$$

В выражении (25)

$$\kappa^2 = \bar{n}_2^2 k_0^2 - \beta^2, \quad (26)$$

где $k_0 \equiv (\omega/\bar{n}_2) \sqrt{\mu_0 \epsilon}$, а β — постоянная распространения. Магнитное поле в направлении z дается выражением

$$\begin{aligned} \mathcal{H}_z(x, z, t) &= (j/\omega\mu_0) (\partial \mathcal{E}_y / \partial x) = \\ &= (-j\kappa/\omega\mu_0) A_e \sin(\kappa x) \exp [j(\omega t - \beta z)]. \end{aligned} \quad (27)$$

Амплитуда удерживаемой волноводом электромагнитной волны спадает за пределами активного слоя. При $|x| > d/2$ решение для поперечного электрического и продольного магнитного полей имеет вид

$$\mathcal{E}_y(x, z, t) = A_e \cos(\kappa d/2) \exp[-\gamma(|x| - d/2)] \exp[j(\omega t - \beta z)], \quad (28)$$

$$\mathcal{H}_z(x, z, t) = (-x/|x|)(j\gamma/\omega\mu_0) A_e \cos(\kappa d/2) \times \\ \times \exp[-\gamma(|x| - d/2)] \exp[j(\omega t - \beta z)], \quad (29)$$

где

$$\gamma^2 \equiv \beta^2 - \bar{n}_1^2 k_0^2. \quad (30)$$

Из уравнений (26) и (30) видно, что, поскольку κ и γ должны быть положительными действительными величинами, условие существования волновых мод представляет собой неравенства $\bar{n}_2 k_0^2 > \beta^2$ и $\beta^2 > \bar{n}_1^2 k_0^2$ или

$$\bar{n}_2 > \bar{n}_1. \quad (31)$$

Этот результат идентичен соотношению (23).

Для определения постоянной распространения используем граничное условие, состоящее в том, что на границе раздела диэлектриков тангенциальная компонента магнитного поля \mathcal{H}_z должна быть непрерывной. Дисперсионное уравнение, полученное из уравнений (27) и (29) с использованием этого условия, имеет вид

$$\operatorname{tg}(\kappa d/2) = \gamma/\kappa = [(\beta^2 - \bar{n}_1^2 k_0^2)/(\bar{n}_2^2 k_0^2 - \beta^2)]^{1/2}. \quad (32)$$

Решение уравнения (32) зависит от аргумента тангенса, который определен с точностью до слагаемого $2\pi m$ (m — целое число). Мода низшего порядка или основной моде соответствует $m = 0$. При $m = 1$ реализуется мода первого порядка, при $m = 2$ — второго и т. д. При заданном числе m уравнение (32) может быть решено численно или графически. Полученное решение можно затем использовать в уравнениях (25) и (29) для определения электрического и магнитного полей.

Найдем теперь выражение для коэффициента оптического ограничения Γ , который представляет собой отношение интенсивности света в активном слое к полной интенсивности падающего света. Поскольку интенсивность света представляется вектором Пойнтинга $\mathcal{E} \times \mathcal{H}$, который пропорционален $|\mathcal{E}_y|^2$, то выражение для коэффициента оптического ограничения симметричного трехслойного диэлектрического волновода может быть

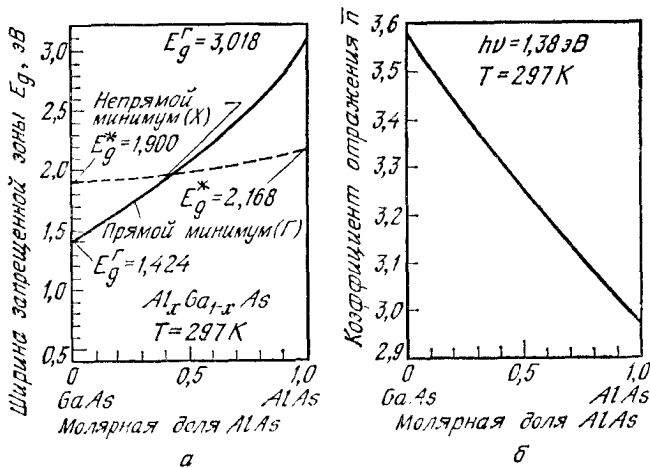


Рис. 31. Зависимость ширины запрещенной зоны (а) от состава $\text{Al}_x\text{Ga}_{1-x}\text{As}$ и показателя преломления (б) от состава для $h\nu = 1,38$ эВ [20].

получено с помощью уравнений (25) и (28). Для четных ТЕ-мод это выражение имеет вид

$$\Gamma = \frac{\int_0^{d/2} \cos^2(\kappa x) dx}{\int_0^{a/2} \cos^2(\kappa x) dx + \int_{d/2}^{\infty} \cos^2(\kappa d/2) \exp[-2\gamma(x-d/2)] dx} = \left\{ 1 + \frac{\cos^2(\kappa d/2)}{\gamma [(d/2) + (1/\kappa) \sin(\kappa d/2) \cos(\kappa d/2)]} \right\}^{-1}. \quad (33)$$

Аналогичные выражения можно получить для нечетных ТЕ-мод, а также для мод с поперечной поляризацией магнитного поля (ТМ-поляризация). Коэффициент оптического ограничения широко используется потому, что он дает представление о доле энергии, которая переносится волной, распространяющейся внутри активного слоя.

Среди лазерных гетероструктур в настоящее время наиболее интенсивно изучается система $\text{GaAs} - \text{Al}_x\text{Ga}_{1-x}\text{As}$. На рис. 31, а приведена зависимость ширины запрещенной зоны $\text{Al}_x\text{Ga}_{1-x}\text{As}$ от молярной доли AlAs . Для лазеров на основе гетероструктур наибольший интерес представляет диапазон $0 < x < 0,35$, в котором ширина прямой запрещенной зоны может быть представлена [20] в виде

$$E_g(x) = 1,424 + 1,247x \text{ [эВ]}. \quad (34)$$

На рис. 31, б приведена зависимость показателя преломления от состава соединения, которая может быть аппроксимирована выражением

$$\bar{n}(x) = 3,590 - 0,710x + 0,091x^2. \quad (35)$$

Например, при $x = 0,3$ ширина запрещенной зоны $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ составляет 1,798 эВ, что на 0,374 эВ больше, чем у GaAs; при этом показатель преломления $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ примерно на 6 % меньше, чем у GaAs.

Зависимости, приведенные на рис. 32, а, иллюстрируют влияние состава на распределение интенсивности $|\mathcal{E}_y|^2$ в трехслойном диэлектрическом волноводе на основе структуры $\text{Al}_x\text{Ga}_{1-x}\text{As} - \text{GaAs} - \text{Al}_x\text{Ga}_{1-x}\text{As}$ в направлении, перпендикулярном плоскости переходов. Расчет кривой проведен с помощью уравнений (25) и (32) для основной моды излучения ($m = 0$) с длиной волны 0,90 мкм (1,38 эВ). При заданной толщине активного слоя d в качестве переменного параметра выступает состав соединения. Значительное улучшение ограничения наблюдается при увеличении x от 0,1 до 0,2. На рис. 32, б показано изменение ограничения в зависимости от d при $x = 0,3$. По мере уменьшения толщины активного слоя свет все глубже проникает в $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ и все меньшая часть общей интенсивности сосредоточивается внутри активного слоя. При увеличении d оказываются разрешенными моды более высокого порядка. Из данных, приведенных на рис. 32, в, следует, что с ростом порядка моды все большая часть света выходит за пределы активной области. Таким образом, для достижения эффективного оптического ограничения предпочтительны моды более низкого порядка.

На рис. 33 приведена зависимость коэффициента оптического ограничения Γ для основной моды от толщины активного слоя d и состава соединения. При $d < \lambda/n_2$ ($\approx 0,5$ мкм) наблюдается резкое уменьшение Γ вследствие того, что толщина активного слоя становится меньше длины волны излучения. Представление той доли интенсивности волноводной моды, которая распространяется внутри активного слоя, с помощью коэффициента оптического ограничения Γ является важной концепцией для понимания взаимосвязи между толщиной активного слоя и пороговой плотностью тока.

Картина дальнего поля¹. Картиной дальнего поля называется распределение интенсивности излучения в свободном пространстве. На рис. 34 схематически представлена картина дальнего поля для излучения ДГ-лазера. Полные углы θ_{\parallel} и θ_{\perp} , измерен-

¹ В литературе употребляется также термин «картина излучения в дальней зоне». — Прим. перев.

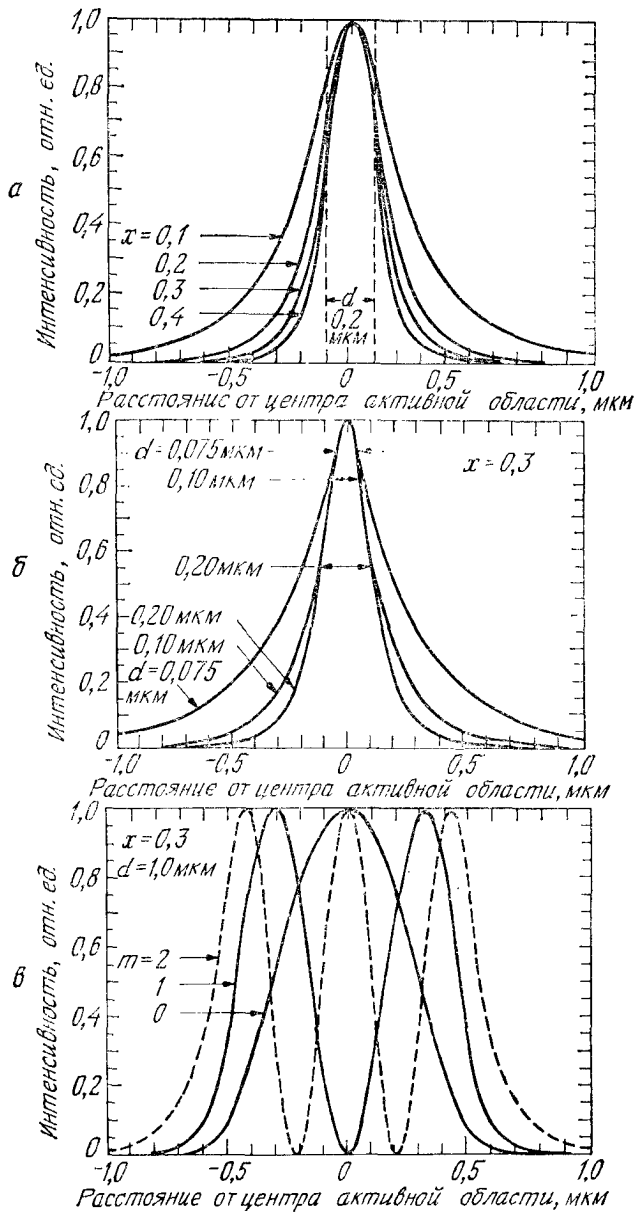


Рис. 32. Распределение интенсивности излучения в волноводе на основе двойной гетероструктуры [20].

а — изменение распределения в зависимости от молярной доли AlAs при $d = 0,2$ мкм; б — изменение распределения в зависимости от d при $x = 0,3$; в — изменение распределения в зависимости от порядка моды при постоянных x и d .

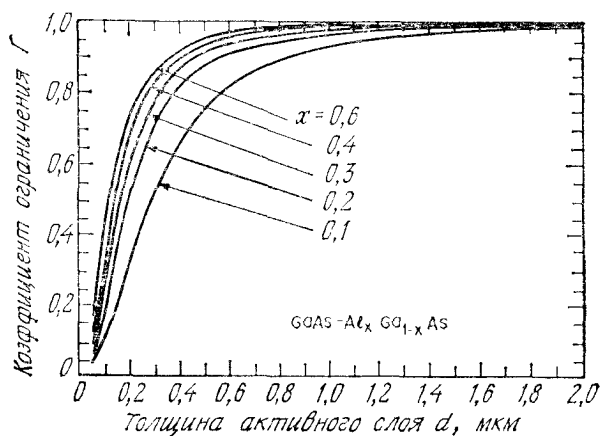


Рис. 33. Зависимость коэффициента оптического ограничения для фундаментальной моды от толщины активного слоя при различных значениях x в симметричном трехслойном диэлектрическом $\text{GaAs} - \text{Al}_x\text{Ga}_{1-x}\text{As}$ -волноводе [20].

ные на уровне половины максимальной интенсивности, характеризуют расходимость луча соответственно в плоскости перехода и перпендикулярно ей. Обычно θ_{\parallel} составляет $\sim 10^\circ$, тогда как θ_{\perp} принимает значительно большие значения ($35\text{--}60^\circ$) и зависит от толщины активного слоя и состава соединения.

Картину дальнего поля можно получить, в первую очередь, исходя из расчета пространственного распределения волн с ТЕ-поляризацией при $z > 0$. Для этого используется волновое уравне-

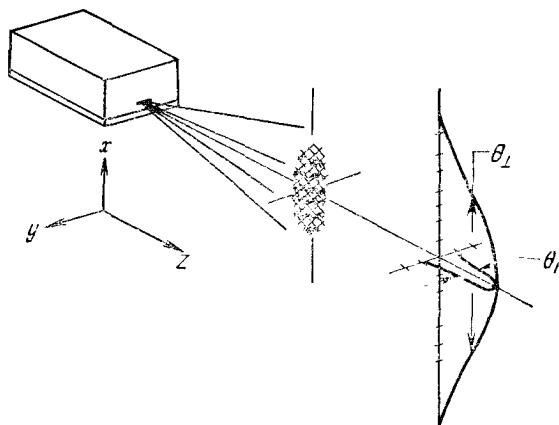


Рис. 34. Картина дальнего поля полоскового ДГ-лазера. Показаны полные углы на уровне половины максимальной интенсивности в плоскости перехода и перпендикулярно ей.

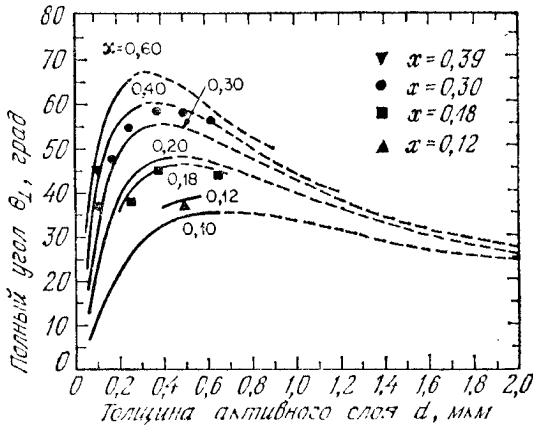


Рис. 35. Зависимость полного угла на уровне половины максимальной интенсивности от толщины активного слоя и состава соединения в GaAs — AlGaAs-лазерах [49].

ние (24), в котором ϵ заменено на ϵ_0 для свободного пространства:

$$\frac{I(\theta_{\perp})}{I(0)} = \frac{\cos^2 \theta_{\perp} \left| \int_{-\infty}^{\infty} \mathcal{E}_y(x, 0) \exp(j \sin \theta_{\perp} k_0 x) dx \right|^2}{\left| \int_{-\infty}^{\infty} \mathcal{E}_y(x, 0) dx \right|^2}. \quad (36)$$

В случае симметричного трехслойного волновода (ДГ-лазер) в уравнение (36) может быть подставлено выражение для электрического поля из уравнений (25) и (28). На рис. 35 приведены данные расчета и экспериментальные результаты для угла θ_{\perp} , полученные при исследовании картины дальнего поля GaAs — AlGaAs-лазера [49]. Сплошными и штриховыми линиями представлены кривые расходимости излучения основной моды, вычисленные из уравнения (36). Часть кривых, изображенных штриховой линией, соответствует толщине активного слоя, при которой возможны моды более высокого порядка. Экспериментальные результаты и данные расчета находятся в хорошем согласии. Для типичного значения толщины активного слоя 0,2 мкм ДГ-лазера на основе GaAs — Al_{0,3}Ga_{0,7}As угол θ_{\perp} составляет 50°.

На рис. 36 изображена базовая структура лазера с плоским волноводом, в котором излучающей является вся площадь перехода. Следует, однако, отметить, что гетеролазеры в основном имеют полосковую геометрию; на рис. 36 показаны две структуры

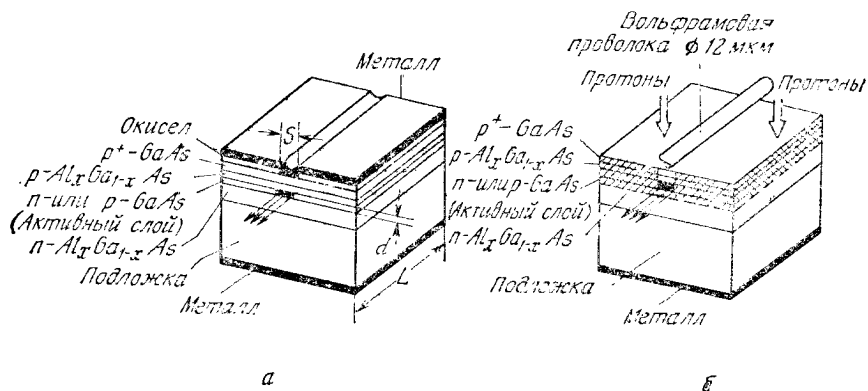


Рис. 36. ДГ-лазеры с полосковой геометрией [50].

а — с использованием оксидной изоляции; *б* — с использованием протонной бомбардировки.

с такой геометрией [50]. В структуре на рис. 36, *а* слой оксида изолирует всю площадь перехода, кроме узкого контакта, который ограничивает излучающую поверхность, расположенную под ним. В структуре на рис. 36, *б* полосковая геометрия формируется в процессе протонной бомбардировки, с помощью которой создаются высокоомные области. Излучающая поверхность сосредоточена в центре области, не подвергнутой бомбардировке. Ширина канала обычно составляет 5—~30 мкм. За счет применения полосковой геометрии удается реализовать ряд преимуществ: 1) уменьшить площадь поперечного сечения излучающей области, что приводит к снижению рабочего тока; 2) устранить возможность возникновения нескольких областей с высокой интенсивностью излучения (так называемых шинуров) при увеличении тока накачки (при ширине канала менее 15 мкм в лазере с полосковой геометрией можно обеспечить одношнуровой режим с излучением основной моды вдоль плоскости перехода); 3) улучшить надежность за счет отдаления большей части периметра перехода от граней кристалла; 4) уменьшить инерционность за счет снижения емкости перехода.

В лазере с полосковой геометрией напряженность электрического поля вдоль направления, параллельного плоскости перехода (ось y), сильно зависит от пространственного изменения диэлектрической постоянной. Для структуры на рис. 37 волновое уравнение для электрического поля, имеющего временную зависимость вида $\exp(j\omega t)$, записывается следующим образом [51]:

$$\nabla^2 \mathcal{E}_y + (k_0^2 \epsilon / \epsilon_0) \mathcal{E}_y = 0. \quad (37)$$

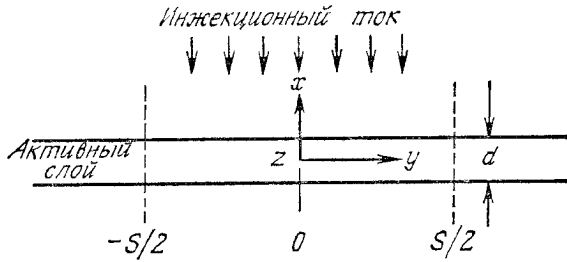


Рис. 37. Система координат для полоскового лазера, полученного протонной бомбардировкой, с шириной полоски S .

В этом уравнении $k_0 = 2\pi/\lambda$, а отношение ϵ/ϵ_0 представляется для области внутри активного слоя в двумерной форме

$$\epsilon(x, y)/\epsilon_0 = [\epsilon(0) - a^2 y^2]/\epsilon_0. \quad (38)$$

В прилегающих слоях оно имеет вид

$$\epsilon(x, y)/\epsilon_0 = \epsilon_1/\epsilon_0. \quad (39)$$

В уравнении (38) $\epsilon(0) = \epsilon_r(0) + j\epsilon_i(0)$ представляет собой комплексную диэлектрическую постоянную активного слоя при $y = 0$, а коэффициент $a - \text{комплексную константу}$, $+ja_i$. Приближенное решение уравнения (37) при использовании диэлектрических постоянных из уравнений (38) и (39) имеет вид

$$\mathcal{E}_x(x, y, z) = \mathcal{E}_y(x) \mathcal{E}_y(y) \exp(-j\beta_z z). \quad (40)$$

Поскольку $\epsilon(x, y)$ слабо зависит от y в плоскости перехода, то компонента \mathcal{E}_y почти не ограничивается вдоль направления y и поэтому может быть представлена полученными выше выражениями (25) и (28). Из уравнения (37) с помощью разделения переменных можно получить

$$[\partial^2 \mathcal{E}_y(x)/\partial x^2] + \beta_x^2 \mathcal{E}_y(x) = 0. \quad (41)$$

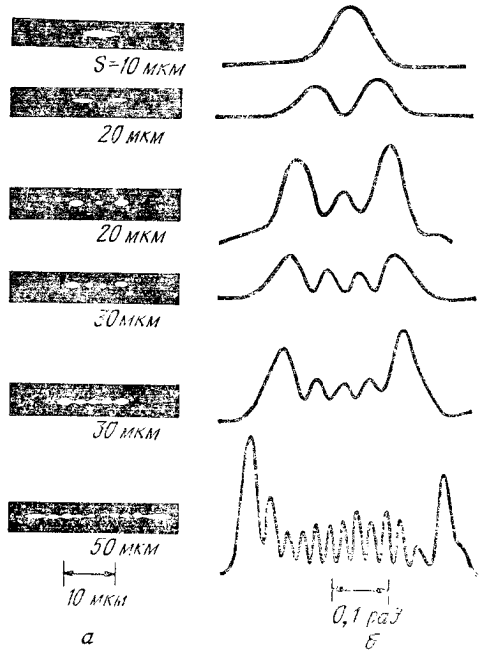
Подстановка уравнений (40) и (41) в уравнение (37) с последующим исключением из него компоненты $\mathcal{E}_y(x)$ путем умножения на сопряженное этой компоненте комплексное выражение и интегрирования по x приводит к дифференциальному уравнению для $\mathcal{E}_y(y)$:

$$[\partial^2 \mathcal{E}_y(y)/\partial y^2] + \{k_0^2 [\Gamma \epsilon(0)/\epsilon_0 + (1 - \Gamma) \epsilon_1/\epsilon_0] - \beta_x^2 - \beta_z^2 - (\Gamma k_0^2 a^2 y^2/\epsilon_0)\} \mathcal{E}_y(y) = 0, \quad (42)$$

в котором Γ — коэффициент оптического ограничения.

Рис. 38. Структура мод вдоль плоскости $p-n$ -перехода в зависимости от ширины полоски S в планарном полосковом ДГ-лазере [52].

a — картины ближнего поля; b — картины дальнего поля.



Распределение поля $\mathcal{E}_y(y)$, полученное при решении уравнения (42), представляет собой функцию Эрмита — Гаусса

$$\mathcal{E}_y(y) = H_p [(\Gamma^{1/2} a k_0' \epsilon_0^{1/2})^{1/2} y] \exp \left[-\frac{1}{2} (\Gamma/\epsilon_0)^{1/2} a k_0 y^2 \right], \quad (43)$$

где H_p — полином Эрмита порядка p , равный

$$H_p(\xi) \equiv (-1)^p \exp(\xi^2) \partial^p \exp(-\xi^2) / \partial \xi^p. \quad (44)$$

Приведем три первых полинома Эрмита: $H_0(\xi) = 1$, $H_1(\xi) = 2\xi$ и $H_2(\xi) = 4\xi^2 - 2$. Таким образом, распределение интенсивности основной моды дается выражением в виде функции Гаусса

$$|\mathcal{E}_y(y)|^2 = \exp \left[-(\Gamma/\epsilon_0)^{1/2} a_r k_0 y^2 \right], \quad (45)$$

из которого следует, что распределение интенсивности вдоль плоскости перехода зависит от a_r .

На рис. 38 показаны картины ближнего и дальнего поля вдоль плоскости перехода для лазера с полосковой геометрией [52]. При ширине канала 10 мкм наблюдается основная мода с гауссовым распределением интенсивности. По мере увеличения ширины канала вдоль плоскости перехода возникают моды более высокого порядка. Эти моды характеризуются распределением интенсивности в виде функций Эрмита — Гаусса, определяемых уравнением (43).

12.4.3. Пороговая плотность тока

В условиях теплового равновесия в основном состоянии находится больше атомов, чем в возбужденном. Если возникает обратная ситуация, то говорят об инверсной населенности. При взаимодействии фотонов, обладающих энергией $h\nu_{12}$, с простой системой (рис. 25), в которой уровень E_2 инверсно заселен по отношению к уровню E_1 , стимулированная эмиссия будет преобладать над поглощением, и в результате систему будет покидать больше фотонов с энергией $h\nu_{12}$, чем входит в нее. Такое явление называют квантовым усилением.

Для того чтобы рассмотреть условия инверсной населенности в полупроводниковых лазерах, обратимся к рис. 39, на котором приведены зависимости энергии от плотности состояний в полупроводнике с прямой зоной [53]. Ситуация, изображенная на рис. 39, а, соответствует условно равновесию при $T = 0$ К для собственного полупроводника, в котором все состояния в валентной зоне заняты (заштрихованная площадь). На рис. 39, б показан случай инверсной населенности при $T = 0$ К, которая возникает, например, при фотовозбуждении электронов фотонами с энергией, превышающей ширину запрещенной зоны E_g . Валентная зона выше уровня E_{FV} свободна от электронов, а зона проводимости заполнена ими до уровня E_{FC} . В такой ситуации фотоны с энергией $h\nu$ в диапазоне $E_g < h\nu < (E_{FC} - E_{FV})$ будут вызывать переходы электронов на нижние уровни, что в свою очередь приведет к возникновению стимулированной эмиссии.

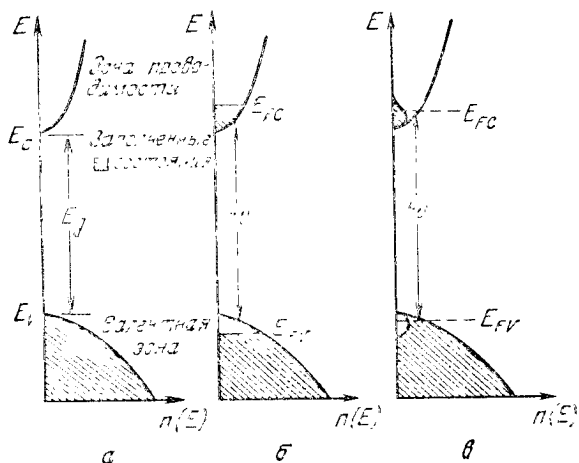


Рис. 39. Зависимость плотности состояний от энергии в полупроводнике [53]. а — состояние равновесия, $T = 0$ К; б — инверсия, $T = 0$ К; в — инверсия, $T > 0$ К.

При конечных температурах распределение носителей «размазано» по энергетическому спектру (рис. 39, в). Несмотря на отсутствие общего теплового равновесия, носители в данной энергетической зоне находятся в тепловом равновесии друг с другом. Вероятность заполнения состояния в зоне проводимости дается распределением Ферми — Дирака

$$F_C(E) = \frac{1}{1 + \exp\left(\frac{E - E_{FC}}{kT}\right)}, \quad (46)$$

где E_{FC} — квазиуровень Ферми для электронов в зоне проводимости. Аналогичное выражение справедливо и для валентной зоны.

Рассмотрим эмиссию фотонов с энергией $h\nu$, возникающую в результате переходов электронов с верхних состояний, которым соответствуют энергии вблизи уровня E в зоне проводимости, на нижние состояния, которым соответствуют энергии вблизи уровня $(E - h\nu)$ в валентной зоне. Скорость такой эмиссии пропорциональна произведению плотности занятых верхних состояний $n_C(E) F_C(E)$ и плотности свободных нижних состояний $n_V(E - h\nu) [1 - F_V(E - h\nu)]$. Полная скорость эмиссии определяется в результате интегрирования по всем энергиям:

$$W_{\text{спонт}}(h\nu) = B \int n_C(E) n_V(E - h\nu) F_C(E) [1 - F_V(E - h\nu)] |\langle M \rangle|^2 dE. \quad (47)$$

Таким же образом можно получить выражение для скорости поглощения:

$$W_{\text{погл}}(h\nu) = B \int n_V(E - h\nu) n_C(E) F_V(E - h\nu) [1 - F_C(E)] |\langle M \rangle|^2 dE. \quad (48)$$

Коэффициент B в этих выражениях определяется соотношением

$$B = (4\pi\hbar^2 q^2 h\nu / m^2 \epsilon_0 h^2 c^3) \text{Vol}, \quad (49)$$

где $\langle M \rangle$ — матричный элемент, а Vol — объем кристалла. Усиление возникает при условии $W_{\text{спонт}} \gg W_{\text{погл}}$. Из уравнений (46) и (48) с соответствующими квазиуровнями Ферми E_{FC} и E_{FV} получим [11]

$$(E_{FC} - E_{FV}) \triangleright h\nu. \quad (50)$$

Это неравенство является необходимым условием для того, чтобы стимулированное излучение преобладало над поглощением.

В полупроводниковом лазере коэффициент усиления g (т. е. приращение потока энергии на единицу длины) определяется структурой энергетических зон и сложным образом зависит от степени легирования, плотности тока, температуры и частоты.

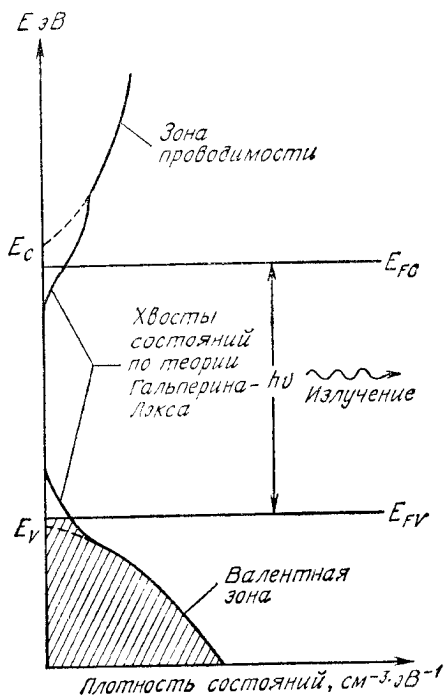


Рис. 40. Распределение плотности состояний при наличии хвостов состояний валентной зоны и зоны проводимости [55].

При возрастании интенсивности возбуждения функции распределения $E_C(E)$ и $F_V(E)$ изменяются таким образом, что E_{FC} увеличивается, а E_{FV} уменьшается. Изменяется также и форма кривой, описывающей зависимость коэффициента усиления от энергии фотонов.

Коэффициент усиления может быть рассчитан для специального распределения плотности состояний [55], при котором как валентная зона, так и зона проводимости имеют хвосты плотности состояний (рис. 40). Такое распределение характерно для большинства ДГ-лазеров на GaAs. Если распределение плотности состояний найдено, то для заданных значений $\Delta E \equiv (E_{FC} - E_{FV})$ и температуры с помощью уравнения (47) можно рассчитать функцию спонтанного излучения и получить зависимость коэффициента усиления $g = -\alpha$ от ΔE . Полученные результаты дают возможность установить также зависимость $g(h\nu)$ от номинальной плотности тока [56] $J_{\text{ном}}$, которая определяется как плотность тока, необходимая для поддержания постоянного возбуждения в активном слое толщиной 1 мкм при квантовой эффективности $\eta = 1$. Действительное значение плотности тока при этом определяется выражением

$$J \text{ (A/cm}^2\text{)} = J_{\text{ном}} d / \eta. \quad (51)$$

На рис. 41 приведена зависимость коэффициента усиления от номинальной плотности тока, рассчитанная для GaAs с концентрацией ионизованных акцепторов $5 \cdot 10^{17} \text{ см}^{-3}$ и концентрацией ионизованных доноров $1 \cdot 10^{17} \text{ см}^{-3}$ [56]. Эта зависимость сверхлинейна при малых значениях коэффициента усиления и линейна при $50 \leq g \leq 400 \text{ см}^{-1}$. Штриховая линия соответствует зависимости, которая может быть описана уравнением

$$g = (g_0/J_0) (J_{\text{ном}} - J_0), \quad (52)$$

где $g_0/J_0 = 5 \cdot 10^{-2} \text{ см} \cdot \text{мкм}/\text{А}$ и $J_0 = 4,5 \cdot 10^3 \text{ А} (\text{см}^2 \cdot \text{мкм})$.

Как уже говорилось выше, при малых токах спонтанное излучение распространяется во всех направлениях. По мере увеличения тока усиление растет (рис. 41) до тех пор, пока не достигается порог лазерной генерации, т. е. не создаются условия, при которых световая волна проходит через резонатор без затухания:

$$R \exp [(\Gamma g - \alpha) L] = 1, \quad (53)$$

$$\text{или } \Gamma g = \alpha + \frac{1}{L} \ln \left(\frac{1}{R} \right), \quad (54)$$

где g — пороговый коэффициент усиления; Γ — коэффициент оптического ограничения; α — потери на излучения на единице длины¹, обусловленные поглощением на свободных носителях и рассеянием на дефектах; L — длина резонатора; R — коэффициент отражения зеркал резонатора (при разных коэффициентах отражения зеркал резонатора $R = \sqrt{R_1 R_2}$). Из урав-

¹ То есть коэффициент поглощения. — Прим. перев.

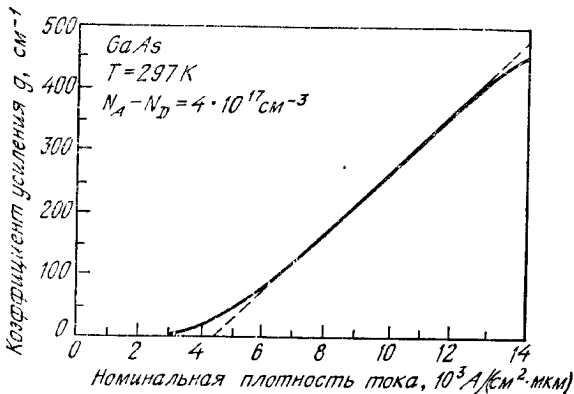


Рис. 41. Зависимость коэффициента усиления от номинальной плотности тока. Штриховой линией представлена линейная зависимость [56].

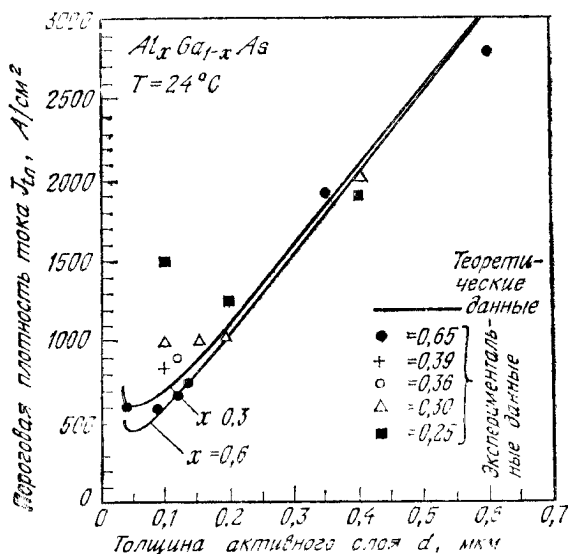


Рис. 42. Сравнение экспериментальной и рассчитанной по уравнению [55] зависимости J_{th} от толщины активного слоя [57].

нений (51), (52) и (54) можно получить выражение для пороговой плотности тока [57]

$$J_{th} \text{ (А/см}^2\text{)} = \frac{J_0 d}{\eta} + \frac{J_0 d}{g_0 \eta \Gamma} \left[\alpha + \frac{1}{L} \ln \left(\frac{1}{R} \right) \right]. \quad (55)$$

Для снижения пороговой плотности тока можно увеличивать η , Γ , L и R , а также уменьшать d и α . На рис. 42 приведены для сравнения экспериментальные результаты [57] и теоретические кривые, рассчитанные по уравнению (55). С уменьшением d пороговая плотность тока уменьшается, достигает минимального значения, а затем снова возрастает. Увеличение J_{th} при очень малых толщинах активного слоя обусловлено ухудшением оптического ограничения. При заданных d уменьшение J_{th} с увеличением x (молярной доли AlAs) происходит вследствие улучшения ограничения. Аналогичные результаты были получены для ДГ-лазеров на основе $\text{InP} - \text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y} - \text{InP}$ [58, 59]. При этом в лазерах с активным слоем из $\text{Ga}_{0,25}\text{In}_{0,75}\text{As}_{0,54}\text{P}_{0,16}$ ($\lambda = 1,23$ мкм) толщиной $\sim 0,2$ мкм минимальное значение J_{th} при комнатной температуре составляет $\sim 1,5$ кА/см².

12.5. РАБОЧИЕ ХАРАКТЕРИСТИКИ ЛАЗЕРОВ

12.5.1. Лазерные структуры

Низкая пороговая плотность в гетеролазерах обеспечивается двумя факторами: 1) ограничением носителей энергетическими барьерами между активной областью и слоем полупроводника с более широкой запрещенной зоной и 2) оптическим ограничением за счет резкого уменьшения показателя преломления за пределами активной области. В разд. 12.4 были рассмотрены ДГ-лазеры, в которых ограничение обоих типов происходит внутри активного слоя на одной и той же толщине.

Для создания гетеролазеров используются и некоторые другие гетероструктуры [60]. На рис. 43, *a* показана последовательность слоев в гетеролазере с раздельным ограничением и приведены зависимости ширины запрещенной зоны, показателя преломления и распределение интенсивности света в направлении, перпендикулярном плоскости перехода. Скачок E_g на границе между GaAs и $\text{Al}_{0,1}\text{Ga}_{0,9}\text{As}$ достаточен для того, чтобы ограничить число носителей внутри слоя толщиной d , в то же время скачок \bar{n} на этой

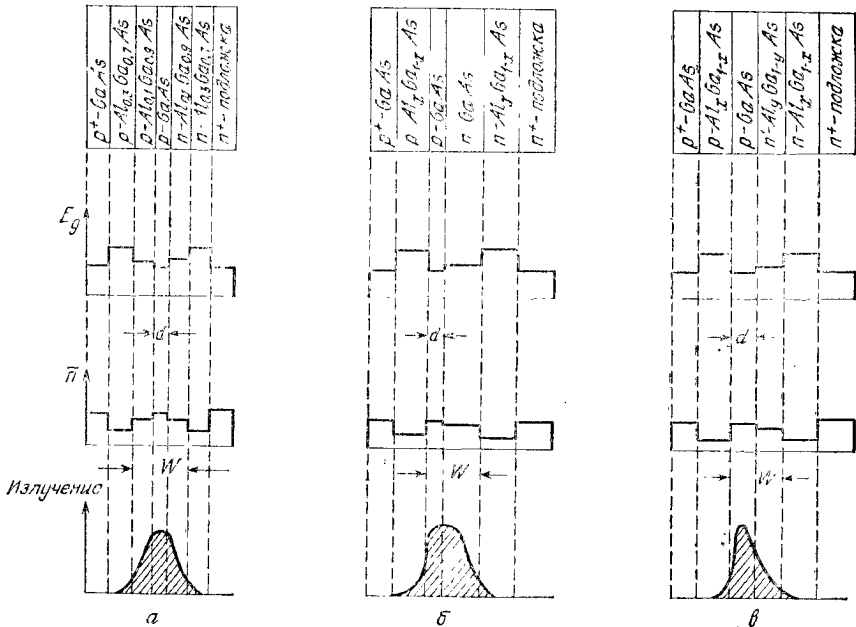


Рис. 43. Энергетическая диаграмма, показатель преломления и интенсивность света в трех типах гетеролазеров [60].

a — ДГ-лазер с раздельным ограничением; *b* — ДГ-лазер с расширенным волноводом; *в* — четырехслойный гетеролазер.

границе не обеспечивает оптического ограничения. Большая величина скачка показателя преломления между $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ и $\text{Al}_{0,1}\text{Ga}_{0,9}\text{As}$ служит для оптического ограничения и формирует оптический волновод толщиной W . В такой структуре была получена низкая пороговая плотность тока $J_{th} \approx 500 \text{ А/см}^2$.

Четырехслойный гетеролазер с расширенным волноводом (рис. 43, б) имеет структуру, аналогичную структуре ДГ-лазера с раздельным ограничением, за исключением того, что в нем в качестве волновода используется $p-n$ -гомопереход, который помещен между двумя гетеропереходами. Большая часть тока, протекающего через переход, обусловлена инжекцией электронов в p -слой, который играет роль активной области. Гетеропереход $p\text{-GaAs}/p\text{-Al}_x\text{Ga}_{1-x}\text{As}$ обеспечивает оптическое и электронное ограничения. Преимущество таких лазеров состоит в том, что в них в отличие от ДГ-лазеров увеличение толщины оптического волновода W достигается без существенного возрастания J_{th} . Поскольку в этих лазерах сечение оптического волновода может быть увеличено, то по сравнению с ДГ-лазерами они могут излучать большую мощность.

На рис. 43, в показан четырехслойный гетеролазер, в котором слой $n\text{-GaAs}$ заменен слоем $n'\text{-Al}_y\text{Ga}_{1-y}\text{As}$ с $y = 0,1-0,2$. Эта структура может работать в режиме генерации основной поперечной моды при сравнительно низкой плотности J_{th} , имеет высокую выходную мощность и относительно небольшую расходимость луча.

Большинство гетеролазеров имеет полосковую геометрию, аналогичную той, которая показана на рис. 36. На рис. 44 приведены поперечные сечения нескольких других полосковых лазерных структур. Для создания структур используются подложки из высококачественных бинарных полупроводников (например, GaAs или InP). Последующие эпитаксиальные слои из различных соединений и с различными легирующими примесями выращиваются методами жидкофазной, газофазной или молекулярно-лучевой эпитаксии. Для каждой структуры на рисунке указано, какой слой является активным, и обозначено сечение излучения.

На рис. 44, а показан полосковый лазер, полученный простым методом [61]. Мезополосковая структура лазера создается путем травления. После нанесения металлических контактов осуществляется завершающая операция скалывания граней, в результате которой формируется резонатор Фабри — Перо. В такой структуре не надо создавать окисную изоляцию или проводить протонную бомбардировку. Боковое ограничение тока достигается за счет высокого контактного сопротивления между металлом и слабелегированным $p\text{-AlGaAs}$. Этот лазер имеет низкую пороговую плотность тока, линейную ватт-амперную характеристику и работает в режиме генерации основной поперечной моды с одной

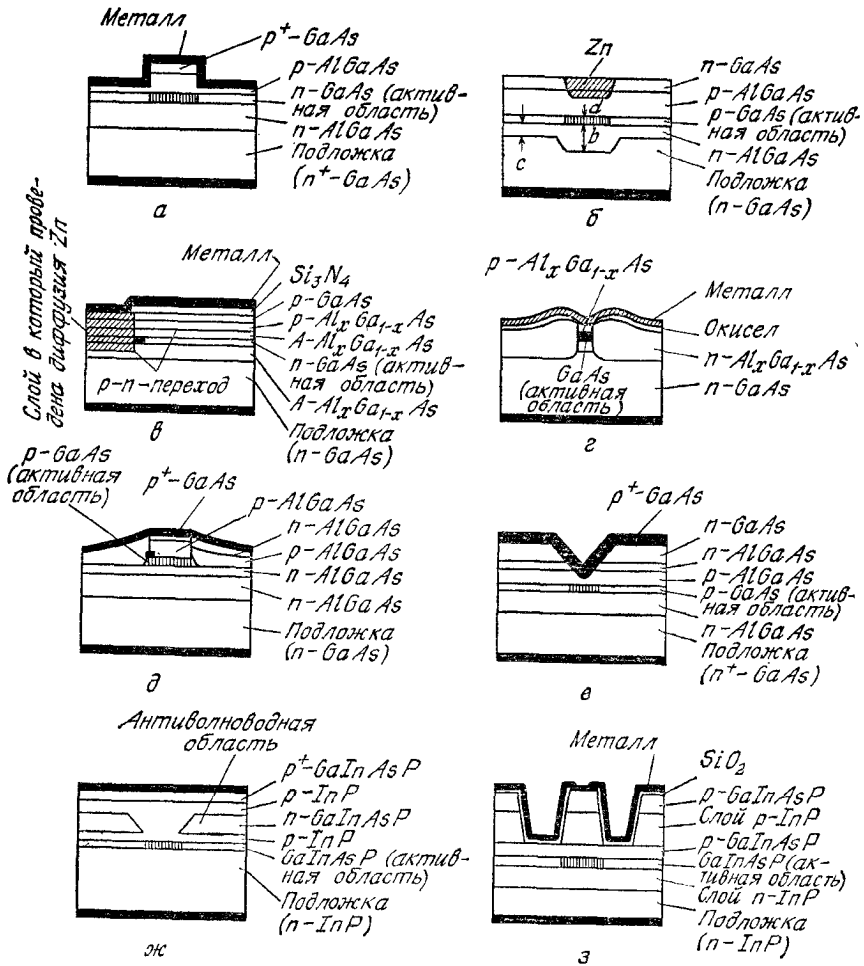


Рис. 44. Поперечные сечения различных гетеролазеров.

продольной модой на длине волны 0,861 мкм (разд. 12.5.2). На рис. 44, б показана структура гетеролазера с каналом в подложке [62]. Формирование канала в подложке проводят перед выращиванием эпитаксиальных слоев. Эта структура также обладает хорошими лазерными характеристиками. На рис. 44, в показана структура, в которой проведена диффузия Zn [63]. Она называется полосковым лазером с поперечным $p-n$ -переходом. Поскольку в этом случае инжекция идет из слоя n -GaAs в слой p -GaAs, полученный диффузией Zn, то такой прибор фактически

представляет собой очень узкий гомолазер. В нем были получены низкие пороговые токи и генерация с одной продольной модой.

В зарощенном гетеролазере [64] активная область полностью окружена слоями из материалов с более широкой зоной и с меньшим показателем преломления (рис. 44, з). Этот лазер характеризуется малым сечением активной области, достигающим 1 мкм^2 . В нем были получены очень низкие пороговые токи (до 15 мА) и почти симметричное распределение интенсивности в дальнем поле. Для создания зарощенного полоскового лазера [65] используется регулярная пятислойная эпитаксиальная структура с соответствующим содержанием Al и легирующих примесей (рис. 44, д). На слое n' -AlGaAs фотолитографическим способом формируется полоска шириной 10 мкм. Затем вместо стравленных областей наращиваются слои p - и n -AlGaAs. Ток ограничивается в центре полоски, так как восстановленный $p-n$ -переход, окружающий центральную область, смещен в обратном направлении. Структура представляет собой четырехслойный лазер, который имеет линейную ватт-амперную характеристику и два симметричных выходных зеркала.

На рис. 44, е показана новая структура полоскового лазера, в котором для ограничения ширины полоски используется V-образная канавка [66]. V-канавка формируется травлением поверхности системы GaAs — AlGaAs; с помощью диффузии Zn прилегающий слой n -AlGaAs преобразуют под V-канавкой в слой p -типа. Глубина диффузии определяется профилем V-канавки. Была получена линейная ватт-амперная характеристика в сочетании с расходимостью светового луча в диапазоне $30-40^\circ$.

На рис. 44, ж показан самосовмещенный GaInAsP — InP-лазер [67]. Электронное и оптическое ограничения обеспечиваются в нем с помощью антиволноводных областей n -GaInAsP, которые расположены за пределами полосковой области и зарощены слоем p -InP, покрывающим активный слой. Носители инжектируются в узкую полосковую область, ограниченную обратносмещенными $p-n-p-n$ -переходами. Прибор стабильно функционирует в режиме генерации основной моды с длиной волны $\sim 1,3 \text{ мкм}$, причем вплоть до 72° генерация осуществляется в непрерывном режиме. На рис. 44, з показан ДГ-лазер с гребенчатым волноводом [68], излучающий на длине волны 1,55 мкм. Ширина полосковой области задается шириной гребенки, которая формируется методом понного травления. Преимущество этой структуры состоит в том, что в процессе ее изготовления можно исключить повреждение активной области. Прибор также функционирует в режиме фундаментальной поперечной моды. Так как он излучает на длине волны 1,55 мкм, его можно использовать в качестве источника в оптических системах связи, работающих в одномодовом режиме и обладающих низкими потерями на больших расстояниях.

Во всех рассмотренных выше структурах обратная связь, необходимая для генерации лазерного излучения, обеспечивается за счет отражения на торцевых гранях резонатора, сформированных скалыванием или полировкой. Обратная связь может быть также получена путем периодического изменения показателя преломления внутри оптического волновода, достигаемого обычно гофрированием границы раздела между двумя диэлектрическими слоями. Пример такой структуры приведен на рис. 45 (вставка).

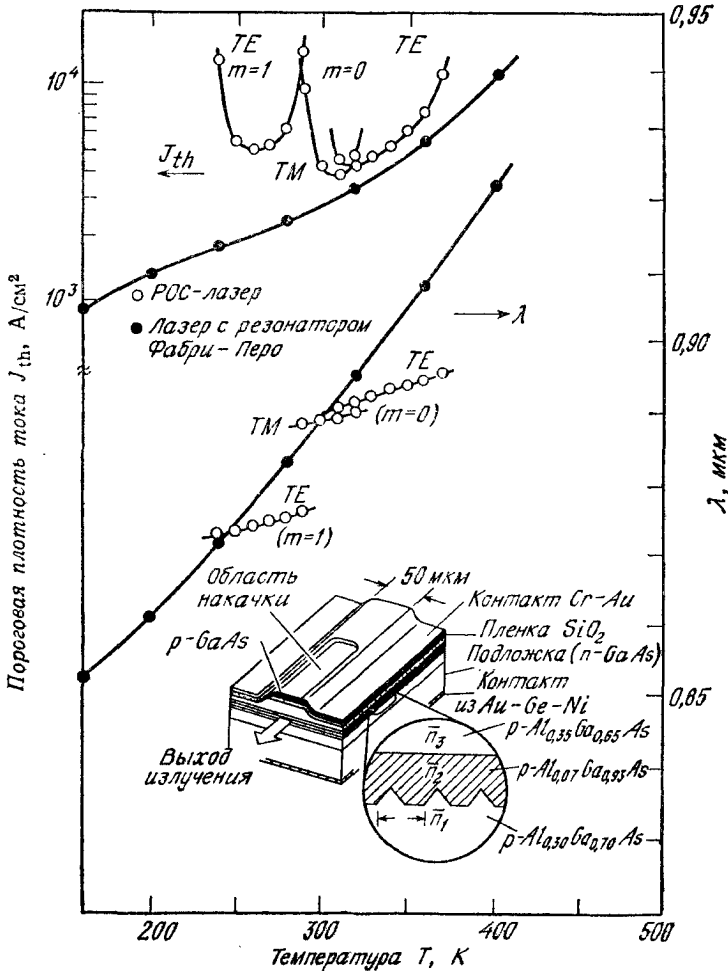


Рис. 45. Зависимость пороговой плотности тока и длины волны генерации от температуры p — n -перехода. На вставке показан лазер с распределенной обратной связью (РОС-лазер) [70].

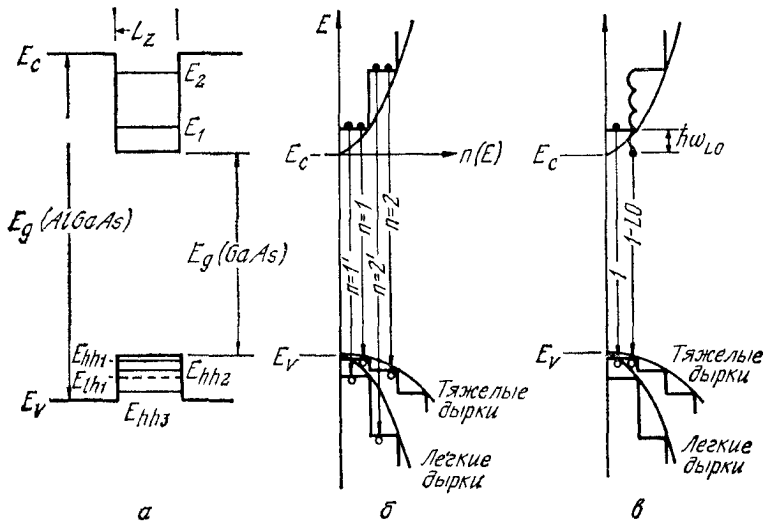


Рис. 46. Зонная диаграмма в гетероструктуре с квантовой ямой (а), диаграмма плотности состояний (б) и фононное рассеяние в гетеролазере с квантовой ямой (в) [78].

Периодическое изменение \bar{n} , заданное конструкцией прибора, вызывает интерференцию. Лазеры, в которых используются такие гофрированные структуры, называются лазерами с распределенной обратной связью (РОС) или с распределенным брэгговским отражателем (РБО) [69]. Эти гетеролазеры могут с успехом применяться в интегральной оптике в качестве источников излучения. На рис. 45 приведены характеристики лазера с распределенной обратной связью и (для сравнения) лазера с резонатором Фабри — Перо [70]. В температурном диапазоне 150—400 К для РОС-лазера наблюдались две ТЕ-моды и одна ТМ-мода. Излучение лазера с резонатором Фабри — Перо связано с температурной зависимостью ширины запрещенной зоны, а изменение длины волны генерации РОС-лазера — с более слабой температурной зависимостью показателя преломления.

В настоящее время созданы полупроводниковые лазеры на основе структур со сверхрешеткой (гл. 2). Такие приборы называются гетеролазерами с квантовыми ямами [78]. Уменьшение толщины активного слоя ДГ-лазера до величины, сравнимой с длиной волны де Бройля для носителей ($\lambda = h/p$, где h — постоянная Планка, а p — импульс), приводит к двумерному квантованию и возникновению дискретных уровней. Энергия уровней зависит от величины потенциальных барьеров на границах конечной

прямоугольной ямы. На рис. 46, а приведена энергетическая диаграмма гетероструктуры $\text{Al}_x\text{Ga}_{1-x}\text{As} - \text{GaAs}$, в которой $L_z \approx 100 \text{ \AA}$. Уровни E_1, E_2 представляют энергетические состояния электронов, $E_{hh1}, E_{hh2}, E_{hh3}$ — состояния тяжелых дырок, E_{lh1}, E_{lh2} — легких дырок.

На рис. 46, б представлена соответствующая диаграмма плотности состояний. Полупараболы, представляющие собой края зоны проводимости E_C и валентной зоны E_V , соответствуют плотностям состояний объемного полупроводника. Ступенчатое изменение плотности состояний характеризует структуру с квантовыми ямами. Междузонные рекомбинационные переходы (правило отбора $\Delta n = 0$) происходят из основного состояния в зоне проводимости (например, E_1) в основное состояние в валентной зоне (например, E_{hh1}). Энергия перехода дается выражением

$$\hbar\omega = E_g(\text{GaAs}) + E_1 + E_{hh1}. \quad (56)$$

Такая рекомбинация в отличие от рекомбинации в объемном полупроводнике с непрерывным энергетическим спектром носителей, обусловленным параболическим изменением плотности состояний, осуществляется при наличии хорошо локализованных энергетических уровней. Рис. 46, в иллюстрирует другую важную особенность гетероструктур с квантовыми ямами, состоящую в том, что носители, инжектированные с большими энергиями, могут возбуждать фононы и рассеивать свою энергию, переходя в конечном счете на состояния с меньшей плотностью. В объемном полупроводнике процесс возбуждения фононов ограничивается уменьшением плотности состояний (особенно у края зоны), тогда как в системе с квантовыми ямами в пределах области с постоянной плотностью состояний такое ограничение отсутствует. В результате этого процесса энергия электрона может оказаться ниже уровня, обусловленного квантовым ограничением, например ниже уровня E_1 (рис. 46, в), что может привести к генерации лазерного излучения с энергией фотонов $\hbar\omega < E_g$ вместо энергии $\hbar\omega > E_g$, которая характеризует рекомбинацию без участия фононов.

Гетеролазеры с квантовыми ямами были изготовлены на основе структуры $\text{Al}_x\text{Ga}_{1-x}\text{As} - \text{GaAs}$ методом химического вакуумного осаждения из металлоорганики. Изготовленные лазеры имели полосковую геометрию, в качестве ограничительных областей применялись слои $\text{Al}_{0,4}\text{Ga}_{0,6}\text{As}$, а нелегированная активная область содержала шесть квантовых ям из слоев GaAs толщиной $L_z \approx 120 \text{ \AA}$, разделенных пятью барьерными слоями $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As}$ толщиной $\sim 120 \text{ \AA}$. Потенциальные преимущества лазеров с квантовыми ямами состоят в высокой квантовой эффективности, низком пороговом токе ($\sim 1 \text{ мА}$ и менее) и слабой чувствительности к изменениям температуры.

12.5.2. Выходная мощность и спектры излучения

На рис. 47 приведена типичная зависимость мощности ДГ-лазера при возрастании тока от низких значений, характерных для спонтанной эмиссии, до значений, превышающих порог лазерной генерации. Вставка на рисунке поясняет схему детектирования, которая использовалась для измерения интенсивности излучения лазерного диода. На начальном участке, соответствующем спонтанной эмиссии, интенсивность излучения (пропорциональная току детектора) медленно растет с увеличением тока, протекающего через диод, а после возбуждения лазерной генерации резко возрастает. Экстраполяция этой кривой к нулевому значению тока I_D , показанная штриховой линией, дает значение порогового тока. Плотность порогового тока J_{th} равна I_{th}/A ,

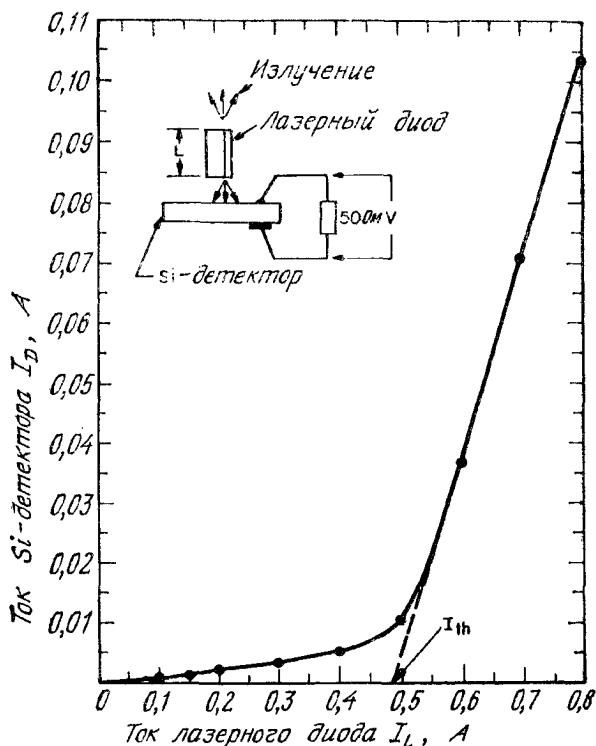


Рис. 47. Зависимость мощности излучения от тока накачки ДГ-лазера на $\text{GaAs} - \text{Al}_x\text{Ga}_{1-x}\text{As}$ при комнатной температуре. На вставке приведена схема измерения [20].

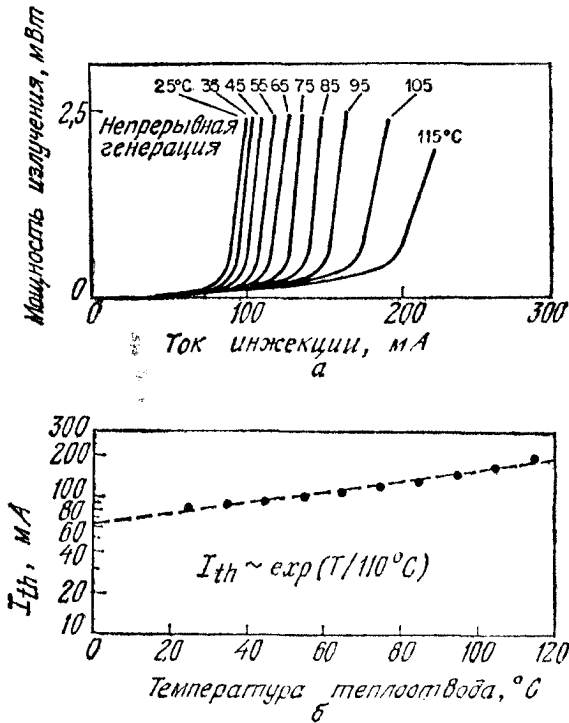


Рис. 48. Зависимость мощности излучения от тока накачки полоскового GaAs — Al_xGa_{1-x}As-гетеролазера с зарощенным активным слоем (а) и зависимость порогового тока в режиме непрерывной генерации от температуры (б) [65].

где A — активная площадь. Из графика на рис. 47 можно получить также дифференциальную квантовую эффективность

$$\eta_D \equiv \Delta I_D / \Delta I_L. \quad (57)$$

Для приведенного примера $\eta_D \approx 30\%$. Температурная зависимость порогового тока полоскового лазера с зарощенным активным слоем для режима непрерывной генерации (рис. 44, д) приведена на рис. 48. Зависимость мощности излучения от тока инжекции при непрерывной генерации (рис. 48, а) для различных температур теплоотвода в диапазоне 25—115°C почти идеально линейна. Зависимость порогового тока от температуры приведена на рис. 48, б. Пороговый ток экспоненциально возрастает с увеличением температуры:

$$I_{th} \sim \exp(T/T_0), \quad (58)$$

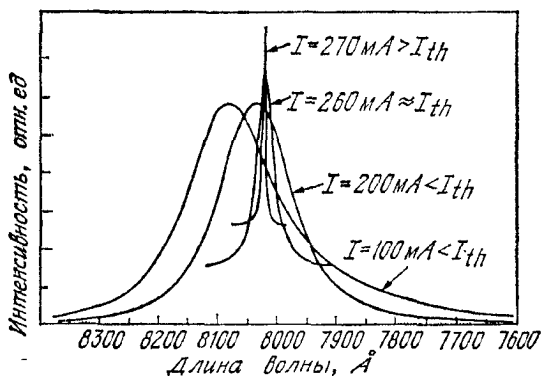


Рис. 49. Спектры излучения диодного лазера при токах накачки ниже порогового, вблизи порогового и выше порогового значений, иллюстрирующие эффект сужения полосы излучения при переходе к режиму лазерной генерации [21].

где T — температура теплоотвода, °C, а показатель T_0 , равный 110 °C, сравним по величине с аналогичным показателем для обычных ДГ-лазеров.

Режим спонтанной эмиссии, наблюдающийся при низких токах, характеризуется широким спектром излучения (полуширина спектра обычно составляет 100—500 Å). При возрастании тока до значений, близких к пороговому, спектр излучения становится уже. На рис. 49 приведены спектры для различных значений тока, иллюстрирующие эффект сужения полосы излучения при переходе к режиму лазерной генерации.

На рис. 50 показан спектр излучения полоскового ДГ-лазера, полученного протонной бомбардировкой, на основе структуры InP — GaInAsP. При токах, незначительно превышающих пороговый ток (145 мА), возникает довольно много линий излучения, которые разделены практически равными интервалами $\Delta\lambda \approx 7,5$ Å. При более высоких значениях тока ($I = 155$ мА) наблюдается тенденция к преобладанию в спектре единственной линии вблизи 1,285 мкм, характерной для одномодовой генерации. Эти линии излучения соответствуют продольным модам. Выведем основные соотношения для этих мод.

Для структуры, изображенной на рис. 29, а, получим основное условие отбора мод в продольном направлении (z -направление), вытекающее из того, что на длине полупроводника L между отражающими плоскостями должно укладываться целое число полуволн:

$$m \left(\frac{\lambda}{2\bar{n}} \right) = L, \quad (59)$$

$$\text{или } m\lambda = 2L\bar{n}, \quad (60)$$

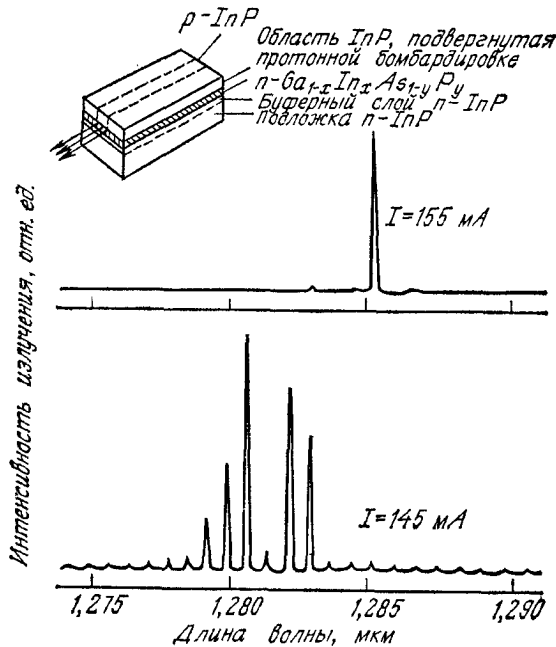


Рис. 50. Спектры излучения ДГ-лазера на InP — GaInAsP, измеренные с высоким разрешением [44].

где \tilde{n} — показатель преломления среды, соответствующий длине волны λ . Интервал $\Delta\lambda$ между этими разрешенными продольными модами представляет собой разность в длинах волн, соответствующих целым числам m и $m + 1$. Дифференцируя уравнение (6) по λ , получим для больших m

$$\Delta\lambda = \frac{\lambda^2 \Delta m}{2\tilde{n}L [1 - (\lambda/\tilde{n})(d\tilde{n}/d\lambda)]}. \quad (61)$$

Член в квадратных скобках связан с дисперсией. В газовых лазерах \tilde{n} почти не зависит от λ и дисперсионный член дает лишь малую поправку. Спектральный интервал между продольными модами обратно пропорционален длине L . У полупроводниковых лазеров вследствие их малых размеров интервал $\Delta\lambda$ значительно больше, чем у газовых лазеров.

На рис. 51, а приведено относительное распределение интенсивности излучения гетеролазера с квантовыми ямами на основе $\text{Al}_{0,3}\text{Ga}_{0,7}\text{As} - \text{GaAs}$ (шесть квантовых и пять разделительных слоев, каждый толщиной $L_z = 120 \text{ \AA}$), работающего в режиме непрерывной генерации при $T = 300 \text{ K}$. Пик линии излучения на длине волны $0,8732 \text{ мкм}$ соответствует переходу с энергией

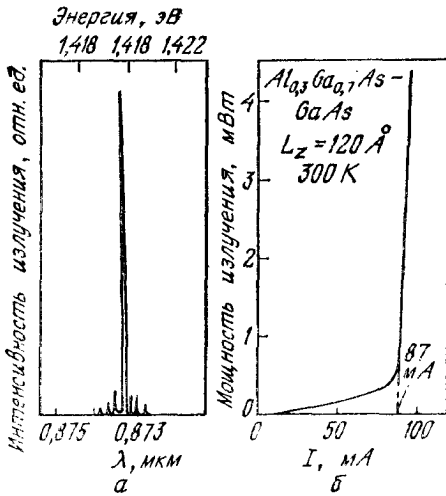


Рис. 51. Спектр излучения AlGaAs — GaAs-гетеролазера с квантовыми ямами, работающего в режиме непрерывной генерации при температуре 300 К (а), и зависимость мощности излучения от тока накачки (б) [78].

$\hbar\omega = (E_g + E_1 + E_{hh1} - \hbar\omega_{L0})$, где $\hbar\omega_{L0}$ — энергия продольного оптического фона (рис. 46, в). На рис. 51, б приведена соответствующая ватт-амперная характеристика. Дифференциальная квантовая эффективность (уравнение (57)) равна 85 %. Температурная зависимость порогового тока описывается уравнением (58) с $T_0 = 220^\circ\text{C}$. Таким образом, лазеры с квантовыми ямами менее чувствительны к температурным изменениям, чем обычные ДГ-лазеры.

12.5.3. Задержка включения и частота модуляции

При пропускании через лазер импульса тока с амплитудой, достаточной для получения стимулированного излучения, генерация излучения обычно начинается с задержкой, равной нескольким наносекундам (вставка на рис. 52). Время задержки t_d связано с временем жизни неосновных носителей.

Для вычисления времени задержки рассмотрим уравнение непрерывности для электронов в полупроводнике p -типа. Предположим, что через активный слой d протекает ток I постоянной величины и концентрация инжектированных электронов n много больше равновесной концентрации. Тогда уравнение непрерывности примет вид

$$dn/dt = I/qAd - n/\tau_n, \quad (62)$$

где A — площадь, а τ_n — время жизни носителей. Первый член уравнения представляет собой постоянную скорость генерации, а второй — скорость рекомбинации. Аналогичное выражение

можно записать для дырок в активном слое n -типа. Решение этого уравнения при $n(0) = 0$ имеет вид

$$n(t) = (\tau_e I / qAd) [1 - \exp(-t/\tau_e)] \quad (63)$$

или

$$t = \tau_e \ln \left[\frac{I}{I - qn(t) Ad / \tau_e} \right]. \quad (64)$$

Когда $n(t)$ достигает порогового значения, $n(t) = n_{th}$ и $I_{th} = qn_{th} Ad / \tau_e$. Так как при $n(t) = n_{th}$, $t = t_d$, то выражение для времени задержки имеет вид

$$t_d = \tau_e \ln \left(\frac{I}{I - I_{th}} \right). \quad (65)$$

Если перед включением импульса через лазер пропускается ток, то решение уравнения (60) при начальном условии $n(0) = I_0 \tau_e / qAd$ будет иметь вид

$$t_d = \tau_e \ln [(I - I_0) / (I - I_{th})]. \quad (66)$$

Экспериментальные результаты, иллюстрирующие изменение задержки включения лазера при увеличении тока сверх его порогового значения при различных концентрациях акцепторной примеси в активном слое, приведены на рис. 52 [71]. В соответствии

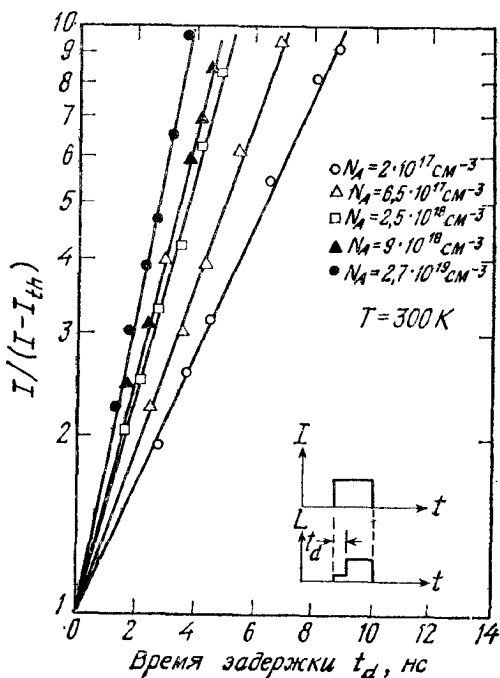


Рис. 52. Зависимость времени задержки от тока. Время задержки t_d показано на вставке [71].

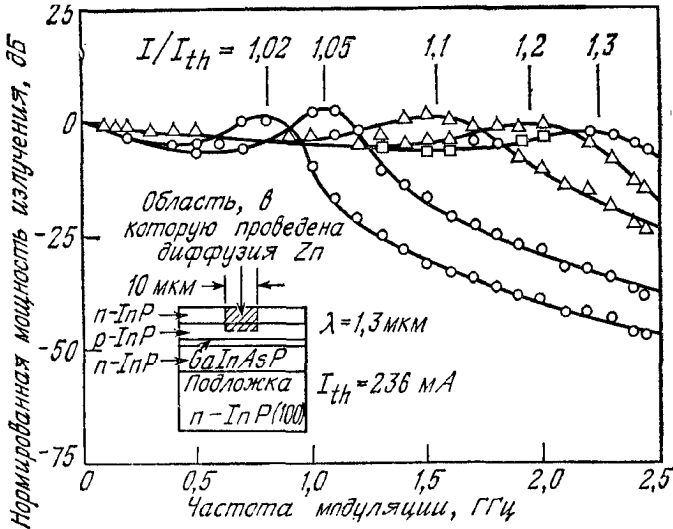


Рис. 53. Зависимость нормированной мощности излучения от частоты модуляции. На вставке показана структура лазера [72].

с уравнением (65) наблюдается логарифмическая зависимость времени задержки t_d от тока; с увеличением N_A время задержки уменьшается (поскольку уменьшается время жизни неосновных носителей).

Источники излучения, предназначенные для волоконно-оптических линий связи, должны обеспечивать возможность осуществления высокочастотной модуляции. ДГ-лазеры на основе GaAs и GaInP обладают хорошими модуляционными характеристиками в пределах гигагерцевого диапазона частот. На рис. 53 приведена зависимость нормированной выходной мощности модулированного излучения от частоты модуляции для ДГ-лазерного диода на основе GaInAsP—InP [72]. Диод излучает на длине волны 1,3 мкм и непосредственно модулируется синусоидальным током, наложенным на постоянный ток накачки. При токе накачки, превышающем значение порогового тока в 1,3 раза, мощность излучения на выходе практически не изменяется.

12.5.4. Лазерная перестройка [73]

Длину волны излучения полупроводникового лазера можно регулировать путем изменения тока диода или температуры теплоотвода, а также с помощью магнитного поля или давления. Такие достоинства полупроводниковых лазеров, как возможность перестройки длины волны узкой линии излучения, высокая стабиль-

ность, низкая потребляемая мощность, простота конструкции, открывают широкие перспективы их применения в промышленности и фундаментальных исследованиях, таких, как молекулярная и атомная спектроскопия, газовая спектроскопия высокого разрешения и контроль загрязнений атмосферы.

На рис. 23 показан диапазон длин волн, который перекрывают лазеры на основе полупроводниковых соединений. Выбирая соответствующий состав соединения для лазера, можно получить излучение с любой требуемой длиной волны из широкого спектрального диапазона 0,3—30 мкм.

Рис. 54 иллюстрирует температурную перестройку излучения ДГ-лазера на основе $\text{PbTe}-\text{Pb}_{1-x}\text{Sn}_x\text{Te}$ [74]. Регулируя температуру теплоотвода в пределах 10—120 К, можно изменять длину волны излучения примерно от 9 до 16 мкм. Это изменение в основном обусловлено эффектом расширения запрещенной зоны в соединениях типа $\text{A}^{\text{II}}\text{B}^{\text{VI}}$ при увеличении температуры.

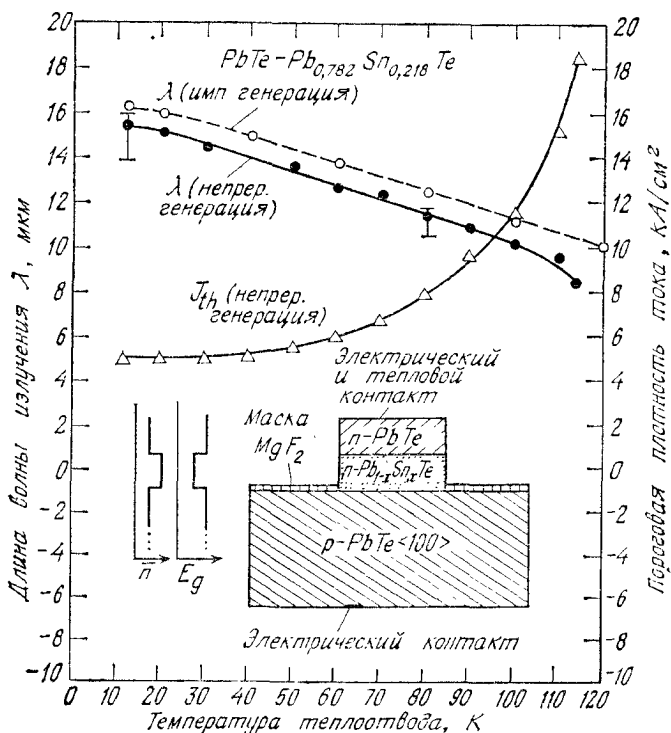


Рис. 54. Изменение длины волны излучения и порогового тока в зависимости от температуры. На вставке показана структура ДГ-лазера на $\text{PbTe}-\text{PbSnTe}$ [74].

Подвергая лазерный диод действию гидростатического давления, можно достичь смещения линии излучения в очень широком диапазоне. Изменение ширины запрещенной зоны в некоторых бинарных соединениях (например, InSb, PbS, PbSe) при изменении гидростатического давления происходит по линейному закону. За счет увеличения гидростатического давления до 14 кбар можно обеспечить перестройку линии излучения лазера на PbSe, работающего при 77 К, в диапазоне 7,5—22 мкм [73].

Диодные лазеры могут также перестраиваться под действием магнитного поля. Для полупроводников с сильной анизотропией эффективных масс положение магнитных энергетических уровней зависит от ориентации приложенного магнитного поля по отношению к оси кристалла. Как зона проводимости, так и валентная зона имеют свои энергии квантования уровней Ландау. При увеличении магнитного поля происходит увеличение энергетических интервалов между разрешенными переходами, вызывающее изменение длины волны излучения. Под действием магнитного поля 10 кГс, приложенного в направлении $\langle 100 \rangle$, лазер на основе $\text{Pb}_{0,79}\text{Sn}_{0,21}\text{Te}$ при $T = 7$ К перестраивается по длине волны излучения от 15 до 14 мкм.

12.5.5. Деградация лазеров [20]

Деградация инжекционных лазеров обусловлена целым рядом механизмов. Выделяют три основных типа деградации: 1) катастрофическое разрушение; 2) образование дефектов темных линий; 3) постепенная деградация.

Катастрофическая деградация происходит под действием больших мощностей излучения, приводящих к непрерывному повреждению зеркал лазера вследствие образования на их поверхности ямок и канавок. В результате модификаций структуры прибора, уменьшающих поверхностную рекомбинацию и поглощение, мощность излучения может увеличиться до предельного значения, при котором происходит разрушение лазера [75].

Дефекты темных линий представляют собой сетку дислокаций, которая может формироваться в процессе работы лазера и внедряться внутрь резонатора. Появившись, она может широко разрастись в течение нескольких часов и вызвать увеличение плотности порогового тока. Для уменьшения вероятности образования дефектов темных линий следует использовать высококачественные эпитаксиальные слои, выращенные на подложках с низкой плотностью дислокаций, а также обеспечивать тщательное соединение лазера с теплопроводом для минимизации напряжений.

ДГ-лазеры могут работать в течение длительного времени с относительно медленной деградацией, если устранить условия,

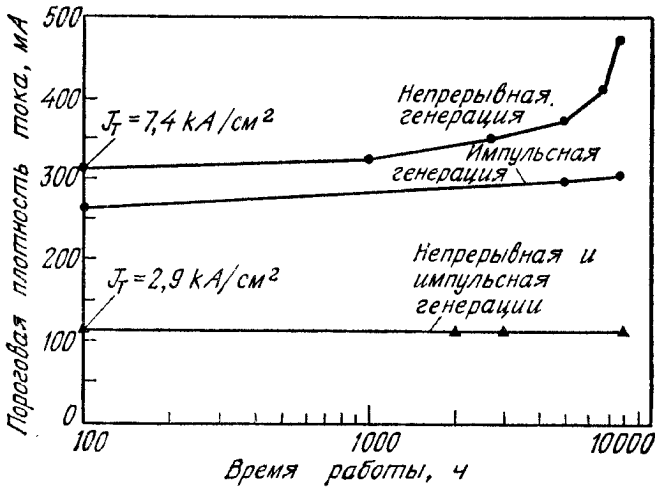


Рис. 55. Зависимость порогового тока от времени работы ДГ-лазера на InP — GaInAsP в режиме непрерывной генерации [44].

которые вызывают мгновенное катастрофическое разрушение и быструю деградацию, обусловленную образованием дефектов темных линий. Так, например, в ДГ-лазерах на основе GaAs—AlGaAs, работающих в режиме непрерывной генерации в течение более $2,6 \cdot 10^4$ ч (3 года) при температуре 30 °С, не обнаруживалось признаков деградации [76]. Экстраполяция этих данных на температуру теплоотвода 22 °С дает срок службы более 10^6 ч (около 100 лет). На рис. 55 приведена зависимость порогового тока от времени работы в режиме непрерывной генерации для ДГ-лазеров на основе GaIn—AsP—InP [44]. После 8000 ч работы характеристики лазера с низким значением порогового тока ($J_{th} = 2,9$ кА/см²) остались прежними, а в лазере с высоким пороговым током существенно увеличилось тепловое сопротивление, но признаки внутренней деградации отсутствовали. Имеются все основания полагать, что срок службы, достигнутый для ДГ-лазеров на GaAs, может быть также получен и для лазеров, излучающих в более длинноволновой области. Лазеры с длительным сроком службы необходимы для волоконно-оптических систем дальней связи, а также для ряда других систем.

ЛИТЕРАТУРА

1. Round H. J. A Note on Carborundum, *Electron World*, 19, 309 (1907).
2. Bergh A. A., Dean P. J. Light-Emitting Diodes, Clarendon, Oxford, 1976. [имеется русский перевод: Берг А., Дин П. Светодиоды, М.: «Мир», 1979.]
8. Gage S., Evans D., Hodapp M., Sorenson H. Optoelectronics Applications Manual McGraw-Hill, N. Y., 1977.

4. Gordon J. P., Zeiger H. J., Townes C. H., Molecular Microwave Oscillator and New Hyperfine Structure in the Microwave Spectrum of NH_3 , *Phys. Rev.*, **95**, 282 (1954).
5. Schawlow A. L., Townes C. H. Infrared and Optical Masers, *Phys. Rev.*, **112**, 1940 (1958).
6. Maiman T. H. Stimulated Optical Radiation in Ruby Masers, *Nature (Lond.)*, **187**, 493 (1960).
7. Aigrain P. (1958), as reported in Proc. Conf. Quantum Electron., Paris, 1963, p. 1762.
8. Басов Н. Г., Вул Б. М., Попов Ю. М. Квантовомеханические полупроводниковые генераторы и усилители электромагнитных колебаний, ЖЭТФ (1960).
9. Nishizawa J. I., Watanabe Y. Japanese Patent (Apr. 1957).
10. Boyle W. S., Thomas D. G., U. S. Patent 3059117 (Oct. 16, 1962).
11. Bernard M. G. A., Duraifourg G. Laser Conditions in Semiconductors, *Phys. Status Solidi*, **1**, 699 (1961).
12. Dumke W. P. Interband Transitions and Maser Action, *Phys. Rev.*, **127**, 1559 (1962).
13. Hall R. N., Genner G. E., Kingsley J. D., Sottys T. J., Carlson R. O. Coherent Light Emission from GaAs Junctions, *Phys. Rev. Lett.*, **9**, 366 (1962).
14. Nathan M. I., Dumke W. P., Burns G., Dill F. J., Jr., Lasher G. J. Stimulated Emission of Radiation from GaAs $p-n$ Junction, *Appl. Phys. Lett.*, **1**, 62 (1962).
15. Quist T. M., Rediker R. H., Keyers R. J., Krag W. E., Lax B., McWhorter A. L., Zeiger H. J. Semiconductor Maser of GaAs, *Appl. Phys. Lett.*, **1**, 91 (1962).
16. Holonyak N., Jr., Bevacqua S. F. Coherent (Visible) Light Emission from Ga(As_{1-x}P_x) Junction, *Appl. Phys. Lett.*, **1**, 82 (1962).
17. Hayashi I., Panish M. B., Foy P. W., Sumski S. Junction Lasers which Operate Continuously at Room Temperature, *Appl. Phys. Lett.*, **17**, 109 (1970).
18. Kroemer H. A Proposed Class of Heterojunction Injection Lasers, *Proc. IEEE*, **51**, 1782 (1963).
19. Алферов Ж. И., Казаринов Р. Ф. Авторское свидетельство № 181737 (1963), Алферов Ж. И. ФТП, т. 1, стр. 436 (1967).
20. Casey H. C., Jr., Panish M. B. Heterostructure Lasers, Academic, N. Y., 1978. [имеется русский перевод: Кейси Х., Панши М. Лазеры на гетероструктурах, М., Мир, 1981.]
21. Kressel H., Butler J. K. Semiconductor Lasers and Heterojunction LEDs, Academic, N. Y., 1977.
22. Goldberg P. Ed., Luminiscence of Inorganic Solids, Academic, N. Y., 1966.
23. Ivey H. F. Electroluminescence and Semiconductor Lasers, *IEEE J. Quantum Electron.*, QE-2, 713 (1966).
24. Mooradian A., Fan H. Y. Recombination Emission in InSb, *Phys. Rev.*, **148**, 873 (1966).
25. Carr W. N. Characteristics of a GaAs Spontaneous Infrared Source with 40 Percent Efficiency, *IEEE Trans. Electron Devices*, ED-12, 531 (1965).
26. Ivey H. F. Electroluminescence and Related Effects, Suppl. 1 to Marton L., Ed., Advances in Electronics and Electron Physics, Academic, N. Y., 1963, p. 205.
27. Wang S. Solid-State Electronics, McGraw-Hill, N. Y., 1966.
28. Eastman P. C., Haering R. R., Barnes P. A. Injection Electroluminescence in Metal — Semiconductor Tunnel Diodes, *Solid State Electron.*, **7**, 879 (1964).
29. Gooch C. N. Injection Electroluminescent Devices, Wiley, N. Y., 1973.
30. Craford M. G. Recent Development in LED Technology, *IEEE Trans. Electron Devices*, ED-24, 935 (1977).
31. Groves W. O., Herzog A. H., Craford M. G. The Effect of Nitrogen Doping on GaAsP Electroluminescent Diodes, *Appl. Phys. Lett.*, **19**, 184 (1971).

32. Carr W. N., Pottman G. E. One Watt GaAs $p-n$ Junction Infrared Source, *Appl. Phys. Lett.*, **3**, 173 (1963).
33. Carr W. N. Photometric Figures of Merit for Semiconductor Luminescent Sources Operating in Spontaneous Mode, *Infrared Phys.*, **6**, 1 (1966).
34. Galginaitis S. V. Improving the External Efficiency of Electroluminescent Diodes, *J. Appl. Phys.*, **36**, 460 (1965).
35. Wight D. R. Green Luminescence Efficiency in GaP, *J. Phys. D*, **10**, 431 (1977).
36. Von Munch W. Silicon Carbide Technology for Blue Emitting Diodes, *J. Electron. Mater.*, **6**, 449 (1977).
37. Geusic J. E., Ostermayer F. W., Marcos H. M., Van Uitert L. G., Van Der Ziel J. P. Efficiency of Red, Green and Blue Infrared-to-Visible Conversion Sources, *J. Appl. Phys.*, **42**, 1958 (1971).
38. Dentaí A. G., Lee T. P., Burrus C. A. Small-Area High-Radiance cw InGaAsP LEDs Emitting at 1.2 to 1.3 μm , *Electron Lett.*, **13**, 484 (1977).
39. Burrus C. A., Miller B. I. Small-Area, Double Heterostructure AlGaAs Electroluminescent Diode Source for Optical-Fiber Transmission Lines, *Opt. Commun.*, **4**, 307 (1971).
40. Ettenberg M., Kressel H., Wittke J. P. Very High Radiance Edge-Emitting LED, *IEEE J. Quantum Electron.*, **QE-12**, 360 (1976).
41. Amann M. C., Proebster W. Small-Area GaAs — GaAlAs Heterostructure LED with Improved Current Confinement, *Electron. Lett.*, **15**, 599 (1979).
- 41a. Hall R. N., *Proc. Inst. Electr. Eng. Lond.*, **B106**, 923 (1959).
42. Ikeda K., Horiuchi S., Tanaka T., Susaki W. Design Parameters of Frequency Response of GaAs — AlGaAs DH LED's for Optical Communications, *IEEE Trans. Electron Devices*, **ED-24**, 10001 (1977).
43. Miya T., Terunuma Y., Hosaka T., Miyashita T. Ultimate Low-Loss Single Mode Fiber at 1.55 μm , *Electron. Lett.*, **15**, 108 (1979).
44. Foyt A. G. 1.0—1.6 μm Sources and Detectors for Fiber Optics Applications, *IEEE Device Res. Conf.*, Boulder, Colo., June 25, 1979.
45. Goodman C. H. L. Devices and Materials for 4 μm -band Fibre-Optical Communication, *Solid State Electron Devices*, **2**, 129 (1978).
46. Lin P. J., Kleinman L. Energy Bands of PbTe, PbSe and PbS, *Phys. Rev.*, **142**, 478 (1966).
47. Levine A. K. Lasers, *Am. Sci.*, **51**, 14 (1963).
48. Panish M. B., Hayashi I., Sumski S. Double-Heterostructure Injection Lasers with Room Temperature Threshold as Low as 2300 A/cm², *Appl. Phys. Lett.*, **16**, 326 (1970).
49. Casey H. C., Jr., Panish M. B., Merz J. L. Beam Divergence of the Emission from Double-Heterostructure Injection Lasers, *J. Appl. Phys.*, **44**, 5470 (1973).
50. D'Asaro L. A. Advances in GaAs Junction Lasers with Stripe Geometry, *J. Lumin.*, **7**, 310 (1973).
51. Paoli T. L. Waveguiding in a Stripe-Geometry Junction Laser, *IEEE J. Quantum Electron.*, **QE-13**, 662 (1977).
52. Yonezu H., Sakuma I., Kobayashi K., Kamejima T., Ueno M., Nantichi Y. A GaAs — Al_xGa_{1-x}As Double Heterostructure Planar Stripe Laser, *Jpn. J. Appl. Phys.*, **12**, 1585 (1973).
53. Nathan M. I. Semiconductor Lasers, *Proc. IEEE*, **54**, 1276 (1966).
54. Lasher G., Stern F. Spontaneous and Stimulated Recombination Radiation in Semiconductors, *Phys. Rev.*, **133**, A553 (1964).
55. Halperin B. I., Lax M. Impurity-Band Tails in the High-Density Limit. I. Minimum Counting Methods, *Phys. Rev.*, **148**, 722 (1966).
56. Stern F. Calculated Spectral Dependence of Gain in Excited GaAs, *J. Appl. Phys.*, **47**, 5382 (1976).
57. Casey H. C., Jr., Room Temperature Threshold-Current Dependence of GaAs — Al_xGa_{1-x}As Double Heterostructure Lasers on x and Active-Layer Thickness, *J. Appl. Phys.*, **49**, 3684 (1978).

58. Nahory R. E., Pollack M. A. Threshold Dependence on Active-Layer Thickness in InGaAsP/InP DH Lasers, *Electron. Lett.*, **14**, 727 (1978).
59. Yana M., Nishi H., Takusagawa M. Theoretical and Experimental Study of Threshold Characteristics in InGaAsP/InP DH Lasers, *IEEE J. Quantum Electron.*, **QE-15**, 571 (1979).
60. Burrus C. A., Casey H. C., Jr., Li T. Y. Optical Sources, in Miller S. E., Chynoweth A. G., Eds., *Optical Fiber Communication*, Academic, N. Y., 1979.
61. Amann M. C. New Stripe-Geometry Laser with Simplified Fabrication Process, *Electron. Lett.*, **15**, 441 (1979).
62. Aiki K., Nakamura M., Kuroda T., Umeda J. Channeled-Substrate Planar Structure AlGaAs Injection Laser, *Appl. Phys. Lett.*, **30**, 649 (1977).
63. Namizaki H. Transverse-Junction-Stripe Lasers with a GaAs $p-n$ Homojunction, *IEEE J. Quantum Electron.*, **QE-11**, 427 (1975).
64. Tsukada T. GaAs — AlGaAs Buried-Heterostructure Injection Lasers. *J. Appl. Phys.*, **45**, 4899 (1974).
65. Tsang W. T., Logan R. A., Van der Ziel J. P. Low-Current-Threshold Stripe-Buried-Heterostructure Lasers with Self-Aligned Current Injection Stripes, *Appl. Phys. Lett.*, **34**, 644 (1979).
66. Gosch J. Simple Technology Irons Out Kinks in Laser Diode Output, *Electronics* **52**, (26), 59 (1979).
67. Nishi H., Yano M., Nishitani Y., Akita Y., Takusagawa M. Self-Aligned Structure InGaAsP/InP Lasers, *Appl. Phys. Lett.*, **35**, 232 (1979).
68. Kaminow I. P., Nahory R. E., Pollack M. A., Stutz L. W., Dewinter J. C. Single-Mode cw Ridge-Waveguide Laser Emitting at 1.55 μm , *Electron. Lett.*, **15**, 763 (1979).
69. Kogelnik H., Shank C. V. Stimulated Emission in a Periodic Structure, *Appl. Phys. Lett.*, **18**, 152 (1971); Coupled-Wave Theory of Distributed Feedback Lasers, *J. Appl. Phys.*, **43**, 2327 (1973).
70. Aiki K., Nakamura M., Umeda J. Lasing Characteristics of Distributed-Feedback GaAs — GaAlAs Diode Lasers with Separate Optical and Carrier Confinement, *IEEE J. Quantum Electron.*, **QE-12**, 597 (1976).
71. Hwang C. J., Dymont J. C. Dependence of Threshold and Electron Lifetime on Acceptor Concentration in GaAs — Ga_{1-x}Al_xAs Lasers, *J. Appl. Phys.*, **44**, 3240 (1973).
72. Akiba S., Sakai K., Yamamoto T. Direct Modulation of InGaAsP/InP Double Heterostructure Lasers, *Electron. Lett.*, **14**, 197 (1978).
73. Melngailis I., Mooradian A. Tunable Semiconductor Diode Lasers and Applications, in Jacobs S., Sargent M., Scott J. F., Seully M. O., Eds., *Laser Applications to Optics and Spectroscopy*. Addison-Wesley Reading, Mass., 1975.
74. Walpole J. N., Calawa A. R., Harman T. C., Groves S. H. Double-Heterostructure PbSnTe Lasers Grown by Molecular-Beam Epitaxy with CW Operation up to 114 K, *Appl. Phys. Lett.*, **28**, 552 (1976).
75. Yonezu H., Sakuma I., Kamojima T., Ueno M., Iwamoto K., Hino I., Hayashi I. High Optical Power Density Emission from a Window Stripe AlGaAs DH Laser, *Appl. Phys. Lett.*, **34**, 637 (1979).
76. Hartman R. L., Schumaker N. E., Dixon R. W. Continuously Operated AlGaAs DH Lasers with 70 °C Lifetimes as Long as Two Years, *Appl. Phys. Lett.*, **31**, 756 (1977).
77. Lo W., Swets D. E. Room Temperature 4.5 μm Light Emitting Diodes, *Appl. Phys. Lett.*, **36**, 450 (1980).
78. Holonyak N., Jr., Kolbas R. M., Dupuis R. D., Dapkus P. D. Quantum-Well Heterostructure Lasers, *IEEE J. Quantum Electron.*, **QE-16**, 170 (1980).

Глава 13

ФОТОДЕТЕКТОРЫ

13.1. ВВЕДЕНИЕ

Фотодетекторы представляют собой полупроводниковые приборы, которые могут детектировать оптические сигналы посредством протекающих в них электронных процессов. Расширение спектрального диапазона излучения когерентных и некогерентных источников света в дальнюю инфракрасную область, с одной стороны, и в ультрафиолетовую — с другой, привело к увеличению потребности в быстродействующих чувствительных фотодетекторах. В общем случае в фотодетекторе протекают три основных процесса: 1) генерация носителей под действием внешнего излучения; 2) перенос носителей и их умножение за счет того или иного механизма усиления тока, характерного для данного полупроводника; 3) взаимодействие тока с внешней цепью, обеспечивающее получение выходного сигнала.

Фотодетекторы играют важную роль в волоконно-оптических системах связи ближнего ИК-диапазона (0,8—1,6 мкм). Они детектируют оптические сигналы, т. е. преобразуют изменения интенсивности излучения в электрические колебания, которые соответствующим образом усиливаются и подвергаются последующей обработке. Фотодетекторы, предназначенные для таких систем, должны обладать высокой чувствительностью в рабочем диапазоне длин волн, высоким быстродействием и низким уровнем шумов. Кроме того, они должны иметь малые размеры, низкие управляющие напряжения или токи, а также высокую надежность в работе.

Быстродействующие фотодетекторы рассмотрены в работах [1—3]. В работе [4] имеется исчерпывающий обзор инфракрасных фотодетекторов, в котором подробно рассмотрены лавинные фотодиоды [5]. Вопросам использования фотодетекторов в составе волоконно-оптических систем связи посвящена работа [6].

13.2. ФОТОРЕЗИСТОР

Фоторезистор представляет собой пластину полупроводника (монокристаллическую или в виде тонкой пленки), на противоположных концах которой созданы омические контакты (рис. 1). При падении излучения на поверхность фоторезистора в последнем генерируются носители вследствие возбуждения либо междоузельных переходов (собственное возбуждение), либо переходов с участием

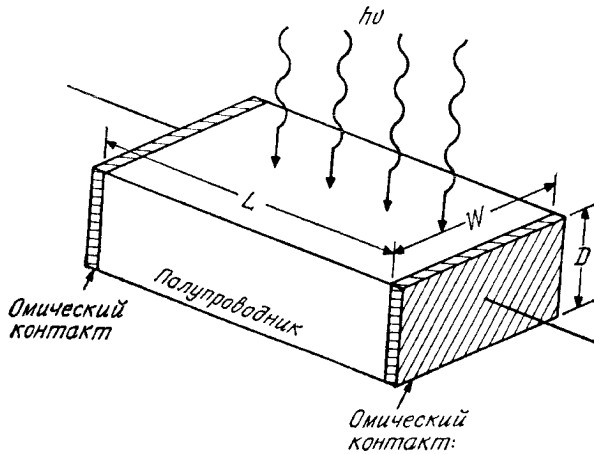


Рис. 1. Фоторезистор, состоящий из пластины полупроводника с омическими контактами.

энергетических уровней в запрещенной зоне (примесное возбуждение), что приводит к увеличению проводимости. Процессы собственного и примесного фотовозбуждения носителей схематически показаны на рис. 2.

Проводимость собственных фоторезисторов описывается формулой $\sigma = q(\mu_n n + \mu_p p)$, и увеличение проводимости под действием освещения в основном связано с увеличением числа носителей. Длинноволновая граница фотопроводимости в этом случае определяется из соотношения

$$\lambda_c = \frac{hc}{E_g} = \frac{1,24}{E_g(\text{эВ})}, \quad (1)$$

где λ_c — длина волны, соответствующая ширине запрещенной зоны E_g . Излучение с длинами волн, меньшими λ_c , поглощается в полупроводнике с образованием электронно-дырочных пар.

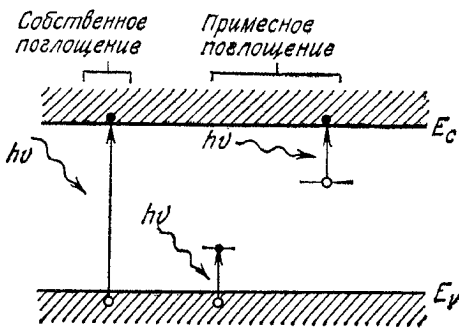


Рис. 2. Процессы собственного (зона — зона) и примесного фотовозбуждения.

В примесном фоторезисторе фотовозбуждение может происходить между краем зоны и энергетическим уровнем в запрещенной зоне. Фотопроводимость может возникать в результате поглощения фотонов с энергией, которая равна или превышает энергию, отделяющую примесный уровень от зоны проводимости или валентной зоны. При этом длинноволновая граница фотопроводимости определяется глубиной залегания примесного уровня в запрещенной зоне.

Работа фотодетекторов вообще и фоторезисторов в частности характеризуется тремя параметрами: 1) квантовой эффективностью или усилением, 2) временем фотоответа и 3) чувствительностью (обнаружительной способностью). Рассмотрим сначала процессы, протекающие в фоторезисторе при воздействии излучения (рис. 1). Пусть в какой-то начальный момент времени $t = 0$ число носителей, возникших в единице объема за счет генерации потоком фотонов, равно n_0 . В последующие моменты времени t число носителей в том же объеме уменьшается за счет рекомбинации по закону $n = n_0 \exp(-t/\tau)$, где τ — время жизни носителей. Другими словами, скорость рекомбинации равна $1/\tau$. Если поток фотонов постоянен и распределен равномерно по поверхности фоторезистора с площадью $A = WL$, то общее число фотонов, падающих на поверхность в единицу времени, равно $P_{opt}/h\nu$, где P_{opt} — мощность падающего излучения и $h\nu$ — энергия фотона.

В стационарном состоянии скорость генерации носителей должна быть равна скорости рекомбинации. Если толщина прибора D значительно больше, чем глубина проникновения света $1/\alpha$, то общая стационарная скорость генерации носителей в единице объема равна

$$G = \frac{n}{\tau} = \frac{\eta (P_{opt}/h\nu)}{WLD}, \quad (2)$$

где η — квантовая эффективность (т. е. число сфотогенерированных носителей, отнесенное к числу падающих фотонов) и n — число носителей в единице объема (плотность носителей). Фототок, протекающий между контактами, равен

$$I_p = (\sigma \mathcal{E}) WD = (q\mu_n n \mathcal{E}) WD = (qnv_d) WD, \quad (3)$$

где \mathcal{E} — электрическое поле внутри фоторезистора и v_d — дрейфовая скорость. Подставляя n из уравнения (2) в уравнение (3), получим

$$I_p = q \left(\eta \frac{P_{opt}}{h\nu} \right) \left(\frac{\mu_n \tau \mathcal{E}}{L} \right). \quad (3a)$$

Определяя исходный фототок как

$$I_{ph} \equiv q \left(\eta \frac{P_{opt}}{h\nu} \right),$$

Таблица 1. Типичные значения коэффициента усиления и времени фотоответа

Фотодетектор	Усиление	Время фотоответа, с	Рабочая температура, К
Фоторезистор	1—10 ⁶	10 ⁻³ —10 ⁻⁸	4,2—300
<i>p-n</i> -переход	1	10 ⁻¹¹	300
<i>p-l-n</i> -переход	1	10 ⁻⁸ —10 ⁻¹⁰	300
Диод со структурой металл—полупроводник	1	10 ⁻¹¹	300
Лавинный фотодиод	10 ² —10 ⁴	10 ⁻¹⁰	300
Биполярный фототранзистор	10 ²	10 ⁻⁸	300
Полевой транзистор	10 ²	10 ⁻⁷	300

из уравнения (3а) получим коэффициент усиления фототока

$$\frac{I_p}{I_{ph}} = \frac{\mu_n \tau \mathcal{E}}{L} = \frac{\tau}{t_r}, \quad (4)$$

где $t_r = L/v_d$ — время пролета носителей. Коэффициент усиления, зависящий от отношения времени жизни и времени пролета, является исключительно важным параметром фоторезисторов. Для образцов с большим временем жизни и малым расстоянием между контактами коэффициент усиления может быть существенно больше единицы. Время фотоответа фоторезистора определяется временем пролета t_r . Поскольку для фоторезисторов характерны большие расстояния между контактами и слабые электрические поля, то их время фотоответа обычно больше аналогичного параметра для фотодиодов. Некоторые типичные значения коэффициента усиления и времени фотоответа приведены в табл. 1 [2, 8].

Рассмотрим теперь оптический сигнал с интенсивностью, модулированной по закону

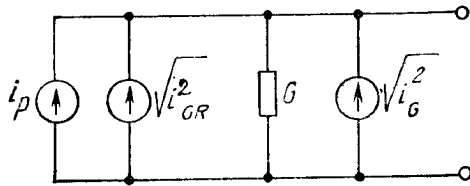
$$P(\omega) = P_{opt}(1 + m e^{i\omega t}), \quad (5)$$

где P_{opt} — средняя мощность оптического сигнала, m — коэффициент модуляции и ω — частота модуляции. Средний ток I_p , обусловленный оптическим сигналом, определяется выражением (3а). Среднеквадратичная мощность модулированного оптического сигнала равна $mP_{opt}/\sqrt{2}$, а среднеквадратичный токовый сигнал может быть представлен в виде [9]

$$i_p \approx \frac{q\eta m P_{opt}}{\sqrt{2} h\nu} \left(\frac{\tau}{t_r} \right) \frac{1}{(1 + \omega^2 \tau^2)^{1/2}}. \quad (6)$$

На рис. 3 показана эквивалентная схема фоторезистора. Проводимость G включает проводимость, обусловленную темновым током,

рис. 3. Эквивалентная схема фотрезистора [9].



и проводимость, вызываемую фототоком и фоновым током. Тепловой шум, обусловленный проводимостью G , равен

$$i_G^2 = 4kTGB, \tag{7}$$

где k — константа Больцмана, T — абсолютная температура и B — ширина полосы. Генерационно-рекомбинационный шум (дробовой шум) определяется выражением [10]

$$\overline{i_{GR}^2} = \frac{\tau}{t_r} \frac{4qI_0B}{1 + \omega^2\tau^2}, \tag{8}$$

где I_0 — стационарный выходной фототок, равный I_p ($=\tau I_{ph}/t_r$). Используя уравнения (6)—(8), можно получить выражение для отношения сигнал/шум:

$$(S/N)_{power} = \frac{i_p^2}{i_{GR}^2 + i_G^2} = \frac{\eta m^2 (P_{opt}/h\nu)}{8B} \left[1 + \frac{kT}{q} \frac{t_r}{\tau} (1 + \omega^2\tau^2) \frac{G}{I_0} \right]^{-1}. \tag{9}$$

Относительным критерием качества фотодетекторов [11] является мощность, эквивалентная шуму (NEP)¹, которая определяется как среднеквадратичная мощность падающего излучения, необходимая для получения отношения сигнал/шум, равного 1, в полосе частот 1 Гц. Выражение для NEP (т. е. $mP_{opt}/\sqrt{2}$) можно получить из уравнения (9), подставив в него $S/N = 1$ и $B = 1$. Для инфракрасных детекторов наиболее употребительным критерием качества является удельная обнаружительная способность D^* [12]:

$$D^* = \frac{A^{1/2}B^{1/2}}{NEP} \text{ [см (Гц)}^{1/2}\text{/Вт)].} \tag{10}$$

Для того чтобы исключить неопределенность в D^* , необходимо указывать, является ли источником излучения абсолютно черное тело или монохроматический источник и при какой частоте осуществляется модуляция. Рекомендуется выражать D^* как $D^*(\lambda, f, 1)$ или $D^*(T, f, 1)$, где λ — длина волны (мкм), f — частота модуляции (Гц), T — температура (К). При этом ширина полосы всегда равна 1 Гц.

Для фоторезистора, чувствительность которого ограничена фоновым излучением, идеальная обнаружительная способность D^*

¹ Сокращенное обозначение словосочетания noise equivalent power (NEP). — Прим. перев.

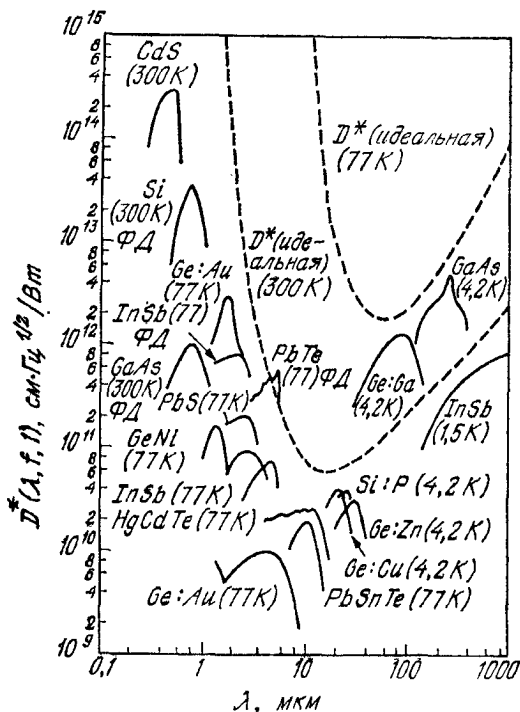


Рис. 4. Зависимость обнаружительной способности D^* от длины волны излучения для различных фоторезисторов и фотодиодов (фотодиоды обозначены ФД). Штриховыми кривыми представлена теоретическая идеальная D^* при 77 и 300 К и угле 2π ер [2, 3].

(рис. 4) при квантовой эффективности, равной 1, определяется выражением [12]

$$D^*(\lambda, f, 1) = \frac{c \exp(\zeta)}{2 \sqrt{\pi \hbar k T} v^2 (1 + 2\zeta + 2\zeta^2)^{1/2}}, \quad (11)$$

где c — скорость света и $\zeta \equiv \hbar\nu/kT$. На рис. 4 представлены также некоторые типичные значения D^* для таких фоторезисторов, как CdS, PbS, InSb, германий, легированный никелем (Ge—Ni), кремний, легированный фосфором Si—P [4, 7] (энергетические уровни примесей см. в гл. 1). Отметим, что для детектирования излучения в среднем, дальнем и сверхдальнем ИК-диапазонах фоторезисторы охлаждаются до низких температур (77 и 4,2 К). При таких температурах уменьшаются тепловые эффекты, вызывающие термическую ионизацию и опустошение энергетических уровней, и увеличиваются усиление и эффективность детектирования. На длинах волн вблизи 0,5 мкм наиболее высокой чувствительностью обладает фоторезистор на CdS, в то время как при $\lambda = 10$ мкм предпочтительны фоторезисторы на HgCdTe [13]. Для детектирования излучения в диапазоне длин волн 100—400 мкм лучше всего выбрать фоторезистор на GaAs [14], так как он

имеет большие значения D^* . Данный фоторезистор обладает широким динамическим диапазоном и вследствие этого может обеспечить хорошие характеристики детектирования сигналов высокой интенсивности. Однако для детектирования слабых сигналов в микроволновом диапазоне частот фотодиоды обеспечивают большее быстродействие и сравнительно более высокое отношение сигнал/шум. Таким образом, применение фоторезисторов в качестве высокочастотных оптических демодуляторов, например смесителей мощных оптических сигналов, ограничено. Вместе с тем они широко используются для детектирования в инфракрасной области спектра, особенно на длинах волн выше нескольких микрометров, где, несмотря на интенсивную работу, пока еще отсутствуют удовлетворительные альтернативные варианты методов детектирования.

13.3. ФОТОДИОД

13.3.1. Общий анализ

В фотодиоде имеется обедненная область полупроводника с сильным электрическим полем, в котором происходит разделение электронно-дырочных пар, возбужденных под действием света. Для работы фотодиода на высоких частотах необходимо обеспечить малые времена пролета, поэтому обедненная область должна быть тонкой. С другой стороны, для увеличения квантовой эффективности (число фотогенерированных электронно-дырочных пар, отнесенное к числу падающих фотонов) обедненный слой должен быть достаточно толстым, чтобы обеспечить поглощение большей части падающего излучения. Таким образом, существует взаимосвязь между быстродействием и квантовой эффективностью.

Фотодиод может работать в фотовольтаическом режиме, характерном для солнечных батарей (гл. 14) — несмещенный диод подсоединяется к нагрузочному сопротивлению. Однако конструктивно фотодиод существенно отличается от солнечной батареи. В работе фотодиода определяющую роль играет лишь излучение, сосредоточенное в узком интервале длин волн в центре оптического диапазона, а солнечные батареи должны обладать высокой чувствительностью в широком диапазоне длин волн солнечного излучения. Фотодиоды имеют малые размеры, необходимые для минимизации емкости перехода, тогда как солнечные батареи относятся к приборам с большой площадью. Одной из наиболее важных характеристик фотодиодов является квантовая эффективность, тогда как качество солнечных батарей в основном оценивается по эффективности преобразования мощности (мощность, выделяемая на нагрузку, отнесенная к мощности падающего солнечного излучения).

При работе в видимом и ближнем инфракрасном диапазонах диоды обычно смещаются в обратном направлении с помощью сравнительно большого напряжения, чтобы уменьшить время пролета носителей и снизить емкость перехода. Напряжение смещения, однако, не настолько велико, чтобы вызвать лавинный пробой. Этим ограничением на величину смещения отличаются рассматриваемые фотодиоды от лавинных фотодиодов, в которых внутреннее усиление достигается как раз за счет ударной ионизации в условиях лавинного пробоя. В семейство фотодиодов входят диоды с $p-n$ -переходом, $p-i-n$ -диоды, диоды со структурой металл—полупроводник (с барьером Шоттки) и диоды с гетеропереходом.

В этом разделе мы кратко рассмотрим общие характеристики фотодиодов: квантовую эффективность, скорость фотоответа и приборные шумы.

Квантовая эффективность, как уже говорилось выше, представляет собой число сфотогенерированных электронно-дырочных пар, отнесенное к числу падающих фотонов:

$$\eta = (I_p/q)/(P_{\text{opt}}/h\nu), \quad (12)$$

где I_p — фототок, обусловленный поглощением падающего оптического излучения с мощностью P_{opt} и длиной волны λ (соответствующей энергии фотона $h\nu$).

Сравнительным критерием качества является чувствительность, которая определяется как отношение фототока к оптической мощности:

$$\mathcal{R} = \frac{I_p}{P_{\text{opt}}} = \frac{\eta q}{h\nu} = \frac{\eta \lambda}{1,24}. \quad (13)$$

Таким образом, для данной квантовой эффективности чувствительность линейно растет с ростом λ . Для идеального фотодиода ($\eta = 1$) $\mathcal{R} = (\lambda/1,24)$ (А/Вт), где длина λ выражена в микрометрах.

Одним из основных факторов, определяющих квантовую эффективность, является коэффициент поглощения. На рис. 5 и 6 приведены зависимости коэффициентов поглощения α от длины волны, измеренные для различных фоточувствительных материалов [2]. Для Ge, Si и полупроводниковых соединений типа $A^{III}B^V$ при увеличении температуры кривые сдвигаются в область больших длин волн. Для некоторых соединений типа $A^{IV}B^{VI}$ (например, PbSe) имеет место обратная ситуация, что обусловлено расширением запрещенной зоны при увеличении температуры. На рисунках указаны также длины волн, соответствующие излучению некоторых важных лазеров.

Вследствие того что α сильно зависит от длины волны, для данного полупроводника область длин волн, в которой можно получить заметный фототок, ограничена. Длинноволновая гра-

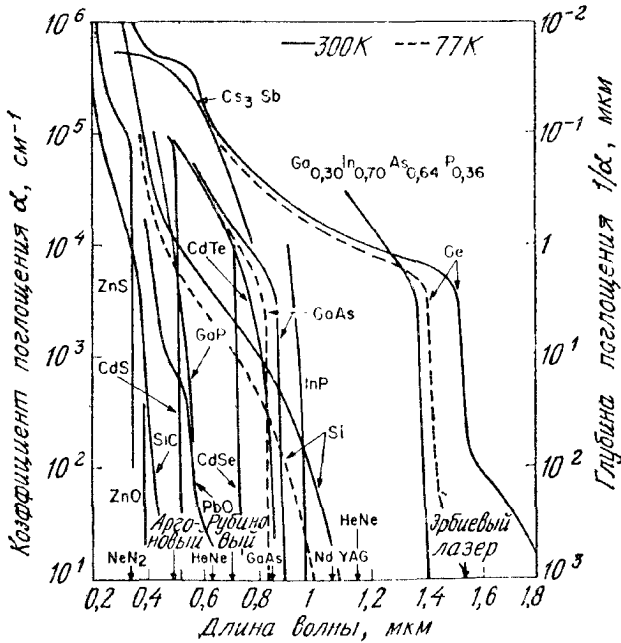


Рис. 5. Зависимость коэффициента поглощения от длины волны излучения для различных фоточувствительных материалов [2].

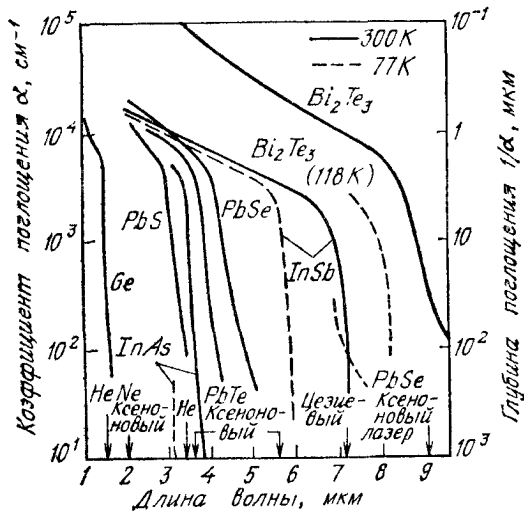


Рис. 6. Зависимость коэффициента поглощения от длины волны излучения для материалов, чувствительных в ИК-области спектра [2].

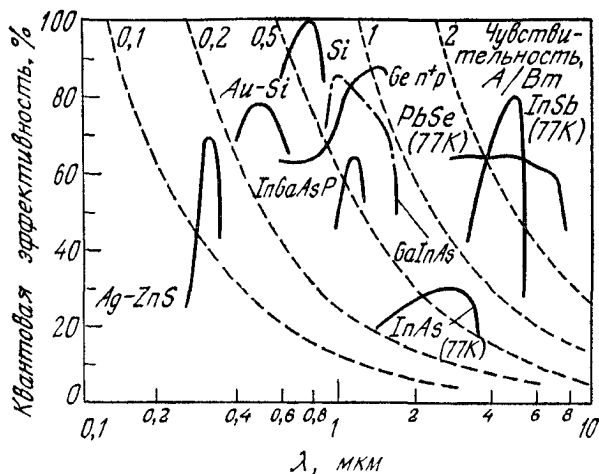


Рис. 7. Зависимости квантовой эффективности (сплошная кривая) и чувствительности (штриховая кривая) различных фотодетекторов от длины волны излучения.

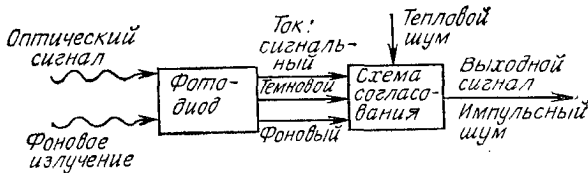
ница λ_c определяется шириной запрещенной зоны полупроводника в соответствии с уравнением (1) и равна $\sim 1,7$ мкм для Ge и $\sim 1,1$ мкм для Si. Для длин волн, больших λ_c , значение α слишком мало, чтобы обеспечить заметное поглощение. Коротковолновая граница фотоответа обусловлена тем, что в области коротких волн коэффициент α очень велик ($\geq 10^5$ см $^{-1}$) и излучение поглощается в непосредственной близости поверхности, где времена рекомбинации малы. Вследствие этого фотоносители рекомбинируют раньше, чем уходят из поверхностного слоя. На рис. 7 приведены типичные зависимости квантовой эффективности от длины волны для некоторых высокочастотных фотодиодов и кривые квантовой эффективности, соответствующие заданным значениям чувствительности. В ультрафиолетовой и видимой областях спектра высокой квантовой эффективности обладают фотодиоды со структурой металл—полупроводник. В ближнем ИК-диапазоне (на длинах волн $\sim 0,8$ — $0,9$ мкм) в кремниевых фотодиодах с просветляющим покрытием может быть достигнута квантовая эффективность 100 %. В области длин волн $1,0$ — $1,6$ мкм высокой квантовой эффективностью характеризуются фотодиоды на Ge, а также на основе тройных (например, GaInAs) и четверных (например, GaInAsP) соединений элементов III и V групп. Для обеспечения высокоэффективной работы в длинноволновой области фотодиоды охлаждаются до низких температур (например, до 77 К).

Скорость фотоответа ограничивается тремя факторами: диффузией носителей, временем дрейфа через обедненную область и емкостью обедненной области. Носители, генерированные за пре-

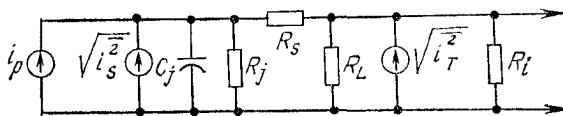
делами обедненной области, должны диффундировать к переходу, что существенно увеличивает время задержки. Чтобы уменьшить влияние диффузии, необходимо формировать переход вблизи поверхности. Основная часть излучения поглощается в том случае, когда обедненная область достаточно широка (порядка $1/\alpha$); при достаточно высоких обратных смещениях дрейф носителей осуществляется со скоростью насыщения. Обедненный слой, однако, не должен быть слишком широким, иначе время переноса будет ограничивать скорость фотосоответа. Кроме того, он не должен быть и слишком узким, иначе за счет большой емкости C будет велика постоянная времени RC (R — сопротивление нагрузки). Оптимальной шириной обедненного слоя считается такая ширина, при которой время переноса равно примерно половине периода модуляции излучения. Например, для частоты модуляции, равной 10 ГГц, оптимальная ширина обедненного слоя в кремнии (скорость насыщения $\sim 10^7$ см/с) составляет ~ 5 мкм.

Для изучения шумовых свойств фотодиода рассмотрим обобщенный процесс детектирования излучения, схематически представленный на рис. 8, а [5]. При поглощении оптического сигнала и фонового излучения в фотодиоде генерируются электронно-дырочные пары. Электроны и дырки разделяются электрическим полем и дрейфуют в противоположные стороны от $p-n$ -перехода. При этом через внешнее нагрузочное сопротивление протекает ток смещения. Для определения генерационного тока, возникающего в результате этого фотоэлектронного процесса, рассмотрим модулированный по интенсивности оптический сигнал, заданный уравнением (5). Средний фототок, обусловленный оптическим сигналом, равен

$$I_P = q\eta P_{\text{opt}}/h\nu. \tag{14}$$



а



б

Рис. 8. Процесс детектирования излучения (а) и эквивалентная схема фотодиода (б) [5].

Среднеквадратичная мощность модулированного оптического сигнала равна $mP_{\text{opt}}/\sqrt{2}$, а среднеквадратичный ток сигнала можно записать в виде

$$i_p = q\eta mP_{\text{opt}}/\sqrt{2} h\nu. \quad (15)$$

Ток, обусловленный фоновым излучением, равен I_B , а темновой ток термической генерации электронно-дырочных пар в обедненной области I_D . Вследствие случайного характера процесса генерации этих токов дробовой шум определяется выражением

$$\langle i_s^2 \rangle = 2q(I_P + I_B + I_D)B, \quad (16)$$

где B — полоса частот.

Эквивалентная схема фотодиода [5] показана на рис. 8, б, где i_p , V_{i_s} , C_j , R_j и R_s определяются параметрами фотодиода. Компонента C_j представляет собой емкость перехода, R_j — сопротивление перехода, а R_s — последовательное сопротивление. Изменяемое сопротивление R_L соответствует внешнему нагрузочному сопротивлению, а R_i — входному сопротивлению усилителя [55]. Все сопротивления вносят в систему дополнительные тепловые шумы. Последовательное сопротивление R_s обычно намного меньше, чем другие сопротивления, поэтому его можно не учитывать. Тепловой шум равен

$$\langle i_T^2 \rangle = 4kT(1/R_{eq})B, \quad (17)$$

где $1/R_{eq} = (1/R_j) + (1/R_L) + (1/R_i)$.

При 100 %-ной модуляции оптического сигнала со средней мощностью P_{opt} выражение для отношения сигнал/шум можно записать в виде

$$(S/N)_{\text{power}} = \frac{i_p^2 R_{eq}}{(\langle i_s^2 \rangle + \langle i_T^2 \rangle) R_{eq}} = \frac{1/2 (q\eta P_{\text{opt}}/h\nu)^2}{2q(I_P + I_B + I_D)B + 4kTB/R_{eq}}. \quad (18)$$

Из этого уравнения следует, что минимальная оптическая мощность, необходимая для получения заданного отношения сигнал/шум, равна

$$(P_{\text{opt}})_{\text{мин}} = \frac{2h\nu B}{\eta} \left(\frac{S}{N} \right) \left\{ 1 + \left[1 + \frac{I_{eq}}{qB(S/N)} \right]^{1/2} \right\}, \quad (19)$$

где $I_{eq} = I_B + I_D + 2kT/qR_{eq}$. В пределе, когда $I_{eq}/qB(S/N) \ll \ll 1$, минимальная оптическая мощность определяется квантовыми шумами, связанными с самим оптическим сигналом. В другом предельном случае, когда $I_{eq}/qB(S/N) \gg 1$, преобладающими становятся шум фонового излучения и тепловой шум эквивалентного

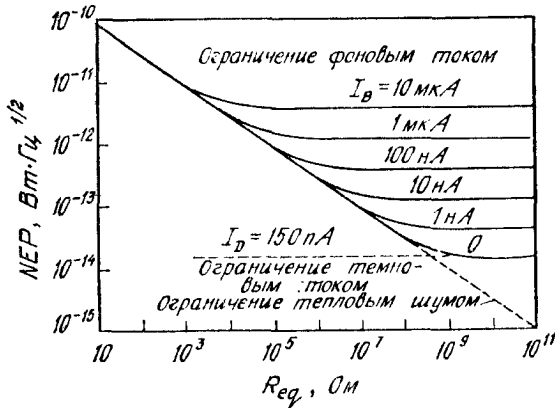


Рис. 9. Зависимость NEP $p-i-n$ -диода от величины нагрузочного сопротивления при темновом токе 150 пА и различных значениях фонового тока.

сопротивления. В этих условиях мощность, эквивалентная шуму (NEP) и равная среднеквадратичной оптической мощности при $S/N = 1$ и $B = 1$ Гц, описывается выражением

$$NEP = \sqrt{2} (h\nu/\eta) (I_{eq}/q)^{1/2} [\text{Вт/см}^2 \text{ Гц}^{1/2}]. \quad (20)$$

Для того чтобы улучшить чувствительность фотодиода, необходимо η и R_{eq} увеличивать, а I_B и I_D уменьшать. Зависимость NEP от величины нагрузочного сопротивления для типичного кремниевого фотодиода, вычисленная при $\eta = 75\%$, $\lambda = 0,77$ мкм и $I_D = 1,5 \cdot 10^{-10}$ А [5], приведена на рис. 9. Из графика следует, что NEP при наличии дробового шума темнового тока или фонового излучения может быть достигнута лишь при достаточно высоких значениях R_{eq} .

13.3.2. $p-i-n$ -Фотодиод

$p-i-n$ -Фотодиод является наиболее распространенным фотодетектором, поскольку толщина обедненной области (слоя собственной проводимости) может быть создана такой, чтобы обеспечить оптимальную квантовую эффективность и быстродействие. На рис. 10 схематически представлены структура $p-i-n$ -диода, его энергетическая диаграмма в условиях обратного смещения и распределение интенсивности излучения [2]. С помощью этой схемы мы проведем анализ некоторых особенностей работы $p-i-n$ -фотодиода. Этот анализ будет справедливым также и для фотодиода с $p-n$ -переходом. Поглощение излучения в полупровод-

нике приводит к генерации электронно-дырочных пар. Пары, создаваемые в обедненной области или на расстоянии диффузионной длины от нее, разделяются электрическим полем, и во внешней цепи во время дрейфа носителей через обедненную область появляется ток.

В стационарных условиях плотность полного тока, протекающего через обратносмещенный обедненный слой, равна [15]

$$J_{\text{tot}} = J_{\text{dr}} + J_{\text{diff}}, \quad (21)$$

где J_{dr} — плотность дрейфового тока, обусловленного генерацией носителей в обедненной области, и J_{diff} — плотность диффузионного тока, обусловленного генерацией носителей в объеме полупроводника за пределами обедненного слоя и их диффузией к обратносмещенному переходу. Выведем выражение для полного тока при условии, что током термической генерации можно пренебречь и что приповерхностный слой p -типа существенно меньше $1/\alpha$. В соответствии с рис. 10, a выражение для скорости генерации электронно-дырочных пар имеет вид

$$G(x) = \Phi_0 \alpha e^{-\alpha x}. \quad (22)$$

Здесь Φ_0 — поток падающих фотонов, приходящийся на единицу площади и равный $P_{\text{opt}}(1-R)/h\nu$, где R — коэффициент

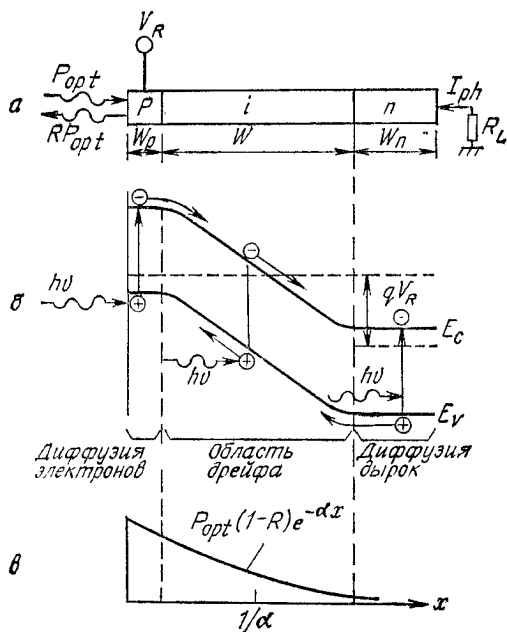


Рис. 10. Принцип работы фотодиода [2].

a — поперечный разрез p - i - n -диода; b — зонная диаграмма в условиях обратного смещения; c — распределение интенсивности излучения.

отражения и A — площадь прибора. Таким образом, дрейфовый ток

$$J_{dr} = -q \int_0^W G(x) dx = q\Phi_0 (1 - e^{-\alpha W}), \quad (23)$$

где W — толщина обедненного слоя. При $x > W$ плотность неосновных носителей (дырок) в объеме полупроводника определяется из одномерного диффузионного уравнения

$$D_p \frac{\partial^2 p_n}{\partial x^2} - \frac{p_n - p_{n0}}{\tau_p} + G(x) = 0, \quad (24)$$

где D_p — коэффициент диффузии дырок, τ_p — время жизни неравновесных носителей и p_{n0} — равновесная плотность дырок. Решение уравнения (24) при граничных условиях $p_n = p_{n0}$ при $x = \infty$ и $p_n = 0$ при $x = W$ имеет вид

$$p_n = p_{n0} - (p_{n0} + C_1 e^{-\alpha W}) e^{(W-x)/L_n} + C_1 e^{-\alpha x}, \quad (25)$$

где $L_p = \sqrt{D_p \tau_p}$ и

$$C_1 \equiv \left(\frac{\Phi_0}{D_p} \right) \frac{\alpha L_p^2}{1 - \alpha^2 L_p^2}. \quad (26)$$

Плотность диффузионного тока, определяемая как $J_{diff} = -qD_p (\partial p_n / \partial x)_{x=W}$, равна

$$J_{diff} = q\Phi_0 \frac{\alpha L_p}{1 + \alpha L_p} e^{-\alpha W} + qp_{n0} \frac{D_p}{L_p}. \quad (27)$$

Плотность полного тока описывается выражением

$$J_{tot} = q\Phi_0 \left(1 - \frac{e^{-\alpha W}}{1 + \alpha L_p} \right) + qp_{n0} \frac{D_p}{L_p}. \quad (28)$$

При нормальных условиях работы член, содержащий p_{n0} , значительно меньше первого члена и полный фототок пропорционален потоку фотонов. Из уравнений (12) и (28) можно вывести выражение для квантовой эффективности

$$\eta = \frac{J_{tot}/q}{P_{opt}/Ah\nu} = (1 - R) \left(1 - \frac{e^{-\alpha W}}{1 + \alpha L_p} \right). \quad (29)$$

Для получения высокой квантовой эффективности желательно, чтобы коэффициент отражения был мал и выполнялось условие $\alpha W \gg 1$. Однако при $W \gg 1/\alpha$ может оказаться значительным время пролета. Рассмотрим теперь влияние времени пролета.

Так как для прохождения носителей через обедненный слой требуется конечное время, то при высокочастотной модуляции интенсивности падающего излучения появляется разность фаз между потоком фотонов и фототоком. Для получения количественного выражения, описывающего этот эффект, обратимся к рис.

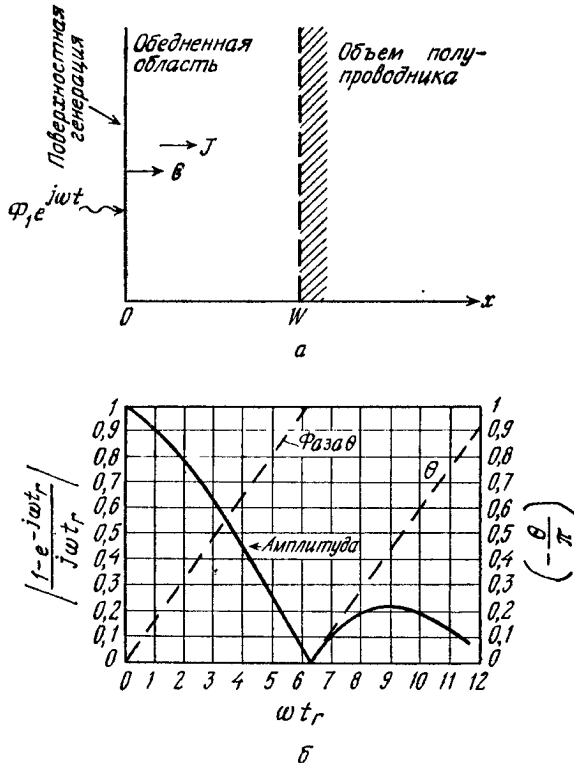


Рис. 11. Представление диода (а), принятое при рассмотрении эффектов, связанных с влиянием времени пролета, и фотоответ (нормированный ток или напряжение) (б) как функция от нормализованной частоты модулятора падающего потока фотонов (здесь $\theta \equiv \omega t_r/2$) [15].

11, а, на котором схематически изображен наиболее простой случай. Предполагается, что приложенное напряжение достаточно велико для того, чтобы область собственной проводимости была приведена в состояние полного обеднения и скорость дрейфа носителей достигла значения скорости насыщения. При плотности потока фотонов, равной $\Phi_1 e^{j\omega t}$ (фотон/с см²), плотность тока проводимости J_{cond} в точке x оказывается равной

$$J_{\text{cond}}(x) = q\Phi_1 e^{j\omega(t-x/v_s)}. \quad (30)$$

Поскольку $\nabla \cdot J_{\text{tot}} = 0$, то можно написать

$$J_{\text{tot}} = \frac{i}{W} \int_0^W \left(J_{\text{cond}} + \epsilon_s \frac{\partial \mathcal{E}}{\partial t} \right) dx, \quad (31)$$

где второй член в скобках представляет собой плотность тока смещения, а ϵ_s и \mathcal{E} — диэлектрическая постоянная и электрическое поле соответственно. Подставляя уравнение (30) в уравнение (31), получим

$$J_{\text{tot}} = \left(\frac{j\omega\epsilon_s V}{W} + q\Phi_1 \frac{1 - e^{-j\omega t_r}}{j\omega t_r} \right) e^{j\omega t}, \quad (32)$$

где V — сумма приложенного и встроенного напряжений и $t_r \equiv W/v_s$ — время пролета носителей через обедненную область.

Из уравнения (32) следует, что плотность тока короткого замыкания ($V = 0$) равна

$$J_{sc} = \frac{q\Phi_1 (1 - e^{j\omega t_r})}{j\omega t_r} e^{j\omega t}. \quad (33)$$

Эффекты, проявляющиеся на высоких частотах и связанные с влиянием времени пролета, иллюстрируются на рис. 11, б, на котором приведены зависимости амплитуды и фазового угла нормализованного тока от нормализованной частоты модуляции. Отметим, что амплитуда переменной составляющей фототока резко уменьшается, когда ωt_r становится больше 1. При $\omega t_r = 2,4$ уменьшение амплитуды в $\sqrt{2}$ раз сопровождается изменением фазового угла на величину $2\pi/5$. Быстродействие фотодетектора, таким образом, ограничивается временем пролета носителей через обедненный слой. Разумный компромисс между быстродействием и квантовой эффективностью достигается при значениях ширины области поглощения от $1/\alpha$ до $2/\alpha$.

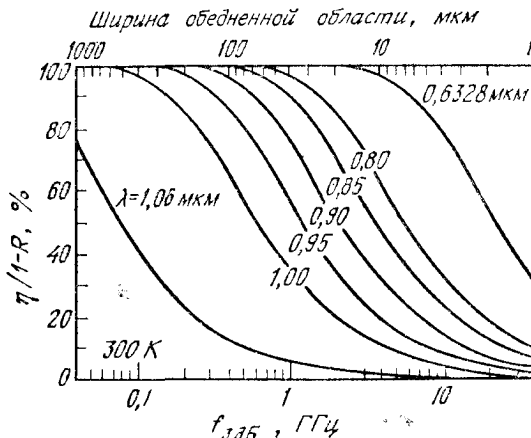


Рис. 12. Зависимость квантовой эффективности кремниевого $p-i-n$ -фотодиода от ширины обедненной области и граничной частоты (на уровне 3 дБ), определяемой временем пролета для различных длин волн. Скорость насыщения равна 10^7 см/с.

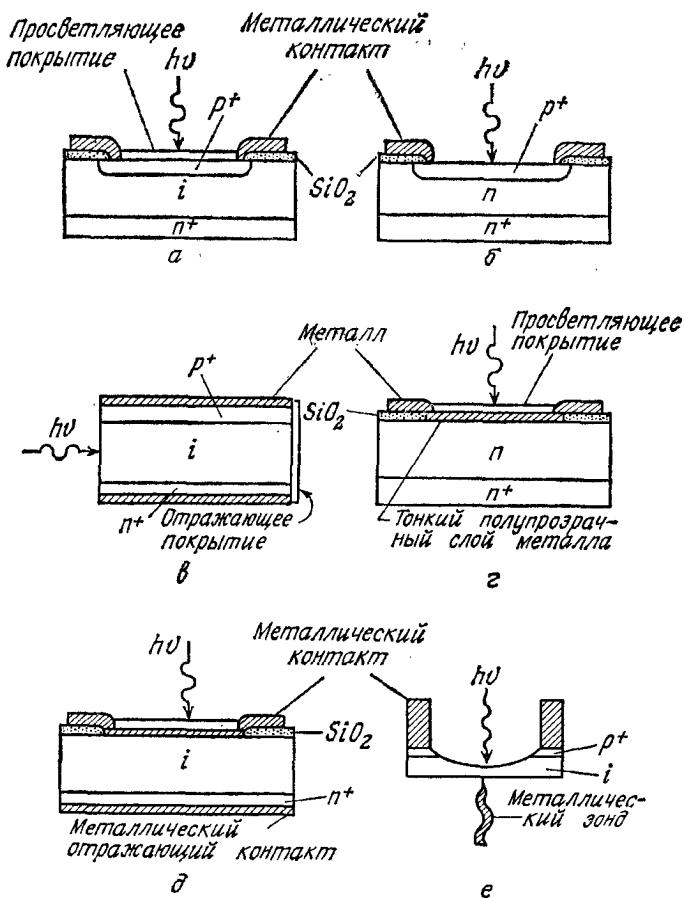


Рис. 13. Конструкции некоторых высокочастотных фотодиодов [2].

a — $p-i-n$ -диод; *б* — диод с $p-n$ -переходом; *в* — $p-i-n$ -диод в освещенном параллельно переходу; *г* — диод со структурой металл — полупроводник; *д* — диод со структурой металл — $i-n$; *е* — полупроводниковый точно-контактный диод.

В $p-i-n$ -диоде ширина i -области выбирается равной $1/\alpha$. Время пролета носителей равно времени их дрейфа через i -область. Из уравнения (33) следует, что ослабление фототока, соответствующее 3 дБ, происходит на частоте

$$f_{3\text{ дБ}} = \frac{2,4}{2\pi l} \approx \frac{0,4v_s}{W} \approx 0,4\alpha v_s. \quad (34)$$

На рис. 12 приведена зависимость внутренней квантовой эффективности, т. е. $\eta/(1-R)$, кремниевого $p-i-n$ -фотодиода от $f_{3\text{ дБ}}$ и ширины обедненной области, вычисленная из уравнения (34)

с учетом данных, представленных на рис. 5. Кривые иллюстрируют связь между быстродействием (частота $f_{з дБ}$) и квантовой эффективностью при различных длинах волн.

Несколько конструкций высокочастотных фотодиодов показаны на рис. 13 [2]. $p-i-n$ -Фотодиод (рис. 13, а) снабжен просветляющим покрытием для увеличения квантовой эффективности. Толщина области собственной проводимости (на практике используется высокоомный слой n -типа, т. е. v -область, или высокоомный слой p -типа, т. е. π -область) выбирается оптимальной для заданной длины волны оптического сигнала и частоты модуляции. В $p-n$ -фотодиоде n -слой имеет более высокую концентрацию легирующей примеси и в результате этого обеднен не полностью (рис. 13, б). Вследствие большего вклада диффузионного тока $p-n$ -фотодиоды обычно обладают меньшим быстродействием, чем $p-i-n$ -фотодиоды. При длинах волн вблизи красной границы чувствительности требуемая толщина поглощения становится очень большой (при $\alpha = 10 \text{ см}^{-1}$ или $1/\alpha = 1000 \text{ мкм}$). В этом случае согласование между квантовой эффективностью и быстродействием может быть достигнуто, если свет входит в полупроводник через грань параллельно переходу (рис. 13, в). Можно также выбрать такие углы падения света, которые обеспечивают его многократное отражение внутри прибора и тем самым приводят к увеличению эффективности толщины поглощения, сохраняя при этом малую длину дрейфа носителей [16, 17]. Три других прибора, представляющие собой фотодиоды с барьером Шоттки, рассмотрены в следующем разделе.

13.3.3. Фотодиод с барьером Шоттки

Фотодиоды с барьером Шоттки могут быть использованы в качестве высокоэффективных фотодетекторов [18, 19]. Зонная энергетическая диаграмма и протекание тока в диоде с барьером Шоттки рассмотрены в гл. 5, а типичная структура изображена на рис. 13, г. С целью исключения больших потерь на отражение и поглощение при освещении диода через металлический контакт следует использовать очень тонкие слои металла ($\sim 100 \text{ \AA}$) и применять просветляющие покрытия. Диод может работать в различных режимах в зависимости от энергии фотонов и прикладываемого смещения:

1. При $E_g \gg h\nu \gg q\phi_{Вн}$ и $V < V_B$ (рис. 14, а) (где V_B — напряжение лавинного пробоя) электроны, возбужденные в металле под действием излучения и преодолевшие барьер, будут входить в полупроводник. Этот эффект часто используется для определения высоты барьера Шоттки и изучения процессов переноса горячих электронов в металлических пленках [20].

2. При $h\nu > E_g$ и $V < V_B$ (рис. 14, б) излучение вызывает генерацию электронно-дырочных пар, и основные характеристики

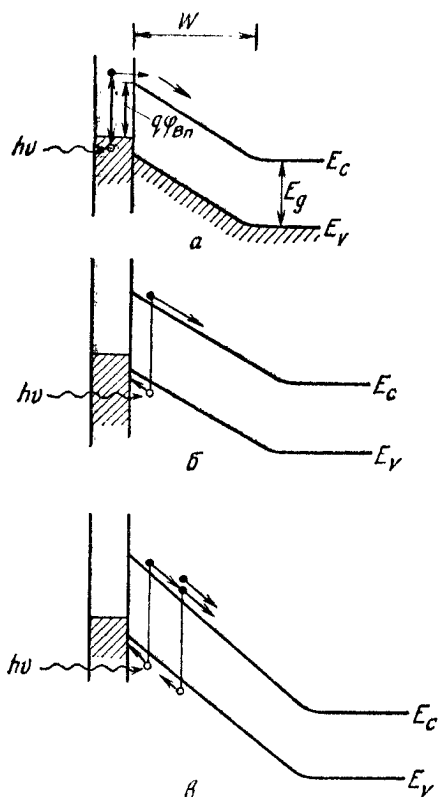


Рис. 14. Фотоэмиссия электронов из металла в полупроводник (а) ($E_g > h\nu > q\Phi_{Bn}$), собственное возбуждение (зона — зона) электронно-дырочной пары (б) ($h\nu > E_g$), а также генерация электронно-дырочной пары и лавинное умножение в условиях большого обратного смещения (в) ($h\nu > E_g$ в $V \approx V_B$).

диода очень похожи на характеристики $p-i-n$ -фотодиода. Квантовая эффективность определяется выражением, идентичным уравнению (29).

3. При $h\nu > E_g$ и $V \approx V_B$ (большое обратное смещение) (рис. 14, в) диод может работать как лавинный фотодиод (последний рассматривается в разд. 13.4).

Фотодиоды с барьером Шоттки особенно эффективны в видимой и ультрафиолетовой областях спектра. Для этих спектральных областей коэффициент поглощения α в наиболее распространенных полупроводниках очень велик ($\sim 10^5 \text{ см}^{-1}$ и более) и эффективная глубина поглощения мала ($1/\alpha \approx 0,1 \text{ мкм}$ и менее). Поэтому оказывается возможным подобрать металл и просветляющее покрытие так, чтобы большая часть падающего излучения поглощалась вблизи поверхности полупроводника.

Структура, показанная на рис. 15, а, имеет тщательно подобранное просветляющее покрытие из ZnS толщиной 500 Å и коэффициентом преломления 2,30. Пленка из золота имеет комплексный

коэффициент преломления $\bar{n} = 0,28 + j0,301$ на длине волны $\lambda = 6328 \text{ \AA}$ (излучение He—Ne-лазера), в то время как на этой длине волны кремниевая подложка имеет коэффициент преломления, равный $3,75 + j0,018$. На рис. 15, б представлены зависимости пропускания, коэффициента отражения и потерь в пленке Au от ее толщины. Видно, что при толщине пленки около 100 \AA более 95 % падающего излучения проникает в кремниевую подложку. Экспериментальные результаты, полученные на диоде, подобном тому, который показан на рис. 13, г, показывают, что полная квантовая эффективность диода на длине волны $\lambda = 6328 \text{ \AA}$ равна 70 % и время нарастания фототоклика достигает 0,1 нс ($1 \text{ нс} = 10^{-9} \text{ с}$) [17].

Квантовая эффективность и коэффициент пропускания Au—n—Si-фотодиода [21] (толщина пленки Au равна 107 \AA) для энергии фотона в диапазоне 1—6 эВ представлены на рис. 16. В видимой области спектра (вплоть до энергии $h\nu = 3,1 \text{ эВ}$) значения η и коэффициента пропускания практически одинаковы. В ультрафиолетовой области они различаются менее чем на 25 %. Найдено, что квантовая эффективность для всех значений энергии фотона не зависит от концентрации легирования при ее изменении от $2 \cdot 10^{13}$ до $2 \cdot 10^{17} \text{ см}^{-3}$. Это означает, что потери, обусловленные

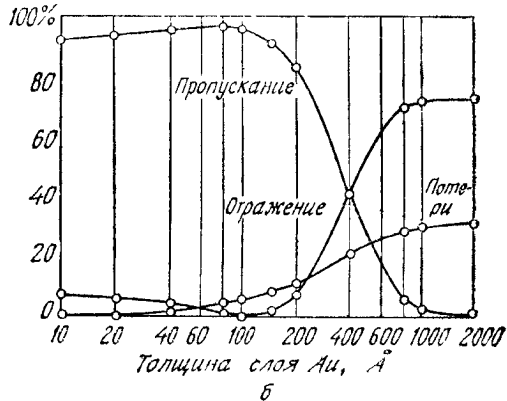
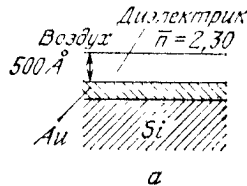


Рис. 15. Структура с просветляющим покрытием из пленки ZnS толщиной 500 \AA (а) и зависимость пропускания, отражения и потерь излучения ($\lambda = 0,6328 \text{ мкм}$) в пленке Au от ее толщины (б) [19].

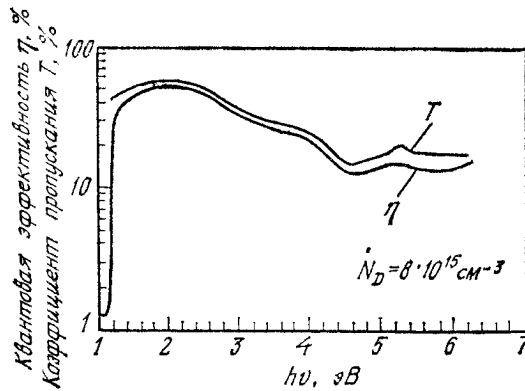


Рис. 16. Квантовая эффективность и коэффициент пропускания в Au — n — Si фотодиоде [21].

диффузией электронов в обедненной области в направлении, противоположном их дрейфу в электрическом поле, или рекомбинацией на ловушках на границе металл—полупроводник, пренебрежимо малы [22].

Используя слабелегированный *i*-слой, можно создать фотодиод со структурой металл—*i*—*n*, аналогичный *p*—*i*—*n*-диоду (рис. 13, *д*). Специальный диод, изображенный на рис. 13, *е*, представляет собой точечно-контактный фотодиод [23]. Активный объем в таких диодах очень мал, и соответственно чрезвычайно малы время пролета и емкость. Поэтому они пригодны для детектирования излучения при высоких частотах модуляции. Применение этих диодов, однако, ограничено лишь теми ситуациями, в которых излучение может быть сфокусировано в пятно диаметром несколько микрон.

13.3.4. Фотодиод с гетеропереходом

Фотодиод можно создать на основе гетероперехода, в котором переход сформирован между двумя полупроводниками с различной шириной запрещенной зоны (гл. 2). Одно из преимуществ такого фотодиода состоит в том, что его квантовая эффективность не зависит критическим образом от расстояния, на котором расположен переход от поверхности, поскольку материал с более широкой запрещенной зоной можно использовать в качестве «окна», пропускающего оптическое излучение. Кроме того, гетеропереход дает уникальную возможность выбора такого сочетания материалов, которое обеспечивает оптимальные для заданной длины волны оптического сигнала значения квантовой эффективности и быстродействия.

Для того чтобы получить гетеропереход с низкими токами утечки, необходимо обеспечить хорошее согласование постоянных решеток обоих полупроводников. Тройные полупроводниковые соединения $Al_xGa_{1-x}As$ (с прямой зоной при $x < 0,4$), выращенные в виде эпитаксиальных слоев на подложке GaAs, формируют гетеропереходы с совершенным согласованием решеток [24] (с постоянной решетки, равной $5,653 \text{ \AA}$). Эти гетеропереходы играют важную роль в оптоэлектронных приборах, работающих в спектральном диапазоне $0,65-0,85 \text{ мкм}$. Фотодетектор с двойным гетеропереходом ($n-Al_{0,24}Ga_{0,76}As/p-GaAs/p-Al_{0,24}Ga_{0,76}As$), снабженный просветляющим покрытием и работающий на длине волны $\lambda = 0,8075 \text{ мкм}$, имеет квантовую эффективность, равную 92% [25]. Каскадная структура [26], состоящая из пары $p-i-n$ -диодов с двойным гетеропереходом из $AlGaAs$, соединенных последовательно с туннельным переходом, также характеризуется высокой эффективностью на длине волны $\lambda = 0,815 \text{ мкм}$ и имеет напряжение холостого хода, равное $1,78 \text{ В}$.

Для больших длин волн ($1-1,6 \text{ мкм}$) могут быть использованы тройные соединения, такие, как $Ga_{0,47}In_{0,53}As$ ($E_g = 0,73 \text{ эВ}$) [59] и четверные соединения, такие, как $Ga_xIn_{1-x}As_yP_{1-y}$ (например, $Ga_{0,27}In_{0,73}As_{0,63}P_{0,37}$ с $E_g = 0,95 \text{ эВ}$) [27], которые имеют совершенное согласование решеток с подложкой InP (постоянная

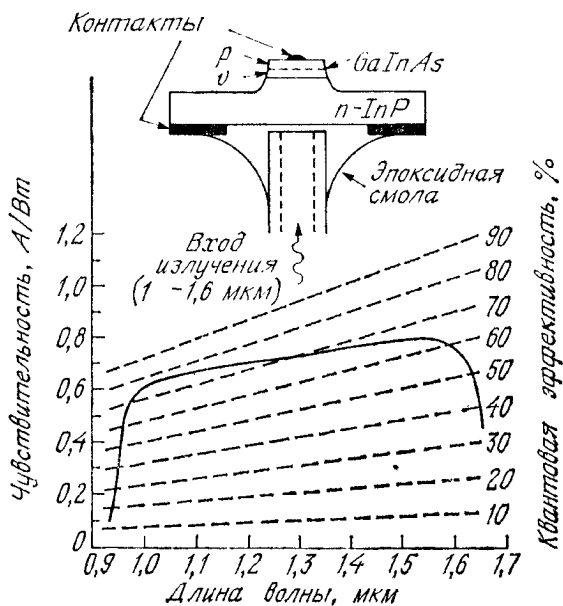


Рис. 17. Зависимость чувствительности и квантовой эффективности $GaInAs-p-i-n$ фотодиода от длины волны [28].

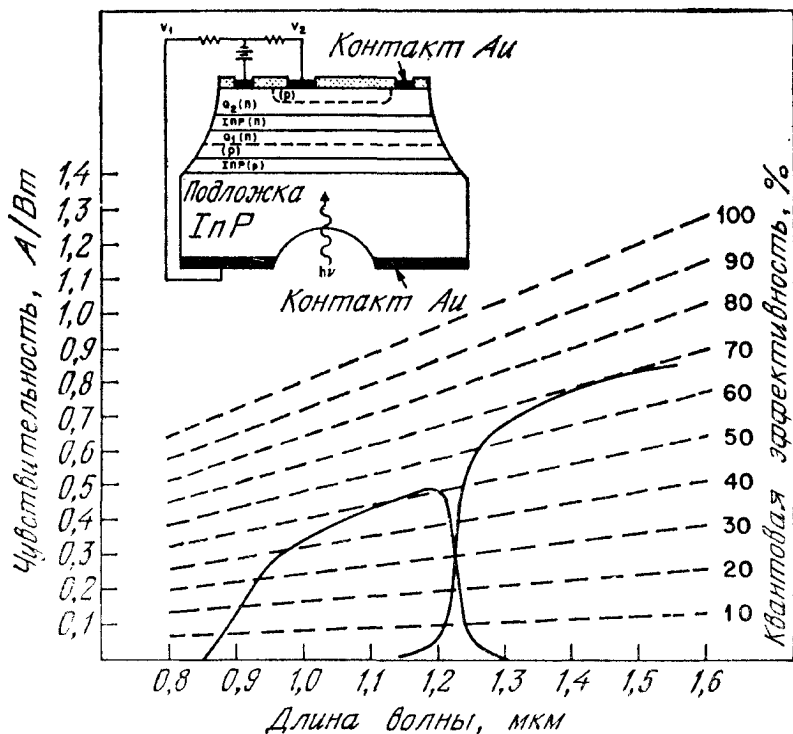


Рис. 18. Зависимость чувствительности и квантовой эффективности спектрального фотодиода от длины волны [56].

решетки $5,8686 \text{ \AA}$, см. рис. 24 гл. 12). На рис. 17 (вставка) изображен фотодиод, освещаемый с обратной стороны, на основе мезоструктуры $p\text{-Ga}_{0.47}\text{In}_{0.53}\text{As}/v\text{-Ga}_{0.47}\text{In}_{0.53}\text{As}/n^+\text{-InP}$. Фотодиод имеет чувствительность более $0,6 \text{ A/Wt}$, которая практически постоянна во всем спектральном диапазоне от $1,0$ до $1,5 \text{ мкм}$ (рис. 17) [28]. Наибольшее значение внешней квантовой эффективности, равное 70% , достигается при длине волны $1,3 \text{ мкм}$; типичные значения квантовой эффективности составляют $55\text{--}60 \%$. Этот прибор, по-видимому, должен иметь более высокие характеристики, чем Ge-фотодиод. Поскольку он обладает прямой запрещенной зоной и, соответственно, большим коэффициентом поглощения вблизи края собственного поглощения, то за счет уменьшения толщины обедненной области можно достичь более высокого быстродействия [29].

Многослойная структура фотодиода с гетеропереходами, в которой одновременно осуществляется детектирование излучения и его разложение на два спектральных интервала, показана на рис. 18

(вставка) [56]. Спектральное разложение обеспечивают два слоя четверного соединения $Ga_xIn_{1-x}As_yP_{1-y}$. Решетки этих слоев согласованы с решеткой InP ; они имеют различный кристаллический состав и, следовательно, характеризуются различными значениями ширины запрещенной зоны. Слой Q_1 имеет более широкую запрещенную зону (1,0 эВ), чем слой Q_2 (0,74 эВ), а подложка прозрачна для излучения падающего с обратной стороны структуры. Перекрытие зон фоточувствительности является следствием оптических перекрестных помех между двумя каналами (рис. 18). Необходимо выбирать такие области применения прибора, в которых спектральный интервал, соответствующий этому перекрытию, не используется. Из графика на рис. 18 следует, например, что на коротковолновой границе интервала перекрытия перекрестные помехи составляют -19 дБ при $\lambda = 1,2$ мкм и уменьшаются до -30 дБ при $\lambda = 1,15$ мкм, а на длинноволновой границе они равны -43 дБ при $\lambda = 1,3$ мкм.

13.4. ЛАВИННЫЙ ФОТОДИОД

Лавинные фотодиоды работают при высоких обратных смещениях, вызывающих лавинное умножение [30]. В результате умножения возникает внутреннее усиление тока. Произведение коэффициента усиления на ширину полосы лавинного фотодиода может достигать более 100 Гц, за счет чего прибор обеспечивает детектирование излучения в микроволновом диапазоне частот модуляции [2, 31]. Квантовая эффективность и быстродействие лавинных фотодиодов определяются так же, как и для обычных фотодиодов. Однако их шумовые свойства и эффект лавинного умножения требуют особого рассмотрения.

13.4.1. Отношение сигнал/шум

Процесс детектирования излучения в лавинном фотодиоде и его эквивалентная схема представлены на рис. 19 [5]. Усиление тока приводит к умножению сигнального, фонового и темнового токов. Среднеквадратичное значение фототока определяется выражением (15), в котором учитывается наличие коэффициента умножения (или коэффициента лавинного умножения) M :

$$i_p = q\eta n P_{opt} M / \sqrt{2} h\nu. \quad (35)$$

Другие элементы эквивалентной схемы на рис. 19, б полностью совпадают с элементами эквивалентной схемы $p-i-n$ -фотодиода. Среднеквадратичное значение дробового шума после умножения определяется выражением

$$\langle i_s^2 \rangle = 2q(I_p + I_B + I_D) \langle M^2 \rangle B, \quad (36)$$

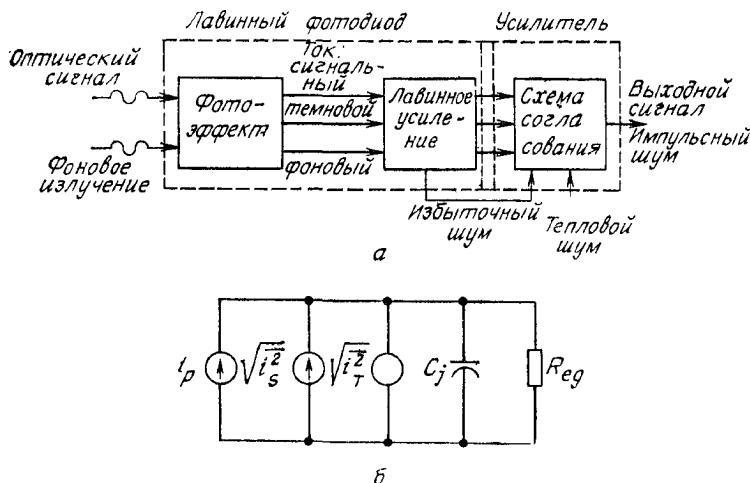


Рис. 19. Процесс детектирования излучения в лавинном фотодиоде (а) и эквивалентная схема лавинного фотодиода (б) [5].

или

$$\langle i_s^2 \rangle = 2q(I_p + I_B + I_D) M^2 F(M) B, \quad (37)$$

где $\langle M^2 \rangle$ — среднеквадратичное значение внутреннего коэффициента усиления и $F(M) = \langle M^2 \rangle / M^2$ — шум-фактор, который представляет собой меру увеличения дробового шума по отношению к идеальному нешумящему усилителю. Тепловой шум, так же как и для $p-i-n$ -фотодиода, определяется уравнением (17).

Отношение сигнал/шум лавинного фотодиода при 100 %-ной модуляции излучения со средней мощностью P_{opt} имеет вид

$$\begin{aligned} S/N &= \frac{1/2 (q\eta P_{opt}/h\nu)^2 M^2}{2q(I_p + I_B + I_D) F(M) M^2 B + 4kTB/R_{eq}} = \\ &= \frac{1/2 (q\eta P_{opt}/h\nu)^2}{2q(I_p + I_B + I_D) F(M) B + 4kTB/R_{eq} M^2}. \end{aligned} \quad (38)$$

Из этого уравнения видно, что лавинное умножение, обеспечивающее уменьшение вклада второго члена в знаменателе, может привести к увеличению отношения сигнал/шум. Ниже показано, что шум-фактор $F(M)$ всегда больше или равен единице и монотонно увеличивается с ростом коэффициента умножения, за исключением тех случаев, когда процесс умножения не вносит дополнительных шумов. Таким образом, существует оптимальное значение коэффициента умножения M , которое обеспечивает максимальное отношение сигнал/шум при заданной оптической мощности. Такое оптимальное умножение имеет место, когда первый и второй члены знаменателя примерно равны.

Используя уравнение (38), можно найти выражение для минимальной оптической мощности P_{opt} , которая необходима для

получения заданного отношения S/N при наличии лавинного умножения. Это выражение имеет вид

$$(P_{\text{opt}})_{\text{мин}} = \frac{2h\nu}{\eta} \left(\frac{S}{N} \right) \left\{ 1 + \left[1 + \frac{I_{eq}}{qBF(M)^2(S/N)} \right]^{1/2} \right\}, \quad (39)$$

где

$$I_{eq} \equiv (I_B + I_D) F(M) + 2kT/qR_{eq}M^2. \quad (40)$$

В условиях, когда $I_{eq}/qBF^2(S/N)$ пренебрежимо мало, превышение минимально обнаружимой мощности над собственным квантовым шумом сигнала при бесшумовом умножении ($F = 2$) составляет 3 дБ. При детектировании сигналов высокой частоты и с широкой полосой пропускания минимально обнаружимая мощность ограничивается тепловым шумом нагрузочного сопротивления и шум-фактором последующей стадии усиления. В этих условиях отношение $I_{eq}/qBF(M)^2(S/N)$ больше единицы, и мощность, эквивалентная шуму, определяется выражением

$$\text{NEP} = \sqrt{2} (h\nu/\eta) [I_{eq}/qF(M)^2]^{1/2}. \quad (41)$$

Поскольку лавинное усиление позволяет значительно уменьшить NEP, лавинные фотодиоды имеют существенное преимущество перед обычными фотодиодами.

13.4.2. Коэффициент умножения

Коэффициент лавинного усиления, называемый также коэффициентом умножения, был введен в гл. 2. Выражение для коэффициента умножения электронов при низких частотах имеет вид

$$M = \left\{ 1 - \int_0^W \alpha_n \exp \left[- \int_0^x (\alpha_n - \alpha_p) dx' \right] dx \right\}^{-1}, \quad (42)$$

где W — ширина обедненного слоя, α_n и α_p — коэффициенты ионизации электронов и дырок соответственно. При независимости от координаты коэффициентах ионизации, что имеет место, например, в $p-i-n$ -диоде, коэффициент умножения электронов, инжектированных в область сильного поля, в точке $x = 0$ равен

$$M = \frac{(1 - \alpha_p/\alpha_n) \exp [\alpha_n W (1 - \alpha_p/\alpha_n)]}{1 - (\alpha_p/\alpha_n) \exp [\alpha_n W (1 - \alpha_p/\alpha_n)]}. \quad (43)$$

Если $\alpha_n = \alpha_p$, выражение для M принимает простую форму

$$M = 1/(1 - \alpha_n W). \quad (44)$$

Если $\alpha_n W = 1$, возникает пробой полупроводника. Восстанавливающийся лавинный процесс создает большое число носителей в области сильного поля в течение длительного времени после прохождения через эту область первичных электронов. Чем выше коэффициент умножения, тем дольше продолжается лавинный

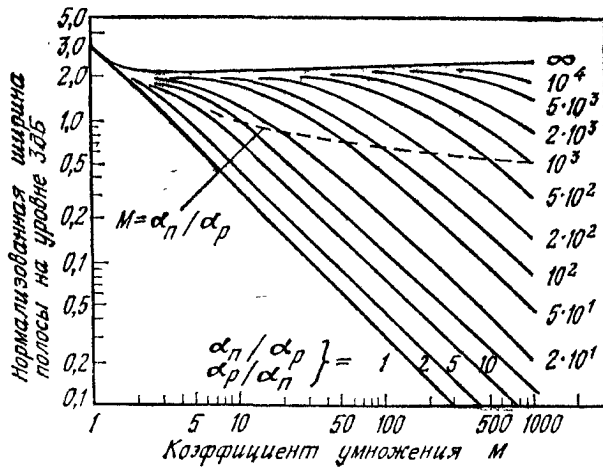


Рис. 20. Теоретическая зависимость граничной частоты (на уровне 3 дБ), умноженной на $2\pi\tau_{av}$ (τ_{av} — среднее время пролета носителей), от низкочастотного коэффициента умножения при различных значениях α_p/α_n (инжекция электронов) и α_n/α_p (инжекция дырок) для лавинного фотодиода [32].

процесс. Эта взаимосвязь устанавливается произведением усиления на ширину полосы. Когда коэффициенты ионизации равны, а $M \rightarrow \infty$, выражение для произведения коэффициента усиления тока на ширину полосы имеет вид

$$\text{Усиление} \times \text{Ширина полосы} = \frac{3}{2\pi\tau_{av}}, \quad (45)$$

где $\tau_{av} = (t_n + t_p)/2$; t_n — время пролета электронов, равное W/v_n (v_n — скорость насыщения электронов), и t_p — время пролета дырок, которое можно записать аналогичным образом.

На рис. 20 приведены значения ширины полосы [32], рассчитанной для идеализированного лавинного $p-i-n$ -фотодиода с постоянной величиной электрического поля в области лавинного умножения. Расчет проведен для ширины полосы B , соответствующей ослаблению 3 дБ и нормализованной на $2\pi\tau_{av}$, в зависимости от низкочастотного коэффициента умножения, причем в качестве параметра взято отношение коэффициентов лавинного умножения. Выше кривой, соответствующей $M = \alpha_n/\alpha_p$, ширина полосы в значительной степени определяется временем пролета носителей и практически не зависит от усиления. Ниже этой кривой наблюдается почти прямолинейная зависимость, соответствующая постоянным значениям произведения усиления на ширину полосы:

$$M(\omega)\omega \approx \frac{1}{(W/v_n)(\alpha_p/\alpha_n)}. \quad (46)$$

Таким образом, для получения больших значений произведения усиления на ширину полосы необходимо, чтобы α_p/α_n было велико, а W — мало.

В реальных приборах максимально достижимое усиление по постоянному току при высоких интенсивностях излучения ограничено последовательным сопротивлением и влиянием пространственного заряда. Эти факторы можно свести к одному эффективному последовательному сопротивлению R . Коэффициент умножения для фотогенерированных носителей может быть описан эмпирическим соотношением [33]:

$$M_{ph} = \frac{I - I_{MD}}{I_p - I_D} = \frac{1}{\left[1 - \left(\frac{V_R - IR}{V_B}\right)^n\right]}, \quad (47)$$

где I — общий усиленный ток, I_p — общий первичный ток (до усиления), I_D и I_{MD} — первичный и усиленный темновые токи соответственно, V_R — напряжение обратного смещения, V_B — напряжение пробоя и n — показатель, зависящий от полупроводникового материала, профиля легирования и длины волны излучения. При высокой интенсивности излучения ($I_p \gg I_D$) и $IR \ll V_B$ максимальная величина коэффициента умножения фототока определяется выражением

$$(M_{ph})_{\max} \approx \frac{I}{I_p} = \frac{1}{1 - \left(\frac{V_R - IR}{V_B}\right)^n} \Bigg|_{V_R \rightarrow V_B} \approx \frac{1}{nIR/V_B}, \quad (48)$$

или

$$(M_{ph})_{\max} = \sqrt{V_B/nI_pR}. \quad (49)$$

При малых фототоках максимальное умножение ограничено темновым током и в уравнении (49) необходимо вместо I_p подставить I_D . Таким образом, весьма важно, чтобы темновой ток был как можно меньше и не ограничивал ни $(M_{ph})_{\max}$, ни минимально обнаружимую мощность (уравнение (39)).

13.4.3. Шум лавинного умножения

Лавинный процесс является по природе статистическим процессом, так как электронно-дырочные пары, генерируемые в данной точке обедненной области, претерпевают неодинаковое умножение. Поскольку коэффициент умножения флуктуирует, среднеквадратичное значение усиления превышает квадрат его среднего значения. Избыточный шум характеризуется шум-фактором $F(M) = \langle M^2 \rangle / M^2$, который зависит от отношения коэффициентов ионизации α_p/α_n и низкочастотного коэффициента умножения M . Если $\alpha_n = \alpha_p$, на каждый фотонотель, возбужденный падающим излучением, в области умножения приходится три носителя (пер-

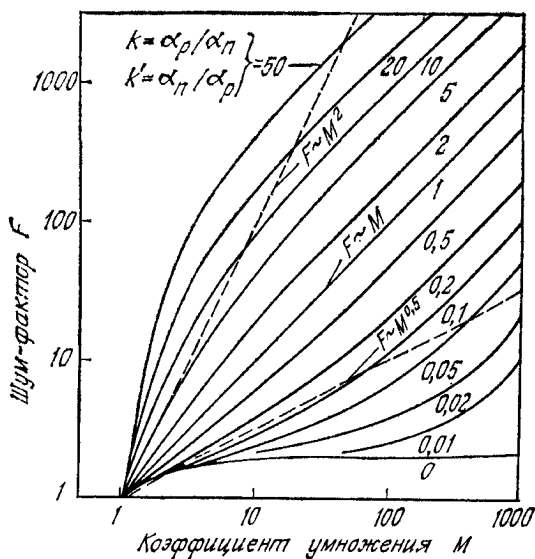


Рис. 21. Теоретическая зависимость шум-фактора от коэффициента умножения для различных значений α_n/α_p и α_p/α_n [34].

вичный носитель и вторичные электрон и дырка). Поэтому флуктуации, изменяющей число носителей на один, соответствует существенное относительное изменение их числа и шум-фактор оказывается большим. Если же один из коэффициентов ударной ионизации пренебрежимо мал (например, $\alpha_p \rightarrow 0$), то на каждый фотоноситель в области умножения приходится около M носителей. В этом случае флуктуация, соответствующая изменению этого числа на один носитель, вызовет лишь незначительное относительное возмущение. Таким образом, следует ожидать, что при существенно различных α_n и α_p значение шум-фактора будет мало.

Если инжектируются только электроны, то шум-фактор можно записать в виде [34]

$$F = kM + (2 - 1/M)(1 - k), \quad (50)$$

где $k \equiv \alpha_p/\alpha_n$ предполагается постоянным во всей области лавинного умножения. Это выражение применимо и для случая инжекции дырок, если k заменить на $k' \equiv \alpha_n/\alpha_p$. Для двух специальных случаев, а именно $\alpha_p = \alpha_n$ (т. е. $k = 1$) и $\alpha_p \rightarrow 0$ (т. е. $k = 0$), из уравнения (50) получаем значения $F = M$ и $F = 2$ (при больших M) соответственно. Зависимость шум-фактора от коэффициента умножения при различных отношениях коэффициентов ионизации

приведена на рис. 21. Видно, что снижение уровня дополнительных шумов достигается при малых значениях k и k' .

На рис. 22 представлены экспериментальные результаты [35], полученные для лавинного Si-фотодиода на частоте 600 кГц и при токе первичной инжекции, равном 0,1 мкА. Светлыми кружками обозначен шум первичного дырочного фототока, возбуждаемого коротковолновым излучением (см. вставку). Темными кружками представлен шум для случая, когда имеет место первичная инжекция электронов. Электронная инжекция характеризуется существенно меньшим шум-фактором, чем дырочная, поскольку в кремнии α_n значительно больше α_p .

Результаты, представленные на рис. 21, могут быть использованы применительно к лавинным $p-i-n$ -фотодиодам и фотодиодам с трехслойной базой, которые имеют постоянное поле в области лавинного умножения. Для обычных лавинных фотодиодов с изменяющимся электрическим полем в области лавинного умножения коэффициенты ионизации в уравнении (50) должны

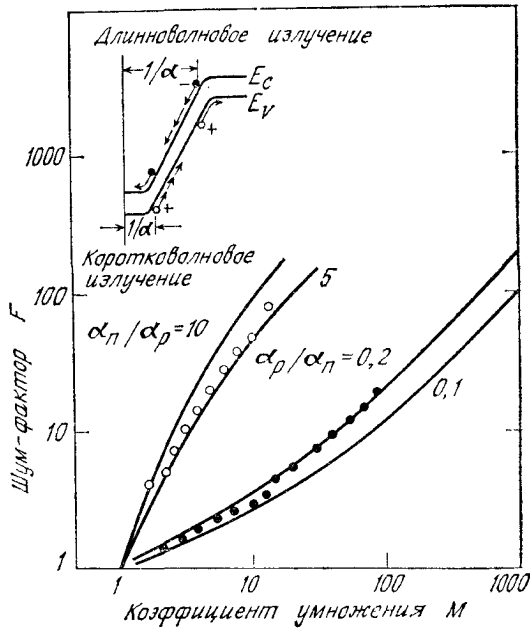


Рис. 22. Результаты экспериментального исследования шум-фактора кремниевого лавинного фотодиода с током первичной инжекции, равным 0,1 мкА. На вставке показана зонная диаграмма лавинного фотодиода в условиях электронной и дырочной инжекции, зависящей от длины волны падающего излучения [35].

— теоретические данные; ● для инжекции электронов; ○ для инжекции дырок.

быть усложнены [36]; k необходимо заменить на k_{eff} , а k' — на k'_{eff} , где

$$k_{\text{eff}} = \frac{\int_0^W \alpha_p(x) M^2(x) dx}{\int_0^W \alpha_n(x) M^2(x) dx}, \quad (51)$$

$$k'_{\text{eff}} = k_{\text{eff}} \left[\frac{\int_0^W \alpha_p(x) M(x) dx}{\int_0^W \alpha_n(x) M(x) dx} \right]^{-2}. \quad (52)$$

Избыточный шум возникает в том случае, когда излучение поглощается как в n -, так и p -областях, прилегающих к переходам, вызывая инжекцию в область лавинного умножения как электронов, так и дырок. Например, для $k_{\text{eff}} = 0,005$ и $M = 10$ значение шум-фактора возрастает примерно от 2 при чисто электронной инжекции до 20 при электронной инжекции, составляющей 10 % [37]. Таким образом, для достижения в лавинных фотодиодах низкого уровня шума и широкой полосы частот необходимо, чтобы коэффициенты ионизации различались как можно более сильно, а лавинный процесс стимулировался носителями, обладающими более высоким коэффициентом ионизации.

13.4.4. Рабочие характеристики

В лавинных фотодиодах необходимо обеспечить пространственно однородное умножение по всей fotocувствительной площади [38]. Должно быть исключено образование микроплазм, представляющих собой локализованные участки, в которых пробивное напряжение меньше, чем во всем переходе. Количество микроплазм в активной области уменьшается при использовании материалов с низкой плотностью дислокаций и конструкций приборов с активной областью, не превышающей по своим размерам область, которая необходима для сбора падающего светового пучка (обычно диаметром от нескольких микрометров до 100 мкм). Избыточные токи утечки по периферии перехода, обусловленные краевым эффектом [39] или сгущением линий электрического поля, устраняются за счет применения охранного кольца или поверхностно-конических структур.

На рис. 23 показаны некоторые конструкции приборов [2]. Охранное кольцо в структуре на рис. 23, а должно иметь низкий градиент примеси на границе p — n -перехода и достаточно большой радиус кривизны, чтобы центральная область резкого p — n^+ -перехода пробивалась раньше, чем область охранного кольца. Охранное кольцо можно использовать и в более сложной n^+ — p — n^+ -структуре (рис. 23, б). Оно может быть применено также для устранения высокого поля на периферии контакта в лавинном фотодиоде Шоттки (рис. 23, в). Мезаструктура (рис. 23, г) и коническая

структура (рис. 23, *д*) характеризуются низким поверхностным полем по всей площади перехода и однородным лавинным пробоем. Для детектирования излучения с длинами волн вблизи края собственного поглощения может быть использован лавинный фотодиод с боковым освещением, с помощью которого можно улучшить квантовую эффективность и отношение сигнал/шум (рис. 23, *е*).

Лавинные фотодиоды изготавливаются на основе различных полупроводников, в том числе Ge, Si, соединений типа $A^{III}B^V$ и других. Основными факторами при выборе того или иного полупроводника являются квантовая эффективность в заданном спектральном диапазоне, быстродействие и шумы. Мы рассмотрим несколько типичных примеров, иллюстрирующих характеристики приборов.

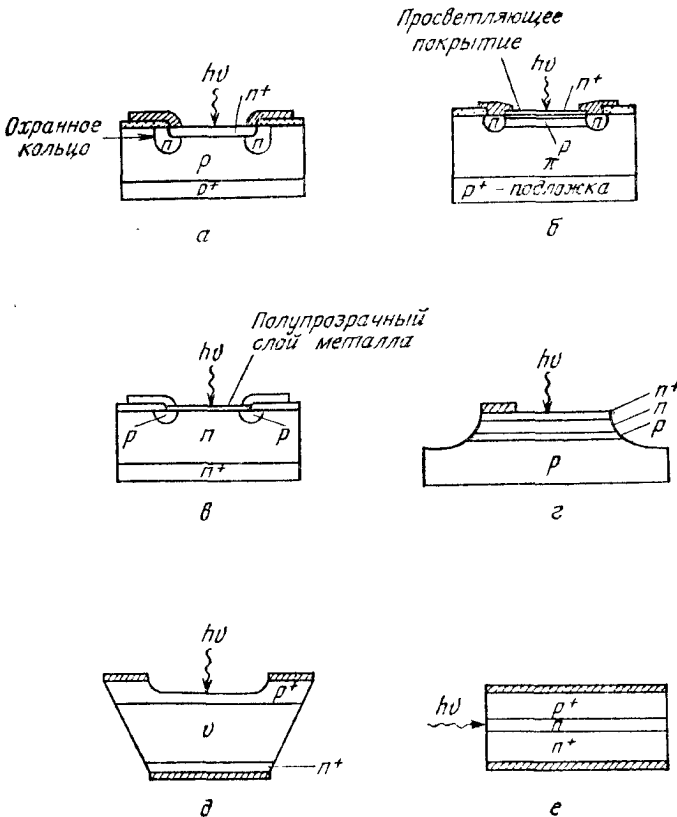


Рис. 23. Структуры некоторых лавинных фотодиодов [2].
 а — структура с охранным кольцом; б — $n^+ - p - n - p^+$ -структура с охранным кольцом; в — диод со структурой металл — полупроводник; г — мез-структура; е — $p^+ - n - n^+$ -структура с боковым освещением.

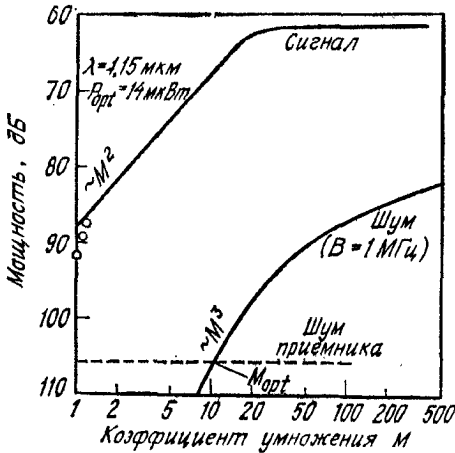


Рис. 24. Мощность сигнала и шума относительно уровня 1 мВт на выходе лавинного Ge-фотодиода при частоте 3 ГГц в полосе 1 МГц. На графике обозначена оптимальная рабочая точка M_{opt} , в которой достигается максимальное значение отношения сигнал/шум и чувствительности [2].

Германиевые лавинные фотодиоды обеспечивают высокую квантовую эффективность в спектральном диапазоне 1—1,6 мкм. Поскольку коэффициенты ионизации электронов и дырок в Ge сравнимы, то шум-фактор близок по величине к $F = M$ (уравнение (50)) и среднеквадратичный дробовой шумовой ток изменяется по закону M^3 (уравнение (37)). Структура, показанная на рис. 23, а, была применена для изготовления высокочастотного Ge-фотодиода [33, 40]. Удельное сопротивление p -слоя составляло 0,5 Ом·см, а его толщина 150 мкм. После диффузии сурьмы для создания охранного кольца глубиной 7,5 мкм проводилась диффузия мышьяка на глубину 0,4 мкм для формирования n^+ -слоя. Прибор имел квантовую эффективность около 80 % во всем спектральном диапазоне от 1 до 1,55 мкм (рис. 7). Плотность тока при комнатной температуре и напряжении смещения, равном $\frac{1}{2} V_B$ (где V_B — пробивное напряжение), составила $\sim 3 \cdot 10^{-4}$ А/см² [33]. Время нарастания импульса фотоотклика при воздействии АИГ-лазера ($\lambda = 1,06$ мкм) равнялось 100 пс.

На рис. 24 представлены результаты исследования [2] выходного сигнала и шума Ge-лавинного фотодиода при частоте модуляции 3 ГГц и первичном токе 5,3 мкА. Отметим, что при $M < 30$ мощность сигнала растет как M^2 , а мощность шума — как M^3 . Это хорошо согласуется с теоретическими предположениями. Отметим также, что наибольшее отношение сигнал/шум (~ 40 дБ) получено при $M \approx 10$, т. е. когда шумы диода примерно равны шумам приемника. При более высоких значениях M отношение S/N уменьшается, так как шум лавины растет быстрее, чем умноженный сигнал.

Кремниевые лавинные фотодиоды особенно эффективны в спектральном диапазоне 0,6—1,0 мкм, в котором для приборов с про-

светляющим покрытием достигнут квантовый выход, близкий к 100 %. Отношение коэффициентов ионизации электронов и дырок ($k = \alpha_p/\alpha_n$) в кремнии сильно зависит от электрического поля; оно изменяется примерно от 0,1 при поле $3 \cdot 10^5$ В/см до 0,5 при поле 6×10^5 В/см. Поэтому для получения минимальных шумов необходимо, чтобы электрическое поле лавинного пробоя было низким и чтобы лавинное умножение стимулировалось электронами.

Идеализированный профиль легирования в $p-\pi-p-\pi-n^+$ -структуре показан на рис. 25, а. Этот профиль аналогичен профилю трехслойной базы в структуре лавинно-пролетного диода (гл. 10). Структура $p-\pi-p-\pi-n^+$ называется структурой со «сквозным проколом» [41], поскольку электрическое поле в ней распространяется на все пространство от n^+ - π -области умножения до p^+ -слоя (рис. 25, б). В области дрейфа с низким электрическим полем носители могут переноситься со скоростью, достигающей их скорости насыщения (10^7 см/с при $\mathcal{E}_D \gg 10^4$ В/см²). Максимальное поле \mathcal{E}_m в области умножения можно регулировать за счет изменения толщины b . Условие пробоя можно записать в виде [42]

$$\alpha_n b = \frac{\ln k}{k-1} \quad (k \equiv \alpha_p/\alpha_n), \quad (53)$$

а пробивное напряжение определяется выражением

$$V_B \simeq \mathcal{E}_m b + \mathcal{E}_D (W - b). \quad (54)$$

С целью оптимизации характеристик прибора для заданной длины волны можно выбрать сначала значение W (из условия $W = 1/\alpha$), а затем независимо подобрать величину b . Как видно из рис. 25, в, большая часть излучения поглощается в π -области; при этом лавинный процесс стимулируется электронами, поступающими в область умножения (рис. 25, з). В приборах с $p^+-\pi-p-\pi-p^+$ -структурой можно получить высокие значения квантовой эффективности, быстродействия и отношения сигнал/шум.

Формирование узкой p^+ -области с целью создания профиля, показанного на рис. 25, а, может быть сопряжено с большими трудностями, поэтому на практике используют приборы со структурой $p^+-\pi-p-n^+$, в которой p -область занимает все пространство до n -подложки. Такой профиль аналогичен профилю двухслойной базы в структуре лавинно-пролетного диода. В кремниевом лавинном фотодиоде со «сквозным проколом» [37], в котором полная ширина обедненного слоя составляла 200 мкм, а π -область имела удельное сопротивление 5000 Ом·см, была получена квантовая эффективность ~ 90 % при $\lambda = 0,9$ мкм. При $M = 100$ шум-фактор принимает значение, равное всего лишь 4, а при $M = 1000$ он возрастает до 20, что находится в соответствии с низкой величиной $k_{\text{эф}} = 0,016$.

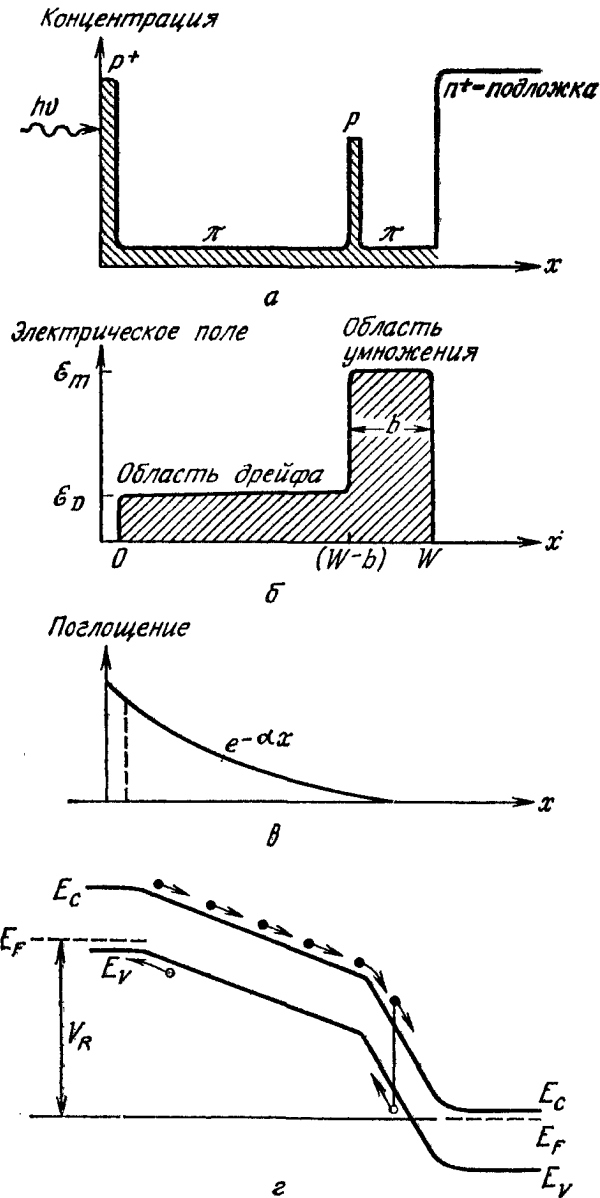


Рис. 25. Лавинный фотодиод со «сквозным проколом» на основе структуры с профилем трехслойной базы.

а — профиль легирования; б — распределение поля; в — поглощение падающего излучения; г — вольтная диаграмма, которая иллюстрирует процесс умножения, стимулируемый электронами.

Если сигнальный фототок существенно больше фонового и темнового токов, то доминирующий дробовой шум в выражении (38) описывается членом $2qI_p M^2 F B$, в котором F определяется выражением (50). Оптимальный коэффициент умножения M_{opt} находится из условия $d(S/N)/dM = 0$. Подставляя M_{opt} в выражение (38), получим максимальное соотношение сигнал/шум, соответствующее условию большого сигнального фототока [43]:

$$(S/N)_{max} \sim (\eta/\sqrt{k_{eff}})^{4/3}. \tag{55}$$

Таким образом, для получения предельно большого отношения S/N необходимо увеличивать $\eta/\sqrt{k_{eff}}$. На рис. 26 приведена зависимость $\eta/\sqrt{k_{eff}}$ от b при различных пробивных напряжениях в кремниевом лавинном фотодиоде с профилем трехслойной базы. Длина волны 0,85 мкм ($\alpha = 600 \text{ см}^{-1}$) соответствует излучению GaAs-лазера. Отметим, что при заданном V_B зависимость имеет максимум, который увеличивается с ростом V_B . Проведенная оценка справедлива и для других примесных профилей.

На рис. 27, а показана структура, позволяющая создавать приборы на кремниевых пластинах большого диаметра и обеспечи-

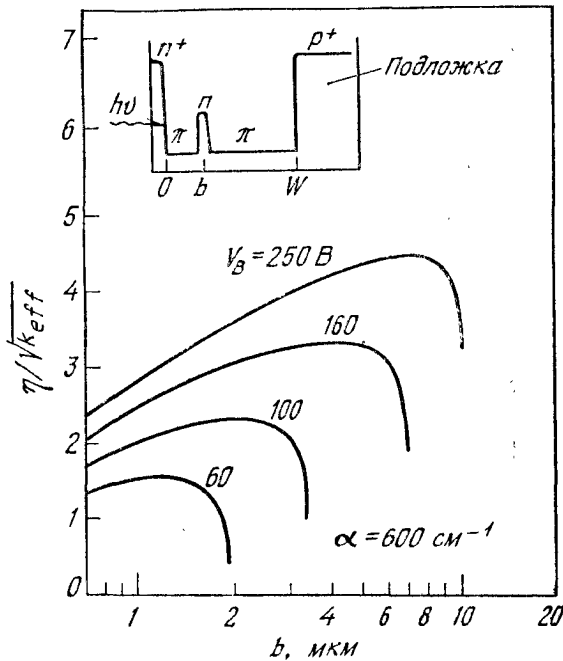


Рис. 26. Зависимость отношения $\eta/\sqrt{k_{eff}}$ от толщины области умножения b для кремниевого лавинного фотодиода с трехслойной базой [43].

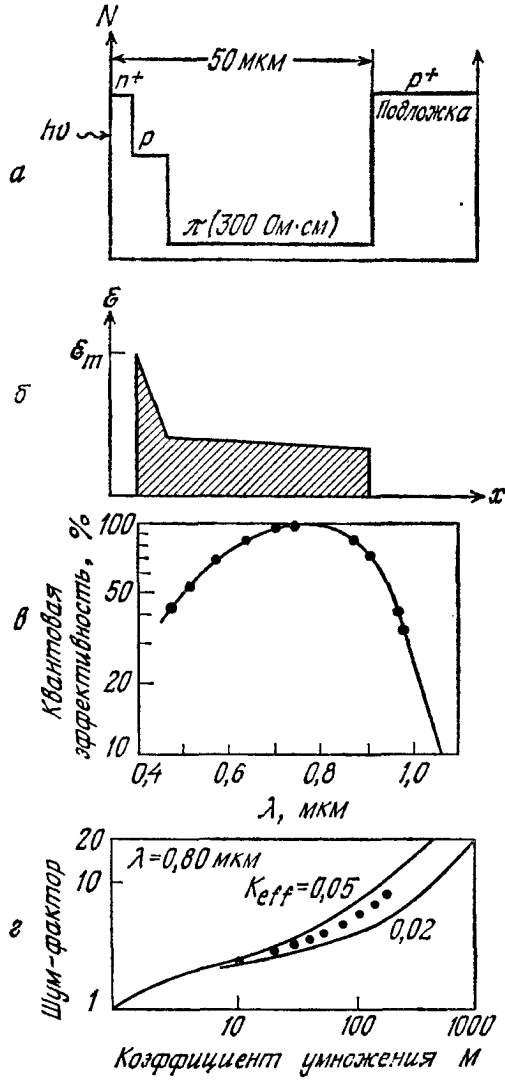


Рис. 27. Кремниевый лавинный фотодиод с двухслойной базой [44].
 а — профиль легирования; б — распределение поля; в — квантовая эффективность; г — шум-фактор [44].

нающая хорошее управление профилем легирования с помощью процессов ионной имплантации и диффузии [44]. Поперечный разрез структуры представлен на рис. 23, б. На рис. 27, б изображено распределение электрического поля, для которого характерно наличие узкой области лавинного умножения и протяженной области дрейфа. В приборе с просветляющим покрытием из пленок $\text{SiO}_2\text{—Si}_3\text{N}_4$ квантовая эффективность близка к 100 % на длине волны 0,8 мкм (рис. 27, в). Поскольку вклад дырок в стимулирование процесса умножения незначителен, то прибор обладает более высоким шум-фактором ($k_{\text{эф}} \approx 0,04$), чем структура, показанная на рис. 25.

Лавинные фотодиоды со структурой металл—полупроводник (с барьером Шоттки) эффективны в видимой и ультрафиолетовой областях спектра. Характеристики лавинных фотодиодов с барьером Шоттки подобны характеристикам фотодиодов с $p\text{—}n$ -переходом [45]. Фотодиод с барьером Шоттки был изготовлен на кремниевой подложке n -типа с удельным сопротивлением 0,5 Ом·см, с тонкой пленкой PtSi ($\sim 100 \text{ \AA}$) на поверхности и охранным кольцом [46]. Поскольку токи утечки по периферии перехода устранены благодаря применению охранного кольца, то в такой структуре может быть получен идеальный обратный ток насыщения с плотностью $2 \cdot 10^{-7} \text{ А/см}^2$. В фотодиодах с барьером Шоттки в режиме лавинного умножения может быть получено 35-кратное усиление амплитуды высокочастотного импульса фототока длительностью 0,8 нс (рис. 28). Разность между фототоком в отсутствие умножения и умноженным фототоком показывает, какое большое усиление по току можно получить для заданной длительности импульса (в рассматриваемом случае 0,8 нс) при определенном уровне освещения. Средняя мощность излучения лазера на длине волны 6328 \AA равна 0,8 мВт, а пиковая мощность в импульсе, соответствующая 0 дБ на рис. 28, примерно равна 7 мВт. Измерения шума лавинного умножения PtSi—Si-диодов показали, что шум умноженного фототока растет по закону $\sim M^3$ при возбуждении светом в видимой области спектра. При уменьшении длины волны преобладает первичный инжекционный фототок электронов и шумы уменьшаются, что находится в хорошем согласии с теорией.

Лавинные фотодиоды с барьером Шоттки, изготовленные на кремниевых подложках n -типа, особенно перспективны для применения в качестве высокочастотных фотодетекторов ультрафиолетового излучения. Ультрафиолетовое излучение, проходящее через тонкие металлические электроды, поглощается в приповерхностном слое кремния толщиной 100 \AA . Вследствие этого умножение носителей в основном стимулируется электронами, что обеспечивает низкий шум и большое произведение коэффициента усиления на ширину полосы. Лавинное умножение фототока повышает чувстви-

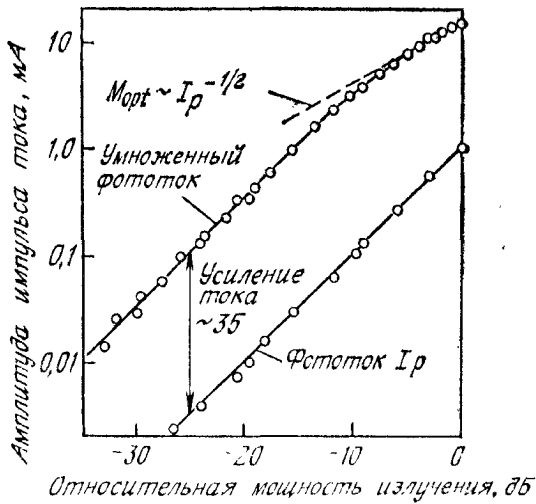


Рис. 28. Зависимость амплитуды импульсов фототока в лавинном фотодиоде с барьером Шоттки от интенсивности излучения при $\lambda = 0,6328$ мкм и длительности световых импульсов 0,8 нс. Нижняя кривая соответствует фототоку при низком обратном смещении. На верхней кривой представлены максимальные значения амплитуды импульсов умноженного фототока, которые достигаются при смещении, равном напряжению пробоя [45].

тельность детектирующих систем, состоящих из фотодиодов с барьером Шоттки и приемника. Кроме того, имеется возможность усиления высокочастотных импульсов фототока.

Лавинные фотодиоды с гетеропереходами, особенно из соединений типа $A^{III}BV$, обладают рядом потенциальных преимуществ по сравнению с приборами на основе Ge и Si. Изменяя состав соединений, можно регулировать спектральную характеристику чувствительности прибора. Поскольку прямозонные соединения типа $A^{III}BV$ имеют высокий коэффициент поглощения, то в приборах на их основе может быть получена высокая квантовая эффективность даже в тех случаях, когда с целью достижения высокого быстродействия в них используются тонкие обедненные слои. Кроме того, прозрачный слой гетероструктуры (поверхностный слой с более широкой запрещенной зоной) может быть выращен по технологии, которая обеспечивает высокочастотные характеристики и минимальные потери, связанные с рекомбинацией фотоносителей.

Лавинные фотодиоды были изготовлены [47, 48] с использованием различных систем из соединений типа $A^{III}BV$, в том числе AlGaAs/GaAs, AlGaSb/GaSb, GaInAs/InP и GaInAsP/InP. На начальной стадии развития таких фотодиодов был получен ряд обнадеживающих результатов по быстродействию и квантовой эффективности. Последние могут превосходить аналогичные харак-

теристики существующих лавинных Ge-фотодиодов, работающих в диапазоне 1—1,6 мкм. Вместе с тем в этом направлении предстоит еще широкие исследования с целью изучения таких параметров материалов, как темновые токи, коэффициенты поглощения, отношения коэффициентов ионизации, а также с целью совершенствования материалов, методов изготовления, конструкций приборов и их надежности.

Последние исследования отношения коэффициентов ионизации в $\text{Al}_x\text{Ga}_{1-x}\text{Sb}$ -переходах показали, что, когда спин-орбитальное расщепление валентной зоны Δ приближается к значению, равному ширине запрещенной зоны (рис. 29, вставка), это отношение может быть очень малым [57]. График на рис. 29 демонстрирует резкое уменьшение k' при $\Delta/E_g \approx 1$. Были получены значения k' менее 0,04, что соответствует значениям шум-фактора менее 5 при $M = 100$.

Ряд лавинных фотодиодов был создан с использованием конструкции, изображенной на рис. 23, г. В качестве подложек использовались полупроводники III—V групп (например, GaAs и InP). Слои тройных или четверных соединений с близкими параметрами решеток выращивались на подложках, например, с помощью методов жидкостной, газовой или молекулярно-лучевой эпитаксии. На рис. 30 (вставка) показана структура $\text{Al}_{0,85}\text{Ga}_{0,15}\text{As}/\text{GaAs}$ [48]. Верхний слой структуры играет роль «окна», прозрачного для падающего излучения в диапазоне 0,5—0,9 мкм. Квантовая эффективность при $\lambda = 0,53$ мкм равна $\sim 70\%$ и может быть увеличена до 95 % при использовании просветляю-

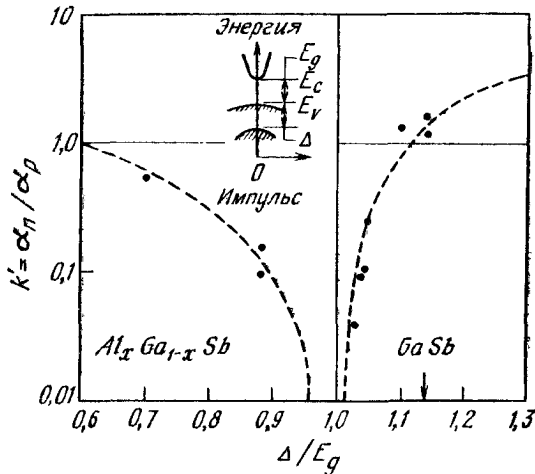


Рис. 29. Зависимость отношения коэффициентов ионизации в $\text{Al}_x\text{Ga}_{1-x}\text{Sb}$ -диоде от Δ/E_g , где Δ — спин-орбитальное расщепление валентной зоны (вставка) [57].

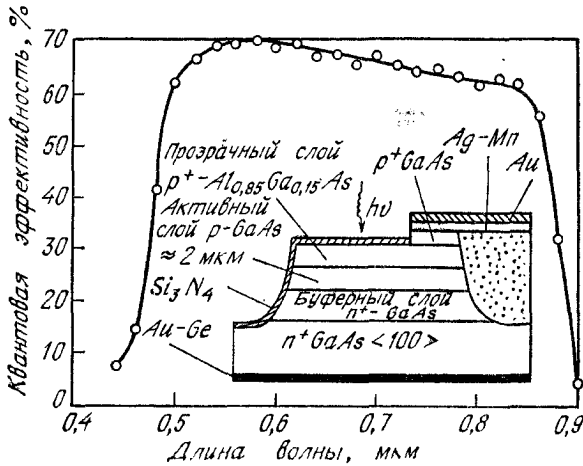


Рис. 30. Зависимость квантовой эффективности от длины волны излучения для лавинного фотодиода с гетеропереходом AlGaAs/GaAs. На вставке показана структура прибора (лавинного Al_{0,85}Ga_{0,15}As/GaAs-фотодиода) [48].

щего покрытия из Si₃N₄ (рис. 30) Темновой ток при напряжении, равном половине пробивного напряжения, не превышал $\sim 10^{-8}$ А/см². Время нарастания импульса составляет 35 пс, а коэффициент умножения на частоте 273 МГц равен 100. Отношение $k = \alpha_p/\alpha_n$ составило 0,83, что, по-видимому, является следствием сравнимых значений коэффициентов ионизации электронов и дырок в GaAs с ориентацией <100>. Необходимо отметить, что в GaAs с ориентацией <111> коэффициент ионизации дырок существенно больше, чем коэффициент ионизации электронов (см. гл. 1). Следовательно, для достижения минимального уровня шума необходимо использовать GaAs с ориентацией <111>, в котором процесс лавинного умножения стимулируется дырками.

На рис. 31 (вставка) представлена гетероструктура [49], в которой планарный $p-n$ -переход сформирован в прозрачном слое InP, отделенном от поглощающего слоя GaInAsP с более узкой запрещенной зоной. В этом диоде обедненная область распространяется из n -InP в слой n -GaInAsP от плоской части перехода, оставаясь внутри InP на периферии перехода. Так как пробивное напряжение V_B пропорционально $E_g^{3/2}$, то его значение в InP примерно на 60 % выше, чем в GaInAsP (при условии, что оба слоя имеют одинаковые уровни легирования). В результате этого в рассматриваемой диодной структуре предотвращается краевой пробой. Уменьшение токов утечки достигается также за счет подавления туннельного пробоя в широкозонном материале (т. е. в InP), соседствующего со слоем с более узкой запрещенной

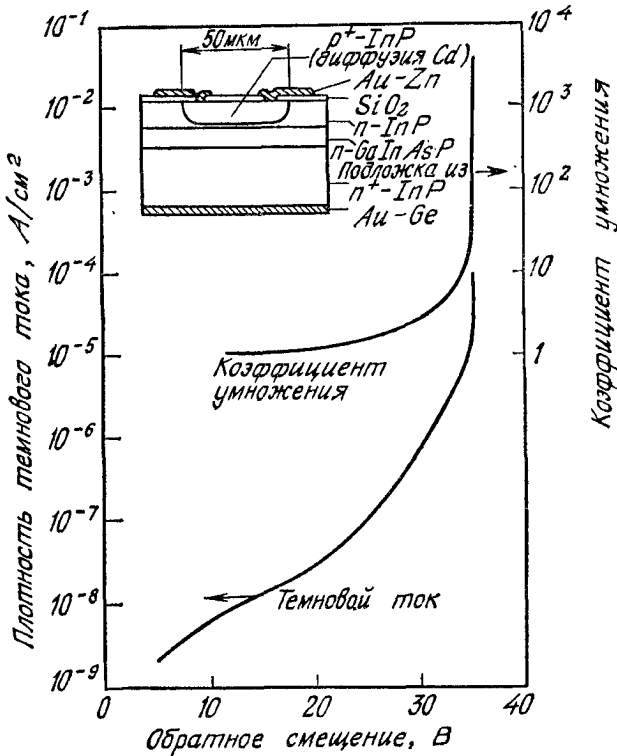


Рис. 31. Зависимость коэффициента умножения и темнового тока от величины обратного смещения для лавинного фотодиода с гетероструктурой GaInAsP [49].

зоной (т. е. GaInAsP). Прибор имеет низкую плотность темнового тока $\sim 10^{-6}$ А/см² при $\frac{1}{2} V_B$ и коэффициент умножения 3000.

На рис. 32, а показана зонная диаграмма лавинного фотодиода в состоянии равновесия, имеющего отдельные области поглощения и умножения [58]. Переход $p^+ - n$ сформирован в слое n -InP (область умножения). Слой GaInAs, выращенный на n -InP, используется в качестве области поглощения. Поскольку в InP коэффициент ионизации дырок в 2—3 раза больше, чем коэффициент ионизации электронов ($k' = 0,4$), то процесс лавинного умножения должен стимулироваться дырками. Для обеспечения инжекции дырок в область умножения через барьер на границе слоев n -InP и n -GaInAs необходимо подобрать уровни легирования и толщины этих слоев таким образом, чтобы обедненная область в n -InP в условиях лавинного пробоя достигала слоя GaInAs (рис. 32, б) Квантовая эффективность прибора составляет 40 %

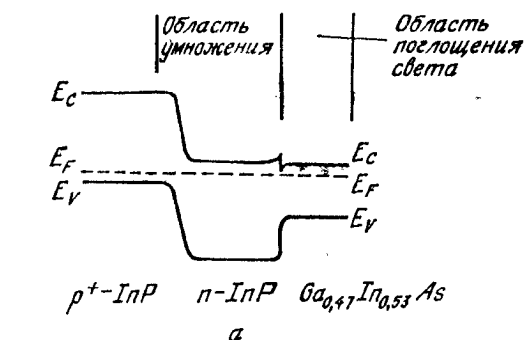
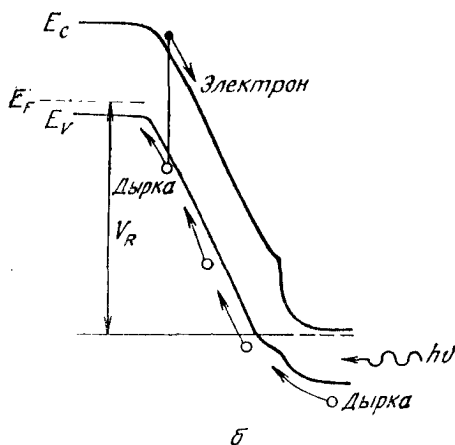


Рис. 32. Зонная диаграмма гетероперехода InP — GaInAs в условиях теплового равновесия (а) и зонная диаграмма при лавинном пробое (б) [58].



при $\lambda = 1,3$ мкм и 50 % при $\lambda = 1,6$ мкм, а шум-фактор на 3 дБ ниже, чем в лавинном Ge-фотодиоде, работающем на длине волны 1,15 мкм.

13.5. ФОТОТРАНЗИСТОР

Все рассмотренные в предыдущих главах биполярные и униполярные транзисторы могут выполнять функции фотодетекторов и, следовательно, называться фототранзисторами. Присущий этим приборам транзисторный эффект обеспечивает высокое усиление. Вместе с тем по сравнению с фотодиодом фототранзистор более сложен в изготовлении и уступает ему в быстродействии вследствие большей площади.

Структура биполярного фототранзистора [51] и его эквивалентная схема приведены на рис. 33, а. Он отличается от обычного биполярного транзистора тем, что переход база — коллектор,

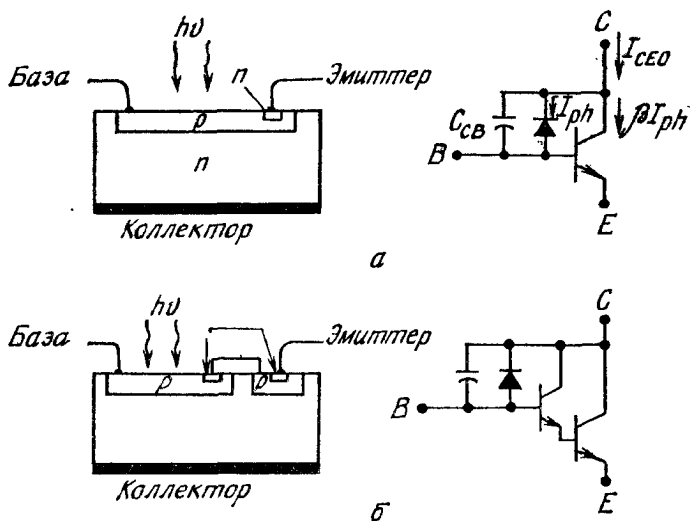


Рис. 33. Биполярный фототранзистор (а) и составной фототранзистор (б) [51].

выполняющий в нем роль чувствительного элемента (на рис. 33, а показан в виде диода с параллельно включенной емкостью), имеет большую площадь. Фототранзисторы особенно эффективны в составе оптронов, поскольку обеспечивают высокий коэффициент преобразования по току (т. е. отношение выходного тока фотодетектора к входному току светодиода), достигающий 50 % и более, тогда как при использовании фотодиодов этот коэффициент не превышает 0,2 %. В режиме работы с плавающей базой фотоносители дают вклад в ток коллектора в виде фототока I_{ph} . Кроме того, дырки, фотогенерированные в базе, а также дырки, приходящие в базу из коллектора, понижают разность потенциалов между собой и эмиттером, что приводит к инжекции электронов через базу в коллектор. Общий коллекторный ток равен

$$I_{CEO} = I_{ph} + h_{FE} I_{ph} = (1 + h_{FE}) I_{ph}, \quad (56)$$

где h_{FE} — статический коэффициент усиления транзистора в схеме с общим эмиттером, который может быть существенно больше единицы. Вследствие этого эффективный квантовый выход фототранзистора в $(1 + h_{FE})$ раз больше, чем квантовый выход фотодиода, которым является переход база — коллектор.

Мощность, эквивалентная шуму, определяется выражением (20), в котором [52]

$$I_{eq} = I_{CEO} (1 + 2h_{fe}^2/h_{FL}), \quad (57)$$

где h_{fe} — дифференциальный коэффициент усиления транзистора в схеме с общим эмиттером. Таким образом, уровень шума прибора и усиление взаимосвязаны.

Биполярный фототранзистор может быть интегрально совмещен с другими приборами. Например, используя дополнительный транзистор, можно сформировать составной фототранзистор с существенно более высоким коэффициентом усиления (рис. 33, б). Быстродействие рассмотренных структур ограничивается большой емкостью перехода база — коллектор и снижается при увеличении усиления за счет эффекта обратной связи. Так, типичное время отклика для фотодиода составляет $\sim 0,01$ мкс, тогда как для биполярного фототранзистора оно равно ~ 5 мкс, а для составного фототранзистора ~ 50 мкс.

Фототранзистор может быть изготовлен на основе гетероструктуры. На рис. 34 (вставка) показана зонная энергетическая диаграмма симметричной структуры $n\text{-AlGaAs}/p\text{-GaAs}/n\text{-AlGaAs}$ [53]. Этот прибор характеризуется высокими значениями напряжения запирания и коэффициента усиления для обеих полярностей смещения и линейной зависимостью тока от напряжения, проходящей через начало координат (рис. 34). Эффективная инжекционная способность обеспечивается за счет того, что слой базы имеет более узкую запрещенную зону, чем слои коллектора и эмиттера. Таким путем устраняется присущее обычным транзисторам ограничение, состоящее в необходимости создания слаболегированного базового слоя, и можно реализовать условия, при которых обед-

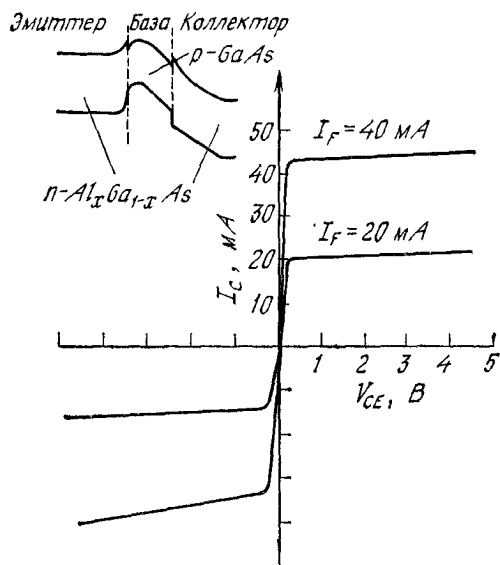
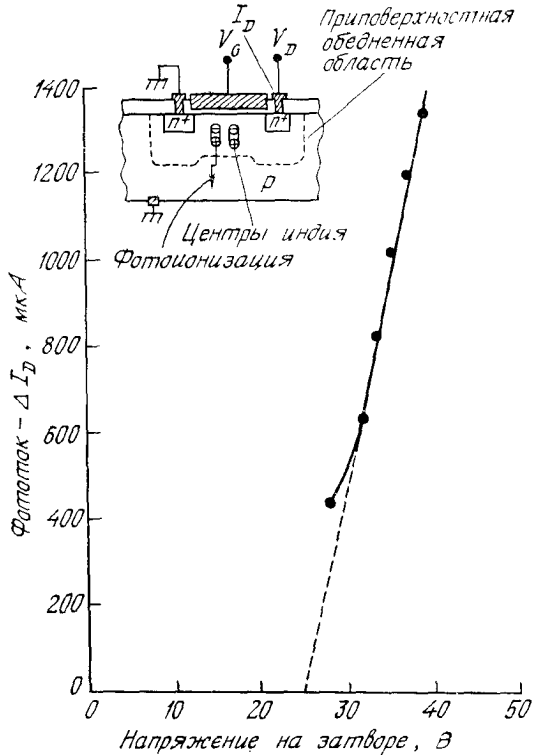


Рис. 34. Вольт-амперная характеристика фототранзистора с двусторонним гетеропереходом [53].

Рис. 35. Зависимость фототока в области насыщения от напряжения на полевом электроде в МОП-транзисторе с подложкой, легированной индием [54].



ненная область запирающего перехода в основном сосредоточена в слое коллектора. Это позволяет получить в одном приборе высокое усиление и высокое напряжение запираания. Прибор обладает симметричным коэффициентом усиления, который при напряжении запираания ± 10 В, соответствующем толщине базы GaAs 2,1 мкм, составляет 180, а при напряжении запираания $\pm 2,6$ В, соответствующем базе толщиной 0,3 мкм, равен 3000.

Структура кремниевого МДП-транзистора с изолированным затвором, чувствительного к ИК-излучению [54], схематически представлена на рис. 35 (вставка). (Характеристики МДП-транзисторов с изолированным затвором рассмотрены в гл. 8.) Подложка р-типа легирована бором, дающим мелкий акцепторный уровень, и индием, дающим глубокий акцепторный уровень (глубина залегания 0,16 эВ от края валентной зоны). В заполненном дырками состоянии центры индия нейтральны, а при возбуждении дырок в валентную зону становятся отрицательно заряженными. Прибор работает следующим образом. Отрицательное напряжение на полевом электроде вызывает обогащение канала; при этом все

центры индия захватывают дырки и переходят в нейтральное состояние. Под действием инфракрасного излучения в диапазоне 2—7 мкм, обеспечивающего энергию фотона, достаточную для фотоионизации центров индия, происходит возбуждение дырок в валентную зону. Этот процесс приводит к увеличению суммарной плотности заряженных акцепторов в области обеднения, что в свою очередь модулирует проводимость поверхностного канала. Выражение для изменения порогового напряжения имеет вид

$$\Delta V_T = 2 \sqrt{q \varepsilon_s \psi_B} (\sqrt{N_A + N_{In}} - \sqrt{N_A}) / C_i, \quad (58)$$

где ψ_B — разность потенциалов между уровнем Ферми и уровнем Ферми собственного полупроводника, C_i — удельная емкость окисла, N_A — концентрация бора и N_{In} — концентрация фотоионизованных центров индия. Ток канала в области насыщения пропорционален $(V_G - V_T)^2$, и, таким образом, изменение тока канала может быть представлено в виде

$$\Delta I_D \sim \Delta V_T (V_G - V_T - \Delta V_T / 2). \quad (59)$$

На рис. 35 приведена зависимость фототока в области насыщения от напряжения на полевом электроде. Полученная зависимость хорошо согласуется с уравнением (59).

Используя вместо индия другие акцепторные примеси, можно обеспечить оптимальное согласование области чувствительности прибора с требуемым спектральным диапазоном. Рассмотренный ИК-детектор может быть интегрально совмещен с элементами накопления и памяти и выходным усилителем, что открывает широкие перспективы для его использования в многоэлементных ИК-фотоприемниках большой степени интеграции.

ЛИТЕРАТУРА

1. Anderson L. K., DiDomenico M., Jr., Fisher M. B. High-Speed Photodetector for Microwave Demodulation of Light, in Young L., Ed., *Advances in Microwaves*, Vol. 5, Academic, N. Y., 1970, pp. 1—122.
2. Melchior H. Demodulation and Photodetection Techniques, in Arecchi F. T., Schulz-Dubois E. O., Eds., *Laser Handbook*, Vol. 1, North-Holland, Amsterdam, 1972, pp. 725—835.
3. Melchior H. Detector for Lightwave Communication, *Phys. Today*, p. 32 (Nov. 1977).
4. Willardson R. K., Bear A. C., Eds. *Semiconductors and Semimetals*, Vol. 12, Infrared Detector II, Academic, N. Y., 1977.
5. Stillman G. E., Wolfe C. M. Avalanche Photodiode, in Ref. 4, pp. 291—394.
6. Lee T. P., Li T. Y. Photodetectors, in Miller S. E., Chynoweth A. G., Eds. *Optical Fiber Communications*, Academic, N. Y., 1979, Chap. 18.
7. Bratt P. R. Impurity Ge and Si Infrared Detectors, in Ref. 4, pp. 39—142.
8. Bube R. H. Comparison of Solid State Photoelectronic Radiation Detectors, *Trans. AIME*, 239, 291 (1967).
9. DiDomenico M., Jr., Svelto O. Solid State Photodetection Comparison between Photodiodes and Photoconductors, *Proc. IEEE*, 52, 136 (1964).

10. Van der Ziel A. Fluctuation Phenomena in Semiconductors, Academic, N. Y., 1959, Chap. 9. [Имеется русский перевод: Ван-дер-Зил А., Флуктуационные явления в полупроводниках, М., 1961.]
11. Ross M. Laser Receivers-Devices, Techniques, Systems, Wiley, N. Y., 1966. [Имеется русский перевод: Росс М. Лазерные приемники, М.: Мир, 1969.]
12. Kruse P. W., McGlauchlin L. D., McQuistan R. B. Elements of Infrared Technology, Wiley, N. Y., 1962.
13. Eisenman W. L., Merriam J. D., Potter R. F. Operational Characteristics of Infrared Photodiode, in Ref. 4, pp. 1—38.
14. Stillman G. E., Wolfe C. M., Dimmock J. O. Far-Infrared Photoconductivity in High Purity GaAs, in Ref. 4, pp. 169—290.
15. Gartner W. W. Depletion-Layer Photoeffects in Semiconductors, *Phys. Rev.*, **116**, 84 (1959).
16. Lee H. S., Sze S. M. Silicon $p-i-n$ Photodetector Using Internal Reflection Method, *IEEE Trans. Electron Devices*, **ED-17**, 342 (1970).
17. Muller J. Thin Silicon Film $p-i-n$ Photodiodes with Internal Reflection, *IEEE Trans. Electron Devices*, **ED-25**, 247 (1978).
18. Ahlstrom E., Gartner W. W. Silicon Surface-Barrier Photocells, *J. Appl. Phys.*, **33**, 2602 (1962).
19. Schneider M. V. Schottky Barrier Photodiodes with Antireflection Coating, *Bell Syst. Tech. J.*, **45**, 1611 (1966).
20. Crowell C. R., Sze S. M. Hot Electron Transport and Electron Tunneling in Thin Film Structures, in Thun R. E., Ed., *Physics of Thin Films*, Vol. 4, Academic, N. Y., 1967, pp. 325—371.
21. Гуткин А. А., Дмитриев В. М., Хайт В. М., *ФТП*, **11** (1977).
22. Lavagna M., Pique J. P., Marfaing Y. Theoretical Analysis of Quantum Yield in Schottky Diodes, *Solid State Electron.*, **20**, 235 (1977).
23. Sharpless W. M. Cartridge-Type Point Contact Photodiode, *Proc. IEEE*, **52**, 207 (1964).
24. Casey H. C., Jr., Panish M. B. Heterostructure Lasers, Part B, Academic, N. Y., 1978, p. 3. [Имеется русский перевод: Кейси Х., Паниш М. Лазеры на гетероструктурах, М.: Мир, 1981.]
25. Miller R. C., Schwartz B., Koszi L. A., Wagner W. R. A High-Efficiency GaAlAs Double-Heterostructure Photovoltaic Detector, *Appl. Phys. Lett.*, **33**, 721 (1978).
26. Ilegems M., Schwartz B., Koszi L. A., Miller R. C. Integrated Multijunction GaAs Photodetector with High Output Voltage, *Appl. Phys. Lett.*, **33**, 629 (1978).
27. Burrus C. A., Dentai A. G., Lee T. P. InGaAsP $p-i-n$ Photodiodes with Low Dark Current and Small Capacitance, *Electron. Lett.*, **15**, 655 (1979).
28. Lee T. P., Burrus C. A., Dentai A. G. InGaAs/InP $p-i-n$ Photodiodes for Lightwave Communications at 0,95 to 1,65 μm Wavelengths, *IEEE J. Quantum Electron.*, **QE-17**, 232 (1981).
29. Ahmad K., Mabbitt A. W. GaInAs Photodiodes, *Solid State Electron.*, **22**, 327 (1979).
30. Johnson K. M. High-Speed Photodiode Signal Enhancement at Avalanche Breakdown Voltage, *IEEE Trans. Electron Devices*, **ED-12**, 55 (1965).
31. Anderson L. K., McMurtry B. J. High Speed Photodetectors, *Proc. IEEE*, **54**, 1353 (1966).
32. Emmons R. B. Avalanche Photodiode Frequency Response, *J. Appl. Phys.*, **38**, 3705 (1967).
33. Melchior H., Lynch W. T. Signal and Noise Response of High Speed Germanium Avalanche Photodiodes, *IEEE Trans. Electron Devices*, **ED-13**, 829 (1966).
34. McIntyre R. J. Multiplication Noise in Uniform Avalanche Diodes, *IEEE Trans. Electron Devices*, **ED-13**, 164 (1966).
35. Baertsch R. D. Noise and Ionization Rate Measurements in Silicon Photodiodes, *IEEE Trans. Electron Devices*, **ED-13**, 987 (1966).

36. McIntyre R. J. The Distribution of Gains in Uniformly Multiplying Avalanche Photodiodes: Theory, *IEEE Trans. Electron Devices*, **ED-19**, 703 (1972).
37. Webb R. P., McIntyre R. J., Conradi J. Properties of Avalanche Photodiodes, *RCA Rec.*, **35**, 234 (1974).
38. Anderson L. K., McMullin P. G., D'Asaro L. A., Goetzberger A., Microwave Photodiodes Exhibiting Microplasma — Free Carrier Multiplication, *Appl. Phys. Lett.*, **6**, 62 (1965).
39. Sze S. M., Gibbons G., Effect of Junction Curvature of Breakdown Voltage in Semiconductors, *Solid State Electron.*, **9**, 831 (1966).
40. Ando H., Kanbe H., Kimura T., Yamaoka T., Kaneda T. Characteristics of Ge Avalanche Photodiodes in the Wavelength Region of 1—1.6 μm , *IEEE J. Quantum Electron.*, **QE-14**, 804 (1978).
41. Ruegg H. W. An Optimized Avalanche Photodiode, *IEEE Trans. Electron Devices*, **ED-14**, 239 (1967).
42. Moll J. Physics of Semiconductors, McGraw-Hill, N. Y., 1964.
43. Kanbe H., Kimura T. Figure of Merit for Avalanche Photodiodes, *Electron. Lett.*, **13**, 262 (1977).
44. Melchior H., Hartman A. R., Schinke A. R., Seidel T. E. Planar Epitaxial Silicon Avalanche Photodiode, *Bell Syst. Tech. J.*, **57**, 1791 (1978).
45. Melchior H., Lepselter M. P., Sze S. M. Metal — Semiconductor Avalanche Photodiode, *IEEE Solid-State Device Res. Conf.*, Boulder, Colo, June 17—19, 1968.
46. Lepselter M. P., Sze S. M. Silicon Schottky Barrier Diode with Near-Ideal I—V Characteristics, *Bell Syst. Tech. J.*, **47**, 195 (1968).
47. Hurwitz C. E., Hsieh J. J. GaInAsP/InP Avalanche Photodiodes, *Appl. Phys. Lett.*, **32**, 487 (1978).
48. Law H. D., Nakano K., Tomasetta L. R. III—V Alloy Heterostructure High Speed Avalanche Photodiodes, *IEEE J. Quantum Electron.*, **QE-15**, 549 (1979).
49. Nishida K., Taguchi K., Matsumoto Y. InGaAsP Heterostructure Avalanche Photodiodes with High Avalanche Gain, *Appl. Phys. Lett.*, **35**, 251 (1979).
50. Shockley W., Sparks M., Teal G. K. $p - n$ Junction Transistor, *Phys. Rev.*, **83**, 151 (1951).
51. Jayson J. S., Knight S. Opto-Isolator, in Wolfe R., Ed., Applied Solid State Science, Vol. 6, Academic, N. Y., 1976, pp. 119—168.
52. DeLaMoneda F. H., Chenette E. R., Van Der Ziel A. Noise in Phototransistors, *IEEE Trans. Electron Devices*, **ED-18**, 340 (1971).
53. Knight S., Dawson L. R., Keramidis U. G., Spencer M. G. An Optically Triggered Double Heterostructure Linear Bilateral Phototransistors, *IEEE Tech. Dig., Int. Electron Device Meet.*, 1977, p. 472.
54. Forbes L., Wittmer L. L., Loh K. W. Characteristic of the Indium-Doped Infrared Sensing MOSFET, *IEEE Trans. Electron Devices*, **ED-23**, 1272 (1976).
55. Smith R. G., Personick S. D. Receiver Design for Optical Communications Systems. Kressel H., Ed., Semiconductor Devices for Optical Communication, Springer-Verlag, N. Y., 1979, Chap. 4.
56. Campbell J. C., Dentai A. G., Lee T. P., Burrus C. A. Improved Two-Wavelength Demultiplexing InGaAsP Photodetector, *IEEE J. Quantum Electron.*, **QE-16**, 601 (1980).
57. Hildebrand O., Kuebart W., Pikuha M. H. Resonant Enhancement of Impact Ionization in $\text{Al}_x\text{Ga}_{1-x}\text{Sb}$ $i - i - n$ Avalanche Photodiodes, *IEEE Device Res. Conf.*, June 23—25, 1980.
58. Susa N., Nakagome H., Mikami O., Ando H., Kanbe H. New InGaAs/InP Avalanche Photodiode Structure for the 1—1.6 μm Wavelength Region *IEEE J. Quantum Electron.*, **QE-16**, 864 (1980).
59. Forrest S. R., Leheny R. F., Nahory R. E., Pollack M. A. $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ Photodiodes with Dark Current Limited by Generation-Recombination and Tunneling, *Appl. Phys. Lett.*, **37**, 322 (1980).