



БИБЛИОТЕЧКА · КВАНТ ·
выпуск 65

М. Е. ЛЕВИНШТЕЙН
Г. С. СИМИН

БАРЬЕРЫ





БИБЛИОТЕЧКА · КВАНТ ·

выпуск 65

**М. Е. ЛЕВИНШТЕЙН
Г. С. СИМИН**

**БАРЬЕРЫ
ОТ КРИСТАЛЛА
ДО ИНТЕГРАЛЬНОЙ
СХЕМЫ**



**МОСКВА «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
1987**

РЕДАКЦИОННАЯ КОЛЛЕГИЯ:

Академик Ю. А. Осипьян (председатель), академик
А. Н. Колмогоров (заместитель председателя), кандидат физ.-мат.
наук А. И. Буздин (ученый секретарь), член-корреспондент АН СССР
А. А. Абрикосов, академик А. С. Боровик-Романов, академик
Б. К. Вайнштейн, заслуженный учитель РСФСР Б. В. Воздвижен-
ский, академик В. Л. Гинзбург, академик Ю. В. Гуляев, академик
А. П. Еринов, профессор С. П. Капица, академик А. Б. Мигдал,
академик С. П. Новиков, академик АПН СССР В. Г. Равумовский,
академик Р. З. Сагдеев, профессор Я. А. Смородинский, академик
С. Л. Соболев, член-корреспондент АН СССР Д. К. Фаддеев

Рецензент

доктор физико-математических наук Р. А. Суриц

Ответственный редактор выпуска Б. И. Шкловский

Левинштейн М. Е., Симин Г. С.

Л 36 Барьеры (От кристалла до интегральной схе-
мы).— М.: Наука. Гл. ред. физ.-мат. лит., 1987.—
320 с.— (Б-чка «Квант», Вып. 65.)

75 к., 134 000 экз.

Популярный рассказ о полупроводниковых приборах — о том, как они устроены, как работают, где применяются, как выращивают и очищают полупроводники, как изготавливают интегральные схемы, как зарождалась полупроводниковая электроника и какие удивительные перемены произошли в мире с ее появлением. В основе работы полупроводниковых приборов лежат свойства кристалла, из которого прибор изготовлен, и свойства барьеров, возникающих на границе между различными частями кристалла. Поняв свойства полупроводников и преодолев барьеры, можно разобраться в том, как работает любой прибор — от простейшего диода до самой сложной интегральной схемы.

Для учащихся старших классов, студентов, преподавателей.

Л 170406000—189
053(02)-87 163-87

ББК 22.37

ОГЛАВЛЕНИЕ

От авторов	6
Введение. Дорогу осилит идущий	7
Часть I. ПОЛУПРОВОДНИКИ	10
Глава 1. Что такое полупроводник	11
Металлы, диэлектрики, полупроводники	12
Собственные полупроводники	16
«Свободные» электроны в кристалле (17). Дырки (18). Генерация и рекомбинация (20). Тепловая генерация электронов и дырок (21). Рекомбинация электронов и дырок (22). Концентрация электронов и дырок в собственном полупроводнике (23).	
Бестигельная зонная плавка	24
Примесные полупроводники	26
Донорная примесь (26). Примесное истощение (28). Зависимость концентрации электронов от температуры (29). Дырки в электронном полупроводнике (30). Акцепторная примесь (32).	
Послушайте теперь ту же мелодию в исполнении фагота Компенсация (37).	33
Moralité	40
Глава 2. Рождение, жизнь и смерть электрона и дырки	41
Рождение и смерть	41
Глубокие центры: компенсаторы, ступеньки, убийцы (41).	
Жизнь и движение	53
Тепловое движение: средняя скорость, шарики для настольного тенниса на подносе и в ведре с водой (53). Движение в электрическом поле: подвижность, зависимость дрейфовой скорости от поля, энергетическая диаграмма (65).	
История самых главных понятий	74
Диффузия (коэффициент диффузии, диффузионный ток, соотношение Эйнштейна, скорость диффузии, диффузия неравновесных носителей) (75).	
Moralité	84

Часть II, БАРЬЕРЫ	86
Глава 3, Барьер на границе кристалла	87
Работа выхода	88
Вернись, я вновь и вновь зову — вернись (91). Двойной заряженный слой (93). Как определяется работа выхода и чему она равна (94). С полупроводниками, как всегда, дело обстоит непросто (99).	
Поверхностные состояния	103
Уровни Тамма (104). Реальная поверхность (104).	
Изгибы зон, поверхностный потенциал	105
Moralité	111
Глава 4. Основные параметры анергетических барьеров в полупроводниках	112
Как внешнее электрическое поле проникает в металл, диэлектрик, полупроводник	113
Металл (113). Диэлектрик (115). Полупроводник (116).	
Закон изменения поля с координатой	117
Распределение поля в барьере. Ширина барьера	120
Трансмутационное легирование полупроводников	125
Несколько слов о прямоугольных треугольниках	127
Moralité	128
Глава 5, $p - n$ переход	129
Способы получения $p - n$ переходов	130
Сплавление (132). Диффузия (133). Ионная имплантация (138).	
Барьер на границе	141
Высота барьера (144). Обедненный слой (147). Распределение поля в барьере. Ширина барьера (149). Чудесное равновесие (154).	
Обратное смещение	161
Высота и форма барьера (163). Обратный ток (164). Фотодиоды (170). Барьерная емкость (175). Варикапы (181). Ударная ионизация (188).	
Первая похвала терпению	198
Прямое смещение	199
Высота барьера (199). Прямой ток (200). Инжекция (205). Снова прямой ток (208). Светодиоды (210). Солнечные батареи (217). Выпрямительные диоды (222).	
Вторая похвала терпению	230
Moralité	231
Часть III. ТРАНЗИСТОРЫ	233
Глава 6. Биполярные транзисторы.	234
Принцип работы биполярного транзистора	234
Усиление по току (238). Притча о главном и деталях (240). Быстродействие транзистора (240).	
Как выглядят биполярные транзисторы и как они изготавливаются	251

Фотолитография	258
Простейшие транзисторные схемы	260
Схема усилителя с общим эмиттером (260). Схема усилителя с общей базой (270).	
Moralité	274
Глава 7. Полевой транзистор	275
На заре туманной юности.	275
Основная идея (275). Простые оценки (277). Старые друзья (278).	
Зрелость и расцвет	279
Полевой транзистор с $p - n$ переходом (279).	
Эпитаксия	281
Усердие все преодолевает: МДП транзисторы	285
Несколько важных деталей	289
Физическая картина работы полевого транзистора в реальных рабочих режимах	290
Основные параметры полевых транзисторов	295
Крутизна (295). Быстродействие (298).	
Полевой транзистор как элемент радиотехнических схем	299
Moralité	300
Глава 8. Транзисторы и жизнь	301
Первая королева	302
Гадкий утенок	305
Да здравствует король!	307
Король исчезает. Да здравствует король!	309
Претенденты на трон.	316
Объемные интегральные схемы (317). Полупроводниковые элементы оптических ЭВМ (317). Биорадиоэлектроника (318).	
Закключение	319

В этой книге рассказано о том, как устроены и как работают полупроводниковые приборы.

Путь к пониманию ведет через барьеры — достаточно крутые и сложные. Мы предлагаем свой маршрут. Разработать его нам помогли наши друзья: Михаил Игоревич Дьяконов и Борис Ионович Шкловский. Очень многие наши коллеги терпеливо отвечали на наши вопросы. Всем им мы глубоко благодарны.

Мы особенно благодарны Марианне Александровне Симиной, которая вместе с нами расставила вдоль маршрута знаки — предупреждающие, запрещающие и предписывающие. Сделала она это, создав образ горного туриста — нашего и вашего спутника в нелегком путешествии через барьеры. Если какой-нибудь раздел текста начинается с рисунка, на котором изображен турист отдыхающий — вас ждет рассказ, не требующий особого напряжения внимания. Если ваш спутник, обливаясь потом, топает вверх под рюкзаком — то же самое ожидает и вас. Наконец, если он переправляется через горный поток, зарубается ледорубом на крутом склоне или залядывает в орлиное гнездо — будьте внимательны, осторожны и терпеливы.

ДОРОГУ ОСИЛИТ ИДУЩИЙ

It's a long way to Tipperary,
It's a long way to go.
It's a long way to little Mary
To the sweetest girl I know *).

Полупроводниковых приборов существует на свете великое множество. В толстенных многотомных справочниках, где перечислены только названия приборов и указаны их самые основные свойства, описание полупроводниковых приборов занимает несколько тысяч страниц. А ведь в этих справочниках перечислены далеко не все приборы. И в них ничего не сказано ни о физических принципах, лежащих в основе работы приборов, ни о схемах, в которых они применяются.

Как научиться плавать в этом море?

Путь до Типперери долг. Но интересен.

Работа великого множества полупроводниковых приборов основана на сравнительно небольшом числе идей. Некоторые из этих идей очень просты и наглядны. Овладение другими требует усилий и терпения. Зато овладев кругом этих идей и основных принципов, можно понять, как устроен и как работает практически любой полупроводниковый прибор.

Осаду полупроводниковых приборов мы начнем, хорошо вооружившись и основательно укрепив тылы.

Сначала, в части I книги, мы повознакомимся с основными свойствами полупроводников. Знание этих свойств совершенно необходимо для понимания работы любого полупроводникового прибора. А для многих приборов и достаточно. В таких приборах используется замеча-

*) Как долг путь до Типперери,

Как еще долго шагать, шагать и шагать.

Еще шагать, шагать и шагать, пока я увижу Мэри,

Самую нежную девушку на свете.

(Старинная ирландская солдатская песня)

тельное свойство полупроводников сильно изменять свое электрическое сопротивление даже при очень слабом изменении внешних условий. Поэтому, измеряя сопротивление полупроводника, можно измерять температуру и освещенность, давление и магнитное поле, напряженность электрического поля и скорость движения жидкости или газа, ускорение или степень задымленности помещения. Такого рода приборы: терморезисторы, фото- и тензорезисторы, датчики Холла и диоды Ганна, широко используются в быту и в промышленности, в научных исследованиях и в военном деле. Каждый год выпускаются сотни тысяч, миллионы и десятки миллионов таких приборов.



Тем не менее в этой книге мы их рассматривать не будем. Во-первых, потому, что эти приборы уже рассмотрены подробно в другой книге Библиотечки «Квант» *). А во-вторых, потому, что все-таки самые важные приборы, в том числе и самый нужный, самый универсаль-

*) Левинштейн М. Е., Сими́н Г. С. Знакомство с полупроводниками. — М.: Наука, 1984. — Библиотечка «Квант», вып. 33.

ный, самый распространенный, самый, самый, самый полупроводниковый прибор — ТРАНЗИСТОР, относится к приборам другого типа.

Конечно, не зная основных свойств полупроводников, нельзя понять, как работает транзистор. Поэтому здесь мы повторяем самое необходимое из того, что было написано в книге «Знакомство с полупроводниками». При этом некоторые разделы мы переработали и сократили. Ну, а в тех случаях, когда автор «лучше выдумать не мог», соответствующие разделы оставлены без изменения.

Однако знакомства только со свойствами полупроводников недостаточно. Необходимо также разобраться в очень интересных, необычных и не всегда простых явлениях, разыгрывающихся не в объеме полупроводника, а на *границах*. В этой книге мы подробно обсудим явления на границе кристалла и окружающей среды и на границе между различными частями одного и того же полупроводника, в которые введены разные примеси.

Вблизи таких границ, как мы увидим, возникают области с очень необычными свойствами — *барьеры*. Свойства этих барьеров и лежат в основе работы диодов и транзисторов. Изучением свойств барьеров мы займемся в части II книги. И там же познакомимся с полупроводниковыми диодами.

В части III рассмотрены принципы работы, устройство и применение транзисторов.

Путь через барьеры сложен: он меньше всего напоминает прогулку по протоптанной дороге. Скорее уместно сравнить его с горным походом. В горах путешественника подстерегают лавины, камнепады, засыпанные снегом трещины... Увы, никто не расставляет вдоль маршрута предостерегающие знаки. Но поскольку наше восхождение предполагается массовым, мы все-таки решили обозначить самые характерные участки маршрута. Такая маркировка подскажет вам, где нужно быть особенно внимательным, а где можно расслабиться и отдохнуть.

ПОЛУПРОВОДНИКИ

Тела, кои в рассуждении способности проводить электричество, занимающие как бы среднее место между проводниками и непроводниками, обыкновенно называются полупроводниками.

Иван Двигубский.
«Начальные основания опытной физики», 1826 г.

Полупроводниками... будут называться проводники ... сопротивление которых очень сильно изменяется с температурой.

И. Кеттисбергер, 1914 г.



Промежуточная — большая, чем у металлов, но значительно меньшая, чем у диэлектриков, величина удельного электрического сопротивления ρ и очень сильное уменьшение сопротивления при нагреве являются самыми характерными признаками полупроводников. Чтобы понять, как возникает такое не очень большое, но и не очень маленькое удельное сопротивление, почему оно сильно зависит от температуры и как эти признаки связаны с другими характерными свойствами полупроводников, необходимо рассмотреть внутреннее строение полупроводниковых кристаллов.

Глава 1

ЧТО ТАКОЕ ПОЛУПРОВОДНИК

Ни крошка с Манежной,
Ни мужу жена...

*А. Вознесенский.
«Песня Офелии»*

Атомы любого вещества электрически нейтральны. Если приложить электрическое поле к объему, в котором находятся только нейтральные частицы, электрический ток через этот объем не потечет. В самом деле, электрический ток есть направленное движение заряженных частиц под действием электрического поля. Нет заряженных частиц — нет и тока. Таким образом, объем, заполненный атомами любого вещества, представляет собой идеальный изолятор.

Хорошо известный пример такого изолятора — воздух. В каждом кубическом сантиметре воздуха находятся $2,7 \cdot 10^{19}$ молекул кислорода (O_2), азота (N_2), водяных паров (H_2O) и прочих газов, которые мы не будем брать в расчет. В каждом атоме кислорода содержатся 18 положительно заряженных протонов и столько же отрицательно заряженных электронов. В каждом атоме азота — 7 протонов и 7 электронов. Казалось бы, заряженных частиц — хоть отбавляй. Но эти заряженные частицы связаны мощными электрическими силами в нейтральные атомы и молекулы, и в результате воздух — прекрасный изолятор. Высоковольтные линии передачи тянутся на многие тысячи километров, передавая энергию с потерями порядка 20%. А хорошо сконструированная батарейка способна годами стоять на воздухе, не разряжаясь. Точно таким же великолепным изолятором будет объем, заполненный газом нейтральных атомов серебра и меди, золота и ртути.

Почему же проводимость этих же веществ в твердом состоянии различается так сильно? Казалось бы, твердые тела состоят из таких же атомов. И все-таки проводимость хорошего проводника (например, серебра) больше, чем проводимость хорошего изолятора (например,

стекла), примерно в 10^{22} раз. В десять тысяч миллиардов миллиардов раз! *)

Более того, атомы одного и того же вещества, например углерода (С), в зависимости от того, в какую кристаллическую решетку твердого тела они соединились, могут образовать и очень хороший проводник — графит, и прекрасный изолятор — алмаз.

Металлы, диэлектрики, полупроводники

Последний пример — с атомами углерода — наводит на мысль, что будет ли твердое тело металлом или диэлектриком зависит не только и, может быть, даже не столько от свойств атомов, составляющих кристалл, сколько

от того, какие отношения сложатся у каждого атома с соседями, т. е. от типа связей, объединяющих атомы в кристаллическую решетку твердого тела. На рис. 1 показана кристаллическая решетка металла, а на рис. 2 — диэлектрика.

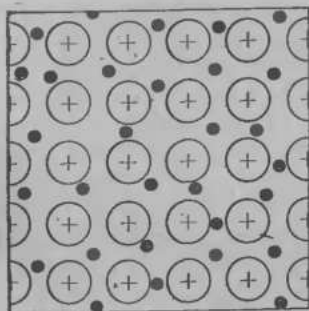


Рис. 1. Схематическое изображение кристаллической решетки металла. Правильная решетка положительно заряженных ионов погружена в «газ» из свободных электронов, не связанных жесткими связями с отдельными ионами

Упрощенное изображение кристаллической решетки на рис. 1 демонстрирует основные особенности металлических кристаллов. Кристаллическую решетку образуют не нейтральные атомы, а положительно заряженные ионы. При образовании решетки каждый атом металла лишился одного валентного электрона, и эти электроны (черные кружки на рис. 1) не принадлежат теперь никакому конкретному иону металла.

Электроны, как принято говорить, обобществлены кристаллом и могут свободно передвигаться под действием внешнего электрического поля. В каждом кубическом сантиметре металла содержится $\sim 10^{22}$ свободных носите-

*) Может быть, разница в проводимости хороших металлов и хороших изоляторов ощутится яснее, если представить себе, что во столько же ($\sim 10^{22}$) раз размер нашей Галактики ($\sim 10^{20}$ м) больше, чем 1 см.

лей тока — электронов. Не мудрено, что металлы являются превосходными проводниками.

Упрощенное изображение кристаллической решетки кремния на рис. 2 демонстрирует совсем другой характер

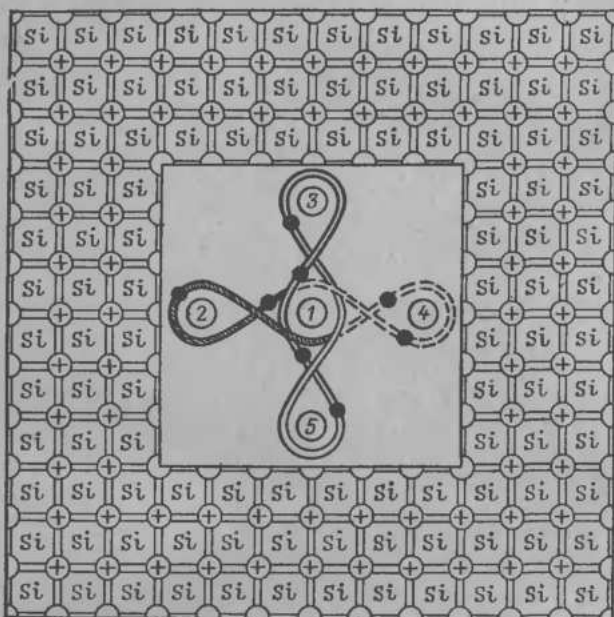


Рис. 2. Схематическое изображение кристаллической решетки кремния (Si). Черточки, связывающие друг с другом атомы Si, изображают электронные связи

связей между атомами, образующими кристалл. Каждый атом связан с ближайшими соседями прочными электронными связями.

В центре рис. 2 изображены несколько более подробно электронные связи произвольно выбранного атома 1 с его четырьмя ближайшими соседями. Вокруг атома по замкнутым (и на самом деле имеющим сложную форму) орбитам вращаются восемь электронов. Четыре электрона из этих восьми, так сказать, искони принадлежали атому 1: кремний, как известно, имеет на внешней электронной оболочке четыре валентных электрона. А откуда взялись остальные четыре? Из рис. 2 видно, что — от соседей по кристаллической решетке. Но атом 1 не отобрал эти электроны у соседей: соседские электроны проводят возле атома 1 лишь часть времени, другую часть они продолжают проводить возле своего исконного хозяина. С другой стороны, валентные электроны атома 1 теперь проводят возле него лишь

часть времени. Другую часть — воле соседних атомов. В результате в среднем отрицательный заряд электронов вокруг атома *1* по-прежнему эквивалентен заряду четырех электронов. Теперь, однако, атом *1* связан с атомами *2*, *3*, *4* и *5* (а те, в свою очередь, со своими соседями и т. д.) прочными электронными связями.

Чтобы разорвать электронные связи между атомами решетки и создать в таком кристалле свободные электроны, способные переносить электрический ток, необходимо затратить энергию. Откуда может взяться эта энергия? Если величина ее не слишком велика, а кристалл достаточно сильно нагрет, то электронные связи могут быть разорваны за счет тепловых колебаний решетки. Поток достаточно энергичных фотонов (или, попросту говоря, свет с подходящей длиной волны) также может разорвать электронные связи и создать в кристалле свободные электроны.

Но при достаточно низкой температуре (и в отсутствие света) кристалл, показанный на рис. 2, является идеальным изолятором: свободных носителей тока в нем нет.

Величина энергии, необходимая для того, чтобы разорвать электронные связи между атомами и создать в неметаллическом кристалле свободный электрон, обозначается символом \mathcal{E}_g . Она является важнейшей характеристикой кристалла. Значение \mathcal{E}_g зависит от структуры кристаллической решетки, от свойств входящих в нее атомов и у различных материалов лежит в пределах от нуля до нескольких десятков электрон-вольт *).

Если значение \mathcal{E}_g велико, то нагревание даже до очень высокой температуры не создает в кристалле сколь угодно заметного числа свободных электронов. При достаточно большом значении \mathcal{E}_g кристалл может даже расплавиться раньше, чем в нем за счет теплового движения возникнет хотя бы один свободный носитель. Такие материалы с большими значениями \mathcal{E}_g являются типичными диэлектриками.

При небольшом значении \mathcal{E}_g нагревание даже до не слишком высокой температуры приведет к разрыву зна-

*) Единица энергии 1 электрон-вольт (1 эВ) равняется $1,6 \times 10^{-19}$ Дж.

Конечно, в обычных житейских масштабах эта энергия ничтожна. Но в том мире, который мы собираемся изучать, мире атомов и электронов, величина 1 эВ выглядит очень солидно. Так, средняя энергия тепловых колебаний частицы при комнатной температуре равняется всего 0,026 эВ. Энергия фотона красного света равна ~1,8 эВ. Кинетической энергии электрона в 1 эВ соответствует скорость его движения ~ $6 \cdot 10^7$ см/с.

чительного числа электронных связей и появлению относительно большого числа свободных носителей. Кристалл приобретает способность проводить электрический ток. Ясно, что «в рассуждении способности проводить электричество» материал с небольшим значением \mathcal{E}_g будет занимать «как бы среднее место между проводниками и непроводниками». Ведь «способность проводить электричество», проводимость, пропорциональна концентрации свободных носителей тока. В проводниках (металлах) концентрация свободных носителей очень велика. Она соответствует ситуации, когда все электронные связи разорваны и все валентные электроны свободны. В «непроводниках» (диэлектриках) носителей тока практически нет.

Ясно также, что удельное сопротивление материала с небольшим значением \mathcal{E}_g может очень сильно изменяться с температурой. В металлах концентрация электронов не зависит от температуры. Даже при абсолютном нуле все электроны остаются свободными и сохраняют способность проводить ток. В типичном диэлектрике с большим значением \mathcal{E}_g тепловое движение не способно разорвать электронные связи и, следовательно, концентрация электронов также не будет изменяться с температурой. А вот в материалах с небольшим значением \mathcal{E}_g , *полупроводниках*, концентрация свободных носителей, пропорциональная числу разорванных электронных связей, будет очень резко расти с увеличением температуры. Соответственно удельное сопротивление ρ будет при нагреве сильно уменьшаться.

Итак, полупроводники — это неметаллические материалы с относительно небольшой величиной энергии \mathcal{E}_g .

Значение \mathcal{E}_g для типичных полупроводников лежит в пределах от нескольких десятых долей электрон-вольт до двух-трех электрон-вольт. Так, для антимонида индия (InSb) значение $\mathcal{E}_g = 0,17$ эВ. Такой небольшой энергией обладают уже кванты инфракрасного излучения с длиной волны $\lambda \approx 10$ мкм. Поэтому из InSb изготавливают приемники инфракрасного излучения, лежащие в основе приборов ночного видения. Для германия (Ge), из которого были изготовлены первые транзисторы и полупроводниковые диоды, $\mathcal{E}_g = 0,72$ эВ. Для кремния (Si) — основного материала современной полупроводниковой электроники, $\mathcal{E}_g = 1,1$ эВ. Арсенид галлия (GaAs) — наиболее перспективный материал полупроводниковой электроники ближайшего будущего, имеет значение $\mathcal{E}_g = 1,4$ эВ. В тройном полупроводниковом соединении

GaAlAs, используемом для создания полупроводниковых светодиодов и лазеров, значение \mathcal{E}_g в зависимости от относительного содержания Al и Ga, лежит в пределах от 1,4 эВ (GaAs) до 2,17 эВ (AlAs). В карбиде кремния (SiC), из которого созданы наиболее надежные и стабильные светодиоды, способные работать при очень высокой температуре, $\mathcal{E}_g \approx 3$ эВ.

Мы перечислили только несколько наиболее важных полупроводниковых материалов. Всего же синтезированы, изучены и используются на практике сотни полупроводниковых соединений.

Собственные полупроводники

Собственными называются полупроводники, в которых концентрация свободных носителей тока определяется только температурой и собственной, присущей данному полупроводнику, величиной энергии связи \mathcal{E}_g .

Может возникнуть законный вопрос: а чем, собственно, еще может определяться концентрация свободных носителей?

Как мы увидим в дальнейшем, одной из самых характерных особенностей полупроводников является очень высокая чувствительность их к невероятно малым количествам примесей. Совершенно ничтожная, неощутимая ни в каких других материалах добавка постороннего вещества в количестве одного атома на миллион или даже на миллиард атомов полупроводника может заметно изменить его свойства и привести к тому, что концентрация свободных носителей будет определяться не собственными свойствами полупроводника, а количеством и характером введенной примеси.

Представим себе, однако, самую простую ситуацию: никаких примесей, никаких дефектов и нарушений кристалл полупроводника не содержит. В идеальной кристаллической решетке, такой, как показана на рис. 2, каждый электрон удерживается на своей орбите, и чтобы освободить его, нужно затратить энергию \mathcal{E}_g . Пусть температура кристалла T . Какая доля электронных связей окажется разорванной? Или, что то же самое, сколько свободных электронов возникнет в кристалле при данной температуре?

Энергия хаотического теплового движения частиц, стремящегося разорвать электронные связи между атомами, по порядку величины равняется, как известно

kT , где k — постоянная Больцмана. (Значение $k = 1,38 \cdot 10^{-23}$ Дж/К $= 8,6 \cdot 10^{-5}$ эВ/К.) При комнатной температуре (300 К) энергия теплового движения $kT \approx 0,026$ эВ. При очень высокой с практической точки зрения температуре $+200$ °С, или ≈ 500 К, энергия $kT = 0,043$ эВ. Между тем, как мы уже знаем, значение \mathcal{E}_g для типичных полупроводников гораздо выше. Может показаться, что в таком случае тепловое движение вообще окажется неспособным разорвать ни одной электронной связи и свободные носители при условии $kT \ll \mathcal{E}_g$ в кристалле не возникнут. Однако это не так.

Дело в том, что величина kT характеризует лишь среднюю тепловую энергию частиц. Однако из-за случайного, хаотического характера теплового движения в кристалле в каждый данный момент найдутся атомы, чья энергия гораздо больше, чем средняя величина. В том числе и такие, чья энергия больше, чем значение \mathcal{E}_g . Таких атомов очень немного. Но они есть — и часть электронных связей между атомами окажется разорванной. Что при этом произойдет, показано на рис.

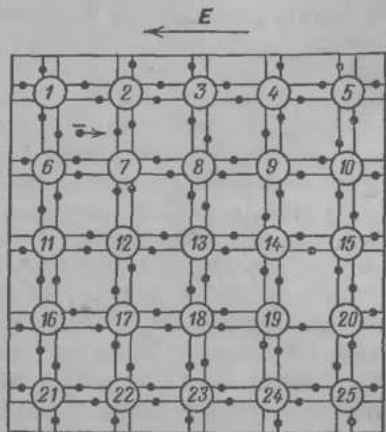


Рис. 3. Хаотическое тепловое движение разорвало одну из электронных связей (между атомами 19 и 20). Образовался свободный электрон (он находится между атомами 1, 2, 6 и 7) и дырка. В электрическом поле E отрицательно заряженный электрон будет двигаться вправо, а дырка — влево

3. Одна из связей (между атомами 19 и 20) разорвана.

Вывранный со своей орбиты электрон находится в данный момент в квадрате, образованном атомами 1, 2, 6 и 7.

«Свободные» электроны в кристалле. Из рис. 3 сразу видно, что, говоря об электроне, вырванном из системы межатомных связей, как о свободном, лучше бы все время, по крайней мере в уме, ставить слово «свободный» в кавычки. В самом деле, по-настоящему свободный электрон, электрон в вакууме, пока на него не действует электрическое поле, не испытывает никаких воздействий. «Свободный» электрон в кристалле находится в сложном элек-

трическом поле, образованном ионами решетки и валентными электронами соседних атомов. Под действием этих полей «свободный» электрон в кристалле совершает непрерывные хаотические перемещения (тем более быстрые и энергичные, чем выше температура кристалла).

Под действием внешнего электрического поля E свободный электрон в вакууме движется с постоянным ускорением $a = eE/m_0$ (m_0 — масса свободного электрона) и все время увеличивает скорость, пока она не приблизится к скорости света и не ограничится эффектами, рассматриваемыми теорией относительности (релятивистскими эффектами). «Свободный» электрон в кристалле под действием внешнего поля может двигаться свободно только в течение очень короткого времени. Затем... или тепловое движение решетки заставит электрон отклониться от траектории, предписываемой ему внешним полем, или электрон натолкнется на атом примеси или на какой-нибудь дефект кристаллической решетки. В результате частых столкновений электрон в кристалле даже в очень сильных электрических полях, как правило, движется вдоль поля со скоростью, не большей чем $(1-3) \cdot 10^5$ м/с. В слабых электрических полях из-за столкновений с решеткой скорость электрона в кристалле (а не ускорение, как в вакууме!) пропорциональна напряженности электрического поля *). Так что, употребляя термины «свободный электрон в кристалле», «свободный носитель заряда», следует помнить, что прилагательное «свободный» означает не тождество со свободным электроном в вакууме, а способность направленно двигаться под действием электрического поля, обуславливая электрический ток. Поэтому очень часто свободные электроны в кристалле называют *электронами проводимости*.



Дырки. Проследим теперь за поведением электронной связи между атомами 19 и 20 (рис. 3), разорванной тепловым движением. Причем будем сравнивать поведение этой связи с поведением свободного электрона.

Если к кристаллу не приложено электрическое поле, свободный электрон хаотически перемещается под действием тепловых колебаний между атомами решетки. А как ведет себя разорванная связь? Любой из электронов, связывающих атомы 19

*) Это и есть закон Ома. Постарайтесь сами установить связь между законом Ома и утверждением, что скорость электрона в кристалле пропорциональна напряженности поля.

и 20 с соседними атомами, может попасть на траекторию, с которой был выбит электрон, и восстановить полноценную связь между атомами 19 и 20. Если это произошло с электроном, связывавшим, например, атомы 14 и 19, то разорванная связь (будем для краткости называть ее дыркой) переместится в положение между атомами 14 и 19. Следующим шагом может быть перемещение дырки в позицию между атомами 9 и 14, или 19 и 18; или 14 и 15; мы видим, что, подобно электрону проводимости, разорванная связь, дырка, может хаотически перемещаться между атомами решетки.

Если к кристаллу приложено внешнее электрическое поле, на хаотическое движение свободного электрона накладывается направленное движение его против поля: возникает электрический ток *). Что в такой ситуации происходит с дыркой? Снова возвратимся к ситуации, показанной на рис. 3. При наличии электрического поля, конечно, на пустую траекторию между атомами 19 и 20 может, вообще говоря, переместиться любой электрон из связывающих атомы 19 и 20 с любым из их соседей. Но теперь, при наличии внешнего поля, наибольшую вероятность заместить разорванную связь имеют электроны, связывающие атомы 18 и 19. Именно эти электроны внешнее поле «тянет» на место разорванной связи. На хаотическое перемещение дырки накладывается направленное перемещение вдоль поля. Это не значит, что замещение обязательно произойдет электроном, связывающим атомы 18 и 19, а на следующем этапе дырка обязательно окажется между атомами 17 и 18. Наложение внешнего поля не прекращает хаотического движения дырки так же, как оно не прекращает хаотического движения электрона. В обоих случаях наложение поля приводит только к появлению в хаотическом движении частиц определенной направленности. И направленность эта (обратим на это особое внимание!) соответствует перемещению дырки в сторону, противоположную направленному движению электрона. Электрон проводимости направленно движется от «минуса» к «плюсу», а разорванная связь, дырка, — от «плюса» к «минусу» вдоль вектора электрического поля.

Еще раз подчеркнем: реально перемещаются электроны. Никакой реальной частицы «дырки» не существует

*) Напомним, что в электрическом поле электрон как отрицательно заряженная частица движется от «минуса» к «плюсу», т. е. против вектора напряженности электрического поля, который всегда направлен от «плюса» к «минусу».

(например, в отличие от электрона, дырку нельзя извлечь из кристалла и изучить ее поведение в вакууме). Чтобы подчеркнуть это обстоятельство, дырку иногда называют *квазичастицей*. Дырка есть просто отсутствие электрона на траектории, связывающей друг с другом атомы кристалла. Но гораздо удобнее следить не за перемещением электронов с одной связи на другую, а за перемещением пустой связи — «дырки».

Итак, если каким-либо образом (например, с помощью кванта света с достаточно большой энергией или за счет теплового движения) разорвать одну из межатомных электронных связей, то у кристалла появится не одна, а две возможности проводить ток. Первая — движение отрицательно заряженного свободного электрона под действием электрического поля. Вторая — направленное перемещение электронов с соседних орбит на место освободившейся пустой связи. Этот второй механизм электропроводности удобнее описывать как движение фиктивной частицы, «дырки», заряженной положительно и движущейся в направлении, противоположном движению электронов.

Генерация и рекомбинация. Хаотическое тепловое движение непрерывно разрывает электронные связи между атомами и создает (генерирует) свободные электроны и дырки. Не приведет ли этот процесс к тому, что все электронные связи окажутся в конце концов разорванными и полупроводник превратится в металл? Нет, не приведет!

Хаотически блуждая по кристаллу, свободный электрон и дырка могут оказаться рядом. Свободный электрон занимает свое место на свободной траектории межатомной связи, и... свободный электрон и дырка одновременно исчезают. Такой процесс называется рекомбинацией, что в переводе с латыни означает «воссоединение».

В установившемся, стационарном состоянии существует динамическое равновесие: сколько электронов и дырок (электронно-дырочных пар) создается каждую секунду в объеме полупроводника за счет тепловой генерации, столько же их будет гибнуть за счет рекомбинации. В полупроводнике при данной температуре устанавливается определенная равновесная концентрация электронов и дырок. Наша задача — определить величину равновесной концентрации в зависимости от значений T и \mathcal{E}_g .

Для этого следует определить, сколько свободных носителей создается в единице объема полупроводника

в единицу времени и сколько их гибнет за счет рекомбинации. В условиях равновесия эти две величины должны быть равны. Приравнивая их друг к другу, мы и определим равновесную концентрацию носителей.

Тепловая генерация электронов и дырок. Мы уже знаем, в чем состоит главная особенность генерации электронно-дырочных пар под действием теплового движения. Для их образования нужна энергия, значительно превышающая среднюю энергию kT .

Ситуации, когда для того, чтобы некоторое событие произошло, нужна большая «порция» энергии $\Delta\mathcal{E}$, значительно превышающая величину kT , встречаются при исследовании громадного числа самых разнообразных физических процессов. С ними приходится сталкиваться при изучении испарения воды и распада атомного ядра, испускания электронов катодом вакуумной лампы и распределения плотности газов в атмосфере. В одном из основных разделов физики — статистической физике, доказываемая, что вероятность такого события всегда пропорциональна $\exp(-\Delta\mathcal{E}/kT)$:

$$w \sim e^{-\Delta\mathcal{E}/kT}. \quad (1)$$

В выражении (1) e — число, равное 2,7183... и называемое основанием натуральных логарифмов.

Зависимость, описываемая выражением (1), называется экспоненциальной или, более фамильярно, экспонентой. Вглядитесь в нее внимательней — она будет сопровождать нас через всю книгу и встречаться очень часто: нам нередко придется иметь дело с событиями, для осуществления которых нужна энергия, значительно бо́льшая, чем средняя тепловая энергия $\sim kT$.

Значения энергий $\Delta\mathcal{E} = \mathcal{E}_g$, требующихся для разрыва электронных связей в различных полупроводниках, нам известны. Следовательно, зная температуру T , можно подсчитать по формуле (1) вероятность образования электронно-дырочной пары в различных материалах. При комнатной температуре ($kT = 0,026$ эВ) для InSb ($\mathcal{E}_g = 0,17$ эВ) эта вероятность пропорциональна $e^{-6,54} \approx 1,44 \cdot 10^{-3}$. Для Si ($\mathcal{E}_g = 1,1$ эВ) вероятность пропорциональна $e^{-42,3} \approx 4,23 \cdot 10^{-19}$, для GaAs ($\mathcal{E}_g = 1,4$ эВ) — $e^{-53,6} \approx 4,12 \cdot 10^{-24}$.

Приведенные примеры наглядно иллюстрируют основную особенность экспоненциальной зависимости: величина экспоненты очень сильно меняется при изменении показателя экспоненты. Увеличение $\Delta\mathcal{E}$ примерно в 1,27 раза при переходе от кремния к арсениду галлия влечет за собой уменьшение вероятности образования электронно-дырочной пары примерно в 100 000 раз! Заметим, что столь же резкое изменение вероятности образования электронов и дырок будет наблюдаться и в одном и том же полупроводнике при

изменении температуры. Например, чтобы уменьшить вероятность образования электронно-дырочной пары в Si в 100 000 раз, достаточно охладить кристалл кремния от комнатной температуры до -78°C , т. е. всего-навсего до температуры «сухого льда», которым охлаждают мороженое.

Итак, число электронно-дырочных пар K_1 , создаваемых каждую секунду в единице объема полупроводника, равняется

$$K_1 = \alpha e^{-E_g/kT}, \quad (2)$$

где α — коэффициент пропорциональности, различный для разных полупроводников.

С другой стороны, каждую секунду в той же единице объема исчезнет за счет рекомбинации некоторое число носителей K_2 . От чего же зависит число рекомбинирующих носителей?

Рекомбинация электронов и дырок. Для того чтобы рекомбинировать, электрон и дырка должны встретиться. Обсудим, от чего зависит частота таких встреч. Проведем следующий мысленный эксперимент. Будем следить за каким-нибудь атомом кристаллической решетки и отмечать появление в его окрестности, скажем, дырки. Ясно, что чем больше дырок имеется в единице объема полупроводника, тем чаще они будут появляться: вероятность появления дырки пропорциональна концентрации дырок p_i . По тем же причинам вероятность появления свободного электрона в окрестностях этого же атома пропорциональна концентрации электронов n_i *). Нас интересует вероятность одновременного появления в окрестностях атома и электрона, и дырки. Эта вероятность (т. е. вероятность их встречи и рекомбинации) пропорциональна произведению концентраций электронов и дырок $n_i p_i$. Таким образом, число носителей, рекомбинирующих каждую секунду в единице объема, равняется

$$K_2 = \beta n_i p_i. \quad (3)$$

Коэффициент пропорциональности β , так же, как и α в формуле (2), различный для разных полупроводников.

В собственных полупроводниках, которые мы изучаем, носители генерируются и рекомбинируют всегда парами. Поэтому $n_i = p_i$. Отсюда следует, что

$$K_2 = \beta n_i^2 = \beta p_i^2. \quad (4)$$

*) Индекс «i» у обозначений n_i и p_i — от английского слова intrinsic (внутренне присущий, собственный).

Концентрация электронов и дырок в собственном полупроводнике. Приравнивая число ежесекундно появляющихся нар K_1 числу гибнущих K_2 , получим

$$\alpha e^{-\mathcal{E}_g/kT} = \beta n_i^2 = \beta p_i^2. \quad (5)$$

Отсюда находим, что

$$n_i = p_i = \sqrt{\alpha/\beta} e^{-\mathcal{E}_g/2kT}$$

или

$$n_i = p_i = (AB)^{1/2} e^{-\mathcal{E}_g/2kT}. \quad (6)$$

Величины A и B измеряются в см^{-3} . Для всех полупроводников, с которыми нам придется иметь дело, значения A и B хорошо известны. При комнатной температуре эти значения лежат в пределах примерно от 10^{17} до 10^{19} см^{-3} . Зная величины A и B , легко рассчитать с помощью формулы (6) значения концентрации свободных носителей в любом полупроводнике при любой температуре. В табл. 1 для некоторых полупроводников приведены значения \mathcal{E}_g и n_i при комнатной температуре (300 K).

Таблица 1

Полупроводник	InSb	Ge	Si	InP
\mathcal{E}_g , эВ	0,17	0,72	1,1	1,3
Собственная концентрация n_i , см^{-3}	$1,3 \cdot 10^{16}$	$2,4 \cdot 10^{13}$	$1,1 \cdot 10^{10}$	$5,7 \cdot 10^7$
Полупроводник	GaAs	GaP	SiC	
\mathcal{E}_g , эВ	1,4	2,3	2,4—3,2	
Собственная концентрация n_i , см^{-3}	$1,4 \cdot 10^7$	0,8	$0,1—2 \cdot 10^{-8}$	

Какие выводы можно извлечь из табл. 1? Прежде всего зададим простой вопрос: легко ли изучать электрические свойства собственных полупроводников? В 1 см^3 различных кристаллов содержится $\sim 10^{22}$ атомов. Чистым мы привыкли называть вещество, в котором содержится один посторонний атом на 1000 собственных (т. е. концентрация примесей $\sim 0,1\%$). С химической точки зрения изумительно чистым будет вещество, в котором один посторонний атом приходится на 100 000 собственных. Тем

не менее это означает, что в каждом кубическом сантиметре вещества будет содержаться $\sim 10^{17}$ посторонних атомов. А теперь представим себе, что примесь, попавшая в полупроводник, способна сравнительно легко отдавать свободный электрон или образовывать дырку. (Чуть дальше мы убедимся, что очень многие примеси обладают этим свойством.) Тогда с химической точки зрения изумительно чистый германий (концентрация примесей всего 0,001%) будет при комнатной температуре содержать созданных примесью электронов примерно в 4000 раз больше, чем собственных. Кремний — в 10 000 000 раз больше. Фосфид галлия — в 10^{17} раз больше!

А ведь примеси содержатся и в исходных материалах, из которых получают полупроводниковые кристаллы, и в стенках печей и приборов, где происходит синтез и очистка полупроводников, и в окружающем воздухе.

Совершенно избавиться от примесей невозможно. Но совместными многолетними усилиями физиков и технологов, металлургов, химиков и многих других специалистов созданы многочисленные методы, позволяющие получать полупроводниковые кристаллы удивительной чистоты.

Бестигельная зонная плавка



Одним из методов получения чистых полупроводниковых кристаллов является метод *бестигельной зонной плавки*.

Впервые метод зонной плавки был применен для очистки германия в 1952 г. американским ученым В. Пфанном. Принцип метода поясняется рис. 4. Стержень из сравнительно грязного полупроводника помещается в поле мощного источника сверхвысокочастотного (СВЧ) излучения, создаваемое индуктором. При включении индуктора небольшая часть стержня, находящаяся внутри индуктора, расплавляется. Петля индуктора медленно движется вдоль стержня (на рис. 4 слева направо). Когда расплавленная часть выходит из петли индуктора, она вновь застывает (кристаллизуется). И вот оказывается, что в этой вновь застывшей части кристалла примесей содержится гораздо меньше, чем в исходном материале.

В основе метода зонной плавки лежит хорошо известное в химии явление сегрегации примесей. То самое явление, благодаря которому лед, кристаллизующийся из соленой морской воды, сам практически не содержит соли. С явлением сегрегации хорошо знакомы полярники. Отправляясь в путешествие по льдам соленого Северного Ледовитого океана, пресную воду они с собой не берут. Достаточно растопить кусок льда (образовавшегося из очень соленой воды!) — и пресная вода к услугам путешественников.

Причина явления сегрегации заключается в том, что когда в процессе остывания расплава начинается образование кристалли-

ческой решетки, чаще всего энергетически более выгодным оказывается размещение в строящейся решетке «своего», а не примесного атома.

Представим себе лужу солевой воды, которая при понижении температуры начала застывать. Образование льда, как правило, начинается с краев лужи. Если отломить от края прозрачную тонкую пластинку льда и попробовать ее на язык, она окажется совершенно пресной. Куда же девалась соль? При выстраивании кристаллической решетки льда она оказалась оттесненной из твердой фазы (льда) в находящуюся рядом жидкую фазу (воду). Строящейся решетке льда оказалось выгоднее забрать из находящегося рядом расплава (воды) «свои» молекулы, оставив чужеродные ионы Na и Cl в жидкой фазе.

Отношение концентрации какой-либо примеси в твердой фазе к концентрации той же примеси в находящейся рядом жидкой фазе называется коэффициентом распределения K_p . Чем меньше величина K_p , тем эффективнее очищается кристалл от примеси в процессе кристаллизации.

Метод зонной плавки чаще всего применяется для очистки кремния — самого важного полупроводникового материала современной полупроводниковой электроники. В таблице приведены коэффициенты распределения для некоторых из важнейших примесей, содержащихся в Si:

Примесь	Cu	Zn	B	Al	Ge	P	Sn	As
K_p	0,0045	10^{-5}	0,8	0,002	0,33	0,35	0,016	0,3

Из таблицы видно, что медь, цинк, алюминий удаляются из растущего монокристалла очень эффективно. За один проход зоны вдоль кристалла Si концентрация Cu уменьшается в $1/K_p = 220$ раз, Al — в 500 раз, Zn — в 10^5 раз. Примеси Ge, P, Sn и As удалить значительно сложнее. Поэтому для получения монокристаллов, отвечающих современным требованиям, иногда приходится осуществлять до 50 проходов зоны вдоль очищаемого стержня.

А вот примесь бора (B) для зонной плавки практически неуязвима. Для борьбы с ней приходится прибегать к специальным мерам. Перед тем как подвергнуть кремний зонной очистке, его расплавляют и помещают в атмосферу влажного водорода, который образует с бором летучее соединение гидрид BH_3 . Бор улетучивается, а остальные примеси зонная плавка уверенно берет на себя.

Исходным материалом, который подвергается очистке методом зонной плавки, является так называемый технический кремний, содержание примесей в котором составляет $\sim 1\%$. Его получают

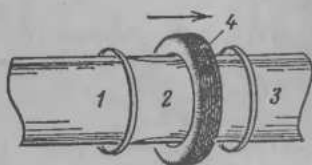


Рис. 4. Схема зонной плавки: 1 — участок кристалла, очищенный прохождением зоны; 2 — расплав (зона); 3 — участок неочищенного полупроводника; 4 — петля индуктора

из природных оксидов кремния — песка или кварца, восстановлением в пламени электрической дуги.

Технический кристалл устанавливается вертикально в камере, заполненной инертным газом или откачанной до глубокого вакуума. Вместо мощного высокочастотного индуктора для создания зоны расплава в некоторых современных установках используется луч мощного лазера. Образовавшаяся капля удерживается силами поверхностного натяжения Si. Таким образом, никакого сосуда (тигла), удерживающего жидкий кремний, не требуется. Эта особенность метода, подчеркнутая в его названии, позволяет избежать загрязнения кремния примесями, неминуемо выделяющимися из стенок тигля при очень высокой (1417 °C) температуре плавления Si.

Методом зонной плавки удается очистить кремний до концентрации $\sim 10^{12}$ атомов примеси в 1 см^3 , т. е. до уровня, соответствующего 1 атому примеси на 10^6 миллиардов атомов Si. И при этом речь идет не об искусстве выращивать уникальные кристаллы, а о промышленной технологии, ориентированной на ежегодный выпуск десятков тонн такого сверхчистого кремния.

Возвращаясь к табл. 1, можно убедиться, что даже в сверхчистом кремнии с содержанием примесей, не превышающим 10^{12} см^{-3} , концентрация примесей приблизительно в 100 раз больше, чем концентрация собственных электронов и дырок при комнатной температуре. Один из самых очевидных выводов, вытекающих из такого сопоставления, состоит в том, что для понимания реальных свойств полупроводников необходимо изучать полупроводники, содержащие примеси.

Примесные полупроводники



Донорная примесь. Начнем с самого простого примера. Пусть в кристалл кремния каким-то образом попал атом мышьяка (As) и занял место в одном из узлов кристаллической решетки, заместив в нем законного хозяина — атом Si. Атом Si имеет четыре валентных электрона, атом As — пять. Четыре валентных электрона мышьяка будут участвовать в связях с соседними атомами кремния. Пятый электрон будет удерживаться атомом As, но гораздо слабее, чем остальные четыре, прочно «связанные» в электронных орбитах, определяемых всей структурой кристалла кремния. Энергия $\Delta \mathcal{E}$, которая нужна, чтобы разорвать связь этого пятого электрона с атомом мышьяка и превратить его в свободный электрон, гораздо меньше, чем энергия \mathcal{E}_g , которую необходимо затратить, чтобы разорвать связь между атомами кремния и образовать электронно-дырочную пару.

Примесь, атомы которой легко отдают электроны, называют донорной. Латинский глагол *donare*, от которого произошло слово «донор», означает дарить, жертвовать.

Мышьяк в кристалле кремния является донорной примесью. Пусть в каждый кубический сантиметр кристалла кремния введено N_d атомов мышьяка. Рассмотрим сначала самую простую ситуацию, когда температура (в кельвинах) кристалла $T = 0$. Ясно, что в этом случае кристалл представляет собой идеальный изолятор: как ни мала энергия $\Delta \mathcal{E}_v$, требуемая для разрыва связи между атомом мышьяка и пятым его электроном, при абсолютном нуле температуры ей взяться неоткуда.

Если температура кристалла отлична от нуля, то равновесная концентрация «примесных» электронов n_d определится из выражения, аналогичного формуле (6):

$$n_d \approx (AN_d)^{1/2} e^{-\Delta \mathcal{E}_v / 2kT}. \quad (7)$$

Здесь A — та же величина, что и в формуле (6). Как мы помним, при комнатной температуре для различных полупроводников ее значение лежит в пределах $10^{17} - 10^{19} \text{ см}^{-3}$. Из формулы (7) видно, что чем больше концентрация введенной в полупроводник донорной примеси N_d , тем больше, при той же температуре, концентрация примесных электронов. Так оно, естественно, и должно быть.

В формуле (7) в показателе экспоненты стоит величина $\Delta \mathcal{E}_v$, а не \mathcal{E}_g , как в формуле (6), определяющей концентрацию собственных носителей. Как мы знаем, $\Delta \mathcal{E}_v \ll \mathcal{E}_g$, а мы уже убедились раньше, как «чувствительна» экспонента к величине показателя.

Энергия ионизации $\Delta \mathcal{E}_v$ (иногда ее называют энергией активации примеси), например, для мышьяка, введенного в кремний, равняется всего лишь 0,05 эВ. В 20 раз меньше, чем энергия \mathcal{E}_g , требуемая для создания электронно-дырочной пары в кремнии!*) В соответствии с формулой (7) (и со здравым смыслом) это означает, что «ионизация» атомов мышьяка, т. е. отрыв «лишнего» пятого электрона начнется при гораздо более низких температурах, чем генерация электронно-дырочных пар. Поэтому при низких температурах примесных носите-

*) Примесь, энергия активации которой $\Delta \mathcal{E}_v$ значительно меньше, чем энергия образования электронно-дырочной пары \mathcal{E}_g , принято называть «мелкой».

лей в полупроводниковом материале, как правило, значительно больше, чем собственных.

Пусть, например, в кремний введена донорная примесь (As) и концентрация примесных атомов $N_d = 10^{15} \text{ см}^{-3}$. (Такое значение N_d типично для многих практически важных случаев.) Для кремния $A \approx B \approx 10^{19} \text{ см}^{-3}$. Используя формулы (6) и (7), легко подсчитать, что при температуре 10 К концентрация примесных электронов составит $n_d \approx 2,5 \cdot 10^4 \text{ см}^{-3}$, а концентрация собственных электронов так мала, что даже в кристалле размером с Галактику не нашлось бы ни одного собственного носителя.

При температуре 50 К величина n_d равняется $\sim 3 \cdot 10^{14} \text{ см}^{-3}$, тогда как концентрация собственных электронов все еще практически равна нулю в любом реальном кристалле. Заметим, что при этой температуре около 1/3 введенных в кристалл атомов примеси оказываются ионизованными.

А при комнатной температуре?

Примесное истощение. А при комнатной температуре формулу (7) использовать для расчетов концентрации примесных электронов нельзя. Дело в том, что формула (7) является не точной, а приближенной. Ею можно пользоваться только тогда, когда рассчитанное по ней значение n_d много меньше, чем концентрация введенных атомов примеси N_d , т. е. когда $kT \ll \Delta \mathcal{E}$.

Если же температура T так велика, что $kT \sim \Delta \mathcal{E}$ (и тем более, если $kT > \Delta \mathcal{E}$), то все примесные атомы оказываются ионизованными и концентрация примесных электронов просто равняется N_d . Это явление носит название «примесное истощение» — термин довольно естественный. При достаточно высокой температуре все атомы примеси отдали «лишние» электроны и источник примесных электронов «истощился». Дальнейший нагрев кристалла не приводит к росту концентрации примесных электронов.

В Si комнатной температуре (300 К) соответствует область примесного истощения. Таким образом, в рассматриваемом нами примере при 300 К $n_d = N_d = 10^{15} \text{ см}^{-3}$. Концентрация же собственных электронов n_i при 300 К равняется всего $\sim 10^{10} \text{ см}^{-3}$ (см. табл. 1) — в 100 000 раз меньше.

А вот при 1000 К ситуация будет совсем иной. Концентрация примесных электронов составит по-прежнему

$\sim 10^{15} \text{ см}^{-3}$, а концентрация собственных, рассчитанная по формуле (6), будет равняться $n_i \sim 1,5 \cdot 10^{16} \text{ см}^{-3}$. Собственных электронов в Si при 1000 К окажется в 15 раз больше, чем примесных.

Сравнивая формулы (6) и (7), можно убедиться, что чем больше энергия \mathcal{E}_g по сравнению с энергией ионизации примеси $\Delta\mathcal{E}$ и чем больше примесных атомов содержит кристалл, тем выше температура, при которой собственная проводимость будет преобладать над примесной.

Зависимость концентрации электронов от температуры. Взгляните на рис. 5. Кривая, которая на нем показана, суммирует многое из того, что мы узнали о свойствах собственных и примесных полупроводников. Эта кривая представляет собой типичную зависимость равновесной концентрации свободных электронов в полупроводнике от температуры.

Правда, для этой зависимости выбраны несколько необычные координаты: по оси ординат откладывается не концентрация n , а величина ее натурального логарифма $\ln n$. По оси абсцисс — не температура, а величина, обратная температуре, — $1/T$. «Читать» такой график непривычно, особенно потому, что продвижению по оси абсцисс вправо соответствует не увеличение, а уменьшение температуры.

(Значение $1/T$ растет — значение T падает.) Почему для этой зависимости выбраны именно такие координаты, мы сейчас обсудим. А пока просто «прочитаем» график качественно. Начнем с области низких температур, т. е. с правой части графика, где значение $1/T$ велико, и будем двигаться по оси абсцисс справа налево.

При низких температурах, как мы видели, концентрация электронов в полупроводнике определяется концентрацией примесных центров. С ростом температуры примесная концентрация растет и на этом участке кривой зависимость $n(T)$ определяется формулой (7). При некоторой температуре концентрация электронов перестает зависеть от температуры. Это область примесного истощения. Все атомы примеси уже ионизованы,



Рис. 5. Зависимость логарифма концентрации электронов от обратной температуры

а собственная концентрация все еще гораздо меньше, чем примесная. Наконец, в области еще более высоких температур начинается очень резкий рост концентрации с дальнейшим повышением температуры. Это область собственной проводимости, где зависимость $n(T)$ определяется формулой (6).

А теперь несколько слов о том, почему для зависимости, показывающей, как связана концентрация электронов в полупроводнике с температурой, выбраны именно такие координаты. Прологарифмируем формулу (6):

$$\ln n_i = \ln A - \frac{\mathcal{E}_g}{2k} \cdot \frac{1}{T}, \quad (8)$$

и обозначим $x = 1/T$, $y = \ln n_i$, $a = \ln A$, $b = \mathcal{E}_g/2k$. Тогда формула (8) приобретет знакомый вид $y = a - bx$ — уравнение прямой линии. Итак, если по оси ординат откладывать $\ln n$, а по оси абсцисс $1/T$, то до тех пор, пока справедливо уравнение (6), зависимость $\ln n(1/T)$ должна представлять собой прямую линию. И что особенно важно, тангенс угла наклона этой прямой $b = \mathcal{E}_g/2k$ просто пропорционален важнейшему параметру полупроводника — \mathcal{E}_g . В ранних работах по полупроводникам такой способ определения \mathcal{E}_g использовался очень часто. Его и сейчас еще иногда используют при изучении свойств новых полупроводников.

Если такую же операцию логарифмирования проделать с уравнением (7), то легко убедиться, что тангенс угла наклона прямой $\ln n(1/T)$ пропорционален величине $\Delta\mathcal{E}$. Значит, построив зависимость концентрации электронов от температуры в координатах $\ln n(1/T)$ в области примесной проводимости, мы легко можем определить величину энергии активации примеси $\Delta\mathcal{E}$.

Дырки в электронном полупроводнике. Полупроводник, в который введена донорная примесь, называют электронным или полупроводником n -типа. Термин «электронный полупроводник» вполне понятен. Что может быть естественней, чем называть так полупроводник, в котором почти все свободные носители тока — электроны? Что касается полупроводника n -типа, то латинская буква « n » — от слова *negativus* (отрицательный) — опять-таки напоминает, что такой полупроводник содержит много отрицательно заряженных частиц — электронов.

Однако рассказ о полупроводниках n -типа будет не полон, если мы забудем поинтересоваться, что произойдет в полупроводнике n -типа с дырками, которые в электронном полупроводнике принято называть *основными* носителями тока. (Электроны, как нетрудно догадаться, в этом случае называются *основными* носителями.)

Прежде всего нужно отметить, что когда атом донора ионизуется и отдает электрон, никакой дырки не образуется. Атом мышьяка в решетке Si, например, отдав пятый валентный электрон, остается связанным с соседними атомами кремния четырьмя «полноценными» связями. Нет никакой свободной связи, на которую могли бы перемещаться электроны с соседних орбит, поэтому дырка не образуется.

На первый взгляд может показаться, что дырок в электронном полупроводнике будет тогда столько же, сколько и в совершенно чистом, собственном. Однако этот первый взгляд неверен. Действительно, при данной температуре T каждую секунду в электронном полупроводнике рвется электронных связей между атомами и образуется электронно-дырочных пар ровно столько же, сколько и в собственном. Но гибнут дырки гораздо чаще. Ведь в электронном полупроводнике свободных электронов гораздо больше, чем в собственном. Значит, и встреча дырки со свободным электроном, встреча, которая заканчивается рекомбинацией (и исчезновением дырки), будет происходить гораздо чаще. В результате дырок в полупроводнике n -типа оказывается меньше, чем в собственном полупроводнике. Существует очень простая формула, связывающая равновесную концентрацию дырок p и электронов n в полупроводнике:

$$p \cdot n = n_i^2 (T). \quad (9)$$

Чем больше электронов, тем меньше дырок (и наоборот). Величина n_i в формуле (9) представляет собой значение собственной концентрации носителей при данной температуре T и определяется выражением (6).

Пользуясь формулами (9) и (6), легко подсчитать, что в кремнии n -типа, например, при комнатной температуре и концентрации электронов 10^{15} см^{-3} концентрация дырок составит $\sim 10^5 \text{ см}^{-3}$ — в десять миллиардов раз меньше, чем электронов! Понятно, какие носители здесь основные, а какие — неосновные.

Наше знакомство с примесными полупроводниками далеко не закончено. Мы познакомились с электронным полупроводником, в который введена донорная примесь. Простое чувство справедливости и настоятельная практическая необходимость требует ознакомления и с другой ситуацией, а именно: в полупроводник введена примесь, которая не отдает (как донорная), а легко отбирает электроны у соседних атомов.

Акцепторная примесь*). На рис. 6 показана кристаллическая решетка кремния, в которой один из узлов занят примесным атомом бора. Бор (В) трехвалентен — на внешней электронной оболочке у него находятся три электрона. Чтобы образовать полноценную связь с соседними атомами кремния, одного электрона ему не хватает.

Сравните рисунок электронных связей вокруг атома В на рис. 6 с конфигурацией электронных связей вокруг атома 19 на рис. 3. Не правда ли, очень схожие ситуации?

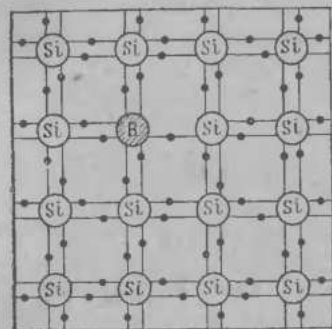


Рис. 6. Атом акцептора (бор) в решетке кремния

И атому кремния 19 на рис. 3, и атому В на рис. 6 не хватает одного электрона. Но имеется и глубокое различие. Ведь все атомы кремния совершенно одинаковы и пустая связь — дырка, которая в данный момент принадлежит атому 19, в следующий момент может оказаться у атома 14, затем у атома 9 и т. д. Для того чтобы дырка перемещалась по кристаллу, никакой энергии не требуется. Атом бора — чужак в решетке кремния.

И чтобы электрон от соседнего атома кремния перешел к бору, необходимо затратить энергию $\Delta\mathcal{E}$. Эта энергия — энергия активации — невелика (для атома В в Si она составляет всего 0,045 эВ), но все-таки не равна нулю, как в случае дырки на рис. 3. Пока этот, пусть небольшой «энергетический барьер» не перейдет, дырка в кристалле не образуется.

Предположим, однако, что тепловые колебания решетки или квант света сообщили необходимую энергию и электрон с соседнего атома перешел к бору. Вот теперь ситуация абсолютно ничем не отличается от той, которая показана на рис. 3. Пустая связь образовалась у атома кремния, и возникла дырка — свободный носитель положительного заряда, способный переносить электрический ток. Акцепторная примесь создает в кристалле полупроводника свободные носители тока — дырки.

*) Латинский глагол *acceptare* означает принимать.

Если температура кристалла отлична от нуля, то равновесная концентрация примесных дырок p_a определится из выражения, аналогичного формуле (6):

$$p_a = (B \cdot N_a)^{1/2} e^{-\Delta \mathcal{E}/2kT}. \quad (10)$$

Величина B здесь та же, что в формуле (6), N_a — концентрация введенных в кристалл акцепторов.

Ситуации с донорной и акцепторной примесью очень похожи. Поскольку величина $\Delta \mathcal{E}$ гораздо меньше, чем энергия образования электронно-дырочной пары \mathcal{E}_g (кстати, обратите внимание, что энергии активации донорной примеси — мышьяка и акцепторной — бора очень близки), то при относительно низких температурах экспонента в формуле (10) будет гораздо больше, чем экспонента в формуле (6). Это значит, что концентрация примесных дырок p_a будет значительно больше, чем концентрация собственных. При более высоких температурах, когда $kT \sim \Delta \mathcal{E}$, наступит явление примесного истощения. Наконец, при еще более высоких температурах собственная концентрация, определяемая формулой (6), превысит примесную.

Появление дырки в примесном полупроводнике не сопровождается появлением электрона проводимости, — это ясно видно из рис. 6. Наоборот, в дырочном полупроводнике (или проводнике p -типа) увеличение концентрации дырок сопровождается уменьшением концентрации электронов. Количественное соотношение между концентрациями основных и неосновных носителей определяется формулой (9) и одинаково справедливо и для n -, и для p -типа полупроводника. Но, разумеется, в полупроводнике p -типа основными носителями будут дырки, а неосновными — электроны.

Послушайте теперь ту же мелодию
в исполнении фагота



Много лет назад, в добрые старые времена, когда магнитофоны были редкостью, и в еще более добрые и более старые времена, когда их вообще не существовало, новые песни разучивались с помощью радио так: сначала песня исполнялась певцом или певицей в сопровождении хора и оркестра. Потом очень медленно, с повторениями, диктовались слова. Затем диктор говорил:

«А теперь послушайте ту же песню в исполнении солиста». Солист пел. Затем — в исполнении хора. Потом — скрипки. Затем можно было услышать ту же мелодию в исполнении фортепьяно, кларнета, а иногда даже и фагота. Это делалось для того, чтобы можно было лучше понять и запомнить мелодию.

И мы сейчас займемся очень похожим делом. Изложим кое-что из того, что нам уже известно, на другом языке, языке *энергетических* диаграмм. Язык энергетических диаграмм широко используется при анализе полупроводниковых приборов. Он позволяет в ясной и удобной форме описать очень сложные ситуации и иногда просто незаменим.

Но привыкать к этому новому языку, как и к любому другому, следует постепенно. В этом параграфе мы изложим на языке энергетических диаграмм самые простые понятия. Тем не менее советуем быть внимательными: когда инструмент энергетических диаграмм начнет использоваться для исполнения сложных «мелодий», привыкнуть к нему будет гораздо труднее.

Мы знаем, что для того чтобы в собственном полупроводнике возникли электрон и дырка, необходимо затратить энергию \mathcal{E}_g . Изобразим процесс возникновения электронно-дырочной пары схематически (рис. 7, а). Будем считать, что пока электрон находится на орбите, связывающей между собой атомы кремния (рис. 2), он имеет энергию \mathcal{E}_v *).

Чтобы создать электрон проводимости и дырку (образовать электронно-дырочную пару), необходимо затратить энергию или, другими словами, преодолеть энергетический барьер высотой \mathcal{E}_g . Энергия \mathcal{E}_g может быть получена от тепловых колебаний решетки или от достаточно энергичного кванта света (фотона). Если энергетический барьер преодолен, электрон попадает на уровень \mathcal{E}_c **). Возникает электрон проводимости (черная точка на рис. 7, а) и дырка (белый кружок со знаком «+»).

Энергетическая диаграмма делает очень наглядным понятие об энергетическом барьере — порции энергии, необходимой для протекания какого-либо процесса.

*) Индекс «v» — от латинского слова *valentia* (сила). Он напоминает, что речь идет о валентных электронах.

***) Индекс «c» — от английского слова *conductivity* (проводимость).

Так же просто представить себе с помощью энергетической диаграммы процесс рекомбинации электрона и дырки в собственном полупроводнике. Этот процесс показан на рис. 7, *а* штриховой стрелкой. При рекомбинации выделяется та же энергия \mathcal{E}_g , которая некогда

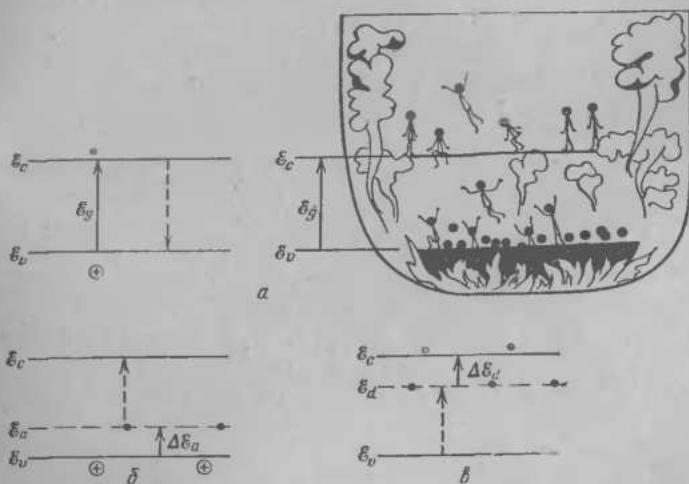


Рис. 7. Энергетические диаграммы, поясняющие процесс образования электронов и дырок в полупроводнике: *а* — образование электронно-дырочной пары в собственном полупроводнике. Только очень немногим электронам под влиянием теплового движения удается разорвать мощные цепи электрического поля, удерживающего валентные электроны на межатомных орбитах (на уровне ε_v). Эти счастливицы попадают в число электронов проводимости. Их концентрация пропорциональна $\exp(-\mathcal{E}_g/kT)$; *б* — образование дырок в полупроводнике, содержащем атомы акцептора; *в* — образование электронов в полупроводнике, содержащем донорную примесь

была затрачена на создание электронно-дырочной пары. Иногда рекомбинация сопровождается рождением фотона с энергией \mathcal{E}_g . Но часто электрон и дырка рекомбинируют, не порождая фотона. Тогда энергия \mathcal{E}_g передается решетке кристалла, нагревая ее.

Рис. 7, *б* иллюстрирует рождение дырки в кристалле, содержащем акцепторные центры. Уровень энергии ε_a , соответствующий акцепторному центру при $T = 0$, пуст, электронов на нем нет. (Помните? — «... как ни мала энергия активации $\Delta\mathcal{E}$, требуемая, чтобы перевести электрон от атома Si к атому В и образовать дырку,

при абсолютном нуле температуры ей взятая неоткуда...») При заданной температуре T , такой, что $kT \ll \ll \Delta \mathcal{E}_a$, концентрация электронов на уровне \mathcal{E}_a определяется формулой (10).

Но ведь формула (10) написана вовсе не для концентрации электронов, а для концентрации дырок? Правильно. Но дырка-то в кристалле p -типа и образуется потому, что электрон с уровня \mathcal{E}_v захватывается на акцепторный уровень \mathcal{E}_a . Сколько электронов захвачено на акцепторный уровень, столько подвижных дырок возникло в кристалле.

Атом акцептора, захвативший электрон, заряжается отрицательно. Удерживается атомом акцептора этот связанный, не способный переносить ток электрон очень прочно. Как ясно видно из рис. 7, а, чтобы захваченный на акцепторный уровень электрон сделать свободным (процесс «освобождение электрона» — перевод его на уровень \mathcal{E}_c — показан на рис. 7, б штриховой стрелкой), необходима энергия $\mathcal{E} = \mathcal{E}_c - \mathcal{E}_a = |\mathcal{E}_g - \Delta \mathcal{E}_a|$. Если $\Delta \mathcal{E}_a \ll \mathcal{E}_g$ (мелкий центр), то энергия, необходимая для освобождения захваченного акцептором электрона, почти так же велика, как и энергия \mathcal{E}_g .

При величине kT порядка, а тем более большей, чем значения $\Delta \mathcal{E}_a$, все акцепторные центры будут заняты электронами, а концентрация дырок будет равна N_a — концентрации атомов акцептора, введенных в кристалл (примесное истощение).

На рис. 7, в показана энергетическая диаграмма электронного полупроводника, содержащего мелкую донорную примесь. При $T = 0$ концентрация электронов проводимости на уровне \mathcal{E}_c равна нулю. Уровень энергии \mathcal{E}_d , соответствующий донорному центру, полностью заполнен электронами. При этом атомы донора, удерживающие электроны, нейтральны. При не слишком высокой температуре T концентрация электронов на уровне \mathcal{E}_c определяется формулой (7). Атомы доноров, отдавшие электроны на уровень \mathcal{E}_c , заряжены положительно. Концентрация этих положительно заряженных донорных центров, естественно, также определяется формулой (7). При $kT \gtrsim \Delta \mathcal{E}_d$ концентрация свободных электронов на уровне \mathcal{E}_c и концентрация положительно заряженных донорных центров равняются N_d (примесное истощение).

Положительно заряженный донорный центр может захватить электрон с уровня \mathcal{E}_v . При этом в кристалле

возникнет дырка. Но чтобы перевести электрон с уровня \mathcal{E}_v на донорный уровень \mathcal{E}_d и таким образом создать в кристалле дырку (этот процесс показан на рис. 7, *в* штриховой стрелкой), необходимо преодолеть энергетический барьер высотой $\mathcal{E}_g - \Delta\mathcal{E}_d$ *).

Обратим внимание на одно важное обстоятельство, особенно наглядно иллюстрируемое энергетическими диаграммами. Если электроны не подвергаются воздействию каких-либо внешних возмущений, то они находятся на самом нижнем из возможных энергетических уровней.

Действительно, в собственном полупроводнике при $T = 0$ и в отсутствие освещения, как мы знаем, нет ни электронов, ни дырок. На языке энергетической диаграммы (рис. 7, *а*) это означает, что все электроны находятся на самом нижнем из возможных энергетических уровней \mathcal{E}_v .

В полупроводнике, содержащем акцепторные центры, при $T = 0$ и в отсутствие освещения все электроны также находятся на уровне \mathcal{E}_v . Уровни \mathcal{E}_a и \mathcal{E}_c (рис. 7, *б*) пусты.

В полупроводнике, содержащем доноры (рис. 7, *в*), при $T = 0$ электроны, как мы знаем, находятся на уровне \mathcal{E}_d . А почему не на уровне \mathcal{E}_v !?

Дело в том, что электроны стремятся занять самый низкий из *возможных* энергетических уровней. В кристалле полупроводника, содержащем доноры, при $T = 0$ уровень \mathcal{E}_v полностью заполнен электронами. Свободных мест на нем нет. И электроны занимают самый низкий из возможных уровней — уровень \mathcal{E}_d . Вспомним еще раз, что это означает физически. Уровень \mathcal{E}_v соответствует энергии электронов, связывающих между собой атомы кремния. При $T = 0$ все электроны находятся на своих орбитах (рис. 2) и свободных мест (дырок), куда могли бы поместиться электроны с донорного уровня \mathcal{E}_d , попросту нет.

Компенсация. Овладев на языке энергетических диаграмм несколькими фразами, решим задачу, ответ для которой может быть получен на этом языке буквально за несколько секунд и к тому же в очень наглядной и запоминающейся форме. Между тем, чтобы получить

*) Переведите эту фразу с языка энергетической диаграммы на язык физических представлений. Если задача покажется сложной, перечитайте выше пункт «Дырки в электронном полупроводнике».

ответ традиционным путем, необходимо провести довольно тонкое физическое рассуждение.

Задача формулируется следующим образом. Пусть в полупроводник одновременно введено строго одинаковое количество мелкой донорной и акцепторной примесей. Например, в кремний в строго одинаковых концентрациях введен мышьяк и бор. Что произойдет?

Естественный ответ: «В полупроводнике возникает большая концентрация свободных носителей обоих знаков — электронов и дырок».

Этот ответ совершенно неправилен. Правильный ответ состоит в том, что в полупроводнике не будет ни примесных электронов, ни примесных дырок — он будет вести себя как собственный полупроводник.

Взглянем на рис. 8, а. Все электроны с донорного уровня \mathcal{E}_d перешли на первоначально свободный нижний акцепторный уровень \mathcal{E}_a . Теперь, чтобы создать

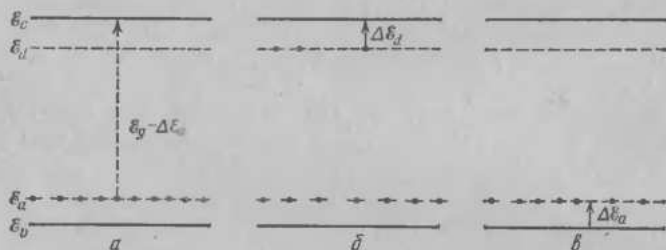


Рис. 8. Компенсация. а — Полная компенсация: концентрация доноров N_d равна концентрации акцепторов N_a . б, в — Частичная компенсация: $N_d > N_a$ (б); $N_d < N_a$ (в)

в кристалле свободный электрон, нужно затратить энергию $\mathcal{E}_g - \Delta\mathcal{E}_a$, почти такую же большую, как и энергия \mathcal{E}_g , необходимая для создания электронно-дырочной пары в собственном полупроводнике. И дырке в кристалле появиться трудно: все свободные места на акцепторном уровне забиты электронами с донорного уровня \mathcal{E}_d . Для электронов с уровня \mathcal{E}_v места на акцепторном уровне нет. Чтобы в кристалле могла появиться дырка, нужно или прямо забросить электрон с уровня \mathcal{E}_v на уровень \mathcal{E}_c (для чего необходима энергия \mathcal{E}_g), или освободить место на уровне \mathcal{E}_a . Для этого надо перебросить электрон с уровня \mathcal{E}_a на уровень \mathcal{E}_c (необходимая энергия $\mathcal{E}_g - \Delta\mathcal{E}_a$ почти так же велика, как и \mathcal{E}_g). Один взгляд на рис. 8, а позволяет понять, что

полупроводник, в который мелкая донорная и акцепторная примеси введены в равной концентрации, будет вести себя почти как собственный. Рассмотренное явление носит название *компенсации*. Введение акцепторной примеси компенсирует действие донорной и наоборот.

Какой физической картине соответствует энергетическая диаграмма рис. 8, а? Давайте рассуждать.

Введенные в кристалл атомы донорной примеси легко отдадут лишние электроны, которые станут свободными. Атомы акцепторной примеси захватят электроны у своих ближайших соседей по решетке и образуют соответствующее количество дырок. Примесные электроны и дырки начнут хаотически блуждать по кристаллу и рано или поздно встретятся и рекомбинируют. В соответствии с условием рассматриваемой задачи количество электронов, которые легко могут стать свободными, в точности равно количеству дырок (поскольку концентрация мелких доноров равна концентрации акцепторов). После того как примесные электроны и дырки встретятся, рекомбинируют и исчезнут, в кристалле не останется больше атомов, которые могут легко образовать электрон или дырку. В самом деле, если рассматривать пример с кремнием, в который введены мышьяк и бор, то у каждого атома мышьяка осталось теперь только по четыре валентных электрона. И эти оставшиеся электроны атом мышьяка удерживает почти так же прочно, как и атом кремния свои четыре валентных электрона. То же можно сказать и об атоме бора. Присоединив к себе четвертый электрон, он образовал с соседними атомами Si почти такие же прочные связи, как и связи между атомами кремния. В результате, чтобы образовать электрон проводимости или дырку, нужно теперь затратить энергию, почти равную энергии образования электронно-дырочной пары в собственном кремнии.

А что произойдет, если концентрации донорной и акцепторной примесей не равны? Рис. 8, б, в позволяют ответить на этот вопрос.

Рис. 8, б соответствует случай, когда донорной примеси больше, чем акцепторной, и донорная примесь *компенсирована частично*. Акцепторный уровень полностью «выведен из строя» (примесные дырки образоваться не могут). А на донорном уровне \mathcal{E}_d еще остались электроны. Их концентрация равняется, очевидно, $N_d - N_a$. Полупроводник будет вести себя как полупроводник *n*-типа, концентрация доноров в котором

равняется $N_d - N_a$. Рис. 8, в иллюстрирует случай, когда $N_a > N_d$. Донорный уровень полностью опустошен, а на акцепторном уровне еще остались свободные места — дырки. Их концентрация равняется $N_a - N_d$. И полупроводник будет вести себя как полупроводник p -типа с концентрацией акцепторов $N_a - N_d$.

Энергетические диаграммы помогут нам разобраться в проблемах генерации, рекомбинации и движения свободных носителей, которые будут рассмотрены в следующей главе. Но полностью значение диаграмм выявится во второй части книги, когда мы будем изучать энергетические барьеры, образующиеся на границах между полупроводником и окружающей средой или на границе между полупроводниками n - и p -типа.

Moralite *)

Полупроводники — это материалы, проводимость которых при абсолютном нуле температуры равна нулю, так как валентные электроны связаны на межатомных орбитах и не способны переносить электрический ток. Однако энергия связи электронов невелика, и при относительно невысоких температурах за счет теплового движения электронные связи разрываются. Возникают свободные электроны (электроны проводимости) и дырки. Концентрация электронов проводимости и дырок экспоненциально возрастает с ростом температуры. Соответственно экспоненциально увеличивается с температурой проводимость полупроводников.

Некоторые примеси (доноры) способны легко отдавать электроны, а другие (акцепторы) — захватывать электроны у атомов полупроводника. Введение таких примесей, даже в ничтожном количестве, способно в громадной степени увеличивать число свободных носителей и, следовательно, проводимость полупроводника.

*) Правлучительное размышление (*фр.*).

Глава 2

РОЖДЕНИЕ, ЖИЗНЬ И СМЕРТЬ ЭЛЕКТРОНА И ДЫРКИ

Природы вековая давящая
Соединяла смерть и бытие
В один клубок, но мысль была
бессильна
Соединить два таинства ее.

Н. Заболоцкий



В предыдущей главе мы познакомились с тем, как рождаются (генерируются) электроны и дырки в собственном полупроводнике и в полупроводнике, содержащем мелкие примесные центры. Нет, однако, никаких оснований ожидать, что любая примесь, введенная в полупроводник, будет вести себя как мелкая, т. е. что энергия активации $\Delta\mathcal{E}$ любого примесного атома во всех случаях будет много меньше, чем энергия \mathcal{E}_g образования собственной электронно-дырочной пары. В этой главе мы изучим, как влияет на процессы генерации и рекомбинации наличие в полупроводнике *глубоких* примесных центров, энергия активации которых сравнима с величиной \mathcal{E}_g . Такие центры играют очень важную роль в работе многих полупроводниковых приборов.

Кроме вопросов, связанных с рождением (генерацией) и смертью (рекомбинацией) электронов и дырок, мы рассмотрим в этой главе основные виды движения носителей: тепловое движение, движение в электрическом поле и диффузию.

Рождение и смерть

Глубокие центры: компенсаторы, ступеньки, убийцы. Для каждого широко используемого полупроводника — кремния, германия, арсенида галлия, фосфида индия, существует длинный список примесей. Для кремния, например, этот список содержит несколько

десятков наименований. Он включает уже знакомые нам мышьяк и бор, но кроме них — фосфор и алюминий, серебро и медь, кадмий и кобальт, золото, железо, кислород, ртуть, платину, молибден, никель, палладий, серу, селен, вольфрам, цинк и многие другие элементы. Каждый из этих элементов характеризуется в качестве примеси в кремнии своей энергией активации $\Delta\mathcal{E}$, и значения этих энергий лежат в пределах от 0,05 эВ (мелкие доноры и акцепторы) до 0,8—0,9 эВ. Последние значения, характерные для самых «глубоких» примесей, уже вполне сравнимы с энергией \mathcal{E}_g в кремнии (1,1 эВ).

Казалось бы, чем «глубже» примесь (т. е. чем больше $\Delta\mathcal{E}$), тем меньшую роль она должна играть. Действительно, из формул (7) и (10) ясно, что с увеличением $\Delta\mathcal{E}$ очень быстро, экспоненциально, падает число свободных электронов и дырок, существующих в кристалле при заданной концентрации легирующей примеси N_d или N_a . Формулы (7) и (10) справедливы для любых значений $\Delta\mathcal{E}$. Однако вывод о том, что на «глубокие» примеси можно не обращать внимания, поспешен и неверен. Глубокие примеси в полупроводниках играют три важнейшие роли — «компенсаторов», «ступенек» и «убийц». Рассмотрим эти роли одну за другой.

Начнем с самой простой и отчасти нам уже знакомой роли *компенсаторов*.

Представим себе, что мы хотим получить полупроводник, например GaAs, с очень высоким удельным сопротивлением. Такая задача не так уж редко встает перед специалистами, изготавливающими полупроводниковые приборы. Пленки или пластины с высоким удельным сопротивлением часто используются в качестве основы, *подложки*, на которой затем наращиваются очень тонкие слои этого же полупроводника с различными легирующими примесями.

Самое высокое удельное сопротивление имеет собственный полупроводник: в нем меньше всего свободных носителей заряда. Однако того, что мы уже знаем о примесях в полупроводниках, вполне достаточно, чтобы оценить трудности получения собственных полупроводников. Ничтожная концентрация мелкой примеси — и... концентрация электронов или дырок возрастает в тысячи, миллионы и даже миллиарды раз. Во всех случаях попытки совершенно избавиться от мелких примесей сопряжены с длительными, дорогими и сложными операциями по очистке материалов. Но ощутимые результаты они при-

посят в настоящее время только для германия и кремния. Для GaAs, например, самая совершенная современная технология обеспечивает уровень концентрации мелкой примеси $N \sim 10^{13} \text{ см}^{-3}$. Эта величина в десять миллиардов раз меньше, чем концентрация атомов галлия и мышьяка в GaAs, однако она все еще в миллион раз больше, чем концентрация собственных электронов и дырок в этом материале при комнатной температуре (см. табл. 1). Как же быть?

А очень просто. Не будем особенно тщательно очищать полупроводник. Пусть в нем содержится 10^{15} или даже 10^{16} см^{-3} мелких доноров. А кроме того, мы... дополнительно введем в полупроводник примесь, и еще в большей концентрации. Скажем, в концентрации 10^{16} или 10^{17} см^{-3} . Но только эта вторая примесь будет глубоким акцептором.

Возникнет эффект, с которым мы уже знакомы, — компенсация. Только не мелкими, а глубокими примесями. Глубокие акцепторы захватят все электроны, созданные мелкими донорами, и теперь освободить электроны

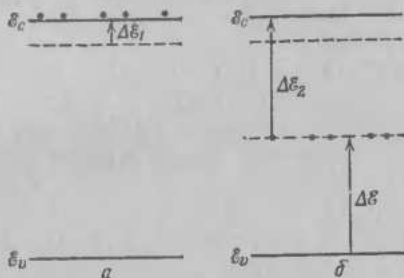


Рис. 9. Компенсация глубоким примесным уровнем. *a* — Полупроводник содержит только мелкую примесь с энергией ионизации $\Delta\epsilon_1$. Температура достаточно велика и все электроны с примесного уровня перешли на уровень ϵ_c (стали свободными). В кристалле много свободных носителей — удельное сопротивление мало. *б* — Кроме мелкого донора в кристалл введен элемент, являющийся глубоким акцептором. Все электроны с мелкого донора захвачены на глубокий акцепторный уровень. Чтобы сделать электроны свободными, нужна большая энергия $\Delta\epsilon_2 \gg kT$. Свободных носителей в кристалле практически нет — удельное сопротивление очень велико

будет не так-то просто. Энергия активации этих электронов с глубоких примесей велика (рис. 9).

Можно, конечно, пытаться компенсировать мелкую донорную примесь мелкой акцепторной (такой случай показан на рис. 8). Но тут нужна предельная тщатель-

ность. Нужно, чтобы величины N_d и N_a равнялись друг другу с большой точностью. Добиться этого очень трудно.

То ли дело глубокая примесь! Даже если ввести ее в несколько раз больше, чем содержится в кристалле мелкой примеси, которую нужно компенсировать, ничего страшного не произойдет. Все носители, созданные мелкой примесью, будут с гарантией захвачены. Сделать эти электроны свободными теперь не так-то просто. Для этого нужна сравнительно большая энергия $\Delta\mathcal{E}_2$ (рис. 9). А как насчет «лишних» акцепторов? Ведь они способны захватывать электроны у атомов полупроводника и образовывать дырки. Способны, но... на то они и глубокие — энергия активации дырки $\Delta\mathcal{E}$ у таких акцепторов велика (рис. 9), и соответственно концентрация дырок окажется очень небольшой.

В случае GaAs именно так и поступают. Берут не особенно тщательно очищенный GaAs и вводят в него в значительной концентрации хром или кислород. Если до введения глубокой примеси GaAs был p -типа, то вводят кислород; в GaAs он является глубоким донором. Если исходный GaAs был n -типа, вводят хром. В GaAs хром является глубоким акцептором. В результате удается получить монокристаллы GaAs с очень высоким удельным сопротивлением — до 10^7 Ом·см. Подложки из такого «полуизолирующего» арсенида галлия широко используются при изготовлении полупроводниковых приборов на основе GaAs.

Итак, в первом из рассмотренных случаев глубокие примеси играют роль положительного героя. Но... «наши недостатки являются продолжением наших достоинств». Эта мудрая поговорка применима не только к людям. В частности, ее полезно вспомнить и при обсуждении свойств глубоких примесей. Представим себе другой, тоже очень распространенный случай. Мы хотим получить не полуизолирующую подложку, а по возможности чистый материал, скажем, n -типа. Мы вводим и вводим в кристалл мелкую донорную примесь, а полупроводник... остается высокоомным. Легко догадаться, в чем здесь дело: в полупроводнике присутствует глубокий акцептор. И пока мы его не «перекомпенсируем», т. е. не введем мелкой примеси больше, чем имеется глубоких акцепторов, повышение уровня легирования ни к чему не приведет.

Между тем полупроводник совсем не безразличен к общему количеству введенной в него примеси. Если примесей ввести слишком много, возникают различные струк-

турные несовершенства, падает скорость движения электронов и дырок в электрическом поле, возникают другие нежелательные явления. Поэтому, если мы хотим получить чистые кристаллы с заданной концентрацией электронов (или дырок), мы в первую очередь должны научиться, когда это необходимо, избавляться от глубоких примесей. Технологи прекрасно справляются с этой задачей в Ge и Si — самых «старых» полупроводниковых материалах. В последние годы большие успехи достигнуты в технологии GaAs. А вот в карбиде кремния, например, минимальный уровень содержания глубоких примесей, которого удастся достигнуть, $\sim 10^{16}$ см⁻³.

Итак, в первой роли, роли *компенсаторов*, глубокие уровни мешают рождению свободных носителей — электронов и дырок.

Во второй своей роли — в роли *ступенек* — глубокие центры выполняют прямо противоположную функцию: помогают рождению дырок и электронов.

Мы знаем, что для того, чтобы в собственном полупроводнике возникли электрон и дырка, необходимо затратить энергию \mathcal{E}_g . Если энергию \mathcal{E}_g нужно получить от тепловых колебаний решетки, а величина \mathcal{E}_g много больше, чем средняя тепловая энергия kT , то такой процесс очень маловероятен. Как мы знаем, он пропорционален $\exp(-\mathcal{E}_g/kT)$ (формула (2)). Из рис. 10, а видно, что если в полупроводнике имеется глубокий примесный центр, то энергию, необходимую для того, чтобы освободить электрон, можно затратить не сразу, а в два этапа. Сначала сообщить энергию $\mathcal{E}_g/2$. При этом электрон перейдет с уровня \mathcal{E}_v на уровень глубокой примеси. Затем сообщить вторую порцию энергии, тоже, естественно, равную $\mathcal{E}_g/2$. Легко понять, что процесс образования электронов и дырок в этом случае будет идти гораздо интенсивнее.



Рис. 10. Глубокие примеси — «ступеньки»

Ситуация очень схожа с изображенной на рис. 10, б. Если требуется преодолеть барьер высотой в 2 м без каких-либо промежуточных ступенек, то на такой подвиг способны лишь несколько десятков человек из нескольких миллиардов жителей нашей планеты. Тот же барьер, разбитый на две метровых ступеньки, способен преодолеть, наверное, каждый десятый.

Заметим, что наиболее эффективен такой двухступенчатый процесс именно для примеси с энергией активации $\mathcal{E}_g/2$.

Действительно, предположим, например, что в полупроводниковом материале имеется глубокий акцепторный уровень. Если энергия активации акцептора меньше, чем $\mathcal{E}_g/2$, то электрон будет сравнительно легко захватываться на него с уровня \mathcal{E}_v , но вот перевести его на уровень \mathcal{E}_c (сделать свободным) будет трудно (рис. 7, б). Полная аналогия с лестницей: если нужно забраться на высоту 2 м и единственная ступенька имеет высоту 30 см, то на эту-то ступеньку взобраться легко, а вот оставшиеся 1 м 70 см преодолеть почти так же трудно, как и исходную высоту. Самый оптимальный вариант — две одинаковые ступеньки (рис. 10, б).

Итак, введение глубоких примесей резко увеличивает скорость генерации электронно-дырочных пар. Но позвольте! Как же быть с тем, что мы уже знаем о свойствах примесей? Как быть с формулой (9), из которой следует, что произведение концентраций электронов и дырок зависит для каждого полупроводника только от температуры, а вовсе не от того, какие в него введены примеси? Как быть с выражениями (7) и (10), из которых следует, что чем больше энергия активации $\Delta\mathcal{E}$, тем меньшую концентрацию электронов (или дырок) будет такая примесь создавать при тех же значениях N_d и T ?

Оснований для беспокойства тем не менее нет никаких. Формулы (7), (9) и (10) справедливы для любых уровней — и глубоких, и мелких. Ведь эти выражения описывают значения концентрации электронов и дырок, которые устанавливаются в результате двух конкурирующих процессов — генерации и рекомбинации.

Введение в полупроводник глубоких центров резко увеличивает скорость генерации электронов и дырок. Но... так же резко увеличивается и скорость рекомбинации через эти примесные центры. В результате формулы (7), (9) и (10) по-прежнему правильно описывают равновесные значения электронов и дырок.

В связи с таким утверждением, естественно, возникают два вопроса. Первый — почему глубокие центры увеличивают скорость рекомбинации электронов и дырок? Второй — если введение глубоких примесей увеличивает в скорость генерации, и скорость рекомбинации электронно-дырочных пар, так что равновесная концентрация электронов и дырок все равно определяется только энергией активации $\Delta\mathcal{E}$, то какое значение имеет, есть в кристалле глубокие центры или нет?

Ответ на первый вопрос будет получен немножко дальше. Там, где мы будем рассматривать выступление глубоких центров в роли «убийц».

Что касается ответа на второй вопрос, то он сводится к следующему: чтобы центры-ступеньки смогли проявить себя, необходимы определенные условия.

До сих пор мы говорили о самой простой — равновесной ситуации, когда число создаваемых тепловым движением носителей в точности равнялось числу носителей, гибнущих за то же время вследствие рекомбинации. В таких условиях проявить себя примеси-ступеньки не в состоянии. Во сколько раз увеличивается, благодаря их присутствию, скорость генерации носителей, во столько же раз растет и скорость рекомбинации. Равновесная концентрация носителей определяется формулами (7), (9) и (10).

Чтобы сделать явной деятельность центров-ступенек, нужно нарушить равновесие между процессами генерации и рекомбинации, создать резко *неравновесную* ситуацию. Удобнее всего это сделать, создав условия, когда процесс генерации будет идти так же, как и раньше в условиях равновесия, а процесс рекомбинации будет полностью устранив. Как этого добиться?

Оказывается, не так уж сложно. Нужно создать в какой-либо области полупроводника сильное электрическое поле. Тогда свободные носители, электроны и дырки, будут очень быстро выноситься из этой области электрическим полем и их концентрация в этой области будет очень маленькой. Раз так, скорость рекомбинации, пропорциональная произведению концентраций электронов и дырок (см. формулу (3)), резко уменьшится — рекомбинировать будет нечему. А процесс генерации? На него электрическое поле влияет слабо. Электрическое поле, способное практически полностью очистить область полупроводника от свободных носителей, на скорость генерации может почти не повлиять.

Электрическое поле, выносящее свободные носители, обуславливает появление тока в цепи. И ток этот тем больше, чем больше свободных носителей генерируется в полупроводнике. Сила тока в цепи прямо пропорциональна скорости генерации электронно-дырочных пар.

Пусть сначала участок сильного поля создан в полупроводнике, в котором нет глубоких центров. Измерим протекающий в таких условиях ток. Затем введем в полупроводник глубокие центры и снова измерим силу тока. Проводя такие эксперименты, можно убедиться, что наличие даже небольшой ($\sim 10^{11} - 10^{12} \text{ см}^{-3}$) концентрации примесей с энергией активации $\Delta\mathcal{E} \approx \frac{\mathcal{E}_g}{2}$ увеличивает ток в тысячи и десятки тысяч раз. С другой стороны, примеси, чья энергия активации заметно отличается от величины $\frac{\mathcal{E}_g}{2}$, влияют на ток гораздо слабее. Мелкие доноры и акцепторы практически не влияют на скорость генерации электронно-дырочных пар.

Итак, в резко неравновесной ситуации, когда процесс рекомбинации электронов и дырок устранен, глубокие центры получают возможность блеснуть в роли *энергетических ступенек*, помогающих рождению электронно-дырочных пар.

В каждом полупроводниковом приборе — диоде или транзисторе — в определенном режиме работы существует область сильного электрического поля. И ток, протекающий через прибор в таком режиме, определяется концентрацией глубоких уровней, введенных в полупроводник.

Обсуждение неравновесной ситуации, в которой глубокие центры выступают в роли ступенек, может подсказать, как создать условия, в которых те же глубокие центры проявят себя в роли *убийц*.

Осветим образец, изготовленный из исследуемого полупроводникового материала, светом, энергия квантов которого $\mathcal{E}_\phi = h\nu$ больше, чем энергия создания электронно-дырочной пары в материале \mathcal{E}_g . В образце возникнут избыточные (по отношению к равновесным, темновым) электроны и дырки. Проводимость образца увеличится.

Выключим свет. Ясно, что через некоторое время концентрация электронов и дырок достигнет равновесного (темнового) значения. Избыточные электроны и дырки рекомбинируют, «вымрут».



Интересно вспомнить, что принцип действия *самых первых* полупроводниковых приборов был основан на изменении проводимости полупроводника под влиянием освещения.

В 1873 г. лондонский инженер У. Смит сообщил о необычном явлении, замеченном его помощником А. Меем. Измеряя сопротивление изготовленной из селена изоляции телеграфного кабеля, Мей заметил, что это сопротивление очень сильно уменьшается, если на селен попадает свет. Изменение сопротивления было заметно даже при освещении материала слабым светом Луны *).

Замечательный немецкий изобретатель Вернер фон Сименс, имя которого носит единица проводимости (сименс), автор десятков изобретений, многие из которых (например, динамомашин) не утратили значения и сегодня, первым нашел практическое применение наблюдению Мея. Уже в 1875 г. на основе селенового светочувствительного элемента он изготовил первый в мире полупроводниковый прибор — *фотометр*.

Принцип устройства фотометра очень прост. Отклонение стрелки гальванометра, включенного в цепь последовательно с селеновым элементом и источником тока, тем больше, чем ярче свет, попадающий на элемент. Таким образом, фотометр позволяет объективно измерять относительную яркость различных источников света. До изобретения Вернера фон Сименса это можно было делать только на глаз.

Может быть, мы лучше оценим значение изобретения фотометра, если вспомним о двух обстоятельствах. Во-первых, о том, что современные фотоэкспонетры, изготавливающиеся сегодня во всем мире в десятках миллионов экземпляров, представляют собой, по существу, тот же фотометр Сименса. А во-вторых, о том, что сравнение яркостей различных звезд и других небесных объектов представляет собой одну из самых важных и самых старых задач древнейшей науки — астрономии. Фотометр Сименса позволил перейти от сравнения яркостей звезд «на глаз» к объективной фотометрии, и благодаря ему был сделан один из важных шагов на пути развития астрономии из науки чисто наблюдательной в науку, где главное место занимают количественные измерения.

Через три года после создания фотометра, в 1878 г., знаменитый американский изобретатель Александр Грэхем Белл, создатель телефона, сообщил, что ему удалось соединить в одном устройстве два замечательнейших изобретения XIX века — телефон и фотометр и создать таким образом новый прибор — *фотофон*, предназначенный для осуществления телефонной связи без проводов.

Первый в истории сеанс беспроводной телефонной связи состоялся между Беллом и его помощником Тейнтером летом 1878 г. Тейнтер находился на крыше Франклинской школы в Вашингтоне, где располагалась «передающая телефонная станция». Устроена она была следующим образом. Солнечный свет фокусировался с помощью линз на небольшое зеркало, соединенное своей тыльной стороной с рупором. Отраженный от зеркала луч направлялся в окно последовательской лаборатории Белла, находившейся

* Так через 56 лет после открытия селен совершенно неожиданно оправдал свое небесное имя, данное ему шведским химиком Я. Берцеллиусом. Открыв в 1817 г. новый элемент, Берцеллиус назвал его в честь Луны (по-гречески — Селена),

напротив на расстоянии 213 м. Здесь располагалась «приемная станция», содержащая селеновый фотозлемент, соединенный с батареей (элементом Даниэля) и телефоном. Подойдя к рупору, Тейнтер произнес: «Мистер Белл, если Вы слышите меня, подойдите к окну и помащите шляпой». Вибрация рупора передалась зеркалу, интенсивность отраженного луча стала изменяться «в такт» со словами



Тейнтера, селеновый элемент преобразовал световые колебания в электрические... Белл подошел к раскрытому окну и радостно замахал шляпой.

Сенсация, вызванная изобретением фотофона, была громкой, но недолгой. Обычный телефон оказался проще, надежнее, дешевле и надолго вытеснил фотофон из практической жизни. Надолго, но не навсегда.

В наши дни идея фотофона возрождена на совершенно новом техническом уровне и получила воплощение в системах волоконно-оптической связи. Так же, как и в фотофоне Белла, сигнал от передающей станции к приемной передается с помощью луча света. Но свет теперь проходит не по воздуху, где он сравнительно быстро затухает, а по оптическому волокну, в котором он может передаваться не ослабевая на очень большие расстояния. Источником света служат не лучи солнца, а полупроводниковые лазеры. Интенсивность света меняется с частотой миллиарды раз в секунду с помощью полупроводниковых модуляторов, а электрические сигналы на передающих и приемных станциях обрабатываются с помощью быстродействующих ЭВМ. Сейчас в промышленно развитых странах проложены уже десятки тысяч километров кабелей волоконно-оптической связи, обслуживающих в год сотни миллионов телефонных разговоров.

А теперь возвратимся к примесям-убийцам. Мы остановились на том, что созданные светом избыточные носители тока после выключения света вымирают. Почему?

Понятно почему — равновесие между генерацией и рекомбинацией после выключения света будет нарушено. Процессы генерации будут создавать в единице объема

в единицу времени определенное число носителей. В равновесии это число в точности равно числу носителей, гибнущих за счет рекомбинации. Но сразу после выключения света в образце существуют избыточные носители: электронов и дырок больше, чем в равновесном состоянии. А скорость рекомбинации пропорциональна числу электронов и дырок в единице объема (формула (3)). Поэтому электроны и дырки будут гибнуть быстрее, чем создаваться: концентрация носителей будет уменьшаться, пока не достигнет равновесного значения, в котором скорости процессов генерации и рекомбинации сравняются.

Время возвращения концентрации к равновесному значению определяется *временем жизни* τ избыточных носителей заряда. Чем скорее концентрация возвращается к темновому значению, чем меньше время жизни, тем быстрее, следовательно, идет рекомбинация.

Измеряя время жизни τ в образцах с различными примесями, можно изучить, как влияют на скорость рекомбинации те или иные примесные центры.

Введение мелких примесных центров будет увеличивать равновесную концентрацию носителей в образце, но практически никак не повлияет на время жизни.

А вот введение глубоких примесей, энергия активации которых $\Delta\mathcal{E}$ близка к $\mathcal{E}_g/2$, немедленно изобличит в них «убийц». Эти примеси, даже в ничтожной концентрации, могут уменьшить время жизни в сотни и тысячи раз. Так, введение золота или платины в кремний в концентрации $\sim 10^{13} \text{ см}^{-3}$ (0,000001%) ускоряет гибель неравновесных носителей в тысячу раз: время жизни неравновесных носителей в кремнии τ уменьшается с $\sim 10^{-3}$ до 10^{-6} с.

Однако даже полностью изобличив убийц, хороший следователь не считает дело законченным. Необходимо еще ответить на вопрос «почему». Почему совершено преступление? Почему именно примеси с большой энергией активации служат активными центрами рекомбинации?

Представим себе электрон и дырку, блуждающие по кристаллу. Чтобы встретиться, рекомбинировать и исчезнуть, им необходимо оказаться совсем рядом, в окрестности одного и того же атома кристаллической решетки. Такая ситуация, конечно, возможна, но осуществляется относительно редко.

А теперь представим себе, что в кристалле имеется примесный центр с большой энергией активации $\Delta\mathcal{E}$. Если в окрестности этого центра появится электрон, он

будет захвачен на примесный центр. Если бы это был центр мелкий ($\Delta\mathcal{E} \lesssim kT$), тепловое движение почти немедленно выбросило бы захваченный электрон. При более глубоком центре (большие значения $\Delta\mathcal{E}$) возможна любопытная ситуация. Пробыв некоторое время на примесном центре, электрон все-таки будет выброшен с центра энергией теплового движения и станет свободным прежде, чем этот центр успеет захватить дырку. В такой ситуации примесные центры принято называть *центрами прилипания*.

И, наконец, по-настоящему глубокий центр, *центр рекомбинации*, будет удерживать захваченный электрон до тех пор, пока в его окрестности не появится дырка. Как только это произошло, электрон и дырка рекомбинируют. Убийца сделал свое дело, истребив электронно-дырочную пару, и... готов снова приняться за него.

Если электроны и дырки должны гибнуть в приборе по возможности быстрее — такое требование часто возникает при создании быстродействующих полупроводниковых приборов, — в материал следует ввести примеси, создающие эффективные центры рекомбинации. Иногда, напротив, электроны и дырки должны жить долго. В этом случае полупроводник подвергается тщательной очистке от примесей.

Нужно еще сказать, что иногда даже при самой тщательной очистке материала от примесей, создающих глубокие уровни, от центров рекомбинации избавиться не удается. Дело в том, что эффективными центрами рекомбинации часто являются не только атомы определенных примесей, но и различного рода структурные нарушения кристаллической решетки полупроводника.

Технологи хорошо знают, например, что если нагреть полупроводник до высокой температуры и быстро его охладить, то время жизни неравновесных носителей τ может уменьшиться в десятки и сотни раз. Даже если нагрев и охлаждение проводились в условиях, исключающих загрязнение материала. Дело в том, что при быстром охлаждении в кристалле возникают внутренние напряжения, так как разные части кристалла остывают с разной скоростью. В результате в кристаллической решетке образуются структурные дефекты (*термодефекты*), обладающие всеми свойствами глубоких центров. Такие дефекты могут или захватывать электроны у соседних атомов полупроводника (*термоакцепторы*), или способствовать рождению свободных электронов (*термодоноры*).

Свойствами глубоких центров обладают не только термодфекты. Для каждого хорошо изученного полупроводника существует целый список дефектов решетки, обладающих свойствами глубоких рекомбинационных центров. Иногда центром рекомбинации может быть даже вакансия. (Вакансией называется пустое место, отсутствие атома в кристаллической решетке на том месте, где ему полагалось бы быть.) На первый взгляд может показаться, что от приписывания пустому месту свойств донора (или акцептора) попахивает мистикой и черной магией. В действительности, конечно, ничего таинственного тут нет. Отсутствие на своем месте атома решетки меняет условия связи между валентными электронами соседних атомов и может или облегчить отрыв электрона от одного из ближайших к вакансии атомов, или побудить атом к захвату электрона у соседей. В первом случае вакансия играет роль донора. Во втором — акцептора.

Таким образом, для того чтобы время жизни τ электронов и дырок было велико, следует позаботиться не только о том, чтобы кристалл был тщательно очищен от примесей, но и о том, чтобы его кристаллическая структура была достаточно совершенна.

Жизнь и движение

В предыдущем разделе мы узнали, как рождаются и погибают дырки и электроны проводимости. Сейчас нам предстоит познакомиться с тем, как они живут.

Для свободного электрона и дырки жизнь — это движение.

Прежде всего — хаотическое *тепловое* движение, которое существует всегда и происходит тем более энергично, чем выше температура кристалла. Если к кристаллу приложено электрическое поле, возникает электрический ток — *направленное вдоль поля* движение свободных носителей заряда. Наконец, если по какой-либо причине в одном месте кристалла носителей окажется больше, чем в другом, возникает *диффузионное* движение носителей, или, как часто говорят, диффузия. Диффузия играет важную роль в работе полупроводниковых приборов.

Тепловое движение: средняя скорость, шарики для настольного тенниса на подносе и в ведре с водой. Свободные электроны и дырки под влиянием тепловых колебаний решетки находятся в состоянии беспорядочного хаотического движения. Средняя энергия теплового дви-

жения равняется $\frac{3}{2}kT$. Приравнивая величину этой энергии кинетической энергии частицы $mv^2/2$, можно найти среднюю скорость v_T хаотического движения электрона (или дырки):

$$v_T = \left(\frac{3kT}{m} \right)^{1/2}. \quad (11)$$

При $T = 300$ К, считая, что масса свободного носителя заряда равняется массе свободного электрона $m_0 = 9,1 \times 10^{-31}$ кг, получим для скорости v_T значение $\sim 10^5$ м/с.

Средняя энергия и средняя скорость теплового движения являются очень важными характеристиками свободных носителей заряда в полупроводнике. Однако понятно, что из-за случайного, хаотического характера теплового движения в каждый данный момент в коллективе свободных носителей (электронов или дырок) всегда присутствуют такие, энергия и скорость которых гораздо выше, чем средние значения энергии и скорости.

Пусть в электронном полупроводнике концентрация свободных электронов равняется n_0 . Сколько в каждом кубическом сантиметре кристалла содержится «горячих» электронов, чья кинетическая энергия значительно превышает среднюю? Или, иными словами, какова концентрация электронов с кинетической энергией $\Delta \mathcal{E} \gg kT$?

Эта концентрация (вспомните о выражении (1)!) может быть подсчитана по приближенной формуле

$$n \approx n_0 e^{-\Delta \mathcal{E}/kT}. \quad (12)$$

Или, если отсчитывать энергию свободных электронов от уровня \mathcal{E}_c ,

$$n = n_0 e^{-(\mathcal{E} - \mathcal{E}_c)/kT}. \quad (12a)$$

Распределение электронов по энергии может быть очень наглядно представлено на энергетической диаграмме (рис. 11). Эта диаграмма имеет не совсем обычный вид. Прежде всего, не видно уровня \mathcal{E}_v , который до сих пор всегда присутствовал на энергетических диаграммах вместе с уровнем \mathcal{E}_c .

Дело в том, что до сих пор мы всегда изображали на энергетических диаграммах процессы, энергия которых по порядку величины равнялась $\mathcal{E}_c - \mathcal{E}_v = \mathcal{E}_g$, т. е. составляла для типичного полупроводника величину ~ 1 эВ. В данном случае нас интересует распределение, в котором характерным масштабом энергии является гораздо меньшая величина порядка kT (0,025 эВ при

комнатной температуре). Мы как бы рассматриваем в микроскоп, что происходит с электронами на уровне \mathcal{E}_c . При этом уровень \mathcal{E}_v оказывается далеко внизу, вне поля зрения «микроскопа».

Электроны на уровне \mathcal{E}_c имеют нулевую кинетическую энергию. Увеличению кинетической энергии соответствует продвижение электрона на энергетической диаграмме вверх. Из рис. 11 видно, что, как и следует из формулы (12), подавляющая часть электронов имеет кинетическую энергию порядка kT . Имеются, однако, и очень

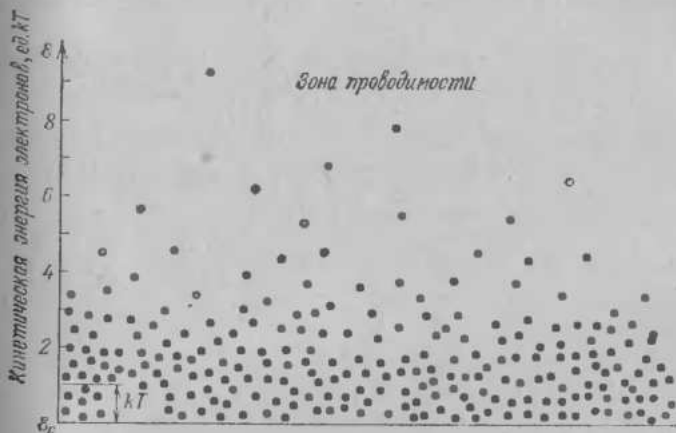


Рис. 11. Распределение по энергии электронов проводимости: шарики для настольного тенниса на трясущемся подносе. (Энергия по оси ординат отложена в единицах kT)

энергичные электроны со значительно большей энергией. Правда, таких электронов мало.

Из рис. 11 видно также, что ставшему свободным электрону проводимости доступна обширная область (зона) энергий $\mathcal{E} > \mathcal{E}_c$. Эту область энергий называют зоной проводимости.

Изображенная на рис. 11 картина очень напоминает движение легких шариков для настольного тенниса на трясущемся подносе. Амплитуда, с которой трясется поднос, тем больше, чем выше температура. Подавляющая часть шариков-электронов вследствие тряски взлетает на высоту, соответствующую энергии $\sim kT$. Но из-за хаотического характера столкновений друг с другом и

с подносом некоторые (очень немногие!) шарики ухитряются приобрести и гораздо большую энергию.

Распределение по энергии присуще, естественно, не только свободным электронам проводимости, но и дыркам. Средняя скорость, с которой будет перемещаться по кристаллу разорванная связь — дырка, зависит от того, как часто будут попадать на место разорванной связи электроны с соседних орбит. Так как процесс замещения пустой связи — процесс случайный, определяемый хаотическим тепловым движением, то ясно, что в каждый данный момент найдутся дырки, которые хаотически мечутся быстрее или медленнее, чем другие.

Средняя энергия теплового хаотического движения дырок, естественно, равняется $\frac{3}{2}kT$, а средняя тепловая скорость описывается формулой (11).

Распределение дырок по энергии описывается приближенным выражением, аналогичным формуле (12):

$$p \approx p_0 e^{-\Delta \mathcal{E}/kT}, \quad (13)$$

или, если отсчитывать энергию дырок от уровня \mathcal{E}_{v1}

$$p \approx p_0 e^{-|\mathcal{E} - \mathcal{E}_{v1}|/kT}. \quad (13a)$$

Внимательного читателя должен заставить насторожиться знак модуля ($| \quad |$), стоящий в показателе экспоненты. Откуда он взялся и что означает?

Чтобы ответить на этот вопрос, необходимо разобраться, как следует изображать распределение дырок по энергии на энергетической диаграмме.



Казалось бы, что же тут думать? Построить картину, аналогичную рис. 11, только энергию откладывать не от уровня \mathcal{E}_c , а от уровня \mathcal{E}_v , вот и все! Увы, такая картина была бы совершенно бессмысленна и равным счетом ничему не соответствовала бы. На самом деле изображение распределения дырок на энергетической диаграмме выглядит совершенно иначе и на первый взгляд довольно странно.

Чтобы понять, как именно, рассмотрим следующую задачу. Пусть в полупроводниковом кристалле поглотился квант света с энергией \mathcal{E}_ϕ , значительно большей, чем энергия \mathcal{E}_g . Энергичный фотон с энергией $\mathcal{E}_\phi > \mathcal{E}_g$ не только разрывает электронные связи. Остаток энергии $\Delta \mathcal{E} = \mathcal{E}_\phi - \mathcal{E}_g$ передается в виде кинетической энергии «новорожденным» носителям — электрону и дырке,

Предположим сначала, что вся избыточная энергия передается новорожденному электрону. Энергетическая диаграмма, соответствующая такому процессу, показана на рис. 12, а. Получив избыточную кинетическую энергию $\Delta\mathcal{E}$, электрон взлетел на энергетической диаграмме в зоне проводимости высоко вверх. Энергетический зазор, т. е. расстояние на энергетической диаграмме между

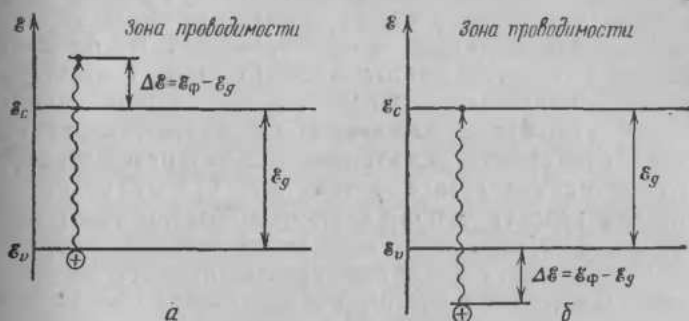


Рис. 12. Чем больше кинетическая энергия электрона $\Delta\mathcal{E}$, тем выше он располагается на энергетической диаграмме (а). Чем больше кинетическая энергия дырки $\Delta\mathcal{E}$, тем ниже на энергетической диаграмме она располагается (б)

электронем и дыркой, равно \mathcal{E}_ϕ , как это и должно быть исходя из закона сохранения энергии.

А теперь, напротив, предположим, что вся избыточная энергия фотона $\Delta\mathcal{E}$ передана не электрону, а дырке. Как изобразить такой процесс на энергетической диаграмме?

По условию, возникший электрон имеет минимально возможную кинетическую энергию. Его, соответственно, на энергетической диаграмме следует поместить в непосредственной близости от уровня ε_c — нижней границы зоны проводимости (рис. 12, б). А куда же поместить на диаграмме энергичную дырку с кинетической энергией $\Delta\mathcal{E}$?

Закон сохранения энергии неумолимо требует, чтобы энергетический зазор между возникшим электронам и дыркой равнялся \mathcal{E}_ϕ . Следовательно? Следовательно, ничего не останется, как поместить дырку ниже уровня ε_v на величину $\Delta\mathcal{E} = \mathcal{E}_\phi - \varepsilon_g$ (рис. 12, б). Чем больше кинетическая энергия возникшей дырки $\Delta\mathcal{E}_x$, тем ниже место, которое она занимает на энергетической диаграмме.

Таким образом, на энергетической диаграмме *росту энергии электрона соответствует движение электрона по шкале энергий вверх, а росту энергий дырки — движение дырки по шкале энергий вниз.*

Конечно, это очень непривычно: частица, набирая кинетическую энергию, движется по шкале энергий не вверх, а вниз. Однако не часто, но время от времени нам все же придется вспоминать, что, как ни удобно пользоваться понятием частица-дырка, дырка является все-таки не подлинной частицей, а *квазичастицей*. И, обсуждая какое-либо сложное понятие или явление, связанное с дырками, бывает очень полезно, а иногда и необходимо вспомнить, что квазичастица-дырка возникает, когда электрон уходит с орбиты межатомной связи, и исчезает, когда электрон на нее возвращается. Вот почему положение дырки на энергетической диаграмме определяется энергией, которая выделится при рекомбинации соответствующей электронно-дырочной пары. Например, при рекомбинации электрона и дырки, показанных на рис. 12, а, б, выделилась бы энергия $\mathcal{E}_g + \Delta\mathcal{E}$, в точности равная, как этого и требует закон сохранения энергии, энергии поглощенного фотона.



С законом сохранения энергии, конечно, спорить не приходится. А все-таки этот фокус с помещением дырки под уровень \mathcal{E}_v вызывает некоторый внутренний протест и требует обсуждения.

Прежде всего потому, что непонятно, чему соответствует область энергий, расположенная ниже уровня \mathcal{E}_v . Когда мы вводили понятие об этом уровне (рис. 7 и относящийся к нему текст), мы понимали под ним энергию электрона, связанного на межатомных орбитах атомов кремния. Мы как бы привяли уровень \mathcal{E}_v за начало отсчета энергии. Действительно, если сообщить электрону на уровне \mathcal{E}_v определенную энергию, он может перейти на акценторный уровень (рис. 8, б). Если сообщить электрону на уровне \mathcal{E}_v энергию \mathcal{E}_g , он может стать свободным (рис. 8, а). Еще большая порция энергии создаст энергичный, «горячий» электрон с большой кинетической энергией (рис. 12, а). С энергиями, большими \mathcal{E}_v , все понятно. А вот чему соответствует область энергий, меньших \mathcal{E}_v ?

Чтобы ответить на этот непростой вопрос, нужно задуматься над тем, что же представляет собой «энергия электрона, связанного на межатомных орбитах атомов кремния».

Представим себе один-единственный изолированный атом. Чтобы сорвать с его внешней электронной оболочки валентный электрон, необходимо затратить определенную энергию, называемую энергией ионизации \mathcal{E}_i (рис. 13, а). Будем теперь подносить к этому атому другой, точно такой же атом. Пока два атома находятся далеко друг от друга, значение энергии ионизации каждого в точности одно и то же и равно \mathcal{E}_i . Если, однако, сблизить атомы на такое расстояние, чтобы стало заметным взаимодействие их электронных

орбит, т. е. так, чтобы электроны могли хотя бы изредка переходить от одного атома к другому, то ситуация изменится. Можно обнаружить, что значения энергии ионизации атомов теперь, во-первых, не равны. В такой ситуации существует не одно, а два значения энергии ионизации (рис. 13, б). Принято говорить, что уровень энергии \mathcal{E}_1 , соответствующий валентному электрону, расщепляется. Во-вторых, ни одно из этих новых значений энергии, вообще

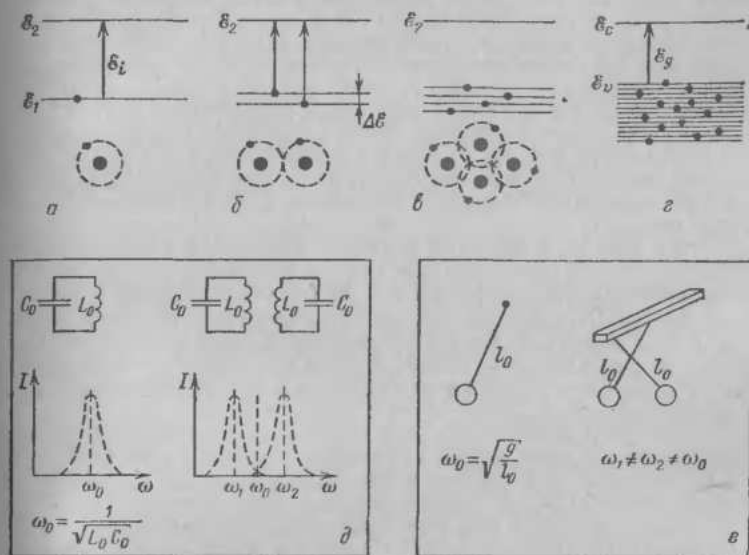


Рис. 13. Расщепление уровней энергии при взаимодействии: *a* — одиночный (изолированный) атом характеризуется энергией ионизации \mathcal{E}_i ; *б* — система из двух взаимодействующих атомов: вследствие взаимодействия уровень \mathcal{E}_1 расщепился; *в, г* — число уровней, возникших в результате расщепления, равно числу взаимодействующих атомов; *д* — расщепление резонансной частоты двух связанных LC-контуров; *е* — система из двух одинаковых связанных маятников характеризуется двумя частотами ω_1 и ω_2 , при этом $\omega_1 \neq \omega_2 \neq \omega_0$

говоря, не равно значению \mathcal{E}_i для изолированного атома. И, наконец, в-третьих, энергетический зазор $\Delta\mathcal{E}$ между расщепленными уровнями тем больше, естественно, чем ближе сдвинуты атомы и чем сильнее взаимодействие их электронных орбит.

Если на небольшом расстоянии друг от друга находятся четыре взаимодействующих атома, уровень \mathcal{E}_1 расщепится на четыре уровня (рис. 13, в). При N взаимодействующих атомах уровень расщепится на N уровней (рис. 13, г).

Расщепление уровней энергии при взаимодействии представляет собой явление, которое наблюдается отнюдь не только в мире атомов и электронов. Это свойство любых колебательных систем,

и с ним, в частности, приходится встречаться в радиотехнике и механике.

На рис. 13, *д* слева показан одиночный LC-контур. При емкости C_0 и индуктивности L_0 круговая частота колебаний контура, соответствующая резонансу, равняется, как известно, $\omega_0 = 1/\sqrt{L_0 C_0}$. Теперь можно взять *точно такой же контур*, с теми же значениями емкости и индуктивности C_0 и L_0 , и разместить его рядом с первым так, чтобы было заметным взаимодействие между индуктивностями контура (т. е. так, чтобы был не слишком мал коэффициент взаимной индукции между индуктивностями контуров. Или, если говорить еще подробнее, так, чтобы магнитный поток, создаваемый одной катушкой индуктивности, некоторой своей частью пронизывал витки соседней катушки). Тогда можно убедиться, что такая система двух связанных контуров имеет резонансную характеристику, показанную на рис. 14, *д* справа. Резонансная частота ω_0 «расщепилась» на две близкие частоты ω_1 и ω_2 . Разница между частотами $\Delta\omega = \omega_2 - \omega_1$ тем больше, чем сильнее связаны взаимодействующие контуры.

На рис. 13, *е* приведен пример расщепления уровней энергии при взаимодействии двух одинаковых маятников длиной l_0 . Если маятники изолированы, круговая частота колебаний обоих равняется $\omega_0 = \sqrt{g/l_0}$ (g — ускорение силы тяжести). Если же обеспечить взаимодействие между маятниками, например подвесив их на одну не очень жесткую балку, то колебания их будут характеризоваться не одной, а двумя частотами ω_1 и ω_2 . При этом $\omega_1 \neq \omega_2 \neq \omega_0$ *).

Теперь, познакомившись с явлениями расщепления уровней энергии при взаимодействии, мы понимаем, что не существует одного-единственного значения «энергии электрона, связанного на межатомных орбитах атомов кремния». Кристалл полупроводника представляет собой систему тесно взаимодействующих атомов с плотностью $\sim 10^{22}$ атомов/см³. Поэтому в каждом кубическом сантиметре кристалла существует громадное ($\sim 10^{22}$) число уровней энергии, соответствующих связанным на орбитах атомов валентным электронам (рис. 13, *з*). Величина расщепления, т. е. энергетический зазор между соседними уровнями, очень мал, так что практически существует непрерывный спектр энергий валентных электронов. Эта область энергий называется *валентной зоной*.

Видно, что уровню \mathcal{E}_v , о котором мы говорили раньше, обсуждая энергетические диаграммы, соответствует самый высокий из возможных уровней валентных электронов, или, как говорят, потолок валентной зоны. Квант света с энергией \mathcal{E}_g может перевести электрон с уровня \mathcal{E}_v на уровень проводимости \mathcal{E}_c . А квант света с большей энергией? Такой квант, если его избыточная энергия $\Delta\mathcal{E} = \mathcal{E}_\phi - \mathcal{E}_g$ будет передана дырке, вырвет электрон не с уровня \mathcal{E}_v , а с более глубоко лежащего уровня валентной зоны (рис. 14, *а*). И чем больше избыточная энергия кванта $\Delta\mathcal{E}$, равная кинетической энергии возникшей дырки, тем ниже место, которое оно занимает на энергетической диаграмме.

*) Попробуйте сами осуществить такой несложный опыт, попятить и проинтерпретировать интересные явления, которые будут при этом наблюдаться. В качестве «не очень жесткой балки» годится обыкновенная линейка или даже бельевая веревка.

Несколько забегаая вперед, мы можем проследить судьбу возникшей таким образом «горячей» дырки (рис. 14, б). Электроны, как мы знаем, стремятся занять самое низкое из возможных энергетических состояний. Как только в глубине валентной зоны возник пустой, не заполненный электроном уровень (дырка), электроны с более высоко лежащих уровней будут стремиться заполнить его

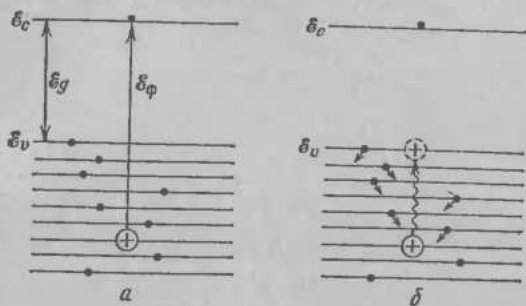


Рис. 14. Возникновение дырки в глубине валентной зоны: а — чем больше избыточная энергия фотона $\Delta\mathcal{E} = \mathcal{E}_\phi - \mathcal{E}_g$, переданная дырке, тем с более глубокого уровня валентной зоны будет вырван электрон; б — растрачивая избыточную кинетическую энергию, дырка «всплывает» к потолку валентной зоны

(рис. 14, б). Свободный уровень (дырка) будет, таким образом, подниматься к потолку валентной зоны. Чем больше кинетическая энергия дырки $\Delta\mathcal{E}$, тем глубже, как мы знаем, она располагается на энергетической диаграмме. Таким образом, ясно, что движению дырки на энергетической диаграмме вверх, к потолку валентной зоны \mathcal{E}_v , будет соответствовать потеря ею кинетической энергии («остывание»).

С учетом сказанного распределение дырок по энергии будет выглядеть так, как показано на рис. 15. Уровня \mathcal{E}_c на энергетической диаграмме не видно: в выбранном масштабе он находится далеко-далеко вверху. Подавляющее большинство дырок имеет кинетическую энергию порядка kT и соответственно располагается в области энергий порядка kT ниже уровня \mathcal{E}_v . Немногие очень энергичные дырки ухитряются приобрести и существенно большую энергию. Они находятся сравнительно глубоко под уровнем \mathcal{E}_v .

Из рис. 15 видно, что дыркам доступна обширная область (зона) энергий ниже уровня \mathcal{E}_v . Эта область энергий называется *валентной зоной*.

Если в полупроводниковом кристалле отсутствуют примеси и дефекты, создающие уровни энергии, лежащие между уровнями \mathcal{E}_c и \mathcal{E}_v , то, как мы хорошо знаем,

электроны не могут иметь энергию, лежащую в области между \mathcal{E}_c и \mathcal{E}_v . Эта область энергий носит поэтому название *запрещенной зоны*.

Картина, похожая на изображенную на рис. 15, будет наблюдаться, если шарики для настольного тенниса поместить в накрытое крышкой ведро с водой и трясти его,

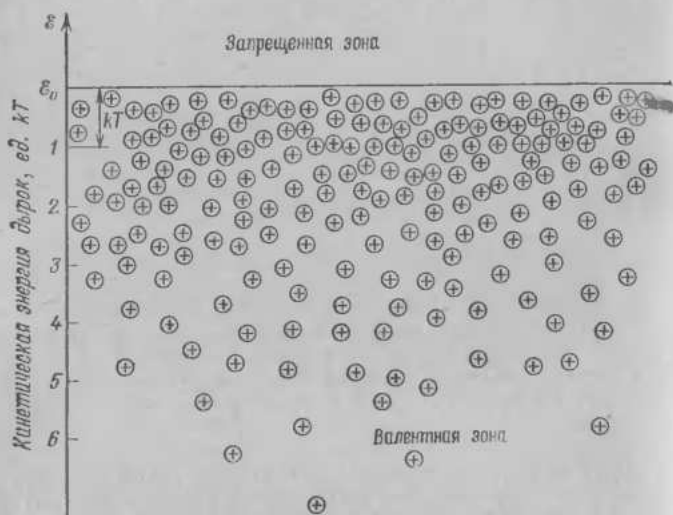


Рис. 15. Распределение дырок по энергии (шарики для настольного тенниса в трясущемся ведре, накрытом крышкой)

Если ведро оставить в покое и не трясти, все шарики (дырки) всплыли бы наверх, к крышке (потолку валентной зоны). Чем сильнее трясется ведро (чем выше температура), тем глубже окажутся наиболее энергичные шарики.

Кинетическая энергия электронов или дырок при некоторых обстоятельствах может быть очень большой.

Осветим, например, пластину кремния зеленым светом ($\lambda = 0,5$ мкм). Энергия фотонов такого света $\mathcal{E}_\phi = hc/\lambda = = 2,48$ эВ *). На образование электронно-дырочной пары уйдет энергия $\mathcal{E}_g = 1,1$ эВ. Таким образом, если вся избыточная энергия передается образовавшемуся элек-

*) С учетом числовых значений постоянной Планка h и скорости света в вакууме c соотношение между энергией фотона \mathcal{E}_ϕ (в эВ) и соответствующей длиной волны λ (в мкм) выражается удобной расчетной формулой: $\mathcal{E}_\phi = 1,24/\lambda$.

трону, его кинетическая энергия составит $\sim 1,4$ эВ, т. е. в 35 раз больше, чем средняя тепловая энергия электронов при комнатной температуре, равная приблизительно 0,04 эВ. Может быть, можно еще нагляднее представить себе, как велика эта энергия, если учесть, что она соответствует энергии электронов при температуре $35 \times 300 \approx 10\,500$ К!

Какая же участь уготована столь горячему и энергичному носителю? На первый взгляд может показаться, что при рекомбинации его с дыркой излучится такой же энергичный фотон, какой погиб, давая ему жизнь.

Такой акт рекомбинации возможен, но очень-очень маловероятен. Из десятков, а иногда и сотен тысяч рожденных светом носителей таким образом рекомбинируют

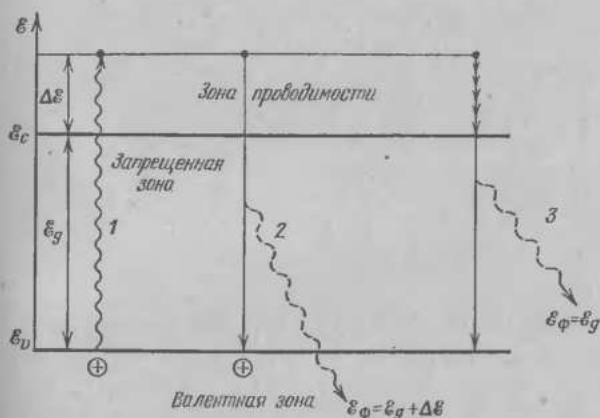


Рис. 16. Рождение и гибель горячего электрона: рождение электрона с большой избыточной энергией $\Delta\mathcal{E}$ при поглощении кванта с энергией $\mathcal{E}_\phi = \mathcal{E}_g + \Delta\mathcal{E}$ (стрелка 1); такой электрон, в принципе, может рекомбинировать с дыркой (стрелка 2); типичный путь рекомбинации показан стрелкой 3

едва ли единицы. Типичный путь рекомбинации таков: электрон сначала растрчивает практически всю избыточную энергию $\Delta\mathcal{E}$ маленькими порциями. И только «охладившись» и попав на уровень \mathcal{E}_c , рекомбинирует. При этом излучается фотон с энергией \mathcal{E}_g (рис. 16).

Так происходит потому, что электроны (и дырки) с громадной частотой $\sim 10^{12} - 10^{13}$ раз в секунду сталкиваются с решеткой кристалла, участвуя в тепловом, хаотическом движении. Электрон, энергия которого равна

$\frac{3}{2}kT$, находится в тепловом равновесии с решеткой: при некоторых соударениях он теряет энергию, при некоторых — приобретает, и в среднем его энергия остается равной $\frac{3}{2}kT$. Совсем иначе обстоит дело с горячим электроном. При каждом столкновении он теряет энергию и поэтому очень быстро «остывает». Ситуация очень напоминает ту, которая складывается, когда очень спешащий и не очень вежливый человек пытается прорваться к эскалатору метро через густую толпу, образовавшуюся в



часы «пик». Он на большой скорости врывается в толпу, иногда даже ухитряется отпихнуть в сторону двух-трех человек... И все. Остывает. Через какую-то секунду он так же медленно переступает, как и его соседи по очереди.

Энергия, которую горячий электрон растрчивает при столкновениях с решеткой, выделяется не в форме света, а в виде тепла и идет на нагревание кристалла.

Величина энергии, которую горячий носитель теряет при каждом столкновении с решеткой, разная у разных полупроводников и, как правило, составляет несколько сотых электрон-вольта. С другой стороны, типичная избыточная энергия горячих носителей редко превышает несколько десятых электрон-вольта. Таким образом, при частоте столкновений $10^{12} - 10^{13}$ раз в секунду носитель «остывает» за время порядка $10^{-11} - 10^{-12}$ с. А время жизни носителей τ — среднее время, которое они проводят в свободном состоянии перед тем, как рекомбинировать, обычно составляет $10^{-8} - 10^{-9}$ с, т. е. в тысячи, миллионы, а иногда даже и миллиарды раз больше. Соответст-

венно подавляющая часть носителей рекомбинирует, успев остыть.

Если в кристалле возникает горячая дырка, то она так же, как и горячий электрон, быстро растрчивает энергию в результате столкновений. Электрон проводимости при этом спускается вниз по шкале энергий, стремясь к минимально возможному для него уровню энергии \mathcal{E}_c (рис. 16). А дырка, наоборот, «всплывает», стремясь к минимально возможному для нее уровню кинетической энергии \mathcal{E}_v (стрелка 2 на рис. 17). Обратите внимание,

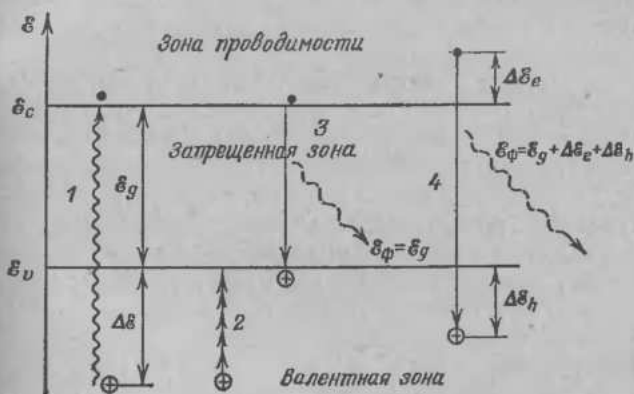


Рис. 17. Рождение и гибель горячей дырки (ср. с рис. 16): рождение горячей дырки (стрелка 1); остывание дырки (стрелка 2); рекомбинация «холодных» носителей (стрелка 3); рекомбинация горячих носителей (стрелка 4)

если электрон и дырка рекомбинируют, когда горячая дырка полностью растратила избыточную энергию и достигла уровня \mathcal{E}_v , — генерируется квант с энергией $\mathcal{E}_g = \mathcal{E}_c - \mathcal{E}_v$ (стрелка 3). Если же рекомбинируют электрон с избыточной энергией $\Delta\mathcal{E}_e$ и дырка с избыточной энергией $\Delta\mathcal{E}_h$ (*), рождается квант с энергией $\mathcal{E}_\phi = \mathcal{E}_g + \Delta\mathcal{E}_e + \Delta\mathcal{E}_h$ (стрелка 4), как это и следует из закона сохранения энергии (и из здравого смысла).

Движение в электрическом поле: подвижность, зависимость дрейфовой скорости от поля, энергетическая диаграмма. Если к полупроводнику приложено электрическое поле, то на хаотическое тепловое движение сво-

*) Индекс «h» — от английского слова hole (дырка).

бодных носителей (электронов и дырок) накладывается направленный дрейф частиц вдоль направления поля. Определим скорость этого направления движения.

Если электрическое поле отсутствует, то электрон (дырка) участвует только в хаотическом тепловом движении. Никакого избранного направления, в котором электрону (дырке) двигаться было бы предпочтительней, чем в другом, не существует. Поэтому, хотя в промежутке между столкновениями электрон (дырка) движется с очень большой скоростью ($\sim 10^5$ м/с, как мы установили в предыдущем пункте), направленное движение отсутствует. Скорость направленного движения равна нулю.

В присутствии электрического поля электроны (дырки) продолжают свою бешеную пляску. По-прежнему с частотой $\sim 10^{12} - 10^{13}$ раз в секунду носители испытывают столкновения. По-прежнему невозможно предсказать, куда полетит носитель после очередного столкновения: вперед, назад, влево, вправо, вверх, вниз — под любым углом, в любом направлении. Но, куда бы они полетели, электрическое поле будет, пусть слабо, тянуть его всегда в одном и том же направлении. В результате возникает направленное движение.

В электрическом поле E на электрон (или дырку) действует сила $F = eE$. Под действием этой силы носитель приобретает вдоль поля ускорение $a = eE/m$. Двигаясь без столкновений, носитель за время t приобретает скорость в направлении поля $v = at = \frac{eE}{m} \cdot t$.

Чтобы правильно вычислить среднюю скорость, приобретаемую носителем в условиях многократных столкновений, нужно помнить о двух обстоятельствах.

Во-первых, как мы уже знаем, непосредственно после столкновения носитель может с равной вероятностью двигаться в любом направлении. Это означает, что скорость направленного движения после столкновения равняется нулю. Во-вторых, поскольку процесс столкновения — процесс случайный, носитель в промежутке между двумя последовательными столкновениями проведет «в свободном полете» разное время. Оба эти обстоятельства наглядно видны из рис. 18, на котором показана зависимость скорости направленного движения носителя от времени.

Средняя скорость направленного движения \bar{v} равняется произведению ускорения на среднее время между

столкновениями τ_0 :

$$\bar{v} = \frac{e\tau_0}{m} E. \quad (14)$$

Таким образом, скорость направленного движения или, как часто говорят, дрейфовая скорость, свободных носителей в кристалле пропорциональна напряженности электрического поля. Коэффициент пропорциональности

$$\mu = \frac{e\tau_0}{m} \quad (15)$$

носит название подвижности,

Подвижность свободных носителей является одной из самых важных характеристик полупроводникового

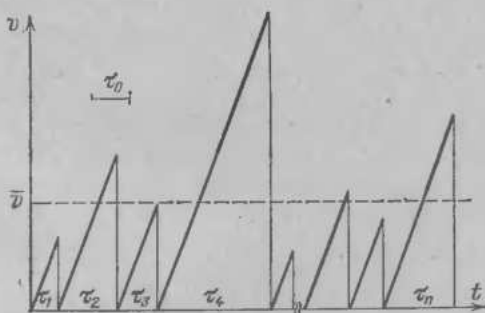


Рис. 18. Зависимость скорости направленного движения электрона (или дырки) в электрическом поле от времени в условиях случайных столкновений

кристалла. Значения подвижности во многом определяют пригодность материала для изготовления полупроводниковых приборов.

Из формул (14) и (15) следует, что

$$\bar{v} = \mu E. \quad (16)$$

Чем больше подвижность, тем больше скорость направленного движения носителей при той же напряженности электрического поля E . Из формулы (16) сразу видно, что размерность подвижности $[\mu] = \text{м}^2/(\text{В}\cdot\text{с})$.

Значения подвижности электронов и дырок наиболее важных полупроводниковых материалов при комнатной температуре приведены в табл. 2.

Зная величину подвижности, с помощью соотношения (16) легко рассчитать, при каких значениях электрическо-

Полупроводники	InSb	Ge	Si	InP	GaAs	GaP	SiC
Подвижность электронов μ_n , $\text{м}^2/(\text{В}\cdot\text{с})$	8	0,39	0,13	0,55	1	0,05	0,02—0,1
Подвижность дырок μ_p , $\text{м}^2/(\text{В}\cdot\text{с})$	0,07	0,19	0,05	0,07	0,04	0,01	$5\cdot 10^{-4}$

го поля скорость направленного движения носителей \bar{v} сравнивается с тепловой скоростью v_T , определяемой формулой (11). Принимая $v_T \sim 10^5$ м/с, найдем, что в электронном германии ($n\text{-Ge}$) это произойдет при $E \approx 2,5 \cdot 10^5$ В/м, в $n\text{-Si}$ — при $E \approx 7 \cdot 10^5$ В/м, в $n\text{-GaAs}$ — при $E \approx 10^5$ В/м. Из табл. 2 видно, что подвижность дырок μ_p , как правило, значительно меньше, чем подвижность электронов μ_n в том же материале. Поэтому, чтобы скорость направленного вдоль поля движения дырок сравнивалась с тепловой, нужны еще более сильные поля. Для дырок в $p\text{-Ge}$ характерное значение напряженности необходимого для этого поля $E \approx 5 \cdot 10^5$ В/м, в $p\text{-Si}$ — $E \approx 2 \cdot 10^6$ В/м, в $p\text{-GaAs}$ — $E \approx 2,5 \cdot 10^6$ В/м.

При анализе работы полупроводниковых приборов приходится встречаться со значениями напряженности поля E , лежащими в диапазоне от 0 до $\sim 10^6$ В/м.

В слабых полях, когда скорость направленного движения мала по сравнению с тепловой, наличие или отсутствие электрического поля никак не сказывается на характере столкновений носителей с кристаллической решеткой. При этом подвижность μ является величиной постоянной, не зависящей от напряженности поля E . Дрейфовая скорость носителей пропорциональна полю E (формула (16)), что соответствует выполнению закона Ома.

В начале книги (см. с. 18) уже упоминалось, что из пропорциональности дрейфовой скорости \bar{v} напряженности электрического поля E следует закон Ома. Установим связь между законом Ома и формулой (16). Вспомним несколько соотношений, известных из школьного курса физики.

Ну, прежде всего, сам закон Ома: $I = U/R$.

Сопротивление образца R зависит от его длины L , площади поперечного сечения S и удельного сопротивления ρ : $R = \rho L/S$. Впрочем, в физике полупроводников чаще предпочитают использовать

не удельное сопротивление, а обратную ему величину — проводимость σ . Соответственно, выражение для сопротивления образца R приобретает вид $R = L/\sigma S$.

Напряженность поля в образце $E = U/L$.

Ток I , текущий через образец, равняется $I = jS$, где j — плотность тока, связанная с концентрацией n_0 и дрейфовой скоростью \bar{v} хорошо известным и вполне понятным соотношением $j = en_0\bar{v}$. Следовательно, $I = en_0\bar{v}S$.

Подставив выражения для напряженности поля, сопротивления и тока в формулу закона Ома, получаем

$$\bar{v} = \frac{\sigma}{en_0} E.$$

Итак, пропорциональность дрейфовой скорости носителей напряженности поля непосредственно вытекает из закона Ома. А из сравнения полученного выражения и формулы (16) следует, что $\mu = \sigma/en_0$, т. е. что проводимость материала σ прямо пропорциональна подвижности носителей в нем.

Из формулы (16) и выражения для плотности тока $j = en_0\bar{v}$ следует еще одно полезное соотношение

$$j = \sigma E. \quad (16a)$$

Плотность тока в любой точке пропорциональна напряженности поля E . Соотношение (16a) часто называют законом Ома в дифференциальной форме.

В сильных электрических полях кинетическая энергия носителей, приобретаемая ими в электрическом поле, может сравняться и даже стать больше, чем энергия хаотического теплового движения.

Частота и характер столкновений, испытываемых носителями, зависит от их энергии. Поэтому в сильных электрических полях частота и характер столкновений электронов и дырок (а следовательно, и подвижность) зависят от напряженности поля.

Во всех практически используемых полупроводниках при комнатной температуре подвижность в сильных полях падает с ростом напряженности поля E . В очень сильных полях величина подвижности становится обратно пропорциональной напряженности поля E : $\mu \sim 1/E$. В соответствии с выражением (16) это означает, что в очень сильных полях дрейфовая скорость носителей не зависит от поля: $\bar{v} = \text{const}$.

Дело в том, что носитель заряда, электрон или дырка, успевший набрать в промежутке между столкновениями определенную, достаточно большую энергию \mathcal{E}_0 , отдает ее решетке полностью при каждом столкновении. Убедимся в том, что в таких условиях время между столкновениями τ_0 будет обратно пропорционально напряженности поля E .

Энергия носителя определяется его скоростью и массой: $\mathcal{E} = mv^2/2$.

Скорость определяется ускорением $a = eE/m$ и временем между столкновениями τ_0 : $v = eE\tau_0/m$.

Подставляя выражение для v в формулу, связывающую энергию и скорость носителей, получим

$$\tau_0 = \sqrt{\frac{2m\epsilon_0}{e^2}} \frac{1}{E}.$$

В соответствии с формулой (14), величина μ в такой ситуации также будет обратно пропорциональна напряженности поля E . А дрейфовая скорость электронов $\bar{v} = \mu E$ будет величиной постоянной, не зависящей от напряженности поля.

Зависимость дрейфовой скорости от напряженности поля для электронов и дырок в германии и кремнии показана на рис. 19. Видно, что во всех случаях в слабых электрических полях скорость пропорциональна полю,

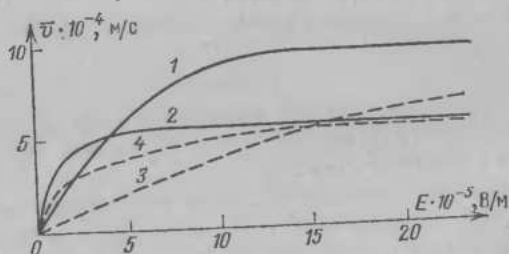


Рис. 19. Зависимость скорости направленного движения (дрейфовой скорости) носителей от напряженности электрического поля в Ge и Si ($T = 300$ K). Сплошные линии — $\bar{v}(E)$ для электронов, штриховые — для дырок: 1, 3 — Si; 2, 4 — Ge

Этот участок зависимости $\bar{v}(E)$ по вполне понятной причине называется *омическим* или *линейным* участком. С ростом поля дрейфовая скорость растет все медленнее и медленнее и в достаточно сильных электрических полях перестает зависеть от напряженности поля E (насыщается). Чем меньше подвижность носителей в слабом электрическом поле, тем в более сильных полях наступает насыщение*). В германии скорость электронов практически перестает зависеть от напряженности поля при $E \gtrsim 5$ кВ/см, скорость дырок — при $E \gtrsim 15$ кВ/см. Значения скорости насыщения v_s ***) и для электронов, и для дырок приблизительно совпадают и равняются $\sim 5 \cdot 10^6$ см/с. В Si скорость насыщения электронов со-

*) Понимаете ли вы, почему?

**) Индекс «s» — от английского слова saturation (насыщение).

ставляет $v_s \sim 10^7$ см/с и достигается при $E \geq 15$ кВ/см. Для дырок в Si значение скорости насыщения также составляет $v_s \sim 10^7$ см/с, но достигается в значительно более сильных полях $E \geq 60$ кВ/см.



Энергетическая диаграмма. В том случае, когда в полупроводнике существует электрическое поле, энергетическая диаграмма, описывающая поведение электронов и дырок, приобретает несколько иной, чем раньше, вид. Чтобы

понять, какой именно и почему, обсудим самый простой случай: к кусочку полупроводника длиной L приложено напряжение U (рис. 20).

В полупроводнике существует везде одинаковое, однородное электрическое поле $E = U/L$. Электроны под

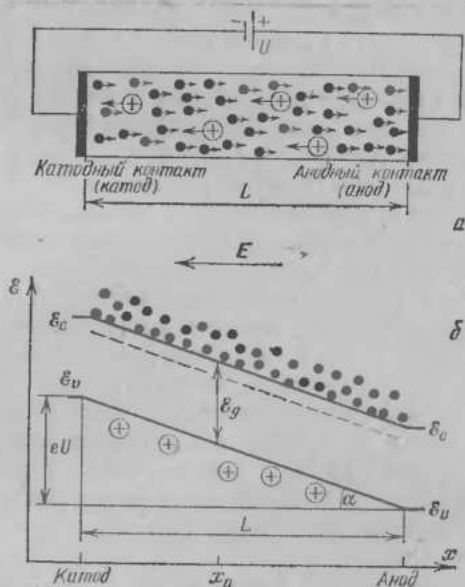


Рис. 20. К образцу длиной L приложено от внешнего источника напряжения U : а — движение электронов и дырок в образце; б — энергетическая диаграмма

действием электрического поля движутся от катода образца к аноду. Дырки — в противоположном направлении, от анода к катоду *).

*) Видно, что на рис. 20, а изображен полупроводник n -типа. Электронов в нем гораздо больше, чем дырок. Но все, что будет говориться дальше об энергетических диаграммах, одинаково применимо и для n -, и для p -типа полупроводников.

Зададимся вопросом: одинакова ли энергия электронов на катоде и на аноде образца?

Ответ очевиден. Конечно, нет. Ведь между катодом и анодом существует разность потенциалов U . Значит, потенциальная энергия электрона, находящегося у катода, отличается от потенциальной энергии электрона у анода на величину $\mathcal{E} = eU$ (e — заряд электрона).

Что служит источником этой разности энергий, на что она тратится и где она больше, на катоде или на аноде? И эти вопросы особой сложности для нас не представляют. Разность энергий создается посторонним источником напряжения, подключенным к образцу и поддерживающим на его концах разность потенциалов U . При движении от катода к аноду электрон, если бы он не сталкивался с решеткой полупроводника, разогнался бы электрическим полем E и в результате, потеряв потенциальную энергию eU , которую он имел у катода, имел бы у анода в точности такую же кинетическую энергию. Но, как мы уже знаем, двигаясь в полупроводнике, электрон очень часто сталкивается с решеткой. В результате никакой дополнительной кинетической энергии при движении к аноду электрон не приобретает, а весь запас его потенциальной энергии eU передается кристаллической решетке полупроводника и выделяется в виде хорошо нам знакомого джоулева тепла *). Наконец, из рассмотренной картины ясно, где энергия больше, а где меньше. Для электрона энергия выше у катода: он растрчивает ее, двигаясь к аноду и не приобретая взамен из-за столкновений с решеткой кинетической энергии. Наоборот, для положительно заряженной дырки, которая в электрическом поле движется от анода к катоду, потенциальная энергия выше у анода. Дырка также растрчивает эту энергию на столкновения с решеткой по мере движения к катодному контакту.

Для случая, показанного на рис. 20,а, потенциал U линейно изменяется с координатой. Поэтому энергетическая диаграмма имеет вид, показанный на рис. 20,б. Энергия, скажем, электрона, находящегося в зоне проводимости на уровне \mathcal{E}_0 на катоде, больше энергии такого же электрона на аноде на величину eU . Так же обстоит дело и с электронами, находящимися, например, в запрещенной зоне на донорном уровне (он показан на рис. 20,б

*) Используя закон Джоуля — Ленца, покажите, что энергия, выделяющаяся в образце при прохождении вдоль него одного электрона, действительно равна eU .

штриховой линией). И на любом другом уровне, глубоком или мелком, донорном или акцепторном, И с электронами на уровне \mathcal{E}_v .

А с дырками?

Энергия дырки, находящейся, например, в валентной зоне на уровне \mathcal{E}_v на катоде, меньше энергии такой же дырки на аноде на величину eU . (Мы, конечно, не забыли, что чем больше энергия дырки, тем ниже она находится на энергетической диаграмме (см. рис. 12).)

Зависимость энергии от координаты $\mathcal{E}(x)$ приобрела вид «горки». Шарики для настольного тенниса (электроны) катятся с этой горки вниз, от катода к аноду. (Те же шарики в ведре с водой (дырки) всплывают по шкале энергий вверх, двигаясь от анода к катоду.)

Очень важно, что параметры, описывающие рождение и смерть электрона и дырки в любой точке кристалла, не зависят от того, существует в образце электрическое поле или нет. Видно, скажем, что в точке x_0 электрону нужна точно такая же, как и раньше, энергия \mathcal{E}_g , чтобы из валентной зоны перейти в зону проводимости. Не изменяются с приложением поля и энергии ионизации донорных и акцепторных уровней $\Delta\mathcal{E}$, и скорости генерации и рекомбинации через глубокие уровни *).

Чем больше приложенное к образцу напряжение, тем круче «горка». Из рис. 20 легко увидеть, что тангенс угла α между линиями диаграммы и осью абсцисс пропорционален напряженности поля E в образце: $\text{tg } \alpha \sim eU/L \sim E$.

А как будет выглядеть энергетическая диаграмма в случае, если поле в образце неоднородно, т.е. напряжен-

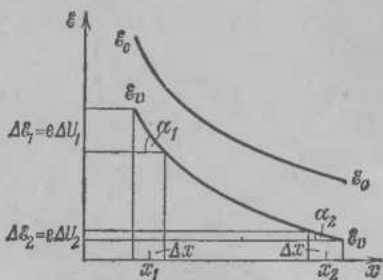


Рис. 21. Пример энергетической диаграммы для случая, когда электрическое поле в образце распределено неоднородно

*) Конечно, если напряженность внешнего электрического поля E достигает очень больших значений (10^6 — 10^8 В/см), сравнимых с межатомными полями, определяющими условия рождения и смерти электронов и дырок, то величины \mathcal{E}_g , $\Delta\mathcal{E}$, скорости генерации и рекомбинации будут очень заметно зависеть от поля E . Но в этой книге мы такие ситуации рассматривать не будем.

ность поля E имеет различную величину в разных точках образца? Такая ситуация очень часто встречается при изучении полупроводниковых приборов.

На рис. 21 показана энергетическая диаграмма для случая, когда напряжение в левой части образца в области точки x_1 спадает значительно быстрее, чем в правой, в районе точки x_2 . Видно, что

$$\operatorname{tg} \alpha_1 \sim e \frac{\Delta U_1}{\Delta x} \sim E_1, \quad \operatorname{tg} \alpha_2 \sim e \frac{\Delta U_2}{\Delta x} \sim E_2. \quad (17)$$

Одного взгляда на энергетическую диаграмму достаточно, чтобы получить наглядное представление о распределении напряженности поля в образце. Однако это не единственное и даже не самое главное достоинство энергетической диаграммы. Как мы увидим во второй части книги, энергетическая диаграмма оказывается особенно полезной, когда существенную роль в движении носителей играет не только электрическое поле, но и *диффузия*.

История самых главных понятий



Как мы убедились, концентрация носителей и их подвижность являются основными характеристиками, описывающими электрические свойства полупроводников. Интересно и поучительно проследить за тем, как возникли эти понятия.

История их возникновения относится к самому началу XIX века, когда знаменитый английский физик и химик Хэмфри Дэви заложил основы электрохимии, проведя первые опыты по разложению веществ электрическим током. Великий ученый сэра Дэви, Майкл Фарадей, изучая законы электролиза, названные впоследствии его именем, предположил, что с атомами растворенного вещества связан определенный заряд, в результате чего эти атомы и движутся в электрическом поле. Заряженные атомы Фарадей назвал *ионами* (что означает по-английски «странники»). Исследование законов и механизма электрического переноса ионов стало одной из главных проблем электрохимии XIX века.

Важный шаг в понимании механизма электролиза сделал в 1838 г. русский физик, в те годы глава кафедры физики Петербургского университета, Эмилий Христианович Ленц. Ленц установил, что электрический ток в электролитах подчиняется закону Ома, т. е. растет пропорционально приложенному напряжению. Как мы теперь понимаем, это обстоятельство указывает на то, что скорости ионов пропорциональны напряженности электрического поля.

Первым, кто сделал такой вывод, был известный немецкий ученый Иоганн Гитторф. В 50-х годах прошлого века он писал: «Поскольку движение ионов подобно движению весьма малых тел в сол-

противляющейся среде, то скорости, а не ускорения, пропорциональны действующим силам. Скорости зависят от самих этих сил и от их способности перемещаться, которую назовем «подвижностью».

Состачественник Гитторфа Фридрих Кольрауш, изучая поведение ионов в электролитах различного состава, ввел в 70-х годах прошлого века прочно укоренившиеся в электрохимии, а затем перешедшие в физику металлов и полупроводников понятия *проводимости* и *концентрации* носителей.

Использование для полупроводников понятий и терминов, возникших в электрохимии, отнюдь не было чисто формальным переносом подходящих названий. Представления о механизме движения ионов в электролитах в течение многих лет, вплоть до появления квантовой теории твердых тел, были основой для понимания физики полупроводников. Образование в полупроводнике свободных электронов объяснялось «диссоциацией» его атомов, движение их в электрическом поле — «движением малых тел в сопротивляющейся среде» и т. д.

Эта первая модель, объяснявшая свойства полупроводников, сыграла важную историческую роль. Именно на ее основе Иоганн Кенигсбергер в 1914 г. впервые определил полупроводники как *класс веществ* и предложил правильную формулу экспоненциального роста их проводимости с температурой, применив к новому классу давно известную экспоненциальную температурную зависимость проводимости электролитов.

Диффузия. Понятие «диффузия» — от латинского *diffusio* (разлитие) — применимо к очень многим явлениям в газах, жидкостях и твердых телах.

Запах одеколона, пролитого в комнате, в конце концов распространится по всей квартире. Даже если отключить батареи, закрыть форточки и законопатить щели, так, чтобы воздух был совершенно неподвижен, молекулы одеколона попадут в другие комнаты благодаря процессу диффузии.

На поверхность полупроводниковой пластины, тщательно очищенной от примесей, наносится слой вещества, содержащего примесь в большом количестве. Через некоторое время эту примесь можно обнаружить в объеме полупроводниковой пластины, далеко от поверхности полупроводника. Примесь попадает в объем благодаря процессу диффузии. Скорость диффузии примесей в полупроводнике очень сильно растет с увеличением температуры. Поэтому при комнатной температуре нужны десятилетия, чтобы примесь проникла в объем полупроводника, а при высокой температуре процесс диффузии может занимать всего лишь несколько часов или даже минут. (И, кстати сказать, диффузионное введение примесей в полупроводник широко используется при изготовлении полупроводниковых приборов.)

Что общего между описанными выше явлениями?

Именно то, что составляет существо процесса диффузии — спонтанное, самопроизвольное, происходящее без каких-либо внешних воздействий перемещение вещества из места, где его много, туда, где его мало.

Процесс диффузии есть прямое следствие хаотического теплового движения атомов (или молекул). Понять, почему совершенно неупорядоченное хаотическое движение частиц приводит к направленному перемещению вещества из области высокой концентрации в область концентрации низкой, очень важно. И очень просто.

Рассмотрим пример с пролитым одеколоном. Мысленно выделим вокруг лужицы пролитого одеколона полусферу, через которую молекулы проходят абсолютно беспрепятственно. Будем подсчитывать молекулы одеколона, прошедшие через поверхность полусферы изнутри наружу и наоборот. Молекулы одеколона под ударом молекул воздуха беспорядочно мечутся во все стороны, хаотически двигаясь во всех направлениях. Те молекулы одеколона, которые находятся поблизости от мысленно выделенной поверхности, получив толчок в нужном направлении, могут пересечь поверхность. Если бы концентрация молекул одеколона была одинакова по обе стороны мысленной «перегородки», то поток молекул, направленных изнутри наружу, был бы в точности равен потоку молекул снаружи внутрь. Сколько молекул одеколона под влиянием случайных блужданий выйдет изнутри за «перегородку», столько же войдет снаружи в выделенный объем.

В разбираемом примере дело обстоит не так, концентрация молекул одеколона n над пролитой лужицей велика и спадает по мере удаления от нее (рис. 22). Снаружи от любой мысленно выделенной поверхности концентрация одеколона ниже, чем внутри. Под влиянием хаотического, случайного движения одна и та же доля молекул от их количества, находящегося непосредственно у поверхности, пересечет поверхность мысленной полусферы



Рис. 22. Распределение молекул одеколона над пролитой лужицей: в области $x > 0$ производная $dn/dx < 0$ (n уменьшается с ростом x), диффузионный поток носителей $J_D > 0$ (направлен вдоль положительного направления оси x); в области $x < 0$ производная $dn/dx > 0$, диффузионный поток $J_D < 0$

изнутри наружу и снаружи внутрь. Одна и та же доля, но от разного количества. Ведь внутри перегородки концентрация выше, чем снаружи. Поэтому больше молекул одеколона пересечет поверхность изнутри, чем снаружи. Образуется направленный изнутри наружу поток молекул одеколона — поток, возникший под влиянием совершенно хаотических блужданий. В этом и состоит процесс диффузии, всегда приводящий к выравниванию концентраций.

Из рассмотренной картины ясно, что поток, вызванный диффузией, будет тем больше, чем резче меняется концентрация вблизи от мысленно выделенной поверхности. Именно от того, насколько резко падает концентрация в рассматриваемой области, и зависит, насколько поток молекул одеколона изнутри наружу превысит встречный поток — снаружи внутрь. Скорость изменения концентрации n с координатой x , естественно, характеризуется производной от концентрации по координате dn/dx . Итак, следует ожидать, что диффузионный поток носителей J_D будет пропорционален dn/dx :

$$J_D = -D \frac{dn}{dx}. \quad (18)$$

Почему перед правой частью уравнения (18) появился знак минус, помогает понять рис. 22. Там, где производная dn/dx положительна, поток направлен против положительного направления оси x , и наоборот.

Коэффициент пропорциональности D носит название *коэффициента диффузии*.

Установим прежде всего размерность коэффициента диффузии. Поток носителей J_D численно равен числу носителей, пересекающих в единицу времени единичную площадь. Таким образом, размерность потока $[J_D] = 1/(m^2 \cdot c)$. Размерность концентрации $[n] = m^{-3}$, а градиента концентрации $[dn/dx]$, очевидно, m^{-4} . Отсюда ясно, что коэффициент диффузии D имеет размерность $[D] = m^2/c$.

От чего зависит коэффициент диффузии? Какими свойствами молекул он определяется?

Установим выражение для коэффициента диффузии D , воспользовавшись так называемыми соображениями размерности. Определение какой-либо величины из размерных соображений при первом знакомстве производит впечатление ловкого фокуса или даже не вполне невинного жульничества. В действительности же это очень

изящный и широко используемый в современной физике метод. О нем рассказывается в статье Ю. Брука и А. Стаценко «Метод размерности помогает решать задачи» (Квант, 1981, № 6).

Итак, от чего же может зависеть коэффициент диффузии?

Естественно ожидать, что коэффициент диффузии будет зависеть от длины свободного пробега молекул l .

Действительно, чем больше l , тем реже сбивают молекулу с «истинного пути» удары других молекул. В рассмотренном нами примере с пролитым одеколоном ясно, что если бы удары молекул воздуха не сообщали движению молекул одеколона характер случайных, хаотических блужданий, то, стартовав с поверхности пролитой лужицы со скоростью v_T , молекула одеколона достигла бы границы комнаты длиной L через время $t \sim L/v_T$. В действительности же она окажется там через гораздо более длительный промежуток времени. Чем чаще будут сбивать ее с пути удары молекул воздуха, чем меньше длина свободного пробега l , тем медленнее будет диффундировать молекула одеколона, тем меньше будет коэффициент диффузии D .

Естественно также ожидать, что коэффициент диффузии будет увеличиваться с ростом тепловой скорости молекул v_T .

От чего еще может зависеть коэффициент диффузии? От времени между столкновениями τ_0 ! Правильно, но сама величина τ_0 выражается через l и v_T : $\tau_0 = l/v_T$.

А от чего еще? Да, пожалуй, больше ни от чего. Может быть, от массы молекул? Вряд ли. Нас ведь интересует не энергия, переносимая молекулами, и не импульс, а просто число молекул.

Итак, предположим, что мы, хорошо представляя себе физическую картину диффузии, правильно определили переменные, от которых зависит коэффициент диффузии D . Тогда осталась главная в методе размерностей, но в нашем случае очень простая задача: так скомбинировать переменные (в нашем случае l и v_T), чтобы комбинация имела размерность определяемой величины. Когда имеешь две величины: одну с размерностью $[v_T] = \text{м/с}$, а другую с размерностью $[l] = \text{м}$, то нетрудно сообразить, что нужно сделать, чтобы получить из них величину D с размерностью $[D] = \text{м}^2/\text{с}$. Перемножить:

$$D \sim lv_T, \quad (19)$$

Подробный расчет дает возможность установить численный коэффициент:

$$D = \frac{1}{3} l v_T, \quad (20)$$

Оставим теперь в покое молекулы одеколона и обратимся к нашим старым знакомым — свободным носителям заряда в полупроводниках.

Представим себе, что в какой-то части полупроводника (для определенности — электронного) концентрация носителей больше, чем в соседних областях. Такая ситуация может возникнуть, например, если эта часть полупроводника нагрета или освещена. Благодаря диффузии, как мы знаем, возникнет поток электронов из области с более высокой в область с более низкой концентрацией. Но направленный поток заряженных носителей по определению представляет собою электрический ток! Плотность такого *диффузионного тока* мы можем легко определить, зная распределение носителей $n(x)$:

$$j_D = e J_D = - e D \frac{dn}{dx}. \quad (21)$$

Плотность диффузионного тока равняется плотности потока носителей, определяемой уравнением (18), умноженной на заряд электрона.

При вычислении диффузионного тока следует помнить о заряде носителей, поведение которых изучается. Если их заряд отрицательный (электроны), то, хотя поток электронов будет направлен в сторону убывания концентрации носителей, диффузионный ток электронов будет направлен в противоположную сторону, в сторону возрастания концентрации: $j_{D_n} = |e| D_n \frac{dn}{dx}$. В этой формуле D_n — коэффициент диффузии электронов.

Для дырок (положительный заряд) диффузионный ток направлен в ту же сторону, что и поток дырок — в сторону уменьшения концентрации: $j_{D_p} = -|e| \cdot D_p \frac{dp}{dx}$, D_p — коэффициент диффузии дырок.

Диффузионный ток — это самый настоящий ток, ничем не уступающий хорошо нам знакомому току, возникающему под действием электрического поля E . При прохождении диффузионного тока выделяется джоулево тепло. Так же, как и ток, возникающий под действием поля (полевой), диффузионный ток вызывает отклонение магнитной стрелки. Чтобы уметь вычислять диффузионный

ток, необходимо знать коэффициенты диффузии электронов и дырок: D_n и D_p .

Между коэффициентом диффузии любых частиц и их подвижностью существует простое соотношение. Перепишем выражение (20) в виде

$$D = \frac{1}{3} v_T^2 \tau_0 \quad (22)$$

и подставим в него значение квадрата тепловой скорости из формулы (11):

$$D = \frac{1}{3} v_T^2 \cdot \tau_0 = \frac{kT}{m} \cdot \tau_0. \quad (23)$$

Сравнив теперь выражение (23) с формулой (15) для подвижности частиц, получим

$$D = \frac{kT}{e} \mu. \quad (24)$$

Зная величину подвижности при данной температуре, из формулы (24) легко найти значение коэффициента диффузии.

Связь между коэффициентом диффузии и подвижностью, устанавливаемая соотношением (24), справедлива не только для электронов и дырок, но и для любых частиц, заряженных и незаряженных*), движущихся в поле силы тяжести или в электрическом поле. Эта связь является наглядным отражением того обстоятельства, что и направленному движению частиц под действием силы, и процессу диффузии мешает один и тот же процесс: столкновения частиц, происходящие через средний промежуток времени τ_0 при средней тепловой скорости частиц v_T .

Связь между коэффициентом диффузии и подвижностью установлена Альбертом Эйнштейном, и соотношение (24) называется соотношением Эйнштейна.

Подставив в формулу (24) численные значения постоянной Больцмана k и заряда электрона e , нетрудно убедиться, что при комнатной температуре (300 К) величина kT/e равняется приблизительно 0,026 В. Таким образом (см. табл. 2), при комнатной температуре коэффициент

*) Для частиц, не обладающих электрическим зарядом, подвижность μ определяют как отношение средней скорости частиц \bar{v} к силе F , под действием которой устанавливается эта скорость: $\mu = \bar{v}/F$. В этом случае связь между коэффициентом диффузии и подвижностью имеет вид $D = kT\mu$.

диффузии электронов в InSb примерно равен $0,2 \text{ м}^2/\text{с}$, в Ge — $0,01 \text{ м}^2/\text{с}$, в GaAs — $0,025 \text{ м}^2/\text{с}$ и т. д.

Обсудим теперь важный вопрос о скорости диффузии. Или, иными словами, определим время t , за которое диффундирующие частицы пройдут расстояние L .

Для решения задачи снова воспользуемся методом размерностей.

Поскольку ответ может, очевидно, определяться только величинами L (с размерностью м) и D (с размерностью $\text{м}^2/\text{с}$), то после некоторого размышления можно убедиться, что единственный способ получить величину, имеющую размерность времени, состоит в том, чтобы разделить квадрат длины L на коэффициент диффузии D . Таким образом, на поставленный вопрос метод размерностей дает следующий ответ:

$$t \sim \frac{L^2}{D},$$

или

$$L \sim \sqrt{Dt}. \quad (25)$$

Соотношение (25) устанавливает очень необычную зависимость между временем диффузии и пройденным за это время путем. Путь при диффузии пропорционален не времени t , а корню квадратному из величины t . Поэтому на первый взгляд может показаться, что диффузия — процесс медленный.

Прделаем, однако, несколько численных оценок. Современные полупроводниковые приборы часто имеют размеры порядка микрометра (10^{-6} м) или даже долей микрометра. За какое время электрон, двигаясь за счет диффузии, пройдет расстояние, равное 10^{-6} м? Как мы знаем, при 300 К коэффициент диффузии электронов в GaAs $D_n = 0,025 \text{ м}^2/\text{с}$. Таким образом, искомое время составит $t \approx L^2/D_n \approx 4 \cdot 10^{-11}$ с. Четыре стомиллиардных доли секунды! Не правда ли, процесс диффузии в таких маленьких приборах не кажется медленным?

Во всех типах диодов и транзисторов свободные носители заряда распределены неравномерно по объему прибора, причем характерная длина L , на которой значение концентрации меняется от очень большой до очень маленькой величины, часто составляет десятые и даже сотые доли микрометра. Вот почему процессы диффузии играют такую важную роль в работе полупроводниковых приборов.

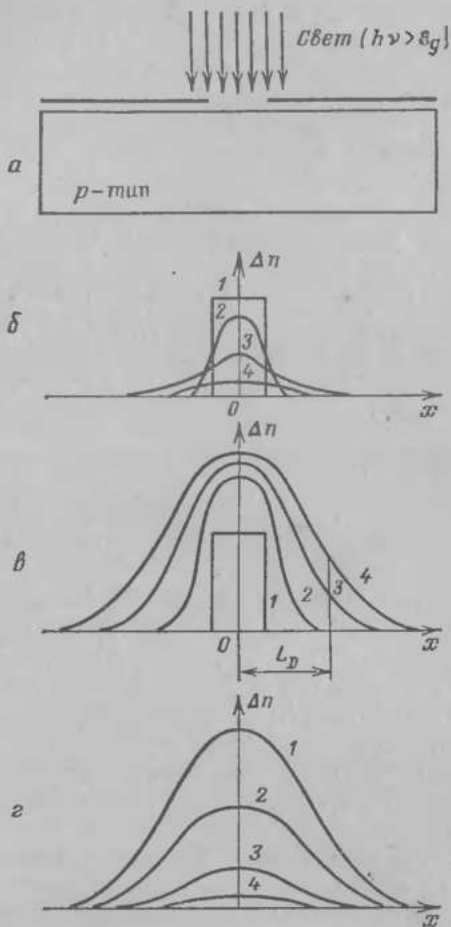


Рис. 23. Носители, возникшие в освещенной области полупроводника, распространяются в область тени, благодаря диффузии: а — образец закрыт металлической пластиной с прорезанной в ней щелью; б — распределение избыточных носителей при короткой вспышке света в моменты времени $t_1 \div t_4$ ($t_1 < t_2 < t_3 < t_4$); в — распределение избыточных носителей в моменты времени $t_1 \div t_4$ ($t_1 < t_2 < t_3 < t_4$) при включении света (свет остается включенным очень долго); г — распределение избыточных носителей при включении света ($t_1 < t_2 < t_3 < t_4$)

Диффузия неравновесных носителей. В некоторых типах приборов самую важную роль играет особый вид диффузии — диффузия избыточных (неравновесных) носителей. Особенность этого вида диффузии в том, что избыточные носители, диффундируя, удаляясь от того места, где они введены в полупроводник, гибнут (рекомбинируют).

Изучая в предыдущем разделе процессы рекомбинации, мы установили, что если создать в полупроводнике, например, освещая его, избыточные, неравновесные носители, то после того, как источник света выключен, концентрация избыточных носителей будет уменьшаться. Характерное время, за которое избыточные носители рекомбинируют, — время жизни τ , зависит от того, в каком полупроводнике они созданы, от типа и концентрации глубоких примесей, содержащихся в материале, и лежит в пределах от $\sim 10^{-10}$ до $\sim 10^{-2}$ с.

Обсудим опыт, схема которого показана на рис. 23.

Полупроводниковый образец освещается светом с энергией кванта $h\nu$, большей, чем энергия образования электронно-дырочной пары \mathcal{E}_g . Поверхность образца закрыта металлической пластиной с прорезанной в ней щелью. Свет сквозь металл не проходит, так что электронно-дырочные пары возникают под действием света только непосредственно под щелью (рис. 23, а).

Включим свет и тут же выключим его (рис. 23, б). Допустим, что вспышка была такой короткой, что возникшие под действием света электроны не успели ни расплыться под влиянием диффузии, ни рекомбинировать. В этом случае распределение электронов сразу же после окончания импульса света будет соответствовать кривой 1. Что произойдет с неравновесными избыточными носителями в дальнейшем?

Во-первых, как мы знаем, неравновесные носители будут рекомбинировать (вымирать) с характерным временем жизни τ . Во-вторых, перемещаться под действием диффузии из области высокой концентрации в область низкой (расплываться). Кривые 2—4 (рис. 23, б) показывают, что происходит с пакетом избыточных носителей благодаря совместному действию процессов рекомбинации и диффузии.

Электроны, возникшие под действием света, рекомбинируют за время порядка τ . В соответствии с выражением (25) это означает, что, умирая, они успевают продвинуться из освещенной области, где они возникли, в область тени

на расстояние порядка диффузионной длины L_D :

$$L_D = \sqrt{D\tau}, \quad (26)$$

Проведем теперь следующий опыт. Включим свет и оставим его включенным неопределенно долго (рис. 23,а). Сразу же после включения света неравновесные электроны существуют только в освещенной области (кривая 1). Затем электронный пакет начинает расплываться благодаря диффузии, однако процесс генерации носителей светом преобладает над процессом рекомбинации. Концентрация в освещенной области продолжает возрастать (кривые 2 и 3). Кривая 4 соответствует стационарному состоянию. Распределение электронов, соответствующее кривой 4, будет сохраняться неизменным, пока свет останется включенным. Число носителей, создаваемых светом, в точности равно числу носителей, погибающих за счет рекомбинации. В каждой точке число носителей, доставляемых диффузией из соседней области более высокой концентрации, равняется сумме числа носителей, уходящих из этой точки в соседнюю область более низкой концентрации, и числа рекомбинирующих носителей.

Созданные в освещенной области неравновесные носители проникают в область тени на расстояние порядка нескольких диффузионных длин L_D .

Рис. 23,б показывает, что произойдет, если выключить свет. Неравновесные носители рекомбинируют с характерным временем τ .

Moralité

Под влиянием тепловых колебаний в полупроводнике непрерывно генерируются свободные носители заряда — электроны проводимости и дырки. Также непрерывно электроны и дырки рекомбинируют. Под действием этих двух процессов — генерации и рекомбинации — в полупроводнике при заданной температуре существует определенная равновесная концентрация свободных носителей. Значение равновесной концентрации электронов и дырок в данном полупроводнике зависит от температуры и от типа и концентрации введенной примеси.

Определенные внешние воздействия, например освещение, могут увеличить уровень концентрации электронов и дырок относительно равновесного значения.

При устранении внешнего воздействия концентрация избыточных, неравновесных носителей уменьшается, стремясь к равновесному значению. Скорость рекомбинации избыточных носителей характеризуется временем жизни носителей τ . Значение τ резко уменьшается при введении в полупроводник примесей, обуславливающих возникновение глубоких уровней.

Глубокие уровни увеличивают не только скорость рекомбинации, но и скорость генерации носителей.

Свободные носители находятся в состоянии беспорядочного хаотического движения, происходящего с большой скоростью. При обычных условиях средняя скорость такого хаотического движения $\sim 10^5$ м/с. Под действием электрического поля на это хаотическое движение накладывается направленный дрейф носителей в электрическом поле. Средняя скорость дрейфа в слабых полях пропорциональна напряженности электрического поля (закон Ома). В сильном поле дрейфовая скорость носителей насыщается. Если концентрация свободных носителей распределена в полупроводнике неоднородно, возникает диффузия — поток носителей из области с большей в область с меньшей концентрацией.

БАРЬЕРЫ

И можно свернуть, обрыв
 обогнуть,
 Но мы выбираем трудный путь,
 Опасный, как военная тропа.

В. Высоцкий

На трудном пути к перевалу, ведущему в цветущую Долину Транзисторов, мы одолели половину подъема. И, как говорят альпинисты и летчики, «набрали приличную высоту». Оглянувшись назад, с этой высоты можно увидеть многое. Мы могли бы, например, детально проанализировать работу таких полупроводниковых приборов, как фоторезисторы, тензорезисторы. Набранная высота позволяет рассчитать все основные параметры этих приборов, познакомиться с их многочисленными разновидностями и с почти бесчисленными практическими приложениями.

Если пройти чуть в сторону и вскарабкаться на соседние скалы, то с них без особого труда можно разглядеть во всех подробностях такие важные и интересные приборы, как полупроводниковые датчики Холла и диоды Ганна. Любуясь красочными и интересными пейзажами, на уже набранной высоте можно было бы приятно провести все оставшееся время.

Мы, однако, «выбираем трудный путь» и продолжаем подъем к перевалу. На нашем пути встают два барьера. Один из них — $p-n$ переход, который является основой самых важных и интересных полупроводниковых приборов. Роль другого — границы между полупроводником и окружающей средой (например, воздухом, вакуумом) — на первый взгляд гораздо скромнее. Тем не менее мы начнем подъем с преодоления именно этого барьера: во-первых, он очень интересен в спортивном отношении, во-вторых, не справившись с ним, к главному барьеру мы просто не сможем подойти.

Глава 3

БАРЬЕР НА ГРАНИЦЕ КРИСТАЛЛА

Как веревочке ни виться,
А конца не миновать.

Размеры кристаллов, с которыми имеет дело современная полупроводниковая электроника, различаются очень сильно. Самые большие полупроводниковые приборы — мощные тиристоры, представляют собой кремниевые диски диаметром более 8 см и толщиной около 1 мм (рис. 24). Рабочий объем такого прибора $\sim 6 \text{ см}^3$. Один

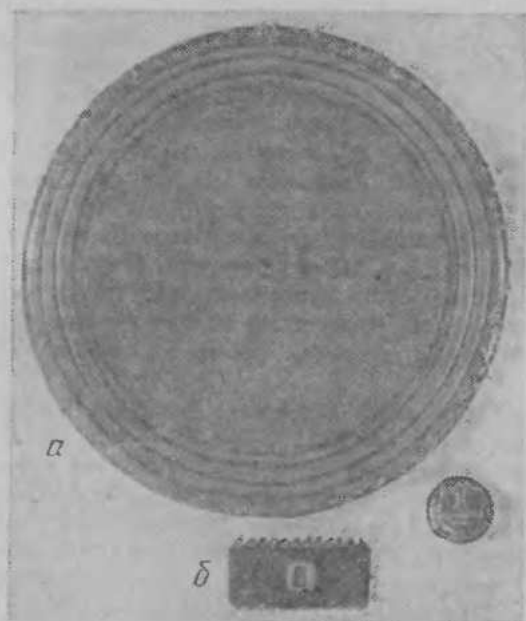


Рис. 24. Мощный тиристор и интегральная схема. *a* — Тиристор. В металлокерамическом корпусе этого тиристора размещается главный элемент прибора, имеющий вид кремниевого диска диаметром ~ 8 см и толщиной ~ 1 мм. *б* — На пластинке кремния размером 3×5 мм (пластина видна сквозь стеклянное окошко) в этой интегральной схеме располагаются более 100 000 кремниевых транзисторов

из самых маленьких приборов — современный арсенид-галлиевый сверхвысокочастотный транзистор, имеет размеры $\sim 0,01 \times 0,5 \times 10$ мкм. Его рабочий объем составляет, таким образом, $\sim 5 \cdot 10^{-14}$ см³. Объемы полупроводникового материала, с которыми работают создатели приборов, различаются, как мы видим, на 14 порядков — в сотни тысяч миллиардов раз. Однако велик кристалл или мал, он всегда имеет конец, границу. И на этой границе разыгрываются удивительные явления.

Работа выхода



Задумаемся прежде всего над следующим обстоятельством. Пусть перед нами лежит кусочек полупроводника или металла и мы знаем, что в нем существует некоторое количество свободных электронов (электронов проводимости). Если, скажем, перед нами кусок металла, то в каждом его кубическом сантиметре существует $\sim 10^{22}$ свободных электронов. С другой стороны, в окружающей среде (воздухе) свободных электронов практически нет (рис. 25, а). Познакомившись с явлением диффузии, мы понимаем, что по всем канонам и правилам должен возникнуть диффузионный поток электронов из металла, где их много, в воздух, где их практически нет. И подобно тому, как испаряется лужица пролитого одеколона, свободные электроны из металла или полупроводника через некоторое время должны были бы исчезнуть, «испариться».

Больше того, мы легко можем рассчитать время, за которое электроны должны были бы уйти из кристалла. Это время по порядку величины равняется времени диффузии электронов через кристалл $t_D \approx L^2/D$, где L — размер кристалла, а D — коэффициент диффузии. В справочниках обычно приводятся значения подвижности электронов в материалах. Скажем, для золота (Au) подвижность $\mu = 30$ см²/(В·с). Используя соотношение Эйнштейна (24), найдем, что коэффициент диффузии электронов в Au равняется приблизительно 0,75 см²/с. Таким образом, из кусочка золота размером в сантиметр электроны проводимости должны были бы исчезнуть за счет диффузии за время порядка $t_D = 1,3$ с.

Ничего подобного, разумеется, в действительности не происходит. И золотые украшения Трои, которым око-

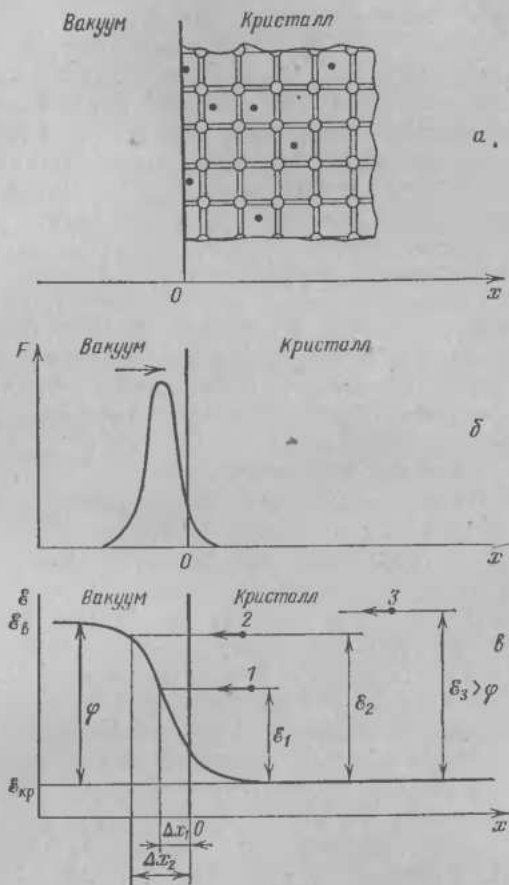


Рис. 25. Барьер на границе кристалла: а — электроны проводимости «заперты» в кристалле и не могут выйти наружу; б — на границе кристалла существует сила, удерживающая электроны в кристалле; в — чтобы извлечь электрон из кристалла, необходимо преодолеть энергетический барьер высотой $\varphi = \epsilon_{\text{в}} - \epsilon_{\text{кр}}$, т. е. затратить работу, равную работе выхода φ . (Электрон с кинетической энергией ϵ_1 сумеет проникнуть за границу кристалла на расстояние не большее, чем Δx_1 , затем удерживающая сила F вновь вернет его в кристалл. Электрон с кинетической энергией ϵ_2 не сможет отойти от границы на расстояние, большее Δx_2 . Электрон с кинетической энергией $\epsilon_3 > \varphi$ может покинуть кристалл)

ло 3 тысяч лет, и даже золотые самородки, пролежавшие на воздухе миллионы лет, содержат электроны проводимости в точности в той же концентрации, что и золотой слиток, только что вынутый из плавильной печи.

Что же мешает электронам уйти из металла? Размышляя, мы неизбежно приходим к выводу, что на границе между кристаллом и окружающей средой (будем пока что для простоты под этой окружающей средой иметь в виду вакуум) на электрон должна действовать сила, препятствующая попыткам электрона покинуть кристалл.

Эта сила действует на электроны вблизи границы и, как в глубине кристалла, так и снаружи, на некотором расстоянии от границы исчезает (рис. 25, б). Ясно, что сила, удерживающая заряженную частицу — электрон, представляет собой электрическое поле. Таким образом, на рис. 25, б фактически показана качественная картина распределения напряженности электрического поля вблизи поверхности кристалла.

Это же самое представление о причине, мешающей электронам вылетать из кристалла наружу, можно очень наглядно выразить на языке энергетической диаграммы (рис. 25, в).

Представим себе, что изнутри кристалла к границе подлетает электрон, скорость которого направлена перпендикулярно границе. Электрон стремится покинуть кристалл. Но как только электрон попадает в область, где действует удерживающая его в кристалле сила $F = eE$ (рис. 25, б), его скорость начинает уменьшаться. Ведь если частица движется в области, где на нее действует сила F , направленная навстречу движению, то, чтобы пройти в этой области расстояние Δx , она должна совершить работу $\Delta A = F \cdot \Delta x = eE \cdot \Delta x$. На эту же величину ΔA уменьшится кинетическая энергия частицы.

Если кинетическая энергия \mathcal{E} подлетевшего к границе электрона невелика, он очень быстро растратит ее, двигаясь в поле тормозящей силы $F = eE$. Скорость его упадет до нуля. Затем электрон повернет обратно, ускорится под действием той же силы и... влетит обратно в кристалл.

Чем больше кинетическая энергия налетающего на границу электрона, тем дальше он сумеет проникнуть в область, где действует тормозящая сила F . Наконец, можно представить себе, что найдется такой энергичный электрон, который сумеет пройти сквозь всю область, где существует тормозящая сила F , и... вылететь наружу. Ми-

нимальная кинетическая энергия, которой должен обладать электрон, чтобы совершить этот подвиг, называется *работой выхода* и обозначается обычно греческой буквой ϕ (рис. 25, *е*). Эта работа, естественно, равна разности между энергией покоящегося электрона в вакууме $\mathcal{E}_в$ и энергией электрона, покоящегося в кристалле $\mathcal{E}_{кр}$. Таким образом, чтобы покинуть кристалл, электрон должен преодолеть энергетический барьер высотой ϕ .

Как возникает сила, удерживающая электрон в кристалле, или, иными словами, какова природа работы выхода?

Вернись, я вновь и вновь зову — вернись. Первые работы, в которых сделана попытка объяснить физическую природу работы выхода, были выполнены в начале нашего века двумя замечательными учеными: немецким — Вальтером Шоттки и американским — Ирвингом Ленгмюром. Идея этих работ очень проста.

Как только электрон вылетает за пределы кристалла, кристалл, до того бывший электрически нейтральным, заряжается положительно. И, естественно, начинает притягивать отрицательно заряженный электрон, стремясь вернуть его обратно (рис. 26, *а*). Чтобы преодолеть это притяжение и улететь, покинувший кристалл электрон должен обладать определенной кинетической энергией. Ее минимальное значение и будет, очевидно, равно работе выхода электрона из кристалла.

Такой механизм формирования энергетического барьера на границе должен, очевидно, присутствовать во всех материалах. И в металлах, и в полупроводниках, и в диэлектриках. Но проявляется этот механизм в разных материалах по-разному.

В металлах, как известно из школьного курса электростатики, электрическое поле в глубь кристалла не проникает. Из-за громадной концентрации в металле свободных электронов положительный заряд, притягивающий улетевший электрон, создается перегруппировкой электронов в очень тонком поверхностном слое. В институтских курсах электростатики доказывается, что действие этого поверхностного положительного заряда на улетевший электрон математически полностью эквивалентно притяжению со стороны положительно заряженной частицы с зарядом, по абсолютной величине равным заряду электрона и находящимся от улетевшего электрона на расстоянии $2\Delta x$. Здесь Δx — расстояние от улетающего электрона до поверхности металла (рис. 26, *б*).

Короче говоря, рассчитать силу, действующую со стороны металла на электрон-беглец, можно очень просто. Нужно только построить зеркальное изображение улетевшего электрона в «металлическом зеркале» и заменить у изображения знак «—» на «+». Поэтому принято говорить, что в металлах рассматриваемый механизм работы выхода обязан своим возникновением *силам зеркального изображения*.

В диэлектриках, где свободных электронов очень мало (практически нет), положительный заряд, действующий

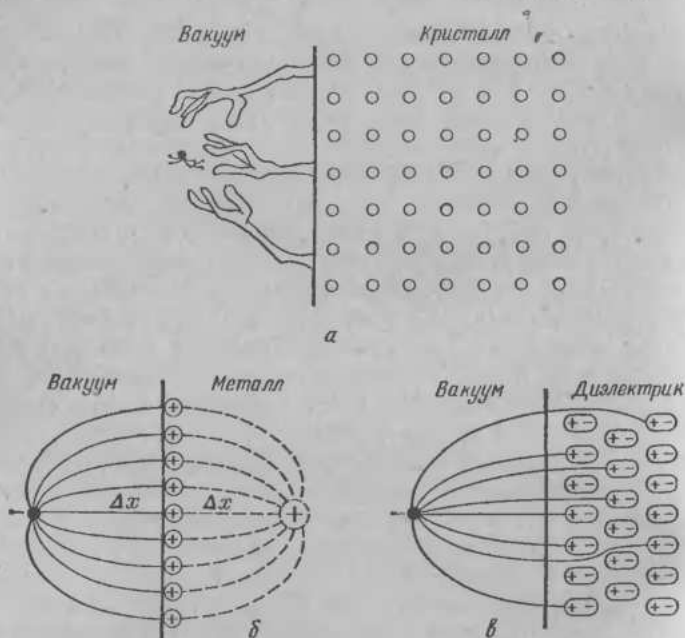


Рис. 26. а — Вернись, я вновь и вновь зову — вернись. б — Силы «зеркального изображения» в металле. в — Поляризационные силы в диэлектрике

на улетающий электрон, создается поляризацией молекул диэлектрика (рис. 26, в). Поэтому в диэлектриках (рис. 26, в) рассматриваемый компонент сил, удерживающих электрон в кристалле, часто называется поляризационными силами.

В полупроводниках относительный вклад того или иного компонента зависит от строения полупроводникового кристалла и концентрации в нем электронов.

Ленгмюр и Шоттки рассчитали ожидаемую величину работы выхода для металлов. Оказалось, что для того чтобы преодолеть силы зеркального изображения и улететь, электрону требуется энергия порядка 1—2 эВ. Примерно такую же высоту энергетического барьера, создаваемого поляризационными силами, дает расчет для диэлектриков и полупроводников. Однако наличие поляризационных сил не единственная причина, мешающая электрону покинуть кристалл.

Двойной заряженный слой. Электроны проводимости непрерывно атакуют границу между твердым телом и вакуумом, стремясь вылететь наружу. Конечно, кинетическая энергия большинства электронов недостаточно велика, чтобы преодолеть барьер на границе и покинуть кристалл. Но благодаря этим атакам за пределами кристалла в каждый данный момент существует отрицательно заряженное электронное облако, а внутри — нескомпенсированный положительный заряд (рис. 27). На границе возникает *двойной заряженный слой*, и на электрон, влетающий в этот слой, действует сила $F = eE$, стремящаяся не выпустить электрон за пределы кристалла.

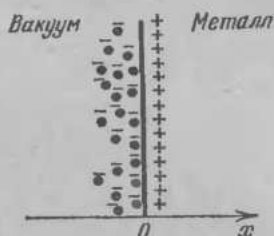


Рис. 27. Двойной заряженный слой на границе металла и вакуума

Изучая барьеры, мы очень часто будем иметь дело с двойными заряженными слоями. Поэтому обсудим немного подробнее действие такого слоя на заряженную частицу (рис. 28). В целом, двойной слой электрически нейтрален. Отрицательный заряд, сосредоточенный в нем, в точности равен положительному (рис. 28, а). Поэтому, скажем, электрон, находящийся вне заряженного слоя (электрон 1 на рис. 28, а), не испытывает с его стороны ни притяжения, ни отталкивания. На электрон 2, попавший в заряженный слой, действуют сила притяжения со стороны положительного заряда справа и сила отталкивания слева (слева у этого электрона отрицательных зарядов больше, чем положительных). Максимальная сила действует, очевидно, на электрон 3. Его отталкивают вправо все отрицательные и тянут в ту же сторону все положительные заряды. Сила, действующая на электрон 4, очевидно, меньше. А на электрон 5, находящийся вне слоя, никакие силы со стороны слоя не действуют.

Ясно, что качественно распределение напряженности электрического поля внутри заряженного слоя будет таким, как на рис. 25, б.

Иногда, не интересуясь распределением поля внутри слоя, можно в грубом приближении считать, что в пределах некоторой области на заряженную частицу действует постоянная сила $F =$

$= eE$. В таком приближении действие двойного заряженного слоя эквивалентно действию конденсатора (рис. 28, б).

Как определяется работа выхода и чему она равна. Изучение явлений на границе кристаллов, приводящих к формированию энергетического барьера и работы выхода, ведется, как уже упоминалось, с начала века. Почти

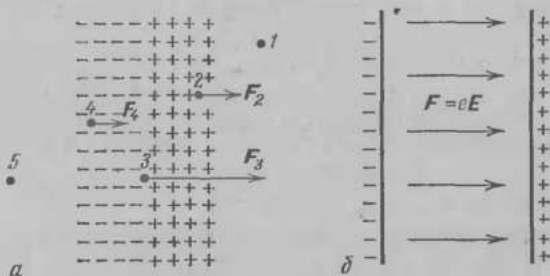


Рис. 28. В целом двойной слой электрически нейтрален: а — на электроны вне двойного слоя (электроны 1 и 5) никакие силы со стороны слоя не действуют, максимальная сила действует на электрон 3; б — действие на заряженную частицу со стороны двойного слоя в грубом приближении можно заменить действием эквивалентного конденсатора с соответствующим зарядом на обкладках

сто лет. И очень усердно продолжается и в настоящее время. Это вызвано не только естественной человеческой любознательностью и желанием понять интересное и сложное явление, но и настоятельной практической потребностью.

Более тридцати лет, с 20-х по 50-е годы нашего столетия, основой электроники была электронная лампа, а основным элементом электронной лампы — накаливаемый катод (источник электронов). Именно катод определяет долговечность и надежность ламп (в частности, кинескопов наших телевизоров), силу тока, который можно пропустить через лампу, и многие другие важные параметры. А параметры катода, в свою очередь, в решающей степени определяются величиной работы выхода материала, из которого он изготовлен.

Процессы на границе твердых тел и окружающей среды лежат в основе многих важнейших химических реакций. Величина работы выхода твердого тела, участвующего в реакции, зависимость работы выхода от температуры и свойств окружающей среды часто определяют не только скорость и направление химической реакции, но и саму возможность ее протекания.

Наконец, свойства границы раздела, поверхности между кристаллом и окружающей средой часто определяют самые важные параметры полупроводниковых приборов. Некоторые полупроводниковые приборы так тонки — их толщина составляет всего лишь несколько атомных слоев, — что весь прибор практически представля-

ет собой «поверхность». Самые современные полупроводниковые приборы — интегральные схемы — выполняются на полупроводниковой пластине, в слое, лежащем всего на несколько микрон или даже на нескольких десятых долей микрометра под поверхностью полупроводника.

Подробные исследования свойств поверхности показали, что в формировании энергетического барьера на поверхности, помимо двух рассмотренных механизмов: поляризационных сил и двойного заряженного слоя, участвует еще и ряд других сложных и тонких механизмов. Некоторые из них до конца не поняты и в настоящее время. Поэтому рассчитать работу выхода для конкретного материала невозможно. Зато ее можно измерить.

Существует несколько способов измерения величины работы выхода. Один из них известен из школьного курса физики: величину работы выхода ϕ можно определить по «красной границе» внешнего фотоэффекта.

Если направить на поверхность металла или полупроводника поток света с достаточно малой длиной волны (большой энергией фотонов \mathcal{E}_ϕ), то с поверхности материала будут вылетать электроны. Скорость вылетевших электронов будет тем больше, чем больше энергия фотонов \mathcal{E}_ϕ . Этот эффект, открытый в 1887 г. немецким физиком Генрихом Герцем, был объяснен в 1905 г. Альбертом Эйнштейном на основе квантовых представлений о природе света. Энергия кванта света (фотона), поглощенного электроном, расходуется на работу выхода ϕ электрона и на приобретение электроном кинетической энергии $mv^2/2$:

$$\mathcal{E}_\phi = \phi + \frac{mv^2}{2}.$$

Если увеличивать длину волны света (уменьшать энергию \mathcal{E}_ϕ), то скорость вылетевших электронов уменьшается. И, наконец, при некоторой критической энергии фотонов \mathcal{E}_ϕ электроны вообще перестают вылетать. Внешний фотоэффект исчезает. Величина \mathcal{E}_ϕ равняется, очевидно, работе выхода ϕ .

Другой широко используемый способ измерения ϕ состоит в исследовании термоэлектронной эмиссии из изучаемого материала.

Суть метода термоэлектронной эмиссии поясняется рис. 29, сходным с рис. 25, а. Он еще раз напоминает, что подавляющее большинство электронов имеет внутри кристалла энергию $\mathcal{E}_{кр}$ и встречает на границе кристалла энергетический барьер высотой $\phi = \mathcal{E}_в - \mathcal{E}_{кр}$. Электроны в металле находятся в энергетической яме глубины ϕ .

Мы знаем, однако, что при конечной температуре T всегда существует некоторое количество очень энергичных электронов, чья энергия значительно больше, чем $\mathcal{E}_{кр}$ (см. формулу (12) и рис. 11).

Электроны, кинетическая энергия которых ($\mathcal{E} - \mathcal{E}_{кр}$) превосходит работу выхода ϕ , могут покинуть кристалл и вылететь наружу. Концентрация таких «суперэнергичных» электронов n может быть вычислена по приближенной формуле $n \approx n_0 \exp(-\phi/kT)$.

Вылетевшие из образца электроны регистрируются следующим образом. Исследуемый металлический или полупроводниковый образец помещается в вакуумную камеру. Делается это для того,

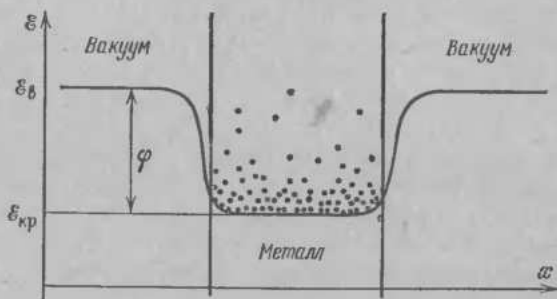


Рис. 29. Подавляющее большинство электронов в кристалле имеет энергию $\mathcal{E}_{кр}$. Эти электроны находятся на дне «энергетической ямы» глубиной ϕ . За счет хаотического теплового движения некоторые, очень немногие, электроны имеют гораздо большую энергию (ср. с рис. 11 и формулой (12))

чтобы движению вылетевших из образца электронов не мешали столкновения с молекулами воздуха. На образец подается «минус» батареи, а рядом с образцом помещается металлическая пластина, на которую подается «плюс». Под действием электрического поля, создаваемого батареями, электроны, вылетевшие из образца (катода), летят к пластине (аноду). Между электродами возникает ток, сила которого, очевидно, пропорциональна числу вылетающих из образца электронов:

$$I \sim \exp(-\phi/kT).$$

Далее снимается зависимость тока от температуры образца и строится зависимость $\ln I$ как функция $1/T$. Тангенс угла наклона этой зависимости к оси абсцисс определит работу выхода ϕ . Взгляните на рис. 5 и формулы (6), (7) и (10). Не правда ли, между определением работы выхода методом термоэлектронной эмиссии и определением энергии образования электронно-дырочной пары по температурной зависимости концентрации имеется определенное сходство? Повятына и глубокая физическая причина, по которой это сходство возникает: в обоих случаях высота энергетического барьера определяется по числу частиц, которые сумели преодолеть барьер за счет хаотического теплового движения.

Существуют и другие методы измерения работы выхода. И какой бы метод ни избрал экспериментатор для определения значения ϕ , одной из главных его забот будет самая тщательная очистка поверхности исследуемого материала от всевозможных примесей и загрязнений.

При исследовании металлов их перед началом измерений многократно нагревают в вакууме до высокой температуры. Посторонние атомы, адсорбированные *) поверхностью металла, при сильном нагреве отрываются от поверхности и удаляются с помощью откачки. Высокая температура способствует также разложению и удалению окисных пленок, возникающих на поверхности металла. С той же целью — очистить поверхность, металлические образцы подвергаются бомбардировке электронным пучком или ионной бомбардировке. Иногда приходится прибегать к многократному плавлению и последующему испарению и конденсации в высоком вакууме.

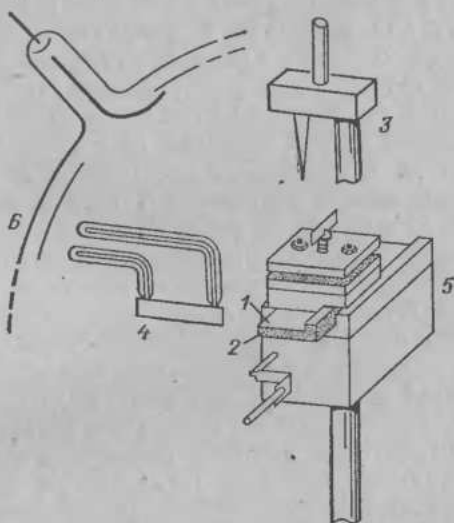


Рис. 30. Устройство для получения чистой поверхности полупроводников методом раскалывания монокристалла в вакууме: 1 — кремниевый образец (кристалл кремния); 2 — поверхность скола; 3 — игла для нанесения царапины («скрайбер»); 4 — вольфрамовый нож, которым производится скалывание; 5 — держатель; 6 — вакуумная колба

Для полупроводников самым надежным методом получения чистой поверхности является раскалывание монокристалла полупроводника в высоком вакууме. Монокристаллы закрепляют в держателе внутри вакуумной колбы (рис. 30). Колбу откачивают. Затем на поверхность

*) Слово «адсорбция» образовано от латинского глагола *sorbere* (глотать) при помощи приставки *ad* (при, к). Оно означает поглощение поверхностью веществ посторонних чужеродных атомов.

кристалла острой иглой наносится царапина и нажатием вдоль царапины кристалл «скалывается». Все операции производятся в вакууме при помощи управляемых снаружи магнитных манипуляторов. Таким методом удается получить идеально гладкую и чистую боковую поверхность кристалла.

Измерение работы выхода производится немедленно после получения свежего скола. Если повторить измерение через час, результат может оказаться совсем иным: хотя в колбе все время поддерживается вакуум, остаточные атомы, неизбежно присутствующие в колбе, оседают на поверхности полупроводника, адсорбируются ею и могут сильно исказить результаты измерений.

Так вот, для металлов в результате кропотливых, тщательных измерений, проделанных и многократно повторенных десятками исследовательских групп, значения работы выхода установлены достаточно точно. В справочниках величина ϕ для металлов не только указывается с точностью до третьей значащей цифры, но для целого ряда металлов иногда указывается, с какой именно грани монокристалла определена работа выхода электрона. Для разных граней величина оказывается несколько различной, и экспериментаторы уверенно измеряют эту, подчас очень небольшую разницу.

Работа выхода в металлах составляет несколько электрон-вольт и лежит в пределах от $\sim 2,1$ эВ (для лития, калия и цезия) до $\sim 5,4$ эВ (для платины). В хорошо всем нам знакомом вольфраме (из вольфрама изготовлены нити накала электрических лампочек) величина ϕ для разных граней составляет $4,3$ — $5,3$ эВ. Для железа $\phi \approx 4,7$ эВ, для золота $\approx 5,3$ эВ.

Иногда величина работы выхода приводится не в электрон-вольтах, а просто в вольтах. Скажем, в некоторых справочниках можно встретить значенье работы выхода ϕ для железа $4,7$ В, для золота $5,3$ В и т. д. Разумеется, говоря строго, составители таких справочников действуют неправильно. Нельзя энергию выразить в единицах разности потенциалов. Но эта неточность, если над ней задуматься, может принести некоторую пользу — прямо по поговорке «нет худа без добра».

В самом деле, эта неточность напоминает нам, что энергию в 1 эВ приобретает электрон, прошедший разность потенциалов в 1 В. Таким образом, ясно, что если разность энергий (высота барьера) между двумя точками ϕ измерена в электрон-вольтах, то она численно равна разности потенциалов U в вольтах между этими точками.

А как обстоит дело с полупроводниками?

С полупроводниками, как всегда, дело обстоит непросто. На рис. 31 показана хорошо нам знакомая энергетическая диаграмма полупроводника. Как всегда, через \mathcal{E}_c обозначена нижняя граница зоны проводимости, через

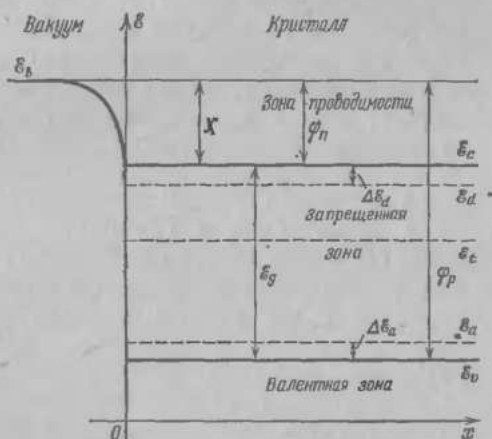


Рис. 31. В полупроводнике электроны могут находиться на самых разных энергетических уровнях. Поэтому работа выхода электрона из полупроводникового кристалла сильно зависит от того, какими примесями легирован кристалл, от уровня легирования и от температуры

\mathcal{E}_v — верхняя граница валентной зоны. Показаны уровни мелких доноров и акцепторов \mathcal{E}_d и \mathcal{E}_a и глубокий уровень \mathcal{E}_t .

Пусть энергия электрона в вакууме равняется $\mathcal{E}_{вк}$, энергия электрона в кристалле полупроводника равняется...

А чему она, собственно говоря, равняется?

В металле все было понятно. Там все электроны всегда были свободны. А в полупроводнике?

В полупроводнике, как мы знаем, ситуации могут быть самыми разными. В полупроводнике *n*-типа при достаточно высокой температуре имеется много электронов проводимости и подавляющее большинство этих электронов имеет энергию $\approx \mathcal{E}_c$ (см. рис. 11 и формулу (12)). Естественно ожидать, аналогично тому, как обстояло дело в металлах, что работа выхода будет равняться в этом случае $\mathcal{E}_v - \mathcal{E}_c$ (рис. 31).

Если охлаждать кристалл *n*-типа, то концентрация электронов в зоне проводимости будет убывать: все боль-

ше и больше электронов будет оседать, связываться на донорных центрах. При $T = 0$, как мы знаем, электронов проводимости нет; все атомы донорной примеси нейтральны и удерживают при себе «лишние» электроны. (Перечитайте еще раз пояснение к рис. 7, е). Итак, наивысший энергетический уровень, на котором в этом случае имеются электроны, — это донорный уровень \mathcal{E}_d . Чтобы вырвать электрон из кристалла, теперь необходима энергия $\mathcal{E}_v - \mathcal{E}_d$, большая, чем в предыдущем случае, на величину энергии ионизации донорного уровня $\Delta\mathcal{E}_d$. Положим, если донорный уровень мелкий и величина $\Delta\mathcal{E}_d$ мала, то разница будет не слишком заметной. Но ведь донорный уровень может быть и глубоким. Энергия его ионизации может составить десятки доли электрон-вольта. В этом случае работа выхода из полупроводника при низкой температуре будет заметно больше, чем при высокой.

А в полупроводнике p -типа? В таком полупроводнике свободных электронов проводимости практически нет при любой температуре. При низкой температуре все электроны связаны на межатомных орбитах и самый высокий энергетический уровень, на котором есть электроны, — это \mathcal{E}_v . При повышении температуры часть электронов переходит на акцепторный уровень \mathcal{E}_a , преодолевая энергетический барьер $\Delta\mathcal{E}_a$ (см. рис. 7, б). Чтобы в полупроводнике p -типа вырвать электрон из кристалла в вакуум и сделать его «по-настоящему свободным» (рис. 32), нужно затратить энергию, большую, чем в полупроводнике n -типа, приблизительно на величину \mathcal{E}_g — ширину запрещенной зоны (рис. 31, 32).

Поскольку в одном и том же полупроводнике работа выхода, как мы видим, сильно зависит от легирования и температуры, то, чтобы сравнивать разные полупроводники по величине работы выхода, вводят величину χ , называемую *электронным сродством* полупроводника. Как видно из рис. 31, $\chi = \mathcal{E}_v - \mathcal{E}_c$ и для данного полупроводника является величиной постоянной. Грубо говоря, величина χ приблизительно равняется работе выхода ϕ_n из полупроводника n -типа. Величина работы выхода из того же полупроводника p -типа, так же грубо говоря, больше на величину \mathcal{E}_g (рис. 31).

Обсуждая величину работы выхода полупроводника, вполне можно обойтись и без грубых оценок. Если известна концентрация и энергия ионизации примесных центров, введенных в полупроводник, величину работы

выхода можно точно рассчитать для любой температуры. Такие расчеты неоднократно проделывались, в частности, для германия и кремния и много раз сравнивались с экспериментальными результатами. И во многих случаях совпадение данных расчета и эксперимента было просто замечательным. Но ... всегда для *косвенных экспериментов*,

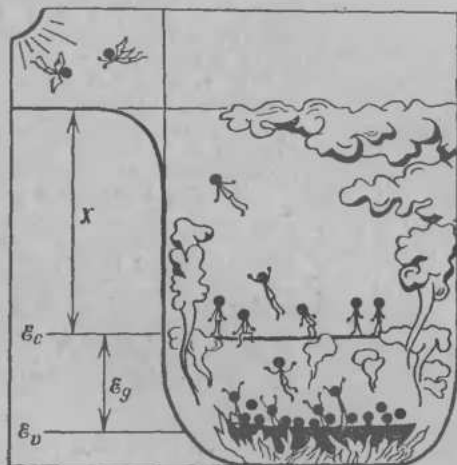


Рис. 32. Сравните этот рисунок с рис. 7. Чтобы перевести электроны в «рай» из «чистилища», нужно преодолеть энергетический барьер высотой χ . Переход из «сада» в «рай» требует преодоления энергетического барьера высотой $\chi + \epsilon_g$

т. е. для таких, в которых измеряется не непосредственное значение работы выхода, а какая-нибудь другая, связанная с работой выхода, величина.

В 1962 г. был выполнен очень изящный и остроумный эксперимент по прямому определению зависимости работы выхода в германии и кремнии от уровня и характера легирования. Выращивались длинные полупроводниковые кристаллы, в которых концентрация примесей и характер легирования от точки к точке менялись.

В принципе, вырастить такой монокристалл не так уж сложно. Можно, например, воспользоваться для этой цели одним из самых распространенных методов получения полупроводниковых монокристаллов — методом Чохральского, суть которого состоит в вытягивании монокристалла из расплава полупроводника. Основные принципы этого метода поясняются рис. 33. В расплав полупроводника сверху опускается на держателе небольшой монокристалл, так называемая затравка. Затем затравка приводится во вра-

щение и медленно поднимается над поверхностью расплава. Силы поверхностного натяжения заставляют некоторую часть жидкости следовать за затравкой. Оказавшись над поверхностью расплава, жидкость кристаллизуется и... становится затравкой для следующей порции жидкого полупроводника. Постепенно на держателе растет массивный монокристалл.

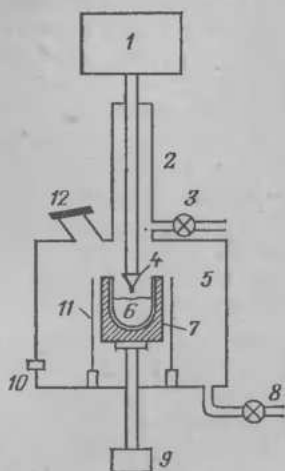


Рис. 33. Схема установки для выращивания полупроводниковых монокристаллов методом Чохральского: 1 — устройство подъема и вращения штока; 2 — внешняя камера; 3 — вход инертного газа; 4 — держатель затравки и затравка; 5 — высокотемпературная камера; 6 — расплав; 7 — тигель из кварца (SiO_2) или нитрида кремния (Si_3N_4); 8 — выход газа; 9 — устройство подъема и вращения тигеля; 10 — датчик температуры; 11 — нагреватели; 12 — смотровое окно

ты, можно рассчитать, как должна меняться работа выхода в кристалле от координаты. Результаты расчета показаны на рис. 34 штриховой прямой.

Затем, скалывая монокристалл слой за слоем (рис. 30), экспериментаторы измерили, как меняется работа выхода. Результат эксперимента, показанный на рис. 34 сплошной кривой, как мы видим, разительно отли-

Чтобы защитить растущий кристалл от загрязнений, выращивание производится в вакууме или в атмосфере инертного газа. Движение держателя вверх и вниз и его вращение производятся с помощью магнитов снаружи установки. Плавление исходного полупроводника производится методом СВЧ нагрева с помощью индуктора, также расположенного снаружи.

Именно с помощью метода Чохральского, а также описанного в гл. 1 метода бестигельной зонной плавки выращиваются громадные монокристаллы кремния, использующиеся для изготовления мощных тиристоров.

Понятно, как с помощью метода Чохральского можно вырастить монокристалл с переменным по длине легированием. Допустим, выращивается монокристалл кремния и исходный расплав содержит донорную примесь (например, мышьяк). Тогда вначале растущий монокристалл будет *n*-типа. По мере вытягивания будем добавлять в расплав акцепторную примесь. Вначале добавление акцептора приведет к частичной компенсации исходной донорной примеси (см. рис. 8). Затем, когда концентрации акцепторной и донорной примеси в расплаве сравняются, произойдет полная компенсация. Из такого расплава будет расти монокристалл со свойствами, близкими к свойствам собственного полупроводника. При дальнейшем добавлении акцепторов из расплава будет вытягиваться полупроводник *p*-типа.

Зная, как меняется концентрация примеси от координаты

чается от ожидаемого. Хотя точность измерений составляла сотые доли, а ожидаемое изменение — несколько десятых долей электрон-вольта, измеренная величина



Рис. 34. Зависимость работы выхода от координаты в монокристалле кремния с переменным по длине легированием: штриховая прямая — расчетная зависимость; сплошная кривая — зависимость, измеренная экспериментально в высоком вакууме послойным скалыванием кристалла (эксперимент Аллена и Гобели, 1962 г.)

работы выхода от координаты не зависела. В любой плоскости кристалла и n - и p -типа она приблизительно равнялась значению, ожидаемому для собственного кремния.

Что же означает этот удивительный результат?

Поверхностные состояния

По правде говоря, результаты Аллена и Гобели никакой сенсации в 1962 г. не вызвали. Они лишь особенно наглядно продемонстрировали тот факт, что сложные явления на границе раздела между полупроводником и окружающей средой не сводятся лишь к образованию на границе энергетического барьера.

Вторым важнейшим фактором, определяющим свойства полупроводника вблизи его границы, являются *поверхностные уровни* или, как их часто называют, *поверхностные состояния*.

С понятием об энергетических уровнях в запрещенной зоне полупроводника мы уже хорошо знакомы. Нам известны уровни донорные и акцепторные, мелкие и глубокие. Мы знаем, что уровни возникают не только при вве-

дении в полупроводник различных примесей. Различного рода структурные нарушения в идеальной периодической решетке полупроводника также могут вызвать появление глубоких или мелких примесных центров (примесных уровней).

Но ведь поверхность полупроводникового кристалла представляет собой, можно сказать, вопиющее нарушение в идеальной периодической решетке. Идеальный строй атомов, расположенных в объеме полупроводника на строго одинаковом друг от друга расстоянии, на поверхности внезапно обрывается (рис. 25, а). Ясно, что условия связи между валентными электронами, принадлежащими приповерхностным атомам, будут несколько иными, чем в объеме.

Уровни Тамма. В 1932 г. Игорь Евгеньевич Тамм, будущий лауреат Нобелевской премии по физике *)_x, теоретически показал, что даже если на поверхности полупроводника нет никаких посторонних атомов, одного только обрыва кристаллической решетки достаточно, чтобы вблизи поверхности возникли дополнительные энергетические уровни. Эти поверхностные состояния носят название *уровней Тамма*.

Несколько позднее, в 1939 г., исходя из совсем других соображений к аналогичному выводу пришел другой будущий лауреат Нобелевской премии — Уильям Шокли. Шокли обратил внимание на то, что у приповерхностных атомов существуют (из-за обрыва решетки) неспаренные валентные электроны (см. рис. 25, а). Условия, в которых находятся эти электроны, могут облегчить отрыв неспаренных электронов от приповерхностных атомов. В этом случае возникают поверхностные донорные состояния. Может возникнуть и обратная ситуация: неспаренные электроны вблизи поверхности приобретут способность связывать (захватывать) электроны проводимости, существующие в объеме кристалла. В этом случае возникают поверхностные акцепторные уровни.

Уровни Тамма существуют на так называемой «атомарно чистой» поверхности, т. е. идеально гладкой поверхности, лишенной каких-либо посторонних атомов и структурных дефектов.

Реальная поверхность. В обычных условиях, т. е. на воздухе, поверхность полупроводника всегда покрыта

*) И. Е. Тамму Нобелевская премия по физике была присуждена (совместно с П. А. Черенковым и И. М. Франком) в 1958 г. за открытие и объяснение эффекта Вавилова—Черенкова.

слоем окислов. Кроме того, на поверхности адсорбируются чужеродные атомы. Наконец, на поверхности всегда существуют структурные дефекты: искажения решетки, возникшие в процессе скалывания или разрезания кристалла, вакансии, т. е. пустые места в узлах решетки, атомы полупроводника, сместившиеся со своих законных мест в узлах решетки в междузлия. Все эти примесные и структурные нарушения также служат источником поверхностных состояний.

Совокупность или, как часто говорят, *спектр поверхностных состояний* реальной поверхности можно изменять, меняя условия окисления и способ обработки поверхности, предотвращая адсорбцию определенных атомов или, напротив, насыщая поверхность атомами определенного вещества.

Так же, как и объемные, поверхностные уровни могут быть донорными и акцепторными, мелкими и глубокими. Количественной мерой концентрации поверхностных уровней служит плотность поверхностных состояний, измеряемая в m^{-2} или в cm^{-2} .

Как мы знаем, в 1 см^3 любого твердого тела содержится $N \approx 10^{22}$ атомов. Это означает, что вдоль ребра куба длиной 1 см располагается $N^{1/3}$, а на поверхности грани $1\text{ см} \times 1\text{ см}$ — $N^{2/3} \approx 10^{15}$ атомов основного вещества. Таким образом, если на каждый атом на поверхности полупроводника приходится по примесному уровню, то такая ситуация соответствует плотности поверхностных состояний $\sim 10^{15}\text{ см}^{-2}$. Плотность поверхностных состояний на реальных поверхностях лежит в пределах от $\sim 10^9$ до $\sim 10^{15}\text{ см}^{-2}$.

Изгибы зон, поверхностный потенциал



В зависимости от того, какой тип поверхностных состояний — донорный или акцепторный — преобладает, поверхность будет либо отдавать электроны в объем полупроводника и заряжаться положительно, либо, наоборот, захватывать электроны из объема и заряжаться отрицательно. Допустим для определенности, что на поверхности полупроводника n -типа преобладают акцепторные состояния. Тогда поверхностные уровни будут захватывать электроны из прилегающей к поверхности части объема полупроводника. В результате поверхность зарядится отрицательно. А в приповерх-

постной области полупроводника возникнет слой, обедненный электронами, т. е. заряженный положительно. Таким образом, вблизи поверхности образуется двойной заряженный слой, поле которого будет отталкивать электроны в глубь кристалла. (Возникшая ситуация напоминает случай, показанный на рис. 27. Только отрицательный заряд теперь образуется не за счет электронов, вылетевших из кристалла, а за счет электронов, захваченных на поверхностные состояния.)

Поле двойного слоя максимально вблизи поверхности и совпадает по мере удаления от нее в глубь кристалла (рис. 35).

Рис. 35 полезно сравнить с рис. 31 и 21. На рис. 31 границы энергетических зон шли не изгибаясь до самой

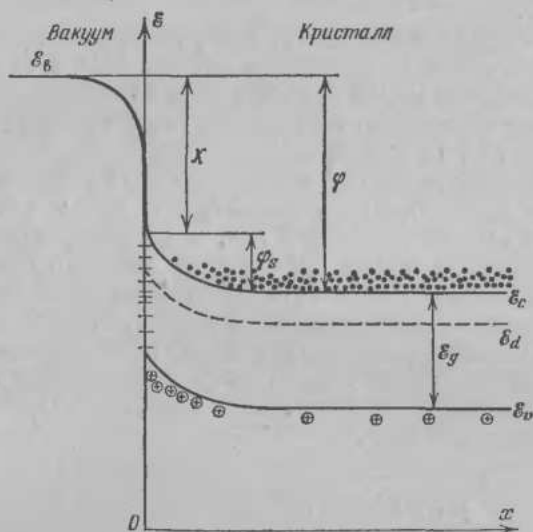


Рис. 35. Энергетическая диаграмма для полупроводника *n*-типа при отрицательном заряде, сосредоточенном на поверхностных состояниях. Черточки на границе раздела кристалл—вакуум соответствуют энергетическому положению поверхностных уровней в запрещенной зоне полупроводника. ε_d — объемный донорный уровень

поверхности кристалла, показывая, что внутри кристалла никакие силы на электроны и дырки не действуют и носителям все равно, где находиться — у поверхности или в глубине кристалла. Теперь, познакомившись с поверхностными состояниями, мы понимаем, что дело обстоит не

так просто. В случае, показанном на рис. 35, электрон отталкивается от границы кристалла, а положительно заряженная дырка притягивается к границе. Энергетическая диаграмма, как всегда, очень наглядно демонстрирует, что происходит. Вблизи границы внутри кристалла образовался, благодаря двойному заряженному слою, энергетический барьер высотой φ_s . Величина φ_s называется *поверхностным потенциалом*.

Для электронов (шариков на трясущемся подносе) барьер представляет собой горку (см. рис. 21), с которой электроны скатываются. Чем выше горка, тем труднее забраться на нее электронам и тем меньшее их количество добирается до вершины. Если в глубине кристалла, вдали от границы, концентрация электронов равняется n_0 , то у самой границы концентрация электронов проводимости составляет (см. формулу (1))

$$n_s \approx n_0 e^{-\varphi_s/kT}, \quad (27)$$

Для дырок, которые на энергетической диаграмме всегда стремятся «всплыть» (см. рис. 14 и 15), область барьера вблизи границы служит местом, где они скапливаются*). Энергия дырок вблизи границы меньше, чем энергия дырок в объеме кристалла на величину φ_s . Поэтому концентрация дырок вблизи поверхности p_s будет равняться

$$p_s \approx p_0 e^{\varphi_s/kT}, \quad (28)$$

где p_0 — концентрация дырок в глубине кристалла.

Призвав на помощь формулы (27), (28) и здравый смысл, мы легко разберемся, к чему может привести наличие на границе кристалла поверхностного потенциала и отсутствующего ему изгиба зон.

Пусть, например, мы исследуем кремний n -типа с концентрацией электронов $n_0 = 10^{15} \text{ см}^{-3}$. Тогда, в соответствии с табл. 1 и формулой (9), равновесная концентрация дырок в таком кристалле $p_0 = n_i^2/n_0 \approx 10^5 \text{ см}^{-3}$. Электронов в объеме кристалла в 10 миллиардов раз больше, чем дырок.

Пусть величина поверхностного потенциала $\varphi_s = = 0,1 \text{ эВ}$. Тогда при комнатной температуре ($kT =$

*) Еще раз вспомним, что это физически означает. Поле поверхностного двойного заряженного слоя отталкивает электроны, но притягивает дырки. Поэтому вблизи поверхности электронов меньше, а дырок больше, чем в объеме.

= 0,026 эВ) концентрация электронов на поверхности $n_s = 2 \cdot 10^{13} \text{ см}^{-3}$, а концентрация дырок $p_s \approx 5 \cdot 10^6 \text{ см}^{-3}$. Электронов вблизи поверхности примерно в 50 раз меньше, чем в объеме, дырок — во столько же раз больше. Но дырок все еще в 400 миллионов раз меньше, чем электронов, их вклад в проводимость ничтожен. Поэтому наличие поверхностного потенциала в этом случае означает появление вблизи поверхности обедненного электронами слоя с удельным сопротивлением, гораздо большим, чем объемное удельное сопротивление. Как мы увидим, возникновение на границах кристалла *обедненных слоев* с высоким удельным сопротивлением лежит в основе работы диодов и транзисторов.

Пусть теперь $\varphi_s = 0,2 \text{ эВ}$. Тогда $n_s \approx 4,5 \cdot 10^{11} \text{ см}^{-3}$, $p_s = 2,2 \cdot 10^8 \text{ см}^{-3}$. Качественно ситуация осталась той же. Электронов на поверхности теперь в две тысячи с лишним раз меньше, чем в объеме, но все еще гораздо больше, чем дырок. На границе кристалла существует, как и в предыдущем случае, обедненный слой. Но удельное сопротивление в слое и его ширина *) больше, чем раньше.

А вот при $\varphi_s = 0,4 \text{ эВ}$ возникает качественно новая ситуация. Теперь $n_s = 2 \cdot 10^8 \text{ см}^{-3}$, $p_s = 5 \cdot 10^{11} \text{ см}^{-3}$. Дырок на поверхности в две с половиной тысячи раз больше, чем электронов! На поверхности полупроводника *n*-типа образовался слой с дырочной проводимостью, так называемый слой с инверсным типом проводимости, или *инверсионный слой*.

На ранних этапах изучения полупроводников, когда о роли поверхностных состояний было мало что известно, возникновение инверсионных слоев на поверхности полупроводниковых кристаллов часто ставило исследователей в тупик. Кристалл, например, выращивался в атмосфере, насыщенной донорной примесью и по всем законам и правилам должен был бы обладать электронной проводимостью. А измерения проводимости (обычно выполняемые на поверхности кристалла) показывают, что кристалл ... *p*-тип! И сейчас еще иногда при неудачно выбранной технологии обработки поверхности возникновение инверсионных каналов препятствует нормальной работе диодов и транзисторов, в особенности высоковольтных.

С другой стороны, как мы увидим в следующей части книги, существуют типы полевых транзисторов, принцип работы которых основан на создании вблизи поверхности управляемого внешним сигналом инверсионного слоя.

*) Ширину обедненного слоя мы научимся вычислять немного дальше.

Обратим внимание на еще одно обстоятельство, к которому приводит наличие поверхностных состояний. Из рис. 35 видно, что наличие поверхностного потенциала изменяет работу выхода полупроводника. В случае, показанном на рис. 35, работа выхода увеличивается на величину φ_s . В самом деле, не будь поверхностных состояний, работа выхода из кристалла n -типа равнялась бы приблизительно величине электронного сродства χ (см. рис. 31). Теперь же, чтобы извлечь электрон из объема кристалла, нужно затратить энергию

$$\varphi = \chi + \varphi_s.$$

Энергетическая диаграмма, как всегда, очень наглядно демонстрирует это обстоятельство. Но, конечно (и тоже, как всегда), необходимо ясно понимать, в чем физическая причина изменения работы выхода. В данном случае причина вполне очевидна. Чтобы вылететь из глубины кристалла наружу, электрон, вдобавок ко всем тем силам, о которых мы говорили, обсуждая работу выхода, должен преодолеть еще и отталкивание со стороны отрицательного заряда на поверхности. Энергия, которая требуется, чтобы преодолеть это отталкивание, как раз и равняется величине φ_s .

Разумеется, возможны и случаи, когда на поверхности полупроводника n -типа будут преобладать донорные поверхностные состояния. Легко понять, к чему это приведет. Отдав электроны, такие состояния окажутся заряженными положительно и будут притягивать электроны из объема кристалла (а дырки — отталкивать). В результате вблизи поверхности кристалла возникнет слой, обогащенный электронами (обогащенный слой). Работа выхода из такого кристалла будет меньше, чем в случае, когда поверхностные состояния отсутствуют. (Понятно, почему?) Рекомендуем вам самостоятельно нарисовать энергетическую диаграмму для полупроводника n -типа с положительным поверхностным зарядом и проанализировать ее так же, как мы это только что сделали вместе для рис. 35.

Рис. 35, если убрать с него изображение донорного уровня \mathcal{E}_d , сильно уменьшить число точек-электронов и увеличить число кружков-дырок, вполне может служить энергетической диаграммой, поясняющей поведение границ зон в дырочном полупроводнике, у которого поверхность заряжена отрицательно. Видно, что наличие поверхностного заряда приведет в этом случае к образованию обогащенного дырками приповерхностного слоя.

Наконец, четвертый возможный случай — дырочный полупроводник, на поверхности которого суммарный заряд поверхностных состояний положителен. Энергетическую диаграмму для этого случая мы не поленились изобразить на рис. 36, но предоставляем читателю проанализировать ее самостоятельно.

Мы убедились, что наличие поверхностных состояний может как уменьшать, так и увеличивать работу выхода. И, конечно, именно наличием поверхностных состояний объясняются парадоксальные на первый взгляд результаты Аллена и Гобели (рис. 34).

Когда в высоком вакууме проводилось скалывание поверхности в полупроводнике n -типа, возникшие на по-

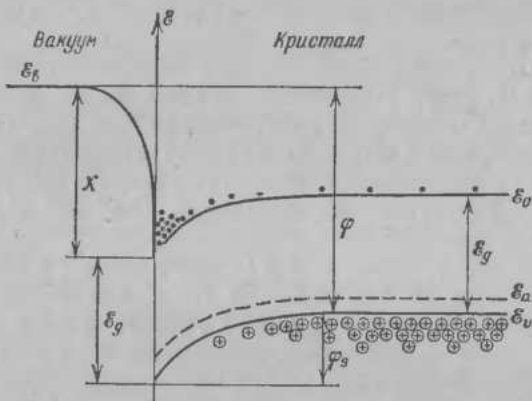


Рис. 36. Энергетическая диаграмма для дырочного полупроводника при положительном заряде поверхностных состояний. Обратите внимание, что изгиб границ вниз приводит к уменьшению работы выхода по сравнению со случаем, когда поверхностных состояний нет. При отсутствии заряда на поверхностных состояниях $\varphi \approx \chi + \epsilon_g$ (рис. 31). В данном случае $\varphi \approx \chi + \epsilon_g - \varphi_s$.

верхности уровни жадно захватывали электроны из объема полупроводника и заряжались отрицательно. При этом работа выхода, как мы теперь понимаем, возросла (рис. 35). Тот же самый спектр поверхностных состояний, возникавший при скалывании в дырочном полупроводнике, так же жадно захватывал из объема дырки (или, что то же самое, отдавал в объем электроны) и заряжался положительно. Работа выхода при этом, естественно, уменьшалась (рис. 36).

В обоих случаях, и для n - и для p -кремния, величина поверхностного потенциала φ_s равнялась приблизительно

$\mathcal{E}_g/2$ *). Тогда для n -кремния работа выхода равнялась $\Phi_n \approx \chi + \Phi_s \approx \chi + \mathcal{E}_g/2$, а для p -кремния — $\Phi_p = \chi + \mathcal{E}_g - \Phi_s = \chi + \mathcal{E}_g/2$. Как мы видим, $\Phi_n \approx \Phi_p$.

Moralité

Вблизи границы, разделяющей кристалл и окружающую среду, существуют силы, препятствующие электронам покидать кристалл и вылетать наружу. Энергия, необходимая для того, чтобы электрон мог преодолеть действие этих сил (работа выхода), составляет несколько электрон-вольт. В полупроводниках величина работы выхода зависит от типа проводимости и уровня легирования полупроводника. Кроме того, она зависит от спектра поверхностных состояний, т. е. от поверхностных уровней, всегда присутствующих на поверхности полупроводника. Отрицательный заряд поверхностных уровней увеличивает работу выхода, положительный — уменьшает.

*) Случай, когда высота поверхностного потенциального барьера близка к величине $\mathcal{E}_g/2$, встречается довольно часто. Ему соответствует ситуация, когда большинство поверхностных уровней расположено близко к середине запрещенной зоны полупроводника.

Глава 4

ОСНОВНЫЕ ПАРАМЕТРЫ ЭНЕРГЕТИЧЕСКИХ БАРЬЕРОВ В ПОЛУПРОВОДНИКАХ

...Бездельники палгали,
 Что верящий в старание — убог,
 Чужд божества. Вздор, боги
 обжигали
 Горшки когда-то. Обожги,
 как бог,
 Но поглотай и ты печную копоть,
 Побей огню поклонов без числа.
 Мастеровой не может не
 работать —
 Он упускает тайну ремесла.

(Из стихов, вспомнившихся в юности)

Эта глава — самая короткая в нашей книге, но она — одна из самых важных.

Мы уже много раз упоминали, что свойства полупроводниковых приборов определяются параметрами энергетических барьеров, возникающих на границе между металлом и полупроводником или между двумя частями полупроводникового кристалла. Самыми важными параметрами барьеров являются их высота, ширина и распределение напряженности электрического поля внутри барьера.

Мы пока что знакомы только с одним энергетическим барьером, существующим внутри полупроводника, — приповерхностным потенциальным барьером, возникающим из-за заряда на поверхностных состояниях (рис. 35, 36). Но уже это знакомство позволяет нам сформулировать вопрос, самый главный и самый важный и для всех остальных барьеров, которые нам предстоит изучить.

Вопрос этот можно сформулировать по-разному. Например, так: пусть в полупроводнике существует энергетический барьер высотой ϕ . Какова ширина этого барьера?

Или так: как глубоко может проникнуть в полупроводник электрическое поле E ?

Внимательно взглянув на рис. 35 и 36, можно понять, что эти вопросы действительно эквивалентны. В самом

деле, ширина барьера и характеризует то расстояние, на которое электрическое поле проникает внутрь полупроводника. Догадаться, каков качественный ответ на поставленный вопрос, можно, вспомнив, как ведет себя электрическое поле, с одной стороны, в металле, с другой — в диэлектрике.

Как внешнее электрическое поле проникает в металл, диэлектрик, полупроводник

Металл. Поместим внутрь конденсатора металлическую пластину так, как это показано на рис. 37, а. Хотя в воздухе, между обкладками конденсатора и металлом, существует электрическое поле E_1 , внутрь металла

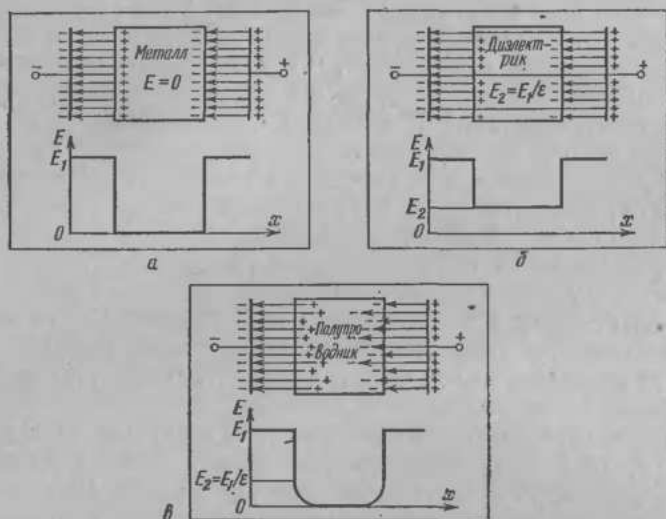


Рис. 37. Распределение поля в металле (а), диэлектрике (б) и полупроводнике (в), помещенных во внешнее электрическое поле

оно, как мы знаем, не проникает (Физика, 9 класс). Поле внутри металла равно нулю. Дело в том, что в металле в громадной концентрации ($\sim 10^{22}$ см $^{-3}$) содержатся свободные электроны и точно в таком же количестве — положительно заряженные ионы решетки (см, рис. 2). Пока электрического поля вблизи металла нет, в каждом, даже очень маленьком объеме внутри металла и на его поверх-

ности число положительных ионов в точности равно числу отрицательно заряженных электронов. Поэтому любая часть металла электрически нейтральна.

В металле, помещенном в электрическом поле, *электронейтральность* вблизи поверхности нарушается. К поверхности, обращенной к положительному электроду внешнего источника, притекают электроны. А от поверхности, обращенной к отрицательному электроду, они оттекают, оставляя вблизи поверхности слой положительно заряженных ионов (рис. 37, а).

Заряд, возникший таким образом на поверхности металла, создает электрическое поле, направленное против вызвавшего его внешнего поля. Внешнее поле полностью экранируется, и в результате поле внутри металла равно нулю *).

Может возникнуть законный вопрос. Если утверждает-ся, что в электрическом поле электронейтральность вблизи поверхности нарушается, то что же тогда означает термин «вблизи»? На какую же все-таки глубину поле проникает в металл? Мы вполне можем ответить на этот вопрос с «цифрами в руках».

Предположим для примера, что под действием внешнего поля, для того чтобы экранировать его, все электроны в металлической пластине отекли из слоя толщиной Δx , равной всего лишь постоянной решетки (т. е. расстоянию между соседними атомами кристаллической решетки, составляющему в твердых телах $\approx 5 \cdot 10^{-10}$ м). Какое поле создается при этом вблизи поверхности металла? Или, что то же самое, какое внешнее поле будет при этом экранировано?

Если из какого-либо участка металла ушли все электроны, то внутри этого участка возник положительный *объемный заряд* с плотностью $\rho = |eN_0|$, где N_0 — концентрация положительно заряженных ионов, в точности равная концентрации электронов n_0 в металле. Объемно-

*) Эту же мысль легко пояснить на другом языке — языке силовых линий электрического поля. Из школьного курса физики (9 класс) мы помним, во-первых, что густота силовых линий пропорциональна напряженности электрического поля, а во-вторых, что силовые линии начинаются на положительных зарядах и оканчиваются на отрицательных. Так вот, как бы ни был густ поток силовых линий, созданных внешним источником вблизи поверхности металла, вследствие большой концентрации заряженных носителей в металле, силовые линии поля всегда замыкаются на поверхностных зарядах и не проникают в глубь проводника (рис. 37, а.)

му заряду ρ , созданному вблизи поверхности в слое толщиной Δx , соответствует плотность поверхностного заряда $\sigma = \rho \Delta x = eN_0 \Delta x$. А плотности поверхностного заряда σ соответствует (Физика, 9 класс) напряженность электрического поля E вблизи поверхности:

$$E = k \cdot 2\pi\sigma = \frac{\sigma}{2\epsilon_0} = \frac{eN_0\Delta x}{2\epsilon_0}. \quad (29)$$

Здесь $k = 1/4\pi\epsilon_0$ — коэффициент пропорциональности в законе Кулона (ϵ_0 — электрическая постоянная).

Подставляя в формулу (29) численные значения входящих в нее величин: $N_0 = 10^{28} \text{ м}^{-3}$, $\Delta x = 5 \cdot 10^{-10} \text{ м}$, $\epsilon_0 = 8,85 \cdot 10^{-12} \text{ Кл}^2/(\text{Н} \cdot \text{м}^2)$, $e = 1,6 \cdot 10^{-19} \text{ Кл}$, получим $E = 4,5 \cdot 10^{10} \text{ В/м}$

Полученное значение E чудовищно велико.

При напряженности поля в десять тысяч раз меньшей ($\sim 3 \cdot 10^6 \text{ В/м}$) воздух теряет электрическую прочность и в нем возникает пробой. Но чтобы экранировать действие даже такого сильного поля, достаточно, как мы видим, чтобы из приповерхностного слоя металла толщиной всего навсего в один атомный слой ушла $3 \cdot 10^6 / 4,5 \cdot 10^{10} \approx 10^{-4}$ — одна десятитысячная доля электронов!

Таким образом, даже самые сильные электрические поля не проникают в металл на глубину, большую, чем постоянная решетки. Да и в пределах этой толщины относительное обеднение электронами со стороны «минуса» внешнего источника (и, соответственно, относительное обогащение со стороны «плюса») не превышает долей процента.

Диэлектрик. На рис. 37, 6 показана помещенная между обкладками конденсатора пластина диэлектрика. Под влиянием внешнего поля диэлектрик *поляризуется* (Физика, 9 класс). На поверхности диэлектрика появляются связанные заряды: отрицательные со стороны положительно заряженной пластины конденсатора и положительные с противоположной стороны. Поле этих связанных зарядов направлено навстречу вызвавшему их внешнему полю. В результате внешнее поле частично экранируется и поле внутри диэлектрика оказывается меньше, чем поле вне его, в ϵ раз (ϵ — диэлектрическая проницаемость диэлектрика).

На языке силовых линий такая ситуация соответствует картине, показанной на рис. 37, 6. Часть силовых линий внешнего поля замыкается на связанных поверхностных зарядах диэлектрика. Но те силовые линии,

которые не замкнулись на поверхностных зарядах и проникли в диэлектрик, пронизывают его насквозь. Внутри диэлектрика не существует свободных зарядов, которые могли бы, перегруппировавшись и изменив свое положение, экранировать действие проникшего в диэлектрик электрического поля.

Полупроводник. В пластине полупроводника, помещенной во внешнее электрическое поле (рис. 37, *е*), с одной стороны, как в диэлектрике, на поверхности возникают за счет поляризации связанные заряды. Таким образом, в полупроводнике на самой границе с вакуумом (или воздухом) напряженность поля E_2 будет в ϵ раз меньше, чем в вакууме (ϵ — диэлектрическая проницаемость полупроводника). С другой стороны, в полупроводнике, как в металле, существуют свободные носители заряда, которые способны, перегруппировавшись, экранировать действие проникшего в полупроводник электрического поля.

В полупроводнике *n*-типа, как мы знаем, существует равновесная концентрация электронов проводимости n_0 , которая в точности равна концентрации положительно заряженных доноров N_d (см. с. 37 и рис. 7). Пока электрическое поле в полупроводнике отсутствует, в каждом элементе объема полупроводника число положительных ионов (доноров) равняется числу отрицательно заряженных электронов. Любая часть полупроводника электрически нейтральна. В полупроводнике, помещенном в электрическом поле....

Дальше, не меняя ни единого слова, мы могли бы переписать кусок текста со с. 114, начиная со слов «электронейтральность нарушается...» и до слов «что, собственно говоря, означает термин «вблизи»...»

На какую же глубину поле проникает в полупроводник? В чем разница между полупроводником и металлом?

В том, что концентрация свободных носителей и соответственно заряженных ионов в полупроводнике, как правило, гораздо меньше, чем в металле. Поэтому, чтобы создать достаточно большой положительный заряд, способный экранировать действие внешнего поля, электроны в полупроводнике *n*-типа должны «оттечь» не из тончайшего слоя, не превышающего по толщине постоянную решетки, а из слоя довольно большой толщины (рис. 37, *е*). Ясно, что толщина этого слоя будет тем больше, чем меньше равновесная концентрация электронов n_0 в полупроводнике.

Концентрация электронов n_0 практически всегда определяется концентрацией введенных в материал доноров N_d (или, как иногда принято говорить, уровнем легирования). Таким образом, глубина, на которую поле проникает внутрь полупроводника, определяется уровнем его легирования.

Уровень легирования полупроводников, используемых для изготовления различных полупроводниковых приборов, лежит в пределах от $\sim 10^{10}$ до $\sim 10^{20}$ см $^{-3}$, т. е. может различаться в десятки миллиардов раз. Соответственно и глубина проникновения поля в полупроводник, или, иными словами, ширина энергетических барьеров в полупроводниках, может меняться в очень широких пределах.

До сих пор мы говорили только об электронном полупроводнике. В дырочном полупроводнике дело обстоит точно так же.

В полупроводнике p -типа, как мы знаем, существует равновесная концентрация положительно заряженных подвижных носителей заряда — дырок, в точности равная концентрации неподвижных отрицательно заряженных акцепторов N_d . Если поместить полупроводник p -типа во внешнее электрическое поле, дырки будут притекать к поверхности полупроводника, обращенной к отрицательному электроду внешнего источника. А у поверхности полупроводника, обращенной к положительному электроду, останется нескомпенсированный объемный заряд отрицательно заряженных акцепторов. Чем меньше концентрация акцепторов, т. е. чем меньше уровень легирования полупроводника акцепторной примесью, тем больше должна быть толщина слоя, чтобы создать необходимый для экранирования внешнего поля заряд.

Закон изменения поля с координатой

В предыдущем разделе, рассматривая качественно картину проникновения поля в различные материалы, мы могли убедиться, что напряженность электрического поля в материале изменяется только там, где имеется объемный заряд — положительный или отрицательный.

Особенно наглядно это утверждение звучит на языке силовых линий. Действительно, если в какой-либо области имеется положительный объемный заряд, он является источником силовых линий. Густота линий поля в этой области увеличивается, т. е. напряженность электрического поля возрастает. В области, где имеется отрицательный объемный заряд, часть силовых линий поля замыкается на нем. Густота линий поля уменьшается — напряженность электрического поля падает. В области, где нет объемного заряда, напряженность поля, очевидно, не меняется.

Язык силовых линий позволяет также четко уяснить, что такое правило справедливо тогда, когда поле изменяется только вдоль одной координаты. Или, как часто выражаются, в одномерном случае.

В более сложных случаях, когда поле может изменяться вдоль двух или всех трех пространственных координат, напряженность поля может меняться и при отсутствии объемного заряда. Например, напряженность поля точечного заряда спадает с удалением от него как квадрат расстояния, даже если никакого объемного заряда в окружающем заряд материале нет. В этом случае густота силовых линий уменьшается не потому, что часть их замыкается на объемных зарядах, а потому, что то же число силовых линий с увеличением расстояния до заряда пронизывает все большую поверхность. В книге мы везде ограничимся рассмотрением только одномерной картины распределения поля. В этом случае напряженность поля изменяется только там, где есть объемный заряд.

Закон, связывающий скорость изменения поля с координатой dE/dx и плотность объемного заряда ρ , имеет вид

$$\frac{dE}{dx} = \frac{\rho}{\epsilon\epsilon_0}, \quad (30)$$

где ϵ — диэлектрическая проницаемость материала, ϵ_0 — электрическая постоянная.

Уравнение (30) играет в физике полупроводниковых приборов исключительно важную роль. Как мы скоро убедимся, именно с его помощью можно определить ширину энергетических барьеров в полупроводниках.

В том разделе физики, где изучаются полупроводниковые приборы, уравнение (30) часто называют *одномерным уравнением Пуассона* *).

Выведем сами уравнение (30).

Рассмотрим очень внимательно, «с большим увеличением» область полупроводника, в которой имеется объемный заряд (рис. 38, а). Пусть для определенности этот заряд положительный. (Для большей наглядности мы можем представлять себе эту область как левую часть полупроводниковой пластины, показанной на рис. 37, в. Из этой части пластины отдели электроды, оставив нескомпенсированным положительный заряд доноров с объемной плотностью ρ .)

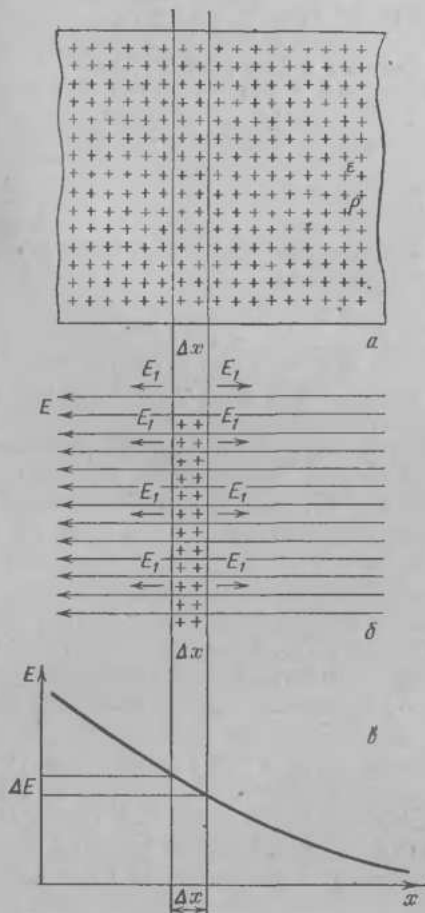
Поле в рассматриваемой области изменяется. Наша задача — определить закон изменения поля.

*) На самом деле такое название не вполне точно. Одномерное уравнение Пуассона, строго говоря, описывает изменение с координатой не напряженности электрического поля E , а потенциала U , и имеет вид

$$\frac{d^2U}{dx^2} = -\frac{\rho}{\epsilon\epsilon_0}.$$

(Можете ли вы сами установить связь между этим уравнением и уравнением (30)?)

Выделим внутри рассматриваемой области узкую полосу толщиной Δx . В пределах этой полосы напряженность электрического поля меняется на величину ΔE . Почему поле меняется,



Фиг. 38. Поле E_x узкой заряженной полоски слева от полоски складывается с внешним полем E , а справа — вычитается из него. На длине Δx поле меняется на величину $\Delta E = 2E_1$

вполне понятно. Заряженная полоска создает вблизи своей поверхности поле E_1 (рис. 38, б). Слева от полоски это поле складывается с внешним полем E , а справа — вычитается из него. Поэтому напряженность поля слева от полоски будет больше, чем справа, на величину $\Delta E = 2E_1$.

При объемной плотности заряда ρ плотность заряда на единицу поверхности σ равняется, естественно, $\sigma = \rho \Delta x$. А полоска с по-

верхностной плотностью заряда σ создает в материале с диэлектрической проницаемостью ϵ поле E_1 , равное

$$E_1 = \frac{\sigma}{2\epsilon\epsilon_0} = \frac{\rho\Delta x}{2\epsilon\epsilon_0}.$$

(Сравните эту формулу с формулой (29).)

Таким образом, $\Delta E = 2E_1 = \rho\Delta x/\epsilon\epsilon_0$, или

$$\frac{\Delta E}{\Delta x} = \frac{\rho}{\epsilon\epsilon_0}.$$

Мы получили уравнение (30).

Распределение поля в барьере.

Ширина барьера



Используем уравнение Пуассона (30) для решения задачи, которую мы поставили перед собой в этой главе. Определим распределение электрического поля и ширину энергетического барьера в области поверхностного потенциала φ_s (рис. 35). Сформулируем сначала несколько предположений о характере барьера и о полупроводнике, в котором он исследуется. Эти предположения не только сильно облегчат расчет, но и будут соответствовать вполне реальным ситуациям, которые очень часто встречаются в самых разных полупроводниковых приборах.

Во-первых, примем, что полупроводник *легирован однородно*, т. е. концентрация легирующей примеси везде одинакова и не зависит от координаты. Если это полупроводник *n*-типа (будем для определенности говорить об электронном полупроводнике), концентрация ионизованных доноров N_d всюду постоянна. Во-вторых, будем считать, что высота потенциального барьера φ_s , с одной стороны, достаточно велика: $\varphi_s \gg kT$, но, с другой стороны, недостаточна для того, чтобы вызвать инверсию проводимости вблизи границы кристалла (именно с такими барьерами нам придется иметь дело особенно часто).

Если эти два условия выполнены, расчет становится очень простым.

В самом деле, потенциальный барьер на границе обусловлен тем, что подвижные носители (в данном случае электроны) оттекли от границы, оставив нескомпенсированным положительный объемный заряд ионизованных доноров (рис. 39, а). Как мы знаем (см. формулу (27) и текст на с. 108), при $\varphi_s \gg kT$ из приповерхностного слоя

уходят практически все электроны. Тогда плотность объемного заряда, естественно, равняется оставшемуся нескомпенсированному заряду ионизованных доноров: $\rho = eN_d$, где N_d — концентрация доноров.

В соответствии с уравнением Пуассона (30) поле внутри поверхностного слоя будет изменяться по закону

$$\frac{dE}{dx} = \frac{\rho}{\epsilon\epsilon_0} = \frac{e}{\epsilon\epsilon_0} N_d. \quad (31)$$

Так как N_d — величина постоянная, то, в соответствии с формулой (31), скорость изменения поля с координатой — тоже величина постоянная, т. е. напряженность

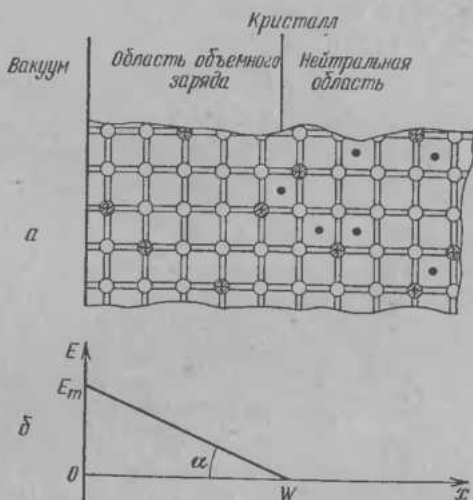


Рис. 39. Формирование приповерхностного потенциального барьера (полупроводник n -типа): a — отрицательный заряд на поверхностных состояниях заставил электроны отечь из приповерхностной области, оставив нескомпенсированным положительный объемный заряд ионизированных доноров; b — поле в приповерхностной области линейно спадает от величины E_m на границе кристалла ($x = 0$) до нуля на границе области объемного заряда ($x = W$), $\operatorname{tg} \alpha = eN_d/\epsilon\epsilon_0$

поля E линейно спадает с координатой x (рис. 39, б). Тангенс угла наклона α как раз и определяется формулой (31): $\operatorname{tg} \alpha = dE/dx = eN_d/\epsilon\epsilon_0$.

Первую задачу — найти распределение поля внутри барьера — мы, таким образом, решили.

Если бы нам была известна величина E_m — поле на границе кристалла, то из уравнения (31) мы легко определили бы ширину барьера W : из уравнения (31) и рис. 39, б видно, что

$$E_m = W \operatorname{tg} \alpha = \frac{eN_d}{\epsilon\epsilon_0} W. \quad (32)$$

Но значения E_m мы не знаем. Нам известно только, что между границей кристалла и точкой $x = W$, где поле обращается в нуль, существует потенциальный барьер высотой φ_s . Или, что то же самое, между этими точками существует разность потенциалов $U_s = \varphi_s/e$. Этого условия оказывается достаточно, чтобы определить и величину E_m , и ширину барьера W !

Напишем сразу ответы:

$$W = \left(\frac{2\epsilon\epsilon_0 U_s}{eN_d} \right)^{1/2}; \quad (33a)$$

$$E_m = \left(\frac{2eN_d U_s}{\epsilon\epsilon_0} \right)^{1/2}, \quad (33b)^*$$

и пригласим самых упорных получить их вместе с нами.

Из школьного курса электростатики известно, что если между точками, находящимися друг от друга на расстоянии Δx , существует поле E , то разность потенциалов между этими точками равняется $\Delta U = E\Delta x$.

Если бы в пределах всего слоя объемного заряда, который мы рассматриваем, поле было бы везде одинаковым, то разность потенциалов между границей кристалла ($x = 0$) и границей области объемного заряда ($x = W$) находилась бы очень просто.

Например, если поле в области объемного заряда всюду равнялось бы E_m , то, очевидно, выполнялось бы равенство $U_s = E_m x$. Такой случай показан на рис. 40 штриховой линией.

Но, как мы знаем, в действительности поле внутри барьера не постоянно. Оно линейно спадает от границы внутрь кристалла. Как же в этом случае найти связь между E_m и U_s ?

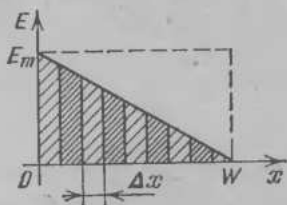


Рис. 40. Разность потенциалов U_s между границей кристалла ($x = 0$) и границей области объемного заряда ($x = W$) равняется площади, ограниченной осями координат и зависимостью $E(x)$. Если поле всюду равно E_m , то $U_s = E_m W$ (штрих). Если поле линейно спадает с координатой, то $U_s = 1/2 E_m W$

*) Проверьте, что из равенств (33б) и (33а) вытекает равенство (32).

Разобьем мысленно область объемного заряда на полоски шириной Δx , такие узкие, чтобы поле в пределах каждой полоски могло приближенно считаться постоянным (рис. 40). Тогда разность потенциалов ΔU между каждыми соседними точками будет равняться значению поля E в данной точке, умноженному на величину Δx , т. е. равняться площади маленького прямоугольника с основанием Δx . А полная разность потенциалов между точками $x = 0$ и $x = W$ равняется сумме площадей всех этих маленьких прямоугольников, т. е. площади заштрихованного треугольника на рис. 40.

Тогда, используя формулу (32), получаем

$$U_s = \frac{E_m W}{2} = \frac{e N_d}{2 \epsilon \epsilon_0} W^2.$$

Отсюда сразу же следует формула (33а).

То же выражение (32) можно переписать в виде $W = \epsilon \epsilon_0 E_m / e N_d$ и из формулы для U_s исключить W :

$$U_s = \frac{E_m W}{2} = \frac{\epsilon \epsilon_0}{2 e N_d} E_m^2.$$

Отсюда сразу следует формула (33б).

Прделаем теперь численные оценки. Пусть высота приповерхностного потенциального барьера составляет несколько десятых электрон-вольта. Примем для определенности $\varphi_s = 0,2$ эВ. Значения диэлектрической проницаемости для наиболее широко распространенных полупроводников различаются не слишком сильно. Для Ge $\epsilon = 16$, для Si $\epsilon = 12,5$, для GaAs $\epsilon = 12$. А вот уровень легирования N_d , как мы знаем, может меняться в полупроводниковых приборах в очень широких пределах: от $\sim 10^{10}$ до $\sim 10^{20}$ см $^{-3}$.

Пусть $N_d = 10^{10}$ см $^{-3}$. Для наглядности будем представлять себе, что барьер высотой $\varphi_s = 0,2$ эВ ($U_s = 0,2$ В) существует на поверхности полуизолирующего GaAs (см. гл. 2). Тогда, подставив в формулы (33) значения $\epsilon = 12$ и $N_d = 10^{16}$ м $^{-3}$, получим $E_m = 2,4 \cdot 10^3$ В/м = 24 В/см, $W = 1,7 \cdot 10^{-4}$ м = $1,7 \cdot 10^{-2}$ см.

А вот если такая же приповерхностная разность потенциалов $U_s = 0,2$ В существует на поверхности высоколегированной GaAs пластины с $N_d = 10^{20}$ см $^{-3}$, то $E_m = 2,4 \cdot 10^6$ В/см, $W = 1,7 \cdot 10^{-7}$ см.

В первом случае поле приповерхностного барьера проникает в глубь полупроводника на толщину порядка 300 000 атомных слоев. Во втором — порядка 3—4 слоев.

Разумеется, если мы имеем дело не с электронным, а с дырочным полупроводником, то все проведенные рассуждения и все полученные формулы остаются справедливыми. Нужно только заменить в них величину N_d на ($-N_a$) — концентрацию мелкой акцепторной примеси.

Может возникнуть законный вопрос: мы все время говорили о ширине и других параметрах обедненных свободными носителями слоев. Между тем, как мы знаем, вполне реально и образование слоев, обогащенных носителями (электронами или дырками). Чем же вызван наш повышенный интерес к обедненным слоям и явное пренебрежение к обогащенным?

При выводе уравнения Пуассона (30) совершенно все равно, какова природа объемного заряда ρ . Возник ли он вследствие обеднения или обогащения участка полупроводника носителями, не играет никакой роли. Поэтому уравнение (30) с одинаковым успехом можно использовать и для расчета параметров обогащенных слоев.

Предпочтительный интерес к обедненным слоям вызван двумя обстоятельствами. Во-первых, именно с такими слоями нам придется иметь дело в подавляющем большинстве случаев, анализируя работу полупроводниковых приборов. А во-вторых, гораздо проще рассчитывать параметры обедненных слоев, чем обогащенных. И вот почему. Обеднение носителями имеет вполне очевидный естественный предел. Если носители ушли все, то ясно, что больше уходить некому. А если какой-либо участок полупроводника полностью обеднен носителями, то плотность объемного заряда ρ в нем равна либо eN_a (полупроводник n -типа), либо $-eN_a$ (p -тип).

При обогащении дело обстоит сложнее. На первый взгляд может даже показаться неясным, имеет ли обогащение избыточными носителями естественный предел. Что мешает избыточным носителям собраться в громадной концентрации в очень узком слое?

Накопление носителей ограничивает уже известный нам процесс диффузии. Чем больше избыточная концентрация свободных носителей и чем уже слой, в котором они сосредоточены, тем больше диффузионный поток носителей, противодействующий дальнейшему их накоплению (см. формулу (18)).

Распределение объемного заряда имеет более сложную зависимость от координаты при обогащении, чем при обеднении. Соответственно более сложный вид имеет распределение поля $E(x)$ и потенциала $U(x)$.

Однако качественный вывод, подкрепленный нами для обедненных слоев строгим расчетом, остается справедливым и для обогащенных слоев. Чем меньше уровень легирования, тем больше ширина слоев и меньше амплитуда поля при той же величине потенциального барьера,

Трансмутационное легирование полупроводников

Из гл. 1 и 2 мы знаем, что концентрация и характер примесей, введенных в полупроводник, в решающей степени определяют его свойства. В этой главе мы познакомились с тем, как сильно влияет уровень и характер легирования полупроводника на параметры возникающих в нем барьеров. Немудрено, что разработка способов введения примесей в полупроводниковые материалы была в центре внимания ученых и инженеров с самого начала развития полупроводниковой технологии.

Методов легирования, т. е. способов введения в полупроводник примесей, существует очень много. С одним из них — введением примесей в расплав при получении монокристаллов (метод Чохральского) — мы уже знакомы.

Познакомимся еще с одним методом легирования — одним из наиболее оригинальных, красивых и неожиданных. Этот метод называется «трансмутационным легированием» *). В словаре современной физики термин «трансмутация» оставлен, главным образом, для тех превращений, которые претерпевают атомные ядра при радиоактивном поглощении и распаде. Трансмутационное легирование полупроводников осуществляется с помощью погружения полупроводниковых материалов в каналы атомных реакторов.

Идея трансмутационного легирования состоит в том, чтобы осуществить в полупроводниковом кристалле ядерную реакцию, превращающую атомы полупроводника в атомы другого вещества, служащего в этом полупроводнике примесным центром (донорным или акцепторным).

Из многочисленных ядерных реакций такого типа самыми удобными оказались реакции, в которых радиоактивное превращение ядер происходит при захвате ими медленных нейтронов с энергией ~100 эВ. Поэтому, говоря о «трансмутационном легировании» или «ядерном легировании», чаще всего имеют в виду нейтронное легирование полупроводников.

Впервые на возможность нейтронного легирования указал в 1950 г. известный американский исследователь полупроводников К. Ларк-Горовиц. Возглавляемая им группа сотрудников осуществила трансмутационное легирование германия. В проделанных экспериментах часть атомов Ge при облучении потоком нейтронов превратилась в атомы галлия (Ga), который в германии является акцепторной примесью.

Природный германий представляет собой смесь изотопов, в которой содержится около 21% изотопа $^{70}_{32}\text{Ge}$. (Напомним, что символ $^{70}_{32}\text{Ge}$ означает, что ядра атомов этого изотопа германия содержат 32 протона и $70 - 32 = 38$ нейтронов.) При облучении германия медленными нейтронами происходят следующие реакции (см. рис. 41):

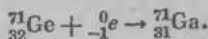
а) ядро изотопа $^{70}_{32}\text{Ge}$ захватывает нейтрон:



*) От латинских слов *trans* (сквозь, через) и *mutatio* (изменение, перемещение).

б) образуется нестабильный изотоп ${}^{71}_{32}\text{Ge}$;

в, г) этот нестабильный изотоп распадается с периодом полураспада в 11 дней, образуя атомы Ga:



Очень интересен механизм, благодаря которому атом ${}^{71}_{32}\text{Ge}$ превращается в галлий. Видно, что для такого превращения ядро германия должно захватить электрон. Откуда же этот электрон берется?

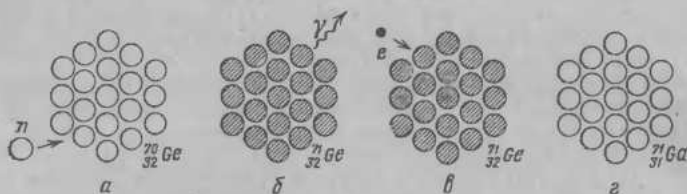


Рис. 41. Схема превращения ядер при трансмутационном легировании германия

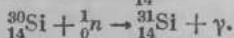
Ядро ${}^{71}_{32}\text{Ge}$ захватывает электрон с внутренней, ближайшей к ядру электронной оболочки атома (K-оболочки). Такой тип реакции поэтому называется K-захватом.

К началу 60-х годов относятся первые сообщения о нейтронном легировании кремния — основного полупроводникового материала тогдашней и современной полупроводниковой электроники. Однако прошло почти 15 лет, пока замечательные свойства этого метода были оценены в полной мере.

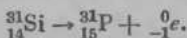
С совершенствованием полупроводниковых приборов появилась нужда в громадных, объемом в десятки тысяч см³, монокристаллах, легированных очень однородно и с высокой точностью по всему объему. Такие кристаллы, как мы знаем, необходимы для создания мощных полупроводниковых приборов. Кроме того, они нужны для современных сверхбольших интегральных схем (СБИС). Однородное и точное легирование совершенно необходимо и для создания полупроводниковых счетчиков ядерных частиц.

Удовлетворить таким строгим требованиям традиционные методы легирования оказались не в состоянии. В то же время нейтронное легирование позволяет получить очень точно и однородно легированные кристаллы. В ядерных реакторах, где проводится нейтронное легирование, плотность потока нейтронов и время облучения контролируются с очень высокой точностью. А вследствие того, что нейтроны слабо взаимодействуют с веществом и обладают высокой проникающей способностью, их поток оказывается однородным и по всей толщине кристалла. В результате, например, при нейтронном легировании неоднородность концентрации примесей в слитке кремния диаметром 10 см и длиной полметра не превышает ~5%. А в таком же слитке, полученном методом зонной плавки, разброс в концентрации примеси составляет 25% — в 5 раз больше!

При поглощении медленного нейтрона изотопом кремния $^{30}_{14}\text{Si}$ образуется нестабильный изотоп $^{31}_{14}\text{Si}$:



Затем этот изотоп распадается с периодом полураспада 2,6 часа, превращаясь в фосфор. Распад сопровождается испусканием электронов (β -излучение):



Фосфор в кремнии, как мы знаем, является мелким донором. Поэтому облучение нейтронами кристаллов кремния n -типа приводит к увеличению их проводимости. А нейтронное легирование дырочного Si приводит к компенсации исходной акцепторной примеси и позволяет получать высокоомные кристаллы, используемые для создания детекторов ядерных частиц.

В последние годы была показана возможность нейтронного легирования арсенида галлия (GaAs) — одного из наиболее перспективных полупроводниковых материалов.

В этом соединении под влиянием нейтронов происходит трансмутация обоих типов атомов, из которых состоит материал: и галлия, и мышьяка.

Несколько слов о прямоугольных треугольниках

Повторим еще раз то, что нам уже известно о связи между высотой барьера, шириной и распределением поля в барьере, пользуясь наглядным языком геометрических построений.

В этой главе уже несколько раз говорилось, что в основном нам придется иметь дело с энергетическими (потенциальными) барьерами, для которых характерно

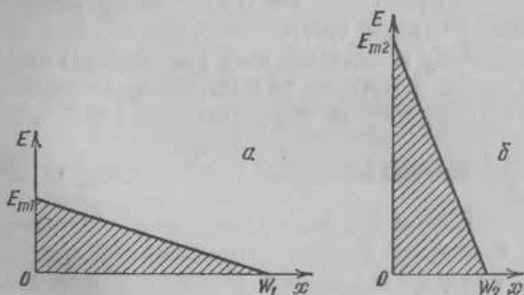


Рис. 42. Распределение поля в барьере $E(x)$ для двух полупроводников. Высота барьера одинакова. Концентрация примесей в случае b в 10 раз больше, чем в случае a

полное обеднение слоя полупроводника свободными носителями. Тогда, как мы знаем, поле линейно спадает от границы в глубь полупроводника (рис. 40),

Те, кто проделал вместе с нами вывод формул (33), знают, что заштрихованная площадь на рис. 40, т. е. площадь, ограниченная зависимостью $E(x)$ и осями координат, численно равняется разности потенциалов U_s . Таким образом, $U_s = \varphi_s/e = 1/2 E_m W$. (Те, кто поленился проследить за выводом, могут убедиться в справедливости этого утверждения, перемножив формулы (33а) и (33б).)

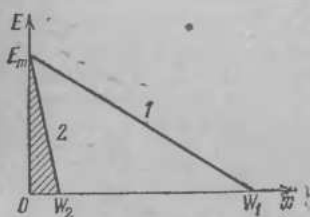


Рис. 43. Распределение поля в барьере $E(x)$ для двух полупроводников. Максимальное поле в барьере одинаково. Концентрация примесей в случае 2 в 10 раз больше, чем в случае 1

$\approx 3,14$ раз меньше, а максимальное поле во столько же раз больше, чем в случае a , как это и следует из (33).

Иногда при исследовании барьеров нам придется сталкиваться с другой ситуацией. Заданной окажется не высота барьера φ , а величина максимального поля в барьере E_m . В этом случае ширина барьера W и высота барьера $\varphi = eU$ обратно пропорциональны концентрации примеси N (рис. 43, формула (32)).

Наглядность, с которой вид распределения $E(x)$ позволяет судить о параметрах барьера, не раз сослужит нам добрую службу в последующих главах.

Moralité

Ширина энергетических (потенциальных) барьеров в полупроводнике определяется уровнем легирования. Для слабо легированных полупроводников она может составлять сотни тысяч атомных слоев (десятки микрометров), для сильно легированных — единицы атомных слоев (тысячные доли микрометра). При заданной высоте барьера $\varphi = eU$ его ширина W обратно пропорциональна, а максимальное поле в барьере E_m прямо пропорционально $N^{1/2}$ (N — концентрация легирующей примеси).

..Нет, никогда не было и не будет впрямь премудрости более достоверной, чем та, которую я поведаю тебе сейчас, о мулла! Но приготовься, чтобы не поразила тебя слишком эта премудрость, ибо от нее легко потерять рассудок — настолько она поразительна, ослепительна и необъятна.

Л. В. Соловьев.

«Повесть о Ходже Насреддине»



Барьер, который мы будем изучать в этой главе, исследуется уже более полувека. Его свойства изучены в тончайших деталях. На его основе созданы десятки самых разнообразных полупроводниковых приборов. Кажется, нет такого вопроса, связанного с этим барьером, на который уже не был получен ответ... И все-таки не проходит года, чтобы на свет не появился еще один, а иногда несколько новых приборов, в основе которых лежат свойства все того же $p - n$ перехода.

Если кристалл полупроводника легирован так, что одна часть его имеет проводимость p -типа (дырочную), а другая — проводимость n -типа (электронную), то на границе между этими частями возникает слой с особыми свойствами, который и называется $p - n$ переходом или, иначе, электронно-дырочным переходом.

Изучая свойства этого замечательного перехода, мы будем придерживаться следующего плана. Сначала мы обсудим методы, с помощью которых получают $p - n$ переходы. Затем — самые основные свойства перехода. А затем мы продолжим обсуждение свойств перехода, параллельно рассматривая работу приборов, основанных на этих свойствах. Каждый из приборов, с которыми мы познакомимся, широко используется и имеет важные и интересные приложения. Но кроме того, знакомство с устройством и принципом работы приборов позволяет лучше узнать и глубже понять свойства $p - n$ перехода.

Первая мысль, которая приходит в голову неискушенному человеку, размышляющему над тем, как изготовить $p-n$ переход, обычно состоит в следующем. Нужно взять два куска полупроводника: один p -, другой n -типа, тщательно их отшлифовать и крепко прижать друг к другу....

Полвека назад, «на заре туманной юности» полупроводниковой технологии, эту мысль неоднократно пытались экспериментально реализовать даже специалисты. Увы, этот простой путь не приводит к успеху. Если все это проделать, в наших руках окажется объект, свойства которого не имеют ничего общего со свойствами $p-n$ перехода, который мы собираемся изучить.

Обратимся к рис. 44, б и отметим следующее очень важное обстоятельство. Кристаллическая решетка полупроводника ничего о $p-n$ переходе не знает, она просто не замечает его. В самом деле, на самой ли границе, разделяющей p - и n -области, слева ли, справа ли от нее, мы видим все ту же кристаллическую решетку кремния, с теми же связями, которые никак не нарушаются тем, что в разные части кристалла введены разные примеси. Только такой переход обладает всеми теми замечательными свойствами, о которых мы только что говорили. И именно такой переход мы намерены изучить.

Между тем, прочитав гл. 3, мы прекрасно понимаем, что на границе каждого из двух кусков полупроводника, p - и n -типа, как бы тщательно ни была обработана их поверхность, возникнут нарушения кристаллической решетки. Образуется спектр поверхностных состояний. Возникнут приповерхностные потенциальные барьеры... Словом, все те явления, которые мы изучили в гл. 3.

Как плотно ни прижимай друг к другу куски p - и n -типа полупроводника, раны, нанесенные кристаллической решетке на поверхности, не залечишь. На границе всегда останутся нарушения кристаллической решетки: окислы, адсорбированные чужеродные атомы...

Проблема создания качественных $p-n$ переходов, у которых все эти и многие другие возможные дефекты отсутствуют, решается вот уже более полувека усилиями тысяч специалистов, работающих во многих странах. Немудрено, что разработано довольно много способов получения переходов и все эти способы доведены до высокого технического совершенства.

Для всех способов, которые мы рассмотрим в этом разделе, характерен один общий прием. Берется кристалл полупроводника p - или n -типа. Затем каким-либо способом на определенную глубину в кристалл вводятся примеси, создающие в данном полупроводнике примесь противоположного характера. Если исходно кристалл был

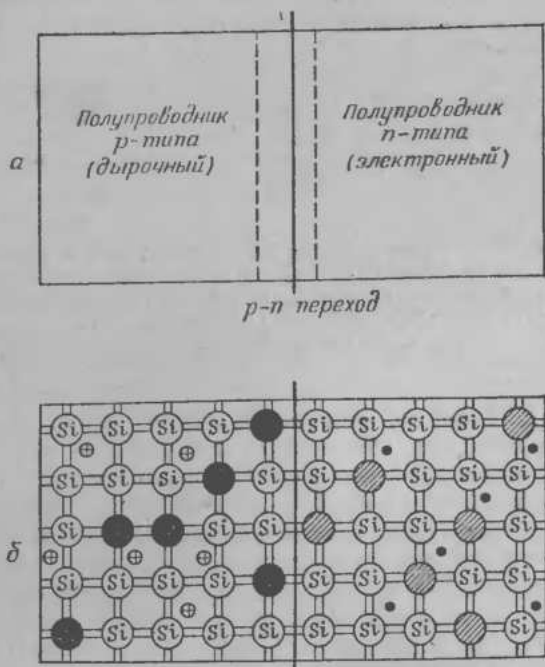


Рис. 44. На границе p - и n -типов полупроводников возникает $p - n$ переход (а). В качестве примера показан кристалл кремния, в левую часть которого введена акцепторная примесь (темные кружки), а в правую — донорная (заштрихованные кружки) (б)

n -типа, то вводится акцепторная примесь, если p -типа — то донорная. Примесь вводится в кристалл в концентрации большей, чем исходная. Поэтому в той части кристалла, куда попала примесь, тип проводимости изменяется. Происходит, как говорят, *перекомпенсация исходной примеси*.

На границе между областью, где тип проводимости в кристалле изменился, и областью, где он остался исходным, возникает $p - n$ переход.

Сплавление. Сплавление — один из самых старых способов создания $p-n$ переходов. Первые $p-n$ переходы были получены этим методом в 1950 г. В основе метода лежит образование сплава полупроводникового материала и какого-либо металла, атомы которого являются донорами или акцепторами в полупроводнике.

Рис. 45, *a* показывает в качестве примера образование сплавного $p-n$ перехода в n -Ge при вплавлении в него

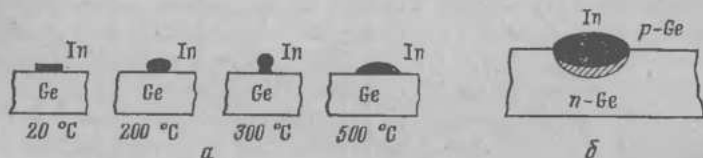


Рис. 45. Изготовление сплавного $p-n$ перехода: *a* — изменение формы таблетки In при увеличении температуры; *b* — приповерхностная область кристалла германия в процессе вплавления индия обогащается атомами акцепторной примеси (In). В зоне застывшего расплава (рекристаллизованной зоне) образуется сильно легированный германий p -типа

индия (In). Таблетка индия, имеющая форму тонкого диска, кладется на очищенную поверхность кристалла германия. Затем кристалл вместе с лежащей на нем таблеткой медленно нагревают. Как только температура превысит температуру плавления металлической таблетки — для индия эта температура составляет 156 °C — таблетка начинает плавиться, превращаясь в каплю жидкости. При повышении температуры форма капли, как видно из рис. 45, изменяется. Силы поверхностного натяжения жидкости стремятся придать капле форму шара, а силы смачивания — размазать каплю по поверхности кристалла. Индий, собравшийся при температуре ≈ 300 °C в шарик, при более высокой температуре (≈ 500 °C) растекается по поверхности германия.

Кристаллический германий при высокой температуре ≥ 500 °C очень хорошо растворяется в жидком индии. Растворение продолжается до тех пор, пока не образуется насыщенный раствор германия в индии. После этого германиевую пластинку начинают медленно охлаждать. С уменьшением температуры количество германия, способного раствориться в индии, также уменьшается; избыточный германий, выпадающий из раствора, вновь кристаллизуется на оставшейся нерастворенной части пластинки (рекристаллизуется). Образовавшийся p -

кристаллизованный слой (рис. 45, б) оказывается обогащенным атомами индия.

Атомы индия в германии являются акцепторами. Концентрация индия в рекристаллизованном слое составляет $\sim 10^{19}$ см⁻³ и значительно превосходит концентрацию донорной примеси в исходном кристалле. Поэтому на границе рекристаллизованной области и нерастворенной части кристалла образуется $p - n$ переход.

В 50-х годах нашего века метод сплавления сыграл важную роль в развитии полупроводниковой электроники. Именно на его основе были созданы первые промышленные германиевые выпрямительные диоды и первые транзисторы, пригодные для широкого практического применения.

Благодаря относительной простоте и высокой надежности метод сплавления используется и при создании современных полупроводниковых приборов. Промышленно выпускаются десятки типов германиевых и кремниевых сплавных диодов самого разнообразного назначения (рис. 46).

Метод сплавления позволяет получать $p - n$ переходы во всех распространенных полупроводниковых материалах. Для германиевых приборов в качестве источника акцепторной примеси используется, как мы уже видели, индий. $p - n$ переходы на основе p -Ge могут быть получены, например, при сплавлении p -Ge с сурьмой.

При создании сплавных $p - n$ переходов в кремнии в качестве металла — источника акцепторной примеси часто используется алюминий (Al), в качестве источника донорной примеси — сплавы олова с сурьмой и золота с сурьмой.

Диффузия. Диффузионный метод изготовления $p - n$ переходов появился вскоре после метода сплавления как некоторая его модификация. Однако уже в первые годы применения диффузии для изготовления полупроводниковых приборов выявились настолько серьезные ее преимущества, что диффузия на долгие годы превратилась в основной метод создания $p - n$ переходов в приборах, а затем и в интегральных схемах.

В гл. 2, обсуждая явление диффузии, мы уже упоминали о том, что благодаря процессу диффузии атомы примеси могут проникать с поверхности в объем полупроводника. Сейчас нам предстоит рассмотреть этот процесс значительно подробнее.

Представим себе, что полупроводниковый кристалл помещен в атмосферу газа, содержащего в большой кон-

центрации атомы вещества, служащего в данном полупроводнике донорной (или акцепторной) примесью. На поверхности любого кристалла, как мы знаем, существует

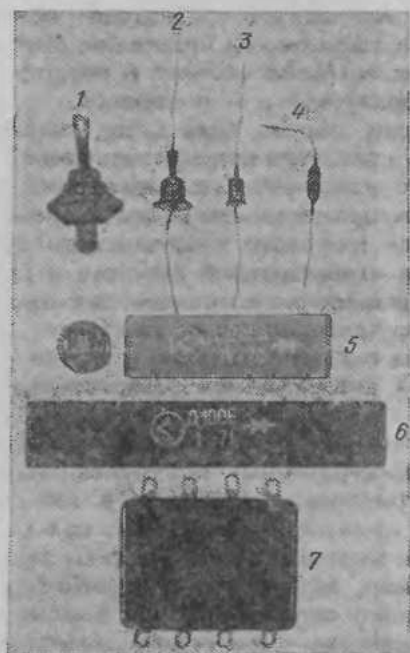


Рис. 46. Примеры германиевых и кремниевых диодов, в которых $p-n$ переход создан методом сплавления: 1 — германиевый диод Д305 предназначен для выпрямления токов до 10 А, напряжений до 50 В; 2 — кремниевый диод Д208: $I \sim 100$ мА, $U \sim 300$ В; 3 — кремниевый сплавной варикап (полупроводниковая емкость, управляемая напряжением), рабочая частота до 50 МГц, $U \sim 50$ В; 4 — кремниевый диод Д223: $I \sim 50$ мА, $U \sim 100$ В, γ до 20 МГц; 5 — выпрямительный «столб» Д1005 (несколько последовательно соединенных сплавных диодов залиты эпоксидной смолой): $I \sim 100$ мА, $U \sim 4000$ В; 6 — выпрямительный столб Д1006: $I \sim 100$ мА, $U \sim 6000$ В; 7 — блок из кремниевых сплавных диодов КЦ 401 Г (внутри блока находятся несколько диодов, соединенных в готовую электрическую схему)

множество нарушений кристаллической решетки. Поэтому атомы примеси из газа-диффузанта легко захватываются поверхностью полупроводника. Поверхность оказывается обогащенной примесью. А что дальше? Как примесь может проникнуть с поверхности в глубь кристалла?

Ведь речь идет не об электронах и дырках — носителях заряда, способных свободно перемещаться внутри кристалла. Мы обсуждаем поведение атомов. Во-первых, атомы самой примеси закреплены на поверхности кристалла. А во-вторых, что пожалуй еще важнее, атомы полупроводника, на место которых должны встать атомы примеси, также закреплены (и притом очень прочно) в узлах кристаллической решетки. Как же в таких условиях возможна диффузия?

Действительно, в абсолютно совершенном кристалле при абсолютном нуле температуры атомная диффузия практически отсутствует.

Но абсолютно совершенных кристаллов не бывает. В любом реальном кристалле, даже при низкой температуре всегда существуют нарушения в идеальной кристаллической решетке. А если кристалл нагреть, то тепловые колебания решетки будут приводить к тому, что некоторые атомы полупроводника в объеме начнут выскакивать со своих законных мест в узлах решетки. На освободившееся место (вакансию) может перепрыгнуть другой атом, в том числе и атом примеси. Выброшенным со своих мест в узлах решетки в междоузлия атомам, и примесным, и собственным, путешествовать по кристаллу уже значительно легче. Таким образом, диффузия примесей осуществляется главным образом по вакансиям, междоузлиям и другим нарушениям в решетке кристалла полупроводника *).

Диффузионный поток примесных атомов направлен, естественно, от поверхности, где примесных атомов много, в глубь полупроводника, где их нет. Величина диффузионного потока, т. е. количество атомов примеси, пересекающих в единицу времени единичную площадь, определяется уравнением (18).

Но значение коэффициента диффузии D , определяющего скорость атомной диффузии, как правило, в миллионы миллиардов раз меньше, чем коэффициент диффузии электронов или дырок.

Коэффициент диффузии свободных носителей при температуре ~ 300 К и выше, как правило, уменьшается с ростом температуры. Коэффициент диффузии примесей с увеличением температуры очень резко, экспоненциально,

*) Подробно о процессах атомной диффузии можно прочитать в книге: Бокштейн В. С. Атомы блуждают по кристаллу. — М.: Наука, 1983. — Библиотечка «Квант», вып. 28.

возрастает (рис. 47). Причины, по которым коэффициенты диффузии свободных носителей и атомов примеси так по-разному зависят от температуры, достаточно очевидны. Для электронов или дырок повышение температуры означает усиление хаотических колебаний атомов в решетке, мешающих и направленному движению носителей вдоль электрического поля, и процессу диффузии (см. комментарий к формуле (24)). Никаких новых возможностей для диффузии свободных носителей увеличение температуры не создает. Иное дело — диффузия примесей. В этом процессе основной эффект, к которому приводит повышение температуры, состоит в экспоненциальном увеличении числа вакансий в междоузлиях, способствующих атомной диффузии *).

На рис. 48 показана схема проведения процесса диффузии. Кварцевая труба 1 помещается в печь, которая создает внутри трубы температуру 1000—1300 °С. В трубе находится кварцевая кассета 3 с закрепленными на ней пластинами кремния. Вещество, атомы которого должны создать необходимую примесь в кремнии (так называемый диффузанта), поступает в

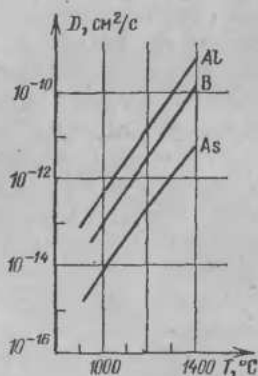


Рис. 47. Зависимость от температуры коэффициентов диффузии примесей в кремнии

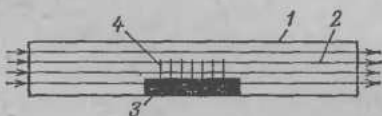


Рис. 48. Схема проведения процесса диффузии: 1 — кварцевая труба; 2 — поток газообразного диффузанта; 3 — кварцевая кассета-держатель, в которой закреплены пластины кремния; 4 — пластины кремния

трубу в виде газа строго контролируемой концентрации. В течение всего времени, пока поверхность полу-

*) Взгляните на формулу (1), затем внимательно посмотрите на рис. 47 и рассчитайте энергию активации $\Delta\mathcal{E}$ процесса, определяющего диффузию примесей в кремнии. (Правильный ответ: ~ 3 эВ.)

проводника обдувается диффузантом, атомы примеси диффундируют в глубь пластины.

Существуют и другие модификации процесса диффузии. В некоторых случаях на поверхность кремниевой пластины наносят слой вещества, содержащего нужную примесь большой концентрации, и диффузия происходит не из газовой, а из твердой фазы.

Расстояние, на которое продвинулись в глубь полупроводника атомы примеси в процессе диффузии, можно приблизительно оценить по формуле (25), зная значение коэффициента диффузии при температуре процесса (рис. 47) и время диффузии. Однако такой грубой оценки при изготовлении $p-n$ переходов для современных приборов оказывается недостаточно. Точный расчет позволяет по известному значению коэффициента диффузии D и времени процесса t определить профиль распределения концентрации введенной примеси, т. е. зависимость концентрации введенной примеси N от расстояния до поверхности полупроводника (рис. 49). Чем больше время диффузии, тем глубже, естественно, проникают примеси в кристалл.

$p-n$ переход при диффузии возникает вследствие того, что кристалл, в который проводится диффузия, имеет исходную объемную концентрацию примеси N_v^*), противоположную по типу примеси, вводимой при диффузии. Эта концентрация, не зависящая от координаты x , показана на рис. 49 штриховой прямой. Если исходная примесь, например, донорного типа, то для создания $p-n$ перехода используют диффузию акцепторов. Для создания перехода в исходном кристалле p -типа — диффузию доноров.

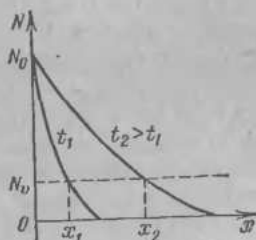


Рис. 49. Распределение диффундировавшей примеси по глубине и образование $p-n$ перехода. Точка $x = 0$ соответствует поверхности полупроводника. При времени диффузии t_1 $p-n$ переход находится на расстоянии x_1 от поверхности. Действительно, правее точки x_1 концентрация исходной примеси больше, чем введенной, и кристалл имеет проводимость исходного типа. Левее точки x_1 концентрация введенной примеси больше, и кристалл имеет проводимость противоположного типа. При времени диффузии t_2 глубина залегания $p-n$ перехода x_2

*) Индекс «v» — от английского слова volume (объем).

Глубина залегания $p - n$ перехода, т. е. расстояние от $p - n$ перехода до поверхности кристалла, зависит, как ясно из рис. 49, не только от профиля концентрации введенной примеси, но и от исходного уровня легирования N_v . Чем больше значение N_v , тем ближе к поверхности будет располагаться $p - n$ переход.

Глубины залегания $p - n$ переходов, изготовляемых методом диффузии, лежат в пределах от долей микрометра до сотен микрометров. Точность, с которой может быть изготовлен $p - n$ переход с желаемым профилем концентрации и глубиной залегания, зависит прежде всего от точности, с которой поддерживается температурный режим в процессе диффузии. В современных диффузионных печах закон изменения температуры со временем (нагрев — выдержка — остывание) контролируется ЭВМ по специальной программе, а температура поддерживается с погрешностью не более $\pm 0,2$ °C, т. е. в сотые доли процента.

Ионная имплантация. Это один из самых современных методов изготовления $p - n$ переходов, позволяющий получать $p - n$ переходы очень маленьких размеров, с точно контролируемой и очень малой (до сотых долей микрометра) глубиной залегания.

Метод ионной имплантации, или, как его иногда называют, метод ионного легирования, состоит во внедрении в полупроводник атомов, которые не мечутся хаотически во всех направлениях, как при диффузии, а движутся строго в одном направлении с определенной и довольно большой скоростью.

Установка для проведения ионной имплантации схематически показана на рис. 50. В камере 1, где происходит процесс, создается и поддерживается глубокий вакуум. Ионы примеси возникают в источнике ионов 2, который устроен следующим образом. Материал примеси в источнике нагревается до такой температуры, чтобы началось испарение атомов с поверхности. Вблизи поверхности горит электрическая дуга, которая ионизирует практически все попадающие в нее частицы.

Образовавшиеся положительные ионы примеси вытягиваются через выходное отверстие источника сильным электрическим полем. Особая система электродов (называемая электростатической линзой) придает пучку нужную иглообразную форму.

Ускоренные полем до большой скорости и сфокусированные ионы попадают в магнитный анализатор 3. Это

важнейший узел, который обеспечивает одно из самых главных достоинств метода ионного легирования: практически полную очистку пучка ионов от посторонних примесей. Принцип работы анализатора во многом аналогичен принципу действия масс-спектрографа. В анализаторе пучок ионов попадает в поперечное по отношению к пучку магнитное поле (силовые линии магнитного поля B на рис. 50 перпендикулярны плоскости рисунка). Как известно, летящие заряженные частицы отклоняются в магнитном поле под действием силы Лоренца (Физика, 9 класс). Радиус траектории ионов, попавших в магнитное поле, зависит от их скорости и от массы ионов. Поэтому, если в пучке окажутся чужеродные частицы, помимо ионов легирующей примеси, они отклонятся в анализаторе не на тот угол, на который отклоняются ионы примеси.

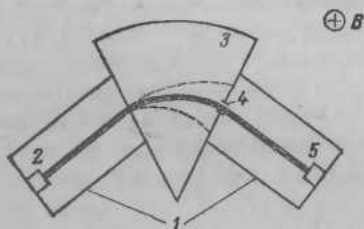


Рис. 50. Схема установки ионной имплантации: 1 — камера, в которой происходит процесс; 2 — источник ионов; 3 — магнитный анализатор с выходной щелью 4; 5 — «мишень» — полупроводниковая пластина, на которой создается $p-n$ переход

В выходной стенке анализатора имеется щель, в которую проходят только те ионы, которые отклонились строго на заданный угол. «Выбирать» тип ионов в пучке очень легко: достаточно изменять величину магнитного поля анализатора.

Вылетевшие из анализатора «отфильтрованные» ионы примеси подлетают к полупроводниковой пластине. При столкновениях с решеткой полупроводника ионы постепенно теряют свою энергию, отдавая ее атомам, попадающим на их пути, и в конце концов останавливаются. Путь, который ион пройдет до полной остановки, или глубина внедрения, зависит от массы и энергии иона. Энергию ионов (обычно она лежит в диапазоне 10—300 кэВ) можно очень точно регулировать, изменяя ускоряющий ионы электрический потенциал. Таким образом, внедрение ионов примеси в решетку происходит на строго заданную глубину. Например, ионы бора B^+ при энергии 20 кэВ внедряются в кремниевую пластину на глубину 0,071 мкм, при энергии 100 кэВ — на 0,327 мкм, а при энергии 300 кэВ — на 0,559 мкм. Более тяжелые ионы фосфора

при тех же значениях энергии пролетают в кристалле кремния 0,026; 0,122 и 0,251 мкм.

Концентрация введенной примеси определяется суммарной дозой облучения, т. е. полным количеством ионов, внедрившихся в кристалл. Эта величина с высокой точностью контролируется временем ионной бомбардировки.

Таким образом, ионная имплантация — один из самых точных методов изготовления $p-n$ переходов (рис. 51). Любая нужная примесь вводится этим методом в полупроводник в строго контролируемом количестве и на заданную глубину, а все посторонние примеси отфильтровываются.

В некоторых установках ионной имплантации на пути ионного пучка помещаются электростатические отклоняющие пластины, такие же, как в электронно-лучевых трубках телевизоров и осциллографов. Подавая на эти пластины напряжение, изменяющееся по заданному закону, вырабатываемому управляющей ЭВМ, можно отклонять ионную «иглу» в нужную точку пластины. Таким образом, ионный луч прямо «рисует» необходимую конфигурацию $p-n$ переходов. Закончив весь «рисунок» для одной примеси и изменив магнитное поле анализатора и энергию пучка, можно сразу же выполнить другой рисунок, создающий другую систему $p-n$ переходов на другой глубине, и т. д. Такие установки используются для изготовления интегральных схем, в которых на одной полупроводниковой пластине размещаются тысячи, десятки тысяч, а иногда

Рис. 51. Пример распределения примеси, введенной в кристалл методом ионной имплантации. Современные методы контроля позволяют точно установить профиль распределения примеси, даже если концентрация уменьшается в 100 000 раз на расстоянии в 1 мкм

даже сотни тысяч $p-n$ переходов.

Не следует думать, что при изготовлении любого полупроводникового прибора обязательно используется только один определенный способ создания $p-n$ перехода. На рис. 52, а показана фотография полупроводникового прибора — тиристора и приведена схема его устройства

(рис. 52, б), из которой видно, что этот прибор состоит из трех чередующихся $p-n$ переходов. Так вот, в тиристоре КУ-103, показанном на рис. 52, а нижний $p-n$

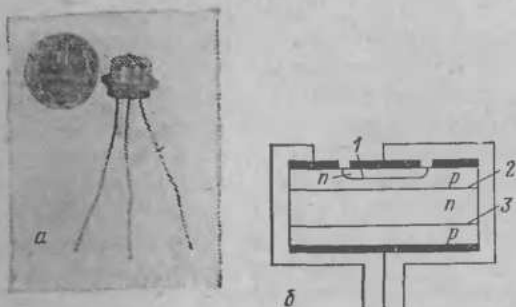


Рис. 52. Полупроводниковый прибор — тиристор типа КУ-103: а — внешний вид прибора; б — схематическое изображение устройства прибора: прибор состоит из трех чередующихся $p-n$ переходов. Верхний переход 1 создан ионной имплантацией, переход 2 — диффузией, 3 — сплавлением

переход изготавливается сплавлением, центральный — диффузией, а верхний — ионной имплантацией.

Барьер на границе



Познакомившись с тем, как создаются $p-n$ переходы, посмотрим теперь, к каким следствиям приводит появление $p-n$ перехода в кристалле.

Самое важное следствие состоит в том, что на границе p - и n -областей возникает энергетический (потенциальный) барьер, свойства которого и определяют все те замечательные особенности $p-n$ перехода, о которых говорилось в начале этой главы.

Обсудим прежде всего, почему между двумя областями одного и того же полупроводника, которые отличаются друг от друга всего лишь характером введенной примеси, возникает потенциальный барьер.

На рис. 53, а слева показан полупроводник p -типа, справа — n -типа. Дырочный полупроводник, как мы знаем, содержит при обычных условиях положительно заряженные дырки — подвижные частицы, способные переносить ток, и отрицательно заряженные акцепторы, непод-

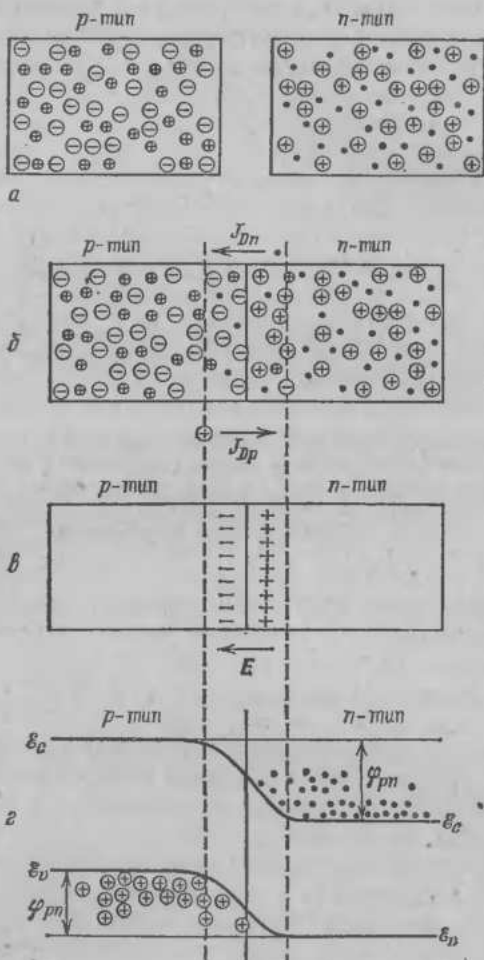


Рис. 53. Образование потенциального барьера на границе $p - n$ перехода: *a* — полупроводники p - и n -типа, в целом нейтральные; *б* — электроны диффундируют в полупроводник p -типа, а дырки — в полупроводник n -типа. Диффузионный поток электронов J_{Dn} приводит к обеднению электронами области n -полупроводника, непосредственно прилегающей к $p - n$ переходу. Диффузионный поток дырок J_{Dp} приводит к обеднению дырками области p -полупроводника, прилегающей к переходу; *в* — у границы $p - n$ перехода возникает двойной заряженный слой (ср. с рис. 28); ϵ — наличие двойного заряженного слоя на границе $p - n$ перехода эквивалентно существованию на границе энергетического барьера. Этот барьер мешает проникновению электронов из полупроводника n -типа в полупроводник p -типа и проникновению дырок в полупроводник n -типа

вижно закрепленные в кристаллической решетке. Электронный полупроводник, наоборот, содержит отрицательно заряженные электроны проводимости и положительно заряженные ионизованные доноры, неподвижно закрепленные в решетке.

На рис. 53, б показано, что произойдет, если мы, на миг превратившись в волшебников, сумеем соединить два этих кусочка полупроводника. (Но не просто прижать друг к другу, а соединить так, чтобы по мановению волшебной палочки с границ этих кусочков исчезли все окислы, адсорбированные атомы, нарушения кристаллической решетки. Чтобы атомы на границе p -полупроводника соединились с атомами на границе n -полупроводника в идеальную кристаллическую решетку.)

Неизбежно должен возникнуть диффузионный поток электронов из полупроводника n -типа в полупроводник p -типа. И наоборот — диффузионный поток дырок из p - в n -полупроводник.

Электроны будут диффундировать из полупроводника n -типа, где их много, в дырочный полупроводник. При этом прилегающая к p — n переходу часть полупроводника p -типа будет заряжаться отрицательно. Прилегающая к p — n переходу часть полупроводника n -типа, наоборот, приобретет положительный заряд, так как в ней остается нескомпенсированным положительный заряд неподвижных доноров.

К точно такому же эффекту приведет и диффузия дырок из p - в n -область. И этот процесс приводит к тому, что прилегающая к границе область p -типа будет заряжаться отрицательно (в ней остается нескомпенсированным отрицательный заряд неподвижных акцепторов), а прилегающая к границе область n -типа — положительно (рис. 53, б).

Понятно, что вблизи p — n перехода образуется двойной заряженный слой, отрицательный со стороны p -области, положительный со стороны n -области (рис. 53, в).

В гл. 3 мы достаточно подробно познакомились со свойствами двойного заряженного слоя и понимаем, что на любой электрон, попавший в такой слой, будет действовать сила, стремящаяся вытолкнуть электрон в n -область (ср. с рис. 28). А дырку, попавшую в двойной слой, электрическое поле слоя, наоборот, будет стремиться вытолкнуть в p -область.

Таким образом, электрическое поле двойного слоя, возникшего на границе p - и n -областей, противодействует процессу диффузии. В самом деле, диффузия стремится

создать поток электронов из n -области в p -область, а поле заряженного слоя — наоборот, втянуть электроны в n -область. Точно так же поле в $p - n$ переходе противодействует и диффузии дырок из p - в n -область.

В конце концов, в результате двух процессов, действующих в противоположные стороны: диффузии и движения зарядов в электрическом поле, устанавливается стационарное (не зависящее от времени) состояние.

В стационарном состоянии диффузионный поток электронов справа налево в точности скомпенсирован потоком электронов в электрическом поле E (полевым потоком) слева направо. Разумеется, то же самое относится и к дыркам: диффузионный поток дырок из p - в n -область полностью скомпенсирован встречным полевым потоком.

Итак, на границе $p - n$ перехода существует двойной заряженный слой, препятствующий проникновению электронов из n - в p -полупроводник, а дырок, наоборот, — из p - в n -полупроводник. Но наличие двойного слоя, как мы знаем из гл. 3, эквивалентно существованию энергетического (потенциального) барьера (рис. 53, z).

Вспомним еще раз, почему наличие двойного слоя между областями эквивалентно существованию между этими областями энергетического барьера.

Внутри слоя действует электрическое поле E , в любой точке направленное в одну сторону (на рис. 53, z — справа налево). Это означает, что заряженная частица, попавшая в слой, в любой точке слоя испытывает действие силы, направленной в одну и ту же сторону (легко убедиться, что эта сила будет стремиться вернуть в n -область любой попавший в нее электрон (ср. с рис. 25), в p -область — дырку). Для того чтобы попасть из n -слоя в p -слоя, электрон должен преодолеть действие этой силы и, следовательно, затратить определенную энергию, т. е. преодолеть энергетический барьер. Естественно, то же самое относится и к дырке в p -области. Чтобы попасть в n -область, ей необходимо затратить ту же самую энергию, равную работе по преодолению силы, препятствующей движению через заряженный слой.

Рис. 53, z поясняет ситуацию на языке энергетической диаграммы. Видно, что электрону для того, чтобы попасть из n -области в p -область, необходимо преодолеть энергетический барьер высотой Φ_{pn} . Такой же барьер необходимо преодолеть и дырке для того, чтобы попасть из p - в n -область $p - n$ перехода.

Высота барьера. Чтобы определить высоту барьера, т. е. величину энергии Φ_{pn} , обсудим еще раз, почему на границе между p - и n -областями возникает энергетический барьер. Но на этот раз с самого начала будем обсуждать задачу на языке энергетических диаграмм.

На рис. 54, а показаны уже хорошо знакомые нам диаграммы (ср. с рис. 31): слева — для p -, справа — для n -типа полупроводника. Если на границах p - и n -кристаллов отсутствуют поверхностные состояния и связанный с ними поверхностный заряд *), то работа выхода из материала n -типа φ_n меньше, чем работа выхода из p -материала φ_p на величину, приблизительно равную ширине запрещенной зоны полупроводника \mathcal{E}_g . То обстоятельство, что φ_n существенно меньше, чем φ_p , означает, конечно, что электронам значительно легче выходить наружу (в вакуум) из n -материала, чем из материала p -типа.

Сближим теперь кристаллы p - и n -типов так, чтобы они могли обмениваться вылетевшими наружу электронами (рис. 54, б).

Естественно, возникнут два потока электронов, направленных навстречу друг другу: поток электронов J_{np} , направленных из кристалла n -типа в кристалл p -типа, и поток электронов J_{pn} — из кристалла p -типа в кристалл n -типа. Понятно также, что $J_{np} > J_{pn}$.

Поскольку электронов в p -материал переходит больше, чем в n -материал, то ясно, что p -область начнет заряжаться отрицательно относительно n -области. Это обстоятельство будет затруднять дальнейший переход электронов из n - в p -область, и, в конце концов, установится стационарное состояние, при котором оба потока выравняются, т. е. поток электронов из p - в n -область будет равен встречному потоку из n - в p -область.

Когда же это произойдет?

Тогда, естественно, когда уровни энергии, с которых осуществляются переходы электронов, выравняются и займут одно и то же положение на энергетической диаграмме!

Пока уровень энергии, на котором находятся электроны, в p -области ниже, чем в n (рис. 54, б), электронам энергетически выгодно переходить из n - в p -область. Стационарное состояние установится только тогда, когда ни переход из n - в p -область, ни переход обратно не будет приводить ни к приобретению энергии электроном, ни к потере энергии — т. е. тогда, когда уровням энергии, на которых находятся электроны в p - и n -области, будет соответствовать одно и то же значение энергии.

*) Мы, конечно, понимаем, что так не бывает. Но, с другой стороны, нетрудно догадаться, почему понадобилась такая идеализация и чем она оправдана. Сейчас мы сдвинем вместе p - и n -области в образуем $p-n$ переход, в котором действительно нет никаких поверхностных состояний между p - и n -областями.

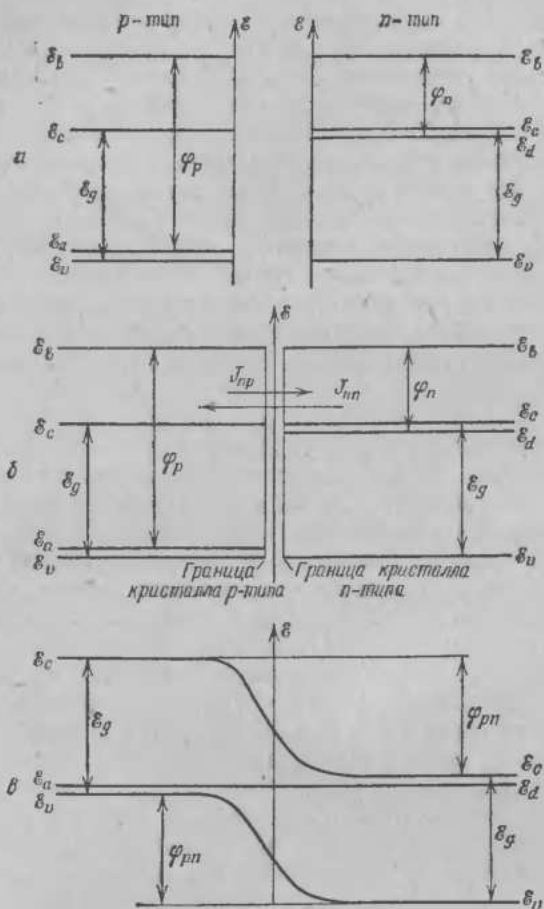


Рис. 54. Образование барьера на границе $p - n$ перехода и его высота φ_{pn} : а — работа выхода из кристалла n -типа φ_n меньше, чем работа выхода из кристалла p -типа φ_p . (Показан случай, когда электроны в кристалле n -типа в основном находятся на мелком донорном уровне ε_d , а в кристалле p -типа — на мелком акцепторном уровне ε_a .); б — поток электронов из кристалла n -типа J_{nn} больше, чем встречный поток из кристалла p -типа J_{np} ; в — потоки электронов выравниваются и стационарное состояние установится тогда, когда сравняются уровни энергии, с которых осуществляется переход электронов из одной части кристалла в другую. При этом возникает энергетический (потенциальный) барьер $\varphi_{pn} \approx \varepsilon_g$

Такому стационарному состоянию соответствует энергетическая диаграмма (рис. 54, *е*), имеющая уже знакомый нам вид (ср. с рис. 53, *з*). Из рис. 54, *е* видно, что теперь уровню \mathcal{E}_a , на котором находятся электроны в *p*-области, и уровню \mathcal{E}_d , на котором находятся электроны в *n*-области, соответствует одно и то же значение энергии. При этом между *p*- и *n*-областями образуется энергетический барьер высотой Φ_{pn} .

Из сопоставления рис. 54, *б* и 54, *е*, очевидно, следует, что

$$\Phi_{pn} = \Phi_p - \Phi_n. \quad (34)$$

В рассматриваемом случае, когда в области *p*-типа электроны находятся на мелком акцепторном уровне, а в области *n*-типа — на мелком донорном (рис. 54, *а*), разность $\Phi_p - \Phi_n$ приблизительно равняется \mathcal{E}_g и, следовательно, $\Phi_{pn} \approx \mathcal{E}_g$.

Обедненный слой. Итак, между *p*- и *n*-областями в *p* — *n* переходе существует энергетический барьер, высота которого составляет $\sim 0,7$ эВ для германиевого, ~ 1 эВ для кремниевого, $\sim 1,4$ эВ для GaAs перехода. Любой электрон, который в стационарном состоянии пожелает проникнуть из *n*-области в *p*, или любая дырка, вознамерившаяся попасть из *p*- в *n*-область, должны этот барьер преодолеть (рис. 54, *е*).

Поскольку подавляющая часть электронов в зоне проводимости (и дырок в валентной зоне) имеет энергию порядка kT (т. е. всего 0,026 эВ при комнатной температуре), ясно, что для большинства носителей барьер на границе *p*- и *n*-областей представляет собой гору чудовищной высоты, взобраться на которую носители не в состоянии.

Мало того, подавляющая часть носителей не способна не только преодолеть весь энергетический барьер, но и хоть сколько-нибудь заметно продвинуться в глубь барьера. Поэтому область энергетического барьера представляет собой слой, практически полностью обедненный свободными носителями, электронами и дырками. Это утверждение так важно, что мы разберем его подробнее.

Начнем с электронов. Рассмотрим рис. 55, *а*. В сущности, на нем показана уже знакомая нам картина. Рис. 55, *а* представляет собой верхнюю половину рис. 54, *е*, и на нем показан профиль дна зоны проводимости $\mathcal{E}_c(x)$ для *p* — *n* перехода. Начало отсчета энергии на рис. 55, *а* совмещено с уровнем \mathcal{E}_c в полупроводнике *n*-типа.

Рассмотрим сначала, что происходит вдали от $p-n$ перехода, в области правее точки x_0 (рис. 55, a), отмечающей границу энергетического барьера. В этой области существует равновесная концентрация электронов n_0 , в

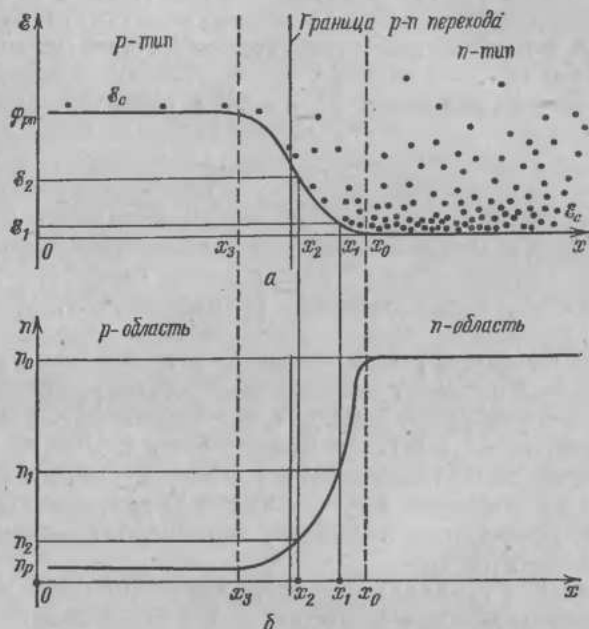


Рис. 55. Концентрация электронов чрезвычайно резко спадает внутри области энергетического барьера. Область барьера представляет собой участок структуры, практически полностью обедненный электронами

точности равная концентрации положительно заряженных понизированных доноров N_d .

Точка x_1 лежит почти у самой границы энергетического барьера. Высота барьера в точке x_1 равняется всего $1/10$ полной высоты барьера φ_{pn} . И все же... И все же в точке x_1 высота барьера \mathcal{E}_1 равняется $\sim 0,1$ эВ для кремниевого и $\sim 0,14$ эВ для GaAs $p-n$ перехода. А это значит, что концентрация электронов в точке x_1 (обозначим ее n_1) гораздо меньше, чем в глубине n -области, правее точки x_0 . В самом деле, как мы знаем (см. формулы (1), (12), а также с. 107), $n_1/n_0 = \exp(-\mathcal{E}_1/kT)$. А следовательно, для кремниевого $p-n$ перехода концентрация электро-

нов в точке x_1 составляет всего $\approx 0,02 n_0$, а для GaAs $p - n$ перехода $\approx 0,004 n_0$.

Как мы видим, уже у самой границы энергетического барьера область перехода оказывается практически полностью обедненной свободными электронами.

Естественно, что в данном случае «чем дальше в лес, тем меньше дров». Чем дальше в глубь барьера, тем меньше концентрация свободных электронов. В точке x_2 , где высота барьера \mathcal{E}_2 равняется приблизительно половине высоты барьера φ_{pn} , концентрация $n_2 \approx n_0 \exp(-\mathcal{E}_2/kT) \approx n_0 \exp(-\varphi_{pn}/2kT) \approx n_0 \exp(-\mathcal{E}_g/2kT)$ равняется $\sim 6 \times 10^{-10} n_0$ для кремниевого и $\sim 2 \cdot 10^{-12} n_0$ для GaAs $p - n$ перехода. Концентрация электронов спадает от точки x_0 к точке x_3 , левой границе барьера, чрезвычайно резко (рис. 55, б).

Точно так же обстоит дело и с дырками (рис. 56). На рис. 56, а изображена нижняя половина рис. 54, в — ход потолка валентной зоны $\mathcal{E}_v(x)$ в $p - n$ переходе. Начало отсчета энергии на рис. 56, а совмещено с потолком валентной зоны \mathcal{E}_v в p -области, а ось энергий направлена не вверх, а вниз. Это сделано, во-первых, для того, чтобы еще раз напомнить, что энергия дырок тем больше, чем ниже они расположены на энергетической диаграмме. Во-вторых, потому, что при таком выборе все формулы, описывающие распределение по энергии дырок в валентной зоне, в точности совпадают с формулами, описывающими распределение электронов в зоне проводимости. И, в-третьих, по причине, которая станет ясной, если вы перевернете рис. 56, а «вверх ногами». Трудно не узнать в перевертыше рис. 55, а, который мы только что обсуждали.

Левее точки x_0 на рис. 56 концентрация дырок p_0 всюду одинакова и равна концентрации отрицательно заряженных акцепторов N_a . В точке x_1 концентрация дырок $p_1 \approx p_0 \exp(-\mathcal{E}_1/kT)$. И если величина \mathcal{E}_1 равна, скажем, 0,2 эВ, то $p_1/p_0 \approx 4 \cdot 10^{-4}$. Дыркам, точно так же, как электронам, трудно проникнуть в глубь $p - n$ перехода. После барьера отбрасывает носители: дырки — в p -область, электроны — в n -область.

Именно поэтому область барьера оказывается обедненной свободными носителями.

Распределение поля в барьере. Ширина барьера. Установив, что область энергетического барьера практически полностью обеднена свободными носителями, можно очень просто рассчитать все основные характеристики

барьера. Справа от границы $p-n$ перехода, в n -области, в каждом кубическом сантиметре полупроводника содержится N_d положительно заряженных доноров. Поскольку,

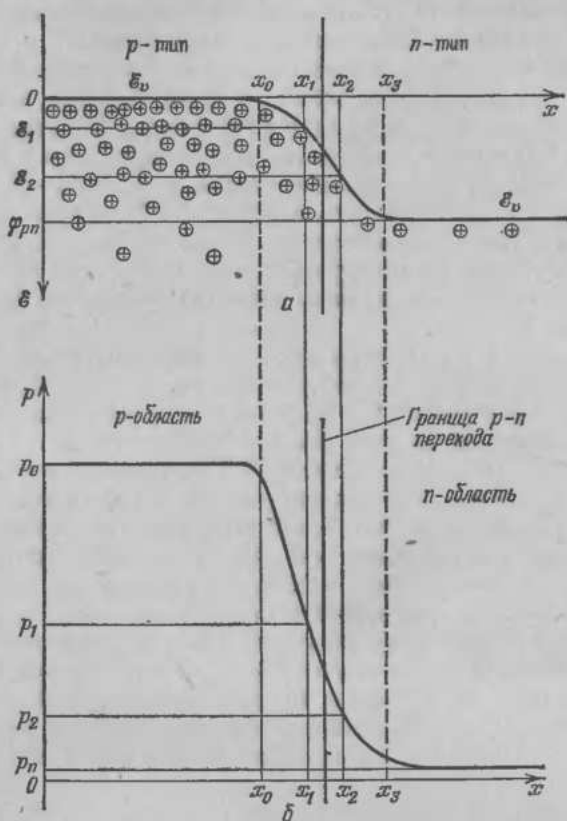


Рис. 56. Дыркам точно так же, как и электронам, трудно проникнуть в область энергетического барьера. Концентрация дырок внутри барьера спадает чрезвычайно резко

как мы только что установили, концентрация электронов и дырок в области барьера ничтожна, ясно, что плотность объемного заряда ρ в барьере будет определяться концентрацией доноров: $\rho = eN_d$. Точно так же в p -области (слева от перехода) плотность объемного заряда определяется концентрацией отрицательно заряженных акцепторов: $\rho = -eN_a$ (рис. 57, б).

В случае, если плотность объемного заряда ρ постоянна, определение распределения поля в барьере $E(x)$ и

ширины барьера W представляет собой простую и уже решенную нами однажды задачу (с. 121, рис. 39). Зависимость $E(x)$ в каждой из «половинок» перехода будет представлять собой прямую. Наклон этой прямой определяется уровнем легирования (формула (31), рис. 39, б).

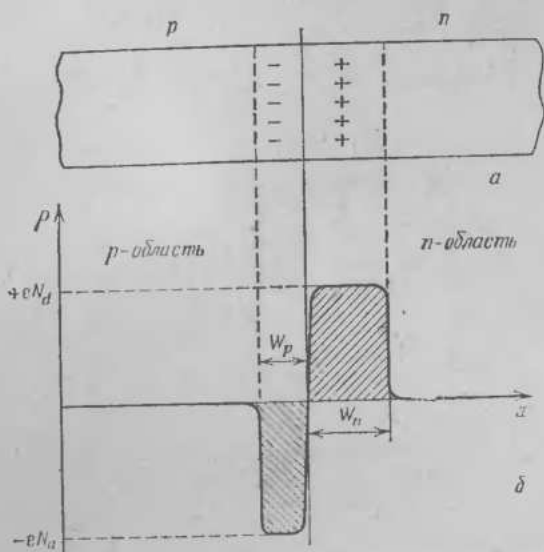


Рис. 57. Распределение объемного заряда в $p-n$ переходе: a — качественное распределение (ср. с рис. 53, a); b — зависимость $\rho(x)$

Зависимость $E(x)$ в $p-n$ переходе показана на рис. 58, б. Напряженность электрического поля максимальна в плоскости $p-n$ перехода и линейно спадает при удалении от перехода. Угол наклона зависимости $E(x)$ в каждой области определяется уровнем легирования (с. 121, рис. 39, б) и тем больше, чем больше концентрация легирующей примеси.

Однако при попытке рассчитать значения E_m и W возникает небольшое затруднение. Если бы распределение поля в $p-n$ переходе имело вид прямоугольного треугольника, то по величине падения напряжения на барьере $U = \varphi/e$ мы легко определили бы и E_m , и W , воспользовавшись формулами (33). Достаточно было бы вместо высоты приповерхностного барьера U_s подставить известное значение напряжения на барьере $U_{pn} = \varphi_{pn}/e \approx \approx \mathcal{E}_g/e$.

Из рис. 58, б, однако, ясно видно, что в случае $p - n$ перехода распределение поля в барьере имеет вид треугольника отнюдь не прямоугольного. Как быть?

Вообще говоря, выход найти очень не сложно. Попробуйте сами догадаться, как обойти эту трудность и рассчитать параметры барьера.

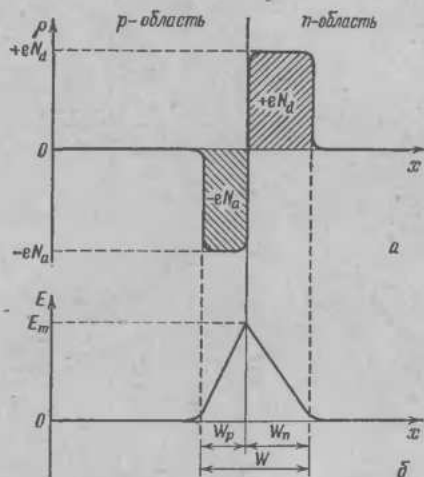


Рис. 58. Распределение объемного заряда $\rho(x)$ (а) и напряженности электрического поля $E(x)$ (б) в $p - n$ переходе

Но, кроме того, в подавляющем большинстве случаев эта трудность вообще не возникает. Дело в том, что почти все используемые на практике $p - n$ структуры являются резко несимметричными. Так называют $p - n$ переходы, в которых одна область легирована гораздо сильнее, чем другая.

На рис. 59 показано распределение поля в $p - n$ переходе, в котором p -область легирована значительно сильнее, чем n -область, т. е. $N_a \gg N_d$. Из рисунка ясно видно, что в этом случае ширина области объемного заряда W_p в p -области значительно меньше, чем W_n (ср. с рис. 58, б), и распределение поля $E(x)$ практически имеет вид прямоугольного треугольника. Практически все напряжение на барьере $U_{pn} = \varphi_{pn}/e \approx 1/2 E_m W_n$ падает на n -области.

В таком случае, разумеется, расчет основных параметров барьера никаких трудностей не представляет (ср. с

формулами (33):

$$W \approx W_n \approx \left(\frac{2\epsilon\epsilon_0 U_{pn}}{eN_d} \right)^{1/2}, \quad E_m \approx \left(\frac{2eN_d U_{pn}}{\epsilon\epsilon_0} \right)^{1/2}. \quad (35)$$

Еще раз напомним, что напряжение на барьере U_{pn} (в вольтах) численно равно энергетической высоте барьера $\approx \mathcal{E}_g$ (в электрон-вольтах).

Чем больше концентрация доноров N_d , тем больше поле в барьере и тем меньше его ширина.

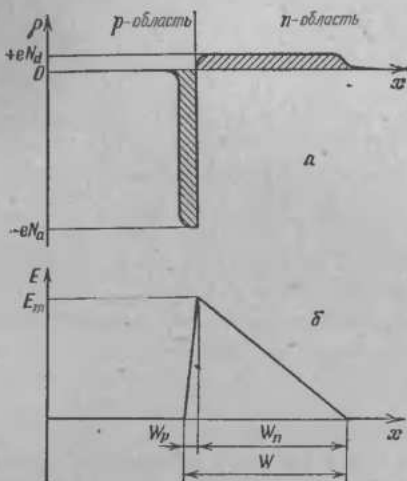


Рис. 59. Распределение объемного заряда $\rho(x)$ (а) и поля $E(x)$ (б) в резко несимметричном $p-n$ переходе (ср. с рис. 58 и 43)

В кремниевом мощном высоковольтном выпрямительном диоде концентрация доноров в n -области $N_d = 10^{13} \text{ см}^{-3}$. (Концентрация акцепторов в p -области $N_a \approx 10^{18} \text{ см}^{-3}$, т. е. в 100 000 раз больше и переход действительно резко несимметричный.) Подставляя в формулу (35) значение $U_{pn} = 1,1 \text{ эВ}$, $\epsilon = 12,5$, $N_d = 10^{19} \text{ м}^{-3}$, получим, что в таком диоде поле в барьере $E_m \approx 1,8 \times 10^3 \text{ В/см}$, а ширина барьера $W \approx W_n \approx 12 \text{ мкм}$ *).

В кремниевых диодах, предназначенных для выпрямления низких напряжений, концентрация доноров в n -области может составлять $\sim 10^{17} \text{ см}^{-3}$. В таком диоде

*). Надеюсь, что вы не поленились нас проверить и убедились, что $1/2 E_m W \approx 1,1 \text{ В}$.

ширина области объемного заряда составляет всего $\sim 0,12$ мкм, а максимальное поле в $p-n$ переходе — $\sim 1,8 \cdot 10^5$ В/см.

Чудесное равновесие. Представим себе, что перед нами лежит крохотный кусочек полупроводника, от которого отходят две металлические проволоочки (рис. 60, а). Пусть

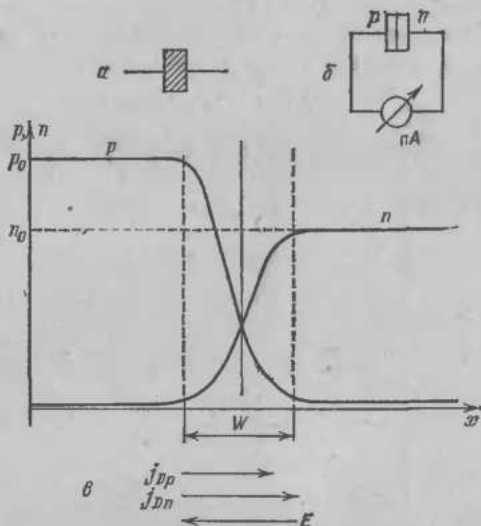


Рис. 60. Если к $p-n$ переходу не приложено внешнее напряжение, ток через него не протекает, хотя в области барьера существуют громадные производные от концентрации по координате dn/dx и dp/dx (ср. с рис. 55, б и 56, б). В любой точке внутри барьера электрическое поле E создает полевой ток, направленный навстречу диффузионному и полностью его компенсирующий

кто-нибудь сообщил нам, что в левую часть кристаллика введена акцепторная примесь, а в правую — донорная. Тогда мы знаем, что внутри кусочка существует $p-n$ переход, и получаем право впрямь именовать его полупроводниковым диодом *).

Кстати, мы вполне можем не полагаться на слова неведомого доброжелателя и проверить факт наличия $p-n$ перехода экспериментально. Для каждого широко используемого полупроводникового материала, в частности для Ge, Si, GaAs, технологи сумели подобрать такие травители, в которых материалы p - и n -типа

*) Полупроводниковым диодом чаще всего принято называть полупроводниковый прибор, который имеет два внешних вывода и внутри которого имеется один энергетический барьер.

страиваются с разной скоростью. Такие травители называют селективными. Если опустить на некоторое время в такой травитель кристаллик полупроводника, содержащий $p-n$ переход, то на месте $p-n$ перехода образуется «ступенька», хорошо различимая в обычный оптический микроскоп.

Давайте замкнем внешние выводы диода через очень чувствительный гальванометр (рис. 60, б). Если в чем-нибудь можно быть уверенным на этом свете, так это в том, что никакого тока гальванометр не регистрирует, какой бы чувствительностью он ни обладал. В самом деле, наличие тока означало бы нарушение закона сохранения энергии: мы получали бы энергию (хотя бы ту, которая идет на джоулев нагрев проводников), не тратя никакой другой.

С другой стороны, если вспомнить то, что мы уже успели узнать о $p-n$ переходе, поневоле хочется воскликнуть вместе с Фамусовым: «Куда как чуден создан свет. Пофилософствуй — ум вскружится...»

В пределах обедненного слоя, на расстоянии в несколько микрометров или даже долей микрометра, концентрация свободных носителей падает в миллионы и миллиарды раз (рис. 60, в). Следовательно, в этой области значения производных от концентрации по координате для электронов dn/dx и для дырок dp/dx очень велики. А следовательно, в соответствии с выражением (21), в области барьера существуют и большие диффузионные токи *). Почему же ток во внешней цепи равен нулю?

В каждой точке барьера электрическое поле создает полевой ток, направленный навстречу диффузионному и в точности его компенсирующий!

Чтобы представить себе, как велики диффузионные и полевые токи, уравнивающие друг друга внутри барьера, оценим, например, плотность диффузионного тока. Мы вполне могли бы, если бы захотели, рассчитать точные зависимости от координаты плотности диффузионного тока дырок j_{Dp} и электронов j_{Dn} . Однако ограничимся более простой задачей и оценим значения j_{Dp} и j_{Dn} по порядку величины. Для этого примем, что внутри обедненного слоя концентрация дырок $p(x)$ и концентрация электронов $n(x)$ спадают линейно, т. е. что в пределах области объемного заряда значения производных dn/dx и dp/dx постоянны.

Для примера предположим, что рассматривается кремниевый $p-n$ переход с концентрацией доноров в слаболегированной об-

*) Напомним, что хотя диффузионные потоки электронов и дырок направлены в $p-n$ переходе навстречу друг другу, диффузионные токи электронов и дырок направлены в одну и ту же сторону и, следовательно, складываются.

ласти 10^{18} см^{-3} . Концентрация акцепторов в сильнолегированной области N_a пусть равняется 10^{18} см^{-3} . Переход резко несимметричный, поэтому $W \approx W_n$. Воспользовавшись формулой (35), легко установить, что $W \approx W_n \approx 0,4 \text{ мкм}$. Средние значения производных dn/dx и dp/dx равняются

$$\frac{dn}{dx} \sim \frac{n_0}{W} \approx \frac{10^{19}}{0,4 \cdot 10^{-4}} \approx 2,5 \cdot 10^{20} \text{ см}^{-4},$$

$$\frac{dp}{dx} \sim \frac{p_0}{W} \approx \frac{10^{18}}{0,4 \cdot 10^{-4}} \approx 2,5 \cdot 10^{22} \text{ см}^{-4}.$$

Чтобы рассчитать величины j_{Dp} и j_{Dn} , нам потребуется вычислить еще значения коэффициентов диффузии электронов и дырок в кремнии: D_n и D_p . Воспользовавшись соотношением Эйнштейна (24) и значениями подвижностей для электронов и дырок, указанными в табл. 2, найдем, что в кремнии при комнатной температуре $D_n = 34 \text{ см}^2/\text{с}$, $D_p = 13 \text{ см}^2/\text{с}$. Затем по формуле (21) определим, что в рассматриваемом случае $j_{Dn} \approx 1400 \text{ А/см}^2$, $j_{Dp} \approx 54 \text{ 000 А/см}^2$. Полная плотность диффузионного тока будет равна $j_D = j_{Dn} + j_{Dp} \approx 5,5 \cdot 10^4 \text{ А/см}^2$.

С ростом уровня легирования плотность диффузионного тока растет, так как значения n_0 и p_0 увеличиваются, а ширина области объемного заряда уменьшается. Для вполне реальных уровней легирования $N \sim 10^{17} + 10^{19} \text{ см}^{-3}$ величина j_D может составлять миллионы ампер/см².

И все-таки стрелка гальванометра, способного зарегистрировать даже ток в 1 пА ($1 \text{ пА} = 10^{-12} \text{ А}$, читается «пикоампер»), не отклонится ни на одно деление. С такой точностью в каждой точке области объемного заряда (барьера) ток, обусловленный полем барьера, компенсирует диффузионный ток.

Следует обратить внимание еще на один чудесный вид равновесия, который обязательно должен существовать для того, чтобы через $p-n$ переход (не подключенный ни в какую электрическую цепь) не протекал ток.

Как мы убедились, существует мощный поток основных носителей электронов, которые стремятся проникнуть из n - в p -область диода и не могут, так как удерживаются сильным полем барьера. Однако если внимательно посмотреть на энергетическую диаграмму, можно увидеть, что существует и тоненький ручеек неосновных носителей — электронов в p -области, которым поле барьера не мешает, а помогает переходить из p - в n -область. В самом деле, обратимся к рис. 55, *a*. Видно, что для электронов в n -области изгиб дна зоны проводимости \mathcal{E}_c соответствует потенциальному барьеру, который нужно преодолеть, чтобы попасть в p -область. А вот для электронов в p -области энергетическая диаграмма соответствует «горке».

попав на которую электрон тут же скатится вниз, в n -область. (Поле барьера, как мы уже много раз убеждались (см., например, рис. 53), направлено так, что стремится вытолкнуть электрон в n -область.)

Разумеется, точно так же обстоит дело и с неосновными носителями — дырками в n -области диода (рис. 56, а). Для них зависимость $\mathcal{E}_v(x)$ также представляет собой горку, попав на которую дырки «всплывают» вверх, в p -область. (Положительно заряженные дырки поле барьера выталкивает в p -область.)

Так как концентрация неосновных носителей мала, плотность тока, обусловленного потоком электронов из p - в n -область, а дырок из n - в p -область, также невелика. Однако, как нам очень скоро предстоит убедиться, этот ток играет очень важную роль в работе диодов и транзисторов.

По причине, которая станет ясной чуть дальше, в следующем пункте, ток, природу которого мы обсуждаем, называется *током насыщения* и обозначается обычно символом I_s . Этот ток, как мы видели, состоит из двух компонентов: электронного (из p -области в n) и дырочного (из n -области в p).

Определим сначала электронный компонент плотности тока насыщения j_{ns} .

На рис. 61 изображена, в сущности, та же самая картина, что и на рис. 55, а. Разница только в том, что, во-

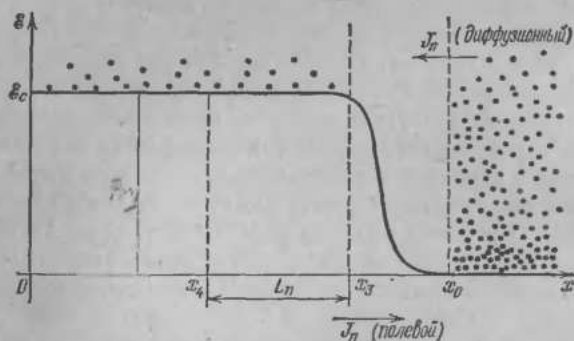


Рис. 61. Для электронов в p -области (неосновных носителей) энергетический барьер $p-n$ перехода представляет собой «горку». Любой электрон, попав из p -области в поле барьера, будет выброшен полем в n -область. Так возникает полевой ток электронов J_n из p - в n -область (нижняя стрелка), т. е. ток насыщения j_{ns} . В стационарном состоянии этот поток компенсируется встречным диффузионным потоком горячих электронов, энергия которых больше, чем высота барьера Φ_{pn} (верхняя стрелка)

первых, для p -области диода отведено больше места (что и естественно, поскольку мы собираемся обсуждать поведение электронов в p -области). А во-вторых, левее точки x_3 , обозначающей, как и на рис. 55, a , границу области объемного заряда, выделена область длиной L_n .

Символом L_n обозначена диффузионная длина электронов, равная, как мы помним (см. формулу (26)), $L_n = \sqrt{D_n \tau_n}$, где D_n — коэффициент диффузии электронов, τ_n — время жизни электронов.

Поговорим сначала о судьбе электронов, существующих в глубине p -области, левее точки x_4 . Каждый из этих электронов, появившись на свет в зоне проводимости (с акцепторного уровня или из валентной зоны, см. рис. 7, b), живет, в среднем, время τ_n . В течение времени жизни он беспорядочно мечется, участвуя в тепловом движении, и удаляется от точки своего рождения на расстояние $\sim L_n = \sqrt{D_n \tau_n}$. Затем он погибает (рекомбинирует).

Число носителей, погибающих в единицу времени в единице объема n^* , будет тем больше, чем больше концентрация электронов n_p и чем меньше время их жизни τ_n : $n^* = n_p / \tau_n$. Но так как в стационарном состоянии концентрация электронов не меняется, то ясно, что точно такое же число носителей n^* и будет генерироваться в единицу времени в единице объема.

А теперь обсудим возможную судьбу электронов, которым суждено было родиться в области между точками x_4 и x_3 . Их участь отличается от участи соседей, родившихся левее точки x_4 , только одним. Родившись и хаотически блуждая в течение времени τ_n по кристаллу, они могут пересечь роковую черту — оказаться правее точки x_3 . Если это произошло, электрическое поле барьера мгновенно подхватит их и выплывнет в n -область. (Они скатятся с потенциальной горки вниз, в точку x_0 .) Вот эти-то родившиеся в p -области на расстоянии $\leq L_n$ от границы области объемного заряда электроны, «переступившие черту», и образуют электронный компонент плотности тока j_{ns} .

Разумеется, не всякий электрон, родившийся между точками x_4 и x_3 , обязательно попадает в течение своей жизни в область объемного заряда. С другой стороны, вполне могут появиться электроны, родившиеся левее точки x_4 , которые случайно прожив существенно дольше, чем среднее время жизни τ_n , добредут до точки x_3 и будут подхвачены полем.

Но чтобы оценить величину j_{ns} , мы предположим, что ни один из электронов левее точки x_4 в область объемного заряда не попадает. А все электроны, родившиеся между точками x_4 и x_3 , на расстоянии диффузионной длины $\sim L_n$ от границы области объемного заряда, будут уноситься полем барьера в n -область.

Тогда число электронов, переносимых через единицу площади барьера из p - в n -область в единицу времени, будет равно числу электронов $n^* = n_p/\tau_n$, рождающихся в единице объема в единицу времени, умноженному на длину L_n . А переносимый этими электронами заряд, т. е. плотность тока j_{ns} , будет равняться

$$j_{ns} = e L_n \frac{n_p}{\tau_n}. \quad (36a)$$

Учитывая, что $n_p = \frac{n_i^2}{N_a}$, формулу для величины j_{ns} часто записывают еще и в виде

$$j_{ns} = e \frac{n_i^2}{N_a} \cdot \frac{L_n}{\tau_n}, \quad (36б)$$

или с учетом, что $L_n = \sqrt{D_n \tau_n}$,

$$j_{ns} = e \frac{n_i^2}{N_a} \cdot \sqrt{\frac{D_n}{\tau_n}}. \quad (36в)$$

Оценим прежде всего, каково по порядку величины значение j_{ns} . Принимая $\tau_n = 10^{-6}$ с, $N_a = 10^{16}$ см⁻³ и пользуясь данными табл. 1 и 2, найдем, что при комнатной температуре для германиевого $p-n$ перехода $j_{ns} \approx 10^{-4}$ А/см², для кремниевого $\sim 10^{-11}$ А/см², для арсенид-галлиевого $\sim 10^{-17}$ А/см².

Это, конечно, не мощные потоки основных носителей, составляющие тысячи и десятки тысяч ампер/см², однако даже для арсенид-галлиевого $p-n$ перехода наличие тока неосновных носителей j_{ns} означает, что каждую секунду через каждый квадратный сантиметр $p-n$ перехода из p - в n -область переходит $j_{ns}/e \approx 10^2$ электронов. А для германиевого $p-n$ перехода $\sim 10^{15}$ электронов.

Между тем условия чудесного равновесия неумолимо требуют, чтобы уход каждого электрона из p - в n -область перехода непременно компенсировался приходом электрона обратно, из n - в p -область диода. Каким же током компенсируется ток j_{ns} ? Откуда берутся электроны, поступающие в стационарном состоянии из n - в p -область?

Как мы уже много раз убеждались, наряду с громадным большинством электронов, энергия которых составляет $\sim kT$, в n -области имеются и очень энергичные, горячие электроны (с. 54, формулы (12)). В том числе и такие, чья энергия \mathcal{E} больше, чем высота энергетического барьера Φ_{pn} . Конечно, концентрация таких электронов n очень мала ($n \approx n_0 \exp(-\Phi_{pn}/kT)$), но ведь и концентрация неосновных носителей, образующих ток насыщения, тоже очень мала! Электроны, энергия которых больше, чем высота барьера Φ_{pn} , поле барьера не в состоянии удержать в n -области. Эти энергичные электроны свободно диффундируют из n - в p -область, и диффузионный ток этих горячих носителей составляет в стационарном состоянии ровно столько электронов в p -область, сколько их уносит ток j_{ns} .

Точно так же, как мы определили электронный компонент j_{ns} тока насыщения, можно определить и дырочный компонент j_{ps} этого тока (желающие могут получить выражение для j_{ps} самостоятельно, используя рис. 56, а):

$$j_{ps} = eL_p \frac{p_n}{\tau_p} = e \frac{n_i^2}{N_d} \frac{L_p}{\tau_p} = e \frac{n_i^2}{N_d} \sqrt{\frac{D_p}{\tau_p}}. \quad (37)$$

Значения времени жизни дырок τ_p в n -области, коэффициента диффузии дырок D_p и диффузионной длины дырок $L_p = \sqrt{D_p \tau_p}$ могут, конечно, отличаться от соответствующих величин для электронов в p -области.

Полная плотность тока насыщения j_s равняется

$$j_s = j_{ns} + j_{ps} = en_i^2 \left(\frac{1}{N_a} \sqrt{\frac{D_n}{\tau_n}} + \frac{1}{N_d} \sqrt{\frac{D_p}{\tau_p}} \right). \quad (38)$$

Если, как это почти всегда и бывает, переход резко несимметричный, то одним из компонентов тока насыщения можно пренебречь. Например, если сильно легированной является p -область перехода, так что $N_a \gg N_d$, то, как видно из формул (37) и (38), $j_{ps} \gg j_{ns}$. Главный вклад в ток насыщения вносят дырки.

Величина j_s пропорциональна квадрату собственной концентрации носителей n_i^2 , поэтому она очень резко зависит от температуры. Из формулы (6) следует, что $j_s \sim n_i^2 \sim \exp(-\mathcal{E}_g/kT)$. Зная значения \mathcal{E}_g (см. табл. 1), легко подсчитать, что если при комнатной температуре (300 К) значение j_s для германиевого диода составляет, скажем, ~ 1 мА/см², то при температуре +80 °С (353 К)

оно составит ~ 60 мА/см², т. е. возрастет в 60 раз. Плотность тока j_s кремниевого диода при увеличении температуры от 300 до 353 К возрастает приблизительно в 600 раз, а арсенид-галлиевого диода — в 3500 раз.

Температура $+80^\circ\text{C}$ соответствует максимально возможной рабочей температуре для германиевых диодов. Кремниевые $p-n$ структуры сохраняют работоспособность до температуры $+150^\circ\text{C}$ (423 К.) При такой температуре плотность тока насыщения в кремниевых диодах возрастает по сравнению с комнатной в $\sim 250\,000$ раз. (Но, однако, все-таки остается меньшей, чем величина j_s в германиевых $p-n$ структурах при комнатной температуре.) Арсенид-галлиевые диоды могут работать при повышении температуры до $+250^\circ\text{C}$ (523 К). При этом величина j_s возрастает, по сравнению с ее значением для комнатной температуры, в $\sim 10^{10}$ раз! И составит (см. пример, иллюстрирующий формулу (36)) всего-навсего $\sim 5 \cdot 10^{-7}$ А/см². Наконец, существуют $p-n$ переходы на основе карбида кремния, сохраняющие работоспособность вплоть до $T \approx 600^\circ\text{C}$! Такие диоды могут совершенно спокойно работать в ванне с расплавленным свинцом. При температуре, близкой к предельной (~ 800 К), ток насыщения возрастает в SiC диодах в $\sim 10^{30}$ раз по отношению к комнатной температуре. И все-таки (проделайте расчет сами, приняв $n_i = 10^{-5}$ см⁻³, $N_a = 10^{17}$ см⁻³, $D_n = 8$ см²/с, $\tau_n = 10^{-8}$ с) значение j_s для SiC диодов при 800 К составляет всего $\sim 5 \cdot 10^{-12}$ А/см².

Чем больше ширина запрещенной зоны \mathcal{E}_g материала, из которого изготовлен диод, тем выше его предельная рабочая температура и тем меньше величина j_s .

Обратное смещение



Налюбовавшись совершенством, с которым компенсируются в $p-n$ переходе потоки электронов и дырок, нарушим чудесное равновесие.

Для этого достаточно приложить к диоду, содержащему $p-n$ переход, напряжение или, как часто говорят, смещение U_0 от внешнего источника (рис. 62). Если «плюс» источника присоединен к n -области диода, а «минус» — к p -области, то принято говорить, что к $p-n$ переходу приложено обратное напряжение (или обратное смещение). Именно такой случай изображен на рис. 62.

Диод, как мы знаем, содержит три области: p -область, n -область и находящуюся между ними область объемного заряда (барьер). По отношению к внешнему источнику U_0 все эти три области включены последовательно. И первый вопрос, на который нам надлежит ответить, — как распределится между этими областями приложенное напряжение.

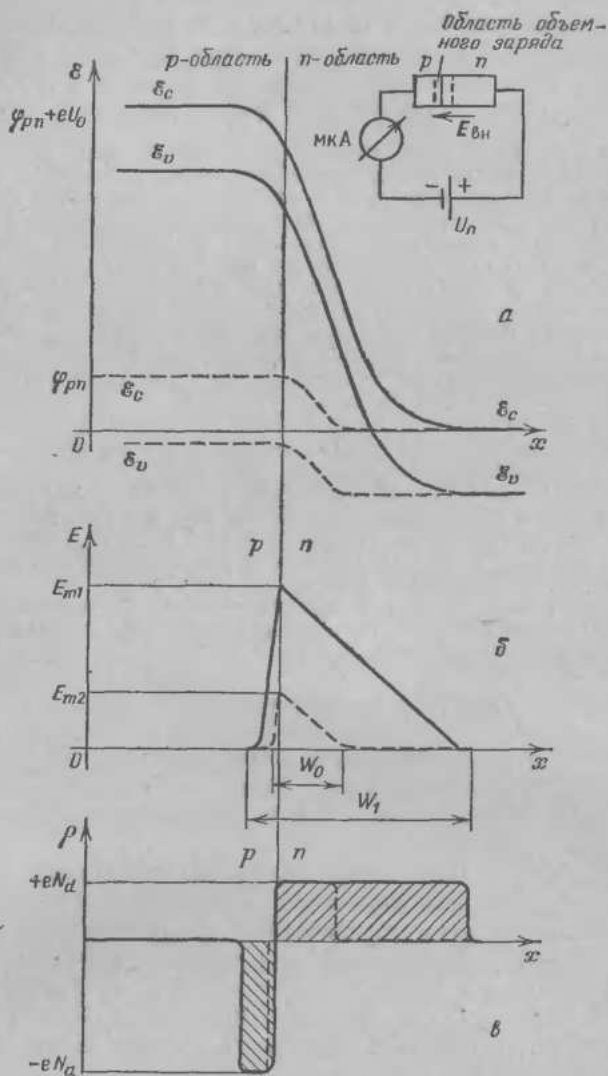


Рис. 62. Энергетическая диаграмма (а), распределение поля (б) и распределение объемного заряда (в) в $p-n$ переходе. Штрих — в случае отсутствия внешнего напряжения. Сплошные кривые — к диоду приложено обратное напряжение U_0 . Высота потенциального барьера увеличивается на величину приложенного напряжения U_0 . Ширина области объемного заряда и величина максимального поля E_m возрастают

Как мы знаем, область объемного заряда очень сильно, практически полностью, обеднена свободными носителями тока — электронами и дырками. Поэтому, естественно, она представляет собой область самого большого сопротивления. И, следовательно, практически все приложенное напряжение U_0 падает на области объемного заряда.

Теперь следует обратить внимание на то, что поле $E_{\text{вн}}$ от внешнего источника U_0 при обратном смещении на диоде совпадает по направлению с полем внутри области объемного заряда (рис. 53, *а*). И, следовательно, падение напряжения от внешнего источника U_0 складывается с напряжением на переходе $U_{pn} = \varphi_{pn}/e$, существовавшем до приложения внешнего смещения. В результате высота потенциального барьера между p - и n -областями становится равной $U_{pn} + U_0$. А высота энергетического барьера соответственно $e(U_{pn} + U_0) = \varphi_{pn} + eU_0$ (рис. 62, *а*).

Высота и форма барьера. Зная высоту барьера, в случае если переход резко несимметричен, мы без труда можем определить и все остальные важнейшие параметры барьера: распределение поля в барьере $E(x)$, величину максимального поля в барьере E_{m1} и ширину барьера W_1 (рис. 62, *б, в*):

$$W_1 \approx W_{n1} \approx \left[\frac{2\epsilon\epsilon_0 (U_{pn} + U_0)}{eN_d} \right]^{1/2} \quad (39)$$

$$E_{m1} \approx \left[\frac{2eN_d (U_{pn} + U_0)}{\epsilon\epsilon_0} \right]^{1/2}.$$

(Ср. с формулой (35).)

Угол наклона зависимостей $E(x)$ к оси абсцисс не меняется с приложением обратного напряжения (рис. 62, *б*).

Видно, что при приложении к диоду обратного напряжения свободные носители заряда вытесняются из областей, прилегающих к $p-n$ переходу: область объемного заряда становится шире (рис. 62, *в*).

Вернувшись к примеру, уже рассматривавшемуся нами раньше (с. 153), рассчитаем значения W и E_m для кремниевого высоковольтного выпрямительного диода, к которому приложено обратное смещение $U_0 = 1000$ В. Из формулы (39) находим $E_{m1} = 54$ кВ/см, $W_1 \approx W_{n1} = 370$ мкм.

С увеличением приложенного обратного напряжения максимальное поле E_m возрастает. Между тем для каждого полупроводникового материала существует предель-

ное значение напряженности электрического поля E_i *), превышать которое «безнаказанно» нельзя. В поле E_i , которое для Si и GaAs составляет $\sim 3 \cdot 10^5$ В/см, электроны и дырки приобретают такую большую энергию, что получают способность при столкновении с атомами кристаллической решетки полупроводника порождать новые электроны и дырки. Это явление, во многом сходное с явлением ионизации газов, называется *ударной ионизацией*. Явление ударной ионизации ограничивает величину максимального обратного напряжения, которое может быть приложено к диоду. Подробно процессы *ударной ионизации* мы обсудим в конце этого раздела.

А пока что обратим внимание на следующее, очень важное обстоятельство. Хотя величина E_i для данного полупроводникового материала практически является константой, максимальное обратное напряжение, которое может быть приложено к $p - n$ переходам, изготовленным из одного и того же материала, может отличаться в тысячи раз.

В самом деле, будем считать, что максимальное обратное напряжение, которое можно приложить к диоду, определяется условием $E_m = E_i$. Определенное таким образом напряжение называется напряжением начала ударной ионизации или, попросту, *напряжением пробоя* $p - n$ перехода. Обычно оно обозначается символом U_i . Из формулы (39) непосредственно вытекает

$$U_i \approx \frac{\epsilon \epsilon_0 E_i^2}{2eN_d} \quad (40)$$

Если $N_d \approx 10^{17}$ см⁻³, то $U_i \approx 3$ В, если $N_d \approx 10^{14}$ см⁻³, то $U_i \approx 3000$ В.

Обратный ток. Среди преподавателей, читающих курс физики полупроводниковых приборов, большой популярностью пользуется анекдот о студенте, очень толково описывавшем на зачете, как работает $p - n$ переход. Студент рассказал о способах получения $p - n$ переходов, о том почему на границе p - и n -областей полупроводника возникает потенциальный барьер. Очень ясно обрисовал, как возникает слой объемного заряда, как в отсутствие внешнего напряжения компенсируются потоки основных и неосновных носителей... Разомлевший преподаватель, уже намереваясь поставить в зачетной книжке «отлично», за-

*) Индекс «i» — от английского слова impact (удар, столкновение).

дал последний дополнительный вопрос: «Чем определяется ток $p - n$ перехода, смещенного в обратном направлении?»

«Чтобы найти ток, протекающий через диод в обратном направлении, — так же бодро и четко ответил студент, глядя прямо в глаза преподавателю, — следует определить по формуле $R_p = \rho_p L/S$ сопротивление p -области перехода, затем по формуле $R_n = \rho_n L/S$ — сопротивление n -области. Затем сложить полученные значения R_p и R_n и вычислить протекающий через переход ток $I = U/(R_p + R_n)$ ».

«Разумеется, — обычно мстительно заключает рассказчик, — этот поросенок вылетел с зачета впереди собственного визга».

Мы сможем лучше понять возмущение преподавателя, взглянув на рис. 63, где показана типичная *вольтамперная характеристика* обратносмещенного германиевого

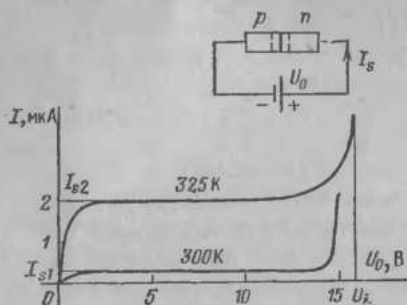


Рис. 63. Вольтамперная характеристика обратносмещенного германиевого $p - n$ перехода

$p - n$ перехода, т. е. зависимость тока I , протекающего через переход, от величины приложенного обратного смещения U_0 .

Первое, что бросается в глаза при взгляде на эту вольтамперную характеристику, — это то, что она не имеет ничего общего с вольтамперной характеристикой обычного активного сопротивления R . Для сопротивления ток пропорционален приложенному напряжению. Вольтамперная характеристика представляет собой прямую линию. Для вольтамперной характеристики обратносмещенного $p - n$ перехода, показанной на рис. 63, характерно *насыщение* тока. При небольшом, порядка десятых долей вольта,

обратном напряжении ток перестает зависеть от напряжения (насыщается). С увеличением температуры ток через переход резко увеличивается.

Чем же объясняется такой вид вольтамперной характеристики обратносмещенного диода?

В обычном кусочке металла или полупроводника ток возникает благодаря направленному движению носителей заряда под действием электрического поля. Чем сильнее поле, тем быстрее направленное движение носителей. Поэтому ток пропорционален напряжению.

В $p-n$ переходе дело обстоит совершенно иначе. Ни электроны в n -области, ни дырки в p -области не способны двигаться туда, куда их гонит электрическое поле. Их движению мешает энергетический барьер на границе p - и n -областей. Даже в отсутствие внешнего напряжения этот барьер препятствовал продвижению электронов из n - в p -область, а дырок — из p - в n -область. Теперь, когда к диоду приложено обратное напряжение, высота барьера увеличилась (рис. 62, *a*) и основные носители — электроны в n -области и дырки в p -области — и подавно не в состоянии преодолеть барьер.

Почему же ток все-таки течет через обратносмещенный переход и что это за ток?

Это наш старый знакомый — ток насыщения I_s ! Ток неосновных носителей: электронов, переходящих из p -области в n , и дырок из n - в p -область (рис. 62).

В самом деле, когда к диоду внешнее напряжение приложено не было, ток насыщения уравнивался встречным потоком горячих основных носителей. Но теперь, когда к диоду приложено обратное смещение и высота барьера увеличилась, эти носители уже не в состоянии преодолеть барьер. Достаточно просто взглянуть на рис. 62, *a*, чтобы немедленно в этом убедиться. Носители, энергия которых превышала высоту барьера, изображенного штриховой линией, преодолеть барьер, изображенный сплошной линией, конечно, не могут.

А вот на ток насыщения, т. е. на поток неосновных носителей, приложенное обратное напряжение никак не влияет. Точно такое же количество электронов, как и раньше, будет генерироваться на расстоянии L_n от перехода и переходить из p - в n -область, образуя электронный компонент тока насыщения I_{ns} . И точно такое же количество дырок будет переходить из n -области в p , образуя дырочный компонент тока насыщения I_{ps} . Встречные компоненты тока подавлены из-за возросшей высоты барьера,

а компоненты тока насыщения не изменились. Вот они и проявились. Протекают через обратносмещенный $p - n$ переход и образуют ток $I_s = I_{ns} + I_{ps}$ (рис. 63), не зависящий от приложенного напряжения.

Понятно, почему достаточно очень небольшого обратного напряжения U_0 , чтобы практически полностью подавить встречные току насыщения потоки горячих основных носителей. Количество горячих основных носителей, способных преодолеть барьер, определяется высотой барьера ϕ и экспоненциально уменьшается с увеличением высоты барьера: $n \sim e^{-\phi/kT}$. Если приложить к барьеру напряжение U_0 , равное всего 0,026 В, то высота барьера увеличится на 0,026 эВ, что равно значению kT при комнатной температуре. Количество электронов и дырок, способных преодолеть барьер, уменьшится в $e \approx 2,7$ раз. Приложенное напряжение $U_0 = 0,1$ В увеличит высоту барьера на $\sim 4 kT$ и, следовательно, уменьшит число носителей, преодолевающих барьер в $e^4 \approx 55$ раз.

Понятно, и почему обратный ток резко возрастает с увеличением температуры: $I_s \sim n_i^2$ (формула (38)).

Из рис. 63 видно, что в области больших напряжений смещения U_0 ток начинает очень резко возрастать с напряжением. (Для данного диода критическое значение напряжения U_i составляет, в зависимости от температуры, 15—17 В.) Этот эффект обусловлен явлением, о котором уже упоминалось, — ударной ионизацией электронов и дырок в $p - n$ переходе.

Вольтамперные характеристики $p - n$ переходов при обратном смещении далеко не всегда похожи на показанные на рис. 63 зависимости $I(U_0)$.

Для кремниевых и арсенид-галлиевых $p - n$ переходов гораздо более характерны зависимости, показанные на рис. 64. Измеренные экспериментально значения обратного тока оказываются гораздо больше (в десятки, сотни, а иногда и миллионы раз), чем величина I_s , рассчитанная по формуле (38). Ток с ростом обратного напряжения U_0 не насыщается, а монотонно возрастает пропорционально $U_0^{1/2}$. О чем свидетельствует такой вид вольтамперных характеристик?

Дело в том, что помимо тока насыщения, который, конечно, существует во всех $p - n$ переходах, и в Ge, и в Si, и в GaAs, и в любом другом, существует еще и другая составляющая обратного тока. Эта другая составляющая обратного тока также существует в $p - n$ переходах, из-

готовленных из любого материала. Только в германиевых переходах она, как правило, мала по сравнению с током насыщения. А в кремниевых и арсенид-галлиевых, наоборот, играет определяющую роль. Полное название этой составляющей звучит довольно длинно: «ток, генерируемый в области объемного заряда». Иногда, для краткости, ее называют просто *генерационным током*.

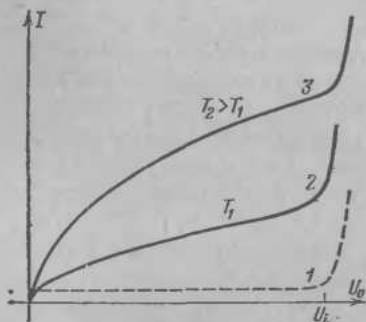


Рис. 64. Качественный вид вольтамперной характеристики, типичный для обратносмещенных кремниевых и арсенид-галлиевых $p-n$ переходов. Штриховая кривая 1 — ход зависимости $I_s(U_0)$, соответствующий току насыщения. Сплошные кривые 2, 3 — реально наблюдаемые вольтамперные характеристики при температурах T_1 и T_2 ($T_2 > T_1$)

До сих пор, обсуждая природу обратного тока, на саму область объемного заряда мы внимания не обращали. Рассматривались только участки диода, отстоящие от области объемного заряда на расстояние L_n (рис. 61). А между тем в области объемного заряда, как и в любой другой области полупроводника, непрерывно происходит генерация электронов и дырок.

Вдали от $p-n$ перехода, там, где электрического поля нет, процесс генерации уравновешивается рекомбинацией. Сколько носителей n^* генерируется в единицу времени, столько и гибнет за счет рекомбинации ($n^* = n/\tau$).

На расстоянии $\sim L_n$ от $p-n$ перехода часть носителей рекомбинирует, а часть попадает в область сильного поля $p-n$ перехода и выбрасывается в соседнюю область, образуя ток насыщения I_s .

А что происходит в самом $p-n$ переходе, в области объемного заряда? Понятно, что любой носитель, электрон или дырка, рожденный в области объемного заряда, будет немедленно подхвачен электрическим полем и выброшен: электроны — в n -область, дырки — в p -область.

Вернемся далеко назад, к с. 47, где мы обсуждали, как создать условия, позволяющие глубоким центрам выступить в роли ступенек. Помните, мы пришли тогда к выводу, что для этого необходимо в какой-либо области полупроводника создать сильное электриче-

ское поле, выносящее из этой области электроны и дырки. В этом случае, как мы установили, любой носитель, появившийся в пределах такой области, будет выноситься из нее полем и давать вклад в ток, протекающий во внешней цепи. Ток этот, естественно, будет тем больше, чем больше свободных носителей генерируется в единицу времени.

А теперь, взглянув на рис. 58, б, или 59, б, или 62, б, мы сейчас же придем к заключению, что область объемного заряда (обедненный слой) $p-n$ перехода отвечает всем условиям, которые предъявлялись к гипотетической области сильного поля. В области объемного заряда очень мало носителей (электронов и дырок), так что рекомбинировать там практически нечему. А генерируются носители в области объемного заряда весьма исправно. Тем энергичней, чем, во-первых, больше концентрация глубоких центров в областях полупроводника p - и n -типа, прилегающих к $p-n$ переходу, и, во-вторых, чем ближе энергетическое положение соответствующих этим центрам глубоких уровней к середине запрещенной зоны. (То есть чем ближе высота энергетической ступеньки, через которую генерируются электроны и дырки, к значению $E_g/2$, — вспомните свойства глубоких-примесей-ступенек!)

Ток, возникающий во внешней цепи вследствие генерации в области объемного заряда, пропорционален ширине области объемного заряда W . Понятно, почему это так: чем больше W , тем шире область, где существует электрическое поле, т. е. тем шире область, где любой порожденный генерацией носитель подхватывается полем и дает вклад в протекающий через диод ток. Поскольку величина W пропорциональна $U_0^{1/2}$ (формула (39)), понятно, что, когда ток, генерируемый в области объемного заряда, вносит основной вклад в обратный ток I_{p-n} перехода, сила тока $I \sim U_0^{1/2}$ (рис. 64, кривые 2, 3).

Как мы знаем, скорость генерации носителей заряда в полупроводнике зависит от концентрации и энергетического положения глубоких уровней, существующих в материале. Такие уровни практически всегда есть в полупроводниках благодаря структурным несовершенствам, дефектам, остаточным примесям, от которых не удается избавиться при очистке. А иногда примеси, создающие глубокие центры, специально вводят в полупроводник, чтобы уменьшить время жизни носителей τ . Скорость генерации, а следовательно, и генерационный ток I_g будут больше всего в том случае, когда глубокий центр, через который идет генерация электронов и дырок, лежит посередине запрещенной зоны (рис. 10).

В этом практически самом важном случае плотность генерационного тока j_g описывается выражением

$$j_g = e \frac{n_i}{\tau} W, \quad (44)$$

Отношение плотности генерационного тока к плотности тока насыщения равняется

$$\frac{j_g}{j_s} = \frac{en_i W}{\tau} \frac{N \sqrt{\tau}}{en_i^2 \sqrt{D}} = \frac{N}{n_i} \frac{W}{\sqrt{D\tau}} = \frac{N}{n_i} \frac{W}{L}. \quad (42)$$

(Для простоты формулу (38) мы написали для случая совершенно симметричного $p - n$ перехода, в котором одинаковы уровни легирования: $N_d = N_a = N$, одинаковы времена жизни электронов и дырок: $\tau_p = \tau_n = \tau$, и даже диффузионные длины электронов и дырок тоже одинаковы: $D_n = D_p = D$, $L_n = L_p = L$.)

Выражение (42) дает возможность сравнить относительные величины j_g и j_s для любого $p - n$ перехода при любой температуре и любом приложенном напряжении. Однако даже просто внимательный взгляд на эту формулу позволяет сразу же установить самый главный признак, характеризующий отношение j_g/j_s . С ростом ширины запрещенной зоны полупроводника \mathcal{E}_g роль генерационного тока чрезвычайно резко, экспоненциально, возрастает. В самом деле, с ростом \mathcal{E}_g чрезвычайно резко уменьшается величина собственной концентрации носителей n_i (см. табл. 1). Вот почему вклад генерационного тока в обратный ток в кремниевых, а в особенности в арсенид-галлиевых $p - n$ переходах несравненно больше, чем в германиевых.

В заключение, как всегда, полезно прочувствовать ситуацию на численном примере. Принимая $N = 10^{14} \text{ см}^{-3}$, $\tau = 10^{-4} \text{ с}$ и пользуясь данными табл. 1 и 2, найдем, что в германиевом $p - n$ переходе с такими параметрами при комнатной температуре генерационный ток I_g при обратном напряжении на диоде $U = 100 \text{ В}$ приблизительно в 6 раз меньше, чем ток насыщения. А вот в кремниевом диоде с аналогичными параметрами, наоборот, генерационный ток в ~ 600 раз больше, чем ток насыщения.

Фотодиоды. В мифах и сказках любого народа живут существа, обладающие способностью видеть в темноте, никогда не спящие, способные за семь верст разглядеть комара на носу. Стоглазый Аргус, приставленный сторожить несчастную возлюбленную Зевса, зоркий охотник — спутник храброго портняжки... и т. п.

В нашем веке созданы сотни устройств, выполняющих не менее трудные задачи. Круглые сутки в течение многих лет несут охрану объектов фотоохранные устройства. Кто бы ни пересек световой луч, опоясывающий по перимет-

ру охраняемую территорию, устройство мгновенно среагирует и подаст сигнал тревоги.

Сотни раз за смену рабочий берет в руки лист металла, подставляет его под штамп, ногой нажимает педаль. Многотонный пресс обрушивается на заготовку. Готовая деталь снимается. Под пресс подставляется новая заготовка... В очередной, сто первый раз — человек не автомат — рабочий замешкался и нажал на педаль, не убрав руку из опасной зоны. Ничего не произошло. Пресс не шелохнулся. Если рука находится в опасной зоне, она пересекает луч света и фотоэлектрическое защитное устройство блокирует механизм, приводящий пресс в движение.

Фотоэлектрические устройства защищают здания от пожаров, обнаруживают наличие в воздухе вредных газов и паров, анализируют прозрачность крови человека, контролируют состояние мостов и плотин, управляют химическими и механическими процессами, контролируют и поддерживают заданный уровень жидкости, определяют параметры частиц в ядерной физике, подсчитывают детали, сортируя их по форме, цвету, массе, размеру, контролируют температуру и выполняют сотни других операций со скоростью, точностью, чувствительностью и неутомимостью, оставляющими далеко позади возможности человеческого глаза.

В основе работы любого фотоэлектрического устройства лежит элемент, электрические параметры которого: сопротивление, ток или напряжение, меняются под действием освещения. Одним из самых распространенных элементов такого рода является фотодиод.

А фотодиод — это... $p - n$ переход, смещенный в обратном направлении.

Как мы хорошо знаем, если на полупроводник падает свет с энергией фотона \mathcal{E}_f , большей, чем ширина запрещенной зоны \mathcal{E}_g , то каждый поглощенный фотон порождает в полупроводнике электронно-дырочную пару (см., например, рис. 7, а). Если электрон и дырка возникли под действием света в области объемного заряда обратнсмещенного $p - n$ перехода, то... мы можем вернуться к с. 168 и дословно переписать текст: «...любой носитель, электрон или дырка, рожденный в области объемного заряда, будет немедленно подхвачен электрическим полем и выброшен: электрон — в n -область, дырка — в p -область».

В темноте, без освещения, генерационный ток $p - n$ перехода (так называемый темновой ток) определяется числом электронов и дырок, возникающих в области объем-

ного заряда за счет тепловой генерации. При освещении обратносмещенного перехода светом с энергией квантов $\mathcal{E}_\phi > \mathcal{E}_g$ ток через переход увеличивается. Возникший под действием света фототок во столько раз больше темнового, во сколько число носителей, создаваемых в области объемного заряда светом, будет больше, чем число электронов и дырок, возникающих за счет тепловой генерации. Фототок, усиленный и преобразованный соответствующей электронной схемой, приводит в действие защитные, охранные, сигнальные, блокировочные и тому подобные устройства.

На рис. 65, а показана типичная конструкция фотодиода. Видно, что это наш старый знакомый: $p-n$ переход, смещенный в обратном направлении. Верхний контакт к диоду выполнен в форме кольца (рис. 65, б).

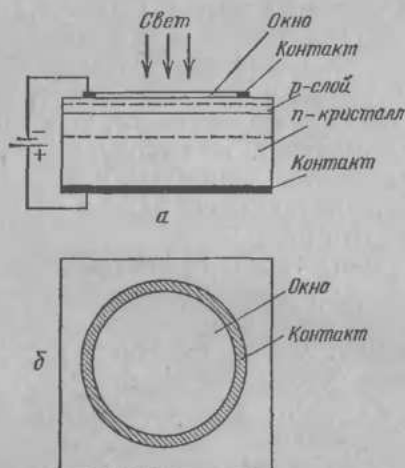


Рис. 65. Конструкция типичного фотодиода

Область $p-n$ перехода внутри кольца, так называемое окно, доступна для света. В фотодиоде плоскость $p-n$ перехода располагается близко к поверхности кристалла; это необходимо для того, чтобы падающий на кристалл свет не успел поглотиться на пути к области объемного заряда. Обычно плоскость $p-n$ перехода располагается в фотодиодах на расстоянии в несколько десятых долей микрометра от поверхности. Такой $p-n$ переход можно изготовить с помощью диффузии или ионной имплантации.

Величина обратного напряжения U_0 , прикладываемого к $p-n$ переходу, составляет обычно 10—30 В. С увеличением обратного напряжения, как мы знаем, область объемного заряда расширяется (рис. 62, формула (39)). Величина U_0 и концентрация примеси в слаболегированной n -области кристалла N_d выбираются такими, чтобы практически весь световой поток, падающий на поверхность

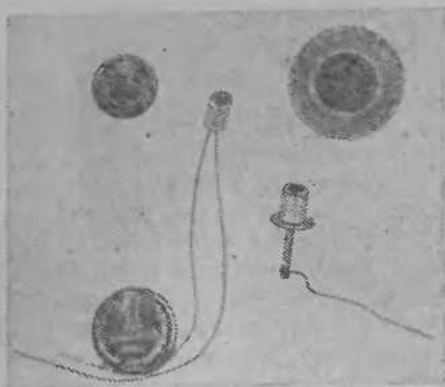


Рис. 66. Некоторые серийно выпускаемые фотодиоды

светодиода, поглотился в области объемного заряда и дал вклад в фототок. Обычно необходимая для этого толщина обедненной области W составляет несколько микрометров.

Промышленностью в настоящее время выпускаются десятки типов фотодиодов. Некоторые из них показаны на рис. 66.

Наибольшее распространение получили фотодиоды на основе Ge и Si. Однако разработаны и используются также фотодиоды из GaAs, InAs, InSb, GaP, тройных полупроводниковых соединений GaInAs, GaAlAs, четверного полупроводника GaInAsP.

Одна из самых важных причин для разработки фотодиодов на основе разных полупроводниковых соединений состоит в том, что фотодиод, изготовленный из определенного материала, способен эффективно реагировать

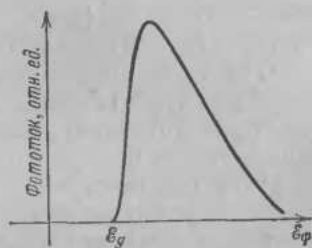


Рис. 67. Качественная зависимость фототока от энергии фотонов падающего на фотодиод света

на свет с длиной волны, лежащей только в определенном диапазоне. Это свойство фотодиодов называют *селективностью* — от латинского *selectio* (выбор, отбор), оно поясняется рис. 67.

На рис. 67 для фотодиода, изготовленного из полупроводника с шириной запрещенной зоны \mathcal{E}_g , показана качественная зависимость фототока от энергии падающих фотонов.

При $\mathcal{E}_\phi < \mathcal{E}_g$ энергия фотонов недостаточна для создания в материале электронно-дырочных пар и фототок поэтому не возникает. При $\mathcal{E}_\phi > \mathcal{E}_g$ фототок сначала быстро увеличивается с ростом \mathcal{E}_ϕ , а затем... начинает уменьшаться. Дело в том, что с увеличением \mathcal{E}_ϕ резко растет коэффициент поглощения света в материале, т. е. падающий световой поток с ростом энергии фотонов поглощается все ближе и ближе к поверхности фотоприемного окна (рис. 65). Как мы знаем (см. гл. 3), на поверхности всегда присутствует большое число дефектов и примесей, создающих эффективные центры рекомбинации. Поэтому электроны и дырки, возникшие под действием света вблизи поверхности, быстро погибают и не дают вклад в фототок.

Таким образом, область максимальной фоточувствительности диода определяется величиной \mathcal{E}_g и соответствует условию $\mathcal{E}_\phi \gtrsim \mathcal{E}_g$. Полупроводники с небольшой шириной запрещенной зоны, так называемые *узкозонные полупроводники*, используются для создания фотодиодов, чувствительных к свету с малой энергией квантов, т. е. с большой длиной волны. Полупроводники с большими значениями \mathcal{E}_g (*широкозонные полупроводники*) используются для создания приемников ультрафиолетового света.

Очень важным достоинством фотодиодов является их высокая скорость срабатывания (быстродействие). Быстродействие фотозлектрических приборов часто оценивают по тому, как быстро спадает фототок через прибор после выключения света.

Если резко «оборвать» световой поток, падающий на приемное окно фотодиода, то фототок прекратится, как только возбужденные светом электроны и дырки будут вынесены полем из области объемного заряда.

Мы уже знаем, что типичные значения напряженности электрического поля в обратносмещенном $p-n$ переходе велики. Они составляют $\sim 10^4 - 10^6$ В/см. В таких сильных полях дрейфовая скорость электронов и дырок v не зависит от поля и составляет $v_s \sim 10^7$ см/с (см. рис. 19).

Путь, который предстоит пройти электронам и дыркам, прежде чем они окажутся вынесенными полем из $p - n$ перехода, равен, очевидно, ширине области объемного заряда W . Как уже говорилось, значение W для типичных фотодиодов составляет несколько микрометров $((1 - 10) \cdot 10^{-4}$ см).

Таким образом, время, определяющее быстродействие фотодиода t_0 , составляет приблизительно

$$t_0 \sim \frac{W}{v_s} \sim 10^{-10} - 10^{-11} \text{ с.}$$

Фотодиод способен реагировать на изменение интенсивности света за стомиллиардные доли секунды!

Высокое быстродействие позволяет использовать фотодиоды в самых скоростных системах передачи и обработки информации. Мы уже упоминали о системах волоконно-оптической связи. Приемниками быстро модулируемых световых сигналов в таких системах служат именно фотодиоды.

С помощью фотодиодов производится ввод информации в ЭВМ (рис. 68). Перфолента, отрезок которой показан на рис. 68, б, протягивается между осветителем и группой фотодиодов. Фотодиод, находящийся перед отверстием в ленте, выдает импульс тока. Там, где отверстия в перфоленте нет, нет и импульса фототока от соответствующего диода. Совокупность электрических импульсов от группы фотодиодов отображает информацию, «записанную» на перфоленте. Фотоэлектрические устройства позволяют вводить в ЭВМ тысячи чисел и команд в секунду.

Сегодня ученые работают над созданием многоэлементных фотоприемников, обладающих «цветным» зрением. Устройство таких приборов похоже на строение сетчатки глаза человека. Каждый из десятков или сотен тысяч элементов, образующих «искусственный глаз», выдает электрические сигналы в соответствии с тем, какой интенсивности и длины волны свет на него попадает. Такие приборы позволяют вводить в ЭВМ сложнейшую информацию без специального кодирования; фактически они смогут непосредственно сообщать машине о том, что они «видят».

Барьерная емкость. Продолжим изучение замечательных свойств $p - n$ перехода. В предыдущих разделах мы довольно много узнали о протекании постоянного тока через диод, к которому приложено обратное смещение. Обсудим теперь вопрос о том, чем будет определяться сопротивление $p - n$ перехода, к которому помимо обратного

го напряжения U_0 приложено еще и небольшое переменное напряжение.

Ответ на этот вопрос прозвучит, возможно, несколько неожиданно. По отношению к переменному напряжению

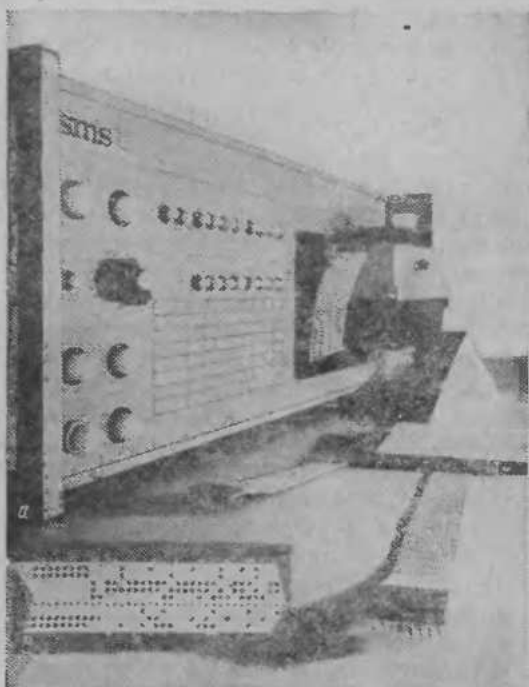


Рис. 68. *a* — Фотография фотосчитывающего устройства ЭВМ. *б* — С помощью такой перфоленты кодируется и при необходимости вводится в ЭВМ вся информация, необходимая для работы машины

обратносмещенный $p-n$ переход будет вести себя, как... конденсатор.

Емкостное сопротивление $p-n$ перехода, как и любого другого конденсатора (Физика, 9 класс), определяется известным выражением $X_C = 1/\omega C$, где $\omega = 2\pi f$ — круговая частота, C — емкость $p-n$ перехода.

Откуда же берется емкость обратносмещенного перехода, какова ее природа и чему она равна?

Легче всего ответить на последний вопрос. Емкость $p-n$ перехода следует определять по формуле емкости

для плоского конденсатора

$$C = \frac{\epsilon\epsilon_0 S}{W}, \quad (43)$$

где W — ширина области объемного заряда, S — площадь перехода.

Так как ширину области объемного заряда определять мы умеем (формула (39)), то, следовательно, умеем определять и емкость перехода при любом приложенном напряжении и любых уровнях легирования.

Ответить на вопрос о природе емкости перехода несколько сложнее, но тоже можно. Вспомним (Физика, 9 класс), почему обыкновенный конденсатор — пластина диэлектрика, снабженная металлическими обкладками и для постоянного тока представляющая бесконечно большое сопротивление, — способен проводить переменный ток. Причина состоит в том, что, когда к конденсатору приложено переменное напряжение, он в определенную часть периода накапливает электрический заряд, а затем, в другую часть периода, отдает его во внешнюю цепь. Периодический заряд и разряд конденсатора поддерживает в цепи переменный ток.

А теперь рассмотрим обратносмещенный $p-n$ переход, к которому наряду с постоянным обратным напряжением U_0 приложено еще и небольшое переменное напряжение с амплитудой ΔU_i (рис. 69). Сплошные кривые на рис. 69 соответствуют распределению объемного заряда ρ в $p-n$ переходе при обратном напряжении U_0 (ср. с рис. 57). Штриховой кривой показано распределение объемного заряда ρ , когда к диоду приложено напряжение $U_0 + \Delta U_0$. Напряжение на диоде стало больше — область объемного заряда расширилась. Свободные носители — электроны в n -области и дырки в p -области — из освобожденной от носителей области объемного заряда вытолкнулись во внешнюю цепь.

В n -области при увеличении напряжения от U_0 до $U_0 + \Delta U_0$ во внешнюю цепь вытолкнулись все электроны, находившиеся при напряжении U_0 на участке между сплошной и штриховой кривыми. В p -области точно так же из участка между сплошной и штриховой кривыми оказались вытесненными дырки.

Точечной кривой показано распределение объемного заряда, когда к диоду приложено напряжение $U_0 - \Delta U_0$. Напряжение на диоде стало меньше — ширина области объемного заряда уменьшилась. А это значит, что из внеш-

пей цепи «затребованы» свободные носители, которые заполнили теперь участки диода, где их не было при напряжении U_0 . Если напряжение уменьшилось от U_0 до $U_0 - \Delta U_0$, то свободные носители, затребованные из внешней цепи, заполнили области между сплошной и точечной кривыми.

Таким образом, ясно, что $p-n$ переход обладает способностью в определенную часть периода накапливать

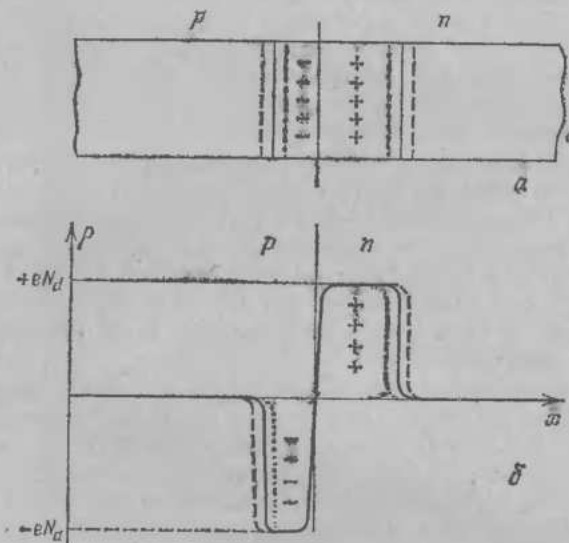


Рис. 69. Распределение объемного заряда в $p-n$ переходе (ср. с рис. 57): сплошная кривая — к диоду приложено обратное напряжение U_0 ; штриховая — обратное напряжение равняется $U_0 + \Delta U_0$; точечная — $U = U_0 - \Delta U_0$

свободные носители заряда, а в другую — отдавать их во внешнюю цепь, т. е. обладает емкостью. Этот тип емкости $p-n$ перехода носит название *барьерной емкости*.

Найдем теперь величину барьерной емкости. Как обычно, величина емкости C определится как отношение заряда Δq , накопленного (или отданного) при изменении напряжения на величину ΔU_0 : $C = \Delta q / \Delta U_0$.

Ограничимся, как и раньше, рассмотрением резко несимметричного $p-n$ перехода, в котором $N_d \ll N_a$. Тогда, как мы знаем, практически все приложенное напряжение падает на слаболегированной области, а полная ши-

рина объемного заряда W практически равняется W_n — ширине обедненной области в n -слое (рис. 59, б).

При увеличении напряжения на величину ΔU_0 ширина области объемного заряда увеличивается на величину ΔW_n . Значение ΔW_n можно определить из формулы (39). При этом во внешнюю цепь будет вытеснен на единицу площади заряд электронов $\Delta q = eN_d \cdot \Delta W_n$. (И точно такой же заряд дырок будет вытеснен из p -области.) Отношение $\Delta q / \Delta U_0$ определит величину емкости $p - n$ перехода на единицу площади C_0 .

Дальше наша дорога разбивается на две. Одна, прямая, — для тех, кто умеет дифференцировать. Другая, в объезд, — для тех, кто чувствует себя в этой области неуверенно. Первая, естественно, короче:

$$C_0 = \frac{dq_0}{dU_0} = eN_d \frac{dW_n}{dU_0} = eN_d \cdot \frac{1}{2} \left(\frac{2e\varepsilon_0}{eN_d} \right)^{1/2} \frac{1}{(U_{pn} + U_0)^{1/2}} = \left[\frac{e\varepsilon_0 N_d}{2(U_{pn} + U_0)} \right]^{1/2}. \quad (44)$$

Производная dW_n/dU_0 получена дифференцированием формулы (39). Дальше, проделав несложные преобразования, получим

$$C_0 = \left[\frac{e\varepsilon_0 N_d}{2(U_{pn} + U_0)} \right]^{1/2} = \frac{\varepsilon_0}{W} z, \quad (45)$$

где U_0 — напряжение на $p - n$ переходе, W — ширина области объемного заряда, определяемая формулой (39).

Второй путь заметно длиннее, но ведет, естественно, к тому же самому результату. Нужно сначала по формуле (39) вычислить ширину объемного заряда W_2 при напряжении на $p - n$ переходе $U_0 + \Delta U_0$. Затем вычесть из нее значение W_1 , соответствующее ширине области объемного заряда при напряжении на переходе U_0 . Так определится величина $\Delta W = W_2 - W_1$. Затем, поделив заряд $\Delta q = eN_d \Delta W$ на величину ΔU_0 , получим значение C_0 . Если все сделано правильно, в итоге получится то же самое выражение (45).

Из формулы (44) следуют два важных вывода.

Во-первых, барьерная емкость $p - n$ перехода зависит от напряжения. Чем больше приложенное к диоду обратное смещение U_0 , тем меньше емкость. Конденсаторы, емкость которых зависит от напряжения, называют *нелинейными*. Поэтому часто говорят, что $p - n$ переход представляет собой нелинейную емкость.

Во-вторых, при равном обратном напряжении на переходе барьерная емкость тем больше, чем выше уровень легирования: $C \sim N_d^{1/2}$.

Посмотрим теперь о каких, по порядку величины, значениях емкости идет речь.

Мощный высоковольтный выпрямительный диод, о котором уже несколько раз упоминалось в этой главе, может иметь площадь $S \sim 10 - 30 \text{ см}^2$. При $S = 20 \text{ см}^2$, $N_d = 10^{13} \text{ см}^{-3}$ и отсутствии внешнего обратного напряжения на диоде емкость диода C составит

$$C \approx S \left(\frac{\epsilon \epsilon_0 N_d}{2U_{pn}} \right)^{1/2} \approx 2 \cdot 10^{-8} \text{ Ф} = 20 \text{ 000 пФ!}$$

Даже при напряжении на структуре $U_0 = 1000 \text{ В}$ емкость диода составит весьма заметную величину: $C \approx 600 \text{ пФ}$. Интересно сравнить емкостное сопротивление диода и его активное сопротивление, определяемое обратным током. При комнатной температуре и напряжении на диоде, скажем, $U_0 = 100 \text{ В}$ типичная величина обратного тока I для кремниевого диода такого типа составляет $\sim 0,5 \text{ мА}$. И, следовательно, активное обратное сопротивление $R = U_0/I \approx 2 \cdot 10^5 \text{ Ом}$. Емкость C при $U = 100 \text{ В}$ равняется $\sim 2 \cdot 10^{-9} \text{ Ф}$. Сравнивая $X_C = 1/\omega C$ и R , найдем, что при $f \geq 400 \text{ Гц}$ сопротивление диода будет определяться не обратным током, а емкостью структуры. Правда, такие мощные диоды предназначены для работы в цепях переменного тока промышленной частоты $f = 50 \text{ Гц}$, и для них наличие даже такой большой емкости особой роли не играет.

Существуют, однако, диоды, предназначенные для работы в цепях с частотой сотни тысяч, миллионы и миллиарды герц. Для таких диодов наличие барьерной емкости играет очень важную, а часто и решающую роль. Иногда емкостный ток, обусловленный барьерной емкостью, мешает диоду выполнять ту работу, для которой он предназначен. В этом случае, естественно, величину C стараются сделать по возможности меньшей. Самый прямой путь к такой цели — уменьшение рабочей площади диодов. Вот почему площадь диодов, предназначенных для работы в цепях сверхвысокой частоты (СВЧ), составляет $10^{-4} - 10^{-6} \text{ см}^2$. Величина барьерной емкости таких $p - n$ структур равняется десятым, а иногда и сотым долям пикофарады.

Существуют, однако, и такие диоды, действие которых прямо основано на свойствах барьерной емкости перехо-

да. Они используются в качестве управляемой напряжением переменной емкости в цепях высокой и сверхвысокой частот.

Варикапы. Название прибора (от английского variable capacitor — переменный конденсатор) очень точно характеризует его назначение. Варикапы используются в качестве конденсаторов переменной емкости, величина которой изменяется в зависимости от приложенного напряжения. Чем больше напряжение, тем меньше емкость.

С конденсаторами переменной емкости имеют дело практически все. Каждый, кто настраивает радиоприемник на нужную станцию, с точки зрения радиотехники занят тем, что, изменяя емкость переменного конденсатора, настраивает входной контур радиоприемника на нужную частоту.

Одно из первых применений варикапов как раз и состояло в замене механически перестраиваемого конденсатора в радиоприемнике. При этом можно было настраиваться на нужную станцию, не вставая с кресла (рис. 70). Сидящий в кресле человек держит в руках небольшую коробочку. Рисунок внизу показывает, что находится в ко-

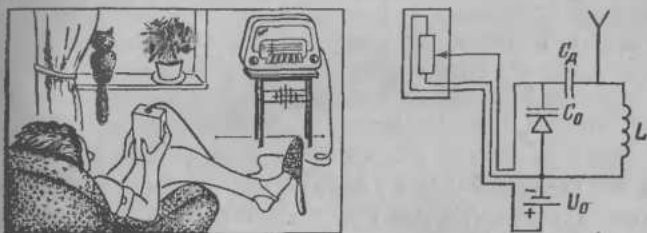


Рис. 70. Замена механически перестраиваемого конденсатора варикапом дает возможность очень просто осуществить дистанционное управление настройкой приемника на нужную станцию. Обратите внимание на условное обозначение варикапа C_0 . Оно содержит общее для всех полупроводниковых диодов обозначение, дополненное символом емкости. Вершине треугольника, обозначающего диод, всегда соответствует n -область p — n перехода, основанию — p -область. Поэтому диод смещен в обратном направлении, когда «плюс» источника подан на вершину треугольника. На рисунке именно такое смещение и подается на варикап. Обратите также внимание на дополнительный конденсатор C_d . Если бы его не было, напряжение, снимаемое с потенциометра, закорачивалось бы индуктивной остью L . Чтобы этот конденсатор не влиял на настройку контура, его емкость должна быть гораздо больше, чем емкость варикапа ($C_d \gg C_0$) — вспомните выражение, описывающее емкость последовательно соединенных конденсаторов (Физика, 9 класс)

робочке, — переменное сопротивление. Из приемника по гибкому шнуру, соединяющему приемник и коробочку, поступает напряжение U_0 . При вращении движка переменного сопротивления изменяется напряжение, подаваемое на находящийся в приемнике варикап. Оно может изменяться от нуля (когда движок потенциометра находится в крайнем нижнем положении) до величины U_0 . Емкость варикапа изменяется — изменяется и резонансная частота контура.

Чтобы схема дистанционной настройки, показанная на рис. 70, была *практически работоспособной*, необходимо, как видно из рисунка, ввести в нее дополнительный конденсатор C_d . Если этого не сделать, варикап окажется замкнутым ничтожным сопротивлением индуктивности L и подать на него постоянное напряжение будет невозможно. Искусство создавать практически работающие устройства, и не только в радиотехнике, в значительной мере зависит от способности предусматривать такие на первый взгляд «непринципиальные» детали.

Для постоянного напряжения батарейки конденсатор C_d представляет собой громадное, практически бесконечно большое сопротивление. Для рабочей частоты, на которую настраивается контур, его сопротивление (емкостное) равняется $X_C = 1/\omega C_d$. Если $C_d \gg C_0$ (C_0 — емкость варикапа), то величина X_C ничтожна по сравнению с емкостным сопротивлением варикапа $X_0 = 1/\omega C_0$ ($X_C \ll X_0$). Поэтому, как говорят, *по переменному току* наличие конденсатора C_d практически не влияет на настройку контура. На практике величину C_d выбирают обычно в сотни раз большей, чем C_0 .

Если бы применение варикапов ограничивалось устройствами дистанционной настройки бытовых радиоприемников, об этом приборе, быть может, не стоило бы и упоминать. В конце концов необходимость вылезти из кресла — не самое суровое испытание, посылаемое судьбой человеку.

В действительности без варикапов невозможна была бы работа телевизоров, в особенности цветных. Полупроводниковые емкости, управляемые напряжением, лежат в основе сверхчувствительных преобразователей постоянного напряжения в переменное. С их помощью осуществляется прием и усиление сверхслабых сигналов, в том числе и сигналов из космоса. Они используются для генерации сигналов с частотой в сотни миллиардов герц.

В основе всех этих применений переменной полупроводниковой емкости лежит то обстоятельство, что вели-

чина этой емкости может быть изменена с помощью приложенного напряжения чрезвычайно быстро.

Чтобы изменить емкость механического конденсатора, нужно повернуть на определенный угол пластины. На это требуются секунды, в лучшем случае — десятые доли секунды.

Емкость варикапа, в зависимости от его конструкции, меняется при изменении приложенного напряжения за тысячные, миллионные, а иногда и миллиардные доли секунды.

Обсудим еще два примера использования варикапа.

Первый из них интересен тем, что услугами схемы, принцип работы которой будет рассмотрен, все мы ежедневно пользуемся несколько сотен раз в день, чаще всего даже не догадываясь об этом. Речь пойдет о схеме автоматической подстройки частоты (АПЧ), использующейся в телевизорах и в хороших радиоприемниках.

Чтобы изображение на экране телевизора было четким и ясным, настройка его контуров на частоту передающей станции должна быть очень точной. Это требование особенно важно для цветных телевизоров, в которых совсем незначительный уход частоты настройки от точного значения вызывает заметное искажение цветного изображения или даже превращение его в черно-белое. Для того чтобы этого не происходило, допускается уход частоты настройки не более чем на 500 кГц. Если учесть, что телепередачи 8-го канала ведутся на частоте $f \approx 200$ МГц, то точность настройки должна быть не хуже $\approx 0,3\%$. А между тем частота настройки зависит от температуры, от напряжения в сети, от влажности воздуха в комнате...

Представьте себе, что телевизор включен в три часа и не выключается до одиннадцати вечера. (Увы, большой фантазии для того, чтобы представить себе такую ситуацию, не требуется.) В четыре часа дня начинают выключаться мощные промышленные потребители электроэнергии. Электрическая сеть, в которую включен телевизор, реагирует на каждое выключение небольшим, но для настройки вполне ощутимым скачком напряжения. Телевизор в процессе работы разогревается — меняется температура корпуса и контуров. Вы открываете в комнате окно — изменяется влажность воздуха. Включение утюга в соседней комнате, сотрясение пола (у соседней танцы), легкий сквозняк... Не будь в телевизоре АПЧ, на каждое такое воздействие телевизор реагировал бы появлением полос, исчезновением изображения, искажением цвета.

Схема АПЧ автоматически поддерживает точную настройку на заданную станцию при любом изменении внешних условий.

Практически пригодная схема АПЧ довольно сложна. Она содержит много «маленьких хитростей» (вроде дополнительного конденсатора C_d на рис. 70). Однако принцип ее работы прост (рис. 71). Схема, называемая *частотным дискриминатором*, непрерывно сравнивает частоты двух сигналов — входного сигнала

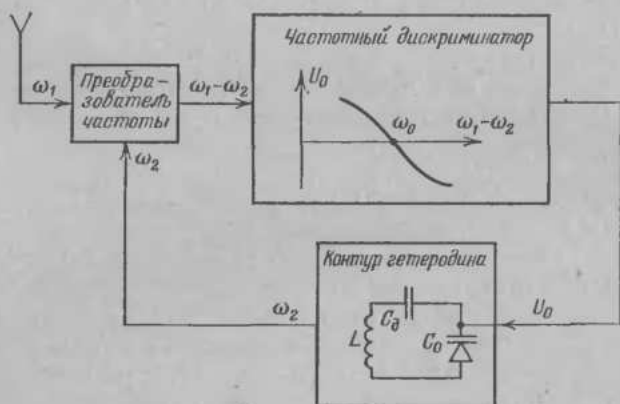


Рис. 71. Принцип работы АПЧ

ω_1 , поступающего на вход телевизора, и второго сигнала ω_2 , вырабатываемого в самом приемнике специальным генератором — гетеродином. Точнее говоря, эта схема сравнивает разность частот $\omega_1 - \omega_2$ с эталонной частотой ω_0 .

Сигнал разностной частоты $\omega_1 - \omega_2$ вырабатывает схема преобразователя частоты, на которую поступают два сигнала. Один — с частотой ω_1 — с приемной антенны, другой — с частотой ω_2 — с выхода гетеродина. Резонансный контур гетеродина содержит варикап.

Сигнал с частотой $\omega_1 - \omega_2$ поступает на вход частотного дискриминатора. Схема частотного дискриминатора вырабатывает напряжение U_0 , величина и знак которого зависят от разности $\omega_1 - \omega_2$. Если $\omega_1 - \omega_2 = \omega_0$, величина U_0 равна нулю.

Если $\omega_1 - \omega_2 > \omega_0$, то это означает, что частота контура ω_2 слишком мала. Ее нужно увеличить. Для этого следует уменьшить емкость варикапа C_0 . Из рис. 71 видно, как справляется с этой задачей схема частотного дискриминатора. При $\omega_1 - \omega_2 > \omega_0$ на выходе дискриминатора вырабатывается отрицательное напряжение. Оно складывается с отрицательным (обратным) напряжением, постоянно подаваемым на варикап. (Этот источник постоянного напряжения на рис. 71 не показан, чтобы не загромождать схему.) Отрицательное напряжение возрастает — уменьшается емкость варикапа C_0 . Частота контура $\omega_2 = 1/\sqrt{LC_0}$ возрастает,

Если $\omega_1 - \omega_2 < \omega_0$, то частоту контура ω_2 следует уменьшить. Схема дискриминатора при $\omega_1 - \omega_2 < \omega_0$ вырабатывает, как видно из рис. 71, положительное напряжение U_0 . Оно вычитается из отрицательного напряжения, постоянно поданного на варикап. Отрицательное сопротивление уменьшается — величина C_0 возрастает — частота ω_2 уменьшается.

Схема АПЧ, изменяя емкость варикапа, автоматически поддерживает разность $\omega_1 - \omega_2$ постоянной и равной эталонной частоте ω_0 , что и необходимо для нормальной работы телевизора.

В качестве второго примера обсудим принцип работы параметрических усилителей сверхвысокочастотного диапазона.

Принцип параметрического усиления был известен задолго до появления варикапов. В основе этого принципа лежит способность конденсатора изменять запасенную в нем энергию при изменении емкости.

Представим себе конденсатор, расстояние между обкладками которого можно изменять. Пусть начальная емкость конденсатора равняется C_1 . Зарядим его до напряжения U_1 . При этом на обкладках конденсатора сосредоточится заряд $q = C_1 U_1$. В конденсаторе окажется запасенной электростатическая энергия $W_1 = C_1 U_1^2 / 2 = q^2 / 2C_1$. Отсоединим теперь конденсатор от источника напряжения и, раздвинув обкладки, увеличим расстояние между пластинами. Заряд q на пластинах остался прежним — ему просто некуда деваться. Емкость конденсатора уменьшилась (обозначим новое значение емкости C_2). Значит..? Значит, возросло напряжение на обкладках конденсатора. Оно стало равным $U_2 = q / C_2$. И увеличилась энергия, запасенная в конденсаторе: $W_2 = C_2 U_2^2 / 2 = q^2 / 2C_2$ *).

Таким образом, периодически изменяя с помощью внешней силы параметр конденсатора — емкость, можно вносить в цепь, содержащую такой переменный конденсатор, энергию.

Как и в любом усилителе, чтобы энергия, вводимая в цепь, пошла впрок и способствовала усилению сигнала, нужно вводить ее в контур в строго определенные промежутки времени и в определенной фазе с усиливаемым сигналом. Хорошо известный каждому пример — качели. Если присесть, когда качели достигают

*) Естественно, возникает вопрос, откуда эта добавочная энергия взялась? Попробуйте ответить на него сами.

О т в е т: чтобы развести пластины конденсатора, нужно преодолеть действующую между ними силу притяжения. Работа по преодолению силы притяжения на пути, равном разности между новыми и старым расстоянием между пластинами, и равняется $W_2 - W_1$. Убедитесь в этом, рассчитав величину работы.

верхнего положения, и распрямляться, когда они проходят через нижнюю точку, т. е. дважды за период изменять параметр, контролирующей частоту колебаний — расстояние от точки подвеса до центра тяжести, — качели можно раскачать до очень большой амплитуды *).

В параметрическом усилителе в LC -контур, настроенный на частоту усиливаемого сигнала, в качестве емкости C включается варикап. Параметр варикапа — емкость дважды за период колебаний усиливаемого сигнала в нужной фазе изменяется внешним источником колебаний. В результате достигается усиление, называемое *параметрическим*. Основное достоинство параметрических усилителей — их рекордно высокая чувствительность (рис. 72).

Параметрические усилители используются для усиления сигналов с частотой в десятки миллиардов герц. Так как частота внешнего источника сигнала, изменяющего емкость «полупроводникового конденсатора», должна быть еще в два раза выше, чем частота усиливаемого сигнала, «полупроводниковый конденсатор» должен изменять емкость за промежутки времен, измеряемые сотыми долями миллиардных долей секунды (10^{-10} — 10^{-11} с). Такие диоды по конструкции заметно отличаются от телевизионных варикапов, и иногда их выделяют в отдельный класс полупроводниковых приборов — *параметрические диоды*. Однако по физическому принципу работы параметрические диоды ничем от варикапов не отличаются.

Пожалуй, самым главным параметром варикапа является коэффициент изменения емкости K_C , определяемый как отношение максимальной емкости варикапа C_{\max} к минимальной C_{\min} : $K_C = C_{\max}/C_{\min}$.

Так как емкость варикапа зависит от приложенного к нему напряжения, то понятно, что величина K_C зависит от диапазона, в котором можно менять отрицательное смещение на варикапе.

Основное условие, ограничивающее этот диапазон, формулируется очень просто: в диапазоне рабочих напряжений варикапа через него не должен протекать большой постоянный ток. Из этого условия и вытекают ограничения на рабочий диапазон напряжений варикапа.

Минимально допустимое обратное напряжение на варикапе U_0 , как правило, допускается равным нулю. При этом напряжение на $p-n$ переходе минимально

*) Проанализируйте, каким образом вносится при этом энергия.



Рис. 72. Антенна радиотелескопа. Сигналы, принятые из дальних областей Вселенной даже такими громадными антеннами, вероятно слабы. Чтобы усилить их, нужны сверхчувствительные усилители. В качестве таких усилителей часто используются параметрические усилители, основным элементом которых является полупроводниковая емкость, периодически изменяемая внешним напряжением

и равняется U_{pn} , а емкость варикапа максимальна. В соответствии с (44) она равняется

$$C_{\max} = S \left(\frac{\epsilon \epsilon_0 N_d}{2U_{pn}} \right)^{1/2} \quad (46)$$

Максимальное напряжение на $p - n$ переходе ограничивается процессом ударной ионизации и описыва-

ется формулой (40). Считая, что $U_i \gg U_{pn}$, из (44) получаем значение минимальной емкости варикапа

$$C_{\min} = S \left(\frac{\epsilon \epsilon_0 N_d}{2U_i} \right)^{1/2} = S \frac{eN_d}{E_i}. \quad (47)$$

Коэффициент изменения емкости K_C , таким образом, будет равен

$$K_C = \frac{C_{\max}}{C_{\min}} = \left(\frac{\epsilon \epsilon_0 E_i^2}{2eU_{pn}N_d} \right)^{1/2}. \quad (48)$$

Видно, что чем меньше концентрация примеси в слаболегированной области, тем больше величина K_C . На практике значение N_d выбирают равным 10^{15} — 10^{16} см⁻³. Величина K_C в реальных варикапах лежит в пределах от ~ 2 до 15.

Ударная ионизация. О явлении ударной ионизации мы уже несколько раз упоминали. Это явление ограничивает величину максимального обратного напряжения, которое может быть приложено к $p-n$ переходу (формула (40)). Однако роль ударной ионизации этим не ограничивается. На ее основе созданы различные полупроводниковые приборы. Познакомимся с явлением ударной ионизации несколько подробнее.

Энергия носителей тока, электронов и дырок, возрастает с ростом напряженности электрического поля E . В очень сильном поле она может стать настолько большой, что носители приобретут способность при столкновении с кристаллической решеткой вырывать из системы межатомных связей (рис. 2) валентные электроны. При каждом таком столкновении, или, как принято говорить, *акте ударной ионизации*, в кристалле образуются электрон и дырка — электронно-дырочная пара.

Обсудим сначала, как всегда, самую простую ситуацию. Пусть в какой-то области полупроводникового кристалла длиной L существует однородное, не зависящее от координаты поле $E \gtrsim E_i$ (рис. 73, а). Условие $E \gtrsim E_i$ означает, что в рассматриваемом поле E носители заряда при столкновении с решеткой полупроводника способны порождать электронно-дырочные пары.

Для начала еще больше упростим ситуацию. Представим себе, что такой способностью — порождать при столкновении с решеткой электроны и дырки — обладает только какой-то один тип носителей, например электроны. А дырки — не обладают,

В рассматриваемую область попадает электрон (рис. 73, б). В сильном электрическом поле он движется слева направо. Рис. 73, б — е по кадрам демонстрируют «кинофильм», иллюстрирующий, что при этом происходит. Пролетев какое-то расстояние, электрон порождает электронно-дырочную пару. Возникают еще один электрон проводимости и дырка (рис. 73, в). Теперь уже слева направо движутся два электрона. Положительно заряженная дырка тоже будет двигаться, но в противоположном направлении — справа налево.

Пролетев приблизительно такое же расстояние, два электрона в актах ударной ионизации породят еще две электронно-дырочные пары (рис. 73, г). Дальнейшее движение слева направо продолжают теперь уже четыре электрона. Рожденные в актах ионизации дырки полетят справа налево. В следующем акте родятся четыре электронно-дырочные пары (рис. 73, д) и т. д.

Один-единственный носитель, попавший в область сильного поля, где возможна ударная ионизация, порождает при движении в этой области нарастающую «лаvinу». Поэтому иногда вместо «ударная ионизация» говорят «лаvinное умножение» носителей. Термин «лаvinное умножение» оправданно вызывает в воображении образ лавины, порожденной маленьким снежным комочком и вырастающей в чудовищную, стремительно несущуюся снежную стену.

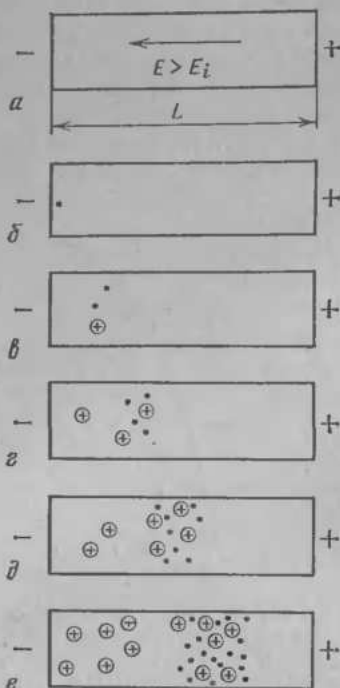


Рис. 73. а — В сильном электрическом поле $E > E_i$ носители заряда приобретают способность порождать при столкновении с решеткой новые электроны и дырки. б — е — Картина лавинного умножения, порожденного одним-единственным электроном, влетевшим в область поля, где возможна ударная ионизация. Предполагается, что ударную ионизацию могут вызывать только электроны

Уровень умножения в области ударной ионизации характеризуется коэффициентом умножения носителей M . Величина этого коэффициента равна отношению числа электронов, вылетающих из области ионизации, к числу влетающих в область электронов. Иными словами, коэффициент умножения равен числу электронов на выходе из области умножения, порожденных одним электроном на входе.

Из рис. 73 ясно, что значение M тем больше, чем больше длина образца L и чем больше носителей создает электрон на единице пройденного им пути.

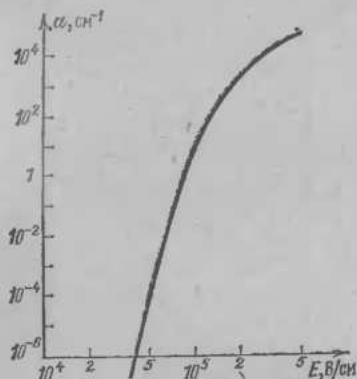


Рис. 74. Зависимость коэффициента ударной ионизации электронов α от напряженности электрического поля E в кремнии

Число носителей, создаваемых электроном за счет ударной ионизации на пути в один сантиметр, называется *коэффициентом ударной ионизации* электронов. Это самая важная при описании процесса ударной ионизации величина. Измеряется она в см^{-1} , а обозначается греческой буквой α .

На рис. 74 показана зависимость коэффициента ударной ионизации электронов α в кремнии от напряженности элект-

рического поля E . Видно, что зависимость эта чрезвычайно резкая.

При $E < 5 \cdot 10^4$ В/см величина α ничтожна. При $E = 5 \cdot 10^4$ В/см электрон способен создать электронно-дырочную пару в кремнии, только пройдя путь $L \sim \sim 1/\alpha \sim 100$ м. Ясно, что ни о каком лавинном умножении при таком поле практически нет и речи. Ведь длина полупроводниковых образцов составляет, как правило, единицы или десятки микрометров, в исключительных случаях — несколько миллиметров.

А вот при $E = 5 \cdot 10^5$ В/см ситуация совсем другая. В таком поле, пройдя путь $L = 1$ мкм = 10^{-4} см, электрон создает в среднем 7 электронно-дырочных пар. А каждый из этих вновь созданных электронов сам создает электронно-дырочные пары. А эти электроны,

в свою очередь, тоже включаются в процесс ударной ионизации (рис. 73). Лавина стремительно нарастает.

Именно из-за очень резкой зависимости коэффициентов ударной ионизации от поля в полупроводниках и оказывается возможным ввести в качестве характеристики полупроводникового материала величину характерного поля ударной ионизации E_i . Если напряженность поля чуть меньше величины E_i , то ударная ионизация практически отсутствует. Чуть больше — лавинное умножение выражено очень сильно. Понятно, и почему для величины E_i в кремнии выбрано значение $E_i \sim 3 \cdot 10^5$ В/см. При таком поле, как видно из рис. 74, $\alpha = 1,5 \cdot 10^4$ см⁻¹. Если длина области L , где существует такое сильное поле, равняется, как это чаще всего бывает, 2—3 мкм, то электрон, пролетая через эту область, успевает создать в ней $N = \alpha L \approx 3-4$ электронно-дырочные пары. Ударная ионизация в таких условиях выражена очень сильно. А вот при поле чуть меньшем, скажем $2 \cdot 10^5$ В/см, $\alpha = 2 \cdot 10^3$ см⁻¹. На пути $L = 2-3$ мкм электрон в таком поле успевает создать всего $N = \alpha L \approx 0,4-0,6$ пары *). Ударная ионизация в таких условиях будет выражена слабо.

Зная величину коэффициента ударной ионизации α , соответствующую данному полю E , и длину области умножения L , можно определить коэффициент умножения носителей M : $M = \exp(\alpha L)$.

Величина α очень резко растет с полем (рис. 76). А коэффициент умножения M экспоненциально возрастает с увеличением α . Поэтому коэффициент умножения с ростом напряженности поля возрастает чрезвычайно резко. Так, при $L = 3$ мкм и $E = 2 \cdot 10^5$ В/см ($\alpha = 2 \cdot 10^3$ см⁻¹) коэффициент умножения $M = 1,8$. А при $L = 3$ мкм и $E = 3 \cdot 10^5$ В/см ($\alpha = 1,5 \cdot 10^4$ см⁻¹) $M = 90$. При $E = 5 \cdot 10^5$ В/см ($\alpha = 7 \cdot 10^4$ см⁻¹) и $L =$

*) Чтобы не попасть в положение ученика, получившего в ответе «два и две трети землекопа», следует ясно отдавать себе отчет в том, что величина α характеризует для каждого данного электрона только вероятность создать в данном поле электронно-дырочную пару. Конечно, создать 0,4—0,6 электронно-дырочной пары невозможно. Полученный ответ просто означает, что если при данном поле через область заданной длины L пройдут 10 электронов, то в среднем будут созданы 4—6 пар. При прохождении 100 электронов появятся 40—60 пар и т. д.

= 3 мкм величина M составит $\sim 4 \cdot 10^{10}$! При возрастании поля всего в 2,5 раза коэффициент умножения увеличится в 20 миллиардов раз!

А теперь взглянем на рис. 75. На нем показана уже знакомая нам картина распределения электрического поля в обратносмещенном $p-n$ переходе (ср. с рис. 62, б). Сплошной линией показана зависимость $E(x)$ при напряжении на переходе, когда максимальное поле $E < E_i$. Ударная ионизация при $E < E_i$ практически отсутствует.

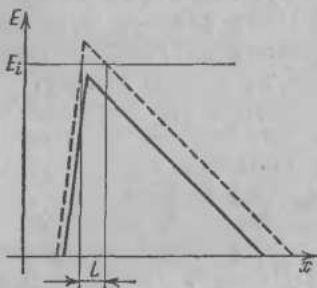


Рис. 75. Распределение напряженности электрического поля в обратносмещенном $p-n$ переходе. Штрихом показан случай $U > U_i$. В небольшой области вблизи $p-n$ перехода длиной L происходит лавинное умножение носителей тока

Через переход протекает небольшой обратный ток $I_{обр}$, определяемый суммой тока насыщения (формула (38)) и генерационного тока (формула (41)). Каждую секунду через единицу площади $p-n$ перехода проходят при этом $N = I/eS$ электронов.

Штрихом на рис. 75 показана картина распределения поля при чуть-чуть большем обратном напряжении. Она почти такая же, как и в предыдущем случае. Но теперь напряженность поля в маленькой области вблизи перехода

(длиной L) превышает значение E_i ! Это значит, что в этой области происходит лавинное умножение. Каждый из N электронов, проходивших через переход при $E < E_i$, положит начало лавине и приведет к появлению MN электронно-дырочных пар. Ток, протекающий через переход, возрастает в M раз.

Мы только что видели, как резко увеличивается коэффициент умножения M даже при небольшом увеличении поля E в области напряжений $U > U_i$, соответствующих наступлению ударной ионизации; обратный ток через $p-n$ переход начинает чрезвычайно резко возрастать с дальнейшим ростом напряжения (рис. 63, 64).

В действительности рост тока с увеличением напряжения при $U > U_i$ будет происходить еще более резко, чем это следует из формулы $M = \exp(\alpha L)$. Ведь на самом деле способностью порождать в акте ударной иониза-

ции электронно-дырочные пары обладают не только электроны, но и дырки. А это обстоятельство приводит совсем к другой картине развития процесса ударной ионизации и к еще более резкой зависимости коэффициента умножения M от поля.

Рис. 76 иллюстрирует картину развития ударной ионизации в случае, когда способностью порождать электронно-дырочные пары обладают носители тока обоих знаков.

Снова начнем рассмотрение с одиночного электрона, влетевшего в область умножения (рис. 76, б). Пролетев расстояние $x \sim 1/\alpha$, электрон порождает электронно-дырочную пару (рис. 76, в). Пока что все выглядит точно так же, как на рис. 73, в. Но уже в следующем кадре «кинофильма» (рис. 76, г) выявляется разница.

Не только два электрона, пролетев слева направо расстояние $x \sim 1/\alpha$, создали две электронно-дырочные пары. Теперь, в отличие от рис. 73, г, и дырка, продвинувшись справа налево, тоже породила пару: электрон и дырку!

В следующем акте ударной ионизации каждый электрон и каждая дырка также произведут на свет новую электронно-дырочную пару (рис. 76, д).

Ясно, что лавинное умножение в этом случае происходит гораздо быстрее, т. е. при том же поле $E > E_i$ коэффициент умножения M будет гораздо больше, чем в случае ударной ионизации одним сортом носителей.

Не удивительно, что при достижении напряжения лавинного пробоя U_i ток в обратном направлении при дальнейшем увеличении обратного напря-

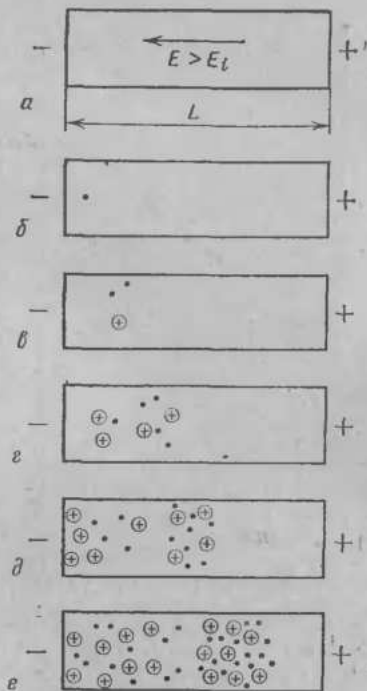


Рис. 76. Картина лавинного умножения в случае, когда ударная ионизация производится и электронами, и дырками (ср. с рис. 73)

жения растет чрезвычайно быстро. При возрастании обратного напряжения на несколько процентов ток через переход может возрасти в сотни тысяч и даже миллионы раз.

На рис. 77 показана типичная вольт-амперная характеристика лавинного диода — одного из приборов, в котором практически используется эффект лавинного умножения в обратносмещенном $p-n$ переходе. Лавинные диоды защищают аппаратуру от внезапных всплесков напряжения в электрических сетях.

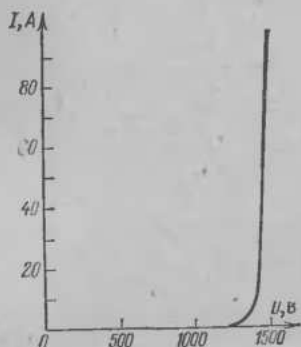


Рис. 77. Типичная вольт-амперная характеристика мощного высоковольтного обратносмещенного $p-n$ перехода (лавинного диода) на участке лавинного пробоя

Внезапное повышение напряжения в сети — крайне нежелательное, но, увы, почти неизбежное явление, с которым приходится считаться.

Несмотря на все меры, которые предпринимаются для защиты от гроз линий электропередач (ЛЭП), 10—15 раз в году в эти колоссальные сооружения, на тысячи километров передающие электрическую энергию при напряжении в сотни тысяч вольт, попадает прямой удар молнии.

Напряжение, при котором происходит пробой атмосферы и возникает молния, лежит в диапазоне 15—300 млн. вольт! Немудрено, что напряжение в ЛЭП, если бы не были предусмотрены схемы защиты от перенапряжений, могло бы очень сильно возрасти.

Перенапряжения в сетях могут возникать и при менее драматических обстоятельствах: запуске или внезапной остановке мощных двигателей или генераторов, обрыве фазы трансформатора, включении или отключении мощных энергоемких потребителей и т. д. Во всех этих и многих других случаях лавинные диоды пропустят через себя избыточный импульс тока и предотвратят повышение напряжения на защищаемой аппаратуре.

Существует довольно много схем защиты с использованием лавинных диодов. Однако все они основаны на одном и том же простом принципе, поясняемом рис. 78.

Схема, показанная на рис. 78, кроме защищаемой аппаратуры (1), включает и два элемента защиты — лавинный диод (2) и сопротивление нагрузки R_n (3). Элементы защиты выбираются так, чтобы при нормальном (номинальном) напряжении в сети $U = U_0$ на них рассеивалась очень маленькая мощность, гораздо меньшая, чем мощность, потребляемая защищаемой аппаратурой. При $U = U_0$ напряжение на лавинном диоде меньше, чем напряжение лавинного пробоя U_i .

Если напряжение в сети возрастает и его значение становится больше, чем величина U_i , ток через лавинный диод резко увеличивается. Увеличивается падение напряжения на сопротивлении R_n . Большая часть избыточного напряжения падает на сопротивлении нагрузки — и защищаемой аппаратуре «достается» гораздо меньше.

Пусть, например, защищаемая аппаратура предназначена для работы при напряжении $U_0 = 1000$ В. Потребляемый ею при этом ток составляет, скажем, $I_0 = 1$ А, а потребляемая мощность $P = U_0 I_0 = 1$ кВт. Аппаратура способна выдержать кратковременное повышение напряжения до 1500 В, но не больше. А напряжение в сети U_s между тем, может иногда возрасти и до гораздо больших значений.

Пусть параллельно защищаемой аппаратуре включен лавинный диод, вольтамперная характеристика которого показана на рис. 77. При номинальном напряжении $U = U_0 = 1000$ В ток, протекающий через лавинный диод I_d , как видно из рис. 77, ничтожен. Он составляет тысячные доли тока, потребляемого защищаемой аппаратурой.

Выберем значение сопротивления нагрузки $R_n = 20$ Ом. Тогда при токе $I_0 = 1$ А на нем падает напряжение $U_n = R_n I_0 = 20$ В, а рассеиваемая на сопротивлении мощность равняется 20 Вт — всего 2 % мощности, потребляемой аппаратурой.

Пусть напряжение в сети U резко возросло. Как изменится напряжение на защищаемой аппаратуре?

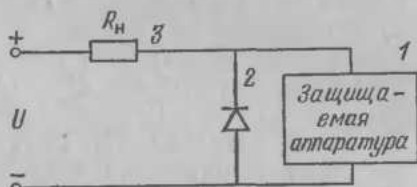


Рис. 78. Схема защиты аппаратуры от перенапряжения с помощью лавинного диода

Каковы будут токи и напряжения в элементах схемы? Существует простой и изящный графический метод, позволяющий наглядно и исчерпывающе ответить на все эти вопросы *). Здесь мы этот метод разбирать не будем, а интересующие нас ответы получим несколько неожиданным методом «от противного».

Предположим, например, для начала, что вследствие увеличения напряжения в сети U величина напряжения на аппаратуре (и соответственно на лавинном диоде — они включены параллельно) возросла до 1400 В. Из рис. 77 видно, что при $U = 1400$ В ток через лавинный диод $I_{\text{л}}$ составит 10 А. Падение напряжения на сопротивлении нагрузки $R_{\text{н}}$ составит теперь $U_{\text{н}} \approx R_{\text{н}} I_{\text{л}} \approx 200$ В. Напряжение в сети U равняется, очевидно, сумме падений напряжений на аппаратуре и на нагрузочном сопротивлении $R_{\text{н}}$, т. е. $U = 1600$ В. Рассуждая «от противного», приходим к заключению, что если напряжение U станет равным 1600 В, то 200 В упадут на сопротивлении $R_{\text{н}}$, а напряжение на лавинном диоде и защищаемой схеме составит 1400 В.

Теперь предположим, что напряжение на аппаратуре и лавинном диоде составляет 1500 В. Из рис. 77 находим, что при этом ток через лавинный диод $I_{\text{л}}$ составляет 100 А. Падение напряжения на сопротивлении нагрузки $U_{\text{н}} \approx R_{\text{н}} I_{\text{л}} \approx 2000$ В. И снова «от противного» приходим к заключению, что при напряжении $U = 3500$ В на сопротивлении нагрузки упадут 2000 В, а напряжение на лавинном диоде и защищаемой схеме составит 1500 В.

Благодаря наличию в схеме лавинного диода в экстремальных условиях, когда перенапряжение в схеме составило $3500 - 1000 = 2500$ В, лавинную долю этого избыточного напряжения — 2000 В — приняло на себя сопротивление нагрузки и только 500 В пришлось на долю защищаемой аппаратуры.

Из разобранных примеров ясно, что уровень, на котором «стабилизируется» напряжение, на защищаемой схеме, грубо говоря, определяется напряжением пробоя лавинного диода U_i . Промышленностью выпускаются лавинные диоды с напряжением пробоя в диапазоне от 10 до 2000 В. При необходимости осуществить защиту

*) Этот метод довольно подробно описан, например, в книге: Левинштейн М. Е., Симин Г. С. Знакомство с полупроводниками. — М.: Наука, 1984. — Библиотечка «Квант», вып. 33, с. 151.

в более высоковольтных сетях лавинные диоды соединяют последовательно (рис. 79).

Эффект ударной ионизации практически используется не только в лавинных диодах. Он лежит также в основе работы лавинно-пролетных диодов — приборов, используемых для генерации и усиления сверхвысокочастотных колебаний. Лавинно-пролетные диоды были созданы в конце 50-х — начале 60-х годов на основе открытого группой советских ученых во главе с Александром Семеновичем Тагером эффекта высокочастотной генерации в обратносмещенных $p-n$ переходах в условиях лавинного пробоя. Лавинно-пролетные диоды являются самыми мощными полупроводниковыми генераторами в диапазоне частот от ~ 30 до ~ 400 ГГц *).

Явление ударной ионизации нашло практическое применение и в получающих все более широкое распространение лавинных фотодиодах. По принципу действия лавинные фотодиоды очень сходны с уже знакомыми нам обычными фотодиодами. Разница состоит только в том, в лавинных фотодиодах напряжение на обратносмещенном $p-n$ переходе выше, так что вблизи $p-n$ перехода напряженность поля E превосходит характерное поле ударной ионизации E_i . В результате в области $p-n$ перехода устанавливается коэффициент умножения $M \gg 1$. (Обычно коэффициент умножения в лавинных фотодиодах составляет 10 — 30.)

Каждый фотон, поглощенный в области объемного заряда, создает электронно-дырочную пару. Но теперь, в условиях лавинного умножения, каждая созданная светом пара, двигаясь через область объемного заряда, создает по пути M других пар. В результате при той же интенсивности падающего света фототок в лавинном фотодиоде оказывается в M раз больше, чем в обычном.



Рис. 79. «Сборка» из четырех последовательно соединенных лавинных диодов. Общее напряжение пробоя 5 кВ

*) Частоте 400 ГГц (400 млрд. герц) соответствует длина электромагнитной волны в вакууме всего около 750 мкм. На таких высоких частотах становится уже неясно, говорить ли о самом коротковолновом участке радиоволн миллиметрового диапазона или об очень длинноволновом инфракрасном излучении. Радиотехника в этой области частот сливается с оптикой!

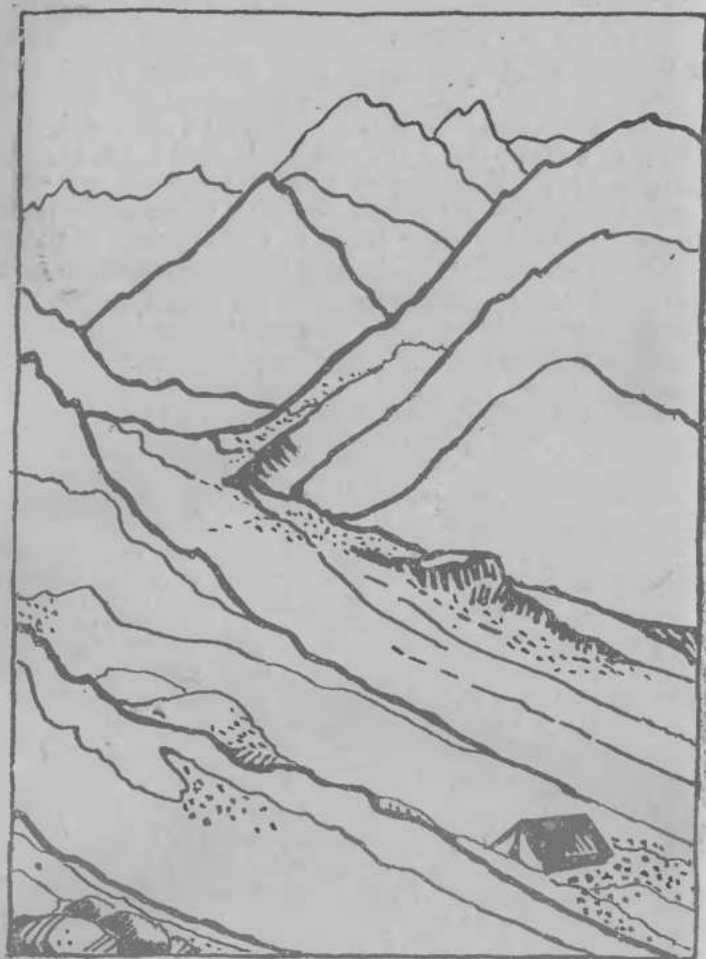


Рис. 80

Тихо, тихо ползи,
Улитка, по склону Фудзи,
Вверх, до самых высот!

Прямое смещение



На рис. 81 показан диод с $p-n$ переходом, к которому приложено прямое смещение. «Плюс» источника подсоединен теперь к p -области диода, «минус» — к n -области (ср. с рис. 62). Нам предстоит убедиться, что свойства диода, включенного в цепь в такой полярности (прямосмещенного диода), разительно отличаются от свойств обратносмещенного перехода. Обратносмещенный диод в цепи постоянного тока представляет собой очень большое сопротивление. Ток через переход, смещенный в обратном направлении, или практически не зависит от приложенного напряжения (рис. 63), или зависит слабо (рис. 64). Прямосмещенный диод представляет собой, как правило, очень маленькое сопротивление. Ток через переход, смещенный в прямом (пропускном) направлении, очень резко, экспоненциально, растет с увеличением прямого смещения.

Такая разница в свойствах обратно и прямо смещенного переходов обуславливается тем, что в прямосмещенном переходе поле $E_{вн}$ от внешнего источника U_0 направлено навстречу полю, существующему внутри области объемного заряда (рис. 53, e). И, следовательно, падение напряжения от внешнего источника U_0 вычитается из барьерной разности потенциалов $U_{pn} = \Phi_{pn}/e$, существовавшей до приложения внешнего смещения. В результате высота потенциального барьера между p - и n -областями становится равной $U_{pn} - U_0$. А высота энергетического барьера соответственно

$$e(U_{pn} - U_0) = \Phi_{pn} - eU_0.$$

Высота барьера. Несмотря на то, что предыдущий абзац с нарочитой скрупулезностью повторяет форму абзаца на с. 163, внимательного читателя он может побудить задать весьма каверзный и очень интересный вопрос.

При обратном смещении на переходе высота барьера на границе между p - и n -областями увеличивалась. Теперь, при прямом смещении, она уменьшается. Если верить только что написанной формуле, то при очень небольших прямых напряжениях ($U_0 \approx 0,7$ В для германиевого, $U_0 \approx 1,1$ В для кремниевое, $U_0 \approx 1,4$ В для арсенид-галлиевого $p-n$ переходов) барьер исчезнет. Высота его станет равной нулю.

А если приложить большее прямое смещение? Напряжение на барьере сменит знак? Станет отрицательным? Что это физически означает?

Все это очень глубокие и интересные вопросы. Ответ на них мы получим чуть дальше, анализируя свойства прямосмещенного перехода.

Пока же для начала будем считать, что приложенное прямое смещение U_0 меньше, чем высота барьера U_{pn} .

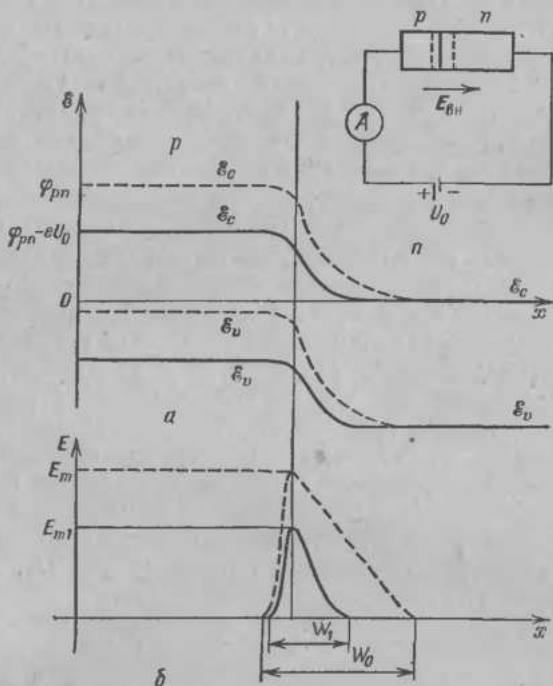


Рис. 81. Энергетическая диаграмма (а) и распределение поля (б) в $p-n$ переходе (ср. с рис. 62). Штрих — в случае отсутствия внешнего напряжения. Сплошные кривые — к диоду приложено прямое смещение

Тогда никаких парадоксов не возникает и высота энергетического барьера при прямом смещении, действительно, равняется $\phi_{pn} - eU_0$ (рис. 81, а). Поскольку уменьшилась высота барьера, уменьшилось также и максимальное поле в барьере E_m и ширина барьера W (рис. 81, б).

Прямой ток. В пункте «Чудесное равновесие» мы обсудили процессы, происходящие в $p-n$ переходе,

к которому никакое внешнее напряжение не приложено. И убедились, что в равновесии, в отсутствие смещения, мощные потоки основных носителей — электронов, стремящихся проникнуть из n - в p -область, и дырок, рвущихся из p - в n -область, — сдерживаются сильным полем барьера.

При прямом смещении высота энергетического барьера понижается. И, не удерживаемые больше полем барьера, электроны врываются (инжектируются *) в p -область, а дырки — в n -область диода.

На рис. 82 штриховой линией показана картина, в точности повторяющая рис. 61. В отсутствие смещения

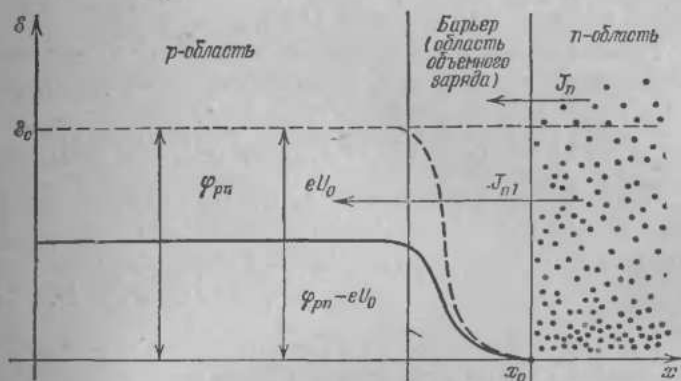


Рис. 82. Штриховая кривая показывает ход энергетического барьера для электронов в равновесии и в точности совпадает с рис. 61. Сплошная кривая — ход энергетического барьера для электронов при прямом смещении U_0 (ср. с рис. 81). Барьер понизился, и из n - в p -область проходит теперь в $\exp(eU_0/kT)$ раз больше электронов, чем в отсутствие смещения

из n - в p -область прорываются только электроны, энергия которых превышает высоту барьера φ_{pn} . Их концентрация $n \approx n_0 \exp(-\varphi_{pn}/kT)$. Образующий ими диффузионный ток J_n в точности равен электронному компоненту тока насыщения J_{ns} и уравнивает его.

Сплошная кривая $\mathcal{E}(x)$ соответствует прямому смещению U_0 . Высота барьера уменьшилась на величину eU_0 . Теперь уже в p -область способны прорваться все электроны, чья энергия превышает высоту барьера ($\varphi_{pn} - eU_0$). Концентрация таких электронов $n_1 \approx$

*) От латинского *injicio* (вбрасываю).

$\approx n_0 \exp\left(-\frac{\Phi_{pn} - eU_0}{kT}\right)$. Таким образом, число электронов, способных преодолеть понизившийся барьер, возросло в $n_1/n = \exp(eU_0/kT)$ раз. Образующий ими диффузионный ток во столько же раз возрос и составляет теперь $J_{n1} = J_{ns} \exp(eU_0/kT)$.

На ток насыщения прямое смещение (так же, как и обратное) влияния не оказывает. Ток электронов из p -области в n (ток неосновных носителей) по-прежнему равен J_{ns} и течет навстречу диффузионному току J_{n1} .

Таким образом, электронный компонент прямого тока $J_n(U_0)$ будет равен

$$J_n(U_0) = J_{n1} - J_{ns} = J_{ns} [\exp(eU_0/kT) - 1]. \quad (49)$$

Точно такое же рассуждение можно проделать и для дырок и убедиться, что дырочный компонент прямого тока $J_p(U_0)$ равняется

$$J_p(U_0) = J_{p1} - J_{ps} = J_{ps} [\exp(eU_0/kT) - 1]. \quad (50)$$

Полный ток через переход, равный сумме электронного и дырочного компонента, составляет

$$\begin{aligned} J(U_0) &= J_p + J_n = [\exp(eU_0/kT) - 1] (J_{ps} + J_{ns}) = \\ &= J_s [\exp(eU_0/kT) - 1]. \end{aligned} \quad (51)$$

Из сравнения формул (49) и (50) видно, что отношение дырочного компонента прямого тока в области перехода к электронному J_p/J_n при любых смещениях равняется отношению токов насыщения: $J_p/J_n = J_{ps}/J_{ns}$. Как мы знаем, в резко несимметричном переходе один из компонентов тока насыщения всегда пренебрежимо мал по сравнению с другим. Например, если сильнолегированной является p -область перехода, то $j_{ps} \gg j_{ns}$. А следовательно, при любых смещениях в таком переходе основной вклад в общий ток в области барьера вносят дырки.

Интересно отметить, что в действительности формула (51) описывает зависимость тока через переход от приложенного напряжения при любых смещениях: и прямых; и обратных *). Только при прямом приложенном напряжении величину U_0 следует считать положительной, а при обратном — отрицательной. При прямом

*) Постарайтесь, исходя из уже рассмотренной физической картины протекания тока через переход, сами понять, почему это так.

смещении плотность прямого тока очень резко, экспоненциально, растет с приложенным напряжением. А при обратном — насыщается на уровне $J = -J_s$, уже при величине обратного смещения U_{0z} равной нескольким kT/e (рис. 83).

При выводе формулы (51) мы, как и прежде, считали, что все приложенное к диоду напряжение U_0 падает на

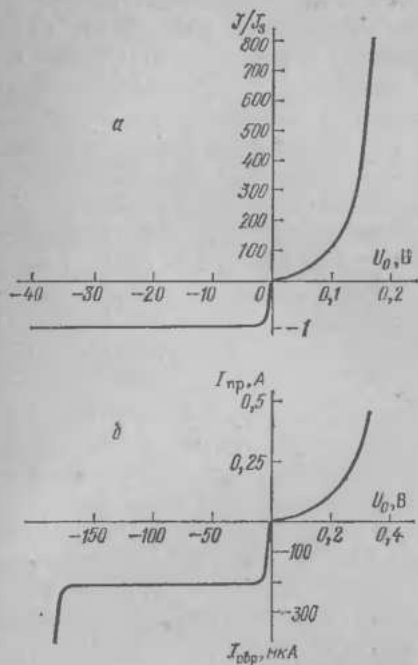


Рис. 83. Вольт-амперная характеристика $p - n$ перехода: a — вольт-амперная характеристика $p - n$ перехода, построенная в соответствии с формулой (51). Положительные значения U_0 соответствуют прямому смещению, отрицательные — обратному. Обратите внимание на разницу в масштабах по осям токов и напряжений, соответствующую прямому и обратному напряжениям; b — вольт-амперная характеристика германиевого диода с $p - n$ переходом

$p - n$ переходе, т. е. что сопротивление $p - n$ перехода гораздо больше, чем все другие, включенные последовательно с ним сопротивления: сопротивление n -области, p -области, сопротивление контактов к диоду и т. д.

При обратном напряжении такое допущение было вполне оправдано. Сопротивление барьера ($p - n$

перехода) было велико уже при нулевом смещении и с ростом обратного напряжения только увеличивалось. При прямом смещении дело обстоит совершенно иначе. С ростом прямого напряжения сопротивление барьера стремительно уменьшается!

Пусть, например, мы исследуем кремниевый диод с $p-n$ переходом площадью $S = 0,1 \text{ см}^2$ с током насыщения $J_s = 10^{-7} \text{ А/см}^2$. Тогда, в соответствии с формулой (51), при прямом смещении на таком диоде, равном $0,01 \text{ В}$, через него протекает ток $I \approx 5 \cdot 10^{-9} \text{ А}$. А сопротивление $p-n$ перехода $R_{pn} = U_0/I \approx 2 \cdot 10^6 \text{ Ом}$. При $U_0 = 0,1 \text{ В}$ ток $I = 5 \cdot 10^{-7} \text{ А}$, а $R_{pn} = 2 \cdot 10^5 \text{ Ом}$.

Проследим за изменением сопротивления R_{pn} при дальнейшем увеличении прямого смещения:

$U_0, \text{ В}$	$I, \text{ А}$	$R_{pn}, \text{ Ом}$	$U_0, \text{ В}$	$I, \text{ А}$	$R_{pn}, \text{ Ом}$
0,2	$2 \cdot 10^{-5}$	10^4	0,6	200	0,003
0,4	$5 \cdot 10^{-2}$	8	0,7	$5 \cdot 10^3$	0,00014

При напряжении на диоде $U_0 = 0,7 \text{ В}$ (если предположить, что все приложенное напряжение упадет на самом $p-n$ переходе (на барьере)) ток через диод составит $\sim 50000 \text{ А}$, плотность тока через диод — 50000 А/см^2 , а сопротивление барьера $R_{pn} = 0,00014 \text{ Ом}$.

Сопротивление $R \sim 10^{-4} \text{ Ом}$ большим никак не назовешь. Такое же, и даже существенно большее значение могут иметь и сопротивление p - и n -областей диода, и даже сопротивление контактов. Что из этого следует?

Прежде всего тот очевидный вывод, что если мы приложим к рассмотренному выше диоду напряжение $U_0 = 0,7 \text{ В}$, то не все это напряжение упадет на $p-n$ переходе. Заметная часть напряжения упадет на включенных последовательно с переходом контактах, n - и p -области (рис. 81, верхняя вставка). Следовательно, высота барьера понизится не на $0,7 \text{ В}$, а на меньшую величину. И, значит, плотность протекающего через диод тока окажется меньше, чем вычисленная по формуле (51).

Когда приложенное к диоду прямое напряжение возрастает, приближаясь к высоте потенциального барьера на $p-n$ переходе, все большая и большая часть напряжения падает не на $p-n$ переходе, а на контак-

тах и на сопротивлении p - и n -областей. В результате ситуация, когда к потенциальному барьеру на границе между p - и n -областями приложено прямое напряжение, превышающее исходную высоту барьера U_{pn} , оказывается невозможной. Самое большее, чего можно добиться, увеличивая прямое напряжение на диоде, — это очень сильно уменьшить высоту барьера (как иногда говорят — «спрямить» барьер). Но достижение даже этого эффекта требует, чтобы плотность тока через диод составляла десятки и даже сотни тысяч ампер/см².

Инжекция. Мы только что установили, что если к диоду прикладывается прямое смещение, высота потенциального барьера понижается и электроны из n -области врываются (инжектируются) в p -область диода. А дырки из p -области инжектируются в n -область. Интересно, полезно и поучительно проследить за судьбой этих инжектированных носителей.

Начнем, например, с дырок, инжектированных в n -область. Оказавшись в n -области, дырки немедленно попадают в положение незваных гостей — избыточных неравновесных носителей, и начинают усиленно вымирать (рекомбинировать) с характерным временем жизни τ_p . Рекомбинируя, они в то же время будут перемещаться под действием диффузии из области p — n перехода, где их больше всего, в глубину n -области. Если прямой ток через переход протекает достаточно долго, то в n -области устанавливается стационарное распределение дырок (рис. 84). Концентрация дырок максимальна у p — n перехода, откуда дырки инжектируются, и спадает в глубину n -области. Характерное расстояние, на которое дырки проникают в n -область, равняется диффузионной длине дырок $L_p = \sqrt{D_p \tau_p}$. Картина очень напоминает рассмотренную в гл. 2 диффузию неравновесных носителей. Только там избыточные носители создавались светом, а здесь — инжекцией через p — n переход (ср. с рис. 23).

Разумеется, точно так же будет обстоять дело и с электронами, инжектированными из n - в p -область (рис. 85).

Из рис. 84, 85 видно, что чем больше прямой ток, протекающий через диод, тем больше концентрация неравновесных носителей, инжектируемых из одной области перехода в другую. Это понятно, так как прямой ток тем больше, чем ниже потенциальный барьер, разделяющий p - и n -области диода. А более низкому потен-

циальному барьеру соответствует и бóльшая концентрация инжектируемых носителей (рис. 82).

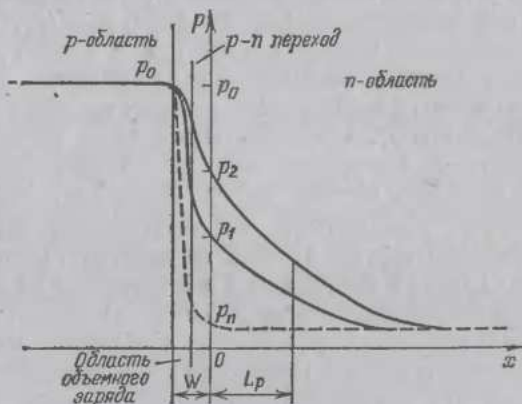


Рис. 84. Распределение дырок, инжектированных в n -область диода при прямом смещении на $p-n$ переходе. Штриховая кривая — ток через переход равен нулю. Концентрация дырок в n -области равняется равновесной величине $p_n = n_i^2/n_0$ (ср. с рис. 60). Сплошные кривые соответствуют двум разным значениям протекающего через диод прямого тока. Чем больше ток, т. е. чем больше напряжение, приложенное к переходу, тем больше дырок инжектируется из p -области в n и тем больше концентрация дырок на границе n -области в точке $x = 0$

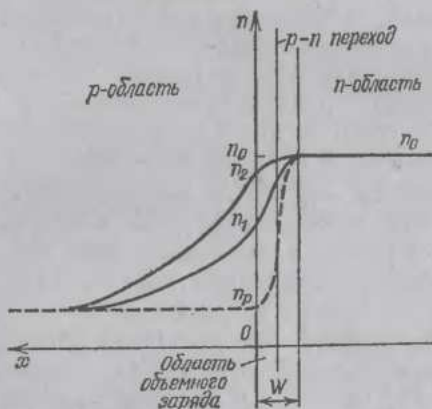


Рис. 85. Распределение электронов, инжектированных в p -область диода при прямом смещении на $p-n$ переходе. Когда ток через переход равен нулю, концентрация равновесных электронов $n_p = n_i^2/p_0$. Чем больше протекающий через переход ток, тем больше концентрация неравновесных электронов (ср. с рис. 84)

Предельно большой плотности тока (сотни тысяч и миллионы ампер/см²), соответствующей практическому спрямлению потенциального барьера, будет соответствовать концентрация дырок p на границе n -области, равная p_0 (рис. 84).

Рис. 86 показывает одновременно распределение и дырок, и электронов в $p-n$ структуре при протекании прямого тока. Как уже несколько раз упоминалось,

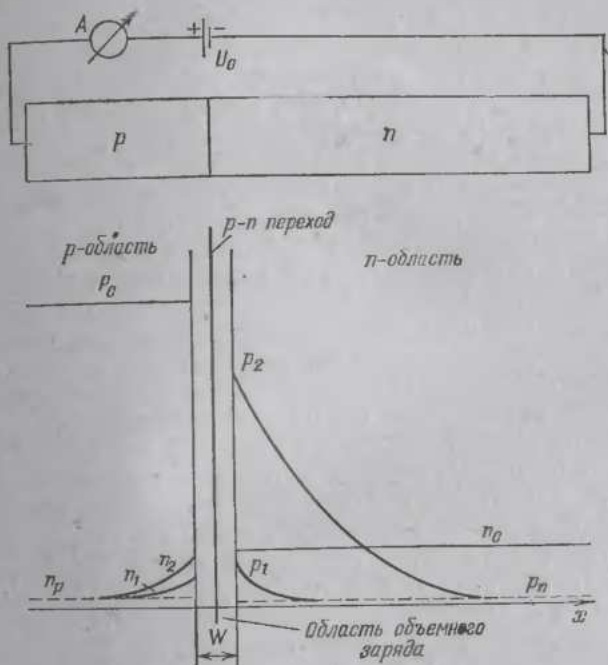


Рис. 86. Распределение электронов и дырок в прямосмещенном $p-n$ переходе. (Ход зависимостей $p(x)$ и $n(x)$ в области объемного заряда не показан, чтобы не загромождать рисунок.) Концентрации на границах p_1 и n_1 соответствуют току I_1 . Концентрации p_2 и n_2 соответствуют большему току I_2

в подавляющем большинстве случаев практически используемые $p-n$ переходы являются резко несимметричными. (Одна из областей, образующих переход, легируется значительно сильнее, чем другая.) В приведенном на рис. 86 примере сильнее легирована p -область.

Из рис. 86 непосредственно следует очень важный вывод. Концентрация носителей, которые слаболегиро-

ванная область способна инжектировать в сильнолегированную, всегда очень мала по сравнению с концентрацией носителей в сильнолегированной области. Так, на рис. 86 даже при самой большой плотности прямого тока концентрация инжектированных в p -область электронов n всегда много меньше, чем величина p_0 .

А вот концентрация носителей, которые сильнолегированная область способна инжектировать в слаболегированную, даже при относительно небольшой плотности тока может стать во много раз больше, чем исходная концентрация носителей в слаболегированной области. Так, на рис. 86 при токе I_2 концентрация дырок на границе n -области p_2 много больше, чем концентрация электронов в n -базе n_0 ($p_2 \gg n_0$).

Вот почему сильнолегированная область диода называется *эмиттером* (от латинского *emitto* — выпускаю). Она способна выпустить (инжектировать) в слаболегированную область такое число неосновных носителей, что их концентрация значительно превысит исходную концентрацию основных носителей.

Слаболегированная область диода называется *базой* (от греческого *basis* — основание, основа). Как мы помним, именно в слаболегированной области перехода, базе, в основном размещается область объемного заряда при обратном смещении (рис. 62). Параметры базы — ее ширина и уровень легирования — определяют напряжение пробоя $p-n$ структуры (формулы (39), (40)). Итак, в соответствии с принятой терминологией, на рис. 86 анализируется работа диода с *p -эмиттером* и *n -базой*.

Интенсивность, или, как говорят, *уровень инжекции*, в каждой из областей диода принято оценивать по отношению концентрации неосновных носителей на границе (где она максимальна) к концентрации основных носителей. Из рис. 86 ясно, что для эмиттерной области диода уровень инжекции всегда будет низким. Концентрация электронов n при любом токе много меньше, чем p_0 .

А вот для базовой области ситуация зависит от тока. Для тока I_1 уровень инжекции низкий $p_1/n_0 < 1$. А для большего тока I_2 реализуется очень высокий уровень инжекции в базе: $p_2/n_0 \gg 1$.

Снова прямой ток. Рассматривая рис. 86, наиболее внимательные читатели, быть может, уже задались вопросом, который мы собираемся здесь рассматривать.

Как мы хорошо знаем из школьного курса физики, при протекании постоянного тока в любом сечении про-

водника сила тока должна быть постоянна. Если через прямосмещенный диод протекает ток I , то его сила одна и та же и в проводах, с помощью которых диод подключен к источнику напряжения U_0 , и в глубине p -области, где есть только дырки, и в глубине n -области, где есть только электроны, и в районе $p-n$ перехода, где вклад в ток дают и электроны, и дырки.

Пользуясь формулой (51), мы можем рассчитать силу протекающего через диод тока I . А знакомство с явлением инжекции поможет нам ответить на вопрос, как протекает ток через диод и какова физическая картина протекания тока в различных частях диода.

Рассмотрим сначала картину протекания тока в базе диода.

Длина показанной на рис. 86 n -базы диода в несколько раз превосходит диффузионную длину дырок L_p , так что в глубине n -базы, вдали от $p-n$ -перехода, никаких дырок нет. Все дырки, инжектированные из p -эмиттера, успевают умереть прежде, чем доберутся до правой части базы, наиболее удаленной от перехода. Ясно поэтому, что ток в глубине n -базы, вдали от перехода, может переноситься только электронами. Если через диод течет ток I , то это означает, что каждую секунду источник смещения U_0 «впихивает» в базу $N = I/e$ электронов. Эти электроны движутся против направления тока, справа налево. И... куда они деваются?

Если бы база представляла собой просто кусочек полупроводника n -типа, как на рис. 20, то все было бы просто. Сколько электронов входило бы в образец справа, столько же за то же время уходило бы из него слева, возвращаясь в источник смещения U_0 .

В базе диода дело обстоит совсем иначе. Из базы в эмиттер вытекает электронный компонент полного тока I_n . Но, как мы уже знаем, если эмиттер легирован значительно сильнее, чем база, то этот электронный компонент составляет ничтожную долю полного тока I , т. е. из базы в эмиттер выходит лишь ничтожная доля электронов, входящих в базу *). Куда же деваются остальные? А остальные погибают, рекомбинируют с дырками.

*) Эта доля, как легко подсчитать (сделайте это сами), равняется отношению концентрации доноров в базе N_d к концентрации акцепторов в эмиттере N_a . Мы уже упоминали, что это отношение, как правило, составляет 10^{-4} — 10^{-5} . Таким образом, только один из 10000 или даже один из 100000 электронов, вошедших в базу, выходит из нее в эмиттер. Этой ничтожной долей часто пренебрегают и просто говорят, что электроны из базы в эмиттер не попадают.

Все дырки, инжектированные из эмиттера в базу, если база достаточно длинная, погибают (рис. 86). Но ведь чтобы рекомбинировать, дырке нужен партнер — электрон (рис. 7,а). Поэтому сколько гибнет каждую секунду дырок, столько же и электронов должно быть поставлено в базу источником смещения U_0 . Как мы знаем, вблизи $p - n$ перехода в базе практически весь ток I переносится дырками, т. е. каждую секунду из эмиттера в базу входит приблизительно $N = I/e$ дырок. И столько же электронов поступает им навстречу из глубины n -базы.

Вопрос о том, как дырки инжектируются из эмиттера в базу, мы уже обсудили несколько раньше, в пункте «Прямой ток». А как же поставляются дырки в эмиттер? Откуда они берутся?

Каждую секунду, «повинуясь притяжению» со стороны «плюса» источника (рис. 86), $N = I/e$ электронов покидают p -область диода и оставляют в ней такое же количество дырок *).

Возникшие дырки движутся слева направо, от контакта к p -эмиттеру в сторону $p - n$ перехода и инжектируются в n -базу. Попав в n -базу, они благодаря диффузии движутся в глубь n -базы и по пути рекомбинируют с приходящими им навстречу электронами.

А можно сказать и иначе. Каждую секунду, «повинуясь отталкиванию» со стороны «минуса» источника, $N = I/e$ электронов входят в n -базу. Покинуть базу они не в состоянии. Притягиваемые отрицательно заряженными электронами, со стороны эмиттера входят дырки и гибнут, рекомбинируя с электронами. Чтобы поддержать их непрерывное поступление в базу, каждую секунду из p -области в металлический контакт выходят $N = I/e$ электронов, оставляя в p -эмиттере такое же количество дырок.

Вот как хитро «замыкается» ток через прямосмещенный диод.

Светодиоды. Самым наглядным и ярким (в прямом и переносном смысле) применением $p - n$ переходов, работающих при положительном смещении, бесспорно,

*) Подробнее рассмотрение того, как это происходит, увело бы нас слишком далеко. Нам пришлось бы преодолеть еще один барьер — барьер металл — полупроводник, существующий на границе между металлическим контактом к диоду и материалом полупроводника.

являются светодиоды. Можно даже сказать, что, в принципе, любой прямосмещенный $p - n$ переход является светодиодом. В самом деле, попавшие из эмиттера в базу носители рекомбинируют. По крайней мере часть носителей рекомбинирует *излучательно*, т. е. с рождением фотона (рис. 7, а, 17). Какая-то часть фотонов, избежав поглощения в самом диоде, сумеет вылететь наружу. . . Вот вам и свет, рожденный протеканием через $p - n$ переход прямого тока.

А дальше... Дальше начинаются всякого рода необходимые уточнения.

Для создания *практически* пригодного светодиода годится вовсе не любой полупроводник. Первое очевидное требование, предъявляемое к полупроводнику, состоит в том, чтобы как можно большая часть носителей в нем гибла с испусканием фотонов (излучательно). Или, выражаясь научно, для изготовления светодиодов пригодны только материалы с достаточно высокой вероятностью излучательной рекомбинации.

Германий и кремний, например, этому требованию не отвечают. В них подавляющая доля электронов и дырок гибнет без испускания фотона, *безизлучательно*, рекомбинируя через глубокие центры (убийцы). А вот в арсениде галлия (GaAs) и тройных полупроводниковых соединениях на его основе: GaAsP и GaAlAs, даже и не особенно хорошо очищенных, вероятность излучательной рекомбинации близка к единице. Почти каждая электронно-дырочная пара гибнет с испусканием фотона. Поэтому большая часть светодиодов выпускается сегодня именно на основе этих трех материалов.

Цвет, а говоря научнее, длина волны света, излучаемого светодиодом, определяется энергией фотона \mathcal{E}_f , испускаемого при рекомбинации. В большинстве случаев эта энергия близка к ширине запрещенной зоны полупроводника \mathcal{E}_g .

Так, длина волны излучения арсенид-галлиевого ($\mathcal{E}_g \approx 1,4$ эВ) светодиода равняется $\lambda = hc/\mathcal{E}_g = = 1,24/1,4 = 0,89$ мкм. Такое излучение для человеческого глаза невидимо. Оно соответствует инфракрасной области спектра, но это вовсе не мешает очень широкому использованию арсенид-галлиевых источников света. Приемниками такого инфракрасного излучения служат кремниевые фотодиоды, для которых энергия фотона $\mathcal{E}_g \approx \approx 1,4$ эВ приблизительно соответствует максимуму их спектральной чувствительности (рис. 67).

Добавление в решетку GaAs атомов фосфора (P) или алюминия (Al) приводит к увеличению ширины запрещенной зоны \mathcal{E}_g (рис. 87). При увеличении в тройном полупроводниковом соединении $\text{GaAs}_{1-x}\text{P}_x$ доли относительного содержания фосфора x ширина запрещенной зоны полупроводника, как видно из рис. 87, увеличивается

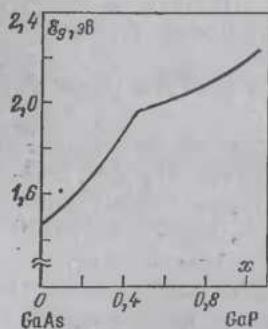


Рис. 87. Зависимость ширины запрещенной зоны \mathcal{E}_g в тройном полупроводниковом соединении $\text{GaAs}_{1-x}\text{P}_x$ от доли фосфора x

в твердый раствор GaAsP или в чистый GaP добавляют специальные примеси (азот N, цинк Zn, кислород O), увеличивающие вероятность излучательной рекомбинации.

Сегодня для создания «разноцветных» светодиодов используется очень широкий набор полупроводниковых материалов. С использованием GaAlAs получают инфракрасные и красные источники света. На основе соединений GaAsP и GaP изготавливают светодиоды с цветом свечения от красного до зеленого. Четверное полупроводниковое соединение GaInAsP позволяет создавать инфракрасные светодиоды с длиной волны $\lambda \sim 1 \div 1,3$ мкм, которые являются наилучшими источниками для волоконно-оптических линий связи. А вот на основе карбида кремния (SiC) оказалось возможным получать светодиоды всех так называемых основных цветов: красного, зеленого и синего.

Основными эти цвета называют потому, что, смешивая их в определенной пропорции, можно получить любой цвет, в том числе и белый. Именно это свойство основных цветов определяет принцип

работы цветного телевизора. Экран кинескопа цветного телевизора покрыт тремя люминофорами, излучающими красный, синий и зеленый свет. Чтобы вызвать свечение люминофоров, их возбуждают электронными пучками, разогнанными до высокой энергии.

Во многих лабораториях мира изготовители светодиодов работают сегодня над созданием плоского цветного телевизионного экрана. Основой такого экрана, быть может, послужат светодиоды основных цветов.

В настоящее время главной областью использования светодиодов являются разнообразные индикаторы. В наш век информационного взрыва появление удобных, универсальных, дешевых и эффективных приборов для индикации информации, какими являются светодиоды, позволяет решить многие важные проблемы.

Простейшие индикаторные приборы представляют собой просто светящиеся «точки» из разноцветных одиночных

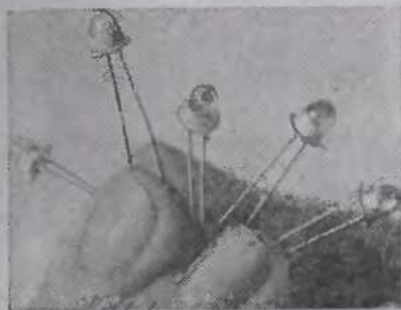


Рис. 88. Жаль, что этот снимок нельзя представить цветным. Светодиоды способны светить красным, зеленым, желтым, синим, фиолетовым светом — всеми цветами радуги

светодиодов (рис. 88). Они могут сообщать о том, включен или выключен телевизор, радиоприемник или какой-нибудь измерительный прибор, закрыты ли двери в автобусе и т. д. В сочетании с электронными схемами светодиоды решают и более сложные задачи: сигнализируют о точной настройке приемника на нужную станцию, о степени разряда батарей микрокалькуляторов, о номере включенного канала телевизора и т. д.

Одно из самых важных применений светодиодов — создание на их основе цифро-буквенных индикаторов. В таких индикаторах светодиоды объединены в группы. Пропуская ток через часть светодиодов, включенных в группу, можно высвечивать любую букву или цифру.

Понять, как это происходит, легче всего, если положить перед собой обычный почтовый конверт (рис. 89). Группе светодиодов соответствует один стандартный прямоугольничек для написания цифры индекса. Каждой из точек,



Рис. 89. Принцип цифровой индикации легко понять на примере написания индекса на почтовом конверте. Чтобы высветить цифру «1», зажигают одну определенную часть светодиодов в группе, «2» — другую и т. д.

соответствующих группе, соответствует отдельный светодиод. Понятно, что, зажигая внутри прямоугольника определенную часть светодиодов, можно высвечивать любую цифру.

Очень важной областью применения светодиодов является их работа в составе так называемой оптоэлектронной пары (оптопары, или оптрона) светодиод — фотодиод.

Внешне оптопары выглядят как обычные полупроводниковые приборы (рис. 90). Из корпуса торчат четыре проводника — два из них выводы светодиода, а другие два — выводы фотодиода. Внутри корпуса на общем диэлектрическом основании напротив друг друга размещены два диода: светодиод и фотодиод. Светодиодом обычно служит диод на основе GaAs, фотодиод изготовлен из Si.

Оптопара работает как управляемый светом ключ (рис. 91)*. Пока фотодиод (правый диод на схеме) не

*) Обратите внимание на условное обозначение оптопар на электрических схемах. В общем корпусе показаны два диода. Стрелки

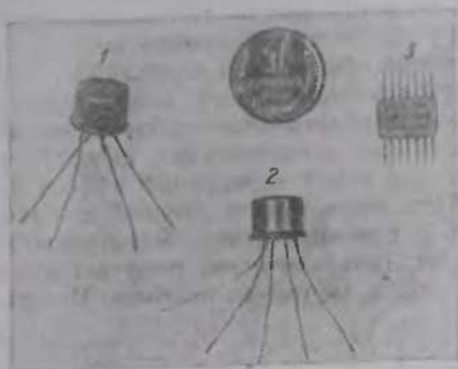


Рис. 90. Оптоэлектронные приборы: 1 — оптопара АОУ103. В корпусе этого оптоэлектронного прибора находится арсенид-галлиевый светодиод (излучатель) и кремниевый тиристор (управляемый светом ключ); 2 — оптопара АОТ102. Арсенид-галлиевый светодиод и кремниевый транзистор; 3 — диодный оптрои с одним арсенид-галлиевым светодиодом и двумя кремниевыми фотодиодами. Применяется в качестве датчика пульса в электронных пульсметрах

освещен, его сопротивление очень велико и через сопротивление нагрузки R_H протекает лишь очень слабый темновой ток фотодиода. При замыкании ключа в цепи светодиода его свет попадает на фотодиод и ток через сопротивление нагрузки R_H резко возрастает.

На первый взгляд кажется, что незачем «огород городить». Чтобы добиться того же эффекта, достаточно убрать и светодиод, и фотодиод, а ключ K перенести в цепь,

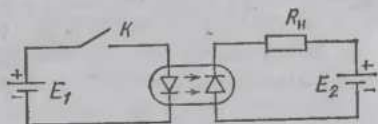


Рис. 91. Схема включения оптопары

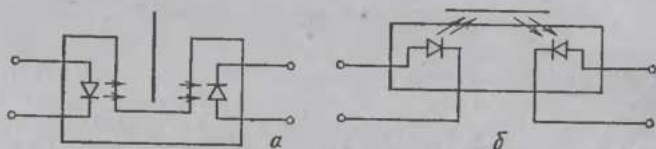
содержащую сопротивление нагрузки и батарею E_2 . Замкнешь ключ — через сопротивление R_H потечет ток $I = E_2/R_H$, разомкнешь — ток будет равен нулю.

И это, конечно, совершенно правильно. Только если представлять себе ключ K как рубильник или выключатель в комнате на стенке. На самом деле, в схемах, где используются оптопары, такими ключами служат полупроводниковые приборы — диоды или транзисторы. Слож-

указывают направление распространения света — от светодиода к фотодиоду.

ные схемы, например электронно-вычислительные машины (ЭВМ), содержат сотни миллионов таких ключей, расположенных очень близко друг к другу. И, увы, случается так, что срабатывание одного ключа вызывает нежелательное и непредусмотренное срабатывание других (из-за помех, случайных электрических шумов).

В оптопарах такой паразитный эффект исключен абсолютно. Ведь связь между входной и выходной цепью замыкается с помощью света. Электрически (или, как часто говорят, гальванически) цепи источников E_1 и E_2 абсолютно друг с другом не связаны. Поэтому оптопары



используются во всех схемах, где опасны электрические наводки, в том числе при передаче информации между блоками ЭВМ.

Интересные задачи решают и такие оптопары, в которых свет от светодиода попадает на фотодиод не непосредственно, а пройдя через препятствие или отразившись от какой-либо поверхности. Их называют оптопарами с открытым оптическим каналом (рис. 92).

Оптопары типа оптического прерывателя (рис. 92, а) служат датчиками наличия объектов на пути светового луча. Они широко используются в разнообразных системах автоматики. Благодаря им увеличилась скорость считывания информации с перфокарт и перфолент в ЭВМ, возрос сигнал, вырабатываемый фотодиодом.

Отражательные оптопары (рис. 92, б) контролируют наличие вблизи прибора отражающих объектов. Некоторые из оптопар этого типа сконструированы так, что вырабатываемый ими сигнал очень сильно зависит от расстояния между прибором и отражателем. Установленные на автоматических станках с программным управлением специальные отражательные оптопары вырабатывают ток 1—2 мкА при изменении расстояния до обрабатываемой детали на 1 мкм! Такие станки имеют уникальную «микронную» точность обработки.

Ну, и наконец, еще раз напомним о том, что именно полупроводниковые источники света служат «передатчиками» в самых современных линиях связи — оптических.

Солнечные батареи. Всему на свете приходит конец. На наших глазах истощаются энергетические ресурсы, накопленные в Земле за сотни миллионов лет: нефть, природный газ и уголь. Уже далеко не такими неисчерпаемыми, как в начале атомной эры, кажутся запасы урана. Все труднее становится добывать железо и медь, серебро и цинк, вольфрам и марганец, никель и золото...

В хозяйственный обиход человечества вовлекаются все более бедные руды. Роятся все более глубокие шахты. Нефтяные вышки и разведочные платформы шагают с суши в зону континентального шельфа. Но каждый шаг в этих направлениях требует все больше энергии на каждый килограмм добываемого металла, на каждый литр нефти и кубометр газа. Откуда же взять необходимую энергию?

И Человечество, вооруженное всеми чудесными и грозными завоеваниями науки, как и во времена Юности и Детства снова поднимает глаза к Солнцу:

Золото лучей его струится к поздрым фараонам.

Да буду облит я лучами твоими каждый день *).

Каждую секунду Земля получает от Солнца в виде излучения $\sim 3 \cdot 10^{17}$ Дж энергии — в 30 000 раз больше, чем производят ее за это же время все действующие электростанции мира: тепловые, гидро- и атомные. И этот чудесный источник неисчерпаем — термоядерная реакция будет стабильно протекать на Солнце еще много миллиардов лет.

Отказываться от столь щедрого источника энергии, естественно, было бы непростительным расточительством. Поэтому способы преобразования солнечной энергии в электрическую исследуются уже очень давно. Первое предложение использовать для этой цели полупроводники относится к 1884 г., когда была сделана попытка изготовить *солнечный элемент* на основе соединения селена с золотом. И вот уже более 100 лет продолжается упорная работа по созданию и совершенствованию полупроводниковых преобразователей солнечной энергии. Как же устроены солнечные элементы и в чем проблема их изготовления?

*) Этот отрывок из египетского гимна взят нами из замечательной книги С. И. Вавилова «Глаз и Солнце».

Солнечный элемент — это все тот же $p - n$ переход, и работа его очень похожа на работу уже известного нам фотодиода. Однако в отличие от фотодиода внешнее смещение к солнечному элементу не прикладывается.

Вернемся к рис. 65 и представим себе, что никакого внешнего смещения к диоду не приложено, никакого источника во внешней цепи нет. А нижний (сплошной)

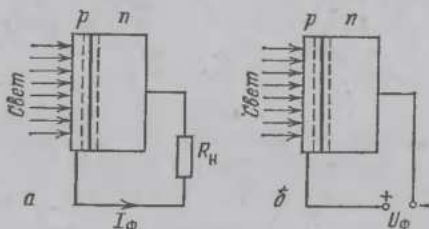


Рис. 93. Возникновение фототока (а) и фото-э.д.с. (б) в $p - n$ -переходе под действием света

и верхний (кольцевой) контакты просто соединены друг с другом через сопротивление нагрузки R_n (рис. 93, а).

Как мы знаем, и в отсутствие смещения между p - и n -областями существует потенциальный барьер. Вблизи $p - n$ перехода имеется область объемного заряда (рис. 58, а), в которой существует электрическое поле (рис. 58, б).

А значит, и без всякого внешнего смещения электронно-дырочные пары, возникшие под действием света в области объемного заряда, будут разделяться полем: дырки будут выбрасываться в p -область, электроны — в n -область. Такое движение дырок и электронов вызовет во внешней цепи ток, направленный от p -области к n -области — фототок. Фототок I_ϕ , протекающий через сопротивление нагрузки R_n (этим сопротивлением может быть нагревательный элемент, двигатель, заряжаемый аккумулятор и т. п.), совершает полезную работу.

Энергия света преобразовалась в электрическую энергию!

Если разомкнуть провод, соединяющий p - и n -области диода, ток в цепи, разумеется, протекать не будет. А между тем если свет непрерывно падает на окно солнечного элемента, то в материале непрерывно создаются электронно-дырочные пары. Они разделяются полем $p - n$ перехода. Дырки выбрасываются в p -область и создают там избыточный положительный заряд. Электроны перемещаются в n -область и создают отрицательный заряд. Между

p - и n -областями элемента возникает разность потенциалов — фото-э. д. с. U_{ϕ} (рис. 93, б).

На первый взгляд может показаться, что если p — n переход освещается достаточно долго, то величина фото-э. д. с. может сделаться сколь угодно большой. В самом деле, чем дольше свет падает на переход, тем больше дырок скапливается в p -, а электронов — в n -области, следовательно, тем больше должна быть разность потенциалов между p - и n -слоями перехода...

На самом деле это не так. Полярность фото-э. д. с. — «плюс» со стороны p - и «минус» со стороны n -области — соответствует, как легко видеть, прямому смещению на p — n переходе (ср. рис. 93, б, 81 и 86). (Именно поэтому рассказ о солнечных элементах помещен в этот раздел.) Прямое смещение на переходе, как мы знаем, понижает высоту барьера и облегчает отток дырок из p -области в n , а электронов — из n -области в p (рис. 81). В результате действия двух противоположных механизмов — накопления носителей под действием света и их оттока из-за снижения высоты барьера — при каждом значении интенсивности света, падающего на элемент, в стационарном состоянии устанавливается определенная величина фото-э. д. с. U_{ϕ} . С увеличением интенсивности света величина U_{ϕ} растет очень слабо, пропорционально логарифму интенсивности. И при самой большой интенсивности света, когда под действием фото-э. д. с. барьер оказывается практически «спрямленным», величина U_{ϕ} равняется высоте барьера на неосвещенном переходе: $U_{\phi \max} \approx U_{pn} = \mathcal{E}_g / e$.

Как мы видим, принцип действия солнечных элементов очень прост. Над чем же более ста лет работают во всем мире разработчики элементов и батарей? Конечно, над повышением надежности, долговечности, над снижением их стоимости, но главным образом — над повышением коэффициента полезного действия элементов.

Если бы каждый квант солнечного света, падающего на фотоприемное окно элемента, мог быть превращен в пару электрон — дырка, дающую вклад в фототок, то такой гипотетический солнечный элемент характеризовался бы 100%-ной эффективностью.

Эффективность (к. п. д.) первых солнечных элементов составляла сотые доли процента. К. п. д. современных элементов вырос в тысячи раз и составляет 10—20%.

Главная трудность на пути к увеличению эффективности солнечных элементов состоит в том, что солнечный свет состоит из фотонов самых разнообразных энергий. В нем содержатся и кванты, соответствующие ультрафиолетовому свету ($\mathcal{E}_{\phi} \gtrsim 3$ эВ), и кванты видимого света, и инфракрасное излучение ($\mathcal{E}_{\phi} \lesssim 1,5$ эВ). Между тем для каждого солнечного элемента, из какого бы полупроводника он ни был изготовлен, существует кривая спектральной фоточувствительности (рис. 67).

Из этой кривой легко понять, что эффективно преобразуются в фототок только кванты света с энергией $\mathcal{E}_\phi \approx \mathcal{E}_g$.

Если выбрать для изготовления элемента материал с большой шириной запрещенной зоны \mathcal{E}_g , то практически весь спектр солнечного света пройдет через такой элемент, не создав электронно-дырочных пар ($\mathcal{E}_\phi < \mathcal{E}_g$).

Если, наоборот, изготовить солнечный элемент из материала с малым значением \mathcal{E}_g , то практически все фотоны, содержащиеся в солнечном свете, будут иметь энергию $\mathcal{E}_\phi \gg \mathcal{E}_g$. Из рис. 67 ясно видно, что такие фотоны дают очень маленький вклад в фототок. Они поглощаются у самой поверхности элемента, вдали от $p - n$ перехода, и порожденные ими электронно-дырочные пары быстро гибнут за счет поверхностной рекомбинации.

Расчеты показывают, что максимальную эффективность преобразования энергии солнечного света в электрическую энергию должны иметь элементы, изготовленные из полупроводника с $\mathcal{E}_g \approx 1,5$ эВ. Вблизи этой оптимальной величины \mathcal{E}_g находятся значения запрещенной зоны хорошо изученных и освоенных в промышленности материалов Si и GaAs. Из них, главным образом, и изготавливают сейчас солнечные элементы.

Предложено множество методов увеличения предельно достижимого к. п. д. Например, путем последовательного соединения элементов с разными \mathcal{E}_g можно поднять к. п. д. примерно до 60%. Это происходит за счет того, что в каждом элементе фототок создается «своими» фотонами с энергией, близкой к величине \mathcal{E}_g этого элемента, и меньшее число фотонов пропадает зря.

Такой «составной» солнечный элемент может быть изготовлен, например, на основе тройного твердого раствора с переменным по глубине содержанием одного из элементов. Взглянем на рис. 87. Видно, что в тройном твердом растворе GaAsP ширина запрещенной зоны \mathcal{E}_g монотонно растет с увеличением содержания фосфора. Можно изготовить кристалл, в котором содержание фосфора, а следовательно, и ширина запрещенной зоны будет зависеть от координаты (*варизонный кристалл*). (Вспомните для сравнения, как получали Аллен и Гобели кристалл кремния с переменным по координате содержанием примесей (рис. 34).) Если на основе такого варизонного кристалла сделать солнечный элемент, то он будет способен эффективно улавливать кванты света в широком диапазоне энергии фотонов \mathcal{E}_ϕ .

Несколько повышает к. п. д. использование различных концентраторов света — линз или зеркал, собирающих свет с большой площади и концентрирующих его на приемное окно солнечного элемента (рис. 94).

Типичное значение фото-а. д. с. солнечного элемента составляет $\sim 0,5$ В, максимальная сила фототока $I_{\phi} \approx 50$ А при рабочей площади элемента $S \sim 2$ см². Для



Рис. 94. Современная солнечная батарея. Выпуклые отражатели концентрируют солнечное излучение на элементы, основу которых составляет контакт GaAs с тройным полупроводниковым соединением GaAlAs

получения больших напряжений (10—20 В) элементы соединяют последовательно, а для увеличения рабочих токов — параллельно. Так получают солнечные батареи.

На рис. 95 показана космическая станция, развернувшая «крылья» солнечных батарей (основного источника питания аппаратуры станции) навстречу Солнцу. Этот снимок, наверное, знакомый каждому, напоминает о самой модной профессии солнечных элементов. Однако уже сейчас солнечные батареи нашли десятки других, менее романтических, но очень нужных применений.

В пустынях Туркмении и Казахстана солнечные батареи приводят в действие насосы, выкачивающие воду из глубоких водоносных слоев. Солнечные батареи обогревают и освещают дома, вращают двигатели, подзаряжают аккумуляторные батареи.

Выпрямительные диоды. Практически вся потребляемая в мире электроэнергия вырабатывается сегодня турбогенераторами электростанций в форме переменного напряжения так называемой промышленной частоты (50

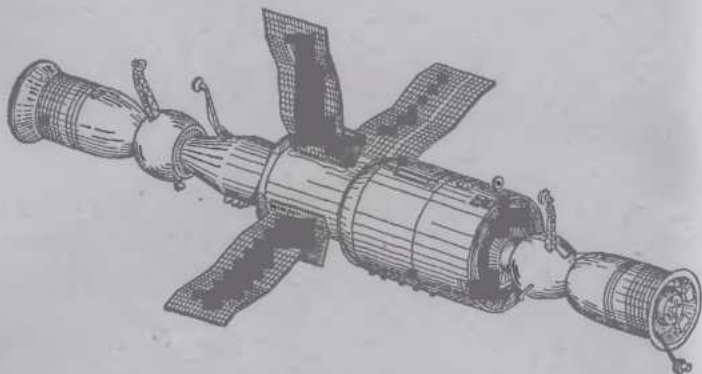


Рис. 95. Солнечные батареи — источник питания космической станции

или 60 Гц). Но во многих случаях использовать электроэнергию в такой форме нельзя. Многие приборы, машины, установки и даже целые отрасли народного хозяйства нуждаются в питании постоянным током.

Постоянным током питаются все виды транспорта на электрической тяге: трамваи, троллейбусы, электропоезда (электрички). Чтобы обеспечить работу электротранспорта, необходимо преобразовать переменное напряжение в постоянное — *выпрямить* токи в тысячи ампер при напряжении в сотни и тысячи вольт. В постоянном токе нуждаются мощные электролизные установки, потребляющие при небольшом напряжении, в несколько вольт, токи в сотни тысяч и миллионы ампер. Мощные современные ЭВМ также питаются постоянным током в тысячи и десятки тысяч ампер при напряжении в несколько вольт. С другой стороны, в радиотехнике и технике СВЧ часто возникает необходимость в выпрямлении сигналов с напряжением в миллионные и даже в миллиардные доли вольта с частотой от долей герца (в сейсмологии)

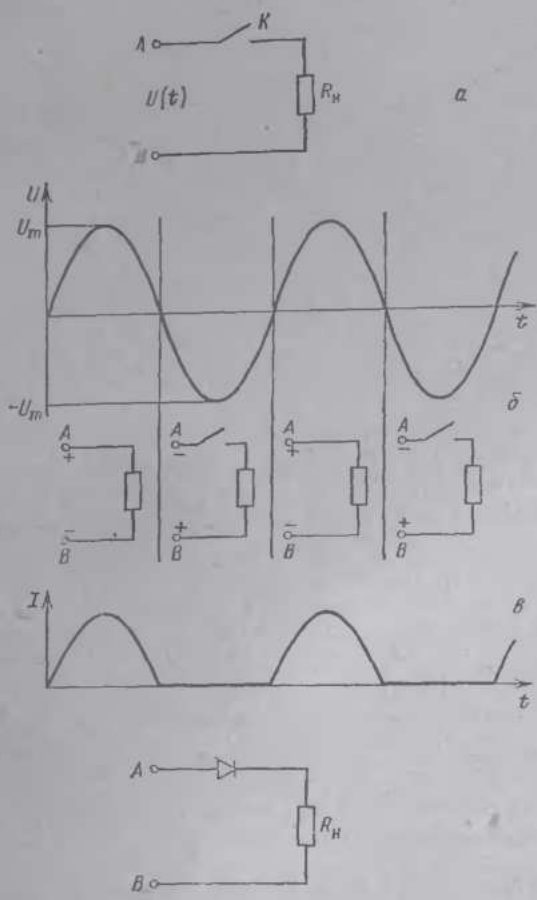


Рис. 96. Принципиальная схема выпрямления переменного тока: а — последовательно с нагрузкой R_n включен ключ K ; б — когда напряжение U на входе схемы положительно, ключ K замкнут. При отрицательной полярности напряжения — разомкнут; в — ток I через нагрузку течет только в одном направлении. В течение положительного полупериода входного напряжения $I = U/R_n$. В течение отрицательного — ток равен нулю; г — диод с $p-n$ переходом играет роль ключа в практических схемах выпрямителей

до тысяч миллиардов герц (в технике СВЧ). Со всеми этими сложными задачами успешно справляются выпрямительные диоды с $p - n$ переходом.

Принцип преобразования переменного тока в постоянный (выпрямления) поясняется рис. 96, а. Последовательно с нагрузкой R_n (так мы будем впредь называть любой объект, нуждающийся в питании постоянным током, — аккумулятор, ЭВМ, установку электролиза, троллейбусную сеть и т. д.) включается ключ K . На вход схемы, между точками A и B , подается переменное напряжение $U(t)$.

Когда напряжение на входе положительно, будем держать ключ замкнутым. Когда напряжение на входе сменит знак — разомкнем ключ (рис. 96, б). В результате зависимость тока, протекающего через нагрузку R_n , от времени будет иметь вид, показанный на рис. 96, в. Когда ключ замкнут, по цепи протекает ток $I = U/R_n$. Когда ключ разомкнут, ток в цепи равен нулю, а все напряжение приложено к месту разрыва цепи — ключу. Как говорят, ключ *блокирует* приложенное напряжение.

Ток, зависимость от времени которого показана на рис. 96, в, уже не переменный. Он всегда течет только в одном направлении — от точки A к точке B . Правда, назвать его постоянным тоже трудно — амплитуда его очень сильно меняется, пульсирует, со временем. Часто такой ток так и называют — пульсирующим. Для некоторых целей, например для электролиза или зарядки аккумуляторных батарей, такой пульсирующий ток вполне пригоден. Если же слишком большие пульсации недопустимы (например, при питании ЭВМ, телевизоров, радиоприемников), они могут быть сглажены с помощью одной из многочисленных схем так называемых фильтров. Простейшим фильтром является конденсатор, включенный параллельно сопротивлению R_n . Работа такого фильтра подробно проанализирована в школьном учебнике (Физика, 10 класс).

Ну вот, а теперь если вместо гипотетического ключа K включить диод с $p - n$ переходом (рис. 96, г), то получится реальная и часто употребляемая на практике схема выпрямителя. Когда напряжение на диоде соответствует прямому смещению, его сопротивление очень мало. По цепи протекает ток $I \approx U/R_n$. Когда напряжение на входе меняет знак и становится отрицательным, диод оказывается смещенным в обратном направлении, его сопротивление очень велико. Диод блокирует приложенное напряжение, и ток, протекающий в цепи, очень мал.

Посмотрим теперь, каким требованиям должен удовлетворять диод с $p - n$ переходом, чтобы успешно работать в качестве выпрямителя,

Понятно, что прежде всего такой диод должен пропускать в прямом направлении такой же ток, какой нужен для питания сопротивления нагрузки R_n . Если этот ток равен нескольким миллиамперам, как при питании трапзисторного приемника или микрокалькулятора, удовлетворить этому требованию нетрудно. Но, как мы видели, необходимая сила тока может составлять сотни, тысячи и даже сотни тысяч ампер.

Чем больше сила тока, который должен выпрямлять диод, тем большую площадь он должен иметь. Это требование вытекает из того простого факта, что диод в процессе работы нагревается. Энергия, выделяющаяся в в диоде в виде тепла тем, естественно, больше, чем больший ток течет через диод. А энергия, которую диод может отдать в окружающее пространство, пропорциональна его площади.

Чем больший ток выпрямляет диод, тем большую энергию необходимо отводить в окружающее пространство и тем большую площадь соответственно должен иметь диод.

Для кремниевых $p - n$ переходов (из кремния изготавливается подавляющее большинство выпрямительных диодов) допустимая плотность прямого тока, рассчитанная из условий отвода тепла, составляет $\sim 50 - 100 \text{ А/см}^2$. Площадь диодов, предназначенных для выпрямления тока в *сильноточных* устройствах, составляет десятки см^2 . Соответственно такие диоды способны выпрямлять токи в тысячи ампер (рис. 97). При необходимости выпрямлять еще бóльшие токи, диоды соединяют между собой параллельно.

Второе требование, которому должен обязательно удовлетворять выпрямительный диод, — способность выдерживать, не пробиваясь, максимальное значение выпрямляемого напряжения U_m . Как мы знаем, чтобы удовлетворить этому требованию, во-первых, концентрация примесей в слаболегированной области диода, база, должна быть достаточно мала (см. формулу (40)). А, во-вторых, длина базы W должна быть достаточно велика, так, чтобы при максимальном обратном напряжении на диоде U_m она была больше (или, по крайней мере, равна) длине области объемного заряда W_1 (формула (39), рис. 62, б, в).

Итак, для того чтобы работать как разомкнутый ключ и выдерживать большое обратное напряжение, диод должен иметь длинную слаболегированную базу. Но

длинная база, изготовленная из полупроводника с малой концентрацией электронов, будет иметь большое омическое сопротивление! На этом большом сопротивлении в ту часть периода, когда диод должен работать как замкнутый ключ и пропускать через себя большой ток,

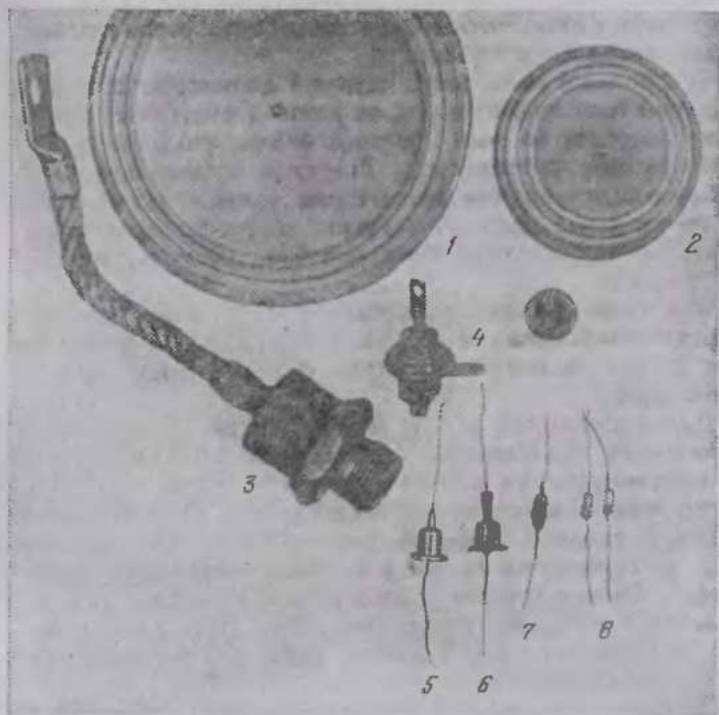


Рис. 97. Выпрямительные диоды, рассчитанные на выпрямление разных токов. 1 — Рабочая площадь этого диода (ДВ2500) составляет $\sim 50 \text{ см}^2$. Он может использоваться для выпрямления токов $I \sim 2500 \text{ А}$. 2 — Выпрямительный диод на ток 1000 А . 3 — Диод ВЛ50: $I \sim 50 \text{ А}$. 4 — Диод КД203: $I \sim 10 \text{ А}$. 5 — Д226: $I \sim 300 \text{ мА}$, 6 — Д208: $I \sim 100 \text{ мА}$. 7 — Д223: $I \sim 50 \text{ мА}$, 8 — Д9: $I \sim 15 + 20 \text{ мА}$

будет падать значительное напряжение. А вот это совершенно недопустимо.

Прежде всего, напряжение, которое падает на ключе (диоде), уменьшает полезное напряжение на сопротивлении нагрузки R_n . И, кроме того, при большом падении напряжения и большом токе на ключе (диоде) выделяется

в виде тепла большая энергия. Как мы знаем, слишком большой перегрев диодов недопустим. Если температура диода превышает определенную для данного материала величину, диод выходит из строя.

Давайте убедимся, что для высоковольтных приборов противоречие, на которое мы обратили внимание, достаточно серьезно. Обсуждая формулу (40), мы уже установили, что для того чтобы диод был способен блокировать напряжение $U_m \approx 3000$ В, концентрация легирующей примеси в базе не должна превышать $N \sim 10^{14}$ см⁻³. Подставив в формулу (39) значения $U_0 = 3 \cdot 10^3$ В и $N = 10^{20}$ м⁻³, получим, что длина базы должна превышать значение $W_1 \approx 2 \cdot 10^{-4}$ м = $2 \cdot 10^{-2}$ см. Как уже упоминалось, плотность прямого тока в выпрямительных диодах составляет $j \approx 50 \div 100$ А/см².

Ну, а теперь — простой подсчет, основанный на знании закона Ома. Как мы знаем, $j = en\bar{v} = en\mu E$. Пусть выпрямительный диод изготовлен из кремния, база диода *n*-типа. Подвижность электронов $\mu_n = 0,13$ м²/(В·с) (табл. 2). Концентрация электронов *n* равна концентрации легирующей примеси *N*. Напряженность поля в базе диода $E = j/eN\mu = (2,5 \div 5) \cdot 10^3$ В/см. А падение напряжения на базе диода $U_6 = EW = 50-100$ В!

Хорош, нечего сказать, «ключ», на котором в замкнутом состоянии падают 100 вольт! Разумеется, практически такой ключ совершенно никуда не годится.

Кажется, что положение безвыходное. Хочешь выпрямлять высокое переменное напряжение — мирись с чудовищными падениями напряжения на диоде в прямом направлении. А между тем выход есть. Больше того, мы уже на самом деле знаем, в чем он состоит: в использовании явления инжекции.

Взглянем еще раз на рис. 86. Видно, что при высоком уровне инжекции концентрация носителей в базе вблизи *p-n* перехода может быть гораздо выше, чем равновесная концентрация носителей n_0 . Чем выше концентрация носителей, тем меньше сопротивление базы и, следовательно, тем меньшее напряжение будет падать на базе при той же плотности тока. Нельзя ли так сконструировать диод, чтобы во время протекания прямого тока вся база диода заполнялась неравновесными носителями, инжектированными из эмиттера, и сопротивление базы резко уменьшалось?

Можно. И легко понять, что для этого необходимо сделать. Нужно, чтобы база диода была не слишком длин-

ной, так, чтобы концентрация дырок, инжектированных эмиттером, не слишком сильно уменьшалась за счет рекомбинации к противоположному концу базы.

Но длиной базы W мы произвольно задаваться не можем! Она, как мы видели, определяется максимальным напряжением U_m , которое должно блокироваться диодом. Для больших значений U_m длина базы также должна быть велика.

Правильно. Длина базы, действительно, определяется блокируемым напряжением. Но (как видно, в частности, из рис. 86) то, насколько спадает в глубину базы концентрация неравновесных носителей, зависит не от самого значения W , а от отношения W/L_p . От того, сколько диффузионных длин дырок L_p укладывается на длине базы W . Даже при большой величине W , если отношение W/L_p невелико, спад концентрации не будет особенно заметен.

Поэтому чем больше блокируемое напряжение, чем длиннее база, тем больше должна быть диффузионная длина дырок L_p , если мы хотим использовать явление инжекции и получать приемлемые прямые падения напряжения при больших плотностях прямого тока.

Считая, что диффузионная длина дырок должна быть порядка длины базы ($W \approx L_p$), и используя формулу (26), легко установить, что для кремниевого диода с длиной базы $W = 200$ мкм время жизни носителей в базе должно быть

$$\tau_p \geq \frac{L_p^2}{D_p} \approx \frac{W^2}{D_p} \geq 4 \cdot 10^{-5} \text{ с.}$$

(Значение D_p для кремния ≈ 10 см²/с.)

Как мы знаем, увеличение времени жизни носителей достигается тщательной очисткой полупроводника от глубоких примесей (гл. 2). Кремний технологи умеют очищать от примесей очень хорошо, и материал с временем жизни $\tau_p \geq 40$ мкс вполне доступен.

Более того, постаравшись, можно получить кремний с временем жизни τ в несколько сотен микросекунд или даже несколько миллисекунд. Значит ли это, что следует стремиться изготовить на основе такого кремния выпрямительные диоды на напряжение в десятки тысяч вольт?

Не будем торопиться с ответом.

Любой ключ, используемый для выпрямления переменного тока, характеризуется еще одним параметром, о котором мы на время предпочли забыть, — быстродействием. Переходить из состояния «включено» в состояние «вы-

«включено» и обратно ключ должен за время, гораздо меньшее, чем период колебаний тока, который он выпрямляет.

Промышленной частоте 50 Гц соответствует период колебаний $T = 1/f = 20$ мс. Если принять, что ключ должен включаться и выключаться за время порядка $1/100$ периода колебаний, то от выпрямительного диода следует потребовать, чтобы он успевал переходить из включенного в выключенное состояние и обратно за время, не превышающее ~ 200 мкс.

Чем же определяется время включения и выключения полупроводникового диода? Да как раз временем жизни неосновных носителей в базе τ !

В самом деле, переход из состояния «выключено» в состояние «включено» соответствует переходу из состояния, когда в базе нет неосновных носителей к состоянию, когда инжектированные из эмиттера носители заполняют всю базу. Когда обратное напряжение на диоде сменяется прямым, носители (дырки) начинают двигаться от эмиттера в глубь базы за счет диффузии и заполняют базу за время $t \sim W^2/D_p$ (см. формулу (25)). Но так как $W^2 \sim L_D^2$, то

$$t \sim \frac{L_D^2}{D_p} \sim \tau_p.$$

Время включения диода оказывается порядка времени жизни неосновных носителей в базе.

Когда же, наоборот, напряжение на диоде с прямого меняется на обратное, то для перехода диода в состояние «выключено» необходимо, чтобы дырки, заполнившие при прямом напряжении базу, исчезли, рекомбинировали. Рекомбинируют носители, по определению, с временем жизни τ_p . Таким образом, и время выключения по порядку величины оказывается равным времени жизни неосновных носителей.

Теперь понятно, почему чем больше напряжение, на которое рассчитан выпрямительный диод, тем меньше предельная частота колебаний, которые он способен эффективно выпрямлять. Больше напряжение — длиннее база. Длиннее база — больше должно быть время жизни носителей τ . Больше τ — медленнее происходит включение и выключение диода. Предельное напряжение диодов, предназначенных для работы в цепях промышленной частоты, составляет приблизительно 10^4 В.

А что же делать, если необходимо выпрямить большее напряжение, скажем, в сотни тысяч или миллионы вольт.

В этом случае диоды соединяют последовательно в так называемые «выпрямительные столбы».

Мы подробно разобрали физические принципы работы высоковольтных диодов — великанов в семействе выпрямительных полупроводниковых диодов. На другом «полюсе» этого обширного семейства располагаются так называемые *детекторные* диоды.

Термин «детектор» происходит от латинского *detectio* (открытие, обнаружение). Так называют выпрямительные диоды, предназначенные для обнаружения сигналов, передаваемых с помощью радиоволн. Их назначение — выпрямлять переменное напряжение очень маленькой амплитуды, но очень высокой (в сотни миллионов герц) частоты.

Предельное быстродействие детекторных диодов с $p - n$ переходом также определяется временем жизни неосновных носителей. Чем оно меньше, тем выше предельная частота, на которой может эффективно работать детектор.

Казалось бы, «ломать — не строить». Уж загрязнить-то полупроводник (уменьшить время жизни носителей в нем) нетрудно. Вообще говоря, это сущая правда. Но если речь идет о получении предельно малых времен жизни, приходится считаться с тем, что при определенной концентрации «грязи» (дефектов, примесей) полупроводник или просто разрушается, или в большой степени утрачивает свои полезные свойства. Поэтому умение уменьшить время жизни, сохранив при этом в неприкосновенности остальные свойства материала, — высокое искусство, в котором технологи постоянно совершенствуются. В настоящее время на основе кремния, арсенида галлия, фосфида индия умеют получать диоды с временем жизни $\tau \sim 10^{-10}$ с. Предельная частота $f_{пр}$, на которой работают детекторные диоды с $p - n$ переходом, составляет приблизительно $1/10$ величины, обратной значению τ (вспомним, что величина τ должна быть не порядка, а значительно меньше, чем период колебаний). Соответственно $f_{пр} \sim \frac{1}{10} \cdot \frac{1}{\tau} \sim 10^9$ Гц.

Вторая похвала терпению

Ну вот, мы и преодолели самый высокий в этой книге барьер. Впереди, правда, знакомство с самым важным и самым интересным полупроводниковым

прибором — транзистором. Но это знакомство не потребует никаких новых физических идей. Мы уже овладели всеми представлениями, необходимыми для того, чтобы понять, как устроен и как работает транзистор.

Какие чувства испытывает путешественник, взошедший на высокий перевал?

В прекрасном сказе Бажова «Васина гора» мальчик спрашивает у старика:

«Дедо, я вот что приметил. Подымется человек на нашу гору... и непременно оглянется, а дальше разница выходит. Один будто и силы небольшой, и на возрасте, а пойдет вперед веселехонек. А другой — случается, по виду могутный — вдруг голову повесит и под гору плетется, как ушиб его кто. Почему такое?»

Дед посоветовал мальчонке самому спросить у прохожих.

«Только вот дивно — ответ тоже на два конца. Те, кто идет дальше веселым, говорят:

— Экую гору одолел, дальше и бояться нечего. Все одолею. Поэтому и весело мне.

Другие стонут:

— Вон на какую гору взобрался, самая бы пора отдохнуть, а еще идти надо.

Эти вот и плетутся, как связанные, смотреть на них тошно...»

Привет дошедшим оптимистам!



Moralité

Между частями полупроводникового кристалла, одна из которых легирована донорной, а другая — акцепторной примесью, возникает энергетический (потенциальный) барьер — $p-n$ переход. В отсутствие внешнего напряжения высота этого барьера приблизительно равняется ширине запрещенной зоны полупроводника \mathcal{E}_g . Внешнее смещение, «плюс» которого приложен к n -области перехода, а «минус» — к p -области (запорное или обратное напряжение), увеличивает высоту барьера. Через $p-n$ переход при этом течет очень небольшой обратный ток. Диод с $p-n$ переходом представляет собой очень большое сопротивление. При прямом смещении («плюс» внешнего напряжения приложен к p -, а «минус» — к n -области диода) высота барьера понижается. Плотность прямого тока очень резко увеличивается с ростом смеще-

ния и даже при относительно небольших напряжениях может достигать очень высоких значений. Диод с $p - n$ переходом представляет собой очень маленькое сопротивление.

Свойства $p - n$ перехода позволяют использовать диоды с $p - n$ переходом в качестве выпрямителей, стабилизаторов, управляемых емкостей (варикапов), преобразователей солнечной энергии, фотодиодов и многих других приборов.

ТРАНЗИСТОРЫ



— Ура! Я нашел нечто действительно великое!

Полупроводниковый триод был изобретен в 1948 г. американскими физиками Дж. Бардином, У. Браттейном и У. Шокли. А в 1956 г. за это изобретение они были награждены почетнейшей международной премией — Нобелевской.

Несколько лет детище американских ученых именовалось просто *транзистором*. Но в 1952 г. один из лауреатов — Уильям Шокли разработал еще один тип полупроводникового триода, по принципу работы сильно отличающийся от первого. Транзистор, изобретенный Шокли, был назван *полевым транзистором*. А за изобретением Бардина, Браттейна и Шокли утвердилось название *биполярный транзистор*.

Хотя физические принципы работы полевого и биполярного транзисторов совершенно разные, их основное назначение, из которого вытекают почти бесчисленные практические применения, одно и то же — усиление электрических сигналов. Сигнал, поступивший на вход транзистора, вызывает появление на выходе гораздо более мощного сигнала.

В гл. 6 мы познакомимся с принципами работы биполярного транзистора. В гл. 7 изложены физические основы работы полевого транзистора. А в гл. 8 описаны некоторые применения этих замечательных приборов.

Глава 6

БИПОЛЯРНЫЕ ТРАНЗИСТОРЫ

От малых причин бывают весьма важные последствия...

Козьма Прутков

Устройство биполярного транзистора схематически показано на рис. 98. Сложностью устройство, как



Рис. 98. Структура биполярного $p-n-p$ транзистора

структуры — эмиттер и база, знакомы нам из предыдущей главы. Слово коллектор происходит от латинского collector (сборщик).

Точно так же, разместив узкую полосу полупроводника p -типа между двумя n -областями, можно получить $n-p-n$ транзистор. Физические принципы работы и $p-n-p$ и $n-p-n$ структур совершенно одинаковы. Тот, кто вместе с нами разберется в работе $p-n-p$ транзистора, сможет, если захочет, без особенного труда самостоятельно проанализировать работу $n-p-n$ транзисторной структуры.

Показанную на рис. 98 структуру можно описать и по-другому — как уже знакомый нам $p-n$ переход, к которому присоединена еще одна p -область. Или как два очень близко расположенных $p-n$ перехода, имеющих общую базу. Каким же образом эта нехитрая структура умудряется выполнять основное свое назначение — усиливать электрические сигналы?

Принцип работы биполярного транзистора

Для того чтобы транзистор мог выполнять свои задачи, к $p-n$ переходам, образующим транзисторную структуру, должны быть приложены определенные напряжения.

Коллекторный переход ($p-n$ переход между базой и коллектором) должен быть смещен в обратном направлении, т. е. к нему должно быть приложено обратное напряжение. *Эмиттерный переход* транзистора ($p-n$ переход между эмиттером и базой) должен быть смещен в прямом направлении.

Появление нужных напряжений на переходах достигается включением транзистора во внешнюю цепь, содержащую источник (или источники) питания, резисторы, конденсаторы и т. д. Существует довольно много разновидностей таких схем; некоторые из них мы ниже обсудим. Но пока, чтобы не отвлекаться, мы сосредоточимся не на способах создания нужных напряжений, а на том, каким образом при выполнении необходимых условий транзистор приобретает способность усиливать сигналы.

Чтобы уяснить принципы работы транзистора, представим себе для начала, что база транзистора не узкая,

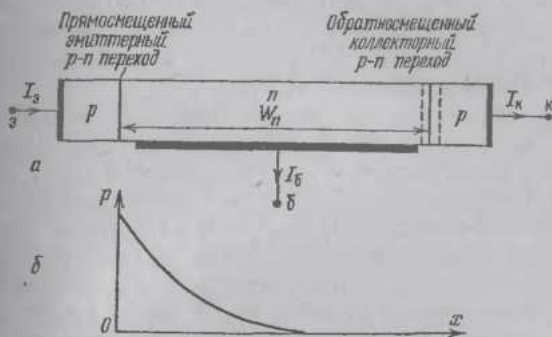


Рис. 99. Схема работы $p-n-p$ структуры с длинной базой. Длина базы W_n значительно больше диффузионной длины дырок L_p в ней ($W_n/L_p \gg 1$): а — $p-n-p$ структура с длинной базой; б — распределение дырок в базе (ср. с рис. 84). Вошедшие из эмиттера дырки рекомбинируют, не достигая коллекторного $p-n$ перехода

а длинная (рис. 99). Кстати, самое время определить, что, собственно говоря, значит «узкая» и «длинная». По сравнению с чем?

Длину n -базы W_n транзистора всегда оценивают по отношению ее к диффузионной длине дырок в базе L_p . В реальных транзисторах отношение W_n/L_p всегда гораздо меньше единицы (база узкая).

Но мы договорились сначала обсудить работу структуры с длинной базой, в которой отношение $W_n/L_p \gg 1$ (рис. 99). О такой структуре мы знаем все. Это просто два диода, один из которых, левый, включен в прямом

направлении, второй, правый, — в обратном. То обстоятельство, что у них общая база, ровно ничего в их работе не меняет. Тем не менее еще раз акцентируем внимание на тех особенностях работы переходов, которые крайне важны при анализе транзистора.

Через обратносмещенный коллекторный переход структуры протекает небольшой обратный ток I_k . Этот ток образован, во-первых, электронами и дырками, родившимися в области объемного заряда (генерационный ток). А во-вторых, неосновными носителями — дырками n -базы и электронами p -коллектора, имевшими неосторожность слишком близко подойти к области перехода (диффузионный ток). Еще раз вспомним (это очень важно!), что любая дырка, родившаяся в n -базе на расстоянии от p — n перехода, меньшем, чем диффузионная длина дырок L_p , имеет серьезные шансы оказаться в пределах области объемного заряда. Если это произойдет, электрическое поле обратносмещенного коллекторного перехода немедленно подхватит такую дырку и вышвырнет ее из базы в коллектор (рис. 56, а *)).

Теперь обратимся к левому, прямосмещенному эмиттерному переходу. Эмиттер транзисторной структуры всегда легирован значительно сильнее, чем база. Поэтому механизм протекания тока через эмиттерный переход структуры в точности такой же, как и подробно проанализированный нами раньше механизм протекания прямого тока через резко несимметричный p — n переход. Как мы помним, протекание прямого тока через такой переход можно описать по-разному. Для анализа работы транзистора удобнее всего стартовать, как мы это делали и раньше, от базового электрода структуры.

Протекание тока I_b через базовый электрод структуры означает, что каждую секунду в базу (из внешнего источника, не показанного на рисунке) входят $N = I_b/e$ электронов. Куда они деваются? В стационарном состоянии накапливаться в базе они не могут, так как стационарное состояние соответствует ситуации, когда ничего, в том числе и концентрация носителей в базе, со временем не изменяется.

Уйти в коллектор электроны не могут, так как сильное поле обратносмещенного коллекторного перехода соз-

*) На рис. 56, а n -область расположена справа, а на рис. 99 — слева. Чтобы не путаться, можно перевернуть с. 150 и рассмотреть рис. 56, а на просвет.

дает для них высокий потенциальный барьер и энергично вталкивает их обратно в базу (рис. 62). Уйти в высоколегированный эмиттер электроны практически тоже не в состоянии (см. с. 209). Что же им остается? Погибнуть в базе, рекомбинируя с дырками, входящими в базу из эмиттера через прямосмешанный эмиттерный $p - n$ переход!

Сколько электронов входят в единицу времени через базовый электрод, столько же и дырок поступает в базу за то же время через $p - n$ переход эмиттер — база. Образно выражаясь, можно представить себе, что поступление в базу отрицательно заряженных электронов провоцирует положительно заряженные дырки выйти из эмиттера и, диффундируя через базу, ценою собственной жизни убить их.

В структуре с длинной базой показанной на рис. 99, б, дырки практически полностью рекомбинируют на расстоянии, равном нескольким диффузионным длинам L_p , и до коллекторного перехода не доходят.

А что происходит в реальной транзисторной структуре с тонкой базой?

Рассматривая процессы в реальной транзисторной структуре с тонкой базой ($W_n/L_p < 1$), важно, прежде всего, ясно представить себе, что эмиттерный и коллекторный переходы в такой структуре нельзя рассматривать как два изолированных $p - n$ перехода.

Почему?

Потому, что, как мы только что выяснили, поступление тока в базовый электрод структуры немедленно вызывает инжекцию дырок из эмиттера в базу. С другой стороны, как мы знаем, любая дырка, появившаяся в базе на расстоянии от коллекторного перехода, меньшем, чем величина L_p , имеет серьезные шансы быть выброшенной полем перехода из базы в коллектор. Но в транзисторной структуре с узкой базой ($W_n/L_p < 1$) каждая дырка в базе находится от коллекторного перехода на расстоянии, меньшем, чем L_p !

Совершенно ясно, что часть дырок, поступивших в базу из эмиттера, непременно попадет в область заряда коллекторного перехода, будет подхвачена полем и выброшена в коллектор. Это, в свою очередь, означает, что коллекторный ток I_K будет теперь определяться не только (и даже не столько) током обратосмещенного коллекторного перехода, но и током дырок, переходящих из базы в коллектор.

Ток дырок, захватываемых коллектором, зависит от концентрации дырок в базе, определяемой темпом поступления дырок в базу, т. е. от тока эмиттера I_3 . А этот ток, в свою очередь, — от величины базового тока I_6 .

Токи, протекающие через все три электрода структуры, оказываются, таким образом, зависимыми друг от друга. Наша задача — установить эту связь.

Усиление по току. Пусть в структуре с узкой базой задан ток базы I_6 . Каждую секунду в базу поступают $N = I_6/e$ электронов. Провоцируемые этими электронами, из эмиттера выходят дырки, чтобы убить их и умереть. Будь база длинной, сколько электронов входило бы в базу, столько дырок поставил бы и эмиттер. Но в структуре с узкой базой дело обстоит не так. Часть электронов, поступивших в базу, будет захвачена полем коллектора и выброшена из базы, не успев прорекомбинировать с электронами.

А между тем все электроны, поступающие в базу, непременно должны прорекомбинировать. Уйти из базы они не могут и накапливаться не должны. Это означает, что эмиттер должен вбрасывать в базу дырок больше, чем поступает в нее электронов. Он должен вбрасывать их столько, чтобы, несмотря на отток дырок в коллектор, обеспечить гибель всех электронов, входящих в базу.

Чем тоньше база, тем большую долю входящих в базу дырок захватывает коллектор. Если база тонкая ($W_n/L_p \ll \ll 1$), то доля дырок α , захватываемая коллектором, описывается простым выражением

$$\alpha \approx 1 - W_n^2/2L_p^2. \quad (52)$$

Чем меньше отношение W_n/L_p , чем ближе коллекторный переход к эмиттерному, тем, естественно, больше доля дырок, которые, не успев рекомбинировать, проходят через базу и выбрасываются в коллектор.

В зависимости от типа и назначения отношение W_n/L_p в транзисторах лежит обычно в пределах от $\sim 0,5$ до $\sim 0,05$. Таким образом, значение коэффициента α , который называется *коэффициентом переноса*, для различных транзисторов может лежать в пределах от $\sim 0,9$ до $\sim 0,999$.

Название «коэффициент переноса» для величины α звучит довольно естественно. В самом деле, ведь она и определяет долю дырок, которые переносятся благодаря процессу диффузии через базу от эмиттера к коллектору.

Итак, коллектор перехватывает (собирает) львиную долю дырок, поступающих из эмиттера, — от 0,9 до 0,999.

Рекомбинирует с электронами, поступающими в базу, лишь малая часть дырок — от 0,1 до 0,001 (рис. 100). Но это заключение, которое в такой формулировке звучит как-то пессимистически, можно выразить и другими словами, звучащими, как победный гимн, заслуживший Нобелевскую премию!

Ток, поступающий в базу транзистора, вызывает появление в эмиттерной и коллекторной цепях тока, в десятки, сотни и даже тысячи раз большего.

Таким образом, если подавать токовый сигнал, нуждающийся в усилении, в базу транзистора, а выходной сигнал регистрировать в коллекторной или эмиттерной цепи, то сигнал окажется усиленным в десятки, сотни и даже тысячи раз.

Коэффициент усиления транзистора по току β определяется отношением тока эмиттера к току базы $\beta = I_E/I_B$, и, естественно, тем больше, чем тоньше база, т. е. чем меньше отношение W_n/L_p .

Определим теперь связь между токами, протекающими через электроды транзистора. Сделать это довольно просто.

Коллекторный ток I_K складывается, как мы знаем, из двух компонентов — обратного тока перехода и тока дырок, поступающих в коллектор из эмиттера. В подавляющем большинстве случаев второй компонент во много раз больше. Тогда

$$I_K \approx \alpha I_E. \quad (53)$$

С другой стороны,

$$I_E = I_B + I_K. \quad (54)$$

Формула (54) просто отражает факт, для нас вполне очевидный, что дырки, вышедшие из эмиттера, или рекомбинируют в базе (I_B), или уходят в коллектор (I_K).

Из (53) и (54) следует

$$I_E = \frac{I_B}{1 - \alpha} = \beta I_B, \quad (55)$$

где коэффициент усиления транзистора по току β

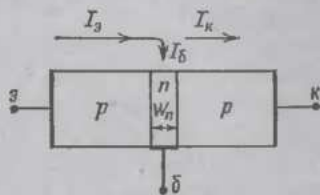


Рис. 100. В транзисторной структуре с узкой базой большая часть дырок, вошедших из эмиттера, проходит в коллектор, образуя коллекторный ток I_K , и лишь малая часть тока эмиттера ответвляется в базу

равняется

$$\beta = \frac{1}{1-\alpha}. \quad (56)$$

При $\alpha = 0,9$ $\beta = 10$; значению $\alpha = 0,999$ соответствует $\beta = 1000$.

Рассмотренный механизм усиления транзистором тока лежит в основе работы биполярного транзистора.

Кстати, почему биполярного? Как мы теперь понимаем, в работе транзистора одинаково важную роль играют носители обоих знаков — и электроны, и дырки. Электроны поступают в n -базу $p-n-p$ транзистора и вызывают в эмиттерной и коллекторной цепях появление дырочного тока (гораздо большего по величине *).

Притча о главном и деталях. Много лет назад в некоем городе два мудреца учили тому, как нужно жить в этом мире. Один мудрец был человеком строгим, а другой — добрым.

Всякий, кто хоть раз задумывался над тем, как нужно жить в этом мире, знает, что думать над этим вопросом можно (и нужно) всю жизнь. И действительно, много лет постигали ученики тайны учителей-мудрецов.

И вот однажды к учителю строгому пришел чужестранец и сказал: «Учитель, я жажду мудрости, света и истины. Я буду самым усердным твоим учеником и готов учиться всю жизнь. Но я прошу тебя — изложи мне главное в твоём учении за то время, которое я способен простоять на одной ноге».

И строгий учитель прогнал чужестранца.

Чужестранец пошел тогда к учителю доброму и попросил о том же. «Становись на одну ногу», — усмехнулся учитель. И когда чужестранец утвердился, как цапля, на одной ноге, мудрец сказал: «Не делай другому того, что ты не хочешь, чтобы делали тебе. Это главное. Все остальное — комментарий. Иди и изучай!»

Мы изучили главное в работе биполярного транзистора — принцип усиления по току. А теперь — некоторые, самые важные из бесчисленных возможных комментариев.

Быстродействие транзистора. Быстродействие транзистора, наряду с коэффициентом усиления по току β , является одной из самых важных его характеристик.

*) В транзисторной структуре $n-p-n$ типа, работу которой, как мы надеемся, вы проанализируете самостоятельно, дырки, поступающие в p -базу, вызывают появление в эмиттерной и коллекторной цепях во много раз большего электронного тока.

Быстродействие транзистора можно характеризовать или предельной частотой $f_{пр}$, которую способна усиливать или генерировать транзисторная структура, или так называемым *временем установления* t_0 . Под временем установления имеется в виду следующее. Пусть в транзистор мгновенно подан входной сигнал. Например, входной ток мгновенно возрос на величину $\Delta I_{вх}$. Если входной сигнал будет подаваться достаточно долго, на выходе установится соответствующий выходной сигнал: выходной ток возрастет на величину $\Delta I_{вых}$. Но произойдет это не мгновенно. Время t_0 , которое происходит от момента подачи входного сигнала до момента установления выходного, и называется (естественно) *временем установления*. Чем меньше время установления t_0 , тем выше частота $f_{пр} : f_{пр} \sim 1/t_0$.

Действительно, представим себе, что мы мгновенно подали входной сигнал $\Delta I_{вх}$ и поддерживаем его на входе в течение времени $t > t_0$ (рис. 101, а). Тогда на выходе успеет установиться соответствующим образом усиленный выходной сигнал $\Delta I_{вых}$ (рис. 101, б). По прошествии времени t выключим выходной сигнал и опять подождем время $t > t_0$ (рис. 101, а). Выходной сигнал успеет отреагировать на выключение входного и полностью исчезнет (рис. 101, б). Понятно, что если периодически включать и выключать входной сигнал с периодом $T \geq 2t_0$, т. е. с частотой $f \leq 1/2t_0$, то такой сигнал транзистор успеет усилить.

Если же частота сигнала $f \geq 1/2t_0$ (рис. 101, в, г), то выходной сигнал не успевает следовать за входным. Реагируя на входной сигнал, выходной ток $I_{вых}$ начинает нарастать. Не успеет он как следует нарасти — входной сигнал требует, чтобы выходной ток начал уменьшаться. В результате при $f \geq 1/2t_0$ выходной ток «дрожит» с небольшой амплитудой. И амплитуда эта тем меньше, чем больше частота. Высокочастотный сигнал с частотой $f \gg 1/t_0$ транзистор усиливать не успевает.

Время установления выходного сигнала t_0 зависит от физических параметров транзистора и от *схемы включения транзистора*, т. е. от того, на какой электрод транзистора подается входной сигнал и с какого снимается выходной.

Рассмотрим сначала, чем определяется время установления t_0 в случае, когда входной сигнал подается в эмиттер транзистора, а выходной — снимается с коллектора (рис. 102).

На первый взгляд может показаться, что изучать работу транзистора в таком режиме — пустая трата времени. О каком усилении тут может идти речь? Мы ведь знаем, что ток коллектора меньше, чем ток эмиттера в α раз (формула (53)). Правда, величина α

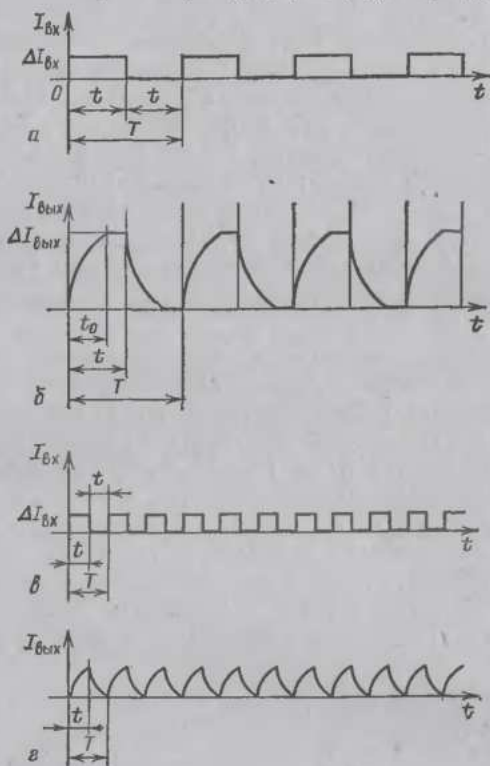


Рис. 101. Если период входного сигнала T значительно больше, чем время установления выходного сигнала t_0 , то сигнал на выходе транзистора $\Delta I_{вх}$ успевает достигать стационарного значения (а, б). Если период входного сигнала T значительно меньше, чем t_0 , сигнал на выходе транзистора не успевает устанавливаться (в, г). Максимальная частота сигнала, которую способен усиливать транзистор, по порядку величины равна $f_{пр} \sim 1/t_0$

не слишком сильно отличается от единицы ($0,9 \leq \alpha \leq 0,999$), но, во всяком случае, меньше единицы! Любое приращение тока в эмиттере $\Delta I_э$ «отзовется» приращением тока в коллекторе $\Delta I_к = \alpha \Delta I_э$, меньшим, чем входной сигнал $\Delta I_э$.

Действительно, при подаче входного сигнала в эмиттер усиление по току отсутствует. Но, как мы увидим дальше, при таком способе включения транзистора может иметь место усиление по напряжению.

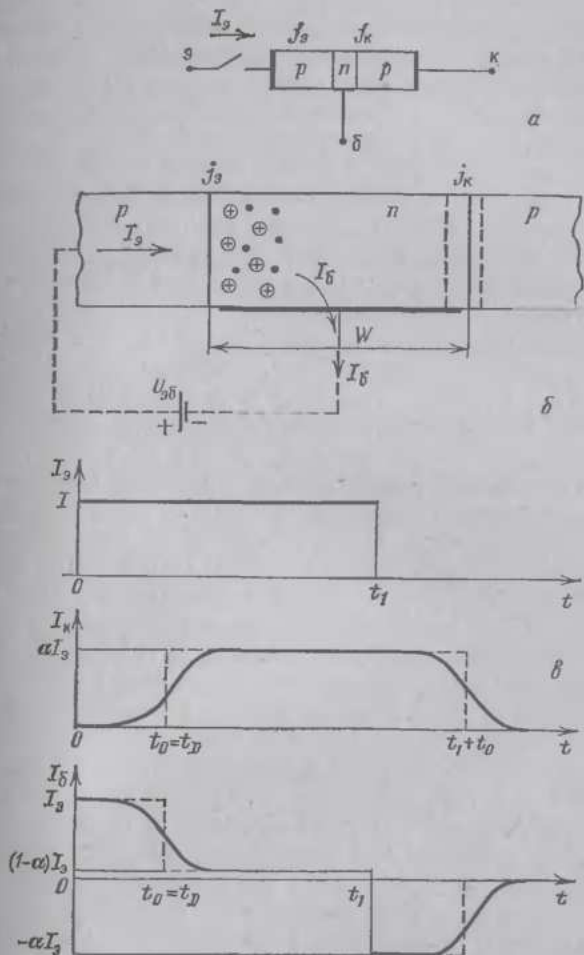


Рис. 102. Процесс установления выходного сигнала в случае, когда входной сигнал подается в эмиттер, а выходной — снимается с коллектора: а — в момент $t = 0$ ключ замыкается и через эмиттерный p — n переход в базу начинают «входить» дырки; б — при $t < t_D \sim W^2/D_D$ дырки не успевают продиффундировать к коллектору. Ток коллектора $I_к$ равен нулю, а $I_э = I_б$; в — зависимости тока от времени: верхняя кривая — тока эмиттера $I_э$ (в рассматриваемой схеме включения ток эмиттера $I_э$ является входным сигналом); средняя — тока коллектора $I_к$ (видно, что время установления t_0 и при включении, и при выключении порядка времени диффузии дырок через базу транзистора $t_D = W^2/D_D$); нижняя — тока базы $I_б$ (в любой момент времени $I_б + I_к = I_э$)

При таком способе включения транзистора время установления, как мы убедимся, оказывается минимальным, т. е. именно такое включение позволяет использовать транзисторы для усиления и генерирования наиболее высокочастотных сигналов.

Пусть в некоторый момент времени $t = 0$ ключ, показанный на рис. 102, а, замыкается и через эмиттерный электрод транзистора начинает течь ток $I_э$. Как мы знаем, это означает, что через эмиттерный $p - n$ переход (обозначенный на рис. 102, а символом $j_э$ *) в базу начинают инжектироваться дырки.

В стационарном состоянии лавинная доля этих дырок будет уходить в коллектор, образуя ток коллектора $I_к = \alpha I_э$. Понятно, однако, что для того чтобы быть втянутыми в коллектор, инжектированные через эмиттер дырки прежде всего должны до него добраться. Так как дырки двигаются (переносятся) через базу благодаря процессу диффузии, на это требуется время $t_D \sim W^2/D_p$ (W — ширина базы, D_p — коэффициент диффузии дырок). А что же будет происходить, пока дырки будут добираться до коллектора, в промежутке времени $0 < t < t_D$?

Рис. 102, б (он, по существу, в увеличенном масштабе и с некоторыми дополнительными деталями воспроизводит рис. 102, а) помогает ответить на этот вопрос. Инжектированные через переход дырки успели пройти только малую часть базы и располагаются вблизи эмиттерного перехода. Коллекторный переход знать не знает и ведать не ведает о том, что происходит в базе. Соответственно ток через коллектор (если пренебречь током обратносмещенного перехода $I_{к0}$) равен нулю. Но течь только через эмиттерный электрод ток, разумеется, не может. Для протекания тока нужна замкнутая цепь. И пока дырки не доберутся до коллектора, цепь замыкается через базовый электрод. В цепи базы протекает ток $I_б$, равный току эмиттера $I_э$.

Что это физически означает? Поступающие через переход дырки вносят в базу положительный заряд. Пытаясь зарядить базу положительно, вошедшие дырки тем самым притягивают к себе электроны из любого источника, который может их поставлять. Повинуясь этому «страстному призыву», в базу входят электроны из базового элек-

*) Символ « j » — от английского $p - n$ junction ($p - n$ переход) — часто используется для обозначения $p - n$ переходов.

трода *). Никаких препятствий для этого нет. И входит их в базу в единицу времени столько же, сколько дырок входит в базу через эмиттерный электрод. Короче говоря, пока «в дело не вступил коллектор», ток протекает через эмиттерный переход точь-в-точь так же, как через прямосмещенный диод.

Но вот проходит время $t_D \approx W^2/D_p$, и дырки добираются до коллектора. Теперь на каждые N дырок, входящих в единицу времени в базу через эмиттер, их доля, равная αN , будет уходить в коллектор. В коллекторном электроде потечет ток $I_k = \alpha I_a$. В базе из N вошедших из эмиттера дырок их будет оставаться лишь $(1 - \alpha) N$. Соответственно, чтобы база не заряжалась ни положительно, ни отрицательно, из базового электрода в единицу времени должно теперь входить $(1 - \alpha) N$ электронов. Поэтому базовый ток уменьшится до величины $I_b = (1 - \alpha) I_a$.

На рис. 102, в штрихом показана идеализированная картина коллекторного I_k и базового I_b токов, которая была бы справедлива, если бы все электроны двигались через базу строго с одной и той же скоростью и достигали бы коллектора в строго определенный момент времени t_D . В этом случае при $t < t_D$ ток через коллектор равнялся бы нулю. Базовый ток I_b был бы равен току эмиттера I_a . В момент времени t_D коллекторный ток увеличился бы скачком до величины αI_a , а базовый — уменьшился до $(1 - \alpha) I_a$.

В действительности, как мы хорошо знаем, процесс диффузии обусловлен случайными блужданиями. Поэтому часть дырок, которым по счастливой случайности удалось избежать столкновений в базе, достигнет коллектора очень быстро. Затем к коллектору подойдут те дырки, которым повезло меньше, и т. д. Время $t_D = W^2/D_p$ характеризует лишь среднее время, за которое к коллектору подойдет основная масса дырок. Поэтому на самом деле зависимости $I_k(t)$ и $I_b(t)$ окажутся не резкими, а «размытыми» (сплошные кривые на рис. 102, в). Характерное время установления коллекторного (выходного) тока t_0 при подаче сигнала в эмиттер оказывается порядка времени диффузионного пролета дырок через базу: $t_0 \sim t_D \sim W^2/D_p$.

*) В конечном итоге, разумеется, цепь база — эмиттер замыкается через источник напряжения $U_{эб}$, смещающий переход j_a в прямом направлении. Включение этого источника показано на рис. 102, б штрихом.

Посмотрим теперь, что произойдет, если в момент t_1 разомкнуть ключ в цепи эмиттера. При этом, конечно, ток эмиттера немедленно обратится в нуль ($I_3 = 0$).

А ток коллектора? А ток коллектора в первый момент ровно ничего не почувствует. В самом деле, размыкание ключа в цепи эмиттера мгновенно прекращает поступление в базу новых дырок. Но возле коллектора сколько в этот момент было дырок, столько и осталось. Соответственно число дырок, попадающих в коллекторный переход j_k и образующих коллекторный ток I_k , не изменилось. И (в идеализированной картине) ток коллектора будет продолжать оставаться неизменным и равным αI_3 , пока последние дырки, находившиеся в момент размыкания ключа у эмиттерного перехода, не продиффундируют через базу к коллектору и не будут «всосаны» им. Понятно, что на это опять-таки потребуется время $t_D \sim W^2/D_p$. Таким образом, и при выключении время установления коллекторного тока $t_0 \sim t_D \sim W^2/D_p$.

А что произойдет в момент выключения с током базы? Как видно из рис. 102, *в*, ток базы в момент $t = t_1$ скачком изменит направление и увеличится от $I_0 = (1 - \alpha) I_3$ до величины αI_3 .

Как это понять? А очень просто. Ток эмиттера обратился в нуль, и поступление дырок в базу прекратилось. К этому моменту концентрация электронов в базе равнялась концентрации дырок. База была электрически нейтральна. И вот, начиная с этого момента, коллектор отсасывает из базы дырки, а новые дырки в базу не входят. В базе остается нескомпенсированный отрицательный заряд «лишних» электронов. Этот отрицательный заряд теперь стремится вычихнуть лишние электроны и... — больше им деться некуда — лишние электроны уходят в базовый электрод. Сколько дырок в единицу времени отсасывает коллектор, столько электронов выталкивается в базовый контакт: $I_0 = I_k = \alpha I_3$.

Итак, когда входной сигнал подается в эмиттер, а выходной снимается с коллектора, и при включении, и при выключении время установления $t_0 \approx W^2/D_p$.

Длина базы первых транзисторов, серийный выпуск которых начался в конце 40-х годов, составляла несколько десятков микрометров. При $W \sim 20$ мкм время установления t_0 равнялось приблизительно $t_D \approx W^2/D_p \approx 4 \cdot 10^{-7}$ с, а предельная рабочая частота $f_{пр} \sim 1/t_0$ не превышала нескольких сотен килогерц. Современные сверхвысокочастотные биполярные транзисторы способны

усиливать и генерировать сигналы с частотой, в десятки тысяч раз большей — вплоть до десятков гигагерц. Рекордно малое время установления t_{02} полученное к

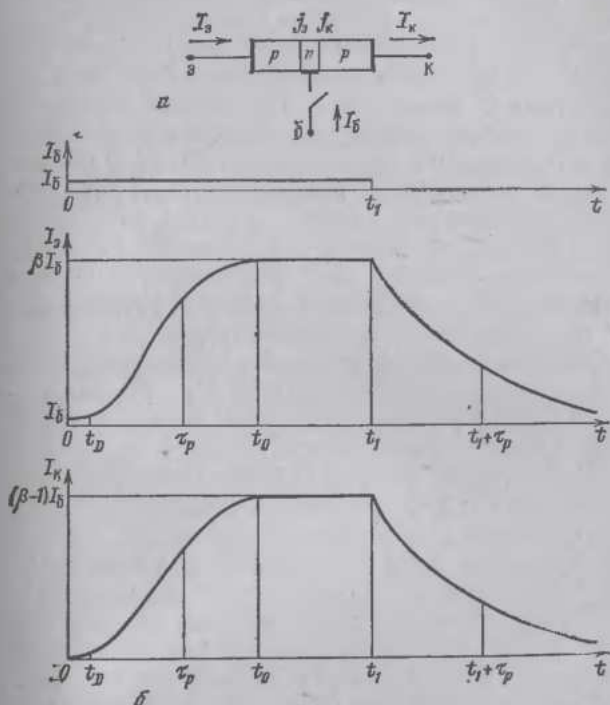


Рис. 103. Процесс установления выходного сигнала в случае, когда входной сигнал подается в базу, а выходной — снимается с коллектора. *a* — В момент $t = 0$ ключ замыкается и через базовый электрод в базу начинают входить электроны. *б* — Зависимости от времени: верхняя кривая — базового тока I_B ; средняя — тока эмиттера I_E ; нижняя — тока коллектора I_K

1986 г. составляет ~ 30 пс. Толщина базы W такого транзистора равняется приблизительно $0,1$ мкм.

Обсудим теперь, чем определяется время установления t_0 в случае, когда входной сигнал подается в базу транзистора, а выходной снимается с коллектора (рис. 103). В таком режиме транзистор способен усиливать и ток, и напряжение, и поэтому именно такая схема включения используется чаще всего. Увы, за все приходится платить, и, как мы сейчас убедимся, в данном случае платить приходится быстродействием. Время установления t_0 при

таким включении оказывается в β раз больше, чем в предыдущем рассмотренном нами случае.

Пусть в момент времени $t = 0$ ключ, показанный на рис. 103, а, замыкается и в базовом электроде транзистора начинает протекать заданный ток I_6 . Это означает, что в базу каждую секунду начинает поступать $N = I_6/e$ электронов. Пытаясь зарядить базу отрицательно, эти электроны немедленно провоцируют положительно заряженные дырки выйти из эмиттера и попытаться уничтожить «виновников возмущения». Пока инжектированные дырки не дойдут до коллектора, чтобы «пресечь возмущение», из эмиттера должно выходить ровно столько же дырок, сколько электронов входит в базу: $I_2 = I_6$.

Рисовать картинку, аналогичную рис. 102, б, нет необходимости: ее пришлось бы просто повторить один к одному. Ситуации полностью аналогичны.

Заметим только, что ток эмиттера I_2 , который пока что протекает в цепи эмиттера, в β раз меньше, чем тот, который будет протекать в стационарном состоянии, когда истечет время установления t_0 . Если $\beta = 100$ (довольно типичное для транзисторов значение), то протекающий на данном этапе ток в 10) раз меньше, чем тот, которому предстоит установиться.

Но вот время диффузии дырок через базу $t_D \sim W^2/D_p$ истекло и первые инжектированные из эмиттера дырки достигли коллектора. Как мы помним, в предыдущем случае, когда сигнал подавался в эмиттер, на этом переходной процесс включения и заканчивался. Устанавливалось стационарное состояние (рис. 102, в). Здесь же главные события с этого момента только начинают развиваться. Именно с этого момента и ток эмиттера, и ток коллектора начинают нарастать. И, как мы знаем, току эмиттера предстоит возрасти еще в β (т. е. в десять, в сто, а то и в тысячу) раз! Наша задача — понять, почему это происходит и как долго будет длиться этот процесс.

Начнем обсуждение с первого вопроса. Пусть для наглядности базовый ток таков, что каждую секунду в базу входит круглым счетом 100 электронов. Пока $t < t_D$, со стороны эмиттера в базу каждую секунду входят сто дырок и база остается электрически нейтральной.

Но вот время t_D истекло, и первые дырки достигли коллектора. Если, опять-таки для наглядности, считать, что $\alpha = 0,99$ ($\beta = 100$), то, начиная с этого момента, из 100 дырок, поставляемых эмиттером, 99 будут утекать в коллектор. В базе остается одна-единственная дырка,

«Караул!» — кричит эмиттер. — «База каждую секунду шлет 100 электронов, а из тех 100 дырок, что я высылал им навстречу, чтобы база осталась нейтральной, коллектор крадет 99. Теперь мне придется высылать 199 дырок!»

Но из посланных 200 дырок 198 опять украдет коллектор, а в базе останется только 2. Эмиттер вышлет 300 — а в базе останется 3. Желая, как теперь модно говорить, возместить недостачу, эмиттер будет все увеличивать и увеличивать поставку дырок в базу, т. е. ток эмиттера, а вслед за ним и ток коллектора будут все расти и расти. До каких пор?

Пока, несмотря на кражу коллектором 99 % дырок, число дырок, остающихся в базе, не сравняется с числом электронов, неумолимо поставляемых в базу. Легко подсчитать, что в рассматриваемом примере ток эмиттера перестанет нарастать, когда эмиттер будет высылать в секунду 10 000 электронов. Только тогда число дырок, остающихся в базе, сравняется с числом электронов, входящих в базу.

Итак, ток эмиттера перестанет нарастать, когда будет выполнено соотношение (55): $I_a = \beta I_b$. При этом ток коллектора I_k также перестанет нарастать и будет (см. рис. 103, б) равен

$$I_k = \alpha I_a = \alpha \beta I_b = \frac{\alpha}{1 - \alpha} I_b = (\beta - 1) I_b. \quad (57)$$

Теперь подумаем над тем, какое время (по порядку величины) уйдет на этот процесс установления стационарного состояния.

Проще всего ответить на этот вопрос, отвлекшись от выяснения сложных отношений между коллектором и эмиттером и сосредоточившись только на судьбе электронов, входящих в базу.

Как мы знаем, ни в коллектор, ни в эмиттер входящие в базу электроны уйти не могут. Их судьба — умереть в базе. Стационарное состояние установится тогда, когда число гибнущих в единицу времени электронов сравняется с числом электронов, поступающих в базу за то же время.

Среднее время жизни избыточных электронов в базе равняется времени жизни дырок в базе τ_p . И понятно, почему. Избыточные электроны «вызывают на себя» дырки из эмиттера и гибнут вместе с ними.

Вспомнив все эти, уже известные нам факты, подумаем теперь над временем установления стационарного состояния в базе. При $t = 0$ в базу начинают втекать электроны. Пока $t \ll \tau_p$, практически ни один электрон не успевает погибнуть, и все входящие электроны накапливаются в базе (вызывая увеличение тока эмиттера и коллектора). Электроны перестанут накапливаться, и стационарное состояние установится, очевидно, через время $t \sim \tau_p$. Это утверждение, можно сказать, просто следует из определения величины τ_p .

Действительно, пока $t \ll \tau_p$, электроны не успевают рекомбинировать и накапливаются. При t , равном нескольким τ_p , совершенно очевидно, что практически все электроны, вошедшие в базу в промежуток времени $0 < t < \tau_p$, успевают умереть. Ясно поэтому, что поступление электронов в базу за счет базового тока компенсируется их гибелью за счет рекомбинации за время установления $t \sim \tau_p$ (рис. 103, б — средняя и нижняя кривые).

Итак, при включении в рассматриваемой схеме время установления $t_0 \sim \tau_p$.

Если в момент t_1 , когда стационарное состояние уже установилось, разомкнуть ключ в цепи базы, то базовый ток, естественно, немедленно обратится в нуль ($I_b = 0$, рис. 103, б). Коллекторный ток I_k в первый момент на это никак не реагирует: концентрация дырок у коллектора, определяющая значение I_k , в первый момент не изменилась. Эмиттерный ток I_a отреагирует на отключение базы мгновенным уменьшением своего значения на величину $\Delta I_a = I_b$. Почему — понятно. Эмиттер посылает в базу столько дырок, сколько их уходит в коллектор (I_k), плюс столько, сколько электронов входит в базу (I_b). В момент t_1 первый компонент (I_k) не изменяется, а второй (I_b) обращается в нуль. Соответственно реагирует и эмиттер.

А затем... Затем, как видно из рис. 103, б, и ток эмиттера, и ток коллектора спадают с характерной постоянной времени, равной времени жизни дырок в базе τ_p .

Дело в том, что после размыкания базового электрода все находящиеся в базе неравновесные, избыточные электроны, накопленные за время включения, оказываются запертыми в базе. Начиная с момента $t = t_1$, их концентрация начинает убывать за счет рекомбинации. И убывает она, как мы только что установили, с постоянной времени, равной τ_p . Поэтому с той же постоянной времени убывает ток коллектора I_k , пропорциональный концен-

трации дырок в базе. И точно так же убывает ток эмиттера I_3 .

Итак, когда входной сигнал подается в базу, а выходной снимается с коллектора и при включении, и при выключении время установления по порядку величины равняется времени жизни неосновных носителей в базе транзистора τ_p .

Отметим, что время τ_p по порядку величины в β раз больше, чем время установления $t_0 = W^2/D_p$ в предыдущем рассмотренном нами случае, когда сигнал подавался в эмиттер транзистора.

Действительно, как мы знаем, $\tau_p = L_p^2/D_p$, где L_p — диффузионная длина дырок. Поэтому отношение $\tau_p/t_0 = L_p^2/W^2$.

С другой стороны, из формулы (52) следует, что

$$\beta = \frac{1}{1-\alpha} \approx \frac{1}{1-1+W^2/2L_p^2} \approx 2 \frac{L_p^2}{W^2}. \quad (58)$$

Когда входной сигнал подается в базу транзистора, а выходной снимается с коллектора и усиление транзистора максимально, быстродействие оказывается приблизительно в β раз меньше, чем предельно возможное.

Как выглядят биполярные транзисторы и как они изготавливаются

Промышленностью выпускаются сотни типов транзисторов: от приборов-великанов, способных усиливать и преобразовывать сигналы с напряжением в сотни вольт и десятки ампер, до транзисторов-малюток, преобразующих сигналы, энергия которых измеряется в фемтоджоулях ($1 \text{ фДж} = 10^{-15} \text{ Дж}$). Рабочая площадь «великанов» (силовых транзисторов) измеряется квадратными сантиметрами. С другой стороны, на кремниевой пластине размером 2×3 мм размещается до 300 000 транзисторов-малюток (рис. 104).

Разумеется, не нужно быть специалистом, чтобы отличить мощный силовой транзистор от транзистора-малютки. Но полагаться только на внешний вид не следует: в одних и тех же стандартных корпусах выпускаются транзисторы с очень разными параметрами. С другой стороны, транзисторы с очень близкими свойствами могут из-

готовляться в совершенно различных корпусах, например пластмассовом и металлокерамическом.

Правильный путь к близкому знакомству с транзистором состоит в следующем. Нужно прочесть на корпусе

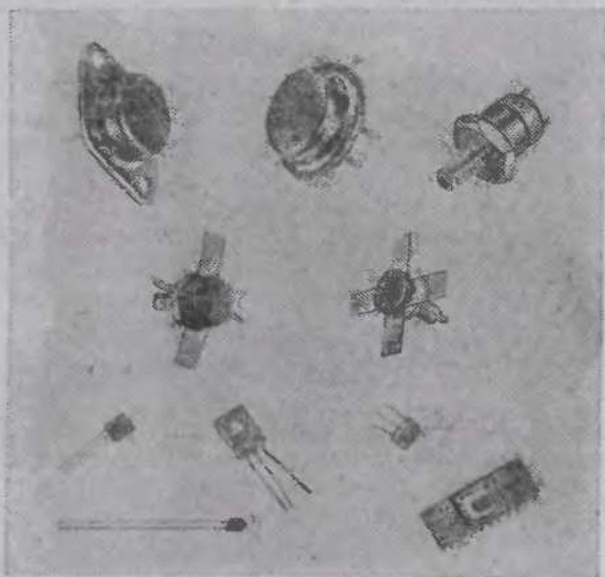


Рис. 104. Различные типы биполярных транзисторов. Верхний ряд — мощные транзисторы, способные выдерживать напряжение между коллектором и базой в сотни вольт и пропускать токи в десятки ампер. Средний ряд — мощные транзисторы, предназначенные для усиления и генерации СВЧ колебаний. Современные биполярные СВЧ транзисторы способны работать на частотах до нескольких десятков гигагерц. Нижний ряд — приборы в пластмассовых корпусах. Их появление привело к резкому удешевлению транзисторов. Подлинную революцию в электронике произвело появление интегральных схем (ИС). Внизу справа показана ИС в корпусе. Основа прибора — кремниевая пластинка размером 2×3 мм (в центре). На такой пластине размещается до 300 000 микротранзисторов, объединенных в сложную электронную схему

буквенно-цифровое обозначение транзистора — так называемую маркировку. А затем... обратиться к справочнику. Под этим кодом в справочнике будут указаны все основные параметры транзистора. Большая и самая основная часть из них окажется нам знакомой. Это предельное напряжение, которое может, не пробиваясь, выдержать обратносмещенный коллекторный переход транзистора, мак-

симальный ток, на который рассчитан прибор, коэффициент усиления по току β , предельная частота усиления и генерации $f_{гр}$.

В справочнике будут указаны и другие параметры прибора. Смысл некоторых из них можно понять, основываясь на том, что мы уже знаем о работе биполярного транзистора. Понимание других требует знакомства с основами схемотехники — науки о том, как создавать на основе приборов с заданными свойствами эффективные, надежные и (относительно) простые схемы нужных устройств.

Представим себе, что мы аккуратно распилили корпус одного из транзисторов, показанных на рис. 104. Обнажилась пластинка полупроводника с тремя выводами — базовым, эмиттерным и коллекторным. Очень аккуратно распилим и ее и протравим в селективном травителе, позволяющем визуализировать (сделать видимыми) p - и n -слои транзисторной структуры.

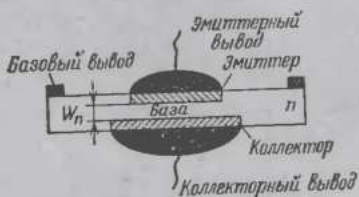


Рис. 105. Сплавной транзистор

Проделав эти delicate операции с транзисторами нескольких десятков типов, мы убедимся, что все транзисторы можно разделить на два больших класса.

Первый класс составляют транзисторные структуры, устройство которых не требует особых комментариев (рис. 105, 106) и очень напоминает схематическое изображение транзистора на рис. 98. Так выглядят сплавные, диффузионные и диффузионно-сплавные транзисторы.

Метод изготовления сплавных транзисторов очень напоминает изготовление диодов (рис. 45). Только к исходной пластине (предположим, n -типа) прижимаются две, а не одна навески, создающие в ней примесь p -типа. Пластина с навесками помещается в печь, где происходит сплавление навесок.

Диаметр пластины — несколько квадратных сантиметров. Диаметр навесок соответствует желаемым размерам коллектора и эмиттера транзистора и составляет обычно несколько десятков микрометров. Поэтому на пластине одновременно размещается несколько десятков или даже сотен навесок. Сплавление всех навесок происходит одновременно.

После сплавления пластину нарезают на отдельные кристаллики-транзисторы, помещают кристаллы в корпуса, приваривают к соответствующим электродам проволочки — выводы. Транзисторы готовы (рис. 105).

Главным параметром, определяющим работоспособность и свойства транзистора, является толщина базы W . Поэтому перед

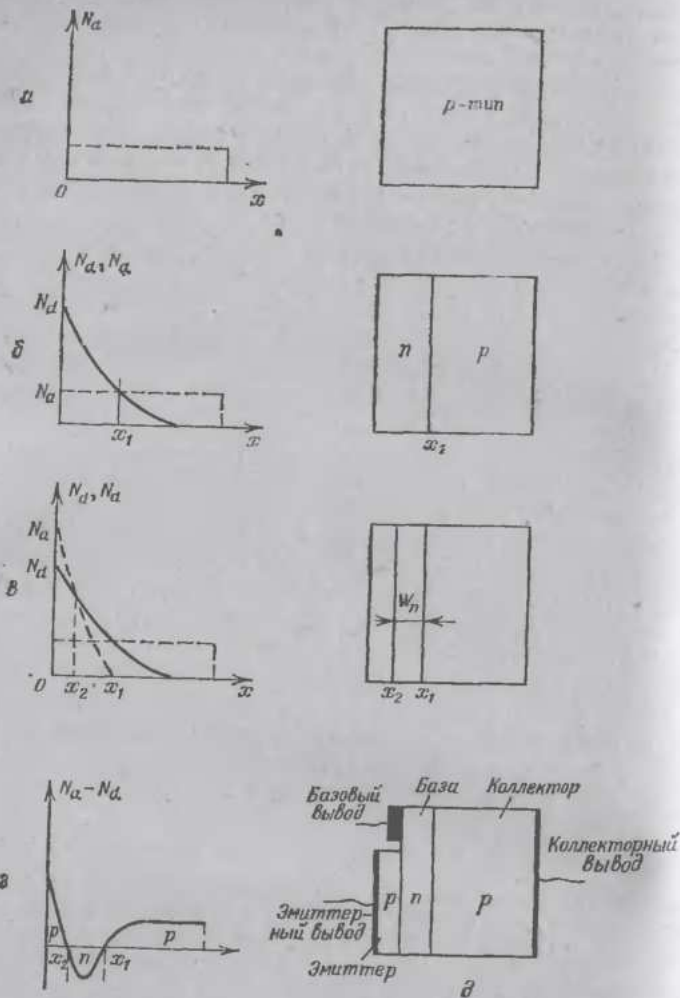


Рис. 106. Изготовление диффузионного транзистора: а — исходная пластина полупроводника p-типа; б — распределение примесей после первой диффузии; в — распределение примесей после второй диффузии; г — часто для наглядности используют график, на котором показывается зависимость $N_a(x) - N_d(x)$; д — диффузионный транзистор

сплавлением толщина исходной пластины тщательно контролируется, а температура и время сплавления поддерживаются с высокой точностью.

Для изготовления диффузионного транзистора (рис. 106) в исходную пластину полупроводника проводится последовательно две диффузии. Чтобы сделать транзистор $p-n-p$ типа, в качестве исходного используется дырочный полупроводник (рис. 106, а). При первой диффузии в кристалл вводится донорная примесь (рис. 106, б). При этом в результате диффузии образуется $p-n$ переход с заданной глубиной залегания x_1 (ср. с рис. 49).

Затем с той же стороны в кристалл проводится вторая диффузия, но уже акцепторной примеси. При этом концентрация акцепторной примеси на поверхности ($x=0$) выбирается большей, чем донорной. А режим диффузии выбирается таким, чтобы глубина залегания второго $p-n$ перехода (точка $x=x_2$) была меньше, чем первого. В результате образуется $p-n-p$ структура (рис. 106, в).

Толщина базы W полученной транзисторной структуры равняется $x_1 - x_2$.

Распределение примеси в диффузионной транзисторной $p-n-p$ структуре нагляднее видно на рис. 106, г. На этом рисунке показана зависимость от координаты разности концентраций акцепторов и доноров в каждой точке структуры $N_a(x) - N_d(x)$ в p -областях структуры $N_a > N_d$, т. е. $N_a - N_d > 0$. В n -базе $N_a - N_d < 0$. Границы эмиттерного (x_2) и коллекторного (x_1) переходов соответствуют точкам компенсации $N_a = N_d$.

После изготовления $p-n-p$ структуры часть эмиттерного слоя стравливается, на все три слоя наносятся омические контакты, привариваются проволочки (рис. 106, д). Структуры помещаются в корпуса. Транзистор готов.

В диффузионно-сплавных транзисторах вместо второй диффузии используется сплавление. Так что базовая область транзистора создается диффузией, а эмиттерная — сплавлением.

Независимо от технологии, по которой изготавливаются транзисторы этого класса, их отличает то обстоятельство, что при создании таких транзисторов технологические операции: сплавление, нанесение металлических контактов, приварка выводов и т. д., производятся с обеих сторон исходной пластины.

Второй класс составляют *планарные* транзисторы (от английского слова *plane* — плоский). При изготовлении планарных транзисторов все без исключения технологические операции производятся только с одной стороны исходной пластины, на одной плоскости, — отсюда и название.

На рис. 107, а показана в разрезе конструкция биполярного транзистора, изготовленного по планарной технологии.

На первый взгляд эта конструкция так сильно отличается от привычного изображения транзистора, что старого знакомого трудно узнать. Тем не менее, взглядевшись вни-

мательно, можно убедиться, что перед нами то же чередование $p - n - p$ областей, что и на рис. 105, 106.

Рис. 107, б показывает, как тот же планарный транзистор выглядит сверху. Особенно ясно видны металлические полоски, с помощью которых осуществляется контакт к эмиттерной (э), базовой (б) и коллекторной (к) областям транзистора.

Существует много разновидностей планарной технологии, но в основе всех их лежит один и тот же принцип, называемый селективным (т. е. избирательным) легированием.

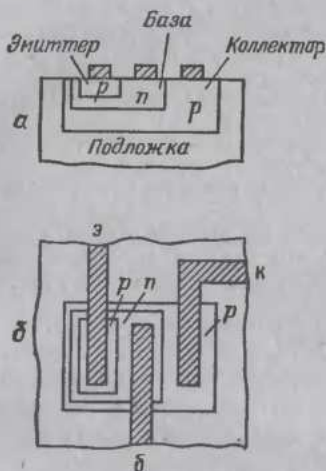


Рис. 107. Планарный транзистор: а — вид транзистора в разрезе; б — вид на транзисторную структуру сверху

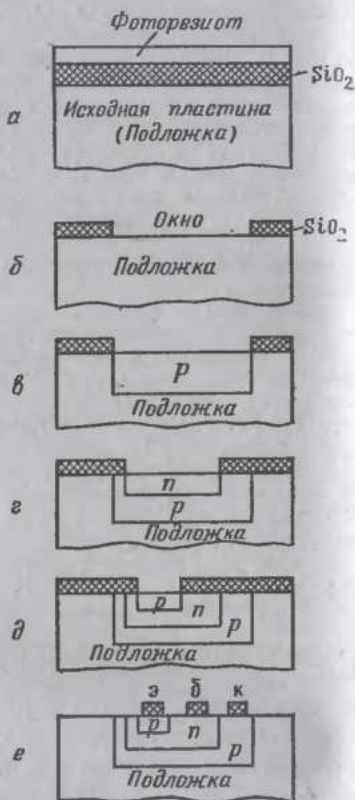


Рис. 108. Принципиальная схема изготовления планарного биполярного транзистора

Одна из широко распространенных технологий, по которой создаются кремниевые биполярные транзисторы, в упрощенном виде показана на рис. 108.

а) Поверхность кремниевой пластины окисляется. На ней образуется слой двуоксида кремния SiO_2 . Затем поверх окисла наносится тонкий слой светочувствительного полимера — фоторезиста. Фоторезист обладает спо-

способностью изменять свои химические свойства под действием света.

К покрытой фоторезистом полупроводниковой пластине прижимается стеклянная пластинка с нанесенным на ней рисунком, составленным из прозрачных и непрозрачных участков — фотошаблон. После этого фоторезист освещается через фотошаблон, и те участки фоторезиста, на которые попадает свет, приобретают способность растворяться в определенных травителях. Эта операция называется *фотолитографией*. Она является одной из самых важных в современной полупроводниковой технологии и чуть дальше мы с ней подробно познакомимся.

б) Пластина с засвеченным фоторезистом погружается в травитель. При этом происходит растворение участков фоторезиста, на которые попадал свет, а также находящихся под ними участков двуокиси кремния. На полупроводниковой пластине в этих местах открываются области чистой поверхности кремния — так называемые окна. Затем в другом травителе, растворяющем незасвеченный фоторезист, но не растворяющем SiO_2 , удаляется оставшийся фоторезист. В результате поверхность пластины представляет собой сочетание окон с участками, защищенными слоем прочной, устойчивой к высокой температуре двуокиси кремния — так называемой маской.

в) Теперь через маску в пластину при высокой температуре проводят диффузию бора (В) или какого-либо другого вещества, образующего в Si акцепторную примесь. Там, где поверхность пластины защищена маской SiO_2 , атомы примеси не достигают кремния. Таким образом, примесь попадает в кремний только в окнах, т. е. селективно (избирательно).

Так формируется будущий коллектор транзистора (ср. рис. 108, в с рис. 107, а).

Если речь идет об изготовлении транзисторов-малюток, число которых на одной пластине достигает нескольких сотен тысяч, то важно подчеркнуть, что на фотошаблоне, который определяет форму окон в окисле сразу «рисуются» все окна — столько, сколько транзисторов должно быть изготовлено на пластине. Таким образом, в результате одного травления получается сразу до сотен тысяч окон. А в результате одной диффузии — сразу сотни тысяч будущих коллекторов.

г) Маска SiO_2 удаляется путем химического травления, и на пластине создается новая маска с окнами, соответствующими расположению базовых областей транзис-

гора. Через эти окна проводится диффузия мышьяка (As) или сурьмы (Sb), являющихся донорами. Концентрация доноров выбирается большей, чем концентрация акцепторов, так что те области полупроводника, куда попадает As и Sb , будут областями n -типа.

Так формируются (сразу все) бавы будущих транзисторов.

д) На поверхности пластины формируется новая маска и новой диффузией акцепторной примеси формируются эмиттерные p -области.

е) На поверхности полученных $p - n - p$ структур (также с помощью соответствующих масок и одновременно на всей пластине) формируются металлические контакты (как правило, из алюминия). Именно эти контакты и видны особенно хорошо на рис. 107, б.

Разумеется, планарная технология позволяет изготавливать и $n - p - n$ транзисторы. При этом меняется лишь очередность введения в кристалл донорной и акцепторной примесей.

При изготовлении транзисторов-малюток и разработке интегральных схем (ИС), когда на одном кристалле изготавливается готовая схема, содержащая до нескольких сотен тысяч транзисторов, планарная технология является единственно возможной.

Действительно, если на одной пластине размещаются десятки и сотни тысяч транзисторов размером несколько микрометров, точно совместить отдельные элементы каждого транзистора с двух сторон пластинки невозможно. Кроме того, планарная технология позволяет использовать сравнительно толстые (толщиной в несколько сотен микрометров) пластины полупроводника. Такие пластины обладают достаточной механической прочностью и удобны в работе. Поскольку полная толщина микротранзистора, включая и эмиттер, и базу, и коллектор, составляет всего доли микрометра, то при двусторонней технологии пришлось бы использовать исходные пластины также толщиной всего в доли микрометра. Работа с пластинами такой толщины практически невозможна.

Фотолитография

Греческое слово «литография» означает искусство рисовать на камне, а также искусство печатать с камня. Поэтому термин «фотолитография» можно перевести, как искусство рисовать по камню светом или искусство светом печатать с камня.

На примере изготовления биполярного транзистора можно видеть, какую важную роль играет процесс фотолитографии в современной полупроводниковой электронике. Именно с помощью фотолитографии на пластине полупроводника (камне) создаются (рисуются) сложнейшие узоры (маски) одновременно для тысяч и сотен тысяч транзисторов (рис. 109).

Фотолитография пришла в микроэлектронику из полиграфии, где использовалась еще с конца прошлого века. С помощью фотолитографии выполненный художником рисунок переносился на так называемый литографический камень — гладкую пластинку из плотного известняка (CaCO_3) (а позднее из алюминия или цинка).

Происходило это следующим образом. Камень (или пластинка) покрывался слоем светочувствительного вещества — альбумина. Не подвергнутый освещению альбумин легко растворялся в воде. Под действием света альбумин, как говорят, «задубливается», т. е. теряет способность растворяться в воде.

Покрытый слоем альбумина камень освещался через подлежащий копированию рисунок. Затем камень (пластину) покрывали краской и промывали водой. Незасвеченный альбумин смывался водой и уносил с собой краску, которой был покрыт. В засвеченных местах альбумин не растворялся и краска оставалась. Так рисунок точно воспроизводился на пластине.

Современный фотолитографический процесс в полупроводниковой технологии, в принципе, почти без изменений воспроизводит известный в полиграфии процесс.

Рисунок, т. е. очертания «окон» для диффузии, вначале фотографируется на фотоластину — фотошаблон. Затем фотошаблон прижимается к полупроводниковой пластине, покрытой слоем фоторезиста — вещества, свойства которого аналогичны свойству альбумина (задубливаться под действием света).

Вот только требования к качеству фоторезиста несравненно выше, чем к альбумину. Прежде всего от резиста требуется очень высокая разрешающая способность. Даже если темный и светлый участки на фотошаблоне находятся друг от друга на расстоянии в доли микрометра, фоторезист должен точно воспроизвести рисунок. Там, где светло — «задублиться», где темно — нет. Кроме того,

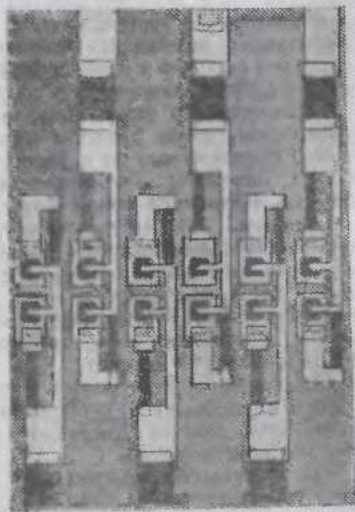


Рис. 109. Память современных быстродействующих ЭВМ выполняется на основе биполярных кремниевых транзисторов. На микрофотографии показан фрагмент большой интегральной схемы (БИС), выполненной на одной кремниевой пластине и содержащей до нескольких сотен тысяч транзисторов

от резиста требуется очень высокая однородность свойств и степень совершенства. Допускается наличие не более одной дефектной точки (цоры, помутнения и т. д.) на площади $\sim 10 \text{ см}^2$.

После того как под действием света произошло задубливание фоторезиста, незасвеченный резист удаляется в травителе — в конфигурации окон точно соответствует рисунку на фотошаблоне.

Для проведения фотолитографии в массовом производстве созданы очень совершенные и высокопроизводительные автоматические установки. В этих установках полупроводниковая пластина автоматически устанавливается на специальный столик, рисунок на фотошаблоне совмещается с рисунком, уже имеющимся после предыдущих операций. Оператор-технолог следит за процессом на экране дисплея и при необходимости вмешивается: дает нужные команды ЭВМ, входящей в состав установки фотолитографии.

Минимальное окно, которое можно сделать с помощью фотолитографии, имеет размеры порядка 1 мкм. И этот минимальный размер определяется не погрешностями работы установки, а физическим явлением — неизбежной дифракцией света на границе темного и светлого полей фотошаблона (Физика, 10 класс).

Однако, как мы уже знаем, размеры некоторых элементов в современных транзисторах составляют десятки доли микрометра. Для создания рисунков-масок с такими размерами используются новые литографические методы — электролитография и рентгенолитография. Новые приставки к слову «литография» указывают на то, какими лучами производится засветка (экспонирование) шаблонов в этих методах.

В электролитографии засветка производится пучком электронов с энергией в несколько десятков тысяч электрон-вольт. Такой пучок можно сфокусировать в пятнышко диаметром в тысячные доли микрометра.

В рентгенолитографии для экспонирования применяют рентгеновское излучение с длиной волны, составляющей тысячные доли микрометра, так что и в этом методе проблемы дифракции не существуют. Практически разрешающая способность этих двух процессов определяется другими техническими факторами и составляет десятки доли микрометра.

Простейшие транзисторные схемы

В этом разделе мы на самых простых примерах познакомимся с тем, как практически реализуется способность транзисторных структур усиливать электрические сигналы.

Схема усилителя с общим эмиттером. На рис. 110 показана одна из самых простых (если не самая простая) схема усилителя на одном биполярном транзисторе типа $p - n - p$.

Три черточки, заключенные в кружок, служат условным обозначением транзистора в электронных схемах. Черточка, в которую упираются две другие, обозначает базу. Наклонная черточка без стрелки — коллектор, наклонная со стрелкой — эмиттер. Если стрелка обращена

в сторону базы — транзистор типа $p - n - p$, если от базы — типа $n - p - n$.

В показанной на рисунке схеме входной сигнал $U_{вх}$ подается в базу транзистора между базой и эмиттером, а выходной $U_{вых}$ снимается с коллектора между коллектором и эмиттером. Эмиттерный электрод, как видно из

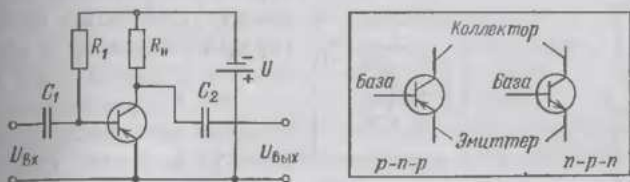


Рис. 110. Схема простейшего усилителя с общим эмиттером на одном биполярном транзисторе $p - n - p$ типа. На вставке справа — условные обозначения биполярных транзисторов в схемах

рисунка, является общим для входного и выходного сигналов. Поэтому показанная на рисунке схема называется *схемой с общим эмиттером*.

В соответствии с тем, что мы уже знаем о свойствах транзисторов, можно утверждать, что в схеме усилителя с общим эмиттером входной сигнал будет в β раз усиливаться по току, а предельная частота сигнала, который такая схема способна усиливать, $f_{пр}$ будет равна по порядку величины $1/\tau$, где τ — время жизни дырок в базе транзистора.

В «Путешествии по Гарцу» Генрих Гейне пишет, что как-то раз портной, рассматривая его гардероб, воскликнул: «В этом сюртуке есть несколько хороших идей!»

В схеме, показанной на рис. 110, также заключено несколько хороших идей. Познакомимся с ними и оценим их последовательно.

Для работы биполярного транзистора в режиме усиления необходимо, чтобы его эмиттерный переход был смещен в прямом направлении, а коллекторный — в обратном. В схеме на рис. 110 выполнены оба эти требования, причем использованы не два (как можно было бы предположить), а всего один источник питания U .

Чтобы яснее представить себе, как задается в схеме режим работы транзистора по постоянному току, сначала рассмотрим транзистор, просто подключенный к источнику питания так, как показано на рис. 111. Чтобы облегчить переход от схематического изображения транзисто-

ра, который мы использовали до тех пор, к общепринятому обозначению транзистора в схемах, справа, на вставке к рис. 111 (в последний раз!), нарисована знакомая $p-n-p$ структура.

Ясно видно, что напряжение источника U приложено к последовательно соединенным коллекторному и эмиттерному переходам и стремится сместить коллекторный переход в обратном, а эмиттерный — в прямом направлении. Насколько оно способно преуспеть в своих стремлениях?

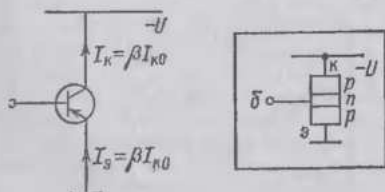


Рис. 111. К транзистору между коллектором и эмиттером подключен источник питания U . «Минус» источника подключен к коллекторному электроду, «плюс» — к эмиттерному. Поскольку в схеме с общим эмиттером все потенциалы отсчитываются от эмиттера, на схеме у эмиттера стоит значок \perp , означающий «общий электрод». Источник питания в схеме не показан. У коллекторного электрода написана величина поданного на него (относительно эмиттера) отрицательного напряжения $-U$. Именно так обозначения чаще всего используются в электронных схемах

Обратносмещенный переход представляет собой, как мы знаем, очень большое сопротивление, прямосмещенный — очень малое. Так как переходы включены последовательно, практически все напряжение источника U оказывается приложенным к коллекторному переходу. Прямосмещенному эмиттерному переходу достаются жалкие крохи. Давайте определим, какие именно.

Для этого зададимся вопросом, который и сам по себе интересен и важен. Какой ток протекает через транзистор с «оторванной базой»? (Именно так часто называют показанную на рис. 111 схему.)

На первый взгляд ответ очевиден. Так как практически все напряжение U приложено к коллекторному переходу, через него будет протекать обратный ток, соответствующий данному напряжению U . Этот ток (условимся обозначать его I_{K0}) и будет течь через транзистор.

На самом деле ответ неправильный. И ошибка оказывается очень большой — в β раз! То есть в десять, в сто, а может быть, и в тысячу.

Если бы напряжение U было приложено между коллектором и базой, ответ был бы совершенно справедливым. Тогда, действительно, протекающий через структу-

ру ток был бы равен току обратносмещенного коллекторного перехода $I_{к0}$. Но в рассматриваемой схеме источник подключен между коллектором и эмиттером, а это полностью меняет дело.

В самом деле, вспомним, что означает протекание через обратносмещенный переход тока $I_{к0}$. Образован ли этот ток диффузионным компонентом или генерационным, все равно, это значит, что коллекторный переход каждую секунду «выбрасывает» в базу $N = I_{к0}/e$ электронов. В диодной структуре эти электроны спокойно уходят затем в базовый электрод. Точно то же самое происходит, если источник питания подключен между базой и коллектором транзистора. Но в рассматриваемом случае дело обстоит не так. Электроны, вброшенные в базу, оказываются в ней запертыми! Уйти через эмиттерный переход к «плюсу» источника они не в состоянии.

Что же происходит?

Как всегда, эмиттер, не желающий, чтобы в базе существовал избыточный отрицательный заряд, пошлет в базу дырки с целью убить избыточные электроны. И, как всегда, из каждых K дырок, которые он пошлет, αK украдет коллектор и только $(1 - \alpha) K$ погибнут, убивая электроны. Ток эмиттера окажется в $\beta = 1/(1 - \alpha)$ раз больше, чем ток $I_{к0}$. И таким же, естественно, будет и ток коллектора.

Ну, а теперь очень легко ответить и на вопрос о напряжении на переходе эмиттер — база. Она будет соответствовать тому напряжению, при котором через переход эмиттер — база протекает ток $I_{э} = \beta I_{к0}$. Если вольтамперная характеристика прямосмещенного перехода эмиттер — база известна, то по заданному току $I_{э}$ легко найти и напряжение $U_{эб}$.

Сила тока $I_{к0}$, даже увеличенная в β раз, как правило, оказывается слишком мала, чтобы обеспечить нормальную работу транзистора. Кроме того, на практике бывает необходимо регулировать протекающий через транзистор ток при одном и том же напряжении питания. Сопротивление R_1 как раз и позволяет задавать силу тока, протекающего через транзистор (рис. 112).

Идея использования сопротивления R_1 особенно наглядна из вспомогательной схемы, показанной на вставке к рис. 112. Если представить себе, что коллекторный переход не подключен к источнику питания, то цепочка «минус» источника питания — сопротивление R_1 — переход эмиттер — база представляет собой просто последова-

тельно соединенные сопротивление и эмиттерный $p-n$ переход, подключенный в прямом направлении.

Как мы знаем, напряжение на прямосмещенном $p-n$ переходе не может превосходить барьерной разности потенциалов $U_{pn} \approx \mathcal{E}_g/e$. На практике напряжение на прямосмещенном переходе эмиттер — база $U_{эб}$ не превосходит $\sim 0,5$ В для германиевых, $\sim 0,8$ для кремниевых и

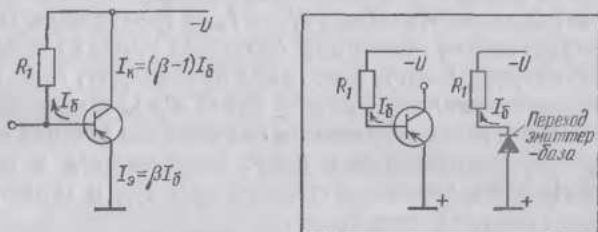


Рис. 112. Сопротивление R_1 определяет силу тока базы $I_b \approx U/R_1$

$\sim 1,1$ В для арсенид-галлиевых транзисторов. А напряжение источника питания U в подавляющем большинстве случаев гораздо больше. Оно, как правило, не меньше 5—10 В (а может, как уже упоминалось, составлять и десятки, и даже сотни вольт). Поэтому ток, протекающий через показанную на вставке цепочку, очень легко оценить:

$$I = \frac{U - U_{эб}}{R_1} \approx \frac{U}{R_1} \quad (59)$$

Но ведь в схеме, показанной на рис. 112, этот ток является базовым током транзистора!

Таким образом, сопротивление R_1 определяет ток базы транзистора I_b в соответствии с формулой (59).

Ток базы задан. Токи эмиттера и коллектора, таким образом, также заданы и определяются выражениями (55) и (53). Изменяя величину сопротивления R_1 , можно изменить в широких пределах ток базы, а следовательно, и токи эмиттера и коллектора.

Обсудим теперь, для чего нужно сопротивление R_n и каким его следует выбирать.

На рис. 113, а показана схема «усилителя», в которой сопротивление R_n отсутствует. Посмотрим, как она будет работать. При отсутствии входного сигнала ($U_{вх} = 0$) через электроды транзистора протекают токи, силы которых определяются сопротивлением R_1 .

Пусть в момент t_1 на вход транзистора поступил отрицательный сигнал $\Delta U_{\text{вх}}$ (рис. 113, *в*). Отрицательный сигнал, поступивший на n -базу эмиттерного $p-n$ перехода, соответствует увеличению прямого смещения на переходе. Базовый ток транзистора увеличится на величину $\Delta I_{\text{бх}} = \Delta U_{\text{вх}}/R_{\text{вх}}$. (Величина $R_{\text{вх}}$ для типового режима работы транзистора обычно приводится в справочнике.) Приращение тока базы $\Delta I_{\text{бх}}$ за счет механизма усиления транзистора по току будет усилено в β раз. Ток

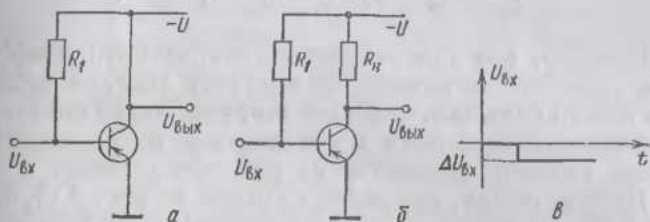


Рис. 113. *а* — Если сопротивление $R_{\text{н}}$ отсутствует, выходное напряжение усилителя $U_{\text{вых}}$ всегда равняется $-U$ и не реагирует на входной сигнал. *б* — При наличии сопротивления $R_{\text{н}}$ увеличение тока, протекающего через транзистор, уменьшает отрицательный потенциал на выходе, а уменьшение тока — увеличивает

эмиттера увеличится на величину $\Delta I_{\text{э}} = \beta \Delta U_{\text{вх}}/R_{\text{вх}}$, а ток коллектора соответственно на $\Delta I_{\text{к}} = (\beta - 1)\Delta U_{\text{вх}}/R_{\text{эх}}$ и . . .

И ничего не произойдет. Напряжение $U_{\text{вых}}$ в схеме на рис. 113, *а* на все эти изменения никак не отреагирует. В самом деле, какой бы ток ни протекал через транзистор, напряжение на выходе транзистора в схеме на рис. 113, *а* всегда равно $-U$. Выход просто подключен параллельно источнику питания.

Чтобы изменение коллекторного тока приводило к изменению выходного напряжения транзистора и, следовательно, к появлению выходного сигнала, необходимо сопротивление нагрузки $R_{\text{н}}$ (рис. 113, *б*).

Теперь, когда в схеме присутствует сопротивление $R_{\text{н}}$, увеличение тока коллектора на $\Delta I_{\text{к}}$ вызовет уменьшение отрицательного напряжения на выходе транзистора на

$$\Delta U_{\text{вых}} = \Delta I_{\text{к}} R_{\text{н}}.$$

Сигнал $\Delta U_{\text{вх}}$, поступивший на вход схемы, окажется усиленным.

Коэффициент усиления схемы по напряжению K_U будет равен, очевидно,

$$K_U = \frac{\Delta U_{\text{вых}}}{\Delta U_{\text{вх}}} = (\beta - 1) \frac{R_{\text{н}}}{R_{\text{вх}}}. \quad (60)$$

Важной характеристикой схемы является коэффициент усиления по мощности K_P , равный отношению мощности выходного сигнала $P_{\text{вых}} = \Delta I_{\text{к}} \Delta U_{\text{вых}}$ к мощности входного $P_{\text{вх}} = \Delta I_{\text{вх}} \Delta U_{\text{вх}}$. Очевидно, что

$$K_P = (\beta - 1) K_U = (\beta - 1)^2 \frac{R_{\text{н}}}{R_{\text{вх}}}. \quad (61)$$

Понятно, что для получения наибольшего коэффициента усиления сопротивление нагрузки выгодно выбирать как можно большим. С другой стороны, легко понять, что особенно усердствовать с увеличением $R_{\text{н}}$ не следует — можно вывести транзистор из рабочего режима.

Действительно, как видно из схемы на рис. 113, б, напряжение на коллекторе транзистора $U_{\text{к}}$ при наличии сопротивления $R_{\text{н}}$ равняется уже не $-U_{\text{к}}$, а меньшей отрицательной величине: $U_{\text{к}} = -(U - I_{\text{к}} R_{\text{н}})$. Если при заданной силе тока, протекающего через транзистор, выбрать величину $R_{\text{н}}$ большей, чем $R = U/I_{\text{к}}$, все напряжение источника питания упадет на сопротивлении $R_{\text{н}}$ и для коллекторного перехода напряжения «не хватит». Нарушится важнейшее условие работы транзистора — наличие отрицательного смещения на коллекторе.

Часто величину $R_{\text{н}}$ выбирают такой, чтобы на сопротивлении $R_{\text{н}}$ при заданном токе $I_{\text{к}}$ падало напряжение, приблизительно равное $U/2$. При отсутствии входного сигнала половина напряжения источника при этом падает на сопротивлении $R_{\text{н}}$, половина — на транзисторе.

Теперь осталось только обсудить назначение емкостей C_1 и C_2 в схеме на рис. 110.

Конденсатор C_1 по постоянному току отделяет анализируемый каскад от источника входного сигнала *). Представим себе, что этого конденсатора в схеме нет, а мы используем наш усилительный каскад для усиления сигнала, снимаемого с высококачественного электродинамиче-

*) Очень редко случается так, что входной сигнал может быть усилен до необходимого значения усилителем на одном транзисторе. Обычно приходится включать последовательно несколько усилительных транзисторных схем-каскадов (от французского cascade — водопад, низвергающийся уступами, ступенями). Усилитель, показанный на рис. 110, — однокаскадный.

ского адаптера хорошего проигрывателя. Сопротивление звукопринимающего узла такого адаптера очень мало — единицы ом. В результате подключения каскада к выходу звукоприемника возникнет схема, показанная на рис. 114,а. База окажется соединенной с эмиттером через очень маленькое сопротивление R_a . Если бы R_a равнялось нулю, то напряжение между эмиттером

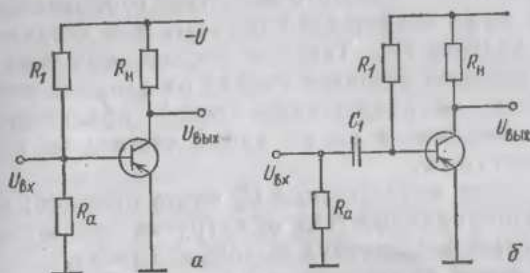


Рис. 114. а — Без конденсатора C_1 подключение усилителя к низкоомному источнику входного сигнала нарушало бы выбранный режим транзистора по постоянному току. б — Наличие конденсатора C_1 устраняет шунтирование схемы по постоянному току. С другой стороны, если емкость конденсатора C_1 достаточно велика ($X_C < R_{вх}$), то для переменного напряжения он представляет собой очень маленькое сопротивление («короткое замыкание»)

и базой $U_{эб}$, задающее ток через транзистор, также оказалось бы равным нулю, несмотря на наличие сопротивления R_1 . В отсутствие входного сигнала ток через транзистор тоже был бы равен нулю. В этом случае отрицательный сигнал с адаптера открывал бы транзистор и усиливался бы им, а положительный — нет. Положительный сигнал с адаптера на n -базу стремился бы уменьшить ток через транзистор. Но уменьшать уже нечего — ток и так равен нулю. И на положительный сигнал транзистор в таком режиме не реагирует. В результате транзистор в таком режиме работает не как усилитель, без искажения усиливающий входной сигнал, а как выпрямитель. Вместо мелодичного звука из динамика, подключенного к выходу такого усилителя, слышно неприятное хрипение — звук сильно искажается.

Даже если R_a не равно нулю, но мало, оно, как говорят, «шунтирует» вход усилителя по постоянному току, изменяя желаемую силу тока, выбранную с помощью сопротивления R_1 . Включение емкости C_1 позволяет избежать этого нежелательного эффекта.

Емкость C_1 выбирается из следующих соображений. Усилитель, как правило, предназначен для усиления сигналов в определенной полосе частот. Например, усилитель проигрывателя — для усиления в полосе звуковых частот 20 Гц — 20 кГц. Чем ниже частота усиливаемого сигнала, тем больше емкостное сопротивление конденсатора $X_C = 1/2\pi f C_1$. Величина C_1 выбирается такой, чтобы на самой нижней частоте емкостное сопротивление конденсатора было значительно меньше, чем входное сопротивление каскада $R_{вх}$. Тогда по постоянному току конденсатор C_1 надежно отделяет каскад от входной цепи, а по переменному он представляет собой практически «короткое замыкание» и ни на какой частоте не влияет на работу усилителя.

Назначение конденсатора C_2 очень похожее: отделить каскад по постоянному току от нагрузки, на которую поступит снятый с усилителя выходной сигнал.

Ну, а теперь от анализа — к синтезу! Рассчитаем параметры усилительного каскада по схеме с общим эмиттером для усиления звуковых сигналов.

Прежде всего зададимся напряжением источника U . Выберем его, скажем, равным 9 В — будем использовать для питания усилителя две плоские батарейки по 4,5 В или шесть круглых по 1,5 В. Именно такими источниками снабжено большинство переносных транзисторных приемников.

Дальше предстоит выбрать транзистор. Поскольку речь идет об усилении слабого сигнала низкой частоты, нам понадобится маломощный низкочастотный транзистор. Раскроем справочник *). К нашим услугам несколько десятков типов таких транзисторов. Нужен известный опыт, чтобы остановиться на каком-то одном. Если опыта нет, можно заглянуть в другой справочник по транзисторным радиоприемникам **) или в паспорт вашего домашнего приемника и посмотреть, на каких транзисторах собран в этом приемнике усилитель низкой частоты. Очень вероятно, что на транзисторах типа МП-41. Они вот уже лет двадцать верой и правдой работают в отечественной бытовой аппаратуре. А может быть, и на транзисторах типа ГТ108 или ГТ109. Нам годится любой из них. Остановимся, например, на транзисторе МП-41.

*) Например: Полупроводниковые приборы: транзисторы / Под ред. Н. Н. Горюнова. — М.: Энергоатомиздат, 1985.

**) Например: Белое И. Ф., Дрызго Е. В. Справочник по транзисторным радиоприемникам. — М.: Сов. радио, 1973.

Коэффициент усиления по току у этого $p - n - p$ транзистора лежит в пределах $\beta = 30 \div 60$. Будем ориентироваться на среднее значение $\beta = 50$.

Из справочника видно, что данные, характеризующие транзистор, приведены для тока эмиттера $I_э = 1$ мА. Следовательно, такой ток является типовым. Нас он вполне устраивает *). Итак, пусть рабочий ток транзистора будет $I_э = 1$ мА.

Сопротивление нагрузки R_n выберем таким, чтобы на нем при токе 1 мА падала половина напряжения источника, т. е. $R_n \approx U/2I_э \approx \approx 4,5$ кОм.

Чтобы ток эмиттера $I_э$ равнялся 1 мА, следует с помощью сопротивления R_1 задать базовый ток $I_б$ в β раз меньшим, т. е. 20 мкА.

Отсюда значение R_1 должно быть выбрано равным $R_1 \approx \approx U/I_б \approx 450$ кОм.

Входное сопротивление транзистора МП-41 в схеме с общим эмиттером составляет ~ 1 кОм. Емкость C_1 выберем из условия, чтобы на частоте $f = 20$ Гц емкостное сопротивление $X_C = 1/2\pi f C_1$ также равнялось 1 кОм. Отсюда $C_1 \approx 10^{-5}$ Ф = 10 мкф. Такой же выберем и емкость конденсатора C_2 .

Усилитель готов (рис. 115)! Если умеете — спаяйте схему и убедитесь, что она работает.

Заметим в заключение, что схема, показанная на рис. 115, обеспечивает усиление по напряжению $K_U \approx$

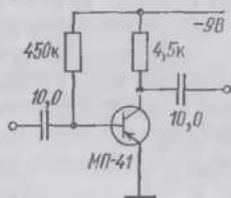


Рис. 115. Каскад усилителя звуковой частоты по схеме с общим эмиттером. Напряжение питания 9 В. Обратите внимание на обозначения величин сопротивлений и емкостей. Вместо кОм пишется просто буква «к». Емкость конденсаторов прямо пишется в мкФ без указания единиц

*) Из данных, приведенных в справочнике, легко установить, в каких пределах можно, если в этом есть необходимость, изменять силу тока, протекающего через транзистор. Предельная рассеиваемая транзистором мощность P_{max} , указанная в справочнике, составляет $P_{max} = 150$ мВт. Это означает, что при напряжении на транзисторе $U = 9$ В протекающий ток не должен быть больше, чем $I_{max} = P_{max}/U \approx 16$ мА. С другой стороны, сила обратного тока коллектора указана в справочнике равной 15 мкА. Рабочий ток транзистора должен, по крайней мере, на порядок превосходить эту величину: $I_{min} \gtrsim 150$ мкА.

$\approx \beta R_H / R_{вх} \approx 200$ и усиление по току $\beta \approx 50$. Таким образом, усиление по мощности составляет $\sim 10^4$.

Схема усилителя с общей базой. В схеме с общей базой входной сигнал подается в эмиттерный электрод транзистора, а выходной снимается с коллектора.

На рис. 116 показана схема с общей базой в наиболее наглядном виде. Так она обычно изображается в учебниках, где поясняется принцип ее работы. Используются два источника питания. Один из них — U_K — смещает коллекторный переход в запертом направлении, другой — U_B — задает прямое смещение на эмиттерном переходе.

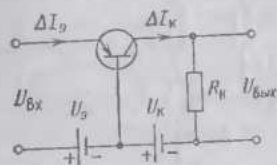


Рис. 116. Схема простейшего усилительного каскада по схеме с общей базой. Использованы два источника питания

Как мы уже знаем, при изменении входного (эмиттерного) тока на величину $\Delta I_э$ коллекторный ток изменяется на $\Delta I_к = \alpha \Delta I_э$. Таким образом, $\Delta I_к < \Delta I_э$. Усиления по току не происходит. Наоборот, по току входной сигнал даже ослабляется, правда, незначительно, всего в α раз.

Усиление в схеме с общей базой происходит только по напряжению. Величина коэффициента усиления по напряжению K_U определяется соотношением между сопротивлением нагрузки R_H и входным сопротивлением транзистора $R_{вх}$. Действительно, амплитуда напряжения входного сигнала $\Delta U_{вх} = \Delta I_э R_{вх}$. Амплитуда выходного напряжения $\Delta U_{вых} = \Delta I_к \cdot R_H = \alpha \Delta I_э R_H$. Коэффициент усиления по напряжению равен

$$K_U = \frac{\Delta U_{вых}}{\Delta U_{вх}} = \alpha \frac{R_H}{R_{вх}} \approx \frac{R_H}{R_{вх}}. \quad (62)$$

Практически таким же будет и коэффициент усиления по мощности K_P :

$$K_P = \frac{\Delta U_{вых}}{\Delta U_{вх}} \cdot \frac{\Delta I_{вых}}{\Delta I_{вх}} = \alpha K_U = \alpha^2 \frac{R_H}{R_{вых}} \approx \frac{R_H}{R_{вых}}. \quad (63)$$

Важно подчеркнуть, что входное сопротивление одного и того же транзистора в схеме с общей базой в β раз меньше, чем входное сопротивление в схеме с общим эмиттером.

Это практически очень важное (взгляните на формулы (60) и (62)!) утверждение прямо вытекает из физического принципа работы транзистора.

Пусть напряжение на эмиттерном переходе изменилось на величину $\Delta U_{эб}$. Тогда ток через переход изменится на величину $\Delta I_э$ независимо от того, в какую схему включен транзистор. Переход эмиттер — база о схеме включения ничего знать не может.

Но... В схеме с общей базой рассмотренная ситуация соответствует изменению входного напряжения на $\Delta U_{эб}$ и входного тока на $\Delta I_э$. Соответственно входное сопротивление равняется $R_{вх} = \Delta U_{эб} / \Delta I_э$.

А в схеме с общим эмиттером изменение входного напряжения такое же — $\Delta U_{эб}$, а изменение входного тока в β раз меньше! Из полного изменения эмиттерного тока $\Delta I_э$, которым эмиттер отвечает на изменение напряжения на переходе, львиная доля — $\alpha \Delta I_э$ — уходит в коллектор. А изменение входного тока — базового, составляет $\Delta I_б = (1 - \alpha) \Delta I_э = \Delta I_э / \beta$. Соответственно $R_{вх} = \Delta U_{эб} / \Delta I_б$ в β раз больше, чем в схеме с общим эмиттером.

Таким образом, если включить уже знакомый нам транзистор МП-41 по схеме с общей базой, его входное сопротивление (при $\beta = 50$) составит всего ≈ 20 Ом. При сопротивлении нагрузки $R_{н} = 4,5$ кОм коэффициент усиления по напряжению (и по мощности) составит, таким образом, ≈ 200 .

Сопоставляя формулы (60) и (62) и учитывая, что в схеме с общей базой величина $R_{вх}$ в β раз меньше, чем в схеме с общим эмиттером, можно убедиться, что коэффициент усиления по напряжению в обеих схемах практически одинаков. Коэффициент усиления по мощности в схеме с общей базой приблизительно в β раз меньше. Зато, как мы знаем (см. с. 240), в схеме с общей базой транзистор сохраняет способность усиливать сигналы до частоты, в β раз большей, чем в схеме с общим эмиттером.

Поэтому, когда речь идет об усилении (или генерации) низких частот, практически всегда используется схема с общим эмиттером. При переходе к более высоким частотам перед инженером встает проблема выбора. Можно использовать один усилительный каскад на более высокочастотном транзисторе (всегда более дорогое) в схеме с общим эмиттером. А можно — несколько каскадов усиления (коэффициент усиления меньше!), собранных на более дешевых низкочастотных транзисторах по схеме с общей базой. При переходе к еще большим, предельно высоким частотам всегда используется схема с общей базой.

На рис. 117 тоже изображена схема с общей базой. На первый взгляд поверить в это довольно трудно. Видно, правда, что входной сигнал действительно подается в эмиттер, но в остальном схема удивительно напоминает только что проанализированную схему с общим эмиттером.

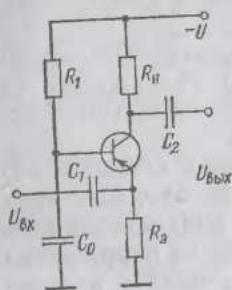


Рис. 117. Схема усилительного каскада по схеме с общей базой. Использован только один источник питания U

Говоря по правде, схема нарочно нарисована так, чтобы она как можно больше напоминала схему с общим эмиттером. И, тем не менее, это — действительно схема с общей базой. А нарисована она так, а не иначе, чтобы подчеркнуть, что при анализе схемы по постоянному току никаких существенных различий между схемами с общим эмиттером и общей базой, действительно, нет.

Проанализируем работу показанной на рис. 117 схемы по постоянному току. Выражение «по постоянному току» означает, что при таком анализе мы будем считать все конденсаторы в схеме имеющими бесконечно большое сопротивление. (А индуктивности, если бы они в нашей схеме встречались — бесконечно малое.) Тогда, очевидно, по постоянному току схема на рис. 117 полностью эквивалентна схеме, показанной на рис. 118, а, и отличается от схемы усилителя с общим эмиттером (рис. 110) только наличием сопротивления R_3 .

Сопротивление R_3 часто используется и в схеме с общим эмиттером. Его введение в схемы улучшает стабильность усилительного каскада, повышает устойчивость работы каскада при изменении температуры, замене одного экземпляра транзистора другим, разряде источника питания и т. д. Введение в схему сопротивления R_3 базируется на общей, очень красивой и очень эффективной в области радиотехники идее отрицательной обратной связи. Однако, увы, изложение этой идеи выходит за рамки книги...

Если величина сопротивления R_3 достаточно мала*), его включение практически не влияет на режим работы транзистора по постоянному току. И, следовательно, мы можем рассчитывать параметры режима — эмиттерный

*) Выражение «достаточно мала» в данном случае означает, что величина R_3 не должна существенно превышать входное сопротивление транзистора в схеме с общей базой,

ток, сопротивление нагрузки R_H , сопротивление R_1 , напряжение на коллекторе — точно так же, как только что рассчитывали эти параметры для схемы с общим эмиттером.

А теперь проанализируем работу каскада по *высокой частоте* (рис. 118, б). Это значит, что теперь, наоборот, сопротивление конденсаторов мы будем считать пренебрежимо малым. При таком взгляде на схему (рис. 118, б) сразу бросается в глаза главный признак схемы с общей

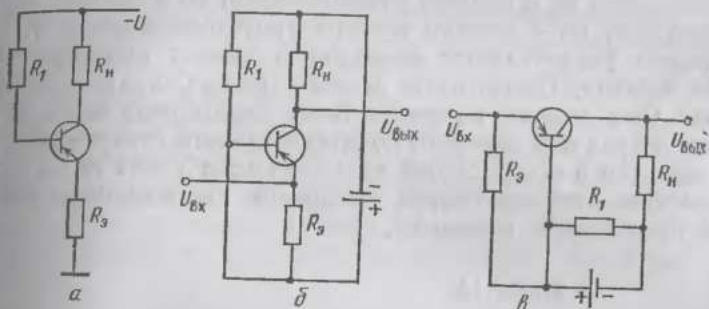


Рис. 118. Схемы, эквивалентные схеме усилительного каскада с общей базой, показанной на рис. 117: а — по постоянному току; б, в — по высокой частоте (схемы б и в не просто эквивалентны, они совершенно одинаковы)

базой: базовый электрод по высокой частоте является общим для входного и выходного сигналов. Еще нагляднее это представлено на рис. 118, в. Показанная на этом рисунке схема в точности совпадает со схемой, приведенной на рис. 118, б. Она просто иначе нарисована *).

С другой стороны, видно, что эта схема практически полностью совпадает с «эталонной» схемой с общей базой (рис. 116). Разница только в том, что положительное смещение на эмиттерном $p - n$ переходе задается не отдельным источником, а с помощью сопротивления R_1 .

На рис. 118, б и в видно также, что в схеме с общей базой без сопротивления R_3 не обойтись. Не будь его, база и эмиттер по высокой частоте (частоте усиливаемого сиг-

*) Рассмотрите внимательно обе схемы и убедитесь в этом. Каждую схему, совершенно ничего в ней не меняя, можно, как правило, нарисовать множеством разных способов. Умение видеть за разными графическими изображениями одну и ту же схему составляет важный элемент навыка «чтения» радиосхем.

пада) оказались бы закороченными (зашунтированными) емкостью C_0 .

Теперь, познакомившись с простейшими схемами усилителей на биполярных транзисторах, можно, если вы почувствовали вкус к схемотехнике, обратиться к анализу реальных схем усилителей телевизоров, радиоприемников, радиол, радиопередатчиков и т. д. Это — целый мир. Со своими законами, своей гармонией, своими критериями пользы, красоты и целесообразности. Разумеется, мы познакомились не с самыми современными, не с самыми стабильными, не с самыми температуроустойчивыми и т. д. схемами усилительных каскадов с общим эмиттером и общей базой. Существуют схемы гораздо более совершенные. Они содержат гораздо более изощренные идеи. Методы их расчета доведены до удивительного совершенства. И все-таки в самой основе всех этих схем лежат те же физические соображения, с которыми мы познакомились на простейших примерах.

Moralité

Трехслойная полупроводниковая структура типа $p - n - p$ или $n - p - n$ при определенных условиях обладает способностью усиливать электрические сигналы. Для этого прежде всего необходимо, чтобы средняя часть структуры, база, была достаточно короткой. Длина базы должна быть меньше, чем диффузионная длина неосновных носителей в ней. Чем короче база, тем больше усиление. Кроме того, чем короче база, тем до более высоких частот сохраняет $p - n - p$ (или $n - p - n$) структура — биполярный транзистор — способность усиливать и генерировать сигналы. Современные биполярные транзисторы усиливают мощность поданного на них сигнала в десятки тысяч раз и сохраняют работоспособность вплоть до частот в десятки миллиардов герц.

Глава 7

ПОЛЕВОЙ ТРАНЗИСТОР

Усердие все преодолевает.

Козьма Прутков

Почти за четверть века до изобретения биполярного транзистора, в 20-х годах нашего столетия, американским исследователем Юлиусом Лилиенфельдом и, почти одновременно, несколькими другими людьми была высказана очень простая, красивая и убедительная идея. Был предложен проект полупроводникового прибора, способного усиливать электрические сигналы.

На заре туманной юности. . .

Основная идея. Основная идея прибора поясняется рис. 119. Прибор представляет собой обычный плоский конденсатор (рис. 119, а). Одной из обкладок конденсатора служит слой металла, другой — пластина полупроводника. Если между обкладками такого конденсатора приложить напряжение, то в зазоре конденсатора возникнет электрическое поле E_1 . На поверхности полупроводниковой пластины, там где она граничит с воздушным зазором, возникает поле $E_2 = E_1/\epsilon$, где ϵ — диэлектрическая проницаемость полупроводника (см. рис. 37, в).

Как мы хорошо знаем (см. гл. 4), возникшее на границе пластины поле проникает в полупроводник на определенную глубину, зависящую от концентрации свободных носителей в полупроводнике. В зависимости от направления это поле может либо вытеснить носители из приповерхностного слоя полупроводника, создав слой, обедненный носителями, либо, наоборот, обогатить приповерхностный слой избыточными носителями заряда.

Если обкладкой конденсатора служит полупроводник n -типа, как показано на рис. 119, б, то при отрицательном потенциале на металлической пластине электроны будут отесняться из приповерхностного слоя полупроводника. При смене полярности приложенного к конденсатору на-

приложения приповерхностный слой полупроводниковой пластины, напротив, обогатится избыточными электронами.

А теперь приложим вдоль полупроводниковой пластины напряжение от отдельного источника U_0 (рис. 119, в). Если приложенное к конденсатору напряжение U_1 имеет такую полярность, что часть полупроводниковой пластины обедняется носителями, то сопротивление пластины

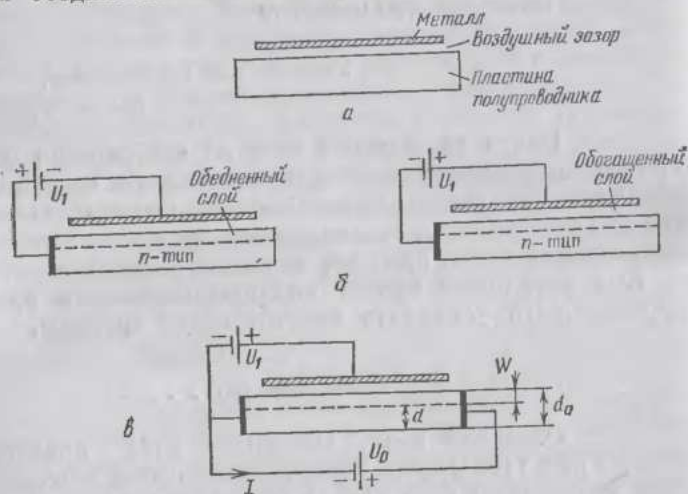


Рис. 119. Идея прибора, предложенного в 1925 г. Ю. Лилиенфельдом: *a* — конденсатор, нижняя обкладка которого изготовлена из полупроводника; *б* — если (для определенности) обкладка изготовлена из полупроводника *n*-типа, то при напряжении на конденсаторе U_1 , когда «плюс» подан на полупроводник, а «минус» на металл, в полупроводнике возникает слой, обедненный электронами (слева), при противоположной полярности приложенного напряжения (справа) в полупроводнике возникает слой, обогащенный электронами; *в* — напряжение U_1 изменяет (модулирует) сопротивление полупроводниковой пластины и, соответственно, модулирует ток I в цепи источника U_0

возрастает и ток I , текущий в цепи источника U_0 , естественно, уменьшается. Наоборот, если часть пластины обогащается свободными носителями, ее сопротивление, естественно, падает и сила тока I увеличивается.

Ток, текущий вдоль полупроводниковой пластины, управляется перпендикулярным току электрическим полем. Если сигнал, управляющий перезарядкой конденсатора U_1 , будет слабее, чем сигнал в цепи питания источника U_0 , то с помощью такого прибора будет реализовано усиление электрических сигналов.

Не правда ли, простая, ясная и изящная идея?

Идея, бесспорно, замечательная. У нее был только один недостаток. Она «не работала». Многочисленные попытки Лилиенфельда и его последователей не увенчались успехом. Им не удалось не только получить усиление, но даже наблюдать сколько-нибудь заметную модуляцию сопротивления полупроводниковой пластины.

Простые оценки. Посмотрим, на что могли рассчитывать в 20-х годах Лилиенфельд и его последователи.

Предположим, что полярность напряжения U_1 , приложенного к полупроводниковому конденсатору, такова, что возникшее в полупроводнике электрическое поле оттесняет носители и создает вблизи поверхности обедненный слой. Тогда, зная поле на поверхности E_1 и концентрацию легирующей примеси N_d , мы можем по формуле (32) определить ширину обедненного слоя:

$$W = \frac{\varepsilon \varepsilon_0 E_1}{e N_d} \quad (64)$$

(см. также рис. 39, б).

Максимальная напряженность поля, которую можно реализовать в полупроводнике, ограничена величиной характерного поля ударной ионизации E_i . Для Ge и Si значение E_i составляет, как мы знаем, $(2 - 3) \cdot 10^5$ В/см.

Что касается величины N_d , то ее минимально возможное значение определяется достигнутым уровнем технологии. В 20-х годах величина $N_d \sim 10^{18}$ см⁻³ служила предметом законной гордости экспериментов. И вполне справедливо. Ведь она соответствует концентрации посторонних примесей в полупроводнике $\lesssim 0,01\%$!

При $E_1 = 3 \cdot 10^7$ В/м, $N_d = 10^{24}$ м⁻³ и $\varepsilon \approx 10$ ширина обедненной области составляет, как нетрудно подсчитать, $\approx 2 \cdot 10^{-8}$ м, т. е. всего 0,02 мкм.

С другой стороны, в те годы практически невозможно было изготовить полупроводниковую пластину толщиной d_0 , меньшей, чем ≈ 50 мкм. Если в пластинке толщиной $d_0 = 50$ мкм очистить от носителей слой W толщиной 0,02 мкм (рис. 119, в), то ее сопротивление увеличится всего-навсего на $W/d_0 = 1/2500$, т. е. на 0,04%.

Рассчитанная величина слишком мала, чтобы практически реализовать эффект усиления. Но измерить изменение сопротивления на 0,04% можно было уже и в 20-е годы абсолютно надежно.

Так вот, экспериментально изменение сопротивления либо вовсе не наблюдалось, либо оказывалось в сотни и тысячи раз меньше рассчитанной величины.

Старые друзья. Если бы уменьшение сопротивления совпадало с расчетным значением, особых оснований для пессимизма не было бы. Всем было ясно, что с развитием технологии уровень очистки полупроводниковых кристаллов будет улучшаться и величину N_d можно будет существенно уменьшить. С уменьшением N_d будет возрастать и ширина слоя W , в пределах которого напряжение U_1 будет модулировать сопротивление полупроводника. С другой стороны, ясно было, что толщину пластины d_0 со временем тоже можно будет научиться делать меньшей.

Но, увы! Измеренное значение было гораздо меньше ожидаемого. А это уже ставило под сомнение правильность идеи и возможность ее практической реализации.

Исследователи повторяли опыты снова и снова. Изменяли геометрию конденсатора, материал изолирующего зазора, тип и характер легирования полупроводника, температуру опыта — и все безрезультатно.

А между тем объяснение в буквальном смысле лежало на поверхности. За обескураживающие и огорчительные результаты опытов ответственность несли старые наши знакомые — поверхностные состояния (с. 103).

Если каждый находящийся на поверхности атом создает один поверхностный уровень, а так оно, к сожалению, чаще всего и бывает, то плотность поверхностных состояний составляет, как мы знаем, $\sim 10^{15}$ см⁻². Каждый из этих поверхностных уровней при случае охотно готов захватить свободный носитель, как только он появится вблизи поверхности.

А теперь представим себе, что поле E_1 , возникшее на поверхности полупроводника под воздействием управляющего напряжения U_1 , подтаскивает к поверхности избыточные электроны. Если бы поверхностных состояний не существовало, эти избыточные электроны увеличили бы концентрацию свободных носителей в приповерхностной области. В результате сопротивление приповерхностной области, а следовательно, и всей полупроводниковой пластины сделалось бы меньше. Однако если концентрация (плотность) поверхностных состояний больше, чем концентрация подтянутых к поверхности избыточных носителей, то все носители захватятся на поверхностные уровни и, следовательно, потеряют способность двигаться под действием поля. Такие связанные на поверхностных состояниях неподвижные носители на сопротивление приповерхностного слоя, естественно, никак не повлияют. Теперь остается сравнить плотность поверхностных

состояний с поверхностной плотностью тех носителей, которые могут быть притянуты к поверхности (или отеснены от поверхности) полем E_1 .

Связь между плотностью поверхностного заряда σ , и полем E_1 устанавливается выражением $\sigma_1 = \epsilon \epsilon_0 E_1$ *). При $\epsilon = 10$ и $E_1 = E_i = 3 \cdot 10^7$ В/м значение $\sigma_1 \approx 2,5 \times 10^{-4}$ Кл/м². Поделив эту величину на заряд электрона $e = 1,6 \cdot 10^{-19}$ Кл, получим поверхностную плотность носителей. Она равняется, как нетрудно подсчитать, $\sim 1,5 \cdot 10^{14}$ м⁻² $\approx 10^{11}$ см⁻² — в 10 000 раз меньше, чем типичная плотность поверхностных состояний на реальной поверхности! Понятно, почему экспериментаторы в течение многих лет терпели неудачу за неудачей.

Зрелость и расцвет

Изобретение в 1948 г. биполярного транзистора не прекратило попыток создать прибор, идею которого мы только что обсудили. Больше того, первый успех в создании *полевого* транзистора (так стали называть этот гипотетический прибор после создания биполярного транзистора) выпал на долю Уильяма Шокли — одного из создателей биполярного прибора.

Полевой транзистор с *p-n* переходом. Идея, высказанная Шокли в 1952 г., была очень красивой, смелой и неожиданный.

Как рассуждали исследователи раньше? Созданию полевого транзистора мешают поверхностные состояния? Ну, что ж, будем пытаться от них избавиться. Испробуем один полупроводник, другой, третий. Попробуем в качестве вещества для диэлектрического зазора один материал, второй . . . десятый. Изменим технологию нанесения диэлектрического покрытия. Испытаем напыление, химическое осаждение, изменим температуру процесса. . .

Как рассудил Шокли? Мешают поверхностные состояния? Ну, так и бог с ними. И с самой поверхностью тоже! Создадим электрод, модулирующий сопротивление полупроводниковой пластины, не над поверхностью, а в глубине пластины, где никаких поверхностных состояний нет.

*) Эту формулу можно получить несколькими способами. Например, непосредственно из выражения (64):

$$\sigma_1 = \rho W = e N_d W = \epsilon \epsilon_0 E_1$$

(ср. с выводом формулы на с. 114).

Рис. 120, *а* поясняет идею Шокли. В тонкой пластине *n*-типа любым из уже хорошо известных нам способов (диффузией, сплавлением, ионной имплантацией) создается сверху область *p*-типа. В глубине пластины, на желаемом и точно регулируемом расстоянии от поверхности, возникает, таким образом, *p* — *n* переход. *p*-область

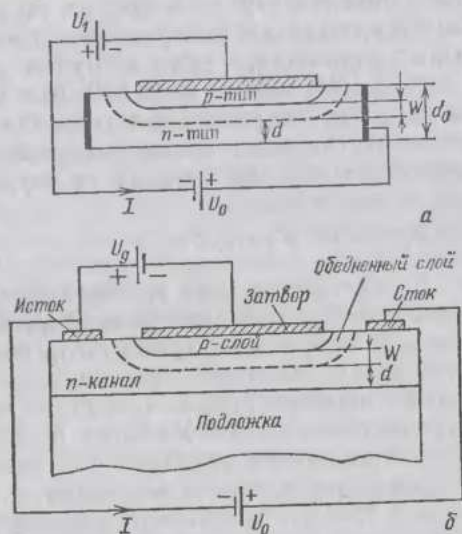


Рис. 120. Полевой транзистор с *p* — *n* переходом: *а* — принципиальная конструкция; *б* — рабочая конструкция транзистора с *n*-каналом

перехода легируется значительно сильнее, чем *n*. Поэтому обедненная область *p* — *n* перехода в основном расположена в *n*-слое. Если к *p* — *n* переходу приложить от источника U_1 обратное напряжение, то обедненный свободными носителями слой, имеющий очень большое сопротивление (слой объемного заряда), будет с ростом напряжения распространяться в глубину *n*-области. Канал d , по которому может протекать ток I от источника U_0 , с ростом напряжения U_1 становится все уже и уже, а сопротивление этого канала все больше и больше. При постоянном напряжении U_0 ток I будет тем меньше, чем больше значение обратного смещения U_1 . Ток, текущий вдоль полупроводниковой пластины, управляется перпендикулярным току электрическим полем! (Ср. с рис. 119, *в*.)

На рис. 120, *б* показана уже не принципиальная, а рабочая конструкция полевого транзистора с *p* — *n* переходом

дом. Транзистор выполнен по планарной технологии. Все три электрода нанесены на одной стороне полупроводниковой пластины, сверху. Электроды, служащие для пропускания тока через прибор, называются *истоком* и *стоком*. Электрод, к которому прикладывается напряжение, модулирующее сопротивление канала, — *затвором*.

Истоком всегда называют тот электрод, через который носители входят в канал. На рис. 120, б изображен полевой транзистор с n -каналом. Носители тока, электроны, входят в канал из электрода, к которому приложен «минус» источника U_0 . Изготавливаются и транзисторы с каналом p -типа. У таких транзисторов к истоку прикладывается «плюс», а к стоку — «минус» источника U_0 . Напряжение, модулирующее сопротивление канала, прикладывается, как правило, между истоком и затвором и называется напряжением затвора U_g^*).

Из рис. 120, б видно, что тонкий слой n -типа, служащий каналом полевого транзистора, выращен на подложке. Непосредственно в работе прибора подложка роли не играет. Она служит основой, на которой изготавливается прибор. Толщина n -слоя, образующего канал, составляет в современных полевых транзисторах десятые доли микрометра. Так что без какой-то относительно толстой подложки не обойтись: пластинка толщиной в доли микрометра слишком хрупка и легко раскалывается. В качестве подложки часто используются высокоомные (полуизолирующие) слои того же полупроводника, из которого изготовлен и канал прибора.

Эпитаксия

Толщина рабочей части многих современных полупроводниковых приборов: полевых и планарных биполярных транзисторов, варикапов, фото- и светодиодов, составляет единицы микрометра или даже десятые доли микрометра. Такие приборы, как правило, изготавливаются на основе очень тонких полупроводниковых пленок, выращенных на сравнительно толстом основании — подложке.

Для выращивания пленок на подложках разработано несколько методов, носящих общее название *эпитаксиальных*. Термин «эпитаксия» составлен из двух греческих слов: «таксис» (расположение в порядке) и «эпи» (на, поверх). Очень удачный термин, напоминающий о том, что речь идет о выращивании поверх подложки монокристаллического (упорядоченного) слоя материала.

*) Индекс « g » — от английского слова gate (затвор, сетка, ворота).

Различают три основных вида эпитаксии: газовую, жидкостную и молекулярно-пучковую.

При *газовой* (или газотранспортной) эпитаксии атомы материала осаждаются на подложку из потока газа. Схема газотранспортного процесса показана на рис. 121. В кварцевой трубе 1 — так называемом реакторе, расположена кассета 2 с набором подложек 3. На подложки, нагретые до определенной температуры, из потока газа 4, проходящего над кассетой, осаждаются атомы материала.

При эпитаксии кремния, например, широко используется реакция восстановления тетрахлорида кремния:



Пары SiCl_4 и водорода подаются в реактор. Образующиеся в результате реакции атомы кремния осаждаются на подложку, имеющую температуру 1150—1200° С, и блуждают по ней некоторое время, пока не найдут себе «подходящее место» и не установят прочные связи с атомами подложки. На подложке начинает расти эпитаксиальная пленка. Скорость роста пленки в процессе газовой эпитаксии составляет около 1 мкм в минуту.

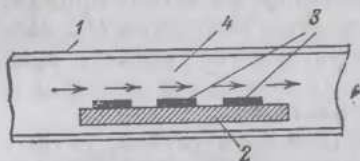
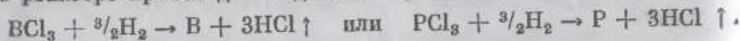


Рис. 121. Схема процесса газовой эпитаксии

С помощью газотранспортной эпитаксии выращиваются слои толщиной от микрометра до нескольких десятков микрометров.

Очень важно, что процесс эпитаксии происходит при температуре значительно более низкой, чем температура плавления, при которой обычно выращиваются кристаллы. Так, температура плавления кремния, например, составляет 1417° С, а температура эпитаксиального роста — почти на 300° С меньше. Более низкая температура процесса приводит к тому, что растущая пленка гораздо меньше загрязняется посторонними примесями. Полупроводниковые материалы рекордной чистоты, как правило, получают именно эпитаксиальными методами.

В процессе эпитаксии возможно строго контролируемое легирование растущего слоя. При газовой эпитаксии, для того чтобы легировать растущую пленку, в реактор вводят пары вещества, содержащего необходимого примесь. Например, при выращивании пленки кремния в реактор подаются хлориды бора (BCl_3) или фосфора (PCl_3) — в зависимости от того, акцепторной или донорной примесью желательнее легировать эпитаксиальный слой. При этом в реакторе происходит выделение примесных атомов:



Пары HCl выносятся из зоны реакции потоком газа.

С помощью газотранспортной эпитаксии выращиваются слои не только кремния, но и германия, арсенида галлия и многих других полупроводниковых материалов.

При *жидкостной* (или жидкофазной) эпитаксии атомы растущего слоя оседают на подложку из расплава или раствора. Раствор (расплав), насыщенный тем полупроводниковым материалом, который необходимо вырастить, наносится на подложку. Затем подложка и расплав охлаждаются. При уменьшении температуры растворимость материала понижается и избыточное количество полупро-

подложка осаждается на подложке. При жидкостной эпитаксии кремния используют растворы Al — Si и Au — Si.

Для выращивания GaAs методом жидкофазной эпитаксии арсенид галлия растворяют в расплавленном галлии. При этом раствор не содержит никаких посторонних примесей. Температура плавления Ga сравнительно невысока (~800° C), поэтому в жидкостном процессе может быть выращен очень чистый материал, не загрязненный примесями реактора, где происходит эпитаксия. Именно с помощью жидкофазной эпитаксии получен GaAs рекордной чистоты с содержанием примесей $\leq 10^{13}$ см⁻³ и очень высокой подвижностью $\mu \geq 20$ м²/(В.с) (при 77 К).

Для легирования растущего слоя в раствор вводится необходимое количество нужной примеси или ее жидкого соединения.

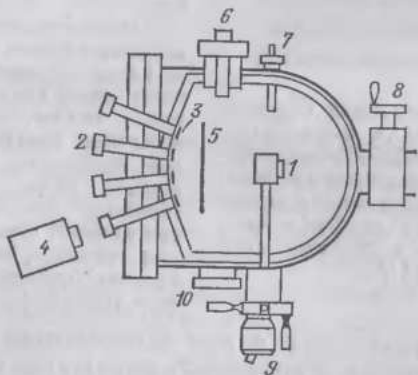


Рис. 122. Схема процесса молекулярно-пучковой эпитаксии: 1 — камера; 2 — испарительные ячейки; 3 — заслонки (Установка молекулярно-пучковой эпитаксии представляет собой сложное инженерное сооружение, включающее комплекс измерительной и анализирующей аппаратуры.); 4 — инфракрасный пирометр для измерения температуры подложки; 5 — главная заслонка; 6 — электронная пушка дифрактометра; 7 — ионный датчик масс-спектрометра; 8 — вакуумный затвор; 9 — манипулятор установки подложки; 10 — приемный экран дифрактометра (установка содержит, также, прибор промышленного телевидения, рентгеновский спектрометр, мини ЭВМ для управления процессом выращивания слоев)

Жидкостная эпитаксия позволяет получить слои полупроводника толщиной от сотых долей микрометра до десятков микрометров.

Третий вид эпитаксии — молекулярно-пучковая (или молекулярно-лучевая) эпитаксия, используется для выращивания очень тонких пленок толщиной от одного атомного слоя до нескольких десятых долей микрометра. Схема процесса иллюстрируется рис. 122.

Подложка закрепляется в камере 1, откачанной до очень низкого давления. В той же камере вокруг подложки полукругом располагаются несколько «отростков» — испарительных ячеек 2. Каждая ячейка отделена от основного пространства камеры заслонкой 3, которая может очень быстро открываться и закрываться по команде, вырабатываемой ЭВМ.

В каждой ячейке располагается вещество, которое понадобится в процессе эпитаксии, и нагреватель. Нагреватель поддерживает в ячейке температуру, достаточную для интенсивного испарения соответствующего материала. Таким образом, в каждой ячейке создается достаточно высокое избыточное давление паров содержащегося в ячейке материала.



Рис. 123. Многослойная структура GaAs — GaAlAs, выращенная методом МВЕ (molecular-beam-epitaxy) — молекулярно-пучковой эпитаксии. Каждая темная полоска — 6 атомных слоев GaAs. Каждая светлая — 4 атомных слоя GaAlAs

По команде ЭВМ открывается заслонка, отделяющая ячейку от камеры, и пучок молекул из ячейки врывается в камеру, летит к подложке и, оседая, кристаллизуется на ней.

Одновременно могут быть открыты несколько заслонок — тогда к подложке поступают сразу молекулы нескольких сортов. С помощью процесса молекулярно-пучковой эпитаксии можно выращивать очень тонкие слои сложных полупроводниковых соединений, строго контролируя толщину, состав и концентрацию примесей в каждом слое (рис. 123).

Полевой транзистор с $p-n$ переходом относится к типу транзисторов с *нормально открытым каналом*. Это означает, что, когда напряжение на затворе не подано ($U_g = 0$), ширина канала d имеет максимальную ширину, а сопротивление канала, наоборот, минимально. С увеличением U_g ширина канала уменьшается и при достаточно большом напряжении U_g может стать равной нулю (канал закрыт для носителей тока).

Важно отметить, что полевой транзистор с $p-n$ переходом никогда не используется в режиме, когда на затвор подан положительный потенциал. На первый взгляд это обстоятельство может показаться непонятным. Действительно, если поменять полярность источника U_g и подать на затвор «плюс» относительно истока, то управляющий $p-n$ переход сместится в прямом направлении. Из p -слоя в n -канал будут инжектироваться дырки. Сопротивление канала понизится. Казалось бы — хорошо! Тем не менее такой режим прямого смещения на затворе не используется никогда. Дело в том, что одним из важнейших достоинств полевого транзистора является его высокое входное сопротивление.

При обратном смещении на затворе ток во входной цепи полевого транзистора, в цепи исток — затвор, является

ся током обратносмещенного $p - n$ перехода и очень мал. В реальных транзисторах он составляет десятки или даже единицы наноампер. Соответственно при напряжении на затворе в единицы вольт входное сопротивление составляет десятки или даже сотни миллионов ом.

Если же использовать полевой транзистор в режиме прямого смещения на затворе, это важнейшее преимущество будет утеряно. При прямом напряжении на $p - n$ переходе входной ток будет велик, а входное сопротивление соответственно мало.

Усердие все превозмогает: МДП транзисторы

Еще через несколько лет после создания полевого транзистора с $p - n$ переходом, в конце 50-х годов, терпение и искусство технологов позволили одержать победу в борьбе, которая длилась более 30 лет. Был найден такой полупроводник, такой материал диэлектрического зазора между металлической и полупроводниковой пластинами (рис. 119) и такой способ нанесения слоя диэлектрика на полупроводник, которые обеспечивали на поверхности полупроводника плотность поверхностных состояний, не превышающую $\sim 10^{10}$ см $^{-2}$, т. е. в 100 000 раз меньшую, чем типичное значение плотности состояний на поверхности полупроводника.

Очень поучительно (и очень, очень важно), что полупроводником, с использованием которого удалось добиться успеха, оказался... кремний. А диэлектриком — окись кремния SiO_2 .

И теперь, когда сотни тысяч неудачных опытов остались позади, процесс создания полевых транзисторов, полностью отвечающих идее Лилиенфельда, выглядит довольно просто. На подложке выращивается тонкий слой кремния с пужным типом легирования и концентрацией — будущий канал. Затем поверхность кремния окисляется. С помощью уже знакомого нам процесса фотолитографии в окисле вскрываются «окна» (с. 256). На эти окна наносятся металлические электроды стока и истока. А там, где окисел остался нетронутым, на месте затвора, сверху тоже наносится металл — транзистор готов (рис. 124). Такие транзисторы называются *МДП полевыми транзисторами* (металл — диэлектрик — полупроводник). Часто, чтобы подчеркнуть, что ди-

электриком служит окисел, их называют МОП *транзисторами* (металл — окисел — полупроводник).

При создании МОП транзисторов часто бывает удобно использовать в качестве подложки не полуизолирующий материал, а сравнительно слаболегированный материал с типом проводимости, противоположным типу проводимости канала, т. е. для создания транзисторов с *n*-каналом использовать слаболегированный кремний *p*-типа

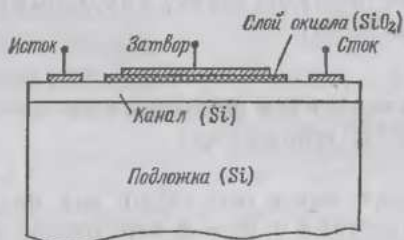


Рис. 124. МДП (МОП) транзистор (ср. с рис. 119, а и 120)

(и, наоборот, для транзисторов с *p*-каналом — подложку *n*-типа).

На рис. 125 показаны две самых важных разновидности МОП транзисторов: *транзистор со встроенным каналом* (рис. 125, а) и *транзистор с индуцированным каналом* (рис. 125, б). Оба показанные на рис. 125 МОП транзистора являются транзисторами с каналами *n*-типа и изготовлены на подложках *p*-типа. Однако точно так же изготавливаются и МОП транзисторы с *p*-каналами на подложках *n*-типа.

Показанный на рис. 125, а транзистор со встроенным каналом по принципу действия и конструкции очень близок к МДП транзисторам, показанным на рис. 119 и 124. Собственно говоря, отличие заключается только в конструкции истока и стока. Как видно из рисунка, для формирования областей стока и истока перед нанесением металлических контактов в *p*-подложке образованы (диффузией или ионной имплантацией) два глубоких *p-n* перехода. «Глубокие» *n*-области служат надежными и качественными контактами к *n*-каналу. А ток между этими *n*-областями течь может практически только по *n*-каналу. В самом деле, цепь исток — подложка — сток представляет собой два последовательно включенных *p-n* перехода. Какова бы ни была полярность напряжения, приложенного между истоком и стоком,

один из этих $p-n$ переходов всегда окажется включенным в обратном направлении, и ток в цепи исток — подложка — сток будет ничтожен.

Понятно, что показанный на рис. 125, а МОП транзистор со встроенным каналом является прибором с нормально открытым каналом. Когда напряжение на

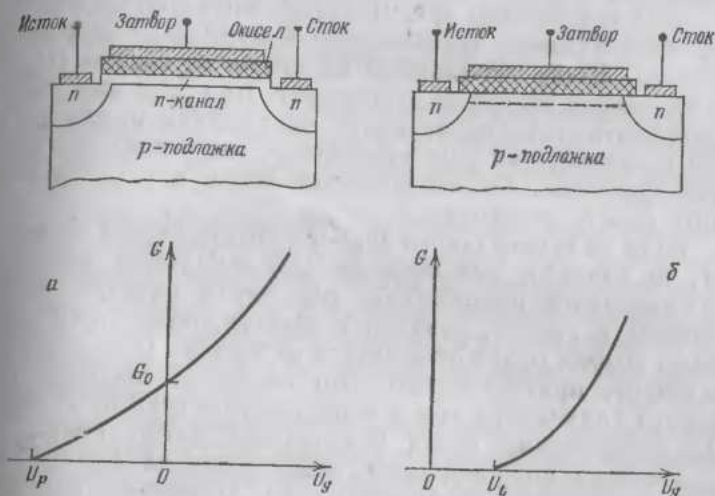


Рис. 125. Разновидности МОП транзисторов: а — транзистор со встроенным каналом, внизу показана качественная зависимость проводимости канала G от напряжения на затворе U_g ; б — транзистор с индуцированным каналом, внизу — качественная зависимость $G(U_g)$. Канал (показан штрихом) возникает только при достаточно большом положительном смещении на затворе U_g

затвор не подано ($U_g = 0$), канал открыт. Его сопротивление R_0 определяется собственными параметрами канала: длиной, толщиной, степенью легирования и подвижностью носителей. При обсуждении параметров полевых транзисторов чаще предпочитают говорить не о сопротивлении, а об обратной сопротивлению величине — проводимости канала $G = 1/R$. При $U_g = 0$ проводимость канала $G = G_0$. Если на затвор подать «плюс», в канале индуцируются дополнительные электроны и проводимость канала увеличится. При отрицательном напряжении на затворе под окислом в канале возникает область, обедненная электронами, и проводимость канала падает. Напряжение на затворе, при котором область объемного заряда, обедненная электро-

нами, перекрывает всю толщину канала, называется *напряжением отсечки* U_p *). При напряжении $U_g = U_p$ проводимость канала обращается в нуль.

МОП транзистор, показанный на рис. 125, б, на первый взгляд выглядит совершенно необычно. Никакого специально сделанного канала между истоком и стоком в таком транзисторе нет. Понятно, что такой транзистор является полевым транзистором с *нормально закрытым каналом*. Когда напряжение на затвор не подано ($U_g = 0$), канала просто не существует. При этом при любой полярности и любой величине напряжения между истоком и стоком, как уже говорилось, ток в цепи исток — сток ничтожен. Он определяется током обратносмещенного $p - n$ перехода.

Если на затвор такого прибора подать «плюс» источника, то вначале, при относительно небольшом величине приложенного напряжения, под слоем окисла, между истоком и стоком, возникнет область, обедненная дырками. Дырочки будут оттесняться от затвора (а электроны, наоборот, притягиваться). Это обеднение дырками никакого влияния на ток в цепи исток — сток не окажет. Он по-прежнему будет практически равен нулю.

Но если положительное напряжение на затворе U_g сделать достаточно большим, то возникнет явление, уже знакомое нам по гл. 3 (см. с. 108), — инверсии типа проводимости. Под затвором возникнет инверсионный слой n -типа, который создает n -канал между областями истока и стока.

Достаточно большое напряжение на затворе создает (индуцирует) канал для протекания тока между областями истока и стока. Поэтому такие приборы и называются транзисторами с *индуцированным каналом*.

Напряжение на затворе, при котором происходит инверсия в области под затвором и образуется канал, называется *пороговым напряжением* и обычно обозначается символом U_t **). При $U_g > U_t$ проводимость канала увеличивается с ростом U_g , так как с увеличением положительного смещения на затворе все больше и больше электронов притягивается к поверхности и концентрация их в канале увеличивается.

Основным преимуществом полевых транзисторов с индуцированным (нормально закрытым) каналом, обес-

*) Индекс «р» — от английского слова pinch-off (отсечка).

***) Индекс «t» — от английского слова threshold (порог).

печившим их очень широкое применение в микропроцессорах и микроЭВМ, является то обстоятельство, что при отсутствии напряжения на затворе они практически не пропускают ток и, следовательно, не потребляют мощность даже при поданном напряжении исток — сток.

Несколько важных деталей

Одно и то же свойство — способность усиливать сигналы, реализуется в биполярном и полевом транзисторах на основе совершенно различных принципов.

В биполярном транзисторе изменение выходного тока в цепи эмиттер — коллектор вызывается изменением входного базового тока. Входной ток протекает через прямосмещенный $p - n$ переход. Поэтому входное сопротивление биполярного транзистора невелико — как правило, оно лежит в пределах от нескольких ом до нескольких килоом.

В полевом транзисторе изменение выходного тока в цепи исток — сток вызывается изменением напряжения на затворе. Входное напряжение U_g приложено к очень большому сопротивлению обратносмещенного $p - n$ перехода или к еще большему сопротивлению пластины изолятора (SiO_2). Поэтому входное сопротивление полевого транзистора любого типа, как правило, очень велико. Оно составляет миллионы, а в некоторых случаях — тысячи миллионов ом (гигаомы).

Такое громадное сопротивление входная цепь полевого транзистора представляет для постоянного тока. Что касается тока переменного, то для оценки входного сопротивления полевого транзистора по переменному току следует учесть, что затвор, совершенно очевидно, представляет собой емкость, величина которой $C_g = \epsilon \epsilon_0 S/t$, где ϵ — диэлектрическая проницаемость изолирующей прослойки, S — площадь затвора, t — толщина прослойки. В МОП транзисторах толщина t слоя окисла, служащего изолятором, составляет $\sim 0,1$ мкм, диэлектрическая проницаемость окисла $\epsilon \approx 4$, площадь затвора $S = LB$, где L — длина, а B — ширина затвора (рис. 126). Длина затвора выпускаемых серийно современных полевых транзисторов составляет 2 — 4 мкм. Однако в лабораториях проходят испытания приборы с гораздо меньшей длиной затвора $\sim 0,1 - 0,2$ мкм.

Ширина затвора B составляет обычно 200—500 мкм. Таким образом, входная емкость затвора лежит в пределах 0,1—1 пФ.

Емкостное сопротивление входной цепи полевого транзистора равняется $X_C = 1/\omega C_g$. При величине входного сопротивления R по постоянному току $\sim 10^7$ Ом

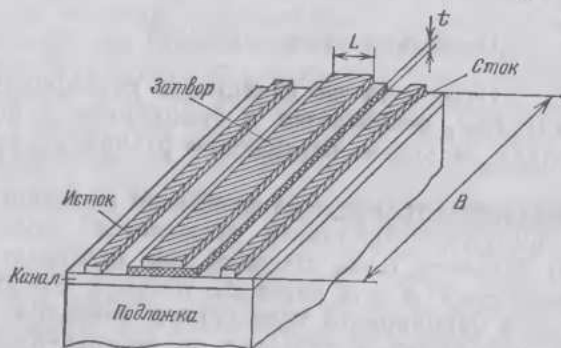


Рис. 126. Исток, затвор и сток полевого транзистора, как правило, представляют собой полоски металла. Размер этих полосок (B) в направлении, перпендикулярном току, гораздо больше, чем длина L электродов в направлении, параллельном току

и $C_g = 1$ пФ емкостное сопротивление станет меньше активного уже на частотах $f \geq 20$ кГц. При дальнейшем увеличении частоты входное сопротивление полевого транзистора будет практически целиком определяться емкостью затвора.

Физическая картина работы полевого транзистора в реальных рабочих режимах

До сих пор, обсуждая принцип действия полевого транзистора, мы, в сущности, представляли себе транзистор как сопротивление, величина которого может быть изменена напряжением, приложенным к затвору.

В самом деле, возьмем в качестве примера полевой транзистор с $p-n$ переходом.

Когда напряжение к затвору не приложено ($U_g = 0$), ширина канала максимальна. Сопротивление канала $R_0 = \rho L/S = \rho L/B \cdot d_0$. Здесь $\rho = 1/\sigma = 1/e\mu n_0$ — удельное сопротивление материала канала, d_0 — пол-

ная ширина канала. Если между истоком и стоком приложить напряжение U , в канале (и в выходной цепи транзистора) будет течь ток $I = U/R_0$.

Если к затвору приложить запирающее напряжение $U_g = U_1$, ширина канала уменьшается на величину области объемного заряда W (см. рис. 120). Зная концентрацию электронов в канале n_0 (она, очевидно, равна концентрации доноров N), величину W легко рассчитать (см. формулу (39)): $W_1 = \left(\frac{2\epsilon\epsilon_0 U_1}{eN} \right)^{1/2}$ *. Теперь канал будет представлять собой сопротивление $R_1 = \rho L/S_1 =$

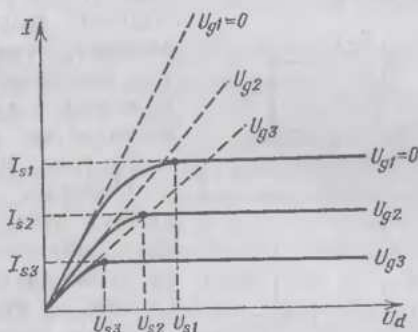


Рис. 127. Вольт-амперные характеристики полевого транзистора при различных напряжениях на затворе U_g . Штриховые прямые — характеристики, соответствующие случаю, когда канал транзистора рассматривается просто как сопротивление, величина которого изменяется в зависимости от напряжения на затворе. Сплошные кривые — реальные вольт-амперные характеристики

$= \rho L/B (d_0 - W_1)$. При напряжении U между истоком и стоком ток в выходной цепи транзистора будет равен $I = U/R_1$. При напряжении на затворе $U_g = U_2 > U_1$ сопротивление канала еще больше возрастает. Наконец, при напряжении отсечки $U_g = U_p$ ширина канала $d = d_0 - W_1$ обратится в нуль. Сопротивление канала делается бесконечно большим.

На рис. 127 штриховыми прямыми показаны вольт-амперные характеристики полевого транзистора в цепи исток — сток, соответствующие рассмотренной картине. Сплошными кривыми показаны реальные вольт-амперные

*) Здесь и дальше мы будем пренебрегать шириной области обедненного слоя при нулевом напряжении на затворе, а в формуле (39) — пренебрегать величиной U_{pn} .

характеристики прибора. Видно, что в области слабых токов реальные характеристики совпадают с идеализированными. А вот при больших токах, т. е. при больших напряжениях на стоке U_d^*), реальные характеристики не имеют со штриховыми прямыми ничего общего. Реальные

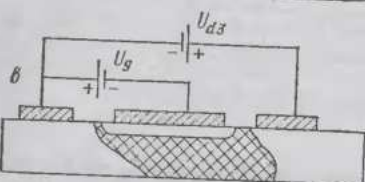
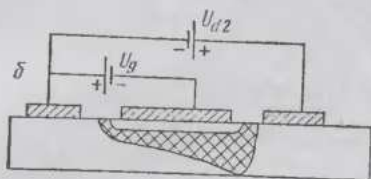
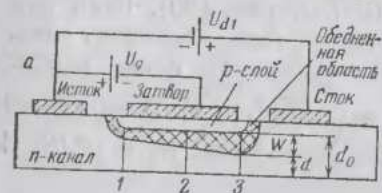


Рис. 128. Механизм насыщения тока в полевом транзисторе:

$U_{d1} < U_{d2} < U_{d3}$
 $(U_{d1} < U_s, U_{d2} = U_s; U_{d3} > U_s)$.
 Напряжение на затворе U_g во всех трех случаях одинаково

На рис. 128, а показана уже знакомая конструкция полевого транзистора с $p-n$ переходом (ср. с рис. 120). К ватвору приложено (относительно истока) небольшое запирающее напряжение U_g . К стоку (опять-таки относительно истока) — напряжение U_d . Знакомая картина.

Раньше мы рассуждали так. Обратное напряжение U_g , приложенное к затвору, создало под затвором область объемного заряда шириной W . В этой области ток не течет. Толщина канала уменьшилась от значения d_0 до значения $d = d_0 - W$. По этому каналу течет ток

ные вольтамперные характеристики насыщаются. При напряжении на стоке $U_d > U_s$ ток перестает зависеть от напряжения на стоке и насыщается на уровне I_s . Чем больше напряжение на затворе, тем меньше и ток насыщения I_s , и напряжение U_s , при котором достигается значение $I = I_s$ (рис. 127).

Полевые транзисторы любого типа в подавляющем большинстве случаев работают в режиме, когда ток в цепи исток — сток не зависит от напряжения стока U_d , т. е. в режиме насыщения. Почему именно в таком режиме, мы обсудим чуть позже. А пока разберемся, отчего реальные вольтамперные характеристики насыщаются.

*) Индекс «d» — от английского слова drain (сток).

$I = U_d/R$, где U_d — напряжение исток — сток (напряжение стока), R — сопротивление канала толщиной d .

Но это — неправда. Или, выражаясь мягче, — не вся правда.

Правда состоит в том, что, когда к стоку приложено напряжение U_d , ширина канала d разная в разных точках. У истока ширина канала максимальна. Наоборот, у стока максимальна ширина области объемного заряда, а ширина канала минимальна.

В самом деле, в любом сечении канала ширина области объемного заряда W определяется разностью потенциалов между потенциалом затвора и потенциалом канала в данном сечении.

Затвор представляет собой металлическую полосу и потенциал ее в любой точке, естественно, одинаков — затвор эквипотенциален.

А канал?

А канал, если к стоку приложено напряжение, эквипотенциальным быть не может: под действием разности потенциалов между стоком и истоком вдоль канала протекает ток. В сечении 1 потенциал канала практически равен потенциалу истока (рис. 128, а). (Так как все напряжение мы считываем как раз от истока, то удобно принять напряжение на истоке равным нулю. Таким образом, в сечении 1 потенциал канала практически равен нулю.) В сечении 3 потенциал канала практически равен потенциалу стока U_d . В сечении 2 потенциал канала имеет промежуточное значение.

Итак, в сечении 1 , у истока, разность потенциалов между каналом и затвором равняется потенциалу затвора U_g . В этом (и только в этом!) сечении ширина объемного заряда будет определяться знакомым выражением

$$W_1 = \left(\frac{2\epsilon\epsilon_0 U_g}{eN} \right)^{1/2}.$$

В сечении 2 потенциал канала имеет относительно сечения 3 положительную величину. (Эта величина составляет некоторую часть от положительного потенциала стока U_d .) Для p — n перехода положительное напряжение, приложенное к n -области, соответствует запирающей полярности. Поэтому в сечении 2 запирающее напряжение между затвором (затвор эквипотенциален!) и каналом больше, чем в сечении 1 . Раз больше запирающее напряжение, значит, шире область объемного заряда W и, значит, меньше ширина канала d .

У стока, в сечении 3, положительный потенциал канала максимален и равен потенциалу стока U_d . Разность потенциалов между затвором и каналом также максимальна и равна $U_g + U_d$ *). Ширина объемного заряда в этом, ближайшем к стоку сечении, максимальна и равняется

$$W_3 = W_{\max} = \left[\frac{2\epsilon\epsilon_0 (U_g + U_d)}{eN} \right]^{1/2}. \quad (65)$$

Итак, наличие напряжения на стоке приводит к тому, что сечение канала становится неодинаковым по длине. Оно максимально у истокового и минимально у стокового конца канала.

Чем больше напряжение на стоке U_d , тем сильнее «перекашивается» канал. В самом деле, ширина объемного заряда у истока (в сечении 1) от значения U_d не зависит. Она определяется только напряжением на затворе U_g . А у стока (в сечении 3) ширина канала с ростом U_d будет становиться все меньше и меньше.

С ростом напряжения U_d сопротивление канала R увеличивается. Действительно, чем больше величина U_d , тем уже становится канал во всех сечениях (кроме сечения 1, у самого истока). А чем уже канал — тем больше сопротивление. Вот почему с ростом напряжения на стоке U_d вольтамперные характеристики (рис. 127) начинают отклоняться от прямой линии и обнаруживают тенденцию к насыщению. С ростом напряжения U_d растет и сопротивление канала, поэтому рост тока с ростом напряжения начинает замедляться.

При напряжении $U_d = U_s$ вольтамперная характеристика насыщается (рис. 127). При этом область объемного заряда у стока практически перекрывает канал (рис. 128, б). С дальнейшим увеличением напряжения U_d ток через транзистор не меняется (рис. 127), а ширина перекрытой области возрастает (рис. 128, в).

Величина напряжения U_d , при котором ток насытится, а канал у стокового конца окажется перекрытым, зависит от напряжения на затворе U_g . Когда напряжение U_d мало и толщина области объемного заряда W

*) Еще раз взгляните на рис. 128, а. Напряжение на затворе отрицательно, а на стоке — положительно. Именно поэтому разность потенциалов между затвором и каналом равна сумме потенциалов $U_g + U_d$. Обдумайте полярность смещений на затворе и стоке в случае транзистора с p -каналом. Убедитесь, что во всех случаях потенциалы на затворе и канале складываются.

практически одинакова вдоль всего канала, ширина канала определяется только величиной U_g . Чем больше U_g , тем уже канал. А чем уже канал, тем при меньших значениях U_d произойдет смыкание области объемного заряда с дном канала у стокового конца транзистора (формула (65), рис. 128, б). И тем меньше будет ток насыщения I_s (рис. 127).

Если подать на затвор такое напряжение, чтобы канал оказался перекрытым даже у истока, т. е. напряжение отсечки $U_p = eNd_0^2/2\epsilon\epsilon_0$, то канал при любом напряжении на стоке U_d окажется полностью запертым. Ток через него протекать не будет.

Основные параметры полевых транзисторов

Одним из самых важных параметров любого прибора, предназначенного для усиления сигналов, является отношение выходного сигнала к входному.

В биполярном транзисторе таким параметром является коэффициент усиления β — отношение приращения выходного тока ΔI_c к величине входного сигнала ΔI_b . В полевом транзисторе изменение выходного тока ΔI достигается изменением напряжения на затворе ΔU_g .

Крутизна. Отношение изменения тока ΔI в цепи сток — исток к изменению напряжения на затворе ΔU_g называется *крутизной*. Для обозначения этого параметра обычно используется символ S :

$$S = \frac{\Delta I}{\Delta U_g}. \quad (66)$$

Измеряется крутизна, очевидно, в Ом^{-1} , или в сименсах (Сим). Эта единица проводимости названа так в честь знаменитого немецкого изобретателя фон Сименса.

Чем больше крутизна, тем большим изменением тока в выходной цепи «отзывается» полевой транзистор на изменение сигнала на затворе. И тем большим в итоге оказывается коэффициент усиления схемы на полевом транзисторе. Поэтому величину S стремятся сделать, по возможности, большей.

Взглянем на рис. 129. На нем показаны уже знакомые нам вольтамперные характеристики полевого транзистора (ср. с рис. 127). Пусть напряжение на затворе $U_g = 0$. Тогда вольтамперной характеристике транзистора соответствует верхняя кривая. Допустим сна-

чала, что напряжение на стоке U_d невелико и ток I , протекающий через транзистор, совсем мал. Такому режиму соответствует точка 1 на верхней кривой.

Изменим напряжение на затворе от $U_g = 0$ до $U_g = U_2$, не меняя напряжения на стоке U_d . Ток через транзистор уменьшится на ΔI_1 . Крутизна транзистора в данном режиме будет равна $S_1 = \Delta I_1 / U_2$.

Пусть снова напряжение на затворе $U_g = 0$, но напряжение на стоке U_d достаточно велико, так что транзистор работает в режиме насыщения. Такому режиму

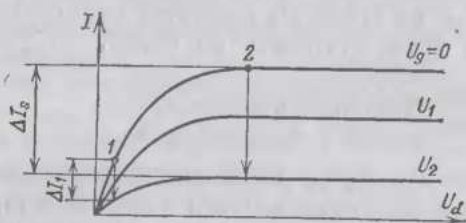


Рис. 129. Вольт-амперные характеристики полевого транзистора (ср. с рис. 127). Одно и то же изменение напряжения на затворе при разных значениях U_d вызывает разное изменение выходного тока. В режиме насыщения изменение выходного тока ΔI_s максимально

соответствует точка 2 на верхней вольт-амперной характеристике.

Опять изменим напряжение на затворе от $U_g = 0$ до $U_g = U_2$. Ток через транзистор уменьшится, но теперь уже на величину $\Delta I_s > \Delta I_1$. Крутизна транзистора в этом режиме $S = \Delta I_s / U_2 > S_1$.

Обратим внимание на следующее важное обстоятельство. Как непосредственно ясно из рис. 129, крутизна полевого транзистора в режиме насыщения максимальна. Вот почему, как уже упоминалось, в подавляющем большинстве случаев полевые транзисторы используют в режиме насыщения.

Крутизна транзистора S зависит от его конструкции и от режима работы: напряжения на затворе U_g и напряжения на стоке U_d . В общем случае формула, описывающая зависимость крутизны от всех параметров, выглядит довольно устрашающе. Однако оценить величину крутизны S в самом важном случае, в режиме насыщения, и оценить довольно точно, можно достаточно просто.

В очередной раз представим себе, что напряжение на затворе $U_g = 0$. Пусть U_d также равняется нулю. Тогда ширина канала во всех сечениях одинакова, максимальна и равняется d_0 . Сопротивление канала $R_0 = \rho L / B d_0$. Оставим напряжение на затворе $U_g = 0$ и будем увеличивать напряжение на стоке U_d . Канал начнет «перекашиваться» — сужаться у стока. Ток с увеличением U_d будет расти, но все медленнее и медленнее. При $U_d = U_{s1}$ ток достигнет максимальной величины I_{s1} (рис. 127, верхняя кривая). При этом, как мы знаем, у стокового конца объемный заряд перекроет всю ширину канала (рис. 128, б).

Можно доказать, что в таком режиме ($U_g = 0$, $U_d = U_{s1}$) сопротивление канала увеличится по отношению к исходной величине приблизительно в три раза: $R_{s1} \approx 3R_0$. Тогда, очевидно, ток насыщения I_{s1} будет равен $I_{s1} = U_d / R_{s1} \approx U_{s1} / 3R_0$.

Приложим теперь к затвору напряжение отсечки $U_g = U_p$. При этом ток, протекающий через транзистор, очевидно, упадет до нуля. А теперь докажем, что, как это ни неожиданно на первый взгляд, $U_p = U_{s1}$!

Доказать это очень просто. Взглянем на формулу (65). При $U_g = 0$ величина напряжения на стоке U_d , необходимая для того, чтобы перекрыть канал, равняется, очевидно,

$$U_d = U_{s1} = \frac{eNW_{\max}^2}{2\epsilon\epsilon_0} = \frac{eNd_0^2}{2\epsilon\epsilon_0}.$$

С другой стороны, напряжение отсечки U_p , по определению, — это такое напряжение на затворе, которое перекрывает канал у истокового конца, где величина напряжения на стоке U_d не оказывает влияния на ширину канала, т. е. это такое напряжение на затворе, при котором в сечении I (рис. 128, а) ширина области объемного заряда $W_1 = d_0$, откуда (см. с. 293)

$$U_g = U_p = \frac{eNd_0^2}{2\epsilon\epsilon_0}.$$

Таким образом, $U_p = U_{s1}$.

Крутизна транзистора S при переходе из режима $U_g = 0$, $I = I_{s1}$ к режиму $U_g = U_p$, $I = 0$ равняется, очевидно,

$$S = \frac{\Delta I}{\Delta U_g} = \frac{I_{s1}}{U_p} = \frac{I_{s1}}{U_{s1}} \approx \frac{1}{3R_0},$$

$$S \approx \frac{1}{3R_0} = \frac{Bd_0}{3\mu L} = \frac{e\mu NBd_0}{3L}. \quad (67)$$

Формула (67), несмотря на приближенный характер, правильно описывает основные зависимости крутизны от параметров транзистора. Крутизна полевого транзистора тем больше, чем меньше длина затвора L и чем выше подвижность электронов μ в канале транзистора.

Первые полевые транзисторы изготавливались из кремния n -типа (подвижность электронов в нем около $0,1 \text{ м}^2/(\text{В}\cdot\text{с})$ (см. табл. 2)), а длина затвора в них составляла десятки микрометров. Характерное значение крутизны транзисторов составляло $\sim 10^{-4}$ Сим.

Успехи современной фотолитографии позволяют изготавливать полевые транзисторы с длиной затвора в десятые доли микрометра, а материалом для транзисторов служит не только кремний, но и арсенид галлия, подвижность электронов в котором составляет $0,5\text{--}0,6 \text{ м}^2/(\text{В}\cdot\text{с})$. Крутизна современных полевых транзисторов $\sim 10^{-2}$ Сим. Рекордные значения крутизны транзисторов, разрабатываемых в лабораториях, доходят до $\sim 1\text{--}2$ Сим.

Быстродействие. Современные арсенид-галлиевые полевые транзисторы являются самыми быстродействующими полупроводниковыми приборами. Именно в полевых транзисторах наблюдалось рекордно малое время установления: ~ 4 пс (1986 г.).

Подробно анализируя работу биполярного транзистора, мы убедились, что минимально возможное время установления определяется в этом приборе временем, которое необходимо, чтобы носители прошли от эмиттера через базу до коллектора.

Точно так же и в полевом транзисторе минимально возможное время установления τ определяется временем движения носителей от истока до стока. Таким образом, $\tau \approx L/v$, где L — длина канала, v — средняя скорость движения носителей вдоль канала. Скорость носителей определяется, конечно, напряженностью поля в канале E . Существуют транзисторы, в которых поле в канале так велико, что скорость носителей имеет максимально возможное значение $v_s \sim 10^7 \text{ см/с}$ (с. 70, рис. 19). Для таких приборов, очевидно,

$$\tau \approx L/v_s, \quad (68)$$

При $L \sim 10^{-4}$ см и $v_s \sim 10^7$ см/с величина $\tau \sim 10^{-11}$ с (~ 10 пс). Такие транзисторы, очевидно, обладают наибольшим быстродействием.

Однако широко используются и такие полевые транзисторы, в которых поле в канале существенно меньше, так что скорость носителей достаточно далека от насыщения и пропорциональна напряженности поля: $v = \mu E$. Такие транзисторы обладают, конечно, меньшим быстродействием, но зато и потребляют значительно меньшую мощность. Последнее обстоятельство особенно важно в схемах ЭВМ, где число транзисторов может достигать десятков миллионов. В этом случае

$$\tau \sim \frac{L}{\mu E}. \quad (69)$$

Полевой транзистор как элемент радиотехнических схем

Как и биполярные, полевые транзисторы используются в схемах в качестве элементов усилителей, генераторов и переключающих устройств. Условное обозначение транзистора в схеме (рис. 130) дает возможность определить тип используемого транзистора.

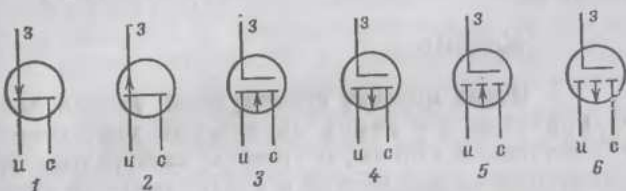


Рис. 130. Условные обозначения различных типов полевых транзисторов (и — исток, с — сток, з — затвор): 1, 2 — транзисторы с управляющим $p-n$ переходом (1 — с n -каналом, 2 — с p -каналом); 3, 4 — МДП транзисторы со встроенным каналом (3 — с n -каналом, 4 — с p -каналом); 5, 6 — МДП транзисторы с индуцированным каналом (5 — с n -каналом, 6 — с p -каналом)

Детали построения схем на полевых и биполярных транзисторах, конечно, несколько различаются, однако основные идеи и общие принципы те же самые. Внимательно разобравшись в идеях, положенных в основу биполярного «сюртука» (с. 260—274), можно легко понять, а то и придумать самому «выкройку» для полевого.

Так же, как и в схеме с биполярным транзистором (рис. 110), последовательно с полевым транзистором

включается сопротивление нагрузки R_H (рис. 131). Коэффициент усиления транзисторного каскада определяется крутизной прибора S и сопротивлением нагрузки R_H .

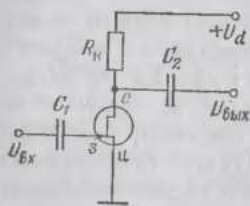


Рис. 131. Схема простейшего усилительного каскада на полевом транзисторе (ср. с рис. 110). Каскад собран на полевом транзисторе с $p-n$ переходом. Канал n -типа. Исходное смещение на затворе $U_g = 0$. Схема с общим истоком

При изменении напряжения на затворе на величину ΔU_g выходной ток транзистора в цепи исток — сток меняется на $\Delta I = S \Delta U_g$. При этом падение напряжения на сопротивлении нагрузки R_H изменяется на $\Delta U_{\text{вых}} = \Delta I R_H = S \Delta U_g \cdot R_H$.

Коэффициент усиления по напряжению

$$K_U = \frac{\Delta U_{\text{вых}}}{\Delta U_{\text{вх}}} = \frac{\Delta U_{\text{вых}}}{\Delta U_g} = S R_H. \quad (70)$$

Величина нагрузки R_H выбирается из тех же соображений, что и в схемах на биполярных транзисторах. Обычно ее значение лежит в пределах от $\sim 10^3$ до $\sim 10^5$ Ом.

При $S \sim 10^{-2}$ Сим коэффициент усиления по напряжению каскада на полевом транзисторе составляет $10-10^3$.

Moralite

Очень простая конструкция: тонкая пластина полупроводника с двумя омическими контактами на концах, истоком и стоком, и третьим электродом, затвором, размещенным посередине и изолированным от полупроводника слоем диэлектрика, способна усиливать электрические сигналы. Усиление возникает за счет того, что напряжение, приложенное между затвором и пластинкой полупроводника, изменяет концентрацию носителей в полупроводнике и модулирует его сопротивление. Поэтому изменяется ток, текущий вдоль пластины, и падение напряжения на последовательно включенном с пластиной сопротивлении нагрузки.

Такой прибор, *полевой транзистор*, обладает очень большим входным сопротивлением и быстродействием, определяемым длиной затвора L и скоростью носителей заряда v в канале. Время установления $\tau \sim L/v$.

Глава 8

ТРАНЗИСТОРЫ И ЖИЗНЬ

Ученик: Вот это лучше,
видно — как и где.

Мефистофель: Суха, мой друг,
теория везде,

А древо жизни пышно зеленеет.

Гете. «Фауст»

Мы начали книгу с утверждения, что полупроводниковых приборов существует великое множество. С некоторыми из них мы познакомились. Свето- и фотодиоды, варикапы, параметрические и выпрямительные диоды — каждый из этих приборов имеет важные практические применения и ежегодно изготавливается десятками миллионов экземпляров. И все-таки не только ни один из этих приборов, но даже и все они вместе взятые по важности и широте применения не могут идти ни в какое сравнение с транзисторами. Ежегодно в мире изготавливаются десятки миллиардов транзисторных структур. На каждого из четырех с лишним миллиардов жителей Земли работают сегодня, в среднем, несколько сотен транзисторов.

Транзисторы, как принято говорить, служат *элементной базой* современной электроники. Иными словами, главными ее элементами.

А между тем и до изобретения транзистора люди слушали радио и смотрели телевизионные передачи, разговаривали по телефону с абонентами, находящимися на другой стороне океана, слушали магнитофонные записи, вели расчеты на ЭВМ, снабжали ракеты головками самонаведения и узлами радиоуправления.

В чем же состоит роль транзисторов? Что изменилось в жизни каждого из нас и в жизни человечества в целом с изобретением транзисторов?

Краткий ответ на эти вопросы состоит в следующем. Транзистор может быть сделан таким маленьким, таким надежным и потребляющим при работе так мало энергии,

что на его основе оказалось возможным создавать схемы и приборы, которые не могут быть выполнены ни на каких других элементах.

Важный и впечатляющий пример — применение транзисторов в медицине. Сегодня десятки тысяч людей носят имплантированные (внедренные) в тело стимуляторы сердечной деятельности (кардиостимуляторы). Не будь транзисторов — эти спасающие жизнь пациентам приборы пришлось бы носить с собой в портфеле или даже катить на тележке.

Исчезла нужда во многих неприятных, болезненных и унизительных процедурах. Чтобы определить уровень кислотности в желудке сегодня многим десяткам миллионов людей, страдающих гастритом или язвенной болезнью, не нужно проделывать процедуру, именуемую в просторечии «глотанием кишки». Пациент проглатывает небольшую таблетку — и транзисторный передатчик передает из желудка большого необходимые данные.

Каждый год появляются десятки новых применений электроники в медицине, и определяющую роль в большинстве из них играют транзисторы.

Вторая часть краткого ответа заключается в том, что благодаря тем же свойствам транзистора: миниатюрности, надежности, ничтожному потреблению энергии, многие схемы и приборы, существовавшие и до изобретения транзисторов, удалось сделать настолько миниатюрными, надежными и дешевыми, что их место и роль в жизни неизмеримо возросли.

Несколько более подробный ответ на поставленные вопросы составляет содержание этой, последней главы.

Первая королева

Труды замечательного русского ученого Александра Степановича Попова и прекрасного итальянского инженера и организатора Гульельмо Маркони превратили электромагнитные волны, предсказанные гением англичанина Джеймса Кларка Максвелла и обнаруженные талантом немца Генриха Герца, в новое средство связи.

Появившаяся в самом конце прошлого века радиосвязь первое время рассматривалась как забавное развлечение. Казалось, она не способна составить серьезную конкуренцию проводной телеграфной связи, получившей к этому времени широкое распространение.

Однако очень скоро отношение к радиосвязи изменилось. И, как это часто бывает, первыми, кто понял незаменимость радиосвязи и неизбежность ее развития, были военные.

В 1900 г. неподалеку от небольшого острова Гогланд в Балтийском море потерпел аварию и сел на камни броненосец «Генерал-адмирал Апраксин». Это был один из нескольких русских военных кораблей, на которых были установлены приборы беспроводной связи, изобретенные и изготовленные преподавателем Минной школы в Кронштадте Александром Степановичем Поповым. Во время проведения спасательных работ между броненосцем и финским городом Кутсали на расстоянии 47 км была установлена и поддерживалась радиосвязь. Наличие связи сыграло решающую роль в спасении гибнущего броненосца. Случай с русским военным кораблем широко освещался в мировой печати. Это было одно из тех событий, которые заставили ученых всего мира работать над изобретением прибора, способного усиливать электрические сигналы и, следовательно, увеличивать дальность радиосвязи.

В 1906 г. американский исследователь Ли де Форест подал заявку на изобретенную им конструкцию электронной лампы — вакуумного триода. В последующие 10 лет конструкция триода, оставаясь в принципе неизменной, впитала в себя множество практических усовершенствований. Электронная лампа на последующие полвека сделалась элементной базой первой электронной революции. Впрочем, в начале века такие сухие выражения были не в ходу. Электронную лампу называли тогда королевой электроники.

Усилители и генераторы на электронных лампах опускались на дно океанов вместе с трансатлантическим телефонным кабелем и взлетали в небо с самолетными радиостанциями. В 1927 г. была создана первая промышленная телевизионная установка, а в середине 30-х годов миллионы зрителей могли наблюдать на телевизионных экранах Олимпийские игры, проходившие в Берлине, или репортажи с международной выставки в Нью-Йорке.

Электронные лампы на многие годы обеспечили прогресс новой науки — радиоэлектроники, уже совершенно необходимой физикам и связистам, медикам и военным, мореплавателям и криминалистам, людям десятков специальностей.

В свою очередь, массовый спрос на электронные лампы способствовал резкому улучшению их параметров. Во всем мире тысячи ученых и инженеров в десятках институтах и лабораториях занимались усовершенствованием

лампы. Для производства ламп конструировались специальные высокопроизводительные высокоточные автоматы, строились специализированные поточные линии и заводы. Лампы становились все миниатюрнее, экономичнее, надежнее, эффективнее и дешевле.

Однако шли годы, и жизнь ставила перед радиоэлектроникой, а радиоэлектроника — перед своей королевой все более сложные задачи.

Устройствам на электронных лампах нужны минуты для того, чтобы прийти в рабочее состояние — лампы должны прогреться. Между тем эти минуты могут решить судьбу самолета, потерявшего управление, или военного корабля, у которого отказал радиолокатор.

Срок службы лампы принципиально ограничен. Раскаленный катод лампы, испускающий электроны, постепенно портится — и лампа выходит из строя.

Пока электронная схема включает только несколько ламп, а срок службы каждой лампы достаточно велик, особых проблем не возникает. Но радиолокационные комплексы к концу второй мировой войны включали в себя тысячи ламп. А спроектированные и начавшие создаваться электронные вычислительные машины — десятки тысяч. Последствия выхода из строя лампы в устройстве, содержащем тысячи и десятки тысяч элементов, или в условиях, когда заменить ее очень трудно (например, в усилителе межконтинентального телефонного кабеля, лежащего на дне океана), могут быть очень серьезными.

Устройство, содержащее десятки тысяч ламп, требует для питания целой большой электростанции. Разработчикам электронных комплексов приходилось ломать голову над тем, как создавать новые, более сложные устройства при минимальном количестве использованных ламп. Например, одну и ту же лампу заставляли выполнять функции и усилителя, и детектора, и генератора сигналов.

На усложняющиеся требования ученые, конструкторы, технологи отвечали разработкой новых, более совершенных ламп. Королева, как это обычно и бывает, не проявляла никакого желания добровольно покинуть трон.

Выдвигались идеи создания ламп «с холодным катодом» — не требующих прогрева. Создавались серии сверхнадежных ламп с гарантированным сроком службы не менее 200 000 часов. Предлагались конструкции сверхминиатюрных и сверхэкономичных ламп.

А в это время...

Гадкий утенок

А в это время отдельные группы оторвавшихся от жизни ученых занимались мало понятной и никому не нужной деятельностью, пытаясь реализовать идеи некоего Лилиенфельда, уже четверть века отвергаемые самой жизнью (см. гл. 7), — создать твердотельный полупроводниковый усилитель электрических сигналов.

В конце концов, как мы знаем, в 1947 г. усилия одной из таких групп (У. Шокли, Дж. Бардин, У. Браттейн) увенчались успехом. Был создан биполярный транзистор.

Но, боже мой! Что это был за хилый и нежизнеспособный уродец!

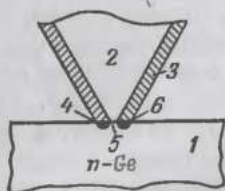


Рис. 132. Конструкция первого биполярного транзистора. В пластинку монокристаллического германия *n*-типа (1) с силой вдавливался пластмассовый треугольник (2), обернутый золотой фольгой (3). На вершине треугольника, вдавленной в пластинку, фольга разрезана бритвой. В месте соприкосновения фольги с поверхностью монокристалла образуются области *p*-типа: эмиттер (4) и коллектор (6), между ними — база (5)

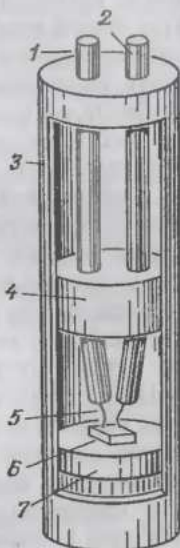


Рис. 133. Первый промышленный биполярный транзистор (1948 г.). Эмиттерный и коллекторный переходы изготавливались вплавлением в германиевый кристалл тонких проволоочек. Диаметр корпуса транзистора 1 см, высота 4 см: 1 — контакт к эмиттеру; 2 — контакт к коллектору; 3 — корпус; 4 — изолирующая прокладка; 5 — две контактные проволоочки; 6 — кристалл германия; 7 — контакт к базе

Конструкция первого биполярного транзистора показана на рис. 132. Понятно, какой надежности, воспроизводимости, устойчивости к внешним воздействиям, долговечности можно ожидать от подобной конструкции.

Через год, в 1948 г., в продажу поступили первые образцы промышленных биполярных транзисторов (рис. 133). Увы, с надежностью, воспроизводимостью, устойчивостью и т. д. дело обстояло немногим лучше, чем в экспериментальной конструкции, показанной на рис. 132. Достаточно было сильно встряхнуть такой «прибор» — и его коэффициент усиления изменялся в несколько раз. Изготавливая транзистор, никогда нельзя было заранее предсказать, какой коэффициент усиления он будет иметь. Стоило слегка повысить или понизить температуру — и транзистор переставал работать.

Где уж было этому уродцу состязаться с королевой, за плечами которой стояли тысячи слуг и 30 лет развития и совершенствования ламповой промышленности.

Гадкий утенок из всех сил старался понравиться и произвести впечатление. Практически одновременно с появлением первых промышленных образцов транзисторов были созданы и продемонстрированы опытные образцы транзисторных радиоприемников, телевизоров, даже усилителей для электрогитары.

Увы, на птичий двор все это не подействовало. Когда фирма «Bell Laboratories», где работали изобретатели транзистора, обратились к военным с запросом, не следует ли засекретить новый прибор, индюки-военные, только бросив взгляд на новорожденного, разрешили фирме публиковать сообщения о нем, где угодно.

Ни публичная демонстрация, ни широкая реклама не привлекли внимания специалистов к новому прибору. Тем не менее небольшие опытные партии транзисторов продолжали выпускаться.

И вскоре появились первые потребители — фирмы, изготавливавшие аппараты для людей с пониженным слухом. Одного достоинства от транзисторов нельзя было отнять — схемы на транзисторах можно было сделать очень миниатюрными. К тому же они питались от источников низкого (2—3 В) напряжения и транзисторный усилитель, размещенный в оправе очков, был совершенно безопасен. Фирма «Bell» продавала транзисторы для этой цели по чрезвычайно низким ценам — в честь Александра Белла, преподавателя школы глухонемых.

Следующими потребителями транзисторов все-таки стали военные. Уже в самом конце 40-х годов начались разработки систем, способных осуществить запуск искусственных спутников Земли. А каждый грамм полезного груза, выводимого на орбиту, в соответствии с рас-

четами, должен был стоить так дорого, что имело явный смысл повозиться хотя и с капризными, но гораздо более легкими, маленькими и экономичными приборами. Военные ведомства в средствах стеснены не были и действовали с присущим им размахом. Они оптом скупали большие партии транзисторов и подвергали их всесторонним и жестким испытаниям. Трясли на вибростендах, грели в термостатах, опускали в воду, испытывали при повышенных нагрузках. В результате отбирался один из 50—100 транзисторов — и на таких вот прошедших огонь и воду приборах собирались схемы.

Тем временем параметры транзисторов быстро улучшались. Появились сплавные транзисторы — достаточно надежные, с воспроизводимыми параметрами, устойчивые к внешним воздействиям. Были разработаны методы серийного производства транзисторов — и они заметно подешевели. Более дешевые и более надежные приборы охотнее покупались теперь уже не только военными, но и гражданскими фирмами, разрабатывавшими портативную радиоаппаратуру.

Транзисторы начали приносить прибыль. Часть этой прибыли можно было направить на разработку новых технологических процессов, создание высокопроизводительных поточных линий, разработку новых типов транзисторов. Круг потребителей более совершенных и более дешевых транзисторов стремительно расширялся.

Да здравствует король!

Через 6 лет после изобретения транзистора, в 1953 г., один из самых популярных в мире журналов по электронике, американский журнал «Electronics»*), писал: «Так как мы вступаем в век транзисторов, необходимо, чтобы инженеры поглубже проанализировали потенциальные возможности этих новых приборов, которые похожи и не похожи на лампы. Схемы могут быть разработаны, даже если представлять себе транзистор просто как элемент, заменяющий лампу. Но более важные схемы появятся, если направленно использовать принципиальные особенности транзисторов».

Период привыкания к транзисторам продолжался первую половину 50-х годов. В начале этого периода раз-

*) С 1961 г. этот журнал под названием «Электроника» переводится на русский язык,

рабатывались транзисторные схемы, действительно, повторявшие аналогичные схемы на лампах. С середины 50-х годов в высших учебных заведениях начинается чтение лекций по новому курсу «Полупроводниковые приборы и транзисторы».

К началу 60-х годов инженеры, наконец, почувствовали главные особенности транзисторов. Малое потребление энергии, ничтожные размеры и масса, высокая надежность в новых разработках используются в полной мере. В 1954 г. в мире выпускалось около 5 млн. транзисторов, в 1958 г. — 200 млн., в 1963 г. — около 1,5 миллиардов! В 1963 г. в мире выпускалось около 2500 типов диодов и 3000 типов транзисторов.

Старой королеве — электронной лампе, после полувекового царствования пришлось покинуть трон *).

Транзисторные усилители и генераторы составили элементную базу новых поколений телевизоров, радиоприемников, магнитофонов. В сочетании с другими полупроводниковыми элементами транзисторы служат точными термометрами и электронными весами. Высоковольтные транзисторы управляют мощными электродвигателями; разработаны специальные транзисторы для управления системой зажигания автомобиля. Сверхчувствительные сверхвысокочастотные транзисторы принимают сигналы искусственных спутников Земли и межпланетных кораблей. Портативные транзисторные аппараты передают по телефону кардиограмму больного врачу-специалисту прямо из дома; миниатюрные датчики контролируют пульс и давление, сообщают о состоянии внутренних органов человека.

За 10 лет, к началу 60-х годов, круг применений транзисторов расширился настолько, что простое перечисление этих применений могло бы заполнить несколько толстых фолиантов.

Одна из наиболее привлекательных возможностей, которые открылись перед электроникой благодаря высокой надежности и эффективности транзистора, была возможность создавать очень сложные схемы, содержащие десятки и даже сотни тысяч элементов. Транзисторы стали элементной базой ЭВМ второго поколения.

*) Однако и в наши дни ежегодно выпускаются сотни миллионов электровакуумных приборов. Не говоря уже об электронно-лучевых трубках телевизоров и осциллографов, лампы остаются незаменимыми там, где нужны большие мощности и очень высокие частоты.

Так, в 1967 г. в СССР была создана (и эксплуатируется до сих пор) одна из лучших машин этого поколения — БЭСМ-6. Эта ЭВМ содержит около 60 000 биполярных транзисторов и более 200 000 диодов и совершает приблизительно 1 млн. операций в секунду.

Такое резкое усложнение схем вызвало к жизни проблему, с которой никогда не приходилось иметь дело ламповой радиоэлектронике. Надежность электронной лампы была настолько ниже надежности любых других элементов радиоаппаратуры, что только числом и надежностью ламп определялась общая долговечность и надежность радиоаппаратуры. Другое дело транзистор!

Он не менее долговечен и надежен, чем любой другой элемент схемы: конденсатор, сопротивление, кусок провода, место соединения элементов (контакт) и т. д.

Жизнь, как и сорок лет назад, ставила перед электроникой новые проблемы, требовавшие создания еще более сложных схем. Но улучшение параметров транзисторов само по себе уже не могло помочь делу. Один оборвавшийся провод, одно бракованное сопротивление, одна плохая пайка (и попробуй найди их в суперсложной схеме)— и устройство не работает.

Король исчезает. Да здравствует король!

В 1952 г. в Вашингтоне проходила ежегодная конференция по элементам электронных схем. Разумеется, основная часть докладов была посвящена ламповым схемам. Но один из участников конференции, инженер из Великобритании Джеффри Даммер, выступил с докладом, никакого отношения к актуальным задачам электроники не имевшим. Он сказал: «С появлением транзистора и полупроводниковой техники вообще можно представить себе электронное оборудование в виде твердого блока, не содержащего соединительных проводов. Блок может состоять из слоев изолирующих, проводящих, выпрямляющих и усиливающих материалов, в которых определенные участки выделены таким образом, чтобы они могли непосредственно выполнять электрические функции».

В 1952 г. голос Даммера прозвучал как глас вопиющего в пустыне. Мало кого интересовали сами транзисторные схемы. А уж проблема их надежности просто никого не волновала.

Но уже через 10 лет, когда сборка компьютеров занимала месяцы и требовала квалифицированного труда тысяч людей, а одно ненадежное соединение грозило вывести из строя сложнейшую аппаратуру, все больше и больше людей...

Вспоминали слова Даммера?

Отнюдь нет. В 1952 г. выступление Даммера не привлекло никакого внимания и было забыто так прочно, что историки науки «раскопали» его только через четверть века. Просто все больше людей приходили к тем же выводам, которые Даммер пророчески предвидел.

К этому времени большинство транзисторов изготавливались уже методами диффузии и вплавления, по *планарной технологии*. В этом случае все транзисторы на полупроводниковой пластине изготавливаются одновременно. Зачем же, сделав транзисторные структуры, разрезать пластину на отдельные элементы, монтировать их в корпуса, приваривать проволочные выводы, делать еще много сложных и трудоемких операций, получать готовые транзисторы... а затем снова с громадными усилиями и сложностями соединять транзисторы в единую схему? Не естественней ли сразу на пластине полупроводника создать между транзисторами все необходимые соединения и сразу получить готовую схему?

Интегральную схему!

Первая интегральная схема (ИС) была выпущена в продажу в 1961 г. Она представляла собой транзисторный переключатель (триггер) и содержала 4 биполярных транзистора и 2 резистора — всего 6 элементов (рис. 134).

Через 2 года, в 1963 г., число элементов на кристалле ИС возросло до 10—20, а сам размер ИС уменьшился до 3—5 мм. К 1967 г. максимальное число элементов на кристалле составляло ~100, к 1970 г. — 1000, к 1975 г. — 30 000, к 1982 г. — 300 000!

На кристаллике кремния площадью несколько квадратных миллиметров разместилась схема, по сложности значительно превосходящая схемы первых ламповых ЭВМ.

С принципом изготовления интегральных схем мы, в сущности, уже знакомы. Он ничем не отличается от описанного в гл. 6 (с. 256) принципа одновременного (группового) изготовления на пластине кремния сразу нескольких десятков или сотен тысяч транзисторов методами планарной технологии. Только соответствующих технологических операций: окисления, вскрытия окон, совмещения новых масок со старыми и т. д. при изготовлении ИС приходится делать гораздо больше. После того как на пластине сформированы все $p-n$ переходы и транзисторы, они закрываются слоем окисла,

в окна вскрываются в форме полосок, соединяющих друг с другом по заданной схеме электроды транзисторов. Если электроды транзисторов должны быть непосредственно соединены между собой, в окна-полоски напыляют или осаждают металл. Часто для этой цели используется алюминий.

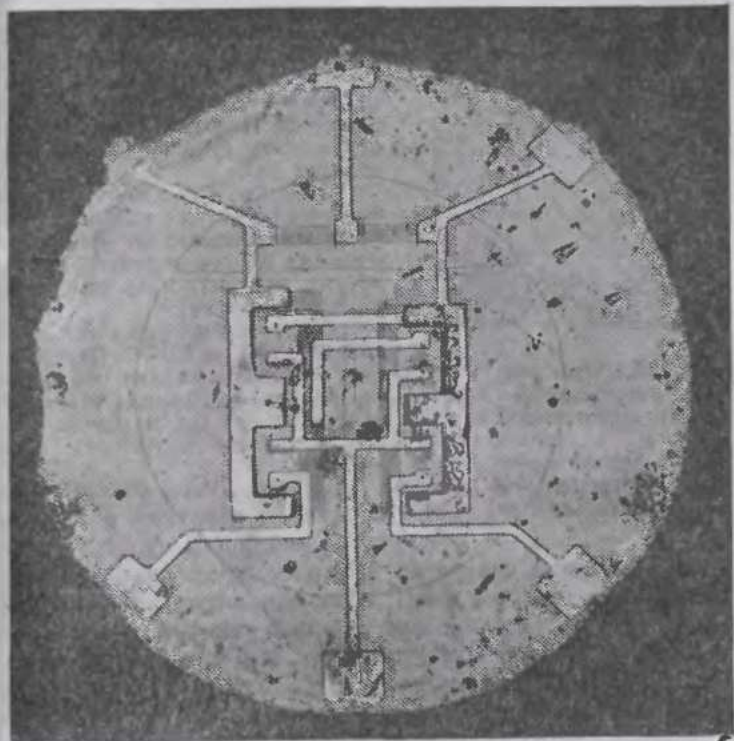


Рис. 134. Первая промышленная интегральная схема. Содержит 6 элементов — 4 биполярных транзистора и 2 резистора (сопротивления). Схема выполнена на пластине кремния диаметром ~ 1 см

Если между электродами должно быть включено сопротивление (резистор), в соответствующие окна-полоски осаждается слой материала с заданным удельным сопротивлением, определенной длины и толщины.

В качестве емкостей в ИС используются обратносмещенные $p-n$ переходы (с. 175). Что касается индуктивностей, то в ИС, по возможности, их стараются избегать.

При сильном увеличении поверхность кремниевой пластины с изготовленной на ней ИС выглядит примерно так, как показано на рис. 135. Электроды транзисторов, слой металлизации, защитный окисел образуют на поверхности сложный геометрический узор.

Изготовление ИС представляет собой очень сложный процесс, в котором заняты тысячи работников высокой квалификации. Проектирование ИС, разработка ее топологии, т. е. точного рисунка элементов и их соединений, изготовление многочисленных фотошаблонов — все это очень трудоемкие операции, требующие высокой точности.

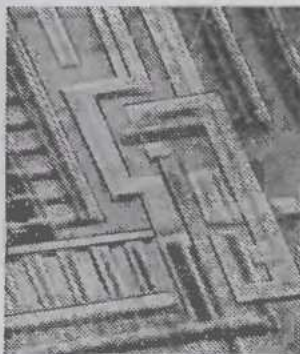


Рис. 135. Фрагмент большой интегральной схемы (БИС), выполненной на основе МДП транзисторов (длина канала самых больших элементов ~ 10 мкм, самых маленьких ~ 1 мкм)

Уровень интеграции, т. е. плотность размещения элементов ИС на пластине полупроводника, таков, что никакие, даже самые искусные человеческие руки не способны осуществить соединения между элементами ИС.

В процессе создания ИС руки человека ни разу не касаются пластины полупроводника. Весь цикл создания ИС осуществляется в замкнутых камерах, куда не попадает ни единая пылинка.

Современная технология создания ИС, БИС и СВИС (сверхбольших интегральных схем) обеспечивает, наряду с малыми размерами, исключительно высокую надежность сложных радиоэлектронных систем.

Первые ИС, содержавшие несколько десятков элементов, изображались на схемах так же, как и обычные схемы на дискретных элементах. Только элементы, входящие в ИС, обводили штриховой линией (рис. 136, а). Однако чем сложнее становились ИС, чем больше элементов они включали, тем проще становились их условные изображения в схемах рис. 136, б. Это изменение в обозначениях отражало очень сложные и глубокие изменения в мышлении инженеров, ученых, производственников — всех, кто имеет отношение к использованию ИС.

Отдельный транзистор перестал быть базовым элементом электроники. На его место пришла ИС.

Может показаться, что это утверждение никакой глубокой мысли не содержит и носит скорее рекламный характер. В самом деле, в основе ИС лежат все те же полевые или биполярные транзисторы. Так стоит ли огород городить и ломиться в открытую дверь?

В этом случае — стоит. Конструктор, проектирующий современную машину, понятия не имеет, сколько в ней винтов, гаек, прокладок и пружин. Думая о конструкции, он мыслит блоками: карбюратор, шасси, двигатель

и т. д. Их надежность, параметры, размеры... Музыкант, обдумывая или исполняя мелодию, не может мыслить отдельными нотами. И вопрос: «Сколько нотных знаков в 40-й симфонии Моцарта?» поставит в тупик самого опытного и квалифицированного музыканта.

Ученый, обдумывающий построение новой ЭВМ, инженер, проектирующий электронный блок, ремонтник, устраняющий неисправность в современной сложной электронной аппаратуре, не думают о транзисторах, составляющих ИС. Они мыслят на языке ИС: какие функции выполняет данная ИС; какие напряжения следует

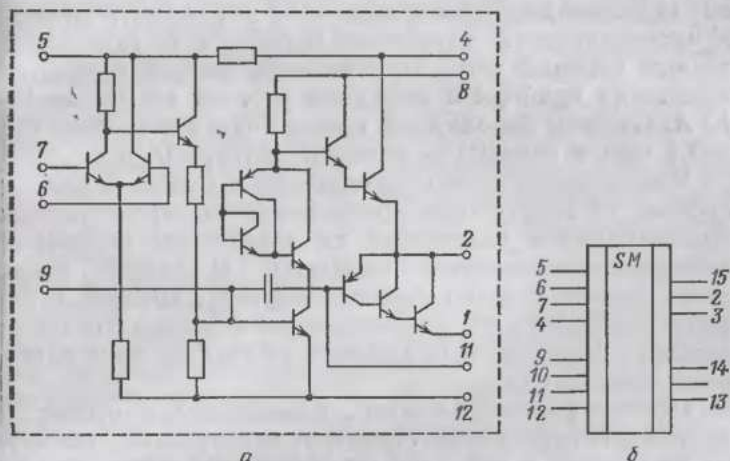


Рис. 136. В процессе усложнения ИС обозначения их в схемах становились все проще: а — принципиальная схема интегрального усилителя К 174 УН5, используемого в современных телевизорах и радиоприемниках; б — сверхскоростной сумматор К 500 ИМ 180 для быстродействующих ЭВМ

подать на ее выводы; какова надежность ИС, ее быстродействие и т. д.

Именно ИС, а не отдельный транзистор является элементной базой, основным функциональным элементом современной электроники.

Итак, король не сошел с трона. Он, в соответствии с лучшими традициями демократии, сделался одним из сотен тысяч граждан.

И тем не менее, да здравствует король!

Все-таки в основе любой ИС лежат транзисторы и улучшение параметров ИС, дальнейший прогресс электроники.

троники прежде всего связаны с улучшением параметров транзисторов — полевых и биполярных.

Совершенствование биполярных и полевых транзисторов идет двумя основными путями. Создаются новые технологические процессы, позволяющие получить транзисторы с возможно меньшей толщиной базы (биполярный транзистор) или возможно более коротким каналом (полевой транзистор). Идет поиск новых материалов с большой подвижностью носителей для изготовления транзисторов.

Современные технологические процессы, как мы знаем, позволяют под контролем ЭВМ наращивать слои полупроводникового материала толщиной в один-единственный атомный слой. Практически созданы и проходят испытания приборы с толщиной рабочих элементов 10—15 Å. Созданы биполярные транзисторы с толщиной базы ~0,1 мкм и полевые — с длиной затвора 0,15 мкм.

Что касается новых материалов с большой подвижностью, то в настоящее время наиболее перспективными представляются разработки на основе так называемых *селективно-легированных структур*. В основе работы таких структур лежат свойства барьера, который в этой книге нам взять не удалось, — так называемого гетеробарьера. Такой барьер возникает на границе двух разных полупроводников.

Однако физическая идея, положенная в основу селективно-легированных структур, так красива, что трудно не сказать о ней хотя бы несколькими словами.

Одной из самых главных, самых важных задач, на решение которой направлены усилия ученых, разрабатывающих новые типы транзисторов, является повышение их быстродействия, увеличение предельной рабочей частоты. В полевых транзисторах (работающих в режиме ненасыщенной скорости) предельная частота тем выше, чем больше подвижность (формула (69)). В биполярных транзисторах предельная частота возрастет с увеличением коэффициента диффузии D , т. е. в соответствии с соотношением Эйнштейна (24), $f_{пр}$ опять-таки увеличивается с ростом подвижности.

Подвижность носителей тока в полупроводниках, как мы знаем, зависит от среднего времени между актами рассеяния τ_0 . Чем больше τ_0 , чем реже столкновения, тем выше подвижность (формула (15)).

Одним из главных источников столкновений являются примеси, существующие в полупроводниковом кристалле. Чем меньше примесей содержит полупроводник, тем выше подвижность *). Таким

*) Подробнее об этом можно прочитать в книге: *Левинштейн М. Е., Симиц Г. С. Знакомство с полупроводниками*, — М.: Наука, 1984. — Библиотечка «Квант», вып. 33.

образом, казалось бы, ясно, в чем должен состоять главный путь усовершенствования полупроводниковых материалов — в максимальной очистке от примесей.

Дело, однако, обстоит не так просто. Для создания полевых и биполярных транзисторов требуется материал с довольно высокой концентрацией носителей ($\sim 2 \cdot 10^{17} \text{ см}^{-3}$). А свободные носители, как мы знаем, возникают при заданной температуре только при введении в полупроводник примеси. Получается заколдованный круг: хочешь иметь в полупроводнике желаемую концентрацию носителей — вводи примеси, а введешь примесь — понизится подвижность.

Идея селективно-легированных структур состоит в том, чтобы пространственно разделить примеси и носители, которые благодаря этим примесям возникают. Примесь (например, донорная) вводится в слой материала 1 (рис. 137). Возникшие благодаря введению этой примеси электроны переходят за счет диффузии в расположенную в непосредственной близости полосу материала 2, в которую примесь не вводится. И полоска материала 2, содержащая электроны, но не содержащая примеси, служит активным элементом транзистора. Например, канал полевого транзистора.

Разумеется, не любые два материала годятся для изготовления такого «бутерброда» или, как часто говорят, «сэндвич-структуры». Заслуга исследователей, предложивших структуру с селективным легированием, и состоит в том, что они поняли, какие именно материалы могут образовать пару с желаемыми свойствами. Ими оказались GaAs и тройное соединение GaAlAs*).

На основе селективно-легированных структур, использующих свойства гетероперехода GaAs/GaAlAs, уже созданы полевые транзисторы с уникальными свойствами. Время переключения их составляет $\sim 4 \text{ пс}$ ($4 \cdot 10^{-12} \text{ с}$), а энергия, потребляемая при переключении, — всего $\sim 10 \text{ фДж}$ (10^{-14} Дж).

Какой большой выигрыш в подвижности позволяет получить использование селективно-легированных структур, можно видеть из рис. 138.

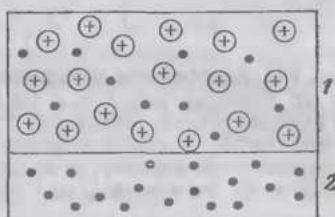


Рис. 137. Селективно-легированная структура. Атомы донорной примеси (обозначены \oplus) вводятся в материал 1 (GaAlAs). Они создают в этом материале свободные электроны (обозначены \bullet). Часть электронов переходит в полосу материала 2 (GaAs), где создается достаточно большая концентрация электронов с высокой подвижностью

*) Гетеропара GaAs — GaAlAs обладает и многими другими замечательными свойствами. Некоторые из них были обнаружены, исследованы и применены для создания полупроводниковых лазеров группой ученых Ленинградского физико-технического института им. А. Ф. Иоффе АН СССР под руководством академика Ж. И. Алферова,

На рис. 139 показан полевой транзистор, выполненный на основе селективно-легированной структуры. При

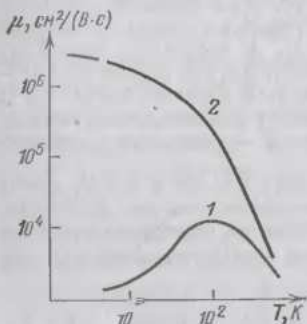


Рис. 138. Зависимость подвижности от температуры: 1 — в GaAs с концентрацией носителей $2 \times 10^{17} \text{ см}^{-3}$; 2 — в селективно-легированной структуре с концентрацией носителей $2 \cdot 10^{17} \text{ см}^{-3}$

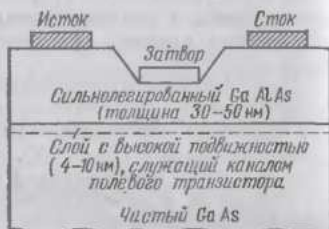


Рис. 139. Полевой транзистор на основе селективно-легированной структуры. На границе сильно легированного GaAlAs и GaAs возникает слой с высокой подвижностью

концентрации электронов в канале полевого транзистора $\sim 2 \cdot 10^{17} \text{ см}^{-3}$ подвижность их выше, чем в арсенид-галлиевом транзисторе, в ~ 2 раза при комнатной температуре (300 K) и в ~ 100 раз при температуре 80 K.

Претенденты на трон

Большинство специалистов в области физики полупроводников и радиоэлектроники сходятся на том, что в ближайшее десятилетие именно ИС на основе полевых и биполярных транзисторов будут составлять элементную базу современной электроники.

Однако и сегодня, как и полвека назад и как это всегда было в истории науки, отдельные группы ученых пытаются реализовать идеи, удачное воплощение которых обещает не улучшение существующих устройств, а создание устройств принципиально новых. Никто сегодня не может сказать, какие из этих идей окажутся ложными и тихо умрут, растеряв своих последователей, какие потребуют для своего воплощения еще многих десятилетий упорной работы, а какие уже в ближайшем будущем приведут к новой революции в электронике.

В заключение книги мы расскажем о нескольких таких идеях, которые сегодня кажутся наиболее перспективными.

Объемные интегральные схемы. Сегодня на кремниевой пластине размером 5×5 мм удается разместить до 500 000 транзисторов. Предполагается, что к концу 80-х годов на кремниевой (или арсенид-галлиевой) пластинке такого же размера будет размещаться несколько миллионов транзисторов. Однако для радиоэлектроники будущего, в частности для ЭВМ будущих поколений, даже такая громадная плотность упаковки окажется недостаточной. Поэтому уже сейчас ведутся работы над созданием интегральных схем нового типа — объемных ИС, каждая из которых должна содержать десятки и сотни миллионов полупроводниковых приборов. Объемная ИС напоминает слоеный пирог. Каждый слой этого пирога — уже знакомая нам плоская (планарная) ИС.

После того как заканчивается формирование обычной плоской ИС на полупроводниковой пластине, изготавливаются все транзисторы, соединения между ними, внешние выводы и т. д., сверху наращивается слой диэлектрика. Затем поверх диэлектрика наращивается слой полупроводника. На этом слое создается вторая ИС и т. д.

В корпус помещается теперь не пластинка, а кубик, содержащий много ИС. При этом соединения между плоскими ИС выполняются внутри кубика сразу в процессе изготовления объемной ИС. В результате число внешних соединений резко уменьшается. Возрастает надежность и уменьшается стоимость схемы.

Полупроводниковые элементы оптических ЭВМ. Одним из самых важных применений ИС, как мы знаем, является их применение в ЭВМ. В современных ЭВМ информация хранится и обрабатывается в виде электрических сигналов. Однако уже с середины 70-х годов начались работы по созданию ЭВМ совсем иного типа — оптических ЭВМ.

В современных ЭВМ «нулю» и «единице» — двум членам, с помощью которых записывается вся хранящаяся и перерабатываемая машиной информация, условно соответствуют закрытое и открытое состояния транзисторов. В проектируемых оптических ЭВМ «единице» соответствует прохождение света через оптическую ячейку, «нулю» — поглощение света в ячейке.

Самый простой образ, соответствующий «0» и «1» в оптических ЭВМ, очень нагляден — это потушенная и зажженная лампочка. Такой «лампочкой» будет служить в оптических ЭВМ полупроводниковый лазер — источник света, интенсивность которого можно изменять

с частотой в миллиарды герц. Однако полупроводниковые лазеры потребляют в сотни тысяч раз больше энергии, чем транзисторы. Поэтому ЭВМ, созданная на основе сотен миллионов лазеров, потребляла бы недопустимо большую мощность.

Предполагается, что луч полупроводникового лазера, введенный в световод, будет направляться к полупроводниковым оптическим ячейкам, прозрачность которых может быть изменена десятки миллиардов раз в секунду с малыми затратами энергии.

Биорадиоэлектроника. Природа предсказывает нам, что увеличение быстродействия — не единственный путь усовершенствования сложных радиоэлектронных устройств и, в частности, ЭВМ. Клетки человеческого мозга — нейроны, переключаются за время, приблизительно в миллион раз большее, чем транзисторы — «клетки» современных ЭВМ. Однако именно Человек изобрел и ЭВМ, и очень многие другие, не менее замечательные вещи.

Как устроен человеческий мозг, почему он (иногда!) работает с такой замечательной эффективностью — эти вопросы интенсивно изучаются и еще очень далеки от разрешения. Однако уже сегодня сделаны первые шаги по созданию биомолекулярных ИС. запатентован метод изготовления ИС на основе белкового материала — полилизина. Созданы молекулярные переключатели, белковые полевые транзисторы, биодатчики сигналов, срабатывающих при ядерных, биологических или химических воздействиях. Предполагается, что плотность упаковки элементов в таких биополимерных структурах приблизится к плотности упаковки структур, ответственных за фотосинтез в зеленых листьях растений, и составит более миллиарда на 1 мм^2 . Есть основания надеяться, что биомолекулярные ИС будут не изготавливать, а выращивать, как живую ткань.

Если эти надежды оправдаются, такие элементы уже не будут иметь отношения к полупроводникам.

Как можно видеть, передовые отряды современной радиоэлектроники ведут разведку в разных направлениях. Штурмовые группы идут по разным маршрутам, пытаются отыскать самые различные подходы к старым перевалам и разведать новые.

Но пока что, в ближайшее десятилетие, именно прогресс в изучении объемных свойств полупроводников и свойств барьеров определит прогресс в области электроники и, во многом, всей современной цивилизации.

ЗАКЛЮЧЕНИЕ

Мы вместе прошли тяжелый и сложный маршрут и преодолели один из самых интересных и трудных барьеров в физике полупроводников. С высоты покоренного нами перевала $p - n$ перехода, как мы убедились, открывается путь к глубокому пониманию многих современных приборов.

В физике полупроводников и полупроводниковой электронике важную роль играют еще два барьера: барьер между полупроводником и металлом и барьер между двумя разными полупроводниками. Если вы когда-нибудь попытаетесь их одолеть, вы убедитесь, что навыки, полученные при нашем совместном восхождении, очень облегчат вашу задачу.

До свидания. Желаем вам удачи и успехов!



Новичок



Разрядник



Мастер спорта

Михаил Ефимович Левинштейн
Григорий Соломонович Симин

БАРЬЕРЫ
ОТ КРИСТАЛЛА ДО ИНТЕГРАЛЬНОЙ СХЕМЫ

СЕРИЯ «БИБЛИОТЕЧКА «КВАНТ», вып. 65

Редактор *Л. А. Панюшкина*
Художественный редактор *Т. Н. Кольченко*
Технический редактор *И. Ш. Аксельрод*
Корректоры *Е. Ю. Рычагова, Н. Д. Храпко*

ИБ № 32499

Сдано в набор 30.07.87. Подписано к печати 23.10.87. Т-19441. Формат 84×108^{7/32}.
Бумага тип. № 3. Гарнитура обыкновенная. Печать высокая.
Усл. печ. л. 16,8. Усл. кр.-отт. 17,22. Уч.-изд. л. 17,99.
Тираж 134 000 экз. Заказ № 706 Цена 75 коп.

Ордена Трудового Красного Знамени издательство «Наука»
Главная редакция физико-математической литературы
117071 Москва В-71, Ленинский проспект, 15

2-я типография издательства «Наука», 121099 Москва, Шубинский пер., 6