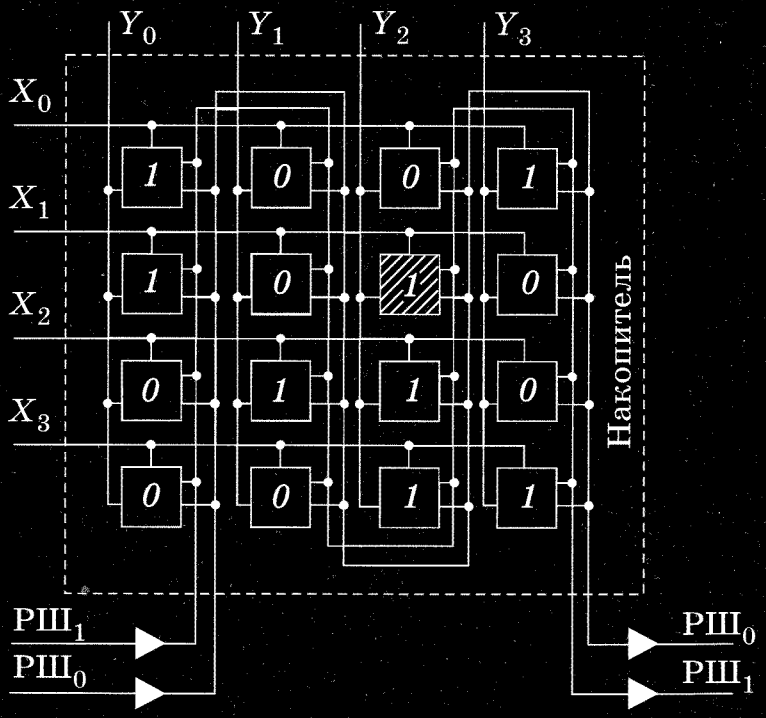
621.38
С 494

ОСНОВЫ МИКРОЭЛЕКТРОНИКИ

И. П. СТЕПАНЕНКО



УДК 621.3.049.77
~~ББК 32.852~~
С 79

Степаненко И. П. Основы микроэлектроники: Учеб. пособие для вузов. — 2-е изд., перераб. и доп. — М.: Лаборатория Базовых Знаний, 2001. — 488 с.: ил.

Рассмотрены основные аспекты микроэлектроники: физические, технологические и схемотехнические. Дается представление об уровне современной микроэлектроники, ее методах, средствах, проблемах и перспективах. Обсуждаются виды интегральных схем и схемотехника цифровых и аналоговых ИС. Во 2-м издании отражены новые фундаментальные достижения в области микроэлектроники, используемые в настоящее время на практике.

Предназначается для студентов радиотехнических и радиофизических специальностей вузов. Может быть полезна широкому кругу специалистов, связанных с созданием и эксплуатацией радиоэлектронной аппаратуры на ИС.

Второе издание данного учебного пособия подготовили:
А. В. Шальнов (гл.1, 2, 3),
В. С. Першенков (гл. 4, 5),
Л. Н. Патрикеев (гл. 6, 7, заключение),
В. В. Беляков (гл. 8, 9, 10)

По вопросам приобретения обращаться:
в Москве
«Лаборатория Базовых Знаний» (095) 955-03-98, e-mail: lbz@aha.ru
в Санкт-Петербурге
«Диалект» (812) 247-93-01, e-mail: dialect@sndlct.ioffe.rssi.ru

ISBN 5-93208-045-0

© Степаненко И.П., 2001
© Лаборатория Базовых Знаний, 2001

Учебное издание

И. П. Степаненко
Основы микроэлектроники

Художник *Н. Лозинская.* Технический редактор *Н. Журавлева.*
Компьютерная верстка *В. Носенко*

Подписано в печать 25.11.00. Формат 60x90$^{1}/_{16}$. Гарнитура Школьная.
Бумага офсетная. Печать офсетная. Усл. печ.л. 30,5.
Тираж 3000 экз. Заказ 3423

Издательство «Лаборатория Базовых Знаний», 2001 г. Адрес для переписки: 103473, Москва, а/я 9. Телефон (095)955-0398. E-mail: lbz@aha.ru

Лицензия на издательскую деятельность №066140 от 12 октября 1998 г.

Отпечатано с готовых диапозитивов в полиграфической фирме «Полиграфист». 160001, г. Вологда, ул. Челюскинцев, 3.

Содержание

Предисловие ко второму изданию 6

Предисловие к первому изданию 7

Глава 1. Предмет микроэлектроники 10

1.1. Введение 10
1.2. Интегральные схемы 13
1.3. Особенности интегральных схем как нового типа электронных приборов 18
1.4. Краткий исторический обзор 20
1.5. Заключение 23
Контрольные вопросы 24

Глава 2. Полупроводники 25

2.1. Введение 25
2.2. Структура полупроводников 25
2.3. Носители заряда 30
2.4. Энергетические уровни и зоны 33
2.5. Распределение носителей в зонах проводимости . . . 36
2.6. Эффект поля 45
2.7. Рекомбинация носителей 53
2.8. Законы движения носителей в полупроводниках . . . 63
Контрольные вопросы 71

Глава 3. Полупроводниковые переходы и контакты . . 73

3.1. Введение 73
3.2. Электронно-дырочные переходы 73
3.3. Контакты полупроводник–металл 89
3.4. Граница полупроводник–диэлектрик 95
Контрольные вопросы 97

Глава 4. Униполярные транзисторы 99

4.1. Введение 99
4.2. МДП-транзисторы 100
4.3. Полевые транзисторы 116
Контрольные вопросы 124

Глава 5. Физические принципы работы биполярного транзистора и тиристора 125

5.1. Введение 125
5.2. Принцип действия 125
5.3. Распределения носителей 130
5.4. Коэффициенты усиления тока 137

5.5. Статические характеристики 145
5.6. Малосигнальные эквивалентные схемы и параметры . 152
5.7. Переходные и частотные характеристики 155
5.8. Тиристор 163
Контрольные вопросы 167

Глава 6. Технологические основы микроэлектроники . 169

6.1. Введение 169
6.2. Подготовительные операции 169
6.3. Эпитаксия 171
6.4. Термическое окисление 173
6.5. Легирование 176
6.6. Травление 184
6.7. Техника масок 187
6.8. Нанесение тонких пленок 193
6.9. Металлизация 200
6.10. Сборочные операции 203
6.11. Технология тонкопленочных гибридных ИС 205
6.12. Технология толстопленочных гибридных ИС 211
Контрольные вопросы 214

Глава 7. Элементы интегральных схем 216

7.1. Введение 216
7.2. Изоляция элементов 218
7.3. Транзисторы n–p–n 227
7.4. Разновидности n–p–n-транзисторов 235
7.5. Транзисторы p–n–p 241
7.6. Интегральные диоды 244
7.7. Полевой транзистор 247
7.8. МДП-транзисторы 248
7.9. Полупроводниковые резисторы 258
7.10. Полупроводниковые конденсаторы 264
7.11. Элементы ИС на полупроводниках группы $A^{III}B^{V}$ . . 271
7.12. Элементы пленочных ИС 274
Контрольные вопросы 280

Глава 8. Основы цифровой схемотехники 282

8.1. Введение 282
8.2. Статический режим простейшего биполярного ключа 284
8.3. Переходные процессы в простейшем биполярном ключе 293
8.4. Ключ с барьером Шоттки 303
8.5. Переключатель тока 305
8.6. МДП-транзисторные ключи 310
8.7. Помехоустойчивость ключей 322
8.8. Бистабильные ячейки и триггеры 326
8.9. Триггер Шмитта 330
Контрольные вопросы 334

Глава 9. Основы аналоговой схемотехники 336
9.1. Введение 336
9.2. Составные транзисторы 337
9.3. Статический режим простейшего усилителя 339
9.4. Переходные процессы в простейшем усилителе . . . 347
9.5. Простейшие усилители на МДП-транзисторах 352
9.6. Дифференциальные усилители 357
9.7. Эмиттерные повторители 373
9.8. Каскод 385
9.9. Выходные каскады 386
9.10. Стабилизаторы напряжения 394
9.11. Стабилизаторы тока 400
Контрольные вопросы 409
Глава 10. Интегральные схемы 411
10.1. Введение 411
10.2. Логические элементы на биполярных транзисторах 412
10.3. Логические элементы на МДП-транзисторах 426
10.4. Логические элементы на совмещенных биполярных и МОП-транзисторах (БиКМОП) 434
10.5. Логические элементы на полевых транзисторах с управляющим переходом металл–полупроводник (МЕП) 435
10.6. Параметры логических элементов 436
10.7. Интегральные триггеры 440
10.8. Запоминающие устройства 446
10.9. Большие и сверхбольшие интегральные схемы 453
10.10. Операционные усилители 460
10.11. Надежность интегральных схем 473
10.12. Заключение 479
Контрольные вопросы 480
Заключение. Перспективы развития микроэлектроники 483
Литература 488

Предисловие ко второму изданию

Со времени выхода в свет первого издания учебного пособия для студентов вузов «Основы микроэлектроники» прошло почти двадцать лет. За эти годы труд Игоря Павловича Степаненко, скончавшегося в 1982 г., оказал добрую помощь нескольким поколениям выпускников русскоязычных вузов и студентам, осваивавшим основы микроэлектроники на английском и испанском языках. К сожалению, в настоящее время книга И. П. Степаненко стала библиографической редкостью.

И. П. Степаненко справедливо считал, что инженеру-физику, специализирующемуся по микроэлектронике, необходимо свободно ориентироваться в трех ее базовых составляющих: физических, технологических и схемотехнических основах микроэлектронных приборов и структур.

Методология изложения и фактический материал «Основ микроэлектроники» бережно сохранены в новом издании, поэтому мы считаем И. П. Степаненко основным автором книги. Настоящее издание подготовлено учениками и соратниками И. П. Степаненко, создавшего в 1965 г. в МИФИ первую в СССР кафедру микроэлектроники. Естественно, что данное издание — не простое повторение материала прошлых лет. В учебное пособие включены новые фундаментальные достижения в области микроэлектроники, используемые сейчас на практике.

В настоящем издании авторы старались учесть новые веяния в организации учебного процесса, в частности, свободное посещение студентами лекций. Так, в конце каждого раздела приведены контрольные вопросы для самостоятельной проработки.

Вся работа по подготовке нового издания выполнена сотрудниками кафедры микроэлектроники МИФИ под руководством профессора А. В. Шальнова.

В. В. Беляков, Л. Н. Патрикеев, В. С. Першенков, А. В. Шальнов

Предисловие к первому изданию

Микроэлектроника как исторический этап развития электроники характерна органическим единством физических, технологических (точнее, конструктивно-технологических) и схемотехнических аспектов. Поэтому полноценная, творческая разработка интегральных схем *невозможна ни силами «классических» физиков-технологов*, способных разрабатывать отдельные полупроводниковые приборы, но не знакомых или слабо знакомых со схемотехникой, ни силами «классических» схемотехников, способных разрабатывать функциональные узлы из элементарных радиодеталей, но слабо знакомых с физикой и тем более технологией полупроводниковых приборов. Специалист, работающий в области микроэлектроники, должен в равной степени владеть ее физическими, технологическими и схемотехническими основами, после чего, в зависимости от конкретных условий, он может получать дополнительную подготовку *в том* или ином направлении.

В настоящее время разработчики интегральных схем выпускаются в рамках нескольких инженерно-физических специальностей (0604, 0629 и др.)[1], сформировавшихся задолго до становления микроэлектроники. Поэтому в учебные планы этих специальностей в разное время были введены такие дополнительные дисциплины, как «Микроэлектроника», «Анализ и расчет интегральных схем», «Расчет и конструирование микроэлементов и микросхем» и др., призванные усовершенствовать профиль специалистов, привести его в соответствие с задачами, решаемыми в области микроэлектроники. Однако, как показывает опыт, ни одна из этих дисциплин не обеспечивает полноценного овладения схемотехническими основами микроэлектроники, а значит единство ее трех основополагающих аспектов остается нерешенной проблемой.

Другим возможным путем для выпуска разработчиков интегральных схем является усиление физико-технологической подготовки инженеров радиотехнического профиля. С этой целью, например, в учебный план специальности 0705 введен курс «Физические основы конструирования и технология РЭА», а курс «Электронные приборы» преобразован в «Электронные приборы и полупроводниковые интегральные схемы». Однако эти курсы не включают в себя схемотехнические

1 В данном предисловии номера специальностей указаны по данным на 1980 г.— год выхода в свет первого издания книги (*Прим. ред.*).

аспекты микроэлектроники, а в классических курсах, посвященных радиоэлектронным схемам, специфика этих аспектов не учитывается.

За последние десять лет для студентов, специализирующихся в области микроэлектроники, выпущено несколько учебных пособий. Не умаляя их достоинств и тем более приоритета, следует все же отметить, что они посвящены либо только физическим, либо физико-технологическим и конструктивным аспектам микроэлектроники. В гораздо меньшей степени они касаются элементов интегральных схем, а главное, опять же совсем не затрагивают схемотехнику.

В данном учебном пособии делается попытка более или менее полно охарактеризовать все составные части микроэлектроники. Особое внимание уделено тем разделам, которые не получили должного отражения в упомянутой учебной литературе: физике транзисторов, транзисторной схемотехнике, а также общей характеристике интегральных схем. Акцент на эти разделы в условиях ограниченного объема книги не позволил столь же подробно изложить конструктивно-технологические аспекты, вопросы метрики и некоторые др. Однако этот недостаток представляется не столь существенным, поскольку дополнительные сведения можно почерпнуть из упомянутой литературы.

Итак, в основу данного учебного пособия по замыслу положен системный подход к изложению материала: отдельные разделы взаимосвязаны, работают друг на друга и в комплексе создают тот общий фундамент, на базе которого возможна дальнейшая специализация студента в области микроэлектроники. Что касается инженеров, которые по роду работы должны преодолеть «односторонность» своей прежней подготовки, то они могут сосредоточить внимание соответственно на физико-технологических или схемотехнических разделах.

Руководствуясь системным подходом, автор включил в учебное пособие все необходимые разделы микроэлектроники, независимо от того, что некоторые из них являются предметами специальных курсов (например, «Физика полупроводников», «Полупроводниковые приборы», «Импульсные схемы», «Усилители» и т. п.). Такую точку зрения разделяют, как нам кажется, и авторы уже выпущенных пособий. Однако автор считал неправильным механически включать в основы микроэлектроники весь материал упомянутых курсов даже в сокращенном виде. Поэтому этот материал был не только тщательно отобран, но и переработан в соответствии с конкретными задачами микроэлектроники, а также с учетом внутренней взаимосвязи разделов.

Например, в гл.2 не включены такие сами по себе важные вопросы, как элементы квантовомеханической теории твердого тела, эффекты Гана и Джозефсона, размерные эффекты в тонких пленках и др. Из числа транзисторов в гл. 4 и 5 рассмотрены только маломощные транзисторы, составляющие основу современных интегральных схем. Что касается транзисторных схем, то их номенклатура резко сокращена и приведена в соответствие с реальными конфигурациями, используе-

мыми в микроэлектронике. Изменена также традиционная последовательность при изучении двух основных классов схем: сначала (гл. 8) рассматриваются ключевые схемы, а затем (гл. 9) — усилительные.

Не менее сложная задача встала при отборе материала и выработке методики его изложения в гл. 10, посвященной реальным интегральным схемам. Здесь, во избежание неоправданных сложностей, пришлось во многих случаях отказаться от рассмотрения полных принципиальных схем и сосредоточить внимание на структурных схемах. Такой подход соответствует методам проектирования современных БИС и тем более аппаратуры на их основе.

«Основы микроэлектроники» — первое из трех учебных пособий, предназначенных для всесторонней подготовки инженеров-разработчиков интегральных схем, т. е. специалистов, хорошо знакомых со всеми тремя аспектами микроэлектроники[1].

Автор надеется, что его замысел, реализованный в данной книге, будет одобрен студентами, заинтересованными кафедрами, а также специалистами, которые воспользуются ею для самоподготовки. Более того, он надеется, что со временем учебный курс «Основы микроэлектроники», построенный в соответствии со структурой данной книги, станет таким же традиционным, как курсы «Полупроводниковые приборы» или «Теоретические основы электротехники». Это, несомненно, способствовало бы как единообразию подготовки специалистов одного и того же профиля, так и взаимопониманию специалистов смежных профилей («элементщиков» и «аппаратурщиков»), неизбежно вступающих в контакт при разработке сложных микроэлектронных систем.

Предлагаемое учебное пособие рассчитано прежде всего на студентов, специализирующихся по микроэлектронике в рамках специальностей 0604, 0629, 0648, 0701, 0705 и др., а также на инженеров, окончивших вузы по этим специальностям более 7–10 лет назад. Оно, вероятно, будет полезно специалистам по радиотехнике, вычислительной технике, кибернетике, автоматике, измерительной технике и другим областям науки и техники, связанным с применением интегральных схем.

Автор выражает искреннюю благодарность проф. С. Я. Шацу, проф. В. Н. Дулину и доц. Ю. Е. Наумову за полезные замечания и советы, а также коллективу кафедры микроэлектроники МИФИ за помощь в работе и оформлении рукописи.

И. П. Степаненко

---

1 Учебные пособия: *Березин А. С., Мочалкина О. Р.* «Технология и конструирование интегральных микросхем», изд. «Радио и связь», 1983 г. и *Алексенко А. Г., Шагурин И. И.* «Микросхемотехника», изд. «Радио и связь», 1990 г. (*Прим. ред.*)

Глава 1

Предмет микроэлектроники

1.1. Введение

Роль микроэлектроники в современной науке и технике трудно переоценить. Она справедливо считается катализатором научно-технического прогресса. Спектр ее применения простирается от фундаментальных исследований до прикладного использования. Микроэлектроника влияет на все народное хозяйство, но не непосредственно, а через целый ряд специфических отраслей, таких как вычислительная техника, информационно-измерительные системы, робототехника, микропроцессоры. Микроэлектроника, очередной исторически обусловленный этап развития электроники и одно из ее основных направлений, обеспечивает принципиально новые пути решения назревших задач.

Электроника — это область науки, техники и производства, охватывающая исследование и разработку электронных приборов и принципов их использования.

Микроэлектроника — это раздел электроники, охватывающий исследования и разработку качественно нового типа электронных приборов — интегральных микросхем — и принципов их применения.

Интегральная микросхема (или просто интегральная схема) есть совокупность, как правило, большого количества взаимосвязанных компонентов (транзисторов, диодов, конденсаторов, резисторов и т.п.), изготовленная в едином технологическом цикле (т.е. одновременно), на одной и той же несущей конструкции — *подложке* — и выполняющая определенную функцию преобразования информации.

Термин «интегральная схема» (ИС)[1] отражает факт объединения (интеграции) отдельных деталей — компонентов — в конструктивно единый прибор, а также факт усложнения вы-

---

1 Integrated Circuit (IC)

полняемых этим прибором функций по сравнению с функциями отдельных компонентов.

Компоненты, которые входят в состав ИС и тем самым не могут быть выделены из нее в качестве самостоятельных изделий, называются *элементами* ИС или *интегральными элементами*. Они обладают некоторыми особенностями по сравнению с транзисторами и т.д., которые изготавливаются в виде конструктивно обособленных единиц и соединяются в схему путем пайки.

В основе развития электроники лежит непрерывное усложнение функций, выполняемых электронной аппаратурой. На определенных этапах становится невозможным решать новые задачи старыми средствами или, как говорят, на основе старой *элементной базы*, например с помощью электронных ламп или дискретных транзисторов. Основными факторами, лежащими в основе смены элементной базы, являются: надежность, габариты и масса, стоимость и мощность.

Особенностью изделий микроэлектроники является высокая степень сложности выполняемых функций, для чего создаются схемы, в которых количество компонентов исчисляется миллионами. Отсюда ясно, что обеспечить надежность функционирования при соединении компонентов вручную — задача невыполнимая. Единственным способом ее решения является применение качественно новых высоких технологий.

Для изготовления интегральных схем используется групповой метод производства и планарная технология.

Групповой метод производства заключается в том, что, во-первых, на одной пластине полупроводникового материала одновременно изготавливается большое количество интегральных схем; во-вторых, если позволяет технологический процесс, то одновременно обрабатываются десятки таких пластин. После завершения цикла изготовления ИС пластина разрезается в двух взаимно-перпендикулярных направлениях на отдельные кристаллы[1], каждый из которых представляет собой ИС.

Планарная технология — это такая организация технологического процесса, когда все элементы и их составляющие создаются в интегральной схеме путем их формирования через плоскость[2].

---

1 По-английски chip (чип).

2 По-английски плоскость — plane.

таллов X на пластине можно достичь двумя путями: увеличением размера пластины и уменьшением размеров отдельных элементов. Эти оба направления используются разработчиками.

В заключение заметим, что все константы, входящие в формулу, не являются ни постоянными, ни зависимыми друг от друга, поэтому анализ на минимум стоимости на самом деле является сложным и многофакторным.

1.2. Интегральные схемы

В процессе развития микроэлектроники (МЭ) номенклатура ИС непрерывно изменялась. Главный тип ИС в настоящее время — полупроводниковые ИС. Им и будет посвящено основное содержание этого пособия.

Классификация ИС. Классификация ИС может производиться по различным признакам, ограничимся здесь лишь одним. По способу изготовления и получаемой при этом структуре различают два принципиально разных типа интегральных схем: полупроводниковые и пленочные.

Полупроводниковая ИС — это микросхема, элементы которой выполнены в приповерхностном слое полупроводниковой подложки (рис. 1.2). Эти ИС составляют основу современной микроэлектроники.

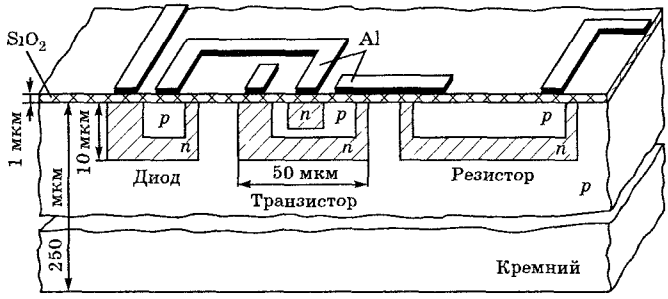


Рис. 1.2. Структура элементов полупроводнниковой ИС

Пленочная ИС — это микросхема, элементы которой выполнены в виде разного рода пленок, нанесенных на поверхность диэлектрической подложки (рис. 1.3). В зависимости от способа нанесения пленок и связанной с этим их толщиной различают *тонкопленочные* ИС (толщина пленок до 1–2 мкм) и *толстопленочные* ИС (толщина пленок от 10–20 мкм и выше).

Одна или несколько технологических операций при изготовлении ИС заключается в соединении отдельных элементов в схему и присоединении их к специальным контактным площадкам. Поэтому необходимо, чтобы выводы всех элементов и контактные площадки находились в одной плоскости. Такую возможность обеспечивает планарная технология.

Финальная операция — *корпусирование* — это помещение ИС в корпус с присоединением контактных площадок к ножкам ИС (рис. 1.1).

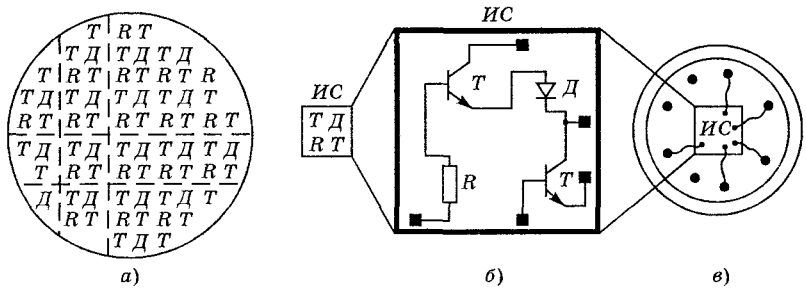


Рис. 1.1. Иллюстрация изготовления простейшей интегральной схемы: *а* — кремниевая пластина с «комплектами» из двух транзисторов, диода и резистора; *б* — межсоединения элементов внутри «комплекта» (планарная технология); *в* — готовая ИС в корпусе

Стоимость D одной ИС (одного кристалла) упрощенно можно вычислить следующим образом:

$$D = \frac{1}{XY}\left\{\frac{A+B}{Z} + C\right\},$$

где A — затраты на научно-исследовательские и опытно-конструкторские работы по созданию ИС; B— затраты на технологическое оборудование, помещение и др.; C — текущие расходы на материалы, электроэнергию, заработную плату, в пересчете на одну пластину; Z — количество пластин, изготовляемых до амортизации основных производственных фондов; X — количество кристаллов на пластине; Y — отношение годных ИС к количеству, запущенному в производство в начале его.

Кроме очевидных комментариев относительно затрат, нужно отметить следующее. Увеличение Y достигается созданием все более современной технологии, пожалуй, наиболее сложной и чистой среди многих новейших производств. Роста числа крис-

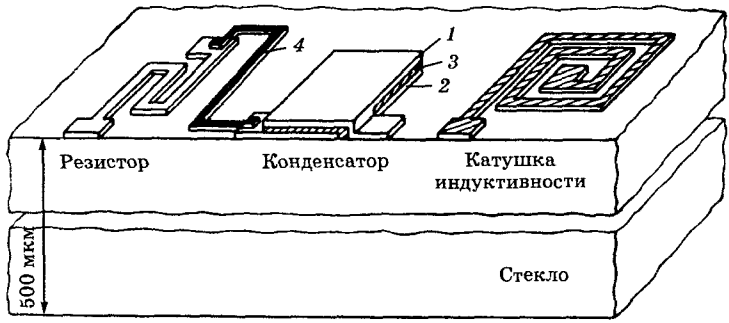


Рис. 1.3. Структура элементов пленочной ИС:
1 — верхняя обкладка; *2* — нижняя обкладка; *3* — диэлектрик; *4* — соединительная металлическая полоска

Поскольку до сих пор никакая комбинация напыленных пленок не позволяет получить активные элементы типа транзисторов, пленочные ИС содержат только пассивные элементы (резисторы, конденсаторы и т.п.). Поэтому функции, выполняемые чисто пленочными ИС, крайне ограничены. Чтобы преодолеть эти ограничения, пленочную ИС дополняют активными компонентами (отдельными транзисторами или ИС), располагая их на той же подложке и соединяя с пленочными элементами. Тогда получается ИС, которую называют гибридной.

Гибридная ИС (или ГИС) — это микросхема, которая представляет собой комбинацию пленочных пассивных элементов и активных компонентов, расположенных на общей диэлектрической подложке. Дискретные компоненты, входящие в состав гибридной ИС, называют *навесными*, подчеркивая этим их обособленность от основного технологического цикла получения пленочной части схемы.

Еще один тип «смешанных» ИС, в которых сочетаются полупроводниковые и пленочные интегральные элементы, называют совмещенными.

Совмещенная ИС — это микросхема, у которой активные элементы выполнены в приповерхностном слое полупроводникового кристалла (как у полупроводниковой ИС), а пассивные нанесены в виде пленок на предварительно изолированную поверхность того же кристалла (как у пленочной ИС).

Совмещенные ИС выгодны тогда, когда необходимы высокие номиналы и высокая стабильность сопротивлений и емко-

стей; эти требования легче обеспечить с помощью пленочных элементов, чем с помощью полупроводниковых.

Во всех типах ИС межсоединения элементов осуществляются с помощью тонких металлических полосок, напыленных или нанесенных на поверхность подложки и в нужных местах контактирующих с соединяемыми элементами. Процесс нанесения этих соединительных полосок называют *металлизацией*, а сам «рисунок» межсоединений — *металлической разводкой*.

Полупроводниковые ИС. В настоящее время различают следующие полупроводниковые ИС: биполярные, МОП (металл-окисел-полупроводник) и БИМОП. Последние представляют собой сочетание первых двух, и в них комбинируются положительные их качества.

Технология полупроводниковых ИС основана на легировании полупроводниковой (кремниевой) пластины поочередно донорными и акцепторными примесями, в результате чего под поверхностью образуются тонкие слои с разным типом проводимости и p–n-переходы на границах слоев. Отдельные слои используются в качестве резисторов, а p–n-переходы — в диодных и транзисторных структурах.

Легирование пластины приходится осуществлять локально, т.е. на отдельных участках, разделенных достаточно большими расстояниями. Локальное легирование осуществляется с помощью специальных *масок* с отверстиями, через которые атомы примеси проникают в пластину на нужных участках. При изготовлении полупроводниковых ИС роль маски обычно играет пленка двуокиси кремния SiO_2, покрывающая поверхность кремниевой пластины. В этой пленке специальными методами гравируется необходимая совокупность отверстий различной формы или, как говорят, необходимый *рисунок* (рис. 1.4). Отверстия в масках, в частности в окисной пленке, называют *окнами*.

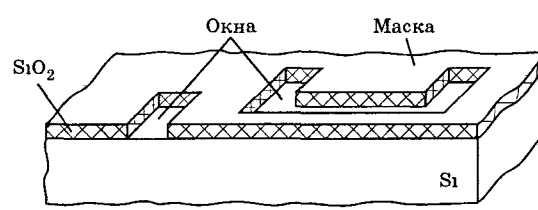


Рис. 1.4. Окисная маска с окнами для локального легирования

Теперь кратко охарактеризуем составные части (элементы) полупроводниковых ИС. Основным элементом биполярных ИС является n–p–n-транзистор: на его изготовление ориентируется весь технологический цикл. Все другие элементы должны изготавливаться, по возможности, одновременно с этим транзистором, без дополнительных технологических операций.

Основным элементом МДП ИС является МДП-транзистор. Изготовление других элементов также подстраивается под базовый транзистор.

Элементы биполярной ИС необходимо тем или иным способом изолировать друг от друга с тем, чтобы они не взаимодействовали через кристалл.

Элементы МОП ИС не нуждаются в специальной изоляции друг от друга, так как взаимодействие между смежными МОП-транзисторами не имеет места. В этом — одно из главных преимуществ МОП ИС по сравнению с биполярными.

Характерная особенность полупроводниковых ИС состоит в том, что среди их элементов отсутствуют катушки индуктивности и, тем более, трансформаторы. Это объясняется тем, что до сих пор не удалось использовать в твердом теле какое-либо физическое явление, эквивалентное электромагнитной индукции. Поэтому при разработке ИС стараются реализовать необходимую функцию без использования индуктивностей, что в большинстве случаев удается. Если же катушка индуктивности или трансформатор принципиально необходимы, их приходится использовать в виде навесных компонентов.

Размеры кристаллов у современных полупроводниковых ИС достигают 20×20 мм^2. Чем больше площадь кристалла, тем более сложную, более многоэлементную ИС можно на нем разместить. При одной и той же площади кристалла можно увеличить количество элементов, уменьшая их размеры и расстояния между ними.

Функциональную сложность ИС принято характеризовать *степенью интеграции*, т.е. количеством элементов (чаще всего транзисторов) на кристалле. Максимальная степень интеграции составляет 10^6 элементов на кристалле. Повышение степени интеграции (а вместе с нею и сложности функций, выполняемых ИС) — одна из главных тенденций в микроэлектронике.

Для количественной оценки степени интеграции используют условный коэффициент $k = \lg N$. В зависимости от его значения интегральные схемы называются по-разному:

$k \le 2$ ($N \le 100$) — интегральная схема (ИС);

$2 \le k \le 3$ ($N \le 1000$) — интегральная схема средней степени интеграции (СИС);

$3 \le k \le 5$ ($N \le 10^5$) — большая интегральная схема (БИС);

$k > 5$ ($N > 10^5$) — сверхбольшая интегральная схема (СБИС).

Ниже приведены английские обозначения и их расшифровки:

IC — Integrated Circuit;

MSI — Medium Scale Integration;

LSI — Large Scale Integration;

VLSI — Very Large Scale Integration.

Кроме степени интеграции, используют еще такой показатель, как *плотность упаковки* — количество элементов (чаще всего транзисторов) на единицу площади кристалла. Этот показатель, который характеризует главным образом уровень технологии, в настоящее время составляет до 500–1000 элементов/мм^2.

Гибридные ИС. Пленочные, а значит, и гибридные ИС в зависимости от технологии изготовления делятся на толсто- и тонкопленочные.

Толстопленочные ГИС (обозначим их ТсГИС) изготавливаются весьма просто. На диэлектрическую пластинку-подложку наносят *пасты* разного состава. Проводящие пасты обеспечивают межсоединения элементов, обкладки конденсаторов и выводы к штырькам корпуса; резистивные — получение резисторов; диэлектрические — изоляцию между обкладками конденсаторов и общую защиту поверхности готовой ГИС. Каждый слой должен иметь свою конфигурацию, свой рисунок. Поэтому при изготовлении каждого слоя пасту наносят через свою маску — *трафарет* — с окнами в тех местах, куда должна попасть паста данного слоя. После этого приклеивают навесные компоненты и соединяют их выводы с *контактными площадками.*

Тонкопленочные ГИС (обозначим их ТкГИС) изготавливаются по более сложной технологии, чем ТсГИС. Классическая тонкопленочная технология характерна тем, что пленки осаждаются на подложку из газовой фазы. Вырастив очередную пленку, меняют химический состав газа и тем самым электрофизические свойства следующей пленки. Таким образом пооче-

редно получают проводящие, резистивные и диэлектрические слои. Конфигурация (рисунок) каждого слоя определяется либо трафаретом, как в случае ТсГИС, либо маской, подобно окисной маске в полупроводниковых ИС (см. рис. 1.4).

Навесные элементы в ТкГИС, как и в ТсГИС, приклеивают на поверхность готовой пленочной части схемы и соединяют с соответствующими контактными площадками элементов.

Степень интеграции ГИС не может оцениваться так же, как в случае полупроводниковых ИС. Тем не менее существует термин *большая ГИС* (или БГИС), который означает, что в состав ГИС в качестве навесных компонентов входят не отдельные транзисторы, а целые полупроводниковые ИС.

1.3. Особенности интегральных схем как нового типа электронных приборов

Поскольку ИС, подобно транзистору, представляет собой конструктивное единое целое, выполняет определенную функцию и должна удовлетворять определенным требованиям при испытаниях, поставках и эксплуатации, она относится к разряду электронных приборов. Однако по сравнению с диодом, транзистором и т.п. ИС является качественно новым типом прибора.

Первая — главная особенность ИС как электронного прибора состоит в том, что она самостоятельно выполняет законченную, часто весьма сложную функцию, тогда как элементарные электронные приборы выполняют аналогичную функцию только в ансамбле с другими компонентами. Например, отдельный транзистор не может обеспечить усиление сигнала или запоминание информации. Для этого нужно из нескольких транзисторов, резисторов и других компонентов собрать (спаять) соответствующую схему. В микроэлектронике же указанные функции выполняются одним прибором — интегральной схемой: она может быть усилителем, запоминающим устройством и т.п.

Второй важной особенностью ИС является то, что повышение функциональной сложности этого прибора по сравнению с элементарными не сопровождается ухудшением какого-либо из основных показателей (надежность, стоимость и т.п.). Более того, все эти показатели улучшаются.

Поскольку габариты и масса простых и средних ИС близки к габаритам и массе дискретных транзисторов, можно считать,

что в первом приближении выигрыш по этим показателям при переходе от дискретных схем к интегральным определяется степенью интеграции и может достигать сотен и тысяч раз[1].

Поскольку надежность работы полупроводникового прибора в аппаратуре определяется прежде всего количеством паяных и (в меньшей степени) сварных соединений, то ИС, у которых межсоединения элементов осуществляются путем металлизации (т.е. без пайки и сварки), обладают заведомо повышенной надежностью по сравнению с дискретными схемами, выполняющими ту же функцию. По мере увеличения степени интеграции этот выигрыш возрастает.

Поскольку все элементы ИС изготавливаются в едином технологическом цикле, то количество технологических операций по их изготовлению не намного превышает количество операций по изготовлению отдельного транзистора. Поэтому стоимость ИС при прочих равных условиях близка к стоимости одного транзистора. Значит, в зависимости от степени интеграции (или, точнее, от плотности упаковки), стоимость одного элемента ИС по сравнению со стоимостью аналогичного дискретного компонента может быть в сотни раз меньше. Такое же соотношение имеет место между стоимостью ИС и стоимостью аналогичной схемы, выполненной на дискретных компонентах.

Третья особенность ИС состоит в предпочтительности активных элементов перед пассивными — принцип, диаметрально противоположный тому, который свойствен дискретной транзисторной технике. В последней активные компоненты, особенно транзисторы, наиболее дорогие, и потому оптимизация схемы при прочих равных условиях состоит в уменьшении количества активных компонентов. В ИС дело обстоит иначе: у них задана стоимость не элемента, а кристалла; поэтому целесообразно размещать на кристалле как можно больше элементов с минимальной площадью. Минимальную площадь имеют активные элементы — транзисторы и диоды, а максимальную — пассивные. Следовательно, оптимальная ИС — это ИС, у которой сведены к минимуму количество и номиналы резисторов и, особенно, конденсаторов.

---

1 Точный выигрыш в габаритах и массе трудно оценить теоретическим расчетом, так как ИС имеют другие типоразмеры корпусов и большее число выводов, чем элементарные компоненты.

Четвертая особенность ИС связана с тем, что смежные элементы расположены друг от друга на расстоянии, измеряемом в мкм или долях мкм. На таких малых расстояниях различие электрофизических свойств материала маловероятно, а следовательно, маловероятен и значительный разброс параметров у смежных элементов. Иначе говоря, параметры смежных элементов взаимосвязаны — коррелированы. Эта корреляция сохраняется и при изменении температуры: у смежных элементов температурные коэффициенты параметров практически одинаковы. Корреляция между параметрами смежных элементов используется при проектировании некоторых ИС с целью снизить влияние разброса параметров и изменений температуры.

Гибридные ИС тоже представляют собой тип электронных приборов. Однако наличие навесных компонентов делает их менее специфичными, чем полупроводниковые ИС. Остается в силе фундаментальная черта всякой ИС — функциональная сложность прибора, что, как и при использовании полупроводниковых ИС, качественно меняет структуру электронной аппаратуры. Спецификой ГИС как прибора могут быть либо высокие номиналы резисторов и конденсаторов, недостижимые в полупроводниковых ИС, либо прецизионность резисторов, либо, наконец, повышенная функциональная сложность. ГИС — это гибкий, дешевый, оперативно проектируемый тип ИС, хорошо приспособленный к решению специальных, частных задач.

1.4. Краткий исторический обзор

История микроэлектроники — часть истории технического прогресса последних лет. Участники ее — в первую очередь творческие работники — ученые, изобретатели и инженеры. Результаты развития микроэлектроники в значительной степени увеличили возможности науки и техники, ускорили многие процессы, связанные с научными исследованиями и разработкой новых высоких технологий, существенно повысили производительность труда во многих отраслях промышленности, фантастически расширили возможности информационных систем и оказали серьезное влияние на прогресс развития человечества в целом. Иногда используется даже новое название современного периода истории человечества — век информатизации, и в этом тоже огромный вклад микроэлектроники.

Первым стартовым шагом к микроэлектронике, несомненно, был переход от электронных вакуумных ламп, в которых применялся принцип управления электронными потоками в вакууме, к твердотельным приборам, где использовалось управление движением подвижных носителей в полупроводнике. Этот шаг был сделан в лаборатории *Bell Telephone Laboratories* в 1948 г. и заключался в изобретении транзистора.

Конкуренция между вакуумными электронными лампами и транзисторами сопровождалась бурными дебатами среди специалистов и, в первую очередь, связана с большим количеством успешно действующей электронной аппаратуры, главным элементом которой была электронная лампа. Производственный цикл такой аппаратуры был хорошо отлажен и она приносила большую прибыль производителям. Еще две основные причины способствовали этому. Главная заключалась в несовершенстве технологии изготовления транзисторов, приводящей к нестабильности характеристик прибора, недостаточной его надежности и высокой стоимости. Другая связана с неприятием нового прибора большинством специалистов, воспитанных на применении в электронике вакуумных ламп. Недостатки первых образцов транзисторов укрепляли уверенность радиоинженеров в правильности их сомнений относительно транзисторов.

Первые транзисторы были точечными и, в дополнение к указанным выше недостаткам, обладали высоким уровнем шумов. Материалом для их изготовления был, в первую очередь, германий. Переход от точечных транзисторов к плоскостным по времени совпал с переходом от германия к кремнию и был осуществлен в 1953 г. на фирме *Texas Instrument Incorporation*. Кремниевый транзистор с выраженными p–n-переходами позволил увеличить мощность транзистора. Позднее процесс изготовления транзистора, основанный на диффузии, позволил наладить групповое производство транзисторов, что немедленно сказалось на уменьшении стоимости этих приборов. В 1956 г. трем американским ученым, Шокли, Бардину и Браттейну, была присуждена Нобелевская премия по физике за исследования полупроводников и открытие транзисторного эффекта. К этому времени уже стало ясно, что во многих отраслях применение транзисторов сильно потеснит или даже заменит электронную лампу.

Второй — решающий шаг — появление интегральных схем. Первая интегральная (биполярная) схема была изготовлена в

1961 г. на фирме *Fairchild Semiconductor* и представляла собой триггер, состоящий из четырех биполярных транзисторов и двух резисторов. Уже в 1963 г. фирмой RCA была выпущена первая МОП интегральная логическая схема, в которой было 16 МОП-транзисторов. Быстрому распространению интегральных схем способствовала хорошо отработанная технологическая база при групповом производстве транзисторов.

Таким образом, с момента изобретения транзистора до изготовления первых интегральных схем, поступивших на рынок, прошло чуть более 10 лет и далее последовало очень быстрое развитие микроэлектроники.

В начале своего пути транзисторы применялись, главным образом, для военных целей, так же как, впрочем, и предшествовавшие им электронные лампы, и последующие за ними интегральные схемы. Кроме того, они способствовали развитию техники радиовещания и телевидения, а впоследствии и компьютерной техники уже для более широкого применения.

Было бы, однако, совершенно неправильным отдать полностью приоритет технологии в микроэлектронике, хотя, несомненно, именно в технологии было занято наибольшее количество сотрудников и вложены наибольшие капиталы. На самом деле понимание физических процессов в электронике начиналось с фундаментальных работ прошлого века, выполненных такими известными учеными как Фарадей и Максвелл в области эксперимента и теории электричества и магнетизма; Попов и Маркони по передаче сигналов через эфир; Флеминг, получивший в начале нашего века патент на электронную лампу, и последующего широкого развития связи, радиовещания и телевидения. Все эти работы подготовили появление новых отраслей промышленности и рынков сбыта для продуктов ее производства.

Для понимания и развития работ в области физики твердого тела исключительную роль сыграли исследования Макса Планка, опубликовавшего в 1906 г. квантовую гипотезу, объяснявшую поведение электронов в твердом теле. Для количественного объяснения этого же явления послужило квантово-механическое уравнение Э. Шредингера, которое было им опубликовано в 1926 г. Таким образом, теоретическое понимание электрических явлений и процессов в полупроводниках предшествовало конкретному применению этих процессов, воплощенных в приборы в виде интегральных схем.

1.5. Заключение

Первые этапы развития микроэлектроники были характерны главным образом прогрессом в области технологии ИС. На этих этапах совершенствовались методы изоляции элементов, методы повышения степени интеграции, способы монтажа навесных компонентов и т.п. Что касается схемотехники (т.е. конфигурации схем, подлежащих интеграции), то на первых порах она заимствовалась из арсенала дискретной транзисторной электроники.

Однако вскоре стало ясно, что качественно новой технологической реализации, свойственной ИС, должны соответствовать адекватные схемные решения. Далеко не все схемы, считавшиеся типичными в дискретной транзисторной электронике, оказались приемлемыми в микроэлектронике. И наоборот, многие схемы, которые в дискретной транзисторной электронике считались «экзотическими» и не имели широкого распространения, в микроэлектронике оказались приемлемыми и даже оптимальными. Поэтому схемотехника ИС отнюдь не совпадает с обычной транзисторной схемотехникой.

В процессе развития микроэлектроники появилось немало специфических элементов ИС, которые не имеют аналогов в транзисторной схемотехнике и не выпускаются в качестве дискретных полупроводниковых приборов (например, многоэмиттерный транзистор, приборы с зарядовой связью и др.). Интегральные схемы, в которых используются такие специфические элементы, не могут быть даже промоделированы на дискретных компонентах.

Все сказанное говорит о том, что микроэлектроника как область науки и техники ни в коем случае не сводится к технологии ИС. Она интегрирует в себе три равнозначных аспекта: *физический, технологический и схемотехнический.* Знание этих трех аспектов микроэлектроники позволяет разработчику ИС трезво оценивать как новые варианты элементной базы или схемных решений (с точки зрения их технологической реализации), так и новые варианты технологических процессов (с точки зрения их пригодности для реализации данных элементов и данных схем).

Контрольные вопросы

1. Что такое интегральная схема?
2. В чем состоят основные преимущества приборов, выполненных на ИС, по сравнению с приборами, выполненными на дискретных схемах?
3. Поясните смысл понятия «микроэлектроника».
4. Дайте классификацию ИС.
5. Какие сложности возникают при применении полупроводниковых ИС?
6. Что такое совмещенная ИС?
7. Как характеризуется функциональная сложность ИС?
8. Как характеризуют уровень технологии изготовления ИС?
9. Дайте определение понятия «биполярная ИС».
10. Дайте определение МДП ИС.
11. Что такое гибридная ИС?
12. Каким образом в ИС достигается высокая надежность?
13. Каковы особенности схемотехнических решений в микроэлектронике?
14. Что называют функциональной сложностью ИС?
15. Перечислите три главных аспекта развития микроэлектроники и покажите связь между ними.
16. В чем сущность группового метода производства?
17. Что достигается увеличением размера кристалла и пластины?
18. В чем заключаются недостатки планарной технологии?

Глава 2

Полупроводники

2.1. Введение

В первом приближении полупроводники выделяют из других веществ по значению удельного электрического сопротивления ρ. Считают, что удельное сопротивление металлов менее 10^{-4} Ом·см, полупроводников — в диапазоне от 10^{-3} до 10^{9} Ом·см, диэлектриков — более 10^{10} Ом·см. Такая чисто количественная классификация весьма условна, особенно применительно к полупроводникам и диэлектрикам, между которыми по существу нет принципиальных различий. Что касается полупроводников и металлов, то главное различие между ними заключается в том, что у металлов удельное сопротивление возрастает с ростом температуры, а у полупроводников — падает.

Из числа полупроводников наиболее подходящим для изготовления интегральных схем оказался *кремний*. Он быстро вытеснил германий. Поэтому ниже в большинстве примеров и иллюстрациях будут использоваться электрофизические параметры кремния. В последнее время все большее количество ИС изготавливается из арсенида галлия.

2.2. Структура полупроводников

Площадь современных ИС порядка 10–500 мм^2, площади их элементов порядка 10^{-2}–10^{-4} мм^2, а линейные размеры отдельных электродов доходят до 0,1 мкм. Ясно, что в пределах таких площадей и расстояний исходная полупроводниковая пластина должна быть достаточно однородной и обладать контролируемыми свойствами. Если же имеются дефекты и неоднородности, то они должны быть локализованы и их должно быть достаточно мало. Тогда в брак уйдет ограниченное количество ИС, которые оказались расположенными на этих дефектных участках. Поэтому однородности и бездефектности полупроводников в микроэлектронике уделяется весьма большое внимание.

Кристаллическая решетка. Полупроводники, как правило, — твердые тела с регулярной кристаллической структурой — *монокристаллы*. Их кристаллическая решетка состоит из множества повторяющихся и примыкающих друг к другу *элементарных ячеек* той или иной формы и размера. В случае простейшей кубической решетки (Ge, NaCl и др.) ребро элементарной ячейки — куба — есть *постоянная решетки a* (0,4–0,6 нм). Кубическая *решетка типа алмаза* (Si, Ge) состоит из тетраэдров (рис. 2.1); расстояние между смежными атомами около 0,25 нм.

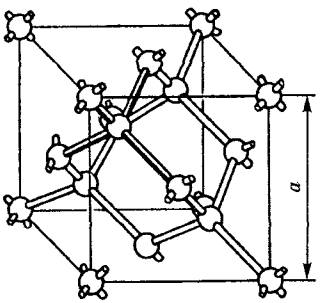


Рис. 2.1. Структура кристаллической решетки типа алмаза

Связь атомов в кристаллической решетке кремния и ряда других полупроводников обусловлена специфическими обменными силами, возникающими в результате попарного объединения валентных электронов у смежных атомов. Такая связь (при которой каждый из атомов остается нейтральным) называется *ковалентной* или просто *валентной*.

Регулярность (периодичность) структуры кристалла приводит к зависимости его свойств от направления в кристаллической решетке — к *анизотропии*. Оценивать направление, т.е. «ориентироваться» в кристаллической решетке принято с помощью *кристаллографических плоскостей*. Эти плоскости обозначают трехзначными *индексами Миллера*. Для обозначения индексы Миллера заключают в круглые скобки: (111), (100) и т.п.

Происхождение индексов Миллера показано на рис. 2.2, *а* применительно к простейшей кубической решетке. Отрезки, отсекаемые данной плоскостью на осях координат, измеряют в единицах постоянной решетки: $x = la$, $y = ma$, $z = na$, где l, m, n — целые числа. Затем обратные величины l^{-1}, m^{-1}, n^{-1} приводят к общему наименьшему знаменателю и знаменатель отбрасывают; тогда числители образуют индексы Миллера для данной плоскости.

Заметим, что каждой кристаллографической плоскости свойственна своя плотность атомов на единицу площади. Например, если «посмотреть» на кристалл с кубической решеткой перпендикулярно плоскостям (100), (110) и (111), то расположение атомов в поле зрения будет таким, как показано на

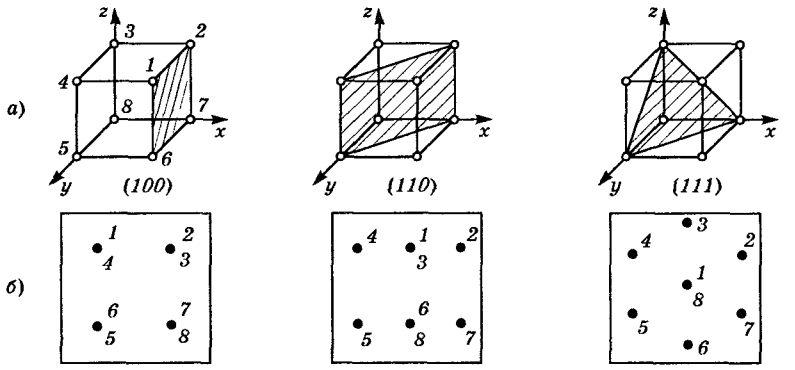


Рис. 2.2. Кристаллографические плоскости: *а* — происхождение индексов Миллера; *б* — расположение атомов в кристаллографических плоскостях

рис. 2.2, *б* (для ясности узловые атомы пронумерованы). Наибольшая плотность атомов соответствует плоскости (111), наименьшая — плоскости (100). У кремния плоскость (111) является *плоскостью спайности*: по ней, как правило, распространяются трещины и происходит раскалывание кристалла.

Для разных кристаллографических плоскостей оказываются разными многие свойства и параметры кристалла: оптические свойства, скорость травления и др. Поэтому пластины для изготовления ИС шлифуют точно по заранее заданной кристаллографической плоскости.

Искажения кристаллической решетки. Структура кристалла никогда не бывает идеальна — ни в объеме, ни тем более на поверхности. Всегда имеются дефекты решетки и дислокации.

Дефекты решетки могут иметь вид пустого узла (*дефект по Шоттки*) или совокупности пустого узла и междуузельного атома (*дефект по Френкелю*). Это — дефекты точечного типа (рис. 2.3, *а*, *б*).

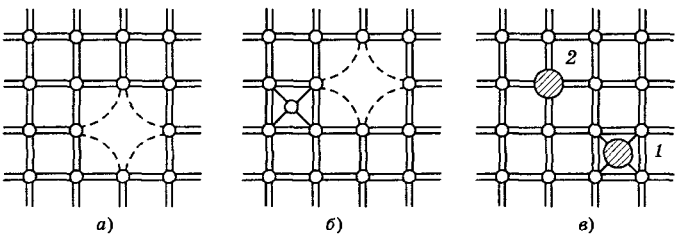


Рис. 2.3. Точечные дефекты кристаллической решетки: *а* — дефект по Шоттки; *б* — дефект по Френкелю; *в* — примесные дефекты

Любой реальный полупроводник содержит примеси — либо паразитные, от которых не удается избавиться при очистке, либо полезные, которые вводятся специально для получения нужных свойств кристалла. Каждый примесный (т.е. чужеродный) атом равносилен точечному дефекту решетки. Примесные атомы (рис. 2.3, *в*) могут располагаться либо в междуузлиях решетки (*примесь внедрения — 1*), либо в самих узлах — вместо основных атомов (*примесь замещения — 2*). Последний вариант более распространен.

Дислокации, т.е. смещения плоскостей решетки, бывают *линейные* (краевые) и *винтовые* (спиральные). Первые — результат неполного (не по всей глубине) сдвига решетки: появляется незаконченная полуплоскость атомов (рис. 2.4, *а*). Вторые — результат полного (по всей глубине) сдвига некоторого участка решетки (рис. 2.4, *б*).

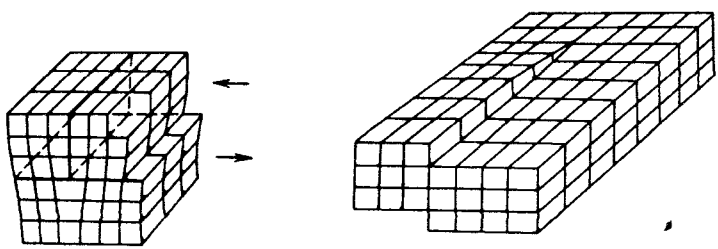


Рис. 2.4. Дислокации в кристаллической решетке:
а — линейные; *б* — винтовые

Наличие дислокаций приводит к дефектам ИС. Поэтому количество дислокаций на пластине полупроводника ограничивают.

Предельным случаем беспорядочных дислокаций можно считать *поликристалл*, состоящий из множества монокристаллических зерен (микрокристаллов) с разной ориентацией, тесно примыкающих друг к другу. В поликристаллах отсутствует регулярность структуры и свойственная ей анизотропия свойств.

Поэтому поликристаллы не стали основой для наиболее ответственных — активных элементов ИС и играют в микроэлектронике вспомогательную роль.

Кроме поликристаллических (зернистых) твердых тел, существуют *аморфные*, т.е. совершенно однородные, бесструктурные. Из-за плохой воспроизводимости и стабильности свойств

аморфные полупроводники на практике находят лишь узкоспециальное применение.

Помимо дислокаций, в пластинах полупроводника имеют место макроскопические дефекты: микротрещины, поры (пузырьки) и т.п. Все это — потенциальные причины брака в ИС.

Поверхность кристалла. У атомов, расположенных на поверхности кристалла, часть ковалентных связей неизбежно нарушается из-за отсутствия «соседей» по другую сторону границы раздела. Количество нарушенных связей зависит от кристаллографической ориентации поверхности. Например, для кремния в плоскости (111) оказывается оборванной одна из четырех связей, а в плоскости (100) — две (рис. 2.5).

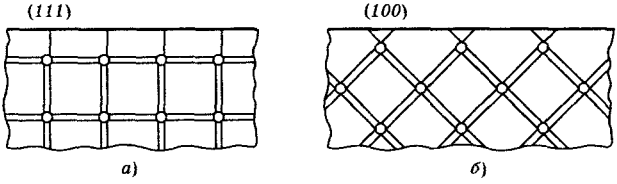


Рис. 2.5. Нарушение ковалентных связей на поверхности кристалла: *а* — в плоскости (100), *б* — в плоскости (111)

Нарушение ковалентных связей влечет за собой нарушение энергетического равновесия на поверхности. Равновесие восстанавливается разными путями: может измениться расстояние между атомами в приповерхностном слое, т.е. структура элементарных ячеек кристалла; может произойти захват — *адсорбция* — чужеродных атомов из окружающей среды, которые полностью или частично восстановят оборванные связи; может образоваться химическое соединение (например, окисел), не имеющее незаполненных связей на поверхности, и т.п. В любом случае *структура тонкого приповерхностного слоя* (толщиной несколько нанометров и менее) *отличается от структуры основного объема кристалла.*

Как следствие, электрофизические параметры приповерхностного слоя заметно отличаются от параметров объема, причем этот вывод не зависит от того, граничит ли кристалл с вакуумом, воздушной средой или другим твердым телом. Поэтому *приповерхностный или граничный слой* (часто говорят просто — поверхность или границу) *следует рассматривать как особую область кристалла.* Эта область играет важную роль в

микроэлектронике, поскольку элементы планарных ИС расположены непосредственно под поверхностью, а размеры рабочих областей часто соизмеримы с толщиной граничных слоев.

Поверхность кристалла, разумеется, может быть загрязнена самыми различными веществами: остатками кислот или щелочей, использованных при ее обработке, жировыми пятнами и т.п.

2.3. Носители заряда

Электропроводность имеет место в том случае, когда имеются свободные носители заряда, которые могут перемещаться под действием электрического поля или градиента концентрации. Рассмотрим происхождение свободных носителей заряда в полупроводниках.

Беспримесный и бездефектный полупроводник с идеальной кристаллической решеткой называют *собственным полупроводником*. При температуре абсолютного нуля в нем нет свободных носителей заряда и он является идеальным изолятором. По мере нагрева в кристалле возникает колебательное движение узловых атомов решетки. В корпускулярной интерпретации носителями энергии механических колебаний решетки являются квантовые частицы *фононы* — аналоги световых фотонов.

С повышением температуры кристалла количество и энергия фононов возрастают и они разрывают ковалентные связи между атомами решетки. При этом одновременно образуются свободные электроны и незаполненные связи — *дырки* вблизи того атома, от которого оторван электрон (рис. 2.6). Процесс образования электронно-дырочных пар под действием фононов носит название *термогенерации*.

Незаполненная связь заполняется одним из валентных электронов смежного атома. На месте этого электрона образуется новая дырка, и этот процесс повторяется. Следовательно, *дырка ведет себя подобно частице с положительным элементарным зарядом*.

Это очень удобное представление позволяет рассматривать электрический ток в полупроводнике состоящим из двух компонент — электронного и дырочного, при этом используются одинаковые, давно известные из электротехники выражения. Правда, некоторые параметры носителей заряда оказываются разными.

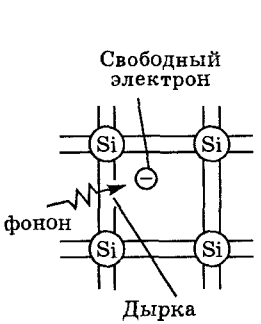


Рис. 2.6. Процесс образования дырки

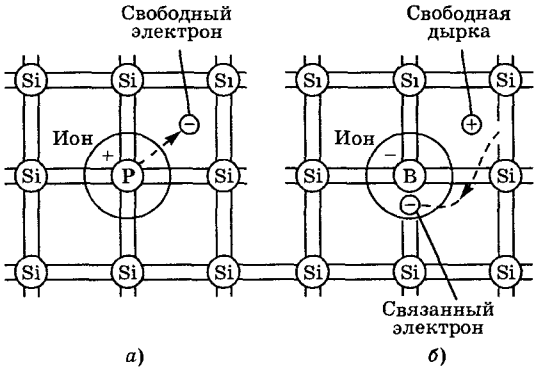


Рис. 2.7. Замещение примесными атомами основных атомов решетки:
а — донорная примесь (образуются свободный электрон и неподвижный положительный ион);
б — акцепторная примесь (образуются свободная дырка и неподвижный отрицательный ион)

Нарушение ковалентных связей и генерация пар электрон–дырка могут происходить как под действием фононов, так и других квантов энергии, например, под действием света, рентгеновских и γ-лучей. Специфика этих факторов по сравнению с нагревом состоит лишь в том, что их действие локальное: оно ограничено по глубине проникающей способностью, а по поверхности — площадью пучка. Иначе говоря, *облучение полупроводника эквивалентно его локальному нагреву*. Если площадь пучка превышает размеры кристалла и кристалл достаточно тонкий (прозрачный для излучения), то результат облучения по существу такой же, как при нагреве.

Проводимость собственного полупроводника, обусловленную парными носителями теплового происхождения, называют *собственной проводимостью*.

Проводимость, обусловленную наличием примесных атомов, называют *примесной проводимостью*. Примеси, характерные для кремния, являются примесями замещения.

Если ввести в кремний атом пятивалентного элемента (например, фосфора, сурьмы или мышьяка), то четыре из пяти валентных электронов этого элемента вступят в связь с четырьмя электронами соседних атомов кремния (рис. 2.7, *а*) и образуют устойчивую оболочку из восьми электронов. Девятый электрон

в этой комбинации оказывается слабо связанным с ядром пятивалентного элемента; он легко отрывается фононами и делается свободным. При этом примесный атом превращается в неподвижный ион с единичным положительным зарядом.

Свободные электроны примесного происхождения добавляются к собственным свободным электронам. Поэтому проводимость полупроводника становится преимущественно электронной. Такие полупроводники называются *электронными* или *n типа*. Примеси, обусловливающие электронную проводимость, называют *донорными*.

Если ввести в кремний атом трехвалентного элемента (например, бора, галлия или алюминия), то все три его валентных электрона вступят в связь с четырьмя электронами соседних атомов кремния (рис. 2.7, *б*).Для образования устойчивой восьмиэлектронной оболочки нужен дополнительный электрон. Таковым оказывается один из валентных электронов, который отбирается от ближайшего атома кремния. В результате образуется незаполненная связь — дырка, а атом примеси превращается в неподвижный ион с единичным отрицательным зарядом.

Дырки примесного происхождения добавляются к собственным дыркам, так что проводимость полупроводника становится преимущественно дырочной. Такие полупроводники называют *дырочными* или *p-типа*[1]. Примеси, обусловливающие дырочную проводимость, называют *акцепторными*.

Отрыв «лишнего» электрона от донора и «недостающего» электрона для акцептора требует некоторой энергии — *энергии ионизации* или *активации* примеси. Поэтому при нулевой температуре $T = 0$ К ионизация не имеет места, но в рабочем диапазоне температур (от −60 °С и выше) и, в частности, при комнатной температуре ($T = 15$–25 °С) примесные атомы ионизированы практически полностью.

Поскольку в примесных полупроводниках концентрации электронов и дырок резко различны, принято называть носителей преобладающего типа *основными*, а носителей другого

1 Буквы *n* (от *negative*) и *p* (от *positive*) общеприняты в полупроводниковой физике и технике для обозначения величин, связанных соответственно с электронами и дырками. Для величин, характеризующих собственный полупроводник, используется буква *i* (от *intrinsic*).

типа — *неосновными*. В полупроводнике n-типа основные носители — электроны, а в полупроводнике p-типа — дырки.

2.4. Энергетические уровни и зоны

Количественный анализ полупроводников и полупроводниковых приборов базируется на зонной теории твердого тела.

Зонная структура. Твердое тело представляет собой множество атомов, *сильно взаимодействующих* друг с другом благодаря малым межатомным расстояниям. Поэтому вместо совокупности дискретных энергетических уровней, свойственных отдельному атому, твердое тело характеризуется совокупностью *энергетических зон*.

Верхняя разрешенная зона называется зоной проводимости, нижняя — валентной зоной. В полупроводниках и диэлектриках они разделены запрещенной зоной. Отличие диэлектриков от полупроводников состоит главным образом в значительно большей ширине запрещенной зоны. При нулевой абсолютной температуре валентная зона всегда полностью заполнена электронами, тогда как зона проводимости либо заполнена только в нижней части, либо полностью пуста. Первый случай свойствен металлам, второй — полупроводникам и диэлектрикам.

При температуре, отличной от абсолютного нуля, ситуация несколько изменяется.

Энергетические диаграммы на рис. 2.8 построены для энергии электрона. Когда энергия электрона увеличивается, электрон занимает более высокое положение в зонной диаграмме. Если же говорить об увеличении энергии дырки, то это будет соответствовать, очевидно, продвижению дырки вглубь валентной зоны. Энергия электрона и дырки измеряется в электрон-вольтах (эВ).

Ширина запрещенной зоны равна

$$\varphi_з = \varphi_c - \varphi_v, \qquad (2.1)$$

где φ_c и φ_v — соответственно энергетические уровни для зоны проводимости и потолка валентной зоны.

На рисунке 2.9 показаны основные параметры зонных диаграмм полупроводников для температуры, отличной от абсо-

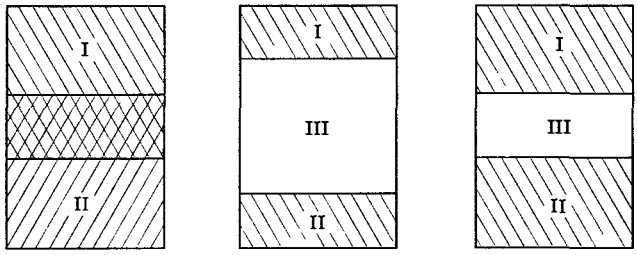


Рис. 2.8. Зонные диаграммы металла, диэлектрика и полупроводника: *а* — металл (зоны перекрыты и даже небольшая добавка энергии приводит к движению электронов); *б* — диэлектрик (очень большой энергетический зазор между зонами, проводимость невозможна); *в* — собственный полупроводник (расстояние между зонами не очень велико — появляются электроны в зоне проводимости и равное количество дырок в валентной зоне, возможна конечная проводимость); I — зона проводимости, II — валентная зона, III — запрещенная зона

лютного нуля. Ширина запрещенной зоны зависит от температуры:

$$\varphi_з = \varphi_{з0} - \varepsilon_з T, \qquad (2.1а)$$

где $\varphi_{з0}$ — ширина зоны при $T = 0$,

T — абсолютная температура,

$\varepsilon_з$ — температурная чувствительность.

Для кремния $\varepsilon_з = 3 \cdot 10^{-4}$ В/°С, а при комнатной температуре $\varphi_з \approx 1{,}11$ В.

Энергию, соответствующую середине зоны, называют электростатическим потенциалом проводника

$$\varphi_E = \tfrac{1}{2}(\varphi_c + \varphi_v). \qquad (2.1б)$$

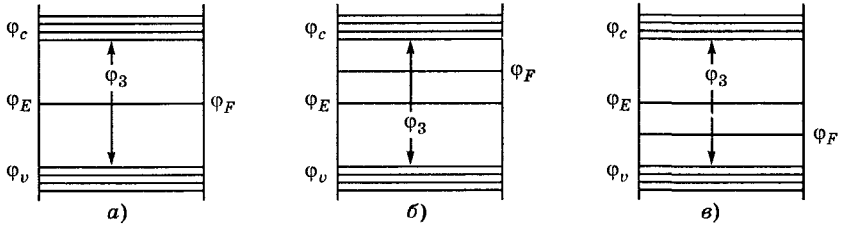


Рис. 2.9. Значения энергий в зонной диаграмме для а) собственного; б) электронного; в) дырочного полупроводников; φ_F — уровень Ферми (см. разд. 2.5)

Приведенные выше качественные соображения относительно примесных полупроводников могут быть проиллюстрированы еще раз качественно на зонных диаграммах. Электрические уровни примесей показаны на зонных диаграммах ⊕ (положительно заряженные ионы-доноры) и ⊖ (отрицательно заряженные ионы-акцепторы), рис. 2.10.

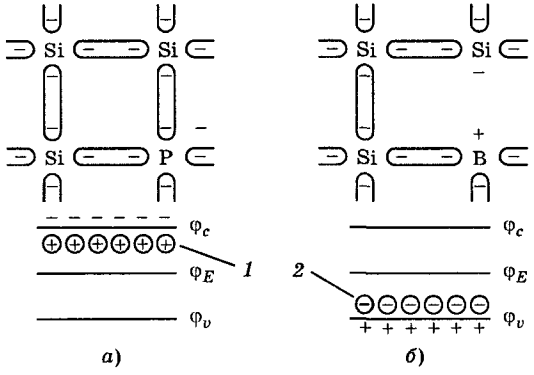


Рис. 2.10. Схематическое атомистическое изображение и зонная диаграмма примесных полупроводников:
а — полупроводник *n*-типа, *б* — полупроводник *p*-типа (1 — ионы доноров; 2 — ионы акцепторов; − — электроны; + — дырки)

В полупроводнике одновременно присутствуют электроны и дырки, порожденные двумя причинами: 1) возбуждением собственного полупроводника n_i и p_i; 2) возбуждением донорных и (или) акцепторных примесей. При этом полные концентрации носителей заряда для примесного полупроводника будут $n = n_n + n_i$ и $p = p_i$ (донорный); $p = p_n + p_i$ и $n = n_i$ (акцепторный); при полной ионизации примесей $n_n = N_Д$ — количеству доноров, а $p_p = N_A$, т.е. количеству акцепторов. Обычно вследствие малой энергии возбуждения эти величины значительно выше собственных концентраций, т.е. $n_n >> n_i$ и $p_n >> p_i$ концентрации основных носителей определяются выражениями $n = N_Д$ (на самом деле $n = N_Д - N_A$, но $N_Д >> N_A$) для донорного полупроводника и $p = N_A$ для акцепторного полупроводника.

2.5. Распределение носителей в зонах проводимости

Разрешенные зоны содержат огромное количество уровней (10^{22}–10^{23} в 1 см^3), на каждом из которых могут находиться электроны. Фактическое же количество электронов зависит от концентрации доноров и от температуры. Чтобы оценить фактическую концентрацию носителей в полупроводнике, нужно знать распределение уровней и вероятность заполнения этих уровней.

Энергетическое распределение электронов в твердом теле определяется статистикой Ферми-Дирака. Принципиальный результат функции распределения Ферми-Дирака дает вероятность того, что электрон занимает уровень, соответствующий потенциалу φ:

$$F_n(\varphi) = \frac{1}{1 + e^{\frac{\varphi - \varphi_F}{\varphi_T}}}, \qquad (2.2)$$

где $\varphi_T = kT$ — температурный потенциал, T — абсолютная температура, k — постоянная Больцмана, φ_F — уровень Ферми.

Можно определить уровень Ферми как потенциал, вероятность заполнения которого электроном равна в точности одной второй. Функции распределения Ферми-Дирака симметричны относительно уровня Ферми. Если энергетические состояния в зоне проводимости и валентной зоне одинаковы, то уровень Ферми находится посередине запрещенной зоны. Это случай собственного полупроводника (рис. 2.11, *а*). В полупроводнике n-типа концентрации электронов в зоне проводимости больше, чем в случае собственного полупроводника (рис. 2.11, *б*), а в полупроводнике p-типа — меньше (см. рис. 2.11, *в*).

Для потенциалов в несколько единиц kT выше или ниже уровня Ферми, когда экспонента значительно больше единицы, распределение Дирака-Ферми можно заменить распределением Максвелла-Больцмана. При этом вероятность заполнения уровня в зоне проводимости определяется как

$$F_n(\varphi) = e^{-\frac{\varphi - \varphi_F}{\varphi_T}} \qquad (2.3)$$

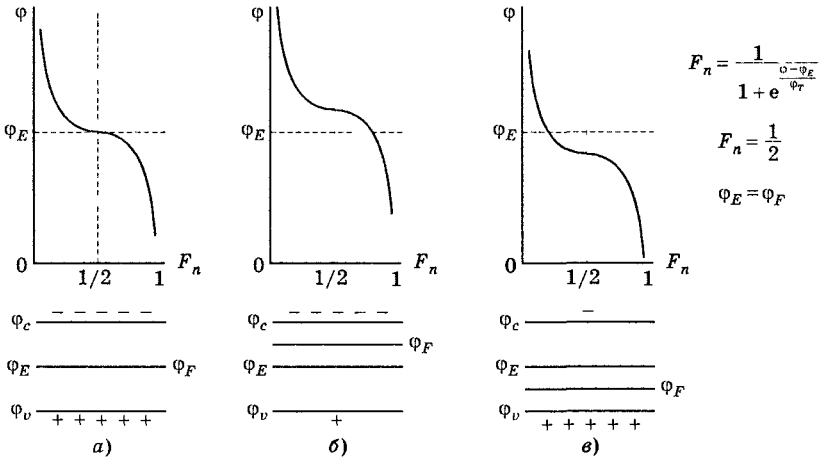


Рис. 2.11. Функция распределения Ферми-Дирака для собственного, *n*- и *p*-типов полупроводников: *а* —собственный полупроводник; *б* — полупроводник *n*-типа; *в* — полупроводник *p*-типа

Вероятность незаполнения уровня в валентной зоне (т.е. наличия дырки на этом уровне) определяется аналогичной функцией

$$F_p(\varphi) = e^{-\frac{\varphi_F - \varphi}{\varphi_T}}. \qquad (2.4)$$

Обозначим через $P(\varphi)$ плотность уровней в зоне проводимости вблизи уровня φ. Тогда $P(\varphi)\,d\varphi$ будет количеством уровней в диапазоне $d\varphi$. Умножив это количество на вероятность заполнения этих уровней $F_n(\varphi)$, получим концентрацию свободных электронов с энергиями от φ до $\varphi + d\varphi$. Полную концентрацию свободных электронов n получим путем интегрирования по всей ширине зоны проводимости. Если принять зависимость $P(\varphi) \sim \sqrt{\varphi}$, то

$$n = N_c e^{-\frac{\varphi_c - \varphi_F}{\varphi_T}}. \qquad (2.5)$$

Здесь N_c — так называемая *эффективная плотность уровней (состояний) в зоне проводимости*:

$$N_c = 0{,}5 \cdot 10^{16} (m_n/m)^{3/2} T^{3/2},$$

где m_n — эффективная масса электрона.

Аналогичным методом получается выражение для концентрации дырок:

$$p = N_v e^{-\frac{\varphi_E - \varphi_v}{\varphi_T}}. \qquad (2.6)$$

Здесь N_v — *эффективная плотность уровней (состояний) в валентной зоне*:

$$N_v = 0{,}5 \ \ 10^{16} (m_p/m)^{3/2} T^{3/2}, \qquad (2.7)$$

где m_p — эффективная масса дырки. Для кремния отношение $N_c/N_v = 2{,}8$. Часто для простоты полагают $N_c = N_v$.

Перемножая левые и правые части в формулах (2.5) и (2.6) и учитывая, что $\varphi_з = \varphi_c - \varphi_v$, нетрудно представить произведение концентраций электронов и дырок следующим образом:

$$np = N_c N_v e^{-\varphi_з/\varphi_T}. \qquad (2.8)$$

Как видим, *при неизменной температуре произведение концентраций — величина постоянная*, т.е. увеличение одной из концентраций сопровождается уменьшением другой.

В собственном полупроводнике концентрации электронов и дырок одинаковы. Обе они обозначаются через n_i и называются *собственными концентрациями*. Подставляя $n = n_i$ и $p = n_i$ в (2.8) и извлекая квадратный корень, получаем выражение для собственной концентрации:

$$n_i = \sqrt{N_c N_v}\, e^{-\varphi_з/2\varphi_T}. \qquad (2.9)$$

Отметим полную зависимость собственной концентрации от ширины запрещенной зоны и температуры.

Соотношение (2.8) часто записывают в более компактной форме через собственную концентрацию:

$$np = n_i^2. \qquad (2.10)$$

Используя выражения (2.5) и (2.6), полагая для простоты $N_c = N_v$ и учитывая, что $\varphi_E = \frac{1}{2}(\varphi_c + \varphi_v)$, нетрудно представить отношение концентраций электронов и дырок в виде:

$$n/p = e^{-\frac{2(\varphi_E - \varphi_F)}{\varphi_T}}. \qquad (2.11)$$

Подставим в левую часть (2.11) значение $p = n_i^2/n$ из (2.10) и прологарифмируем обе части; тогда уровень Ферми запишется через концентрацию свободных электронов следующим образом:

$$\varphi_F = \varphi_E + \varphi_T \ln(n/n_i). \qquad (2.12а)$$

Если подставить в (2.11) значение $n = n_i^2/p$ из (2.10), то уровень Ферми запишется через концентрацию дырок:

$$\varphi_F = \varphi_E - \varphi_T \ln(p/n_i). \qquad (2.12б)$$

Вторые члены в правых частях (2.12), характеризующие концентрации носителей, называются *химическим потенциалом*. Следовательно, уровень Ферми является суммой электрического и химического потенциалов. Отсюда еще одно его название — *электрохимический потенциал*.

Одно из фундаментальных положений в физике полупроводников формулируется следующим образом: *уровень Ферми одинаков во всех частях равновесной системы, какой бы разнородной она ни была*. Это положение можно записать в виде двух равносильных выражений:

$$\varphi_F = \text{const},$$
$$\text{grad}\,(\varphi_F) = 0. \qquad (2.12в)$$

Из условия (2.12в) в одномерном случае следует, что если концентрация электронов изменяется вдоль координаты x, то возникает электрическое поле

$$E = \varphi_T \frac{dn/dx}{n}.$$

Таким образом, в неоднородно-легированных полупроводниках смещение подвижных носителей, обусловленное градиентом концентрации, уравновешивается внутренним электрическим полем. Иногда его называют встроенным, а возникающее при этом равновесие называют больцмановским.

Движение носителей в электрическом поле называют *дрейфом*. Плотность дрейфового тока определяется известным выражением

$$j = \sigma E, \qquad (2.13)$$

где σ — удельная проводимость.

Поскольку в полупроводниках имеется два типа подвижных носителей, удельная проводимость складывается из двух составляющих — электронной и дырочной:

$$\sigma = qn\mu_n + qp\mu_p, \tag{2.14}$$

где μ_n и μ_p — *подвижности* соответствующих носителей.

Главной составляющей в формуле (2.14) является та, которая связана с основными носителями. Составляющая, связанная с неосновными носителями, обычно не существенна. В собственном полупроводнике обе составляющие равноценны.

Для оценки удельной проводимости, а значит, и дрейфового тока необходимо прежде всего знать концентрации электронов и дырок.

Казалось бы, значения n и p можно найти по формулам (2.5) и (2.6). Однако для этого нужно знать положение уровня Ферми в запрещенной зоне. Между тем уровень Ферми, как и химический потенциал, является функцией концентрации. Следовательно, расчет концентрации должен предшествовать определению уровня Ферми. Эту задачу можно решить, руководствуясь *условием нейтральности.*

Исходя из условия нейтральности, запишем для электронного полупроводника следующее соотношение:

$$n = N_д^* + p, \tag{2.15}$$

где $N_д^*$ — концентрация положительных донорных ионов (имеется в виду эффективная концентрация). Выражая концентрацию дырок через концентрацию электронов с помощью (2.9) и решая получившееся квадратное уравнение относительно n, находим концентрацию электронов в виде:

$$n_n = \sqrt{\left(\frac{N_д^*}{2}\right)^2 + n_i^2} + \frac{N_д^*}{2}. \tag{2.16а}$$

Аналогичным путем можно найти концентрацию дырок в дырочном полупроводнике:

$$p_p = \sqrt{\left(\frac{N_а^*}{2}\right)^2 + n_i^2} + \frac{N_а^*}{2}. \tag{2.16б}$$

Индексы n и p означают принадлежность к полупроводнику с соответствующим типом проводимости.

Рабочий температурный диапазон примесных полупроводников ограничен снизу температурой полной ионизации примесей (для кремния −70...−100 °С), а сверху — критической температурой, при которой примесный полупроводник превращается в собственный. В этом диапазоне формулы (2.16) можно упростить, заменяя эффективную концентрацию примесных ионов N^* эффективной концентрацией примесных атомов N (поскольку в рабочем диапазоне практически все атомы примеси ионизированы) и пренебрегая собственной концентрацией n_i (поскольку в рабочем диапазоне она существенно меньше концентрации примеси). Тогда концентрации основных носителей запишутся в виде:

$$n_n = N_д, \tag{2.17а}$$

$$p_p = N_а. \tag{2.17б}$$

Концентрации неосновных носителей легко определить, воспользовавшись соотношением (2.10):

$$p_p = n_i^2/N_д, \tag{2.18а}$$

$$n_p = n_i^2/N_а. \tag{2.18б}$$

Критическую температуру $T_{кр}$ можно найти следующим образом:

$$T_{кр} = 5800\frac{\varphi_з}{\ln\left(\frac{\sqrt{N_c N_v}}{aN}\right)}, \tag{2.19}$$

где N — концентрация примеси, а N_c и N_v зависят от температуры.

Полагая $a = 0{,}1$ и $N = 10^{16}$ см$^{-3}$ и задаваясь параметрами для кремния $\varphi_з = 1{,}11$ В, $N_c = 2{,}8 \cdot 10^{19}$ см$^{-3}$ и $N_v = 1{,}0 \cdot 10^{19}$ см$^{-3}$, получим

$$T_{кр} = 273\left(\frac{10}{4{,}5\ln\rho} - 1\right), \tag{2.20}$$

где удельное сопротивление $\rho = 0{,}85$ Ом · см соответствует концентрации $N = 10^{16}$ см$^{-3}$.

Из выражений (2.18) следует, что при низких температурах концентрация неосновных носителей очень мала. Так, если $N_д = 2 \cdot 10^{17}$ см$^{-3}$, то при комнатной температуре в кремнии концентрация дырок согласно (2.18а) составляет всего 2000 см$^{-3}$, т.е. на 14 порядков (!) меньше, чем концентрация электронов. Однако с повышением температуры концентрация неосновных носителей возрастает очень резко — пропорционально n_i^2, т.е. несравненно быстрее, чем даже концентрация собственных носителей. Так, в кремнии рост температуры на 50°С сопровождается увеличением концентрации неосновных носителей примерно на 3 порядка.

Столь же сильно влияют на концентрацию неосновных носителей такие факторы, как свет и разного рода ионизирующие излучения. Поэтому в тех полупроводниковых приборах и элементах ИС, работа которых основана на неосновных носителях, эти факторы необходимо по возможности исключать. С другой стороны, в специальных случаях влияние этих факторов можно использовать для построения фоточувствительных приборов, дозиметров излучения и т.п.

Как известно, в свободном пространстве (вакууме) электроны под действием электрического поля совершают равноускоренное движение.

В твердом теле движущиеся электроны непрерывно испытывают столкновения с узлами кристаллической решетки, примесями и дефектами, т.е., как говорят, испытывают *рассеяние.* Равноускоренное движение под действием поля возможно только в коротких интервалах между столкновениями, на *длине свободного пробега*. После каждого столкновения электрон, грубо говоря, должен заново «набирать» скорость. В результате средняя дрейфовая скорость электронов и дырок оказывается вполне определенной величиной, пропорциональной напряженности поля:

$$\overline{v} = \mu E.$$

Коэффициент пропорциональности μ есть *подвижность* носителей, измеряемая в единицах см^2/В с. При напряженности поля 1 В/см подвижность численно равна скорости.

В связи с различием эффективных масс электронов и дырок их подвижности тоже различаются. *Как правило, подвижность электронов больше подвижности дырок* (у кремния почти в 3

раза). Чем больше подвижность, тем больше скорость движения носителей и тем выше быстродействие полупроводникового прибора.

Подвижность зависит от ряда факторов, важнейшими из которых являются температура, концентрация примесей и напряженность поля. Эти зависимости необходимо иметь в виду при разработке полупроводниковых приборов и элементов ИС.

Зависимость подвижности от температуры определяется механизмом рассеяния носителей. Если преобладает рассеяние на узлах решетки, то

$$\mu_L = \mu_{0L}(T_0/T)^c, \tag{2.21}$$

если же преобладает рассеяние на ионах примеси, то

$$\mu_I = \mu_{0I}(T/T_0)^{3/2}. \tag{2.22}$$

Значения μ_0 относятся к исходной (например, комнатной) температуре T_0, а значения μ — к температуре T (имеются в виду абсолютные температуры)[1]. Показатель степени c зависит от материала и типа проводимости. Для кремния (электронного и дырочного) $c \approx 5/2$.

Учет двух процессов приводит к следующему выражению для подвижности

$$\frac{1}{\mu} = \frac{1}{\mu_L} + \frac{1}{\mu_I}. \tag{2.23}$$

Результирующая подвижность близка к меньшей из двух составляющих μ_L и μ_I. Для кремния при температурах $T > 0$ °С меньшей оказывается составляющая μ_L; поэтому зависимость $\mu(T)$ описывается формулой (2.21): *подвижность уменьшается с ростом температуры*. При температурах $T < -50$ °С меньшей оказывается составляющая μ_I; поэтому зависимость $\mu(T)$ описывается формулой (2.22); *подвижность уменьшается с уменьшением температуры*. В диапазоне рабочих температур −60...+125 °С подвижность может изменяться в 4–5 раз, что, конечно, существенно.

---

1 Индексы L и I происходят от Lattice (решетка) и Ion (ион).

Запишем общее выражение для удельной проводимости применительно к собственному и примесным полупроводникам.

Для собственного полупроводника $n = p = n_i$ и, следовательно,

$$\sigma_i = qn_i(\mu_n + \mu_p). \qquad (2.24)$$

Температурная зависимость собственной проводимости определяется температурной зависимостью собственной концентрации n_i, см. (2.9). Эта зависимость, как уже отмечалось, очень сильная — экспоненциальная. На рис. 2.12,*а* функция σ_i $(1/T)$ для кремния представлена в полулогарифмическом масштабе; приведен также масштаб в °С. Как видим, в рабочем диапазоне −60...+125 °С собственная проводимость кремния меняется на 5 порядков. У материалов с меньшей шириной запрещенной зоны (например, у германия) изменение σ_i будет меньше, хотя сами значения σ_i будут больше из-за большей собственной концентрации.

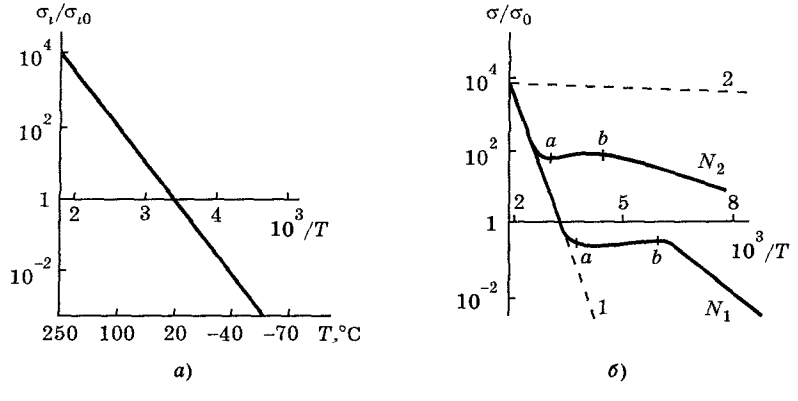


Рис. 2.12. Зависимость относительной удельной проводимости кремния от температуры, σ_0 и σ_0 — проводимости при +20°С
а — собственный кремний, *б* — примесный кремний

Для примесных полупроводников — электронного и дырочного — пренебрегая в выражении (2.14) составляющими, связанными с неосновными носителями, и используя соотношения (2.17), получаем

$$\sigma_n = qN_д\mu_n; \qquad (2.25а)$$

$$\sigma_p = qN_а\mu_p. \qquad (2.25б)$$

В рабочем диапазоне температуры концентрации $N_д$ и $N_а$ можно считать постоянными. Следовательно, в этом диапазоне температурная зависимость проводимости примесного полупроводника определяется температурной зависимостью подвижности.

На рис. 2.12, *б* показаны две кривые $(\sigma/\sigma_0)(1/T)$ для разной концентрации примеси ($N_2 > N_1$). Для сравнения штриховой кривой *1* показана часть функции σ_i $(1/T)$, заимствованной из рис. 2.12, *а*. Точки *а* соответствуют критической температуре (2.19), при которой примесный полупроводник превращается в собственный, поэтому левее точек *а* кривые σ сливаются с штриховой кривой σ_i. Точки *b* соответствуют температуре ионизации примеси; правее этих точек (т.е. при более низких температурах) концентрация ионизированных примесных атомов убывает и соответственно уменьшается удельная проводимость.

Как видим, в рабочем диапазоне зависимость $\sigma(T)$ для примесных полупроводников несравненно слабее, чем для собственного. Кроме того, она имеет «обратный» характер: с ростом температуры σ не увеличивается, а уменьшается.

Штриховой кривой *2* показана зависимость $\sigma(1/T)$ для очень большой концентрации примеси, т.е. для вырожденного полупроводника. Зависимость эта очень слабая, что подтверждает близость вырожденных полупроводников к металлам и оправдывает их название — полуметаллы.

2.6. Эффект поля

Эффектом поля называют *изменение концентрации носителей* (а, значит, и проводимости) *в приповерхностном слое полупроводника под действием электрического поля*. Слой с повышенной (по сравнению с объемом) концентрацией основных носителей называют *обогащенным*, а слой с пониженной их концентрацией — *обедненным*.

Пусть между металлической пластинкой и полупроводником, разделенными диэлектриком (например, воздухом) задано напряжение U (рис. 2.13). Ясно, что в системе МДП (металл — диэлектрик — полупроводник) протекание тока невозможно. Поэтому такая система равновесна и представляет собой своеобразный конденсатор, у которого одна из обкладок полупровод-

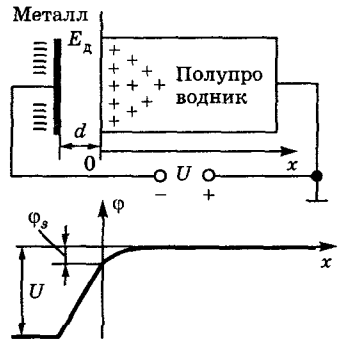


Рис. 2.13. Эффект поля в структуре металл–диэлектрик–полупроводник

никовая. На этой обкладке будет наведен такой же заряд, как и на металлической. Однако *в отличие от металла заряд в полупроводнике не сосредоточивается на поверхности, а распространяется на некоторое расстояние в глубь кристалла.*

Электрическое поле, созданное напряжением U, распределяется между диэлектриком и полупроводником. Поле в диэлектрике $E_д$ постоянное (так как в диэлектрике нет объемных зарядов), а поле в полупроводнике заведомо непостоянное, так как заряд спадает от поверхности в глубь полупроводника.

Знак заряда в полупроводнике зависит от полярности приложенного напряжения. При отрицательной полярности (рис. 2.13) наведенный заряд положительный. В дырочном полупроводнике положительный заряд обусловлен дырками, которые притянулись к поверхности, а в электронном полупроводнике — ионами доноров, от которых оттолкнулись электроны, компенсировавшие их заряд. Значит, в первом случае происходит обогащение, а во втором — обеднение приповерхностного слоя основными носителями. При положительной полярности напряжения, наоборот, в электронном полупроводнике происходит обогащение приповерхностного слоя электронами, а в дырочном — обеднение дырками и «обнажение» отрицательных акцепторных ионов.

Протяженность подвижных зарядов в обогащенном слое называют *длиной Дебая* или *дебаевской длиной*, а протяженность неподвижных ионных зарядов — *глубиной обедненного слоя*. Обе эти величины рассматриваются ниже. Обогащенные и обедненные слои оказываются тем тоньше, чем больше концентрация примеси, а значит, и концентрация основных носителей. Иначе говоря, тонкие слои свойственны низкоомным полупроводникам, а толстые — высокоомным.

Если принять потенциал в объеме полупроводника равным нулю, то потенциал поверхности будет отличен от нуля благо-

даря наличию зарядов между объемом и поверхностью. Разность потенциалов между поверхностью и объемом называют *поверхностным потенциалом* и обозначают через φ_s.

Следует заметить, что в отсутствие внешнего напряжения поверхностный потенциал не падает до нуля, а имеет конечную равновесную величину φ_{s0}. Она обусловлена наличием поверхностных состояний, которые способны захватывать или отдавать электроны на сравнительно длительное время. Еще одним фактором, влияющим на величину φ_{s0}, является *контактная разность потенциалов между металлом и полупроводником*. Внешнее напряжение, необходимое для того, чтобы скомпенсировать равновесный поверхностный потенциал, называется *напряжением спрямления зон* и обозначается через U_F (от Flat Band — плоские зоны).

Как уже отмечалось, электрическое поле распределяется между диэлектриком и полупроводником. Поле в диэлектрике возрастает при уменьшении расстояния d. Расстояние d не может быть произвольно малым: при условии $d < 10$ нм диэлектрик становится проницаемым для подвижных носителей благодаря *туннельному эффекту*. При этом структура МДП перестает быть аналогом конденсатора: обмен носителями через диэлектрик вызывает протекание тока, а значит, нарушает равновесное состояние. Распределение потенциала в области объемного заряда можно оценить с помощью одномерного уравнения Пуассона

$$\frac{d^2\varphi}{dx^2} = -\frac{\lambda}{\varepsilon_0 \varepsilon}, \qquad (2.26)$$

где λ — плотность заряда; ε_0 — электрическая постоянная; ε — относительная диэлектрическая проницаемость полупроводника.

В общем случае плотность заряда в полупроводнике записывается следующим образом:

$$\lambda = q(p + N_д^* - n - N_а^*), \qquad (2.27)$$

где $N_д^*$ и $N_а^*$ — концентрации ионизированных примесей.

Концентрации свободных носителей в правой части (2.27) связаны с величиной электростатического потенциала φ_E. Чтобы установить эту связь, воспользуемся выражениями (2.12).

Пусть в глубине полупроводника, где заряды и поле отсутствуют, концентрации электронов и дырок равны n_0 и p_0, а электростатический потенциал равен φ_{E0}. Соответствующие величины вблизи поверхности обозначим просто через n, p и φ_E. Подставим в (2.12а) сначала значения n_0 и φ_{E0}, а затем n и φ_E и приравняем правые части (поскольку в равновесной системе φ_F = const). В результате получим

$$\varphi_E - \varphi_{E0} = \varphi_t \ln(n_0/n).$$

Примем для простоты $\varphi_{E0} = 0$ (это соответствует заземлению полупроводника на рис. 2.13). Тогда концентрация n выразится через φ_E следующим образом:

$$n = n_0 e^{-\varphi_E/\varphi_T}. \qquad (2.28а)$$

Аналогичным путем, с помощью (2.12б), можно выразить концентрацию p через φ_E:

$$p = p_0 e^{\varphi_E/\varphi_T}. \qquad (2.28б)$$

Заменим в формулах (2.28) величину φ_E на $-\varphi$, т.е. перейдем от энергетических потенциалов к электрическим. После этого, подставляя концентрации n и p в правую часть (2.27) и далее — плотность заряда λ в уравнение Пуассона, получам нелинейное дифференциальное уравнение, которое в общем случае не имеет аналитического решения. Однако в двух важных частных случаях, когда можно пренебречь либо концентрацией ионизированных примесей (обогащенные слои), либо концентрацией свободных носителей (обедненные слои), аналитические решения существуют. Эти случаи рассмотрены ниже.

Эффект поля в собственном полупроводнике. Подставим в правую часть (2.27) концентрации (2.28) и заменим φ_E на $-\varphi$ Далее, учитывая, что полупроводник собственный, положим $n_0 = p_0 = n_i$ и $N_д^* = N_а^* = 0$. Тогда плотность заряда приводится к виду

$$\lambda = -2qn_i \operatorname{sh}(\varphi/\varphi_T).$$

Подставим полученное значение λ в правую часть (2.26), поделим обе части на φ_T и введем безразмерную переменную $\Phi = \varphi/\varphi_T$. После этого уравнение Пуассона примет вид:

$$\frac{d^2\Phi}{dx^2} = \frac{l}{l^2_{Di}} \operatorname{sh}(\Phi), \qquad (2.29)$$

где

$$l_{Di} = \sqrt{\frac{\varepsilon_0 \varepsilon \varphi_T}{2qn_i}} \qquad (2.30)$$

— дебаевская длина в собственном полупроводнике. Для кремния $l_{Di} \approx 14$ мкм.

Рассмотрим простейший случай, когда $|\varphi_s| < \varphi_T$, т.е. $|\Phi| < 1$. В этом случае можно положить $\mathrm{sh}\,\Phi \approx \Phi$ и (2.29) превращается в линейное дифференциальное уравнение 1-го порядка. Для граничных условий $\varphi(\infty) = 0$ и $\varphi(0) = \varphi_s$ решение имеет вид:

$$\varphi(x) \approx \varphi_s e^{-x/l_{Di}}. \qquad (2.31)$$

Из (2.31) следует, что дебаевская длина — это расстояние, на котором потенциал уменьшается в e раз по сравнению с максимальным значением φ_s на поверхности.

Зная функцию $\varphi(x)$, легко получить функции $E(x)$, $\lambda(x)$, $n(x)$ и $p(x)$. Все они показаны на рис. 2.14 для той же полярности напряжения, что и на рис. 2.13. Показана также зонная диаграмма полупроводника, где кривая $\varphi_E(x)$, а значит, и все другие энергетические уровни являются зеркальным отображением кривой $\varphi(x)$. Искривление энергетических зон вблизи границы полупроводник-диэлектрик — характерная особенность эффекта поля.

Если изменить полярность напряжения, то знак объемного заряда изменится и зоны искривятся в другую сторону — «вниз». Однако при обеих полярностях приповерхностный слой в собственном полупроводнике оказывается обогащенным (либо электронами, либо дырками).

Величину поверхностного потенциала можно найти из условия непрерыв-

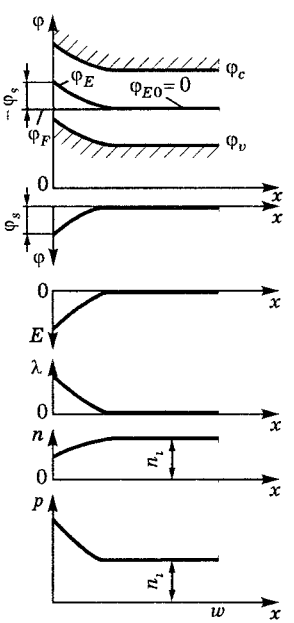


Рис. 2.14. Эффект поля в собственном полупроводнике. зонная диаграмма, распределение потенциала, поля, заряда и концентраций носителей

ности электрической индукции на границе полупроводник-диэлектрик:

$$\varepsilon_{\text{п}} E(0) = \varepsilon_{\text{д}} E_{\text{д}}(0), \qquad (2.32\text{а})$$

где $\varepsilon_{\text{п}}$ и $\varepsilon_{\text{д}}$ — относительные диэлектрические проницаемости полупроводника и диэлектрика.

Поле в диэлектрике постоянное, поэтому (рис. 2.13)

$$E_{\text{д}}(0) = (U - \varphi_s)/d; \qquad (2.32\text{б})$$

поле в полупроводнике на границе с диэлектриком (рис. 2.14) определяется функцией $\varphi(x)$:

$$E(0) = -\frac{\mathrm{d}\varphi}{\mathrm{d}x}(0). \qquad (2.32\text{в})$$

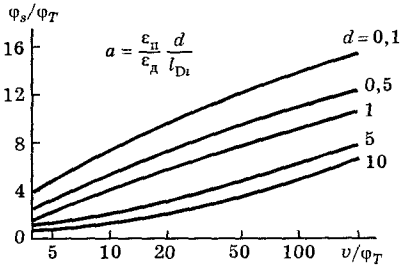


Рис. 2.15. Зависимость поверхностного потенциала в собственном полупроводнике от *толщины диэлектрика* и напряжения на металлическом электроде

Опуская математические выкладки, приведем зависимость $\varphi_s(U)$ в виде кривых на рис. 2.15. Из этих кривых видно, что поверхностный потенциал составляет тем большую долю приложенного напряжения, чем тоньше диэлектрик (чем меньше параметр a). При всех реальных значениях толщины диэлектрика и приложенного напряжения поверхностный потенциал не превышает нескольких десятых долей вольта.

Примесный полупроводник. Особенностью эффекта поля в примесных полупроводниках по сравнению с собственными является возможность получения как обогащенных, так и обедненных слоев.

Режим обогащения соответствует такой полярности приложенного напряжения, при которой основные носители притягиваются к поверхности. Этот случай близок к рассмотренному на рис. 2.14, но отличается меньшим искривлением зон (рис. 2.16, *а*). Меньшее искривление зон обусловлено тем, что примесный полупроводник богат подвижными носителями и

потому даже небольшой поверхностный потенциал обеспечивает необходимый заряд вблизи поверхности.

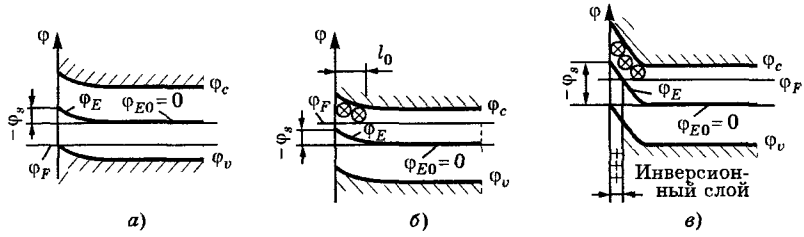


Рис. 2.16. Эффект поля в примесных полупроводниках: *а* — режим обогащения; *б* — режим обеднения, *в* — образование инверсионного слоя

При условии $\varphi_s < 2\varphi_T$ потенциал в примесном полупроводнике описывается выражением (2.31), но дебаевская длина имеет вид

$$l_D = \sqrt{\frac{\varepsilon_0 \varepsilon \varphi_T}{qN}}, \qquad (2.33)$$

где N — концентрация ионизированной примеси (донорной или акцепторной).

Поскольку $N \gg n_i$, дебаевская длина в примесных полупроводниках гораздо меньше, чем в собственных. Кроме того, она практически не зависит от материала. Полагая $N = 10^{16}$ см$^{-3}$, получаем из (2.33) типичное значение $l_D \approx 0{,}04$ мкм. Как видим, поле проникает в примесные полупроводники на ничтожную глубину.

Если применить формулу (2.33) к металлам (хотя это не совсем правомерно), то при свойственных им огромных концентрациях свободных носителей 10^{22}–10^{23} см$^{-3}$ дебаевская длина l_D лежит в пределах десятых долей нанометра, что соответствует 1–2 межатомным расстояниям. Подобная оценка хорошо иллюстрирует тот известный факт, что заряды в металле всегда сосредоточены на поверхности, *внутри металла заряды и электрические поля отсутствуют*.

Режим обеднения соответствует такой полярности приложенного напряжения, при которой основные носители отталкиваются от поверхности. В этом случае поверхностный потенциал может иметь гораздо большие значения, чем в режиме обо-

гащения (рис. 2.16,б). Отталкивание основных носителей, как уже отмечалось, приводит к появлению некомпенсированного объемного заряда примесных ионов.

Предположим, что граница обедненного слоя резкая и расположена на расстоянии l_0 от поверхности. Плотность объемного заряда в обедненном слое примем постоянной и равной qN, где N — концентрация ионизированной примеси. Подставляя значение $\lambda = qN$ в уравнение Пуассона (2.26) и используя граничные значения $E(l_0) = 0$ и $\varphi(l_0) = 0$, получаем после двукратного интегрирования:

$$\varphi = (\lambda/2\varepsilon_0\varepsilon)(x - l_0)^2.$$

Положив в этом выражении $x = 0$ и $\varphi(0) = \varphi_s$, найдем протяженность (толщину) обедненного слоя:

$$l_0 = \sqrt{\frac{2\varepsilon_0\varepsilon|\varphi_s|}{qN}}. \qquad (2.34)$$

Хотя структура выражений (2.34) и (2.33) одинакова, между ними есть и существенная разница: дебаевская длина зависит только от свойств материала, тогда как толщина объемного заряда зависит еще и от приложенного напряжения, поскольку от него зависит потенциал φ_s (см. рис. 2.15). Обычно величина l_0 в несколько раз превышает величину l_D.

С ростом напряжения основные носители продолжают отталкиваться (а обедненный слой расширяться), но одновременно к поверхности притягиваются неосновные носители. Когда нарастающий заряд неосновных носителей превысит заряд оставшихся основных, *изменится тип проводимости приповерхностного слоя*. Этот случай характеризуют термином *инверсия типа проводимости*, а слой, образованный неосновными носителями, называют *инверсионным слоем* (рис. 2.16, *в*).

С точки зрения зонной теории образование инверсионного слоя объясняется тем, что вблизи поверхности уровень электростатического потенциала пересекает уровень Ферми. Тем самым на приповерхностном участке уровень Ферми оказывается в той половине запрещенной зоны, которая соответствует преобладанию неосновных носителей. Толщина инверсионного слоя составляет всего 1–2 нм, т.е. 3–4 постоянных решетки.

Из рис. 2.16, *в* видно, что инверсионный слой образуется при значении поверхностного потенциала $-(\varphi_F - \varphi_{E0})$. Дальнейшее увеличение внешнего напряжения сопровождается дальнейшим увеличением потенциала φ_s до тех пор, пока уровень Ферми не пересечет границу разрешенной зоны (рис. 2.16, *в* — валентной). После этого граничный слой превращается в полуметалл, а потенциал φ_s практически не меняется и сохраняет значение

$$\varphi_{sm} = -2(\varphi_F - \varphi_{E0}). \qquad (2.35)$$

В обычных случаях максимальный поверхностный потенциал составляет 0,6–1,0 В.

2.7. Рекомбинация носителей

Процессы генерации и рекомбинации носителей заряда неотъемлемы друг от друга, хотя и противоположны по содержанию. Хотя в последующем изложении будут обсуждаться только процессы рекомбинации, на трех иллюстрациях приводятся оба процесса.

На рис. 2.17 условно изображены процессы генерации и рекомбинации, совершающиеся непосредственным переходом из зоны в зону.

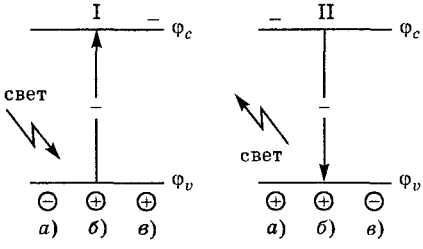


Рис. 2.17. Прямые переходы из зоны в зону I — генерация носителей заряда, II — рекомбинация носителей заряда (*а* — начальное состояние, *б* — переход, *в* — конечное состояние)

Рассмотрим случай, когда примесь характеризуется *глубокими* уровнями, расположенными вблизи середины запрещенной зоны (на рис. 2.18 и 2.19). Энергия активации в этом случае достаточно большая, поэтому атомы таких примесей практически не ионизируются, соответственно не меняются концентрации свободных носителей. Тем не менее роль глубоких уровней может быть весьма существенной: они представляют собой так называемые ловушки или центры захвата подвижных носителей.

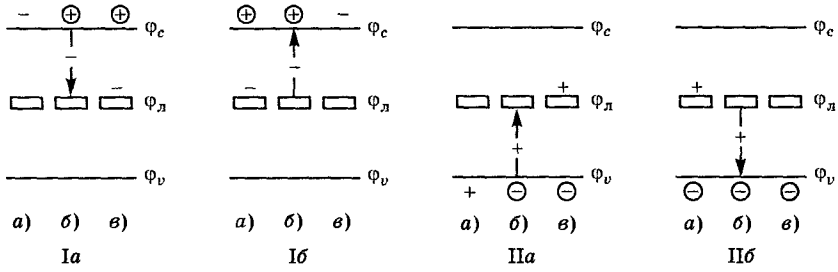


Рис. 2.18. I*а* — захват электрона; I*б* — *отдача электрона;* II*а* — захват дырки; II*б* — отдача дырки (*а* — начальное состояние; *б* — переход; *в* — конечное состояние; $\varphi_л$ — энергетический уровень ловушки)

На рис. 2.18 показаны процессы захвата и отдачи электрона и дырки ловушкой при одноступенчатых процессах.

Электрон, попавший из разрешенной зоны на ловушку, находится на ней в течение некоторого времени — времени релаксации. На рис. 2.19 показаны двухступенчатые процессы.

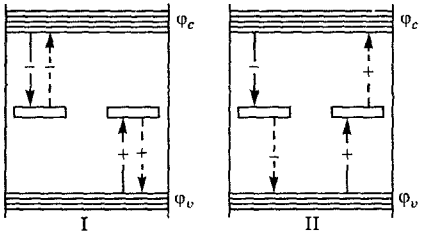


Рис. 2.19. I — двухступенчатая рекомбинация; II — двухступенчатая генерация

Вероятность двухступенчатых процессов гораздо больше, чем одноступенчатых, рассмотренных выше. Поэтому в присутствии ловушек процессы генерации-рекомбинации идут значительно интенсивнее и время жизни носителей оказывается значительно меньше.

Захват электронов глубокими уровнями особенно характерен для поверхности полупроводника, которая богата поверхностными состояниями. В зависимости от времени релаксации поверхностные состояния делят на *быстрые* и *медленные*. Быстрые состояния характерны временами релаксации порядка 10^{-6} с, а медленные — временами порядка 10^{-3} с и более, вплоть до нескольких секунд. Рекомбинация, противодействуя накоплению носителей, обусловливает их равновесные концентрации. Благодаря рекомбинации свободные носители имеют конечное *время жизни*, а этот параметр во многом определяет длительность переходных процессов.

Механизмы рекомбинации. Различают *непосредственную рекомбинацию* и *рекомбинацию на примесных центрах.*

Непосредственной рекомбинацией называют переход электрона из зоны проводимости непосредственно в валентную зону, где он занимает один из вакантных уровней, т.е. «уничтожает» дырку. Разумеется, при таком переходе должна выделяться энергия $q\varphi_з$ — такая же, какая была ранее затрачена на перевод электрона из валентной зоны в зону проводимости. Энергия может выделяться либо в виде фотона (*излучательная рекомбинация*), либо в виде фонона (*безызлучательная рекомбинация*). В большинстве полупроводников, в том числе и в кремнии, вероятность излучательной рекомбинации на несколько порядков меньше, чем безызлучательной [1].

Однако вероятность безызлучательной непосредственной рекомбинации сама по себе тоже очень мала, поскольку сравнительно большая энергия $q\varphi_з$ (порядка 1 эВ) редко может воплотиться в одном фононе, а ее одновременное распределение между двумя фононами маловероятно. Таким образом, *непосредственная рекомбинация в целом не является главным механизмом рекомбинации в полупроводниках.*

Главную роль играет рекомбинация на примесных центрах. Речь идет о глубоких уровнях, расположенных вблизи середины запрещенной зоны, которые называют *ловушками* (см. рис. 2.18). Данный вид рекомбинации — двухэтапный: сначала электрон переходит из зоны проводимости на уровень ловушки, а затем с уровня ловушки в валентную зону. На каждом этапе выделяется энергия, близкая к $\frac{1}{2}q\varphi_з$, т. е. вдвое меньше, чем при непосредственной рекомбинации. Это обстоятельство резко повышает вероятность передачи энергии фонону, что и объясняет преимущественное значение данного механизма рекомбинации.

Роль ловушек могут играть не только примесные атомы, но и различные дефекты решетки. Поэтому повышенная скорость рекомбинации свойственна, в частности, поликристал-

---

1 Это объясняется тем, что электрон, возвращаясь в валентную зону, должен отдать не только энергию, но и импульс. Поскольку фотон не способен воспринять сколько-нибудь заметный импульс, необходимо, чтобы в процесс включалась еще третья частица — фонон, а такая комбинация встречается крайне редко.

лам (у которых дефектами являются все грани между отдельными зернами) и приповерхностным слоям любого монокристаллического полупроводника (где неизбежны нарушения периодичности решетки и разрывы ковалентных связей).

Рекомбинация равновесных носителей. Вероятность непосредственной рекомбинации электрона с одной из дырок в единицу времени можно записать следующим образом:

$$r = \sigma_{\text{эфф}} v_T,$$

где $\sigma_{\text{эфф}}$ — *эффективное сечение захвата*; v_T — средняя тепловая скорость электронов [1]. Величину r называют *коэффициентом рекомбинации*. Умножив коэффициент r на концентрацию дырок, получим полную вероятность рекомбинации электрона в единицу времени (с любой из имеющихся дырок). Обратная величина будет средним интервалом между актами рекомбинации, т.е. *средним временем жизни* электронов при непосредственной рекомбинации:

$$\tau_n = 1/(rp_0). \qquad (2.36а)$$

Путем аналогичных рассуждений получаем среднее время жизни дырок:

$$\tau_p = 1/(rn_0). \qquad (2.36б)$$

В формулах (2.36) индекс «0» присвоен равновесным концентрациям. Таким обозначением будем пользоваться и в дальнейшем.

Если умножить вероятность рекомбинации одного электрона rp_0 на концентрацию электронов n_0, получится общее количество актов рекомбинации в единицу времени, т.е. *скорость непосредственной рекомбинации*:

$$R_0 = rn_0p_0. \qquad (2.37)$$

---

1 Очевидно, что неподвижный электрон никогда не «встретится» с дыркой; чем больше скорость электрона, тем более вероятна такая «встреча». Что касается сечения захвата, то оно характеризует тот объем вокруг дырки, попав в который, электрон неизбежно притянется к ней, несмотря на инерцию своего движения.

Из формул (2.36) ясно, что равновесные времена жизни электронов и дырок в общем случае резко различны из-за различия концентраций n_0 и p_0, причем время жизни неосновных носителей всегда меньше, чем основных.

Заменяя в правой части (2.36) произведение rn_0 на $1/\tau_p$ или rp_0 на $1/\tau_n$, запишем скорость рекомбинации еще в одной распространенной форме:

$$R_0 = p_0/\tau_p = n_0/\tau_n. \qquad (2.38)$$

Непосредственная рекомбинация. В неравновесном состоянии полупроводника концентрации свободных носителей отличаются от равновесных значений:

$$n = n_0 + \Delta n; \qquad (2.39а)$$

$$p = p_0 + \Delta p. \qquad (2.39б)$$

Неравновесные концентрации n и p могут быть больше и меньше равновесных, т.е. знаки приращений Δn и Δp в формулах (2.39) могут быть как положительными, так и отрицательными. Приращения Δn и Δp называют *избыточными концентрациями*. Для сохранения нейтральности полупроводника избыточные концентрации электронов и дырок должны быть одинаковыми:

$$\Delta n = \Delta p. \qquad (2.40а)$$

Более того, при изменениях избыточных концентраций нейтральность тоже должна сохраняться, отсюда — условие равенства скоростей изменения концентраций:

$$dn/dt = dp/dt. \qquad (2.40б)$$

Из выражений (2.40) следует, что *анализировать порознь поведение избыточных электронов и избыточных дырок не имеет смысла, так как функции* $\Delta n(t)$ и $\Delta p(t)$ совпадают. Поэтому ниже рассматривается только поведение электронов.

Предположим, что по какой-то причине нарушено равенство между скоростями генерации и рекомбинации электронов. Тогда электроны будут накапливаться (или рассасываться) со скоростью, равной разности скоростей генерации и рекомбинации:

$$dn/dt = g - r(np), \qquad (2.41)$$

где g — скорость генерации; $r(np)$ — скорость рекомбинации.

Запишем скорость генерации в виде:

$$g = g_0 + \Delta g = r(n_0 p_0) + \Delta g, \tag{2.42}$$

где g_0 — равновесное значение, равное равновесной скорости рекомбинации (2.37).

Преобразуем рекомбинационный член в правой части (2.41), подставляя (2.39) и учитывая (2.40):

$$r(np) = r[n_0 p_0 + \Delta n(n_0 + p_0) + \Delta n^2].$$

Положим $\Delta n << n_0 + p_0$; это дает основание пренебречь членом Δn^2 и линеаризовать уравнение (2.41). Далее концентрации n_0 и p_0 в круглых скобках выразим через времена жизни с помощью (2.36), тогда

$$r(np) = r(n_0 p_0) + \Delta n\left(\frac{1}{\tau_n} + \frac{1}{\tau_p}\right).$$

Наконец, введем *эквивалентное время жизни* τ в форме соотношения

$$\frac{1}{\tau} = \frac{1}{\tau_n} + \frac{1}{\tau_p}. \tag{2.43}$$

После этого скорость рекомбинации запишется следующим образом:

$$r(np) = r(n_0 p_0) + \Delta n/\tau.$$

Подставляя в (2.41) полученное значение $r(np)$, а также скорость генерации g из (2.42), получаем *уравнение накопления* избыточных носителей в виде:

$$dn/dt = \Delta g - \Delta n/\tau. \tag{2.44}$$

Если принять $\Delta g = 0$, получаем *уравнение рассасывания*:

$$dn/dt = -\Delta n/\tau. \tag{2.45}$$

Решением уравнения рассасывания является экспоненциальная функция

$$\Delta n(t) = \Delta n(0) e^{-t/\tau}, \tag{2.46}$$

где $\Delta n(0)$ — начальное значение избыточной концентрации. Зависимость (2.46) позволяет определить время жизни как *интервал, в течение которого избыточная концентрация уменьшается в* e *раз.*

Из структуры выражения (2.43) следует, что величина τ близка к минимальной из двух ее составляющих τ_n и τ_p. Следовательно, *эквивалентное время жизни избыточных носителей определяется временем жизни неосновных носителей.* У электронных полупроводников $\tau \approx \tau_p$, у дырочных $\tau \approx \tau_n$.

Рекомбинация на ловушках. При *ловушечном механизме* рекомбинации скорость рассасывания избыточных носителей описывается *формулой Шокли–Рида*:

$$\frac{dn}{dt} = -\frac{np - n_0 p_0}{(n + n_t)\tau_p + (p + p_t)\tau_n}. \qquad (2.47)$$

Здесь n_t и p_t — параметры с размерностью концентрации, зависящие от расположения уровня ловушек в запрещенной зоне[1]; τ_p и τ_n — времена жизни неосновных носителей:

$$\tau_n = 1/(r_n N_t); \qquad (2.48а)$$

$$\tau_p = 1/(r_p N_t), \qquad (2.48б)$$

где N_t — концентрация ловушек.

Формулы (2.48) имеют ту же структуру, что и (2.36). Однако при непосредственной рекомбинации времена жизни различались из-за различия в концентрациях носителей, а в данном случае они различаются из-за различия в коэффициентах рекомбинации.

Приравнивая правые части (2.47) и (2.45), легко получить время жизни τ. Подставляя значения n и p из (2.39), учитывая (2.40а) и полагая $\Delta n \ll n_0 + p_0$ [такое же условие было принято при выводе (2.44)], получаем

$$\tau = \frac{n_0 + n_t}{n_0 + p_0}\tau_p + \frac{p_0 + p_t}{n_0 + p_0}\tau_n. \qquad (2.49)$$

---

1 Концентрации n_t и p_t на несколько порядков меньше концентрации основных носителей. Однако собственную концентрацию они могут превышать в десятки раз.

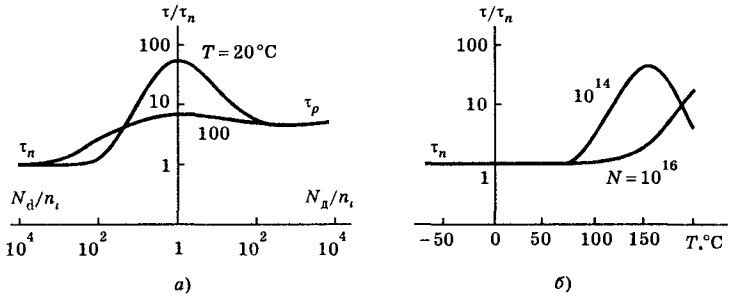


Рис 2.20. Зависимость времени жизни носителей от концентрации примеси (*а*) и от температуры (*б*)

ция $\tau(T)$ не нарастающая, а спадающая, так как концентрация n_t увеличивается медленнее, чем n_i. На рис. 2.20, *б* показаны примеры функции $\tau(T)$ для разных концентраций примеси. Как видим, *температурная зависимость времени жизни наиболее существенна в слабо легированных полупроводниках* (для кремния при $N < 10^{14}$–10^{15} см$^{-3}$). В сильно легированных (низкоомных) полупроводниках эта зависимость второстепенная.

В заключение отметим, что типичные значения времени жизни для кремния составляют 0,1–1 мкс. Если специально легировать кремний «ловушечной» примесью (чаще всего золотом), то время жизни уменьшается до 10 нс и менее.

Поверхностная рекомбинация. В принципе процессы рекомбинации в приповерхностном слое полупроводника не отличаются от процессов в объеме. Однако приповерхностный слой имеет особую зонную структуру, а значит, и количественно иные параметры, чем объем. Это нельзя не учитывать при анализе и проектировании полупроводниковых приборов и ИС, тем более, что активные области ИС расположены вблизи поверхности. Учитывая специфику приповерхностного слоя, обозначим поверхностное время жизни через τ_s, а объемное через τ_v.

Если рабочий участок элемента ИС расположен полностью в приповерхностном слое или полностью в объеме кристалла, то при анализе нужно использовать соответственно параметр τ_s или τ_v. Однако, если (как обычно бывает) рабочий участок «выходит на поверхность», т.е. частично расположен в объеме, а частично в приповерхностном слое, то принято пользоваться

Для электронного полупроводника при выполнении свойственных ему неравенств $n_0 >> p_0$ и $n_0 >> n_t, p_t$ из (2.49) следует: $\tau \approx \tau_p$. Для дырочного полупроводника при выполнении аналогичных неравенств следует: $\tau = \tau_n$. Значит, при ловушечной рекомбинации, как и при непосредственной, *время жизни избыточных носителей определяется временем жизни неосновных носителей.*

Зависимость времени жизни от концентрации ловушек следует из выражений (2.48): чем больше концентрация ловушек, тем меньше время жизни.

Зависимость времени жизни от концентрации примеси рассмотрим на примере электронного полупроводника, у которого время жизни характеризуется первым слагаемым в правой части (2.49):

$$\tau \approx \frac{n_0 + n_t}{n_0 + p_0} \tau_p. \qquad (2.50)$$

Если концентрация доноров достаточно велика, то $n_0 >> p_0 n_t$. Тогда время жизни не зависит от концентрации примеси: $\tau \approx \tau_p$. С уменьшением концентрации доноров неравенство $n_0 >> n_t$ нарушается и время жизни растет. В пределе, когда концентрация доноров приближается к нулю, полупроводник становится собственным и время жизни достигает максимального значения:

$$\tau \approx \frac{n_t}{2n_i} \tau_p >> \tau_p. \qquad (2.51)$$

Аналогичные результаты получаются для дырочного полупроводника. Характер функции $\tau(N)$ показан на рис. 2.20, *а*. Как видим, *у сильно легированных полупроводников время жизни меньше, чем у слабо легированных и собственного.*

Зависимость времени жизни от температуры обусловлена резким возрастанием концентрации n_t по закону, близкому к (2.9). Когда величина n_t делается сравнимой с n_0, время жизни начинает увеличиваться, а при условии $n_t > n_0$ зависимость $\tau(T)$ практически совпадает с экспоненциальной функцией $n_t(T)$. Быстрый рост времени жизни замедляется вблизи критической температуры, когда полупроводник превращается в собственный. Для собственного полупроводника согласно (2.51) функ-

так называемым *эффективным временем жизни* τ, которое определяется следующим образом:

$$1/\tau = (1/\tau_s + 1/\tau_v). \qquad (2.52)$$

Именно этот параметр имеют в виду, когда анализируют транзисторы и другие элементы ИС.

Поскольку, как правило, $\tau_s < \tau_v$ (из-за большой концентрации ловушек вблизи поверхности), то эффективное время жизни ближе к величине τ_s. Последняя, однако, труднее поддается расчету и измерению, чем τ_v. Поэтому широкое распространение для характеристики поверхностной рекомбинации получил особый параметр — *скорость поверхностной рекомбинации* s (см/с), которая легче поддается измерению, чем время жизни τ_s. Скорость поверхностной рекомбинации существенно зависит от способа и качества обработки поверхности кристалла. Ее типичные значения лежат в широких пределах от 100 до 10^4 см/с и более.

Физический смысл параметра s состоит в следующем. Если избыточная концентрация носителей создана вблизи приповерхностного слоя (где интенсивность рекомбинации выше, чем в объеме), то значительная часть избыточных носителей направится к поверхности, чтобы скомпенсировать убыль носителей в приповерхностном слое. Следовательно, между объемом и приповерхностным слоем возникает поток носителей, скорость которого и определяется параметром s.

Связь между скоростью поверхностной рекомбинации s и приповерхностным временем жизни τ_s в общем случае установить трудно. Эта задача решена только в двух частных, хотя и важных случаях: для бруска бесконечной длины и для тонкой пластинки, у которой толщина равна d, а площадь бесконечно большая. В последнем, практически наиболее интересном случае получается соотношение:

$$\tau_s = d^2/(4\eta D), \qquad (2.53)$$

где D — коэффициент диффузии носителей, а величина η определяется трансцендентным уравнением

$$\eta \operatorname{tg}\eta = sd/(2D).$$

При условии $s < D/d$ соотношение между величинами s и τ_s переходит в явное:

$$\tau_s = d/2s. \tag{2.54}$$

На практике токи, связанные с движением избыточных носителей из объема к поверхности, удобно рассчитывать без использования параметра τ_s, поскольку плотность тока, обусловленного таким движением, весьма просто связана со скоростью поверхностной рекомбинации:

$$j_s = qs\Delta n. \tag{2.55}$$

Вообще говоря, этот ток является паразитным и его следует уменьшать, уменьшая скорость поверхностной рекомбинации всеми доступными средствами.

2.8. Законы движения носителей в полупроводниках

В общем случае движение носителей заряда обусловлено двумя процессами: диффузией под действием градиента концентрации и дрейфом под действием градиента электрического потенциала. Поскольку имеются два типа носителей — электроны и дырки, полный ток состоит из четырех составляющих:

$$j = (j_n)_{\text{др}} + (j_n)_{\text{диф}} + (j_p)_{\text{др}} + (j_p)_{\text{диф}}, \tag{2.56}$$

где индексы «др» и «диф» относятся соответственно к дрейфовым и диффузионным составляющим тока.

При анализе удобнее пользоваться не токами, а плотностями токов j, что и сделано в формуле (2.56). Там, где это не вызывает недоразумений, будем для краткости называть величину j током.

Составляющие тока. В одномерном случае, когда движение носителей происходит только вдоль оси x, дрейфовые составляющие записываются следующим образом:

$$(j_n)_{\text{др}} = qn\mu_n E = -qn\mu_n(\partial\varphi/\partial x); \tag{2.57а}$$

$$(j_p)_{\text{др}} = qn\mu_p E = -qn\mu_p(\partial\varphi/\partial x). \tag{2.57б}$$

Для диффузионных составляющих нужно вместо градиента электрического потенциала использовать градиенты химическо-

го потенциала соответствующих носителей. Химические потенциалы — это вторые слагаемые в правых частях формул (2.12). Продифференцируем эти слагаемые по x и подставим результаты вместо величины E в выражения (2.57). Тогда диффузионные составляющие токов запишутся следующим образом:

$$(j_n)_{\text{диф}} = q\mu_n\varphi_T \frac{dn}{dx} = qD_n \frac{dn}{dx}; \qquad (2.58а)$$

$$(j_p)_{\text{диф}} = -q\mu_p\varphi_T \frac{dp}{dx} = -qD_p \frac{dp}{dx}. \qquad (2.58б)$$

Коэффициенты D_n и D_p, которые вошли в выражения (2.58), называются *коэффициентами диффузии* электронов и дырок. Эти величины играют при диффузии ту же роль, что и подвижности при дрейфовом механизме движения. Связь между коэффициентами диффузии и подвижностями выражается *формулой Эйнштейна*:

$$D = \varphi_T\mu. \qquad (2.59)$$

Сравнивая выражения (2.57) и (2.58), можно заметить, что дрейфовые составляющие токов пропорциональны концентрациям носителей, тогда как диффузионные не зависят от концентраций, а определяются только градиентами концентрации.

Уравнения непрерывности. Выражения (2.57) и (2.58), говорят о том, что для оценки полного тока (2.56) необходимо, кроме распределения потенциала $\varphi(x)$, знать еще и распределения концентраций $n(x)$ и $p(x)$.

В общем случае концентрации зависят не только от координаты, но и от времени, т.е. мы имеем дело с функциями двух переменных: $n(x;t)$ и $p(x;t)$. Эти функции являются решениями так называемых *уравнений непрерывности потока*. Для электронов и дырок уравнения непрерывности записываются в следующем общем виде:

$$\frac{dn}{dt} = \Delta g - \frac{n - n_0}{\tau} + \frac{1}{q} div(j_n); \qquad (2.60а)$$

$$\frac{dp}{dt} = \Delta g - \frac{p - p_0}{\tau} + \frac{1}{q} div(j_p), \qquad (2.60б)$$

где $n - n_0 = \Delta n$ и $p - p_0 = \Delta p$ — избыточные концентрации, Δg — скорость генерации носителей под действием внешних факторов, например, света; τ — время жизни избыточных носителей.

Легко заметить, что уравнения непрерывности обобщают уравнение накопления носителей, которое было получено при анализе процессов рекомбинации [см. (2.44)]. Обобщение состоит в том, что при наличии тока в правую часть, помимо генерационного и рекомбинационного членов, добавляется еще один фактор, изменяющий концентрацию: дивергенция вектора потока (j/q).

В одномерном случае

$$div\left(\frac{j}{q}\right) = \frac{1}{q}\frac{\partial}{\partial x} j.$$

Используя эту операцию для всех четырех составляющих тока, описываемых выражениями (2.57) и (2.58), подставляя полученные значения в (2.60) и опуская генерационный член Δg, запишем уравнения непрерывности в следующей форме:

$$\frac{\partial n}{\partial t} = -\frac{n - n_0}{\tau} + D_n \frac{\partial^2 n}{\partial x^2} + \mu_n E \frac{\partial n}{\partial x} + n\mu_n \frac{\partial E}{\partial x}; \qquad (2.61а)$$

$$\frac{\partial p}{\partial t} = -\frac{p - p_0}{\tau} + D_p \frac{\partial^2 p}{\partial x^2} - \mu_p E \frac{\partial p}{\partial x} - p\mu_p \frac{\partial E}{\partial x}. \qquad (2.61б)$$

Последние слагаемые в правых частях (2.61) связаны с наличием объемных зарядов[1]. В условиях нейтральности эти слагаемые отсутствуют. Третьи слагаемые не связаны с наличием объемных зарядов и в ряде случаев должны учитываться, например, если имеется внутреннее электрическое поле, свойственное неоднородным полупроводникам.

1 При наличии объемных зарядов, помимо уравнений непрерывности, приходится для анализа привлекать еще уравнение Пуассона (2.26).

Если поле отсутствует или его влиянием можно заведомо пренебречь, полагаем $E = 0$. Тогда общие уравнения непрерывности (2.61) упрощаются:

$$\frac{\partial n}{\partial t} = -\frac{n - n_0}{\tau} + D_n \frac{\partial^2 n}{\partial x^2}; \qquad (2.62а)$$

$$\frac{\partial p}{\partial t} = -\frac{p - p_0}{\tau} + D_p \frac{\partial^2 p}{\partial x^2}. \qquad (2.62б)$$

Уравнения (2.62) называются *уравнениями диффузии*. Они широко используются при анализе полупроводниковых приборов.

Диффузия носителей. Пусть на поверхность полупроводника падает рассеянный пучок света (рис. 2.21). Тогда в тонком приповерхностном слое, в который проникает свет, будут генерироваться электронно-дырочные пары. Между поверхностью и объемом возникнут градиенты концентрации электронов и дырок и избыточные носители начнут диффундировать вглубь полупроводника. Такое совместное движение обоих типов носителей называют *биполярной* или *амбиполярной* диффузией.

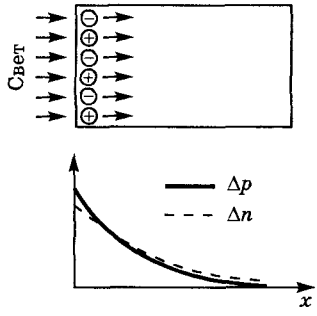


Рис. 2.21. Биполярная диффузия; эффект Дембера

Если бы подвижности (а значит, и коэффициенты диффузии) у электронов и дырок были одинаковы, то они двигались бы в виде единого нейтрального потока. На самом же деле подвижности носителей различны, поэтому у электронного потока будет тенденция «обогнать» дырочный поток. В результате небольшого взаимного сдвига потоков образуется небольшой объемный заряд и соответствующее электрическое поле; это поле тормозит поток электронов и ускоряет поток дырок. В конце концов устанавливается стационарный режим, при котором избыточные электроны и дырки распределены в виде сдвинутых относительно друг друга «облачков». Эти «облачка» двигаются синхронно, так что результирующий ток отсутствует. Концентрация носителей в «облачках» спадает с удалением от поверхности благодаря рекомбинации.

Описанные явления известны под названием *эффекта Дембера*, а электрическое поле и разность потенциалов, свойственные этому эффекту, называют *дембероВским полем* и *дембероВским напряжением*. Эффект Дембера существенен только при больших избыточных концентрациях и больших удельных сопротивлениях полупроводников.

На практике главную роль играет *монополярная диффузия*[1], характерная тем, что в приповерхностный слой полупроводника тем или иным способом вводится только один тип носителей — неосновных (рис. 2.22). Процесс введения неосновных носителей называется *инжекцией*.

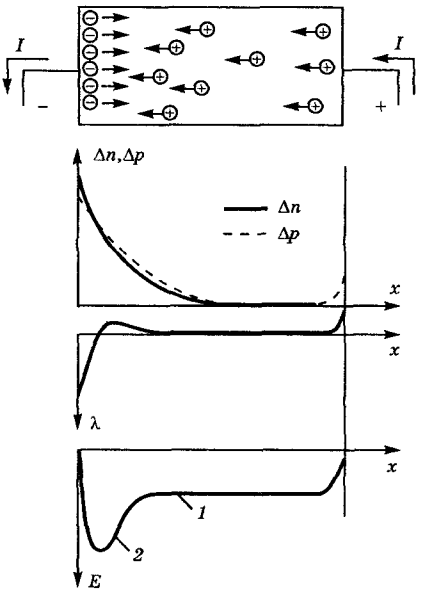


Рис. 2.22. Монополярная диффузия; инжекция носителей: *1* — омическое поле; *2* — дембероВское поле

Пусть для определенности осуществляется инжекция электронов в дырочный полупроводник. Инжектированные электроны благодаря градиенту концентрации будут диффундировать вглубь кристалла, т.е. появится электронный ток. Заряд избыточных электронов практически мгновенно (со временем диэлектрической релаксации) будет компенсирован таким же зарядом дырок, притягиваемых из глубоких слоев. В результате вблизи инжектирующей поверхности образуется квазинейтральное электронно-дырочное «облачко», почти такое же, как при биполярной диффузии. Несмотря на это внешнее сходство, монополярная диффузия принципиально отличается от биполярной следующими особенностями:

---

1 Учитывая большую распространенность монополярной диффузии, прилагательное «монополярная» обычно опускают.

- наличие тока предполагает, что полупроводник является элементом замкнутой цепи; значит, помимо дембероровского поля (сосредоточенного вблизи инжектирующей поверхности) во всей толще полупроводника действует «обычное» — омическое поле, обусловленное приложенным напряжением (рис. 2.22);
- потоки электронов и дырок направлены в разные стороны: электроны двигаются вглубь кристалла, а дырки в сторону инжектирующей поверхности — в район электронно-дырочного «облачка», где происходит интенсивная рекомбинация и необходимо пополнение основных носителей;
- в связи с постоянством полного тока его электронная и дырочная составляющие меняются в разные стороны: с удалением от поверхности электронный ток убывает (из-за рекомбинации), а дырочный ток растет; поэтому вдали от поверхности дырочная составляющая — главная и имеет чисто дрейфовый характер (дырки двигаются в поле, созданном внешним напряжением); наоборот, в непосредственной близости от поверхности ток почти чисто электронный и обусловлен диффузией, так как напряженность поля здесь близка к нулю (рис. 2.22).

Строгое решение задачи о распределении носителей при диффузии затруднительно. Обычно эту задачу решают в так называемом *диффузионном приближении* и для малых избыточных концентраций или, как говорят, для низкого уровня инжекции.

Диффузионное приближение означает, что для анализа используются уравнения диффузии (2.62), несмотря на заведомое наличие электрического поля. Справедливость диффузионного приближения оценивают в каждом конкретном случае.

Уровнем инжекции принято называть отношение концентрации избыточных носителей к равновесной концентрации основных носителей или, что то же самое — к концентрации примеси:

$$\delta = \Delta n / N. \qquad (2.63)$$

Низким уровнем инжекции считается значение $\delta << 1$. При этом действительно неравенство

$$\Delta n << n_0 + p_0, \qquad (2.64)$$

которое уже использовалось при выводе формул (2.43) и (2.49). Условие низкого уровня инжекции обеспечивает линейность уравнений диффузии. В условиях нейтральности $\Delta n \approx \Delta p$, поэтому выражения (2.63) и (2.64) действительны для обоих типов носителей.

Анализ процессов диффузии. Если интерес представляют только избыточные концентрации (как обычно и бывает), то достаточно использовать одно из двух уравнений диффузии, поскольку второе дает такой же результат в силу условия нейтральности ($\Delta n \approx \Delta p$). На самом деле функции $\Delta n(x)$ и $\Delta p(x)$ несколько различаются из-за различия в коэффициентах диффузии D_n и D_p, т.е. из-за эффекта Дембера. Однако в диффузионном приближении электрическими полями (в том числе и дембероовским) пренебрегают.

Найдем избыточную концентрацию Δn из уравнения (2.62а). Для этого подставим $n = n_0 + \Delta n$ и опустим индекс n у коэффициента диффузии. Кроме того, поделим обе части на D. Тогда уравнение диффузии принимает вид

$$\frac{\partial^2(\Delta n)}{\partial x^2} - \frac{\Delta n}{L^2} = \frac{1}{D}\frac{\partial(\Delta n)}{\partial t}. \qquad (2.65)$$

Стационарный вариант уравнения получается, если в правой части положить $\partial(\Delta n)/\partial t = 0$:

$$\frac{\mathrm{d}^2(\Delta n)}{\mathrm{d}x^2} - \frac{\Delta n}{L^2} = 0. \qquad (2.66)$$

Величина L, входящая в уравнения (2.65) и (2.66), выражается следующим образом:

$$L = \sqrt{D\tau}. \qquad (2.67)$$

Это — так называемая *средняя диффузионная длина*. Она характеризует то среднее расстояние, на которое носители успевают продиффундировать за время жизни. Поэтому отношение L/τ есть *средняя скорость диффузии носителей*. Диффузионная длина — одна из фундаментальных величин в полупроводниковой физике и технике. Для кремния типичные значения L составляют 5–20 мкм в зависимости от времени жизни.

Стационарное уравнение (2.66) — это обыкновенное линейное уравнение 2-го порядка. Его решение представляет собой сумму экспонент:

$$\Delta n(x) = A_1 e^{x/L} + A_2 e^{-x/L}, \qquad (2.68)$$

где коэффициенты A_1 и A_2 определяются из граничных условий. Примем $\Delta n(\infty) = 0$, т.е. предположим, что вдали от инжектирующей поверхности избыточные концентрации отсутствуют и эта часть полупроводника находится в равновесном состоянии. При таком граничном условии $A_1 = 0$. Полагая $x = 0$, получаем $A_2 = \Delta n(0)$; следовательно, распределение избыточной концентрации экспоненциальное (рис. 2.23, *а*):

$$\Delta n(x) = \Delta n(0) e^{-x/L}. \qquad (2.69)$$

Из этого выражения и из рис. 2.23, *а* следует, что на расстоянии диффузионной длины избыточная концентрация уменьшается в e раз. На участке длиной (3–4) L концентрация уменьшается в 20–50 раз, т.е. становится пренебрежимо малой по сравнению с граничной.

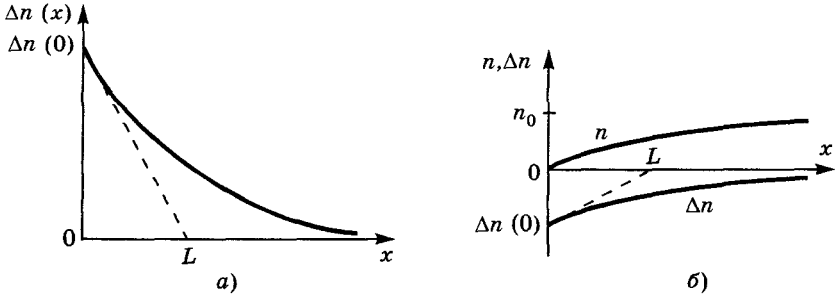


Рис. 2.23. Стационарное распределение избыточных носителей при инжекции (*а*) и экстракции (*б*)

Дифференцируя (2.69), получаем градиент концентрации

$$\frac{d(\Delta n)}{dx} = -\frac{\Delta n(0)}{L} e^{-x/L}. \qquad (2.70а)$$

Как видим, градиент концентрации, а значит, и диффузионный ток спадают по мере удаления от поверхности вглубь

кристалла. Градиент имеет максимальное (по модулю) значение при $x = 0$, т.е. на инжектирующей поверхности:

$$\left.\frac{\mathrm{d}(\Delta n)}{\mathrm{d}x}\right|_{x=0} = -\frac{\Delta n(0)}{L}. \qquad (2.70б)$$

Нестационарное уравнение (2.65) — это линейное дифференциальное уравнение 2-го порядка в частных производных.

Качественные выводы относительно переходного процесса состоят в следующем.

В начальный момент, когда $L(0) = 0$, градиент концентрации вблизи инжектирующей поверхности согласно (2.70б) оказывается бесконечно большим, и, следовательно, избыточные носители будут диффундировать в кристалл с очень большой скоростью. По мере нарастания величины $L(t)$ градиент концентрации на поверхности уменьшается и скорость диффузии становится все меньше. В конце концов при $t \approx 2\tau$ достигается стационарный режим. Исходя из таких соображений, на рис. 2.24 показаны примерные кривые распределения концентрации для нескольких моментов переходного процесса.

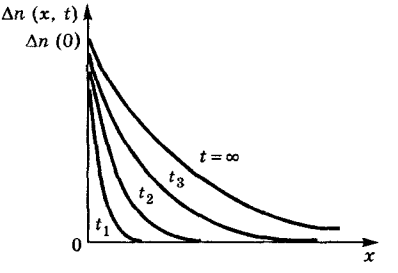


Рис. 2.24. Распределение инжектированных носителей во время переходного процесса

До сих пор избыточная концентрация Δn считалась положительной, поскольку в результате инжекции появляются дополнительные носители. Однако может быть и обратный процесс, когда из приповерхностного слоя полупроводника отбирается часть равновесных носителей.

Очевидно, что избыточная концентрация в случае экстракции будет отрицательной (рис. 2.23, *б*), так как количество носителей уменьшается по сравнению с равновесным состоянием.

Контрольные вопросы

1. Для чего служат индексы Миллера?
2. Какие вы знаете дефекты кристаллической решетки?
3. В чем заключается специфика поверхности кристалла?
4. Что такое примесь замещения?

5. Что такое основные и неосновные носители?
6. Нарисуйте энергетические зоны металла, полупроводника и диэлектрика. В чем их отличие?
7. Напишите условия электронейтральности.
8. Чем отличается распределение Ферми—Дирака от распределения Максвелла—Больцмана? Почему можно использовать второе распределение без больших погрешностей?
9. Что такое собственная концентрация?
10. Сформулируйте основное положение физики полупроводников в равновесной системе.
11. Запишите уравнение для дрейфа носителей.
12. Как подвижность зависит от температуры?
13. В чем отличие зависимости проводимостей собственного и примесного полупроводника от температуры?
14. Что называют эффектом поля и какое практическое использование в микроэлектронике он имеет?
15. Опишите явление эффекта поля в собственном полупроводнике.
16. Чем отличается режим обогащения от режима обеднения и как можно их достичь в примесном полупроводнике?
17. Поясните образование инверсионного слоя.
18. Расскажите о различных типах рекомбинации заряда.
19. Чем отличается непосредственная рекомбинация от рекомбинации на примесных центрах?
20. Какая разница между излучательной и безызлучательной рекомбинациями?
21. Что такое скорость рекомбинации?
22. Чем определяется эквивалентное время жизни неосновных носителей?
23. Что такое поверхностная рекомбинация и какой параметр вводится для ее определения?
24. Напишите выражение для дрейфовых составляющих тока.
25. Запишите формулы для диффузионных составляющих токов.
26. Что такое уравнение непрерывности потока?
27. В чем заключается разница между биполярной и амбиполярной диффузией? Как определить среднюю скорость диффузии?

Глава 3

Полупроводниковые переходы и контакты

3.1. Введение

Однородные полупроводники и однородные полупроводниковые слои находят весьма узкое применение: они непосредственно используются только в виде разного рода резисторов. Основные же элементы ИС (и основная масса дискретных полупроводниковых приборов) представляют собой сугубо неоднородные структуры. Два важнейших варианта таких структур — так называемый p–n-переход (контакт двух полупроводников с разным типом проводимости) и структура МП (контакт металла с полупроводником).

Ниже наиболее подробно рассматриваются p–n-переходы, без которых немыслима современная микроэлектроника, а также контакты МП, которые могут выполнять как функции диодов, так и функции обычных омических контактов. Последние неизбежны в любой ИС. Заключительный параграф посвящен явлениям, которые обусловлены контактом полупроводника с диэлектриком (в первую очередь — контакту кремния с двуокисью кремния). Эти явления влияют на характеристики p–n-переходов и особенно существенны при использовании эффекта поля.

3.2. Электронно-дырочные переходы

Комбинация двух полупроводниковых слоев с разным типом проводимости (рис. 3.1, *а*) обладает выпрямляющими или вентильными свойствами: она гораздо лучше пропускает в одном направлении, чем в другом. Полярность напряжения, соответствующая бóльшим токам, называется прямой, а меньшим — обратной. Обычно пользуются терминами *прямое* и *обратное напряжение*, *прямой* и *обратный ток*.

Выпрямительные свойства рассматриваемой структуры позволяют использовать ее в качестве полупроводникового диода. На рис. 3.1, *б* показаны символическое обозначение диода, направление прямого тока и полярность прямого напряжения.

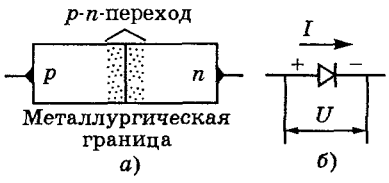


Рис. 3.1. Полупроводниковый диод: *а* — упрощенная структура; *б* — условное обозначение

Поверхность, по которой контактируют слои *p* и *n*, называется *металлургической границей*, а прилегающая к ней область объемных зарядов — электронно-дырочным переходом или *p–n-переходом*. Два других (внешних) контакта в диоде — невыпрямляющие, поэтому их называют омическими.

Структура *p–n*-перехода. Электронно-дырочные переходы классифицируют по резкости металлургической границы и по соотношению удельных сопротивлений слоев.

Ступенчатыми переходами называют переходы с идеальной границей, по одну сторону которой находятся доноры с постоянной концентрацией $N_д$, а по другую — акцепторы с постоянной концентрацией $N_а$. Такие переходы наиболее просты для анализа.

Плавными переходами называют такие, у которых в районе металлургической границы концентрация одного типа примеси постепенно уменьшается, а другого типа — растет. Сама металлургическая граница в этом случае соответствует равенству примесных концентраций ($N_д = N_а$).

По соотношению концентраций примесей в *p*- и *n*-слоях переходы делят на *симметричные*, *несимметричные* и *односторонние*. Симметричные переходы характерны условием $N_{дn} \approx N_{ap}$, где $N_{дn}$ и N_{ap} — концентрации примесей в соответствующих слоях.

Симметричные переходы не типичны для полупроводниковой техники.

В случае резкой асимметрии, когда концентрации примесей (а значит, и основных носителей) различаются на 1–2 порядка и более, переходы называют односторонними и обозначают символами $n^+–p$ или $p^+–n$, где верхний индекс «+» соответствует слою со значительно большей концентрацией.

На рис. 3.2 показана электрическая структура p–n-перехода. Для наглядности разница в концентрациях основных носителей n_{n0} и p_{p0} принята меньшей, чем это имеет место в действительности.

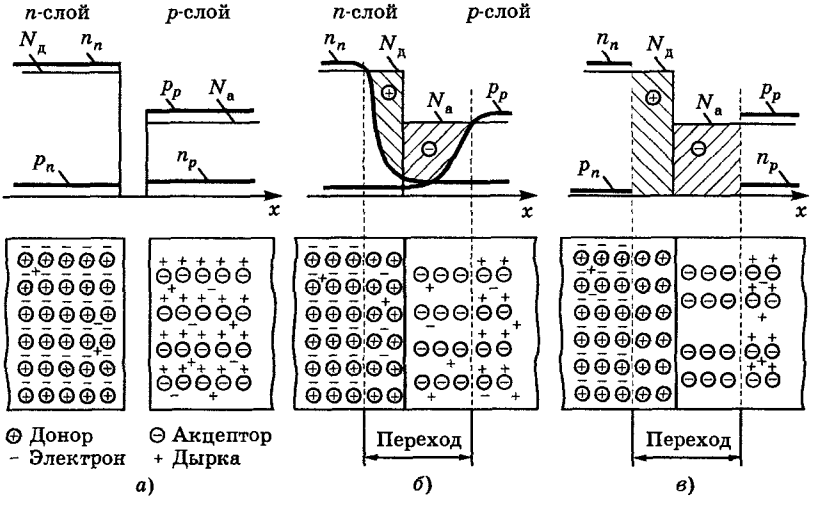


Рис. 3.2. Электрическая структура p–n-перехода: *а* — начальное состояние слоев; *б* — объемные заряды в реальном переходе; *в* — объемные заряды в идеализированном переходе

На самом деле они отличаются на несколько порядков, поэтому на графиках чаще всего для концентраций используется логарифмический масштаб. Поскольку концентрация электронов в слое n значительно больше, чем в слое p, часть электронов диффундирует из слоя n в слой p. При этом в слое p вблизи металлургической границы окажутся избыточные электроны. Эти электроны будут рекомбинировать с дырками до тех пор, пока не будет выполнено условие равновесия (2.10). Соответственно концентрация дырок в этой области уменьшится и «обнажатся» некомпенсированные отрицательные заряды акцепторных атомов. Слева от металлургической границы «обнажатся» некомпенсированные положительные заряды донорных атомов, от которых ушли электроны (рис. 3.2, *б*). Аналогичные рассуждения можно провести для дырок, которые диффундируют из слоя p в слой n. Однако в одностороннем переходе, в котором $p_{p0} \ll n_{n0}$, перемещение дырок мало существенно, так

как их градиент концентрации несравненно меньше градиента концентрации электронов.

Образовавшиеся объемные заряды и связанные с ними поля обеспечивают равновесие в области p–n-перехода. Область объемных зарядов называют обедненным слоем, имея в виду резко пониженную концентрацию подвижных носителей в обеих ее частях.

В большинстве случаев p–n-переход можно идеализировать так, как показано на рис. 3.2, *в*, т.е. полностью пренебречь наличием свободных носителей в переходе и считать границы перехода идеально резкими. Такая идеализация упрощает решение многих задач.

Переход в целом нейтрален: положительный заряд в левой части равен отрицательному заряду в правой части. Однако плотности зарядов резко различны (из-за различия в концентрациях примесей). Поэтому различны и протяженности обедненных слоев: в слое с меньшей концентрацией примеси (в нашем случае в p-слое) область объемного заряда значительно шире. Как говорят, несимметричный переход сосредоточен в высокоомном слое.

На рис. 3.3 показано распределение носителей в полулогарифмическом масштабе, более удобном для количественных оценок и сравнений. Следует обратить внимание на тот факт, что внутри p–n-перехода имеется участок с собственной, т.е. минимальной концентрацией носителей. Поэтому область перехода является наиболее высокоомной частью всей диодной структуры. Удельное сопротивление в этой области на несколько порядков превышает удельные сопротивления нейтральных n- и p-областей.

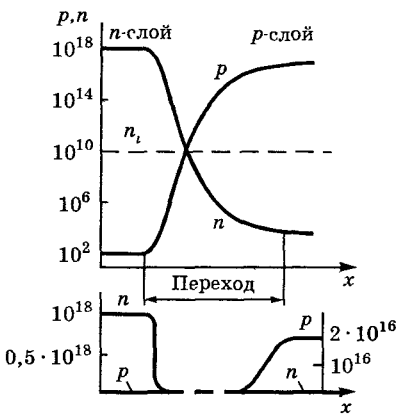


Рис. 3.3. Распределение носителей в несимметричном переходе (полулогарифмический и линейный масштабы)

На рис. 3.4 показаны зонные диаграммы p–n-перехода до и после воображаемого «соприкосновения» слоев. Как видим, единство уровня Ферми в

равновесном p–n-переходе приводит к искривлению зон в районе металлургической границы. В результате образуются разность потенциалов (потенциальный барьер) и электрическое поле, свойственное равновесию.

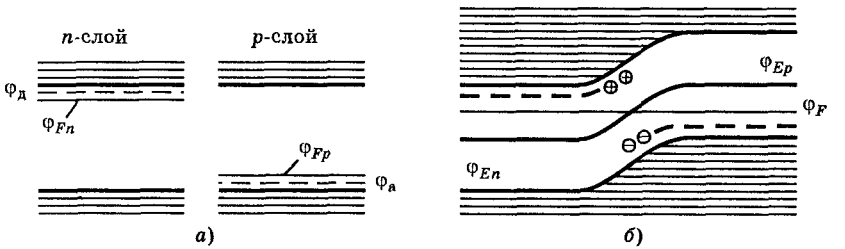


Рис. 3.4. Зонная диаграмма слоев (*а*) и p–n-перехода в равновесном состоянии (*б*)

Анализ равновесного p–n-перехода. Высота равновесного потенциального барьера определяется разностью электростатических потенциалов в p- и n-слоях (рис. 3.4, *б*):

$$\Delta\varphi_0 = \varphi_{Ep} - \varphi_{En}. \qquad (3.1)$$

Потенциалы φ_{Ep} и φ_{En} легко получить из выражений (2.12), подставляя соответственно $p = p_{p0}$ и $n = n_{n0}$ (индексы n и p обозначают принадлежность к тому или иному слою, а индекс 0 — равновесное состояние). Тогда

$$\Delta\varphi_0 = \varphi_T \ln(n_{n0} p_{p0}/n_i^2). \qquad (3.2а)$$

Если положить $n_{n0} = N_д$ и $p_{p0} = N_а$ (где $N_д$ и $N_а$ — эффективные концентрации примесей), то

$$\Delta\varphi_0 = \varphi_T \ln(N_д N_а/n_i^2). \qquad (3.2б)$$

Очевидно, что при прочих равных условиях равновесная высота потенциального барьера тем выше, чем меньше собственная концентрация (т.е. чем больше ширина запрещенной зоны полупроводника). Подставляя в (3.2б) $N_д = 10^{19}$ см$^{-3}$, $N_а = 10^{16}$ см$^{-3}$ и значение $n_i = 2 \cdot 10^{10}$ для кремния, получаем для комнатной температуры $\Delta\varphi_0 = 33\varphi_T \approx 0{,}83$ В.

Используя соотношение (2.10), заменим в формуле (3.2а) одну из концентраций основных носителей (n_{n0} или p_{p0}) на концентрацию неосновных (p_{n0} или n_{p0}). При этом оказывается,

что равновесная *высота потенциального барьера определяется отношением концентраций однотипных носителей* (электронов или дырок) *по обе стороны перехода*, на его границах:

$$\Delta\varphi_0 = \varphi_T \ln(n_{n0}/n_{p0}); \qquad (3.3а)$$

$$\Delta\varphi_0 = \varphi_T \ln(p_{p0}/p_{n0}). \qquad (3.3б)$$

Эти варианты записи мы используем позднее при анализе неравновесного состояния перехода.

Чтобы рассчитать равновесную ширину перехода, воспользуемся идеализированным распределением зарядов (см. рис. 3.2, *в*). При таком распределении (рис. 3.5, *а*) плотности заряда в каждой из двух частей перехода постоянны (рис. 3.5, *б*):

в левой части (в n^+-слое) $\lambda_n = qN_д$;

в правой части (в p-слое) $\lambda_p = -qN_а$.

Подставляя эти значения в уравнение Пуассона (2.26) и интегрируя его дважды для каждой из двух частей перехода, получаем линейное распределение напряженности E и квадратичное распределение электрического потенциала φ (рис. 3.5, *в*, *г*).

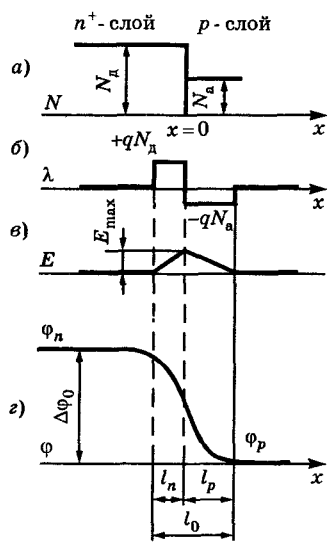


Рис. 3.5. Распределение концентрации примесей (*а*), плотности объемного заряда (*б*), напряженности поля (*в*) и потенциала (*г*) в ступенчатом n^+–p-переходе

Функция $E(x)$ имеет вид:

$$E(x) = \frac{qN_д}{\varepsilon_0\varepsilon}(l_n + x); \quad x \le 0$$

$$E(x) = \frac{qN_а}{\varepsilon_0\varepsilon}(l_p - x); \quad x \ge 0. \qquad (3.4)$$

Приравнивая значения $E(x)$ при $x = 0$ (на металлургической границе), получаем соотношение между составляющими ширины перехода в n- и p-слоях:

$$l_n/l_p = N_а/N_д. \qquad (3.5)$$

В несимметричном (и тем более одностороннем) переходе типа n^+–p выполняется неравенство $N_д >> N_а$.

Значит, $l_n << l_p$ и полная ширина перехода близка к составляющей в высокоомном слое: $l_0 \approx l_p$.

Функция $\varphi(x)$ имеет вид:

$$\varphi(x) = \varphi_n - \frac{qN_д}{2\varepsilon_0\varepsilon}(x + l_n)^2; \quad x \le 0;$$

$$\varphi(x) = \varphi_p + \frac{qN_а}{2\varepsilon_0\varepsilon}(x - l_p)^2; \quad x \ge 0.$$

Приравнивая значения $\varphi(x)$ при $x = 0$ и учитывая, что $\varphi_n - \varphi_p = \Delta\varphi_0$, получаем:

$$\Delta\varphi_0 = \frac{qN_д}{2\varepsilon_0\varepsilon} l_n^2 + \frac{qN_а}{2\varepsilon_0\varepsilon} l_p^2.$$

Для несимметричных переходов одним из слагаемых можно пренебречь. Например, для n^+–p-перехода, у которого $l_n << l_p$, можно пренебречь первым слагаемым и положить $l_p = l_0$. Поэтому, опуская для общности индексы, запишем ширину потенциального барьера в несимметричном переходе следующим образом:

$$l_0 = \sqrt{\frac{2\varepsilon_0\varepsilon\Delta\varphi_0}{qN}}, \tag{3.6}$$

где N — концентрация примеси в высокоомном слое перехода. Полагая $\Delta\varphi_0 = 0{,}8$ В и $N = 10^{16}$ см$^{-3}$, получаем для кремния $l_0 \approx 0{,}3$ мкм.

При анализе плавных переходов считают, что на коротком участке (в пределах ширины перехода) распределение эффективной концентрации примеси линейное. Тогда и плотности объемных зарядов в переходе можно считать линейными функциями координаты x (рис. 3.6). В этом случае решение уравнения Пуассона приводит к квадратичной функции $E(x)$ и кубической функции $\varphi(x)$. Используя эти функции так же, как и выше для ступенчатого перехода, можно получить ширину равновесного плавного перехода в следующем виде:

$$l_0 = \sqrt[3]{\frac{9\varepsilon_0\varepsilon\Delta\varphi_0}{qN'}}, \tag{3.7}$$

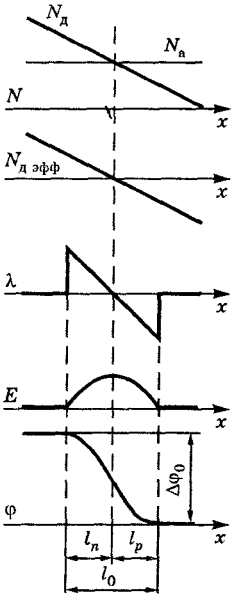


Рис. 3.6. Распределение концентрации примесей, эффективной концентрации донорной примеси, плотности объемного заряда, напряженности поля и потенциала в плавном n^+–p-переходе

где N' — градиент эффективной концентрации. Поскольку градиент одинаков в обеих частях перехода, то и ширина l_0 делится поровну между n- и p-слоями, т.е. плавный переход симметричен.

Что касается высоты равновесного потенциального барьера, то она определяется формулой (3.2б), если под $N_д$ и $N_а$ понимать эффективные концентрации примесей на соответствующих границах перехода.

Анализ неравновесного p–n-перехода. Если подключить источник э.д.с. U между p- и n-слоями, то равновесие перехода нарушится и в цепи потечет ток. Выше уже подчеркивалось, что удельное сопротивление обедненного слоя намного выше, чем удельные сопротивления нейтральных слоев. Поэтому внешнее напряжение практически полностью падает на переходе, а значит, изменение высоты потенциального барьера равно значению приложенной э.д.с.

Когда э.д.с. U приложена плюсом к p-слою, высота барьера уменьшается (рис. 3.7, *а*):

$$\Delta\varphi = \Delta\varphi_0 - U. \qquad (3.8)$$

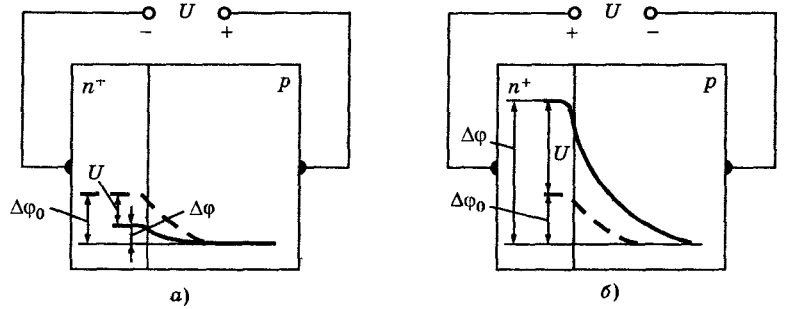


Рис. 3.7. Смещение перехода в прямом (*а*) и обратном (*б*) направлениях

Напряжение такой полярности является прямым. При отрицательном потенциале на p-слое высота барьера увеличивается (рис. 3.7, *б*) и знак минус в формуле (3.8) следует изменить на плюс. Напряжение такой полярности является обратным. В дальнейшем прямые напряжения будем считать положительными, а обратные — отрицательными.

Вместе с высотой потенциального барьера меняются его ширина и граничные концентрации носителей.

Подставляя значение $\Delta\varphi$ из (3.8) в (3.6), получаем ширину неравновесного барьера в виде:

$$l = \sqrt{\frac{2\varepsilon_0\varepsilon(\Delta\varphi_0 - U)}{qN}}. \qquad (3.9)$$

Как видим, переход сужается при прямом напряжении ($U>0$) и расширяется при обратном ($U<0$).

Изменение высоты потенциального барьера сопровождается, вообще говоря, изменением всех четырех граничных концентраций. Однако поскольку концентрации основных носителей значительно больше, чем неосновных, можно считать, что меняются только последние. Поэтому заменим в правых частях формул (3.3) концентрации n_{p0} на n_p и p_{n0} на p_n, а в левых частях — величину $\Delta\varphi_0$ на $\Delta\varphi$. После этого, подставляя значения $\Delta\varphi_0$ из (3.3), нетрудно установить связь между граничными концентрациями неосновных носителей в равновесном и неравновесном состояниях перехода:

$$n_p = n_{p0}\mathrm{e}^{U/\varphi_T}; \qquad (3.10а)$$

$$p_n = p_{n0}\mathrm{e}^{U/\varphi_T}. \qquad (3.10б)$$

При прямых напряжениях граничные концентрации оказываются больше равновесных. Значит, в каждом из слоев появляются избыточные носители, т.е. происходит инжекция (см. рис. 2.23а). При обратных напряжениях граничные концентрации уменьшаются по сравнению с равновесными, т.е. имеет место экстракция (см. рис. 2.23б).

Избыточные концентрации на границах перехода найдем, вычитая из n_p и p_n, соответственно n_{p0} и p_{n0}:

$$\Delta n_p = n_{p0}(\mathrm{e}^{U/\varphi_T} - 1); \qquad (3.11а)$$

$$\Delta p_n = p_{n0}(\mathrm{e}^{U/\varphi_T} - 1). \qquad (3.11б)$$

Поделив (3.11а) на (3.11б), заменив в правой части концентрации неосновных носителей концентрациями основных с помощью соотношения (2.9) и считая концентрации основных носителей равными концентрациям соответствующих примесей, получим

$$\Delta n_p / \Delta p_n = N_д / N_а . \qquad (3.12)$$

Отсюда следует, что у несимметричных переходов концентрация избыточных носителей в высокоомном слое (с малой концентрацией) гораздо больше, чем в низкоомном. Можно сказать, что в несимметричных переходах инжекция имеет односторонний характер: главную роль играют носители, инжектируемые из низкоомного (сильно легированного) слоя в высокоомный.

Инжектирующий слой (с меньшим удельным сопротивлением) называют *эмиттером*, а слой с большим удельным сопротивлением, в который инжектируются неосновные для него носители, — *базой*.

При обратных напряжениях, т.е. в режиме экстракции, граничные концентрации согласно (3.10) меньше равновесных и могут быть сколь угодно малыми. При этом избыточные концентрации согласно (3.11) — отрицательны, по модулю они не превышают равновесных значений n_{p0} и p_{n0}.

Вольт-амперная характеристика *p–n*-перехода. На рис. 3.8 показана структура токов в переходе. В общем случае (рис. 3.8, *а*) ток состоит из *электронной* и *дырочной* составляющих (нижние индексы *n* и *p*), а каждая из них, в свою очередь, из *инжекционной* и *рекомбинационной* составляющих (верхние индексы «и» и «р»). Последние обусловлены рекомбинацией носителей в области объемного заряда, на пути в смежный слой. Рекомбиционные составляющие электронного и дырочного токов, разумеется, одинаковы. Эти составляющие иногда играют важную роль, однако при выводе идеализированной вольт-амперной характеристики мы ими пренебрежем и будем пользоваться структурой, показанной на рис. 3.8, *б*.

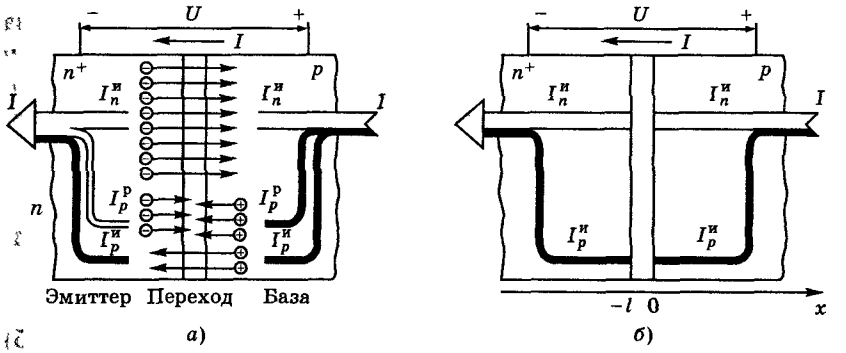


Рис. 3.8. Структура тока в n^+–p-переходе с учетом (*а*) и без учета (*б*) рекомбинации в области объемного заряда

Эта структура характерна тем, что электронная и дырочная составляющие одинаковы в обоих слоях, т.е. не претерпевают изменений в области перехода.

Для расчета инжекционых составляющих воспользуемся тем обстоятельством, что на границах перехода электрическое поле равно нулю (рис. 3.5, *в*), а значит, токи инжектированных носителей чисто диффузионные. Диффузионные токи рассчитываются по формулам (2.58). Запишем граничные градиенты концентраций в форме (2.70б), считая, что ширина перехода $l = 0$ (см. выше), т.е. базовая и эмиттерная границы перехода совпадают. Тогда:

$$\left.\frac{\mathrm{d}(\Delta n)}{\mathrm{d}x}\right|_{x=0} = -\frac{\Delta n_p}{L_n}; \qquad \left.\frac{\mathrm{d}(\Delta p)}{\mathrm{d}x}\right|_{x=0} = \frac{\Delta p_n}{L_p}.$$

Знак плюс для дырочного градиента обусловлен тем, что дырки инжектируются из базы в эмиттер, т.е. в направлении отрицательных значений x (см. рис. 3.8). Подставляя эти градиенты в (2.58) и используя значения (3.11), получаем электронную и дырочную составляющие тока в виде

$$j_n = -\frac{qD_n}{L_n} n_{p0}\left(\mathrm{e}^{U/\varphi_T} - 1\right); \tag{3.13а}$$

$$j_p = -\frac{qD_p}{L_p} p_{n0}\left(\mathrm{e}^{U/\varphi_T} - 1\right). \tag{3.13б}$$

Знаки минус обусловлены только тем, что ось x направлена от отрицательного полюса э.д.с. U к положительному (рис. 3.8). С физической точки зрения токи протекают от «плюса» к «минусу», т.е. являются положительными. Суммируя величины j_n и j_p, умножая их сумму на площадь перехода S и опуская знак минус, запишем ВАХ p–n-перехода в виде

$$I = I_0(e^{U/\varphi_T} - 1), \tag{3.14}$$

где

$$I_0 = \frac{qD_nS}{L_n} n_{p0} + \frac{qD_pS}{L_p} p_{n0}. \tag{3.15}$$

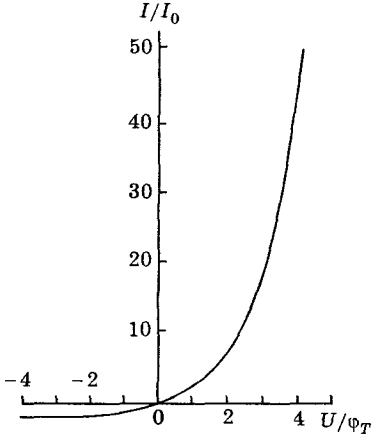


Рис 3.9. Вольт амперная характеристика идеализированного p–n перехода (диода)

Формула (3.14) — одна из важнейших в транзисторной электронике. Ее начальный участок представлен на рис. 3.9 в относительных единицах. ВАХ, определяемую формулой (3.14), называют *идеальной*, так как при ее выводе были опущены многие привходящие факторы. Реальные ВАХ отличаются от идеальной, но соответствующие выражения оказываются не только менее наглядными, чем (3.14), но и сложными для практического использования. Поэтому при анализе и расчете полупроводниковых приборов всегда пользуются формулой (3.14).

Ток I_0, определяющий «масштаб» ВАХ, называется *тепловым током*, поскольку он имеет тепловое происхождение и, как увидим, сильно зависит от температуры. Из формулы (3.14) видно, что при достаточно большом обратном напряжении (а именно, при $|U| > 3\varphi_T$) величина обратного тока равна $-I_0$ и не зависит от напряжения. Таким образом, можно сказать, что тепловой ток определяет величину обратного тока для идеальной ВАХ. Кроме того, тепловой ток влияет на многие важные параметры p–n-переходов и транзисторов; поэтому рассмотрим его подробнее.

В несимметричных переходах составляющие тока j_n и j_p существенно различаются из-за различия избыточных концентраций [см. (3.12)]. Соответственно различаются и составляющие теплового тока. Для n^+–p-перехода главной является электронная составляющая, т.е. первое слагаемое в правой части (3.15). Запишем эту составляющую, заменяя n_{p0} на n_i^2/p_{p0} согласно (2.10) и полагая $p_{p0} = N_a$. Для общности опустим индексы, тогда

$$I_0 = (qDS/LN)n_i^2. \qquad (3.16)$$

Если $N = 10^{16}$ см$^{-3}$ и $L = 10$ мкм, то для кремния при комнатной температуре плотность теплового тока $j_0 \approx 2\ 10^{-10}$ А/см^2. У современных интегральных транзисторов площади баз составляют не более 210^{-5} см^2, а площади эмиттеров — до 10^{-6} см^2 и менее. Поэтому при комнатной температуре типичным значением теплового тока у интегрального кремниевого перехода можно считать $I_0 = 10^{-15}$ А.

Температурная зависимость теплового тока определяется величиной n_i^2. Используя формулу (2.9), запишем тепловой ток в виде

$$I_0 = I_{00} e^{\varphi_a/\varphi_T}. \qquad (3.17)$$

Небольшие изменения температуры приводят к резкому изменению теплового тока.

Прямая ветвь ВАХ. При напряжениях $U > 0$ зависимость $I(U)$ настолько крутая, что получить нужный ток, задавая напряжение, очень трудно: малейшее изменение напряжения вызывает существенное изменение тока. Поэтому для p n-переходов характерен режим заданного прямого тока. Чтобы исследовать зависимость $U(I)$, запишем ВАХ (3.14) в следующей форме:

$$U = \varphi_T \ln\left(\frac{I}{I_0} + 1\right).$$

Для кремниевых переходов с их ничтожным тепловым током можно пользоваться выражением

$$U = \varphi_T \ln(I/I_0). \qquad (3.18)$$

Если диапазон изменения прямых токов составляет до 2-х порядков и более, то прямое напряжение в этом диапазоне может меняться существенно. Однако на практике диапазон токов редко бывает таким широким. Тогда прямое напряжение в рабочем диапазоне меняется очень незначительно.

Различают два режима работы перехода — нормальный, когда ток порядка миллиампера, и микрорежим, когда ток порядка микроампера.

В зависимости от диапазона токов прямые напряжения несколько различаются, но в пределах диапазона их можно считать постоянными и рассматривать как своего рода параметр открытого кремниевого перехода. Для этого параметра введем специальное обозначение U^* и будем называть его *напряжением открытого перехода.* При комнатной температуре будем считать, что в нормальном режиме $U^* = 0{,}7$ В, а в микрорежиме $U^* = 0{,}5$ В. Напряжение U^* зависит от температуры при неизменном токе.

Одной из особенностей реальной ВАХ является омическое падение напряжения в слое базы. Действительно, если базовый слой имеет достаточно большое сопротивление $r_б$, то внешнее напряжение, вообще говоря, не полностью падает на p-n-переходе, а распределяется между ним и слоем базы. Тогда прямое напряжение вместо формулы (3.18) будет выражаться суммой

$$U = \varphi_T \ln(I/I_0) + Ir_б. \qquad (3.19)$$

Один из важнейших параметров прямой ветви ВАХ — *дифференциальное сопротивление перехода.* Для начального (невырожденного) участка его легко получить из выражения (3.18):

$$r_{p-n} = \mathrm{d}U/\mathrm{d}I = \varphi_T/I. \qquad (3.20)$$

Физический смысл этого параметра становится ясным, если заменить дифференциалы $\mathrm{d}U$ и $\mathrm{d}I$ конечными приращениями, тогда

$$\Delta U = \Delta I r_{p-n}.$$

Значит, r_{p-n} *есть сопротивление для приращений тока* ΔI, *малых по сравнению с постоянной составляющей тока* I, *определяющей величину* r_{p-n}.

Типичным значением, которое полезно запомнить, является $r_{p-n} = 25$ Ом, соответствующее току $I = 1$ мА.

Обратная ветвь ВАХ. Как отмечалось, реальный обратный ток перехода намного превышает величину I_0, предсказываемую выражением (3.15). Причиной этого является, прежде всего, генерация электронно-дырочных пар в области обратносмещенного перехода. Составляющую обратного тока, обусловленную этим явлением, называют *током термогенерации.*

Процессы генерации и рекомбинации носителей происходят во всех частях диода — как в нейтральных слоях n и p, так и в области перехода. В равновесом состоянии скорости генерации и рекомбинации везде одинаковы, поэтому направленных потоков носителей нет. Когда к переходу приложено обратное напряжение, область перехода дополнительно обедняется носителями. Поэтому рекомбинация здесь замедляется и процесс генерации оказывается неуравновешенным. Избыточные генерируемые носители уносятся электрическим полем в нейтральные слои: электроны в n-слой, дырки в p-слой. Эти потоки и образуют ток термогенерации I_G:

$$I_G = I_{G0} e^{-\varphi_з/2\varphi_T}. \qquad (3.21)$$

Ток термогенерации зависит от обратных напряжения и температуры, он на 4–5 порядков больше тока I_0.

Пробой перехода. Различают три вида (механизма) пробоя p-n-перехода при достаточно больших обратных напряжениях: *туннельный, лавинный и тепловой.* Первые два связаны с увеличением напряженности электрического поля в переходе, а третий — с увеличением рассеиваемой мощности и соответственно температуры.

В основе туннельного пробоя лежит туннельный эффект, т.е. «просачивание» электронов сквозь тонкий потенциальный барьер (рис. 3.10).

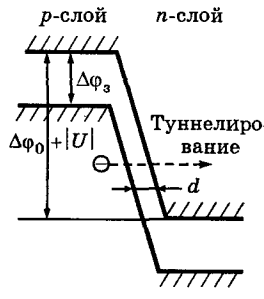


Рис. 3.10. Туннельный пробой (зонная диаграмма)

В основе лавинного пробоя лежит «размножение» носителей в сильном электрическом поле, действующем в области перехода. Электрон и дырка, ускоренные полем на длине свободного пробега, могут разорвать одну из ковалентных связей нейтрального атома полупроводника. В результате рождается

новая пара электрон–дырка и процесс повторяется уже с участием новых носителей. Обратный ток при этом, естественно, возрастает. При достаточно большой напряженности поля, когда исходная пара носителей в среднем порождает более одной новой пары, ионизация приобретает лавинный характер. Граничное значение пробивного напряжения, выше которого пробой лавинный, а ниже — туннельный, составляет около 5 В.

В основе теплового пробоя лежит саморазогрев перехода при протекании обратного тока. С ростом температуры обратные токи резко возрастают, соответственно увеличивается мощность, рассеиваемая в переходе; это вызывает дополнительный рост температуры и т.д. Характерной особенностью ВАХ при тепловом пробое является наличие участка с отрицательным дифференциальным сопротивлением: $\mathrm{d}U/\mathrm{d}I < 0$.

Барьерная емкость. Полупроводниковый диод инерционен по отношению к достаточно быстрым изменениям тока или напряжения, поскольку новое распределение носителей устанавливается не сразу. Как известно, внешнее напряжение меняет ширину перехода, а значит, и величину объемных зарядов в переходе. Кроме того, при инжекции или экстракции меняются заряды в области базы. Следовательно, диод обладает емкостью, которую можно считать подключенной параллельно p–n-переходу.

Эту емкость принято разделять на две составляющие: *барьерную емкость*, отражающую перераспределение зарядов в переходе, и *диффузионную емкость*, отражающую перераспределение в базе. Такое разделение в общем условное, но оно удобно на практике, поскольку соотношение обеих емкостей различно при разных полярностях приложенного напряжения. При прямом напряжении главную роль играют избыточные заряды в базе и соответственно — диффузионная емкость. При обратном напряжении избыточные заряды в базе малы и главную роль играет барьерная емкость.

Заметим заранее, что обе емкости нелинейны: диффузионная емкость зависит от прямого тока, а барьерная — от обратного напряжения. Нелинейность диффузионной емкости настолько существенна, что ее использование при решении практически важных задач не упрощает анализ по сравнению с непосредственным использованием уравнений непрерывности. Поэтому мы эту емкость не рассматриваем. Что касается барьерной емкости,

то она, несмотря на нелинейность, является удобным параметром и заслуживает рассмотрения.

Примем распределение зарядов в переходе ступенчатым. Будем по-прежнему считать переход несимметричным, типа n^+–p. Тогда протяженность отрицательного заряда в p-базе можно считать равной всей ширине перехода: $l_p = l$. Запишем модуль этого заряда:

$$|Q| = qNSl,$$

где N — концентрация примеси в базе; S — площадь перехода. Такой же (но положительный) заряд будет в эмиттерном n^+-слое.

Представим, что эти заряды расположены на обкладках воображаемого конденсатора. Емкость конденсатора обычно определяют как Q/U. Однако в данном случае заряд обусловлен не только внешним напряжением U, но и равновесной высотой барьера $\Delta\varphi_0$, см. (3.9). Поэтому заряд $|Q|$ нужно делить на величину $\Delta\varphi = \Delta\varphi_0 - U$. В транзисторной электронике чаще пользуются дифференциальной барьерной емкостью, которую получают, дифференцируя заряд $|Q|$ по напряжению $\Delta\varphi$. Учитывая (3.9), получаем удельную барьерную емкость (на единицу площади):

$$C_{б0} = \sqrt{\frac{0{,}5\varepsilon_0\varepsilon qN}{\Delta\varphi_0 - U}}. \qquad (3.22)$$

Так как напряжение обратное, т.е. имеет отрицательный знак, удобно пользоваться модулем обратного напряжения:

$$C_{б0} = \sqrt{\frac{0{,}5\varepsilon_0\varepsilon qN}{\Delta\varphi_0 + |U|}}. \qquad (3.23)$$

3.3. Контакты полупроводник–металл

Исторически первыми полупроводниковыми приборами были диоды, основанные на контакте полупроводника с металлом (*точечно-контактные диоды*). Их применение основывалось на экспериментально обнаруженном факте — выпрямлении слабых переменных сигналов при соприкосновении металлической иглы с кристаллами некоторых естественных полупроводниковых минералов. В интегральных схемах кон-

такты металла с полупроводником находят двоякое применение: либо в качестве невыпрямляющих — омических контактов (соединения элементов ИС), либо в качестве специфических выпрямляющих контактов (диодов Шоттки).

Структура и свойства контактов металл–полупроводник зависят в первую очередь от взаимного расположения уровней Ферми в том и другом слое. На рис. 3.11 вверху показаны зонные диаграммы разделенных слоев, а внизу — зонные диаграммы соответствующих контактов (после «соприкосновения» слоев и установления равновесия).

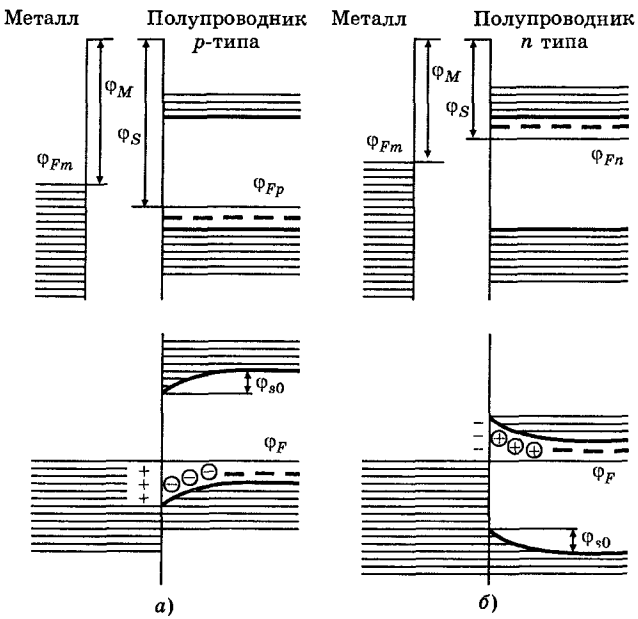


Рис. 3.11. Зонные диаграммы выпрямляющих контактов металла с полупроводником *а* — контакт с полупроводником *p* типа, *б* — контакт с полупроводником *n*-типа

Выпрямляющие контакты. На рис. 3.11, *а* показаны зонные диаграммы для случая, когда $\varphi_{Fm} > \varphi_{Fp}$. Такое соотношение означает, что заполненность зоны проводимости в полупроводнике меньше, чем заполненность такого же энергетического участка в металле. Поэтому после «соприкосновения» слоев часть электронов перейдет из металла в полупроводник *p*-типа. Появление дополнительных электронов в приповерхностном

слое полупроводника приводит к усиленной рекомбинации. В результате уменьшается количество основных носителей — дырок, и вблизи границы с металлом «обнажаются» некомпенсированные отрицательные ионы акцепторов. Появляется электрическое поле, которое препятствует дальнейшему притоку электронов и обеспечивает больцмановское равновесие в области контакта. Энергетические уровни оказываются искривленными «вниз».

На рис. 3.11, *б* показаны зонные диаграммы для случая $\varphi_{Fm} < \varphi_{Fn}$, когда после соприкосновения слоев электроны переходят из полупроводника *n*-типа в металл. Соответственно вблизи границы с металлом «обнажаются» некомпенсированные положительные ионы доноров, а зоны искривляются «вверх».

Область искривления зон (т.е. область объемных зарядов) в обоих случаях имеет протяженность, определяемую формулой (2.34), обычно 0,1–0,2 мкм.

Контакты такого рода в настоящее время создаются напылением металла на полупроводник в вакууме.

Обмен электронами между металлом и полупроводником обычно характеризуют не разностью «исходных» уровней Ферми, а разностью *работ выхода*. Работой выхода электрона из твердого тела называют энергию, необходимую для вылета за пределы кристалла (т.е. для термоэмиссии). На зонных диаграммах работа выхода есть энергетическое «расстояние» между уровнем свободного электрона вне твердого тела и уровнем Ферми. На рис. 3.11 работы выхода из металла и из полупроводника обозначены соответственно через φ_M и φ_S. Разность работ выхода $\varphi_{MS} = \varphi_M - \varphi_S$, выраженную в вольтах, называют *контактной разностью потенциалов*.

В зависимости от соотношения работ выхода φ_M и φ_S электроны переходят в тот или иной слой. Если $\varphi_M < \varphi_S$ (т.е. $\varphi_{MS} < 0$), то электроны переходят из металла в полупроводник (рис. 3.11, *а*), если же $\varphi_M > \varphi_S$ (т.е. $\varphi_{MS} > 0$), то электроны переходят из полупроводника в металл (рис. 3.11, *б*). Такой критерий более нагляден, чем использованный в начале раздела, тем более, что контактные разности потенциалов для типовых комбинаций металлов и полупроводников приводятся в литературе.

Степень искривления энергетических зон вблизи поверхности (рис. 3.11) характеризуется величиной равновесного повер-

хностного потенциала φ_{S0}. Если пренебречь ролью поверхностных состояний, то величина φ_{S0} будет равна контактной разности потенциалов φ_{MS}.

Оба контакта, показанные на рис. 3.11, характерны наличием обедненных слоев в приконтактном слое полупроводника. Здесь концентрация основных носителей меньше по сравнению с равновесной, сохранившейся вдали от контакта. Следовательно, такой приконтактный слой обладает повышенным удельным сопротивлением и поэтому определяет сопротивление всей системы.

Потенциальный барьер в приконтактном слое называют *барьером Шоттки.* Его высота φ_{S0} является аналогом величины $\Delta\varphi_0$ в p–n-переходе. В зависимости от полярности приложенного внешнего напряжения потенциал φ_S и соответственно сопротивление приконтактного слоя будут меняться.

Так, если напряжение приложено плюсом к металлу, а минусом к полупроводнику, то потенциальный барьер в контакте на рис 3.11, *а* повышается. Тогда приконтактный слой еще больше обедняется основными носителями — дырками и, следовательно, будет иметь повышенное сопротивление по сравнению с равновесным. Значит, напряжение с этой полярностью является для данного контакта обратным. В контакте на рис. 3.11, *б* при той же полярности напряжения потенциальный барьер понижается, приконтактный слой обогащается основными носителями — электронами и его сопротивление будет меньше равновесного. Значит, напряжение с этой полярностью является для данного контакта прямым.

Таким образом контакты, показанные на рис. 3.11, обладают выпрямляющими свойствами и могут быть основой диодов. Диоды, использующие барьеры Шоттки, называют *диодами Шоттки.*

Вариантом выпрямляющих контактов является контакт, в котором вблизи границы с металлом образуется инверсионный слой, т.е. слой с противоположным типом проводимости.

Зонная диаграмма контакта, содержащего инверсионный p-слой, показана на рис. 3.12. Этот случай характерен для сильного искривления зон, т.е. для больших контактных разностей потенциалов φ_{MS}, когда вблизи границы уровень электростатического потенциала пересекает уровень Ферми. Толщина инверсионного слоя, как уже отмечалось, не превышает 1–2 нм.

Диоды Шоттки. Важнейшей особенностью диодов Шоттки по сравнению с p–n-переходом является *отсутствие инжекции неосновных носителей*. Эти диоды, как говорят, *работают на основных носителях*. Отсюда следует, что у диодов Шоттки *отсутствует диффузионная емкость*, связанная с накоплением и рассасыванием неосновных носителей в базе. Отсутствие диффузионной емкости существенно повышает быстродействие диодов при изменениях токов и напряжений, в том числе при переключениях с прямого направления на обратное и с обратного на прямое. Время таких переключений определяется только барьерной емкостью и у диодов с малой площадью может составлять десятые и сотые доли наносекунды. Соответствующие рабочие частоты лежат в пределах 3–15 ГГц.

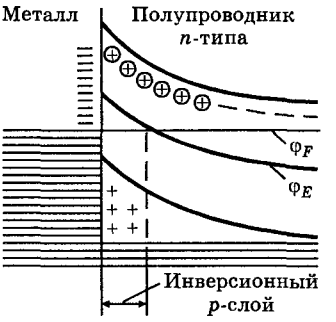


Рис. 3.12. Зонная диаграмма контакта, при котором образуется инверсионный слой

Не менее важной особенностью диодов Шоттки является значительно меньшее прямое напряжение по сравнению с напряжением на p–n-переходе. Это объясняется тем, что ВАХ у диодов Шоттки описывается той же формулой (3.14), что и у p–n-переходов, но тепловой ток существенно больше, поскольку диффузионная скорость D/L, характерная для p–n-перехода [см. (3.16)], у диода Шоттки заменяется на среднюю тепловую скорость носителей v_T. Последняя превышает величину D/L примерно на 3 порядка. В таком же отношении различаются и тепловые токи. Тогда из формулы (3.18) следует, что прямое напряжение у диодов Шоттки будет примерно на 0,2 В меньше, чем p–n-перехода. Типичным для диодов Шоттки являются прямые напряжения 0,4 В. Что касается обратных токов, то они могут составлять, в зависимости от площади, до 10^{-11}–10^{-12} А, т.е. близки к реальным обратным токам кремниевых p–n-переходов, определяемым термогенерацией.

Еще одна особенность диодов Шоттки состоит в том, что их прямая ВАХ строго подчиняется экспоненциальному закону (3.14) в очень широком диапазоне токов — на протяжении нескольких декад, например, от 10^{-12} до 10^{-4} А. Отсюда следует

возможность использования диодов Шоттки в качестве прецизионных логарифмирующих элементов в соответствии с соотношением (3.18).

Качественные барьеры Шоттки образуются в кремнии при контакте с такими металлами, как молибден, нихром, золото, платина (точнее, сплав платины с кремнием — силицид платины), а также алюминий — основной материал для металлизации в ИС.

Невыпрямляющие коитакты. Пусть для контакта металла с полупроводником p-типа имеет место неравенство $\varphi_{MS} > 0$, а для контакта с полупроводником n-типа — неравенство $\varphi_{MS} < 0$ (рис. 3.13). Как уже известно, в первом случае электроны будут переходить из полупроводника в металл и зоны искривятся «вверх», а во втором случае электроны будут переходить из металла в полупроводник и зоны искривятся «вниз».

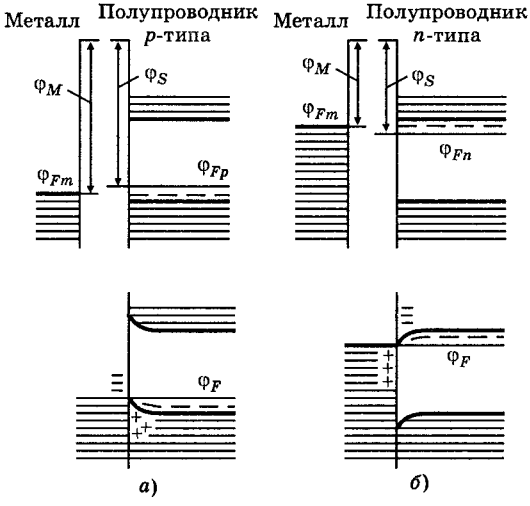


Рис. 3.13. Зонные диаграммы невыпрямляющих контактов металла с полупроводником: *а* — контакт с полупроводником p-типа; *б* — контакт с полупроводником n-типа

В таких контактах вблизи границы в полупроводнике накапливаются основные носители, т.е. получаются обогащенные слои. Их протяженность составляет сотые доли микрометра. Как видно из рис. 3.13, мы приняли очень небольшие контактные разности потенциалов, поэтому искривление зон незначи-

тельное и полупроводники остаются невырожденными. Если принять значения $\varphi_{MS} = 0{,}1–0{,}2$ В и более, то искривление зон будет значительно сильнее и вблизи границы уровень Ферми будет проходить через соответствующую разрешенную зону. На этом участке полупроводник превращается в полуметалл с ничтожным удельным сопротивлением.

Наличие обогащенного слоя означает, что сопротивление системы в целом определяется нейтральным слоем полупроводника и, следовательно, не зависит ни от величины, ни от полярности приложенного напряжения. Такие невыпрямляющие комбинации металла с полупроводником называют *омическими контактами*.

Омические контакты осуществляются в местах присоединения выводов к полупроводниковым слоям. Получение омических контактов — задача не менее важная, чем получение выпрямляющих контактов. Помимо двусторонней проводимости, важным свойством омического контакта является ничтожное время жизни избыточных носителей в обогащенном слое. Поэтому при анализе полупроводниковых приборов обычно считают, что концентрации избыточных носителей на омическом контакте равны нулю.

В микроэлектронике в настоящее время наиболее распространенным металлом для омических контактов является алюминий.

3.4. Граница полупроводник–диэлектрик

Приповерхностный слой полупроводника — это особая область. Имеет кристаллическую структуру, содержит адсорбированные примеси, характерен наличием особых энергетических уровней и т.п. Соответственно, приповерхностному слою свойственны свои значения подвижности, времени жизни и других электрофизических параметров.

Общая характеристика границы $Si–SiO_2$. Свойства среды, с которой граничит полупроводник, оказывают влияние на свойства приповерхностного слоя. Примером могут служить границы (контакты) полупроводников с металлами, рассмотренные в предыдущем разделе. Как было показано, наличие металла на поверхности полупроводника приводит к образованию в последнем обедненных или обогащенных слоев. Аналогичные

процессы имеют место на границе полупроводника с диэлектриком.

Особый интерес представляет граница кремния с двуокисью кремния, поскольку поверхность всех современных полупроводниковых ИС защищается окисным слоем. Кроме того, в структурах МДП (рис. 2.13), выполненных на основе кремния, в качестве диэлектрика тоже, как правило, используется слой SiO_2. Поэтому ниже, говоря о границе полупроводник–диэлектрик, будем подразумевать структуру $Si–SiO_2$.

Главная особенность слоев (пленок) двуокиси кремния, используемых в ИС, состоит в том, что они *всегда содержат примеси донорного типа*. Наиболее распространенными примесями такого типа являются натрий, калий и водород. Все они содержатся в типовых растворах, которыми обрабатывают поверхности кремния и его окисла.

Донорные примеси, свойственные пленке SiO_2, сосредоточены вблизи границы с кремнием. Поэтому в пленке SiO_2 на границе с кремнием образуется тонкий слой положительно заряженных донорных атомов, а отданные ими электроны переходят в приповерхностный слой кремния. Последствия такого перехода зависят как от типа проводимости полупроводника, так и от концентрации донорных примесей в диэлектрике. Поскольку донорные атомы сосредоточены в очень тонком слое диэлектрика, объемная концентрация ($см^{-3}$) оказывается неудобным параметром и вместо нее используют поверхностную концентрацию ($см^{-2}$). Характерными значениями поверхностной концентрации доноров в двуокиси кремния являются $N_{д\,SiO_2} = (0{,}5–2{,}0) \cdot 10^{-12}$ $см^{-2}$.

Если кремний имеет проводимость n-типа, то электроны, перешедшие в него из окисла, обогащают его приповерхностный слой основными носителями: образуется так называемый *n-канал* (рис. 3.14, *а*). Если же кремний имеет проводимость p-типа, то электроны, перешедшие в него из окисла, либо обедняют приповерхностный слой, «обнажая» отрицательные ионы акцепторов (рис. 3.14, *б*), либо образуют наряду с обедненным слоем тонкий инверсионный n-слой (рис. 3.14, *в*).

Возникновение незапланированных каналов под слоем SiO_2 в приборах, работающих как с использованием $p–n$-переходов, так и МДП, может нарушить структуру этих приборов и нормальную работу элементов ИС.

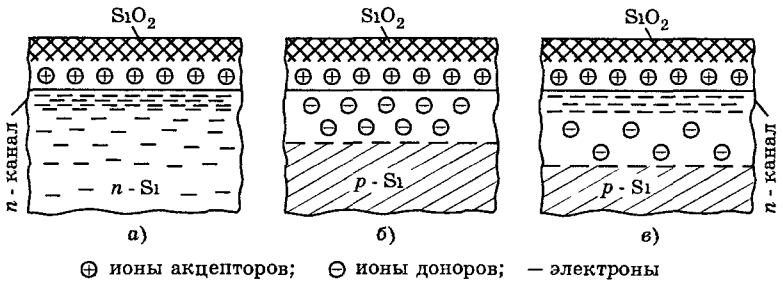


Рис. 3.14. Приповерхностная структура кремния на границе с окислом: *а* — обогащенный слой, *б* — обедненный слой, *в* — обедненный слой с инверсионным каналом

Контрольные вопросы

1. Что такое металлургическая граница?
2. Дайте классификацию переходов.
3. Поясните роль обедненного слоя.
4. Чем определяется величина потенциального барьера перехода?
5. В чем разница между несимметричным и плавным переходом?
6. Что такое прямое напряжение и как изменяется величина потенциального барьера в результате его воздействия?
7. Поясните влияние обратного напряжения на величину потенциального барьера.
8. Какова зависимость ширины перехода от полярности и величины приложенного напряжения?
9. Поясните структуру токов в n–p-переходе.
10. Нарисуйте ВАХ идеализированного p–n-перехода.
11. В чем отличие нормального режима работы p–n-перехода от микрорежима?
12. Что называют напряжением открытого перехода U^* и каковы его величины при нормальном и микрорежиме?
13. Какова величина диффузионного сопротивления p–n-перехода?
14. Что такое ток термогенерации?
15. Охарактеризуйте виды пробоя p–n-перехода.
16. В чем отличие диффузионной емкости от барьерной?

17. Нарисуйте зонную диаграмму выпрямляющего контакта металла с полупроводником n-типа.
18. Нарисуйте зонную диаграмму выпрямляющего контакта металла с полупроводником p-типа.
19. Что такое барьер Шоттки?
20. Каковы основные качества диода Шоттки?
21. В чем преимущества диода Шоттки по сравнению с обычным p–n-переходом?
22. В чем заключается влияние границы полупроводник–диэлектрик на работу МДП–структуры?

Глава 4

Униполярные транзисторы

4.1. Введение

Транзисторами называют полупроводниковые усилительные приборы, т.е. приборы, которые способны усиливать электрическую мощность[1]. Транзисторы имеют очень много конструктивно-технологических разновидностей, но по принципу действия их делят на два основных класса: *униполярные* и *биполярные*.

Работа униполярных транзисторов основана на использовании только одного типа носителей — основных (или электронов, или дырок). Процессы инжекции и диффузии в таких транзисторах практически отсутствуют, во всяком случае они не играют принципиальной роли. Основным способом движения носителей является дрейф в электрическом поле.

Для того чтобы управлять током в полупроводнике при постоянном электрическом поле, нужно менять либо удельную проводимость полупроводникового слоя, либо его площадь. На практике используется и тот, и другой способ, причем в основе обоих способов лежит эффект поля. Поэтому униполярные транзисторы обычно называют *полевыми транзисторами*. Проводящий слой, по которому проходит рабочий ток, называют *каналом*. Отсюда еще одно название такого класса транзисторов — *канальные транзисторы*.

Каналы могут быть приповерхностными и объемными. Приповерхностные каналы представляют собой либо обогащенные слои, обусловленные наличием донорных примесей в диэлектрике, либо инверсионные слои, образующиеся под действием внешнего поля. Объемные же каналы представляют собой участки однородного полупроводника, отделенные от поверхности обедненным слоем. Оба варианта каналов, а также способы их использования показаны на рис. 4.1.

---

1 Именно усиление мощности (а не порознь напряжения и тока) есть тот критерий, по которому прибор можно относить к усилительным. Например, трансформатор усиливает напряжение (за счет тока) или ток (за счет напряжения), но не усиливает мощность и поэтому не относится к усилительным приборам.

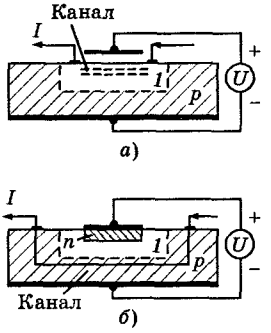


Рис 4.1. Принцип использования каналов в униполярных транзисторах:
а — приповерхностный *n*-канал;
б — объемный *p*-канал;
1 — обедненный слой

Транзисторы с приповерхностным каналом (рис. 4.1, *а*) имеют классическую структуру металл—диэлектрик—полупроводник (МДП). Их называют *МДП-транзисторами*. В частном случае, если диэлектриком является окисел (двуокись кремния SiO_2), используется название *МОП-транзисторы*.

Транзисторы с объемным каналом (рис. 4.1, *б*) характерны тем, что обедненный слой создается с помощью *p–n*-перехода. Поэтому их часто называют *полевыми транзисторами с p–n-переходом*. Однако это название длинное и неудобное. Поэтому ниже мы будем называть их просто *полевыми транзисторами*.

Несмотря на различие в структуре, полевые и МДП-транзисторы имеют много общего. Обоим транзисторам свойственна явно выраженная управляющая цепь (с источником напряжения U), четко отделенная от управляемой цепи, в которой протекает рабочий ток I. Управляющая цепь практически не потребляет тока, поскольку в нее входит либо участок диэлектрика (рис. 4.1, *а*), либо обратносмещенный *n*-переход (рис. 4.1, *б*). Направление электрического поля, создаваемого управляющим напряжением, перпендикулярно направлению тока. Наряду с общими чертами каждый из этих двух транзисторов обладает, конечно, и рядом отличительных особенностей.

4.2. МДП-транзисторы

Реальная структура МДП-транзистора с *n*-каналом, выполненного на основе полупроводника, показана на рис. 4.2. Металлический электрод, создающий эффект поля, называют *затвором* (З). Два других электрода называют *истоком* (И) и *стоком* (С). Эти электроды в принципе обратимы. Стоком является тот из них, на который (при соответствующей полярности напряжения) поступают рабочие носители канала. Если канал *n*-типа, то рабочие носители — электроны и полярность стока положительная. Исток обычно соединяют с основной пластиной полупроводника, которую называют *подложкой* (П).

Принцип действия. В идеальном случае, когда равновесный потенциал поверхности равен нулю ($\varphi_{S0}=0$), МДП-транзистор с n-каналом работает следующим образом. Пусть затвор соединен с истоком, т.е. $U_{зи}=0$. При этом канал отсутствует и на пути между стоком и истоком оказываются два встречновключенных p–n^+-перехода. Поэтому при подаче напряжения $U_{си}$ ток в цепи стока будет ничтожно мал.

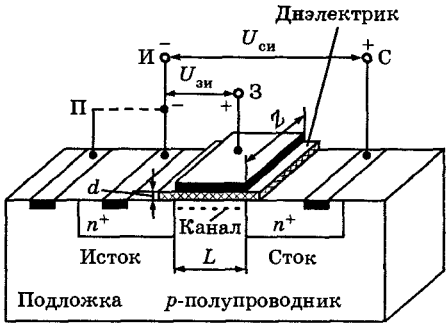


Рис. 4.2. Структура МДП-транзистора с индуцированным n-каналом

Если подать на затвор отрицательное напряжение, то приповерхностный слой обогатится дырками; при этом ток в рабочей цепи мало изменится. Если же подавать на затвор все большее положительное смещение $U_{зи} > 0$, то сначала образуется обедненный слой (объемный заряд акцепторов), а затем — инверсионный слой электронов, т.е. проводящий канал. После этого ток стока принимает конечное значение и зависит от напряжения на затворе. Это и есть рабочий режим МДП-транзистора. Поскольку входной ток (в цепи затвора) ничтожно мал, получается значительное усиление мощности.

Каналы, отсутствующие в равновесном состоянии и образующиеся под действием внешнего напряжения, называют *индуцированными*. Толщина индуцированного канала практически неизменная (1–2 нм), поэтому модуляция его проводимости обусловлена изменениями концентрации носителей. Напряжение на затворе, при котором образуется канал, называют *пороговым напряжением* и обозначают U_0. Длина канала L равна расстоянию между слоями истока и стока, а ширина Z — протяженности этих слоев (рис. 4.2).

Если выбрать подложку n-типа, а слои истока и стока сделать p^+-типа, то получится МДП-транзистор с индуцированным p-каналом. Он характерен обратными полярностями порогового и рабочих напряжений: $U_0 < 0$, $U_{зи} < 0$, $U_{си} < 0$.

Электронные схемы, в которых используется сочетание транзисторов с n- и p-каналами, называют *комплементарными схемами*.

Подложку МДП-транзисторов стараются делать из материала с высоким удельным сопротивлением, с тем чтобы облегчить образование канала и увеличить пробивное напряжение переходов истока и стока.

В принципе механизм работы и свойства МДП-транзисторов одинаковы. Однако есть и некоторые различия. Во-первых, n-канальные транзисторы более быстродействующие, так как подвижность их рабочих носителей — электронов — примерно в три раза выше, чем дырок. Во-вторых, у n- и p-канальных транзисторов структура приповерхностного слоя в равновесном состоянии оказывается разной, и это отражается на величине порогового напряжения.

Различие в структуре приповерхностного слоя объясняется разным влиянием положительного заряда, который обычно имеется в окисле. В подложке n-типа этот заряд создает обогащенный слой, который препятствует образованию p-канала, соответственно пороговое напряжение у p-канальных транзисторов увеличивается. В подложке p-типа тот же заряд создает обедненный слой, т.е. способствует образованию n-канала; соответственно пороговое напряжение у n-канальных транзисторов уменьшается.

Иногда положительный заряд в окисле может образовать не только обедненный, но и инверсионный слой, т.е. n-канал. Поскольку такой канал существует при нулевом напряжении на затворе, его уже нельзя считать индуцированным (т.е. наведенным полем затвора). Значит, величина порогового напряжения теряет свой обычный смысл. В транзисторах этого типа канал называют *встроенным* (т.е. имеющимся «заранее»), а вместо порогового напряжения вводят параметр *напряжение отсечки*. Это — напряжение, при котором электроны равновесного инверсионного слоя отталкиваются от поверхности и встроенный канал исчезает[1]. Вообще говоря, наличие встроенного канала не является препятствием для использования МДП-транзисторов. Такие транзисторы работают при обеих полярностях

---

1 Для того, чтобы предотвратить образование равновесного канала, при изготовлении n-канальных МДП-транзисторов приходится использовать специальные меры по обработке поверхности кремния и диэлектрика, а также использовать подложку с повышенной концентрацией акцепторов. Все это делает технологию n-канальных транзисторов несколько сложнее, чем p-канальных.

напряжения затвора: при отрицательной полярности канал обедняется носителями и ток стока уменьшается, при положительной полярности канал обогащается и ток увеличивается. Однако транзисторы с индуцированным каналом имеют гораздо большее распространение, хотя они работают только при одной полярности напряжения на затворе — той, при которой возникает канал. В тех сравнительно редких случаях, когда встроенный канал желателен, его обычно специально осуществляют в виде тонкого приповерхностного слоя с помощью ионного легирования (рис. 4.3).

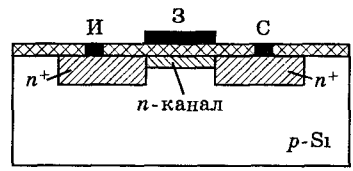


Рис. 4.3. Структура МДП-транзистора со встроенным n-каналом

В дальнейшем рассматриваются только транзисторы с индуцированным n-каналом.

Пороговое напряжение. Напряжение затвора наводит в полупроводнике тем больший удельный заряд (на единицу площади), чем больше удельная емкость между металлом и поверхностью полупроводника. Значит, *удельная емкость затвор—канал определяет управляющую способность затвора* и поэтому является одним из важных параметров МДП-транзистора. Эта емкость имеет вид

$$C_0 = \varepsilon_0 \varepsilon_д / d, \tag{4.1}$$

где d — толщина диэлектрика; $\varepsilon_д$ — его диэлектрическая проницаемость. Уменьшение величины d желательно, но ограничено пробоем диэлектрика. Типичные значения толщины двуокиси кремния составляют d = 0,01–0,1 мкм. Если положить d = 0,03 мкм, и $\varepsilon_д$ = 3,9 (для SiO_2), то $C_0 \approx 1000$ пФ/мм^2.

Пороговое напряжение U_0 можно разделить на две составляющие (рис. 4.4)[1].

$$U_0 = U_{0F} + U_{0B}. \tag{4.2}$$

Составляющая U_{0F} — это *напряжение спрямления зон*; оно сводит к нулю равновесный поверхностный потенциал φ_{s0}, т.е.

---

1 На зонной диаграмме положительные значения электрических потенциалов откладываются вниз.

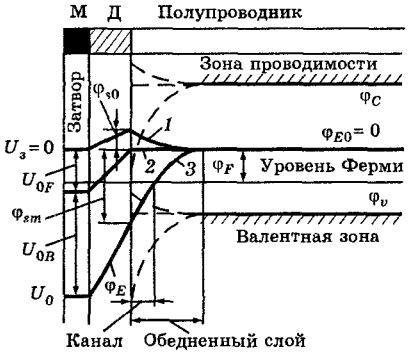


Рис. 4.4. Зонные диаграммы МДП-транзистора при напряжениях от 0 до U_0

ликвидирует начальное искривление зон (ср. кривые *1* и *2*). На рис. 4.4 начальное искривление принято противоположным тому, которое необходимо для образования канала.

Составляющая U_{0B} — это *напряжение изгиба зон*: оно обеспечивает изгиб зон в сторону, необходимую для образования канала (кривая *3*) и создает поверхностный потенциал φ_{sm}, при котором уровень электростатического потенциала пересекает уровень Ферми.

Таким образом, напряжение U_{0F} характеризует «подготовленность» полупроводника к образованию канала; если $\varphi_{s0}=0$, то и $U_{0F}=0$, а если равновесные зоны искривлены вниз, то $U_{0F}<0$. Что касается напряжения U_{0B}, то оно определяет значение порогового напряжения в «идеальных» условиях, когда поверхностный потенциал равен нулю.

Напряжение U_{0F} выражается следующим образом:

$$U_{0F}=\varphi_{MS}+Q_{0s}/C_0. \qquad (4.3а)$$

где Q_{0s} — равновесный удельный заряд поверхности, включающий в себя заряд поверхностных состояний и заряд, обусловленный ионами примесей в диэлектрике; φ_{MS} — контактная разность потенциалов между металлом и диэлектриком. Величина Q_{0s} получается экспериментально и составляет обычно $5\cdot10^{-9}-5\cdot10^{-8}$ Кл/см^2.

Напряжение U_{0B} выражается следующим образом:

$$U_{0B}=\varphi_{sm}+\frac{a}{C_0}\sqrt{\varphi_{sm}}, \qquad (4.3б)$$

где

$$a=\sqrt{2q\varepsilon_0\varepsilon_п N} \qquad (4.4)$$

— коэффициент, характеризующий влияние объемного заряда в подложке ($\varepsilon_п$ — диэлектрическая проницаемость полупроводника; N — концентрация примеси).

Обычно полагают $\varphi_{sm} = 2\varphi_F$, где φ_F — модуль разности между уровнем Ферми и уровнем электростатического потенциала в объеме полупроводника (рис. 4.4). Например, если $N = 10^{16}$ см$^{-3}$, то $\varphi_F \approx 0{,}3$ В и, значит, $\varphi_{sm} = 0{,}6$ В; согласно (4.4) $a \approx 5 \cdot 10^{-8}$ Ф $\cdot$ В$^{1/2}\cdot$см^2. Полагая $C_0 = 10^{-7}$ Ф/см^2 (см. выше), из (4.3б) получаем: $U_{0B} \approx 1{,}0$ В. Практически значения полного порогового напряжения лежат в пределах $U_0 = 0{,}5$–$1{,}5$ В.

Статические характеристики. Рассмотрим влияние тока на структуру канала. Если напряжение $U_{си} = 0$, то поверхность полупроводника эквипотенциальна, поле в диэлектрике однородное и толщина образовавшегося канала h одинакова на всем протяжении (рис. 4.5, *а*). Если же $U_{си} > 0$, то протекает ток и потенциал поверхности возрастает от истока к стоку. Значит, разность потенциалов между затвором и поверхностью в направлении стока уменьшается. Соответственно уменьшаются напряженность поля в диэлектрике и удельный заряд электронов в канале. Поэтому сечение канала вблизи точки $x = L$ сужается (рис. 4.5, *б*).

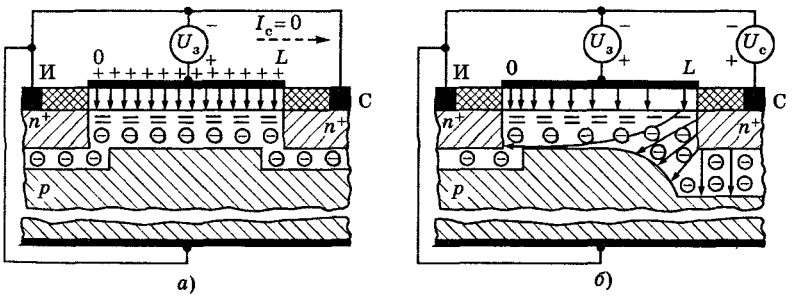


Рис 4.5. Распределение поля и зарядов в МДП-транзисторе при нулевом (*а*) и наибольшем положительном (*б*) напряжениях на стоке

При некотором критическом напряжении на стоке, которое называют *напряжением насыщения*, разность потенциалов между затвором и поверхностью в точке $x = L$ делается равной пороговому напряжению. Образуется так называемая «горловина» канала (рис. 4.6, *а*).

Напряжение насыщения имеет вид

$$U_{сн} = U_{зи} - U_0. \qquad (4.5)$$

При напряжениях $U_{си} > U_{сн}$ слой объемного заряда, который до сих пор отделялся от поверхности каналом, «выходит на поверхность» на участке ΔL, а «горловина» канала соответственно сдвигается в точку L' (рис. 4.6, *б*)[1]. Вследствие этого происходит укорочение канала на величину ΔL; потенциал «горловины» в точке L' сохраняет значение $U_{сн}$, которое было в начале насыщения.

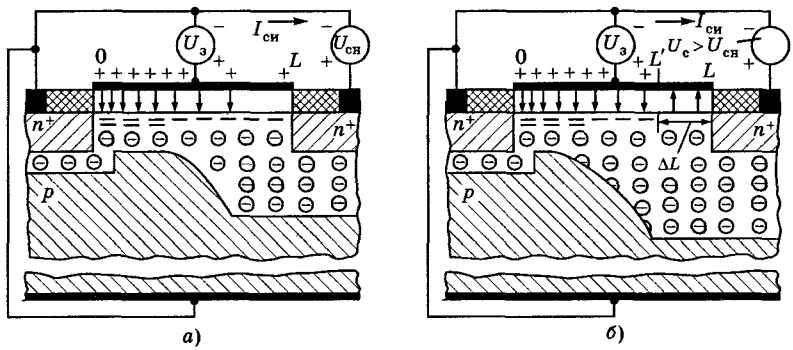


Рис 4.6. Распределение поля и зарядов в МДП-транзисторе: *а* — на границе насыщения ($U_{си} = U_{сн}$); *б* — в области насыщения ($U_{си} > U_{сн}$)

Величина ΔL (приповерхностная ширина объемного заряда) зависит от разности напряжений на этом участке $U_{си} - U_{сн}$. Зависимость эта такая же, как зависимость ширины p–n-перехода от обратного напряжения: $\Delta L \sim \sqrt{U_{си} - U_{сн}}$.

После образования «горловины» канала ток в рабочей цепи практически перестает зависеть от напряжения на стоке — наступает насыщение тока (рис. 4.7, *а*), откуда и название напряжения $U_{сн}$.

Анализ, выполненный с учетом описанных процессов, приводит к выражению для ВАХ, которое неудобно для инженерных расчетов (из-за наличия членов в степени 3/2). Поэтому на практике пользуются аппроксимациями ВАХ, из которых наиболее простой и распространенной является следующая:

$$I_с = b[(U_{зи} - U_0)U_{си} - \tfrac{1}{2}U_{си}^2]. \qquad (4.6)$$

1 Процессы образования «горловины» канала и ее сдвига просматриваются гораздо яснее в полевых транзисторах, где канал несравненно толще (рис. 4.13)

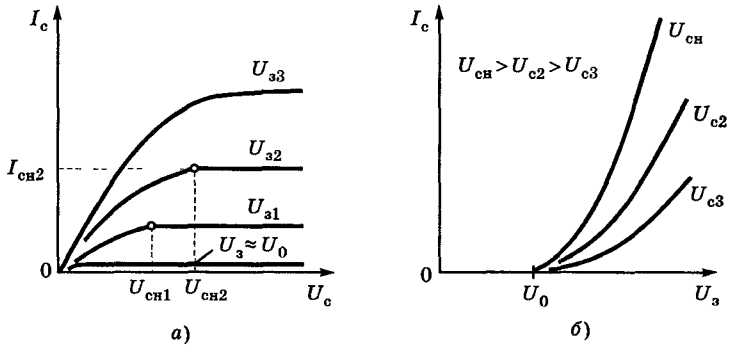


Рис. 4.7. Статические характеристики МДП-транзистора:
а — выходные; *б* — передаточные

Здесь b — *удельная крутизна* МДП-транзистора (один из его основных параметров):

$$b = \mu C_0 \frac{Z}{L} = \frac{\varepsilon_0 \varepsilon_д}{d} \frac{Z}{L}, \qquad (4.7)$$

где μ — приповерхностная подвижность носителей (она обычно в 2-3 раза меньше объемной); Z — ширина канала. При значениях $\mu = 550$ см^2/ (В · с), $Z/L = 10$ и $C_0 = 10^{-7}$ Ф/см^2 получаем типичное значение $b \approx 0{,}5$ мА/В^2.

Выражение (4.6) действительно только при условии $U_{си} < U_{сн}$, т.е., как говорят, на начальных — *крутых* участках ВАХ (см. рис. 4.7,*а*). Если же $U_{си} > U_{сн}$, то ток не меняется и остается равным тому значению, которое он имеет при $U_{си} = U_{сн}$. Поэтому, подставляя (4.5) в (4.6), получаем выражение для области насыщения, т.е. для *пологих* участков ВАХ:

$$I_с = \tfrac{1}{2} b (U_{зи} - U_0)^2. \qquad (4.8)$$

Этому выражению соответствует на рис 4.7,*б* кривая с параметром $U_{сн}$.

Обычно *номинальным током* МДП-транзистора считается ток при напряжении $U_{зи} = 2U_0$, т.е.

$$I_{с\,ном} = \tfrac{1}{2} b U_0^2. \qquad (4.9)$$

Как видим, чем меньше пороговое напряжение, тем меньше рабочий ток. Номинальному режиму, т.е. значению $U_{зи} = 2U_0$, согласно (4.5) соответствует напряжение насыщения $U_{сн} = U_0$. Следовательно, малые значения U_0 обеспечивают и малые токи, и малые рабочие напряжения транзистора.

Выражения (4.6) и (4.8) имеют широкое распространение благодаря своей простоте и наглядности. Однако они дают значительную погрешность при расчетах, если концентрация примеси в подложке превышает 10^{15} см$^{-3}$, что обычно имеет место. Поэтому в тех случаях, когда это необходимо, вместо (4.6) пользуются более точной аппроксимацией:

$$I_c = b[(U_{зи} - U_0)U_{си} - \tfrac{1}{2}(1 + \eta)U_{си}^2], \tag{4.10}$$

где поправочный коэффициент η имеет вид

$$\eta = \frac{1}{3}\frac{a/C_0}{\sqrt{\varphi_{sm}}}. \tag{4.11}$$

Например, если $a/C_0 = 0{,}5$ В$^{1/2}$ (как было выше) и $\varphi_{sm} = 0{,}6$ В, то $\eta \approx 0{,}22$.

Дифференцируя (4.10) по $U_{си}$ и полагая $dI_c/dU_{си} = 0$, находим напряжение насыщения:

$$U_{сн} = \frac{1}{1 + \eta}(U_{зи} - U_0). \tag{4.12}$$

Оно оказывается меньше, чем при расчете по (4.5). Подставляя (4.12) в (4.10), получаем уточненную ВАХ для пологой области — области насыщения

$$I_c = \frac{1}{2}\frac{b}{1 + \eta}(U_{зи} - U_0)^2. \tag{4.13}$$

До сих пор считалось, что исток соединен с подложкой. Бывают случаи, когда подложка имеет отрицательный потенциал $U_{пи}$ относительно истока (например, в интегральных схемах, у которых подложка общая для всех транзисторов)[1]. Тогда напряжение, падающее на слое объемного заряда, увеличивается,

---

1 Положительное напряжение на подложке (в случае n-канального транзистора) недопустимо, так как при этом p–n-переход истока будет работать в прямом включении и будет иметь место инжекция электронов в подложку, т.е. нарушится принцип работы униполярных транзисторов.

и это приводит к поправке в выражении (4.3б) для напряжения изгиба зон:

$$U_{0B} = \varphi_{sm} + \frac{a}{C_0}\sqrt{\varphi_{sm} + |U_{пи}|}\,. \qquad (4.14)$$

При этом напряжение $U_{пи}$, естественно, войдет и в выражение (4.8). Соответственно ток I_c, вообще говоря, будет функцией двух напряжений: $U_{зи}$ и $U_{пи}$, т.е. возможно *двойное управление током*.

С учетом влияния подложки характеристику (4.13) можно заменить следующей:

$$I_c = \frac{1}{2}\frac{b}{1+\eta}(U_{зи} - U_0 - \frac{2}{3}\eta|U_{пи}|)^2. \qquad (4.15)$$

Как видим, наличие напряжения между подложкой и истоком равносильно увеличению порогового напряжения.

В заключение рассмотрим начальные крутые участки ВАХ, которые широко используются в ключевых (импульсных) схемах. Полагая

$$U_{си} << U_{зи} - U_0,$$

можно пренебречь квадратичным членом в выражении (4.6) и получить линейную зависимость

$$I_c = b(U_{зи} - U_0)U_{си}. \qquad (4.16)$$

Соответствующее семейство ВАХ показано на рис. 4.8. Коэффициент при $U_{си}$ в правой части (4.16) называется проводимостью канала, а обратная величина — *сопротивлением канала*:

$$R_0 = \frac{1}{b(U_{зи} - U_0)}. \qquad (4.17)$$

Как видим, сопротивление канала можно регулировать в широких пределах, меняя напряжение на затворе. Такая возможность используется

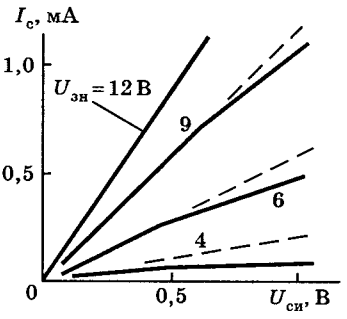


Рис. 4.8. Начальные квазилинейные участки выходных характеристик МДП-транзистора

на практике. Если положить $U_{зи} - U_0 = 4$ В и $b = 0{,}5$ мА/В^2, то $R_0 = 0{,}5$ кОм.

Малосигнальные параметры. В усилительной технике используются пологие участки ВАХ — область насыщения. Этой области свойственны наименьшие нелинейные искажения сигналов и оптимальные значения малосигнальных параметров, существенных для усиления.

Малосигнальными параметрами МДП-транзистора являются:

крутизна $S = \left.\dfrac{dI_c}{dU_{зи}}\right|_{U_{си} = \text{const}}$;

внутреннее сопротивление $r_c = \left.\dfrac{dU_{си}}{dI_c}\right|_{U_{зи} = \text{const}}$;

коэффициент усиления[1] $k = \left.\dfrac{dU_{си}}{dU_{зи}}\right|_{I_c = \text{const}}$.

Эти три параметра связаны соотношением

$$k = Sr_0. \tag{4.18}$$

Крутизна в области насыщения легко определяется из выражения (4.8):

$$S = b(U_{зи} - U_0). \tag{4.19а}$$

Как видим, крутизна пропорциональна параметру b. Название последнего (удельная крутизна) обусловлено тем, что при $U_{зи} - U_0 = 1$ В величина b численно равна крутизне. С помощью выражений (4.19а) и (4.8) легко установить связь крутизны с рабочим током:

$$S = \sqrt{2bI_c}. \tag{4.19б}$$

Например, при $b = 0{,}5$ мА/В^2 и $I_c = 1$ мА получаем: $S = 1{,}0$ мА/В.

Если использовать более точную формулу (4.13), то крутизна будет меньше, чем при расчете по формулам (4.19), поскольку величина b заменяется на $b/(\eta + 1)$.

Внутреннее сопротивление на пологом участке ВАХ обусловлено зависимостью длины канала от стокового напряжения

---

1 Обычно коэффициент усиления электронных приборов обозначают через μ, но это обозначение в теории МДП-транзисторов совпадает с обозначением подвижности, и мы его избегаем.

(рис. 4.6, *б*). Рост напряжения $U_{си}$ сопровождается увеличением ширины стокового перехода ΔL и соответственно уменьшением длины канала L'. При этом возрастает удельная крутизна b, а вместе с нею и ток стока $I_с$ Внутреннее сопротивление равно

$$r_c = \left(L\sqrt{\frac{2qN}{\varepsilon_0 \varepsilon_п}} \right) \frac{\sqrt{U_c}}{I_c}. \qquad (4.20)$$

Полагая $N = 10^{16}$ см$^{-3}$, $U_с = 4$ В, $I_с = 1$ мА, $L = 10$ мкм, получаем $r_с = 100$ кОм.

Перемножив (4.20) и (4.19б), получим коэффициент усиления k. Он не зависит от длины канала; его типичные значения составляют 50–200 (в зависимости от ширины канала Z).

Выше было отмечено, что МДП-транзистор может управляться не только напряжением затвора, но и напряжением подложки. Дифференцируя (4.15) по $|U_{пи}|$, получаем *крутизну по подложке*:

$$S_п = -\frac{2}{3}\frac{\eta}{1+\eta} b \left(U_{зи} - U_0 - \frac{2}{3}|U_{пи}| \right). \qquad (4.21)$$

Знак минус говорит о том, что ток I_c уменьшается с увеличением напряжения $|U_{пи}|$. Дифференцируя (4.15) по $U_{зи}$, получаем *крутизну по затвору*:

$$S_з = \frac{b}{1+\eta} \left(U_{зи} - U_0 - \frac{2}{3}\eta|U_{пи}| \right). \qquad (4.22)$$

Как видим, наличие напряжения $|U_{пи}|$ приводит к уменьшению крутизны $S_з$.

Отношение крутизны $S_п$ и $S_з$ находится в прямой зависимости от коэффициента η, т.е. в конечном счете определяется толщиной диэлектрика и концентрацией примеси в подложке. Обычно $|S_п| < S_з$.

В любом случае предпочтительно управление по затвору, потому что при этом входное сопротивление, определяемое диэлектриком, несравненно больше (при управлении по подложке входное сопротивление определяется обратным током истокового p–n-перехода).

В заключение отметим, что рассматривавшееся во всех предыдущих разделах включение МДП-транзистора с общим истоком ОИ (рис 4.9, *а*) — наиболее распространенное, но не единственно возможное. Иногда используется включение с общим затвором ОЗ (рис. 4.9, *б*). Оно характерно весьма низким входным сопротивлением (близким к величине $1/S$) и потому находит применение только в некоторых специальных схемах.

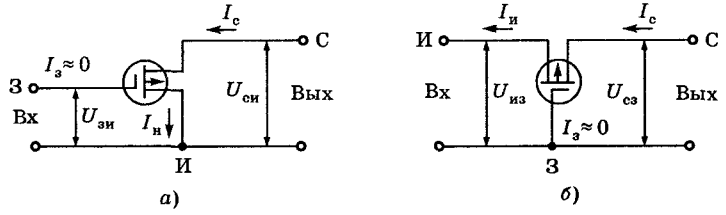


Рис. 4.9. Включение МДП-транзистора с общим истоком (*а*) и с общим затвором (*б*)

Стабильность параметров. При заданных напряжениях на затворе и стоке ток стока зависит от температуры. Эта зависимость проявляется через параметры b и U_0. Функция $b(T)$ обусловлена температурной зависимостью подвижности носителей, а функция $U_0(T)$ — температурной зависимостью уровня Ферми [см. (4.3б), где $\varphi_{sm} = 2\varphi_F$].

С ростом температуры и удельная крутизна, и пороговое напряжение уменьшаются, причем уменьшение этих параметров влияет на ток в противоположных направлениях, см. (4.8) и (4.13). Существует такое значение тока I_c, при котором влияния зависимостей $b(T)$ и $U_0(T)$ уравновешиваются. Это стабильное значение называют *критическим током*. Наличие критического тока — важная отличительная черта МДП-транзисторов, она обеспечивает возможность температурной стабилизации схем простейшим путем — выбором рабочего тока.

Из условия $dI_c/dT = 0$ (с учетом производных $\partial b/\partial T$ и $\partial U_0/\partial T$) можно получить напряжение на затворе, соответствующее критическому току:

$$U_{зи\,кр} - U_0 = (0{,}8\text{–}2{,}4)\ \text{В} \qquad (4.23)$$

(минимальное значение соответствует концентрации примеси в подложке 10^{18} см$^{-3}$, максимальное — концентрации 10^{15} см$^{-3}$). Обычно критический ток в 5—10 раз меньше номинального, определяемого формулой (4.9).

В диапазоне $I_c > I_{c\,кр}$ (в частности, при номинальном токе) температурный коэффициент тока положительный, а в диапазоне $I_c < I_{c\,кр}$ (микрорежим) — отрицательный. Температурную нестабильность тока принято характеризовать не приращением тока ΔI_c, а эквивалентным приращением $\Delta U_{зи}$, которое получается из очевидного соотношения: $\Delta U_{зи} = \Delta I_c / S$. Для токов, близких к критическому, характерны температурные чувствительности ±0,5 мВ/°С, для «сверхкритических» токов они составляют +(8–10) мВ/°С, а для «субкритических» – (4–6) мВ/°С.

Крутизна МДП-транзистора зависит от температуры через те же параметры b и U_0, что и ток. Поэтому наряду с понятием критического тока существует понятие *критической крутизны*, для которой влияния зависимостей $b(T)$ и $U_0(T)$ уравновешиваются. Критическая крутизна получается при токе, меньшем критического.

Тот факт, что главная рабочая часть МДП-транзистора — канал — граничит непосредственно с инородной средой — диэлектриком, оказывает значительное влияние на стабильность параметров. Главное проявление нестабильности состоит в изменениях порогового напряжения. Эти изменения обусловлены в первую очередь изменениями равновесного поверхностного заряда Q_{0s} [см. (4.3а)].

При протекании тока неизбежно происходит обмен электронами между каналом и ловушками, имеющимися в диэлектрической пленке. Важным следствием такого обмена являются флуктуации тока — одна из главных составляющих собственных шумов транзистора. Эта составляющая относится к категории избыточных шумов, т.е. шумов не неизбежных, обусловленных не дискретной структурой потока носителей, а «привходящими» обстоятельствами, в данном случае — наличием близлежащего диэлектрика. Повышенный уровень собственных шумов — один из недостатков МДП-транзисторов.

Переходные и частотные характеристики. Малосигнальная эквивалентная схема МДП-транзистора показана в общем виде на рис. 4.10, *а*. Поскольку подразумевается работа транзистора на пологих участках ВАХ, в качестве сопротивления канала использована величина r_c. Элементами, отражающими усилительную способность транзистора, являются источники тока $S_з U_{зи}$ и $S_п U_{пи}$. Сопротивления $R_{зи}$ и $R_{зс}$ — это сопротивления диэлектрика затвора: ими обычно пренебрегают, так как они

имеют значения 10^{13}–10^{14} Ом и более. Сопротивления $R_{пи}$ и $R_{пс}$ — это обратные сопротивления p–n-переходов истока и стока; их значения составляют 10^{10}–10^{11} Ом. Емкости $C_{пи}$ и $C_{пс}$ — это барьерные емкости тех же переходов; их значения зависят прежде всего от площадей истока и стока. Если, например, размеры обоих этих электродов составляют 20×40 мкм^2, то при удельной емкости 150 пФ/мм^2 получаем $C_{пи} = C_{пс} = 0{,}12$ пФ. Наконец, емкости $C_{зи}$ и $C_{зс}$ — это емкости металлического электрода затвора относительно слоев истока и стока.

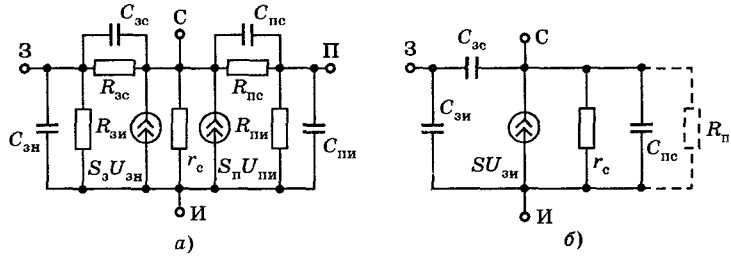


Рис. 4.10. Малосигнальные эквивалентные схемы МДП-транзистора: *а* — полная; *б* — упрощенная при $U_{пи} = 0$

В том наиболее распространенном случае, когда исток соединен с подложкой, источник тока $S_п U_{пи}$ отсутствует, а сопротивление $R_{пи}$ и емкость $C_{пи}$ оказываются закороченными. Если, кроме того, пренебречь сопротивлениями диэлектрика $R_{зи}$ и $R_{зс}$, получаем для данного случая эквивалентную схему, показанную на рис. 4.10, *б* (индекс «з» у крутизны для простоты опущен). Эта схема служит основой большинства практических расчетов.

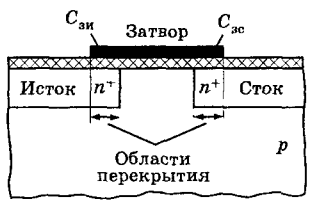


Рис. 4.11. Перекрытие затвора; емкости перекрытия

Происхождение емкостей $C_{зи}$ и $C_{зс}$ показано на рис. 4.11. Они обусловлены так называемым *перекрытием областей истока и стока затвором* (коротко — *перекрытием затвора*). Имеется в виду, что по технологическим причинам часто не удается расположить электрод затвора точно между слоями n^+, как показано на идеализированной структуре (рис. 4.2). Тогда между краями затвора и этими слоями образуются паразитные

емкости $C_{зи}$ и $C_{зс}$. Обычно эти емкости в несколько раз меньше барьерных, но их роль (особенно емкости $C_{зс}$) весьма существенна.

Емкость между затвором и каналом ($C_з$) на рис. 4.10 не показана, так как вносимая ею инерционность отражена комплексным характером крутизны (см. ниже).

Инерционность МДП-транзисторов по отношению к быстрым изменениям управляющего напряжения $U_{зн}$ обусловлена двумя факторами: перезарядом емкости затвора $C_з$ и перезарядом межэлектродных емкостей.

Первый фактор можно пояснить следующим образом. Скачок напряжения $U_{зи}$ вызывает изменение поля в диэлектрике вблизи истока. До тех пор, пока это изменение не распространится до стока, ток $I_с$ остается неизменным. Время распространения определяется скоростью заряда емкости $C_з$ через сопротивление канала.

Второй фактор связан с тем, что если даже ток $I_с$ возрастает скачком, то напряжение $U_с$, а значит, и ток во внешней цепи, будут нарастать плавно — по мере перезаряда межэлектродных емкостей. Скорость этого перезаряда зависит от внешних сопротивлений, т.е. не определяется свойствами собственно транзистора. Однако при прочих равных условиях она тем больше, чем меньше межэлектродные емкости. В этом смысле значения емкостей транзистора являются показателем его быстродействия.

Из сказанного ясно, что относительная роль обоих инерционных факторов в принципе неоднозначна и во многом зависит от схемы. Вместе с тем ясно, что первый фактор (время заряда емкости $C_з$) является лимитирующим: он определяет предельное быстродействие МДП-транзистора в режиме короткого замыкания цепи стока (когда влияние межэлектродных емкостей отсутствует).

Цепь затвора, строго говоря, представляет собой систему с распределенными параметрами. В инженерной практике целесообразно аппроксимировать ее простой RC-цепью в виде емкости затвора $C_з$ и сопротивления канала R_0.

Сопротивление канала выражается формулой (4.17), а емкость затвора легко записать, зная площадь затвора (ZL) и его удельную емкость (4.1):

$$C_з = \frac{\varepsilon_0 \varepsilon_д}{d} ZL. \qquad (4.24)$$

Заряд и разряд RC-цепи описываются простейшей экспоненциальной функцией. Такой же функцией будет описываться крутизна транзистора, поскольку она характеризует изменения тока I_c при заданном скачке напряжения $U_{зи}$. Следовательно, в операторной форме крутизну можно записать следующим образом:

$$S(s) = \frac{S}{1 + s\tau_S}, \tag{4.25}$$

где $\tau_S = C_з R_0$ — *постоянная времени крутизны*. В комплексной форме крутизна будет иметь вид

$$\dot{S} = \frac{S}{1 + j\omega/\omega_S}, \tag{4.26}$$

где $\omega_S = 1/\tau_S$ — граничная частота крутизны. Модуль и фаза выражения (4.26) будут соответственно амплитудно-частотной и фазо-частотной характеристиками крутизны.

Постоянную времени τ_S легко получить, умножая емкость затвора (4.24) на сопротивление канала (4.17). С учетом (4.7) получаем:

$$\tau_S = \frac{L^2}{\mu(U_{зи} - U_0)}. \tag{4.27}$$

Например, если L = 10 мкм, μ = 500 см^2/В · с и $U_{зи} - U_0 = 4$ В, то τ_S = 0,5 нс. Тогда $f_S = (1/2\pi)\omega_S \approx 300$ МГц.

Из выражения (4.27) очевидна предпочтительность n-канала перед p-каналом (бóльшая подвижность μ), а также определяющая роль длины канала. У современных МДП-транзисторов удается делать длину канала менее 1 мкм. При этом $\tau_S < 0{,}01$ нс и $f_S > 15$ ГГц. *Такие значения параметров часто позволяют пренебречь инерционностью крутизны и считать, что инерционность МДП-транзистора обусловлена только межэлектродными и паразитными емкостями.*

4.3. Полевые транзисторы

Идеализированная структура современного полевого транзистора показана на рис. 4.12. Здесь металлический контакт вместе со слоем p^+ играет роль затвора, но затвор отделен от полупроводника n-типа не диэлектриком, как в случае МДП-тран-

зисторов, а обедненным слоем p–n-перехода. На переход задается обратное напряжение.

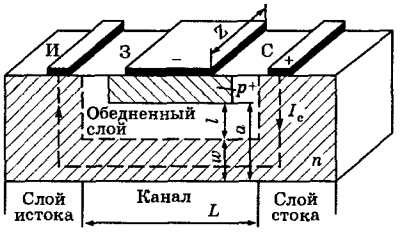


Рис. 4.12. Структура полевого транзистора

Вообще говоря, p^+-слой не обязателен: обедненный слой может иметь место и при непосредственном контакте металла с полупроводником. Транзисторы с такой структурой называют *полевыми транзисторами с барьером Шоттки*. Основные свойства обеих разновидностей одинаковы, поэтому ниже рассматриваются только транзисторы с p–n-переходом, анализ которых более нагляден.

Ниже будет показано, что для нормальной работы полевого транзистора толщина рабочего слоя под затвором (величина a на рис. 4.12) должна составлять не более нескольких микрон. Кристаллы полупроводника такой толщины оказываются непригодными из-за механической хрупкости. Поэтому структуру на рис. 4.12 следует понимать как тонкий рабочий n-слой, расположенный на более толстой «несущей» пластине, которая на рисунке не показана.

Принцип действия. На p–n-переход затвора задается обратное напряжение и глубина обедненного слоя меняется. Чем больше обратное напряжение, тем глубже обедненный слой и тем соответственно меньше толщина канала w. Таким образом, меняя обратное напряжение на затворе, можно менять поперечное сечение, а значит, и сопротивление канала. При наличии напряжения на стоке будет меняться ток канала, т.е. выходной ток транзистора.

Усиление мощности обеспечивается малой величиной входного тока. У полевых транзисторов входным током является обратный ток p–n-перехода затвора. Для кремниевых p–n-переходов небольшой площади обратный ток составляет до 10^{-11} А и менее.

Определим зависимость толщины и сопротивления канала от управляющего напряжения на затворе при нулевом напряжении на стоке. Толщину канала согласно рис. 4.12 можно записать следующим образом:

$$w = a - l,$$

где a — расстояние от «дна» n-слоя до металлургической границы перехода. Пренебрегая равновесной высотой потенциального барьера, получаем зависимость толщины канала от напряжения на затворе:

$$w = a - \sqrt{\frac{2\varepsilon_0 \varepsilon U_{зи}}{qN}}. \qquad (4.28)$$

Под $U_{зи}$ здесь и ниже понимается модуль напряжения на затворе.

Из условия $w = 0$ легко найти *напряжение отсечки*, при котором обедненный слой перекрывает весь канал и ток в канале прекращается:

$$U_{зо} = (qN/2\varepsilon_0\varepsilon)\, a^2. \qquad (4.29)$$

Например, если $N = 5 \cdot 10^{15}$ см$^{-3}$ и $a = 2$ мкм, то $U_{зо} = 12{,}5$ В. С учетом высоты равновесного барьера напряжение отсечки будет несколько меньше.

Как видим, *толщина рабочего слоя и концентрация примеси в нем должны быть достаточно малы.* В противном случае напряжение отсечки будет настолько большим, что полное управление током (начиная с нулевого значения) окажется практически невозможным.

Используя величину $U_{зо}$, толщину канала можно записать в следующей форме:

$$w = a\left(1 - \sqrt{\frac{U_{зи}}{U_{зо}}}\right). \qquad (4.30)$$

Такая толщина сохраняется по всей длине канала. Сопротивление канала в этом случае равно

$$R_0 = \frac{\rho L}{aZ}\left(1 - \sqrt{\frac{U_{зи}}{U_{зо}}}\right)^{-1}, \qquad (4.31)$$

где Z — ширина канала (см. рис. 4.12); ρ — удельное сопротивление n-слоя. При $\rho = 1$ Ом · см, $a = 2$ мкм и $U_{зи} = 0$ получается минимальное значение $R_{0мин} = 0{,}5$ кОм. При $U_{зи}/U_{зо} = 0{,}5$ сопротивление R_0 возрастает до 1,8 кОм.

Статические характеристики. Если подано напряжение $U_{си}$, то через канал протекает ток и поверхность канала, прилегающая к обедненному слою, уже не будет эквипотенциальной. Соответственно напряжение на p–n-переходе будет меняться вдоль оси x, возрастая вблизи стока. Значит, и ширина обедненного слоя перехода будет увеличиваться в направлении от истока к стоку (рис. 4.13, *а*).

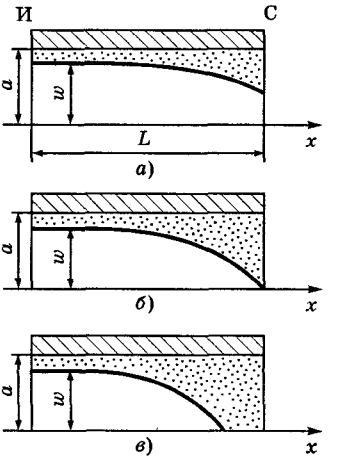


Рис. 4.13. Сечение канала полевого транзистора в ненасыщенном режиме (*а*), на границе насыщения (*б*) и в режиме насыщения (*в*)

Когда разность потенциалов $U_{си} - U_{зи}$ (где $U_{зи} < 0$) сделается равной напряжению отсечки $U_{зо}$, толщина канала вблизи стока станет равной нулю, т.е. образуется «горловина» канала (рис. 4.13, *б*). В отличие от случая $U_{си} = U_{зо}$ это не приводит к отсечке тока, так как само образование «горловины» есть следствие увеличения тока. Вместо отсечки тока происходит отсечка его приращений, т.е. *насыщение тока.*

Образование «горловины» канала знакомо по МДП-транзисторам. В дальнейшем, когда $U_{си} - U_{зи} > U_{зо}$, «горловина» перемещается к истоку, а длина канала несколько уменьшается (рис. 4.13, *в*). Эти явления также свойственны МДП-транзисторам.

Из приведенного описания следует, что напряжение насыщения для полевых транзисторов выражается следующим образом:

$$U_{сн} = U_{зо} - U_{зи}, \quad \text{где } U_{зи} < 0. \tag{4.32}$$

Семейство стоковых ВАХ (рис. 4.14, *а*) имеет сходство с аналогичным семейством для МДП-транзисторов (рис. 4.7, *а*). Однако с ростом напряжения на затворе (по модулю) ток стока в данном случае не возрастает, а уменьшается. Можно сказать, что полевому транзистору свойствен режим обеднения, подобно МДП-транзистору со встроенным каналом.

Семейство стоко-затворных ВАХ (рис. 4.14, *б*) отличается от аналогичного семейства МДП-транзисторов (рис. 4.7, *б*) прежде

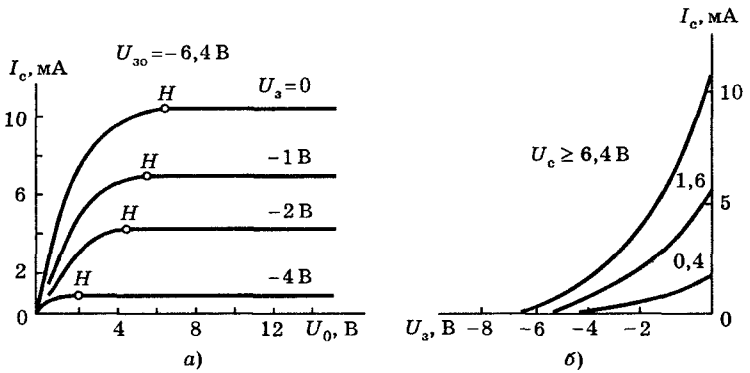


Рис. 4.14. Статические характеристики полевого транзистора: *а* — выходные; *б* — передаточные

всего тем, что ток протекает при нулевом напряжении на затворе. Условно можно сказать, что напряжение отсечки у полевого транзистора эквивалентно отрицательному пороговому напряжению у МДП-транзистора.

Важная особенность ВАХ на рис. 4.14, *б* состоит также в том, что напряжение на затворе может иметь только одну полярность, в данном случае — отрицательную. В противном случае напряжение на p–n-переходе будет прямым, начнется инжекция неосновных носителей и транзистор перестанет быть униполярным прибором. Заметим, что у МДП-транзисторов со встроенным каналом (рис. 4.3), которые во многом аналогичны полевым транзисторам, ограничение на полярность управляющего напряжения не имеет места, так как затвор отделен от канала диэлектриком.

Аналитические выражения для ВАХ полевого транзистора: на крутом участке

$$I_c = \frac{1}{R_{0\text{мин}}}\left[U_{\text{си}} + \frac{2}{3}\frac{U_{\text{зи}}^{3/2} - (U_{\text{зи}} + U_{\text{си}})^{3/2}}{U_{\text{зо}}^{1/2}}\right]; \qquad (4.33)$$

на пологом участке

$$I_c = \frac{1}{R_{0\text{мин}}}\left[\frac{1}{3}U_{\text{зо}} - U_{\text{зи}}\left(1 - \frac{2}{3}\sqrt{\frac{U_{\text{зи}}}{U_{\text{зо}}}}\right)\right], \qquad (4.34)$$

где $R_{0\text{мин}}$ — сопротивление канала при $U_{\text{зи}} = 0$ [см. (4.31)]

Выражение (4.34) хорошо аппроксимируется квадратичной зависимостью, аналогичной зависимости (4.8) для МДП-транзисторов:

$$I_c = \tfrac{1}{2} b(U_{зо} - U_{зи})^2. \qquad (4.35)$$

Здесь коэффициент b, аналогичный удельной крутизне МДП-транзистора, имеет вид:

$$b = 4\varepsilon_0 \varepsilon \mu Z/(3aL). \qquad (4.36)$$

Например, при $\mu = 1500$ см^2/В · с, $Z/L = 10$ и $a = 2$ мкм получаем $b = 0{,}12$ мА/В^2. Заметим, что в этом примере мы использовали значение подвижности, свойственное объему полупроводника, так как у полевых транзисторов канал не граничит с поверхностью.

Полевым транзисторам, как и МДП-транзисторам, свойственно понятие критического тока, при котором зависимость тока от температуры в принципе отсутствует.

У полевых транзисторов наличие критического тока обусловлено противоположным влиянием функций $b(T)$ и $U_{зо}(T)$. Функция $b(T)$ связана с температурной зависимостью подвижности, как и у МДП-транзисторов. Что касается функции $U_{зо}(T)$, то из (4.29) она не вытекает. Однако, если при выводе выражения (4.29) использовать более точную зависимость, то в него войдет равновесная высота барьера в p–n-переходе; последняя зависит от температуры. Именно с учетом этой зависимости и получается величина критического тока.

Из условия $dI_c/dT = 0$ можно найти напряжение на затворе, соответствующее критическому току:

$$U_{зо} - U_{зикр} \approx 0{,}65\,\text{В} \qquad (4.37)$$

[ср. с (4.23)]. Обычно значения критического тока лежат в области микрорежима.

Малосигнальные параметры и эквивалентная схема. Если воспользоваться аппроксимацией (4.35), то на пологом участке крутизна будет равна

$$S = b(U_{зо} - U_{зи}), \qquad (4.38)$$

а зависимость крутизны от тока выражается формулой

$$S = \sqrt{2bI_c}.$$

Дифференциальное сопротивление стока r_c обусловлено той же причиной (модуляцией длины канала) и имеет те же значения, что и у МДП-транзисторов [см. (4.20)].

Малосигнальная эквивалентная схема полевого транзистора показана на рис. 4.15. Элементы этой схемы в сущности те же, что и у МДП-транзистора (см. рис. 4.10): r_c — дифференциальное сопротивление канала на пологом участке ВАХ; $SU_{зи}$ — источник тока, отражающий усилительные свойства транзистора; $R_{зи}$ и $R_{зс}$ — обратные сопротивления p–n-перехода; $C_{зи}$ и $C_{зс}$ — барьерные емкости боковых частей p–n-перехода (рис. 4.12).

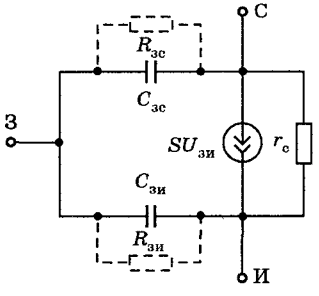


Рис. 4.15. Малосигнальная эквивалентная схема полевого транзистора

Инерционность изменений тока, как и у МДП-транзисторов, характеризуется постоянной времени крутизны τ_S. Этот параметр также представляет собой произведение сопротивления канала на емкость затвор–канал. Поскольку сечения канала и обедненного слоя разные на разных участках (см. рис. 4.13), воспользуемся средними значениями w и l. А именно, примем $w_{ср} = l_{ср} = \frac{1}{2}a$. Тогда средняя емкость и среднее сопротивление канала запишутся следующим образом:

$$\overline{C}_з = \frac{\varepsilon_0 \varepsilon (ZL)}{\frac{1}{2}a}; \qquad (4.39а)$$

$$\overline{R}_0 = \rho \frac{L}{\frac{1}{2}aZ}. \qquad (4.39б)$$

Соответственно средняя постоянная времени крутизны будет иметь вид:

$$\tau_S = 4\varepsilon_0 \varepsilon \rho L^2 / a^2. \qquad (4.40)$$

Эту величину можно привести к такой же форме (4.27), как у МДП-транзисторов. Для этого подставим в (4.40) значение a^2 из (4.29) и учтем соотношение $qN\mu = \sigma = 1/\rho$; тогда

$$\tau_S = 2L^2/(\mu U_{зо}). \qquad (4.41)$$

Таким образом, *в принципе переходные и частотные характеристики полевых и МДП-транзисторов могут быть одинаковыми.* Однако практически длину канала у полевых транзисторов не удается сделать такой же малой, как у современных МДП-транзисторов. Поэтому в настоящее время быстродействие полевых транзисторов значительно ниже.

Естественно, что полевые транзисторы уступают МДП-транзисторам и по величине входного сопротивления: оно определяется обратным током p–n-перехода и обычно не превышает 10^{11} Ом. С ростом температуры это сопротивление быстро уменьшается и на границе рабочего диапазона (+125 °С) может дойти до 10^7 Ом и менее.

Важными достоинствами полевых транзисторов являются высокая стабильность характеристик во времени и малый уровень собственных шумов. Причина этих достоинств в том, что канал отделен от поверхности обедненным слоем, играющим роль диэлектрика. В результате на границе канала с таким «диэлектриком» отсутствуют дефекты кристаллической решетки, поверхностные каналы и загрязнения — все то, что у МДП-транзисторов является причиной нестабильности и шумовых флуктуаций. По той же причине не происходит уменьшения подвижности, свойственного МДП-транзисторам.

Единственный неизбежный тип шумов у полевого транзистора — это тепловой шум, который свойственен каналу, как и любому резистору. Тепловой шум оценивают по формуле Найквиста:

$$U^2_{\text{тш}} = 4kTR\Delta f,$$

где Δf — полоса частот. Подставляя значения $R_{0\text{мин}} = 0{,}5$ кОм и $\Delta f = 1$ Гц, получаем $U^2_{\text{тш}} \approx 3$ нВ.

В приведенном анализе рассматривалась только активная область полевого транзистора — его канал. Влияние пассивных областей (слоев истока и стока, рис. 4.12) сводится к дополнению эквивалентной схемы резисторами $R_{\text{и}}$ и $R_{\text{с}}$, включенными последовательно с истоком и стоком. Сопротивления этих резисторов обычно составляют не более 10–20 Ом, так что их влияние (по сравнению с сопротивлением канала) мало существенно.

Контрольные вопросы

1. Чем отличаются МДП-транзисторы со встроенным и индуцированным каналом?
2. Как связана удельная емкость затвора с толщиной подзатворного диэлектрика?
3. Что такое пороговое напряжение МДП-транзистора?
4. Как влияют заряды в окисле и на поверхностных состояниях на пороговое напряжение?
5. Чему равен поверхностный потенциал при пороговом напряжении?
6. Чему равно напряжение спрямления зон?
7. Чему равна разность потенциалов затвор-исток на границе насыщения?
8. С чем связан наклон ВАХ в области насыщения?
9. В каком режиме МДП-транзистор может использоваться в качестве омического сопротивления?
10. Дайте определение крутизны МДП-транзистора.
11. Как зависит внутреннее сопротивление МДП-транзистора в пологой области от тока стока?
12. Как соотносятся крутизны по затвору и подложке?
13. Чему соответствует критический ток МДП-транзистора?
14. В чем состоит причина нестабильности параметров МДП-транзистора?
15. Как связана постоянная времени крутизны с длиной канала МДП-транзистора?
16. Дайте определение напряжения отсечки полевого транзистора.
17. Как соотносятся входные сопротивления МДП- и полевого транзистора?
18. Как изменяется длина канала полевого транзистора в пологой области при увеличении напряжения на стоке?
19. Сравните быстродействие МДП- и полевых транзисторов.
20. Сравните уровень шумов МДП- и полевых транзисторов.
21. Каков порядок величины сопротивлений пассивных областей полевого транзистора?

Глава 5

Физические принципы работы биполярного транзистора и тиристора

5.1. Введение

Биполярный транзистор, наряду с МДП-транзистором, является одним из основных твердотельных приборов, используемых в микроэлектронике.

В основе работы биполярных транзисторов лежит инжекция неосновных носителей. Поэтому неотъемлемой составной частью биполярных транзисторов являются p–n-переходы. Термин «биполярный» призван подчеркнуть роль обоих типов носителей заряда (электронов и дырок) в работе этого класса транзисторов: инжекция неосновных носителей сопровождается компенсацией их заряда основными носителями.

Предметом данной главы являются изучение физических процессов в биполярном транзисторе, а также анализ его основных характеристик и параметров. В конце главы рассмотрены физические основы работы тиристора, представляющего собой соединение не двух, а трех p–n-переходов. Этот прибор часто встречается в микроэлектронных структурах как «паразитный» элемент, способный приводить к сбоям в работе микросхем.

5.2. Принцип действия

Биполярный транзистор представляет собой совокупность двух встречно включенных взаимодействующих p–n-переходов. Взаимодействие переходов обеспечивается тем, что они расположены достаточно близко друг от друга — на расстоянии, меньшем диффузионной длины носителей.

Структура транзистора. У реальных транзисторов площади обоих p–n-переходов существенно различаются. Это различие проиллюстрировано на рис. 5.1, *а*: переход n_1–p имеет гораздо меньшую площадь, чем n_2–p. Кроме того, у большинства транзисторов один из крайних слоев (а именно, слой с меньшей площадью — n_1) легирован гораздо сильнее, чем второй (n_2). Таким образом, транзистор является асимметричным прибором.

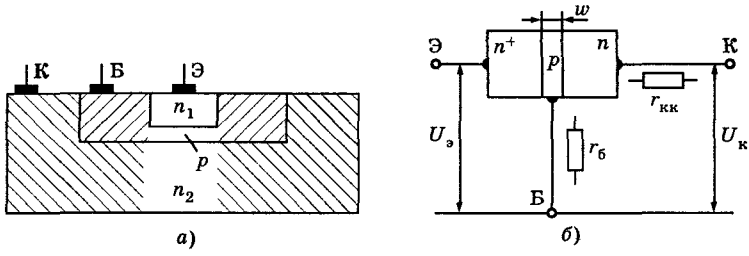


Рис. 5.1. Структура биполярного транзистора:
а — реальная; *б* — идеализированная (без пассивных областей)

Асимметрия транзистора отражается в названиях крайних слоев: сильно легированный слой с меньшей площадью (n_1) называют *эмиттером*, а слой с большей площадью (n_2) — *коллектором*. Соответственно различают эмиттерный и коллекторный переходы (n_1–p и n_2–p). Смысл этих названий поясняется ниже. Средний слой транзистора называют *базой*.

Каждый из p–n-переходов транзистора имеет *донную* и *боковые* части.

Рабочей или, как говорят, *активной* областью транзистора является область, расположенная под донной частью эмиттерного перехода (на рис. 5.1, *а* эта область не заштрихована). Остальные (заштрихованные) участки структуры являются *пассивными*, т.е. в известной мере паразитными. Пассивные участки можно в первом приближении моделировать резисторами, подключенными к рабочим слоям базы и коллектора.

На рис. 5.1, *б* активная область транзистора показана в горизонтальном положении. Здесь же показаны места подключения резисторов $r_б$ и $r_{кк}$, характеризующих пассивные участки. Эмиттерному — высоколегированному — слою присвоен верхний индекс «+». Структура, показанная на рис. 5.1, *б*, служит основой при анализе транзисторов.

Взаимодействие между эмиттерным и коллекторным переходами обеспечивается малой шириной базы w. У современных транзисторов она обычно не превышает 1 мкм, тогда как диффузионная длина L лежит в пределах 5–10 мкм.

Основные свойства транзистора определяются процессами в базе. Если база однородная, то движение носителей в ней чисто диффузионное. Если же база неоднородная, то в ней, как известно, есть внутреннее электрическое поле, и тогда движение носителей будет комбинированным: диффузия сочетается с

дрейфом. Транзисторы с однородной базой называют *бездрейфовыми* (или диффузионными), а с неоднородной — *дрейфовыми*. Последние имеют в настоящее время наибольшее распространение в интегральных схемах.

Транзистор, показанный на рис. 5.1, характерен тем, что его крайние слои (эмиттер и коллектор) имеют проводимость n-типа, а средний слой (база) — проводимость p-типа. Транзисторы с такой структурой называют n–p–n-транзисторами. В микроэлектронике они играют главную роль и будут основой последующего анализа. Однако используются и транзисторы, у которых эмиттер и коллектор имеют проводимость p-типа, а база — проводимость n-типа. Транзисторы с такой структурой называют p–n–p-транзисторами. По принципу действия они ничем не отличаются от n–p–n-транзисторов, однако им свойственны другие полярности рабочих напряжений (см. ниже), а также ряд количественных особенностей.

Способы включения. Зонная диаграмма n–p–n-транзистора с однородной базой в равновесном состоянии показана на рис. 5.2. При равновесии уровень Ферми постоянен по всей структуре. Электростатический потенциал (и связанный с ним уровень E_i) изменяется таким образом, чтобы скомпенсировать диффузионное движение носителей, концентрация которых различна в n и p областях. При нормальном включении транзистора на эмиттерном переходе действует прямое смещение, а на коллекторном — обратное. При этом потенциальный барьер эмиттерного перехода уменьшается, а коллекторного — увеличивается (рис. 5.3). Электроны инжектируются из эмиттера в базу, проходят ее почти без рекомбинации (поскольку ширина базы мала) и беспрепятственно попадают в коллектор, находящийся под положительным потенциалом (см. зонную диаграм-

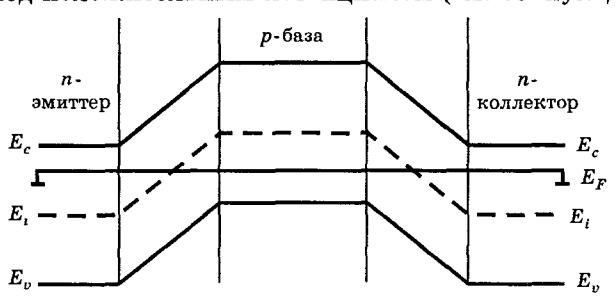


Рис. 5.2. Зонная диаграмма n–p–n-транзистора в равновесном состоянии

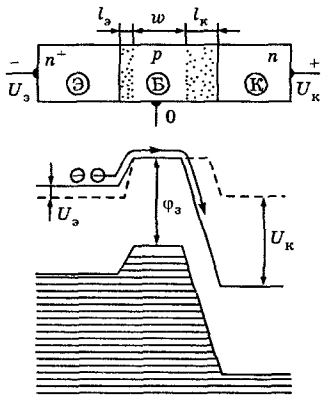


Рис. 5.3. Зонная диаграмма транзистора при нормальном включении

му транзистора на рис. 5.3). Таким образом, при нормальном включении коллектор *собирает* поступившие в базу неосновные носители, откуда — его название.

Ясно, что при указанной полярности напряжения коллектор способен собирать только электроны. Поэтому важно, чтобы ток эмиттера содержал в основном электронную составляющую. Именно поэтому эмиттер легируют значительно сильнее, чем базу. Легирование коллекторной области связано с получением необходимого пробойного напряжения обратносмещенного коллекторного перехода. Как правило, коллектор легирован значительно слабее эмиттера.

В нормальном включении токи коллектора и эмиттера почти одинаковы с точностью до небольшого тока базы. Последний компенсирует убыль основных носителей (дырок) в базе из-за рекомбинации (которая имеет место даже при очень малой толщине базы), а также в результате инжекции дырок из базы в эмиттер.

Сопротивление обратносмещенного коллекторного перехода очень велико — несколько мегом и более. Поэтому в цепь коллектора можно включать весьма большие сопротивления нагрузки, не изменяя величину коллекторного тока. Соответственно в цепи нагрузки может выделяться значительная мощность. Сопротивление прямосмещенного эмиттерного перехода, напротив, весьма мало (например, при токе 1мА оно составляет всего 25 Ом). Поэтому при почти одинаковых токах мощность, потребляемая в цепи эмиттера, оказывается несравненно меньше, чем мощность, выделяемая в цепи нагрузки. Следовательно, транзистор способен усиливать мощность, т.е. является усилительным прибором.

Несмотря на асимметрию транзистора, эмиттер и коллектор в принципе можно поменять ролями: задать на коллекторный переход прямое напряжение, а на эмиттерный — обратное. Такой режим работы называют *инверсным включением*. Передача тока при инверсном включении значительно хуже, чем при

нормальном. Причины этого следующие. Во-первых, в связи со слабым легированием коллектора, мала электронная составляющая коллекторного тока. Во-вторых, площадь реального коллектора значительно больше площади эмиттера (рис. 5.1, *а*), поэтому на эмиттер попадает лишь небольшая часть электронов, инжектированных коллектором.

Особое место в работе транзистора занимает *режим двойной инжекции* или, менее точно, *режим насыщения*. Режим двойной инжекции характерен тем, что на обоих переходах — эмиттерном и коллекторном — действуют прямые напряжения. При этом и эмиттер, и коллектор инжектируют носители в базу навстречу друг другу и одновременно каждый из них собирает носители, дошедшие от другого.

До сих пор мы задавали напряжение на эмиттере и коллекторе относительно базы (рис. 5.1, *б*). Такое включение транзистора называют *включением с общей базой* (или *схемой с общей базой*) и обозначают ОБ. Напомним, что задавать прямое напряжение на p–n-переходе практически невозможно; как правило, задается прямой ток. Значит, *для включения ОБ характерна заданная величина тока эмиттера.*

Схема ОБ позволяет хорошо раскрыть физику транзистора; есть у нее и некоторые другие положительные особенности. Но тот факт, что она не обеспечивает усиления тока и обладает малым входным сопротивлением (равным сопротивлению эмиттерного перехода), делает ее не оптимальной для большинства применений. Поэтому главную роль в транзисторной технике играет другое включение — *с общим эмиттером*, которое обозначают ОЭ. *Для включения ОЭ характерна заданная величина тока базы.* Оба включения показаны на рис. 5.4, *а, б* с использованием схемотехнических символов, присвоенных n–p–n-транзистору.

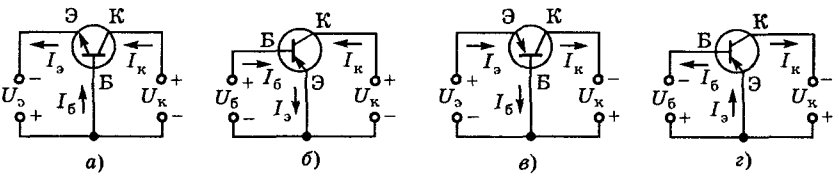


Рис. 5.4. Включение n–p–n-транзистора с общей базой (*а*) и с общим эмиттером (*б*) и включение p–n–p-транзистора с общей базой (*в*) и с общим эмиттером (*г*).

На рис. 5.4, *в*, *г* показаны аналогичные включения *p–n–p*-транзисторов. Как уже отмечалось, этим транзисторам свойственны обратные полярности рабочих напряжений. Соответственно обратные направления имеют и рабочие токи.

Использование *p–n–p*-транзисторов открывает возможность сочетания *n–p–n*- и *p–n–p*-транзисторов в одной и той же схеме. Такое сочетание в ряде случаев обеспечивает упрощение структуры и оптимизацию параметров соответствующих схем. Транзисторы *n–p–n* и *p–n–p* в таких схемах, а также сами схемы такого типа называют *комплементарными* (дополняющими), как и в МДП-микросхемах.

5.3. Распределения носителей

Для того чтобы рассчитывать токи, межэлектродные напряжения и избыточные заряды в транзисторе, необходимо знать распределения избыточных концентраций, т.е. функции $\Delta n(x)$ или $\Delta p(x)$. Эти функции рассматриваются ниже применительно к главному элементу ИС — транзистору *n–p–n*-типа.

В одном и тоже слое функции Δn и Δp совпадают в силу принципа квазинейтральности, поэтому ниже приводятся выражения только для избыточных концентраций неосновных носителей (электронов).

Случай нормального включения бездрейфового транзистора. В стационарном режиме концентрация носителей, инжектированных в базу, описывается диффузионным уравнением

$$D_{\text{б}} \frac{\mathrm{d}^2 n_p}{\mathrm{d}x^2} - \frac{n_p - n_0}{\tau_{\text{б}}} = 0, \qquad (5.1)$$

где $D_{\text{б}}$ и $\tau_{\text{б}}$ — коэффициент диффузии и время жизни электронов в базе; n_p — концентрация электронов; $n_0 = n_i^2/N_{\text{б}}$ — равновесная концентрация электронов; $N_{\text{б}}$ — концентрация акцепторов в базе.

При записи граничных условий для нормального включения транзистора примем, что на коллекторном переходе задано обратное напряжение $U_{\text{к}}$, а на эмиттерном переходе — прямой ток $I_{\text{э}}$, точнее — его электронная составляющая $I_{\text{эп}}$.

Концентрация электронов у коллекторного перехода находится из граничного условия Шокли

$$n_p(w) = n_0 \mathrm{e}^{\frac{U_{\text{к}}}{\varphi_{\text{т}}}}. \qquad (5.2)$$

Так как коллекторный переход имеет обратное смещение, то величина $n_p(w)$ близка к нулю и избыточная концентрация электронов $\Delta n_б = n_p(w) - n_0 = -n_0$. Поскольку в базе p-типа равновесная концентрация электронов очень мала, пренебрежем величиной n_0 и запишем первое граничное условие в виде

$$\Delta n_б(w) = 0. \tag{5.3}$$

Электроны, инжектированные в базу, распространяются в ней под действием диффузии, поэтому для электронной составляющей можно записать

$$I_{эп} = -qD_б S \frac{\mathrm{d}\Delta n_б}{\mathrm{d}x}\Big|_{x=0}, \tag{5.4}$$

где S — площадь эмиттерного перехода. Знак минус в правой части (5.4) отражает тот факт, что у n–p–n-транзисторов прямой, т.е. положительный ток эмиттера означает инжекцию электронов в базу; при этом градиент их концентрации должен быть отрицательным.

Из (5.4) легко найти градиент концентрации электронов и соответственно второе граничное условие:

$$\frac{\mathrm{d}(\Delta n_б)}{\mathrm{d}x}\Big|_{x=0} = -\frac{I_{эп}}{qD_б S}. \tag{5.5}$$

Используя граничные условия (5.3) и (5.5), можно найти решение диффузионного уравнения в следующем виде:

$$\Delta n_б(x) = I_{эп} \frac{L}{qD_б S} \frac{\mathrm{sh}[(w-x)/L]}{\mathrm{ch}(w/L)}. \tag{5.6}$$

Поскольку в транзисторах выполняется неравенство $w << L$, полученное выражение можно упростить, воспользовавшись соотношениями, действительными для малых аргументов: $\mathrm{sh}(z) \approx z$ и $\mathrm{ch}(z) \approx 1$. Тогда

$$\Delta n_б(x) = I_{эп} \frac{w}{qD_б S}\left(1 - \frac{x}{w}\right). \tag{5.7}$$

Как видим, *бездрейфовым транзисторам с однородной базой свойственно почти линейное распределение избыточных носителей* (рис. 5.5).

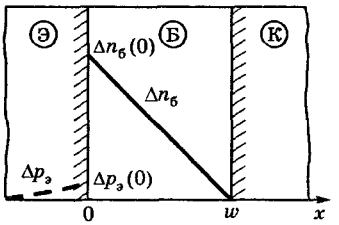


Рис. 5.5. Распределение электронов в базе бездрейфового n–p–n-транзистора

Интегрируя функцию $\Delta n_б(x)$ в пределах от 0 до w и умножая интеграл на площадь S и элементарный заряд q, получим величину избыточного заряда в базе:

$$\Delta Q_б = I_{эп}(w^2/2D_б). \qquad (5.8)$$

Как видим, избыточный заряд пропорционален току эмиттера, а при заданном токе уменьшается с уменьшением толщины базы.

Обратимся теперь к распределению дырок, инжектированных из базы в эмиттер. Поскольку эмиттерный слой значительно толще базового, ему свойственно неравенство $w_э >> L_э$, где $w_э$ — толщина эмиттерного слоя, $L_э$ — диффузионная длина носителей (дырок). При выполнении этого неравенства распределение избыточных носителей будет таким же, как при бесконечно длинной базе, т.е. экспоненциальным. Заменяя в формуле (5.6) электронную составляющую тока на дырочную, концентрацию $\Delta n_б$ на $\Delta p_э$ и полагая $w \to \infty$, получаем:

$$\Delta p_э(x) = I_{эр}\frac{L_э}{qD_э S}e^{-x/L_э}. \qquad (5.9)$$

Расстояние x отсчитывается здесь от границы перехода вглубь эмиттера (см. штриховую кривую на рис. 5.5). Соотношение граничных концентраций $\Delta n_б(0)$ и $\Delta p_э(0)$ определяется граничным условием Шокли.

Интегрируя (5.9) в пределах от $x = 0$ до $x = \infty$, умножая результат на S и q, получим величину избыточного заряда в эмиттере:

$$\Delta Q_э = I_{эр}\tau_э, \qquad (5.10)$$

где $\tau_э$ — время жизни неосновных носителей в эмиттерном слое.

Случай нормального включения дрейфового транзистора. Концентрация носителей в неоднородной базе описывается уравнением непрерывности с учетом диффузионного и дрейфового движения. В общем случае распределение получается гро-

моздким и ненаглядным. Для тонкой базы оно существенно упрощается:

$$\Delta n_б(x) = I_{эп}\frac{w}{qD_б S}\frac{1-\mathrm{e}^{-2\eta[1-(x/w)]}}{2\eta}, \qquad (5.11)$$

где $\eta = 1/2\ \ln[N_б(0)/N_б(w)]$; $N_б(0)$ и $N_б(w)$ — концентрации примеси на эмиттерной и коллекторной границах базы. Чем больше разница в концентрациях примеси, тем больше величина η. Поэтому ее называют *коэффициентом неоднородности базы*. На практике типичны значения $\eta = 2$–3.

Распределение носителей, описываемое выражением (5.11), показано на рис. 5.6 в относительном масштабе. За единицу масштаба принята граничная концентрация носителей у бездрейфового транзистора, получаемая из (5.7) при $x = 0$

$$\Delta n_D(0) = I_{эп}(w/qD_б S)$$

(индекс D символизирует чисто диффузионное движение). Из рис. 5.6 видно, что с ростом коэффициента неоднородности распределение носителей в базе дрейфового транзистора все больше отличается от линейного. При одинаковых значениях эмиттерного тока избыточные концентрации у дрейфовых транзисторов значительно меньше, чем у бездрейфовых.

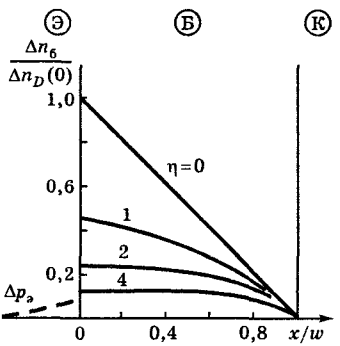


Рис. 5.6. Распределение электронов в базе дрейфового транзистора

Интегрируя функцию (5.11) и умножая на S и q, получаем величину избыточного заряда в базе:

$$\Delta Q_б = I_{эп}\frac{w^2}{2D_б}\left(\frac{2\eta - 1 + \mathrm{e}^{-2\eta}}{2\eta^2}\right). \qquad (5.12а)$$

Сложная функция, стоящая в круглых скобках, хорошо аппроксимируется простой функцией $(\eta + 1)^{-1}$. Поэтому запишем избыточный заряд в виде

$$\Delta Q_б = I_{эп}[w^2/2(\eta + 1)D_б]. \qquad (5.12б)$$

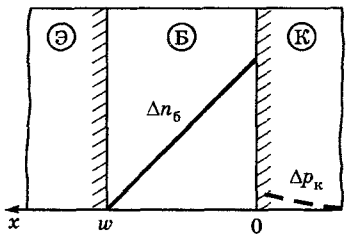


Рис. 5.7. Распределение электронов в базе бездрейфового транзистора при инверсном включении

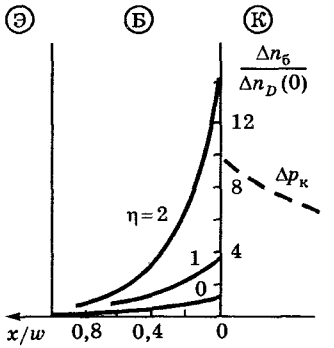


Рис. 5.8. Распределение электронов в базе дрейфового транзистора при инверсном включении

Распределение дырок, инжектированных из базы в эмиттер, показано на рис. 5.6 штриховой линией. Это распределение и соответствующий избыточный заряд можно оценить по формулам (5.9) и (5.10).

Случаи инверсного включения и режима двойной инжекции. При инверсном включении бездрейфового транзистора распределения носителей и в базе, и в коллекторе получаются практически такими же, как при нормальном включении (ср. рис. 5.7 и 5.5).

У дрейфовых транзисторов инверсное включение характерно качественно иным распределением носителей в базе. Это связано с тем, что поле в базе является для электронов, инжектируемых коллектором, не ускоряющим, а тормозящим. Изменяя знак коэффициента η в выражении (5.11), получаем следующее распределение избыточных носителей в базе:

$$\Delta n_б(x) = I_{кп} \frac{w}{qD_б S} \frac{e^{2\eta[1-(x/w)]} - 1}{2\eta}, \qquad (5.13)$$

где $I_{кп}$ — электронная составляющая коллекторного тока, а координата x отсчитывается от коллектора к эмитттеру. Соответствующие кривые распределения показаны на рис. 5.8. Сравнивая с рис. 5.6, видим, что при инверсном включении избыточные концентрации у дрейфового транзистора значительно больше, чем у бездрейфового. Соответственно больше оказыва-

ются и избыточные заряды. Изменяя знак коэффициента η в (5.12а), получаем величину избыточного заряда в базе:

$$\Delta Q_{б} = I_{кп} \frac{w^2}{2D_{б}} \frac{e^{2\eta} - 2\eta - 1}{2\eta^2}.$$

Для избыточных концентраций и зарядов в коллекторе действительны выражения (5.9) и (5.10), если заменить индексы «э» на «к». Поскольку у дрейфовых транзисторов коллекторный переход почти симметричный, граничные концентрации электронов и дырок почти одинаковы (см. штриховую линию на рис. 5.8). Соответственно сравнимыми будут электронная и дырочная составляющие коллекторного тока, а также избыточные заряды в коллекторе и базе.

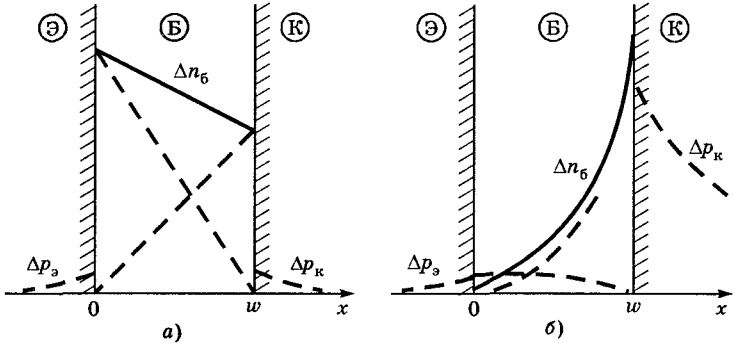


Рис. 5.9. Распределение электронов в базе в режиме двойной инжекции (штриховые линии — составляющие нормального и инверсного включений): *а* — у бездрейфового транзистора, *б* — у дрейфового транзистора

В режиме двойной инжекции, когда оба перехода — эмиттерный и коллекторный — работают в прямом направлении, распределение избыточных носителей в базе можно оценить приближенно, суммируя распределения при нормальном и инверсном включениях (рис. 5.9). При этом у бездрейфовых транзисторов (рис. 5.9, *а*) распределение приобретает трапецеидальную форму, а избыточный заряд оказывается значительно больше, чем при нормальном включении. Что касается дрейфовых транзисторов (рис. 5.9, *б*), то у них результирующее распределение и результирующий избыточный заряд близки к величинам, свойственным инверсному включению.

Суммарный избыточный заряд, накапливающийся во всех трех слоях транзистора, можно выразить через ток базы. Действительно

$$I_б = I_{бр} + I_{эр} + I_{кр},$$

где $I_{бр}$ — ток рекомбинации дырок в базе, $I_{эр}$ и $I_{кр}$ — дырочные составляющие токов эмиттера и коллектора. Очевидно, что

$$I_{бр} = \frac{\Delta Q_б}{\tau_б}, \quad I_{эр} = \frac{\Delta Q_э}{\tau_э}, \quad I_{кр} = \frac{\Delta Q_к}{\tau_к}.$$

Поэтому

$$I_б = \frac{\Delta Q_б}{\tau_б} + \frac{\Delta Q_э}{\tau_э} + \frac{\Delta Q_к}{\tau_к}.$$

Если для простоты положить времена жизни во всех трех слоях транзистора одинаковыми и равными τ, то суммарный заряд будет связан с током базы элементарным соотношением:

$$\Delta Q = I_б \tau. \qquad (5.14)$$

При неодинаковых временах жизни пропорциональность между суммарным зарядом и током базы сохраняется, но выражение будет более громоздким. В тех случаях, когда зарядами $\Delta Q_э$ и $\Delta Q_к$ можно пренебречь, выражение (5.14) используется для оценки избыточного заряда в базе.

До сих пор мы подразумевали идеализированную структуру транзистора. В реальной структуре площадь коллекторного перехода значительно больше площади эмиттерного. Поэтому при инверсном включении коллектор инжектирует электроны не только в активную, но и в пассивную область базы. Суммарный заряд, накапливающийся в обеих областях, разумеется, больше, чем при нормальном включении.

Поскольку пассивная область базы значительно толще активной, а граничные концентрации избыточных электронов одинаковы, распределение носителей в пассивной области оказывается более пологим. Различие в распределениях создает на границе активной и пассивной областей базы градиент концентрации электронов. В результате часть электронов, инжектированных в пассивную область с площади S_2 (рис. 5.10), отклоня-

ется от прямолинейной траектории и попадает не на поверхность, а на боковые части эмиттера. При этом задача анализа делается неодномерной, т.е. сильно усложняется. Однако суммарный заряд избыточных носителей, накапливающийся в активной и пассивной областях базы, по-прежнему определяется формулой (5.14).

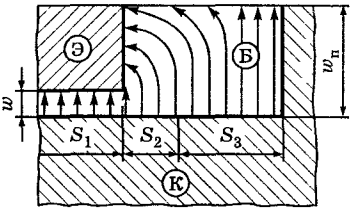


Рис. 5.10. Траектории электронов, инжектированных в базу при инверсном включении

5.4. Коэффициенты усиления тока

В обычных транзисторных схемах выходной (управляемой) величиной является либо коллекторный, либо эмиттерный ток, а входной (управляющей) — либо ток базы, либо ток эмиттера. Связь между выходными и входными токами характеризуется коэффициентами усиления.

Общие определения. Связь между коллекторным и эмиттерным токами можно записать в виде[1]:

$$I_к = \alpha I_э. \qquad (5.15)$$

Здесь α — *коэффициент усиления эмиттерного тока* — один из основных параметров транзистора. Этот параметр особенно удобен тогда, когда ток эмиттера можно считать заданной величиной, например, в схеме ОБ. Коэффициент α близок к единице. У интегральных транзисторов он обычно составляет 0,99–0,995.

Чтобы установить связь между током коллектора и током базы, подставим в (5.15) значение $I_э = I_к + I_б$. Тогда искомое соотношение легко приводится к виду:

$$I_к = B I_б. \qquad (5.16)$$

Здесь B — *коэффициент усиления базового тока*:

$$B = \alpha/(1 - \alpha). \qquad (5.17)$$

---

1 При условии $I_э = 0$ (при оборванном эмиттере) в цепи коллектора, находящегося под обратным напряжением, протекает обратный ток перехода. Однако обратный ток даже при высоких температурах значительно меньше рабочих токов, и его можно не учитывать.

Этот параметр, широко используемый в транзисторной электронике, особенно удобен тогда, когда задан ток базы, прежде всего в схеме ОЭ. Типичные значения коэффициента B лежат в пределах 100—300. Коэффициент усиления B тем больше, чем ближе коэффициент α к единице.

Запишем коэффициент α в следующем виде:

$$\alpha = \frac{I_к}{I_э} = \frac{I_{эп}}{I_э}\frac{I_к}{I_{эп}}.$$

Каждый из двух множителей в правой части имеет свой физический смысл и свое название.

Первый множитель

$$\gamma = I_{эп}/I_э = I_{эп}/(I_{эп} + I_{эр}) \qquad (5.18)$$

называется *коэффициентом инжекции*. Он характеризует долю полезной — электронной — составляющей в общем токе эмиттера[1]. Только эта составляющая, как уже отмечалось, способна дойти до коллектора и составить коллекторный ток.

Второй множитель

$$\varkappa = I_к/I_{эп} \qquad (5.19)$$

называется *коэффициентом переноса*. Он характеризует долю инжектированных носителей, избежавших рекомбинации на пути к коллектору. Только такие носители образуют коллекторный ток.

Таким образом, коэффициент усиления эмиттерного тока можно записать в виде

$$\alpha = \gamma\varkappa . \qquad (5.20)$$

Учитывая определяющую роль параметра α в работе транзистора, рассмотрим его составляющие.

Коэффициент переноса. Для того чтобы найти коэффициент $\varkappa$ из определения (5.19), нужно предварительно рассчитать ток $I_к$. Для бездрейфовых транзисторов это можно сделать, воспользовавшись распределением (5.6). Определив градиент кон-

1 Напомним, что мы рассматриваем n–p–n-транзистор У p–n–p-транзистора полезной составляющей будет дырочная

центрации при $x = w$ и умножив на площадь перехода S, найдем ток $I_к$. После этого из (5.19) получаем:

$$\varkappa = \frac{1}{\mathrm{ch}(w/L)} = \mathrm{sech}\frac{w}{L} \tag{5.21}$$

(знак «минус» опущен, поскольку отрицательному градиенту концентрации электронов соответствует положительное направление тока $I_к$). Выражение (5.21) — одно из фундаментальных в теории транзисторов.

Учитывая соотношение $w << L$, можно разложить правую часть (5.21) в ряд с точностью до двух первых членов и получить более удобное выражение:

$$\varkappa = \frac{1}{1 + 1/2(w/L)^2}. \tag{5.22а}$$

Поскольку второй член знаменателя много меньше единицы, можно воспользоваться еще одним приближением:

$$\varkappa = 1 - 1/2(w/L)^2. \tag{5.22б}$$

Например, если w/L = 0,1–0,2, то $\varkappa$ = 0,98–0,995.

Выражения (5.22) ясно показывают, что *коэффициент переноса тем ближе к единице, чем больше диффузионная длина и чем меньше ширина базы.* Однако увеличение диффузионной длины (т.е. времени жизни) сопровождается, как увидим, ухудшением частотных свойств транзистора. Поэтому *главным направлением в развитии транзисторов является уменьшение ширины базы.*

Для дрейфовых транзисторов коэффициент переноса получается аналогичным способом и имеет аналогичную структуру:

$$\varkappa = \frac{1}{1 + \dfrac{1}{2(\eta + 1)}\left(\dfrac{w}{L}\right)^2} \tag{5.23а}$$

или

$$\varkappa = 1 - \frac{1}{2(\eta + 1)}\left(\frac{w}{L}\right)^2. \tag{5.23б}$$

Отличие от формул (5.22) состоит в дополнительном множителе $(\eta + 1)^{-1}$. Например, если в бездрейфовом транзисторе $\varkappa = 0{,}98$–$0{,}995$, то в дрейфовом транзисторе при той же ширине базы и $\eta = 3$ получится $\varkappa = 0{,}995$–$0{,}999$. С физической точки зрения увеличение коэффициента переноса у дрейфовых транзисторов объясняется тем, что носители в ускоряющем поле двигаются быстрее и тем самым уменьшается вероятность их рекомбинации.

Коэффициент инжекции. Поделим числитель и знаменатель в правой части (5.18) на ток $I_{эп}$. Далее подставим токи $I_{эп}$ и $I_{эр}$ из выражений (5.7) и (5.9), положив в них $x = 0$. Отношение граничных концентраций $\Delta p_э(0)/\Delta n_б(0)$ заменим отношением концентраций примесей $N_б/N_э$ согласно условиям Шокли. Тогда коэффициент инжекции для дрейфового транзистора запишется следующим образом:

$$\gamma = \left(1 + \frac{D_э}{D_б}\frac{w}{L_э}\frac{N_б}{N_э}\frac{1 - e^{-2\eta}}{2\eta}\right)^{-1}. \qquad (5.24а)$$

При $\eta > 1$ экспоненциальным членом можно пренебречь. Для бездрейфовых транзисторов, полагая $\eta = 0$, получаем:

$$\gamma = \left(1 + \frac{D_э}{D_б}\frac{w}{L_э}\frac{N_б}{N_э}\right)^{-1}. \qquad (5.24б)$$

Как видим, коэффициент инжекции тем ближе к единице, чем меньше ширина базы и чем больше разница между граничными концентрациями примесей в эмиттерном и базовом слоях. Поэтому *эмиттерный слой всегда стараются легировать как можно сильнее*, так что он, как правило, оказывается полуметаллом. При этом расчетные значения γ могут составлять 0,9999 и более.

Формулы (5.24) выведены из условия, что токи $I_{эп}$ и $I_{эр}$ чисто инжекционные, т.е. потери носителей в области эмиттерного перехода отсутствуют. В микрорежиме, при малых токах, такое предположение не оправдано и приходится учитывать рекомбинацию в переходе. Тогда соотношение между электронной и дырочной составляющими эмиттерного тока дополнительно меняется в пользу дырочной составляющей. Иначе говоря, коэффициент инжекции дополнительно уменьшается.

Тот факт, что рекомбинация становится существенной именно в области малых токов, объясняется следующим. Рекомбинационные потери носителей зависят от объема перехода, а значит, сравнительно постоянны. Поэтому на фоне больших потоков носителей роль таких потерь незначительна, а с уменьшением потоков она возрастает. Большая доля рекомбинационных потерь приходится и на приповерхностный слой. Следовательно, *качество обработки поверхности оказывает первостепенное влияние на величину коэффициента инжекции в области малых токов.*

Типичные значения коэффициента инжекции с учетом рекомбинации в эмиттерном переходе составляют 0,99–0,997 в нормальном токовом режиме и 0,98–0,985 в микрорежиме.

Коэффициенты усиления при нормальном и инверсном включениях. Перемножая коэффициенты переноса и инжекции, пренебрегая членом второго порядка малости и разлагая в ряд с точностью до членов первого порядка, получаем коэффициент усиления эмиттерного тока:

$$\alpha = 1 - \frac{1}{2(\eta+1)}\left(\frac{w}{L_б}\right)^2 - \frac{D_э}{D_б}\frac{w}{L_э}\frac{N_б}{N_э}\frac{1-e^{-2\eta}}{2\eta}. \qquad (5.25)$$

Подставляя значение α в выражение (5.17) и пренебрегая членом второго порядка малости, получаем коэффициент усиления базового тока. Запишем его в виде:

$$\frac{1}{B} = \frac{1}{2(\eta+1)}\left(\frac{w}{L_б}\right)^2 + \frac{D_э}{D_б}\frac{w}{L_э}\frac{N_б}{N_э}\frac{1-e^{-2\eta}}{2\eta}. \qquad (5.26)$$

Зависимость (5.26) позволяет сделать следующие выводы:

- коэффициент усиления транзистора возрастает с уменьшением ширины базы;
- при сравнительно большой ширине базы определяющую роль играет коэффициент переноса, при достаточно малой — коэффициент инжекции;
- при прочих равных условиях коэффициент усиления дрейфового транзистора больше, чем бездрейфового.

Коэффициенты усиления при инверсном включении не поддаются строгому анализу, так как процессы движения носителей при этом двумерные (рис. 5.10). Многие носители, инжек-

тированные коллектором в пассивную область базы, не попадают на эмиттер, а рекомбинируют в слое базы и на поверхности. Поэтому инверсный коэффициент переноса $\varkappa_I$ в значительной мере определяется отношением площади эмиттера (S_1) к площади коллектора ($S_1+S_2+S_3$) и может быть существенно меньше единицы. К тому же если коллекторный переход почти симметричный, то коэффициент инжекции γ_I тоже будет небольшим. В таких условиях инверсный коэффициент усиления α_I может составлять 0,7 и менее даже у бездрейфовых транзисторов. У дрейфовых транзисторов благодаря влиянию тормозящего поля он еще меньше.

Заметим, однако, что поскольку заметная часть носителей, инжектированных в пассивную область базы, все же доходит до эмиттера со стороны его боковой поверхности, то при оценке коэффициента $\varkappa_I$ следует пользоваться не площадью донной части эмиттера (S_1 на рис. 5.10), а некоторой эффективной площадью S_1+S_2, которая может в несколько раз превышать донную. Это способствует увеличению коэффициента $\varkappa_I$. С другой стороны, длина траекторий у носителей, попадающих на боковую часть эмиттера, больше, чем w, что приводит к уменьшению $\varkappa_I$.

Таким образом, *инверсные параметры* α_I *и* B_I *всегда меньше нормальных*. Однако в зависимости от структуры транзистора их значения могут меняться в довольно широких пределах. Так, коэффициент усиления B_I обычно не превышает 0,5–1,5, но в специальных структурах составляет до 5–10 и более.

Режимная и температурная зависимости. Коэффициенты α и B зависят от рабочего режима транзистора (т.е. от тока и напряжения на коллекторном переходе), а также от температуры. Зависимости коэффициента B показаны соответственно на рис. 5.11 и 5.12.

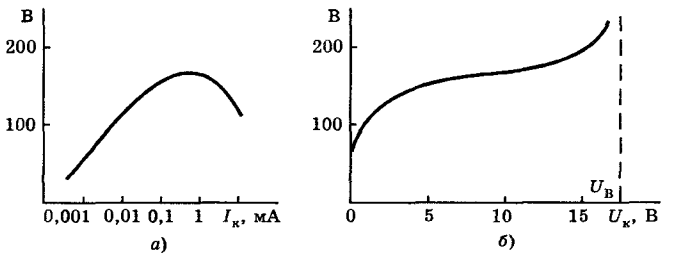


Рис. 5.11. Зависимость коэффициента усиления тока от рабочего режима: *а* — от коллекторного тока; *б* — от коллекторного напряжения

Спад коэффициента усиления в области малых токов объясняется уменьшением коэффициента инжекции из-за рекомбинации в эмиттерном переходе и поверхностной рекомбинации. Спад в области больших токов объясняется увеличением удельной проводимости базы при высоких концентрациях избыточных носителей. Этот случай равносилен увеличению концентрации примеси в базе, что согласно (5.24) приводит к уменьшению коэффициента инжекции.

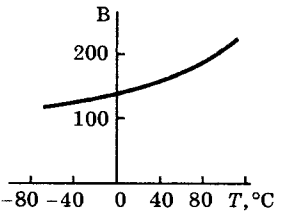


Рис. 5.12. Зависимость коэффициента усиления тока от температуры

В аналитическом виде зависимость B(I) в области малых токов можно записать в следующем виде:

$$B_2 = B_1 \sqrt[N]{\frac{I_{к2}}{I_{к1}}}, \qquad (5.27)$$

где значение B_1 соответствует току $I_{к1}$, а B_2 — току $I_{к2}$. Показатель степени N — своеобразная характеристика электрофизических свойств транзистора, его способности работать в микрорежиме. В настоящее время можно считать $N \approx 6$, т.е. зависимость B от тока сравнительно слабая. Некоторое время назад характерными значениями были $N \approx 3$ и даже $N \approx 2$. Если положить $I_{к2}/I_{к1} \approx 10^{-3}$, то при $N = 6$ получаем $B_2 \approx 0{,}3\ B_1$; при $N = 2$ значение B_2 будет на порядок меньше.

Зависимость B от напряжения $U_к$ обусловлена, во-первых, так называемым *эффектом Эрли*, и, во-вторых, предпробойными явлениями в коллекторном переходе.

Сущность эффекта Эрли состоит в следующем. При изменениях обратного коллекторного напряжения меняется ширина коллекторного перехода $l_к$ (см. рис. 5.3). Изменения величины $l_к$ приводят к изменению ширины базы w: если переход расширяется, то база сужается и наоборот. В наихудшем случае $\Delta w = -\Delta l_к$ (прокол базы). Изменения ширины базы влияют на целый ряд параметров транзисторов, поэтому эффект Эрли часто приходится принимать во внимание.

С увеличением напряжения $U_к$ ширина базы из-за эффекта Эрли уменьшается, а значит, коэффициент B растет. Когда напряжение $U_к$ приближается к пробивному, ток коллектора, а

значит, и коэффициент усиления дополнительно возрастают благодаря ударной ионизации в коллекторном переходе. В этом диапазоне напряжений коэффициент усиления можно записать в виде

$$B = M\alpha/(1 - M\alpha), \tag{5.28}$$

где M — коэффициент ударной ионизации.

При условии $M\alpha = 1$, когда $B \to \infty$, наступает специфическая разновидность пробоя, свойственная включению ОЭ, т.е. режиму заданного тока базы. Приравнивая значение $M = 1/(1 - (U/U_M)^n)$ значению $1/\alpha$, нетрудно получить напряжение такого пробоя (рис. 5.11, *б*):

$$U_B = U_M \sqrt[n]{1 - \alpha}, \tag{5.29}$$

где U_M — напряжение лавинного пробоя (при котором $M = \infty$); n — показатель степени, значение которого для кремния лежит в диапазоне от 3 до 5.

Напряжение U_B значительно меньше напряжения лавинного пробоя U_M, свойственного включению ОБ, т.е. режиму заданного тока эмиттера. Например, если $\alpha = 0{,}99$ и $n = 3$, то $U_B \approx 0{,}2\,U_M$.

Пробой может иметь место не только в результате лавинной ионизации, но и в результате сужения базы по мере роста коллекторного напряжения (эффект Эрли, см. выше). Если коллекторный переход расширится настолько, что ширина базы сделается равной нулю, то переходы транзистора сомкнутся и ток будет беспрепятственно проходить из эмиттера в коллектор, т.е. наступит пробой. Такой эффект называют *эффектом смыкания*, а соответствующее напряжение — *напряжением смыкания*. Анализ показывает, что напряжение смыкания имеет вид:

$$U_w = (qN_б/2\varepsilon_0\varepsilon)w_0^2, \tag{5.30}$$

где $N_б$ — концентрация примеси в базе, w_0 — ширина базы при $U_к = 0$. Такой вид пробоя свойствен транзисторам с особо тонкой базой. Например, если $N_б = 10^{16}$ см$^{-3}$ и $w_0 = 0{,}7$ мкм, то $U_w = 3{,}5$ В.

Зависимость коэффициента B от температуры обусловлена главным образом температурной зависимостью времени жизни $\tau(T)$. С ростом температуры время жизни растет, а вместе с ним

растет диффузионная длина $L_б$, т.е. уменьшается рекомбинация в базе во время пролета носителей. Это приводит к увеличению коэффициента B. Кроме того, с ростом времени жизни замедляется рекомбинация и в эмиттерном переходе, а это способствует увеличению коэффициента инжекции и коэффициента B.

5.5. Статические характеристики

Исходя из того, что биполярный транзистор есть совокупность двух встречно-включенных взаимодействующих p–n-переходов, можно представить его в виде эквивалентной схемы или, как говорят, физической модели. Одна из таких моделей показана на рис. 5.13. Ее называют *моделью Молла–Эберса.*

Модель Молла—Эберса. Эта модель характеризует только активную область транзистора. Если дополнить модель резисторами, отражающими наличие пассивных областей базы и коллектора, ее использование сильно усложняется, и она теряет свою наглядность.

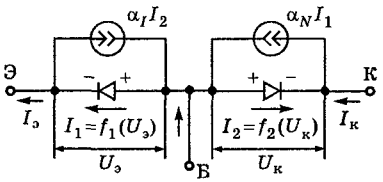


Рис. 5.13. Физическая модель биполярного транзистора

Модель Молла—Эберса хорошо отражает обратимость транзистора — принципиальную равноправность обоих его переходов. Эта равноправность особенно ярко проявляется в режиме двойной инжекции, когда на обоих переходах действуют прямые напряжения. В таком режиме каждый из переходов одновременно инжектирует носители в базу и собирает носители, дошедшие от другого перехода. Токи инжектируемых носителей обозначены через I_1 и I_2, а токи собираемых — через $\alpha_N I_1$ и $\alpha_I I_2$, где α_N и α_I — коэффициенты передачи тока соответственно при нормальном и инверсном включениях. Токи $\alpha_N I_1$ и $\alpha_I I_2$ обеспечиваются в рассматриваемой модели с помощью источников (генераторов) тока[1].

1 Источник или генератор тока — понятие, широко используемое в теории электрических цепей. Это — дуальный аналог источника или генератора э. д. с. Последнему в идеальном случае свойственно нулевое внутреннее сопротивление, а идеальному источнику тока бесконечно большое. он «жестко» задает ток в цепи независимо от ее сопротивления.

Запишем соотношения, вытекающие из рис. 5.13:

$$I_э = I_1 - \alpha_I I_2; \tag{5.31а}$$

$$I_к = \alpha_N I_1 - I_2. \tag{5.31б}$$

Будем считать, что ВАХ каждого из p–n-переходов описывается выражениями:

$$I_1 = I'_{э0}\,(e^{U_э/\varphi_T} - 1); \tag{5.32а}$$

$$I_2 = I'_{к0}\,(e^{U_к/\varphi_T} - 1), \tag{5.32б}$$

где $I'_{э0}$ и $I'_{к0}$ — тепловые токи соответствующих переходов. Каждый из них можно измерить, задавая обратное напряжение $|U| > 3\varphi_T$ на данном переходе и закорачивая второй переход. Однако на практике принято измерять тепловые токи, не закорачивая, а обрывая цепь второго перехода. Соответствующие значения обозначают через $I_{э0}$ и $I_{к0}$.

С помощью формул (5.31) нетрудно установить связь между тепловыми токами, измеренными в режиме холостого хода и в режиме короткого замыкания второго перехода:

$$I'_{э0} = \frac{I_{э0}}{1 - \alpha_N \alpha_I}; \tag{5.33а}$$

$$I'_{к0} = \frac{I_{к0}}{1 - \alpha_N \alpha_I}. \tag{5.33б}$$

Именно величины $I_{э0}$ и $I_{к0}$ принято называть тепловыми токами переходов в транзисторах.

Подставляя токи I_1 и I_2 из (5.32) в соотношения (5.31), найдем аналитические выражения для статических ВАХ транзистора:

$$I_э = I'_{э0}\,(e^{U_э/\varphi_T} - 1) - \alpha_I I'_{к0}\,(e^{U_к/\varphi_T} - 1); \tag{5.34а}$$

$$I_к = \alpha_N I'_{э0}\,(e^{U_э/\varphi_T} - 1) - I'_{к0}\,(e^{U_к/\varphi_T} - 1). \tag{5.34б}$$

Ток базы легко записать как разность токов $I_э$ и $I_к$:

$$I_б = (1 - \alpha_N) I'_{э0}\,(e^{U_э/\varphi_T} - 1) + (1 - \alpha_I) I'_{к0}\,(e^{U_к/\varphi_T} - 1). \tag{5.34в}$$

Выражения (5.34) называются *формулами Молла—Эберса*. Они являются математической моделью транзистора и составляют основу для анализа его статических режимов.

Следует подчеркнуть, что в формулах (5.34) положительными считаются прямые напряжения (независимо от того, что у n–p–n-транзисторов фактическая полярность прямых напряжений на эмиттере и коллекторе относительно базы — отрицательная). Кроме того, необходимо иметь в виду, что параметры $I'_{э0}$ и $I'_{к0}$ в формулах (5.34) — это именно тепловые токи, а не обратные токи переходов, которые в случае кремния намного превышают тепловые. Только тогда, когда на оба перехода заданы обратные напряжения, формулы (5.34) теряют силу и обратные токи следует оценивать с учетом тока термогенерации.

Можно показать, что в транзисторах выполняется соотношение

$$\alpha_N I_{э0} = \alpha_I I_{к0}; \qquad (5.35)$$

оно позволяет упрощать выражения, полученные на основе формул (5.34).

Характеристики схемы ОБ. Как известно, для схемы ОБ (рис. 5.4, *а*) характерны заданные значения тока эмиттера и коллекторного напряжения. Поэтому характеристиками схемы ОБ называют функции $I_к$ ($I_э$; $U_к$) и $I_э$ ($U_э$; $U_к$), представленные семействами кривых. Одно из таких семейств, представляющее функцию $I_к$ ($U_к$) с параметром $I_э$ (рис. 5.14, *а*), называют *семейством выходных или коллекторных характеристик*; второе, представляющее функцию $I_э$ ($U_э$) с параметром $U_к$ (рис. 5.14, *б*) — *семейством входных или эмиттерных характеристик*. Оба семейства легко получаются из формул (5.34) и записываются в виде

$$I_к = \alpha_N I_э - I_{к0}(e^{U_к/\varphi_T} - 1); \qquad (5.36)$$

$$U_э = \varphi_T \ln\left[\frac{I_э}{I'_{э0}} + 1 + \alpha_N(e^{U_к/\varphi_T} - 1)\right]. \qquad (5.37)$$

Семейство эмиттерных кривых (5.37) записано в форме $U_э(I_э)$, поскольку заданной величиной (аргументом) является ток эмиттера.

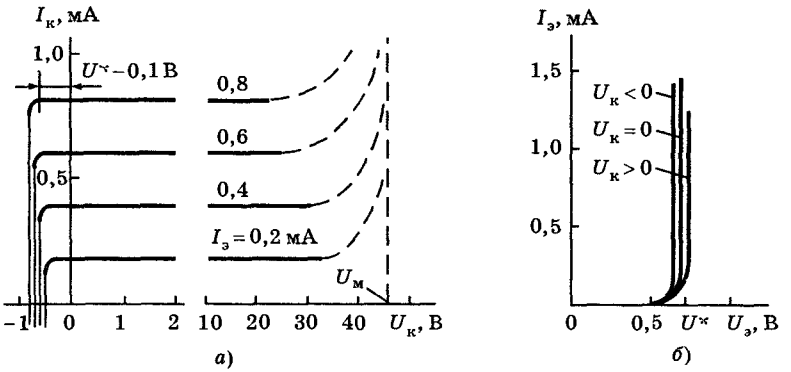


Рис. 5.14. Статические выходные (*а*) и входные (*б*) характеристики транзистора при включении ОБ

Из рис. 5.14, *а* ясно видны два резко различных режима работы транзистора: нормальный *активный режим*, соответствующий обратным напряжениям на коллекторном переходе (первый квадрант) и *режим двойной инжекции* или *насыщения*, соответствующий прямым напряжениям на коллекторном переходе (второй квадрант). Активный режим характерен для усилительных схем, а режим двойной инжекции — для ключевых (импульсных) схем.

Для активного режима формулы (5.36) и (5.37) упрощаются, так как при $|U_к| > 3\varphi_T$ исчезают экспоненциальные члены. Если, кроме того, пренебречь током $I_{к0}$ и величиной $1 - \alpha_N$, то формула (5.36) переходит в (5.15):

$$I_к = \alpha_N I_э, \qquad (5.38а)$$

а формула (5.37) — в ВАХ обычного перехода

$$U_э = \varphi_T \ln(I_э / I'_{э0}). \qquad (5.38б)$$

Из выражений (5.38) следует, что *в активном режиме коллекторное напряжение не влияет ни на выходную — коллекторную, ни на входную — эмиттерную характеристики.*

Этот вывод справедлив в большинстве практических случаев. Однако в принципе и коллекторный ток, и эмиттерное напряжение несколько зависят от коллекторного напряжения. Это значит, что выходные характеристики на рис. 5.14, *а* имеют конечный наклон (определяемый сопротивлением коллек-

торного перехода), а входные характеристики несколько смещаются при изменениях коллекторного напряжения (см. рис. 5.14, *б*). Причиной обеих зависимостей является эффект Эрли. Его влияние на наклон выходных характеристик рассмотрено в разделе 5.6. Что касается сдвига входных характеристик, то в этом случае эффект Эрли проявляется в следующем. Изменение коллекторного напряжения приводит к изменению ширины базы. Поскольку ток эмиттера, а значит, и градиент концентрации носителей заданы, изменение ширины базы приводит к изменению граничной концентрации носителей (см. рис. 5.5), а это согласно граничному условию Шокли неизбежно связано с изменением напряжения на переходе.

Поскольку (5.38б) соответствует ВАХ обычного перехода, то *в рабочем диапазоне токов напряжение $U_э$ остается почти неизменным; его можно считать параметром кремниевого транзистора U^*.* Для нормального токового режима (0,1–1 мА) $U^* \approx 0{,}7$ В, для микрорежима (1–10 мкА) $U^* \approx 0{,}5$ В. Температурная зависимость эмиттерного напряжения (как и для обычных диодов) составляет для кремниевых транзисторов от $-1{,}5$ мВ/°С до -2 мВ/°С.

Для режима двойной инжекции характерен спад коллекторного тока при неизменном эмиттерном токе. Это — результат встречной инжекции со стороны коллектора. Важно отметить, что у кремниевых транзисторов заметное уменьшение тока $I_к$ наступает не при $U_к = 0$, а при достаточно больших прямых напряжениях $U_к$. Это объясняется тем, что практическое отпирание кремниевого p–n-перехода (в данном случае коллекторного) наступает лишь при прямых напряжениях 0,4–0,6 В.

Характеристики схемы ОЭ. Как известно, для схемы ОЭ (рис. 5.4, *б*) характерен заданный ток базы. Поэтому выходные (коллекторные) характеристики представляют функцию $I_к$ ($I_б$; $U_{кэ}$), а входные (базовые) — функцию $I_б$ ($U_б$; $U_{кэ}$). Эти характеристики, которые нетрудно получить с помощью формул Молла—Эберса, показаны на рис. 5.15. Главной особенностью выходных характеристик является то, что они полностью расположены в первом квадранте.

Оценим напряжение, при котором наступает спад коллекторного тока. Запишем для режима двойной инжекции:

$$U_{кэ} = U_э - U_к, \qquad (5.39)$$

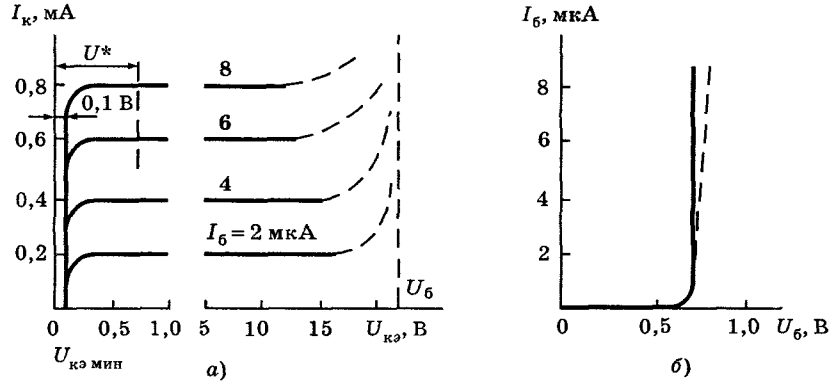


Рис. 5.15. Статические выходные (*а*) и входные (*б*) характеристики транзистора при включении ОЭ

где под $U_э$ и $U_к$ понимаются прямые напряжения. Формально границе активного режима соответствует значение $U_к = 0$. При этом согласно (5.39) выходное напряжение еще сравнительно велико — оно равно напряжению на открытом эмиттерном переходе: $U_{кэ} = U^* = 0{,}7$ В. Заметный спад тока наступает лишь тогда, когда прямое напряжение $U_к$ достигает напряжения отпирания $U^* - 0{,}1$ В. При этом выходное напряжение составляет: $U_{кэ} = U^* - (U^* - 0{,}1\ \text{В}) \approx 0{,}1$ В (рис. 5.15, *а*).

Минимальное значение выходного напряжения получается при нулевом токе коллектора (рис. 5.15, *а*). Чтобы определить величину $U_{кэ\ мин}$, разрешим систему уравнений (5.34б) и (5.34в) относительно напряжений на переходах:

$$U_э = \varphi_T \ln\left[\frac{I_б + (1-\alpha_I)I_к}{I_{э0}} + 1\right]; \qquad (5.40а)$$

$$U_к = \varphi_T \ln\left[\frac{\alpha_N I_б - (1-\alpha_N)I_к}{I_{к0}} + 1\right]. \qquad (5.40б)$$

Далее, пренебрегая единицами в квадратных скобках, подставляя напряжения $U_э$ и $U_к$ в выражение (5.39), получаем выходное напряжение в следующем общем виде:

$$U_{кэ} = \varphi_T \ln\left[\frac{\alpha_N I_б + (1-\alpha_I)I_к}{\alpha_I\alpha_N I_б - (1-\alpha_N)I_к}\right]. \qquad (5.40в)$$

Полагая $I_к = 0$, найдем минимальное выходное напряжение:

$$U_{кэ\ мин} = \varphi_T \ln(1/\alpha_I). \qquad (5.41)$$

Напряжение $U_{кэ\ мин}$ весьма мало. Так, если $\alpha_I = 0{,}5$ (этому значению соответствует $B_I = 1$), то $U_{кэ\ мин} \approx 0{,}7\varphi_T \approx 15$ мВ.

Наклон ВАХ в схеме ОЭ значительно больше, а сопротивление, характеризующее этот наклон, значительно меньше, чем в схеме ОБ. Эта особенность объясняется тем, что приращение $\Delta U_{кэ}$ частично падает на эмиттерном переходе, т.е. вызывает приращение $\Delta U_э$, соответственно получается приращение $\Delta I_э$ и дополнительное приращение тока $I_к$. В предпробойной области наклон ВАХ быстро возрастает. Напряжение пробоя в схеме ОЭ меньше, чем в схеме ОБ [см. (5.29)].

В заключение отметим важную особенность тока базы. Из выражения (5.34в) следует, что в нормальном активном режиме (т.е. при $|U_к| > 3\varphi_T$) ток базы можно записать в виде

$$I_б = I_s(e^{U_э/\varphi_T} - 1), \qquad (5.42а)$$

где $I_s = (1 - \alpha_N)I'_{э0}$. Однако фактически, с учетом рекомбинации в эмиттерном переходе и на поверхности, ток базы описывается несколько иначе (см. штриховую кривую на рис. 5.15, *б*):

$$I_б = I_s(e^{U_э/m\varphi_T} - 1), \qquad (5.42б)$$

где $m > 1$. Величину m, характеризующую отличие реального тока от идеального, называют m-фактором. Этот параметр очень удобен для оценки качества эмиттерного перехода, а вместе с тем — уровня собственных шумов, стабильности и надежности транзистора. Естественно, что m-фактор связан с показателем N в формуле (5.27), поскольку показатель N характеризует тот же круг явлений, но применительно к коэффициенту инжекции. Связь между параметрами m и N следующая: $N = m/(m-1)$. Прогресс в отношении показателя N, отмеченный выше, обусловлен уменьшением m-фактора от значения $m \approx 2$ до $m \approx 1{,}2$.

5.6. Малосигнальные эквивалентные схемы и параметры

Большому классу так называемых линейных электронных схем свойствен такой режим работы транзистора, при котором на фоне сравнительно больших постоянных токов и напряжений действуют малые переменные составляющие. Именно эти составляющие представляют в таких схемах основной интерес.

Запишем напряжения и токи в виде

$$U = U^0 + \Delta U; \quad I = I^0 + \Delta I,$$

где U^0 и I^0 — постоянные составляющие; ΔU и ΔI — переменные составляющие, много меньшие постоянных.

Постоянные и переменные составляющие анализируются и рассчитываются раздельно. При анализе постоянных составляющих, как мы уже и делали, используется нелинейная физическая модель Молла—Эберса. При анализе переменных составляющих использование нелинейной модели не имеет смысла, так как *связь между малыми приращениями определяется не самими функциями,* а их производными[1]. Поэтому для анализа переменных составляющих пользуются специальными — малосигнальными моделями (*эквивалентными схемами*), состоящими из линейных элементов. Эти элементы отображают те производные, которые связывают между собой малые приращения токов и напряжений.

Для заданного эмиттерного тока (условие, свойственное включению ОБ) малосигнальную эквивалентную схему транзистора легко получить из рис. 5.13, заменяя эмиттерный и коллекторный диоды их дифференциальными сопротивлениями. Поскольку в линейных электронных схемах режим двойной инжекции недопустим, можно исключить из схемы источник тока $\alpha_I I_2$. С другой стороны, учет сопротивления базового слоя

---

1 Например, малое приращение эмиттерного тока связано с малыми приращениями эмиттерного и коллекторного напряжений соотношением

$$\Delta I_э = \frac{\partial I_э}{\partial U_э}\Delta U_э + \frac{\partial I_э}{\partial U_к}\Delta U_к,$$

где $I_э\,(U_э; U_к)$ — функция (5.34а).

не усложняет анализа малосигнальной схемы; поэтому добавим в схему сопротивление $r_б$. Можно было бы учесть и сопротивление коллекторного слоя, но его роль оказывается несущественной. Таким образом, малосигнальная (и, добавим, низкочастотная) эквивалентная схема транзистора при заданном токе эмиттера принимает такой вид, как показано на рис. 5.16. Емкости $C_э$ и $C_к$ будут учтены позднее.

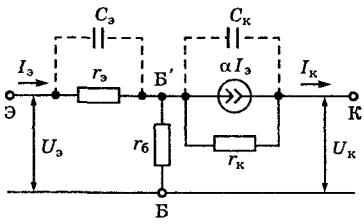


Рис. 5.16. Малосигнальная модель транзистора при включении ОБ

Положительное направление тока эмиттера выбрано произвольно, поскольку знак приращения $\Delta I_э$ может быть любым. Обозначения Δ для простоты опущены.

Заметим, что коэффициент α в малосигнальной схеме (индекс N опущен) является дифференциальным, в отличие от интегрального, которым мы пользовались до сих пор. Дифференциальный коэффициент α определяется как производная $dI_к/dI_э$, тогда как интегральный коэффициент α есть отношение $I_к/I_э$. Оба коэффициента несколько различаются, но это различие не существенно.

Дифференциальное *сопротивление эмиттерного перехода* $r_э$ выражается по аналогии с обычным диодом формулой

$$r_э = \varphi_T/I_э, \tag{5.43}$$

где $I_э$ — постоянная составляющая тока. При токе 1 мА сопротивление $r_э$ составляет 25 Ом.

Дифференциальное *сопротивление коллекторного перехода* $r_к$ обусловлено эффектом Эрли. Это сопротивление можно вычислить, подставляя в (5.15) значение $\alpha = \varkappa$ из (5.22б), дифференцируя ток $I_к$ по ширине базы w и считая $dw = -dl_к$ (приращение $dl_к$ легко связать с $dU_к$). В результате вычислений получаем:

$$r_к \approx \left(\frac{L^2}{w}\sqrt{\frac{2qN}{\varepsilon_0\varepsilon}}\right)\frac{\sqrt{U_к}}{I_э}, \tag{5.44}$$

где $U_к$ —модуль обратного напряжения. Следует обратить внимание на то, что сопротивление $r_к$, как и $r_э$, обратно пропорцио-

нально постоянной составляющей тока. Кроме того, оно несколько возрастает с увеличением напряжения, однако эта зависимость мало существенна. Для ориентировки подставим в (5.44) значения $L = 10$ мкм, $w = 1$ мкм, $N = 10^{16}$ см$^{-3}$ и $U_к = 4$ В. Тогда $r_к \approx 10^3/I_э$; при токе 1 мА получается $r_к = 1$ МОм.

Сопротивление базы $r_б$, вообще говоря, является суммой сопротивлений активной и пассивной областей базы. Расчет этих сопротивлений затрудняется сложной траекторией базового тока, сложной геометрией базового слоя, а также его неоднородностью. Типичными для планарных транзисторов можно считать значения $r_б = 50$–200 Ом.

Когда заданной исходной величиной является ток базы (при включении ОЭ), целесообразна другая эквивалентная схема (рис. 5.17), в которой источник тока в коллекторной цепи управляется током базы. Поскольку схема малосигнальная, используется не интегральный коэффициент усиления В, а дифференциальный, для которого принято особое обозначение β. Связь между малосигнальными параметрами β и α определяется по аналогии с общим выражением:

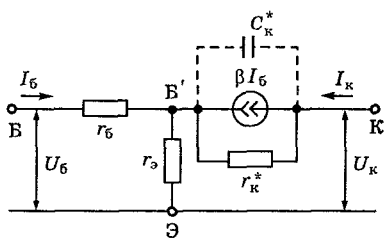


Рис. 5.17. Малосигнальная модель транзистора при включении ОЭ

$$\beta = \alpha/(1 - \alpha). \qquad (5.45)$$

В области малых токов коэффициент β несколько больше, а в области больших токов несколько меньше, чем В. В целом же значения β близки к значениям В.

Заменяя источник тока $\alpha I_э$ на $\beta I_б$, необходимо одновременно заменить сопротивление коллекторного перехода $r_к$ на на меньшую величину

$$r_к^* = (1 - \alpha) r_к = r_к/(\beta + 1). \qquad (5.46)$$

Величина $r_к^*$ получается из следующих соображений. Для того чтобы обе эквивалентные схемы были равноценны, они, как четырехполюсники, должны иметь одинаковые параметры

в режимах холостого хода и короткого замыкания. Напряжения холостого хода в схемах на рис. 5.16 и 5.17 близки соответственно к $\alpha I_э r_к$ и $\beta I_б r_к^*$. Приравнивая эти значения и учитывая, что в режиме холостого хода $I_э \approx I_б$, получаем (5.46). Причина уменьшения сопротивления в схеме ОЭ пояснялась ранее. Если выше мы получили для $r_к$ значение 1 МОм, то значение $r_к^*$ при $\beta = 100$ составит всего 10 кОм.

5.7. Переходные и частотные характеристики

Инерционность транзистора при быстрых изменениях входных токов обусловлена пролетом инжектированных носителей через базу, а также перезарядом барьерных емкостей эмиттерного и коллекторного переходов. Относительная роль этих факторов зависит от ширины базы, режима транзистора и сопротивления внешних цепей.

Рассмотрим сначала процессы в базе, пренебрегая влиянием емкостей. Роль последних будет рассмотрена позже. Кроме того, при анализе переходных процессов всегда пренебрегают сопротивлением коллекторного перехода $r_к$.

Процессы в базе при включении ОБ. Пусть транзистор включен по схеме ОБ и пусть сначала эмиттерный ток равен нулю, а на коллектор подано постоянное обратное напряжение. При этом в цепи коллектора протекает ничтожный ток термогенерации, которым можно пренебречь. Транзистор, как говорят, находится в *режиме отсечки*. В некоторый момент времени зададим скачком эмиттерный ток $I_э$ (рис. 5.18). Для простоты положим $\gamma = 1$, т.е. пренебрежем дырочной составляющей эмиттерного тока.

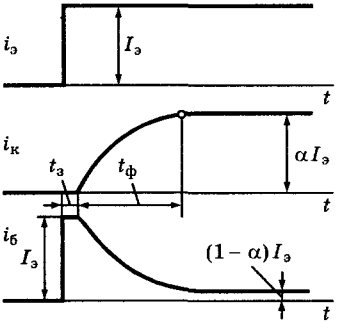


Рис. 5.18. Переходные процессы при включении ОБ

Инжектированные электроны распространяются вглубь базы постепенно. Они достигают коллектора только через время $t_з$, которое называется *временем задержки*. После этого коллекторный ток начинает нарастать, но нарастает тоже постепенно, потому что скорость диффузии — величина средняя. У отдельных носите-

лей скорости существенно различаются, так что носители, попавшие в базу одновременно, доходят до коллектора за разное время. В результате фронт коллекторного тока оказывается плавным и имеет конечную длительность $t_ф$.

При постоянном токе эмиттера функцию $i_к(t)$ удобно записать в виде $i_к = \alpha(t)I_э$, где $\alpha(t)$ — *переходная характеристика коэффициента* α. Именно она является предметом анализа переходных процессов в схеме ОБ. Параметром, определяющим длительность этих процессов, является постоянная времени τ_α. О ней подробно сказано ниже.

В интервале $t_з$, когда коллекторный ток еще отсутствует, ток базы равен току эмиттера $I_э$. Затем, по мере нарастания коллекторного тока, ток базы уменьшается до установившегося значения $(1 - \alpha)I_э$. Получается характерный *выброс* базового тока.

Одновременно с током коллектора нарастают избыточные заряды в базе. В качестве 1-го приближения (оно практически оправдано) примем, что коллекторный ток и избыточные заряды нарастают по экспоненциальному закону с постоянной времени τ_α. Моделью такого переходного процесса является процесс заряда конденсатора в простейшей *RC*-цепочке под действием ступеньки тока (рис. 5.19). Установившееся значение заряда в такой цепочке имеет следующий вид:

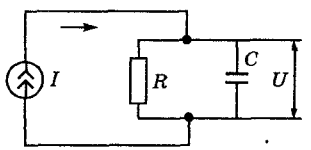


Рис. 5.19. *RC*-цепочка, моделирующая процесс накопления заряда в базе

$$Q = CU = C(IR) = I\tau,$$

где $\tau = RC$ — постоянная времени. Такую же форму имеют выражения (5.8) и (5.12б) для установившихся избыточных зарядов в базе.

Отсюда следует полезный вывод: *постоянную времени* τ_α *можно получить, поделив установившееся значение избыточного заряда* $\Delta Q_б$ *на заданный ток эмиттера* $I_э$.

Введем следующие обозначения для величин в правых частях (5.8) и (5.12б):

$$t_D = w^2/(2D); \tag{5.47}$$

$$t_{пр} = w^2/[2(\eta + 1)D] = t_D/(\eta + 1). \tag{5.48}$$

Величину t_D называют средним *временем диффузии*, а величину $t_{пр}$ — *временем пролета*. Время пролета является обобщением времени диффузии на случай, когда в базе есть ускоряющее поле. При условии $\eta = 0$ время пролета делается равным времени диффузии.

Таким образом, приняв экспоненциальное приближение для переходных характеристик, получаем

$$\tau_\alpha = t_{пр} \tag{5.49а}$$

или, в случае бездрейфовых транзисторов,

$$\tau_\alpha = t_D. \tag{5.49б}$$

Операторное изображение для коэффициента α запишется в форме

$$\alpha(\mathbf{s}) = \alpha/(1 + \mathbf{s}\tau_\alpha), \tag{5.50}$$

т.е. переходная характеристика коэффициента передачи тока имеет вид:

$$\alpha(t) = \alpha(1 - e^{-t/\tau_\alpha}). \tag{5.51}$$

Более строгое выражение может быть получено подстановкой в

$$\varkappa = \operatorname{sech}\frac{w}{L}$$

вместо L величины $L(\mathrm{s})$

$$L(s) = \frac{L}{\sqrt{1 + s\tau}}.$$

Оно же будет изображением $\alpha(\mathbf{s})$, поскольку выше было принято $\gamma = 1$. Переходная характеристика, соответствующая такому изображению, показана на рис. 5.20 штриховой линией. Ее аналитическое выражение слишком сложно для практических применений. Поэтому при подстановке $L(\mathbf{s})$ воспользуемся аппроксимацией (5.22а). Тогда

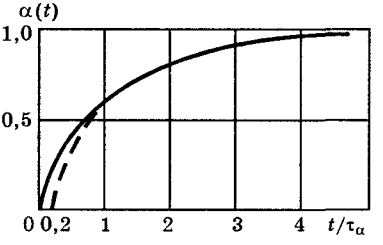


Рис. 5.20. Переходные характеристики коэффициента передачи α

изображение $\alpha(s)$ совпадает с (5.50). Это подтверждает, что экспоненциальная аппроксимация вполне пригодна на практике.

Недостаток аппроксимации (5.50) состоит лишь в том, что она не отражает наличия задержки $t_з$ (см. рис. 5.18). В тех случаях, когда задержка существенна, пользуются более точным изображением

$$\alpha(s) = \alpha e^{-st_з}/(1 + s\tau_\alpha). \tag{5.52}$$

Входящие сюда параметры имеют следующие значения:

$$\tau_\alpha \approx 0{,}8t_{пр}, \tag{5.53а}$$

$$t_з \approx 0{,}2t_{пр}. \tag{5.53б}$$

Оригиналом изображения (5.52) является экспоненциальная функция (5.51), сдвинутая относительно момента $t = 0$ на интервал $t_з$.

Частотные характеристики коэффициента α получим, заменяя в выражениях (5.50) или (5.52) оператор s на $j\omega$:

$$\dot{\alpha} = \frac{\alpha}{1 + j(\omega/\omega_\alpha)} \tag{5.54а}$$

или

$$\dot{\alpha} = \frac{\alpha e^{-j\omega t_з}}{1 + j(\omega/\omega_\alpha)}, \tag{5.54б}$$

где $\omega_\alpha = 1/\tau_\alpha$ — угловая *граничная частота.*

Следует иметь в виду, что комплексная величина α может использоваться только в малосигнальной эквивалентной схеме (рис. 5.16), т.е. при анализе переменных составляющих. Полные токи в транзисторе не могут быть синусоидальными из-за выпрямляющих свойств p–n-перехода.

Амплитудно-частотные характеристики коэффициента α, представленные выражениями (5.54), имеют одинаковую форму:

$$\alpha(\omega) = \frac{\alpha}{\sqrt{1 + (\omega/\omega_\alpha)^2}}, \tag{5.55}$$

хотя значения ω_α несколько разные [см. (5.49а) и (5.53а)].

Фазочастотные характеристики оказываются существенно разными. Для выражения (5.54а)

$$\varphi(\omega) = -\text{arctg}(\omega/\omega_\alpha), \qquad (5.56а)$$

т.е. фазовый сдвиг ограничен [$\varphi(\infty) = -90°$], а $\varphi(\omega_\alpha) = -45°$. Для выражения (5.54б)

$$\varphi(\omega) = -\omega t_з - \text{arctg}(\omega/\omega_\alpha). \qquad (5.56б)$$

В этом случае фазовый сдвиг не ограничен [$\varphi(\infty) = -\infty$], а $\varphi(\omega_\alpha) = -59°$. Выражение (5.56б) гораздо более точное, тогда как (5.56а) дает большую погрешность уже при частотах, меньших граничной. Частотные характеристики коэффициента α показаны на рис. 5.21.

Процессы в базе при включении ОЭ. Пусть транзистор включен по схеме ОЭ и задана ступенька базового тока $I_б$ (рис. 5.22). При этом функцию $\imath_к(t)$ можно записать в виде $\imath_к(t) = B(t)\, I_б$, где $B(t)$ — *переходная характеристика коэффициента усиления* В. Она и ее постоянная времени $\tau_в$ являются в

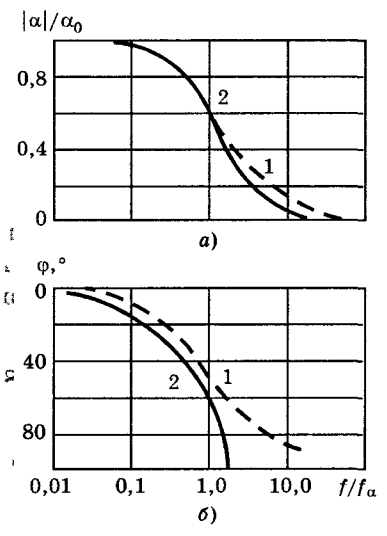


Рис. 5.21. Частотные характеристики коэффициента передачи α: *а* — амплитудно-частотные; *б* — фазочастотные; *1* — аппроксимация (5.54а); *2* — аппроксимация (5.54б)

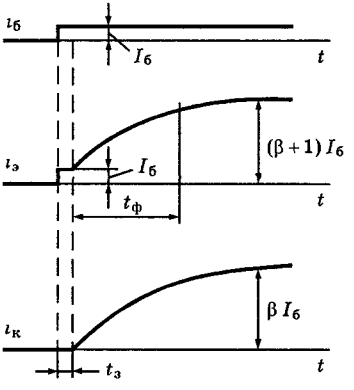


Рис. 5.22. Переходные процессы при включении ОЭ

данном случае предметом анализа. Для простоты опять будем считать $\gamma = 1$.

Операторное изображение B(s) легко получить, подставляя $\alpha(s)$ в выражение $B = \alpha/(1 - \alpha)$. При этом легко убедиться, что переходный процесс сохраняет экспоненциальный характер, но постоянная времени τ_B будет гораздо больше, а именно

$$\tau_B = \tau_\alpha/(1 - \alpha) = (B + 1)\tau_\alpha. \qquad (5.57)$$

Для того чтобы выяснить физический смысл величины τ_B, проведем следующие рассуждения.

Ток $I_б$ задает скорость нарастания положительного заряда в базе. Значит, в момент поступления ступеньки $I_б$ в базе начинает повышаться концентрация дырок. Соответственно повышается потенциал базы и это вызывает отпирание эмиттерного перехода. Начинается инжекция электронов, заряд которых поддерживает квазинейтральность базы. Значит, в первый момент имеет место равенство $I_э = I_б$, как и в схеме ОБ.

Через время задержки $t_з$ инжектированные носители доходят до коллектора и появляется коллекторный ток. В схеме ОБ нарастание коллекторного тока сопровождалось уменьшением тока базы. Однако в схеме ОЭ базовый ток задан, поэтому возрастание коллекторного тока (связанное с уходом электронов из базы) вызывает дополнительное нарастание тока эмиттера (т.е. приток новых электронов, необходимых для поддержания квазинейтральности).

Такое совместное увеличение токов $I_к$ и $I_э$ продолжается до тех пор, пока в базе не накопится настолько большой избыточный заряд $\Delta Q_б$, что скорость его рекомбинации уравновесит ток базы. Условие равновесия имеет вид:

$$\Delta Q_б \approx I_б\tau. \qquad (5.58)$$

Отсюда, руководствуясь тем, что $\Delta Q_б/I_б$ есть постоянная времени переходного процесса, приходим к выводу: *в схеме ОЭ постоянная времени равна времени жизни носителей в базе*

$$\tau_B = \tau. \qquad (5.59)$$

Итак, операторный коэффициент усиления в схеме ОЭ имеет вид:

$$B(s) = B/(1 + s\tau_B), \qquad (5.60)$$

где постоянная времени τ_B в десятки и более раз превышает величину τ_α. Можно показать, что *большой коэффициент усиления* B *в схеме ОЭ достигается ценой ухудшения переходных и частотных свойств*. На фоне большой постоянной времени τ_B задержка $t_з$ не существенна и ее не учитывают.

Заменяя в выражении (5.60) B на β и s на $j\omega$, получаем малосигнальную частотную характеристику

$$\dot{\beta} = \frac{\beta}{1 + j(\omega/\omega_\beta)}, \qquad (5.61)$$

где $\omega_\beta = 1/\tau_\beta$ — верхняя граничная частота (можно считать $\tau_\beta = \tau_B$). Соответственно амплитудно- и фазочастотная характеристики будут иметь вид:

$$\beta(\omega) = \frac{\beta}{\sqrt{1 + (\omega/\omega_\beta)^2}}; \qquad (5.62а)$$

$$\varphi(\omega) = -\text{arctg}(\omega/\omega_\beta). \qquad (5.62б)$$

Поскольку коэффициент β весьма велик, усилительная способность транзистора сохраняется при частотах, значительно превышающих граничную частоту ω_β. При $\omega > 3\omega_\beta$ в выражении (5.62а) можно пренебречь единицей, тогда

$$\beta(\omega) \approx \beta\omega_\beta/\omega.$$

Полагая $\beta(\omega) = 1$, найдем частоту, при которой коэффициент β уменьшается до единицы и транзистор теряет усилительные свойства:

$$\omega_T \approx \beta\omega_\beta = \beta/\tau. \qquad (5.63)$$

Частоту ω_T называют *предельной частотой коэффициента усиления тока*. Учитывая соотношение (5.57), приходим к выводу, что предельная частота практически совпадает с граничной частотой ω_α.

Влияние барьерных емкостей. Начнем с влияния емкости эмиттерного перехода (рис. 5.16). Поскольку барьерная емкость обусловлена изменениями ширины перехода, т.е. перемещениями основных носителей, ток перезаряда этой емкости не имеет отношения к инжекции и не входит в состав коллектор-

ного тока. Следовательно, та часть тока эмиттера, которая ответвляется в барьерную емкость, приводит к уменьшению коэффициента инжекции.

Распределение тока эмиттера между барьерной емкостью и переходом зависит от соотношения сопротивлений этих двух цепей. Ограничимся переменными составляющими. Тогда за сопротивление перехода следует принять величину $r_э$, а за сопротивление емкости ее операторный импеданс $1/(sC_э)$. Запишем часть тока $I_э$, которая ответвляется в p–n-переход:

$$I_{эр-п} = I_э \frac{1/(sC_э)}{r_э + 1/(sC_э)}.$$

Именно эта величина подразумевалась в определении коэффициента инжекции под током $I_э$. Заменяя ток $I_э$ на $I_{эр-п}$, нетрудно получить

$$\gamma(s) = I_{эп}/I_э = \gamma/(1 + s\tau_\gamma), \qquad (5.64)$$

где

$$\tau_\gamma = r_э C_э \qquad (5.65)$$

— постоянная времени эмиттерного перехода, она же — *постоянная времени коэффициента инжекции*.

Если $r_э = 25$ Ом и $C_э = 1$ пФ, то $\tau_\gamma = 0{,}025$ нс. В этом случае обычно $\tau_\gamma \ll t_{пр}$, так что постоянную времени τ_γ можно не учитывать или учитывать в виде дополнительной задержки. Однако с уменьшением тока сопротивление $r_э$ растет и величина τ_γ становится сравнимой с $t_{пр}$. Поэтому в микрорежиме с определенным приближением можно считать

$$\tau_\alpha \approx \tau_\gamma + t_{пр}. \qquad (5.66)$$

Тем самым *роль эмиттерной барьерной емкости сводится к увеличению постоянной времени* τ_α.

Теперь рассмотрим роль коллекторной барьерной емкости $C_к$ (рис. 5.16). Если осуществить на выходе короткое замыкание и по-прежнему пренебречь сопротивлением $r_к$, то емкость $C_к$ окажется соединенной параллельно с сопротивлением $r_б$. Постоянная времени такой цепочки называется *постоянной времени базы*:

$$\tau_б = r_б C_к. \qquad (5.67)$$

Например, если $r_б = 100$ Ом и $C_к = 1$ пФ, то $\tau_б = 0{,}1$ нс. Ток $\alpha I_э$ распределяется между внешней цепью (куда вошло сопротивление $r_б$) и емкостью $C_к$. Следовательно, на высоких частотах внешний ток $I_к$ будет всегда меньше, чем $\alpha I_э$. В частности, если положить $\tau_\alpha = 0$, то именно *постоянная времени базы ограничивает предельное быстродействие транзистора.*

Если в цепь коллектора включено сопротивление $R_к$, то в предыдущих рассуждениях нужно заменить сопротивление $r_б$ на $r_б + R_к$. Обычно $R_к >> r_б$, поэтому инерционность токораспределения в коллекторной цепи будет определяться не постоянной времени $\tau_б$, а постоянной времени $C_к R_к$. Несмотря на то, что величина $C_к R_к$ не является параметром транзистора, (она зависит от внешнего элемента $R_к$), ее удобно рассматривать как составную часть параметра τ_α. Для этого введем понятие *эквивалентной постоянной времени* $\tau_{\alpha oe}$, которую по аналогии с (5.66) определим в виде суммы

$$\tau_{\alpha oe} = \tau_\alpha + C_к R_к . \qquad (5.68)$$

Если $C_к R_к > 3\ \tau_\alpha$ (что часто бывает), то величина $\tau_{\alpha oe}$ практически не зависит от процессов в базе.

Для схемы ОЭ эквивалентную постоянную времени τ_{oe} найдем с помощью соотношения (5.57), умножив обе части (5.68) на B + 1. Тогда

$$\tau_{oe} = \tau_B + C_к^* R_к , \qquad (5.69)$$

где

$$C_к^* = (B + 1) C_к \qquad (5.70)$$

— эквивалентная емкость коллекторного перехода (рис. 5.17).

Постоянные времени $\tau_{\alpha oe}$ и τ_{oe} — наиболее универсальные параметры, характеризующие инерционность биполярного транзистора.

5.8. Тиристор

Тиристор представляет собой четырехслойную структуру *p–n–p–n* (рис. 5.23, *а*), вольтамперная характеристика которой имеет участок отрицательного сопротивления (рис. 5.23, *б*). Появление участка отрицательного сопротивления качественно

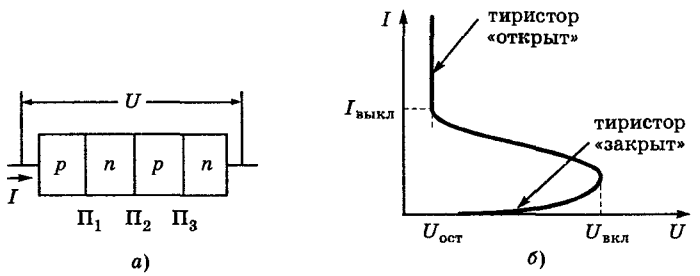


Рис. 5.23. Четырехслойная p–n–p–n-структура (*а*) и ее ВАХ (*б*)

можно объяснить, используя двухтранзисторную модель тиристора (рис. 5.24). Действительно p–n–p–n-структура может быть представлена как соединение двух биполярных транзисторов p–n–p и n–p–n-типа. Обозначим коэффициент усиления p–n–p-транзистора α_p, а n–p–n-транзистора — α_n. Ток центрального перехода тиристора $П_2$ складывается из токов коллектирования p–n–p и n–p–n-транзисторов. Если ток во внешней цепи равен I, то до центрального перехода доходит ток $\alpha_p I + \alpha_n I = (\alpha_p + \alpha_n)I$. По условию непрерывности тока суммарный ток через центральный переход должен быть равен I. Если сумма коэффициентов передачи $\alpha_p + \alpha_n < 1$, то через переход $П_2$ должен протекать дополнительный ток, совпадающий по направлению с током во внешней цепи и равный $(1 - \alpha_p - \alpha_n)I$. Такое протекание тока соответствует обратному смещению перехода $П_2$ и закрытому состоянию тиристора. Если сумма коэффициентов передачи больше единицы ($\alpha_p + \alpha_n > 1$), то через центральный переход должен протекать дополнительный ток $(\alpha_p + \alpha_n - 1)I$, но направление этого тока противоположно направлению тока во внешней цепи. Это соответствует прямому смещению центрального перехода и открытому состоянию тиристора. При этом на тиристоре падает напряжение, примерно равное напряжению одного *обычного* p–n-перехода. Это напряжение называется остаточным напряжением $U_{ост}$. Обычно $U_{ост} \approx (0{,}7–1)$ В. Известно, что коэффициенты усиления биполярных транзисторов увеличива-

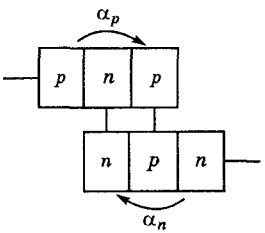


Рис. 5.24. Двухтранзисторная модель тиристора

ются при увеличении тока эмиттера. Поэтому при возрастании тока во внешней цепи тиристора сумма коэффициентов передачи увеличивается от значения $\alpha_p + \alpha_n < 1$ до значения $\alpha_p + \alpha_n > 1$ и напряжение на тиристоре сначала возрастает, а затем уменьшается. Ток выключения $I_{выкл}$ соответствует условию $\alpha_p + \alpha_n = 1$.

Если в область базы n–p–n-транзистора втекает внешний управляющий ток, то условие равенства суммы коэффициентов передачи единице (условие $\alpha_p + \alpha_n = 1$) выполняется при меньших токах во внешней цепи (рис. 5.25, *а*). Это приводит к уменьшению тока выключения и напряжения включения. Вольтамперные характеристики тиристора для нескольких значений тока управления показаны на рис. 5.25, *б*. При некотором значении тока управления, называемым током управления спрямления $(I_{упр})_{спр}$, участок отрицательного сопротивления исчезает и тиристор при любом токе находится в открытом состоянии.

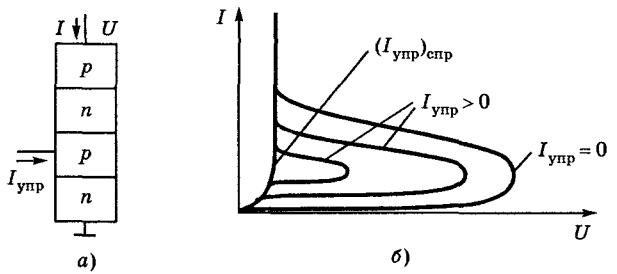


Рис. 5.25. Управление тиристором по p-базе (*а*) и ВАХ для разных токов управления (*б*)

Рассмотрим включение тиристора в цепь с источником напряжения E и нагрузкой $R_н$ (рис. 5.26, *а*). Предположим, что линия нагрузки пересекает ВАХ тиристора в трех точках (рис. 5.26, *б*). В закрытом состоянии на тиристоре падает напряжение питания E, так как ток через тиристор очень мал. Очевидно, что в закрытом состоянии должно выполняться условие $E < U_{вкл}$. В открытом состоянии ток через тиристор в основном определяется величиной сопротивления нагрузки

$$I_{отк} = \frac{E - U_{ост}}{R_н} \approx \frac{E}{R_н},$$

так как обычно $U_{ост} << E$.

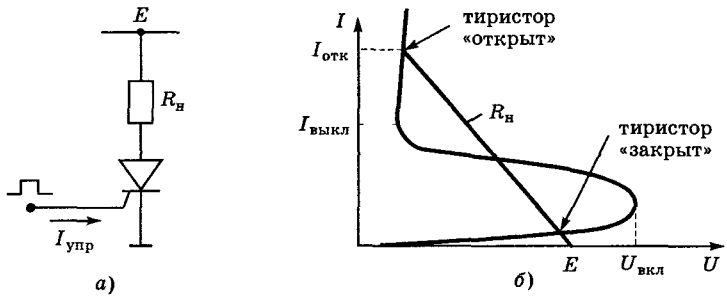


Рис. 5.26. Включение тиристора во внешнюю цепь (*а*), ВАХ тиристора и линия нагрузки (*б*)

Переход из закрытого состояния в открытое может произойти при подаче в цепь управления импульса управляющего тока. После прекращения возбуждения по цепи управления тиристор остается в состоянии «открыто».

Именно эффект открывания тиристора является причиной, почему этот прибор рассматривается в курсе «Основы микроэлектроники». Как самостоятельный прибор тиристор используется в микроэлектронных структурах относительно редко (он применяется в качестве ключа в мощных специализированных микросхемах малой степени интеграции). Однако структура p–n–p–n очень часто формируется как паразитная структура, при изготовлении МОП- и биполярных микросхем. На рис. 5.27 показано поперечное сечение обычного КМОП-инвертора. Сток p-канального транзистора обычно подключается к источнику питания, а исток n-канального транзистора — к земле. Нетрудно видеть, что между питанием и землей имеется паразитная четырехслойная p^+–n–p–n^+-структура, которая должна находиться в закрытом состоянии. Если этот паразитный тиристор откроется, например, при воздействии помехи, то через структуру потечет очень большой ток, так как сопротивление нагрузки в этом случае близко к нулю. Это может привести к выгоранию структуры, т.е. выходу из строя микросхемы, так как после открывания паразитный тиристор остается в этом состоянии бесконечно долго. Описанный эффект носит название эффекта «защелки» и для его устранения используется комплекс конструктивно-технологических и схемотехнических мер. Конструктивно-технологические решения направлены на уменьшение коэффициентов усиления (снижение времени жизни,

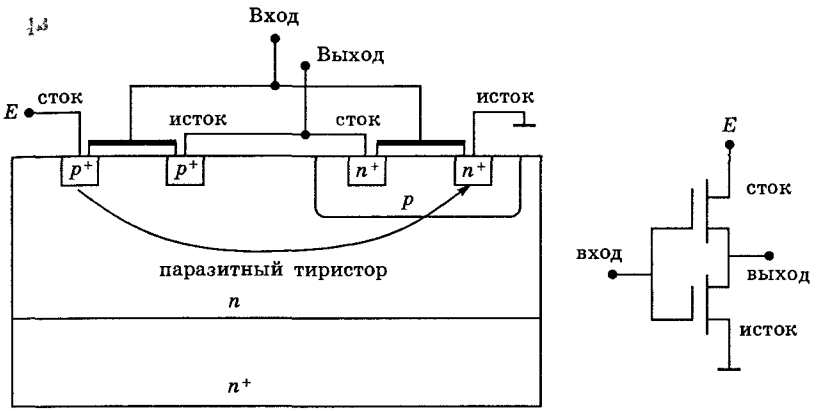


Рис. 5.27. Поперечное сечение КМОП-инвертора

выбор соответствующей топологии), с тем чтобы их сумма $\alpha_p + \alpha_n$ не превышала единицы при любых токах. В этом случае паразитный тиристор не может устойчиво находиться в открытом состоянии. Схемотехнические меры сводятся к подавлению влияния электрических помех.

Контрольные вопросы

1. Чем отличаются бездрейфовые и дрейфовые биполярные транзисторы?
2. Нарисуйте зонную диаграмму n–p–n-транзистора в равновесном состоянии.
3. Какое смещение имеют эмиттерный и коллекторный переходы при нормальном включении?
4. Назовите основные схемы включения биполярных транзисторов.
5. Что такое режим двойной инжекции?
6. Как распределены избыточные носители в базе бездрейфового транзистора при нормальном включении?
7. Как влияет коэффициент неоднородности базы на распределение электронов в n–p–n дрейфовом транзисторе?
8. Нарисуйте распределение избыточных носителей в базе и коллекторе n–p–n бездрейфового транзистора при инверсном включении.

9. Через какой параметр связаны ток базы и избыточный заряд в базе?
10. Дайте определение коэффициенту переноса и коэффициенту инжекции.
11. Как связаны коэффициенты усиления эмиттерного и базового тока?
12. Как связан коэффициент переноса с шириной базы?
13. Как влияет уровень легирования эмиттера на величину коэффициента инжекции?
14. Объясните зависимость коэффициента усиления от тока эмиттера.
15. Объясните зависимость коэффициента усиления от коллекторного напряжения.
16. Как изменяется коэффициент усиления с увеличением температуры?
17. В чем заключается эффект Эрли?
18. Что такое напряжение смыкания?
19. Сравните наклон выходных характеристик транзистора при включении ОЭ и ОБ.
20. Как смещаются входные характеристики транзистора в схеме ОБ при увеличении напряжения на коллекторе?
21. Как смещаются входные характеристики транзистора в схеме ОЭ при увеличении напряжения на коллекторе?
22. Напишите выражение для дифференциального сопротивления эмиттерного перехода.
23. Каков порядок величины сопротивления базы?
24. Какая постоянная времени определяет инерционность транзистора в схеме ОБ?
25. Какая постоянная времени определяет инерционность транзистора в схеме ОЭ?
26. Почему ВАХ тиристора имеет участок отрицательного сопротивления?
27. Что такое эффект «защелки» в КМОП-инверторе?

Глава 6

Технологические основы микроэлектроники

6.1. Введение

Технология полупроводниковых ИС развилась на основе планарной технологии транзисторов, а последняя, в свою очередь, впитала в себя весь предшествующий опыт производства полупроводниковых приборов. Поэтому чтобы разбираться в технологических циклах изготовления ИС, необходимо ознакомиться с типовыми технологическими процессами, из которых эти циклы складываются. Технология ГИС также зародилась не на пустом месте, а обобщила и развила те методы нанесения пленок, которые ранее использовались в радиотехнической промышленности, машиностроении и оптике.

6.2. Подготовительные операции

Монокристаллические слитки кремния, как и других полупроводников, получают обычно путем кристаллизации из расплава — *методом Чохральского*. При этом методе стержень с затравкой (в виде монокристалла кремния) после соприкосновения с расплавом медленно поднимают с одновременным вращением (рис. 6.1). При этом вслед за затравкой вытягивается нарастающий и застывающий слиток.

Кристаллографическая ориентация слитка (его поперечного сечения) определяется кристаллографической ориентацией затравки. Чаще других используются слитки с поперечным сечением, лежащим в плоскости (111) или (100) (см. раздел 2.2).

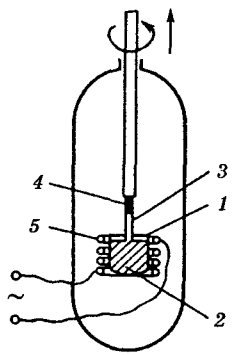


Рис. 6.1. Схема выращивания монокристалла методом Чохральского: *1* — тигель; *2* — расплав полупроводника; *3* — выращиваемый монокристалл; *4* — затравка; *5* — катушка высокочастотного нагрева

Типовой диаметр слитков составляет в настоящее время 150 мм, а максимальный может достигать 300 мм и более. Длина слитков может достигать 3 м и более, но обычно она в несколько раз меньше.

Слитки кремния разрезают на множество тонких пластин (толщиной 0,4–0,5 мм), на которых затем изготавливают интегральные схемы или другие приборы. Во время резки слиток прочно закрепляют, причем очень важно обеспечить перпендикулярное расположение слитка относительно режущих полотен или дисков с тем, чтобы пластины имели необходимую кристаллографическую ориентацию.

Поверхность пластин после резки весьма неровная: размеры царапин, выступов и ямок намного превышают размеры будущих элементов ИС. Поэтому перед началом основных технологических операций пластины многократно шлифуют, а затем полируют. Цель шлифовки, помимо удаления механических дефектов, состоит в том, чтобы обеспечить необходимую толщину пластины (150–250 мкм), недостижимую при резке, и параллельность плоскостей. Шлифовку осуществляют на вращающихся шлифовальных кругах. Шлифующим агентом являются суспензии из микропорошков, размер зерен которых выбирают все меньшим при каждом цикле шлифовки, вплоть до 0,3–0,5 мкм.

По окончании шлифовки на поверхности все же остается механически нарушенный слой толщиной несколько микрон, под которым расположен еще более тонкий, так называемый физически нарушенный слой. Последний характерен наличием «незримых» искажений кристаллической решетки и механических напряжений, возникающих в процессе шлифовки.

Полировка состоит в удалении обоих нарушенных слоев и снижении неровностей поверхности до уровня, свойственного оптическим системам — сотые доли микрона. Помимо механической (с помощью еще более мелкозернистых суспензий), используется химическая полировка (травление), т.е. по существу растворение поверхностного слоя полупроводника в тех или иных реактивах. Выступы и трещины на поверхности стравливаются быстрее, чем основной материал, и в целом поверхность выравнивается.

Достигаемая в процессе шлифовки и полировки параллельность плоскостей пластины составляет единицы и даже доли микрона на сантиметр длины.

Важным процессом в полупроводниковой технологии является также очистка поверхности от загрязнений органическими веществами, особенно жирами. Очистку и обезжиривание проводят в органических растворителях (толуол, ацетон, этиловый спирт и др.) при повышенной температуре.

Травление, очистка и многие другие процессы сопровождаются отмывкой пластин в *деионизованной* воде. Деионизация осуществляется в специальных установках путем пропускания предварительно дистиллированной воды через гранулированные смолы, в которых благодаря химическим реакциям происходит связывание растворенных ионов. Степень деионизации оценивается по удельному сопротивлению воды, которое обычно лежит в пределах 10–20 МОм · см и выше (удельное сопротивление бидистиллированной воды не превышает 1–2 МОм · см).

6.3. Эпитаксия

Эпитаксией называют процесс наращивания монокристаллических слоев на подложку, при котором кристаллографическая ориентация наращиваемого слоя повторяет кристаллографическую ориентацию подложки.

В настоящее время эпитаксия обычно используется для получения тонких рабочих слоев однородного полупроводника на сравнительно толстой подложке, играющей роль несущей конструкции.

Типовой — *хлоридный* процесс эпитаксии применительно к кремнию состоит в следующем (рис. 6.2). Монокристаллические кремниевые пластины загружают в тигель «лодочку» и помещают в кварцевую трубу. Через трубу пропускают поток водорода, содержащий небольшую примесь тетрахлорида кремния $SiCl_4$. При высокой температуре (около 1200 °С), которая обеспечивается высокочастотным нагревом тигля, на поверхности пластин происходит реакция

$$SiCl_4 + 2H_2 = Si + 4HCl.$$

В результате реакции на подложке постепенно осаждается слой чистого кремния, а пары HCl уносятся потоком водорода. Эпитаксиальный слой осажденного кремния монокристалличен и имеет ту же кристаллографическую ориентацию, что и подложка. Химическая реакция, благодаря подбору температуры,

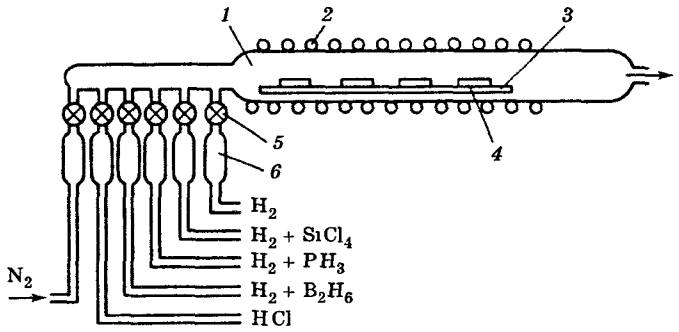


Рис. 6.2. Схема хлоридного процесса эпитаксии: *1* — кварцевая труба; *2* — катушка ВЧ нагрева; *3* — тигель с пластинами; *4* — пластина кремния; *5* — вентиль для перекрытия соответствующего газа; *6* — измеритель скорости потока

происходит только на поверхности пластины, а не в окружающем пространстве.

Процесс, проходящий в потоке газа, называют *газотранспортной реакцией,* а основной газ (в данном случае водород), переносящий примесь в зону реакции, — *газом-носителем.*

Если к парам тетрахлорида кремния добавить пары соединений бора (B_2H_6) или фосфора (PH_3), то эпитаксиальный слой будет иметь уже не собственную, а соответственно дырочную или электронную проводимость, поскольку в ходе реакции в осаждающийся кремний будут внедряться акцепторные атомы бора или донорные атомы фосфора.

В установке, показанной на рис. 6.2, предусмотрены некоторые дополнительные операции: продувка трубы азотом и неглубокое травление поверхности кремния в парах HCl (с целью очистки). Эти операции проводятся до основных.

Таким образом, эпитаксия позволяет выращивать монокристаллические слои любого типа проводимости и любого удельного сопротивления на подложке, обладающей тоже любым типом и величиной проводимости (рис 6.3).

Эпитаксиальная пленка может отличаться от подложки по химическому составу. Способ получения таких пленок называют *гетероэпитаксией,* в отличие от *гомоэпитаксии,* описанной выше. Конечно, при гетероэпитаксии материалы пленки и подложки должны по-прежнему иметь одинаковую кристаллическую решетку. Например, можно выращивать кремниевую пленку на сапфировой подложке.

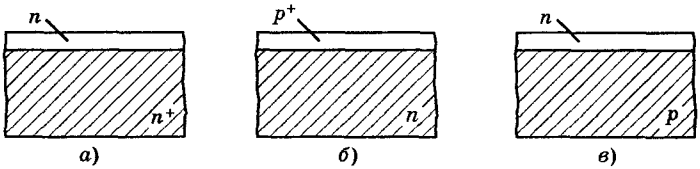


Рис. 6.3. Примеры эпитаксиальных структур:
а — пленка *n*-типа на n^+-подложке; *б* — пленка p^+-типа на *n*-подложке; *в* — пленка *n*-типа на *p*-подложке

Граница между эпитаксиальным слоем и подложкой не получается идеально резкой, так как примеси в процессе эпитаксии частично диффундируют из одного слоя в другой. Это обстоятельство затрудняет создание сверхтонких (менее 1 мкм) и многослойных эпитаксиальных структур. Основную роль в настоящее время играет однослойная эпитаксия. Она существенно пополнила арсенал полупроводниковой технологии; получение таких тонких однородных слоев (1–10 мкм), какие обеспечивает эпитаксия, невозможно иными средствами.

В заключение заметим, что помимо описанной газовой эпитаксии, существует жидкостная эпитаксия, при которой наращивание монокристаллического слоя осуществляется из жидкой фазы, т.е. из раствора, содержащего необходимые компоненты.

6.4. Термическое окисление

Окисление кремния — один из самых характерных процессов в технологии современных ИС. Получаемая при этом пленка двуокиси кремния (SiO_2) выполняет несколько важных функций, в том числе:

- функцию защиты — *пассивации* поверхности и, в частности, защиты вертикальных участков *p–n*-переходов, выходящих на поверхность (рис. 6.4, *а*);
- функцию маски, через окна которой вводятся необходимые примеси (рис. 6.4, *б*);
- функцию тонкого диэлектрика под затвором МОП-транзистора (рис. 6.4, *в*).

Такие широкие возможности двуокиси кремния — одна из причин того, что кремний стал основным материалом для изготовления полупроводниковых ИС.

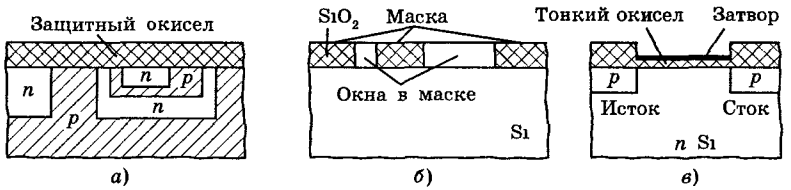


Рис 6 4. Функции двуокисной пленки кремния
а — пассивация поверхности, *б* — маска для локального легирования, *в* — тонкий подзатворный окисел

Поверхность кремния всегда покрыта «собственной» окисной пленкой, получающейся в результате «естественного» окисления при самых низких температурах. Однако эта пленка имеет слишком малую толщину (около 5 нм), чтобы выполнять какую-либо из перечисленных функций. Поэтому в технологии ИС пленки SiO_2 получают искусственным путем.

Искусственное окисление кремния осуществляется обычно при высокой температуре (1000–1200 °С). Такое термическое окисление можно проводить в атмосфере кислорода (*сухое окисление*), в смеси кислорода с парами воды (влажное окисление) или просто в парах воды[1].

Во всех случаях процесс проводится в окислительных печах. Основу таких печей составляет, как и при эпитаксии, кварцевая труба, в которой размещается «лодочка» с пластинами кремния, нагреваемая либо токами высокой частоты, либо иным путем. Через трубу пропускается поток кислорода (сухого или увлажненного) или пары воды, которые реагируют с кремнием в высокотемпературной зоне. Получаемая таким образом пленка SiO_2 имеет аморфную структуру.

Механизм окисления имеет два варианта. Первый вариант состоит из следующих этапов: 1) диффузия атомов кремния через уже имеющуюся пленку окисла к поверхности, 2) адсорбция молекул кислорода поверхностью из газовой фазы, 3) собственно окисление, т.е. химическая реакция. В этом случае пленка нарастает над исходной поверхностью кремния. Второй вариант состоит из следующих этапов: 1) адсорбция кислорода поверхностью уже имеющегося окисла, 2) диффузия кислорода

1 При влажном окислении кислород предварительно пропускают через сосуд с водой — так называемый барботер

через окисел к еще не окисленному кремнию, 3) собственно окисление. В этом случае пленка нарастает вглубь от исходной поверхности кремния. На практике оба механизма сочетаются, но главную роль обычно играет второй.

Очевидно, что скорость роста окисла со временем должна убывать, так как новым атомам кислорода приходится диффундировать через все более толстый слой окисла. Полуэмпирическая формула, связывающая толщину окисной пленки со временем термического окисления, имеет вид:

$$d \approx k\sqrt{t},$$

где k — параметр, зависящий от температуры и влажности кислорода.

Сухое окисление идет в десятки раз медленнее влажного. Например, для выращивания пленки SiO_2 толщиной 0,5 мкм в сухом кислороде при 1000 °С требуется около 5 ч, а во влажном — всего 20 мин. С уменьшением температуры на каждые 100 °С время окисления возрастает в 2–3 раза.

В технологии ИС различают «толстые» и «тонкие» окислы SiO_2. Толстые окислы (d = 0,5–0,8 мкм) выполняют функции защиты и маскировки, а тонкие (d = 0,05–0,15 мкм) — функции подзатворного диэлектрика в МОП-транзисторах.

Одной из важных проблем при выращивании пленки SiO_2 является обеспечение ее однородности. В зависимости от качества поверхности пластины, от чистоты реагентов и режима выращивания в пленке возникают те или иные *дефекты*. Распространенным типом дефектов являются микро- и макропоры, вплоть до сквозных отверстий (особенно в тонком окисле).

Качество окисной пленки повышается с уменьшением температуры ее выращивания, а также при использовании сухого кислорода. Поэтому тонкий подзатворный окисел, от качества которого зависит стабильность параметров МОП-транзистора, получают сухим окислением. При выращивании толстого окисла чередуют сухое и влажное окисление: первое обеспечивает отсутствие дефектов, а второе позволяет сократить время процесса.

6.5. Легирование

Внедрение примесей в исходную пластину (или в эпитаксиальный слой) путем диффузии при высокой температуре является исходным и до сих пор основным способом легирования полупроводников с целью создания диодных и транзисторных структур. Этому способу мы уделим главное внимание. Однако за последние годы широкое распространение получил и другой способ легирования — ионная имплантация, — который рассматривается в конце раздела.

Способы диффузии. Диффузия может быть общей и локальной. В первом случае она осуществляется по всей поверхности пластины (рис. 6.5, *а*), а во втором — на определенных участках пластины через окна в маске, например, в слое SiO_2 (рис. 6.5, *б*).

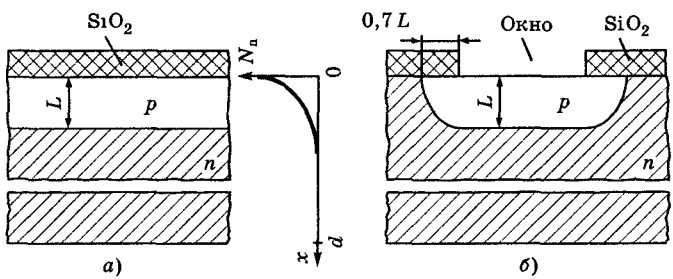


Рис. 6.5. Общая (*а*) и локальная (*б*) диффузия примеси в кремний

Общая диффузия приводит к образованию в пластине тонкого диффузионного слоя, который отличается от эпитаксиального неоднородным (по глубине) распределением примеси (см. кривую $N_a(x)$ на рис. 6.5).

В случае локальной диффузии примесь распространяется не только вглубь пластины, но и во всех перпендикулярных направлениях, т.е. под маску. В результате этой так называемой *боковой диффузии* участок p–n-перехода, выходящий на поверхность, оказывается «автоматически» защищенным окислом (рис. 6.5, *б*). Соотношение между глубинами боковой и основной — «вертикальной» — диффузии зависит от ряда факторов, в том числе от глубины диффузионного слоя L. Типичным для глубины боковой диффузии можно считать значение 0,7 L.

Диффузию можно проводить однократно и многократно. Например, в исходную пластину n-типа можно во время 1-й диффузии внедрить акцепторную примесь и получить p-слой , а затем во время 2-й диффузии внедрить в полученный p-слой (на меньшую глубину) донорную примесь и тем самым обеспечить трехслойную структуру. Соответственно различают двойную и тройную диффузию.

При проведении многократной диффузии следует иметь в виду, что концентрация каждой новой вводимой примеси должна превышать концентрацию предыдущей, в противном случае тип проводимости не изменится, а значит, не образуется p–n-переход. Между тем концентрация примеси в кремнии (или другом исходном материале) не может быть сколь угодно большой: она ограничена особым параметром — *предельной растворимостью примеси.* Предельная растворимость зависит от температуры. При некоторой температуре она достигает максимального значения $N_{\text{пр макс}}$, а затем снова уменьшается. Максимальные предельные растворимости вместе с соответствующими температурами приведены в табл. 6.1.

Таблица 6.1. Максимальная предельная растворимость типичных примесей в кремнии

Примесь	As	P	B	Sb
$N_{\text{пр}}$, см$^{-3}$	$20 \cdot 10^{20}$ (1150 °C)	$13 \cdot 10^{20}$ (1150 °C)	$5 \cdot 10^{20}$ (1200 °C)	$0{,}6 \cdot 10^{20}$ (1300 °C)

Следовательно, если проводится многократная диффузия, то для последней диффузии нужно выбирать материал с максимальной предельной растворимостью. Поскольку ассортимент примесных материалов ограничен, не удается обеспечить более 3-х последовательных диффузий.

Примеси, вводимые путем диффузии, называют *диффузантами* (бор, фосфор и др.). Источниками диффузантов являются их химические соединения. Это могут быть и жидкости (BBr_3, POCl), и твердые тела (B_2O_3, P_2O_5), и газы (B_2H_6, PH_3).

Внедрение примесей обычно осуществляется с помощью газотранспортных реакций — так же, как при эпитаксии и окислении. Для этого используются либо однозонные, либо двухзонные *диффузионные печи.*

Двухзонные печи используются в случае твердых диффузантов. В таких печах (рис. 6.6) имеются две высокотемпературные зоны, одна — для испарения источника диффузанта, вторая — собственно для диффузии. Пары источника диффузанта, полученные в 1-й зоне, примешиваются к потоку нейтрального газа-носителя (например, аргона) и вместе с ним доходят до 2-й зоны, где расположены пластины кремния. Температура во 2-й зоне выше, чем в 1-й. Здесь атомы диффузанта внедряются в пластины, а другие составляющие химического соединения уносятся газом-носителем из зоны.

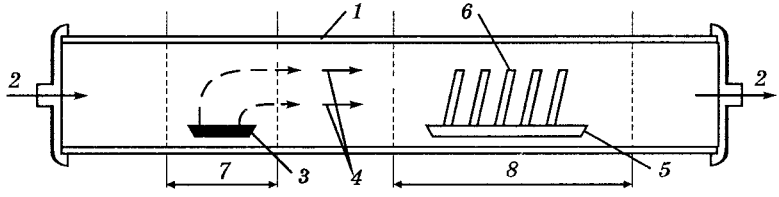


Рис. 6.6. Схема двухзонной диффузионной печи: 1 — кварцевая труба; 2 — поток газа-носителя; 3 — источник диффузанта; 4 — пары источника диффузанта; 5 — тигель с пластинами; 6 — пластина кремния; 7 — первая высокотемпературная зона; 8 — вторая высокотемпературная зона

В случае жидких и газообразных источников диффузанта нет необходимости в их высокотемпературном испарении. Поэтому используются однозонные печи, в которые источник диффузанта поступает уже в газообразном состоянии.

При использовании жидких источников диффузанта диффузию проводят в окислительной среде, добавляя к газу-носителю кислород. Кислород окисляет поверхность кремния, образуя окисел SiO_2, т.е. в сущности — стекло. В присутствии диффузанта (бора или фосфора) образуется *боросиликатное* или *фосфорно-силикатное стекло*. При температуре выше 1000° эти стекла находятся в жидком состоянии, покрывая поверхность кремния тонкой пленкой, так что диффузия примеси идет, строго говоря, из жидкой фазы. После застывания стекло защищает поверхность кремния в местах диффузии, т.е. в окнах окисной маски. При использовании твердых источников диффузанта — окислов — образование стекол происходит в процессе диффузии без специально вводимого кислорода.

Теоретические основы диффузии. Теория диффузии основана на двух *законах Фика*. 1-й закон Фика связывает плотность

потока частиц J с градиентом их концентрации. В одномерном случае

$$J = -D\,(\mathrm{d}N/\mathrm{d}x), \tag{6.1а}$$

где D — коэффициент диффузии, N — концентрация.

2-й закон Фика характеризует скорость накопления частиц (в нашем случае — атомов примеси):

$$\partial N/\partial t = D\,(\partial^2 N/\partial x^2). \tag{6.1б}$$

Из уравнения (6.1б) можно найти функцию $N\,(x,\,t)$, т.е. *распределение концентрации* $N(x)$ в любой момент времени. Для этого нужно задаться двумя граничными условиями.

Пусть координата $x = 0$ соответствует той плоскости пластины, через которую вводится примесь (рис. 6.5). Тогда координата противоположной плоскости равна толщине пластины d. На практике глубина диффузионных слоев всегда меньше толщины пластины (см. рис. 6.5): поэтому можно положить $N(d) = 0$. С математической точки зрения удобнее считать пластину бесконечно толстой и в качестве 1-го граничного условия принять

$$N(\infty, t) = 0. \tag{6.2}$$

2-е граничное условие имеет два варианта, которые соответствуют двум разновидностям реального технологического процесса.

1) *Случай неограниченного источника примеси.* В этом случае диффузант непрерывно поступает к пластине, так что в ее приповерхностном слое концентрация примеси поддерживается постоянной.

Граничное условие для этого варианта имеет вид:

$$N\,(0,t) = N_s = \mathrm{const}, \tag{6.3а}$$

где N_s —*поверхностная* (точнее — приповерхностная) концентрация. Обычно количество поступающего диффузанта обеспечивает режим предельной растворимости, т.е. $N_s = N_{\text{пр макс}}$.

2) *Случай ограниченного источника примеси.* В этом случае сначала в тонкий приповерхностный слой пластины вводят некоторое количество атомов диффузанта, а потом источник диффузанта отключают и атомы примеси перераспределяются по

глубине пластины при неизменном их общем количестве. Первую стадию процесса называют «загонкой», вторую — «разгонкой» примеси.

Для этого варианта можно записать условие в виде

$$\int_0^\infty N(x)\,dx = Q = \text{const}, \tag{6.3б}$$

где Q — количество атомов примеси на единицу площади (задается на этапе «загонки»).

Решая уравнение (6.1б) при граничных условиях (6.2) и (6.3а), получаем распределение концентрации при неограниченном источнике (рис. 6.7, *а*):

$$N(x, t) = N_s \operatorname{erfc}(x/2\sqrt{Dt}), \tag{6.4а}$$

где erfc (z) — *дополнительная функция ошибок*, близкая к экспоненциальной функции e^{-z}.

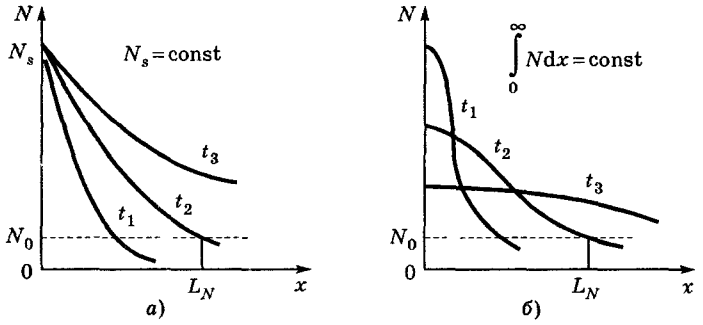


Рис. 6.7. Распределение примеси при диффузии из неограниченного (*а*) и ограниченного (*б*) источников для разных моментов времени

Решая уравнение (6.1б) при условиях (6.2) и (6.3б), получаем распределение концентрации при ограниченном источнике (рис. 6.7, *б*):

$$N(x, t) = \frac{Q}{\sqrt{\pi}\sqrt{Dt}} e^{-x^2/(4Dt)}. \tag{6.4б}$$

В данном случае распределение описывается *функцией Гаусса*, которая характерна нулевой начальной производной, нали-

чием точки перегиба и почти экспоненциальным «хвостом» после этой точки.

Под глубиной диффузионного слоя (*глубиной диффузии*) понимают координату $x = L_N$, при которой концентрация введенной примеси N равна концентрации исходной примеси N_0 (рис. 6.7). Величину L_N нетрудно найти из выражений (6.4), полагая в левой части $N = N_0$.

Аппроксимируя функцию (6.4а) экспонентой, получаем для неограниченного источника

$$L_N \approx 2\sqrt{Dt}\ln(N_s/N_0);$$

логарифмируя обе части (6.4б), получаем для ограниченного источника

$$L_N = 2\sqrt{Dt}\ln\frac{Q}{N_0\sqrt{\pi Dt}}.$$

Оба выражения имеют одинаковую структуру и позволяют сделать два важных общих вывода:

- *время проведения диффузии пропорционально квадрату желательной глубины диффузии*; поэтому получение глубоких диффузионных слоев требует большого времени; в ИС глубина рабочих диффузионных слоев обычно лежит в пределах 1–4 мкм;
- при заданной глубине диффузионного слоя *изменения коэффициента диффузии эквивалентны изменениям времени процесса.*

Второй вывод заслуживает более детального рассмотрения. На рис. 6.8 показана температурная зависимость коэффициентов диффузии для некоторых материалов, применяемых в технологии ИС. Как видим, эта зависимость экспоненциальная, т.е. весьма сильная: при $\Delta T = 100$ °С коэффициент диффузии меняется на порядок, а при $\Delta T = \pm 1$ °С — на ± 2,5%.

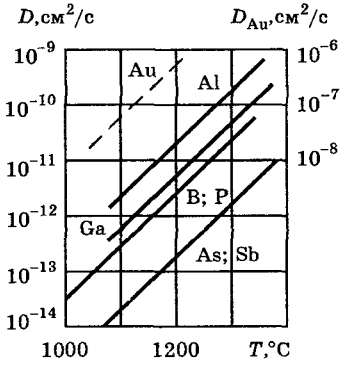


Рис. 6.8. Зависимость коэффициентов диффузии типичных (для кремния) примесей от температуры. Для золота масштаб на 3 порядка меньше

Последняя цифра, казалось бы, невелика, но ее истинное значение можно проиллюстрировать следующим простым расчетом. Если $\Delta D/D = 2{,}5\%$, то разброс глубины диффузии составит $\pm$ 1,25% или, при $L_N = 4$ мкм, около $\pm$ 0,05 мкм. Значит, ширина базы w, равная $L_{N6} - L_{N\text{э}}$, может получиться с ошибкой 0,1 мкм, что при $w = 0{,}5$ мкм составит 20%. Поскольку коэффициент β и предельная частота f_T обратно пропорциональны величине w^2, их разброс превысит 40%.

Из сказанного ясна необходимость прецизионной регулировки температуры в диффузионных печах. Допустимая нестабильность температуры составляет до $\pm$ 0,2 °С (т.е. сотые доли процента).

Ионная имплантация. Ионной имплантацией называют метод легирования пластины (или эпитаксиального слоя) путем бомбардировки ионами примеси, ускоренными до энергии, достаточной для их внедрения вглубь твердого тела.

Ионизация атомов примеси, ускорение ионов и фокусировка ионного пучка осуществляются в специальных установках типа ускорителей частиц в ядерной физике. В качестве примесей используются те же материалы, что и при диффузии.

Глубина внедрения ионов зависит от их энергии и массы. Чем больше энергия, тем больше получается толщина имплантированного слоя. Однако с ростом энергии возрастает и количество *радиационных дефектов* в кристалле, т.е. ухудшаются его электрофизические параметры. Поэтому энергию ионов ограничивают величиной 100–150 кэВ. Нижний уровень составляет 5–10 кэВ. При таком диапазоне энергии глубина слоев лежит в пределах 0,1–0,4 мкм, т.е. она значительно меньше типичной глубины диффузионных слоев.

Концентрация примеси в имплантированном слое зависит от плотности тока в ионном пучке и времени процесса или, как говорят, от *времени экспозиции*. В зависимости от плотности тока и желательной объемной концентрации время экспозиции составляет от нескольких секунд до 3-5 мин и более (иногда до 1–2 ч). Разумеется, чем больше время экспозиции, тем, опять же, больше количество радиационных дефектов.

Типичное распределение примеси при ионной имплантации показано на рис. 6.9 сплошной кривой. Как видим, это распределение существенно отличается от диффузионного наличием максимума. Вблизи максимума кривая хорошо аппроксимируется функцией Гаусса, см. (6.4б).

Поскольку площадь ионного пучка меньше площади пластины (а иногда и кристалла), то приходится *сканировать* пучок, т.е. плавно или «шагами» перемещать его (с помощью специальных отклоняющих систем) поочередно по всем «строкам» пластины, на которых расположены отдельные ИС.

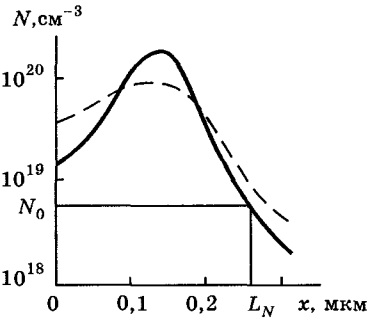


Рис. 6.9. Распределение примесей при ионной имплантации

По завершении процесса легирования пластину обязательно подвергают *отжигу* при температуре 500–800 °С для того, чтобы упорядочить кристаллическую решетку кремния и устранить (хотя бы частично) неизбежные радиационные дефекты. При температуре отжига процессы диффузии несколько меняют профиль распределения (см. штриховую кривую на рис. 6.9).

Ионная имплантация так же, как диффузия, может быть общей и локальной (избирательной). В последнем, более типичном случае облучение (бомбардировка) проводится через маски, в которых длина пробега ионов должна быть значительно меньше, чем в кремнии. Материалом для масок могут служить распространенные в ИС двуокись кремния и алюминий. При этом важным достоинством ионной имплантации является то, что ионы, двигаясь по прямой линии, внедряются только вглубь пластины, а аналогия боковой диффузии (под маску) практически отсутствует.

В принципе ионную имплантацию, как и диффузию, можно проводить многократно, «встраивая» один слой в другой. Однако сочетание энергий, времен экспозиции и режимов отжига, необходимое для многократной имплантации, оказывается затруднительным. Поэтому ионная имплантация получила главное распространение при создании тонких одинарных слоев.

Главными преимуществами ионной имплантации являются: низкая температура процесса и его хорошая контролируемость. Низкая температура обеспечивает возможность проведения ионной имплантации на любом этапе технологического цикла, не вызывая при этом дополнительной диффузии примесей в ранее изготовленных слоях.

6.6. Травление

Обычно травление ассоциируется с использованием специальных растворов — травителей для общего или локального удаления поверхностного слоя твердого тела на ту или иную глубину. Действительно, жидкие травители остаются главным средством для достижения указанной цели. Однако в технологии микроэлектроники появились и другие средства, выполняющие ту же задачу. Поэтому в общем случае травление можно рассматривать как не механические способы изменения рельефа поверхности твердого тела.

Классический процесс химического травления состоит в химической реакции жидкого травителя с твердым телом с образованием растворимого соединения; последнее смешивается с травителем и в дальнейшем удаляется вместе с ним. Переход поверхностного слоя твердого тела в раствор означает удаление этого слоя. Однако, в отличие от механического удаления, травление обеспечивает гораздо большую прецизионность процесса: стравливание происходит плавно — один мономолекулярный слой за другим. Подбирая травитель, его концентрацию, температуру и время травления, можно весьма точно регулировать толщину удаляемого слоя. Например, при химической полировке пластины кремния (см. раздел 6.2), используя соответствующий травитель, можно обеспечить скорость травления 0,1 мкм/мин, т.е. за 20–30 с снять слой толщиной всего 40–50 нм.

Для большей равномерности травления и удаления продуктов реакции с поверхности ванночку с раствором вращают в наклонном положении (*динамическое травление*) или вводят в раствор ультразвуковой вибратор (*ультразвуковое травление*).

Конечно, травление подчиняется законам физической химии, но в реальных условиях имеется столько привходящих обстоятельств, что рецептура травителей для каждого материала подбирается не расчетным путем, а экспериментально.

Характерной особенностью локального травления (через защитную маску) является так называемое *подтравливание* (рис. 6.10, *а*) — эффект, в какой-то мере аналогичный боковой диффузии (рис.6.5, *б*). Он выражается в том, что травление идет не только вглубь пластины, но и в стороны — под маску. В результате стенки вытравленного рельефа оказываются не со-

всем вертикальными, а площадь углубления — несколько больше площади окна в маске.

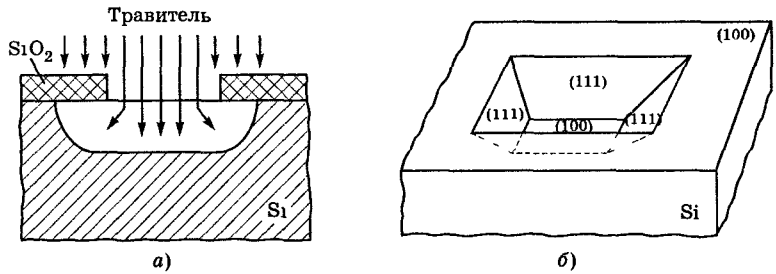


Рис. 6.10. Локальное травление кремния:
а — изотропное; *б* — анизотропное

Электролитическое травление отличается тем, что химическая реакция жидкости с твердым телом и образование растворимого соединения происходят в условиях протекания тока через жидкость, причем твердое тело играет роль одного из электродов — анода. Значит, твердое тело в данном случае должно обладать достаточной электропроводностью, что, конечно, ограничивает круг используемых материалов. Преимуществом электролитического травления является возможность регулировать скорость травления путем изменения тока в цепи и прекращать процесс путем его отключения.

Так называемое *ионное травление* (один из специфических процессов в микроэлектронике) не связано с использованием жидкостей. Пластина кремния помещается в разреженное пространство, в котором, невдалеке от пластины, создается тлеющий разряд. Пространство тлеющего разряда заполнено *квазинейтральной электронно-ионной плазмой*. На пластину относительно плазмы подается достаточно большой отрицательный потенциал. В результате положительные ионы плазмы бомбардируют поверхность пластины и слой за слоем выбивают атомы с поверхности, т.е. *травят* ее[1]. Аналогичным способом достигается очистка поверхности от загрязнений — *ионная очистка*. Структура ионно-плазменных установок описана в разделе 6.9.

---

1 Напряжение, свойственное ионному травлению (2–3 кэВ), значительно меньше ускоряющих напряжений при ионной имплантации, поэтому внедрение ионов в пластину не происходит.

Ионное травление, как и химическое, может быть общим и локальным. Несомненным преимуществом локального ионного травления является отсутствие «подтравливания» под маску: стенки вытравленного рельефа практически вертикальны, а площади углублений равны площади окон в маске.

Общее преимущество ионного травления заключается в его универсальности (не требуется индивидуального кропотливого подбора травителей для каждого материала), а общий недостаток — в необходимости дорогостоящих установок и значительных затрат времени на создание в них нужного вакуума.

За последние годы разработаны и широко используются методы так называемого *анизотропного травления*. Эти методы основаны на том, что скорость химической реакции, лежащей в основе классического травления, зависит от кристаллографического направления. Наименьшая скорость свойственна направлению (111), в котором плотность атомов на единицу площади максимальна (рис. 2.2), а наибольшая — направлению (100), в котором плотность атомов минимальна. Поэтому при использовании специальных анизотропных травителей скорость травления оказывается разной в разных направлениях и боковые стенки лунок приобретают определенный рельеф — *огранку*. Пример огранки при травлении в плоскости (100) показан на рис. 6.10, *б*. Как видим, в данном случае травление идет параллельно плоскостям (111), поскольку в направлении (111), перпендикулярном этой плоскости, скорость травления намного меньше, чем в других направлениях.

Углы, под которыми вытравливаются боковые стенки лунок, строго определены и поддаются расчету [например, на рис. 6.10, *б* угол между плоскостями (100) и (111) составляет 61,5°]. Поэтому вместе с методом масок метод анизотропного травления дает разработчику ИС возможность проектировать рельеф отверстий не только по плоскости, но и по глубине.

Тот факт, что плоскость (111) как бы «непроницаема» для травителя, обеспечивает еще одно преимущество анизотропного травления: если края окон в маске ориентированы по осям (100), то отсутствует явление подтравливания, свойственное изотропному травлению (рис. 6.10, *а*). Соответственно при анизотропном травлении наружные размеры лунок могут практически совпадать с размерами окон в маске.

6.7. Техника масок

В технологии полупроводниковых приборов важное место занимают маски: они обеспечивают локальный характер напыления, легирования, травления, а в некоторых случаях и эпитаксии. Всякая маска содержит совокупность заранее спроектированных отверстий — окон. Изготовление таких окон есть задача *литографии* (гравировки). Ведущее место в технологии изготовления масок сохраняет *фотолитография*, которой мы уделим главное внимание.

Фотолитография. В основе фотолитографии лежит использование материалов, которые называют *фоторезистами*. Это разновидность фотоэмульсий, известных в обычной фотографии. Фоторезисты чувствительны к ультрафиолетовому свету, поэтому их можно обрабатывать в не очень затемненном помещении.

Фоторезисты бывают негативные и позитивные. Негативные фоторезисты под действием света полимеризуются и становятся устойчивыми к травителям (кислотным или щелочным). Значит, после локальной засветки будут вытравливаться незасвеченные участки (как в обычном фотонегативе). В позитивных фоторезистах свет, наоборот, разрушает полимерные цепочки и, значит, будут вытравливаться засвеченные участки.

Рисунок будущей маски изготавливается в виде так называемого *фотошаблона* (рис. 6.11). Фотошаблон представляет собой толстую стеклянную пластину, на одной из сторон которой нанесена тонкая непрозрачная пленка с необходимым *рисунком* в виде прозрачных отверстий. Размеры этих отверстий (элементов рисунка) в масштабе 1:1 соответствуют размерам будущих элементов ИС, т.е. могут составлять как десятки и единицы микронов, так и доли микрона (в случае т. н. субмикронной технологии). Поскольку ИС изготавливаются групповым методом (см. раздел 1.2), на фотошаблоне по «строкам» и «столбцам» размещается множество однотипных рисунков. Размер каждого рисунка соответствует размеру будущего кристалла ИС.

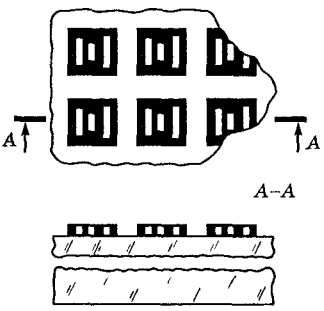


Рис. 6.11. Фрагмент фотошаблона (в плане и в разрезе)

Процесс фотолитографии для получения окон в окисной маске SiO_2, покрывающей поверхность кремниевой пластины, состоит в следующем (рис. 6.12).

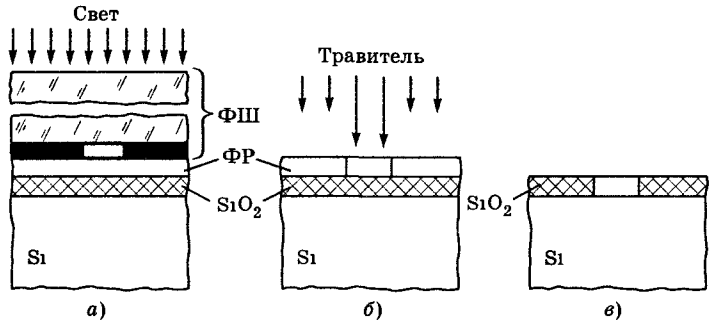


Рис. 6.12. Этапы процесса фотолитографии. *а* — экспозиция фоторезиста через фотошаблон; *б* — локальное травление двуокиси кремния через фоторезистную маску, *в* — окисная маска после удаления фоторезиста

На окисленную поверхность пластины наносится капля фоторезиста (ФР). С помощью центрифуги каплю распределяют тонким слоем (около 1 мкм) по всей поверхности. Полученную пленку фоторезиста высушивают до затвердевания. На пластину, покрытую фоторезистом, накладывают фотошаблон ФШ (рисунком к фоторезисту) и экспонируют его в лучах кварцевой лампы (рис. 6.12, *а*). После этого фотошаблон снимают.

Если используется позитивный фоторезист (см. выше), то после проявления и закрепления (которое состоит в задублировании и термообработке фоторезиста) в нем получаются окна на тех местах, которые соответствуют прозрачным участкам на фотошаблоне[1]. Как говорят, мы *перенесли рисунок* с фотошаблона на фоторезист. Теперь слой фоторезиста представляет собой маску, плотно прилегающую к окисному слою (рис. 6.12, *б*).

Через фоторезистную маску производится травление окисного слоя вплоть до кремния (на кремний данный травитель не действует). В результате рисунок с фторезиста переносится на

1 Если используется негативный фоторезист, то и фотошаблон должен быть негативным, т.е. будущим отверстиям в окисной маске должны соответствовать непрозрачные участки

окисел. После удаления (стравливания) фоторезистной маски конечным итогом фотолитографии оказывается кремниевая пластина, покрытая окисной маской (рис. 6.12, *в* и 6.4, *б*). Через окна в этой маске можно осуществлять диффузию, ионную имплантацию, травление и т.п.

В технологических циклах изготовления диодов, транзисторов и тем более ИС процесс фотолитографии используется многократно (отдельно для получения базовых слоев, эмиттеров, омических контактов и т.д.). При этом возникает так называемая проблема совмещения фотошаблонов. Суть этой проблемы иллюстрируется на рис. 6.13.

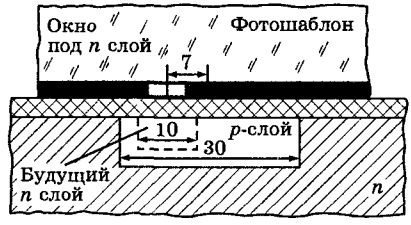


Рис. 6.13. Совмещение фотошаблона с рисунком на поверхности ИС

Пусть с помощью предыдущей фотолитографии и диффузии в пластине выполнен p-слой шириной 30 мкм, а с помощью следующей фотолитографии и диффузии нужно внутри p-слоя получить n-слой шириною 10 мкм (показан штрихами), смещенный на 7 мкм относительно центра p-слоя. Для этого рисунок 2-го фотошаблона необходимо совместить с уже существующим рельефом (т.е. с границами p-слоя) с точностью не менее 0,1 от минимального размера изображения.

При многократном использовании фотолитографии (в технологии ИС до 15–20 раз) допуск на совмещение доходит до сотых долей микрона. Техника совмещения состоит в том, что на фотошаблонах делают специальные «отметки» (например, крестики или квадраты), которые переходят в рисунок на окисле и просвечивают сквозь тонкую пленку фоторезиста. Накладывая очередной фотошаблон, аккуратнейшим образом (под микроскопом) совмещают отметки на окисле с аналогичными отметками на фотошаблоне.

Рассмотренный процесс фотолитографии характерен для получения окисных масок на кремниевых пластинах с целью последующей локальной диффузии. В этом случае фоторезистная маска (рис. 6.12, *б*) является промежуточной, вспомогательной, так как она не выдерживает высокой температуры, при которой проводится диффузия. Однако в некоторых случаях, когда процесс идет при низкой температуре, фоторезистные маски могут быть основными — рабочими. Примером может слу-

жить процесс создания металлической разводки в полупроводниковых ИС (см. раздел 6.9).

Фотошаблоны. Первым этапом процесса изготовления фотошаблонов является конструирование и вычерчивание послойной топологии фотошаблонов. Ранее, когда размеры элементов топологии составляли единицы микрон, а количество элементов на кристалле не превышало десятков тысяч, фотошаблоны вычерчивались на специальных координатографах в масштабе 100:1 и 300:1. В настоящее время при субмикронных размерах элементов и их количестве на кристалле до десятков миллионов, фотошаблоны изготавливаются только с помощью компьютерных комплексов, использующих методы электронно-лучевой литографии в масштабах 1:1 и 4:1. При этом объем данных при проектировании уже в настоящее время достигает 0,5–1 Гбайт, а в перспективе 10–100 Гбайт.

Следующим этапом является так называемый промежуточный отсъем оригинала, т.е. его фотографирование на стеклянную фотопластинку с необходимым уменьшением размеров и, если необходимо, — мультиплицированием (см. рис. 6.11). Редуцирование размеров при финальном отсъеме зависит от масштаба при промежуточном отсъеме. Мультиплицирование осуществляется на специальных фотоштампах, «размножающих» изображение кристалла на всю подложку фотошаблона размером до 150×150 мм^2.

Фотошаблоны изготавливаются комплектами по числу операций фотолитографии в технологическом цикле. В пределах каждого комплекта фотошаблоны согласованы, т.е. обеспечивают взаимную ориентацию и совмещение с заданной точностью.

Одним из слабых мест классической фотолитографии является механический контакт фотошаблона с пластиной, покрытой фоторезистом. Такой контакт никогда не может быть совершенным и сопровождается разного рода искажениями рисунка. Конкурирующим методом является *проекционная фотолитография*, при которой рисунок фотошаблона проектируется на пластину с помощью специальной оптической системы.

Новые решения и тенденции. Описанные выше процессы до сих пор не потеряли своего значения в качестве одной из основ микроэлектронной технологии. Однако по мере повышения степени интеграции и уменьшения размеров элементов ИС возник ряд проблем, которые частично уже решены, а частично находятся в стадии изучения.

Одно из принципиальных ограничений фотолитографии касается разрешающей способности, т.е. минимальных размеров в создаваемом рисунке маски. Из-за дифракции света минимальный размер изображения на кристалле (при длине волны ультрафиолетового источника засветки фотошаблона 0,5–0,2 мкм) не может быть с допустимой точностью +10% менее 1,0–0,4 мкм. Между тем уже в настоящее время при создании больших и сверхбольших ИС такие размеры элементов оказываются недостаточно малыми.

Наиболее очевидный путь для повышения разрешающей способности литографии — использование при экспозиции более коротковолновых излучений, например, мягкого рентгеновского (с длинами волн 1–2 нм). Одной из наиболее трудноразрешимых проблем рентгенолитографии является создание рентгеношаблонов с масштабом М1:1 и рентгенорезистов. Другой тяжелой задачей является поиск адекватных методов совмещения шаблонов.

За последние годы разработаны методы *электронной литографии*. Их сущность состоит в том, что сфокусированный пучок электронов *сканируют* (т.е. перемещают «построчно») по поверхности пластины, покрытой резистом, и управляют интенсивностью пучка в соответствии с заданной программой. В тех точках, которые должны быть «засвечены», ток пучка максимален, а в тех, которые должны быть «затемнены», — минимален или равен нулю. Диаметр пучка электронов находится в прямой зависимости от тока в пучке: чем меньше диаметр, тем меньше ток. Однако с уменьшением тока растет время экспозиции. Поэтому повышение разрешающей способности (уменьшение диаметра пучка) сопровождается увеличением длительности процесса.

Одна из разновидностей электронной литографии основана на отказе от резисторных масок и предусматривает воздействие электронного пучка непосредственно на окисный слой SiO_2. Оказывается, что в местах «засветки» этот слой в дальнейшем травится в несколько раз быстрее, чем в «затемненных» участках.

Что касается проблемы совмещения рисунков, то ее стараются решать путем *самосовмещения*. Этот принцип можно охарактеризовать как использование ранее полученных структурных элементов в качестве масок для получения последующих

элементов. Примерами могут служить изопланарная технология (рис. 7.10) и технология МОП-транзисторов с самосовмещенным затвором (рис. 7.30 и 7.31).

Реальный прогресс и ближайшие перспективы промышленных методов литографии показаны на рис. 6.14.

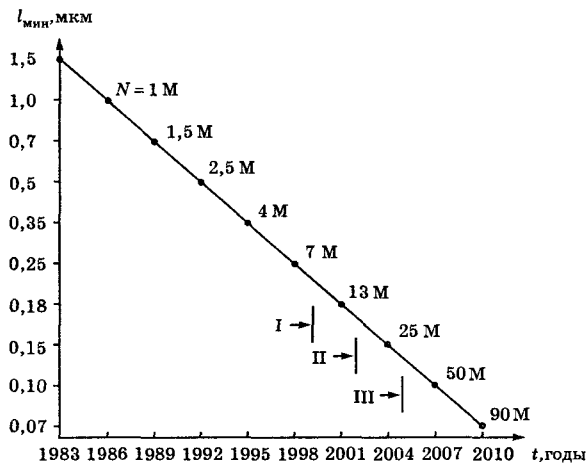


Рис. 6.14. Прогресс и перспективы промышленных методов литографии. I — оптическая литография с длиной волны 365 нм, 248 нм и 193 нм. II — рентгенолитография или прямое получение рисунка с помощью электронного луча. III — электронно-лучевая проекционная литография. N — количество логических элементов микропроцессора на 1 см^2 кристалла БИС

Если считать, что темпы развития микроэлектроники до 2010 г. не изменятся, то каждые три года, по-прежнему, минимальный размер элементов будет уменьшаться с коэффициентом 0,7. Для достижения таких разрешений, естественно, понадобятся и новые маскирующие материалы (резисты), и новые производительные установки экспонирования, и новые, интегрированные технологические процессы, отличающиеся от современных повышенными экологическими требованиями.

На рис. 6.14 показаны наиболее вероятные методы, уже апробированные в научных лабораториях, которые могут обеспечить требуемые разрешения на пластинах кремния диаметром до 300 мм. При этом еще одной сложнейшей проблемой, решение которой предстоит найти, является поиск экономичных методов совмещения и контроля жестких допусков на совмещение.

6.8. Нанесение тонких пленок

Тонкие пленки не только являются основой тонкопленочных ГИС, но широко используются и в полупроводниковых интегральных схемах. Поэтому методы получения тонких пленок относятся к общим вопросам технологии микроэлектроники.

Существуют три основных метода нанесения тонких пленок на подложку и друг на друга: *термическое* (вакуумное) *напыление*, *ионно-плазменное напыление* и *электрохимическое осаждение*. Ионно-плазменное напыление имеет две разновидности: *катодное напыление* и собственно ионно-плазменное.

Термическое (вакуумное) напыление. Схема этого метода показана на рис. 6.15. Металлический или стеклянный колпак *1* расположен на опорной плите *2*. Между ними находится прокладка 3, обеспечивающая поддержание вакуума после откачки подколпачного пространства. Подложка *4*, на которую проводится напыление, закреплена на держателе *5*. К держателю примыкает нагреватель *6* (напыление проводится на нагретую подложку). Испаритель *7* включает в себя нагреватель и источник напыляемого вещества. Поворотная заслонка *8* перекрывает поток паров от испарителя к подложке: напыление длится в течение времени, когда заслонка открыта.

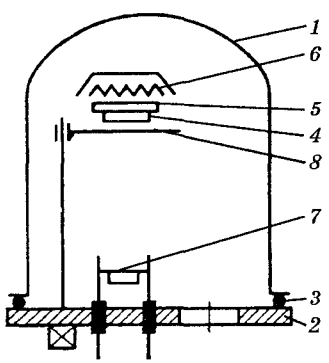


Рис. 6.15. Схема установки термического напыления

Нагреватель обычно представляет собой нить или спираль из тугоплавкого металла (вольфрам, молибден и др.), через которую пропускается достаточно большой ток. Источник напыляемого вещества связывается с нагревателем по-разному: в виде скобок («гусариков»), навешиваемых на нить накала; в виде небольших стержней, охватываемых спиралью, в виде порошка, засыпанного в тигель, нагреваемый спиралью, и т.п. Вместо нитей накала в последнее время используют нагрев с помощью электронного луча или луча лазера.

На подложке создаются наиболее благоприятные условия для конденсации паров, хотя частично конденсация происходит и на стенках колпака. Слишком низкая температура подложки препятствует равномерному распределению адсорбируемых атомов: они группируются в «островки» разной толщины, часто не связанные друг с другом. Наоборот, слишком высокая температура подложки приводит к отрыву только что осевших атомов, к их «реиспарению». Поэтому для получения качественной пленки температура подложки должна лежать в некоторых оптимальных пределах (обычно 200–400 °С). Скорость роста пленок в зависимости от ряда факторов (температура нагревателя, температура подложки, расстояние от испарителя до подложки, тип напыляемого материала и др.) лежит в пределах от десятых долей до десятков нанометров в секунду.

Прочность связи — сцепления пленки с подложкой или другой пленкой — называется *адгезией*. Некоторые распространенные материалы (например, золото) имеют плохую адгезию с типичными подложками, в том числе с кремнием. В таких случаях на подложку сначала наносят так называемый *подслой*, характерный хорошей адгезией, а затем на него напыляют основной материал, у которого адгезия с подслоем тоже хорошая. Например, для золота подслоем могут быть никель или титан.

Для того чтобы атомы газа, летящие от испарителя к подложке, испытывали минимальное количество столкновений с атомами остаточного газа и тем самым минимальное рассеяние, в подколпачном пространстве нужно обеспечивать достаточно высокий вакуум. Критерием необходимого вакуума может служить условие, чтобы средняя длина свободного пробега атомов в несколько раз превышала расстояние между испарителем и подложкой. Однако этого условия часто недостаточно, так как любое количество остаточного газа чревато загрязнением напыляемой пленки и изменением ее свойств. Поэтому в принципе вакуум в установках термического напыления должен быть как можно более высоким. В настоящее время вакуум ниже 10^{-6} мм рт. ст. считается неприемлемым, а в ряде первоклассных напылительных установок он доведен до 10^{-11} мм рт. ст.

Главными достоинствами рассмотренного метода являются его простота и возможность получения исключительно чистых пленок (при высоком вакууме). Однако у него есть и серьезные

недостатки: трудность напыления тугоплавких материалов и трудность (а иногда невозможность) воспроизведения на подложке химического состава испаряемого вещества. Последнее объясняется тем, что при высокой температуре химические соединения диссоциируют, а их составляющие конденсируются на подложке раздельно. Естественно, имеется вероятность того, что новая комбинация атомов на подложке не будет соответствовать структуре исходной молекулы.

Катодное напыление. Схема этого метода показана на рис. 6.16. Здесь большинство компонентов те же, что и на рис. 6.15. Однако отсутствует испаритель: его место по расположению (и по функции) занимает катод 6, который либо состоит из напыляемого вещества, либо электрически контактирует с ним. Роль анода выполняет подложка вместе с держателем.

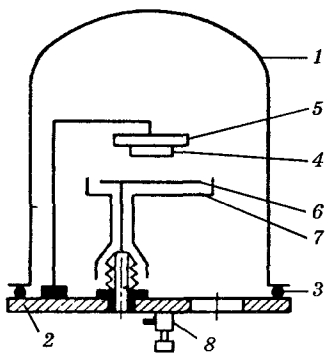


Рис. 6.16. Схема установки катодного напыления

Подколпачное пространство сначала откачивают до 10^{-5}–10^{-6} мм рт. ст., а затем в него через штуцер *8* вводят некоторое количество очищенного нейтрального газа (чаще всего аргона), так что создается давление 10^{-1}–10^{-2} мм рт. ст. При подаче высокого (2–3 кВ) напряжения на катод (анод заземлен из соображений электробезопасности) в пространстве анод-катод возникает аномальный тлеющий разряд, сопровождающийся образованием квазинейтральной электронно-ионной плазмы.

Специфика аномального тлеющего разряда состоит в том, что в прикатодном пространстве образуется настолько сильное электрическое поле, что положительные ионы газа, ускоряемые этим полем и бомбардирующие катод, выбивают из него не только электроны (необходимые для поддержания разряда), но и нейтральные атомы. Тем самым катод постепенно разрушается. В обычных газоразрядных приборах разрушение катода недопустимо (поэтому в них используется нормальный тлеющий разряд), но в данном случае выбивание атомов из катода является полезным процессом, аналогичным испарению.

Важным преимуществом катодного напыления по сравнению с термическим является то, что распыление катода не связано с высокой температурой. Соответственно отпадают трудности при напылении тугоплавких материалов и химических соединений (см. последний абзац предыдущего раздела).

Однако в данном методе катод (т.е. напыляемый материал), будучи элементом газоразрядной цепи, должен обладать высокой электропроводностью. Такое требование ограничивает ассортимент напыляемых материалов. В частности, оказывается невозможным напыление диэлектриков, в том числе многих окислов и других химических соединений, распространенных в технологии полупроводниковых приборов.

Это ограничение в значительной мере устраняется при использовании так называемого *реактивного* (или *химического*) катодного напыления, особенность которого состоит в добавлении к основной массе инертного газа небольшого количества активных газов, способных образовывать необходимые химические соединения с распыляемым материалом катода. Например, примешивая к аргону кислород, можно вырастить на подложке пленку окисла. Примешивая азот или монооксись углерода, можно получить нитриды или карбиды соответствующих металлов. В зависимости от парциального давления активного газа химическая реакция может происходить либо на катоде (и тогда на подложке осаждается уже готовое соединение), либо на подложке-аноде.

Недостатками катодного напыления в целом являются некоторая загрязненность пленок (из-за использования сравнительно низкого вакуума), меньшая по сравнению с термическим методом скорость напыления (по той же причине), а также сложность контроля процессов.

Ионно-плазменное напыление. Схема этого метода показана на рис. 6.17. Главная его особенность по сравнению с методом катодного напыления состоит в том, что в промежутке между электродом *9* — *мишенью* (с нанесенным не нее напыляемым материалом) и подложкой *4* действует независимый, «дежурный» газовый разряд. Разряд имеет место между электродами *6* и *7*, причем тип разряда — несамостоятельный дуговой. Для этого типа разряда характерны: наличие специального источника электронов в виде накаливаемого катода (*6*), низкие рабочие напряжения (десятки вольт) и большая плотность элект-

ронно-ионной плазмы. Подколпачное пространство, как и при катодном напылении, заполнено нейтральным газом, но при более низком давлении (10^{-3}–10^{-4} мм рт. ст.).

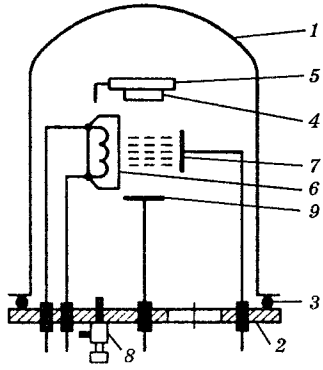


Рис. 6.17. Схема установки ионно-плазменного напыления

Процесс напыления состоит в следующем. На мишень относительно плазмы (практически — относительно заземленного анода 7) подается отрицательный потенциал (2–3 кВ), достаточный для возникновения аномального тлеющего разряда и интенсивной бомбардировки мишени положительными ионами плазмы. Выбиваемые атомы мишени попадают на подложку и осаждаются на ней. Таким образом, принципиальных различий между процессами катодного и ионно-плазменного напыления нет. Различаются лишь конструкции установок: их называют соответственно двух- и трехэлектродными.

Начало и конец процесса напыления определяются подачей и отключением напряжения на мишени. Если предусмотреть механическую заслонку (см. рис. 6.15), то ее наличие позволяет реализовать важную дополнительную возможность: если до начала напыления закрыть заслонку и подать потенциал на мишень, то будет иметь место ионная очистка мишени (см. раздел 6.6). Такая очистка полезна для повышения качества напыляемой пленки. Аналогичную очистку можно проводить на подложке, подавая на нее (до напыления пленки) отрицательный потенциал.

При напылении диэлектрических пленок возникает затруднение, связанное с накоплением на мишени положительного заряда, препятствующего дальнейшей ионной бомбардировке. Это затруднение преодолевается путем использования так называемого *высокочастотного* ионно-плазменного напыления. В этом случае на мишень наряду с постоянным отрицательным напряжением подается переменное напряжение высокой частоты (около 15 МГц) с амплитудой, несколько превышающей постоянное напряжение. Тогда во время большей части периода

результирующее напряжение отрицательно; при этом происходит обычный процесс распыления мишени и на ней накапливается положительный заряд. Однако во время небольшой части периода результирующее напряжение положительно; при этом мишень бомбардируется электронами из плазмы, т.е. распыления не происходит, но зато компенсируется накопленный положительный заряд[1].

Вариант реактивного (химического) ионно-плазменного напыления открывает те же возможности получения окислов, нитридов и других соединений, что и реактивное катодное напыление (см. предыдущий раздел).

Преимущества собственно ионно-плазменного метода по сравнению с катодным состоят в большей скорости напыления и большей гибкости процесса (возможность ионной очистки, возможность отключения рабочей цепи без прерывания разряда и др.). Кроме того, на качестве пленок сказывается более высокий вакуум.

Анодирование. Один из вариантов химического ионно-плазменного напыления называют *анодированием.* Этот процесс состоит в окислении поверхности металлической пленки (находящейся под положительным потенциалом) отрицательными ионами кислорода, поступающими из плазмы газового разряда. Для этого к инертному газу (как и при чисто химическом напылении) следует добавить кислород. Таким образом, анодирование осуществляется не нейтральными атомами, а ионами.

Химическое напыление и анодирование, вообще говоря, проходят совместно, так как в газоразрядной плазме (если она содержит кислород) сосуществуют нейтральные атомы и ионы кислорода. Для того чтобы анодирование превалировало над чисто химическим напылением, подложку располагают «лицом» (т.е. металлической пленкой) в сторону, противоположную катоду, с тем чтобы на нее не попадали нейтральные атомы.

По мере нарастания окисного слоя ток в анодной цепи падает, так как окисел является диэлектриком. Для поддержания тока нужно повышать питающее напряжение. Поскольку часть этого напряжения падает на пленке, процесс анодирования

---

1 Если подавать на мишень чисто переменное напряжение, то за время положительного полупериода заряд электронов благодаря их большей подвижности превысит заряд ионов, и мишень приобретет отрицательный потенциал.

протекает в условиях большой напряженности поля в окисной пленке. В результате и в дальнейшем, при эксплуатации, она обладает повышенной электрической прочностью.

К числу других преимуществ анодирования относятся бóльшая скорость окисления (поскольку поле в пленке окисла ускоряет взаимные перемещения атомов металла и кислорода) и возможность управления процессом путем изменения тока в цепи разряда. Качество окисных пленок, получаемых методом анодирования, выше, чем при использовании других методов.

Электрохимическое осаждение. Этот метод получения пленок отличается от предыдущих тем, что рабочей средой является жидкость. Однако характер процессов сходен с ионно-плазменным напылением, поскольку и плазма, и электролит представляют собой квазинейтральную смесь ионов и неионизированных молекул или атомов. А главное, осаждение происходит так же постепенно (послойно), как и напыление, т.е. обеспечивает возможность получения тонких пленок.

Электрохимическое осаждение исторически развилось значительно раньше всех других рассмотренных методов — еще в XIX веке. Уже десятки лет назад оно широко использовалось в машиностроении для разного рода тонких гальванических покрытий (никелирование, хромирование и т.п.). В микроэлектронике электрохимическое осаждение не является альтернативой термическому и ионно-плазменному напылению; оно дополняет их и сочетается с ними.

В основе электрохимического осаждения лежит электролиз раствора, содержащего ионы необходимых примесей. Например, если требуется осадить медь, используется раствор медного купороса, а если золото или никель — растворы соответствующих солей.

Ионы металлов имеют в растворе положительный заряд. Поэтому, чтобы осадить металлическую пленку, подложку следует использовать как катод. Если подложка является диэлектриком или имеет низкую проводимость, на нее предварительно наносят тонкий металлический подслой, который и служит катодом. Подслой можно нанести методом термического или ионно-плазменного напыления.

Чтобы осуществить электрохимическое анодирование, окисляемую пленку металла следует использовать как анод, а электролит должен содержать ионы кислорода.

Большое преимущество электрохимического осаждения перед напылением состоит в гораздо большей скорости процесса, которая к тому же легко регулируется изменением тока. Поэтому основная область применения электролиза в микроэлектронике — это получение сравнительно толстых пленок (10–20 мкм и более). Качество (структура) таких пленок хуже, чем при напылении, но для ряда применений они оказываются вполне приемлемыми.

6.9. Металлизация

В полупроводниковых ИС процесс металлизации призван обеспечить омические контакты со слоями полупроводника, а также рисунок межсоединений и контактных площадок.

Основным материалом для металлизации служит алюминий. Он оказался оптимальным в силу следующих положительных качеств: малое удельное сопротивление ($1{,}7\,10^{-6}$ Ом·см); хорошая адгезия к окислу SiO_2 (металлизация осуществляется по окислу); возможность сварных контактов с алюминиевой и золотой проволокой (при осуществлении внешних выводов); отсутствие коррозии; низкая стоимость и др.

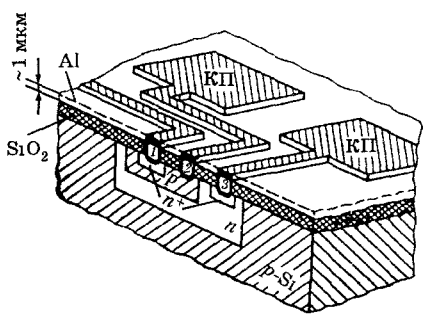


Рис. 6.18. Получение металлической разводки методом фотолитографии

При создании металлической разводки сначала на всю поверхность ИС напыляют сплошную пленку алюминия толщиной 0,1–1 мкм (рис. 6.18). Эта пленка контактирует со слоями кремния в специально сделанных (с помощью предыдущей фотолитографии) окнах в окисле (*1*, *2* и *3* на рис. 6.18). Основная же часть алюминиевой пленки лежит на поверхности окисла. Покрывая пленку алюминия фоторезистом, экспонируя его через соответствующий фотошаблон и проявляя, получают фоторезистную маску, которая защищает будущие полоски металлизации и контактные площадки (КП) от травителя. После вытравливания алюминия с незащищенных

участков и удаления фоторезиста остается запланированная металлическая разводка (на рис. 6.18 ее рельеф, прилегающий к контактам *1*, *2*, *3*, заштрихован).

Минимальная ширина полосок в современных ИС соответствует предельному разрешению литографии. Естественно, что для достижения необходимых допусков на ширину металлизации ее толщина, как правило, не может быть более 1/10 от минимальной ширины проводника. Расстояние между соседними проводниками — с целью ослабления паразитной емкостной связи между ними — выбирают более 1,5 мкм. Для проводника шириной 1 мкм, его толщина 0,1 мкм, а расстояние между ближайшими проводниками более 1,5 мкм.

Погонное сопротивление полоски шириной 10 мкм и толщиной 1 мкм составляет около 2 Ом/мм. Для контактных площадок, к которым в дальнейшем присоединяются внешние выводы, типичны размеры 100×100 мкм. Присоединение внешних выводов непосредственно к полоскам металлизации невозможно из-за малой ширины.

Разумеется, рисунок межсоединений предполагает отсутствие пересечений, т.е. коротких замыканий. Однако в ИС с высокой степенью интеграции не удается спроектировать металлическую разводку так, чтобы избежать пересечений. В этих случаях используется *многослойная* или *многоуровневая разводка*, т.е. несколько «этажей» металлизации, разделенных изолирующими слоями. Необходимые соединения между разными уровнями осуществляются через специальные окна в изолирующих слоях (рис. 6.19, *а*). Изоляцию между слоями обычно обеспечивают путем напыления диэлектрика по завершении очередной металлической разводки. В качестве диэлектрика чаще всего используют моноокись кремния SiO. Количество «этажей» при многоуровневой металлизации для современных БИС лежит в пределах от двух до четырех.

Некоторые предприятия для создания многоуровневой разводки используют алюмоксидную технологию. В этой технологии роль изоляции между соседними проводниками выполняют слои «пористого» Al_2O_3, а роль межслойной изоляции толщиной порядка 0,1 мкм играют слои «плотного» окисла, образуемого в результате анодирования первичного слоя Al (рис. 6.19, *б*). Отличительной особенностью этой технологии является планарность многоуровневой разводки.

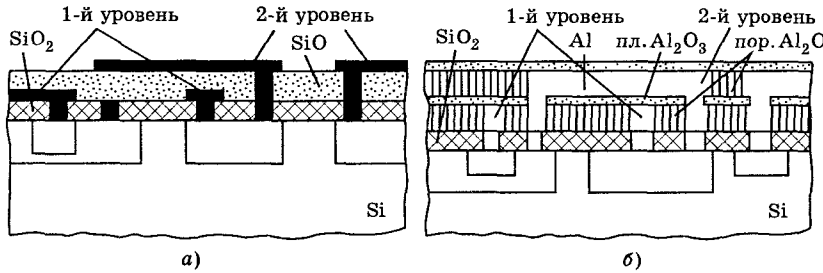


Рис. 6.19. Многослойная металлическая разводка

Проблема омических контактов при использовании алюминия состоит в следующем. Если пленку алюминия просто напылить на поверхность кремния, то образуются барьеры Шоттки (см. раздел 3.3), причем барьер на границе с n-слоем является не омическим, а выпрямляющим. Чтобы избежать барьеров Шоттки, алюминий *вжигают* в кремний при температуре около 600 °С, близкой к температуре эвтектики сплава Al–Si. При такой температуре на границе алюминиевой пленки с кремнием образуется слой, в котором растворен практически весь прилегающий алюминий. После застывания сплав представляет собой кремний, легированный алюминием; концентрация последнего составляет около $5 \cdot 10^{-18}$ см$^{-3}$.

Поскольку алюминий является акцептором по отношению к кремнию, возникает новая проблема: предотвращение образования p–n-переходов в n-слоях. Действительно, если концентрация доноров в n-слое меньше $5 \cdot 10^{18}$ см$^{-3}$, то атомы алюминия создадут в нем приповерхностный p-слой. Чтобы этого избежать, область n-слоя вблизи контакта специально легируют, превращая ее в n^+-слой с концентрацией доноров 10^{20} см$^{-3}$ и более (см. рис. 6.18). Тогда концентрация алюминия оказывается недостаточной для образования p-слоя, и p–n-переход не образуется.

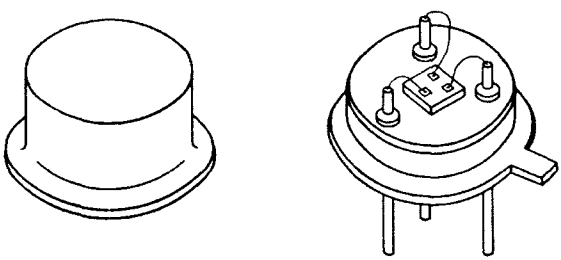
Если n-слой с самого начала сильно легирован (например, эмиттерный слой транзистора), то дополнительного легирования не требуется. Не возникает проблем и при контакте алюминия с p-слоями, так как растворение в них алюминия приводит к образованию приповерхностных p^+-слоев, что способствует повышению качества омического контакта.

6.10. Сборочные операции

После того как все основные технологические этапы (включая металлизацию) закончены, пластина кремния, содержащая сотни и тысячи ИС, поступает на операции тестового контроля электрических параметров. На этом этапе отбраковываются и специальным лаком маркируются все ИС, параметры которых не соответствуют требуемым значениям. Трудоемкость всех контрольных операций (на всех этапах производства ИС) достигает до 30% себестоимости работ по созданию схем. После маркировки отбракованных ИС пластина кремния разделяется на отдельные кристаллы.

Разделение осуществляется методом скрайбирования, т.е., по сути, процарапыванием вертикальных и горизонтальных рисок в промежутках между соседними чипами (см. рис. 1.1). Процарапывание осуществляют либо с помощью алмазного резца (наподобие того, как это делает стекольщик, разрезая стекло), либо с помощью лазерного луча. После скрайбирования пластину ломают на отдельные кристаллы и годные чипы монтируются в корпусах.

Сборка кристалла в корпусе начинается с операции, которую называют *посадкой на ножку* (под ножкой имеют в виду дно корпуса). При этом кристалл приклеивается или припаивается (легкоплавким припоем) в средней части ножки (на рис. 6.20 показан простой транзистор). Затем контактные площадки на кристалле соединяются со штырьками — выводами корпуса. Соединения осуществляются с помощью тонких (15–30 мкм) алюминиевых или золотых проволочек, которые одним концом закрепляются на контактных площадках, а другим на торцах штырьков.



Крышка корпуса Ножка корпуса

Рис. 6.20. Монтаж кристалла на ножке корпуса

Надежный электрический контакт между металлическими деталями (в данном случае контакт проволочек со штырьками и контактными площадками) может быть обеспечен разными методами. Наибольшее распространение в настоящее время имеет метод *термокомпрессии*, т.е. сочетание достаточного давления (прижатия деталей друг к другу) с повышенной температурой (200–300 °С), способствующей взаимной диффузии атомов из одной детали в другую.

Термокомпрессия, в свою очередь, реализуется в виде разных конструктивных вариантов. На рис. 6.21 показан простейший пример. С помощью клинообразного пуансона, изготовленного из твердого материала (например из алмаза), проволочка прижимается к поверхности контактной площадки. При этом проволочка деформируется (несколько расплющивается). Непродолжительного удержания расплющенной проволочки на площадке достаточно для образования их прочного соединения с малым переходным сопротивлением. При значительном количестве внешних контактов операция разводки кристалла на ножке осуществляется только в автоматизированном процессе.

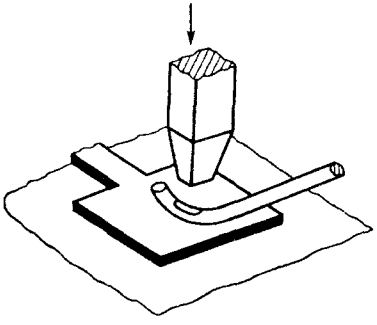


Рис. 6.21. Термокомпрессия

По окончании монтажа кристалла на ножке следует *корпусирование*, т.е. окончательное внешнее оформление транзистора. Ножка корпуса соединяется с крышкой (рис. 6.20) путем горячей или холодной сварки (последняя по существу близка к термокомпресии). Корпусирование предполагает также защиту кристалла от влияния внешней среды, поэтому его проводят либо в вакууме, либо в среде инертного газа (азот, аргон). Бескорпусные варианты транзисторов описаны в следующем разделе.

Главная специфика сборочных операций применительно к интегральным схемам состоит в том, что корпуса ИС многовыводные: у простых ИС количество выводов составляет 8–14, а у больших доходит до сотен и более. Номенклатура корпусов для ИС довольно разнообразна: наряду с круглым корпусом, похо-

жим на транзисторный (см. рис. 6.20), используются прямоугольные корпуса: металлические или пластмассовые, с выводами, лежащими в плоскости корпуса или перпендикулярными ей (рис. 6.22). Выбор корпуса в значительной степени зависит от назначения аппаратуры и способов ее конструирования.

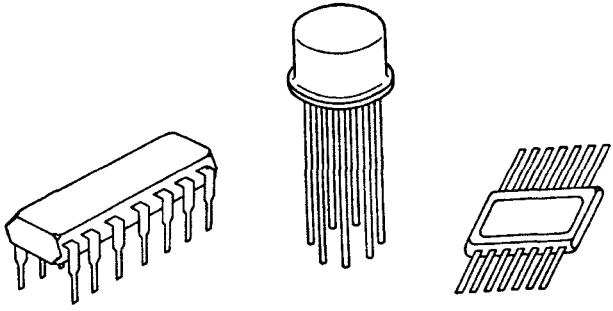


Рис. 6.22. Типичные корпуса интегральных схем

6.11. Технология тонкопленочных гибридных ИС

Согласно определению, приведенному в разделе 1.2, гибридные ИС представляют собой совокупность пленочных пассивных элементов и навесных активных компонентов. Поэтому технологию тонкопленочных ГИС можно разбить на технологию тонкопленочных пассивных элементов и технологию монтажа активных компонентов.

Изготовление пассивных элементов. Тонкопленочные элементы ГИС осуществляются с помощью технологических методов, описанных в разделе 6.8, т.е. путем локального (через маски) термического, катодного или ионно-плазменного напыления того или иного материала на диэлектрическую подложку.

В качестве масок длительное время использовались *накладные металлические трафареты.* Такие трафареты представляли собой тонкую биметаллическую фольгу с отверстиями-окнами. Основу трафарета составлял слой электрохимически нанесенного никеля толщиной 10–20 мкм. Последний определял размеры окон, т.е. рисунок трафарета, а слой бериллиевой бронзы выполнял роль несущей конструкции.

Серьезные недостатки металлических накладных трафаретов заключаются в том, что, во-первых, в процессе напыления пленок происходит напыление на сами трафареты, что меняет их толщину и по-

ний, которых мы не будем касаться. Заметим лишь, что с помощью фотолитографии не удается получать многослойные структуры типа конденсаторов. Однако это ограничение не очень существенно, так как в последнее время предпочитают использовать в ГИС навесные конденсаторы (ради экономии площади).

Для резистивных пленок чаще всего используют хром, нихром (Ni — 80%, Cr — 20%) и кермет из смеси хрома и монооксиси кремния (1:1). Метод напыления для этих материалов — термический (вакуумный). Омические контакты к резистивным пленкам (полоскам) осуществляются так, как показано на рис. 6.23.

Для обкладок конденсаторов используют алюминий, причем до напыления нижней обкладки (прилегающей к подложке) приходится предварительно напылять тонкий *подслой* из сплава CrTi, так как адгезия алюминия непосредственно с подложкой оказывается недостаточной.

Для диэлектрических слоев пленочных конденсаторов по совокупности требований (большая диэлектрическая проницаемость ε, малый тангенс угла потерь $tg\delta$, большая пробивная напряженность и др.) наибольшее распространение имеют монооксись кремния SiO и монооксись германия GeO. Особое место среди диэлектриков занимают окислы Ta_2O_5 и Al_2O_3, которые получают не методом напыления, а методом анодирования нижних металлических обкладок (Ta или Al).

Для проводниковых пленок и омических контактов используют, как правило, либо золото с подслоем CrTi, либо медь с подслоем ванадия (назначение подслоев — улучшить адгезию с подложкой). Толщина проводящих пленок и контактных площадок обычно составляет 0,5–1 мкм. Размеры контактных площадок от 100×100 мкм и более.

Толщина наносимых пленок контролируется в процессе напыления. Для этого используются несколько методов. Один из них, пригодный только в случае резистивных пленок, состоит в использовании так называемого *свидетеля*. Свидетель представляет собой вспомогательный (не входящий в структуру ГИС) слой, напыляемый одновременно с рабочими слоями, но расположенный на периферии подложки и снабженный двумя заранее предусмотренными внешними выводами. Через эти выводы осуществляется контроль сопротивления свидетеля в про-

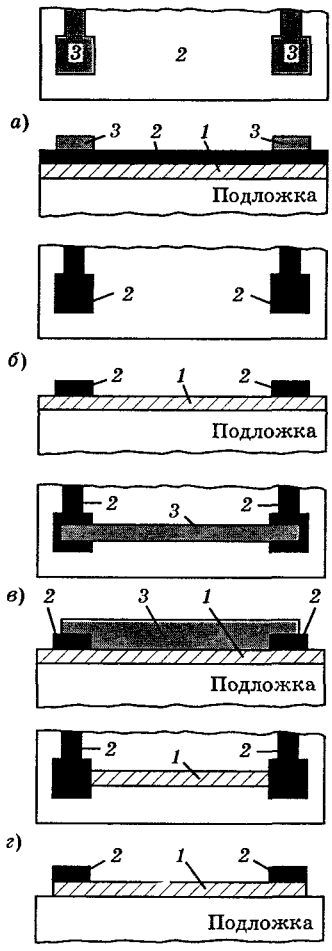


Рис. 6.23. Получение тонкопленочных резисторов методом фотолитографии:
а — фоторезистная маска (3) под рисунок проводящего слоя; *б* — готовый рисунок проводящего слоя (2);
в — фоторезистная маска (3) под рисунок резистивного слоя; *г* — готовый резистор с проводящими выводами

степенно приводит их в негодность. Во-вторых, металлические трафареты мало пригодны при катодном и ионно-плазменном напылении, так как металл искажает электрическое поле и, следовательно, влияет на скорость напыления. Поэтому в последние годы от металлических накладных трафаретов практически отказались и используют для получения необходимого рисунка фотолитографию — метод, заимствованный из технологии полупроводниковых ИС.

Фотолитографию осуществляют следующим образом. На подложку наносят сплошные пленки необходимых материалов, например, резистивный слой и поверх него — проводящий слой. Затем поверхность покрывают фоторезистом и с помощью соответствующего фотошаблона создают в нем рисунок для проводящего слоя (например, для контактных площадок будущего резистора, рис. 6.23, *а*). Через окна в фоторезистной маске проводят травление проводящего слоя, после чего фоторезист удаляют. В результате на пока еще сплошной поверхности резистивного слоя получаются готовые контактные площадки (рис. 6.23, *б*). Снова наносят фоторезист и с помощью другого фотошаблона создают рисунок полоски резистора (рис. 6.23, *в*). Затем проводят травление, удаляют фоторезист и получают готовую конфигурацию резистора с контактными площадками (рис. 6.23, *г*).

Конечно, важно, чтобы травитель, действующий на проводящий слой, не действовал на резистивный и наоборот. Имеется и еще ряд ограниче-

ных элементов. Только после контроля пластины скрайбируются и снабжаются навесными компонентами. ГИС могут, как и ИС, помещаться в корпусе или (в бескорпусном исполнении) герметизироваться в составе аппаратуры.

Монтаж навесных компонентов. В качестве навесных компонентов используются бескорпусные диоды, транзисторы, резистивные и конденсаторные сборки, а также бескорпусные ИС и БИС. Рассмотрим пример.

Простейшим вариантом бескорпусного транзистора является кристалл, полученный после скрайбирования, к трем контактным площадкам которого присоединены тонкие проволочные выводы и который защищен от внешней среды каплей эпоксидной смолы, обволакивающей кристалл со всех сторон. Такой транзистор приклеивается к подложке вблизи тех пленочных элементов, с которыми он должен быть соединен, после чего проволочные выводы транзистора методом термокомпрессии присоединяются к соответствующим контактным площадкам на подложке.

Имеется два других варианта бескорпусных транзисторов, монтаж которых осуществляется иначе. Первый вариант называется транзистором с *шариковыми выводами* (рис. 6.24, *а*). Шарики диаметром 50–100 мкм связаны с контактными площадками транзистора, а через них — с тем или иным слоем кремния: эмиттерным, базовым или коллекторным. Материалом для шариков служит золото, медь или сплав Sn–Sb. Из того же материала на диэлектрической подложке пленочной ГИС делаются *контактные столбики* высотой 10–15 мкм и диаметром 150–200 мкм, расположенные в точном соответствии с расположением шариков на кристалле кремния (рис. 6.24, *б*). Соединение шариков со столбиками осуществляется *методом перевернутого монтажа* (англ. термин flip-chip): кристалл транзистора переворачивается «вверх ногами», т.е. шариками вниз, и накладывается шариками на столбики подложки (рис. 6.24, *в*). Сочетая давление на кристалл с повышением температуры (т.е. в сущности используя термокомпрессию), обеспечивают прочное соединение шариков со столбиками. Как видим, метод перевернутого монтажа существенно экономит площадь подложки ГИС и время, необходимое на разводку активных компонентов. Главная трудность состоит в совмещении шариков со столбиками, поскольку кристалл при наложении перевернут «вверх ногами» и закрывает от оператора места соединения.

цессе напыления. Геометрия свидетеля известна. Поэтому, когда его сопротивление достигает значения, соответствующего необходимой толщине, напыление прекращают (перекрывают заслонку). Толщина рабочих слоев будет такой же, как у свидетеля, так как они напылялись в одинаковых условиях.

Другой способ контроля состоит в использовании в качестве свидетеля тонкой кварцевой пластины, которая через внешние выводы присоединена к колебательному контуру генератора колебаний. Как известно, кварцевая пластина обладает свойствами колебательного контура, причем резонансная частота однозначно связана с толщиной пластины. В процессе напыления толщина пластины меняется и меняется частота генератора. Изменения частоты легко измерить и остановить процесс напыления в нужный момент.

Подложки тонкопленочных ГИС должны прежде всего обладать хорошими изолирующими свойствами. Кроме того, желательны малая диэлектрическая проницаемость, высокая теплопроводность, достаточная механическая прочность. Температурный коэффициент расширения должен быть близким к температурным коэффициентам расширения используемых пленок. Типичные параметры подложек следующие: $\rho = 10^{14}$ Ом·см; $\varepsilon = 5–15$; $\mathrm{tg}\delta = (2–20)\ 10^{-4}$; $\mathrm{TK}L = (5–7)\ 10^{-6}$.

В настоящее время наибольшее распространение в качестве подложек имеют *ситалл* и *керамика*. Ситалл представляет собой кристаллическую разновидность стекла (обычное стекло аморфно), а керамика — смесь окислов в стекловидной и кристаллической фазах (главные составляющие Al_2O_3 и SiO_2).

Толщина подложек составляет 0,5–1 мм в зависимости от площади. Площадь подложек у ГИС иногда больше площади кристаллов у полупроводниковых ИС. Стандартные размеры подложек лежат в пределах от 12×10 до 96×120 мм. Требования к гладкости поверхности примерно такие же, как и в случае кремния: допустимая шероховатость не превышает 25×50 нм (класс шероховатости 12–14).

Обычно ГИС, как и полупроводниковые ИС, изготавливаются групповым методом на ситалловых или иных пластинах большой площади. По завершении основных технологических операций, связанных с получением пленочных пассивных элементов и металлической разводки, проводится выходной тестовый контроль и, если необходимо, подгонка параметров пассив-

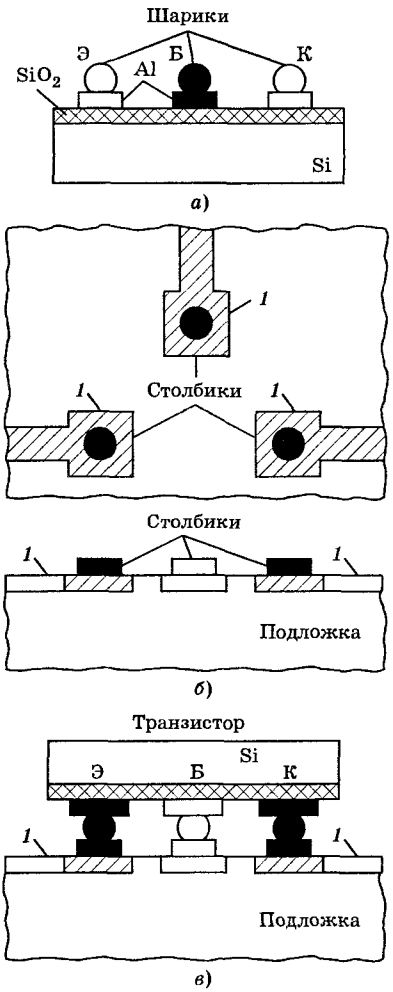


Рис. 6.24. Монтаж бескорпусных транзисторов с шариковыми выводами: *а* — транзистор с шариковыми выводами; *б* — контактные столбики на подложке пленочной ИС; *в* — соединение шариков с контактными столбиками; *1* — контактные площадки на подложке и выводы от них

Трудность совмещения контактных площадок кристалла и подложки облегчается при использовании второго варианта бескорпусных транзисторов — транзистора с *балочными выводами* (рис. 6.25, *а*). Здесь контактные площадки продлены за пределы кристалла и нависают над его краями на 100–150 мкм, откуда и название — *балки*. Толщина балок (10–15 мкм) значительно больше толщины металлической разводки на кристалле. Поэтому их получают не напылением, а электрохимическим осаждением золота (с подслоем из титана). Длина балочных выводов 200–250 мкм (включая выступ), а ширина такая же, как у обычных контактных площадок (50–200 мкм).

Получение балок основано на сквозном травлении кремния через фоторезистную маску, нанесенную на нижнюю поверхность пластины (рис. 6.25, *б*). При сквозном травлении одновременно с получением балок достигается разделение пластины на отдельные кристаллы без механического скрайбирования. До начала травления пластина приклеивается верхней (лицевой) поверхностью к стеклу. Чтобы сократить время травления и избежать бокового растравливания пластины, ее (после приклеивания к стеклу) сошлифовывают от обычной

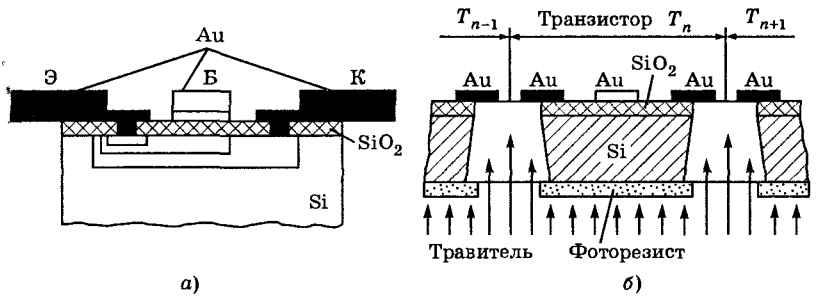


Рис. 6.25. Бескорпусные транзисторы с балочными выводами: *а* — транзистор с балочными выводам1и; *б* — получение балок и разделение транзисторов на пластине

толщины 200–300 мкм до 50 мкм. По окончании травления клей растворяют и разделенные кристаллы отпадают от стекла.

Монтаж навесных компонентов с балочными выводами может осуществляться так же, как и в случае шариковых выводов — методом перевернутого монтажа. При этом выступающие балки хорошо видны и их совмещение с контактными площадками на подложке не представляет затруднений. Можно монтировать кристалл и «лицом вверх», но тогда в подложке следует предусмотреть углубление для кристалла.

Изготовление шариковых и балочных выводов сложнее и дороже, чем проволочных, но они обеспечивают существенное упрощение и удешевление сборочных операций и в 4–10 раз позволяют уменьшить площадь на подложке ГИС, отводимую для размещения навесных компонентов. К сожалению, надежность таких соединений ниже, чем в случае использования гибких выводов.

6.12. Технология толстопленочных гибридных ИС

Пассивные элементы толстопленочных ГИС получаются локальным нанесением на подложку полужидких *паст-стеклоэмалей* с последующим их высушиванием и вжиганием в подложку. Следовательно, в данном случае пленки приобретают свою толщину сразу, а не постепенно — слой за слоем — как при тонкопленочной технологии.

Последовательность технологических операций при нанесении толстых пленок следующая:

а) нанесение слоя пасты на подложку через маску — накладной трафарет (отсюда название — *метод трафаретной печати*);

б) выжигание (испарение) растворителя при температуре 300–400 °С и тем самым превращение пасты из полужидкого состояния в твердое;

в) вжигание затвердевшего вещества пасты в подложку — *спекание* — при температуре 500–700 °С (в зависимости от состава пасты).

Операция вжигания — самая ответственная в технологическом цикле; она требует высокой стабилизации температуры: с точностью ±1 °С.

В основе всех паст-стеклоэмалей лежит так называемая *фритта* — тончайший порошок стекла, к которому, в зависимости от назначения пасты, примешивается порошок резистивного, проводящего или диэлектрического материала. Дисперсная (т.е. совершенно однородная) смесь фритты и примесного материала приобретает вязкость при добавлении специальных органических веществ и растворителей. На этапе выжигания (см. выше) растворитель испаряется, а органические вещества связывают частицы порошка в единую компактную массу.

Для проводящих паст примесью обычно служит серебро или золото, для резистивных — смесь серебра и палладия (1:1), а для диэлектрических — титанат бария с высокой диэлектрической проницаемостью. Варьируя материал и процентное содержание примесей, можно изменять электрические параметры пленок в очень широких пределах (см. раздел 7.11).

Масками для нанесения паст на подложку служат сетчатые трафареты (рис. 6.26, *а*). Они представляют собой тонкую сетку из капрона или нержавеющей стали, натянутую на дно рамки[1]. Размер ячеек сетки — около 100 мкм, диаметр нитей — около 50 мкм. Большая часть сетки покрыта пленкой, но в пленке имеются окна. Рисунок окон получают методом фотолитографии, вытравливая отверстия в пленке. Учитывая ячеистую структуру сетки, размеры окон трудно сделать менее 10–200 мкм. Это предопределяет минимальные размеры элементов толстопленочных ГИС и ширину линий.

---

1 Исходным материалом для сеток был шелк и потому методику нанесения паст через сетчатые трафареты часто называют *шелкографией*.

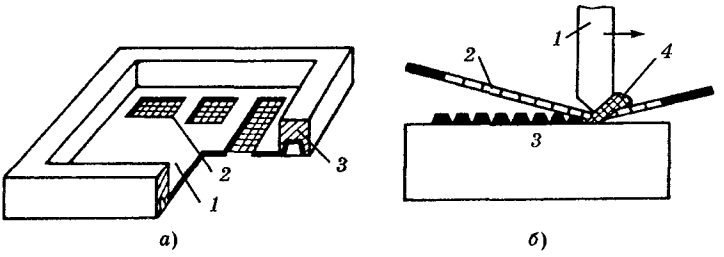


Рис. 6.26. Метод локального нанесения пасты: *а* — сетчатый трафарет; *б* — продавливание пасты через трафарет: *1* — ракель; 2 — сетка; *3* — подложка; *4* — паста

Рамка с трафаретом заполняется пастой и размещается над подложкой на расстоянии 0,5–1 мм. После этого на сетку опускается специальный нож — *ракель*, который, перемещаясь вдоль рамки, продавливает пасту через отверстия в сетке (рис. 6.26, *б*). Несмотря на простоту идеи продавливания, эта операция — прецизионная; на качество будущей пленки и повторяемость результатов оказывают влияние угол заточки ракеля, его наклон относительно подложки, скорость перемещения и другие факторы.

Вообще говоря, сетка на трафарете не обязательна: можно продавливать пасту и через сплошные отверстия. Однако качество пленок при этом хуже, так как сетка обеспечивает более однородные слои в результате слияния отдельных «капель», прошедших через ячейки сетки. Толщина получаемых пленок зависит от диаметра нитей и размеров ячеек. Обычно она составляет 20–40 мкм.

К подложкам для толстопленочных ГИС предъявляются в общем те же требования, что и для тонкопленочных. Особое внимание часто уделяется повышенной теплопроводности, так как толстопленочный вариант ГИС характерен для мощных схем. Поэтому распространены высокоглиноземистые керамики (96% Al_2O_3) и бериллиевые керамики (99,5% BeO); последние в 7–10 раз превышают глиноземистые по теплопроводности, но уступают им по прочности. Важная отличительная особенность подложек для толстопленочных ГИС состоит в том, что их поверхность должна быть достаточно шероховатой, чтобы обеспечить необходимую адгезию с веществом пасты. Степень шероховатости характеризуется неровностями до 1–2 мкм.

Методы монтажа навесных компонентов те же, что и у тонкопленочных ГИС, но размеры контактных площадок делают больше: 400×400 мкм.

В целом толстопленочная технология характерна простотой и низкой стоимостью ИС. Однако по сравнению с тонкопленочной технологией плотность компоновки оказывается меньше (из-за большей ширины линий), а разброс параметров — больше (из-за неконтролируемой толщины пленок).

Контрольные вопросы

1. Почему для изготовления большинства полупроводниковых приборов используются монокристаллические материалы?
2. Перечислите основные технологические операции получения монокристаллического кремния.
3. Для чего при производстве ИС на монокристаллических подложках выращивают эпитаксиальные слои кремния?
4. Назовите акцепторные примеси для легирования кремния. Чем они отличаются друг от друга? Какие донорные примеси кремния вы знаете?
5. Назовите основные операции технологического процесса изготовления ИС.
6. Какие способы окисления кремния вам известны? В чем их принципиальные отличия?
7. Напишите уравнение двух законов Фика.
8. В чем качественное отличие распределения примеси при термической диффузии из неограниченного и ограниченного источников диффузанта?
9. Как распределяются примеси при ионной имплантации? Какие факторы оказывают влияние на характер этого распределения?
10. Назовите известные вам травители пленок SiO_2. А для кремния?
11. Что такое анизотропное травление? Для решения каких задач в производстве ИС применяют анизотропное травление?
12. Какие методы изготовления фотошаблонов вам знакомы? Сравните их.

13. Какие источники экспонирования применяются при литографии?
14. Почему разрешающая способность литографии зависит от длины волны источника экспонирования?
15. Каковы в настоящее время пределы разрешения литографии?
16. Какие методы нанесения тонких пленок металлов и диэлектриков вам знакомы? В чем их принципиальные отличия?
17. Какие металлы и почему используются в качестве подслоя при нанесении металлических пленок?
18. Назовите наиболее распространенные материалы для изготовления проводников ИС.
19. Как при изготовлении проводников ИС избежать образования p–n-переходов и барьеров Шоттки?
20. Чем обусловлены максимально достижимые пределы миниатюризации проводников ИС?
21. Какие типы корпусов ИС вам знакомы? Отличаются ли они по надежности?
22. Что такое шариковые и балочные выводы?
23. Назовите области применения гибридных пленочных и толстопленочных схем.

Глава 7

Элементы интегральных схем

7.1. Введение

Напомним, что *элементами ИС* (как полупроводниковых, так и гибридных) называют их неделимые составные части — те, которые нельзя отдельно специфицировать и поставить как отдельные изделия. Одна из особенностей элементов ИС по сравнению с аналогичными дискретными приборами или электрорадиоэлементами состоит в том, что они имеют электрическую связь с общей подложкой, а иногда и друг с другом. Поэтому математические и физические модели (эквивалентные схемы) элементов ИС несколько отличаются от моделей дискретных аналогов.

Вторая важная особенность элементов ИС по сравнению с дискретными приборами связана с тем, что все элементы ИС получаются в едином технологическом процессе. Например, все резисторы пленочной ИС получаются одновременно и, следовательно, характеризуются одинаковой толщиной и одинаковым удельным сопротивлением (разница может быть только в длине и ширине слоя). Или в полупроводниковой ИС рабочий слой резистора получается одновременно с базовым слоем транзистора и, следовательно, имеет те же электрофизические параметры, что и базовый слой. Иначе говоря, *при изготовлении элементов ИС имеется меньше «степеней свободы», чем при изготовлении их дискретных аналогов*: можно варьировать главным образом конфигурацией элементов ИС «в плане», т.е. их длиной и шириной, а не глубиной слоев и их электрофизическими параметрами. В результате *параметры элементов ИС в значительной мере коррелированы* (взаимосвязаны) *и ограничены*, чего нет у дискретных компонентов.

Отметим, однако, что указанные особенности прежде всего характерны для элементов БИС и СБИС, при изготовлении которых из-за проблем совмещения всегда приходится «экономить» на количестве фотолитографий. Если проблема ограни-

чения количества фотолитографий не превалирует над остальными, параметры элементов ИС могут значительно отличаться.

Помимо указанных двух особенностей, следует отметить, что в процессе развития микроэлектроники появились такие элементы ИС, которые не имеют аналогов в дискретной электронике: многоэмиттерные и многоколлекторные транзисторы, транзисторы с барьером Шоттки, трехмерные элементы и др.

Традиционные компоненты — диоды, конденсаторы и т.п. — изменились конструктивно, изменился диапазон их параметров. В полупроводниковых ИС отсутствуют аналоги таких традиционных компонентов, как катушки индуктивности и тем более трансформаторы.

Компонентами ИС, как отмечалось в гл. 1, называют такие составные части гибридных микросхем, которые можно специфицировать отдельно и поставлять в виде отдельных изделий. Компоненты ГИС представляют собой навесные детали, отличающиеся от «обычных» дискретных компонентов лишь конструктивным оформлением (бескорпусные диоды, транзисторы и ИС).

Главными элементами биполярных полупроводниковых ИС являются n–p–n-транзисторы. Именно на них ориентируются при разработке новых технологических циклов, стараясь обеспечить оптимальные параметры этих транзисторов. Технология всех других элементов (p–n–p-транзисторов, диодов, резисторов и т.п.) должна «приспосабливаться» к технологии n–p–n-транзистора. Такое «приспособление» означает, что для изготовления других элементов следует по возможности избегать дополнительных технологических операций: желательно использовать те же рабочие слои (коллекторный, базовый и эмиттерный), которые необходимы для n–p–n-транзистора. Отсюда такая терминология: «в качестве резистора используется базовый слой» или «рабочий слой резистора получается на этапе базовой диффузии».

Главными элементами МДП-транзисторных ИС до последнего времени были МДП-транзисторы с индуцированным p-каналом. Соответственно на них ориентировался технологический цикл и к этому циклу «приспосабливалась» технология остальных элементов. В последние годы, после того как удалось преодолеть трудности изготовления качественных n-канальных МДП-транзисторов и разработать технологию комплементар-

ных МДП (КМДП) схем, последняя, по существу, заняла главное место в технологии ИС.

7.2. Изоляция элементов

На рис. 7.1 показаны два n–p–n-транзистора и диод, выполненные в общей кремниевой подложке n-типа. Как видим, коллекторы транзисторов и катод диода оказались принудительно связаны через подложку. Такие связи, как правило, неприемлемы: они не соответствуют желательной конфигурации схемы. Следовательно, *элементы биполярных полупроводниковых ИС нужно изолировать друг от друга с тем, чтобы необходимые соединения осуществлялись только путем металлической разводки.*

В случае изготовления на одной подложке МДП-транзисторов истоки и стоки смежных элементов (рис. 7.2) оказываются разделенными встречно-включенными p–n-переходами и такая связь не столь губительна, как в биполярных элементах. Однако с ростом степени интеграции и «сближением» элементов, обратные токи разделительных p–n-переходов растут и принуждают разработчиков ИС искать способы изоляции не только биполярных, но и МДП элементов.

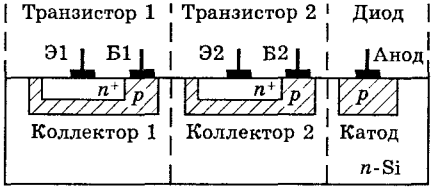


Рис. 7.1. Внутренняя связь элементов биполярных ИС через подложку в отсутствие изоляции

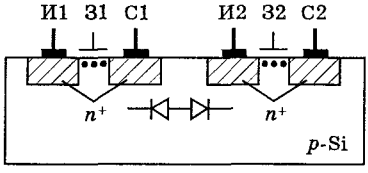


Рис. 7.2. Отсутствие связи между элементами МДП-транзисторных ИС

Сравнительная оценка способов изоляции. Все известные способы изоляции можно разделить на два главных типа: *изоляцию обратносмещенным p–n-переходом и изоляцию диэлектриком.* Оба типа изоляции проиллюстрированы на рис. 7.3.

Обедненный слой p–n-перехода, особенно при большом обратном смещении, имеет очень высокое удельное сопротивление, близкое к удельному сопротивлению диэлектриков. Поэтому указанные два типа

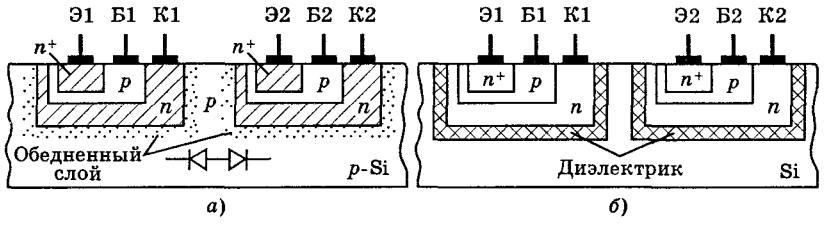


Рис. 7.3. Основные методы изоляции элементов ИС:
а — с помощью *p*–*n*-переходов; *б* — с помощью диэлектрика

изоляции различаются не столько удельным сопротивлением изолирующего слоя, сколько его структурой. Изоляцию *p*–*n*-переходом относят к однофазным способам, имея в виду, что материал по обе стороны и в пределах изолирующего слоя один и тот же — кремний. Изоляцию диэлектриком относят к двухфазным способам, имея в виду, что материал (фаза) изолирующего слоя отличается от материала подложки — кремния. Учитывая, что при разработке ИС происходит постоянное снижение рабочих напряжений, изоляция *p*–*n*-переходом применяется все реже и реже.

Из рис. 7.3, *а* ясно, что *изоляция p–n-переходом сводится к осуществлению двух встречно-включенных диодов между изолируемыми элементами* — так же, как в МДП-транзисторных ИС (рис. 7.2). Для того, чтобы оба изолирующих диода находились под обратным смещением (независимо от потенциалов коллекторов), на подложку задают максимальный отрицательный потенциал от источника питания ИС[1].

Изоляция *p*–*n*-переходом хорошо вписывается в общий технологический цикл биполярных ИС, однако ее недостатки — наличие обратных токов в *p*–*n*-переходах и наличие барьерных емкостей.

Изоляция диэлектриком более совершенная и «радикальная» (рис. 7.3, *б*). При комнатной температуре токи утечки в диэлектрике на 3–5 порядков меньше, чем обратные токи *p*–*n*-перехода. Что касается паразитной емкости, то, разумеется, она имеет место и при диэлектрической изоляции. Однако ее легко сделать меньше барьерной, выбирая материал с малой диэлектрической проницаемостью и увеличивая толщину диэлектрика. Как правило, паразитная емкость при диэлектриче-

1 Так же поступают в МДП-транзисторных ИС.

ской изоляции на порядок меньше, чем при изоляции p–n-переходом.

Изоляция p–n-переходом. Изолирующие переходы можно получить по-разному. Так, в чисто планарной технологии в свое время использовались методы *тройной диффузии* (рис. 7.4, *а*) и *встречной* (или двусторонней) *диффузии* (рис. 7.4, *б*). Обоим этим методам свойственны серьезные недостатки.

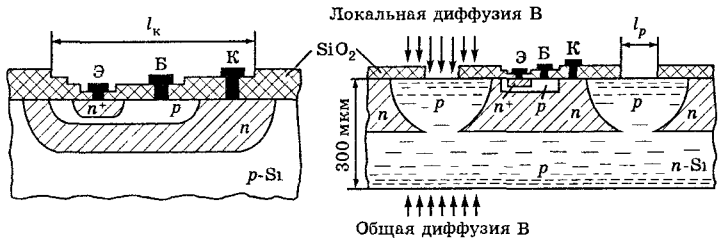


Рис. 7.4. Возможные варианты изоляции элементов с помощью переходов: *а* — метод тройной диффузии ($l_к$ — размер окна под коллекторную диффузию); *б* — метод встречной (двусторонней) диффузии (l_p — размер окна под разделительную диффузию)

В структуре на рис. 7.4, *а* коллекторный n-слой, получаемый на этапе 1-й диффузии, является неоднородным: концентрация примеси возрастает от донной части к поверхности. Поэтому на границе с базовым слоем концентрация примеси оказывается достаточно большой и соответственно пробивное напряжение коллекторного перехода получается сравнительно низким. Кроме того, сложна сама технология тройной диффузии.

В структуре на рис. 7.4, *б* изоляция элементов достигается общей диффузией акцепторной примеси через нижнюю поверхность пластины n-типа и локальной диффузией той же примеси через верхнюю поверхность. Глубина обеих диффузий составляет половину толщины пластины, так что обе диффузионные области смыкаются. В верхней части пластины образуются «островки» исходного кремния n-типа, которые являются коллекторными слоями будущих транзисторов. При данном методе, в отличие от предыдущего, коллекторный слой оказывается однородным. Главный недостаток этого метода — необходимость проведения глубокой диффузии (100–150 мкм). Время такой диффузии составляет 2–3 суток и более, что экономически невыгодно. Кроме того, из-за боковой диффузии (см. рис. 6.5, *б*) протяженность изолирующих p-слоев на поверхности оказывается порядка толщины пластины, т.е. превышает размеры обычных транзисторов. Соответственно уменьшается коэффициент использования площади.

В настоящее время вместо монолитной пластины n-типа используют тонкий эпитаксиальный n-слой, выращенный на подложке (рис. 7.5). В этом случае проблема изоляции существенно упрощается: диффузия, обеспечивающая образование коллекторных «островков» (ее называют *разделительной* или *изолирующей диффузией*), проводится только через одну (верхнюю) поверхность пластины на глубину, равную толщине эпитаксиального слоя (обычно не более 5–10 мкм). При такой малой глубине время диффузии не превышает 2–3 ч, а расширение изолирующих p-слоев на поверхности, обусловленное боковой диффузией, в несколько раз меньше, чем при методе встречной диффузии (рис. 7.4, *б*). При этом коэффициент использования площади кристалла имеет приемлемую величину.

«Островки» n-типа, оставшиеся в эпитаксиальном слое после разделительной диффузии, называют *карманами*. В этих карманах на последующих этапах технологического процесса осуществляют необходимые элементы ИС, в первую очередь — транзисторы[1].

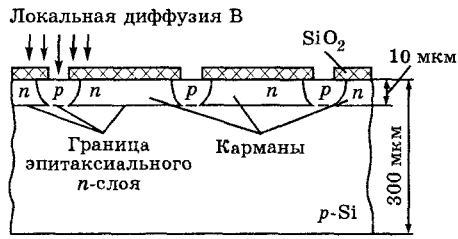


Рис. 7.5. Основной способ изоляции элементов с помощью переходов в планарно-эпитаксиальных ИС

Простейшие карманы, показанные на рис. 7.5, находят ограниченное применение. Транзисторы, изготовленные в таких карманах (рис. 7.6, *а*), характерны большим горизонтальным сопротивлением коллекторного слоя $r_{кк}$ (100 Ом и более). Уменьшать удельное сопротивление эпитаксиального слоя нерационально: при этом уменьшается пробивное напряжение коллекторного перехода и возрастает коллекторная емкость. Поэтому типовым решением является использование так называемого скрытого n-слоя, расположенного на дне кармана. Роль такого низкоомного слоя очевидна из рис. 7.6, *б*.

---

1 Слой n^+ под коллекторным электродом, получаемый на этапе эмиттерной диффузии (т.е. одновременно с эмиттерным n^+-слоем), предотвращает образование паразитного p–n-перехода при вжигании алюминия в n-слой (см. с. 202)

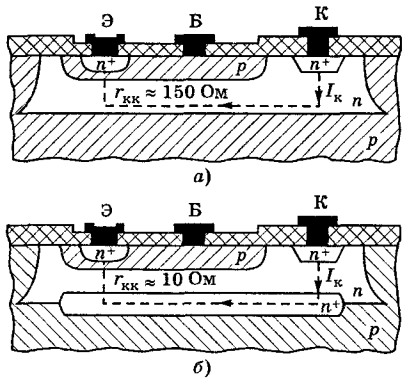


Рис. 7.6. Структура интегрального n–p–n-транзистора: *а* — без скрытого слоя, *б* — со скрытым n -слоем

Скрытые слои получают диффузией, которая проводится до наращивания эпитаксиального слоя. Во время эпитаксии донорные атомы скрытого слоя под действием высокой температуры диффундируют в нарастающий эпитаксиальный n-слой. В итоге скрытый слой частично расположен в эпитаксиальном, и дно кармана в этом месте оказывается «приподнятым» на несколько микрон. Для того, чтобы предотвратить чрезмерное распространение доноров из скрытого слоя в эпитаксиальный (что может привести к смыканию скрытого n^+-слоя с базовым p-слоем), выбирают для скрытого слоя диффузант со сравнительно малым коэффициентом диффузии — сурьму или мышьяк.

Использование скрытого n^+-слоя обеспечивает не только меньшее сопротивление коллектора (первоначальное назначение скрытого слоя), но и некоторые другие преимущества, в том числе большой инверсный коэффициент усиления транзистора и меньший избыточный заряд в коллекторном слое, накапливающийся в режиме двойной инжекции.

Разделительная диффузия в эпитаксиальный слой является в настоящее время наиболее простым и распространенным вариантом изоляции p–n-переходом. Однако используются и более сложные варианты этого метода, к числу которых относится *изоляция коллекторной диффузией* (ИКД, см. рис. 7.7).

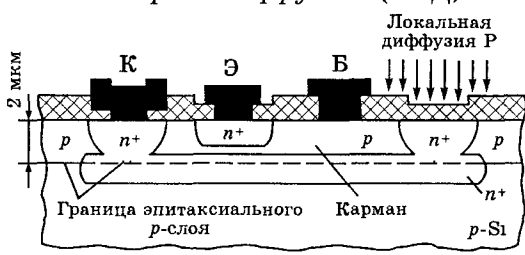


Рис. 7.7. Метод изоляции элементов коллекторной диффузией

Метод ИКД характерен тем, что эпитаксиальный слой (толщиной 2–3 мкм) имеет проводимость p-типа. В этом слое расположены заранее созданные скрытые n^+-слои. Разделительная диффузия в данном случае осуществляется с помощью донорной примеси (фосфора); глубина диффузии соответствует расстоянию от поверхности до скрытого слоя. В результате получаются карманы p-типа (будущие базовые слои), а скрытый n^+-слой вместе с разделительными n^+-слоями образует область коллектора. Как видим, в данном случае разделительные слои выполняют полезную функцию и, следовательно, не влияют на коэффициент использования площади. Последний при методе ИКД оказывается значительно больше, чем в основном варианте (рис. 7.5). Однако из-за большой концентрации примеси в n^+-слоях методу ИКД свойственны меньшие пробивные напряжения коллекторного перехода и большие значения коллекторной емкости. Кроме того, чтобы сделать базу неоднородной и тем самым создать в ней внутреннее ускоряющее поле, в p-карманы приходится дополнительно проводить диффузию акцепторной примеси.

Изоляция диэлектриком. Исторически первым способом изоляции диэлектриком был так называемый *эпик-процесс*. Этапы этого процесса показаны на рис. 7.8. Исходная пластина кремния n-типа покрывается тонким (2–3 мкм) эпитаксиальным слоем n^+-типа (рис. 7.8, *а*). Через маску в пластине вытравливают канавки глубиной 10–15 мкм, после чего всю рельефную поверхность окисляют (рис. 7.8, *б*)[1]. Далее на окисленную рельефную поверхность напыляют толстый (200–300 мкм) слой поликристаллического кремния (рис. 7.8, *в*). После этого исходную пластину n-типа сошлифовывают на всю толщину вплоть до дна канавок. В результате получаются карманы n-типа (со скрытым n^+-слоем), расположенные уже в поликристаллической подложке (рис. 7.8, *г*). Изоляция элементов обеспечивается окисным слоем SiO_2 (ср. с рис. 7.3, *б*). Основную трудность в эпик-процессе представляет прецизионная шлифовка монокристаллической пластины: при толщине сошлифовываемого слоя 200–300 мкм погрешность шлифовки по всей поверхности должна лежать в пределах 1–2 мкм.

---

1 Можно использовать как изотропное, так и анизотропное травление (см. раздел 6 6). Особенности последнего рассмотрены ниже.

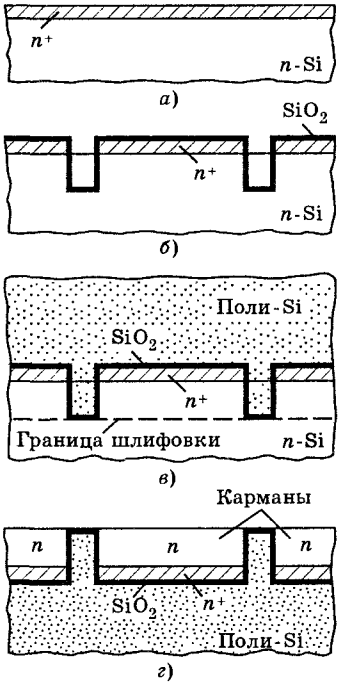


Рис. 7.8. Метод диэлектрической изоляции (эпик-процесс): *а* — исходная структура; *б* — травление канавок и окисление; *в* — напыление полнкристаллического кремния; *г* — конечная структура (карманы со скрытым n^+-слоем)

Если на 2-ом этапе (рис. 7.8, *б*) напылять не полупроводниковый слой (поликристаллический кремний), а диэлектрический — керамику, получается вариант *керамической изоляции*. Этот вариант обеспечивает лучшую гальваническую и емкостную «развязку» элементов, но он сложнее и дороже.

Большое распространение получила так называемая технология *кремний на сапфире* (КНС, англ. SOS — Silicon On Sapphire). Ее основные этапы показаны на рис. 7.9. Сапфир имеет такую же структуру кристаллической решетки, как и кремний. Поэтому на сапфировой пластине (подложке) можно нарастить эпитаксиальный слой кремния (рис. 7.9, *а*), а затем протравить этот слой насквозь до сапфира, так чтобы образовались кремниевые «островки — карманы» для будущих элементов ИС (рис. 7.9, *б*). Эти карманы с нижней стороны изолированы друг от друга сапфиром — диэлектриком, а с боковых сторон — воздухом. Поэтому технологию КНС часто относят к классу *воздушной изоляции*. Недостатком этого метода является рельефность поверхности, которая затрудняет осуществление металлической разводки.

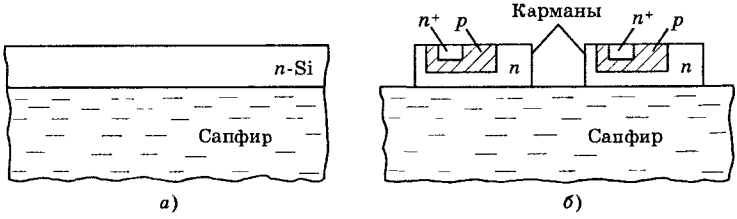


Рис. 7.9. Технология «кремний на сапфире» (КНС): *а* — исходная структура; *б* — рельефные карманы

Комбинированные способы изоляции. В настоящее время самым распространенным комбинированным методом является так называемая *изопланарная технология* (изопланар). В ее основе лежит локальное сквозное прокисление тонкого (2–3 мкм) эпитаксиального слоя кремния n-типа (рис. 7.10). В результате локального прокисления эпитаксиальный n-слой оказывается разделенным на отдельные карманы n-типа — так же, как при методе разделительной диффузии (рис. 7.5). Однако в данном случае боковые изолирующие слои не полупроводниковые, а диэлектрические (окисные). Что касается донных частей карманов, то они по-прежнему разделены встречно-включенными p–n-переходами. Именно поэтому изопланар относится к комбинированным методам.

Каждый карман в свою очередь разделен окислом на две части (*1* и *2* на рис. 7.10, *а*). В главной части (*1*) осуществляются база и эмиттер транзистора, а во второй (*2*) — омический контакт коллектора (рис. 7.10, *б*). Обе части связаны через скрытый n^+-слой. Таким образом устраняются все четыре боковые (вертикальные) стенки коллекторного перехода, что способствует уменьшению коллекторной емкости.

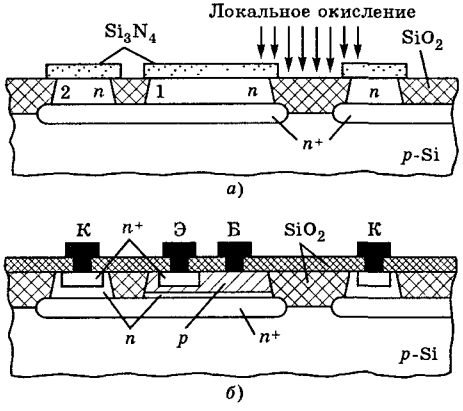


Рис. 7.10. Изопланарная технология: *а* — структура до базовой диффузии; *б* — конечная структура транзистора

Локальное прокисление эпитаксиального слоя нельзя проводить через окисную маску, так как при высокой температуре и при наличии кислорода ее толщина будет расти за счет толщины n-слоя. Поэтому в изопланаре для локального прокисления используют маски из нитрида кремния.

По сравнению с классическим методом разделительной диффузии изопланарный метод изоляции обеспечивает большую плотность компоновки элементов (т.е. лучшее использование площади), а также более высокие частотные и переходные характеристики транзисторов.

Метод изоляции V-*канавками* показан на рис. 7.11, *а*. В данном случае вместо сквозного прокисления эпитаксиального слоя используется его сквозное протравливание методом анизотропного травления (см. раздел 6.6). При этом поверхность кристалла должна иметь ориентацию (100), а травление идет по плоскостям (111), см. рис. 6.10, *б*. Размеры окна в маске делают такими, что грани (111) «сходятся» чуть ниже границы эпитаксиального слоя и образуют канавки V-образной формы, откуда и название метода. Зависимость между шириной и глубиной канавки строго определенная: $l/d = \sqrt{2}$. При глубине 4–5 мкм ширина канавки составит всего 6–7 мкм, т.е. потери площади под изоляцию получаются весьма незначительными. Недостатком метода является необходимость использования плоскости (100), которой свойственна повышенная плотность поверхностных дефектов (см. рис. 2.5).

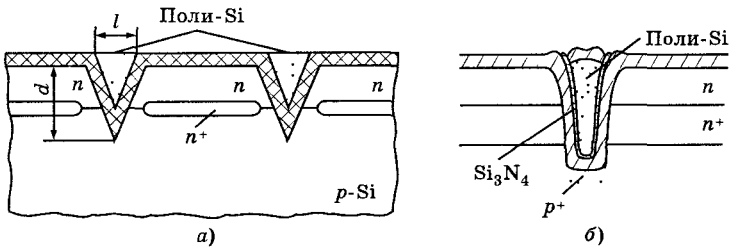


Рис. 7.11. Метод изоляции V-канавками (*а*) и U-канавками (*б*)

Рельеф, полученный в результате травления, окисляется, как и при эпик-процессе. Однако, в отличие от эпик-процесса, последующее напыление поликристаллического кремния имеет целью лишь выровнять рельеф поверхности для облегчения металлической разводки. Для этого достаточно заполнить поликремнием только канавки.

При использовании метода реактивного ионного травления можно дополнительно уменьшить ширину канавки. Такой метод изоляции получил название *изоляции* U-*образными канавками* (рис. 7.11, *б*).

7.3. Транзисторы *n–p–n*

Поскольку *n–p–n*-транзисторы составляют основу биполярных ИС, мы рассмотрим их наиболее подробно, включая технологию изготовления. При этом будем считать, что изоляция осуществлена методом разделительной диффузии. Особенности, обусловленные другими методами изоляции, в необходимых случаях оговариваются.

Распределение примесей. На рис. 7.12 показано распределение примесей в слоях интегрального транзистора со скрытым n^+-слоем (см. рис. 7.6, *б*). Следует обратить внимание на то, что распределение эффективной концентрации акцепторов в базовом слое оказывается немонотонным. Соответственно немонотонным оказывается и распределение дырок. Справа от точки максимума градиент концентрации дырок отрицательный и внутреннее поле (по отношению к инжектированным электронам) является ускоряющим. Это характерно для всех дрейфовых транзисторов. Однако слева от точки максимума градиент концентрации положительный, а значнт поле является тормозящим. *Наличие участка с тормозящим полем приводит к некоторому увеличению результирующего времени пролета носи-*

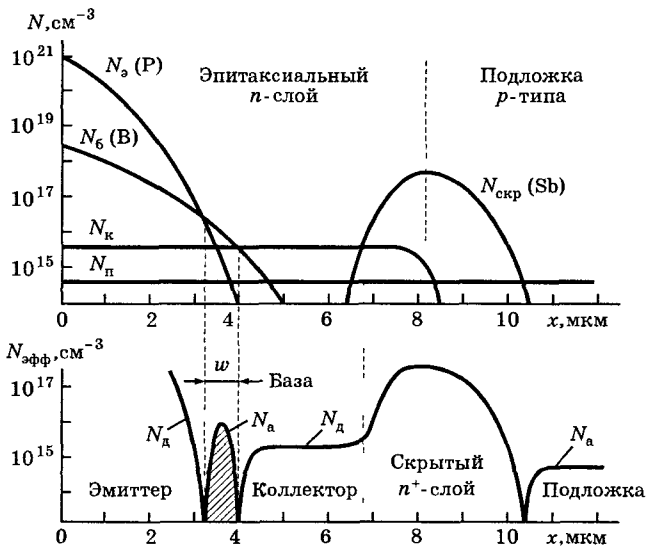


Рис. 7.12. Распределение концентрации примесей в структуре интегрального *n–p–n*-транзистора и распределение эффективных концентраций

телей через базу. Однако расчеты показывают, что это увеличение составляет всего 20–30% и для приближенных оценок может не учитываться.

Конфигурации и рабочие параметры. Конфигурации интегральных транзисторов (в плане) имеют несколько вариантов. Два из них показаны на рис. 7.13.

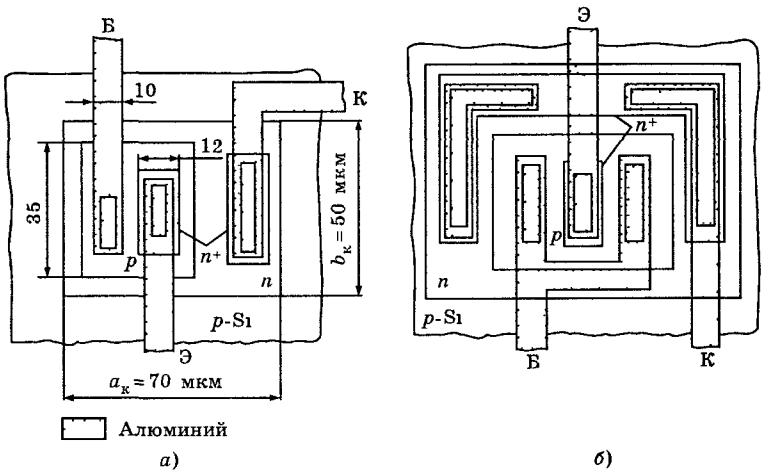


Рис. 7.13. Конфигурация (топология) транзисторов.
а — асимметричная, *б* — симметричная

Первая конфигурация (рис. 7.13, *а*) называется асимметричной: в ней коллекторный ток протекает к эмиттеру только в одном направлении: на рис. 7.13, *а* — справа. Вторая конфигурация (рис. 7.13, *б*) называется симметричной: в ней коллекторный ток протекает к эмиттеру с трех сторон. Соответственно сопротивление коллекторного слоя $r_{кк}$ оказывается примерно в 3 раза меньше, чем у асимметричной конфигурации.

Вторая конфигурация характерна также тем, что контактное окно и металлизация коллектора разбиты на две части. При такой конструкции облегчается металлическая разводка: алюминиевая полоска (например, эмиттерная на рис. 7.13, *б*) может проходить над коллектором по защитному окислу, покрывающему поверхность ИС (подробнее см. разд. 7.9).

Для примера на рис. 7.13, *а* приведены относительные размеры слоев интегрального n–p–n-транзистора для минимального литографического разрешения, равного 10 мкм. Для этого

случая в табл. 7.1 приведены типичные параметры этих слоев, а в табл. 7.2 — типичные параметры транзисторов.

Таблица 7.1. Типичные параметры слоев интегрального $n-p-n$-транзистора

Наименование слоя	N, см $^{-3}$	d, мкм	ρ, Ом·см	R_s, Ом/□
Подложка p-типа	$1{,}5\,10^{15}$	300	10	—
Скрытый n^+-слой	—	5–10	—	8–20
Коллекторный n-слой	10^{16}	10–15	0,5	500
Базовый p-слой	$5\cdot10^{18}$	2,5	—	200
Эмиттерный n^+-слой	10^{21}	2	—	5–15

Примечание: N — концентрация примеси (для диффузных базового и эмиттерного слоев — поверхностная концентрация); d — глубина слоя; ρ — удельное сопротивление материала; R_s — удельное сопротивление слоя.

Таблица 7.2. Типичные параметры интегральных $n-p-n$-транзисторов

Параметр	Номинал	Допуск δ, %
Коэффициент усиления B	100–200	±30
Предельная частота f_T, МГц	200–500	±20
Коллекторная емкость $C_к$, пФ	0,3–0,5	±10
Пробивное напряжение $U_{кб}$, В	40–50	±30
Пробивное напряжение $U_{эб}$, В	7–8	±5

Величину R_s, фигурирующую в табл. 7.1, называют *удельным сопротивлением слоя*. Происхождение этого параметра следующее. Пусть имеется прямоугольная полоска материала длиной a, шириной b и толщиной d. Если ток протекает вдоль полоски (т.е. параллельно ее поверхности), то сопротивление полоски можно записать в виде:

$$R = \rho(a/bd) = R_s(a/b), \qquad (7.1)$$

где $R_s = \rho/d$. Если слой неоднороден по толщине (например, если он получен диффузией примеси), то величина R_s запишется в общем виде:

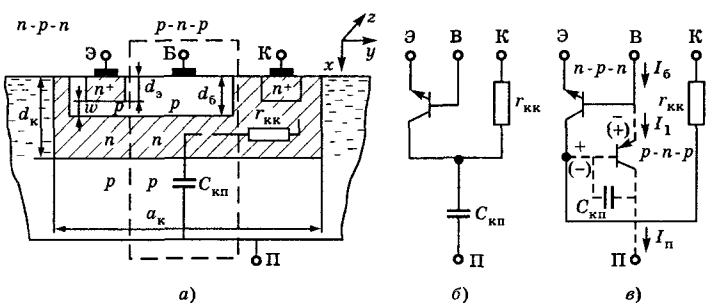
$$R_s = \left[\int_0^d \sigma(x)\,dx\right]^{-1},$$

где $\sigma(x) = 1/\rho(x)$ — удельная проводимость материала в плоскости, расположенной на расстоянии x от поверхности.

При условии $a = b$ прямоугольная полоска принимает квадратную форму, а ее сопротивление делается равным R_s. Значит, величину R_s можно определить как *продольное сопротивление слоя или пленки квадратной конфигурации*. Чтобы подчеркнуть последнюю оговорку, вместо истинной размерности «Ом» пишут «Ом/□» (читается: «Ом на квадрат»). Зная величину R_s, легко рассчитать сопротивление слоя или пленки прямоугольной конфигурации по известным значениям a и b.

Из табл. 7.2 видно, что пробивное напряжение эмиттерного перехода в 5–7 раз меньше, чем коллекторного. Эта особенность, свойственная всем дрейфовым транзисторам, связана с тем, что эмиттерный переход образован более низкоомными слоями, чем коллекторный. При включении транзистора с общим эмиттером пробойное напряжение коллекторного перехода уменьшается в соответствии с (5.29). Если база достаточно тонкая ($w < 1$ мкм), то пробой обычно обусловлен эффектом смыкания, а напряжение пробоя характеризуется выражением (5.30).

Паразитные параметры. На рис. 7.14, *а* показана упрощенная структура интегрального n–p–n-транзистора, выполненного по методу разделительной диффузии. Особенность интегрального транзистора состоит в том, что его структура (с учетом подложки) — четырехслойная: наряду с рабочими эмиттерным и коллекторным переходами имеется третий (паразитный) переход между коллекторным n-слоем и подложкой p-типа. Наличие скрытого n^+-слоя (не показанного на рис. 7.14, *а*) не вносит принципиальных изменений в структуру.



а) *б*) *в*)

Рис. 7.14. Интегральный n–p–n-транзистор:
а — упрощенная структура с выделенным паразитным p–n–p-транзистором; *б* — упрощенная модель; *в* — полная модель

Подложку ИС (если она имеет проводимость p-типа) *присоединяют к самому отрицательному потенциалу*. Поэтому напряжение на переходе «коллектор—подложка» всегда обратное или (в худшем случае) близко к нулю. Следовательно, этот переход можно заменить барьерной емкостью $C_{кп}$, показанной на рис. 7.14, *а*.

Вместе с горизонтальным сопротивлением коллекторного слоя $r_{кк}$ емкость $C_{кп}$ образует *RC*-цепочку, которая подключена к активной области коллектора. Тогда эквивалентная схема интегрального *n–p–n*-транзистора имеет такой вид, как показано на рис. 7.14, *б*.

Цепочка $r_{кк}$– $C_{кп}$, *шунтирующая коллектор, — главная особенность интегрального n–p–n-транзистора*. Эта цепочка, естественно, ухудшает его быстродействие и ограничивает предельную частоту и время переключения.

Поскольку подложка находится под неизменным потенциалом, ее можно считать заземленной по переменным составляющим. Поэтому, дополняя малосигнальную эквивалентную схему ОБ (рис. 5.16) цепочкой $r_{кк}C_{кп}$ и пренебрегая сопротивлением $r_б$, приходим к выводу, что емкость $C_{кп}$ складывается с емкостью $C_к$, а сопротивление $r_{кк}$ — с внешним сопротивлением $R_к$ (см. с. 163). Соответственно эквивалентная постоянная времени (5.68) запишется следующим образом:

$$\tau_{\alpha oe} = \tau_\alpha + (C_к + C_{кп})(r_{кк} + R_к). \qquad (7.2)$$

Из выражения (7.2) очевидно, что паразитные параметры $C_{кп}$ и $r_{кк}$ ограничивают быстродействие интегрального транзистора в идеальных условиях, когда $\tau_\alpha = 0$, $C_к = 0$ и $R_к = 0$. В этом случае эквивалентная постоянная времени $\tau_{\alpha oe}$ равна *постоянной времени подложки*:

$$\tau_п = C_{кп} r_{кк}. \qquad (7.3)$$

Например, если $C_{кп} = 2$ пФ и $r_{кк} = 100$ Ом, получаем $\tau_п = 0{,}2$ нс, соответствующая граничная частота $f_п = 1/2\pi\tau_п \approx 800$ МГц. С учетом параметров τ_α, $C_к$ и при наличии внешнего сопротивления $R_к$ эквивалентная постоянная времени возрастает, а граничная частота уменьшается.

Значение $r_{кк} = 100$ Ом, использованное в предыдущем примере, характерно для транзисторов без скрытого n^+-слоя. При наличии скрытого слоя типичны значения $r_{кк} = 10$ Ом. Тогда

постоянная времени $\tau_п$ оказывается на порядок меньше и влияние подложки становится мало существенным.

Соотношение между емкостями $C_{кп}$ и $C_к$ зависит в первую очередь от соотношения площадей соответствующих переходов и концентраций примеси в слоях подложки и коллектора. Обычно $C_{кп} = (2–3)C_к$.

При расчете емкости $C_{кп}$ следует учитывать не только донную часть перехода коллектор–подложка, но и его боковые (вертикальные) части (см. рис. 7.14, *а*). Удельная емкость боковых частей больше, чем донной, поскольку концентрация акцепторов в разделительных слоях возрастает в направлении от дна перехода к поверхности (на рис. 7.14, *а* эта концентрация характеризуется густотой штриховки). Типичное значение удельной емкости для донной части составляет $C_{0x} = 100$ пФ/мм^2, а для боковых частей $C_{0y,z} = 150–250$ пФ/мм^2. Обычно все три составляющие емкости $C_{кп}$ оказываются почти одинаковыми и лежат в пределах 0,5–1,5 пФ.

Пассивную область базы вместе с лежащими под ней областями коллектора и подложки можно представить как некий *паразитный* $p–n–p$-транзистор. На рис. 7.14, *а* структура такого транзистора обведена штриховой линией, а эквивалентная схема, характеризующая взаимосвязь рабочего $n–p–n$-транзистора с паразитным, показана на рис. 7.14, *в*.

Если $n–p–n$-транзистор работает в нормальном активном режиме ($U_{кб} > 0$), то паразитный транзистор находится в режиме отсечки ($U_{эб} < 0$, см. знаки без скобок). В этом случае коллекторный переход паразитного транзистора представлен емкостью $C_{кп}$ (рис. 7.14, *б*). Если же $n–p–n$-транзистор работает в инверсном режиме или в режиме двойной инжекции ($U_{эб} < 0$), то паразитный $p–n–p$-транзистор находится в активном режиме ($U_{эб} > 0$, см. знаки в скобках). При этом в подложку уходит ток $I_п = \alpha_{p–n–p}I_1$, где I_1 — часть базового тока (рис. 7.14, *в*).

Утечка базового тока в подложку ухудшает параметры транзистора в режиме двойной инжекции (см. раздел 8.2). Поэтому транзисторы, предназначенные для работы в таком режиме, специально легируют золотом. Атомы золота играют в кремнии роль ловушек, т.е. способствуют уменьшению времени жизни носителей. Соответственно коэффициент $\alpha_{p–n–p}$ уменьшается до значений менее 0,1, и утечкой тока в подложку можно пренебречь.

В случае диэлектрической изоляции паразитный $p–n–p$-транзистор отсутствует, но емкость $C_{кп}$ сохраняется. Она, как уже отмечалось, меньше, чем при изоляции $p–n$-переходом. Если ди-

электриком является двуокись кремния, то удельная емкость при толщине 1 мкм составляет около 35 пФ/мм^2.

Типовой технологический цикл. Промышленность поставляет разработчику ИС готовые пластины кремния, прошедшие механическую и химическую обработку. Поэтому будем считать, что в начале технологического цикла имеется пластина кремния p-типа с отполированной поверхностью, покрытой тонким, естественным слоем окисла. В этой пластине групповым методом нужно изготовить транзисторы со структурой, показанной на рис. 7.6, *б*. Последовательность операций будет следующая[1].

1) Общее окисление пластины.

2) 1-я фотолитография: создание окон в окисле «под скрытые n^+-слои».

3) 1-я диффузия (создание скрытых n^+-слоев, рис. 7.15, *а*); диффузант — мышьяк или сурьма.

4) Стравливание окисла со всей поверхности.

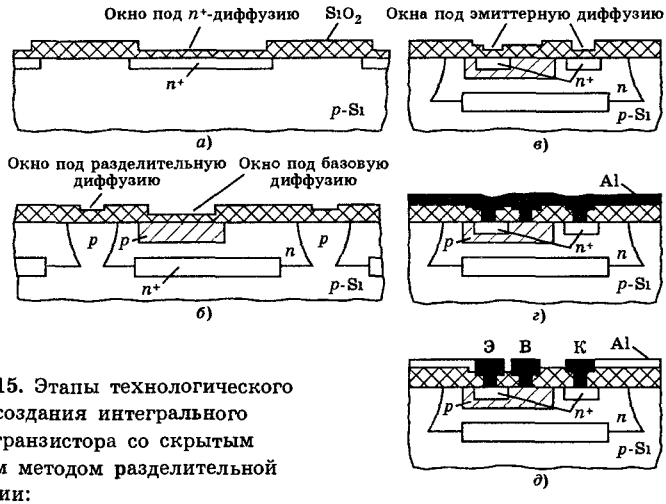


Рис. 7.15. Этапы технологического цикла создания интегрального n–p–n-транзистора со скрытым n^+-слоем методом разделительной диффузии:
а — создание скрытых слоев; *б* — создание базовых слоев в эпитаксиальных карманах; *в* — создание эмиттерных слоев и слоев под омические контакты коллекторов; *г* — общая металлизация; *д* — создание рисунка металлической разводкой

1 В приводимом перечне опущены многочисленные операции очистки и отмывки пластины, а также нанесения и удаления фоторезиста.

5) Наращивание эпитаксиального n-слоя (при этом скрытый n^+-слой несколько диффундирует как в подложку, так и в эпитаксиальный слой; см. с. 222).

6) Общее окисление.

7) 2-я фотолитография: создание окон в окисле «под разделительную диффузию».

8) 2-я диффузия (создание разделительных p-слоев и соответственно изолированных n-карманов в эпитаксиальном слое, рис. 7.5); диффузант — бор.

9) 3-я фотолитография: создание окон в окисле «под базовую диффузию».

10) 3-я диффузия (создание базовых p-слоев, рис. 7.15, *б*); диффузант — бор. Диффузия двухстадийная («загонка» и «разгонка», см. с. 179–180).

11) 4-я фотолитография: создание окон в окисле «под эмиттерную диффузию и омические контакты коллекторов».

12) 4-я диффузия (создание n^+-слоев, рис. 7.15, *в*); диффузант — фосфор. Иногда эта диффузия тоже двухстадийная.

13) 5-я фотолитография: создание окон в окисле «под омические контакты».

14) Общее напыление алюминия на пластину (рис. 7.15, *г*).

15) 6-я фотолитография: создание окон в фоторезисте «под металлическую разводку».

16) Травление алюминия через фоторезистную маску, снятие фоторезиста (рис. 7.15, *д*).

17) Термическая обработка для вжигания алюминия в кремний.

Сборочные операции мы опускаем: они были рассмотрены в разделе 6.10. К рассмотренному технологическому циклу необходимо сделать несколько дополнительных замечаний.

В последнее время промышленность начала выпускать пластины с уже осуществленными эпитаксиальным слоем и скрытым n^+-слоем. В таком случае первые пять операций отпадают.

В п.10 отмечено, что диффузия бора на этапе базовой диффузии — двухстадийная («загонка» и «разгонка»). Такое, казалось бы, усложнение процесса имеет серьезные основания и является общепринятым.

Действительно, для того чтобы коэффициент инжекции эмиттерного перехода составлял не менее 0,999, концентрация примеси в эмиттерном слое должна превышать концентрацию в базовом слое не менее

чем в 100 раз [см. (5.24)]. Между тем предельные растворимости бора и фосфора при оптимальных температурах различаются всего в 3 раза (см. табл. 6.1). Для того чтобы преодолеть это противоречие, нужно понизить приповерхностную концентрацию бора. Это можно сделать несколькими способами.

Можно проводить диффузию бора при столь низкой температуре, при которой его предельная растворимость будет в 100 раз меньше, чем у фосфора; однако тогда коэффициент диффузии уменьшится на несколько порядков и диффузию придется проводить в течение ряда суток или даже недель. Можно понизить температуру в зоне источника диффузанта и таким образом создать «диффузантный голод» вблизи поверхности пластины; однако этот процесс трудно контролируется. Таким образом, двухстадийная диффузия оказывается оптимальным решением: во время «разгонки» поверхностную концентрацию легко уменьшить в десятки раз и более (см. рис. 6.7, *б*).

Температуру «разгонки» делают на 150–200 °С больше, чем температура «загонки», чтобы повысить коэффициент диффузии примеси и сократить время процесса. Типичное время «загонки» составляет 20-40 мин (при температуре 1000–1050 °С), а «разгонки» — несколько часов (при температуре около 1200 °С).

Внедрение фосфора на этапе эмиттерной диффузии (п.12) — последняя высокотемпературная операция в цикле (температура выбирается на 100–150 °С ниже температуры «разгонки» бора, чтобы не изменить глубину коллекторного p–n-перехода). Длительность этой операции определяет толщину n^+-слоя, а значит, и ширину базы транзистора. У современных планарных транзисторов типичная ширина базы составляет 0,4–0,5 мкм.

В заключение заметим, что в результате неоднократных операций фотолитографии, окисления и диффузии рельеф окисной пленки перед металлизацией оказывается сложным — многоступенчатым. В ряде случаев это затрудняет получение хорошей адгезии алюминия с поверхностью окисла. Обычно на рисунках, иллюстрирующих структуру (разрез) транзисторов или ИС, многоступенчатость окисла для простоты не показывают.

7.4. Разновидности n–p–n-транзисторов

В процессе развития микроэлектроники появились некоторые разновидности n–p–n транзисторов, не свойственные дискретным электронным схемам и не выпускаемые в виде дискретных приборов. Ниже рассматриваются наиболее важные из этих разновидностей.

Многоэмиттерный транзистор. Структура многоэмиттерного транзистора (МЭТ) показана на рис. 7.16, *а*. Такие транзисторы составляют основу весьма распространенного класса цифровых ИС — так называемых схем ТТЛ (см. разд. 10.2). Количество эмиттеров может составлять 5–8 и более.

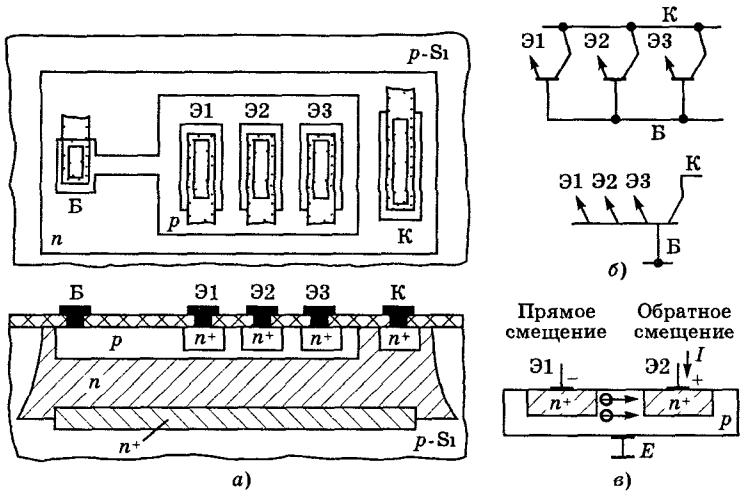


Рис. 7 16 Многоэмиттерный транзистор *а* — топология и структура, *б* — схемные модели, *в* — взаимодействие между смежными эмиттерами

В первом приближении МЭТ можно рассматривать как совокупность отдельных транзисторов с соединенными базами и коллекторами (рис. 7.16, *б*). Особенности МЭТ как единой структуры следующие (рис. 7.16, *в*).

Во-первых, каждая пара смежных эмиттеров вместе с разделяющим их p-слоем базы образует *горизонтальный* (иногда говорят — *продольный*) транзистор типа n^+–p–n^+. Если на одном из эмиттеров действует прямое напряжение, а на другом обратное, то первый будет инжектировать электроны, а второй будет собирать те из них, которые инжектированы через боковую поверхность эмиттера и прошли без рекомбинации расстояние между эмиттерами. Такой *транзисторный эффект* является для МЭТ паразитным: в обратносмещенном переходе, который должен быть запертым, будет протекать ток. Чтобы избежать горизонтального транзисторного эффекта, расстояние между эмиттерами, вообще говоря, должно превышать диффузионную длину носителей в базовом слое. Если транзистор легирован зо-

лотом (см. с. 232), то диффузионная длина не превышает 2–3 мкм и практически оказывается достаточным расстояние 10–15 мкм.

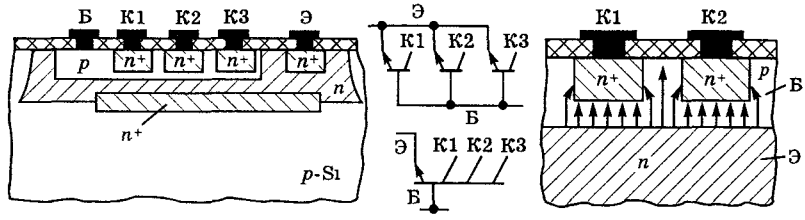
Во-вторых, важно, чтобы МЭТ имел как можно меньший инверсный коэффициент передачи тока. В противном случае в инверсном режиме, когда эмиттеры находятся под обратным напряжением, а коллектор под прямым, носители, инжектируемые коллектором, будут в значительной мере достигать эмиттеров, и в цепи последних, несмотря на их обратное смещение, будет протекать ток — паразитный эффект, аналогичный отмеченному выше.

Как известно, инверсный коэффициент передачи всегда меньше нормального из-за различий в степени легирования и в площадях эмиттера и коллектора (см. с. 142). Чтобы дополнительно уменьшить инверсный коэффициент α_I в МЭТ, искусственно увеличивают сопротивление пассивной базы, удаляя омический базовый контакт от активной области транзистора (рис. 7.16, *а*). При такой конфигурации сопротивление узкого «перешейка» между активной областью и базовым контактом может составлять 200–300 Ом, а падение напряжения на нем от базового тока 0,1–0,15 В. Значит, прямое напряжение на коллекторном переходе (в инверсном режиме) будет в активной области на 0,1–0,15 В меньше, чем вблизи базового контакта. Соответственно инжекция электронов из коллектора в активную область базы будет незначительной и паразитные токи через эмиттеры будут практически отсутствовать.

Многоколлекторные n–p–n-транзисторы. Структура многоколлекторного транзистора (МКТ), показанная на рис. 7.17, *а*, не отличается от структуры МЭТ. Различие состоит лишь в использовании структуры. Можно сказать, что МКТ — *это МЭТ, используемый в инверсном режиме.* общим эмиттером является эпитаксиальный n-слой, а коллекторами служат высоколегированные n^+-слои малых размеров. Такое решение составляет основу одного из популярных классов цифровых ИС — так называемых схем инжекционной логики И^{2}Л (см. раздел 10.3)[1]. Эквивалентная схема МКТ показана на рис. 7.17, *б*.

---

1 Специфичность МКТ в схемах И^{2}Л — способ питания, но их основные свойства такие же, как рассматриваемые ниже



а) *б*) *в*)

Рис. 7.17. Многоколлекторный транзистор: *а* — структура; *б* — схемные модели; *в* — траектория движения инжектированных носителей

Главной проблемой при разработке МКТ является увеличение нормального коэффициента передачи тока от общего n-эмиттера (инжектора) к каждому из n^+-коллекторов. Естественно, что эта проблема — обратная той, которая решалась в случае МЭТ, когда коэффициент передачи от n-слоя к n^+-слоям старались уменьшать.

В данном случае желательно, чтобы скрытый n^+-слой располагался как можно ближе к базовому или просто контактировал с ним (как, например, при ИКД технологии). Тогда этот высоколегированный n^+-слой, будучи эмиттером, обеспечит высокий коэффициент инжекции. Что касается коэффициента переноса, то для его повышения n^+-коллекторы следует располагать как можно ближе друг к другу, сокращая тем самым площадь пассивной области базы. Оба эти пути, конечно, ограничены конструктивно-технологическими факторами. Тем не менее, даже при сравнительно разреженном расположении коллекторов, можно получить коэффициенты передачи на всю совокупность коллекторов $\alpha = 0{,}8–0{,}9$ или коэффициенты усиления $B = 4–10$. Этого достаточно для функционирования схем $И^2Л$, если число коллекторов не превышает 3–5[1].

На рис. 7.17, *в* показаны траектории движения инжектированных носителей в базе. Как видим, носители двигаются так, что их доля, попадающая на коллекторы, существенно больше, чем если ее рассчитывать по формальному отношению площади коллектора к площади эмиттера. Именно поэтому реальный ко-

1 Коэффициент усиления в расчете на один коллектор равен общему коэффициенту усиления, поделенному на число коллекторов. Для приведенных значений общего коэффициента B коэффициент усиления на один коллектор превышает единицу, что и требуется в схемах $И^2Л$.

эффициент B имеет те сравнительно большие значения, которые приведены выше. Следовательно, при расчете коэффициентов α и B нужно использовать не геометрические, а эффективные площади, о чем уже говорилось на с. 142 применительно к инверсному включению транзисторов.

Из рис. 7.17, *в* видно также, что средняя длина траектории носителей значительно превышает толщину активной базы w. Поэтому среднее время диффузии будет значительно меньше, чем у МЭТ и отдельных транзисторов [см. (5.47)]. Разница во временах пролета еще больше, поскольку *в МКТ поле базы для инжектированных носителей является не ускоряющим, а тормозящим.* Время пролета $t_{пр}$ составляет не менее 5–10 нс, а соответствующая предельная частота $f_т$ — не более 20–50 МГц (ср. с параметрами в табл. 7.2).

С другой стороны, коллекторная емкость $C_к$ у МКТ значительно меньше, чем у МЭТ и обычных транзисторов, из-за малой площади n^+-коллектора. Поэтому в выражениях (5.68) и (5.69) членами $C_к R_к$ и $C^* R_к$ часто можно пренебречь.

Транзистор с барьером Шоттки. Назначение и принцип действия такого транзистора (ТШ) рассмотрены в разделе 8.4. На рис. 7.18 показана структура интегрального ТШ. Здесь очень изящно решена задача сочетания транзистора с диодом Шоттки: алюминиевая металлизация, обеспечивающая омический контакт с p-слоем базы, проделана в сторону коллекторного n-слоя. На первый взгляд, коллекторный слой оказался закороченным со слоем базы. На самом же деле алюминиевая полоска образует с p-слоем базы невыпрямляющий, омический контакт, а с n-слоем коллектора выпрямляющий контакт Шоттки (см. раздел 3.3). Поэтому эквивалентная схема такой структуры соответствует схеме на рис. 8.12.

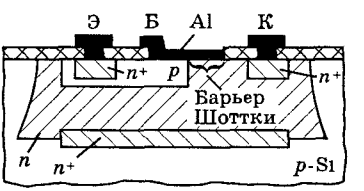


Рис. 7.18. Интегральный транзистор с барьером Шоттки

Разумеется, структурное решение, показанное на рис. 7.18, можно использовать не только в простейшем транзисторе, но и в МЭТ. В обоих случаях отсутствуют накопление и рассасывание избыточных зарядов и получается существенный (в 1,5–2 раза) выигрыш во времени переключения транзисторов из полностью открытого в запертое состояние.

Супербета транзистор. Это название присвоено транзисторам со сверхтонкой базой: $w = 0{,}2–0{,}3$ мкм. При такой ширине базы коэффициент усиления базового тока составляет $B = 3000–5000$ и более, откуда и следует их название.

Получение сверхтонкой базы представляет серьезную технологическую проблему. Во-первых, ширина базы есть разность глубин базового и эмиттерного слоев: $w = d_б - d_э$. Если допуск на ширину базы составляет ± 10%, т.е. 0,02 мкм, то при глубине базового слоя $d_б = 2$ мкм глубина эмиттерного слоя должна составлять $d_э = 1{,}8 \pm 0{,}02$ мкм. Значит, эмиттерная диффузия должна осуществляться с допуском ± 1,25%, что лежит на пределе технологических возможностей. Во-вторых, когда в процессе диффузии эмиттерного слоя его металлургическая граница приближается к металлургической границе коллекторного слоя на расстояние 0,4 мкм, наступает так называемый *эффект оттеснения* коллекторного перехода: дальнейшая диффузия атомов фосфора в эмиттерном слое сопровождается диффузией (с той же скоростью) атомов бора в базовом слое. Можно сказать, что эмиттерный слой «продавливает» металлургическую границу ранее полученного базового слоя (рис. 7.19). При этом толщина базы сохраняет значение около 0,4 мкм. Для того, чтобы преодолеть отмеченные трудности и обеспечить воспроизводимость ширины базы, потребовались многолетние усилия физиков-технологов. *Большой коэффициент усиления у супербета транзисторов покупается ценой очень малого пробивного напряжения* (1,5–2 В). Это — результат эффекта смыкания переходов, свойственного транзисторам с тонкой базой (см. разд. 4.4). Поэтому *супербета транзисторы являются не универсальными, а специализированными элементами ИС.* Их главная область применения — входные каскады операционных усилителей (раздел 10.10).

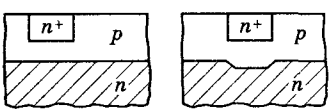


Рис. 7.19. Иллюстрация проблемы получения сверхтонкой базы

Необходимо заметить, что дальнейшее уменьшение ширины базы до 0,1 мкм и менее связано уже не столько с технологическими, сколько с принципиальными физическими проблемами. А именно, если принять среднюю концентрацию акцепторов в базе равной $8 \cdot 10^{15}$ см$^{-3}$ (см. рис. 7.12), то на 1 см длины их приходится $2 \cdot 10^5$. При ширине базы 0,1 мкм (т.е. 10^{-5} см)

оказывается, что в базе располагаются всего два слоя акцепторных атомов. При этом теряет смысл понятие градиента концентрации примеси (и связанное с ним понятие внутреннего поля), качественно меняются процессы рассеяния и характер движения носителей в базе. Тем самым классическая теория транзисторов в значительной мере теряет силу.

7.5. Транзисторы p–n–p

Получение p–n–p-транзисторов с такими же высокими параметрами, как и n–p–n-транзисторы, в едином технологическом цикле остается до сих пор нерешенной задачей. Поэтому все существующие варианты интегральных p–n–p-транзисторов существенно уступают n–p–n-транзисторам по коэффициенту усиления и предельной частоте.

Как известно, при прочих равных условиях p–n–p-транзисторы уступают n–p–n-транзисторам по предельной частоте примерно в 3 раза из-за меньшей подвижности дырок по сравнению с электронами. Поэтому в данном случае, говоря о меньшей предельной частоте p–n–p-транзисторов, мы имеем в виду, что не удается обеспечить те «равные условия», при которых различие было бы только в 3 раза.

На ранней стадии развития ИС в качестве p–n–p-транзисторов использовались p–n–p-структуры, образованные слоями базы, коллектора и подложки (рис. 7.20, *а*). Такие транзисторы обычно называют *паразитными* по аналогии с теми, которые входят в состав транзисторов n–p–n (см. рис. 7.14, *а*). Создание паразитного p–n–p-транзистора не требует дополнительных технологических операций, однако его параметры оказываются крайне низкими из-за большой ширины базы (сравнимой с глубиной эпитаксиального слоя) и слабой степени легирования эмиттера. Практически толщина базы составляет около 10 мкм, что соответствует предельной частоте f_T = 1–2 МГц, а коэффициент усиления B обычно не превышает 2–3.

Не менее существенным недостатком паразитных p–n–p-транзисторов является то, что разделительный p^+-слой связан с подложкой, а через нее с другими разделительными слоями. Следовательно, *коллекторы всех таких p–n–p-транзисторов оказываются соединенными друг с другом*, а это сильно ограничивает сферу их применения (ср. с рис. 7.1).

Вводя дополнительную технологическую операцию — глубокую диффузию акцепторов (рис. 7.20, *б*), можно получить меньшую толщину базы и улучшить параметры (B = 8–10 и f_T = 4–6 МГц). Однако

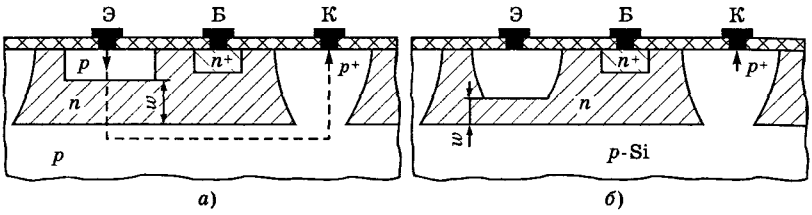


Рис. 7.20. Паразитные p–n–p-транзисторы: *а* — эмиттер, полученный на этапе базовой диффузии n–p–n-транзистора; *б* — эмиттер, полученный специально глубокой диффузией бора

увеличивается время проведения диффузии и сохраняется недостаток, обусловленный связью коллекторов через подложку.

В настоящее время основным структурным вариантом p–n–p-транзистора является горизонтальный p–n–p-транзистор (рис. 7.21). Он изолирован от других элементов, имеет гораздо лучшие параметры, чем паразитный p–n–p-транзистор, а его технология полностью вписывается в классический технологический цикл с разделительной диффузией.

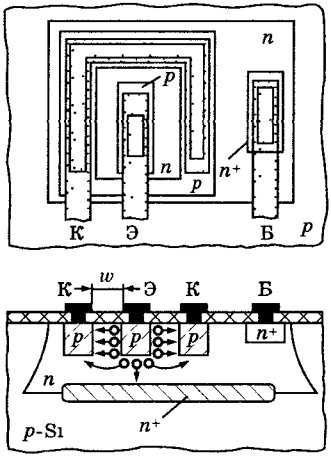


Рис. 7.21. Горизонтальный p–n–p-транзистор (топология и структура)

Эмиттерный и коллекторный слои получаются на этапе базовой диффузии, причем коллекторный слой охватывает эмиттер со всех сторон. Это позволяет собирать инжектированные дырки со всех боковых частей эмиттерного слоя. Приповерхностные боковые участки p-слоев характерны повышенной концентрацией примеси, что способствует увеличению коэффициента инжекции. Поскольку базовая диффузия сравнительно мелкая (2–3 мкм), ширину базы (т.е. расстояние между p-слоями) удается сделать порядка 3–4 мкм. В результате предельная частота может составлять до 20–40 МГц, а коэффициент усиления до 50.

Из рис. 7.21 видно, что горизонтальный p–n–p-транзистор (как и паразитный) является бездрейфовым, так как его база однородная — эпитаксиальный n-слой. Этот фактор вместе с

меньшей подвижностью дырок предопределяет примерно на порядок худшие частотные и переходные свойства p–n–p-транзистора даже при той же ширине базы, что и у дрейфового n–p–n-транзистора. Из рис. 7.21 также видно, что для увеличения коэффициента передачи эмиттерного тока желательно, чтобы площадь донной части эмиттерного слоя была мала по сравнению с площадью боковых частей. Значит, эмиттерный слой нужно делать как можно более узким.

Заметим, что горизонтальному p–n–p-транзистору свойственна электрофизическая симметрия, так как слои эмиттера и коллектора однотипные. В частности, это означает, что пробивные напряжения эмиттерного и коллекторного переходов — одинаковы (обычно 30-50 В). Близкими оказываются также нормальный и инверсный коэффициенты усиления тока B_N и B_I.

Горизонтальная структура позволяет легко осуществить многоколлекторный p–n–p-транзистор: достаточно разделить кольцевой p-коллектор (рис. 7.21) на n частей и сделать отдельные выводы от каждой части.

Коэффициент усиления по каждому из коллекторов будет примерно в n раз меньше, чем для единого коллектора, но все коллекторы будут действовать «синхронно», а нагрузки, присоединенные к ним, будут изолированы — «развязаны» друг от друга.

Главные недостатки горизонтального p–n–p-транзистора — сравнительно большая ширина базы и ее однородность. Эти недостатки можно устранить в вертикальной структуре (рис. 7.22), но ценой дополнительных технологических операций. Из рис. 7.22 следуют две такие операции: глубокая диффузия p-слоя и заключительная диффузия p^{++}-слоя. Последняя операция весьма проблематична, так как для получения p^{++}-слоя необходим акцепторный материал, у которого предельная растворимость больше, чем у фосфора, используемого для получения n^+-слоя. Поскольку такие материалы практически отсутствуют, верхнюю — наиболее легированную часть n^+-слоя необходимо стравить до осуществления p^{++}-слоя, что дополнительно усложняет технологический цикл.

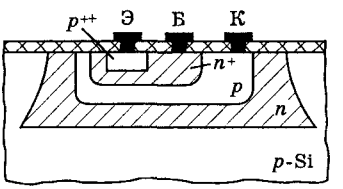


Рис. 7.22. Вертикальный p–n–p-транзистор

Большие возможности для получения качественных p–n–p-транзисторов открывает технология КНС («кремний на сапфире», см. разд. 7.2). В этом случае (рис. 7.23) p–n–p-транзистор изготавливается по существу отдельно от n–p–n-транзисторов, начиная с этапа эпитаксии p-слоя (эпитаксия n- и p-слоев осуществляется локально, через разные маски). Поэтому и ширина базы, и степень легирования эмиттерного слоя могут быть оптимизированы. Однако необходимость в локальной эпитаксии и два дополнительных процесса диффузии сильно усложняют и удорожают производство.

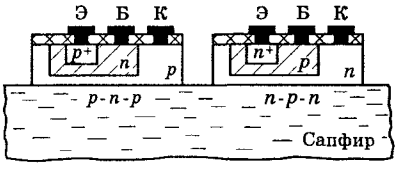


Рис. 7.23. Транзистор p–n–p-типа, полученный по технологии «кремний на сапфире» наряду с n–p–n-транзистором

7.6. Интегральные диоды

В качестве диода можно использовать любой из двух p–n-переходов, расположенных в изолирующем кармане: эмиттерный или коллекторный. Можно также использовать их комбинации. Поэтому по существу *интегральный диод представляет собой диодное включение интегрального транзистора.*

Пять возможных вариантов диодного включения транзистора показаны на рис. 7.24. В табл. 7.3 приведены типичные параметры этих вариантов. Для них приняты следующие обозначения: до черточки стоит обозначение анода, после черточки — катода; если два слоя соединены, их обозначения пишутся слитно. Из табл. 7.3 видно, что варианты различаются как по статическим, так и по динамическим параметрам.

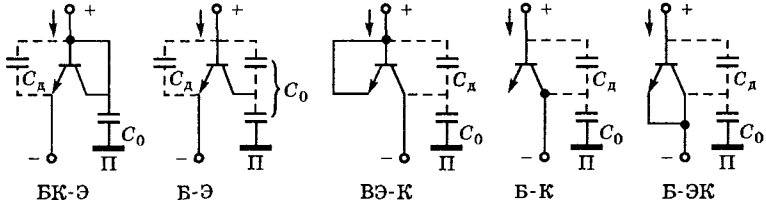


Рис. 7.24. Интегральные диоды (диодные включения транзистора)

Пробивные напряжения $U_{пр}$ зависят от используемого перехода: они меньше у тех вариантов, в которых используется эмиттерный переход (см. табл. 7.2).

Обратные токи $I_{обр}$ (без учета токов утечки) — это токи термогенерации в переходах. Они зависят от объема перехода и, следовательно, меньше у тех вариантов, у которых используется только эмиттерный переход, имеющий наименьшую площадь.

Емкость диода $C_д$ (т.е. емкость между анодом и катодом) зависит от площади используемых переходов; поэтому она максимальна при их параллельном соединении (вариант Б–ЭК). Паразитная *емкость на подложку* C_0 шунтирует на «землю» анод или катод диода (считается, что подложка заземлена). Емкость C_0, как правило, совпадает с емкостью $C_{кп}$, с которой мы встретились при рассмотрении n–p–n-транзистора (рис. 7.14, *б*). Однако у варианта Б–Э емкости $C_{кп}$ и $C_к$ оказываются включенными последовательно и результирующая емкость C_0 минимальна.

Таблица 7.3. Типичные параметры интегральных диодов

Параметр	Тип диода				
	БК–Э	Б–Э	БЭ–К	Б–К	Б–ЭК
$U_{пр}$, В	7–8	7–8	40–50	40–50	7–8
$I_{обр}$, нА	0,5–1	0,5–1	15–30	15–30	20–40
$C_д$, пФ	0,5	0,5	0,7	0,7	1,2
C_0, пФ	3	1,2	3	3	3
$t_в$, нс	10	50	50	75	100

Время восстановления обратного тока $t_в$ (т.е. время переключения диода из открытого в закрытое состояние) минимально у варианта БК–Э; у этого варианта заряд накапливается только в базовом слое (так как коллекторный переход закорочен). У других вариантов заряд накапливается не только в базе, но и в коллекторе, так что для рассасывания заряда требуется большее время.

Сравнивая отдельные варианты, приходим к выводу, что в целом *оптимальными вариантами являются БК–Э и Б–Э.* Малые пробивные напряжения этих вариантов не играют существенной роли в низковольтных ИС. Чаще всего используется вариант БК–Э.

Помимо собственно диодов, в ИС часто используются *интегральные стабилитроны*. Они также осуществляются в нескольких вариантах, в зависимости от необходимого напряжения стабилизации и температурного коэффициента.

Если необходимы напряжения 5–10 В, то используют обратное включение диода Б–Э в режиме пробоя; при этом температурная чувствительность составляет + (2–5) мВ/°С. Если необходимы напряжения 3–5 В, то применяют либо обратное включение диода БЭ–К, используя эффект смыкания (см. с. 144), либо обратное включение p–n-перехода, специально образованного в разделительном слое (рис. 7.25, *а*). В последнем случае n^+-слой получают на этапе эмиттерной диффузии. Поскольку приповерхностная часть разделительного слоя сильно легирована, переход имеет структуру p^+–n^+ и ему свойствен туннельный — низковольтный пробой. Температурная чувствительность составляет – (2–3) мВ/°С.

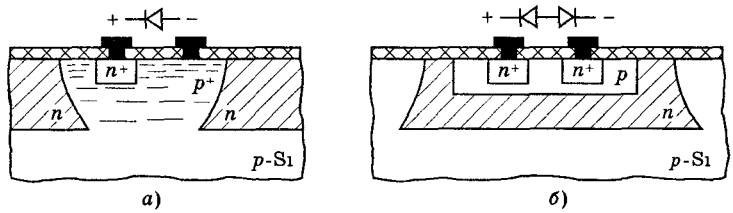


Рис. 7.25. Интегральные стабилитроны.
а — на основе разделительного слоя, *б* — на основе базового слоя (с применением температурной компенсации)

Широкое распространение имеют стабилитроны, рассчитанные на напряжения, равные или кратные напряжению на открытом переходе $U^* \approx 0{,}7$ В. В таких случаях используют один или несколько последовательно включенных диодов БК–Э, работающих в прямом направлении. Температурная чувствительность в этом случае составляет – (1,5–2) мВ/°С.

Если в базовом слое осуществить два p–n-перехода (рис. 7.25, *б*), то при подаче напряжения между n^+-слоями один из переходов работает в режиме лавинного пробоя, а второй — в режиме прямого смещения. Такой вариант привлекателен малой температурной чувствительностью (± 1 мВ/°С и менее), так как температурные чувствительности при лавинном пробое и при прямом смещении имеют разные знаки.

7.7. Полевой транзистор

Полевые транзисторы (ПТ), рассмотренные в разделе 4.3, хорошо вписываются в общую технологию биполярных ИС и потому часто изготавливаются совместно с биполярными транзисторами на одном кристалле. Типичные структуры ПТ, расположенные в изолированных карманах, показаны на рис. 7.26.

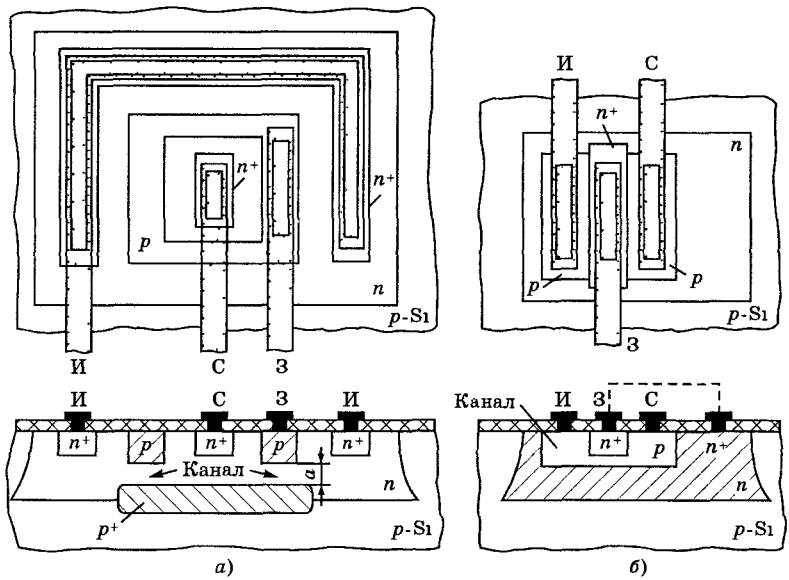


Рис. 7.26. Интегральные полевые транзисторы с n-каналом (*а*) и с p-каналом (*б*)

В структуре, показанной на рис. 7.26, *а*, p-слой затвора образуется на этапе базовой диффузии, а n^+-слои, обеспечивающие омический контакт с областями истока и стока, — на этапе эмиттерной диффузии. Заметим, что p-слой затвора окружает сток со всех сторон, так что ток между истоком и стоком может протекать только через управляемый канал.

В n-карманах, предназначенных для ПТ, вместо скрытого n^+-слоя осуществляется скрытый p^+-слой. Назначение этого слоя — уменьшить начальную толщину канала a и тем самым напряжение отсечки [см.(4.29)]. Осуществление скрытого p^+-слоя связано с дополнительными технологическими операциями. Для того чтобы скрытый p^+-слой проник в эпитаксиаль-

ный слой достаточно глубоко, в качестве акцепторного диффузанта используют элементы с большим коэффициентом диффузии (бор или галлий).

На подложку, а значит, и на p^+-слой задают постоянный (максимально отрицательный) потенциал; поэтому они не выполняют управляющих функций.

Структура, показанная на рис. 7.26, *б*, совпадает со структурой обычного n–p–n-транзистора. Роль канала играет участок базового p-слоя, расположенный между n^+- и n-слоями. Если при совместном изготовлении ПТ и биполярного транзистора не использовать дополнительных технологических процессов, то толщина канала будет равна ширине базы n–p–n-транзистора (0,5–1 мкм). При такой малой толщине канала получаются большой разброс параметров ПТ и малое напряжение пробоя. Поэтому целесообразно пойти на усложнение технологического цикла, осуществляя p-слой ПТ отдельно от базового p-слоя, с тем чтобы толщина канала была не менее 1–2 мкм. Для этого проводят предварительную диффузию p-слоя ПТ до базовой диффузии. Тогда во время базовой диффузии p-слой ПТ дополнительно расширяется, и его глубина оказывается несколько больше глубины базового слоя.

Для того чтобы области истока и стока соединялись только через канал, n^+-слой делают более широким (в плане), чем p-слой (рис. 7.26, *б*). В результате n^+-слой контактирует с эпитаксиальным n-слоем и вместе они образуют «верхний» и «нижний» затворы. В нижней части рис. 7.26, *б* контакт между «верхним» и «нижним» затворами условно показан штриховой линией. Подложка p-типа присоединяется к максимальному отрицательному потенциалу.

7.8. МДП-транзисторы

Вообще говоря, совместное изготовление МДП- и биполярных транзисторов на одном кристалле, в едином технологическом цикле возможно, но является специальным случаем. Как правило, биполярные и МДП-транзисторные ИС разрабатываются и изготавливаются раздельно. Эти два типа ИС предназначены либо для решения разных функциональных задач, либо для решения одной и той же задачи, но с использованием преимуществ соответствующего класса транзисторов. Главную роль в современной микроэлектронике играют МДП-транзисторы, в

которых диэлектриком является SiO_2, их называют МОП-транзисторами.

Простейший МОП-транзистор. Поскольку интегральные МДП-транзисторы, как отмечалось в разделе 7.2, не нуждаются в изоляции, их структура внешне не отличается от структуры дискретных вариантов. На рис. 7.27 воспроизведена структура МОП-транзистора с индуцированным n-каналом, которая была подробно проанализирована в разделе 4.2. Отметим особенности этого транзистора как элемента ИС.

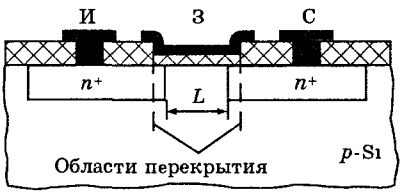


Рис. 7.27. Интегральный МОП-транзистор (с перекрытием затвора)

Из сравнения рис. 7.6, 7.21 и 7.22 очевидна, прежде всего, технологическая простота МОП-транзистора по сравнению с биполярными: необходимы всего лишь один процесс диффузии и четыре процесса фотолитографии (под диффузию, под тонкий окисел, под омические контакты и под металлизацию). Технологическая простота обеспечивает меньший брак и меньшую стоимость.

Отсутствие изолирующих карманов способствует лучшему использованию площади кристалла, т.е. повышению степени интеграции элементов. Однако, с другой стороны, отсутствие изоляции делает подложку общим электродом для всех транзисторов. Это обстоятельство может привести к различию параметров у внешне идентичных транзисторов. Действительно, если на подложку задан постоянный потенциал, а истоки транзисторов имеют разные потенциалы (такое различие свойственно многим схемам), то будут разными и напряжения $U_{пи}$ между подложкой и истоками. Согласно (4.15) это равносильно различию пороговых напряжений МДП-транзисторов.

Как известно, главным фактором, лимитирующим быстродействие МДП-транзисторов, обычно являются паразитные емкости (см. с. 114–115). Металлическая разводка, используемая в ИС, гораздо компактнее проволочного монтажа, свойственного узлам и блокам, выполненным на дискретных компонентах. Поэтому паразитные емкости интегрального МОП-транзистора меньше, чем дискретного, а его быстродействие соответственно в несколько раз выше.

Паразитные емкости МДП-транзистора были показаны на рис. 4.10. Барьерные емкости переходов истока и стока ($C_{\text{пи}}$ и $C_{\text{пс}}$) рассчитываются по формуле (3.25); при размерах n^+-слоев 20×40 мкм эти емкости лежат в пределах 0,04–0,10 пФ.

Удельная емкость металлизации определяется элементарной формулой

$$C_{0\text{м}} = \varepsilon_0 \varepsilon / d, \tag{7.4}$$

где d — толщина защитного окисла, ε — его диэлектрическая проницаемость. Подставляя $\varepsilon = 4{,}5$ и $d = 0{,}7$ мкм, получаем типичное значение $C_{0\text{м}} \approx 60$ пФ/мм^2. При ширине полоски металлизации 15 мкм погонная емкость составит 0,9 пФ/мм. Как видим, полоски длиною всего 50–100 мкм могут иметь емкость 0,05–0,09 пФ, сравнимую с емкостями переходов $C_{\text{пи}}$ и $C_{\text{пс}}$. Еще больший вклад вносят контактные площадки: при площади 100×100 мкм^2 их емкость составляет около 0,6 пФ.

Емкости перекрытия $C_{\text{зи}}$ и $C_{\text{зс}}$ (см. рис. 4.11) не поддаются сколько-нибудь точному расчету, так как площадь перекрытия характерна большим разбросом из-за неровности краев металлизации затвора и границ диффузионных слоев (рис. 7.27). Порядок величин можно оценить, полагая толщину тонкого окисла $d = 0{,}12$ мкм. Тогда из (7.4) получаем удельную емкость тонкого окисла $C_0 \approx 350$ пФ/мм^2. При ширине истока и стока 40 мкм и перекрытии 2 мкм получим средние значения $C_{\text{зи}} = C_{\text{зс}} \approx 0{,}03$ пФ. Эти значения меньше, чем значения других емкостей. Поэтому, в частности, емкостью $C_{\text{зи}}$ часто пренебрегают. Однако емкость $C_{\text{зс}}$, представляющая собой *емкость обратной связи* между выходом транзистора (стоком) и входом (затвором), проявляется во многих схемах не как таковая, а в виде гораздо большей эквивалентной емкости $KC_{\text{зс}}$ (благодаря так называемому *эффекту Миллера*, см. разд. 9.5). Множитель K есть коэффициент усиления схемы: он может составлять от нескольких единиц до нескольких десятков и более. Поэтому эквивалентная емкость обратной связи $KC_{\text{зс}}$ может достигать значений 0,1–0,5 пФ, т.е. превышать все остальные емкости.

В комплементарных МОП-транзисторных ИС (КМОП) (см. с. 101) на одном и том же кристалле необходимо изготовлять транзисторы обоих типов: с n- и с p-каналом. При этом один из типов транзисторов нужно размещать в специальном кармане. Например, если в качестве подложки используется p-кремний,

то n-канальный транзистор можно осуществить непосредственно в подложке, а для p-канального транзистора потребуется карман с электронной проводимостью (рис. 7.28, *а*). Получение такого кармана в принципе несложно, но связано с дополнительными технологическими операциями (фотолитография, диффузия доноров и др.). Кроме того, затрудняется получение низкоомных p^+-слоев в верхней (сильно легированной) части n-кармана. Другим способом изготовления КМОП-транзисторов на одной подложке является КНС технология (см. разд. 7.2). В этом случае на сапфировой подложке создаются «островки» кремния с собственной проводимостью, после чего в одних «островках» проводится диффузия донорной примеси и получаются n-канальные транзисторы, а в других — диффузия акцепторной примеси и получаются p-канальные транзисторы (рис. 7.28, *б*). Хотя количество технологических операций и в этом случае больше, чем при изготовлении транзисторов одного типа, зато отпадают трудности, связанные с получением низкоомных слоев истока и стока (см. выше).

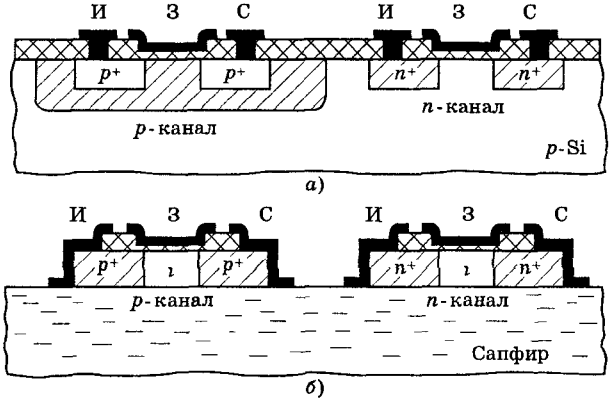


Рис. 7.28. Комплементарные МОП-транзисторы:
а — использование изолирующего n-кармана;
б — использование воздушной изоляции (технология КНС)

Что касается сочетания МОП-транзисторов с биполярными, то в принципе оно осуществляется просто (рис. 7.29): n-канальные транзисторы изготавливаются непосредственно в p-подложке на этапе эмиттерной диффузии, а p-канальные — в изолирующих карманах на этапе базовой диффузии.

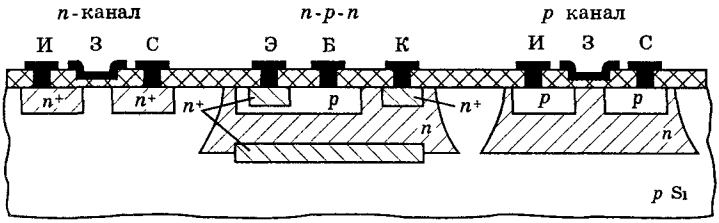


Рис. 7.29. Сочетание биполярных и МОП транзисторов на одном кристалле

В процессе развития микроэлектроники усовершенствование МОП-транзисторов происходило по двум главным направлениям: повышение быстродействия и снижение порогового напряжения. В основе последней тенденции лежало стремление снизить рабочие напряжения МОП-транзисторов и рассеиваемую ими мощность. Поскольку полная мощность кристалла ограничена, уменьшение мощности, рассеиваемой в одном транзисторе, способствует повышению степени интеграции, а уменьшение напряжений питания облегчает совместную работу МОП-транзисторных и низковольтных биполярных ИС без специальных согласующих элементов.

Способы повышения быстродействия. Повышение быстродействия МОП-транзисторов связано прежде всего с уменьшением емкостей перекрытия. Существенное (примерно на порядок) уменьшение емкостей перекрытия достигается при использовании *технологии самосовмещенных затворов*. Общая идея такой технологии состоит в том, что слои истока и стока осуществляются не до, а после осуществления затвора. При этом затвор используется в качестве маски при получении слоев истока и стока, а значит, края затвора и этих слоев будут совпадать и перекрытие будет отсутствовать.

Один из вариантов МОП-транзистора с самосовмещенным затвором показан на рис. 7.30. Последовательность технологических операций при этом следующая. Сначала проводится диффузия n^+-слоев, причем расстояние между ними делается заведомо больше желательной длины канала. Затем осуществляется тонкое окисление на участке между n^+-слоями и частично над ними. Далее на тонкий окисел напыляется алюминиевый электрод затвора, причем его ширина меньше расстояния между n^+-слоями. Наконец, проводится ионное легирование (имплантация атомов фосфора) через маску, образуемую алю-

миниевым затвором и толстым защитным окислом. Атомы фосфора проникают в кремний через тонкий окисел и «продлевают» n^+-слои до края алюминиевой полоски так, что края затвора практически совпадают с краями истока и стока. Имплантированные слои легированы несколько слабее, чем диффузионные; поэтому для них использовано обозначение n вместо n^+. Глубина имплантации также несколько меньше, чем глубина диффузии, и составляет 0,1–0,2 мкм.

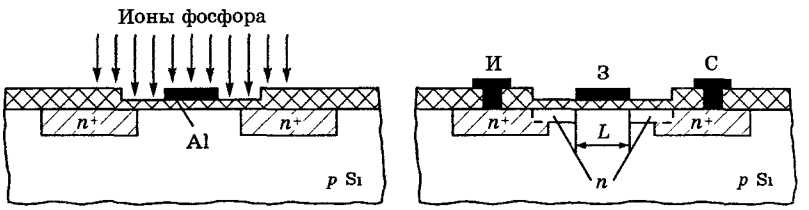


Рис. 7.30. МОП-транзистор с самосовмещенным затвором, полученный методом ионной имплантации

Другой вариант МОП-транзистора с самосовмещенным затвором показан на рис. 7.31. В этом варианте сначала вытравливают окно в окисле с размерами, соответствующими всей структуре транзистора. Затем в средней части окна проводят тонкое окисление кремния в виде полоски, ширина которой равна длине будущего канала L. Далее на эту полоску напыляют поликристаллический слой кремния той же ширины, но более длинный, выходящий за границы исходного окна в окисле (рис. 7.31, *а*). Удельное сопротивление напыляемого кремния делается достаточно малым, так что *слой поликристаллического кремния выполняет роль металлического затвора* в обычных МОП-транзисторах. На следующем этапе проводят мелкую диффузию донорной примеси через маску, образуемую полоской поликристаллического кремниевого затвора и защитным окислом, окружающим окно; в результате получаются n^+-слои истока и стока, края которых почти совпадают с краями кремниевого затвора. Далее всю поверхность кристалла окисляют и в этом окисле, как обычно, делают окна для омических контактов, в том числе для контакта с кремниевым затвором. В заключение осуществляют металлическую разводку. Из рис. 7.31, *б* видно, что поликристаллический кремниевый затвор оказывается «погруженным» в слой защитного окисла;

омический контакт к нему располагается за пределами рабочей области транзистора.

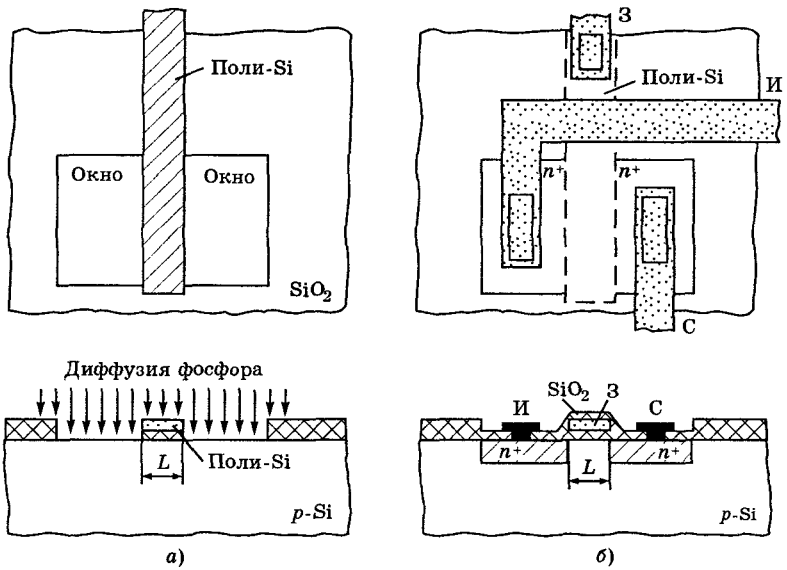


Рис. 7.31. МОП-транзистор с самосовмещенным поликремниевым затвором: *а* — этап диффузии доноров через маску, включающую слой поликристаллического кремния; *б* — готовая структура (после нанесения защитного окисла и металлизации)

Уменьшение паразитных емкостей МОП-транзисторов и прежде всего емкости перекрытия $C_{эс}$ выдвигает на первый план задачу уменьшения постоянной времени крутизны τ_S. При малых емкостях она становится главным фактором, ограничивающим быстродействие.

Переход от транзисторов с *p*-каналом к транзисторам с *n*-каналом позволил уменьшить значение τ_S примерно в 3 раза благодаря увеличению подвижности носителей. Дальнейшее уменьшение величины τ_S требует уменьшения длины канала L. Этот путь наиболее радикально реализуется *методом двойной диффузии*. Структура МОП-транзистора, полученная этим методом, показана на рис. 7.32.

Эта структура аналогична структуре классического *n–p–n*-транзистора (рис. 7.14, *а*) с той, однако, существенной разницей, что эмиттерный n^+-слой (в данном случае это слой истока) имеет почти такую же площадь, что и базовый *p*-слой

(в данном случае это слой канала). Для того, чтобы обеспечить точное «вписывание» n^+-слоя в p-слой, диффузию доноров для n^+-слоя осуществляют через то же самое окно в окисле, через которое до этого осуществляли диффузию акцепторов для p-слоя. Тем самым устраняется необходимость в совмещении фотошаблонов для двух последовательных фотолитографий, а значит, и ошибка совмещения, которая могла бы привести к сдвигу n^+-слоя относительно n-слоя. В результате расстояние между n^+- и n-слоями, т.е. толщина p-слоя, может иметь примерно те же значения, что и ширина базы w у n–p–n-транзистора (см. рис. 7.14, *а*): до 1 мкм и менее.

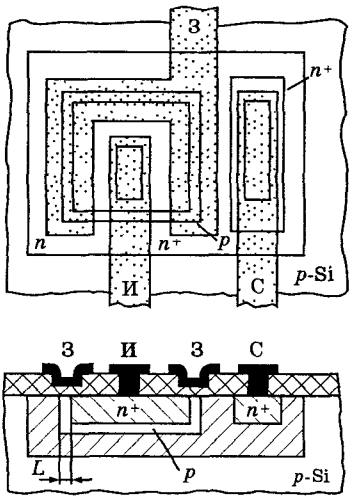


Рис. 7.32. МОП-транзистор, полученный методом двойной диффузии

Вблизи поверхности расстояние между n^+- и n-слоями играет роль длины канала L (рис. 7.32). При значениях $L \leq 1$ мкм (вместо 4–5 мкм у наиболее совершенных МОП-транзисторов, полученных по обычной технологии) постоянная времени τ_S согласно (4.27) может быть менее 0,005 нс, а граничная частота f_S более 30 ГГц.

Способы уменьшения порогового напряжения. Транзисторы со структурой, показанной на рис. 7.31, обычно называют МОП-транзисторами *с кремниевым затвором*. Такие транзисторы характерны не только малой емкостью перекрытия, но и малым пороговым напряжением: 1–2 В вместо обычных 2,5–3,5 В. Это объясняется тем, что материал затвора и подложки — один и тот же — кремний. Следовательно, контактная разность потенциалов между ними (φ_{MS}) равна нулю, что и приводит к уменьшению порогового напряжения [см. (4.3а)]. Примерно такой же результат дает использование молибденового затвора.

Помимо контактной разности потенциалов, для уменьшения порогового напряжения можно варьировать и другими параметрами, входящими в выражения (4.3). Например, можно за-

менить тонкий окисел SiO_2 тонким напыленным слоем нитрида кремния Si_3N_4, у которого диэлектрическая проницаемость ($\varepsilon \approx 7$) примерно в полтора раза больше, чем у двуокиси кремния ($\varepsilon = 4{,}5$). Это приводит к увеличению удельной емкости C_0, а значит к уменьшению соответствующих слагаемых порогового напряжения. Нитрид кремния в качестве подзатворного диэлектрика обеспечивает также дополнительные преимущества: меньшие шумы, бóльшую временну́ю стабильность ВАХ и повышенную радиационную стойкость МДП-транзистора.

Можно вместо пластин кремния с традиционной кристаллографической ориентацией (111) использовать пластины с ориентацией (100). При этом увеличивается плотность поверхностных состояний (рис. 2.5), а вместе с нею и заряд захватываемых ими электронов. Соответственно возрастает отрицательное слагаемое Q_{0S}/C_0 в формуле (4.3а), а алгебраическая сумма обоих слагаемых, т.е. напряжение U_{0F}, уменьшается.

Обратное влияние оказывает введение в тонкий окисел акцепторных атомов: они захватывают из приповерхностного слоя кремния часть электронов, порожденных донорными примесями, которые всегда присутствуют в окисле (см. с. 95-96). В результате заряд Q_{0s} уменьшается. Введение акцепторов в окисел можно осуществлять с помощью ионной имплантации.

Сочетая перечисленные методы, можно обеспечить пороговые напряжения практически любой сколь угодно малой величины. Следует, однако, иметь в виду, что слишком малые значения порогового напряжения (0,5–1 В и менее) в большинстве случаев неприемлемы по схемотехническим соображениям (малая помехоустойчивость).

МНОП-транзистор. Особое место среди МДП-транзисторов занимает так называемый *МНОП-транзистор*, у которого диэлектрик имеет структуру «сэндвича», состоящего из слоев нитрида и окисла кремния (рис. 7.33, *а*). Слой окисла получается путем термического окисления и имеет толщину 2–5 нм, а слой нитрида — путем реактивного напыления и имеет толщину 0,05–0,1 мкм, достаточную для того, чтобы пробивное напряжение превышало 50–70 В.

Главная особенность МНОП-транзистора состоит в том, что его пороговое напряжение можно менять, подавая на затвор короткие (100 мкс) импульсы напряжения разной полярности, с большой амплитудой (30-50 В). Так, при подаче импульса + 30 В

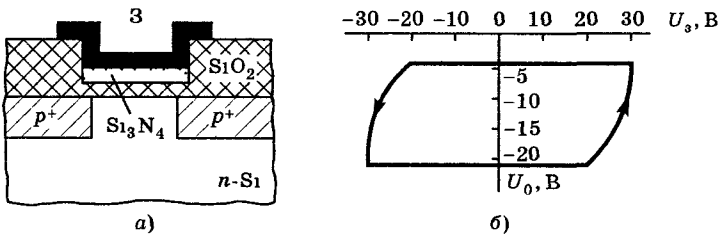


Рис. 7.33. МОП-транзистор с индуцированным p-каналом: *а* — структура; *б* — зависимость порогового напряжения от напряжения затвора

устанавливается пороговое напряжение $U_0 = -4$ В (рис. 7.33, *б*). Это значение сохраняется при дальнейшем использовании транзистора в режиме малых сигналов ($U_з \leq \pm 10$ В); в таком режиме МНОП-транзистор ведет себя как обычный МДП-транзистор с индуцированным p-каналом. Если теперь подать импульс -30 В, то пороговое напряжение сделается равным $U_0 = -20$ В и, следовательно, сигналы $U_з \leq \pm 10$ В не смогут вывести транзистор из запертого состояния. Как видим, благодаря гистерезисной зависимости U_0 ($U_з$) МНОП-транзистор можно с помощью больших управляющих импульсов переводить из рабочего в запертое состояние и обратно. Эта возможность используется в интегральных запоминающих устройствах.

В основе работы МНОП-транзистора лежит накопление заряда на границе нитридного и оксидного слоев. Это накопление есть результат неодинаковых токов проводимости в том и другом слоях. Процесс накопления описывается элементарным выражением

$$dQ/dt = I_{SiO_2} - I_{Si_3N_4},$$

где оба тока зависят от напряжения на затворе и меняются в процессе накопления заряда. При большом отрицательном напряжении на границе накапливается положительный заряд. Это равносильно введению доноров в диэлектрик и сопровождается увеличением отрицательного порогового напряжения. При большом положительном напряжении $U_з$ на границе накапливается отрицательный заряд. Это приводит к уменьшению отрицательного порогового напряжения.

При малых напряжениях $U_з$ токи в диэлектрических слоях уменьшаются на 10–15 порядков (!), так что *накопленный за-*

ряд сохраняется в течение тысяч часов. Вместе с ним сохраняется и пороговое напряжение.

Смешанные монолитные ИС на МОП и биполярных структурах. Биполярные структуры обеспечивают высокоточное преобразование аналоговых величин и обладают большими управляющими токами. КМОП-схемы имеют большую степень интеграции, малую потребляемую мощность и эффективно используются в запоминающих устройствах. В настоящее время активно развивается технология, позволяющая интегрально объединить биполярные и КМОП-схемы в одном кристалле и таким образом существенно расширить функциональные возможности ИС, БИС и особенно СБИС. На рис. 7.34 приведен пример интеграции $И^2Л$-структур с КМОП-структурами с биполярными p-карманами. Аналогичное совмещение возможно в случае КМОП-структур с биполярными n-карманами.

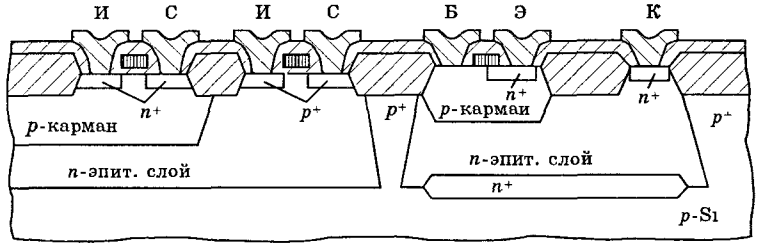


Рис. 7.34. КМОП-структура с p-карманами и кремниевыми затворами

7.9. Полупроводниковые резисторы

Первоначально в полупроводниковых ИС применялись только *диффузионные резисторы* (ДР), основу которых составлял один из диффузионных слоев, расположенных в изолированном кармане. В настоящее время большое распространение получили также ионно-имплантированные резисторы.

Диффузионные резисторы. Для диффузионных резисторов чаще всего используется полоска базового слоя с двумя омическими контактами (рис. 7.35, *а*). Для такой полосковой конфигурации сопротивление ДР согласно (7.1) записывается в виде

$$R = R_s\,(a/b), \qquad (7.5а)$$

где R_s — удельное сопротивление слоя (см. с. 229-230), а размеры a и b показаны на рис. 7.35.

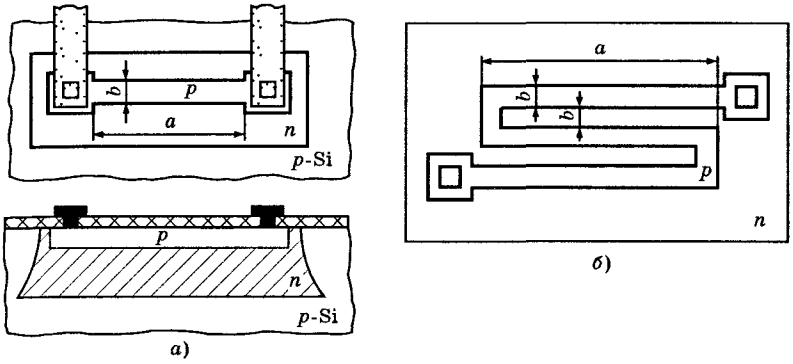


Рис. 7.35. Диффузионный резистор с полосковой (*а*) и с зигзагообразной конфигурацией (*б*)

И длина, и ширина полоскового ДР ограничены. Длина a не может превышать размеров кристалла. Ширина b ограничена возможностями фотолитографии, боковой диффузией, а также допустимым разбросом (10–20%).

Подставляя в (7.5а) значения $R_s = 200$ Ом/□ и $a/b = 100$, получаем типичное значение максимального сопротивления $R_{макс} = 20$ кОм. Это значение можно повысить в 2-3 раза, используя не полосковую, а *зигзагообразную конфигурацию ДР* (рис. 7.35, *б*). В этом случае сопротивление записывается в более общем виде

$$R = R_s\left(\frac{\sum_i a_i}{b} + n + 1{,}3\right). \qquad (7.5б)$$

Здесь n — количество «петель» (на рис. 7.35, *б* $n = 2$), а слагаемое 1,3 учитывает неоднородность ДР в районе омических контактов.

Количество «петель» в конечном счете ограничено площадью, отводимой под ДР. Обычно $n \le 3$, в противном случае площадь резистора может достигать 15–20 % площади всего кристалла. Максимальное сопротивление при $n = 3$ не превышает 50–60 кОм.

Температурный коэффициент сопротивления ДР, выполненного на основе базового слоя, составляет 0,15–0,30 %/°С, в зависимости от значения R_s. Разброс сопротивлений относительно расчетного номинала составляет ± (15–20) %. При этом сопротивления резисторов, расположенных на одном кристалле, меняются в одну и ту же сторону. Поэтому отношение сопротивлений сохраняется с гораздо меньшим допуском (± 3 % и менее), а температурный коэффициент для отношения сопротивлений не превышает ± 0,01 %/°С. Эта особенность ДР играет важную роль и широко используется при разработке ИС.

Если необходимые номиналы сопротивлений превышают 50–60 кОм, можно использовать так называемые *пинч-резисторы*. Структура пинч-резистора показана на рис. 7.36. По сравнению с простейшим ДР пинч-резистор имеет меньшую площадь сечения и большее удельное сопротивление (так как используется донная, т.е. слабо легированная часть *p*-слоя). Поэтому у пинч-резисторов удельное сопротивление слоя обычно составляет 2–5 кОм/□ и более, в зависимости от толщины. При таком значении R_s максимальное сопротивление может достигать значений 200–300 кОм даже при простейшей полосковой конфигурации.

Недостатками пинч-резисторов являются: больший разброс номиналов (до 50 %) из-за сильного влияния изменения толщины *p*-слоя, больший температурный коэффициент сопротивления (0,3–0,5 %/°С) из-за меньшей степени легирования донной части *p*-слоя, нелинейность вольт-амперной характеристики при

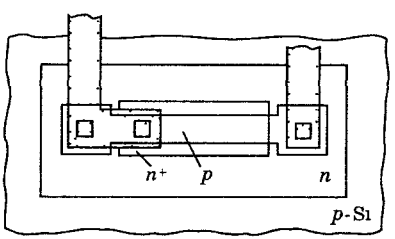


Рис. 7.36. Пинч-резистор

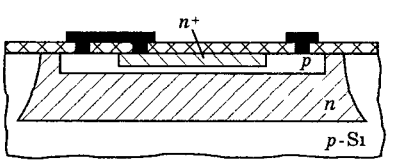


Рис. 7.37. Туннельное пересечение

напряжениях более 1–1,5 В. Последняя особенность вытекает из аналогии между структурами пинч-резистора и полевого транзистора (см. рис. 7.26, *б*). ВАХ пинч-резистора совпадает с ВАХ полевого транзистора (рис. 4.14, *а*), если напряжение на затворе последнего положить равным нулю (поскольку у пинч-резистора слои n^+ и p соединены друг с другом металлизацией). Пробивное напряжение пинч-резисторов определяется пробивным напряжением эмиттерного перехода (обычно 5–7 В).

Если необходимые номиналы сопротивлений составляют 100 Ом и менее, то использование базового слоя в ДР нецелесообразно, так как согласно (7.5а) ширина резистора должна быть меньше его длины, что конструктивно трудно осуществить. Для получения ДР с малыми номиналами сопротивлений используют низкоомный эмиттерный слой. При значениях R_s = 5–15 Ом/□, свойственных этому слою (см. табл. 7.1), удается получить минимальные сопротивления 3–5 Ом с температурным коэффициентом 0,01–0,02 %/°С.

Используя эмиттерный n^+-слой, можно решить еще одну задачу, возникающую при конструировании ИС: осуществить так называемые *«туннельные» пересечения* (рис. 7.37). Речь идет о возможности изоляции двух взаимно перпендикулярных полосок металлизации, одна из которых (*1*) идет полностью по защитному окислу, а вторая (*2*) частично расположена под первой в виде низкоомного n^+-участка — «туннеля». Пример «туннельного» пересечения можно видеть на рис. 7.13, *б*, где в качестве «туннеля» используется участок коллекторного n^+-слоя. «Туннельное» пересечение не является универсальным решением, потому что n^+-участок имеет все же заметное сопротивление (3–5 Ом). Поэтому, в частности, *«туннельные» пересечения неприемлемы в цепях питания*, где протекают достаточно большие токи.

Ионно-легированные резисторы. За последнее время все большее распространение получают *ионно легированные резисторы*, которые в отличие от ДР получаются не диффузией, а локальной ионной имплантацией примеси (см. разд. 6.5).

Структура ионно-легированного резистора такая же, как ДР (рис. 7.38), но глубина имплантированного p-слоя значительно меньше глубины базового слоя и составляет всего 0,2–0,3 мкм. Кроме того, ионная имплантация позволяет обеспечить сколь угодно малую концентрацию примеси в слое. Оба фактора способствуют получению весьма высоких удельных сопротивлений слоя — до 10–20 кОм/□. При этом номиналы сопротивлений

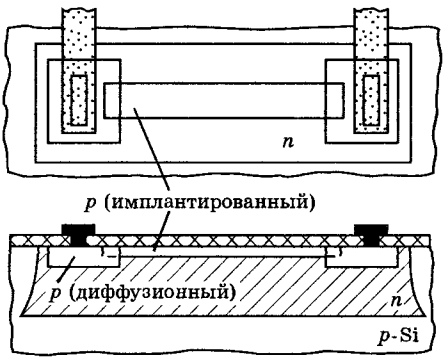


Рис. 7.38. Ионно-легированный резистор

могут составлять сотни килоом. ТКС меньше, чем у ДР, и лежит в пределах 3–5%/°С, а разброс сопротивлений не превышает ± (5–10)%.

Поскольку толщина имплантированного слоя мала, к нему трудно осуществить омические контакты. Поэтому по краям резистивного слоя на этапе базовой диффузии осуществляют узкие диффузионные p-слои, к которым омический контакт осуществляется обычным способом.

Эквивалентные схемы. Характерной особенностью любого интегрального резистора является наличие у него паразитной емкости относительно подложки или изолирующего кармана. В простейшем ДР (рис. 7.35) такой паразитной емкостью является барьерная емкость перехода между рабочим p-слоем и эпитаксиальным n-слоем кармана [1].

Строго говоря, совокупность резистора и паразитной емкости представляет собой распределенную RC-линию (рис. 7.39). Однако для приближенных расчетов удобнее пользоваться эквивалентными схемами с сосредоточенными постоянными: П-образной (рис. 7.40, *а*) или Т-образной (рис. 7.40, *б*). На этих схемах R — сопротивление резистора, $\overline{C}$ — усредненная емкость перехода.

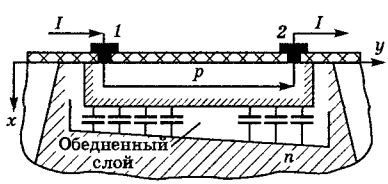


Рис. 7.39. Физическая модель интегрального резистора в виде распределенной RC-линии

1 На n-слой задается максимальный положительный потенциал от источника питания, поэтому все участки p–n-перехода находятся под обратным напряжением.

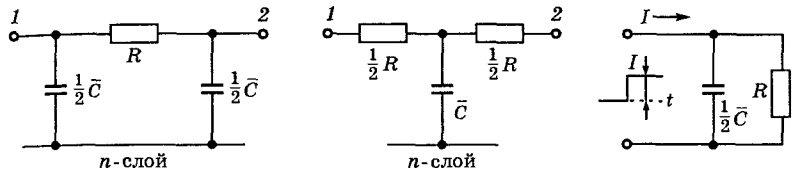


Рис. 7.40. Эквивалентные схемы интегрального резистора: *а*) П-образная; *б*) Т-образная; *в*) П-образная при $U_2 = \text{const}$

Необходимость усреднения емкости объясняется следующим. При протекании тока через резистор потенциал p-слоя оказывается разным в разных точках. Поскольку потенциал n-слоя постоянный, напряжение на p–n-переходе будет меняться вдоль оси y, а следовательно, будет меняться и барьерная емкость.

В типичном случае, когда один из зажимов резистора (например, *2*) находится под постоянным потенциалом, а через другой зажим (*1*) задается ступенька тока, эквивалентная П-образная схема сводится к простейшей RC-цепочке, показанной на рис. 7.40, *в*. Сущность переходного процесса состоит в плавном изменении напряжения на резисторе при скачкообразном изменении тока. Постоянная времени, определяющая длительность переходного процесса, имеет вид

$$\tau = \tfrac{1}{2} R\bar{C}, \qquad (7.6а)$$

а соответствующая граничная частота

$$f_{гр} = 1/(2\pi\tau) = 1/(\pi R\bar{C}). \qquad (7.6б)$$

Для значений $R = 10$ кОм и $\bar{C} = 1{,}3$ пФ получаем: $\tau = 6{,}5$ нс и $f_{гр} \approx 25$ МГц. Это значит, что в данном примере *резистор выполняет свою функцию* (т.е. имеет чисто активное сопротивление) *только до частот* 10–15 МГц. При более высоких частотах его сопротивление становится комплексным и работа схемы, в которой он используется, может существенно измениться.

Рассмотренные эквивалентные схемы действительны и для других вариантов резисторов: когда рабочими являются эмиттерный или коллекторный слой, а также при диэлектрической изоляции элементов. Однако количественные результаты оказываются разными. Например, при использовании диэлектрической изоляции постоянная времени может быть в несколько раз меньше.

7.10. Полупроводниковые конденсаторы

В биполярных полупроводниковых ИС роль конденсаторов играют обратносмещенные p–n-переходы. У таких конденсаторов хотя бы один из слоев является диффузионным, поэтому их называют *диффузионными конденсаторами* (ДК).

Диффузионный конденсатор. Типичная структура ДК, в котором используется переход коллектор–база, показана на рис. 7.41. Емкость такого конденсатора в общем случае имеет вид:

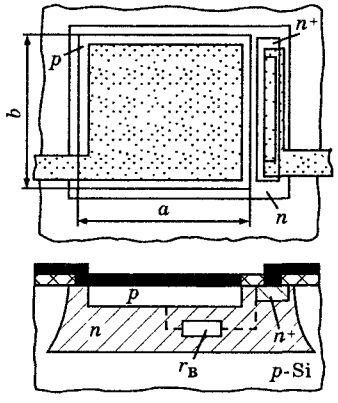


Рис. 7.41. Диффузионный конденсатор

$$C = C_{01}(ab) + C_{02}2(a+b)d, \qquad (7.7а)$$

где C_{01} и C_{02} — удельные емкости донной и боковой частей p–n-перехода. Соотношение слагаемых в правой части (7.7а) зависит от отношения a/b, т.е. от конфигурации ДК. *Оптимальной конфигурацией является квадрат* ($a = b$): при этом «боковая» составляющая емкости оказывается в десятки раз меньше донной. Пренебрегая боковой составляющей, т.е. вторым слагаемым в (7.7а), и полагая $a = b$, получаем:

$$C = C_{01}(ab) = C_{01}a^2. \qquad (7.7б)$$

Например, если $C_{01} = 150$ пФ/мм^2 и $C = 100$ пФ, то $a \approx 0{,}8$ мм. Как видим, размеры конденсатора получились сравнимыми с размерами кристалла.

Для того чтобы суммарная площадь всех конденсаторов, входящих в состав ИС, не превышала 20–25 % площади кристалла, необходимо ограничить суммарную емкость конденсаторов величиной

$$C_{макс} = (0{,}2 - 0{,}25)C_{01}S_{кр},$$

где $S_{кр}$ — площадь кристалла. Если $S_{кр} = 2$–9 мм^2 и $C_{01} = 150$ пФ/мм^2, то $C_{макс} = 50$–300 пФ.

Используя не коллекторный, а эмиттерный p–n-переход, можно обеспечить в 5–7 раз большие значения максимальной емко-

сти. Это объясняется большей удельной емкостью эмиттерного перехода, поскольку он образован более низкоомными слоями.

Основные параметры ДК, включая технологический разброс номиналов δ, температурный коэффициент емкости ТКЕ [1], пробивное напряжение $U_{пр}$ и добротность Q, приведены в табл. 7.4 для обоих вариантов ДК — с использованием коллекторного и эмиттерного переходов. Как видим, основное преимущество при использовании эмиттерного перехода — бóльшие значения максимальной емкости. По пробивному напряжению и добротности (см. ниже) этот вариант уступает варианту с использованием коллекторного перехода.

Таблица 7.4. Типичные параметры интегральных конденсаторов

Тип конденсатора	C_0, пФ/мм^2	$C_{макс}$, пФ	δ, %	ТКЕ, %/°С	$U_{пр}$, В	Q (1 МГц)
Переход БК	150	300	± 20	− 0,1	50	50–100
Переход БЭ	1000	1200	± 20	− 0,1	7	1–20
МОП-структура	300	500	± 25	0,02	20	200

Необходимым условием для нормальной работы ДК является обратное смещение p–n-перехода. Следовательно, *напряжение на ДК должно иметь строго определенную полярность.*

Емкость ДК согласно (3.25) зависит от напряжения. Это значит, что ДК, вообще говоря, является *нелинейным конденсатором* с вольт-фарадной характеристикой $C(U)$. Нелинейные конденсаторы находят применение в специальных узлах радиотехнической аппаратуры: параметрических усилителях, умножителях частоты и др. В таких узлах нелинейность ДК оказывается полезной. Однако чаще требуются *линейные* конденсаторы с постоянной емкостью, которые способны пропускать без искажения переменные сигналы и «блокировать» (т.е. не пропускать) постоянные составляющие сигналов. ДК успешно выполняет такую функцию при наличии постоянного смещения E, превышающего амплитуду переменного сигнала.

---

1 Зависимость емкости ДК от температуры обусловлена функцией $\Delta\varphi_0(T)$, которая входит в выражение (3.25).

Пусть, например, полное обратное напряжение на ДК имеет вид

$$U = E + U_m \sin \omega t ,$$

где E = const и $U_m \ll E$. В этом случае напряжение U меняется в пределах от $E + U_m$ до $E - U_m$, т.е. весьма незначительно. Соответственно емкость ДК остается практически постоянной и равной значению $C(E)$. Переменная составляющая тока будет определяться реактивной проводимостью этой емкости, как в «обычном» конденсаторе:

$$I_m = Y_C U_m ,$$

где $Y_C = \omega\ C(E)$.

Важной особенностью ДК является возможность менять значение емкости, меняя смещение E. Следовательно, ДК можно использовать не только в качестве «обычного» конденсатора с постоянной емкостью, но и в качестве *конденсатора с электрически управляемой емкостью* или, как говорят, *конденсатора переменной емкости*. Такие конденсаторы необходимы, например, для настройки колебательных контуров в радиотехнике. Электрическая регулировка емкости, разумеется, предпочтительнее обычной механической. Однако диапазон электрической регулировки ограничен: меняя смещение E от 1 до 10 В, можно согласно (3.25) изменить емкость ДК всего в 2–2,5 раза.

Добротность. Важным параметром всякого конденсатора, в том числе ДК, является высокочастотная добротность $Q_в$. Она характеризует потери мощности при протекании емкостного тока и определяется как отношение реактивного сопротивления конденсатора к активному:

$$Q = \frac{X_C}{r_в} = \frac{1}{\omega C r_в}, \qquad (7.8а)$$

где $r_в$ — сопротивление потерь на высоких частотах (рис. 7.42, *а*). Чем меньше активная мощность по сравнению с реактивной, тем больше добротность. Например, если C = 100 пФ, $r_в$ = 20 Ом и f = 1 МГц, то $Q_в \approx 75$. У идеального конденсатора $r_в = 0$ и $Q_в = \infty$.

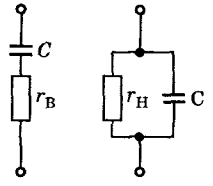


Рис. 7.42. Физические модели диффузионных конденсаторов на ВЧ (*а*) и на НЧ (*б*)

Главным источником потерь в ДК являются горизонтальные сопротивления нижних слоев, входящих в состав p–n-переходов. Для перехода БК — это сопротивление коллекторного слоя (рис. 7.41), а для перехода БЭ — базового. При наличии скрытого n^+-слоя сопротивление $r_в$ для перехода БК значительно меньше, чем при использовании коллекторного (табл. 7.4).

Из выражения (7.8а) очевидно, что добротность возрастает с уменьшением частоты. Однако при достаточно низких частотах становится существенным другой тип потерь, которым на высоких частотах можно пренебречь. Речь идет о сопротивлении потерь $r_н$, обусловленном обратным током p–n-перехода, т.е. о сопротивлении утечки. Это сопротивление шунтирует емкость ДК (рис. 7.42, *б*). Поэтому *добротность на низких частотах* определяют как отношение реактивной проводимости конденсатора к активной:

$$Q_н = \frac{\omega C}{1/r_н} = \omega C r_н, \qquad (7.8б)$$

где $r_н$ — сопротивление потерь (утечки) на низких частотах.

Поскольку сопротивление $r_н$ определяется обратным током перехода, оно обратно пропорционально площади перехода. Тем самым произведение $Cr_н$, а значит, и низкочастотная добротность не зависят от площади. Типичное значение $Q_н$ на частоте 500 Гц составляет 50–100.

В том диапазоне частот, в котором значения обеих добротностей $Q_в$ и $Q_н$ превышают 100–200, ДК представляет собой почти идеальную емкость, т.е. на эквивалентных схемах (рис. 7.42) можно не учитывать сопротивление потерь. Из приведенных примеров следует, что ДК является почти идеальным в диапазоне частот 500 Гц – 500 кГц.

Эквивалентная схема. Специфической особенностью ДК как элемента ИС является наличие у него паразитной емкости. При использовании перехода БК — это барьерная емкость между коллекторным слоем и подложкой $C_{пар} = C_{кп}$ (см. рис. 7.14). *Наличие паразитной емкости приводит к неполной передаче напряжения через ДК в нагрузку.*

Действительно, из эквивалентной схемы на рис. 7.43 видно, что ДК вместе с паразитной емкостью образует емкостной дели-

тель напряжения. Поэтому на выход проходит только часть входного напряжения $U_{вх}$:

$$U_{вых} = U_{вх} \frac{X_{пар}}{X_{пар} + X_C},$$

где $X_{пар}$ и X_C — реактивные сопротивления паразитной и рабочей емкостей $C_{пар}$ и C. Подставляя $X_{пар} = 1/\omega C_{пар}$ и $X_C = 1/\omega C$, запишем коэффициент передачи напряжения в виде

$$U_{вых}/U_{вх} = C/(C + C_{пар}). \qquad (7.9)$$

Коэффициент передачи будет близок к единице, если выполняется неравенство $C_{пар} < C$. Однако площади обоих конденсаторов (рабочего и паразитного) почти одинаковы (рис. 7.43).

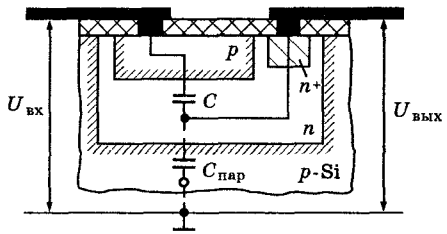


Рис. 7.43. Роль паразитной емкости при передаче переменного напряжения через диффузионный конденсатор

Более того, площадь паразитного конденсатора $C_{кп}$ даже несколько больше площади рабочего. Поэтому емкости $C_{пар}$ и C различаются только благодаря различию удельных емкостей переходов БК и КП и различию напряжений на этих переходах. Расчеты показывают, что в реальных структурах ИС паразитную емкость $C_{пар}$ не удается сделать меньше $(0{,}15–0{,}2)C$. Соответственно коэффициент передачи не превышает 0,8–0,9. Аналогичные выводы и аналогичная эквивалентная схема действительны и для ДК, использующего переход БЭ.

В последнее время в качестве накопительных конденсаторов в элементах памяти широкое применение находят структуры на основе U-канавок (рис. 7.11). По существу, использование таких конденсаторов эквивалентно «вертикальному» интегрированию структур, когда область накопления заряда располагается не на поверхности кристалла, а в его объеме. Это позволяет резко увеличить степень интеграции.

МОП-конденсатор. Интегральным конденсатором, принципиально отличным от ДК, является МОП-конденсатор. Его типичная структура показана на рис. 7.44. Здесь над эмиттерным

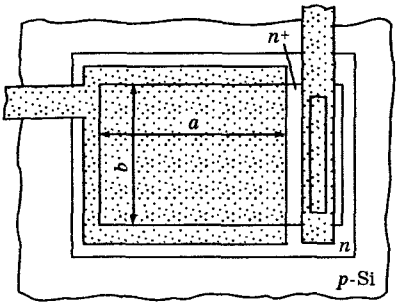


Рис. 7.44. МОП-конденсатор с диэлектриком

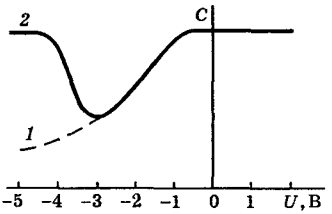
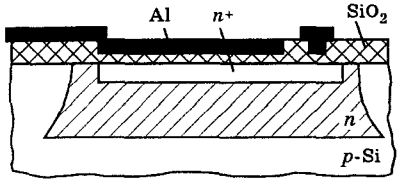


Рис. 7.45. Вольт-фарадная характеристика МОП-конденсатора: *1* — без образования инверсионного слоя; *2* — при образования инверсионного слоя



n^+-слоем с помощью дополнительных технологических процессов выращен слой тонкого (0,08–0,12 мкм) окисла. В дальнейшем, при осуществлении металлической разводки, на этот слой напыляется алюминиевая верхняя обкладка конденсатора. Нижней обкладкой служит эмиттерный n^+-слой.

Удельная емкость МОП-конденсатора выражается формулой (7.4) и обычно составляет около 350 пФ/мм^2. Основные параметры МОП-конденсаторов приведены в табл. 7.4.

Важным преимуществом МОП-конденсаторов по сравнению с ДК является то, что они работают при любой полярности напряжения, т.е. аналогичны «обычному» конденсатору. Однако МОП-конденсатор, как и ДК, тоже нелинейный; пример вольт-фарадной характеристики (которую обычно называют C–U характеристикой) показан на рис. 7.45.

Зависимость $C(U)$ обусловлена тем, что емкость МОП-конденсатора, вообще говоря, представляет собой последовательное соединение двух емкостей: емкости диэлектрика (о которой говорилось выше) и емкости обедненного слоя, который может образоваться в приповерхностной области полупроводника. В конденсаторе, показанном на рис. 7.44, при нулевом и положительном напряжениях на металлической обкладке приповерхностная область оказывается обогащенной электронами, т.е. обедненный слой отсутствует. Соответственно емкость конденсатора определяется диэлектриком и имеет максимальное значение.

При отрицательных напряжениях постепенно образуется обедненный слой. Его глубина растет с ростом напряжения, а емкость соответственно уменьшается. Это приводит к уменьшению результирующей емкости МОП-конденсатора (кривая *1*). При достаточно большом отрицательном напряжении вблизи поверхности образуется инверсионный дырочный слой, т.е. проводящий канал. Тогда емкость обедненного слоя оказывается «отключенной» от емкости диэлектрика, и результирующая емкость МОП-конденсатора снова приближается к начальному значению (кривая *2*).

Для того чтобы влияние обедненного слоя было незначительным, необходимо, чтобы емкость этого слоя была большой по сравнению с емкостью диэлектрика. Такое требование удовлетворяется при большой концентрации примеси в полупроводнике. Именно поэтому в качестве полупроводниковой «обкладки» МОП-конденсатора используется n^+-слой. Одновременно, благодаря малому сопротивлению этого слоя, обеспечивается высокая добротность конденсатора.

Паразитная емкость МОП-конденсаторов учитывается с помощью уже известной эквивалентной схемы (рис. 7.43), где под емкостью $C_{пар}$ следует понимать емкость между n-карманом и p-подложкой. Коэффициент передачи (7.9) в данном случае составляет не менее 0,9–0,95.

Важной особенностью МОП-конденсаторов является зависимость их емкости от частоты. Такая зависимость обусловлена влиянием быстрых поверхностных состояний на границе полупроводник—диэлектрик. Перезаряд этих состояний является

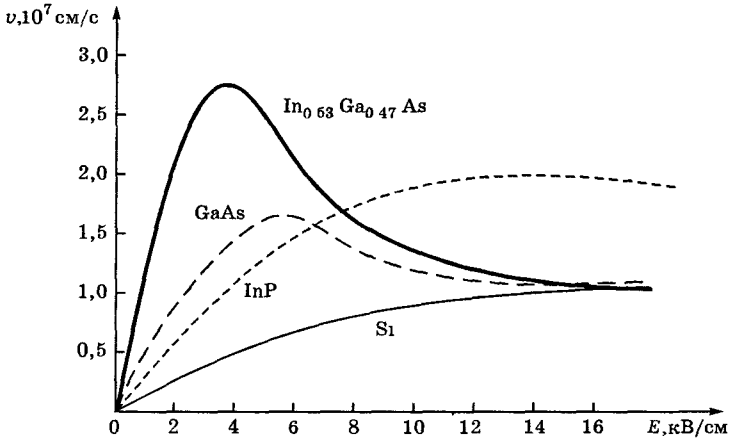


Рис. 7.46. Зависимость дрейфовой скорости электронов проводимости от напряженности электрического поля

инерционным процессом и происходит с постоянной времени порядка 0,1 мкс. Поэтому с ростом частоты емкость МОП-конденсатора уменьшается и достигает установившегося значения лишь при частотах более нескольких мегагерц.

В заключение заметим, что в МОП-транзисторных ИС, в отличие от биполярных, изготовление МОП-конденсаторов не связано с дополнительными технологическими процессами: тонкий окисел для конденсаторов получается на том же этапе, что и тонкий окисел под затвором, а низкоомный полупроводниковый слой — на этапе легирования истока и стока. Изолирующие карманы в МОП-технологии, как известно, отсутствуют.

7.11. Элементы ИС на полупроводниках группы $A^{III}B^{V}$

Полупроводники группы $A^{III}B^{V}$ как материалы для ИС. Сравнительный анализ свойств полупроводников группы $A^{III}B^{V}$ и кремния позволяет выявить три основных различия при создании на их основе активных элементов ИС:

1. Сильно отличаются зависимости дрейфовой скорости электронов проводимости, особенно при низких полях.

2. Из полупроводников группы $A^{III}B^{V}$ легко изготовить полуизолирующую подложку.

3. Эти полупроводники (арсенид галлия, фосфид индия, арсенид индия и др.) образуют твердые растворы не только двух-, но и трех- и четырехкомпонентные с широким диапазоном изменения их электрофизических свойств.

На рис. 7.46 приведена зависимость дрейфовой скорости электронов проводимости от напряженности электрического поля для двух- и трехкомпонентных полупроводников $A^{III}B^{V}$ и кремния. Причиной того, что зависимость скорости дрейфа электронов проводимости от поля в GaAs и InP имеет не монотонный, как у кремния, а экстремальный характер, кроется в различии зонной структуры этих материалов. Зонная структура GaAs приведена на рис. 7.47.

Абсолютный минимум зоны проводимости GaAs находится в т. Г. Во втором минимуме в т. X энергия на 0,36 эВ выше. Эффективная масса электронов проводимости в т. Г мала ($0{,}068m$, где m —масса покоя свободных электронов в вакууме) и поэтому дрейфовая подвижность электронов проводимости велика.

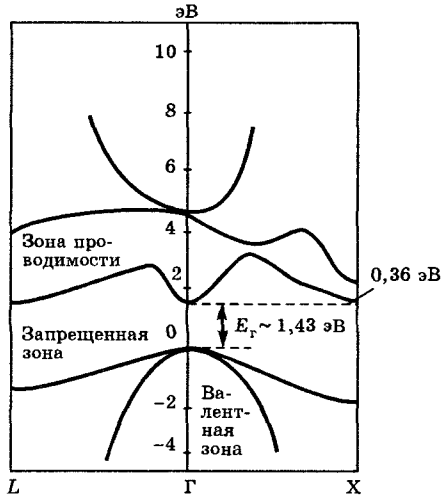


Рис. 7.47. Зонная структура GaAs

Как видно из рис. 7.47, в т. X имеет место второй минимум энергии. Однако здесь эффективная масса электронов значительно выше (1,2 m), и дрейфовая скорость электронов проводимости уменьшается. Этим объясняется участок с отрицательной дифференциальной проводимостью на рис. 7.46. Этот эффект получил название эффекта Ганна и используется для генерации высокочастотных сигналов. В ИС на GaAs этот эффект нежелателен и для его подавления необходимо выполнить условие

$$\frac{\Delta E}{E_\Gamma} > 1. \qquad (7.10)$$

Здесь ΔE — энергетический зазор между т. Г и т. X, а E_Γ — ширина запрещенной зоны в т. Г зоны проводимости.

Элементы ИС на полупроводниках группы $A^{III}B^{V}$. Элементы ИС на $A^{III}B^{V}$ можно грубо разделить на следующие группы: полевые транзисторы с затвором Шоттки, гетеробиполярные транзисторы и МДП-транзисторы.

Несмотря на то, что технологию изготовления материалов $A^{III}B^{V}$ и приборов на их основе по целому ряду причин нельзя считать столь же отработанной, как для кремниевых структур, исследования и разработки уже продемонстрировали возможности создания сверхбыстродействующих цифровых схем, конкурирующих с кремниевыми. Можно предположить, что главную роль в конкурентной борьбе будут играть полевые транзисторы с затвором Шоттки, т.к. они просты в изготовлении и потребляют незначительную энергию.

Структура и принцип работы полевого транзистора с затвором Шоттки (ЗШП-транзистора). Структура GaAs—ЗШП-транзистора изображена на рис. 7.48, *а*.

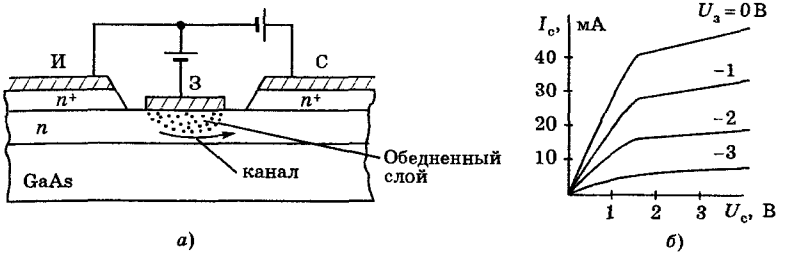


Рис. 7.48. Структура (*а*) и статические ВАХ (*б*) ЗШП-транзистора

На полуизолирующей пластине GaAs с высоким удельным сопротивлением (порядка 10^8 Ом·см) формируют тонкий электропроводный слой n-типа, который называют активным слоем. В активном слое расположены области истока, затвора и стока. Для получения хороших омических контактов на активном слое под электродами истока и стока располагаются n^+-области. Электрод затвора образует с активным слоем контакт Шотки (см. раздел 3.3). Активный слой обычно формируется ионной имплантацией, например кремния. Толщина этого слоя 0,1–0,3 мкм, а концентрация примеси около $3 \cdot 10^{17}$ $см^{-3}$. Чем короче длина затвора, тем выше быстродействие прибора, в настоящее время длина затвора составляет 0,2–1,0 мкм. Омические электроды изготавливают из сплава Al–Ge–Ni, а электрод затвора — из Al или силицида вольфрама.

Если к стоку приложить положительное напряжение, а к затвору отрицательное, обедненный слой под затвором расширится, а канал между истоком и стоком сузится и ток в цепи от истока к стоку изменится. Поскольку выходной сигнал стока можно регулировать малым входным сигналом затвора, такой прибор можно использовать как усилительный элемент. На рис. 7.48, *б* приведены статические характеристики такого прибора.

Длина канала в ЗШП-транзисторах является важнейшим параметром, определяющим быстродействие элемента. При использовании технологии самосовмещенной ионной имплантации при образовании слоев n^+-типа оказывается возможным управлять величиной зазора между ними, что обеспечивает снижение последовательного сопротивления исток-сток и позволяет формировать затворы длиной около 0,1 мкм.

Транзисторы с высокой подвижностью электронов. Известны различные варианты конструкции транзисторов с высокой подвижностью электронов (ВПЭТ). Общее представление о структуре ВПЭТ с одним гетеропереходом дает рис. 7.49. Исток, сток и затворный электрод формируются на поверхности GaAs. Повышение подвижности электронов в канале достигается за счет использования гетероперехода GaAs с материалом с большой шириной запрещенной зоны ($Al_{0,3}$ $Ga_{0,7}$ As) с введенной в него донорной примесью.

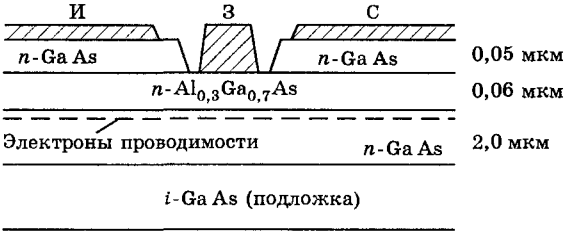


Рис. 7.49. Структура ВПЭТ на GaAs

Создание такой структуры возможно методами либо высококонтролируемого эпитаксиального наращивания слоев $A^{III}B^{V}$, либо методами молекулярно-лучевой эпитаксии. При реализации таких приборов удается достигнуть значений подвижности электронов до 10^5 см^2/В · с при 77 К и до $2 \cdot 10^5$ см^2/В · с при 4,2 К. Для сравнения отметим, что в обычных ЗШП-транзисторах подвижность электронов в слоях n-типа GaAs составляет всего $5 \cdot 10^3$ см^2/В · с, т.е. меньше на два порядка.

7.12. Элементы пленочных ИС

Как известно, промышленная технология активных пленочных элементов (диодов и транзисторов) пока не разработана. Поэтому ниже рассматриваются только пассивные элементы: резисторы, конденсаторы и катушки индуктивности. Эти элементы могут изготавливаться как по тонкопленочной, так и по толстопленочной технологии. Конфигурации тонко- и толстопленочных элементов одинаковы, но их конкретные геометрические размеры (при заданных электрических параметрах) могут существенно различаться в связи с использованием совершенно разных материалов.

Пленочные элементы нет необходимости изолировать друг от друга, так как все они выполняются на диэлектрической подложке. Поскольку подложка сравнительно толстая (не менее 500 мкм), а расстояния между элементами сравнительно большие, паразитные емкости практически отсутствуют и их учет на эквивалентных схемах обычно не имеет смысла.

Резисторы. Структура и конфигурации пленочного резистора показаны на рис. 7.50. Как видим, в общем случае конфигурация пленочного резистора такая же, как диффузионного (рис. 7.35), т.е. зигзагообразная. В частности, она может быть полосковой. Поэтому расчет сопротивления можно проводить по формулам (7.5). Удельное сопротивление слоя зависит от толщины слоя и материала. Типичные значения R_s приведены в табл. 7.5. Там же приведены типичные значения других параметров резисторов: максимального и минимального номиналов сопротивления, разброса номиналов (δ), температурного коэффициента (ТКС) и временного дрейфа сопротивления (за 1000 ч при температуре + 70 °С).

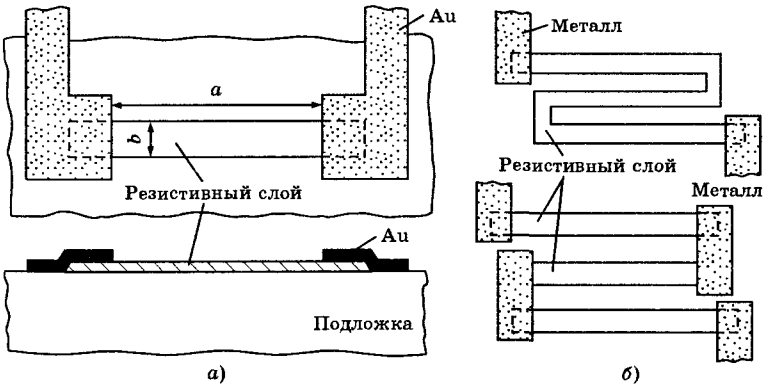


Рис. 7.50. Пленочные резисторы полосковой (*а*) и зигзагообразной (*б*) конфигурации

Значение разброса (допуска) δ приведены для двух случаев: когда отсутствует специальная подгонка (юстировка) резисторов после их изготовления (б/подг.) и после такой подгонки (с подг.)[1]. Значения ТКС также приведены для двух величин: для

1 Способы подгонки рассмотрены ниже.

сопротивления отдельного резистора (R) и для отношения сопротивлений двух резисторов (R_1/R_2).

Таблица 7.5. Типичные параметры пленочных резисторов

Тип резистора	R_s, Ом/□	$R_{макс}$, Ом	$R_{мин}$, Ом	δ, %		ТКС, 10^{-4}/°С		$\Delta R(t)$, % (1000 ч, +70°С)
				б/подг.	с подг.	R	R_1/R_2	
Тонкопленочный	10–300	10^6	10	±5	±0,05	0,25	±0,05	0,005
Толстопленочный	5,0–10^6	$5 \cdot 10^8$	0,5	±15	±0,2	2	±0,1	0,05

Из табл. 7.5 можно сделать следующие общие выводы:

- диапазон сопротивлений пленочных резисторов несравненно шире, чем полупроводниковых (диффузионных и ионно-легированных);
- тонкопленочная технология обеспечивает более высокую прецизионность и стабильность резисторов;
- подгонка обеспечивает существенное уменьшение разброса (допусков) сопротивлений; следовательно, возможность такой подгонки является важным преимуществом пленочных резисторов;
- отношение сопротивлений, как и в случае полупроводниковых ИС, характеризуется меньшим разбросом и меньшим ТКС, чем отдельное сопротивление.

Подгонку резисторов можно осуществлять разными способами. Простейший, исторически первый способ состоит в частичном механическом соскабливании резистивного слоя до того, как поверхность ИС защищается тем или иным покрытием. Более совершенными являются методы частичного удаления слоя с помощью электрической искры, электронного или лазерного луча. Разумеется, все эти способы позволяют только увеличивать сопротивление резистора. Наиболее совершенный и гибкий метод состоит в пропускании через резистор достаточно большого тока. При токовой подгонке одновременно идут два процесса:

окисление поверхности резистивного слоя и упорядочение его мелкозернистой структуры. Первый процесс способствует увеличению, а второй — уменьшению сопротивления. Подбирая силу тока и атмосферу, в которой ведется подгонка, можно обеспечить изменение сопротивления и в ту, и в другую сторону на ± 30% с погрешностью (по отношению к желательному номиналу) до долей процента.

Конденсаторы. Структура и конфигурация типичного пленочного конденсатора показаны на рис. 7.51. Удельная емкость конденсатора определяется по формуле (7.4), где толщина диэлектрической пленки d существенно зависит от технологии: для тонких пленок d = 0,1–0,2 мкм, для толстых d = 10–20 мкм. Поэтому при прочих равных условиях удельная емкость толстопленочных конденсаторов меньше, чем тонкопленочных. Однако различие в толщине диэлектрика может компенсироваться благодаря различию диэлектрических проницаемостей материалов.

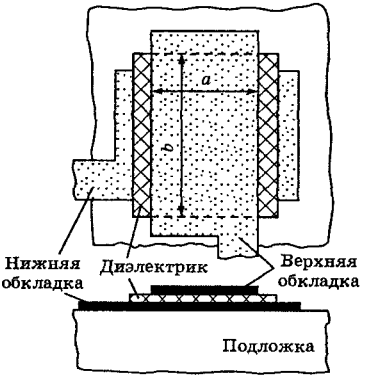


Рис. 7.51. Пленочный конденсатор

У тонкопленочных конденсаторов удельная емкость не пропорциональна диэлектрической проницаемости используемого материала, так как учитывается еще его пробивная напряженность. Материал с высоким значением ε может иметь малую пробивную напряженность. Тогда при заданном пробивном напряжении толщину диэлектрического слоя необходимо увеличивать, и выигрыш в удельной емкости оказывается меньше ожидаемого.

При выборе диэлектрика для высокочастотных конденсаторов (как тонко-, так и толстопленочных) приходится дополнительно учитывать потери энергии в диэлектрике[1]. Что касается омических потерь в обкладках пленочных конденсаторов, то

1 Эти потери характеризуют параметром «тангенс дельта» (tgδ) — тангенсом угла между векторами полного тока и реактивной составляющей тока через конденсатор при заданной частоте. Если потери незначительны, то $\mathrm{tg}\delta \approx \delta << 1$.

они гораздо меньше, чем у полупроводниковых конденсаторов, потому что в качестве обкладок используются металлические слои с высокой проводимостью.

В табл. 7.6 приведены типичные параметры пленочных конденсаторов. Для сравнения воспроизведены также параметры близких им по структуре МОП-конденсаторов. Из таблицы можно сделать следующие общие выводы:

- удельные емкости пленочных конденсаторов (при надлежащем выборе диэлектрика) в несколько раз и даже на порядок превышают удельную емкость МОП-конденсаторов и тем более диффузионных конденсаторов;
- максимальные емкости пленочных конденсаторов могут быть на несколько порядков больше, чем емкости полупроводниковых конденсаторов, главным образом благодаря большей площади (поскольку площадь подложек гибридных ИС значительно превышает площадь кристаллов полупроводниковых ИС);
- толстопленочные конденсаторы незначительно уступают тонкопленочным по большинству параметров, за исключением, может быть, температурного коэффициента;
- для высокочастотных тонкопленочных конденсаторов оптимальным диэлектриком является моноокись кремния; близкими к ней параметрами обладает также моноокись германия.

Таблица 7.6. Типичные параметры пленочных конденсаторов

Тип конденсатора		C_0, пФ/мм^2	$C_{макс}$, пФ (S = 25 мм^2)	δ, %	ТКЕ, %/°С	Q (10 МГц)
Тонкопленочный	SiO	60	1500	± 15	0,2	200
	Al_2O_3	1500	$4 \cdot 10^4$	± 15	0,03	30
	Ta_2O_5	4000	10^5	± 15	0,02	30
Толстопленочный		—	10^4	± 20	± 0,05... ± 0,15	—
МОП		350	200	± 20	0,02	10

Следует заметить, что в последнее время, в связи с наличием миниатюрных дискретных конденсаторов (в том числе с весьма большой емкостью — до нескольких микрофарад), наблюдается тенденция к отказу от пленочных конденсаторов и замене их навесными конденсаторами.

Катушки индуктивности. Как уже отмечалось, возможность осуществлять катушки индуктивности методами микроэлектроники является одним из достоинств пленочной технологии. Такие катушки представляют собой плоские спирали, обычно прямоугольной конфигурации (рис. 7.52). Для уменьшения сопротивления в качестве материала используется золото. Ширина металлической полоски составляет 30–50 мкм, просвет между витками 50–100 мкм. При таких геометрических размерах удельная индуктивность лежит в диапазоне 10–20 нГн/мм^2, т.е. на площади 25 мм^2 можно получить индуктивность 250–500 нГн.

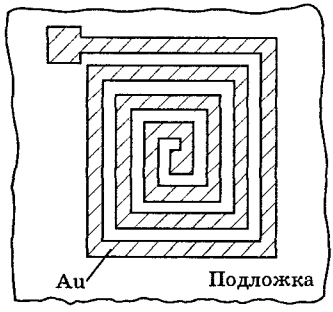


Рис. 7.52. Пленочная катушка индуктивности

Добротность катушек индуктивности на высокой частоте определяется выражением

$$Q_в = \omega L / r_в, \qquad (7.11)$$

где $r_в$ — сопротивление высокочастотных потерь. Например, на частоте 100 МГц добротность может иметь значение $Q_в = 50$. В отличие от добротности конденсатора [см. (7.8а)] добротность катушки возрастает с увеличением частоты. Поэтому пленочные катушки могут успешно работать в диапазоне сверхвысоких частот (СВЧ), при частотах 3–5 ГГц. При этом число витков составляет 3–5.

В связи с разработкой микроминиатюрных проволочных катушек применение пленочных катушек, особенно на частотах менее 50–100 МГц, ограничивается, и предпочтение, как и в случае конденсаторов, отдается навесным компонентам.

Контрольные вопросы

1. Дайте определение элементам интегральных схем.
2. Какие основные методы изоляции элементов вы знаете?
3. В чем их принципиальные достоинства и недостатки?
4. В каком направлении продолжается прогресс в разработке методов изоляции элементов?
5. Нарисуйте конфигурацию и назовите характер и параметры распределения примесей в транзисторной n–p–n-структуре.
6. Почему пробивное напряжение коллекторного перехода n–p–n-транзистора больше эмиттерного?
7. Нарисуйте структуру интегрального n–p–n- (p–n–p)-транзистора и укажите на ней все паразитные параметры. Какими из этих параметров ограничивается время переключения интегральных транзисторов?
8. Какие разновидности интегральных биполярных транзисторов вы знаете?
9. Чем определяется проблематика создания сверхтонкой базы биполярных транзисторов?
10. Перечислите основные операции создания структур биполярных транзисторов.
11. Нарисуйте несколько вариантов диодного включения транзисторов.
12. Чем отличаются структуры полевого и МДП-транзисторов?
13. Нарисуйте структуру комплементарной пары МОП-транзисторов.
14. В чем заключается принцип самосовмещения при изготовлении МОП-транзисторов? При изготовлении каких еще других интегральных структур используется этот принцип?
15. Как изготавливается поликремниевый затвор? Каковы преимущества использования поликремниевой металлизации?
16. С какой точностью и какого порядка величины можно изготовить диффузионный резистор? Какими методами можно поднять пределы значений резисторных структур?
17. Какие структуры используются в качестве конденсаторных? Что ограничивает предельные параметры полупроводниковых конденсаторов?

18. Почему использование материалов группы $A^{III}B^{V}$ позволяет повысить быстродействие ИС?
19. Нарисуйте структуру полевого транзистора с затвором Шотки на GaAs. Поясните принципы его создания и работы.
20. Какими средствами можно повысить быстродействие простого полевого транзистора с затвором Шоттки?
21. Какие материалы используются для создания пленочных резисторов? Какие предельные параметры пленочных резисторов можно обеспечить? С какой точностью?
22. Какие материалы используются и какого порядка предельные параметры можно получить при создании пленочных конденсаторов?
23. Можно ли изготовить по технологии кремниевых ИС катушки индуктивности?
24. Какого порядка величины и добротности можно изготовить индуктивности по тонкопленочной и толстопленочной технологии?

Глава 8

Основы цифровой схемотехники

8.1. Введение

Схемотехника микроэлектронных устройств, возникшая на базе дискретной транзисторной электроники, является непрерывно развивающейся инженерной дисциплиной. В ходе более чем 30-летнего развития происходил своего рода «естественный отбор» из огромного количества схемных решений. В последующих главах рассматриваются схемы, составляющие «золотой фонд», который наиболее часто используется в современной схемотехнике и является основой ее дальнейшего развития.

Классификация электронных схем. В настоящее время принято делить электронные схемы на два класса: *цифровые и аналоговые*.

В основе цифровых схем лежат простейшие *транзисторные ключи* — аналоги металлических контактов. Ключи характерны двумя устойчивыми состояниями: разомкнутым и замкнутым. На основе простейших ключей строятся более сложные схемы: логические, бистабильные, триггерные (спусковые) и др.

В основе аналоговых схем лежат простейшие усилительные ячейки — *каскады* или *ступени*. На основе усилительных каскадов строятся сложные (многокаскадные) усилители, стабилизаторы напряжения и тока, преобразователи частоты (модуляторы и детекторы), генераторы синусоидальных колебаний и ряд других схем. Такие схемы иногда называют линейными или квазилинейными, хотя это название подходит только для усилителей и стабилизаторов, а для остальных схем является условным.

Специфику цифровых и аналоговых схем удобно пояснить с помощью так называемых *передаточных характеристик*, описывающих зависимость выходной величины от входной. Для определенности примем, что эти величины — напряжения.

На рис. 8.1 показаны типичные передаточные характеристики схем. Цифрой *1* обозначена характеристика *инвертирую-*

щих схем, у которых низким входным напряжениям соответствуют высокие выходные, а цифрой *2* — характеристика *неинвертирующих* схем, у которых низким входным напряжениям соответствуют низкие выходные. Инвертирующие схемы имеют большое распространение. Такие же характеристики свойственны и простейшим ключам, и простейшим усилительным каскадам. Однако использование этих характеристик в том и другом классе схем принципиально различается.

В транзисторном ключе два его устойчивых состояния (разомкнутое и замкнутое) соответствуют точкам A и B (рис. 8.1). В точке A ключ разомкнут и на нем падает большое напряжение, а в точке B ключ замкнут и падение напряжения на нем близко к нулю. Входные и выходные сигналы (напряжения) в ключе принимают только два значения: либо $U_{вхA}$ и $U_{выхA}$, либо $U_{вхB}$ и $U_{выхB}$. Форма передаточной характеристики между точками A и B несущественна; если она меняется (как показано штриховой линией), то выходные сигналы остаются практически неизменными. Отсюда следует, что ключи, а значит, и цифровые схемы мало чувствительны к разбросу (допускам) параметров, к их температурной зависимости (температурному дрейфу), к изменению параметров со временем (временному дрейфу), а также к внешним электромагнитным помехам (наводкам) и собственным шумам. Последний вывод иллюстрируется на рис. 8.1 тем, что небольшие колебания напряжения ΔU_B около точки B (это могут быть шумы или наводки) практически не меняют значения выходного сигнала, а значит, не влияют на работу ключа.

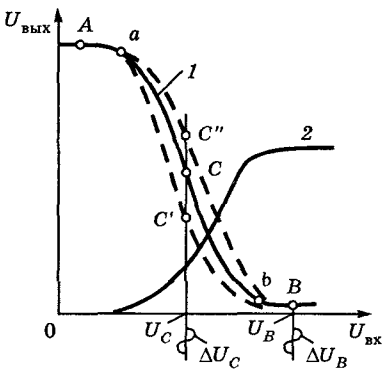


Рис. 8.1. Передаточные характеристики электронных схем:
1 — инвертирующих;
2 — неинвертирующих

В усилительном каскаде используется непрерывный рабочий участок передаточной характеристики между точками a и b. Входные и выходные сигналы могут принимать любые значения в пределах этого участка и связаны друг с другом функциональной зависимостью $U_{вых} = f(U_{вх})$. Очевидно, что любая

«деформация» характеристики на участке $a—b$, по каким бы причинам она ни произошла, будет непосредственно отражаться на указанной функциональной зависимости и на работе схемы. Например, при одном и том же входном сигнале $U_{вхC}$ выходной сигнал может принимать различные значения $U_{вых\,C'}$, $U_{вых\,C''}$. Отсюда следует, что усилительный каскад, а значит, и аналоговые схемы чувствительны к разбросу параметров, к их температурному и временному дрейфу, а также к шумам и наводкам. Последний вывод иллюстрируется на рис. 8.1 тем, что небольшие колебания напряжения ΔU_C около точки C вызывают заметные изменения выходного сигнала в соответствии с функцией $U_{вых} = f(U_{вх})$.

Внутри каждого класса электронных схем — цифровых и аналоговых — разумеется, имеет место более детальная классификация, прежде всего по выполняемым функциям.

Мы начнем изучение схемотехнических основ микроэлектроники с транзисторных ключей и ключевых схем, как более простых, хотя исторически первыми развивались аналоговые схемы (именно они составляли основу радиотехники — первой области применения электронных приборов).

8.2. Статический режим простейшего биполярного ключа

В статическом режиме ключа, т. е. в двух его устойчивых точках (A и B на рис. 8.1), идеальными параметрами обладают металлические контакты, замыкаемые и размыкаемые путем механического перемещения. У таких механических ключей остаточный ток в разомкнутом состоянии определяется качеством изоляции и обычно не превышает 10^{-12} А. В замкнутом состоянии остаточное напряжение на контакте составляет доли микровольта при токах порядка 1 мА. По этим параметрам механические ключи пока остаются вне конкуренции.

Однако в динамическом режиме, т. е. при переключении из одной устойчивой точки в другую, механические ключи значительно уступают электронным по максимальной частоте переключений, надежности контакта и сроку службы. Эти показатели оказали решающее влияние на замену механических ключей электронными в цифровых устройствах.

Рабочие точки. На рис. 8.2 показана схема простейшего транзисторного ключа. Транзистор включен по схеме с общим эмиттером (ОЭ). Управляемой (прерываемой) является коллекторная цепь с источником питания $E_к$ и нагрузкой в виде резистора $R_к$. В управляющей (базовой) цепи включен источник управляющего напряжения $E_б$ и последовательное сопротивление $R_б$.

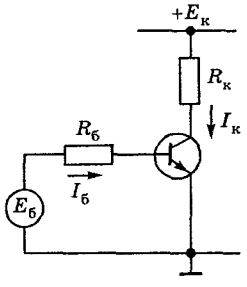


Рис. 8.2. Простейший транзисторный ключ

Если напряжение $E_б$ имеет отрицательную полярность, то эмиттерный переход смещен в обратном направлении, транзистор заперт, и остаточный ток в цепи нагрузки очень мал. Соответственно напряжение на ключе $U_{кэ}$ близко к $E_к$.

Если напряжение $E_б$ имеет положительную полярность и достаточно велико, то транзистор открыт, в цепи нагрузки протекает ток $I_к$, и остаточное напряжение на ключе может быть близким к нулю.

Из сказанного следует, что рассматриваемый ключ является инвертирующей схемой, так как увеличение входного напряжения $E_б$ от отрицательных значений к положительным сопровождается уменьшением выходного напряжения $U_{кэ}$ от $E_к$ до малого остаточного напряжения.

Остаточный ток и остаточное напряжение — главные статические параметры ключа. Рассмотрим их подробнее.

В запертом состоянии ключа, строго говоря, должно выполняться условие $E_б \leq 0$. Однако учитывая, что ток p–n-перехода экспоненциально зависит от приложенного напряжения, для каждого конкретного диапазона режимных токов можно определить достаточно резкую границу области прямых смещений, при которых этим током можно пренебречь. Поэтому кремниевый n–p–n-транзистор ключа может считаться практически запертым, если $E_б < (0{,}5–0{,}6)$ В. При этом токи всех трех электродов транзистора обычно не превышают долей микроампера. Поэтому падениями напряжения на сопротивлениях $R_б$ и $R_к$ можно пренебречь и считать $U_б = E_б$ и $U_{кэ} = E_к$. Запертому состоянию ключа соответствуют точки A на рис. 8.3.

Когда напряжение $E_б$ достигает значения U^*, транзистор отпирается. Начинает протекать базовый ток $I_{б1}$ и пропорциона-

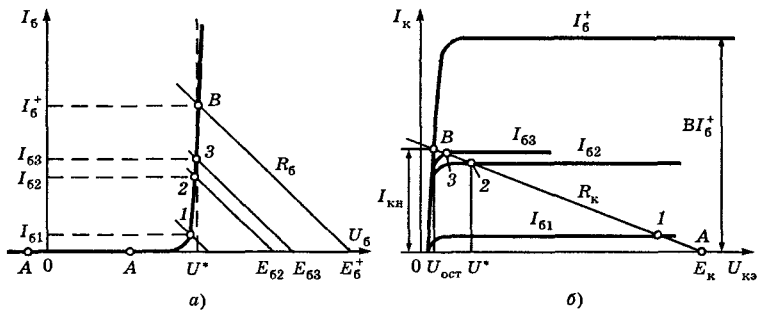


Рис. 8.3. Расположение рабочих точек ключа на статических характеристиках транзистора. *а* — входной, *б* — выходной

льный ему коллекторный ток $I_{к1}$, а потенциал коллектора соответственно уменьшается (см. точки *1* на рис. 8.3). При дальнейшем росте напряжения $E_б$ потенциал базы $U_б$ остается равным U^* (рис. 8.3, *а*), но токи продолжают расти, а потенциал коллектора — падать[1].

В точке *2*, при токе $I_{б2}$, потенциал коллектора $U_к$ делается равным U^*, а напряжение на коллекторном переходе $U_{кб} = U_к - U_б$ — равным нулю. При еще больших токах напряжение $U_{кб}$ становится отрицательным, т. е. прямым, и транзистор работает в режиме двойной инжекции. Однако до тех пор, пока прямое напряжение на коллекторном переходе остается меньше напряжения отпирания (0,5–0,6) В, инжекция коллектора несущественна и ток коллектора продолжает расти пропорционально току базы.

Только в точке *3*, при токе $I_{б3}$, когда прямое напряжение $U_{кб}$ достигает напряжения отпирания (а потенциал коллектора соответственно падает до 0,1 В), инжекция коллектора начинает препятствовать дальнейшему увеличению коллекторного тока и этот ток остается далее практически неизменным.

---

1 На рис. 8.3 и далее для нахождения рабочих точек используется *метод линий нагрузки*. Это способ графического определения тока и потенциала в общей точке последовательно включенных элементов с использованием ВАХ этих элементов. По оси напряжений откладывается заданное напряжение (в данном случае величины $E_б$ или $E_к$) и из этой точки проводится вольт-амперная характеристика нагрузки (в данном случае — резисторов $R_б$ или $R_к$). Точка пересечения обеих характеристик — основной и нагрузочной — определяет ток и напряжение в той или иной цепи.

Такой максимальный ток коллектора называют *током насыщения* и обозначают $I_{кн}$. Соответственно и режим двойной инжекции, характерный для открытого состояния ключа, обычно называют *режимом насыщения* транзистора. Открытому (насыщенному) состоянию ключа соответствуют точки B на рис. 8.3. Управляющие ток и напряжение в открытом состоянии обозначены через $I_б^+$ и $E_б^+$.

Остаточное напряжение на ключе в точке B содержит две составляющие:

$$U_{ост} = U_{кэ} + I_{кн} r_{кн}, \qquad (8.1)$$

где $U_{кэ}$ — разность напряжений на переходах, $I_{кн} r_{кн}$ — падение напряжения на омическом сопротивлении коллекторной области транзистора.

Для определения первой составляющей воспользуемся физической моделью биполярного транзистора в виде эквивалентной схемы, представленной на рис. 8.4: $I_{дэ} = \frac{I_{к0}}{B_N}(e^{U_{бэ}/\varphi_T} - 1)$ — ток через эмиттерный диодный элемент, $I_{дк} = \frac{I_{к0}}{B_I}(e^{U_{бк}/\varphi_T} - 1)$ — ток через коллекторный диодный элемент, $I_{кэ} = I_{к0}(e^{U_{бэ}/\varphi_T} - e^{U_{бк}/\varphi_T})$ — ток управляемого источника, $C_э = C_{эб} + I_{дэ}\frac{\tau_э}{\varphi_T}$ — емкость эмиттерного диодного элемента, $C_к = C_{кб} + I_{дк}\frac{\tau_к}{\varphi_T}$ — емкость коллекторного диодного элемента, где $I_{к0}$ — параметр модели управляемого источника тока, B_N и B_I — нормальный и инверсный статические коэффициенты усиления тока базы, $\tau_э$ и $\tau_к$ — эффективное время жизни неосновных носителей в эмиттерном и коллекторном переходах, $C_{эб}$ и $C_{кб}$ — барьерные емкости переходов.

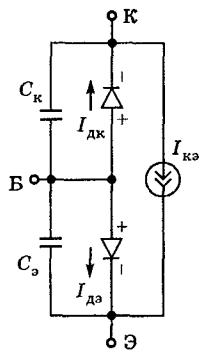


Рис. 8.4. Эквивалентная схема биполярного транзистора

Токи базы и коллектора при заданных напряжениях $U_{бэ}$ и $U_{кэ}$ определяются выражениями:

$$I_б^+ = \frac{I_{к0}}{B_N} e^{\frac{U_{бэ}}{\varphi_T}} + \frac{I_{к0}}{B_I} e^{\frac{U_{бэ}-U_{кэ}}{\varphi_T}},$$

$$I_{кн} = I_{к0}\left(e^{\frac{U_{бэ}}{\varphi_T}} - e^{\frac{U_{бэ}-U_{кэ}}{\varphi_T}}\right) - \frac{I_{к0}}{B_I} e^{\frac{U_{бэ}-U_{кэ}}{\varphi_T}}$$

Откуда, исключая $U_{бэ}$, получаем зависимость $U_{кэ}$ от токов базы и коллектора:

$$U_{кэ} = \varphi_T \ln\left[\left(1+\frac{1}{B_I}\right)\left(\frac{\frac{I_б^+}{I_{кн}}+\frac{1}{B_I+1}}{\frac{I_б^+}{I_{кн}}-\frac{1}{B_N}}\right)\right]. \qquad (8.2)$$

Например, если $B_N = 100$, $B_I = 1$ и $I_{кн} = I_б^+$, то $U_{кэ} \approx 27$ мВ. Если $I_{кн} = 10\, I_б^+$, то $U_{кэ}$ возрастает до 65 мВ. Заметим, что уменьшению составляющей $U_{кэ}$ способствует прежде всего увеличение инверсного коэффициента усиления B_I. Так, если в приведенном примере положить $B_I = 3$, то значения $U_{кэ}$ составят 12 и 40 мВ.

Вторая (омическая) составляющая меняется в широких пределах в зависимости от тока насыщения и структуры транзистора. Например, если $I_{кн} = 100$ мкА и $r_{кк} = 10$ Ом (структура со скрытым слоем), то $I_{кн} r_{кк} = 1$ мВ , т. е. вторая составляющая много меньше первой. Если же $I_{кн} = 2$ мкА и $r_{кк} = 100$ Ом (структура без скрытого слоя), то $I_{кн} r_{кк} = 200$ мВ , т. е. вторая составляющая заметно превышает первую. Именно поэтому в ключевых ИС всегда используют скрытые слои. Обычно полное остаточное напряжение составляет 50–100 мВ.

Критерий насыщения. Из рис. 8.3 нетрудно получить следующие выражения для отпирающего тока базы и тока насыщения коллектора:

$$I_б^+ = (E_б^+ - U^*)/R_б; \qquad (8.3а)$$

$$I_{кн} = (E_к - U_{ост})/R_к \approx E_к/R_к. \qquad (8.3б)$$

Поскольку напряжение U^* слабо зависит от тока, можно считать, что в режиме насыщения токи ключа определяются внешними цепями ($E_б$, $R_б$ и $E_к$, $R_к$). Иначе говоря, при расчете ключа токи $I_б^+$ *и* $I_{кн}$ можно считать заданными, независимыми величинами, а напряжения — функциями токов.

Формальным критерием насыщения ключа (т. е. режима двойной инжекции) является условие, чтобы напряжение на коллекторном переходе было прямым. Однако в тех случаях, когда заданы токи, такой критерий неудобен. Вместо него используют так называемый токовый критерий насыщения:

$$B_N\, I_б^+ > I_{кн}. \qquad (8.4)$$

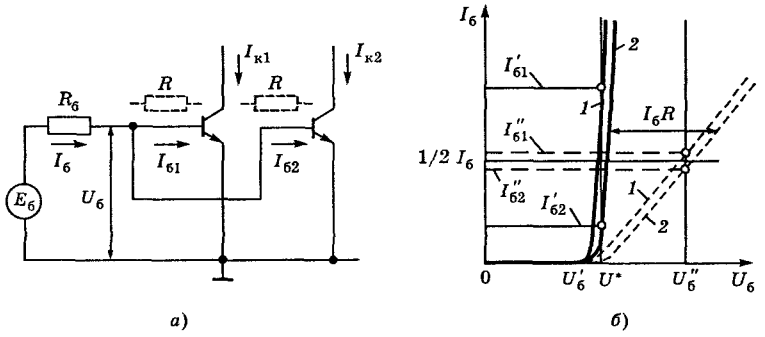
Суть этого критерия состоит в том, что ток базы должен превышать значение, при котором ток коллектора достигает значения $I_{кн}$. Неравенство (8.4) должно быть достаточно сильным, чтобы неизбежные изменения входящих в него величин не приводили к выходу ключа из насыщения и соответственно к резкому возрастанию остаточного напряжения.

Силу неравенства (8.4) характеризуют особым параметром — *степенью насыщения*:

$$S = B_N I_б^+ / I_{кн}. \qquad (8.5)$$

Значение $S = 1$ соответствует границе с активным режимом, значение $S = \infty$ — нулевому коллекторному току, а значение $S = B_N$ — равенству базового и коллекторного токов. На рис. 8.3, *б* степень насыщения составляет $S \approx 2$.

Параллельное соединение ключей. На практике часто один источник напряжения $E_б$ управляет несколькими ключами. Тогда участки база–эмиттер этих ключей оказываются вклю-



а) *б*)

Рис. 8.5. Параллельное соединение ключей:
а — схема; *б* — распределение токов

ченными параллельно. На рис. 8.5. показана схема двух параллельно включенных ключей. Очевидно, что общий ток $I_б$ распределяется (делится) между базами:

$$I_б = I_{б1} + I_{б2}.$$

Если транзисторы и их режимы идентичны, то общий ток делится между базами поровну:

$$I_{б1} = I_{б2} = \tfrac{1}{2} I_б.$$

Если же входные ВАХ неидентичны, то распределение тока $I_б$ при одном и том же напряжении U^* может быть весьма неравномерным. Причиной различия входных ВАХ может быть характер нагрузки. Пусть один транзистор находится в активном режиме, а другой — в глубоком насыщении, что, согласно 8.2, соответствует остаточному напряжению

$$U_{кэ} \approx \varphi_T \ln\left(1 + \frac{1}{B_I}\right).$$

Токи базы транзисторов определяются выражениями:

$$I_{б1} = \frac{I_{ко}}{B_N} e^{\frac{U_{бэ}}{\varphi_T}},$$

$$I_{б2} = \frac{I_{ко}}{B_N} e^{\frac{U_{бэ}}{\varphi_T}} + \frac{I_{ко}}{B_I} e^{\frac{U_{бэ}}{\varphi_T}} \frac{1}{\left(1 + \frac{1}{B_I}\right)}, \qquad (8.6)$$

т.е.

$$\frac{I_{б2}}{I_{б1}} = 1 + \frac{B_N}{1 + B_I}.$$

При $B_N >> B_I$ бóльшая часть полного тока базы параллельно включенных транзисторов попадает в насыщенный транзистор и симметрия распределения базового тока нарушается. Выравнивание токов базы может быть достигнуто включением одинаковых сопротивлений последовательно с базами каждого транзистора (на рис. 8.5 показаны штриховыми линиями).

Последовательная ключевая цепочка. Отдельные ключи используются главным образом в аналоговых схемах. Для цифровых схем характерна совместная работа нескольких ключей — *ключевые цепочки*. В таких цепочках каждым ключом управляет предыдущий и сам он, в свою очередь, управляет последующим.

Рассмотрим последовательную цепочку ключей на рис. 8.6. Ключ $T4$, показанный штриховой линией, пока не будем учитывать. Если транзистор $T1$ открыт и насыщен, то потенциал $U_{к1}$ и равный ему потенциал $U_{б2}$ близки к нулю, а значит, транзистор $T2$ заперт. Тогда в базу транзистора $T3$ протекает ток от

источника $E_к$ через резистор $R_к$, и транзистор $T3$ открыт. Таким образом, *для последовательной цепочки характерно чередование открытых и закрытых ключей.*

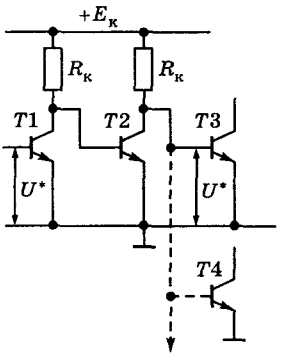


Рис. 8.6. Ключевая цепочка

Эквивалентные схемы для коллекторных цепей насыщенного и запертого транзисторов показаны на рис. 8.7. Эти схемы позволяют определить режим базовых цепей управляемых транзисторов. Сравнивая схему на рис. 8.7, *а* с общей схемой на рис. 8.2, запишем параметры базовой цепи запертого ключа:

$$E_б^- = U_{ост} \approx 0; \quad R_б \approx 0, \tag{8.7а}$$

где $E_б^-$ — «запирающее» напряжение. Сравнивая схему на рис. 8.7, *б* с общей схемой на рис. 8.2, запишем параметры базовой цепи открытого ключа:

$$E_б^+ \approx E_к; \quad R_б \approx R_к. \tag{8.7б}$$

Подставляя значения (8.7б) в выражение (8.3а), получаем величину отпирающего тока базы:

$$I_б^+ = (E_к - U^*)/R_к. \tag{8.8}$$

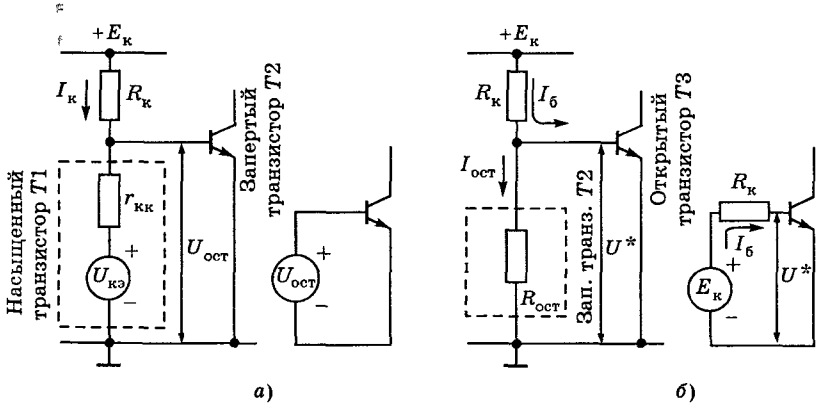


Рис. 8.7. Эквивалентные схемы ключа:
а — в запертом состоянии; *б* — в открытом (насыщенном) состоянии

Коллекторный ток насыщенного транзистора выражается той же формулой (8.3б), что и в «изолированном» ключе:

$$I_{кн} = (E_к - U_{ост})/R_к \approx E_к/R_к. \qquad (8.9)$$

Как видим, *коллекторный и базовый токи в последовательной цепочке почти одинаковы.*

Подставляя (8.8) и (8.9) в критерий насыщения (8.4), получаем ограничение на коэффициент усиления:

$$B > E_к/(E_к - U^*). \qquad (8.10)$$

Даже при минимальных напряжениях питания $E_к = 1–1{,}2$ В получаем $B > 2{,}4–3{,}3$, что легко выполнить в обычных рабочих режимах.

Степень насыщения ключа в последовательной цепочке согласно (8.5) записывается следующим образом:

$$S = B\,(E_к - U^*)/E_к. \qquad (8.11)$$

При обычных напряжениях питания (3–5 В) степень насыщения может достигать значений 50–100 и более.

В заключение подчеркнем, что напряжение на ключе, работающем в цепочке, меняется в более узких пределах, чем на отдельном, изолированном ключе. Действительно, у изолированного ключа в запертом состоянии имеем: $U_к = E_к$ (см. точку A на рис. 8.3, *б*), тогда как у запертого ключа, работающего в цепочке ($T2$ на рис. 8.6), $U_к = U^*$. Эта особенность обусловлена тем, что коллектор запертого транзистора соединен с базой следующего, открытого транзистора.

Нагрузочная способность ключа. Типичным для ключевых схем является сочетание последовательного и параллельного соединения ключей. А именно, в последовательной цепочке каждый транзистор может управлять не одним, а несколькими параллельно включенными ключами. Например, на рис. 8.6 штриховыми линиями показано, что транзистор $T2$ управляет не только ключом $T3$, но одновременно и ключом $T4$, а в общем случае еще рядом ключей.

Нагрузочной способностью называют количество параллельно включенных ключей, которыми способен управлять данный ключ. Обозначим это количество через n.

Предположим, что суммарный отпирающий ток (8.8) делится поровну между n базами; тогда в каждом из параллельно включенных ключей

$$I_{б1}^{+} = \frac{1}{n}\frac{E_к - U^*}{R_к}.$$

Ток $I_{б1}^{+}$ должен удовлетворять критерию насыщения (8.4), где коллекторный ток $I_{кн}$ по-прежнему определяется выражением (8.9). Из условия (8.4) нетрудно получить принципиальное ограничение на нагрузочную способность:

$$n < \frac{E_к - U^*}{E_к}\text{B}. \qquad (8.12а)$$

На самом деле ограничение должно быть более жестким, так как нужно обеспечить не просто насыщение (т. е. $S > 1$), а минимальную степень насыщения $S_{мин}$. Тогда, учитывая (8.5), нагрузочная способность будет ограничена более сильным неравенством:

$$n < \frac{\text{B}}{S_{мин}}\frac{E_к - U^*}{E_к}. \qquad (8.12б)$$

Например, если $E_к = 3$ В, B = 100 и $S_{мин} = 4$, то $n \le 18$. Задаваясь значением $S_{мин}$, необходимо, в частности, учитывать разброс базовых токов из-за неидентичности входных ВАХ.

8.3. Переходные процессы в простейшем биполярном ключе

Переходные процессы происходят под действием ступенчатых изменений входного сигнала. Временны́е диаграммы токов и напряжений в ключе показаны на рис. 8.8.

Как видим, выходные импульсы ($i_к$ и $u_к$) сдвинуты относительно входного ($e_б$), а их фронт и срез имеют конечную длительность. Соответственно различают *задержки* фронтов и срезов импульсов ($t_{зф}$ и $t_{зс}$) и длительности или *времена фронтов* и *срезов* ($t_ф$ и $t_с$). Приращения коллекторного напряжения связаны с приращениями коллекторного тока соотношением

$$u_к = - i_к R_к .$$

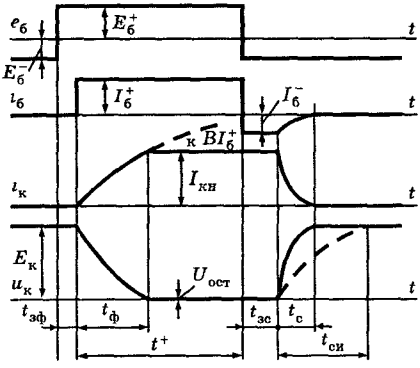


Рис. 8.8. Переходные процессы в ключе на биполярном транзисторе

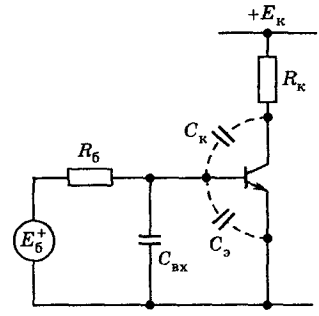


Рис. 8.9. Эквивалентная схема ключа на этапе задержки фронта

Задержка фронта. Этот первый этап переходного процесса обусловлен зарядом входной емкости запертого транзистора (рис. 8.9). Заряд емкости начинается после того, как управляющее напряжение скачком меняет свое значение от $E_б^-$ до $E_б^+$. Процесс заряда описывается уравнением

$$u_б(t) = E_б^+(1 - e^{-t/\tau_c}) - E_б^- e^{-t/\tau_c},$$

где $\tau_c = C_{вх}R_б$ — постоянная времени заряда.

Когда напряжение $U_б$, нарастая, становится равным напряжению U^*, отпирается эмиттерный переход транзистора и этап заряда оканчивается. Значит, время задержки $t_{зф}$ можно найти, полагая $u_б(t) = U^*$; оно выражается следующим образом:

$$t_{эф} = \tau_c \ln \frac{E_б^+ + E_б^-}{E_б^+ - U^*}. \qquad (8.13)$$

Например, если $E_б^- = 0$ [см. (8.7а)] и $E_б^+ = 3$ В, то $t_{зф} \approx 0{,}25\tau_c$.

Входную емкость обычно принимают равной сумме барьерных емкостей эмиттерного и коллекторного переходов:

$$C_{вх} = C_{эб} + C_{кб}. \qquad (8.14)$$

Полагая $C_{вх} = 2$ пФ и $R_б = 2$ кОм, получаем $\tau_c = 4$ нс. Тогда для использованных выше значений $E_б^-$ и $E_б^+$ получаем $t_{зф} \approx 1$нс.

Формирование фронта. Нарастание коллекторного тока и спад коллекторного напряжения на 2-м этапе переходного процесса происходят в условиях заданного тока базы $I_б^+$ [см. (8.3а)]. При анализе следует учитывать одновременные процессы заряда диффузионной емкости эмиттерного перехода и барьерной емкости перехода коллектор–база.

Предположим, что основное влияние на формирование фронта оказывает перезаряд диффузионной емкости базовым током $I_б^+$. Рассматривая переходный процесс установления тока через эмиттерный диод $I_{дэ}$ с учетом диффузионной емкости эмиттера $C_{э\,дифф} = I_{дэ}\tau_э/\varphi_T$ (см. модель биполярного транзистора, рис. 8.4), нетрудно получить

$$i_{дэ}(t) = I_б^+\left(1 - e^{-t/\tau_э}\right).$$

Следовательно,

$$i_к(t) = I_б^+ B_N\left(1 - e^{-t/\tau_э}\right). \qquad (8.15)$$

Асимптотическим значением тока при $t \to \infty$ является $i_к(\infty) = B_N I_б^+$ (см. рис. 8.8). Однако оно не может быть достигнуто, так как в момент $t_ф$ ток достигает значения $I_{кн}$, после чего транзистор переходит из активного режима в режим насыщения и, следовательно, теряет силу исходное соотношение $I_к = B_N I_б$. Время фронта $t_ф$ легко найти из (8.15), подставляя $i_к(t) = I_{кн}$; оно выражается следующим образом:

$$t_ф = \tau_э \ln \frac{1}{1 - \dfrac{I_{кн}}{B_N I_б^+}}. \qquad (8.16)$$

Типичными значениями времени жизни $\tau_э$ можно считать 100 нс (в отсутствие примеси золота) и 10 нс (при наличии примеси золота). Пусть, например, $B_N = 100$ и $\tau_э = 100$ нс. Тогда, если $I_б^+ = 0{,}1\, I_{кн}$, то $t_ф \approx 10$ нс, если же $I_б^+ = I_{кн}$, то $t_ф = 1$ нс.

Рассматривая перезаряд барьерной емкости коллекторного перехода, следует рассмотреть систему уравнений для токов базы и коллектора в процессе переключения:

$$I_б^+ = I_{дэ} - C_к \frac{dU_{кб}}{dt},$$

$$I_к = \frac{E_к - U_{кэ}}{R_к} = I_{дэ} \; B_N + C_к \frac{dU_{кб}}{dt}.$$

Учитывая, что $\frac{dU_{кб}}{dt} \approx \frac{dU_{кэ}}{dt}$, и исключая $I_{дэ}$, приходим к дифференциальному уравнению

$$(B_N + 1)C_к R_к \frac{dU_{кэ}}{dt} = (E_к - I_б^+ B_N R_к) - U_{кэ},$$

решением которого является:

$$U_{кэ}(t) = E_к - I_б^+ B_N R_к \left(1 - e^{-t/(B_N+1)C_к R_к}\right).$$

Значение $U_{кэ} \approx 0$ достигается в момент $t_ф$, который может быть определен из (8.17):

$$t_ф = (B_N + 1)C_к R_к \ln \frac{1}{1 - \frac{I_{кн}}{B_N I_б^+}}. \qquad (8.17)$$

Это выражение аналогично (8.16), но длительность процесса здесь определяется постоянной времени коллекторной емкости $(B_N + 1)C_к R_к$ [1]. При типичных значениях $C_к = 0{,}5$ пФ, $B_N = 100$ и $R_к = 2$ кОм она составляет около 100 нс. Тем самым эта постоянная времени всегда существенна, а в транзисторах, легированных золотом, является определяющей.

Эффективная постоянная времени на этапе формирования фронта определяется единой постоянной времени для процессов перезаряда диффузионной емкости эмиттерного перехода и барьерной емкости коллекторного перехода:

$$\tau_{эфф} = \tau_э + (B_N + 1)C_к R_к.$$

Из выражений (8.16) и (8.17) следует: *чем больше отпирающий ток, тем меньше время фронта.*

---

1 Множитель $(B_N + 1)$ при постоянной времени $C_к R_к$ связан с проявлением так называемого *эффекта Миллера* при наличии связи (в нашем случае емкостной) между входом и выходом ключа

Этап накопления заряда. После того как транзистор начал работать в режиме насыщения, заметных внешних изменений в схеме ключа не происходит. Однако продолжается накопление заряда в диффузионных емкостях эмиттерного и коллекторного перехода.

В начале этапа, когда транзистор находится на границе активного режима, заряд определяется током через эмиттерный диод: $Q = I_{дэ}\ \tau_э$. Подставляя $I_{дэ} = \frac{I_{кн}}{B_N}$ и учитывая (8.5), запишем *граничный заряд*

$$Q_{гр} = \frac{I_б^+}{S} \cdot \tau_э. \qquad (8.18)$$

В конце этапа *стационарный* заряд определяется распределением тока базы между коллекторным и эмиттерным переходами ($I_б^+ = I_{дэ} + I_{дк}$). Отношение этих токов в режиме насыщения имеет вид:

$$\frac{I_{дэ}}{I_{дк}} = \frac{B_I}{B_N} e^{\frac{U_{кэ}}{\varphi_т}},$$

или, с учетом (8.2) и (8.5),

$$\frac{I_{дэ}}{I_{дк}} = \frac{S\frac{B_I + 1}{B_N} + 1}{S + 1} \approx \frac{1}{S + 1}.$$

С учетом распределения токов стационарный заряд $Q^+ = I_{дэ} \cdot \tau_э + I_{дк} \cdot \tau_к$ запишем в следующем виде:

$$Q^+ = \frac{I_б^+}{S + 2}\tau_э + \frac{I_б^+}{S + 2}(S + 1)\tau_к. \qquad (8.19)$$

Здесь первый член — это заряд, накопленный в эмиттерном p–n-переходе, второй — в коллекторном. Коллекторный заряд превышает эмиттерный в $(S + 1)\tau_к/\tau_э$ раз. Учитывая, что обычно эффективное время жизни в коллекторном переходе существенно выше, чем в эмиттерном, приходим к выводу: *в насыщенном режиме биполярного транзистора заряд неосновных носителей накапливается в основном в коллекторном переходе.*

Поскольку токи базы и коллектора на данном этапе не меняются, заряд накапливается лишь благодаря термогенерации носителей, а значит, скорость накопления определяется временем жизни. Процесс накопления заряда — экспоненциальный и описывается выражением

$$Q(t) = Q(0)e^{-t/\bar{\tau}} + Q(\infty)(1 - e^{-t/\bar{\tau}}), \qquad (8.20)$$

где $Q(t)$ и $Q(\infty)$ — начальное и установившееся значения заряда, а $\bar{\tau} = Q^+/I_б^+$ — эффективная постоянная времени перезаряда, определяемая из (8.19). Учитывая, что при достаточно высоких степенях насыщения $Q_{гр} << Q^+$, пренебрежем в выражении (8.20) зарядом $Q(0)$; далее положим, что процесс накопления заканчивается на уровне 0,95 $Q(\infty)$. Тогда из выражения (8.20) получаем *время накопления*:

$$t_н = 3\bar{\tau}. \qquad (8.21)$$

Например, если $\bar{\tau} = 30$ нс, то $t_н \approx 90$ нс.

Для того чтобы накопление успело закончиться, длительность отпирающего импульса t^+ (рис. 8.8) должна превышать время накопления $t_н$. В противном случае в момент обратного переключения накопленный заряд будет меньше установившегося значения $Q(\infty)$. А именно, если, как и выше, пренебречь зарядом $Q(0)$, то в момент t^+ заряд будет иметь величину

$$Q(t^+) \approx I_б^+\bar{\tau}(1 - e^{-t^+/\bar{\tau}}). \qquad (8.22)$$

Эта величина в случае коротких импульсов может составлять малые доли $Q(\infty)$.

Задержка среза. *Заряд, накопленный в слоях и переходах транзистора, не может меняться мгновенно.* Соответственно, не могут меняться мгновенно и напряжения на эмиттерном и коллекторном переходах. Значит, в момент переключения управляющего напряжения $E_б$ от значения $E_б^+$ до значения $E_б^-$ (в частности, до 0) на обоих p–n-переходах сохранятся прямые смещения, близкие к U^*. При этом коллекторный ток не меняется и остается равным $I_{кн}$. Что касается базового тока, то он принимает значение

$$I_б^- = (E_б^- - U^*)/R_б. \qquad (8.23)$$

Если ключ работает в последовательной цепочке, то запирание происходит в результате насыщения предыдущего ключа. Тогда запирающий ток $I_б^-$ можно найти из эквивалентной схемы, показанной на рис. 8.10. В этой схеме насыщенный (предыдущий) транзистор представлен сопротивлением $r_{кк}$ и напряжением $U_{кэ}$ (ср. с рис. 8.7, *а*), а у запирающего транзистора выделено сопротивление базы $r_б$. Пренебрегая в первом приближении током $I_б^+$, напряжением $U_{кэ}$ и сопротивлением $r_{кк}$, получаем начальное значение запирающего тока:

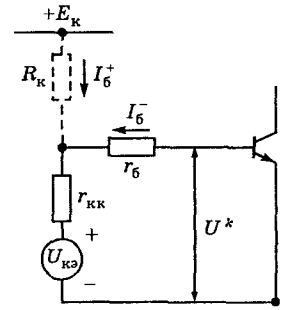


Рис. 8.10. Эквивалентная схема ключа на этапе рассасывания (при насыщении предыдущего ключа)

$$I_б^- = -U^*/r_б. \qquad (8.24)$$

Пусть, например, $r_б = 100$ Ом, тогда $I_б^- \approx -7$ мА; это значение может в несколько раз превышать значение отпирающего тока $I_б^+$, особенно при работе ключей в микрорежиме. Если не делать указанных выше пренебрежений, ток $I_б^-$ будет несколько меньше.

Скачок базового тока от значения $I_б^+$ до значения $I_б^-$ влечет за собой плавное уменьшение или, как говорят, *рассасывание* заряда от значения Q^+ до значения $Q_{гр}$ [см. (8.19) и (8.18)].

Рассасывание происходит в тех же условиях, что и накопление — при неизменных внешних токах. Поэтому скорость процесса рассасывания определяется той же постоянной времени $\bar{\tau}$, а уравнение рассасывания совпадает с уравнением накопления (8.20). Однако начальное и установившееся значения заряда будут иными.

А именно, начальный заряд на этапе рассасывания равен конечному заряду на этапе накопления. При достаточно длинном отпирающем импульсе этот заряд описывается формулой (8.19). Поэтому примем

$$Q(0) = I_б^+ \bar{\tau}. \qquad (8.25а)$$

Установившийся (асимптотический) заряд по-прежнему определяется током базы, т. е. в данном случае — запирающим током $I_б^-$:

$$Q(\infty) = I_б^- \bar{\tau}. \qquad (8.25б)$$

Поскольку $I_б^- < 0$, величина $Q(\infty)$ оказывается отрицательной. Это значит, что ее в данном случае следует рассматривать не как реальный заряд, а лишь как асимптотическое значение.

Окончание этапа рассасывания характеризуется тем, что концентрация избыточных носителей на коллекторной границе базы падает до нуля и на коллекторном переходе восстанавливается обратное напряжение. Только после этого может начаться уменьшение коллекторного тока и формирование среза импульса.

Время задержки среза $t_{эс}$ или время рассасывания $t_р$ есть тот интервал, в течение которого заряд уменьшается от начального значения $Q(0)$ до величины остаточного заряда $Q_{ост}$, соответствующего началу запирания коллекторного перехода.

Обычно остаточный заряд значительно меньше граничного заряда (8.18), а последний, в свою очередь, намного меньше заряда, накапливающегося в режиме насыщения. Поэтому в первом приближении можно пренебречь остаточным зарядом и определить время рассасывания из уравнения (8.20), положив $Q(t) = 0$. Тогда с учетом выражений (8.25) получаем:

$$t_р = \bar{\tau} \ln\left(1 + \frac{I_б^+}{|I_б^-|}\right), \qquad (8.26)$$

где $|I_б^-|$ — модуль запирающего тока. Например, если отношение токов лежит в пределах 0,2 — 5,0, то время рассасывания составляет (0,2–0,6) $\bar{\tau}$ или, при $\bar{\tau} = 30$ нс, от 6 до 20 нс.

Из формулы (8.26) видно, что *время рассасывания, т. е. задержка среза, тем меньше, чем меньше отпирающий ток* $I_б^+$ и *чем больше запирающий ток* $I_б^-$. Поскольку малая величина отпирающего тока приводит к увеличению длительности фронта [см. (8.17)], желательно в первую очередь увеличивать запирающий ток. Этот путь широко используется при проектировании ключей.

Формирование среза. По окончании этапа рассасывания начинается последний этап переходного процесса — запирание транзистора. Этот этап труднее других поддается количественному анализу, так как величина остаточного заряда сравнительно мала, а форма распределения носителей в базе — сложная.

Рассмотрим сначала простейший случай, когда запирающий ток настолько мал, что в начале данного этапа распределение носителей соответствует нормальному активному режиму (рис. 8.11, *а*). Тогда начальный заряд равен граничному: $Q(0) = Q_{гр}$; величина $Q_{гр}$ определяется формулой (8.18). Дальнейшее уменьшение заряда идет с такой же постоянной времени, как и уменьшение коллекторного тока, т. е. $\tau_э$ [см. (8.15)]. Асимптотическое значение заряда остается тем же, что и на этапе рассасывания: $Q(\infty) = Q^-$; величина Q^- определяется выражением (8.25б). Формирование среза заканчивается тогда, когда коллекторный ток, а вместе с ним и заряд делаются равными нулю.

Подставляя в (8.20) значение $Q(t) = 0$, заменяя в показателях экспонент $\bar{\tau}$ на $\tau_э$ и используя приведенные выше значения $Q(0)$ и $Q(\infty)$, получаем длительность среза в виде:

$$t_c = \tau_э \ln\left(1 + \frac{I_б^+}{|I_б^-|S}\right). \qquad (8.27)$$

Начальные условия, принятые при выводе (8.27), характерны для достаточно малых запирающих токов, много меньших тока насыщения. Такой случай редко имеет место на практике. Гораздо чаще запирающий ток сравним с током насыщения или превышает его.

На рис. 8.11, *б* показаны начальные распределения носителей при достаточно большом запирающем токе, всего лишь в несколько раз меньшем тока $I_{кн}$. Для сравнения штриховыми линиями повторены распределения, показанные на рис. 8.11, *а*. Как видим, площади под кривыми в данном случае заметно меньше, чем в активном режиме. Тем самым подтверждается, что

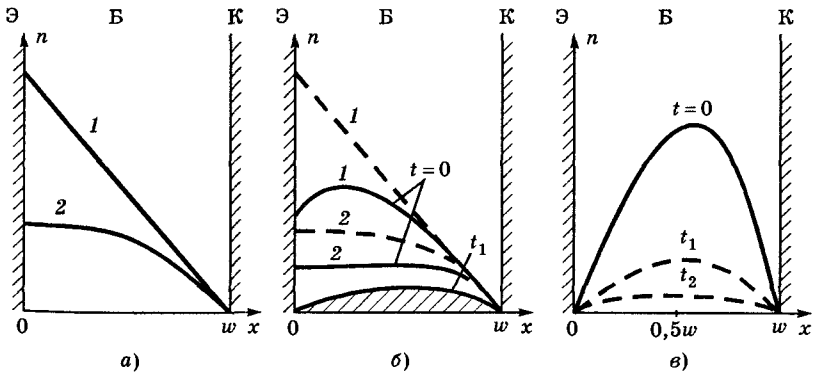


Рис. 8.11. Распределение концентрации носителей в базе на этапе запирания при малом (*а*), среднем (*б*) и большом (*в*) запирающих токах:
1 — бездрейфовый транзистор; *2* — дрейфовый транзистор

остаточный заряд в конце этапа рассасывания меньше граничного. Соответственно длительность среза будет меньше, чем при расчете по формуле (8.27). Однако главная особенность данного случая состоит не в этом чисто количественном отличии, а в изменении структуры среза.

А именно, в некоторый момент времени t_1 концентрация носителей на эмиттерной границе базы тоже делается равной нулю. После этого оба перехода — коллекторный и эмиттерный — работают при обратном смещении. Формально такой режим работы следует считать *отсечкой*. Однако, в отличие от «обычной» отсечки, когда в базе отсутствует избыточный заряд, а токи переходов ничтожно малы, в данном случае, несмотря на обратные смещения, в базе сохранился некоторый остаточный заряд (соответствующий заштрихованной площади), а токи имеют вполне конечные значения. Такой режим называют режимом *динамической отсечки*.

В режиме динамической отсечки все три тока транзистора спадают до нуля с постоянной времени отсечки $\tau_{отс}$, значение которой приведено ниже. Постоянная времени $\tau_{отс}$ значительно меньше, чем постоянная времени $\tau_э$, которая определяет процесс запирания до момента t_1. Поэтому конечный участок среза будет более крутым, чем начальный.

Как уже не раз отмечалось, запирающий ток $I_б^-$ обычно сравним с током $I_{кн}$ или превышает его. В случае таких больших токов распределение носителей в конце этапа рассасывания и начале этапа запирания получается таким, как показано на рис. 8.11, *в* (кривая $t = 0$). Это распределение уже соответствует режиму динамической отсечки, а значит, весь срез формируется с постоянной времени $\tau_{отс}$ и длительность среза $t_с$ оказывается весьма незначительной (рис. 8.8).

Постоянную времени отсечки можно грубо оценить, исходя из того, что избыточным носителям, расположенным в средней части базы, необходимо пройти расстояние $\frac{1}{2}w$, для того чтобы уйти из базы через тот или иной переход (рис. 8.11, *в*). Соответствующее время в 4 раза меньше времени пролета через всю базу. Поэтому запишем постоянную времени отсечки в виде:

$$\tau_{отс} = 0{,}25t_{пр} + C_к R_к. \qquad (8.28)$$

Первое слагаемое в большинстве случаев несущественно, и постоянная времени $\tau_{отс}$ определяется постоянной времени коллекторной цепи. Например, если $t_{пр} = 0{,}2$ нс, $C_к = 0{,}5$ пФ и $R_к = 1$ кОм, то $\tau_{отс} \approx C_к R_к = 0{,}5$ нс.

Время среза на уровне 0,1 от начального тока $I_{кн}$ выражается соотношением

$$t_с = 2{,}3\tau_{отс}. \qquad (8.29)$$

При $\tau_{отс} = 0{,}5$ нс получаем $t_с \approx 1{,}1$ нс.

Влияние емкости нагрузки. Если ключ работает на емкостную нагрузку (например, на вход следующего — запертого ключа, см. рис. 8.9), то длительности фронта и среза тока и напряжения могут существенно различаться. Например, во время среза ток по-прежнему уменьшается с малой постоянной времени $\tau_{отс}$, тогда как напряжение нарастает с гораздо большей постоянной времени $C_н R_к$, где $C_н$ — емкость нагрузки (см. штриховую кривую на рис. 8.8).

При условии $C_н >> C_к$, которое часто имеет место, срез импульса тока можно считать практически вертикальным, тогда как срез напряжения имеет конечную длительность, характеризуемую величиной (см. рис. 8.8)

$$t_{сн} \approx 2{,}3 C_н R_к. \qquad (8.30)$$

Аналогичное различие между длительностями фронтов тока и напряжения имеет место при отпирании ключа, если емкость нагрузки превышает коллекторную емкость. В этом случае фронт импульса тока тоже можно считать практически вертикальным, а длительность фронта импульса напряжения описывается выражением (8.30).

8.4. Ключ с барьером Шоттки

Одной из главных проблем при повышении быстродействия ключей является уменьшение задержки среза, т. е. времени рассасывания избыточного заряда Для этого, как видно из (8.26), нужно уменьшать отпирающий ток $I_б^+$, т. е. степень насыщения S. Однако при этом согласно (8.17) возрастает длительность фронта. Кроме того, в реальных условиях степень насыщения должна превышать минимальное значение $S_{мин}$ (см. разд. 8.2). Иначе малейшее уменьшение коэффициента B_N или тока $I_б^+$ пе-

реводит транзистор в активный режим, а это сопровождается увеличением остаточного напряжения на ключе.

Общепризнанным способом предотвратить насыщение транзистора и в то же время избежать отмеченных осложнений является использование в ключе нелинейной обратной связи. Этот способ состоит в том, что между коллектором и базой транзистора включается диод Шоттки (рис. 8.12), для которого характерно отсутствие накопления заряда неосновных носителей и малое напряжение отпирания.

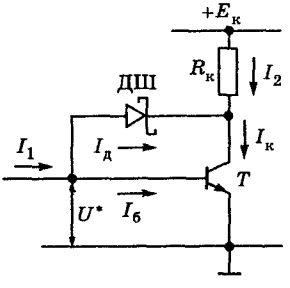


Рис. 8.12. Ключ с нелинейной обратной связью (транзистор с барьером Шоттки)

Когда транзистор заперт или работает в активном режиме, потенциал коллектора положителен относительно базы; следовательно, диод находится под обратным смещением и не влияет на работу ключа. Когда в процессе формирования фронта потенциал коллектора относительно базы проходит через нуль и делается отрицательным, диод отпирается и на нем устанавливается прямое напряжение $U_д^*$. Если это напряжение меньше 0,5 В (что характерно для диодов Шоттки), то коллекторный переход практически заперт, а значит, исключается режим двойной инжекции и накопление избыточного заряда, свойственные режиму насыщения. Соответственно при запирании ключа будут отсутствовать этап рассасывания избыточного заряда и задержка среза. Рассмотренная комбинация обычного транзистора и диода Шоттки получила название *транзистора с барьером Шоттки*. Легко заметить, что отсутствие насыщения в транзисторе с барьером Шоттки обусловлено меньшим прямым напряжением на диоде ($U_д^*$) по сравнению с прямым напряжением на кремниевом p–n-переходе (U^*).

Остаточное напряжение на ключе с барьером Шоттки несколько больше, чем на обычном ключе. А именно:

$$U_{ост} = U^* - U_д^* = 0{,}2\text{–}0{,}3 \text{ В}.$$

Однако этот недостаток окупается более высоким быстродействием, поскольку транзистор все время работает в активном режиме.

Следует подчеркнуть, что несмотря на отсутствие насыщения, ключ с барьером Шоттки мало чувствителен к изменениям коэффициента B_N и отпирающего тока, поскольку остаточное напряжение слабо зависит от этих величин, а значит, ток I_2 сохраняет свое значение, определяемое выражением (8.3б).

Задержка фронта $t_{зф}$ и длительность фронта $t_ф$ — оказываются такими же, как и в насыщенном ключе. Однако процесс запирания происходит иначе.

Когда управляющий ток I_1 принимает значение I_1^-, ток $I_д$ в первый момент не меняется, а ток $I_б$ изменяется на ту же величину, что и ток I_1:

$$\Delta I_б = \Delta I_1 = I_1^- - I_1^+ < 0.$$

Поскольку транзистор находится в активном режиме, коллекторный ток под действием скачка $\Delta I_б$ уменьшается с постоянной времени $\tau_э$. При этом приращения $\Delta I_к$ полностью идут через диод $Д$ и уменьшают ток $I_д$:

$$i_д(t) = I_д - B_N|\Delta I_б|(1 - e^{-t/\tau_э}), \qquad (8.31)$$

где $|\Delta I_б|$ — модуль приращения базового тока.

Внешний ток I_2 начинает уменьшаться только тогда, когда запирается диод. Полагая левую часть (8.31) равной нулю, легко найти время задержки среза.

Если, кроме того, положить $1/B_N << 1 - I_1^-/I_1^+$ (что обычно имеет место), то, разлагая логарифм в ряд, получим:

$$t_{зс} \approx \frac{\tau_э}{B_N} \frac{1}{1+|I_1^-|/I_1^+}. \qquad (8.32)$$

Пусть, например, $|I_1^-|/I_1^+ = 1$; тогда $t_{зс} = \tau_э/2B_N$. Эта величина у современных транзисторов редко превышает 0,5 нс.

8.5. Переключатель тока

Переключателем тока называют симметричную схему (рис. 8.13), в которой заданный ток I_0 протекает через ту или иную ее ветвь в зависимости от потенциала $U_б$ на одном из входов. Потенциал E на втором входе поддерживается неизменным.

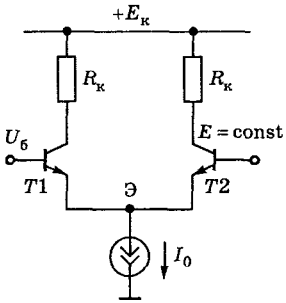


Рис. 8.13. Переключатель тока

Таким образом, первая особенность переключателя тока состоит в том, что управление осуществляется не током (как в простейшем ключе), а напряжением. Вторая особенность переключателя состоит в использовании ненасыщенного (активного) режима транзисторов в открытом состоянии, что обеспечивает повышенную скорость переключения в связи с отсутствием задержки на рассасывание.

Статический режим. Положим сначала $U_б = E$. При этом открыты оба транзистора и в каждом протекает ток $0{,}5\,I_0$. Потенциал эмиттеров меньше потенциала E на величину U^*: $U_э = E - U^*$.

Уменьшим потенциал $U_б$ на величину $\delta \geq 0{,}1$ В. Поскольку потенциал $U_э$ не меняется, напряжение $U_{бэ1}$ уменьшится тоже на величину δ. При этом ток транзистора $T1$ уменьшится в десятки раз. Значит, при входном сигнале $U_б \leq E - \delta$ транзистор $T1$ будет заперт, а через транзистор $T2$ будет протекать полный ток I_0. Величину

$$E_б^- = E - \delta \qquad (8.33а)$$

назовем *запирающим потенциалом.*

Если, наоборот, увеличить потенциал $U_б$ на величину δ, то вместе с ним на ту же величину возрастет потенциал эмиттеров и соответственно уменьшится напряжение $U_{бэ2}$; при этом резко уменьшится ток транзистора $T2$. Значит, при входном сигнале $U_б \geq E + \delta$ транзистор $T2$ можно считать запертым, а весь ток I_0 протекает через транзистор $T1$. Величину

$$E_б^+ = E + \delta \qquad (8.33б)$$

назовем *отпирающим потенциалом.*

Таким образом, *перепад потенциала* $\Delta U_б = \pm\delta$ *около средней величины* E *обеспечивает переключение тока* I_0 *из одного транзистора в другой.* Соотношения между отпирающим и запирающим потенциалами следующие:

$$E_б^+ - E_б^- = 2\delta; \qquad (8.34а)$$

$$\tfrac{1}{2}(E_б^+ + E_б^-) = E. \qquad (8.34б)$$

Остановимся на режиме открытого транзистора. Полагая, что открытый транзистор работает в активном режиме и учитывая, что ток эмиттера задан ($I_э = I_0$), получаем для токов коллектора и базы известные соотношения:

$$I_к = \alpha I_0; \tag{8.35а}$$

$$I_б = (1 - \alpha) I_0 = I_0/(B + 1). \tag{8.35б}$$

Потенциал коллектора определяется выражением

$$U_к = E_к - \alpha I_0 R_к. \tag{8.36}$$

Для того, чтобы обеспечить активный режим, т. е. избежать насыщения, нужно выполнить условие $U_{кб} \geq 0$ или $U_к > U_б$. Подставляя в последнее неравенство $U_к$ из (8.36) и $U_б = E_б^+$, получим

$$E_к - \alpha I_0 R_к \geq E_б^+. \tag{8.37а}$$

На практике данное нестрогое неравенство можно заменить равенством, так как небольшие отрицательные значения $U_{кб}$ не вызывают существенной инжекции в коллекторном переходе, а значит, и накопления избыточного заряда. Тогда условие ненасыщенного режима можно записать в следующем виде:

$$E_к - \alpha I_0 R_к = E_б^+ = E + \delta. \tag{8.37б}$$

Остаточное напряжение на транзисторе при условии граничного режима ($U_{кб} = 0$) равно напряжению на эмиттерном переходе:

$$U_{ост} = U^*. \tag{8.38}$$

Как видно из (8.36), потенциал коллектора открытого транзистора не зависит ни от входного сигнала $E_б^+$, ни от изменений коэффициента B (поскольку изменения B практически не меняют значения $\alpha \approx 1$). При стабилизированном напряжении питания $E_к$ единственным параметром, влияющим на величину $U_к$, является сопротивление $R_к$.

Последовательное соединение переключателей. Как и простейшие ключи, переключатели тока обычно работают совместно в последовательных цепочках. При этом управляющий сиг-

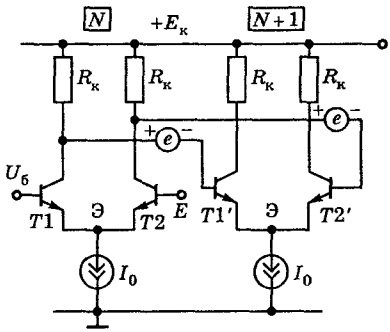


Рис. 8.14. Цепочка переключателей тока

нал поступает на данный переключатель с коллектора предыдущего, а выходной сигнал данного переключателя поступает на базу следующего (рис. 8.14). Легко убедиться, что *непосредственное соединение переключателей в таких цепочках невозможно*. Приходится соединять их друг с другом через дополнительные согласующие схемы, которые называют *схемами смещения уровня*.

Простейшим способом смещения уровня является включение э.д.с. e между смежными переключателями (рис. 8.14). В этом случае при запертом транзисторе $T1$ (в N-м переключателе) потенциал базы транзистора $T1'$ [в $(N+1)$-м переключателе] будет иметь величину

$$(U'_{б1})^{+} = E_к - e. \qquad (8.39а)$$

Эта величина должна превышать значение $E_б^{+}$ с тем, чтобы транзистор $T1'$ был открыт. Выполнение условия $(U'_б)^{+} > E_б^{+}$ не вызывает затруднений.

Если же транзистор $T1$ (в N-м переключателе) открыт, то потенциал базы транзистора $T1'$ составляет

$$(U'_{б1})^{-} = E_к - \alpha I_0 R_к - e. \qquad (8.39б)$$

Эта величина должна быть меньше $E_б^{-}$ с тем, чтобы транзистор $T1'$ был заперт. Такое условие накладывает определенное ограничение на э.д.с. e. Действительно, подставим в (8.39б) значение $E_к - \alpha I_0 R_к$ из (8.37б); тогда

$$(U'_{б1})^{-} = E + \delta - e.$$

Подставляя в неравенство $(U'_{б1})^{-} \le E_б^{-}$ полученную величину $(U'_{б1})^{-}$ и значение $E_б^{-}$ из (8.33а), нетрудно получить указанное ограничение:

$$e \ge 2\delta. \qquad (8.40)$$

Источник тока I_0 можно осуществлять разными способами. Наиболее простым и исторически первым является использование резистора R_0 (рис. 8.15). Если открыт транзистор $T2$, то ток I_{02} определяется соотношением

$$I_{02} = U_э/R_0 = (E - U^*)/R_0.$$

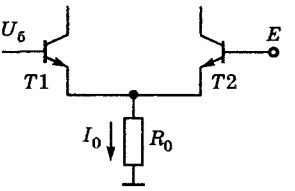


Рис. 8.15. Переключатель тока с резистором в качестве источника тока I_0

Если открыт транзистор $T1$, то ток I_{01} имеет несколько большее значение:

$$I_{01} = U_э/R_0 = (E_б^+ - U^*)/R_0.$$

Подставляя (8.37б), нетрудно получить:

$$I_{01} = I_{02} + (\delta/R_0).$$

Как видим, ток I_0 при переключении не остается постоянным, а меняется на величину δ/R_0. Для того, чтобы это непостоянство было пренебрежимо малым, должно выполняться условие

$$\frac{\delta}{R_0} << \frac{E - U^*}{R_0} \text{ или } E - U^* >> \delta.$$

Например, если δ = 0,1 В, то напряжение E должно превышать 1,7 В.

Переходные процессы. Будем считать, что управляющие сигналы поступают от источника э.д.с. с нулевым внутренним сопротивлением. В реальных схемах такое предположение обычно оправдано.

Пусть в исходном состоянии схемы (рис. 8.13) на входе действует потенциал запирания $E_б^-$ и, значит, транзистор $T1$ заперт. При поступлении сигнала, равного потенциалу отпирания $E_б^-$, первым этапом переходного процесса будет, как и в простейшем ключе, заряд входной емкости. Анализ, аналогичный проведенному в разделе 8.4, приводит к выражению:

$$t_{эф} = \tau_c \ln 2 \approx 0{,}7\tau_c, \qquad (8.41)$$

где постоянная времени $\tau_c = r_б C_{вх}$. Если принять $r_б$ = 100 Ом и $C_{вх}$ = 2 пФ, то $t_{эф} \approx 0{,}15$ нс. С учетом конечного внутреннего сопротивления источника сигнала задержка будет соответственно больше. Формула (8.41) остается в силе для любых сигналов, симметричных относительно потенциала E.

После отпирания транзистора $T1$ через его эмиттерный переход протекает неизменный ток I_0, а потенциал базы сохраняет неизменное значение $E_б^+$. Эти условия означают, что *по существу транзистор включен по схеме с общей базой* (ОБ), хотя внешне кажется, что он включен по схеме ОЭ.

В первый момент, когда коллекторный ток еще равен нулю, весь эмиттерный ток I_0 протекает через базу: $I_б(0) = I_0$. Значение $I_б(0)$ может намного превышать установившееся значение (8.35б). По мере нарастания коллекторного тока ток базы уменьшается.

Как известно, ток коллектора в схеме ОБ меняется с эквивалентной постоянной времени

$$\tau_{экв} = \frac{\tau_э}{B_N + 1} + C_к R_к. \qquad (8.42)$$

Поскольку в переключателях тока транзисторы работают в активном режиме, фронт и срез оказываются экспоненциальными. Их длительности на уровне (0,1–0,9) $\Delta I_к$ (где $\Delta I_к = \alpha I_0$) одинаковы и составляют:

$$t_ф = t_c = 2{,}2\tau_{экв}. \qquad (8.43)$$

Из выражения (8.42) очевидно, что второе слагаемое ($C_к R_к$) желательно делать близким к первому или меньше его. Таким образом, значение сопротивления $R_к$ целесообразно выбирать из условия

$$R_к \le \frac{\tau_э}{B_N + 1}\frac{1}{C_к}. \qquad (8.44)$$

Например, если $B_N = 100$, $\tau_э = 20$ нс и $C_к = 0{,}5$ пФ, то $R_к \le 0{,}4$ кОм. Накопление и рассасывание носителей, а значит, и задержка среза в ненасыщенном ключе отсутствуют.

8.6. МДП-транзисторные ключи

Ключи на МДП-транзисторах, как и на биполярных, в статическом режиме характеризуются *остаточным током* (в запертом состоянии) и *остаточным напряжением* (в открытом состоянии). Известны три разновидности МДП-транзисторных ключей: с *резисторной* нагрузкой, с *динамической* (транзистор-

ной) нагрузкой и *комплементарные* (КМДП) ключи. Последние выполнены на комплементарных транзисторах, т. е. на транзисторах с каналами противоположного типа проводимости. Рассмотрим поочередно статические параметры перечисленных трех разновидностей.

Ключ с резисторной нагрузкой. Схема такого ключа, выполненного на n-канальном транзисторе, показана на рис. 8.16, $а$[1]. Для запирания ключа на затвор транзистора подается напряжение $E_з^- < U_0$, где U_0 — пороговое напряжение.

В запертом транзисторе остаточный ток есть обратный ток стокового p–n-перехода, поскольку этот переход работает при обратном смещении, близком к $E_с$. Следовательно, ток $I_{ост}$ составляет не более 10^{-9}–10^{-10} А (при условии, что поверхность кристалла хорошо обработана и отсутствуют приповерхностные каналы). На вольт-амперной характеристике запертому состоянию ключа соответствует точка A (рис. 8.16, б). При указанных

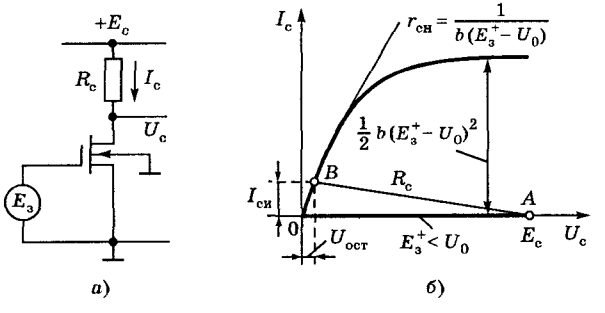


Рис. 8.16. МДП-транзисторный ключ с резисторной нагрузкой: $а$ — схема; $б$ — расположение рабочих точек на выходной характеристике

значениях остаточного тока падением напряжения $I_с R_с$ можно пренебречь и считать, что максимальное напряжение на запертом ключе $U_{макс} = E_с$.

Для отпирания ключа на затвор подается напряжение $E_з^+ > U_0$. Это напряжение должно быть достаточно большим, чтобы рабочая точка B (рис. 8 16, $б$) соответствовала как можно меньшему остаточному напряжению. Тогда рабочий ток откры-

1 В случае p-канального транзистора все напряжения в последующем анализе следует считать модулями отрицательных величин.

того ключа (ток насыщения) определяется, как у биполярного ключа, внешними элементами схемы:

$$I_{\text{сн}} = (E_{\text{с}} - U_{\text{ост}})/R_{\text{с}} \approx E_{\text{с}}R_{\text{с}} . \qquad (8.45)$$

Рабочая точка B в открытом состоянии ключа лежит на начальном, квазилинейном участке характеристики МДП-транзистора. Поэтому остаточное напряжение можно записать, умножая ток насыщения (8.45) на сопротивление канала $r_{\text{сн}}$. Полагая $U_{\text{зи}} = E_{\text{з}}^{+}$, получаем

$$U_{\text{ост}} = \frac{E_{\text{с}}}{R_{\text{с}}} \cdot \frac{1}{b(E_{\text{з}}^{+} - U_0)}, \qquad (8.46)$$

где b — удельная крутизна.

При совместной работе ключей в последовательной цепочке отпирающий сигнал $E_{\text{з}}^{+}$ поступает от предыдущего (запертого) ключа; в этом случае $E_{\text{з}}^{+} = E_{\text{с}}$.

Если принять $b = 0{,}1$ мА/В^2, $U_0 = 2{,}5$ В, $R_{\text{с}} = 50$ кОм и $E_{\text{з}}^{+} = E_{\text{с}} = 7{,}5$ В, то $U_{\text{ост}} = 300$ мВ. Такое значение $U_{\text{ост}}$ сравнительно велико, причем пути его уменьшения в данной схеме ограничены, поскольку и увеличение $R_{\text{с}}$, и увеличение b в конечном счете означают увеличение площади, занимаемой схемой, а это в полупроводниковых ИС нежелательно.

Однако следует подчеркнуть, что принципиальных ограничений на величину $U_{\text{ост}}$ в МДП-транзисторных ключах нет; *остаточное напряжение можно сделать сколь угодно малым*, увеличивая сопротивление $R_{\text{с}}$ и напряжение $E_{\text{з}}^{+}$. Это — одно из важнейших преимуществ МДП-транзисторных ключей перед биполярными, у которых величина $U_{\text{ост}}$ принципиально ограничена напряжением $U_{\text{кэ}}$ [см. (8.2)].

Ключ с динамической нагрузкой. Схема такого ключа, выполненного на однотипных транзисторах, показана на рис. 8.17, *а*. Роль динамической нагрузки выполняет транзистор $T2$, у которого затвор соединен со стоком и который, тем самым, является двухполюсником — резистором. В схеме с динамической нагрузкой транзистор $T2$ называют *нагрузочным*, а $T1$ — *активным*.

Вольт-амперную характеристику резистора $T2$ можно получить из следующих соображений. Поскольку при соединении затвора со стоком получается $U_{\text{зи}2} = U_{\text{си}2}$, то, очевидно, спра-

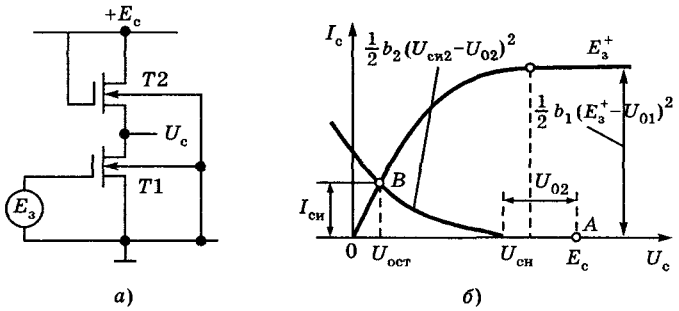


Рис. 8.17. МДП-транзисторный ключ с динамической нагрузкой: *а* — схема, *б* — расположение рабочих точек на выходной характеристике

ведливо неравенство $U_{зи2} - U_0 < U_{си2}$. Это неравенство означает, что транзистор $T2$ работает на пологом участке характеристики. Для этого участка действительна формула $I_с = \frac{1}{2} b(U_{зи} - U_0)^2$. Подставляя в нее $U_{зи} = U_{си2}$, запишем ВАХ резистора в виде

$$I_{с2} = \tfrac{1}{2} b_2 (U_{си2} - U_{02})^2. \qquad (8.47)$$

Как видим, эта ВАХ — параболическая, т. е. нелинейная.

В запертом состоянии ключа, когда на затвор подано напряжение $E_з^- < U_0$, остаточный ток имеет примерно то же значение, что и в резисторном ключе (10^{-9}–10^{-10} А и менее), а максимальное выходное напряжение близко к напряжению питания: $U_{макс} \approx E_с$ (см. точку A на рис. 8.17, *б*). Точное положение точки A определяется пересечением обратных характеристик стоковых p–n-переходов активного и нагрузочного транзисторов.

В открытом состоянии ключа, когда на затвор подано напряжение $E_з^+ > U_0$, рабочая точка B лежит на квазилинейном участке характеристики активного транзистора $T1$. Остаточное напряжение в этой точке, как обычно, мало. Поэтому питающее напряжение можно считать полностью приложенным к нагрузочному транзистору $T2$. Тогда ток насыщения определяется из формулы (8.47), если положить $U_{си2} = E_с$:

$$I_{си} = \tfrac{1}{2} b_2 (U_с - U_{02})^2. \qquad (8.48)$$

Умножая ток $I_{сн}$ на сопротивление канала и полагая $U_{зи} = E_з^+$, получаем остаточное напряжение в виде[1]:

$$U_{ост} = \frac{b_2}{2b_1} \frac{(E_с - U_{02})^2}{E_э^+ - U_{01}}. \qquad (8.49)$$

Поскольку на практике всегда выполняется условие $E_з^+ \le E_с$, нетрудно сделать следующий важный вывод: для того, чтобы остаточное напряжение было мало, *в ключе с динамической нагрузкой должно выполняться соотношение* $b_2 << b_1$, *т. е. транзисторы должны быть существенно различными.*

Напомним, что удельная крутизна b определяется в первую очередь геометрией транзистора, а именно, отношением ширины к длине канала Z/L. Значит, у активного транзистора отношение Z/L должно быть как можно большим, а у нагрузочного как можно меньшим. В обоих случаях ограничения накладываются конструктивно-технологическими факторами. Если обеспечить отношение $b_1/b_2 = 50–100$ (что вполне реально), то остаточное напряжение может лежать в пределах 50–100 мВ.

Комплементарный ключ. Схема такого ключа показана на рис. 8.18. Пусть в исходном состоянии управляющее напряжение равно $E_з = 0$. Тогда

$$U_{зи1} = 0; \; U_{зи2} = -E_с.$$

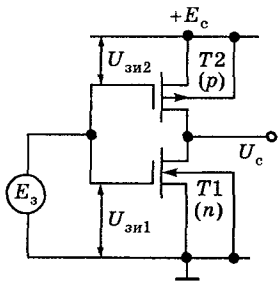


Рис. 8.18. Комплементарный МДП-транзисторный ключ

Значит, n-канальный транзистор $T1$ заперт, а p-канальный транзистор $T2$ открыт (считаем, что $E_с > |U_{02}|$).

Ток в общей цепи определяется запертым транзистором $T1$ и составляет величину $I_{ост}$. Открытый транзистор $T2$, как и в предыдущих схемах, работает на квазилинейном участке характеристики. Умножая остаточный ток первого

1 Пороговые напряжения транзисторов для общности приняты разными. В интегральных схемах это различие неизбежно из-за различия напряжений между истоками и общей подложкой.

транзистора на сопротивление канала второго и полагая $U_{зи} = E_c$, получаем напряжение на открытом транзисторе $T2$:

$$|U_{си2}| = \frac{I_{ост\,1}}{b_2(E_c - |U_{02}|)}. \qquad (8.50)$$

Если принять $I_{ост1} = 10^{-9}$ А, $b_2 = 0{,}1$ мА/В^2 и $E_c - |U_{02}| = 5$ В, то $|U_{си2}| = 2$ мкВ.

Пусть теперь управляющее напряжение принимает положительное значение $E_з^+ = E_c$. Тогда

$$U_{зи1} = E_c > U_{01};\quad U_{зи2} = 0.$$

Значит, теперь n-канальный транзистор $T1$ открыт, а p-канальный транзистор $T2$ заперт. При этом ток в общей цепи по-прежнему остается на уровне $I_{ост}$, хотя транзисторы «поменялись местами».

Таким образом, важнейшей особенностью комплементарных ключей является то, что они *практически не потребляют мощности в обоих состояниях*. Соответственно, эти два состояния можно называть «закрытым» и «открытым» только условно — по отношению к одному из транзисторов (например, n-канальному).

Однако устойчивые состояния различаются весьма четко по уровню выходного напряжения. Так, выше было показано, что при низком значении $E_з^+$, когда транзистор $T1$ заперт, напряжение $U_{си2}$ на открытом транзисторе $T2$ ничтожно мало, а значит, выходное напряжение равно напряжению питания:

$$U_{макс} = E_c. \qquad (8.51а)$$

При высоком значении $E_з^+$, когда открыт транзистор $T1$, на нем падает такое же ничтожное напряжение. Величину $U_{си1}$ можно найти из выражения (8.50), заменив индексы в правой части. Это и будет остаточное напряжение на ключе:

$$U_{ост\,1} = I_{ост\,2} / [b_1(E_c - U_{01})]. \qquad (8.51б)$$

Остаточное напряжение может иметь такие же крайне малые значения, как и в приведенном выше примере: до единиц микровольт и менее. Малые остаточные напряжения являются еще одним важным преимуществом комплементарных ключей.

Если напряжение питания E_c превышает сумму пороговых напряжений обоих транзисторов, то имеется интервал управляющих сигналов

$$U_{01} < E_з < E_c - |U_{02}|,$$

в котором открыты оба транзистора. Тогда общий ток в цепи будет иметь конечное (иногда достаточно большое) значение, которое можно вычислить по известным формулам. Однако для КМДП-ключей типичны низкие напряжения питания, близкие к сумме пороговых напряжений, так что заметного возрастания тока в период переключения обычно не происходит.

Переходные процессы. Инерционность МДП-транзисторных ключей обусловлена главным образом перезарядом емкостей, входящих в состав комплексной нагрузки. Инерционность канала, характеризуемую постоянной времени τ_s, при необходимости можно учесть, складывая τ_s с постоянной времени перезаряда емкостей.

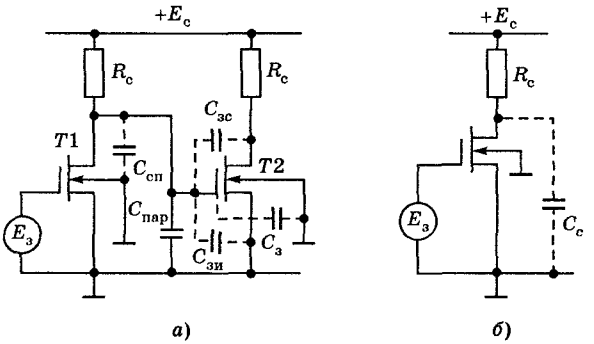


Рис. 8.19. Паразитные емкости МДП-транзисторного ключа:
а — составляющие емкости; *б* — результирующая емкость

На рис. 8.19, *а* показан ключ на транзисторе $T1$ (с резисторной нагрузкой), работающий в последовательной цепочке на второй аналогичный ключ. На рис. 8.19, *б* показана эквивалентная схема ключа, в которой все частные емкости заменены одной суммарной емкостью C_c:

$$C_c = C_з + C_{сп} + C_{пар} + C_{зи} + KC_{зс}. \qquad (8.52)$$

Типичные значения суммарной емкости составляют C_c = 1–3 пФ. Она включает следующие компоненты: $C_з$ — емкость затвор–канал; в отличие от других емкостей она являет-

ся имманентной, т. е. *органически свойственной МДП-транзистору* по самому принципу его действия (удельная емкость затвор-канал определяет величину удельной крутизны); $C_{сп}$ — емкость сток-подложка (барьерная емкость стокового p–n-перехода); $C_{пар}$ — паразитная емкость монтажных соединений относительно подложки (в интегральных схемах это емкость металлической или поликремниевой разводки); $C_{зи}$, $C_{зс}$ — емкости затвора относительно областей истока и стока.

Происхождение коэффициента K связано с так называемым *эффектом Миллера* (см. разд. 9.5). Коэффициент K может иметь значения от нескольких единиц до 10–20 и более. При этом роль емкости $C_{зс}$ часто оказывается доминирующей.

Пусть в исходном состоянии транзистор открыт и на нем падает небольшое остаточное напряжение. При поступлении запирающего сигнала $E_з^-$ ток в транзисторе уменьшается до нуля с весьма малой постоянной времени τ_s — практически мгновенно. После запирания транзистора (т. е. запирания ключа) емкость $C_с$ заряжается от источника питания $E_с$ через резистор $R_с$ с постоянной времени $\tau_с = C_с R_с$ (рис. 8.20, *а*).

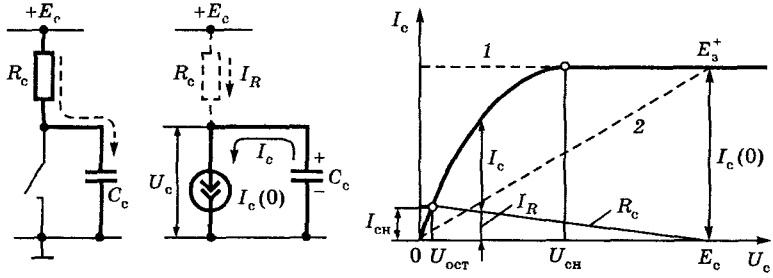


Рис. 8.20. К расчету переходных процессов в МДП-транзисторном ключе: *а* — эквивалентная схема после запирания ключа; *б* — эквивалентная схема после отпирания ключа; *в* — ход процесса отпирания на выходных характеристиках

Процесс заряда описывается простейшей экспоненциальной функцией

$$u_с(t) = E_с(1 - e^{-t/\tau_с}).$$

Время заряда, т. е. длительность фронта напряжения на уровне $0{,}9E_с$, составляет

$$t_ф = 2{,}3\, R_с C_с. \qquad (8.53а)$$

Например, если $R_c = 50$ кОм и $C_c = 3$ пФ, то $\tau_c = 150$ нс и $t_ф \approx 350$ нс.

Если в формуле (8.53а) заменить сопротивление R_c отношением $E_c/I_{сн}$ [см. (8.45)], то длительность фронта запишется в следующем более общем виде:

$$t_ф = 2{,}3\ (E_c C_c / I_{сн}). \qquad (8.53б)$$

Из такой формы записи ясно, что длительность фронта предопределяется желательной величиной рабочего тока.

Отпирание ключа и формирование среза импульса напряжения протекают несколько сложнее. После подачи отпирающего сигнала $E_з^+$ ток I_c практически мгновенно (с постоянной времени τ_s) достигает значения

$$I_c(0) = \tfrac{1}{2} b (E_з^+ - U_0)^2 \qquad (8.54)$$

(рис. 8.20, *б*). Этим током начинает разряжаться емкость C_c. По мере разряда емкости напряжение на стоке U_c уменьшается. До тех пор, пока оно остается больше напряжения насыщения $U_{сн}$, равного $E_з^+ - U_0$, транзистор работает на пологом участке характеристики и ток сохраняет значение $I_c(0)$ (рис. 8.20, *в*).

Когда напряжение U_c, уменьшаясь, делается меньше напряжения насыщения, ток I_c начинает падать, стремясь в пределе к значению $I_c(\infty) = I_{сн}$. Соответственно на этом этапе переходного процесса нужно было бы учесть нелинейную зависимость $I_c(U_c)$. Однако, чтобы избежать математических трудностей, воспользуемся двумя простейшими аппроксимациями, одна из которых дает заведомо заниженное, а другая — заведомо завышенное значение длительности среза.

В обоих случаях пренебрежем током I_R через нагрузочный резистор. Такое пренебрежение вполне оправдано, так как почти на всем протяжении переходного процесса ток I_R незначителен по сравнению с емкостным током I_C (рис. 8.20, *в*).

В качестве 1-й аппроксимации примем $I_c = I_c(0) = \text{const}$ (емкость разряжается неизменным током, см. штриховую линию *1* на рис. 8.20, *в*). Тогда, поделив начальный заряд $Q = E_c C_c$ на разрядный ток $I_c(0)$, получаем:

$$t_c = E_c C_c / I_c(0).$$

В качестве 2-й аппроксимации примем $I_c = U_c/R_{ср}$, где $R_{ср} = E_c/I_c(0)$ — среднее сопротивление в период разряда (см. штриховую линию *2* на рис. 8.20, *в*). В этом случае разряд носит обычный экспоненциальный характер

$$u_c(t) = E_c e^{-t/\tau},$$

где $\tau = C_c R_{ср}$. Считая, что разряд заканчивается на уровне 0,1 E_c, получаем:

$$t_c = 2{,}3\tau = 2{,}3\ [E_c C_c / I_c(0)].$$

Для расчетов можно принять промежуточное из двух полученных значений:

$$t_c = 1{,}5\ [E_c C_c / I_c(0)], \qquad (8.55)$$

где ток $I_c(0)$ определяется выражением (8.54).

Например, если $E_з^+ = E_c = 7{,}5$ В; $U_0 = 2{,}5$ В; $b = 0{,}1$ мА/В^2 и $C_c = 3$ пФ, то $I_c(0) = 1{,}25$ мА и $t_c \approx 25$ нс.

Как видим, *срез положительного импульса значительно короче фронта*. В общем случае такой вывод вытекает из структуры выражений (8.53б) и (8.55), которые в принципе различаются только значениями токов. Из рис. 8.20, *в* ясно, что $I_c(0) >> I_{сн}$. Отсюда неизбежно следует, что $t_c << t_ф$.

Таким образом, быстродействие данного типа ключей определяется длительностью фронта. Для уменьшения времени $t_ф$ необходимо уменьшать сопротивление R_c , а это приводит к росту остаточного напряжения на ключе [см. (8.46)]. Следовательно, возможности повышения быстродействия ограничены.

Общий вид переходных процессов в рассмотренной схеме показан на рис. 8.21.

В ключе с динамической нагрузкой (рис. 8.17, *а*) формирование среза происходит так же, как в ключе с резисторной нагрузкой, а время t_c определяется формулой (8.55). Такое совпадение объясняется тем, что при выводе формулы (8.55) мы пренебрегли током нагрузки I_R и

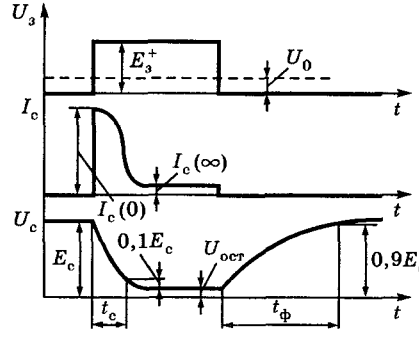


Рис. 8.21. Переходные процессы в МДП транзисторном ключе

тем самым исключили специфику нагрузки. Ток $I_c(0)$, входящий в формулу (8.55), является в данном случае начальным током активного транзистора: по аналогии с (8.54) запишем:

$$I_c(0) = \tfrac{1}{2} b_1 (E_э^+ - U_{01}). \qquad (8.56)$$

Формирование фронта происходит в период заряда емкости C_c через нелинейную динамическую нагрузку. Учитывая параболический характер ВАХ (8.47), можно заранее ожидать, что заряд емкости будет происходить медленнее, чем при резисторной нагрузке, а время $t_ф$ будет больше. Чтобы избежать математических трудностей, заменим параболическую ВАХ линейной с сопротивлением $E_c/(\frac{1}{2} I_{сн})$. Тогда вместо формулы (8.53б) получим

$$t_ф = 2{,}3 E_c C_c/(\tfrac{1}{2} I_{сн}), \qquad (8.57)$$

где ток насыщения $I_{сн}$ выражается формулой (8.48).

Заметим, что в данном ключе в состав нагрузочной емкости (8.52) следует дополнительно включить емкость $C_{зн2}$, а в интегральных схемах (учитывая общность подложки у обоих транзисторов) — еще и емкость $C_{ип2}$.

Поскольку по сравнению с резисторным ключом время $t_ф$ возросло, а время t_c осталось тем же, можно сделать вывод, что *в ключах с динамической нагрузкои*, как и в резисторных ключах, *быстродействие определяется длительностью фронта.*

Используя выражения (8.47) и (8.56), легко показать, что отношение $t_ф/t_c$ определяется в первую очередь отношением удельных крутизн b_1/b_2. Попытка уменьшить величину b_1/b_2 (с целью выравнивания фронта и среза) приводит к увеличению остаточного напряжения [см. (8.49)]. Поэтому повышение быстродействия ключа требует увеличения удельной крутизны обоих транзисторов, а это, как известно, сопряжено с увеличением их размеров, т. е. площади. В интегральных схемах такой путь, естественно, ограничен.

В комплементарном ключе переходные процессы характерны тем, что заряд и разряд нагрузочной емкости $C_н$ происходят примерно в одинаковых условиях. Это объясняется симметрией схемы по отношению к запирающему и отпирающему сигналам.

Заряд емкости происходит через открытый транзистор $T2$ при запертом $T1$ (рис. 8.18), а разряд — через открытый транзистор $T1$ при запертом $T2$. В обоих случаях транзистор, открывшийся после очередного переключения, сначала работает в режиме насыщения со сравнительно большим током $I_c(0)$, а затем, по мере заряда или разряда емкости, напряжение на стоке падает ниже значения $U_{сн}$ и ток начинает уменьшаться. Следовательно, механизм обоих процессов (заряда и разряда) тот же, который был рассмотрен при анализе разряда в ключе с резисторной нагрузкой (рис. 8.20, *б* и *в*).

Соответственно, длительности фронта и среза определяются однотипными выражениями, аналогичными (8.55):

$$t_ф = 1{,}5\frac{E_c C_c}{I_{c2}(0)} = \frac{3E_c C_c}{b_2(E_c - |U_{02}|)^2}; \qquad (8.58а)$$

$$t_c = 1{,}5\frac{E_c C_c}{I_{c1}(0)} = \frac{3E_c C_c}{b_1(E_c - U_{01})^2}. \qquad (8.58б)$$

В формулах (8.58) индексы *1* и *2* призваны подчеркнуть различие параметров n- и p-канального транзисторов. Однако это различие второстепенное. Практически длительности фронта и среза оказываются одинаковыми.

Если, как и выше, положить U_0 = 2,5 В, b = 0,1 мА/В^2, C_c = 3 пФ и E_c =7,5 В, то $t_ф = t_c \approx 25$ нс.

Сравнивая полученные значения с предыдущими, видим, что быстродействие комплементарного ключа оказывается почти на порядок выше, чем у двух других типов ключей. Это преимущество сохраняется и при уменьшении напряжения питания.

Другой очень важной чертой комплементарного ключа является очень малое потребление энергии от источника питания в статическом режиме. Заметная энергия, обусловленная необходимостью перезаряда емкостей, потребляется в процессе переключения. Также при переключении из-за конечной длительности фронта входного сигнала на короткое время оба транзистора ключа оказываются открытыми, что приводит к короткому импульсу тока потребления от источника питания. Использование комплементарных ключей при очень пологом фронте входного сигнала может приводить к тепловому разрушению транзисторов при протекании большого сквозного тока.

Для всех трех типов ключей *главным путем повышения быстродействия является уменьшение суммарной емкости* C_c. При заданной емкости *быстродействие повышается с увеличением токов*, в частности, — с увеличением питающего напряжения.

8.7. Помехоустойчивость ключей

Помимо полезных (управляющих) сигналов, на ключи и ключевые схемы всегда воздействуют паразитные сигналы, обусловленные внешними электромагнитными помехами («наводками») или же внутренними процессами, например, связями через общий источник питания. Поэтому рабочие сигналы должны превышать уровень помех, а к малым паразитным сигналам ключ должен быть по возможности нечувствителен, т. е. не должен реагировать на них так, как он реагирует на полезный сигнал.

Индифферентность ключа (или другой схемы) к паразитным сигналам называют *помехозащищенностью* или *помехоустойчивостью*.

Помехоустойчивость принято измерять абсолютной величиной сигнала (обычно в вольтах), который еще не вызывает ложного отпирания или запирания ключа. Помехоустойчивость по отношению к сигналам положительной и отрицательной полярности может быть существенно разной. Анализ помехоустойчивости предполагает, что ключ работает в последовательной цепочке.

Последовательная цепочка ключей должна работать как одно целое; изменение состояния 1-го ключа должно сопровождаться изменением состояния всех остальных ключей, вплоть до последнего. Оказывается, для этого входной сигнал должен превышать так называемый *порог чувствительности*; в противном случае сигналы в цепочке «затухают» и ключи, удаленные от первого, не меняют своего состояния. Напротив, сигнал помехи должен быть меньше порога чувствительности.

Для того, чтобы оценить порог чувствительности, рассмотрим предварительно общую методику определения рабочих точек в последовательной цепочке ключей. Задача ставится таким образом: известна рабочая точка n-го ключа; необходимо с помощью передаточной характеристики найти рабочую точку $n+1$-го ключа. Используются три способа решения этой за-

дачи, показанные на рис. 8.22, причем последующие рассуждения основаны на очевидном соотношении

$$U_{\text{вых } n} = U_{\text{вх } (n+1)}, \tag{8.59}$$

где n — номер ключа в цепочке.

На рис. 8.22, *а* показан непосредственный способ решения указанной задачи. Пусть задана рабочая точка n. Проектируя ее на ось ординат (стрелка *1*), получаем значение $U_{\text{вых n}}$. Откладывая это значение по оси абсцисс (стрелка *2*), получаем согласно (8.59) значение $U_{\text{вх (n+1)}}$. Наконец, проводя вертикаль из этой абсциссы (стрелка *3*), находим рабочую точку $(n+1)$. Далее процедура повторяется.

На рис. 8.22, *б* показан второй, более удобный и наглядный способ решения, основанный на использовании биссектрисы, проведенной из начала координат. Биссектриса является геометрическим местом точек, характеризуемых равенством $U_{\text{вых}} = U_{\text{вх}}$. Значит, проектируя точку n на биссектрису (стрелка *1*), получаем точку n', абсцисса которой есть $U_{\text{вх (n +1)}}$. Проводя из точки n' вертикаль до пересечения с передаточной характеристикой (стрелка *2*), получаем точку $(n+1)$. Именно этот способ будет использоваться в дальнейшем.

На рис. 8.22, *в* показан третий способ решения, аналогичный методу линий нагрузки. Здесь передаточная характеристика последующего ключа (*2*) отложена в зеркальном отображении и повернута на 90° относительно характеристики предыдущего ключа (*1*). Точка n на кривой *1* проектируется по горизонтали на кривую *2* (стрелка *1*), и это дает рабочую точку $(n+1)$. Затем точка $(n+1)$ проектируется по вертикали снова на кривую *1* (стрелка *2*), и это дает точку $(n+2)$. Далее процесс повторяется.

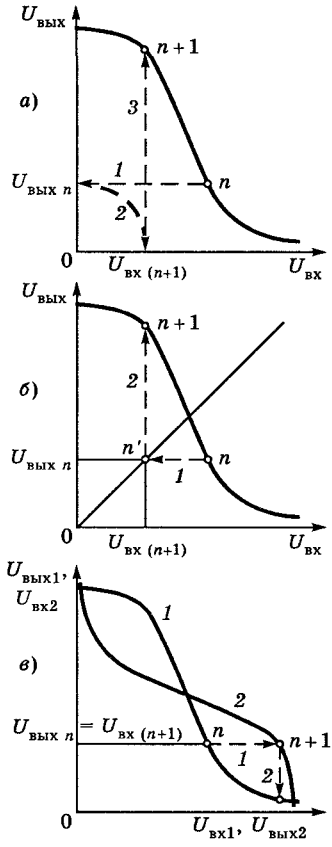


Рис. 8.22. Методы определения рабочих точек в последовательной цепочке ключей:
а — непосредственный метод;
б — метод биссектрисы;
в — метод линии нагрузки

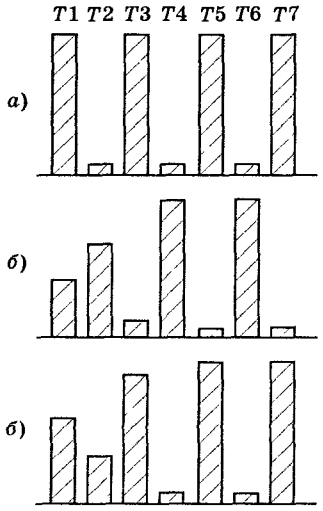


Рис. 8.23. Выходные напряжения ключей в последовательной цепочке:
а — в исходном состоянии;
б — при сигнале выше порогового;
в — при сигнале ниже порогового

Предположим, что в последовательной цепочке (см. рис. 8.6) ключ $T1$ заперт, т. е. на его выходе напряжение близко к напряжению питания (точка A на рис. 8.1). Тогда ключ $T2$ открыт и на его выходе остаточное напряжение близко к нулю (точка B на рис. 8.1). Соответственно ключ $T3$ заперт, ключ $T4$ открыт и т.д. Выходные напряжения ключей в исходном состоянии показаны на рис. 8.23, *а*.

Подадим теперь на вход ключа $T1$ управляющий сигнал $U_{вх1}$ (рис. 8.24, *а*). При этом рабочая точка ключа переместится из положения A в положение *1*. Проектируя точку *1* на биссектрису и проводя вертикаль из точки *1'* до передаточной характеристики, получаем рабочую точку ключа $T2$ (точка *2*). Проектируя точку *2* на биссектрису и проводя из точки *2'* вертикаль до пересечения с передаточной характеристикой, получаем точку *3* — рабочую точку ключа $T3$. Продолжая процедуру, убеждаемся, что рабочие точки нечетных ключей, начиная с $T5$, совпадают с точкой B (открытое состояние), а рабочие точки четных ключей, начиная с $T4$, — с точкой A (запертое состояние). Иначе говоря, сигнал $U_{вх1}$ достаточен для управления цепочкой, так как под его действием все ключи (кроме нескольких первых) изменяют свое состояние (рис. 8.23, *б*).

Теперь рассмотрим случай, когда при том же исходном состоянии цепочки подан меньший управляющий сигнал $U_{вх2}$ (рис. 8.24, *б*). Используя прежнюю процедуру, получаем рабочие точки *1*, *2*, *3* и т.д. и убеждаемся, что данный сигнал недостаточен для управления цепочкой, так как изменяется состояние только нескольких первых ключей (и то не полностью), а состояние остальных остается прежним (рис. 8.23, *б*).

Нетрудно догадаться, что критерием достаточной величины сигнала является условие $U_{вх} > U_C$. Поэтому напряжение U_C,

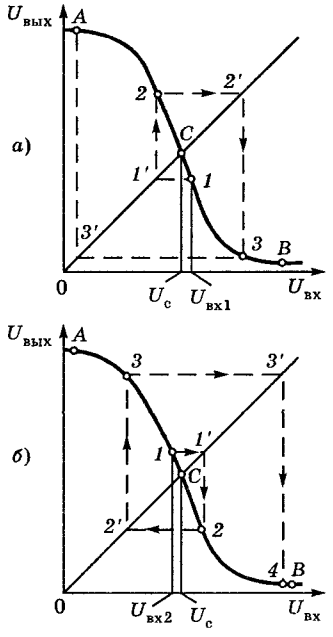
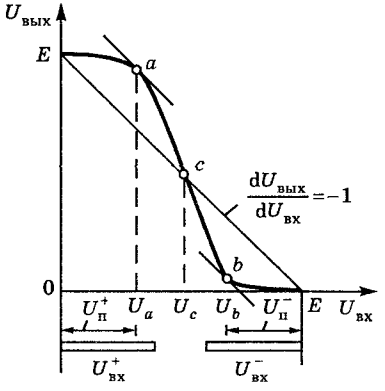


Рис. 8.24. Расчет состояния ключей в цепочке при сигнале выше (*а*) и ниже (*б*) порога чувствительности

Рис. 8.25. Оценка помехоустойчивости ключа



соответствующее точке пересечения биссектрисы с передаточной характеристикой, называют *порогом чувствительности* (переключения).

На первый взгляд, порог чувствительности непосредственно определяет помехоустойчивость ключевой цепочки. А именно, если импульс положительной помехи удовлетворяет условию $U_{вх}^{+} < U_С$, а импульс отрицательной помехи — условию $U_{вх} < (E - U_С)$, то состояние цепочки (за исключением первых нескольких ключей) не будет меняться (рис. 8.25)[1]. На самом деле указанные условия необходимы, но недостаточны.

Достаточными условиями помехоустойчивости (по причинам, которые пояснены ниже) являются следующие:

$$U_{вх}^{+} < U_a; \qquad U_{вх}^{-} < E - U_b.$$

---

1 Импульс отрицательной помехи целесообразно отсчитывать от устойчивой точки открытого ключа (точка *B* на рис. 8.1), поскольку такой импульс способствует запиранию ранее открытого ключа. Значит, величину $U_п$ нужно сравнивать с величиной $E - U_С$, а не с $U_С$.

Здесь U_a и U_b — абсциссы точек a и b, характерных тем, что для них модуль производной $dU_{вых}/dU_{вх}$ равен единице. Эта производная есть не что иное, как *дифференциальный коэффициент усиления* ключа. Если $U^+_{вх} < U_a$, то коэффициент усиления $K < 1$, т. е. входной сигнал в цепочке не усиливается, а ослабляется, и нет опасности ложного переключения. Если же $U^+_{вх} > U_a$, то $K > 1$ и «верхушка» импульса $U^+_{вх} - U_a$ усиливается. При наличии обратных связей это может стать причиной ложного переключения (ложного потому, что сам по себе входной сигнал $U^+_{вх}$ меньше порога чувствительности). Аналогично обстоит дело и с отрицательной помехой.

Учитывая все сказанное, помехоустойчивость ключа оценивают величинами

$$U^+_{п} = U_a; \tag{8.60а}$$

$$U^-_{п} = -(E - U_b). \tag{8.60б}$$

Обе эти величины показаны на рис. 8.25.

8.8. Бистабильные ячейки и триггеры

Простейшие ключи, рассмотренные в предыдущих разделах, составляют основу всей цифровой схемотехники. Они широко используются как самостоятельно (в качестве прерывателей тока и разного рода коммутаторов), так и в составе некоторых стандартных функциональных узлов, прежде всего — в бистабильных ячейках. Для бистабильных ячеек характерно использование не только прямых связей между ключами (как в последовательной цепочке), но и положительных обратных связей.

Схема и принцип действия. В последовательной цепочке ключей (рис. 8.6) каждый ключ «окружен» ключами, находящимися в противоположном состоянии. Значит, в произвольной паре смежных ключей (T_n и T_{n+1} на рис. 8.26) выходное напряжение ключа T_{n+1} такое же, как входное напряжение ключа T_n. Поэтому, изолировав рассматриваемую пару от предыдущих и

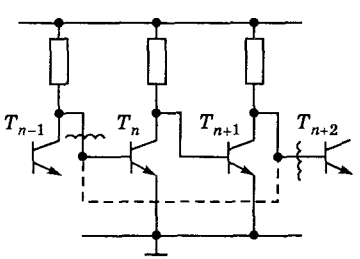


Рис. 8.26. Происхождение бистабильной ячейки

последующих звеньев цепочки и соединив выход $(n + 1)$-го ключа со входом n-го (см. штриховую линию на рис. 8.26), мы не изменим состояния пары. Это устойчивое состояние может иметь два варианта: T_n заперт, T_{n+1} открыт или, наоборот, T_n открыт, T_{n+1} заперт. Такие электронные схемы, имеющие два равноценных варианта устойчивых состояний, называют *бистабильными ячейками* (БЯ) или *триггерами*[1].

Если изобразить БЯ, отвлекаясь от ее «происхождения», т. е. в виде самостоятельной схемы (рис. 8.27), то этой схеме, как видим, свойственны симметричная конфигурация и наличие перекрестных (взаимных) обратных связей. Устойчивые состояния БЯ характерны тем, что один из ключей заперт, а второй открыт и насыщен. Иначе говоря, бистабильной ячейке свойственна электрическая асимметрия. Покажем, что электрическая симметрия в БЯ невозможна.

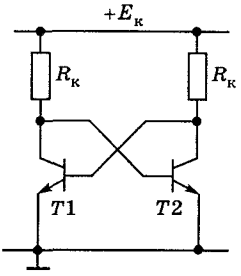


Рис. 8.27. Бистабильная ячейка

Доказательство проведем от противного. Пусть схема БЯ находится в симметричном состоянии, когда оба транзистора (рис. 8. 27) открыты и работают на границе активного режима[2]. Напряжения на обоих коллекторах и обеих базах одинаковы и близки к U^*; коллекторный ток пропорционален току базы:

$$I_к = BI_б .$$

Пусть теперь в результате неизбежных флюктуаций (внутренних или внешних) напряжение на одной из баз, например на базе транзистора $T1$, изменилось на малую величину $\Delta U_{б1}$. Тогда токи изменяются следующим образом:

$$\Delta I_{б1} = \Delta U_{б1}/R_{вх}; \quad \Delta I_{к1} = B\Delta I_{б1}.$$

1 Строго говоря, термины «триггер» (англ. trigger) и «бистабильная ячейка» (англ. flip-flop) — не синонимы. Бистабильная ячейка составляет лишь основу всякого триггера, а триггеры различаются способами управления бистабильной ячейкой (см. ниже). Однако на практике термин «триггер» часто используют и для БЯ.

2 Запертое состояние обоих транзисторов невозможно, так как при этом коллекторные потенциалы были бы равны $+ E_к$, т. е. превышали бы напряжение отпирания U^*. Невозможно и насыщенное состояние обоих транзисторов, так как тогда коллекторные потенциалы были бы меньше U^*.

Здесь $R_{вх}$ — входное сопротивление открытого транзистора. Часть приращения $\Delta I_{к1}$ ответвится в цепь базы транзистора $T2$; тогда

$$\Delta I_{б2} = -m\Delta I_{к1}; \quad \Delta I_{к2} = B\Delta I_{б2},$$

где $m < 1$. Аналогично часть приращения $\Delta I_{к2}$ ответвится в цепь базы транзистора $T1$ в виде дополнительного приращения базового тока:

$$\Delta I'_{б1} = -m\Delta I_{к2} = m^2B^2\Delta I_{б1}.$$

При обычных значениях $m \approx 0{,}5$ дополнительное приращение $\Delta I'_{б1}$, получившееся при обходе цепи, будет значительно превышать исходное приращение $\Delta I_{б1}$. Следующее приращение $\Delta I''_{б1}$ окажется во столько же раз больше $\Delta I'_{б1}$ и т. д. Значит, *реакция схемы на малейшую исходную флюктуацию состоит в усилении последней.*

Лавинообразный процесс нарастания токов в одной половине БЯ и уменьшения токов в другой половине называют *регенерацией*. Регенерация заканчивается запиранием одного из ключей и насыщением второго. В нашем примере при положительной флюктуации $\Delta U_{б1}$ запирается транзистор $T2$, а при отрицательной — транзистор $T1$.

Поскольку знак флюктуации — величина случайная, то и результаты лавинообразного процесса (запирание транзистора $T2$ или $T1$) равновероятны. Значит, при анализе БЯ можно с равным основанием считать исходным любое из двух возможных устойчивых состояний.

Цель управления бистабильной ячейкой состоит в том, чтобы с помощью внешних сигналов задавать то или иное из двух устойчивых состояний.

Установка состояния БЯ. Для реализации режима установки, параллельно каждому из транзисторов, входящих в состав БЯ, подключается еще один транзисторный ключ ($T3$ или $T4$, рис. 8.28, *а*). Эти ключи управляются внешним сигналом — током базы, принимающим одно из двух значений: $I_б^+$ или 0. Функции управляющих ключей аналогичны функциям металлических контактов: они могут быть либо замкнуты, либо разомкнуты.

Пусть в исходном состоянии БЯ транзистор $T1$ заперт, $T2$ открыт и насыщен, а оба ключа $T3$ и $T4$ заперты. Если отпереть транзистор $T4$ током $I_б^+$, то состояние БЯ не изменится, так как напряжение $U_{к2}$ в исходном состоянии уже было близко к нулю. Если же отпереть транзистор $T3$, то потенциал $U_{к1}$ падает до нуля; вместе с ним падает до нуля потенциал базы

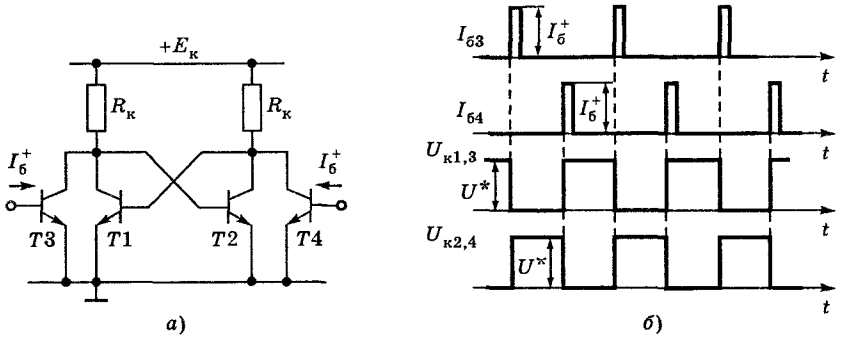


Рис. 8.28. Схема (*а*) и временные диаграммы (*б*) симметричного триггера с раздельными входами установки

$U_{б2}$, а значит, транзистор $T2$ запирается. При этом благодаря регенерации транзистор $T1$ отпирается до насыщения. После того как достигнуто новое устойчивое состояние, ключ $T3$ теряет свое управляющее действие; его «размыкание» и «замыкание» не меняют потенциалов $U_{к1}$ и $U_{б2}$, близких к нулю. Чтобы вернуть БЯ в исходное состояние, нужно теперь отпереть транзистор $T4$.

Необходимо подчеркнуть, что *одновременное поступление отпирающих импульсов на оба входа в данной схеме недопустимо.*

Действительно, если отпирающие сигналы действуют одновременно, то на базах обоих транзисторов $T1$ и $T2$ будет нулевой потенциал и оба эти транзистора будут заперты. По окончании сигналов оба транзистора откроются, т. е. БЯ окажется временно в симметричном состоянии. Из этого состояния, как показано выше, БЯ с равной вероятностью может перейти в любое из двух устойчивых состояний. Значит, *результат одновременного действия отпирающих сигналов оказывается неоднозначным*, что неприемлемо в цифровых схемах. Рассмотренный вариант БЯ, работающий в режиме раздельных входов, получил название RS-триггера[1].

---

1 Буква S — начальная в слове set (установление), буква R — начальная в слове reset (переключение). Следовательно, сигнал на входе S устанавливает начальное (исходное) состояние БЯ, а сигнал на входе R меняет это состояние.

Переходные процессы. Разумеется, переход БЯ из одного устойчивого состояния в другое происходит не мгновенно, поскольку каждому ключу, входящему в ее состав, свойственна инерционность при изменении токов и напряжений.

Переходный процесс протекает следующим образом. Пусть в исходном состоянии транзистор $T1$ заперт, а $T2$ открыт (рис. 8.28), и пусть в момент $t = 0$ скачком отпирается (до насыщения) управляющий ключ $T3$. Тогда на 1-м этапе переходного процесса происходит запирание транзистора $T2$, т. е. рассасывание избыточного заряда в его слоях, а затем быстрый спад до нуля коллекторного тока (см. разд. 8.3). Далее, на 2-м этапе, заряжается входная емкость транзистора $T1$. Когда напряжение $U_{б1}$ достигает значения U^*, транзистор $T1$ отпирается. На 3-м этапе происходит нарастание тока $I_{к1}$; этот этап заканчивается насыщением транзистора $T1$. На 4-м этапе происходит накопление избыточного заряда в слоях насыщенного транзистора (см. разд. 8.3).

Из сказанного следует, что общее время переключения триггера ($t_п$) складывается из следующих основных составляющих: времени рассасывания $t_р$, задержки фронта $t_{зф}$, длительности фронта $t_ф$ и времени накопления $t_н$:

$$t_п = t_р + t_{зф} + t_ф + t_н. \qquad (8.61)$$

Значения перечисленных составляющих определяются соответствующими выражениями в разделе 8.3.

Напомним, что в случае достаточно коротких входных сигналов время накопления согласно (8.22) определяется длительностью входного сигнала. В пределе, если $t^+ = t_{зф} + t_ф$, накопления избыточного заряда не происходит и $t_н = 0$. Таким образом, максимальная рабочая частота триггера определяется временем переключения при условии $t_н = 0$:

$$F_{макс} = (t_р + t_{зф} + t_ф)^{-1}. \qquad (8.62)$$

Например, если $t_р = 5$ нс, $t_{зф} = 3$ нс и $t_ф = 2$ нс, то $F_{макс} = 100$ МГц.

8.9. Триггер Шмитта

В основе триггера Шмитта лежит переключатель тока, рассмотренный в разделе 8.5. Для того, чтобы подчеркнуть эту

преемственность, в схеме триггера Шмитта (рис. 8.29) переключатель тока изображен непрерывными линиями, а делитель напряжения R_1, R_2 — штриховыми. С этой же целью потенциал $U_{б2}$ обозначен через E, хотя в данном случае величина E не постоянная.

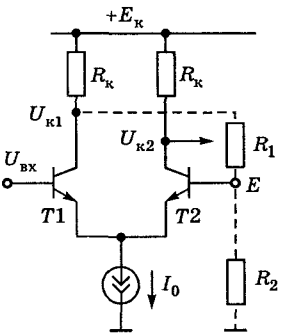


Рис. 8.29. Триггер Шмитта

Для упрощения анализа идеализируем делитель R_1, R_2, считая, что он не потребляет тока и только передает часть напряжения $U_{к1}$ на базу транзистора $T2$:

$$E = \gamma U_{к1}, \quad \text{где } \gamma = R_2/(R_1 + R_2).$$

Пусть в исходном состоянии транзистор $T1$ заперт, а $T2$ открыт, но работает в активном режиме. Такому состоянию отвечают следующие значения потенциалов:

$$U_{к1} = E_к; \quad E = \gamma E_к; \quad U_{к2} = E_к - I_0 R_к.$$

Активный режим транзистора $T2$ подразумевает выполнение условия $U_{к2} \geq E$, откуда легко получить ограничение на величину $I_0 R_{к2}$.

Исходное состояние сохраняется до тех пор, пока транзистор $T1$ остается запертым, т. е. при входных напряжениях, меньших E. Обозначим через $U^+_{вх}$ напряжение отпирания транзистора $T1$:

$$U^+_{вх} = E - \delta = \gamma E_к - \delta, \qquad (8.63)$$

где $\delta \approx 0{,}1$ В. Если $E_к = 5$ В и $\gamma = 1/2$, то $U^+_{вх} \approx 2{,}4$ В.

Пусть входной сигнал $U_{вх}$ немного превысил напряжение отпирания и в результате появилось малое приращение тока $\Delta I_{к1}$. Тогда получается следующая цепочка взаимодействий:

$$\Delta U_{к1} = -\Delta I_{к1} R_к; \quad \Delta E = \gamma \Delta U_{к1}; \quad \Delta U_э = \Delta E$$

(последнее равенство предполагает, что прямое напряжение на эмиттерном переходе не меняется и остается равным U^*). Под

действием приращения $\Delta U_э$ возникает дополнительное приращение коллекторного тока:

$$\Delta I'_{к1} = -S\Delta U_э = \gamma S R_к \Delta I_{к1},$$

где S —крутизна транзистора[1].

Если произведение $\gamma S R_к$ превышает единицу, то дополнительное приращение $\Delta I'_{к1}$, получившееся при обходе замкнутой цепи, будет больше исходного приращения $\Delta I_{к1}$. Значит, в схеме развивается *регенеративный* (лавинообразный) процесс, в результате которого ток I_0 переходит в транзистор $T1$, а транзистор $T2$ запирается.

Напряжение $U^+_{вх}$, при котором происходит скачкообразное переключение тока I_0 из транзистора $T2$ в $T1$, называют *напряжением срабатывания триггера.*

После срабатывания потенциалы в схеме имеют вид:

$$U_{к1} = U_{к0}; \quad E = \gamma U_{к0}; \quad U_{к2} = E_к,$$

где $U_{к0}$ — коллекторный потенциал открывшегося транзистора $T1$. В зависимости от параметров I_0 и $R_{к1}$ транзистор $T1$ после срабатывания может работать как в активном режиме, так и в режиме насыщения. Типичным является активный режим. В таком случае

$$U_{к0} = E_к - I_0 R_{к1}, \tag{8.64}$$

причем $U_{к0} \geq U^+_{вх}$ (поскольку потенциал базы сохраняет свое значение в процессе срабатывания).

Чтобы вернуть триггер в исходное состояние, нужно уменьшить входной сигнал до значения, близкого к E, при котором начинает отпираться транзистор $T2$.

Обозначим через $U^-_{вх}$ напряжение отпирания транзистора $T2$:

$$U^-_{вх} = E + \delta = \gamma U_{к0} + \delta. \tag{8.65}$$

---

1 Знак минус в соотношении между приращениями $\Delta I_{к1}$ и $\Delta U_э$ объясняется тем, что у *n–p–n*-транзисторов ток возрастает при *отрицательном* приращении эмиттерного потенциала.

Если положить $\gamma = 1/2$ и $U_{к0} = U^{+}_{вх} = 2{,}4$ В [см. (8.63)], то $U^{-}_{вх} \approx 1{,}3$ В.

Как только напряжение $U_{вх}$ сделается меньше $U^{-}_{вх}$, транзистор $T2$ начинает отпираться и развивается регенеративный процесс, подобный описанному выше. В результате триггер скачком возвращается в исходное состояние, в котором транзистор $T1$ заперт, а транзистор $T2$ — открыт.

Напряжение $U^{-}_{вх}$, при котором происходит скачкообразное переключение тока из транзистора $T1$ в $T2$, называют *напряжением отпускания триггера*.

Пренебрегая малой величиной δ и учитывая равенство $U_{к0} = U^{+}_{вх}$, находим из (8.65), что напряжение отпускания меньше напряжения срабатывания. Такое соотношение принципиально для триггера Шмитта.

Часто вместо терминов «напряжение срабатывания» и «напряжение отпускания» используют более общие термины: *порог срабатывания* и *порог отпускания*. В соответствии с этими терминами триггер Шмитта называют пороговым устройством.

Выходной сигнал в триггере Шмитта снимают с коллектора транзистора $T2$. Поскольку этот коллектор — «свободный» (т. е. не связан с цепью обратной связи), а транзистор $T2$ работает в ненасыщенном режиме, длительность переходных процессов оказывается предельно короткой — гораздо меньшей, чем у БЯ. Это одно из важных преимуществ триггера Шмитта перед другими триггерами.

На рис. 8.30 показана передаточная характеристика триггера Шмитта, где $U_{вых} = U_{к2}$. Как видим, наличие двух пороговых напряжений приводит к появлению на характеристике

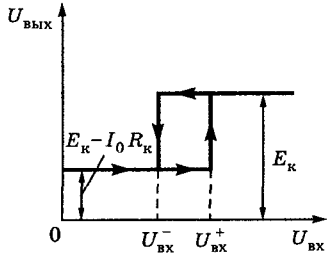


Рис. 8.30. Передаточная характеристика триггера Шмитта

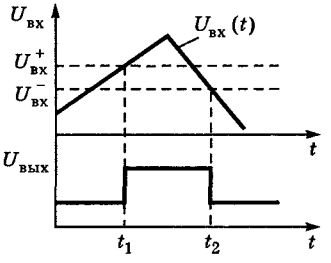


Рис. 8.31. Временны́е диаграммы при использовании триггера Шмитта в качестве дискриминатора амплитуды и формирователя импульсов

«петли гистерезиса», ширина которой составляет $U^+_{вх} - U^-_{вх}$. При полученных выше значениях $U^+_{вх}$ и $U^-_{вх}$ ширина «петли гистерезиса» составляет около 1,1 В.

На рис. 8.31 иллюстрируется типичный способ использования триггера Шмитта в качестве порогового устройства. При плавном нарастании $U_{вх}$, начиная с нуля, схема находится в исходном состоянии ($T1$ заперт). Как только $U_{вх}$ достигает значения $U^+_{вх}$, происходит срабатывание триггера, транзистор $T2$ запирается и на выходе получается положительный скачок напряжения. Если входной сигнал не достигает порога срабатывания, то выходной сигнал (скачок) отсутствует. Следовательно, триггер Шмитта позволяет «сортировать» входные сигналы по их амплитуде: больше $U^+_{вх}$ или меньше $U^+_{вх}$. Устройства такого рода называют *амплитудными анализаторами* или *дискриминаторами* (различителями).

Если амплитуда входных сигналов заведомо больше, чем порог срабатывания, то функция триггера Шмитта меняется: он превращается в так называемый *формирователь импульсов*. Эта функция означает, что плавные входные сигналы с меняющейся амплитудой «превращаются» в стандартные (по амплитуде) прямоугольные импульсы с короткими фронтами. Такое преобразование часто необходимо на практике.

В заключение отметим, что при достаточно больших входных сигналах ($U_{вх} > U^+_{вх}$) транзистор $T1$ оказывается в режиме насыщения. Это не приводит к отпиранию транзистора $T2$, т. е. не нарушает работу триггера, но сопровождается резким возрастанием входного тока триггера и ухудшением его быстродействия (из-за необходимости рассасывания накопленного избыточного заряда).

Контрольные вопросы

1. Чем отличаются передаточные характеристики инвертирующих и неинвертирующих электронных схем?
2. Чем определяется остаточное напряжение простейшего ключа на биполярном транзисторе?
3. Сформулируйте критерий насыщения ключа на биполярном транзисторе и дайте определение степени насыщения.
4. Чем может быть обусловлено неравномерное распределение тока базы в параллельно соединенных ключевых транзисторах? Как достигается выравнивание базовых токов?

5. Может ли простейший ключ на биполярном транзисторе быть использован в качестве усилителя?
6. Какими параметрами определяется нагрузочная способность ключа? Почему необходим запас по нагрузочной способности?
7. Какими динамическими параметрами характеризуются переходные процессы в простейшем биполярном ключе?
8. Каково влияние барьерной и диффузионной емкостей коллекторного перехода на переходные процессы в простейшем биполярном ключе?
9. Объясните принцип работы ключа на биполярном транзисторе с барьером Шоттки.
10. Дайте сравнительную характеристику простейшего ключа на биполярном транзисторе и ключа с барьером Шоттки.
11. Как исключается насыщение транзисторов в переключателе тока? Чем обусловлено его высокое быстродействие?
12. Опишите принцип действия и основные характеристики МДП-транзисторного ключа с резистивной нагрузкой.
13. Почему МДП-транзисторный ключ с динамической нагрузкой предпочтительнее ключа с резистивной нагрузкой?
14. Опишите принцип работы и характеристики ключа на комплементарных МДП-транзисторах.
15. Чем обусловлена инерционность МДП-транзисторных ключей? Какова роль емкости затвор-сток в переходных процессах?
16. Какие параметры характеризуют помехоустойчивость ключей? Проанализируйте простейший биполярный и МДП-транзисторный ключи с точки зрения помехоустойчивости.
17. Объясните происхождение и принцип действия бистабильной ячейки.
18. Сколько последовательных инвертирующих ключевых ячеек должно быть в кольцевой цепочке, чтобы она имела два устойчивых состояния?
19. Как осуществляется установка состояния триггера? Существуют ли ограничения на комбинации сигналов установки?
20. Опишите принцип работы и назначение триггера Шмитта. Что такое порог срабатывания и порог отпускания?

Глава 9

Основы аналоговой схемотехники

9.1. Введение

Общее представление об особенностях аналоговых схем было дано в разделе 8.2. Номенклатура аналоговых схем весьма обширная и разнообразная. До зарождения микроэлектроники трудно было представить, что в области аналоговых схем можно найти небольшое число типовых структур, которые, подобно ключу и переключателю тока в цифровых схемах, послужили бы основой для всех или большинства остальных схем. Однако с развитием микроэлектроники такие типовые структуры были найдены, и они кратко охарактеризованы в следующей главе.

Усилители и раньше играли ведущую роль в аналоговой технике, выполняя самую распространенную функцию — усиление мощности слабых сигналов. В эпоху микроэлектроники эта роль стала еще более важной, поскольку основная разновидность аналоговых ИС — операционный усилитель способен выполнять самые разнообразные функции, отнюдь не только усилительные. Поэтому ниже усилителям уделено главное внимание.

Усилители можно классифицировать по разным признакам. По полосе (диапазону) усиливаемых частот различают *широкополосные* и *узкополосные* (*избирательные*) усилители. Важнейшей разновидностью широкополосных усилителей являются *усилители постоянного тока*, способные усиливать сигналы сколь угодно низкой частоты. По мощности различают маломощные и мощные усилители. Особенностью мощных усилителей являются повышенные требования к коэффициенту полезного действия, т.е. мощность выходного сигнала должна быть соизмерима с мощностью, потребляемой от источника питания. По назначению различают линейные, логарифмические, дифференцирующие, электрометрические и другие усилители.

В дискретной транзисторной схемотехнике широкое распространение имела классификация усилителей по типу связей между отдельными узлами. Различали емкостную, трансформаторную, гальваническую (непосредственную) связи. Наибо-

лее популярной была емкостная связь, с которой и начинали изучение усилителей. Однако в полупроводниковых ИС, как известно, осуществление конденсаторов связано с трудностями, а осуществление трансформаторов — невозможно. Поэтому на первый план выдвинулись усилители с непосредственной связью каскадов.

Поскольку непосредственная (безреактивная) связь обеспечивает усиление сколь угодно медленных сигналов, усилители с такой связью являются усилителями постоянного тока, которые уже упоминались выше. Эти усилители составляют основу интегральной аналоговой техники.

9.2. Составные транзисторы

Во многих аналоговых ИС применяются комбинации из нескольких (обычно двух) транзисторов, соединенных между собой так, что их можно рассматривать как единое целое — *составной транзистор*. Составные транзисторы обладают такими свойствами, которые трудно или невозможно получить в транзисторах с обычной структурой.

Среди составных транзисторов наибольшее распространение имеет так называемая *пара Дарлингтона* (рис. 9.1). Ее главная особенность — исключительно большая величина коэффициента усиления базового тока.

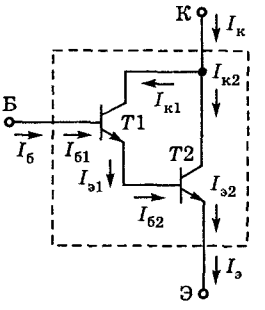


Рис. 9.1
Пара Дарлингтона

Действительно, из рис. 9.1 следует:

$$I_{б2} = I_{э1} = (B_1 + 1)\, I_б;\quad I_к = B_1 I_б + B_2 I_{б2}.$$

Подставляя значение $I_{б2}$ во второе равенство и деля обе части на $I_б$, получаем эквивалентный коэффициент усиления пары Дарлингтона:

$$B = B_1 + B_2 + B_1 B_2. \qquad (9.1а)$$

Во всех практических случаях первые два члена в правой части несущественны и эквивалентный коэффициент усиления можно записать в виде

$$B = B_1 B_2. \qquad (9.1б)$$

Если составляющие B_1 и B_2 имеют значения 100—200, то расчетный коэффициент B составляет $(1—4)\cdot10^4$. Примерно таким же будет и дифференциальный коэффициент усиления β.

Следует, однако, иметь в виду, что транзисторы в паре Дарлингтона работают при существенно отличающихся величинах эмиттерных токов: эмиттерный ток $I_{э2}$ превышает ток $I_{э1}$ примерно в B_2 раз. Коэффициент усиления B биполярного транзистора сильно зависит от тока эмиттера. Если транзисторы $T1$ и $T2$ однотипны, условиям оптимальности тока для обоих транзисторов удовлетворить невозможно. На практике для устранения этого противоречия в качестве транзистора $T2$ используют структуру в виде параллельно включенных транзисторов, однотипных с $T1$, число которых приблизительно равно B_1. В этом случае коэффициенты усиления $T1$ и структуры $T2$ равны, т.к. все транзисторы работают при одинаковых токах эмиттера. В дискретной схемотехнике такой подход неприемлем, однако в интегральной схемотехнике — это типовой прием, обеспечивающий выравнивание токов.

На рис. 9.2 показана схема еще одного составного транзистора, который можно назвать *составным p–n–p-транзистором.*

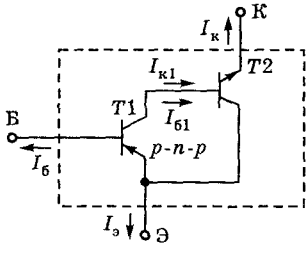


Рис. 9.2. Составной *p–n–p*-транзистор

В данном случае речь идет о соединении двух транзисторов разного типа: *p–n–p* и *n–p–n*. Направления результирующих токов, как видно из рисунка, соответствуют транзистору *p–n–p*-типа. Что касается коэффициента усиления, то он определяется выражением $B = B_1 + B_1B_2$, которое практически совпадает с выражениями (9.1) для пары Дарлингтона.

Интегральные *p–n–p*-транзисторы уступают *n–p–n* транзисторам, в частности, по коэффициенту усиления B. В этом отношении составной *p–n–p*-транзистор обладает явными преимуществами, так как его коэффициент усиления превышает коэффициент усиления *n–p–n*-транзистора, входящего в состав схемы. Быстродействие составного транзистора такое же, как у составляющего *p–n–p*-транзистора, т. е. хуже, чем у *n–p–n*-транзистора.

9.3. Статический режим простейшего усилителя

Усилители обычно состоят из нескольких однотипных элементарных ячеек, которые называют *усилительными каскадами* (или ступенями). Усилительный каскад может содержать один, два и более транзисторов, т.е. сам по себе может представлять достаточно сложную схему. Однако каскад, даже сложный, нельзя разделить на более простые компоненты без утраты его специфических свойств. В этом смысле каскад и называют элементарной ячейкой усилителя.

Схема и режим покоя. Простейший усилитель постоянного тока, выполненный на одном транзисторе, показан на рис. 9.3, *а*. Для таких усилителей характерно двуполярное питание, т.е. использование двух источников питания с напряжениями $+E_к$ и $-E_э$ относительно земли.

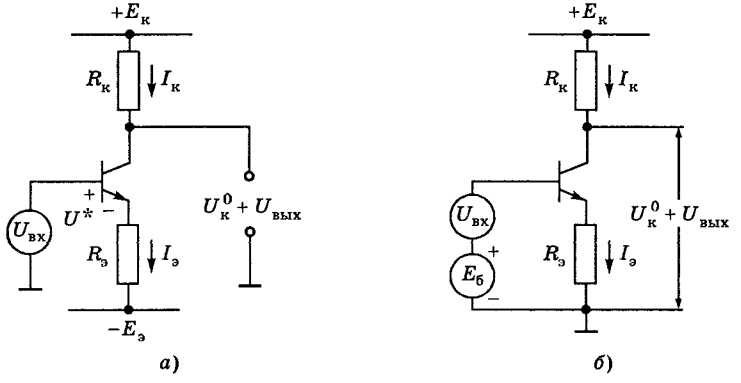


Рис. 9.3. Варианты усилительных каскадов с двуполярным (*а*) и однополярным (*б*) питанием

В принципе можно использовать и однополярное питание (рис. 9.3,б). Однако в таком варианте возникают серьезные затруднения. Во-первых, требуется специальный источник смещения $E_б$. Во-вторых, источник сигнала не имеет заземленной точки, что исключает использование большинства типичных источников сигнала и резко повышает уровень паразитных наводок (помех) на входе усилителя. Если поменять местами источники сигнала и смещения, то высокий уровень помех остается, а реализовать схемным путем незаземленный источник смещения практически невозможно.

Пусть входной сигнал $U_{вх}$ равен нулю. Тогда в схеме протекают *постоянные составляющие* токов, обусловленные источниками $E_к$ и $E_э$. Режим отсутствия сигнала принято называть *режимом покоя* усилителя.

При наличии входного сигнала к постоянным составляющим добавляются *переменные составляющие*, пропорциональные величине $U_{вх}$. Таким образом, в рабочем режиме полные величины напряжений и токов можно записать в виде: $U = U^0 + \Delta U$; $I = I^0 + \Delta I$. Здесь верхний индекс «0» присвоен постоянным составляющим, а переменные составляющие (которые обычно малы по сравнению с постоянными) обозначены как приращения.

Как уже отмечалось в разделе 5.6, анализ постоянных и переменных составляющих можно проводить независимо. Рассмотрим постоянные составляющие, свойственные режиму покоя.

Положим $U_{вх} = 0$ и изобразим эквивалентную схему усилителя так, как показано на рис. 9.4. Здесь для транзистора использована упрощенная модель Молла–Эберса, соответствующая нормальному активному режиму (ср. с рис. 5.13), а в цепь базы для общности введено сопротивление $R_б$. Оно включает в себя внутреннее сопротивление базы $r_б$, а также сопротивление источника сигнала или предыдущего каскада.

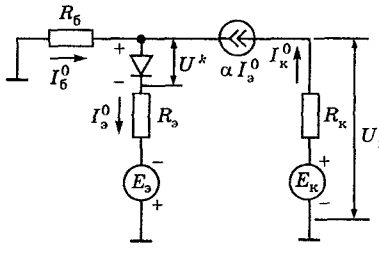


Рис. 9.4. Эквивалентная схема каскада для постоянных составляющих

Обойдя входной контур схемы на рис. 9.4, получаем уравнение

$$I_б^0 R_б + U^* + I_э^0 R_э - E_э = 0.$$

Подставляя $I_б^0 = (1-\alpha) I_э^0$, легко найти ток эмиттера:

$$I_э^0 = \frac{E_э - U^*}{R_э + (1-\alpha) R_б}. \qquad (9.2а)$$

Потенциал коллектора имеет вид

$$U_к^0 = E_к - I_к^0 R_к, \qquad (9.2б)$$

где $I_к^0 = \alpha I_э^0$.

Величины $I_э^0$ и $U_к^0$ задаются заранее. Их совокупность определяет, как говорят, *рабочую точку* транзистора в режиме покоя. Напряжение питания $E_к$ обычно тоже задано; тогда из выражения (9.2б) однозначно получается необходимое сопротивление $R_к$.

Что касается величин $E_э$ и $R_э$, то обе они должны быть достаточно большими с тем, чтобы неизбежные изменения параметров α и U^* не оказывали заметного влияния на ток $I_э^0$. Можно сказать, что выбор значений $E_э$ и $R_э$ определяется желательной стабильностью рабочей точки транзистора при изменениях температуры и других факторов.

Сопротивление $R_э$ выбирают из условия

$$R_э >> (1 - \alpha) R_б. \qquad (9.3)$$

Например, если $R_б$ = 2 кОм и α = 0,99 (т.е. B = 100), то сопротивление $R_э$ должно быть не менее 200 Ом.

Выбрав сопротивление $R_э$, легко найти величину $E_э$ из (9.2а). Может оказаться, что значение $E_э$ недостаточно велико, чтобы в нужной степени предотвратить влияние изменений U^*. Тогда нужно увеличить $E_э$ и соответственно $R_э$. Обычно $E_э$ > 2–3 В.

Дифференциальные параметры. Сигнал $U_{вх}$ вызывает изменения потенциалов и токов в схеме, т.е. «порождает» переменные составляющие. Чтобы найти эти составляющие, воспользуемся малосигнальной моделью транзистора (рис. 5.16). Ограничиваясь пока областью достаточно низких частот, будем считать α действительной величиной и пренебрежем коллекторной емкостью. Пренебрежем также сопротивлением коллекторного перехода $r_к$: его учет не вносит существенных поправок в результаты анализа. Тогда эквивалентная схема каскада будет такой, как показано на рис. 9.5. Здесь учтено внутреннее сопротивление источника сигнала $R_г$, а величины токов и напряжений записаны без знака Δ.

Из рис. 9.5 следует:

$$U_{вх} = I_б(R_г + r_б) + I_э(R_э + r_э).$$

Подставляя $I_б = (1 - \alpha) I_э$, легко найти ток эмиттера:

$$I_э = \frac{U_{вх}}{R_э + r_э + (1 - \alpha)(R_г + r_б)}. \qquad (9.4)$$

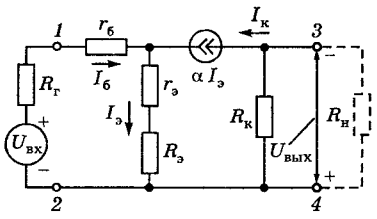


Рис. 9.5. Малосигнальная эквивалентная схема каскада в области низких частот

Учитывая условие (9.3), можно без большой погрешности положить $I_э = U_{вх}/R_э$. Зная ток $I_э$, легко определить все другие токи и напряжения в схеме.

Коэффициенты (размерные и безразмерные), связывающие переменные составляющие между собой и с входным сигналом, называют *дифференциальными параметрами* усилителя.

Главный из этих параметров — *коэффициент усиления* — определяется как отношение выходного и входного сигналов:

$$K = U_{вых}/U_{вх}.$$

Выходным сигналом принято считать переменную составляющую коллекторного напряжения $\Delta U_к$ (поэтому на рис. 9.3 потенциал коллектора записан как $U_к^0 + U_{вых}$). Из рис. 9.5 следует: $U_{вых} = -\alpha I_э R_к$. Подставляя ток $I_э$ из (9.4) и деля обе части на $U_{вх}$, находим коэффициент усиления в общем виде:

$$K = -\frac{\alpha R_к}{R_э + r_э + (1 - \alpha)(R_г + r_б)}. \qquad (9.5а)$$

Пренебрегая согласно (9.3) двумя последними членами в знаменателе, получаем простое выражение, вполне пригодное для всех практических расчетов:

$$K = -\alpha (R_к/R_э). \qquad (9.5б)$$

Знак «минус» свидетельствует о различии полярностей выходного и входного сигналов или (при синусоидальном сигнале) — о сдвиге фаз выходного и входного сигналов на 180°.

Из выражения (9.5б) следует, что сопротивление $R_к$ желательно делать большим, а сопротивление $R_э$ малым. Однако в реальной схеме сопротивление $R_к$ определяется напряжением

питания и рабочей точкой транзистора [см. (9.2б)], а сопротивление $R_э$ должно удовлетворять условию стабильности (9.3). Поэтому практически коэффициент усиления в рассматриваемом каскаде не превышает значений 4–5.

Ограниченность коэффициента усиления становится особенно очевидной, если преобразовать выражение (9.5а), подставив в него значения $R_к$ и $R_э$ из (9.2). Тогда, полагая $r_э << R_э$, получим:

$$K = -(E_к - U_к^0)/(E_э - U^*). \qquad (9.6)$$

Пусть, например, $E_к = 12$ В, $E_э = 3$ В и $U_к^0 = 2$ В; при этом получается $K \approx 4{,}5$. Полезно заметить, что коэффициент усиления не зависит от рабочих токов и что при равенстве напряжений питания $E_к$ и $E_э$ он делается близким к единице.

Если на выходе усилителя включена внешняя нагрузка $R_н$ (на рис. 9.5 показана штриховой линией), то в формулах (9.5) нужно заменить $R_к$ на эквивалентное сопротивление $R_к \parallel R_н$, где $\parallel$ — символ параллельного соединения.

Следующий важный параметр усилителя — *входное сопротивление* — определяется выражением

$$R_{вх} = U_{вх}/I_{вх},$$

где $I_{вх}$ — переменная составляющая базового тока, а напряжение $U_{вх}$ считается приложенным непосредственно к базе. Значит, при расчете входного сопротивления нужно полагать $R_г = 0$.

По отношению к источнику сигнала входное сопротивление играет роль нагрузки. Поэтому чем оно больше, тем меньше нагружен источник сигнала и тем лучше передается его напряжение на вход каскада.

Из рис. 9.5, полагая $R_г = 0$, получаем:

$$U_{вх} = I_б r_б + I_э (R_э + r_э).$$

Подставляя $I_э = (\beta + 1) I_б$ и деля обе части на $I_б = I_{вх}$, находим входное сопротивление в общем виде:

$$R_{вх} = r_б + (\beta + 1)(R_э + r_э). \qquad (9.7а)$$

Практически сопротивлениями $r_б$ и $r_э$ можно пренебречь. Тогда

$$R_{вх} \approx (\beta + 1) R_э. \qquad (9.7б)$$

Например, если $\beta = 100$ и $R_э = 2$ кОм, то $R_{вх} \approx 200$ кОм.

Заметим, что с ростом сопротивления $R_э$ входное сопротивление не может возрастать неограниченно, как следует из формул (9.7). О причинах ограничения говорится в разделе 9.4.

Третий важный параметр усилителя — *выходное сопротивление* — определяется выражением

$$R_{вых} = (U_{вых})_{хх}/(I_{вых})_{кз},$$

где $(U_{вых})_{хх}$ — выходное напряжение при холостом ходе каскада (т.е. в отсутствие внешней нагрузки $R_н$), а $(I_{вых})_{кз}$ — выходной ток при коротком замыкании выходных зажимов (имеется в виду короткое замыкание для переменных составляющих).

Выходное сопротивление характеризует нагрузочную способность каскада: чем оно меньше, тем больший ток можно отбирать во внешнюю нагрузку и тем меньше может быть внешнее сопротивление.

По физическому смыслу выходное сопротивление схемы — это дифференциальное сопротивление, которое можно измерить со стороны выходных зажимов в отсутствие входного сигнала ($U_{вх} = 0$) и при отключенной внешней нагрузке ($R_н = \infty$). В таких же условиях осуществляют и теоретический расчет величины $R_{вых}$.

Для схемы, показанной на рис 9.5, нет необходимости проводить специальные расчеты. Поскольку входная часть схемы отделена от выхода генератором тока и поскольку этот генератор при $U_{вх} = 0$ бездействует, можно сразу записать:

$$R_{вых} = R_н. \qquad (9.8)$$

Учет сопротивления коллекторного перехода (т.е. внутреннего сопротивления генератора тока) не оказывает никакого практического влияния на полученный результат.

Дрейф постоянных составляющих. В режиме покоя (при $U_{вх} = 0$) токи и напряжения усилителя могут меняться под действием изменений температуры, напряжений питания или других факторов. Получающиеся при этом медленные неконтролируемые приращения токов и напряжений называют *дрейфом постоянных составляющих*.

Если входной сигнал меняется достаточно быстро (с достаточно высокой частотой), то отличить его от дрейфа не представляет труда. Однако в усилителях постоянного тока, характеризующихся медленными изменениями сигнала, дрейф (т.е. паразитные приращения) в общем случае неотличим от приращений, вызванных полезным сигналом. Следовательно, величина дрейфа ставит предел чувствительности усилителя постоянного тока — его способности усиливать слабые сигналы.

Поскольку в микроэлектронике усилители постоянного тока занимают ведущее место, проблема дрейфа приобретает особую остроту.

Пусть в результате неконтролируемых процессов получились приращения напряжений $\Delta E_к$ и $\Delta E_э$, а также параметров транзистора ΔU^* и $\Delta\alpha$. Тогда из (9.2б) следует, что коллекторный потенциал $U_к^0$ получит приращение

$$\Delta U_к^0 = \Delta E_к + \Delta I_к^0 R_к$$

[знак «плюс» объясняется тем, что приращения $\Delta E_к$ и $\Delta I_к^0 R_к$ независимы — *некоррелированы* и потому в худшем случае могут складываться]. Приращение коллекторного тока можно записать, учитывая соотношение $I_к^0 = \alpha I_э^0$:

$$\Delta I_к^0 = \Delta\alpha\, I_э^0 + \alpha\Delta I_э^0.$$

Наконец, приращение эмиттерного тока можно получить из (9.2а):

$$\Delta I_э^0 = \frac{\Delta E_э + \Delta U^* + \Delta\alpha\, I_э^0 R_б}{R_э + (1-\alpha)R_б}.$$

Производя подстановки и деля $\Delta U_к^0$ на коэффициент усиления в форме (9.5а), получаем напряжение, которое называют *приведенным дрейфом*

$$U_{др} = \frac{\Delta U_к^0}{|K|}.$$

Приведенный дрейф удобен для оценки минимального входного сигнала, еще различимого на фоне дрейфа.

Если пренебречь небольшой разницей в знаменателях выражений (9.2а) и (9.5а) и перейти от α к более удобной величине B, то приведенный дрейф можно записать в следующем виде:

$$U_{др} = \Delta U^* + \left(\Delta E_э + \frac{\Delta E_к}{|K|}\right) + \frac{\Delta B}{B}\frac{I_э^0(R_э + R_б)}{B+1}. \qquad (9.9)$$

Это выражение говорит о том, что из четырех причин дрейфа (ΔB, $\Delta E_э$, $\Delta E_к$ и ΔU^*) принципиально неустранимой является составляющая ΔU^*. Остальные можно в принципе уменьшить схемным путем: стабилизируя напряжения питания и уменьшая рабочий ток.

Для количественной оценки дрейфа примем: B = 100; $\Delta B/B = 0{,}5$; $I_э^0 = 1$ мА; $R_э + R_б = 3$ кОм; $E_э = 3$ В; $E_к = 12$ В; $|K| = 5$. Тогда 3-я составляющая дрейфа будет иметь значение 15 мВ. Для того чтобы составляющие ΔE были меньше, необходимо стабилизировать напряжения питания до уровня 0,1 %. Что касается неустранимой составляющей ΔU^*, то ее значение прежде всего зависит от диапазона изменения температуры. Считая температурную чувствительность $\varepsilon = 1{,}5$ мВ/°С и полагая $\Delta T = 100$ °С, получаем $\Delta U^* = 150$ мВ. Как видим, эта составляющая самая значительная. Таким образом, в рассмотренном примере приведенный дрейф ограничивает входные сигналы снизу значениями порядка 100 мВ.

Каскадирование. Коэффициент усиления простейшего одно-каскадного усилителя ограничен и обычно не превышает значений 4–5. Поэтому получить нужный коэффициент усиления (часто — несколько тысяч и десятков тысяч) можно только соединением нескольких усилительных каскадов в последовательную цепочку. В такой цепочке выходное напряжение предыдущего каскада является входным для следующего. Результирующий коэффициент усиления будет произведением коэффициентов усиления каждого каскада:

$$K = K_1 K_2 \ldots K_N \qquad (9.10а)$$

или, в случае идентичных каскадов,

$$K = K_1^N. \qquad (9.10б)$$

Здесь N — число каскадов, K_1 — коэффициент усиления одного каскада.

Процесс объединения одиночных каскадов в последовательную цепочку называют *каскадированием*.

Из формулы (9.10б) легко найти необходимое количество каскадов, если заданы значения K и K_1:

$$N = \lg K / \lg K_1.$$

Пусть, например $K = 10^4$ и $K_1 = 4$; тогда $N \approx 7$. Как видим, количество каскадов в усилителях такого типа может быть весьма значительным.

При каскадировании простейших усилителей возникают специфические проблемы, которые дополнительно осложняют получение больших коэффициентов усиления. На этих проблемах мы не останавливаемся, так как сами простейшие усилители в микроэлектронике не популярны.

9.4. Переходные процессы в простейшем усилителе

Внешне транзистор в простейшем усилителе включен по схеме с общим эмиттером (ОЭ): входной сигнал подается на базу, выходной снимается с коллектора. Однако *по существу транзистор работает по схеме с общей базой* (ОБ), поскольку и в режиме покоя, и при наличии сигнала заданной величиной оказывается не ток базы, а ток эмиттера. Для переменной составляющей этот вывод следует из выражения (9.4) с учетом условия (9.3).

Значит, подавая на вход усилителя сигнал $U_{вх}$ ступенчатой формы, мы тем самым задаем ступеньку эмиттерного тока $\Delta I_э$. При этом переходные характеристики коллекторного тока и коллекторного напряжения будут определяться переходной характеристикой коэффициента α. С учетом влияния коллекторной емкости нужно вместо постоянной времени τ_α пользоваться эквивалентной постоянной времени $\tau_{\alpha oe}$ [см. (5.68)]. Подставляя изображение $\alpha(s)$ в (9.5б), получим изображение коэффициента усиления $K(s)$, на основании которого далее легко найти переходную и частотные характеристики. Все они в относительном масштабе будут совпадать с характеристиками коэффициента α, рассмотренными в разд. 5.7. Поэтому нет необходимости повторять соответствующие формулы и рисунки.

Вместо этого рассмотрим случай, когда ток эмиттера нельзя считать заданным, т. е. когда слагаемое $(1 - \alpha)(R_г + r_б)$ в знаменателе (9.5а) сравнимо с двумя первыми слагаемыми. Этот случай не характерен для простейшего каскада, но важен в методическом отношении как основа для анализа других вариантов усилителей.

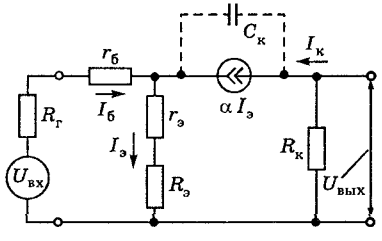


Рис. 9.6. Малосигнальная эквивалентная схема каскада в области высоких частот

Постоянная времени в области высоких частот. Эквивалентная схема усилителя в области высоких частот (малых времен) показана на рис. 9.6. Коллекторную емкость $C_к$ не будем учитывать отдельно: ее влияние отразим, как и в разд. 5.7, с помощью эквивалентной постоянной времени

$$\tau_{\alpha oe} = \tau_\alpha + C_к R_к. \qquad (9.11)$$

Для коэффициента α примем простейшее изображение (5.50) и подставим его в выражение (9.5а). Умножая числитель и знаменатель на двучлен $1 + s\tau_{\alpha oe}$ и приводя знаменатель к форме $A(1 + sa)$, можно записать изображение коэффициента усиления каскада в следующем виде:

$$K(s) = K/(1 + s\tau_в). \qquad (9.12)$$

Здесь $\tau_в$ — постоянная времени каскада в области высоких частот (малых времен). Ее легко представить в такой форме:

$$\tau_в = \tau_{\alpha oe}/(1 - \alpha\gamma_э), \qquad (9.13)$$

где $\gamma_э$ — коэффициент токораспределения во входной цепи:

$$\gamma_э = (R_г + r_б)/(R_э + r_э + R_г + r_б). \qquad (9.14)$$

Коэффициент $\gamma_э$ характеризует долю тока $\alpha I_э$, которая ответвляется из источника тока в цепь эмиттера в процессе нарастания коэффициента α (в первый момент $\alpha = 0$).

Если $R_г + r_б << R_э + r_э$, как до сих пор считалось, то $\gamma_э \approx 0$. Это значит, что ответвления тока $\alpha I_э$ в цепь эмиттера практически не происходит: почти весь ток $\alpha I_э$ поступает в низкоомную

цепь базы. Соответственно исходная ступенька эмиттерного тока $I_э(0)$ остается неизменной и переходный процесс развивается с постоянной времени $\tau_в = \tau_{\alpha oe}$. Об этом говорилось в начале раздела.

Заметное возрастание величины $\tau_в$ по сравнению с $\tau_{\alpha oe}$ наступает лишь при сравнимых значениях сопротивлений в эмиттерной и базовой цепях. В пределе, если $R_г + r_б >> R_э + r_э$, коэффициент $\gamma_э$ близок к единице. Тогда постоянная времени $\tau_в$ делается равной $\tau_{\alpha oe}/(1 - \alpha)$, а это есть эквивалентная постоянная времени схемы ОЭ [см. (5.69)]:

$$\tau_{oe} = \tau + (\beta + 1)\, C_к R_к. \qquad (9.15)$$

Такая трансформация постоянной времени $\tau_в$ вполне естественна, поскольку условие $\gamma_э = 1$ означает, что весь ток $\alpha I_э$ идет в цепь эмиттера, а следовательно, ток базы остается неизменным — заданным. При этом транзистор следует считать включенным по схеме ОЭ, а такому включению, как известно, свойственна постоянная времени τ_{oe}.

Заметим, что постоянную времени $\tau_в$ можно было с самого начала выразить через τ_{oe}. Для этого достаточно в формуле (9.13) заменить α на $\beta/(\beta + 1)$ и умножить числитель и знаменатель на $\beta + 1$. После элементарных преобразований, учитывая (5.57), получим:

$$\tau_в = \tau_{oe}/(1 + \beta\gamma_б). \qquad (9.16)$$

Здесь $\gamma_б$ — коэффициент токораспределения, характеризующий ответвление тока $\alpha I_э$ в базовую цепь:

$$\gamma_б = 1 - \gamma_э = \frac{R_э + r_э}{R_э + r_э + R_г + r_б}. \qquad (9.17)$$

С увеличением коэффициента $\gamma_б$ постоянная времени $\tau_в$ уменьшается и в пределе, при $\gamma_б = 1$, достигает максимального значения $\tau_{\alpha oe}$.

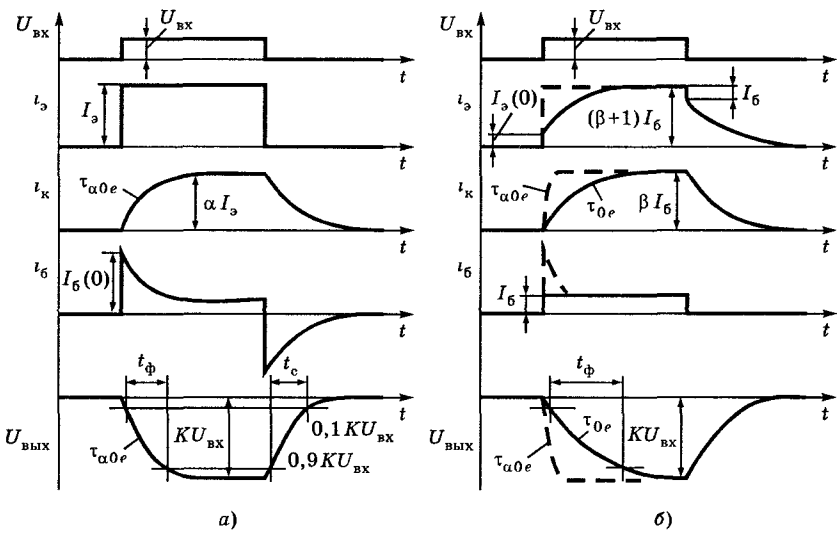
Из всего сказанного следует, что *быстродействие усилительного каскада повышается* (т.е. постоянная времени $\tau_в$ становится меньше) *при условии как можно меньшего сопротивления в базовой цепи и как можно большего — в эмиттерной*. Значит, чем меньше сопротивление источника сигнала, тем выше быстродействие усилителя.

Переходные и частотные характеристики. Умножая изображение коэффициента усиления (9.12) на $U_{вх}$, получаем изобра-

жение выходного напряжения. Оригиналом этого изображения будет переходная характеристика

$$U_{\text{вых}}(t) = KU_{\text{вх}}(1 - e^{-t/\tau_{\text{в}}}); \qquad (9.18)$$

она показана на рис 9.7 в двух вариантах: при низкоомном источнике сигнала, когда фактически задан ток эмиттера, т.е. $\gamma_{\text{э}} \approx 0$ и $\tau_{\text{в}} \approx \tau_{\alpha oe}$ (рис. 9.7, *а*), и при высокоомном источнике сигнала, когда задан ток базы, т.е. $\gamma_{\text{э}} \approx 1$ и $\tau_{\text{в}} \approx \tau_{oe}$ (рис. 9.7, *б*). Для наглядности установившиеся значения токов и выходного напряжения приняты одинаковыми, причем на рис. 9.7, *б* штрихами воспроизведены кривые, соответствующие заданному току эмиттера (рис. 9.7, *а*).



а) *б*)

Рис. 9.7. Переходные процессы в каскаде
а — при заданном токе эмиттера, *б* — при заданном токе базы

По окончании входного сигнала напряжение $U_{\text{вых}}$ падает до нуля с той же постоянной времени $\tau_{\text{в}}$. Поэтому длительности фронта и среза равны друг другу. Отсчитывая эти длительности между уровнями 0,1 и 0,9 от установившегося значения $KU_{\text{вх}}$, получаем:

$$t_{\text{ф}} = t_{\text{с}} = 2{,}2\ \tau_{\text{в}}. \qquad (9.19)$$

Пусть $\tau_{\alpha oe} = 5$ нс и $\gamma_э = 0{,}4$; тогда $\tau_в \approx 8$ нс и $t_ф = t_с \approx 17{,}5$ нс.

Частотные характеристики коэффициента усиления получаются путем замены оператора s на $j\omega$ в изображении (9.12):

$$K = \frac{K}{1 + j\omega/\omega_в}, \qquad (9.20)$$

где $\omega_в = 1/\tau_в$ — граничная угловая частота коэффициента усиления на уровне 0,7 (или 3 дБ).

Модуль и фаза комплексного коэффициента усиления K представляют собой соответственно амплитудно-частотную и фазо-частотную характеристики:

$$K(\omega) = \frac{K}{\sqrt{1 + (\omega/\omega_в)^2}}; \qquad (9.21а)$$

$$\varphi(\omega) = -\mathrm{arctg}(\omega/\omega_в). \qquad (9.21б)$$

Форма этих характеристик такая же, как у коэффициента α (см. рис. 5.21).

Пусть, как и в предыдущем примере, $\tau_{\alpha oe} = 5$ нс и $\gamma_э = 0{,}4$. Тогда $f_в = (1/2\pi)\,\omega_в \approx 20$ МГц.

Особо следует остановиться на входном базовом токе, который определяет входное сопротивление усилителя. В первый момент после подачи входного сигнала, когда $i_к = 0$, ток базы равен току эмиттера (рис. 9.7, *а*):

$$I_б(0) = I_э.$$

Ток $I_б(0)$ в $\beta + 1$ раз превышает установившееся значение. Соответственно начальное значение входного сопротивления в $\beta + 1$ раз меньше установившегося значения (9.7). В дальнейшем базовый ток спадает, а входное сопротивление возрастает вплоть до установившихся значений. Однако, поскольку установившееся значение тока базы в β раз меньше, чем тока коллектора, длительность переходного процесса оказывается во столько же раз больше, т.е. примерно $\beta t_ф$.

В заключение заметим, что с учетом сопротивления внешней нагрузки $R_н$ в выражениях для постоянных времени $\tau_{\alpha oe}$ и τ_{oe} вместо сопротивления $R_к$ должна стоять меньшая величина $R_к \parallel R_н$.

9.5. Простейшие усилители на МДП-транзисторах

Используются два схемных варианта таких усилителей — с резистивной и с динамической нагрузкой (рис. 9.8). В усилительных каскадах МДП-транзисторы всегда работают на пологих участках характеристик, где крутизна и коэффициент усиления транзистора имеют максимальные значения.

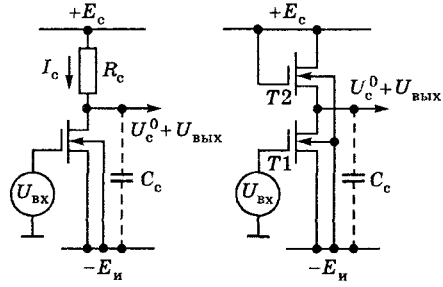


Рис. 9.8. Усилительные каскады на МДП-транзисторах:
а — с резистивной нагрузкой; *б* — с динамической нагрузкой

Усилитель с резистивной нагрузкой. В усилителе с резистивной нагрузкой (рис. 9.8, *а*) режим покоя характеризуется следующими потенциалами:

$$U_{и}^{0} = -E_{и}; \tag{9.22а}$$

$$U_{с}^{0} = E_{с} - I_{с}^{0} R_{с}. \tag{9.22б}$$

Для того чтобы транзистор был открыт, напряжение $U_{зи}^{0}$ должно превышать пороговое напряжение; значит, в данной схеме должно выполняться условие $E_{и} > U_{0}$. Потенциал $U_{с}^{0}$ удобно делать равным нулю. Это облегчает каскадирование усилителей: можно непосредственно соединять сток предыдущего каскада с затвором следующего.

Ток покоя $I_{с}^{0}$ легко записать, подставляя $U_{зи}^{0} = U_{з}^{0} - U_{н}^{0}$ в формулу (4.8):

$$I_{с}^{0} = 1/2\ b\,(E_{и} - U_{0})^{2}. \tag{9.23}$$

Отсюда, задаваясь током $I_{с}^{0}$, легко найти необходимое значение $E_{и}$.

Если напряжения питания $E_с$ и $E_и$ стабилизированы, то дрейф постоянных составляющих $I_с^0$ и $U_с^0$ обусловлен в первую очередь дрейфом параметров U_0 и b. Как известно, (см. с. 112), существует критическое значение тока $I_с$, при котором температурный дрейф минимален (в узком диапазоне температур он близок к нулю). При токах, больших критического, температурный коэффициент тока положительный, а при токах, меньших критического, — отрицательный.

Перейдем к оценке коэффициента усиления. Если положить дифференциальное сопротивление стока на пологом участке бесконечно большим ($r_с = \infty$), то из малосигнальной эквивалентной схемы на рис. 9.9 следует: $I_с = SU_{вх}$ и, следовательно,

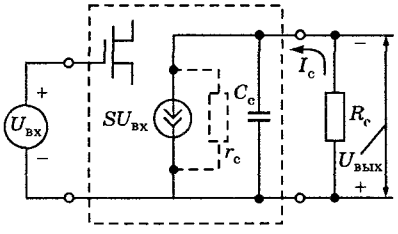


Рис. 9.9. Эквивалентная схема стоковой цепи усилительного каскада

$$U_{вых} = - I_с R_с = - SR_с U_{вх}.$$

Тогда коэффициент усиления будет иметь вид:

$$K = U_{вых}/U_{вх} = - SR_с. \qquad (9.24)$$

Если же сопротивление $r_с$ имеет конечную величину, сравнимую с $R_с$, то полный ток $SU_{вх}$ распределяется между ветвями $R_с$ и $r_с$. При этом ток стока оказывается равным

$$I_с = SU_{вх} r_с /(r_с + R_с).$$

Соответственно

$$U_{вых} = - I_с R_с = - S\,(r_с \parallel R_с)\, U_{вх}.$$

Тогда коэффициент усиления можно записать в следующем общем виде:

$$K = - \frac{\mu}{1 + r_с/R_с}, \qquad (9.25)$$

где $\mu = Sr_с$ — собственный коэффициент усиления транзистора [см. (4.18) и сноску на с. 110].

Из выражения (9.25) видно, что максимальный коэффициент усиления $|K| = \mu$ получается при условии $R_c >> r_c$. На практике это условие невыполнимо; получается большое падение напряжения $I_c^0 R_c$ и соответственно требуется большое напряжение питания E_c [см. (9.22б)]. Поэтому обычно $R_c \le (0{,}2–0{,}3)\, r_c$ и соответственно $|K| \le 0{,}2\, \mu$. При этом, как легко показать, можно пользоваться формулой (9.24).

Усилитель с динамической нагрузкой. В усилителе с динамической нагрузкой (рис. 9.8, *б*) нагрузочный транзистор $T2$ работает на пологом участке характеристик. Поэтому его сопротивление малым сигналам можно найти, дифференцируя ток I_{c2} по напряжению $U_{си2}$ в формуле (8.47). Учитывая (4.19а), получаем:

$$R_c = \mathrm{d}U_{c2}/\mathrm{d}I_{c2} = 1/S_2, \qquad (9.26)$$

где S_2 — крутизна транзистора $T2$. Внутреннее сопротивление r_{c2} при этом считается бесконечно большим; его учет обычно не существен.

Заменяя в выражении (9.25) сопротивление R_c на $1/S_2$, коэффициент μ на μ_1 и подставляя $r_c = \mu_1/S_1$, получаем:

$$K = -\frac{\mu_1}{\mu_1(S_2/S_1) + 1} \approx -\frac{S_1}{S_2}$$

(неравенство $\mu_1(S_2/S_1) >> 1$ обосновано ниже).

Поскольку токи обоих транзисторов одинаковы, отношение S_1/S_2 согласно (4.19б) запишется следующим образом:

$$\frac{S_1}{S_2} = \sqrt{\frac{b_1}{b_2}}.$$

Введем коэффициент B, который с учетом (4.7) характеризует геометрию транзисторов:

$$B = \frac{b_1}{b_2} = \frac{Z_1/L_1}{Z_2/L_2}. \qquad (9.27а)$$

При одинаковых длинах каналов

$$B = Z_1/Z_2 \qquad (9.27б)$$

Тогда коэффициент усиления можно записать в виде

$$K = -\sqrt{B}. \tag{9.28}$$

Следовательно, *коэффициент усиления определяется размерами каналов у активного и нагрузочного транзисторов*, прежде всего, — отношением ширин каналов. Отношение Z_1/Z_2 трудно сделать более 50–100, поэтому коэффициент усиления, как правило, составляет всего несколько единиц.

Заметим, что коэффициент усиления связан с режимом покоя усилителя. Действительно, приравнивая друг другу токи обоих транзисторов и используя коэффициент B, легко получить соотношение

$$\frac{U^0_{\text{зи2}} - U_0}{U^0_{\text{зи1}} - U_0} = \sqrt{B} = |K|$$

Подставим сюда значения $U^0_{\text{зи1}} = E_{\text{и}}$ и $U^0_{\text{зи2}} = E_{\text{с0}} - U^0_{\text{с}}$, которые очевидны из рис. 9.8, *б*. Тогда связь между коэффициентом усиления и режимом покоя можно представить в виде:

$$\frac{E_{\text{с}} - (U^0_{\text{с}} + U_0)}{E_{\text{и}} - U_0} = |K|. \tag{9.29}$$

Это выражение и по форме, и по существу аналогично выражению (9.6) для усилителя на биполярных транзисторах. Ясно, что напряжение $E_{\text{и}}$ должно быть существенно меньше $E_{\text{с}}$; однако оно должно заметно превышать U_0 во избежание нестабильности.

Переходные процессы. В усилителях на МДП-транзисторах переходные процессы связаны с перезарядом паразитной емкости $C_{\text{с}}$, подключенной к стоку активного транзистора (показана штриховой линией на рис. 9.8). Структура этой емкости та же, что и в МДП-транзисторных ключах [см. (8.52)].

Пусть на вход усилителя подан ступенчатый сигнал. При этом ток стока изменится практически мгновенно (с очень малой постоянной времени τ_S), а напряжение на стоке будет меняться экспоненциально, с постоянной времени $\tau_{\text{с}}$.

Операторное изображение коэффициента усиления нетрудно получить из (9.24), заменяя $R_{\text{с}}$ полным сопротивлением

$$Z_{\text{с}} = R_{\text{с}} \parallel \frac{1}{sC_{\text{с}}}.$$

После элементарных преобразований коэффициент усиления приводится к виду

$$K(s) = \frac{K}{1 + s\tau_c}, \qquad (9.30)$$

где $\tau_c = C_c R_c$. В случае динамической нагрузки под R_c понимается дифференциальное сопротивление $1/S_2$ (9.26).

Изображению (9.30) соответствует простейшая экспоненциальная функция, которая не раз встречалась выше [см., например, (9.18)]. Длительности фронта и среза определяются выражением типа (9.19).

Комплексный коэффициент усиления также имеет традиционную структуру, вытекающую из (9.30):

$$K = \frac{R}{1 + j\omega/\omega_c}, \qquad (9.31)$$

где $\omega_c = 1/\tau_c$.

Пусть, например, $R_c = 20$ кОм, $C_c = 3$ пФ; тогда $\tau_c = 60$ нс, $t_ф = 130$ нс и $f_c \approx 1{,}2$ МГц.

Эффект Миллера. Этот эффект упоминался в разд. 8.3 в связи с постоянной времени коллекторной емкости и в разд. 8.6 в связи со структурой суммарной емкости МДП-транзистора (8.52).

Сущность эффекта Миллера состоит в том, что эквивалентная входная проводимость активного четырехполюсника, обусловленная обратной связью, отличается от реальной проводимости, включенной в цепь обратной связи.

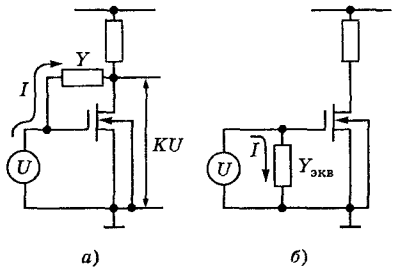


Рис. 9.10. Эффект Миллера:
а — схема с реальной проводимостью; *б* — схема с эквивалентной проводимостью

Рассмотрим конкретную схему (рис. 9.10, *а*), в которой между выходом (стоком) и входом (затвором) имеется комплексная проводимость Y. Подадим на вход переменную составляющую напряжения U. Тогда на выходе получится напряжение KU, где K — коэффициент усиления.

Таким образом, разность потенциалов на проводимости Y составит

$$U - KU = U\,(1 - K).$$

Эта разность потенциалов вызовет протекание тока

$$I = U\,(1 - K)\,Y.$$

Поскольку ток отбирается от источника входного сигнала, отношение I/U есть *эквивалентная входная проводимость* $Y_{экв}$ (рис. 9.10, *б*). Ее значение $Y_{экв} = Y\,(1 - K)$.

В рассматриваемой схеме и в большинстве других случаев коэффициент усиления отрицательный. Тогда удобнее использовать запись:

$$Y_{экв} = (|K| + 1)\,Y.$$

Как видим, при условии $|K| > 1$ эквивалентная проводимость может намного превышать реальную.

На практике проводимость Y чаще всего представляет собой емкость (в данном случае это емкость $C_{зс}$). Поэтому обычно эффект Миллера ассоциируется с кажущимся увеличением входной емкости:

$$C_{экв} = (|K| + 1)\,C \qquad (9.32)$$

[в формуле (8.52) величина $|K| + 1$ была обозначена через K]. Однако, если коэффициент усиления схемы положителен, но меньше единицы (например у повторителя, см. разд. 9.7), то эквивалентная входная емкость определяется величиной $(1 - K)\,C$ и оказывается значительно меньше реальной емкости между входом и выходом.

Подчеркнем, что эффект Миллера имеет общее значение: он не ограничен однокаскадными схемами и типом активных приборов.

9.6. Дифференциальные усилители

Схема дифференциального усилителя (ДУ) показана на рис. 9.11. Он состоит из двух одинаковых (симметричных) плеч, каждое из которых содержит транзистор и резистор. В общей эмиттерной цепи действует источник тока I_0. Выходным напряжением является разность коллекторных потенциалов, а входным — разность базовых потенциалов.

В целом структура ДУ такая же, как у переключателя тока (см. рис. 8.13), но режим работы другой: ни один из транзисторов не заперт, оба они работают в активном режиме. Использование источника тока I_0 обеспечивает стабильность рабочей точки — токов $I_э^0$ и напряжений $U_к^0$.

Принцип действия. В основе ДУ лежит идеальная симметрия обоих его плеч, т.е. идентичность параметров транзисторов $T1$, $T2$ и равенство сопротивлений $R_{к1}$, $R_{к2}$. При этом в отсутствие сигнала токи и коллекторные потенциалы будут одинаковы, а выходное напряжение будет равно нулю. В силу симметрии нулевое значение $U_{вых}$ сохраняется при одновременном и одинаковом изменении токов в обоих плечах, какими бы причинами такое изменение ни вызывалось. Следовательно, *в идеальном ДУ дрейф выходного напряжения отсутствует*, хотя в каждом из плеч он может быть сравнительно большим.

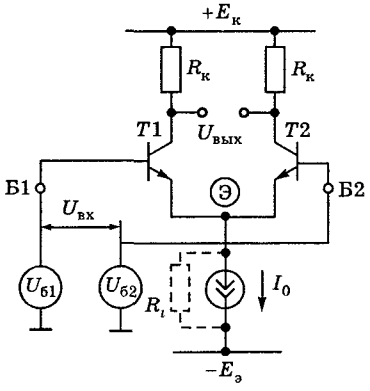


Рис. 9.11. Дифференциальный усилитель

Подадим на базы одинаковые напряжения ($\Delta U_{б1} = \Delta U_{б2}$). Такие сигналы называют синфазными. Под действием синфазных сигналов потенциал эмиттеров изменится на такую же величину, как и потенциалы баз: $\Delta U_э = \Delta U_б$ (поскольку напряжения на эмиттерных переходах U^* можно считать неизменными). Если источник тока I_0 идеальный (т. е. $R_i = \infty$), то приращение $\Delta U_э$ не вызовет изменения токов в ветвях ДУ. Коллекторные потенциалы не изменятся и выходное напряжение останется равным нулю. Если же $R_i \neq \infty$, то появится приращение тока ΔI_0, но оно поровну распределится между обеими ветвями ДУ и коллекторные потенциалы изменятся одинаково. Соответственно и в этом случае $U_{вых} = 0$. Значит, *в идеальном ДУ синфазные сигналы не влияют на выходное напряжение.*

Теперь подадим на базы напряжения равной величины, но противоположных знаков ($\Delta U_{б1} = -\Delta U_{б2}$). Такие сигналы называют *дифференциальными*. Их разность, по определению, является входным сигналом ДУ:

$$U_{вх} = \Delta U_{б1} - \Delta U_{б2}.$$

В силу симметрии сигнал $U_{вх}$ поделится поровну между обоими эмиттерными переходами: на одном из них напряжение U^* увеличится на $\frac{1}{2} U_{вх}$, а на другом уменьшится на ту же величину. Соответственно приращения токов и коллекторных потенциалов в плечах ДУ будут одинаковыми по величине, но разного знака. В результате появится выходное напряжение

$$U_{вых} = \Delta U_{к1} - \Delta U_{к2}.$$

Как видим, *идеальный ДУ реагирует только на дифференциальный сигнал*, отсюда — название этого типа усилителей.

Поскольку дифференциальный сигнал делится поровну между эмиттерными переходами, потенциал средней точки, т.е. потенциал эмиттеров, остается неизменным. Следовательно, *при анализе дифференциальных сигналов можно считать потенциал* $U_э$ заданным, *а точку Э заземленной для переменных составляющих.*

Любую комбинацию напряжений $\Delta U_{б1}$ и $\Delta U_{б2}$ можно представить в виде суммы синфазной и дифференциальной составляющих (рис. 9.12, а):

$$\Delta U_{б1} = U_{вх\,с} + \tfrac{1}{2} U_{вх\,д}; \qquad (9.33а)$$

$$\Delta U_{б2} = U_{вх\,с} - \tfrac{1}{2} U_{вх\,д}. \qquad (9.33б)$$

Удобство такого представления — в том, что действие каждой из составляющих можно анализировать раздельно.

Синфазная и дифференциальная составляющие входного сигнала согласно (9.33) выражаются следующим образом:

$$U_{вх\,с} = \tfrac{1}{2} (\Delta U_{б1} + \Delta U_{б2}); \qquad (9.34а)$$

$$U_{вх\,д} = \Delta U_{б1} - \Delta U_{б2}. \qquad (9.34б)$$

Выходное напряжение тоже можно представить как сумму синфазной и дифференциальной составляющих (рис. 9.12, *б*):

$$U_{вых\,с} = \tfrac{1}{2} (\Delta U_{к1} + \Delta U_{к2}); \qquad (9.35а)$$

$$U_{вых\,д} = \Delta U_{к1} - \Delta U_{к2}, \qquad (9.35б)$$

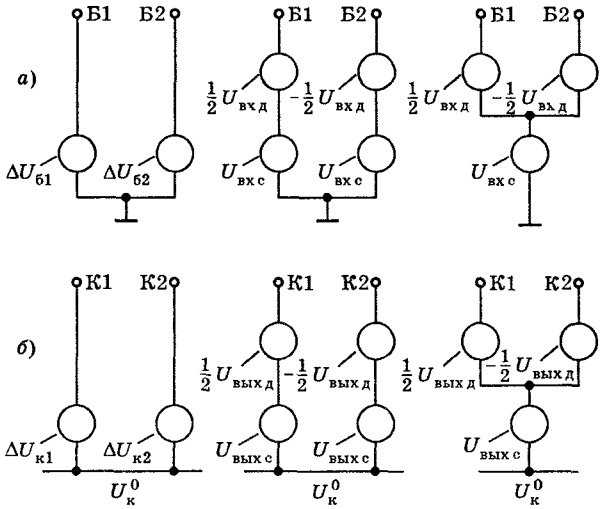


Рис. 9.12. Синфазные и дифференциальные составляющие входного (*а*) и выходного (*б*) сигналов

где $\Delta U_к$ — приращения коллекторных потенциалов относительно потенциала покоя $U_к^0$. На рис. 9.11 дифференциальные составляющие $U_{вх\,д}$ и $U_{вых\,д}$ записаны без индекса «д».

Важную роль в работе ДУ играет постоянство тока I_0. Если источник тока идеальный (т.е. $R_i = \infty$, см. рис. 9.11) то синфазная составляющая сигнала вызывает только приращение эмиттерного потенциала: $\Delta U_э = \Delta U_б = U_{вх\,с}$. Токи в плечах и коллекторные потенциалы остаются неизменными.

Если же источник тока не идеален, т.е. имеет конечное сопротивление R_i, то приращение $\Delta U_э$ вызывает приращение тока $\Delta I_0 = \Delta U_э / R_i$. Это приращение делится между обоими плечами ДУ и вызывает приращения коллекторных потенциалов $\Delta U_{к1}$ и $\Delta U_{к2}$. В том случае, когда плечи идентичны, эти приращения одинаковы: $\Delta U_{к1} = \Delta U_{к2}$; тогда на выходе согласно (9.35) получается только синфазная составляющая. В том случае, когда плечи неидентичны, приращения коллекторных потенциалов неодинаковы: $\Delta U_{к1} \neq \Delta U_{к2}$; тогда на выходе наряду с синфазной получается паразитная дифференциальная составляющая.

Поскольку в идеальном ДУ синфазный сигнал на входе не должен вызывать синфазного и тем более дифференциального сигнала на выходе, к внутреннему сопротивлению источника тока предъявляют самые высокие требования.

Тот факт, что работа ДУ зиждется на идентичности его плеч, объясняет популярность этих усилителей (и основанных на них схем) в микроэлектронике. Только в ИС, где элементы расположены друг от друга на расстояниях десятков микрометров, можно обеспечить идентичность параметров, температурных коэффициентов и т.п. Кроме того, в микроэлектронике не критично количество элементов, способствующих повышению качества схемы.

Коэффициенты усиления. В реальном ДУ, в котором оба плеча неидентичны, а источник тока имеет конечное сопротивление, имеют место влияние синфазной составляющей входного сигнала на дифференциальную составляющую выходного сигнала и влияние дифференциальной составляющей входного сигнала на синфазную составляющую выходного сигнала.

В общем случае соотношение между синфазными и дифференциальными составляющими можно записать с помощью двух уравнений:

$$U_{\text{вых с}} = K_{\text{сс}}U_{\text{вх с}} + K_{\text{сд}}U_{\text{вх д}}; \qquad (9.36\text{а})$$

$$U_{\text{вых д}} = K_{\text{дс}}U_{\text{вх с}} + K_{\text{дд}}U_{\text{вх д}}. \qquad (9.36\text{б})$$

Здесь коэффициенты K являются коэффициентами передачи соответствующих составляющих со входа на выход. В идеальном ДУ взаимные коэффициенты $K_{\text{сд}}$ и $K_{\text{дс}}$ равны нулю.

Рассмотрим главный параметр ДУ — *коэффициент усиления дифференциальной составляющей* $K_{\text{дд}}$. Его часто называют просто *коэффициентом усиления* и обозначают через K.

Как уже подчеркивалось в предыдущем разделе, потенциал эмиттера при подаче дифференциального сигнала остается неизменным, а значит, для переменных составляющих его следует считать равным нулю. Поэтому коэффициент усиления каждого плеча можно получить из (9.5а), полагая $R_{\text{э}} = 0$. Поскольку в каждом плече усиливается сигнал $\frac{1}{2}U_{\text{вх}}$, а на выходе усиленные сигналы складываются, коэффициент усиления ДУ равен коэффициенту усиления отдельного плеча.

Полагая $R_{\text{э}} = 0$ в выражении (9.5а), получаем:

$$K = \frac{\alpha R_{\text{к}}}{r_{\text{э}} + (1-\alpha)(R_{\text{г}} + r_{\text{б}})}. \qquad (9.37\text{а})$$

Очевидно, что коэффициент усиления ДУ значительно больше, чем у простейшего усилителя. Разница может составлять

десятки раз. Следовательно, помимо отсутствия (или гораздо меньшего) дрейфа, *ДУ свойствен гораздо больший коэффициент усиления*, что является его вторым важным преимуществом.

В случае низкоомных источников сигнала ($R_г$ менее 1 кОм) и небольших рабочих токов (менее 1 мА) вторым слагаемым в знаменателе (9.37а) можно пренебречь; тогда

$$K = -\alpha\,(R_к/r_э). \qquad (9.37б)$$

Подставляя сюда значение $R_к$ из (9.2б) и значение $r_э = \varphi_T/I_э^0$ из (5.43), получаем коэффициент усиления в следующей форме:

$$K = -(E_к - U_к^0)/\varphi_T. \qquad (9.38)$$

Пусть, например, $E_к = 12$ В и $U_к^0 = 2$ В, тогда $K = -400$. Как видим, коэффициент усиления ДУ, как и у простейших усилителей, связан с напряжением питания и коллекторным напряжением в режиме покоя [ср. с (9.6) и (9.29)]. Однако знаменатель в выражении (9.38) имеет такое малое значение, которое недостижимо в простейших усилителях по соображениям стабильности. Как и в других рассмотренных схемах, коэффициент усиления (при заданном $U_к^0$) не зависит от рабочего тока. От температуры он зависит непосредственно через величину φ_T.

Коэффициент усиления синфазной составляющей согласно (9.36) определяется как

$$K_{сс} = \frac{U_{вых\ с}}{U_{вх\ с}}\Bigg|_{U_{вх\ д}=0}.$$

Следовательно, на рис. 9.11 нужно мысленно соединить обе базы и подать на них сигнал $U_{вх\ с}$. Полагая $U^* = \text{const}$, получаем $\Delta U_э = U_{вх\ с}$. При этом, если сопротивление источника тока равно R_i, ток I_0 изменяется на величину $\Delta I_0 = U_{вх\ с}/R_i$, а коллекторные потенциалы — на величину $-\alpha(\frac{1}{2}\Delta I_0)R_к$. Тогда

$$K_{сс} = -(\alpha R_к)/2R_i. \qquad (9.39)$$

Обычно $R_к/R_i < 1$, а значит, и $K_{сс} < 1$.

Коэффициент $K_{сд}$ согласно (9.36) характеризует влияние дифференциальной составляющей сигнала на синфазную составляющую выходного напряжения:

$$K_{сд} = \frac{U_{вых\ с}}{U_{вх\ д}}\Bigg|_{U_{вх\ с}=0}.$$

Поскольку дифференциальный сигнал распределяется между обоими эмиттерными переходами поровну, главной причиной изменения среднего коллекторного потенциала является неидентичность коэффициентов усиления плеч. Поэтому можно считать

$$K_{сд} \approx \Delta K,$$

где $\Delta K = K_1 - K_2$. Умножим и поделим правую часть на средний коэффициент усиления

$$K = \tfrac{1}{2}(K_1 + K_2).$$

Далее, считая, что главной причиной неидентичности коэффициентов усиления является различие сопротивлений $R_к$, положим $\Delta K/K = \Delta R_к/R_к$. Тогда

$$K_{сд} = K(\Delta R_к/R_к). \qquad (9.40)$$

Например, если $\Delta R_к/R_к = 0{,}02$, то $K_{сд} = 0{,}02$ К. Значит, *изменение постоянной составляющей коллекторных потенциалов под действием дифференциального сигнала в десятки раз меньше выходного напряжения.*

Коэффициент подавления синфазной составляющей. Коэффициент $K_{дс}$ согласно (9.36) характеризует влияние синфазной составляющей входного сигнала на дифференциальную составляющую выходного сигнала. Это влияние весьма существенно, так как слагаемое $K_{дс}U_{вх\ с}$ неотличимо от слагаемого $K_{дд}U_{вх\ д}$, т.е. равносильно ложному сигналу. Поскольку на практике составляющая $U_{вх\ с}$ может в тысячи раз превышать составляющую $U_{вх\ д}$, то значение $K_{дс}$ должно быть меньше $K_{дд}$ на несколько порядков. Соотношение модулей этих двух величин принято характеризовать *коэффициентом подавления синфазной составляющей*, выраженным в децибелах[1]:

$$K_п = 20 \lg \left|\frac{K_{дд}}{K_{дс}}\right|. \qquad (9.41)$$

---

1 Английский термин. Common Mode Rejection Ratio (CMRR). Официальное обозначение этого коэффициента. $K_{осл\ сф}$ (коэффициент ослабления синфазной составляющей).

Например, если $|K_{дд}/K_{дс}| = 10^4$, то $K_п = 80$ дБ.

Распространенным типом синфазной составляющей являются различные (внутренние и внешние) помехи и наводки, действующие одновременно на оба входа. Поэтому *увеличение коэффициента $K_п$ — один из главных путей повышения помехоустойчивости ДУ.*

Для оценки коэффициента подавления предположим, что сопротивление источника тока, как и выше, равно R_i, а неидентичность плеч ДУ определяется соотношениями

$$\alpha_1 = \alpha + \Delta\alpha; \quad \alpha_2 = \alpha - \Delta\alpha;$$

$$R_{к1} = R_к + \Delta R_к; \quad R_{к2} = R_к - \Delta R_к,$$

где α и $R_к$ — средние значения.

Как уже отмечалось, приращение тока I_0 под действием синфазного сигнала составляет $U_{вх\,с}/R_i$. Пусть это приращение распределяется поровну между эмиттерными переходами. Тогда изменения коллекторных потенциалов можно записать следующим образом:

$$\Delta U_{к1} = -\alpha_1(U_{вх\,с}/2R_i)R_{к1};$$

$$\Delta U_{к2} = -\alpha_2(U_{вх\,с}/2R_i)R_{к2}.$$

Приравнивая разность этих приращений первому слагаемому в правой части (9.36б) и подставляя приведенные выше значения α_1, α_2, $R_{к1}$ и $R_{к2}$, нетрудно записать

$$K_{дс} = -\alpha\frac{R_к}{R_i}\left(\frac{\Delta\alpha}{\alpha} + \frac{\Delta R_к}{R_к}\right).$$

Коэффициент подавления в линейной (не логарифмической) форме найдем, поделив коэффициент усиления $K_{дд}$ на коэффициент $K_{дс}$. Используя (9.37б), представим коэффициент подавления в следующей общей форме:

$$K_п = \frac{1}{\delta}\frac{R_i}{r_э}. \qquad (9.42)$$

Здесь δ — *коэффициент асимметрии* ДУ, т.е. сумма относительных разбросов параметров в его плечах:

$$\delta = \frac{\Delta\alpha}{\alpha} + \frac{\Delta R_к}{R_к}.$$

В случае необходимости эту сумму можно дополнить разбросом других параметров транзисторов.

При выводе выражения (9.42) знаки приращений $\Delta\alpha$ и $\Delta R_к$ были положительными, так как считалось, что $R_{к1} > R_{к2}$ и $\alpha_1 > \alpha_2$. Соответственно коэффициент асимметрии оказался арифметической суммой относительных приращений. На практике оба приращения независимы и неконтролируемы, т.е. могут иметь разные величины и знаки. Ориентируясь на наихудший случай, коэффициент асимметрии рассчитывают как сумму модулей максимально возможных или наиболее вероятных относительных приращений. При этом в реальном ДУ коэффициент δ может быть меньше, а коэффициент $K_п$ больше расчетного значения.

Интегральные ДУ, в связи с меньшим коэффициентом асимметрии, обеспечивают более высокие коэффициенты подавления, чем ДУ, выполненные на дискретных транзисторах.

Из выражения (9.42) следует важный вывод: *коэффициент подавления находится в прямой зависимости от сопротивления источника тока* R_i. Следовательно, это сопротивление должно быть как можно больше. Простейшие — резисторные источники тока (см. рис. 9.15) в ДУ непригодны.

Входное сопротивление. Различают входные сопротивления ДУ для дифференциальной и синфазной составляющих сигнала. Эти сопротивления существенно различны.

Для дифференциальной составляющей входное сопротивление равно удвоенному входному сопротивлению каждой половины ДУ. Используя выражение (9.7а) при $R_э = 0$, получаем:

$$R_{вх\,д} = 2\,[(\beta + 1)\,(r_э + r_б)]. \qquad (9.43)$$

Пусть, например, $\beta = 100$, $r_э = 25$ Ом и $r_б = 150$ Ом; тогда $R_{вх\,д} = 5{,}35$ кОм. Сопротивление $r_э$ обратно пропорционально току покоя $I_э^0$. Поэтому для увеличения входного сопротивления целесообразно использовать ДУ в режиме малых токов — в микрорежиме. Кроме того, целесообразно использовать транзисторы с высокими значениями β, например, пары Дарлингтона (см. разд. 9.2). Так, если $I_э^0 = 50$ мкА и $\beta = 2000$, то $r_э = 0{,}5$ кОм и $R_{вх\,д} \approx 2$ МОм.

Для синфазной составляющей входное сопротивление определяется сопротивлением источника тока R_i. Полагая $\Delta U_э = U_{вх\,с}$, получаем приращение тока $\Delta I_0 = U_{вх\,с}/R_i$. Соответственно

$\Delta I_э = \frac{1}{2}\Delta I_0$ и $\Delta I_б = \frac{1}{2}(1-\alpha)\Delta I_0$. Деля $U_{вх\,с}$ на $2\Delta I_б$ и переходя от коэффициента α к β, находим синфазное входное сопротивление:

$$R_{вх\,с} = (\beta + 1)\, R_t. \qquad (9.44)$$

Поскольку $R_t >> r_э$, сопротивление $R_{вх\,с}$ намного превышает $R_{вх\,д}$.

Динамический диапазон. Под динамическим диапазоном понимается выраженное в децибелах отношение максимального и минимального входных сигналов. Минимальный сигнал ограничивается собственными шумами, а максимальный — нелинейными искажениями, т.е. искажениями формы сигнала. Приближенно можно оценить максимально допустимый сигнал, пользуясь критериями отсечки или насыщения транзистора.

Пусть в режиме покоя $U_к^0 = \frac{1}{2}E_к$. При положительной полярности входного сигнала потенциал $U_к$ уменьшается вплоть до 0 (после чего наступает насыщение транзистора), а при отрицательной полярности — увеличивается вплоть до $E_к$ (после чего наступает запирание транзистора). Таким образом, в обоих случаях максимальное приращение $\Delta U_к$ составляет $\frac{1}{2}E_к$. Деля эту величину на коэффициент усиления в форме (9.38), получаем максимально допустимый входной сигнал:

$$U_{вх\,макс} = \varphi_T. \qquad (9.45)$$

Практически сигналы величиной φ_T неприемлемы, так как при этом на краях диапазона существенно меняется ток эмиттера, а вместе с ним сопротивление $r_э$ и коэффициент усиления. В результате возникают большие нелинейные искажения. *Чтобы искажения сигнала были невелики, амплитуды сигналов должны лежать в пределах* 0,5 φ_T.

Синфазные сигналы могут иметь гораздо большие значения, чем дифференциальные, поскольку коэффициент $K_{сс}$ значительно меньше, чем $K_{дд}$. Запишем соотношение между потенциалом коллектора и синфазным сигналом, полагая $U_к^0 = \frac{1}{2}E_к$:

$$U_к = \tfrac{1}{2}E_к + K_{сс}U_{вх\,с}.$$

Подставляя в левую часть $U_к = U_{вх\ с}$ (условие насыщения) или $U_к = E_к$ (условие отсечки), получаем соответственно максимальные положительный и отрицательный синфазные сигналы:

$$U^{+}_{вх\ с\ макс} = E_к/(1 - K_{сс}); \tag{9.46а}$$

$$U^{-}_{вх\ с\ макс} = E_к/2K_{сс}, \tag{9.46б}$$

где $K_{сс} < 0$ [см. (9.39)]. Поскольку обычно $|K_{сс}| < 1$, легко видеть, что синфазные входные сигналы могут составлять несколько вольт, вплоть до напряжений, близких к $E_к$.

Точностные параметры. К точностным параметрам ДУ относятся паразитные напряжения и токи, которые имеют место в режиме покоя, но оказывают влияние на выходной сигнал в процессе усиления переменной составляющей.

Неизбежная асимметрия плеч в реальных ДУ является причиной того, что в режиме покоя на выходе имеется конечная разность — разбаланс потенциалов: $U^0_{к1} - U^0_{к2}$. Такому разбалансу соответствует на входе некоторый мнимый дифференциальный сигнал

$$U_{см} = (U^0_{к1} - U^0_{к2})/K, \tag{9.47}$$

который называют *напряжением смещения нуля*. Чтобы устранить разбаланс выходных потенциалов, нужно подать на вход дифференциальный сигнал величиной $U_{см}$, но обратного знака.

Напряжение смещения состоит из нескольких слагаемых, каждое из которых зависит от разброса соответствующих величин: эмиттерных токов $I_э$, коллекторных сопротивлений $R_к$ и др.

Разброс эмиттерных токов (при одинаковых напряжениях $U_э$) обусловлен разбросом тепловых токов в эмиттерных переходах [см. (5.38б)]: в транзисторе с меньшим током $I_{э0}$ будет меньше и ток $I_э$. Для того, чтобы выровнять эмиттерные токи, на вход ДУ следует подать «выравнивающий» дифференциальный сигнал $U_{см1}$. Этот сигнал должен иметь такую полярность и такую величину, чтобы напряжение $U_э$ в транзисторе с меньшим током $I_{э0}$ увеличилось, а в транзисторе с большим током $I_{э0}$ уменьшилось и чтобы при этом эмиттерные токи стали одинаковыми:

$$U_{э1} = U_э + \tfrac{1}{2}U_{см1} = \varphi_T \ln (I_э/I_{э01});$$

$$U_{э2} = U_э - \tfrac{1}{2}U_{см1} = \varphi_T \ln (I_э/I_{э02}).$$

Вычитая второе равенство из первого, получаем:

$$U_{см1} = \varphi_T \ln (I_{э02}/I_{э01}). \qquad (9.48)$$

Например, если тепловые токи различаются на 20%, то $U_{см1} \approx 5$ мВ.

Вторая по важности составляющая напряжения смещения обусловлена разбросом сопротивлений $R_к$. Пусть токи в плечах одинаковы. Тогда разность коллекторных потенциалов в режиме покоя составит:

$$U^0_{к1} - U^0_{к2} = \alpha I_0 \Delta R_к .$$

Деля эту разность на коэффициент усиления (9.37б) и подставляя $r_э = \varphi_T(\frac{1}{2} I_0)$, получаем вторую составляющую напряжения смещения в виде:

$$U_{см2} = 2\ \varphi_T(\Delta R_к/R_к). \qquad (9.49)$$

Например, при $\Delta R_к/R_к = 0{,}02$ получаем $U_{см2} \approx 1$ мВ. Меньшее значение этой составляющей по сравнению с первой [см. пример к (9.48)] типично для ДУ. Другие составляющие, связанные с разбросом коэффициента α, сопротивления $r_к$ и т.п., еще менее существенны.

Необходимо иметь в виду, что напряжение смещения зависит от температуры. Эта зависимость характеризуется температурной чувствительностью ε_U, которая обычно измеряется в мкВ/°С.

Для главной составляющей напряжения смещения $U_{см1}$, обусловленной разбалансом эмиттерных токов, температурную чувствительность можно оценить как разность температурных чувствительностей напряжений $U_{э1}$ и $U_{э2}$ при выровненных эмиттерных токах. Известно, что температурная чувствительнось при заданном токе пропорциональна разности величин ширины запрещенной зоны и прямого смещения и обратно пропорциональна температуре. Откуда нетрудно получить:

$$\varepsilon_U = \varepsilon_1 - \varepsilon_2 = U_{см1}/T. \qquad (9.50)$$

Например, если $U_{см1} = 5$ мВ и $T = 300$ К, то $\varepsilon_U \approx 17$ мкВ/°С.

Из (9.50) видно, что температурная чувствительность ε_U уменьшается вместе с уменьшением напряжения смещения.

Однако при $U_{см1} < 1$ мВ такая пропорциональность нарушается, так как главной составляющей напряжения смещения становится величина $U_{см2}$, обусловленная разбалансом сопротивлений $R_к$ [см. (9.49)]. Температурная чувствительность напряжения $U_{см2}$ определяется температурным коэффициентом сопротивлений $R_к$, т.е. зависит от типа и структуры резисторов.

Помимо начального разбаланса коллекторных потенциалов, имеет место также начальный разбаланс входных (базовых) токов $\Delta I_{вх}$. Этот параметр называют *током смещения* или просто *разностью входных токов*.

В идеальном ДУ токи в обоих плечах одинаковы, т.е. $\Delta I_{вх} = 0$. В реальном ДУ эту разность можно записать в виде

$$\Delta I_{вх} = \frac{I_{э1}}{B_1 + 1} - \frac{I_{э2}}{B_2 + 1} \approx \frac{I_{э1}}{B_1} - \frac{I_{э2}}{B_2}.$$

Используя наиболее неблагоприятные соотношения ($I_{э1} > I_{э2}$ и $B_1 < B_2$), нетрудно получить:

$$\Delta I_{вх} = \frac{I_0}{B}\left(\frac{\Delta I_э}{I_э} + \frac{\Delta B}{B}\right), \qquad (9.51)$$

где $I_э$ и B — средние значения ($I_э = \frac{1}{2} I_0$). Как видим, ток смещения уменьшается с уменьшением рабочего тока ДУ и с увеличением коэффициента усиления B.

Значимость параметра $\Delta I_{вх}$ состоит в том, что ток смещения, протекая через сопротивление источника дифференциального сигнала $R_{г\,д}$, дает на нем падение напряжения $\Delta I_{вх} R_{г\,д}$. Это напряжение равносильно напряжению смещения, т.е. ложному сигналу. Например, если $\Delta I_{вх} = 20$ нА и $R_{г\,д} = 100$ кОм, то $\Delta I_{вх} R_{г\,д} = 2$ мВ.

Важную роль в ДУ играет еще один точностной параметр — *средний входной ток*:

$$I_{вх\,ср} = \tfrac{1}{2}\,(I_{б1} + I_{б2}).$$

При оценке этого параметра можно использовать средние значения базовых токов $I_{б1} = I_{б2} \approx I_э/B$. Тогда

$$I_{вх\,ср} = I_0/2B. \qquad (9.52)$$

Пути для уменьшения тока $I_{вх\,ср}$ те же, что и для $\Delta I_{вх}$.

Из сравнения (9.51) и (9.52) следует, что ток смещения значительно меньше среднего тока (обычно на порядок).

Значимость тока $I_{\text{вх ср}}$ состоит в том, что он, протекая через сопротивление источника синфазного сигнала $R_{\text{г с}}$, дает на нем падение напряжения, равносильное синфазному сигналу. Будучи умножен на коэффициент $K_{\text{дс}}$, этот сигнал вызывает на выходе начальный разбаланс потенциалов.

И средний входной ток, и разность входных токов зависят от температуры, т.е. характеризуются своими температурными чувствительностями. Из выражений (9.51) и (9.52) ясно, что эти температурные чувствительности связаны прежде всего с зависимостью B (T). Температурные чувствительности токовых параметров пропорциональны самим параметрам, так что уменьшение последних влечет за собой повышение температурной стабильности.

Переходные процессы. Характер переходных процессов в ДУ такой же, как в простейшем каскаде (рис. 9.7). Остается в силе и анализ, выполненный в разделе 9.4. Однако количественные параметры — постоянная времени и длительность фронта — оказываются хуже, чем в простейшем каскаде.

Запишем постоянную времени $\tau_{\text{в}}$ в форме (9.16):

$$\tau_{\text{в}} = \tau_{oe}/(1 + \beta\gamma_{\text{б}}). \qquad (9.53)$$

В выражении (9.17) для коэффициента токораспределения $\gamma_{\text{б}}$ положим $R_{\text{э}} = 0$; тогда

$$\gamma_{\text{б}} = r_{\text{э}}/(r_{\text{э}} + R_{\text{г}} + r_{\text{б}}). \qquad (9.54)$$

Как видим, при прочих равных условиях коэффициент $\gamma_{\text{б}}$ в ДУ меньше, чем в простейшем усилителе, а следовательно, постоянная времени $\tau_{\text{в}}$ больше. Например, если $r_{\text{э}} = 25$ Ом, $r_{\text{б}} = 150$ Ом, $\beta = 100$ и $R_{\text{г}} = 1$ кОм и если в простейшем каскаде $R_{\text{э}} = 2$ кОм, то быстродействие ДУ оказывается в 20 раз ниже, чем у простейшего каскада.

Если уменьшать рабочий ток, то сопротивление $r_{\text{э}}$ согласно (5.43) возрастает и вместе с ним возрастает коэффициент $\gamma_{\text{б}}$. В пределе, при $\gamma_{\text{б}} \approx 1$, постоянная времени $\tau_{\text{в}}$ будет в $\beta + 1$ раз меньше $\tau_{\text{ое}}$, т.е. превращается в эквивалентную постоянную времени $\tau_{\alpha oe}$ [см. (9.11)]. Однако это отнюдь не означает существенного изменения быстродействия, поскольку с уменьшением

тока согласно (9.2б) должно увеличиваться сопротивление $R_к$, а оно определяет величину $\tau_{дое}$.

Реальный путь для повышения быстродействия — тот, который был сформулирован в разделе 9.4: уменьшение сопротивления источника сигнала и, в первую очередь, улучшение высокочастотных параметров транзистора.

Дифференциальный усилитель на МДП-транзисторах. Схема простейшего МДП-транзисторного ДУ показана на рис. 9.13. Плечами этого ДУ служат простейшие каскады с динамической нагрузкой (см. рис. 9.8, *б*).

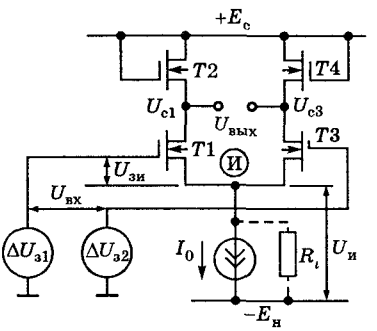


Рис. 9.13. Дифференциальный усилитель на МДП транзисторах

Поскольку дифференциальный сигнал $U_{вх}$ делится поровну между участками затвор–исток активных транзисторов ($T1$ и $T3$), можно считать, что коэффициент усиления ДУ такой же, как у отдельного плеча:

$$K = -S_1/S_2 = \sqrt{B}, \qquad (9.55)$$

где параметр $B = b_1/b_2$ характеризует геометрию транзисторов $T1$ и $T2$ [см. (9.27)]. Как и в простейшем каскаде, коэффициент усиления ограничен и обычно не превышает значений 5–7, что намного меньше, чем у биполярных ДУ.

Коэффициент усиления синфазной составляющей определим с помощью следующих соотношений. Синфазный сигнал вызывает приращения напряжения на общем истоке и на участках затвор–исток (рис. 9.13):

$$U_{вх\,с} = \Delta U_{зи} + \Delta U_и;$$

напряжение $\Delta U_{эи}$ вызывает приращение тока $I_с$:

$$\Delta I_с = S\Delta U_{зи},$$

а напряжение $\Delta U_и$ — приращение тока I_0:

$$\Delta I_0 = \frac{\Delta U_и}{R_i} = \frac{U_{вх\,с} - \Delta U_{зи}}{R_i}.$$

Приращение ΔI_0 распределяется поровну между обеими ветвями ДУ; следовательно, $\Delta I_c = \frac{1}{2}\Delta I_0$. Из этого равенства, подставляя приведенные выше значения токов, получаем:

$$\Delta U_{зи} = U_{вх\,с}/(2SR_1 + 1). \qquad (9.56)$$

Умножая $\Delta U_{зи}$ на K и деля на $U_{вх\,с}$, находим коэффициент усиления синфазной составляющей:

$$K_{сс} = K/(2SR_1 + 1). \qquad (9.57)$$

Очевидно, что коэффициент $K_{сс}$ меньше, чем K (обычно в десятки раз) и составляет доли единицы.

Коэффициент усиления $K_{дс}$, лежащий в основе коэффициента подавления синфазной составляющей, получается следующим образом. Пусть плечи ДУ несимметричны, так что $K_1 \neq K_3$ (индексы относятся к транзисторам $T1$ и $T3$). Тогда дифференциальный выходной сигнал, обусловленный синфазным входным сигналом, можно записать в виде:

$$U_{вых\,д} = (K_1 - K_3)\,\Delta U_{зи} = (K_1 - K_3)\,\frac{U_{вх\,с}}{2SR_1 + 1},$$

где для $\Delta U_{зи}$ использовано выражение (9.56).

Полагая $K_1 = K + \Delta K$ и $K_3 = K - \Delta K$ (где K — среднее значение) и деля $U_{вых\,д}$ на $U_{вх\,с}$, получаем:

$$K_{дс} = 2\Delta K/(2SR_1 + 1). \qquad (9.58)$$

Коэффициент подавления найдем, поделив K на $K_{дс}$. Запишем его в общей форме, аналогичной (9.42):

$$K_п = (2SR_1 + 1)/2\delta. \qquad (9.59)$$

Здесь δ — коэффициент асимметрии, который в данном случае имеет вид

$$\delta = \Delta K/K.$$

Используя (9.55), можно представить коэффициент δ в виде суммы относительных разбросов [см. замечания к (9.42)].

Из выражения (9.59) ясно, что основными путями для увеличения $K_п$ являются увеличение крутизны активных транзисторов и сопротивления источника тока. Если, например, $S = 1$

мА/В, $R_г$ = 0,1 МОм и δ = 0,04, то $K_п$ = 2500 (т.е. около 70 дБ). Обычно этот параметр меньше, чем у биполярных ДУ.

Входные сопротивления (как для дифференциального, так и для синфазного сигналов) можно считать практически бесконечно большими: обычно они составляют 10^{10}–10^{12} Ом. Соответственно пренебрежимо малы входные токи. Отсюда следует, что *такие параметры ДУ, как ток смещения* (разность входных токов), *средний входной ток и их температурный дрейф не являются ограничивающими факторами при использовании МДП транзисторных ДУ.*

Если у биполярных ДУ главная составляющая напряжения $U_{см}$ обусловлена разбросом тепловых токов эмиттера (см. разд. 9.6), то у МДП-транзисторных ДУ главная составляющая обусловлена разбросом пороговых напряжений и удельных крутизн активных транзисторов, т.е. параметров, зависящих не только от геометрии и электрофизических свойств кристалла, но и от состояния его поверхности. Как известно, состояние поверхности труднее контролировать, чем свойства объема и геометрию. Поэтому значения $U_{см}$ получаются больше, чем у биполярных ДУ.

Переходные характеристики МДП-транзисторных ДУ такие же, как у простейших каскадов, образующих плечи ДУ. Соответствующие выражения были приведены в разд. 9.5.

9.7. Эмиттерные повторители

Повторителями принято называть усилители с коэффициентом усиления, близким к единице, не меняющие полярность входного сигнала и обладающие повышенным входным и пониженным выходным сопротивлениями (по сравнению с простейшими усилительными каскадами).

Классическая схема *эмиттерного повторителя* показана на рис. 9.14, *а*, а его малосигнальная эквивалентная схема — на рис. 9.14, *б*. Легко заметить, что эмиттерный повторитель отличается от простейшего усилителя (см. рис. 9.3, *а*) только тем, что выходное напряжение снимается не с коллектора, а с эмиттера и что в цепи коллектора отсутствует резистор $R_к$.

Коэффициент передачи. Из рис. 9.14, *а* видно, что если $R_г = 0$ и $U^* = \text{const}$, то входной сигнал полностью передается на

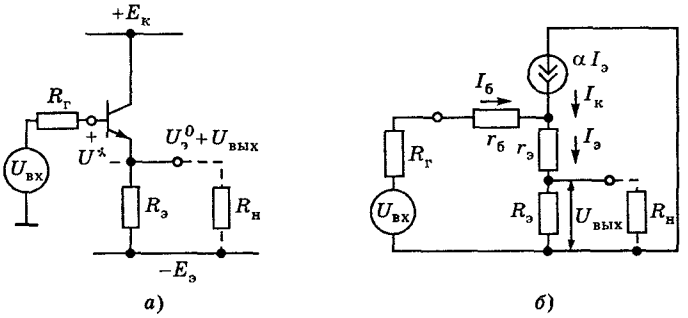


Рис. 9.14. Принципиальная (*а*) и малосигнальная эквивалентная (*б*) схемы эмиттерного повторителя

выход: $U_{вых} = U_{вх}$ и $K = 1$. С учетом сопротивлений $R_г$, $r_б$ и $r_э$ из рис. 9.14, *б* получаем соотношение:

$$U_{вых} = U_{вх} - I_б (r_б + R_г) - I_э r_э,$$

где $U_{вых} = I_э R_э$ и $I_б = (1 - \alpha) I_э$.

Из этого соотношения легко найти величину $I_э$, а затем выразить $U_{вых}$ через $U_{вх}$. В результате коэффициент передачи повторителя запишется в следующем общем виде:

$$K = \frac{R_э}{R_э + r_э + (1 - \alpha)(r_б + R_г)}. \qquad (9.60)$$

Например, если $R_г = 0$; $R_э = 5$ кОм, $r_э = 25$ Ом; $r_б = 150$ Ом и $\beta = 100$, то $K \approx 0{,}995$. Если $R_г = 2$ кОм, то коэффициент передачи несколько уменьшается: $K \approx 0{,}991$.

При подключении внешнего нагрузочного сопротивления $R_н$ (на рис. 9.14 показано штриховой линией) коэффициент передачи дополнительно уменьшается, поскольку сопротивление $R_э$ заменяется на $R_э \parallel R_н$.

Из (9.60) видно, что коэффициент передачи повторителя положительный, т.е. *повторитель не меняет полярности сигнала*, или, в случае синусоидальных сигналов, *не меняет их фазы* (конечно, при достаточно низких частотах).

Несмотря на единичный коэффициент передачи напряжения, повторитель относится к классу усилителей, поскольку он усиливает ток. Это ясно из того, что между выходным (эмиттерным) и входным (базовым) токами существует хорошо известное соотношение: $I_э = (\beta + 1) I_б$, где $\beta >> 1$.

Входное сопротивление. Входное сопротивление повторителя, изображенного на рис. 9.14, определяется теми же формулами (9.7), что и для простейшего усилителя. В общем случае

$$R_{вх} = (\beta + 1)(R_э + r_э) + r_б, \qquad (9.61а)$$

а в частном случае, когда сопротивлениями $r_э$ и $r_б$ можно пренебречь,

$$R_{вх} = (\beta + 1)R_э. \qquad (9.61б)$$

С учетом внешнего нагрузочного сопротивления входное сопротивление уменьшается.

В отличие от усилительного каскада сопротивление $R_э$, а вместе с ним и входное сопротивление повторителя, можно увеличивать, практически не изменяя коэффициента передачи [ср. (9.5а) и (9.60)]. Однако с ростом $R_э$ необходимо увеличивать напряжение питания $E_э$ с тем, чтобы сохранить желательное значение тока покоя $I_э^0$ [см. (9.2а)]. Поэтому такой путь повышения входного сопротивления ограничен. На практике для повышения $R_{вх}$ используют либо источник тока в эмиттерной цепи, либо составной транзистор.

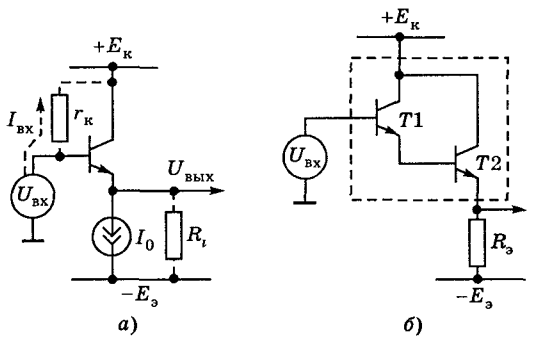


Рис. 9.15 Повторители с повышенным входным сопротивлением *а* — с нагрузкой в виде источника тока, *б* — с использованием пары Дарлингтона

На рис. 9.15, *а* показана схема повторителя, в которой резистор $R_э$ заменен на источник тока I_0. Если источник тока идеальный ($R_i = \infty$), то согласно (9.61) получается $R_{вх} = \infty$. На самом же деле входное сопротивление имеет конечное значение, обусловленное сопротивлением коллекторного перехода (на рис. 9.15, *а* показано штриховой линией). До сих пор это сопротивление не

учитывалось, так как в обычных усилителях его роль не существенна. Однако в данном случае, когда цепь эмиттера «оборвана» для переменных составляющих тока (поскольку ток эмиттера жестко задан), сопротивление $r_к$ является единственным элементом, через который может протекать входной ток. Следовательно, *максимальное значение входного сопротивления повторителя* (и любого другого усилителя)

$$R_{вх\ макс} = r_к. \qquad (9.62)$$

При токе $I_э = 1$ мА значения $r_к$ составляют 2–3 МОм. С уменьшением тока сопротивление $r_к$ возрастает [см. (5.44)], но предел ставится поверхностными утечками коллекторного перехода.

При конечном сопротивлении источника тока R_t входное сопротивление повторителя будет меньше $r_к$. Его можно оценить как параллельное соединение величин $r_к$ и $(\beta + 1)R_t$.

На рис 9.15,*б* показан повторитель, выполненный на паре Дарлингтона (см. рис. 9.1, *а*). Как известно, составному транзистору свойственны весьма большие значения коэффициента усиления: $\beta \approx \beta_1\beta_2$ [см. (9.1)]. Поэтому согласно (9.61) в повторителе по схеме Дарлингтона *легко получить большое входное сопротивление при сравнительно малом сопротивлении* $R_э$. Например, если $R_э = 2$ кОм и $\beta = 2000$, то расчетное значение $R_{вх} \approx 4$ МОм. Реальное значение $R_{вх}$, как и в предыдущей схеме, ограничено сопротивлением $r_к$ [см. (9.62)].

Выходное сопротивление. Выходное сопротивление повторителя можно найти из общего определения, приведенного в разд. 9.3. Положим сначала $R_н = \infty$ (рис. 9.14), тогда

$$(U_{вых})_{хх} = KU_{вх},$$

где K — коэффициент усиления (9.60). Затем положим $R_н = 0$, т.е. закоротим сопротивление $R_э$. Тогда

$$(I_{вых})_{кз} = \frac{U_{вх}}{r_э + (1 - \alpha)(R_г + r_б)}$$

[см., например, (9.4)]. Деля напряжение холостого хода на ток короткого замыкания и подставляя значение K, нетрудно представить выходное сопротивление в следующем виде:

$$R_{вых} = R_э \parallel [r_э + (1 - \alpha)(R_г + r_б)]. \qquad (9.63а)$$

Внешняя составляющая $R_э$ обычно мало существенна. Также мало существенна и составляющая $(1 - \alpha) r_б$ по сравнению с $r_э$. Поэтому в реальных схемах повторителей можно пользоваться упрощенным выражением

$$R_{вых} = r_э + \frac{1}{\beta + 1} R_г. \qquad (9.63б)$$

Как видим, в общем случае выходное сопротивление зависит от сопротивления источника входного сигнала. Однако при достаточно больших значениях β (например, при использовании пары Дарлингтона) вторым слагаемым в (9.63б) можно пренебречь. Тогда выходное сопротивление минимально и определяется только сопротивлением эмиттерного перехода:

$$R_{вых\ мин} = r_э. \qquad (9.64)$$

Легко заметить, что отношение входного и выходного сопротивлений у повторителя несравненно больше, чем у простейшего усилительного каскада и у ДУ, у которых это отношение не превышает величины $\beta + 1$. Отношение максимального входного сопротивления (9.62) к минимальному выходному (9.64) составляет $r_к/r_э$ (обычно более 50000). Это отношение не зависит от рабочего тока, поскольку и $r_к$, и $r_э$ обратно пропорциональны току.

В связи с большим различием входного и выходного сопротивлений повторитель широко используют в качестве *буферного каскада* — так называемого *трансформатора сопротивлений*. Эта функция повторителя проиллюстрирована на рис. 9.16.

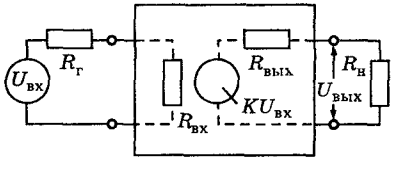


Рис. 9.16. Повторитель в качестве буферного каскада

Пусть имеются источник сигнала $E_г$ с сопротивлением $R_г$ и нагрузка с сопротивлением $R_н$, причем $R_г >> R_н$. Тогда, подключая нагрузку непосредственно к источнику $E_г$, получим очень малый коэффициент передачи напряжения:

$$\frac{U_{вых}}{E_г} = \frac{R_н}{R_г + R_н} << 1.$$

Если же между источником $E_г$ и нагрузкой включить буферный повторитель (рис. 9.16) и если при этом выполняются характерные для повторителя соотношения $R_{вх} >> R_г$ и $R_{вых} << R_н$, то коэффициент передачи напряжения будет близок к единице. Таким образом, использование повторителя позволяет как бы искусственно уменьшить сопротивление источника сигнала или увеличить сопротивление нагрузки, откуда и термин — «трансформатор сопротивлений».

В частности, используя буферные повторители, можно уменьшить сопротивление $R_г$ на входе дифференциального усилителя и тем самым повысить его быстродействие (см. разд. 9.6). Буферные повторители находят также широкое применение в случае емкостной нагрузки (кабели, длинные провода, развитая сеть металлической разводки в ИС). На достаточно высоких частотах сопротивление такой нагрузки оказывается малым, и ее непосредственное подключение к высокоомному источнику сигнала приводит к снижению коэффициента передачи, как показано выше. Буферный повторитель позволяет улучшить передачу сигнала на высоких частотах и расширить полосу усиливаемых частот.

Переходные процессы. Эквивалентная схема для анализа переходных процессов показана на рис. 9.17. Формальный анализ этой схемы несложен, но приводит к громоздким и мало наглядным выражениям. Поэтому полезнее рассмотреть особенности переходного процесса и привести приближенные расчетные формулы. Пренебрежем сначала емкостями $C_к$ и $C_н$ и подадим на вход ступенчатый сигнал $U_{вх}$. Поскольку в первый момент коллекторный ток не меняется (т.е. генератор тока бездействует), схема превращается в пассивную резисторную цепочку. Отсюда следует, что первое время после подачи сигнала усиление мощности отсутствует.

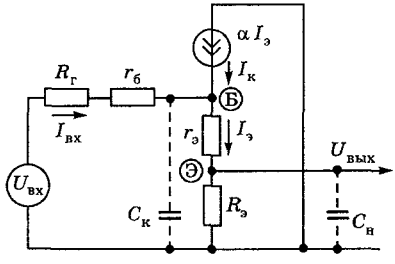


Рис. 9.17. Эквивалентная схема повторителя для области малых времен

Начальные значения токов и выходного напряжения имеют вид:

$$I_{вх}(0) = I_б(0) = I_э(0) = \frac{U_{вх}}{R_г + r_б + r_э + R_э}; \qquad (9.65а)$$

$$U_{вых}(0) = I_э(0)R_э = U_{вх}\frac{R_э}{R_г + r_б + r_э + R_э}. \qquad (9.65б)$$

Соответственно начальный коэффициент передачи

$$K(0) = R_э/(R_э + r_э + r_б + R_г). \qquad (9.66)$$

В дальнейшем, под действием скачка $I_э(0)$, ток коллектора начинает возрастать. Приращения тока $I_к$ распределяются между эмиттерной и базовой цепями обратно пропорционально их сопротивлениям.

Если сопротивление источника сигнала достаточно мало ($R_г << R_э$), то приращения $\Delta I_к$ практически полностью идут в базовую цепь. При этом ток эмиттера, а значит, и выходное напряжение почти не меняются и остаются близкими к начальным значениям (см. сплошные линии на рис. 9.18, *а*). Соответственно длительность фронта $t_{ф1}$, отсчитанная на уровнях 0,1–0,9 от установившегося значения, оказывается равной нулю, т.е. фронт формируется мгновенно. Что касается коллекторной и базовой цепей, то в них переходные процессы выражены достаточно ярко: ток коллектора возрастает от нуля до установившегося значения $\alpha I_э$, а ток базы падает от начального значения $I_э$ до $(1 - \alpha)I_э$. Как видим, отношение токов $I_э$ и $I_б$ возрастает, т.е. через некоторое время повторитель начинает усиливать мощность. Поскольку эмиттерный ток остается почти постоянным, коллекторный и базовый токи меняются с постоянной времени τ_α.

Если условие $R_г << R_э$ не выполняется, то переходный процесс претерпевает некоторые изменения (см. штриховые кривые на рис. 9.18, *а*). А именно, начальные скачки эмиттерного тока и выходного напряжения уменьшаются, а последующие приращения коллекторного тока ответвляются не только в цепь базы, но в значительной мере и в цепь эмиттера. Соответственно ток $I_э$ и напряжение $U_{вых}$ заметно возрастают во время переходного процесса. В этом случае, как видно из рисунка,

длительность фронта $t_{ф2}$ имеет конечное значение. Постоянная времени, с которой меняются коллекторный ток и другие величины, определяется выражением (9.13). Однако в нем следует заменить величину $\tau_{\alpha ое}$ на τ_α поскольку в начале анализа мы пренебрегли емкостью $C_к$. Таким образом,

$$\tau_в = \tau_\alpha(1 - \alpha\gamma_э). \qquad (9.67)$$

Обычно величина $\tau_в$ не превышает (2–5) τ_α.

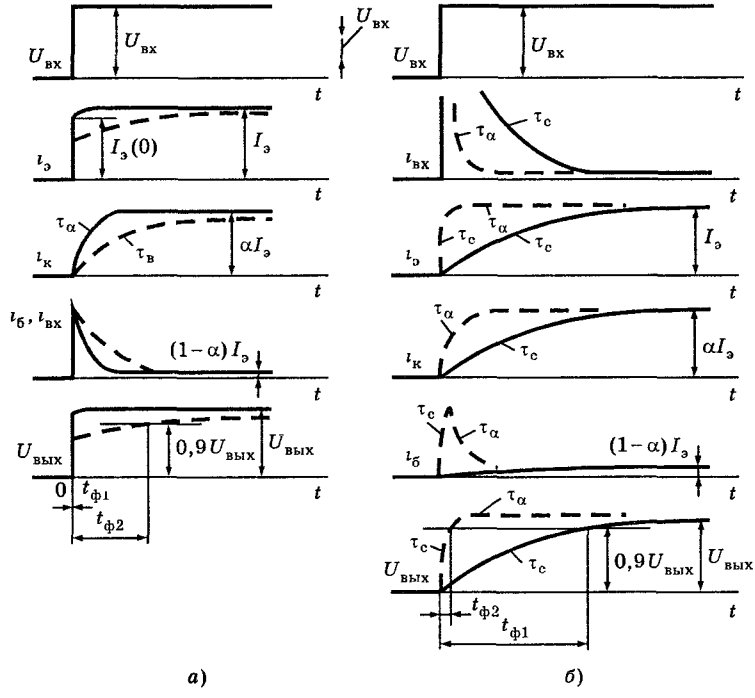


Рис. 9.18. Переходные процессы в повторителе без учета (*а*) и с учетом (*б*) коллекторной емкости

Теперь учтем емкость $C_к$, но сначала положим $\tau_\alpha = 0$, т.е. пренебрежем инерционностью процессов в базе. Тогда при поступлении ступенчатого входного сигнала входной ток в первый момент полностью протекает через емкость, а токи транзистора и выходное напряжение не меняются. По мере заряда емкости токи транзистора и выходное напряжение плавно возрастают до установившихся значений (см. сплошные кривые на рис. 9.18, *б*).

Постоянная времени этого переходного процесса определяется емкостью $C_к$ и шунтирующими ее сопротивлениями (см. рис. 9.17):

$$\tau_с = C_к[(R_г + r_б) \parallel (R_э + r_э)]. \qquad (9.68)$$

Фронты импульсов будут экспоненциальными; их длительность на уровне 0,1–0,9 от установившегося значения составит:

$$t_ф = 2{,}2\tau_с. \qquad (9.69)$$

Типичными значениями постоянной времени являются $\tau_с \approx 0{,}5$–1 нс. В случае идеального источника сигнала ($R_г = 0$) постоянная времени практически совпадает с постоянной времени базы (5.67): $\tau_с = C_к r_б$. Значения этой величины лежат в пределах 0,1–0,2 нс и менее.

Начальный «пик» входного тока (рис. 9.18, *б*) намного превышает установившееся значение тока базы. Действительно, в первый момент напряжение на емкости $C_к$ не меняется, а значит, начальный входной ток определяется входным напряжением и сопротивлением в цепи базы:

$$I_{вх}(0) = U_{вх}/(R_г + r_б).$$

Например, если $U_{вх} = 0{,}1$ В и $R_г + r_б = 1$ кОм, то $I_{вх}(0) = 0{,}1$ мА, тогда как в установившемся режиме, при $R_{вх} = 100$ кОм, получаем $I_{вх} = 1$ мкА — значение, в 100 раз меньшее.

Покажем теперь, что формула (9.69) пригодна для приблизительной оценки длительности фронта с учетом обоих инерционных факторов: коллекторной емкости и времени пролета носителей в базе. Для этого рассмотрим два крайних случая: $\tau_с > 2\tau_в$ и $\tau_с < 0{,}5\tau_в$. Первый из них характерен для сравнительно высокоомных, а второй — для сравнительно низкоомных источников сигнала.

В первом случае постоянная времени $\tau_в$ мало существенна. Следовательно, переходный процесс характеризуется постоянной времени $\tau_с$, а длительность фронта — формулой (9.69).

Во втором случае переходный процесс оказывается «комбинированным». Действительно, в этом случае токи $I_б$ и $I_э$ нарастают с постоянной времени $\tau_с$, т.е. быстрее, чем ток $I_к$, который нарастает с постоянной времени $\tau_в$. Поэтому можно считать, что на первом этапе токи $I_б$ и $I_э$ вместе достигают

максимального значения (9.65а), а на втором этапе ток $I_к$, ответвляясь в низкоомную цепь базы, уменьшает базовый ток до установившегося значения. Ток эмиттера, а значит, и выходное напряжение, на втором этапе почти не меняются. Соответственно длительность фронта определяется длительностью первого этапа, т.е. формулой (9.69).

В промежуточном случае ($\tau_с \approx \tau_в$) переходный процесс оказывается более сложным, но расчеты подтверждают, что и в этом случае формула (9.69) дает приемлемые результаты.

Емкость нагрузки, если она достаточно мала, а именно, если $C_н R_э < \tau_с$, можно считать подключенной параллельно коллекторной емкости. Тогда длительности фронта и среза соответственно возрастают. Если же нагрузочная емкость велика, то длительности фронта и среза могут оказаться разными (см. ниже).

Блокировка повторителя. Специфический переходный процесс имеет место в повторителе при достаточно большой нагрузочной емкости $C_н$ (рис. 9.17). В первый момент после поступления ступенчатого входного сигнала напряжение на емкости $C_н$ не меняется, а значит, не меняется и потенциал эмиттера. Поэтому, если сигнал $U_{вх}$ имеет отрицательную полярность и превышает величину U^* (рис. 9.19), то напряжение на эмиттерном переходе становится обратным и транзистор запирается.

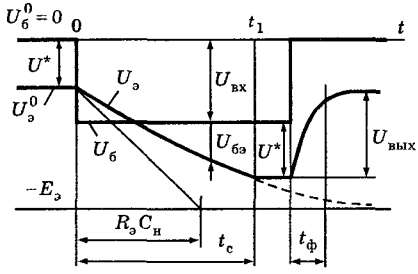


Рис. 9.19. Переходный процесс в повторителе при большой нагрузочной емкости

После этого емкость $C_н$ разряжается через сопротивление $R_э$ с большой постоянной времени $C_н R_э$. Транзистор снова открывается только тогда, когда напряжение база–эмиттер достигает значения U^* (момент t_1 на рис. 9.19). Описанное явление (т.е. временное запирание транзистора при больших отрицательных сигналах) получило название *блокировки повторителя*.

Как известно, для запирания кремниевых транзисторов обратное смещение на эмиттерном переходе не обязательно: достаточно, чтобы прямое смещение было на 0,1–0,2 В меньше

величины U^*. Поэтому явление блокировки может иметь место даже при малых отрицательных сигналах 0,2–0,3 В.

В результате блокировки длительность среза превышает длительность фронта. Последняя по-прежнему определяется формулой (9.69). Что касается длительности среза, то ее легко найти из выражения, описывающего процесс разряда нагрузочной емкости через сопротивление $R_э$ (рис. 9.19):

$$u_э(t) = -U^* e^{-t/\tau_н} - E_э(1 - e^{-t/\tau_н}),$$

где $\tau_н = C_н R_э$. Подставляя в левую часть $U_э = -U_{вх} - U^*$ (условие отпирания эмиттерного перехода в момент t_1, см. рис. 9.19), получаем:

$$t_с = \tau_н \ln\left(1 - \frac{U_{вх}}{E_э - U^*}\right)^{-1}. \qquad (9.70)$$

Например, если $\tau_н = 40$ нс, $E_э = 2$ В и $U_{вх} = 0{,}5$ В, то $t_с \approx 24$ нс. Полученное значение $t_с$ намного больше типичных значений $t_ф$.

Следует подчеркнуть, что блокировка повторителя и различие $t_ф$ и $t_с$ подразумевают ступенчатый характер сигналов. В противном случае емкость $C_н$ успевает «отследить» входной отрицательный сигнал и блокировка не имеет места. Критерием блокировки может служить неравенство

$$t_{с\,вх} < r_б C_н,$$

где $t_{с\,вх}$ — время нарастания среза входного импульса.

Схемы сдвига уровня. В многокаскадных усилителях на базу каждого следующего каскада поступает не только полезный сигнал, но и постоянная составляющая напряжения с коллектора предыдущего каскада. Таким образом, постоянная составляющая «накапливается», возрастает от каскада к каскаду, что вызывает определенные затруднения при разработке последних — выходных каскадов. Поэтому часто возникает задача устранить постоянную составляющую на входе очередного каскада, но по возможности без изменений передать переменную составляющую — сигнал. Именно такую задачу и решают так называемые *схемы сдвига уровня*.

Простейшей схемой сдвига уровня является эмиттерный повторитель. Действительно, у него уровень выходного (эмиттерного) потенциала ниже уровня базового потенциала на величину U^*, а сигнал передается с коэффициентом $K \approx 1$.

Эмиттерный повторитель лежит в основе других, более сложных схем сдвига уровня. Например, если нужно понизить уровень входного сигнала на величину $2U^*$, то можно либо использовать повторитель по схеме Дарлингтона (см. рис. 9.15, *б*), либо включить в эмиттерную цепь простейшего повторителя прямосмещенный диод (см. ниже).

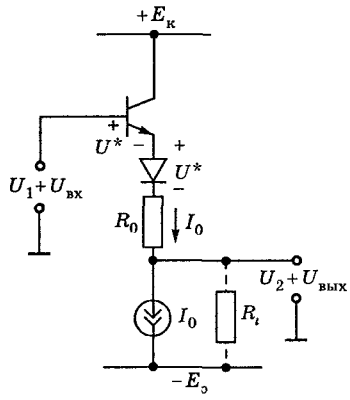


Рис. 9.20. Схема сдвига уровня

Иногда требуется сместить уровень на величину, не кратную U^*, например, на 2,5 В. Тогда используется универсальная схема сдвига уровня, показанная на рис. 9.20. В общем случае в такой схеме может быть не один, а n последовательно включенных диодов. Соотношение между входными и выходными уровнями имеет вид:

$$U_1 - U_2 = (n + 1)U^* + I_0R_0. \qquad (9.71)$$

Варьируя значения n, I_0 и R_0, можно обеспечить любой сдвиг уровня. Так, если желательный сдвиг составляет $U_1 - U_2 = 2{,}5$ В, то нужно выбрать $n = 2$; тогда $(n + 1)U^* = 2{,}1$ В и $I_0R_0 = 0{,}4$ В; при токе $I_0 = 1$ мА получаем необходимое значение сопротивления $R_0 = 400$ Ом.

Коэффициент передачи переменной составляющей в схеме на рис. 9.20 зависит в первую очередь от внутреннего сопротивления источника тока. Если $R_t = \infty$, то $K = 1$ независимо от структуры эмиттерной и базовой цепей. Если же сопротивление R_t имеет конечное значение, его следует подставить вместо $R_э$ в формулу (9.60). Кроме того, в этой формуле нужно заменить сопротивление $r_э$ на $(n + 1)r_э + R_0$. Тогда

$$K = \frac{R_t}{R_t + (n+1)r_э + R_0 + (1-\alpha)(r_б + R_г)}. \qquad (9.72)$$

Как правило, сопротивление R_t лежит в пределах от 100 кОм и выше, а другие сопротивления в знаменателе (9.72) не превышают 1–2 кОм. Поэтому коэффициент передачи сигнала оказывается весьма близким к единице. При наличии сопротивления нагрузки $R_н$ необходимо в формуле (9.72) заменить R_t на меньшую величину $R_t \parallel R_н$. Тогда коэффициент передачи соответственно уменьшится.

9.8. Каскод

Под каскодом понимается схема, в которой два транзистора соединены последовательно, так что через них протекает один и тот же ток (рис. 9.21). Такое соединение можно рассматривать как одно целое (см. штриховую линию), т.е. как вариант составного транзистора. Коэффициент передачи α в таком составном транзисторе легко найти из следующих очевидных соотношений:

$$I_{к2} = \alpha_2 I_{э2} = \alpha_2 I_{к1} = \alpha_2(\alpha_1 I_{э1}).$$

Деля $I_{к2}$ на $I_{э1}$, получаем:

$$\alpha = \alpha_1 \alpha_2.$$

Значит, коэффициент передачи эмиттерного тока при каскодном соединении ненамного отличается от коэффициента передачи в одном (активном) транзисторе $T1$. Отсюда ясно, что и коэффициент усиления напряжения в каскодной схеме будет практически таким же, как и в простейшем усилителе [см. (9.5б)]:

$$K \approx -\alpha_1 \alpha_2 (R_к / R_э).$$

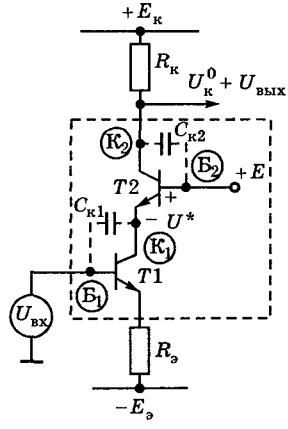


Рис. 9.21. Каскод

Не давая выигрыша по коэффициенту усиления (а также по входному и выходному сопротивлениям), каскод, однако, имеет важное преимущество по сравнению с простейшим усилителем. Это преимущество состоит в *отсутствии связи между точкой вы-*

хода (коллектор $К_2$) и *точкой входа* (база $Б_1$), тогда как в простейшем усилителе выходная и входная точки связаны через емкость $С_к$ и сопротивление $r_к$. Такая обратная связь в ряде случаев осложняет работу усилителей. В частности, она приводит к увеличению входной емкости каскада (эффект Миллера, см. разд. 9.5). При наличии индуктивной составляющей в сопротивлении нагрузки такая связь может привести к самовозбуждению усилителя, т.е. превращению его в генератор[1].

Причиной, по которой выход и вход в каскоде *развязаны* друг от друга, является то, что промежуточная точка схемы (база $Б_2$) находится под неизменным потенциалом E. Величину E можно считать напряжением питания для транзистора $Т1$. В таком случае нагрузкой этого транзистора является весьма малое сопротивление эмиттерного перехода $r_{э2}$. Значит, транзистор $Т1$ работает практически в режиме короткого замыкания коллекторной цепи. Соответственно его коэффициент усиления близок к нулю, эффект Миллера отсутствует и входная емкость равна проходной емкости $С_{к1}$.

Выходное напряжение каскода вызывает протекание тока через емкость $С_{к2}$. Однако этот ток замыкается «на землю» через источник постоянного смещения E и не попадает в коллекторную цепь транзистора $Т1$. Поэтому между выходом и входом каскода отсутствует обратная связь, а значит, опасность самовозбуждения (см. замечания выше) устраняется или по крайней мере существенно уменьшается. Благодаря такой особенности каскод находит широкое применение в резонансных (контурных) усилителях.

9.9. Выходные каскады

Задача выходных каскадов — обеспечить заданную (достаточно большую) мощность и, значит, — достаточно большие напряжения и токи в нагрузке. Коэффициент усиления напряжения является для выходных каскадов второстепенным параметром. На первый план выдвигаются такие параметры, как

---

1 Индуктивная составляющая неизбежна, когда в качестве нагрузки используются колебательные контуры, что характерно для избирательных — резонансных усилителей Поэтому в резонансных каскадах проблема *развязки* выхода и входа стоит особенно остро

коэффициент полезного действия (η) и *коэффициент нелинейных искажений* ($K_{н\,н}$).

Выходные каскады обычно потребляют основную часть мощности усилителя; поэтому высокий к. п. д. означает эффективное использование источника питания. Это особенно важно в интегральных схемах, в которых мощность, рассеиваемая кристаллом, ограничена. Что касается коэффициента нелинейных искажений, то он характерен для выходных каскадов потому, что получение заданной достаточно большой мощности при высоком к.п.д. неизбежно связано с использованием таких токов и напряжений, амплитуды которых сравнимы со значениями постоянных составляющих.

Параметры выходных каскадов. Коэффициент полезного действия определяется как отношение мощности выходного сигнала к мощности, отбираемой от источника питания:

$$\eta = \frac{\frac{1}{2} U_{\text{вых}\,m} I_{\text{вых}\,m}}{E I_{\text{ср}}}, \qquad (9.73)$$

где $I_{\text{вых}\,m}$, $U_{\text{вых}\,m}$ — амплитуды выходных величин.

Поскольку к.п.д. зависит от выходной мощности, принято понимать под к.п.д. его максимальное значение, соответствующее максимальной выходной мощности.

Коэффициент нелинейных искажений характеризует отличие формы выходного сигнала от формы входного. Это отличие обусловлено нелинейностью передаточной характеристики каскада (рис. 8.1). Нелинейные искажения характерны появлением в выходном сигнале новых гармоник, отсутствующих во входном сигнале. Поэтому нелинейные искажения принято оценивать, задавая на входе усилителя чисто синусоидальный сигнал и вычисляя (или измеряя) гармоники, появившиеся на выходе.

Энергетической характеристикой нелинейных искажений служит отношение суммарной мощности высших гармоник, начиная со *2*-й, к мощности *1*-й гармоники, частота которой равна частоте входного сигнала.

Коэффициентом нелинейных искажений принято называть корень квадратный из отношения суммарной мощности высших гармоник на выходе усилителя к мощности *1*-й гармоники. Если сопротивление нагрузки одинаково для всех гармо-

ник, то мощности пропорциональны квадратам токов или напряжений. Тогда коэффициент нелинейных искажений запишется в одной из следующих двух форм:

$$K_{\text{н и}} = \frac{\sqrt{\Sigma I_m^2}}{I_1} = \frac{\sqrt{\Sigma U_m^2}}{U_1}, \qquad (9.74)$$

где m — номер гармоники, начиная со *2*-ой.

Допустимое значение коэффициента $K_{\text{н и}}$ диктуется конкретными требованиями к той или иной аппаратуре. Например, при воспроизведении звука желательно иметь $K_{\text{н и}} < 2$–3%. В измерительных устройствах значения $K_{\text{н и}}$ должны быть существенно меньше.

Аналитическая оценка нелинейных искажений возможна только тогда, когда передаточная характеристика задана в виде математической функции. Чаще коэффициент нелинейных искажений оценивают графически (по известной передаточной характеристике) или экспериментально (с использованием специальных измерительных приборов).

Классы усиления. В зависимости от того, как расположена на передаточной характеристике исходная рабочая точка (точка покоя), различают несколько *классов усиления*: А, В, АВ и др. Эти классы различаются максимальными значениями к.п.д. и величинами нелинейных искажений.

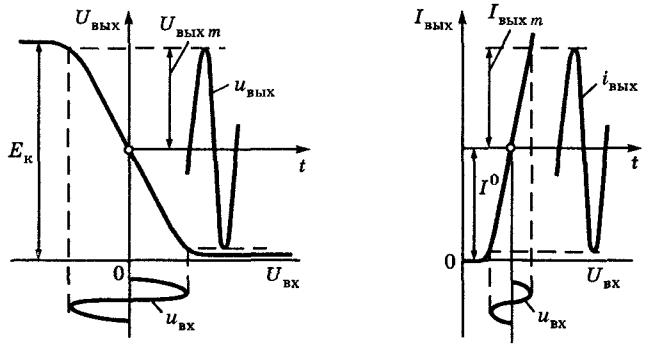
Класс А характерен тем, что рабочая точка в режиме покоя расположена в середине квазилинейного участка передаточной характеристики (рис. 9.22). Очевидно, что нелинейные искажения в этом случае будут минимальными, так как обе полуволны входного сигнала укладываются в пределы квазилинейного участка. Что касается максимального к.п.д., то он получается из формулы (9.73), если подставить в нее максимальные значения амплитуд напряжения и тока, соответствующие границам квазилинейного участка:

$$U_{\text{вых m}} = \tfrac{1}{2} E_{\text{к}};\ I_{\text{вых m}} = I^0 = I_{\text{ср}};$$

тогда

$$\eta_m = 1/4 \text{ (т.е. } 25\ \%). \qquad (9.75)$$

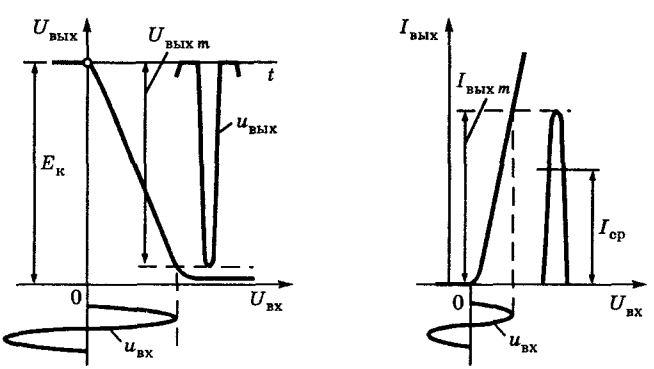
При использовании трансформаторной связи транзистора с нагрузкой максимальный к.п.д. удваивается: $\eta_m = 50\ \%$. Однако в интегральных схемах использование трансформаторов (в качестве навесных компонентов) нежелательно.



а) *б)*

Рис. 9.22. Режим класса А. Расчет выходных напряжения (*а*) и тока (*б*).

Класс В характерен тем, что рабочая точка в режиме покоя расположена на границе квазилинейного участка, а именно, на той, которая соответствует запертому состоянию транзистора (рис. 9.23). Ясно, что в этом случае усиливаются только положительные полуволны входного сигнала. Поэтому выходное напряжение оказывается существенно несинусоидальным, т.е. содержит большое число высших гармоник. Анализ показывает, что коэффициент нелинейных искажений в классе В неза-



а) *б)*

Рис. 9.23. Режим класса В. Расчет выходных напряжения (*а*) и тока (*б*).

висимо от амплитуды сигнала составляет около 70 %, что, конечно, неприемлемо. Поэтому в таком простейшем варианте класс В не используется.

Для класса В характерна так называемая *двухтактная схема*, состоящая из двух усилителей, один из которых усиливает положительную полуволну сигнала, а другой — отрицательную. В нагрузке эти полуволны складываются и образуют полную синусоиду (рис. 9.24). Чтобы усилить отрицательные полуволны, нужно либо включить на вход усилителя *2* фазоинверсный элемент (например, трансформатор), либо использовать в усилителе *2* транзистор другого типа проводимости (*p–n–p*). Примеры реализации двухтактной схемы приведены ниже.

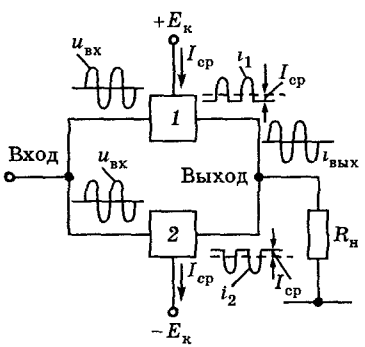


Рис. 9.24. Принцип действия двухтактного усилителя

Коэффициент полезного действия двухтактной схемы класса В можно оценить, исходя из того, что мощность, рассеиваемая в каждой из ее половин, одинакова. Следовательно, достаточно рассчитать мощность в течение одного полупериода. Для этого в формулу (9.73) нужно подставить $U_{\text{вых m}} = E_{\text{к}}$ (см. рис. 9.23, *а*), $E = E_{\text{к}}$ и $I_{\text{ср}} = (2/\pi)I_{\text{вых m}}$ (среднее значение синусоидального тока за половину периода, см. рис. 9.23, *б*). Тогда максимальный к.п.д. принимает следующее значение:

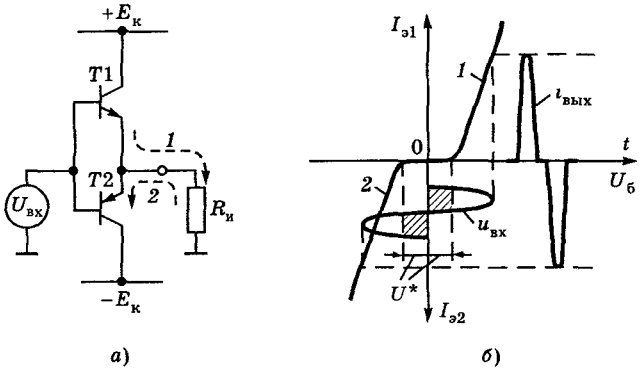
$$\eta_m = \pi/4 \text{ (т.е. } 78\%). \qquad (9.76)$$

Значит, если, например, усилитель выполнен на кристалле, способном рассеять мощность 300 мВт, то в нагрузке можно выделить полезную мощность более 1 Вт.

Под классом АВ понимают промежуточный случай между классами А и В: исходная рабочая точка лежит не на границе отсечки, а в области прямых смещений эмиттерного перехода, но при токах, значительно меньших, чем в классе А (пример будет приведен ниже).

Схемы выходных каскадов. В микроэлектронике класс А не используется из-за низкого к.п.д. Наиболее популярны двухтактные усилители классов В и АВ.

На рис. 9.25,*а* показана простейшая двухтактная схема класса В, выполненная на комплементарных транзисторах. Нагрузка включена в эмиттерную цепь транзисторов; следовательно, они работают в режиме повторителей напряжения. Усиление мощности обусловлено усилением тока.



а) *б*)

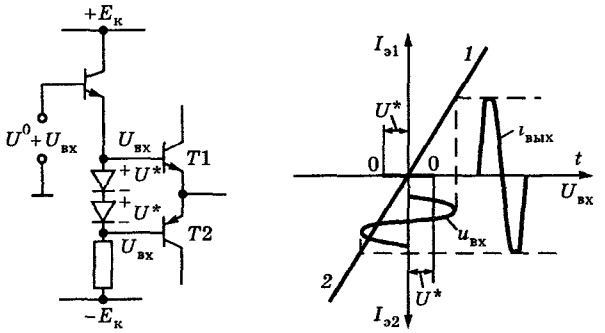
Рис. 9.25. Схема (*а*) и передаточные характеристики (*б*) двухтактного усилителя класса *В* на комплементарных транзисторах

В режиме покоя оба транзистора заперты, поскольку напряжения на эмиттерных переходах равны нулю. Во время положительной полуволны входного сигнала открыт *n*–*p*–*n*-транзистор *T*1, ток через нагрузку протекает по штриховой стрелке *1*. Во время отрицательной полуволны открыт *p*–*n*–*p*-транзистор *T*2 и ток протекает по штриховой стрелке *2*. Таким образом, выходной сигнал двуполярный, как и входной. Коэффициент усиления мощности близок к отношению эмиттерного и базового токов, т.е. к $B + 1$.

К сожалению, такой простой схеме свойственны сравнительно большие нелинейные искажения. Их причина связана с наличием «пятки» на входной ВАХ кремниевых транзисторов. Форму выходного сигнала легко получить графически с помощью так называемых *совмещенных передаточных характеристик* (рис. 9.25, *б*). На этих характеристиках кривая *1* соответствует транзистору *T*1, а кривая *2* — транзистору *T*2. Как видим, длительность положительной и отрицательной полуволн на выходе меньше полупериода, и эти полуволны отделены друг от друга небольшими горизонтальными участками (не усиливаемые части полусинусоид заштрихованы). Очевидно,

что такие искажения будут особенно существенны при малых входных сигналах с амплитудой, сравнимой с напряжением U^*.

Для устранения отмеченного недостатка усилитель несколько усложняют, вводя раздельные (индивидуальные) смещения на базы транзисторов с помощью схемы сдвига уровня (рис. 9.26, *а*). Совмещенные характеристики для этого варианта показаны на рис. 9.26, *б*. В данном случае используется режим класса АВ.



а) *б*)

Рис. 9.26. Схема (*а*) и передаточные характеристики (*б*) двухтактного усилителя класса АВ на комплементарных транзисторах

Поскольку параметры интегральных n–p–n- и p–n–p-транзисторов значительно различаются, в качестве транзистора $T2$ на рис. 9.25 и 9.26 часто используют составной p–n–p-транзистор (рис. 9.2). Это способствует симметрии выходного каскада и уменьшению нелинейных искажений.

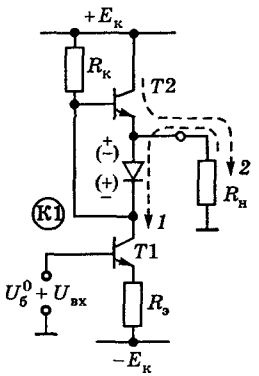


Рис. 9.27. Двухтактный усилитель класса АВ на транзисторах одного типа

Если по тем или иным причинам выходной каскад нужно строить на однотипных (не комплементарных) транзисторах, схема каскада выглядит иначе (рис. 9.27). Здесь транзистор $T1$ открыт в течение обоих полупериодов. В режиме покоя ток $I^0_{к1}$ и сопротивление $R_к$ выбираются так, чтобы потенциал коллектора $U^0_{к1}$ был равен нулю. При этом диод $Д$ и транзистор $T2$ заперты; ток в нагрузке отсутствует.

Во время положительной полуволны входного сигнала потенциал $U_{к1}$ падает, открывается диод $Д$, и через нагрузку протекает ток по штриховой стрелке *1*. Транзистор $T2$ остается запертым, так как прямое напряжение U^* на диоде создает на эмиттерном переходе обратное смещение (см. знаки «+» и «−» без скобок). Во время отрицательной полуволны потенциал $U_{к1}$ повышается, отпирается транзистор $T2$, и ток через нагрузку протекает по штриховой стрелке *2*. При этом диод заперт, так как прямое напряжение U^* на эмиттерном переходе создает на диоде обратное смещение (см. знаки «+» и «−» в скобках).

Для того чтобы открылся диод $Д$ (положительная полуволна) или транзистор $T2$ (отрицательная полуволна), потенциал $U_{к1}$ должен измениться на величину $\pm U^*$ по сравнению с потенциалом покоя. Значит, минимальная амплитуда входного сигнала, на которую реагирует данный усилитель, составляет

$$U^*/K,$$

где K — коэффициент усиления каскада на транзисторе $T1$. Коэффициент усиления может быть достаточно большим, поэтому обычно нет необходимости использовать класс АВ, т.е. вводить дополнительные смещения подобно тому, как это было на рис. 9.26.

Очевидно, что полуволны напряжения на нагрузке будут одинаковыми только в том случае, если будут одинаковыми коэффициенты усиления положительного и отрицательного входного сигнала. При положительном сигнале коэффициент усиления K^+ пропорционален сопротивлению $R_и$, которое в этом полупериоде является коллекторной нагрузкой транзистора $T1$. При отрицательном сигнале коэффициент усиления K^- пропорционален сопротивлению $R_к$ (транзистор $T2$ в этом полупериоде выполняет функцию эмиттерного повторителя). Следовательно, равенство $K^+ = K^-$ требует выполнения равенства $R_к = R_н$.

Недостатком схемы является то, что изменение сопротивления $R_н$ сопровождается изменением амплитуды отрицательной полуволны выходного напряжения (когда ток нагрузки протекает по стрелке *1*), поскольку при этом изменяется коэффициент усиления K^+. Чтобы избежать этого недостатка, схему нужно дополнить буферным повторителем.

9.10. Стабилизаторы напряжения

Стабилизаторы напряжения необходимы прежде всего в качестве источников питания ИС. Примером могут служить усилители постоянного тока, в которых, при питании от нестабилизированных источников, существенное значение приобретает один из компонентов дрейфа (см. разд. 9.3). Кроме того, простейшие стабилизаторы напряжения широко используются в качестве источников смещения во многих аналоговых схемах, в том числе в операционных усилителях.

Параметры стабилизаторов. На скелетной схеме (рис. 9.28) величина U_1 — нестабилизированное (входное), а величина U_2 — стабилизированное (выходное) напряжение; I_1 и I_2 — входной и выходной токи стабилизатора; $R_н = U_2/I_2$ — сопротивление нагрузки.

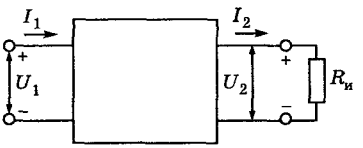


Рис. 9.28. Скелетная схема стабилизатора напряжения

Конечно, выходное напряжение стабилизатора не может быть абсолютно стабильным. Приращения ΔU_2 малы (т.е. лежат в допустимых пределах), но все же зависят от приращений входного напряжения и выходного тока[1]. Поэтому запишем относительную нестабильность выходного напряжения в виде суммы двух слагаемых:

$$\frac{\Delta U_2}{U_2} = \frac{\Delta U_2^U}{U_2} + \frac{\Delta U_2^I}{U_2}, \qquad (9.77)$$

где первое слагаемое обусловлено нестабильностью входного напряжения, а второе — нестабильностью выходного тока.

Составляющие относительной нестабильности принято выражать через два основных параметра стабилизаторов — *коэффициент стабилизации и выходное сопротивление*[2]:

---

1 Помимо этих зависимостей, имеет место температурный и временной дрейф выходного напряжения, аналогичный дрейфу в усилителях постоянного тока (см. разд. 9.3). Этот тип нестабильности далее не рассматривается.

2 Знак минус в правой части (9.78б) означает, что увеличение выходного тока ($\Delta I_2 > 0$) сопровождается уменьшением выходного напряжения ($\Delta U_2 < 0$).

$$\frac{\Delta U_2^U}{U_2} = \frac{1}{K_{ст}} \frac{\Delta U_1}{U_1}; \qquad (9.78а)$$

$$\frac{\Delta U_2^I}{U_2} = -R_{вых} \frac{\Delta I_2}{U_2}. \qquad (9.78б)$$

Из выражения (9.78а) следует, что коэффициент стабилизации определяется следующим образом:

$$K_{ст} = \frac{U_2}{U_1} \frac{\Delta U_1}{\Delta U_2} \qquad (9.79)$$

(при этом ток I_2 считается постоянным). Что касается выходного сопротивления, то оно определяется общим выражением (см. разд. 9.3):

$$R_{вых} = (U_{вых})_{хх}/(I_{вых})_{кз}. \qquad (9.80)$$

При использовании (9.80) нужно представить стабилизатор малосигнальной эквивалентной схемой. Ясно, что качество стабилизатора тем выше, чем больше коэффициент стабилизации и чем меньше выходное сопротивление.

Диодные стабилизаторы. Простейшая схема диодного стабилизатора с использованием источника тока и кремниевого стабилитрона показана на рис. 9.29, *а*. На рис. 9.29, *б* показана малосигнальная эквивалентная схема. Из рисунка следует, что выходное напряжение определяется номинальным напряжением стабилитрона: $U_2 = U_д$. Обычно $U_д > 5$–6 В при лавинном пробое и $U_д = 2$–5 В при туннельном пробое (см. разд. 3.2).

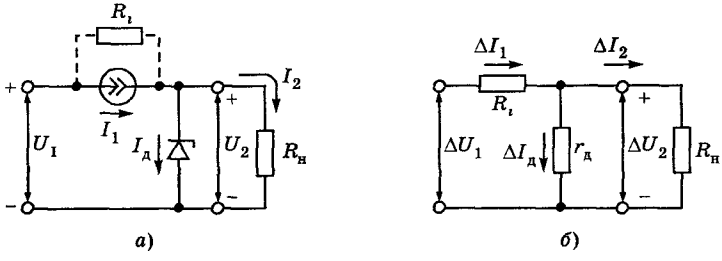


Рис. 9.29. Диодный стабилизатор с использованием стабилитрона

Поскольку ток I_1 задан, увеличение выходного тока сопровождается уменьшением тока диода. При коротком замыкании выходных зажимов ($R_н = 0$) получаем: $I_д = 0$, т.е. диодный стабилизатор «не боится» коротких замыканий на выходе. Эта особенность свойственна всем стабилизаторам *параллельного типа*, у которых регулирующий элемент включен параллельно нагрузке.

Выходное сопротивление диодного стабилизатора согласно (9.80) равно

$$R_{вых} = r_д \parallel R_i \approx r_д. \qquad (9.81)$$

Значение $r_д$ обычно составляет 10–20 Ом при номинальном токе 5–10 мА и несколько возрастает с уменьшением тока.

Коэффициент стабилизации в случае идеального источника тока ($R_i = \infty$) оказывается бесконечно большим, поскольку в этом случае изменения входного напряжения ΔU_1 ни в какой степени не передаются на выход. При конечном значении R_i приращения ΔU_2 и ΔU_1 связаны коэффициентом передачи резистивного делителя напряжения (рис. 9.29, *б*):

$$\Delta U_2 = \Delta U_1 \frac{r_д}{R_i + r_д} \approx \Delta U_1 \frac{r_д}{R_i}.$$

Подставляя отношение $\Delta U_2/\Delta U_1$ в выражение (9.79), получаем

$$K_{ст} = \frac{U_2}{U_1}\frac{R_i}{r_д}. \qquad (9.82)$$

Например, если $U_2/U_1 = 0{,}8$; $R_i = 50$ кОм и $r_д = 10$ Ом, то $K_{ст} = 4000$.

Такое значение коэффициента стабилизации вполне приемлемо в большинстве случаев. Главную же проблему представляет сравнительно большое выходное сопротивление стабилизатора.

Вместо стабилитрона (т.е. диода, работающего в режиме пробоя) можно использовать диод, работающий в режиме прямого тока (рис. 9.30). В этом случае выходное напряжение $U_2 = U^*$, а выходное сопротивление

$$R_{вых} = r_д = (\varphi_T/I_д) + r_б, \qquad (9.83)$$

где первое слагаемое есть дифференциальное сопротивление перехода, а второе — сопротивление базового слоя. С ростом тока нагрузки ток $I_д$ уменьшается и выходное сопротивление существенно возрастает. Минимальное значение $R_{вых}$ соответствует режиму холостого хода ($I_2 = 0$), но в любом случае оно ограничено значением $r_б$ (обычно не менее 2–5 Ом).

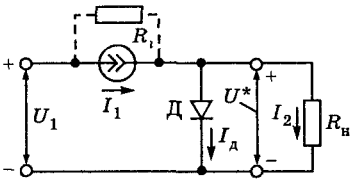


Рис. 9.30. Диодный стабилизатор с использованием прямого включения диода

Коэффициент стабилизации получим, подставляя в (9.82) величину $r_д$ из (9.83) и $U_2 = U^*$:

$$K_{ст} = \frac{U^*}{U_1} \frac{R_i}{\varphi_т / I_д + r_б}. \qquad (9.84)$$

Например, если $R_i = 50$ кОм, $I_д = 1$ мА, $r_б = 5$ Ом и $U_1 = 5$ В, то $K_{ст} \approx 220$ и $R_{вых} = 30$ Ом.

По сравнению с параметрами предыдущей схемы выходное сопротивление получилось в 3 раза больше, а коэффициент стабилизации в 20 раз меньше.

В обеих рассмотренных схемах вместо источника тока можно использовать резистор, который называют *балластным* (рис. 9.31). Выходное сопротивление в этом варианте практически не отличается от предыдущих, но коэффициент стабилизации имеет свою специфику. Она обусловлена тем, что в данном случае входное и выходное напряжение связаны друг с другом:

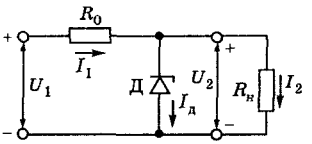


Рис. 9.31. Диодный стабилизатор с балластным резистором

$$U_1 = U_2 + I_1 R_0.$$

Подставляя это значение U_1 в (9.82) и заменяя R_i на R_0, получаем:

$$K_{ст} = \frac{U_2}{U_2 + I_1 R_0} \frac{R_0}{r_д}. \qquad (9.85а)$$

Как видим, с увеличением сопротивления R_0 коэффициент стабилизации данной схемы, в отличие от предыдущих, стремится к предельной величине:

$$K_{ст\ макс} \doteq U_2 / I_1 r_д. \qquad (9.85б)$$

Пусть значения U_2 и $r_д$ такие же, как и в предыдущем примере (0,7 В и 30 Ом), и пусть $I_1 = 2$ мА. Тогда $K_{ст\ макс} \approx 12$, т.е. несравненно меньше, чем при использовании источника тока.

Температурная чувствительность выходного напряжения у диодных стабилизаторов определяется температурной чувствительностью диодов. Комбинируя стабилитроны, работающие в режиме лавинного пробоя (у которых ТКН > 0) с диодами, работающими при прямом смещении (у которых ТКН < 0), можно обеспечить температурную чувствительность до 0,01%/°С и менее.

Транзисторные стабилизаторы. Принципиальная схема простейшего транзисторного стабилизатора напряжения показана на рис. 9.32, *а*, а его малосигнальная эквивалентная схема — на рис. 9.32, *б*. Стабилизатор имеет структуру эмиттерного повторителя: нагрузка включена в цепь эмиттера, а на базу вместо переменного сигнала подано постоянное *опорное напряжение* $U_{оп}$. Источником опорного напряжения обычно служит диодный стабилизатор.

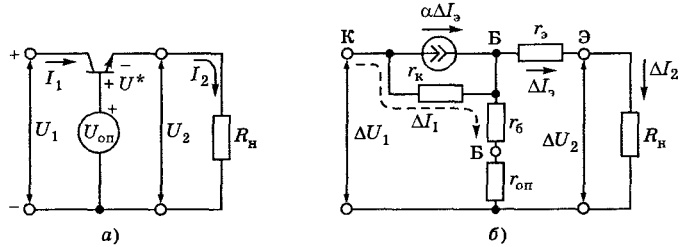


Рис. 9.32. Принципиальная (*а*) и эквивалентная (*б*) схемы транзисторного стабилизатора на основе эмиттерного повторителя

Из рисунка следует: $U_2 = U_{оп} - U^*$, т.е. выходное напряжение определяется значением опорного напряжения.

Если пренебречь базовым током, то входной и выходной токи практически одинаковы: $I_к \approx I_2$. Следовательно, увеличение тока нагрузки влечет такое же увеличение коллекторного тока и тем самым — мощности, рассеиваемой в транзисторе. Очевидно, что короткое замыкание на выходе недопустимо, так как приводит к перегрузке транзистора. Этот вывод относится ко всем стабилизаторам *последовательного типа*, у которых регулирующий элемент включен последовательно с нагрузкой. Во избежание перегрузки мощные стабилизаторы последовательного типа снабжают специальной системой защиты.

Выходное сопротивление данного стабилизатора такое же, как у эмиттерного повторителя. Учитывая, что роль сопротивления $R_г$ играет малое сопротивление опорного элемента $r_{оп}$, можно принять выражение (9.64):

$$R_{вых} = r_э. \tag{9.86}$$

Отсюда следует, что выходное сопротивление растет с уменьшением тока и по мере приближения к режиму холостого хода может достигать неприемлемых значений. Чтобы сгладить зависимость выходного сопротивления от тока нагрузки, можно включить параллельно нагрузке (до выходных зажимов) постоянный шунт. Этот шунт обеспечит некоторый остаточный эмиттерный ток даже при холостом ходе (т.е. при $I_2 = 0$).

Если пренебречь сопротивлением коллекторного перехода, т.е. положить $r_к = \infty$, то приращения ΔU_1 не будут проникать в базовую и эмиттерную цепи транзистора, поскольку в коллекторной цепи находится идеальный источник тока (рис. 9.32, *б*). В этом случае $K_{ст} = \infty$.

С учетом сопротивления $r_к$ приращение выходного напряжения ΔU_2 в зависимости от ΔU_1 можно записать в виде следующих соотношений (рис. 9.32, *б*)[1]:

$$\Delta U_2 = \Delta U_{б'} \frac{R_н}{R_н + r_э} \approx \Delta U_{б'};$$

$$\Delta U_{б'} = \Delta U_1 \frac{r_б + r_{оп}}{r_к + r_б + r_{оп}} \approx \Delta U_1 \frac{r_б + r_{оп}}{r_к},$$

где в правых частях стоят коэффициенты передачи соответствующих резистивных делителей напряжения.

Подставляя отношение $\Delta U_2/\Delta U_1$ в (9.79), получаем коэффициент стабилизации:

$$K_{ст} = \frac{U_2}{U_1} \frac{r_к}{r_б + r_{оп}}. \tag{9.87}$$

Пусть, например, $U_2/U_1 = 0{,}8$; $r_к = 250$ кОм; $r_б = 150$ Ом; $r_{оп} = 20$ Ом. Тогда $K \approx 1200$.

---

1 Источник тока $\alpha\Delta I_э$ считается бездействующим (отсутствующим), так как приращение $\Delta I_э$ заведомо очень мало (поскольку $U_2 \approx$ const)

Таким образом, *простой однотранзисторный стабилизатор обеспечивает достаточно высокий коэффициент стабилизации, но не позволяет кардинально решить проблему уменьшения выходного сопротивления.* Современное решение этой проблемы с использованием операционного усилителя рассмотрено в гл. 10.

9.11. Стабилизаторы тока

Выше не раз использовались источники тока в виде условных элементов электрических цепей. Стабилизаторы тока, рассматриваемые в данном разделе, представляют собой реальное схемное воплощение таких источников.

Задача стабилизаторов тока — поддерживать неизменным выходной ток I_2 при изменениях нагрузки $R_н$ и входного (питающего) напряжения U_1 (рис. 9.33). В дальнейшем напряжение на стабилизирующем элементе обозначается через $U_т$.

Параметры стабилизаторов. Идеальный стабилизатор тока имеет горизонтальную вольт-амперную характеристику, показанную на рис. 9.34 штриховой линией *1*. Характеристики реальных стабилизаторов (кривая *2*) отличаются от идеальной ограниченным рабочим участком и конечным дифференциальным сопротивлением R_i на рабочем участке. Как правило, сопротивление R_i зависит от напряжения $U_т$. Однако учет такой нелинейности не оправдан на практике. Поэтому под R_i всегда понимают некоторую усредненную величину.

Рабочий участок ограничен точками *a* и *b*, в которых наступает заметное уменьшение сопротивления R_i. У биполярных транзисторов точка *a* соответствует переходу в режим насыщения, а у МДП-транзисторов — на крутой участок характеристики. Точка *b* у обоих типов транзисторов соответствует началу пробоя.

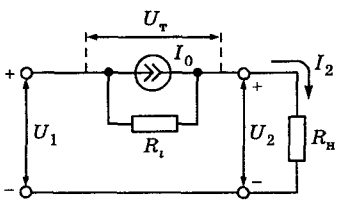


Рис. 9.33 Скелетная схема стабилизатора тока

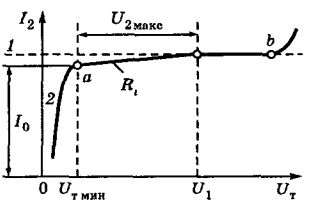


Рис. 9.34. Вольт-амперная характеристика стабилизатора тока

Напряжение $U_{\text{т мин}}$ в точке *а* есть то минимальное напряжение на стабилизирующем элементе, при котором он еще выполняет свои функции. У биполярных *n*–*p*–*n*-транзисторов минимальное напряжение можно считать равным 0,2 В для включения ОЭ и −0,5 В для включения ОБ (см. рис. 5.14, *а* и 5.15, *а*). У МДП-транзисторов минимальное напряжение близко к напряжению насыщения $U_{\text{сн}}$ (см. рис. 4.7, *а*). Последнее, в зависимости от тока и удельной крутизны транзистора, лежит в пределах от долей вольта до нескольких вольт.

В большинстве электронных схем нагрузка стабилизаторов тока нелинейна. Для нелинейных элементов сопротивление является лишь условным символом, обозначающим отношение напряжения к току. Поэтому нелинейную выходную цепь удобнее характеризовать не сопротивлением $R_{\text{н}}$, а выходным напряжением U_2 при заданном токе.

Номинальным режимом стабилизатора тока является режим короткого замыкания, когда $U_2 = 0$. При этом, как видно из рис. 9.33, на стабилизирующем элементе падает максимальное напряжение, равное напряжению U_1. Разумеется, напряжение U_1 должно лежать в пределах рабочего участка (рис. 9.34).

Ток короткого замыкания считается номинальным током стабилизатора. Этот ток складывается из тока I_0 и тока U_1/R_i, обусловленного конечным сопротивлением стабилизирующего элемента (см. рис. 9.33). Во всех реальных стабилизаторах вторая составляющая пренебрежимо мала, так что номинальный выходной ток I_2 считают равным току идеального стабилизирующего элемента I_0.

В режиме, отличном от короткого замыкания, на выходе стабилизатора имеется конечное напряжение U_2. Чем оно больше, тем меньше напряжение $U_{\text{т}}$ на стабилизирующем элементе. Максимальное напряжение на нагрузке соответствует значению $U_{\text{т мин}}$.

Величины I_0, U_1 и U_2 в общем случае некоррелированы, т.е. меняются независимо друг от друга. Поэтому для наихудшего случая приращение выходного тока является арифметической суммой трех приращений (рис. 9.33):

$$\Delta I_2 = \Delta I_0 + \frac{\Delta U_1}{R_i} + \frac{\Delta U_2}{R_i}.$$

Деля левую часть на ток I_2, а правую — на близкий к нему ток I_0, запишем относительную нестабильность в виде

$$\frac{\Delta I_2}{I_2} = \frac{\Delta I_0}{I_0} + \frac{\Delta U_1}{E_i} + \frac{\Delta U_2}{E_i}. \qquad (9.88)$$

Здесь $E_i = I_0 R_i$ — эквивалентное напряжение, определяющее качество стабилизации по выходному и входному напряжениям. Чем больше значение E_i, тем выше качество стабилизации.

На практике роль стабилизирующих элементов выполняют транзисторы. В случае биполярных транзисторов, включенных по схеме ОБ, ток I_0 является коллекторным током, а сопротивление R_i — сопротивлением коллекторного перехода $r_к$. Как известно [см. (5.44)], сопротивление $r_к$ обратно пропорционально току $I_к$. Поэтому их произведение, т.е. напряжение E_i, не зависит от тока и может рассматриваться как параметр транзистора. Типичные значения этого параметра составляют 2–3 кВ. Для включения ОЭ, при котором сопротивление $r_к^* = r_к(1-\alpha)$, значения E_i в десятки раз меньше.

В случае МДП-транзисторов произведение $I_c r_c$ согласно (4.20) тоже является параметром. Типичные значения E_i составляют 200–500 В.

Во многих схемах в качестве стабилизирующего элемента используют обыкновенный резистор (см. рис. 8.15). При анализе этого варианта нужно на рис. 9.33 положить $I_0 = 0$. Тогда

$$I_2 = (U_1 - U_2)/R_i.$$

Для того чтобы стабилизированный ток слабо зависел от выходного напряжения, должно выполняться условие $U_2 << U_1$; при этом $I_2 \approx U_1/R_i$.

Относительная нестабильность по аналогии с (9.88) имеет вид:

$$\frac{\Delta I_2}{I_2} = \frac{\Delta U_1}{U_1} + \frac{\Delta U_2}{U_1}.$$

Значение питающего напряжения U_1 обычно не превышает 10–15 В, т.е. оно несравненно меньше эквивалентного напряжения E_i у нелинейных стабилизирующих элементов [см. (9.88)]. Соответственно при прочих равных условиях нестабильность тока будет во много раз больше.

Простейшие стабилизаторы. На рис. 9.35, *а* показан типовой вариант стабилизатора тока, в котором E_0 — источник стабилизированного напряжения. В данном случае такой источник выполнен на стабилитроне, питаемом через балластное сопротивление (рис. 9.31).

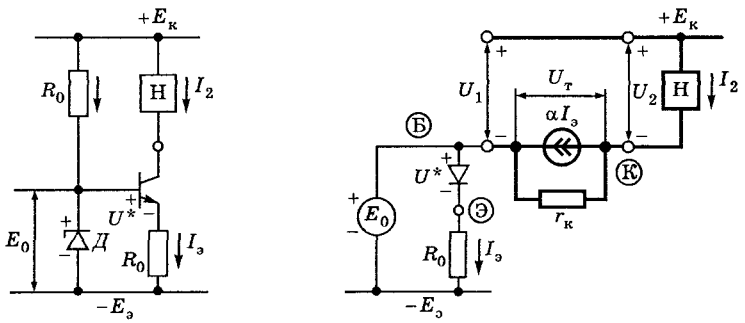


Рис. 9.35. Принципиальная (*а*) и эквивалентная (*б*) схемы простейшего стабилизатора тока на *n*–*p*–*n*-транзисторе

На рис. 9.35, *б* показана эквивалентная схема стабилизатора, где выходная цепь обведена жирными линиями. Обозначения основных величин те же, что и на рис. 9.33; элемент Н символизирует нелинейную нагрузку. Сравнивая выходную цепь на рис. 9.35, *б* со схемой на рис. 9.33, получаем:

$$I_0 = \alpha I_э = \alpha \frac{E_0 - U^*}{R_0}; \qquad (9.89а)$$

$$R_i = r_к; \qquad (9.89б)$$

$$U_1 = E_к + E_э - E_0. \qquad (9.89в)$$

Из выражений (9.89) можно сделать следующие выводы.

Стабилизированный ток I_0 задается цепью E_0, R_0. Стабильность тока определяется в первую очередь стабильностью величин E_0 и U^*. В частности, если температурные чувствительности этих величин имеют разные знаки, то температурная чувствительность тока будет превышать каждую из них.

Внутреннее сопротивление R_i возрастает с уменьшением тока [см. (5.44)]. Значение R_i, приведенное в (9.89б), соответствует абсолютно неизменному току эмиттера. В реальных случа-

ях, когда сопротивление R_0 имеет конечную величину, приращение ΔI_2 распределяется между эмиттерной и базовой цепями. В эмиттерной цепи получается приращение тока

$$\Delta I_э = \gamma_э \Delta I_2,$$

где коэффициент $\gamma_э$ определяется формулой (9.14). Роль сопротивления $R_г$ в данном случае играет сопротивление опорного элемента $Д$ на рис. 9.35, *а*. Таким образом, приращение выходного тока складывается из двух составляющих:

$$\Delta I_2 = \Delta U_2 / r_к + \alpha\gamma_э \Delta I_2.$$

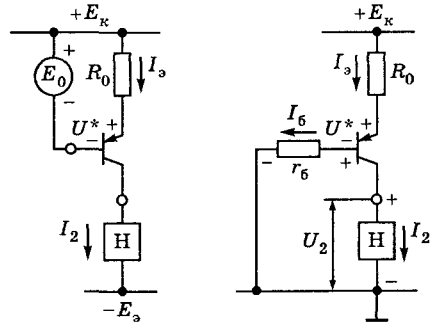
Отсюда легко получается выходное сопротивление в общем виде:

$$R_i = \Delta U_2 / \Delta I_2 = r_к(1 - \alpha\gamma_э). \qquad (9.90)$$

Максимальное выходное напряжение в данной схеме может быть близким к величине U_1 [см. (9.89в)], поскольку транзистор сохраняет усилительные свойства вплоть до значений $U_{кб}$, близких к нулю.

Нестабильность тока I_2 можно оценить по общей формуле (9.88), используя выражения (9.89).

Если нагрузка присоединяется не к положительному, а к отрицательному источнику питания или «к земле», то стабилизатор строится на основе p–n–p-транзистора (рис. 9.36, *а*). Эта схема аналогична предыдущей и характеризуется теми же выражениями.



а) *б*)

Рис. 9.36. Простейшие стабилизаторы тока на p–n–p-транзисторе: *а* — основной вариант, *б* — вариант с заземленной базой

На практике встречается вариант стабилизатора, показанный на рис. 9.36, *б*. Этот вариант характерен однополярным питанием и нулевым потенциалом базы (с точностью до небольшого падения напряжения $I_б r_б$). В результате потенциал коллектора оказывается положительным относительно базы, т.е. коллекторный переход смещен в прямом направлении. При этом формально транзистор работает в режиме двойной инжекции. Однако, как известно, небольшие прямые смещения коллекторного перехода (меньше, чем U^* – 0,1 В) не приводят к заметному уменьшению коллекторного тока. Следовательно, в области достаточно малых выходных напряжений (десятые доли вольта) данный стабилизатор работает нормально.

Отражатели тока. Широкое распространение в аналоговых интегральных схемах получил стабилизатор тока, известный под названием «токовое зеркало» или *отражатель тока* (рис. 9.37). Легко установить формальное сходство между схемой отражателя тока и схемой простейшего стабилизатора: вместо стабилитрона (рис. 9.35, *а*) в данном случае используются резистор R_1 и прямосмещенный p–n-переход; роль последнего играет транзистор $T1$, включенный по диодной схеме БК–Э. Такая трансформация источника E_0 приводит к повышению гибкости схемы и улучшению ряда параметров.

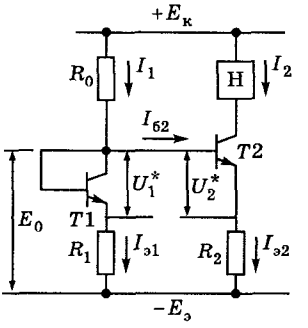


Рис. 9.37. Отражатель тока

Из рис. 9.37 следует равенство

$$U_1^* + I_{э1}R_1 = U_2^* + I_{э2}R_2. \qquad (9.91)$$

На этом равенстве основывается работа отражателя тока.

Омические слагаемые в (9.91) обычно не превышают значений U^*. Поэтому, в зависимости от рабочих токов, сопротивления R_1 и R_2 лежат в пределах от сотен ом до 10–20 кОм.

Пренебрежем для простоты малым током $I_{б2}$; тогда $I_{э1} = I_1$ и $I_{э2} = I_2$. Кроме того, предположим, что сопротивления R_1 и R_2 одинаковы, а транзисторы $T1$ и $T2$ идентичны (в ИС идентичность транзисторов легко обеспечивается благодаря близкому расположению элементов на одном кристалле). При таких

условиях слагаемые в левой и правой частях равенства (9.91) будут одинаковы, откуда следует: $I_2 = I_1$.

Значит, в рассмотренном варианте схемы выходной ток I_2 повторяет или отражает входной ток I_1, откуда и название «отражатель» тока.

Входной ток согласно рис. 9.37 выражается следующим образом:

$$I_1 = (E_к - E_0)/R_0.$$

Если $E_к >> E_0$, то ток I_1 определен внешними параметрами $E_к$ и R_0. Во многих случаях ток I_1 поступает от тех или иных каскадов, входящих в состав сложного устройства, а значит, может считаться величиной заданной.

Если сопротивления R_1 и R_2 сделать неодинаковыми, то неодинаковыми будут и эмиттерные токи. Поскольку напряжение $U_{эб}$ слабо зависит от тока, будем по-прежнему считать $U_1^* = U_2^*$. Тогда из (9.91) следует:

$$I_2 = I_1(R_1/R_2). \qquad (9.92)$$

Как видим, ток I_2 может «отражать» ток I_1 как в «увеличенном», так и в «уменьшенном» масштабе, в зависимости от соотношения сопротивлений. Этот масштаб обычно не превышает нескольких единиц, в противном случае резистор с большим номиналом занимает слишком значительную площадь.

Из (9.92) следует, что *выходным током I_2 можно управлять*, меняя входной ток I_1 тем или иным способом. Такая возможность — одно из проявлений гибкости отражателя тока.

Учет тока $I_{б2}$ приводит к тому, что эмиттерные токи не точно равны токам I_1 и I_2. А именно,

$$I_{э1} = I_1 - I_{б2}; \qquad I_{э2} = I_2 + I_{б2}.$$

При этом в выражение (9.92) нужно внести соответствующие поправки. При больших значениях B, когда $I_{б2} << I_{э2}$, такие поправки несущественны.

Необходимо подчеркнуть, что в выражение (9.92) не входят ни напряжение U^*, ни коэффициент усиления B. Это значит, что работа отражателя тока в первом приближении не зависит от изменения этих параметров, т.е. прежде всего — от изменений температуры. На самом деле такая зависимость (хотя и весьма слабая) имеет место, поскольку коэффициент B не войдет в вы-

ражение (9.92) при учете тока $I_{б2}$, а напряжение U^* — при учете разности $U_1^* - U_2^*$, обусловленной различием эмиттерных токов.

Чтобы обеспечить особо малые выходные токи (например, при работе ДУ в микрорежиме) сопротивление R_1 делают равным нулю. Для такого варианта отражателя тока (рис. 9.38, *а*) формула (9.92) недействительна, так как пренебрегать различием напряжений U_1^* и U_2^* в этом случае нельзя. Воспользуемся для напряжений U_1^* и U_2^* выражением (5.38б). Тогда из (9.91) при $R_1 = 0$ нетрудно получить следующую трансцендентную зависимость между выходным и входным токами:

$$I_2 = (\varphi_T / R_2)\ln(I_1/I_2). \qquad (9.93а)$$

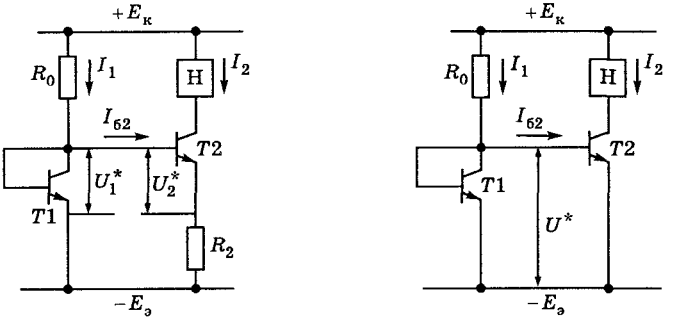


Рис. 9.38. Варианты отражателей тока:
а — однорезисторный; *б* — безрезисторный

Более наглядной является аппроксимация, выражающая связь токов в явном виде:

$$I_2 \approx \sqrt{\left(\frac{\varphi_T}{R_2}\right) I_1}. \qquad (9.93б)$$

Как видим, в данной схеме зависимость тока I_2 от I_1 значительно слабее, чем в предыдущей, т.е. управление выходным током затруднительно.

Выражения (9.93) легко использовать для расчета необходимого сопротивления R_2, если задаться желательными значениями токов. Пусть, например, $I_1 = 0{,}5$ мА и $I_2 = 10$ мкА; тогда из (9.93а) следует: $R_2 \approx 10$ кОм. При этом падение напряжения $I_2 R_2$ (т.е. разность $U_1^* - U_2^*$) составит около 100 мВ.

Из выражений (9.93) видно, что зависимость отношения токов от коэффициента В по-прежнему отсутствует, но имеет место непосредственная зависимость тока I_2 от температуры через температурный потенциал φ_T. Для уменьшения этой зависимости желательно, чтобы резистор R_2 имел такой же температурный коэффициент, как и φ_T, а именно 0,33 %/ °С. Такие значения ТКС легко обеспечить в интегральных резисторах.

Еще один вариант отражателя тока (рис. 9.38, *б*) характерен отсутствием резисторов, т.е. минимальной площадью на кристалле. Однако если $I_1 \neq I_2$, то один из транзисторов приходится делать с большей площадью, и тогда экономия по сравнению с резисторным вариантом в значительной мере утрачивается.

Полагая $R_1 = R_2 = 0$ в выражении (9.91) и подставляя значения U_1^* и U_2^* из (5.38б), легко убедиться, что отношение токов I_2/I_1 пропорционально отношению тепловых токов $I_{э02}/I_{э01}$. Тепловые токи при прочих равных условиях пропорциональны площадям переходов. В интегральных схемах «прочие равные условия» (т.е. идентичность электрофизических параметров) обеспечиваются близким расположением транзисторов. Учитывая все сказанное, можно записать для схемы на рис. 9.38, *б*:

$$I_2 = I_1(S_2/S_1), \tag{9.94}$$

где S_1 и S_2 —площади эмиттерных переходов.

В данном варианте, как и в основном (рис. 9.37), влияние изменений коэффициента В и напряжения U^* в пером приближении отсутствует. В небольшой степени такое влияние имеет место, если учесть ток $I_{б2}$, о чем говорилось выше. Чем больше площадь S_2, а значит и ток I_2, тем больше базовый ток и тем больше погрешность при использовании выражения (9.94). На практике отношение токов и площадей редко превышает несколько единиц и тогда погрешность лежит в пределах до 2–5 %.

В заключение отметим, что выходное сопротивление R_i в отражателе тока может заметно отличаться от сопротивления $r_к$, которое было принято для простейшего стабилизатора [см. (9.89б)]. Дело в том, что у простейшего стабилизатора сопротивление эмиттерной цепи всегда намного больше, чем базовой, поскольку в эмиттерную цепь включено достаточно большое сопротивление $R_э$. Поэтому коэффициент токораспределения $\gamma_э$ в формуле (9.90) обычно не превышает 0,1–0,2. В отражателях тока сопротивления эмиттерной и базовой цепей транзистора

$T2$ могут находиться в разных отношениях, в частности, может иметь место неравенство $R_б > R_э$. В таком случае выходное сопротивление транзистора $T2$, а значит, и внутреннее сопротивление отражателя тока следует рассчитывать по формуле (9.90).

Например, в схеме на рис. 9.38, *б* под $R_б$ следует понимать сопротивление $r_{э1}$, а под $R_э$ — сопротивление $r_{э2}$. Оба они согласно (5.43) обратно пропорциональны соответствующим эмиттерным токам, практически — токам I_1 и I_2. Поэтому, если для простоты положить $\alpha = 1$ и $r_{б2} = 0$, то внутреннее сопротивление (9.90) после простых преобразований принимает вид:

$$R_i \approx \frac{r_к}{1 + I_2/I_1}.$$

Как видим, с ростом отношения I_2/I_1 сопротивление R_i существенно уменьшается. Этот вывод носит общий характер.

Контрольные вопросы

1. Опишите признаки, по которым могут быть классифицированы электронные усилители.
2. Что такое составной транзистор? Опишите принцип действия и характеристики пары Дарлингтона.
3. Опишите принцип действия и характеристики составной транзисторной структуры на транзисторах разной проводимости.
4. Какие параметры определяют рабочую точку простейшего усилительного каскада на биполярном транзисторе?
5. Перечислите дифференциальные параметры усилителя. Как практически измеряются эти параметры?
6. При каком условии при расчете коэффициента усиления простейшего усилительного каскада можно пренебречь сопротивлением генератора и сопротивлением базы транзистора?
7. Чем определяется выходное сопротивление простейшего усилительного каскада в области низких частот?
8. Чем определяется постоянная времени переходного процесса простейшего усилителя в области высоких частот?

9. Опишите принцип действия и характеристики простейших усилительных каскадов на МДП-транзисторах.
10. Какими параметрами характеризуется реальный дифференциальный усилитель? Что такое синфазная и дифференциальная составляющие входного сигнала?
11. Чем обусловлен более высокий коэффициент усиления дифференциального усилителя по сравнению с простейшим усилительным каскадом?
12. Опишите принцип действия и характеристики эмиттерного повторителя. От каких параметров зависит коэффициент передачи эмиттерного повторителя?
13. Для каких целей используются эмиттерные повторители, как соотносятся их входное и выходное сопротивления?
14. Опишите особенности переходного процесса в эмиттерном повторителе при скачкообразном увеличении и уменьшении входного сигнала.
15. Как реализуются схемы сдвига уровней в многокаскадных усилителях?
16. Что такое каскод? Для минимизации какого эффекта используется данная структура?
17. Опишите схемы, принцип действия, режимы и основные характеристики выходных усилительных каскадов.
18. Какими параметрами характеризуются стабилизаторы напряжения? Что такое стабилизаторы параллельного и последовательного типа?
19. Какими параметрами характеризуются стабилизаторы тока? Опишите схемы и принцип действия простейших стабилизаторов.
20. Опишите принцип работы и характеристики «токового зеркала» на биполярных и МДП-транзисторах. Для каких целей используются такие каскады?

Глава 10

Интегральные схемы

10.1. Введение

В предыдущих главах мы подробно познакомились с физическими, конструктивно-технологическими и схемотехническими основами микроэлектроники. В данной заключительной главе рассмотрим примеры простейших ИС и некоторые проблемы, связанные с их разработкой.

Первыми, появившимися в 1960 г., были простейшие цифровые ИС, выполненные на биполярных транзисторах. В середине 60-х годов началось развитие аналоговых ИС на биполярных транзисторах и цифровых ИС на МДП-транзисторах. Первая половина 70-х годов ознаменовалась появлением некоторых новых, специфичных для микроэлектроники схемных решений (приборы с зарядовой связью, схемы с инжекционным питанием и др.), резким возрастанием степени интеграции, а также установлением своего рода «динамического равновесия» между основными классами ИС — биполярными и МДП-транзисторными, полупроводниковыми и гибридными, — которые до этого нередко рассматривались как альтернативные. 80-е и начало 90-х годов характерны главным образом дальнейшим ростом степени интеграции и превращением ИС, по сложности выполняемых функций, в крупные аппаратурные узлы. Существенное развитие получили аналого-цифровые ИС, совмещающие функции преобразования и обработки цифровых и аналоговых сигналов.

Подробное описание современных больших ИС — их структуры, принципов построения, номенклатуры, назначения и т.п. — в рамках данной книги не только невозможно, но и нецелесообразно. Поэтому ниже дается представление лишь об элементарных «кирпичиках», из которых складываются современные цифровые ИС и которые во многом определяют их качество. Что касается аналоговых ИС, то их особенности лучше всего характеризуются на примере операционного усилителя.

10.2. Логические элементы на биполярных транзисторах

Логическими элементами или *логическими вентилями* называют электронные схемы, выполняющие простейшие логические операции. Прежде чем рассматривать схемные варианты логических элементов, остановимся на функциях, которые они выполняют.

Логические функции. Логические функции и логические операции над ними — предмет *алгебры логики* или булевой алгебры. В основе алгебры логики лежат логические величины, которые будем обозначать A, B, C и т.д. Логическая величина характеризует два взаимоисключающие понятия: есть и нет, черное и нечерное, включено и выключено и т.п. Если одно из значений логической величины обозначено через A, то второе обозначают через $\overline{A}$ (не A).

Для операций с логическими величинами удобно использовать двоичный код, полагая $A = 1$, $\overline{A}=0$ или, наоборот, $A = 0$, $\overline{A} = 1$. При этом одна и та же схема может выполнять как логические, так и арифметические операции (в двоичной системе счисления).

Если понятие «не A» обозначить особой буквой, например B, то связь между значениями B и A будет иметь вид:

$$B = \overline{A}. \qquad (10.1)$$

Это — простейшая логическая функция, которую называют *отрицанием*, *инверсией* или *функцией НЕ*. Схему, обеспечивающую выполнение такой функции, называют инвертором или схемой НЕ. Ее условное обозначение показано на рис. 10.1, *а*. Функция инверсии характеризуется кружком на выходной стороне прямоугольника.

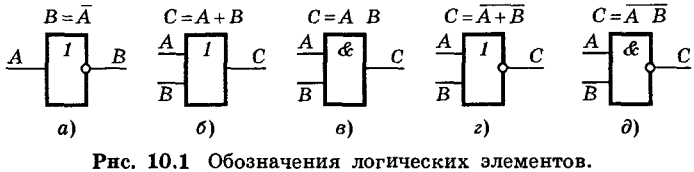


Рис. 10.1 Обозначения логических элементов.
а — НЕ, *б* — ИЛИ, *в* — И, *г* — ИЛИ-НЕ, *д* — И-НЕ

Функция отрицания является функцией одного аргумента (одной переменной). Приведем примеры логических функций двух аргументов.

Логическое сложение, дизъюнкция или функция ИЛИ:

$$C = A + B. \qquad (10.2)$$

Эта функция определяется следующим образом: $C = 1$, если $A = 1$ или $B = 1$ или и $A = 1$, и $B = 1$. Обозначение схемы ИЛИ показано на рис. 10.1, *б*.

Логическое умножение, конъюнкция или функция И:

$$C = AB. \tag{10.3}$$

Эта функция определяется следующим образом: $C = 1$, только если одновременно и $A = 1$, и $B = 1$. Обозначение схемы И показано на рис. 10.1, *в*.

Сочетание функции ИЛИ с инверсией (рис. 10.1, *г*) приводит к комбинированной *функции ИЛИ–НЕ*:

$$C = \overline{A + B}. \tag{10.4}$$

Аналогично, сочетание функции И с инверсией (рис. 10.1, *д*) приводит к комбинированной *функции И–НЕ*:

$$C = \overline{AB}. \tag{10.5}$$

Функции ИЛИ–НЕ и И–НЕ — самые распространенные, так как на их основе можно реализовать любую другую логическую функцию. Разумеется, количество аргументов, а значит, и количество входов у соответствующих схем может быть равно трем, четырем и более.

В схемах, реализующих логические функции, т. е. в логических элементах, логические нули и единицы обычно представлены разными значениями напряжения: напряжением или *уровнем нуля* U^0 и напряжением или *уровнем единицы* U^1.

Если уровень единицы больше уровня нуля, говорят, что схема работает в «положительной логике», в противном случае она работает в «отрицательной логике». Никакой принципиальной разницы между положительной и отрицательной логиками нет. Более того, как будет показано ниже, одна и та же схема может работать и в той, и в другой логике. В дальнейшем будем считать логику положительной, что соответствует ее практической распространенности.

Разность уровней единицы и нуля называют *логическим перепадом*:

$$U_л = U^1 - U^0. \tag{10.6}$$

Естественно, что логический перепад должен быть достаточно большим, чтобы «единицы» и «нули» четко отличались друг от друга и чтобы случайные помехи не «превращали» один уровень в другой.

Интегральные логические элементы (ИЛЭ) составляют основу или, как говорят, *элементную базу* более сложных ИС и аппаратуры в целом. При этом параметры ИЛЭ оказывают непосредственное влияние на параметры узлов и субсистем. Иначе говоря, выбор типа ИЛЭ в значительной мере предопределяет качественные показатели аппаратуры.

Схемные варианты ИЛЭ принято называть *транзисторными логиками* и обозначать буквами ТЛ с некоторыми дополнениями, характеризующими специфику того или иного варианта. Однако такая система обозначений выдерживается не строго.

Резисторно-транзисторная логика (РТЛ). В основе ИЛЭ этого типа лежит параллельное соединение обычных транзисторных ключей и использование общей коллекторной нагрузки. Элемент РТЛ показан на рис. 10.2 сплошными линиями. Штриховыми линиями показаны транзисторы, входящие в состав других аналогичных логических элементов. Предыдущие ИЛЭ, представленные транзисторами $T3$ и $T4$, управляют ключами $T1$ и $T2$, а транзистор $T5$ представляет последующие логические элементы, являющиеся нагрузкой для данного.

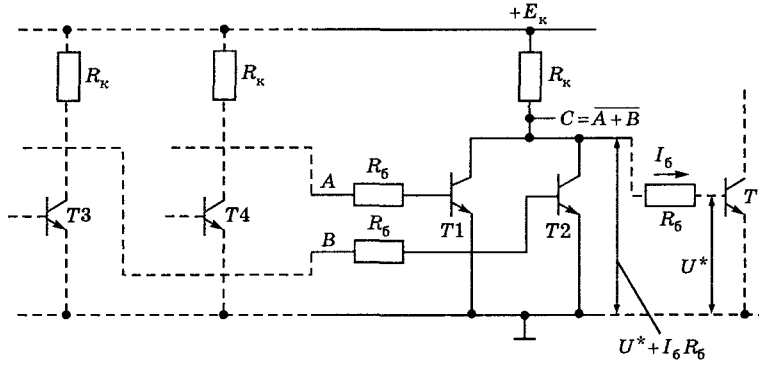


Рис. 10.2. Логический элемент РТЛ

Нетрудно убедиться, что в положительной логике схема выполняет функцию ИЛИ–НЕ. Действительно, если на оба входа A и B подан низкий уровень $U^0 \approx 0$, то оба транзистора заперты, ток течет через резистор $R_к$ в базу $T5$ и на выходе ИЛЭ получается напряжение

$$U^1 = E_к - I_б R_к. \qquad (10.7а)$$

Если же на одном из входов ИЛЭ действует высокий уровень $U^1 = U^*$, то соответствующий транзистор открыт, а при достаточно большом токе базы — насыщен. Тогда на выходе ИЛЭ получаем низкий уровень, равный остаточному напряжению:

$$U^0 = U_{ост}. \qquad (10.7б)$$

Такой же уровень получается при отпирании обоих транзисторов. Таким образом, удовлетворяется формула (10.4).

В отрицательной логике ($U^0 = U^*$; $U^1 = U_{ост}$) схема выполняет функцию И–НЕ: высокий уровень на выходе (U^0) получается только при условии, что оба транзистора заперты, т.е. на оба входа поступили низкие уровни (U^1). *Изменение характера выполняемой функции при замене положительной логики на отрицательную (и наоборот) является общим свойством ИЛЭ*, и мы не будем в дальнейшем иллюстрировать его для каждой схемы.

Учитывая приведенные выше значения логических уровней, можно записать логический перепад для схемы РТЛ в виде:

$$U_л = E_к - I_б R_б - U_{ост}. \qquad (10.8)$$

Следует отметить, что логическая функция в РТЛ-элементе выполняется «по выходу», т.е. сначала осуществляется инверсия входных сигналов на обычных транзисторных ключах, затем на нагрузочном резисторе $R_к$ реализуется логическая функция И. В такой интерпретации логическая функция данного элемента $C = \overline{A} \cdot \overline{B}$. Данная запись эквивалентна функции $C = \overline{A + B}$, которая обычно и используется для описания логической функции РТЛ-элемента.

Схемы РТЛ использовались на первом этапе развития микроэлектроники. Однако в полупроводниковых ИС с высокой степенью интеграции они оказались неперспективными в связи с большим количеством резисторов, занимающих большую площадь, и низким быстродействием.

Транзисторно-транзисторная логика (ТТЛ). Схемы элементов ТТЛ показаны на рис. 10.3. В них логическая функция осуществляется многоэмиттерным транзистором T_0 (рис. 10.3, *а*) или транзисторами T_{01} и T_{02} (рис. 10.3, *б*), а транзистор $T1$ выполняет функцию инверсии.

В современных интегральных схемах используют оба варианта входных каскадов (рис. 10.3, *а* и *б*). ТТЛ-элемент с раздель-

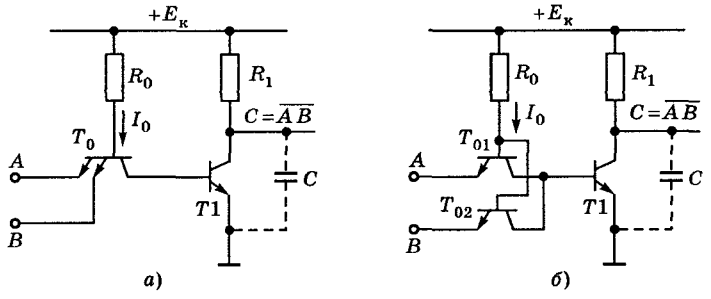


Рис. 10.3. Логические элементы ТТЛ: *а* — использующий входной многоэмиттерный транзистор; *б* — использующий раздельные входные транзисторы

ными входными транзисторами, хотя и занимает большую площадь, но позволяет исключить ряд нежелательных эффектов, свойственных многоэмиттерным транзисторам, а также обеспечивает большую гибкость при проектировании топологии элемента.

Для описания работы ТТЛ-элемента воспользуемся схемой с раздельными входными транзисторами (рис. 10.3, *б*).

Работа элемента с входным многоэмиттерным транзистором может быть описана в терминах, учитывающих, что многоэмиттерная транзисторная структура получена в результате объединения баз и коллекторов нескольких однотипных транзисторов.

Пусть в исходном состоянии напряжения на логических входах равны нулю: $U_A = U_B = U^0 = 0$. Тогда эмиттеры транзисторов T_{01} и T_{02} находятся под нулевым напряжением, их эмиттерные переходы смещены прямо и ток $I_0 = (E_к - U^*)R_0$ поровну распределяется между базами транзисторов. Входные транзисторы и резистор R_0 работают в состоянии $U_A = U_B = 0$ как обычный транзисторный ключ с «оторванным» коллектором ($R_к = \infty$). Выходное напряжение ключа определяется из (8.2) при условии $I_{кн} = 0$:

$$U_{кэ} = \varphi_T \ln\left(1 + \frac{1}{B_I}\right).$$

Это напряжение не превышает (2–3) φ_T, и учитывая, что оно приложено к эмиттерному переходу выходного транзистора, транзистор $T1$ надежно заперт.

Когда транзистор $T1$ заперт, выходное напряжение имеет максимальную величину, равную напряжению питания. Очевидно, что это есть уровень логической единицы, т. е.

$$U^1 = E_к.$$

Пусть теперь на вход A подано напряжение U^1. Поскольку напряжение на втором входе осталось равным нулю, транзистор T_{02} по-прежнему открыт, а значит, напряжение на базе транзистора $T1$ по-прежнему близко к нулю. Транзистор T_{01} оказался включенным в инверсном активном режиме (к эмиттеру приложено напряжение U^1, близкое к $E_к$, коллектор под потенциалом, близким к нулю, а база под потенциалом прямосмещенного эмиттерного перехода U^* транзистора T_{01}). Можно показать, что такое включение T_{01} приводит к небольшому [на $(1-2)\varphi_T$] возрастанию напряжения на базе $T1$.

Никаких других изменений в схеме не происходит и на выходе по-прежнему имеем уровень $E_к$. То же самое будет, если подать напряжение U^1 на вход B.

Если же подать напряжение U^1 на оба логических входа, то оба транзистора T_{01} и T_{02} оказываются в инверсном активном режиме и ток I_0, усиленный в $(B_I + 1)$ раз, втекает в базу транзистора $T1$. При этом транзистор отпирается, потенциал базы делается равным U^*, а потенциал коллектора (при условии насыщения) равным $U_{ост}$. Это значение есть уровень логического нуля, т.е.

$$U^0 = U_{ост} \approx 0{,}1 \text{ В}.$$

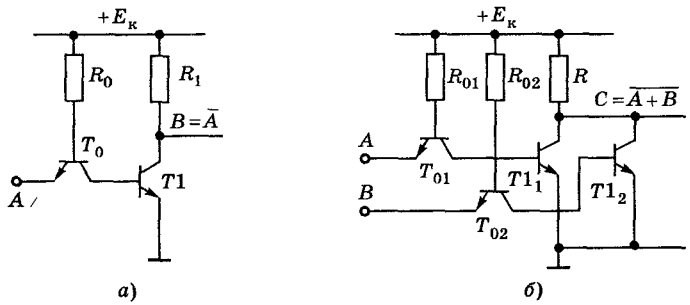
Таким образом, уровень выходного напряжения меняется с U^1 до U^0 только при поступлении уровня U^1 на все входы. Значит, схема ТТЛ в положительной логике выполняет функцию И–НЕ. Это еще одна важная особенность данной схемы по сравнению с ранее рассмотренными, которые выполняли функцию ИЛИ–НЕ. Однако в отрицательной логике схема ТТЛ тоже выполняет функцию ИЛИ–НЕ.

Следует подчеркнуть, что схеме ТТЛ свойствен большой логический перепад, причем в реальной цепочке, где каждый ИЛЭ связан с предыдущим и последующим, он будет таким же, как у изолированной схемы. Для логического перепада получаем:

$$U_л = E_к - U_{ост} \approx E_к.$$

Усиление тока I_0 в (B_I +1) раз (см. выше) является одной из проблем ТТЛ-элементов, так как приращение тока $\Delta I = B_I \cdot I_0$ приводит к значительному возрастанию входного тока при уровне U^1 на входе. Поэтому при разработке ТТЛ-элементов стремятся максимально снизить инверсный коэффициент усиления транзисторов входного каскада ($B_I < 0{,}01$). Наиболее эффективно такое снижение достигается путем шунтирования коллекторных переходов диодами Шоттки (см. разд. 8.4).

На рис. 10.4, *а* показан простой инвертор, осуществленный с помощью одноэмиттерного транзистора, а на рис. 10.4, *б* — схема ИЛИ–НЕ, осуществленная с помощью двух параллельно включенных инверторов. Эти примеры иллюстрируют гибкость и универсальность основной схемы ТТЛ.



а) *б*)

Рис. 10.4. Варианты использования схемы ТТЛ:
а — инвертор; *б* — схема ИЛИ–НЕ

Одним из существенных недостатков простейшей схемы ТТЛ, показанной на рис. 10.3, является ее малая нагрузочная способность: при подключении нескольких нагрузок возрастает суммарная нагрузочная емкость (показана штриховой линией) и постоянная времени CR_1, с которой эта емкость заряжается. Для того чтобы ускорить заряд емкости и повысить нагрузочную способность ИЛЭ, вместо простого инвертора, состоящего из одного транзистора, используется сложный инвертор, состоящий из трех транзисторов и диода (рис. 10.5). Такую схему принято обозначать как ТТЛ-3.

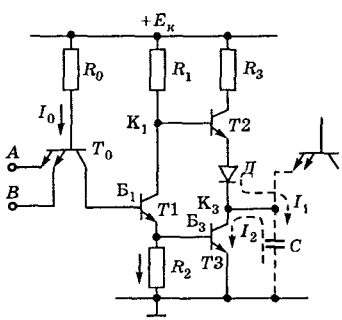


Рис. 10.5. Логический элемент ТТЛ со сложным инвертором

В сложном инверторе транзисторы $T1$ и $T3$ включены по схеме Дарлингтона (см. рис. 9.1.); поэтому их можно рассматривать как одно целое; они вместе запираются и вместе отпираются. Наличие резистора R_2 не принципиально: через него ответвляется часть тока $I_{э1}$, что позволяет уменьшить ток $I_{б3}$, а значит, степень насыщения транзистора $T3$. Совокупность транзисторов $T2$ и $T3$ можно рассматривать как вариант двухтактного выходного каскада (см. рис. 9.25), в котором транзисторы работают поочередно: $T2$ заряжает конденсатор C (ток I_1), а $T3$ ее разряжает (ток I_2). Диод $Д$ обеспечивает запирание транзистора $T2$ при открытом и насыщенном $T3$, а резистор R_3 ограничивает ток во время тех интервалов, когда $T2$ открывается, а $T3$ еще не заперся.

На смену первичным элементам ТТЛ, описанным выше, в конце 70-х годов пришли ТТЛ-микросхемы с диодами Шоттки (ТТЛШ). При этом диоды Шоттки (ДШ) использовались как для минимизации накопления заряда в насыщенных транзисторах, так и в качестве «ускоряющих» емкостей в ключевых элементах. Одна из разновидностей ТТЛШ-элементов использует матрицу диодов Шоттки вместо многоэмиттерного транзистора.

В целом, развитие ТТЛ-элементов определялось совершенствованием процессов их изготовления. Внедрение изопланарной технологии привело к появлению новых вариантов ТТЛШ-элементов, получивших условные названия FAST, AS и ALS. FAST — это начальные буквы слов Fairchild Advanced Shottky TTL. Сокращение AS происходит от слов Advanced (с опережением) и Shottky, а буква L в наименовании ALS обозначает слово Low (низкий), т. е. это маломощный вариант микросхем типа AS.

Интегральная инжекционная логика ($И^2Л$). Схемы $И^2Л$, появившиеся позднее других, не имеют аналогов в дискретных транзисторных схемах, т. е. специфичны именно для интегрального исполнения. Однако по существу они являются еще одной, наиболее совершенной модификацией схемы РТЛ. Имеет смысл пояснить эту преемственность, так как внешне схемы $И^2Л$ выглядят несколько необычно и непосредственно не ассоциируются со схемами РТЛ.

В схемах $И^2Л$ удаляют (закорачивают) резисторы в базовых цепях и заменяют резисторы $R_к$ генераторами постоянного тока I^* (см. рис. 10.6, *а* и 10.2). Такая замена не влияет на принцип действия ИЛЭ. Преимущества, связанные с заменой резистора на генератор тока, выяснятся позднее. А сейчас формально переместим все генераторы тока

«вправо» — к базам нагрузочных транзисторов (см. рис. 10.6, *б*). Очевидно, что это никак не повлияет на работу схемы. Наконец, переключим верхние зажимы генераторов тока с шины $+E_к$ на «земляную» шину (рис. 10.6, *в*). Согласно теории цепей, такое переключение не отразится на режиме транзисторов, поскольку ток генератора тока не зависит от наличия или отсутствия последовательно включенных источников э.д.с. (в данном случае источника $E_к$).

Как видим, *характерной особенностью схем И²Л является индивидуальное питание базы каждого транзистора от «своего» генератора тока*. Часто эти генераторы для простоты опускают, и тогда схема принимает несколько странный вид (рис. 10.6, *г*): в ней как бы отсутствуют источники питания.

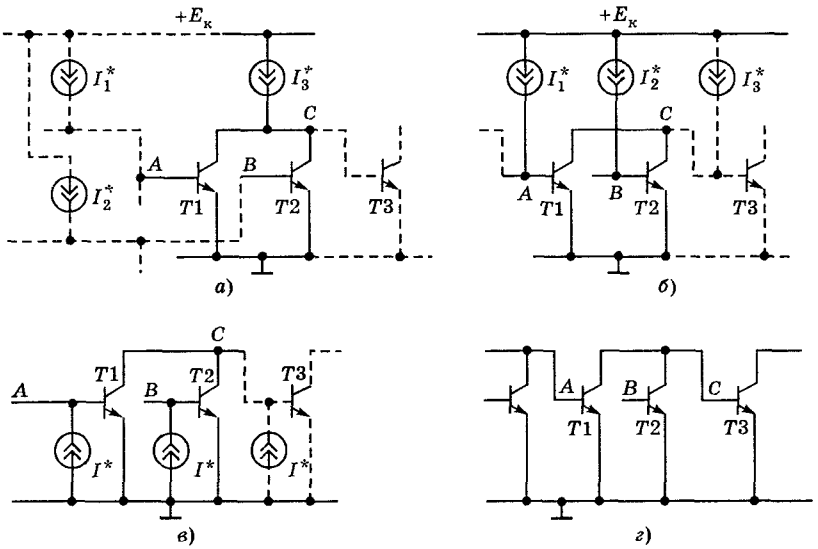


Рис. 10.6. Происхождение логического элемента И²Л: *а* — замена резистора генератором тока в схеме РТЛ; *б* — перемещение генераторов тока в базовые цепи следующих логических элементов; *в* — включение генератора тока между базой и общей «земляной» шиной; *г* — схема И²Л (генераторы тока не обозначены)

Индивидуальные генераторы тока осуществляются с помощью *p–n–p*-транзисторов, включенных по схеме ОБ (рис. 10.7, *а*). Такой вариант известен из разд. 9.11 (см. рис. 9.36, *б*). Как было показано, в таком варианте схемы *p–n–p*-транзистор формально работает в режиме двойной инжекции (насыщения), поскольку потенциал коллектора больше

нуля. Однако, если напряжение $U_к$ хотя бы на 0,1 В меньше, чем напряжение $U_э$, то инжекция коллекторного перехода практически отсутствует и коллекторный ток остается неизменным, равным $\alpha I_э$. В схемах И^{2}Л меньшее значение $U_к$ по сравнению с $U_э$ обеспечивается благодаря малому значению B_I у p–n–p-транзистора.

Из рис. 10.7, *а* видно, что эмиттеры и базы всех p–n–p-транзисторов оказываются соединенными друг с другом. Поэтому все p–n–p-транзисторы можно заменить одним многоколлекторным транзистором, что характерно для схем И^{2}Л (рис. 10.7, б). Характерно также использование многоколлекторных n–p–n-транзисторов.

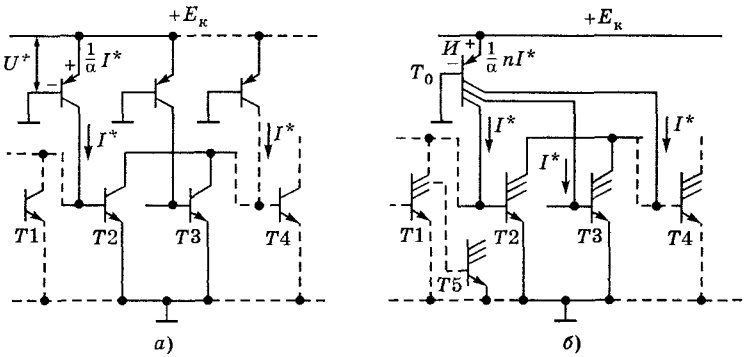


Рис. 10.7. Осуществление генераторов тока с помощью индивидуальных p–n–p-транзисторов (*а*) и одного многоколлекторного p–n–p-транзистора (*б*)

Базы транзисторов $T2$ и $T5$ на рис. 10.7, *б* управляются от разных коллекторов транзистора $T1$. Следовательно, базы «развязаны» друг от друга, а значит, исключена возможность неравномерного распределения тока между ними. В результате, использование многоколлекторных логических транзисторов существенно облегчает осуществление разветвленных связей между ИЛЭ в сложных цифровых ИС.

Термин «инжекционное питание», принятый для данного типа ИЛЭ, оправдан тем, что питающие токи I^* получаются благодаря инжекции дырок через эмиттерный переход p–n–p-транзистора. Эмиттер, выполняющий функцию питания, принято называть *инжектором* и обозначать буквой *И* (рис. 10.7, *б*). Поскольку напряжение $E_к$ подключено непосредственно к инжекторному переходу, получаем $E_к = U^*$. Малое

напряжение питания — одно из важных преимуществ схем И^{2}Л. Легко заметить, что это преимущество есть следствие замены резисторов генераторами тока.

Как известно, подключать напряжение непосредственно к p–n-переходу нежелательно, так как при этом трудно обеспечить стабильное значение тока. Поэтому на практике последовательно с инжектором включают небольшое сопротивление. Тогда напряжение питания несколько превышает U^* и составляет обычно 1–1,5 В.

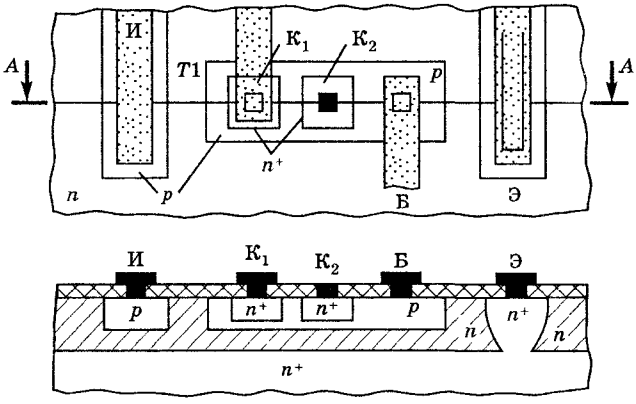
При логическом анализе схем И^{2}Л обычно пользуются эквивалентной схемой на рис. 10.6, *в*, считая, что база транзистора либо замкнута на «землю» (если в предыдущем ИЛЭ один из транзисторов насыщен, $U^0 = U_{ост}$), либо «оборвана» (если в предыдущем ИЛЭ все транзисторы заперты, $U^1 \approx U^*$). В первом случае данный транзистор заперт, а ток I^* протекает через транзисторы предыдущего ИЛЭ. Во втором случае ток I^* полностью протекает в базу данного транзистора и насыщает его. Пусть, например, на рис. 10.7, *б* насыщен транзистор $T2$. Тогда в его базе протекает ток I^* от нижнего коллектора T_0 (так как транзистор $T1$ заперт), а в коллекторе — ток I^* от верхнего коллектора T_0 (так как транзистор $T4$ тоже заперт; на его базе действует малое остаточное напряжение насыщенного транзистора $T2$). Следовательно, $I_{б2} = I_{к2} = I^*$. Условие насыщения (8.4) при этом имеет вид $BI^* \geq I^*$ или

$$B \geq 1.$$

Как видим, *требования к коэффициенту усиления базового тока минимальные*. Обычно это условие выполняется даже в микрорежиме, при токах 5–10 мкА. Это еще одно важное преимущество схем И^{2}Л, которое, наряду с малым напряжением питания, способствует резкому уменьшению потребляемой мощности. Логический перепад для И^{2}Л около 0,7 В.

Оригинальность схемотехнического решения сочетается в схемах И^{2}Л с оригинальностью технологического решения. На рис. 10.8 показана структура И^{2}Л-вентиля.

Роль эмиттера, общего для всех n–p–n-транзисторов, играет эпитаксиальный n-слой вместе с n^+-подложкой (последняя в сущности обеспечивает омический контакт с n-слоем). В схемах И^{2}Л, вообще говоря, нет необходимости изолировать n–p–n-транзисторы друг от друга, так как общность эмиттерно-



Разрез по линии A–A

Рис. 10.8. Структура и топология схемы И^{2}Л

го слоя не только не противопоказана, но и необходима по самой структуре схемы (рис. 10.6 и 10.7).

Инжектор осуществлен в виде длинной p-полоски, выполненной на этапе базовой диффузии. Базой p–n–p-транзистора является эпитаксиальный n-слой, а коллекторами — базовые p-слои n–p–n-транзисторов. Следовательно, p–n–p-транзистор имеет горизонтальную структуру (см. рис. 7.21) и однородную базу, т.е. является бездрейфовым.

Транзисторы n–p–n представляют собой «обращенные» многоэмиттерные транзисторы.

В заключение просуммируем преимущества И^{2}Л: отсутствие изолирующих карманов (т.е. экономия площади); отсутствие резисторов (и связанные с этим экономия площади, уменьшение напряжения питания, мощности, времени задержки); малая емкость коллектора (из-за малой площади n^+-слоев).

Эмиттерно-связанная логика (ЭСЛ). В основе схемы ЭСЛ (рис. 10.3) лежит переключатель тока, рассмотренный в разделе 8.5. Отличие от простого переключателя состоит в том, что в одно из плеч включено параллельно несколько транзисторов ($T1$ и $T2$ на рис. 10.9). Эти транзисторы равноправны в том смысле, что отпирание любого из них (или всех вместе) приводит к переключению тока I_0 из правого плеча схемы в левое. Поэтому ЭСЛ, как и РТЛ, выполняет функцию ИЛИ–НЕ.

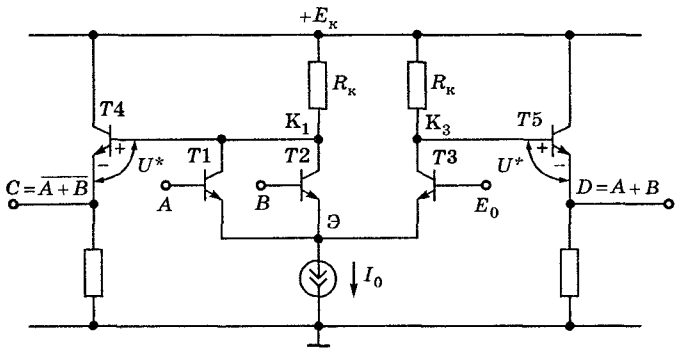


Рис. 10.9. Логический элемент ЭСЛ

Повторители $T4$ и $T5$ смещают уровни коллекторных потенциалов на величину U^*. Как известно, без такого смещения переключатели тока не могут работать совместно (см. рис. 8.14, где величина смещения обозначена через e). Подставляя $e = U^*$ в (8.40) и полагая $\delta \approx 0{,}1$ В, убеждаемся, что при наличии повторителей цепочка ЭСЛ работоспособна.

Пусть на оба логических входа поданы запирающие нулевые уровни U^0. Тогда в соответствии с функцией ИЛИ–НЕ напряжение на выходе C будет единичным уровнем U^1. Его легко найти, учитывая, что транзисторы $T1$ и $T2$ заперты, т.е. $U_{к1} = E_к$. Вычитая напряжение на эмиттерном переходе $T4$, получаем:

$$U^1 = E_к - U^*. \qquad (10.9а)$$

Пусть теперь на один из входов, например A, подан отпирающий единичный уровень U^1. Тогда напряжение на выходе C примет значение U^0. Положим, как и в разделе 8.5, что открытый транзистор находится на границе насыщения, т.е. $U_{к1} = U_{б1} = U^1$. Вычитая U^* и подставляя (10.9а), получаем:

$$U^0 = E_к - 2U^*. \qquad (10.9б)$$

Используя выражения (10.9), найдем значение логического перепада:

$$U_л = U^* \approx 0{,}7 \text{ В}. \qquad (10.10)$$

Как видим, это значение практически не отличается от значения, свойственного схеме $И^2Л$.

Напряжение смещения E_0 выбирается равным полусумме уровней U^0 и U^1, т. е.

$$E_0 = E_к - \tfrac{3}{2}U^*. \qquad (10.11)$$

При этом уровни U^0 и U^1 расположены симметрично относительно E_0 на «расстоянии» $\pm\frac{1}{2}U^*$.

Рабочий ток I_0 оценивается из равенства $U_{к1} = U^1$, обеспечивающего граничный режим открытого транзистора. Подставляя в это равенство $U_{к1} = E_к - \alpha I_0 R_к$ и значение U^1 из (10.9а), получаем:

$$\alpha I_0 R_к = U^*. \qquad (10.12)$$

Сопротивление $R_к$ выбирают из условия (8.44), обеспечивающего минимальное время переключения. Обычно значения $R_к$ лежат в пределах 0,5–2 кОм, а ток I_0 — в пределах 0,5–2 мА. При расчете мощности нужно, конечно, учитывать и токи в повторителях.

Используя выражения (10.9б) и (10.11), легко показать, что в исходном состоянии, когда на обоих входах задан уровень U^0, эмиттерные переходы транзисторов $T1$ и $T2$ находятся под прямым смещением $+\frac{1}{2}U^* = 0{,}35$ В. Такое смещение меньше напряжения отпирания, составляющего около 0,6 В, поэтому транзисторы практически заперты. Однако помехоустойчивость схемы, определяемая разностью между напряжением отпирания и прямым смещением, составляет всего 0,25 В.

Для повышения помехоустойчивости можно пойти на то, чтобы открытые транзисторы работали не на границе насыщения, а в квазинасыщенном режиме, т.е. при небольшом (0,2–0,3 В) прямом смещении коллекторного перехода. Можно также заменить повторители более сложными схемами сдвига уровня (см. рис. 9.20). Применение последних равносильно увеличению напряжения U^*. Однако такие схемы содержат дополнительные p–n-переходы или резисторы, а следовательно, занимают большую площадь.

До сих пор транзистор $T3$ считался вспомогательным — «хранящим» ток I_0, когда логические транзисторы заперты. Однако легко заметить, что он вместе с повторителем $T5$ тоже выполняет логическую функцию. А именно, в исходном состоянии, когда $U_A = U_B = U^0$, транзистор $T3$ открыт, а значит, на выходе D получается напряжение U^0. Если хотя бы на одном из логических входов действует напряжение U^1, то транзистор $T3$ обесточен и, следовательно, $U_D = U^1$. Как видим, одновременно

с функцией ИЛИ–НЕ, выполняемой на выходе C, выполняется функция ИЛИ на выходе D. Поэтому говорят, что схеме ЭСЛ свойственна *функция ИЛИ–НЕ/ИЛИ*.

В заключение отметим, что на практике в схемах ЭСЛ заземляют не отрицательный, а положительный полюс источника питания, так что все рабочие потенциалы на рис. 10.3 оказываются отрицательными. Разумеется, это не меняет принципа действия схемы и основных соотношений, однако заземление положительной шины питания существенно уменьшает влияние проходящих по ней помех на величины уровней U^0 и U^1.

10.3. Логические элементы на МДП-транзисторах

В настоящее время в логических схемах используются только МДП-транзисторы с окисным диэлектриком SiO_2. Поэтому ниже рассматриваются ИЛЭ на МОП-транзисторах. В основе МОП-транзисторной логики (МОПТЛ) лежат МОП-транзисторные ключи-инверторы, рассмотренные в разделе 8.6.

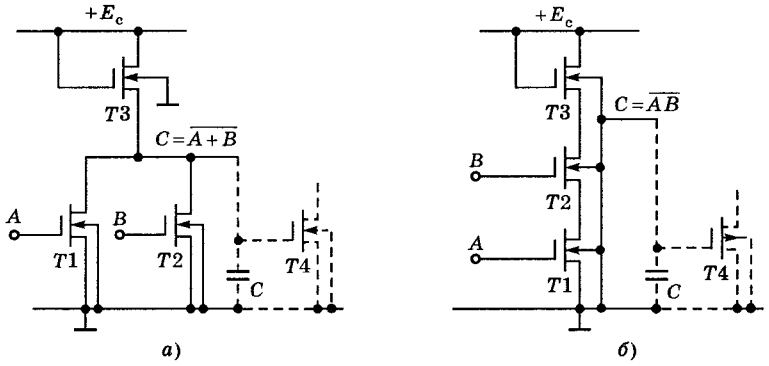


Рис. 10.10. Логические элементы МОПТЛ на ключах одного типа проводимости с параллельным (*а*) и последовательным (*б*) включением логических транзисторов

МОП-транзисторая логика на ключах одного типа проводимости. Как и в разделе 8.6, положим в основу транзисторы с индуцированным n-каналом (поскольку им свойственно положительное напряжение питания, более удобное при анализе).

Анализ МОП-транзисторных ИЛЭ проще, чем биполярных, потому что входные (затворные) цепи практически не потребляют тока. Следовательно, при работе в цепочке *отдельные*

ИЛЭ функционируют независимо друг от друга и каждый из них можно анализировать, отвлекаясь от влияния предыдущего и последующего ИЛЭ. В частности, *логические уровни* U^0 *и* U^1 *не зависят от нагрузки* и остаются такими же, как и при холостом ходе. Влияние последующего (нагрузочного) ИЛЭ сводится лишь к увеличению выходной емкости данного ИЛЭ.

На рис. 10.10 показаны два типовых варианта ИЛЭ, выполненных на МОП-транзисторах с n-каналом. В обоих вариантах используется динамическая нагрузка, так как использование нагрузочных резисторов приводит к резкому увеличению площади и лишает МОПТЛ одного из ее основных преимуществ — высокой степени интеграции.

ИЛЭ, показанный на рис. 10.10, *а*, построен по тому же принципу, что и РТЛ (рис. 10.2): логические транзисторы $T1$ и $T2$ включены параллельно, отпирание каждого из них приводит к снижению выходного уровня, т.е. схема выполняет функцию ИЛИ–НЕ.

Логические уровни в МОПТЛ соответствуют выходным напряжениям открытого и запертого ключа (рис. 8.17, *б*). В открытом состоянии остаточное напряжение на ключе описывается формулой (8.49). При соответствующей геометрии транзисторов оно имеет столь же малое значение (0,05–0,15 В), как и в биполярных ключах. Поэтому можно считать

$$U^0 = U_{\text{ост}} \approx 0{,}1 \text{ В}. \qquad (10.13\text{а})$$

Напомним, что малое остаточное напряжение подразумевает как можно меньшую ширину канала нагрузочного транзистора по сравнению с шириной канала активного транзистора. Заметим также, что в схеме МОПТЛ остаточное напряжение убывает пропорционально числу открытых логических транзисторов, так как параллельное соединение транзисторов равносильно увеличению удельной крутизны b_1 в формуле (8.49).

В закрытом состоянии ключа выходное напряжение близко к напряжению питания:

$$U^1 \approx E_{\text{с}}. \qquad (10.13\text{б})$$

Соответственно логический перепад составляет:

$$U_{\text{л}} = E_{\text{с}} - U_{\text{ост}} \approx E_{\text{с}}. \qquad (10.14)$$

Обычно напряжение питания в МОПТЛ выбирается в 3–4 раза больше порогового напряжения. Следовательно, если $U_0 = 1{,}5–3$ В, то логический перепад (5–10 В) намного превышает значения, свойственные схемам $И^2Л$, ЭСЛ и даже ТТЛ (при напряжении питания 4–5 В).

Еще одним достоинством МОПТЛ является повышенная помехоустойчивость. В самом деле, на входе запертого транзистора действует уровень U^0. Значит, для отпирания транзистора необходимо напряжение $U_0 - U^0$, близкое к пороговому, т. е. 1,5–3 В, тогда как у биполярных ИЛЭ, рассмотренных выше, оно составляет $(1–2)U^*$, т. е. 0,7–1,4 В.

ИЛЭ, показанный на рис. 10.10, *б*, характерен не параллельным, а последовательным включением логических транзисторов. Поэтому протекание тока в цепи, а значит, и низкий уровень U^0 выходного напряжения возможны только при отпирании всех (в данном случае обоих) логических транзисторов. Это имеет место при подаче уровня U^1 на все логические входы. Отсюда ясно, что данный ИЛЭ выполняет функцию И–НЕ.

Уровень U^1 в данной схеме такой же, как в предыдущей, но уровень U^0 больше — он пропорционален числу последовательно включенных логических транзисторов и может составлять 0,2–0,5 В и более. Соответственно меньше будет логический перепад. При напряжениях питания 10 В и более этот недостаток мало существен, однако в низковольтных схемах с малыми пороговыми напряжениями повышенный уровень U^0 представляет определенную проблему.

Быстродействие МОПТЛ, как и простейших МОП-ключей, ограничивается скоростью перезаряда выходной емкости C (на рис. 10.10 показана штриховой линией). Величина этой емкости пропорциональна количеству нагрузочных ИЛЭ, а также зависит от емкости металлизации.

Для повышения быстродействия нужно увеличивать рабочие токи транзисторов, т.е. их удельные крутизны. Однако это означает увеличение ширины канала, т.е. площади, занимаемой транзисторами. Кроме того, увеличение рабочих токов связано с увеличением потребляемой мощности, а это — еще одно препятствие на пути повышения степени интеграции. Из-за отмеченных затруднений реальные МОПТЛ оказываются менее быстродействующими, чем биполярные.

МОП-транзисторная логика на комплементарных транзисторах (КМОП-логика или КМОПТЛ). Простейшие КМОП-ключи были рассмотрены в разд. 8.6 (рис. 8.18). Их основное достоинство состоит в том, что изменение выходного напряжения не связано с изменением тока: он остается близким к нулю. Это преимущество — ничтожную потребляемую мощность — сохраняют и КМОП логические элементы. Два типовых варианта таких ИЛЭ показаны на рис. 10.11. По принципу действия они аналогичны вариантам, показанным на рис. 10.10.

На рис. 10.11 видна закономерность структуры КМОП логических элементов: *параллельное соединение одного типа транзисторов сопровождается последовательным соединением транзисторов другого типа*. Выполняемая логическая функция определяется включением транзисторов «нижнего этажа» (ср. с рис. 10.10). В рассматриваемой схеме — это n-канальные транзисторы. Если изменить полярность питающего напряжения, то в «нижнем этаже» окажутся транзисторы с p-каналом.

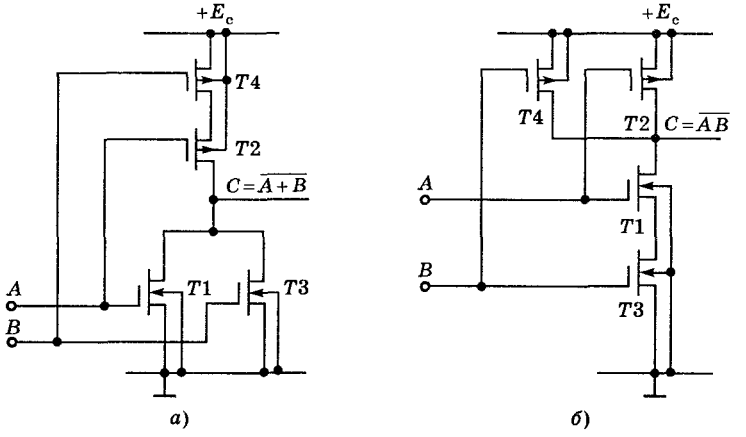


Рис. 10.11. КМОП логические элементы на комплементарных ключах: *а* — элемент ИЛИ-НЕ; *б* — элемент И-НЕ

Пусть в схеме на рис. 10.11, *а* на оба логических входа поданы уровни $U^0 < U_0$. Тогда в n-канальных транзисторах $T1$ и $T3$ канал отсутствует, т. е. они заперты. В p-канальных транзисторах $T2$ и $T4$ образуются каналы, так как разность потенциалов $U^0 - E_c = U_{зи}$ превышает (по модулю) пороговое напряжение. Однако, поскольку через каналы протекают ничтожные токи

запертых транзисторов $T1$ и $T3$, падение напряжения на каналах тоже получается ничтожным [см. (8.51б)]. Поэтому можно считать, что выходное напряжение равно напряжению питания. Это и есть уровень единицы:

$$U^1 = E_c. \qquad (10.15а)$$

Если подать на вход A уровень U^1, то в транзисторе $T1$ образуется канал, а в транзисторе $T2$ канал исчезает, т.е. этот транзистор запирается. Ничтожный остаточный ток транзистора $T2$, протекая через канал $T1$, дает на нем практически нулевое падение напряжения. Поэтому в данном случае можно считать

$$U^0 = 0. \qquad (10.15б)$$

Соответственно логический перепад составляет

$$U_л = E_c. \qquad (10.16)$$

Помимо высокой экономичности, дополнительными преимуществами КМОП-логики по сравнению с МОПТЛ являются малые рабочие напряжения (до $2U_0$ и менее) и более высокое быстродействие (см. конец разд. 8.6). Схема на рис. 10.11, *б* характеризуется аналогичными параметрами. Ее работа анализируется путем таких же рассуждений, какими мы пользовались выше.

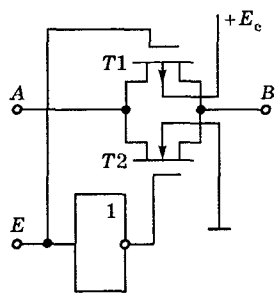


Рис. 10.12. Ключ коммутации на комплементарных МОП-транзисторах

Еще одной разновидностью элементов, выполненных на комплементарных транзисторах, являются ключи коммутации. Поскольку при размыкании и замыкании канала полевого транзистора управляющий сигнал гальванически не связан с коммутируемой цепью, такой ключ может замыкать и разрывать последовательные электрические цепи. На рис. 10.12 показана типовая схема КМОП-ключа коммутации.

Ключ имеет вход (A) и выход (B) сигнала, а также вход (E) разрешения прохождения сигнала (названия «вход» и «выход» условные, т.к. ключ симметричен относительно этих функциональных выводов). В отличие от КМОП-инвертора, на затворы p- и n-каналь-

ного транзисторов ключа подаются противоположные логические уровни, так что оба транзистора или одновременно закрыты (при высоком уровне на входе E), или одновременно открыты (при низком уровне на входе разрешения E). Канал проводимости в таком ключе двунаправленный и может пропускать логические уровни с амплитудой от 0 до $+ E_c$. За счет одновременного использования в ключе n- и p-канальных транзисторов удается снизить зависимость последовательного сопротивления ключа коммутации от уровня коммутируемого сигнала.

Динамическая МОП-транзисторная логика (ДМОПТЛ). Общая особенность МОП-транзисторных ИЛЭ — то, что в цепочке они гальванически изолированы друг от друга благодаря огромному входному сопротивлению со стороны затвора — позволяет строить особый класс ИЛЭ — *схемы динамического типа*. Проиллюстрируем эту возможность на примере схемы, показанной на рис. 10.10, *а*.

Отключим затвор транзистора $T3$ от шины питания и будем подавать на него управляющие импульсы с амплитудой $U_{ти} = E_c$ (рис. 10.13, *а*). Эти импульсы называют *тактовыми*, а режим работы ИЛЭ при наличии тактовых импульсов — *синхронным* режимом. Ранее рассмотренные ИЛЭ работали в *асинхронном* режиме. Очевидно, что в отсутствие тактового импульса, когда $U_{ти} = 0$, транзистор $T3$ заперт независимо от состояния логических транзисторов. При этом мощность от источника питания не потребляется.

Когда поступает тактовый импульс, затвор нагрузочного транзистора оказывается фактически присоединенным к шине питания, т.е. схема приобретает такую же конфигурацию, как на рис. 10.10, *а*. Значит, во время действия тактового импульса синхронный ИЛЭ работает так же, как асинхронный, и потребляет или не потребляет мощность в зависимости от состояния логических транзисторов.

Из сказанного следует, что синхронный режим обеспечивает выигрыш в потребляемой мощности. Этот выигрыш определяется *скважностью* тактовых импульсов, т.е. отношением периода их повторения T к длительности импульса $t_{ти}$. Чем больше скважность $T/t_{ти}$, тем больше выигрыш в мощности. Однако величина $t_{ти}$ ограничена временем перезаряда паразитной емкости C, а период T — желательной скоростью передачи информации (частотой опроса ИЛЭ).

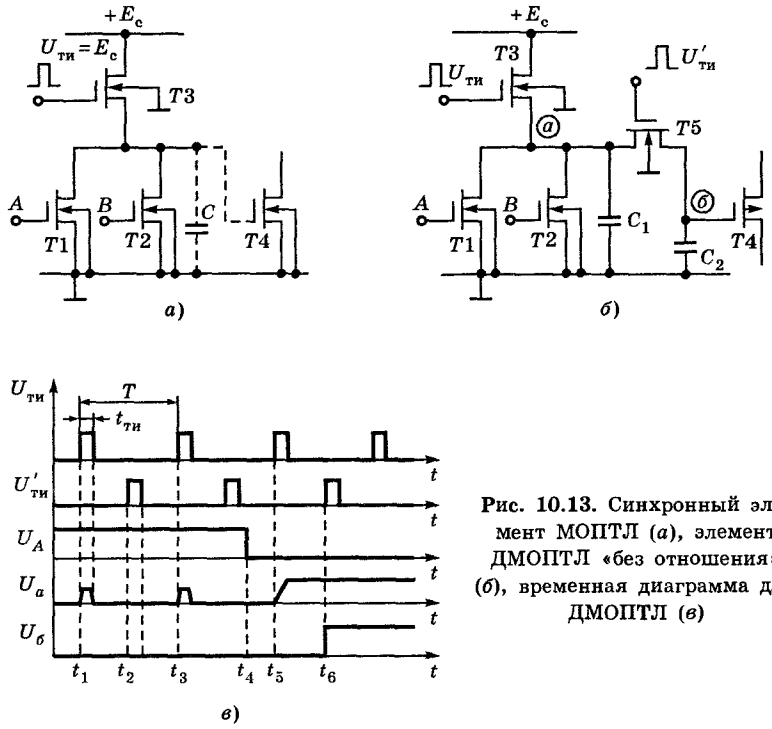


Рис. 10.13. Синхронный элемент МОПТЛ (*а*), элемент ДМОПТЛ «без отношения» (*б*), временная диаграмма для ДМОПТЛ (*в*)

Для построения элемента ДМОПТЛ простейшую схему (рис. 10.13, *а*) нужно дополнить ключом $T5$, который отделяет выход данного ИЛЭ от входа следующего (рис. 10.13, *б*). Ключ $T5$ вместе с конденсаторами C_1 и C_2 образует схему памяти. Емкости могут быть либо «паразитными» (в сумме равными емкости C на рис. 10.13, *а*), либо специально осуществленными на кристалле. Ключ $T5$ управляется вспомогательными тактовыми импульсами ($U'_{\text{ти}}$), сдвинутыми относительно основных ($U_{\text{ти}}$). Работа схемы иллюстрируется временны́ми диаграммами на рис. 10.13, *в*.

Положим для простоты, что на входе B действует неизменное запирающее напряжение U^0, а на входе A напряжение принимает значение U^1 или U^0. Пусть в исходном состоянии $U_A = U^1$, т.е. транзистор $T1$ открыт. Тогда в интервале между тактовыми импульсами $U_{\text{ти}}$, когда транзистор $T3$ заперт, через $T1$ протекает ничтожный ток и на выходе (в точке $а$) остаточное напряжение практически равно нулю [см. (8.51б)].

При поступлении очередного тактового импульса $U_{ти}$ (момент t_1) транзистор $T3$ открывается и в точке *а* устанавливается остаточное напряжение, зависящее от отношения удельных крутизн активного и нагрузочного транзисторов [см (8.49)]. Если размеры транзисторов сравнимы, это напряжение может быть достаточно большим. По окончании тактового импульса $U_{ти}$ напряжение U_0 снова падает до нуля.

При поступлении тактового импульса $U'_{ти}$ (момент t_2) отпирается ключ $T5$. При этом емкости C_1 и C_2 оказываются соединенными параллельно и на них устанавливается одинаковое напряжение, в данном случае — равное нулю. По окончании импульса $U'_{ти}$ ключ $T5$ запирается и емкость C_2 сохраняет нулевое напряжение несмотря на то, что в момент t_3 при очередном импульсе $U_{ти}$ напряжение U_a снова временно возрастает.

Таким образом, на входе следующего ИЛЭ (в точке *б*) уровень U^0 равен нулю независимо от отношения удельных крутизн b_1/b_3 в предыдущем ИЛЭ. Это обстоятельство позволяет выбирать отношение b_1/b_3 близким к единице, т.е. уменьшать размеры активного транзистора до размеров нагрузочного. Тем самым обеспечивается существенная экономия площади на кристалле, что является одним из важных преимуществ ДМОПТЛ.

Схемы ДМОПТЛ, построенные по описанному принципу, называют схемами «без отношения», а простейшую схему (рис. 10.13, *а*), в которой должно выполняться неравенство $b_1/b_3 >> 1$, схемой «с отношением».

Если на вход A подать напряжение $U_A = U^0$ (момент t_4), то транзистор $T1$ запирается, и при поступлении очередного тактового импульса $U_{ти}$ (момент t_5) напряжение U_0 поднимается до уровня $U^1 = E_c$. Этот уровень в дальнейшем сохраняется, так как транзистор $T1$ остается запертым. При очередном отпирании ключа $T5$ (момент t_6) конденсатор C_2 зарядится до такого же напряжения[1] и сохранит его после запирания ключа $T5$.

Из приведенного описания следует, что в схеме «без отношения» информация передается от одного ИЛЭ к другому с запаздыванием — сдвигом на один такт (на один период тактовых импульсов).

---

1 Вообще говоря, при подключении емкости C_2 к C_1 напряжение уменьшается, так как заряд распределяется между емкостями. Это уменьшение не существенно, если $C_1 >> C_2$.

10.4. Логические элементы на совмещенных биполярных и МОП-транзисторах (БиКМОП)

Продолжительное время, до начала 80-х годов, развитие биполярных и МОП цифровых интегральных схем шло параллельно и представляло два конкурирующих направления. Основным достоинством биполярных схем являлось высокое быстродействие (ТТЛШ и ЭСЛ), а МОП-схем — высокая степень интеграции и низкая потребляемая мощность (КМОП). Поэтому усилия разработчиков биполярных логических ИС были направлены на снижение потребляемой мощности при заданном быстродействии, а разработчиков МОП-ИС — на повышение быстродействия с сохранением низкого значения потребляемой мощности.

Решение задач, стоящих перед обоими направлениями, биполярных и МОП-интегральных схем, было найдено путем синтеза и создания БиКМОП-логического элемента. Этот шаг был обеспечен разработкой технологии, позволяющей совместить на одном кристалле биполярные и МОП-транзисторные структуры с необходимыми электрическими характеристиками.

На рис. 10.14 показаны варианты БиКМОП-логических элементов. Поясним работу БиКМОП-логики на примере простого инвертора, элемента НЕ (рис. 10.14, *а*). В статическом состоянии этот элемент работает как обычный КМОП-инвер-

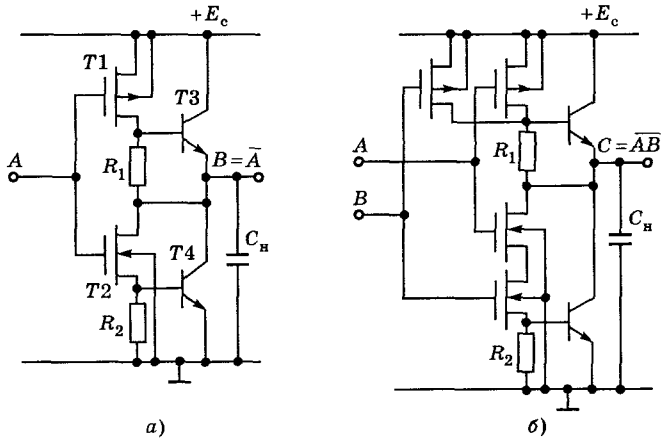


Рис. 10.14. БиКМОП-логические элементы:
а — инвертор, *б* — схема И–НЕ

тор ($T1$, $T2$), у которого последовательно с каналом каждого из транзисторов включены резисторы R_1, R_2, сопротивления которых сравнимы с сопротивлениями каналов соответствующих транзисторов в открытом состоянии. Статические токи в КМОП-инверторе отсутствуют, напряжения на резисторах R_1 и R_2 равны нулю и биполярные транзисторы $T3$ и $T4$, эмиттерные переходы которых подсоединены к этим резисторам, закрыты, т.е. биполярные транзисторы в статическом состоянии не влияют на характеристики БиКМОП-вентиля.

Пусть в исходном состоянии на входе инвертора (рис. 10.14, *а*) низкий логический уровень; это соответствует открытому состоянию *p*-канального транзистора $T1$ и высокому уровню на выходе инвертора, т.е. конденсатор нагрузки $C_н$ заряжен до уровня $E_с$. После поступления на вход высокого логического уровня транзистор $T1$ запирается, а $T2$ отпирается и начинается разряд емкости $C_н$ через транзистор $T2$ на параллельно соединенные резистор R_2 и эмиттерный переход биполярного транзистора $T4$. Протекание части разрядного тока через эмиттерный переход приводит к открыванию $T4$ и значительному ускорению перезаряда емкости нагрузки. Процесс переключения при переходе входного сигнала из высокого логического уровня в низкий происходит аналогично (открывание $T3$ вызывается зарядным током, протекающим по цепи «источник питания — $T1$ — эмиттерный переход $T3$»).

Из приведенного описания следует, что использование биполярных транзисторов позволяет нейтрализовать влияние емкости нагрузки, являющейся одним из основных факторов, ограничивающих быстродействие КМОП-ИС.

10.5. Логические элементы на полевых транзисторах с управляющим переходом металл–полупроводник (МЕП)

Схема логического элемента ИЛИ–НЕ на МЕП-транзисторах (рис. 10.15) аналогична схеме элемента на *n*-канальных МОП-транзисторах (см. рис. 10.10, *а*). Полевые транзисторы типа металл–полупроводник аналогичны транзисторам с управляющим *p–n*-переходом, рассмотренным в разделе 4.3, но обедненный слой формируется на границе металл–полупроводник (транзисторы такой структуры называют *полевыми тран-*

зисторами с барьером Шоттки). В качестве нагрузочного транзистора используется нормально открытый МЕП-транзистор, а в качестве управляющих — нормально закрытые с пороговым напряжением около +0,2 В.

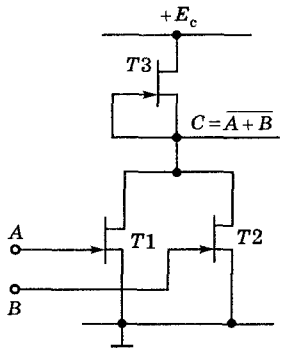


Рис. 10.15. Логический элемент ИЛИ–НЕ на МЕП-транзисторах

Если оба входных напряжения меньше порогового, транзисторы $T1$ и $T2$ закрыты, и уровень выходного сигнала близок к уровню напряжения питания. При появлении на одном из входов напряжения выше порогового, соответствующий ключевой транзистор открывается, и выход элемента переходит к низкому логическому уровню. Особенностью данного типа логических элементов является то, что транзисторы $T1$ и $T2$ работают при прямом смещении на управляющем переходе металл–полупроводник, которое для барьера Шоттки обычно не превышает 0,5 В. Данное ограничение определяет низкие значения напряжения питания (обычно около 0,8 В) и логического перепада (около 0,5 В).

Быстродействие МЕП-логического элемента, как и простейших МОП-ключей, определяется временем перезарядки нагрузочного конденсатора. МЕП-логический элемент имеет существенно более высокое быстродействие при одинаковой потребляемой мощности и линейных размерах. Это объясняется меньшими значениями логического перепада и емкости затвор–исток. Интегральные схемы на МЕП-логических элементах обычно изготавливаются на подложках из арсенида галлия (GaAs). Подвижность электронов в канале GaAs МЕП-транзистора выше, чем у кремниевого МДП-транзистора, что обеспечивает возможность использования транзисторов с меньшей шириной канала и, соответственно, меньшей емкостью затвор–исток.

10.6. Параметры логических элементов

Все логические элементы характеризуются определенным набором — *номенклатурой* параметров, приводимых в справочниках и другой документации. Большинство параметров имеет четкое официальное определение, что обеспечивает однозначность измерительных методик и возможность сравнения

разных типов ИЛЭ. Полная номенклатура параметров слишком велика. Для первого знакомства рассмотрим ограниченное число параметров — тех, которые наиболее существенны для сравнительных оценок.

Средняя статическая потребляемая мощность. Она определяется следующим образом:

$$P_{\text{стат ср}} = \tfrac{1}{2}(P^0_{\text{пот}} + P^1_{\text{пот}}), \qquad (10.17)$$

где $P^0_{\text{пот}}$ и $P^1_{\text{пот}}$ — мощности, потребляемые ИЛЭ соответственно в состоянии логического нуля и логической единицы.

Определение средней мощности исходит из того, что в сложной многоэлементной аппаратуре в среднем половина ИЛЭ находится в состоянии «1», а половина — в состоянии «0».

Среднее время задержки распространения сигнала. Определение этого параметра следующее:

$$t_{\text{зд р ср}} = \tfrac{1}{2}(t^{1,0}_{\text{зд р}} + t^{0,1}_{\text{зд р}}), \qquad (10.18)$$

где $t^{1,0}_{\text{зд р}}$ и $t^{0,1}_{\text{зд р}}$ — времена задержки между фронтами выходного и входного импульсов при включении (когда $U_{\text{вых}}$ меняется от U^0 до U^1) и при выключении (когда $U_{\text{вых}}$ меняется от U^1 до U^0). Времена задержки измеряются либо на уровне 0,5 от полной амплитуды импульса, либо на уровне порога чувствительности.

Поскольку переходный процесс зависит от характера нагрузки, время задержки оценивают при определенных условиях на выходе, считая что данный ИЛЭ нагружен входной цепью такого же ИЛЭ.

Средняя работа переключения ($A_{\text{ср}}$). Этот параметр — комбинированный: он представляет собой произведение средней мощности, потребляемой ИЛЭ во время переключения, на среднюю продолжительность одного переключения, т.е. одновременно характеризует и экономичность ИЛЭ, и его быстродействие. В настоящее время именно по этому параметру производят первоочередное сравнение разных типов ИЛЭ и особенно оценку перспективности новых типов.

Средняя динамическая потребляемая мощность ($P_{\text{дин ср}}$) равна произведению средней работы переключения на частоту переключений.

Средняя продолжительность переключения, как правило, пропорциональна емкости нагрузки $C_{\text{н}}$, а средняя мощность — квадрату на-

пряжения питания E. Тогда для средней работы переключения можно записать:

$$A_{ср} \sim E^2 C_н. \qquad (10.19а)$$

Учитывая, что емкость в первую очередь пропорциональна площади транзистора, и считая для наглядности конфигурацию транзистора квадратной, перепишем (10.19а) в виде

$$A_{ср} \sim E^2 a^2, \qquad (10.19б)$$

где a — линейные размеры транзистора.

Из выражений (10.19) видно, что *работа переключения характеризует физико-технологический и схемотехнический уровень интегральных схем*, так как площадь прибора и рабочее напряжение в конечном итоге зависят от типа прибора, его структуры, разрешающей способности фотолитографии и других аналогичных факторов.

При заданном физико-технологическом и схемотехническом уровне, считая заданной величину $A_{ср}$, можно осуществить различные варианты ИЛЭ, но они будут характеризоваться либо *относительно высоким быстродействием при низкой экономичности*, либо, наоборот, *высокой экономичностью при относительно низком быстродействии*. Такая обратная зависимость хорошо известна из практики разработки интегральных схем и прослеживается в табл. 10.1.

Таблица 10.1. Основные параметры интегральных логических элементов

Тип логики	$P_{пот\ ср}$, мВт	$t_{зд\ ср}$, нс	$A_{ср}$, пДж	$U_{п\ ст}$, В	$K_{об}$	$K_{раз}$
ТТЛ	1–20	5–20	50–100	0,8–1	2–5	10
ТТЛШ		2–10	10–50	0,5–0,8		
ЭСЛ	20–50	0,5–2	20–50	0,2–0,3	2–5	10–20
$И^2Л$	0,01–0,1	10–100	0,2–2	0,02–0,05	1	3–5
МОП	1–10	20–200	50–200	2–3	2–5	10–20
КМОП	0,01–0,1	10–50	0,5–5,0	1–2		
БиКМОП	0,01–0,1	2–10	2–20	1–2	2–5	10–100
МЕП	0,1–0,5	0,15–0,5	0,1–0,5	0,1–0,2	2–5	2–5

Кардинальный прогресс в физике, технологии и схемотехнике ИС должен сопровождаться уменьшением работы переключения. Этот тезис хорошо подтверждается и опытом повышения плотности компоновки элементов (что привело к снижению работы переключения с 50–100 до 0,1–2 пДж), и разработкой БиКМОП-логики (табл. 10.1), и исследованиями в области принципиально новых электронных приборов, отличных от транзисторов. Например, использование так называемого *эффекта Джозефсона* позволяет в принципе довести работу переключения до 10^{-4} пДж и менее.

Статическая помехоустойчивость. Этот параметр $U_{\text{п ст}}$ представляет собой максимально допустимое напряжение статической помехи, при которой еще не происходит изменения выходных уровней ИЛЭ. Под статической помехой понимают паразитные напряжения и токи, длительность которых значительно больше времени переключения ИЛЭ из одного устойчивого состояния в другое. Механизм действия таких помех и способ их аналитической оценки были рассмотрены в разделе 8.7.

Иногда помехоустойчивость оценивают по отношению к логическому перепаду в виде параметра

$$K_{\text{п ст}} = U_{\text{п ст}}/U_{\text{л}}, \qquad (10.20)$$

который называют коэффициентом помехоустойчивости.

Коэффициент объединения по входу $K_{\text{об}}$ — число логических входов ИЛЭ. Увеличение коэффициента $K_{\text{об}}$ (например, количества транзисторов в РТЛ или количества эмиттеров в ТТЛ) ограничено не только конструктивно-технологическими факторами, но и взаимодействием между входами, а также увеличением времени задержки (поскольку каждый логический вход привносит дополнительную емкость).

Коэффициент разветвления по выходу $K_{\text{раз}}$ — допустимое число нагрузок, подключаемых к выходу данного ИЛЭ. Под каждой нагрузкой понимается вход такого же ИЛЭ. Увеличение коэффициента $K_{\text{раз}}$ ограничено рабочим током ИЛЭ, разбросом параметров, а также увеличением суммарной выходной емкости и связанного с нею времени задержки.

При необходимости увеличить коэффициенты $K_{\text{об}}$ и $K_{\text{раз}}$ к ИЛЭ (соответственно на входах или выходе) подключают специальные буферные каскады, которые называют *расширителями по входу и по выходу.*

Типичные параметры основных современных ИЛЭ приведены в табл. 10.1. Эти данные иллюстрируют те особенности отдельных типов ИЛЭ, которые отмечались выше в соответствующих разделах. Наиболее быстродействующими являются сейчас схемы МЕП, ЭСЛ и ТТЛ с барьерами Шоттки (ТТЛШ), наиболее экономичными — схемы ДМОП, а особенно КМОП и БиКМОП.

10.7. Интегральные триггеры

В разделе 8.8 было показано, что основу триггеров составляют ключи, охваченные положительной обратной связью. Из рис. 8.28, *а* видно, что RS-триггер состоит из двух пар транзисторов. Один из транзисторов в паре — управляющий, а другой входит в петлю обратной связи. Теперь, когда мы знакомы с логическими элементами, легко заметить, что каждая пара в RS-триггере представляет собой двухвходовую схему ИЛИ–НЕ типа РТЛ.

Вывод, который мы только что сделали, носит общий характер: *любой триггер является совокупностью нескольких определенным образом соединенных логических элементов.* При этом количество используемых ИЛЭ, а также способы их соединения могут быть различными. Соответственно различными будут и функции, выполняемые триггерами. Поэтому наряду с RS-триггерами существует весьма много других разновидностей. Что касается базиса, т.е. типа используемых ИЛЭ, то от него зависят такие основные параметры, как быстродействие, потребляемая мощность, нагрузочная способность и др.

Поскольку принципиальные схемы, параметры и особенности ИЛЭ подробно рассмотрены в предыдущих разделах, ниже, при рассмотрении триггеров, мы ограничимся, как правило, структурными схемами, чтобы не рассеивать внимание на детали принципиальных схем, тем более, что последние при большом числе используемых ИЛЭ имеют достаточно сложный внешний вид.

RS-триггер. Логическая структура RS-триггера, выполненного на схемах ИЛИ–НЕ, показана на рис. 10.16, *а*. На рис. 10.16, *б* приведены его логическая формула и символическое обозначение. Верхними индексами n и $n+1$ обозначены значения величины Q до и после поступления управляющих сигналов (т.е. в n- и $n+1$-м тактах).

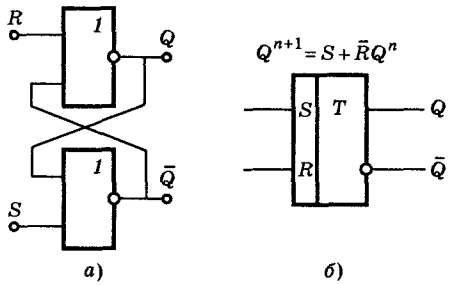


Рис. 10.16. RS-триггер, составленный из двух элементов ИЛИ–НЕ (*а*) и условное обозначение RS-триггера (*б*)

Как известно, уровни напряжений на обоих выходах триггера различны и одновременно изменяются на противоположные. Поэтому на символе триггера один из выходов обозначен через Q, а второй через $\overline{Q}$ (факт инверсии, как обычно, отражен кружком на стороне прямоугольника). Выход Q считается *главным*: значениями Q характеризуют состояние триггера в целом. Так, если говорят: «триггер находится в состоянии '1'», то это значит, что $Q = 1$ (соответственно $\overline{Q} = 0$).

Для того чтобы не повторяться в дальнейшем, проверим логическую формулу (рис. 10.16, *б*):

если $S = 0$, $R = 0$, то $Q^{n+1} = 0 + 1 \cdot Q^n = Q^n$;

если $S = 0$, $R = 1$, то $Q^{n+1} = 0 + 0 \cdot Q^n = 0$;

если $S = 1$, $R = 0$, то $Q^{n+1} = 1 + 1 \cdot Q^n = 1 + Q^n = 1$.

Еще один возможный набор ($S = 1$ и $R = 1$) рассмотрим чуть позднее.

Исходя из перечисленных наборов, сделаем следующий вывод: сигналы на S- и R-входах (под сигналами понимаются значения «1») *обеспечивают однозначное состояние триггера*. Сигнал $S = 1$ означает $Q = 1$; сигнал $R = 1$ означает $Q = 0$. По окончании сигнала принятое состояние сохраняется; оно сохраняется и при повторении такого же сигнала.

Теперь обратимся к набору $S = 1$ и $R = 1$. Каким бы ни было предыдущее состояние триггера, при таком наборе входных сигналов уровни на обоих выходах согласно рис. 10.16, *а* получаются одинаковыми: $Q = \overline{Q} = 0$. Уже одно это обстоятельство говорит о ненормальности ситуации. Однако главное противоречие состоит в том, что по окончании сигналов S и R триггер оказывается в неопределенном состоянии: на обеих парах входов в первый момент действуют нулевые сигналы. Под действием внутренних флюктуаций триггер с равной вероятностью может перейти в любое из двух устойчивых состояний $Q = 1$ или $Q = 0$. Об этом подробно говорилось в разделе 8.8. Поэтому на-

бор $S = 1$, $R = 1$ в данном триггере является запрещенным (с точки зрения правильности его функционирования), т.е. он не должен встречаться при использовании триггера в узлах аппаратуры. Такой запрет понятен и из общих соображений: нельзя одновременно подавать на триггер противоположные команды «установить единицу (S)» и «установить нуль (R)».

Разумеется, RS-триггеры приходится строить не только на схемах ИЛИ–НЕ, но и на схемах И–НЕ (например, при использовании базиса ТТЛ). Переход от схем ИЛИ–НЕ к схемам И–НЕ осуществляется путем замены всех входных и выходных величин на их отрицания. Тогда RS-триггер, выполненный на схемах И–НЕ, будет иметь структуру, показанную на рис. 10.17, *а*. Логическая формула для этой структуры остается такой же, как и на рис. 10.16, *б*. Однако по сравнению с рис. 10.16, *а* изменилось положение главного выхода Q, а на входы необходимо подавать инвертированные значения величин S и R. Если инвертор включить в состав схемы ИЛЭ, то изменится ее символ — появится кружок на входе (рис. 10.17, *б*).

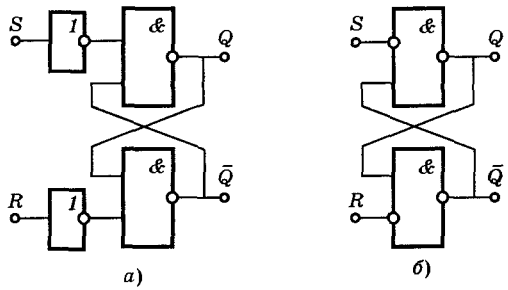


Рис. 10.17. RS-триггер, составленный из двух элементов И–НЕ: *а* — схема, полученная из схемы рис. 10.16, *а* путем инверсии входных и выходных величин; *б* — та же схема с внутренними входными инверторами

Принципиальные схемы RS-триггеров, выполненных в базисах КМОП и ТТЛ, представлены на рис. 10.18. В первой схеме (рис. 10.18, *а*) используются ИЛЭ типа ИЛИ–НЕ, и она работает в положительной логике, во второй (рис. 10.18, *б*) используются ИЛЭ типа И-НЕ, и она работает в отрицательной логике.

RST-триггер. Рассмотренный выше RS-триггер относится к классу *асинхронных*, у которых изменение состояния происходит тогда, когда меняется уровень на соответствующем входе. Большее распространение имеют *синхронные триггеры* или RST-триггеры, у которых состояние может меняться только

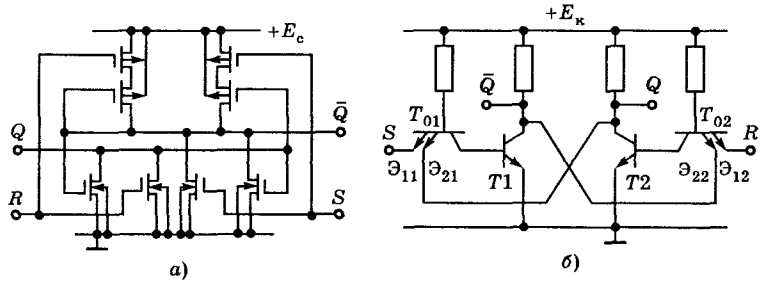


Рис. 10.18. RS-триггеры, выполненные в базисе КМОП (*а*) и ТТЛ (*б*)

при поступлении специальных *тактовых импульсов* (см. для сравнения разд. 10.3). В промежутках между тактовыми импульсами изменения уровней на входах S и R не вызывают изменения состояния триггера, а лишь «программируют» то состояние, которое он примет при поступлении очередного тактового импульса.

В обозначении RST-триггера буква T связана с тактовыми импульсами, но сам тактовый вход на схемах обозначают обычно буквой C (от английского clock — времязадающий). Структура, символ и временная диаграмма работы RST-триггера показаны на рис. 10.19.

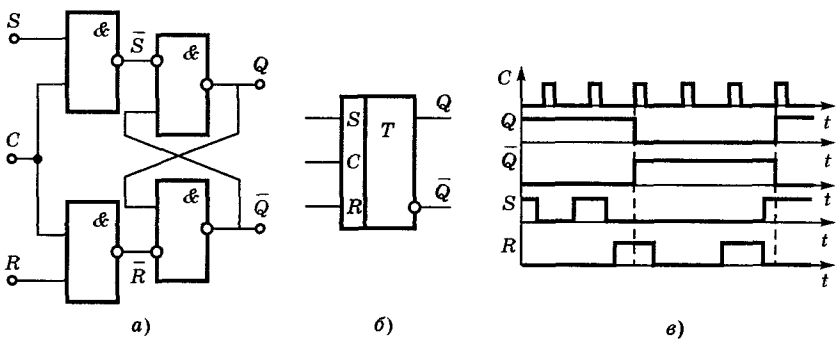


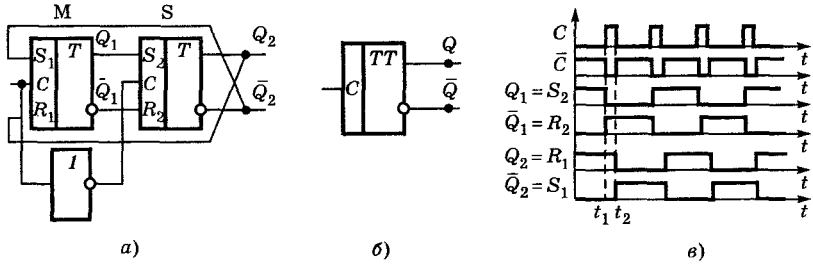
Рис. 10.19. RST-триггер *а* — структурная схема, *б* — обозначение, *в* — временные диаграммы

Как видим, в основе структуры лежит известный уже RST-триггер, выполненный на основе схем И–НЕ (см. рис. 10,17, *а*). Входы этого триггера управляются с помощью еще двух схем И–НЕ, которые инвертируют уровни S и R при наличии тактового импульса C. В отсутствие тактового импуль-

са, при C = 0, схемы И–НЕ бездействуют и состояние RST-триггера не меняется. Влияние тактового импульса можно охарактеризовать логической формулой

$$Q^{n+1} = C(S + \overline{R}Q^n). \qquad (10.21)$$

Т-триггер. Структура интегрального триггера с общим входом (Т-триггера), а также его символ и временная диаграмма показаны на рис. 10.20. Из рисунка видно, что Т-триггер состоит из двух RST-триггеров (М и S) и инвертора, через который подается тактовый импульс на S-триггер. Триггер М называют *ведущим*, а триггер S — *ведомым*[1].



а) *б)* *в)*

Рис. 10.20. Т-триггер: *а* — структурная схема; *б* — обозначение; *в* — временные диаграммы

В интервале между тактовыми импульсами (когда $C = 0$) уровни на выходах обоих триггеров одинаковы: $Q = Q_1$. Пусть, например, $Q = Q_1 = 1$ (см. исходное состояние на рис. 10.20, *в*). Для того чтобы очередной тактовый импульс изменил состояние ведущего триггера (т.е. обеспечил значение $Q_1 = 0$), нужно предварительно задать на его входах соответствующие уровни: $S_1 = 0$ и $R_1 = 1$. Эту функцию выполняют перекрестные обратные связи с выхода ведомого триггера на вход ведущего. Действительно, из рис. 10.20, *а* следует: $S_1 = \overline{Q} = 0$ и $R_1 = Q = 1$. Таким образом, *в интервале между тактовыми импульсами ведущий триггер подготовлен к изменению своего состояния*, и это состояние меняется при поступлении очередного тактового импульса (момент t_1 на рис. 10.20, *в*).

Что касается ведомого триггера, то его состояние во время тактового импульса не может измениться, поскольку в это

1 Соответствующие английские термины: Master (хозяин) и Slave (раб).

время на его тактовом входе действует запрещающий сигнал $\overline{C} = 0$. Следовательно, во время тактового импульса выходные величины Q и $\overline{Q}$ не меняются, а значит, не меняются и входные величины S_1 и R_1. Последнее обстоятельство обеспечивает надежное переключение ведущего триггера.

Из рис. 10.20, *а* видно, что переключение ведущего триггера сопровождается изменением уровней S_2 и R_2 на входах ведомого триггера. Поэтому *к концу тактового импульса ведомый триггер оказывается подготовленным к переключению в новое состояние*, соответствующее новому состоянию ведущего. Такое переключение происходит по окончании тактового импульса, когда на тактовом входе ведомого триггера восстанавливается разрешающий сигнал $C = 1$ (момент t_2 на рис. 10.20, *в*).

Таким образом в Т-триггере каждый тактовый импульс вызывает переход в новое устойчивое состояние, но с запаздыванием на величину, равную длительности тактового импульса. Логическое уравнение Т-триггера можно записать в виде:

$$Q^{n+1} = CQ^n + \overline{C}\overline{Q}^n, \qquad (10.22)$$

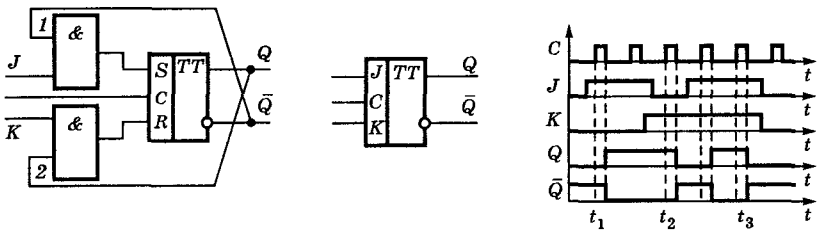
где CQ^n — значение во время тактового импульса, а $\overline{C}\overline{Q}^n$ — по его окончании.

JK-триггер. JK-триггеры (читается «джи-ка» триггеры)[1] являются наиболее универсальными: в них входы J и K (подобно входам R и S) задают желательное состояние, но, в отличие от RS-триггеров, допустим набор $J = 1$, $K = 1$. По принципу действия JK-триггеры относятся к категории синхронных, т. е. их выходные уровни устанавливаются только при поступлении тактовых импульсов C. Структура, символ и временная диаграмма JK-триггера показаны на рис. 10.21.

В основе JK-триггера, как видим, лежит структура Т-триггера, построенного по принципу «мастер—слэйв», но входы S и R связаны с выходами перекрестными связями не непосредственно, как в Т-триггере, а через две схемы И, на один из входов которых подаются управляющие уровни J и K. При наборе $J = 1$, $K = 1$ схемы И превращаются в повторители для уровней, поданных на входы *1* и *2*. Поэтому при таком наборе

---

1 Буква J — начальная в англ. слове jump (прыжок, переброс), а буква K — начальная в англ. слове keep (держать, сохранять).



а) *б*) *в*)

Рис. 10.21. JK-триггер: *а* — структурная схема; *б* — обозначение; *в* — временные диаграммы

JK-триггер работает как T-триггер (интервал t_2–t_3 на рис. 10.21, *в*). При остальных наборах величин J и K схема работает как RST-триггер (интервал t_1–t_2), где уровень J устанавливает значение $Q = 1$, а уровень K — значение $Q = 0$.

Логическое уравнение JK-триггера имеет вид:

$$Q^{n+1} = J\overline{Q}^n + \overline{K}Q^n. \qquad (10.23)$$

В частности, при $J = 1$, $K = 1$ получаем $Q^{n+1} = \overline{Q}^n$ [режим T-триггера, см. (10.22) при $C = 1$].

10.8. Запоминающие устройства

В цифровой аппаратуре, и прежде всего в ЭВМ, системы памяти или запоминающие устройства (ЗУ) занимают важнейшее место. Системы памяти делятся на *внешние* и *внутренние*. Внешние ЗУ до сих пор реализуются на магнитных лентах и магнитных или оптических дисках. Внутренние ЗУ (т.е. ЗУ, конструктивно объединенные с электронными блоками) на ранних этапах развития ЭВМ выполнялись на ферритовых сердечниках с прямоугольной петлей гистерезиса, а с конца 70-х годов — в основном на ИС.

Внешние магнитные ЗУ характеризуются неопределенно длительным сохранением информации при отсутствии питания, а также практически неограниченной *емкостью* (объемом) *памяти* в битах[1].

---

1 Битом информации называют значения «0» или «1» в двоичной системе счисления

Внутренние ЗУ предназначены в основном для хранения промежуточных данных в процессе выполнения арифметических или логических операций (*оперативные* ЗУ или ОЗУ), а также для хранения небольших стандартных программ, необходимых при использовании данного цифрового устройства для решения типовых задач (*постоянные* ЗУ или ПЗУ).

Оперативные запоминающие устройства (ОЗУ) характерны возможностью быстрого попеременного ввода и вывода (записи и считывания) информации, причем и для записи, и для считывания доступна любая отдельная ячейка ОЗУ. Поэтому синонимом ОЗУ является *память с произвольной выборкой* (английское RAM — Random Access Memory).

Постоянные запоминающие устройства (ПЗУ) в отличие от ОЗУ используются в основном для считывания записанной в них информации. Запись же осуществляется либо «раз и навсегда», либо во всяком случае весьма редко. Поэтому этот класс ЗУ называют в зарубежной литературе ROM (Read-Only-Memory — память только для считывания).

Оперативные ЗУ. Всякое ОЗУ состоит из двух основных частей: *накопителя и схем управления* или, как говорят, *периферии*. Накопитель — это основная часть ОЗУ, где хранятся данные (двоичные коды). Периферия же предназначена для ввода и вывода этих данных. В нее входят дешифраторы, усилители, регистры, разного рода ключи, коммутаторы и другие схемы общего назначения. Мы их не будем рассматривать. Сосредоточим внимание на наиболее специфичной части ОЗУ — накопителе.

Накопитель состоит из *запоминающих ячеек* (ЗЯ); каждая из них хранит один бит информации («0» или «1»). Естественно, что основу ЗЯ составляют бистабильные ячейки, основным свойством которых является наличие двух устойчивых состояний $Q = 1$ или $Q = 0$.

На рис. 10.22 показана типичная *матричная* организация ОЗУ, где отдельная ЗЯ расположена в узлах «решетки», образованной *адресными шинами* X и Y. Количество ячеек равно произведению количества горизонтальных шин на количество вертикальных (например, 4×4 = 16 ячеек). Каждая ЗЯ связана с одной горизонтальной и одной вертикальной адресными шинами. Поэтому, если подаются напряжения на определенные горизонтальную и вертикальную шины (например, X_1 и Y_2), то

к периферии подключается вполне определенная ЗЯ (в нашем примере $ЗЯ_{12}$, которая на рис. 10.22 заштрихована). В эту ячейку (с уникальным адресом X_1Y_2) можно либо записать необходимый бит информации («0» или «1»), либо считать информацию из ячейки (в ячейке $ЗЯ_{12}$ записана «1»).

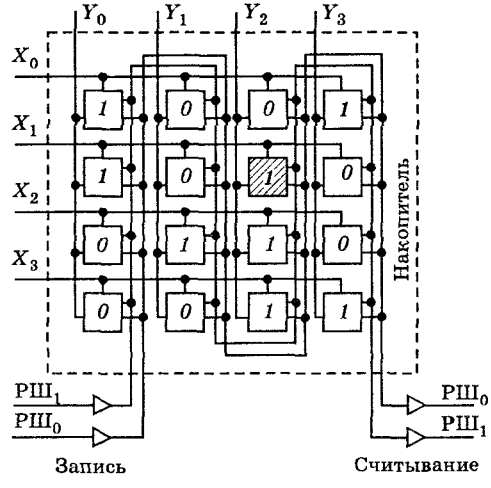


Рис. 10.22. Оперативное ЗУ с матричной структурой

И запись, и считывание осуществляется с помощью *разрядных шин* $РШ_1$ и $РШ_0$, которые подключены ко всем ЗЯ[1]. Индексы разрядных шин в известной мере условны: и на шину единиц ($РШ_1$) и на шину нулей ($РШ_0$) могут подаваться уровни как U^0 так и U^1. Поэтому индексация отнюдь не предопределяет логического уровня на той или иной шине, а является лишь условным обозначением того, что одна из них ($РШ_1$) подключена к главным выходам триггеров Q, а другая к их «напарникам» $\overline{Q}$. При записи выходы считывания отключаются, и на разрядные шины подают один из двух наборов (01 или 10), в зависимости от того, какое число желательно хранить в ЗЯ (на рис. 10.22 в клетках указаны уровни на главных выходах ячеек). При считывании входы записи отключаются и уровни, ха-

1 На входах и выходах разрядных шин треугольниками показаны коммутирующие устройства, которые подключают (или отключают) разрядные шины к управляющим устройствам (см. ниже).

рактеризующие состояние ЗЯ, поступают через усилители в соответствующие внешние узлы.

Запоминающие ячейки, используемые в накопителях ОЗУ, чрезвычайно разнообразны. Примеры таких ячеек показаны на рис. 10.23.

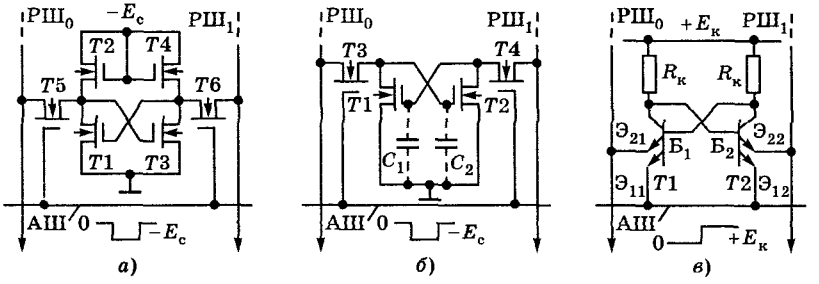


Рис. 10.23. Запоминающие ячейки оперативного ЗУ:
а — статическая на однотипных МОП-транзисторах; *б* — динамическая на однотипных МОП-транзисторах; *в* — статическая на многоэмиттерных биполярных транзисторах

Ячейка на однотипных МОП-транзисторах с *p*-каналом (рис. 10.23, *а*) имеет классическую структуру RS-триггера с управляющими ключами $T5$ и $T6$. Эти ключи нормально заперты и ЗЯ отключена от разрядных шин. При поступлении отрицательного импульса $-E_с$ на адресную шину ключи $T5$ и $T6$ отпираются и подключают ячейку к разрядным шинам. При этом на разрядные шины поступят уровни Q и $\overline{Q}$, записанные в ячейке. В режиме записи на адресную шину также подается импульс $-E_с$, а на разрядные шины подаются необходимые (взаимно противоположные) уровни, которые приводят ячейку в соответствующее состояние. Таким образом, импульс на адресной шине в обоих режимах играет роль тактового импульса.

На рис. 10.23, *б* показана ЗЯ динамического типа, в которой хранение бита информации осуществляется емкостями C_1 и C_2 (обычно это паразитные емкости МОП-транзисторов). Методика записи и считывания — та же, что и в предыдущей ячейке. Пусть при записи на разрядные шины $РШ_1$ и $РШ_0$ поданы соответственно уровни $-E_с$ и 0. Уровень $-E_с$ через ключ $T4$ поступит на затвор транзистора $T1$, и этот транзистор будет открыт. На затвор транзистора $T2$ поступит уровень 0, и этот транзистор будет заперт. Напряжения на емкостях будут иметь значе-

ния $U_{c1} = -E_c$, $U_{c2} = 0$. Если остаточный ток запертого транзистора $T2$ достаточно мал, то емкость C_1 будет разряжаться весьма медленно и, следовательно, напряжения $-E_c$ и 0 будут сохраняться на выходах ячейки (на стоках) длительное время. За это время можно несколько раз считывать информацию из ячейки (хотя при считывании емкость дополнительно шунтируется цепями считывания и ее разряд ускоряется). Для того чтобы поддерживать напряжение на емкости, несмотря на неизбежный ее разряд, осуществляют *регенерацию*, т.е. периодически производят запись того же кода. Динамические ЗЯ значительно экономичнее статических, так как у них отсутствует источник питания и, следовательно, в режиме хранения они не потребляют мощность.

В целом ЗЯ на МОП-транзисторах экономичнее и компактнее, чем ЗЯ на биполярных транзисторах, но уступают им по быстродействию. Поэтому, хотя МОП-транзисторные ОЗУ имеют большее распространение, важное место занимают и ОЗУ на биполярных транзисторах. Пример ячейки такого ОЗУ, выполненной на многоэмиттерных транзисторах, показан на рис. 10.23, *в*. Принципы записи и считывания здесь такие же, как и в предыдущих ЗУ, если не считать положительных полярностей логических уровней и адресного (тактового) импульса.

Пусть в режиме хранения транзистор $T2$ заперт, а $T1$ насыщен, тогда $U_{б1} = U^*$ и $U_{б2} \approx 0$. Если на разрядных шинах задан небольшой (0,1–0,2 В) «дежурный» потенциал, то эмиттерный переход $Э_{21}$ будет практически заперт и весь ток будет проходить через $Э_{11}$; в транзисторе $T2$ будут заперты оба эмиттерных перехода.

При считывании, когда на адресную шину подается положительное напряжение $E_к$, эмиттер $Э_{11}$ запирается и ток транзистора $T1$ протекает в шину $РШ_0$ через эмиттер $Э_{21}$, который остался под низким потенциалом; шина $РШ_1$ остается обесточенной.

При записи одновременно с адресным импульсом подается напряжение $+E_к$ на ту разрядную шину, которая связана с транзистором, подлежащим запиранию. В нашем примере, если подать напряжение $E_к$ на шину $РШ_1$, то транзистор $T2$ останется запертым и состояние ЗЯ не изменится. Если же подать напряжение $E_к$ на шину $РШ_0$, то окажутся запертыми оба эмиттера транзистора $T1$. Тогда ток потечет через базу транзистора $T2$ в эмиттер $Э_{22}$, который находится под низким потенциалом шины $РШ_1$. При этом транзистор $T2$ откроется, т. е. состояние ЗЯ изменится на обратное.

В номенклатуру параметров интегральных ОЗУ входят следующие основные величины.

Информационная емкость ОЗУ в битах — параметр, характеризующий степень интеграции элементов на кристалле.

Удельная мощность ОЗУ, т. е. общая мощность, потребляемая в режиме хранения, отнесенная к 1 биту.

Минимальный период обращения ($T_{\text{обр мин}}$) — минимально допустимый интервал между началом одного цикла считывания и началом второго. Величина, обратная $T_{\text{обр мин}}$, называется *максимальной частотой обращения*. При записи оба эти параметра могут быть несколько иными.

Удельная стоимость одного бита информации, т. е. общая стоимость кристалла, поделенная на информационную емкость. Этот параметр — один из решающих при сравнительных оценках.

МОП-транзисторные ОЗУ в целом превышают биполярные по информационной емкости, удельной мощности и удельной стоимости, но значительно уступают по быстродействию. Особое место среди биполярных ОЗУ занимают схемы в базисе $И^2Л$, которые по емкости и удельной мощности приближаются к МОП-транзисторным. Среди последних минимальная удельная мощность свойственна КМОП-схемам, а минимальная стоимость — динамическим типам ОЗУ. Среди биполярных разновидностей максимальным быстродействием характерны ОЗУ в базисе ЭСЛ.

Постоянные запоминающие устройства (ПЗУ). Как уже отмечалось, этот тип ЗУ характерен тем, что запись информации производится либо «раз и навсегда», либо представляет собой специальную, редко применяемую операцию. Поэтому использование ПЗУ состоит в считывании информации.

Типичная схема диодного ПЗУ показана на рис. 10. 24. Структура схемы уже известная — матричная; строки образуются адресными шинами, а столбцы — разрядными. Каждая адресная шина хранит определенный код: 0011, 0100 и т. д., как показано на рисунке. Запись кода (в принципе —

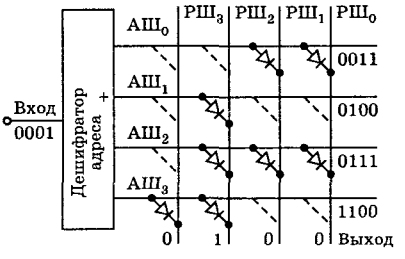


Рис. 10.24. Диодное постоянное ЗУ

однократная) осуществляется с помощью диодов, которые присоединены между адресными шинами и теми разрядными шинами, на которых (при считывании) должна быть логическая 1.

Пусть дешифратор адреса выбрал адресную шину $АШ_1$ (знак «+» на рис. 10.24). Тогда напряжение с этой адресной шины поступает на разрядную шину $РШ_2$, а на разрядных шинах $РШ_0$, $РШ_1$ и $РШ_3$ напряжение будет равно нулю. Значит, при параллельном считывании информации со всех четырех разрядных шин получим код (слово) 0100, записанный в выбранной строке.

При проектировании интегральных матриц ПЗУ нецелесообразно размещать диоды сразу в тех узлах, где они образуют необходимые коды. Номенклатура матриц с разными вариантами кодовых наборов оказывается слишком широкой, а тираж каждого варианта — слишком малым, не оправданным с экономической точки зрения. Поэтому диоды изготавливают во всех узлах матрицы и в таком однородном виде матрицу поставляют заказчику. Каждый заказчик сам записывает в ПЗУ нужные ему коды. Для этого он (с помощью специальных устройств) *пережигает* выводы — перемычки тех диодов, которые находятся в местах расположения логических нулей. Такие диоды с перегоревшими выводами показаны на рис. 10.24 штриховыми линиями. Пережигание выводов осуществляется индивидуально, путем пропускания через соответствующий диод большого тока, на который диод (в нормальном режиме) не рассчитан.

Использование МОП-транзисторов открывает дополнительные возможности при использовании ПЗУ. Во-первых, запись информации может осуществляться не пережиганием выводов, а более «тонким» способом. Однородная матрица ПЗУ (с МОП-транзисторами во всех узлах) поставляется заказчику в виде «полуфабриката»: без металлизации затворов. Заказчик сам осуществляет последнюю фотолитографию. При этом он использует такой фотошаблон, который обеспечивает металлизацию затворов только у тех транзисторов, которые должны передавать «1» на разрядную шину. У остальных транзисторов затворы не будут присоединены к адресным шинам и эти транзисторы будут бездействовать.

Во-вторых, использование МОП-транзисторов позволяет осуществить так называемые *полупостоянные* или *программируе-*

мые ПЗУ (их обозначают ППЗУ), в которых можно время от времени менять записанную информацию. Такая возможность очень полезна для заказчика, хотя ее реализация и связана с некоторыми техническими трудностями. Общая идея, лежащая в основе ППЗУ, состоит в том, чтобы *обратимо менять пороговое напряжение МОП-транзистора.*

Управление пороговым напряжением достигается в настоящее время двумя способами. Оба они основаны на введении дополнительных зарядов в диэлектрик. Первый способ состоит в использовании МНОП-транзисторов, описанных в разд. 7.8. У этих транзисторов введение и выведение зарядов осуществляется с помощью коротких высоковольтных импульсов разной полярности, подаваемых на затвор (рис. 7.33, *б*). Второй способ, практикуемый при использовании «обычных» МОП-транзисторов с однослойным диэлектриком, состоит в следующем. Прикладывая к затвору достаточно большое напряжение, вызывают лавинный пробой в диэлектрике, в результате чего в нем накапливаются электроны. Соответственно меняется пороговое напряжение. Заряд электронов сохраняется в течение весьма длительного времени, как и у МНОП-транзисторов (в случае необходимости заряд можно периодически восстанавливать — регенерировать). Для того чтобы осуществить перезапись информации, нужно удалить электроны из диэлектрика. Это достигается путем освещения кристалла ультрафиолетовым светом, который вызывает фотоэффект, т. е. выбивание электронов из диэлектрика.

10.9. Большие и сверхбольшие интегральные схемы

Тенденция к повышению степени интеграции наблюдалась с самого зарождения микроэлектроники. Сначала в каждом корпусе размещались отдельные ИЛЭ. Затем, увеличив количество выводов, стали размещать по несколько ИЛЭ в одном корпусе. Это позволило сократить общее количество корпусов в аппаратуре, но не привело к какому-либо новому этапу в развитии ИС. Качественно новый этап начался лишь после того, как простые ИС, расположенные на одном кристалле, стали объединять в сложные функциональные комплексы путем металлической разводки — так же, как в самих ИС объединяются отдель-

ные элементы. На этом этапе появились сначала *средние* (СИС), а затем *большие интегральные схемы* (БИС) и *сверхбольшие интегральные схемы* (СБИС). Можно сказать, что в основе БИС лежит интеграция простых ИС.

Общая характеристика БИС. Классическим примером простых ИС являются логические вентили типа ТТЛ, КМОП, ЭСЛ и др. Промежуточное место между простыми ИС и СИС занимают JK-триггеры, состоящие из 8–10 логических вентилей. Собственно к СИС относятся сумматоры, счетчики, ОЗУ и ПЗУ емкостью 256–1024 бит и др. Примерами БИС могут служить схемы памяти на 4 бит и более, арифметико-логические и управляющие устройства ЭВМ, цифровые фильтры. Наибольшая степень интеграции свойственна однородным структурам — ЗУ и составляет в настоящее время десятки миллионов элементов на кристалле. ИС со степенью интеграции более 100 тыс. элементов называют сверхбольшими.

Использование БИС сопровождается резким улучшением всех основных показателей по сравнению с аналогичным функциональным комплексом, выполненным на отдельных ИС. Действительно, интеграция ИС на одном кристалле приводит к уменьшению количества корпусов, числа сборочных и монтажных операций, количества внешних — наименее надежных — соединений. Все это, как обычно, способствует уменьшению размеров, массы, стоимости и повышению надежности. Дополнительными преимуществами от интеграции ИС являются уменьшение общего количества контактных площадок (а значит, экономия площади), сокращение длины соединений (что способствует повышению быстродействия и помехоустойчивости), а также меньший разброс параметров, поскольку все ИС расположены на одном кристалле и изготовлены в едином технологическом цикле.

Повышение степени интеграции можно обеспечить двумя путями. *увеличивая плотность упаковки элементов* (т. е. уменьшая их площадь, включая площадь металлической разводки) и *увеличивая размеры кристалла.* И тот, и другой путь связан с решением сложных технологических задач: необходимо повысить разрешающую способность литографии, стабилизировать технологические режимы, обеспечить особую гигиену производства, снизить плотность дефектов на поверхности кремния и т. п. Некоторые проблемы конструирования и контроля БИС рассматриваются в следующем разделе.

На этапе развития БИС возник также ряд схемотехнических проблем, не менее серьезных, чем технологические. Более того, обе категории проблем тесно переплелись друг с другом, т. е. *вопросы «как делать» и «что делать» должны при разработке БИС решаться одновременно и комплексно.*

БИС на базовых матричных кристаллах. Один из первоочередных вопросов, возникающих при разработке БИС, — это вопрос об их технически и экономически целесообразной сложности. Необходимо сочетать достаточную сложность (чтобы реализовать преимущества высокой степени интеграции) с достаточной универсальностью (чтобы обеспечить экономически оправданный объем выпуска). Практика показывает, что такого компромисса можно достигнуть, обеспечивая *элементную избыточность и многофункциональность.*

Действительно, если количество ИС, расположенных на кристалле, больше, чем нужно для осуществления некоторой определенной функции, то, значит, выполняя разные металлические разводки, можно с помощью одного и того же набора ИС осуществлять БИСы с разными функциями. Данный принцип положен в основу БИС на базовых матричных кристаллах. Базовый матричный кристалл (БМК) — это набор топологических ячеек, расположенных в виде матрицы, между элементами которой отсутствуют соединения.

Топологическая ячейка состоит из отдельных полупроводниковых элементов. Для получения БИС с заданными функциями отдельные элементы внутри ячеек и сами ячейки соединяются токоведущими дорожками. Для размещения проводников (трассировки) используют системы автоматизированного проектирования (САПР).

В конкретной БИС, изготовленной на БМК, обычно используют не все топологические ячейки, что определяется функциональными особенностями микросхемы.

Другой принцип формирования заданной структуры БИС на БМК состоит в первоначальном создании кристалла, в котором выполнены все возможные межсоединения элементов. Далее, в нужных местах межсоединений делаются разрывы путем локального удаления материала межсоединения методом лазерного разрушения.

Программируемые логические матрицы. Существенным недостатком описанных выше БИС на БМК является то, что фор-

мирование структуры БИС может быть выполнено только в рамках микроэлектронного производства в соответствии с логической структурой, определенной разработчиком. Существенно бóльшую гибкость обеспечивают программируемые пользователем логические матрицы (ПЛМ). Это специализированные БИС, внутренняя структура которых, подобно обычным БМК, состоит из матрицы базовых логических ячеек и межсоединений, но конфигурация отдельных ячеек и связей между ними осуществляется с помощью специальной системы, расположенной на этом же кристалле.

Обычно для задания конфигурации предусмотрены специальные выводы БИС, и разработчик может устанавливать и изменять логическую функцию БИС. Для разных типов ПЛМ установленная конфигурация может либо сохраняться при отключении питания (подобно постоянным запоминающим устройствам, см. разд. 10.8), либо требуется перезагрузка при каждом новом включении. Управление перезагрузкой может выполнять сама схема ПЛМ, считывая информацию о конфигурации из внешнего ПЗУ.

Проблемы повышения степени интеграции. Опыт разработки БИС выявил ряд общих проблем, которые ограничивают повышение степени интеграции и которые, следовательно, нужно так или иначе решать в процессе дальнейшего развития микроэлектроники.

Проблема теплоотвода. При заданных размерах элементов повышение степени интеграции может достигаться увеличением плотности компоновки, т. е. сближением элементов на кристалле. При этом неизбежно возрастает удельная мощность, рассеиваемая на единице площади. При современных конструкциях кремниевых ИС допустимая удельная мощность на кристалле без дополнительного теплоотвода не превышает 5 Вт/см^2. Значит, допустимая мощность для кристалла площадью 20 мм^2 составляет не более 1 Вт. При средней мощности 0,5 мВт, потребляемой одним ИЛЭ, на указанном кристалле удастся разместить не более 2000 логических элементов.

Естественным путем для преодоления этого ограничения является использование микрорежима транзисторов и таких схем, которым микрорежим свойствен. Например, для того чтобы на той же площади 20 мм^2 разместить 10000 вентилей, нужно использовать ИЛЭ с потребляемой мощностью не более 0,1 мВт, т. е. ИЛЭ типа КМОП (см. табл. 10.1).

Конечно, может оказаться, что при данных размерах кристалла желательную степень интеграции нельзя осуществить ни на одном из имеющихся элементных базисов. Тогда приходится идти на увеличение площади ИС. В принципе этот путь открывает широкие возможности, но практически он тоже ограничен.

Ограничение накладывается неизбежными дислокациями на поверхности полупроводника (см. разд. 2.2). Любая дислокация в пределах БИС означает негодность транзистора или отдельной ИС, соответственно негодной может оказаться и БИС в целом. Поэтому увеличение площади кристалла сопровождается увеличением процента брака и уменьшением процента выхода годных БИС.

В отдельных случаях используют искусственное охлаждение БИС или подложек, на которых они монтируются. Однако такое решение — не универсальное и экономически далеко не всегда оправдано.

Проблема межсоединений. Внутренняя структура БИС настолько сложна, что конструктор не может за разумное время спроектировать топологию (расположение) элементов и рисунок оптимальных межсоединений (с минимальной суммарной длиной и минимальным количеством пересечений). Для этого нужно сравнить тысячи вариантов, и это практически можно выполнить только при использовании систем автоматизированного проектирования.

Опыт показывает, что в большинстве БИС не удается расположить разводку межсоединений в одной плоскости без пересечений. Поэтому для БИС характерна *многослойная разводка*, расположенная обычно в 2-х или 3-х плоскостях. Изоляция слоев друг от друга и необходимые соединения между разводками разных слоев представляют собой особую технологическую проблему, специфичную для БИС.

Проблема контроля параметров. Электрический контроль параметров БИС до ее помещения в корпус осуществляется с помощью измерительных зондов, прижимаемых к контактным площадкам, т. е. к будущим внешним выводам. Зонды представляют собой тонкие металлические проволочки, острие которых имеет диаметр 5–10 мкм. Зонды объединяются в зондовые головки — своего рода проволочные «щетки», в которых каждый зонд соприкасается с соответствующей контактной площадкой,

имеющей размеры 100 × 100 мкм. Количество внешних выводов у БИС значительно больше, чем у простых ИС, в силу большей сложности выполняемых функций. Оно может составлять от десятков до нескольких сотен. Если для иллюстрации принять 50 выводов и учесть, что на каждом выводе может быть два значения выходной величины («0» или «1»), то для полноценной проверки функционирования БИС (только в статике) потребуется $2^{50} \approx 10^{15}$ измерений. При длительности каждого измерения 1 мкс контроль одной БИС займет около 25 лет.

Следовательно, помимо автоматизации контроля, нужно упростить и его методику. По необходимости измерения должны быть выборочными: количество измерений, свидетельствующих о работоспособности БИС (с определенной вероятностью), обычно лежит в пределах 200–300.

Отбор контролируемых параметров, последовательность и правила (алгоритмы) их испытания, а также разработка соответствующей аппаратуры и программ (для использования ЭВМ) представляют нередко задачу, не менее сложную, чем проектирование самой БИС.

Физические ограничения на размеры элементов. В современных БИС размеры отдельных участков доходят до 2–5 мкм и имеется тенденция к дальнейшему уменьшению размеров. На этом пути, однако, возникают некоторые принципиальные ограничения. В упрощенном виде они проиллюстрированы на рис. 10.25.

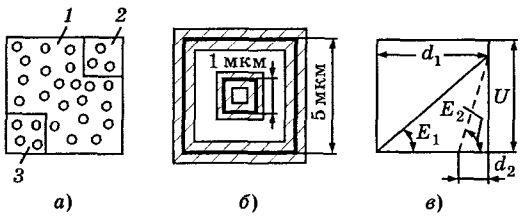


Рис. 10.25. Ограничения на минимальные размеры элементов ИС: *а* — флюктуация распределения примеси; *б* — технологические допуски; *в* — рост напряженности поля

Во-первых, с уменьшением площади начинает сказываться неравномерное (статистическое) распределение примеси в полупроводнике. Пусть квадраты на рис. 10.25, *а* иллюстрируют конфигурацию эмиттерного слоя. При большой площади (квадрат *1*) количество атомов примеси в двух одинаковых квадратах

будет практически одинаковым. При малой площади (квадраты *2* и *3*) количество атомов может заметно различаться (три и четыре атома на рис. 10.25, *а*). Соответственно будут различаться концентрации примеси в эмиттерах, а значит, и коэффициент инжекции [см. (5.24)]. Анализ показывает, что этот фактор становится существенным при стороне квадрата менее 1 мкм.

Во-вторых, с уменьшением линейных размеров возрастает роль технологических допусков (рис. 10.25, *б*). Так, если погрешность фотолитографии составляет ±0,2 мкм, то при линейных размерах 5 мкм (большой квадрат) площади элементов будут различаться незначительно (на 20%), а при размерах 1 мкм (малый квадрат) — в 2,3 раза.

В-третьих, с уменьшением линейных размеров возрастают напряженности электрических полей в полупроводниковых слоях (рис. 10.25, *в*). При одном и том же напряжении $U = 0{,}2$ В напряженность поля в слое толщиной 5 мкм сравнительно невелика (400 В/см), а в слое толщиной 0,2 мкм она составляет 10^4 В/см, т. е. превышает критическую напряженность. Соответственно полупроводниковый слой приобретает нелинейные свойства.

Можно также показать, что при линейных размерах менее 1–2 мкм определенную роль начинают играть шумовые флюктуации, влияние космического излучения и естественный радиационный фон Земли.

Учитывая, что возможности обычной фотолитографии тоже лежат в пределах 0,7–1 мкм, можно сказать, что и физические и технологические аспекты делают область размеров менее 1 мкм особой областью и позволяют говорить о «субмикронной микроэлектронике» как о самостоятельном научно-техническом и технологическом направлении.

Большие гибридные интегральные схемы (БГИС). Этот тип БИС не является альтернативой полупроводниковым БИС. Скорее его можно считать адекватным конструктивным решением при разработке современной микроэлектронной аппаратуры.

Основное различие между простой ГИС и БГИС состоит в том, что у первых типичными навесными компонентами являются бескорпусные транзисторы и диоды, тогда как у вторых — бескорпусные ИС и БИС. Кроме того, БГИС имеют более сложную структуру, чем ГИС. В частности, им часто свойственна многоуровневая металлизация, как и полупроводниковым БИС.

Чаще всего БГИС содержит только металлическую разводку и навесные бескорпусные ИС и БИС. Поэтому понятие ГИС (подразумевающее наличие пленочных пассивных элементов) часто сводится к понятию тонко- или толстопленочной *коммутационной платы*, главная задача которой — объединить множество ИС и БИС в единый функциональный комплекс. Такой комплекс называют *микросборкой*.

Коммутационная плата является микроэлектронным аналогом печатной платы, которая до последнего времени была основной конструктивной единицей радиоэлектронной аппаратуры. Что касается микросборки, то ее качественная особенность по сравнению с узлами, выполненными на печатных платах, состоит в том, что она представляет собой законченный прибор (суперкомпонент электронных схем), снабженный отдельным корпусом и характеризуемый единой технической документацией. Функциональная сложность такого электронного прибора намного выше, чем БИС и даже СБИС. Из микросборок, как «суперкомпонентов», можно на печатной плате монтировать сверхсложные узлы аппаратуры, а часто и все устройство в целом.

10.10. Операционные усилители

Аналоговые устройства по сравнению с цифровыми характерны гораздо большим разнообразием по виду сигналов, по выполняемым функциям, а также по назначению и внутренней структуре. Поэтому унификация элементной базы в области аналоговых устройств возможна только на основе многофункциональных узлов.

В настоящее время *основными аналоговыми функциями* принято считать: усиление, сравнение, ограничение, перемножение и частотную фильтрацию сигналов. Каждая из этих функций, вообще говоря, выполняется особым классом аналоговых ИС. Однако все эти специализированные ИС, как правило, происходят от основного, наиболее универсального и многофункционального узла — *операционного усилителя* (ОУ), которому и посвящен данный раздел.

Общая характеристика. Операционным усилителем (ОУ) принято называть усилитель постоянного тока с дифференциальным входом и однотактным выходом, характерный высоким коэффициентом усиления, а также большим входным и малым вы-

ходным сопротивлениями[1]. Условное обозначение ОУ показано на рис. 10.26.

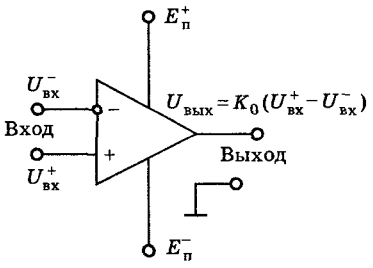


Рис. 10.26. Условное обозначение операционного усилителя

Сигнал не обязательно должен быть дифференциальным, его можно подавать на один из входов ОУ, заземляя второй. В зависимости от полярности сигналов на выходе и входе один из входов называют *инвертирующим* (обозначается «−»), а другой *неинвертирующим* (обозначается «+»). Иногда инвертирующий вход обозначается кружком, как и в логических вентилях (рис. 10.26). Выходное напряжение операционного усилителя $U_{вых}$ определяется разностью напряжений на неинвертирующем и инвертирующем входах, умноженной на положительный коэффициент усиления K_0. На практике ОУ чаще всего бывает охвачен цепью глубокой обратной связи. Именно в совокупности с цепями обратной связи ОУ выполняет те разнообразные *операции*, откуда и происходит его название. Типичная схема ОУ с активной отрицательной обратной связью показана на рис. 10.27, где сопротивление R_1 включает в себя сопротивление источника питания $R_г$. В таком варианте ОУ выполняет функцию стабильного усиления. Если входное сопротивление ОУ существенно превышает значение сопротивлений R_1 и R_2, то входным током операционного усилителя $I_{вх}$ можно пренебречь и считать $I_1 = I_2$. Последнее равенство можно записать в виде

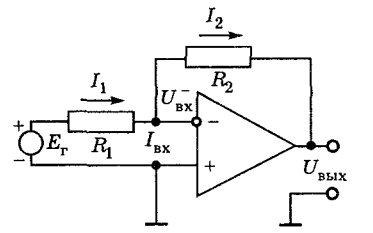


Рис. 10.27. Типичная схема включения ОУ

$$(E_г - U_{вх}^-)/R_1 = (U_{вх}^- - U_{вых})/R_2 .$$

---

1 Понятия «высокий», «большое» и «малое» применительно к параметрам ОУ будут уточнены позднее.

Подставляя $U_{вых} = -K_0 U_{вх}^-$ и проводя простые преобразования, получаем коэффициент усиления схемы:

$$K = \frac{U_{вых}}{E_г} = -\frac{R_2/R_1}{1 + (1 + R_2/R_1)/K_0}. \qquad (10.24)$$

Если коэффициент усиления ОУ достаточно велик, то вторым слагаемым в знаменателе (10.24) можно пренебречь, тогда

$$K = -R_2/R_1. \qquad (10.25)$$

Выражение (10.25) является фундаментальным для ОУ. Оно показывает, что *при определенных условиях коэффициент усиления схемы зависит только от параметров цепи обратной связи и не зависит от параметров самого ОУ.* В частности, коэффициент усиления схемы не зависит от температуры, от напряжений питания и от изменений коэффициентов β, какими бы причинами эти изменения ни вызывались. Заменяя активные сопротивления R_2 и R_1 комплексными, можно получить желательные переходные и частотные характеристики, по-прежнему не зависящие от параметров ОУ.

Уточним условия, при которых действительно выражение (10.25). Прежде всего оно базируется на высоких значениях коэффициента усиления K_0. А именно, согласно (10.24) должно выполняться неравенство

$$K_0 >> (R_2/R_1) + 1 = K + 1. \qquad (10.26)$$

Следовательно, *коэффициент усиления ОУ должен намного превышать желательный коэффициент усиления схемы.* Например, если желательно иметь $K = 100$, то коэффициент K_0 должен быть больше 10^3–10^4.

С ростом частоты величина K_0 неизбежно уменьшается, что приводит к нарушению неравенства (10.26). Поэтому *чем выше граничная частота коэффициента K_0, тем в более широком диапазоне частот действительно выражение* (10.25) со всеми его преимуществами.

Из рис. 10.26 видно, что интегральный ОУ, помимо входных и выходных зажимов имеет выводы для подачи двухполярного напряжения питания. Иногда ОУ также имеет дополнительные выводы для коррекции характеристик.

Основные параметры ОУ. Для того чтобы ОУ имел дифференциальный вход, его первый каскад должен быть дифференциальным усилителем (см. раздел 9.6). В зависимости от коэффициента усиления первого дифференциального каскада (ДК)

за ним следует либо второй ДК, либо сразу схемы сдвига уровня и другие промежуточные каскады, которые в конечном счете призваны связать ДК с мощным выходным каскадом. Последний практически всегда строится по двухтактной схеме класса В (см. раздел 10.9).

Наличие входного ДК предопределяет номенклатуру параметров ОУ: она практически такая же, как у отдельного дифференциального усилителя. Эти параметры были подробно рассмотрены в разделе 9.6. К их числу относятся: коэффициент усиления K_0, коэффициент ослабления (подавления) синфазной составляющей $K_{\text{ос сф}}$, напряжение смещения $U_{\text{см}}$ и его температурная чувствительность $\varepsilon_{\text{см}}$, средний входной ток $I_{\text{вх ср}}$ и разность входных токов $\Delta I_{\text{вх}}$. Кроме того, конечно, указываются напряжения питания $E_{\text{п}}$, потребляемые ток $I_{\text{пот}}$ и мощность $P_{\text{пот}}$, максимальные входные и выходные напряжения, максимальный выходной ток и некоторые другие.

Входное и выходное сопротивления не всегда включают в число основных параметров, но о них можно судить по значениям входных и выходных токов.

Быстродействие ОУ принято характеризовать *скоростью нарастания выходного напряжения* $v_{U\text{вых}}$; этот параметр измеряется при ступенчатом входном сигнале с максимально допустимой амплитудой[1]. Реже используют максимальную частоту или частоту единичного усиления f_1, при которой коэффициент усиления падает до значения $K_0 = 1$.

Развитие интегральных ОУ происходило определенными этапами, каждый из которых характеризовался своеобразными схемотехническими решениями и специфической элементной базой. Можно условно назвать три этапа и три поколения интегральных ОУ. Усредненные параметры этих трех поколений приведены в табл. 10.2. Все три первых поколения можно отнести к универсальным ОУ. Четвертое поколение относится к специализированным ОУ. Это значит, что параметры, приведенные для 4-го поколения в табл. 10.2, не сочетаются в одном и том же ОУ; они относятся к разным ОУ и являются «рекордными». Например, один тип ОУ может иметь минимальное напряжение смещения, но низкое быстродействие, другой — мак-

---

1 Английский термин Slew Rate (SR).

симальное быстродействие, но большую потребляемую мощность и т. п.

Таблица 10.2. Типичные параметры интегральных ОУ

Поколение ОУ	1-е	2-е	3-е	4-е	
				ср.	макс. или мин.
K_0, В/мВ	25–50	100	10^2–10^3	10^3	10^5
$K_{ос\ сф}$, дБ	70–85	80–90	90–100	120	140
$U_{см}$, мВ	2–3	1 2	0,5–1	0,25–0,5	0,01
$\varepsilon_{см}$, мкВ/°С	3–5	3	2–3	0,5	0,1
$I_{вх\ ср}$, нА	500	100	2–5	0,1	0,001
$\Delta I_{вх}$, нА	200	20–50	0,1–0,2	0,005	0,001
$v_{U вых}$, В/мкс	0,5–1	0,5	0,2–2	50–70	1000
$P_{пот}$, мВт	10–50	5–20	0,2–5	0,08–0,15	0,001
$E_п$, В	±5...±20	±3...±20	±2.. ±15	±(2–5)	±1,5

Из табл. 10.2 следует, что основными тенденциями в развитии универсальных ОУ были повышение коэффициента усиления и, главное, уменьшение входных токов. Последнее способствовало не только увеличению входного сопротивления, но и улучшению точностных показателей (см. разд. 9.6).

Типичными параметрами современных универсальных ОУ можно считать:

$K_0 = 10^5$–10^6; $K_{ос\ сф} = 80$–100 дБ;

$U_{см} = 1$–3 мВ; $\varepsilon_{см} = 2$–3 мкВ/°С;

$I_{вх\ ср} = 5$–50 нА; $\Delta I_{вх} = 1$–10 нА.

Что касается 4-го (специализированного) поколения ОУ, то у него существенные достижения по отдельным параметрам достигаются разными средствами; некоторые из них описаны ниже.

Как известно, в интегральном исполнении количество транзисторов мало критично и потому разработчик может позволить себе вводить в ОУ дополнительные элементы и каскады ради повышения качества усилителя. Кроме того, близкое рас-

положение элементов на одной подложке способствует большей симметрии плеч ДК, а значит, улучшению точностных показателей схемы. Поэтому *параметры интегральных ОУ всегда существенно выше, чем у аналогичных дискретных схем.*

Схемные решения. Каждое поколение интегральных ОУ характерно своими схемными решениями. В ОУ первого поколения использовались биполярные транзисторы только одного типа (n–p–n) и диффузионные резисторы в качестве нагрузок дифференциального каскада. Операционные усилители второго поколения характеризовались целым рядом усовершенствований. Во-первых, в них наряду с транзисторами n–p–n использовались и транзисторы p–n–p; это, в частности, облегчало построение выходных каскадов (см. рис. 9.25). Во-вторых, наряду с простыми диффузионными резисторами использовались пинч-резисторы, имеющие более высокие номиналы сопротивлений (см. разд. 7.9). В-третьих, наряду с биполярными транзисторами, во входных ДК иногда использовались полевые транзисторы с p–n-переходом Они уступали биполярным по усилительным и частотным свойствам, но обеспечивали резкое уменьшение входных токов и увеличение входного сопротивления. Однако главная специфика 2-го поколения ОУ состояла в замене резисторных нагрузок ДК динамическими. Примеры динамических нагрузок были рассмотрены выше применительно к МДП-транзисторам (см. рис. 9.13).

Во многих операционных усилителях ДК используют только на входе устройства, остальные каскады усиления выполняют по однофазной схеме. На рис. 10.28 приведена упрощенная схема типового ДК с однофазным выходом на биполярных транзисторах, использующего динамическую нагрузку в виде отражателя тока в коллекторных цепях.

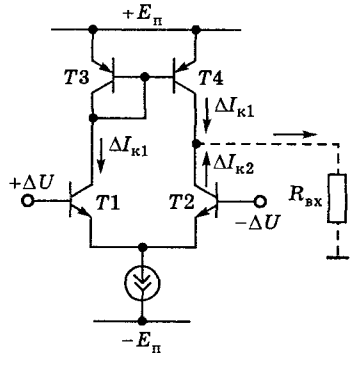


Рис. 10.28. Дифференциальный каскад с динамической нагрузкой

Поскольку транзисторы $T2$ и $T4$ соединены «коллектор к коллектору», можно сказать, что нагрузкой транзистора $T2$ является сопротивление $r_{к4}$, а нагрузкой транзистора $T4$ — сопротивление $r_{к2}$. Оба эти сопротивления очень велики, особенно в микрорежиме [см. (5.44)].

Практически нагрузкой обоих транзисторов является меньшее сопротивление $R_{вх}$ — входное сопротивление следующего каскада (на рис. 10.28 показано штриховой линией).

Пусть на базы транзисторов $T1$ и $T2$ поступают соответственно сигналы $+\Delta U$ и $-\Delta U$. При этом коллекторный ток $I_{к1}$ изменится на величину $\Delta I_{к1} = \alpha(\Delta U/r_э)$. Приращение $\Delta I_{к1}$ будет повторено отражателем тока в коллекторной цепи транзистора $T4$ и даст приращение коллекторного потенциала $\Delta U_{к2} \approx \Delta I_{к1} R_{вх}$. Точно такая же величина получится под действием приращения $\Delta I_{к2}$, обусловленного сигналом $-\Delta U$. В результате получаем на выходе: $\Delta U_{к2} = 2\Delta U(R_{вх}/r_э)$, где $R_{вх}/r_э$ — коэффициент усиления. Таким образом, *использование отражателя тока позволяет не только получить высокий коэффициент усиления (до нескольких тысяч), но и удвоить сигнал на однотактном выходе ДК.*

3-е поколение интегральных ОУ характерно использованием во входных ДК супербета транзисторов (см. разд. 7.4). Высокие значения коэффициента β, свойственные таким транзисторам, в сочетании с микрорежимом обеспечивают практически такие же малые входные токи, как и полевые транзисторы, но быстродействие и усиление оказываются выше.

Способы уменьшения дрейфа. Симметрия интегрального ДК в значительной мере решает проблему дрейфа. Однако в ряде специальных случаев, особенно в прецизионных измерительных устройствах, эта проблема сохраняет свое значение. Рассмотрим два способа повышения температурной стабильности, характерные для интегрального исполнения.

Первый способ состоит в стабилизации температуры кристалла, на котором расположен ОУ. При этом температурный дрейф будет отсутствовать или во всяком случае существенно уменьшится. Структурная схема стабилизации температуры показана на рис. 10.29. На том же кристалле, на котором расположен ОУ, расположены дополнительные стабилизирующие элементы и узлы: датчики температуры Д (обычно прямосмещенные p–n-переходы), вспомогательный усилитель У и тепловыделяющие элементы ТЭ (мощные транзисторы). Запишем следующее соотношение:

$$\Delta T_{ИС} = \Delta T_0 + \Delta P_{ст} R_t, \qquad (10.27)$$

где $\Delta T_{ИС}$ — изменение температуры кристалла, ΔT_0 — изменение внешней температуры, $\Delta P_{ст}$ — приращение мощности в тепловыделяющих элементах, R_t — тепловое сопротивление кристалла. Учитывая, что тепловыделяющие элементы могут только нагревать кристалл, но не охлаждать его, нетрудно сделать вывод, что *данная ИС всегда работает при максимально допустимой температуре*. Это — один из недостатков такого метода. Вторым недостатком является необходимость дополнительной площади (до 20%) под элементы стабилизатора.

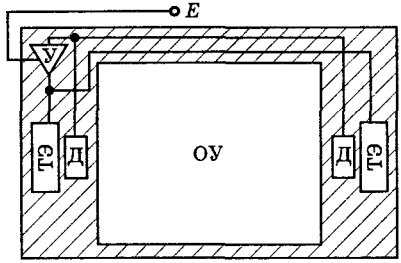


Рис. 10.29. Структурная схема стабилизации температуры кристалла

Обозначим температурную чувствительность датчиков через ε (В/°С), коэффициент усиления вспомогательного усилителя через $K_у$ и крутизну ВАХ тепловыделяющих элементов через $S_{ТЭ}$. Тогда приращение мощности $\Delta P_{ст}$ под действием приращения $\Delta T_{ИС}$ запишется в виде

$$\Delta P_{ст} = (\varepsilon \Delta T_{ИС}) K_у S_{ТЭ} E,$$

где E — напряжение питания. Подставляя величину $\Delta P_{ст}$ в (10.27), получаем:

$$\Delta T_{ИС} = \Delta T_0 / (K_{ст} + 1), \qquad (10.28)$$

где коэффициент стабилизации:

$$K_{ст} = -\varepsilon K_у S_{ТЭ} E R_t.$$

Очевидно, что знак коэффициента усиления $K_у$ должен быть противоположным знаку ε . Для типичных значений ε = −2,2 мВ/°С; $S_{ТЭ}$ = 40 мА/В; E = 10 В; $K_у$ = −200 и R_t = 0,3 °С/мВт получаем: $K_{ст}$ = 50. Значит, в диапазоне ΔT_0 = 150 °С температура кристалла будет меняться всего на 3 °С (при этом его рабочая температура будет около +120 °С).

Радикальным способом устранить влияние дрейфа (но не сам дрейф) является отказ от обычных *схем прямого усиления* и использование *усилителей типа М–ДМ* (название происходит от слов модуляция—демодуляция). В обычных усилителях медленно меняющиеся сигналы неотделимы от дрейфа, потому что совпадают их частотные спектры. Идея М–ДМ-усилителей состоит в том, что полезный сигнал модулируется, т. е. его спектр смещается в область сравнительно высоких частот при

сохранении амплитуды (рис. 10.30). После этого можно использовать обычный усилитель (У), не опасаясь того, что сигнал и дрейф «смешаются». На выходе усилителя сигнал демодулируется, т. е. восстанавливается его исходный спектр.

Усилители типа М–ДМ в дискретных схемах не получили распространения, поскольку, во-первых, требуют дополнительных активных элементов, а во-вторых, нуждаются в высококачественных модуляторах — прерывателях сигнала. В интегральном исполнении усилители М–ДМ оказываются достаточно простыми, а высококачественная модуляция и демодуляция достигаются использованием МДП-транзисторных ключей, характерных отсутствием остаточного напряжения в замкнутом состоянии (см. разд. 8.6). Управление ключами осуществляется с помощью генератора прямоугольных импульсов, расположенного на том же кристалле (ГПИ на рис. 10.30).

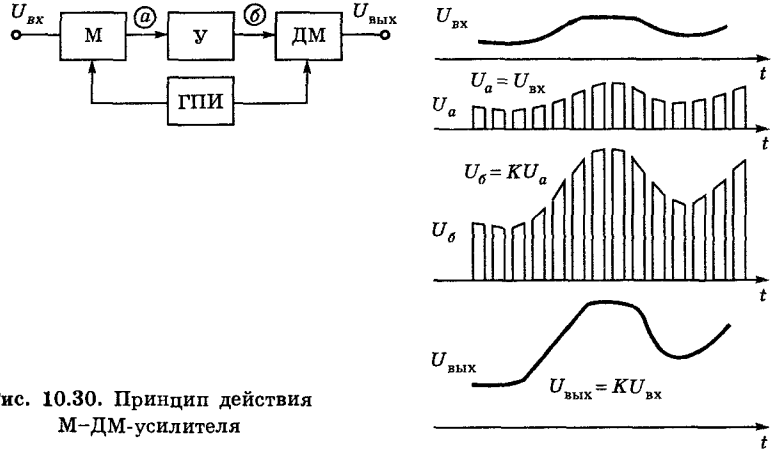


Рис. 10.30. Принцип действия М–ДМ-усилителя

Усилители типа М–ДМ обеспечивают не только малый температурный дрейф (0,2 мкВ/°С, см. табл. 10.2), но и пониженный уровень низкочастотных шумов.

Многофункциональность ОУ. Прежде чем рассмотреть примеры использования ОУ, заметим, что анализ и расчет многих схем, содержащих ОУ, упрощается, если считать входное напряжение ОУ равным нулю. Такой прием оправдан тогда, когда ОУ заведомо работает в нормальном линейном режиме. При этом входное напряжение в K_0 раз меньше выходного. Последнее обычно не превышает 5–10 В. Следовательно, значения $U_{вх}$,

как правило, составляют всего десятки микровольт и менее, т. е. на несколько порядков меньше других напряжений в схемах.

В ряде случаев ОУ работает в нелинейном режиме, когда коэффициент K_0 перестает быть коэффициентом пропорциональности между входным и выходным напряжениями. Действительно, выходное напряжение не может превышать напряжение питания. Поэтому, если подать на вход достаточно большой сигнал $U_{вх} > E_п/K_0$, то выходной (или ближайший к нему) транзистор окажется либо запертым, либо насыщенным (см. разд. 9.6). При этом выходное напряжение принимает одно из предельных значений $+E_1$ или $-E_2$ и не зависит от входного сигнала.

На рис. 10.31 показана типичная схема стабилизатора напряжения. Считая $U_{вх} = 0$ (см. выше), можно записать: $U_a = U_{оп}$, где $U_{оп}$ — опорное напряжение. Выражая U_a через выходное напряжение в виде $U_a = \gamma U_a$, получаем:

$$U_2 = U_{оп}/\gamma,$$

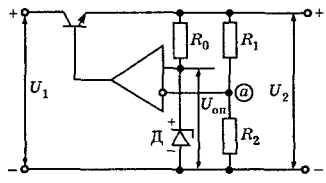


Рис. 10.31. Стабилизатор напряжения с использованием ОУ

где $\gamma = R_2/(R_1+R_2)$. Меняя сопротивления R_1 и R_2, можно регулировать выходное напряжение.

Использование ОУ позволяет решить одну из главных проблем — резко уменьшить выходное сопротивление стабилизатора по сравнению с простейшей схемой (рис. 9.32). Действительно, зададим на выходе приращение ΔU_0. Усиленное приращение $K_0\Delta U_0$, поступая на базу транзистора, вызовет приращение эмиттерного тока $\Delta I_э \approx K_0\Delta U_0/r_э$. Деля ΔU_0 на $\Delta I_э$, получаем выходное сопротивление в виде

$$R_{вых} \approx r_э/K_0.$$

Расчетное значение $R_{вых}$ может составить тысячные доли ома и менее. В этом случае фактическое значение часто определяется сопротивлением металлизации или проводников. Значения коэффициента стабилизации оказываются больше, чем в простейшей схеме.

На рис. 10.32 показана схема так называемого *компаратора*, обеспечивающего сравнение двух напряжений.

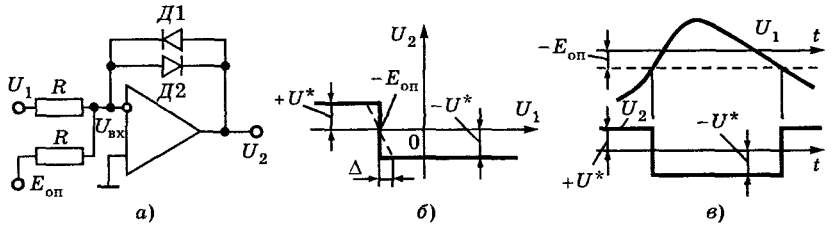


Рис. 10.32. Компаратор напряжений: *а* — схема; *б* — передаточная характеристика; *в* — функция сравнения

Пусть сначала $U_1 = -E_{оп}$, где $E_{оп}$ — постоянное опорное напряжение. Тогда, при одинаковых сопротивлениях R, потенциал на инвертирующем входе будет равен полусумме величин U_1 и $E_{оп}$, т. е. нулю. Соответственно $U_2 = 0$ и оба диода заперты. Если теперь увеличить входное напряжение на величину ΔU_1, то потенциал инвертирующего входа станет положительным и на выходе появится отрицательное напряжение U_2. При этом откроется диод $Д2$. Как известно, напряжение на открытом диоде — величина практически постоянная, равная U^*. Принимая $U_{вх} = 0$ (см. начало раздела), приходим к выводу: после отпирания диода $Д2$ выходное напряжение равно $-U^*$ независимо от значения U_1. Если $\Delta U_1 < 0$, то отпирается диод $Д1$, выходное напряжение делается равным $+U^*$ и тоже не зависит от U_1.

Величину $U_{вх}$, при которой отпирается тот или иной диод, называют чувствительностью компаратора Δ. Ее легко оценить, поделив выходное напряжение U^* на коэффициент усиления K_0:

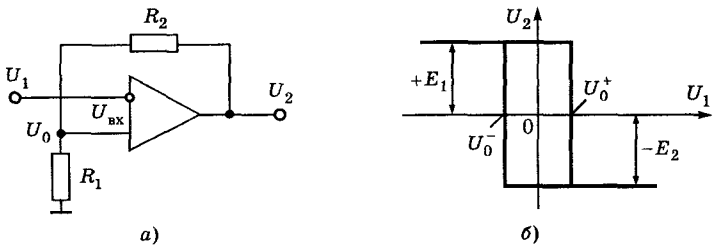
$$\Delta = U^*/K_0.$$

Например, если $K_0 = 10^5$, то $\Delta \approx 7$ мкВ. Значит, *выходное напряжение фиксируется на уровнях* $\pm U^*$ *при крайне незначительном отклонении напряжения* U_1 от $-E_{оп}$. На графике, где масштаб напряжения U_1 составляет вольты или десятые доли вольта, интервал Δ изобразить нельзя и характеристика компаратора будет иметь ступенчатый характер (рис. 10.32, *б*). На этом и основано использование компараторов: они служат для сравнения меняющихся напряжений с эталонным напряжением $E_{оп}$ и фиксации момента их равенства. Фиксация равенства характеризуется резким изменением полярности выходного напряжения (рис. 10.32, *в*). В частном случае, когда $E_{оп} = 0$, компаратор называют нуль-индикатором.

Компаратор является простейшим элементом, выполняющим *аналого-цифровое преобразование*. Он сравнивает два напряжения (входные аналоговые величины) и выдает на выходе сигнал, подчиняющийся законам двоичной логики.

Если вместо одиночных диодов $Д1$ и $Д2$ использовать последовательные диодные цепочки, то выходное напряжение компаратора будет соответственно больше. Однако оно не может превышать предельных значений $+E_1$ и $-E_2$, о которых сказано выше. Если вообще не включать диодов в цепь обратной связи, то выходные уровни компаратора будут равны $+E_1$ и $-E_2$, а чувствительность составит $+E_1/K_0$ и $-E_2/K_0$.При этом значения U_1, отличающиеся от $-E_{оп}$ на величину, большую чувствительности, соответствуют нелинейному режиму ОУ.

На рис. 10.33, *а* показано пороговое устройство, которое работает аналогично триггеру Шмитта (см. раздел 8.9).



а) *б)*

Рис. 10.33. Пороговое устройство: *а* — схема; *б* — передаточная характеристика

Обозначим коэффициент передачи цепи обратной связи:

$$\gamma = U_0/U_2 = R_1/(R_1 + R_2).$$

Предположим, что в начальном состоянии все потенциалы в схеме (U_1, U_0 и U_2) равны нулю. Можно убедиться, что это состояние неустойчиво: малейшая флюктуация вызывает лавинообразный процесс, в результате которого выходное напряжение приобретает одно из двух предельных значений: $+E_1$ или $-E_2$. Делая управляющее напряжение U_1 отрицательным, мы еще больше увеличим отрицательное значение $U_{вх}$ и поэтому не изменим выходного напряжения: оно останется равным E_1. При положительных значениях U_1, меньших U_0^+, выходное напряжение тоже не изменится, так как разность $U_1 - U_0^+$ останется отрицательной. И только при условии $U_1 \approx U_0^+$ входное напря-

жение станет близким к нулю, а выходное напряжение начнет уменьшаться. Соответственно начнет уменьшаться потенциал $U_0 = \gamma U_{вых}$, а вместе с ним и входное напряжение. Получается лавинообразный процесс, в результате которого выходное напряжение примет второе устойчивое значение $-E_2$. При этом потенциал U_0 станет равным $U_0^- = -\gamma E_2 < 0$. Поскольку управляющее напряжение U_1 за время лавинообразного процесса не изменилось и осталось равным U_0^+, входное напряжение запишется следующим образом:

$$U_{вх} = U_2 - U_0^- = \gamma(E_1 + E_2).$$

Дальнейшее увеличение U_1 не влияет на величину U_2.

Если теперь уменьшать величину U_1 до нуля и в сторону отрицательных значений, то выходное напряжение не изменится до тех пор, пока потенциал U_1 не сделается почти равным U_0^-. После этого опять произойдет лавинообразный процесс и выходное напряжение вернется к значению E_1. Функция $U_2(U_1)$ показана на рис. 10.33, *б*. В отличие от триггера Шмитта передаточная характеристика данной схемы двуполярная (а при равенстве $E_1 = E_2$ еще и симметричная относительно оси U_2).

Пороговое устройство служит основой для построения широкого класса импульсных схем: генераторов прямоугольных и пилообразных колебаний, формирователей импульсов и т. п.

В заключение рассмотрим варианты схемы ОУ с обратной связью (рис. 10.27). Если сделать $R_2 = R_1$, то получится схема *инвертирующего повторителя* с коэффициентом передачи $K = -1$. Входное сопротивление такого повторителя равно R_1.

Если заменить сопротивление R_2 конденсатором C, то операторный коэффициент передачи согласно (10.25) будет иметь вид

$$K(\mathbf{s}) = -1/(\mathbf{s}CR_1).$$

Как известно, $1/\mathbf{s}$ есть оператор интегрирования. Следовательно, ОУ с такой цепью обратной связи является *интегрирующим усилителем*, у которого связь между выходным и входным напряжениями имеет вид:

$$u_{вых}(t) \sim \int u_{вх}(t)\, dt.$$

Если заменить сопротивление R_1 конденсатором C, то согласно (10.25)

$$K(\mathrm{s}) = -\mathrm{s}CR_2.$$

Оператор s есть оператор дифференцирования, т. е. в данном случае ОУ превращается в *дифференцирующий усилитель*, у которого

$$u_{\text{вых}}(t) \sim (\mathrm{d}/\mathrm{d}t)\, u_{\text{вх}}(t).$$

Аналогичным путем можно строить схемы, выполняющие другие операции.

10.11. Надежность интегральных схем

В гл. 1 мы уже подчеркивали, что повышенная надежность, свойственная ИС, — один из главных факторов, обеспечивших развитие микроэлектроники как особого научно-технического направления.

Надежность — не интуитивное понятие: ей свойственны количественные параметры, которые в совокупности характеризуют качество ИС. В основе понятия надежности лежит безотказность выполнения заданных функций в оговоренных режимах эксплуатации и в течение заданного интервала времени.

Отказом ИС (или других приборов) считается либо полное нарушение их работоспособности, либо только уход некоторых параметров за допустимые, заранее заданные нормы. Соответственно различают *полные* (катастрофические) и *условные* (постепенные) *отказы*. Примером полного отказа может служить короткое замыкание коллектора с базой, а примером условного отказа — уменьшение коэффициента β в 3 раза.

Причины отказов ИС. В основе полных отказов лежат короткие замыкания и обрывы в тех или иных частях ИС, а в основе условных отказов — медленные изменения электропроводности и других электрофизических свойств отдельных участков ИС.

Если короткие замыкания или обрывы имеют место уже до контроля, то эти дефекты проявляются в виде брака во время контроля. Если же замыкания и обрывы во время контроля находятся в стадии «зарождения», то контроль их не обнаружит и они будут потенциальными факторами ненадежности. Перечислим типичные процессы, приводящие к коротким замыканиям и обрывам.

Короткое замыкание возникает либо в результате соприкосновения соединительных проводников между собой или с корпусом под действием механических вибраций или ударов, либо в результате локального перегрева и проплавления перегретого участка, либо в результате проникновения проводящего вещества через поры в диэлектрике. Последний случай часто имеет место в тонком окисле МОП-транзистора, если в корпус ИС проникает влага. Что касается локального перегрева, то он характерен лишь для мощных транзисторов, а в ИС и тем более БИС практически не наблюдается.

Обрывы могут возникать как под действием механических сил (вибраций, ударов), так и в результате электрохимических и химических процессов. В первом случае, как правило, нарушаются электрические связи соединительных проволочек с контактными площадками на кристалле или со штырьками корпуса. Что касается электрохимических и химических процессов, то они проявляются в нескольких вариантах.

Во-первых, имеет место *электрохимическая коррозия* металлических пленок и контактных соединений. Например, при наличии следов влаги и соляной кислоты алюминиевая металлизация превращается в окисел алюминия (Al_2O_3), причем соляная кислота не расходуется, а служит лишь материалом для промежуточных реакций.

Во-вторых, образуются *интерметаллические соединения*. Это явление особенно характерно для контактов разнородных металлов, например, золотой проволочки с алюминиевой контактной площадкой. В свое время такие контакты послужили причиной массовых отказов ИС благодаря образованию так называемой «пурпурной чумы» — непроводящего порошкообразного соединения на границе Al и Au, имеющего красный цвет. Предотвращение такого соединения требует соблюдения строгого температурного режима при термокомпрессии.

В-третьих, имеет место *электромиграция* — перемещение атомов металла (алюминия) в прилегающие области под действием электрического поля и повышенной температуры. В результате толщина металлической дорожки уменьшается и обрыв получается как следствие локального перегрева и «перегорания» дорожки.

Теперь рассмотрим процессы, лежащие в основе постепенных (условных) отказов, т. е. в сущности — временно́го дрейфа

параметров ИС. Конечно, между процессами, вызывающими полные и условные отказы, нет четкой границы. Можно сказать, что *полный отказ есть результат лавинообразного накопления тех изменений, которые до этого проявлялись в виде дрейфа параметров.*

Главную роль в появлении условных отказов играют процессы на границе кремния с защитным окислом. Это — образование инверсионных и обогащенных слоев — каналов — под действием ионов, расположенных в окисле (см. раздел 3.4). Такие каналы оказывают непосредственное влияние на обратные токи p–n-переходов и величину пробивного напряжения. Нестабильность обоих этих параметров обусловлена миграцией ионов в окисле. В свою очередь, миграция ионов обусловлена их диффузией (особенно при повышенной температуре) и дрейфом в электрических полях.

Электрические поля в окисле неизбежны, поскольку он граничит с полупроводниковыми слоями и металлическими дорожками, которые находятся под разными потенциалами. И толщина окисла и ширина p–n-переходов (т. е. расстояния между нейтральными p- и n-слоями) составляют доли микрона; поэтому напряженность поля даже при разности потенциалов 0,5–1 В достигает 10^4 В/см и более.

Направление поля может быть как продольным (параллельным поверхности раздела), так и поперечным (перпендикулярным поверхности). Следовательно, миграция ионов в окисле происходит в обоих направлениях, так что меняется не только электропроводность, но и протяженность каналов.

Приповерхностные явления оказывают также влияние на величину коэффициентов усиления B и β. Действительно, согласно (5.26) коэффициент B зависит от величин $L_б$ и $L_э$. В свою очередь, эти величины зависят от эффективных времен жизни носителей в слоях базы и эмиттера. Эффективное время жизни складывается из объемной и поверхностной составляющих. Поскольку миграция ионов в окисле меняет структуру приповерхностного слоя, а значит, и скорость поверхностной рекомбинации, то коэффициенты усиления базового тока подвержены временно́му дрейфу.

Методы оценки надежности. До последнего времени основным методом оценки надежности ИС был статистический метод, в основе которого лежат испытания партий приборов на срок службы. Если в партии из N шт. за время t произошло n

отказов, то вероятность отказа в единицу времени оценивают как

$$\lambda\ [1/\text{ч}] = n/(Nt). \qquad (10.29)$$

Величину λ называют *средней частотой* или *интенсивностью отказов*.

Зная величину λ, можно оценить вероятность безотказной работы ИС в течение заданного времени эксплуатации:

$$P = e^{-\lambda t}. \qquad (10.30)$$

Из выражения (10.30) следует: какой бы малой ни была величина λ, с течением времени вероятность безотказной работы приближается к нулю.

Средним временем безотказной работы или *средним сроком службы* прибора принято считать величину, получаемую из условия $\lambda t = 1$:

$$t_{\text{ср}} = 1/\lambda. \qquad (10.31)$$

Например, если $\lambda = 10^{-5}$ 1/ч, то $t_{\text{ср}} = 10^5$ ч (т. е. около 10 лет).

Вообще говоря, величина λ — не постоянная: она меняется с течением времени (рис. 10.34). На кривой $\lambda(t)$ различают три характерных участка: участок I, на котором выявляются грубые ошибки при сборке, загрязнения поверхности и т. п.; участок II, на котором λ = const, т. е. отказы обусловлены случайными, неконтролируемыми причинами; участок III, на котором λ снова возрастает в результате неизбежного старения прибора, т. е. проявления тех химических и физико-химических процессов, от которых не избавлена ни одна реальная структура и которые связаны с принципом действия прибора. В случае ИС такими принципиальными факторами являются взаимная диффузия разнородных материалов, радиационные дефекты, обусловленные космическим излучением, и т. п.

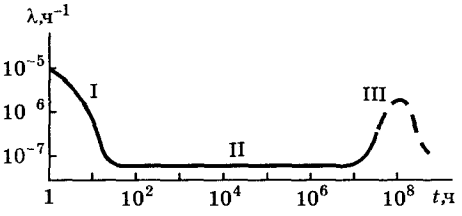


Рис. 10.34. Зависимость интенсивности отказов от времени

Средний срок службы (10.31) соответствует границе между участками II и III. Участок I обычно устраняется путем тренировки приборов у производителя до «выпуска в свет», т. е. до поставки потребителю. Тренировка состоит в том, что после узаконенных испытаний — механических, электрических и климатических (на температуру, влажность и т. п.) — приборы работают в течение нескольких десятков или сотен часов в нормальных эксплуатационных условиях, и отказавшие за это время приборы устраняются.

В настоящее время интенсивность отказов ИС и БИС лежит в пределах 10^{-8}–10^{-9} 1/ч. Для достоверной оценки величины λ нужно при испытаниях «дождаться» хотя бы 2–3 отказов. Тогда из выражения (10.29) при $n = 2$–3 следует, что время испытаний для партии $N = 10^3$ шт. составит десятки лет. Ставить же на испытания партии ИС в количестве 10^4–10^5 шт. экономически невыгодно.

В таких случаях используется *метод ускоренных испытаний*, основанных на *законе Аррениуса*. Этот закон гласит, что скорость химических и физико-химических процессов связана с температурой экспоненциальной зависимостью:

$$v \sim e^{-W_a/kT},$$

где W_a — энергия активации процесса. Отсюда следует, что средний срок службы при повышенной температуре будет существенно меньше, чем при нормальной:

$$t_y = t_н \exp[-(W_a/k)(T_н^{-1} - T_y^{-1})], \qquad (10.32)$$

где индексы «н» и «у» относятся к нормальной и повышенной абсолютным температурам.

Проводя ускоренные испытания при повышенной температуре, форсируют отказы прибора, добиваясь их появления за разумное время. Полученное значение λ_y пересчитывают к нормальной температуре с помощью соотношений (10.31) и (10.32). Используя, например, для испытаний ИС температуру +250 °С, можно ускорить оценку величины λ в сотни раз. Однако при значениях $\lambda \le 10^{-9}$ 1/ч и такое ускорение оказывается недостаточным. Отсюда следует, что *на этапе современной микроэлектроники обычные статистические методы оценки надежности оказываются неприемлемыми.*

Поэтому за последние 5–10 лет большое внимание уделялось разработке *физических методов* оценки и прогнозирования надежности. Под такими методами понимаются индивидуальные исследования структуры готовых ИС с целью выявления дефектов, чреватых отказом, а также исследования отказавших ИС с целью выяснения причин отказа и внесения соответствующих усовершенствований в технологию их производства.

В отличие от статистических методов, которые относятся к категории *разрушающих* (поскольку в их основе лежит отказ прибора), физические методы являются *неразрушающими*, а часто и бесконтакными. К их числу относятся тепловидение (обследование в инфракрасных лучах), рентгеноскопия, электронная микроскопия, а также измерение избыточных шумов, которые характеризуют качество контактов. Все эти методы связаны с использованием сложного, дорогостоящего оборудования. Поэтому их еще нельзя считать установившимися. Однако учитывая неприемлемость статистических методов, они, по-видимому, займут со временем ведущее место при оценке надежности БИС.

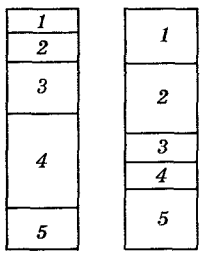


Рис. 10.35. Соотношение количества отказов, вызванных разными причинами, в простых ИС (*а*) и больших ИС (*б*):
1 — дефекты металлизации; 2 — погрешности диффузии; 3 — дефекты кристалла и окисла; 4 — неправильное применение; 5 — прочие причины

Интенсивность отказов снижается с повышением степени интеграции, поскольку производству БИС свойствен более высокий технологический уровень. Одновременно меняется роль различных факторов отказов (рис. 10.35): дефекты металлизации и погрешности диффузии, которые у простых ИС занимали незначительное место, у БИС выступают на первый план. Наоборот, ошибки, связанные с неправильным применением, в случае БИС отступают на второй план (поскольку резко уменьшается количество внешних соединений).

Говоря о статистическом методе оценки надежности, мы подразумевали, что результаты испытаний конкретной партии приборов в виде формулы (10.29) действительны и для других аналогичных партий. Это справедливо только в том случае, когда другие партии изготавливают-

ся по той же технологии, что и испытанная партия. Отсюда следует важный вывод: *высокая надежность ИС обеспечивается в первую очередь стабильностью технологического цикла.* Любое (даже прогрессивное) изменение технологического цикла может вызвать (хотя бы временное) снижение надежности ИС.

В заключение подчеркнем условность понятия «постепенный» отказ. В зависимости от типа аппаратуры изменение β на 40% может быть либо неприемлемым, либо приемлемым, т. е. может быть или не быть «отказом». Таким образом ИС, считающиеся отказавшими (т. е. непригодными) в одной аппаратуре, могут использоваться в другой еще длительное время.

10.12. Заключение

Из всего, что мы узнали в данной главе, можно сделать следующие общие выводы относительно ближайших перспектив развития микроэлектроники.

Повышение степени интеграции, характерное для последних лет, не является самоцелью. С одной стороны, оно обеспечивает получение новых электронных приборов (БИС и СБИС), функции которых соответствуют бывшим узлам в дискретной электронной аппаратуре. С другой стороны, рост потенциальных возможностей новых электронных приборов обуславливает создание принципиально новых устройств, принципов обработки информации и управления, использование которых в рамках традиционных технологий было нецелесообразно.

Опыт разработки радиоэлектронной аппаратуры показывает, что чисто электронные узлы занимают в ней по количеству и значимости выполняемых функций сравнительно небольшое место: 30–40%. Остальные 60–70% относятся к электрическим, электромеханическим и механическим узлам (световые индикаторы, гальванические батареи, трансформаторы, реле, электродвигатели, электромагниты, разъемы, переключатели и т. п.). Прогресс микроэлектроники явно опережает прогресс в смежных областях техники. На нынешнем этапе становится остро необходимым «подтягивание тылов», т. е. распространение методов и средств микроэлектроники на смежные области с целью резкого улучшения габаритно-массовых, стоимостных и надежностных показателей неэлектронных узлов. Такое на-

правление, получившее название *комплексной микроминиатюризации*, уже находится в стадии становления: появились оптоэлектронные бесконтактные реле, интегральные (пленочные) магнитозаписывающие головки и др.

В области самих электронных узлов чисто количественное наращивание степени интеграции не может быть беспредельным. Качественно новым решением, которое, в частности, обеспечит повышенную надежность — «живучесть», может быть отказ от традиционных схемных элементов (в первую очередь — транзисторов) и использование объемных эффектов в твердом теле.

Классическим примером прибора, основанного на таких эффектах, является кварцевый резонатор, который, будучи однородной структурой, выполняет функции колебательного контура, состоящего из элементов L, C и R.

Научно-техническое направление, связанное с отказом от компонентной структуры ИС и основанное на использовании объемных эффектов, получило заранее условное название *функциональной электроники*.

Учитывая все сказанное, можно сделать вывод, что современная микроэлектроника, основы которой мы рассмотрели, вероятно, является некоторым законченным этапом, после которого рано или поздно наступит новый этап развития электроники, обладающий своими качественными особенностями. Однако этот новый этап, несомненно, впитает в себя те физические, технологические и схемотехнические принципы, которые свойственны современной микроэлектронике.

Контрольные вопросы

1. Перечислите и опишите свойства простейших логических операций, выполняемых логическими элементами.
2. Опишите принцип действия элементов резисторно-транзисторной логики. Какую элементарную функцию выполняет элемент РТЛ?
3. Как реализуются логические функции И-НЕ и ИЛИ-НЕ в транзисторно-транзисторной логике?
4. Чем определяются входные токи ТТЛ-элемента при высоком и низком логических уровнях на входе?

5. Опишите работу логического элемента ТТЛ со сложным инвертором. Перечислите типы и особенности современных разновидностей ТТЛ-элементов.
6. Опишите принцип действия и характеристики $И^2Л$-элементов. Как реализуются логические функции на таких элементах?
7. Как реализуются логические функции в схемах на переключателях тока (ЭСЛ)? Почему в схемах на ЭСЛ-элементах обеспечивается высокое быстродействие?
8. Опишите принцип действия логических элементов на МОП-транзисторных ключах с резистивной и динамической нагрузкой.
9. Как реализуются простейшие логические функции в КМОП-логических элементах? Как строятся логические элементы на КМОП-ключах коммутации?
10. Опишите принцип действия БиКМОП-логического элемента. Какие параметры удается улучшить при использовании этих элементов по сравнению с КМОП-логическими элементами?
11. Перечислите справочные параметры, характеризующие интегральные логические элементы. Что такое работа переключения?
12. Опишите основные разновидности интегральных триггеров. Приведите пример реализации записи информации в триггер.
13. Опишите основные разновидности (по выполняемым функциям) запоминающих устройств и элементную базу, на основе которой они могут быть реализованы.
14. Какую структуру имеют оперативные запоминающие устройства с произвольной выборкой? Опишите реализацию запоминающих ячеек на различных типах логических элементов.
15. Какие физические принципы, схемотехнические и технологические приемы используются для записи информации в интегральные микросхемы постоянных запоминающих устройств?
16. Как классифицируются интегральные микросхемы по степени интеграции? Что такое ИС на базовых матричных кристаллах и программируемые логические матрицы?

17. Какие физические и конструктивно-технологические ограничения препятствуют повышению степени интеграции? Как могут быть преодолены некоторые из этих ограничений?
18. Какие функции выполняет идеальный операционный усилитель? Приведите схему и характеристики инвертирующего усилителя на базе ОУ.
19. Опишите основные статические и динамические параметры реальных интегральных операционных усилителей. Как обеспечиваются точностные параметры ОУ?
20. В чем причины отказов интегральных микросхем? Проведите классификацию методов испытаний, используемых для отбраковки и определения показателей надежности ИС.

Заключение

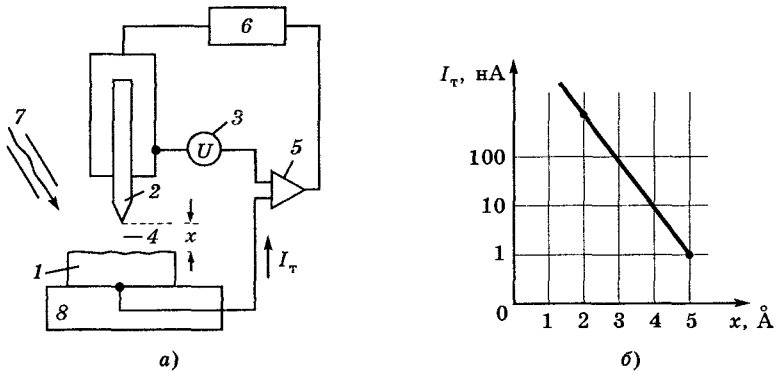
Перспективы развития микроэлектроники

В течение всех лет победоносного развития микроэлектроники постоянно велись и ведутся поиски создания альтернативной элементной базы. Многие ученые предсказывали, что на смену полупроводниковой микроэлектронике придет функциональная электроника, одноэлектроника, оптоэлектроника, фотоника, квантовая и, наконец, биоэлектроника. Во всех перечисленных направлениях к настоящему времени достигнуты обнадеживающие результаты. Однако ни по одному из этих направлений не создано технологической базы, обеспечивающей экономически конкурентное производство высоконадежной элементной базы.

Выше на рис. 6.14 приведен прогноз совершенствования технологических методов создания планарно-эпитаксиальных структур. Дальнейшее уменьшение размеров методами электронной и рентгеновской литографии ограничивается прежде всего явлениями существенного разупорядочения материалов за пределами окон в фоторезисте. Поэтому 2010–2015 годы можно считать вершиной развития микроэлектроники.

В 1985 г. американские кристаллографы Дж. Карл и Г. А. Хауптман стали лауреатами Нобелевской премии за выдающиеся достижения в разработке прямых методов определения структуры кристаллов. С этого времени началась история бурного развития исследований, создания лабораторного и промышленного приборостроения и нанометровых структур, основанных на использовании туннельной микроскопии.

На рис. З.1, *а* приведена схема разработанной российскими специалистами М. А. Ананяном и П. Н. Лускиновичем нанотехнологической установки. В качестве подложки могут быть использованы любые проводящие материалы с тщательно отполированной поверхностью. Зонд представляет собой металлическую иглу, как правило, из твердосплавного материала, с заточенной методами ионного травления вершиной. С микроскопической точки зрения радиус кривизны вершины зонда определяется размерами единичного атома, находящегося на вершине зонда.



а) *б)*

Рис. 3.1. Нанотехнологическая установка: *а* — схема нанотехнологической установки на основе туннельного микроскопа; *б* — зависимость величины туннельного тока от зазора.

1 — подложка, *2* — зонд, *3* — источник питания, *4* — зазор между зондом и подложкой, *5* — усилитель туннельного тока, *6* — динамический регулятор зазора на основе пьезоманипуляторов, *7* — приспособление для напуска газообразных и жидких реактивов, *8* — система прецизионного позиционирования подложки

Если к зонду по отношению к подложке приложить некоторое напряжение, то при уменьшении величины зазора x до размеров порядка единиц ангстрем, через зазор начинает протекать туннельный ток (рис. 3.1, *б*). Важно отметить, что величина зазора значительно меньше величин межатомных и межмолекулярных расстояний в окружающем зазор газе (20–80 Å). Поэтому можно считать, что туннельный ток в зазоре практически протекает в вакууме. При этом напряженность электрического поля в зазоре, даже при слабых управляющих напряжениях порядка милливольт, достигает весьма значительных величин порядка 10^6 В/см и выше.

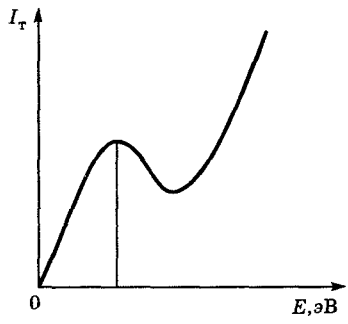
Как видно из рис. 3.1, *б*, ток в зазоре при стабилизированном управляющем напряжении линейно зависит от величины зазора. При изменении величины зазора на 1 Å величина тока изменяется в 10 раз. Измеряя туннельный ток, можно с помощью пьезопреобразователей регулировать или стабилизировать величину зазора с точностью не ниже 0,1 Å. При указанных величинах электрических полей диаметр пучка туннельных электронов, протекающих в вакууме между зазором и подложкой, составляет порядка 1,0–1,5 Å.

В описываемой нанотехнологической установке предусмотрена возможность откачки и напуска в активный объем необходимых жидких или газообразных реактивов. Естественно, что вся конструкция технологической камеры изготовлена из коррозионно-стойких материалов. Это обстоятельство существенно отличает нанотехнологическую установку от туннельного микроскопа. Отметим также, что во избежание влияния внешних сейсмических и акустических воздействий, вся установка снабжена системой пассивной, а в ряде случаев и активной, виброзащиты.

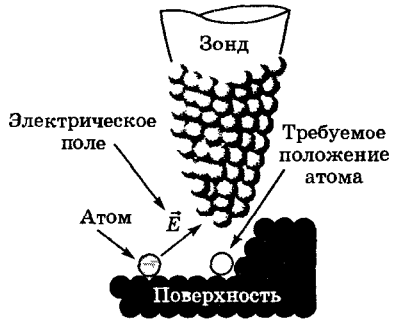
С помощью линейных пьезоманипуляторов подложка может перемещаться относительно острия зонда в пределах 10×10 мм^2 с точностью не менее 0,1 Å.

На рис. 3.2, *а* показана типичная вольтамперная характеристика, снятая для некоторого образца при постоянной величине зазора. При энергии электронов, меньшей энергии тепловых колебаний атомов материала подложки (порядка 25 мэВ), можно исследовать атомарную структуру поверхности подложки, не разрушая ее. При энергиях, равных или несколько больших энергии межатомных связей атомов поверхности подложки, на вольтамперной характеристике появляются различные нелинейности, позволяющие снять туннельную спектрограмму подложки и определить ее химический состав. При энергии пучка, равной энергии межатомных связей, можно «возбудить» отдельный атом, находящийся на поверхности, «оторвать» его от нее и «перенести», перемещая подложку, в некоторое новое положение. Снижая энергию возбуждения, можно «пришить» этот перемещенный атом к поверхности в новом положении (рис. 3.2, *б*).

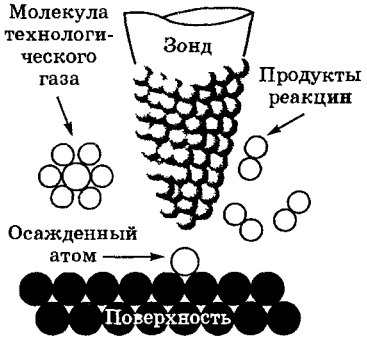
Если в активную область установки ввести молекулы технологического газа (рис. 3.2, *в*), то под действием сильного электрического поля эти молекулы прежде всего ионизируются и далее на поверхность подложки можно осадить необходимый атом, выбранный таким образом, чтобы он образовал с атомами подложки прочно соединенный радикал. Наращивая осажденные атомы и перемещая подложку, можно вырастить на ней прочно закрепленные дорожки проводников или отдельные группы атомов с поперечными размерами атомарной величины (порядка 20 Å). Такие проводники и группы атомов можно назвать квантовыми проводниками и квантовыми точками.



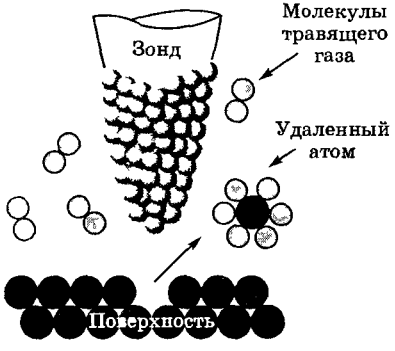
а)



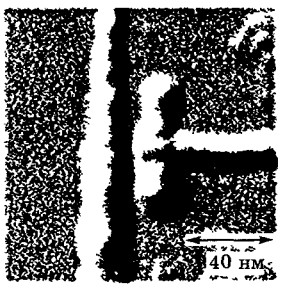
б)



в)



г)



д)

Рис. 3.2. Основные нанотехнологические операции:
а — зависимость туннельного тока от свойств материала подложки и энергии электронов, б — фиксация и перемещение атомов; *в* — осаждение атомов из окружающего зонд газа; *г* — травление подложки, *д* — пример нанотехнологической структуры — полевой транзистор

Напуская в технологический объем газы-травители (рис. 3.2, *г*), можно обеспечить активацию химических реакций «захвата» и удаления с поверхности некоторых атомов, создавая «канавки» нанометровых размеров.

На рис. 3.2, *д* в качестве примера реализации некоторых нанотехнологических операций приведена туннельно-микроскопическая фотография полевого транзистора. Если на управляющем электроде (затворе), расположенном на фотографии справа, отсутствует заряд, то по левому проводнику ток может проходить беспрепятственно — транзистор открыт. Если на затвор подать запирающее напряжение, то поле перекрывает канал, и транзистор оказывается закрытым.

Чрезвычайно важно отметить, что при поперечных размерах квантовых проводников порядка 20 Å в них за счет поперечного квантования электронов значительно уменьшается рассеяние энергии и, следовательно, резко повышается быстродействие. При размерах полевого транзистора, приведенного на рис. 3.2, *д*, его быстродействие лежит в терагерцовом диапазоне.

Отметим еще одну принципиальную особенность отечественной нанотехнологической установки. С ее помощью можно наращивать на подложке не только продольные квантовые проводники, но и последовательно формировать трехмерные элементы. Это открывает практически неограниченные возможности разрешения проблемы «тирании» проводников. На основе трехмерных связей могут быть реализованы не только апробированные в микроэлектронике элементы, но и весьма экзотические нейристорные структуры.

Магистральным путем решения проблемы повышения производительности однозондовых нанотехнологических установок является создание многозондовых машин. По оценкам специалистов уже к 2005 году удастся разработать установки, обеспечивающие сборку атомов со скоростью один кубический дециметр вещества в час при стоимости не выше одного доллара.

Формирование электрических сигналов с временем фронта 10^{-14} с и распространение их по двухпроводным нанопроводникам, являющимся, по существу, металло-оптическими волноводами, обеспечивает реальную интеграцию в единой среде всей гаммы электронных и оптоэлектронных схем.

Высокая степень интеграции наноэлектронных структур, быстродействие, трехмерная сборка элементов и уменьшенное энергорассеяние закладывают фундамент для приоритетного развития на их основе быстродействующих устройств обработ-

ки информации. В частности, уже в ближайшие годы могут быть промышленно реализованы элементы памяти со сверхвысокой плотностью (10^{12} бит/см^2) записи информации, что в тысячи раз превосходит плотность записи на традиционных лазерных дисках.

Учитывая резкий рост публикаций по нанотехнологиям, структурам и приборам и широчайший характер практических направлений исследований, можно с уверенностью сказать, что одним из ближайшим продолжений развития микроэлектроники является *наноэлектроника*.

Нанотехнология обеспечивает не только успехи в развитии элементной базы информационного приборостроения. Уже в настоящее время нанотехнологические разработки используются в медицине, робототехнике, машиностроении, атомной энергетике, оборонных системах и многих других областях. Не случайно в подавляющем большинстве развитых стран огромное внимание уделяется поддержке национальных программ по нанотехнологиям. Начало XXI века будет характеризоваться бурным развитием нанотехнологий вообще и *наноэлектроники* в частности.

Литература

1. Степаненко И.П. Основы микроэлектроники: Учебное пособие для вузов. М.: Сов. Радио, 1980.
2. Росадо Л. Физическая электроника и микроэлектроника. М.: Высшая школа, 1991.
3. Зи С. Физика полупроводниковых приборов. Т.1,2. М.: Мир, 1984.
4. Grove A.S. Physics and Technology of Semiconductor Devices. John Wiley & Sons, New York, 1967.