Цифровые Интегральные Схемы

МЕТОДОЛОГИЯ ПРОЕКТИРОВАНИЯ

titula-Cifrovie integral-shemi.i1 1 25.06.2007 14:20:03

Digital Integrated Circuits

A DESIGN PERSPECTIVE

SECOND EDITION

JAN M. RABAEY
ANANTHA CHANDRAKASAN
BORIVOJE NICOLIC

Prentice Hall Electronic and VLSI series
Charles G. Sodini, series editor

Pearson Education, Inc.
Upper Saddle River, New Jersey 07458

titula-Cifrovie integral-shemi.i2 2 25.06.2007 14:20:04

Цифровые Интегральные Схемы

МЕТОДОЛОГИЯ ПРОЕКТИРОВАНИЯ

ЖАН М. РАБАИ
АНАНТА ЧАНДРАКАСАН
БОРИВОЖ НИКОЛИЧ

Серия книг Prentice Hall по электронике и СБИС
Редактор серии Чарльз Г. Содини

ВИЛЬЯМС

Москва • Санкт-Петербург • Киев
2007

titula-Cifrovie integral-shemi.i3 3 25.06.2007 14:20:04

ББК (Ж/О)32.844.1
Р12
УДК 621.382.2/.3

Издательский дом "Вильямс"

Зав. редакцией *С.Н. Тригуб*

Перевод с английского и редакция *А.В. Назаренко*

По общим вопросам обращайтесь в Издательский дом "Вильямс" по адресу:
info@williamspublishing.com, http://www.williamspublishing.com
115419, Москва, а/я 783; 03150, Киев, а/я 152

Рабаи, Жан М., **Чандракасан**, Ананта, **Николич**, Боривож.
Р12 Цифровые интегральные схемы, 2-е издание.: Пер. с англ. — М. : ООО "И.Д. Вильямс", 2007. — 912 с. : ил. — Парал. тит. англ.

ISBN 978-5-8459-1116-2 (рус.)

Книга будет чрезвычайно полезна разработчикам цифровых интегральных схем и инженерам, желающим лучше разобраться в структуре СБИС. В ней предлагается множество полезных примеров, которые вы можете использовать в своей работе, и учтены основные тенденции, существующие в современном мире разработки КМОП-устройств. Книгу можно использовать как учебник по микроэлектронике при подготовке студентов соответствующих специальностей или как справочник по интегральным схемам, который удобно всегда иметь под рукой. Она хорошо структурирована и достаточно понятна, хотя базовые знания о предметной области и используемом в ней математическом аппарате все же желательны.

ББК (Ж/О)32.844.1

Все названия программных продуктов являются зарегистрированными торговыми марками соответствующих фирм.

Никакая часть настоящего издания ни в каких целях не может быть воспроизведена в какой бы то ни было форме и какими бы то ни было средствами, будь то электронные или механические, включая фотокопирование и запись на магнитный носитель, если на это нет письменного разрешения издательства Prentice Hall, Inc.

Authorized translation from the English language edition published by Prentice Hall, Inc., Copyright © 2003, 1996 by Pearson Education, Inc.

All rights reserved. No part of this book shall be reproduced, stored in a retrieval system, or transmitted by any means-electronic, mechanical, photocopying, recording, or otherwise-without written permission from the publisher, except for the inclusion of brief quotations in a review.

Russian language edition published by Williams Publishing House according to the Agreement with R&I Enterprises International, Copyright © 2007

ISBN 978-5-8459-1116-2 (рус.)
ISBN 0-13-090996-3 (англ.)

© Издательский дом "Вильямс", 2007
© by Pearson Education, Inc., 2003, 1996

Оглавление

Предисловие 17

I. Структуры 27

1. Введение 29

2. Технология производства 67

А. Топология ИС 103

3. Приборы 111

Б. Моделирование схемы 181

4. Проводник 187

II. Перспектива схемы 233

5. КМОП-инвертор 235

6. Разработка комбинационных логических элементов на КМОП 299

В. Моделирование сложных логических схем 385

Г. Техники разработки топологии для сложных логических элементов 395

7. Разработка последовательных логических схем 401

III. Перспектива системы 459

8. Стратегии реализации цифровых интегральных схем 461

Д. Описание характеристик логических и последовательных ячеек 519

Е. Синтез проекта 527

9. Вопросы межсоединений 537

10. Синхронизация в цифровых схемах 593

Ж. Верификация проекта 669

11. Проектирование арифметических блоков 675

12. Проектирование запоминающих устройств и матричных структур 751

3. Приемочные испытания произведенных схем 867

IV. Приложения 889

А. Ответы на задачи 891

Б. Справочная информация 901

Предметный указатель 905

Содержание

Предисловие 17
Что в этой книге нового 17
Сохранение духа первого издания 18
Как пользоваться этой книгой 18
Общедоступный Web-ресурс 20
Уникальные особенности данной книги 20
Краткий обзор содержания 21
Слова признательности 24
От издательства 26
I. Структуры 27

1. Введение 29
1.1. Взгляд в историю 30
1.2. Главное в проектировании цифровых интегральных схем 33
1.3. Меры качества разработки цифровой интегральной схемы 44
1.3.1. Стоимость 44
1.3.2. Функциональные возможности и помехоустойчивость 48
1.3.3. Быстродействие 58
1.3.4. Потребляемая мощность 61
1.4. Резюме 63
1.5. Для любознательных 63

2. Технология производства 67
2.1. Введение 68
2.2. Изготовление КМОП ИС 68
2.2.1. Кремниевая пластина 69
2.2.2. Фотолитография 70
2.2.3. Некоторые повторяющиеся шаги технологического процесса 74
2.2.4. Схема упрощенного КМОП-процесса 76
2.3. Нормы проектирования: между разработчиком и технологом 80
2.4. Корпусирование интегральных схем 84
2.4.1. Материалы корпусов 87
2.4.2. Уровни межсоединений 87
2.4.3. Учет тепловыделения при корпусировании 94

2.5. Тенденции в технологии интегральных схем 96
2.5.1. Краткосрочная перспектива 97
2.5.2. Долгосрочная перспектива 99
2.6. Резюме 100
2.7. Для любознательных 101

А. Топология ИС 103
А.1. Редактор топологии 104
А.2. Символьная топология 105
А.3. Проверка соблюдения норм проектирования 106
А.4. Восстановление принципиальной электрической схемы 109
А.5. Для любознательных 110

3. Приборы 111
3.1. Введение 112
3.2. Диод 113
3.2.1. Беглое описание принципа работы: обедненная область 113
3.2.2. Статическое поведение 116
3.2.3. Динамическое поведение или переходные процессы 119
3.2.4. Реальный диод. Вторичные эффекты 123
3.2.5. SPICE-модель диода 125
3.3. Полевой МОП-транзистор 127
3.3.1. Краткое описание принципов работы прибора 127
3.3.2. МОП-транзистор в статических условиях 128
3.3.3. Реальный МОП-транзистор: некоторые вторичные эффекты 159
3.3.4. SPICE-модели МОП-транзисторов 163
3.4. Несколько слов об отклонениях в технологическом процессе 166
3.5. Взгляд в перспективу: масштабирование технологии 169
3.6. Резюме 176
3.7. Для любознательных 177

Б. Моделирование схемы 181
Б.1. Свойства схемного моделирования 182
Б.2. Режимы моделирования 183
Б.3. Развертка постоянной составляющей 183
Б.4. Анализ переходных процессов 183
Б.5. Модели устройств 184
Б.6. Для любознательных 185
Б.7. Литература 185

4. Проводник 187
4.1. Введение 188
4.2. На первый взгляд 188
4.3. Параметры соединения — емкость, сопротивление и индуктивность 191
4.3.1. Емкость 191
4.3.2. Сопротивление 198
4.3.3. Индуктивность 202
4.4. Модели электрических проводников 204
4.4.1. Идеальный проводник 205
4.4.2. Модель с сосредоточенными параметрами 205
4.4.3. Сосредоточенная RC-модель 207
4.4.4. Распределенная RC-цепь 211
4.4.5. Линия передачи 215
4.5. SPICE-модели проводов 226
4.5.1. Распределенные *RC*-линии в SPICE 226
4.5.2. SPICE-модели линии передачи 228
4.5.3. Перспектива: взгляд в будущее 228
4.6. Резюме 230
4.7. Для любознательных 231
II. Перспектива схемы 233
5. КМОП-инвертор 235
5.1. Введение 236
5.2. Статический КМОП-инвертор — интуитивная перспектива 236
5.3. Расчет устойчивости КМОП-инвертора (статическое поведение) 241
5.3.1. Порог переключения 241
5.3.2. Запас помехоустойчивости 245
5.3.3. Еще раз об устойчивости 249
5.4. Быстродействие КМОП-инвертора: динамическое поведение 252
5.4.1. Расчет емкости 252
5.4.2. Задержка распространения: анализ в первом приближении 258
5.4.3. Задержка распространения с точки зрения разработки 263
5.5. Мощность, энергия и задержка энергии 275
5.5.1. Динамическое потребление мощности 275
5.5.2. Статическое потребление 286
5.5.3. Собираем все вместе 289
5.5.4. Анализ потребления мощности с помощью SPICE 292
5.6. Перспектива: масштабирование технологии и ее влияние на метрики инвертора 294
5.7. Резюме 297
5.8. Для любознательных 298

6. Разработка комбинационных логических элементов на КМОП 299
6.1. Введение 300
6.2. Разработки статических КМОП-устройств 301
6.2.1. КМОП ИС 301
6.2.2. Стандартизованная логика 332
6.2.3. Логика на проходных транзисторах 338
6.3. Разработка динамических КМОП-элементов 356
6.3.1. Динамическая логика: основные принципы 356
6.3.2. Скорость и рассеяние мощности в динамических логических элементах 359
6.3.3. Вопросы сохранности сигнала в динамических схемах 363
6.3.4. Каскадное соединение динамических элементов 370
6.4. Перспективы 378
6.4.1. Как выбрать стиль разработки логики 378
6.4.2. Разработка логики со сниженным потреблением напряжения 379
6.5. Резюме 382
6.6. Для любознательных 383

В. Моделирование сложных логических схем 385
В.1. Представление цифровых данных в виде непрерывной сущности 386
В.2. Представление данных в виде дискретной сущности 386
В.3. Использование моделей более высокого уровня 391

Г. Техники разработки топологии для сложных логических элементов 395
Г.1. Разработка топологии согласно Уэйнбергеру и концепции стандартных ячеек 395
Г.2. Планирование топологии с использованием путей Эйлера 396

7. Разработка последовательных логических схем 401
7.1. Введение 402
7.1.1. Временны*е* метрики последовательных схем 403
7.1.2. Классификация элементов памяти 405
7.2. Статические защелки и регистры 407
7.2.1. Принцип бистабильности 407
7.2.2. Защелки, основанные на мультиплексоре 409
7.2.3. Регистр “master–slave”, управляемый фронтом 411
7.2.4. Низковольтные статические защелки 418
7.2.5. Статические SR-триггеры — запись данных методом грубой силы 420
7.3. Динамические защелки и регистры 423
7.3.1. Динамические регистры на проходных логических элементах, управляемые фронтом 424
7.3.2. Тактируемая КМОП-схема — подход, нечувствительный к расфазировке 426
7.3.3. Действительно однофазный тактируемый регистр 430

7.4. Альтернативные стили регистров 435
7.4.1. Импульсные регистры 435
7.4.2. Регистры, основанные на усилителях считывания 438
7.5. Конвейер: подход к оптимизации последовательных цепей 440
7.5.1. Конвейер на защелках или на регистрах? 442
7.5.2. NORA–КМОП — стиль логики конвейерных структур 443
7.6. Небистабильные последовательные схемы 446
7.6.1. Триггер Шмитта 446
7.6.2. Последовательные схемы с одним устойчивым состоянием 449
7.6.3. Автоколебательные схемы 451
7.7. Перспектива: выбор стратегии тактирования 453
7.8. Резюме 455
7.9. Для любознательных 456

III. Перспектива системы 459

8. Стратегии реализации цифровых интегральных схем 461
8.1. Введение 462
8.2. От индивидуального подхода к проектированию к полуиндивидуальному и использованию структурированных массивов 467
8.3. Индивидуальная разработка 468
8.4. Методология проектирования на основе элементов 469
8.4.1. Стандартный элемент 470
8.4.2. Скомпилированные элементы 476
8.4.3. Макроэлементы, мегаэлементы и интеллектуальная собственность 478
8.4.4. Полуиндивидуальный процесс проектирования 483
8.5. Подходы к реализации, основанные на матрицах 486
8.5.1. Матрицы с предварительной диффузией (или программируемым фотошаблоном) 487
8.5.2. Матрицы с предварительным монтажом 493
8.6. Перспектива — платформа реализации будущего 512
8.7. Резюме 514
8.8. Для любознательных 515

Д. Описание характеристик логических и последовательных ячеек 519
Д.1. Важность и сложность описания характеристик библиотеки 519
Д.2. Описание характеристик логических ячеек 520
Д.3. Описание характеристик регистров 523
Д.4. Список использованной литературы 526

E. Синтез проекта 527
E.1. Синтез схем 528
E.2. Синтез логики 529
E.3. Синтез архитектуры 532
E.4. Для любознательных 536

9. Вопросы межсоединений 537
9.1. Вступление 538
9.2. Паразитная емкость 538
9.2.1. Емкость и надежность — перекрестные помехи 538
9.2.2. Емкость и быстродействие в КМОП 542
9.3. Паразитное сопротивление 555
9.3.1. Сопротивление и надежность — омическое падение напряжения 555
9.3.2. Электромиграция 558
9.3.3. Сопротивление и быстродействие — *RC*-задержка 560
9.4. Паразитная индуктивность* 566
9.4.1. Индуктивность и надежность — скачок напряжения на $L\frac{di}{dt}$ 566
9.4.2. Индуктивность и быстродействие — эффекты линии передачи 573
9.5. Улучшенные техники межсоединений 577
9.5.1. Цепи с уменьшенным размахом 579
9.5.2. Техники передачи на переключателях тока 585
9.6. Перспектива: сети на кристалле 587
9.7. Резюме 588
9.8. Для любознательных 589

10. Синхронизация в цифровых схемах 593
10.1. Введение 594
10.2. Классификация цифровых систем с позиции синхронизации 595
10.2.1. Синхронное межсоединение 595
10.2.2. Мезохронное межсоединение 596
10.2.3. Плезиохронное межсоединение 596
10.2.4. Асинхронное межсоединение 597
10.3. Синхронное проектирование — глубокий анализ перспектив 598
10.3.1. Основы синхронного тактирования 598
10.3.2. Источники расфазировки и дрожания 606
10.3.3. Технологии разводки тактовых сигналов 614
10.3.4. Тактирование на основе защелок* 623
10.4. Проектирование схем с внутренней синхронизацией* 627
10.4.1. Логические схемы с внутренней синхронизацией — асинхронный подход 627
10.4.2. Генерация сигнала завершения 631
10.4.3. Передача сигналов с самосинхронизацией 635
10.4.4. Практические примеры логики с внутренней синхронизацией 641

10.5. Схемы синхронизации и арбитры* 644
10.5.1. Схемы синхронизации — концепция и реализация 644
10.5.2. Арбитры 650
10.6. Синтез тактовых сигналов и синхронизация с использованием ФАПЧ* 650
10.6.1. Суть 652
10.6.2. Основные составляющие схем ФАПЧ 654
10.7. Перспективы 659
10.7.1. Распределенное тактирование и использование систем автоподстройки по задержке 659
10.7.2. Оптическое распределение тактового сигнала 661
10.7.3. Синхронное или асинхронное проектирование 663
10.8. Резюме 663
10.9. Для любознательных 664

Ж. Верификация проекта 669

11. Проектирование арифметических блоков 675
11.1. Введение 676
11.2. Информационные каналы в архитектурах цифровых процессоров 677
11.3. Сумматор 678
11.3.1. Двоичный сумматор: определения 678
11.3.2. Полный сумматор: вопросы проектирования схем 682
11.3.3. Двоичный сумматор: проектирование логики 689
11.4. Умножитель 707
11.4.1. Умножитель: определения 707
11.4.2. Генерация частичных произведений 708
11.4.3. Накопление частичных произведений 710
11.4.4. Конечное суммирование 715
11.4.5. Резюме 716
11.5. Сдвиговый регистр 716
11.5.1. Многорегистровое циклическое сдвиговое устройство 717
11.5.2. Логарифмический сдвиговый регистр 719
11.6. Другие арифметические операторы 720
11.7. Компромиссы между мощностью и скоростью в информационных каналах* 722
11.7.1. Технологии уменьшения мощности (время проектирования) 725
11.7.2. Управление мощностью во время выполнения 737
11.7.3. Уменьшение мощности в холостом (или спящем) режиме 743
11.8. Перспектива: проектирование как поиск компромиссов 745
11.9. Резюме 746
11.10. Для любознательных 747

12. Проектирование запоминающих устройств и матричных структур 751

12.1. Введение 752
12.1.1. Классификация запоминающих устройств 753
12.1.2. Архитектуры и основные составляющие запоминающих устройств 757

12.2. Сердечник запоминающего устройства 764
12.2.1. Постоянные запоминающие устройства 764
12.2.2. Энергонезависимые оперативные запоминающие устройства 780
12.2.3. Оперативные запоминающие устройства (ОЗУ) 792
12.2.4. Ассоциативные запоминающие устройства 807

12.3. Периферийные схемы запоминающих устройств* 810
12.3.1. Декодеры адреса 810
12.3.2. Усилители считывания 818
12.3.3. Опорное напряжение 826
12.3.4. Задающие устройства/буферы 831
12.3.5. Синхронизация и контроль 831

12.4. Надежность и выход готовых запоминающих устройств* 834
12.4.1. Отношение сигнал/шум 835
12.4.2. Выход годных запоминающих устройств 841

12.5. Рассеяние мощности в запоминающих устройствах* 844
12.5.1. Источники рассеяния мощности в запоминающих устройствах 844
12.5.2. Секционирование памяти 846
12.5.3. Решение проблемы рассеяния активной мощности 846
12.5.4. Рассеяние хранимой информации 848
12.5.5. Резюме 851

12.6. Разбор примеров запоминающих устройств 851
12.6.1. Программируемая логическая матрица (ПЛМ) 851
12.6.2. Статическое ОЗУ емкостью 4 Мбит 854
12.6.3. Флэш-память емкостью 1 Гбит на основе элементов НЕ-И 855

12.7. Перспективы: тенденции и эволюция в сфере полупроводниковых ЗУ 859

12.8. Резюме 861

12.9. Для любознательных 863

3. Приемочные испытания произведенных схем 867

Введение 867

Процедура тестирования 869

Проектирование с возможностью тестирования 870

Для любознательных 886

IV. Приложения 889

А. Ответы на задачи 891

Б. Справочная информация 901

Значения некоторых физических констант и коэффициентов материалов 901

Формулы и уравнения 901

Модели КМОП-устройств 903

Комбинационная КМОП-логика 903

Модели межсоединений 904

Предметный указатель 905

Предисловие

Что в этой книге нового

Пере вами второе издание книги *Цифровые интегральные схемы. Методология проектирования*. За шесть лет, прошедших с момента выхода в свет первого издания, цифровые интегральные схемы существенно эволюционировали и видоизменились. Технология изготовления ИС продолжила свой путь в область все более малых размеров. Со времени выхода первого издания размер минимального топологического элемента уменьшился почти в десять раз и теперь приближается к диапазону 100 нм. Подобное уменьшение оказывает двойное воздействие на проектирование цифровых интегральных схем. Во-первых, резко возросла сложность систем, которые теперь могут быть размещены в одном кристалле микросхемы. Решение связанных с этим проблем привело к появлению новых стратегий проектирования и реализации таких проектов. В то же время переход в область субмикронных размеров изменяет поведение приборов и выдвигает на передний план ряд новых аспектов, которые сказываются на надежности, стоимости, быстродействии и величине рассеиваемой мощности цифровых ИС. Именно углубленное рассмотрение этих вопросов и отличает данное издание от первого.

Беглого взгляда на оглавление достаточно, чтобы заметить, что в книге содержится подробное описание поведения субмикронных приборов, методов оптимизации схем, моделирования и оптимизации межсоединений, сохранения целостности сигналов, синхронизации и организации временных диаграмм, кроме того, значительное внимание уделено рассеиваемой мощности. Все названные темы иллюстрируются примерами современных конструктивных решений. Кроме того, поскольку на рынке цифровых ИС более 99% — это МОП-схемы, из книги были исключены разделы, посвященные описанию технологий кремниевых биполярных схем и схем на основе GaAs (однако интересующийся читатель может найти старые главы с описанием этих технологий на Web-сайте данной книги). Учитывая важность методологии в современном процессе проектирования, в состав книги были введены специальные разделы-вставки, в каждом из которых подчеркивается тот или иной конкретный аспект процесса проектирования. Это новое издание существенно переработано, и связано это прежде всего с появлением двух соавторов Ананты Чандракасана (Anantha Chandrakasan) и Боривожа Николича (Borivoje Nicolić), которые внесли новый более широкий взгляд на процесс проектирования цифровых ИС, а также на последние тенденции и проблемы в этой области.

Сохранение духа первого издания

Вводя эти изменения, авторы тем не менее хотели сохранить дух и цели первого издания, а именно: создание мостика между чисто **схемотехническим и системным подходами** к проектированию цифровых устройств. Начинается книга с описания принципов работы электронных приборов и углубленного анализа главного элемента всех цифровых устройств: инвертора. Затем полученные знания распространяются на вопросы проектирования более сложных модулей: вентилей, регистров, контроллеров, сумматоров, устройств умножения и запоминающих устройств. В ходе изложения поднимаются все те вопросы, которые с неизбежностью встают перед разработчиками современных сложных микросхем: какие параметры разработки определяющие, на какую часть проекта следует обратить особое внимание, а что можно проигнорировать? Понятно, что естественной реакцией на возрастающую сложность цифровых систем является упрощение. Однако чрезмерное упрощение может привести к потере схемой работоспособности, поскольку было опущено влияние таких общесистемных эффектов, как временные требования, система межсоединений и потребляемая мощность. Чтобы не угодить в эту западню, важно выполнять проектирование с учетом как схемотехнических, так и общесистемных аспектов. Именно такой подход и взят за основу в данной книге, которая предоставит читателю знания и умение, необходимые в борьбе со сложностью современных систем и основанные на использовании как аналитического, так и экспериментального методов.

Как пользоваться этой книгой

В основном материал, изложенный в данной книге, предназначен для изучения на старших курсах в группах по специальности “Проектирование цифровых устройств”. В дополнение к этому ядру в книгу включены главы и разделы, посвященные передовым направлениям в данной области. В процессе разработки структуры данной книги быстро стало очевидным, что в такой обширной теме, как проектирование цифровых устройств, очень сложно определить подмножество, которое бы удовлетворяло потребностям читателей всех категорий. С одной стороны, новичкам в данной области требуется подробное описание базовых концепций. С другой стороны, отзывы читателей и рецензентов первого издания показали, что крайне желательно и даже необходимо, чтобы в книгу были включены материалы, дающие углубленное и расширенное описание передовых направлений и современного положения дел. Следование подобному видению привело к появлению книги, содержание которой выходит за рамки односеместрового курса. Посвященный описанию передовых направлений материал может быть использован в качестве основы для изучения предмета на последнем курсе обучения. Широкий охват тем и включение в состав книги описания современных подходов делает ее полезным справочным пособием для профессиональных инженеров. Кроме того, предполагается, что студенты, приступающие к изучению данного курса, уже знакомы с основами проектирования логических схем.

Организация материала такова, что если придерживаться определенного порядка следования, то его можно проходить и читать в разной последовательности. Ядром книги являются главы 5–8. Главы 1–4 можно рассматривать в качестве вводных. В ответ на часто упоминаемое требование в главе 2 дан краткий курс технологии изготовления полупроводниковых приборов. Что касается главы 3, то студенты, уже знакомые с основами физики полупроводниковых приборов, могут ее просто быстро пролистать. Авторы настоятельно рекомендуют это сделать, так как именно в ней вводится ряд важных обозначений и базовых положений. Кроме того, в ней же приводится описание подхода к первичному моделированию субмикронных транзисторов, допускающему выполнение анализа вручную. Чтобы подчеркнуть важность системы межсоединений при проектировании современных цифровых устройств, описание подходов к моделированию межсоединений было сдвинуто в материале книги вперед, в главу 4.

Главы 9–12 являются наиболее передовыми по своему содержанию и могут быть использованы для придания курсу определенной направленности. Например, курс с акцентом на схемотехнические аспекты может включать в себя базовые главы, дополненные материалом из глав 9 и 12. Курс с упором на проектирование цифровых систем должен включать дополнение в виде частей глав 9–11. Все эти главы с материалом, посвященным передовым подходам, могут составить основу курса для групп последнего года обучения или преддипломного уровня. В тексте книги главы с передовым материалом отмечены звездочкой.

Ниже приведено несколько возможных способов организации курса для групп старшего курса. В *методических материалах для преподавателей*, размещенных на Web-сайте книги, приведено несколько полных учебных планов, основанных на курсах, которые читают в некоторых институтах.

Базовый курс схемотехники (при минимальном предварительном знакомстве с физикой полупроводниковых приборов): 1, 2.1–2.3, 3–8, (9.1–9.3, 12).

Несколько более усложненный курс схемотехники: 1 (2, 3), 4–9, 10.1–10.3, 10.5-10.6, 12.

Курс проектирования систем: 1 (2, 3), 4–9, 10.1–10.4, 11, 12.1-12.2.

Вставки приводятся в составе тех глав, к которым они относятся.

Для обеспечения последовательного изучения материала в каждой из глав сначала указывается перечень излагаемых в ней тем, за которым следует их подробное и углубленное обсуждение. В разделе “Перспективы” обсуждается связь изложенных концепций с реальными разработками и возможные пути их развития в будущем. Каждая глава заканчивается разделом “Резюме” с кратким перечислением тем, раскрытых в тексте, за которым следуют разделы “Для любознательных” и “Список литературы”. В этих разделах приводится пространный список ссылок и указателей для тех читателей, которые хотят получить более подробную информацию по тем или иным вопросам, затронутым в главе.

Как и подразумевает название этой книги, одной из ее целей является выделение этапа проектирования в процессе создания цифровых интегральных схем. Чтобы добиться наглядности и дать возможность увидеть реальные перспективы, в книге были введены примеры практически воплощенных конструктивных решений и топологий. Их анализ помогает получить ответы на такие вопросы, как: "Насколько в действительности позволяет сберечь площади или мощности или не потерять в быстродействии использование конкретной методики?" В целях имитации реального процесса проектирования в книге также широко используются результаты, получаемые с помощью инструментальных средств разработки, например, с помощью средств моделирования на схемном и ключевом уровнях, а также с помощью систем генерации и редактирования топологии. Очень часто там, где необходимо проверить результаты расчетов, выполненных вручную, проиллюстрировать новую концепцию или исследовать сложное поведение, не поддающееся анализу вручную, используется компьютерный анализ.

Наконец, для облегчения процесса изучения в книге содержится большое количество примеров. Кроме того, в каждой главе дается несколько *заданий или вопросов на сообразительность* (ответы на них можно найти в конце книги), которые заставляют думать и разбираться в процессе чтения.

Общедоступный Web-ресурс

На Web-сайте `http://bwrc.eecs.berkeley.edu/IcBook/index.htm` выложены задания по разработке с решениями, а также полный набор слайдов, содержащих наиболее важные рисунки и графики, приведенные в книге.

Кроме всего прочего, на Web-сайте также можно найти примеры лабораторных работ по программированию и схемотехнике, дополнительную информацию и массу полезных ссылок на другие Web-сайты.

Уникальные особенности данной книги

- Объединяет схемотехнический и системный взгляды на задачу проектирования. Обеспечивает глубокое понимание проблемы разработки сложных цифровых интегральных схем, одновременно готовя разработчика к решению новых сложных задач, которые, возможно, уже ждут его.
- Основная направленность книги — проектирование. В ней представлены основные проблемы, возникающие при проектировании, и приводятся методические указания по их решению. Представленные методики иллюстрируются примерами реальных конструктивных решений и результатами полного анализа с использованием пакета SPICE.
- Это первая книга по проектированию интегральных схем, посвященная исключительно приборам с субмикронными размерами. Для облегчения понимания связанных с этим особенностей была разработана простая модель транзистора, получившая название *унифицированной модели МОП-транзистора* и позволяющая проводить анализ вручную.
- Уникальна в показе самых современных методик проектирования сложных быстродействующих интегральных схем и схем с малым энергопотреб-

лением. Быстродействие и мощность рассматриваются как равноправные составляющие.

- Рассматриваются такие критически важные вопросы проектирования систем, как целостность сигналов, рассеиваемая мощность, система межсоединений, корпусирование, организация временной диаграммы работы и синхронизация.
- Приводится уникальное описание самых современных методологий проектирования и инструментальных средств. Одновременно обсуждаются вопросы их применения с точки зрения разработчика.
- Дается анализ возможных путей развития технологии интегральных схем в будущем.
- Прекрасные иллюстрации и полезные проектно-ориентированные вставки.
- Разделы "Для любознательных" и "Список литературы" дают пространный список ссылок и указателей для тех читателей, которые хотят получить более детальную информацию по тем или иным вопросам, затронутым в книге.
- На Web-сайте авторов доступен широкий пакет методических материалов для преподавателей, в состав которого входят программные средства проектирования, слайдовые заготовки, сборники расчетных заданий и заданий по проектированию, примеры топологии реальных схем, а также лабораторные работы по схемотехнике и программированию.

Краткий обзор содержания

Уже беглого взгляда на оглавление достаточно, чтобы понять, как порядок следования глав и раскрытия материала соответствует декларируемой авторами методологии проектирования. Книга начинается с описания моделей полупроводниковых приборов и затем осуществляется постепенный переход на более высокие уровни абстракций: инвертор, сложные логические вентили (И-НЕ, ИЛИ-НЕ, ИСКЛЮЧАЮЩЕЕ ИЛИ), функциональные (сумматоры, устройства умножения, схемы сдвига, регистры) и системные (шины данных, контроллеры, запоминающие устройства) модули. Для каждого из этих уровней идентифицируются главенствующие с точки зрения проектирования параметры и строятся упрощенные модели, не учитывающие малозначащие детали. Хотя такой многоуровневый подход к моделированию и является наилучшим для разработчика подходом, позволяющим хоть как-то справиться со сложностью проектируемой системы, все же он не свободен от некоторых существенных недостатков. Это иллюстрируется в главах 9 и 10, в которых рассматриваются такие связанные с глобальным воздействием на разработку в целом темы, как паразитные элементы в системе межсоединений и организация временной диаграммы работы кристалла. Чтобы еще больше подчеркнуть дихотомичность схемотехнического и системного подходов к проектированию, книга была разделена на две основные части: часть II (главы 4–7), в которой излагается в основном схемотехнический аспект проектирования цифровых интегральных схем, и часть III (главы 8–12), отражающую в боль-

шей степени системный подход в проектировании. Часть I (главы 1–4) содержит описание необходимых основ (метрики качества разработки, процесса изготовления, моделей приборов и межсоединений).

Глава 1 служит в качестве общего *введения*. После исторического обзора возникновения и проектирования цифровых интегральных схем вводятся понятия иерархического проектирования и многоуровневых абстракций. Здесь же приводится описание основных метрик, которые дают возможность получить количественную оценку стоимости, надежности и быстродействия разработки.

Глава 2 содержит краткий и компактно изложенный курс *технологии изготовления МОП-структур*. Четкое представление об основных этапах технологического процесса помогает формировать трехмерный образ МОП-транзистора, что очень важно при идентификации источников паразитных связей прибора. Кроме того, процесс изготовления несет ответственность за многие изменения в параметрах приборов. Далее в главе вводится понятие норм проектирования, которые служат интерфейсом между разработчиком и производителем. Завершает главу обзор методов корпусирования кристаллов ИС — это часто пропускаемый, но критически важный элемент цикла проектирования цифровых ИС.

В **главе 3** содержится краткое описание основных строительных блоков любой разработки: *полупроводниковых приборов*. Основная цель этой главы заключается в том, чтобы дать интуитивное понимание принципов работы МОП-транзистора, а также представить различные модели приборов, широко используемых в последующих главах. Основное внимание уделяется современным субмикронным структурам и их моделированию. Читатели, уже знакомые с физикой полупроводниковых приборов, могут просто быстро пролистать эту главу.

В **главе 4** дается подробный анализ линий связи и сопутствующих паразитных элементов, играющих основную роль. Каждый из паразитных элементов (емкость, сопротивление и индуктивность) рассматриваются по очереди. Также здесь приводятся описания моделей для анализа вручную и с помощью компьютера.

Глава 5 посвящена рассмотрению основы цифровых устройств — *инвертора*. Сначала приводится описание ряда фундаментальных свойств цифровых вентилей. Для наиболее распространенной структуры; статического комплементарного МОП-вентиля, выводятся параметры, помогающие в получении количественной оценки быстродействия и надежности. Описываемые в этой главе методики и подходы критически важны, поскольку они снова и снова повторяются при анализе других, в том числе и более сложных вентильных структур.

В **главе 6** полученные ранее базовые знания переносятся на задачу проектирования простых и сложных цифровых КМОП-вентилей, например структур с функциями ИЛИ-НЕ и И-НЕ. Здесь же показано, что в зависимости от определяющих разработку ограничений (надежность, площадь, быстродействие или мощность) определенную привлекательность могут иметь не только комплементарные статические вентили, но и другие вентильные КМОП-

структуры. Анализируются и сравниваются свойства ряда семейств современных логических вентилей. Здесь же представлены методики оптимизации быстродействия и потребляемой мощности сложных вентилей.

Глава 7 посвящена обсуждению того, как функция памяти может быть реализована с использованием либо положительной обратной связи, либо с помощью хранения заряда. Кроме анализа традиционных бистабильных триггеров в этой главе также приводится описание других последовательных логических схем: астабильных и моностабильных мультивибраторов. Во всех предшествующих главах рассматривались исключительно *комбинационные схемы*, т.е. схемы, не запоминающие историю системы. В противоположность этому *последовательные логические схемы* способны запоминать и сохранять предыдущее состояние.

До **главы 8** представлен схемотехнический подход к проектированию цифровых устройств. Процесс анализа и оптимизации был ограничен уровнем отдельного вентиля. В этой главе предпринимается следующий шаг вперед и анализируется вопрос объединения вентилей в строительные блоки системы. Соответственно системно-ориентированная часть книги начинается с обсуждения *методологий проектирования*. Автоматизация проектирования — это единственный способ справиться с постоянно возрастающей сложностью разрабатываемых систем. В главе 8 обсуждаются наиболее распространенные методы завершения больших разработок за ограниченный период времени. Значительная часть главы посвящена рассмотрению различных методологий реализации разработок, доступных на сегодняшний день проектировщикам. Заказной подход против полузаказного, однократное программирование против неизменяемого фотошаблонного решения, регулярная матричная организация против произвольного размещения — вот лишь некоторые из вопросов, которые поднимаются в данной главе.

Глава 9 возвращается к обсуждению влияния *системы межсоединений* на функциональность и характеристики цифрового вентиля. Каждая связь вносит свои эффекты от паразитных емкости, сопротивления и индуктивности, которые приобретают все возрастающее значение при масштабировании технологических процессов в сторону уменьшения размеров. Здесь приводится описание подходов к решению задачи минимизации влияния этих паразитных элементов системы межсоединений на рабочие характеристики, величину рассеиваемой мощности и надежность интегральной схемы. В этой главе также рассматриваются такие важные вопросы, как организация системы распределения питания и цепи ввода-вывода.

В **главе 10** обсуждается важность наложения строгих временных ограничений на порядок следования событий переключения для обеспечения правильной работы последовательных логических схем. В отсутствие подобных временны*х* ограничений в ячейки памяти могут заноситься неправильные данные. В большинстве цифровых устройств для реализации правильной временно*й* диаграммы работы используется синхронный подход, основанный на тактировании. В этой главе рассматриваются различные методы организации временны*х* диаграмм и тактирования. Здесь же анализируется влияние на поведение синхронных цифровых схем таких важных эффектов, как перекос

фазы синхросигналов. Далее синхронный подход сравнивается с альтернативными методами, к которым относится, например, использование самосинхронизирующихся схем. Завершается глава описанием схем синхронизации и генераторов тактовых импульсов.

Глава 11 посвящена проектированию различных *арифметических блоков*: сумматоров, устройств умножения и сдвиговых регистров. Содержание этой главы очень важно, поскольку показывает, как описанные в главах 5 и 6 методики проектирования переносятся на следующий уровень абстрагирования. Далее вводится понятие критического пути, которое затем широко используется для анализа быстродействия и в процессе оптимизации. Выводятся высокоуровневые модели быстродействия устройств. Эти модели помогают разработчику получать основные данные о работоспособности и качестве проектируемого модуля без необходимости обращаться к углубленному и подробному анализу схемного решения.

Наконец, в **главе 12** рассматриваются различные классы схем памяти и их реализация. В тех случаях, когда требуется хранение больших объемов данных, разработчик цифровых микросхем обращается к специальным схемным модулям, называемым *схемами памяти*. Полупроводниковые запоминающие устройства обеспечивают очень высокую плотность хранения, правда, достигается это за счет ухудшения некоторых фундаментальных свойств цифровых вентилей. Средством получения быстрых и надежных схем памяти является правильная реализация периферийных цепей: дешифраторов, усилителей считывания, выходных формирователей и схем управления, которые также подробно описываются в данной главе. Поскольку главным моментом в разработке схем памяти является обеспечение их бесперебойной работы при всех рабочих условиях, в конце главы приводится подробное обсуждение темы надежности запоминающих устройств. Данную главу, как и предыдущую, можно не включать в учебный план невыпускных курсов.

Слова признательности

Авторы хотели бы поблагодарить всех тех, кто внес вклад в возникновение, создание и корректировку этой рукописи. Прежде всего, спасибо всем студентам-выпускникам, которые годами помогали привести текст к теперешнему его виду. Также спасибо студентам курсов eecs141 и eecs241 университета в Беркли и курса 6.374 Массачусетсского технологического института, которые мучались в группах с экспериментальными программами, основанными на материале данной книги. Огромную помощь в окончательном определении направленности данного нового издания и доводке текста также оказали отзывы студентов, преподавателей и инженеров со всего мира. Непрекращающийся поток писем по электронной почте указывает нам на то, что мы на правильном пути.

В частности, нам хотелось бы поблагодарить Мари-Джейн Ирвин (Mary-Jane Irwin), Виджаи Нараянана (Vijay Narayanan), Эби Фрейдман (Eby Freidman), Фреда Розенбергера (Fred Rosenberger), Вайану Бурлесан (Wayan Burleson), Шехара Боркара (Shekhar Borkar), Иво Болсенса (Ivo Bolsens),

Дуана Бонинга (Duane Boning), Оливера Френцу (Oliver Franza), Лайонела Кимерлинга (Lionel Kimerling), Джози Аммер (Josie Ammer), Майка Шитса (Mike Sheets), Тюфана Каралара (Tufan Karalar), Хуифанга Кина (Huifang Qin), Рета Девиса (Rhett Davis), Натана Чана (Nathan Chan), Джеба Дюранта (Jeb Durant), Андрея Владимиреску (Andrei Vladimirescu), Раду Златановича (Radu Zlatanovici), Ясухиси Шимазаки (Yasuhisa Shimazaki), Фуджио Ишихару (Fudjio Ishihara), Деяна Марковича (Dejan Markovic), Владимира Стояновича (Vladimir Stojanovic), Сеонг Хван Чо (Seong Hwan Cho), Джеймса Као (James Kao), Тревиса Симпкинса (Trevis Simpkins), Сиву Нарендру (Siva Narendra), Джеймса Гудмана (James Goodman), Вадима Гутника (Vadim Gutnik), Теодороса Константакопулоса (Theodoros Konstantakopoulos), Рекса Мина (Rex Min), Викаса Михротру (Vikas Mehrotra) и Пауля-Питера Сотириадиса (Paul-Peter Sotiriadis). Мы очень признательны за их помощь, предложения и отзывы. Конечно же, мы остаемся благодарны и всем тем, кто помог в разработке и написании первого издания.

Мы также чрезвычайно благодарны сотрудникам издательства Prentice Hall. Труд этих людей позволил превратить сырую рукопись в прекрасную книгу. Прежде всего хотелось бы произнести слова благодарности в адрес издателя Тома Роббинса (Tom Robbins), производственного редактора Дениэля Сандина (Danial Sandin) и управляющего редактора Девида Джорджа (Devid George). Отдельные слова благодарности Бренде Ванони (Brenda Vononi) из Беркли за ее неоценимую помощь в редактировании копий и при оформлении содержания Web-сайта. Очень кстати также пришелся опыт в создании Web-сайтов Кэрол Сайти (Carol Sitea).

Хотелось бы подчеркнуть роль компьютерных средств в создании данной книги. Все макеты были полностью подготовлены с использованием издательской системы FrameMaker (Adobe Systems). Графики создавались в основном с помощью пакета MATLAB. Для создания Web-страниц в качестве инструментального средства был выбран пакет Frontpage компании Microsoft. Моделирование схем выполнялось на системе HSPICE (Avant), а все рисунки топологий генерировались с использованием системы физического проектирования Cadence.

Наконец, несколько слов признательности тем людям, которые вынуждены были выносить процесс создания этой книги: Кателийн (Kathelijn), Картияяни (Karthiyayani), Критивасан (Krithivasan) и Ребекка (Rebecca). Хотя процесс написания нового издания менее болезнен, чем первого, мы все время недооцениваем, во что это выливается, особенно в свете наших остальных повседневных обязанностей. И именно они постоянно поддерживали и вдохновляли нас в течение всего времени написания данной книги.

Жан М. Рабаи
Ананта Чандракасан
Боривож Николич
Беркли, Калистога, Кембридж

От издательства

Вы, читатель этой книги, и есть главный ее критик и комментатор. Мы ценим ваше мнение и хотим знать, что было сделано нами правильно, что можно было сделать лучше и что еще вы хотели бы увидеть изданным нами. Нам интересно услышать и любые другие замечания, которые вам хотелось бы высказать в наш адрес.

Мы ждем ваших комментариев и надеемся на них. Вы можете прислать нам бумажное или электронное письмо либо просто посетить наш Web-сервер и оставить свои замечания там. Одним словом, любым удобным для вас способом дайте нам знать, нравится или нет вам эта книга, а также выскажите свое мнение о том, как сделать наши книги более интересными для вас.

Посылая письмо или сообщение, не забудьте указать название книги и ее авторов, а также ваш обратный адрес. Мы внимательно ознакомимся с вашим мнением и обязательно учтем его при отборе и подготовке к изданию последующих книг.

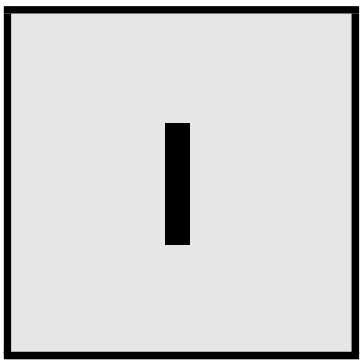
Наши электронные адреса:

E-mail: `info@williamspublishing.com`
WWW: `http://www.williamspublishing.com`

Наши почтовые адреса:

из России: 115419, Москва, а/я 783
из Украины: 03150, Киев, а/я 152

ЧАСТЬ



Структуры

”Уровень сложности при минимальной стоимости компонентов возрастал примерно вдвое за год. Нет никаких сомнений, что в ближайшее время эта скорость сохранится, если не увеличится. В более длительной перспективе скорость роста сложности несколько более неопределенна, хотя нет никаких оснований полагать, что она не останется постоянной по крайней мере еще лет десять. Это означает, что к 1975 году количество компонентов в интегральной схеме с минимальной стоимостью будет составлять 65 000”.

Гордон Мур, *Gramming more Components onto Integrated Circuits*, 1965

ГЛАВА 1

Введение

Эволюция конструктивного исполнения цифровых схем

Главные моменты в проектировании цифровых схем

Как измерять качество разработки

Полезные ссылки

1.1. Взгляд в историю
1.2. Главное в проектировании цифровых интегральных схем
1.3. Меры качества разработки цифровой интегральной схемы
1.3.1. Стоимость
1.3.2. Функциональные возможности и помехоустойчивость
1.3.3. Быстродействие
1.3.4. Потребляемая мощность
1.4. Резюме
1.5. Для любознательных

1.1. Взгляд в историю

Идея цифровой обработки данных оказала огромное влияние на наше общество. Оно достаточно долго привыкало к цифровым компьютерам. Но теперь в результате постоянной эволюции мэйнфреймы и мини-компьютеры превратились в персональные и переносные компьютеры, которые сегодня уже широко используются в повседневной жизни. Более значительной, однако, является наблюдаемая тенденция перехода к цифровым решениям во всех других областях электроники. Одной из первых несвязанных с вычислительной техникой отраслью, где поняли преимущества цифровой обработки данных в сравнении с аналоговой, было приборостроение. Вскоре за ним последовали и системы управления. Совсем недавно мы стали свидетелями перехода на цифровой формат телекоммуникационного оборудования и бытовой электроники. Все шире телефонные данные передаются и обрабатываются в цифровом виде в проводных и беспроводных сетях. Компакт-диски привели к революции в мире звукозаписи, и за ними по пятам уже идет цифровое видео.

Идея создания вычислительных машин, использующих специальным образом закодированные данные, отнюдь не является плодом нашего времени. Еще в начале XIX столетия Бэббидж (Babbage) предвидел возникновение больших механических вычислительных устройств, которые он назвал *разностными машинами* [Swade93]. Хотя в этих машинах использовалась десятичная система счисления, а не двоичное представление, которое общепринято в современной электронике, лежащие в основе идеи были весьма сходными. Разработанная в 1834 году аналитическая машина воспринималась как вычислительное устройство общего применения и имела возможности, удивительно похожие на те, которыми обладают современные компьютеры. Кроме того, что она выполняла некоторый базовый набор операций (сложение, вычитание, умножение и деление) в произвольной последовательности, циклограмма ее работы имела два режима работы, которые назывались хранением и обработкой (исполнением), что также не очень отличается от принципов работы современных компьютеров. Для ускорения выполнения дополнительных операций в ней даже использовался конвейер! К сожалению, сложность и стоимость конструкции сделали саму идею непрактичной. Например, разностная машина I (часть которой показана на рис. 1.1) содержала 25 000 механических деталей и стоила 17 470 фунтов стерлингов (и это в ценах 1834 года!).

Электрическое решение в стоимостном плане оказалось более эффективным. Первые цифровые электронные системы основывались на использовании ключей с магнитным управлением (или реле). Их применение ограничивалось в основном реализацией простых логических цепей. Подобные системы и сегодня еще используются в системах безопасности поездов. Эра цифровых электронных вычислений по-настоящему началась с появления электровакуумных ламп. Хотя первоначально они использовались почти исключительно для аналоговой обработки, очень скоро стало очевидным, что вакуумные лампы также полезны и для выполнения цифровых вычислений. Вскоре были созданы и первые полнофункциональные компьютеры. Кульми-

Рис. 1.1. Работающая часть разностной машины I Бэббиджа (1832), которая считается первым известным автоматическим калькулятором (взято из книги [Swade93] с разрешения Лондонского музея науки)

нацией эры компьютеров на лампах была разработка таких вычислительных машин, как ENIAC (предназначалась для расчетов артиллеристских таблиц стрельбы) и UNIVAC I (первый коммерчески успешный компьютер). Чтобы получить представление о таком понятии, как *степень интеграции*, скажем, что ENIAC был 24,4 метра длиной, 2,6 метра высотой и около 2 метров шириной; содержал он 18 000 ламп. Однако очень быстро стало понятно, что подобная технология достигла своих пределов. Проблемы с надежностью и огромной потребляемой мощностью сделали реализацию более мощных машин экономически и практически нежизнеспособной.

Все изменилось с изобретением в 1947 году в Bell Telephone Laboratories *транзистора* [Bardeen48] и последующим представлением Шокли в 1949 году биполярного транзистора [Schockley49][1]. Но только в 1956 году это привело к появлению первого логического цифрового вентиля, сделанного Харрисом [Harris56] на основе дискретных компонентов. В 1958 году у Джека Килби (Jack Kilby) из Texas Instruments зародилась идея создания *интегральной схемы* (ИС), в которой все компоненты, пассивные и активные, интегрировались бы в одной полупроводниковой подложке, — за это открытие он был удостоен Нобелевской премии. В результате появился первый коммерческий набор интегральных логических вентилей, получивший название *семейства Micrologic компании Fairchild* [Norman60]. Первым же действительно имев-

[1] Захватывающий обзор эволюции цифровых интегральных схем можно найти в книге Мерфи [Murphy93]. (Большая часть данных, которые приведены в данном разделе, была взята именно из этой книги.) В нем также приводятся основополагающие публикации в области цифровых ИС.

шим коммерческий успех семейством логических ИС было семейство ТТЛ-схем (транзистор-транзисторная логика), впервые появившееся на рынке в 1962 году [Beeson62]. Со временем были разработаны и другие семейства с более высокими характеристиками. В качестве примеров можно привести современные ключевые схемы, в которых уже были реализованы первые субнаносекундные вентили. Вершиной биполярных ИС стало появление семейства ЭСЛ-схем (эмиттерно-связанная логика) [Masaki74]. Однако ТТЛ-схемы обладали тем преимуществом, что обеспечивали более высокую плотность интеграции, и именно они были основой первой революции в области интегральных схем. Фактически как раз производство ТТЛ ИС послужило толчком к появлению таких первых больших полупроводниковых компаний, как Fairchild, National и Texas Instruments. Семейство было настолько успешным, что занимало бо́льшую долю рынка полупроводниковых приборов вплоть до 1980-х годов.

Но в конце концов биполярная цифровая логика проиграла сражение за первенство в мире цифровых приборов, и произошло это по тем же причинам, которые привели к уходу со сцены электронных ламп: большая потребляемая мощность в пересчете на один вентиль устанавливает верхний предел на количество вентилей, надежно размещаемых на одном кристалле, в одном корпусе, на плате или в блоке. Хотя и предпринимались попытки создания семейств низкопотребляющих биполярных ИС с высокой плотностью интеграции (например, *И*^{2}*Л-логика — интегральная инжекционная логика* [Hart72]), власть захватил подход, основанный на использовании цифровых МОП ИС.

Базовые принципы конструкции полевого МОП-транзистора (MOSFET) (первоначально имевшего название *полевой транзистор с изолированным затвором*, IGFET) были изложены в патенте Дж. Лилиенфелда (J. Lilienfeld) (Канада) еще в 1925 году и независимо от него англичанином О. Хеллом (O. Hell) в 1935 году. Однако недостаточность знаний о материалах и проблемы с нестабильностью подзатворного диэлектрика надолго задержали начало практического использования этого прибора. Когда к началу 1970-х годов эти проблемы были решены, начался взлет цифровых интегральных МОП-схем. Как ни удивительно, первыми появились логические вентили со структурой КМОП [Wanlass63], и эта тенденция продолжалась до конца 1960-х годов, но сложность процесса изготовления отсрочила широкое использование таких микросхем более чем на два десятилетия. Первые же практические МОП ИС были реализованы на основе p-канальной логики и использовались в основном в калькуляторах. Второй этап революции в области цифровых интегральных схем начался, когда компанией Intel были представлены первые микропроцессоры: 4004 в 1972 году [Faggin72] и 8080 в 1974 [Shima74]. Эти процессоры были реализованы на основе n-канальной логики, которая в сравнении с p-канальной обладала более высоким быстродействием. Одновременно МОП-технология позволила получить первые плотноупакованные микросхемы полупроводниковой памяти. Например, первый кристалл МОП-памяти емкостью 4K появился на рынке в 1970 году [Hoff70].

Все эти события послужили началом поражающей воображение эволюции ИС в сторону все возрастающей плотности интеграции и быстродействия,

революции, которая и сейчас продолжает идти полным ходом. Однако путь к современным уровням интеграции не был гладким. В конце 1970-х годов n-канальная логика начала испытывать трудности по той же причине, которая сделала непривлекательной и нежизнеспособной биполярные схемы с высокой плотностью упаковки, — из-за потребляемой мощности. Осознание этого факта вместе с прогрессом в технологии изготовления наконец склонили чашу весов в пользу КМОП-технологии, и такое положение сохраняется до настоящего момента. Достаточно интересно отметить, что вопрос о потребляемой мощности быстро становится главным и для КМОП ИС, только на этот раз похоже, что под рукой нет технологии, которая бы сняла эту проблему.

Хотя подавляющее большинство современных интегральных схем изготавливаются по МОП-технологии, когда на кону стоит быстродействие, в игру вступают и другие технологии. Примером может послужить БиКМОП-технология, которая объединяет в одном кристалле биполярные и МОП-транзисторы. Если же необходимо добиться еще более высокого быстродействия, то возникают такие технологии, как кремний-германиевая, или даже технология с использованием сверхпроводящих материалов. Но на общей сцене цифровых интегральных схем роль этих технологий весьма незначительна. И, учитывая все продолжающийся рост быстродействия КМОП-схем, эта роль будет продолжать уменьшаться со временем — поэтому в данное книге основное внимание будет уделяться только КМОП ИС.

1.2. Главное в проектировании цифровых интегральных схем

За последние два десятилетия плотность интеграции и быстродействие интегральных схем претерпели поражающие воображение изменения. В 1960 году Гордон Мур, тогда работавший в корпорации Fairchild Corporation и позднее ставший одним из основателей компании Intel, предсказал, что количество транзисторов, размещаемых на одном кристалле микросхемы, будет расти со временем по экспоненте. Это предсказание, позднее названное *законом Мура*, оказалось удивительно прозорливым [Moore65]. Его справедливость лучше всего иллюстрируется рядом графиков. На рис. 1.2 показаны графики зависимости плотности интеграции логических ИС и схем памяти в функции времени. Как можно видеть, величина степени интеграции приблизительно удваивается каждые один или два года. В результате с 1970 года плотность интеграции схем памяти, например, возросла более чем тысячекратно.

Интересное явление представляют собой микропроцессоры. С момента их появления в начале 1970-х годов быстродействие и плотность интеграции микропроцессоров росли с постоянной и предсказуемой скоростью. На рис. 1.3 показаны цифры, соответствующие количеству транзисторов в наиболее значимых разработках. Барьер миллион транзисторов/чип был преодолен в конце 1980-х годов. За последнее десятилетие тактовая частота удваивалась каждые три года и теперь уже достигла гигагерцевого диапазона.

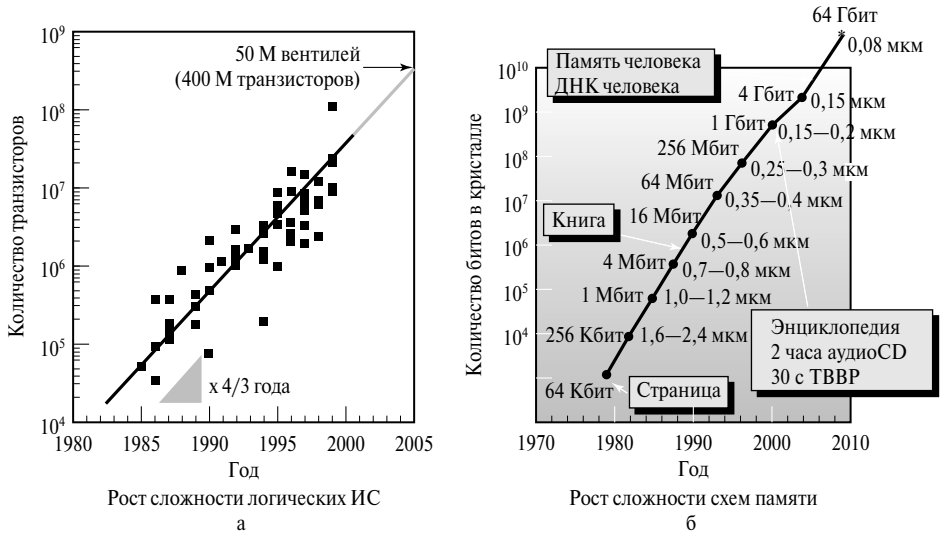


Рис. 1.2. Рост степени интеграции логических ИС и схем памяти в функции времени

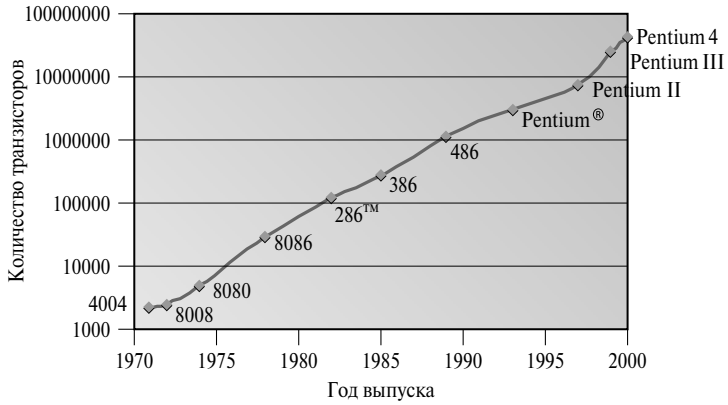


Рис. 1.3. История эволюции количества транзисторов в микропроцессорах (согласно [Intel01])

Этот факт иллюстрируется на рис. 1.4, на котором показана тенденция роста быстродействия микропроцессоров вплоть до начала XXI столетия. Важно то, что по состоянию на данный момент эта тенденция не проявляет никаких признаков замедления.

Читателя не удивит то существенное влияние, которое эти революционные изменения оказали на методику проектирования интегральных схем. Первые разработки делались полностью вручную. Каждый транзистор вырисовывался и оптимизировался отдельно и тщательно вписывался в свою среду. Это хорошо иллюстрирует рис. 1.5, *а*, на котором показан внешний

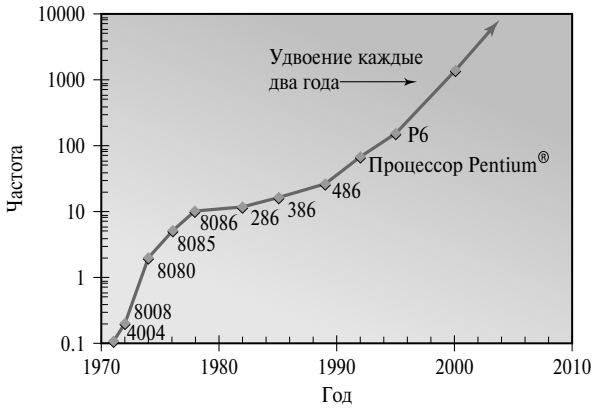
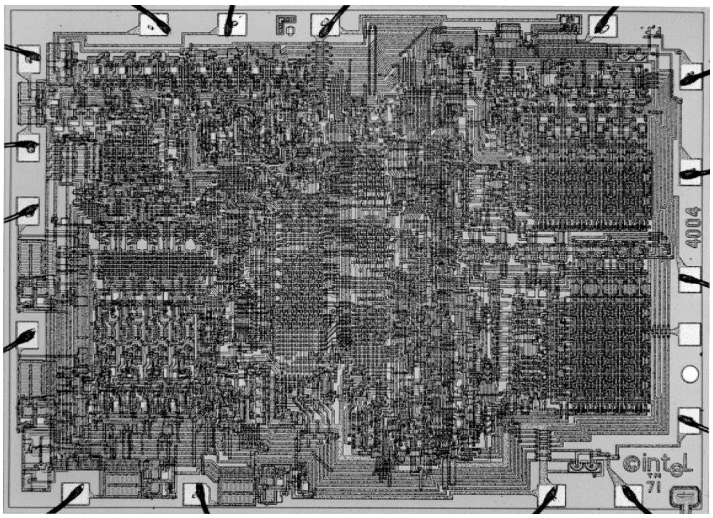


Рис. 1.4. Тенденция роста быстродействия микропроцессоров на начало XXI века (приводится с разрешения компании Intel)

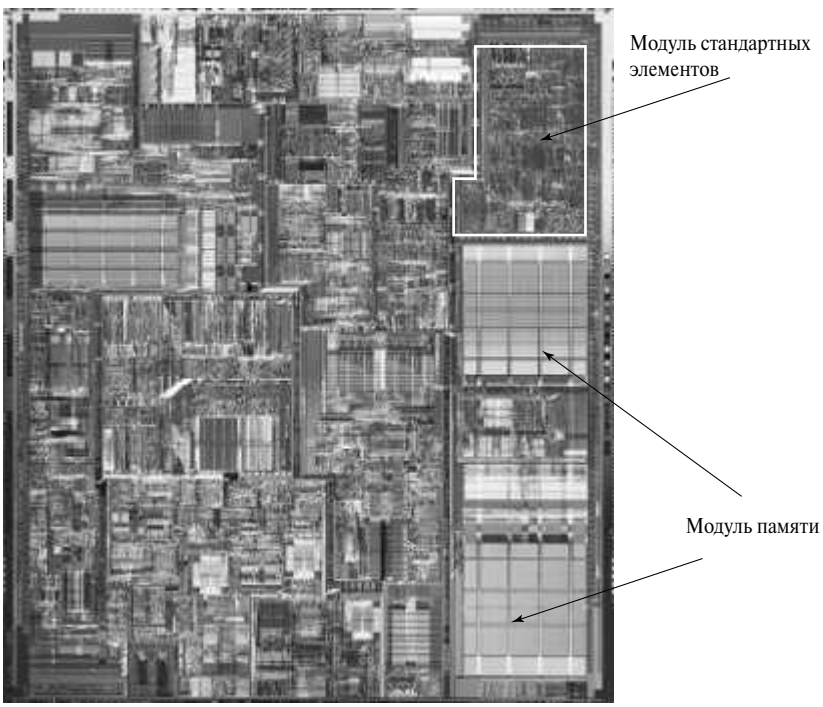
вид кристалла микропроцессора Intel 4004. Совершенно очевидно, что такой подход не годится, когда необходимо создать и объединить в систему более миллиона приборов. В условиях быстрого развития технологии проектирования критически важным моментом в достижении коммерческого успеха изделия становится время до его появления на рынке.

Как результат, разработчики начали все в большей степени придерживаться жестких методологий проектирования и стратегий, которые были более податливы автоматизации процесса проектирования. Влияние такого подхода очень хорошо видно на примере топологии одного из самых последних микропроцессоров компании Intel — Pentium®4, показанной на рис. 1.5, *б*. Вместо использования индивидуализированного подхода ранних разработок теперь схема строится иерархическим образом: процессор представляет собой набор модулей, каждый из которых, в свою очередь, состоит из ячеек. В целях уменьшения трудозатрат на проектирование и увеличения шанса получения работоспособного решения с первого раза ячейки как можно чаще используются повторно. Иерархический подход, оказавшийся единственно возможным, является ключевой составляющей успеха цифровых интегральных схем, и это объясняет тот факт, что большие аналоговые схемы так и не получили развития.

Очевиден следующий вопрос: почему такой подход оказался жизнеспособным в мире цифровых приборов и не прижился (или прижился в меньшей степени) при проектировании аналоговых устройств? В данном случае краеугольной и наиболее важной в плане преодоления проблемы сложности оказалась концепция *абстрагирования*. На каждом уровне конструирования можно абстрагироваться от внутренних деталей строения сложного модуля и заменить их *черным ящиком*, или *моделью*. Такая модель содержит фактически всю информацию, необходимую для проектирования блока, принадлежащего следующему уровню иерархии. Например, после того, как разработчик создал модуль умножителя, характеристики этого модуля могут быть описаны



Микропроцессор 4004
а



Микропроцессор Pentium® 4
б

Рис. 1.5. Сравнение методологий проектирования микропроцессоров Intel 4004 (1971) и Pentium® 4 (2000) (воспроизведено с разрешения Intel)

с высокой точностью и введены в модель. В общем случае характеристики модуля только в весьма незначительной степени зависят от способа его использования в более крупной системе. Поэтому в большинстве случаев он может рассматриваться как черный ящик с известными характеристиками. Поскольку у проектировщика системы отсутствует настоятельная необходимость лезть внутрь такого ящика, сложность проектирования значительно снижается. Влияние подобного подхода "разделяй и властвуй" огромно. Вместо того чтобы иметь дело с мириадами элементов, разработчику достаточно рассматривать всего лишь горстку компонентов, каждый из которых в плане работы и стоимости характеризуется небольшим набором параметров.

Это похоже на работу разработчика программного обеспечения, использующего библиотеку подпрограмм, например драйверов ввода-вывода. Тому, кто пишет большую программу, вовсе не надо лезть внутрь библиотечных подпрограмм. Единственное, что его волнует, так это получение ожидаемого результата после обращения к одному из таких программных модулей. Можно только представить себе, во что бы превратилось написание программ, если бы проходилось извлекать из жесткого диска каждый бит по отдельности и обеспечивать его правильность, а не просто полагаться на удобные операторы "file open" и "get string".

Обычно используемые при проектировании цифровых устройств уровни абстрагирования в порядке возрастания степени абстрагирования выглядят следующим образом: прибор, цепь, вентиль, функциональный модуль (например, сумматор) и система (например, процессор) (рис. 1.6). Полупроводниковый прибор представляет собой элемент с очень сложным поведением. Но ни один схемотехник в процессе разработки вентиля не будет серьезно углубляться в уравнения физики твердого тела, описывающие поведение приборов. Вместо этого он просто воспользуется упрощенной моделью, которая адекватно описывает входные/выходные характеристики транзистора. Например, вентиль И достаточным образом описывается реализуемым им булевым выражением ($Z = A \cdot B$), размерами охватывающего его прямоугольника, положением входных и выходных выводов и величиной задержки выходного сигнала относительно входного.

Подобная философия проектирования сделала возможным появление развитых *систем автоматизированного проектирования (САПР)*, ориентированных на цифровые интегральные схемы, — без них никогда бы не был достигнут тот уровень сложности, который мы имеем на сегодняшний день. Современные системы САПР имеют в своем составе средства для моделирования на различных иерархических уровнях, средства верификации проекта, генерации топологии и синтеза разработки в целом. Обзор существующих инструментальных средств и методологий проектирования приводится в главе 8.

Кроме того, чтобы избежать перепроектирования и повторной верификации часто используемых ячеек, например базовых вентилей и арифметических модулей и модулей памяти, разработчики прибегают к созданию *библиотек ячеек*. Такие библиотеки содержат не только топологии, но также обеспечивают полное документирование и описание характеристик поведе-

ния ячеек. Использование библиотек ячеек отчетливо просматривается, например, в топологии процессора Pentium®4 (см. рис. 1.5, *б*). Блоки целочисленных вычислений и вычислений с плавающей запятой, которые упоминаются лишь для краткости перечисления, содержат большие области, спроектированные с использованием одного из основанных на применении ячеек подходов, называемого методом *стандартных ячеек*. Логические вентили размещаются в строках ячеек равной высоты и соединяются между собой с помощью связей, проходящих в специальных каналах трассировки. При условии наличия библиотеки ячеек топология такого блока может быть сгенерирована автоматически.

Приведенный выше анализ демонстрирует, что автоматизация процесса проектирования и практика модульного проектирования эффективно справляются с некоторыми проблемами, связанными со сложностью разрабатываемых систем и присущими современным разработкам цифровых интегральных схем. Это приводит к неизбежно возникающему вопросу: если автоматизация проектирования решает все проблемы, то зачем тогда вообще переживать по поводу проектирования цифровых интегральных схем? Придется ли разработчику когда-либо волноваться из-за транзисторов и паразитных связей, или наименьшими рассматриваемыми им объектами проектирования будут вентиль и модуль?

Истина состоит в том, что реальность более сложная и существует множество причин, по которым глубокое понимание мельчайших деталей структуры и поведения цифровых схем еще долгое время будет оставаться важным активом разработчика.

- Во-первых, кто-то по-прежнему должен будет *проектировать и реализовывать* библиотеки модулей. Полупроводниковые технологии продолжают с каждым годом совершенствоваться. До тех пор, пока кто-нибудь не разработал надежного метода "портирования" ячейки из одной технологии в другую, каждое изменение в технологии, а это происходит примерно каждые два года, будет требовать перепроектирования библиотеки.
- Создание адекватной *модели* ячейки или модуля требует глубокого понимания того, как они работают внутри. Например, чтобы идентифицировать главные параметры, определяющие быстродействие, сначала необходимо выделить критический путь, задающий временную диаграмму функционирования.
- Библиотечный подход хорошо работает в тех случаях, когда налагаемые на разработку ограничения (быстродействие, стоимость или потребляемая мощность) не слишком жесткие. Тут возможны разнообразные *специфичные для конкретного применения проектные решения*, когда в качестве главной цели выступает получение более интегрированного варианта системы, а требования по быстродействию не выходят за рамки возможностей технологии. К сожалению, для большого количества других продуктов, например процессоров, успех зависит от быстродействия, и поэтому разработчики стремятся выйти на пределы возможностей технологии. В таких обстоятельствах иерархический подход становится несколько

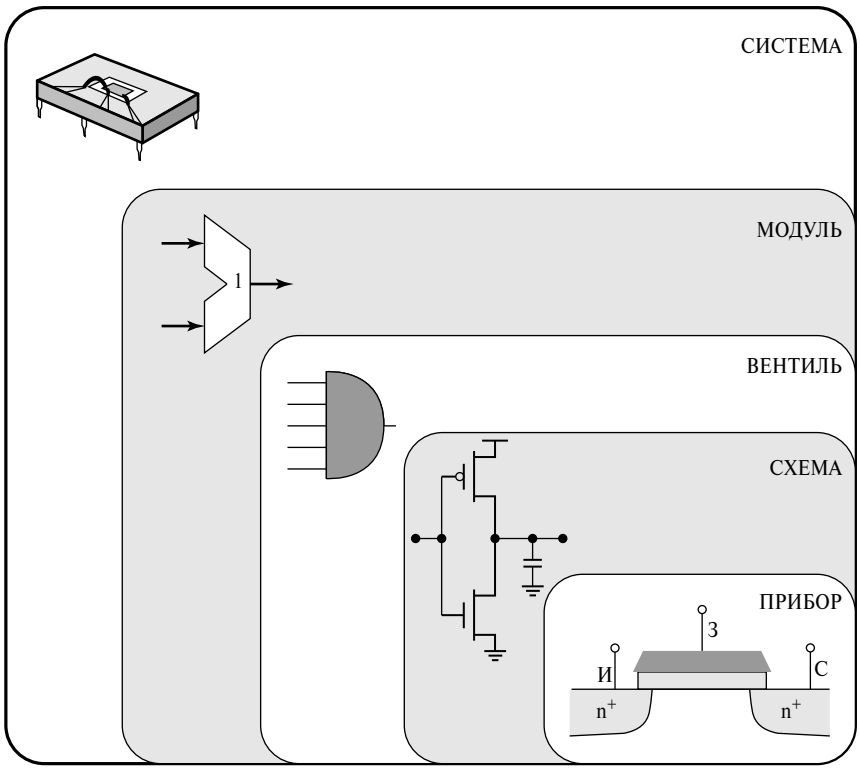


Рис. 1.6. Уровни абстрагирования при проектировании цифровых интегральных схем

менее привлекательным. Если вернуться к упомянутой ранее аналогии с методологиями разработки программного обеспечения, то в тех случаях, когда критически важным является время исполнения, программист склоняется к "индивидуализации" подпрограмм; компиляторы — или средства проектирования — еще не достигли того уровня, который могут дать упорный человеческий труд и изобретательность.

- Еще более важным является то наблюдение, что основанный на абстрагировании подход корректен только до определенной степени. Например, быстродействие сумматора может существенным образом зависеть от способа его подключения к окружению. Вносят вклад в общую задержку и сами межсоединения, поскольку они обладают паразитными емкостью, сопротивлением и даже индуктивностью. В будущем по мере масштабирования технологий в сторону уменьшения размеров влияние паразитных элементов межсоединений будет только возрастать.
- Масштабирование, как правило, усиливает и некоторые другие недостатки основанной на абстрагировании модели. Некоторые объекты проектирования приобретают *глобальный* или *внешний* характер. Примерами глобальных факторов являются тактовые сигналы, используемые для синхрони-

зации цифровой системы, и линии подвода питания. Рост размеров цифровых микросхем оказывает заметное влияние на эти глобальные сигналы. Например, подключение большего количества ячеек к линии питания может вызвать заметное падение напряжения на ней, что может замедлить работу всех подсоединенных к ней ячеек. Проблемы, связанные с распределением тактовых сигналов, синхронизацией работы микросхемы и распределением напряжения питания, приобретают все более и более важное значение. Их решение требует хорошего понимания всех особенностей проектирования цифровых микросхем.

- Еще одним следствием развития технологии является появление со временем *новых особенностей проектирования* и новых ограничений. Примером этого является периодическое возникновение в качестве ограничивающего фактора величины рассеиваемой мощности, о чем уже говорилось в историческом обзоре. В качестве другого примера можно привести изменение в пропорциях величин паразитных элементов прибора и межсоединений. Чтобы преодолеть эти непредвиденные факторы, необходимо по крайней мере уметь смоделировать и проанализировать их воздействие, что опять же требует глубокого понимания топологии цепи и ее поведения.
- Наконец, если что-то может пойти не так, то обычно так и происходит. Изготовленная микросхема не всегда имеет на выходе сигналы с той формой, которую можно было бы ожидать по результатам предварительного моделирования. Отклонения могут вызываться изменениями параметров процесса изготовления, индуктивностью корпуса или плохим моделированием тактового сигнала. *Устранение ошибок проектирования* требует квалификации в области схемотехники.

По всем этим причинам авторы свято верят, что глубокие знания в области методик проектирования цифровых интегральных схем и соответствующих подходов являются ценным активом любого разработчика цифровых систем. Даже если ему и не придется каждый день иметь дело со схемотехническими деталями поведения элементов микросхем, четкое представление о них несомненно поможет сориентироваться в неожиданных обстоятельствах и определить в процессе анализа разработки доминирующие эффекты.

Пример 1.1. Иерархический подход не годится при проектировании системы распределения тактовых импульсов

Для иллюстрации некоторых вопросов, поднятых ранее, исследуем влияние недостатков иерархического подхода на проектирование системы распределения наиболее важного глобального сигнала: тактовых импульсов. Функцией тактового сигнала в цифровой интегральной схеме является установление определенного порядка следования множества событий, имеющих в ней место. Эту задачу можно сравнить с функцией дорожного светофора, который определяет те автомобили, которым разрешается движение. Он также гарантирует завершение всех операций до начала следующих — сигнал светофора должен достаточно долго быть зеленым, чтобы дать возможность ма-

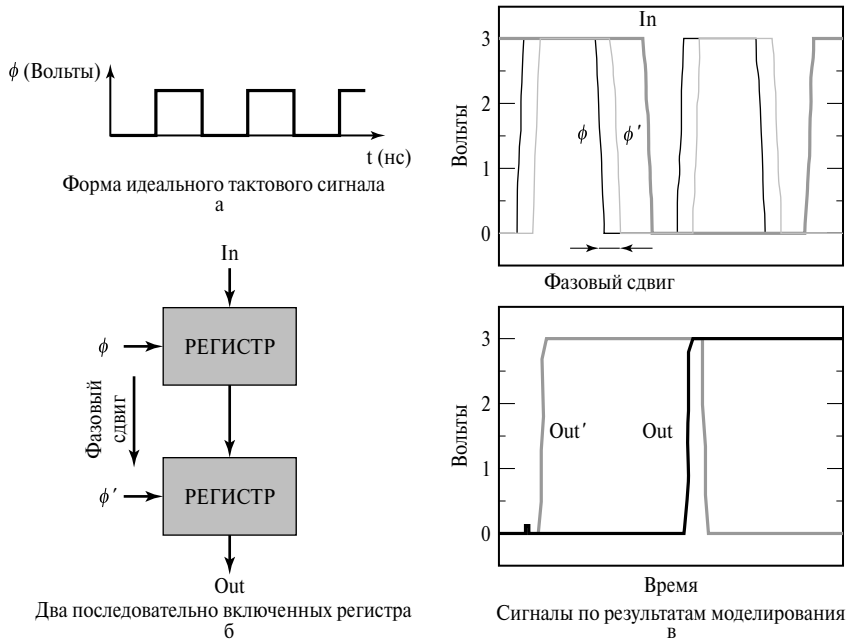


Рис. 1.7. Влияние рассогласованности тактовых сигналов

шине или пешеходу пересечь дорогу. В идеальных условиях тактовый сигнал представляет собой прямоугольный периодический сигнал, моменты изменения уровня которого синхронизированы в пределах проектируемой микросхемы (рис. 1.7, *а*). В свете нашей аналогии, переключения сигналов светофоров на дороге должны быть синхронизированы таким образом, чтобы обеспечить максимальную пропускную способность и одновременно исключить столкновения. Важность концепции согласованности тактовых сигналов иллюстрируется на примере двух последовательно включенных регистров, работающих по переднему фронту тактового сигнала ϕ (см. рис. 1.7, *б*). При нормальной работе входной сигнал *In* заносится в первый регистр по переднему фронту сигнала ϕ и появляется на выходе через промежуток времени, точно соответствующий одному периоду тактового сигнала. Это подтверждается результатами моделирования, показанными на рис. 1.7, *в* (сигнал *Out*).

Вследствие задержки, возникающей из-за различий в трассировке линий тактового сигнала, может случиться так, что тактовые сигналы окажутся рассогласованными по отношению друг к другу. Как результат, регистры интерпретируют время, маркерами которого являются тактовые импульсы, поразному. Рассмотрим случай, когда тактовый сигнал для второго регистра задерживается или сдвигается по фазе на величину δ. Передний фронт задержанного тактового сигнала ϕ' будет отодвигать во времени момент записи во второй регистр сигнала на его входе. Теперь, если время, необходимое для распространения сигнала от выхода первого регистра до входа второго, меньше величины задержки тактового сигнала, второй регистр будет подхваты-

вать ошибочный сигнал. Это вызывает преждевременное изменение сигнала на выходе, что отчетливо видно из результатов моделирования, когда сигнал Out' переходит в состояние с высоким логическим уровнем по первому фронту нарастания тактового сигнала ϕ', а не по второму, как это должно было бы быть. В терминах нашей аналогии с дорожным движением автомобили, проехавшие первый светофор, натыкаются на автомобили, еще не покинувшие второй светофор.

Рассогласование, или фазовый сдвиг тактовых сигналов, представляет собой важный пример того, как глобальные сигналы могут влиять на функционирование иерархически спроектированной системы. По сути, фазовый сдвиг тактовых сигналов является одной из наиболее серьезных проблем, с которыми приходится сталкиваться разработчику больших быстродействующих систем.

Пример 1.2. Иерархический подход не годится при проектировании систем разводки питания

Если тактовый сигнал является одним примеров глобального сигнала, пересекающего иерархические границы кристалла, то система разводки питания является вторым. В цифровой системе требуется, чтобы на каждый вентиль подавалось стабильное напряжение постоянного тока. Для обеспечения правильности работы это напряжение должно быть стабильным в пределах нескольких сотен милливольт. Система разводки питания должна обеспечить подобную стабильность напряжения в условиях очень больших изменений величины тока. Резистивный характер внутрикристалльной разводки и индуктивность выводов корпуса микросхемы делают эту задачу нетривиальной. Например, величина среднего тока, который должен подводиться в микропроцессор с напряжением питания 1 В и потребляемой мощностью 100 Вт, составляет 100 А! Пиковое же значение тока может быть вдвое выше, причем величина требуемого тока может изменяться от нуля до пикового значения за очень короткий промежуток времени: в диапазоне от 1 нс и меньше. Это дает значение скорости изменения тока 100 ГА/с, а это число поражает.

Рассмотрим проблему сопротивления разводки питания. При токе 100 А сопротивление дорожки всего в 1,25 мОм приводит к 5%-ному падению напряжения питания (при напряжении питания 2,5 В). Расширение дорожки снижает сопротивление и, следовательно, величину падения напряжения. Но если при прямолинейном проектировании такое увеличение размеров дорожек системы разводки питания реализуется достаточно просто, то при иерархическом проектировании все оказывается значительно сложнее. Рассмотрим, например, два блока, показанные на рис. 1.8, *а* [Saleh01]. Если система подвода питания блока А проектируется отдельно, то дополнительная нагрузка, возникающая из-за присутствия блока Б, не учитывается. Если линии питания к блоку Б прокладываются сквозь блок А, то для блока Б будет иметь место повышенное значение падения напряжения IR, поскольку подводимая мощность сначала потребляется в блоке А и только затем достигает блока Б.

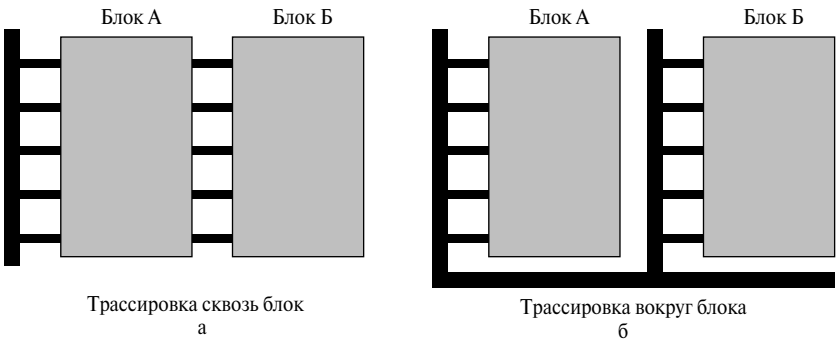


Рис. 1.8. Проектирование системы разводки питания

Поскольку общая величина падения напряжения IR зависит от сопротивления на участке от ножки микросхемы до блока, то можно проложить дорожки питания вокруг блока и подводить мощность к каждому блоку отдельно, как показано на рис. 1.8, *б*. В идеале главные каналы подвода питания должны быть достаточно широкими, чтобы справиться с пропусканием токов, протекающих во всех его ответвлениях. Хотя подобная трассировка дорожек системы питания легче поддается управлению и расчету, ее реализация требует большей площади. Широкие металлические дорожки канала подвода питания должны иметь достаточные размеры, чтобы пропускать все токи для каждого из блоков. Это требование заставляет проектировщиков отводить площадь под шины системы питания, забирая тем самым площадь, доступную под трассировку сигнальных линий связи.

С увеличением количества добавляемых блоков фактическое значение падения напряжения определяется сложным взаимодействием блоков. К примеру, не всегда легко определить путь, по которому потечет ток, если между источником питания и потребляющим вентилем одновременно существует множество параллельных путей. Кроме того, токи в различных модулях редко достигают своих пиковых значений в одно и то же время. Все эти соображения превращают задачу проектирования системы разводки питания в очень непростую работу. Ее решение требует наличия методологического подхода, который бы позволял преодолевать искусственные границы, порождаемые иерархическим проектированием.

Цель данной книги как раз и заключается в том, чтобы *перекинуть мостик между абстрактным взглядом на проектирование и лежащими на более низком уровне цифровыми схемами и их особенностями*. Начинается книга с описания принципов работы электронных приборов и углубленного анализа главного элемента всех цифровых устройств: инвертора. Затем полученные знания распространяются на вопросы проектирования более сложных объектов: вентилей, информационных каналов, регистров, котроллеров и запоминающих устройств. При проектировании каждого из перечисленных вы-

ше модулей перед разработчиком постоянно встает задача определения ключевых для данной разработки параметров, выявления той части разработки, на которой следует сосредоточить усилия по оптимизации, и выделения специфических свойств, отличающих исследуемый модуль (например, память) от всех других.

Также большое внимание в книге уделяется и другим (глобальным) аспектам проектирования современных цифровых интегральных схем, в частности *рассеиваемой мощности*, *системе межсоединений*, *организации временной диаграммы работы* и *синхронизации*.

1.3. Меры качества разработки цифровой интегральной схемы

Данный раздел посвящен описанию базовых свойств разработки цифровой интегральной схемы. Эти свойства помогают в получении количественной оценки качества разработки с различных углов зрения: по стоимости, функциональным возможностям, помехоустойчивости, быстродействия и энергопотребления. Какая из этих метрик является наиболее важной, зависит от конкретного применения разработки. Например, чистое быстродействие является критически важным свойством для сервера. С другой стороны, энергопотребление — доминирующая метрика для таких мобильных устройств, как сотовые телефоны. Описываемые свойства одинаково важны на всех уровнях иерархии разработки: системном, кристальном, модульном и вентильном. Для обеспечения совместимости определений на всех ступенях иерархии предлагается следовать подходу снизу вверх: в начале даются определения базовых метрик качества простого инвертора, которые постепенно будут распространены на более сложные функциональные узлы: модули и кристалл микросхемы.

1.3.1. Стоимость

Общую стоимость любого изделия можно разделить на две составляющие: восполняемые расходы, или *переменную стоимость*, и невосполняемые расходы, или *фиксированную стоимость*.

Фиксированная стоимость

Величина фиксированной стоимости не зависит от объема продаж или количества проданных изделий. Важной составляющей фиксированной стоимости интегральной схемы являются временные и трудовые затраты на получение законченной разработки. Стоимость разработки в сильной степени зависит от сложности разработки, агрессивности спецификаций и производительности разработчика. Последняя может быть улучшена за счет использования передовых методологий, позволяющих автоматизировать большинство этапов процесса проектирования. Снижение стоимости разработки в условиях посто-

янного роста сложности ИС представляет собой одну из основных проблем, всегда стоящих в полупроводниковой промышленности.

Дополнительно необходимо учитывать и *косвенную стоимость*, включающую накладные расходы компании, которые не могут быть непосредственно отнесены на стоимость одного изделия. В косвенную стоимость среди прочего входят расходы компании на исследования и разработки, производственное оборудование, маркетинг и организацию продаж, а также расходы на строительство.

Переменная стоимость

Переменная стоимость — это стоимость, которая непосредственно переносится на изготавливаемое изделие, и поэтому ее величина пропорциональна объему производства. Переменная стоимость включает в себя стоимость используемых в изделии комплектующих, стоимость сборки и стоимость испытаний. Общая стоимость интегральной схемы равна:

$$\text{стоимость ИС} = \text{переменная стоимость ИС} + \\ + \left(\frac{\text{фиксированная стоимость}}{\text{объем производства}}\right) \tag{1.1}$$

Влияние фиксированной стоимости более заметно для мелкосерийных изделий. Это объясняет, почему есть смысл иметь многочисленные группы разработчиков, работающих в течение нескольких лет над одним, но сверхуспешным изделием, например микропроцессором.

Хотя за последние несколько десятилетий стоимость производства одного транзистора падала по экспоненте, основное уравнение определения переменной стоимости не изменилось:

$$\text{переменная стоимость} = \\ = \left(\frac{\text{ст. кристалла} + \text{ст. испытаний кристалла} + \text{ст. корпусирования}}{\text{конечный процент выхода годных кристаллов}}\right) \tag{1.2}$$

Как будет показано в главе 2, процесс изготовления ИС позволят формировать на одной *пластине* группу идентичных кристаллов (рис. 1.9). После завершения изготовления пластина разделяется на *кристаллы*, которые после тестирования по отдельности корпусируются. Здесь мы будем рассматривать только стоимость кристалла. Стоимости корпусирования и испытаний являются предметом обсуждения последующих глав.

Величина стоимости кристалла зависит от количества кристаллов на пластине и процента тех, которые функционируют. Последний параметр носит название *процента выхода годных кристаллов*, и в результате получается следующее выражение:

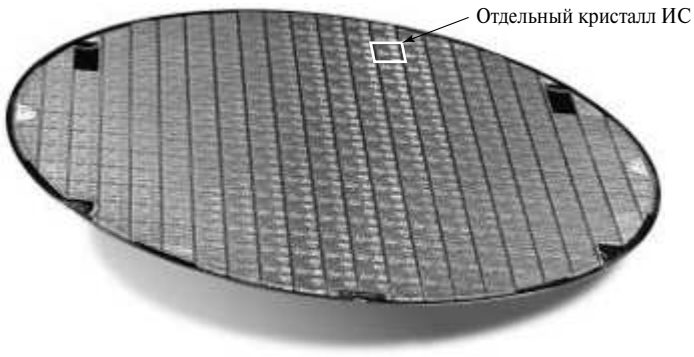


Рис. 1.9. Внешний вид пластины после обработки. Каждый квадратик — это кристалл микросхемы; в данном случае — микропроцессора Duron компании AMD™ (перепечатано с разрешения компании AMD)

$$\text{стоимость кристалла} = \frac{\text{стоимость пластины}}{\text{количество кристаллов на пластине} \times \text{выход годных кристаллов}}. \quad (1.3)$$

Количество кристаллов на пластине по сути равно площади пластины, деленной на площадь кристалла. Реальная ситуация несколько более сложная, поскольку пластины круглые по форме, а кристаллы — квадратные. Поэтому кристаллы, расположенные по периметру пластины, теряются. С годами размер пластины постоянно увеличивался, обеспечивая рост количества кристаллов, получаемых за один технологический цикл. Выражение (1.3) также содержит первое указание на то, что стоимость микросхемы зависит от площади кристалла: увеличение площади кристалла просто означает уменьшение количества кристаллов, умещающихся на пластине.

Реальное соотношение между стоимостью и площадью более сложно и зависит от процента выхода годных кристаллов. Как материал подложки, так и процесс изготовления вносят дефекты, которые могут вызывать отказ кристалла. Если предположить случайное распределение дефектов по пластине и то, что выход годных кристаллов обратно пропорционален сложности процесса изготовления, то можно получить следующее выражение:

$$\text{выход годных кристаллов} = \left(1 + \frac{\text{кол. дефектов на ед. площади} \times \text{пл. кристалла}}{\alpha}\right)^{-\alpha}, \quad (1.4)$$

где α — параметр, зависящий от сложности процесса изготовления и примерно пропорциональный количеству используемых фотошаблонов. Для современных сложных процессов изготовления КМОП ИС хорошим оценочным

значением является $\alpha = 3$. Величина количества дефектов на единицу площади представляет собой меру вероятности отказов, вызываемых материалом и производственным процессом. На сегодня типичными являются значения в пределах от 0,5 до 1 дефекта/см^2, но они сильно зависят от степени отработанности процесса.

Пример 1.3. Выход годных кристаллов

Предположим, что диаметр пластины составляет 12 дюймов, площадь кристалла = 2,5 см^2, плотность дефектов = 1 дефект/см^2 и $\alpha = 3$. Определим выход кристаллов для такого производственного цикла изготовления КМОП-схем.

Количество кристаллов на пластине можно оценить с помощью следующего выражения, учитывающего потерю кристаллов по периметру пластины:

$$\text{количество кристаллов на пластине} = \frac{\pi \times (\text{диам. пластины} \;/\; 2)^2}{\text{пл. кристалла}} - \frac{\pi \times \text{диам. пластины}}{\sqrt{2 \times \text{пл. кристалла}}}.$$

Для данного примера это означает наличие 252 ($= 296 - 44$) потенциально пригодных кристаллов. Процент выхода годных кристаллов можно вычислить с помощью выражения (1.4), и он равен 16%! Таким образом, в среднем только 40 кристаллов будут полностью функционирующими.

Вывод таков, что количество пригодных кристаллов на пластине и, следовательно, стоимость одного кристалла является сильной функцией площади кристалла. Если для небольших разработок выход годных кристаллов обычно очень высок, то после превышения некоторого порога его величина быстро падает. Держа в уме ранее выведенные уравнения и используя типовые значения параметров, можно заключить, что стоимость кристалла пропорциональна четвертой степени его площади:

$$\text{стоимость кристалла} = f(\text{площадь кристалла})^4. \qquad (1.5)$$

Площадь является тем параметром, который непосредственно контролируется разработчиком и представляет собой главную меру стоимости. Таким образом, малая площадь является желательным свойством цифрового вентиля. Чем меньше вентиль, тем выше плотность интеграции и меньше размер кристалла. Меньшие вентили также обычно обладают бо́льшим быстродействием и потребляют меньше энергии — величина общей емкости, которая представляет собой один из доминирующих параметров, определяющих быстродействие, тоже часто масштабируется в соответствии с величиной площади.

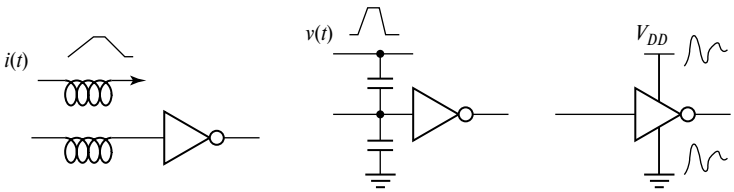


Рис. 1.10. Источники шума в цифровых микросхемах

Индикативным показателем ожидаемой площади кристалла является *количество транзисторов в вентиле*, но на ее величину могут оказывать влияние и другие параметры. Например, сложная картина соединений между транзисторами может привести к тому, что доминирующей окажется площадь разводки. Также на стоимость микросхемы влияет вентильная сложность, выражаемая через количество транзисторов, и регулярность структуры межсоединений. Сложные структуры труднее реализуемы и, как правило, требуют бо́льших затрат дорогого времени разработчика. Простота и регулярность — вот то, что требуется в чувствительных к стоимости разработках.

1.3.2. Функциональные возможности и помехоустойчивость

Очевидно, что главным требованием к цифровой микросхеме является выполнение ею той функции, для которой она была спроектирована. Измеряемое поведение изготовленной схемы обычно имеет отклонения от ожидаемой реакции. Одной из причин подобного отклонения от нормы являются вариации производственного процесса. Размеры и параметры приборов изменяются от партии к партии и даже на одной пластине или кристалле. Такие изменения могут существенным образом повлиять на электрическое поведение микросхемы. Вторым источником отклонений в характере реакции ИС являются возмущения, вызываемые источниками шума внутри или вне чипа. В контексте цифровых микросхем слово *шум* означает *нежелательные изменения напряжений и токов в логических узлах*. Шум может попадать в схему многими путями. Некоторые примеры источников шума в цифровых микросхемах приведены на рис. 1.10. Скажем, две дорожки, размещенные в интегральной микросхеме рядом друг с другом, взаимодействуют между собой через механизм взаимной емкости и индуктивности. Таким образом, изменение напряжения или тока в одной дорожке может повлиять на сигналы в соседней. Шум в шинах питания и земли вентиля также оказывает влияние на уровень сигналов в вентиле.

Шум в цифровой системе большей частью генерируется в ней самой, и величина шума пропорциональна размаху сигнала. Емкостная и индуктивная перекрестная помеха и внутренне генерируемый шум в цепях питания и земли — вот примеры такового. Другие источники шума, например шум на входных ножках питания и земли, являются внешними по отношению к микро-

схеме, и величина их шума не связана с уровнями сигналов. Для таких источников шум выражается непосредственно в вольтах или амперах. Шум от источников, являющихся функцией уровня сигнала, лучше выражать в долях или процентах от уровня сигнала. Шум представляет собой главную заботу в процессе разработки цифровых схем. Задача нейтрализации влияния всех этих возмущений — это одна из основных проблем, требующих решения в процессе проектирования быстродействующих цифровых микросхем и неоднократно рассматриваемая в данной книге.

Мерой помехоустойчивости схемы по отношению к вариациям процесса изготовления и шумовым возмущениям являются параметры вентиля в установившемся состоянии (также называемые *статическими характеристиками*). Определения и вывод выражений для этих параметров требуют предварительного понимания того, как цифровые сигналы представляются в мире электронных схем.

Цифровые схемы выполняют операции над *логическими* (или *булевыми*) переменными. Логическая переменная x может принимать только два дискретных значения:

$$x \in \{0, 1\}\,.$$

Например, инверсия (т.е. функция, которую реализует инвертор) означает воплощение следующего составного соотношения между двумя булевыми переменными x и y:

$$y = \overline{x}:\ \{x = 0\ \Rightarrow y = 1;\ x = 1\ \Rightarrow y = 0\}\,. \tag{1.6}$$

Однако логическая переменная — это всего лишь математическая абстракция. Физически такая переменная представляется электрической величиной. Чаще всего это напряжение в узле, которое не дискретно по своей природе и может принимать непрерывный ряд значений из определенного диапазона. Электрическое напряжение превращается в дискретную переменную путем назначения каждому логическому состоянию *номинального уровня напряжения*: $1 \Leftrightarrow V_{OH}$, $0 \Leftrightarrow V_{OL}$, где V_{OH} и V_{OL} соответствуют высокому и низкому логическим состояниям. Если на вход инвертора подать напряжение V_{OH}, то на его выходе установится напряжение V_{OL} и наоборот. Разность между двумя этими напряжениями называется *логическим перепадом* или *размахом сигнала* V_{sw}:

$$V_{OH} = \overline{(V_{OL})},$$
$$V_{OL} = \overline{(V_{OH})}.$$

Передаточная характеристика по напряжению

Предположим теперь, что логическая переменная *in* играет роль входного сигнала для инвертирующего вентиля, порождающего переменную *out*. Электрическая функция вентиля лучше всего описывается передаточной характеристикой по напряжению (иногда ее называют передаточной характеристикой по постоянному току), представляющую собой график зависимости выходного напряжения в функции входного напряжения $V_{out} = f(V_{in})$. Пример передаточной характеристики по напряжению инвертора показан на рис. 1.11. С ее помощью можно легко определить величины номинальных напряжений высокого и низкого состояний: $V_{OH} = f(V_{OL})$ и $V_{OL} = f(V_{OH})$. Другой представляющей интерес точкой на передаточной характеристике является *напряжение порога вентиля* или *переключения* V_M (не путать с пороговым напряжением транзистора), которое определяется как $V_M = f(V_M)$. Напряжения V_M также можно определить графически. Ему будет соответствовать точка пересечения передаточной характеристики и прямой, задаваемой уравнением $V_{out} = V_{in}$. Пороговое напряжение вентиля лежит точно посередине характеристики переключения, которая получается, если выход вентиля накоротко соединить с его входом. Эта точка будет представлять собой особый интерес при рассмотрении схем с обратной связью (также называемых *последовательными схемами*).

Даже если на вход вентиля подавать напряжение с идеальным номинальным значением, то сигнал на выходе очень часто будет отклоняться по величине от ожидаемого номинального значения. Эти отклонения могут вызываться шумом или нагрузкой выхода вентиля (т.е. другими вентилями, подключенными к его выходу). На рис. 1.12 показано, как в реальности логический уровень представляется не одними номинальными напряжениями, а диапазоном приемлемых напряжений, разделенных зоной неопределенности. Области приемлемых значений напряжений высокого и низкого логических состояний ограничиваются напряжениями с уровнями V_{IH} и V_{IL} соответственно. По определению они соответствуют точкам, в которых коэффициент усиления ($= dV_{out}/dV_{in}$) равен -1 (см. рис. 1.12, *б*). Зона между напряжениями V_{IH} и V_{IL} называется *областью неопределенности* (иногда ее еще называют *переходной*). Если необходимо гарантировать нормальную работу схемы, то устоявшиеся значения сигналов не должны попадать в эту область.

Запас устойчивости по шуму

Для того чтобы вентиль был помехоустойчив и нечувствителен к шуму, существенно, чтобы интервалы “0” и “1” были как можно бо́льшими. Мера чувствительности вентиля к шуму задается значениями запасов устойчивости по шуму NM_L (*запас устойчивости по шуму низкого логического состояния*) и NM_H (*запас устойчивости по шуму высокого логического состояния*), которые квантуют размеры допустимых интервалов “0” и “1” соответственно и устанавливают фиксированный порог на величину шума:

$$NM_L = V_{IL} - V_{OL},$$
$$NM_H = V_{OH} - V_{IH}.$$

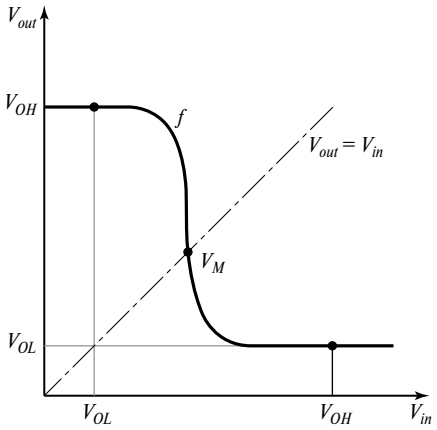
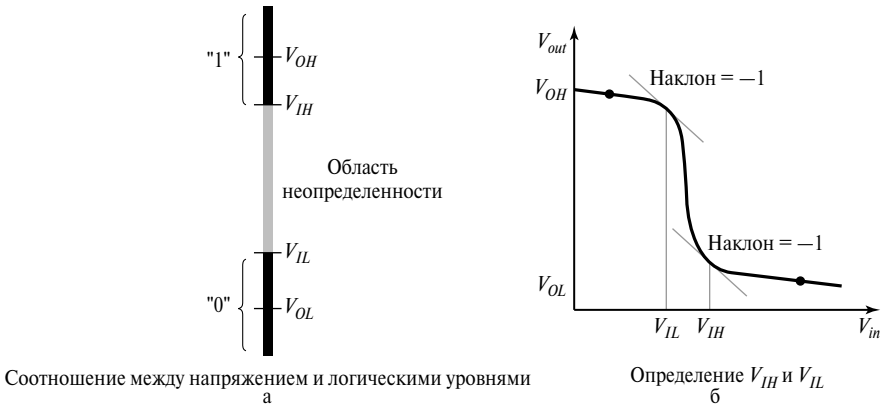


Рис. 1.11. Передаточная характеристика по напряжению инвертора



Соотношение между напряжением и логическими уровнями
а

Определение V_{IH} и V_{IL}
б

Рис. 1.12. Отображение логических уровней на область напряжений

Запасы устойчивости по шуму определяют уровни шума, который еще не будет приводить к ошибкам при каскадном включении вентилей, показанном на рис. 1.13. Совершенно очевидно, что для обеспечения работоспособности схемы эти запасы должны быть больше нуля и предпочтительно как можно бо́льшими.

Свойство регенеративности

Большой запас устойчивости по шуму желателен, но это не является единственным требованием. Предположим, что сигнал возмущается шумом и отличается по величине от номинальных уровней напряжения. Пока сигнал не выходит за пределы запаса устойчивости по шуму, вентиль продолжает работать нормально, хотя его выходное напряжение уже отличается от номинального. Это отклонение суммируется с шумом, инжектируемым в выходном узле, и передается на вход следующего вентиля. Влияние различных источни-

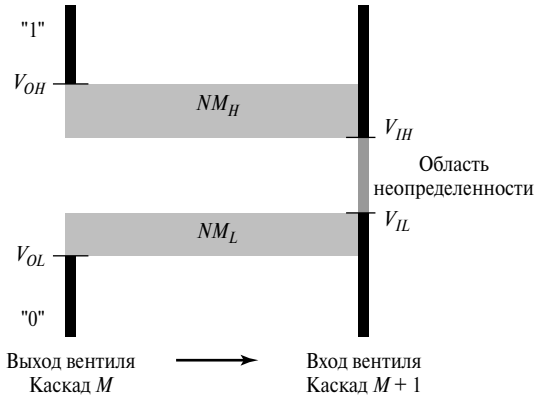


Рис. 1.13. Каскадное включение инвертирующих вентилей: определение запасов устойчивости по шуму

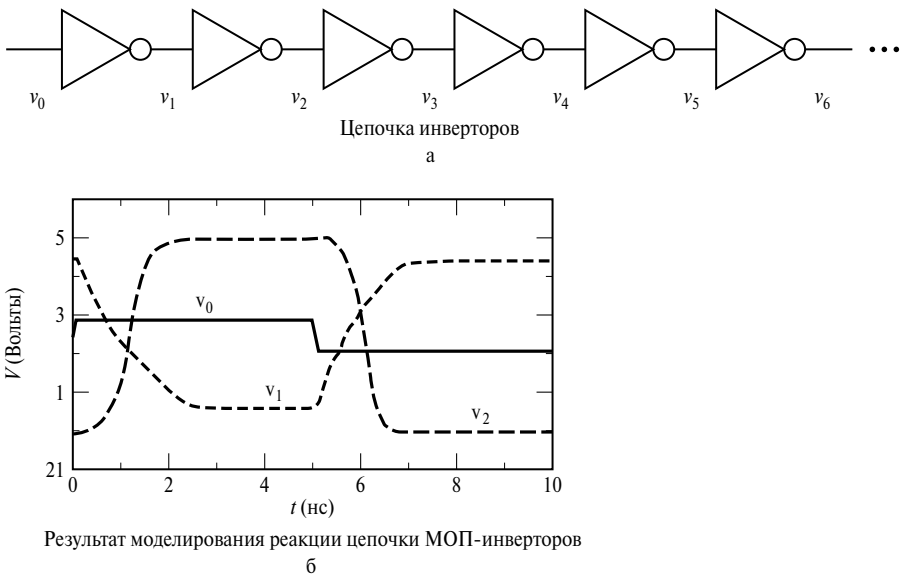


Рис. 1.14. Свойство регенеративности

ков шума могут накапливаться и в конце концов заставить уровень сигнала сместиться в зону неопределенности. К счастью, этого не происходит, если вентиль обладает *свойством регенеративности*, которое гарантирует постепенное возвращение возмущенного сигнала к одному из номинальных уровней напряжения после прохождения через несколько логических каскадов. Это свойство можно понимать следующим образом.

Пусть входное напряжение v_{in} ($v_{in} \in$ “0”) подается на вход цепочки из N инверторов (рис. 1.14, *а*). Если предположить, что количество инверторов в цепочке четное, то выходное напряжение $v_{out}(N \to \infty)$ будет равно

V_{OL} тогда и только тогда, когда инвертор обладает свойством регенеративности. Аналогично, если на вход цепочки инверторов подается входное напряжение $v_{in}(v_{in} \in$ "1″), то величина выходного напряжения будет приближаться к значению номинального напряжения V_{OH}.

Пример 1.4. Свойство регенеративности

Концепция регенерации иллюстрируется на рис. 1.14, *б*, который представляет собой результат моделирования переходных процессов в цепочке из КМОП-инверторов. На вход цепочки поступает сигнал в виде перепада напряжения с уменьшенной амплитудой, что могло быть вызвано наличием шума. Вместо размаха от напряжения питания до напряжения земли сигнал v_0 умещается в диапазоне от 2,1 до 2,9 В. Из результатов моделирования видно, что это отклонение от нормы быстро исчезает по мере продвижения сигнала по цепочке; например, сигнал v_1 уже имеет размах от 0,6 до 4,45 В. Еще дальше сигнал v_2 уже изменяет свою амплитуду между номинальными значениями V_{OL} и V_{OH}. Совершенно очевидно, что используемые в этом примере инверторы обладают свойством регенеративности.

Условия, при которых вентиль обладает свойством регенеративности, могут быть интуитивно определены на основе анализа простого случая. На рис. 1.15, *а* показан вид передаточной характеристики по напряжению инвертора $V_{out} = f(V_{in})$, а также приведен вид обратной функции $finv()$, которая меняет местами оси x и y и определяется следующим образом:

$$in = f(out) \Rightarrow in = finv(out). \tag{1.7}$$

Предположим, что на первый инвертор цепочки подается напряжение v_0, имеющее отклонение от номинального. Напряжение на выходе этого инвертора равно $v_1 = f(v_0)$, и оно будет входным для следующего инвертора. Графически это соответствует $v_1 = finv(v_2)$. Как показано стрелками, после прохождения нескольких каскадов величина напряжения сигнала постепенно приближается к номинальному значению. В ситуации, изображенной на рис. 1.15, *б*, напряжение сигнала стремится не к номинальным уровням, а к некоторой точке, соответствующей промежуточному уровню. Поэтому такая характеристика не имеет свойства регенеративности. Различие между этими двумя случаями заключается в усилительных характеристиках вентилей. Для обеспечения регенерации передаточная характеристика по напряжению должна иметь переходную область (или зону неопределенности) с коэффициентом усиления *больше 1* по абсолютной величине, ограниченную с двух сторон разрешенными зонами с коэффициентом усиления *меньше 1*. Такой вентиль имеет две стабильных рабочих точки. Из сказанного выше становится понятным определения уровней V_{IH} и V_{IL}, которые служат границей между разрешенными и переходной зонами.

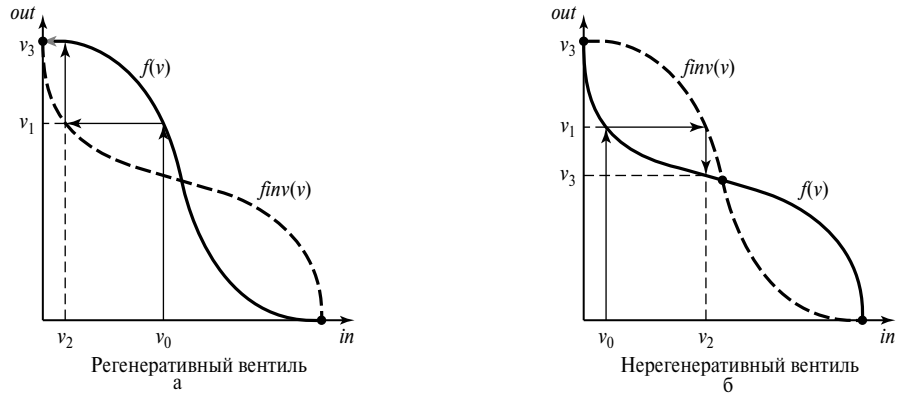


Рис. 1.15. Условия наличия свойства регенеративности

Невосприимчивость к шуму

Хотя запас устойчивости к шуму и является информативным средством для количественного измерения помехоустойчивости схемы при наличии шума, этого параметра недостаточно. Он выражает способность схемы "пересиливать" источник шума. С другой стороны, *невосприимчивость к шуму* выражает возможности системы правильно обрабатывать и передавать информацию в присутствии шума [Dally98]. Многие цифровые схемы с малым запасом устойчивости к шуму обладают очень хорошей невосприимчивостью к шуму, поскольку они не пересиливают, а скорее *подавляют источник шума*. Такие микросхемы обладают тем свойством, что только незначительная часть потенциально вредоносного сигнала от источника шума попадает в важные узлы схемы. Если выражаться более точным языком, то у таких микросхем значение передаточной функции между источником шума и сигнальным узлом значительно меньше единицы. Схемы, не обладающие таким свойством, *восприимчивы* к шуму.

Чтобы исследовать невосприимчивость вентиля к шуму, необходимо рассмотреть бюджет шума, определяющий мощность шума, поступающего от различных источников. Как уже говорилось ранее, источники шума можно разделить на следующие два типа источников:

- уровень шума от которых *пропорционален* размаху сигнала V_{sw}. Их влияние на сигнальный узел определяется выражением gV_{sw};
- уровень шума от которых *фиксирован*. Их влияние на сигнальный узел можно записать в виде fV_{Nf}, где V_{Nf} — это амплитуда шума от источника; f — передаточная функция от источника шума к сигнальному узлу.

Для простоты предположим, что запас устойчивости по шуму равен половине размаха сигнала (как для высокого, так и для низкого логических состояний). Чтобы схема работала правильно, запас устойчивости по шуму должен быть больше, чем сумма уровней шума, поступающего от всех источников:

$$V_{NM} = \frac{V_{SW}}{2} \geq \sum_i f_i V_{Nfi} + \sum_j g_j V_{sw}. \tag{1.8}$$

Для заданного множества источников шума можно вывести значение минимального размаха сигнала, при котором система еще будет работоспособной:

$$V_{sw} \geq \frac{2\sum_i f_i V_{Nfi}}{1 - 2\sum_j g_j}. \tag{1.9}$$

Отсюда понятно, что размах сигнала (и запас устойчивости по шуму) должен быть достаточно большим, чтобы преодолеть влияние от фиксированных источников (fV_{Nf}). С другой стороны, чувствительность к внутренним источникам зависит главным образом от способности вентиля к подавлению шума, которая пропорциональна значению коэффициентов усиления g_j. При больших значениях коэффициентов усиления увеличение размаха сигнала не дает ничего хорошего в плане подавления шума, так как шум растет пропорционально. В последующих главах будут рассмотрены семейства дифференциальных логических микросхем, которые способны подавлять бо́льшую часть внутренне генерируемого шума и, таким образом, работать при очень малых значениях запаса устойчивости по шуму и размаха сигнала.

Направленность

Свойство направленности требует, чтобы вентиль был *однонаправленным*, т.е. изменения уровня сигнала на выходе не должны проявляться на любом из входов той же микросхемы, состояние которых не изменяется. В противном случае переходные процессы на выходе будут отражаться на входах вентиля в виде сигнала шума, нарушая целостность полезного сигнала.

В реальных вентилях полной направленности не удается достичь никогда. Исключить наличие некоторой обратной связи от изменений уровней выходного сигнала невозможно. Типичным примером такой обратной связи является емкостная связь между входами и выходами. Важно минимизировать такое воздействие, чтобы оно не оказывало влияния на логические уровни сигналов на входе.

Разветвление по входу и выходу

Разветвлением по выходу называется количество вентилей нагрузки N, которые подключаются к выходу вентиля, играющего роль источника сигнала (рис. 1.16). Увеличение разветвления по выходу вентиля может влиять на логические уровни его сигнала на выходе. Из теории работы аналоговых усилителей известно, что этот эффект минимизируется путем увеличения, насколько это возможно, входного сопротивления вентилей нагрузки (при этом минимизируются входные токи) и путем поддержания выходного сопротивления вентиля возбуждения на низком уровне (чем снижается влияние токов

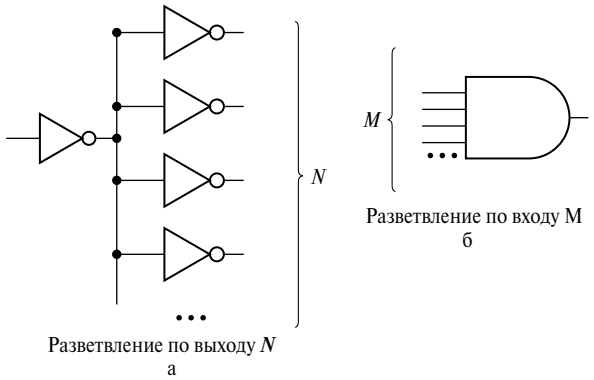


Рис. 1.16. Определения разветвления по выходу и входу цифрового вентиля

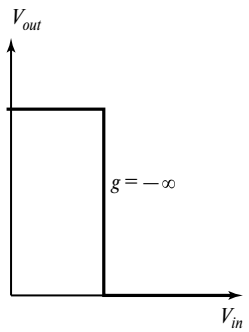


Рис. 1.17. Идеальная передаточная характеристика по напряжению

нагрузок на выходное напряжение). При высоком разветвлении по выходу каждая дополнительная нагрузка может сильно ухудшить динамические характеристики возбуждающего вентиля. По этой причине для многих базовых и библиотечных компонентов задается максимальное значение коэффициента разветвления по выходу, при котором еще гарантируется, что статические и динамические характеристики элемента будут удовлетворять определенным в спецификациях требованиям.

Под *разветвлением по входу* понимается *количество входов* вентиля (см. рис. 1.16, *б*). Вентили с большим разветвлением по входу обычно более сложные, что часто приводит к снижению их статических и динамических характеристик.

Идеальный цифровой вентиль

На основе всех приведенных выше рассуждений можно дать определение *идеального* цифрового вентиля с точки зрения статических характеристик. Модель идеального инвертора важна, поскольку позволяет получить меру, с помощью которой можно судить о качестве реальных приборов.

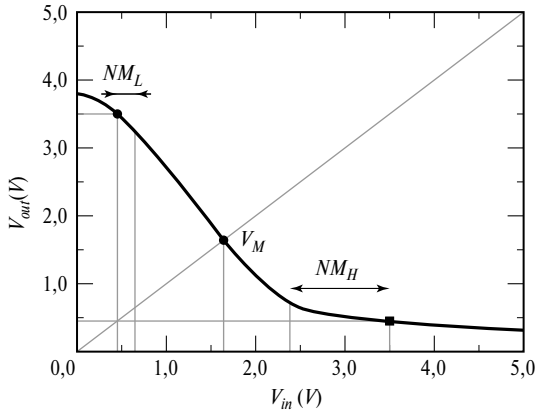


Рис. 1.18. Передаточная характеристика n-МОП-инвертора 1970-х годов

Передаточная характеристика по напряжению такого идеального инвертора показана на рис. 1.17 и обладает следующими свойствами: коэффициент усиления в переходной области равен бесконечности, точка порога вентиля располагается в середине размаха логических уровней, а величина запаса устойчивости по шуму в высоком и низком логических состояний равна половине размаха. Входной и выходной импедансы у идеального вентиля равны бесконечности и нулю соответственно (т.е. вентиль не имеет ограничений на разветвление по выходу). Хотя такую идеальную передаточную характеристику по напряжению, к сожалению, нельзя получить в реальных схемах, некоторые из них, например статические КМОП-инверторы, обладают характеристиками, достаточно близкими к ней.

Пример 1.5. Передаточная характеристика по напряжению

На рис. 1.18 приведен пример передаточной характеристики по напряжению реальной, но уже устаревшей вентильной структуры. Значения статических параметров получены из графика.

$$\begin{aligned} V_{OH} &= 3,5 \text{ В}; \quad V_{OL} = 0,45 \text{ В}; \\ V_{IH} &= 2,35 \text{ В}; \quad V_{IL} = 0,66 \text{ В}; \\ V_M &= 1,64 \text{ В}; \\ NM_H &= 1,15 \text{ В}; \quad NM_L = 0,21 \text{ В}. \end{aligned}$$

Очевидно, что приведенная характеристика далека от идеальной: она асимметрична, имеет очень низкое значение запаса устойчивости по шуму в низком логическом состоянии NM_L и размах напряжения 3,05 В, что значительно ниже максимально достижимого значения в 5 В (это напряжение источника питания для данного прибора).

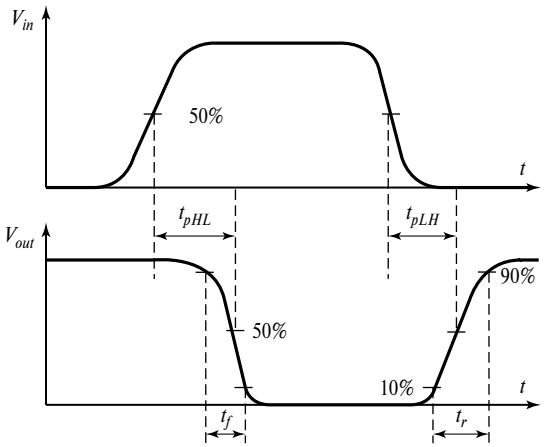


Рис. 1.19. Определение задержек распространения и времен фронтов нарастания и спада

1.3.3. Быстродействие

С точки зрения разработчика системы, быстродействие цифровой схемы определяет ее вычислительные возможности. Например, микропроцессоры часто характеризуются количеством выполняемых за одну секунду команд. Такая метрика быстродействия зависит как от архитектуры процессора — например, ряд команд может выполняться параллельно, — так и от фактического схемотехнического решения логических цепей. Хотя первый фактор и является критически важным, он не является предметом рассмотрения данной книги. Читатель может найти множество превосходных книг, посвященных этой теме (например, [Hennssy02]). Если же остановиться только на чисто схемотехническом аспекте проектирования, то наиболее часто быстродействие выражается через продолжительность периода тактовых импульсов (*время такта*) или скорость их поступления (*частота тактовых импульсов*). Минимальная величина периода тактовых импульсов для данной технологии и схемотехнических решений определяется рядом факторов, к которым можно отнести время прохождения сигнала через логическую схему, время записи и считывания данных из регистра и неопределенность момента прихода тактовых импульсов. Однако в основе анализа общего быстродействия лежит быстродействие отдельного вентиля.

Задержка распространения вентиля t_p определяет быстроту, с которой он реагирует на изменение состояния на его входе (входах). Выражает она ту *задержку, которую испытывает сигнал при прохождении через вентиль*. Измеряется же она между точками перехода входным и выходным сигналом уровня 50%, как показано на рис. 1.19 для инвертирующего вентиля[2]. Поскольку у вентиля разные времена реакции на передний и задний фронты

[2] Причиной измерения на уровне 50% служит предположение, что порог переключения V_M обычно лежит посередине размаха логического сигнала.

входного сигнала, то необходимо учитывать две задержки распространения. Задержка t_{pLH} определяет время реакции вентиля для случая перехода его выходного сигнала из низкого в высокое логическое состояние (положительный фронт), тогда как задержка t_{pHL} относится к случаю перехода из высокого логического состояния в низкое (отрицательный фронт). Задержка распространения t_p определяется как среднее между этими двумя задержками:

$$t_p = \frac{t_{pLH} + t_{pHL}}{2}. \tag{1.10}$$

ВНИМАНИЕ

Заметьте, что задержка распространения t_p, в отличие от параметров t_{pLH} и t_{pHL}, представляет собой искусственную метрику вентиля и сама по себе не имеет физического смысла. Она используется главным образом для сравнения различных полупроводниковых технологий, схемотехнических решений и стилей в проектировании логических функций.

Величина задержки распространения зависит не только от технологии получения микросхемы и ее топологии, но и от многих других факторов. Наиболее важно то, что задержка является функцией *наклона* фронтов входного и выходного сигналов вентиля. Чтобы иметь возможность количественной оценки этих свойств, вводятся понятия *времени фронта нарастания* и *времени фронта спада*, t_r и t_f, которые скорее представляют собой метрики, применимые в отношении формы отдельных сигналов, а не вентилей (см. рис. 1.19), и определяют быстроту перехода сигнала от одного уровня к другому. Неопределенность в определении фактического начала и окончания перехода устраняется за счет того, что, как показано на рисунке, времена фронтов нарастания и спада определяются по точкам, соответствующим 10 и 90% амплитуды сигнала. Величина времени фронта нарастания/спада обусловливается в основном "силой" возбуждающего вентиля и его нагрузкой.

При сравнении быстродействия вентилей, реализуемых с использованием разных технологий или схемотехнических решений, важно не запутывать картину влиянием таких параметров, как коэффициенты нагрузки, разветвления по входу и выходу. Необходим некий единообразный способ измерения величины t_p, который бы позволял судить о технологиях, ставя их в одинаковые условия. Стандартной схемой измерения задержки распространения де-факто является *кольцевой генератор*, состоящий из нечетного числа инверторов, включенных в замкнутое кольцо (рис. 1.20). Благодаря нечетности количества инверсий такая схема не имеет устойчивой рабочей точки и генерирует сигнал. Период колебаний T определяется величиной задержки распространения перепада сигнала сквозь всю цепочку или $T = 2 \times t_p \times N$, где N — количество инверторов в цепочке. Коэффициент 2 появляется из-за того, что полный период требует наличия как перехода из низкого состояния

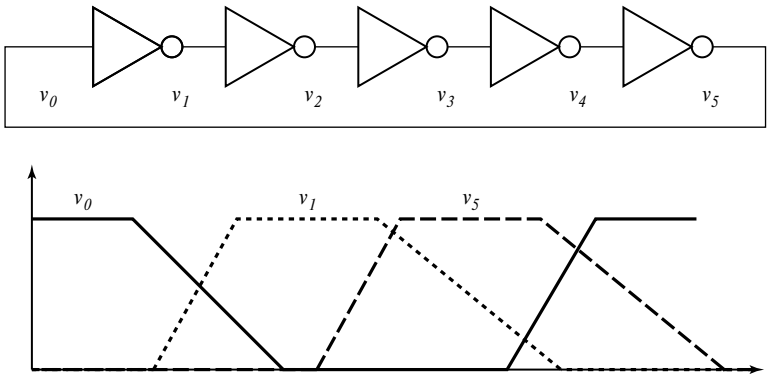


Рис. 1.20. Схема кольцевого генератора для измерений задержки распространения

в высокое, так и из высокого в низкое. Следует отметить, что это уравнение справедливо только в случае выполнения условия $2Nt_p \gg t_f + t_r$. Если это условие не выполняется, то схема может не генерировать сигнал: одна "волна" сигналов, распространяющихся по цепочке, будет накрывать предшествующую и, в конце концов, подавит генерацию. Как правило, кольцевой генератор работоспособен, если содержит по крайней мере пять каскадов.

ВНИМАНИЕ

Необходимо быть чрезвычайно осторожными с результатами, полученными в ходе измерений с использованием кольцевого генератора. Величина t_p в 20 пс совсем не означает, что схема, построенная на таких вентилях, будет работать на частоте 50 ГГц. Полученные по методу генератора результаты в основном полезны в плане получения количественной оценки отличий между различными технологиями изготовления и топологиями вентилей. Генератор представляет собой идеализированную схему, в которой каждый вентиль имеет коэффициенты разветвления по входу и выходу, точно равные 1, а паразитные нагрузки минимальны. В более реалистичных цифровых схемах значения коэффициентов разветвления по входу и выходу выше, да и задержки межсоединений не являются пренебрежимо малыми. Кроме того, функциональность вентилей в них также значительно сложнее, чем у простого инвертора. Как следствие, значение достижимой тактовой частоты в среднем в 50–100 раз ниже, чем значение частоты, полученное по результатам измерений с помощью кольцевого генератора. Но это в среднем; тщательно оптимизированные разработки могут существенно ближе приближаться по рабочей частоте к идеальному значению.

Пример 1.6. Задержка распространения *RC*-цепи первого порядка

Очень часто цифровые схемы моделируют RC-цепями первого порядка, аналогичными показанной на рис. 1.21. Таким образом, весьма интересно посмотреть на время задержки распространения такой цепи.

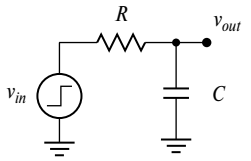


Рис. 1.21. RC-цепь первого порядка

Если на вход такой цепи подать перепад напряжения (v_{in} ступенчато изменяется от 0 до V), то, как известно, переходной процесс в ней описывается экспоненциальной функцией следующего вида (где $\tau = RC$ — постоянная времени цепи):

$$v_{out}(t) = (1 - e^{-t/\tau})V. \tag{1.11}$$

Отсюда легко вычислить время, требующееся для достижения уровня 50%: $t = ln(2)\tau = 0,69\tau$. Аналогично, для достижения уровня 90% потребуется время $t = ln(9)\tau = 2,2\tau$. Стоит запомнить эти цифры, поскольку они часто будут использоваться во всей книге.

1.3.4. Потребляемая мощность

Потребляемая микросхемой мощность определяет количество энергии, расходуемое на выполнение одной операции, и количество рассеиваемого ею тепла. Эти факторы в значительной степени влияют на большое количество критически важных конструктивных решений: мощность источника питания, требуемое время жизни аккумуляторной батареи, размер шин питания, метод корпусирования и требования к системе охлаждения. Таким образом, рассеиваемая мощность представляет собой важный параметр разработки, который определяет ее реализуемость, стоимость и надежность. В мире быстрых вычислений пределы потребляемой мощности, устанавливаемые типом корпуса и возможностями системы теплоотвода, определяют количество схем, которые могут быть интегрированы в одном кристалле, и с какой частотой они смогут переключаться. С увеличением популярности распределенных и мобильных вычислений энергетические ограничения устанавливают жесткие пределы на объем вычислений, который может быть выполнен при заданном минимальном времени между перезарядкой аккумуляторной батареи.

В зависимости от решаемой конструктивной проблемы могут рассматриваться различные меры рассеиваемой мощности. Например, пиковая мощность P_{peak} важна для определения размеров шин питания. При рассмотрении требований к системе охлаждения или емкости аккумулятора более интересна величина средней рассеиваемой мощности P_{av}. Эти меры определяются следующим образом:

$$P_{peak} = i_{peak} V_{supply} = max[p(t)];$$
$$P_{av} = \frac{1}{T}\int_0^T p(t)dt = \frac{V_{supply}}{T}\int_0^T i_{supply}(t)dt,$$

где $p(t)$ — мгновенное значение мощности; i_{supply} — ток, отбираемый от источника напряжения V_{supply} за период времени $t \in [0, T]$; i_{peak} — максимальное значение тока i_{supply} за период.

Далее рассеиваемую мощность можно разделить на *статическую* и *динамическую* составляющие. Последняя рассеивается только во время переходных процессов, когда вентиль переключается. Она связана с зарядом емкостей и возникновением временных путей протекания тока между шинами питания; поэтому ее величина пропорциональна частоте переключений: *чем больше количество событий переключения, тем выше динамическая потребляемая мощность*. С другой стороны, статическая составляющая присутствует даже тогда, когда переключений нет вообще, и обусловливается наличием статических проводящих путей между шинами питания или токов утечки. Она есть всегда, даже когда схема находится в режиме ожидания, и минимизация влияния этого источника потребления является стоящей целью.

Задержка распространения и потребляемая мощность вентиля взаимосвязаны: задержка распространения определяется главным образом скоростью, с которой заданное количество энергии может быть запасено в емкостях вентиля. Чем быстрее передается энергия (или чем выше потребление мощности), тем быстрее вентиль. Для конкретной технологии и топологии вентиля произведение потребляемой мощности и задержки распространения в общем случае есть величина постоянная. Это произведение называется *произведением мощность-задержка* и может рассматриваться в качестве меры качества переключающегося прибора. Произведение мощность-задержка представляет собой *энергию*, потребляемую вентилем в течение *одного события переключения*. Схемой для измерения величины произведения мощность-задержка опять же может быть кольцевой генератор.

Идеальный вентиль обладает высоким быстродействием и потребляет мало энергии. Произведение *энергия-задержка* представляет собой объединенную метрику, которая сводит эти два элемента воедино и часто используется в качестве основной метрики качества. Таким образом, должно быть понятно, что произведение *энергия-задержка* эквивалентно произведению *мощность-задержка*.

Пример 1.7. Рассеивание энергии в RC-цепи первого порядка

Давайте снова рассмотрим *RC*-цепь первого порядка (см. рис. 1.21). При подаче на вход цепи перепада напряжения (напряжение v_{in} скачком меняет свою величину от 0 до V) ей сообщается от источника сигнала определенное количество энергии. Общее количество энергии, передаваемой от источника (с момента начала переходного процесса до его полного окончания), можно легко вычислить:

$$E_{in} = \int_0^\infty i_{in}(t)v_{in}(t)dt = V\int_0^\infty C\frac{dv_{out}}{dt}dt = (CV)\int_0^V dv_{out} = CV^2. \qquad (1.12)$$

Интересно заметить, что величина энергии, требующейся для заряда конденсатора от 0 до напряжения V перепадом напряжения, является функцией высоты перепада и емкости и не зависит от величины сопротивления. Мы можем вычислить ту часть подводимой энергии, которая запасается в конденсаторе после завершения переходного процесса:

$$E_C = \int_0^\infty i_C(t)v_{out}(t)dt = \int_0^\infty C\frac{dv_{out}}{dt}v_{out}dt = C\int_0^V v_{out}dv_{out} = \frac{CV^2}{2}. \qquad (1.13)$$

Эта величина составляет половину энергии, поставляемой источником. Для тех, кому интересно, а что же произошло со второй половиной, простой анализ показывает, что за время переходного процесса эквивалентное количество энергии рассеивается резистором в виде тепла. Читателю предлагается самостоятельно продемонстрировать, что во время фазы разряда энергия, первоначально запасенная в конденсаторе, также рассеется резистором и перейдет в тепло.

1.4. Резюме

В данной вводной главе было рассказано об истории и тенденциях в области проектирования цифровых интегральных схем. Также были представлены важные количественные метрики, используемые для оценки качества разработки: стоимость, функциональные возможности, помехоустойчивость, быстродействие и рассеиваемая энергия/мощность.

1.5. Для любознательных

Разработка цифровых интегральных схем является темой многих книг и монографий. Чтобы помочь читателю в поиске дополнительной информации по какой-нибудь отдельной теме, ниже приводится обширный список различных ссылок. Поскольку последние разработки в области цифровых интегральных схем чаще всего публикуются в технических журналах и трудах конференций, ниже также приводится перечень наиболее важных из них.

Журналы

IEEE Journal of Solid-State Circuits
IEICE Transactions on Electronics (Japan)
Proceedings of The International Solid-State and Circuits Conference (ISSCC)
Proceedings of the VLSI Circuits Symposium
Proceedings of the Custom Integrated Circuits Conference (CICC)
European Solid-State Circuits Conference (ESSCIRC)

МОП ИС

Annaratone, *Digital CMOS Circuit Design*, Kluwer, 1986.

T. Dilliger, *VLSI Engineering*, Prentice Hall, 1988.

M. Elmasry, ed., *Digital MOS Integrated Circuits*, IEEE Press, 1981.

M. Elmasry, ed., *Digital MOS Integrated Circuits II*, IEEE Press, 1982.

L. Glasser and D. Dopperpuhl, *The Design and Analysis of VLSI Circuits*, Addison-Wesley, 1985.

A. Kung and Leblebici, *CMOS Digital Integrated Circuits*, 2^{nd} Ed., McGraw-Hill, 1999.

C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, 1980.

K. Martin, *Digital Integrated Circuit Design*, Oxford University Press, 2000.

D. Pucknell and K. Eshraghian, *Basic VLSI Design*, Prentice Hall, 1988.

M. Shoji, *CMOS Digital Circuit Technology*, Prentice Hall, 1988.

J. Uyemura, *Circuit Design for CMOS VLSI*, Kluwer, 1992.

H. Veendrick, *Deep-Submicron CMOS ICs: From Basics to ASICS*, Second Edition, Kluwer Academic Publishers, 2000.

N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*, Addison-Wesley, 1985, 1993.

Проектирование быстродействующих ИС

K. Bernstein et al, *High Speed CMOS Design Styles*, Kluwer Academic, 1998.

A. Chandrakasan, F. Fox, and W. Bowhill, ed., *Design of High-Performance Microprocessor Circuits*, IEEE Press, 2000.

M. Shoji, *High-Speed Digital Circuits*, Addison-Wesley, 1996.

Проектирование ИС с низким энергопотреблением

A. Chandrakasan and R. Brodersen, ed., *Low-Power Digital CMOS Design*, IEEE Press, 1998.

M. Pedram and J. Rabaey, ed., *Power-Aware Design Methodologies*, Kluwer Academic, 2002.

J. Rabaey and M. Pedram, ed., *Low-Power Design Methodologies*, Kluwer Academic, 1996.

G. Yeap, *Practical Low-Power CMOS Design*, Kluwer Academic, 1998.

Проектирование схем памяти

K. Itoh, *VLSI Memory Chip Design*, Springer, 2001.

B. Keeth and R. Baker, *DRAM Circuit Design*, IEEE Press, 1999

B. Prince, *Semiconductor Memories*, Wiley, 1991.

B. Prince, *High Performance Memories*, Wiley, 1996.

D. Hodges, *Semiconductor Memories*, IEEE Press, 1972.

Межсоединения и корпусирование

H. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley, 1990.

W. Dally and J. Poulton, *Digital Systems Engineering*, Cambridge University Press, 1998.

E. Friedman, ed., *Clock Distribution Networks in VLSI Circuits and Systems*, IEEE Press, 1995.

J. Lau et al, ed., *Electronic Packaging: Design, Materials, Process, and Reliability*, McGraw-Hill, 1998.

Средства и методологии проектирования

V. Agrawal and S. Seth, *Test Generation for VLSI Chips*, IEEE Press, 1988.

D. Clein, *CMOS IC Layout*, Newnes, 2000.

G. De Micheli, *Synthesis and Optimization of Digital Circuits*, McGraw-Hill, 1994.

S. Rubin, *Computer Aids for VLSI Design*, Addison-Wesley, 1987.

J. Uyemura, *Physical Design of CMOS Integrated Circuits Using L-Edit*, PWS, 1995.

A. Vladimirescu, *The Spice Book*, John Wiley and Sons, 1993.

W. Wolf, *Modern VLSI Design*, Prentice Hall, 1998.

Разработка биполярных и БиКМОП ИС

A. Alvarez, *BiCMOS Technology and Its Applications*, Kluwer, 1989.

M. Elmasry, ed., *BiCMOS Integrated Circuit Design*, IEEE Press, 1994.

S. Embabi, A. Bellaouar, and M. Elmasry, *Digital BiCMOS Integrated Circuit Design*, Kluwer, 1993.

Литература общего плана

J. Buchman, *CMOS/TTL Digital Systems Design*, McGraw-Hill, 1990.

H. Haznedar, *Digital Micro-Electronics*, Benjamin/Cummings, 1991.

D. Hodges and H. Jackson, *Analysis and Design of Digital Integrated Circuits*, 2nd ed., McGraw-Hill, 1988.

M. Smith, *Application-Specific Integrated Circuits*, Addison-Wesley, 1997.

R.K. Watts, *Submicron Integrated Circuits*, Wiley, 1989.

1.5.1. Список использованной литературы

[Bardeen48] J. Bardeen and W. Brattain, "The Transistor, a Semiconductor Triod",*Phys. Rev.*, vol. 74, p. 230, July 15, 1948.

[Beeson62] R. Beeson and H. Ruegg, "New Forms of All Transistor Logic", *ISSCC Digest of Technical Papers*, p. 10-11, Feb. 1962.

[Dally98] B. Dally, *Digital Systems Engineering*, Cambridge University Press, 1998.

[Faggin72] F. Faggin, M.E. Hoff, Jr., H. Feeney, S. Mazor, M. Shima, "The MCS-4 — An LSI MicroComputer System", *1972 IEEE Region Six Conference Record, San Diego*, CA, p. 1–6, April 1972.

[Harris56] J. Harris, "Direct-Coupled Transistor Logic Circuitry in Digital Computers", *ISSCC Digest of Technical Papers*, p. 9, Feb. 1956.

[Hart72] C. Hart and M. Slob, "Integral Injection Logic — A New Approach to LSI", *ISSCC Digest of Technical Papers*, p. 92-93, Feb. 1972.

[Hoff70] E. Hoff, "Silicon-Gate Dynamic MOS Cram 1024 Bits on Chip", *Electronics*, p. 68–73, Augest 3, 1970.

[Intel01] "Moore's Law", `http://www.intel.com/research/silicon/mooreslaw.htm`.

[Masaki74] A. Masaki, Y. Harada and T. Chiba, "200-Gate ECL Master-Slice LSI", *ISSCC Digest of Technical Papers*, p. 62-63, Feb. 1974.

[Moore65] G. Moore, "Cramming more Components into Integrated Circuits", *Electronics*, vol. 38, Nr. 8, April 1965.

[Murphy93] B. Murphy, "Perspectives on Logic and Microprocessors", *Commemorative Supplement to the Digest of Technical Papers, ISSCC*, pp. 49–51, San Francisco, 1993.

[Norman60] R. Norman, J. Last and I. Haas, "Solid-State Micrologic Elements", *ISSCC Digest of Technical Papers*, p. 82-83, Feb. 1960.

[Hennessy02] J. Hennessy, D. Patterson, and David Goldberg, *Computer Architecture — A Quantitative Approach*, Third Edition, Morgan Kaufmann Publishers, 2002.

[Saleh01] R. Saleh, M. Benoit, and P. McCrorie, "Power Distribution Planning", Simplex Solutions, `http://www.simplex.com/wt/sec.php?page_name=wp_powerplan`.

[Schokley49] W. Schokley, "The Theory of pn-Junctions in Semiconductors and pn-Junction Transistors", *BSTJ*, vol. 28, p. 435, 1949.

[Shima74] M. Shima, F. Faggin and S. Mazor, "An N-Channel, 8-bit Single-Chip Microprocessor", *ISSCC Digest of Technical Papers*, p. 56-57, Feb. 1974.

[Swade93] D. Swade, "Redeeming Charles Babbage's Mechanical Computer", *Scientific American*, p. 86–91, Feb. 1993.

[Wanlass63] F. Wanlass, and C. Sah, "Nanowatt Logic Using Field-Effect Metal-Oxide-Semiconductor Triodes", *ISSCC Digest of Technical Papers*, p. 32-33, Feb. 1963.

1.5.2. Упражнения

За последней версией сборника задач и упражнений обратитесь, пожалуйста, по адресу `http://bwrc.eecs.berkley.edu/IcBook`. Сделав упражнения доступными в электронном, а не печатном виде, авторы смогли обеспечить динамичную среду, способную отслеживать быстрые изменения в современных технологиях проектирования цифровых интегральных схем.

ГЛАВА 2

Технология производства

Краткое описание технологического процесса производства

Нормы проектирования

Корпусирование ИС

Тенденции в технологии интегральных схем

2.1. Введение
2.2. Изготовление КМОП ИС
2.2.1. Кремниевая пластина
2.2.2. Фотолитография
2.2.3. Некоторые повторяющиеся шаги технологического процесса
2.2.4. Схема упрощенного КМОП-процесса
2.3. Нормы проектирования: между разработчиком и технологом
2.4. Корпусирование интегральных схем
2.4.1. Материалы корпусов
2.4.2. Уровни межсоединений
2.4.3. Учет тепловыделения при корпусировании
2.5. Тенденции в технологии интегральных схем
2.5.1. Краткосрочная перспектива
2.5.2. Долгосрочная перспектива
2.6. Резюме
2.7. Для любознательных

2.1. Введение

Большинству разработчиков цифровых схем никогда не придется сталкиваться с деталями технологического процесса, лежащего в основе возникновения полупроводниковой революции. Тем не менее некоторое понимание тех шагов, которые приводят к получению работающего кремниевого чипа, весьма полезно для того, чтобы почувствовать физический смысл ограничений, которых должен придерживаться разработчик, а также разобраться во взаимосвязи между процессом производства и такими параметрами, как, скажем, стоимость.

В данной главе приводится краткое описание шагов и методик, использующихся в технологическом процессе изготовления современных интегральных схем. Здесь авторы не ставят перед собой целью дать подробное описание технологии изготовления, изучение которой само по себе заслуживает полного курса [Plummer00]. Скорее тут предпринимается попытка представить общую картину последовательности и взаимодействия различных шагов процесса. Читатель узнает, что именно набор *оптических масок*, или *фотошаблонов*, является основным интерфейсом между внутренними особенностями процесса изготовления и разработкой, которую пользователь хотел бы видеть воплощенной в виде кремниевой структуры. Эти маски определяют рисунки, которые, будучи перенесенными на различные слои полупроводникового материала, формируют элементы электронных приборов и электрические соединения. Как следствие, для получения полностью работоспособной схемы эти рисунки должны удовлетворять некоторым ограничениям по минимальной ширине и расстоянию между ними. Такой набор ограничений называется *нормами проектирования* и играет роль контракта между разработчиком интегральной схемы и технологом. Если разработчик придерживается этих правил, то он получает гарантию того, что схему можно изготовить. В данной главе дается краткий обзор общих норм проектирования, которые приходится учитывать при разработке современных КМОП-схем, а также приводится описание вариантов *корпусирования* ИС. Корпус является интерфейсом между реализованной в кристалле полупроводника микросхемой и внешним миром и поэтому оказывает огромное влияние на быстродействие, надежность, долговечность и стоимость интегральной схемы.

2.2. Изготовление КМОП ИС

На рис. 2.1 показан упрощенный вид поперечного сечения типового КМОП-инвертора. Процесс получения КМОП ИС требует формирования в объеме одной и той же пластины кремния как n-канальных (n-МОП), так и p-канальных (p-МОП) транзисторов. Чтобы иметь возможность размещения приборов обоих типов, необходимо создать специальные области, называемые *карманами*, тип полупроводникового материала в которых противоположен типу канала. МОП-транзистор с p-каналом должен создаваться либо в подложке n-типа, либо в n-кармане, тогда как n-МОП-транзистор раз-

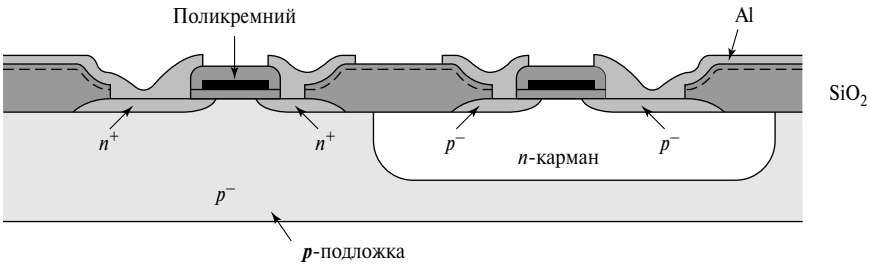


Рис. 2.1. Поперечное сечение структуры, получаемой в ходе КМОП-процесса с n-карманами

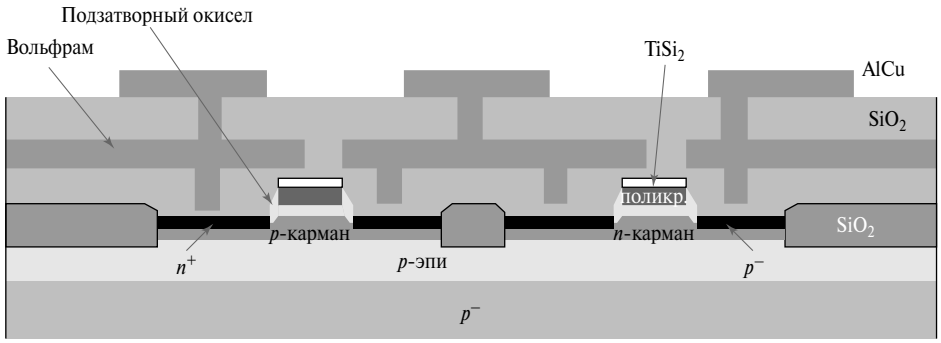


Рис. 2.2. Поперечное сечение современной КМОП-структуры, получаемой по двухкарманной технологии

мещается либо в подложке p-типа, либо в p-кармане. Поперечное сечение, показанное на рис. 2.1, соответствует КМОП-процессу с n-карманами, когда n-МОП-транзисторы формируются в подложке p-типа, а p-МОП-приборы размещаются в n-кармане. В современных процессах все чаще используется *двухкарманный* подход, когда в эпитаксиально выращенном слое формируются как n-карманы, так и p-карманы (рис. 2.2).

КМОП-процесс требует большого количества шагов, каждый из которых состоит из последовательности базовых операций. В ходе процесса изготовления ряд этих шагов и/или операций часто периодически повторяется. Прежде чем приступить к описанию общей последовательности этапов процесса, рассмотрим исходный материал, а затем подробно остановимся на некоторых наиболее часто повторяющихся операциях.

2.2.1. Кремниевая пластина

Базовым материалом для процесса изготовления является монокристаллическая слаболегированная *пластина*. Такие пластины обычно имеют диаметр от 4 до 12 дюймов (10 и 30 см соответственно) и толщину до 1 мм. Получаются они путем нарезания монокристаллического слитка на тонкие пластинки (рис. 2.3). Исходная пластина p^--типа может быть легированной до уровня

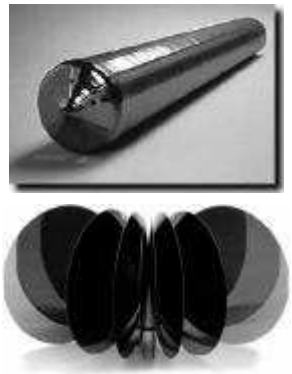


Рис. 2.3. Нарезание монокристаллического слитка на тонкие пластины

примерно 2×10^{21} см$^{-3}$. Перед передачей пластин в производство часто ее поверхность легируется до более высокого уровня, и поверх этого сильнолегированного слоя наращивается монокристаллический *эпитаксиальный слой* с проводимостью противоположного типа. Важной метрикой является плотность дефектов исходного материала. Высокая плотность дефектов приводит к увеличению доли неработоспособных микросхем и, как следствие, к увеличению стоимости конечного продукта.

2.2.2. Фотолитография

В ходе выполнения каждого шага процесса определенные области кристалла микросхемы маскируются с использованием соответствующего фотошаблона, так что требуемая обработка может производиться только на оставшейся незакрытой части чипа. Обработка может быть различной, включая окисление, травление, осаждение металла или поликремния и ионную имплантацию. Метод, который позволяет осуществлять селективное маскирование, называется *фотолитографией* и постоянно используется в процессе изготовления ИС. На рис. 2.4 графически показаны различные операции, выполняемые в ходе типового процесса фотолитографии. Можно выделить следующие этапы.

1. *Формирование окисного слоя.* Этот этап необязателен и заключается в формировании на всей поверхности пластины тонкого слоя SiO_2 путем выдержки пластины в атмосфере из смеси кислорода и водорода с высокой степенью очистки при температуре около 1000°C. Оксид используется как изолирующий слой или подзатворный диэлектрик транзисторов.
2. *Нанесение фоторезиста.* Светочувствительный полимер (подобный латексу) путем вращения пластины наносится равномерным слоем толщиной примерно 1 мкм. Исходный материал растворим в органическом растворителе, но обладает тем свойством, что под воздействием света по-

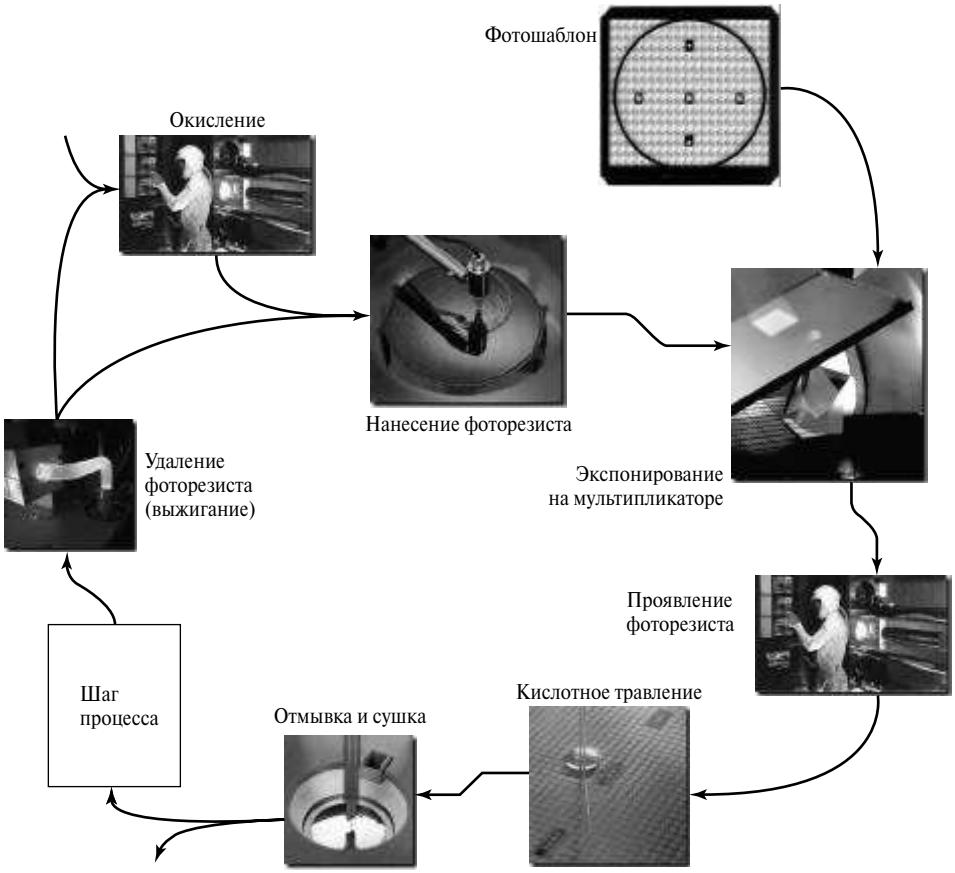


Рис. 2.4. Типичная последовательность работ в фотолитографичном цикле

перечные связи в полимере сшиваются, делая его нерастворимым. Фоторезист такого типа называется *негативным*. Позитивный фоторезист имеет противоположное свойство: первоначально нерастворимый, он становится растворимым после освещения. При использовании позитивного и негативного резистов одна и та же маска может иногда участвовать в двух этапах технологического процесса, делая доступными для обработки взаимодополняющие области кристалла микросхемы. Поскольку с уходом технологий в область все меньших размеров стоимость одного фотошаблона увеличивается достаточно быстро, то уменьшение количества шаблонов — задача высокоприоритетная.

3. *Пошаговое экспонирование.* Стеклянная маска (или фотооригинал), содержащая рисунок структур, которые надо перенести на кремний, устанавливается в непосредственной близости от поверхности пластины. Маска непрозрачна в тех областях, которые должны будут подвергнуться обработке, и прозрачна в других (предполагается, что используется нега-

тивный фоторезист). Стеклянную маску можно рассматривать как негативное изображение одного слоя микросхемы. Затем совмещенные маска и пластина подвергаются воздействию ультрафиолетового света. В тех местах, где маска прозрачна, фоторезист становится нерастворимым.

4. *Проявление и дубление фоторезиста.* Для удаления необлученных светом областей фоторезиста пластина проявляется в кислотном или щелочном растворе. После удаления необлученных участков фоторезиста пластина подвергается "мягкому дублению" при низкой температуре, в результате которого увеличивается стойкость оставшегося фоторезиста.
5. *Кислотное травление.* В областях пластины, незакрытых фоторезистом, выполняется селективное удаление материала. Осуществляется это в зависимости от подлежащего удалению материала путем использования различных кислотных, щелочных или солевых растворов. Бо́льшая часть работ с химикатами проводится в больших установках с жидкостными ваннами, в которых под конкретные задачи готовятся специальные растворы. Вследствие опасной природы некоторых из растворителей бо́льшую важность приобретают вопросы, связанные с техникой безопасности и воздействием на окружающую среду.
6. *Отмывка* и *сушка.* В специальной установке отмывки и сушки осуществляется очистка пластины в деионизованной воде с последующей сушкой азотом. Микроскопический масштаб современных полупроводниковых приборов означает, что даже мельчайшие частички пыли или грязи могут вызвать повреждение схемы. Чтобы этого не происходило, все этапы обработки выполняются в сверхчистых помещениях, где количество пылинок в одном кубическом футе поддерживается на уровне от 1 до 10. Кроме того, там, где это возможно, операции выполняются с использованием автоматических и роботизированных установок. Это объясняет тот факт, что стоимость современных производственных помещений достигает нескольких миллиардов долларов. Но даже при всем этом во избежание загрязнения и для удаления остатков от предшествующих операций обработки пластины должны постоянно подвергаться очистке.
7. *Различные обработки.* Открытые области теперь могут подвергаться различной обработке, например, ионной имплантации, плазменному травлению или осаждению металла. Эти операции будут рассмотрены в последующих разделах.
8. *Удаление фоторезиста (или сжигание).* Для селективного удаления остатков фоторезиста без повреждения слоев приборов используется обработка в высокотемпературной плазме.

На рис. 2.5 показаны этапы фотолитографического процесса на примере создания рисунка в слое SiO_2. Показанная на рисунке последовательность шагов обработки позволяет создать рисунок исключительно в одном слое полупроводникового материала и может показаться очень сложной. Но читатель должен помнить, что эта последовательность операций переносит рисунок на всю поверхность пластины. Таким образом, этот процесс в высшей степени

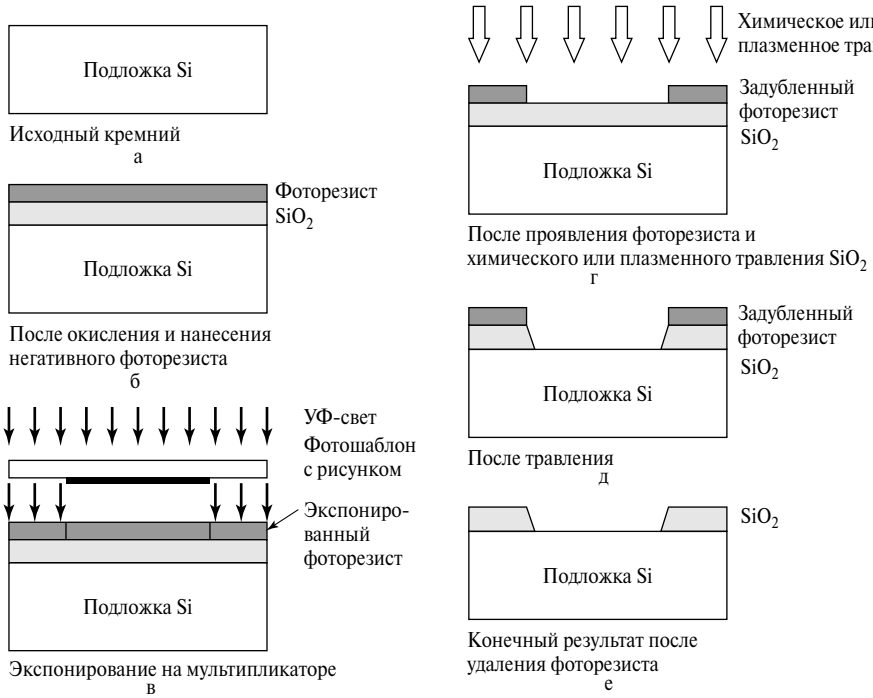


Рис. 2.5. Этапы обработки при создании рисунка в слое SiO_2

параллельный и позволяет формировать на полупроводниковой пластине сотни миллионов элементов одновременно. Параллельность и масштабируемая природа процесса оптической фотолитографии — это именно то, что делает возможным дешевое изготовление сложных полупроводниковых схем и лежит в основе экономического успеха полупроводниковой промышленности.

Продолжающееся уменьшение размера минимального топологического элемента в интегральных схемах является постоянной проблемой для разработчиков производственного оборудования для полупроводниковой промышленности. Особенно это справедливо в отношении фотолитографического оборудования. Размер минимального воспроизводимого элемента уже едва превышает длину волны источников света, так что задача достижения необходимой степени разрешения и точности становится все более и более трудной. Инженерам-электронщикам удалось довести процесс оптической литографии по крайней мере до размеров 100 нм (или 0,1 мкм). Метод *оптической коррекции масок* позволяет заранее вносить деформацию формы топологического элемента, учитывающую дифракционные явления, возникающие при переносе изображения с размерами, близкими к длине волны имеющегося источника света. Но это существенно увеличивает стоимость изготовления масок. В обозримом будущем могут потребоваться другие решения, способные обеспечить более высокую степень разрешения, например, переход на использование глубокого ультрафиолета, рентгеновского излучения или электронного

луча. И хотя эти методы полностью работоспособны, они непривлекательны с экономической точки зрения.

2.2.3. Некоторые повторяющиеся шаги технологического процесса

Диффузия и ионная имплантация

Многие шаги процесса изготовления интегральных схем требуют изменения концентрации легирующей примеси на некоторых участках материала. В качестве примеров можно привести создание областей стока и истока, карманов и контактов к подложке, легирование поликремния и подгонку порогов приборов. Существуют два подхода к введению легирующих примесей: диффузия и ионная имплантация. Оба метода предусматривают вскрытие подвергающихся легированию областей, при этом остальная поверхность пластины остается закрытой слоем буферного материала, в роли которого обычно выступает SiO_2.

В случае *диффузионной имплантации* пластины помещают в кварцевую трубу, находящуюся в разогретой печи. В трубу подается газ, содержащий легирующую примесь. Высокая температура печи, обычно от 900 до 1100°C, вызывает диффузию легирующей примеси в открытые области пластины в вертикальном и горизонтальном направлениях. Получающаяся в результате концентрация примеси максимальна у поверхности и уменьшается с глубиной, следуя гауссовому распределению.

При *ионной имплантации* легирующая примесь вводится в материал в виде ионов. Система ионной имплантации направляет качающийся луч очищенных ионов на поверхность материала. Скорость ионов определяет глубину их проникновения в материал, тогда как ток луча и время облучения определяют дозу. Метод ионной имплантации позволяет независимо управлять глубиной и дозой. Именно по этой причине ионная имплантация в значительной степени вытеснила диффузию при изготовлении современных полупроводниковых приборов.

Однако ионная имплантация, к сожалению, не свободна от побочных эффектов, среди которых одним из наиболее важных является повреждение кристаллической решетки. Столкновения атомов во время имплантации с высокой энергией приводят к смещению атомов подложки, вызывая тем самым появление в материале дефектов. Эта проблема в значительной степени решается введением последующего этапа *отжига*, во время которого пластина нагревается примерно до 1000°C и выдерживается при этой температуре в течение 15–30 минут, после чего медленно охлаждается. Нагрев вызывает тепловые колебания атомов, в результате чего связи восстанавливаются.

Осаждение

Любой процесс изготовления КМОП-структур требует повторяющегося осаждения на всю поверхность пластины слоя материала, который может играть роль буфера, быть изолирующим или проводящим слоем. Мы уже рассматривали процесс окисления, который позволяет выращивать слой SiO_2. Для осаждения других материалов требуются другие методы. Например, при

формировании рисунка в базовом окисле и как ограничитель имплантации в качестве удаляемого буфера используется нитрид кремния (Si_3N_4). Нитрид кремния повсеместно осаждается с использованием процесса, называемого *химическим осаждением из паровой фазы*. Этот процесс основан на газотранспортной реакции, энергия для которой подводится за счет нагрева до температуры около 850°C.

С другой стороны, поликремний осаждается посредством химического осаждения, когда над нагретой до 650°C пластиной, покрытой SiO_2, пропускается газ силан. В результате реакции получается некристаллический, или аморфный, материал, называемый *поликремнием*. Чтобы увеличить проводимость материала, после осаждения необходим этап имплантации.

Алюминиевые слои межсоединений обычно наносятся с использованием процесса, известного как *распыление*. Алюминий испаряется в вакууме, при этом необходимое тепло подводится либо с помощью электронного луча, либо путем ионной бомбардировки. Другие металлы для создания межсоединений, например медь, требуют применения других методов осаждения.

Травление

После того как материал осажден, для селективного формирования рисунка, например разводки или контактных отверстий, используется процесс травления. Мы уже рассматривали процесс *мокрого травления*, в ходе которого используются кислотные или щелочные растворы. Например, для травления SiO_2 обычно используется плавиковая кислота с буфером, в качестве которого выступает фтористый аммоний.

В последние годы широко применяется *сухое* или *плазменное травление*. Пластина помещается под колпак установки травления, и на нее подается отрицательное смещение. Колпак нагревается до 100°C и откачивается до давления 7,5 Па, после чего заполняется положительно заряженной плазмой (обычно это смесь азота, хлора и трихлорида бора). Противоположный электрический заряд заставляет быстро движущиеся молекулы плазмы выстраиваться в вертикальном направлении, в результате чего они начинают действовать по типу "микропескоструйного" аппарата с химическим и физическим действием, удаляя материал. Плазменное травление обладает преимуществом направленности, позволяя получать структуры с четкими вертикальными стенками.

Планаризация

Для надежного осаждения слоя материала на поверхность полупроводника важно, чтобы поверхность была как можно более плоской. Если не предпринять специальных действий, то в процессе изготовления КМОП-схем, когда множество металлических слоев системы межсоединений с перекрывающимися рисунками накладываются друг на друга, это определенно будет порождать проблемы. Поэтому перед осаждением следующего слоя металла поверх изолирующего слоя SiO_2 вводится этап *химико-механической планаризации*. В этом процессе для придания поверхности прибора плоскостности и уменьшения высоты ступенек используется суспензия, представляющая собой взвесь в жидкости абразивной составляющей, например окиси алюминия или моноокиси кремния.

2.2.4. Схема упрощенного КМОП-процесса

Укрупненная схема возможного процесса изготовления КМОП ИС приведена на рис. 2.6. Процесс начинается с задания *активных областей*, т.е. областей, в которых будут формироваться транзисторы. Все остальные области кристалла микросхемы будут закрываться толстым слоем двуокиси кремния (SiO_2), называемым *защитным окислом*. Этот слой оксида играет роль изоляции между соседними приборами и либо выращивается (как в процессе, показанном на рис. 2.1), либо осаждается в вытравленные канавки (см. рис. 2.2), — отсюда и название: *изоляция канавками*. Дополнительная изоляция обеспечивается обратно смещенным np-диодом, образующимся в результате введения под защитный окисел дополнительной p^+-области, называемой *противоканальной* (или *охранной*) областью. На следующем этапе методом ионной имплантации формируются слаболегированные p- и n-карманы. Для формирования в p-кармане n-МОП-транзистора ионной имплантацией (или диффузией) в слаболегированной подложке p-типа формируются сильно легированные области истока и стока n-типа. Области стока и истока разделяются тонким слоем SiO_2, называемым *подзатворным окислом*, который сам закрывается проводящим поликристаллическим кремнием (или поликремнием). Проводящий материал образует *затвор* транзистора. МОП-транзисторы с каналом p-типа формируются аналогичным образом (только p- и n-области меняются местами). Для обеспечения электрических соединений между транзисторами поверх этих приборов осаждается несколько разделенных диэлектриком слоев металла (чаще всего это алюминий).

Более детальное разбиение технологического маршрута на операции обработки и результаты их влияния на полупроводниковый материал показаны на рис. 2.7. Хотя в свете изложенного выше смысл большинства операций должен быть понятен из рисунка, относительно отдельных операций все же стоит дать небольшой комментарий. Процесс начинается с p-подложки, на поверхности которой сформирован слаболегированный эпитаксиальный p^--слой (рис. 2.7, *а*). Затем осаждается тонкий слой SiO_2, который будет служить в качестве подзатворного окисла, и поверх него осаждается удаляемый более толстый слой нитрида кремния (рис. 2.7, *б*). С использованием фотошаблона, рисунок на котором представляет собой негатив по отношению к рисунку активных областей, методом плазменного травления формируются канавки, изолирующие приборы друг от друга (рис. 2.7, *в*). После противоканальной имплантации канавки заполняются SiO_2 и выполняется ряд шагов для обеспечения плоскостности поверхности (включая подтравливание окисла с использованием шаблона-негатива слоя активных областей и химико-механическую планаризацию). В этот момент удаляется и вспомогательный слой нитрида (рис. 2.7, *г*). С использованием фотошаблона слоя n-карманов вскрываются области, соответствующие их расположению на кристалле (остальная часть поверхности остается закрытой буферным материалом), после чего выполняется последовательность операций имплантации и отжига, позволяющих получить заданный уровень легирования в кармане. Затем для подгонки величины порогового напряжения p-МОП-транзисторов выполняется вторая имплантация. В ходе данной имплантации

Рис. 2.6. Упрощенная последовательность этапов процесса изготовления КМОП ИС по двухкарманной технологии

изменяется уровень легирования только в той области, которая непосредственно примыкает к подзатворному диэлектрику (рис. 2.7, *д*). Аналогичные операции (только используется другая легирующая примесь) выполняются и для создания p-карманов и подгонки порогового напряжения n-МОП-транзисторов (рис. 2.7, *е*). После этого химически осаждается тонкий слой поликремния и с помощью фотошаблона слоя поликремния в нем создается рисунок. Поликремний используется как в качестве материала электрода затвора транзисторов, так и в качестве среды для создания межсоединений (рис. 2.7, *ж*). Далее последовательно следуют операции ионной имплантации, в ходе которых легируются области истока и стока p-МОП-транзисторов (p^+) и n-МОП-транзисторов (n^+) соответственно (рис. 2.7, *з*), после завершения которых стравливается тонкий подзатворный окисел, незакрытый поликремнием[1]. Эти же операции имплантации используются для легирования поверхности поликремния для уменьшения величины его удельного сопротивления. У нелегированного поликремния удельное сопротивление очень высоко. Обратите внимание, что поликремниевый затвор, конфигурация которого формируется до легирования, фактически задает точное расположение области канала и, следовательно, расположение областей истока и стока. Такая процедура, называемая *процессом самосовмещения*, позволяет добиться

[1] В более современные процессы на этом этапе еще включаются дополнительные операции имплантации для создания слаболегированных областей стока и формирования затворных прокладок. Здесь ради простоты эти операции опущены.

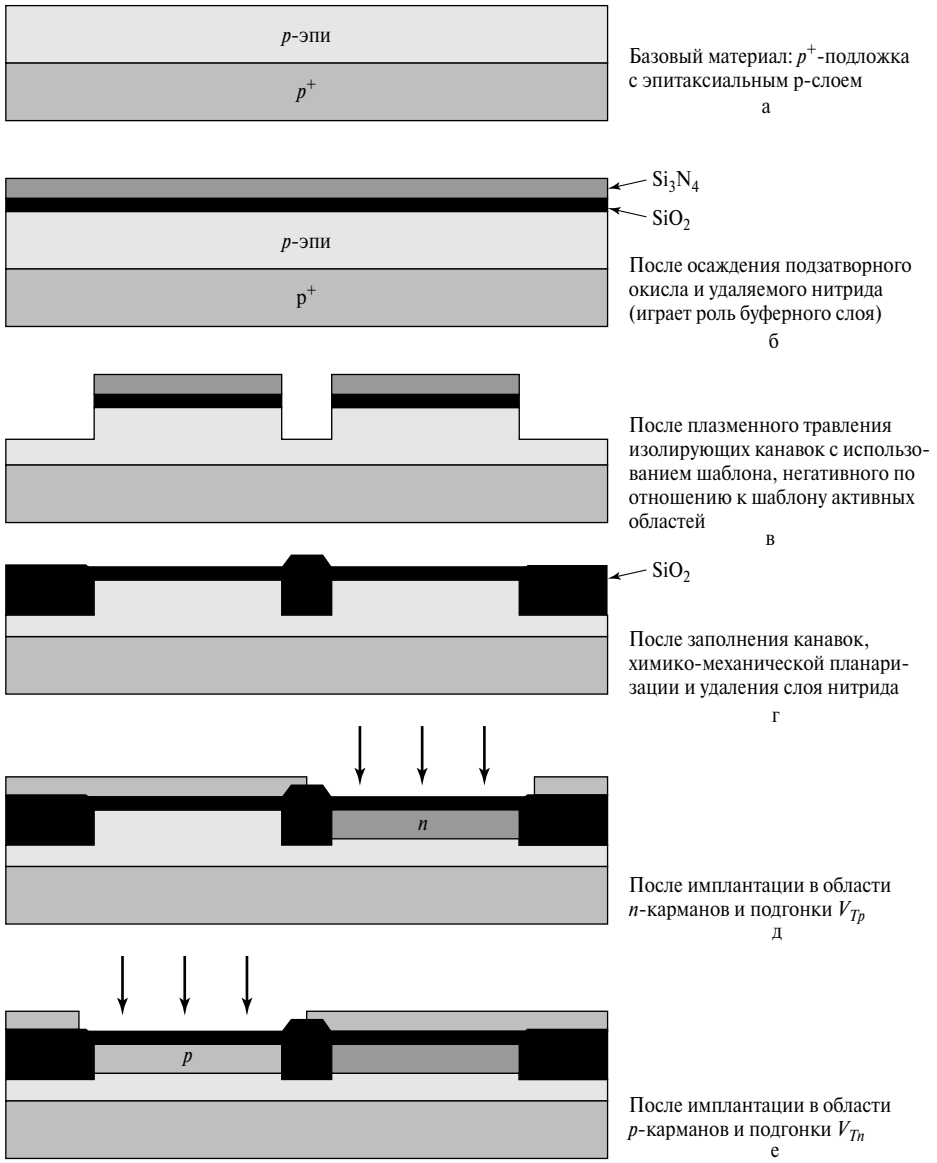


Рис. 2.7. Диаграмма процесса изготовления n-МОП и p-МОП транзисторов по двухкарманной КМОП-технологии. Следует помнить, что рисунки для наглядности стилизованы и характеристические отношения не соответствуют реальности

очень точного позиционирования двух областей относительно затвора. Метод самосовмещения очень полезен в плане снижения паразитных емкостей транзистора. Процесс продолжается осаждением металлических слоев системы межсоединений. На этом этапе повторяется выполнение следующих операций (рис. 2.7, *и–л*): осаждение изолирующего материала (чаще всего это

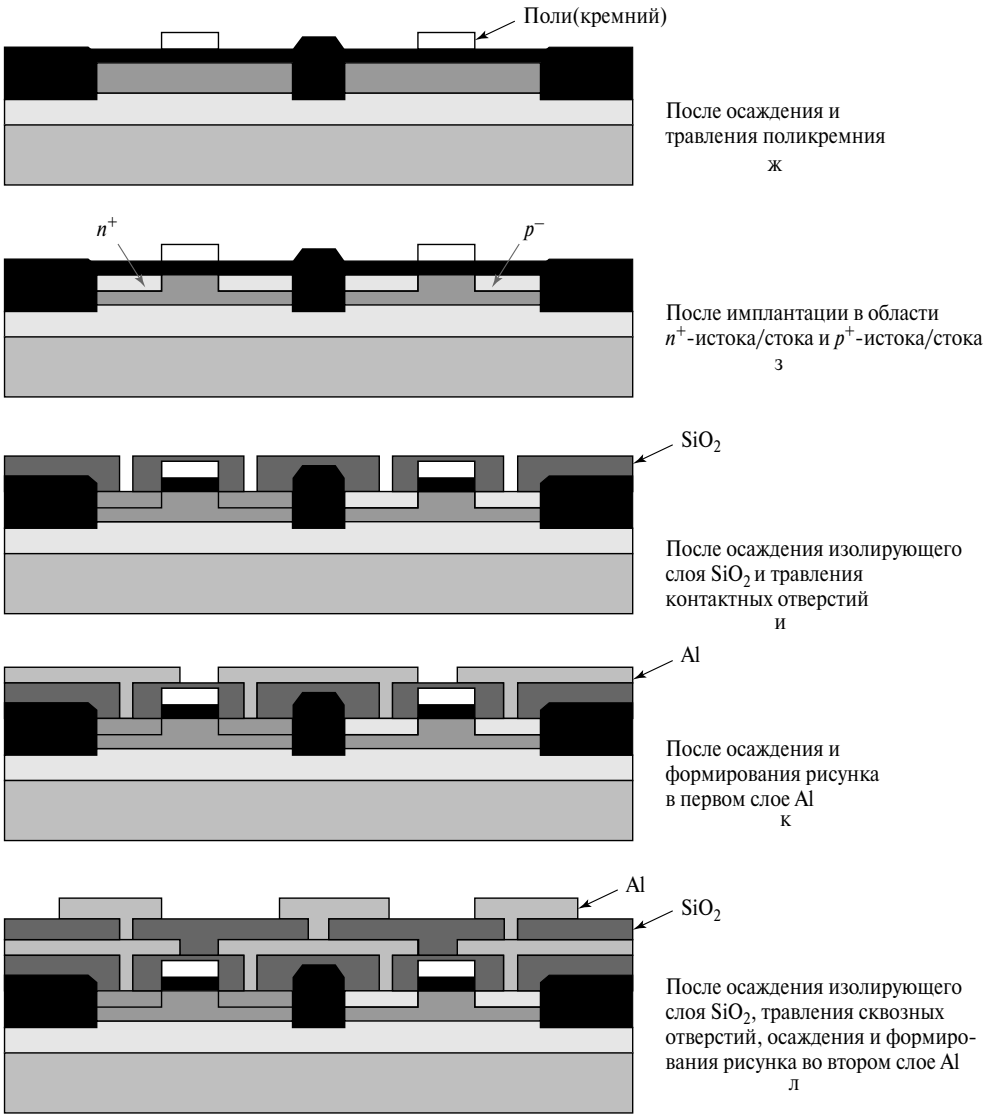


Рис. 2.7. *Окончание*

SiO_2), травление контактных или сквозных отверстий, осаждение металла (чаще всего это алюминий и медь, хотя для нижних слоев часто применяется и вольфрам) и создание рисунка межсоединений. Промежуточные этапы планаризации с использованием *химико-механической полировки* обеспечивают достаточную плоскостность поверхности даже при наличии нескольких слоев межсоединений. После осаждения последнего слоя металла в целях защиты поверхность закрывается слоем пассивации или *остекловывается*. Обычно это слой SiO_2, получаемый методом осаждения из паровой фазы, хотя часто

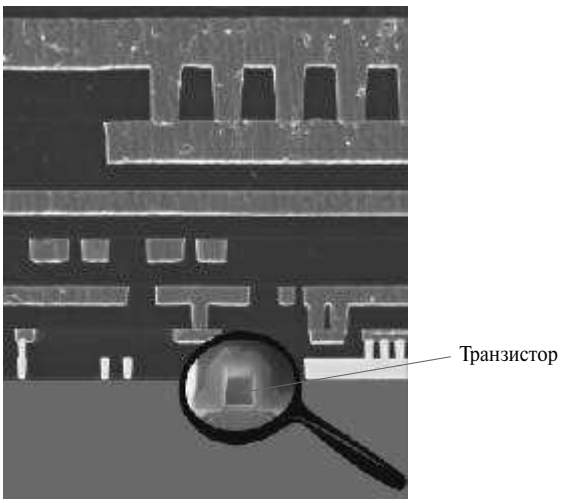


Рис. 2.8. Поперечное сечение структуры, формируемой в ходе современного процесса создания КМОП ИС

еще добавляется слой нитрида кремния, поскольку он более влагонепроницаем. На последнем этапе вытравливаются окна к контактным площадкам, использующимся для термокомпрессионной приварки проволочных выводов.

Поперечное сечение конечного продукта показано на рис. 2.8. Обратите внимание на то, что транзистор занимает только незначительную часть высоты структуры. Бо́льшую часть в вертикальном направлении занимают слои системы межсоединений.

2.3. Нормы проектирования: между разработчиком и технологом

С усложнением процессов требовать от разработчика понимания всех деталей процесса изготовления и умения интерпретировать взаимосвязи между различными масками — это верный путь к неприятностям. Целью задания набора норм проектирования является обеспечение возможности легкого перевода концепции схемы на язык фактических геометрических структур в кремнии. Нормы проектирования выступают в роли интерфейса или даже контракта между разработчиком схемы и технологом.

Разработчики микросхем, как правило, стремятся к получению более плотноупакованной меньшей по размеру разработки, поскольку это дает более высокое быстродействие и более высокую плотность интеграции. Технологи, напротив, хотят иметь воспроизводимый процесс с высоким значением процента выхода годных кристаллов. Следовательно, нормы проектирования представляют собой попытку добиться компромисса.

Нормы проектирования обеспечивают набор указаний для следования при проектировании различных масок, необходимых в процессах формирования рисунка. Они включают в себя ограничения и требования к минимальному размеру и минимальному расстоянию между объектами в пределах одного или в разных слоях.

Основной единицей в определении норм проектирования является *минимальная ширина линии*. Этот термин означает минимальный размер на фотошаблоне, который может быть надежно перенесен на полупроводниковый материал. В общем случае величина минимальной ширины линии определяется разрешающей способностью процесса формирования рисунка, который чаще всего основан на оптической литографии. В более совершенных подходах используются электронно-лучевые, рентгеновские источники и источники глубокого ультрафиолета, каждый из которых обеспечивает более высокое разрешение, но на текущий момент они менее привлекательны с экономической точки зрения.

Даже при одинаковом минимальном размере нормы проектирования обычно отличаются у различных компаний и процессов. Это превращает задачу переноса существующей разработки на другой процесс в весьма трудоемкую процедуру. Одним из подходов к решению этой проблемы является использование совершенных систем автоматизированного проектирования, которые позволяют осуществлять миграцию между совместимыми процессами. Другой подход заключается в использовании *масштабируемых норм проектирования*. В соответствии с последним подходом, известность которому дали работы Мида (Mead) и Конвея (Conway) [Mead80], все нормы задаются в виде функции одного параметра, часто обозначаемого λ. Нормы выбираются таким образом, что разработка с легкостью переносится на весь срез промышленных процессов. Масштабирование минимального размера выполняется путем простого изменения значения параметра λ. Это приводит к *линейному масштабированию* всех размеров. Для данного процесса устанавливается конкретное значение λ, и все конструктивные размеры последовательно переводятся в абсолютные цифры. Обычно величина минимальной ширины линии для процесса устанавливается равной 2λ. Например, для 0,25-микронного процесса (т.е. для процесса с минимальной шириной линии 0,25 мкм) значение λ равно 0,125 мкм.

Данный подход, хотя и является привлекательным, имеет два недостатка.

1. Линейное масштабирование возможно только в ограниченном диапазоне размеров (например, в пределах от 0,25 до 0,18 мкм). При масштабировании в более широком диапазоне отношения между различными слоями обычно изменяются нелинейным образом, что не может быть адекватно отражено при линейном масштабировании норм.
2. Масштабируемые нормы проектирования консервативны: они отражают срез различных технологий и поэтому должны представлять нормы для наихудшего случая всего набора. Вследствие чего имеем завышение размеров и меньшую плотность интеграции разработки.

По этим и другим причинам в промышленности избегают пользоваться масштабируемыми нормами проектирования[2]. Поскольку в промышленных разработках основной целью является плотность упаковки микросхемы, то большинство компаний, выпускающих полупроводниковые устройства, обычно используют *нормы в микронах*, когда нормы проектирования выражаются в абсолютных размерах и поэтому позволяют использовать особенности конкретного процесса в максимальной степени. Масштабирование и перенос разработки с одной технологии на другую при использовании подобных норм представляет собой более сложную задачу и должны выполняться либо вручную, либо с использованием развитых САПР.

В этой книге в качестве основной среды реализации был выбран "добрый, старый" 0,25-микронный КМОП-процесс. Последующая часть данного раздела посвящена краткому введению и обзору норм проектирования для этого процесса, относящихся к классу норм в микронах. Полный набор норм проектирования включает в себя следующие составляющие: набор слоев, отношения между объектами, принадлежащими одному слою, и отношения между объектами в различных слоях. Каждая из них будет рассмотрена по очереди.

Послойное представление

Концепция слоев позволяет перевести неудобоваримый язык комплекта фотошаблонов, использующихся в современном КМОП-процессе, на язык набора концептуальных уровней топологии, которые для разработчика схем значительно нагляднее. С точки зрения разработчика, все КМОП-схемы имеют в основе следующие элементы.

- *Подложка* и/или *карманы*, которые могут быть p-типа (для n-МОП-приборов) и n-типа (для p-МОП).
- *Диффузионные области* (n^+ и p^+), которые задают области, где могут быть сформированы транзисторы. Эти области часто называют *активными областями*. Для создания контактов к карманам или к подложке необходима диффузия примесей разного типа. Такие области называются *выборочными*.
- Один или несколько слоев *поликремния*, которые используются для создания электродов затворов транзисторов (но используются также и в качестве слоев системы межсоединений).
- Ряд *слоев металла* межсоединений.
- Слои *контактов* и *сквозных отверстий*, которые обеспечивают соединение между слоями.

Топология представляет собой комбинацию многоугольников, каждый из которых относится к определенному слою. Функциональность схемы определяется выбором слоев и взаимодействием объектов, принадлежащих различным слоям. Например, МОП-транзистор образуется поперечным сечени-

[2] Хотя это не совсем так, поскольку лямбда-нормы по-прежнему полезны при оценке влияния масштаба технологии на площадь разработки.

ем диффузионного слоя и слоя поликремния. Межсоединение между двумя слоями металла формируется поперечным сечением двух слоев металла и дополнительным слоем контактов. Чтобы сделать эти связи более наглядными, каждому слою назначается стандартный цвет (или тип штриховки при черно-белом отображении).

Внутрислойные ограничения

Первый набор норм задает минимальные размеры объектов в каждом слое, а также минимальные расстояния между объектами в пределах одного слоя. Все расстояния выражаются в микронах.

Межслойные ограничения

Межслойные нормы обычно более сложные. Поскольку тут участвует несколько слоев, то представить их смысл и функциональное значение в наглядной форме труднее. Чтение топологии требует умения переводить двухмерную картину топологического чертежа в трехмерную структуру реального прибора. Это требует некоторой практики.

Приведем эти нормы в виде набора из трех отдельных групп.

1. *Нормы проектирования транзистора.* Транзистор формируется перекрытием слоя активных областей и слоя поликремния. Из внутрислойных норм проектирования уже ясно, что минимальная длина транзистора равна 0,24 мкм (минимальная ширина поликремния), а его ширина составляет по крайней мере 0,3 мкм (минимальная ширина диффузионной области). Дополнительные нормы включают расстояние между активной областью и границей кармана, перекрытие затвором активной области и перекрытие активной областью затвора.
2. *Нормы проектирования контактов и сквозных отверстий.* Контакт (который обеспечивает межсоединение между металлом и активной областью или поликремнием) или сквозное отверстие (которое обеспечивает контакт между двумя металлами) формируется путем перекрытия двух соединяемых слоев и обеспечения между ними контактного отверстия, заполняемого металлом. В рассматриваемом процессе минимальный размер контактного отверстия составляет 0,3 мкм, при этом слои поликремния и диффузии должны выступать за область контактного окна еще по крайней мере на 0,14 мкм. Это приводит к тому, что минимальная площадь контакта устанавливается равной 0,44 мкм $\times$ 0,44 мкм. А это больше, чем минимально возможный размер транзистора! Вот почему при трассировке следует избегать лишних переходов между слоями межсоединений. Дополнительно на рисунке показаны минимальные расстояния между контактами и сквозными отверстиями, а также их соотношения с окружающими слоями.
3. *Нормы проектирования контактов к карманам и подложке.* Чтобы цифровая микросхема была устойчивой к ошибкам, очень важно, чтобы карманы и подложка имели хорошее соединение с источником питания. Если этого не сделать, то может образоваться резистивный путь между контактом к подложке транзистора и шиной питания, что, в свою очередь, может привести к проявлению разрушительных паразитных эффектов, напри-

мер эффекта защелки. Поэтому рекомендуется делать множество контактов к подложке (карману), которые были бы распределены по всей площади области. Для получения омического контакта между шиной питания, реализуемой в слое металла, и материалом p-типа необходимо сформировать диффузионную область p^+-типа. Делается это с помощью *выборочного* слоя, который изменяет тип диффузии.

Рассмотрим процесс с n-карманами, который предусматривает формирование p-МОП-транзисторов в карманах n-типа, создаваемых диффузией в материале p-типа. Номинальный диффузионный слой — это слой p^+. Чтобы изменить полярность диффузии, вводится выборочный n-слой, который обеспечивает формирование диффузионных n^+-областей контактов к n-карману и n^+-областей истока и стока n-МОП-транзисторов в подложке.

Верификация топологии

Фундаментальным требованием к процессу проектирования является обеспечение соблюдения всех без исключения норм проектирования. Если этого не делается, то практически всегда результатом будет нефункционирующая разработка. Но и выполнение этого требования для сложной микросхемы, которая может содержать миллионы транзисторов, задача не из простых, особенно если принять во внимание сложность некоторых наборов норм проектирования. В прошлом целые коллективы разработчиков тратили часы, разглядывая чертежи топологии размером с комнату. Теперь бо́льшую часть этой работы выполняют компьютеры. На сегодня этап *компьютерной проверки соблюдения норм проектирования* является неотъемлемой частью цикла проектирования каждого передаваемого в производство кристалла ИС. Ряд инструментальных средств проектировании топологии даже обеспечивает *интерактивную* проверку соблюдения норм в фоновом режиме непосредственно в процессе разработки концепции.

Пример 2.1. Пример топологии

На рис. 2.9 показан пример законченной топологии инвертора. Чтобы сделать процесс более наглядным, также приведен вид осевого поперечного сечения структуры и электрической принципиальной схемы.

В качестве упражнения читателю предлагается самостоятельно определить размеры n-МОП- и p-МОП-транзисторов.

2.4. Корпусирование интегральных схем

Корпус ИС играет основную роль в работе и обеспечении быстродействия компонента. Кроме осуществления подвода сигналов и напряжения питания к кристаллу микросхемы и от него, он также отводит тепло, генерируемое микросхемой, и выполняет роль механической опоры. Наконец, он еще защищает кристалл от воздействия факторов внешней среды, например влажности.

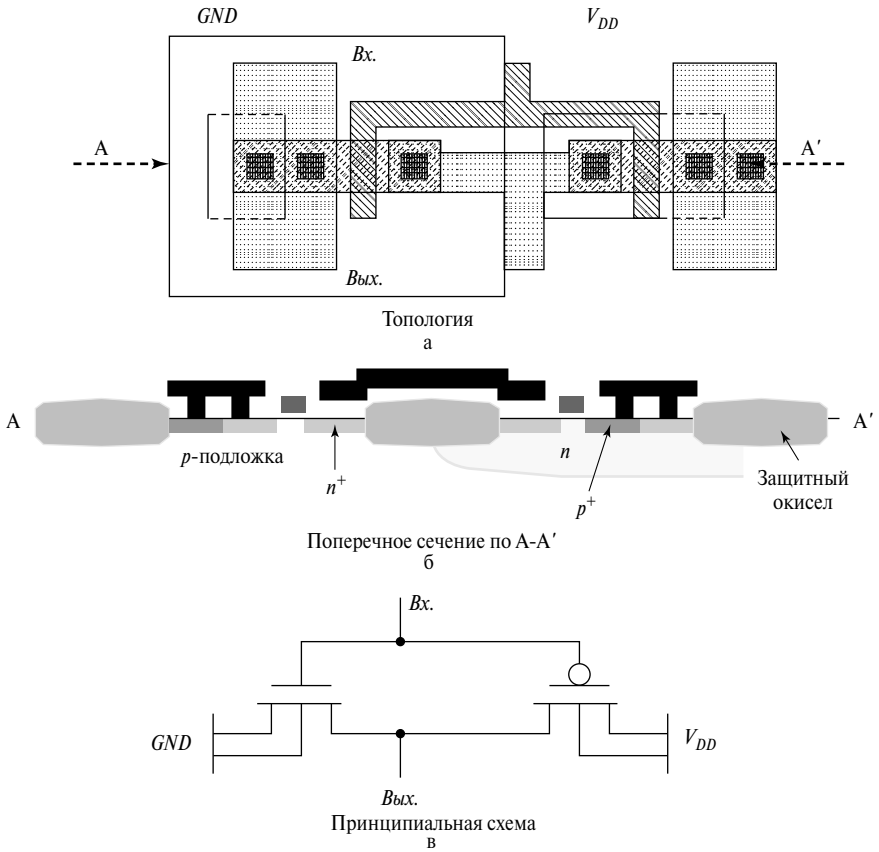


Рис. 2.9. Пример детально прорисованной топологии, включая поперечное сечение структуры и эквивалентной электрической принципиальной схемы

Кроме того, технология корпусирования оказывает существенное влияние на быстродействие и рассеивание мощности микропроцессоров и сигнальных процессоров. Благодаря уменьшению внутренних задержек сигналов и внутрикристальных емкостей из-за масштабирования технологии с течением времени это влияние становится все более заметным. На настоящее время до 50% задержки быстродействующего компьютера обусловливается задержками корпусов, и, как ожидается, эта цифра будет увеличиваться. В последние годы резко увеличился объем исследовательских работ в области создания быстродействующих корпусов с меньшими паразитными емкостями и индуктивностями. Увеличивающаяся сложность систем, размещаемых на одном кристалле, также трансформировалась в проблему увеличения количества ножек ввода-вывода, поскольку количество внешних связей, выходящих от кристалла микросхемы, примерно пропорционально сложности схемы чипа. Эта взаимосвязь впервые была сформулирована сотрудником компании IBM Э. Рентом (E. Rent) (опубликовано в работе [Landman71]) и выражена в ви-

Таблица 2.1. Константы Рента для систем различных классов ([Bakoglu90])

Область применения	β	***K***
Статическая память	0,12	6
Микропроцессор	0,45	0,82
Вентильная матрица	0,5	1,9
Быстродействующих компьютер (однокристальный)	0,63	1,4
Быстродействующий компьютер (одноплатный)	0,25	82

де эмпирической формулы, называемой соответственно *законом Рента*. Эта формула связывает количество ножек ввода-вывода микросхемы с ее сложностью, измеряемой количеством вентилей. Записывается она следующим образом:

$$P = KxG^{\beta}, \tag{2.1}$$

где K — среднее количество каналов ввода-вывода у одного вентиля; G — количество вентилей; β — показатель степени Рента; P — количество ножек ввода-вывода микросхемы. Значение β лежит в пределах от 0,1 до 0,7 и сильно зависит от области применения, архитектуры и организации микросхемы, что подтверждают данные, приведенные в табл. 2.1. Понятно, что микропроцессоры в плане количества ножек ввода-вывода ведут себя совсем не так, как схемы памяти.

Наблюдаемая скорость увеличения количества выводов интегральных схем составляет от 8 до 11% в год, и прогнозируется, что к 2010 году потребуются корпуса с количеством выводов 2000 и более. По этим причинам традиционные корпуса с двухрядным расположением выводов под монтаж в сквозные отверстия были заменены другими подходами, в частности, были разработаны методы поверхностного монтажа, матрицы шариковых выводов и многокристальные модули. Разработчику микросхем полезно знать о возможных вариантах, каждый из которых имеет свои “за” и “против”.

В силу своей многофункциональности хороший корпус должен удовлетворять большому количеству разнообразных требований.

- **Электрические требования**. Выводы должны иметь низкую емкость и индуктивность (как для внутренних связей, так и для наружных). Высокое значение характеристического импеданса должно быть оптимизировано под работу линий передачи. Следует отметить, что собственный импеданс интегральной схемы высок.
- **Механические и тепловые свойства**. Скорость отвода тепла должна быть как можно большей. Механическая надежность требует хорошего согласования тепловых свойств чипа и носителя кристалла. Долгосрочная надежность требует прочного соединения кристалла с корпусом, а также корпуса с платой.
- **Низкая стоимость**. Стоимость является одним из самых важных соображений в любом проекте. Например, хотя керамические корпуса обладают

значительно более высокими характеристиками в сравнении с пластмассовыми, но они также и существенно более дорогие. Увеличение способности корпуса к отбору тепла обычно тоже приводит к увеличению его стоимости. Самый дешевый пластмассовый корпус способен рассеивать до 1 Вт мощности. Несколько более дорогие, но тоже относящиеся к более низкому классу пластмассовые корпуса могут рассеивать до 2 Вт. Более высокие уровни рассеивания требуют применения более дорогих керамических корпусов. Чипы, рассеивающие более 20 Вт, требуют установки радиаторов. Для еще более высоких уровней рассеивания необходимо применение таких экстремальных методов, как установка вентиляторов, воздуховодов, организация водяного охлаждения аппаратуры или использование тепловых трубок.

Главным фактором в уменьшении стоимости платы является плотность установки корпусов. Увеличение количества внешних выводов требует либо увеличения размера корпуса, либо уменьшения шага между выводами. Оба подхода заметно влияют на экономические показатели корпуса.

Корпуса можно классифицировать различными способами: по материалу изготовления, по количеству уровней межсоединений и по способам отвода тепла. В данном коротком разделе будут очень кратко рассмотрены каждый из этих подходов.

2.4.1. Материалы корпусов

Наиболее широко используемые для изготовления корпусов материалы — это керамика и полимеры (пластмассы). Последние обладают тем преимуществом, что существенно дешевле, но обладают ужасными тепловыми свойствами. Например, керамика на основе Al_2O_3 (окиси алюминия) проводит тепло лучше, чем SiO_2 и полиимид, в 30 и 100 раз, соответственно. Кроме того, величина ее коэффициента теплового расширения значительно ближе к величинам коэффициентов теплового расширения наиболее широко применяемых металлов. Недостатком окиси алюминия и других керамик является высокая диэлектрическая проницаемость, приводящая к высоким значениям емкостей соединений.

2.4.2. Уровни межсоединений

Традиционный подход к корпусированию предусматривает использование стратегии двухуровневых межсоединений. Сначала кристалл прикрепляется к индивидуальному носителю кристалла или основанию. Тело корпуса имеет внутреннюю полость, куда устанавливается чип. Эта полость имеет достаточно места для большого количества соединений к внешним выводам (или ножкам) микросхемы. Выводы составляют второй уровень межсоединений и обеспечивают подключение чипа к среде глобальных межсоединений, которой обычно является печатная плата. Сложные системы содержат еще большее количество уровней межсоединений, поскольку платы объединяются

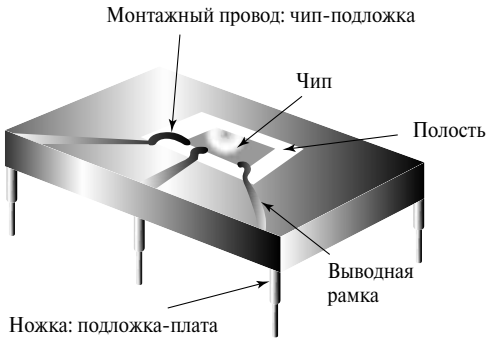


Рис. 2.10. Иерархия межсоединений при традиционном корпусировании ИС

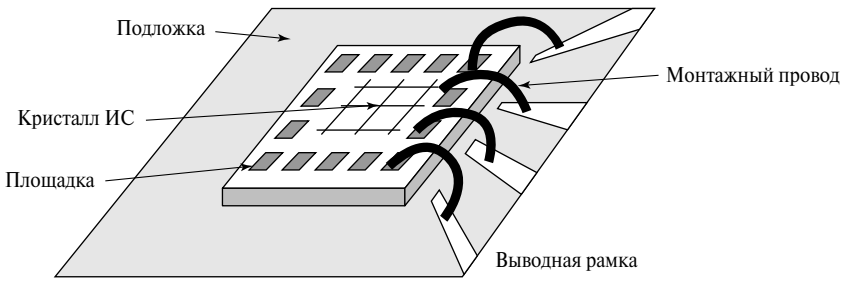


Рис. 2.11. Проволочный монтаж

с помощью задних стенок или плоских кабелей. Первые два уровня в иерархии межсоединений показаны в виде чертежа на рис. 2.10. В последующих разделах приводится краткий обзор методов создания соединений, используемых на первом и втором уровнях иерархии, а также бегло рассматриваются более совершенные подходы к корпусированию.

Первый уровень межсоединений: кристалл–подложка корпуса

Долгое время для обеспечения электрических соединений между кристаллом и корпусом использовался метод *термокомпрессионной сварки микропроволоки*. В соответствии с этим подходом сначала кристалл с помощью теплопроводящего клея прикрепляется обратной стороной к основанию корпуса. На следующем этапе контактные площадки кристалла по отдельности подсоединяются к выводной рамке с помощью алюминиевой или золотой микропроволоки. Работа используемой для этой цели установки термокомпрессионной сборки сильно похожа на работу швейной машинки. Пример термокомпрессионной сборки показан на рис. 2.11. Хотя процесс сборки микропроволокой в значительной степени автоматизирован, он имеет ряд существенных недостатков.

1. Проволочки должны подсоединяться последовательно. При увеличении количества выводов это приводит к увеличению времени производства.

Таблица 2.2. Типовые величины емкости и индуктивности корпусов и методов сборки (по данным из работ [Steidel83], [Franzon93] и [Harper00])

Тип корпуса	*Емкость, пФ*	*Индуктивность, нГн*
68-выводной пластмассовый DIP-корпус	4	35
68-выводной керамический DIP-корпус	7	20
300-выводная матрица шариковых выводов	1–5	2–15
Проволочный монтаж	0,5–1	1–2
Столбик припоя	0,1–0,5	0,01–0,1

2. С увеличением количества выводов очень не просто найти способ размещения проволочек, который бы исключал их перемыкание.
3. Из-за промышленного подхода и нерегулярности топологии сложно спрогнозировать величины паразитных емкостей и индуктивностей.

Микропроволочки обладают ужасными электрическими свойствами, например, высокой собственной индуктивностью (5 нГн и более) и взаимной индуктивностью между путями соседних кристаллов. Индуктивность проволочки обычно составляет около 1 нГн/мм, тогда как индуктивность одного вывода корпуса лежит в диапазоне от 7 до 40 нГн, что зависит от типа корпуса и расположения вывода относительно края корпуса [Steidel83]. Типовые значения паразитных емкости и индуктивности для ряда широко используемых корпусов приведены в табл. 2.2.

Из-за упомянутых недостатков исследовались и альтернативные методы сборки. В соответствии с одним из подходов, называемым *монтажом на гибкий носитель*, кристалл соединяется с металлической выводной рамкой, которая вытравливается на поверхности полимерной пленки, обычно из полиимида (рис. 2.12, *а*). Соединение контактных площадок чипа и разводки на полиимидной пленке осуществляется с помощью столбиков припоя (см. рис. 2.12, *б*). После этого лента может присоединяться к телу корпуса с использованием ряда различных методов. Один из возможных подходов может быть основан на использовании разъемов нажимного действия.

К преимуществам процесса сборки на гибкий носитель можно отнести его высокую степень автоматизированности. Для автоматического перемещения пленки используется перфорация. Все соединения формируются одновременно. Литографический подход к формированию разводки на пленке помогает получать малый шаг, что позволяет увеличивать количество выводов. Исключение длинных проволочных соединений улучшает электрические характеристики. Например, смонтированная на 48-миллиметровой ленте с двумя уровнями металла микросхема имеет следующие электрические характеристики: $L = 0,3$–$0,5$ нГн, $C = 0,2$–$0,3$ пФ и $R = 50$–200 Ом [Doane93].

Еще один подход заключается в том, что кристалл переворачивается лицевой частью вниз и присоединяется непосредственно к основанию корпуса с помощью столбиков припоя. Преимуществом этого метода, называемого монтажом *методом перевернутого кристалла*, являются превосходные электрические характеристики (рис. 2.13). Кроме того, вместо создания всех со-

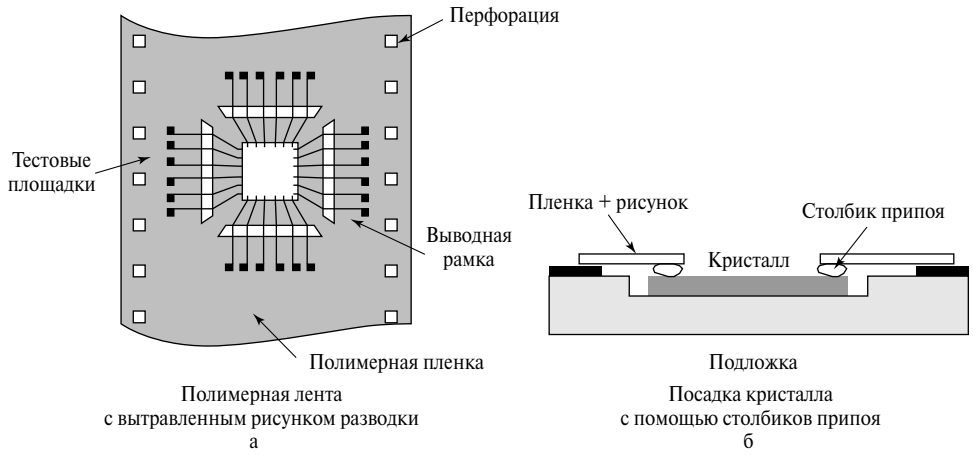


Рис. 2.12. Метод сборки на гибкий носитель

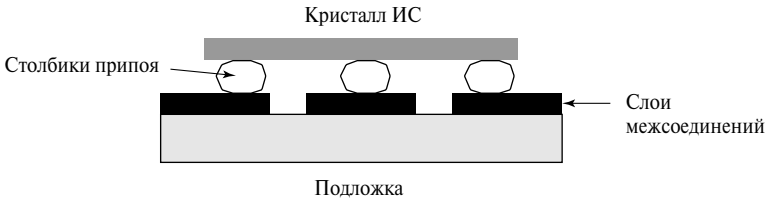
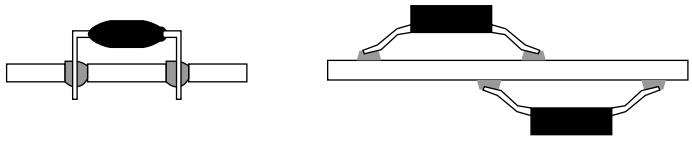


Рис. 2.13. Сборка методом перевернутого кристалла

единений ввода-вывода по краю кристалла этот метод позволяет размещать контактные площадки в любом месте чипа, что может помочь справиться с проблемой распределения питания и сигнала синхронизации, так как качество материалов разводки на основании (подложке) (например, Cu или Au) обычно выше, чем Al на чипе.

Второй уровень межсоединений: основание корпуса–плата

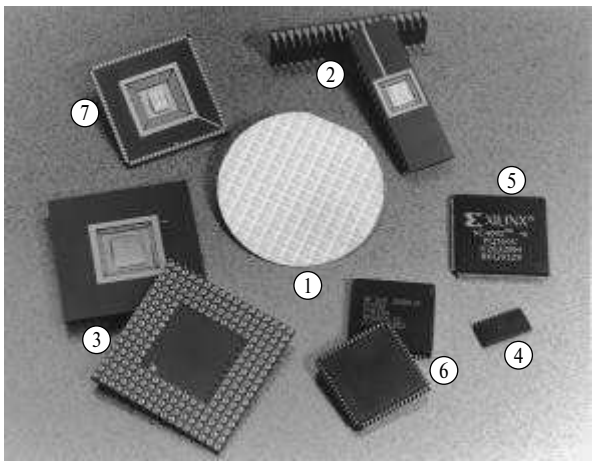
Для обеспечения соединения корпуса с платой долгое время основным методом был *монтаж с установкой в сквозные отверстия*. Печатная плата изготавливается путем объединения в стопку слоев меди и стеклотекстолита. При монтаже методом установки в сквозные отверстия, в плате просверливаются отверстия и с помощью процесса анодного осаждения покрываются медью. Ножки корпуса вставляются в эти отверстия, после чего электрическое соединение обеспечивается припоем (рис. 2.14, *а*). Самым предпочтительным корпусом в этом классе был корпус с двухрядным расположением выводов или DIP-корпус, показанный на рис. 2.15 под номером 2. Но когда количество выводов становится больше 64, плотность установки DIP-корпусов быстро ухудшается. Эта проблема частично решается при использовании корпусов типа *"матрица выводов"* (Pin Grid Array — PGA), у которых ножки располагаются не только по периферии, но и по всей нижней поверхности



Монтаж в сквозные отверстия
а

Поверхностный монтаж
б

Рис. 2.14. Методы установки корпусов на плату



1 – голый кристалл ИС
2 – DIP-корпус
3 – матрица выводов
4 – малогабаритный корпус
5 – плоский квадратный корпус
6 – пластмассовый носитель кристалла
7 – безвыводной носитель кристалла

Рис. 2.15. Типы широко используемых корпусов

(см. рис. 2.15, п. 3). Матрицы выводов позволяют перейти в область больших количеств ножек (возможно получение более 400 выводов).

Подход на основе монтажа в сквозные отверстия обеспечивает механически надежное и прочное соединение. Однако дается это ценой плотности компоновки. По механическим соображениям сквозные отверстия не могут располагаться с шагом меньше 2,54 мм. Но даже при этом матрицы выводов с большим количеством ножек обычно существенно ослабляют плату. Кроме того, сквозные отверстия ограничивают плотность компоновки на плате, блокируя дорожки, которые в ином случае могли бы проходить под ними, что приводит к удлинению межсоединений. Поэтому матрицы выводов с большим количеством ножек требуют введения дополнительных слоев разводки для обеспечения связей с каждой из этой массы ножек. Наконец, хотя паразитные емкость и индуктивность у матрицы выводов немного ниже, чем у DIP-корпуса, они по-прежнему еще велики.

Многие из недостатков установки в сквозные отверстия устраняются при использовании *метода поверхностного монтажа*. Чип устанавливается на поверхность платы методом пайки, и при этом не требуется наличия каких-либо сквозных отверстий (см. рис. 2.14, *б*). Плотность компоновки увеличивается, так как исключается наличие сквозных отверстий, что обеспечивает больше места для разводки; уменьшается шаг выводов и чипы могут уста-

Таблица 2.3. Параметры носителей кристаллов различных типов

Тип корпуса	*Расстояние между выводами (типовое)*	*Количество выводов (максимальное)*
С двухрядным расположением выводов	2,54 мм	64
Матрица выводов	2,54 мм	больше 300
Миниатюрный корпус ИС	1,27 мм	28
Носитель кристалла с выводами	1,27 мм	124
Безвыводной носитель кристалла	0,75 мм	124

навливаться на обе стороны платы. Кроме того, устранение сквозных отверстий улучшает механическую прочность платы. К негативным последствиям можно отнести то, что поверхностное соединение делает соединение чипа с платой более слабым. Установка компонента на плату становится не только достаточно сложным занятием, но и требует применения более дорогостоящего оборудования, так как простого паяльника теперь уже недостаточно. Наконец, тестирование платы усложняется, поскольку ножки корпуса теперь недоступны с обратной стороны платы. Съем сигналов становится затруднительным или почти невозможным.

В настоящее время находит применение большое количество корпусов под поверхностный монтаж с различным количеством и шагом выводов. Три типа из них изображены на рис. 2.15: *миниатюрный корпус* с выводами типа крыло чайки, *пластмассовый носитель кристалла* с J-выводами и *безвыводной носитель кристалла*. Некоторые данные о наиболее важных параметрах ряда корпусов приведены в табл. 2.3.

Но даже поверхностный монтаж не способен удовлетворить потребность в еще большем количестве выводов. Дело усугубляется из-за увеличения количества ножек для подвода питания: работа на низких напряжениях питания требует использования такого же количества ножек питания и земли, как и для ввода-вывода сигналов! Когда количество выводов превышает 300, то предпочтительной средой межсоединений между корпусом и платой становятся шариковые выводы из припоя, заменяющие собой ножки. Пример реализации такого подхода к корпусированию показан на рис. 2.16. Подобный корпус получил название керамической *матрицы шариковых выводов* (Ball Grid Array — BGA). Столбики припоя используются для обеспечения электрического контакта как кристалла к основанию корпуса, так и корпуса к плате. Матричное расположение соединений обеспечивает постоянство плотности ножек ввода-вывода вне зависимости от общего количества выводов у корпуса. Минимальная величина шага между столбиками припоя может достигать 0,8 мм, так что вполне реально получение корпуса, имеющего несколько тысяч входов-выходов.

Многокристальные модули: кристалл–плата

Многоуровневая иерархия межсоединений в корпусе становится неприемлемой в современных сложных разработках, что обусловливается их высокой степенью интеграции, большим количеством сигналов и растущими требованиями к быстродействию. Поэтому основное направление развития состоит

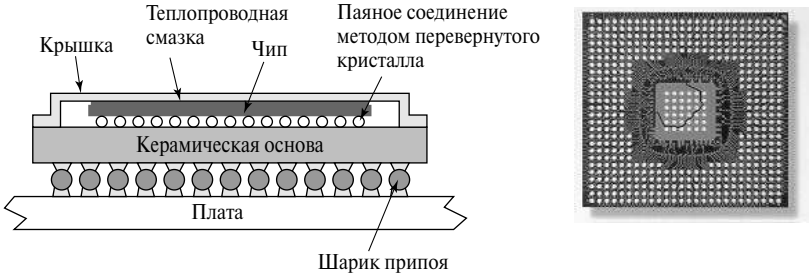


Рис. 2.16. Корпусирование с использованием матрицы шариковых выводов: а) поперечное сечение; б) вид корпуса снизу

в уменьшении количества уровней. На сегодня основное внимание уделяется методам корпусирования, позволяющим исключить первый уровень иерархии. Исключение одного уровня в иерархии корпусирования за счет установки кристаллов микросхем непосредственно на коммутационную плату, будь то печатная плата или специальная подложка, обеспечивает существенный выигрыш в тех случаях, когда ключевым параметром разработки является быстродействие или плотность упаковки. Подобный подход к сборке получил название *метод многокристальных модулей* (МКМ). Он позволяет значительно увеличить плотность компоновки и поднять уровень общего быстродействия системы.

Для непосредственной установки кристаллов микросхем на подложку могут использоваться уже упомянутые ранее методы сборки, включая монтаж методом термокомпрессионной сварки, монтаж на гибкий носитель и метод перевернутого кристалла, хотя применение последних двух предпочтительнее. Сама подложка в зависимости от требуемых механических, электрических, тепловых и экономических параметров может изготавливаться из разнообразных материалов. Это могут быть текстолитовые подложки (подобные печатным платам), металлические, керамические и кремниевые. Подложки из кремния выгодно отличает то, что они обеспечивают идеальное согласование механических и тепловых свойств с материалом кристалла микросхемы.

Основными преимуществами МКМ является увеличение плотности компоновки и быстродействия. Пример многокристального модуля, реализованного на кремниевой подложке (обычно такой подход называют *кремний-на-кремнии*), показан на рис. 2.17. Этот модуль, представляющий собой процессор для авиационного оборудования и изготавливаемый компанией Rockwell International, содержит 53 ИС и 40 дискретных компонентов, установленных на подложке с размерами $2,2'' \times 2,2''$ и алюминий-полиимидной системой межсоединений. Поскольку в данном случае для создания рисунка на подложке используются те же методы, что и при изготовлении ИС, то ширина дорожек межсоединений на подложке всего на порядок величины больше, чем на самих кристаллах микросхем. Модуль имеет 180 внешних выводов. Быстродействие увеличивается за счет исключения уровня кристалл–носитель с присущими ему паразитными связями, а также за счет уменьшения длины

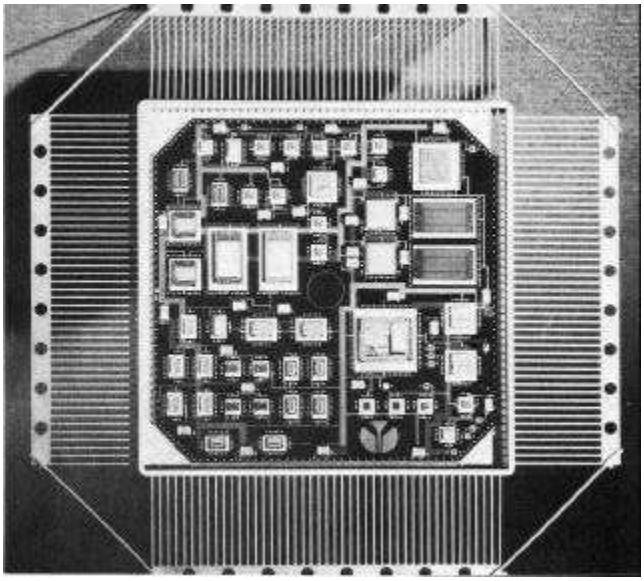


Рис. 2.17. Модуль процессора для авиационной аппаратуры. Перепечатано с разрешения компании Rockwell International, Inc.

глобальных коммутационных дорожек связи, что является результатом увеличения плотности компоновки. Например, собственные емкость и индуктивность столбика припоя равны всего 0,1 пФ и 0,001 нГн соответственно. Технология МКМ также позволяет существенно уменьшить потребляемую мощность, поскольку благодаря снижению величины емкостной нагрузки становятся излишними мощные выходные каскады и связанная с ними рассеиваемая мощность. Одновременно уменьшается и динамическая рассеиваемая мощность, расходуемая на переключение больших емкостей нагрузки.

Хотя технология МКМ и дает ряд очевидных преимуществ, все же основным ее недостатком являются плохие экономические показатели. Эта технология требует выполнения ряда сложных операций, что делает процесс дорогим. Данный подход до недавнего времени был оправдан только в тех случаях, когда важным является либо получение высокой плотности компоновки, либо достижение рекордных показателей быстродействия. В последние годы влияние экономики становится не столь определяющим, и многокристальные сборки нашли свое место в нескольких дешевых быстродействующих изделиях. Подобная тенденция получила название стратегии *система в корпусе* (system-in-a-package — SIP).

2.4.3. Учет тепловыделения при корпусировании

С увеличением потребляемой мощности интегральных схем становится все более важным эффективный отвод генерируемого чипами тепла. Большое количество механизмов возникновения отказов в ИС становятся более актив-

ными при повышенных температурах. Примерами таковых являются токи утечки в обратно смещенных диодах, электромиграция и захват ловушками горячих электронов. Для предотвращения отказов температура кристалла должна оставаться в определенных пределах. Скажем, коммерческие приборы поддерживают рабочий температурный диапазон от 0° до 70°C. Детали военного назначения более требовательны и требуют обеспечения работоспособности в температурном диапазоне от −55° до +125°C.

Эффективность охлаждения корпуса зависит от теплопроводности (сопротивления) материала корпуса, состоящего из основания и тела, структуры корпуса и эффективности теплопередачи от корпуса в охлаждающую среду. При стандартном подходе в качестве среды охлаждения используется неподвижный или циркулирующий воздух. Эффективность теплопередачи может быть улучшена путем добавления к корпусу металлического теплоотвода с ребрами. При использовании более сложных подходов к компоновке аппаратуры, используемых, например, в мэйнфреймах или суперкомпьютерах, для достижения еще более высокой эффективности охлаждения сквозь тонкие каналы в корпусе принудительно прокачивается воздух, жидкость или инертный газ.

Зная тепловое сопротивление корпуса Θ, выражаемое в °C/Вт, можно легко получить температуру корпуса чипа, если воспользоваться *уравнением теплового потока*:

$$\Delta T = T_{chip} - T_{env} = \Theta Q, \tag{2.2}$$

где T_{chip} и T_{env} — температура чипа и окружающей среды соответственно. Q представляет собой величину потока тепла (в ваттах). Заметьте, насколько близко уравнение теплового потока напоминает закон Ома. Тепловой поток и разность температур эквивалентны току и разности напряжений соответственно. Тепловое моделирование чипа, его корпуса и окружения — задача очень сложная. Более подробное обсуждение данной темы читатель может найти в [Lau98, глава 3].

Пример 2.2. Теплопроводность корпуса

Возьмем в качестве примера 40-выводной DIP-корпус, который имеет тепловое сопротивление 38° и 25°C/Вт при естественном конвекционном и принудительном воздушном охлаждении соответственно. Это означает, что DIP-корпус способен рассеивать 2 ватта (3 ватта) мощности при естественном (принудительном) воздушном охлаждении и поддерживать при этом разность температур между кристаллом и окружающей средой ниже 75°C. Для сравнения, величина теплового сопротивления корпусов типа “матрица выводов” лежит в диапазоне от 15 до 30°C/Вт.

Поскольку методы корпусирования, обеспечивающие пониженное тепловое сопротивление, неприемлемо дороги, то удержание рассеиваемой интегральной схемой мощности в определенных рамках является экономической

необходимостью. Увеличение уровней интеграции и быстродействия превращает эту задачу в далеко не тривиальную. В этом контексте интересное соотношение было выведено Нагатой [Nagata92]. Оно позволяет определять границу степени интеграции и уровня быстродействия в функции тепловых параметров. Можно записать:

$$\frac{N_G}{t_p} \leq \frac{\Delta T}{\Theta E}, \tag{2.3}$$

где N_G — количество вентилей в кристалле микросхемы; t_p — задержка распространения; ΔT — максимальная разность температур кристалла и окружающей среды; Θ — тепловое сопротивление между ними; E — энергия переключения каждого вентиля.

Пример 2.3. Тепловые границы степени интеграции

Для $\Delta T = 100°\text{C}$, $\Theta = 2,5°$ C/Вт и $E = 0,1$ пДж выражение (2.1) дает следующий результат: $N_G/t_p \leq 4 \times 10^5$ вентилей/нс. Другими словами, в случае, если все вентили переключаются одновременно и имеют время переключения 1 нс, то максимальное количество вентилей на кристалле микросхемы должно быть менее 400 000. Это эквивалентно рассеиваемой мощности 40 Вт.

К счастью, в реальных системах не все вентили переключаются одновременно. С учетом характера работы в конкретной схеме максимальное количество вентилей может быть значительно больше. Например, экспериментально было найдено, что отношение средней величины периода переключения к задержке распространения в мини-компьютерах и больших ЭВМ лежит в пределах от 20 до 200 [Masaki92].

Тем не менее выражение (2.1) демонстрирует, что рассеивание тепла и тепловые параметры накладывают на величину степени интеграции существенные ограничения. Поэтому все бо́льшую значимость приобретают такие подходы к проектированию, которые обеспечивают малое потребление либо за счет уменьшения величины E, либо за счет снижения коэффициента использования.

2.5. Тенденции в технологии интегральных схем

Современный технологический процесс изготовления КМОП ИС достаточно близко соответствует диаграмме, описанной в предыдущем разделе, хотя некоторые шаги могут меняться местами, может использоваться однокарманный подход, вместо заполнения канавок защитный окисел может выращиваться, а также могут вводиться и дополнительные шаги, например формирование низколегированной области стока. Также с целью увеличения проводимости широко распространенной практикой стало покрытие поликремниевых межсоединений и областей истока и стока *силицидами*, например, $TiSi_2$

(см. рис. 2.2). Данная дополнительная операция вводится между шагами (и) и (к) описанного процесса. В настоящее время уже готовятся к внедрению или близки к этому ряд важных модификаций или усовершенствований процесса, которые заслуживают внимания. Однако, кроме них, в ближайшее десятилетие никаких коренных изменений в описанной технологии КМОП ИС не ожидается.

2.5.1. Краткосрочная перспектива

Использование меди и диэлектриков с низким значением диэлектрической проницаемости

В данной книге постоянным рефреном будет проходить тема роста влияния межсоединений на общее быстродействие проектируемой микросхемы. Технологи постоянно оценивают возможные альтернативы традиционной комбинации алюминий-SiO_2 в качестве проводника и диэлектрика, которая была нормой в последние несколько десятков лет. В 1998 году инженеры из компании IBM наконец продемонстрировали подход, который сделал использование меди в качестве материала для межсоединений в КМОП-процессе жизнеспособным и экономически целесообразным [Geppert98]. Медь обладает удельным сопротивлением, которое значительно ниже, чем у алюминия. Недостатком ее является то, что она легко диффундирует в кремний, вызывая ухудшение характеристик приборов. Покрытие меди слоем буферного материала, например нитрида титана, который предотвращает диффузию, решает эту проблему, но требует специального процесса осаждения. В представленном компанией IBM так называемом *процессе двойной инкрустации* (Dual Damascene process) (рис. 2.18) для металлизации используется подход, в соответствии с которым металлом заполняются вытравленные в диэлектрике канавки, после чего выполняется химико-механическая полировка. Это отличается от традиционного подхода, когда сначала осаждается сплошной слой металла, а затем лишние участки удаляются путем травления.

В дополнение к межсоединениям из материалов с более низким удельным сопротивлением, начиная с поколения 0,18-мкм КМОП-процессов, в промышленном производстве также стали использоваться изолирующие материалы с диэлектрической проницаемостью и, следовательно, емкостью ниже, чем у SiO_2.

Кремний на диэлектрике

Хотя он уже и известен достаточно давно, но, похоже, сейчас есть неплохие шансы, что описанный в предыдущем разделе традиционный КМОП-процесс (также известный под названием *объемный* КМОП-процесс) может быть заменен КМОП-процессом *кремний на диэлектрике*. Главное его отличие заключается в исходном материале: транзисторы, произведенные по технологии кремний на диэлектрике (КНД), формируются в очень током слое кремния, осажденного поверх толстого слоя изолирующего SiO_2 (рис. 2.19). Основными преимуществами КНД-процесса являются уменьшение паразитных связей и лучшие переключательные характеристики транзистора. Исследователями из компании IBM было продемонстрировано, что простой пере-

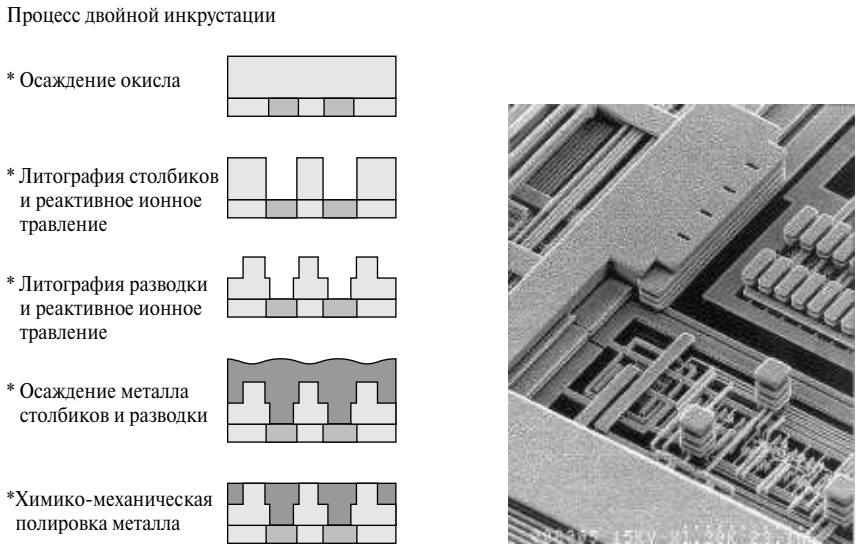


Рис. 2.18. Процесс инкрустации (из [Gepper98]): а) этапы процесса; б) микрофотография межсоединений после удаления диэлектрика

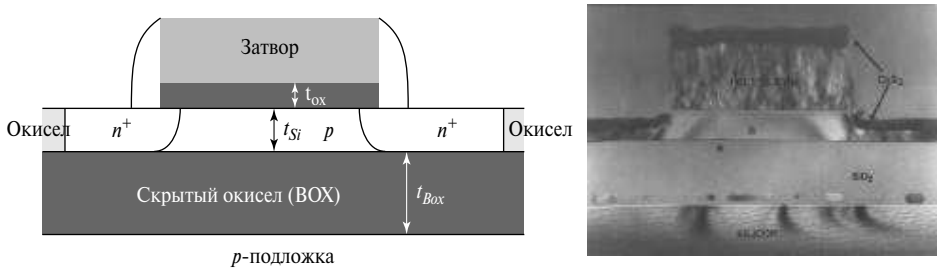


Рис. 2.19. Процесс “кремний на диэлектрике”: а) схематическое изображение структуры; б) полученный с помощью сканирующего электронного микроскопа вид поперечного сечения [Eaglesham99]

вод микросхемы с традиционного объемного КМОП-процесса изготовления на КНД-процесс при сохранении всех других конструктивных и технологических параметров, к примеру длины канала и толщины окисла, увеличивает быстродействие на 22% [Allen99]. Долгое время изготовление КНД-подложек высокого качества по экономически целесообразной цене было основным препятствием на пути крупномасштабного внедрения этого процесса. Но к концу 1990-х годов картина изменилась, и теперь КНД-процесс уверенно движется к завоеванию главенствующего положения.

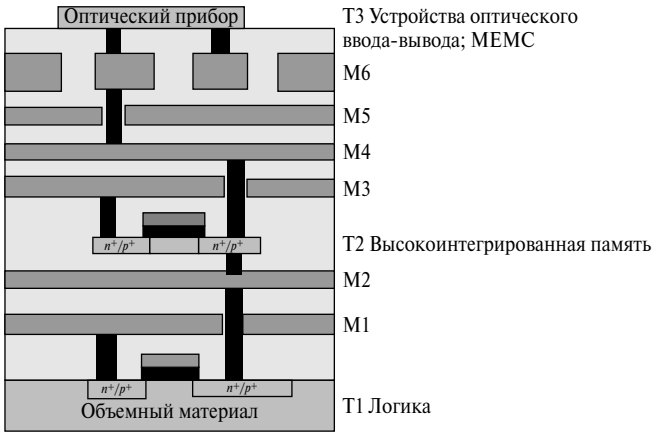


Рис. 2.20. Пример истинной трехмерной интеграции. Дополнительные активные слои (Т*), в которых реализуются высокоинтегрированная память и устройства ввода-вывода, чередуются со слоями металлических межсоединений (М*)

2.5.2. Долгосрочная перспектива

Продление жизни КМОП-технологии далее следующего десятилетия и переход к длинам канала менее 100 нм потребует переработки технологического процесса и структуры прибора. Уже сейчас мы становимся свидетелями появления широкого диапазона новых приборов (например, органических транзисторов, молекулярных ключей и квантовых приборов). Хотя на данный момент сложно делать прогнозы, какой из подходов будет доминирующим в следующем веке, все же об одной интересной разработке стоит упомянуть.

Истинно трехмерные интегральные схемы

Подача и вывод сигналов из вычислительных элементов в заданные моменты времени представляют собой одну из основных проблем, вызываемых продолжающимся ростом степени интеграции. Одним из возможных способов ее решения является введение дополнительных активных слоев и размещение их между слоями металлических межсоединений (рис. 2.20). Это позволит разместить высокоинтегрированную память непосредственно над логическим процессором, реализованном по объемной КМОП-технологии, что уменьшит расстояние между вычислительным устройством и устройством хранения и таким образом снизит величину задержки [Souri00]. Кроме того, приборы с различными рабочими напряжениями, быстродействием и материалом подложки могут размещаться в разных слоях. Например, место над активным слоем может быть зарезервировано за оптическими приемо-передатчиками, которые способны решить проблему ввода-вывода, или за микроэлектро-механическими системами (МЕМС), способными выполнять функции датчиков или радиоинтерфейса.

Хотя этот подход и выглядит многообещающе, чтобы сделать его по-настоящему жизнеспособным, необходимо еще решить ряд проблем и преодолеть несколько фундаментальных препятствий. Как отводить выделяемое

тепло — это один из наиболее насущных вопросов; получение гарантированного выхода годных кристаллов — второй. Исследователи демонстрируют значительный прогресс в этих вопросах, так что 3D-интеграция, возможно, уже не за горами. Но до получения окончательного решения можно использовать некоторые промежуточные подходы. Одной из альтернатив является так называемая *2,5-мерная интеграция*. Суть подхода заключается в соединении двух прошедших полную обработку пластин, на которых схемы изготавливаются на поверхности таким образом, что отдельные кристаллы перекрывают друг друга. Для обеспечения электрического соединения двух чипов после создания металлизации протравливаются сквозные отверстия. К преимуществам такой технологии можно отнести подобность электрических свойств приборов во всех активных слоях и независимость от температур обработки, поскольку чипы могут изготавливаться отдельно и только потом объединяться. Главным недостатком этой технологии является отсутствие точности (в наилучшем случае точность совмещения составляет ± 2 мкм), что ограничивает межчиповые связи только шинами глобальных межсоединений.

Из данного футуристического прогноза ясно видится, что границы между чипом, подложкой, корпусом и платой размываются. Так что разработчикам таких *систем на кристалле* или *систем в корпусе* придется учитывать все эти аспекты одновременно.

2.6. Резюме

В данной главе вид процессов изготовления и корпусирования КМОП ИС представлен с "высоты птичьего полета".

- Процесс изготовления интегральных схем требует выполнения многих шагов, каждый из которых состоит из последовательности базовых операций. Ряд таких шагов и/или операций, например экспонирование и проявление при фотолитографии, осаждение материалов и травление, весьма часто повторяются в ходе технологического процесса.
- *Оптические маски*, или *фотошаблоны*, представляют собой центральный интерфейс между особенностями процесса изготовления и разработкой, которую пользователь хочет видеть воплощенной в виде кремниевой структуры.
- *Набор норм проектирования* определяет ограничения в терминах минимальной ширины и расстояний между топологическими элементами, которых следует придерживаться, чтобы проектируемая микросхема была полностью функционирующей. Нормы проектирования выступают в роли контракта между разработчиком схемы и технологом.
- Корпус представляет собой интерфейс между схемой, реализованной в кремниевом кристалле, и внешним миром и, будучи таковым, оказывает существенное влияние на быстродействие, надежность, долговечность и стоимость интегральной схемы.

2.7. Для любознательных

За последние несколько десятилетий было выпущено множество книг, посвященных изготовлению полупроводниковых приборов. Прекрасный обзор текущего положения дел в области изготовления КМОП ИС содержится в книге *Технология кремниевых СБИС* Дж. Пламмера, М. Дила и П. Гриффина (*Silicon VLSI Technology*, J. Plummer, M. Deal, P. Griffin) [Plummer00]. Наглядный обзор различных шагов процесса изготовления можно найти в [Fullman99]. Другими источниками информации являются такие периодические издания, как *IEEE Transactions on Electron Devices* и *Technical Digest of IEDM conference*. Имеется и ряд великолепных источников подробной информации о современных методах корпусирования. В качестве примеров можно привести работы [Doan93], [Harper00] и [Lau98].

2.7.1. Список использованной литературы

[Allen99] D. Allen et al., "A 0.2 μm 1.8 V SOI 550 MHz PowerPC Microprocessor with Copper Interconnects", *Proceedings IEEE ISSCC Conference*, vol. XLII, p. 438-439, February 1999.

[Bakoglu90] H. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, Addicon-Wesley, 1990.

[Doane93] D. Doan, ed., *Multichip Module Technologies and Alternatives*, Van Nostrand-Reinhold, 1993.

[Eaglesham99] D. Eaglesham, "0.18 μm CMOS and Beyond", *Proceedings 1999 Design Automation Conference*, p. 703–708, June 1999.

[Franzon93] P. Franzon, "Electrical Design of Digital Multichip Modules", in [Doane93], p. 525–568, 1993.

[Fullman99] Fullman Kinetics, "The Semiconductor Manufacturing Process", *http://www.fullman-kinetics.com/semiconductors/semiconductors.html*, 1999.

[Geppert98] L. Geppert, "Technology — 1998 Analysis and Forecast", *IEEE Spectrum*, vol. 35, no. 1, p.23, January 1998.

[Harper00] C. Harper, ed., *Electronic Packaging and Interconnection Handbook*, McGraw-Hill, 2000.

[Landman71] b. Landman and R. Russo, "On a Pin versus Block Relationship for Partitions of Logic Graphs", *IEEE Trans. on Computers*, vol. C-20, p. 1469–1479, December 1971.

[Lau98] J. Lau et al., *Electronic Packaging — Design, Materials, Process, and Reliability*, McGraw-Hill, 1998.

[Masaki92] A. Masaki, "Deep-Submicron CMOS Warms Up to High-Speed Logic", *Circuits and Devices Magazine*, November 1992.

[Mead80] C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, 1980.

[Nagata92] M. Nagata, "Limitations, Innovations, and Challenges of Circuits and Devices into a Half Micrometer and Beyond", *IEEE Journal of Solid State Circuits*, vol. 27, no. 4, p. 465–472, April 1992.

[Plummer00] J. Plummer, M. Deal, and P. Griffin, *Silicon VLSI Technology*, Prentice Hall, 2000.

[Steidel83] C. Steidel, *Assembly Techniques and Packaging*, in [Sze83], p. 551–598, 1983.

[Souri00] S.J. Souri, K. Banerjee, A. Mehrota and K.C. Saraswat, "Multiple Si Layer ICs: Motivation, Performance Analysis, and Design Implications", *Proceedings 37th Design Automation Conference*, p. 213–220, June 2000.

Вставка

A

Топология ИС

Создание пригодной к производству топологии

Верификация топологии

Возрастающая сложность интегральных схем сделала роль автоматизированных средств проектирования незаменимой и подняла уровень абстракций, с которыми работает разработчик, на еще бо́льшую высоту. Тем не менее в тех случаях, когда на первом месте стоит быстродействие или плотность упаковки, у разработчика нет другого выбора, кроме как вернуться к ручной прорисовке топологии схемы и физическому проектированию. Трудозатратная природа такого подхода, называемого *проектированием под заказ*, трансформируется в высокую стоимость и длительное время до выхода разработки на рынок. Поэтому он может быть оправдан только в следующих случаях.

- Заказной блок может использоваться многократно, например в качестве библиотечной ячейки.
- Повышенная стоимость одного блока может компенсировать бо́льши́м объемом производства. Примерами приложений подобного класса являются микропроцессоры и полупроводниковые схемы памяти.
- Стоимость не относится к главным критериям[1]. В качестве примеров можно привести схемы для космической аппаратуры и научного приборостроения.

[1] Что случается все реже.

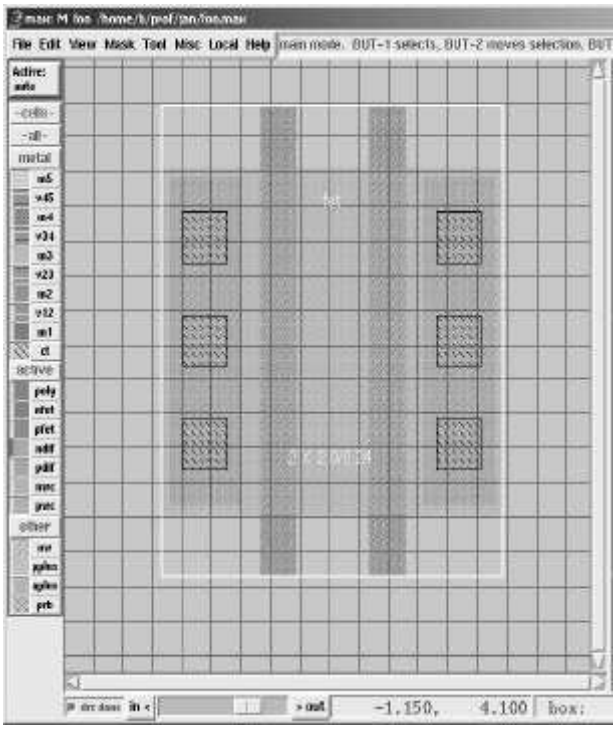


Рис. А.1. Внешний вид окна пакета **max**. Здесь показана топология двух последовательно включенных *n*-МОП-транзисторов. Меню слева позволяет выбирать слой, в котором будет размещаться конкретный многоугольник

В условиях непрерывного прогресса в области автоматизированных средств проектирования доля заказного проектирования год от года уменьшается. Даже в быстродействующих микропроцессорах большие блоки проектируются автоматически с использованием метода полузаказного проектирования. Вручную делаются только наиболее критичные с точки зрения быстродействия модули, например блоки выполнения целочисленных операций и операций с плавающей запятой.

Хотя степень автоматизации в процессе проектирования под заказ минимальна, некоторые из инструментальных средств проектирования доказали свою незаменимость. Вместе со средствами моделирования схемы эти программы составляют ядро любой среды автоматического проектирования и являются первыми средствами, с которыми столкнется честолюбивый разработчик микросхем.

Редактор топологии

Редактор топологии является основным рабочим инструментом разработчика и служит главным образом для генерации физического представления разработки, в частности топологии микросхемы. Фактически каждый поставщик средств автоматического проектирования предлагает какой-нибудь продукт в этой области. Наиболее известным является пакет MAGIC, разработанный в Калифорнийском университете в Беркли [Ousterhout84] и получивший очень широкое распространение. И хотя пакет MAGIC не устоял в ходе эволюции технологии программного обеспечения и интерфейса пользователя, некоторые из его потомков выжили. В этой книге использован пакет генерации топологии **max** — потомок MAGIC, разработанный в компании MicroMagic [mmi00]. Типичный вид экрана пакета **max** показан на рис. А.1 и иллюстрирует базовую функцию редактора топологии: размещение многоугольников в фотошаблонах различных слоев таким образом, чтобы в результате получить функционирующую физическую структуру (иногда едко называемую *растаскиванием многоугольников*).

Символьная топология

Поскольку физическое проектирование занимает основную часть времени при разработке новой ячейки или компонента, то методы ускорения этого процесса всегда были востребованы. За прошедшие годы широкую популярность приобрел подход под названием *символьное проектирование топологии*. В соответствии с этой методологией проектирования разработчик использует только стенографическое обозначение топологической структуры. Это обозначение указывает только *относительное* позиционирование различных проектируемых компонентов (транзисторов, контактов, разводки). *Абсолютные* же координаты этих элементов определяются автоматически редактором с использованием *компактора* [Hsueh79, Weste93]. Компактор переводит нормы проектирования в набор ограничений на положение компонентов и решает задачу оптимизации, пытаясь минимизировать площадь или какую-либо другую стоимостную функцию.

Пример символьного обозначения топологии микросхемы, называемого столбиковой диаграммой, показан на рис. А.2. Различные топологические объекты безразмерны, поскольку важна только их компоновка. Преимущество этого подхода состоит в том, что разработчику нет необходимости беспокоиться о нормах проектирования, так как компактор гарантирует физическую корректность конечной топологии. Таким образом, он может не заниматься утомительными манипуляциями с многоугольниками. Другим плюсом символьного подхода является наличие возможности автоматической подгонки шага ячеек к окружающей среде. Например, автоматическое согласование шага ячеек — привлекательная функция для генератора модулей. Рассмотрим ситуацию, изображенную на рис. А.3 (из работы [Croes88]), когда исходные ячейки имеют различную высоту, и расположения точек сопряжения

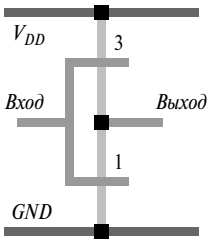


Рис. А.2. Столбиковое представление КМОП-инвертора. Числа соответствуют отношениям ширина/длина каналов транзисторов

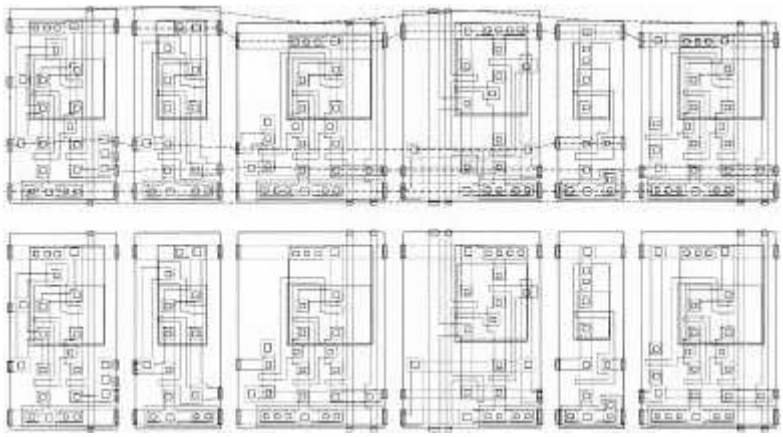


Рис. А.3. Автоматическое согласование шага канала данных в ячейках, спроектированных с использованием метода символьной топологии

не согласуются. Создание электрических соединений между такими ячейками потребовало бы дополнительной разводки. Символьный подход позволяет ячейкам автоматически подстроиться друг под друга и организовать связи между ними без каких-либо дополнительных накладных расходов.

Недостатком символьного подхода является то, что результат на выходе фазы уплотнения часто непредсказуем. Полученная топология может быть менее компактной по сравнению с результатом, получаемым вручную. Именно это не позволило данному подходу стать основным направлением в развитии инструментальных средств для проектирования топологии. Тем не менее с годами методы символьного проектирования топологии значительно улучшились и стали очень полезным инструментом для создания предварительной компоновки топологии новых ячеек. Но что более важно, они послужили хорошей основой для разработки методов автоматической генерации ячеек, описание которых приведено в главе 8.

Проверка соблюдения норм проектирования

В главе 2 было сказано, что нормы проектирования представляют собой набор топологических ограничений, гарантирующих, что после изготовления разработанная микросхема будет работать надлежащим образом и не иметь закороток или разрывов. Главным требованием к физической топологии микросхемы является соблюдение норм проектирования. Проверить это можно путем использования специальной *программы контроля норм проектирования*, входом для которой является физическая топология схемы и описания норм проектирования, представленные в форме *файла технологии*. Поскольку сложная микросхема может содержать миллионы многоугольников, относительное расположение которых должно быть проверено, то для хорошей программы контроля наиболее важным свойством является эффективность. Верификация большого кристалла может занимать часы или даже дни вычислений. Один из способов ускорения процесса заключается в сохранении иерархии разработки и на физическом уровне. Например, если ячейка используется в микросхеме неоднократно, то проверять ее следует только один раз. Кроме ускорения процесса, использование иерархии может сделать сообщения об ошибках более информативными, сохраняя информацию о структуре микросхемы.

Программы контроля норм проектирования существуют в двух форматах. 1. Программа *интерактивного* контроля исполняется параллельно с редактором топологии и сигнализирует о нарушениях норм непосредственно в процессе создания топологии ячейки. Пакет **max**, например, имеет встроенные средства контроля соблюдения норм проектирования. (Пример работы интерактивной системы контроля норм показан на рис. А.4.) 2. Программа *пакетного* контроля норм проектирования используется в качестве верификатора на этапе завершения разработки и работает с полностью законченным кристаллом перед отправкой описаний фотошаблонов производителю.

Восстановление принципиальной электрической схемы

Другим важным инструментов в методологии заказного проектирования является программа восстановления принципиальной схемы по физической топологии. Сканируя различные слои с учетом их взаимодействия, программа восстановления схемы воссоздает транзисторы, включая размеры приборов и их межсоединения. Получаемая принципиальная схема может использоваться для проверки того, что спроектированное изделие действительно реализует ожидаемую функцию. Кроме того, получающаяся в результате принципиальная электрическая схема содержит точную информацию о таких паразитных элементах, как емкость и сопротивление диффузионных областей и разводки. Это позволяет более точно выполнять моделирование и анализ. Сложность процесса восстановления в значительной степени зависит от желаемой информации. Большинство программ восстановления электри-

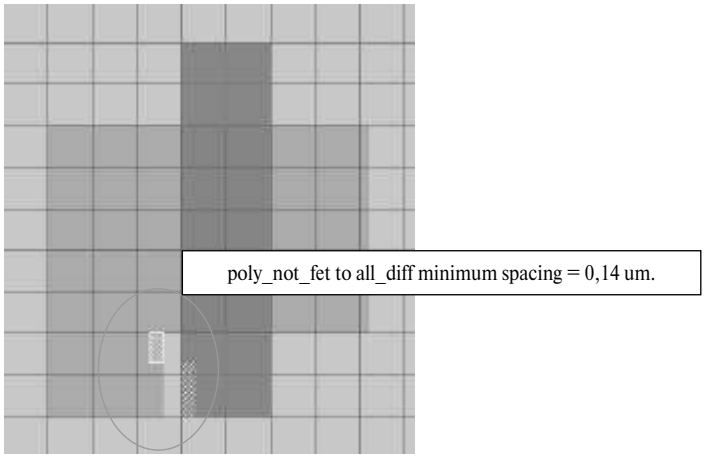


Рис. А.4. Интерактивная проверка норм проектирования. Белые точки указывают место нарушения нормы проектирования. Описание нарушенной нормы может быть получено простым щелчком мышью

ческих схем генерируют схему включения транзисторов и емкости межсоединений относительно земли (*GND*) или других узлов цепи. По более высокой стоимости уже можно получить учет сопротивления разводки, что стало необходимостью при проектировании фактически всех быстродействующих микросхем. Использование специальных алгоритмов помогло уменьшить сложность получаемых принципиальных схем. Для сверхбыстродействующих микросхем было бы желательным и восстановление индуктивностей, но, к сожалению, это требует трехмерного анализа и в настоящее время возможно только для небольших микросхем.

Для любознательных

Более подробную информацию о редакторах топологии MAGIC и **max** можно найти на Web-сайте, посвященном данной книге. Также издано большое количество хороших книг по вопросам генерации и верификации топологии, которые могут оказать существенную помощь начинающему разработчику. В качестве нескольких примеров можно назвать монографии [Clein00], [Uyemura95] и [Wolf94], содержащие подробное и хорошо иллюстрированное описание и обсуждение предмета.

Список использованной литературы

[Clein00] D. Clein, *CMOS IC Layout — Concepts, Methodologies, and Tools*, Newnes, 2000.

[Croes88] K.Croes, H. De Man, and P. Six, "CAMELEON: A Process-Tolerant Symbolic Layout System", *Journal of Solid State Circuits*, vol. 23, no. 3, p. 705–713, June 1988.

[Hsueh79] M. Hsueh and D. Pederson, "Computer-Aided Layout of LSI Building Blocks", *Proceedings ISCAS Conf.*, p. 474–477, Tokyo, 1979.

[mmi00] MicroMagic, Inc., `http://www.micromagic.com`

[Ousterhout84] J. Ousterhout, G. Hamachi, R. Mayo, W. Scott, and G. Taylor, "Magic: A VLSI Layout System", *Proc., 21st Design Automation Conference*, p. 152–159, 1984.

[Uyemura95] J. Uyemura, *Physical Design of CMOS Integrated Circuits Using L-EDIT*, PWS Publishing Company, 1995.

[Weste93] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design — A Systems Perspective*, Addison-Wesley, 1993.

[Wolf94] W. Wolf, *Modern VLSI Design — A Systems Approach*, Prentice Hall, 1994.

ГЛАВА

3

Приборы

Качественный анализ работы МОП-приборов

Простые модели компонентов для анализа вручную

Полные SPICE-модели компонентов

Влияние отклонений в технологическом процессе

3.1. Введение
3.2. Диод
3.2.1. Краткое описание принципа работы: обедненная область
3.2.2. Статическое поведение
3.2.3. Динамическое поведение или переходные процессы
3.2.4. Реальный диод. Вторичные эффекты
3.2.5. SPICE-модель диода
3.3. Полевой МОП-транзистор
3.3.1. Беглое описание принципов работы прибора
3.3.2. МОП-транзистор в статических условиях. Динамическое поведение
3.3.3. Реальный МОП-транзистор: некоторые вторичные эффекты
3.3.4. SPICE-модель МОП-транзистора
3.4. Несколько слов об отклонениях в технологическом процессе
3.5. Взгляд в перспективу: масштабирование технологии
3.6. Резюме
3.7. Для любознательных

3.1. Введение

В технике есть хорошо известная посылка, что построение концепции сложной конструкции без понимания принципов работы низлежащих строительных блоков — это верный путь к неудаче. Вышесказанное справедливо и в отношении разработки цифровых схем. Основными строительными блоками современных цифровых микросхем являются полупроводниковые приборы: если конкретнее, то это МОП-транзисторы и, в меньшей степени, паразитные диоды и межсоединения. В мире цифровых интегральных схем роль полупроводниковых приборов признавалась давно. Межсоединения же начали играть доминирующую роль только недавно после значительного масштабирования полупроводниковой технологии.

Дать читателю *необходимые знания для понимания работы* этих компонентов — основная цель следующих двух глав. В намерения авторов не входит детальное и глубокое изложение физики полупроводниковых приборов и влияния межсоединений. Для этого читатель может найти множество прекрасных книг по полупроводниковым приборам, ссылки на некоторые из которых приведены в конце главы. В данном случае основная цель заключается в том, чтобы описать принципы, лежащие в основе функционирования приборов, и выделить те свойства и параметры, которые особенно важны при проектировании цифровых вентилей, а также познакомить с принятыми условными обозначениями.

Второй важной функцией данной главы является знакомство с *моделями*. Учет всех физических аспектов каждого компонента при проектировании сложной цифровой микросхемы приводит к ненужному росту сложности, быстро становящейся "неудобоваримой". Такой подход подобен рассмотрению молекулярной структуры бетона при постройке моста. Для разрешения этого вопроса обычно используются абстракции или модели поведения компонентов. Для каждого компонента можно придумать широкий спектр моделей, отражающих компромисс между точностью и сложностью. Простые модели первого порядка полезны для выполнения анализа вручную. Они имеют ограниченную точность, но помогают понять работу схемы и выделить ее главные параметры. Если необходимо получить более точные результаты, то используются сложные модели второго и более высоких порядков и компьютерное моделирование. В данной главе будут представлены как модели первого порядка для ручного анализа, так и модели более высокого порядка для моделирования каждого из интересующих компонентов.

Обычно разработчики воспринимают выдаваемые моделями параметры компонентов как нечто незыблемое. Однако следует понимать, что это только номинальные значения, тогда как реальные значения параметров изменяются не только в зависимости от рабочей температуры, но и от партии к партии и даже в пределах одной пластины. Для раскрытия этой темы в состав главы включено краткое обсуждения *отклонений технологического процесса* и их влияния.

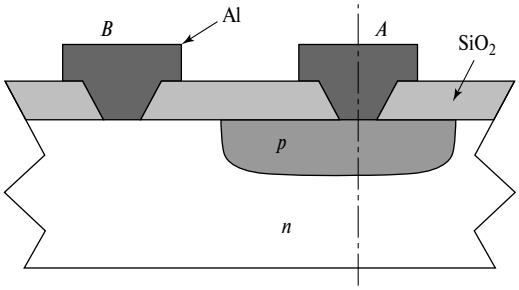
3.2. Диод

Хотя в принципиальных схемах современных цифровых вентилей диоды редко указываются в явном виде, они по-прежнему вездесущи. Каждый МОП-транзистор содержит в неявном виде несколько обратносмещенных диодов, которые оказывают непосредственное влияние на поведение прибора. В частности, зависящие от напряжения емкости, вносимые этими паразитными элементами, играют важную роль в поведении цифрового МОП-вентиля при переключении. Диоды также используются для защиты входов приборов от статического электричества. По этим причинам вполне уместно дать краткий обзор базовых свойств и уравнений, описывающих поведение диода. Дабы не увеличивать объем книги, основное внимание уделено тем аспектам, учет которых наиболее важен при проектировании цифровых МОП-схем: если говорить конкретно, то в основном речь пойдет о работе диода в режиме обратного смещения.

3.2.1. Беглое описание принципа работы: обедненная область

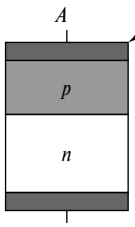
Диод на pn-переходе является простейшим из полупроводниковых приборов. На рис. 3.1, *а* показано поперечное сечение типового pn-перехода. Он состоит из двух гомогенных областей материала с проводимостью p- и n-типа, разделенных областью перехода от одного типа легирования к другому, которая по предположению обладает очень малой толщиной. Такой прибор называется *ступенчатым* или *резким переходом*. Материал p-типа легирован *акцепторными* примесями (например, бором), в результате чего в нем доминирующими или основными носителями являются дырки. Аналогично, легирование кремния *донорными* примесями (например, фосфором или мышьяком) приводит к получению материала n-типа, в котором основными носителями являются электроны. Доступ к p- и n-выводам прибора обеспечивают контакты из алюминия. Условное обозначение диода, использующееся в принципиальных схемах, показано на рис. 3.1, *в*.

Чтобы понять поведение диода с pn-переходом, часто обращаются к одномерному упрощению прибора (см. рис. 3.1, *б*). Сведение вместе материалов p- и n-типа приводит к возникновению на границе высокого градиента концентраций. Концентрация электронов изменяется от очень высокого значения в материале n-типа до очень низкого значения в материале p-типа. Обратное справедливо для концентрации дырок. Этот градиент вызывает *диффузию* электронов из n-области в p-область и дырок из p-области в n-область. Когда дырки покидают материал p-типа, они оставляют за собой неподвижные отрицательно заряженные ионы акцепторной примеси. Как следствие, материал p-типа в окрестности pn-границы заряжается отрицательно. Аналогично, диффундирующие электроны оставляют после себя положительно заряженные ионы донорной примеси, и на n-стороне границы возникает положительный заряд. Область у перехода, из которой ушли основные носители, оставив неподвижные ионы донорной и акцепторной примесей, называется



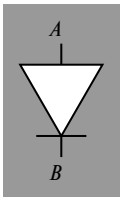
Поперечное сечение *pn*-перехода в интегральной схеме

а



Одномерное представление

б



Условное обозначение диода

в

Рис. 3.1. Диод с резким pn-переходом и его условное обозначение

обедненной областью или *областью пространственного заряда*. Заряды создают на границе электрическое поле, направленное от n-области к p-области. Это поле противодействует диффузии дырок и электронов, поскольку заставляет электроны дрейфовать из p-области в n-область и дырки — из n-области в p-область. При равновесии заряд в обедненной области порождает такое электрическое поле, что дрейфовый ток оказывается равным и противоположным по направлению диффузионному току, и в результате полный ток равен нулю.

Результаты вышеприведенного анализа в графическом виде показаны на рис. 3.2, на котором изображены направления токов, плотности зарядов, электрическое поле и электростатический потенциал в резком pn-переходе при нулевом смещении. В изображенном приборе p-материал легирован сильнее, чем материал n-типа, или $N_A > N_D$, где N_A и N_D — концентрации акцепторной и донорной примесей соответственно. Поэтому концентрация заряда в обедненной области выше на p-стороне перехода. На рис. 3.2 также видно, что при нулевом смещении на переходе существует напряжение ϕ_0, называемое *контактной разностью потенциалов*. Величина этого напряжения равна:

$$\phi_0 = \phi_T \ln \left[\frac{N_A N_D}{n_i^2} \right], \tag{3.1}$$

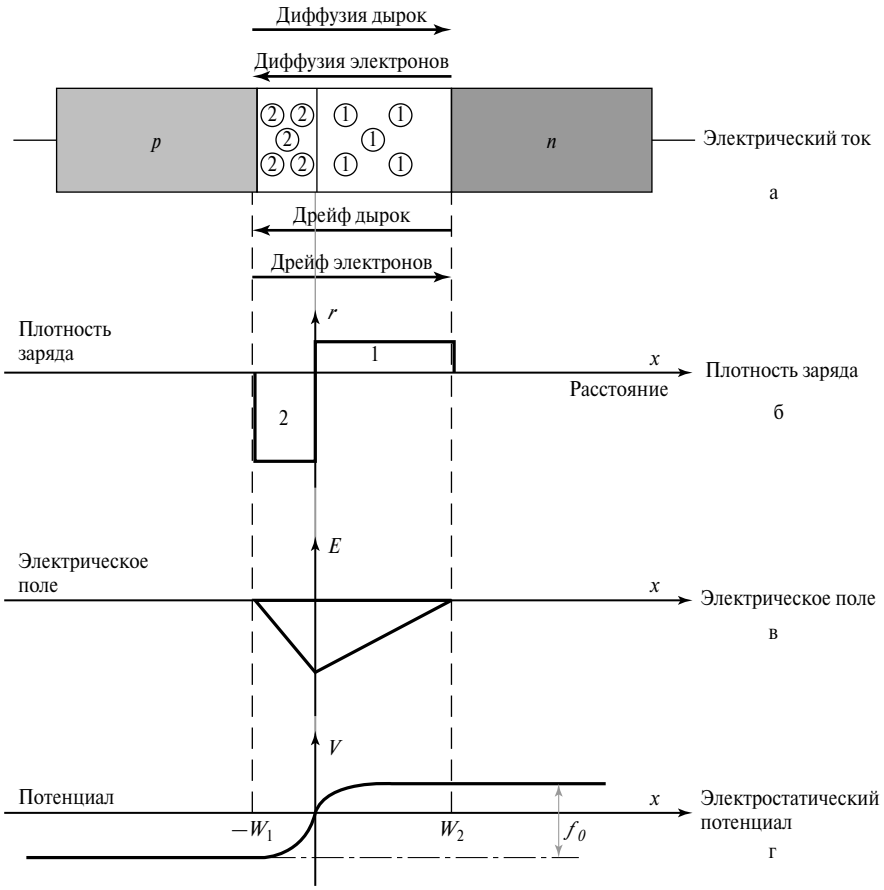


Рис. 3.2. Резкий pn-переход в условиях равновесия

где ϕ_T — *тепловой потенциал*:

$$\phi_T = \frac{kT}{q} = 26 \text{ мВ при } 300 \text{ К}. \tag{3.2}$$

Величина n_i представляет собой концентрацию собственных носителей в чистом полупроводнике и для кремния при 300 К приблизительно равна $1,5 \times 10^{10}$ см$^{-3}$.

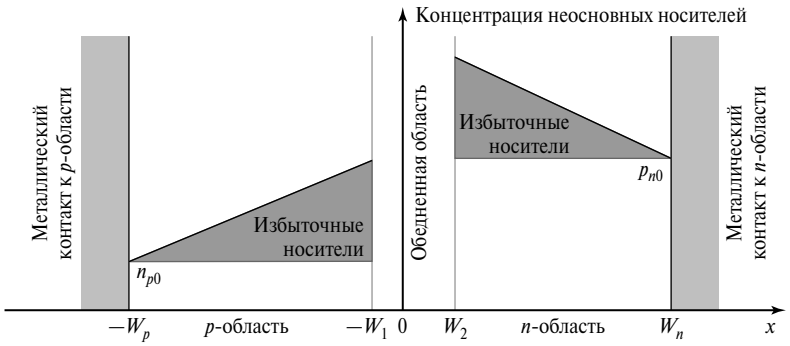


Рис. 3.3. Концентрация неосновных носителей в нейтральных областях рядом с резким pn-переходом, находящимся в режиме прямого смещения

Пример 3.1. Контактная разность потенциалов *pn*-перехода

Резкий переход характеризуется концентрациями примесей порядка $N_A = 10^{15}$ атомов/см$^{-3}$ и $N_D = 10^{16}$ атомов/см$^{-3}$. Вычислим величину контактной разности потенциалов при 300 К.

Из уравнения (3.1) имеем:

$$\phi_0 = 26 \ln \left[\frac{10^{15} \times 10^{16}}{2,25 \times 10^{20}} \right] \text{мВ} = 638 \text{ мВ}.$$

3.2.2. Статическое поведение

Уравнение идеального диода

Предположим теперь, что к переходу в прямом направлении прикладывается напряжение V_D — потенциал p-области повышается относительно n-зоны. Прикладываемый потенциал снижает высоту потенциального барьера. В результате поток подвижных носителей через переход увеличивается, так как диффузионный ток преобладает над дрейфовой составляющей. Эти носители проходят сквозь обедненную зону и инжектируются в нейтральные n- и p-области, где они становятся неосновными носителями, как показано на рис. 3.3. Если предположить, что в нейтральных областях отсутствует градиент напряжения, что приблизительно справедливо для большинства современных приборов, то эти неосновные носители благодаря наличию градиента концентрации будут диффундировать сквозь область, пока не рекомбинируют с основными носителями. В итоге через диод будет протекать ток из p-области в n-область, и в этом случае говорят, что диод находится в режиме *прямого смещения*.

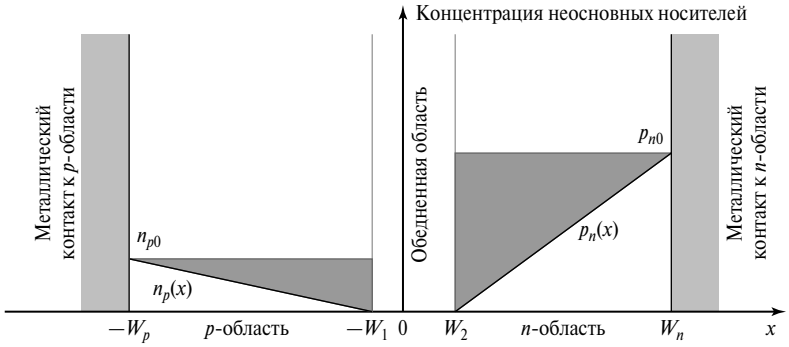


Рис. 3.4. Концентрация неосновных носителей в нейтральных областях рядом с pn-переходом, находящимся в режиме обратного смещения

Напротив, если к диоду прикладывается напряжение V_D в обратном направлении, т.е. когда потенциал p-области понижается относительно n-зоны, то высота потенциального барьера увеличивается. Это приводит к снижению диффузионного тока, и доминирующей становится дрейфовая составляющая. Ток течет из n-области в p-область. Поскольку количество неосновных носителей в нейтральных областях (электронов в p-зоне и дырок в n-области) очень мало, то этой дрейфовой составляющей тока можно практически пренебречь (рис. 3.4). С достаточной точностью можно заявить, что в *режиме обратного смещения* диод работает как непроводящий или запирающий прибор. Таким образом, диод представляет собой однонаправленный проводник.

Наиболее важным свойством тока диода является его *экспоненциальная зависимость* от прикладываемого напряжения смещения. Этот факт проиллюстрирован на рис. 3.5, *а*, где приведен график зависимости тока диода I_D в функции напряжения смещения V_D. Еще более очевиден экспоненциальный характер тока при прямом смещении на рис. 3.5. *б*, где ток показывается в логарифмическом масштабе. При повышении напряжения прямого смещения на 60 мВ ($= 2,3\phi_T$) ток увеличивается в 10 раз. При малых уровнях напряжения ($V_D < 0,15$ В) можно видеть отклонение от экспоненциальной зависимости, которое обусловливается рекомбинацией дырок и электронов в обедненной области.

Поведение диода как в режиме прямого смещения, так и в режиме обратного смещения лучше всего описывается с помощью хорошо известного *уравнения идеального диода*, связывающего ток через диод I_D с напряжением смещения диода V_D:

$$I_D = I_S \left(e^{\,V_D/\phi_T} - 1 \right). \tag{3.3}$$

Экспоненциальный характер поведения диода иллюстрирует рис. 3.5. ϕ_T — это тепловой потенциал из выражения (3.2), равный при комнатной температуре 26 мВ. I_S представляет собой постоянную величину, называемую *током насыщения* диода. Ток насыщения пропорционален площади диода и является функцией уровней легирования и ширины нейтральных областей. Чаще всего величина тока I_S определяется эмпирически. Стоит упомянуть, что в реальных приборах обратный ток значительно выше тока насыщения I_S. Обусловлено это тепловой генерацией электрон-дырочных пар в обедненной области. Присутствующее электрическое поле растаскивает эти носители из области, что приводит к появлению дополнительной составляющей тока. Для типовых переходов в кремнии номинальное значение тока насыщения лежит в диапазоне 10^{-17} А/мкм^2, тогда как фактические значения обратных токов примерно на три порядка выше. Как следствие, реальные значения токов утечки обратносмещенных диодов необходимо определять путем прямых измерений на реальных приборах.

Модели для ручного анализа

Полученные уравнения относительно тока и напряжения могут быть использованы для формирования простых моделей, полезных при ручном анализе

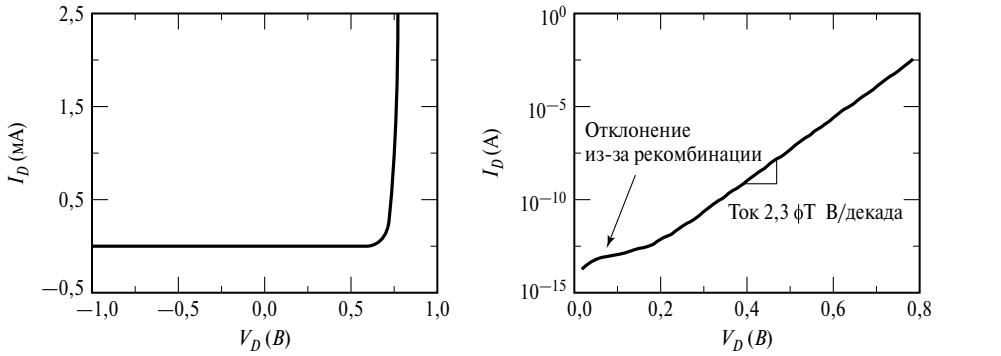


Рис. 3.5. Ток диода в функции напряжения смещения V_D

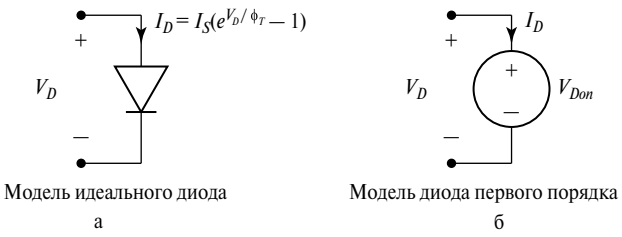


Рис. 3.6. Модели диода

диодных цепей. Одна из моделей, показанная на рис. 3.6, *а*, основана на уравнении идеального диода (3.3). Хотя эта модель и дает точные результаты, ее недостаток заключается в ее сильной нелинейности, что не дает возможности выполнять быстрый прикидочный анализ условий работы цепи на постоянном токе. Часто используемая упрощенная модель получается из рассмотрения показанного на рис. 3.5 графика тока диода. У "полностью проводящего" диода величина падения напряжения на диоде V_D лежит в узком диапазоне значений: где-то между 0,6 и 0,8 В. Поэтому в первом приближении разумно предположить, что падение напряжения на проводящем диоде V_{Don} постоянно. Хотя величина V_{Don} зависит от тока I_S, обычно предполагается значение 0,7 В, что приводит к модели, показанной на рис. 3.6, *б*, в соответствии с которой проводящий диод заменяется источником фиксированного напряжения, а непроводящий — разрывом цепи.

Пример 3.2. Анализ цепи с диодом

Рассмотрим простую цепь, показанную на рис. 3.7, и предположим, что $V_S = 3$ В, $R_S = 10$ кОм и $I_S = 0,5 \times 10^{-16}$ А. Ток диода и напряжения в цепи связаны следующим уравнением:

$$V_S - R_S I_D = V_D.$$

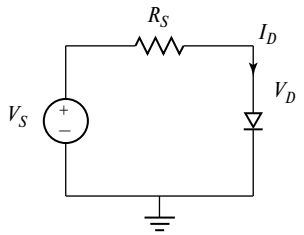


Рис. 3.7. Простая цепь с диодом

Если подставить уравнение идеального диода и решить (помучавшись) нелинейное уравнение, воспользовавшись числовым или итерационным методом, то в результате получается следующее решение: $I_D = 0,224$ мА и $V_D = 0,757$ В. Упрощенная модель с $V_{Don} = 0,7$ В дает похожие результаты ($V_D = 0,7$ В и $I_D = 0,23$ мА), но при значительно меньших усилиях. Таким образом, для расчета цепи с диодом в первом приближении вполне имеет смысл использовать именно эту модель.

3.2.3. Динамическое поведение или переходные процессы

До сих пор рассматривались в основном статические характеристики диода или поведение в установившемся состоянии. Такую же важность при проектировании цифровых схем имеет и реакция прибора на изменения условий смещения. Переходные или динамические характеристики определяют максимальное быстродействие, с которым может работать прибор. Поскольку рабочий режим диода является функцией величины зарядов, присутствующих как в нейтральных областях, так и в области пространственного заряда, то его динамическое поведение в значительной степени определяется тем, насколько быстро эти заряды могут перемещаться.

Тут, конечно, есть большой соблазн перейти к подробному анализу переключательного поведения диода в режиме прямого смещения, но в данном случае это просто излишне усложнит рассмотрение вопроса. Фактически в работающей цифровой интегральной МОП-схеме все диоды находятся в режиме обратного смещения, и предполагается, что они в нем и остаются при всех обстоятельствах. Прямое смещение может возникнуть только в исключительных случаях. Примером такого случая может быть наличие больших положительных (или отрицательных) выбросов на шинах питания микросхемы. Поскольку это оказывает пагубное влияние на работу схемы в целом, то при всех обстоятельствах следует исключать возможность их появления. Поэтому в последующем материале основное внимание будет уделено исключительно рассмотрению заряда в обедненной области, который, собственно, и управляет динамическим поведением диода в режиме обратного смещения.

Емкость обедненной области

В идеальной модели в обедненной области нет подвижных носителей, и ее заряд определяется неподвижными ионами донорной и акцепторной примесей. График распределения заряда при нулевом смещении уже был показан на рис. 3.2. Эту картину можно легко распространить и на случай наличия смещения. На интуитивном уровне достаточно просто убедиться в следующем: в условиях прямого смещения потенциальный барьер понижается, что означает уменьшение пространственного заряда, необходимого для создания разности потенциалов. Это соответствует уменьшению ширины обедненной области. С другой стороны, при обратном смещении потенциальный барьер увеличивается. Соответственно увеличивается пространственный заряд и обедненная область становится шире. Данные рассуждения подтверждаются приведенными ниже хорошо известными выражениями для обедненной области[1]. Один из выводов имеет критически важное значение: поскольку должен соблюдаться принцип общей нейтральности, то величины общего заряда ионов акцепторов и ионов доноров должны быть численно равны.

[1] Вывод этих выражений, справедливых для резкого перехода, либо достаточно прост, либо его можно найти в литературе по полупроводниковым приборам, например в [Howe97].

1. Заряд в обедненной области (при положительном смещении V_D положительно).

$$Q_j = A_D \sqrt{\left(2\varepsilon_{Si} q \frac{N_A N_D}{N_A + N_D}\right)(\phi_0 - V_D)}. \tag{3.4}$$

2. Ширина области обеднения.

$$W_j = W_2 - W_1 = \sqrt{\left(\frac{2\varepsilon_{Si}}{q}\frac{N_A + N_D}{N_A N_D}\right)(\phi_0 - V_D)}. \tag{3.5}$$

3. Максимальная напряженность электрического поля.

$$E_j = \sqrt{\left(\frac{2q}{\varepsilon_{Si}}\frac{N_A N_D}{N_A + N_D}\right)(\phi_0 - V_D)}. \tag{3.6}$$

В приведенных выше выражениях ε_{Si} обозначает диэлектрическую проницаемость кремния, которая в 11,7 раза превышает проницаемость вакуума, равную $1,053 \times 10^{-10}$ Ф/м. Величина отношения ширины обедненной области на n-стороне к ширине обедненной области на p-стороне определяется отношением уровней легирования: $W_2/(-W_1) = N_A/N_D$.

Абстрактно обедненную область можно представить себе в виде конденсатора, хотя и с весьма специфическими характеристиками. Поскольку область пространственного заряда содержит небольшое количество подвижных носителей, то она играет роль изолятора с диэлектрической проницаемостью полупроводникового материала. В качестве обкладок конденсатора выступают n- и p-области. Небольшое изменение прикладываемого к переходу напряжения dV_D вызывает изменение пространственного заряда dQ_j. Поэтому емкость обедненного слоя можно записать в следующем виде:

$$\begin{aligned} C_j &= \frac{dQ_j}{dV_D} = A_D \sqrt{\left(\frac{\varepsilon_{Si} q}{2}\frac{N_A N_D}{N_A + N_D}\right)(\phi_0 - V_D)^{-1}} \\ &= \frac{C_{j0}}{\sqrt{1 - V_D/\phi_0}}, \end{aligned} \tag{3.7}$$

где C_{j0} — емкость при нулевом смещении, являющаяся функцией только физических параметров прибора. Это дает следующее соотношение:

$$C_{j0} = A_D \sqrt{\left(\frac{\varepsilon_{Si} q}{2}\frac{N_A N_D}{N_A + N_D}\right)\phi_0^{-1}}. \tag{3.8}$$

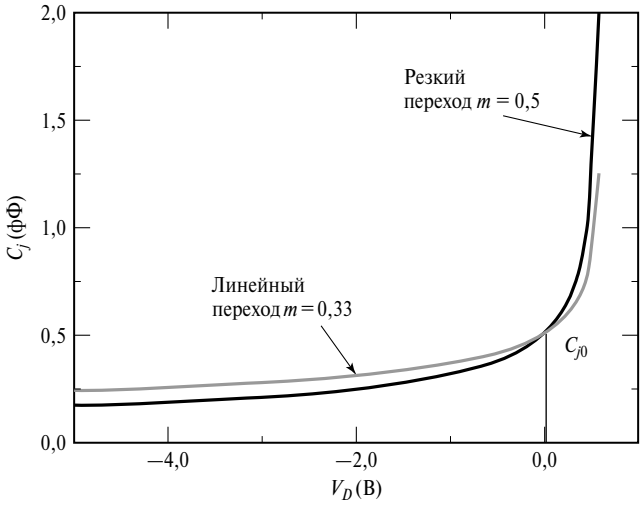


Рис. 3.8. Емкость перехода (в фФ/мкм^2) в функции прикладываемого напряжения смещения

Стоит обратить внимание, что точно такое же значение емкости получается, если воспользоваться стандартным выражением для емкости конденсатора с параллельными обкладками $C_j = \varepsilon_{Si} A_D / W_j$ (W_j задается выражением (3.5)). Обычно коэффициент A_D опускается, и емкости C_j и C_{j0} выражаются в единицах емкость/единица площади.

График зависимости емкости перехода типового кремниевого диода, который можно найти в МОП-схемах, в функции напряжения смещения показан на рис. 3.8. Можно видеть наличие сильной *нелинейной зависимости*. Обратите внимание, что величина емкости уменьшается с увеличением обратного смещения: обратное смещение в 5 В приводит к уменьшению емкости более чем в два раза.

Пример 3.3. Емкость перехода

Рассмотрим следующий кремниевый диод на *pn*-переходе: $C_{j0} = 2 \times 10^{-3}$ Ф/м^2, $A_D = 0,5$ мкм^2 и $\phi_0 = 0,64$ В. При обратном смещении $-2,5$ В емкость перехода будет равна $0,9 \times 10^{-3}$ Ф/м^2 ($0,9$ фФ/мкм^2) или в целом для диода — $0,45$ фФ.

Выражение (3.7) справедливо только при условии, что *pn*-переход является резким, т.е. переход от *n*-материала к *p*-материалу не имеет протяженности. У *pn*-переходов в реальных интегральных схемах это часто не так, и переход от *n*-материала к *p*-материалу может быть плавным. В подобных случаях линейное распределение примесей в переходе является лучшей аппроксимацией, чем ступенчатая функция резкого перехода. Анализ линейного плавного перехода показывает, что выражение для емкости перехода (3.7)

справедливо и для этого случая, но только с той разницей, что изменяется показатель степени в знаменателе. Более общее выражение для емкости перехода имеет вид

$$C_j = \frac{C_{j0}}{(1 - V_D/\phi_0)^m}, \tag{3.9}$$

где m — так называемый *коэффициент плавности*, который равен $1/2$ для резкого перехода и $1/3$ для линейного или плавного перехода. Оба случая показаны на рис. 3.8.

Емкость обедненной области для случая большого сигнала

Рис. 3.8 ясно свидетельствует о том факте, что емкость перехода является зависящим от напряжения параметром, значение которого значительно изменяется с изменением положения рабочей точки прибора. В цифровых схемах рабочие напряжения обычно быстро меняются в широком диапазоне значений. При этих обстоятельствах более привлекательно заменить зависящую от напряжения нелинейную емкость C_j эквивалентной линейной емкостью C_{eq}. Емкость C_{eq} определяется как такая, которая при заданном диапазоне изменения напряжения от высокого уровня V_{high} до низкого уровня V_{low} обеспечивает передачу такого заряда, какой бы прогнозировался в случае использования нелинейной модели:

$$C_{eq} = \frac{\Delta Q_j}{\Delta V_D} = \frac{Q_j(V_{high}) - Q_j(V_{low})}{V_{high} - V_{low}} = K_{eq} C_{j0}. \tag{3.10}$$

Объединение выражения (3.4) (преобразованного с учетом коэффициента плавности m) и выражения (3.9) дает значение коэффициента K_{eq}:

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1 - m)} \left[(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m} \right]. \tag{3.11}$$

Пример 3.4. Средняя емкость перехода

Диод из примера 3.3 переключается в диапазоне напряжений от 0 до $-2,5$ В. Вычислим среднюю емкость перехода ($m = 0,5$).

При заданном диапазоне напряжений и $\phi_0 = 0,64$ В оценка значения K_{eq} дает результат 0,622. Таким образом, средняя емкость равна 1,24 фФ/мкм^2.

3.2.4. Реальный диод. Вторичные эффекты

На практике ток диода меньше предсказываемого уравнением идеального диода. Не все прикладываемое напряжение смещения оказывается непосредственно на переходе: всегда есть какое-то падение напряжения на нейтральных областях. К счастью, сопротивление нейтральных зон мало (от 1 до 100 Ом в зависимости от уровней легирования), и падение напряжения становится заметным только при больших токах ($>$ 1мА). Этот эффект может быть смоделирован путем добавления резистора последовательно с контактами к n- и p-областям диода.

При изложении выше также предполагалось, что при достаточно высоком обратном смещении обратный ток достигает постоянного значения, которое по сути равно нулю. Когда обратное смещение начинает превышать определенный уровень, называемый *напряжением пробоя*, обратный ток резко увеличивается (рис. 3.9). В диодах, которые можно найти в типовых КМОП-схемах, это увеличение тока вызывается *лавинным пробоем*. Рост обратного смещения поднимает величину напряженности электрического поля в переходе. Как следствие, пересекающие обедненную область носители ускоряются, приобретая высокую скорость. При критической напряженности поля E_{crit} носители приобретают достаточно энергии, чтобы при столкновениях с неподвижными атомами кремния образовывать электронно-дырочные пары. В свою очередь, эти носители перед выходом из обедненной области порождают новые носители. Значение E_{crit} при концентрации примесей порядка 10^{16} см$^{-3}$ составляет примерно 2×10^5 В/см. Хотя лавинный пробой сам по себе не является разрушительным и этот эффект исчезает при снятии обратного смещения, продолжительное удержание диода в условиях лавинного пробоя не рекомендуется, так как высокие уровни токов и связанное с ними выделение тепла может вызвать необратимое повреждение структуры. Следует сказать, что лавинный пробой не единственный механизм пробоя, встречающийся в диодах. В диодах с высоким уровнем легирования может иметь место и другой механизм, называемый *зеннеровским пробоем*. (Дальнейшее рассмотрение этого явления выходит за рамки рассмотрения данной книги.)

Наконец, необходимо отметить, что на ток диода влияет и рабочая температура, причем характер этого влияния двоякий.

1. Тепловой потенциал ϕ_T, стоящий в показателе степени экспоненты в уравнении тока, линейно зависит от температуры. Рост ϕ_T вызывает уменьшение тока.
2. Ток насыщения I_S тоже зависит от температуры, поскольку величина концентрации носителей при тепловом равновесии повышается с увеличением температуры. Теоретически ток насыщения приблизительно удваивается через каждые 5°C. Экспериментально было доказано, что обратный ток удваивается через каждые 8°C.

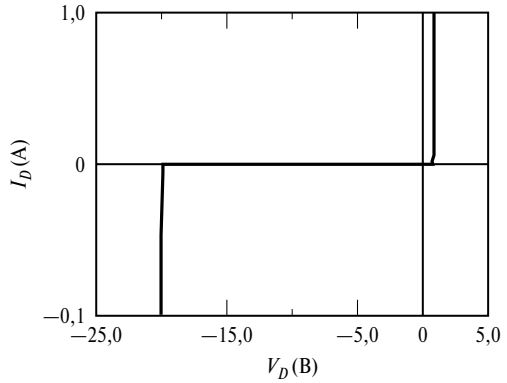


Рис. 3.9. Вольт-амперная характеристика диода на pn-переходе с участком пробоя на обратной ветви (напряжение пробоя = 20 В)

Такая двойная зависимость оказывает значительное влияние на работу цифровой схемы. Во-первых, могут существенно увеличиваться уровни токов (и, следовательно, потребляемая мощность). Например, при 300 К и прямом смещении 0,7 В ток увеличивается примерно на 6% на каждый градус Цельсия и удваивается через каждые 12°C. При фиксированной величине тока напряжение на диоде V_D уменьшается на 2 мВ через каждый градус Цельсия. Во-вторых, функционирование цифровых схем сильно зависит от работы обратно смещенных диодов в качестве изолирующих элементов. Рост температуры вызывает увеличение токов утечки и уменьшает качество изоляции.

3.2.5. SPICE-модель диода

В предыдущих разделах была представлена модель для ручного анализа тока диода. Для более сложных цепей или в тех случаях, когда требуется более точное моделирование диода с учетом эффектов второго порядка, ручной расчет становится затруднительным, и требуется привлечение компьютерного моделирования. Хотя за последние несколько десятилетий было разработано множество различных программ моделирования схем, определенно наиболее успешной является пакет SPICE, разработанный в Калифорнийском университете в Беркли [Nagel75]. Моделирование интегральной схемы, содержащей активные приборы, требует наличия математической модели этих приборов. Далее будем называть такую модель *SPICE-моделью*. Точность моделирования непосредственно зависит от качества модели. Например, нельзя ожидать увидеть результат действия эффекта второго порядка, если этот эффект не отражен в модели прибора. Создание точных и эффективных в вычислительном плане SPICE-моделей было длительным и утомительным процессом, и он ни в коей мере не окончен. Некоторое время назад все крупные компании-производители полупроводников устройств разрабатывали собственные модели, заявляя, что они более точные, имеют более высокую вычислительную эффективность или помехоустойчивость. К счастью, в последнее время в области моделирования наметилась некоторая консолидация.

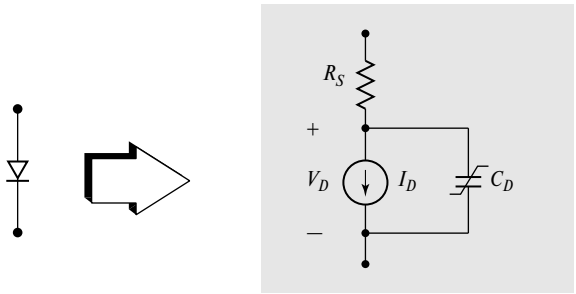


Рис. 3.10. SPICE-модель диода

Стандартная SPICE-модель диода проста (рис. 3.10). Статическая характеристика диода моделируется нелинейным источником тока I_D, который описывается уравнением, представляющим собой модифицированный вариант уравнения идеального диода:

$$I_D = I_S(e^{\,V_D/n\phi_T} - 1). \tag{3.12}$$

Дополнительный параметр n называется *коэффициентом эмиссии*. Для большинства обычных диодов он равен 1, хотя для некоторых может быть несколько больше 1. Резистор R_S моделирует последовательное сопротивление, вносимое нейтральными областями по обе стороны перехода. При более высоких уровнях тока это сопротивление приводит к тому, что внутреннее напряжение диода V_D отличается от внешнего прикладываемого напряжения, и величина тока оказывается ниже той, которую можно было бы ожидать из уравнения идеального диода.

Динамическое поведение диода моделируется нелинейной емкостью C_D, которая объединяет в себе два эффекта, связанных с накоплением заряда в диоде: пространственный заряд (или заряд обедненной области) и заряд избыточных неосновных носителей. В этой главе рассматривался только первый из них, так как второй имеет место только при прямом смещении. Модель нелинейной емкости описывается следующим выражением:

$$C_D = \frac{C_{j0}}{(1 - V_D/\phi_0)^m} + \frac{\tau_T I_S}{\phi_T} e^{\,V_D/n\phi_T}. \tag{3.13}$$

Список параметров, использующихся в модели диода, приведен в табл. 3.1. Кроме названия параметра, его символа и имени в SPICE в таблице также приводится его значение по умолчанию, используемое пакетом SPICE в том случае, когда параметр не задается в явном виде. Эта таблица не является полной: существуют и другие параметры, учитывающие такие эффекты второго порядка, как пробой, высокие уровни инжекции и шум. Для краткости список параметров был ограничен только теми из них, которые представляют непосредственный интерес в контексте данной книги. Полное описание моделей приборов и порядок использования пакета SPICE читатель может найти

Таблица 3.1. Параметры SPICE-модели диода первого порядка

Параметр	Обозначение	Название SPICE	Единицы	Значение по умолчанию
Ток насыщения	I_S	IS	A	1.0E-14
Коэффициент эмиссии	n	N	–	1
Последовательное сопротивление	R_S	RS	Ом	0
Время перехода	τ_T	TT	с	0
Емкость соединения с нулевым смещением	C_{j0}	CJ0	Ф	0
Коэффициент плавности	m	M	–	0.5
Контактная разность потенциалов	ϕ_0	VJ	В	1

в многочисленных монографиях, посвященных данному предмету (см., например, [Banhzaf92], [Thorpe92] и [Vladimirescu93]).

3.3. Полевой МОП-транзистор

Вне всякого сомнения, полевой транзистор со структурой металл–окисел–полупроводник (полевой МОП-транзистор или просто МОП-транзистор для краткости) является рабочей лошадкой современных цифровых микросхем. Его главным достоянием с точки зрения цифровой обработки является прекрасная работа в качестве ключа, а также малое количество паразитных эффектов. Другие его важные преимущества состоят в том, что достигаемая на его основе высокая плотность интеграции совместно с относительно "простым" процессом изготовления делают возможным экономически оправданное производство больших и сложных схем.

Следуя подходу, использованному при рассмотрении диода, в данном разделе мы ограничимся только общим обзором поведения транзистора и его параметров. После небольшого обсуждения основных принципов работы прибора будет представлено аналитическое описание транзистора как с точки зрения его поведения в статических условиях (в устоявшемся состоянии), так и в динамике (переходные процессы). В конце раздела будут перечислены некоторые эффекты второго порядка и дано представление о существующих SPICE-моделях МОП-транзистора.

3.3.1. Краткое описание принципов работы прибора

Полевой МОП-транзистор — это прибор с четырьмя выводами. Напряжение, подаваемое на вывод *затвора*, определяет наличие и величину тока, протекающего между портами *истока* и *стока*. *Подложка* представляет собой четвертый вывод транзистора. Его функция вторична, поскольку он служит только для модулирования характеристик и параметров прибора.

На самом простом уровне транзистор можно считать ключом. Если прикладываемое к затвору напряжение выше некоторого заданного (так называемого *порогового напряжения* V_T), то между стоком и истоком образуется про-

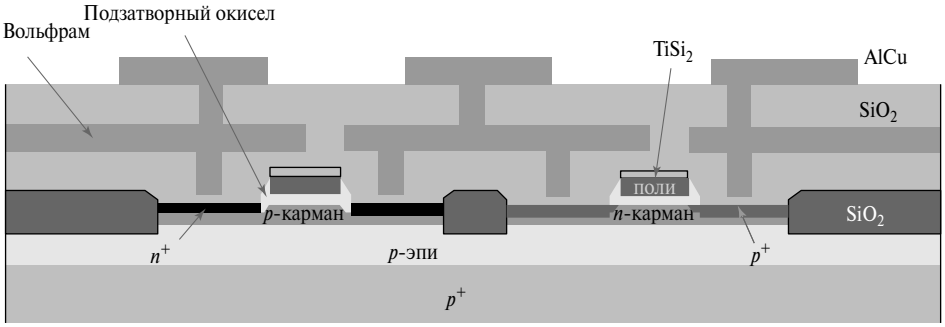


Рис. 3.11. Поперечное сечение структуры, формируемой при двухкарманном КМОП-процессе

водящий канал. При наличии разности потенциалов между стоком и истоком между ними начинает течь ток. Проводимость канала модулируется напряжением затвора: чем больше разность потенциалов между истоком и затвором, тем меньше сопротивление и тем больше ток. Если напряжение на затворе меньше порогового, то канал отсутствует и ключ считается разомкнутым.

Различают два типа полевых МОП-транзисторов. n-МОП-тразистор содержит n^+-области истока и стока, погруженные в подложку p-типа. Ток переносится электронами, двигающимися по каналу n-типа между истоком и стоком. Это отличается от диода на pn-переходе, в котором ток переносится как электронами, так и дырками. МОП-приборы могут также изготавливаться с использованием подложки n-типа и p^+-областей истока и стока. В таком транзисторе ток переносится дырками, двигающимися в канале p-типа. Подобный прибор называется МОП-транзистором с p-каналом или p-МОП-транзистором. В технологии комплементарных МОП-транзисторов (КМОП) присутствуют приборы обоих типов. Вид поперечного сечения структуры, формируемой в результате современного двухкарманного КМОП-процесса, уже приводился в главе 2 и здесь воспроизводится просто для удобства (рис. 3.11).

Условные обозначения различных МОП-транзисторов показаны на рис. 3.12. Как уже упоминалось ранее, транзистор является четырехпортовым прибором с выводами затвора, истока, стока и подложки (см. рис.3.12, *а* и *в*). Поскольку подложка в общем случае подключается к источнику напряжения постоянного типа, одному для всех приборов одинакового типа (к земле (GND) у n-МОП-транзисторов, к напряжению V_{dd} у p-МОП-транзистора), то чаще всего она на принципиальной схеме не показывается (см. рис. 3.12, *б* и *г*). **Если четвертый вывод не показан, то предполагается, что подложка подключается к соответствующему выводу источника питания.**

3.3.2. МОП-транзистор в статических условиях

При выводе модели МОП-транзистора основное внимание будет сосредоточено на n-МОП-приборе. Но все те же доводы справедливы также и для p-МОП-прибора и будут рассмотрены в конце раздела.

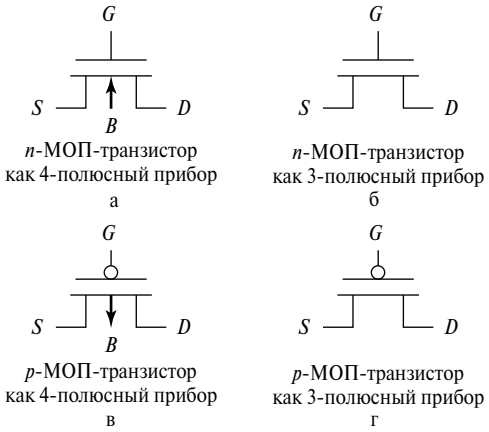


Рис. 3.12. Условные обозначения МОП-транзисторов

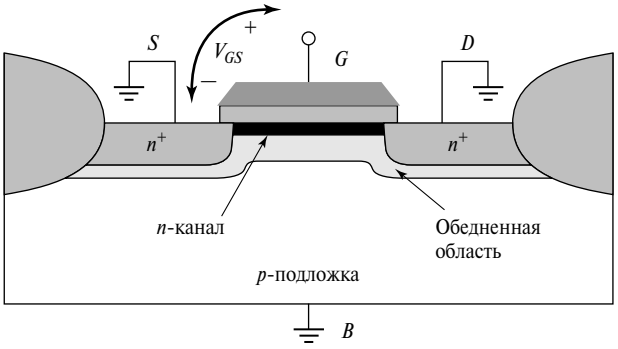


Рис. 3.13. МОП-транзистор n-типа с обедненной областью и индуцированным каналом при положительном напряжении V_{GS}

Пороговое напряжение

Рассмотрим сначала случай, когда $V_{GS} = 0$, а сток, исток и подложка подключены к земле. Цепь между истоком и стоком представляет собой два встречно включенных pn-перехода (подложка-исток и подложка сток). В упомянутых условиях оба перехода имеют нулевое смещение и могут считаться непроводящими, что в результате приводит к наличию между истоком и стоком чрезвычайно высокого сопротивления.

Теперь предположим, что к затвору приложено положительное (относительно истока) напряжение, как показано на рис. 3.13. Затвор и подложка образуют обкладки конденсатора с подзатворным окислом в качестве диэлектрика. Положительное напряжение на затворе приводит к накоплению положительного и отрицательного зарядов на электроде затвора и в подложке соответственно. Последний первоначально проявляется в отталкивании подвижных дырок. В результате под затвором формируется обедненная

область, аналогичная той, что имеет место в диоде на pn-переходе. Следовательно, для ширины и величины пространственного заряда на единицу площади справедливы подобные выражения. Сравните приведенные ниже выражения с выражениями (3.4) и (3.5).

$$W_d = \sqrt{\frac{2\varepsilon_{Si}\phi}{qN_A}} \tag{3.14}$$

и

$$Q_d = \sqrt{2qN_A\varepsilon_{Si}\phi}. \tag{3.15}$$

В этих выражениях N_A — уровень легирования подложки; ϕ — напряжение на обедненном слое (т.е. потенциал на границе раздела окисел–кремний).

С увеличением напряжения на затворе потенциал на поверхности кремния достигает в некоторой точке критической величины, и поверхность полупроводника меняет свой тип проводимости (т.е. становится материалом n-типа). Эта точка отмечает момент возникновения явления *сильной инверсии*, которое имеет место при напряжении, равном удвоенному *потенциалу Ферми* (см. выражение (3.15)) (для типовой подложки из кремния p-типа $\phi_F \approx -0,3$ В):

$$\phi_F = \phi_T \ln\left(\frac{N_A}{n_i}\right). \tag{3.16}$$

Дальнейшее увеличение напряжения на затворе не приводит к дальнейшему изменению ширины обедненного слоя, но имеет результатом появление в тонком инверсионном слое непосредственно под окислом дополнительных электронов. Они втягиваются в инверсионный слой из сильнолегированной n^+-области истока. Таким образом, между областями истока и стока образуется сплошной канал n-типа, проводимость которого модулируется напряжением затвор-исток.

В присутствии инверсионного слоя величина хранящегося в обедненной области заряда фиксирована и равна:

$$Q_{B0} = \sqrt{2qN_A\varepsilon_{Si}\,|2\phi_F|}. \tag{3.17}$$

Эта картина несколько изменяется, когда между истоком и подложкой прикладывается напряжение смещения подложки V_{SB} (для приборов с n-каналом напряжение V_{SB} обычно положительно). Это приводит к увеличению потенциала поверхности, необходимого для возникновения сильной инверсии, который становится равным $|-2\phi_F + V_{SB}|$. Теперь величину заряда в обедненной области можно описать выражением вида.

$$Q_B = \sqrt{2qN_A\varepsilon_{Si}\,(|(-2)\,\phi_F + V_{SB}|)}. \tag{3.18}$$

Напряжение затвор-исток V_{GS}, при котором наступает сильная инверсия, называется *пороговым напряжением* V_T. VT является функцией нескольких параметров, бо́льшая часть которых представляет собой константы материала, например, разность работ выхода материалов затвора и подложки, толщина окисла, положение уровня Ферми, величина заряда примесей, захваченных на поверхности между каналом и подзатворным окислом, доза ионной имплантации в процессе подгонки величины порога. Из предыдущих рассуждений также понятно, что на величину порога влияет напряжение исток-подложка V_{SB}. Обычно вместо сложного (и едва ли точного) аналитического выражения для порога используется эмпирический параметр V_{T_0}, который представляет собой пороговое напряжение при $V_{SD} = 0$ и является функцией главным образом процесса изготовления. В этом случае величина порогового напряжения при различных смещениях подложки может быть описана следующим выражением:

$$V_T = V_{T0} + \gamma \left(\sqrt{|(-2)\,\phi_F + V_{SB}|} - \sqrt{|2\phi_F|} \right). \tag{3.19}$$

Параметр γ (гамма) называется *коэффициентом влияния подложки*. Он отражает влияние изменений напряжения V_{SB}. Заметьте, что пороговое напряжение **положительно** для типовых ***n*-МОП**-приборов и **отрицательно** для нормальных ***p*-МОП**-транзисторов.

На рис. 3.14 показан график зависимости порогового напряжения от смещения кармана для n-МОП-транзистора при типовых значениях $|-2\phi_F| = 0,6$ В и $\gamma = 0{,}4$ $\text{В}^{0,5}$. Отрицательное смещение кармана или подложки вызывает увеличение порогового напряжения от 0,45 В до 0,85 В. Заметьте также, что для n-МОП-транзисторов напряжение V_{SB} всегда должно быть больше $-0,6$ В. Если это не так, то диод исток-подложка оказывается прямосмещенным, что нарушает работу транзистора.

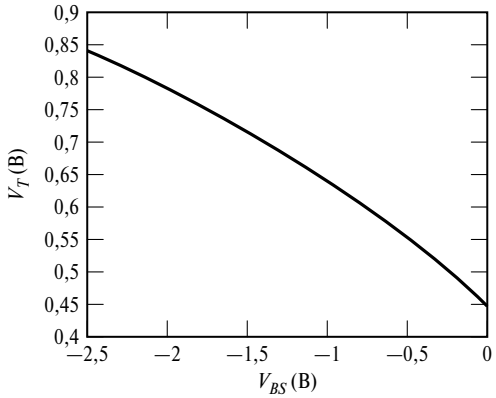


Рис. 3.14. Влияние смещения подложки на пороговое напряжение

Пример 3.5. Пороговое напряжение p-МОП-транзистора

МОП-транзистор с каналом p-типа имеет пороговое напряжение $-0,4$ В, при этом коэффициент влияния подложки равен $-0,4$. Вычислим пороговое напряжение при $V_{SB} = -2,5$ В. $2\phi_F = 0,6$ В.

Используя выражение (3.9), получаем $V_T(-2,5) = -0,4 - 0,4 \times ((2,5 + 0,6)^{0,5} - 0,6^{0,5}) = -0,79$ В, что вдвое больше, чем при нулевом смещении!

Работа в резистивном режиме

Предположим теперь, что $V_{GS} > V_T$ и что между истоком и стоком приложено небольшое напряжение V_{DS}. Разность потенциалов вызывает протекание тока I_D от стока к истоку (рис. 3.15). Используя простой анализ, можно получить выражение первого порядка функции зависимости тока от напряжений V_{GS} и V_{DS}.

В точке x вдоль канала напряжение равно $V(x)$, и напряжение затвор-канал в этой точке равно $V_{GS} - V(x)$. Если предположить, что это напряжение превышает пороговое напряжение вдоль всего канала, то величину заряда в индуцированном канале на единицу площади в точке x можно вычислить, используя следующее выражение:

$$Q_i(x) = -C_{ox}[V_{GS} - V(x) - V_T], \tag{3.20}$$

где C_{ox} обозначает емкость на единицу площади, вносимую подзатворным окислом, которая равна:

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}, \tag{3.21}$$

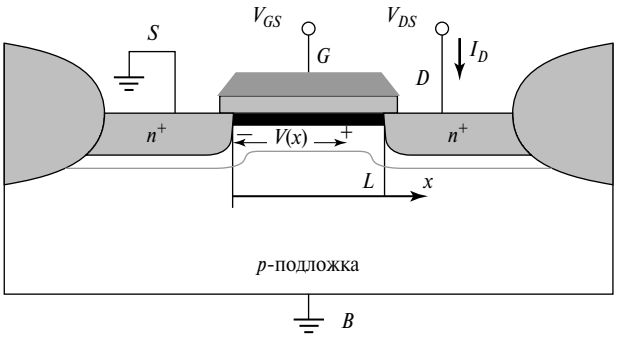


Рис. 3.15. МОП-транзистор с n-каналом при наличии напряжений смещения

где $\varepsilon_{ox} = 3,97 \times \varepsilon_0 = 3,5 \times 10^{-11}$ Ф/м — диэлектрическая проницаемость окисла; t_{ox} — толщина окисла. В современных процессах последняя составляет 10 нм (= 100 А) или меньше. При толщине окисла 5 нм емкость окисла получается равной 7 фФ/мкм^2.

Ток задается произведением дрейфовой скорости носителей v_n на имеющийся заряд. В силу принципа сохранения заряда он постоянен вдоль всей длины канала. Пусть W — это ширина канала в направлении, перпендикулярном направлению протекания тока. Тогда выражение для тока имеет вид:

$$I_D = -v_n(x)Q_i(x)W. \tag{3.22}$$

Скорость электронов связана с напряженностью электрического поля через параметр, называемый *подвижностью* μ_n (измеряется в м^2/В·с). Подвижность является сложной функцией кристаллической структуры и локального электрического поля. Как правило, для подвижности используется эмпирическое выражение:

$$v_n = -\mu_n \xi(x) = \mu_n \frac{dV}{dx}. \tag{3.23}$$

Объединим выражения (3.20)–(3.23). В результате получается уравнение:

$$I_D dx = \mu_n C_{ox} W (V_{GS} - V - V_T) dV. \tag{3.24}$$

Интегрирование по длине канала L дает выражение, связывающее ток и напряжения в транзисторе:

$$I_D = k_n' \frac{W}{L} \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] = k_n \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right]. \tag{3.25}$$

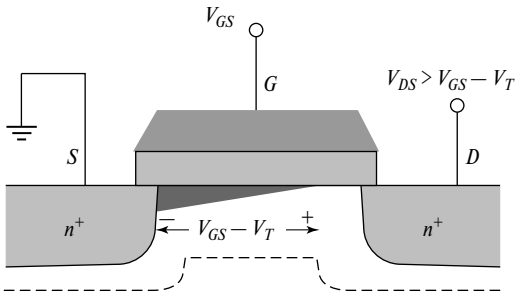


Рис. 3.16. n-МОП-транзистор в режиме смыкания канала

Здесь k'_n называется *параметром крутизны процесса* и задается выражением

$$k'_n = \mu_n C_{ox} = \frac{\mu_n \varepsilon_{ox}}{t_{ox}}. \tag{3.26}$$

Произведение крутизны процесса k'_n на отношение (W/L) n-МОП-транзистора называется *коэффициентом усиления* прибора k_n. При малых значениях V_{DS} квадратичным членом в выражении (3.23) можно пренебречь, и в результате наблюдается линейная зависимость между V_{DS} и I_D. Таким образом, рабочая область, в которой справедливо выражение (3.23), называется *резистивной* или *линейной областью*. Одним из ее основных свойств является наличие сплошного проводящего канала между областями истока и стока.

ПРИМЕЧАНИЕ.

Параметры W и L в выражении (3.23) представляют *эффективные ширину и длину канала* транзистора. Эти значения отличаются от топологических размеров из-за таких эффектов, как боковая диффузия в областях стока и истока (L) и боковое нарастание изолирующего защитного окисла (W). Далее в книге обозначения W и L всегда будут обозначать эффективные размеры, а подстрочный индекс d будет показывать, что речь идет о топологическом размере. Приведенные ниже выражения связывают два параметра, где ΔW и ΔL представляют собой параметры технологического процесса:

$$\begin{aligned} W &= W_d - \Delta W; \\ L &= L_d - \Delta L. \end{aligned} \tag{3.27}$$

Область насыщения

При дальнейшем увеличении напряжения исток-сток предположение о том, что напряжение в канале больше порогового на протяжении всей длины ка-

нала, перестает быть справедливым. Это происходит, когда $V_{GS} - V(x) < V_T$. В этой точке индуцированный заряд равен нулю, и проводящий канал исчезает или *смыкается*. Данное явление проиллюстрировано на рис. 3.16, на котором (несколько утрированно) показано постепенное уменьшение толщины канала от истока к стоку вплоть до наступления смыкания. В окрестности области стока канал уже отсутствует. Понятно, что для того, чтобы это явление имело место, необходимо, чтобы условие смыкания удовлетворялось у области стока или:

$$V_{GS} - V_{DS} \leq V_T. \tag{3.28}$$

В этих обстоятельствах транзистор переходит в область *насыщения*, и выражение (3.25) больше не является справедливым. Разность потенциалов вдоль индуцированного канала (от точки смыкания до истока) фиксируется на уровне $V_{GS} - V_T$, и соответственно ток остается постоянным по величине (или насыщается). Замена в выражении (3.25) V_{DS} на $V_{GS} - V_T$ дает значение тока стока в режиме насыщения. Стоит отметить, что в первом приближении ток больше не является функцией V_{DS}. Также заметьте, что зависимость тока стока от управляющего напряжения V_{GS} теперь носит квадратичный характер:

$$I_D = \frac{k'_n}{2} \frac{W}{L} \left(V_{GS} - V_T\right)^2. \tag{3.29}$$

Модуляция длины канала

Уравнение (3.29) предполагает, что в режиме насыщения транзистор работает как идеальный источник тока: ток между выводами истока и стока постоянен и не зависит от напряжения на выводах. Но это не совсем так. В действительности эффективная длина проводящего канала модулируется прикладываемым напряжением V_{DS}: увеличение V_{DS} вызывает рост обедненной области у стокового перехода, уменьшая тем самым эффективную длину канала. Как видно из выражения (3.29), ток увеличивается, когда длина канала L уменьшается. Поэтому более точное описание тока МОП-транзистора имеет вид

$$I_D = I'_D \left(1 + \lambda V_{DS}\right), \tag{3.30}$$

где I'_D — ранее полученное выражение для тока (т.е. выражение (3.29)); λ — эмпирический параметр, называемый *коэффициентом модуляции длины канала*. Аналитическое выражение для λ оказалось сложным и не очень точным. В общем случае величина коэффициента λ обратно пропорциональна длине канала. В более коротких транзисторах обедненная область перехода стока составляет большую долю канала, и эффект модуляции канала более заметен. Поэтому если необходимо иметь высокоимпедансный источник тока, то лучше использовать транзисторы с большой длиной канала.

Рис. 3.17. Эффект насыщения скорости

Насыщение скорости

Поведение транзисторов с очень короткими каналами (такие приборы получили название *короткоканальных*) значительно отличается от описанных выше резистивного режима и режима насыщения. Главный виновник этого — *эффект насыщения скорости*. Выражение (2.23) говорит о том, что скорость носителей пропорциональна напряженности электрического поля и не зависит от величины этой напряженности. Другими словами, подвижность носителей является константой. Однако при высоких напряженностях поля (горизонтального) носители не следуют этой линейной модели. Фактически, когда напряженность электрического поля вдоль канала достигает некоторого критического значения ξ_c, скорость носителей из-за эффектов рассеивания (столкновений, испытываемых носителями) насыщается. Это проиллюстрировано на рис. 3.17.

Величины скоростей насыщения для электронов и дырок примерно равны и составляют 10^5 м/с. Значение критической напряженности поля, при котором наступает насыщение, зависит от уровней легирования и приложенного вертикального электрического поля. Для электронов оно лежит в пределах от 1 до 5 В/мкм. Это означает, что в n-МОП-приборе с длиной канала 0,25 мкм для достижения точки насыщения достаточно иметь между стоком и истоком напряжение всего около 2 В. Данное условие с легкостью выполняется в современных короткоканальных приборах. Дыркам в кремнии n-типа для достижения насыщения требуется несколько более высокая напряженность поля. Поэтому в p-МОП-транзисторах эффект насыщения скорости проявляется в меньшей степени.

Этот эффект имеет заметное влияние на работу транзистора. Проиллюстрируем это с помощью упрощенных выкладок, позволяющих получить характеристику прибора в режиме насыщения скорости [Ko89]. Функцию зависимости скорости от напряженности электрического поля, показанную в виде графика на рис. 3.17, грубо можно аппроксимировать следующими условными равенствами:

$$\begin{aligned} v &= \frac{\mu_n \xi}{1 + \xi/\xi_c} \quad \text{при } \xi \le \xi_c \\ &= v_{sat} \quad \text{при } \xi \ge \xi_c. \end{aligned} \tag{3.31}$$

Требование неразрывности двух областей диктует выполнение равенства $\xi_c = 2v_{sat}/\mu_n$. Если теперь переписать выражения (3.20) и (3.22) с учетом измененной формулы для скорости, то это приведет к получению следующего модифицированного выражения для тока стока в резистивной области:

$$\begin{aligned} I_D &= \frac{\mu_n C_{ox}}{1 + (V_{DS}/\xi_c L)} \left(\frac{W}{L}\right) \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] = \\ &= \mu_n C_{ox} \left(\frac{W}{L}\right) \left[(V_{GS} - V_T) V_{DS} - \frac{V_{DS}^2}{2} \right] \kappa (V_{DS}) . \end{aligned} \tag{3.32}$$

Сомножитель $\kappa(V)$ представляет собой меру степени насыщения скорости и определяется следующим образом:

$$\kappa(V) = \frac{1}{1 + (V/\xi_c L)}. \tag{3.33}$$

Отношение V_{DS}/L можно интерпретировать как среднюю напряженность поля в канале. В случае приборов с длинным каналом (большое значение L) или при малых значениях V_{DS} коэффициент κ приближается к 1, и выражение (3.32) упрощается до традиционного уравнения тока в резистивном режиме работы. Для короткоканальных транзисторов значение κ меньше 1, и это означает, что обеспечиваемый ток будет меньше нормально ожидаемого.

С увеличением напряжения исток-сток напряженность электрического поля в канале в конце концов достигает критического значения, и носители у стока приобретают скорость, соответствующую скорости насыщения. Напряжение стока насыщения V_{DSAT} можно вычислить, приравняв ток стока при насыщении к току, задаваемому выражением (3.32), при $V_{DS} = V_{DAST}$ [Toh88]. Первый вычисляется с использованием выражения (3.22) в предположении, что дрейфовая скорость находится в области насыщения и равна v_{sat}. Имеем:

$$\begin{aligned} I_{DSAT} &= v_{sat} C_{ox} W (V_{GT} - V_{DSAT}) = \\ &= \kappa(V_{DSAT}) \mu_n C_{ox} \frac{W}{L} \left[V_{GT} V_{DSAT} - \frac{V_{DSAT}^2}{2} \right], \end{aligned} \tag{3.34}$$

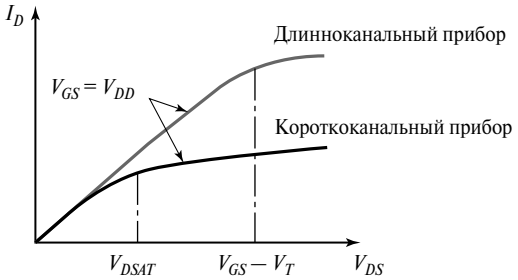


Рис. 3.18. Короткоканальные устройства обладают расширенной областью насыщения

где V_{GT} представляет собой сокращенное обозначение разности $V_{GS} - V_T$. После сокращений получаем:

$$V_{DSAT} = \kappa(V_{GT})V_{GT}. \tag{3.35}$$

Дальнейшее увеличение напряжения исток-сток не приводит к получению дополнительного тока (в первом приближении), и ток транзистора насыщается на уровне I_{DSAT}. Это приводит к двум наблюдениям.

- Для короткоканального транзистора и для достаточно больших значений V_{GT} значение $\kappa(V_{GT})$ существенно меньше 1, и, таким образом, $V_{DSAT} < V_{GT}$. Прибор входит в насыщение до того, как напряжение V_{DS} достигает уровня $V_{GS} - V_T$. Поэтому у короткоканального прибора наблюдается наличие протяженной области насыщения, и он более часто работает в режиме насыщения, чем его длинноканальный аналог (рис. 3.18).
- Ток насыщения I_{DAST} имеет *линейную зависимость* от напряжения затвор-исток V_{GS}, тогда как у длинноканального прибора она носит квадратичный характер. Это снижает величину тока, которую может обеспечить транзистор при заданном значении напряжения управления. С другой стороны, в субмикронных приборах снижение рабочего напряжения не оказывает такого заметного влияния на работу, как это имеет место в длинноканальном транзисторе.

В предыдущих выражениях игнорируется тот факт, что при дальнейшем увеличении напряжения V_{DS} все в большей части канала скорость носителей достигает насыщения. С точки зрения моделирования это выглядит так, словно с увеличением V_{DS} эффективная длина канала уменьшается подобно тому, как это происходит в результате эффекта модулирования длины канала. Возникающее в результате увеличение тока легко учитывается введением дополнительного множителя $(1 + \lambda V_{DS})$. Тем, кого интересует более углубленное описание эффектов короткого канала в МОП-транзисторах, можно порекомендовать обратиться, среди прочих, к [Ko89][2].

[2] Ради простоты в данном разделе сознательно было опущено рассмотрение такого эффекта, как снижение подвижности из-за наличия вертикального электрического поля. Авторы посчитали, что это можно сделать без ущерба для последующего рассмотрения вопросов моделирования.

Насыщение скорости: возвращение к теме

К сожалению, описывающие поведение тока стока уравнения (3.32) и (3.34) представляют собой сложные выражения относительно напряжений VGS и VDS, что делает их довольно громоздкими для выполнения приближенного анализа вручную. Значительно более простую модель можно получить, если сделать два предположения.

1. Скорость носителей насыщается резко при достижении напряженности поля ξ_c, и ее поведение описывается следующими выражениями:

$$\begin{aligned} v &= \mu_n \xi && \text{при } \xi < \xi_c \\ &= v_{sat} = \mu_n \xi_c && \text{при } \xi \geq \xi_c. \end{aligned} \tag{3.36}$$

2. Напряжение исток-сток V_{DSAT}, при котором достигается критическая величина напряженности электрического поля и в игру вступает эффект насыщения скорости, является *константой* и приближенно описывается выражением вида:

$$V_{DSAT} = L\xi_c = \frac{Lv_{sat}}{\mu_n}. \tag{3.37}$$

Из выражения (3.35) видно, что при больших значениях V_{GT} это предположение выглядит разумным.

В таких обстоятельствах уравнения тока для резистивной области остаются неизменными относительно модели длинноканального транзистора. Как только достигается значение напряжения V_{DSAT}, ток резко насыщается. Величина тока I_{DSAT} в этой точке может быть вычислена путем подстановки в уравнение тока для резистивной области значения напряжения насыщения (см. выражение (3.25)):

$$\begin{aligned} I_{DSAT} &= I_D(V_{DS} = V_{DSAT}) = \\ &= \mu_n C_{ox} \frac{W}{L} \left[(V_{GS} - V_T) V_{DSAT} - \frac{V_{DSAT}^2}{2} \right] = \\ &= v_{sat} C_{ox} W \left(V_{GS} - V_T - \frac{V_{DSAT}}{2} \right). \end{aligned} \tag{3.38}$$

Конечно, эта модель дает результат только в первом приближении и носит эмпирический характер. Упрощенное моделирование поведения скорости приводит к значительным отклонениям в переходной зоне между линейной областью и областью насыщения. Тем не менее, если тщательно подобрать значения параметров модели, то можно получить приличное согласование с эмпирическими данными для других рабочих областей, что и будет показано позднее. Что важно, так это схожесть уравнений со знакомыми выражениями для модели транзистора с длинным каналом, и кроме того, они дают

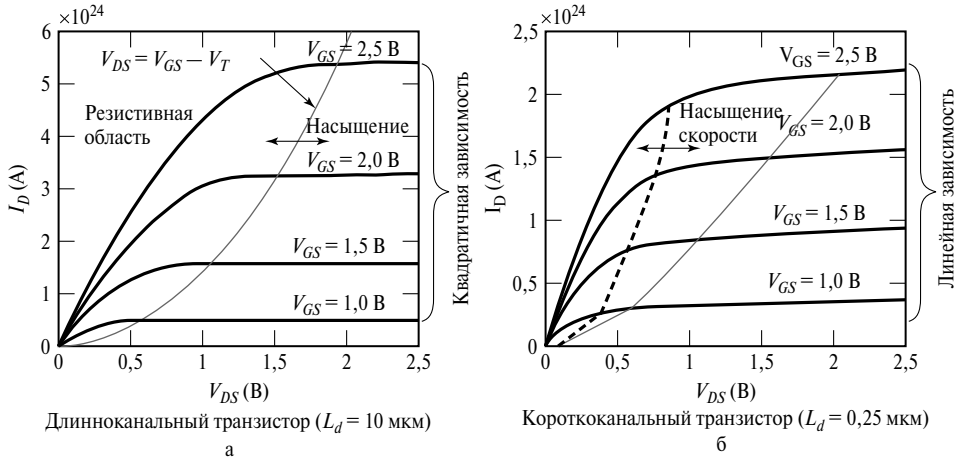


Рис. 3.19. ВАХ длинно- и короткоканального n-МОП-транзисторов, изготовленных по 0,25-микронной КМОП-технологии. Значение отношения (W/L у обоих транзисторов одинаково и равно 1,5. Обратите внимание на разницу масштабов по оси y

в руки разработчику цифровых схем так необходимый ему инструмент для интуитивного понимания и интерпретации получаемых результатов.

Графики зависимости тока стока от напряжений

Поведение МОП-транзистора в различных режимах работы лучше всего можно понять, если проанализировать кривые зависимости $I_D - V_{DS}$, которые представляют собой графики зависимости тока стока I_D от напряжения исток-сток V_{DS} при напряжении исток-затвор V_{GS} в качестве параметра. На рис. 3.19 показан вид этих кривых для двух n-МОП-транзисторов, изготовленных по одной технологии и с одинаковым значением отношения W/L. Можно было бы ожидать, что у обоих приборов будут одинаковые вольт-амперные характеристики (ВАХ). Однако основное отличие состоит в том, что первый прибор имеет длинный канал ($L = 10$ мкм), тогда как второй транзистор является короткоканальным ($L = 0,25$ мкм), и в нем проявляется эффект насыщения скорости.

Рассмотрим сначала длинноканальный прибор. В резистивной области транзистор ведет себя как управляемый напряжением резистор, тогда как в области насыщения он уже выступает в роли управляемого напряжением источника тока (если проигнорировать эффект модуляции длины канала). Переход между двумя областями разграничивается кривой $V_{DS} = V_{GS} - V_T$. Квадратичная зависимость I_D от V_{GS} в области насыщения, что типично для длинноканального прибора, ясно просматривается по расстоянию между различными кривыми семейства. У короткоканального же прибора, характеристики которого показаны на рис. 3.19, *б*, очевидна линейная зависимость тока насыщения от V_{GS}. Отметим также, как эффект насыщения скорости приводит к переходу прибора в режим насыщения при значительно мень-

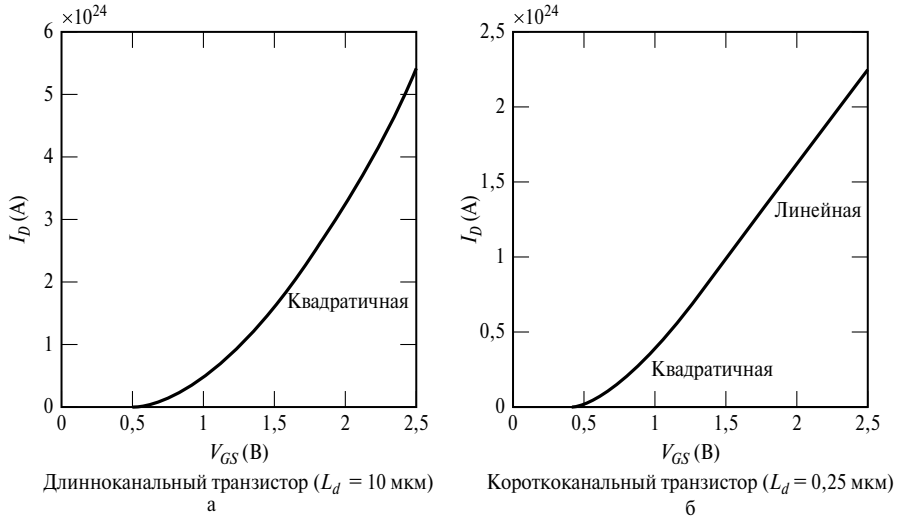


Рис. 3.20. $I_D - V_{GS}$-характеристики n-МОП-транзисторов с длинным и коротким каналом (0,25-микронная КМОП-технология). Для обоих транзисторов $W/L = 1,5$ и $V_{DS} = 2,5$ В

ших значениях V_{DS}. Штриховой линией показана граница зоны, в которой согласно выражению (3.29) начинает вносить свой вклад эффект насыщения скорости, тогда как тонкой линией отмечена зона начала традиционного насыщения ($V_{DS} = V_{GT}$). Это приводит к значительному снижению уровня тока при более высоких напряжениях. Например, при $V_{GS} = 2,5$ В, $V_{DS} = 2,5$ В значение тока стока короткоканального транзистора составляет всего 40% от соответствующей величины для прибора с более длинным каналом (220 мкА по сравнению с 540 мкА).

Отличие характера зависимости от V_{GS} для длинно- и короткоканальных приборов еще более заметно, если взять другой набор смоделированных кривых, представляющих собой графики зависимости тока стока I_D в функции напряжения исток-затвор V_{GS} при напряжении исток-сток V_{DS}, гарантирующем насыщение (рис. 3.20). При больших значениях V_{GS} ясно видны квадратичная и линейная зависимости.

Все полученные уравнения справедливы также и для p-МОП-транзисторов. Единственное отличие состоит в том, что **для p-МОП-транзисторов полярности всех напряжений и токов противоположны**. Это иллюстрируется на рис. 3.21, на котором показаны $I_D - V_{DS}$-характеристики p-МОП-транзистора с минимальными размерами, изготовленного с использованием 0,25-микронного КМОП-процесса. Все кривые расположены в третьем квадранте, так как I_D, V_{DS} и V_{GS} отрицательны. Заметьте, что в данном случае влияние эффекта насыщения скорости менее заметно, чем в n-МОП-приборах. Это можно отнести на счет более высокого значения критической напряженности поля из-за меньшей подвижности дырок по сравнению с электронами.

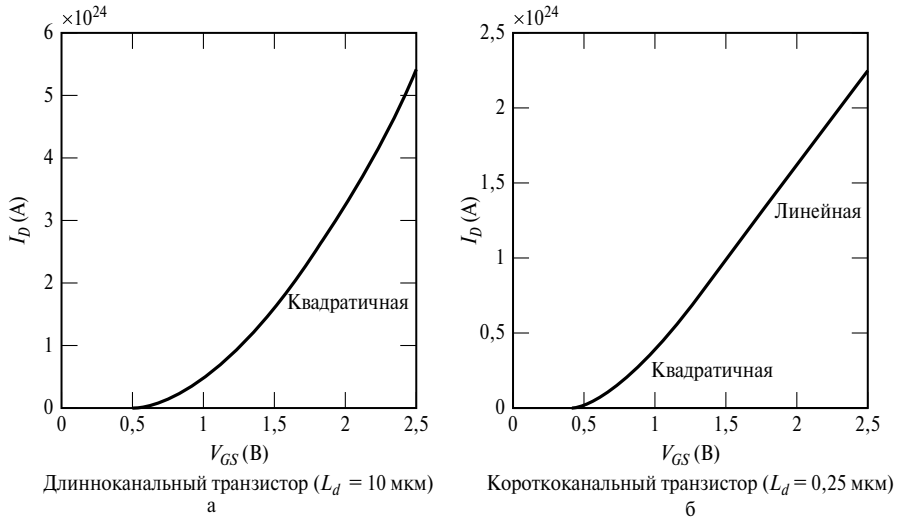


Рис. 3.21. ВАХ ($W_d = 0,375$ мкм, $L_d = 0,25$ мкм) p-МОП-транзистора (0,25-микронная КМОП-технология). Максимальный ток составляет всего 40% от того, что дает аналогичный n-МОП-транзистор

Проводимость в предпороговой области

Более внимательное рассмотрение $I_D - V_{DS}$-кривых позволяет выявить тот факт, что при $V_{GS} = V_T$ ток вовсе не падает резко до 0. Другими словами, становится очевидным, что при напряжениях ниже порогового МОП-транзистор уже частично проводит ток. Этот эффект получил название *предпороговой* или *слабой инверсии*. Начало сильной инверсии означает, что имеется большое количество носителей, которые могут участвовать в проводимости, но это совсем не подразумевает, что при напряжениях затвор-исток ниже V_T совсем не может течь ток, хотя в подобных условиях уровни тока малы. Резюмируя, можно сказать, что переход транзистора из "включенного" состояния в "выключенное" является не резким, а постепенным.

Чтобы исследовать этот эффект более подробно, перерисуем приведенную на рис. 3.20, *б* кривую в логарифмическом масштабе, как показано на рис. 3.22. Это подтверждает, что ток не падает до 0 сразу после выполнения условия $V_{GS} < V_T$, а спадает по экспоненте подобно тому, как это имеет место в биполярном транзисторе[3]. В отсутствие проводящего канала n^+ (исток)-p (подложка)-n^+ (сток) области образуют паразитный биполярный транзистор. Ток в этой области приблизительно может быть описан следующим выражением:

$$I_D = I_S e^{\frac{V_{GS}}{nkT/q}} \left(1 - e^{-\frac{V_{DS}}{kT/q}}\right)(1 + \lambda V_{DS}) \qquad (3.39)$$

[3] Более подробную информацию о биполярных транзисторах можно найти на Web-сайте данной книги.

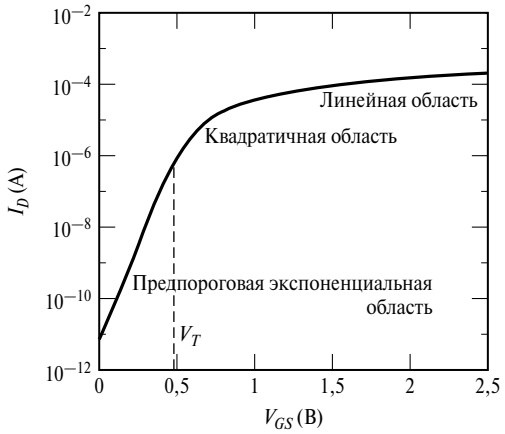


Рис. 3.22. Зависимость тока I_D от напряжения V_{GS} (в логарифмическом масштабе), из которой виден экспоненциальный характер поведения тока в предпороговой области

где I_S и n — эмпирические параметры, при этом $n \geq 1$ и обычно имеет значение около 1,5.

Для большинства цифровых применений наличие предпорогового тока нежелательно, поскольку это является отклонением от идеального ключевого поведения, которое хотелось бы иметь у МОП-транзистора. Предпочтительно, чтобы после уменьшения напряжения исток-затвор ниже V_T ток спадал как можно быстрее. Таким образом, величина (обратная) скорости спада тока при V_{GS} ниже V_T является мерой качества прибора. Количественно этот показатель выражается через *коэффициент спада* S, который показывает, на сколько надо уменьшить напряжение V_{GS}, чтобы ток упал в 10 раз. Из выражения (3.39) находим:

$$S = n\left(\frac{kT}{q}\right)\ln(10), \tag{3.40}$$

где S выражается в мВ/порядок. У идеального транзистора с самым крутым спадом, какой только возможен, $n = 1$, и оценка величины $(kT/q)ln(10)$ при комнатной температуре дает значение 60 мВ/порядок. Это означает, что предпороговый ток уменьшается в 10 раз при уменьшении напряжения V_{GS} на 60 мВ. К сожалению, у реальных приборов значение n больше 1, и ток спадает с меньшей скоростью (90 мВ/порядок при $n = 1,5$). Скорость спада еще больше уменьшается с ростом рабочей температуры; большинство интегральных схем работают при температурах, значительно превышающих комнатную. Значение n определяется топологией прибора и структурой. Поэтому для его уменьшения требуется использование иных технологических процессов, например технологии “кремний на диэлектрике”.

Наличие предпорогового тока имеет несколько важных последствий. В общем случае крайне желательно, чтобы при $V_{GS} = 0$ ток через транзистор был как можно ближе к нулю. Это особенно важно в *динамических схемах*, принцип работы которых основан на хранении заряда, и само их функционирование может нарушаться из-за предпороговой утечки. В результате наличие предпорогового тока требует наложения жестких требований на положение нижней границы значений пороговых напряжений приборов.

Пример 3.6. Спад в предпороговой области

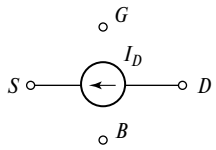
В примере, показанном на рис. 3.22, значение коэффициента спада составляет 89,5 мВ/порядок (в диапазоне напряжений от 0,2 до 0,4 В). Это эквивалентно значению n-фактора 1,49.

Модели для анализа вручную

Из вышеописанного материала понятно, что субмикронный транзистор представляет собой чрезвычайно сложный прибор. Его поведение существенно нелинейно и определяется большим количеством эффектов второго порядка. К счастью, были разработаны точные модели для моделирования схем, которые сделали возможным удивительно точное прогнозирование поведения приборов с разнообразными размерами, формой и в разных режимах работы, что и будет обсуждаться в этой главе позднее. Но эти модели, хотя и превосходны с точки зрения точности, не дают разработчику интуитивного понимания поведения схемы и того инструмента, который бы позволил выявлять ключевые параметры разработки. Такое понимание необходимо в процессе анализа и оптимизации разрабатываемой микросхемы. Разработчик, у которого нет четкого видения того, что определяет и управляет работоспособностью схемы, по необходимости должен полагаться на длительный процесс оптимизации методом проб и ошибок, который часто дает ужасное решение.

Но тут возникает вопрос: как втиснуть поведение МОП-транзистора в простую и понятную аналитическую модель, которая не приводит к безнадежно сложным уравнениям и тем не менее описывает все наиболее существенные детали поведения прибора? Оказывается, что выведенные выше выражения можно объединить в одно, которое будет удовлетворять поставленным целям. Модель представляет транзистор в виде одного источника тока (рис. 3.23), величина которого задается выражениями, приведенными на рисунке. Читатель может сам проверить, что в зависимости от рабочих условий модель сводится либо к выражению (3.23), либо к выражению (3.29), либо к выражению (3.31) (скорректированному с учетом модулирования длины канала)[4]. Таким образом, мы имеем:

[4] Унифицированная модель также включает в себя коррекцию на модулирование длины канала и в линейной области. Этот коэффициент, который сам по себе мал, служит исключительно для целей моделирования и не несет в себе никакого физического смысла. Следует также помнить, что функция min для p-МОП-транзистора превращается в операцию max.



$I_D = 0$ при $V_{GT} \le 0$

$I_D = k' \frac{W}{L}\left(V_{GT} V_{min} - \frac{V^2_{min}}{2}\right)(1 + \lambda V_{DS})$ при $V_{GT} \ge 0$

где $V_{min} = \min(V_{GT}, V_{DS}, V_{DSAT})$,

$V_{GT} = V_{GS} - V_T$,

и $V_T = V_{T0} + \gamma(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|})$

Рис. 3.23. Унифицированная модель МОП-транзистора для выполнения анализа вручную

$$I_D = 0 \text{ при } \quad V_{GT} \le 0;$$
$$I_D = k' \frac{W}{L}\left(V_{GT} V_{\min} - \frac{V^2_{\min}}{2}\right)(1 + \lambda V_{DS}) \text{ при } \quad V_{GT} \ge 0,$$

где $V_{\min} = \min(V_{GT}, V_{DS}, V_{DSAT})$,

$$V_{GT} = V_{GS} - V_T$$

и $V_T = V_{T0} + \gamma\left(\sqrt{|-2\phi_F + V_{SB}|} - \sqrt{|-2\phi_F|}\right)$.

На рис. 3.24 показано, как **унифицированная модель** делит общее рабочее пространство транзистора на три области, линейную, насыщения скорости и насыщения, и идентифицирует границы между ними.

Помимо того, что модель представляет собой функцию напряжений на четырех выводах транзистора, в ней еще используется набор из пяти параметров: V_{T0}, γ, V_{DSAT}, k' и λ. В принципе значения этих параметров можно определить, исходя из технологического процесса и уравнений, описывающих физику работы прибора. Однако сложность прибора делает результат такой работы весьма сомнительным. Более разумно выбрать значения так, чтобы получалось хорошее согласование с реальными характеристиками прибора. И что наиболее важно, модель должна давать наилучшее согласование в тех областях, которые представляют собой наибольший интерес. Для цифровых схем это область высоких значений V_{GS} и V_{DS}. Быстродействие цифровых МОП-схем определяется главным образом максимально возможным током (т.е. током, получаемым при $V_{GS} = V_{DS} =$ напряжение источника питания). Поэтому важно иметь хорошее согласование именно для этой области. **Наконец, заметьте, что для типового *n*-МОП-прибора все пять параметров положительны, тогда как для типового *p*-МОП-транзистора они отрицательны**.

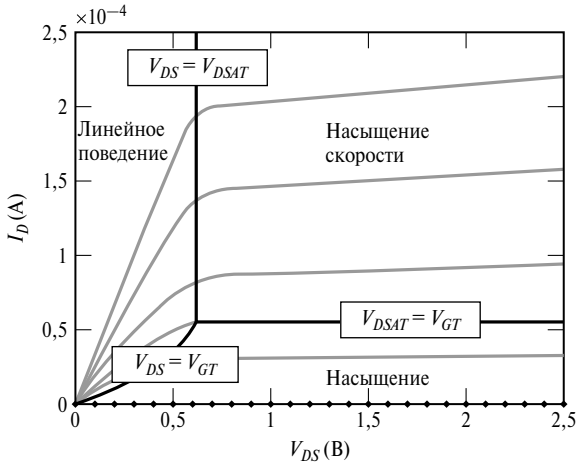


Рис. 3.24. Границы рабочих областей в соответствии с тем, как они задаются моделью для ручного анализа

Пример 3.7. Модель для ручного анализа транзисторов, получаемых по 0,25-микронной КМОП-технологии

(Реализацию модели с использованием MATLAB можно найти на Web-сайте данной книги.) На основе полученного моделированием ВАХ $I_D = f(V_{DS})$ и $I_D = f(V_{GS})$ (см. рис. 3.19 и 3.20) транзистора, реализуемого с использованием нашего базового 0,25-микронного КМОП-процесса ($W = 0,375$ мкм, $L = 0,25$ мкм), были подобраны значения параметров прибора, обеспечивающие хорошее согласование в области напряжений $V_{DS} = 2,5$ В, $V_{GS} = 2,5$ В; 2,5 В — типовое значение напряжения источника питания для данного процесса. Получающиеся в результате характеристики n-МОП-транзистора показаны на рис. 3.25, где они также сравниваются со значениями, получаемыми в результате моделирования. Можно видеть хорошее соответствие везде, за исключением переходной области между резистивным режимом и режимом с насыщением скорости. Подобное расхождение, являющееся результатом использования простой модели поведения скорости, описываемой выражением (3.36), приемлемо, поскольку имеет место в диапазоне более низких значений V_{DS}. Таким образом, наша модель, хотя и является простой, все же способна дать приличное описание общего поведения прибора.

Модель транзистора для ручного анализа

В табл. 3.2 сведены полученные значения параметров для n-МОП- и p-МОП-транзисторов минимального размера, получаемых с использованием нашего типового 0,25-микронного КМОП-процесса. В последующих главах эти значения будут использоваться в качестве типовых значений параметров модели.

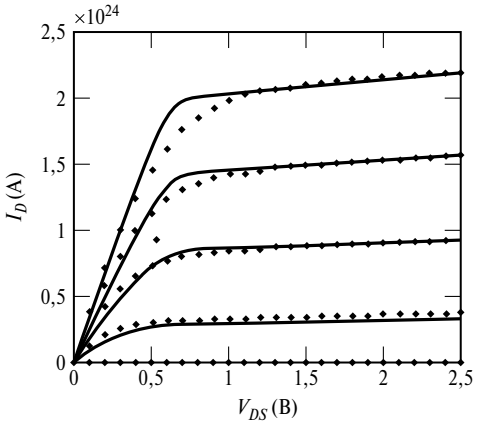


Рис. 3.25. Соответствие результатов моделирования n-МОП-транзистора минимального размера ($W = 0,375$ мкм, $L = 0,25$ мкм) с использованием простой модели (сплошная линия) и SPICE-модели (точечная линия). Видно расхождение в зоне перехода от резистивного режима в режим с насыщением скорости

Таблица 3.2. Параметры модели для ручного анализа транзисторов (минимального размера), получаемых с использованием типового 0,25-микронного КМОП-процесса

Параметр	V_{T0}, В	γ, $В^{0,5}$	V_{DSAT}, В	k', $А/В^2$	λ, $В^{-1}$
n-МОП	$0,43$	$0,4$	$0,63$	115×10^{-6}	$0,06$
p-МОП	$-0,4$	$-0,4$	-1	-30×10^{-6}	$-0,1$

Предостережение. Представленная здесь модель была выведена на основании характеристик единичного прибора с минимальными длиной и шириной канала. Попытка экстраполяции его поведения на приборы с существенно отличающимися значениями W и L, скорее всего, приведет к заметным ошибкам. К счастью, в цифровых схемах обычно используются приборы с минимальными размерами, поскольку они обеспечивают минимальность площади микросхемы. Согласование для таких транзисторов обычно достаточно приемлемое. Однако для приборов с заметно отличающимися ширинами каналов все-таки желательно использовать другой набор значений параметров.

Представленная модель в виде *источника тока* будет весьма полезной при анализе основных свойств и метрик простого цифрового вентиля, но все же ее нелинейность делает ее непривлекательной для чего-либо более сложного. Поэтому вводится еще более простая модель, которая обеспечивает как линейность, так полную прозрачность. Она основана на основополагающем предположении для всех разработок цифровых схем, что транзистор представляет собой не что иное, как *ключ* с бесконечным сопротивлением в со-

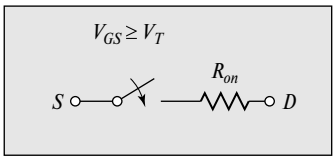


Рис. 3.26. Модель n-МОП-транзистора в виде ключа

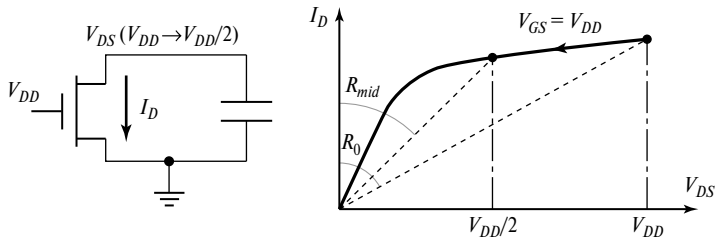
стоянии "выключено" и конечным сопротивлением в состоянии "включено" (рис. 3.26).

Основная проблема с этой моделью состоит в том, что R_{on} (сопротивление во включенном состоянии) меняется во времени, нелинейно и зависит от рабочей точки транзистора. При исследовании цифровых схем в переходном режиме, т.е. во время переключения из одного логического состояния в другое, заманчиво предположить, что R_{on} представляет собой постоянное и линейное сопротивление R_{eq}, величина которого выбирается такой, чтобы конечный результат был подобен получаемому с исходным транзистором. В этом отношении разумным подходом будет использование среднего значения сопротивления в рассматриваемой рабочей области, или еще проще — среднее значение величин сопротивления в конечных точках перехода. Последнее предположение справедливо в том случае, если сопротивление не проявляет сильной нелинейности в интервале усреднения. Соответственно, можно записать:

$$R_{eq} = \text{среднее}_{t=t_1 \ldots t_2} \left(R_{on}(t)\right) = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} R_{on}(t)dt = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \frac{V_{DS}(t)}{I_D(t)} dt \approx$$
$$\approx \frac{1}{2} \left(R_{on}(t_1) + R_{on}(t_2)\right). \qquad (3.41)$$

Пример 3.8. Эквивалентное сопротивление при разряде/заряде емкости

Один из наиболее часто встречающихся сценариев в современных цифровых схемах — это разряд конденсатора от напряжения V_{DD} до напряжения земли GND через n-МОП-транзистор при напряжении на его затворе V_{DD} или наоборот — заряд конденсатора до напряжении V_{DD} через p-МОП-транзистор при напряжении на его затворе, равном напряжению земли GND. Особый интерес представляет собой точка, в которой напряжение на конденсаторе достигает половины напряжения питания ($V_{DD}/2$), что связано, если вспомнить, с определением задержки распространения, приведенным в главе 2. Предполагая, что величина напряжения источника питания значительно выше значения напряжения насыщения скорости V_{DSAT}, будет справедливо утверждать, что в течение всего переходного процесса транзистор остается в режиме насыщения скорости. График подобного сценария для случая разряда конденсатора от напряжения V_{DD} до напряжения $V_{DD}/2$ через n-МОП-транзистор показан на рис. 3.27.



Схематическое представление
а

Траектория движения по I_D-V_{DS} кривой
б

Рис. 3.27. Разряд конденсатора через n-МОП-транзистор: а) принципиальная схема; б) траектория движения рабочей точки по вольт-амперной кривой. Мгновенное значение сопротивления транзистора равно (V_{DS}/I_D) и визуально представляет собой угол по отношению к оси y

С помощью выражения (3.41) можно вывести формулу для эквивалентного сопротивления, представляющего собой среднее сопротивление прибора в рассматриваемом интервале:

$$R_{eq} = \frac{1}{-V_{DD}/2} \int\limits_{V_{DD}}^{V_{DD}/2} \frac{V}{I_{DSAT}(1+\lambda V)} dV \approx \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left(1 - \frac{7}{9}\lambda V_{DD}\right),$$
$$\text{где } I_{DSAT} = k' \frac{W}{L} \left((V_{DD} - V_T) V_{DSAT} - \frac{V_{DSAT}^2}{2} \right). \tag{3.42}$$

Подобный результат можно получить и просто путем усреднения значений сопротивления в конечных точках переходной области (упростив полученное выражение путем разложения в ряд Тейлора):

$$R_{eq} = \frac{1}{2} \left(\frac{V_{DD}}{I_{DSAT}(1 + \lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1 + \lambda V_{DD}/2)} \right) \approx$$
$$\approx \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left(1 - \frac{5}{6}\lambda V_{DD}\right). \tag{3.43}$$

На основании этих выражений можно сделать три важных вывода.

1. Величина сопротивления обратно пропорциональна отношению (W/L) прибора. Удвоение ширины транзистора приводит к уменьшению его сопротивления вдвое.
2. При $V_{DD} \gg V_T + V_{DSAT}/2$ сопротивление становится фактически независящим от напряжения источника питания. Это подтверждается рис. 3.28, на котором показан график (по результатам моделирования) зависимости эквивалентного сопротивления в функции напряжения источника питания V_{DD}. С ростом напряжения наблюдается только весьма незначительное улучшение величины сопротивления за счет эффекта модуляции длины канала.

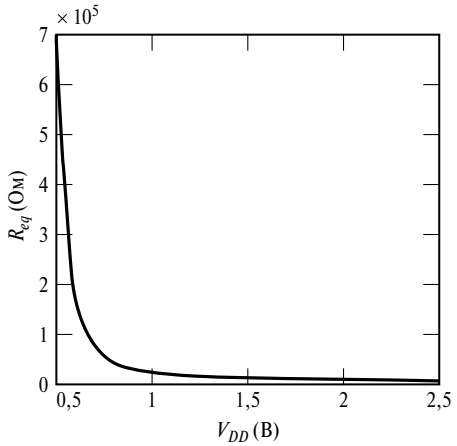


Рис. 3.28. Результат моделирования величины эквивалентного сопротивления n-МОП-транзистора минимального размера, получаемого с использованием $0,25$-микронного КМОП-процесса, в функции $V_{DD}(V_{GS} = V_{DD}, V_{DS} = V_{DD} \rightarrow V_{DD}/2)$

3. При достижении напряжения питания величины V_T величина сопротивления резко возрастает.

Модель эквивалентного сопротивления

В табл. 3.3 приведены полученные с помощью моделирования числовые значения эквивалентного сопротивления транзисторов для нашего типового 0,25-микронного КМОП-процесса. Эти данные будут полезны при анализе быстродействия КМОП-вентилей в последующих главах.

Таблица 3.3. Эквивалентное сопротивление R_{eq} $(W/L = 1)$ n-МОП и p-МОП-транзисторов, получаемых в результате 0,25-микронного КМОП-процесса (при $L = L_{min}$ по чертежу). Для больших транзисторов R_{eq} делить на W/L

V_{DD}, В	1	1,5	2	2,5
n-МОП, кОм	35	19	15	13
p-МОП, кОм	115	55	38	31

Динамическое поведение

Динамическая реакция полевого МОП-тразистора является исключительно функцией времени, требующегося для разряда/заряда собственных паразитных емкостей прибора и дополнительной емкости, вносимой линиями межсоединений и нагрузкой (рассматривается в главе 4). Четкое представление о природе и поведении собственных емкостей весьма важно для разработчика высококачественных цифровых интегральных схем. Существуют три источника их возникновения: базовая МОП-структура, заряд в канале и обеднен-

ные области обратносмещенных pn-переходов стока и истока. За исключением емкостей МОП-структуры все остальные конденсаторы нелинейны, и их емкость изменяется в зависимости от прикладываемого напряжения, что делает их анализ трудной задачей. В последующих разделах все составляющие будут рассмотрены отдельно.

Емкости МОП-структуры

Затвор МОП-транзистора изолирован от проводящего канала подзатворным окислом, который обладает емкостью на единицу площади, равной $C_{ox} = \varepsilon_{ox}/t_{ox}$. Раньше уже было сказано, что с точки зрения ВАХ величину C_{ox} полезно иметь как можно большей, или, другими словами, необходимо делать толщину окисла очень малой. Общая величина этой емкости называется емкостью затвора C_g и может быть разбита на две составляющие, каждая из которых ведет себя по-разному. Очевидно, что одна часть C_g вносит свой вклад в заряд канала. Вторая часть зависит исключительно от топологической структуры транзистора, и рассматривается ниже в настоящем разделе.

Рассмотрим транзисторную структуру, изображенную на рис. 3.29. В идеале диффузионные области истока и стока должны заканчиваться непосредственно под границей окисла затвора. В реальности же как исток, так и сток обычно заходят под окисел на расстояние x_d, называемое *боковой диффузией*. Поэтому эффективная длина канала транзистора L становится меньшей, чем длина по чертежу L_d (или исходно запроектированной длины транзистора), на величину $\Delta L = 2x_d$. Это также приводит к увеличению паразитной емкости между затвором и истоком (стоком), которая носит название *емкости перекрытия*. Эта емкость линейна и имеет фиксированное значение:

$$C_{GSO} = C_{GDO} = C_{ox}x_dW = C_oW. \tag{3.44}$$

Поскольку величина x_d определяется технологией, то принято объединять ее с емкостью окисла в параметр, называемый емкостью перекрытия на единицу длины транзистора C_o (если быть более точным, то в параметры C_{gso} и C_{gdo}).

Емкость канала

Вероятно, наиболее значимый паразитный элемент МОП-структуры — емкость затвор-канал C_{GC} — изменяется как по величине, так в плане перераспределения между тремя ее составляющими C_{GCS}, C_{GCD} и C_{GCB} (емкости затвор-исток, затвор-сток и затвор-подложка соответственно) в зависимости от рабочей области и напряжений на выводах транзистора. Изменение перераспределения поясняется на простых схемах, изображенных на рис. 3.30. Когда транзистор находится в области отсечки (см. рис. 3.30, *а*), канал отсутствует, и общая емкость C_{GC} определяется емкостью между затвором и подложкой. С переходом в резистивную область (см. рис. 3.30, *б*) формируется инверсный слой, который играет роль проводника между истоком и стоком. Как следствие, $C_{GCB} = 0$, поскольку электрод подложки экранируется от

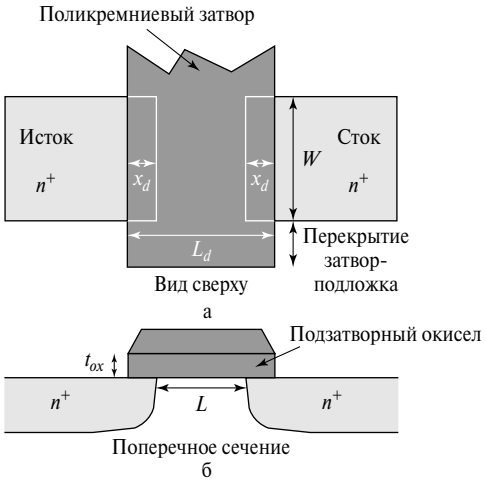


Рис. 3.29. Емкости перекрытия полевого МОП-транзистора

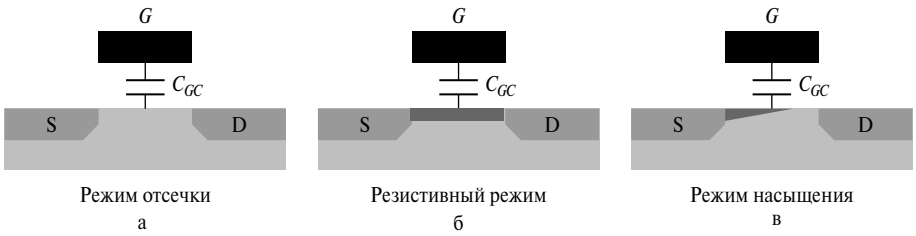


Рис. 3.30. Емкость затвор-канал и влияние режима работы на ее перераспределение относительно трех других выводов прибора

затвора каналом. Симметрия требует равномерного распределения емкости между истоком и стоком. Наконец, в режиме насыщения (см. рис. 3.30, *в*) канал смыкается. Емкость между затвором и стоком приблизительно равна нулю, как и емкость затвор-подложка. Поэтому вся емкость сосредоточивается между затвором и истоком.

Реальную величину общей емкости затвор-канал и ее перераспределение между тремя составляющими более наглядно можно проследить с помощью нескольких графиков. На первом графике (рис. 3.31, *а*) представлена зависимость емкости в функции V_{GS} при $V_{DS} = 0$. Когда $V_{GS} = 0$, транзистор находится в выключенном состоянии, канал отсутствует и общая емкость, равная WLC_{ox}, сосредоточена между затвором и подложкой. С увеличением V_{GS} под затвором формируется обедненная область. Это выглядит как увеличение толщины подзатворного диэлектрика, что приводит к уменьшению емкости. Как только транзистор открывается ($V_{GS} = V_T$), образуется канал и емкость C_{GCB} падает до 0. При $V_{DS} = 0$ прибор работает в резистивном режиме, и емкость делится поровну между истоком и стоком ($C_{GCS} = C_{GCD} =_{ox} /2$).

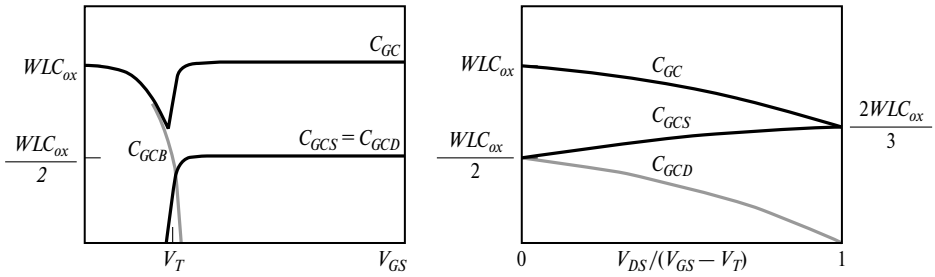


Рис. 3.31. Перераспределение емкости затвор-канал в функции V_{GS} и V_{DS} (из [Dally98])

Таблица 3.4. Среднее распределение емкости канала МОП-транзистора в различных рабочих областях

Рабочая область	C_{GCB}	C_{GCS}	C_{GCD}	C_{GC}	C_G
Отсечки	$C_{ox}WL$	0	0	$C_{ox}WL$	$C_{ox}WL + 2C_oW$
Резистивная	0	$C_{ox}WL/2$	$C_{ox}WL/2$	$C_{ox}WL$	$C_{ox}WL + 2C_oW$
Насыщения	0	$(2/3)C_{ox}WL$	0	$(2/3)C_{ox}WL$	$(2/3)C_{ox}WL+2C_oW$

Стоит отложить в памяти наличие сильной флуктуации емкости канала в окрестности $V_{GS} = V_T$. Разработчик, которого интересует хорошая линейная емкость, должен избегать работы в этой области.

Как только транзистор включается, распределение емкости затвора начинает зависеть от степени насыщения, измеряемой отношением $V_{DS}/(V_{GS}-V_T)$. Как показано на рис. 3.31, *б*, емкость C_{GCD} с увеличением степени насыщения постепенно падает до 0, тогда как емкость C_{GCS} увеличивается до уровня $2/3\ C_{ox}WL$. Это также означает, что общая емкость затвора уменьшается с увеличением уровня насыщения.

Из изложенного выше понятно, что составляющие емкости затвора нелинейны и меняют свою величину в зависимости от рабочих напряжений. Чтобы можно было провести анализ в первом приближении, обычно принимается кусочно-линейная модель, когда в каждой из рабочих областей емкость считается постоянной. Предлагаемые значения емкостей сведены в табл. 3.4.

Пример 3.9. Применение пакета моделирования для выяснения поведения емкости

Определение величины паразитных емкостей МОП-транзистора в заданном режиме работы представляет собой весьма трудоемкую задачу и требует знания ряда определяемых технологией параметров, которые часто недоступны в явном виде. К счастью, при наличии SPICE-модели транзистора нужные данные можно получить, выполнив простое моделирование. Предположим, что необходимо узнать величину общей емкости затвора транзистора,

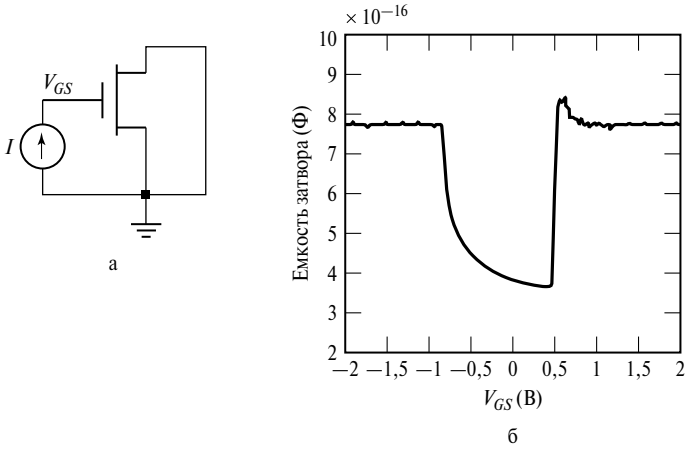


Рис. 3.32. Результат моделирования емкости затвора МОП-транзистора: а) используемая для анализа схема; б) график зависимости емкости для изготавливаемого по 0,25-микронной технологии n-МОП-транзистора минимального размера

изготавливаемого по определенной технологии, в функции напряжения V_{GS} (при $V_{DS} = 0$). Эту информацию даст моделирование цепи, показанной на рис. 3.32, *а*. Собственно, соотношение

$$I = C_G(V_{GS})\frac{dV_{GS}}{dt}$$

справедливо, и его можно записать в виде

$$C_G(V_{GS}) = I/\left(\frac{dV_{GS}}{dt}\right).$$

Моделирование переходного процесса даст значение V_{GS} в функции времени, которое путем выполнения простых математических операций может быть преобразовано в емкость. На рис. 3.32, *б* показан полученный моделированием график зависимости емкости затвора, получаемого по 0,25-микронной технологии n-МОП-транзистора минимального размера в функции напряжения V_{GS}. На графике ясно видно снижение величины емкости при приближении значения V_{GS} к пороговому напряжению V_T и разрыв при V_T, который и предсказывался на основании графиков, приведенных на рис. 3.31.

Емкость переходов

В последнюю емкостную компоненту дают вклад обратносмещенные pn-переходы исток-подложка и сток-подложка. Как уже обсуждалось, емкость

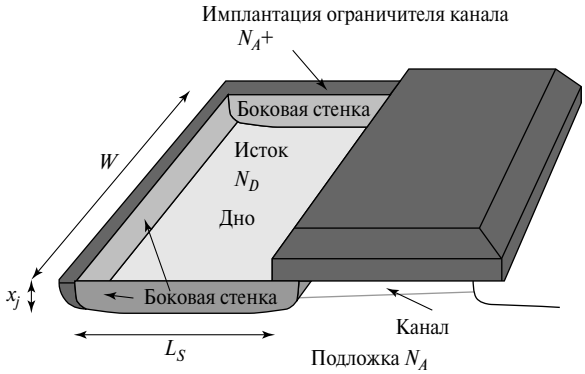


Рис. 3.33. Подробная схема истокового перехода

обедненной области нелинейна и уменьшается с ростом обратного смещения. Чтобы разобраться в составляющих емкости перехода (часто называемой *диффузионной емкостью*), необходимо внимательно рассмотреть область истока (стока) и ее окружение. На рис. 3.33 видно, что переход состоит из двух компонентов.

- *Донный плоский переход*, который образуется на границе раздела области истока (с уровнем легирования N_D) и подложки (с уровнем легирования N_A). Общая емкость обедненной области этой компоненты задается выражением $C_{bottom} = C_j W L_S$, где C_j — емкость единицы площади перехода, определяемая выражением (3.8). Поскольку донный плоский переход обычно является резким, то значение коэффициента плавности m приближается к 0,5.
- *Переход боковой стенки*, образуемый областью истока с уровнем легирования N_D и p^+-областью ограничения канала с уровнем легирования N_A^+. Уровень легирования области ограничения канала больше, чем у подложки, что приводит к большей величине емкости единицы площади. Как правило, переход на боковой стенке плавный со значением коэффициента плавности, лежащим в диапазоне от 0,33 до 0,5. Величина его емкости задается формулой $C_{sw} = C''_{sw} x_j (W + 2 \times L_s)$. Обратите внимание, что емкость четвертой боковой стенки истока не учитывается, так как она представляет собой проводящий канал[5]. Поскольку глубина залегания перехода x_j является параметром технологии, то обычно ее объединяют с емкостью C'_{jsw} в емкость единицы длины периметра $C_{jsw} = C'_{jsw} x_j$. Теперь выражение для общей емкости перехода может быть записано в следующем виде:

[5] Дабы все было полностью корректно, диффузионную емкость перехода исток(сток)-канал следует учитывать. Однако из-за уровней легирования и малой площади при выполнении приближенного анализа этой компонентой почти всегда можно пренебречь. В SPICE-модели этот переход учитывается путем ввода параметра C_{JSWG}.

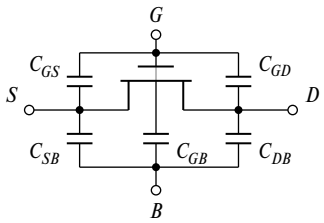


Рис. 3.34. Модель емкостей полевого МОП-транзистора

$$C_{diff} = C_{bottom} + C_{sw} = C_j \times \text{площадь} + C_{jsw} \times \text{периметр} = \\ = C_j L_S W + C_{jsw}\left(2L_S + W\right). \tag{3.45}$$

Так как все эти емкости малосигнальные, то обычно их линеаризируют и используют средние значения, как в выражении (3.9).

Задача 3.1. Использование пакета моделирования для определения емкости стока

Используя пакет моделирования (см. рис. 3.32), нарисуйте простую схему, которая поможет определить величину емкости стока n-МОП-транзистора в различных режимах работы.

Модель емкостей прибора

Рассмотренные составляющие емкости можно объединить в одну модель емкостей МОП-транзистора, которая показана на рис. 3.34. Все ее элементы легко идентифицируются на основе вышеизложенного. Уравнение имеет вид

$$C_{GS} = C_{GCS} + C_{GSO};\ C_{GD} = C_{GCD} + C_{GDO};\ C_{GB} = C_{GCB} \\ C_{SB} = C_{Sdiff};\ C_{DB} = C_{Ddiff}. \tag{3.46}$$

Разработчикам быстродействующих и низкопотребляющих интегральных схем очень важно хорошо знать эту модель и чувствовать относительные пропорции ее составляющих.

Пример 3.10. Емкости МОП-транзистора

Рассмотрим n-МОП-транзистор со следующими параметрами: $t_{ox} = 6$ нм, $L = 0,24$ мкм, $W = 0,36$ мкм, $L_D = L_S = 0,625$ мкм, $C_O = 3 \times 10^{-10}$ Ф/м, $C_{j0} = 2 \times 10^{-3}$ Ф/м и $C_{jsw0} = 2,75 \times 10^{-10}$ Ф/м. Определим значения всех соответствующих емкостей при нулевом смещении.

Таблица 3.5. Емкостные параметры n-МОП- и p-МОП-транзисторов, изготавливаемых по 0,25-микронной КМОП-технологии

	C_{ox}, фФ/мкм^2	C_O, фФ/мкм^2	C_j, фФ/мкм^2	m_j	ϕ_b, В	C_{jsw}, фФ/мкм	m_{jsw}	ϕ_{bsw}, В
n-МОП	6	0,31	2	0,5	0,9	0,28	0,44	0,9
p-МОП	6	0,27	1,9	0,48	0,9	0,22	0,32	0,9

Емкость затвора на единицу площади определяется легко, как $(\varepsilon_{ox}/t_{ox})$, и равна 5,7 фФ/мкм^2. Тогда емкость затвор-канал C_{GC} равна $WLC_{ox} = 0,49$ фФ. Чтобы определить общую емкость затвора, необходимо прибавить емкости перекрытия истока и стока, каждая из которых равна $WC_O = 0,105$ фФ. Таким образом, общая емкость затвора будет составлять 0,7 фФ.

Диффузионная емкость включает в себя донную емкость и емкость боковой стенки. Первая равна $C_{j0}L_DW = 0,45$ фФ, а емкость боковой стенки при нулевом смещении оценивается как $C_{jsw0}(2L_D + W) = 0,44$ фФ. В результате это дает значение общей емкости сток(исток)-подложка 0,89 фФ.

Похоже, диффузионная емкость преобладает над емкостью затвора. Но это наихудший случай. С увеличением обратного смещения перехода, что является нормальным режимом работы МОП-схем, величина диффузионной области значительно уменьшается. Кроме того, грамотное конструирование может помочь уменьшить величину $L_D(L_S)$. В общем случае вклад диффузионных емкостей самое большее равен и очень часто меньше емкости затвора.

Данные для проектирования: емкости МОП-транзистора

В табл. 3.5 сведены значения параметров, необходимых для оценки величин паразитных емкостей МОП-транзисторов, получаемых с использованием типового 0,25-микронного КМОП-процесса.

Сопротивление истока/стока

Дальнейшее влияние на быстродействие КМОП-схемы могут оказывать другие паразитные элементы, в частности последовательные сопротивления областей стока и истока, показанные на рис. 3.35. Их влияние усиливается с масштабированием транзисторов в сторону уменьшения размеров, так как это приводит к уменьшению глубины залегания переходов и размеров контактных окон. Сопротивление области стока (истока) можно записать в виде следующего выражения:

$$R_{S,D} = \frac{L_{S,D}}{W} R_{\square} + R_C \tag{3.47}$$

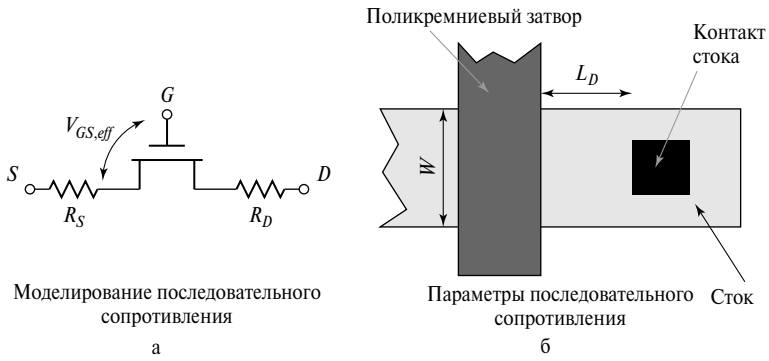


Рис. 3.35. Последовательное сопротивление стока и истока

где R_C — сопротивление контакта; W — ширина транзистора; $L_{S,D}$ — длина области истока или стока (см. рис. 3.35, *б*). Величину $R_\square$ называют *поверхностным сопротивлением* единицы поверхности диффузионной области истока-стока. Ее значение лежит в диапазоне от 20 до 100 Ом/□. Обратите внимание, что сопротивление квадратной области материала есть величина постоянная и не зависит от размера квадрата.

Последовательное сопротивление вызывает ухудшение быстродействия прибора, поскольку уменьшает ток стока при заданном управляющем напряжении. Таким образом, поддержание его значения на возможно низшем уровне является важной целью при проектировании как для разработчика приборов, так и для разработчика микросхем. Одним из возможных решений, которое стало популярным во многих современных технологических процессах, является покрытие областей стока и истока материалом с низким удельным сопротивлением, например титаном или вольфрамом. Этот процесс называется *силицидацией* и позволяет эффективно снизить величину поверхностного сопротивления до значений от 1 до 4 Ом/□[6]. Другой возможностью является изготовление транзистора с большей чем нужно шириной, что должно быть очевидным из выражения (3.40). Если в технологическом процессе предусмотрено использование силицидации, то при достаточном внимании к топологии паразитное сопротивление теряет свою важность. Но читателю следует помнить, что несерьезное отношение к топологии может привести к величинам сопротивлений, существенно ухудшающим быстродействие прибора.

[6] Процесс силицидации также используется для снижения сопротивления поликремниевого затвора. Это будет обсуждаться в главе 4.

3.3.3. Реальный МОП-транзистор: некоторые вторичные эффекты

В работе современных транзисторов могут наблюдаться некоторые важные отклонения от модели, представленной в предыдущих разделах. Эти расхождения становятся особенно заметными, как только размеры транзистора начинают заходить в дальнюю субмикронную область. В этот момент предположение о том, что работа транзистора адекватно описывается одномерной моделью, в соответствие с которой все токи текут по поверхности кремния и электрические поля направлены вдоль нее, перестает быть справедливым. Здесь уже более подходящими становятся двух- или даже трехмерные модели. Этот вопрос уже затрагивался на примере в разделе 3.2.2, когда обсуждалось уменьшение подвижности носителей.

Понимание сути некоторых из этих эффектов второго порядка и их влияния на поведение прибора очень важно при проектировании современных цифровых микросхем, поэтому эта тема заслуживает рассмотрения. Но прежде всего стоит предостеречь от попыток учета всех этих эффектов при выполнении прикидочного анализа, поскольку в результате модели схемы потеряют свою простоту и прозрачность. Разумнее на первом этапе анализа и проектирования МОП-схем использовать идеализированную модель. Влияние отклонений от идеальности можно исследовать на втором этапе с использованием компьютерного моделирования и более точных моделей транзистора.

Отклонения порогового напряжения

Выражение (3.18) постулирует, что пороговое напряжение является функцией только технологии изготовления и прикладываемого к подложке напряжения смещения V_{BS}. Поэтому его величина может считаться константой для всех n-МОП- и p-МОП-транзисторов в разрабатываемой схеме. Но с уменьшением размеров прибора такая модель становится неточной, а пороговое напряжение становится функцией L, W и V_{SB}. Неожиданно приобретают значение двухмерные эффекты второго порядка, которыми можно было пренебрегать в приборах с длинным каналом.

Например, при традиционном подходе к выводу формулы для напряжения V_{TO} предполагается, что обедненная область канала образуется исключительно под влиянием поданного на затвор напряжения и что весь заряд обедненной области под затвором обусловлен полевым эффектом в МОП-структуре. В этом случае игнорируется присутствие обедненных областей перехода истока и обратно смещенного перехода стока, которые становятся относительно более важными с уменьшением длины канала. Поскольку часть области под затвором уже обеднена (благодаря действию полей истока и стока), то для возникновения сильной инверсии оказывается достаточным меньшее пороговое напряжение. Другими словами, в приборах с коротким каналом напряжение V_{TO} уменьшается с уменьшением L (рис. 3.36, *а*). Подобного эффекта можно добиться и путем подъема напряжения исток-сток, так как это увеличивает ширину обедненной области перехода стока. В результате порог уменьшается с увеличением напряжения V_{DS}. Этот эффект,

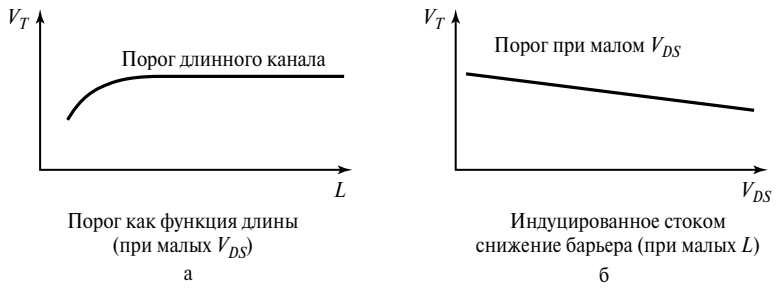


Рис. 3.36. Отклонения в поведении порогового напряжения

известный как *индуцированное стоком понижение барьера*, приводит к тому, что пороговое напряжение становится функцией рабочих напряжений (см. рис. 3.36, *б*). При достаточно высоких напряжениях на стоке области истока и стока могут даже закоротиться, и нормальная работа транзистора прекращается. Резкое увеличение тока в результате действия этого эффекта, известного под названием *прокола*, может привести к необратимому повреждению прибора, поэтому его следует избегать. Эффект прокола накладывает ограничение на верхнюю границу напряжений, которые могут прикладываться к транзистору.

Поскольку большинство транзисторов в цифровой микросхеме проектируется с минимальной длиной канала, то отклонение порогового напряжения в функции длины почти однородно по величине для всей разработки и поэтому не является столь уж большой проблемой, исключая, правда, увеличение предпороговых токов утечки. Большую неприятность представляет собой индуцируемое стоком понижение барьера, так как этот эффект зависит от рабочего напряжения. Например, он является проблемой в схемах динамической памяти. Ток утечки ячейки, т.е. предпороговый ток транзистора выборки, становится функцией напряжения на линии данных и зависит от конфигурации данных. С точки зрения ячейки памяти этот эффект проявляется как зависящий от данных источник шума.

Необходимо также упомянуть, что порог МОП-транзистора подвержен воздействию *эффекта узкого канала*. Обедненная область не заканчивается резко на границах транзистора, а несколько заходит под защитный окисел. Напряжение на затворе должно поддерживать этот дополнительный заряд обедненной области для установления проводящего канала. Данный эффект пренебрежимо мал в широких транзисторах, но становится значительным при малых значениях W, где он приводит к увеличению порогового напряжения. Для транзисторов с малыми геометрическими размерами, т.е. с малыми значениями W и L, эффекты короткого и узкого канала могут компенсировать друг друга.

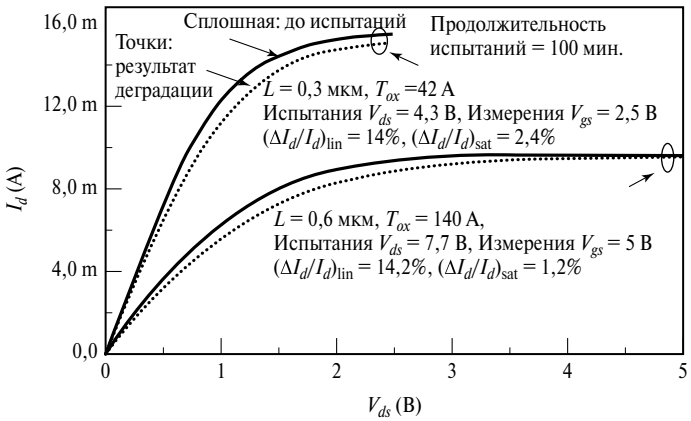
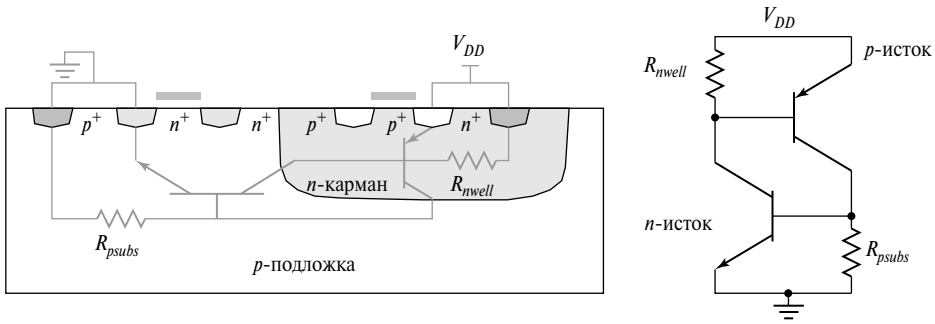


Рис. 3.37. Влияние горячих носителей вызывает деградацию ВАХ n-МОП-транзистора после эксплуатации в ужесточенных условиях (из [McGaughy98])

Влияние горячих носителей

Кроме изменения в зависимости от конструкции, пороговое напряжение в приборах с коротким каналом также имеет тенденцию к *дрейфу во времени*. Это является результатом действия эффекта *горячих носителей* [Hu92]. За последние несколько десятков лет размеры приборов непрерывно уменьшались, а вот напряжение питания и рабочие напряжения не уменьшались с соответствующим масштабом. Возникающее в результате увеличение напряженности электрического поля вызывает увеличение скорости электронов, которые при достижении достаточно высоких уровней энергии могут покидать кремний и туннелировать в подзатворный окисел. Захваченные в окисле электроны изменяют пороговое напряжение, обычно увеличивая его для n-МОП-транзисторов и снижая для p-МОП-транзисторов. Чтобы электрон стал горячим, необходимо наличие напряженности поля с величиной по крайней мере 10^4 В/см. Это условие с легкостью выполняется в приборах с длиной канала 1 мкм и менее. Горячие электроны могут приводить к проблемам с долгосрочной надежностью, когда схема деградирует или даже отказывает после использования некоторого времени. Данный эффект проиллюстрирован на рис. 3.37, где показана деградация ВАХ n-МОП-транзистора после эксплуатации в ужесточенных условиях. Поэтому современные технологии изготовления КМОП-микросхем предусматривают использование специально разработанных конструкций областей стока и истока, которые гарантируют ограничение пиковых значений напряженности поля, предотвращая тем самым приобретение носителями скоростей, достаточных для того, чтобы превратить их в горячие. Введение пониженных напряжений питания отчасти можно отнести на счет необходимости держать под контролем влияние горячих электронов.



Причина возникновения эффекта защелки
а

Эквивалентная схема
б

Рис. 3.38. Эффект защелки в КМОП-структуре

Эффект защелки в КМОП-структуре

МОП-структура содержит ряд собственных биполярных транзисторов. Особенно неприятны они в КМОП-конструкциях, где комбинация карманов и подложки приводит к формированию паразитных n-p-n-p-структур. Срабатывание подобных тиристороподобных приборов вызывает закорачивание шин питания V_{DD} и V_{SS}, что, как правило, имеет результатом разрушение микросхемы или, в лучшем случае, отказ системы, который может быть восстановлен только путем отключения питания.

Рассмотрим структуру с n-карманом, показанную на рис. 3.38, *а*[7]. n-p-n-p-структура образуется истоком n-МОП-транзистора, p-подложкой, n-карманом и истоком p-МОП-транзистора. Эквивалентная схема этой структуры приведена на рис. 3.38, *б*. Когда один из биполярных транзисторов оказывается прямосмещенным (например, из-за тока, протекающего по карману), он начинает поставлять ток в базу второго транзистора. Положительная обратная связь увеличивает ток до тех пор, пока схема не откажет либо перегорит.

Из приведенного выше анализа вывод для разработчика очевиден: чтобы избежать защелкивания, сопротивления R_{nwell} и R_{psubs} должны быть минимальны. Этого можно добиться размещением большого количества контактов к карману и подложке рядом с контактами к истокам n-МОП/p-МОП-приборов. Приборы, через которые проходят большие токи (например, транзисторы в цепях ввода-вывода), должны окружаться *охранными кольцами*. Такие располагающиеся по кругу вокруг транзистора контакты к карману/подложке дополнительно снижают сопротивление и уменьшают коэффициент усиления паразитных биполярных структур. Эффект защелки был особенно критичен в первых КМОП-процессах. В последние годы нововведения

[7] Данное описание механизмов, приводящих к возникновению эффекта защелки, по необходимости весьма поверхностно. Интересующиеся читатели могут найти подробное обсуждение этого вопроса в литературе, посвященной физике работы полупроводниковых приборов, например в [Streetman95].

в технологию изготовления и применение улучшенных методов проектирования практически исключили риски, связанные с эффектом защелки.

3.3.4. SPICE-модели МОП-транзисторов

Сложность поведения короткоканального МОП-транзистора и наличие в нем большого количества паразитных эффектов привели к разработке массы моделей с разной степенью точности и вычислительной эффективностью. В общем случае более высокая точность также означает бо́льшую сложность и, как результат, большее время счета. В данном разделе будут кратко рассмотрены характеристики наиболее популярных моделей полевого МОП-транзистора, а также на примере будет показано, каким образом МОП-транзистор вводится в описание схемы.

SPICE-модели

Пакет SPICE имеет три встроенные модели полевого МОП-транзистора, выбираемых по параметру LEVEL (УРОВЕНЬ). К сожалению, с переходом на короткоканальные транзисторы все эти модели неприменимы. Их следует использоваться только при выполнении прикидочного анализа, и поэтому здесь мы ограничимся только кратким рассмотрением их основных свойств.

- В SPICE-модели первого уровня (LEVEL 1) реализована модель Шихмана–Ходжеса (Shichman-Hodges model), в основе которой лежат квадратичные выражения для длинноканального транзистора, выведенные ранее в этой главе. Она совершенно не учитывает эффекты короткого канала.
- Модель второго уровня (LEVEL 2) основана на геометрических размерах, и уравнения прибора выводятся в ней через детальное рассмотрение физики работы. Она учитывает такие эффекты, как насыщение скорости, снижение подвижности и индуцированное стоком понижение барьера. К сожалению, учет всех трехмерных эффектов, проявляющихся в современных субмикронных приборах, на чисто физической основе становится сложным и неточным.
- Модель третьего уровня (LEVEL 3) — полуэмпирическая. В ней используется комбинация эмпирических и аналитических выражений, для определения параметров которых используются данные измерений на реальных приборах. Она достаточно хорошо работает при длинах канала до 1 мкм.

В ответ на неадекватность встроенных моделей поставщики пакета SPICE и производители полупроводниковых приборов представили широкий спектр точных, хотя и сделанных под себя, моделей. Полное описание каждой из них потребовало бы значительно больше места, чем может позволить данная книга, так что заинтересованный читатель может обратиться к соответствующей литературе по этой теме (например, к [Vladimirescu93]).

SPICE-модель BSIM3V3

Запутанная ситуация, связанная с необходимостью использования каждым производителем разных моделей, к счастью, частично была разрешена с при-

Таблица 3.6. Категории параметров модели BSIMv3 и некоторые важные параметры

Категория параметров	***Описание***
Параметры управления	Выбор уровня моделей подвижности, емкости и шума: LEVEL, MOBMOD, CAPMOD
Параметры по постоянному току	Параметры для вычисления порогового напряжения и токов: VTH0, K1, U0,VSAT, RSH
Параметры по переменному току и емкостные параметры	Параметры для расчета емкостей: CGS(D)O, CJ, MJ, CJSW, MJSW
dW и dL	Параметры для определения значений эффективной длины и ширины канала
Параметры процесса	Такие параметры процесса, как толщина окисла и концентрации легирующих примесей: TOX, XJ, GAMMA1, NCH, NSUB
Температурные параметры	Номинальная температура и температурные коэффициенты различных параметров прибора: TNOM
Кассета	Границы размеров прибора, в пределах которых модель справедлива: LMIN, LMAX, WMIN, WMAX
Шумовые параметры	Параметры модели шума

нятием в качестве отраслевого стандарта моделирования субмикронных полевых МОП-транзисторов модели BSIM3v3. **B**erkeley **S**hort-Channel **I**GFET **M**odel (Берклиевская модель короткоканального полевого транзистора с изолированным затвором), или BSIM, является аналитически простой моделью и основывается на "небольшом" количестве параметров, значения которых обычно определяются из эмпирических данных. Ее популярность и точность сделали ее естественным выбором для получения всех результатов моделирования, представленных в данной книге.

В полном виде модель BSIMv3 (обозначаемая как LEVEL 49) содержит свыше 200 параметров, большинство из которых связано с моделированием эффектов второго порядка. К счастью, разработчику цифровых схем совсем не обязательно разбираться в сложностях всех этих параметров. Поэтому в данной книге приводится только краткий перечень категорий параметров (табл. 3.6). Особого внимания заслуживает категория "*Кассета*". Получение единого набора параметров, который был бы приемлемым для приборов со всеми возможными размерами, почти невозможно. Вместо этого производители создают ряд моделей, каждая из которых справедлива только в ограниченной области размеров, границы которой задаются параметрами LMIN, LMAX, WMIN и WMAX (эта область и называется кассетой). Право выбора правильной кассеты для конкретного транзистора обычно оставляют за пользователем.

Полное описание параметров модели и уравнений заинтересованный читатель сможет найти в документации BSIMv3 на Web-сайте данной книги (`http://bwrc.eecs.berkeley.edu/IcBook`). Там же можно найти и модель LEVEL 49 для нашего типового 0,25-микронного КМОП-процесса.

Таблица 3.7. SPICE-параметры транзистора

Наименование параметра	*Символ*	*SPICE-имя*	*Единицы измерения*	*Значение по умолчанию*
Длина по чертежу	L	L	м	–
Эффективная ширина	W	W	м	–
Площадь истока	*AREA*	AS	$м^2$	0
Площадь стока	*AREA*	AD	$м^2$	0
Периметр истока	*PERIM*	PS	м	0
Периметр стока	*PERIM*	PD	м	0
Количество квадратов диффузионной области истока		NRS	–	1
Количество квадратов диффузионной области стока		NRD	–	1

Описание транзистора

Параметры, которые могут задаваться для отдельного транзистора, перечислены в табл. 3.7. Не все эти параметры должны задаваться для каждого транзистора. Для пропущенных параметров пакет SPICE сам устанавливает значения, принимаемые по умолчанию (которые часто равны нулю!).

ВНИМАНИЕ.

Трудно ожидать точных результатов моделирования, если введенное разработчиком описание схемы не содержит необходимых деталей. Например, при выполнении анализа быстродействия необходимо точно задать площадь и периметр областей стока и истока прибора. В отсутствии этой информации, которая используется при вычислении паразитных емкостей, результаты моделирования переходных процессов будут иметь мало пользы. Аналогично, часто необходимо скрупулезно задавать значение сопротивления истока и стока. Для точного учета паразитных последовательных сопротивлений истока и стока каждого транзистора значения параметров NRS и NRD умножаются на величину поверхностного сопротивления, задаваемую в модели транзистора.

Пример 3.11. Описание КМОП-инвертора в пакете SPICE

В данном примере приводится описание в рамках пакета SPICE КМОП-инвертора, состоящего из n-МОП- и p-МОП-транзистора. Транзистор M1 (представляющий собой n-МОП-прибор) описывается моделью (и относится к кассете) типа *nmos*.1, выводы стока, затвора, истока и подложки которого подключены к узлам *nvout*, *nvin*, 0 и 0 соответственно. Длина его затвора равна минимально допустимой для данной технологии (0,25 мкм). Символ + в начале второй строки указывает, что данная строка является продолжением предыдущей.

p-МОП-прибор относится к типу *pmos*.1, подключается к узлам *nvout*, *nvin*, *nvdd* и *nvdd* (сток, затвор, исток и подложка) и в три раза шире, что уменьшает последовательные сопротивления стока и истока, но увеличивает паразитные диффузионные емкости, поскольку площадь и периметр областей стока и истока при этом увеличиваются.

Наконец, строка *.lib* представляет собой ссылку на файл, в котором содержатся модели транзисторов.

```
M1 nvout nvin 0 0 nmos.1 W=0.375U L=0.25U
+AD=0.24P PD=1.625U AS=0.24P PS=1.625U NRS=1 NRD=1
M2 nvout nvin nvdd nvdd pmos.1 W=1.125U L=0.25U
+AD=0.7P PD=2.375U AS=0.7P PS=2.375U NRS=0.33 NRD=0.33
.lib 'c:\Design\Models\cmos025.1'
```

3.4. Несколько слов об отклонениях в технологическом процессе

Выше предполагалось, что прибор адекватно моделируется одним набором параметров. В реальности же параметры транзистора могут изменяться от платины к пластине или даже от транзистора к транзистору на одном кристалле в зависимости от расположения. Подобный наблюдаемый случайный разброс между предположительно идентичными приборами в основном является результатом действия двух факторов.

1. Отклонения параметров технологических процессов, например концентрации примесей, толщины окисла и глубины диффузии, вызываемые неоднородностью условий в ходе осаждения и/или диффузии примесей. Это приводит к разбросу в значениях поверхностного сопротивления и параметров транзистора, например порогового напряжения.
2. Отклонения размеров приборов, вызываемых в основном ограниченной разрешающей способностью процесса фотолитографии, что приводит к разбросу значений отношения (W/L) МОП-транзисторов и ширины линий межсоединений.

Заметьте, что достаточно большое количество таких отклонений совершенно не коррелирует друг с другом. Например, разброс значений длины МОП-транзистора никак не связан с разбросом значений порогового напряжения, поскольку они формируются в ходе разных операций технологического процесса.

- Подобные отклонения в технологическом процессе оказывают влияние на параметры, определяющие быстродействие микросхемы, например, на ток транзистора и его параметры.
- Пороговое напряжение V_T может изменяться по многим причинам, среди которых изменение толщины окисла, концентрации примесей в подлож-

ке, поликремнии и в областях имплантации, а также изменение плотности поверхностного заряда. Точное управление величиной порогового напряжения является важным по многим причинам. И если в прошлом пороги могли изменяться вплоть до 50% от своего номинального значения, то современные технологические процессы изготовления цифровых микросхем позволяют контролировать их с точностью до 25–50 мВ.

- Основной причиной разброса значений крутизны процесса k'_n является изменение в толщине окисла. Могут наблюдаться и отклонения в величине подвижности носителей, но в меньшей степени.
- Отклонения значений W и L. В основном они обусловлены процессом фотолитографии. Кстати, обратите внимание, что отклонения значений ширины W совершенно не коррелируют с отклонениями значений длины L, поскольку первые определяются операцией травления защитного окисла, тогда как вторые — точностью воспроизведения размеров в слое поликремния и процессами диффузии при формировании областей стока и истока.

Измеримым эффектом от отклонений в технологическом процессе может быть значительный разброс в поведении схемы в сравнении с номинальной или ожидаемой реакцией, причем отклонения поведения могут быть как в положительную, так и в отрицательную сторону. Все это ставит перед разработчиком важную экономическую дилемму. Предположим, например, что вас попросили спроектировать микропроцессор, работающий на тактовой частоте 3 ГГц. С экономической точки зрения важно, чтобы большинство изготавливаемых кристаллов удовлетворяло данному требованию по быстродействию. Одним из способов достижения этой цели является проектирование микросхемы в предположении худшего случая значения всех возможных параметров приборов. Хотя подобный подход и безопасен, он чрезмерно консервативен и приводит к получению схем с завышенным запасом и, следовательно, невыгодных экономически.

Чтобы помочь разработчику в принятии решения относительно величины закладываемого в разработку запаса прочности, обычно изготовители приборов в дополнение к номинальной модели еще предоставляют так называемые быстрые и медленные модели прибора. В результате значения токов могут быть больше (или меньше) ожидаемых в соответствии с правилом разброса 3σ.

Пример 3.12. Отклонения процесса изготовления МОП-транзистора

Чтобы проиллюстрировать возможное влияние отклонений в технологическом процессе на быстродействие МОП-прибора, рассмотрим n-МОП-транзистор минимального размера, получаемый в рамках нашей типовой 0,25-микронной КМОП-технологии. (В главе 5 будет показано, что быстродействие прибора пропорционально току стока, который он может обеспечить.)

Предположим сначала, что $V_{GS} = V_{DS} = 2,5$ В. Из полученных ранее результатов моделирования уже известно, что при этих условиях ток стока

составляет 220 мкА. Теперь заменим номинальную модель на быструю и медленную модели, в которых модифицируются такие параметры транзистора, как длина и ширина (±10%), пороговое напряжение (±60 мВ) и толщина окисла (±5%). Моделирование дает следующие результаты:

быстрая: $I_d = 265$ мкА: +20%;

медленная: $I_d = 182$ мкА: −17%.

Теперь сделаем следующий шаг. Напряжение источника питания, подводимое к микросхеме, тоже не является константой. Например, напряжение батареи может значительно падать с приближением ее времени жизни к концу. На практике вполне можно ожидать наличия отклонений напряжения в диапазоне ±10%:

быстрая $+V_{dd} = 2,75$ В: $I_d = 302$ мкА: +37%;

медленная $+V_{dd} = 2,25$ В: $I_d = 155$ мкА: −30%.

Таким образом, в экстремальных случаях уровни тока и соответственно быстродействия прибора могут различаться почти на 100%. Чтобы гарантировать для предполагаемой к изготовлению микросхемы удовлетворение требования по быстродействию при всех обстоятельствах, необходимо сделать транзистор на 42% шире (= 220 мкА / 155 мкА), чем это требуется для номинального случая. В результате это приведет к сильному проигрышу по площади.

К счастью, на практике такие условия наихудшего (или наилучшего) случая встречаются крайне редко. Вероятность того, что все параметры одновременно будут соответствовать наихудшему случаю, очень мала, и большинство спроектированных изделий обычно показывает быстродействие, величина которого лежит вокруг центрального значения, соответствующего номинальным условиям. Искусство *проектирования технологичных изделий* как раз и состоит в умении выбрать номинальные условия таким образом, чтобы подавляющее большинство (например, 98%) изготавливаемых микросхем удовлетворяло требованиям спецификации по быстродействию и при этом накладные расходы по площади оставались минимальными.

Для достижения этой цели имеются специализированные инструментальные средства проектирования. Например, анализ с применением метода Монте-Карло [Jensen91] позволяет смоделировать микросхему, когда значения параметров прибора выбираются в широком диапазоне случайным образом. Результатом такого анализа является график распределения значений параметров разработки (скажем, быстродействия или устойчивости к шуму), который может помочь в оценке экономической пригодности выбранного номинального решения. Примеры подобных графиков распределения показаны на рис. 3.39. Они графически демонстрируют влияние разброса значений эффективной длины канала транзистора и порогового напряжения p-МОП-транзистора на быстродействие ячейки сумматора. Как видно на рисунке, отклонения в технологическом процессе могут существенным образом влиять на параметры быстродействия разработки.

Из всех предыдущих рассуждений можно сделать один важный вывод о том, что к моделированию с помощью пакета SPICE следует относиться

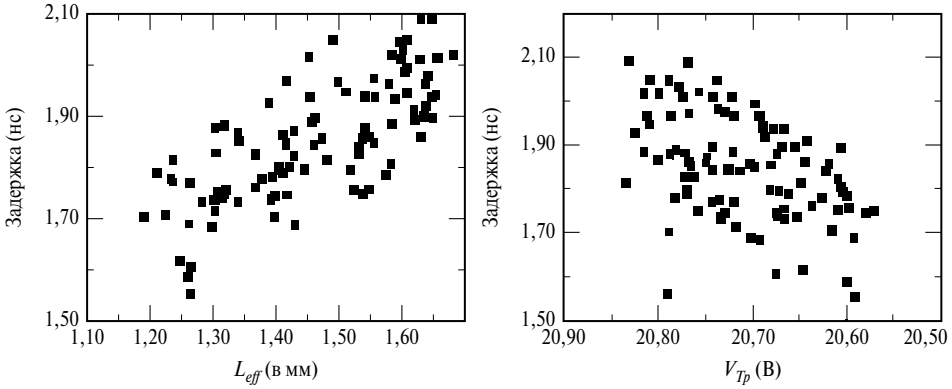


Рис. 3.39. Графики распределения значений быстродействия микросхемы сумматора в функции изменяющихся параметров прибора, полученные в результате анализа с использованием метода Монте-Карло. Микросхема выполнялась по 2 мкм (номинальное значение) КМОП-технологии. (Перепечатано с разрешения Эрика Боскина (Erik Boskin), UCB и корпорации ATMEL corp.)

с осторожностью. Представленные в модели параметры прибора отражают средние значения, измеренные на партии обрабатываемых пластин. Реальные микросхемы обязательно будут показывать результаты, отличающиеся от результатов моделирования, и дело тут вовсе не в несовершенстве подхода к моделированию. Более того, надо также помнить, что разброс температуры по кристаллу может быть еще одним источником отклонений в значениях параметров. Поэтому задача оптимизации МОП-схемы с помощью пакета SPICE до уровня пикосекунд или микровольт не стоит затрачиваемых на ее решение усилий.

3.5. Взгляд в перспективу: масштабирование технологии

За последние несколько десятилетий наблюдалось заметное увеличение плотности интеграции и вычислительной сложности цифровых интегральных схем. Как уже упоминалось ранее, приложения, считавшиеся ранее невероятными, сегодня уже забыты. Основой этой революции является постоянное совершенствование технологии изготовления приборов, которое позволяет стабильно уменьшать величину минимальной нормы проектирования, представляемой для МОП-схем минимальной длиной канала транзистора, реализуемой на кристалле микросхемы. Для иллюстрации этого утверждения на рис. 3.40 показан график изменения (средних) минимальных размеров прибора, начиная с 1960-х годов с проекцией на XXI столетие. Наблюдаемая скорость уменьшения составляет примерно 13% в год, или в два раза меньше за каждые пять лет. Более того, нет никаких признаков замедления, и это захватывающее дух движение вперед может продолжаться и в обозримом будущем.

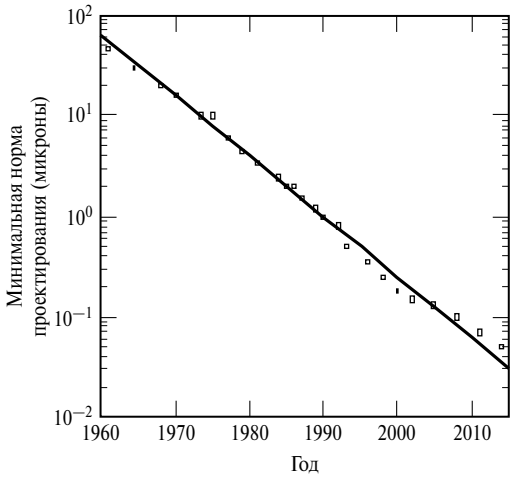


Рис. 3.40. Изменение (средней) минимальной длины канала МОП-транзистора во времени. Точки соответствуют фактическим или прогнозируемым (2000 год и далее) значениям. Сплошная линия отражает сценарий масштабирования, в соответствии с которым минимальная норма проектирования уменьшается вдвое за каждые пять лет

Вопрос состоит в том, как это устойчивое уменьшение минимальной нормы проектирования влияет на рабочие характеристики и свойства МОП-транзистора и косвенно на такие критически важные показатели качества разработки цифровой микросхемы, как частота переключения и рассеиваемая мощность? Темой настоящего раздела как раз и является прогнозирование этого поведения в первом приближении, получившем название *масштабирование*. В данном исследовании в дополнение к минимальному размеру прибора в качестве второй независимой переменной следует рассматривать напряжение питания. В зависимости от характера изменения этих двух независимых переменных относительно друг друга получаются разные сценарии масштабирования (см. [Dennard74, Baccarani84]).

В табл. 3.8 анализируются три модели. С целью облегчения интерпретации результатов предполагается, что все размеры приборов изменяются в одинаковом масштабе S ($S > 1$ в случае уменьшения). Сюда входят длина и ширина транзистора, толщина окисла и глубина залегания перехода. Аналогично предполагается, что все напряжения, включая напряжение питания и пороговое напряжение, тоже изменяются с одинаковым масштабом U. Соотношения, определяющие поведение при масштабировании зависимых переменных, приведены во втором столбце. Обратите внимание, что данный анализ касается только короткоканальных приборов с линейной зависимостью между управляющим напряжением и током насыщения (в соответствии с выражением (3.31)). Рассмотрим каждый сценарий по очереди.

Таблица 3.8. Сценарии масштабирования короткоканальных приборов

Параметр	*Соотношение*	*Полное масштабирование*	*Общее масштабирование*	*Масштабирование с постоянным напряжением*
W, L, t_{ox}		$1/S$	$1/S$	$1/S$
V_{DD}, V_T		$1/S$	$1/U$	1
N_{SUB}	V/W_{depl}^2	S	S^2/U	S^2
Площадь/прибор	*WL*	$1/S^2$	$1/S^2$	$1/S^2$
C_{ox}	$1/t_{ox}$	S	S	S
C_{gate}	C_{ox}*WL*	$1/S$	$1/S$	$1/S$
k_n, k_p	$C_{ox}W/L$	S	S	S
I_{sat}	C_{ox}*WV*	$1/S$	$1/U$	1
Плотность тока	I_{sat}/*площадь*	S	S^2/U	S^2
R_{on}	V/I_{sat}	1	1	1
Собственная задержка	$R_{on}C_{gate}$	$1/S$	$1/S$	$1/S$
P	$I_{sat}V$	$1/S^2$	$1/U^2$	1
Плотность мощности	*P/площадь*	1	S^2/U^2	S^2

Полное масштабирование или масштабирование с постоянной величиной напряженности электрического поля

В этой идеальной модели напряжения и размеры масштабируются с одинаковым коэффициентом S. Цель заключается в том, чтобы сохранить картину напряженностей электрических полей в масштабируемом приборе такой же, какой она существует в исходном приборе. Сохранение постоянства напряженностей электрического поля гарантирует физическую целостность прибора и позволяет избежать пробоя или других вторичных эффектов. Такое масштабирование приводит к увеличению плотности приборов (*площадь*), быстродействия (*собственная задержка*) и снижению потребляемой мощности (P). Влияние полного масштабирования на параметры прибора и микросхемы отражено в третьем столбце табл. 3.8. В качестве меры быстродействия использована постоянная времени внутренней задержки, которая равна произведению емкости затвора на сопротивление во включенном состоянии (“On”). Анализ показывает, что благодаря одновременному масштабированию перепада напряжения и уровня тока сопротивление во включенном состоянии остается неизменным. Улучшение быстродействия является следствием исключительно уменьшения емкости. Результаты ясно показывают наличие выигрыша от масштабирования: быстродействие микросхемы растет линейно, тогда как мощность вентиля падает по квадратичному закону![8]

[8] При заполнении табл. 3.8 были сделаны определенные предположения.
1. Предполагалось, что подвижность носителей не изменяется при масштабировании.
2. Уровень легирования подложки N_{sub} масштабируется таким образом, что максимальная ширина обедненного слоя уменьшается с коэффициентом масштабирования S.
3. Также предполагалось, что величина задержки прибора определяется в основном собственной емкостью (емкостью затвора) и что другие емкости прибора (например, диффузионные емкости) масштабируются соответствующим образом. Это предположение примерно спра-

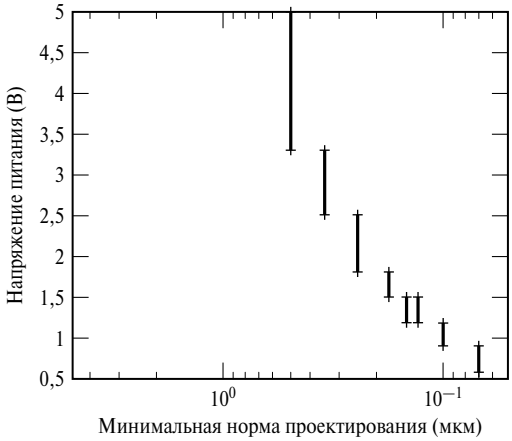


Рис. 3.41. Эволюция минимального и максимального напряжения питания в цифровых интегральных схемах как функции размера элемента

Масштабирование с постоянным напряжением

В реальной жизни вариант с полным масштабированием нежизнеспособен. Для сохранения совместимости новых приборов с существующими компонентами напряжения не могут масштабироваться произвольным образом, так как обеспечение нескольких источников питания существенно удорожает систему. В результате напряжения не изменялись вместе с нормами проектирования, а разработчики следуют установленным стандартами уровням напряжений питания и сигналов. Как показано на рис. 3.41, вплоть до начала 1990-х годов для всех цифровых компонентов стандартом де-факто было напряжение питания 5 В, и работала *модель масштабирования с фиксированным напряжением.*

Новые стандарты 3,3 и 2,5 В начали вступать в силу только с появлением 0,5-микронной КМОП-технологии. Сегодня наблюдается уже более согласованное изменение напряжений и размеров приборов. Причину такого изменения модели работы частично можно объяснить с помощью модели масштабирования с постоянным напряжением, приведенной в пятом столбце табл. 3.8. Для прибора, работающего в режиме насыщения скорости, сохранение напряжения неизменным при масштабировании размеров прибора не приводит к выигрышу по быстродействию, как в случае полного масштабирования, а, наоборот, дает существенный проигрыш в мощности. Положительный эффект от увеличения тока просто уничтожается более высоким уровнем напряжения, что приводит только к ухудшению ситуации с рассеиваемой мощностью. Этот сценарий сильно отличается от ситуации, когда транзистор работает в режиме длинного канала и ток является квадратичной функцией напряже-

---

ведливо в случае полного масштабирования, но не верно в случае масштабирования с постоянным напряжением, при котором C_{diff} масштабируется с коэффициентом $1/\sqrt{S}$.

ния (вспомните выражение (3.32)). В таких обстоятельствах сохранение напряжения на постоянном уровне дает явный выигрыш по быстродействию, поскольку при этом имеется чистое уменьшение величины сопротивления во включенном состоянии. Другие причины для масштабирования напряжения питания вместе с технологией связаны с такими физическими явлениями, как влияние горячих носителей и пробой окисла. Именно эти эффекты внесли существенный вклад в то, что модель масштабирования с постоянным напряжением перестала поддерживаться.

Задача 3.2. Масштабирование длинноканальных приборов

Продемонстрируйте, что для длинноканального транзистора полное масштабирование напряжения дает снижение собственной задержки с коэффициентом S^2, но увеличивает рассеиваемую одним прибором мощность в S раз. Перезаполните табл. 3.8 в предположении, что ток является квадратичной функцией напряжения (выражение (3.29)).

ВНИМАНИЕ.

Обрисованная в предыдущем разделе картина представляет собой модель первого порядка приближения. Увеличение напряжения питания все же дает некоторый выигрыш по быстродействию в короткоканальных транзисторах. Это четко видно на рис. 3.28 и в табл. 3.3, в которых отражено снижение величины сопротивления во включенном сопротивлении с увеличением напряжения питания, даже в диапазоне высоких значений напряжения. Тем не менее этот эффект, который обусловлен в основном модуляцией длины канала, вторичен и значительно меньше, чем это имеет место в длинноканальных приборах.

Читатель должен помнить об этом предупреждении при изучении вопроса масштабирования. Главная цель здесь — рассмотрение основных тенденций, что подразумевает игнорирование таких вторичных эффектов, как снижение подвижности, наличие последовательного сопротивления и т.д.

Общее масштабирование

На рис. 3.41 показано, что напряжение питания, хотя и уменьшается, но не так быстро, как возможности технологии. Например, при масштабировании технологии от 0,5-микронной до 0,1-микронной значение максимального напряжения питания снижается только с 5 до 1,5 В. Естественно, встает вопрос, почему не следовать модели полного масштабирования, если сохранение высокого уровня напряжения не дает каких-либо убедительных преимуществ? Подобное отклонение обусловлено следующими моментами.

- Некоторые внутренние напряжения прибора, например ширина запрещенной зоны кремния и контактная разность потенциалов, являются параметрами материала и поэтому не могут масштабироваться.

- Масштабирование порогового напряжения транзистора ограничено. Слишком сильное понижение порога делает сложным полное выключение прибора. Ситуация усугубляется наличием большого технологического разброса значений порогового напряжения, причем даже в пределах одной пластины.

По этим и другим причинам необходима более общая модель масштабирования, когда напряжения и размеры масштабируются независимо. Такая модель общего масштабирования отражена в четвертом столбце табл. 3.8, в соответствии с которой размеры прибора масштабируются с коэффициентом S, а напряжения уменьшаются с коэффициентом U. Если напряжение удерживается неизменным, то модель масштабирования вырождается в модель с фиксированным напряжением. Обратите внимание, что модель общего масштабирования предлагает сценарий поведения быстродействия, идентичный сценарию моделей полного масштабирования и масштабирования с фиксированным напряжением, но поведение рассеиваемой мощности представляет собой нечто среднее между первыми двумя моделями (при $S > U > 1$).

Проверка справедливости модели

В табл. 3.9 сведены характеристики нескольких наиболее современных КМОП-процессов и прогнозов на будущее. Обратите внимание на постоянное снижение рабочих напряжений с уменьшением размеров приборов. Как и предсказывается моделью масштабирования, величина максимального обеспечиваемого прибором тока на единицу длины остается приблизительно постоянной. Поддержание такого уровня нагрузочной способности при уменьшенном напряжении питания требует агрессивного снижения порогового напряжения, что трансформируется в быстрое увеличение предпорогового тока утечки.

На основе приведенных в таблице данных можно вполне обоснованно сказать, что как плотность интеграции, так и быстродействие будет продолжать расти. Как долго они будут вести себя подобным образом, неизвестно, и этот вопрос является предметом многих исследований. Экспериментальные субдвадцатинанометровые-КМОП-приборы показали свою работоспособность в лабораторных условиях и имели токовые характеристики, удивительно близко совпадающие с характеристиками современных транзисторов. Эти транзисторы, хотя и работают на тех же принципах, что и привычные МОП-приборы, весьма отличаются от них структурой и требуют существенного изменения в *подходах к проектированию*. На рис. 3.42 показан пример структуры потенциального транзистора, разработанного в Беркли и названного *ножевым двухзатворным полевым транзистором* (FinFET dual-gate transistor), который оказался работоспособным вплоть до очень малых длин канала.

Еще одним вариантом является *вертикальный транзистор*. Хотя введение дополнительных слоев металлизации превратило интегральную микросхему в истинно трехмерное изделие, сам транзистор по-прежнему размещается главным образом в горизонтальной плоскости. Это заставляет разработ-

Таблица 3.9. Прогноз развития технологии полевых МОП-транзисторов для быстродействующих логических микросхем (по данным [SIA01])

Год появления	**2001**	**2003**	**2005**	**2007**	**2010**	**2013**	**2016**
Длина канала по чертежу (нм)	90	65	45	35	25	18	13
Физическая длина канала (нм)	65	45	32	25	18	13	9
Подзатворный окисел (нм)	2,3	2,0	1,9	1,4	1,2	1.0	0,9
V_{DD} (В)	1,2	1,0	0,9	0,7	0,6	0,5	0,4
n-МОП I_{Dsat}(мкА/мкм)	900	900	900	900	1200	1500	1500
n-МОП I_{leak} (мкА/мкм)	0,01	0,07	0,3	1	3	7	10

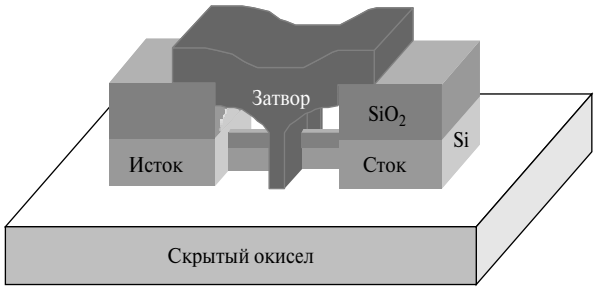


Рис. 3.42. Двухзатворный ножевой полевой транзистор с длиной канала 25 нм [Huang99]

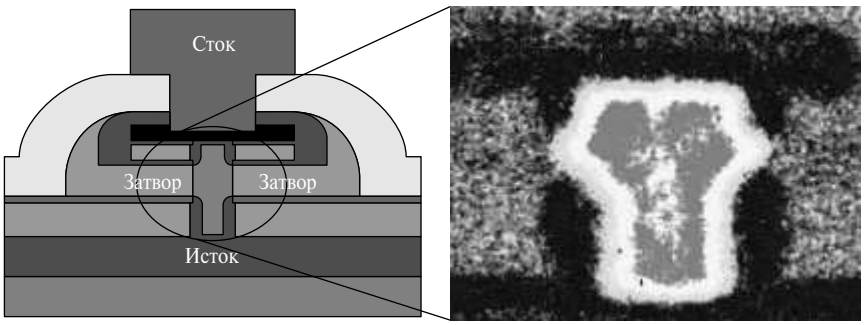


Рис. 3.43. Вертикальный транзистор с двойным затвором. На фотографии справа показано увеличенное изображение области канала

чика приборов заниматься одновременной оптимизацией плотности упаковки и параметров, связанных с быстродействием. Если же развернуть прибор таким образом, чтобы сток оказался вверху, а исток внизу, то эти две проблемы можно будет решать по отдельности: плотность упаковки по-прежнему будет определяться размерами в горизонтальной плоскости, тогда как быстродействие будет зависеть от расстояний по вертикали (рис. 3.43). Работающие приборы такого типа были изготовлены с длиной канала существенно ниже 0,1 мкм [Eaglesham00].

Похоже, что уже на подходе интегральные схемы, объединяющие в себе более миллиарда транзисторов и работающие с тактовыми частотами в десятки гигагерц. Вопрос о том, будет ли этот процесс продолжаться, пока остается открытым. Хотя это и может быть технологически возможным делом, но прежде чем участвовать в таком мероприятии, следует учесть множество других факторов. Стоит поинтересоваться, а будет ли изготовление такой микросхемы экономически выгодным. Стоимость современных полупроводниковых заводов составляет более 5 миллиардов долларов, и, как ожидается, эта стоимость будет расти с дальнейшим уменьшением норм проектирования. Свою роль играют и конструктивные соображения. Уровень потребляемой мощности у такой компоненты может оказаться совершенно запредельным, да и растущая роль паразитных элементов межсоединений может поставить предел верхней границы быстродействия. Наконец, желательный уровень степени интеграции могут определять чисто системные соображения. В общем, совершенно очевидно, что в будущем в области проектирования полупроводниковых микросхем по-прежнему придется сталкиваться с весьма непростыми проблемами.

3.6. Резюме

В данной главе было представлено достаточно подробное описание принципов работы полевого МОП-транзистора — полупроводникового прибора, который является основой фактически всех современных цифровых интегральных схем. Кроме попытки дать интуитивное понимание его поведения, были описаны разнообразные подходы к моделированию, включая как простые модели, пригодные для выполнения приближенного ручного анализа работы схемы, так и сложные SPICE-модели. Эти модели будут широко использоваться в последующих главах при рассмотрении основных строительные блоков цифровых микросхем. Первым в данной главе было приведено краткое описание поведения полупроводникового диода, являющегося одним из основных паразитных элементов в КМОП-конструкциях.

- Статическое поведение диода на pn-переходе хорошо описывается уравнением идеального диода, которое постулирует, что ток является *экспоненциальной функцией подаваемого напряжения смещения.*
- В режиме обратного смещения пространственный заряд в обедненной области диода может моделироваться *нелинейной зависящей от напряжения емкостью.* Это особенно важно, поскольку всегда присутствующие в МОП-транзисторе переходы исток-подложка и сток-подложка работают именно в этом режиме. Была представлена линеаризованная укрупненная модель емкости обедненного слоя, пригодная для ручного анализа.
- Полевой МОП-транзистор представляет собой прибор, управляемый напряжением, в котором управляющий электрод затвора изолируется от проводящего канала емкостью слоя SiO_2. На основе величины напряжения исток-затвор относительно порогового напряжения V_T были выделены три рабочие области: *отсечки*, *линейная* и *насыщения.* Наиболее привле-

кательным свойством МОП-транзистора, которое делает его особенно пригодным для разработки цифровых схем, является его способность работать в качестве ключа, управляемого напряжением: при низком управляющем напряжении ключ находится в непроводящем состоянии (разомкнут), а при высоком управляющем напряжении образуется проводящий канал, и ключ может рассматриваться замкнутым. Подобное поведение с двумя состояниями хорошо согласуется с концепцией двоичной логики.

- Продолжающееся уменьшение размеров прибора до субмикронного диапазона привело к существенным отклонениям от традиционной модели длинноканального транзистора. Наиболее важным из них является возникновение эффекта *насыщения скорости*, который изменяет характер зависимости тока транзистора от управляющего напряжения с *квадратичного* на *линейный*. Были представлены модели этого эффекта, а также модели других паразитных элементов второго порядка. Одним из наиболее важных эффектов является *предпороговая проводимость*, которая приводит к тому, что прибор пропускает даже тогда, когда управляющее напряжение падает ниже порогового.
- Динамика работы МОП-транзистора определяется главным образом *емкостями прибора*. Основной вклад вносят емкость затвора и емкость, создаваемая областями обеднения переходов истока и стока. Минимизация величины этих емкостей является основным требованием для конструкции быстродействующих МОП-микросхем.
- Были представлены SPICE-модели и их параметры для всех приборов. Было отмечено, что эти модели отражают среднее поведение и могут изменяться в пределах одной пластины и даже кристалла.
- Ожидается, что по крайней мере все следующее десятилетие МОП-транзистор будет доминировать на сцене цифровых интегральных схем. Постоянное масштабирование приведет к появлению к 2010 году приборов с размерами около 0,03 мкм и интегральных схем с количеством транзисторов в одном кристалле более 1 миллиарда.

3.7. Для любознательных

Теме полупроводниковых приборов посвящено множество книг, повторных изданий, учебных пособий и журнальных статей. Основным журналом является *IEEE Transactions on Electron Devices*, и именно в нем описывается большинство последних конструкций приборов и подходов к их моделированию. Другим ценным источником информации являются труды Международного форума по электронным приборам (International Electron Devices Meeting — IEDM). В перечисленных ниже книгах (например, [Streetman95] и [Pierret96]) и журнальных статьях содержатся прекрасные описания упомянутых полупроводниковых приборов или рассматриваются другие темы, затронутые в данной главе.

3.7.1. Список использованной литературы

[Baccarani84] G. Baccarani, V. Wordeman, and R. Dennard, "Generalized Scaling Theory and Its Application 1/4 Micrometer MOSFET Design", *IEEE Trans. Electron Devices*, 1984 ED-31(4), p. 452.

[Banzhaf92] W. Banzhaf, *Computer Aided Analysis Using PSPICE*, 2nd ed., Prentice Hall, 1992.

[Dennard74] R. Dennard et al., "Design of Ion-Implanted MOSFETS with Very Small Physical Dimensions", *IEEE Journal of Solid-State Circuits*, 1974, SC-9, p. 256–258.

[Eaglesham99] D. Eaglesham, "0,18 μm CMOS and Beyond", *Proceedings 1999 Design Automation Conference*, June 1999, p. 703–708.

[Howe97] R. Howe and S. Sodini, *Microelectronics: An Integrated Approach*, Prentice Hall, 1997.

[Hu92] C. Hu, "IC Reliability Simulation", *IEEE Journal of Solid State Circuits*, March 1992, vol. 27, no. 3, p. 241–246.

[Huang99] X. Huang, W.C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y. K. Choi, K. Asano, V. Subramanian, T.J. King, J. Bokor, and C. Hu, "Sub 50-nm FinFET: PMOS", *International Electron Devices Meeting*, 1999, p. 67–70.

[Jensen91] G. Jensen et al., "Monte Carlo Simulation of Semiconductor Devices", *Computer Physics Communications*, August 1991, 67, p. 1–61.

[Ko89] P.Ko, "Approaches to Scaling", in *VLSI Electronics: Microstructure Science*, vol. 18, chapter 1, p. 1–37, Academic Press, 1989.

[McGaughy98] J.F. Chen, B.W. McGaughy, and C. Hu, "Statistical Variation of NMOSFET Hot-Carrier Lifetime and Its Impact on Digital Circuit Reliability", *International Electron Device Meeting Technical Digest (IEDM)*, 1995, p. 29–32.

[Nagel75] L. Nagel, "SPICE2: a Computer Program to Simulate Semiconductor Circuits", memo ERL-M520, Dept. Elect. And Computer Science, University of California at Berkeley, 1975.

[Pierret96] R. Pierret, *Semiconductor Device Fundamentals*, Addison-Wesley, 1996.

[SIA01] *International Technology Roadmap for Semiconductors*, `http://www.sematech.org`, 2001.

[Streetman95] B. Streetman, *Solid State Electronic Devices*, Prentice Hall, 1995.

[Thorpe92] T. Thorpe, *Computerized Circuit Analysis with SPICE*, John Wiley and Sons, 1992.

[Toh88] K. Toh, P. Ko, and R. Meyer, "An Engineering Model for Short-Channel MOS Devices", *IEEE Journal of Solid State Circuits*, August 1988, vol. 23, no. 4, p. 950–958.

[Vladimirescu93] A. Vladimirescu, *The SPICE Book*, John Wiley and Sons, 1993.

Упражнения

Обновляемый сборник задач и упражнений по диодам и МОП-приборам предоставлен на Web-сайте `http://bwrc.eecs.berkley.edu/IcBook`.

Вставка

Б

Моделирование схемы

Свойства схемного моделирования

Модели схемного моделирования

Первое, чего ожидает разработчик от автоматизации проектирования, — это доступности быстрых и точных инструментов. Первым инструментом автоматизированного проектирования (Computer-Aided Design — CAD), АП, получившим широкое распространение, стал пакет SPICE, несомненно являющийся наиболее используемым средством компьютерной помощи в проектировании (цифровых) схем. Изначально пакет SPICE разрабатывался в Калифорнийском университете в Беркли [Nagel75], где было последовательно выпущено три его версии (Spice-1, 2 и 3). Позже появились различные модификации SPICE (например, PSPICE или HSPICE), предложившие коммерческую поддержку университетскому продукту. Кроме того, в последнее время мы наблюдаем развитие множества конкурирующих инструментов моделирования схем (таких как SPECTRE [Spectre] и ELDO [Eldo]), которые благодаря внедрению новых алгоритмов и современных технологий программирования предлагают лучшее быстродействие и точность.

В данной книге мы используем моделирование в основном для иллюстрации основных концепций цифровых схем и проверки моделей, разработанных вручную. Упражнения, предлагаемые для самостоятельного решения, также сильно связаны с моделированием схем. Таким образом, вы должны хорошо разбираться в свойствах и особенностях моделирования и анализа на определенном уровне абстракции.

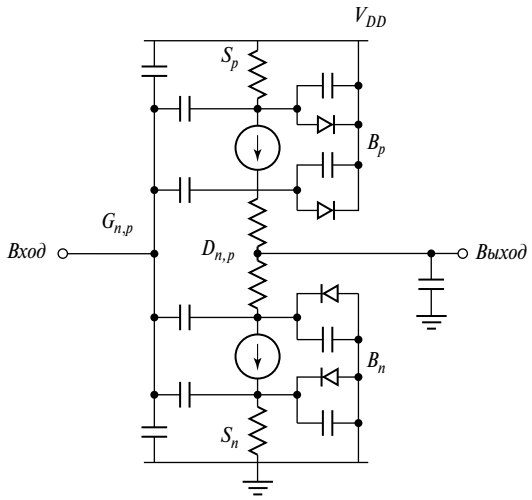


Рис. Б.1. Пример моделирования схемы КМОП-инвертора

Свойства схемного моделирования

Сформулируем для начала несколько свойств схемного моделирования.

- При анализе цифровой сети с использованием средства моделирования получающиеся *сигналы* напряжения и сила тока представляются в виде *непрерывных волн*.
- При анализе переходных процессов *время* полагается *непрерывной* переменной, и с практической точки зрения ее можно считать таковой. В действительности средство моделирования запускается на цифровом компьютере и вычисляет ограниченное число моментов времени, получая промежуточные данные посредством интерполяции.

Моделирование переходных процессов означает решение дифференциальных уравнений в каждый момент времени. Ситуацию усложняет еще и то, что точное моделирование полупроводниковых устройств требует введения нелинейности, например, в уравнения для тока в МОП-устройствах или диодно-конденсаторной модели. Например, на рис. Б.1 показана сеть, которую нужно рассчитать при анализе переходных процессов КМОП-инвертора. Все конденсаторы и источники тока в данной модели являются сильно нелинейными. Аналитическое решение набора дифференциальных уравнений, описывающих подобную сеть, обычно основывается на вычислительно емком итеративном подходе. На каждом шаге по времени текущее значение узла оценивается по значениям на предыдущих шагах. Эта оценка итеративно улучшается, пока не будет получена приемлемая точность. Чем строже заданный критерий ошибок, тем точнее будет конечный результат, однако тем больше потребуется итераций.

Режимы моделирования

SPICE предлагает множество режимов моделирования, предназначенных для охвата различных аспектов дизайна.

С точки зрения разработки цифровых схем особый интерес представляют два режима, которые кратко описаны ниже.

Развертка постоянной составляющей

В данном режиме пакет моделирования схем рассчитывает развертку постоянной составляющей.

При расчете передаточной функции напряжения схем нас интересует набор точек смещения для некоторого диапазона входных значений. Получить соответствующую кривую можно с помощью функции развертки постоянной составляющей. При развертке источника пакет начинает с одного значения, рассчитывает точку смещения постоянной составляющей, затем немного увеличивает значение, находит соответствующую точку смещения и т.д.

Описанный процесс продолжается до обработки последнего значения. В целом одна команда дает тот же результат, что и множество отдельных операций смещения. Типичная команда развертки постоянной составляющей выглядит примерно следующим образом:

```
.DC VIN 0 0.5 2.5
```

В приведенном примере источник напряжения `VIN` “заметается” с 0 до 2,5 В с шагом 0,5 В.

Анализ переходных процессов

Это самый распространенный тип анализа в моделировании цифровых схем. Анализ переходных процессов моделирует операции вашей схемы с течением времени. Он является идеальным инструментом для определения таких параметров переходных процессов, как задержка распространения, время нарастания и спада, рассеиваемая мощность. Анализ инициируется с помощью следующей строки:

```
.TRAN 1ps 250ps
```

В данном случае задано время моделирования (250 пс) и указано, как часто нужно записывать данные (каждую пикосекунду). Последняя величина не связана прямо с моментами времени, используемыми пакетом для расчетов, — пакет определяет их, исходя из сложности схемы.

SPICE (или эквивалентный пакет) предлагает множество дополнительных инструментов, позволяющих получать необходимую информацию. Если вас интересует этот вопрос — зайдите на Web-сайт данной книги, где предлагается множество полезных подсказок и другой информации.

Модели устройств

Решение уравнений в различных пакетах выполняется по похожим алгоритмам, но большее влияние на ценность пакета оказывают устойчивость, эффективность и точность моделей устройств. Например, досконально разработанные очень точные модели могут оказаться абсолютно бесполезными из-за плохой сходимости. Сложность поведения короткоканального МОП-транзистора и присущее ему множество паразитных эффектов привело к разработке большого набора моделей, характеризующихся различной степенью точности и вычислительной эффективности (некоторые из этих моделей были представлены в предыдущей главе).

Осознав это, компании-производители полупроводниковых структур решили выпускать собственные модели своих транзисторов. В результате появилась масса моделей, объективно судить о качестве которых было затруднительно (разумеется, каждая компания заявляла, что именно ее модель — самая лучшая). Подобная ситуация была неудобна и для потребителей, которым было довольно сложно разобраться в мешанине процессов и производителей. К счастью, ситуация кардинально изменилась после разработки в Калифорнийском университете в Беркли модели BSIM (Berkeley Short-Channel IGFET Model — Берклиевская модель короткоканального полевого транзистора с изолированным затвором). Через некоторое время модель BSIM3v3 была принята в качестве промышленного стандарта и теперь широко используется всеми производителями полупроводниковых структур. В 2000 году была выпущена более полная модель BSIM4. Кроме того, существуют модели BSIM, учитывающие альтернативные процессы, например SOI (BSIMSOI). Подробнее о BSIM можно узнать на Web-сайте Калифорнийского университета ([Bsim34]).

Предупреждение

Неопытные дизайнеры часто обладают почти фанатичной верой в результаты моделирования. В частности, нередко можно увидеть дизайнера, оптимизирующего задержки распространения схем КМОП в субпикосекундном диапазоне. Тем не менее следует помнить, что результат моделирования может отклоняться от реального результата из-за наличия множества источников ошибок: неточностей в моделях устройств, отклонений параметров устройств от заданных значений, паразитных сопротивлений и емкостей. Реальное и предсказанное поведение схемы может отличаться из-за изменения хода процессов вследствие изменений температуры. Таким образом, дизайнеры должны проводить четкую границу между проектными ограничениями и результатами моделирования. “Ширина” этой границы зависит от усилий, затраченных на моделирование, — насколько точными и тщательно разработанными являются модели устройства, насколько точно описаны паразитные нагрузки. Как правило, нормальным считается допуск до 10%.

Хотя для разработчика цифровых схем моделирование достаточно долго было любимой рабочей лошадкой, с появлением бо́льших схем оно стало непрактичным из-за вычислительной сложности. Следовательно, в настоящее время оно используется преимущественно для анализа критических участков структур. Для общего анализа применяются пакеты моделирования более высокого уровня. Эти пакеты жертвуют быстродействием в пользу точности или абстракции. Подробнее данные пакеты рассмотрены в следующих вставках.

Для любознательных

Пакет SPICE рассмотрен во многих учебниках (например, [Thorpe92, Vladimirescu93]). Кроме того, в Интернете можно найти большой объем информации, подобной руководствам пользователей и справочным материалам (например, [Rabaey96]).

Литература

[BSIM34] "BSIM, The Berkeley Short-Channel IGFET Model", `http://www-device.eecs.berkeley.edu/bsim3`.

[Eldo] "Eldo AMS Simulation", `http://www.mentor.com/eldo/overview.html`, Mentor Graphics.

[Huang99] X. Huang, W. C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y. K. Choi, K. Asano, B. Subramanian, T.J. King, J bokor, and C. Hu, "Sub 50-nm FinFET: PMOS", International Electron Devices Meeting, 1999, p. 67–70.

[Nagel75] L. Nagel, "SPICE2: a Computer Program to Simulate Semiconductor Circuits", Memo ERL-M520, Dept. Elect. and Computer Science, University of California at Berkeley, 1975.

[Rabaey96] J. Rabaey (et al), "The Spice Page", `http://bwrc.eecs.berkeley.edu/ICbook/spice`, 1996.

[Spectre] "Spectre Circuit Simulator", `http://www.cadence.com/datasheets/spectre_cir_sim.html`, Cadence.

[Thorpe92] T. Thorpe, *Computerized Circuit Analysis with SPICE*, John Wiley and Sons, 1992.

[Vladimirescu93] A. Vladimirescu, *The SPICE Book*, John Wiley and Sons, 1993.

ГЛАВА 4

Проводник

Определение и численный расчет параметров межсоединения

Модели схем для соединительных проводников

Подробные модели проводов для SPICE

Масштабирование технологии и его влияние на соединения

4.1. Введение
4.2. На первый взгляд
4.3. Параметры соединения — емкость, сопротивление и индуктивность
4.3.1. Емкость
4.3.2. Сопротивление
4.3.3. Индуктивность
4.4. Модели электрических проводников
4.4.1. Идеальный проводник
4.4.2. Модель с сосредоточенными параметрами
4.4.3. Сосредоточенная RC-модель
4.4.4. Распределенная RC-цепь
4.4.5. Линия передачи
4.5. SPICE-модели проводов
4.5.1. Распределенные RC-линии в SPICE
4.5.2. SPICE-модели линии передачи
4.5.3. Перспектива: взгляд в будущее
4.6. Резюме
4.7. Для любознательных

4.1. Введение

За все время существования интегральных схем внутрикристальные межсоединения считались "вторым сортом" и рассматривались только в частных случаях или при выполнении высокоточного анализа. С введением субмикронных полупроводниковых технологий данная ситуация стала быстро меняться. Для паразитных эффектов, вводимых проводниками, характерен эффект масштабирования, отличный от того, что наблюдается в активных устройствах, подобных транзисторам, причем значимость этих эффектов увеличивается при уменьшении размеров устройства и повышении скорости переключения схемы. Фактически данные эффекты начинают подавлять такие важные характеристики интегральных схем, как скорость, энергопотребление и надежность. Ситуация усугубляется еще и тем, что улучшение технологии делает создание еще бо́льших кристаллов интегральных схем экономически оправданным, а это приводит к увеличению средней длины соединяющих проводников (межсоединений) и связанных с ними паразитных эффектов. Исходя из сказанного, можно сделать вывод, что аккуратный и глубокий анализ роли и поведения межсоединений в полупроводниковой технологии является не только желательным, но и необходимым.

4.2. На первый взгляд

Разработчик электронной схемы может по-разному представлять соединения между различными устройствами, составляющими эту схему. Современные процессы позволяют получать множество слоев алюминия или меди и по крайней мере один слой поликремния. В качестве связующего звена могут применяться даже высоколегированные диффузионные слои n^+ или p^+, обычно используемые для реализации областей истока и стока. Все эти проводники показываются на схематической диаграмме электронной схемы простыми линиями, на первый взгляд не имеющими никакого влияния на быстродействие цепи. Из приводившегося ранее описания процесса производства интегральной схемы должно быть ясно, что такое представление является слишком упрощенным. На самом деле проводка современных интегральных схем формирует сложную геометрию, вводящую паразитные емкости, индуктивности и сопротивления, имеющие сложное влияние на поведение схемы.

1. Все они увеличивают задержку распространения (т.е. снижают быстродействие).
2. Все они влияют на рассеяние энергии и энергетический спектр.
3. Все они приводят к появлению дополнительных источников шума, влияющих на надежность схемы.

Разработчик может решить обезопасить себя, включить в анализ все паразитные эффекты и разработать процесс оптимизации. Такой консервативный подход не является конструктивным (а часто он вообще нереален). Прежде

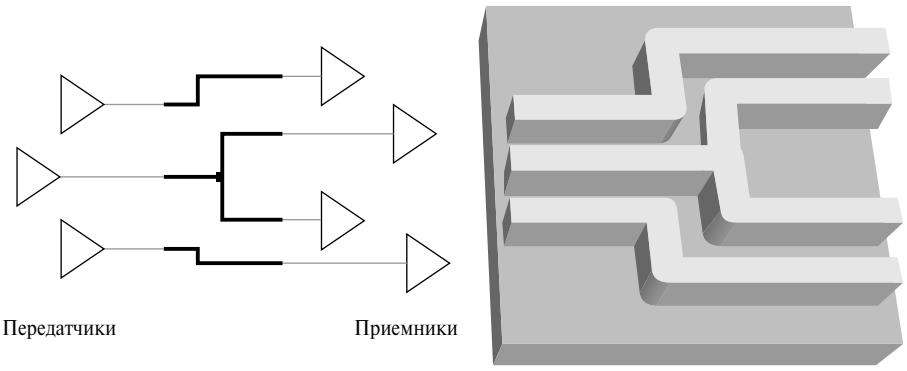


Рис. 4.1. Схематическое и физическое представление проводки магистральной сети. Физическое представление показано для области, выделенной на рисунке слева

всего, "полная" модель невероятно сложна, поэтому применяется только для очень маленьких топологий. Следовательно, она совершенно бесполезна для современных интегральных схем, содержащих миллионы узлов. Более того, такой подход неудачен тем, что "за деревьями не видно леса". Поведение схемы в данном узле определяется только несколькими основными параметрами. Учитывая все возможные эффекты, мы можем скрыть основную картину и превратить процессы оптимизации и разработки в игру проб и ошибок, а не в элегантный целенаправленный поиск.

Таким образом, разработчику нужно иметь четкое представление о паразитных эффектах проводки, их относительной важности и способах моделирования. Сказанное лучше всего проиллюстрировать на простом примере (рис. 4.1). Каждый проводник магистральной сети соединяет передатчик (или передатчики) с набором приемников и реализуется как цепь проводников, имеющих различную длину и геометрию. Предположим, что все сегменты реализованы на одном слое межсоединений и изолированы от кремниевой подложки и друг от друг посредством слоя диэлектрического материала. (Помните, что в реальности все может быть гораздо сложнее.)

Полномасштабная модель схемы, учитывающая паразитные сопротивления, емкости и индуктивности межсоединений показана на рис. 4.2, *а*. Обратите внимание на то, что дополнительные элементы схемы не расположены в разных физических точках, а распределены по длине проводника. Такое представление необходимо, когда длина проводника становится гораздо больше ее толщины. Кроме названных элементов, в схеме еще присутствует паразитное взаимодействие проводников, приводящее к интерференционному взаимодействию сигналов различных шин, которое не учитывается на исходной схеме.

Анализировать поведение такой схемы, моделирующей только часть реальной цепи, очень сложно и неудобно. К счастью, задачу можно существенно упростить, если принять следующие допущения.

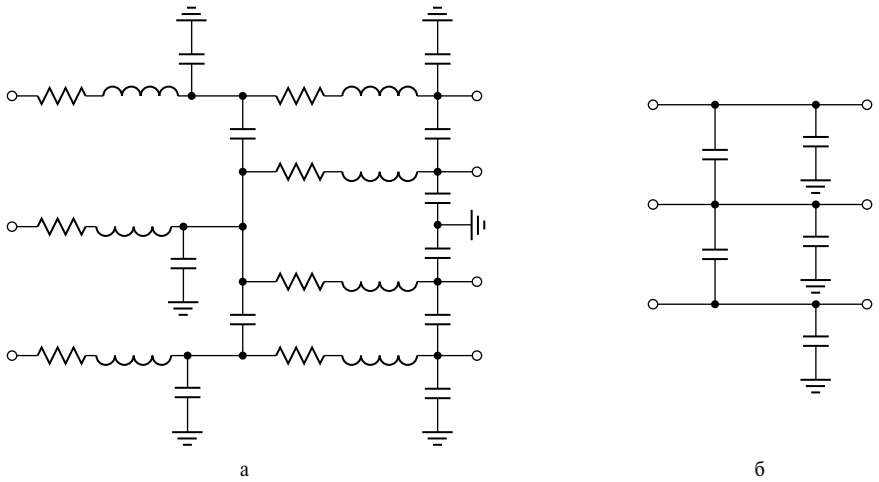


Рис. 4.2. Модели проводников для схемы, показанной на рис. 4.1: а) в модели учитываются все паразитные нагрузки проводников (исключая внутреннее сопротивление и взаимоиндукцию); б) в модели учитываются только емкости

- Индуктивные эффекты можно игнорировать, если сопротивление проводника достаточно велико (это справедливо, например, для длинных алюминиевых проводников с небольшим поперечным сечением или для сигналов, время нарастания и спада которых мало).
- Если проводники короткие с большим поперечным сечением или материал межсоединений имеет низкое сопротивление, можно использовать модель, учитывающую только сопротивления (см. рис. 4.2, *б*).
- Наконец, если соседние провода достаточно разнесены или сближаются только на небольшом участке, перекрестная емкость проводников и все паразитные емкости можно смоделировать как емкость относительно земли.

Очевидно, последние случаи гораздо легче моделировать, анализировать и оптимизировать. Опытный дизайнер знает, как разделить главные и второстепенные эффекты. В данной главе мы представим основные технологии оценки значений различных параметров соединения; простые модели, позволяющие оценить их влияние; а также набор эмпирических правил, позволяющих определить, когда и где следует рассматривать конкретную модель или эффект.

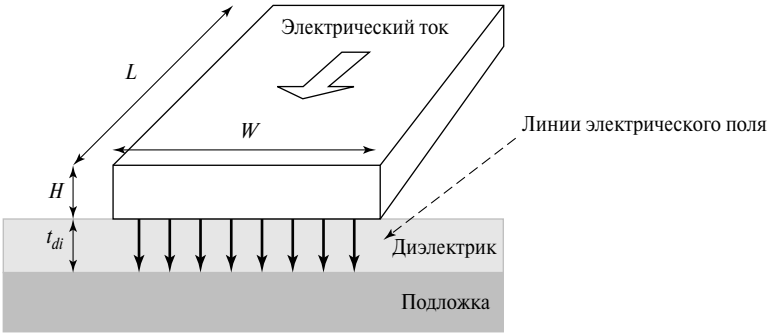


Рис. 4.3. Моделирование соединительного проводника через емкость параллельных пластинок

4.3. Параметры соединения — емкость, сопротивление и индуктивность

4.3.1. Емкость

Точное моделирование емкости (емкостей) проводника в современных интегральных схемах является нетривиальной задачей, и поиск ее удачных решений продолжается даже сейчас. Задача усложняется еще и тем, что структура соединений современных интегральных схем является трехмерной, что отчетливо видно по поперечному сечению интегральной схемы, показанному на рис. 2.8. Емкость такого проводника является функцией его формы, окружающих элементов, расстояния до подложки и окружающих проводников. Вместо того чтобы "зарываться" в сложнейшие уравнения и модели, дизайнер обычно использует сложные инструменты-зонды, позволяющие получать точные значения межсвязных емкостей готового слоя. Большинство производителей полупроводников также предоставляют эмпирические данные (измерения, полученные на тестовых кристаллах), описывающие вклады различных емкостей. Чтобы просто получить представление о природе межсвязной емкости и ее параметрах, а также оценить, как изменится емкость проводников с введением новых технологий, до сих пор применяются простые модели первого порядка.

Рассмотрим для примера прямоугольный проводник, размещенный над полупроводниковой подложкой (рис. 4.3). Если ширина проводника значительно больше толщины изолирующего материала, можно предположить, что линии напряженности электрического поля ортогональны обкладкам конденсатора, емкость которого можно смоделировать с помощью *модели конденсатора с параллельными обкладками*. В таком случае общую емкость проводника можно аппроксимировать следующим образом[1]:

[1] Чтобы отличать удельные (на единицу длины) параметры проводника от сосредоточенных значений, будем обозначать первые строчными буквами, а вторые — прописными.

Таблица 4.1. Относительная диэлектрическая проницаемость некоторых распространенных материалов

Материал	ϵ_r
Вакуум	1
Аэрогели	$\sim 1,5$
Полиимиды (органические)	3–4
Диоксид кремния	$3,9$
Стеклоэпоксид (печатная плата)	5
Нитрид кремния (Si_3N_4)	$7,5$
Глинозем (корпус)	$9,5$
Кремний	$11,7$

$$C_{int} = \frac{\varepsilon_{di}}{t_{di}} WL, \tag{4.1}$$

где W и L — ширина и длина проводника; t_{di} и e_{di} представляют толщину слоя диэлектрика и его диэлектрическую проницаемость. В качестве диэлектрического материала в интегральных схемах используется SiO_2, хотя существуют и другие материалы с более низкой диэлектрической проницаемостью, а следовательно, более низкой емкостью. В качестве примеров таких материалов, завоевывающих все бо́льшую популярность, можно привести органические полиимиды и аэрогели. Величина ε обычно выражается в виде произведения двух членов: $\varepsilon = \varepsilon_r \varepsilon_0$, где $\varepsilon_0 = 8,854 \times 10^{-12}$ Ф/м — диэлектрическая проницаемость вакуума, а ε_r — относительная диэлектрическая проницаемость материала. Относительные диэлектрические проницаемости нескольких диэлектриков, используемых в интегральных схемах, приведены в табл. 4.1. Другими словами, в уравнении (4.1) говорится, что емкость пропорциональна перекрытию проводников и обратно пропорциональна расстоянию между ними.

Как и в предыдущих примерах, приведенная модель в действительности является слишком упрощенной. Чтобы минимизировать сопротивление проводником при переносе технологии на более мелкий масштаб, поперечное сечение проводника ($W \times H$) желательно поддерживать максимально большим. (Данное требование объясняется в следующем разделе.) С другой стороны, небольшие значения W дают более плотное расположение проводников, а следовательно, менее жесткие требования к размерам. В результате в последние годы наблюдается тенденция к уменьшению отношения W/H, так что в сложных процессах оно уже составляет меньше единицы. Данный момент отчетливо виден на поперечном сечении, представленном на рис. 2.8. В таком случае описанная ранее модель параллельных пластинок становится неточной. Теперь емкость между боковыми стенками проводников и подложкой (*краевая емкость*) уже нельзя игнорировать и следует учитывать ее вклад в общую емкость. Описанный эффект показан на рис. 4.4, *а*. Представить точную модель для такой геометрии довольно сложно, поэтому будем придерживаться принятой практики и используем упрощенную модель, аппроксимирующую емкость суммой двух компонентов (см. рис. 4.4, *б*): емкости параллельных

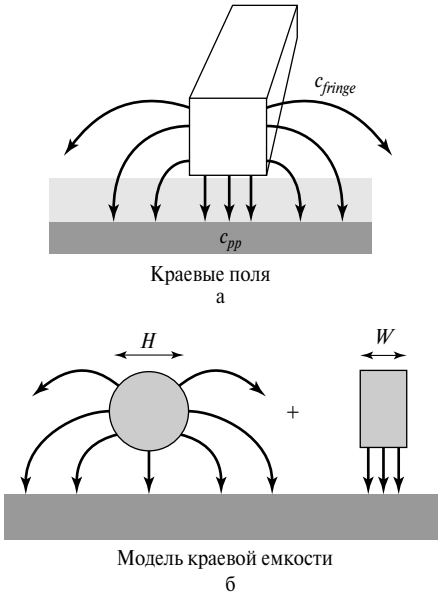


Рис. 4.4. Краевая емкость. В представленной модели емкость разделяется на два вклада: емкость параллельных пластинок и краевую емкость, моделируемую цилиндрическим проводником, диаметр которого равен толщине провода

пластинок, определяемой ортогональным полем между проводником ширины w и пластиной заземления, и параллельно подключенной сосредоточенной емкости, моделируемой цилиндрическим проводником с диаметром, равным толщине межсоединения H. В результате получаем следующую аппроксимацию, сравнительно простую и дающую хорошие результаты на практике:

$$C_{wire} = C_{pp} + C_{fringe} = \frac{W\varepsilon_{di}}{t_{di}} + \frac{2\pi\varepsilon_{di}}{\log(t_{di}/H)}, \tag{4.2}$$

где $w = W - H/2$ — хорошая аппроксимация толщины конденсатора с параллельными обкладками. Существует множество более точных моделей (например, [Vdmeijs84]), которые мы не будем рассматривать из-за их излишней сложности.

Чтобы проиллюстрировать важность краевого компонента, рассмотрим рис. 4.5, где представлена зависимость емкости проводки от (W/t_{di}) (или от (W/H)). Для бо́льших значений W/H суммарная емкость близка к значениям модельного конденсатора с параллельными обкладками. Если W/H меньше 1,5, краевой компонент в действительности становится доминирующим. При небольшой длине провода краевая емкость может увеличить общую емкость более чем в 10 раз. Интересно отметить, что в том случае, если ширина линии меньше толщины диэлектрика общая емкость выходит на постоянное значение, равное приблизительно 1 пФ/см. Другими словами, емкость уже не является функцией ширины.

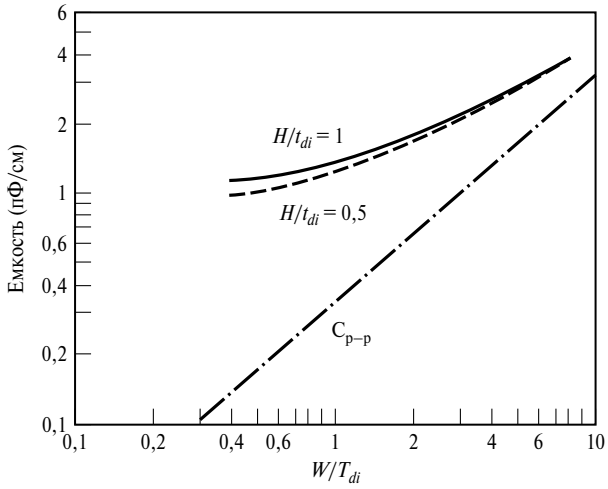


Рис. 4.5. Емкость соединительного проводника как функция W/t_{di} с учетом краевых эффектов (согласно [Schaper83]). Рассмотрены два значения H/t_{di}. В качестве диэлектрика использован диоксид кремния с $\epsilon_r = 3,9$

До этого момента мы ограничивали анализ разбором простого прямоугольного проводника, помещенного на пластину заземления. Такая структура, именуемая *микрополосковой линией*, считается хорошей моделью полупроводниковых межсоединений, когда число уровней межсоединений равно 1 или 2. Современные процессы позволяют получать гораздо больше уровней межсоединений, которые, кроме того, располагаются довольно плотно. В подобной ситуации предполагать, что проводник полностью изолирован от окружающих структур и имеет емкостное соединение только с землей, безосновательно. Сказанное иллюстрируется на рис. 4.6, где обозначены емкостные компоненты проводника, внедренного в иерархию межсоединений. Все проводники связаны не только с заземленной подложкой, но и с соседними проводниками того же и соседних уровней. В первом приближении это не меняет общую емкость данного проводника. Основным отличием является то, что не все емкостные компоненты проводника заканчиваются на заземленной подложке — большое их число соединяется с другими проводниками, уровни напряжения которых колеблются в очень широких пределах. Позже мы увидим, что данные *"плавающие" конденсаторы* являются не только источником шума (*перекрестные помехи*), но и могут отрицательно влиять на быстродействие схемы.

В целом перекрестная емкость проводников начинает доминировать в многоуровневых связанных структурах. Данный эффект более заметен для проводников более высоких уровней, поскольку эти проводники более удалены от подложки. Повышающийся вклад перекрестной емкости проводников в общую емкость с уменьшением размера элементов графически представлен на рис. 4.7. На этом графике, представляющем зависимость емкостных компонентов набора параллельных проводников, расположенных над плоско-

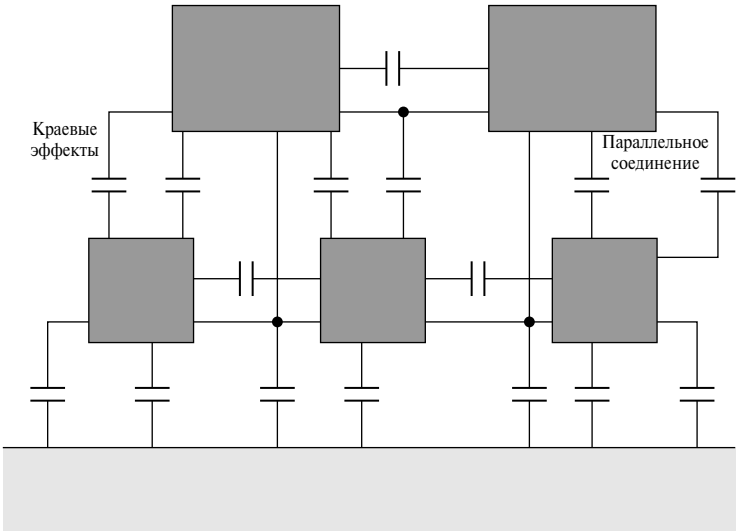


Рис. 4.6. Емкостная связь проводников в иерархии межсоединений

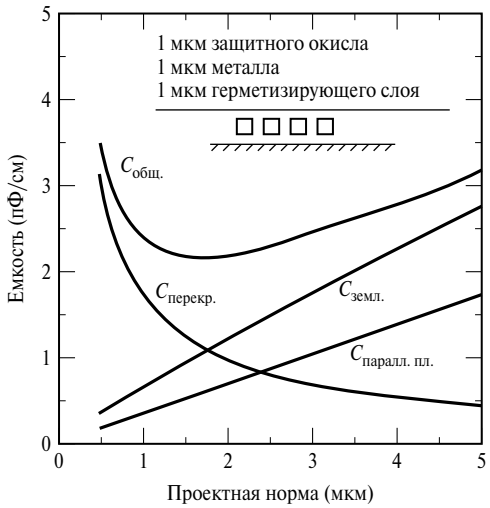


Рис. 4.7. Емкость межсоединения как функция проектных норм. Состоит из емкости относительно земли и перекрестной емкости проводников (согласно [Schaper83])

стью заземления, предполагается, что толщина диэлектрика и проводников постоянна, а остальные размеры изменяются. Когда W становится меньше $1,75H$, перекрестная емкость проводников начинает доминировать.

Таблица 4.2. Поверхностная и краевая емкость для типичного 0,25-микронного КМОП-процесса. Строки таблицы представляют верхнюю, а столбцы — нижнюю обкладку конденсатора. Поверхностная емкость выражена в аФ/мкм^2, а краевая емкость (в четных строках) — в аФ/мкм

	Защитный слой	Активный слой	Поликремний	Al1	Al2	Al3	Al4
Поликремний	88						
	54						
Al1	30	41	57				
	40	47	54				
Al2	13	15	17	36			
	25	27	29	45			
Al3	8,9	9,4	10	15	41		
	18	19	20	27	49		
Al4	6,5	6,8	7	8,9	15	35	
	14	15	15	18	27	45	
Al5	5,2	5,4	5,4	6,6	9,1	14	38
	12	12	12	14	19	27	52

Информация о проектировании емкости межсоединений

В табл. 4.2 представлены типичные емкости межсоединений для стандартного 0,25-микронного КМОП-процесса. Процесс поддерживает один уровень поликремния и пять уровней алюминия. Четыре первых уровня металла имеют одинаковую толщину и используют похожий диэлектрик, тогда как проводники на пятом металлическом уровне примерно вдвое толще и внедрены в диэлектрик с более высокой диэлектрической проницаемостью. Если проводники помещаются на толстый защитный слой окисла, используемого для разделения различных транзисторов, необходимые значение берутся из столбца “Защитный слой”. Проводники, проложенные над активной областью, имеют более высокую емкость, представленную в столбце “Активный слой”. Помните, что указанные значения представляют усредненные данные. Чтобы получить более точные результаты для реальных структур, необходимы сложные трехмерные модели, учитывающие окружение проводника. Характерные значения для емкостей параллельных проводников, помещенных на одном уровне на минимальном расстоянии друг от друга (согласно проектным нормам), представлены в табл. 4.3. Обратите внимание на то, что здесь учитываются вклады обоих компонентов: того, что моделируется параллельными пластинками, и краевого компонента. Повторимся, емкости сильно зависят от топологии системы. Например, наличие на соседнем уровне заземленного слоя ограничивает бо́льшую часть краевого поля и эффективно уменьшает перекрестную емкость (хотя общая емкость, вводимая проводником, может даже немного увеличиться). Поликремниевые проводники обычно демонстрируют более низкую перекрестную емкость из-за своей меньшей толщины. С другой стороны, толстые провода Al5 дают наиболее высокую межпроводниковую емкость. Исходя из этого рекомендуется либо разделять проводники на этом слое расстоянием, превышающем минимально допустимое,

Таблица 4.3. Перекрестная емкость на единицу длины проводника для различных уровней межсоединений типичного 0,25-микронного КМОП-процесса. Удельные емкости рассчитаны для минимально разнесенных проводников и представлены в аФ/мкм

Слой	Поликремний	Al1	Al2	Al3	Al4	Al5
Емкость	40	95	85	85	85	115

либо использовать их для передачи глобальных сигналов, нечувствительных к интерференции (например, применять их в качестве шин питания.)

Пример 4.1. Емкость металлического провода

Некоторые глобальные сигналы, например тактовые импульсы, распространяются по всему кристаллу, причем длина соответствующих проводников может быть значительной. Если размер кристалла интегральной схемы составляет порядка 1-2 см, проводники могут иметь длину до 10 см и приличную емкость. Рассмотрим, например, алюминиевый провод длиной 10 см и толщиной 1 мкм, проложенный на первом алюминиевом слое. Для расчета общей емкости используем данные из табл. 4.2:

Поверхностная емкость (плоский конденсатор)	$(0,1 \times 10^6 \text{ мкм}^2) \times 30 \text{ аФ/мкм}^2 = 3 \text{ пФ}$
Краевая емкость	$2 \times (0,1 \times 10^6 \text{ мкм}) \times 40 \text{ аФ/мкм} = 8 \text{ пФ}$
Общая емкость	11 пФ

Обратите внимание на множитель 2 при расчете краевой емкости — он нужен для учета двух сторон проводника.

Предположим теперь, что вдоль первого проводника проложен второй, и их разделяет минимально допустимое расстояние. Используя табл. 4.3, мы можем определить, что второй провод будет соединяться с первым емкостью, равной

$$C_{\text{inter}} = (0,1 \times 10^6 \text{ мкм}) \times 95 \text{ аФ/мкм} = 9,5 \text{ пФ},$$

что практически равно общей емкости относительно земли!

С помощью похожего примера можно показать, что замена провода на Al4 уменьшит емкость относительно земли до 3,45 пФ (0,65 пФ поверхностной емкости и 2,8 пФ краевой), тогда как перекрестная емкость останется приблизительно прежней (примерно 8,5 пФ).

Таблица 4.4. Сопротивление распространенных проводников (при 20°C)

Материал	ρ (Ом·м)
Серебро (Ag)	$1,6 \times 10^8$
Медь (Cu)	$1,7 \times 10^8$
Золото (Au)	$2,2 \times 10^8$
Алюминий (Al)	$2,7 \times 10^8$
Вольфрам (W)	$5,5 \times 10^8$

4.3.2. Сопротивление

Сопротивление проводника пропорционально его длине L и обратно пропорционально площади его поперечного сечения A. Сопротивление прямоугольного проводника, показанного на рис. 4.3, выражается следующей формулой:

$$R = \frac{\rho L}{A} = \frac{\rho L}{HW}, \tag{4.3}$$

где константа c — удельное сопротивление материала (выражается в Ом·м). Удельные сопротивления некоторых распространенных материалов приведены в табл. 4.4. В интегральных схемах чаще всего используется алюминий (из-за его дешевизны и совместимости со стандартным процессом производства интегральных схем). К сожалению, алюминий имеет слишком большое сопротивление по сравнению с такими материалами, как медь. Таким образом, поскольку требования быстродействия становятся все более и более важными, в современных процессах в качестве проводника все интенсивнее используется медь.

Поскольку для данной технологии H — константа, уравнение уравнение (4.3) можно переписать в следующем виде:

$$R = R_{\square}\frac{L}{W}, \tag{4.4}$$

где

$$R_{\square} = \frac{\rho}{H} \tag{4.5}$$

представляет собой *удельное поверхностное сопротивление* материала, измеряемое в Ом/□ (читается как "Ом на единицу площади"). Таким образом, из уравнения (4.4) видно, что сопротивление квадратного проводника не зависит от его абсолютного размера. Чтобы получить сопротивление проводника, умножьте удельное поверхностное сопротивление на отношение его длины L к ширине W.

Проектные нормы относительно сопротивления межсоединений

Типичные значения удельных поверхностных сопротивление различных материалов, используемых в полупроводниковых схемах, представлены в табл. 4.5.

Таблица 4.5. Удельное поверхностное сопротивление для типичного 0,25-микронного КМОП-процесса

Материал	Удельное поверхностное сопротивление (Ом/□)
Диффузия n- или p-кармана	1000–1500
Диффузный слой n^+, p^+	50–150
Диффузный слой n^+, p^+ с силицидом	3–5
Поликремний n^+, p^+	150–200
Поликремний n^+, p^+ с силицидом	4–5
Алюминий	0,05–0,1

Изучая таблицу, можно сделать вывод, что алюминий является наиболее предпочтительным материалом для длинных проводников. Поликремний должен использоваться только для локальных межсоединений. Хотя удельное поверхностное сопротивление диффузных уровней (n^+, p^+) сравнимо с сопротивлением поликремния, использование диффузионных проводников рекомендуется избегать из-за большой емкости и соответствующей RC-задержки.

Сложные процессы также позволяют использовать силицидированные поликремниевые и диффузионные уровни. Силицид — это составной материал, получаемый соединением кремния и тугоплавкого металла. В результате получается материал с высокой проводимостью, который может выдерживать высокотемпературные процессы без плавления. В качестве примеров силицидов можно привести WSi_2 (кремнистый вольфрам), $TiSi_2$ (кремнистый титан), $PtSi_2$ (кремнистый свинец) и TaSi (кремнистый тантал). WSi_2, например, обладает сопротивлением 130 мкОм-см, что приблизительно в восемь раз меньше, чем у поликремния. Чаще всего силициды используются в конфигурации, называемой *полисидом*, который представляет собой простую многоуровневую комбинацию поликремния и силицида. Типичный полисид состоит из нижнего слоя поликремния и верхнего слоя силицида, объединяя лучшие свойства обоих материалов — хорошее сцепление и покрытие (поликремний) и высокую проводимость (силицид). На рис. 4.8 показан канальный полевой униполярный МОП-транзистор с полисидным затвором. Достоинством силицидных затворов является уменьшенное сопротивление затвора. Подобным образом, силицидированные области истока и стока уменьшают истоковое и стоковое сопротивление устройства.

Переходы между слоями трассировки добавляют к проводнику дополнительное сопротивление, называемое *контактным* (или *сопротивление перехода*). В связи с этим рекомендуется по возможности располагать сигнальные шины на одном уровне и избегать лишних контактов или сквозных отверстий. Контактное сопротивление можно уменьшить, увеличив контактные отверстия. К сожалению, современные тенденции заключаются в концентрации элементов по периметру контактного отверстия. Данный эффект, именуемый *сжатием тока*, вводит практическое ограничение на размер контакта. Для типичного 0,25-микронного процесса обычными являются следующие контактные сопротивления (для контактов минимального размера): 5,20 Ом для переходов металла (или поликремний)/n^+, p^+, а также переходов металл/поликремний; 1,5 Ом — для отверстий (контактов металл–металл).

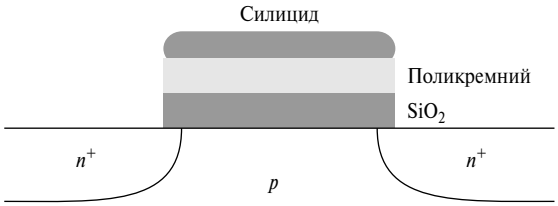


Рис. 4.8. Канальный полевой униполярный МОП-транзистор с полисидным затвором

Пример 4.2. Сопротивление металлического провода

Вернемся еще раз к алюминиевому проводу, рассмотренному в примере 4.1 (длина 10 см, толщина 1 мкм, проложен на первом алюминиевом слое). Считая, что удельное поверхностное сопротивление алюминия равно 0,075 Ом/□, можно вычислить общее сопротивление проводника:

$$R_{\text{wire}} = 0,075 \text{ Ом/}\square \times (0,1 \times 10^6 \text{ мкм})/(1 \text{ мкм}) = 7,5 \text{ кОм}.$$

Если сделать тот же провод из поликремния с удельным поверхностным сопротивлением 175 Ом/□, общее сопротивление возрастет до 17,5 МОм, что, разумеется, является неприемлемым. Силицидированный поликремний с удельным сопротивлением 4 Ом/□ будет лучшей альтернативой, хотя он также имеет сравнительно высокое сопротивление 400 кОм.

До этого момента мы считали сопротивление полупроводникового провода постоянной линейной функцией длины. Такое допущение действительно справедливо для большинства полупроводниковых схем. Тем не менее при очень высоких частотах появляется дополнительное явление (*скин-эффект*), проявляющееся в том, что сопротивление начинает зависеть от частоты. Высокочастотный ток протекает преимущественно по поверхности провода, причем плотность тока экспоненциально падает с увеличением глубины проводника. *Глубина скин-слоя* δ определяется как расстояние, на котором ток в e раз меньше номинального значения, и представляется следующей формулой:

$$\delta = \sqrt{\frac{\rho}{\pi f \mu}}, \tag{4.6}$$

где f — частота сигнала; μ — магнитная проницаемость окружающего диэлектрика (обычно равная магнитной проницаемости вакуума, или *магнитной постоянной*, $\mu = 4\pi \times 10^7$ Гн/м). Для алюминиевого провода при частоте 1 ГГц, глубина скин-слоя составляет 2,6 мкм. В связи с этим, естественно, возникает вопрос, что это означает и должны ли мы как-то учитывать это при разработке современных цифровых схем.

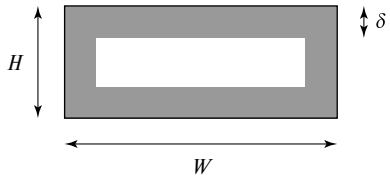


Рис. 4.9. Скин-эффект ограничивает протекание тока поверхностью проводника

Скин-эффект можно аппроксимировать, предположив, что ток протекает равномерно по внешней оболочке прямоугольного проводника, имеющей толщину δ (рис. 4.9). Считая теперь, что общее поперечное сечение проводника приблизительно ограничено величиной $2(W + H)\delta$, получаем следующее выражение для сопротивления (на единицу длины) при высоких частотах ($f > f_s$):

$$r(f) = \frac{\sqrt{\pi f \mu \rho}}{2(H + W)}. \tag{4.7}$$

Повышение сопротивления при высоких частотах может вызвать дополнительное поглощение (а следовательно, искажение) сигнала, передаваемого по проводу. Чтобы определить, с какого момента начинается скин-эффект, можно найти частоту f_s, при которой глубина скин-эффекта равна половине большего размера (W или H) проводника. Ниже f_s ток проводится всем проводником, и сопротивление равно (постоянному) низкочастотному сопротивлению провода. Значение f_s находится из уравнения (4.6):

$$f_s = \frac{4\rho}{\pi \mu (\max(W, H))^2}. \tag{4.8}$$

Пример 4.3. Скин-эффект и алюминиевые провода

Чтобы определить, какое влияние имеет скин-эффект в современных интегральных схемах, проанализируем алюминиевый провод с сопротивлением $2,7 \times 10^8$ Ом·м, внедренный в диэлектрик (SiO_2) с магнитной проницаемостью $4\pi \times 10^7$ Гн/м. Из уравнения (4.8) находим, что, для того, чтобы эффект был заметен при частоте 1 ГГц, больший размер провода должен быть равен по крайней мере 5,2 мкм. Сказанное подтверждается более точными результатами моделирования, показанными на рис. 4.10, где представлено увеличение сопротивления, вызванное скин-эффектом, для алюминиевых проводников разной ширины. При частоте 1 ГГц и проводе шириной 20 мкм можно наблюдать 30%-ное увеличение сопротивления, тогда как увеличение сопротивления для провода толщиной 1 мкм не превышает 2%.

Итак, сделаем выводы. Скин-эффект следует учитывать только для широких проводов. Поскольку в кристалле самыми высокочастотными явля-

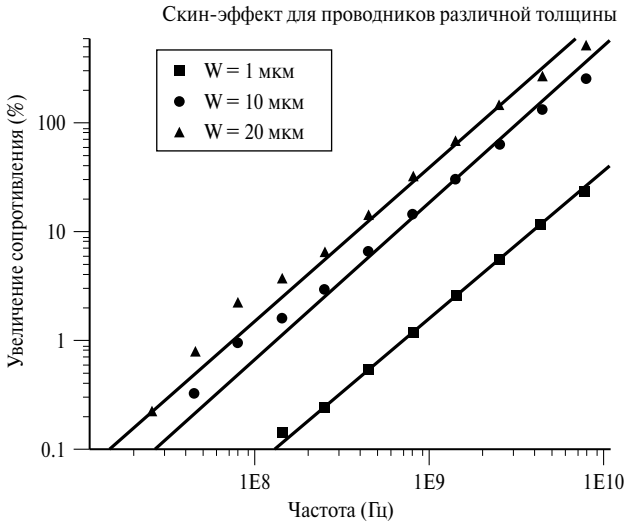


Рис. 4.10. Увеличение сопротивления вследствие скин-эффекта в зависимости от частоты и ширины провода. Все расчеты выполнялись для проводников толщиной 0,7 мкм [Sylvester97]

ются сигналы тактовых импульсов (к тому же соответствующие шины достаточно широкие, чтобы ограничивать сопротивление), наиболее вероятным местом проявления скин-эффекта являются именно эти шины. Действительно, данный вопрос является достаточно серьезным при работе в гигагерцовом диапазоне, поскольку тактовые импульсы определяют общее быстродействие кристалла (время цикла, число команд, выполняемых за секунду, и т.д.). Еще один набор вопросов связан с тем, что использование более подходящих проводников (например, медных) может привести к смещению области скин-эффекта в сторону более низких частот.

4.3.3. Индуктивность

Разработчики интегральных схем обычно относятся к индуктивности как к понятию, которое рассматривается в школьном курсе физики, но к их нынешней работе не имеет никакого отношения. Да, так действительно было в первое десятилетие развития цифровых интегральных схем, однако с появлением соединительных материалов с малым сопротивлением и увеличением частот коммутации до сверхгигагерцового диапазона индуктивность (даже в кристалле) стала играть весьма важную роль. Последствиями наличия внутрикристальной индуктивности стали эффекты “звона” и перегрузки, отражение сигналов из-за рассогласования импедансов, индуктивная связь линий, а также коммутационные помехи, вызванные перепадами напряжения (Ldi/dt).

Индуктивность участка цепи всегда можно рассчитать по определению, в котором говорится, что изменение силы тока, проходящего через катушку индуктивности, порождает скачок напряжения:

Таблица 4.6. Диэлектрические константы и скорость распространения электромагнитных волн для различных материалов, используемых в электрических схемах. Относительная магнитная проницаемость μ_r большинства диэлектриков приблизительно равна 1

Диэлектрик	ϵ_r	Скорость распространения (см/нс)
Вакуум	1	30
SiO_2	3,9	15
Печатная плата (эпоксидное стекло)	5,0	13
Глинозем (керамический корпус)	9,5	10

$$\triangle V = L\frac{di}{dt}. \tag{4.9}$$

Индуктивность провода можно рассчитать непосредственно из его геометрии и окружения. В частности, можно использовать тот факт, что емкость c и индуктивность l (на единицу длины) проводника связаны следующим соотношением:

$$cl = \varepsilon\mu, \tag{4.10}$$

где ε и μ являются соответственно диэлектрической и магнитной проницаемостью окружающего диэлектрика. Чтобы приведенное выражение было справедливым, проводник должен быть полностью окружен однородной диэлектрической средой. На практике так бывает не всегда. Тем не менее, даже если провод внедрен в среду, состоящую из различных диэлектрических материалов, можно ввести "средние" диэлектрические константы, что позволит использовать уравнение (4.10) для расчета приблизительного значения индуктивности.

Здесь можно отметить еще несколько интересных соотношений, следующих из законов Максвелла. Константа, равная произведению диэлектрической и магнитной проницаемости, также определяет скорость, с которой электромагнитная волна может распространяться в среде:

$$\nu = \frac{1}{\sqrt{lc}} = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c_0}{\sqrt{\varepsilon_r\mu_r}}, \tag{4.11}$$

где c_0 равно скорости света в вакууме (30 см/нс). Скорости распространения электромагнитных волн в других средах, используемых в производстве электрических цепей, приводятся в табл. 4.6. В частности, скорость распространения электромагнитной волны в SiO_2 вдвое меньше, чем в вакууме.

Пример 4.4. Индуктивность полупроводникового провода

Рассмотрим провод Al1 (т.е первый слой алюминия), реализованный по 0,25-микронной технологии КМОП и проложенный поверх защитного окисла.

Используя табл. 4.2, можем найти емкость проводника на единицу длины:

$$c = (W \times 30 + 2 \times 40) \text{ аФ/мкм}.$$

Из уравнения (4.10) можно вывести индуктивность на единицу длины провода, предполагая, что однородный диэлектрик состоит из SiO_2 (следите за использованием правильных единиц измерения!):

$$l = (3.9 \times 8,854 \times 10^{-12}) \times (4\pi 10^{-7})/c.$$

Для провода с размерами 0,4, 1 и 10 мкм получаем следующее:

$$W = 0,4 \text{ мкм} : c = 92 \text{ аФ/мкм}; l = 0,47 \text{ пГн/мкм};$$
$$W = 1 \text{ мкм} : c = 110 \text{ аФ/мкм}; l = 0,39 \text{ пГн/мкм};$$
$$W = 10 \text{ мкм} : c = 380 \text{ аФ/мкм}; l = 0,11 \text{ пГн/мкм}.$$

Считая, что удельное поверхностное сопротивление равно 0,075 Ом/□, можем также определить сопротивление провода:

$$r = 0,075/W \text{ Ом/мкм}.$$

Интересно отметить, что индуктивная часть импеданса провода становится равной значению резистивного компонента при частоте 30,6 ГГц (для проводника толщиной 1 мкм), что видно из следующего выражения:

$$\omega l = 2\pi f l = r.$$

Для очень широких проводов данная частота снижается примерно до 11 ГГц. Для проводников с меньшей емкостью и сопротивлением (например, более толстых проводов, расположенных на верхних соединительных слоях), эта частота может снижаться до 500 МГц, особенно если используются лучшие соединительные материалы (например, медь). Таким образом, видим, что в интегральных схемах индуктивность проводников следует учитывать только при разработке самых современных цепей.

4.4. Модели электрических проводников

В предыдущих разделах мы ввели электрические свойства соединительных проводов (емкость, сопротивление и индуктивность) и представили несколько простых соотношений и технологий, используемых для расчета требуемых значений по геометрии и топологии межсоединений. Названные паразитные элементы влияют на электрическое поведение схемы и вносят вклад в ее задержку, рассеяние мощности и надежность. Для изучения этих эффектов

требуется введение электрических моделей, рассчитывающих и аппроксимирующих реальное поведение проводника как функцию его параметров. Такие модели могут быть как очень простыми, так и очень сложными в зависимости от изучаемых эффектов и требуемой точности. В данном разделе мы разработаем модели для ручного анализа, а затем рассмотрим, как анализировать соединительные проводники с помощью пакета SPICE.

4.4.1. Идеальный проводник

На схемах проводники представляются простыми линиями, с которыми не соотнесены никакие параметры или паразитные нагрузки. Данные проводники не влияют на электрическое поведение схемы. Изменение напряжения на одном конце провода немедленно распространяется на другие концы, даже если они достаточно разнесены в пространстве. Следовательно, можно предположить, что в любой момент времени на всех участках проводника присутствует одно и то же напряжение, а значит, весь провод является эквипотенциальной областью. Хотя такая модель идеального проводника является упрощенной, она находит свое применение, особенно на ранних стадия процесса разработки, когда дизайнер желает сконцентрировать внимание на свойствах и поведении соединяемых транзисторов. Кроме того, при рассмотрении таких небольших компонентов схемы, как логические вентили, проводники обычно настолько малы, что их паразитной нагрузкой можно пренебречь. Их учет может необоснованно затруднить анализ. Тем не менее очень часто паразитная нагрузка проводов играет важную роль, поэтому необходимы и более сложные модели.

4.4.2. Модель с сосредоточенными параметрами

Паразитная нагрузка проводника распределена вдоль его длины и не является сосредоточенной в одном месте. Но при наличии одного доминирующего паразитного компонента, когда взаимодействие компонентов незначительно или исследуется только один аспект поведения схемы, часто бывает полезно сосредоточить распределенную нагрузку в одном элементе схемы. Преимуществом такого подхода является возможность описания эффекта паразитной нагрузки с помощью обычного дифференциального уравнения. Как будет показано ниже, описание распределенного элемента требует системы уравнений в частных производных.

Пока резистивный компонент провода мал, а частоты коммутации — низкие или средние, целесообразно рассматривать только емкостный компонент провода и сосредоточивать распределенную емкость в одном конденсаторе, как показано на рис. 4.11. Обратите внимание на то, что в данной модели провод по-прежнему представляет эквипотенциальную область и сам провод не вводит никакой задержки. Единственное влияние на быстродействие оказывает эффект зарядки конденсатора. Приведенная модель сосредоточенной емкости является простой, но эффективной, и ее часто выбирают при анализе соединительных проводников в цифровых интегральных схемах.

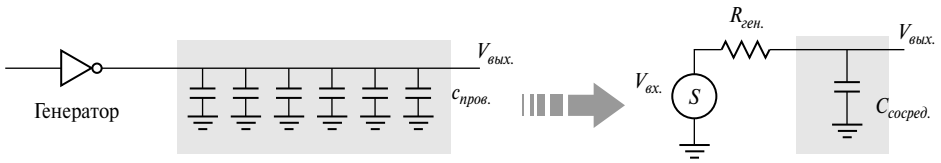


Рис. 4.11. Модели проводника: с распределенной и сосредоточенной нагрузкой. $C_{\text{сосред.}} = L \times c_{\text{пров.}}$, где L — длина проводника, а $c_{\text{пров.}}$ — емкость на единицу длины. Генератор моделируется источником напряжения и сопротивлением источника $R_{\text{ген.}}$

Пример 4.5. Модель провода с сосредоточенной емкостью

Рассмотрим схему, представленную на рис. 4.11, и будем считать, что генератор с сопротивлением 10 кОм используется для питания 10-сантиметрового провода Al1 толщиной 1 мкм. В примере 4.1 мы выяснили, что общая сосредоточенная емкость данного провода равна 11 пФ.

Работа данной простой RC-цепи описывается следующим дифференциальным уравнением (подобным выражению, выведенному в примере 4.1):

$$C_{lumped}\frac{dV_{\text{вых.}}}{dt} + \frac{V_{\text{вых.}} - V_{\text{вх.}}}{R_{driver}} = 0.$$

При подаче на вход скачкообразного сигнала (когда $V_{\text{вх.}}$ меняется с 0 до V) переходная функция, как известно, является экспоненциальной и описывается следующим выражением:

$$V_{\text{вых.}}(t) = (1 - \exp^{-t/\tau})V,$$

где $\tau = R_{\text{ген.}} \times C_{\text{сосред.}}$ — постоянная времени рассматриваемой сети.

Время, требуемое для достижения "точки 50%", вычисляется как $t = \ln(2)\tau = 0,69\tau$. Подобным образом, для перехода от 10 до 90% требуется время $t = \ln(9)\tau = 2,2\tau$. Подставляя данные этой задачи, получаем:

$$t_{50\%} = 0,69 \times 10 \text{ кОм} \times 11 \text{ пФ} = 76 \text{ нс};$$
$$t_{90\%} = 2,2 \times 10 \text{ кОм} \times 11 \text{ пФ} = 242 \text{ нс}.$$

Подобные значения неприемлемы даже для самых медленных цифровых схем. Чтобы избавиться от данного узкого места, применяются специальные технологии, например уменьшение сопротивления генератора, которое будет рассмотрено в главах 5 и 9.

Хотя сосредоточенная модель конденсатора очень популярна, иногда полезно представить сосредоточенные модели провода относительно сопротивления либо индуктивности. Это часто требуется при изучении распределительной питающей электрической сети. Как источники паразитного шума можно рассматривать и сопротивление, и индуктивность проводов шины питания, которые вводят перепады напряжения в шине питания.

4.4.3. Сосредоточенная RC-модель

Встроенные в кристалл металлические проводники, длина которых превышает несколько миллиметров, имеют существенное сопротивление. В связи с этим предположение об эквипотенциальной области, использованное в модели с сосредоточенной емкостью, уже не является справедливым, и нужно принимать резистивно-емкостную (RC) модель.

Существует несколько подходов к сосредоточенной RC-модели. В первом из них все сопротивление проводника сосредоточивается на одном сегменте и обозначается R, вся емкость C присваивается одному конденсатору. Такая простая модель, именуемая *сосредоточенной RC-моделью*, является довольно пессимистичной и неточной для длинных соединительных проводов, которые гораздо точнее представляются *распределенной RC-моделью*. Кроме того, из-за указанных ниже причин перед анализом распределенной модели полезно проанализировать и смоделировать сосредоточенные RC-цепи.

- Распределенная RC-модель достаточно сложная, для нее не существует аналитического решения. Поведение распределенной RC-линии можно достаточно точно смоделировать с помощью простой RC-цепи.
- При изучении переходных процессов в сложной транзисторной схеме принято сокращать схему до RC-цепи. Средства эффективного анализа подобных сетей и предсказания их характеристик в первом приближении должен иметь любой разработчик.

В примере 4.5 мы анализировали простую сеть “сопротивление–конденсатор”. Поведение такой сети полностью описывает одно дифференциальное уравнение, и его переходная характеристика моделируется экспоненциальной зависимостью с одной постоянной времени (или полюсом сети). К сожалению, расчет точных сигналов для сети с большим числом конденсаторов и резисторов очень быстро становится безнадежно сложным: описание ее поведения требует набора обычных дифференциальных уравнений, и теперь сеть может описываться несколькими постоянными времени (или полюсами и нулями). К счастью, нас спасают методы моделирования с помощью SPICE, например методы расчета задержки по *формуле Элмора* [Elmore48].

Рассмотрим резисторно-емкостную сеть, представленную на рис. 4.12. Подобная схема называется *RC-деревом* и обладает перечисленными ниже свойствами.

- Единственный узел входа (отмечен буквой s на рис. 4.12).
- Все конденсаторы включены между узлом и землей.
- Цепь не содержит ни одного резистивного контура (из-за этого она и является деревом).

Интересным следствием подобной топологии является существование единственного резистивного маршрута между узлом-источником s и любым узлом i сети. Общее сопротивление вдоль этого маршрута называется *со-*

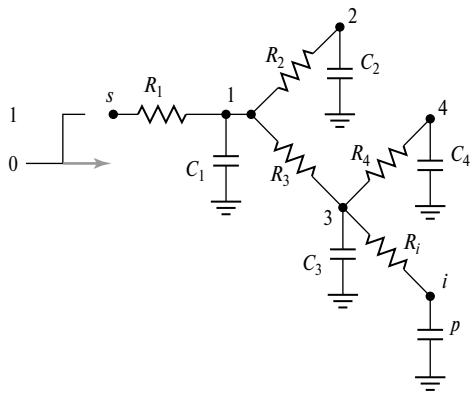


Рис. 4.12. RC-цепь с древовидной структурой

противление тракта R_{ii}. Например, сопротивление тракта между исходным узлом s и узлом 4 на рис. 4.12 равно:

$$R_{44} = R_1 + R_3 + R_4.$$

Определение сопротивления тракта можно расширить на *сопротивление тракта совместного использования* R_{ik}, которое представляет сопротивление, создаваемое совместно на тракте от корневого узла s до узлов k и i:

$$R_{ik} = \sum R_j \Rightarrow (R_j \in [path(s \rightarrow i) \cap path(s \rightarrow k)]). \tag{4.12}$$

Для цепи, представленной на рис. 4.12, $R_{i4} = R_1 + R_3$, а $R_{i2} = R_1$.

Предположим теперь, что все N узлов сети изначально разряжены на землю и в момент времени $t = 0$ на узел s подается ступенчатое напряжение. В таком случае Элморову задержку в узле i можно представить следующим выражением:

$$\tau_{Di} = \sum_{k=1}^{N} C_k R_{ik}. \tag{4.13}$$

Элморова задержка эквивалентна постоянной времени первого порядка данной сети (или первому моменту импульсной характеристики). Разработчик должен помнить, что эта постоянная времени представляет простую аппроксимацию реальной задержки между узлом-источником и узлом i. Тем не менее во многих случаях данная аппроксимация достаточно разумна и приемлема. Она представляет собой мощный механизм, с помощью которого разработчик может быстро оценить задержку сложной цепи.

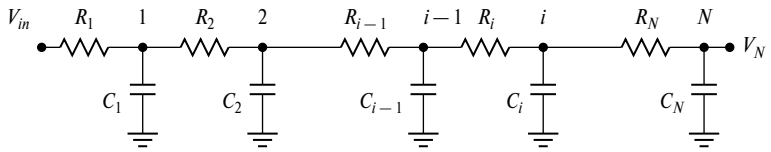


Рис. 4.13. RC-цепь

Пример 4.6. RC-задержка сети с древовидной структурой

Используя уравнение (4.13), можно рассчитать Элморову задержку для узла i сети, представленной на рис. 4.12: $\tau_{Di} = R_1C_1 + R_1C_2 + (R_1 + R_3)C_3 + (R_1 + R_3)C_4 + (R_1 + R_3 + R_i)C_i$.

Как частный случай RC-цепи с древовидной структурой, рассмотрим простую неразветвленную RC-цепочку, показанную на рис. 4.13. Мы проанализируем данную сеть потому, что такая структура очень часто встречается в цифровых схемах, а также потому, что она представляет собой аппроксимацию резистивно-емкостного проводника. Элморова задержка данной сети находится с помощью уравнения (4.13):

$$\tau_{Dn} = \sum_{i=1}^{N} C_i \sum_{j=1}^{i} R_j = \sum_{i=1}^{N} C_i R_{ii}. \tag{4.14}$$

Другими словами, сопротивление совместно используемого тракта заменяется сопротивлением тракта. Рассмотрим в качестве примера узел 2 RC-цепи, представленной на рис. 4.13. Его постоянная времени состоит из двух компонентов — вкладов узлов 1 и 2. Компонент, порождаемый узлом 1, составляет C_1R_1, где R_1 — общее сопротивление между узлом и источником, а вклад узла 2 равен $C_2(R_1 + R_2)$. Эквивалентная постоянная времени в узле 2 равна $C_1R_1 + C_2(R_1 + R_2)$. Постоянная времени τ_i узла i рассчитывается аналогично:

$$\tau_{Di} = C_1R_1 + C_2(R_1 + R_2) + \ldots + C_i(R_1 + R_2 + \ldots + R_i).$$

Пример 4.7. Постоянная времени резистивно-емкостного провода

Модель, представленную на рис. 4.13, можно использовать в качестве аппроксимации резистивно-емкостного проводника. Итак, провод общей длины L разбит на N идентичных сегмента, каждый из которых имеет длину L/N. Следовательно, сопротивление и емкость каждого сегменты равны rL/N и cL/N соответственно. Используя формулу Элмора, можно вычислить главную постоянную времени для данного проводника:

$$\tau_{DN} = \left(\frac{L}{N}\right)^2 (rc + 2rc + \ldots + Nrc) = (rcL^2)\frac{N(N+1)}{2N^2} = RC\frac{N+1}{2N}, \quad (4.15)$$

где R $(= rL)$ и C $(= cL)$ — общие сосредоточенные сопротивление и емкость провода. Для очень больших значений N данная модель асимптотически приближается к распределенной RC-цепи. В подобном случае уравнение (4.15) упрощается до следующего вида:

$$\tau_{DN} = \frac{RC}{2} = \frac{rcL^2}{2}. \quad (4.16)$$

Из уравнения (4.16) можно сделать два важных вывода.

- Задержка провода квадратично зависит от его длины! Это означает, что удвоение длины провода в четыре раза увеличивает задержку распространения сигнала.
- Задержка распределенной RC-линии равна половине задержки, предсказываемой сосредоточенной RC-моделью. В последней общее сопротивление и емкость объединены в одном элементе, а постоянная времени равна RC (тот же результат можно получить, положив $N = 1$ в уравнении (4.15)). Это подтверждает сделанное ранее наблюдение относительно того, что сосредоточенная модель представляет пессимистическую оценку задержки резистивного проводника.

Предупреждение.

Помните, что RC-цепь характеризуется целым набором постоянных времени. Формула Элмора определяет значение только доминирующей постоянной, а следовательно, представляет собой всего лишь аппроксимацию первого порядка.

Формула Элмора для задержки доказала свою полезность. Помимо того, что она позволяет анализировать проводники, формула может использоваться для аппроксимации задержки распространения сложных транзисторных сетей. В модели транзистора, представленного в форме ключа, транзисторы заменяются эквивалентными сопротивлениями в открытом состоянии. В таком случае расчет задержки распространения сводится к анализу получающейся RC-цепи. Более точные нижняя и верхняя границы сигналов напряжения в RC-дереве рассчитываются в [Rubinstein83]. На основе данных границ были созданы многие автоматизированные средства анализа на коммутируемом и функциональном уровнях [Horowitz83]. Интересно также отметить [Lin84], что экспоненциальная кривая напряжения с постоянной времени, рав-

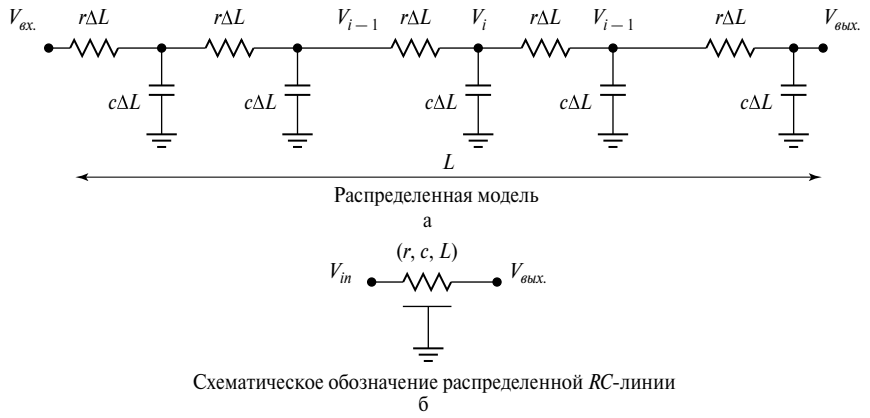


Рис. 4.14. Распределенная RC-цепь и ее схематическое обозначение

ной элморовой задержке, всегда располагается между данными минимальной и максимальной границами, что подтверждает справедливость аппроксимации, предложенной Элмором.

4.4.4. Распределенная RC-цепь

Ранее мы показывали, что сосредоточенная RC-модель является пессимистической для резистивно-емкостного провода и более подходящей является распределенная RC-модель (рис. 4.14, *а*). Как и ранее, L представляет общую длину провода, а r и c — сопротивление и емкость на единицу длины. Схематическое представление распределенной RC-цепи представлено на рис. 4.14, *б*.

Напряжение на узле i данной сети можно определить, решив следующую систему уравнений в частных производных:

$$c\triangle L\frac{\partial V_i}{\partial t} = \frac{(V_{i+1} - V_i) + (V_{i-1} - V_i)}{r\triangle L}. \tag{4.17}$$

Правильное поведение распределенной RC-цепи получается при асимптотическом устремлении ΔL к нулю. При $\Delta L \to 0$ уравнение (4.17) переходит в известное уравнение диффузии:

$$rc\frac{\partial V}{\partial t} = \frac{\partial^2 V}{\partial x^2}. \tag{4.18}$$

В этом уравнении V — напряжение в конкретной точке провода; x — расстояние между данной точкой и источником сигнала. Аналитического решения данного уравнения не существует, однако можно вывести приближенные формулы, подобные представленной в уравнении (4.19) [Bakoglu90]:

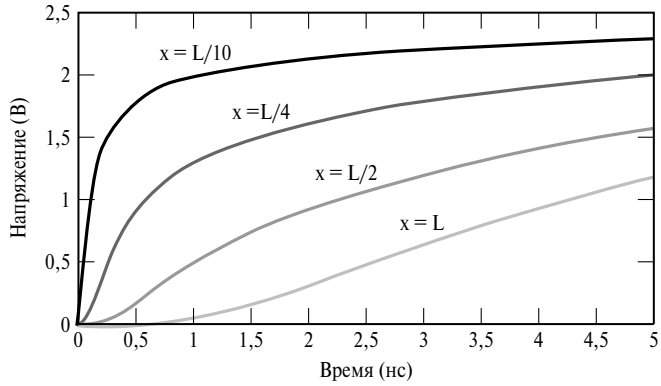


Рис. 4.15. Смоделированная импульсная характеристика резистивно-емкостной цепи как функция времени и положения

$$V_{\text{вых.}}t = 2erfc\left(\sqrt{\frac{RC}{4t}}\right), \tau \ll RC \tag{4.19}$$
$$= 1,0 - 1,366e^{-2,5359\frac{t}{RC}} + 0,366\exp^{-9,4641\frac{t}{RC}}, \tau \gg RC.$$

Приведенные уравнения сложно использовать для анализа обычных схем. Тем не менее известно, что распределенную RC-цепь можно аппроксимировать сосредоточенной RC-лесенкой, которую легко использовать в автоматизированном анализе.

На рис. 4.15 показана реакция провода на ступенчатый ввод, когда как функция времени в различимых точках проводника изображена форма сигнала напряжения. Обратите внимание на то, что ступенчатый сигнал "рассеивается" от начала до конца проводника, сигнал быстро ослабевает, что приводит к существенной задержке для длинных проводников. Питания подобных RC-цепей и минимизация задержки и ухудшения сигнала является одной из самых мудреных проблем в современной разработке цифровых интегральных схем. Именно поэтому в данной книге ему уделяется значительное внимание.

Некоторые важные контрольные точки импульсной характеристики сосредоточенной и распределенной RC-модели провода приведены в табл. 4.7. Например, задержка распространения (определяемая при 50% конечного значения) сосредоточенной сети равна $0,69RC$ (что неудивительно). С другой стороны, распределенная сеть имеет задержку всего лишь около $0,38RC$, где R и C — общее сопротивление и емкость провода. Отметим, что данный результат, полученный при моделировании, согласуется с уравнением (4.16).

Пример 4.8. ***RC*-задержка алюминиевого провода**

Вернемся к нашему алюминиевому проводу Al1 длиной 10 см, шириной 1 мкм, рассмотренному в примере 4.1. Тогда мы нашли следующее значения для r и c:

Таблица 4.7. Интересные точки импульсной характеристики сосредоточенной и распределенной RC-цепи

Диапазон напряжения	Сосредоточенная RC-цепь	Распределенная RC-цепь
$0 \to 50\%(t_p)$	$0,69RC$	$0,38RC$
$0 \to 63\%(\tau)$	RC	$0,5RC$
$10\% \to 90\%(t_r)$	$2,2RC$	$0,9RC$
$0\% \to 90\%$	$2,3RC$	$1,0RC$

$$c = 110 \text{ аФ/мкм}; \quad r = 0,075 \text{ Ом/мкм}.$$

Используя данные из табл. 4.7, находим задержку распространения сигнала по проводнику:

$$t_p = 0,38RC = 0,38 \times (0,075 \text{ Ом/мкм}) \times (110 \text{ аФ/мкм}) \times (10^5 \text{ мкм})^2 = 31,4 \text{ нс}.$$

Кроме того, можно рассчитать задержку распространения идентичного провода, сделанного из поликремния и Al5 (пятый металлический слой). Необходимые емкости берутся из табл. 4.2, а сопротивления считаются равными соответственно 150 и 0,0375 Ом/мкм для поликремния и Al5:

Поликремний: $t_p = 0,38 \times (150 \text{ Ом/мкм}) \times (88 + 2 \times 54 \text{ аФ/мкм}) \times \times (10^5 \text{ мкм})^2 = 112 \text{ мкс!}$

Al5: $t_p = 0,38 \times (0,0375 \text{ Ом/мкм}) \times (5,2 + 2 \times 12 \text{ аФ/мкм}) \times \times (10^5 \text{ мкм})^2 = 4,2 \text{ нс}.$

Очевидно, выбор соединительного материла и слоя очень сильно влияет на задержку провода.

При анализе соединительной сети разработчик должен решить, следует ли учитывать эффекты RC-задержек или же можно просто рассмотреть упрощенную модель с сосредоточенной емкостью. В данном случае полезными оказываются нижеприведенные эмпирические правила.

Эмпирические правила разработки

- RC-задержки следует рассматривать только в том случае, когда t_{pRC} сравнимо или больше t_{pgate} питающего элемента.

Таким образом, мы можем определить критическую длину:

$$L_{crit} = \sqrt{\frac{t_{pgate}}{0,38rc}}. \tag{4.20}$$

RC-задержки начинают играть решающую роль для соединительных проводов, длина которых больше критической (L_{crit}). Реальное значение L_{crit} зависит от размера питающего элемента и выбранного соединительного материала.

- *RC*-задержки следует рассматривать только в том случае, когда время нарастания (спада) на входе цепи меньше RC (времени нарастания (спада) линии).

Другими словами, задержки следует учитывать, когда

$$t_{rise} < RC, \tag{4.21}$$

где R и C — соответственно общее сопротивление и емкость провода. Если данное условие не выполняется, изменение сигнала происходит за время, большее задержки распространения, поэтому для описания цепи достаточно модели с сосредоточенной емкостью.

Пример 4.9. RC или сосредоточенная C?

Представленные правила можно проиллюстрировать с помощью простой схемы, показанной на рис. 4.16. Здесь предполагается, что схему питания можно смоделировать как источник напряжения с конечным сопротивлением R_s. Применяя формулу Элмора[2], получаем, что общую задержку распространения сети можно аппроксимировать следующим выражением:

$$\tau_D = R_s C_w + \frac{R_w C_\omega}{2} = R_s C_w + 0,5 r_w c_w L^2.$$

Кроме того,

$$t_p = 0,69 R_s C_w + 0,38 R_s C_w,$$

где $R_w = rL$ и $C_w = cL$. Задержка, вводимая сопротивлением провода, начинает доминировать при $(R_w C_w)/2 \geq R_s C_w$ или $L \geq 2R_s/r$. Далее предполагаем, что генератор имеет сопротивление 1 кОм и нагружает провод Al1 толщиной 1 мкм ($r = 0,075$ Ом/мкм). При таких данных получаем, что критическая длина равна 2,67 см.

---

[2] *Подсказка*: на рис. 4.13 замените провод сосредоточенной *RC*-цепью и примените уравнение Элмора к получающейся сети.

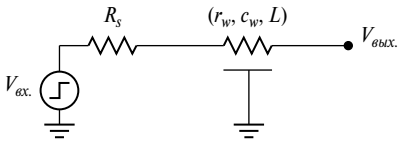


Рис. 4.16. RC-цепь длины L, питаемая источником с сопротивлением R_s

4.4.5. Линия передачи

Когда скорость коммутации схемы становится достаточно большой, а качество соединительного материала — настолько высоким, что сопротивление провода входит в допустимые границы, основным параметром, определяющим задержку, становится индуктивность, и необходимо рассматривать эффекты линии передачи. Точнее, в описанном случае время нарастания и время спада сигнала сравнимы с временем прохода сигнала по линии, которое определяется скоростью света. С появлением медных соединений и высоких скоростей коммутации (ставших возможными с развитием субмикронных технологий) вопрос эффектов линии передачи привлек еще большее внимание, и его необходимо изучать при разработке наиболее быстрых КМОП-схем.

В данном разделе мы сначала проанализируем модель линии передачи. Далее мы применим эту модель к современным полупроводниковым технологиям и определим, когда данные эффекты действительно следует учитывать в процессе разработки.

Модель линии передачи

Подобно сопротивлению и емкости соединительной линии, индуктивность проводника также распределена по всей его длине. Распределенная RLC-модель провода, именуемая *моделью линии передачи*, наиболее точно аппроксимирует реальную ситуацию. Основное свойство линии передачи заключается в том, что сигнал распространяется по связующей среде как *волна*. В распределенной RC-модели сигнал *диффундировал* от источника до цели, при этом его поведение описывалось уравнением диффузии (4.18). В волновой модели сигнал распространяется посредством попеременной передачи энергии между электрическим и магнитным полем, или между емкостным и индуктивным режимами.

Рассмотрим точку x вдоль линии передачи, представленной на рис. 4.17, в момент времени t. Согласно данным, можно записать следующую систему уравнений:

$$\begin{aligned} \frac{\partial \nu}{\partial x} &= -ri - l\frac{\partial i}{\partial t}; \\ \frac{\partial i}{\partial x} &= -g\nu - c\frac{\partial \nu}{\partial t}. \end{aligned} \tag{4.22}$$

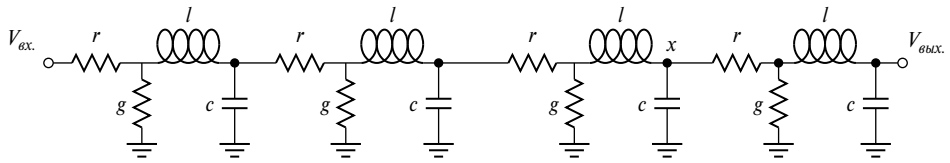


Рис. 4.17. Линия передачи с потерями

Если предположить, что проводимость утечки g равна 0 (что справедливо для большинства диэлектриков), и убрать ток i, получим *волновое уравнение*:

$$\frac{\partial^2 \nu}{\partial x^2} = rc\frac{\partial \nu}{\partial t} + lc\frac{\partial^2 \nu}{\partial t^2}, \tag{4.23}$$

где r, c и l — соответственно сопротивление, емкость и индуктивность на единицу длины.

Чтобы лучше понять поведение линии передачи, предположим для начала, что сопротивление линии мало. В таком случае подходит упрощенная емкостно-индуктивная модель, именуемая *линией передач без потерь*. Названная модель применима к проводникам на уровне печатной платы. Из-за высокой проводимости использованных медных соединительных элементов сопротивление линии передачи можно игнорировать. С другой стороны, сопротивление играет важную роль в интегральных схемах, поэтому следует рассмотреть более сложную модель, называемую *линией передачи с потерями*. (Модель с потерями кратко рассматривается в конце главы.)

Линия передач без потерь

Для линии передач без потерь уравнение (4.23) упрощается до *уравнения идеальной волны:*

$$\frac{\partial^2 \nu}{\partial x^2} = lc\frac{\partial^2 \nu}{\partial t^2} + \frac{1}{\nu^2}\frac{\partial^2 \nu}{\partial t^2}. \tag{4.24}$$

Единичный сигнал, поданный на вход линии передачи без потерь, распространяется по ней со скоростью υ, определяемой уравнением (4.11), которое повторно приводится ниже.

$$\upsilon = \frac{1}{\sqrt{lc}} = \frac{1}{\sqrt{\varepsilon\mu}} = \frac{c_0}{\sqrt{\varepsilon_r \mu_r}}. \tag{4.25}$$

Несмотря на то что значения l и c зависят от геометрической формы провода, их произведение равно константе и является функцией только окружающей среды. Задержка распространения на единицу длины провода (t_p) линии передачи является величиной, обратной к скорости:

$$t_p = \sqrt{lc}. \tag{4.26}$$

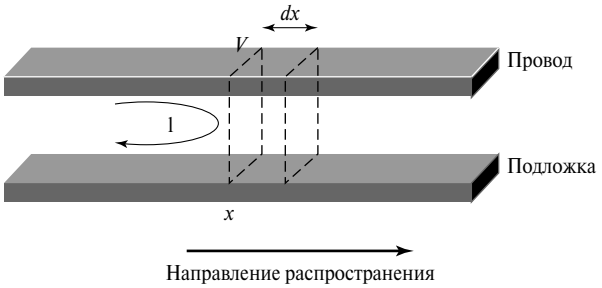


Рис. 4.18. Распространение скачка напряжения вдоль линии передачи без потерь

Проанализируем теперь, как волна распространяется вдоль линии передачи без потерь. Предположим, что на вход подан скачок напряжения V, и сигнал распространился до точки x, как показано на рис. 4.18. Справа от x все токи равны 0, тогда как напряжение в линии с левой стороны равно V. Чтобы волна распространилась на дополнительное расстояние dx, должна зарядиться дополнительная емкость cdx. Таким образом, сила тока должна быть равна:

$$i = \frac{dq}{dt} = c\frac{dx}{dt}\nu = cv\nu = \sqrt{\frac{c}{l}}\nu, \tag{4.27}$$

поскольку скорость распространения сигнала dx/dt равна v. Это означает, что сигнал воспринимает оставшуюся часть линии как действительный импеданс Z_0, где

$$Z_0 = \frac{V}{I} = \sqrt{\frac{l}{c}} = \frac{\sqrt{\varepsilon\mu}}{c} = \frac{1}{cv}. \tag{4.28}$$

Данный импеданс, называемый *характеристическим импедансом линии*, является функцией диэлектрической среды и геометрии проводника изолятора (уравнение (4.28)); оно не зависит от длины проводника и частоты. Такая линия произвольной длины с постоянным действительным импедансом невероятно привлекательна, поскольку она существенно упрощает разработку схемы задающего генератора. Типичные значения характеристических импедансов проводников в полупроводниковых схемах принадлежат диапазону от 10 до 200 Ом.

Пример 4.10. Скорость распространения сигналов

Из информации, представленной в табл. 4.6, видно, что для прохождения сигнала от источника до цели по 20-сантиметровому проводу, проложенному по эпоксидной печатной плате, требуется 1,5 нс. Если эффекты линии передачи влияли на кремниевые интегральные схемы, то 10-сантиметровый провод сигнал проходил бы за 0,67 нс.

Предупреждение

Характеристический импеданс провода является функцией общей топологии соединений. Электромагнитные поля являются сложной соединительной структурой, как правило, — несимметричной, и на них очень сильно влияют такие факторы, как тракт возврата тока. Общее решение последней проблемы до сих пор не найдено, также не существует аналитических решений соответствующих уравнений. Следовательно, точный расчет индуктивности и характеристического импеданса по-прежнему активно изучается. Для некоторых упрощенных структур были выведены различные аппроксимации. Например, характеристические импедансы трехпластинчатой линии передачи (провод, внедренный между двух пластин заземления) аппроксимируются следующими уравнениями:

$$Z_0(triplate) \approx 94\sqrt{\frac{\mu_r}{\varepsilon_r}} \ln\left(\frac{2t+W}{H+W}\right) \text{ Ом} \tag{4.29}$$

и

$$Z_0(microstrip) \approx 60\sqrt{\frac{\mu_r}{0,475\varepsilon_r + 0,67}} \ln\left(\frac{4t}{0,536W + 0,67H}\right) \text{ Ом.} \tag{4.30}$$

Оконечная нагрузка

На поведение линии передачи очень сильное влияние оказывает оконечная нагрузки цепи. Оконечная нагрузка определяет, какая часть волны будет отражена по достижении конца провода. Данный параметр представляется *коэффициентом отражения* ρ, который определяет связь между напряжениями и силами тока падающей и отраженной волн. Итак,

$$\rho = \frac{V_{refl}}{V_{inc}} = \frac{I_{refl}}{I_{inc}} = \frac{R - Z_0}{R + Z_0}, \tag{4.31}$$

где R — значение оконечного сопротивления. Общие напряжения и силы тока на выходе являются суммой падающей и отраженной волн:

$$\begin{aligned} V &= V_{inc}(1+\rho); \\ I &= I_{inc}(1-\rho). \end{aligned} \tag{4.32}$$

Как показано на рис. 4.19, можно рассмотреть три интересных частных случая. В случае *а* оконечное сопротивление равно характеристическому импедансу линии. Оконечная нагрузка выглядит как бесконечное продолжение линии, и сигнал не отражается. В частности, это видно из значения ρ, которое равно 0. В случае *б* оконечной нагрузкой является разомкнутый контур

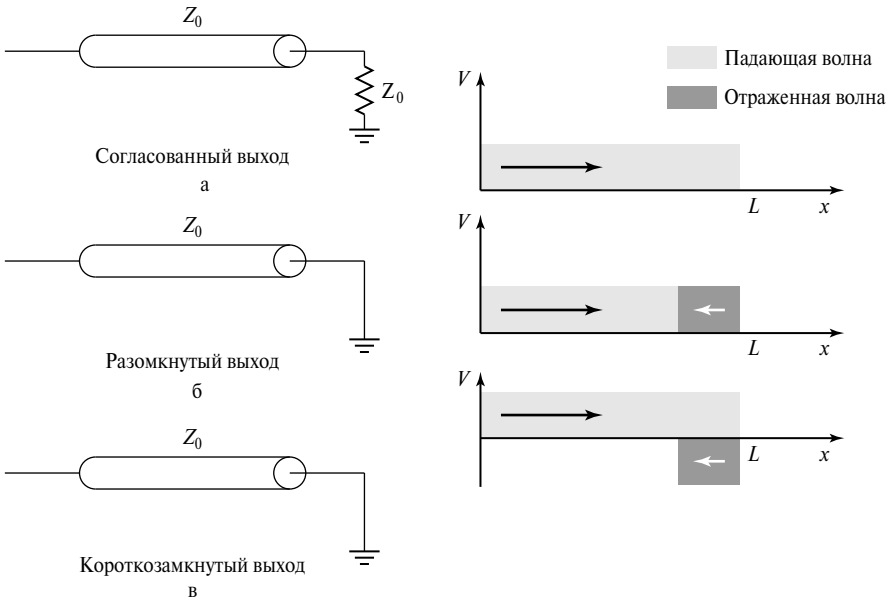


Рис. 4.19. Поведение различных оконечных нагрузок цепи

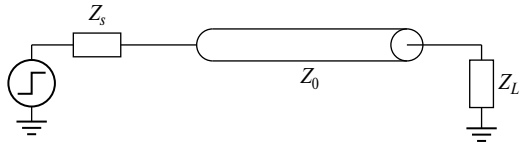


Рис. 4.20. Линия передачи с оконечными импедансами

($R = \infty$) и $\rho = 1$. Общий сигнал напряжения после отражения равен удвоенному падающему сигналу, что следует из уравнения (4.32). Наконец, в случае *в*, когда оконечной нагрузкой является цепь короткого замыкания, $R = 0$ и $\rho = -1$. Суммарное напряжение на конце линии после отражения равно нулю.

Далее можно исследовать переходной режим полной линии передачи. На него влияют характеристический импеданс цепи, последовательно включенный импеданс источника Z_S и нагрузочный импеданс Z_L на конце линии, как показано на рис. 4.20.

Прежде всего рассмотрим ситуацию, когда проводник открыт на конце, или при $Z_L = \infty$ и $\rho_L = 1$. Падающая волна полностью отражается без опрокидывания фазы. Предполагая, что импеданс источника является резистивным, возможны три сценария, схематически изображенных на рис. 4.21: $R_S = 5Z_0, R_S = Z_0$ и $R_S = 1/5Z_0$.

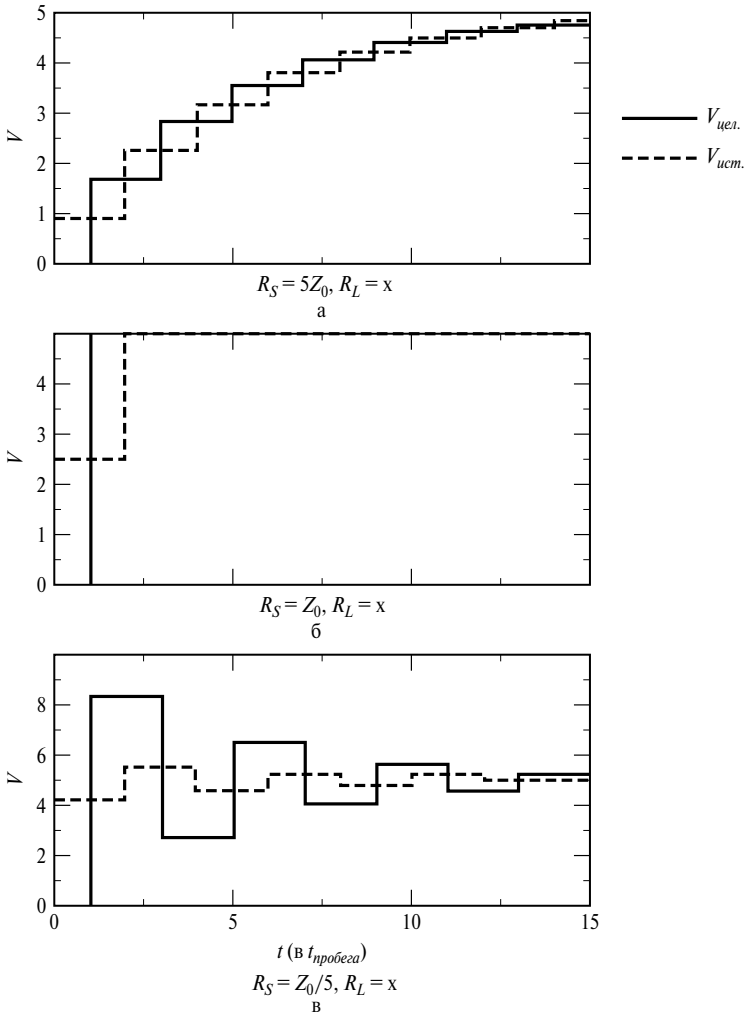


Рис. 4.21. Переходная характеристика линии передачи

1. Большое сопротивление источника — $R_S = 5Z_0$ (см. рис. 4.21, *а*)

Линия передачи питается лишь небольшой долей поступающего сигнала $V_{\text{вх.}}$, равного 5 В. Данная доля определяется резистивным делителем, зависящем от сопротивления источника и характеристического импеданса Z_0:

$$V_{source} = (Z_0/(Z_0 + R_s))V_{\text{вх.}} = 1/6 \times 5 \text{ D} = 0,83 \text{ В}. \qquad (4.33)$$

Данный сигнал достигает конца линии через L/v, где L — длина провода, и полностью отражается, что эффективно удваивает амплитуду волны ($V_{dest} = 1,67$ В). Время, требуемое волне на распространение от одного конца

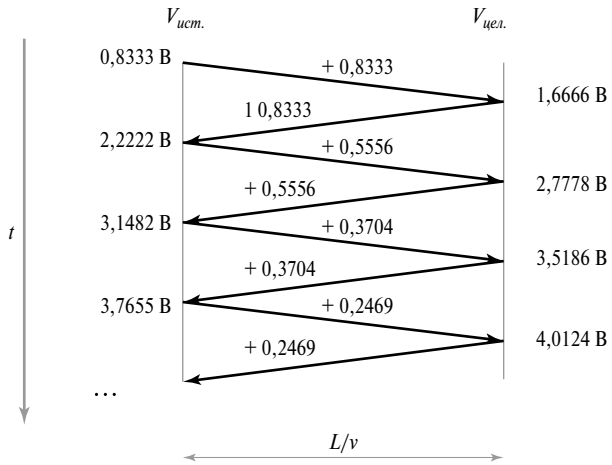


Рис. 4.22. Ступенчатая диаграмма для $R_S = 5Z_0$ и $R_L = \infty$. $V_{step} = 5$ В (как на рис. 4.21, *а*)

провода до другого, называется *временем пробега* (или *пролета*) и составляет $t_{flight} = L/v$. Когда отраженная волна достигает исходного узла, происходит практически то же самое. Падающая волна отражается с амплитудой, определяемой коэффициентом отражения источника, который в данном случае равен $2/3$:

$$\rho_S = \frac{5Z_0 - Z_0}{5Z_0 + Z_0} = \frac{2}{3}. \tag{4.34}$$

Амплитуда напряжения в двух краевых узлах постепенно достигает конечного значения $V_{\text{вх.}}$. Тем не менее общее время нарастания во много раз больше L/v.

При наличии многократного отражения (как в данном случае) отслеживать каждую волну в линии и общие уровни напряжения слишком утомительно. Поэтому часто используется графическая конструкция, именуемая *ступенчатой диаграммой* (рис. 4.22). Диаграмма содержит значения напряжений на обоих концах, а также значения амплитуд падающей и отраженной волн. Напряжении в линии на конце провода равно сумме предыдущего напряжения, падающей и отраженной волн.

2. Небольшое сопротивление источника — $R_S = Z_0/5$ (см. рис. 4.21, *в*).

В линию поступает большая доля входного сигнала. На конце амплитуда волны удваивается, что приводит к существенной перегрузке. В исходном узле фаза сигнала обращается ($\rho_S = -2/3$). Отражаясь, сигнал путешествует между двумя концами линии, постепенно затухая.

3. Согласованное сопротивление источника — $R_S = Z_0$ (см. рис. 4.21, *б*).

Половина входного сигнала вводится источником в линию. Отражение на конце линии удваивает сигнал, который немедленно достигает конечного значения. Очевидно, что подобная схема является наиболее эффективной.

Обратите внимание на то, что в предыдущем анализе мы рассматривали идеальные условия, поскольку предполагали, что входной сигнал имеет нулевое время нарастания. Реальные сигналы значительно более гладкие, что видно по смоделированной характеристике на рис. 4.23 (для $R_S = Z_0/5$ и $t_r = t_{flight}$).

Задача 4.1. Характеристика линии передачи

Выведите ступенчатую диаграмму рассмотренной выше линии передачи при $R_S = Z_0/5$, $R_L = \infty$ и $V_{step} = 5$ В. Кроме того, попробуйте решить обратную задачу: предположите, что последовательное сопротивление источника равно нулю, и рассмотрите различные импедансы нагрузки.

Пример 4.11. Емкостная оконечная нагрузка

В цифровых МОП-схемах оконечные нагрузки, как правило, имеют емкостную природу. Можно рассмотреть, как это влияет на поведение линии передачи и когда следует учитывать емкость нагрузки.

Характеристический импеданс линии передачи определяет ток, который может заряжать нагрузку с емкостью C_L. С точки зрения нагрузки линия ведет себя как сопротивление со значением Z_0. Следовательно, переходная характеристика на конденсаторе описывается постоянной времени Z_0C_L. Сказанное иллюстрируется на рис. 4.24, где показана смоделированная переходная характеристика с линией передачи, которая завершается последовательно, имеет характеристический импеданс 50 Ом и нагружена емкостью 2 пФ. По характеристике видно, как выход нарастает до конечного значения, равного постоянной времени 100 пс (= 50 Ом × 2 пФ) после задержки, равной времени пробега линии.

Данная асимптотическая характеристика обладает несколькими любопытными артефактами. После времени $2t_{flight}$ происходит якобы необычное понижение напряжения на входе. Это можно объяснить следующим образом. После достижения конца линии падающая волна отражается. Отраженная волна также достигает своего конечного значения асимптотически. Поскольку V_{dest} изначально равно 0 (а не 5 В, как можно было бы предположить), отражается сигнал $-2,5$ В, а не $2,5$ В. Из-за этого сигнал в линии передачи временно падает до 0 В, что видно по смоделированной характеристике. По мере того как выходной узел постепенно доходит до своего конечного значения, описанный эффект сходит на нет.

Задержка распространения линии равна сумме времени пробега (50 пс) и времени, которое требуется для зарядки конденсатора ($0,69Z_0C_L = 69$ пс). Именно такой результат дает моделирование. В общем случае можно говорить, что емкостную нагрузку следует рассматривать только в том случае,

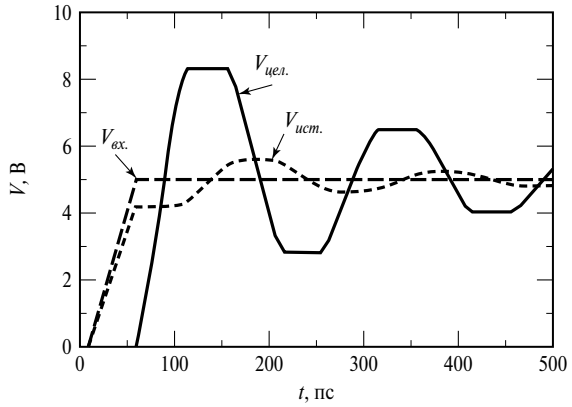


Рис. 4.23. Смоделированная переходная характеристика линии передачи без потерь для конечного времени нарастания входа ($R_S = Z_0/5$, $t_r = t_{flight}$)

если ее значение сравнимо с общей емкостью линии передачи (или больше его) [Bakoglu90].

Линия передачи с потерями

Если провода платы и модулей настолько толстые и широкие, что их можно рассматривать как линии передачи без потерь, аналогичное допущение для внутрикристальных соединений будет несправедливым, поскольку здесь сопротивление проводника является важным фактором. В подобном случае необходимо применять модель линии передачи с потерями. Мы не будем сильно углубляться в поведение линии с потерями, просто качественно обсудим влияние резистивных потерь на линию передачи.

На импульсной характеристике *RLC-линии* с потерями распространение волны объединяется с диффузным компонентом. Сказанное демонстрируется на рис. 4.25, где представлена характеристика *RLC*-линии передачи как функция расстояния от источника. Единичный импульс распространяется по линии как волна, однако амплитуда этой волны постепенно уменьшается (затухает):

$$\frac{V_{step}(x)}{V_{step}(0)} = \exp^{-\frac{r}{2Z_0}x}. \tag{4.35}$$

За прохождением волны следует диффузная релаксация до значения, установившегося в точке x. Чем дальше точка от источника, тем больше похожа характеристика на поведение распределенной *RC*-линии. Фактически резистивный эффект начинает доминировать, и цепь ведет себя как распределенная *RC*-линия, когда R ($= rL$, общее сопротивление линии) $\gg 2Z_0$. При $R = 5Z_0$ до конца линии доходит всего 8% исходного импульса. При таких параметрах линию точнее моделировать как распределенную *RC*-цепь.

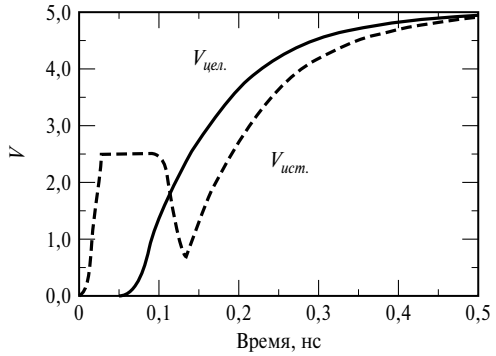


Рис. 4.24. Линия передачи, оканчивающаяся емкостью: $R_S = 50$ Ом, $R_L = \infty$, $C_L = 2$ пФ, $Z_0 = 50$ Ом, $t_{flight} = 50$ пс

Помните, что реальные провода на кристаллах, платах или подложках на самом деле гораздо сложнее, чем предсказывает данный простой анализ. Например, ответвления проводов, часто называемые *линейными выводами*, вызывают дополнительное отражение и могут влиять на форму и задержку сигнала. Поскольку анализ данных эффектов очень сложен, единственными приемлемым подходом является использование компьютерного анализа и технологий моделирования. Более подробное обсуждение данных эффектов можно найти в [Bakoglu90] и [Dally98].

Эмпирическое правило разработки

Вернемся к вопросу, когда ст*о*ит рассматривать эффекты линии передачи. Из сказанного выше можно сделать два важных вывода.

- Эффекты линии передачи следует рассматривать тогда, когда время нарастания или спада входного сигнала (t_r, t_f) меньше времени пробега линии передачи (t_{flight}).

Отсюда следует эмпирическое правило, определяющее, когда следует рассматривать эффекты линии передачи:

$$t_r(t_f) < 2,5 t_{flight} = 2,5 \frac{L}{\upsilon}. \tag{4.36}$$

Для внутренних межсоединений кристалла, имеющих максимальную длину 1 см, эффекты линии передачи необходимо рассматривать только при $t_r < 150$ пс. На уровне монтажной платы, когда провода могут иметь длину до 50 см, задержку линии передачи следует учитывать только при $t_r < 8$ нс. Данное условие легко удовлетворяется с помощью современных процессов и технологий упаковки. Отметим, что игнорирование индуктивного компонента задержки распространения может привести к общей оптимистичной оценке задержки.

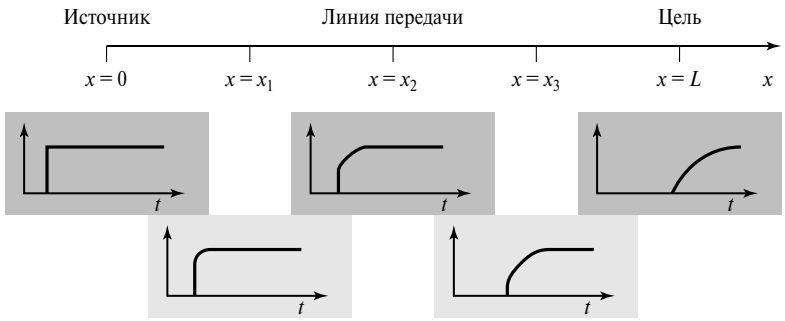


Рис. 4.25. Импульсная характеристика линии передачи с потерями

- Эффекты линии передачи следует рассматривать только тогда, когда общее сопротивление провода ограничено:

$$R < 5Z_o. \tag{4.37}$$

Если данное условие не выполняется, достаточно распределенной RC-модели.

Оба приведенных условия можно свести в следующее ограничение на длину провода:

$$\frac{t_r}{2.5}\frac{1}{\sqrt{lc}} < L < \frac{5}{r}\sqrt{\frac{l}{c}}. \tag{4.38}$$

- Линия передачи считывается линией без потерь, когда общее сопротивление значительно меньше характеристического импеданса или когда

$$R < \frac{Z_0}{2}. \tag{4.39}$$

Пример 4.12. Эффекты линии передачи

Рассмотрим еще раз наш провод Al1. Используя данные из примера 4.4 и уравнение (4.28), можно аппроксимировать значение Z_0 для различных длин проводов:

$$W = 0,1 \text{ мкм} : c = 92 \text{ аФ/мкм}; Z_0 = 74 \text{ Ом};$$
$$W = 1,0 \text{ мкм} : c = 110 \text{ аФ/мкм}; Z_0 = 60 \text{ Ом};$$
$$W = 10 \text{ мкм} : c = 380 \text{ аФ/мкм}; Z_0 = 17 \text{ Ом}.$$

Используя уравнение (4.37), мы можем вывести для провода шириной 1 мкм максимальную длину, для которой необходимо учитывать эффекты линии передачи:

$$L_{max} = \frac{5Z_0}{r} = \frac{5 \times 60\Omega}{0,075 \text{ Ом/мкм}} = 4000 \text{ мкм}.$$

Из уравнения (4.36) находим, что соответствующее максимальное время нарастания (или спада) входного сигнала равно:

$$t_{rmax} = 2.5 \times (4000 \text{ мкм})/(15 \text{ см/нс}) = 67 \text{ пс}.$$

Получить подобное значение с помощью современных технологий сложно, поэтому для описанных проводов лучше применять модель с сосредоточенной емкостью. Эффекты линии передачи более вероятны в более широких проводах. Для провода шириной 10 мкм получаем максимальную длину 11,3 мм, что соответствует максимальному времени нарастания 188 пс.

Предположим теперь, что медный провод используется на уровне 5 с характеристическим импедансом 200 Ом и сопротивлением 0,025 Ом/мкм. Получающаяся максимальная длина провода равна 40 мм. Если время нарастания меньше 670 пс, будут наблюдаться эффекты линии передачи.

Помните, однако, что значения Z_0, выделенные в данном примере, являются всего лишь аппроксимациями. При реальной разработке следует использовать более сложные выражения или эмпирические данные.

4.5. SPICE-модели проводов

В предыдущих разделах мы обсуждали различные паразитные нагрузки межсоединений, вводя для каждой из них простые модели. В современном мире разработчик может понять полное и точное влияние этих эффектов только с помощью подробного моделирования. В данном разделе мы введем модели, предлагаемые SPICE для описания емкостных, резистивных и индуктивных паразитных нагрузок.

4.5.1. Распределенные RC-линии в SPICE

Из-за важности распределенной RC-цепи в современном дизайне большинство пакетов моделирования схем имеют встроенные высокоточные распределенные RC-модели. Например, пакет SPICE3 поддерживает равномерно распределенную модель RC-линии (URC). Данная модель аппроксимирует RC-цепь как сеть сосредоточенных RC-сегментов с внутренне генерируемыми узлами. В качестве параметров модели используются длина провода L и (дополнительно) число сегментов, используемых в модели.

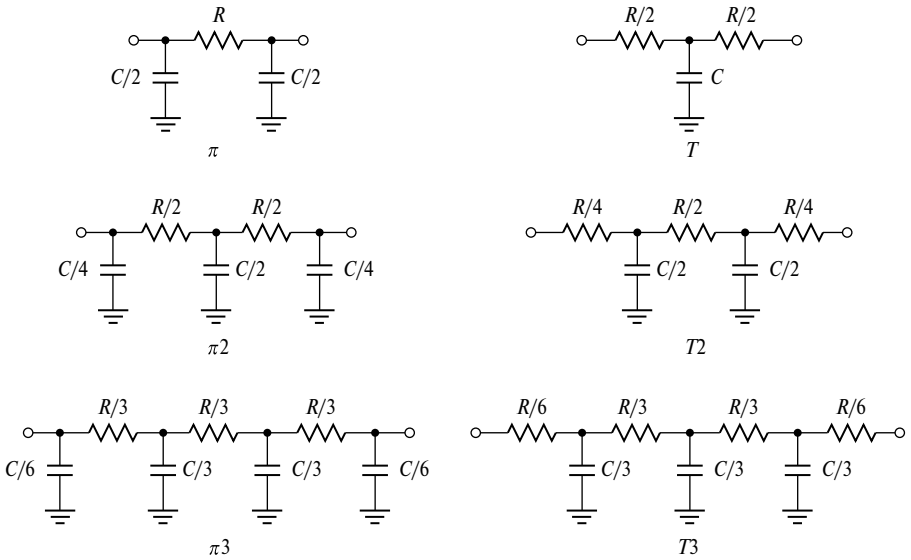


Рис. 4.26. Модели распределенной RC-цепи

Пример 4.13. Модель URC (SPICE3)

Ниже приведен типичный пример команды SPICE3, моделирующей распределенную RC-цепь.

```
U1 N1=1 N2=2 N3=0 URCMOD L=50m N=6
MODEL URCMOD URC(RPERL=75K CPERL=100pF)
```

N1 и N2 представляют концы линии, а N3 — узел, с которым связаны емкости. RPERL и CPERL обозначают удельные (на 1 м) сопротивление и емкость.

Если используемый вами пакет не поддерживает распределенную RC-модель или если вычислительная сложность подобных моделей сильно замедляет процесс моделирования, можно самостоятельно построить простую, но точную модель, аппроксимировав распределенную RC-линию сосредоточенной RC-цепью с ограниченным числом элементов. Несколько таких аппроксимаций показаны на рис. 4.26 упорядоченными по возрастанию точности и сложности. Точность модели определяется числом каскадов. Например, ошибка модели $\pi 3$ не превышает 3%, чего обычно достаточно для большинства приложений.

4.5.2. SPICE-модели линии передачи

SPICE поддерживает модель линии передачи без потерь. Характеристики этой линии определяются характеристическим импедансом Z_0, а длину линии можно определить в двух формах. Во-первых, можно прямо определить задержку передачи TD, которая эквивалентна времени пробега. Во-вторых, можно задать частоту F и NL — безразмерную нормированную электрическую длину линии, которая измеряется относительно длины волны линии при частоте F. Справедливо следующее соотношение:

$$NL = F \cdot TD. \tag{4.40}$$

Моделей для линии передачи с потерями пакет в настоящее время не предлагает. При необходимости потери можно смоделировать, разбив длинную линию передачи на ряд коротких участков и включив в каждый из них последовательное сопротивление. Используя данную аппроксимацию, будьте аккуратны. Во-первых, ее точность ограничена; во-вторых, может пострадать скорость моделирования, поскольку SPICE выбирает шаг по времени, меньший или равный половине значения TD. Для небольших линий передачи такой шаг может оказаться намного меньше, чем необходимо для анализа транзистора.

4.5.3. Перспектива: взгляд в будущее

Придерживаясь подхода, выбранного при рассмотрении МОП-транзистора, стоит исследовать, как будут эволюционировать параметры провода при дальнейшем масштабировании технологии. Итак, по мере уменьшения размеров транзистора размеры межсоединений также должны уменьшаться, чтобы использовать все возможности процесса масштабирования.

Если подходить к задаче прямолинейно, необходимо масштабировать все размеры провода с одним коэффициентом S, считая, что речь идет о транзисторах (*идеальное масштабирование*). Это может оказаться невозможным, по крайней мере, для одного размера провода: его длины. Можно догадаться, что длина *локальных межсоединений* (проводов, соединяющих близко расположенные транзисторы) масштабируется так же, как указанные транзисторы. С другой стороны, *глобальные межсоединения*, обеспечивающие связи между большими модулями и контурами входа-выхода, имеют другое поведение относительно масштабирования. В качестве примеров подобных проводов можно привести тактовые импульсы и сигнал шины данных и команд. На рис. 4.27 представлена гистограмма с распределением длин провода в реальной микропроцессорной схеме, содержащей примерно 90 000 элементов [Davis98]. Хотя большинство проводов имеют длину всего в несколько элементов, цепочка таких проводов может превышать длину 500 элементов.

Средняя длина данных длинных проводов пропорциональна размеру кристалла (или сложности) конкретной схемы. В современных схемах наблюдается интересная тенденция — хотя размеры транзисторов последовательно уменьшаются в течение нескольких последних десятилетий, размеры кри-

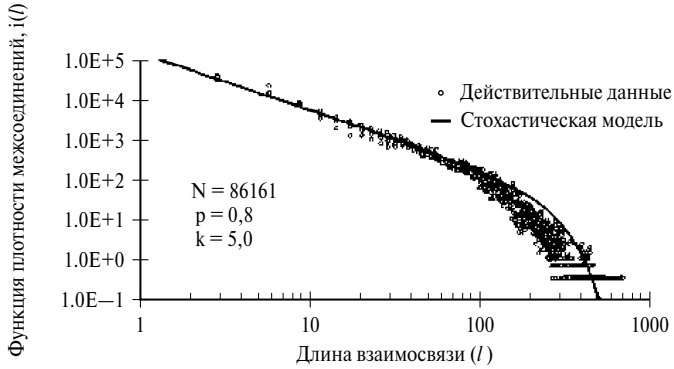


Рис. 4.27. Распределение длин проводов в современном микропроцессоре как функция расстояния между элементами

Таблица 4.8. Идеальное масштабирование свойств проводников

Параметр	Формула	Локальный провод	Провод с постоянной длиной	Глобальный провод
W, H, t		$1/S$	$1/S$	$1/S$
L		$1/S$	1	$1/S_C$
C	LW/t	$1/S$	1	$1/SC$
R	L/WH	S	S^2	S^2/S_C
CR	L^2/Ht	1	S^2	S^2/S_C^2

сталла так же постепенно растут. Фактически, размер типичного кристалла (корень квадратный из площади кристалла) увеличивается на 6% за год, удваиваясь каждые десять лет. С начала 1960-х чипы выросли с 2×2 мм приблизительно до 2×2 см в 2000. Хотя темпы роста размеров кристалла и замедляются, в ближайшие годы он вряд ли остановится полностью.

Таким образом, изучая поведение длины провода относительно масштабирования, мы должны различать локальные и глобальные проводники. При последующем анализе мы будем рассматривать три режима: локальные проводники ($S_L = S > 1$), проводники с постоянной длиной ($S_L = 1$) и глобальные проводники ($S_L = S_C < 1$).

Предположим теперь, что все остальные размеры соединяющей структуры (W, H, t) масштабируются с коэффициентом S. В результате получается зависимость, проиллюстрированная в табл. 4.8. Обратите внимание на то, что все сказанное является всего лишь приближением первого порядка, цель которого — изучить общие тенденции. Различные более сложные эффекты (например, краевая емкость) при этом игнорируются, а также не рассматриваются технологические прорывы (например, изобретение новых соединительных и изолирующих материалов).

Из приведенного примера можно сделать удивительный вывод, что масштабирование технологии не уменьшает задержку проводов (воплощенную в постоянной времени RC). Постоянная задержка предсказывается локальными проводами, тогда как задержка глобальных проводов увеличивается на

50% в год (для $S = 1,15$ и $S_C = 0,94$). Данное положение дел сильно отличается от ситуации с задержкой элемента, которая уменьшается из года в год. Из сказанного понятно, почему задержки проводов начинают играть доминирующую роль в современной разработке цифровых интегральных схем.

Тем не менее идеальное масштабирование имеет свои проблемы, поскольку приводит к быстрому увеличению сопротивления проводом. Это объясняет привлекательность других технологий масштабирования межсоединений. Например, одним из подходов является масштабирование толщины провода с другим коэффициентом. Из модели "с постоянным сопротивлением", приведенной в табл. 4.9, видно, что получится, если вообще не масштабировать толщину провода. Хотя данный подход, похоже, имеет положительное влияние на быстродействие, он выводит на первый план компоненты краевой и перекрестной емкости. В результате мы вводим дополнительный коэффициент масштабирования емкости ϵ_c (> 1), который учитывает горизонтальную природу емкости, проявляющуюся, когда при постоянной высоте уменьшаются ширина провода и размер элементов.

Такой сценарий масштабирования предлагает несколько более оптимистичную перспективу, если предполагать, что $\epsilon_c < S$. Однако задержка будет значительно возрастать для средних и длинных проводов, независимо от сценария масштабирования. Чтобы эти задержки не стали чрезмерными, необходимо улучшать технологию межсоединений. Как вариант, можно использовать лучшие соединительные (Cu) и изоляционные материалы (полимеры и воздух). Другая возможность — различать локальные и глобальные провода. В первом случае жизненно важны степень интеграции и низкая емкость; во втором — контроль над сопротивлением. Чтобы удовлетворить данным противоречивым требованиям, в современных технологиях межсоединений объединяют плотную и тонкую сетку проводки на низких металлических уровнях с толстыми и разнесенными в пространстве проводами на более высоких уровнях (рис. 4.28). Несмотря на подобный прогресс, очевидно, что в последующие годы межсоединения будут играть доминирующую роль как в быстродействующих, так в низкоэнергетичных цифровых схемах. Продолжение пути цифровых полупроводников вполне может оказаться связанным с введением альтернативных и инновационных подходов, подобных трехмерной интеграции, рассмотренной в главе 2.

4.6. Резюме

В данной главе представлен четкий и глубокий анализ роли и поведения соединительного провода в современной полупроводниковой технологии. Основной целью является определение доминирующих параметров, задающих значения паразитных нагрузок провода (емкости, сопротивления и индуктивности), и представление достаточных моделей, помогающих анализировать и оптимизировать сложные цифровые схемы.

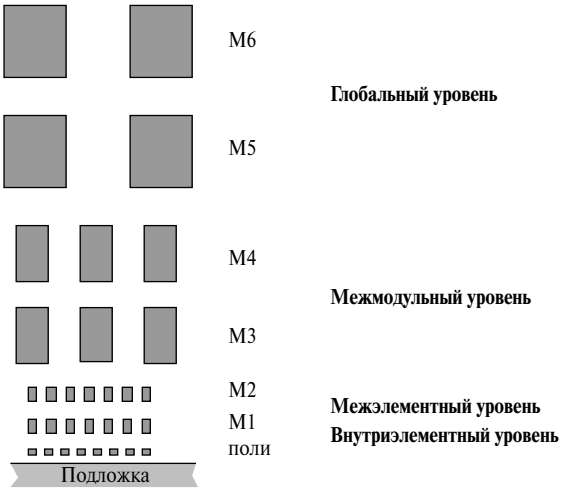


Рис. 4.28. Иерархия межсоединений в 0,25-микронном КМОП-процессе

Таблица 4.9. Масштабирование свойств проводов с “постоянным сопротивлением”

Параметр	Формула	Локальный провод	Провод с постоянной длиной	Глобальный провод
W, t		$1/S$	$1/S$	$1/S$
H		1	1	1
L		$1/S$	1	$1/S_C$
C	$\epsilon_c LW/t$	ϵ_c/S	ϵ_c	ϵ_c/S_C
R	L/WH	1	S	S/S_C
CR	L^2/Ht	ϵ_c/S	$\epsilon_c S$	$\epsilon_c S/S_C^2$

4.7. Для любознательных

Тема межсоединений и их моделирования обсуждается довольно живо, ей уделяется много внимания в журналах и на конференциях. Кроме того, данному вопросу посвящены многие учебники. В частности, для последующего изучения можно порекомендовать [Bakoglu90], [Tewksbury94] и [Dally98].

4.7. Список использованной литературы

[Bakoglu90] H. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, Addison-Wesley, 1990.

[Dally98] B. Dally, *Digital Systems Engineering*, Cambridge University Press, 1998.

[Davis98] J. Davis and J. Meindl, “Is Interconnect the Weak Link?” *IEEE Circuits and Systems Magazine*, March 1998, p. 30–36.

[Elmore48] E. Elmore, "The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers", *Journal of Applied Physics*, January 1948, p. 55–63.

[Horowitz83] M. Horowitz, "Timing Models for MOS Circuits", Ph.D. diss., Stanford University, 1983.

[Lin84] T. Lin and C. Mead, "Signal delay in general RC networks", *IEEE Transactions on Computer-Aided Design*, 1984, vol. 3, no. 4, p. 321–349.

[Rubinstein83] J. Rubinstein, P. Penfield, and M. Horowitz, "Signal Delay in RC Networks", IEEE Transactions on Computer-Aided Design, July 1983, vol. CAD-2, p. 202–211.

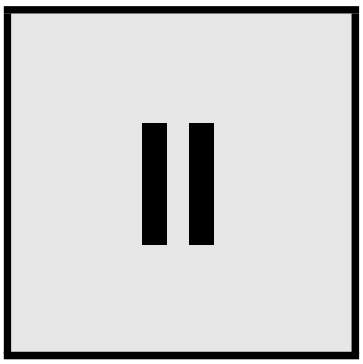
[Schaper83] L. Schaper and D. Amey, "Improved Electrical Performance Required for Future MOS Packaging", *IEEE Trans. on Components, Hybrids and Manufacturing Technology*, September 1983, vol. CHMT-6, p. 282–289.

[Sylvester97] D. Sylvester, "High-Frequency VLSI Interconnect Modeling", *Project Report EE241*, UCB, May 1997.

[Tewksbury94] S. Tewksbury, ed., *Microelectronics System Interconnections.Performance and Modeling*, IEEE Press, 1994.

[Vdmeijs84] N. Van De Meijs and J. Fokkema, "VLSI Circuit Reconstruction from Mask Topology", *Integration*, 1984, vol. 2, no. 2, p. 85–119.

ЧАСТЬ



Перспектива схемы

"Хороший ученый — это человек с оригинальными идеями. Хороший инженер — это человек, создающий проект, работающий с максимально малым числом оригинальных идей".

Фриман Дайсон (Freeman Dyson), *Disturbing the Universe*, 1979

"Проект — это то, что остается у разработчика, когда у него уже нет ни времени, ни денег".

Джеймс Пул (James Poole), *The Fifth 637 Best Things Anybody Ever Said*, 1993

ГЛАВА 5

КМОП-инвертор

Количественное описание целостности, быстродействия и энергетических метрик инвертора

Оптимизация структуры инвертора

5.1. Введение
5.2. Статический КМОП-инвертор — интуитивная перспектива
5.3. Расчет устойчивости КМОП-инвертора (статическое поведение)
5.3.1. Порог переключения
5.3.2. Запас помехоустойчивости
5.3.3. Еще раз об устойчивости
5.4. Быстродействие КМОП-инвертора: динамическое поведение
5.4.1. Расчет емкости
5.4.2. Задержка распространения: анализ в первом приближении
5.4.3. Задержка распространения с точки зрения разработки
5.5. Мощность, энергия и задержка энергии
5.5.1. Динамическое потребление мощности
5.5.2. Статическое потребление
5.5.3. Собираем все вместе
5.5.4. Анализ потребления мощности с помощью SPICE
5.6. Перспектива: масштабирование технологии и ее влияние на метрики инвертора
5.7. Резюме
5.8. Для любознательных
5.9. Список использованной литературы

5.1. Введение

Инвертор является сердцем всех цифровых разработок. Как только вы поймете его принцип действия и предлагаемые им возможности, разработка более сложных структур, подобных логическим элементам, сумматорам, умножителям и микропроцессорам, существенно упростится. Электрическое поведение названных сложных схем можно вывести почти целиком, экстраполируя результаты, полученные для инверторов. Анализ инверторов можно расширить на объяснение поведения более сложных логических элементов (НЕ-И, НЕ-ИЛИ или ИСКЛЮЧАЮЩЕЕ ИЛИ), которые, в свою очередь, являются строительными блоками умножителей и процессоров.

В этой главе мы сфокусируем внимание на одной реализации инвертора — статическом КМОП-инверторе. Пожалуй, в настоящее время это самый популярный инвертор, следовательно, он заслуживает особого внимания. Данный элемент мы рассмотрим относительно различных метрик разработки, сформулированных в главе 1.

- *Стоимость*, выраженная через сложность и площадь.
- *Целостность и устойчивость*, выраженная через статическое (или стационарное) поведение.
- *Быстродействие*, определяемое динамической (или переходной) характеристикой.
- *Энергетический КПД*, определяемый потребляемой энергией и мощностью.

Используя данный анализ, мы разработаем модель элемента и определим его расчетные параметры. Мы создадим метод, который позволит выбирать значения параметров, чтобы получающийся проект удовлетворял требуемым спецификациям. Хотя численное значение любого из параметров можно легко определить для любой заданной технологии, мы также обсудим, как на них влияет *масштабирование технологии*.

В данной главе мы сконцентрируем внимание исключительно на КМОП-инверторе, а в следующей рассмотрим, как та же методология применяется к другим топологиям элементов.

5.2. Статический КМОП-инвертор — интуитивная перспектива

На рис. 5.1 показана схема статического КМОП-инвертора. Чтобы понять его работу, можно воспользоваться простой моделью МОП-транзистора в виде ключа, введенной в главе 3 (см. рис. 3.26). Действительно, приведенный транзистор представляется собой обычный ключ с бесконечным сопротивлением в состоянии "выключено" (при $|V_{GS}| < |V_T|$) и конечным сопротивлением в состоянии "включено" (при $|V_{GS}| > |V_T|$). В результате получа-

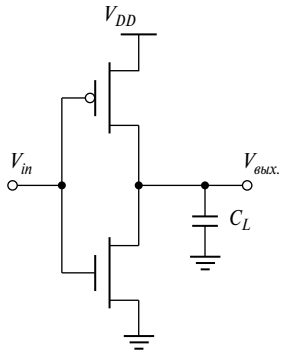


Рис. 5.1. Статический КМОП-инвертор. Напряжение источника питания равно V_{DD}

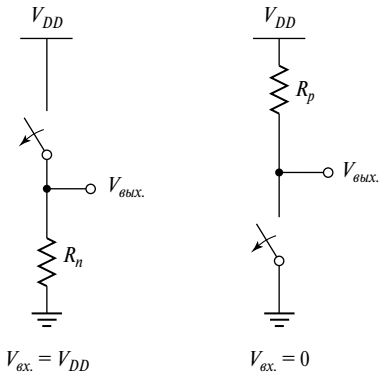


Рис. 5.2. Модель КМОП-инвертора в виде ключа

ем следующую интерпретацию инвертора. Когда $V_{\text{вх.}}$ высоко и равно V_{DD}, n-МОП-транзистор включен, а p-МОП-транзистор выключен. Таким образом, эквивалентную схему можно представить так, как показано на рис. 5.2, *а*. Между $V_{\text{вых.}}$ и узлом-землей существует прямой тракт, в результате значение напряжения устанавливается равным 0 В. С другой стороны, когда на вход поступает низкое напряжение (0 В), n-МОП-транзистор выключен, а p-МОП-транзистор включен. Из эквивалентной схемы, представленной на рис. 5.2, *б*, видно, что между V_{DD} и $V_{\text{вых.}}$ существует прямой тракт, что дает высокое напряжение на выходе. Очевидно, что представленный элемент работает как инвертор.

Используя представление в виде ключа, можно вывести несколько других важных свойств статического КМОП-инвертора.

- Высокий и низкий уровень выходного напряжения равны V_{DD} и GND соответственно; другими словами, размах напряжения равен напряжению источника питания. В результате получаем высокий запас помехоустойчивости.

- Логические уровни не зависят от относительных размеров устройства, поэтому транзисторы могут иметь минимальный размер. Элементы с подобным свойством называются *безотносительными*. Это отличает их от *относительных логических элементов*, где уровни логики определяются относительными размерами составляющих их транзисторов.
- В стационарном состоянии всегда существует тракт с конечным сопротивлением между выходом и V_{DD} либо землей. Следовательно, грамотно спроектированный КМОП-инвертор имеет *низкий выходной импеданс*, что делает его менее чувствительным к шуму и возмущениям. Типичные значения выходного сопротивления принадлежат килоомному диапазону.
- *Входное сопротивление* КМОП-инвертора очень высоко, поскольку вентиль МОП-транзистора является практически идеальным диэлектриком и не пропускает постоянную составляющую входного сигнала. Поскольку входной узел инвертора соединяется только с вентилями транзистора, стационарный входной ток практически равен нулю. Теоретически один инвертор может питать бесконечное число элементов (или иметь бесконечную нагрузочную способность по выходу), при этом функционируя в полном объеме; тем не менее (как станет понятно чуть позже) увеличение нагрузочной способности по выходу также увеличивает задержку распространения. Хотя нагрузочная способность по выходу никак не влияет на стационарный режим, она ухудшает переходную характеристику.
- При работе в стационаром режиме (т.е. когда вход и выход постоянны) между шинами питания и заземления не существует прямого тракта. Отсутствие протекания тока (токи утечки не учитываем) означает, что элемент не потребляет мощности в статическом состоянии.

Отступление

Предыдущее наблюдение, хотя и кажется очевидным, является чрезвычайно важным; оно — одна из основных причин использования КМОП в цифровых технологиях. В 1970–начале 1980-х гг. ситуация было совершенно иной. Все первые микропроцессоры (например, Intel 4004) были реализованы с использованием чистой технологии n-МОП. Отсутствие в подобной технологии дополняющих (комплементарных) элементов (подобных n-МОП- и p-МОП-транзисторам) делало реализацию инверторов с нулевым потреблением энергии в стационарном режиме нетривиальной задачей. Получавшееся в результате потребление энергии жестко ограничивало сверху число логических элементов, которые можно было интегрировать в один кристалл; следовательно, переход на КМОП в 1980-х годах, когда масштабирование технологии сделало возможным более плотную интеграцию, был предсказуемым.

Природу и форму передаточной функции по напряжению (Voltage-Transfer Characteristic — VTC) можно исследовать графически, наложив передаточные функции силы тока n-МОП- и p-МОП-устройств. Подобная гра-

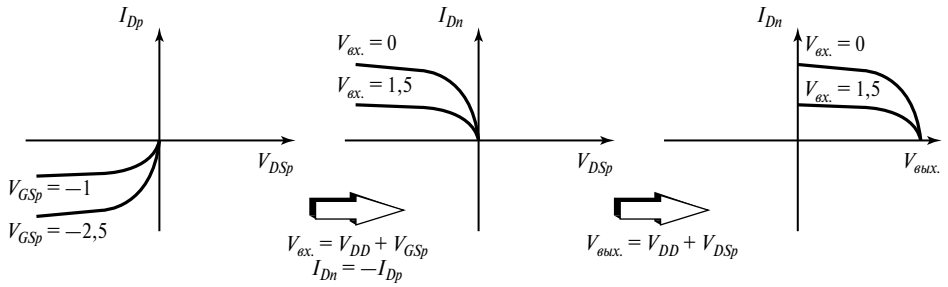


Рис. 5.3. Преобразование ВАХ p-МОП-транзистора в общую систему координат (предполагается, что $V_{DD} = 2,5$ В)

фическая конструкция называется *графиком линии нагрузки*. Она требует, чтобы вольт-амперные характеристики n-МОП- и p-МОП-устройств были сведены в одной системе координат. Выберем в качестве переменных напряжение на входе $V_{\text{вх.}}$, напряжение на выходе $V_{\text{вых.}}$ и ток стока n-МОП-транзистора I_{DN}. Вольт-амперную характеристику p-МОП-транзистора можно преобразовать в выбранную систему координат с помощью следующих соотношений (индексы n и p обозначают n-МОП- и p-МОП-устройства соответственно):

$$\begin{aligned} I_{DSn} &= -I_{DSn}; \\ B_{GSn} &= V_{\text{вх.}}; V_{GSp} = V_{\text{вх.}} - V_{DD}; \\ B_{DSn} &= V_{\text{вых.}}; V_{DSp} = V_{\text{вых.}} - V_{DD}. \end{aligned} \tag{5.1}$$

Линии нагрузки p-МОП-устройства получаются зеркальным отражением относительно оси x и горизонтальным смещением на V_{DD}. Схематически данная процедура показана на рис. 5.3, где демонстрируются этапы преобразования исходных вольт-амперных характеристик p-МОП-транзистора в общее пространство с координатами $V_{\text{вх.}}$, $V_{\text{вых.}}$ и I_{Dn}.

Получающиеся в результате линии нагрузки представлены на рис. 5.4. Чтобы получить правильную рабочую точку постоянной составляющей, сила тока в устройствах n-МОП и p-МОП должна быть одинаковой. Графически это означает, что точка постоянного тока должна располагаться на пересечении соответствующих линий нагрузки. На графике отмечено несколько таких точек (при $V_{\text{вх.}} = 0$, $0,5$, 1, $1,5$, 2 и $2,5$ В). Видно, что все рабочие точки расположены либо на высоком, либо на низком уровне выходного сигнала. Таким образом, передаточная функция инвертора имеет очень узкую область перехода. Это следует из высокого коэффициента усиления переходного процесса при переключении, когда n-МОП- и p-МОП-транзисторы одновременно включены и находятся в режиме насыщения. В данной рабочей области небольшое изменение входного напряжения приводит к большому изменению выхода. Исходя из всего сказанного, можно получить передаточную функцию по напряжению, приведенную на рис. 5.5.

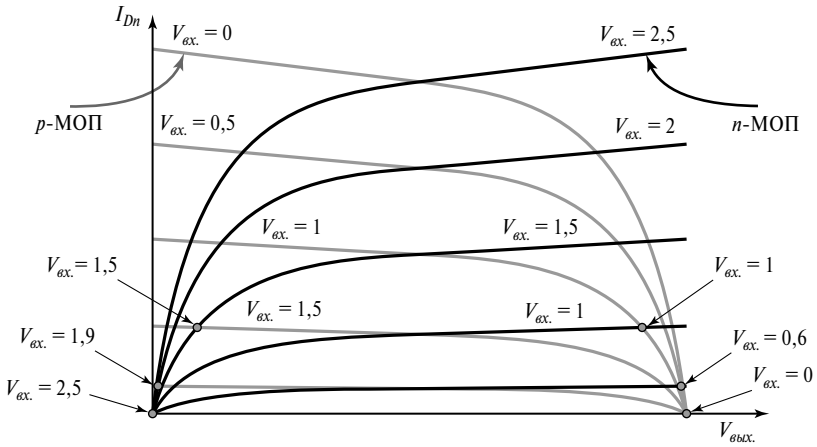


Рис. 5.4. Кривые нагрузки для n-МОП- и p-МОП-транзисторов статического КМОП-инвертора ($V_{DD} = 2,5$ В). Кружочками отмечены рабочие точки для различных входных напряжений

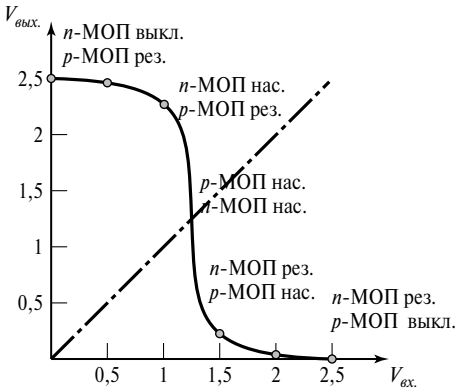


Рис. 5.5. Передаточная функция по напряжению статического КМОП-инвертора, выведенная из рис. 5.4 ($V_{DD} = 2,5$ В). Для каждой рабочей области обозначены режимы транзисторов — выкл. (выключенный), рез. (резистивный) или нас. (насыщенный)

Прежде чем мы перейдем к аналитическим деталям работы КМОП-инвертора, стоит качественно проанализировать переходное поведение элемента. На данную характеристику влияет в основном выходная емкость элемента C_L, которая складывается из диффузионной емкости стока n-МОП- и p-МОП-транзисторов, емкости соединительных проводов и входной емкости нагрузочных элементов. Предполагая, что транзисторы переключаются мгновенно, мы можем приблизительно представить переходную характеристику, снова используя упрощенную модель ключа (рис. 5.6, *а*). Время срабатывания элемента определяется как время, требуемое для зарядки конденсатора с емкостью C_L через резистор с сопротивлением R_p. В примере 4.5 мы на-

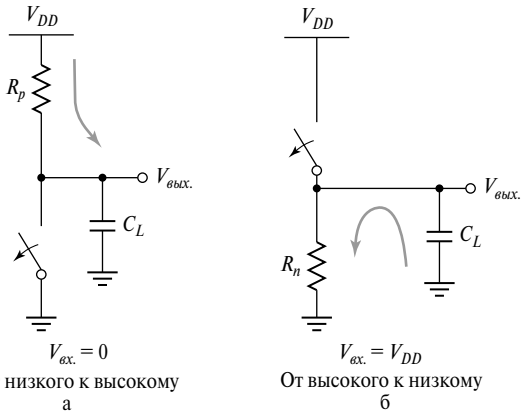


Рис. 5.6. Модели динамического поведения статического КМОП-инвертора, представленного в виде ключа

шли, что задержка распространения подобной сети пропорциональна постоянной времени R_pC_L. **Следовательно, чтобы создать быстрый элемент, нужно либо поддерживать малой выходную емкость, либо уменьшать сопротивление транзистора во включенном состоянии.** Чтобы реализовать второй сценарий, необходимо увеличивать отношение W/L устройства. Подобные соображения справедливы для изменения уровня сигнала от высокого к низкому (см. рис. 5.6, *б*), которое характеризуется постоянной времени R_nC_L. Следует помнить, что сопротивление n-МОП- и p-МОП-транзисторов во включенном состоянии не является постоянным; на самом деле оно — нелинейная функция напряжения внутри транзистора. Из-за этого точное определение задержки распространения усложняется. (Подробное описание того, как следует анализировать и оптимизировать быстродействие статического КМОП-инвертора, приводится в разделе 5.4.)

5.3. Расчет устойчивости КМОП-инвертора (статическое поведение)

В предыдущем качественном обсуждении была обрисована общая форма характеристики передачи по напряжению статического КМОП-инвертора и найдены значения V_{OH} и V_{OL}, равные соответственно V_{DD} и GND. Теперь нужно определить точные значения V_M, V_{IH} и V_{IL}, а также запас помехоустойчивости.

5.3.1. Порог переключения

Порог переключения V_M определяется как точка, в которой $V_{\text{вх.}} = V_{\text{вых.}}$. Его значение можно получить графически, найдя пересечения передаточной функции по напряжению с линией, определяемой уравнением $V_{\text{вх.}} = V_{\text{вых.}}$

(см. рис. 5.5). В данной области и p-МОП, и n-МОП насыщены, поскольку $V_{DS} = V_{GS}$. Аналитическое выражение для V_M можно получить, приравняв силы тока, протекающего через транзисторы. Решим эту задачу для случая, когда напряжение источника питания достаточно высоко, чтобы можно было считать, что устройства находятся в режиме насыщения скорости (или $V_{DSAT} < V_M - V_T$). Более того, мы игнорируем эффекты модуляции длины канала. Итак, получим следующее:

$$\begin{aligned} k_n V_{DSATn}\left(V_M - V_{Tn} - \frac{V_{DSATn}}{2}\right) + \\ + k_p V_{DSATp}\left(V_M - V_{DD} - V_{Tp} - \frac{V_{DSATn}}{2}\right) = 0. \end{aligned} \tag{5.2}$$

Выражая отсюда V_M, находим, что

$$\begin{aligned} V_M &= \frac{\left(V_{Tn} + \frac{V_{DSATn}}{2}\right) + r\left(V_{DD} + V_{Tp} + \frac{V_{DSATp}}{2}\right)}{1+r}, \\ \text{где } r &= \frac{k_p V_{DCATp}}{k_n V_{DCATn}} = \frac{\upsilon_{satp} W_p}{\upsilon_{satn} W_n}, \end{aligned} \tag{5.3}$$

в предположении, что толщина окисла в p-МОП- и n-МОП-транзисторах одинакова. Для больших значений V_{DD} (по сравнению с пороговым напряжением и напряжением насыщения) уравнение (5.3) можно упростить:

$$V_M \approx \frac{r V_{DD}}{1+r}. \tag{5.4}$$

Из уравнения (5.4) следует, что порог переключения задается отношением r, которое сравнивает относительные напряжения питания p-МОП- и n-МОП-транзисторов. В общем случае желательно, чтобы напряжение V_M располагалось около середины доступного размаха напряжения (или в точке $V_{DD}/2$), поскольку это даст сравнимые значения нижней и верхней границ запаса помехоустойчивости. Это требует, чтобы r приблизительно равнялось 1, что эквивалентно такому изменению размеров p-МОП-устройства, чтобы $(W/L)_p = (W/L)_n \times (V_{DSATn}k'_n)/(V_{DSATp}k'_p)$. Чтобы поднять V_M, требуется более высокое значение r, т.е. более широкий p-МОП-транзистор. С другой стороны, увеличение влияния n-МОП-транзистора приближает порог переключения к GND.

Из уравнения (5.2) можно вывести требуемое отношение размеров p-МОП- и n-МОП-транзисторов, чтобы порог переключения имел искомое значение V_M:

$$\frac{(W/L)_p}{(W/L)_n} = \frac{k'_n V_{DSATn}(V_M - V_{Tn} - V_{DSATn}/2)}{k'_p V_{DSATp}(V_{DD} - V_M + V_{Tp} + V_{DSATp}/2)}. \tag{5.5}$$

Используя данное выражение, помните, что предположение о том, что оба устройства находится в режиме насыщения скорости, выполняется и для выбранной рабочей точки.

Задача 5.1. Порог переключения инвертора для длинноканальных устройств, или низкие напряжения питания

Предыдущие выражения были выведены из предположения, что транзисторы находятся в режиме насыщения скорости. Когда p-МОП- и n-МОП-транзисторы являются длинноканальными устройствами или когда напряжение источника питания низкое, насыщение скорости не происходит ($V_M - V_T < V_{DSAT}$). В таких случаях справедливо следующее выражение для V_M:

$$V_M = \frac{V_{Tn} + r(V_{DD} + V_{Tp})}{1 + r}, \text{ где } r = \sqrt{\frac{-k_p}{k_n}}. \tag{5.6}$$

Выведите эту формулу.

Максимизация запаса помехоустойчивости

Если желательно получить максимальный запас помехоустойчивости и симметричные характеристики, при разработке статических КМОП-схем рекомендуется находить компромисс между направляющими силами транзисторов, делая p-МОП-сегмент более широким, чем n-МОП-сегмент. Требуемое соотношение между указанными размерами находится из уравнения (5.5).

Пример 5.1. Порог переключения КМОП-инвертора

Найдем такие размеры p-МОП- и n-МОП-транзисторов, чтобы порог переключения КМОП-инвертора, реализованного с помощью традиционного 0,25-микронного КМОП-процесса, располагался посредине между шинами питания. Используем параметры процесса, указанные в примере 3.7, и предположим, что напряжение питания равно 2,5 В. Устройство минимального размера описывается отношением ширины к длине, равным 1,5. Используя уравнение (5.5), находим, что

$$\frac{(W/L)_p}{(W/L)_n} = \frac{115 \times 10^{-6}}{30 \times 10^{-6}} \times \frac{0,63}{1,0} \times \frac{(1,25 - 0,43 - 0,63/2)}{(1,25 - 0,4 - 1,0/2)} = 3,5.$$

На рис. 5.7 показан порог переключения как функция отношения размеров p-МОП- к n-МОП-транзистору, полученный путем моделировании. Полученное при модуляции значение отношения 3,4 для порога переключения в 1,25 В согласуется со значением, предсказываемым уравнением (5.5).

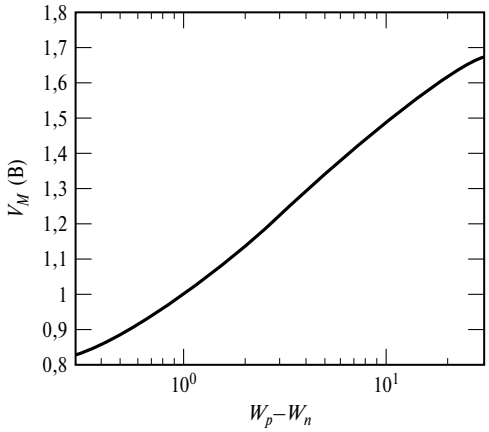


Рис. 5.7. Смоделированный порог переключения инвертора как функция отношения размеров p-МОП-транзистора к n-МОП-транзистору (0,25 мкм КМОП, $V_{DD} = 2,5$ В)

Анализ кривой, представленной на рис. 5.7, позволяет сделать несколько интересных наблюдений.

1. Значение V_M относительно нечувствительно к характеристическому отношению размеров устройства. Это означает, что небольшое изменение данного параметра (например, с 3,5 до 3 или 2,5) не сильно повлияет на передаточную функцию. Следовательно, в промышленности принято делать ширину p-МОП-транзистора несколько меньше, чем требуется для точной симметрии. Если в условии предыдущего примера установить отношение равным 3, 2,5 и 2, получим пороги переключения 1,22, 1,18 и 1,13 В соответственно.
2. Изменение отношения W_p к W_n приводит к смещению области перехода передаточной функции. Увеличение ширины p-МОП- или n-МОП-транзистора сдвигает V_M к V_{DD} или GND соответственно. Данное свойство может быть очень полезным, поскольку в определенных системах желательны асимметричные передаточные функции. Сказанное иллюстрируется на рис. 5.8. Входной сигнал $V_{\text{вх.}}$ имеет очень шумное нулевое значение. Пропуская данный сигнал через симметричный инвертор, мы получим ошибочные значения (рис. 5.8, *а*). Для решения этой проблемы можно увеличить порог инвертора, что даст правильную характеристику (см. рис. 5.8, *б*). Позже мы рассмотрим другие схемы, в которых желательны инверторы с асимметричными порогами переключения. Тем не менее значительно изменить порог переключения нелегко, особенно, если отношение питающего напряжения к порогу транзистора относительно мало (в нашем случае — $2,5/0,4 = 6$). Чтобы изменить порог на 1,5 В, потребуется характеристическое отношение 11, а дальнейшее увеличение порога будет неоправданно догорим. Помните, что график на рис. 5.7 изображен в полулогарифмическом формате.

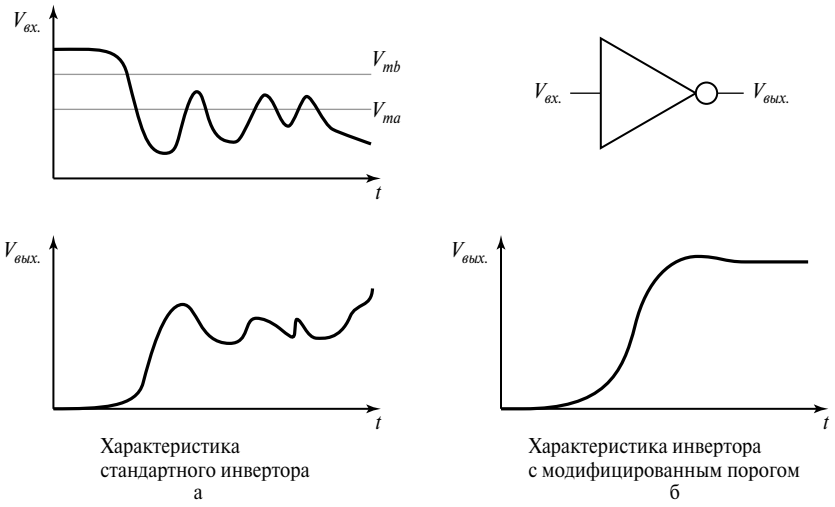


Рис. 5.8. Изменение порога инвертора может повысить надежность работы схемы

5.3.2. Запас помехоустойчивости

По определению V_{IH} и V_{IL} являются рабочими точками инвертора, когда $\frac{dV_{\text{вых.}}}{dV_{\text{вх.}}} = -1$. Используя терминологию разработки аналоговых схем, можно сказать, что это две точки, в которых коэффициент усиления g усилителя, сформированного инвертором, равен -1. Хотя мы можем вывести аналитические выражения для V_{IH} и V_{IL}, они будут громоздкими и с их помощью будет трудно понять, посредством каких параметров действительно можно настраивать запас помехоустойчивости.

Существует более простой подход — использовать кусочно-линейную аппроксимацию передаточной функции напряжения, как показано на рис. 5.9. Область перехода аппроксимируется прямой линией, коэффициент усиления на которой равен коэффициенту g при пороге переключения V_M. Чтобы определить точки V_{IH} и V_{IL}, находится пересечение линий V_{OH} и V_{OL}. Ошибка, вводимая при этом, мала и принадлежит допустимому диапазону, задаваемому требованиями к исходному проекту. Описанный подход дает следующие выражения для ширины и области перехода: $V_{IH} - V_{IL}$, V_{IH}, V_{IL}, запас помехоустойчивости — от NM_H до NM_L:

$$\begin{aligned} V_{IH} - V_{IL} &= -\frac{(V_{OH} - V_{OL})}{g} = \frac{-V_{DD}}{g}; \\ V_{IH} &= V_M - \frac{V_M}{g}; \quad V_{IL} = V_M + \frac{V_{DD} - V_M}{g}; \\ NM_H &= V_{DD} - V_{IH}; \quad NM_L = V_{IL}. \end{aligned} \tag{5.7}$$

Из приведенных выражений понятно, что высокое усиление в области перехода крайне желательно. В граничном случае бесконечного усиления запас

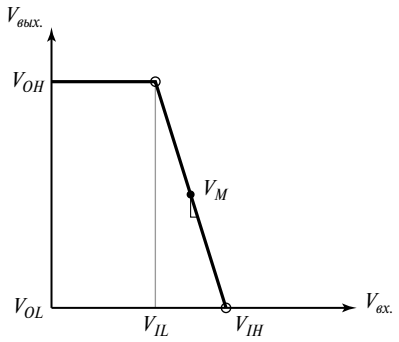


Рис. 5.9. Кусочно-линейная аппроксимация передаточной характеристики по напряжению упрощает расчет V_{IL} и V_{IH}

помехоустойчивости сводится к $V_{OH} - V_M$ и $V_M - V_{OL}$ для NM_H и NM_L соответственно и захватывает весь размах напряжения.

Итак, нам осталось определить усиление в средней точке статического КМОП-инвертора. Снова будем предполагать, что p-МОП- и n-МОП-транзисторы находятся в режиме насыщения скорости. На рис. 5.4 видно, что усиление сильно зависит от огибающих силы тока в области насыщения. Следовательно, при анализе мы не можем игнорировать эффект модуляции длины канала, поскольку иначе получим бесконечное усиление. Итак, чтобы получить усиление, продифференцируем уравнение (5.8), справедливое в окрестности порога переключения, по $V_{\text{вх.}}$:

$$k_n V_{DSATn}(V_{\text{вх.}} - V_{Tn} - \frac{V_{DSATn}}{2})(1 + \lambda_n V_{\text{вых.}}) + \tag{5.8}$$
$$+ k_p V_{DSATp}(V_{\text{вх.}} - V_{DD} - V_{Tp} - \frac{V_{DSATp}}{2})(1 + \lambda_p V_{\text{вых.}} - \lambda_p V_{DD}) = 0.$$

Дифференцируя и выражая $dV_{\text{вых.}}/dV_{\text{вх.}}$, получаем следующее:

$$\frac{dV_{\text{вых.}}}{dV_{\text{вх.}}} = \tag{5.9}$$
$$\frac{k_n V_{DSATn}(1 + \lambda_n V_{\text{вых.}}) + k_p V_{DSATp}(1 + \lambda_p V_{\text{вых.}} - \lambda_p V_{DD})}{\lambda_n k_n V_{DSATn}(V_{\text{вх.}} - V_{Tn} - \frac{V_{DSATn}}{2}) + \lambda_p k_p V_{DSATp}(V_{\text{вх.}} - V_{DD} - V_{Tp} - \frac{V_{DSATp}}{2})}.$$

Игнорируя некоторые члены второго порядка и полагая $V_{\text{вх.}} = V_M$, получаем формулу для усиления:

$$g = -\frac{1}{I_D(V_M)} \frac{k_n V_{DSATn} + k_p V_{DSATp}}{\lambda_n - \lambda_p} \approx \tag{5.10}$$
$$\approx \frac{1 + r}{(V_M - V_{Tn} - \frac{V_{DSATn}}{2})(\lambda_n - \lambda_p)},$$

где $I_D(V_M)$ — сила тока, проходящего через инвертор при $V_{\text{вх.}} = V_M$. Усиление практически полностью определяется параметрами технологии (особенно модуляцией длины канала). Разработчик может повлиять на него только незначительно, выбирая напряжение источника питания и размеры транзисторов.

Пример 5.2. Передаточная функция напряжения и запас помехоустойчивости КМОП-инвертора

Предположим, что инвертор разработан с учетом 0,25-мкм КМОП-технологии с характеристическим отношением размеров p-МОП- к n-МОП-транзистору равным 3,4 и минимальным размером n-МОП-транзистора ($W = 0,375$ мкм, $L = 0,25$ мкм, $W/L = 1,5$). Рассчитаем вначале усиление при V_M ($= 1,25$ В):

$$\begin{aligned} I_D(V_M) &= 1,5 \times 115 \times 10^{-6} \times 0,63 \times (1,25 - 0,43 - 0,63/2) \times (1 + 0,06 \times 1,25) \\ &= 59 \times 10^{-6} \text{ А} \\ g &= -\frac{1}{59 \times 10^{-6}} \frac{1,5 \times 115 \times 10^{-6} \times 0,63 + 1,5 \times 3,4 \times 30 \times 10^{-6} \times 1,0}{0,06 + 0,1} = \\ &= -27,5. \end{aligned}$$

В результате получаем следующие значения V_{IL}, V_{IH}, NM_L, NM_H:

$$V_{IL} = 1,2 \text{ В}, V_{IH} = 1,3 \text{ В}, NM_L = NM_H = 1,2.$$

На рис. 5.10 для данного инвертора представлена смоделированная передаточная функция по напряжению, ее производная и усиление. Полученная характеристика близка к идеальной. Реальные значение V_{IL} и V_{IH} равны $1,03$ и $1,45$ В соответственно, что дает запас помехоустойчивости от $1,03$ до $1,05$ В. Данные значения меньше предсказанных, что можно объяснить двумя причинами.

- Уравнение (5.10) завышает оценку усиления. Как видно на рис. 5.10, *б*, максимальное усиление (при V_M) равно всего 17. При таком более низком значении усиления V_{IL} и V_{IH} равны соответственно 1,17 и 1,33 В[1].
- Самое важное отклонение вызвано кусочно-линейной аппроксимацией передаточной функции, оптимистичной относительно реального запаса помехоустойчивости. Тем не менее полученные выражения исключительно полезны в роли оценок в первом приближении, а также позволяют определить существенные параметры и их влияние.

Чтобы завершить данный пример, мы смоделировали выходное сопротивление инвертора в низком и высоком выходных состояниях. Были получены

[1] Кроме того, уравнение (5.10) не совсем справедливо в данном случае. Внимательный читатель мог заметить, что для указанных условий p-МОП-транзистор работает в режиме насыщения, а не насыщения скорости. Впрочем, данный фактор имеет минимальное влияние на результат.

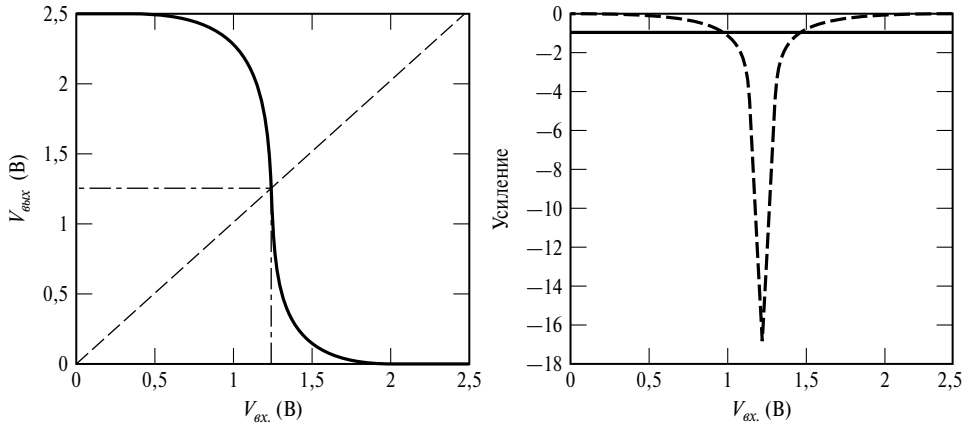


Рис. 5.10. Смоделированная передаточная функция по напряжению (а) и усиление по напряжению КМОП-инвертора (б) (0,25-микронный процесс, V_{DD} = 2,5 В)

невысокие значения — 2,4 и 3,3 кОм соответственно. Выходное сопротивление является хорошей мерой чувствительности элемента относительно шума на выходе, и его желательно уменьшать, насколько это возможно.

Отступление

Как ни странно (а возможно, абсолютно закономерно), но статический КМОП-инвертор можно использовать как аналоговый усилитель, поскольку он дает достаточно высокое усиление в области перехода. Тем не менее, как видно из графика, представленного на рис. 5.10, *б*, эта область очень узкая. Кроме того, инвертор не подходит по другим параметрам, важным для усилителя, например, подавлению шумов источника питания. Однако с помощью указанного факта можно продемонстрировать значительные отличия между цифровой и аналоговой разработкой. Если бы разработчик аналоговых устройств сместил усилитель к середине переходной области, чтобы получить максимальную линейность, то разработчик цифровых устройств предпочел бы работу в максимально нелинейных областях, что дает четко определяемые и разделяемые низкий и высокий сигналы.

Задача 5.2. Запас помехоустойчивости инвертора для длинноканальных устройств

Выведите выражение для усиления и запаса помехоустойчивости, предполагая, что *p*-МОП- и *n*-МОП-транзисторы являются длинноканальными устройствами (или что источник питания дает низкое напряжение), так что насыщение скорости не происходит.

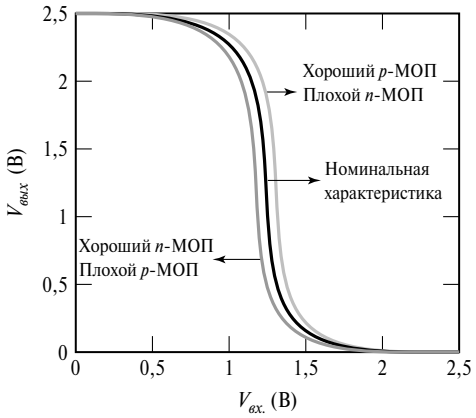


Рис. 5.11. Влияние колебания параметров устройства на передаточную функцию напряжения статического КМОП-инвертора. "Хорошее" устройство имеет меньшую толщину окисла (-3 нм), меньшую длину (-25 нм), большую ширину ($+30$ нм) и меньший порог (-60 мВ). "Плохой" транзистор характеризуется противоположными значениями

5.3.3. Еще раз об устойчивости

Колебания устройства

При разработке элемента для номинальных рабочих условий и типичных параметров устройства следует всегда помнить, что реальная рабочая температура может меняться в широких пределах и что параметры устройства после изготовления будут отклоняться от номинальных значений, использованных в процессе оптимизации разработки. К счастью, статическая характеристика статического КМОП-инвертора, похоже, относительно нечувствительна к данным колебаниям, и элемент сохраняет функциональность в широком диапазоне рабочих условий. Сказанное подтверждается рис. 5.7, где показано, что колебания размеров устройства имеют лишь незначительное влияние на порог переключения инвертора. Чтобы лишний раз подтвердить устойчивость элемента, повторно смоделируем передаточную функцию напряжения, заменив номинальные устройства их самыми худшими или самыми лучшими вариантами. Два граничных случая изображены на рис. 5.11: лучший, чем ожидалось, n-МОП-транзистор, объединенный с гораздо худшим p-МОП-транзистором, и противоположный сценарий. Сравнивая получающиеся кривые с номинальной характеристикой, видим, что работа элемента никоим образом не затрагивается и что колебания в основном приводят к смещению порога переключения. Такое устойчивое поведение, обеспечивающее функционирование элемента в широком диапазоне условий, дало значительный вклад в популярность статического КМОП-инвертора.

Масштабирование напряжения источника питания

В главе 3 мы отмечали, что уменьшение напряжения питания происходит примерно с той же скоростью, с которой в ходе улучшения технологии уменьшаются размеры устройств. В то же время пороговые напряжения устройства

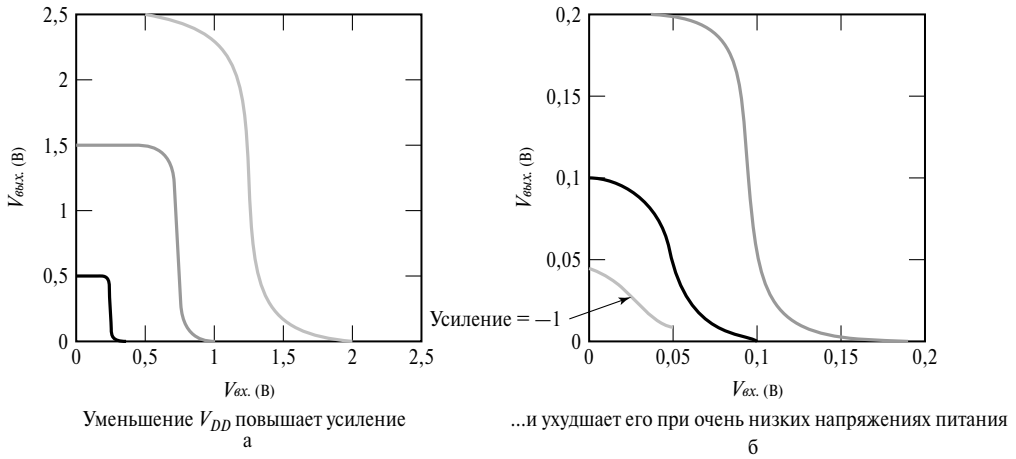


Рис. 5.12. Передаточная характеристика напряжения как функция напряжение источника питания (КМОП-технология 0,25 мкм)

остаются практически неизменными. Можно задаться вопросом, как эта тенденция влияет на работоспособность КМОП-инвертора. Будет ли инвертор продолжать работать при масштабировании напряжений и существуют ли потенциальные пределы масштабирования питания?

Первая подсказка относительно того, что может произойти, заключена в уравнении (5.10), из которого следует, что усиление инвертора в области перехода в действительности увеличивается с уменьшением напряжения питания! Обратите внимание на то, что для фиксированного отношения характеристических размеров транзистора r V_M приблизительно пропорционально V_{DD}. Изображая (нормированную) передаточную функцию по напряжению для различных напряжений источника питания, мы не только подтверждаем сделанное наблюдение, но и показываем, что инвертор нормально работает для напряжений, близких к пороговому напряжению составляющих его транзисторов (рис. 5.12, *а*). При напряжении 0,5 В (всего на 100 мВ выше порога транзистора) ширина области перехода составляет всего 10% напряжения источника питания (при максимальном усилении в 35), тогда как для 2,5 В оно увеличивается до 17%. Если статическая характеристика улучшается таким образом, тогда почему бы не использовать все цифровые схемы при низких напряжениях питания? В качестве объяснения можно привести три важные причины.

- Неразборчивое уменьшение питающего напряжения положительно влияет на рассеяние энергии, однако чрезвычайно вредно для задержки элемента, как будет показано в следующих разделах.
- Как только напряжение источника питания и внутреннее напряжение становятся сравнимыми, статическая характеристика становится очень чувствительной к колебаниям параметров устройства (например, порога транзистора).

- Масштабирование питающего напряжения означает уменьшение амплитуды сигнала. Хотя обычно это помогает уменьшить внутренний шум системы (вызванный, например, перекрестными помехами), структура становится более чувствительной к внешним источникам шума (масштабирование которых не происходит).

Чтобы лучше разобраться в вопросе потенциальных ограничений технологии масштабирования напряжения, рассмотрим передаточную функцию напряжения одного и того же инвертора при последовательно уменьшающихся напряжениях источника питания: 200, 100 и 50 мВ (см. рис. 5.12, *б*). Порог транзистора во всех трех случаях одинаковый. Как ни удивительно, но мы по-прежнему получаем характеристику инвертора, хотя питающего напряжения недостаточно даже для включения транзистора! Для объяснения данного факта необходимо рассмотреть работу транзистора при напряжениях ниже порогового. Подпорогового тока достаточно для переключения элемента между низким и высоким уровнями, а также для обеспечения достаточного усиления для получения приемлемой передаточной функции. Низкое значение тока переключения приводит к очень медленной работе, которая, впрочем, в определенных приложениях может быть вполне приемлемой (например, в часах).

Примерно при 100 мВ мы замечаем сильное ухудшение характеристики элемента. V_{OL} и V_{OH} уже не совпадают с шинами питания, и усиление области перехода приближается к 1. Последнее является основной накладкой: чтобы получить достаточное усиление для использования в цифровых схемах, необходимо, чтобы питающее напряжение по крайней мере в два раза превышало $\phi_T = kT/q$ (при комнатной температуре равно 25 мВ) — тепловое напряжение, введенное в главе 3 [Swanson72]. Из сказанного следует, что ниже данного напряжения следует учитывать еще и тепловой шум, который вполне может привести к неустойчивой работе. Таким образом, можно записать следующее соотношение:

$$V_{DDmin} > 2 \ldots 4 \frac{kT}{q}. \tag{5.11}$$

В уравнении (5.11) представлена фактическая нижняя граница, до которой можно масштабировать напряжения источника питания. Следовательно, единственный способ получить КМОП-инверторы, работающие при напряжении, меньшем 100 мВ, — это уменьшить окружающую температуру (другими словами, охладить схему).

Задача 5.3. Минимальное напряжение питания КМОП-инвертора

Когда напряжение источника питания становится ниже порогового напряжения, транзистор работает в подпороговой области и представляется экспоненциальной вольт-амперной характеристикой (как описывается в уравнении (3.39)). Выведите выражение для усиления инвертора при таких условиях (n-МОП- и p-МОП-транзисторы считайте симметричными, максимальное

усиление соответствует $V_M = V_{DD}/2$). Из полученной формулы видно, что минимальное напряжение является функцией параметра n транзистора:

$$g = -\left(\frac{1}{n}\right)(e^{V_{DD}/2\phi_T} - 1). \tag{5.12}$$

Согласно приведенному выражению, при $V_{DD} = 48$ мВ усиление падает до -1 (при $n = 1,5$ и $\phi_T = 25$ мВ).

5.4. Быстродействие КМОП-инвертора: динамическое поведение

Из представленного выше качественного анализа следует, что задержка распространения КМОП-инвертора определяется временем, необходимым для зарядки и разрядки конденсатора емкостью C_L через p-МОП- и n-МОП-транзисторы соответственно. Из данного наблюдения следует, что **для получения быстродействующих КМОП-схем необходимо максимально уменьшать** C_L. Таким образом, нам стоит вначале изучить основные компоненты нагрузочной емкости, а лишь после этого начинать глубокий анализ задержки распространения. Помимо указанного подробного анализа, в данном разделе мы представим краткую информацию о технологиях, которые может использовать разработчик для оптимизации быстродействия инвертора.

5.4.1. Расчет емкости

Вручную проанализировать МОП-схему, когда каждая емкость рассматривается отдельно, практически невозможно. Задача усложняется еще и тем, что в модели МОП-транзистора имеется множество нелинейных емкостей. Чтобы сделать анализ более реальным, предположим, что все емкости сосредоточены в одном большом конденсаторе с емкостью C_L, расположенном между $V_{\text{вых.}}$ и GND. Помните, что это существенное упрощение реальной ситуации, даже когда речь идет о простом инверторе.

Схематическое представление каскадной пары инвертора представлено на рис. 5.13. Здесь изображены все емкости, влияющие на переходную характеристику узла $V_{\text{вых.}}$. Изначально предполагается, что вход $V_{\text{вх.}}$ питается *идеальным источником напряжения с нулевыми временами нарастания и спада.* Учитывая только емкости, соединенные с выходным узлом, C_L можно разбить на следующие компоненты.

Емкость затвора-стока C_{gd12}

Во время первой половины (до 50%) переходного процесса на выходе узлы M_1 и M_2 находятся либо в режиме отсечки, либо в режиме насыщения. В таком случае единственный вклад в C_{gd12} дают перекрывающиеся емкости M_1 и M_2. Емкость канала МОП-транзисторов здесь значения не имеет, поскольку она

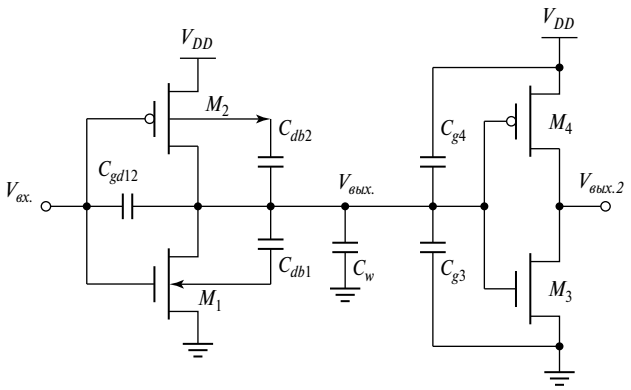


Рис. 5.13. Паразитные емкости, влияющие на переходную характеристику каскадной пары инвертора

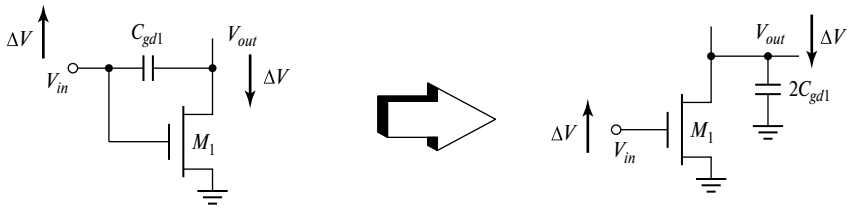


Рис. 5.14. Эффект Миллера. Конденсатор, на оба контакта которого подаются идентичные по амплитуде, но противоположные по знаку напряжения, можно заменить конденсатором на землю, значение которого в два раза больше исходного значения

полностью располагается между затвором и объемным материалом (режим отсечки) либо между затвором и истоком (режим насыщения) (см. главу 3).

Модель сосредоточенной емкости требует, чтобы данный конденсатор из плавающего затвора и стока был заменен емкостью на землю. Это достигается путем учета так называемого эффекта Миллера. В процессе перехода из низкого состояния в высокое или из высокого в низкое оконечные нагрузки данного конденсатора "затвор-сток" перемещаются в противоположных направлениях (рис. 5.14). Таким образом, изменение напряжения конденсатора с плавающим затвором вдвое больше реальной амплитуды выходного напряжения. Чтобы представить идентичную нагрузку на выходном узле, емкость на землю должна иметь значение, вдвое большее "плавающей" емкости.

Для конденсаторов, образованных затвором и стоком, используем следующее уравнение: $C_{gd} = 2C_{GD0}W$ (C_{GD0} — перекрывающаяся емкость на единицу ширины, используемая в модели SPICE). Подробное обсуждение эффекта Миллера можно найти в специальных учебниках, например [Sedra87, p. 57][2].

[2] Эффект Миллера, обсуждаемый в данном контексте, является упрощенной версией общего аналогового случая. В цифровом инверторе крупномасштабное усиление между входом и выходом всегда равно −1.

Диффузионные емкости C_{db1} и C_{db2}

Емкость между стоком и объемным материалом образована обратносмещенным pn-переходом. К сожалению, подобный конденсатор является достаточно нелинейным и сильно зависит от прикладываемого напряжения. В главе 3 мы утверждали, что лучшим подходом к упрощению анализа является замена нелинейного конденсатора линейным, имеющим такое же изменение заряда в искомом диапазоне напряжений. Чтобы связать линеаризованный конденсатор со значением емкости pn-перехода при нулевом смещении, вводится коэффициент K_{eq}:

$$C_{eq} = K_{eq} C_{j0}, \tag{5.13}$$

где C_{j0} — емкость pn-перехода на единицу площади при нулевом смещении. Повторим для удобства уравнение (3.11):

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1-m)} [(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}], \tag{5.14}$$

где ϕ_0 — встроенная контактная разность потенциалов; m — коэффициент выравнивания перехода. Обратите внимание на то, что напряжение на pn-переходе для обратносмещенных переходов является отрицательным.

Пример 5.3. K_{eq} для КМОП-инвертора 2,5 В

Рассмотрим изображенный на рис. 5.13 инвертор, разработанный с использованием традиционной 0,25-микронной КМОП-технологии. Существенные емкостные параметры этого процесса сведены в табл. 3.5.

Итак, проанализируем сначала n-МОП-транзистор (см. C_{db1} на рис. 5.13). Задержка распространения определяется временем между 50%-ными переходами входа и выхода. Для КМОП-инвертора это момент времени, когда $V_{\text{вых.}}$ достигает значения $1,25$ В, поскольку полный размах напряжения на выходе составляет $2,5$ В. Таким образом, мы линеаризуем емкость pn-перехода в интервале $\{2,5$ В, $1,25$ В$\}$ для перехода из высокого состояния в низкое и в интервале $\{01,25$ В$\}$ — для перехода из низкого состояния в высокое.

В процессе перехода на выходе из высокого состояния в низкое $V_{\text{вых.}}$ изначально равно $2,5$ В. Из-за того что объемный материал n-МОП-устройства соединен с землей, это дает обратное напряжение 2,5 В на стоковом переходе транзистора, или $V_{high} = -2,5$ В. В точке 50% $V_{\text{вых.}} = 1,25$ В или $V_{low} = -1,25$ В. Подставляя в формулу (5.14) величины, соответствующие верхней плате и боковой стенке диффузионной емкости, получаем следующие данные:

нижняя плата: $K_{eq}(m = 0,5; \phi_0 = 0,9) = 0,57$;
боковая стенка: $K_{eqsw}(m = 0,44; \phi_0 = 0,9) = 0,61$.

В процессе изменения уровня сигнала от низкого к высокому V_{low} и V_{high} равны 0 и $-1,25$ В соответственно, что дает более высокие значения K_{eq}:

нижняя плата: $K_{eq}(m = 0,5; \phi_0 = 0,9) = 0,79$;
боковая стенка: $K_{eqsw}(m = 0,44; \phi_0 = 0,9) = 0,81$.

У p-МОП-транзистора наблюдаем обратное поведение, поскольку на его подложку подается 2,5 В. Следовательно, при изменении уровня сигнала от высокого к низкому ($V_{low} = 0$, $V_{high} = -1,25$ В) получаем такие значения:
нижняя плата: $K_{eq}(m = 0,48; \phi_0 = 0,9) = 0,79$;
боковая стенка: $K_{eqsw}(m = 0,32; \phi_0 = 0,9) = 0,86$.

Наконец, при изменении уровня сигнала от низкого к высокому ($V_{low} = -1,25$ В, $V_{high} = -2,5$ В) получаем:
нижняя плата: $K_{eq}(m = 0,48; \phi_0 = 0,9) = 0,59$;
боковая стенка: $K_{eqsw}(m = 0,32; \phi_0 = 0,9) = 0,7$.

Используя данный подход, можно заменить емкость pn-перехода линейным компонентом и рассматривать ее как емкость любого другого устройства. Результатом линеаризации является небольшая ошибка сигналов напряжения и силы тока. На задержки логики упрощение влияет незначительно.

Емкость разводки C_w

Емкость, порожденная разводкой, зависит от длины и ширины соединительных проводов и является функцией расстояния разветвления по выходу от питающего элемента и числа элементов на выходе. Как отмечалось в главе 4, при масштабировании технологии важность данного компонента увеличивается.

Емкость затвора разветвления по выходу C_{g3} и C_{g4}

Мы предполагаем, что емкость разветвления по выходу равна общей емкости затвора нагружающих элементов M_3 и M_4. Следовательно,

$$C_{fan-out} = C_{gate}(NMOS) + C_{gate}(PMOS) = \\ = (C_{GSOn} + C_{GDOn} + W_n L_n C_{ox}) + (C_{GSOp} + C_{GDOp} + W_p L_p C_{ox}). \quad (5.15)$$

Приведенное выражение дважды упрощает ситуацию.

- В нем предполагается, что все компоненты емкости затвора включены между $V_{\text{вых.}}$ и GND (или V_{DD}); кроме того, игнорируется эффект Миллера емкостей, образованных затвором-стоком. На точность это влияет незначительно, поскольку мы спокойно можем предполагать, что связующий затвор не переключается, пока не будет достигнута точка 50%, следовательно, в интересующем нас интервале V_{out2} остается постоянным.
- Второе приближение заключается в том, что емкость канала связующего затвора считается постоянной в рассматриваемом интервале. Это не совсем тот случай, который мы рассматривали в главе 3. Общая емкость канала является функцией рабочего режима и изменяется примерно от $(2/3)WLC_{ox}$ (насыщение) до максимального WLC_{ox} (линейный режим и режим отсечки). Кроме того, как видно на рис. 3.31, непосредственно перед

Таблица 5.1. Параметры транзисторов инвертора

	W/L	AD (мкм^2)	PD (мкм)	AS (мкм^2)	PS (мкм^2)
n-МОП	$0,375/0,25$	$0,3(19\lambda^2)$	$1,875(15\lambda^2)$	$0,3(19\lambda^2)$	$1,875(15\lambda^2)$
p-МОП	$1,125/0,25$	$0,7(45\lambda^2)$	$2,375(19\lambda^2)$	$0,7(45\lambda^2)$	$2,375(19\lambda^2)$

включением транзистора происходит перепад емкости затвора. Во время первой половины перехода можно предполагать, что одно из нагружающих устройств находится в линейном режиме, тогда как другой транзистор переходит из выключенного состояния в режим насыщения. Если игнорировать изменение емкости, мы получим пессимистическую оценку с ошибкой, составляющей примерно 10% (для анализа в первом приближении этого вполне достаточно).

Пример 5.4. Емкости КМОП-инвертора, реализованного по 0,25-микронной технологии

С помощью 0,25-микронной КМОП-технологии был разработан симметричный КМОП-инвертор минимального размера, схема которого показана на рис. 5.15. Напряжение источника питания V_{DD} равно 2,5 В. Размеры транзисторов, области диффузии и периметры элементов, найденные по схеме, представлены в табл. 5.1. Выведем в качестве примера площадь и периметр стока n-МОП-транзистора. Область стока формируется контактом "металл-диффузионная область", площадь которого равна $4\times4\lambda^2$, и прямоугольником между контактом и затвором, площадь которого равна $3\times1\lambda^2$. В результате общая площадь составляет $19\lambda^2$, или 0,30 мкм^2 ($\lambda = 0,125$ мкм). Периметр области стока состоит из следующих компонентов (считаем против часовой стрелки): $5+4+4+1+1=15\lambda$, или $PD = 15\times0,125 = 1,875$ мкм. Обратите внимание на то, что мы не учитываем сторону периметра стока, соседствующую с затвором, поскольку она не считается частью боковой стенки. Подобным образом находятся площадь и периметр стока p-МОП-транзистора (из-за прямоугольной формы задача является элементарной): $AD = 5\times9\lambda^2 = 45\lambda^2$, или 0,7 мкм^2; $PD = 5+9+5 = 19\lambda$, или 2,375 мкм.

Приведенную физическую информацию можно объединить с выведенными ранее аппроксимациями и оценить C_L. Емкостные параметры нашего обобщенного процесса см. в табл. 3.5. Повторим их еще раз для удобства.
Перекрывающаяся емкость: $CGD0$ (n-МОП) = 0,31 фФ/мкм;
$CGDO$ (p-МОП) = 0,27 фФ/мкм.
Емкость с нижней стенкой: CJ (n-МОП) = 2 фФ/мкм^2; CJ(p-МОП) = 1,9 фФ/мкм^2.
Емкость с боковой стенкой: $CJSW$ (n-МОП) = 0,28 фФ/мкм;
CJSW(p-МОП) = 0,22 фФ/мкм.
Емкость затвора: C_{ox} (n-МОП) = C_{ox} (p-МОП) = 6 фФ/мкм^2.

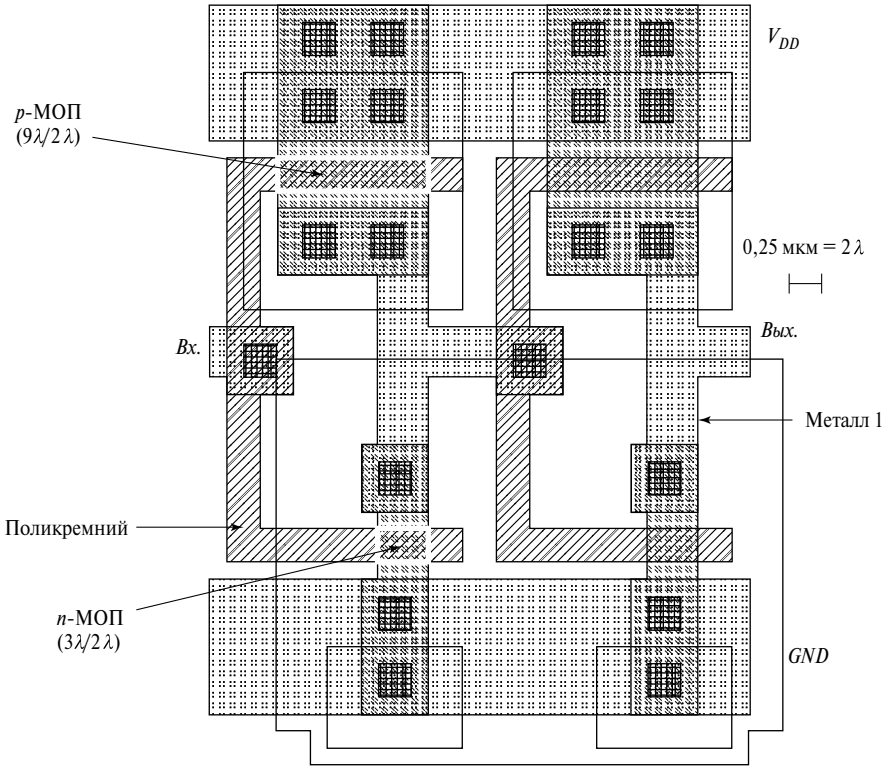


Рис. 5.15. Схема двух соединенных инверторов минимального размера, реализованных с использованием норм проектирования КМОП

Наконец, мы должны также рассмотреть емкость, порождаемую проводником, соединяющим затворы и реализованным на первом металлическом и поликремниевом уровнях. Точные значения данных паразитных емкостей обычно находятся с помощью программы моделирования. Изучая представленную схему, мы можем оценить параметры в первом приближении. Получается, что области проводника (металл-1 и поликремний), не участвующие в активной диффузии, равны $42\lambda^2$ и $72\lambda^2$ соответственно. Используя параметры межсоединения, приведенные в табл. 4.2, находим, что емкость проводника равна (обратите внимание на то, что в данном простом примере мы игнорируем краевую емкость; из-за небольшой длины провода ее вклад мал по сравнению с другими составляющими):

$$C_{wire} = 42/8^2 \text{ мкм}^2 \times 30 \text{ аФ/мкм}^2 + 72/8^2 \text{ мкм}^2 \times 88 \text{ аF/мкм}^2 = 0,12 \text{ фФ}.$$

Все полученные компоненты сведены в табл. 5.2. Для расчета диффузионных емкостей мы использовали значения K_{eq}, полученные в примере 5.3. Обратите внимание на то, что емкость нагрузки практически равномерно распределена между двумя основными компонентами: внутренней емкостью,

Таблица 5.2. Компоненты C_L (при изменении уровня сигнала от высокого к низкому и от низкого к высокому)

Конденсатор	Формула	Значение (фФ) (H → L)	Значение (фФ) (L → H)
C_{gd1}	$2CGD0_nW_n$	$0,23$	$0,23$
C_{gd2}	$2CGD0_pW_p$	$0,61$	$0,61$
C_{db1}	$K_{eqn}AD_nCJ+$ $K_{eqswn}PD_nCJSW$	$0,66$	$0,90$
C_{db2}	$K_{eqp}AD_pCJ+$ $K_{eqswp}PD_pCJSW$	$1,5$	$1,15$
C_{g3}	$(CGD0_n + CGSO_n)W_n+$ $C_{ox}W_nL_n$	$0,76$	$0,76$
C_{g4}	$(CGD0_p + CGSO_p)W_p+$ $C_{ox}W_pL_p$	$2,28$	$2,28$
C_w	По результатам моделирования	$0,12$	$0,12$
C_L	$\sum$	$6,1$	$6,0$

образованной диффузионной и перекрывающейся емкостью, и внешней нагрузкой, в которой учитывается вклад проводника и соединяющего затвора.

5.4.2. Задержка распространения: анализ в первом приближении

Чтобы рассчитать задержку распространения инвертора, можно проинтегрировать ток зарядки (разрядки) конденсатора. В результате мы получим следующее выражение:

$$t_p = \int_{\nu_1}^{\nu_2} \frac{C_L(\nu)}{i(\nu)} d\nu, \tag{5.16}$$

где i — зарядный (разрядный) ток; ν — напряжение, приложенное к конденсатору; ν_1 и ν_2 — исходное и конечное напряжения соответственно. Точный расчет по этой формуле затруднителен, поскольку и $C_L(\nu)$, и $i(\nu)$ являются нелинейными функциями ν. Поэтому мы обратимся к упрощенной модели инвертора в виде ключа (см. рис. 5.6) и выведем разумную оценку задержки распространения, приемлемую для ручного анализа. Зависимость от напряжения сопротивления во включенном состоянии и нагружающего конденсатора находится путем их замены постоянным линейным элементом, значение которого усреднено по интересующему нас интервалу. Именно это значение нагружающего конденсатора было рассчитано в предыдущем разделе. Выражение для среднего сопротивления МОП-транзистора во включенном состоянии было выведено в примере 3.8; ниже мы его повторяем для удобства.

$$R_{eq} = \frac{1}{V_{DD}/2} \int_{V_{DD}/2}^{V_{DD}} \frac{V}{I_{DSAT}(1+\lambda V)} dV \approx \frac{3}{4} \frac{V_{DD}}{I_{DSAT}} \left(1 - \frac{7}{9}\lambda\ V_{DD}\right),$$

$$\text{где } I_{DSAT} = k' \frac{W}{L} \left((V_{DD} - V_T)V_{DSAT} - \frac{V_{DSAT}^2}{2}\right) \quad (5.17)$$

Теперь расчет задержки распространения получающейся схемы достаточно прост — все свелось к анализу RC-цепи первого порядка, идентичной рассмотренной в примере 4.5. Тогда мы нашли, что задержка распространения такой сети, на вход которой подан скачок напряжения, пропорциональна постоянной времени сети, образованной резистором утечки (согласующим выходным резистором) и емкостью нагрузки. Следовательно,

$$t_{pHL} = \ln(2) R_{eqn} C_L = 0,69 R_{eqn} C_L. \quad (5.18)$$

Подобным образом мы можем получить задержку распространения для изменения уровня сигнала от низкого к высокому. Имеем:

$$t_{pLH} = 0,69 R_{eqp} C_L, \quad (5.19)$$

где R_{eqp} — эквивалентное сопротивление p-МОП-транзистора во включенном состоянии в интересующем нас интервале. В данном случае мы предполагаем, что эквивалентная нагрузочная емкость идентична для изменения уровня сигнала от низкого к высокому и от высокого к низкому (это было показано в примере в предыдущем разделе). Общая задержка распространения инвертора определяется как среднее двух значений:

$$t_p = \frac{t_{pHL} + t_{pLH}}{2} = 0,69 C_L \left(\frac{R_{eqn} + R_{eqp}}{2}\right). \quad (5.20)$$

Очень часто бывает желательно, чтобы затвор имел одинаковые задержки распространения для нарастающего и спадающего входа. Данному условию можно удовлетворить, сделав приблизительно равными сопротивления n-МОП- и p-МОП-транзисторов во включенном состоянии. Помните, что данное условие идентично требованию наличия симметричной передаточной функции напряжения.

Пример 5.5. Задержка распространения КМОП-инвертора, реализованного по 0,25-микронной технологии

Чтобы рассчитать задержки распространения КМОП-инвертора, представленного на рис. 5.15, используем уравнения (5.18) и (5.19). Нагрузочная емкость C_L уже была рассчитана в примере 5.4, а эквивалентные сопротивления транзисторов во включенном состоянии для общего 0,25-микронного

КМОП-процесса были представлены в табл. 3.3. При напряжении источника питания в 2,5 В нормированные сопротивления n-МОП- и p-МОП-транзисторов во включенном состоянии равны 13 и 31 кОм соответственно. По схеме мы можем определить отношения (W/L) транзисторов — 1,5 для n-МОП- и 4,5 — для p-МОП-транзистора. Мы предполагаем, что разница между изображенными и эффективными размерами настолько мала, что ее можно проигнорировать. В результате получаем следующее значения задержек:

$$t_{pHL} = 0,69 \times \left(\frac{13 \text{ кОм}}{1,5}\right) \times 6,1 \text{ фФ} = 36 \text{ пс};$$
$$t_{pLH} = 0,69 \times \left(\frac{13 \text{ кОм}}{4,5}\right) \times 6,0 \text{ фФ} = 29 \text{ пс}$$

и

$$t_p = \left(\frac{36 + 29}{2}\right) = 32,5 \text{ пс}.$$

Точность данного анализа проверяется путем моделирования средствами SPICE схемы цепи, сформированной согласно рис. 5.15. Рассчитанная переходная характеристика цепи изображена на рис. 5.16; из нее следует, что задержки распространения равны 39,9 пс (изменение уровня сигнала от высокого к низкому, HL) и 31,7 пс (противоположный переход, LH). Результаты, полученные вручную, хороши, если учитывать огромное число сделанных упрощений. В частности, обратите внимание на перегрузки в смоделированных выходных сигналах. Они вызваны емкостями затвора-стока транзисторов инвертора, которые соединяют крутой скачок напряжения на входе непосредственно с выходом (до того, как транзисторы хотя бы начнут реагировать на изменения входа). Очевидно, подобные перегрузки отрицательно влияют на быстродействие элемента и объясняют, почему смоделированные задержки больше оценок.

Предупреждение

Возможно, после данного примера вам кажется, что ручной анализ всегда дает близкие оценки реальной характеристики. Не заблуждайтесь, так бывает не всегда! Модели первого (да и более высоких) порядков точности могут давать результаты, сильно отклоняющиеся от реальных. Задача ручного анализа — почувствовать поведение схемы и определить основные параметры. Когда требуются количественные данные, необходимо подробное моделирование. Рассматривайте предыдущий пример просто как случайную удачу.

Как правило, к этому моменту у разработчиков уже возникает вопрос, как можно повлиять на задержку элемента (или оптимизировать ее). Чтобы ответить на него, необходимо явно выделить параметры, управляющие за-

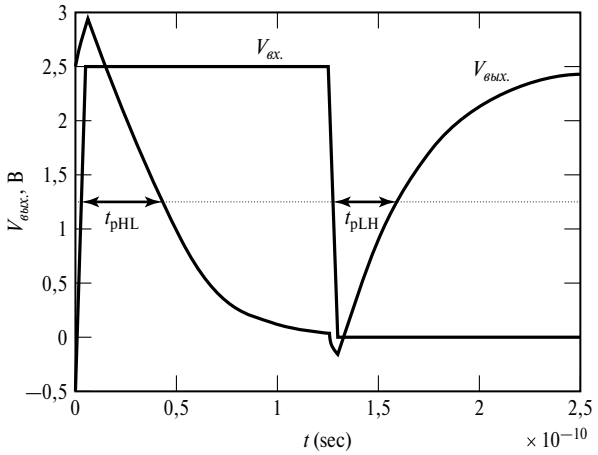


Рис. 5.16. Смоделированная переходная характеристика инвертора, представленного на рис. 5.15

держкой, расписав R_{eq} в уравнении задержки. Объединяя уравнение (5.18) и уравнение (5.17) и предполагая (пока), что коэффициент модуляции длины канала λ можно игнорировать, получаем следующее выражение для t_{pHL} (аналогичный анализ можно выполнить и для t_{pLH}):

$$t_{pHL} = 0,69\frac{3}{4}\frac{C_L V_{DD}}{I_{DSATn}} = 0,52\frac{C_L V_{DD}}{(W/L)_n k'_n V_{DSATn}(V_{DD} - V_{Tn} - V_{DSATn}/2)}. \quad (5.21)$$

В большей части разработок напряжение источника питания выбирается достаточно большим, чтобы $V_{DD} \gg V_{Tn} + V_{DSATn}/2$. При таких условиях задержка становится практически независящей от напряжения источника питания:

$$t_{pHL} \approx 0,52\frac{C_L}{(W/L)_n k'_n V_{DSATn}}. \quad (5.22)$$

Мы используем первое приближение и видим, что увеличение напряжения источника питания улучшает (хотя и незначительно) быстродействие из-за ненулевого коэффициента модуляции длины канала. Данный анализ подтверждается рис. 5.17, где задержка распространения инвертора изображена как функция напряжения источника питания. После всего сказанного не удивляет, что данная кривая практически идентична форме, представленной на рис. 3.28, где эквивалентное сопротивления МОП-транзистора во включенном состоянии изображено как функция V_{DD}. Хотя задержка относительно нечувствительна к изменениям напряжения для бо́льших значений V_{DD}, начиная примерно с $\approx 2V_T$, наблюдаем резкий максимум. Очевидно, что этой рабочей области следует избегать, если основной целью разработки является получения высокого быстродействия.

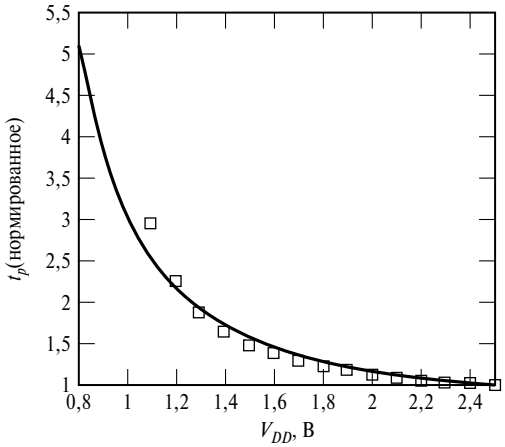


Рис. 5.17. Задержка распространения КМОП-инвертора как функция напряжения источника питания (нормирована относительно задержки при 2,5 В). Квадратиками показаны значения задержки, предсказанные уравнением (5.21). Обратите внимание на то, что данное уравнение справедливо только при насыщении скорости (отсюда и отклонения при низких напряжениях источника питания)

Методология проектирования

Из сказанного выше можно сделать вывод, что задержку распространения элемента можно минимизировать с помощью перечисленных ниже способов.

- *Уменьшить* C_L. Помните, что на емкость нагрузки влияют три основных фактора: внутренняя диффузионная емкость самого элемента, емкость межсоединений и разветвление по выходу. За счет тщательно спроектированной схемы диффузионную емкость и емкость межсоединений можно уменьшить. **Диффузионные области стока принято делать насколько возможно маленькими.**
- *Увеличить отношение* W/L *транзистора.* Данный инструмент оптимизации быстродействия является наиболее мощным и эффективным из всех, находящихся в распоряжении разработчика. Тем не менее, применяя данный подход, следует быть очень осторожным. Увеличение размера транзистора также приводит увеличению диффузионной емкости, а следовательно, и C_L. Фактически, после того как внутренняя емкость (т.е. диффузионная емкость) превысит внешнюю нагрузку, сформированную проводкой и разветвлением по выходу, увеличение размера элемента уже не будет уменьшать задержку (будет просто увеличиваться площадь элемента). Данный эффект называется *самонагрузкой*. Кроме того, широкие транзисторы обладают большей емкостью затвора, что увеличивает коэффициент разветвления по выходу питающего элемента и неблагоприятно влияет на его скорость.

- *Увеличить* V_{DD}. Как показано на рис. 5.17, задержку элемента можно регулировать, изменяя напряжение источника питания. Такая возможность позволяет разработчику повышать быстродействие за счет рассеяния энергии, как будет показано в следующем разделе. Тем не менее увеличение напряжения источника питания выше определенного предела дает лишь минимальное улучшение, а следовательно, его следует избегать. Кроме того, существует напряжение, выше которого в субмикронных процессах подниматься нецелесообразно из соображений надежности (пробой оксида, эффект горячих электронов).

Задача 5.4. Задержка распространения как функция тока зарядки/разрядки

До этого момента задержку распространения мы выражали как функцию эквивалентного сопротивления транзисторов. Существует и другой подход: заменить транзистор источником тока, сила тока которого равна среднему току зарядки/разрядки в интересующем нас интервале. Используя описанный альтернативный подход, выведите формулу для задержки распространения.

5.4.3. Задержка распространения с точки зрения разработки

Изучая выведенные формулы для задержки распространения, можно сделать несколько интересных выводов и наметить ряд компромиссов. Более того, с помощью этих формул можно разработать общий (невероятно полезный) подход к выбору размеров транзисторов.

Характеристическое отношение размеров n-МОП- и p-МОП-транзисторов

До этого момента мы последовательно увеличивали ширину p-МОП-транзистора, чтобы его сопротивление согласовывалось с сопротивлением выходного n-МОП-устройства. Как правило, для этого требуется, чтобы отношение ширины p-МОП-транзистора к ширине n-МОП-транзистора составляло 3–3,5. Смысл подобной конфигурации заключается в том, чтобы создать инвертор с симметричной передаточной функцией по напряжению и сравнять задержки распространения при изменении уровня сигнала от высокого к низкому и от низкого к высокому. Однако из этого не следует, что такое соотношение даст минимальную общую задержку распространения. Если вопросы симметрии и уменьшения запаса помехоустойчивости не являются первостепенными, мы можем ускорить инвертор, уменьшив ширину p-МОП-транзистора!

Смысл сказанного состоит в том, что хотя увеличении ширины p-МОП-транзистора улучшает t_{pLH} инвертора путем увеличения тока зарядки, оно также ухудшает t_{pHL}, вводя бо́льшую паразитную емкость. При наличии двух

противоречивых эффектов должно существовать отношение размеров транзисторов, оптимизирующее задержку распространения инвертора. Данное оптимальное соотношение можно найти, используя простой метод анализа. Рассмотрим два идентичных каскадных КМОП-инвертора. Приблизительная емкость нагрузки первого элемента выражается следующим образом:

$$C_L = (C_{dp1} + C_{dn1}) + (C_{gp2} + C_{gn2}) + C_W, \tag{5.23}$$

где C_{dp1} и C_{dn1} — эквивалентные диффузионные емкости стоков p-МОП- и n-МОП-транзисторов первого инвертора; C_{gp2} и C_{gn2} — емкости затворов второго инвертора; C_W — емкость разводки.

Когда p-МОП-устройства в β раз больше n-МОП-транзисторов ($\beta = (W/L)_p/(W/L)_n$), все емкости транзисторов масштабируются примерно одинаково ($C_{dp1} \approx \beta C_{dn1}$ и $C_{gp2} \approx \beta C_{gn2}$). Таким образом, в данном случае уравнение (5.23) можно переписать следующим образом:

$$C_L = (1 + \beta)(C_{dn1} + C_{gn2}) + C_W. \tag{5.24}$$

Отталкиваясь от уравнения (5.20), можно вывести следующее выражение для задержки распространения:

$$\begin{aligned} t_p &= \frac{0,69}{2}((1 + \beta)(C_{dn1} + C_{gn2}) + C_W)\left(R_{eqn} + \frac{R_{eqn}}{\beta}\right) = \\ &= 0,345((1 + \beta)(C_{dn1} + C_{gn2}) + C_W)R_{eqn}\left(1 + \frac{r}{\beta}\right), \end{aligned} \tag{5.25}$$

где $r(= R_{eqp}/R_{eqn})$ представляет отношение сопротивлений p-МОП- и n-МОП-транзисторов одинакового размера. Чтобы найти оптимальное значение β, необходимо приравнять к нулю $\partial t_p/\partial\beta$, откуда следует, что

$$\beta_{opt} = \sqrt{r\left(1 + \frac{C_W}{C_{dn1} + C_{gn2}}\right)}. \tag{5.26}$$

Последняя формула означает, что, когда емкостью разводки можно пренебречь ($C_{dn1} + C_{gn2} \gg C_W$), β_{opt} равно $\sqrt{r}$, тогда как в некаскадном случае это значение равно r. Если емкость разводки доминирует, необходимо использовать бо́льшие значения β. Таким образом, можно сделать вывод, что меньшие размеры устройств (а следовательно, меньшая площадь) позволяют получить более быструю структуру за счет симметрии и запаса помехоустойчивости.

Пример 5.6. Выбор размеров КМОП-инвертора по идентичному затвору

Рассмотрим еще раз пример стандартной разработки. Используя значения эквивалентных сопротивлений (см. табл. 3.3), находим, что значение

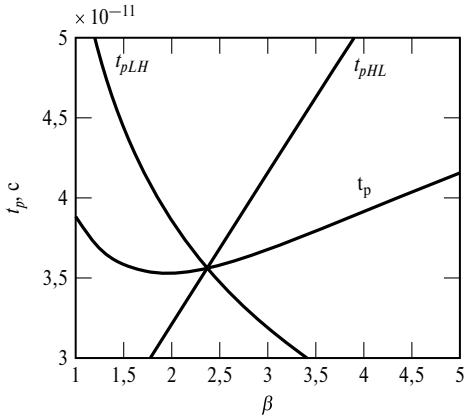


Рис. 5.18. Задержка распространения КМОП-инвертора как функция отношения β размеров p-МОП-транзистора к n-МОП-транзистору

$\beta = 2,4$ (= 31 кОм/13 кОм) дает симметричную переходную характеристику. Далее из уравнения (5.26) следует, что отношение размеров устройства, дающее оптимальное быстродействие, должно быть равно 1,6. Данные результаты подтверждаются рис. 5.18, где изображена смоделированная задержка распространения как функция параметра β. На графике отчетливо видно, как изменение β позволяет находить компромисс между t_{pLH} и t_{pHL}. Оптимальная точка соответствует примерно значению $\beta = 1,9$, что несколько больше, чем предсказывалось. Кроме того, видно, что задержки нарастания и спада идентичны в предсказанной точке $\beta = 2,4$. Данная рабочая точка является оптимальным выбором, когда главным требованием к разработке является минимальная задержка в худшем случае[3].

Выбор размеров инверторов для получения максимального быстродействия

В данном анализе мы будем рассматривать *симметричный инвертор*, размеры p-МОП- и n-МОП-транзисторов которого выбраны так, чтобы задержки нарастания и спада были идентичными. Емкость нагрузки инвертора можно разделить на внутренний и внешний компонент, или $C_L = C_{int} + C_{ext}$. Здесь C_{int} представляет самонагружающую или внутреннюю выходную емкость инвертора и соотносится с диффузионными емкостями n-МОП- и p-МОП-транзисторов, а также с емкостью перекрытия затвора-стока (Миллера). C_{ext} — это внешняя емкость нагрузки, приписываемая емкостям разводки

[3] Возможно, вам интересно, почему мы не всегда рассматриваем худший вариант задержек нарастания и спада как основную меру быстродействия. Дело в том, что при каскадном объединении инвертирующих элементов в более сложную логическую сеть становится понятно, что более содержательной мерой является среднее двух указанных задержек (за нарастающим переходом одного элемента следует спадающий переход другого).

и разветвления по выходу. Обозначив через R_{eq} эквивалентное сопротивление элемента, задержку распространения можно выразить следующим образом:

$$\begin{aligned} t_p &= 0,69 R_{eq}(C_{int} + C_{ext}) = \\ &= 0,69 R_{eq} C_{int}(1 + C_{ext}/C_{int}) = t_{p0}(1 + C_{ext}/C_{int}), \end{aligned} \tag{5.27}$$

где $t_{p0} = 0,69 R_{eq} C_{int}$ представляет задержку инвертора, нагруженного только внутренней емкостью ($C_{ext} = 0$), и называется *внутренней*, *собственной* или *ненагруженной задержкой*.

Далее необходимо разобраться, как изменение размеров транзистора влияет на быстродействие элемента. Чтобы ответить на этот вопрос, необходимо установить связь между различными параметрами в уравнении (5.27) и масштабным коэффициентом S, связывающим размеры транзисторов нашего инвертора с размерами некоторого эталонного элемента (обычно инвертора минимального размера). Внутренняя емкость C_{int} состоит из диффузионной емкости и емкости Миллера (обе пропорциональны ширине транзисторов). Следовательно, $C_{int} = SC_{iref}$. Сопротивление элемента соотносится с сопротивлением эталонного элемента как $R_{eq} = R_{ref}/S$. Теперь уравнение (5.27) можно переписать в следующем виде:

$$\begin{aligned} t_p &= 0,69(R_{ref}/S)(SC_{iref})(1 + C_{ext}/(SC_{iref})) = \\ &= 0,69 R_{ref} C_{iref} \left(1 + \frac{C_{ext}}{SC_{iref}}\right) = t_{p0} \left(1 + \frac{C_{ext}}{SC_{iref}}\right). \end{aligned} \tag{5.28}$$

Из выполненного анализа следуют два важных вывода.

- Внутренняя задержка инвертора t_{p0} не зависит от масштабирования размеров элемента и определяется исключительно используемой технологией и схемой инвертора. При отсутствии нагрузки увеличение скорости работы элемента полностью нейтрализуется увеличившейся емкостью.
- Бесконечно увеличивая S, получаем максимальное повышение быстродействия, перекрывающее влияние любой внешней нагрузки и уменьшающее задержку до уровня внутренней. Помимо этого, любой масштабный коэффициент S, достаточно превышающий (C_{ext}/C_{int}), дает похожие результаты при значительном усилении.

Пример 5.7. Масштабирование устройства для получения оптимального быстродействия

Рассмотрим улучшение быстродействия, которое можно получить путем пропорционального уменьшения размеров прибора, схема которого использовалась в примере 5.5. Из табл. 5.2 находим, что $C_{int}/C_{ext} \approx 1,05$ ($C_{int} = 3,0$ фФ, $C_{ext} = 3,15$ фФ). Отсюда следует, что максимальное усиление быстродействия равно $2,05$. Используя масштабный коэффициент 10, мы получим порядка 10% данного оптимального быстродействия, тогда как

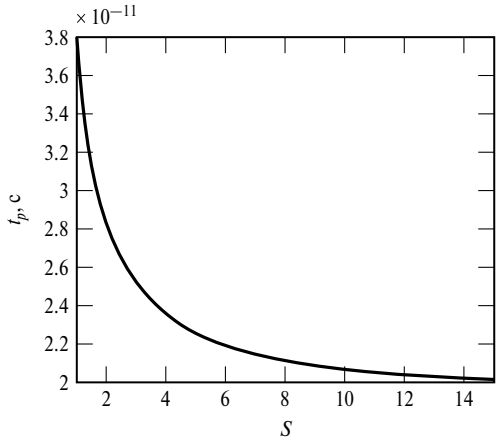


Рис. 5.19. Увеличение быстродействия инвертора путем масштабирования n-МОП- и p-МОП-транзисторов с одинаковым коэффициентом S при фиксированной нагрузочной способности на выходе (инвертор, представленный на рис. 5.15)

б*о*льшие размеры устройств дадут пренебрежимо малое повышение быстродействие.

Сказанное подтверждается результатами моделирования, которые предсказывают максимальное улучшение быстродействия, равное 1,9 ($t_{p0} = 19,3$ пс). Изучая график, представленный на рис. 5.19, видим, что значительная доля данного улучшения достигается уже при $S = 5$, а при использовании масштабных коэффициентов, б*о*льших 10, дополнительное усиление практически не заметно.

Масштабирование цепочки инверторов

Хотя увеличение размера инвертора уменьшает его задержку, оно также увеличивает его входную емкость. Само по себе масштабирование элемента без учета его влияния на задержку предыдущих элементов представляет исключительно академический интерес. Следовательно, гораздо важнее определить оптимальный размер элемента, **внедренного в реальную среду**. В качестве хорошего начального примера можно рассмотреть простую цепочку инверторов. Чтобы определить нагрузку на входе, необходимо вывести соотношение между емкостью элемента на входе C_g и внутренней выходной емкостью инвертора. Обе указанные величины пропорциональны размеру элемента. Следовательно, независимо от этого размера, справедливым будет следующее соотношение:

$$C_{int} = \gamma C_g. \tag{5.29}$$

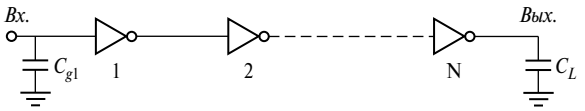


Рис. 5.20. Цепочка из N инверторов с фиксированной емкостью на входе и выходе

В уравнении (5.29) γ — это коэффициент пропорциональности, зависящий только от выбранной технологии и, как мы видели в предыдущих примерах, близкий к 1 для большинства субмикронных процессов. Переписывая уравнение (5.28), получаем

$$t_p = t_{p0}\left(1 + \frac{C_{ext}}{\gamma C_g}\right) = t_{p0}(1 + f/\gamma). \tag{5.30}$$

Таким образом, задержка инвертора является **функцией только отношения между его внешней емкостью нагрузки и внутренней емкостью**. Данное отношение называется *эффективным коэффициентом разветвления по выходу* f.

Рассмотрим схему, изображенную на рис. 5.20. Наша цель — минимизировать задержку распространения сигнала по цепочке инверторов, если входная емкость первого инвертора (обычно элемент минимального размера) равна C_{g1}, а емкость нагрузки на конце цепочки фиксирована и равна C_L.

Используя выражение для задержки j-го инвертора каскада[4]

$$t_{p,j} = t_{p0}\left(1 + \frac{C_{g,j+1}}{\gamma C_{g,j}}\right) = t_{p0}(1 + f_j/\gamma), \tag{5.31}$$

можем вывести общую задержку цепочки:

$$t_p = \sum_{j=1}^{N} t_{p,j} = t_{p0}\sum_{j=1}^{N}\left(1 + \frac{C_{g,j+1}}{\gamma C_{g,j}}\right), \text{ где } C_{g,N+1} = C_L. \tag{5.32}$$

Данное уравнение содержит $N-1$ неизвестных: $C_{g,2}, C_{g,3}, \ldots, C_{g,N}$. Чтобы найти минимальную задержку, необходимо взять $N-1$ частных производных и приравнять их к нулю ($\partial t_p/\partial C_{g,i} = 0$). В результате получаем следующий набор условий:

$$C_{g,j+1}/C_{g,j} = C_{g,j}/C_{g,j-1}, \text{ где } (j = 2 \ldots N). \tag{5.33}$$

Другими словами, оптимальный размер каждого инвертора является геометрическим средним размеров соседних инверторов:

[4] Здесь мы игнорируем емкость разводки, что на данный момент является достаточно оправданным допущением.

$$C_{g,j} = \sqrt{C_{g,j-1}C_{g,j+1}}. \tag{5.34}$$

Это означает, что каждый инвертор масштабируется с коэффициентом f относительно предыдущего элемента, имеет такое же эффективное разветвление по выходу ($f_j = f$), а следовательно, и такую же задержку. Зная $C_{g,1}$ и C_L, можем рассчитать масштабный коэффициент:

$$f = \sqrt[N]{C_L/C_{g,1}} = \sqrt[N]{F} \tag{5.35}$$

и минимальную задержку распространения сигнала по цепочке:

$$t_p = Nt_{p0}(1 + \sqrt[N]{F}/\gamma). \tag{5.36}$$

Здесь F представляет собой *общее эффективное разветвление по выходу* схемы и имеет значение $C_L/C_{g,1}$. Обратите внимание на то, что связь между t_p и F сильно зависит от числа каскадов. Как и ожидалось, при наличии одного каскада зависимость линейна. Введение второго каскада превращает ее в зависимость вида "квадратный корень из F" и т.д. Далее возникает естественный вопрос, как выбрать число каскадов, чтобы минимизировать задержку для данного значения F.

Выбор правильного числа каскадов цепочки инверторов

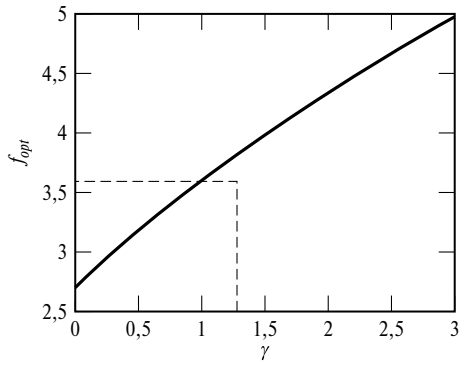
Изучая уравнение (5.36), мы можем подобрать для данного F ($= f^N$) наиболее подходящее число каскадов. Когда число каскадов слишком велико, доминирует первый компонент уравнения, представляющий внутреннюю задержку каскадов. Когда число каскадов мало, большим становится эффективное разветвление по выходу и доминирует уже второй компонент. Чтобы найти оптимальное значение, необходимо продифференцировать выражение для минимальной задержки по числу каскадов и приравнять результат к нулю. Сделав это, получаем:

$$\gamma + \sqrt[N]{F} - \frac{\sqrt[N]{F}\ln F}{N} = 0 \tag{5.37}$$

или

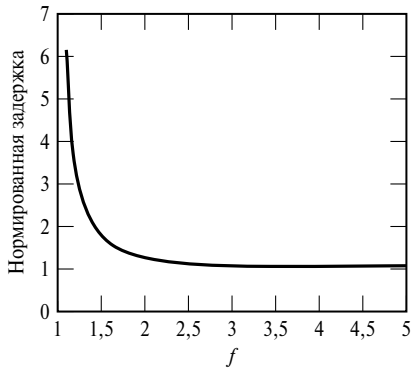
$$f = e^{(1+\gamma/f)}.$$

Аналитическое решение уравнения (5.35) существует только при $\gamma = 0$, т.е. тогда, когда мы игнорируем самонагрузку и емкость нагрузки определяется только разветвлением по выходу. При таких упрощениях находим, что оптимальное число каскадов равно $N = \ln(F)$ и эффективное разветвление по выходу каждого каскада равно $f = e = 2,71828$. Таким образом, при данной схеме последовательные каскады масштабируются по экспоненте [Mead80].



Оптимальное эффективное разветвление по выходу (или масштабный коэффициент инвертора) как функция коэффициента самозагрузки γ в цепи инвертора

а



Нормированная задержка распространения как функция эффективного разветвления по выходу f для $\gamma = 1$

б

Рис. 5.21. Оптимизация числа каскадов в цепочке инверторов

Если учитывать самонагрузку, уравнение можно будет решить только численно. Результаты соответствующих расчетов приведены на рис. 5.21, *а*. Для типичного случая $\gamma \approx 1$ оптимальный коэффициент масштабирования близок к 3,6. На рис. 5.21, *б* для $\gamma = 1$ изображена (нормированная) задержка распространения цепочки инверторов как функция эффективного разветвления по выходу. Выбор значений разветвления по выходу, б*ольших* оптимального значения, не сильно влияет на задержку и уменьшает требуемое число буферных каскадов и общую площадь реализации. Как правило, **оптимальное значение разветвления по выходу выбирается равным 4**. С другой стороны, использование слишком большого числа каскадов ($f < f_{opt}$) крайне негативно сказывается на задержке, а следовательно, этого рекомендуется избегать.

Пример 5.8. Влияние добавочных буферных каскадов

В табл. 5.3 для различных значений F ($\gamma = 1$) перечислены значения $t_{p,opt}/t_{p0}$ для структуры без буфера, двухкаскадной схемы и оптимизированной цепочки инверторов. Обратите внимание на впечатляющее ускорение, которое дают выстроенные каскадом инверторы при очень больших емкостях нагрузки.

Приведенный выше анализ можно расширить не только на цепочки инверторов, но и на сети инверторов, содержащие реальные разветвления по выходу (один из примеров такой сети показан на рис. 5.22). Все, что нам для этого потребуется, — это переписать выражение для C_{ext} с учетом дополнительных разветвлений по выходу.

Таблица 5.3. Зависимость отношения t_{opt}/t_{p0} от x для различных конфигураций цепочки

F	Без буфера	Два каскада	Цепочка инверторов
10	11	8,3	8,3
100	101	22	16,5
1000	1001	65	24,8
10 000	10 001	202	33,1

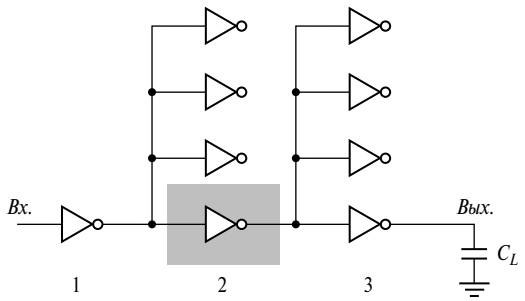


Рис. 5.22. Сеть инверторов, в которой каждый элемент имеет разветвление по выходу на 4 элемента (сигнал, поданный на вход, проходит через древовидную структуру, порождая на выходе 16 сигналов)

Задача 5.5. Масштабирование сети инверторов

Определите размеры инверторов цепи, изображенной на рис. 5.22, минимизировав задержку между узлами *Вых.* и *Вх.* Считайте, что $C_L = 64C_{g,1}$.

Подсказки. Вначале определите соотношение размеров устройств, при котором задержка минимальна. У вас должна получиться следующая формула:

$$\frac{4C_{g,2}}{C_{g,1}} = \frac{4C_{g,3}}{C_{g,2}} = \frac{C_L}{C_{g,3}}.$$

Как найти реальные размеры элементов ($C_{g,3} = 2,52C_{g,2} = 6,35C_{g,1}$, где $2,52 = 16^{1/3}$), понятно. Если масштабировать цепочку инверторов "в лоб" без учета дополнительного разветвления по выходу, масштабный коэффициент будет равен не 2,52, а 4.

Время нарастания/спада входного сигнала

Все представленные выше выражения выводились в предположении, что входной сигнал инвертора резко меняется с 0 до V_{DD} или наоборот. Предполагается, что в любой момент процесса зарядки/разрядки всегда включено

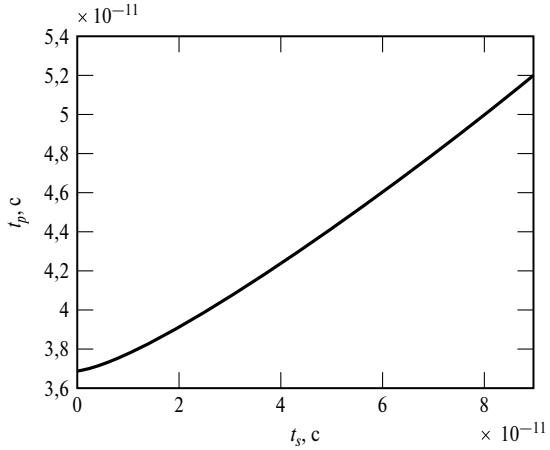


Рис. 5.23. Задержка распространения t_p как функция огибающей входного сигнала (10–90% времени нарастания или спада) для инвертора минимального размера с нагрузкой в виде одного элемента

одно устройство. В реальной жизни входной сигнал изменяется постепенно и в определенный промежуток времени p-МОП- и n-МОП-транзисторы вместе проводят ток. Это влияет на общий ток, доступный для зарядки/разрядки, и задержку распространения. На рис. 5.23 задержка распространения инвертора минимального размера изображена как функция наклона входного сигнала (получено с помощью SPICE). На графике видно, что после момента $t_s > t_p$ ($t_s = 0$) время t_p увеличивается (приблизительно) линейно с увеличением наклона.

Хотя мы можем вывести аналитическое выражение, описывающее связь между огибающей входного сигнала и задержкой распространения, это потребует значительных усилий, а практическая ценность полученного результата будет небольшой. С точки зрения дизайна гораздо более ценным будет связать непосредственное влияние конечной огибающей на быстродействие в нашем случае — при ограниченной питающей способности предыдущего элемента. Если последняя будет иметь бесконечные возможности, время нарастания/спада сигнала на выходе будет нулевым и не будет влиять на быстродействие рассматриваемого элемента. Прелесть данного подхода заключается в том, что он позволяет уяснить: *никакой элемент не разрабатывается сам по себе, на быстродействие влияют как разветвление по выходу, так и питающие возможности элементов, обеспечивающих вход для данного.* В результате мы возвращаемся к выражению для задержки распространения инвертора i в цепочке инверторов [Hedenstierna87]:

$$t_p^i = t_{step}^i + \eta t_{step}^{i-1}. \tag{5.38}$$

В уравнении (5.38) утверждается, что задержка распространения инвертора i равна задержке того же самого элемента с импульсным входом ($t_s^i tep$) (т.е. бесконечным наклоном сигнала на входе), дополненной долей задержки предыдущего $(i-1)$ элемента при импульсном входе. Доля η является эмпирической константой, типичное значение которой составляет порядка 0,25. Достоинством приведенного выражения является его простота, несмотря на которую оно представляет все соотношения, необходимые для расчета задержек любых сложных схем.

Пример 5.9. Задержка инвертора, внедренного в сеть

Рассмотрим схему, изображенную на рис. 5.22. С помощью уравнений (5.31) и (5.38) мы можем вывести выражения для задержки инвертора каскада 2, отмеченного серым прямоугольником:

$$t_{p,2} = t_{p0}\left(1 + \frac{4C_{g,3}}{\gamma C_{g,2}}\right) + \eta t_{p0}\left(1 + \frac{4C_{g,2}}{\gamma C_{g,1}}\right).$$

Анализируя общую задержку распространения так, как в задаче 5.5, получаем следующее требование к размерам, дающим минимальную задержку:

$$\frac{4(1+\eta)C_{g,2}}{C_{g,1}} = \frac{4(1+\eta)C_{g,3}}{C_{g,2}} = \frac{C_L}{C_{g,3}}.$$

Если предположить, что $\eta = 0,25$, f_2 и f_1 будут равны 2,47.

Вопрос проектирования

На практике выгодно, чтобы время нарастания сигнала было меньшим или равным задержкам распространения элементов. Причем сказанное справедливо не только с точки зрения быстродействия, но и с точки зрения потребляемой мощности, как будет показано ниже. Одной из основных задач разработки, нацеленной на получение быстродействующих схем, является получение времен нарастания и спада сигнала, имеющих приблизительно равные малые значения; часто данная задача именуется *проектированием наклона* (slope engineering).

Задача 5.6. Влияние наклона входного сигнала

Определите, как уменьшение напряжения питания сказывается на влиянии наклона входного сигнала на задержку распространения (усиливает или ослабляет). Аргументируйте свой ответ.

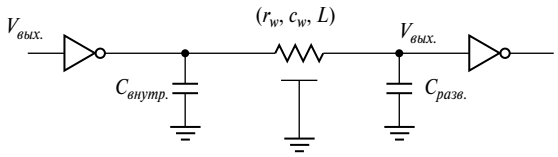


Рис. 5.24. Инвертор, питающий единственный элемент через провод длиной L

Задержка при наличии (длинных) соединительных проводов

До этого момента соединительный провод мы практически не рассматривали. Тем не менее, когда элементы удаляются друг от друга, емкость и сопротивление проводника игнорировать уже нельзя, поскольку теперь они являются факторами, определяющими переходную характеристику. Выведенные ранее выражения для задержки можно переписать с учетом данных дополнительных вкладов, используя технологии моделирования, представленные в предыдущей главе. Анализ, подробно расписанный в примере 4.9, применим и к нашей текущей задаче. Рассмотрим схему, представленную на рис. 5.24, где инвертор питает единственную выходную нагрузку через проводник длиной L. Питающий элемент представлен единственным сопротивлением R_{dr}, представляющим собой среднее между R_{eqn} и R_{eqp}. В C_{int} и C_{fan} учитывается внутренняя емкость питающего элемента и входная емкость элемента с разветвлением на выходе соответственно.

Задержку распространения схемы можно найти, использовав формулу Элмора:

$$\begin{aligned} t_p &= 0,69R_{dr}C_{int} + (0,69R_{dr} + 0,38R_w)C_w + 0,69(R_{dr} + R_w)C_{fan} = \\ &= 0,69R_{dr}(C_{int} + C_{fan}) + 0,69(R_{dr}c_w + r_wC_{fan})L + 0,38r_wc_wL^2. \end{aligned} \tag{5.39}$$

Коэффициент 0,38 позволяет учесть тот факт, что провод представляет распределенную задержку. C_w и R_w обозначают общую емкость и сопротивление провода соответственно. Выражение для задержки содержит компонент, линейно зависящий от длины провода, а также квадратичный компонент. Именно из-за этого квадратичного компонента задержка, порождаемая проводом, быстро становится основным фактором, определяющим задержку длинных проводов.

Пример 5.10. Задержка инвертора при наличии межсоединения

Рассмотрим схему, представленную на рис. 5.24, и предположим, что параметры устройства имеют такие же значения, как и в примере 5.5: $C_{int} = 3$ фФ; $C_{fan} = 3$ фФ; $R_{dr} = 0,5(13/1,5 + 31/4,5) = 7,8$ кОм. Провод реализован в металле и имеет ширину 0,4 мкм (минимально допустимая). В результате получаем следующие значения параметров: $c_w = 92$ аФ/мкм; $r_w = 0,19$ Ом/мкм (пример 4.4). С помощью уравнения (5.39) мы можем рассчитать, при какой длине провода задержка межсоединения сравняется

с внутренней задержкой, определяемой исключительно паразитными нагрузками устройства. Решая приведенное ниже квадратное уравнение, получаем единственное (имеющее смысл) решение:

$$6,6 \times 10^{-18} L^2 + 0,5 \times 10^{-12} L = 32,29 \times 10^{-12},$$
$$\text{или } L = 65 \text{ мкм}.$$

Итак, мы видим, что дополнительная задержка порождается исключительно линейным членом уравнения (точнее, дополнительной емкостью, вводимой проводом). Квадратичный член (распределенная задержка провода) начинает доминировать только при гораздо б*о*льших длинах проводов (больше 7 см). Данный фактор вызван высоким сопротивлением питающих транзисторов (минимального размера). При использовании более широких транзисторов наблюдается совсем другой баланс. Проанализируйте, например, решение той же задачи, когда питающие транзисторы имеют в 100 раз б*о*льшую ширину.

5.5. Мощность, энергия и задержка энергии

Выше мы показали, что статический КМОП-инвертор с его практически идеальной передаточной характеристикой по напряжению, полным перепадом логических уровней и высоким запасом помехоустойчивости предлагает невероятную устойчивость, которая значительно упрощает процесс разработки и открывает двери автоматизированному проектированию. Другой привлекательной чертой статического КМОП-инвертора является почти нулевая потребляемая мощность при работе в стационарном режиме. Такая комбинация устойчивости и низкой статической мощности сделала КМОП предпочтительной технологией во многих современных цифровых разработках. Рассеяние мощности в КМОП-схеме происходит “благодаря” динамическим процессам, порождаемым зарядкой и разрядкой емкостей.

5.5.1. Динамическое потребление мощности

Динамическое рассеяние из-за заряжающихся и разряжающихся емкостей

При каждой зарядке конденсатора C_L через p-МОП-транзистор его напряжение поднимается с 0 до V_{DD}, и из источника питания получается определенная энергия. Часть этой энергии рассеивается на p-МОП-устройстве, а оставшаяся часть запасается в конденсаторе. В процессе изменения уровня сигнала от высокого к низкому данный конденсатор разряжается и запасенная энергия рассеивается в n-МОП-транзисторе[5].

[5] Обратите внимание на то, что данная модель является упрощением реальной схемы. В действительности емкость нагрузки состоит из нескольких компонентов, одна часть которых расположена между выходным узлом и заземлением, другая — между выходным узлом и V_{DD}. У этой второй группы цикл зарядки/разрядки не совпадает по фазе с разрядкой конденсаторов на землю

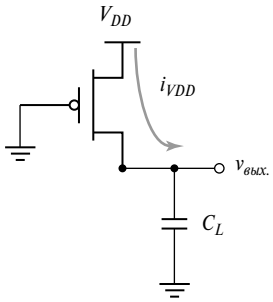


Рис. 5.25. Эквивалентная схема для изменения уровня сигнала от низкого к высокому

Мы можем вывести точную меру описанного потребления энергии. Рассмотрим вначале изменение уровня сигнала от низкого к высокому. Сначала предположим, что входной сигнал имеет нулевые времена нарастания и спада (другими словами, n-МОП- и p-МОП-устройства никогда не находятся одновременно во включенном состоянии). При таких условиях справедлива эквивалентная схема, представленная на рис. 5.25. Энергию E_{VDD}, взятую от источника в ходе данного процесса, а также энергию E_C, запасенную конденсатором к концу процесса, можно найти, проинтегрировав мгновенную мощность по интересующему нас интервалу времени:

$$\begin{aligned} E_{VDD} &= \int_0^\infty i_{VDD}(t) V_{DD} dt = V_{DD} \int_0^\infty C_L \frac{d\nu_{\text{вых.}}}{dt} dt = \\ &= C_L V_{DD} \int_0^{V_{DD}} d\nu_{\text{вых.}} = C_L V_{DD}^2 \end{aligned} \tag{5.40}$$

и

$$\begin{aligned} E_C &= \int_0^\infty i_{VDD}(t) \nu_{\text{вых.}} dt = \int_0^\infty C_L \frac{d\nu_{\text{вых.}}}{dt} \nu_{\text{вых.}} dt = \\ &= C_L \int_0^{V_{DD}} \nu_{\text{вых.}} d\nu_{\text{вых.}} = \frac{C_L V_{DD}^2}{2}. \end{aligned} \tag{5.41}$$

Соответствующие сигналы $v_{\text{вых.}}(t)$ и $i_{VDD}(t)$ показаны на рис. 5.26.

Такие же результаты можно получить, заметив, что при изменении уровня сигнала от низкого к высокому ($L \rightarrow H$) в C_L загружается заряд $C_L V_{DD}$. Выдача такого заряда требует от источника энергии, равной $C_L V_{DD}^2$

(т.е. они заряжаются, когда $V_{\text{вых.}}$ падает, и разряжаются, когда $V_{\text{вых.}}$ растет.) Хотя таким образом доставка энергии от источника разбивается на две фазы, это не влияет на общее рассеяние, поэтому результаты, представленные в данном разделе, все же справедливы.

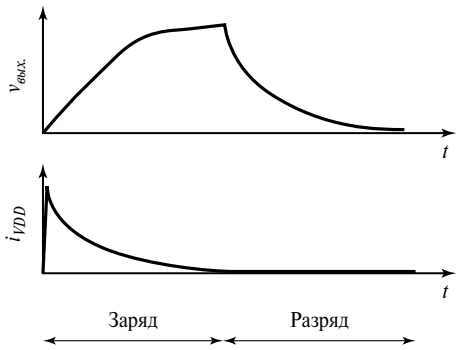


Рис. 5.26. Выходное напряжение и сила тока в процессе зарядки (разрядки) C_L

$(= Q \times V_{DD})$. Энергия, запасенная в конденсаторе, равна $C_L V_{DD}^2/2$. Это означает, что в C_L запасена только половина энергии, поставленной источником. Вторая половина была рассеяна p-МОП-транзистором. Обратите внимание, что данная диссипация энергии не зависит от размера (а следовательно, сопротивления) p-МОП-устройства! В ходе фазы разрядки $(H \rightarrow L)$ заряд снимается с конденсатора и его энергия рассеивается в n-МОП-устройстве. Как и ранее, все это происходит независимо от размера устройства. Подведем итоги: каждый цикл переключения (состоящий из переходов $L \rightarrow H$ и $H \rightarrow L$) забирает фиксированную энергию, равную $C_L V_{DD}^2$. Чтобы рассчитать потребление мощности, необходимо учесть, насколько часто переключается устройство. Если устройство включается и выключается $f_{0\rightarrow1}$ раз в секунду, потребление мощности равно

$$P_{dyn} = C_L V_{DD}^2 f_{0\rightarrow1}, \tag{5.42}$$

где $f_{0\rightarrow1}$ представляет частоту энергопотребляющих переходов (для статического КМОП-инвертора это переходы $0 \rightarrow 1$).

Технологические прорывы приводят к тому, что значение $f_{0\rightarrow1}$ вырастает еще больше (по мере уменьшения t_p). В то же время суммарная емкость элементов на кристалле (C_L) увеличивается, поскольку на него помещается все больше и больше элементов. Рассмотрим, например, КМОП-кристалл, обработанный по 0,25-микронной технологии, с тактовой частотой 500 МГц и средней емкостью нагрузки 15 фФ/элемент, предполагая, что среднее число разветвлений по выходу равно 4. Потребление мощности на элемент при источнике питания 2,5 В равно приблизительно 50 мВт. Для структуры с одним миллионом элементов (считая, что в каждом такте выполняется переход) потребление мощности составит порядка 50 Вт! К счастью, данная оценка является пессимистичной. В реальной жизни не все элементы интегральной схемы переключаются с полной частотой 500 МГц. Реальная активность схемы значительно ниже.

Пример 5.11. Емкостное рассеяние энергии инвертора

Емкостное рассеяние КМОП-инвертора, приведенного в примере 5.4, рассчитать довольно просто. Из табл. 5.2 находим, что емкость нагрузки равна 6 фФ. При напряжении источника питания 2,5 В энергия, необходимая для зарядки и разрядки этой емкости, равна:

$$E_{dyn} = C_L V_{DD}^2 = 37,5 \text{ фДж}.$$

Предположим, что инвертор переключается с максимальной возможной частотой ($T = 1/f = t_{pLH} + t_{pHL} = 2t_p$). При $t_p = 32,5$ пс (см. пример 5.5) находим, что динамическое рассеяние мощности схемы равно:

$$P_{dyn} = E_{dyn}/(2t_p) = 580 \text{ мкВт}.$$

Разумеется, в реальной цепи инвертор редко переключается с максимальной частотой, а даже если бы и переключался, то выход не менялся бы от минимального до максимального, захватывая весь диапазон. Таким образом, рассеяние мощности должно быть значительно ниже. При частоте 4 ГГц ($T = 250$ пс) рассеяние уменьшается до 150 мВт. Сказанное подтверждается результатами моделирования, которое дает 155 мВт.

Расчет рассеяния в сложной схеме затрудняется фактором $f_{0\to1}$, также называемым *активностью переключения*. Хотя для инвертора рассчитать активность переключения достаточно легко, при рассмотрении более сложных элементов и схем ситуация будет не такой радужной. Одна из проблем заключается в том, что активность переключения сети является функцией природы и статистических характеристик входных сигналов: если входные сигналы остаются неизменными, переключения не происходят и динамическое потребление мощности равно нулю! С другой стороны, быстро меняющиеся сигналы инициируют множество переключений, а следовательно — рассеяние. Существуют и другие факторы, влияющие на активность переключения, — общая топология сети и реализуемая функция. Сказанное можно записать следующим образом:

$$P_{dyn} = C_L V_{DD}^2 f_{0\to1} = C_L V_{DD}^2 P_{0\to1} f = C_{EFF} V_{DD}^2 f, \tag{5.43}$$

где f теперь представляет максимальную возможную частоту поступления входных сигналов (практически всегда это тактовая частота) и $P_{0\to1}$ — вероятность того, что событие дает переход $0 \to 1$ (или энергопотребляющее событие) на выходе элемента. $C_{EFF} = P_{0\to1} C_L$ называется *эффективной емкостью* и представляет среднюю емкость, переключаемую на каждом такте. В качестве примера отметим, что при активности 10% ($P_{0\to1} = 0,1$) среднее потребление мощности снижается до 5 Вт.

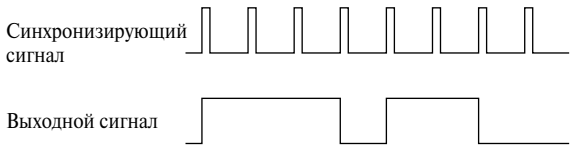


Рис. 5.27. Тактовые импульсы и сигнал

Пример 5.12. Активность переключения

Рассмотрим сигналы, изображенные на рис. 5.27: верхний представляет идеальный тактовый сигнал, нижний — сигнал на выходе элемента. Энергопотребляющие переходы происходят два раза из восьми возможных, следовательно, вероятность перехода равна 0,25 (или 25%).

Технологии разработки для получения низкого потребления энергии/мощности

С увеличением сложности цифровых интегральных схем ожидается, что в ближайшем будущем проблема потребляемой мощности станет еще острее, чем сейчас. В частности, из-за этого все более и более привлекательной становится возможность использования низких напряжений источников питания. **Уменьшение V_{DD} квадратично влияет на P_{dyn}.** Например, уменьшение V_{DD} с 2,5 до 1,25 В в нашем примере уменьшает рассеяние энергии с 5 до 1,25 Вт. При этом предполагается возможность поддержания той же тактовой частоты. Изучая рис. 5.17, видим, что данное предположение не настолько нереалистично, как могло показаться, если напряжение источника питания значительно выше порогового напряжения. Как только V_{DD} достигает значения $2V_T$, быстродействие падает невероятно быстро.

Если нижний предел возможного напряжения источника питания определяется внешними условиями (как часто бывает в реальных разработках) или если ухудшение быстродействия из-за понижения напряжения питания недопустимо, единственная оставшаяся возможность уменьшения рассеяния — понизить эффективную емкость. Для этого можно использовать оба фактора влияния: физическую емкость и активность переключения.

Уменьшить активность переключения можно на уровне логики и архитектурной абстракции; подробнее этот вопрос рассмотрен в главе 11. *Уменьшать физическую емкость* полезно в любом случае; это может также улучшить быстродействие схемы. Поскольку бо́льшая часть емкости в комбинационной логической схеме порождается емкостями транзисторов (емкость затвора и диффузионная емкость), имеет смысл свести данные вклады к минимуму, уменьшив таким образом потребляемую мощность. Это означает, что тогда, когда это возможно или оправдано, транзисторы должны иметь *минимальный размер*. Разумеется, это влияет на быстродействие схемы, однако этот эффект можно компенсировать, используя технологии ускорения на

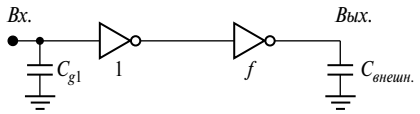


Рис. 5.28. КМОП-инвертор, питающий внешнюю емкость нагрузки C_{ext} и питаемый элементом минимального размера

логическом или архитектурном уровне. Единственный случай, когда размер транзистора ст*о*ит увеличивать, — это когда емкость нагрузки подавляется внешними емкостями (например, емкостью разводки или разветвления по выходу). Это противоречит общей манере разработки, принятой в библиотеках схемных элементов, когда транзисторы обычно делают большими, чтобы учесть все требования к нагрузке и быстродействию.

Сделанные наблюдения ставят перед нами интересную проектную задачу. Предположим, что нам нужно минимизировать рассеяние энергии схемы с заданной допустимой нижней границей быстродействия. Кажется, что оптимальный вариант — насколько возможно понизить напряжение источника питания и компенсировать падение быстродействия увеличением размеров транзисторов. Однако последнее действие приведет еще и к увеличению емкости. Можно предсказать, что при достаточно низком напряжении источника питания этот фактор начнет играть решающую роль и приведет к увеличению энергии (что еще сильнее понизит напряжение источника питания).

Пример 5.13. Масштабирование транзисторов с целью минимизации энергии

Чтобы проанализировать задачу масштабирования транзисторов с целью минимизации энергии, изучим простой случай — статический КМОП-инвертор, питающий внешнюю емкость нагрузки C_{ext} (рис. 5.28). Чтобы учесть эффекты нагрузки на входе, предположим, что сам инвертор питается устройством минимального размера. Наша цель — минимизировать рассеяние энергии во всей цепи, не опускаясь при этом ниже заданного быстродействия. Нашими степенями свободы являются масштабный коэффициент f инвертора и напряжение источника питания V_{dd} цепи. Задержка распространения оптимизированной схемы не должна быть выше, чем у исходной, параметры которой — $f = 1$ и $V_{dd} = V_{ref}$.

Используя подход, введенный в разделе 5.4.3, “Масштабирование цепочки инверторов”, можем вывести следующее выражение для задержки распространения схемы:

$$t_p = t_{p0}\left(\left(1+\frac{f}{\gamma}\right)\left(1+\frac{F}{f\gamma}\right)\right), \tag{5.44}$$

где $F = (C_{ext}/C_{g1})$ — общее эффективное разветвление по выходу; t_{p0} — внутренняя задержка инвертора, зависимость которой от V_{DD} аппроксимируется

следующим выражением, выведенным из уравнения (5.21):

$$t_{p0} \sim \frac{V_{DD}}{V_{DD} - V_{TE}}, \tag{5.45}$$

где $V_{TE} = V_T + V_{DSAT}/2$. Когда общая емкость цепи известна, рассеяние энергии для одного изменения уровня сигнала на входе находится легко:

$$E = V_{dd}^2 C_{g1}((1+\gamma)(1+f)+F). \tag{5.46}$$

Согласно требованиям к быстродействию, задержка распространения масштабированной схемы не должна превышать задержку эталонной схемы ($f = 1$, $V_{dd} = V_{ref}$). Чтобы упростить последующий анализ, предположим, что внутренняя выходная емкость элемента равна его емкости управляющего электрода, или $\gamma = 1$. В таком случае

$$\frac{t_p}{t_{pref}} = \frac{t_{p0}\left(2+f+\frac{F}{f}\right)}{t_{p0ref}(3+F)} = \left(\frac{V_{DD}}{V_{ref}}\right)\left(\frac{V_{ref}-V_{TE}}{V_{DD}-V_{TE}}\right)\left(\frac{2+f+\frac{F}{f}}{3+F}\right) = 1. \tag{5.47}$$

Уравнение (5.47) задает связь между масштабным коэффициентом f и напряжением источника питания, изображенную на рис. 5.29, *а* для различных значений F. Приведенные кривые имеют четкий минимум. Увеличение размера инвертора, начиная с минимального, сначала увеличивает быстродействие, а значит, позволяет понизить напряжение источника питания. Это происходит до тех пор, пока мы не дойдем до оптимального коэффициента $f = \sqrt{F}$ (если вы внимательно читали предыдущие разделы, то такое значение не должно вас удивить). Дальнейшее увеличение размеров устройства приводит только к увеличению параметра самонагрузки, ухудшает быстродействие и требует повышения напряжения источника питания. Кроме того, обратите внимание на случай $F = 1$. Видим, что вариант, предложенный в качестве исходного, действительно является лучшим решением; любое масштабирование только увеличивает самонагрузку.

Зная зависимость $V_{DD}(f)$, можно найти энергию масштабированной схемы (нормированный относительно эталонной схемы) как функцию масштабного коэффициента f:

$$\frac{E}{E_{ref}} = \left(\frac{V_{DD}}{V_{ref}}\right)^2 \left(\frac{2+2f+F}{4+F}\right). \tag{5.48}$$

Мы можем найти аналитическое выражение для оптимального масштабного коэффициента, но оно будет сложным и громоздким. Графический подход почти так же эффективен. Полученные в результате кривые изображены на рис. 5.29, *б*, изучая который можно сделать несколько выводов[6].

[6] Некоторые из данных выводов мы рассмотрим в главе 11 в более широком контексте.

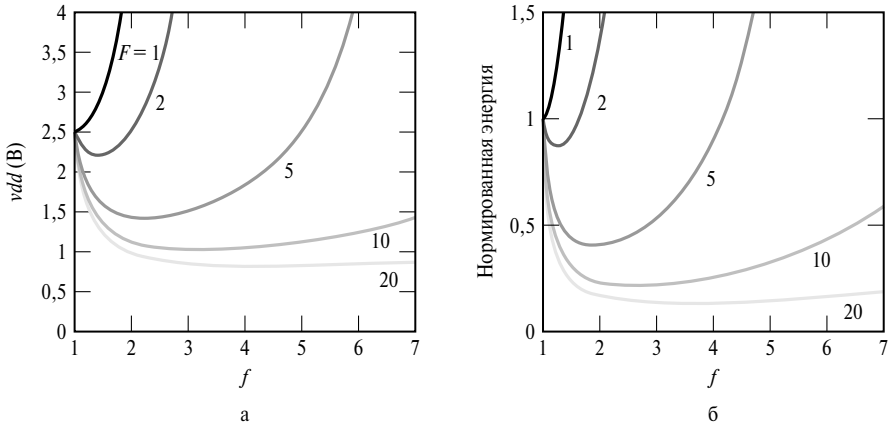


Рис. 5.29. Масштабирование инвертора с целью минимизации потребления энергии: а) требуемое напряжение источника питания как функция масштабного коэффициента f для различных значений общего эффективного разветвления по выходу F; б) энергия масштабированной цепи (нормирована относительно эталонного случая) как функция f. $V_{ref} = 2,5$ В, $V_{TE} = 0,5$ В

- **Масштабирование устройства плюс уменьшение напряжения источника питания является очень эффективным подходом, позволяющим уменьшить потребление энергии логической схемы.** Особенно сказанное справедливо для сетей с большими эффективными разветвлениями по выходу, когда реально можно добиться уменьшения потребления энергии в 10 раз. Для меньших значений F усиление также получается достаточно большим. Единственным исключением является случай $F = 1$, когда наиболее эффективным является устройство минимального размера.
- Чрезмерное увеличение размеров транзисторов свыше оптимального размера требует огромных энергетических ресурсов. К сожалению, во многих современных разработках это является нормой.
- Оптимальный масштабный коэффициент с точки зрения энергии меньше коэффициента, оптимального с точки зрения быстродействия (особенно для больших значений F). Например, при разветвлении на выходе 20, $f_{opt}(energy) = 3,53$, а $f_{opt}(performance) = 4,47$. После того как V_{DD} начинает приближаться к V_{TE}, увеличение размеров устройства приводит только к минимальному уменьшению напряжения источника питания, приводя таким образом к минимальному выигрышу энергии.

Рассеяние, вызванное токами прямого пути

В реальных схемах предположение о нулевых временах нарастания и спуска входного сигнала является несправедливым. Конечный наклон входного сигнала приводит к возникновению прямого пути тока между V_{DD} и землей

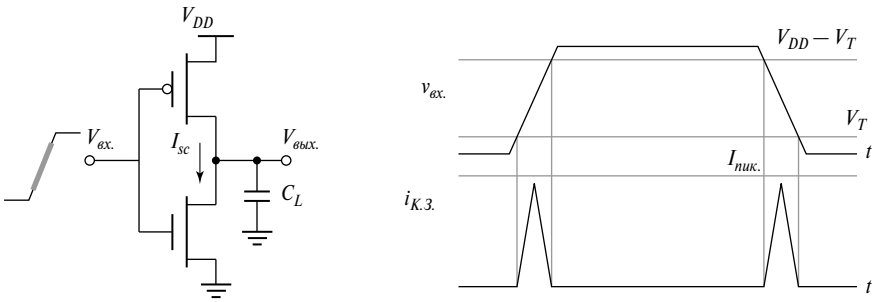


Рис. 5.30. Ток короткого замыкания во время перехода

на небольшой период времени при переключении, когда n-МОП- и p-МОП-транзистор одновременно проводят ток. Это иллюстрируется на рис. 5.30. Делая (разумное) предположение, что получающиеся в результате всплески тока можно аппроксимировать треугольниками и что характеристика инвертора симметрична в периоды нарастания и спада сигнала, можем рассчитать энергию, потребляемую за период переключения:

$$E_{dp} = V_{DD}\frac{I_{peak}t_{sc}}{2} + V_{DD}\frac{I_{peak}t_{sc}}{2} = t_{sc}V_{DD}I_{peak}. \tag{5.49}$$

Среднее потребление энергии рассчитывается следующим образом:

$$P_{dp} = t_{sc}V_{DD}I_{peak}f = C_{sc}V_{DD}^2 f \tag{5.50}$$

Рассеяние энергии на прямом пути пропорционально активности переключения (подобно емкостному рассеянию мощности). В формуле (5.50) t_{sc} представляет время, в течение которого оба устройства проводят ток. При линейном нарастании входного сигнала данное время разумно аппроксимируется уравнением (5.51), где t_s представляет время перехода 0–100%:

$$t_{sc} = \frac{V_{DD} - 2V_T}{V_{DD}}t_s \approx \frac{V_{DD} - 2V_T}{V_{DD}} \times \frac{t_r(f)}{0,8}. \tag{5.51}$$

I_{peak} определяется током насыщения устройств, а следовательно, прямо пропорционально размерам транзисторов. Максимальный ток также **сильно зависит от отношения входного и выходного наклонов сигналов**. Данное отношение лучше всего иллюстрируется с помощью следующего простого анализа. Рассмотрим статический КМОП-инвертор с переходом $0 \rightarrow 1$ на входе. Предположим вначале, что емкость нагрузки очень велика, так что выходное время спада значительно больше входного времени нарастания (рис. 5.31, *а*). В таком случае вход проходит через переходную область до того, как выход начинает меняться. Поскольку в течение этого периода времени напряжение истока-стока p-МОП-устройства приблизительно равно 0, устройство выключается, даже не успев провести ток. В данном случае

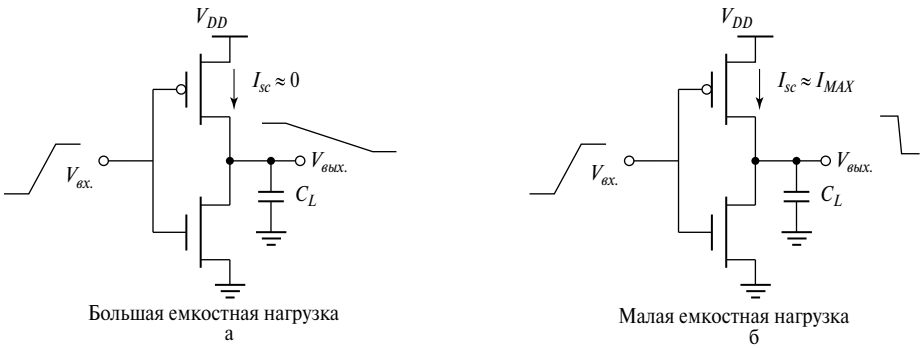


Рис. 5.31. Влияние емкости нагрузки на ток короткого замыкания

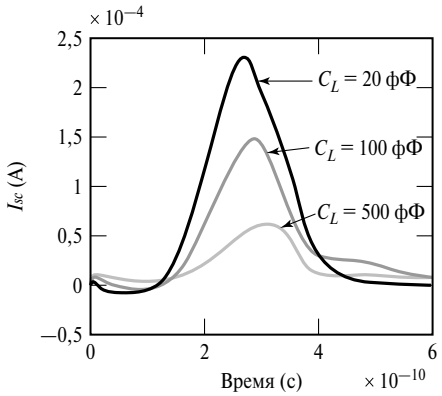


Рис. 5.32. Ток короткого замыкания через n-МОП-транзистор КМОП-инвертора как функция емкости нагрузки (при фиксированном входном наклоне сигнала, равном 500 пс)

ток короткого замыкания близок к нулю. Рассмотрим теперь обратную ситуацию, когда выходная емкость очень мала и выходное время спада значительно меньше входного времени нарастания (см. рис. 5.31, *б*). Напряжение стока-истока p-МОП-устройства равно V_{DD} для большей части периода перехода, что гарантирует максимальный ток короткого замыкания (равный току насыщения p-МОП-транзистора). Очевидно, что такие условия представляют наихудший сценарий. Вывод, сделанный из предыдущего анализа, подтверждается на рис. 5.32, где изображен ток короткого замыкания через n-МОП-транзистор в процессе изменения уровня сигнала от низкого к высокому как функция емкости нагрузки.

Из приведенного анализа можно сделать вывод, что для минимизации рассеяния короткого замыкания выходное время нарастания/спада необходимо сделать больше, чем входное время нарастания/спада. С другой стороны, если сделать выходное время нарастания/спада слишком большим, это замедлит схему и может породить токи короткого замыкания в элементах на

выходе. Таким образом, мы имеем идеальный пример того, как локальная оптимизация без учета глобальной картины может привести к несовершенному решению.

Технологии проектирования

Существует более практичное правило, применение которого глобально оптимизирует потребление мощности ([Veendrick84]).

> Рассеяние мощности из-за токов короткого замыкания минимизируется путем согласования времен нарастания/спада входных и выходных сигналов. На уровне всей схемы это означает, что времена нарастания/спада всех сигналов необходимо поддерживать постоянными в определенном диапазоне.

Сделав идентичными времена нарастания на входе и выходе, мы не только получим оптимальное решение для данного элемента, но и сможем держать в рамках токи короткого замыкания по всей схеме. Сказанное иллюстрируется рис. 5.33, где рассеяние энергии короткого замыкания инвертора (нормированное относительно рассеяния при нулевом времени нарастания на входе) изображено как функция отношения r времени нарастания/спада сигнала на входе и выходе. Когда емкость нагрузки слишком мала для данного размера инвертора ($r > 2 \ldots 3$ для $V_{DD} = 5$ В), основной вклад в мощность дает ток короткого замыкания. Для очень больших значений емкости все рассеяние мощности связано с зарядкой и разрядкой емкости нагрузки. Если сравнять времена нарастания/спада на входе и выходе, бо́льшая часть рассеяния мощности будет относиться к динамической мощности рассеяния, а с токами короткого замыкания будет связана лишь малая доля (меньше 10%).

Кроме того, **влияние тока короткого замыкания уменьшается при понижении напряжения источника питания**, как следует из уравнения (5.51). В граничном случае, когда $V_{DD} < V_{Tn} + |V_{Tp}|$, рассеяние короткого замыкания полностью исчезает, поскольку оба устройства никогда не будут одновременно находиться во включенном состоянии. Поскольку пороговые напряжения уменьшаются медленнее, чем напряжение источника питания, в глубокой субмикронной области рассеяние мощности из-за короткого замыкания становится менее существенным. При напряжении источника питания 2,5 В и порогах порядка 0,5 В для 10%-ного увеличения рассеяния требуется, чтобы отношение наклонов на входе и выходе составляло 2.

Наконец, стоит отметить, что рассеяние мощности из-за короткого замыкания можно смоделировать, как следует из уравнения (5.50), последовательно включив с C_L емкость нагрузки $C_{sc} = t_{sc} I_{peak} / V_{DD}$. Значение данной емкости короткого замыкания является функцией V_{DD}, размеров транзисторов и отношения наклонов сигнала на входе и выходе.

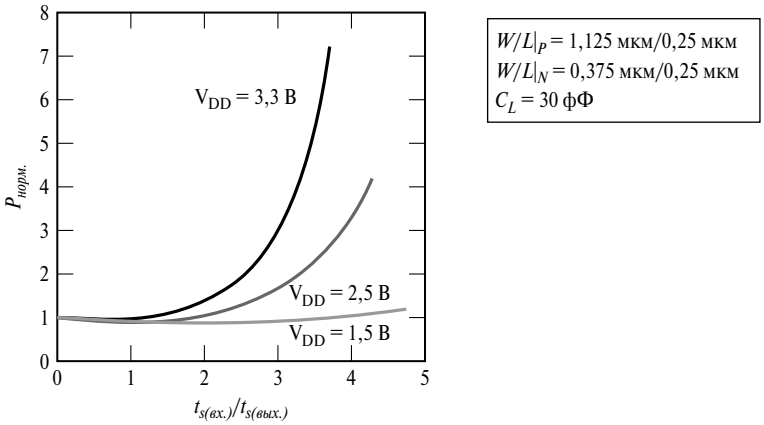


Рис. 5.33. Рассеяние мощности статического КМОП-инвертора как функция отношения между временами нарастания/спада на входе и выходе. Мощность нормирована относительно рассеяния при нулевых временах нарастания/спада на входе. При малых значениях данного параметра связь входа-выхода приводит к небольшому дополнительному рассеянию

5.5.2. Статическое потребление

Статическое (или стационарное) рассеяние энергии схемы выражается следующим соотношением:

$$P_{stat} = I_{stat} V_{DD}, \tag{5.52}$$

где I_{stat} — ток, протекающий между шинами питания при нулевой активности переключения.

В идеальном случае ток покоя КМОП-инвертора равен нулю, поскольку p-МОП- и n-МОП-устройства никогда не находятся одновременно во включенном состоянии при работе в стационарном режиме. К сожалению, еще существует ток утечки, протекающий через обратносмещенные диодные переходы транзисторов, расположенные между истоком или стоком и подложкой, как показано на рис. 5.34. В общем случае данный вклад очень мал, и его можно игнорировать. При размерах, характерных для рассматриваемых устройств, ток утечки на единицу площади стока обычно составляет порядка 10–100 пА/мкм^2 при комнатной температуре. Для кристалла, содержащего 1 миллион элементов, каждый из которых имеет площадь 0,5 мкм^2 и работает при питающем напряжении 2,5 В, в худшем случае потребление мощности из-за утечки диода равно 0,125 мВ, что, очевидно, пренебрежимо мало.

Тем не менее следует помнить, что токи утечки через переходы вызываются носителями, генерируемыми под воздействием температуры. Их значение экспоненциально увеличивается с повышением температуры перехода. При 85°C (широко распространенная верхняя граница температуры перехода в коммерческом аппаратном обеспечении) токи утечки увеличиваются примерно в 60 раз по сравнению со значениями при комнатной температуре. Следовательно, общую рабочую температуру желательно поддерживать как можно

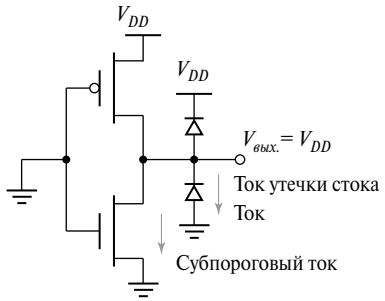


Рис. 5.34. Источники токов утечки в КМОП-инверторе ($V_{вх.} = 0$ В)

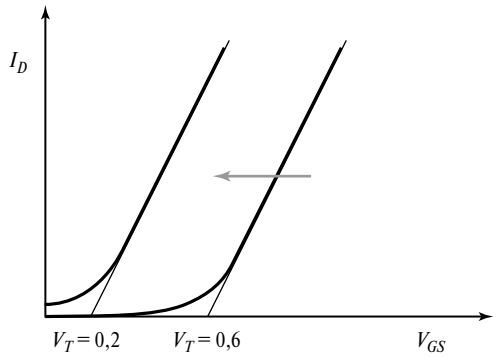


Рис. 5.35. Уменьшение порога приводит к увеличению подпорогового тока при $V_{GS} = 0$

более низкой. Поскольку температура сильно зависит от рассеянного тепла и механизмов его устранения, сказанное возможно только при ограничении рассеяния мощности схемы или при использовании корпусов кристаллов ИС, поддерживающих эффективный отвод тепла.

Новым источником тока утечки является подпороговый ток транзисторов. Как обсуждалось в главе 3, МОП-транзистор может иметь ток стока-истока, даже если V_{GS} меньше порогового напряжения (рис. 5.35). Чем ближе пороговое напряжение к нулю вольт, тем больше ток утечки при $V_{GS} = 0$ В и тем больше статическое потребление мощности. Чтобы компенсировать этот эффект, обычно пороговое напряжение устройства стараются поддерживать достаточно большим. В стандартных процессах значения V_T никогда не бывают меньше 0,5–0,6 В, а в некоторых случаях они даже значительно выше (0,75 В).

Как видно на рис. 3.41, данный подход ставит под сомнение уменьшение напряжение питания, которое обычно используется при масштабировании технологии в глубокую субмикронную область. Ранее мы отмечали (см. рис. 5.17), что масштабирование напряжения питания при неизменном пороговом напряжении приводит к существенному снижению быстродействия, особенно когда V_{DD} приближается к $2V_T$. Один из способов борь-

бы с данной проблемой заключается в пропорциональном уменьшении пороговых напряжений устройства. Вследствие этого кривая, изображенная на рис. 5.17, смещается влево (это означает, что ухудшение быстродействия при уменьшении напряжения питания уменьшается). К сожалению, возможные значения пороговых напряжений ограничены снизу величиной допустимого субпорогового тока утечки (см. рис. 5.35). Таким образом, выбор порогового напряжения представляет собой компромисс между быстродействием и статическим рассеянием мощности. Тем не менее постоянное масштабирование напряжения питания, предсказанное для следующего поколения КМОП-технологий, вынуждает еще больше снижать пороговые напряжения и приводит к тому, что подпороговые условия становятся основным источником рассеяния мощности. Таким образом, более привлекательными становятся технологические процессы, содержащие устройства с более резкими запирающими характеристиками. В качестве примера можно привести технологию КНД (кремний на диэлектрике, Silicon-on-Insulator — SOI), когда изготовленные МОП-транзисторы характеризуются коэффициентами спада, близкими к идеальным 60 мВ/декада.

Пример 5.14. Влияние снижения порога на быстродействие и статическое рассеяние мощности

Рассмотрим n-МОП-транзистор минимального размера, реализованный по 0,25-мкм КМОП-технологии. В главе 3 мы вывели, что коэффициент спада S для данного устройства равен 90 мВ/декада. Ток транзистора в выключенном состоянии (при $V_{GS} = 0$) для $V_T \approx 0,5$ В равен примерно 10^{-11} А (см. рис. 3.22). Уменьшение порога на 200 мВ до 0,3 В увеличивает ток нерабочего состояния транзисторов в 170 раз! Предполагая, что структура из миллиона элементов использует напряжение питания 1,5 В, мы получаем, что статическое рассеяние мощности равно $10^6 \times 170 \times 10^{-11} \times 1,5 = 2,6$ мВ. Дальнейшее уменьшение порога до 100 мВ приводит к неприемлемому рассеянию, достигающему почти 0,5 Вт! При таком напряжении источника уменьшение порога соответствует улучшению быстродействия на 25 и 40% соответственно.

Данная нижняя граница порогов является в некотором смысле искусственной. Дело в том, что считать ток утечки статической КМОП-цепи нулевым неправильно. Очевидно, что наличие токов утечки ухудшает запас помехоустойчивости, поскольку логические уровни перестают быть равными шинам питания, но до тех пор, пока запас помехоустойчивости принадлежит допустимому диапазону, данный момент несуществен. Разумеется, токи утечки приводят к увеличению статического рассеяния мощности. Чтобы компенсировать его, используется понижение напряжения питания, которое допускается из-за сниженных порогов без ущерба быстродействию и приводит к квадратичному уменьшению динамической мощности. Для 0,25-микронного КМОП-процесса следующие схемы имеют одинаковое быстродействие: напряжение источника питания 3 В, $V_T = 0,7$ В; напряжение источника питания $0,45$ В $V_T = 0,1$ В. При этом динамическое потребление мощности у

второй схемы в 45 раз меньше [Liu93]! Повторимся, выбор правильных значений напряжений источника питания и порога является компромиссным. Оптимальная рабочая точка зависит от активности схемы. При наличии заметного статического рассеяния энергии необходимо, чтобы неактивные модули получали меньше питания (выключались), иначе статическое рассеяние мощности может быть основным источником потерь. *Режим пониженного потребления энергии* можно реализовать, отсоединив блок от шины питания или понизив питающее напряжение.

5.5.3. Собираем все вместе

Теперь общее потребление мощности КМОП-инвертора можно выразить как сумму трех его компонентов:

$$P_{tot} = P_{dyn} + P_{dp} + P_{stat} = (C_L V_{DD}^2 + V_{DD} I_{peak} t_s) f_{0\to1} + V_{DD} I_{peak}. \qquad (5.53)$$

В типичных КМОП-цепях доминирующим фактором, без сомнения, является емкостное рассеяние. Потребление на прямом пути можно удержать в приемлемых пределах, тщательно продумав структуру устройства, поэтому данный фактор можно не учитывать. Утечку в настоящее время можно игнорировать, однако в обозримом будущем ситуация может существенно измениться.

Произведение мощности на задержку, или расход энергии на одну операцию

В главе 1 мы ввели параметр PDP — произведение мощности на задержку (Power-Delay Product), который использовали в качестве меры качества логического элемента:

$$PDP = P_{av} t_p. \qquad (5.54)$$

Как следует из размерности (В × с = Дж), PDP представляет меру энергии. Предполагая, что элемент переключается с максимальной возможной частотой $f_{max} = 1/(2t_p)$, и игнорируя вклады в потребление энергии статического тока и тока прямого пути, находим, что

$$PDP = C_L V_{DD}^2 f_{max} t_p = \frac{C_L V_{DD}^2}{2}. \qquad (5.55)$$

В данном уравнении PDP представляет потребление энергии на переключение (т.е. для перехода $0 \to 1$ или $1 \to 0$). Как правило, нас интересует энергия, потребляемая за цикл переключения E_{av}. Поскольку каждый цикл инвертора содержит переход $0 \to 1$ и $1 \to 0$, E_{av} вдвое превышает PDP.

Произведение энергия-задержка

Приемлемость PDP в качестве меры качества технологического процесса или топологии элемента является спорной. Мы измеряем энергию, необходимую для переключения элемента, — довольно важную его характеристику. Однако для данной структуры это число можно сделать сколь угодно малым, уменьшив напряжение питания. С такой точки зрения оптимальное напряжение, питающее схему, будет наименьшим возможным, при котором схема еще работает. Однако, как обсуждалось ранее, это происходит за счет быстродействия. Более удачная метрика должна объединять меру быстродействия и энергии. Именно это дает произведение энергии на задержку (Energy-Delay Product — EDP):

$$EDP = PDP \times t_p = P_{av} t_p^2 = \frac{C_L V_{DD}^2}{2} t_p. \tag{5.56}$$

Проанализируем зависимость напряжения от EDP. Более высокое напряжение питания уменьшает задержку, для низких напряжений имеем противоположный эффект. Следовательно, должна существовать оптимальная рабочая точка. Предполагая, что n-МОП- и p-МОП-транзисторы имеют сравнимые пороги и напряжения насыщения, выражение (5.21) для задержки распространения мы можем упростить до такой формулы:

$$t_p \approx \frac{\alpha C_L V_{DD}}{V_{DD} - V_{TE}}, \tag{5.57}$$

где $V_{Te} = V_T + V_{DSAT}/2$; α — технологический параметр. Объединяя уравнение (5.56) и (5.57)[7], получаем:

$$EDP = \frac{\alpha C_L^2 V_{DD}^3}{2(V_{DD} - V_{Te})}. \tag{5.58}$$

Чтобы найти оптимальное напряжение питания, необходимо продифференцировать уравнение (5.58) по V_{DD} и приравнять результат к 0. Получаем:

$$V_{DDopt} = \frac{3}{2} V_{TE}. \tag{5.59}$$

Интересным результатом проведенного анализа является низкое значение напряжения питания, которое одновременно оптимизирует быстродействие и потребление энергии. Для субмикронных технологий, пороги которых лежат в районе 0,5 В, оптимальное питание составляет порядка 1 В.

[7] Данное уравнение является точным только в том случае, когда устройства остаются в режиме насыщения скорости, что, очевидно, несправедливо для низких напряжений питания. Из-за этого анализ несколько неточен, однако на общий результат это влияет не сильно.

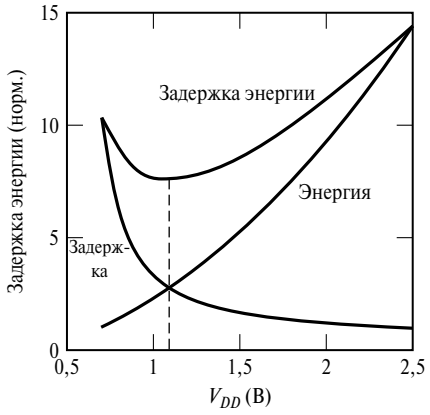


Рис. 5.36. Нормированные задержка, энергия и произведение энергии на задержку для КМОП-инвертора, реализованного по 0,25-микронной технологии

Пример 5.15. Оптимальное напряжение питания для КМОП-инвертора, реализованного по 0,25-микронной технологии

Зная технологические параметры нашего общего КМОП-процесса, представленного в главе 3, можно следующим образом рассчитать значение V_{TE}:

$$V_{Tn} = 0,43 \text{ В}, V_{Dsatn} = 0,63 \text{ В}, V_{TEn} = 0,74 \text{ В};$$
$$V_{Tp} = -0,4 \text{ В}, V_{Dsatp} = -1 \text{ В}, V_{TEp} = -0,9 \text{ В};$$
$$V_{TE} \approx (V_{TEn} + |V_{TEp}|)/2 = 0,8 \text{ В}.$$

Следовательно, $V_{DDopt} = (3/2) \times 0,8$ В$= 1,2$ В. Данный результат подтверждается графиками, представленными на рис. 5.36 (результаты моделирования нормированной задержки, энергии и произведения энергии на задержку). Из данных графика отчетливо видно, что выбор подходящих значений энергии и задержки является компромиссом. Предсказываемое оптимальное напряжение питания составляет 1,1 В.

Предупреждение

Хотя в предыдущем примере показано существование напряжения питания, минимизирующего произведение энергии элемента на задержку, данное напряжение не обязательно представляет оптимальное напряжение для поставленной задачи проектирования. Например, в некоторых проектах требуется обеспечить минимальное быстродействие, что требует более высокого напряжения (за счет энергии). Подобным образом можно получить структуру с более низким потреблением энергии, работая при меньшем напряжении и получив общее требуемое быстродействие системы путем использования таких архитектурных технологий, как конвейерная или параллельная обработка.

5.5.4. Анализ потребления мощности с помощью SPICE

Определение среднего потребления мощности мы дали в главе 1, ниже оно повторяется для удобства.

$$P_{av} = \frac{1}{T}\int_0^T p(t)dt = \frac{V_{DD}}{T}\int_0^T i_{DD}(t)dt, \tag{5.60}$$

где T — интервал усреднения; V_{DD} и i_{DD} — напряжение и сила тока источника напряжения соответственно. В некоторых реализациях SPICE существуют встроенные функции, измеряющие среднее значение сигнала в цепи. Например, команда `.MEASURE TRAN I(VDD) AVG` (пакет HSPICE) рассчитывает площадь под вычисленной переходной характеристикой (I(VDD)) и делит ее на указанный период времени. В уравнении (5.60) заложены идентичные действия. К сожалению, другие реализации SPICE не настолько мощные. Правда, это не так плохо, как кажется, если вспомнить, что пакет SPICE в действительности является пакетом для решения дифференциальных уравнений. Мы легко можем создать небольшую цепь, действующую как интегратор, выходной сигнал которого будет представлять среднюю мощность.

Рассмотрим, например, цепь, изображенную на рис. 5.37. Ток, поступающий от источника, измеряется токоуправляющим источником тока и накапливается (интегрируется) в конденсаторе C. Сопротивление R добавлено только из соображений сходимости постоянного тока, и для минимизации утечки его следует выбирать как можно бо́льшим. При правильном выборе параметров элемента напряжение на выходе (P_{av}) будет равно среднему потреблению мощности. Работа цепи описывается следующим уравнением (предполагается, что первоначальное напряжение на конденсаторе C равно нулю):

$$C\frac{dP_{av}}{dt} = ki_{DD}$$

или

$$P_{av} = \frac{k}{C}\int_0^T i_{DD}dt. \tag{5.61}$$

Приравнивая уравнение (5.60) к уравнению (5.61), получаем необходимые условия на эквивалентные параметры цепи: $k/C = V_{DD}/T$. При таких условиях изображенная эквивалентная цепь представляет удобные средства отслеживания средней мощности цифровой схемы.

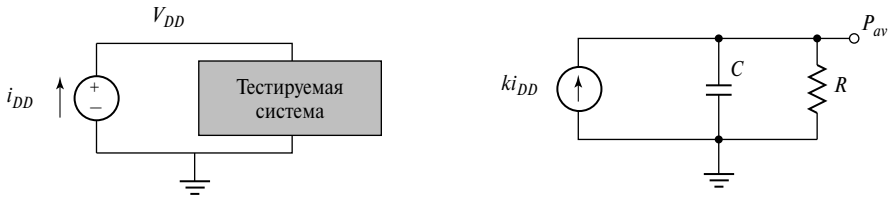


Рис. 5.37. Эквивалентная схема, используемая для моделирования средней мощности в SPICE

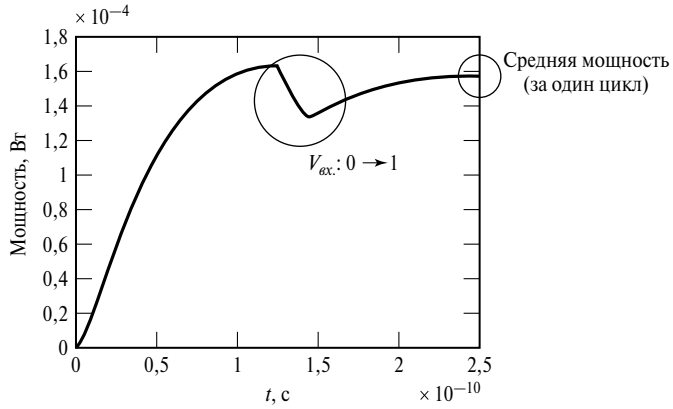


Рис. 5.38. Расчет потребления мощности с помощью пакета SPICE

Пример 5.16. Средняя мощность инвертора

Проанализируем среднюю потребляемую мощность инвертора, рассмотренного в примере 5.4, с использованием описанной техники для периода в 250 пс ($T = 250$ пс, $k = 1$, $V_{DD} = 2,5$ В, поскольку $C = 100$ пФ). Получающееся потребление мощности приведено на рис. 5.38, из которого следует, что средняя потребляемая мощность составляет примерно 157,3 мВ. Используя команду `.MEAS AVG` (SPICE), получаем значение 160,3 мВ, что показывает приблизительную эквивалентность обоих методов. Данные числа соответствуют энергии 39 фДж (что близко к полученному в примере 5.11 значению 37,5 фДж). Изучая график, обратите внимание на падение мощности при изменении уровня сигнала от высокого к низкому. Это явление вызвано подачей тока на источник питания, когда на небольшое время выход превышает V_{DD} из-за емкостной связи между входом и выходом (это следует из переходной характеристики, показанной на рис. 5.16).

Таблица 5.4. Сценарии масштабирования для короткоканальных устройств (S и U представляют параметры масштабирования технологии и напряжения соответственно)

Параметр	Формула	Полное масштабирование	Общее масштабирование	Масштабирование с постоянным напряжением
Площадь-прибор	$W - L$	$1/S^2$	$1/S^2$	$1/S^2$
Собственная задержка	$R_{on}C_{gate}$	$1/S$	$1/S$	$1/S$
Собственная энергия	$C_{gate}V_2$	$1/S^3$	$1/SU^2$	$1/S$
Собственная мощность	Энергия-задержка	$1/S^2$	$1/U^2$	1
Плотность мощности	P-площадь	1	S^2/U^2	S^2

5.6. Перспектива: масштабирование технологии и ее влияние на метрики инвертора

В разделе 3.5 мы рассматривали влияние масштабирования технологии на некоторые важные параметры разработки — площадь, задержку и мощность. Для ясности повторим некоторые важные позиции табл. 3.8.

Справедливость теоретических расчетов можно проверить, обратившись к истории и проследив тенденции развития интегральных схем за последние несколько десятков лет. На рис. 5.39 видно, что задержка элемента действительно экспоненциально растет со скоростью, примерно равной 13% за год (или удваивается каждые пять лет). Данная скорость согласуется с данным, представленными в табл. 5.4, поскольку, как видно из рис. 3.40, S в среднем равно примерно 1,15. Задержка элемента НЕ-И с двумя входами и разветвлением по выходу равным 4 увеличилась с десятков наносекунд (1960-е годы) до десятых наносекунды (2000 год), а в будущем (к 2010 году) прогнозируется ее рост до нескольких десятков пикосекунд.

До недавнего времени уменьшение рассеяния не относилось к числу главных приоритетов разработки. Соответственно, было собрано мало статистики по рассеянию на один элемент или конструкцию. Рассмотрим график, изобра-

Рис. 5.39. Эволюция задержки элемента (согласно [Dally98])

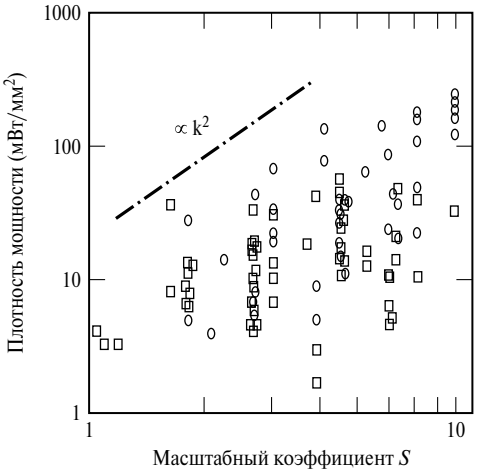


Рис. 5.40. Эволюция плотности мощности в микропроцессорах и ЦОС-процессорах как функция масштабного коэффициента S ([Kuroda95]). S нормировано на 1 до процесса 4 мкм

женный на рис. 5.40, где представлена плотность мощности, измеренная для большого числа конструкций, произведенных между 1980 и 1995 гг. Хотя наблюдаемый разброс достаточно велик даже для фиксированной технологии, из него видно, что плотность мощности растет примерно как S^2. Это согласуется со сценарием масштабирования с постоянным напряжением, представленным в табл. 5.4. Ожидается, что в последующие годы сценарий будет ближе к модели полного масштабирования (которая предсказывает постоянную плотность мощности) из-за ускоренного масштабирования напряжения питания и повышения внимания к технологиям разработки с пониженным потреблением мощности. Но даже в таком случае рассеяние мощности на один кристалл будет расти из-за использования кристаллов все больших размеров.

Следует отметить, что представленная модель масштабирования имеет один большой недостаток. Предсказания быстродействия и мощности дают чисто "внутренние" значения, учитывающие исключительно параметры устройства. В главе 4 мы уже говорили, что соединительные провода имеют своеобразное поведение при масштабировании и что паразитные нагрузки провода могут являться определяющим фактором быстродействия. Подобным образом, зарядка и разрядка емкостей проводников может являться основным фактором, определяющим энергетический бюджет. Чтобы получить более четкое представление о происходящем, необходимо построить комбинированную модель, в которой учитывается масштабирование устройства и масштабирование проводников. Влияние емкости проводников и ее масштабирования на общее поведение схемы представлено в табл. 5.5 (мы использовали модель с фиксированным сопротивлением, представленную в главе 4). Кроме того, мы предположили, что сопротивление питающего элемента подавляет сопротивление проводов (что, очевидно, справедливо для коротких и средних проводов).

Таблица 5.5. Сценарии масштабирования емкости проводов (S и U представляют параметры масштабирования технологии и напряжения соответственно; S_L — коэффициент масштабирования длины проводов; ϵ_c — влияние краевой и перекрестной емкости)

Параметр	Формула	Общее масштабирование
Емкость провода	WL/t	ϵ_c/S_L
Задержка провода	$R_{on}C_{int}$	ϵ_c/S_L
Энергия провода	$C_{int}V^2$	$\epsilon_c/S_L U^2$
Задержка провода/собственная задержка		$\epsilon_c S/S_L$
Энергия провода/собственная энергия		$\epsilon_c S/S_L$

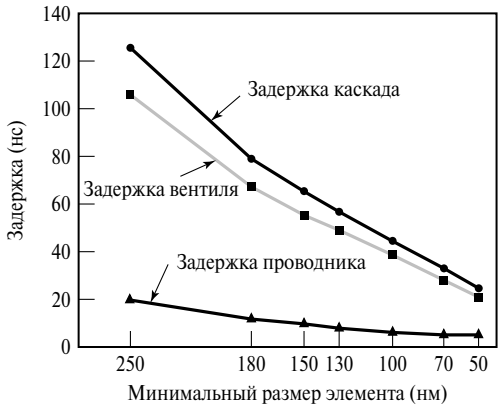


Рис. 5.41. Эволюция отношения задержки провода к задержке элемента с развитием технологии (согласно [Fisher98])

Приведенная модель предсказывает, что задержка (и рассеяние энергии), вызываемая межсоединениями, становится все более важной по мере масштабирования технологии. Данное влияние ограничивается увеличением задержки с ростом ε_c для коротких проводов ($S = S_L$), но для средних и длинных проводов ($SL < S$) оно становится гораздо существеннее. Данные выводы подтверждаются многочисленными исследованиями, один из примеров которых показан на рис. 5.41. Вопрос о том, как в действительности будет эволюционировать соотношение между параметрами провода и внутренними вкладами, является спорным, поскольку ответ на него зависит от целого набора независимых параметров — архитектуры системы, методологии разработки, масштабирования транзисторов, а также используемых материалов межсоединений. Предсказывать что-либо даже на ближайшее будущее довольно затруднительно. Единственное, что ясно, — повышенное внимание к межсоединениям абсолютно закономерно, поскольку оно может изменить основоположные принципы разработки и оптимизацию интегральных схем следующего поколения (см., например, [Sylvester98]).

5.7. Резюме

В данной главе представлен строгий и подробный анализ статического КМОП-инвертора. Ключевые характеристики этого элемента кратко сформулированы ниже.

- Статический КМОП-инвертор объединяет нагрузочный p-МОП-участок с разгрузочным n-МОП-устройством. Как правило, p-МОП-транзистор делают шире n-МОП-транзистора из-за возможности питания более низким током.
- Данный элемент имеет практически идеальную передаточную функцию по напряжению. Перепад логических уровней равен напряжению источника питания и не зависит от размеров транзисторов. Запас помехоустойчивости симметричного инвертора (когда p-МОП- и n-МОП-транзисторы имеют одинаковые характеристики питания током) приближается к $V_{DD}/2$. На стационарную характеристику разветвление по выходу не влияет.
- Основным фактором, определяющим задержку распространения КМОП-инвертора, является время, требуемое для зарядки или разрядки емкости нагрузки C_L. В первом порядке это можно аппроксимировать следующим образом:

$$t_p = 0,69 C_L \left(\frac{R_{eqn} + R_{eqp}}{2} \right).$$

  Самым эффективным средством реализации быстродействующих схем является поддержание малой емкости нагрузки. Если основное влияние на задержку оказывает внешняя (или нагрузочная) емкость разветвления по выходу и разводки, то быстродействие можно увеличить за счет масштабирования размеров транзисторов.
- Рассеяние мощности связано в основном с динамической мощностью рассеяния при зарядке и разрядке емкости нагрузки и выражается формулой $P_{0\to1} C_L V_{DD}^2 f$. Рассеяние пропорционально активности в сети. Рассеяние из-за токов прямого пути, которое происходит при переключении, можно ограничить, тщательно спроектировав фронт (наклон) сигнала. Статическое рассеяние обычно можно игнорировать, однако в будущем из-за подпороговых токов оно может стать довольно важным фактором, определяющим приоритеты при разработке схем.
- Масштабирование технологии является эффективным средством уменьшения площади, задержки распространения и энергопотребления элемента. Если одновременно с этим масштабировать и напряжение источника питания, эффект может быть просто поразительным.
- Как правило, связующий компонент вносит весьма существенный вклад в бюджет задержки и быстродействия.

5.8. Для любознательных

Работа КМОП-инвертора рассматривается во многих публикациях и учебниках. Практически в любой книге по цифровой разработке можно найти главу, посвященную анализу стандартного инвертора. Полный список рекомендуемой литературы см. в главе 1. Ссылки на некоторые работы, упоминавшиеся в данной главе, приведены ниже.

5.9. Список использованной литературы

[Dally98] W. Dally and J. Poulton, *Digital Systems Engineering*, Cambridge University Press, 1998.

[Fisher98] P. D. Fisher and R. Nesbitt, "The Test of Time: Clock-Cycle Estimation and Test Challenges for Future Microprocessors", *IEEE Circuits and Devices Magazine*, 1998, 14(2), p. 37–44.

[Hedenstierna87] N. Hedenstierna and K. Jeppson, "CMOS Circuit Speed and Buffer Optimization", *IEEE Transactions on CAD*, March 1987, vol. CAD-6, no. 2, p. 270–281.

[Kuroda95] T. Kuroda and T. Sakurai, "Overview of low-power ULSI circuit techniques", *IEICE Trans. on Electronics*, April 1995, vol. E78-C, no. 4, p. 334–344.

[Liu93] D. Liu and C. Svensson, "Trading speed for low power by choice of supply and threshold voltages", *IEEE Journal of Solid-State Circuits*, Jan. 1993, vol. 28, no. 1, p. 10–17.

[Mead80] C. Mead and L. Conway, *Introduction to VLSI Systems*, Addison-Wesley, 1980.

[Sedra87] A. Sedra and K. Smith, *MicroElectronic Circuits*, Holt, Rinehart and Winston, 1987.

[Swanson72] R. Swanson and J. Meindl, "Ion-Implanted Complementary CMOS transistors in Low-Voltage Circuits", *IEEE Journal of Solid-State Circuits*, April 1972, vol. SC-7, no. 2, p.146–152.

[Sylvester98] D. Sylvester and K. Keutzer, "Getting to the Bottom of Deep Submicron", *Proceedings ICCAD Conference*, p. 203, San Jose, November 1998.

[Veendrick84] H. Veendrick, "Short-Circuit Dissipation of Static CMOS Circuitry and its Impact on the Design of Buffer Circuits", *IEEE Journal of Solid-State Circuits*, 1984, vol. SC-19, no. 4, p. 468–473.

Упражнения

Последнюю версию сборника задач и упражнений можно найти по адресу `http://bwrc.eecs.berkley.edu/IcBook`. Сделав упражнения доступными в электронном, а не в печатном виде, авторы смогли обеспечить динамичную среду, способную отслеживать быстрые изменения в современных технологиях проектирования цифровых интегральных схем.

ГЛАВА

6

Разработка комбинационных логических элементов на КМОП

Семейства КМОП-логики

Оптимизация логического элемента с позиции площади, скорости, энергии или надежности

Технологии проектирования маломощных и быстродействующих схем

6.1. Введение
6.2. Разработка статических КМОП-устройств
6.2.1. КМОП ИС
6.2.2. Стандартизованная логика
6.2.3. Логика на проходных транзисторах
6.3. Разработка динамических КМОП-элементов
6.3.1. Динамическая логика: основные принципы
6.3.2. Скорость и рассеяние мощности в динамических логических элементах
6.3.3. Вопросы сохранности сигнала в динамических схемах
6.3.4. Каскадное соединение динамических элементов
6.4. Перспективы
6.4.1. Как выбрать стиль разработки логики
6.4.2. Разработка логики со сниженным потреблением напряжения
6.5. Резюме
6.6. Для любознательных

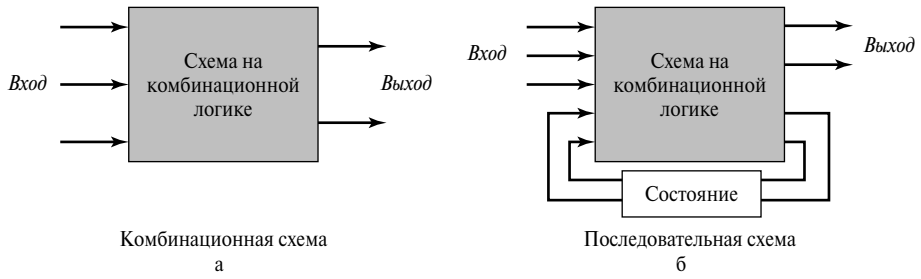


Рис. 6.1. Высокоуровневая классификация логических схем

6.1. Введение

В предыдущей главе были изложены проектные соображения, касающиеся разработки схемы простого инвертора. Расширим теперь это обсуждение на производство любых логических элементов, например НЕ-ИЛИ, НЕ-И и ИСКЛЮЧАЮЩЕЕ ИЛИ. Основное внимание будет обращено на *комбинационную логику* или *нерегенеративные схемы*, т.е. цепи, обладающие следующим свойством: в любой момент времени связь выхода схемы с ее входными сигналами описывается каким-то булевым выражением (предполагается, что нестационарные эффекты, присутствующие в схеме, фиксированы). Никакой специальной обратной связи (умышленного соединения выходов с входами) нет.

В общем, мы не рассматриваем класс устройств, называемых *последовательными*, или *регенеративными*, у которых выход является функцией не только текущих входных данных, но и предыдущих значений входных сигналов (рис. 6.1). Подобного функционирования можно добиться, подав один или несколько выходов на некоторые входы. В результате схема "запомнит" прошедшие события и будет обладать *историей*. Последовательная схема включает комбинационную логику, хранящую информацию о состоянии. В качестве примеров подобных устройств можно привести регистры, счетчики, осцилляторы и ячейки памяти. Подробно последовательные схемы рассмотрены в следующей главе.

Данную логическую функцию можно реализовать множеством различных способов. Как и для инвертора, основными метриками разработки выступают площадь, скорость, энергия и мощность. В зависимости от приложения акцент может ставиться на различные метрики. Например, скорость переключения (коммутации) цифровых схем является основной метрикой в быстродействующем процессоре, тогда как для устройства, работающего от батареи, важнейшим параметром является рассеяние энергии. В последнее время еще одним важным моментом стало рассеяние мощности, поэтому пониманию источников мощности и подходам к работе с мощностью уделяется значительное внимание. Помимо названных метрик очень важными соображениями являются устойчивость к шуму и надежность. Ниже будет показано, что определенные стили логики могут существенно повысить быстродействие, однако при этом они более чувствительны к шуму.

6.2. Разработки статических КМОП-устройств

Наиболее распространенным стилем логики является статическая комплементарная МОП-структура. Статический стиль КМОП в действительности является расширением статического КМОП-инвертора с множественными входами. Напомним, что основным достоинством КМОП-структуры является устойчивость (т.е. низкая чувствительность к шуму), хорошее быстродействие и низкое потребление мощности без статического рассеяния мощности. Большинство этих свойств переносится и на большие логические элементы с разветвлением по входу, реализованные с использованием подобной топологии схемы.

Комплементарные МОП-структуры относятся к широкому классу логических схем, называемых *статическими* схемами, в которых **в каждый момент времени каждый выход элемента соединяется либо с V_{DD}, либо с V_{SS} через тракт с малым сопротивлением**. Кроме того, предполагается, что выходы элементов в любой момент времени описываются булевой функцией, реализованной схемой (игнорируя переходные эффекты во время переключения). Это отличается от класса *динамических* схем, которые полагаются на временные запоминающие устройства значений сигнала в емкости узлов схемы с высоким импедансом. Достоинством последнего подхода является то, что элемент получается проще и быстрее. Тем не менее его структура и работа являются более сложными и подвержены ошибкам из-за повышенной чувствительности к шуму.

В данном разделе мы последовательно рассмотрим разработку различных разновидностей статических схем, включая комплементарные МОП-схемы, стандартизованную логику (псевдо-n-МОП и DCVSL) и логику на проходных транзисторах. Кроме того, мы изучим тему масштабирования с целью уменьшения напряжений питания и пороговых напряжений.

6.2.1. КМОП ИС

Концепция

Статический КМОП-элемент является комбинацией двух цепей задания режима — PUN (Pull-Up Network) и PDN (Pull-Down Network), как показано на рис. 6.2. На рисунке показан логический элемент общего вида с N входами, где все входы распределяются по цепям PUN и PDN. Функцией цепи PUN является обеспечение соединения между выходом и V_{DD} всегда, когда выход логического элемента предполагается равным 1 (на основе входов). Подобным образом функцией цепи PDN является соединение выхода с V_{SS}, когда выход логического элемента предполагается равным 0. Данные цепи создаются по взаимоисключающему принципу, так что в стационарном режиме ток проводит ровно одна цепь. Таким образом, после завершения переходов всегда существует тракт между V_{DD} и выходом F для высокого выхода (“один”) или между V_{SS} и F при низком выходе (“нуль”). Сказанное можно сформулировать и другим способом: выходной узел всегда является узлом с низким импедансом в стационарном состоянии.

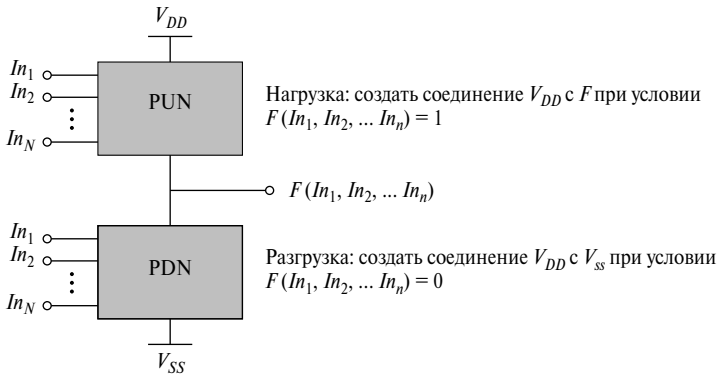
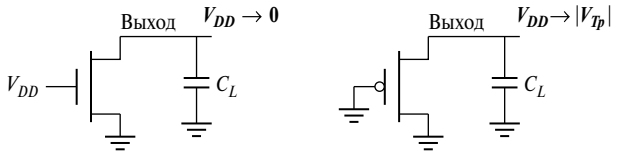


Рис. 6.2. Комплементарный логический элемент как объединение нагрузочной и разгрузочной сети

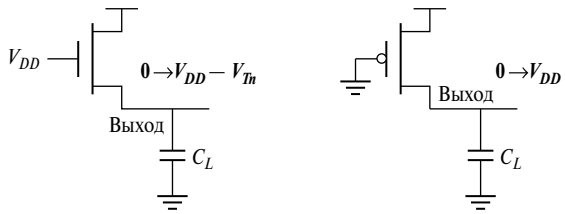
При создании данных цепей разработчик должен иметь в виду следующие соображения.

- Транзистор можно рассматривать как ключ, управляемый отпирающим импульсом. Ключ n-МОП включен, когда управляющий сигнал имеет высокий уровень, и выключен, когда управляющий сигнал имеет низкий уровень; p-МОП-транзистор действует как обратный ключ, включенный при низком управляющем сигнале и выключенный при высоком.
- Разгрузочная сеть создается с использованием n-МОП-устройств, а p-МОП-транзисторы применяются в нагрузочной сети. Основной причиной такого выбора является то, что n-МОП-транзисторы дают "сильные нули", а p-МОП-устройства генерируют "сильные единицы". Чтобы проиллюстрировать сказанное, рассмотрим пример, приведенный на рис. 6.3. На рис. 6.3, *а* выходная емкость изначально заряжена до V_{DD}. Далее показаны два возможных сценария разрядки. Из первого рисунка следует, что n-МОП-устройство опускает выход до уровня GND, тогда как p-МОП-транзистор не опустит выход ниже уровня $|V_{Tp}|$ — p-МОП-устройство просто выключается в этой точке и прекращает давать вклад в ток разрядки. Таким образом, в качестве устройств разгрузочной сети предпочтительнее n-МОП-транзисторы. Подобным образом, на рис. 6.3, *б* показаны два альтернативных подхода к зарядке конденсатора, когда выход изначально равен GND. Ключ p-МОП успешно заряжает выход до уровня V_{DD}, тогда как n-МОП-устройство не может поднять выход выше $V_{DD} - V_{Tn}$. Таким образом, в нагрузочной сети предпочтительнее использовать p-МОП-транзисторы.
- При создании логических функций можно разработать набор правил (рис. 6.4). В частности, n-МОП-устройства, соединенные последовательно, соответствуют функции И. Если все входы высокие, последовательное соединение проводит ток и значение на одном конце цепочки передается на другой конец. Подобным образом, n-МОП-транзисторы, соеди-



Разгрузка узла посредством использования переключателей *n*-МОП и *p*-МОП

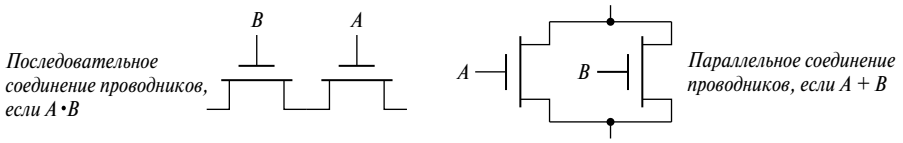
а



Нагрузка узла посредством использования переключателей *n*-МОП и *p*-МОП

б

Рис. 6.3. Простые примеры, иллюстрирующие, почему n-МОП-транзисторы должны использоваться для перевода логического состояния на низкий уровень, а p-МОП-транзисторы — для перевода на высокий уровень



Последовательное соединение

а

Параллельное соединение

б

Рис. 6.4. Правила использования логических элементов n-МОП-транзисторов — последовательно соединенные устройства реализуют функцию И, а устройства, включенные параллельно, — функцию ИЛИ

ненные параллельно, представляют функцию ИЛИ. Между выходом и входом существует проводящий тракт, если высоким является хотя бы один из входов. Используя подобные аргументы, можно сформулировать правила проводимости для p-МОП-сетей. Последовательное включение p-МОП-устройств, если оба входа низкие, представляет функцию ИЛИ-НЕ ($\overline{A} \cdot \overline{B} = \overline{A + B}$); когда p-МОП-транзисторы включены параллельно, они представляют функцию И-НЕ ($\overline{A} + \overline{B} = \overline{A \cdot B}$).

- Используя теорему де Моргана (De Morgan) ($\overline{A + B} = \overline{A} \cdot \overline{B}$ и $\overline{A \cdot B} = \overline{A} + \overline{B}$), можно показать, что нагрузочная и разгрузочная сети комплементарной МОП-структуры являются дуальными. Это означает, что параллельное соединение транзисторов в нагрузочной сети соответствует последовательному соединению соответствующих устройств в разгрузочной сети и наоборот. Следовательно, для создания КМОП-элемента одна из сетей (например, PDN) реализуется с использованием и параллельных, и последовательных соединений. Другая сеть (т.е. PUN) получается путем

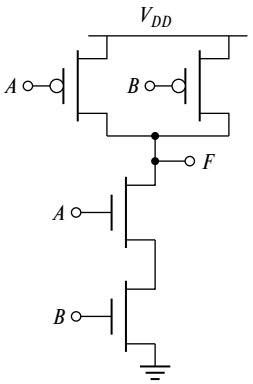


Рис. 6.5. Элемент НЕ-И с двумя входами в комплементарном статическом МОП-стиле

использования принципа дуальности — прохода по иерархическому представлению, замены последовательных подцепей параллельными, а параллельных — последовательными. Полный КМОП-элемент создается за счет объединения цепи PDN с PUN.

- Комплементарный элемент природно является инвертирующим, релизующим только такие функции, как НЕ-И, НЕ-ИЛИ и НЕ-ИЛИ. Реализация неинвертирующих булевых функций (таких, как И, ИЛИ или исключающее ИЛИ) в одном каскаде нереальна и требует введения дополнительного каскада инвертора.
- Число транзисторов, требуемых для реализации логического элемента с N входами, равно $2N$.

Пример 6.1. Элемент И-НЕ с двумя входами

На рис. 6.5 показан элемент И-НЕ с двумя входами ($F = \overline{A \cdot B}$). Разгрузочная сеть состоит из двух n-МОП-устройств, соединенных последовательно, которые проводят, когда сигналы A и B имеют высокий уровень. Нагрузочная сеть является дуальной к данной и состоит из двух включенных параллельно p-МОП-транзисторов. Это означает, что F равно 1, если $A = 0$ или $B = 0$, что эквивалентно функции $F = \overline{A \cdot B}$. Таблица истинности для данного простого элемента НЕ-И с двумя входами приведена в табл. 6.1. Можно проверить, что выход F всегда соединен с V_{DD} или с землей, но никогда — с обоими одновременно.

Пример 6.2. Создание сложных КМОП-схем

Используя комплементарную КМОП-логику, рассмотрим создание сложного КМОП-элемента, действие которого представляется функцией $F = \overline{D + A \cdot (B + C)}$. Как показано на рис. 6.6, *а*, первым этапом разработки логического элемента является создание разгрузочной сети (помним, что после-

Таблица 6.1. Таблица истинности для элемента НЕ-И с двумя входами

A	B	F
0	0	1
0	1	1
1	0	1
1	1	0

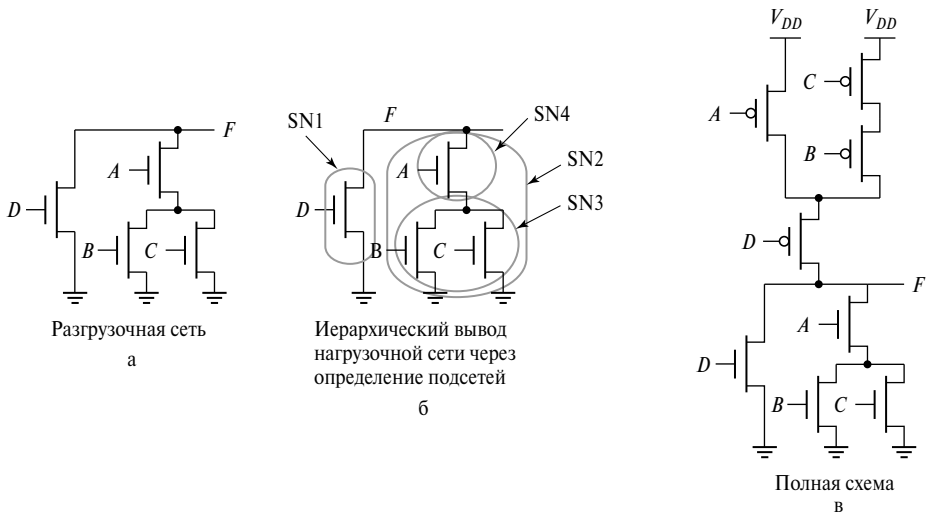


Рис. 6.6. Сложный комплементарный МОП-элемент

довательно включенные n-МОП-устройства реализуют функцию И, а включенные параллельно — функцию ИЛИ). Следующий этап — иерархический вывод нагрузочной сети из соображений дуальности. Разгрузочная сеть разбивается на меньшие сети (т.е. подмножества разгрузочной сети), называемые *подсетями*, что облегчает вывод нагрузочной сети. На рис. 6.6, *б* показаны подсети (subnet — SN) для разгрузочной сети. На верхнем уровне SN1 и SN2 включены параллельно, так что в дуальной сети они будут идти последовательно. Поскольку SN1 состоит из одного транзистора, она отображается непосредственно в нагрузочную сеть. С другой стороны, мы должны последовательно применить правила дуальности к SN2. Внутри SN2 мы имеем последовательное соединение подсетей SN3 и SN4, поэтому в нагрузочной сети они будут представлены параллельно. Наконец, внутри SN3 устройства соединены параллельно, поэтому в нагрузочной сети они будут включены последовательно. Полный элемент показан на рис. 6.6, *в*. Можно проверить, что для любой возможной комбинации входов всегда существует путь либо к V_{DD}, либо к земле.

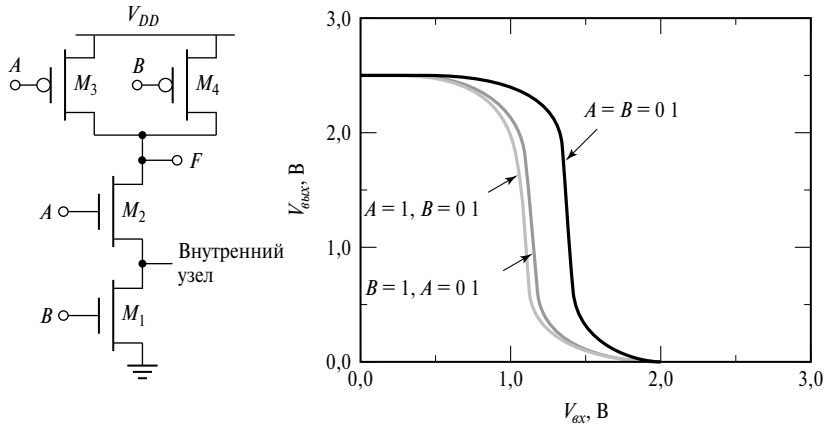


Рис. 6.7. Передаточная функция напряжения элемента НЕ-И с двумя входами зависит от данных (n-МОП-устройства имеют размеры 0,5/0,25 мкм, а p-МОП-устройства — 0,75/0,25 мкм)

Статические свойства комплементарных МОП-элементов

Комплементарные МОП-элементы наследуют все прекрасные свойства стандартного КМОП-инвертора. Полный размах сигнала составляет $V_{OH} = V_{DD}$ и $V_{OL} = GND$. Кроме того, схемы не имеют статического рассеяния энергии, поскольку проектируются так, чтобы разгрузочная и нагрузочная сети были взаимно исключающими. Анализ передаточной характеристики напряжения постоянного тока и запаса помехоустойчивости здесь будет сложнее, чем для инвертора, поскольку указанные параметры **зависят от поданных на вход элемента наборов данных**.

Рассмотрим статический элемент НЕ-И с двумя входами, показанный на рис. 6.7. Переключение выходного сигнала элемента с высокого уровня на низкий происходит при трех возможных комбинациях входных сигналов: (а) $A = B = 0 \rightarrow 1$, (б) $A = 1$, $B = 0 \rightarrow 1$ и (в) $B = 1, A = 0 \rightarrow 1$. Получающиеся в результате передаточные функции напряжения существенно отличаются. Сильное отличие случая (а) от двух других (б и в) объясняется тем, что при $A = B = 0$ оба транзистора нагрузочной сети включены одновременно, т.е. имеют сильную нагрузку. В случаях (б) и (в) включено только одно из нагрузочных устройств. Смещение функции влево объясняется более слабой нагрузочной сетью.

Разница между (случаями б и в) проистекает исключительно от состояния внутреннего узла *int* между двумя n-МОП-устройствами. Чтобы n-МОП-устройства включились, оба напряжения затвор-исток должны превышать V_{Tn} при $V_{GS2} = V_A - V_{DS1}$ и $V_{GS1} = V_B$. Пороговое напряжение транзистора M_2 будет больше, чем у транзистора M_1, из-за влияния подложки. Пороговые напряжения двух устройств представляются следующими уравнениями:

$$V_{Tn2} = V_{Tn0} + \gamma \left((\sqrt{|2\phi_f| + V_{int}}) - \sqrt{|2\phi_f|} \right); \qquad (6.1)$$

$$V_{Tn1} = V_{Tn0}. \tag{6.2}$$

В случае (б) устройство M_3 *выключено* и напряжение элемента M_2 устанавливается равным V_{DD}. В первом приближении M_2 можно рассматривать как резистор, последовательно подключенный к M_1. Поскольку напряжение питания M_2 велико, данное сопротивление мало и имеет небольшой эффект на передаточную функцию напряжения. В случае (в) транзистор M_1 выступает как резистор, приводя к увеличению V_T на M_2 из-за влияния подложки. Как видно на графике, общее воздействие при этом достаточно мало.

Проектные соображения

Из предыдущего обсуждения следует вынести одну важную мысль: **запас помехоустойчивости зависит от входа/набора данных**. В примере 6.2 мы видели, что сбой в одном из двух входных каналов имеет бо́льшие шансы породить ложный переход на выходе, чем одновременный сбой в обоих каналах. Следовательно, в первом случае имеем меньших запас помехоустойчивости по малому шуму. В связи с этим часто поступают следующим образом: при разработке логических элементов, подобных схемам НЕ-И и НЕ-ИЛИ, соединяют вместе все входы. К сожалению, это не самый худший из возможных сценариев статического поведения; в любом случае зависимость функционирования схемы от данных следует изучать очень тщательно.

Задержка распространения комплементарных МОП-элементов

Расчет задержки распространения выполняется по схеме, использованной для статического инвертора. С целью анализа задержки каждый транзистор моделируется как резистор, последовательно соединенный с идеальным ключом. Поскольку значение сопротивления зависит от напряжения источника питания, необходимо использовать эквивалентное большое сопротивление сигнала, масштабированное с коэффициентом, равным отношению длины устройства к его ширине. Логика преобразуется в эквивалентную RC-цепь, учитывающую эффект внутренних емкостей узлов. Рассмотрим рис. 6.8, где изображены элемент НЕ-И с двумя входами и эквивалентная RC-модель в виде ключа. Обратите внимание на включение в схему внутренней емкости узла C_{int} (порожденной областями истока/стока и перекрывающейся емкости элементов M_2 и M_1). Хотя это и усложняет анализ, внутренняя емкость узла может играть весьма значительную роль, которой нельзя пренебрегать в определенных структурах (например, в элементах с большим разветвлением по входу). В первом приближении влияние внутренней емкости мы будем игнорировать.

Простой анализ модели показывает, что, подобно запасу помехоустойчивости, **задержка распространения зависит от структуры входных данных**. Рассмотрим, например, изменение уровня сигнала от низкого к

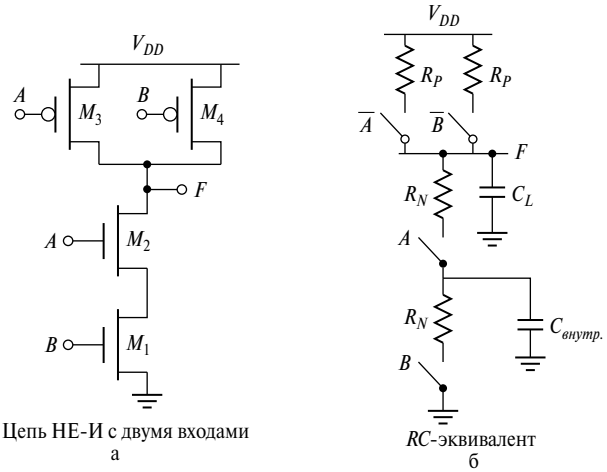


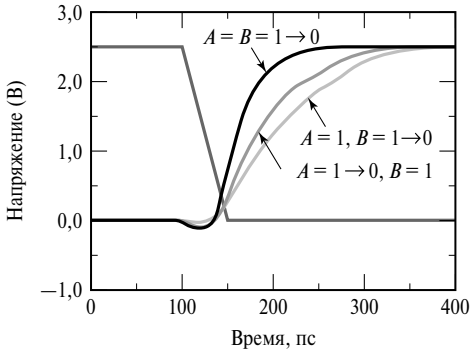
Рис. 6.8. Эквивалентная RC-модель элемента НЕ-И с двумя входами

высокому и изменение уровня сигнала от низкого к высокому. Возможны три сценария зарядки выхода до V_{DD}. Если на оба входа поступают низкие сигналы, включены два p-МОП-устройства. В этом случае задержка равна $0,69 \times (R_p/2) \times C_L$, поскольку два резистора включены параллельно. Это не самое худшее быстродействие при изменении уровня сигнала от низкого к высокому (самая плохая получается, когда включено только одно устройство, и равняется $0,69 \times R_p \times C_L$). Для разгрузочного тракта выход разряжается, только когда A и B переключены в высокое состояние и задержка (в первом приближении) равна $0,69 \times (2R_N) \times C_L$. Другими словами, последовательное добавление устройства замедляет цепь, поэтому, чтобы избежать падения быстродействия, устройства нужно делать шире. При масштабировании транзисторов элемента с несколькими входами необходимо подобрать такую комбинацию входов, которая инициирует самый худший вариант, и проводить расчеты, исходя из соответствующего значения задержки.

Чтобы элемент НЕ-И имел такую же задержку разгрузки (t_{phl}), как и инвертор минимального размера, n-МОП-устройства в разгрузочной сети должны быть вдвое шире, тогда эквивалентные сопротивления разгрузочной сети НЕ-И будут такими же, как у инвертора. Устройства p-МОП могут оставаться незаряженными[1].

Выполняя данный анализ в первом приближении, мы предполагали, что дополнительную емкость, вводимую из-за увеличения ширины транзисторов, можно игнорировать. В общем случае такое предположение не очень хорошее, однако сейчас оно позволяет разумно оценить, как следует масштабировать устройства.

[1] В глубоко субмикронном процессе транзисторы не нужно делать настолько широкими. Для элемента НЕ-И с двумя входами n-МОП-транзисторы должны быть шире не в два, а в полтора раза.



Структура входных данных	Задержка (пс)
$A = B = 0 \rightarrow 1$	69
$A = 1, B = 0 \rightarrow 1$	62
$A = 0 \rightarrow 1, B = 1$	50
$A = B = 1 \rightarrow 0$	35
$A = 1, B = 1 \rightarrow 0$	76
$A = 1 \rightarrow 0, B = 1$	57

Рис. 6.9. Пример, демонстрирующий зависимость задержки от структуры входных сигналов

Пример 6.3. Зависимость задержки от структуры входных данных

Рассмотрим элемент НЕ-И, показанный на рис. 6.8, *а*. Предположим, что n-МОП- и p-МОП-устройства имеют размеры 0,5/0,25 мкм и 0,75/0,25 мкм соответственно. При таком выборе размеров мы должны получить примерно равные времена нарастания и спада для худшего сценария (поскольку эффективное сопротивление разгрузочной сети равно сопротивлению нагрузочной).

На рис. 6.9 представлена задержка при изменении уровня сигнала с низкого до высокого для различных структур входных данных. Как и ожидалось, ситуация, когда на обоих входах уровень сигнала меняется с высокого на низкий ($A = B = 1 \rightarrow 0$), дает меньшую задержку по сравнению со случаем, когда понижение уровня сигнала происходит только на одном входе. Обратите внимание на то, что худший сценарий повышения уровня сигнала зависит от того, на каком входе (A или B) уровень сигнала понижается. Причина здесь кроется во внутренней емкости узла разгрузочного участка (т.е. истока транзистора M_2). В случае, когда $B = 1$, а на входе A уровень сигнала понижается ($1 \rightarrow 0$), нагрузочное p-МОП-устройство должно зарядить только емкость выходного узла (транзистор M_2 выключен). С другой стороны, в случае, когда $A = 1$, а на входе B выполняется переход $1 \rightarrow 0$, нагрузочное p-МОП-устройство должно зарядить две емкости (выходную и внутреннюю), что замедляет переход.

Различные суммарные задержки распространения сигнала через данную цепь показаны в таблице, представленной на рис. 6.9. Масштабирование транзистора первого порядка действительно дает примерно равные задержки нарастания и спада. Здесь важно отметить, что задержка распространения при изменении уровня сигнала от низкого к высокому зависит от исходного состояния внутренних узлов. Например, когда на обоих входах происходит переход $0 \rightarrow 1$, важно установить состояние внутреннего узла. Худшим является случай, когда внутренний узел изначально заряжен до $V_{DD} - V_{Tn}$, что можно получить, выполнив на входе A переход $1 \rightarrow 0 \rightarrow 1$, чтобы на входе B при этом происходил только переход $0 \rightarrow 1$. При такой подготовке внутренний узел будет инициализирован правильно.

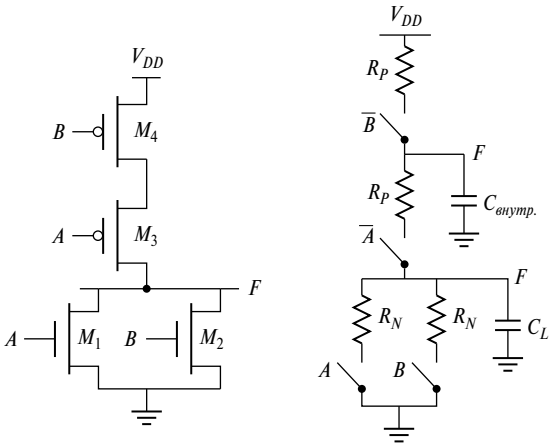


Рис. 6.10. Выбор размеров элемента НЕ-ИЛИ

Из данного примера вы должны вынести, что оценка задержки может быть очень сложной и требовать тщательного рассмотрения внутренних емкостей и структур данных. В любом случае в соответствующем пакете обязательно нужно смоделировать наихудший сценарий. Помните, что лобовой метод перебора всех возможных входных структур может не сработать, поскольку, кроме этого, важно учитывать состояние внутренних узлов.

На рис. 6.10 показана КМОП-реализация элемента НЕ-ИЛИ ($F = \overline{A + B}$). На выход данной сети поступает сигнал высокого уровня тогда и только тогда, когда на оба входа, A и B, подан сигнал низкого уровня. Худший сценарий разгрузочного перехода реализуется тогда, когда включено только одно из n-МОП-устройств (т.е. если либо A, либо B получают на вход сигнал высокого уровня). Предположим, что нам необходимо так выбрать размеры элемента НЕ-ИЛИ, чтобы он имел приблизительно такую же задержку, как и инвертор со следующими размерами транзисторов: n-МОП — 0,5/0,25 мкм, p-МОП — 1,5/0,25 мкм. Поскольку разгрузочный тракт в худшем случае представляет собой одно устройство, n-МОП-устройства (M_1 и M_2) могут иметь такую же ширину, как и n-МОП-устройство инвертора. Чтобы на выходе получился сигнал высокого уровня, оба устройства должны включиться. Так как сопротивления складываются, устройства должны быть в два раза больше p-МОП-транзистора инвертора (т.е. M_3 и M_4 должны иметь размер 3/0,25 мкм). Поскольку p-МОП-устройства характеризуются меньшей подвижностью, чем n-МОП-устройства, последовательного включения устройств следует по возможности избегать. При реализации общих логических схем, очевидно, реализация элемента НЕ-И является более предпочтительной, чем реализация элемента НЕ-ИЛИ.

Задача 6.1. Выбор размеров транзисторов в комплементарных МОП-элементах

Определите размеры транзисторов, показанных на рис. 6.6, *в*, чтобы время t_{plh} и t_{phl} было приблизительно таким, как у инвертора со следующими размерами: n-МОП — 0,5/0,25 мкм, p-МОП — 1,5/0,25 мкм.

До этого момента при анализе задержки распространения мы игнорировали эффект внутренних емкостей узла. Часто такое предположение разумно для анализа в первом приближении. Тем не менее в более сложных логических элементах с большими разветвлениями по входу собственные емкости узлов могут быть значительными. Рассмотрим, например, элемент НЕ-И с четырьмя входами, показанный на рис. 6.11 вместе с эквивалентной RC-моделью, которая включает внутренние емкости узлов. Внутренние емкости состоят из емкости pn-перехода транзисторов, а также емкостей затвора-стока и затвора-истока. Последние согласно равенству Миллера превращаются в емкости на землю. Анализ задержки для подобной цепи включает расчет распределенных RC-цепей, с которым мы уже сталкивались при анализе задержки соединительных сетей. Рассмотрим разгрузочную задержку схемы. Разрядка выхода происходит, когда на все входы подается сигнал высокого уровня. Прежде чем поднимать уровень сигнала на входе, нам нужно правильно выставить исходные состояния внутренних узлов (т.е. внутренние узлы должны зарядиться до $V_{DD}V_{TN}$).

Задержку распространения можно рассчитать, используя модель Элмора:

$$\begin{aligned} t_{pHL} =& 0,69(R_1 \cdot C_1 + (R_1 + R_2) \cdot C_2 + \\ &+ (R_1 + R_2 + R_3) \cdot C_3 + (R_1 + R_2 + R_3 + R_4) \cdot C_L). \end{aligned} \tag{6.3}$$

Обратите внимание на то, что сопротивление транзистора M_1 появляется во всех членах, т.е. данное устройство особенно важно при расчетах, направленных на минимизацию задержки. Предполагая, что все n-МОП-устройства имеют одинаковый размер, уравнение (6.3) можно упростить до следующего вида:

$$t_{pHL} = 0,69R_N(C_1 + 2 \cdot C_2 + 3 \cdot C_3 + 4 \cdot C_L). \tag{6.4}$$

Пример 6.4. КМОП-элемент НЕ-И с четырьмя входами

В данном примере мы вычислим *собственную* (или *ненагруженную*) задержку распространения элемента НЕ-И с четырьмя входами с использованием расчетов вручную и моделирования. Схема элемента показана на рис. 6.12. Предположим, что для всех n-МОП-устройств $W/L = 0,5/0,25$ мкм, а для всех p-МОП-устройств $W/L = 0,375/0,25$ мкм. Размеры устройств выбраны так, что в худшем случае время нарастания и спада были в первом приближении равны (внутренние емкости узлов игнорируются).

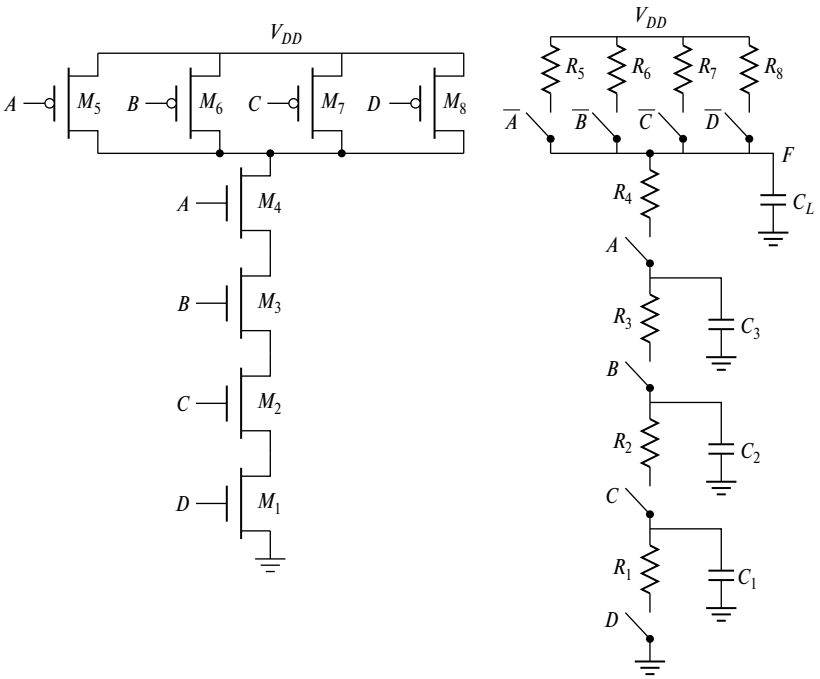


Рис. 6.11. Элемент НЕ-И с четырьмя входами и его RC-модель

С помощью техник, подобных использованным в главе 5 для КМОП-инвертора, мы можем рассчитать по схеме значения емкостей. Обратите внимание на то, что в нагрузочном тракте p-МОП-устройства совестно используют терминал стока, чтобы уменьшить общий паразитный вклад. Применяя стандартные правила разработки, находим, что площадь и периметр различных устройств имеют значения, представленные в табл. 6.2.

В данном примере мы сконцентрируем внимание на разгрузочной задержке и рассчитаем емкости для изменения на выходе уровня сигнала от высокого к низкому. Когда на выходе происходит изменение сигнала от V_{DD} до 0, на внутренних узлах напряжение меняется всего лишь с $V_{DD} - V_{Tn}$ до GND. Для указанного изменения напряжения мы должны линеаризовать внутреннюю емкость pn-перехода, однако для упрощения анализа мы используем для внутренних узлов такое же значение K_{eff}, как и для выходного узла.

Предполагается, что выход подается на один инвертор минимального размера. Малое влияние внутриячеечной трассировки игнорируется. Различные вклады в емкость сведены в табл. 6.3. Для pn-переходов n-МОП- и p-МОП-транзисторов используются значения $K_{eq} = 0,57$ и $K_{eqsw} = 0,61$; $K_{eq} = 0,79$ и $K_{eqsw} = 0,86$ соответственно. Обратите внимание на то, что емкость затвора-стока удваивается для выходного и всех внутренних узлов с целью учета эффекта Миллера. (Здесь игнорируется тот факт, что внутренние узлы имеют немного меньший размах напряжения из-за падения порога.)

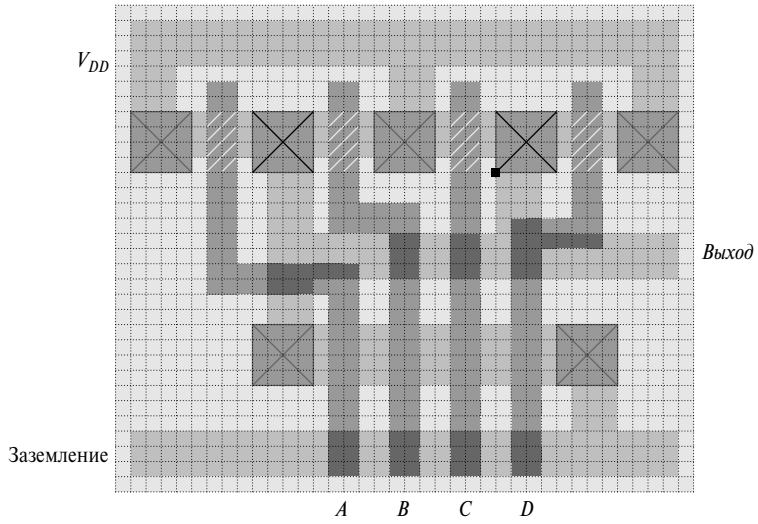


Рис. 6.12. Схема элемента НЕ-И с четырьмя входами на КМОП

Таблица 6.2. Площадь и периметр транзисторов элемента НЕ-И с четырьмя входами

Транзистор	W (мкм)	AS (мкм^2)	AD (мкм^2)	PS (мкм)	PD (мкм)
1	0,5	0,3125	0,0625	1,75	0,25
2	0,5	0,0625	0,0625	0,25	0,25
3	0,5	0,0625	0,0625	0,25	0,25
4	0,5	0,0625	0,3125	0,25	1,75
5	0,375	0,297	0,172	1,875	0,875
6	0,375	0,172	0,172	0,875	0,875
7	0,375	0,172	0,172	0,875	0,875
8	0,375	0,297	0,172	1,875	0,875

Используя уравнение (6.4), мы следующим образом рассчитываем задержку распространения:

$$t_{pHL} = 0,69\left(\frac{13\text{ кОм}}{2}\right)(0,85\text{ фФ}+2{\cdot}0,85\text{ фФ}+3{\cdot}0,85\text{ фФ}+4{\cdot}3,47\text{ фФ}) = 85\text{ пс}.$$

Смоделированная задержка для данного перехода равна 86 пс! Несмотря на все предположения и линеаризацию, расчеты вручную дают довольно точную оценку. Например, мы предположили, что емкость затвора-истока (или затвора-стока) состоит только из перекрывающегося компонента. Это не совсем справедливо, поскольку при изменении уровня сигнала вместо них следует учитывать некоторые другие вклады (в зависимости от рабочей области). Повторимся, цель ручного анализа — не получить абсолютно точную оценку задержки, а разобраться, какие факторы влияют на задержку, и помочь определить стартовые размеры транзистора. Точный временной анализ

Таблица 6.3. Расчет емкостей для изменения на выходе уровня сигнала от высокого к низкому. В таблице представлена собственная задержка элемента без дополнительной нагрузки. Любая емкость на выходе просто добавится к члену C_L

Конденсатор	Вклады ($H \to L$)	Значение (фФ) ($H \to L$)
C_1	$C_{d1} + C_{s2} + 2 * C_{gd1} + 2 * C_{gs2}$	$(0,57 * 0,0625 * 2 + 0,61 * 0,25 * 0,28)$ + $(0,57 * 0,0625 * 2 + 0,61 * 0,25 * 0,28)$+ $2 * (0,31 * 0,5) + 2 * (0,31 * 0,5) = 0,85$ фФ
C_2	$C_{d2} + C_{s3} + 2 * C_{gd2} + 2 * C_{gs3}$	$(0,57 * 0,0625 * 2 + 0,61 * 0,25 * 0,28)$ + $(0,57 * 0,0625 * 2 + 0,61 * 0,25 * 0,28)$+ $2 * (0,31 * 0,5) + 2 * (0,31 * 0,5) = 0,85$ фФ
C_3	$C_{d3} + C_{s4} + 2 * C_{gd3} + 2 * C_{gs4}$	$(0,57 * 0,0625 * 2 + 0,61 * 0,25 * 0,28)$ + $(0,57 * 0,0625 * 2 + 0,61 * 0,25 * 0,28)$+ $2 * (0,31 * 0,5) + 2 * (0,31 * 0,5) = 0,85$ фФ
C_L	$C_{d4} + 2 * C_{gd4} + C_{d5} + C_{d6} + C_{d7} + C_{d8}$ $+2 * C_{gd5} + 2 * C_{gd6} + 2 * C_{gd7} + 2 * C_{gd8}$ $= C_{d4} + 4 * C_{d5} + 4 * 2 * C_{gd6}$	$(0,57 * 0,3125 * 2 + 0,61 * 1,75 * 0,28) + 2 *$ $(0,31 * 0,5) + 4 * (0,79 * 0,171875 * 1,9 +$ $0,86 * 0,875 * 0,22) + 4 * 2 * (0,27 * 0,375) =$ $3,47$ фФ

и оптимизация транзистора обычно выполняются с использованием SPICE. Задержка при изменении уровня сигнала с низкого до высокого, полученная по результатам моделирования, равна 106 пс.

Хотя комплементарные МОП-структуры — очень устойчивые и простые средства для реализации логических элементов, с увеличением сложности элемента (т.е. *разветвления по входу*) возникают две серьезные проблемы. Во-первых, число транзисторов, требующихся для реализации элемента с N входами, равно $2N$. В результате можно получить очень большую площадь реализации. Во-вторых, задержка распространения комплементарного МОП-элемента быстро ухудшается с увеличением разветвления по выходу. Фактически *ненагруженная собственная задержка* элемента в худшем случае является *квадратичной функцией разветвления по входу.*

- Большое число транзисторов ($2N$) увеличивает общую емкость элемента. Для элемента с N входами *собственная емкость* линейно увеличивается с ростом разветвления по входу. Рассмотрим, например, элемент НЕ-И, показанный на рис. 6.11. При линейном увеличении числа p-МОП-устройств, соединенных с выходным узлом, мы ожидаем, что задержка элемента при изменении уровня сигнала от низкого к высокому будет увеличиваться линейно с ростом разветвления по входу, — при линейном росте емкости нагрузочное сопротивление остается без изменений.
- Последовательное соединение транзисторов в нагрузочной либо разгрузочной сети элемента приводит к дополнительному замедлению. Мы знаем, что *распределенная RC-цепь* в разгрузочной сети, показанной на рис. 6.11, дает задержку, квадратично зависящую от числа элементов в цепочке. Следовательно, задержка при изменении уровня сигнала от высокого к низкому должна быть квадратичной функцией разветвления по входу.

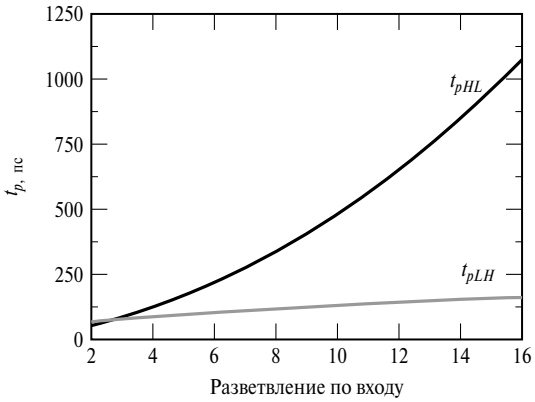


Рис. 6.13. Задержка распространения КМОП-элемента НЕ-И как функция разветвления по входу. Разветвление по выходу инвертора считается равным 1, все разгрузочные транзисторы имеют минимальный размер

На рис. 6.13 представлена (собственная) задержка распространения элемента НЕ-И как функций разветвления по входу в предположении, что разветвление по выходу инвертора фиксирована (n-МОП — 0,5 мкм и p-МОП — 1,5 мкм). Как и предсказывалось, t_{pLH} является линейной функцией разветвления по входу, тогда как одновременное увеличение сопротивления разгрузки и емкости нагрузки дает примерно квадратическую зависимость t_{pHL}. Элементы, разветвление по входу которых больше или равно 4, становятся чересчур медленным, поэтому использовать их не следует.

Технологии разработки для больших разветвлений по входу

Разработчик имеет в своем распоряжении несколько технологий, позволяющих уменьшить задержку схем с большим разветвлением по входу.

- **Масштабирование транзистора.** Наиболее очевидным решением проблемы является увеличение размеров транзистора. Это уменьшает сопротивление устройств, соединенных последовательно, и уменьшает постоянные времени. Тем не менее увеличение размеров устройства дает бо́льшие паразитные емкости, которые не только влияют на *задержку распространения* рассматриваемого элемента, но также представляют бо́льшую нагрузку для предыдущего элемента. Следовательно, данную технологию необходимо использовать с осторожностью. Если емкость нагрузки определяется в основном собственной емкостью элемента, увеличение ширины устройства создает только эффект “самонагрузки”, и на *задержку распространения* это не влияет. Масштабирование эффективно только тогда, когда нагрузка определяется в основном разветвлением по выходу. В следующем разделе рассматривается более комплексный подход к масштабированию транзисторов в сложных комбинационных КМОП-сетях.

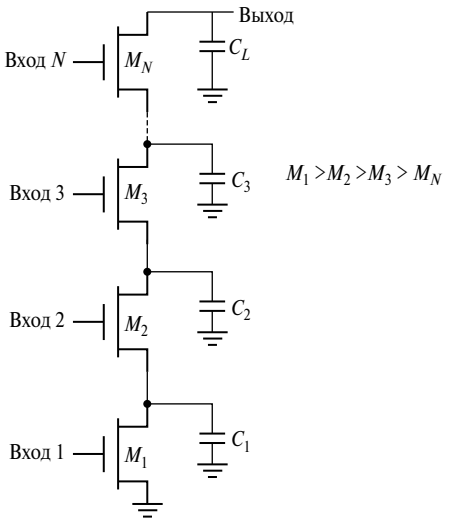


Рис. 6.14. Прогрессивное масштабирование транзисторов в большой цепочке транзисторов с дополнительной нагрузкой в виде собственных емкостей

- **Прогрессивное масштабирование транзисторов.** Альтернативой равномерного (пропорционального) масштабирования всех транзисторов является использование прогрессивного масштабирования (рис. 6.14). Возвращаясь к уравнению (6.3), видим, что сопротивление $M_1(R_1)$ входит в уравнение задержки N раз, сопротивление $M_2(R_2)$ — $N-1$ раз, и т.д. Из уравнения видно, что сопротивление R_1 нужно сделать самым маленьким, R_2 — чуть бо́льшим, и т.д. Следовательно, нам выгодно прогрессивное масштабирование транзисторов: $M_1 > M_2 > M_3 > M_N$. Данный подход уменьшает доминирующее сопротивление, при этом удерживая рост емкости в пределах дозволенного. Прекрасный разбор оптимального масштабирования транзисторов в сложной сети дан в [Shoji88, p. 131–143]. В любом случае вам следует знать об одном важном подводном камне данного подхода. Хотя прогрессивное масштабирование транзисторов сравнительно просто выглядит на диаграмме, в реальной жизни все не так просто. Довольно часто из соображений проектных норм разработчик вынужден раздвигать транзисторы, что приводит к росту внутренней емкости. Помните, что это может свести на нет весь выигрыш от масштабирования!
- **Переупорядочение входа.** Некоторые сигналы в сложных блоках комбинационной логики могут быть более критичными, чем другие. Не все входы элемента поступают в один момент времени (например, из-за задержек распространения предыдущих логических элементов). Входной сигнал элемента называется *критическим*, если это последний сигнал из всех входов, требуемый для приобретения постоянного значения. Путь через логику, определяющий предельную скорость структуры, называется *критическим путем*. Поместив транзисторы критического пути ближе к выходу элемента, мы можем ускорить его работу, как видно из рис. 6.15.

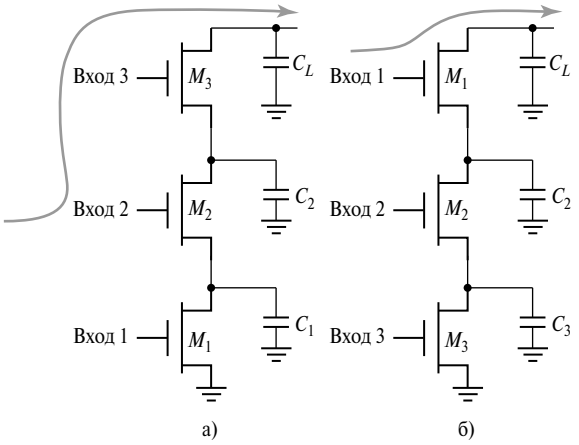


Рис. 6.15. Влияние порядка транзисторов на задержку. Сигнал In_1 является критическим

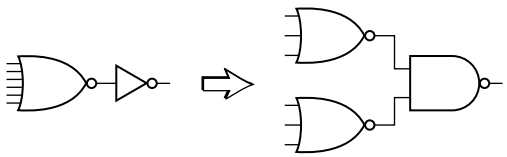


Рис. 6.16. С помощью реструктуризации логики можно уменьшить разветвление элемента по входу

Предположим, что сигнал In_1 — критический. Далее предположим, что In_2 и In_3 имеют высокий уровень, а In_1 выполняет переход $0 \rightarrow 1$. Допустим, что изначально емкость C_L заряжена до высокого уровня. В первом случае пути к земле не существует, пока не включится транзистор M_1, что, к сожалению, является самым последним событием. Следовательно, задержка между поступлением сигнала In_1 на вход и изменением выхода определяется временем, требуемым для разрядки C_L, C_1 и C_2. Во втором случае C_1 и C_2 уже разряжены при изменении In_1. Все, что требуется, — разрядить C_L, т.е. задержка в этом случае будет меньше.

- **Реструктуризация логики.** Переупорядочив уравнение логической схемы, мы можем снизить требования к разветвлению по входу, а следовательно, уменьшить задержку элемента, как показано на рис. 6.16. Из-за квадратичной зависимости задержки схемы от разветвления по входу элемент НЕ-ИЛИ с шестью входами будет крайне медленным. Разбив его на два элемента с тремя входами, получим существенное ускорение работы, которое с запасом компенсирует дополнительную задержку, появившуюся из-за превращения инвертора в элемент НЕ-И с двумя входами.

Оптимизация быстродействия в комбинационных сетях

Ранее мы установили, что минимизация задержки распространения элемента самого по себе представляет чисто академический интерес. Масштабирование устройств следует производить в контексте той схемы, где планируется использовать устройство. В главе 5 мы разработали методологию соответствующих расчетов для инвертора. Кроме того, мы нашли, что оптимальное разветвление по выходу цепочки инверторов, питающей нагрузку C_L, равно $(C_L/C_{in})^{1/N}$, где N — число каскадов цепочки; C_{in} — входная емкость первого элемента цепочки. Если мы вольны выбирать число каскадов, тогда число разветвлений по выходу на каскад стоит положить равным 4. Можно ли распространить данный результат на задачу определения размеров любого комбинационного пути, если требуется получить минимальную задержку? Развивая описанный ранее подход на случай сложных логических сетей, мы находим, что это действительно возможно [Sutherland99][2].

Для этого мы модифицируем стандартное уравнение задержки инвертора, которое мы ввели в главе 5,

$$t_p = t_{p0}\left(1 + \frac{C_{ext}}{\gamma C g}\right) = t_{p0}(1 + f/\gamma), \tag{6.5}$$

к такому виду:

$$t_p = t_{p0}(p + gf/\gamma). \tag{6.6}$$

В этих уравнениях t_{p0} представляет собственную задержку инвертора, а f — *эффективное разветвление по выходу*, определяемое как отношение внешней нагрузки к внутренней емкости элемента. В таком контексте f также называют *электрическими трудозатратами* (electrical effort), а p представляет отношение собственной (без нагрузки) задержки сложного элемента и простого инвертора и является функцией топологии элемента и стиля разработки топологии. Внутренняя задержка элементов с несколькими входами, которые имеют более сложную структуру, выше задержки инвертора. Для примера в табл. 6.4 перечислены значения p для некоторых стандартных элементов (предполагаются простые стили разработки топологии и игнорируются эффекты второго порядка, подобные собственным емкостям узлов).

Параметр g называется *логическими трудозатратами* (logical effort) и выражает тот факт, что для данной нагрузки сложные элементы должны работать интенсивнее, чем инвертор, чтобы дать подобную характеристику. Другими словами, логические трудозатраты логического элемента показывают, насколько хорош данный элемент в производстве такого же тока на выходе, как и инвертор, при условии, что все входы элемента могут дать только такую же входную емкость, как и инвертор. Сказанное можно сформулировать и иначе: логические трудозатраты показывают, насколько больше

[2] Подход, введенный в данном разделе, формально был представлен в [Sutherland99], где данная тема рассматривалась всесторонне. Изложение, представленное в данном разделе, является беглым обзором общего подхода.

Таблица 6.4. Оценки собственных задержек различных типов логических схем (предполагается использование простых топологий и фиксированного отношения размеров p-МОП к n-МОП

Тип элемента	p
Инвертор	1
НЕ-И с n входами	n
НЕ-ИЛИ с n входами	n
n-канальный мультиплексор	$2n$
ИСКЛЮЧАЮЩЕЕ ИЛИ, НЕ-ИСКЛЮЧАЮЩЕЕ ИЛИ	$n2^{n-1}$

Таблица 6.5. Логические трудозатраты распространенных логических элементов (отношение размеров p-МОП-транзистора к n-МОП-транзистору равно 2)

	Число входов			
Тип элемента	1	2	3	n
Инвертор	1			
НЕ-И		$4/3$	$5/3$	$(n+2)/3$
НЕ-ИЛИ		$5/3$	$7/3$	$(2n+1)/3$
Мультиплексор		2	2	2
ИСКЛЮЧАЮЩЕЕ ИЛИ		4	12	

входной емкости должен иметь логический элемент, чтобы подать на выход такой же ток, как и инвертор. Логические трудозатраты являются очень полезным параметром, поскольку зависят только от топологии схемы. Значения этого параметра для некоторых распространенных логических элементов представлены в табл. 6.5.

Пример 6.5. Логические трудозатраты сложных элементов

Рассмотрим логический элемент, показанный на рис. 6.17. Предположим, что отношение размеров p-МОП-транзистора к n-МОП-транзистору равно, входная емкость симметричного инвертора минимального размера втрое превышает емкость n-МОП-транзистора минимального размера (обозначается C_{unit}). Размеры транзисторов, входящих в элементы НЕ-И и НЕ-ИЛИ, выбираются так, чтобы эквивалентные сопротивления были равны сопротивлению инвертора (при расчете используется технология, описанная ранее). Благодаря этому входная емкость элемента НЕ-И с двумя входами увеличивается до $4C_{unit}$, или $4/3$ емкости инвертора. Входная емкость элемента НЕ-ИЛИ с двумя входами равна $5/3$ от емкости инвертора. Аналогично, чтобы элементы НЕ-И и НЕ-ИЛИ имели одинаковую входную емкость, они должны получать в $4/3$ и $5/3$ раз меньше напряжения, чем инвертор. Это влияет на компонент задержки, соответствующий нагрузке, увеличивая его во столько же раз (коэффициент *логических трудозатрат*). Следовательно, $g_{NAND} = 4/3$ и $g_{NOR} = 5/3$.

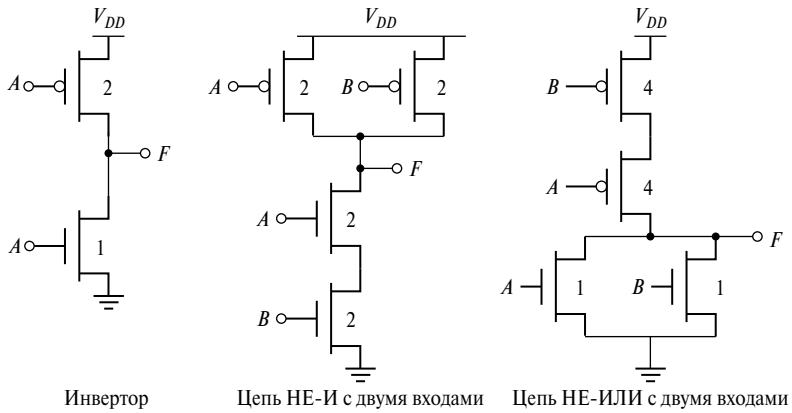


Рис. 6.17. Логические трудозатраты элементов НЕ-И и НЕ-ИЛИ с двумя входами

Рис. 6.18. Задержка как функция разветвления по выходу для инвертора и схемы НЕ-И с двумя входами

Модель задержки логического элемента, представленная в уравнении (6.6), является простой линейной зависимостью. На рис. 6.18 данная зависимость представлена графически: для инвертора и вентиля НЕ-И с двумя входами задержка показана как функция разветвления по выходу. Наклон линии равен логической работе вентиля; отрезок, отсекаемый на координатной оси, — внутренняя задержка. На графике видно, что мы можем регулировать задержку, настраивая эффективное разветвление по выходу (с помощью выбора размеров транзисторов) или выбирая логический вентиль с другой логической работой. Обратите внимание на то, что разветвление по выходу и логическая работа одинаково влияют на задержку. Произведение данных двух параметров, $h = fg$, называется *трудозатратами логического элемента* (gate effort).

Суммарную задержку тракта, проходящего через комбинационный логический элемент, можно выразить следующим образом:

$$t_p = \sum_{j=1}^{N} t_{p,j} = t_{p0} \sum_{j=1}^{N} \left(p_j + \frac{f_j g_j}{\gamma} \right). \tag{6.7}$$

Чтобы определить минимальную задержку тракта, воспользуемся процедурой, подобной использованной в главе 5 при рассмотрении цепочки инверторов. Находя $N-1$ частных производных и приравнивая их к нулю, мы находим, что *все каскады должны иметь одинаковые трудозатраты элементов*:

$$f_1 g_1 = f_2 g_2 = \ldots = f_N g_N. \tag{6.8}$$

Логические трудозатраты тракта в сети G находятся как произведение логических трудозатрат всех элементов вдоль тракта:

$$G = \prod_{1}^{N} g_i. \tag{6.9}$$

Кроме того, мы можем определить *эффективное разветвление по выходу всего тракта* (или *электрические трудозатраты* F, выражающееся через емкость нагрузки последнего элемента тракта и входную емкость первого элемента:

$$F = \frac{C_L}{C_{g1}}. \tag{6.10}$$

Чтобы выразить F через эффективные разветвления по выходу отдельных элементов, мы должны ввести еще один параметр, чтобы учесть логическое разветвление по выходу в сети. Когда разветвление по выходу происходит на выходе узла, часть доступного управляющего тока направляется вдоль тракта, который мы анализируем, а часть — отводится от этого тракта. Таким образом, мы приходим к понятию *трудозатрат на ветвление* (branching effort) b логического элемента на данном тракте:

$$b = \frac{C_{on-path} + C_{off-path}}{C_{on-path}}, \tag{6.11}$$

где $C_{on-path}$ — емкость нагрузки логического элемента вдоль рассматриваемого тракта; $C_{off-path}$ — емкость соединений, открывающих новый тракт. Обратите внимание на то, что если тракт не разветвляется (как в случае цепочки элементов), трудозатраты на ветвление равны 1. Трудозатраты на разветвление тракта определяются как произведение трудозатрат на разветвление всех отдельных каскадов вдоль рассматриваемого тракта:

$$B = \prod_{1}^{N} b_i. \tag{6.12}$$

Теперь мы можем выразить электрические трудозатраты тракта через электрические трудозатраты и трудозатраты на ветвление отдельных каскадов:

$$F = \prod_{1}^{N} \frac{f_i}{b_i} = \frac{\prod f_i}{B}. \tag{6.13}$$

Наконец, мы можем определить суммарные трудозатраты тракта H. Используя уравнение (6.13), получаем следующее выражение:

$$H = \prod_{1}^{N} h_i = \prod_{1}^{N} g_i f_i = GFB. \tag{6.14}$$

С этого момента мы будем придерживаться той же схемы анализа, что и для цепочки инверторов. Итак, трудозатраты логического элемента, минимизирующие задержку тракта, равны:

$$h = \sqrt[N]{H}, \tag{6.15}$$

а минимальная задержка вдоль тракта равна:

$$D = t_{p0}\left(\sum_{j=1}^{N} p_j + \frac{N(\sqrt[N]{H})}{\gamma}\right). \tag{6.16}$$

Обратите внимание на то, что собственная задержка тракта является функцией типов логических элементов тракта и на нее не влияет масштабирование элементов. Следовательно, масштабные коэффициенты отдельных элементов цепочки s_i можно вывести путем последовательного рассмотрения схемы от входа к выходу (или наоборот). Будем предполагать, что питающая способность элемента единичного размера такая же, как у инвертора минимального размера. Используя определение логических трудозатрат, можно сказать, что входная емкость этого элемента в g раз больше, чем у эталонного инвертора, которая равна C_{ref}. Обозначив через s_1 масштабный коэффициент первого элемента цепочки, получаем, что входная емкость цепочки C_{g1} равна $g_1 s_1 C_{ref}$. Учитывая трудозатраты на разветвление, находим, что входная емкость элемента 2 в (f_1/b_1) больше этого значения, т.е. равна:

$$g_2 s_2 C_{ref} = \left(\frac{f_1}{b_1}\right) g_1 s_1 C_{ref}. \tag{6.17}$$

Для i-го элемента цепочки получаем следующее значение:

$$s_i = \left(\frac{g_1 s_1}{g_i}\right) \prod_{j=1}^{i-1} \left(\frac{f_j}{b_j}\right). \tag{6.18}$$

Пример 6.6. Выбор размеров комбинационных логических элементов для получения минимальной задержки

Рассмотрим логическую цепь, показанную на рис. 6.19. Она может представлять критический тракт в более сложном логическом блоке. Выход данной цепи нагружен емкостью, в пять раз большей входной емкости первого элемента — инвертора минимального размера. Таким образом, эффективное разветвление по выходу тракта равно: $F = C_L/C_{g1} = 5$. Используя данные табл. 6.5, находим, что логические трудозатраты тракта равны:

$$G = 1 \times \frac{5}{3} \times \frac{5}{3} \times 1 = \frac{25}{9}.$$

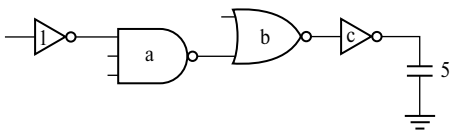


Рис. 6.19. Критический тракт вычислительной сети

Поскольку сеть не разветвляется, $B = 1$. Следовательно, $H = GFB = 125/9$ и оптимальные трудозатраты каскада h равны: $\sqrt[4]{H} = 1,93$. Учитывая типы логических элементов, мы находим следующие коэффициенты разветвления по выходу: $f_1 = 1,93$; $f_2 = 1,93 \times (3/5) = 1,16$; $f_3 = 1,16$; $f_4 = 1,93$. Обратите внимание на то, что коэффициенты инверторов больше, чем у более сложных элементов, поскольку они работают лучше при больших управляющих нагрузках.

Наконец, используя уравнение (6.18), находим размеры элементов (относительно устройств минимального размера). В итоге получаем следующие значения: $a = f_1 g_1/g_2 = 1,16$; $b = f_1 f_2 g_1/g_3 = 1,34$; $c = f_1 f_2 f_3 g_1/g_4 = 2,60$.

Приведенные расчеты не обязаны быть очень точными. Как обсуждалось в главе 5, масштабирование элемента до очень большого или очень малого размера с коэффициентом масштабирования 1,5 по-прежнему дает цепи, задержка которых отличается от минимальной не более чем на 5%. Следовательно, “сделанные левой ногой” расчеты с использованием описанной техники являются достаточно эффективными.

Задача 6.2. Масштабирование сети инверторов

Вернитесь к задаче 5.5 и решите ее с использованием подхода, учитывающего трудозатраты на ветвление.

Потребление мощности в КМОП-элементах

Источники потребления мощности в КМОП-инверторах подробно рассматривались в главе 5. Многие из поднятых вопросов напрямую переносятся на сложные КМОП-элементы. Рассеяние мощности сильно зависит от размеров транзистора (которые влияют на физическую емкость), времен нарастания/спада сигнала на выходе и выходе (которые определяют мощность короткого замыкания) порогов и температур устройств (которые влияют на мощность утечки) и активности переключения. Динамическое потребление мощности выражается как $\alpha_{0\to1}C_LV_{DD}^2f$. Усложнение элемента сказывается преимущественно на *активности переключения* $\alpha_{0\to1}$, которая состоит из двух компонентов: статического, зависящего только от топологии логической сети, и динамического, который порождается нестационарным поведением цепи.

Логическая функция

Активность переключения (или переходов) сильно зависит от логической функции, которую воплощает элемент. Для статических КМОП-элементов со статически независимыми входами статическая вероятность перехода представляет собой произведение вероятности p_0 того, что выход будет равен *нулю* в одном такте, на вероятность p_1 того, что в следующем такте выход будет равен *единице*:

$$\alpha_{0\to1} = p_0 \cdot p_1 = p_0 \cdot (1 - p_0). \tag{6.19}$$

Предполагая, что входы схемы независимы и распределены равномерно, любой статический элемент с N входами имеет следующую вероятность перехода:

$$\alpha_{0\to1} = \frac{N_0}{2^N} \cdot \frac{N_1}{2^N} = \frac{N_0 \cdot (2^N - N_0)}{2^{2N}}, \tag{6.20}$$

где N_0 — число *нулевых* позиций; N_1 — число *единичных* позиций в столбце выхода таблицы истинности функции. Чтобы проиллюстрировать сказанное, рассмотрим статический элемент НЕ-ИЛИ с двумя входами, таблица истинности которого приведена в табл. 6.6. Предположим, что во время одного такта возможен только один переход, а входы элемента НЕ-ИЛИ равновероятны (другими словами, четыре возможных комбинации входов A и B — 00, 01, 10, 11 — имеют одинаковую вероятность).

Таблица 6.6. Таблица истинности элемента НЕ-ИЛИ с двумя входами

A	B	Выход
0	0	1
0	1	0
1	0	0
1	1	0

Из табл. 6.6 и уравнения (6.20) получаем, что вероятность перехода на выходе КМОП-элемента НЕ-ИЛИ с двумя входами выражается следующим образом:

$$\alpha_{0\to 1} = \frac{N_0 \cdot (2^N - N_0)}{2^{2N}} = \frac{3 \cdot (2^2 - 3)}{2^{2\cdot 2}} = \frac{3}{16}. \tag{6.21}$$

Задача 6.3. Элемент ИСКЛЮЧАЮЩЕЕ ИЛИ с N входами

Предполагая, что входы элемента ИСКЛЮЧАЮЩЕЕ ИЛИ с N входами некоррелированны и равновероятны, выведите формулу для активности переключения этой схемы.

Статистические характеристики сигнала

Активность переключения логического элемента сильно зависит от статистических характеристик входного сигнала. Использовать для расчета активности равномерное распределение входов нецелесообразно, поскольку распространение сигнала через логические элементы может существенно модифицировать его статистические характеристики. Рассмотрим, например, статический элемент НЕ-ИЛИ с двумя входами и обозначим символами p_a и p_b вероятности того, что входы A и B равны *единице*. Далее предположим, что входы не коррелируют между собой. Вероятность получения 1 на выходе равна:

$$p_1 = (1 - p_a)(1 - p_b). \tag{6.22}$$

Следовательно, вероятность перехода $0 \to 1$ выражается так:

$$\alpha_{0\to 1} = p_0 p_1 = (1 - (1 - p_a)(1 - p_b))(1 - p_a)(1 - p_b). \tag{6.23}$$

На рис. 6.20 вероятность перехода представлена как функция функция p_a и p_b. Посмотрите, как этот график переходит в зависимость для простого инвертора, если одну из входных вероятностей положить равной 0. Из приведенного графика видно, что статистические характеристики сигнала и их влияние на переключения можно использовать для воздействия на рассеяние энергии.

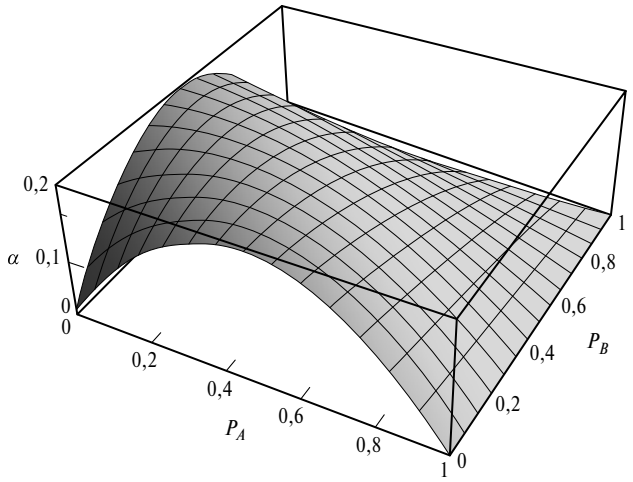


Рис. 6.20. Активность переходов элемента НЕ-ИЛИ с двумя входами как функция входных вероятностей (p_A, p_B)

Таблица 6.7. Вероятности переходов на выходе статических логических элементов

	$0 \to 1$
И	$(1 - p_A p_B) p_A p_B$
ИЛИ	$(1 - p_A)(1 - p_B)[1 - (1 - p_A)(1 - p_B)]$
ИСКЛЮЧАЮЩЕЕ ИЛИ	$[1 - (p_A + p_B - 2p_A p_B)](p_A + p_B - 2p_A p_B)$

Задача 6.4. Рассеяние мощности базовых логических элементов

Выведите выражения для вероятностей переходов $0 \to 1$ на выходах базовых логических элементов (И, ИЛИ, ИСКЛЮЧАЮЩЕЕ ИЛИ). В результате должны получиться зависимости, представленные в табл. 6.7.

Корреляция сигналов

Расчет активности переключения дополнительно усложняется тем, что сигналы коррелируют в пространстве и времени. Даже если первичные входы логической сети не коррелируют, распространение сигналов по сети приводит к их взаимной корреляции ("окрашиванию"). Лучше всего сказанное иллюстрирует простой пример. Рассмотрим схему, приведенную на рис. 6.21, *а*, и предположим, что первичные входы A и B не коррелируют и равны 1. Вероятность $\mathbf{1}(\mathbf{0})$ на узле C равна 1/2, а вероятность перехода $0 \to 1$ — 1/4. Вероятность того, что на узле Z произойдет переход с потреблением мощности, определяется с использованием формулы для элемента И из табл. 6.7:

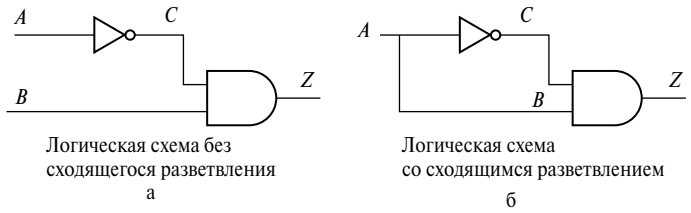


Рис. 6.21. Пример, демонстрирующий эффект корреляции сигналов

$$p_{0\to 1} = (1 - p_a p_b)p_a p_b = (1 - 1/2 \cdot 1/2)1/2 \cdot 1/2 = 3/16. \qquad (6.24)$$

Расчет вероятностей производится достаточно просто. Вначале последовательно (от входного узла к выходному) рассчитываются вероятности различных значений сигнала и вероятности переходов. Данный подход, однако, имеет два основных недостатка: он не работает для схем с обратной связью (поскольку разрабатывался для последовательных цепей), и в нем предполагается, что вероятности сигналов на входах всех элементов независимы. В реальных схемах такое предположение редко бывает справедливым, поскольку сходящиеся разветвления часто вызывают перекрестное воздействие сигналов друг на друга. Например, входы элемента И, показанного на рис. 6.21, *б* (C и B), взаимозависимы, поскольку оба являются функцией сигнала A. В таких случаях описанный выше подход просто не работает. Проход цепи от входов к выходу дает для узла Z вероятность перехода $3/16$, как и для предыдущего случая. Очевидно, данное значение является ошибочным, поскольку эквивалентная цепь представляется зависимостью $Z = C \cdot B = A \cdot \overline{A} = 0$, следовательно, переход вообще произойти не может.

Чтобы прогрессивный анализ давал точные результаты, необходимо учитывать взаимозависимость сигналов. Для этого можно использовать условные вероятности. Для элемента И Z равно 1 тогда и только тогда, когда B и C равны 1. Следовательно,

$$p_Z = p(Z = 1) = p(B = 1, C = 1), \qquad (6.25)$$

где $p(B = 1, C = 1)$ представляет вероятность того, что B и C одновременно равны 1. Если сигналы B и C независимы, то вероятность $p(B = 1, C = 1)$ можно разложить на $p(B = 1) \cdot p(C = 1)$, и мы приходим к полученному ранее выражению для элемента И: $p_Z = p(B = 1) \cdot p(C = 1) = p_B p_C$. Если два сигнала зависимы (как на рис. 6.21, *б*), необходимо рассчитывать условную вероятность:

$$p_Z = p(C = 1|B = 1) \cdot p(B = 1). \qquad (6.26)$$

Первый множитель в уравнении (6.26) представляет вероятность того, что $C = 1$ при условии $B = 1$. Данное дополнительное условие необходимо из-за того, что C зависит от B. Изучая сеть, находим, что данная вероятность

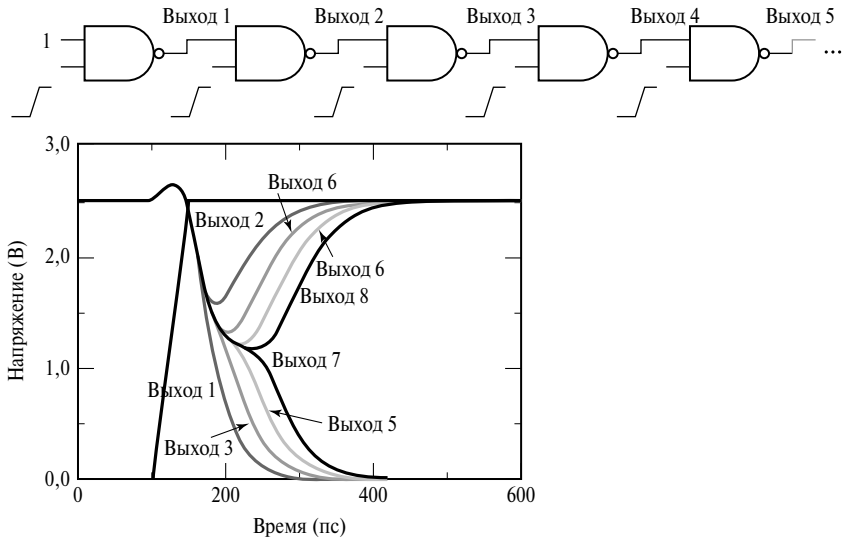


Рис. 6.22. Импульсные переходы в цепи элементов НЕ-И

равна 0, поскольку C и B логически обратны друг другу, т.е. вероятность сигнала для узла Z равна $p_Z = 0$.

Системный расчет подобных выражений для больших сетей со сходящимися разветвлениями достаточно сложен, особенно если сети содержат контуры обратной связи. Таким образом, необходима компьютерная поддержка расчетов. На самом деле программа анализа должна обрабатывать типичную последовательность входных сигналов, поскольку рассеяние мощности сильно зависит от статистических характеристик данных сигналов.

Динамические или импульсные переходы

При анализе в предыдущем разделе вероятностей перехода сложных многокаскадных логических сетей мы игнорировали тот факт, что элементы имеют ненулевые задержки распространения. В реальной ситуации конечная задержка распространения сигнала от одного логического блока к следующему может вызвать побочные переходы, именуемые *импульсными переходами* (glitch) или *динамическими паразитными импульсами* (dynamic hazard): за время одного такта правильный логический уровень сигнала устанавливается на узле после нескольких переходов.

Типичный пример импульсных помех показан на рис. 6.22, где приведена смоделированная характеристика цепи элементов НЕ-И для всех входов, одновременно изменяющихся с 0 на 1. Изначально все выходы равны 1, поскольку один из входов равен 0. В случае рассматриваемого перехода все нечетные биты должны изменить свое значение на 0, а четные — сохранить значение 1. Тем не менее из-за конечной задержки распространения четные биты выхода в старшем разряде начинают разряжаться и напряжение падает. Когда правильный входной сигнал пройдет через сеть, выход переходит на высокий уровень (1). Импульсные переходы в четных битах вызывают

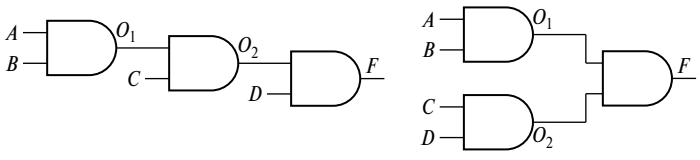


Рис. 6.23. Несколько примеров, демонстрирующих влияние топологии цепи на активность

дополнительное рассеяние мощности, помимо того, что требуется для строгой реализации логической функции. Хотя в данном примере рассмотрены частичные импульсные помехи (т.е. не с полной амплитудой), они делают существенный вклад в рассеяние мощности. Поскольку длинные цепочки логических элементов часто входят в такие важные структуры, как сумматоры и умножители, импульсный компонент может легко стать доминирующим источником потребления мощности.

Технологии разработки с уменьшением активности переключения

Динамическую мощность логического элемента можно уменьшить, минимизировав физическую емкость и активность переключения. Минимизировать физическую емкость можно различными способами, в том числе выбирая стиль схемы, размеры транзисторов, расположение и трассировку, а также оптимизируя архитектуру. С другой стороны, активность переключения можно минимизировать на всех уровнях абстракции дизайна, и именно этим мы займемся в данном разделе. Логические структуры можно оптимизировать так, чтобы минимизировать и основоположные переходы (необходимые для реализации данной функции), и паразитные переходы.

1. **Реструктуризация логики.** Изменение топологии логической сети может уменьшить вносимое ею рассеяние мощности. Рассмотрим, например, две альтернативные реализации функции $F = A \cdot B \cdot C \cdot D$ (рис. 6.23). Не будем учитывать импульсные помехи и предположим, что все основные входные сигналы (A, B, C, D) не коррелируют и распределены равномерно (т.е. $p_1(a, b, c, d) = 0,5$). Используя выражения из табл. 6.7, мы можем рассчитать активность для двух топологий, как показано в табл. 6.8. Из результатов видно, что в целом для случайных входных сигналов цепная структура обладает более низкой активностью переключения, чем древовидная реализация. Тем не менее, как отмечалось ранее, необходимо рассмотреть нестационарные эффекты, чтобы правильно подобрать компромиссное распределение мощности. В данном примере древовидная структура практически не имеет импульсной активности, поскольку для всех элементов тракты прохождения сигналов сбалансированы.
2. **Упорядочение входов.** Рассмотрим две статических логических цепи (рис. 6.24). Вероятности того, что A, B и C равны 1, указаны на рисунке. Поскольку в обеих цепях реализованы идентичные логические функции,

Таблица 6.8. Вероятности для древовидной и последовательной топологии

	O_1	O_2	F
p_1 (цепь)	1/4	1/8	1/16
$p_0 = 1 - p_1$ (цепь)	3/4	7/8	15/16
$p_0 \to 1$ (цепь)	3/16	7/64	15/256
p_1 (дерево)	1/4	1/4	1/16
$p_0 = 1 - p_1$ (дерево)	3/4	3/4	15/16
$p_0 \to 1$ (дерево)	3/16	3/16	15/256

очевидно, что активность на выходном узле Z в обоих случаях равна. Разница заключается в активности на промежуточном узле. Для первой схемы эта активность равна: $(1-0,5\times0,2)(0,5\times0,2) = 0,09$. Во втором — вероятность выполнения перехода $0 \to 1$ равна: $(1-0,2\times0,1)(0,2\times0,1) = 0,0196$, что существенно ниже, чем в первом случае. Из полученных результатов можно сделать вывод, что сигналы с высокой частотой перехода (т.е. сигналы с вероятностью, близкой к $0,5$) выгодно подавать в сеть ближе к ее выходу. Для реализации этого, как правило, достаточно переупорядочить входные сигналы.

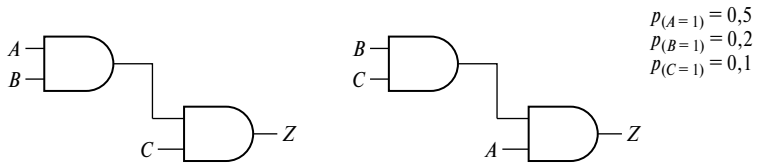


Рис. 6.24. Переупорядочение входов влияет на активность схемы

3. **Временно́е уплотнение.** Для минимизации области реализации часто используется техника временно́го уплотнения одного аппаратного ресурса (логического блока или шины) по нескольким функциям. К сожалению, минимальная площадь не всегда приводит к самой низкой активности переключения. Рассмотрим, например, передачу двух входных битов (A и B), используя любой подход: либо выделения ресурсов, либо временно́го уплотнения, как показано на рис. 6.25. В первом приближении, если игнорировать расходы, которые вводит мультиплексор, будет казаться, что степень уплотнения по времени не должна влиять на управляющую емкость, поскольку решение с временны́м уплотнением характеризуется половинной физической емкостью, переключаемой с удвоенной частотой (предполагается фиксированная пропускная способность).

   Если передаваемые данные — случайные, тогда разницы в использовании архитектур нет никакой. Однако если информационные сигналы обладают какими-либо особыми свойствами (например, временно́й корреляцией), рассеяние энергии схемы с временны́м уплотнением может быть значительно выше. Предположим, например, что сигнал A всегда (или

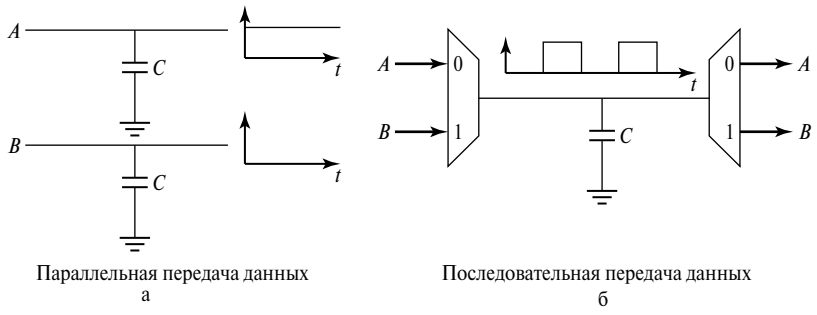


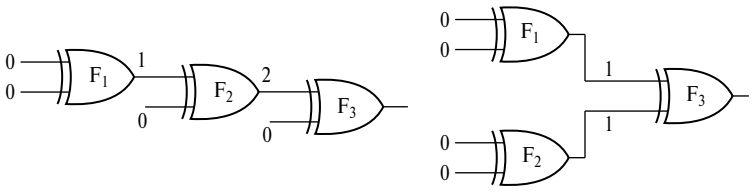
Рис. 6.25. Параллельная и последовательная шины данных с временны*м* уплотнением

б*о*льшую часть времени) равен 1, а B — 0. При параллельной реализации схемы емкость переключения очень мала, поскольку мало число переходов в битах данных. В то же время в схеме с временны*м* уплотнением сигнал шины постоянно изменяется между значениями 0 и 1. Следовательно, в цифровых системах необходимо избегать временно*го* уплотнения потоков данных с очень разными характеристиками.

4. **Уменьшение импульсных помех путем балансировки трактов сигналов.** Появление импульсных помех в схеме в основном происходит из-за рассогласования длин трактов прохождения сигналов в сети. Если все входные сигналы элемента изменяются одновременно, импульсной помехи не будет. С другой стороны, если входные сигналы изменяются в разное время, может возникнуть динамический паразитный импульс. Как правило, подобное рассогласование времен поступления сигналов возникает из-за того, что сигналы проходят пути различной длины относительно первичных входов сети. Сказанное демонстрируется на рис. 6.26. Предположим, что элемент ИСКЛЮЧАЮЩЕЕ ИЛИ обладает единичной задержкой. Первая сеть (а) страдает от импульсных помех, вызванных сильным расхождением времени прибытия сигналов на вход элемента. Например, для элемента F_3 один входной сигнал устанавливается в момент времени 0, а второй — только в момент времени 2. Если перепроектировать цепь так, чтобы время поступления сигналов на логический элемент совпадало, мы существенно уменьшим число ненужных переходов (цепь *б*).

Резюме

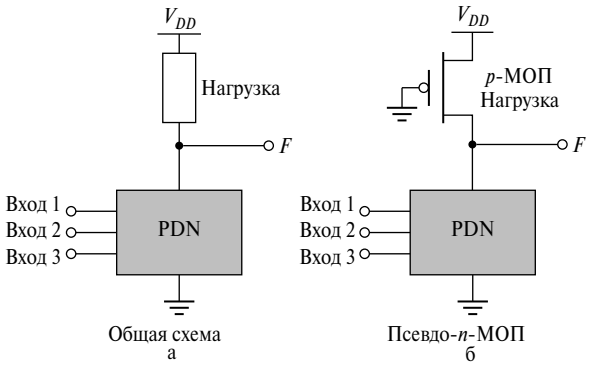
Стиль КМОП-логики, описанный в предыдущем разделе, является очень гибким и масштабируемым с точки зрения технологии, однако требует использования $2N$ транзисторов для реализации логического элемента с N входами. Кроме того, он характеризуется большой емкостью нагрузки, поскольку каждый элемент через *разветвление по выходу* питает два устройства (p-МОП- и n-МОП-транзистор). Таким образом, мы приходим к необходимости альтернативных семейств логики, которые либо проще, либо быстрее.



Цепь, чувствительная к импульсным помехам
а

Цепь, нечувствительная к помехам
б

Рис. 6.26. Импульсные помехи влияют на согласование длин трактов сигналов. Цифрами обозначено время поступления сигнала на узел



Общая схема
а

Псевдо-*n*-МОП
б

Рис. 6.27. Стандартизованный логический элемент

6.2.2. Стандартизованная логика

Концепция

Стандартизованная логика (ratioed logic) представляет собой попытку уменьшить число транзисторов, необходимых для реализации данной логической функции, часто за счет уменьшения устойчивости и дополнительного рассеяния мощности. Задачей нагрузочной сети в КМОП-устройстве является обеспечение условного тракта между V_{DD} и выходом, когда разгрузочная сеть *выключена*. В стандартизованной логике вся нагрузочная сеть заменяется одним безусловным нагрузочным устройством, которое нагружает выход до более высокого уровня, как показано на рис. 6.27, *а*. Вместо объединения активной нагрузочной и разгрузочной сети описанный элемент состоит из n-МОП-разгрузочной сети, которая реализует *логическую функцию*, и простого *нагрузочного устройства*. Пример стандартизованной логики, использующей заземленную p-МОП-нагрузку, показан на рис. 6.27, *б*; данное устройство называется псевдо-n-МОП-элементом.

Очевидным достоинством псевдо-n-МОП-элемента является уменьшенное число транзисторов ($N+1$ по сравнению с $2N$ для КМОП-элемента). Номинальное высокое выходное напряжение (V_{OH}) для данного элемента равно V_{DD}, поскольку разгрузочные устройства *выключаются*, когда выход нагружается до высокого уровня (предполагается, что V_{OL} меньше V_{Tn}). С дру-

гой стороны, *номинальное низкое выходное напряжение не равно 0 В* из-за наличия связи между устройствами разгрузочной сети и заземленным устройством нагрузки p-МОП-элемента. Это приводит к уменьшению запаса помехоустойчивости и, что более важно, статического потребления мощности. Масштабирование нагрузочного устройства относительно разгрузочных устройств можно использовать для нахождения компромиссного значения параметров — *запаса помехоустойчивости*, *задержки распространения* и *рассеяния мощности*. Поскольку размах напряжения на выходе и общее функционирование логического элемента зависят от отношения размеров n-МОП- и p-МОП-транзисторов, цепь называется *стандартизованной* (ratioed). Логической противоположностью являются *нестандартизованные* (ratioless) схемы, подобные КМОП-элементам, где низкий и высокий уровни сигнала не зависят от размеров транзисторов.

Расчет статической передаточной функции псевдо-n-МОП-элемента выполняется почти по такой же схеме, что и для аналогичных КМОП-устройств. Для расчета значения V_{OL} приравниваются токи, проходящие через питающие и нагрузочные устройства при $V_{\text{вх.}} = V_{DD}$. В данной рабочей точке разумно предположить, что n-МОП-устройство работает в линейном режиме (следовательно, в идеальном случае выход должен быть близким к 0 В), тогда как p-МОП-нагрузка находится в режиме насыщения:

$$k_n \left((V_{DD} - V_{Tn}) V_{OL} - \frac{V_{OL}^2}{2} \right) + \\ + k_p \left((-V_{DD} - V_{Tp}) \cdot V_{DSATp} - \frac{V_{DSATp}^2}{2} \right) = 0. \qquad (6.27)$$

Предполагая, что напряжение V_{OL} мало относительно питающего напряжения элемента $(V_{DD} - V_T)$, а V_{Tn} по модулю равно V_{Tp}, V_{OL} можно аппроксимировать следующим образом:

$$V_{OL} \approx \frac{k_p (V_{DD} + V_{Tp}) \cdot V_{DSATp}}{k_n (V_{DD} - V_{Tn})} \approx \frac{\mu_p \cdot W_p}{\mu_n \cdot W_n} \cdot V_{DSATp}. \qquad (6.28)$$

Чтобы максимально понизить V_{OL}, p-МОП-устройство должно иметь размер, гораздо меньший, чем у разгрузочных n-МОП-устройств. К сожалению, это отрицательно влияет на *задержку распространения* при зарядке выходного узла, поскольку ток, обеспечиваемый p-МОП-устройством, ограничен. Основным недостатком псевдо-n-МОП-элемента является статическая мощность, рассеиваемая при низком уровне выходного сигнала через тракт постоянного тока, существующий между V_{DD} и землей. Статическое потребление мощности в режиме выходного сигнала с низким выходом рассчитывается достаточно просто:

$$P_{low} = V_{DD} I_{low} \approx V_{DD} \cdot \left| k_p \left((-V_{DD} - V_{Tp}) \cdot V_{DSATp} - \frac{V_{DSATp}^2}{2} \right) \right|. \qquad (6.29)$$

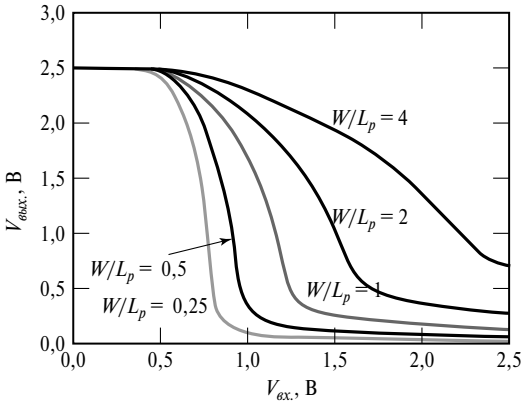


Рис. 6.28. Передаточная функция напряжения псевдо-n-МОП-инвертора как функция размеров p-МОП-транзистора

Пример 6.7. Псевдо-n-МОП-инвертор

Рассмотрим простой псевдо-n-МОП-инвертор (когда разгрузочная сеть, изображенная на рис. 6.27, вырождается до одного транзистора), n-МОП-транзистор которого характеризуется размерами 0,5/0,25 мкм. В данном примере мы изучим влияние масштабирования p-МОП-устройства на различные параметры схемы. Отношение W/L заземленного p-МОП-транзистора может иметь значение 4, 2, 1, 0,5 или 0,25. Устройства с $W/L < 1$ характеризуются длиной, большей ширины. Передаточная функция напряжения для пяти указанных отношений размеров показана на рис. 6.28.

Данные по номинальному выходному напряжению (V_{OL}), статическому рассеянию мощности и задержки распространения при изменении уровня сигнала от низкого к высокому представлены в табл. 6.9. По определению задержка распространения при изменении уровня сигнала от низкого к высокому — это время, которое требуется для достижения напряжения 1,25 В от значения V_{OL} (которое в данном инверторе не равно 0 В). Из приведенных данных четко виден компромисс между статическими и динамическими свойствами. Большее нагружающее устройство не только улучшает быстродействие, но и повышает статическое рассеяние мощности и понижает запас помехоустойчивости путем увеличения V_{OL}. Обратите внимание на то, что простая модель первого порядка, предсказывающая значение V_{OL}, является довольно эффективной. Для отношения размеров p-МОП-транзистора $W/L = 4$ V_{OL} равно $(30/115)(4)(0,63)$ В $= 0,66$ В.

Использование псевдо-n-МОП-элемента ограничивает его статическое рассеяние мощности. Тем не менее, когда основным фактором является размер, меньшее число транзисторов псевдо-n-МОП-элемента по сравнению с КМОП-элементами кажется достаточно привлекательным. По этой причине псевдо-n-МОП-элементы все еще используются в схемах с большим разветв-

Таблица 6.9. Быстродействие псевдо-n-МОП-инвертора

Размер	V_{OL}	Статическое рассеяние мощности	t_{plh}
4	0,693 В	564 мВ	14 пс
2	0,273 В	298 мВ	56 пс
1	0,133 В	160 мВ	123 пс
0,5	0,064 В	80 мВ	268 пс
0,25	0,031 В	41 мВ	569 пс

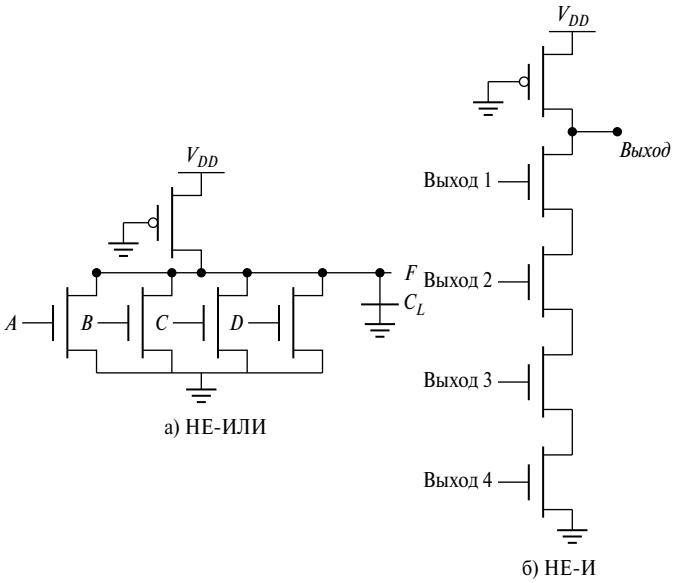


Рис. 6.29. Псведо-n-МОП-элементы НЕ-ИЛИ и НЕ-И с четырьмя входами

лением по входу. На рис. 6.29 для примера показаны схемы псевдо-n-МОП-элементов НЕ-ИЛИ и НЕ-И.

Задача 6.5. НЕ-И или НЕ-ИЛИ на псевдо-n-МОП-транзисторах

Если требуемую функцию можно реализовать с помощью логических элементов НЕ-ИЛИ либо НЕ-И на псевдо-n-МОП-транзисторах, какие устройства вы выберете?

Как создать еще лучшую нагрузку?

Существует стиль стандартизованных логических элементов, позволяющий полностью убрать токи покоя и обеспечить размах, равный напряжению питания. Подобные элементы объединяют две концепции: *дифференциальную*

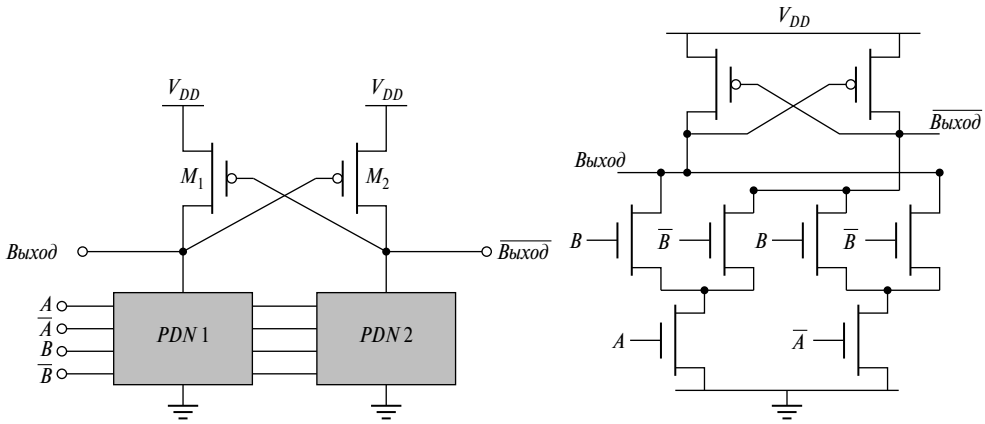


Рис. 6.30. Логический элемент DCVSL

логику и *положительную обратную связь*. Дифференциальный логический элемент требует, чтобы каждый вход обеспечивался в комплементарном формате, выход также является комплементарным. Механизм обратной связи гарантирует отключение нагрузочного устройства, когда оно не нужно. Пример подобного семейства логических элементов, именуемый *дифференциальной каскодной логикой переключения напряжения* (Differential Cascode Voltage Switch Logic — DCVSL), представлен на рис. 6.30, *а* [Heller84].

Разгрузочные сети 1 и 2 используют n-МОП-устройства и являются взаимно исключающими, т.е. когда сеть 1 проводит ток, сеть 2 выключена, а когда сеть 2 включена, отключается сеть 1. Таким образом, одновременно реализуются и требуемая логическая функция, и обратная к ней. Предположим теперь, что для определенного набора входных сигналов сеть 1 проводит ток, а сеть 2 — нет, и что Out и $\overline{Out}$ имеют изначально высокий и низкий уровень соответственно. *Включение* сети 1 приводит к понижению уровня сигнала Out, хотя между M_1 и сетью 1 все еще продолжается конкуренция. Сигнал Out находится в состоянии с высоким импедансом, поскольку M_2 и сеть 2 *отключены*. Сигнал сети 1 должен быть достаточно сильным, чтобы опустить сигнал Out ниже $V_{DD} - |V_{Tp}|$, точки, в которой *включается* M_2 и $\overline{Out}$ начинает заряжаться до уровня V_{DD}, что со временем выключит устройство M_1. В свою очередь, это позволит Out разрядиться до уровня земли. Пример работающего по подобному принципу элемента ИСКЛЮЧАЮЩЕЕ ИЛИ/НЕ-ИЛИ показан на рис. 6.30, *б*. Обратите внимание на то, что одни и те же транзисторы могут совместно использоваться двумя разгрузочными сетями, что уменьшает расходы на реализацию.

Размах выходного сигнала получившейся цепи равен входному напряжению, а статическое рассеяние мощности отсутствует: в установившемся режиме никакие разгрузочные сети и нагрузочные устройства одновременно проводить ток не могут. Тем не менее схема все равно относится к стандар-

тизованным, поскольку масштабирование p-МОП-устройства относительно разгрузочных устройств критично влияет не только на быстродействие, но и на функциональные возможности. Помимо проблемы увеличения сложности структуры, описанному стилю разработки присуща проблема рассеяния мощности из-за перекрестных токов. В процессе перехода существует промежуток времени, когда p-МОП и разгрузочная сеть включены одновременно, т.е. создают тракт короткого замыкания.

Пример 6.8. Переходная характеристика DCVSL

На рис. 6.31 показан пример переходной характеристики для логического элемента И/НЕ-И, реализованного на DCVSL. Обратите внимание на то, что по мере снижения уровня сигнала Out до $V_{DD} - |V_{Tp}|$ узел $\overline{Out}$ начинает быстро заряжаться до V_{DD}. Задержка между входом и Out составляет 197 пс, а между входом и $\overline{Out}$ — 321 пс. Статический КМОП-элемент И (НЕ-И плюс инвертор) имеет задержку 200 пс.

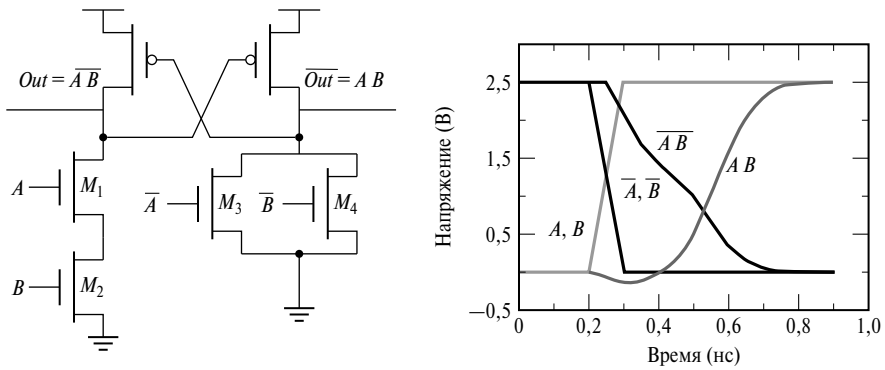


Рис. 6.31. Переходная характеристика простого DCVSL-элемента И/НЕ-И. Транзисторы M_1 и M_2 характеризуются размерами 1/0,25 мкм, M_3 и M_4 — 0,5/0,25 мкм, а p-МОП-устройства, связанные перекрестными обратными связями, — 1,5/0,25 мкм

Вопрос дизайна: несимметричная или дифференциальная схема?

DCVSL-элемент дает *дифференциальные* (или *комплементарные*) *выходы*. И выходной сигнал (V_{out}), и его обращенное значение ($\overline{V_{out}}$) доступны одновременно. Это выгодно, поскольку для получения комплементарного сигнала не требуется дополнительный инвертор. Известно, что дифференциальная реализация сложной функции может уменьшить число требуемых логических элементов вдвое! Кроме того, это часто уменьшает число элементов на критическом тракте. Наконец, данный подход предотвращает некоторые проблемы рассинхронизации, вводимые дополнительными инверторами. Например,

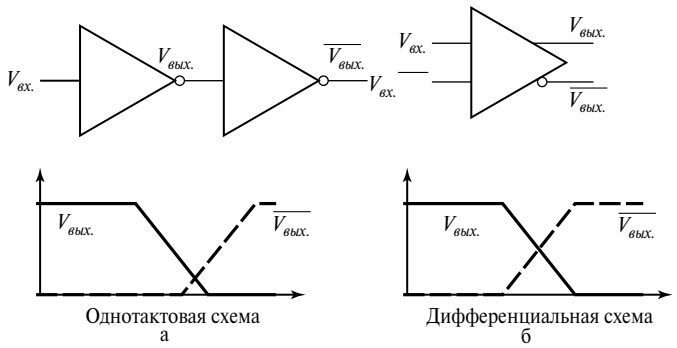


Рис. 6.32. Преимущества несимметричных логических элементов (*а*) по сравнению с дифференциальными (*б*)

в логической структуре часто получается так, что одновременно требуется и сигнал, и обратный к нему. Если комплементарный сигнал генерируется с помощью инвертора, инвертированный сигнал задерживается относительно оригинала (рис. 6.32, *а*). Это приводит к проблемам синхронизации, особенно если речь идет о высокоскоростных разработках. Семейства логических элементов с возможностью дифференциального выхода снимают эту проблему либо полностью, либо в большей степени (см. рис. 6.32, *б*).

Учитывая сказанное, возникает вопрос: если дифференциальная логика настолько хороша, почему она не используется повсеместно? Проблема в том, что из-за природы дифференциальной логики практически удваивается число соединительных проводников, что часто приводит к громоздким структурам и издержкам на реализацию отдельных элементов. Кроме того, дифференциальная логика характеризуется высокой рассеиваемой мощностью.

6.2.3. Логика на проходных транзисторах

Основы схем на проходных транзисторах

Популярной и широко используемой альтернативой КМОП-устройств является *логика на проходных транзисторах* (pass-transistor logic), уменьшающая число транзисторов, требуемых для реализации логического элемента, позволяя первичным входам питать терминалы затвора, а также терминалы истока-стока [Radhakrishnan85]. Это отличается от семейств логики, которые мы рассматривали ранее, и где первичные входы могли питать только терминалы затвора в канальных полевых униполярных МОП-транзисторах.

На рис. 6.33 показана реализации функции И, построенной по описанному принципу с использованием только n-МОП-транзисторов. В данном элементе, если вход B имеет высокий уровень, верхний транзистор включается и копирует вход A на выход F. Если вход B имеет низкий уровень, нижний проходной транзистор включается и передает значение 0. На первый взгляд ключ, питаемый сигналом $\overline{B}$, кажется избыточным. Его присутствие необхо-

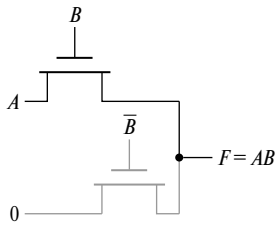


Рис. 6.33. Реализация логического элемента И с транзисторными трактами

димо, чтобы гарантировать, что элемент является статическим — в любом случае к шинам питания должен существовать тракт с низким импедансом (в рассматриваемой схеме — когда сигнал B имеет низкий уровень).

Потенциальное достоинство данного подхода заключается в том, что для реализации данной логической функции требуется меньше транзисторов. Например, реализации элемента И, изображенного на рис. 6.33, требует 4 транзистора (включая инвертор, необходимый для инвертирования сигнала B), тогда как КМОП-реализация потребует 6 транзисторов. Помимо всего прочего, меньшее число устройств — это еще и меньшая емкость.

К сожалению, как обсуждалось ранее, n-МОП-устройство эффективно при передаче нулевого сигнала, но при подъеме напряжения на узле до V_{DD} ведет себя довольно плохо. Когда проходной транзистор повышает напряжение на узле, выход заряжается только до $V_{DD} - V_{Tn}$. Фактически ситуация ухудшается тем, что устройства испытывают влияние подложки, поскольку при увеличении напряжения возникает большое напряжение между истоком и подложкой. Рассмотрим случай, когда проходной транзистор заряжает узел, напряжения на затворе и стоке которого установлено равным V_{DD}. Обозначим исток проходного n-МОП-транзистора как x. Узел x зарядится до напряжения $V_{DD} - V_{Tn}(V_x)$. В результате получаем такую формулу:

$$V_x = V_{DD} - (V_{tn0} + \gamma((\sqrt{|2\phi_f| + V_x}) - \sqrt{|2\phi_f|})). \qquad (6.30)$$

Пример 6.9. Перепад напряжения для схем с проходными транзисторами

Переходная характеристика, показанная на рис. 6.34, соответствует n-МОП-транзистору, заряжающему конденсатор. Напряжение стока n-МОП-транзистора равно V_{DD}, а напряжение затвора линейно растет с 0 В до V_{DD}. Предполагается, что изначально напряжение на узле x равно 0 В. Видим, что в начале выход заряжается быстро, однако конец перехода довольно медленный. Когда выход достигает значения $V_{DD} - V_{Tn}$, напряжение между затвором и истоком существенно падает и ток, доступный для зарядки узла x также существенно уменьшается. Расчеты с помощью уравнения (6.30) дают вы-

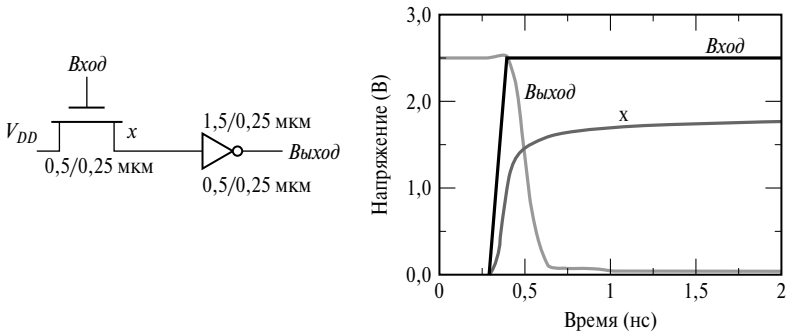


Рис. 6.34. Переходная характеристика при зарядке узла с использованием N устройств. Обратите внимание на медленную зависимость после первоначальной быстрой реакции. $V_{DD} = 2,5$ В

ходное напряжение 1,8 В, что достаточно близко к значению, полученному с помощью моделирования.

Предупреждение

Предыдущий пример демонстрирует, что проходные логические элементы нельзя соединять каскадом, подавая выход затвора одного проходного транзистора на вход затвора другого проходного транзистора. Сказанное иллюстрируется на рис. 6.35, *а*, где выход M_1 (узел x) подается на вход другого МОП-устройства. Узел x может зарядиться до напряжения $V_{DD} - V_{Tn1}$. Если размах напряжения на узле C равен напряжению питания, узел Y заряжается только до напряжения на узле x минус V_{Tn2}, т.е. до $V_{DD} - V_{Tn1} - V_{Tn2}$. С другой стороны, у схемы, приведенной на рис. 6.35, *б*, выход каскада M_1 (т.е. x) соединяется с M_2, и падение порога происходит только один раз. Следовательно, именно такой способ соединения проходных элементов является правильным.

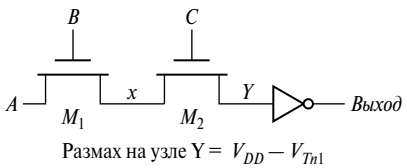


Рис. 6.35. Выходной терминал проходного транзистора (сток-исток) нельзя соединять с другими входными терминалами, чтобы не вызвать многократное падение порогового значения

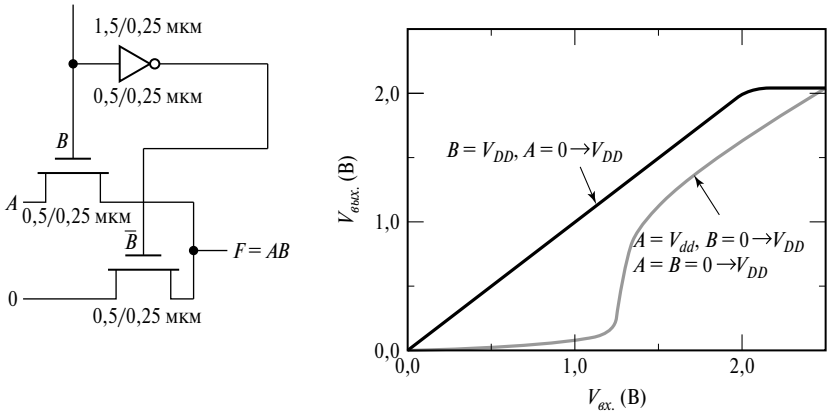


Рис. 6.36. Передаточная функция напряжения для элемента И на проходных транзисторах, показанного на рис. 6.33

Пример 6.10. Передаточная функция напряжения для элемента И на проходных транзисторах

Передаточная функция напряжения элемента на проходных транзисторах очень мало похожа на функции для КМОП-элементов. Рассмотрим, например, элемент И, показанный на рис. 6.36. Точно так же, как и для КМОП-устройств, передаточная функция напряжения логического элемента на проходных транзисторах зависит от данных. В случае, когда $B = V_{DD}$, верхний проходной транзистор включен, а нижний — выключен. В данном случае выход просто следует за входом A, пока этот вход на станет достаточно высоким, чтобы выключить верхний транзистор (т.е. пока напряжение не достигнет значения $V_{DD} - V_{Tn}$). Рассмотрим далее случай, когда $A = V_{DD}$, а сигнал B выполняет переход $0 \rightarrow 1$. Поскольку порог инвертора равен $V_{DD}/2$, нижний проходной транзистор включается и выход остается близким к нулю. Как только нижний проходной транзистор отключается, выход становится равным входу B минус падение порога. Подобное поведение наблюдается и в случае, когда на входах A и B происходит переход $0 \rightarrow 1$.

Отметим, что чисто проходной транзисторный элемент не является регенеративным. После прохождения сигнала через ряд последовательных каскадов будет наблюдаться постепенное его ухудшение. Это можно компенсировать периодической вставкой в цепь КМОП-инвертора. При включении в тракт передачи сигнала инвертора передаточная функция напряжения для этой системы становится ближе к функции для КМОП-элементов.

Из-за уменьшенного перепада напряжения использование проходных транзисторов требует для зарядки узла меньшей энергии переключения. Рассмотрим схему на проходных транзисторах (см. рис. 6.34) и будем предполагать, что напряжение стока равно V_{DD}, а напряжение на затворе меняется на

V_{DD}. Выходной узел заряжается с 0 В до $V_{DD} - V_{Tn}$ (предполагается, что узел x изначально имел напряжение 0 В), а энергия, полученная от источника для зарядки выхода проходного транзистора, выражается следующим образом:

$$E_{0\to 1} = \int_0^T P(t)dt = V_{DD} \int_0^T i_{supply}(t)dt = \\ = V_{DD} \int_0^{(V_{DD}-V_{Tn})} C_L dV_{\text{вых.}} = C_L \cdot V_{DD} \cdot (V_{DD} - V_{Tn}). \tag{6.31}$$

Хотя схема требует меньше энергии на переключение, она также может потреблять статическую мощность, когда выход находится на высоком уровне, — сниженного уровня напряжения может быть недостаточно для выключения p-МОП-транзистора последующего КМОП-инвертора.

Дифференциальная логика на проходных транзисторах

Для разработки быстродействующих структур широко используется семейство дифференциальных логических элементов на проходных транзисторах (Differential Pass-Transistor Logic — CPL или DPL). Основная идея (подобно идее DCVSL) заключается в получении обычного и комплементарного (дополнительного) входа и выдаче обычного и комплементарного выхода. Несколько элементов CPL (И/НЕ-И, ИЛИ/НЕ-ИЛИ и ИСКЛЮЧАЮЩЕЕ ИЛИ/НЕ-ИСКЛЮЧАЮЩЕЕ ИЛИ) показано на рис. 6.37. Ниже описаны некоторые интересные свойства, общие для всех элементов.

- Поскольку схемы являются *дифференциальными*, всегда доступны комплементарные входы и выходы. Хотя генерация дифференциальных сигналов требует дополнительных цепей, преимуществом дифференциального стиля разработки является то, что некоторые сложные логические элементы (например, ИСКЛЮЧАЮЩЕЕ ИЛИ или сумматоры) можно реализовать эффективно с помощью небольшого числа транзисторов. Более того, доступность всех сигналов с двумя полярностями позволяет обойтись без дополнительных инверторов, что часто удобно для статических КМОП-устройств или псевдо-n-МОП-устройств.
- CPL принадлежит к классу *статических* элементов, поскольку узлы, определяющие выход, всегда связаны либо с V_{DD}, либо с землей посредством тракта с низким сопротивлением. С точки зрения сопротивления шуму это является серьезным достоинством.
- Структура является в высокой степени модульной. По сути, все элементы используют одинаковую топологию. Меняются только входы. Благодаря этому разработка библиотеки логических элементов очень упрощается. Более сложные элементы можно создать путем каскадного соединения стандартных модулей на проходных транзисторах.

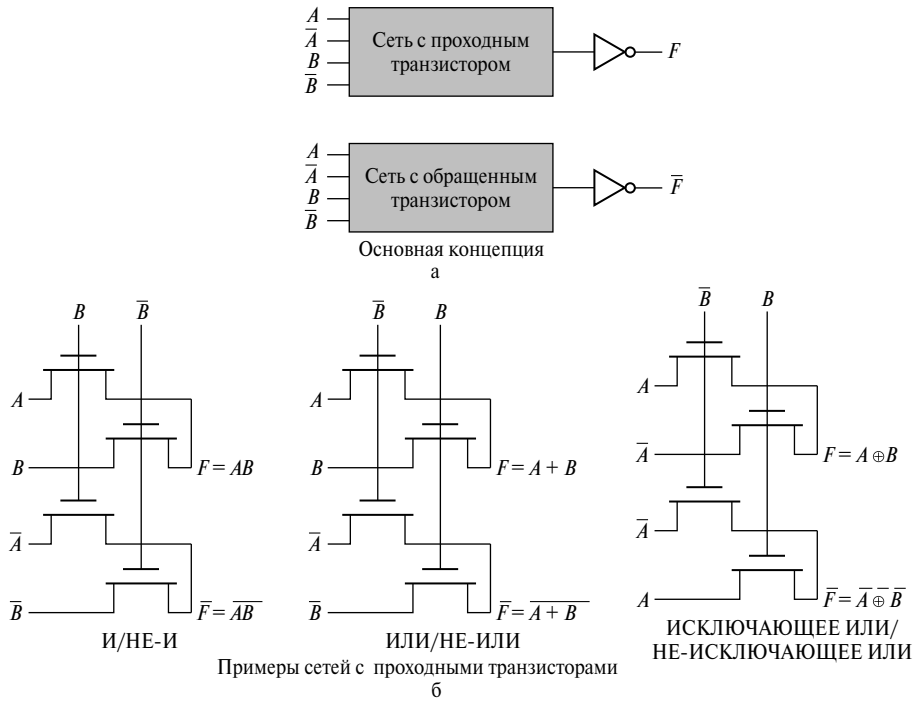


Рис. 6.37. Логическая схема на комплементарных проходных транзисторах (Complementary Pass-Transistor Logic — CPL)

Пример 6.11. Реализация на CPL схемы НЕ-И с четырьмя входами

Рассмотрим реализацию с использованием CPL элемента И/НЕ-И с четырьмя входами. Согласно ассоциативности булевой операции И ($A \cdot B \cdot C \cdot D = (A \cdot B)(C \cdot D)$), для реализации логического элемента был выбран двухкаскадный подход (рис. 6.38). Общее число транзисторов (включая заключительный буфер) равно 14, что существенно больше, чем у обсуждавшихся ранее элементов[3]. Этот фактор плюс сложные требования к трассировке соединений делают выбранный стиль разработки не очень эффективным для данного логического элемента. Тем не менее следует учитывать, что данная структура одновременно реализовывает функции И и НЕ-И, что может уменьшить число транзисторов в глобальной схеме.

[3] Приведенная конфигурация схемы приемлема только при использовании проходных транзисторов с нулевым порогом. Если это не так, она прямо противоречит концепциям, введенным на рис. 6.35.

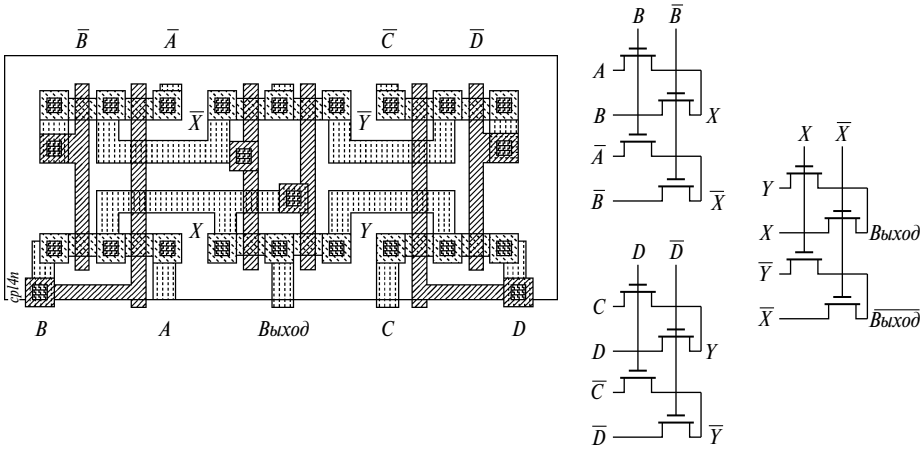


Рис. 6.38. Топология и схема реализованного на CPL элемента НЕ-И с четырьмя входами. Последний каскад инвертора опущен

В целом CPL является концептуально простым и модульным стилем разработки логических элементов. Его применимость сильно зависит от реализуемой логической функции. Доступность простой схемы ИСКЛЮЧАЮЩЕГО ИЛИ и легкость реализации мультиплексоров делают этот стиль привлекательным для разработки таких структур, как сумматоры и умножители. Данный стиль дает несколько очень быстрых и эффективных реализаций [Yano90]. Рассматривая возможность использования CPL, разработчик не должен игнорировать неявные затраты на трассировку соединений, которые очевидны при изучении схемы, изображенной на рис. 6.38.

Устойчивая и эффективная разработка с использованием проходных транзисторов

К сожалению, дифференциальные логические схемы на проходных транзисторах, как и несимметричные, характеризуются статическим рассеянием мощности и пониженным запасом помехоустойчивости, поскольку высокий уровень сигнала на входе восстанавливающего сигнал инвертора заряжается только до $V_{DD} - V_{Tn}$. На данный момент разработано несколько решений данной проблемы, которые мы и рассмотрим ниже.

Решение 1: восстановление уровня. Распространенным решением проблемы падения напряжения является использование *схемы восстановления уровня* — p-МОП-устройства, собранного в цепи обратной связи (см. рис. 6.39). Затвор p-МОП-устройства соединяется с выходом инвертора, его сток — со входом инвертора, а исток — с V_{DD}. Предположим, что напряжение на узле X равно 0 В (на выходе — V_{DD}, а устройство M_r выключено), $B = V_{DD}$ и $A = 0$. Если на входе A происходит переход $0 \rightarrow V_{DD}$, узел M_n заряжает узел X только до $V_{DD} - V_{Tn}$. Тем не менее этого достаточно, чтобы переключить выход инвертора на низкий уровень, включив устройство обратной связи M_r и зарядив узел X до V_{DD}. Таким образом, статическое рассеяние

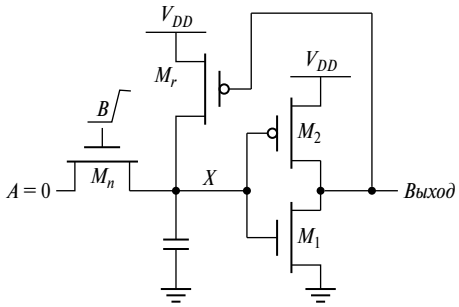


Рис. 6.39. Задача масштабирования транзистора в схеме восстановления уровня

мощности инвертора исчезает. Более того, не может существовать статического пути тока через схему восстановления уровня и проходной транзистор, поскольку фиксатор уровня активен только тогда, когда сигнал A имеет высокий уровень. В общем, достоинством приведенной схемы является то, что все уровни напряжения либо равны уровню GND, либо V_{DD} и статического потребления мощности не происходит.

Хотя приведенное решение выглядит привлекательным с точки зрения устранения статического рассеяния мощности, оно также несколько усложняет структуру схемы (схема является стандартизованной). Проблемы начинают возникать при переходе на узле X сигнала с высокого на низкий уровень (рис. 6.40). Сеть с проходным транзистором пытается разгрузить узел X, тогда как фиксатор уровня поднимает X до значения V_{DD}. Следовательно, разгрузочная сеть, обозначенная на схеме как M_n, должна быть сильнее нагрузочного устройства M_r, чтобы переключить узел X (и выход). Чтобы правильно представить функцию, реализованную схемой, требуется аккуратный расчет размеров транзисторов. Обозначим через R_1 эквивалентное сопротивление транзистора M_1 во включенном состоянии, через R_2 — аналогичную величину для транзистора M_2 и через R_r — для M_r. Когда сопротивление R_r слишком мало, напряжение на узле X невозможно опустить ниже порога переключения инвертора. Следовательно, инвертор никогда не переключится на V_{DD}, и вентиль будет заперт в одном состоянии. Чтобы разрешить проблему, можно так подобрать размеры транзисторов M_n и M_r, чтобы напряжение на узле X упало ниже порога инвертора V_M, который является функцией R_1 и R_2. Данного условия достаточно, чтобы гарантировать переключение выходного напряжения V_{out} на V_{DD} и выключение фиксирующего транзистора.

Пример 6.12. Масштабирование фиксатора уровня

Анализ схемы целиком является нетривиальной задачей, поскольку восстанавливающий транзистор действует как устройство обратной связи. Чтобы упростить схему и сделать ее пригодной для ручного анализа, можно при определении точки переключения открыть контур обратной связи и зазем-

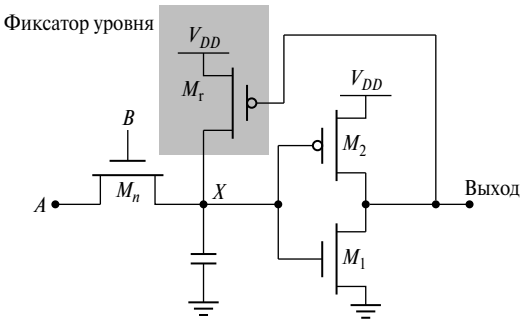


Рис. 6.40. Фиксатор уровня

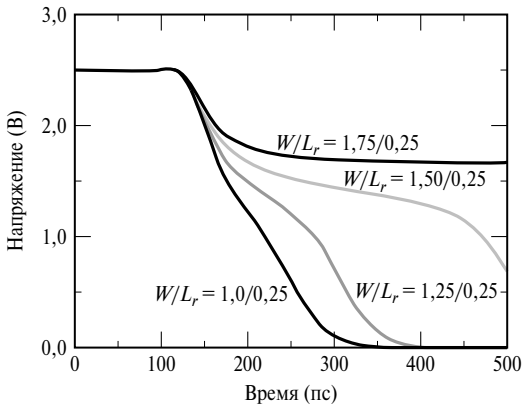


Рис. 6.41. Переходная характеристика схемы, изображенной на рис. 6.39. Слишком большой фиксатор уровня дает неправильную оценку

лить затвор восстанавливающего транзистора (данное предположение является разумным, поскольку обратная связь активизируется только тогда, когда инвертор начинает переключаться). Следовательно, M_r и M_n формируют конфигурацию, которая в сумме дает псевдо-n-МОП-транзистор, в котором M_r выступает как нагрузочный транзистор, а M_n — как разгрузочная (до уровня земли) сеть. Предположим, что размеры транзисторов M_1 и M_2 выбраны так, чтобы порог переключения инвертора был равен $V_{DD}/2$ (n-МОП — 0,5/0,25 мкм и p-МОП — 1,5/0,25 мкм). Таким образом, напряжение на узле X необходимо опустить ниже уровня $V_{DD}/2$, чтобы переключить инвертор и выключить M_r. Сказанное подтверждается рис. 6.41, на котором показана переходная характеристика для разных размеров фиксатора уровня при фиксированных размерах транзистора M_n (0,5/0,25 мкм). Как видно из результатов моделирования, для размеров, бо́льших 1,5/0,25 мкм, узел X невозможно опустить ниже порога переключения инвертора и переключить выход.

Еще одним моментом, заслуживающим рассмотрения, является влияние фиксатора уровня на скорость переключения устройства. Добавление восста-

навливающего устройства увеличивает емкость на внутреннем узле X, замедляя работу вентиля. Кроме того, это негативно влияет на время нарастания сигнала вентиля. Транзистор M_r, восстанавливающий уровень, препятствует понижению напряжения на узле X перед собственным выключением. С другой стороны, фиксатор уровня уменьшает время спада, поскольку p-МОП-транзистор после включения ускоряет процесс нагрузки.

Задача 6.6. Масштабирование устройств в проходных транзисторах

Рассмотрим схему, представленную на рис. 6.39, и предположим, что разгрузочное устройство состоит из шести соединенных последовательно проходных транзисторов, каждый из которых имеет размер 0,5/0,25 мкм (вместо транзистора M_n). Определите максимальный размер W/L фиксатора уровня, при котором устройство функционирует правильно.

На рис. 6.42 показана модификация концепции фиксатора уровня. В таком виде концепция применима к дифференциальным сетям и называется *логикой на проходных транзисторах с восстановлением амплитуды* (swing-restored pass-transistor logic). Вместо простого инвертора на выходе сети на проходных транзисторах для восстановления уровня и улучшения быстродействия применяются два инвертора, включенных встречно-параллельно. Входы подаются на терминалы затвора и истока-стока (как и в обычных сетях на проходных транзисторах). На рис. 6.42 показан простой вентиль "ИСКЛЮЧАЮЩЕЕ ИЛИ/ИСКЛЮЧАЮЩЕЕ НЕ-ИЛИ" с тремя переменными, A, B и C. Комплементарную сеть можно оптимизировать, реализовав совместное использование транзисторов основным и дополнительным выходами. Используя данное семейство логических элементов, следует помнить, что *при каскадном объединении элементов могут потребоваться промежуточные буферы.* В противном случае состязание устройств восстановления уровня, принадлежащих к различным элементам, может негативно сказаться на быстродействии.

Решение 2: транзисторы с несколькими порогами

Технологическим решением проблемы падения напряжения, связанной с логикой на проходных транзисторах, является использование устройств с несколькими порогами. Использование в проходных n-МОП-транзисторах устройств *с нулевым порогом* позволяет устранить большинство случаев падения напряжения и передать сигнал, близкий к V_{DD}. Все устройства, отличные от проходных транзисторов (т.е. инверторы), реализуются с использованием стандартных устройств с высоким порогом. В настоящее время все большее распространение получают транзисторы с несколькими порогами, кроме того, немного модифицируются существующие технологические процессы. Даже если имплантируемые устройства были тщательно откалиброваны и их порог точно равен нулю, влияние подложки по-прежнему не позволит получить полную амплитуду, равную V_{DD}.

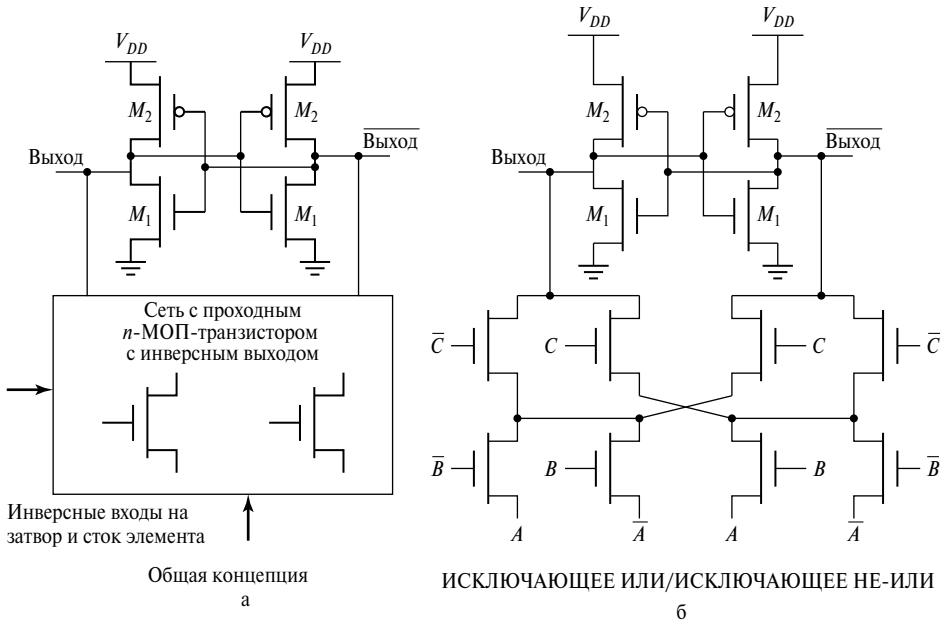


Рис. 6.42. Логика на проходных транзисторах с восстановлением амплитуды [Landman91], [Parameswar96]

Использование транзисторов с нулевым порогом может отрицательно сказаться на потреблении мощности из-за подпороговых токов, проходящих через проходные транзисторы, даже если V_{GS} не превышает V_T. Сказанное иллюстрируется на рис. 6.43, где обозначен потенциальный паразитный тракт постоянного тока. Хотя подобные каналы утечки не критичны, когда устройство постоянно переключается, в стационарном состоянии они могут стать причиной значительного потребления энергии.

Решение 3: логика на передаточных элементах

Наиболее распространенным решением проблемы падения напряжения является использование *передаточных логических элементов* (transmission gate)[4]. Данная техника основана на комплементарных свойствах n-МОП- и p-МОП-транзисторов: n-МОП-устройства пропускают устойчивый нулевой сигнал, но слабый единичный, а p-МОП-транзисторы — сильный единичный и слабый нулевой. В связи с этим идеальный вариант решения проблемы — использовать n-МОП-транзистор для разгрузки, и p-МОП-транзистор — для нагрузки. Передаточный вентиль объединяет лучшие черты обоих устройств, располагая n-МОП-устройство параллельно с p-МОП-устройством, как показано на рис. 6.44, *а*. Управляющие сигналы передаточного элемента (C и $\bar{C}$)

[4] Передаточный элемент является одним из возможных решений. Существуют и другие стили сетей на проходных транзисторах, объединяющих n-МОП- и p-МОП-транзисторы. Например, в [Bernstein98, p. 84] рассмотрена логика на двойных проходных транзисторах (Double Pass-Transistor Logic — DPL).

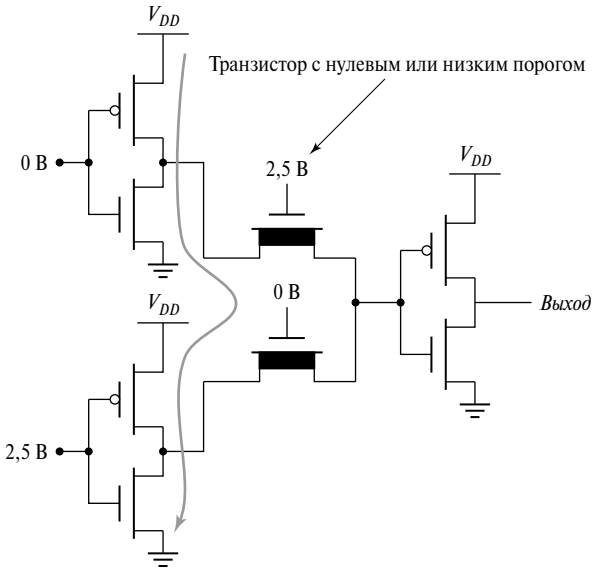


Рис. 6.43. Статическое потребление мощности при использовании проходных транзисторов с нулевым порогом

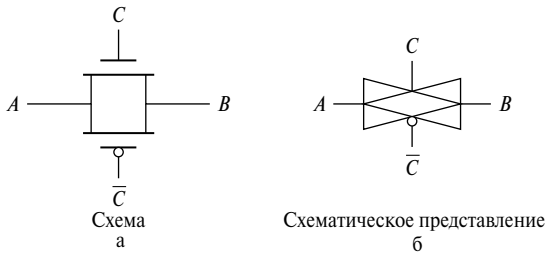


Рис. 6.44. Передаточный КМОП-элемент

являются комплементарными. Передаточный вентиль действует как двунаправленный переключатель, управляемый сигналом C. При $C = 1$ оба КМОП-транзистора включены, что позволяет сигналу пройти через элемент. Короче говоря,

$$A = B, \text{ если } C = 1. \tag{6.32}$$

С другой стороны, если $C = 0$, оба транзистора запираются, создавая таким образом разомкнутый контур между узлами A и B. Принятое обозначение передаточного вентиля показано на рис. 6.44, *б*.

Рассмотрим передаточный вентиль, показанный на рис. 6.45, *а*, и исследуем зарядку узла B до напряжения V_{DD}. Напряжение на узле A равно V_{DD}, а передаточный логический элемент открыт ($C = 1$ и $\bar{C} = 0$). Если бы в схеме

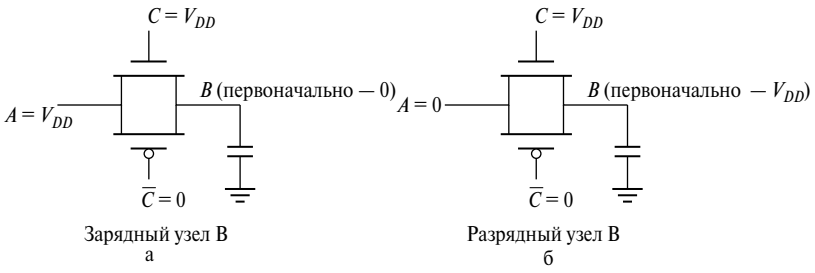


Рис. 6.45. Передаточный вентиль позволяет получить размах, равный напряжению питания

присутствовало только проходное n-МОП-устройство, узел B зарядился бы только до $V_{DD} - V_{Tn}$, после чего n-МОП-устройство закрылось бы. Однако, поскольку p-МОП-устройство присутствует и "включено" ($V_{GSp} = -V_{DD}$), выход постепенно меняется до V_{DD}. На рис. 6.45, *б* показан противоположный случай, т.е. разрядка узла B до 0. Изначально напряжение на узле B равно V_{DD}, а напряжение на узле A — 0. Сам по себе p-МОП-транзистор может только разгрузить узел B до V_{Tp}, после чего он отключается. Включенное параллельно n-МОП-устройство при этом остается включенным (поскольку его напряжение $V_{GSn} = V_{DD}$) и разгружает узел B до уровня земли. Таким образом, хотя передаточный вентиль требует двух транзисторов и больше управляющих сигналов, он позволяет добиться размаха, равного напряжению питания.

Передаточные вентили можно использовать для эффективного создания сложных логических элементов. На рис. 6.46 показан пример простого инвертирующего мультиплексора с двумя входами. Данный элемент выбирает один из входов (A или B) на основе значения контрольного сигнала S, что эквивалентно реализации следующей булевой функции:

$$\bar{F} = (A \cdot S + B \cdot \bar{S}). \tag{6.33}$$

Комплементарная реализация элемента требует восьми транзисторов.

Другим примером эффективного использования передаточных вентилей является популярная схема исключающего ИЛИ, приведенная на рис. 6.47. Полная реализация данного логического элемента требует всего шести транзисторов (включая инвертор, используемый для генерации $\bar{B}$), по сравнению с двенадцатью транзисторами, требуемыми для комплементарной реализации. Чтобы понять работу данной схемы, достаточно проанализировать по отдельности $B = 0$ и $B = 1$. Для случая $B = 1$ транзисторы M_1 и M_2 действуют как инвертор, а передаточный вентиль M_3/M_4 выключен; следовательно, $F = \bar{A}B$. В противоположном случае M_1 и M_2 заперты и работает передаточный элемент (или $F = A\bar{B}$). Комбинируя оба случая, получаем функцию исключающее ИЛИ. Обратите внимание, что, независимо от значений A и B, узел F всегда соединен либо с V_{DD}, либо с землей, а следовательно, является узлом с низким импедансом. Вообще, при разработке сетей на статических

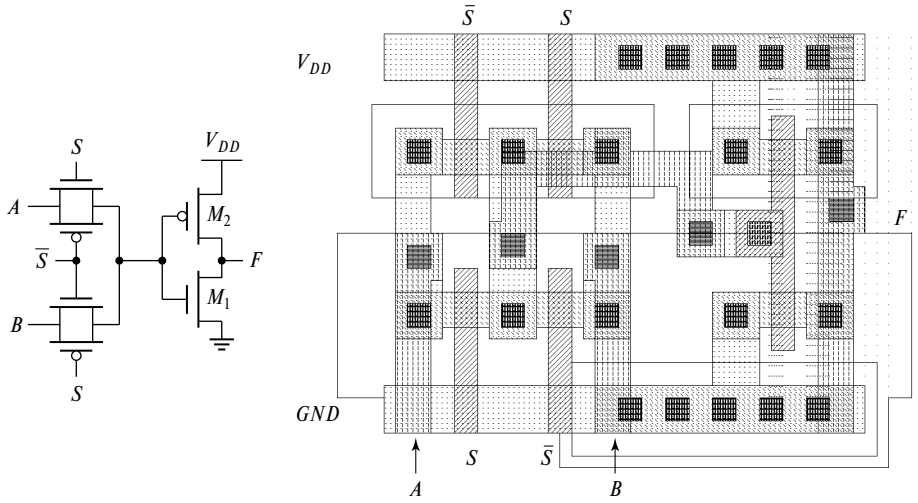


Рис. 6.46. Мультиплексор на передаточных элементах и его схема

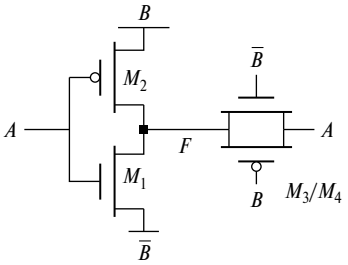


Рис. 6.47. Схема ИСКЛЮЧАЮЩЕЕ ИЛИ на передаточных вентилях

проходных транзисторах следует всегда придерживаться правила низкого импеданса. Отметим также, что логика на передаточных вентилях применяется для эффективной разработки в быстрых сумматоров и регистров.

Быстродействие логики на проходных транзисторах и передаточных вентилях

К сожалению, проходные транзисторы и передаточные вентили не являются идеальными переключателями, кроме того, они обладают последовательным сопротивлением. Чтобы количественно описать сопротивление, рассмотрим схему, приведенную на рис. 6.48, где происходит зарядка узла с 0 В до V_{DD}. При обсуждении будем использовать определение сопротивления, предполагающее большой сигнал, — напряжение, поданное на переключатель, деленное на ток утечки. Эффективное сопротивление переключателя моделируется параллельным соединением сопротивлений R_n и R_p n-МОП и p-МОП-устройств, определяемых как $(V_{DD} - V_{out})/I_{Dn}$ и $(V_{DD} - V_{out})/(-I_{Dp})$ соответ-

ственно. Очевидно, токи, проходящие через устройства, зависят от значения V_{out} и режимов работы транзисторов. В ходе изменения уровня сигнала от низкого к высокому проходные транзисторы последовательно проходят через ряд режимов работы. Для низких значений V_{out} n-МОП-устройство насыщается, и сопротивление можно аппроксимировать следующим образом:

$$R_p = \frac{V_{\text{вых.}} - V_{DD}}{I_{Dp}} = \frac{V_{\text{вых.}} - V_{DD}}{k_p \cdot \left((-V_{DD} - V_{Tp})(V_{\text{вых.}} - V_{DD}) - \frac{(V_{\text{вых.}} - V_{DD})^2}{2}\right)} \approx$$
$$\approx \frac{1}{k_p(-V_{DD} - V_{Tp})}. \tag{6.34}$$

Сопротивление повышается при увеличении значений V_{out} и стремится к бесконечности, когда V_{out} достигает $V_{DD} - V_{Tn}$ и устройство выключается. Подобным образом, мы можем проанализировать поведение p-МОП-транзистора. При малых значениях V_{out} p-МОП-устройство насыщается, но при V_{out} стремящемся к V_{DD} устройство попадает в линейный режим работы. В результате мы получаем следующую аппроксимацию сопротивления:

$$R_p = \frac{V_{\text{вых.}} - V_{DD}}{I_{Dp}} = \frac{V_{\text{вых.}} - V_{DD}}{k_p \cdot \left((-V_{DD} - V_{Tp})(V_{\text{вых.}} - V_{DD}) - \frac{(V_{\text{вых.}} - V_{DD})^2}{2}\right)} \approx$$
$$\approx \frac{1}{k_p(-V_{DD} - V_{Tp})}. \tag{6.35}$$

Смоделированное значение $R_{eq} = R_p||R_n$ как функция V_{out} показано на рис. 6.48. Видно, что R_{eq} является относительно постоянным (в данном случае — $\approx$ 8 кОм). Сказанное справедливо и для других случаев (например, при разрядке C_L). Таким образом, при анализе сетей на передаточных вентилях можно предполагать, что сопротивление переключателя является константой.

Задача 6.7. Эквивалентное сопротивление при разрядке

Определите эквивалентное сопротивление, смоделировав изменение уровня сигнала от высокого к низкому в передаточном вентиле. (Другими словами, получите график, подобный показанному на рис. 6.48.)

Кроме того, важным параметром является задержка, связанная с цепочкой из n передаточных вентилей, подобной показанной на рис. 6.49. Такая конфигурация часто встречается в таких схемах, как сумматоры и мультиплексоры. Предположим, что все передаточные вентили включены и на вход подан единичный скачок напряжения. Чтобы проанализировать задержку распространения данной сети, заменим передаточные вентили эквивалентными сопротивлениями R_{eq}. В результате получим цепь, показанную на рис. 6.49, *б*.

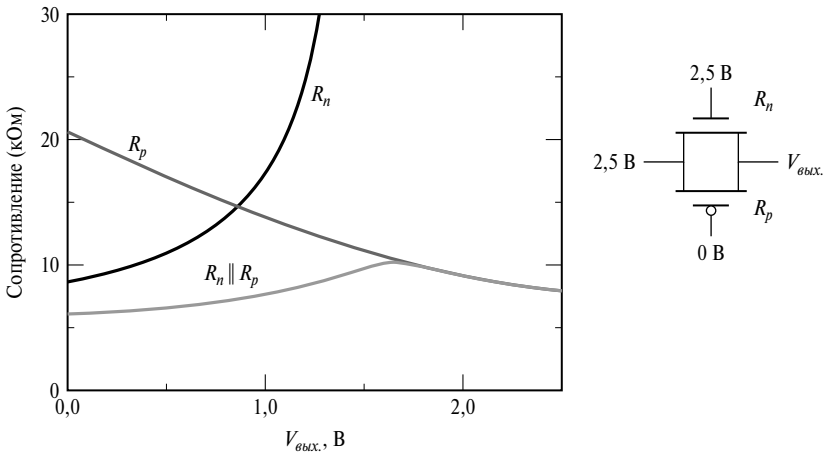


Рис. 6.48. Смоделированное эквивалентное сопротивление передаточного вентиля для изменения уровня сигнала от низкого к высокому (при $(W-L)_n = (W-L)_p = 0,5/0,25$ мкм). Подобная характеристика общего сопротивления получается и при изменении уровня сигнала от высокого к низкому

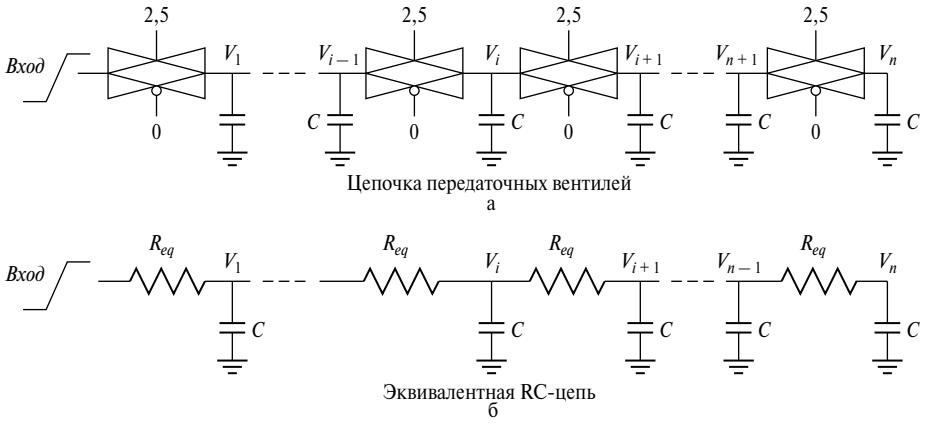


Рис. 6.49. Оптимизация скорости в цепях с передаточными вентилями

Задержку цепи из n передаточных вентилей, соединенных последовательно, можно оценить с помощью аппроксимации Элмора (см. главу 4):

$$t_p(V_n) = 0,69 \sum_{k=0}^{n} CR_{eq}k = 0,69CR_{eq}\frac{n(n+1)}{2}. \qquad (6.36)$$

Это означает, что задержка распространения пропорциональна n^2 и быстро увеличивается с ростом числа переключателей в цепочке.

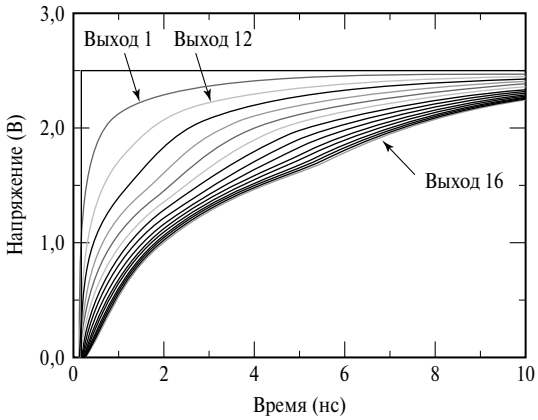


Рис. 6.50. Оптимизация скорости в сетях с передаточными вентилями

Пример 6.13. Задержка цепочки передаточных вентилей

Рассмотрим шестнадцать соединенных каскадом передаточных ветилей минимального размера, среднее сопротивление которых равно 8 кОм. Емкость узла состоит из емкости двух n-МОП- и p-МОП-устройств (емкость перехода и емкость управляющего электрода). Поскольку входы элементов считаются фиксированными, эффект Миллера отсутствует. Можно рассчитать, что для изменения уровня сигнала от низкого к высокому емкость равна приблизительно 3,6 фФ. Следовательно, задержка равна:

$$\begin{aligned} t_p &= 0,69 \cdot CR_{eq}\frac{n(n+1)}{2} = \\ &= 0,69 \cdot (3,6 \text{ фФ})(8 \text{ кОм})\left(\frac{16(16+1)}{2}\right) \approx 2,7 \text{ нс}. \end{aligned} \tag{6.37}$$

Переходная характеристика данной цепи показана на рис. 6.50. По результатам моделирования задержка равна 2,7 нс. Видим, что простая RC-модель весьма точно предсказала задержку. Также очевидно, что использование длинных цепочек проходные транзисторов приводит к существенному увеличению задержки.

Наиболее распространенным средством борьбы с длительными задержками является разрыв цепочки после каждой группы из m переключателей (рис. 6.51). Предположим, что задержка распространения каждого буфера равна t_{buf}; тогда общая задержка распространения сети с передаточными вентилями и буферами будет рассчитываться следующим образом:

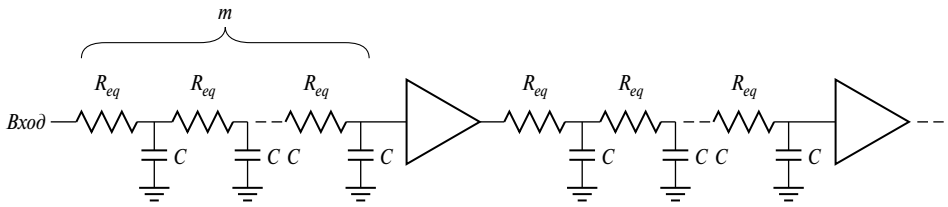


Рис. 6.51. Использование буферов для разрыва длинных цепочек передаточных вентилей

$$t_p = 0,69\left[\frac{n}{m}CR_{eq}\frac{m(m+1)}{2}\right] + \left(\frac{n}{m} - 1\right)t_{buf} = \\ = 0,69\left[CR_{eq}\frac{n(m+1)}{2}\right] + \left(\frac{n}{m} - 1\right)t_{buf}. \tag{6.38}$$

Получающаяся задержка характеризуется всего лишь линейной зависимостью от числа переключателей n, тогда как в цепи без буферов зависимость является квадратичной по n. Чтобы найти оптимальное число переключателей m_{opt} между буферами, можно приравнять производную $\frac{\partial t_p}{\partial m}$ к 0, откуда следует:

$$m_{opt} = 1,7\sqrt{\frac{t_{pbuf}}{CR_{eq}}}. \tag{6.39}$$

Очевидно, число переключателей на сегмент растет с увеличением значения t_{buf}. В современных технологиях m_{opt} обычно равно 3 или 4. Отметим, что в приведенном анализе игнорировался тот факт, что время t_{pbuf} само является функцией нагрузки m. Более строгий анализ, учитывающий этот факт, будет представлен в главе 9.

Пример 6.14. Цепочка передаточных вентилей

Рассмотрим ту же цепочку из шестнадцати передаточных вентилей. Буферы, показанные на рис. 6.51, можно реализовать как инверторы (и использовать их вместо двух инверторов, соединенных каскадом). В некоторых случаях может потребоваться дополнительный инвертор для получения правильной полярности. Предположив, что масштаб инверторов выбран так, что размер n-МОП-транзистора — 0,5/0,25 мкм, а p-МОП-транзистора — 0,5/0,25 мкм, из уравнения (6.39) получим, что инвертор необходимо добавлять после каждого третьего передаточного вентиля. По результатам моделирования задержка, получающаяся при размещении инверторов через каждые два вентиля, равна 154 пс; через каждые три вентиля — 154 пс; через каждые четыре вентиля — 164 пс. В целом добавление буферизующих инверторов уменьшает задержку почти в два раза.

Предупреждение

Хотя многие схемы, рассмотренные в предыдущих разделах, выглядят очень интересными, и кажется, что они по многим параметрам предпочтительнее КМОП-элементов, ни одна из них не обладает такой устойчивостью и легкостью разработки, как комплементарные МОП-устройства. Следовательно, использовать их необходимо умеренно и осторожно. Если на схему не налагаются чрезмерно ограничивающие требования, касающиеся площади, сложности и скорости, рекомендуется использовать КМОП-элементы.

6.3. Разработка динамических КМОП-элементов

Как отмечалось ранее, реализация статической КМОП-логики с разветвлением по входу N требует $2N$ устройств. Было представлено несколько подходов, позволяющих уменьшить число транзисторов, необходимых для реализации данной логической функции: псевдо-n-МОП-элементы, логика на проходных транзисторах и т.д. Использование концепции псевдо-n-МОП-элементов для реализации логического элемента с N входами требует всего $N+1$ транзисторов, но, к сожалению, этот подход характеризуется статическим рассеянием мощности. В данном разделе представлен альтернативный стиль разработки логических элементов — с использованием *динамической логики*. Этот подход позволяет получить подобный результат, но без статического потребления мощности. Суть его заключается в добавлении входного синхроимпульса и использовании последовательности фаз *предварительной зарядки* и *оценки*.

6.3.1. Динамическая логика: основные принципы

Базовая структура динамического логического элемента (n-типа) показана на рис. 6.52, *а*. Разгрузочная сеть создается точно так же, как и при использовании КМОП-элементов. В данном случае работа схемы разбивается на две основные фазы — *предварительную зарядку* и *оценку*, причем режим работы определяется *синхронизирующим сигналом* CLK.

Предварительная зарядка

Когда $CLK = 0$, выходной узел Out предварительно заряжается p-МОП-транзистором M_p до V_{DD}. В это время n-МОП-транзистор оценки M_e выключен, поэтому тракт разгрузки закрыт. Благодаря транзистору оценки статическое потребление мощности в период предварительной зарядки отсутствует (т.е. отсутствует ток, который мог бы протекать между устройствами питания, если бы одновременно были включены устройства разгрузки и предварительной зарядки).

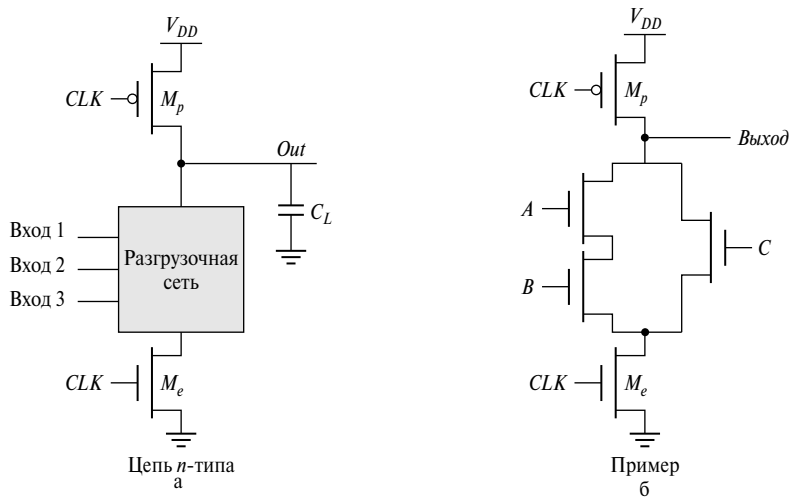


Рис. 6.52. Базовая концепция динамического элемента

Оценка

При $CLK = 1$ транзистор предварительной зарядки M_p выключен, а включен транзистор оценки M_e. Выход условно разряжается на основе входных значений и топологии разгрузочной сети. Если входы таковы, что разгрузочная сеть проводит ток, тогда межд узлами Out и GND существует тракт с низким сопротивлением, и выход разряжается на землю. Если разгрузочная сеть выключена, значение, полученное при предварительной зарядке, хранится в выходной емкости C_L, которая представляет собой комбинацию емкостей переходов, емкости проводки и входной емкости элементов разветвления по выходу. В фазе оценки единственный возможный тракт между выходным узлом и шиной питания — к земле. Следовательно, как только узел Out разрядится, его нельзя будет зарядить снова до следующей операции предварительной зарядки. Таким образом, **входы логического элемента могут в процессе оценки измениться не более одного раза**. Обратите внимание на то, что в процессе оценки выход может находиться в *состоянии с высоким импедансом*, если разгрузочная сеть при этом выключена. Такое поведение принципиально отличается от поведения статического аналога схемы, который всегда имеет тракт с низким импедансом между выходом и одной из шин питания.

Рассмотрим в качестве примера схему, показанную на рис. 6.52, *б*. В фазе предварительной зарядки ($CLK = 0$) выход независимо от входных значений заряжается до V_{DD}, поскольку устройство оценки в это время выключено. В фазе оценки ($CLK = 1$) создается проводящий тракт между узлами Out и GND тогда (и только тогда), когда $A \cdot B + C$ равно TRUE. В противном случае выход остается в заряженном состоянии (V_{DD}). Таким образом, в схеме реализована следующая логическая функция:

$$Out = \overline{CLK} + \overline{(A \cdot B + C)} \cdot CLK. \tag{6.40}$$

Динамические логические элементы обладают рядом важных свойств. Опишем их.

- Логическая функция реализуется разгрузочной n-МОП-сетью. Эта сеть создается так же, как и для статических КМОП-элементов.
- *Число транзисторов* (для сложных логических элементов) существенно меньше, чем для статического случая: $N + 2$ по сравнению с $2N$.
- Схема является *нестандартизованной*. Масштабирование p-МОП-устройства предварительной загрузки не влияет на правильное функционирование элемента. Это устройство можно сделать большим, чтобы улучшить время изменения уровня сигнала от низкого к высокому (разумеется, за счет времени изменения уровня сигнала от высокого к низкому). Тем не менее здесь следует еще учитывать рассеяние мощности, поскольку большее устройство предварительной зарядки прямо увеличивает рассеяние мощности тактового сигнала.
- Схема потребляет только *динамическую мощность*. В идеальном случае между V_{DD} и GND вообще не существует тракта тока покоя. При этом, правда, общее рассеяние мощности может быть существенно выше, чем у статического логического элемента.
- Динамические логические элементы имеют *более высокую скорость переключения*. Это объясняется двумя основными причинами. Первая (очевидная) — меньшая емкость нагрузки, порожденная меньшим числом транзисторов на логический элемент, и нагрузка каждого *разветвления по входу* единственным транзистором. В результате мы получаем *меньшие логические трудозатраты*. Например, логические трудозатраты динамического элемента НЕ-ИЛИ с двумя входами равны $2/3$, что значительно меньше значения $5/3$, которым характеризуется статический КМОП-аналог. Вторая причина — динамический логический элемент не имеет тока короткого замыкания, и весь ток, обеспечиваемый разгрузочными устройствами, идет на разрядку емкости нагрузки.

Низкий и высокий уровень напряжения (V_{OL} и V_{OH}) определяются легко — это GND и V_{DD}, причем они не зависят от размеров транзисторов. Другие параметры передаточной функции напряжения существенно отличаются от значений для статических элементов. Запас помехоустойчивости и пороги переключения определяются как статические величины, не зависящие от времени. Кроме того, для работы динамического элемента необходима периодическая последовательность фаз предварительной зарядки и оценки. Следовательно, чистый статический анализ здесь неприменим. В фазе оценки разгрузочная сеть динамического инвертора начинает проводить ток, когда входное напряжение превышает пороговое значение (V_{Tn}) разгрузочного n-МОП-транзистора. Следовательно, будет разумно предположить, что порог переключения (V_M), а также напряжения V_{IH} и V_{IL} равны V_{Tn}. Таким образом, мы получаем низкое значение NM_L.

Вопрос разработки

Динамический логический элемент можно реализовать и с использованием дуального подхода — когда выходной узел через n-МОП-транзистор предварительной зарядки соединяется с землей, а нагрузочная сеть оценки реализуется на p-МОП. В таком случае схема будет работать согласно описанным выше принципам: в фазе предварительной зарядки выходной узел разряжается на землю; в фазе оценки выход заряжается до V_{DD} при выполнении предопределенного условия. Недостатком данного динамического элемента p-типа является более медленная работа по сравнению с элементом n-типа, вызванная меньшим управляющим током p-МОП-транзисторов.

6.3.2. Скорость и рассеяние мощности в динамических логических элементах

Основными достоинствами динамической логики являются повышенная скорость и уменьшенная площадь реализации. Поскольку для реализации нужной логической функции требуется меньше устройств, общая емкость нагрузки также будет существенно меньше. Исходя из этого, при анализе поведения схемы при переключении выявляется несколько интересных особенностей. После фазы предварительной зарядки выходной сигнал имеет высокий уровень. При низком входном сигнале никакого дополнительного переключения не происходит. В результате $t_{pLH} = 0$! С другой стороны, изменение уровня сигнала от высокого к низкому требует разрядки выходной емкости через разгрузочную сеть. Следовательно, время t_{pHL} пропорционально C_L и характеристикам разгрузочной сети, чувствительных к току. Наличие транзистора оценки несколько замедляет работу элемента, поскольку вводит дополнительное последовательное сопротивление. Если же убрать этот транзистор, функциональность схемы нарушена не будет, но может появиться статическое рассеяние мощности и потенциальное падение быстродействия.

Сделанный выше анализ нельзя назвать совсем честным, поскольку мы проигнорировали влияние времени предварительной зарядки на скорость переключения вентиля. Временем предварительной зарядки считается время, требуемое для зарядки емкости C_L через p-МОП-транзистор предварительной зарядки. В этот промежуток времени логическую функцию, реализованную в элементе, использовать невозможно. Тем не менее довольно часто цифровые системы можно спроектировать так, чтобы время предварительной зарядки совпадало с периодом выполнения других системных функций. Например, предварительная зарядка арифметического блока в микропроцессоре совпадает с декодированием инструкции. Разработчик должен помнить об этой "мертвой зоне" динамической логики и внимательно изучить все преимущества и недостатки ее использования, учитывая общесистемные требования.

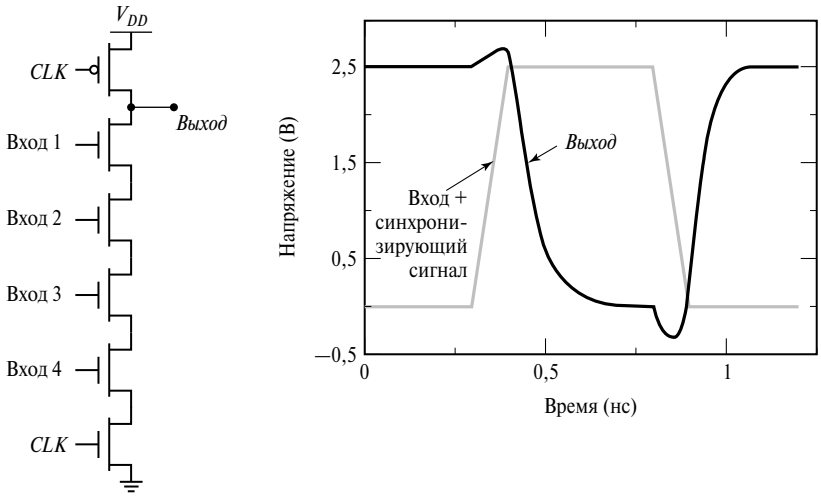


Рис. 6.53. Схема и переходная характеристика динамического элемента НЕ-И с четырьмя входами

Таблица 6.10. Параметры постоянного и переменного тока динамического элемента НЕ-И с четырьмя входами

Число транзисторов	V_{OH}	V_{OL}	V_M	NM_H	NM_L	t_{pHL}	t_{pLH}	t_{pre}
6	2,5 В	0 В	V_{TN}	$2,5 - V_{TN}$	V_{TN}	110 пс	0 пс	83 пс

Пример 6.15. Динамический элемент НЕ-И с четырьмя входами

На рис. 6.53 показана структура элемента НЕ-И с четырьмя входами, разработанного с использованием динамической логики. Вследствие динамической природы вентиля передаточная функция напряжения рассчитывается не так, как при традиционном подходе. Как обсуждалось ранее, мы предполагаем, что порог переключения вентиля равен порогу разгрузочного n-МОП-транзистора. Это, как показано в табл. 6.10, приводит к асимметричному запасу помехоустойчивости.

Динамическое поведение вентиля моделировалось с помощью SPICE. Предполагалось, что при высоком уровне тактового импульса на все входы подаются сигналы высокого уровня. Нарастанию тактового сигнала соответствует разрядка выходного узла. Получающаяся в результате переходная характеристика показана на рис. 6.53, задержки распространения указаны в табл. 6.10. Длительность фазы предварительной зарядки можно регулировать, изменяя размер p-МОП-транзистора предварительной зарядки. Размер p-МОП-транзистора не следует делать слишком большим, поскольку это и замедляет работу логического элемента, и увеличивает емкостную нагрузку на шину тактовых сигналов. В больших схемах последний фактор может быть очень важным, поскольку в этом случае нагрузка на шину становится чрезмерной и трудноуправляемой.

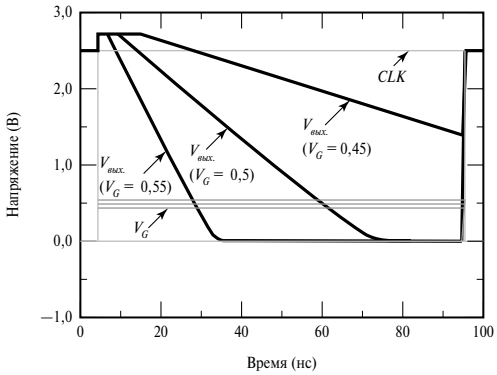


Рис. 6.54. Влияние помехи на входе на выход. Порог переключения зависит от времени, отведенного на оценку. Если фаза оценки коротка, допустима большая помеха

Как отмечалось ранее, параметры статического элемента зависят от времени. Чтобы проиллюстрировать это, рассмотрим элемент НЕ-И с четырьмя входами, в котором все частичные входы связаны и в них происходит изменение уровня сигнала от низкого к высокому. На рис. 6.54 показана смоделированная переходная характеристика выходного напряжения для трех различных переходов на входе — с 0 до 0,45, 0,5 и 0,55 В соответственно. В предыдущем анализе мы определили порог переключения динамического элемента как порог устройства. Однако величина, на которую падает входное напряжение, сильно зависит от входного напряжения и *доступного времени оценки*. Если время оценки мало, напряжение шумов, которое может повредить сигнал, должно быть больше. Другими словами, порог переключения действительно зависит от времени.

Из приведенных рассуждений может показаться, что динамическая логика очень выгодна с точки зрения потребляемой мощности. Действительно, во-первых, физическая емкость ниже, поскольку динамическая логика требует для реализации данной функции меньше транзисторов. Кроме того, нагрузкой на каждое разветвление по выходу является один транзистор, а не два. Во-вторых, динамические логические элементы *по построению* предполагают не более одного перехода за такт. В динамических логических элементах не наблюдаются динамические помехи. Наконец, в-третьих, с динамическими логическими элементами можно забыть о мощности короткого замыкания, поскольку при оценке значения логической функции нагрузочный тракт не включается.

Хотя, в общем-то, приведенные аргументы справедливы, динамические логические элементы обладают и рядом недостатков: (1) мощность тактового сигнала динамического элемента может быть значительной, в частности, из-за того, что этот сигнал гарантированно выполняет один переход в каждом такте; (2) число транзисторов больше минимального, требуемого для реализации логической функции; (3) при добавлении устройств борьбы с утечкой возможно появление мощности короткого замыкания (подробнее этот момент

обсуждается ниже); (4) (и самое важное) динамическая логика обычно характеризуется большей активностью переключения из-за периодических операций *предварительной зарядки* и *разрядки*. Ранее мы показали, что вероятность перехода для статического элемента равна $p_0 p_1 = p_0(1 - p_0)$. Для динамической логики вероятность перехода на выходе не зависит от состояния (истории) входов, а зависит от вероятностей сигналов. Для динамического элемента n-типа выход совершает переход $0 \rightarrow 1$ в фазе предварительной зарядки только в том случае, если он был разряжен в предшествовавшей фазе оценки. Следовательно, вероятность перехода $0 \rightarrow 1$ для динамического элемента n-типа выражается следующим образом:

$$a_{0\rightarrow 1} = p_0, \tag{6.41}$$

где p_0 — вероятность того, что выход равен нулю. Данное число всегда больше или равно $p_0 p_1$. Для равновероятных входов вероятность перехода для вентиля с N входами равна:

$$a_{0\rightarrow 1} = \frac{N_0}{2^N}, \tag{6.42}$$

где N_0 — число нулевых позиций в таблице истинности логической функции.

Пример 6.16. Оценка активности динамической логической схемы

Чтобы проиллюстрировать увеличение активности в динамическом элементе, рассмотрим снова элемент НЕ-ИЛИ с двумя входами. Его динамическая реализация показана на рис. 6.55, где также показан статический аналог схемы. Для равновероятных входов существует вероятность 75%, что выходной узел динамического элемента будет разряжен сразу после фазы предварительной зарядки, откуда следует, что активность такого элемента равна 0,75 (т.е. $P_{NOR} = 0,75 C_L V_{dd}^2 f_{clk}$). Активность соответствующей статической реализации немного меньше — 3/16. Для динамического элемента НЕ-И вероятность перехода равна 1/4 (поскольку вероятность разрядки выхода составляет 25%), тогда как для статической реализации получается значение 3/16. Хотя данные примеры иллюстрируют, что активность переключения динамической логики в общем случае выше, чем активность статической, следует отметить, что динамическая логика имеет меньшую физическую емкость. Таким образом, при анализе динамического рассеяния мощности необходимо учитывать оба фактора.

Задача 6.8. Расчет активности

Для динамического элемента НЕ-И с четырьмя входами рассчитайте активность при следующих предположениях относительно входов: они независимы и $p_{A=1} = 0,2$; $p_{B=1} = 0,3$; $p_{C=1} = 0,5$; $p_{D=1} = 0,4$.

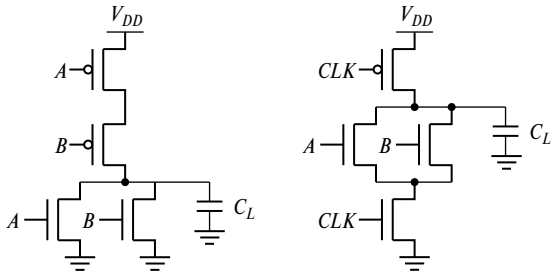


Рис. 6.55. Статический и динамический (n-типа) элементы НЕ-ИЛИ

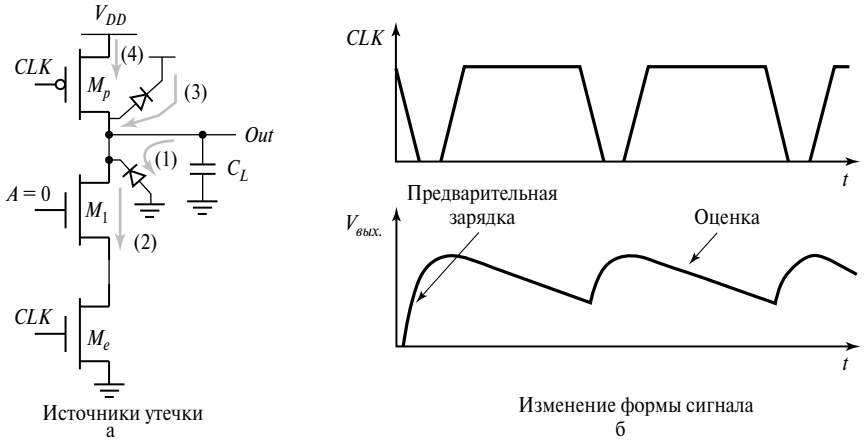


Рис. 6.56. Источники утечки в динамических схемах

6.3.3. Вопросы сохранности сигнала в динамических схемах

Очевидно, динамическая логика позволяет получить более быстродействующие решения по сравнению со статическими схемами. Тем не менее существует несколько важных соображений, которые необходимо учитывать при разработке динамических схем. Эти соображения касаются утечки и разделения заряда, емкостной связи и проникания тактового сигнала. В данном разделе мы рассмотрим эти факторы подробнее.

Утечка заряда

Работа динамического элемента зависит от динамического хранения выходного значения в конденсаторе. Если разгрузочная сеть находится в *выключенном состоянии* (идеальный случай), в фазе оценки выход должен оставаться в состоянии предварительной зарядки с напряжением V_{DD}. Однако данный заряд постепенно просачивается из-за токов утечки, что со временем приводит к неправильной работе вентиля. Источники утечки для стандартной динамической схемы на инверторах показаны на рис. 6.56, *а*.

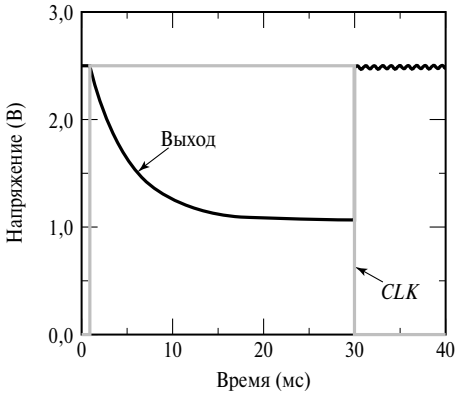


Рис. 6.57. Влияние утечки заряда: выходное напряжение устанавливается равным промежуточному значению, определяемому резистивным делителем устройств нагрузки и разгрузки

Источниками 1 и 2 являются *обратносмещенный диод* и *подпороговая утечка* разгрузочного n-МОП-устройства M_1 соответственно. Заряд, запасенный в C_L, будет медленно просачиваться через указанные каналы утечки, вызывая ухудшение высокого уровня сигнала (см. рис. 6.56, *б*). Следовательно, динамическим схемам требуется минимальная тактовая частота; на практике она обычно равна нескольким килогерцам. Из-за этого динамические технологии малопривлекательны для таких медленных продуктов, как часы или процессоры, использующие условные тактовые импульсы (когда значение минимальной тактовой частоты является неопределенным). Обратите внимание на то, что p-МОП-устройство предварительной зарядки также вносит определенный ток утечки из-за обратносмещенного диода (источник 3) и подпороговой проводимости (источник 4). В некоторой степени ток утечки p-МОП-транзистора нейтрализует утечку разгрузочного тракта. В результате выходное напряжение задается резистивным делителем, состоящим из нагрузочного и разгрузочного трактов.

Пример 6.17. Утечка в динамических схемах

Рассмотрим простой инвертор, все устройства которого имеют размер 0,5/0,25 мкм. Предположим, что в фазе оценки вход остается низким. В идеальном случае выход должен оставаться в состоянии предварительной зарядки — V_{DD}. Однако, как видно на рис. 6.57, выходное напряжение падает. Как только выход упадет ниже порога переключения подключенного к нему логического элемента, выход будет интерпретирован как низкий уровень напряжения. Обратите внимание на то, что из-за тока утечки, порождаемого нагрузочным p-МОП-транзистором, выход опускается до напряжения, являющегося промежуточным между высоким и низким уровнями.

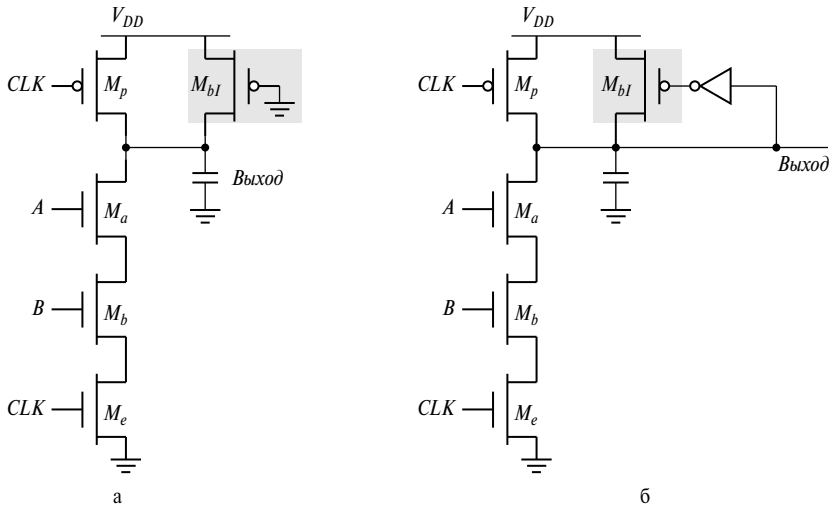


Рис. 6.58. Статические делители напряжения компенсируют утечку заряда

Утечка вызывается высокоимпедансным состоянием выходного узла в фазе оценки, когда разгрузочный тракт выключен. Проблему утечки можно решить, уменьшив выходной импеданс выходного узла в фазе оценки. Это часто реализуется путем добавления *стабилизирующего резистора* (bleeder transistor), как показано на рис. 6.58, *а*. Единственной функцией стабилизирующего нагрузочного n-МОП-транзистора является компенсация потери заряда вследствие утечки через разгрузочный тракт. Чтобы избежать проблем масштабирования, связанных с данным стилем логических схем, а также сопутствующего статического потребления мощности, сопротивление стабилизатора должно быть большим (другими словами, устройство должно быть малым). Это позволяет (мощным) разгрузочным устройствам понижать напряжение на узле Out значительно ниже порога переключения следующего логического элемента. Довольно часто стабилизатор реализуется в цепи обратной связи, чтобы вообще убрать статическое рассеяние мощности (см. рис. 6.58, *б*).

Деление заряда

Другим важным моментом при разработке динамических логических элементов является влияние деления заряда. Рассмотрим схему, показанную на рис. 6.59. В фазе предварительной зарядки выходной узел предварительно заряжается до V_{DD}. Предположим, что в фазе предварительной зарядки все входы установлены равными 0, а емкость C_a разряжена. Кроме того, предположим, что в течение фазы оценки вход B остается равным нулю, а на входе A происходит переход $0 \rightarrow 1$, что включает транзистор M_a. Заряд, изначально запасенный в конденсаторе C_L, перераспределяется между C_L и C_a. Это приводит к падению выходного напряжения, которое невозможно

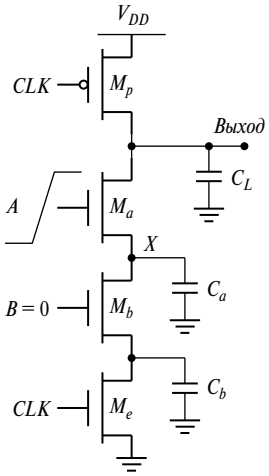


Рис. 6.59. Разделение заряда в динамических сетях

восстановить из-за динамической природы схемы. Это влияние на выходное напряжение рассчитать довольно легко. Учитывая сделанные ранее предположения, справедливы следующие начальные условия: $V_{out}(t = 0) = V_{DD}$ и $V_X(t = 0) = 0$. В результате необходимо рассмотреть два сценария.

1. $\triangle V_{out} < V_{Tn}$. В этом случае конечное значение V_X равно $V_{DD} - V_{Tn}(V_X)$. Далее из закона сохранения заряда получаем:

$$C_L V_{DD} = C_L V_{\text{вых.}}(\text{final}) + C_a[V_{DD} - V_{Tn}(V_X)]$$
$$\text{или} \tag{6.43}$$
$$\triangle\ V_{\text{вых.}} = V_{\text{вых.}}(\text{final}) + (-V_{DD}) = -\frac{C_a}{C_L}[V_{DD} - V_{Tn}(V_X)].$$

2. $\triangle V_{out} > V_{Tn}$. В таком случае V_{out} и V_X достигают одинакового значения:

$$\triangle\ V_{\text{вых.}} = -V_{DD}\left(\frac{C_a}{C_a + C_L}\right). \tag{6.44}$$

Чтобы определить, какой из двух сценариев справедлив, используется отношение емкостей. Граничной между двумя случаями является ситуация, когда в уравнении (6.44) $\triangle V_{out}$ равно V_{Tn}, откуда получаем следующее:

$$\frac{C_a}{C_L} = \frac{V_{Tn}}{V_{DD} - V_{Tn}}. \tag{6.45}$$

Случай 1 имеет место, когда отношение (C_a/C_L) меньше величины, определенной в формуле (6.45). Если это не так, справедливо уравнение (6.44). В любом случае желательно, чтобы значение $\triangle V_{out}$ было меньше $|V_{Tp}|$. Выход

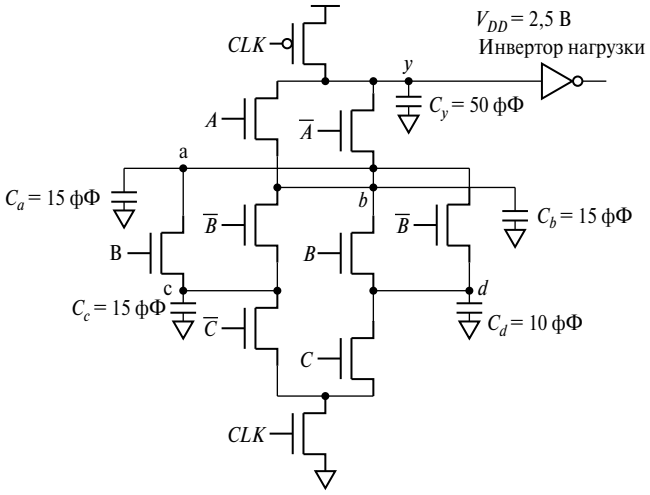


Рис. 6.60. Пример, иллюстрирующий эффект разделения заряда в динамических логических схемах

динамического вентиля может подаваться на вход статического инвертора, и тогда низкий уровень V_{out} приведет к статическому потреблению мощности. Кроме того, если выходное напряжение упадет ниже порога переключения вентиля, который оно питает, схема будет функционировать неверно.

Пример 6.18. Разделение заряда

Рассмотрим влияние разделения заряда на динамический логический элемент, показанный на рис. 6.60, в котором реализована функция ИСКЛЮЧАЮЩЕЕ ИЛИ с тремя входами $y = A \oplus B \oplus C$. Первое, с чем мы должны разобраться, — какие условия приводят к наступлению наихудшего падения напряжения на узле y. Для упрощения будем игнорировать нагружающий инвертор и предположим, что все входы имеют низкий уровень в фазе предварительной зарядки и что напряжение на всех изолированных внутренних узлах (V_a, V_b, V_c и V_d) изначально равно 0 В. Изучая таблицу истинности для данной логической функции, видим, что для 4 из 8 случаев выход останется высоким. Худшее изменение выхода происходит тогда, когда в фазе оценки на выходной узел подается максимальная величина внутренней емкости. Это происходит в случае $\bar{A}BC$ или $A\bar{B}C$. Чтобы найти изменение напряжения, приравняем внутренний заряд к конечному заряду согласно уравнению (6.44), что дает худшее из возможных изменений — $30/(30 + 50) \times 2,5$ В $= 0,94$ В. Чтобы гарантировать, что схема работает правильно, порог переключения связующего инвертора не должен превышать $2,5 - 0,94 = 1,56$ В.

Наиболее распространенным и эффективным подходом к решению проблемы перераспределения заряда является предварительная зарядка внутрен-

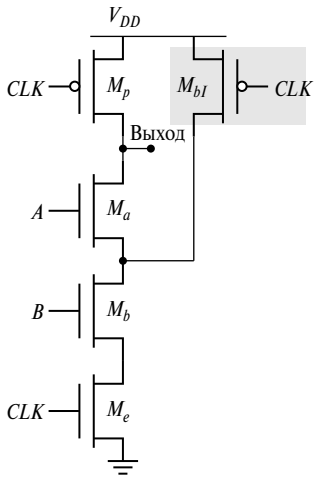


Рис. 6.61. Решение проблемы деления заряда путем предварительной зарядки внутренних узлов. Может использоваться n-МОП-транзистор предварительной зарядки, но это потребует инвертированного синхронизирующего сигнала

них критических узлов, как показано на рис. 6.61. Поскольку внутренние узлы заряжаются до V_{DD} в фазе предварительной зарядки, разделение заряда не происходит. Платой за данное решение, очевидно, является увеличение площади и емкости схемы.

Емкостная связь

Из-за относительно высокого импеданса выходного узла схема становится очень чувствительной к перекрестным эффектам. Проводка, проложенная поверх динамического узла (или рядом с ним), может образовывать с ним емкостную связь и разрушать состояние плавного узла. Другим не менее важным видом емкостной связи является *обратносмещенное связывание* (или связь типа *"выход на вход"*). Рассмотрим схему, показанную на рис. 6.62, *а*, в которой динамический вентиль НЕ-И с двумя входами питает статический вентиль НЕ-И. Переход на входе In статического вентиля может привести к переходу на нижний уровень выходного вентиля (Out_2). Данный переход на выходе образует емкостную связь с другим входом вентиля (динамический узел Out_1) через емкости затвора-истока и затвора-стока транзистора M_4. Результат моделирования этого эффекта показан на рис. 6.62, *б*. На приведенных графиках видно, как связывание приводит к значительному падению выходного напряжения динамического вентиля Out_1. В свою очередь, это не дает выходу статического элемента НЕ-И спокойно снизиться до 0 В, и происходит небольшое рассеяние статической мощности. Если падение напряжения достаточно велико, схема может работать неправильно, и выход элемента НЕ-И не станет низким. Следовательно, при разработке динамических цепей необходимо обратить особое внимание на минимизацию емкостной связи.

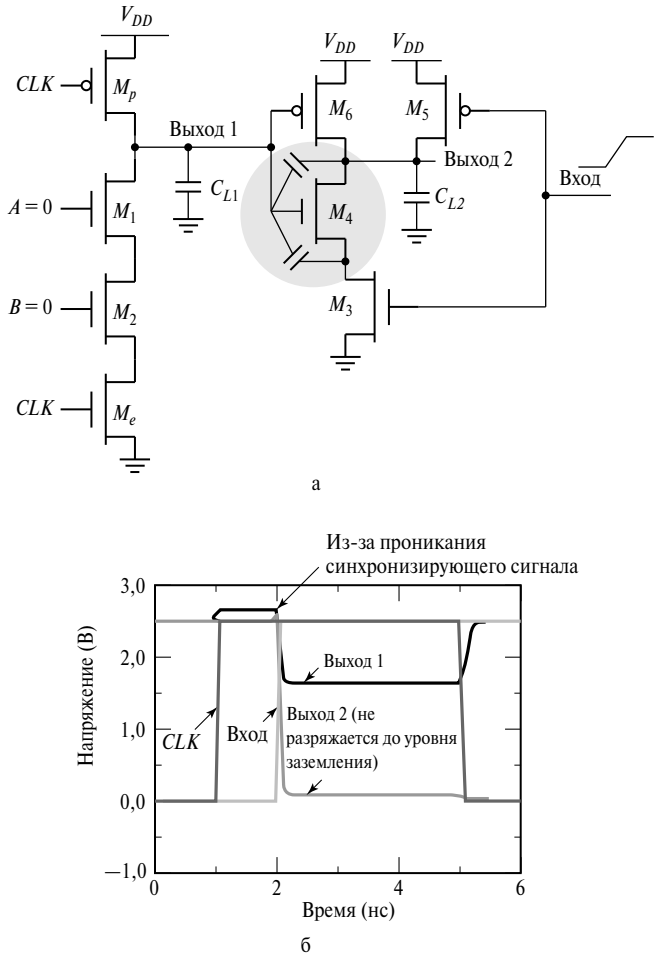


Рис. 6.62. Пример, демонстрирующий эффект обратносмещенного связывания: *а*) схема; *б*) результаты моделирования

Проникание тактового сигнала

Частным случаем емкостной связи является *проникание тактового сигнала* (clock feedthrough) — эффект, порождаемый емкостной связью между входным тактовым сигналом устройства предварительной зарядки и выходом динамического узла. Емкость связи состоит из емкости затвора-стока устройства предварительной зарядки и включает емкости перекрытия и канала. Данная емкостная связь приводит к повышению выхода динамического узла выше значения V_{DD} при изменении уровня тактового сигнала от низкого к высокому (предполагается, что разгрузочная сеть при этом выключена). Следовательно, нарастание и спад тактового сигнала соединяются в узле сигнала, что видно на рис. 6.62, *б*.

Опасность проникания тактового сигнала заключается в том, что оно может привести к тому, что плоскостные диоды транзистора предварительной зарядки, являющиеся в нормальном состоянии обратносмещенными, превратятся в прямосмещенные. Это приводит к инжекции электронов в подложку, причем эти электроны могут собраться ближайшим узлом с высоким импедансом, находящимся в состоянии **1**, и со временем вызвать неправильное функционирование схемы. Другим результатом этой инжекции может стать "защелкивание" КМОП-транзистора. В любом случае высокоскоростные динамические схемы необходимо тщательно промоделировать, чтобы гарантировать, что эффекты проникания тактового сигнала удерживаются в приемлемых пределах.

Из всех приведенных выше соображений видно, что проектирование динамических схем является очень сложным делом и требует особого внимания. Следовательно, разработчик должен использовать такие схемы только тогда, когда требуется высокое быстродействие или доступны хорошие средства автоматизации проектирования.

6.3.4. Каскадное соединение динамических элементов

Вопросы целостности сигнала порождают одну опасность, которая усложняет разработку динамических схем: простое каскадное соединение динамических элементов с целью получения многоуровневых логических структур не работает. Для иллюстрации этой проблемы используем два соединенных каскадом динамических инвертора n-типа, показанных на рис. 6.63, *а*. В фазе предварительной зарядки (т.е. когда $CLK = 0$) выходы обоих инверторов предварительно заряжаются до значения V_{DD}. Предположим, что на первичном входе In происходит переход $0 \rightarrow 1$ (см. рис. 6.63, *б*). На нарастающем фронте тактового сигнала выход Out_1 начинает разряжаться. Второй выход должен оставаться в предварительно заряженном состоянии — V_{DD}, — поскольку его ожидаемое значение равно 1 (выход Out_1 в фазе оценки переходит в 0). Тем не менее существует конечная задержка распространения входа, влияющая на разрядку Out_1 до уровня земли. Следовательно, второй выход также начинает разряжаться. Пока Out_1 превышает порог переключения второго вентиля (который приблизительно равен V_{Tn}), существует проводящий тракт между Out_2 и землей, и на выходе Out_2 заряд теряется. Проводящий тракт перекрывается только тогда, когда Out_1 достигает значения V_{Tn} и отключает разгрузочный n-МОП-транзистор. В результате выход Out_2 устанавливается на промежуточном уровне напряжения. Правильный уровень восстановлен не будет, поскольку динамические элементы зависят от запоминающих конденсаторов, тогда как статические элементы обеспечивают восстановление постоянной составляющей. Потеря заряда приводит к уменьшению запаса помехоустойчивости и потенциальной неверной работе.

Проблема каскадного соединения возникает из-за того, что выходы всех логических элементов (а следовательно, и входы следующих каскадов) предварительно заряжаются до значения 1. Это может привести к непреднамеренной разрядке в начале каждого цикла оценки. Чтобы снять этот вопрос,

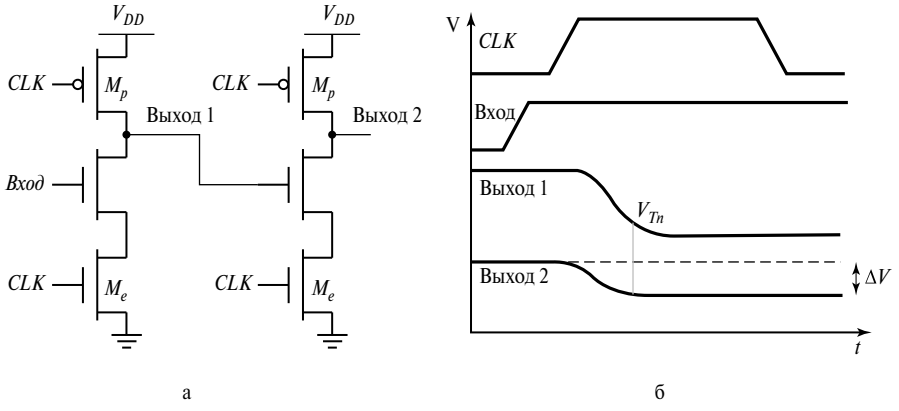


Рис. 6.63. Каскад динамических блоков n-типа

в фазе предварительной зарядки необходимо установить все входы равными 0. В таком случае все транзисторы в разгрузочной сети выключаются после предварительной зарядки, и в фазе оценки не может произойти никакой непреднамеренной разрядки запоминающих конденсаторов. Другими словами, правильная работа гарантируется тогда, когда в фазе оценки входы могут сделать только один переход $0 \rightarrow 1$[5]. Транзисторы включаются только тогда, когда они потребуются — т.е. не более одного раза за такт. В настоящее время описанный принцип реализован во множестве стилей разработки, два наиболее важных из них мы рассмотрим ниже.

Логика домино

Концепция

Логический модуль "домино" [Krambeck82] состоит из динамического логического блока n-типа, за которым следует статический инвертор (рис. 6.64). В фазе предварительной зарядки выход динамического вентиля n-типа заряжается до V_{DD}, и выход инвертора устанавливается равным 0. В фазе оценки динамический вентиль разряжается при выполнении заданного условия, и выход инвертора выполняет условный переход $0 \rightarrow 1$. Если предположить, что все входы элемента "домино" являются выходами других элементов "домино"[6], тогда все входы гарантированно будут установлены на 0 в конце фазы предварительной зарядки, и единственными переходами в фазе оценки будут переходы $0 \rightarrow 1$. Следовательно, сформулированное выше правило будет выполняться. Введение в схему статического инвертора полезно еще и тем, что разветвление по выходу вентиля питается статическим инвертором с низкоимпедансным выходом, который увеличивает помехоустойчивость. Кроме того, буфер уменьшает емкости динамического выходного узла, разделяя внут-

[5] Здесь мы игнорируем влияние рассмотренных ранее эффектов разделения заряда и утечки.
[6] Необходимо, чтобы все остальные входы, не подпадающие под эту классификацию (например, первичные входы), в фазе оценки оставались постоянными.

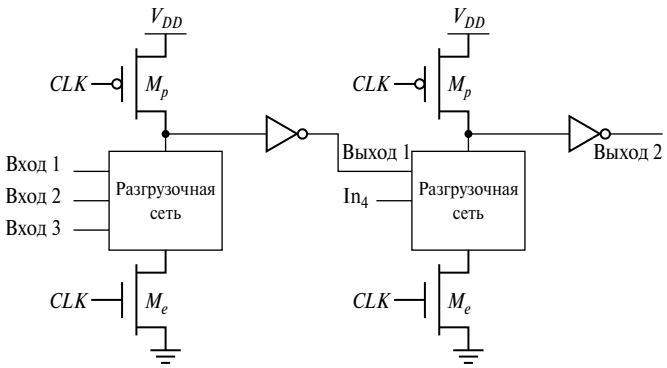


Рис. 6.64. КМОП-логика “домино”

реннюю емкость и емкость нагрузки. Наконец, инвертор можно использовать для питания цепи деления напряжения, что позволит решить проблему утечки и перераспределения заряда, как показано во втором каскаде схемы, приведенной на рис. 6.64.

Рассмотрим теперь работу цепочки элементов “домино”. В фазе предварительной зарядки все входы получают значение 0. В фазе оценки выход первого блока “домино” либо остается равным 0, либо выполняет переход $0 \rightarrow 1$, что затрагивает второй элемент. Данный эффект может распространяться по всей цепочке, подобно ряду падающих костяшек домино (отсюда и название). КМОП-схемы “домино” обладают следующими свойствами.

- Поскольку каждый динамический элемент имеет статический инвертор, реализовать можно только неинвертирующую логику. Хотя, как обсуждается в следующем разделе, данный ограничивающий фактор можно убрать, он является наибольшим ограничением, из-за которого чистая схема “домино” используется редко.
- Можно получить очень высокие скорости: существует только задержка нарастания, тогда как время t_{pHL} равно нулю. Размер инвертора можно выбрать согласующимся с *разветвлением по выходу*, которое обычно намного меньше, чем у статической КМОП-логики, поскольку для вентиля разветвления по выходу необходимо учесть емкость только одного элемента.

Поскольку входы элемента “домино” являются низкими в фазе предварительной зарядки, появляется искушение убрать транзистор оценки, поскольку это уменьшает нагрузку на тактовый сигнал и увеличивает скорость разгрузки. Но если убрать устройство оценки, то мы затронем и цикл предварительной зарядки, поскольку в таком случае предварительная зарядка должна будет точно так же распространяться по логической сети. Рассмотрим логическую схему, показанную на рис. 6.65, в которой были удалены устройства оценки. Если первичный вход In_1 равен 1 в фазе оценки, выход каждого динамического вентиля оценивается как 0, а выход каждого статического

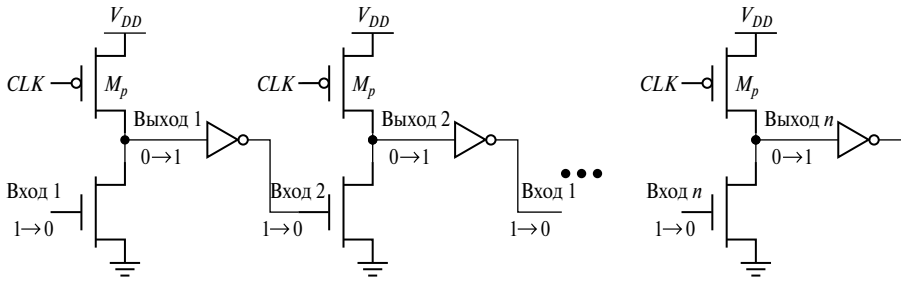


Рис. 6.65. Эффект распространения предварительного заряда при удалении из схему транзистора оценки. Помимо всего прочего, приведенная схема характеризуется еще и статическим рассеянием мощности

инвертора равен 1. Во время спада тактового сигнала начинается операция предварительной зарядки. Далее предположим, что на входе In_1 сигнал переходит с высокого на низкий уровень. Вход второго вентиля изначально является высоким, и перевод In_2 в низкое состояние происходит с задержкой, равной времени распространения сигнала через два элемента. Все это время второй вентиль не может предварительно зарядить свой выход, поскольку устройству предварительной зарядки мешает разгрузочная сеть. Подобным образом, третий вентиль должен ожидать завершения зарядки второго, а лишь после этого начинать свою предварительную зарядку и т.д. Следовательно, время, требуемое для предварительной зарядки логической схемы, равно времени критического пути этой схемы. Другим важным недостатком указанного технического решения является дополнительное рассеяние мощности, когда одновременно включены устройства нагрузки и разгрузки. Следовательно, устройства оценки все же нужны и их стоит использовать.

Борьба с невозможностью реализации инвертирующей логики

Основным ограничением логики "домино" является возможность реализации только неинвертирующей логики. Данное требование ограничивает широкое использование чистой логики "домино". Тем не менее существует несколько способов, позволяющих обойти его. Один подход к решению проблемы показан на рис. 6.66 — реорганизация логики с использованием простых булевых преобразований (например, закона Де Моргана). К сожалению, подобная оптимизация не всегда возможна, поэтому приходится использовать более универсальные схемы.

Универсальным (но дорогим) подходом к решению описанной проблемы является использование дифференциальной логики. Логика *"домино на дуальных шинах"*(dual-rail domino) по своей сути похожа на рассмотренную ранее структуру DCVSL, только вместо статической p-МОП-нагрузки с перекрестными обратными связями она использует предварительно заряженную нагрузку. На рис. 6.67 показана схема дифференциального логического вентиля И/НЕ-И. Обратите внимание на то, что все входы поступают от других дифференциальных вентилей "домино". В фазе предварительной зарядки они имеют низкий уровень, а в фазе оценки выполняют условный переход $0 \rightarrow 1$. Используя дифференциальную логику "домино", можно реализовать любую

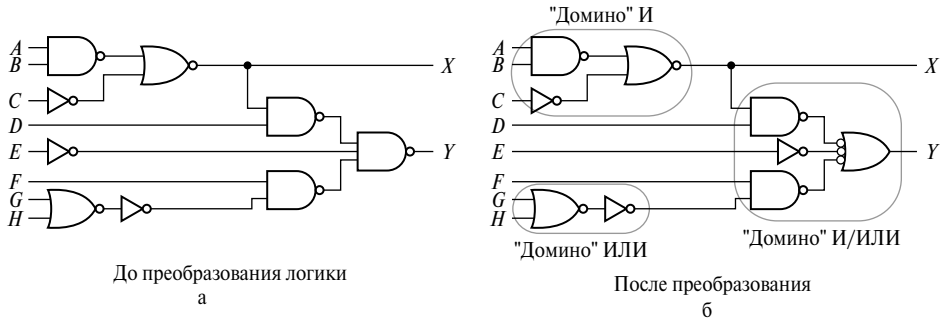


Рис. 6.66. Реструктуризация логики позволяет реализовать неинвертирующую логику “домино”

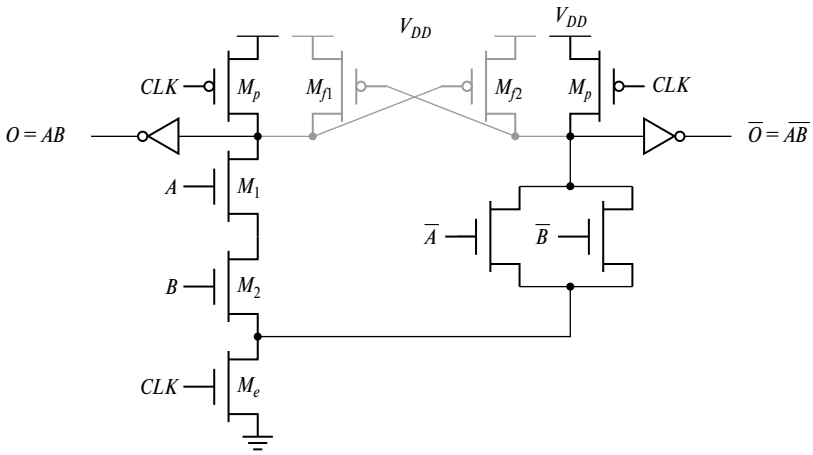


Рис. 6.67. Простой логический вентиль “домино” на дуальных шинах (дифференциальная логика)

функцию. Платой за это является увеличенное рассеяние мощности, поскольку независимо от входных значений в каждом такте будет выполняться переход — либо O, либо $\bar{O}$ должны совершить переход $0 \rightarrow 1$. Функция транзисторов M_{f1} и M_{f2} заключается в том, чтобы схема оставалась статической при высоком тактовом сигнале на протяжении больших промежутков времени (фактически мы имеем *стабилизатор*). Обратите внимание на то, что данная схема не является стандартизованной даже при наличии нагрузочных p-МОП-устройств! Из-за своего высокого быстродействия описанный дифференциальный подход очень популярен и используется в нескольких коммерческих микропроцессорах.

Оптимизация вентилей “домино”

Логические вентили “домино” можно оптимизировать по нескольким направлениям. Наиболее очевидная оптимизация быстродействия включает масштабирование транзисторов в статическом инверторе. При включении в схемы “домино” устройств оценки предварительная зарядка всех вентилей

происходит параллельно, следовательно, время операции предварительной зарядки равно задержке распространения через два вентиля — зарядка выхода динамического вентиля до V_{DD} и перевод выхода инвертора на низкий уровень. Критический путь в фазе оценки проходит через разгрузочный путь динамического вентиля и нагрузочный p-МОП-транзистор статического инвертора. Следовательно, чтобы ускорить работу схемы в фазе оценки, отношение размеров транзисторов статического инвертора необходимо сделать настолько большим, чтобы его порог переключения был близок к V_{DD}. Этого можно добиться, использовав малый (минимального размера) n-МОП-транзистор и большое p-МОП-устройство. Очевидно, n-МОП-транзистор минимального размера влияет только на время предварительной зарядки, которое обычно ограничено из-за параллельной предварительной зарядки всех вентилей. Единственным недостатком использования большого отношения размеров является снижение запаса помехоустойчивости. Следовательно, при выборе масштаба устройства разработчик должен рассматривать и его влияние на запас помехоустойчивости и быстродействие.

Существует несколько вариаций логики "домино" [Bernstein98]. Одной из них, направленной на уменьшение площади реализации, является *логика "домино" с несколькими выходами*. Основная концепция этого типа схем иллюстрируется на рис. 6.68. Свою роль здесь сыграл тот факт, что определенные выходы являются подмножествами других выходов (это позволяет генерировать несколько логических функций с помощью одного логического вентиля). В данном примере сигнал $O3 = C + D$ используется во всех трех выходах, следовательно, он реализуется внизу разгрузочной сети. Поскольку $O2$ равно $B \cdot O3$, мы можем повторно использовать логику, реализующую $O3$. Обратите внимание на то, что для получения правильных результатов внутренние узлы необходимо предварительно зарядить до значения V_{DD}. Если считать, что внутренние узлы предварительно заряжаются до V_{DD}, число устройств, питающих устройства предварительной зарядки, не уменьшается. Однако число транзисторов оценки при этом существенно уменьшается, поскольку они распределяются по нескольким выходам. Кроме того, описанный подход позволяет уменьшить коэффициент разветвления по выходу (также за счет повторного использования транзисторов для реализации нескольких функций).

Другой оптимизацией общего вентиля "домино" является *составная логика "домино"* (рис. 6.69), которая позволяет минимизировать число транзисторов. Вместо того чтобы каждый динамический вентиль питал статический инвертор, можно объединить выходы нескольких динамических вентилей с помощью сложного статического КМОП-элемента, как показано на рис. 6.69. Выходы трех динамических структур (реализующих функции $O1 = \overline{ABC}$, $O2 = \overline{DEF}$ и $O3 = \overline{GH}$) объединяются с использованием сложного статического КМОП-элемента, реализующего функцию $O = \overline{(O1 + O2)O3}$. В итоге вся схема реализует логическую функцию $O = ABCDEF + GH$.

Составные схемы "домино" являются очень полезным инструментом создания сложных динамических вентилей. Большие динамические стеки заменяются параллельными структурами с небольшими разветвлениями по входу и сложными КМОП-вентилями. Например, схему "домино" И с большим раз-

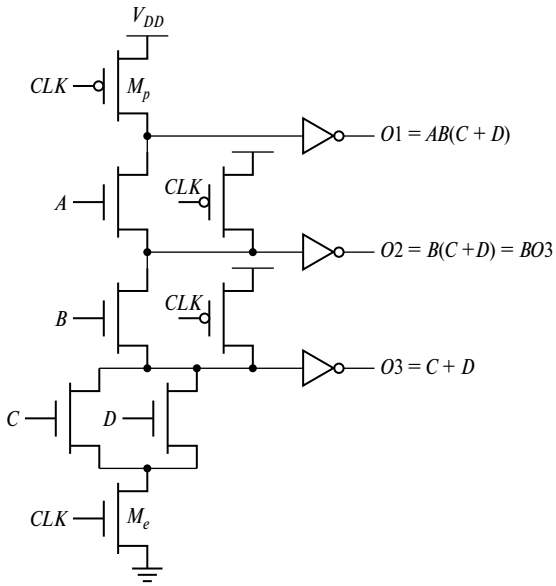


Рис. 6.68. Логика “домино” с несколькими выходами

ветвлением по входу можно реализовать как набор параллельных динамических структур НЕ-И с меньшими разветвлениями по входу, объединяемых посредством статического вентиля НЕ-ИЛИ. Кроме того, при разработке составных схем “домино” необходимо рассмотреть проблему обратносмещенного соединения. Следует уделить особое внимание тому, чтобы на работу динамических узлов не влияло соединение между выходом статических вентилей и выходом динамических узлов.

np-КМОП

Альтернативным подходом к каскадному соединению динамической логики является использование *np*-КМОП-схем, в которых применяются два типа динамической логики (с управляющим электродом *n*-типа и *p*-типа) и отсутствует дополнительный статический инвертор на критическом пути логики “домино”. В логическом вентиле *p*-типа *p*-МОП-устройства используются для построения нагрузочной логической схемы, включающей *p*-МОП-транзистор оценки ([Gonçalvez83], [Friedman84], [Lee86]) (рис. 6.70). *n*-МОП-транзистор предварительной зарядки переводит выход на низкий уровень в фазе предварительной зарядки. В фазе оценки выход в зависимости от входов выполняет переход $0 \rightarrow 1$.

np-КМОП-логика построена на использовании дуальности логических элементов с *n*- и *p*-деревом, что позволяет снять проблему каскадного соединения. Если вентили с *n*-деревьями контролируются тактовым сигналом CLK, а вентили с *p*-деревьями контролируются с использованием $\overline{CLK}$, деревья элементов с управляющим электродом *n*-типа могут непосредственно

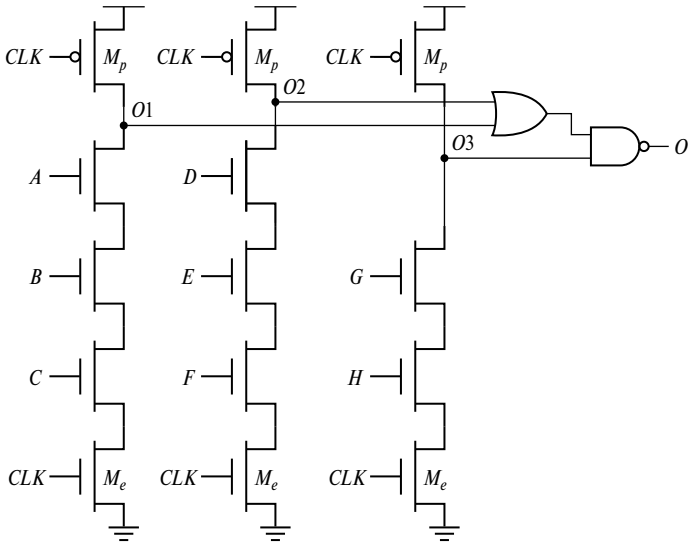


Рис. 6.69. В составной логике “домино” на выходе динамических вентилей используются сложные статические вентили

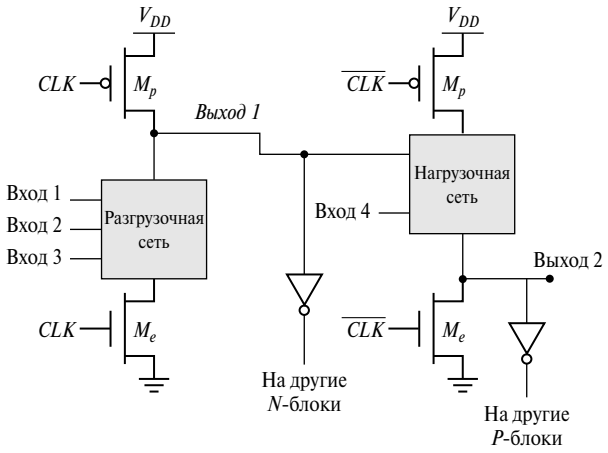


Рис. 6.70. Стиль схем np-КМОП-логики

питать деревья элементов p-типа и наоборот. Подобно элементам “домино”, выходы n-деревьев должны проходить через инвертор при соединении с другим элементом n-дерева. В фазе предварительной зарядки ($CLK = 0$) выход n-дерева (Out_1) заряжается до значения V_{DD}, тогда как выход элемента с p-деревом (Out_2) предварительно разряжается до 0 В. Поскольку элемент n-типа соединяет нагрузочные p-МОП-устройства, нагрузочная сеть p-дерева в это время выключается. В фазе оценки выход n-вентиля может выполнить только переход $1 \rightarrow 0$, при выполнении заданного условия включая

некоторые транзисторы p-дерева. Таким образом, гарантируется, что случайной разрядки Out_2 не произойдет. Подобным образом, блоки n-деревьев могут без проблем располагаться после p-деревьев, поскольку входы вентилей с управляющим электродом n-типа предварительно заряжаются до значения 0. Недостаток np-КМОП-логики заключается в том, что блоки p-типа медленнее модулей n-типа из-за меньшего тока питания p-МОП-транзисторов в логической схеме. Выравнивание задержек распространения требует дополнительной площади.

6.4. Перспективы

6.4.1. Как выбрать стиль разработки логики

В предыдущих разделах мы обсудили несколько подходов к реализации логических вентилей с использованием технологии КМОП. Все стили разработки имеют свои достоинства и недостатки. Выбор конкретного из них зависит от того, какое требование к разработке является основным: легкость проектирования, устойчивость, площадь реализации, скорость или рассеяние мощности. Ни один стиль не позволяет одновременно оптимизировать все названные параметры. Даже больше — предпочтительный выбор может отличаться для различных логических функций.

Достоинством статического подхода является устойчивость при наличии помех. Это немного облегчает процесс проектирования и позволяет использовать в нем различные средства автоматизации разработки. Данный стиль является наиболее универсальным. Тем не менее легкость разработки имеет свою цену: для сложных логических элементов с большим разветвлением по входу технология КМОП становится чересчур дорогой с точки зрения площади и быстродействия. Следовательно, потребовалась альтернативная статическая логика. Простым и быстрым решением являются псевдо-n-МОП-элементы, недостаток которых — сниженная помехоустойчивость и статическое рассеяние мощности. Для реализации определенных типов схем (например, мультиплексоров и элементов с большим количеством вентилей ИСКЛЮЧАЮЩЕЕ ИЛИ, подобных сумматорам) удачной альтернативой является логика на проходных транзисторах.

С другой стороны, динамическая логика позволяет реализовать быстрые и маленькие сложные элементы. Однако и это имеет свою цену. Из-за паразитных эффектов (например, перераспределения заряда) проектирование таких схем является очень рискованным. Утечка заряда приводит к периодическому обновлению, что, в свою очередь, определяет минимальную возможную рабочую частоту схемы.

В настоящее время наблюдается тенденция к более интенсивному использованию статической КМОП-логики. Поддержанию этой тенденции способствует и более широкое распространение инструментов автоматизированного проектирования. Данные инструменты позволяют оптимизировать схему на уровне логики, а не на уровне схемы, и делают акцент на надежности ее работы. Еще одним аргументом в пользу статической КМОП-логики являет-

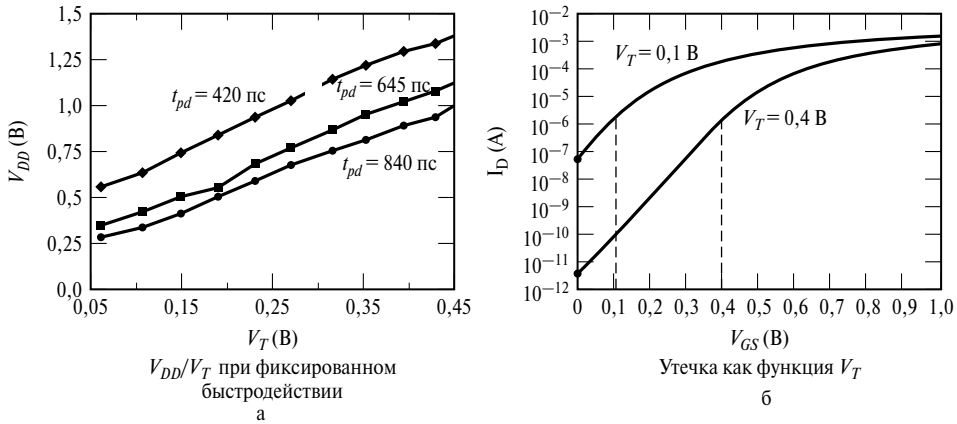


Рис. 6.71. Масштабирование напряжения

ся то, что она лучше поддается масштабированию напряжения, чем другие технологии, рассмотренные в этой главе.

6.4.2. Разработка логики со сниженным потреблением напряжения

В главе 3 мы указывали, что напряжение питания для КМОП-процессов будет в текущем десятилетии продолжать падать и к 2010 году может понизиться до 0,6 В. Чтобы в таких условиях сохранить имеющееся быстродействие, необходима возможность масштабирования порогов устройств. На рис. 6.71, *а* показан график зависимости V_T от V_{DD}, требуемой для поддержания данного уровня быстродействия (предполагается, что характеристики устройства остаются неизменными).

Указанный компромисс имеет и свои недостатки. Уменьшение порогового напряжения экспоненциально увеличивает подпороговые токи утечки, как следует из уравнения (3.39) (ниже оно повторяется для удобства):

$$I_{leakage} = I_S 10^{\frac{V_{GS}-V_{th}}{S}} \left(1 - 10^{-\frac{n\ V_{DS}}{S}}\right). \qquad (6.46)$$

В уравнении (6.46) S — это *коэффициент спада* устройства. Подпороговая утечка инвертора — это ток n-МОП-транзистора при $V_{\text{вх.}} = 0$ В и $V_{\text{вых.}} = V_{DD}$ (или ток p-МОП-транзистора при $V_{\text{вх.}} = V_{DD}$ и $V_{\text{вых.}} = 0$). Экспоненциальное увеличение тока утечки для повышенных порогов иллюстрируется на рис. 6.71, *б*.

Особенно токи утечки важны при проектировании схем, характеризующихся резкими периодами вычислительной активности, разделенными длительными периодами неактивности. Например, процессор сотового телефона неактивен бо́льшую часть времени. В это время система (в идеальном случае) должна потреблять нулевую (или почти нулевую) энергию. Это возможно

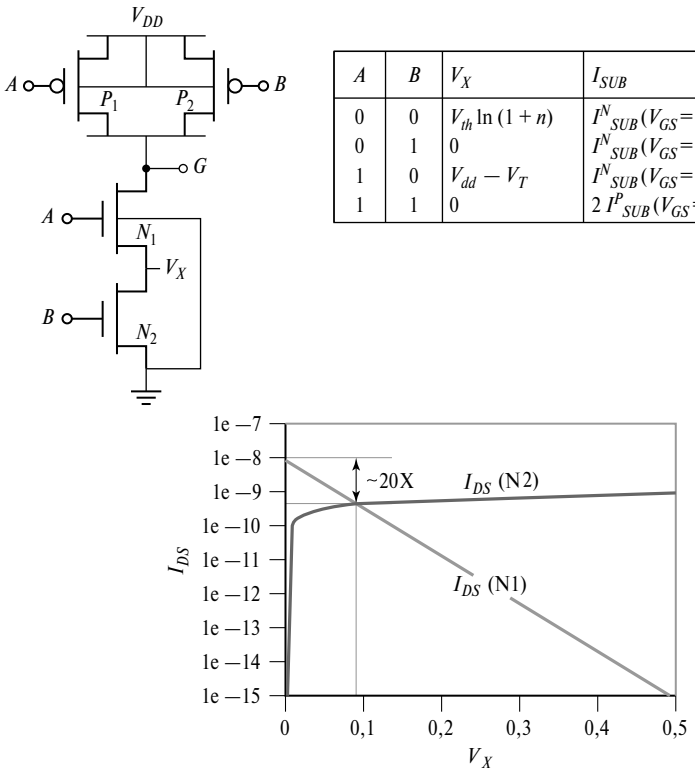
только в том случае, если утечка мала, т.е. устройство имеет высокое пороговое напряжение. Это противоречит описанному выше сценарию масштабирования, когда высокое быстродействие при низком напряжении питания означает уменьшенные пороги. Чтобы удовлетворить противоречивым требованиям высокого быстродействия в активные периоды и низкой утечки в ненагруженном режиме, КМОП-процессы были несколько модифицированы путем внедрения технологий контроля утечки. Большинство процессов с характерным размером порядка 0,18 мкм или ниже поддерживают устройства с различными порогами (обычно устройство с низким порогом для быстродействующих схем и транзистор с высоким порогом для контроля утечки). Существует и другой подход, популярность которого постоянно растет, — динамический контроль над пороговым напряжением устройства путем использования влияния подложки. Отметим, что для применения этих подходов для контроля над отдельными устройствами необходимо использование двухкарманного процесса (см. рис. 2.2).

Правильная разработка позволяет уменьшить ток утечки, который зависит от топологии схемы и значений сигналов, поданных на вход вентиля. Поскольку V_T зависит от смещения подложки (V_{BS}), подпороговая утечка МОП-транзистора зависит не только от напряжения питания V_{GS}, но и от напряжения смещения подложки. В инверторе с $In = 0$ подпороговая утечка инвертора определяется n-МОП-транзистором с его $V_{GS} = V_{BS} = 0$ В. В более сложных КМОП-вентилях, например, показанном на рис. 6.72 вентиле НЕ-И с двумя входами, ток утечки зависит от вектора входов. Подпороговый ток утечки данного вентиля минимален при $A = B = 0$. В таком случае промежуточный узел X имеет следующее напряжение:

$$V_X \approx V_{th} \ln(1 + n). \tag{6.47}$$

Ток утечки вентиля определяется самым верхним n-МОП-транзистором с $V_{GS} = V_{BS} = -V_X$. Очевидно, подпороговая утечка в описанном случае меньше, чем у инвертора. Данное уменьшение утечки из-за размещения транзисторов в несколько ярусов называется *эффектом тяги*. В таблице, представленной на рис. 6.72, анализируются компоненты утечки для вентиля НЕ-И с двумя входами при различных входных условиях.

В реальной жизни ситуация выглядит даже еще лучше. В короткоканальных МОП-транзисторах подпороговые токи утечки зависят не только от напряжения питания затвора (V_{GS}) и смещения подложки (V_{BS}), но и от напряжения стока (V_{DS}). Пороговое напряжение короткоканального МОП-транзистора уменьшается с увеличением V_{DS} из-за *индуктированного стоком понижения барьера* (Drain-Induced Barrier Lowering — DIBL). Типичные значения DIBL составляют порядка 20–150 мВ изменения V_T при изменении напряжения V_{DS}. Из-за этого влияние эффекта тяги является еще бо́льшим для короткоканальных транзисторов. Промежуточное напряжение уменьшает напряжение стока-истока самого верхнего устройства, увеличивая его порог, а следовательно, уменьшая утечку.



A	B	V_X	I_{SUB}
0	0	$V_{th} \ln(1+n)$	$I^N_{SUB}(V_{GS} = V_{BS} = -V_X)$
0	1	0	$I^N_{SUB}(V_{GS} = V_{BS} = 0)$
1	0	$V_{dd} - V_T$	$I^N_{SUB}(V_{GS} = V_{BS} = 0)$
1	1	0	$2\,I^P_{SUB}(V_{GS} = V_{BS} = 0)$

Рис. 6.72. Уменьшение подпороговой утечки в вентили НЕ-И с двумя входами из-за образования тяги для различных входных условий. На графике представлены смоделированные линии нагрузки логического элемента для случая $A = B = 0$

Пример 6.19. Эффект тяги в вентиле НЕ-И с двумя входами

Рассмотрим еще раз вентиль НЕ-И с двумя входами, изображенный на рис. 6.72, *а*, когда транзисторы N_1 и N_2 выключены ($A = B = 0$). Из смоделированных линий нагрузки, показанных на рис. 6.72, *в*, видно, что в установившемся состоянии V_X примерно равно 100 мВ. Следовательно, подпороговая утечка в установившемся состоянии порождается напряжением $V_{GS} = V_{BS} = -100$ мВ и $V_{DS} = V_{DD} - 100$ мВ, что в 20 раз меньше, чем утечка автономного n-МОП-транзистора с $V_{GS} = V_{BS} = 0$ мВ и $V_{DS} = V_{DD}$ [Ye98].

Подведем итоги: подпороговая утечка в сложных многоуровневых схемах может быть существенно меньше, чем в отдельных устройствах. Отметим, что максимальное снижение утечки происходит, когда все транзисторы *выключены* и напряжение промежуточного узла достигает установившегося значения. Исследование этого эффекта требует тщательного выбора входных сигналов для каждого вентиля в стационарном или дежурном режиме.

Задача 6.9. Расчет V_X

В уравнении (6.47) рассчитывается напряжение промежуточного узла для вентиля НЕ-И с двумя входами с уровнем ошибок, не превышающим 10%, при $A = B = 0$. Выведите уравнение (6.47), предполагая, что (1) зависимости V_T и I_S от N_1 и N_2 приблизительно идентичны; (2) n-МОП-транзисторы имеют одинаковые размеры; (3) $n < 1,5$.

6.5. Резюме

В этой главе мы тщательно проанализировали поведение и быстродействие комбинационных цифровых КМОП-схем относительно площади, скорости и мощности. Сделаем основные выводы.

- *Статические комплементарные КМОП-схемы* объединяют разгрузочную и нагрузочную сети, только одна из которых активизирована в конкретный момент времени.
- Быстродействие КМОП-вентиля сильно зависит от *разветвления по входу*. Для регулирования разветвления по входу используются масштабирование транзисторов, переупорядочение входов и секционирование. Скорость схемы также является линейной функцией разветвления по выходу. При больших разветвлениях по выходу в схему необходимо включать дополнительные буферы.
- *Стандартизованная логика* состоит из активной разгрузочной (нагрузочной) сети, соединенной с устройством нагрузки. Это позволяет существенно снизить сложность вентиля за счет статического потребления мощности и асимметричной характеристики. Для получения приемлемого запаса помехоустойчивости необходим аккуратный выбор размеров транзистора. Наиболее популярные подходы к разработке схем данного класса заключаются в использовании псевдо-n-МОП-элементов и дифференциальных DCVSL-схем, требующих комплементарных сигналов.
- При использовании *логики на проходных транзисторах* логический элемент реализуется как простая сеть-переключатель. Таким образом, некоторые логические функции удается реализовать очень просто. Длинных каскадов переключателей следует избегать из-за квадратичного роста задержки от числа элементов в цепочке. Логика, основанная только на проходных n-МОП-транзисторах, дает даже еще более простые структуры, однако может характеризоваться статическим потреблением мощности и уменьшенным запасом помехоустойчивости. Данную проблему можно решить, добавив в схему транзистор восстановления уровня.
- Работа *динамической логики* основана на хранении заряда на емкостном узле и условной разрядке этого узла в зависимости от входов. В результате была разработана двухфазная схема, в которой реализована фаза пред-

варительной зарядки с последующим этапом оценки. Динамическая логика — это "размен" запаса помехоустойчивости на быстродействие. Она чувствительна к таким паразитным эффектам, как утечка, перераспределение заряда и проникание тактового сигнала. Кроме того, возможны проблемы с каскадным соединением динамических элементов; следовательно, проектировать их следует очень аккуратно.

- *Потребление мощности* логической схемы сильно зависит от активности переключения сети. Данная активность является функцией статистики входов, топологии сети и стиля логики. Влияние таких источников потребления мощности, как импульсные помехи и токи короткого замыкания, можно минимизировать с помощью тщательного проектирования и выбора размеров транзисторов.
- Для работы при низком напряжении необходимо масштабирование напряжения. Особое внимание при этом следует уделять контролю над утечкой.

6.6. Для любознательных

Тема стилей (К)МОП-логики широко освещается в литературе. Данному вопросу посвящено огромное множество работ. Одними из самых подробных являются [Weste93] и [Chandrakasan01]. Если вас интересуют детали сложного процесса разработки быстродействующих схем, обратитесь к книгам [Shoji96] и [Bernstein98], в которых очень подробно рассмотрены оптимизация и анализ цифровых МОП-схем. Тема минимизации мощности сравнительно новая, довольно подробно она рассмотрена в [Chandrakasan95], [Rabaey95] и [Pedram02].

Новые достижения в сфере МОП-логики обычно публикуются в материалах *Proceedings of The International Solid-State and Circuits Conference* и *Proceedings of the VLSI Circuits Symposium*, а также в *IEEE Journal of Solid State Circuits* (особенно в ноябрьских выпусках).

6.7. Список использованной литературы

[Bernstein98] K. Bernstein et al., *High-Speed CMOS Design Styles*, Kluwer Academic Publishers, 1998.

[Chandrakasan95] A. Chandrakasan and R. Brodersen, *Low Power Digital CMOS Design*, Kluwer Academic Publishers, 1995.

[Chandrakasan01] A. Chandrakasan, W. Bowhill, and F. Fox, ed., *Design of High-Performance Microprocessor Circuits*, IEEE Press, 2001.

[Goncalvez83] N. Goncalvez and H. De Man, "NORA: A Racefree Dynamic CMOS Technique for Pipelined Logic Structures", *IEEE Journal of Solid State Circuits*, June 1983, vol. SC-18, no. 3, p. 261–266.

[Heller84] L. Heller et al., "Cascade Voltage Switch Logic: A Differential CMOS Logic Family", *Proc. IEEE ISSCC Conference*, February 1984, p. 16-17.

[Krambeck82] R. Krambeck et al., "High-Speed Compact Circuits with CMOS", *IEEE Journal of Solid State Circuits*, June 1982, vol. SC-17, no. 3, p. 614–619.

[Landman91] P. Landman and J. Rabaey, "Design for Low Power with Applications to Speech Coding", *Proc. International Micro-Electronics Conference*, Cairo, December 1991.

[Parameswar96] A. Parameswar, H. Hara, and T. Sakurai, "A Swing Restored Pass-Transistor Logic-Based Multiply and Accumulate Circuit for Multimedia Applications", *IEEE Journal of Solid State Circuits*, June 1996, vol. SC-31, no. 6, p. 805–809.

[Pedram02] M. Pedram and J. Rabaey, ed., *Power-Aware Design Methodologies*, Kluwer, 2002.

[Rabaey95] J. Rabaey and M. Pedram, ed., *Low Power Design Methodologies*, Kluwer, 1995.

[Radhakrishnan85] D. Radhakrishnan, S. Whittaker, and G. Maki, "Formal Design Procedures for Pass-Transistor Switching Circuits", *IEEE Journal of Solid State Circuits*, April 1985, vol. SC-20, no. 2, p. 531–536.

[Shoji88] M. Shoji, *CMOS Digital Circuit Technology*, Prentice Hall, 1988.

[Shoji96] M. Shoji, *High-Speed Digital Circuits*, Addison-Wesley, 1996.

[Sutherland99] I. Sutherland, B. Sproull, and D. Harris, *Logical Effort*, Morgan Kaufmann, 1999.

[Weste93] N. Weste and K. Eshragian, *Principles of CMOS VLSI Design: A Systems Perspective*, Addison-Wesley, 1993.

[Yano90] K. Yano et al., "A 3–8 ns CMOS 16 × 16 b Multiplier Using Complimentary Pass-Transistor Logic", *IEEE Journal of Solid State Circuits*, April 1990, vol. SC-25, no. 2, p. 388–395.

[Ye98] Y. Ye, S. Borkar, and V. De, "A New Technique for Standby Leakage Reduction in High-Performance Circuits", *Symposium on VLSI Circuits*, 1998, p. 40-41.

Упражнения

Последнюю версию сборника задач и упражнений можно найти по адресу `http://bwrc.eecs.berkley.edu/IcBook`.

Вставка

B

Моделирование сложных логических схем

Временное моделирование и моделирование на уровне переключателей

Логическое и функциональное моделирование

Поведенческое моделирование

Языки моделирования на уровне регистровых передач

Хотя моделирование схемы с помощью пакета, подобного SPICE, является чрезвычайно полезной составляющей набора инструментов разработчика, оно обладает одним существенным недостатком. Если учитывать все нюансы и эффекты второго порядка полупроводниковых устройств, то данный процесс может быть чересчур длительным. Если вы не собираетесь несколько дней ждать результаты компьютерных расчетов сложной схемы, этот путь невыгоден. Несмотря на то что компьютеры становятся все быстрее, а средства моделирования — лучше, сложность схем все равно увеличивается быстрее. Чтобы справиться с проблемой сложности, разработчик может задать точность моделирования и прибегнуть к более высоким уровням представления. Ниже мы рассмотрим различные уровни абстракции, доступные разработчику, и их влияние на точность моделирования.

Самый простой способ разобраться во множестве подходов к моделированию и уровней абстракции — это определить, как представляются данные и функции времени — в аналоговой форме, как непрерывные переменные, как дискретные сигналы или как абстрактные модели данных.

Представление цифровых данных в виде непрерывной сущности

Во вставке Б мы установили, что моделирование схемы, по сути, выполняется аналоговыми средствами. Напряжение, сила тока и время считаются аналоговыми переменными. Из-за такой точной модели и нелинейности большинства устройств моделирование занимает довольно много времени.

Разработчики сделали довольно много, чтобы уменьшить время вычислений за счет обобщения. Рассмотрим цифровую МОП-схему. Из-за прекрасной изоляции МОП-элемента схему довольно часто удается разбить на ряд фрагментов, мало взаимодействующих между собой. Далее для данного промежутка времени мы можем рассчитать каждый такой фрагмент отдельно, считая, что входы от других фрагментов известны или фиксированы. Получающиеся в результате сигналы можно уточнять итеративно. Такой *релаксационный* подход вычислительно эффективнее, чем традиционная техника, поскольку в нем отсутствуют дорогие операции обращения матриц, однако применять его можно только к МОП-схемам [White87]. Если схема содержит цепи обратной связи, фрагменты могут стать большими и производительность моделирования снизится.

Другой подход заключается в уменьшении сложности используемой эквивалентной схемы транзистора. Например, линеаризация модели приводит к существенному снижению вычислительной сложности. Кроме того, можно использовать упрощенную модель таблицы соответствий. Хотя такой подход приводит к уменьшению точности описания сигналов, он по-прежнему позволяет хорошо оценить такие временн*ы*е параметры, как задержка распространения и время нарастания/спада. Именно поэтому соответствующие инструменты часто называют *временны*ми. Серьезным их преимуществом является скорость выполнения, которая может на один-два порядка превышать скорость моделирования средств, подобных SPICE. Другое преимущество заключается в том, что в отличие от рассмотренных ниже инструментов средства временн*о*го моделирования учитывают такие эффекты второго порядка, как утечка, падение порога и импульсные помехи. В качестве примеров средств данного класса можно привести набор инструментов NanoSim (ранее — TimeMill/PowerMill) производства Synopsys [TimeMill]. Для справки: точность инструментов подобного типа обычно всего на 5–10% меньше точности полнофункциональных средств моделирования схем.

Представление данных в виде дискретной сущности

В цифровых схемах нас обычно не интересует реальное значение функции напряжения, важно знать только цифровое значение, которое она представляет. Следовательно, мы можем разработать такое средство моделирования, чтобы сигналы данных принадлежали диапазону 0 или 1. Сигналы, которые не согласуются ни с одним условием, обозначаются как X (т.е. неопределенные).

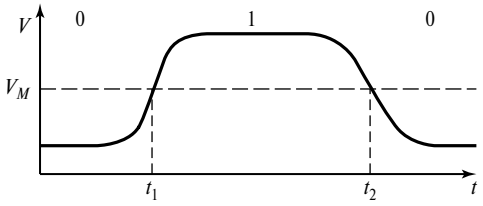


Рис. В.1. Дискретизация функции времени

Такое третичное представление $\{0, 1, X\}$ широко используется в моделировании, как на уровне устройств, так и на уровне отдельных логических элементов. Дополняя данное множество допустимых значений данных, мы можем получить более подробную информацию, при этом сохраняя возможность описания сложных структур. Возможными исключениями являются значение Z для узла с тремя состояниями, соответствующее высокоимпедансному состоянию, значения R и F, соответствующие переходам в более высокое и более низкое состояние. В некоторых коммерческих продуктах число возможных состояний сигнала доходит до десятка.

Хотя дискретизировав пространство представлений данных, мы существенно повысим производительность средства моделирования, аналогичные преимущества можно получить, сделав время дискретной переменной. Рассмотрим форму кривой напряжения, изображенную на рис. В.1, которая представляет сигнал на входе инвертора с порогом переключения V_M. Разумно предположить, что выход инвертора меняет свое значение в течение промежутка времени, равного задержке распространения, после того как его вход проходит через значение V_M. Если вам не требуется знать точную форму сигнала, достаточно оценить ее значение только в нужные моменты времени — t_1 и t_2.

Подобным образом, значение выходного сигнала интересует нас в моменты времени $t_1 + t_{pHL}$ и $t_2 + t_{pLH}$. Программа моделирования, оценивающая состояние элемента только в те моменты, когда на одном из его входов происходит какое-либо событие, называется программой, *управляемой событиями*. Порядок оценки определяется путем помещения спроектированных событий во временную очередь и последовательной их обработки. Предположим, что сигнал, изображенный на рис. В.1, подается на вход логического элемента. Событие должно произойти в момент времени t_1. При обработке этого события для узлов разветвления по выходу планируется новое событие в момент времени $t_1 + t_{pHL}$, которое затем помещается во временную очередь. Очевидно, подобный подход, управляемый событиями, эффективнее, чем пошаговый подход программ моделирования электрических схем. Чтобы учесть влияние разветвления по выходу, задержку распространения схемы можно выразить через внутреннюю задержку (t_{in}) и параметр, зависящий от нагрузки (t_l):

$$t_{pLH} t_{inLH} + t_{lLH} \times C_L. \tag{В.1}$$

Нагрузку C_L можно подставлять в абсолютных единицах (в пФ) или как функцию числа элементов разветвления по выходу. Обратите внимание на то, как сильно данное уравнение похоже на модель *логических трудозатрат*, которую мы ввели в предыдущей главе.

Предыдущий подход хотя и предлагает существенные преимущества с точки зрения быстродействия, обладает и недостатком — события могут произойти в любой момент времени. Можно сделать еще одно упрощение, задав более мелкую дискретизацию времени и позволив событиям происходить только в моменты, кратные некоторому *единичному времени*. Примером такого подхода служит модель с *единичными задержками*, когда каждая схема характеризуется единичной задержкой. Наконец, наиболее простой является модель с *нулевой задержкой*, в которой предполагается, что все логические элементы не имеют задержек. При таком подходе время идет от одного тактового события до другого, а все события считаются происходящими одновременно в момент смены такта. Описанные концепции можно применить на множестве уровней абстракции, что привело к возникновению большого числа подходов к моделированию.

Моделирование на уровне переключателей

Одной из основных преград высоким скоростям моделирования является нелинейная природа полупроводниковых устройств. Справиться с этой помехой позволяет модель на уровне переключателей [Bryant81], в которой поведение транзисторов аппроксимируется линейным сопротивлением, значение которого является функцией режима работы. В нерабочем режиме сопротивление устанавливается равным бесконечности, иначе оно равно сопротивлению прибора во включенном состоянии (рис. В.2). Получающаяся в результате сеть является нестационарной линейной сетью резисторов и конденсаторов, анализировать которую можно более эффективно. При расчете резистивной цепи определяются установившиеся значения сигналов и обычно применяется модель $\{0, 1, X\}$. Например, если общее сопротивление между узлом и землей существенно меньше сопротивления с узлом V_{DD}, состояние узла устанавливается нулевым. Распределение событий по времени можно разрешить, проанализировав RC-цепь. Кроме того, могут использоваться и более простые модели, например модель с единичными задержками.

Пример В.1. Моделирование на уровне переключателей или на уровне цепей

Четырехбитовый сумматор моделируется с использованием средства моделирования на уровне переключателей IRSIM ([Salz89]). Полученные результаты изображены на рис. В.3. Изначально все входы ($IN1$ и $IN2$), а также входной сигнал переноса CIN установлены равными 0. После 10 нс все входы $IN2$ и CIN равны 1. В приведенном окне изображены входные сигналы, вектор выхода OUT[0–3], а также старшие биты выхода OUT[2] и OUT[3]. Выход сходится к правильному значению 0000 по прошествии переходного перио-

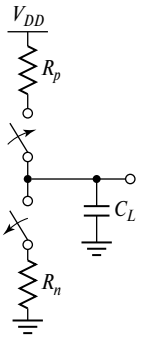


Рис. В.2. Модель КМОП-инвертора на уровне переключателей

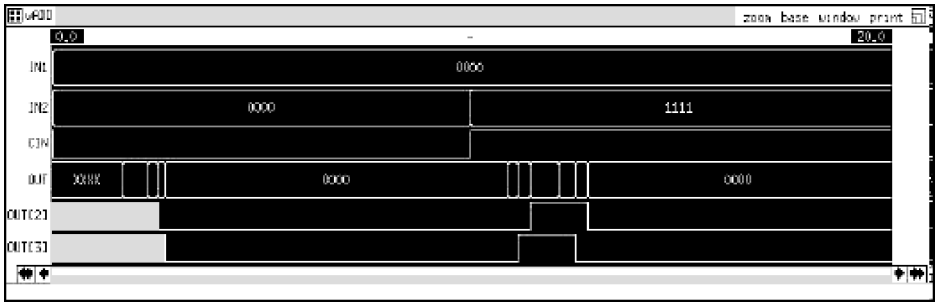
да. Обратите внимание, что для данных предполагается наличие всего двух уровней — 0 и 1. Изменения выходных сигналов происходят с полным размахом, хотя в действительности это могут быть лишь небольшие отклонения. Во время переходов сигнал помечается как X, т.е. неопределенный. Чтобы лучше понять приведенные данные, на рис. В.3 приведены результаты моделирования в SPICE для тех же входных векторов. Видно, что распределение событий во времени с помощью IRSIM (основанное на *RC*-модели) является относительно точным и достаточным для оценки в первом приближении.

Моделирование на уровне логических элементов

Программы моделирования на уровне логических элементов используют те же значения сигналов, что и инструменты моделирования на уровне переключателей, однако в этом случае основными примитивами являются не транзисторы, а логические элементы. Данный подход разрешает моделировать более сложные схемы за счет увеличения степени детализации и универсальности. Например, такие распространенные СБИС-структуры, как шины с тремя состояниями и проходные транзисторы, очень сложно рассматривать на таком уровне. Поскольку уровень логических элементов удобен для многих разработчиков, указанные средства моделирования были весьма популярными до появления инструментов синтеза логики, которые сместили фокус разработки на функциональный или поведенческий уровень абстракции. Интерес к моделированию логических схем был настолько велик, что для ускорения процесса моделирования были даже разработаны специальные и довольно дорогие аппаратные ускорители (например, [Agrawal90]).

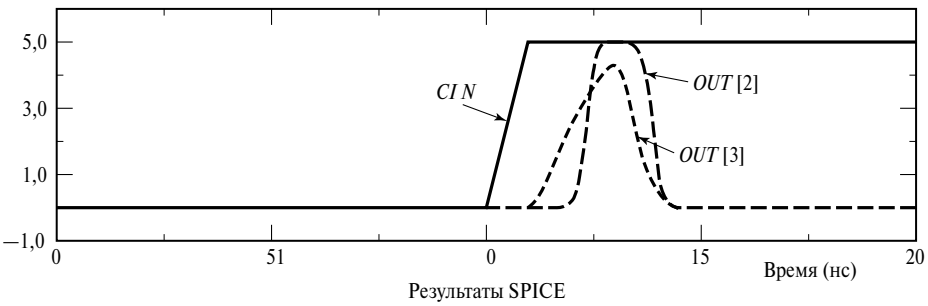
Функциональное моделирование

Функциональное моделирование можно считать простым расширением логического. Примитивные элементы входного описания могут иметь любую



Результаты IRSIM

а



Результаты SPICE

б

Рис. B.3. Сравнение моделирования на уровне схем и на уровне переключателей

сложность. Например, в качестве основы моделирования могут применяться логические элемент НЕ-И, умножитель или статическое ОЗУ. Функциональные возможности одного из указанных сложных элементов можно описать, используя современный язык программирования или специализированный аппаратный описательный язык. Например, для описания выходных значений модуля через его входы средство моделирования THOR использует язык программирования C [Thor88].

Язык *SystemC* [SystemC] очень похож по синтаксису и семантике на C, однако, кроме того, он обладает рядом конструкций и типов данных, учитывающих особенности цифровых разработок, например одновременность процессов. С другой стороны, существует язык *VHDL* (VHSIC Hardware Description Language — аппаратный описательный язык сверхбыстродействующих интегральных схем) [VHDL88], специально предназначенный для описания аппаратных разработок.

В *структурном режиме* язык VHDL описывает проект как набор соединений функциональных модулей. Такое описание часто называется *таблицей соединений* (netlist). Например, на рис. B.4 показано описание 16-битового накапливающего сумматора, состоящего из регистра и сумматора.

В свою очередь сумматор и регистр можно описать как набор таких компонентов, как полный сумматор или ячейки регистра. Можно также применить альтернативный *поведенческий режим* языка, независимо от выбран-

```
entity accumulator is
   port (--  определение входных и выходных контактов
     DI: in bit_vector(15 downto 0) --  вектор, шириной 16 бит
     DO: inout bit_vector(15 downto 0);
     CLK: in bit
   );
end accumulator;

architecture structure of accumulator is
   component reg --   определение портов регистра
     port (
       DI : in bit_vector(15 downto 0);
       DO : out bit_vector(15 downto 0);
       CLK : in bit
     );
   end component;
   component add --   определение портов сумматора
     port (
       IN0 : in bit_vector(15 downto 0);
       IN1 : in bit_vector(15 downto 0);
       OUT0 : out bit_vector(15 downto 0)
     );
   end component;
--   определение структуры накапливающего сумматора
signal X : bit_vector(15 downto 0); begin
  add1 : add
    port map (DI, DO, X); --   определение связности портов
  reg1 : reg  port map (X, DO, CLK);
end structure;
```

Рис. B.4. Функциональное описание накапливающего сумматора на VHDL

ной реализации описывающий функциональные возможности модуля через соотношения ввода-вывода. В качестве примера рассмотрите рис. B.5, где описывается, что выход сумматора является суммой его входов с дополнением до двух.

Используя средства функционального моделирования, вы можете выбирать множество моделей задержек, например, включать описание задержки между входным и выходным сигналами в поведенческое описание модуля. Довольно часто применяется модель с нулевой задержкой, поскольку она дает наибольшую скорость моделирования.

Использование моделей более высокого уровня

При создании основ цифровой системы, например проигрывателя компакт-дисков или встроенного микроконтроллера, разработчик редко мыслит категориями битов. Вместо этого он представляет поток данных по шинам в форме целочисленных слов или слов в формате с плавающей запятой, а сигналы, передаваемые по шинам команд, — в виде конечного набора командных слов (*ACC*, *RD*, *WR*, *CLR* и др.). Очевидно, что моделирование дискретного проекта на этом уровне абстракции более понятно, при этом позволяет

```
entity add is
  port (
    IN0 : in bit_vector(15 downto 0);
    IN1 : in bit_vector(15 downto 0);
    OUT0 : out bit_vector(15 downto 0)
  );
end add;

architecture behavior of add is begin
  process(IN0, IN1)
    variable C : bit_vector(16 downto 0);
    variable S : bit_vector(15 downto 0);
  begin
    loop1:
    for i in 0 to 15 loop
    S(i) := IN0(i) xor IN1(i) xor C(i);
    C(i+1):= IN0(i) и IN1(i) or C(i) и (IN0(i) or IN1(i));
    end loop loop1;
    OUT0 <= S;
  end process;
end behavior;
```

Рис. B.5. Поведенческое описание 16-битового сумматора

существенно выиграть с точки зрения скорости моделирования. Поскольку 64-битовая шина теперь обрабатывается как один объект, анализ ее значения требует всего одного действия, а не 64, как на логическом уровне. Недостаток данного подхода заключается в снижении точности синхронизации событий. Поскольку шина считается одним объектом, ей можно приписать только одну глобальную задержку, хотя на логическом уровне для разных битов задержка элементов шины может быть различной.

Кроме того, принято различать *функциональное* (или *структурное*) и *поведенческое* описания. На функциональном уровне описание отображает предполагаемую структуру аппаратного обеспечения. Описание на поведенческом уровне только имитирует связь между входом и выходом реальной структуры.

Наиболее популярными языками на данном уровне абстракции являются VHDL и VERILOG. VHDL позволяет вводить определенные пользователем типы данных, например, 16-битовые слова с дополнением до двух или перечислимые наборы команд. Многие разработчики предпочитают использовать традиционные подходы и описывают поведенческие модели в первом приближении с помощью C или C++.

Преимуществом описанного подхода является его гибкость, однако он требует, чтобы пользователь определял все типы данных и, по сути, писал полный сценарий моделирования.

```
entity accumulator is
  port (
    DI : in integer;
    DO : inout integer := 0;
    CLK : in bit
  );
end accumulator;

architecture behavior of accumulator is
begin
  process(CLK)
  variable X : integer := 0; --  промежуточная переменная
  begin
    if CLK = '1' then
      X <= DO + D1;
      DO <= X;
    end if;
  end process;
end behavior;
```

Рис. В.6. Накапливающий сумматор из примера В.2

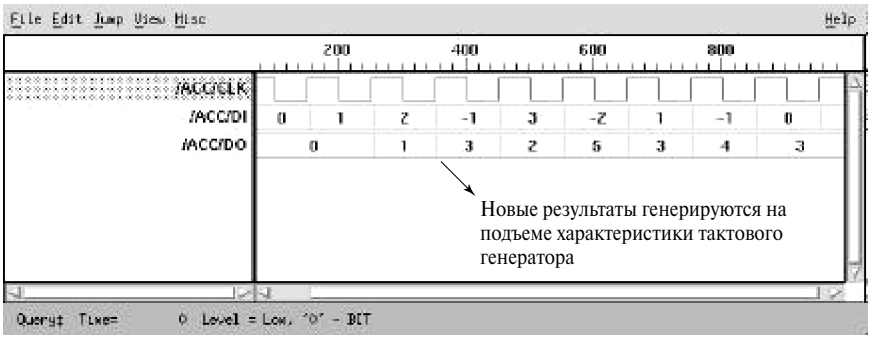


Рис. В.7. Отображение результатов моделирования, полученных на поведенческом уровне. Средство WAVES (и программа моделирования VHDL) входят в набор инструментов Synopsis VHDL. (*Перепечатано с разрешения Synopsys.*)

Пример В.2. VHDL-описание на поведенческом уровне

Чтобы сравнить функциональный и поведенческий режимы описания и использование высокоуровневых моделей данных, вернемся к примеру с накапливающим сумматором (рис. В.6). В данном случае мы используем чисто поведенческое описание, в котором для представления работы модуля применяются целочисленные типы данных.

Результаты моделирования, выполненного на данном уровне абстракции, показаны на рис. В.7. Даже для данного небольшого примера производительность моделирования с точки зрения времени центрального процессора получается в три раза выше, чем было получено с помощью структурного описания, представленного на рис. В.4.

Литература

[Agrawal90] P. Agrawal and W. Dally, "A Hardware Logic Simulation System", *IEEE Trans. Computer-Aided Design*, CAD-9, January 1990, no. 9, p. 19–29.

[Bryant81] R. Bryant, *A Switch-Level Simulation Model for Integrated Logic Circuits*, Ph. D. diss., MIT Laboratory for Computer Science, report MIT/LCS/TR-259, March 1981.

[Salz89] A. Salz and M. Horowitz, "IRSIM: An Incremental MOS Switch-Level Simulator", *Proceedings of the 26th Design Automation Conference*, 1989, p. 173–178.

[SystemC] *Everything You Wanted to Know about SystemC*, `http://www.systemc.org/`

[Thor88] R. Alverson et al., "THOR User's Manual", Technical Report CSL-TR-88-348 and 349, Stanford University, January 1988.

[TimeMill] `http://www.synopsys.com/products/mixedsignal/mixedsignal.html`

[VHDL88] VHDL Standards Committee, IEEE Standard VHDL Language Reference Manual, IEEE standard 1076-1077, 1978.

[White87] J. White and A. Sangiovanni-Vincentelli, *Relaxation Techniques for the Simulation of VLSI Circuits*, Kluwer Academic, 1987.

Вставка

Г

Техники разработки топологии для сложных логических элементов

Разработка топологии согласно Уэйнбергеру и концепции стандартных ячеек

Подход с использованием графов Эйлера

В главе 6 мы подробно обсудили, как создавать схемы сложных логических элементов и выбирать размеры транзисторов. Последним этапом процесса проектирования является разработка топологии для логического элемента или ячейки; другими словами, мы должны определить точную форму различных многоугольников, составляющих топологию элемента. На состав топологии очень сильно влияет *принцип межсоединений*. Как связана ячейка с общей структурой кристалла и как она сообщается с соседними ячейками? Если вы думаете об этих вопросах с самого начала разработки, конечный проект будет иметь меньшую паразитную емкость.

Разработка топологии согласно Уэйнбергеру и концепции стандартных ячеек

Ниже мы рассмотрим два важных подхода к разработке топологии, хотя, в общем-то, их существует огромное множество. Согласно *подходу Уэйнбергера* (Weinberger) [Weinberger67], шины данных (входов и выходов) трассируются (в металле) параллельно шинам питания и перпендикулярно диффузионным областям, как показано на рис. Г.1. Транзисторы формируются на пересечении поликремниевых сигнальных проводников (соединенных с горизонтальными металлическими проводниками) и диффузионных областей.

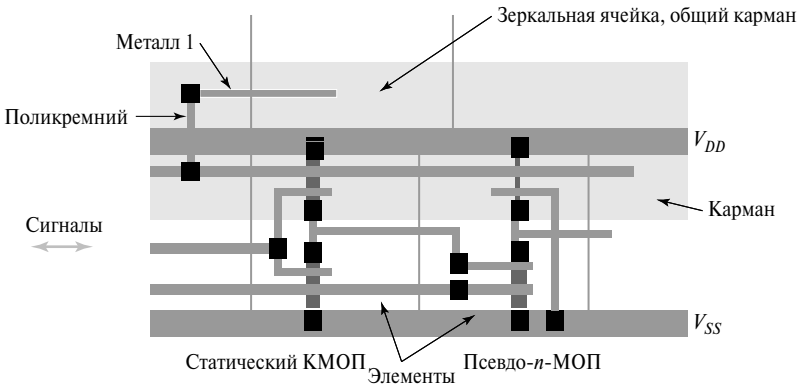


Рис. Г.1. Подход Уэйнбергера к топологии сложных элементов (с использованием единственного металлического слоя)

Использование подхода принципа "над ячейкой" делает технику Уэйнбергера особенно удачной для разрядно-модульных информационных каналов. Хотя данная техника все еще используется в различных проектах, в целом она потеряла свою привлекательность в эпоху стиля стандартных ячеек.

Согласно *концепции стандартных ячеек*, сигналы трассируются в поликремнии перпендикулярно шинам электропитания (рис. Г.2). Обычно данный подход дает плотную топологию для статических КМОП-элементов, поскольку вертикальный поликремниевый провод может служить входом и n-МОП-, и p-МОП-транзисторов. Пример ячейки, реализованной с использованием концепции стандартных ячеек, показан на рис. 6.12. Межсоединения, связывающие ячейки, обычно устанавливаются посредством так называемых трассировочных каналов, как показано на рис. Г.2. Концепция стандартных ячеек весьма популярна и в наши дни из-за высокой степени автоматизации. (Подробное описание средств автоматизации процесса проектирования, поддерживающих концепцию стандартных ячеек, приводится в главе 8.)

Планирование топологии с использованием путей Эйлера

Поскольку данная стратегия планирования топологии распространена довольно широко, нам стоит потратить немного времени и разобраться, как в подобную структуру можно эффективно отобразить сложную булеву функцию. Из-за соображений плотности n-МОП- и p-МОП-транзисторы желательно реализовать в виде неразрывной цепочки устройств, с примыкающими соединениями исток-сток и выравненными соединениями затворов соответствующих n-МОП- и p-МОП-транзисторов. Данный подход требует наличия в обоих карманах единственной диффузионной полоски. Чтобы добиться этого, требуется очень аккуратное упорядочение входных клемм. Сказанное иллюстрируется на рис. Г.3, где показана реализация логической функции $x = (a + b) \cdot c$. В первой версии принят порядок $\{acb\}$. Легко видеть, что решения с единственной диффузионной полоской не существует. Переупорядо-

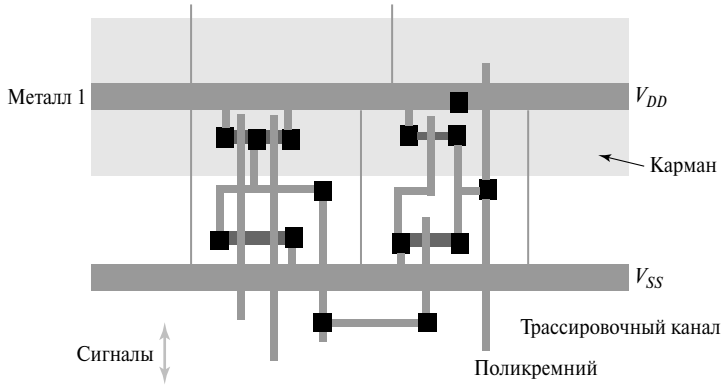


Рис. Г.2. Использование концепции стандартных ячеек для разработки топологии сложных элементов

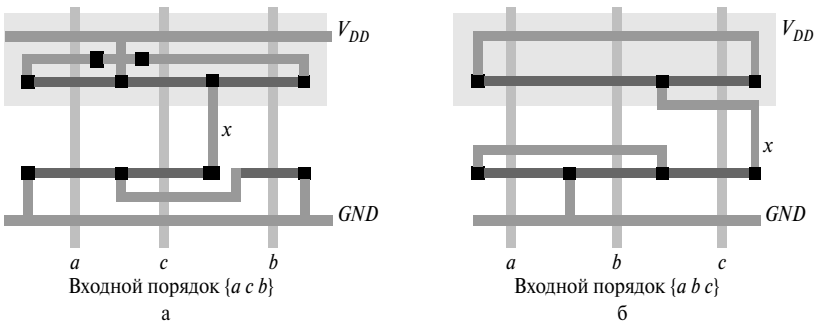


Рис. Г.3. Штриховая диаграмма для $x = (a + b) \cdot c$

чивая клеммы (например, используя порядок $\{abc\}$), мы получаем возможное решение, показанное на рис. Г.3, *б*. Обратите внимание на то, что "топологии" на рис. Г.3 не представляют реальные геометрии фотошаблонов, скорее это символические изображения топологий логических элементов. Проводники и транзисторы представляются как безразмерные объекты, и их расположение является относительным, а не абсолютным. Подобные абстрактные представления называются *штриховыми диаграммами* и часто используются на этапе планирования логического элемента перед определением реальных размеров. В данной книге штриховые диаграммы используются при обсуждении топологий логических элементов или стратегий разработки топологий схем.

К счастью, существует систематический подход, позволяющий так рассчитывать перестановки входных клемм, чтобы сложные функции можно было реализовать с использованием непрерывных диффузионных полосок, минимизирующих площадь [Uehara81]. К достоинствам систематической природы описанной технологии также относится легкость автоматизации. Концептуально ее можно разбить в два этапа.

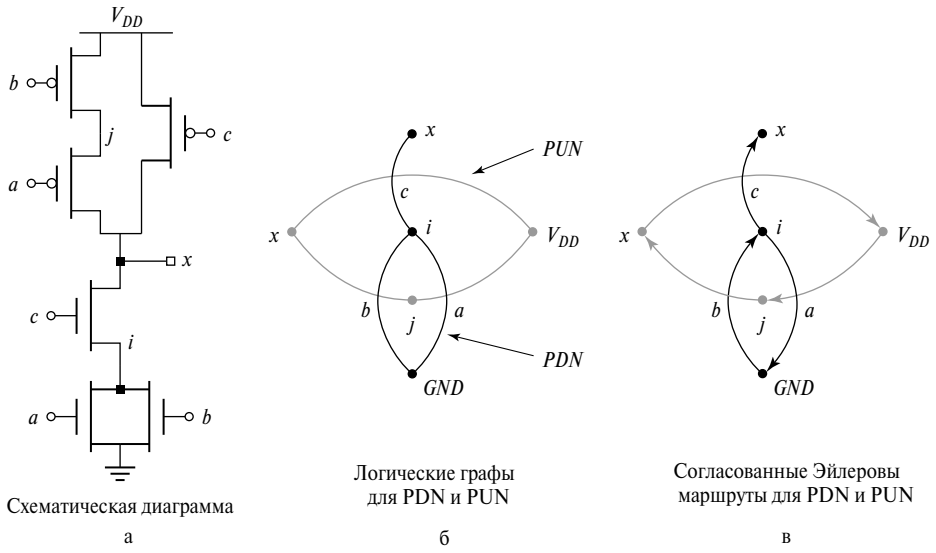
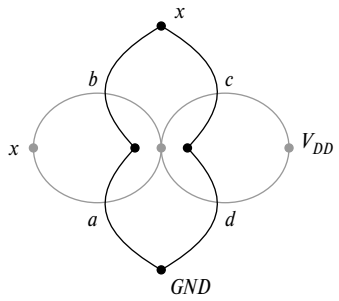


Рис. Г.4. Принципиальная схема, логический граф и пути Эйлера для функции $x = (a + b) \cdot c$

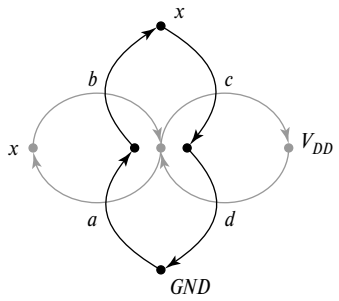
1. **Построение логического графа.** Логический граф транзисторной сети (или переключательной функции) представляет собой граф, вершины которого являются узлами (сигналами) сети, а ребра представляют транзисторы. Каждое ребро обозначает сигнал, управляющий соответствующим транзистором. Поскольку нагрузочная и разгрузочная сети статического КМОП-элемента являются дуальными, дуальными также являются и соответствующие графы — т.е. параллельное соединение заменяется последовательным и наоборот. Сказанное демонстрируется на рис. Г.4, где представлены логические графы нагрузочной и разгрузочное сетей булевой функции $x = (a+b)\cdot c$ (обратите внимание, что данный подход можно использовать для расчета дуальных сетей).
2. **Определение путей Эйлера.** Путь Эйлера на графе определяется как такой путь через все вершины, при котором все они посещаются один раз. Найти такой путь очень важно, поскольку упорядочение входов, позволяющее использовать непрерывную диффузионную полоску n-МОП- (или p-МОП-) транзисторов, возможно только тогда, когда на логическом графе, представляющей цепь PDN (PUN), существует путь Эйлера. Ниже приведены причины поиска такого пути.

Чтобы сформировать непрерывную полоску диффузии, все транзисторы должны проходиться последовательно, т.е. сток одного устройства должен быть истоком другого. Это эквивалентно обходу логического графа вдоль пути Эйлера. Учтите, что пути Эйлера не уникальны: может существовать несколько различных решений.



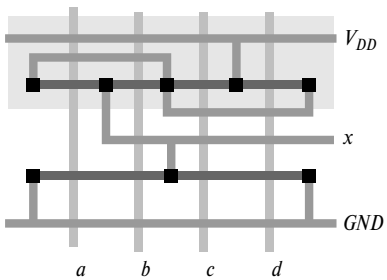
Логические графы для $\overline{(ab + cd)}$

а



Эйлеровы маршруты

б



Штриховая диаграмма для упорядочения {*a b c d*}

в

Рис. Г.5. Расчет топологии для функции $x = (ab + cd)$

Последовательность ребер в пути Эйлера соответствует порядку входов в топологии логического элемента. Чтобы получить одинаковый порядок в нагрузочной и разгрузочной сетях (это необходимо, если требуется использовать одну поликремниевую полоску для всех входных сигналов), пути Эйлера должны быть *непротиворечивыми*, т.е. иметь ту же последовательность.

Непротиворечивые пути Эйлера для схемы, приведенной на рис. Г.4, *а*, показаны на рис. Г.4, *в*. Топология, соответствующая этому решению, представлена на рис. Г.3, *б*. Изучая логическую диаграмму функции, видим, что $\{acb\}$ является путем Эйлера для цепи PUN, но не PDN. Следовательно, решение с одной диффузионной полоской является несуществующим (рис. Г.3, *а*).

Пример Г.1. Расчет топологии сложного логического элемента

Выведем в качестве примера топологию схемы для следующей логической функции:

$$x = \overline{ab + cd}.$$

На рис. Г.5, *а* и *б* показаны логическая функция и один непротиворечивый путь Эйлера. Соответствующая топология представлена на рис. Г.5, *в*.

Следует помнить, что существование непротиворечивых путей Эйлера зависит от способа построения булевых выражений (и соответствующих логических графов). Например, для функции $\overline{x} = a + b \cdot c + d \cdot e$ нельзя найти ни одного непротиворечивого пути Эйлера, однако для функции $\overline{x} = b \cdot c + a + d \cdot e$ имеется единственное решение. Иногда для нахождения набора непротиворечивых путей требуется предварительная реструктуризация функции. Это может быть полный перебор всех существующих вариантов. К счастью, в [Uehara81] был предложен простой алгоритм, позволяющий этого избежать. Впрочем, в данной книге мы его обсуждать не будем, если вас интересует эта тема — обращайтесь к первоисточнику.

Наконец, стоит отметить, что представленные стратегии разработки топологий не являются единственно возможными. Например, иногда эффективнее использовать несколько диффузионных полосок, выстроенных вертикально в ряд. В таком случае одна поликремниевая входная шина может использоваться в качестве входа нескольких транзисторов. Это бывает выгодно при определенных структурах логических элементов (например, ИСКЛЮЧАЮЩЕЕ НЕ-ИЛИ), следовательно, подход рекомендуется выбирать на основе изучения конкретной ситуации.

Для любознательных

Хороший обзор технологий генерации ячеек можно найти в [Rubin87, p. 116–128]. Кроме того, заслуживают внимания указанные ниже работы в данной области.

[Clein00] D. Clein, *CMOS IC Layout*, Newnes, 2000.

[Rubin87] S. Rubin, *Computer Aids for VLSI Design*, Addison-Wesley, 1987.

[Uehara81] T. Uehara and W. Van Cleemput, "Optimal Layout of CMOS Functional Arrays", *IEEE Trans. on Computers*, May 1981, vol. C-30, no. 5, p. 305–311.

[Weinberger67] A. Weinberger, "Large Scale Integration of MOS Complex Logic: A Layout Method", *IEEE Journal of Solid State Circuits*, 1967, vol. 2, no. 4, p. 182–190.

ГЛАВА

7

Разработка последовательных логических схем

Технологии реализации регистров, защелок, триггеров, вибраторов, генераторов импульсов и триггеров Шмитта

Статическая и динамическая реализации

Выбор стратегии синхронизации

7.1. Введение

7.1.1. Временны*е* метрики последовательных схем

7.1.2. Классификация элементов памяти

7.2. Статические защелки и регистры

7.2.1. Принцип бистабильности

7.2.2. Защелки, основанные на мультиплексоре

7.2.3. Регистр “master–slave”, управляемый фронтом

7.2.4. Низковольтные статические защелки

7.2.5. Статические SR-триггеры — запись данных методом чистой силы

7.3. Динамические защелки и регистры

7.3.1. Динамические регистры на проходных логических элементах, управляемые фронтом

7.3.2. Тактируемая КМОП-схема — подход, нечувствительный к расфазировке

7.3.3. Действительно однофазный тактируемый регистр

7.4. Альтернативные стили регистров
7.4.1. Импульсные регистры
7.4.2. Регистры, основанные на усилителях считывания
7.5. Конвейер: подход к оптимизации последовательных цепей
7.5.1. Конвейер на защелках или регистрах?
7.5.2. NORA–КМОП — стиль логики конвейерных структур
7.6. Небистабильные последовательные схемы
7.6.1. Триггер Шмитта
7.6.2. Последовательные схемы с одним устойчивым состоянием
7.6.3. Автоколебательные схемы
7.7. Перспектива: выбор стратегии тактирования
7.8. Резюме
7.9. Для любознательных

7.1. Введение

Как обсуждалось ранее, комбинационные логические схемы обладают следующим свойством: выход логического блока является функцией только *текущих* входных значений, при этом предполагается, что прошло достаточно времени для установления режима логического элемента. В настоящее же время практически все полезные системы требуют запоминания информации о состоянии, т.е. требуется другой класс схем, называемых *последовательными логическими схемами*. В этих схемах выход зависит не только от *текущих* значений на входах, но и от *предшествовавших* входных значений. Другими словами, последовательная схема помнит часть истории системы — или обладает *памятью*.

На рис. 7.1 показана функциональная схема стандартного *конечного автомата* (Finite-State Machine — FSM), состоящего из комбинационной логики и резисторов, которые хранят состояние системы. Система, изображенная на рисунке, принадлежит к классу *синхронных* последовательных систем, в которых все регистры управляются одним глобальным тактовым сигналом. Выходы конечного автомата являются функцией текущих *входов* и *текущего состояния*. *Следующее состояние* определяется на основе *текущего состояния* и текущих *входов* и подается на входы регистров. В фазе нарастания тактового сигнала биты *следующего состояния* копируются на выходы регистров (после некоторой задержки распространения) и начинается новый цикл. Таким образом, регистр игнорирует изменения входных сигналов до начала следующей фазы нарастания тактового сигнала. Вообще, регистры могут *активизироваться положительным фронтом* (когда входные данные копируются во время нарастания тактового сигнала) или *отрицательным фронтом* (когда входные данные копируются при спаде тактового сигнала, что обозначается небольшим кружочком на тактовом входе).

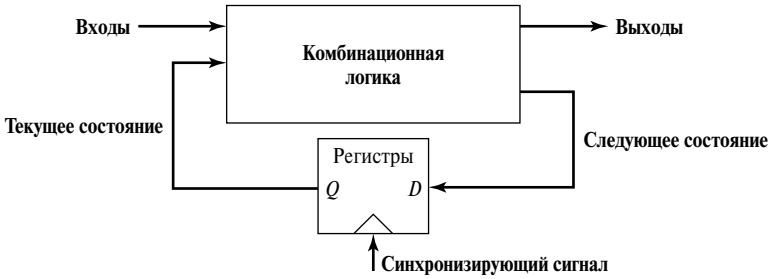


Рис. 7.1. Функциональная схема конечного автомата с регистрами, активизируемыми положительным фронтом тактового сигнала

В данной главе обсуждается КМОП-реализация важнейших последовательных компоновочных блоков. Существует множество последовательных примитивов и методологий синхронизации; так что при разработке современных цифровых схем очень важно выбрать правильные составляющие, поскольку это может оказать существенное влияние на быстродействие, мощность и/или сложность разработки. Прежде чем мы приступим к подробному обсуждению различных вариантов разработки, напомним существенные метрики дизайна и классификацию последовательных элементов.

7.1.1. Временные метрики последовательных схем

Регистр характеризуется тремя важными временны́ми параметрами, показанными на рис. 7.2. *Временем установки* (t_{su}) называется временной интервал, в течение которого входные данные (D) должны оставаться стабильными до тактового перехода (т.е. перехода $0 \rightarrow 1$ для *регистра, активизируемого положительным фронтом*). *Временем удержания* (t_{hold}) называется временной интервал, в течение которого входные данные должны оставаться стабильными после фронта синхроимпульса. Предполагая, что время *установки* и *удержания* выдержано правильно, данные на входе D будут скопированы на выход Q после наихудшей задержки распространения (относительно фронта синхроимпульса) t_{c-q}.

Обладая информацией о временны́х метриках регистров и комбинационных логических блоков, мы можем вывести временны́е условия на уровне системы (см. рис. 7.1). В синхронных последовательных схемах переключения происходят одновременно в ответ на входной тактовый сигнал. Результаты операций ожидают следующих переходов тактового сигнала, после чего передаются на следующий каскад. Другими словами, следующий цикл не может начаться, если не были завершены все текущие вычисления. Следовательно, *такт* (clock period) T, характеризующий работу последовательной схемы, должен перекрывать наиболее длительную задержку для всех каскадов цепи. Предположим, что наибольшая задержка распространения логики равна t_{plogic}, а минимальная — t_{cd}. Минимальный такт T, требуемый для правильной работы последовательной схемы выражается следующим образом:

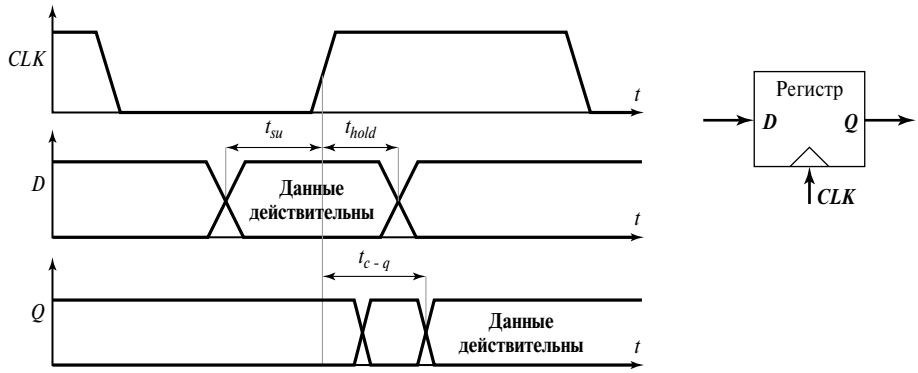


Рис. 7.2. Определение времени установки, времени удержания и задержки распространения синхронного регистра

$$T \geq t_{c-q} + t_{plogic} + t_{su}. \tag{7.1}$$

Время удержания регистра налагает еще одно условие на правильное функционирование схемы:

$$t_{cdregister} + t_{cdlogic} \geq t_{hold}, \tag{7.2}$$

где $t_{cdregister}$ — минимальная *задержка распространения* регистра. Выполнение данного условия гарантирует, что входные данные последовательных элементов будут достаточно долго храниться после фронта синхроимпульса и не будут слишком быстро модифицированы новой порцией данных.

Как видно из уравнения (7.1), значения временны*х* параметров регистра необходимо минимизировать, поскольку они непосредственно влияют на скорость возможной работы последовательной схемы. Фактически современные быстродействующие системы характеризуются очень малой глубиной логики, и *задержка распространения* и *время установки* регистра вносят большой вклад в определение времени такта. Например, максимальная глубина логики микропроцессора DEC Alpha EV6 [Gieseke97] равна 12 логическим элементам, и служебные издержки, связанные с регистром, дают время, составляющее примерно 15% такта. Вообще, удовлетворить требованиям уравнения (7.2) не сложно, хотя в случае, если между регистрами располагается небольшое число логических элементов (или их нет вообще), возможно возникновение определенных проблем[1].

[1] Проблемы также возникают, когда тактовые сигналы различных регистров работают не в фазе из-за расфазировки синхронизирующих импульсов. Подробнее этот вопрос рассмотрен в главе 10.

7.1.2. Классификация элементов памяти

Приоритетная и фоновая память

На высоком уровне память делится на фоновую (background) и приоритетную (foreground). Память, внедренная в логику, называется *фоновой* и довольно часто реализуется в виде отдельных регистров или блоков регистров. Большие массивы централизованных ячеек памяти называются *фоновой памятью*. Фоновая память, рассмотренная в главе 12, позволяет достичь большей плотности за счет эффективного использования матричных структур и небольшого снижения быстродействия и устойчивости в обмен на уменьшение размера. В этой главе мы сосредоточимся на приоритетной памяти.

Статическая и динамическая память

Память может быть статической или динамической. Статическая память сохраняет свое состояние, пока на нее подается питание. Для создания статической памяти применяется *положительная обратная связь* или регенерация, а топология схемы включает намеренное соединение выхода и входах комбинационной схемы. Статическая память наиболее полезна в том случае, когда в течение длительных промежутков времени регистр обновляться не будет. Хорошим примером статических данных являются конфигурационные данные, загружаемые во время включения питания. Данное условие также справедливо для большинства процессоров, использующих условную синхронизацию (т.е. стробированные генераторы тактовых импульсов), когда для неиспользуемых модулей тактовый сигнал отключается. В таком случае нельзя сказать точно, как часто будут переключаться регистры, и для хранения информации о состоянии необходима статическая память. Память, основанная на положительной обратной связи, относится к классу элементов, называемых *мультивибраторами*. Наиболее популярным представителем мультивибраторов является *бистабильная* схема, однако широко применяются и другие элементы, например *моностабильные* и *автоколебательные* схемы.

Динамическая память хранит данные в течение небольшого периода времени, возможно, нескольких миллисекунд. Ее действие основано на принципах временного *накопления заряда* на паразитных конденсаторах, связанных с МОП-устройствами. Точно так же, как и в случае с рассмотренной ранее динамической логикой, конденсаторы необходимо периодически обновлять, чтобы компенсировать утечку заряда. Динамическая память обычно проще статической, что приводит к значительно большему быстродействию и меньшему рассеянию мощности. Такая память наиболее полезна в схемах с информационными каналами, требующими высокого быстродействия и периодической синхронизации. Динамические схемы можно использовать даже при условной синхронизации, если состояние можно отбросить, когда модуль переходит в нерабочий режим.

Защелки и регистры

Защелка (latch) является необходимым компонентом регистра, *инициируемого фронтом*. Это *чувствительная к уровню* схема, передающая D входов на

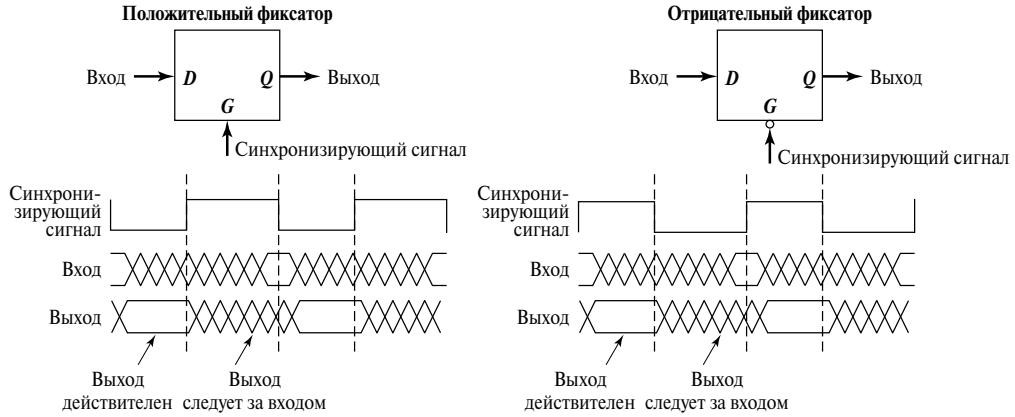


Рис. 7.3. Временны*е* диаграммы для положительной и отрицательной защелок

Q выходов при высоком уровне тактового сигнала. В таких случаях говорится, что данная защелка находится в *прозрачном режиме*. При низком уровне тактового сигнала входные данные, выбранные на заднем фронте тактового сигнала, удерживаются устойчивыми на выходе всей фазы; в таком случае защелка находится в режиме *удержания*. Входы должны быть устойчивыми на протяжении небольшого периода на спаде тактового сигнала, чтобы согласовываться с требованиями, касающимися времени установки и удержания. Защелка, действующая при таких условиях, называется *положительной*. Соответственно *отрицательная защелка* передает D входов на Q выходов при низком уровне тактового сигнала. Сигналы, соответствующие положительной и отрицательной защелкам, показаны на рис. 7.3. В настоящее время существует множество вариантов практической реализации защелок.

В отличие от защелок, *чувствительных к уровню*, регистры, *переключаемые фронтом*, всего лишь выбирают вход в момент, когда тактовый сигнал совершает переход, т.е. при переходе $0 \rightarrow 1$ для регистра, *управляемого положительным фронтом*, и при переходе $1 \rightarrow 0$ — для регистра, *управляемого отрицательным фронтом*. Как правило, такие регистры создаются для использования примитивов защелок, изображенных на рис. 7.3. Например, довольно часто используется конфигурация *ведущий–ведомый* (master–slave), когда положительная и отрицательная защелки соединяются последовательно. Кроме того, регистры можно создавать, используя однократные генераторы тактовых сигналов (“импульсные” регистры) или другие специальные структуры. Примеры таких структур приводятся далее в этой главе.

В литературе, посвященной последовательным схемам, приводится множество определений различных типов элементов памяти (т.е. регистров, триггеров и защелок), часто противоречащих друг другу. Чтобы избежать путаницы, мы приведем строгие определения устройств, которых будем придерживаться в данной книге.

- Элемент памяти, *управляемый фронтом тактового сигнала*, будем называть *регистром*.
- Защелкой будем считать устройство, *чувствительное к уровню*.
- Любой *бистабильный* компонент, сформированный перекрестным соединением элементов, будем именовать *триггером*[2].

7.2. Статические защелки и регистры

7.2.1. Принцип бистабильности

В статической памяти с помощью положительной обратной связи создается *бистабильная схема* — цепь, имеющая два устойчивых состояния, которые представляют 0 и 1. Принцип работы данной цепи показан на рис. 7.4, *а*, где изображены два включенных последовательно инвертора, а также типичная передаточная характеристика подобных схем. Кроме того, изображены передаточные характеристики первого инвертора (т.е. зависимость V_{o1} от V_{i1}) и второго инвертора (зависимость V_{o2} от V_{o1}). Последний график повернут относительно оси, чтобы акцентировать внимание на том, что $V_{i2} = V_{o1}$. Предположим теперь, что выход второго инвертора V_{o2} соединен с входом первого (V_{i1}), что на рис. 7.4, *а* обозначено пунктирными линиями. Как видно из объединенной передаточной характеристики, суммарная схема имеет всего три возможных рабочих точки (A, B и C). Относительно данной схемы можно сделать следующее важное предположение (справедливость которого доказать относительно легко).

Когда коэффициент усиления инвертора в переходной области больше 1, единственными устойчивыми рабочими точками являются A и B; C является метастабильной рабочей точкой.

Предположим, что пара инверторов с перекрестными обратными связями имеет рабочую точку C. Даже небольшое отклонение от этой точки (возможно, вызванное шумом) усиливается и *распространяется* по всей схеме. Это вызвано тем, что коэффициент усиления петли больше 1. Данный эффект демонстрируется на рис. 7.5, *а*. К V_{i1} (в точке C) применяется небольшое отклонение δ. Данное отклонение усиливается инвертором. Увеличенное отклонение подается на второй инвертор и усиливается еще раз. Рабочая точка смещается от C, пока не достигнет одной из рабочих точек A или B. Таким образом, C является неустойчивой рабочей точкой. Любое отклонение (даже минимальное) приводит к удалению рабочей точки от ее исходного смещения. Шансы на то, что пара инверторов с перекрестными обратными связями сместится в точку C и останется в ней, невероятно малы. Рабочие точки, обладающие данным свойством, называются *метастабильными*.

С другой стороны, A и B — это устойчивые рабочие точки, что видно на рис. 7.5, *б*. В данных точках *коэффициент усиления замкнутой цепи имеет*

[2] Регистр, управляемый фронтом тактового импульса, часто также называют триггером. В данном тексте под триггером понимается *только* бистабильный элемент.

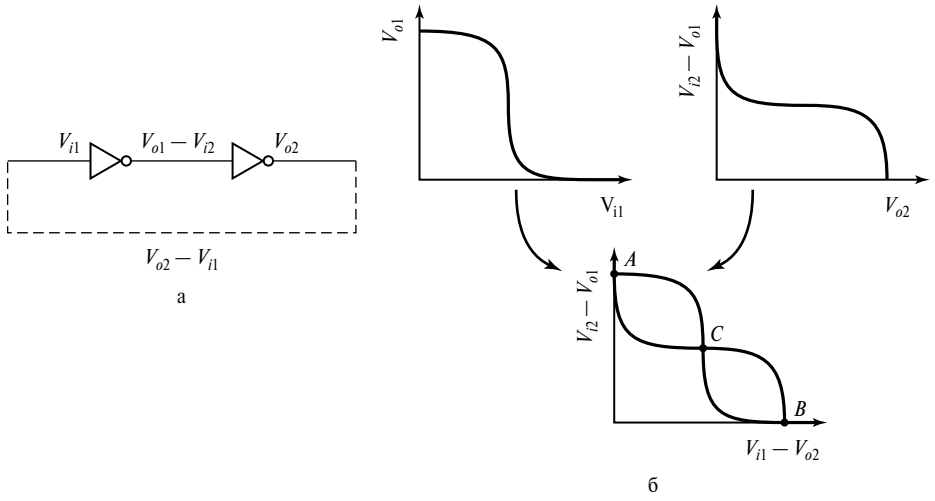


Рис. 7.4. Два инвертора, включенных последовательно (*а*), и их передаточные функции (*б*)

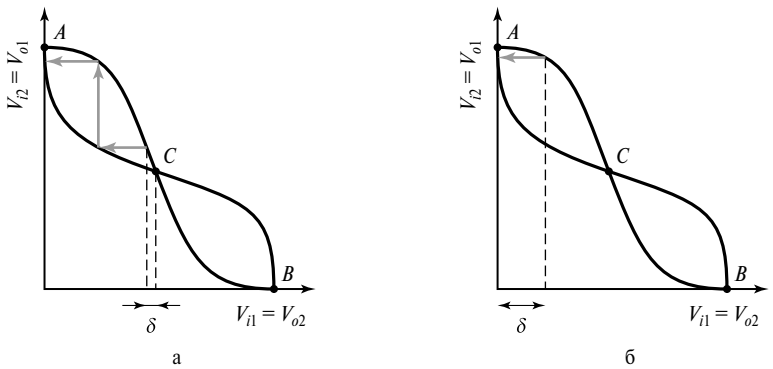


Рис. 7.5. Метастабильная и стабильная рабочие точки

значение, гораздо меньшее единицы. Даже сравнительно большое отклонение от рабочей точки постепенно уменьшается и исчезает.

Таким образом, перекрестное соединение двух инверторов дает *бистабильную* схему, т.е. схему с двумя устойчивыми состояниями, соответствующими двум логическим состояниям. Данная схема используется как ячейка памяти, хранящая значение 1 или 0 (соответствующие позициям A и B).

Чтобы изменить записанное значение, необходимо уметь переводить систему из состояния A в состояние B и наоборот. Поскольку устойчивость подразумевает, что коэффициент усиления G меньшее единицы, мы можем этого добиться, временно сделав позицию A (или B) неустойчивой, увеличив G до значения, превышающего 1. Как правило, для этого на V_{i1} или V_{i2} подается запускающий импульс. Например, предположим, что система нахо-

дится в положении A ($V_{i1} = 0$, $V_{i2} = 1$). Если перевести V_{i1} в состояние 1, оба инвертора некоторое время будут включены одновременно и коэффициент усиления контура G будет больше 1. Положительная обратная связь регенерирует эффект запускающего импульса и схема переходит в другое состояние (в данном случае — B). Ширина запускающего импульса должна быть лишь немногим больше общей задержки распространения контура (которая, в свою очередь, равна удвоенной средней задержке распространения инверторов).

Подытожим, бистабильная схема имеет два устойчивых состояния. При отсутствии запускающего импульса схема остается в одном состоянии (предполагается, что все это время цепь присоединена к источнику питания), а следовательно, помнит заданное значение. Довольно часто бистабильную схему также называют *триггером* (flip-flop). Триггер полезен только тогда, когда существует возможность его перевода из одного состояния в другое. Вообще, для реализации описанных ниже возможностей применяются два различных подхода.

- **Отсечение контура обратной связи**. Как только контур обратной связи размыкается, в Out (или Q) можно легко записать новое значение. В таком случае говорят, что защелка *основана на мультиплексоре*, поскольку логическое выражение синхронной защелки идентично уравнению мультиплексора:

$$Q = \overline{Clk} \cdot Q + Clk \cdot In. \tag{7.3}$$

  Этот подход наиболее популярен в современных защелках, следовательно, в данном разделе мы будем рассматривать преимущественно его.
- **Перегрузить контур обратной связи**. Подавая запускающий импульс на вход триггера, в ячейку памяти принудительно вводится новое значение, перегружающее уже имеющееся в ней. Чтобы это стало возможным, необходимо аккуратно выбрать размер транзисторов в контуре обратной связи и входной схеме. Данный подход был популярен на заре цифровых разработок, однако сейчас постепенно выходит из моды. Тем не менее он является основным при реализации статических фоновых ячеек памяти (которые мы подробнее рассмотрим в главе 12). Краткий обзор памяти этого типа приводится ниже в данной главе.

7.2.2. Защелки, основанные на мультиплексоре

Наиболее устойчивой и распространенной техникой построения защелок является использование мультиплексоров на передаточных вентилях. На рис. 7.6 показана реализация положительной и отрицательной статических защелок, основанных на мультиплексорах. Для отрицательной защелки при низком уровне тактового сигнала на входе мультиплексора выбирается значение 0, и вход D передается на выход. При высоком уровне тактового сигнала на входе мультиплексора выбирается сигнал 1, который соединяется с выходом защелки. Обратная связь гарантирует устойчивый выход до тех пор,

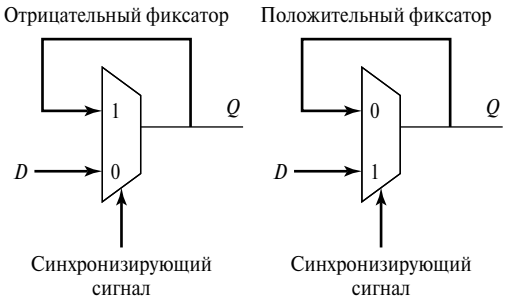


Рис. 7.6. Положительная и отрицательная защелки, основанные на мультиплексорах

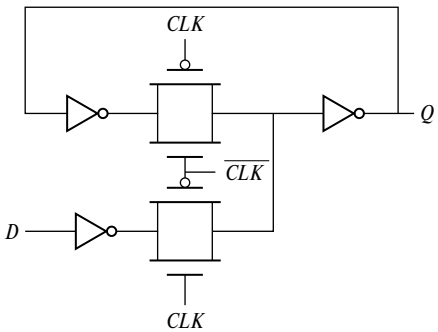
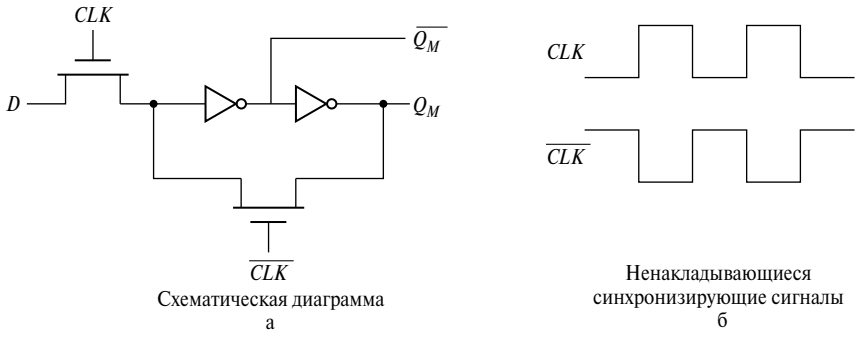


Рис. 7.7. Реализация положительной защелки на уровне транзисторов, созданная с использованием передаточных вентилей

пока тактовый сигнал имеет высокий уровень. Подобным образом в положительной защелке вход D выбирается, когда тактовый сигнал имеет высокий уровень, а при низком уровне тактового сигнала удерживается постоянное значение выхода (с помощью обратной связи).

Реализации положительной защелки, основанной на мультиплексорах, на уровне транзисторов показана на рис. 7.7. При высоком уровне сигнала CLK нижний передаточный вентиль включен и защелка является "прозрачной", т.е. вход D копируется на выход Q. В этой фазе контур обратной связи разомкнут, поскольку верхний передаточный вентиль выключен. Следовательно, правильное функционирование схемы возможно и без тщательного масштабирования транзисторов. Число транзисторов, управляемых тактовым сигналом, является очень важной метрикой с точки зрения мощности, поскольку *коэффициент активности* схемы тактовых импульсов равен 1. С этой точки зрения приведенная в качестве иллюстрации реализации защелки не очень эффективна: она представляет нагрузку в четыре транзистора на сигнал CLK.

Нагрузку на тактовый сигнал можно уменьшить до двух транзисторов, реализовав мультиплексоры, использующие только проходные n-МОП-транзисторы, как показано на рис. 7.8. При высоком уровне сигнала CLK



Схематическая диаграмма
а

Ненакладывающиеся
синхронизирующие сигналы
б

Рис. 7.8. Реализация n-МОП-защелки с использованием в мультиплексорах только проходных n-МОП-транзисторов

защелка выбирает вход D, тогда как низкий уровень тактового сигнала активизирует контур обратной связи и переводит защелку в режим удержания. Использование только проходные n-МОП-транзисторов хотя и является привлекательным с точки зрения простоты, но приводит к передаче более плохого высокого уровня напряжения $V_{DD} - V_{Tn}$ на вход первого инвертора. Это влияет как на запас помехоустойчивости, так и на быстродействие при переключении (особенно при низких значениях V_{DD} и высоких V_{Tn}). Кроме того, это приводит к статическому рассеянию мощности на первом инверторе, поскольку максимальное входное напряжение инвертора равно равно $V_{DD} - V_{Tn}$ и p-МОП-устройство инвертора никогда не выключается полностью.

7.2.3. Регистр “master–slave”, управляемый фронтом

Наиболее распространенным подходом к построению регистра, *управляемого фронтом*, является использование конфигурации *“master–slave”*, как показано на рис. 7.9. Регистр состоит из последовательного соединения отрицательной защелки (главный каскад) и положительной (подчиненный). В данной реализации использованы защелки, основанные на мультиплексорах, хотя можно использовать любую защелку. В фазе низкого тактового сигнала главный каскад является прозрачным, и вход D передается на выход главного каскада Q_M. В этой фазе подчиненный каскад находится в режиме удержания, сохраняя свое предыдущее состояние с помощью обратной связи. На нарастающем фронте тактового сигнала главный каскад прекращает выбирать вход, и выборку начинает подчиненный каскад. В фазе высокого тактового сигнала подчиненный каскад выбирает выход главного (Q_M), тогда как сам главный каскад остается в режиме удержания. Поскольку сигнал Q_M является постоянным в фазе высокого уровня тактового сигнала, выход Q совершает всего один переход за такт. Значение Q равно значению D непосредственно перед нарастающим фронтом тактового сигнала, что дает эффект *управления положительным фронтом*. Регистр, управляемый отрицательным фронтом, можно создать, используя тот же принцип и просто изменив порядок положительной и отрицательной защелок (т.е. поставив первой положительную защелку).

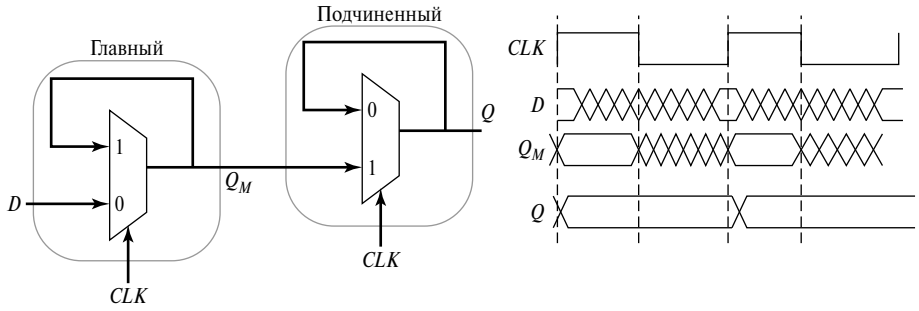


Рис. 7.9. Регистр, управляемый положительным фронтом, основанный на конфигурации “master–slave”

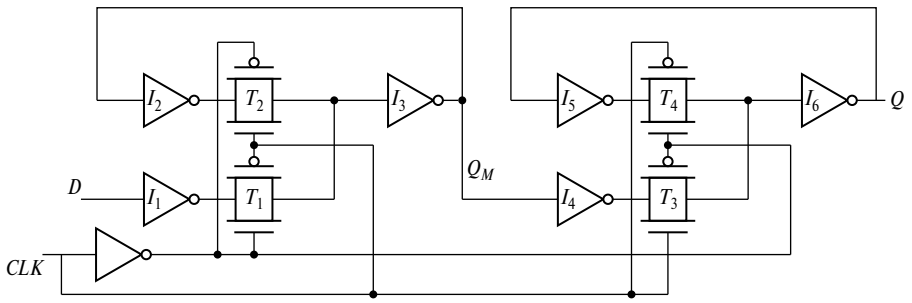


Рис. 7.10. Регистр “master–slave”, управляемый положительным фронтом и созданный с использованием мультиплексоров

Полная реализация регистра “master–slave”, управляемого положительным фронтом тактового сигнала, показана на рис. 7.10. Мультиплексор реализован с использованием передаточных вентилей, как обсуждалось в предыдущем разделе. При низком уровне тактового сигнала ($\overline{CLK} = 1$) транзистор T_1 включен, а T_2 — выключен, и вход D выбирается на узел Q_M. В этой фазе транзистор T_3 выключен, а T_4 — включен. Инверторы с перекрестными обратными связями (I_5, I_6) удерживают состояние подчиненной защелки. Когда уровень тактового сигнала начинает расти, главный каскад перестает выбирать вход и переходит в режим удержания. Транзистор T_1 выключен, T_2 — включен, и инверторы с перекрестными обратными связями I_2 и I_3 удерживают состояние Q_M. Кроме того, транзистор T_3 включен, а T_4 — выключен, и сигнал Q_M копируется на выход Q.

Задача 7.1. Оптимизация регистра “master–slave”

Инверторы I_1 и I_4 (рис. 7.10) можно убрать без ущерба для функциональных возможностей регистра. Дают ли эти инверторы какую-либо пользу?

Временны*е* свойства регистров “master–slave”, основанных на мультиплексорах

Регистры характеризуются тремя важными временны*ми* параметрами: временем установки, временем удержания и задержкой распространения. Для правильной разработки схем необходимо понимать факторы, влияющие на эти временны*е* параметры и тренировать интуицию, которая позволит оценивать их на глаз. Предположим, что задержка распространения каждого инвертора равна t_{pd_inv}, а задержка распространения передаточного вентиля составляет t_{pd_tx}. Предположим также, что задержка загрязнения равна 0 и что инвертор, получающий $\overline{CLK}$ по сигналу CLK, также имеет нулевую задержку.

Временем установки называется время до нарастающего фронта тактового сигнала, в течение которого входные данные D должны иметь приемлемое значение, т.е. имеется в виду, как долго до нарастающего фронта тактового сигнала вход D должен быть стабильным, чтобы сигнал Q_M правильно выбрал его значение. Для регистра на передаточных вентилях, основанного на мультиплексорах, вход D должен пройти через I_1, T_1, I_3 и I_2 до начала нарастающего фронта тактового сигнала. В таком случае напряжение на обоих терминалах передаточного вентиля T_2 будет иметь одинаковое значения. В противном случае существует вероятность того, что пара инверторов I_2 и I_3 с перекрестными обратными связями получит неверное значение. Следовательно, время установки равно: $3 \times t_{pd_inv} + t_{pd_tx}$.

Задержка распространения — это время, требующееся значению Q_M, чтобы дойти до выхода Q. Обратите внимание на то, что поскольку мы включили задержку I_2 во время установки, выход инвертора I_4 станет приемлемым до начала нарастающего фронта тактового сигнала. Следовательно, задержка t_{c-q} представляет собой просто задержку распространения через T_3 и I_6 ($t_{c-q} = t_{pd_tx} + t_{pd_inv}$).

Время удержания представляет собой время, в течение которого вход должен быть стабильным после нарастающего фронта тактового сигнала. В данном случае передаточный вентиль T_1 выключается при изменении уровня тактового сигнала с низкого на высокий. Поскольку и вход D, и сигнал CLK проходят через инверторы перед вентилем T_1, любое изменение входа после перехода тактового сигнала на высокий уровень на выход не влияет. Следовательно, время удержания равно 0.

Пример 7.1. Временной анализ с использованием SPICE

Чтобы найти время установки регистра с помощью SPICE, будем постепенно смещать вход относительно фронта тактового импульса, пока схема не перестанет работать правильно. На рис. 7.11 показано смоделированное время установки при смещении 210 и 200 пс. Для случая 210 пс выбирается правильное значение входа D (в данном случае выход Q сохраняет значение V_{DD}). Для смещения в 200 пс на выход распространяется неправильное значение, поскольку выход Q меняет свое значение на 0. Узел Q_M начинает переход в высокое состояние, и выход инвертора I_2 (вход передаточного

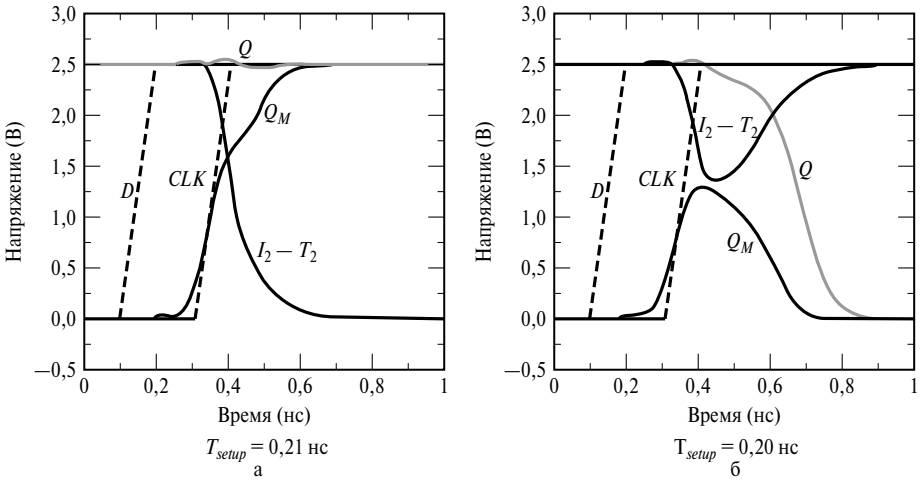


Рис. 7.11. Моделирование времени установки

вентиля T_2) начинает падать. Тем не менее тактовый сигнал активизируется до того, как на двух узлах вдоль передаточного вентиля T_2 устанавливается одинаковое значение. Это приводит к тому, что в главную защелку записывается неправильное значение. Таким образом, время установки для этого регистра равно 210 пс.

Подобным образом можно смоделировать и время удержания. Входной сигнал D снова смещается относительно тактового сигнала, пока схема не прекратит работу. Приведенная схема имеет время удержания 0 (т.е. входы можно менять на фронте тактового сигнала). Наконец, чтобы найти задержку распространения, отметим, что входы меняют свое значение по крайней мере в течение времени, равного времени установки, а задержка измеряется от точки 50% фронта CLK до точки 50% выхода Q. Согласно результатам моделирования (рис. 7.12), $t_{c-q(lh)}$ равно 160 пс, а $t_{c-q(hl)}$ — 180 пс.

Недостатком регистра на передаточных вентилях является высокая емкостная нагрузка тактового сигнала. Нагрузка на тактовый сигнал из расчета на один регистр является важным параметром, поскольку она непосредственно влияет на рассеяние мощности в сети генератора тактовых импульсов. Если не учитывать служебные издержки, требуемые для инверсии тактового сигнала (поскольку служебные издержки инвертора можно распределить на несколько битов регистра), то нагрузка на тактовый сигнал для каждого регистра составляет восемь транзисторов. Уменьшить эту нагрузку можно за счет устойчивости, сделав схему относительной (ratioed). Например, на рис. 7.13 показано, как можно убрать передаточный вентиль обратной связи, напрямую соединив инверторы.

Платой за уменьшение нагрузки на тактовый сигнал является усложнение структуры. Передаточный вентиль (T_1) и его управляющий сигнал должны

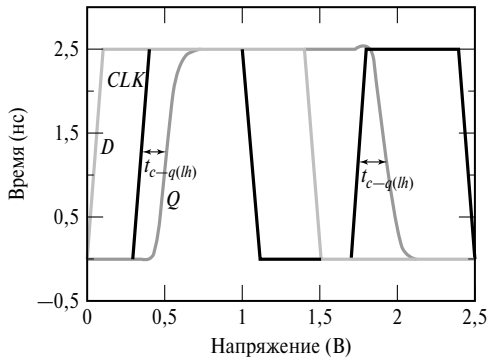


Рис. 7.12. Моделирование задержки распространения регистра на передаточных вентилях

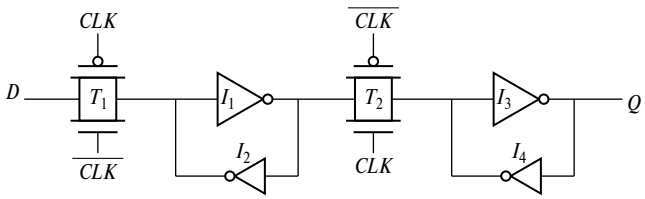


Рис. 7.13. Уменьшение нагрузки на тактовый сигнал статического регистра “master–slave”

пересиливать инвертор обратной связи (I_2), чтобы переключить состояние инвертора перекрестной связи. Требования к размерам передаточных вентилей можно вывести с помощью анализа, подобного использованному в главе 6 при масштабировании схемы восстановления уровня. Для выполнения перехода вход инвертора I_1 необходимо опустить ниже порога переключения этого инвертора. Если в передаточных вентилях необходимо использовать устройства минимального размера, то транзисторы инвертора I_2 должны быть даже еще слабее. Это можно организовать, сделав их длину канала больше минимальной. Вообще, использование в передаточных вентилях устройств минимального (или близкого к минимальному) размера желательно, поскольку это уменьшает рассеяние мощности в защелках и сети распределения тактового сигнала.

Другой проблемой описанной схемы является *обратная проводимость* — второй каскад может влиять на состояние первой защелки. Когда подчиненный каскад включен (рис. 7.14), комбинация блоков T_2 и I_4 может влиять на данные, хранимые в защелке $I_1 - I_2$. К счастью, если I_4 является слабым устройством, это не представляет особой проблемы.

Неидеальные тактовые сигналы

До этого момента мы предполагали, что $\overline{CLK}$ является идеальной инверсией сигнала CLK, или, другими словами, что задержка распространения гене-

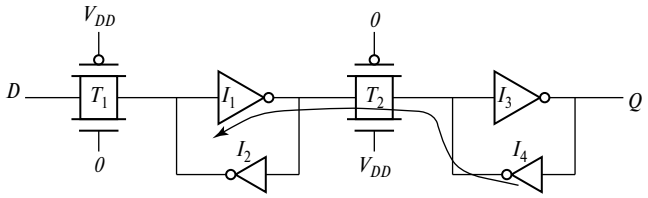


Рис. 7.14. В передаточном вентиле возможна обратная проходимость

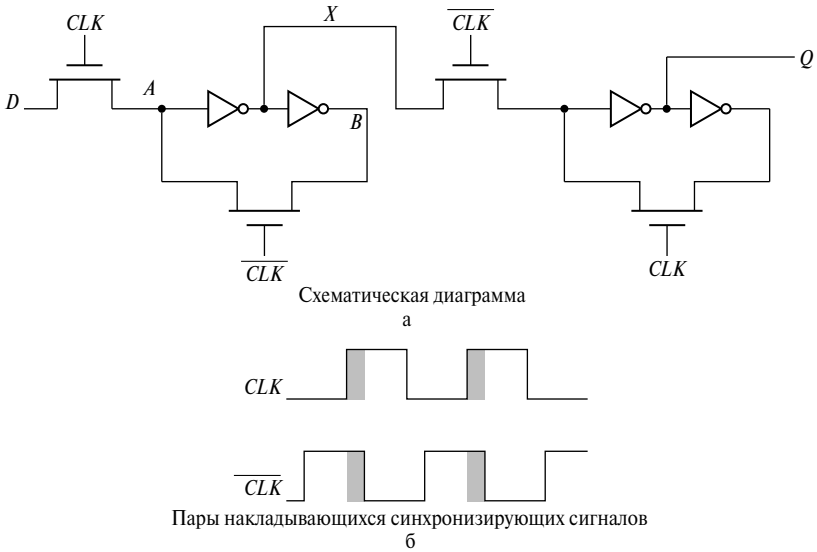
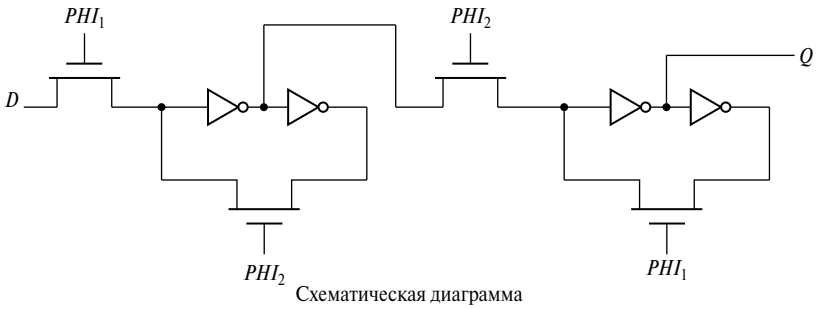


Рис. 7.15. Регистр “master–slave”, основанный только на проходных n-МОП-транзисторах

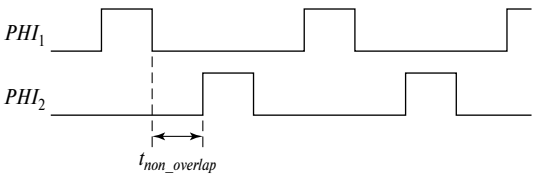
рирующего инвертора равна нулю. Даже если бы это было возможно, такое предположение не совсем оправдано. Проводники, по которым доставляются два тактовых сигнала, могут немного отличаться; кроме того, емкости нагрузки могут быть разными из-за данных, хранимых в соответствующих защелках. Данный эффект, именуемый *расфазировкой тактовых импульсов*, представляет серьезную проблему, приводящую к совмещению двух тактовых сигналов, как показано на рис. 7.15, *б*. Перекрывание тактовых сигналов вызывает сбои двух типов, которые мы проиллюстрируем для отрицательного n-МОП-регистра “master–slave” (см. рис. 7.15, *а*).

1. При повышении уровня тактового сигнала подчиненный каскад должен прекратить выборку выхода главного каскада и перейти в режим удержания. Однако, поскольку некоторое время (*период перекрытия*) оба сигнала CLK и $\overline{CLK}$ имеют высокий уровень, в это время проводят оба выбирающих проходных транзистора, следовательно, имеется прямой тракт между входом D и выходом Q. В результате данные на выходе могут из-



Схематическая диаграмма

а



Двухфазные ненакладывающиеся синхронизирующие сигналы

б

Рис. 7.16. Псевдостатический двухфазный D-регистр

мениться на нарастающем фронте тактового импульса, что нежелательно для регистра, управляемого отрицательным фронтом. Данная ситуация называется *состязанием*, когда значение выхода Q является функцией того, *когда* вход D поступает на узел X — до или после заднего фронта сигнала $\overline{CLK}$. Если значение узла X выбирается в метастабильном состоянии, выход переключится на значение, определяемое шумом, присутствующим в системе.

2. Основным достоинством регистра, основанного на мультиплексорах, является то, что контур обратной связи открыт во время периода выборки, следовательно, масштабирование устройств не имеет критического влияния на функциональные возможности. Тем не менее, если сигналы CLK и $\overline{CLK}$ перекрываются, узел A может управляться как сигналом D, так и сигналом B, что приводит к неопределенному состоянию.

Описанных проблем можно избежать, использовав два неперекрывающихся тактовых сигнала PHI_1 и PHI_2 (рис. 7.16) и выдержав между ними "время неперекрывания", $t_{non_overlap}$, достаточно большое, чтобы даже при наличии задержек в доставке тактовых сигналов перекрытие не происходило. В указанный промежуток времени триггер находится в высокоимпедансном состоянии — контур обратной связи разомкнут, усиление контура равно нулю, и вход отключен от выхода. Если это состояние продлится слишком долго, оно будет разрушено утечкой, поэтому его называют *псевдостатическим*: регистр обладает свойствами статической и динамической памяти в зависимости от состояния тактового сигнала.

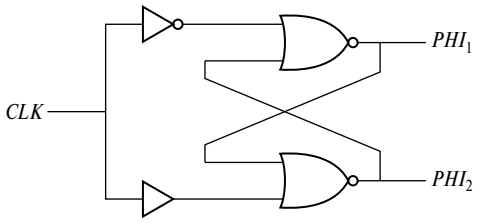


Рис. 7.17. Схема, генерирующая двухфазный неперекрывающийся тактовый сигнал

Задача 7.2. Генерация неперекрывающихся тактовых импульсов

На рис. 7.17 показана одна из возможных реализаций схемы генерации тактовых импульсов, генерирующей двухфазный неперекрывающийся тактовый сигнал. Предполагая, что каждый вентиль характеризуется единичной задержкой, найдите временну*ю* связь между входным тактовым импульсом и двумя выходными тактовыми импульсами. Чему равен период неперекрывания? Как увеличить его при необходимости?

7.2.4. Низковольтные статические защелки

Одним из необходимых требований к работе при низких напряжениях является возможность масштабирования напряжения питания. К сожалению, определенные защелки не функционируют при понижении питания. Например, без масштабирования порогов устройств проходных n-МОП-транзисторы (см., например, рис. 7.16) не масштабируются так же, как напряжение питания, из-за падения внутреннего порога. При очень низких напряжениях питания вход инвертора нельзя увеличить выше порога переключения, что приводит к неправильной оценке. Даже в схемах на передаточных вентилях уменьшение напряжения питания приводит к существенному снижению быстродействия.

Таким образом, масштабирование, направленное на уменьшение напряжения питания, требует использования устройств с более низкими порогами. Однако это связано с экспоненциальным увеличением мощности подпороговой утечки (см. главу 6). Если регистры постоянно находятся в работе, энергия утечки обычно незначительна по сравнению с мощностью переключения. Тем не менее при использовании условных тактовых импульсов регистры могут длительное время находиться в нерабочем режиме, и энергия утечки, расходуемая регистрами может быть значительной.

Существует множество решений проблемы большой утечки в периоды простоя. Один из вариантов — использовать устройства с несколькими порогами, как показано на рис. 7.18 [Mutoh95] (показана только отрицательная защелка). Закрашенные инверторы и передаточные вентили реализованы на устройствах с низкими порогами. Для устранения утечки инверторы с низким порогом управляются устройствами с высоким порогом.

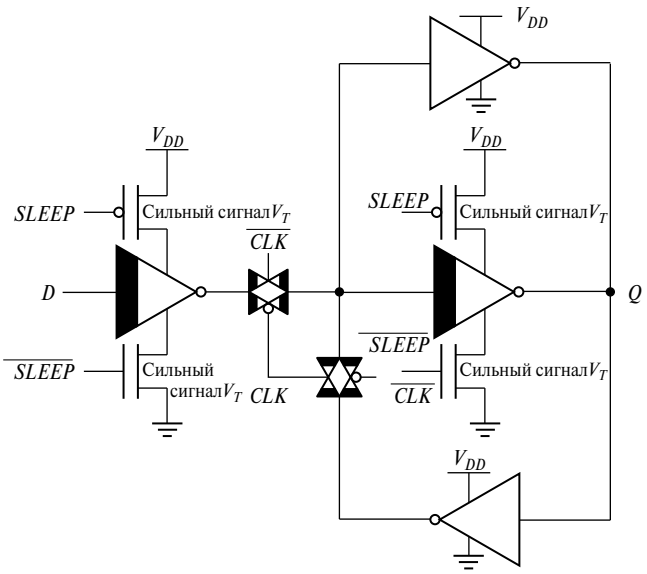


Рис. 7.18. Решение проблемы утечки путем использования КМОП-элементов с несколькими порогами

В нормальном режиме работы спящие устройства включаются. При низком уровне тактового сигнала вход D выбирается и передается на выход. При высоком тактовом сигнале защелка находится в режиме удержания. Передаточный вентиль обратной связи проводит ток, и перекрестная обратная связь активна. Для хранения состояния в то время, когда защелка находится в *холостом* (или *спящем*) режиме, в схему вводится дополнительный инвертор (параллельно инвертору с низким порогом). Таким образом, высокопороговые устройства, соединенные последовательно с низкопороговым инвертором, выключаются (сигнал $SLEEP$ является высоким), устраняя утечку. При этом предполагается, что во время, когда защелка находится в спящем состоянии, тактовый сигнал удерживается на высоком уровне. В это время низкопороговый передаточный вентиль обратной связи включен и высокопороговые устройства с перекрестными обратными связями поддерживают состояние защелки.

Задача 7.3. Минимизация числа транзисторов в КМОП-регистре с несколькими порогами

В отличие от комбинационной логики для устранения утечки в низкопороговых защелках требуются и n-МОП, и p-МОП-устройства. Объясните, почему это так. *Подсказка*: уберите из правого низкопорогового инвертора, изображенного на рис. 7.18, n-МОП- или p-МОП-транзистор с высоким значением V_T и изучите потенциальные пути утечки.

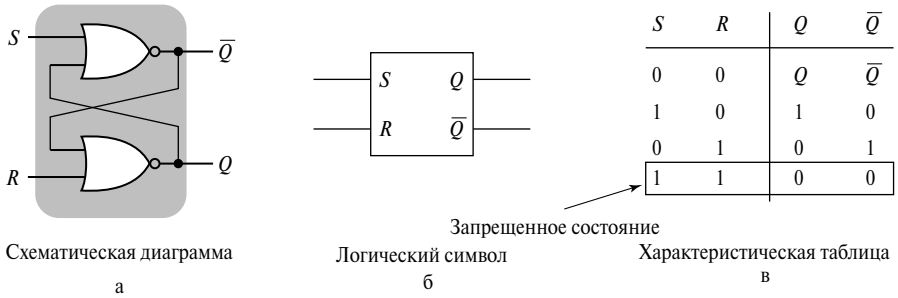


Рис. 7.19. SR-триггер на основе схемы НЕ-ИЛИ

7.2.5. Статические SR-триггеры — запись данных методом грубой силы

Традиционный способ заставить бистабильный элемент изменить состояние — перегрузить контур обратной связи. Простейшей реализацией этой идеи является *SR-триггер* (от "set-reset" — установка-сброс), показанный на рис. 7.19, *а*. Данная схема похожа на схему из пары инверторов с перекрестными обратными связями, в которой вместо инверторов использованы вентили НЕ-ИЛИ. Второй вход вентилей НЕ-ИЛИ соединяется со входами триггеров (S и R), что позволяет принудительно перевести выходы Q и $\bar{Q}$ в нужное состояние. Данные выходы являются комплементарными (исключение составляет только случай $SR = 11$). Когда оба сигнала S и R равны 0, триггер находится в состоянии покоя и оба выхода сохраняют свои значения. (Вентиль НЕ-ИЛИ один из входов которого равен 0, выглядит как инвертор, а вся структура в целом похожа на инвертор с перекрестными обратными связями.) Если на вход S подан положительный импульс (или сигнал 1), выход Q переводится в состояние 1 ($\bar{Q}$, соответственно, — в 0) и наоборот: подача на вход R импульса 1 сбрасывает триггер, и выход Q становится равным 0.

Указанные результаты обычно сводят в *характеристическую таблицу* триггера, показанную на рис. 7.19, *в*. По сути, это таблица истинности вентиля, в которой перечислены выходные состояния как функции всех возможных входных условий. Когда оба сигнала S и R имеют высокий уровень, оба выхода Q и $\bar{Q}$ переводятся в 0. Поскольку это не соответствует нашему представлению о том, что Q и $\bar{Q}$ — это комплементарные сигналы, такой входной режим считается запрещенным. К тому же он представляет еще одну проблему: когда входные триггеры возвращаются к нулевым уровням, получающееся состояние защелки является непредсказуемым и зависит от того, какой вход перешел в низкое состояние последним. Наконец, на рис. 7.19, *б* показано схематическое условное обозначение SR-триггера.

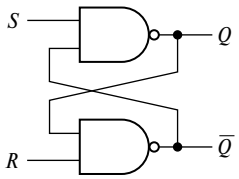


Рис. 7.20. *SR*-триггер на основе схемы НЕ-И

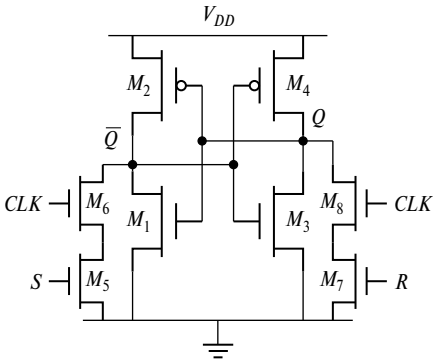


Рис. 7.21. Стандартизованная КМОП-SR-защелка

Задача 7.4. *SR*-триггер на основе вентилей НЕ-И

SR-триггер также можно реализовать с использованием структуры НЕ-И с перекрестными обратными связями, как показано на рис. 7.20. Найдите таблицу истинности для данной реализации.

Рассмотренный выше *SR*-триггер является асинхронным, что противоречит синхронной методологии, согласно которой разрабатывается более 99% современных интегральных схем. Поэтому на рис. 7.21 показана синхронизированная версия защелки. Она состоит из пары инверторов с перекрестными обратными связями, а также из четырех дополнительных транзисторов, переводящих триггер из одного состояния в другое и обеспечивающих синхронизацию. В стационарном состоянии один инвертор находится в высоком состоянии, а другой — в низком. Статического тракта между V_{DD} и землей не существует. Тем не менее, чтобы гарантировать, что триггер, когда это потребуется, может переходить из одного состояния в другое, необходим правильный выбор размеров транзисторов.

Пример 7.2. Масштабирование транзисторов синхронизируемой *SR*-защелки

Рассмотрим ситуацию, когда сигнал Q высокий, и на вход подается импульс R. Чтобы переключить защелку, мы должны опустить Q ниже порога переключения инвертора $M_1 - M_2$. После этого положительная обрат-

ная связь вынудит триггер инвертировать состояния. Данное требование заставляет нас увеличить размеры транзисторов M_5, M_6, M_7 и M_8. Комбинация транзисторов M_4, M_7 и M_8 формирует стандартизованный инвертор. Предположим, что пара инверторов с перекрестными обратными связями спроектирована так, что порог инвертора V_M равен $V_{DD}/2$. При использовании 0,25-микронной КМОП-технологии были выбраны следующие размеры транзисторов: $(W/L)_{M_1} = (W/L)_{M_3}$ =(0,5 мкм/0,25 мкм) и $(W/L)_{M_2} = (W/L)_{M_4}$ =(1,5 мкм/0,25 мкм). Предполагая, что $Q = 0$, мы определяем минимальные размеры транзисторов M_5, M_6, M_7 и M_8, при которых устройство способно переключаться.

Чтобы переключить защелку из состояния $Q = 0$ в состояние $Q = 1$, необходимо, чтобы низкий уровень стандартизованного псевдо-n-МОП-инвертора $(M_5 - M_6) - M_2$ был ниже порога переключения инвертора $M_3 - M_4$, который равен $V_{DD}/2$. Здесь разумно предположить, что при $V_{\bar{Q}} > V_M$, V_Q равно 0 и затвор транзистора M_2 заземлен. Чтобы найти граничное условие на размеры транзисторов, можно приравнять силу тока в инверторе для $V_{\bar{Q}} = V_{DD}/2$, как показано в уравнении (7.4) (модуляция длины канала здесь игнорируется). Сила тока определяется через ток насыщения, поскольку $V_S = V_{DD} = 2,5$ В и $V_M = 1,25$ В. Мы предполагаем, что M_5 и M_6 имеют одинаковые размеры, а эффективное отношение размеров этих последовательно включенных устройств равно W/L_{5-6}. В таком случае разгрузочную сеть можно смоделировать одним транзистором M_{5-6}, длина которого равна удвоенной длине отдельных устройств:

$$\begin{aligned} k'_n \left(\frac{W}{L}\right)_{5-6} \left((V_{DD} - V_{Tn})V_{DSATn} - \frac{V^2_{DSATn}}{2}\right) = \\ = -k'_p \left(\frac{W}{L}\right)_2 \left((-V_{DD} - V_{Tp})V_{DSATp} - \frac{V^2_{DSATp}}{2}\right). \end{aligned} \tag{7.4}$$

Если использовать параметры для процесса 0,25-мкм, уравнение (7.4) даст следующее условие: эффективное отношение $(W/L)_{M_{5-6}} \geq 2,26$. Отсюда следует, что отношение размеров отдельных устройств M_5 или M_6 должно быть больше (приблизительно) 4,5. На рис. 7.22, *а* показан график напряжения постоянного тока $V_{\bar{Q}}$ как функции размеров отдельных устройств M_5 и M_6. Обратите внимание на то, что если отношение размеров отдельных устройств больше 3, этого достаточно для перевода напряжения $\bar{Q}$ через порог переключения инвертора. Разница между ручным анализом и моделированием объясняется эффектами второго порядка — модуляцией длины канала и DIBL. На рис. 7.22, *б* изображена переходная характеристика для различных размеров устройств, из которой видно, что для перегрузки обратной связи и переключения состояния защелки требуется отношение W/L, более 3.

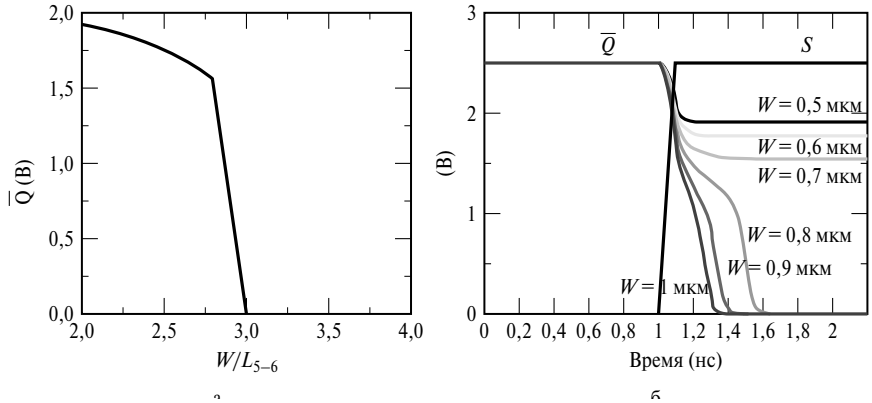


Рис. 7.22. Масштабирование SR-триггера: *а* — зависимость выходного напряжения постоянного тока от размеров разгрузочного устройства M_{5-6} (W/L_2 = 1,5/0,25 мкм); *б* — переходная характеристика, из которой видно, что для переключения SR-триггера отношение W/L транзисторов M_5 и M_6 должно быть больше 3

7.3. Динамические защелки и регистры

Память в статической последовательной схеме основана на том, что пара инверторов с перекрестными обратными связями дает бистабильный элемент, а следовательно, ее можно использовать для запоминания двоичных значений. Данный подход полезен тем, что записанное значение хранится до тех пор, пока на элемент подается напряжение (отсюда и название — *статическая* память). Тем не менее основным недостатком статического элемента является его сложность. Когда регистры используются в постоянно переключающихся вычислительных структурах (например, в конвейерном информационном канале), требование хранения состояние на протяжении длительного периода времени можно сильно ослабить.

В результате мы получаем класс схем, основанных на временном хранении заряда паразитными конденсаторами. Принцип работы таких схем идентичен принципу, использованному в динамической логике, — заряд, хранимый конденсатором, можно использовать для представления логического сигнала. Отсутствие заряда обозначает 0, а его присутствие — 1. К сожалению, идеальных конденсаторов не существует, поэтому всегда присутствует некоторая утечка заряда. Следовательно, удерживать хранимое значение можно только на протяжении ограниченного промежутка времени (обычно это несколько миллисекунд). Если требуется обеспечить целостность сигнала, необходимо периодическое *обновление* значения. В результате мы получаем класс устройств, называемых *динамической* памятью. Отметим также, что считывание с конденсатора значения записанного сигнала, при котором заряд не разрушается, требует наличия устройства с высоким входным сопротивлением.

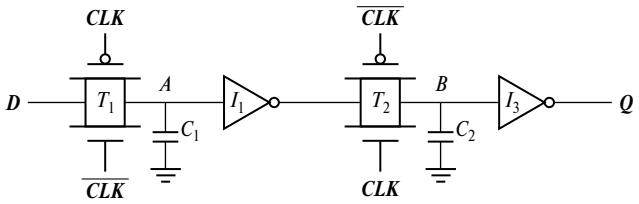


Рис. 7.23. Динамический регистр, управляемый фронтом

7.3.1. Динамические регистры на проходных логических элементах, управляемые фронтом

На рис. 7.23 показан полностью динамический регистр, управляемый положительным фронтом, основанный на концепции "master–slave". При $CLK = 0$ входные данные выбираются в запоминающий узел 1, эквивалентная емкость которого равна C_1 (образована емкостью логического элемента I_1, емкости перехода T_1 и перекрывающейся емкости логического элемента T_1). В это время вспомогательный каскад находится в режиме удержания, а узел 2 — в высокоимпедансном (плавающем) состоянии. На нарастающем фронте тактового сигнала передаточный вентиль T_2 включается, и значение, выбираемое на узле 1 непосредственно перед нарастающим фронтом, распространяется на выход Q (обратите внимание на то, что узел 1 является устойчивым во время высокой фазы тактового сигнала, поскольку первый передаточный элемент выключен). Теперь узел 2 хранит инвертированную версию сигнала в узле 1. Отметим, что такая реализация регистра, управляемого фронтом, очень эффективна, поскольку требует всего восьми транзисторов. Дискретизаторы можно реализовать с использованием проходных n-МОП-транзисторов, что даст даже еще более простую реализацию на шести транзисторах. Разумеется, с точки зрения систем, требующих низкой мощности и высокого быстродействия, уменьшение числа транзисторов очень привлекательно.

Время установки этой схемы равно задержке передаточного вентиля; оно соответствует времени, требуемому узлу 1 для выборки входа D. Время удержания приблизительно равно нулю, поскольку передаточный вентиль выключен на фронте тактового импульса и дальнейшие изменения входов игнорируются. Задержка распространения (t_{c-q}) равна двум задержкам инвертора плюс задержка передаточного вентиля T_2.

Относительно подобного динамического регистра следует сделать одно важное замечание: запоминающие узлы (т.е. состояние) необходимо обновлять через определенные интервалы, чтобы предотвратить потери, вызванные утечкой заряда, утечкой диода или подпороговыми токами. В схемах информационных каналов частота обновления не имеет особого значения, поскольку регистры периодически включаются/выключаются и запоминающие узлы постоянно обновляются.

Важным параметром при анализе схем описанного типа является перекрывание тактовых импульсов. Рассмотрим тактовые сигналы, показанные на рис. 7.24. Во время перекрывания (0,0) n-МОП-транзистор элемента T_1 и

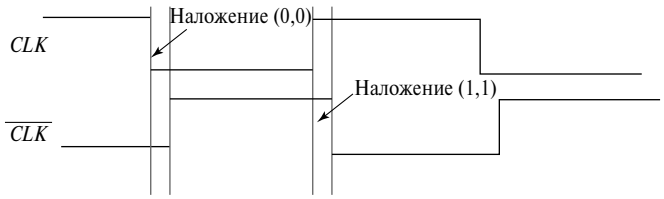


Рис. 7.24. Перекрывание двух тактовых сигналов

p-МОП-транзистор элемента T_2 включены одновременно, что создает для данных прямой путь между входом D регистра и его выходом Q. Другими словами, происходит *состязание*. Если период большой, выход Q может измениться на отрицательном фронте, что, очевидно, нежелательно для регистра, управляемого положительным фронтом. То же справедливо и для перекрывания (1,1), когда p-МОП-транзистор элемента T_1 и n-МОП-транзистор элемента T_2 создают прямой путь от входа к выходу. Чтобы избежать подобных проблем, вводится условие на время *удержания*, т.е. данные должны быть устойчивыми во время перекрывания (1,1). Проблема с перекрыванием (0,0) решается с помощью достаточной задержки между входом D и узлом B, гарантирующей, что новые данные, выбираемые основным каскадом, не распространяются на вспомогательный каскад. В общем случае должно быть достаточно собственной задержки одного инвертора. Таким образом, условие на период перекрывания выражается следующим образом:

$$t_{overlap0-0} < t_{T1} + t_{I1} + t_{T2}. \tag{7.5}$$

Подобным образом, условие для перекрывания (1,1) записывается так:

$$t_{hold} > t_{overlap1-1}. \tag{7.6}$$

Предупреждение

Динамические схемы, описанные в данном разделе, очень привлекательны с точки зрения сложности, быстродействия и мощности. К сожалению, их применимость ограничена из-за соображений устойчивости. В полностью динамической схеме, подобной показанной на рис. 7.23, сеть, имеющая емкостную связь с внутренним запоминающим узлом, может стать источником сильных помех и разрушить состояние. Особенно это важно в потоках ASIC, где связь между сигнальными сетями и внутренними динамическими узлами контролируется очень слабо. Другую проблему порождают токи утечки: большинство современных процессоров требует, чтобы "часы" можно было замедлять или совсем останавливать (чтобы экономить мощность в периоды низкой активности). Наконец, внутренние динамические узлы не отслеживают изменения напряжения источника питания. Например, когда в схеме, изображенной на рис. 7.23, сигнал CLK имеет высокий уровень, узел A сохраняет

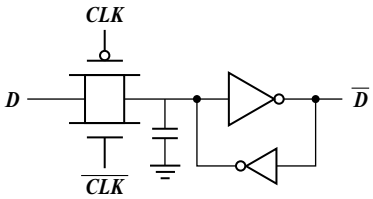


Рис. 7.25. Как сделать из динамической защелки псевдостатическую

свое состояние, однако не видит изменений источника питания, наблюдаемых узлом I_1. В результате мы получаем меньший запас помехоустойчивости.

Большинство описанных проблем можно решить, добавив инвертор со слабой обратной связью и сделав схему *псевдостатической* (рис. 7.25). Хотя такое решение немного увеличивает задержку, оно существенно увеличивает запас помехоустойчивости. Если регистры не планируется использовать в среде, где контролируется буквально каждый аспект быстродействия (например, в быстродействующем заказном информационном канале), их следует делать псевдостатическими или статическими. Сказанное справедливо для всех защелок и регистров, рассмотренных в данном разделе.

7.3.2. Тактируемая КМОП-схема — подход, нечувствительный к расфазировке

Регистр ТКМОП

На рис. 7.26 показан оригинальный регистр, управляемый положительным фронтом, основанный на концепции “master–slave” и нечувствительный к перекрыванию тактовых сигналов. В литературе данная схема называется *регистром* C^2MOS (Clocked CMOS — тактируемый КМОП-регистр) [Suzuki73], работает она в двух фазах.

1. $CLK = 0$ ($\overline{CLK} = 1$). Первый элемент с тремя состояниями включен, и главный каскад действует как инвертор, выбирающий инвертированную версию сигнала D на внутреннем узле X. Главный каскад находится в режиме оценки. В это время дополнительный каскад находится в низкоимпедансном режиме или в режиме удержания. Оба транзистора, M_7 и M_8, выключены и разъединяют выход и вход. Выход Q сохраняет свое предыдущее значение, записанное в выходном конденсаторе C_{L2}.
2. При $CLK = 1$ роли каскадов меняются. Главный каскад находится в режиме удержания (блок M_3–M_4 выключен), а второй каскад оценивает значение (блок M_7–M_8 включен). Значение, хранимое в C_{L1}, распространяется до выходного узла через вспомогательный каскад, который действует как инвертор.

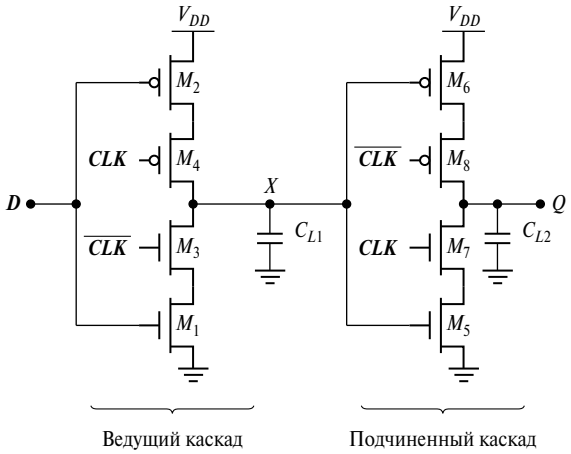


Рис. 7.26. Регистр "master–slave" C^2MOS, управляемый положительным фронтом

В общем, схема действует как регистр "master–slave", управляемый положительным фронтом, — очень похоже на представленный выше регистр, основанный на проходных элементах. При этом, правда, имеется одно существенное отличие.

Регистр C^2MOS с тактированием $CLK - \overline{CLK}$ нечувствителен к перекрыванию, если время нарастания и спада фронтов тактовых импульсов достаточно малы.

Чтобы доказать приведенное утверждение, рассмотрим случаи перекрывания (0-0) и (1-1) (см. рис. 7.24). В случае перекрывания (0-0) схема упрощается до сети, показанной на рис. 7.27, *а*, где в этой фазе *включены* оба p-МОП-устройства. Для правильной работы никакие из новых выбранных данных не должны во время перекрывания успеть дойти до выхода Q, поскольку на отрицательном фронте регистра, управляемого положительным фронтом, не должны происходить изменения данных. Действительно, новые данные выбираются на узле X через последовательность p-МОП-устройств M_2–M_4, и узел X может совершить во время перекрывания переход $0 \rightarrow 1$. Тем не менее эти данные не могут дойти до выхода, поскольку n-МОП-устройство M_7 выключено. В конце периода перекрывания $\overline{CLK} = 1$, и оба устройства, M_7 и M_8, выключены, т.е. вспомогательный каскад переводится в режим удержания. Следовательно, любые новые данные, выбранные на отрицательном фронте тактового сигнала не видны на выходе вспомогательного каскада Q, поскольку вспомогательный каскад выключен до следующего положительного фронта тактового сигнала. Поскольку цепь состоит из последовательного соединения инверторов, для распространения сигнала тре-

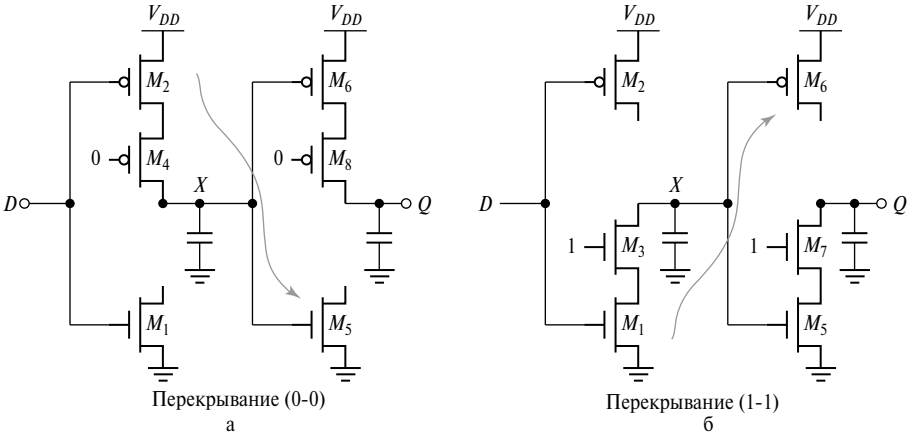


Рис. 7.27. C^2MOS-триггер в периоды перекрывания. Стрелками показано, что надежного тракта передачи сигнала между In и D не существует

буется одна нагрузочная сеть, за которой следует одна разгрузочная сеть, или наоборот, что в описанной ситуации нереально.

Более спорную ситуацию имеем для перекрывания (1-1), когда включены оба n-МОП-устройства, M_3 и M_7 (см. рис. 7.27, *б*). Нас интересует все тот же вопрос: могут ли данные, выбранные во время перекрывания (сразу после того, как тактовый сигнал вырос), попасть на выход Q. Регистр, управляемый положительным фронтом, может передать на выход только данные, присутствовавшие на входе перед началом положительного фронта. Если вход D изменится во время перекрывания, узел X может совершить переход $1 \rightarrow 0$, но дальше это изменение не пойдет. Однако после завершения периода перекрытия p-МОП-устройство M_8 включается, и 0 попадает на выход, что, разумеется, нежелательно. Проблема решается наложением на выходные данные D условия на время удержания; другими словами, данные D должны быть стабильными во время перекрывания.

Подведем итоги: C^2MOS-защелка нечувствительна к перекрыванию тактовых сигналов, поскольку эти перекрывания активизируют либо нагрузочную, либо разгрузочную сети защелки, но не обе одновременно. Если же процессы нарастания и спада тактового сигнала идут достаточно медленно, существует отрезок времени, когда одновременно проводят и n-МОП-, и p-МОП-транзисторы. В результате создается проводящий тракт между входом и выходом, что может разрушить состояние схемы. Результаты моделирования показывают, что схема работает правильно, пока время нарастания (или спада) тактового сигнала меньше примерно пятикратного значения задержки распространения регистра. Данный критерий не очень строг, и на практике ему достаточно просто удовлетворить. Влияние времени нарастания и спада иллюстрируется на рис. 7.28, где показана смоделированная переходная характеристика C^2MOS-D-триггера для времени нарастания/спада, равного соответственно 0,1 и 3 нс. Видно, что для медленных тактовых сигналов существует возможность возникновения *состязания.*

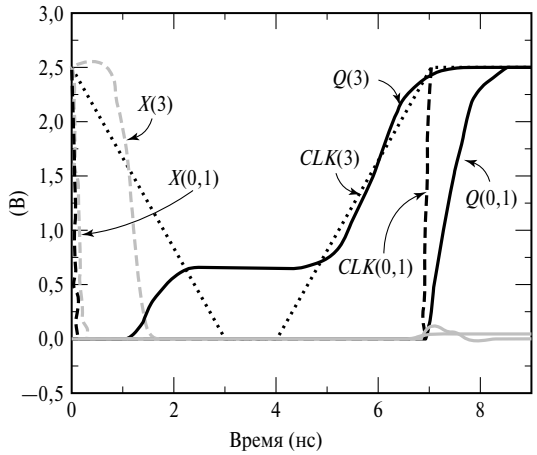


Рис. 7.28. Переходная характеристика C^2MOS-триггера для времен нарастания/спада тактовых сигналов, равных 0,1 и 3 нс (предполагается, что $In = 1$)

Двухфронтовые регистры

До этого момента мы акцентировали внимание на регистрах, управляемых фронтом тактового импульса, которые выбирают входные данные только на одном фронте тактового сигнала (положительном или отрицательном). В принципе можно создать последовательную схему, которая будет выбирать вход на обоих фронтах. Преимущество такой схемы состоит в том, что при той же функциональной пропускной способности используется меньшая тактовая частота (половина оригинальной), что позволяет сэкономить мощность в сети распределения тактового сигнала. На рис. 7.29 показана модификация C^2MOS-регистра, допускающая выборку на обоих фронтах. Эта схема состоит из двух параллельных управляемых фронтом регистров "master–slave", выходы которых уплотняются с помощью элементов с тремя состояниями.

При высоком уровне тактового сигнала положительная защелка, составленная из транзисторов M_1–M_4, выбирает инвертированный вход D на узле X. Узел Y остается в устойчивом состоянии, поскольку устройства M_9 и M_10 выключены. На отрицательном фронте тактового сигнала верхняя вспомогательная защелка M_5–M_8 включена и направляет инвертированное значение X на выход Q. В низкой фазе нижняя главная защелка (M_1, M_4, M_9, M_{10}) включена и выбирает инвертированный вход D на узле Y. Обратите внимание на то, что устройства M_1 и M_4 используются повторно, что уменьшает нагрузку на вход D. На положительном фронте нижняя вспомогательная защелка проводит сигнал и передает инвертированную версию Y на узел Q. Таким образом, данные меняются на обоих фронтах. Обратите внимание на то, что вспомогательные защелки действуют комплементарно, т.е. только одна из них включена в течение каждой фазы тактового сигнала.

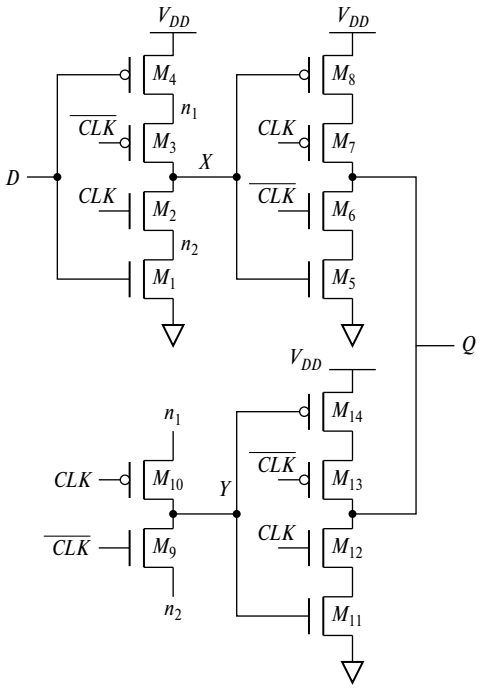


Рис. 7.29. Регистр на основе C^2MOS, управляемый обоими фронтами

Задача 7.5. Двухфронтовые регистры

Определите, как использование двухфронтовых регистров влияет на рассеяние мощности в сети распределения тактового сигнала.

7.3.3. Действительно однофазный тактируемый регистр

В описанных выше системах с двухфазным тактированием необходимо внимательно следить за трассировкой двух тактовых сигналов, чтобы минимизировать их перекрывание. Хотя C^2MOS-регистры нечувствительны к расфазировке, можно разработать регистры, которые будут использовать только однофазный тактовый сигнал. *Действительно однофазный тактируемый регистр* (True Single-Phase Clocked Register — TSPCR), предложенный Юаном (Yuan) и Свенссоном (Svensson), использует один тактовый сигнал [Yuan89]. Стандартные схемы однофазных положительной и отрицательной защелок показаны на рис. 7.30. Рассмотрим положительную защелку. Когда сигнал CLK имеет высокий уровень, защелка находится в прозрачном режиме и соответствует последовательному соединению двух инверторов; защелка

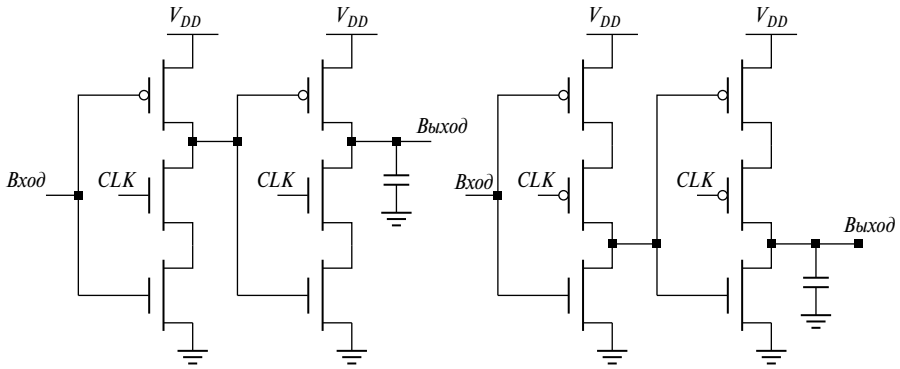


Рис. 7.30. Действительно однофазные защелки

является неинвертирующей и передает сигнал с входа на выход. С другой стороны, когда $CLK = 0$, оба инвертора отключены, и защелка находится в режиме удержания. Активны только нагрузочные сети, а разгрузочные сети в это время отключены. В этом режиме в результате использования двухкаскадного подхода никакой сигнал не может пройти со входа защелки на выход. Регистр можно организовать, последовательно соединив положительную и отрицательную защелки. В таком случае нагрузка на тактовый сигнал будет примерно такой же, как у обычного регистра на передаточных вентилях или C^2MOS-регистра. Основным преимуществом предложенной схемы является использование одной фазы, недостатком — небольшое увеличение числа транзисторов, поскольку теперь их требуется двенадцать.

Напомним, что динамическую схему, подобную изображенной на рис. 7.30, следует использовать осторожно. При низком уровне тактового сигнала (для положительных защелок) выходной узел может "плавать" и получать другие сигналы. Кроме того, может произойти перераспределение заряда, если выходной узел соединится с передаточными элементами. Динамические узлы необходимо изолировать с помощью статических инверторов или делать псевдостатическими с целью улучшения помехоустойчивости.

TSPC-регистры, как и многие другие семейства защелок, предлагают дополнительное преимущество, которое мы еще не рассмотрели: возможность внедрения в защелки функциональных возможностей логического элемента. Это позволяет получить меньшую общую задержку в сложной схеме с защелками. Суть внедрения логики в защелку объясняется на рис. 7.31, *а*, а на рис. 7.31, *б* приведен пример положительной защелки, не только выполняющей функции регистра, но и применяющей операцию И к сигналам In_1 и In_2.

Хотя время установки данной защелки больше, чем у конструкции, показанной на рис. 7.30, общее быстродействие цифровой схемы (т.е. длительность такта последовательной схемы) повысилась: увеличение времени установки обычно намного меньше задержки элемента И. Данный подход внедрения логики в защелки широко использовался в разработке микропроцессоров Alpha EV4 [Dobberpuhl92] и многих других быстродействующих процессоров.

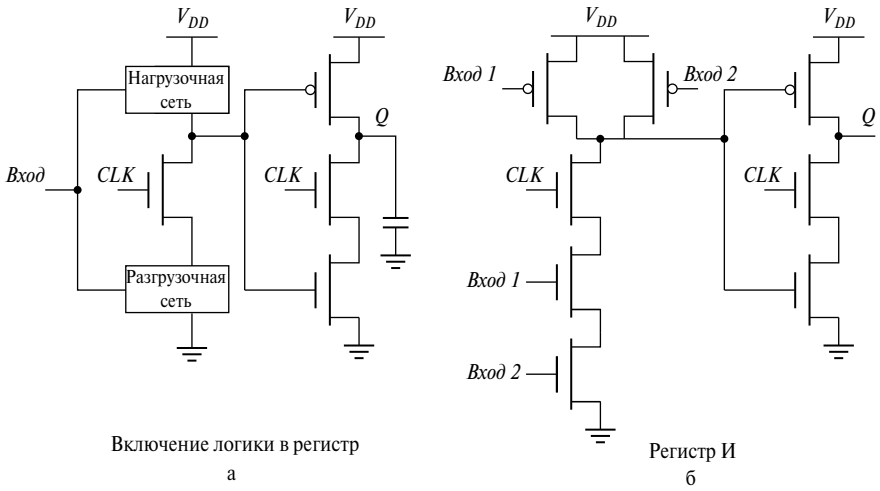


Рис. 7.31. Добавление логики в TSPC-регистр

Пример 7.3. Влияние логики, внедренной в защелки, на быстродействие

Рассмотрим внедрение элемента И в TSPC-защелку (см. рис. 7.31, *б*). В технологическом процессе 0,25-мкм время установки такой схемы (предполагается использование устройств минимального размера) равно 140 пс. При обычном подходе — элемент И, за которым следует положительная защелка, — эффективное время установки равно 600 пс (элемент И и защелку мы считаем "черным ящиком", выполняющим обе указанные функции). Таким образом, внедрение логики в защелку дает значительное увеличение быстродействия.

Как показано на рис. 7.32, сложность TSPC-защелки можно уменьшить еще больше, оставив под управлением тактового сигнала только первый инвертор. Из-за меньшего числа транзисторов данные схемы обладают дополнительным преимуществом: нагрузка на тактовый сигнала уменьшается вдвое. С другой стороны, не на всех узлах защелки размах напряжения равен полному перепаду логических уровней. Например, максимальное напряжение на узле A (при $V_{\text{вх.}} = 0$ В) для положительной защелки равно $V_{DD} - V_{Tn}$, что дает меньшее питание для выхода n-МОП-транзистора и снижение быстродействия. Подобным образом, напряжение на узле A (при $V_{\text{вх.}} = V_{DD}$) для отрицательной защелки опускается только до $\mid V_{Tp} \mid$. Помимо всего прочего, это также ограничивает возможности масштабирования V_{DD} в защелке.

На рис. 7.33 показана структура специализированного однофазного регистра, управляемого фронтом. При $CLK = 0$ входной инвертор выбирает инвертированный вход D на узле X. Второй (динамический) инвертор находится в режиме предварительной зарядки, когда M_6 заряжает узел Y до

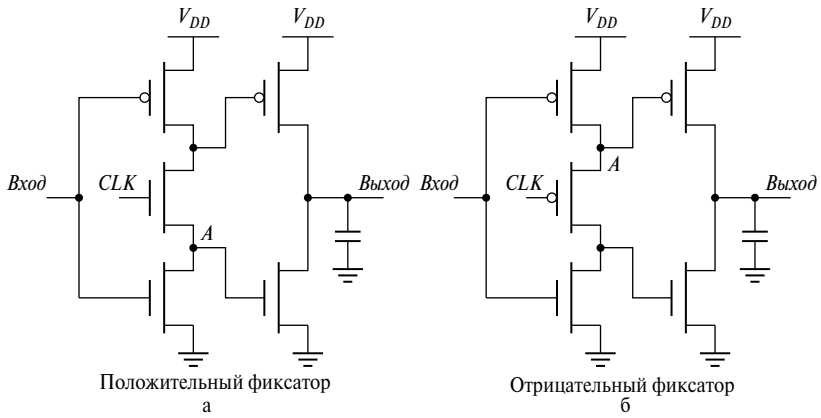


Рис. 7.32. Упрощенная TSPC-защелка

напряжения V_{DD}. Третий инвертор находится в режиме удержания, поскольку M_8 и M_9 выключены. Следовательно, в низкой фазе тактового сигнала вход последнего (статического) инвертора сохраняет свое предыдущее значение, и выход Q стабилен. На нарастающем фронте тактового сигнала динамический инвертор $M_4 - M_6$ находится в режиме оценки. Если узел X имеет высокий уровень напряжения во время нарастания тактового сигнала, то узел Y разряжается. Третий инвертор M_7–M_9 в высокой фазе включен и значение, имеющееся на узле Y, передается на выход Q. Обратите внимание на то, что в положительной фазе тактового сигнала узел X переходит в низкое состояние, если вход D меняется на высокий уровень. Следовательно, вход не должен менять свое значение, пока значение, бывшее на узле X до начала нарастания тактового сигнала, не дойдет до узла Y. Таким образом, мы получаем время удержания регистра. По сути, задержка распространения регистра равна времени прохождения сигнала через три инвертора, поскольку значение на узле X должно дойти до выхода Q. Наконец, время установки представляет собой время, в течение которого узел X не должен менять свое значение, т.е. оно равно одной задержке инвертора.

Предупреждение

Подобно C^2MOS-защелке, защелка TSPC работает неверно при недостаточно крутом наклоне тактового сигнала. Медленно меняющиеся тактовые сигналы приводят к тому, что тактируемые n-МОП- и p-МОП-транзисторы включаются одновременно, из-за чего состояния становятся неопределенными и возникают состязания. Таким образом, наклон тактового сигнала следует контролировать. При необходимости для обеспечения качества тактовых сигналов можно вводить в схему дополнительные буферы.

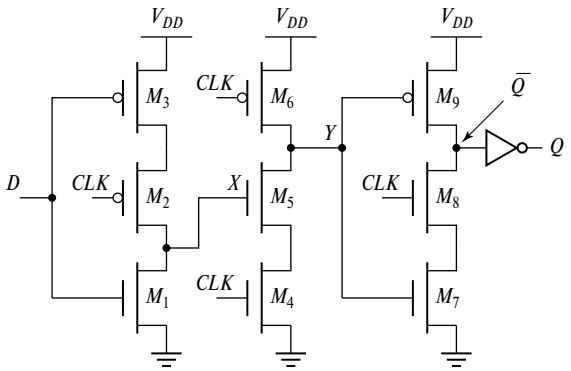
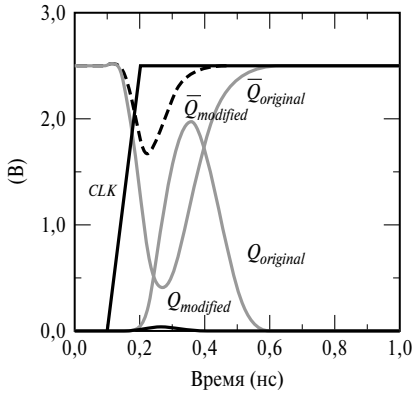


Рис. 7.33. Реализация TSPC-регистра, управляемого положительным фронтом

Пример 7.4. TSPC-регистр, управляемый фронтом

Для достижения правильной работы TSPC-регистра необходимо правильное масштабирование транзисторов. При неправильном подборе на выходе возможны импульсные помехи, вызванные *состязанием* при переходе тактового сигнала с низкого уровня на высокий. Рассмотрим ситуацию, когда D имеет низкий уровень и $\bar{Q} = 1$ ($Q = 0$). Когда CLK также имеет низкий уровень, узел Y предварительно заряжается до высокого уровня, включая транзистор M_7. Когда CLK переходит с низкого уровня на высокий, узлы Y и $\bar{Q}$ одновременно начинают разряжаться (через M_4–M_5 и M_7–M_8 соответственно). Как только Y опускается до определенного уровня, узел $\bar{Q}$ начинает заряжаться через M_9. По сути, похожая последовательность событий наблюдается при последовательном соединении динамических логических элементов. На рис. 7.34 показана переходная характеристика схемы, изображенной на рис. 7.33, для различных размеров устройств двух последних каскадов.



	M_4, M_5	M_7, M_8
Первоначальная ширина	0,5 мкм	2 мкм
Модифицированная ширина	1 мкм	1 мкм

Рис. 7.34. Иллюстрация к вопросу масштабирования TSPC-регистра (изображенного на рис. 7.33)

Данные сбои могут породить неисправимые ошибки, поскольку могут создавать нежелательные события (например, когда выход защелки используется как тактовый импульс на входе другого регистра). Кроме того, они уменьшают задержку загрязнения регистра. Проблему можно решить путем масштабирования относительных "интенсивностей" разгрузочных трактов через M_4–M_5 и M_7–M_8, чтобы узел Y разрядился гораздо быстрее, чем $\bar{Q}$. Для этого разгрузочный тракт M_7–M_8 делается менее, а тракт M_4–M_5 — более интенсивным.

7.4. Альтернативные стили регистров

7.4.1. Импульсные регистры

До этого момента для создания регистра, управляемого фронтом, мы использовали конфигурацию "master–slave". Совершенно иным подходом к созданию регистра является использование *импульсных сигналов*. Их идея заключается в том, чтобы создать короткий импульс вокруг нарастающего (или спадающего) фронта тактового сигнала. Данный импульс действует как тактовый вход защелки (как, например, на рис. 7.35, *а*), выбирая вход только в коротком промежутке времени ("окне"). Таким образом, поскольку время размыкания (т.е. прозрачный период) защелки является очень малым, состязания не происходит. В результате объединения схемы генерации импульсов и защелки получается регистр, управляемый положительным фронтом.

На рис. 7.35, *б* приведен пример схемы, генерирующей короткий импульс на каждом положительно фронте тактового сигнала [Kozu96]. При $CLK = 0$ узел X заряжается до V_{DD} (M_N выключен, поскольку $CLKG$ имеет низкий уровень). На нарастающем фронте тактового сигнала существует короткий промежуток времени, когда оба входа элемента И имеют высокий уровень, что переводит сигнал $CLKG$ также на высокий уровень. Это, в свою очередь, активизирует M_N, нагружая X и со временем понижая $CLKG$ (см. рис. 7.35, *в*). Длительность импульса контролируется задержкой элемента И и двумя инверторами. Обратите внимание на то, что также существует задержка между положительными фронтами тактового входа (CLK) и импульсного тактового входа ($CLKG$), которая равна задержке элемента И и двух инверторов. Если все регистры на чипе будут использовать одинаковый механизм генерации тактовых импульсов, данная задержка выборки не будет иметь значения. Тем не менее из-за изменений параметров процесса и значений нагрузки задержки в схеме генерации тактовых импульсов могут быть различными. Этот момент следует обязательно учитывать при верификации временны*ы*х параметров и анализе расфазировки тактовых импульсов (подробнее об этом — в главе 10).

Если время установки и время удержания измеряются относительно положительного фронта импульсного тактового сигнала, время установки практически равно нулю, время удержания равно длительности импульса, а задерж-

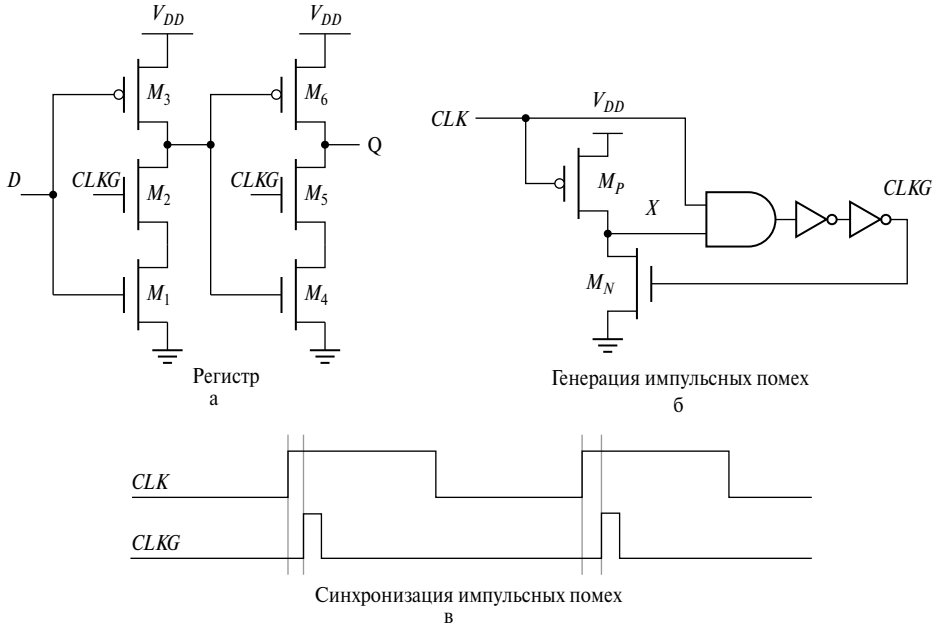


Рис. 7.35. Генерация импульса и регистр на основе TSPC

ка распространения (t_{c-q}) — двум задержкам логического элемента. Достоинством описанного подхода является *сниженная нагрузка на тактовый сигнал и меньшее чисто требуемых транзисторов*. Для уменьшения амортизации схему генерации импульсов можно использовать для питания нескольких битов регистра. Недостаток данной схемы — существенное усложнение верификации. Чтобы схема работала правильно, необходимо смоделировать все узлы, чтобы убедиться, что тактовый импульс существует всегда (т.е. что схема генерации импульсов работает надежно). Несмотря на повышенную сложность, описанные регистры являются альтернативой обычных схем, используемой в некоторых быстродействующих процессорах (см., например, [Kozu96]).

Другой вариант импульсного регистра показан на рис. 7.36 (используется в процессоре AMD-K6 [Partovi96]). При низком уровне тактового сигнала M_3 и M_6 выключены, а устройство P_1 включено. Узел X предварительно заряжается до V_{DD}, выходной узел (Q) отсоединен от X и хранит свое предыдущее состояние. Сигнал $\overline{CLKD}$ представляет собой задержанную и инвертированную версию CLK. На положительном фронте тактового сигнала M_3 и M_6 включаются, тогда как устройства M_1 и M_4 включены на протяжении короткого промежутка времени, определяемого задержкой трех инверторов. В это время схема пропускает сигнал и входные данные D выбираются защелкой. Как только сигнал $CLKD$ начинает спадать, узел X отсоединяется от входа D и либо остается в текущем состоянии, либо начинает предварительную зарядку до V_{DD} через p-МОП-устройство P_2. На отрицательном фронте тактового сигнала узел X удерживает значение V_{DD}, и выход поддерживается стабильным с помощью инверторов с перекрестными обратными связями.

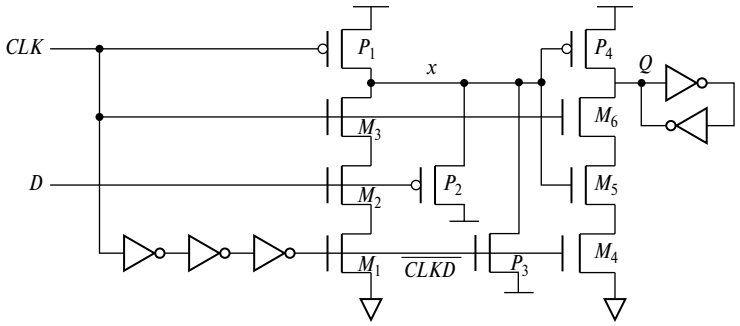


Рис. 7.36. Регистр, управляемый положительным фронтом

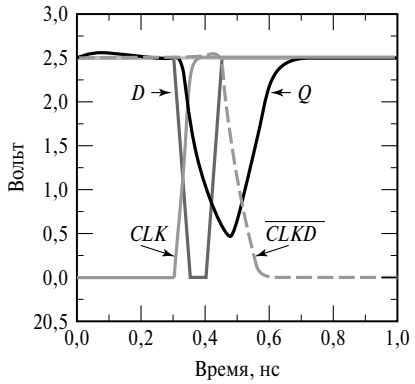


Рис. 7.37. Согласно моделированию, импульсный регистр имеет отрицательное время установки

Обратите внимание на то, что в данной схеме также используется *генератор импульсов*, интегрированный в регистр. Период прозрачности также определяет время удержания регистра. "Окно" должно быть достаточно широким, чтобы входные данные были переданы на выход Q. В приведенном примере время установки может быть отрицательным. Так бывает, когда окно прозрачности больше, чем задержка между входом и выходом. Это удобно, поскольку данные могут поступить на регистр даже после того, как уровень тактового сигнала начал расти, т.е. получается, что это время как бы заимствуется из предыдущего такта.

Пример 7.5. Время установки импульсного регистра

Импульсный регистр, показанный на рис. 7.36, является прозрачным во время перекрывания (1-1) сигналов CLK и $\overline{CLKD}$. В результате входные данные могут в действительности меняться после положительного фронта тактового сигнала, что даст отрицательное время установки (рис. 7.37). Вход D переходит в низкое состояние после положительного фронта тактового сиг-

нала и в высокое состояние перед отрицательным фронтом сигнала $\overline{CLKD}$ (т.е. в фазе прозрачности). Посмотрите, как выход следует за входом. Выход Q переходит в правильное состояние V_{DD}, если вход D имеет правильное значение некоторое время перед отрицательным фронтом $\overline{CLKD}$. При использовании отрицательного времени установки монотонное поведение выхода гарантировать невозможно, т.е. выход может давать несколько переходов на положительном фронте, а следовательно, выход регистра не следует использовать для управления динамической логикой или в качестве тактового сигнала других регистров.

Задача 7.6. Преобразование импульсного регистра в условный импульсный регистр

Модифицируйте схему, изображенную на рис. 7.36, введя в нее дополнительный вход *Enable*. Ваша задача — преобразовать устройство в условный регистр, который будет фиксировать данные только при получении разрешающего сигнала.

7.4.2. Регистры, основанные на усилителях считывания

Помимо конфигурации "master–slave" и управления импульсами, существует еще одна техника создания регистра, управляемого фронтом, которая основана на *усилителях считывания* (sense amplifier) (см. рис. 7.38 и [Montanaro96])[3]. Схемы усиления считывания принимают небольшие входные сигналы и усиливают их до полного размаха сигнала. Эти устройства широко используются в сердечниках запоминающих устройство и шинных усилителях с малым размахом либо для повышения быстродействия, либо для снижения рассеяния мощности. Существует множество технологий создания данных усилителей. Распространенным подходом является использование обратной связи, например, посредством ряда инверторов с перекрестными обратными связями. В схеме, показанной на рис. 7.38, используется предварительно заряженный входной усилитель, выбирающий входной сигнала на нарастающем фронте тактового сигнала. Выходы переднего каскада подаются на SR-триггер НЕ-И с перекрестными обратными связями, который хранит данные и гарантирует, что дифференциальные выходы переключаются только один раз за такт. Отметим, что в данной реализации дифференциальные входы не обязательно должны иметь полный размах.

Основой переднего каскада является инвертор с перекрестными обратными связями (M_5–M_8), выходы которого (L_1 и L_2) предварительно заряжаются с помощью устройств M_9 и M_{10} в низкой фазе тактового сигнала. В результате p-МОП-транзисторы M_7 и M_8 выключены и триггер НЕ-И сохраняет свое предыдущее состояние. Транзистор M_1 подобен ключу оценки в динамиче-

[3] По сути, работа данных регистров, основанных на усилителях считывания, подобна действию импульсных регистров, т.е. первый каскад генерирует импульс, а второй — фиксирует его.

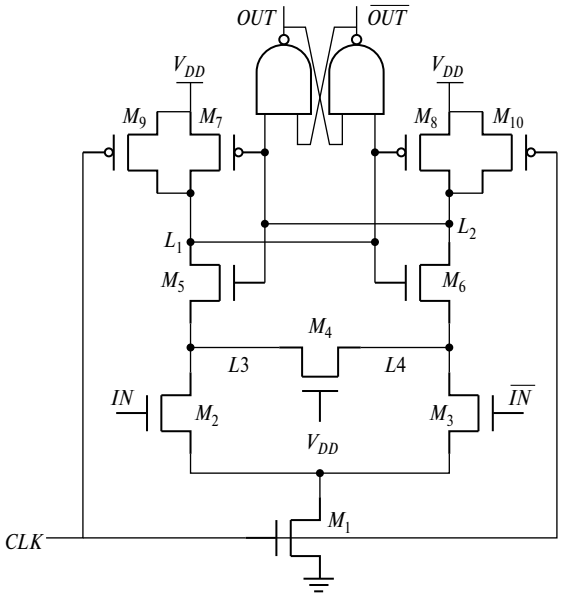


Рис. 7.38. Регистр, управляемый положительным фронтом, основанный на усилителе считывания

ских схемах, он включается, чтобы гарантировать, что дифференциальные входы не влияют на выход в низкой фазе тактового сигнала. На нарастающем фронте тактового сигнала транзистор оценки включается, активизирует пару дифференциальных входов (M_2 и M_3), и разница между входными сигналами усиливается на выходных узлах L_1 и L_2. Пара инверторов с перекрестными обратными связями перебрасывается в одно из устойчивых состояний согласно значению входов. Например, если IN равно 1, то L_1 устанавливается в 0, а L_2 остается со значением V_{DD}. Из-за усиливающей способности входного каскада вход совсем не обязательно доводить до полной амплитуды V_{DD}, что позволяет подавать на вход сигналы с неполным размахом.

Закорачивающий транзистор M_4 используется для заземления либо узла L_3, либо узла L_4. Таким образом, учитывается возможность изменения входами своих значений после положительного фронта сигнала CLK, которое приведет к тому, что L_3 либо L_4 остается в высокоимпедансном состоянии и хранит низкий логический уровень напряжения. Без тракта утечки данный узел будет заряжаться токами утечки. В результате защелка может изменить состояние до следующего нарастающего фронта CLK! Чтобы лучше понять сказанное, удобно использовать рис. 7.39.

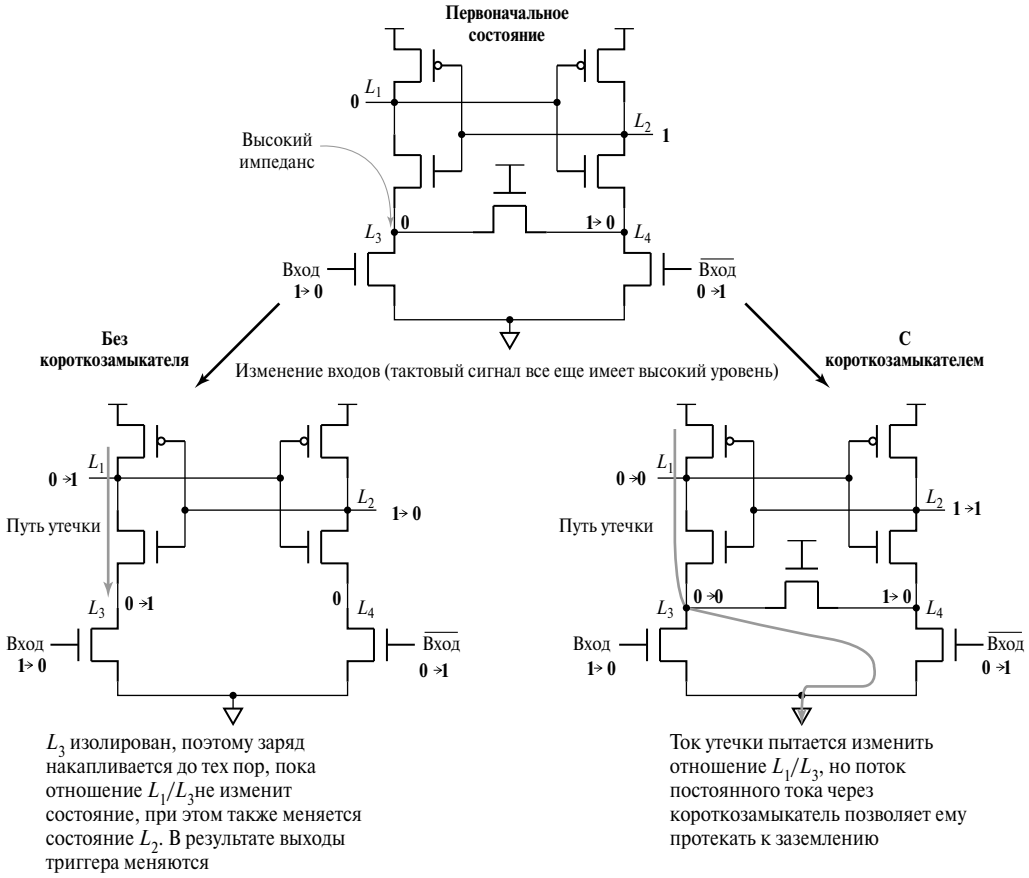


Рис. 7.39. Объяснение необходимости закорачивающего транзистора M_4

7.5. Конвейер: подход к оптимизации последовательных цепей

Конвейерный подход представляет собой популярную технологию разработки, которая часто используется для ускорения работы информационных каналов в цифровых процессорах. Объясним суть подхода на примере, изображенном на рис. 7.40, *а*. Данная схема предназначена для расчета $log(|a+b|)$, где сигналы a и b представляют собой потоки чисел (т.е. обработке подвергаются большие наборы входных значений). Минимальный такт T_{min}, требуемый для обеспечения правильной работы схемы, находится следующим образом:

$$T_{min} = t_{c-q} + t_{pd,logic} + t_{su}, \tag{7.7}$$

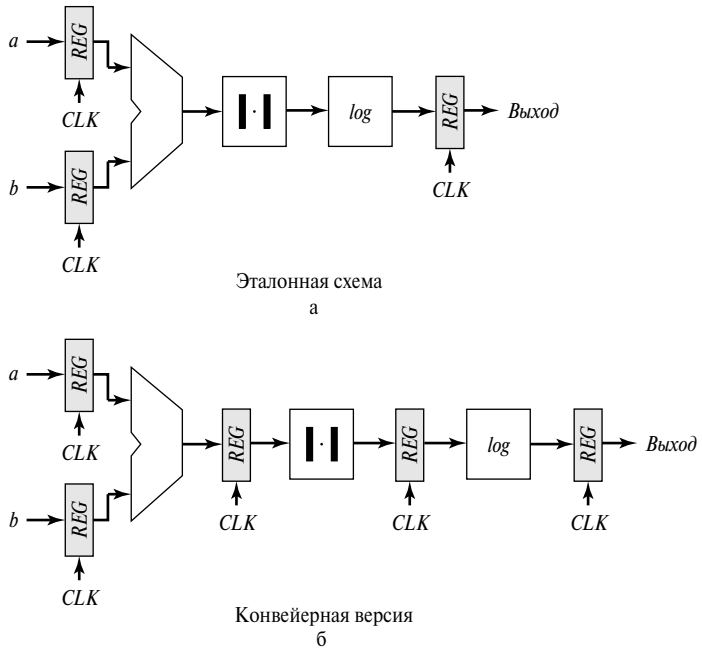


Рис. 7.40. Информационный канал для расчета $\log(|a+b|)$

где t_{c-q} и t_{su} — задержка распространения и время установки регистра соответственно. Мы предполагаем, что регистры представляют собой D-регистры, управляемые фронтом. Через $t_{pd,logic}$ обозначена максимальная задержка прохождения сигнала через комбинационную сеть, состоящую из реализаций функций суммы, модуля и логарифма. В обычных системах (не использующих современные технологии) последняя задержка обычно гораздо больше задержек, связанных с регистрами, поэтому именно она определяет быстродействие схемы. Предположим, что все логические модули имеют равные задержки распространения. В таком случае каждый блок будет активен только треть времени такта (здесь мы игнорируем задержку регистра). Например, сумматор активен в первой трети такта, после чего две трети периода находится в холостом состоянии (не выполняет полезных расчетов). Чтобы полнее использовать доступные ресурсы и увеличить функциональную пропускную способность, применяется технология конвейерной обработки. Предположим, что между логическими блоками мы ввели регистры, как показано на рис. 7.40, *б*. В результате, как видно из табл. 7.1, вычисления для одного набора входных данных распределяются по нескольким тактам. Результат расчетов для первого набора (a_1, b_1) появляется на выходе только через три такта. Однако в это время схема успевает выполнить часть вычислений для следующих наборов данных (a_2, b_2) и (a_3, b_3).

Преимущество конвейерной обработки станет понятным, если посмотреть на величину минимального такта модифицированной схемы. Блок комбина-

Таблица 7.1. Пример конвейерных вычислений

Такт	Сумма	Модуль	Логарифм
1	$a_1 + b_1$		
2	$a_2 + b_2$	$\|a_1 + b_1\|$	
3	$a_3 + b_3$	$\|a_2 + b_2\|$	$\log(\|a_1 + b_1\|)$
4	$a_4 + b_4$	$\|a_3 + b_3\|$	$\log(\|a_2 + b_2\|)$
5	$a_5 + b_5$	$\|a_4 + b_4\|$	$\log(\|a_3 + b_3\|)$

ционной схемы был разбит на три секции, каждая из которых обладает меньшей задержкой распространения, чем исходная функция. Таким образом, мы эффективно уменьшили значение минимально допустимого такта:

$$T_{min,pipe} = t_{c-q} + max(t_{pd,add}, t_{pd,abs}, t_{pd,log}) + t_{su}. \tag{7.8}$$

Предположим, что все логические блоки имеют приблизительно равные задержки распространения и что служебные издержки регистра малы по сравнению с задержками логических блоков. При таких предположениях конвейерная сеть втрое превосходит по быстродействию исходный вариант (т.е. $T_{min,pipe} = T_{min}/3$). Ценой (сравнительно малой) увеличения быстродействия является два дополнительных регистра и увеличенное время ожидания[4]. Из сказанного должно быть понятно, почему конвейерная обработка используется в реализациях многих быстродействующих информационных каналов.

7.5.1. Конвейер на защелках или на регистрах?

Конвейерные схемы можно создавать, используя вместо регистров, управляемых фронтом, уровневые защелки. Рассмотрим конвейерную схему, изображенную на рис. 7.41. Система конвейера реализована с использованием положительной и отрицательной защелок, основанных на проходных транзисторах, т.е. между главной и вспомогательной защелками системы "master–slave" вводится логическая схема. Ниже без потери общности мы будем описывать систему с двухфазным тактированием через сигналы CLK-$\overline{CLK}$. Системы, основанные на защелках, предлагают гораздо бо́льшую гибкость в реализации конвейерных систем, чем системы на регистрах, при этом обладая еще и бо́льшим быстродействием. Конвейер работает правильно, когда тактовые сигналы CLK и $\overline{CLK}$ не перекрываются. Входные данные выбираются на узле C_1 на отрицательном фронте сигнала CLK и начинаются вычисления логического блока F; результат этих вычислений сохраняется на узле C_2 на отрицательном фронте сигнала $\overline{CLK}$, после чего начинаются вычисления блока G. Правильная работа схемы достигается за счет того, что тактовые сигналы

[4] Под *временем ожидания* (latency) мы понимаем число тактов, требуемых данным для прохождения со входа на выход. В рассматриваемом примере конвейерная обработка увеличивает время ожидания с 1 до 3 тактов. Как правило, увеличение времени ожидания является допустимым, однако если неаккуратно учесть все связанные с ним факторы, это может привести к глобальному ухудшению быстродействия.

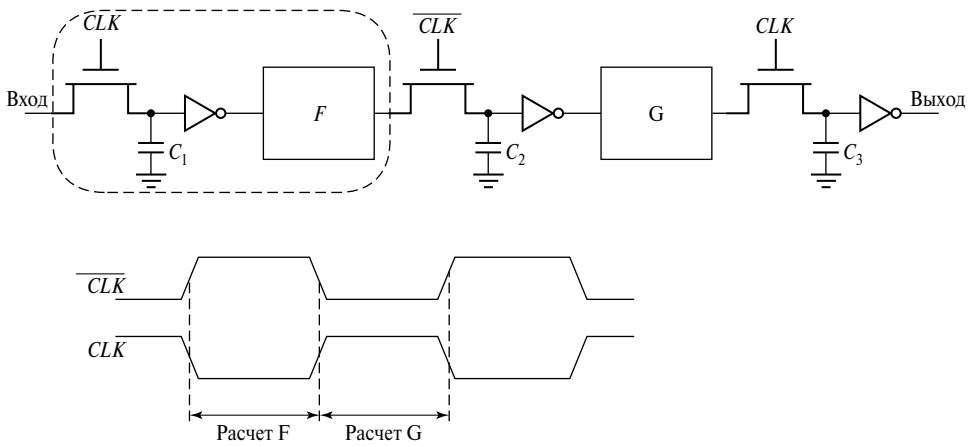


Рис. 7.41. Работа двухфазной конвейерной схемы с использованием динамических регистров

не перекрываются. Значение, записанное в C_2 в конце низкой фазы сигнала CLK, представляет собой результат передачи предыдущего входа (записанного на узле C_1 на спаде сигнала CLK) через логическую функцию F. Если сигналы CLK и $\overline{CLK}$ перекрываются, то следующий вход уже подан на F и его влияние может распространиться на C_2 до понижения уровня сигнала $\overline{CLK}$ (предполагается, что задержка загрязнения узла F мала). Другими словами, между предыдущим и текущим входом происходит *состязание*. Какое значение победит — зависит от логики и часто является функцией входов. Последний фактор несколько усложняет обнаружение и устранение условий состязания.

7.5.2. NORA–КМОП — стиль логики конвейерных структур

Как показано на рис. 7.42, конвейерную схему на основе защелок можно реализовать и с использованием C^2MOS-защелок. Ее действие подобно работе схемы, рассмотренной в разделе 7.5.1, только предлагаемая топология обладает дополнительным важным свойством.

В конвейерной схеме на основе C^2MOS состязания невозможны, если все логические функции F (реализованные с использованием статической логики), действующие между защелками, являются неинвертирующими.

Аргументы в защиту подобного утверждения аналогичны аргументам, использованным при создании C^2MOS-регистра. В фазе перекрывания (0-0) сигналов CLK и $\overline{CLK}$ все C^2MOS-защелки упрощаются до чистых нагрузочных сетей (см. рис. 7.27). Как показано на рис. 7.43, единственная возможность возникновения состязания реализуется, когда логическая функция F является инвертирующей (F на рисунке заменена статическим КМОП-инвертором). Подобные соображения можно привести и для перекрывания (1-1).

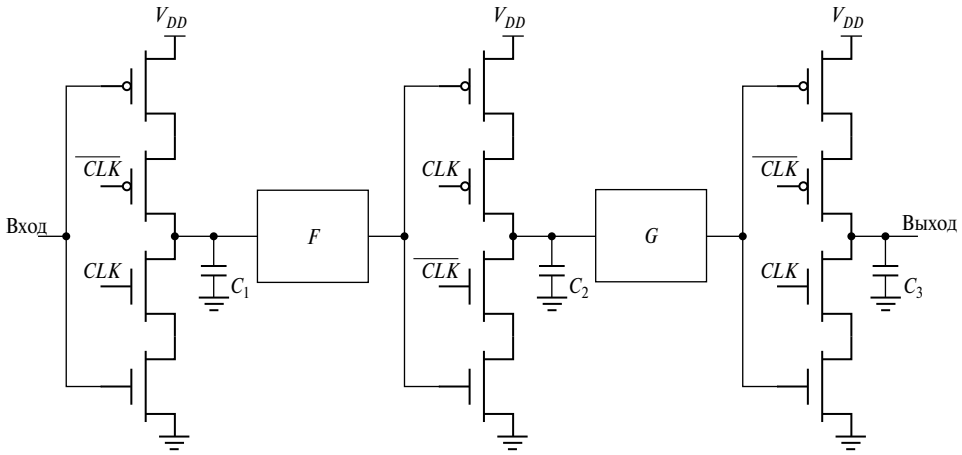


Рис. 7.42. Конвейерный информационный канал, использующий C^2MOS-защелки

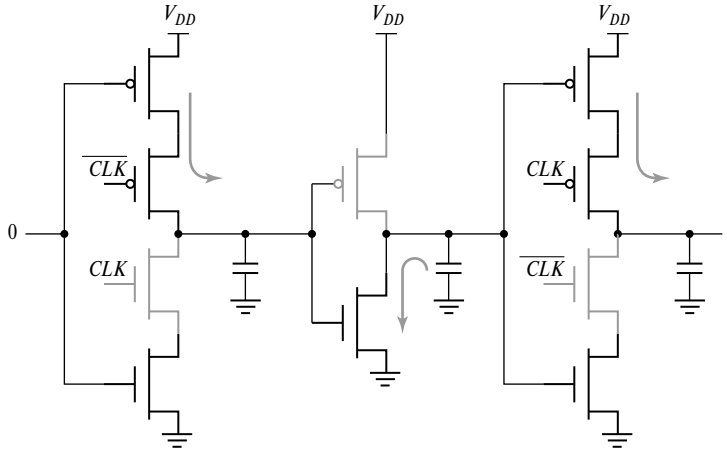


Рис. 7.43. Потенциальное состязание в фазе перекрывания (0-0) при использовании структуры, основанной на C^2MOS

На основе данной концепции был разработан стиль логических схем, названный NORA–CMOS [Goncalves83]. Он объединяет конвейерные регистры C^2MOS и динамические блоки логических функций NORA. Каждый модуль состоит из блока комбинационной логики, который может представлять собой смесь статической и динамической логики, за которым следует C^2MOS-защелка. Логика и защелка тактируются так, чтобы они одновременно находились либо в режиме оценки, либо в режиме удержания (предварительной зарядки). Блок, который в фазе $CLK = 1$ находится в режиме оценки, называется *модулем* CLK, его противоположность именуется *модулем* $\overline{CLK}$. Примеры модулей обоих типов показаны на рис. 7.44, *а* и *б* соответственно. Данные о рабочих режимах модулей представлены в табл. 7.2.

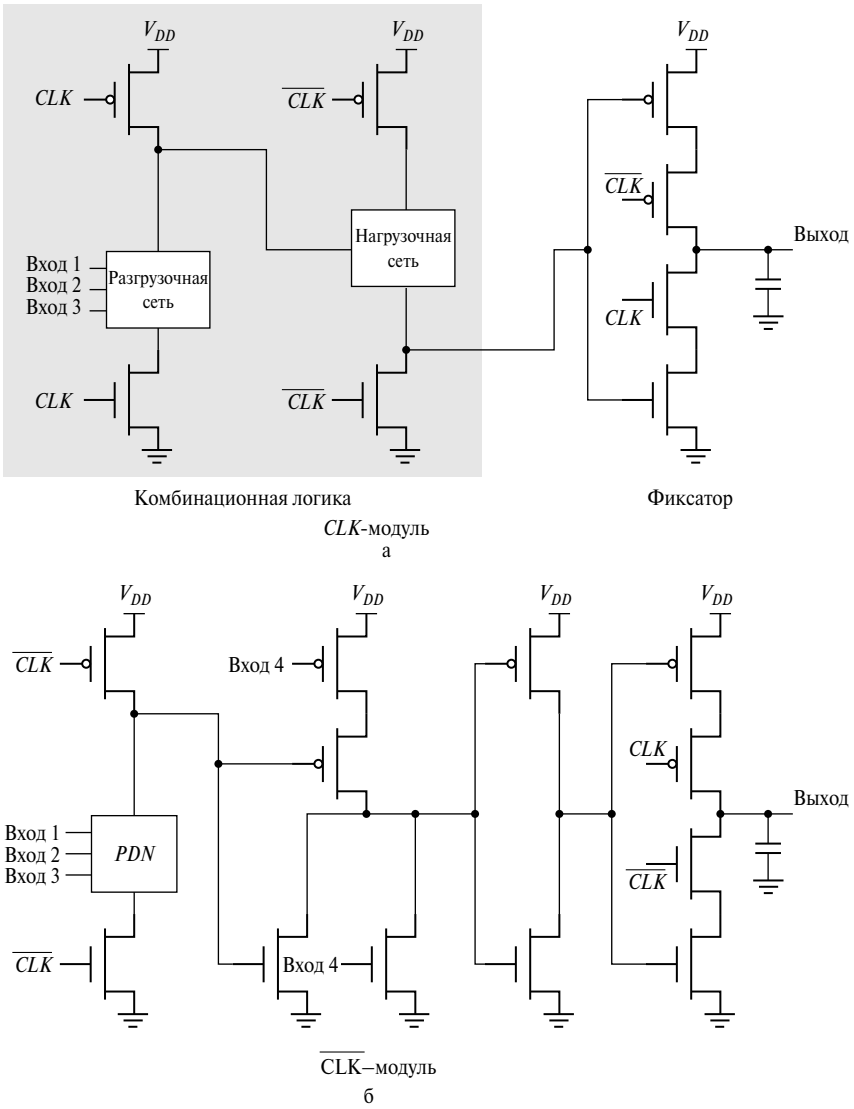


Рис. 7.44. Примеры модулей NORA–CMOS

Информационный канал NORA состоит из цепочки чередующихся модулей CLK и $\overline{CLK}$. Когда модули одного класса находятся в режиме предварительной зарядки и их выходная защелка находится в режиме удержания, сохраняя предыдущее выходное значение, модули другого класса находятся в режиме оценки. Данные передаются конвейером из модуля в модуль. NORA предлагает богатый выбор стилей разработки. Например, можно свободно смешивать динамическую и статическую логику, динамические блоки CLK_p и CLK_n можно соединять каскадом или конвейером. Хотя в данном стиле

Таблица 7.2. Рабочие режимы логических модулей NORA

	Блок CLK		Блок $\overline{CLK}$	
	Логика	Защелка	Логика	Защелка
$CLK = 0$	Предварительная зарядка	Удержание	Оценка	Оценка
$CLK = 1$	Оценка	Оценка	Предварительная зарядка	Удержание

отсутствует дополнительный инвертор, требуемый в схемах "домино", здесь также существует множество правил, которых следует придерживаться для достижения надежной работы без состязаний. Таким образом, из-за дополнительных сложностей стиль NORA применяется только в быстродействующих приложениях.

7.6. Небистабильные последовательные схемы

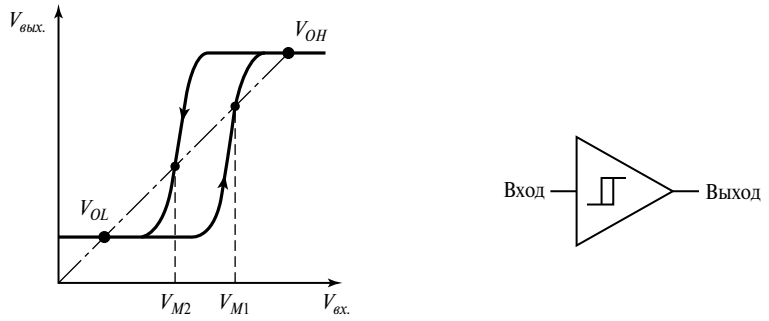
В предыдущих разделах мы сосредоточили внимание на одном типе последовательного элемента: защелке (и ее "собрате" — регистре). Важнейшее свойство такой схемы заключается в том, что она имеет два устойчивых состояния, следовательно, является *бистабильной*. Отметим, что бистабильный элемент — это не единственная последовательная схема, представляющая практический интерес. К другим важным регенеративным схемам относятся *автоколебательные схемы* (astable circuit) и *ждущие мультивибраторы* (monostable circuit). Автоколебательные схемы действуют как осцилляторы и могут, например, использоваться для внутрикристальной генерации тактового сигнала. Ждущие мультивибраторы (или *одновибраторы*) действуют как генераторы импульсов. Еще одной интересной регенеративной схемой является триггер Шмитта, передаточная функция которого имеет петлю гистерезиса — порог переключения триггера Шмитта зависит от направления перехода (от низкого уровня к высокому или от высокого к низкому). В зашумленных средах данное свойство может оказаться весьма полезным.

7.6.1. Триггер Шмитта

Определение

Триггер Шмитта [Schmitt38] — это устройство, обладающее двумя важными свойствами.

1. На медленно меняющийся входной сигнал он реагирует *быстрым переходом на выходе*.
2. Передаточная характеристика по напряжению устройства дает *различные пороги переключения для положительного и отрицательного входных сигналов*. Сказанное иллюстрируется на рис. 7.45, где показана типичная передаточная характеристика по напряжению триггера Шмитта (и его обозначение в схеме). Пороги переключения для изенения уровня сигнала



Передаточная характеристика по напряжению
а

Схематическое изображение
б

Рис. 7.45. Неинвертирующий триггер Шмитта

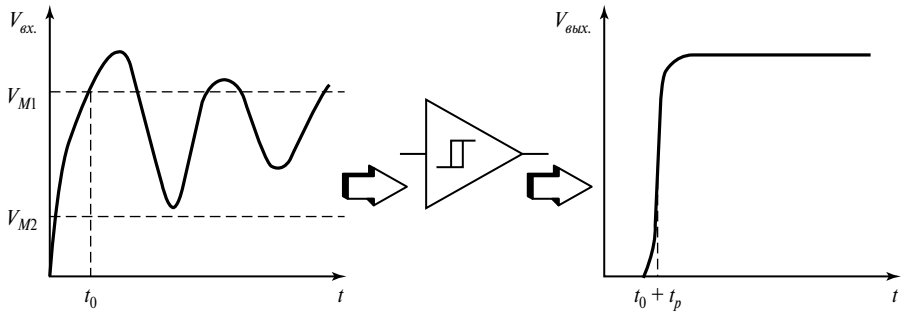


Рис. 7.46. Подавление шума с использованием триггера Шмитта

от низкого к высокому и от высокого к низкому обозначаются V_{M+} и V_{M-} соответственно. *Напряжение гистерезиса* определяется как разность этих двух величин.

Одной из основных сфер применения триггера Шмитта является преобразование зашумленного или медленно меняющегося входного сигнала в четкий цифровой выходной сигнал (рис. 7.46). Обратите внимание на то, как гистерезис подавляет колебания сигнала. В то же время должен быть виден быстрый переход выходного сигнала с низкого уровня на высокий (или с высокого на низкий). В общем случае полезны сигналы с крутым фронтом; например, это помогает уменьшить потребление мощности, подавив токи прямого тракта. Секретом концепции триггера Шмитта является использование положительной обратной связи.

КМОП-реализация

Одна из возможных КМОП-реализаций триггера Шмитта показана на рис. 7.47. Идея, воплощенная в данной схеме, заключается в том, что порог переключения КМОП-инвертора определяется отношением (k_n/k_p) p-МОП-

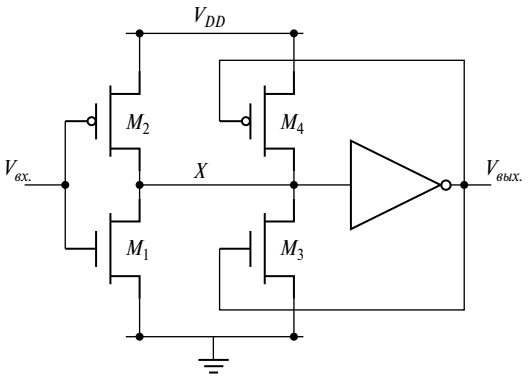


Рис. 7.47. Реализация триггера Шмитта в КМОП

и n-МОП-транзисторов. Увеличение отношения приводит к повышению порога, а уменьшение — снижает V_M. Подстройка отношения в зависимости от направления перехода приводит к смещению порога переключения и гистерезису. На практике указанная подстройка производится с помощью обратной связи.

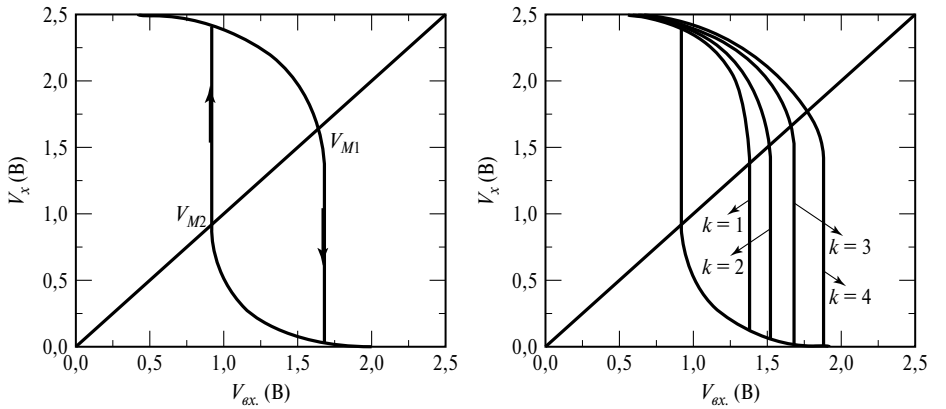
Предположим, что $V_{\text{вх.}}$ изначально равно 0, так что $V_{\text{вых.}}$ также равно 0. Контур обратной связи отпирает p-МОП-транзистор M_4, тогда как транзистор M_3 закрыт. Входной сигнал эффективно соединяется с инвертором, состоящим из двух параллельно включенных p-МОП-транзисторов (M_2 и M_4), которые действуют как нагрузочная сеть, и одного n-МОП-транзистора (M_1) в разгрузочной цепи. Таким образом, эффективное отношение транзисторов становится равным $k_{M1}/(k_{M2} + k_{M4})$, и порог переключения поднимается.

Как только инвертор переключается, контур обратной связи выключает M_4 и активизируется n-МОП-устройство M_3. Данное дополнительное разгрузочное устройство ускоряет переход и создает чистый выходной сигнал с крутым фронтом.

Похожее поведение наблюдается при изменении уровня сигнала от высокого к низкому. В этом случае разгрузочная сеть первоначально состоит из параллельно включенных транзисторов M_1 и M_3, а нагрузочная сеть формируется транзистором M_2. В результате порог переключения понижается до V_{M-}.

Пример 7.6. КМОП-триггер Шмитта

Рассмотрим триггер Шмитта, изображенный на рис. 7.47, где размеры M_1 и M_2 равны 1 мкм/0,25 мкм и 3 мкм/0,25 мкм соответственно. Инвертор спроектирован так, что порог переключения равен примерно $V_{DD}/2$ ($= 1,25$ В). На рис. 7.48, *а* триггер Шмитта смоделирован в предположении, что устройства M_3 и M_4 имеют размеры 0,5 мкм/0,25 мкм и 1,5 мкм/0,25 мкм соответственно. Как видно на графике, схема характеризуется эффектом гистерезиса. Точка переключения с низкого уровня на высокий ($V_{M-} = 0,9$ В) находится ниже $V_{DD}/2$, а порог переключения из высокого состояния в низкое ($V_{M+} = 1,6$ В)



Передаточная характеристика по напряжению с гистерезисом
а

Влияние изменения коэффициента M_4 p-МОП-устройства. Ширина равна $k \times 0{,}5$ мкм
б

Рис. 7.48. Моделирование триггера Шмитта

больше $V_{DD}/2$. Точку переключения можно сместить, изменив размеры M_3 и M_4. Например, чтобы модифицировать переход сигнала с низкого уровня на высокий, необходимо изменить p-МОП-устройство. Порог переключения с высокого уровня на низкий удерживается постоянным путем фиксации ширины устройства M_3 на уровне 0,5 мкм. Ширина устройства M_4 изменяется как $k \times 0,5$ мкм. На рис. 7.48, *б* показано увеличение порога переключения с ростом значений k.

Задача 7.7. Альтернативный КМОП-триггер Шмитта

На рис. 7.49 показана другая версия КМОП-триггера Шмитта. Обсудите работу этого элемента и выведите формулы для V_{M-} и V_{M+}.

7.6.2. Последовательные схемы с одним устойчивым состоянием

Элемент с одним устойчивым состоянием (*ждущий мультивибратор*) представляет собой схему, генерирующую импульс предопределенной длительности всякий раз, когда схема, находящаяся в состоянии покоя, активизируется импульсом или изменением уровня сигнала. Название элемента объясняется тем, что он имеет единственное устойчивое состояние (состояние покоя). Активизирующее событие (либо изменение уровня сигнала, либо импульс) временно переводит элемент в другое квазиустойчивое состояние. Это означает, что со временем схема вернется в исходное состояние по прошествии времени, определяемого ее параметрами. Вообще, подобные схемы полезны при генерации импульсов заданной длительности, что требуется во многих

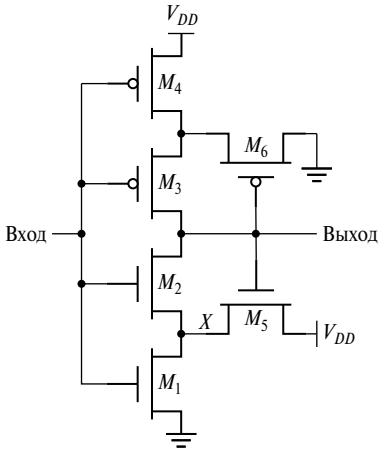


Рис. 7.49. Альтернативный КМОП-триггер Шмитта

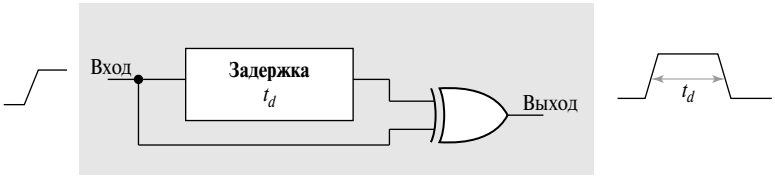


Рис. 7.50. Ждущий мультивибратор, активизируемый переходами

приложениях. Выше мы уже показывали, как схемы с одним устойчивым состоянием используются в создании импульсных регистров. Другим распространенным примером их применения являются схемы *детектирования адресных переходов* (Address Transition Detection — ATD), применяющиеся в тактовых генераторах статических элементов памяти. Эта схема детектирует изменение сигнала или группы сигналов, например адресной шины или шины данных, и генерирует импульс, инициализирующий последующую схему.

Наиболее распространенным подходом к реализации ждущих мультивибраторов является использование простого элемента задержки для управления длительностью импульса. Суть концепции иллюстрируется на рис. 7.50. В состоянии покоя оба входа элемента ИСКЛЮЧАЮЩЕЕ ИЛИ идентичны, и выход имеет низкий уровень. Переход на входе приводит к тому, что входы логического элемента временно становятся разными, и выход переходит на высокий уровень. После задержки t_d (вводится элементом задержки) данное возмущение устраняется, и выход снова переходит на низкий уровень. Таким образом, создается импульс длительности t_d. Схему задержки можно реализовать множеством различных способов, например, в виде RC-цепи или цепи стандартных элементов.

7.6.3. Автоколебательные схемы

Автоколебательная система не имеет устойчивых состояний (поэтому также называется *неустойчивой*). Ее выход осциллирует между двумя квазиустойчивыми состояниями, а период колебаний определяется топологией и параметрами схемы (задержка, электропитание и т.д.). Одной из основных сфер применения осцилляторов является внутрикристальная генерация тактовых сигналов. (Подробнее это будет рассмотрено в следующих главах.)

Простым примером автоколебательной системы является кольцевой генератор, состоящий из нечетного числа инверторов, соединенных в кольцеобразную цепь. Из-за нечетного числа инверторов устойчивой рабочей точки не существует, и схема осциллирует с периодом, равным $2 \times t_p \times N$, где N — число инверторов в цепи; t_p — задержка распространения одного инвертора.

Пример 7.7. Кольцевой генератор

На рис. 7.51 показана смоделированная характеристика кольцевого генератора, состоящего из пяти каскадов (все элементы используют устройства минимального размера). Наблюдаемый период колебаний приблизительно равен 0,5 нс, что соответствует задержке распространения элемента в 50 пс. Путем снятия сигнала в различных точках цепи были получены различные фазы осциллирующего сигнала. (На графике представлены фазы 1, 3 и 5.) С помощью простых логических операций из данных элементарных сигналов можно получить большой набор тактовых сигналов с различными рабочими циклами и фазами.

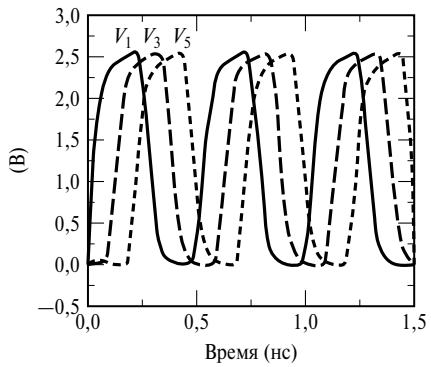


Рис. 7.51. Смоделированные сигналы пятикаскадного кольцевого генератора. Показаны выходы каскадов 1, 3 и 5

Кольцевой генератор, сформированный каскадными инверторами, генерирует сигнал фиксированной частоты, определяемый задержкой инвертора в КМОП-процессе. Во многих приложениях частоту генератора необходимо

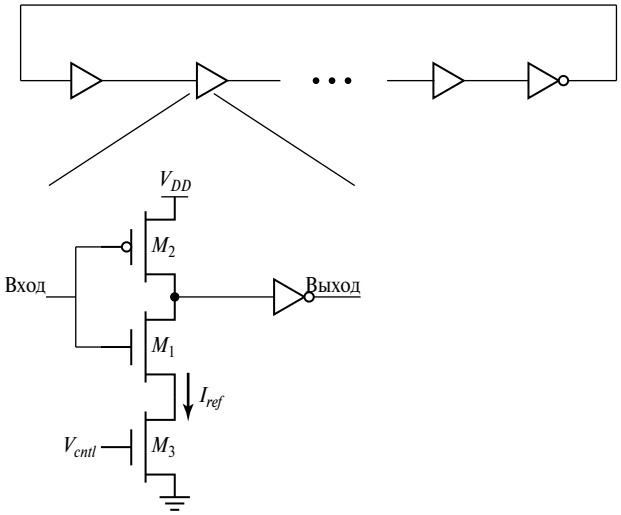


Рис. 7.52. Генератор, управляемый напряжением, основанный на инверторах, поглощающих ток

контролировать. В качестве примера подобной схемы можно привести *генератор, управляемый напряжением*, ГУН (voltage-controlled oscillator — VCO), частота колебаний которого является функцией (как правило, нелинейной) управляющего напряжения. Стандартный кольцевой генератор можно преобразовать в ГУН, заменив стандартный инвертор *инвертором, поглощающим ток*, подобным изображенному на рис. 7.52 [Jeong87]. В качестве механизма управления задержкой каждого применяется ограничение тока, доступного для разрядки емкости нагрузки элемента.

В данной схеме с модифицированным инвертором максимальной разрядный ток инвертора ограничивается путем добавления дополнительных последовательных устройств. Обратите внимание на то, что изменением уровня сигнала от низкого к высокому на инверторе также можно управлять, добавив последовательно к M_2 p-МОП-устройство. Дополнительный n-МОП-транзистор M_3 контролируется аналоговым управляющим напряжением V_{cntl}, которое определяет доступный разрядный ток. Понижение V_{cntl} приводит к уменьшению разрядного тока и, следовательно, увеличивает t_{pHL}. Возможность менять величину задержки распространения на каскад позволяет контролировать частоту кольцевой структуры. Управляющее напряжение обычно задается с помощью технологий обратной связи. В условиях низкого рабочего тока инвертор, поглощающий ток, обладает большим временем спада на выходе, что может привести к большому току короткого замыкания. Данная проблема решается подачей выхода схемы на КМОП-инвертор или (что даже лучше) на триггер Шмитта. Чтобы обеспечить осцилляцию сигнала, на выходе схеме требуется еще один инвертор.

Пример 7.8. Моделирование инвертора, поглощающего ток

На рис. 7.53 показана смоделированная задержка инвертора, поглощающего ток, как функция управляющего напряжения V_{cntl}. Задержку инвертора можно изменять в довольно широких пределах. Когда управляющее напряжение меньше порога, устройство вступает в подпороговую область. Это приводит к большим колебаниям задержки распространения, поскольку управляющий ток экспоненциально зависит от управляющего напряжения. При работе в данной области задержка очень чувствительна к колебаниям управляющего напряжения, а следовательно, к шуму.

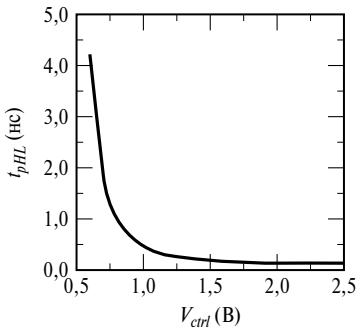


Рис. 7.53. Задержка t_{pHL} инвертора, поглощающего ток, в зависимости от управляющего напряжения

Как показано на рис. 7.54, *а*, существует и другой подход к реализации элемента задержки — с помощью дифференциального элемента. Поскольку элемент задержки обеспечивает и инвертированный, и неинвертированный выход, это позволяет реализовать генератор с четным числом каскадов. На рис. 7.54, *б*, например, показан двухкаскадный дифференциальный ГУН, в котором контур обратной связи с помощью задержки двух элементов (неинвертирующего и инвертирующего) обеспечивает сдвиг фазы на 180°, позволяющий создать осцилляции. Смоделированные сигналы данного двухкаскадного ГУН показаны на рис. 7.54, *в*. Синфазный и квадратурный выходы схемы доступны одновременно. Отметим, что дифференциальный ГУН обладает лучшей устойчивостью к типичным помехам (например, к шуму источника питания), чем обычный кольцевой генератор. Тем не менее он потребляет больше мощности из-за повышенной сложности и тока покоя.

7.7. Перспектива: выбор стратегии тактирования

На ранних этапах проектирования чипа необходимо принять одно важное решение — выбрать подходящую методологию тактирования. Надежная син-

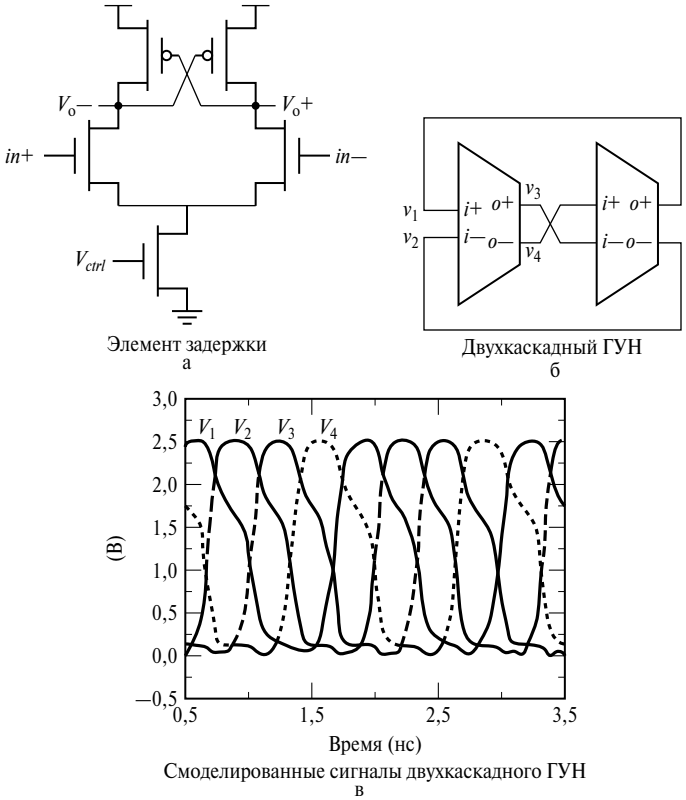


Рис. 7.54. Дифференциальный элемент задержки и топология ГУН

хронизация различных операций, происходящих в сложной схеме, является одной из самых интригующих и сложных задач, с которыми придется столкнуться разработчику следующего десятилетия. Выбор правильной схемы тактирования влияет на функционирование, скорость и возможности цепи.

В этой главе было представлено несколько широко используемых схем тактирования. Наиболее надежной и простой по сути схемой является двухфазный подход, основанный на концепции "master–slave". Наиболее распространенным выбором является использование регистров на основе мультиплексоров и локальная генерация двух фаз путем инвертирования тактового сигнала. Более экзотические схемы, подобные импульсным регистрам, также используются на практике, хотя и требуют тонкой настройки и оправданы только в определенных ситуациях. Например, это может быть потребность в отрицательном времени установки, позволяющая справиться с расфазировкой тактовых импульсов.

Таким образом, общей тенденцией в разработке быстродействующих КМОП-СБИС схем является *использование простых схем тактирования*, даже если это происходит за счет быстродействия. В большинстве методологий автоматизированного проектирования, например, методе стандартных элементов, принят подход с использованием однофазных ячеек, управляемых

фронтом, основанных на статических триггерах. Однако тенденция использования простых схем тактирования прослеживается и в быстродействующих разработках (например, микропроцессорах). Кроме того, типичным является и использование между логическими элементами защелок, позволяющее повысить общее быстродействие схемы.

7.8. Резюме

В данной главе рассматривались последовательные цифровые схемы. Были освещены следующие темы.

- Перекрестное соединение двух инверторов дает *бистабильную* систему, также именуемую *триггером*. Третья потенциальная рабочая точка этой системы метастабильна, т.е. любое отклонение от нее переводит триггер в одно из устойчивых состояний.
- *Защелкой* называется *чувствительный к уровню* запоминающий элемент, который выбирает данные в одной фазе и хранит их в другой. С другой стороны, *регистр* выбирает данные на положительном или отрицательном фронте управляющего сигнала. Регистр характеризуется тремя важными параметрами: временем установки, временем удержания и задержкой распространения. Данные параметры необходимо аккуратно оптимизировать, поскольку часто именно они определяют минимальный возможный такт работы схемы.
- Регистры могут быть *статическими* или *динамическими*. Статический регистр хранит свое состояние, пока на него подается питание. Он идеален для хранения редко используемой информации (например, регистры реконфигурации или память для хранения управляющей информации). Статические регистры либо создаются на основе мультиплексоров, либо запись данных в них производится путем вытеснения старой информации.
- Динамическая память основана на временном хранении заряда на конденсаторах. Основными ее достоинствами являются меньшая сложность, более высокое быстродействие и меньшее потребление мощности. Тем не менее заряд динамического узла со временем утекает, следовательно, для динамических схем определена минимальная возможная тактовая частота. В любом случае чисто динамическая память используется редко. Регистры обычно делают псевдостатическими, защищая их от емкостной связи и других источников помех, наводимых схемой.
- Регистры можно также создавать с использованием *импульсного* подхода. Для выборки входа на фронте тактового импульса применяются специальные короткие импульсы. При создании регистров применяются также схемы на основе усилителей считывания; как правило, их стоит использовать для достижения высокого быстродействия или использования сигналов с неполным размахом.

- Выбор *стиля тактирования* очень важен. Двухфазные схемы могут иметь проблемы с состязанием сигналов. Для решения этих проблем применяются технологии, подобные C^2MOS. Как вариант, можно также использовать истинное однофазное тактирование. Впрочем, в любом случае время нарастания тактовых сигналов необходимо аккуратно оптимизировать, чтобы устранить возможность состязания.
- Объединяя динамическую логику с динамическими защелками, можно получить очень быстрые вычислительные структуры. Один из примеров такого подхода, стиль NORA, очень эффективен в *конвейерных информационных каналах*.
- *Ждущие мультивибраторы* имеют одно устойчивое состояние; следовательно, их можно использовать в качестве генераторов импульсов.
- *Автоколебательные мультивибраторы*, или осцилляторы, не имеют устойчивого состояния. Наиболее известным представителем этого класса является кольцевой генератор.
- Для *триггеров Шмитта* характерно наличие гистерезиса в статической характеристике и быстрых переходов в переходной характеристике. Преимущественно используются для подавления шума.

7.9. Для любознательных

Основные сведения по последовательным логическим элементам можно найти во многих учебниках (см., например, [Mano82] и [Hill74]). Разработка последовательных схем щедро описана в большинстве современных руководств по цифровым схемам. Подробные обзоры вопросов и дизайнерских решений, касающихся быстродействующих последовательных элементов, представлены в [Partovi01] и [Bernstein98].

7.10. Список использованной литературы

[Bernstein98] K. Bernstein et al., *High-Speed CMOS Design Styles*, Kluwer Academic Publishers, 1998.

[Dopperpuhl92] D. Dopperpuhl et al., "A 200 MHz 64-b Dual Issue CMOS Microprocessor", *IEEE Journal of Solid-State Circuits*, Nov 1992, vol. 27, no. 11, p. 1555–1567.

[Gieseke97] B. Gieseke et al., "A 600 MHz Superscalar RISC Microprocessor with Out-of-Order Execution", *IEEE International Solid-State Circuits Conference*, Feb. 1997, p. 176-177.

[Goncalves83] N. Goncalves and H. De Man, "NORA: a racefree dynamic CMOS technique for pipelined logic structures", *IEEE Journal of Solid-State Circuits*, June 1983, vol. SC-18, no. 3, p. 261–266.

[Hill74] F. Hill and G. Peterson, *Introduction to Switching Theory and Logical Design*, Wiley, 1974.

[Jeong87] D. Jeong et al., "Design of PLL-based clock generation circuits", *IEEE Journal of Solid-State Circuits*, April 1987, vol. SC-22, no. 2, p. 255–261.

[Kozu96] S. Kozu et al., "A 100 MHz 0,4 W RISC Processor with 200 MHz Multiply-Adder, using Pulse-Register Technique", *IEEE ISSCC*, February 1996, p. 140-141.

[Mano82] M. Mano, *Computer System Architecture*, Prentice-Hall, 1982.

[Montanaro96] J. Montanaro et al., "A 160-MHz, 32-b, 0,5-W CMOS RISC Microprocessor", *IEEE Journal of Solid State Circuits*, November 1996 p. 1703.

[Mutoh95] S. Mutoh et al., "1-V Power Supply High-Speed Digital Circuit Technology with Multithreshold-Voltage CMOS", *IEEE Journal of Solid State Circuits*, August 1995, p. 847–854.

[Partovi96] H. Partovi, "Flow-Through Latch and Edge-Triggered Flip-Flop Hybrid Elements", *IEEE ISSCC*, February 1996, p. 138-139.

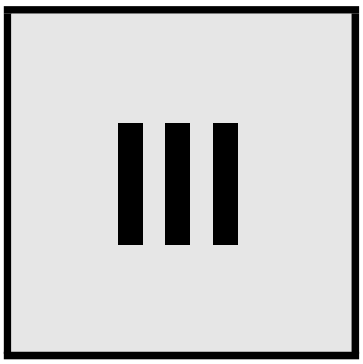
[Partovi01] H. Partovi, "Clocked Storage Elements", in *Design of High-Performance Microprocessor Circuits*, Chandakasan et al., ed., 2001, Ch. 11, p. 207–233.

[Schmitt38] O. H. Schmitt, "A Thermionic Trigger", *Journal of Scientific Instruments*, January 1938, vol. 15, p. 24–26.

[Suzuki73] Y. Suzuki, K. Odagawa, and T. Abe, "Clocked CMOS calculator circuitry", *IEEE Journal of Solid State Circuits*, December 1973, p. 462–469.

[Yuan89] J. Yuan and Svensson C., "High-Speed CMOS Circuit Technique", *IEEE JSSC*, February 1989, vol. 24, no. 1, p. 62–70.

ЧАСТЬ



Перспектива системы

По моему мнению, искусство должно упрощать. Действительно, это очень близко к результату высокохудожественного процесса; найти, без какого набора форм и деталей может обойтись конечный результат, при этом сохранив его дух.

Уилла Сиберт Кэсер (Willa Sibert Cather),

On the Art of Fiction (1920)

Простота и уравновешенность — вот чем измеряется истинная ценность произведения искусства.

Френк Ллойд Райт (Frank Lloyd Wright)

ГЛАВА

8

Стратегии реализации цифровых интегральных схем

Полуиндивидуальная и структурированная методология проектирования

Специализированные интегральные схемы и "системы на кристалле"

Настраиваемое аппаратное обеспечение

8.1. Введение
8.2. От индивидуального подхода к проектированию к полуиндивидуальному и структурированным массивам
8.3. Индивидуальная разработка
8.4. Методология проектирования на основе элементов
8.4.1. Стандартный элемент
8.4.2. Скомпилированные элементы
8.4.3. Макроэлементы, мегаэлементы и интеллектуальная собственность
8.4.4. Полуиндивидуальный процесс проектирования
8.5. Подходы к реализации, основанные на матрицах
8.5.1. Матрицы с предварительной диффузией (или программируемым фотошаблоном)
8.5.2. Матрицы с предварительным монтажом
8.6. Перспектива — платформа реализации будущего
8.7. Резюме
8.8. Для любознательных

Рис. 8.1. Зазор производительности проектирования. Развитие технологии (рост числа транзисторов на кристалл) опережает производительность разработки (число транзисторов, спроектированных одним разработчиком за месяц). *Источник: SIA [SIA97]*

8.1. Введение

Из-за невероятного роста сложности современных интегральных схем задача их проектирования также является весьма нетривиальной. Разработать схему, содержащую несколько миллионов транзисторов, и обеспечить ее корректную работу настолько сложно, что без помощи компьютера и четких методологий проектирования это сделать практически невозможно. Фактически развитие технологии часто опережает полосу поглощения сообщества разработчиков. Сказанное подтверждает рис. 8.1, где показано, что сложность интегральных схем (число транзисторов) растет быстрее, чем производительность разработчика, создавая “технологический зазор”. Одним из способов преодоления этого разрыва является постоянное увеличение размеров команд, работающих над одним проектом. Данная тенденция отчетливо прослеживается в сфере разработки быстродействующих процессоров, где команда в 500 и более разработчиков уже давно никого не удивляет.

Разумеется, с точки зрения долгосрочной перспективы данный подход нельзя назвать удовлетворительным — только представьте, что все инженеры мира начали работать над одним проектом. К счастью, примерно раз в десятилетие мы наблюдаем внедрение новаторской методологии проектирования, которая дает мгновенный скачок в функции производительности проектирования, помогая временно заполнить “брешь производительности”. Изучая данные, накопившиеся за минувшие сорок лет, мы можем выделить несколько таких скачков производительности. На начальном этапе развития интегральных схем в 1970-х нормой было чистое индивидуальное проектирование. Позже отображение функции в кремние стало облегчаться за счет программируемых логических матриц (Programmable Logic Array — PLA), стандартных ячеек, макроэлементов, компиляторов модулей, логических матриц и интегральных схем с изменяемой структурой. В этой главе мы опишем несколько

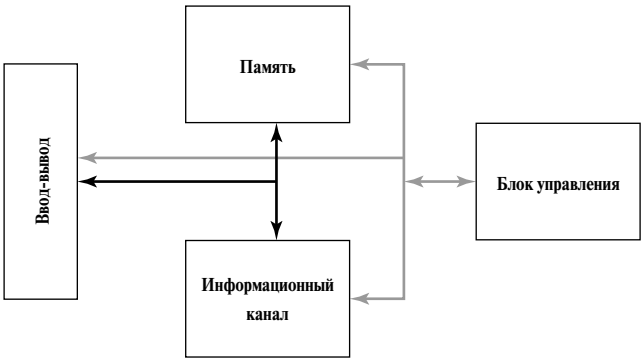


Рис. 8.2. Схемы типичного цифрового процессора. Стрелками показаны возможные межсоединения

широко распространенных подходов к реализации проектов. Из-за экстенсивной природы области мы в принципе не сможем охватить все аспекты в одной главе (на это потребовалась бы отдельная книга). Поэтому мы рассмотрим данную тему с точки зрения пользователя, чтобы понять, что предлагают различные методологии проектирования и чего от них можно ожидать.

Выбор способа отображения функции в полупроводниковую интегральную схему очень сильно зависит от самой функции. Рассмотрим, например, простой цифровой процессор, изображенный на рис. 8.2. Подобный процессор может быть мозгом персонального компьютера или сердцем проигрывателя компакт-дисков либо сотового телефона. Он состоит из нескольких компоновочных блоков, в той или иной форме присутствующих практически во всех цифровых процессорах.

- Ядром процессора является *информационный канал*; именно в нем происходят все вычисления. Другие блоки процессора являются вспомогательными: они либо хранят результаты, которые дает информационный канал, либо помогают определить, что произойдет в следующем цикле. Типичный информационный канал состоит из набора стандартных комбинационных функций, выполняющих логические (И, ИЛИ, исключающее ИЛИ) или арифметические операции (сложение, умножение, сравнение, сдвиг). Промежуточные результаты хранятся в регистрах. В общем, существуют различные стратегии реализации информационных каналов — структурированные индивидуальные ячейки либо автоматизированные стандартные ячейки, жесткий монтаж либо программируемые логические схемы. На выбор платформы реализации влияют преимущественно компромиссы между различными метриками проекта — площадь, скорость, энергия, время разработки и возможность повторного использования.
- *Блок управления* определяет, какие действия производятся в процессоре в любой заданный момент времени. Контроллер представляет собой подобие конечного автомата. Он состоит из регистров и логики, а следовательно, является последовательной схемой. Логика может реализовы-

ваться различными способами — либо в виде соединений основных логических элементов (стандартных ячеек), либо более структурировано — с использованием программируемых логических матриц и памяти для хранения команд.

- *Модуль памяти* используется как централизованное место хранения данных. В настоящее время существует огромное множество различных классов ячеек памяти. Основные различия этих классов заключаются в способе доступа к данным: "только чтение" или "чтение–запись", последовательный или произвольный доступ, одно- или многопортовый доступ. Другой критерий разделения ячеек памяти на классы связан с их возможностями хранения данных. Динамические запоминающие структуры должны периодически обновляться, а статические хранят информацию, пока на них подается питание. Наконец, энергонезависимая память (например, флэш-память) хранит информацию даже при выключении источника питания. Отметим, что в одном процессоре может использоваться несколько классов запоминающих устройств. Например, для хранения данных может применяться оперативное запоминающее устройство (ОЗУ), а для хранения команд — постоянное запоминающее устройство (ПЗУ).
- *Соединительная сеть* связывает между собой различные модули процессора, а *схема входа-выхода* соединяет процессор с внешним миром. Довольно долгое время межсоединения были пережитком процесса разработки. К сожалению, провода, образующие соединительную сеть, никак нельзя назвать идеальными, поскольку они вводят емкостную, резистивную и индуктивную нагрузку на питающую схему. По мере роста размеров кристаллов также увеличивается длина соединительных проводов, что приводит к повышению данных паразитных нагрузок. В настоящее время разрабатываются методологии автоматизированного или структурированного проектирования, которые облегчают развертывание подобных соединительных структур. В качестве примеров можно привести *внутрикристальные шины*, *решетчатые соединительные структуры* и даже целые *сети на кристалле*. Некоторые компоненты соединительной сети обычно рассматриваются отдельно от блок-схем, подобных показанной на рис. 8.2, поскольку являются критично важными для работы схемы. В частности, это относится к сетям питания и синхронизации. Правильное планирование этих "обслуживающих" сетей может потребовать много времени, однако оно позволяет добиться правильной работы интегральной схемы.

Структура, приведенная на рис. 8.2, может многократно повторяться на одном кристалле. Для примера на рис. 8.3 показана система, реализованная на одном кристалле и включающая все функции, требуемые для реализации цифрового телевизора высокой четкости. Здесь представлены два процессора, блоки памяти, специализированные ускорители таких функций, как кодирование/декодирование MPEG и фильтрация данных, а также ряд периферийных устройств. В других приложениях, например беспроводных трансиверах или блоках чтения/записи на жесткий диск, могут использоваться даже несколько масштабируемым аналоговых модулей.

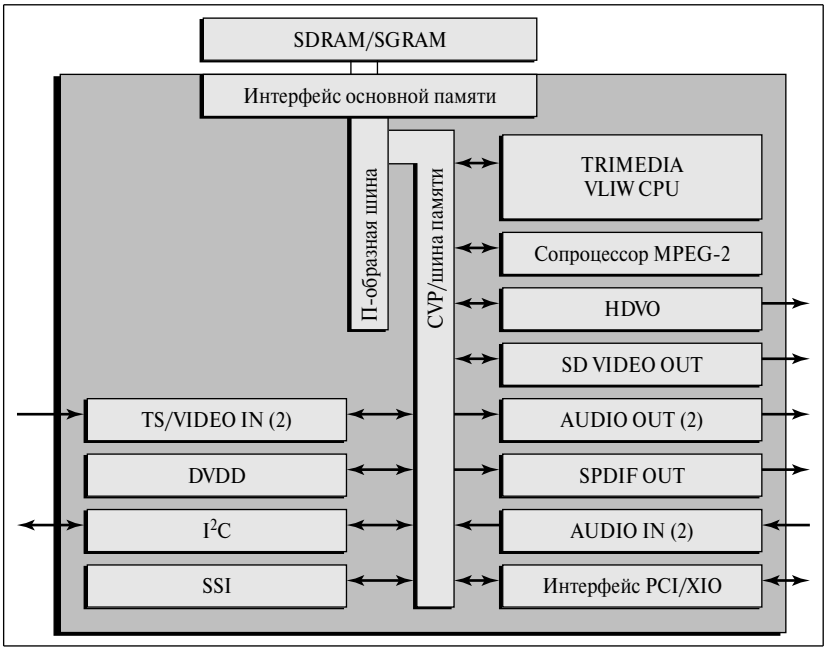


Рис. 8.3. Система “Nexperia” [Philips99]. Данный простой кристалл объединяет универсальное микропроцессорное ядро, процессор сигналов VLIW (very large instruction word — с командными словами сверхбольшой длины), запоминающую систему, сопроцессор MPEG, несколько ускорителей, периферийные устройства ввода-вывода, а также две системные шины

Выбор эффективного подхода к реализации сильно зависит от функции рассматриваемых модулей. Например, блоки памяти обычно имеют правильную и строгую структуру. Следовательно, в этом случае лучше всего будет использовать модульный компилятор, выстраивающий ячейки в виде массива. С другой стороны, контроллеры обычно являются неструктурированными, поэтому желательны другие подходы к реализации. Выбор стратегии реализации может очень сильно сказаться на качестве конечного продукта. Задача разработчика — выбрать стиль, удовлетворяющий спецификациям и ограничениям конкретного продукта. То, что хорошо для одного проекта, может абсолютно не подходить для другого.

Пример 8.1. Компромисс между эффективностью использования энергии и гибкостью

Проект, предполагающий гибкость (или возможность программирования), очень привлекателен с точки зрения приложения. Он допускает динамическое связывание, т.е. изменение приложения после производства кристалла. Гибкость разрешает повторно использовать один проект в нескольких приложениях или обновлять микропрограммное обеспечение компонента на месте, уменьшая риск для производителя. Если же компонент реализован жестко, все его параметры полностью определены во время производства, и после этого модифицировать их нельзя никак.

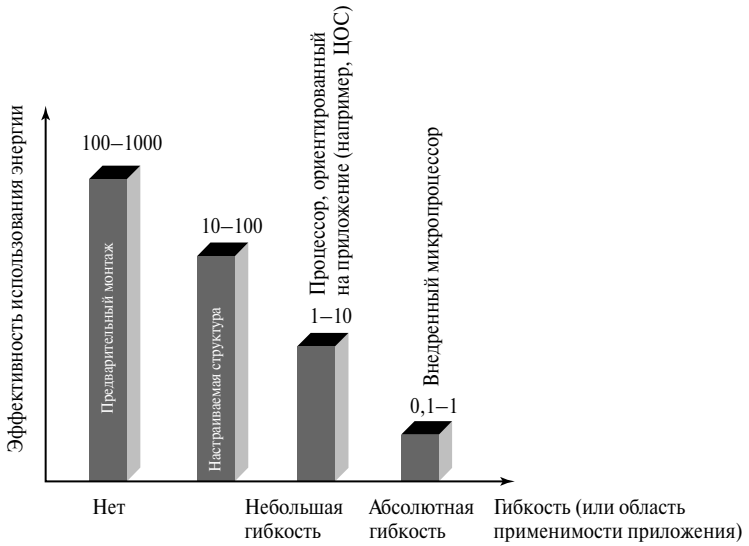


Рис. 8.4. Компромисс между гибкостью и эффективностью использования энергии (представленный в миллионах операций в секунду на милливатт или миллионах операций на миллиджоуль энергии) для различных стилей реализации. Числа представлены для КМОП-процесса 0,25 мкм [Rabaey00]

В связи с этим возникает вопрос: почему гибкие (или программируемые) компоненты не используются во всех разработках? Как всегда — потому что даром ничего не дается. Ценой гибкости является сниженное быстродействие и эффективность использования энергии. Возможность программирования означает дополнительные служебные издержки реализации. Например, чтобы один информационный канал программируемого процессора мог выполнять несколько функций, процессор должен хранить как инструкции, так и декодер инструкций. Большинство разработчиков не знают о большой цене гибкой, поэтому на рис. 8.4 представлена *эффективность использования энергии* (число операций, которые можно выполнить при выделенной энергии) различных стилей реализации схем в зависимости от их *гибкости* (т.е. количества приложений, в которых можно использовать эти схемы). Мы видим, что разница составляет *три порядка*! Очевидно, что когда эффективность использования энергии имеет первостепенное значение, предпочтительна жесткая реализация или ограниченная гибкость (например, настраиваемые или параметризуемые модули).

В данной и трех последующих главах мы обсудим соответственно техники реализации произвольной логики и контроллеров (данная глава), межсоединений (глава 9), информационных каналов (глава 11), и запоминающих устройств (глава 12). Обратите внимание на то, что выбор стратегии реализации может значительно повлиять на качество конечного продукта. Отдельное внимание при разработке сложных схем, состоящих из большого числа блоков, необходимо обратить на такие вопросы, как синхронизация и тактирова-

ние (глава 10), а также проектирование сети питания (глава 9). Распределение тактовых сигналов и питания является одной из важнейших проблем разработки современных процессоров. В дополнениях, которыми перемежаются главы этой книги, рассмотрены многие аспекты разработки данных сложных компонентов и представляются современные средства автоматизации проектирования, которыми может пользоваться разработчик. В дополнениях Е–З обсуждаются синтез дизайна, верификация и тестирование соответственно.

8.2. От индивидуального подхода к проектированию к полуиндивидуальному и структурированным массивам

Жизнеспособность микроэлектронного проекта зависит от нескольких (часто противоречащих друг другу) факторов: быстродействия (скорости или потребляемой мощности), цены и объема производства. Например, конкурентоспособный процессор должен иметь хорошее быстродействие и низкую цену. Достичь этого можно только за счет больших объемов продаж. Благодаря этому высокая стоимость производства (сопутствующая быстродействующей разработка) амортизируется большим числом составляющих. Другой сценарий представляют приложения, подобные разработке суперкомпьютеров и некоторых приложений защиты. Помимо того, что основной целью является максимально возможное быстродействие, часто желательны еще и быстродействующие заказные технологии. Объем производства в этом случае мал, однако стоимость электронных компонентов составляет только часть общей стоимости системы, а следовательно, не имеет первостепенной важности. Наконец, в большинстве заказных приложений основной задачей является уменьшение размера системы посредством интеграции. В подобных случаях стоимость разработки можно существенно уменьшить за счет использования технологий автоматизации проектирования, которые не дают максимального быстродействия, но минимизируют время, требуемое на разработку схемы. Как отмечалось в главе 1, стоимость полупроводникового устройства представляет собой сумму двух компонентов.

- *Единовременные затраты на проектирование.*
- *Стоимость производства одного устройства*, которая зависит от сложности процесса, площади разработки и используемого процесса.

Данные экономические соображения инициировали создание множества различных подходов к реализации — от быстродействующих ручных разработок до полностью программируемых разработок, дающих низкое или среднее быстродействие. Краткий обзор существующих методологий представлен на рис. 8.5. В следующем разделе мы вначале обсудим методологию индивидуального проектирования, а затем перейдем к полуиндивидуальным разработкам и подходам, основанным на матрицах.

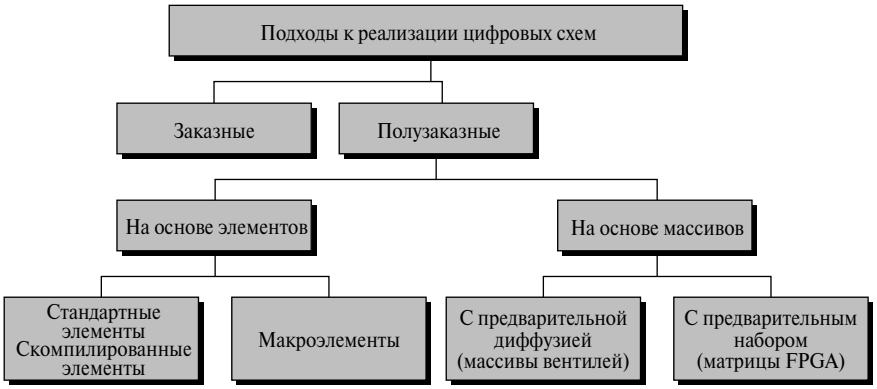


Рис. 8.5. Обзор подходов к реализации цифровых интегральных схем (согласно [DeMicheli94])

8.3. Индивидуальная разработка

Когда основным критерием разработки является быстродействие или плотность монтажа схемы, единственным разумным выбором кажется проектирование топологии схемы вручную. Действительно, такой подход был единственно возможным в первые годы цифровой микроэлектроники, что явно видно на структуре микропроцессора Intel 4004 (см. рис. 8.5, *а*). Трудоемкая природа индивидуальной разработки приводит к высокой цене и большому *времени выхода на рынок*. Следовательно, экономически ее можно оправдать только при выполнении следующих условий.

- Заказной блок можно повторно использовать (например, как библиотечный элемент).
- Стоимость можно амортизировать большим тиражом (например, речь идет о микропроцессорах или полупроводниковых ячейках памяти).
- Главный критерий разработки — это не цена (например, проектируется суперкомпьютер или гиперсуперкомпьютер).

Вследствие постоянного прогресса в сфере автоматизации проектирования доля индивидуальных разработок в общем количестве проектов постоянно снижается. Даже в наиболее современных быстродействующих микропроцессорах, подобных Intel Pentium® 4 (рис. 8.6), практически все фрагменты были спроектированы автоматически с использованием полуиндивидуальных подходов. Вручную разрабатывались только модули, имеющие наибольшее влияние на быстродействие (схемы фазовой автоподстройки частоты и буферы синхроимпульсов). Фактически единственной областью, в которой до сих пор процветает индивидуальная разработка, является проектирование библиотечных элементов.

Степень автоматизации разработки в индивидуальном процессе минимальна, хотя некоторые инструменты проектирования доказали здесь свою

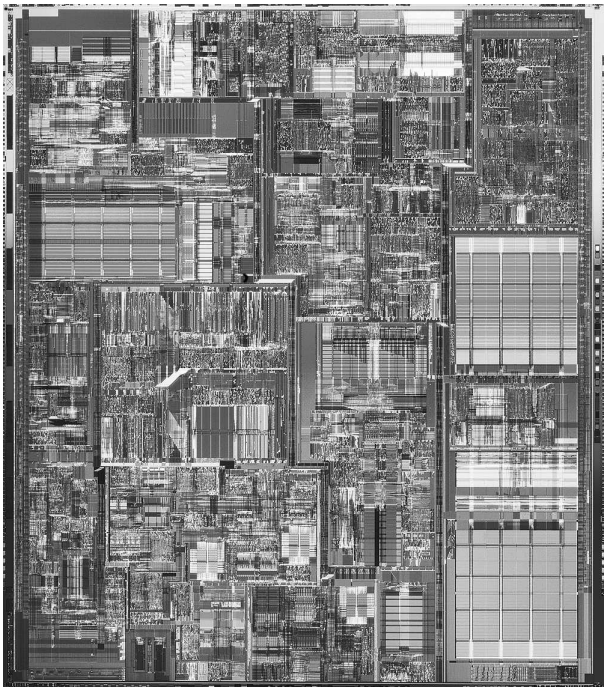


Рис. 8.6. Микрофотография кристалла процессора Intel Pentium®4. Кристалл содержит 42 миллиона транзисторов, разработанных по 0,18-микронному КМОП-процессу. Первое поколение данных процессоров работало с тактовой частотой 1,5 ГГц (*Перепечатано с разрешения Intel Corp.*)

незаменимость. Объединяя их с обширным набором средств верификации и моделирования, редакторами топологии, программами контроля проектных норм и электрических параметров (см. вставку А) получим основу типичной среды индивидуального проектирования. Если вы интересуетесь этой темой, прекрасное описание возможностей и задач индивидуального проектирования можно найти в [Grundman97].

8.4. Методология проектирования на основе элементов

Поскольку индивидуальная разработка оказывается чрезмерно дорогой, было разработано множество подходов, сокращающих и автоматизирующих процесс разработки. Ценой этой автоматизации является меньшая степень интеграции и/или быстродействие. Похоже, справедливым оказывается следующее правило: **чем меньше время проектирования, тем дороже оно обходится**. В данном разделе мы обсудим несколько подходов к проектированию, которые для каждой новой структуры по-прежнему требуют полного прохода по производственному процессу. В следующем разделе рассмотрен

подход, основанный на матрицах, который позволяет сократить время разработки и уменьшить цену, требуя при этом небольшого числа дополнительных этапов обработки или вообще не требуя их.

Смысл проектирования на основе ячеек состоит в уменьшении трудозатрат на реализацию путем повторного использования конечной библиотеки элементов. Достоинством этого подхода является то, что базовые элементы для данной технологии требуется спроектировать один раз, после чего их можно многократно использовать в последующих разработках. Недостаток же подхода заключается в том, что ограниченная природа библиотечного элемента снижает возможность тонкой настройки структуры. Подходы, основанные на элементах, можно условно разделить на несколько классов в зависимости от степени детализации модулей.

8.4.1. Стандартный элемент

Использование стандартных элементов подразумевает стандартизацию разработки на уровне логического вентиля. В этом случае библиотека содержит большой набор логических элементов с различными разветвлениями по входу и выходу. Помимо основных логических функций (инвертор, И/НЕ-И, ИЛИ/НЕ-ИЛИ, ИСКЛЮЧАЮЩЕЕ ИЛИ/ИСКЛЮЧАЮЩЕЕ НЕ-ИЛИ, триггеры), типичная библиотека содержит более сложные функции — И-ИЛИ-НЕ, мультиплексор, полный сумматор, компаратор, счетчик, кодеры и декодеры. Проект представляется в виде схемы, содержащей только элементы, присутствующие в библиотеке, или генерируется автоматически согласно командам высокоуровневого описательного языка. Подобная высокая степень автоматизации возможна благодаря наложению строгих ограничений на схему. При использовании этого подхода элементы располагаются в рядах, разделенных трассировочными каналами, как показано на рис. 8.7. Эффективность схемы повышается, когда все библиотечные элементы имеют одинаковую высоту. Ширина элементов может меняться согласно их сложности. Как показано на схеме, технологию стандартных элементов можно комбинировать с другими подходами, вставляя в структуру дополнительные модули (например, ячейки памяти и умножители), которые нельзя легко или эффективно реализовать с помощью парадигмы стандартных логических элементов.

На рис. 8.8, *а* приведен пример структуры, реализованной с использованием раннего стиля стандартных элементов. Значительная часть площади отведена на трассировку сигналов. Таким образом, видно, что важнейшей задачей инструментов размещения и связывания стандартных элементов является минимизация служебных издержек, связанных с межсоединениями. Один из подходов, направленных на минимизацию длин проводников, заключается в введении проходных ячеек (см. рис. 8.7), что позволяет соединять элементы, принадлежащие различным рядам, без прокладывания тракта в обход всего ряда. Еще сильнее сократить объем проводки можно за счет введения большего числа уровней межсоединений. Современные КМОП-процессы допускают использование семи или даже большего числа металлических уровней, что позволяет практически устранить потребность в трассировочных кана-

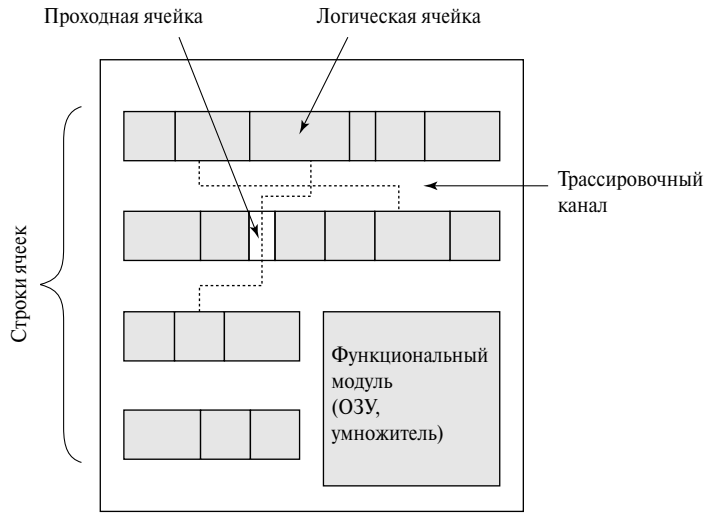


Рис. 8.7. Схема использования стандартных элементов

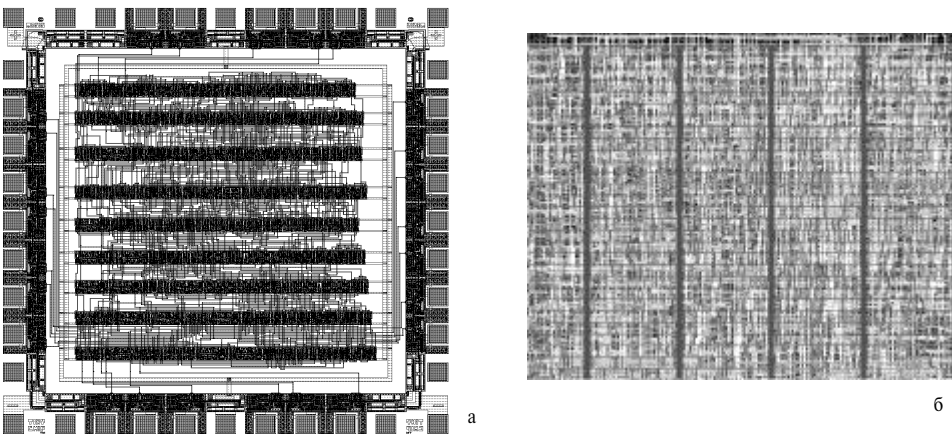


Рис. 8.8. Эволюция разработок на основе стандартных элементов: *а* — структура, реализованная по трехуровневой технологии; значительную долю кристалла занимают трассировочные каналы; *б* — проект, реализованный по семиуровневой технологии; трассировочные каналы практически исчезли, все межсоединения проходят поверх логических элементов

лах. Практически все сигналы можно провести над элементами, т.е. возможна действительно трехмерная схема. Для примера на рис. 8.8, *б* показан фрагмент структуры из стандартных элементов, реализованной с использованием семи металлических уровней. Плотность структуры составляет более 90%, т.е. практически весь кристалл заполнен логическими элементами, и лишь небольшая его доля используется для соединений.

Проектирование библиотеки стандартных элементов является трудоемким процессом, который, к счастью, можно амортизировать большим числом

разработок. В связи с этим разработчику приходится решать нетривиальную задачу определения составляющих библиотеки. Например, необходимо определить, что лучше: небольшая библиотека, большинство элементов в которой имеет ограниченное разветвление по входу, или большая библиотека с множеством версий одного и того же логического элемента (например, наличие элементов НЕ-И с двумя, тремя и четырьмя входами, каждый из которых представлен несколькими разновидностями разного размера)? Поскольку разветвление по выходу и емкость нагрузки, порождаемые проводкой, заранее неизвестны, принято, чтобы каждый элемент имел хороший диапазон возможностей питания (т.е. включал большие выходные транзисторы). Такое решение упрощает процедуру разработки, но отрицательно сказывается на площади реализации и потребляемой мощности. В современных библиотеках реализуется множество разновидностей каждого элемента, размеры которых выбираются с учетом различных питающих напряжений, а также быстродействия и уровня потребляемой мощности. Задача разработчика — задать требуемые скорость и площадь реализации, по которым пакет подберет требуемые элементы нужного размера.

Чтобы можно было использовать библиотечный подход, необходима подробнейшая документация по библиотеке элементов. Информация должна включать не только схему, описание функциональных возможностей и расположения выходов, но и точно характеризовать задержку и потребление мощности элемента в зависимости от емкости нагрузки и входных времен нарастания и спада сигнала. Получение этой информации занимает значительную часть времени генерации всей библиотеки. Подробно об описании логических и последовательных элементов рассказано в дополнении Д.

Пример 8.2. Стандартный элемент НЕ-И с тремя входами

Чтобы проиллюстрировать некоторые из представленных выше наблюдений, рассмотрим разработку стандартного логического элемента НЕ-И с тремя входами, реализованного по 0,18-микронной КМОП-технологии (рис. 8.9). В действительности библиотека содержит пять версий элемента, поддерживающих емкости нагрузки от 0,18 до 0,72 пФ и имеющих площадь от 16,4 до 32,8 мкм^2. Показанный элемент является самым медленным энергосберегающим вариантом из этой группы, в котором для уменьшения утечки применяются транзисторы с высоким порогом. Отношение размеров W/L n-МОП- и p-МОП-транзисторов разгрузочной (нагрузочной) сети приблизительно равно 8.

Обратите внимание на то, как стратегия разработки топологии согласуется с подходом, представленным на рис. Г.2. Линии питания распределены горизонтально и совместно используются элементами одного ряда. Входные сигналы передаются вертикально по поликремнию. Входы/выходы расположены в теле ячейки (это видно по штырьковому выводу на схеме), что согласуется со стандартной на сегодняшний день методологией “проводки над элементом”.

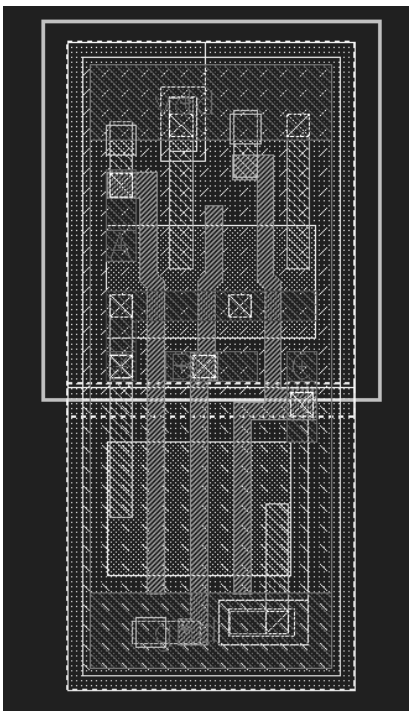


Рис. 8.9. Стандартный элемент НЕ-И с тремя входами.
(*Перепечатано с разрешения ST Microelectronics.*)

Использование библиотеки стандартных элементов весьма популярно и используется для реализации практически всех логических элементов в современных интегральных схемах. Единственными исключениями являются проекты, в которых требуется максимальное быстродействие или минимальное энергопотребление, или целевая функции имеет очень правильную структуру (например, память или умножитель). Успех подхода с использованием стандартных элементов стал возможен благодаря ряду событий, некоторые из которых описаны ниже.

- **Повышение качества инструментов автоматического размещения элементов и трассировки**, а также доступность нескольких уровней трассировки. Фактически во многих исследованиях было показано, что в современных условиях при создании сложных несимметричных логических схем автоматизированный подход успешно соперничает с ручной разработкой или дает существенно лучшие результаты. В прошлом ситуация была иной, и автоматизированному проектированию сопутствовали существенные служебные издержки.
- **Появление сложных инструментов синтеза логических схем**. Возможность синтеза логических схем позволяет рассматривать проектирование на высоком уровне абстракции с использованием булевых уравнений,

конечных автоматов или языков моделирования на уровне регистровых передач, подобных VHDL или Verilog. Инструменты синтеза автоматически транслируют эту спецификацию в таблицу соединений логических элементов, минимизируя указанную функцию стоимости — площадь, задержку или потребляемую мощность. Первые инструменты синтеза, использовавшиеся в начале 1980-х, фокусировались в основном на двухуровневой логике. Это хотя и позволяло в какой-то мере автоматизировать разработку, но в то же время ограничивало эффективность использования площади и быстродействие генерируемых схем. Реально полезными такие средства автоматизированной генерации стали с появлением возможности *синтеза многоуровневой логики* в конце 1980-х. В наши дни практически никто из разработчиков не использует библиотеки стандартных элементов без обращения к автоматическому синтезу. Более подробное описание процесса синтеза проекта приводится во вставке E.

Историческая перспектива: программируемая логическая матрица

На заре проектирования МОП ИС разработка (и оптимизация) логических элементов была ручной трудоемкой работой. Предпочтительным инструментом в то время были карты Карно и таблицы Куина–Мак-Класки. В конце 1970-х был сделал первый шаг к автоматизации утомительного процесса разработки логических схем, инициированный следующими важными событиями.

- Вместо того чтобы использовать для разработки каждой новой схемы индивидуальный подход, была придумана правильная структура, названная программируемой логической матрицей, ПЛМ (Programmable Logic Array — PLA). Эта методология позволяет автоматически генерировать двухуровневые логические схемы и, что более важно, результат этой операции предсказуем с точки зрения площади и быстродействия.
- С появлением автоматизированных средств синтеза логических элементов [Brayton84] стала возможной трансляция любого возможного булева выражения в оптимизированную двухуровневую (сумма произведений или произведение сумм) логическую структуру. Вскоре после этого появились и инструменты синтеза последовательных схем. Идея структурированной разработки быстро получила распространение и уже в середине 1980-х годов была взята на вооружение основными компаниями-разработчиками микропроцессоров, включая Intel и DEC. Хотя ПЛМ редко используются в современных полуиндивидуальных разработках, данная тема заслуживает отдельного внимания.

Новую концепцию проще всего объяснить на примере. Рассмотрим приведенные ниже логические функции, уравнения которых мы преобразовали в формат “сумма произведений”.

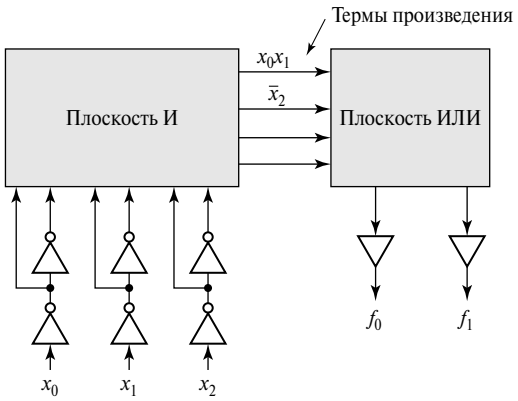


Рис. 8.10. Обычная двухуровневая реализация булевых функций

$$f_0 = x_0x_1 + \bar{x}_2; \qquad (8.1)$$
$$f_1 = x_0x_1x_2 + \bar{x}_2 + \bar{x}_0x_1.$$

Одно из важных достоинств данного представления заключается в том, что правильную реализацию легко себе представить (рис. 8.10). В первом слое элементов реализуются операции И, а во втором слое — функции ИЛИ. Таким образом, ПЛМ — это обычная прямоугольная макроячейка, состоящая из массива транзисторов, выровненного так, что строки соответствуют множителям, а столбцы — входам и выходам. Столбцы входов и выходов разбивают матрицу на два подмассива, называемых плоскостями И и ИЛИ соответственно. Схему, изображенную на рис. 8.10, нельзя реализовать "в лоб", поскольку одноуровневые логические функции в КМОП всегда являются инвертирующими. С помощью простых булевых манипуляций уравнение (8.1) можно легко преобразовать в формат НЕ-ИЛИ/НЕ-ИЛИ.

$$\bar{f}_0 = \overline{\overline{(\bar{x}_0 + \bar{x}_1)} + \bar{x}_2}; \qquad (8.2)$$
$$\bar{f}_1 = \overline{\overline{(\bar{x}_0 + \bar{x}_1 + \bar{x}_2)} + \bar{x}_2 + \overline{(x_0 + \bar{x}_1)}}.$$

Задача 8.1. Двухуровневое представление логики

Уравнение (8.1) можно представить и в формате НЕ-И/НЕ-И. Вообще, представление НЕ-ИЛИ/НЕ-ИЛИ предпочтительнее, поскольку элементы НЕ-И с большим разветвлением по входу характеризуются медленной работой. Тем не менее конфигурация НЕ-И/НЕ-И является очень плотной, а это можно использовать для снижения потребления мощности. Выведите представление НЕ-И/НЕ-И для уравнения (8.2).

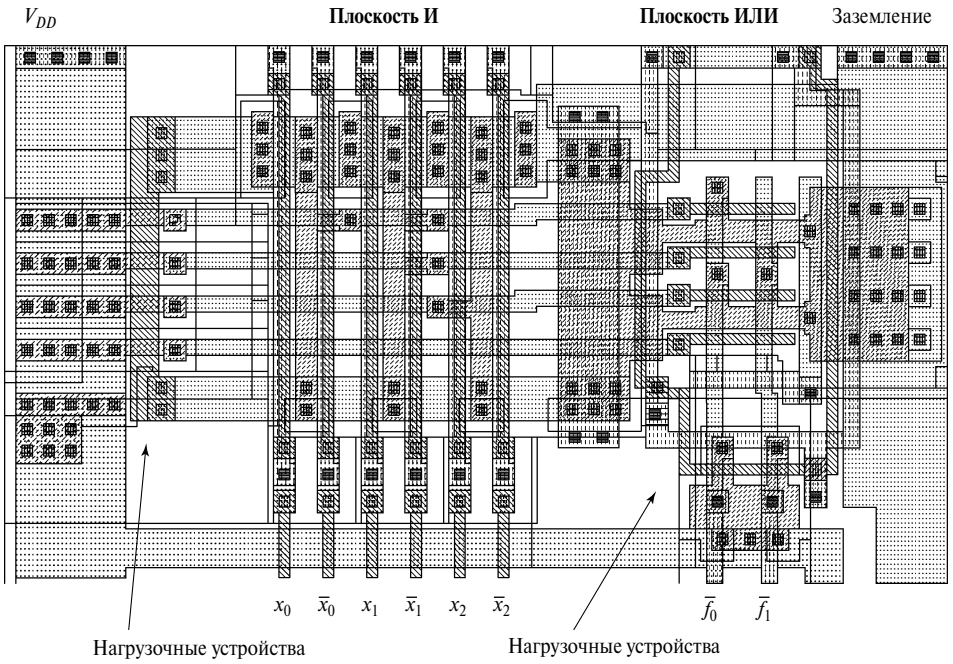


Рис. 8.11. ПЛМ-схема, реализующая уравнение (8.2)

Таким образом, трансляция набора двухуровневых логических функций в физическую структуру сводится к задаче "программирования" — т.е. выбору расположения транзисторов на плоскостях И и ИЛИ. Эту задачу очень просто автоматизировать, именно этим и объясняется первоначальный успех ПЛМ. Например, на рис. 8.11 показана автоматически сгенерированная ПЛМ-реализация логических функций, описываемых уравнением (8.2). К сожалению, правильная структура, хотя и является предсказуемой, несет очень много служебных издержек с точки зрения площади и задержки (это видно и из схемы), что совершенно недопустимо в мире полуиндивидуального проектирования. Если вам все же интересно, как на самом деле реализуются плоскости И и ИЛИ, обратитесь к главе 12, где обсуждается реализация ПЛМ на уровне транзисторов.

8.4.2. Скомпилированные элементы

Стоимость реализации и описания библиотеки элементов не следует недооценивать. Современные библиотеки содержат от 700 до 1000 и даже более элементов. Данные элементы необходимо перепроектировать при каждом переходе на новую технологию. Более того, изменения происходят и в процессе разработки одного поколения технологии. Например, для повышения выхода готовых изделий часто меняется минимальная толщина металлического слоя

или нормы, касающиеся контактов. В результате всю библиотеку необходимо снова пересмотреть и описать. Кроме того, даже объемная библиотека является дискретной, т.е. число вариантов структур, которые можно создать с ее помощью, ограничено. Если основным критерием разработки является быстродействие или мощность, необходимы заказные элементы с оптимизированными размерами транзисторов. Поскольку доля служебных издержек, вводимых межсоединениями, становится в проектах все больше, с точки зрения быстродействия и потребляемой мощности возникает потребность в элементах с настраиваемыми мощностями питания [Sylvester98], и в этом состоит задача автоматизированной (или компилируемой) генерации элементов.

Существует множество автоматизированных подходов, позволяющих на лету генерировать схемы элементов по данным таблицам соединений транзисторов, однако получить с их помощью высококачественную автоматическую схему невозможно. Ранние подходы отталкивались от фиксированных топологий. Более поздние допускали бо́льшую свободу в расположении транзисторов (например, [Hill85]). В настоящее время существует несколько коммерческих средств генерации элементов, позволяющих получать схемы с такой плотностью, которую может дать разработчик ([Cadabra01, Prolific01]).

Пример 8.3. Автоматизированная генерация элементов

Проиллюстрируем ход типичного процесса генерации элемента на примере простого инвертора (используется средство Abracad [Cadabra01]).

- Вначале разрабатывается схема элемента. Отправной точкой процесса автоматизированной генерации структуры является таблица соединений Spice. Генератор изучает эту таблицу и начинает работать с геометрией транзисторов. В нашем случае (КМОП-инвертор) элемент содержит всего два транзистора (рис. 8.12, *а*).
- Инструмент работает по той же схеме, что и человек. Вначале транзисторы располагаются в архитектуре элемента согласно предопределенным топологическим правилам (см. рис. 8.12, *б*). Данная архитектура является общей для всех элементов библиотеки и задает их высоту, шины электропитания, расположение контактов, трассировку и стили организации контактов.
- Между точками схемы символически проводятся соединения (рис. 8.12, *в*).
- Трассировка переупорядочивается, и ячейка делается более компактной согласно нормам проектирования и настройкам библиотеки (см. рис. 8.12, *г*).
- На последнем этапе ячейка очищается от ошибок и получается окончательная схема (см. рис. 8.12, *д*).

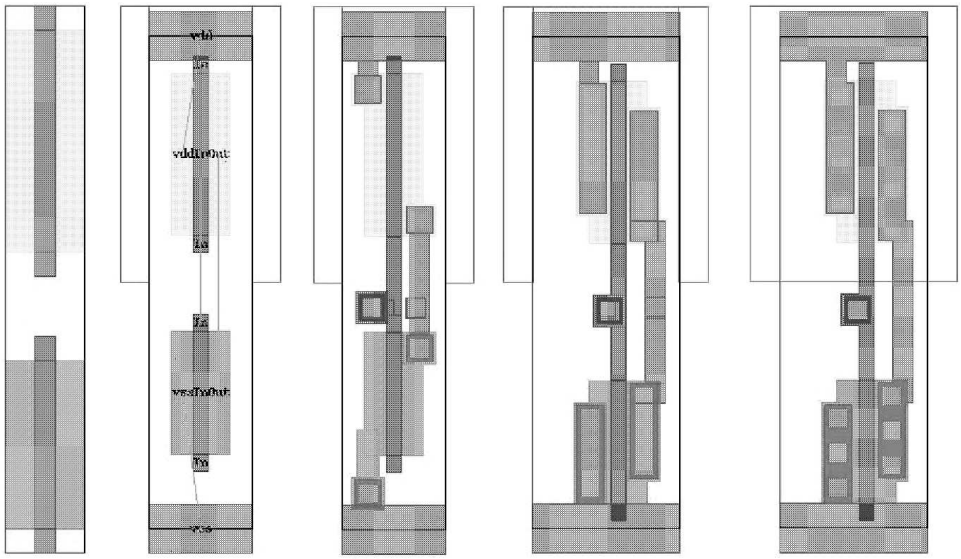


Рис. 8.12. Автоматическая генерация схемы: *а* — исходная геометрия транзисторов; *б* — размещение транзисторов с указанием соединений; *в* — ячейка с соединениями; *г* — компактная ячейка; *д* — окончательный вариант

8.4.3. Макроэлементы, мегаэлементы и интеллектуальная собственность

Стандартизация на уровне логических элементов кажется привлекательной, когда требуется реализовывать произвольные логические функции, однако они оказываются неэффективными для более сложных структур — умножителей, информационных трактов, запоминающих устройств, внедренных микропроцессоров и цифровых процессоров. Учитывая специфическую природу данных блоков, можно получить реализации, демонстрирующие намного лучшие результаты, чем позволяют стандартные интегральные схемы. Элементы, сложность которых выше, чем у составляющих типичной библиотеки стандартных элементов, называются *макроячейками* (а иногда *мегаячейками*). Макроячейки можно условно разделить на два типа — *жесткие* и *мягкие*.

Жесткие макроячейки представляют модуль с данными функциональными возможностями и предопределенной физической структурой. Относительное положение транзисторов и проводки внутри модуля фиксировано. По сути, жесткие макроячейки — это заказная реализация заданной функции. В некоторых случаях ячейка параметризована, т.е. существуют версии с несколько отличными свойствами (либо же эти версии можно при необходимости сгенерировать). Примерами макроячеек являются запоминающие устройства и умножители: макроэлемент “жесткий умножитель” позволяет сгенерировать не только умножитель 32×16, но и умножитель 8×8.

Плюс жесткой макроячейки заключается в наличии всех достоинств индивидуального проекта: плотной схемы, оптимизированного и предсказуемого быстродействия, хороших характеристик с точки зрения рассеяния мощности. Замкнув функцию в макромодуль, мы сможем многократно использовать ее впоследствии в других проектах (это помогает быстрее окупить средства, вложенные в первоначальный проект). Минусом жесткой макроячейки является сложность переноса проекта на другие технологии. При появлении новой технологии обязательно требуется существенная переработка блока. По этой причине жесткие макроячейки используются все реже и реже; в последнее время о них вспоминают преимущественно тогда, когда автоматическая генерация проекта дает плохие результаты или просто невозможна. Хорошим примером жестких макроячеек являются встроенные запоминающие устройства и микропроцессоры. Обычно они предоставляются производителем интегральных схем (тем же, что поставляет библиотеку стандартных элементов) или производителем полупроводниковых приборов, который может предложить особо востребованную функцию (например, стандартный микропроцессор или процессор цифровой обработки сигналов).

Если макроэлемент можно параметризовать, то для создания реального топологического чертежа применяется генератор, именуемый *компилятором модуля*. Правильные структуры (ПЛМ, запоминающие устройства и умножители) создаются путем соединения готовых базовых ячеек в двухмерный массив. Все межсоединения создаются с помощью примыканий, и если ячейки спроектированы правильно, дополнительная трассировка требуется в минимальном объеме или вообще не требуется (это, в свою очередь, минимизирует паразитные емкости). Примером подобной конфигурации является программируемая логическая матрица, изображенная на рис. 8.11. Всю матрицу можно создать с помощью минимального числа ячеек. Генератор представляет собой простую программу, определяющую относительное расположение различных базовых ячеек в матрице.

Пример 8.4. Макромодуль "запоминающее устройство"

На рис. 8.13 приведен пример "жесткой" макроячейки памяти. Блок статического ОЗУ 256×32 генерируется параметризованным генератором модулей. Помимо создания топологии, генератор также выдает информацию по правильному тактированию и распределению мощности. Современные генераторы запоминающих устройств также включают небольшой "запас избыточности" на случай дефектов.

Мягкая макроячейка представляет собой модуль с данными функциональными возможностями, не имеющий конкретной физической реализации. Расположение элементов и проводки у различных мягких макроячеек с аналогичными свойствами может быть разным. Это означает, что временны*е* параметры можно определить только после завершения этапов синтеза и размещения/трассировки, т.е. процесс является непредсказуемым. Впрочем, зная внутреннюю структуру модуля и наложив в процессе генерации физической

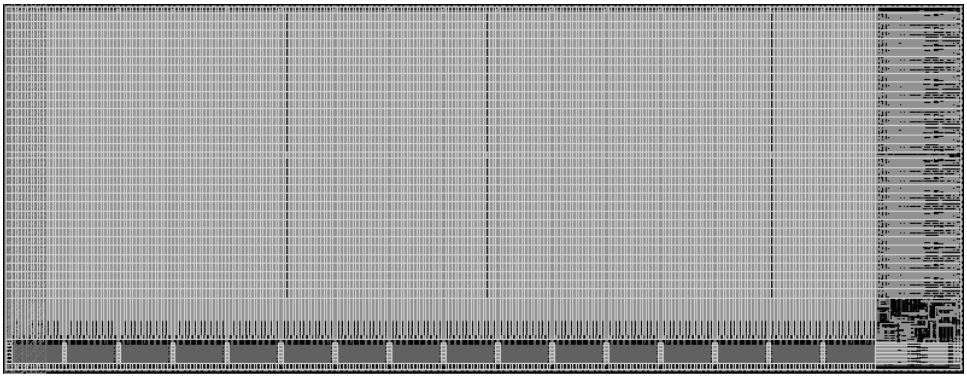


Рис. 8.13. Параметризуемая жесткая макроячейка — запоминающее устройство. Приведенный образец может хранить 256×32 (или 8192) бит. Декодеры расположены внизу. Все восемь адресных битов, а также 32 порта ввода-вывода данных располагаются на правой стороне ячейки. Общая площадь модуля памяти, реализованного по 0,18-микронной КМОП-технологии, равна всего 0,094 мм^2 (*Перепечатано с разрешения ST Microelectronics.*)

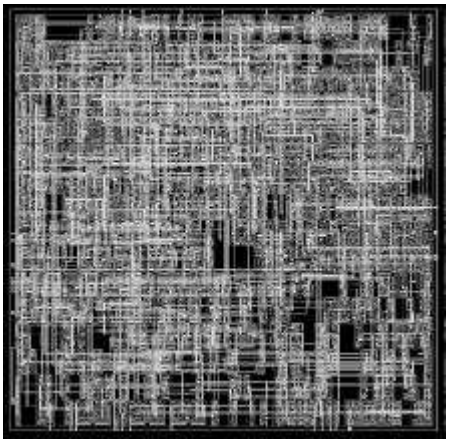
топологии жесткие ограничения на расположение и тактирование элементов, можно получить мягкую макроячейку с хорошими временны́ми характеристиками. Хотя в данном случае мы удаляемся от преимуществ индивидуальных проектов и полагаемся на полуиндивидуальные процессы разработки, мягкие макроячейки обладают одним несомненным достоинством: их можно переносить на другие технологии и процессы. Таким образом, стоимость разработки амортизируется большим набором проектов. В генераторах мягких макроячеек используются различные стили разработки в зависимости от типа целевой функции. Практически все эти генераторы можно отнести к структурным: для данной функции и значений требуемых параметров генератор дает таблицу соединений, в которой перечисляются задействованные стандартные элементы и соединения между ними. Кроме того, он предоставляет набор ограничения на тактирование, которым должны удовлетворять инструменты расположения и трассировки. Достоинство данного подхода заключается в том, что генератор использует знания о функции, которую требуется реализовать, на их основе создавая интеллектуальные структуры, более эффективные, чем те, что могут предложить средства синтеза логических элементов. Например, несколько десятилетий велись исследования, которые бы позволили получить быстрые и емкие умножители[1]. Генератор умножителей представляет собой воплощение лучших из этих разработок в автоматизированном инструменте.

Пример 8.5. Макромодуль — умножитель

На рис. 8.14 показаны два экземпляра модуля — умножителя 8×8 с различными характеристическими отношениями. Модули сгенерированы с ис-

[1] Более подробно структура умножителей описана в главе 11, где рассматривается проектирование арифметических структур.

пользованием средства ModuleCompiler от Synopsys [ModuleCompiler01]. Как видно на приведенной схеме, для генерации оригинала фотошаблона применяется распространенная методология стандартных элементов. Задачей генератора макромодуля является перевод компактного входного описания в оптимизированное соединение стандартных элементов, удовлетворяющее условиям, наложенным на временны*е* параметры. Достоинством такого "мягкого" подхода является возможность быстрой генерации модулей с различными характеристическими отношениями. Кроме того, относительно просто можно обеспечить переносимость на другую технологию производства. Доступность макромодулей существенно изменила сферу полуиндивидуального проектирования XXI столетия. При том что сложность интегральных схем растет экспоненциально со временем, создание новых схем с нуля становится непрактичным, неэкономичным и неправдоподобным. Все чаще и чаще схемы создаются на основе готовых блоков, сложность которых постоянно растет, а функциональные возможности расширяются. Обычно эти модули покупаются в готовом виде у сторонних производителей согласно условиям лицензионных соглашения (или за роялти). Макромодули, распространяемые по такому принципу, называются *частными* модулями. Чем-то это похоже на ситуацию в мире программного обеспечения, где в большом проекте часто задействует набор программных библиотек, предполагающих повторное использование. В качестве хороших примеров распространенных частных модулей можно привести встроенные микропроцессоры и микроконтроллеры, процессоры цифровой обработки сигналов, интерфейсы шин (например, PCI) и несколько специализированных функциональных модулей (модули БПФ и фильтры для приложений цифровой обработки сигналов, кодеры коррекции ошибок для беспроводной связи, модули кодирования/декодирования MPEG для видеоприложений). Понятно, что для того, чтобы модуль, защищенный законами об интеллектуальный собственности, был полезен, требуется не только "железо", но и соответствующее программные средства (например, компиляторы и отладчики, если речь идет о встроенных процессорах), модели прогнозирования и испытательные стенды. Наличие стендов очень важно, поскольку они предлагают конечному пользователю единственную возможность проверить, что модуль дает обещанные функциональные возможности и быстродействие. Проектирование системы на чипе очень быстро превращается в задачу повторного использования модулей различной степени интеграции. На низком уровне мы имеем библиотеку стандартных элементов; на высоком — такие функциональные модули, как умножители, информационные тракты и запоминающие устройства; далее мы получаем встроенные процессоры и, наконец, приходим к специализированным мегаячейкам. Поскольку с каждым годом в один кристалл внедряется все больше и больше функциональных возможностей, уже никого не удивляет, что типичная специализированная интегральная схема состоит из комбинации стилей разработки и модулей — моря стандартных элементов, в котором плавает несколько жестких или мягких макроячеек.



```
string mat = "booth";
directive (multtype = mat);
output signed [16] Z = A * B;
```

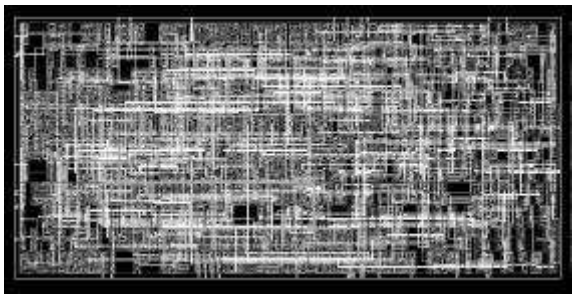


Рис. 8.14. "Мягкие" макромодули запоминающих устройств. На обоих слоях реализованы умножители 8×8, имеющие, правда, разные характеристические размеры. В рамке показано краткое входное описание схемы для компилятора

Пример 8.6. Процессор для беспроводной связи

На рис. 8.15 показана интегральная схема, реализующая стек протоколов для домашней беспроводной системы связи [Silva01]. Бо́льшая часть площади занята встроенным микропроцессором (процессор Tensilica Xtensa [Xtensa01]) и его системой памяти. Данный процессор разрешает гибкую реализацию высших уровней стека протоколов (уровень приложений/сетевой уровень) и позволяет изменять функциональные возможности чипа после его изготовления. Модули памяти генерируются с использованием компиляторов модулей, предоставленных производителем процесса. Ядро процессора автоматически генерируется согласно высокоуровневому описанию на Verilog; его физическая реализация строится на основе стандартных элементов. Преимуществом подхода "мягкое ядро" является возможность настройки набора команд процессора под конкретное приложение; кроме того, сам процессор легко перевести на новую технологию и процесс производства.

Реализация на микропроцессоре частей протокола, предполагающих большой объем вычислений (MAC/физический уровень), потребовала бы очень высоких тактовых частот и привела бы к ненужному рассеянию мощности на чипе. Но поскольку данные функции постоянны и обычно не требуют гибкой реализации, они реализуются как модуль-ускоритель на стандартных элементах. Жесткая реализация позволяет справиться с огромным объемом вычислений при относительно низком потреблении мощности и невысокой тактовой частоте. Разработчик системы на кристалле постоянно сталкивается с проблемой выбора: какой параметр будет критическим после производства — гибкость или высокое быстродействие при низких уровнях мощности. К счастью, существуют инструменты, которые помогают разработчику изучить все пространство проектных параметров и проанализировать существующие компромиссы [Silva01]. Обратите также внимание на то, что кристалл

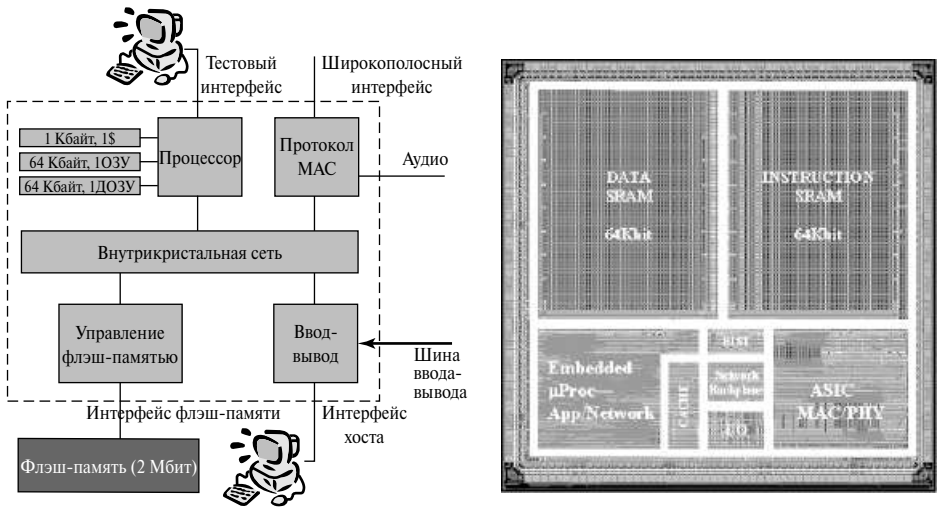


Рис. 8.15. Процессор беспроводной связи — пример гибридной методологии проектирования специализированных интегральных схем. Процессор объединяет встроенный микропроцессор и его систему памяти со специализированными аппаратными ускорителями и модулями ввода-вывода. Кроме того, обратите внимание на встроенный сетевой модуль [Silva01]

содержит набор интерфейсов ввода-вывода, а также встроенный сетевой модуль, который позволяет гармонично управлять потоком информации между процессором и различными ускорителями/модулями ввода-вывода.

Процесс генерации макромодуля зависит от природы блока (жесткая или мягкая), а также от того, какие элементы используются в качестве составляющих. Несколько наиболее распространенных подходов к решению этой задачи кратко рассмотрены ниже.

8.4.4. Полуиндивидуальный процесс проектирования

Итак, мы определили составляющие компоненты методологии проектирования на основе стандартных элементов. В данном разделе мы обсудим, как они связаны между собой. На рис. 8.16 представлена традиционная последовательность действий при проектировании полуиндивидуальной схемы. Рассмотрим кратко эти составляющие процесса проектирования.

1. **Выбор проекта** заключается в вводе начальной информации в систему проектирования специализированной ИС. Здесь могут использоваться различные методы, в том числе структурные схемы и блок-схемы; аппаратные описательные языки (Hardware Description Language — HDL), подобные VHDL, Verilog и производные от C (например, SystemC); поведенческие описательные языки плюс высокоуровневый синтез; а также улучшенные частные модули.

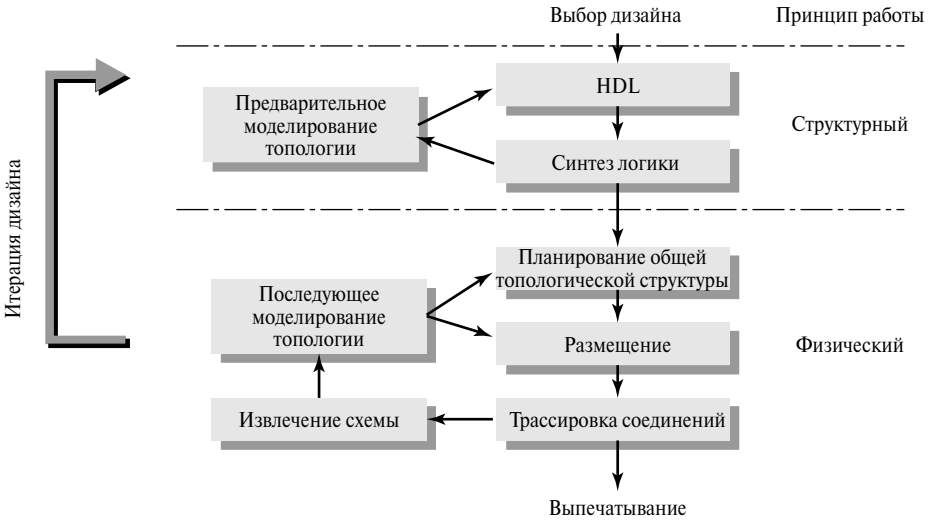
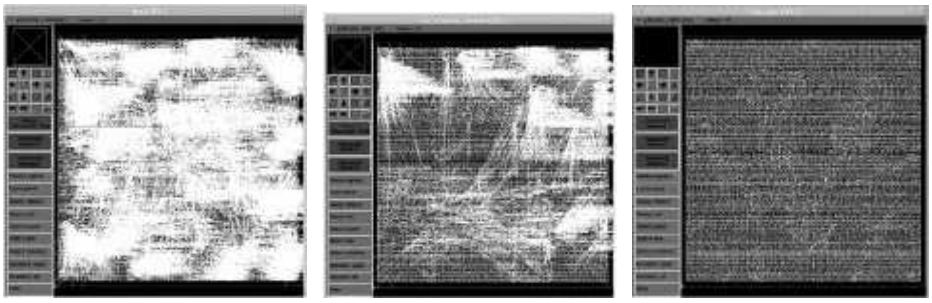


Рис. 8.16. Процесс разработки полуиндивидуальной (или специализированной) ИС

2. Инструменты **синтеза логики** транслируют модули, описанные с помощью HDL, в таблицу соединений. После этого к ним можно добавить таблицы соединений повторно используемых или сгенерированных макроэлементов и получить полную таблицу соединений проекта.
3. *Предварительное моделирование и верификация топологии.* Проверяется точность проекта. Согласно оценкам паразитных нагрузок и параметров топологии выполняется анализ быстродействия. Если проект оказывается нефункциональным, необходимы дополнительные итерации (выбор другого дизайна или повтор этапа синтеза логики).
4. **Планирование общей топологической структуры**. На основе оцененных размеров модулей создается общая схема использования кристалла. Также на этом этапе разрабатываются сети питания и синхронизации.
5. **Размещение**. Выбирается точное расположение элементов.
6. **Трассировка соединений**. Прокладываются соединения между элементами и блоками.
7. **Извлечение**. Согласно фактической топологии (точные размеры устройств, паразитные нагрузки, емкость и сопротивление проводников) генерируется модель кристалла.
8. **Заключительное моделирование и верификация**. Функциональные возможности и быстродействие кристалла проверяется при наличии паразитных нагрузок. Если получающаяся структура не удовлетворяет сформулированным требованиям, могут потребоваться дополнительные итерации, включающие этапы планирования общей топологической структуры, размещения и трассировки соединений. Очень часто это не решает проблему, поэтому приходится возвращаться на этап структурной разработки.



Первоначальный проект
а

Промежуточный проект
б

Окончательный проект
в

Рис. 8.17. Процесс проектирования. Белыми линиями обозначены соединения, нарушающие условия, наложенные на временны*е* параметры. На каждой итерации процесса разработки эти ошибки устраняются путем оптимизации логики, введения буферов, выбора положений элементов или оптимального маршрута соединений. Конечным результатом является проект, не содержащий ошибок [Avanti01]

9. **Выпечатывание**. Как только получена структура, удовлетворяющая всем требованиям, генерируется двоичный файл, который содержит всю информацию, необходимую для генерации фотошаблона. Затем данный файл отправляется производителю ИС. Этот важный этап в жизни кристалла называется *выпечатыванием*.

Хотя описанный процесс проектирования хорошо служил многие годы, он оказался недостаточным, когда КМОП-технология достигла 0,25-микронного рубежа. После эволюции технологии в глубокую субмикронную область весьма важной стала роль паразитных нагрузок (особенно порождаемых межсоединениями). Модели прогнозирования, используемые инструментами структурного синтеза и генерации логики, давали недостаточно точные оценки этих паразитных составляющих. Из-за этого шансы того, что сгенерированный проект с первой попытки будет удовлетворять условиям, наложенным на временны*е* параметры, были очень малыми (рис. 8.17, *а*). Таким образом, разработчик (или команда разработчиков) должен был выполнять ряд дорогих итераций процессов синтеза/генерации топологии, пока не получал приемлемый фотошаблон, удовлетворяющий всем наложенным ограничениям (см. рис. 8.17, *б*, *в*). Каждая итерация могла требовать нескольких дней — на одну трассировку сложного кристалла с помощью самых мощных компьютеров уходит порядка недели! Понятно, что с уменьшением масштаба технологии число требуемых итераций только растет. Наличие данной проблемы сделало очевидной потребность в новых решениях и изменении методологии проектирования.

Как правило, для решения данной проблемы предлагается более тесная интеграция процессов логической и физической разработки. Если, например, средство синтеза логики дополнительно выполняет какие-то расчеты по расположению элементов (или управляет им), это позволяет получить более точные оценки параметров схемы. На рис. 8.18 приведен пример среды

Рис. 8.18. Комплексный синтез с размещением и трассировкой уменьшает число итераций при работе в глубокой субмикронной области

разработки, в которой РТЛ-синтез совмещается с расположением и трассировкой элементов в первом приближении. Получающаяся в результате таблица соединений передается средству оптимизации, которое выполняет более точное размещение элементов и трассировку соединений, при котором не нарушаются условия на временны*е* параметры. Хотя такой подход позволяет достаточно успешно снизить количество итераций, реализовать его достаточно сложно с точки зрения разработчика инструментов проектирования. Поскольку с каждым качественным скачком технологии число паразитных эффектов увеличивается, экспоненциально усложняется и процесс оптимизации проекта, в котором необходимо учитывать все эти эффекты. В результате могут потребоваться новые подходы к разработке. В последующих главах мы рассмотрим дизайнерские решения, которые могут помочь решить некоторые из этих проблем. Например, это может быть использование правильных и предсказуемых структур как на логическом, так и на физическом уровне.

8.5. Подходы к реализации, основанные на матрицах

Хотя автоматизация позволяет уменьшить время разработки, она никак не влияет на время процесса производства. Все рассмотренные до этого момента методологии проектирования требуют полного прохода по процессу производства, что может требовать от трех недель до нескольких месяцев и существенно замедлить выход конечного продукта. Кроме того, при постоянно растущей стоимости фотошаблона использование специализированного процесса является дорогим, поэтому применять его следует лишь тогда, когда это экономически выгодно.

Таким образом, было разработано несколько альтернативных подходов к реализации, которые не требуют полного цикла процесса производства либо совершенно не требуют специальной обработки. Преимуществом такого подхода является низкие единовременные затраты на проектирование, следовательно, они выгодны для небольших серий. Платой за это является более низкое быстродействие, меньшая плотность интеграции или более высокое рассеяние мощности.

8.5.1. Матрицы с предварительной диффузией (или программируемым фотошаблоном)

При данном подходе производители выпускают пачки пластин, содержащих массивы элементарных ячеек или транзисторов. Все этапы производства, требуемые для создания транзисторов, стандартизованы и выполняются безотносительно к конечному приложению.

Чтобы превратить эти пластины с базовыми кристаллами в реальную структуру, необходимо добавить только требуемые межсоединения, что и определит общую функцию кристалла при минимуме этапов металлизации. Рисунок разводки можно спроектировать и нанести на пластины гораздо быстрее, чем при полной схеме производства, что сокращает время производства до недели (или даже меньшего срока).

Описанный подход часто называют разработкой на основе *матрицы логических элементов* или *матрицы с "морем" элементов* в зависимости от стиля пластины с предварительной диффузией. Чтобы проиллюстрировать данную концепцию, рассмотрим элементарную ячейку матрицы логических элементов, показанную на рис. 8.19, *а*. Она состоит из четырех n-МОП- и четырех p-МОП-транзисторов, поликремниевых соединений, а также шин питания и заземления. Существуют две возможные точки контакта на диффузионной поверхности и две возможные точки контакта для поликремниевых полосок. Эту ячейку, которая на данный момент не реализует никакой логической функции, можно превратить в реальную схему, нанеся несколько проводников на металлический уровень и контактные отверстия. Сказанное иллюстрируется на рис. 8.19, *б*, где ячейка преобразования в элемент НЕ-ИЛИ с четырьмя входами.

Изначально в подходе с использованием *матрицы логических элементов* ячейки в строках разделяются каналами проводки, как показано на рис. 8.20, *а*. В общем, по внешнему виду это похоже на результат использования методологии стандартных элементов. С введением дополнительных металлических слоев от каналов трассировки можно избавиться, осуществив соединение поверх основных элементов (возможно, при этом какая-то ячейка останется незадействованной). Такая безканальная архитектура, также именуемая *"морем" элементов* (sea of gates) (рис. 8.20, *б*), позволяет получать бо́льшую плотность и добиваться огромной степени интеграции — миллионы логических элементов на одном кристалле. Другим достоинством плотной стыковки логических элементов является возможность настройки контактного слоя между первым металлическим слоем и диффузным и/или поликрем-

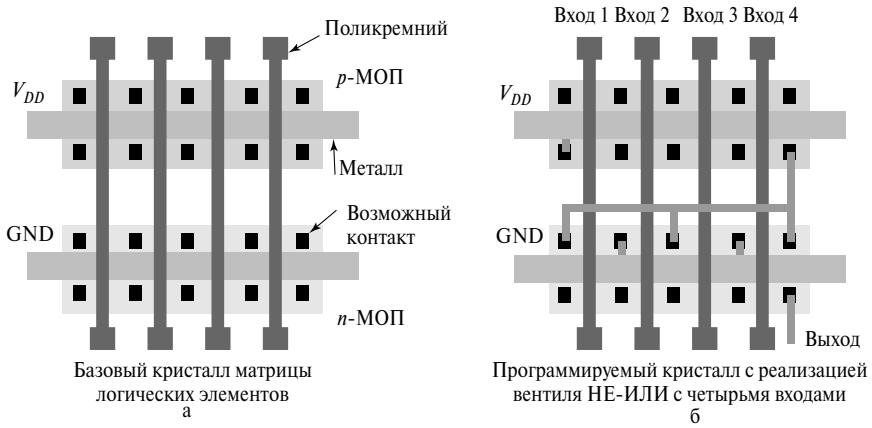


Рис. 8.19. Пример использования матрицы логических элементов

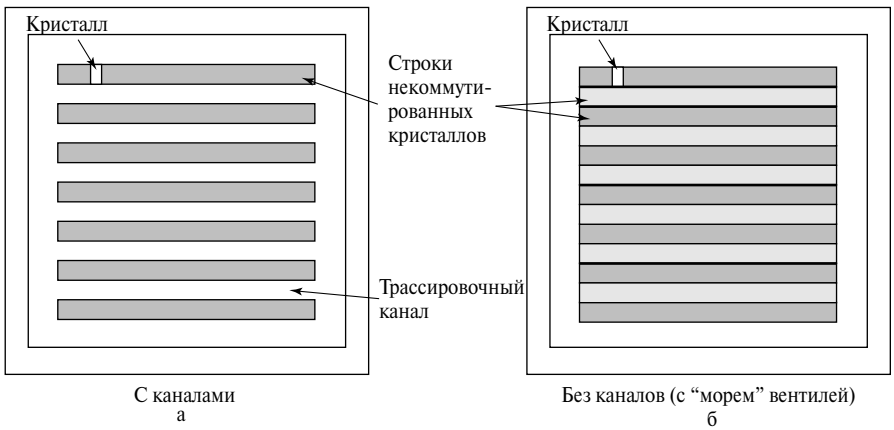


Рис. 8.20. Архитектуры матриц логических элементов

ниевым слоем, тогда как при использовании стандартного подхода (матрица базовых элементов) все контакты предопределены (см. рис. 8.19, *а*). Эта дополнительная гибкость позволяет добиваться еще большего уменьшения размеров ячеек.

Основная трудность при разработке шаблона матрицы элементов (или “моря” элементов) заключается в определении связи между примитивным элементом и размером отдельных транзисторов. Чтобы минимизировать число элементов, пожертвованных на создание межсоединений, необходимо наличие достаточного числа проводящих трактов. Размер элемента нужно выбирать так, чтобы готовые транзисторы можно было использовать с максимальной эффективностью в большом диапазоне разработок. Например, конфигурация, показанная на рис. 8.19, хорошо подходит для реализации элементов с четырьмя входами, однако для реализации элементов с двумя вхо-

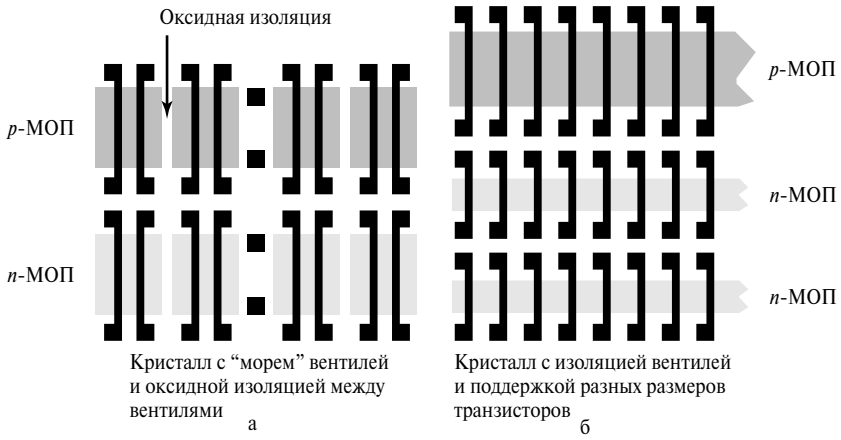


Рис. 8.21. Примеры примитивных элементов (метод "моря" элементов) (согласно [Veendrick92])

дами применять ее невыгодно. Реализация триггера требует нескольких элементов. Для примера на рис. 8.21 в упрощенном формате показано несколько альтернативных структур элементов. При одном подходе каждый элемент содержит ограниченное число транзисторов (от четырех до восьми). Элементы обособлены посредством *изоляции окислом* (или *за счет геометрии*). Второй подход заключается в применении длинных рядов транзисторов, когда одна диффузионная область совместно используется всеми транзисторами. При такой архитектуре некоторые устройства необходимо выключить электрически, чтобы обеспечить изоляцию соседних элементов, соединив n-МОП- и p-МОП-транзисторы с GND и V_{DD} соответственно. Такая техника называется *изоляцией*. Описанный подход нерационально использует часть транзисторов для обеспечения изоляции, однако в целом дает более высокую плотность транзисторов.

На рис. 8.22 показан стандартный компонент матрицы элементов при использовании подхода "изоляция элементов" (согласно [Smith97]). Толщина ячейки равна ширине одного соединительного проводника, а сама ячейка содержит один p- и один n-канальный транзистор. Кроме того, на рисунке показана стандартная ячейка, содержащая все возможные положения контактов. Всего предусмотрен 21 контакт в вертикальном направлении, т.е. ячейка имеет ширину в 21 канал.

Стоит отметить, что ячейка, изображенная на рис. 8.21, *б*, содержит два ряда меньших n-МОП-транзисторов, которые при необходимости можно соединить параллельно. Меньшие транзисторы полезны при реализации запоминающих устройств или логики на проходных транзисторах. Очевидно, здесь возникает проблема масштабирования транзисторов в ячейках. Из-за того, что методология проектирования, основанная на матрицах элементов, сильно зависит от межсоединений, задержка распространения определяется преимущественно емкостью межсоединений. Из-за этого кажется, что нам выгодны устройства большего размера, которые в холостом режиме да-

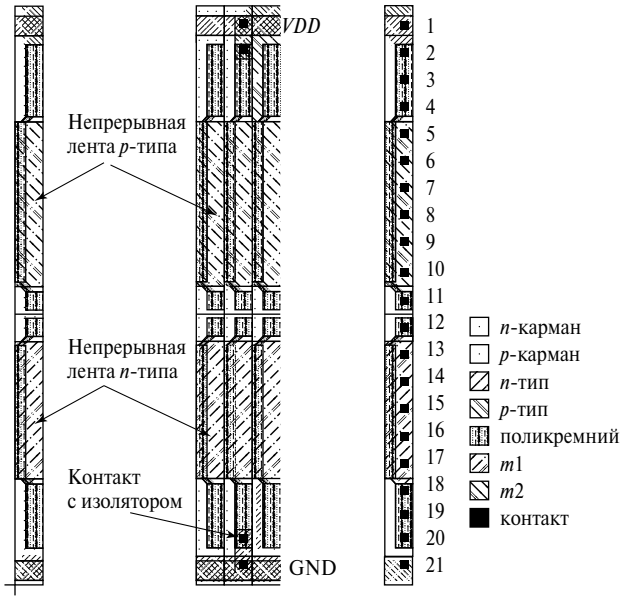


Рис. 8.22. Базовая ячейка в матрице элементов с изолированными затворами (согласно [Smith97])

ют бо́льшие потери на единицу площади. С другой стороны, бо́льшие транзисторы можно создать путем параллельного соединения нескольких меньших устройств.

Отображение логической структуры в массив ячеек выполняется преимущественно автоматически путем синтеза логики с последующим расположением элементов и трассировкой. Конечная плотность и быстродействие реализации "моря" элементов сильно зависят от качества соответствующих средств автоматического проектирования. Кроме того, коэффициенты использования структур типа "море" элементов сильно зависят от типа реализуемого приложения. Для правильных структур (например, запоминающих устройств) можно получить коэффициенты использования, близкие к 100%. Для других приложений коэффициенты использования могут быть существенно ниже (меньше 75%) из-за ограничений, связанных с проводкой. На рис. 8.23 приведен пример макроячейки триггера, реализованной с помощью библиотеки матриц элементов с изолированными элементами.

Подобно сценариям, реализуемым на основе стандартных ячеек, разработчики матриц "моря" элементов обнаружили, что проект с большим числом логических элементов имеет большие требования к памяти. Реализация соответствующих запоминающих устройств на основе базовых ячеек матрицы логических элементов возможна, однако не очень эффективна. Более эффективный подход заключается в выделении какой-то площади для специализированных модулей памяти. Комбинация матриц логических элементов с фиксированными макроячейками называется *матрицей с внедренными логическими элементами*. Кроме того, идеальными кандидатами на внедрение являются другие модули, например микропроцессоры и микроконтроллеры.

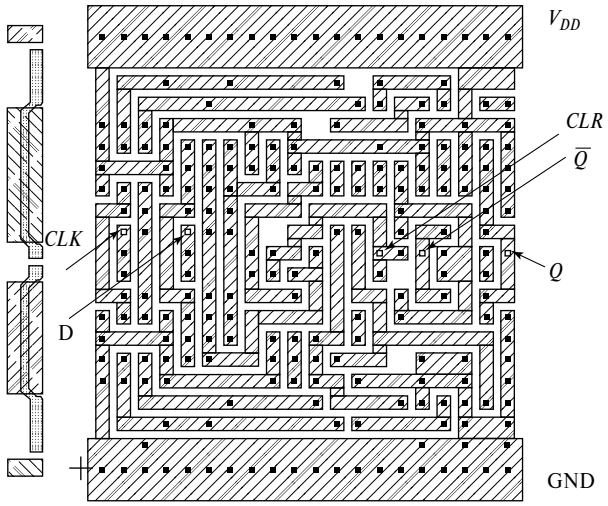


Рис. 8.23. Триггер, реализованный с помощью библиотеки матриц элементов с изолированными элементами. Слева показана базовая ячейка (согласно [Smith97])

Пример 8.7. "Море" логических элементов

На рис. 8.24 приведен пример реализации "моря" логических элементов. Максимальная емкость матрицы составляет 300 тысяч логических элементов, сама матрица реализована по 0,6-микронной КМОП-технологии. В левой верхней части матрицы реализована запоминающая подсистема, поэтому здесь схема имеет правильный модульный вид. В остальной части матрицы реализована произвольная логика.

Матрицы логических элементов или стандартные ячейки?

В 1980–1990-х годах, когда подавляющее большинство кристаллов содержало менее 50 тысяч логических элементов, один цикл проектирования длился несколько недель или месяцев. В таких условиях двух-трехнедельное сокращение времени цикла было существенным и компенсировало увеличение площади разработки. Теперь, когда речь идет о процессах в глубокой субмикронной области, а число логических элементов давно исчисляется миллионами, времени на разработку требуется гораздо больше, поэтому незначительное сокращение цикла разработки уже не так существенно, как ранее. Более того, самой трудоемкой и затратной частью процесса производства полупроводникового устройства стала металлизация, что еще сильнее уменьшило привлекательность матриц логических элементов. Появился другой альтернативный вариант быстрого получения прототипа — рассматриваемые в следующем разделе предварительно спаянные матрицы, которые отобрали у матриц логических элементов значительную часть их рынка.

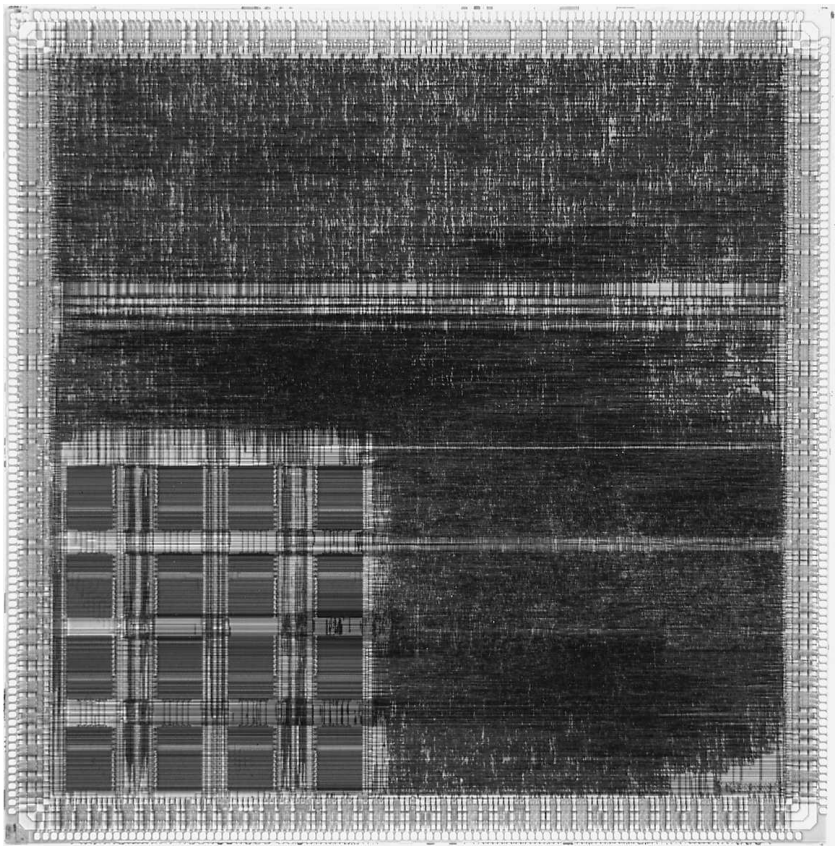


Рис. 8.24. Микрофотография кристалла с матрицей логических элементов (LEA300K) (*Перепечатано с разрешения LSI Logic.*)

Тем не менее идею логического модуля с программируемым фотошаблоном еще рано предавать забвению. Достоинством разработок с правильной и фиксированной структурой является возможность быстрого и точного расчета коэффициентов нагрузки, паразитных нагрузок проводки и шума перекрестных соединений. Если использовать подход, основанный на стандартных ячейках, то по большому счету эти значения будут известны только после расположения, трассировки и выщелачивания. В принципе можно рассмотреть возможность заполнения фрагментов большого кристалла правильной логической матрицей, состоящей из нескоммутированных логических ячеек, на которые наложена сетка проводки. Реальное программирование модуля выполняется путем расположения сквозных отверстий в предопределенных местах. Как показано на рис. 8.25, использование структуры с узлами, программируемыми посредством сквозных отверстий, позволяет учесть большое число рисунков разводки на правильной сетке разводки. По мнению авторов данной книги, матрицы с предварительной диффузией еще просуществуют какое-то время.

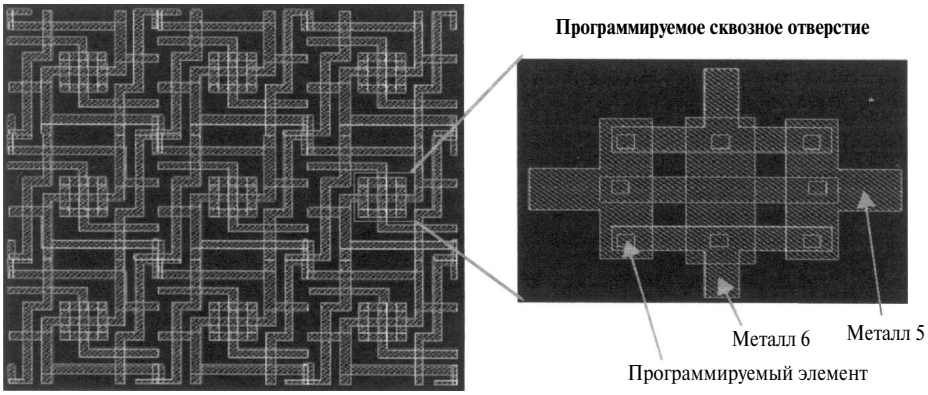


Рис. 8.25. Матрица логических элементов, программируемая посредством сквозных отверстий. Сквозные отверстия используются для превращения универсальной сетки разводки в конкретный рисунок, что позволяет получать предсказуемые матрицы логических элементов [Pileggi02]

8.5.2. Матрицы с предварительным монтажом

Хотя матрицы с предварительной диффузией открывают путь к быстрой реализации, более эффективным было бы вообще полное устранение этапов специализированного производства. В результате мы приходим к концепции предварительно обработанного кристалла, который можно запрограммировать "на месте" (т.е. за пределами фабрики по производству полупроводниковых устройств) и реализовать набор заданных булевых функций. Подобная программируемая матрица с предварительным монтажом ячеек называется *программируемой пользователем матрицей логических элементов* (field-programmable gate array — FPGA). Преимуществом данного подхода является полное отделение процесса производства от фазы реализации, что позволяет разделить его стоимость между большим числом проектов. Собственно реализацию может выполнить пользователь за разумно короткое время. Основным недостатком данной техники является снижение быстродействия и плотности структуры по сравнению с более специализированными подходами.

При реализации набора булевых функций на основе правильной матрицы ячеек без каких бы то ни было этапов обработки необходимо решить два основных вопроса.

1. Как реализовать "программируемую" логику — т.е. логику, которую можно переделать под реализацию любой возможной булевой функции?
2. Как и где хранить *программу* (или *конфигурацию*), которая задает выполнение программируемой матрицей определенной логической функции?

Ответ на второй вопрос зависит от используемой технологии запоминания. Поскольку эти технологии мы будем подробно рассматривать позже, здесь ограничимся только поверхностным обзором. В целом можно выделить три различные техники.

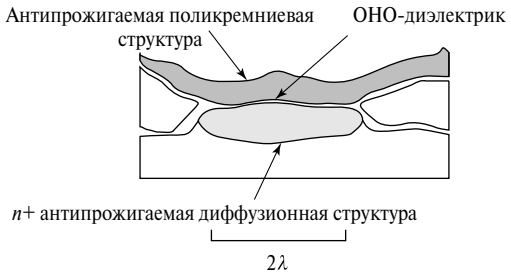


Рис. 8.26. Пример антипрожигаемой перемычки. Слой толщиной 10 нм диэлектрика "оксид–нитрид–оксид" размещается между проводящими поликремниевым и диффузионным слоями. По умолчанию цепь разомкнута; в таком состоянии она остается, пока через нее не пропустить большой программирующий ток. Тогда диэлектрик плавится и формируется постоянное соединение с фиксированным сопротивлением (согласно [Smith97])

- **FPGA с однократной записью или пережигаемыми перемычками**. Матрица логических элементов сводится к реализации определенной функции путем пережигания плавких перемычек или замыкания накоротко антипрожигаемых перемычек. Плавкой перемычкой (fuse) называется соединительный элемент, который по умолчанию замкнут накоротко. При подаче большого тока элемент плавится, т.е. цепь размыкается. Антипрожигаемая перемычка имеет противоположные характеристики. Пример реализации программируемой матрицы с антипрожигаемыми перемычками показан на рис. 8.26 [El-Ayat89]. Преимуществом однократной записи являются очень маленькие служебные издержки на площадь под память программы (т.е. собственно перемычки). Однако есть и серьезный недостаток — однократное программирование. Исправление или расширение схемы невозможно, так как каждый новый проект требует новых компонентов.
- **Энергонезависимая FPGA**. Программа хранится в энергонезависимой памяти, которая сохраняет информацию даже при отключении питания. В качестве примеров таких матриц можно привести электрически стираемое программируемое постоянное запоминающее устройство, ЭСППЗУ (Electrically Erasable Programmable Read-Only Memory — EEPROM), или флэш-память. После программирования логика сохраняет свои функциональные возможности и работает, пока не будет перепрограммирована. Недостаток этого подхода заключается в том, что энергонезависимая память требует специальных этапов производственного процесса (например, размещения сверхтонких окислов). Кроме того, для программирования и стирания ячеек памяти требуются высокие напряжения (больше 10 В). Генерация этих высоких напряжений и распределение их по матрице логических элементов дополнительно усложняют структуру.
- **Энергозависимая FPGA.** В данном популярном подходе к программированию матриц логических элементов для хранения программы применяются энергозависимые статические ОЗУ (оперативные запоминающие устройства). Поскольку данные ячейки памяти теряют записанную инфор-

мацию при выключении FPGA, при включении блока требуется повторная загрузка конфигурации из внешнего постоянного запоминающего устройства. Чтобы запрограммировать компонент во время запуска, необходимые данные последовательно вводятся в элемент по одной шине. Вообще, на практике можно считать, что ячейки ОЗУ FPGA в это время конфигурируются как гигантский сдвиговый регистр. Как только все ячейки памяти будут загружены, начинается нормальная работа. Время задания конфигурации пропорционально числу программируемых элементов. В современных больших FPGA (часто более миллиона логических элементов) это время может быть значительным. Поэтому в последнее время наблюдается тенденция к переходу на параллельный программный интерфейс, позволяющий одновременно программировать несколько ячеек.

В отличие от энергонезависимых устройств энергозависимые FPGA не налагают никаких специальных требований на процесс производства, и их можно реализовать с помощью обычного КМОП-процесса. Кроме того, в фазе создания прототипа разработчики могут повторно использовать уже готовые кристаллы. Логику можно модифицировать и обновлять после доставки пользователю — клиенту можно отправить новый файл конфигурации, с помощью которого он обновит кристалл (это гораздо дешевле, чем отправлять новый кристалл). Мало того, логику можно модифицировать динамически на лету (в процессе выполнения). Данный процесс называется *реконфигурацией* и стал довольно популярным в конце 1990-х. В некотором смысле он породил парадигму, ставшую весьма популярной как в мире программирования (представленном микропроцессорами), так и в области проектирования логических схем.

Если со вторым вопросом все более-менее понятно, то ответ на первый не будет таким коротким. Реализация сложной схемы, допускающая программирование, требует, чтобы настраиваемыми были и логические функции, и их соединения. В последующих разделах мы обсудим различные способы реализации программируемой логики, после чего представим обзор программируемых межсоединений. В завершение мы рассмотрим несколько способов объединения этих составляющих.

Программируемая логика

В настоящее время существуют два подхода к современной программируемой логике: основанный на матрицах и основанный на ячейках.

Программируемая логика на основе матриц

Ранее мы обсуждали, как с помощью *программируемой логической матрицы*, ПЛМ (Programmable Logic Array — PLA), реализовать произвольные булевы логические функции в виде регулярной структуры. Подобный подход можно применить и к программируемым пользователем логическим устройствам. Рассмотрим, например, логическую структуру, приведенную на рис. 8.27. Кружочком (○) на пересечении соединений обозначено программируемое соединение, т.е. точка, которая либо активизирована, либо нет. Изучая диаграмму, мы видим, что она эквивалентна ПЛМ, в которой плоскости

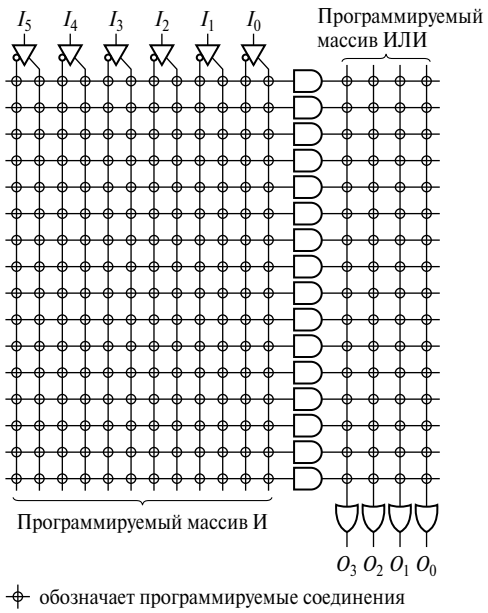


Рис. 8.27. Логическая матрица, программируемая плавкими перемычками

И и ИЛИ можно программировать, избирательно активизируя соединения. Такой подход позволяет реализовать произвольные логические функции в двухуровневом формате суммы произведений. Плоскость И создает требуемые конъюнктивные формы, а плоскость ИЛИ формирует выход, суммируя выбранные произведения. Чтобы включить данную входную переменную (например, I_1) в заданную конъюнктивную форму, требуется просто включить выключатель на пересечении входного сигнала и конъюнктивной формы. Подобным образом конъюнктивная форма включается в выход путем замыкания соответствующего соединения на плоскости ИЛИ. В целом функциональные возможности ПЛМ ограничиваются числом входов, выходов и элементарных конъюнктивных форм.

Можно предсказать, что данная схема будет использоваться с несколькими вариациями, часть из которых представлена на рис. 8.28. Точкой (•) на пересечении двух линий обозначена неплавкая жесткая связь. Первая структура описывает архитектуру ППЗУ, в которой плоскость И фиксирована, и на ней представлены все возможные конъюнктивные формы. Вторая структура, называемая *программируемой матричной логикой*, ПМЛ (Programmable Array Logic — PAL), представляет другой край спектра возможных решений, когда фиксируется плоскость ИЛИ, а плоскость И является программируемой. Архитектура ПЛМ является наиболее общей, позволяющей реализовывать произвольные логические функции. С другой стороны, структуры ППЗУ и ПМЛ хотя и имеют меньшую гибкость, но это компенсируется большей плотностью и быстродействием. Выбор нужной структуры сильно зависит от природы реализуемых булевых функций. В литературе все эти

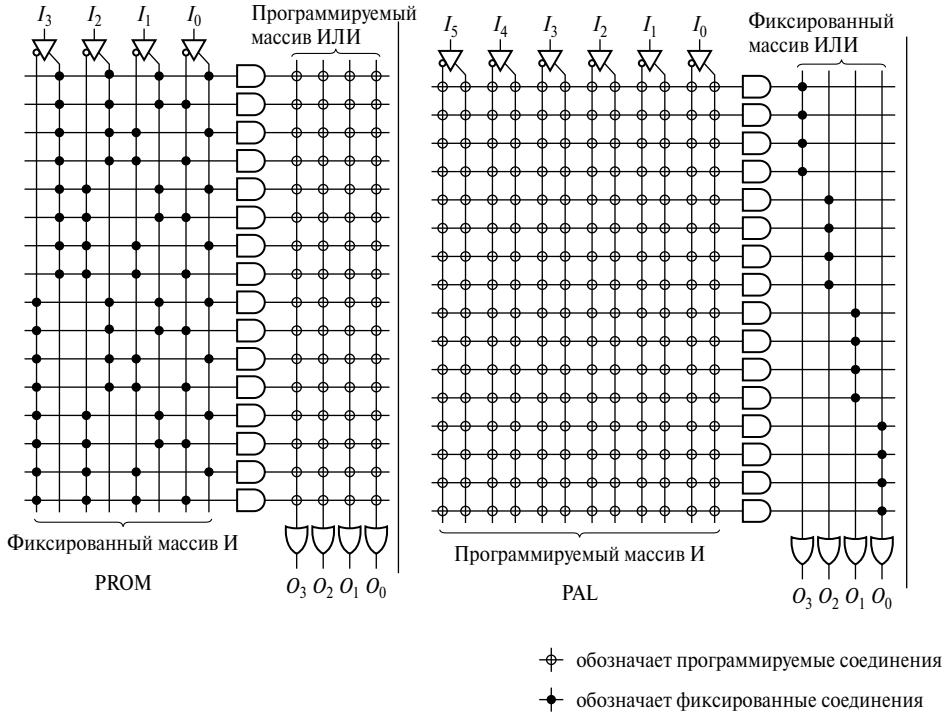


Рис. 8.28. Альтернативные программируемые логические устройства (ПЛУ) на основе плавких перемычек

подходы обычно называются общим термином *программируемые логические устройства*, ПЛУ (Programmable Logic Devices — PLD).

В эпоху большой плотности интеграции одноматричная природа структур ПЛМ, ППЗУ и ПМЛ, (см. рис. 8.27 и 8.28) становится все менее и менее привлекательной. Во-первых, реализация очень сложных логических функций на одной большой матрице приводит к потере плотности и быстродействия. Во-вторых, показанные массивы реализуют только комбинационную логику. Для реализации полной последовательной структуры требуется наличие регистров и/или триггеров. Таким образом, мы приходим к необходимости следующих модификаций.

1. Разбиение матрицы на несколько меньших фрагментов, часто называемых макроячейками.
2. Введение в структуру триггеров и обеспечение потенциальной обратной связи выходных сигналов с входными.

Один из вариантов реализации указанных модификаций показан на рис. 8.29. ПМЛ состоит из k макроячеек, каждая из которых может выбирать из i входов и давать максимум j конъюнктивных членов. Каждая макроячейка состоит из одного программируемого регистра — его можно сконфигурировать как D-, T-, J-K- или тактируемый S-R-триггер. k выходных сигналов

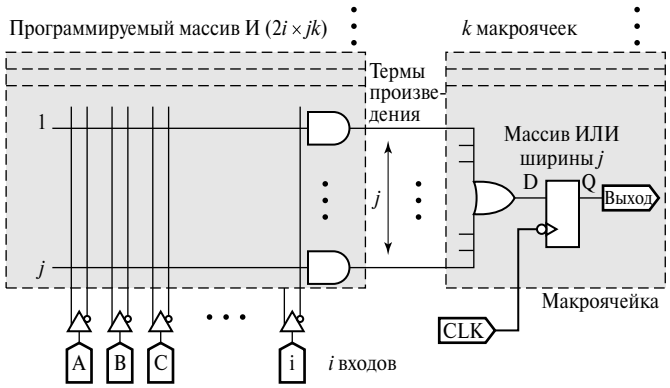


Рис. 8.29. Схематическое представление ПМЛ с i входами, j конъюнктивными членами/макроячейками и k макроячейками (или выходами) [Smith97]

возвращаются на входную шину, т.е. они являются подмножеством i входных сигналов.

Использование подхода ПЛМ к настраиваемой логике имеет два преимущества.

- Структура является правильной, благодаря чему очень легко можно оценить ее паразитные параметры и предсказать площадь, скорость и рассеяние мощности.
- Она обеспечивает эффективную реализацию логических функций, которые хорошо представляются двухуровневым описанием. К этой категории относятся, в частности, функции с большим разветвлением по входу. Соответствующие конечные автоматы применяются в контроллерах и синтезаторах.

С другой стороны, недостатком матричной структуры являются очень большие служебные издержки. Каждый промежуточный узел имеет масштабируемую емкость, которая отрицательно влияет на быстродействие и потребляемую мощность. Особенно это относится к матрицам, некоторые фрагменты которых используются не на полную мощность, т.е. когда активно задействована лишь часть конъюнктивных форм.

Пример 8.8. Пример программируемой макроячейки

На рис. 8.30 приведен пример программирования модуля ППЗУ. Цель программирования структуры — реализовать логическую функцию, использованную ранее при обсуждении ПЛМ (уравнение (8.1)):

$$f_0 = x_0x_1 + \bar{x}_2;$$
$$f_1 = x_0x_1x_2 + \bar{x}_2 + \bar{x}_0x_1.$$

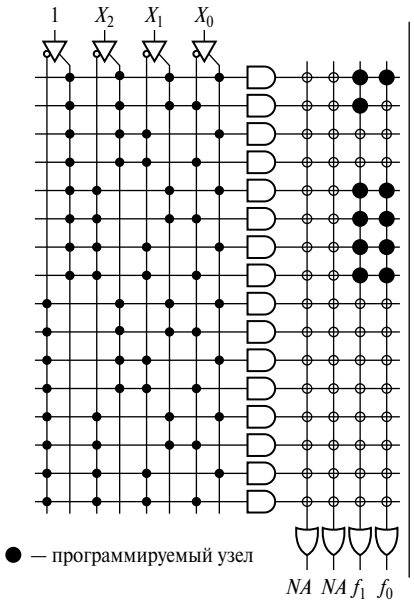


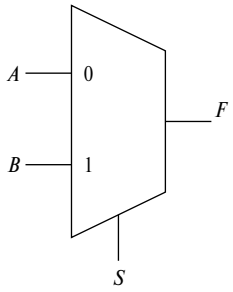
Рис. 8.30. Программирование ППЗУ

Отметим, что используется только часть матрицы, поскольку число входных (3) и выходных (2) переменных меньше размерности матрицы (4×4). Неиспользуемым входным переменным присваивается постоянное значение (0 или 1). Большие точки на выходных плоскостях представляют программируемые узлы. Читателю предлагается самостоятельно повторить упражнение, выполненное для модулей ПЛМ и ПМЛ (см. рис. 8.28 и 8.29).

Программируемая логика, основанная на ячейках

Использование подхода "сумма произведений" дает правильные структуры, очень эффективные для логических функций, имеющих большое разветвление по входу (например, конечные автоматы). С другой стороны, он не очень хорош для логики с большим разветвлением по выходу или многоуровневой реализации логики. (Рассмотрите, например, такие арифметические операции, как сложение и умножение.) Таким образом, можно разработать другие подходы, более согласующиеся с многоуровневым подходом, пользующимся благосклонностью у разработчиков, использующих методы стандартных элементов и "моря" элементов.

Существует множество способов разработки логического блока, который можно настроить на реализацию широкого диапазона логических функций. Как вариант, *в качестве генераторов функции можно использовать мультиплексоры*. Рассмотрим мультиплексор с двумя входами, изображенный на рис. 8.31 и реализующий логическую функцию F:



а

Конфигурация			
A	***B***	***S***	***F*** =
0	0	0	0
0	X	1	X
0	Y	1	Y
0	Y	X	XY
X	0	Y	$X\overline{Y}$
Y	0	X	$\overline{X}Y$
Y	1	X	$X+Y$
1	0	X	$\overline{X}$
1	0	Y	$\overline{Y}$
1	1	1	1

б

Рис. 8.31. Использование мультиплексора с двумя входами (*а*) в качестве настраиваемого логического блока. Правильно соединяя входы A, B и S с выходными переменными X или Y либо с сигналами 0 или 1, можно получить 10 различных логических функций (*б*)

$$F = A \cdot \bar{S} + B \cdot S. \tag{8.3}$$

Правильно выбирая соединения между переменными X и Y и входными портами A, B и S мультиплексора, мы можем запрограммировать устройство на выполнение полезных логических операций с одним или несколькими входными сигналами (см. рис. 8.31).

Объединяя несколько мультиплексоров, можно получать более сложные логические элементы. Рассмотрим, например, логическую ячейку, показанную на рис. 8.32, которая используется в семействе Actel ACT матриц FPGA. Ячейка состоит из трех мультиплексоров с двумя входами и двух элементов НЕ-ИЛИ с двумя входами. Ее можно запрограммировать на реализацию любой логической функции с двумя или тремя входами, некоторых булевых функций с четырьмя входами и защелки.

Пример 8.9. Программируемая логическая ячейка

Можно доказать, что запрограммировав описанным ниже способом логическую ячейку (рис. 8.32), мы превратим ее в элемент ИСКЛЮЧАЮЩЕЕ ИЛИ с двумя входами. Предположим, что мультиплексоры выбирают нижний входной сигнал при высоком управляющем сигнале. В таком случае имеем следующее:

$$A = 1; \quad B = 0; \quad C = 0; \quad D = 1; \quad SA = SB = In1; \quad S0 = S1 = In2.$$

В качестве упражнения определите, как нужно запрограммировать ячейку, чтобы получить функцию ИСКЛЮЧАЮЩЕЕ НЕ-ИЛИ с двумя входами. Элемент И с тремя входами можно реализовать следующим образом:

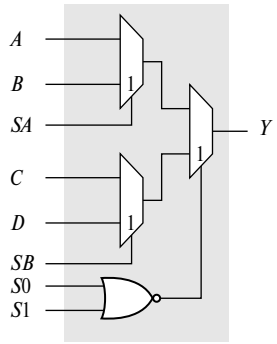


Рис. 8.32. Логическая ячейка (используется в FPGA с плавкими перемычками от Actel)

$$A = 0; \quad B = In1; \quad C = 0; \quad D = 0; \quad SA = In2; \quad SB = 0; \quad S0 = S1 = In3.$$

Наконец, наибольшей функцией, которую можно реализовать с помощью приведенной схемы, является мультиплексор с четырьмя входами. Сигналы A, B, C и D используются как входы, а SA, SB и $(S0 + S1)$ являются управляющими сигналами.

Подход "мультиплексор как функциональный блок" предлагает возможность настройки путем программирования межсоединений. Совершенно другая стратегия принята в методе *таблицы соответствий* (lookup table — LUT). Чтобы для конкретной функции сконфигурировать полностью программируемый модуль с разветвлением по входу i, программируется запоминающее устройство (называемое таблицей соответствий), в котором записана таблица истинности данной функции. Входные переменные выступают как управляющие входы мультиплексора, который и выбирает соответствующее значение из памяти. Описанный подход иллюстрируется на рис. 8.33 для ячейки с двумя входами. Чтобы реализовать функцию ИСКЛЮЧАЮЩЕЕ ИЛИ, в таблицу соответствий загружается входной столбец таблицы истинности функции ИСКЛЮЧАЮЩЕЕ ИЛИ, т.е. "0 1 1 0". При подаче на вход значения "0 0" мультиплексор выбирает первое значение таблицы ("0") и т.д. Благодаря этому подходу можно реализовать любую логическую функцию с двумя входами — для этого достаточно лишь перепрограммировать память.

Точно так же, как и при использовании подхода, основанного на мультиплексорах, мы можем создавать и более сложные логические элементы. Для этого нужно либо объединить несколько таблиц соответствий, либо увеличить их размеры, либо использовать оба подхода. Кроме того, дополнительные функциональные возможности можно получить, вводя в структуру триггеры.

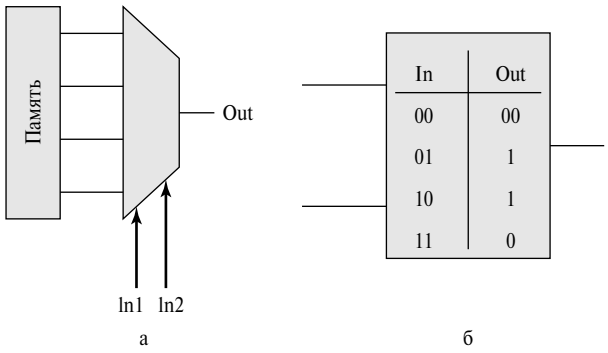


Рис. 8.33. Настраиваемая логическая ячейка, основанная на таблице соответствий: *а* — схематическое изображение; *б* — реализация функции ИСКЛЮЧАЮЩЕЕ ИЛИ

Пример 8.10. Логическая ячейка, основанная на таблицах соответствий

На рис. 8.34 показана базовая ячейка, именуемая *настраиваемым логическим блоком* (Configurable Logic Block — CLB) и используемая в серии FPGA Xilinx 4000 [Xilinx4000]. В ней объединены две таблицы соответствий с четырьмя входами, выходы которых подаются на таблицу соответствий с тремя входами. Кроме того, ячейка содержит два триггера, входами которых могут быть любые выходы таблицы соответствий F, G или H, а также дополнительный внешний вход D_{in}; выходы триггеров снимаются с выходных контактов XQ и YQ. Выходы X и Y таблиц соответствий позволяют создавать более сложные комбинационные функции. Ячейка также имеет четыре дополнительных входа ($C1...C4$), которые можно использовать либо как входы, либо как сигналы установки/сброса и запуска синхронизатора триггеров.

Программируемые межсоединения

До этого момента мы сравнительно подробно обсудили, как сделать логику программируемой. В связи с этим возникает также вопрос, как сделать изменяемыми или программируемыми соединения, связывающие эти логические элементы. Чтобы использовать все возможности доступных логических элементов, соединительная сеть должна быть гибкой и не содержать узких мест, снижающих общее быстродействие схемы. Другим необходимым параметром является скорость, поскольку при использовании описываемого стиля разработки фактором, определяющим быстродействие, часто выступает задержка соединения. В то же время следует помнить, что платой за программируемое соединение является значительное увеличение площади и потребляемой мощности. Фактически основное рассеяние мощности в архитектурах, допускающих программирование пользователем, происходит за счет связующей сети [George01].

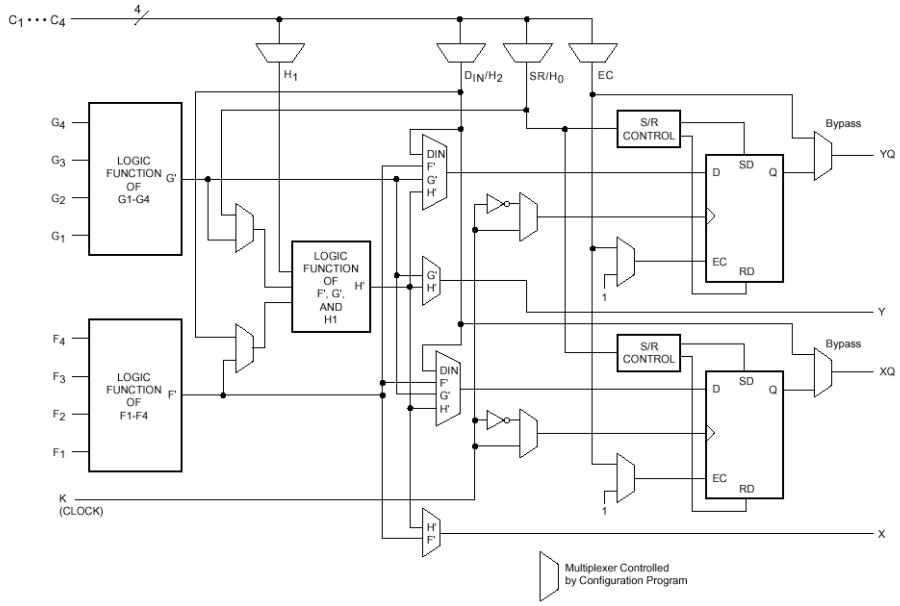


Рис. 8.34. Упрощенная блок-схема настраиваемого логического блока серии XC4000 (ОЗУ и функции логического переноса не показаны) [Xilinx4000]

Повторимся, мы должны различать программирование с помощью фотошаблона, однократное программирование и перепрограммирование. Кроме того, желательно различать локальные межсоединения (между ячейками) и глобальные сигналы (например, тактовые импульсы), которые необходимо распространить по всему кристаллу с низкой задержкой. На локальном уровне программируемую проводку можно разделить на два основных класса: матрицы и трассировка пересечений каналов.

Программируемая проводка на основе матриц

В этом подходе проводка группируется в трассировочные каналы, каждый из которых содержит полную сеть горизонтальных и вертикальных проводов. Таким образом, соединительный провод можно запрограммировать в структуру, замкнув накоротко некоторые точки пересечения горизонтальных и вертикальных проводов (рис. 8.35). Сказанное можно реализовать, добавив в каждую точку пересечения проходной транзистор. Замыкание соединения означает повышение управляющего сигнала — запись значения “1” в соответствующую ячейку памяти M (рис. 8.36). Данный подход относится к разряду дорогих, и его использование не рекомендуется, поскольку он требует большого числа транзисторов и управляющих сигналов. Кроме того, большое число транзисторов, соединенных с каждым проводом, приводит к большому разветвлению по выходу, а это, в свою очередь, ведет к увеличению задержки и потребления мощности. Более эффективным программируемым соединением является плавкая перемычка. В таком случае каждый трассировочный канал представляет собой полносвязную сеть горизонтальных и вертикаль-

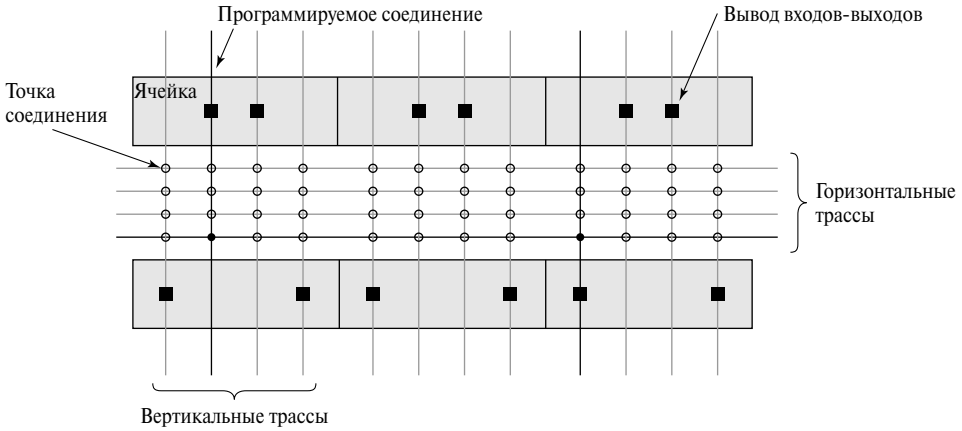


Рис. 8.35. Программируемая проводка, созданная по принципу матрицы

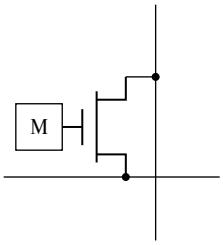


Рис. 8.36. Программируемая точка соединения. Состояние соединения управляется ячейкой памяти. Записанные значения 0 и 1 означают разомкнутую и замкнутую цепь соответственно. Ячейка памяти может быть энергонезависимой (электронно-перепрограммируемое ПЗУ) или энергозависимой (статическое ОЗУ)

ных соединительных проводов; если какое-то из соединений не требуется, соответствующая перемычка просто пережигается. К сожалению, соединительные сети обычно очень скудно заполнены, что требует разрыва очень большого числа соединений, из-за чего программирование занимает довольно много времени.

Чтобы обойти эту проблему, можно использовать *антипрожигаемую перемычку* (см. рис. 8.26). Антипрожигаемые перемычки нужно активизировать только тогда, когда в трассировочном канале нужно соединение (а это — малая часть общей сети). Обратите внимание на рис. 8.35, где для установки требуемого соединения достаточно пережечь всего две антипрожигаемых перемычки. Учтите, что на рисунке не показана схема программирования. Напомним также, что операция пережигания является необратимой. В целом использование проводки в форме матриц стало наиболее удачным подходом к однократному программированию соединений FPGA. Тем не менее он не допускает исправление или расширение схемы, и каждое изменение структуры требует новых компонентов. Следовательно, реальная возможность (пере)программирования требует более эффективной стратегии трассировки.

Программируемая проводка на основе пересечения каналов

Осознав, что полносвязная сетка проводки является чересчур избыточной, мы можем представить более эффективный подход к программируемой проводке. Ограничив ресурсы трассировки и число точек пересечения, мы по-прежнему сможем управлять проводкой в требуемых межсоединениях, при этом существенно снизив общие служебные издержки. Недостаток данного подхода состоит в том, что иногда трассировка межсоединения невозможна. Чаще всего это можно исправить путем преобразования структуры, например, выбрав для реализации данной функции другую группу логических ячеек.

Большое число локальных межсоединений можно учесть, сформировав сеть, подобную той, что создается для соединения соседних ячеек. Например, выход каждой логической ячейки (Logic Cell — LC) можно распространить на соседние ячейки — северную, восточную, южную и западную. Чтобы учесть соединения непересекающихся ячеек или обеспечить глобальные соединения, между ячейками, содержащими фиксированное число нескоммутированных вертикальных и горизонтальных трассировочных проводников, располагаются трассировочные каналы (рис. 8.37). На пересечении горизонтальных и вертикальных проводников создаются ОЗУ-программируемые коммутационные матрицы (S-блоки), которые управляют трассировкой данных. Входы и выходы ячеек соединяются с глобальное соединительной сетью посредством RAM-программируемых точек пересечения (C-блоки). Схематически сказанное представлено на рис. 8.38, где показана реализация на уровне транзисторов коммутационных и соединительных блоков. Не забудьте, что реализация коммутатора на одном проходном транзисторе сопровождается падением порогового напряжения. Соответствующее уменьшение размаха сигнала выгодно с точки зрения потребляемой мощности, но отрицательно сказывается на быстродействии. Соответственно могут потребоваться специальные технологии проектирования — устройства с нулевым порогом, восстановители уровня или усилители управляющих сигналов.

Архитектура типа “сеть” представляет собой гибкое и масштабируемое средство соединения большого числа компонентов. Это довольно эффективно для локальных соединений, когда число коммутаторов, проходимых одним межсоединением, мало, и, кроме того, мало разветвление по выходу. Тем не менее сеть не очень хорошо подходит для глобальных межсоединений. Задержка, порождаемая большим набором коммутаторов и большой емкостной нагрузкой, становится непомерной. Таким образом, большинство архитектур FPGA, основанных на сетях, предлагают альтернативные ресурсы, которые могут повысить эффективность глобальной проводки. Один из подходов к решению этой задачи показан на рис. 8.39. Помимо стандартной проводки, соединяющей S-блоки, сеть также включает проводники, соединяющие S-блоки, отстоящие на два блока друг от друга. Исключение одного S-блока из межсоединения уменьшает сопротивление. Подобным образом мы можем включить длинные провода, которые соединяют каждый четвертый, восьмой или шестнадцатый S-блок. По сути, мы создаем несколько перекрывающих-

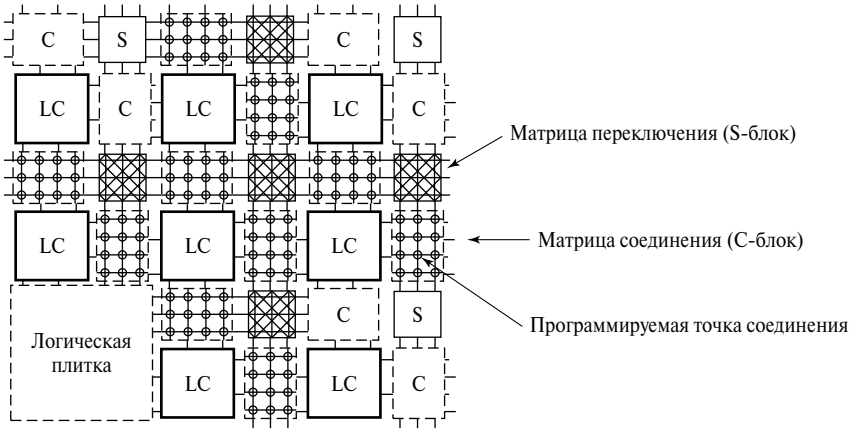


Рис. 8.37. Программируемая соединительная сеть (*Перепечатано с разрешения Андре Дегона (Andre Dehon) и Джона Вавржинека (John Wawrzyniek)*).

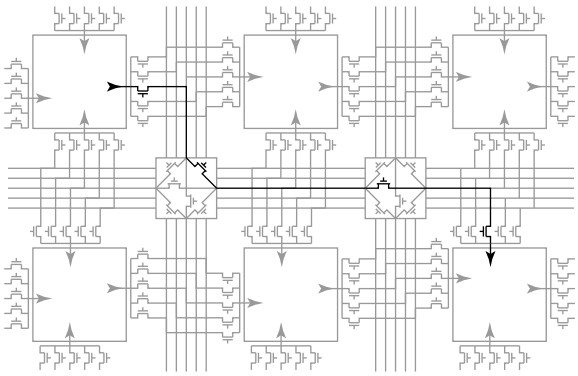


Рис. 8.38. Принципиальная схема программируемой сети разводки на уровне транзисторов (*Перепечатано с разрешения Андре Дегона (Andre Dehon) и Джона Вавржинека (John Wawrzyniek)*.)

ся сетей с разной зернистостью (одинарный шаг, двойной шаг и т.д.). Самые длинные проводники отображаются в сеть с наибольшим шагом.

Собираем все в единое целое

Итак, теперь мы можем собрать полностью программируемую матрицу логических элементов, объединив подходы с использованием логических элементов и межсоединений. Реализовать это можно с помощью множества альтернативных архитектур. При выборе конкретной реализации прежде всего нужно выбрать стиль конфигурации (однократная запись, энергонезависимое устройство, энергозависимая память). Этот выбор налагает определенные ограничения на типы ячеек и межсоединений, которые можно использовать. Полный обзор всех возможных вариантов мы приводить не будем,

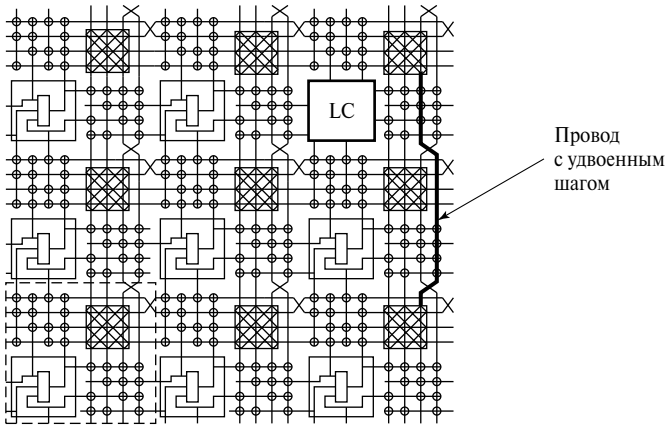


Рис. 8.39. Программируемая соединительная архитектура на основе сети, на которую наложена еще одна сеть 2×2 (*Перепечатано с разрешения Андре Дегона (Andre Dehon) и Джона Вавржинека (John Wawrzyniek).*)

ограничимся только двумя популярными (и показательными) архитектурами. Более подробную информацию по этому вопросу можно найти в [Trimberger94], [Smith97], [Betz99] и [George01].

Серия Altera MAX [Altera01]

Семейство устройств MAX (рис. 8.40) относится к классу энергонезависимых FPGA (часто называемых также *EPLD — Electrically Programmable Logic Devices*, логические устройства с электрическим программированием). В качестве основного логического модуля в данных структурах применяется модуль ПМЛ (см. рис. 8.29). Этот модуль (на языке Altera он называется LAB — Logic Array Block — блок логической матрицы) незначительно отличается для разных представителей семейства: широкая программируемая матрица И, за которой следует узкая фиксированная матрица ИЛИ и программируемая инверсия. Обычно блок LAB содержит 16 макроячеек.

Основные отличия логических устройств связаны с архитектурой межсоединений блоков LAB. Наименьшие устройства (MAX5000, MAX7000) используют архитектуру трассировки, основанную на концепции матрицы. Основная магистраль трассировочного канала формируется выходами всех макроячеек, дополненного прямыми входами кристалла. Эти входы могут соединяться с входами блоков LAB через программируемые точки коммутации. Достоинство этой архитектуры, называемой *программируемой матрицей межсоединений* (Programmable Interconnect Array — PIA), заключается в ее простоте; кроме того, задержка трассировки сигнала между блоками является предсказуемой и фиксированной (рис. 8.41). Недостатком архитектуры является не очень удобное масштабирование. Именно из-за этого более крупные представители серии (MAX9000) прибегают к помощи другой схемы. Когда число макроячеек достигает значения 560, одноканальный поток “выдыхается” и становится чересчур медленным. В таких случаях вместо него применяется архитектура, основанная на сетке. Отдельные макроячейки мо-

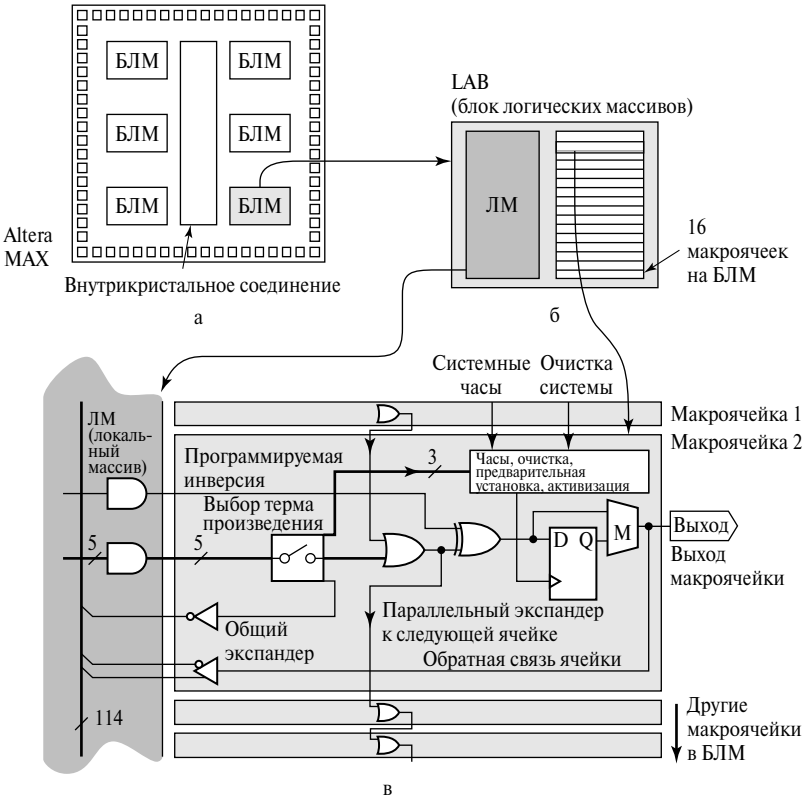


Рис. 8.40. Архитектура Altera MAX: *а* — организация логики и межсоединений; *б* — модуль LAB; *в* — макроячейка из семейства MAX. Расширители увеличивают число доступных произведений путем дополнительного прохода через логическую матрицу (согласно [Smith97])

гут соединяться и с каналами-"строками", и с каналами-"столбцами", которые могут быть достаточно широкими (от 48 до 96 проводов).

Подход EPLD позволяет объединять до 15 тысяч логических элементов; обычно он используется, когда требуется высокое быстродействие. Когда требуется реализовать более сложные функции, необходимо задействовать другие архитектуры.

Серия Xilinx XC40xx

В данном популярном семействе ОЗУ-программируемых устройств объединяются концепция таблицы соответствий (для реализации логических элементов) и соединительная сеть. Наибольший элемент серии (XC4085) поддерживает почти 100 тысяч логических элементов, используя для этого массив CLB размером 56×56. Архитектура CLB была показана на рис. 8.34. Здесь стоит также отметить, что CLB можно сконфигурировать как массив запоминающих устройств чтения/записи, использующий таблицы соответствий в блоках F' и G'. В зависимости от выбранного режима один блок CLB можно настроить как массив 16×2, 32×1 или 16×1 бит. Данная особенность полез-

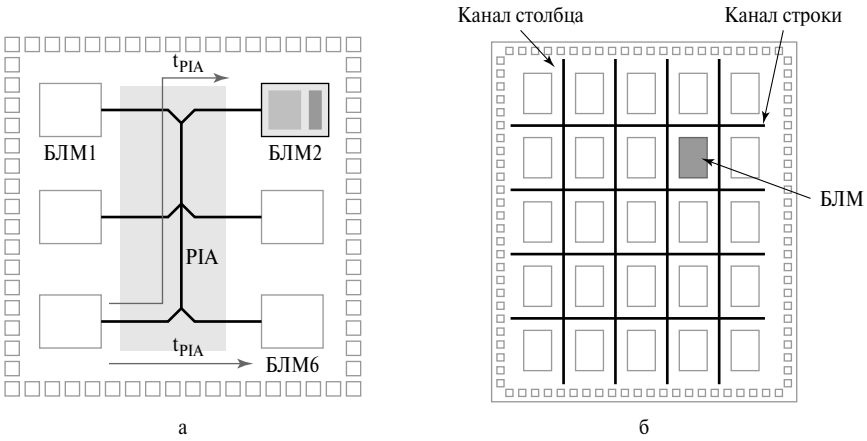


Рис. 8.41. Архитектура межсоединений, используемая в серии Altera MAX: *а* — архитектура на основе матрицы, применяющаяся в MAX 3000-7000; *б* — архитектура на основе сети, применяющаяся в MAX9000

на, особенно из-за того, что большие модули логических элементов обычно требуют сравнимых объемов памяти.

Архитектура межсоединений также достаточно богата и объединяет в себе несколько типов проводки, как показано на рис. 8.42. Наложенная сеть состоит из сегментов проводов длины 1, 2 и 4. Некоторые компоненты также поддерживают прямые соединения, которые связывают соседние блоки CLB без использования глобальной проводки. Сигналы, проходящие по прямым соединениям, испытывают минимальную задержку, поскольку в этом случае разветвление по выходу мало. Такие *прямые соединения* особенно эффективны в реализациях быстрых арифметических модулей, обладающих большим числом критических локальных соединений. Чтобы обеспечить глобальные межсоединения, предусмотрены *длинные линии*, которые формируют сеть металлических соединительных сегментов, идущих по всей длине или ширине матрицы. Данные линии предназначены для больших разветвлений по выходу, чувствительных ко времени сигнальных сетей или сетей, развернутых на большой площади (например, шин). Кроме того, предоставляются специальные проводники для трассировки тактовых синхроимпульсов.

Отметим, что при обсуждении настраиваемых структур мы еще не рассматривали архитектуру ввода-вывода. Для обеспечения максимальной практичности необходимо, чтобы выводы входов-выходов компонента были гибкими и обеспечивали широкий диапазон возможностей с точки зрения направления, логических уровней и напряжения питания. На рис. 8.43 показан один возможный стиль блока ввода-вывода (Input/Output Block — IOB), используемый в серии XC4000. Его можно запрограммировать как порт ввода, вывода или двунаправленный порт. Он содержит триггер, который можно запрограммировать в качестве триггера, активизируемого фронтом, или чувствительного к уровню. Блок управления скоростью нарастания выходного напряжения обеспечивает различные напряжения питания и позволяет уменьшить время нарастания и спада некритических сигналов.

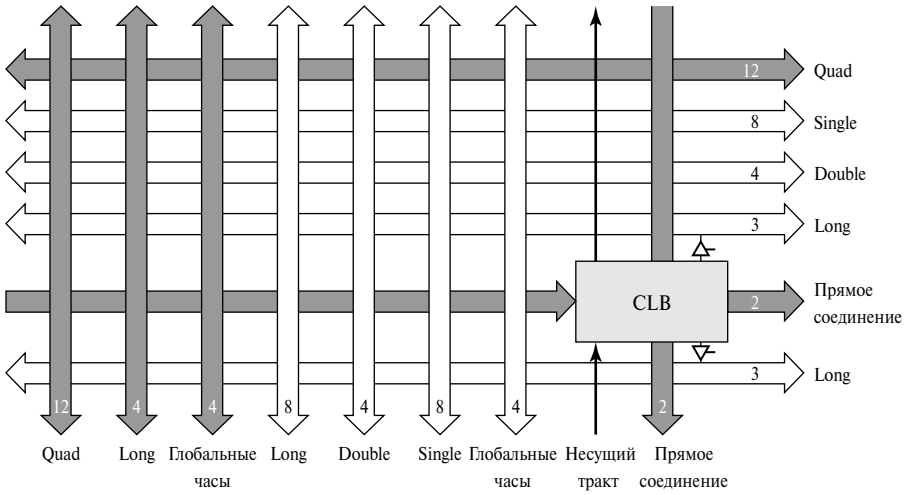


Рис. 8.42. Архитектура межсоединений серии Xilinx XC4000. Числа, указанные на диаграмме, обозначают количество соответствующих ресурсов

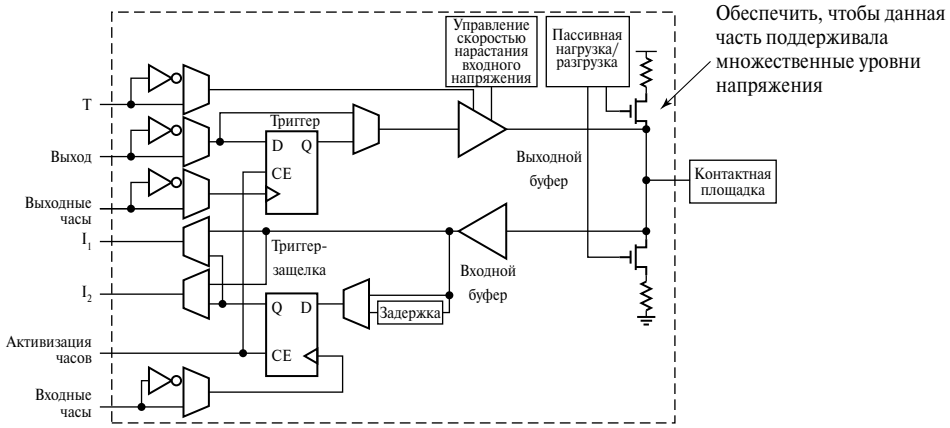


Рис. 8.43. Программируемый блок ввода-вывода серии XC4000

Пример 8.11. Сложность и быстродействие FPGA

Чтобы представить, чего можно добиться с помощью энергозависимых программируемых пользователем компонентов, рассмотрим устройство Xilinx 4025. Оно состоит приблизительно из 1000 блоков CLB, организованных в массив 32×32. Приблизительно это соответствует схеме с 25 тысячами логических элементов. Кристалл содержит 422 Кбит ОЗУ, используемого преимущественно для программирования. Отдельный блок CLB может функционировать на частоте 250 МГц. Если учитывать соединительную сеть и попытаться получить более сложные логические конфигурации (например, сумматоры), можно добиться частот порядка 20–50 МГц. Чтобы представить себе сложность интеграции, считайте, что 32-битовый сумматор требу-

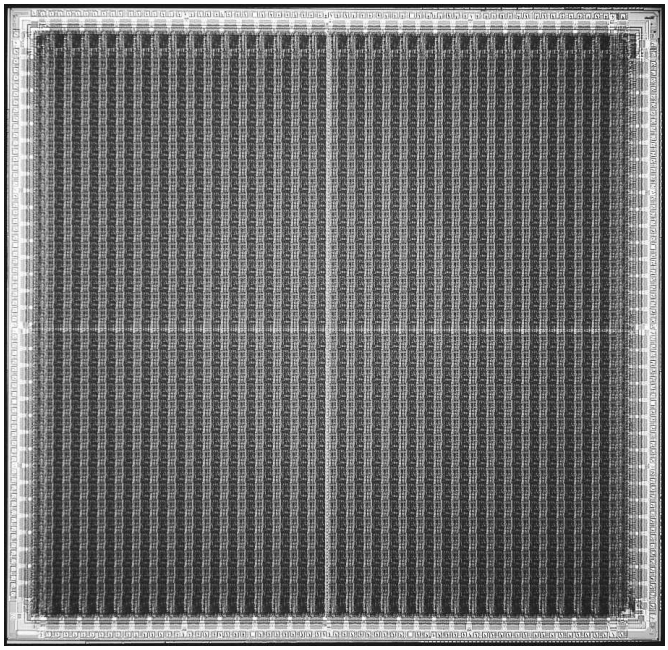


Рис. 8.44. Микрофотография кристалла энергозависимой FPGA XC4025 (Перепечатано с разрешения Xilinx, Inc.)

ет приблизительно 62 блоков CLB. На рис. 8.44 показана микрофотография кристалла XC4025, на которой легко прослеживают горизонтальные и вертикальные трассировочные каналы.

Логические матрицы с предварительным монтажом быстро захватили значительную часть рынка логических компонентов. С их появлением завершилась эпоха проектирования логических схем с помощью дискретных компонентов (представленная семейством ТТЛ-схем). Существует распространенное мнение, что влияние этих компонентов будет увеличиваться при дальнейшем масштабировании технологии. Тем не менее для успеха данного подхода требуется нетривиальная программная поддержка, облегчающая размещение ячеек, трассировку сигналов и синтез. Кроме того, не следует недооценивать служебные издержки, сопутствующие гибкости. Программируемая логика по крайней мере в 10 раз менее эффективна с точки зрения потребляемой энергии и быстродействия, чем ASIC[2]-решения. Следовательно, пока что ее сфера применения — это разработка прототипов и мелкомасштабных приложений. Впрочем, гибкость и возможность повторного использования все же весьма привлекательны. Таким образом, в ближайшие годы компоненты, программируемые пользователем, наверняка будут разрабатываться и использоваться весьма интенсивно.

---

[2] Applications Specific Integrated Circuit — специализированная интегральная схема.

8.6. Перспектива — платформа реализации будущего

Разработчик современных передовых систем на кристалле может выбирать из множества доступных концепций реализации. В конечном счете выбор подхода определяется большим набором факторов:

- соображения быстродействия, мощности и стоимости;
- сложность проекта;
- возможность тестирования
- время выхода на рынок;
- неопределенность рынка или поздние изменения проекта;
- диапазон приложений, охваченных проектом;
- наличие у команды опыта подобных разработок.

На первый взгляд кажется, что некоторые из указанных факторов аргументируют за более гибкие программируемые компоненты, которые можно использовать повторно и модифицировать даже после производства. В то же время чаще всего победителями становятся более сбалансированные решения. Слишком большая гибкость часто приводит к неэффективным и дорогим решениям, которые быстро уходят с рынка. Самой сложной задачей современной разработки является нахождение разумного компромисса между двумя крайностями.

Исходя из этого, логично предположить, что платформа реализации, которая появится в будущем, будет комбинировать в себе все рассмотренные стратегии, обеспечивая эффективность и гибкость реализации, когда и где это нужно. Система на кристалле объединяет встроенные микропроцессоры с подсистемами памяти, процессоры для цифровой обработки сигналов, жесткие аппаратные ускорители в стиле ASIC, настраиваемые модули и гибкие логические элементы, реализованные в стиле FPGA. Способ объединения этих составляющих определяется требованиями приложения и рынком.

Пример 8.12. Примеры гибридных платформ реализации

На рис. 8.45 показаны две платформы реализации для беспроводных приложений. Первое устройство, Virtex-II Pro от Xilinx [XilinxVirtex01], построено на базе большой матрицы FPGA. В центр матрицы встроен микропроцессор PowerPC, обеспечивающий эффективную реализацию функций приложения и внешнего контроля. Чтобы обеспечить более высокое быстродействие приложений обработки сигналов, добавлен массив умножителей 18×18. Данный специализированный компонент предлагает высокое быстродействие, более выгоден с точки зрения потребляемой мощности и занимаемой площади, чем чистая FPGA-реализация той же самой функции. Наконец, в схеме присутствует несколько 3,125-гигагерцовых трансиверов, что позволяет реализовать высокоскоростную последовательную связь с кристаллом.

Рис. 8.45. Примеры гибридных платформ реализации

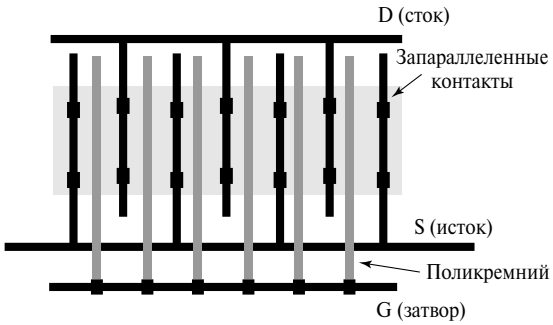


Рис. 9.9. Для реализации очень широких транзисторов удобно разместить параллельно несколько меньших транзисторов

вызванное внутренними емкостями, уменьшается до незначительного уровня. Общее рассеяние энергии буфера и нагрузки уменьшается на 24% (это достаточно много для данного размера емкостной нагрузки). В конце концов, задача разработчика — **спроектировать схему, работающую с заданной, а не с максимальной скоростью**!

Реализация широких транзисторов

Даже полученный перепроектированный буфер требует широких транзисторов, а при разработке таких устройств следует соблюдать особую осторожность, поскольку большое значение W приводит к очень длительной организации связи между элементами. Длинные поликремниевые соединения обычно имеют высокое сопротивление, которое ухудшает быстродействие переключений. Данную проблему можно решить следующим образом: создать широкий транзистор, соединив параллельно несколько меньших транзисторов (рис. 9.9).

Сопротивление затвора уменьшается с помощью низкоомной металлической обходной перемычки, соединяющей закорачивающие поликремниевые участки. На рис. 9.10 приведен пример задающего устройства контактной площадки, реализованного с использованием описанных техник. Решая, стоит ли разбивать очень широкий транзистор на меньшие устройства, используйте следующее эмпирическое правило: RC-задержка поликремниевого затвора должна быть существенно меньше, чем задержка переключения всего цифрового элемента.

Разработка надежных входных и выходных площадок

Разработка задающих устройств контактных площадок очевидно является критичной и нетривиальной задачей, которая дополнительно усложняется соображениями, касающимися шума и надежности. Например, большие переходные токи, возникающие при переключении огромных выходных емкостей, могут привести к “защелкиванию” выхода переключателя. Данного деструк-

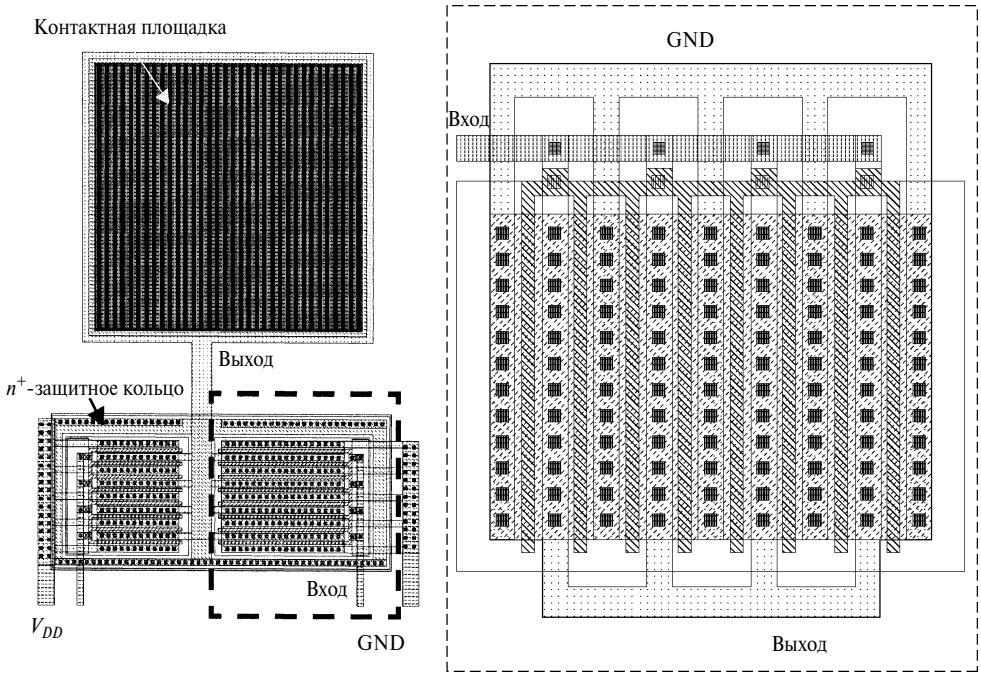


Рис. 9.10. Схема последнего каскада задающего устройства контактной площадки. Справа представлено увеличенное изображение n-МОП-транзистора, расположенного между шинами GND и Out

тивного эффекта позволяет избежать большое число контактов с карманом и подложкой с защитными кольцами.

Защитные кольца представляют собой заземленные p^+-диффузионные области в p-карманах и соединенные с источником питания n^+-диффузионные области в n-карманах, которые применяются для сбора инжектированных неосновных носителей заряда перед тем, как они достигнут базы паразитных биполярных транзисторов. Данные кольца должны окружать n-МОП- и p-МОП-транзисторы последнего каскада выходного задающего устройства контактной площадки.

Разработчик входной контактной площадки сталкивается немного с другими проблемами. Вход первого каскада входного буфера непосредственно соединяется с цепью внешней нагрузки, а следовательно, он чувствителен к любым колебаниям напряжения в связанных с ним входных контактах. Человек, проходящий по синтетическому ковру при относительной влажности воздуха порядка 80% или более, может накапливать потенциал порядка 1,5 кВ (возможно, вы помните, как в таких случаях с ваших пальцев срываются искры при контакте с металлическими предметами). То же самое справедливо и для сборных механизмов. Выход затвора МОП-транзистора имеет очень большое входное сопротивление (10^{12}–10^{13} Ом). Напряжение, при котором пробивается оксидный слой затвора, равно порядка 10–20 В и становится меньше при уменьшении толщины оксида. Следовательно, человек или

машина, заряженные до высокого статического потенциала, могут легко вызвать пробой входных транзисторов при соединении с входными контактами. В процессе производства и сборки данное явление, называемое *электростатическим разрядом*, смертельно опасно для многих схем.

Для ограничения потенциально разрушительного напряжения применяется комбинация сопротивления и диодной фиксации уровня. На рис. 9.11, *а*, например, показана типичная *схема электростатической защиты*. Защитные диоды D_1 и D_2 включаются, когда напряжение на узле X превышает V_{DD} или опускается ниже земли. Резистор R используется для ограничения пикового тока, проходящего через диод в случае необычных колебаний напряжения. В настоящее время для реализации резисторов разработчики предпочитают использовать резисторы с p-диффузионной областью в n-кармане и n-диффузионной областью в p-кармане; типичное сопротивление в таких случае может изменяться от 200 Ом до 3 кОм. Разработчик должен помнить, что получающаяся постоянная времени может ограничивать быстродействие в высокоскоростных схемах. Топология типичной входной площадки показана на рис. 9.11, *б*.

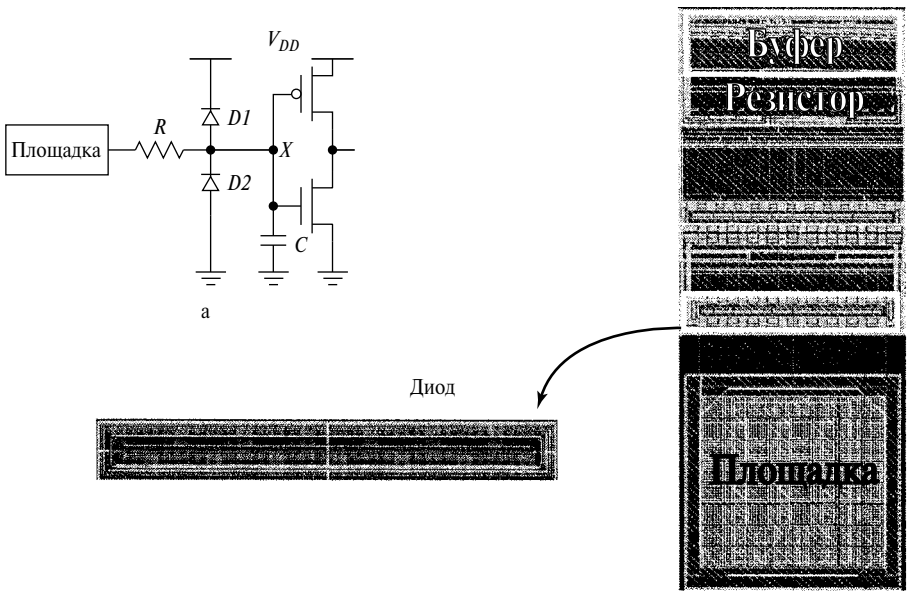


Рис. 9.11. Схема защиты входа (*а*) и топология входной контактной площадки (*б*)

Похоже, что с появлением каждой новой технологии задача разработчика контактных площадок только усложняется. К счастью, существует несколько современных технологий корпусирования, уменьшающих емкость

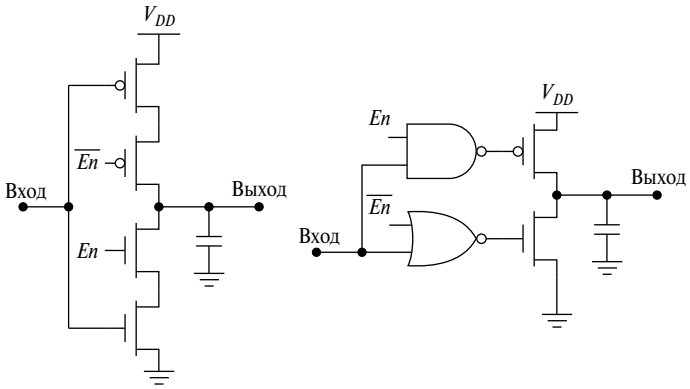


Рис. 9.12. Две возможные реализации буфера с тремя состояниями. Сигнал $E_n = 1$ включает буфер

вне кристалла. Например, такие технологии, как *матрицы шариковых выводов* и кристаллы на плате (см. главу 2), помогают существенно снизить требования к задающим устройствам вне кристалла.

Некоторые интересные схемы задающих генераторов

До этого момента большинство наших задающих генераторов представляли собой простые инверторы. Иногда бывают нужны и другие функции, например *буфер с тремя состояниями*. Кроме того, в большинстве цифровых систем необходимы *шины* — общие пучки проводов, соединяющие набор передающих и принимающих устройств (процессоров, запоминающих устройств, дисков и устройств ввода-вывода). Когда одно устройство передает информацию через шину, все остальные передающие устройства должны отключаться. Такую схему можно реализовать, поместив выходные буферы данных устройств в высокоомное состояние Z, которое эффективно отключает элемент от выходного провода. Описанный буфер имеет три состояния — 0, 1 и Z, — отсюда и его название.

Реализация на КМОП инвертора с тремя состояниями не вызывает особых трудностей. При одновременном включении n-МОП- и p-МОП-транзистора получается плавающий выходной узел, отсоединенный от входа. Для примера на рис. 9.12 показаны две возможные реализации буфера с тремя состояниями. Первая реализация достаточно проста[2], а вторая лучше подходит для управления большими емкостями. Из-за возможного огромного увеличения площади наличия в выходном каскаде многоярусных транзисторов следует избегать.

[2] Обратите внимание на то, как буфер с тремя состояниями похож на C^2MOS-защелку.

9.3. Паразитное сопротивление

9.3.1. Сопротивление и надежность — омическое падение напряжения

Ток, проходящий через резистивный провод, приводит к омическому падению напряжения, которое ухудшает логические уровни сигнала. Особенно это важно в схемах разводки питания, где сила тока может легко доходить до нескольких ампер, как показано на рис. 9.13 для семейства процессоров Compaq (ранее — Digital Equipment Corporation) Alpha.

Рассмотрим теперь провод V_{DD} или GND длиной 2 см, по которому проходит ток силой 1 мА на 1 мкм ширины. Такой ток близок к максимальному, который может выдержать алюминиевый провод (это связано с *электромиграцией*; см. следующий раздел). Предполагая, что удельное поверхностное сопротивление слоя равно 0,05 Ом/□, сопротивление данного провода (на 1 мкм ширины) равно 1 кОм. Ток 1 мА/мкм приведет к падению напряжения в 1 В. Изменившееся значение напряжения источника питания уменьшает запас помехоустойчивости и изменяет логические уровни в зависимости от расстояния от клемм источника питания. Сказанное можно продемонстрировать на примере схемы, изображенной на рис. 9.14, где инвертор расположен далеко от источника питания и выводы заземления расположены близко от источника питания. Разница в логических уровнях, вызванная падением напряжения (равно IR) в шинах питания, может частично включить транзистор M_1. Это может привести к случайной разрядке предварительно заряженного динамического узла X или к ненулевому потреблению мощности в статическом режиме, если подключенный элемент является статическим. Короче говоря, импульсы тока от внутрикристальных логических элементов и запоминающих устройств, а также контактов ввода-вывода может приводить к падению напряжения в схемах разводки питания и является основным источником внутрикристального шума в цепях питания. Помимо того что это угрожает надежности, падение напряжение на IR также влияет на

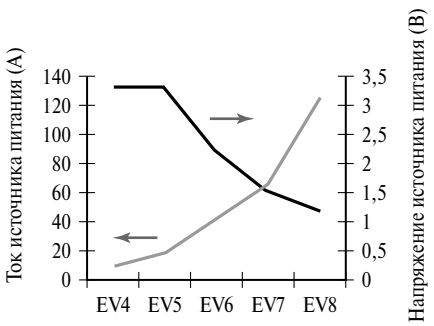


Рис. 9.13. Эволюция силы тока и напряжения источника питания для различных поколений семейств быстродействующих микропроцессоров Alpha от Compaq. Даже несмотря на снижение напряжения источника питания, суммарная сила тока в новых поколениях микропроцессоров превышает 100 А [Herrick00]

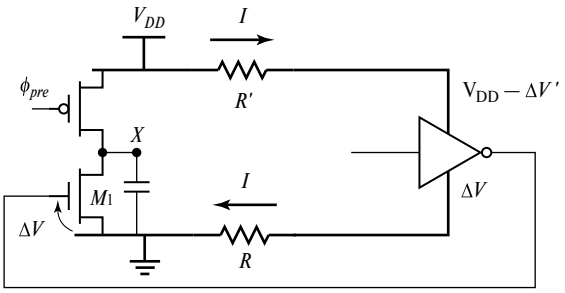


Рис. 9.14. Омическое падение напряжения в шинах питания снижает запас помехоустойчивости

быстродействие системы, поскольку небольшое падение напряжения источника питания может привести к существенному увеличению задержки.

Наиболее очевидным решением данной проблемы является уменьшение максимального расстояния между контактами источника питания и точкой подключения схемы к источнику питания. Легче всего этого добиться с помощью структурированной топологии схемы разводки питания. На рис. 9.15 для примера показано несколько внутрикристальных схем разводки питания с периферическими соединениями. Во всех решениях питание и земля подводятся к кристаллу посредством контактных площадок, расположенных по четырем сторонам кристалла. Выбор конкретного подхода зависит от числа грубых металлических слоев (т.е. толстых, с большим шагом, самых верхних металлических слоев), которые можно выделить под разводку питания. При первом подходе (*а*) питание и земля трассируются вертикально (или горизонтально) на одном и том же слое. Питание подводится с двух сторон кристалла. Локальные шины питания соединяются перемычками с этой верхней сетью, а затем трассируются на более низкие металлические слои. При втором подходе (*б*) для разводки питания применяются два грубых металлических слоя, и питание подводится с четырех сторон кристалла. Данный подход использовался в процессорах Alpha (Compaq) поколения EV5 [Herrick00], где сети разводки питания и распределения тактового сигнала занимали в общей сложности более 90% третьего и четвертого алюминиевых слоев. Более агрессивным решением является использование двух сплошных металлических плоскостей для распределения V_{dd} и GND (*в*). Преимуществом данного подхода является существенно уменьшение сопротивления сети. Металлические плоскости также действуют как экраны между слоями передачи данных, что приводит к снижению перекрестных помех. Кроме того, это помогает уменьшить внутрикристальную индуктивность. Разумеется, данный подход оправдан только при наличии достаточного числа металлических слоев.

Масштабирование сети электропитания является нетривиальной задачей. Электрическую сеть можно смоделировать как сеть резисторов (проводов) и источников тока (логические элементы), содержащих сотни миллионов элементов. Очень часто существует большое число маршрутов между контактами источника питания и контактами модуля кристалла или логического элемента. Хотя в общем случае ток проходит по пути наименьшего сопротив-

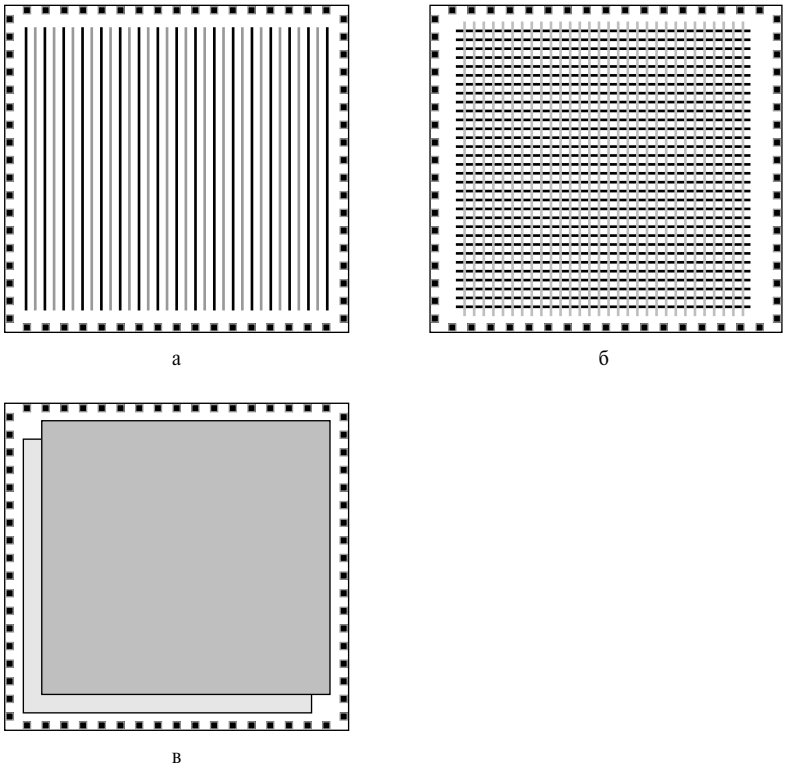
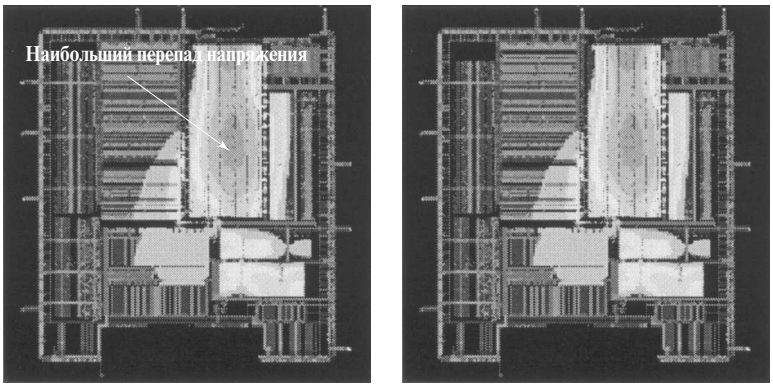


Рис. 9.15. Внутрикристальные цепи разводки питания: *а* — одноуровневая электрическая сеть; *б* — двухслойная сеть; *в* — двойная плоскость разводки питания

ления, точный поток зависит от таких факторов, как поток тока от соседних модулей, совместно с данным использующих одну и ту же сеть. Анализ дополнительно усложняется тем фактом, что пиковые токи, порождаемые присоединенными модулями, распределены во времени. Падение напряжения на IR — это динамическое явление! Наибольшую нагрузку на электрическую сеть могут вызывать одновременные переключения, например, инициированные тактовыми сигналами. “Спрос” на ток, поставляемый электрической сетью, обычно достигает своего максимума после тактовых переходов или при переключении больших задающих устройств. В то же время анализ наихудшего случая с учетом всех пиковых токов чаще всего приводит к грубому пересмотру размеров проводов.



а б

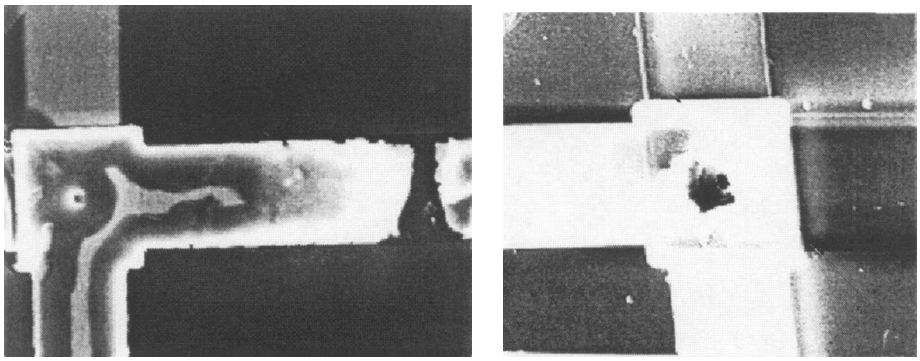
Рис. 9.16. Смоделированное падение напряжения в схеме разводки питания для двух вариантов сложной цифровой интегральной схемы. Введение дополнительной шины к верхним слоям энергосистемы (*б*) помогает существенно уменьшить падение напряжения в правом верхнем модуле по сравнению с первоначальным проектом (*а*) [Cadence-Power]

Таким образом, должно быть очевидно, что разработчику цепи разводки питания просто необходимы средства автоматизированного проектирования. Учитывая сложность современных интегральных схем, анализ на уровне транзисторов всех требований этой цепи нереален. В то же время разбиение задачи на меньшие составляющие может не дать точной картины. Изменения электрической сети в одном сегменте обычно имеют глобальное влияние. Сказанное иллюстрируется на рис. 9.16, где приводится смоделированное падение напряжения сложной цифровой схемы. Первая реализация (см. рис. 9.16, *а*) претерпевает падение напряжение, превышающее допустимый уровень, в правом верхнем модуле структуры, поскольку только верхняя часть сети электропитания питает верхние большие устройства. Нижняя часть модуля не соединена непосредственно с сеткой. Как показано на рис. 9.16, *б*, решению проблемы способствует добавление всего одной дополнительной шины.

Следовательно, точную картину падения напряжения нельзя получить, не проверив работу всего кристалла как единого объекта. Любое средство, используемое для этой цели, должно позволять анализировать миллионы резисторных сетей. К счастью, в настоящее время существует несколько эффективных и достаточно точных средств анализа электрических сетей [Cadence-Power, RailMill], объединяющих динамический анализ требований к току отдельных модулей схемы с подробным моделированием электросети. В общем-то, в современных условиях такие инструменты просто незаменимы.

9.3.2. Электромиграция

Плотность тока (ток на единицу площади) в металлическом проводе ограничена эффектом *электромиграции*. Постоянный ток в металлическом проводнике, проходящий на протяжении длительного периода времени, приво-



Обрыв цепи
а

Обрыв в контактном штепселе
б

Рис. 9.17. Повреждения, связанные с электромиграцией (*Перепечатано с разрешения Н. Ченга (N. Cheung) и А. Тао (A. Tao), Калифорнийский университет в Беркли).)*

дит к переносу ионов металла. В конечном счете это приводит к разрыву провода или короткому замыканию с другим проводом. Сбои подобного типа происходят только после того, как устройство поработает некоторое время. Несколько примеров повреждений, вызванных миграцией, показаны на рис. 9.17. Обратите внимание, что на первом рисунке отчетливо виден бугорок в направлении потока электронов.

Скорость электромиграции зависит от температуры, кристаллической структуры и средней плотности тока. Разработчик схемы может контролировать только последний фактор. Как правило, для предотвращения миграции достаточно удерживать ток в пределах ниже 0,5–1 мА/мкм. С помощью данного параметра можно определить минимальную ширину провода в сетях питания и заземления. По сигнальным шинам обычно передается переменный ток, и они менее чувствительны к миграции. Кроме того, двунаправленный поток электронов обычно аннулирует любой ущерб, нанесенный миграцией кристаллической структуре. Большинство компаний, основываясь на измерениях и прошлом опыте, предоставляют разработчикам несколько строгих директив относительно размеров проводов. Как показывают результаты исследований, многие из этих правил слишком консервативны [Tao94].

Правило проектирования

Эффекты электромиграции пропорциональны среднему потоку тока через провод, а омическое падение напряжения является функцией пикового тока.

На уровне технологии следует очень аккуратно следить за уменьшением риска миграции. Как вариант, к алюминию можно добавить какой-нибудь легирующий элемент (например, Cu или Tu), чтобы предотвратить движение ионов алюминия. Другой подход состоит в управлении неоднородностью

ионов. Например, очень помогает введение новых соединительных материалов — при использовании не алюминиевых, а медных межсоединений, предполагаемое время жизни провода увеличивается в 100 раз.

9.3.3. Сопротивление и быстродействие — RC*-задержка*

В главе 4 мы установили, что задержка провода квадратично растет с его длиной. Удвоение длины провода увеличивает его задержку в 4 раза! Следовательно, основной вклад в задержку сигналов, проходящих по длинным проводам, вносит *RC-эффект*. В современных условиях эта проблема становится еще острее из-за роста средних длин глобальных проводников (см. рис. 4.27), при том, что средняя задержка отдельных топологических элементов уменьшается. Это приводит к довольно странным ситуациям, когда для прохождения сигнала с одной стороны кристалла на другую требуется несколько тактов [Dally01]. В таких условиях основной задачей является обеспечение точной синхронизации и правильной работы схемы. В данном разделе мы обсудим несколько технологий разработки, которые могут уменьшить задержку, вызванную сопротивлением провода.

Лучшие соединительные материалы

Первым действием, направленным на уменьшение *RC*-задержек, является выбор лучших соединительных материалов, когда это возможно и целесообразно. Введение силицидов и меди помогает уменьшить сопротивление поликремниевых и металлических проводников соответственно, а использование диэлектрических материалов с меньшей диэлектрической проницаемостью уменьшает емкость. В современных КМОП-технологиях (начиная с 0,18-микронного поколения) и медь, и диэлектрики с низкой диэлектрической проницаемостью используются довольно широко. Тем не менее разработчику следует помнить, что новые материалы предлагают только временную передышку на одно-два поколения технологии и не решают основную проблему задержки больших проводов. Как правило, единственный способ справиться с подобными проблемами — это разработать совершенную новую технологию проектирования.

Пример 9.5. Влияние современных соединительных материалов

Удельное сопротивление меди примерно в 1,6 раза меньше, чем у алюминия (для чистых материалов; см. табл. 4.4). Из-за осаждения и других артефактов производства эффективное удельное сопротивление внутрикристальных медных проводов составляет приблизительно $2,210^{-8}$ Ом·м.

Отказаться от использования длинных поликремниевых проводов иногда бывает тяжело. Типичный пример такой ситуации — адресные шины в запоминающем устройстве, которые должны соединять большое число транзисторных элементов. Использование поликремниевых проводов существенно

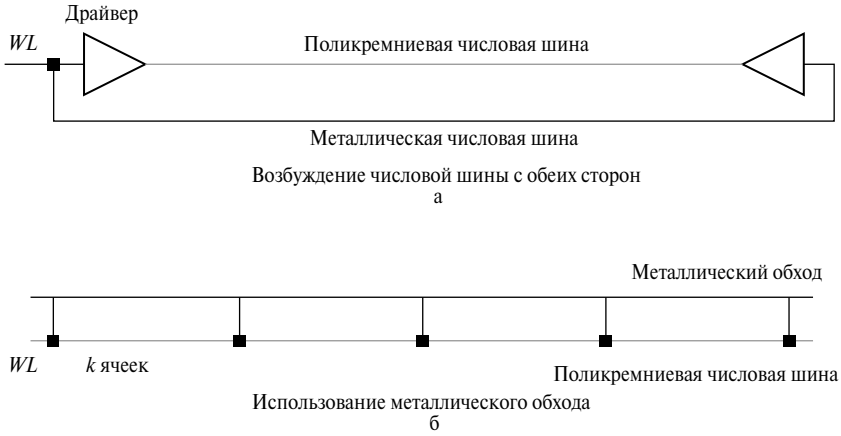


Рис. 9.18. Подходы к уменьшению задержки адресной шины

повышает плотность запоминающего устройства, поскольку не требует дополнительных металлических контактов. К сожалению, использование только поликремния приводит к чрезмерной задержке распространения. Одним из возможных решений этой проблемы является питание адресной шины с обоих концов (рис. 9.18, *а*). Таким образом, наихудшая задержка эффективно уменьшается в 4 раза. Другая возможность заключается в использовании дополнительного металлического провода, называемого *обходом*, который располагается параллельно поликремниевому проводу и соединяет его с каждой k-й ячейкой (см. рис. 9.18, *б*). Теперь задержка определяется существенно меньшими поликремниевыми сегментами и пропорциональна $(k/2)^2$. Создание контактов с каждой k-й ячейкой помогает сохранить плотность реализации на прежнем уровне. Например, если создать соединение с обходной линией для каждой 16-й ячейки адресной шины из 1024 ячеек, задержка уменьшится приблизительно в 4000 раз.

Лучшие стратегии межсоединений

Поскольку длина провода является основным фактором, влияющим на задержку и энергопотребление межсоединения, любой подход, помогающий уменьшить длину провода, изначально очень важен. Ранее уже отмечалось, что добавление большего числа слоев межсоединений позволяет уменьшить среднюю длину проводников, поскольку уменьшает перегруженность каналов трассировки, и межсоединения в среднем становятся ближе к прямым маршрутам между источником и пунктом назначения. Пока что в современных инструментах трассировки используется “манхеттенский” подход, приносящий с собой существенное увеличение служебных издержек, которое часто недооценивают. При трассировке “в манхеттенском стиле” межсоединения вначале прокладываются вдоль одного из двух предпочтительных направлений, после чего прокладываются соединения в другом направлении (см. рис. 9.19, *а*). Очевидно, что трассировка вдоль диагональных направ-

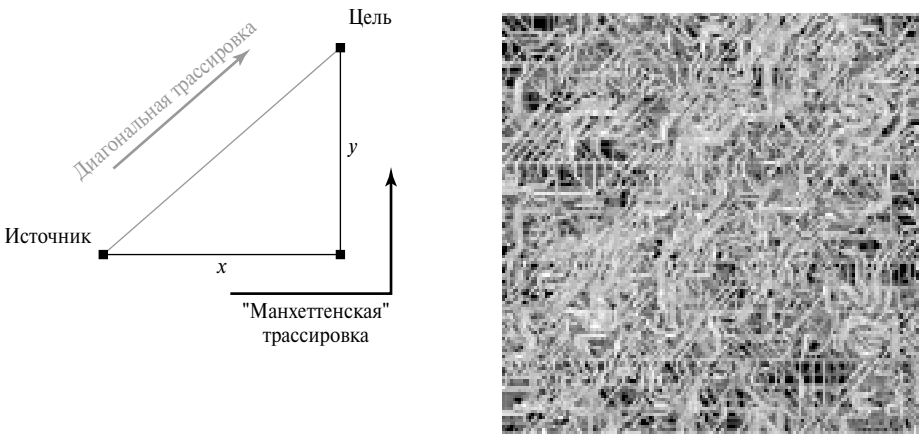


Рис. 9.19. Разница между "манхеттенской" и диагональной трассировкой. При "манхеттенской" трассировке предпочтительными считаются направления вдоль ортогональных осей, а при диагональной допускается использование линий под углом 45° (*а*); пример топологии, использующей линии под углом в 45° (*б*)

лений существенно уменьшит длину провода (в лучшем случае на 29%). По иронии судьбы, линии под углом в 45° были очень популярны на заре развития цифровых интегральных схем, однако они вышли из моды из-за своей сложности, влияния на используемые инструменты и вопросов, связанных с созданием фотошаблонов. Недавно было показано, что все эти проблемы можно решить и что линии под углом в 45° вполне реальны [Cadence-X]. Их влияние на длину проводки вполне ощутимо: уменьшение средней длины провода на 20%! В свою очередь, это повышает быстродействие, снижает рассеяние мощности и уменьшает используемую площадь кристалла. На простой топологии, показанной на рис. 9.19, *б*, демонстрируется структура проводки, плотность которой выше, чем у чисто "манхеттенской схемы".

Введение повторителей

В сфере разработки наиболее популярным подходом к уменьшению задержки распространения длинных проводов является введение в соединительную линию (рис. 9.20) промежуточных буферов (или *повторителей*). При уменьшении длины соединения в m раз ее задержка распространения уменьшается квадратично, и для длинных проводов этого вполне достаточно, чтобы оправдать дополнительную задержку, вводимую повторителями [Cong99, Adler00]. Предполагая, что повторители имеют фиксированную задержку t_{pbuf}, мы можем рассчитать оптимальное число требуемых повторителей m_{opt} с помощью схемы, подобной использованной при оптимизации цепочки передаточных вентилей. В результате получаем следующее соотношение:

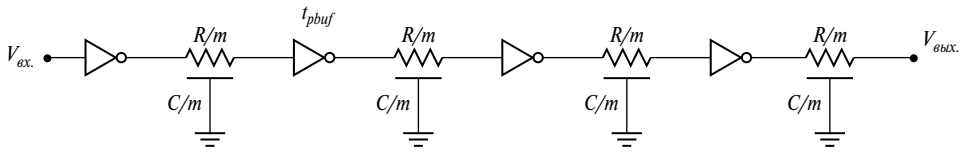


Рис. 9.20. Уменьшение RC-задержки межсоединения с помощью повторителей

$$m_{opt} = L\sqrt{\frac{0,38rc}{t_{pbuf}}} = \sqrt{\frac{t_{pwire(unbuffered)}}{t_{pbuf}}}. \tag{9.7}$$

Соответствующая минимальная задержка для провода равна:

$$t_{p,opt} = 2\sqrt{t_{pwire(unbuffered)}t_{pbuf}}. \tag{9.8}$$

Оптимальным является состояние, когда задержка отдельных сегментов провода равна задержке повторителя.

Пример 9.6. Уменьшение задержки с помощью повторителей

В примере 4.8 мы нашли, что задержка распространения провода Al-1 длиной 10 см и толщиной 1 мкм равна 31,4 нс. Из уравнения (9.7) следует, что для минимизации задержки необходимо разбить этот провод на 18 сегментов (значение t_{pbuf} предполагается фиксированным и равным 0,1 нс). В результате общая задержка составит 3,5 нс, так что улучшение получено довольно существенное. Подобным образом, задержки распространения поликремниевого провода и провода Al-5 одинаковой длины уменьшаются до 212 нс (с 112 мс) и 1,3 нс (с 4,2 нс); число каскадов после модификации составляет 1058 и 6 соответственно.

Предыдущий анализ является упрощенным и оптимистичным в том смысле, что t_{pbuf} представляет собой функцию емкости нагрузки. Для уменьшения задержки необходимо масштабирование повторителей. Более точное выражение для задержки соединительной цепочки получается, если смоделировать повторитель как RC-цепь и использовать формулу Элмора. Обозначив через R_d и C_d соответственно входное сопротивление и емкость повторителя минимального размера, получим следующее выражение:

$$\begin{aligned} t_p = \; & m\left(0,69\frac{R_d}{s}\left(s\gamma C_d + \frac{cL}{m} + sC_d\right) + \right. \\ & \left. +0,69\left(\frac{rL}{m}\right)(sC_d) + 0,38rc\left(\frac{L}{m}\right)^2\right). \end{aligned} \tag{9.9}$$

В данной формуле γ представляет отношение внутренних выходной и входной емкостей повторителя, как определяется в уравнении (5.29). Приравняв к нулю $\partial t_p/\partial m$ и $\partial tp/\partial s$, получим оптимальные значения m, s и $t_{p,min}$:

$$\begin{aligned} m_{opt} &= L\sqrt{\frac{0,38rc}{0,69R_dC_d(\gamma+1)}} = \sqrt{\frac{t_{pwire}(unbuffered)}{t_{p1}}}; \\ s_{opt} &= \sqrt{\frac{R_dc}{rC_d}}; \\ t_{p,min} &= (1,38+1,02\sqrt{1+\gamma})L\sqrt{R_dC_drc}. \end{aligned} \tag{9.10}$$

В данных формулах $t_{p1} = 0,69R_dC_d(\gamma+1) = t_{p0}(1+1/\gamma)$ представляет задержку инвертора для единичного разветвления по выходу ($f = 1$). Из уравнения (9.10) отчетливо видно, как введение повторителей линеаризует задержку провода. Кроме того, **для данной технологии и данного слоя трассировки существует оптимальная длина сегментов проводов между повторителям**. Данную *критическую длину* можно представить следующим выражением:

$$L_{crit} = \frac{L}{m_{opt}} = \sqrt{\frac{t_{p1}}{0,38rc}}. \tag{9.11}$$

Можно показать, что задержка сегмента критической длины всегда равна

$$t_{p,crit} = \frac{t_{p,min}}{m_{opt}} = 2\left(1+\sqrt{\frac{0,69}{0,38(\gamma+1)}}\right)t_{p1} \tag{9.12}$$

и **не зависит от слоя трассировки**. Для типичного значения $\gamma = 1$, $t_{p,crit} = 3,9t_{p1} = 7,8t_{p0}$. Введение дополнительных повторителей для уменьшения задержки имеет смысл только тогда, когда провод по крайней мере вдвое больше критической длины.

Пример 9.7. Минимизация задержки провода (продолжение)

В главе 5 (пример 5.5) мы определили, что задержка t_{p1} инвертора минимального размера 0,25-микронного КМОП-процесса равна 32,5 пс при (средних) значениях R_d и C_d, равных 7,8 кОм и 3 фФ соответственно, и $\gamma = 1$. Для провода Al-1 ($c = 110$ аФ/мкм; $r = 0,075$ Ом/мкм) это соответствует оптимальному масштабному коэффициенту $s_{opt} = 62$. После добавления 31 такого масштабированного инвертора минимальная задержка 10-сантиметрового провода из примера 9.6 даст более реалистичные 3,9 нс. При добавлении же 31 повторителя минимального размера задержка в действительности увеличится до 61 нс из-за плохой нагрузочной способности повторителя.

Критическая длина провода первого алюминиевого слоя составляет 3,2 мм, а для поликремниевого и пятого алюминиевого слоев — 54 мкм и 8,8 мм соответственно.

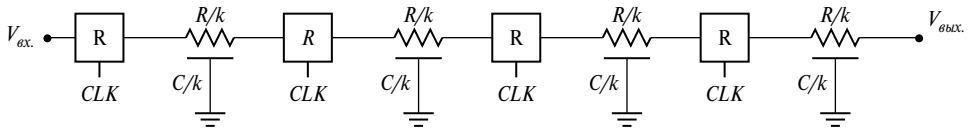


Рис. 9.21. "Конвейер" проводов повышает пропускную способность межсоединений

Таким образом, введение повторителей оказалось настолько эффективным (и необходимым) средством борьбы с большими задержками проводов, что современные средства автоматизации выполняют эту задачу автоматически.

Оптимизация архитектуры межсоединений

Даже при добавлении буферов задержку резистивного провода нельзя уменьшить ниже минимального уровня, следующего из уравнения (9.10). Следовательно, длинные провода часто характеризуются задержкой, превышающей такт, заданный для проекта. Например, 10-сантиметровый провод первого алюминиевого слоя (Al-1) из примера 9.7 характеризуется минимальной задержкой 3,9 нс (даже после оптимального добавления буферов и масштабирования), а 0,25-микронный КМОП-процесс, который мы преимущественно рассматриваем в данной книге, может поддерживать тактовую частоту, превышающую 1 ГГц (т.е. тактовый интервал составляет меньше 1 нс). Таким образом, задержка провода сама по себе становится фактором, ограничивающим быстродействие, потенциально достижимое интегральной схемой. Единственный путь избавиться от этого узкого места — это решить проблему на уровне архитектуры системы.

Одной из популярных технологий повышения быстродействия является *конвейерная прокладка проводов*. Концепция конвейерной обработки вводилась в главе 7 как средство улучшения пропускной способности логических модулей с длинными критическими путями. Как показано на рис. 9.21, подобный подход можно использовать для повышения пропускной способности провода. Проводник разбивается на k сегментов посредством регистров или защелок. Это хотя и не уменьшает задержку распространения сигнала по сегменту провода (прохождение сигнала через провод занимает k тактов), но помогает увеличить пропускную способность этого сегмента, поскольку в любой момент времени провод обрабатывает k сигналов одновременно. Задержку отдельных отрезков можно дополнительно оптимизировать, добавив повторители и снизив ее до уровня, меньшего одного тактового интервала.

В качестве примера мы привели всего одну из множества технологий, имеющихся в распоряжении архитектора кристалла и позволяющих решить проблему задержки. Самое важное, что вы должны вынести из приведенного обсуждения, — проводники нужно учитывать в самом начале процесса разработки, допустимый ранее принцип "об этом я подумаю позже" в современных условиях уже не действует.

9.4. Паразитная индуктивность*

Помимо паразитных сопротивления и емкости, межсоединения также обладают паразитной индуктивностью. Важным источником паразитной индуктивности являются монтажные проводы и корпус кристалла интегральной схемы. Даже для КМОП-разработок средней скорости ток, идущий через соединения ввода-вывода, может быстро меняться, приводя к скачкам напряжения, а также затухающим колебаниям и перегрузке, чего не бывает в RC-цепях. При высоких скоростях переключения свой вклад в общую картину могут давать эффекты линии передачи и распространения волн. Оба эффекта подробно рассмотрены в данном разделе, в частности, предлагаются дизайнерские решения, позволяющие устранить их.

9.4.1. Индуктивность и надежность — скачок напряжения на $L\frac{di}{dt}$

При каждом переключении переходной ток поступает из шин питания (или в шины питания) и заряжает (или разряжает) емкости контура, как показано на рис. 9.22. Соединения V_{DD} и V_{SS} связываются с внешними источниками питания посредством монтажных проводов и выводов корпуса, несущих последовательную индуктивность, игнорировать которую нельзя. Таким образом, изменение переходного тока создает разность напряжений между внешним и внутренним (V'_{DD}, GND') напряжениями источника питания. Данная ситуация особенно существенна на выходных контактных площадках, где из-за питания больших внешних емкостей возникают большие выбросы тока. Колебания внутренних напряжений питания влияют на логические уровни и приводят к уменьшению запаса помехоустойчивости.

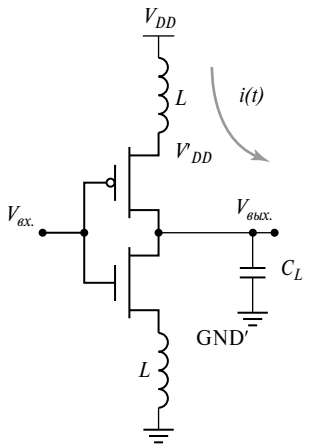


Рис. 9.22. Индуктивная связь между внешним и внутренним источниками напряжения

Пример 9.8. Шум, индуцированный индуктивной монтажной проводкой и выводами корпуса интегральной схемы

Предположим, что схема, приведенная на рис. 9.22, представляет собой последний каскад задающего генератора выходной контактной площадки, питающего емкость нагрузки 10 пФ с размахом напряжения 2,5 В. Инвертор масштабирован так, что время нарастания и спада 10–90% выходного сигнала (t_r, t_f) равно 1 нс. Поскольку соединения с источником питания и землей подключены к внешним источникам питания посредством соответствующих выводов, оба соединения обладают последовательной индуктивностью L. В традиционной структуре корпуса со сквозными отверстиями типичной является индуктивность порядка 2,5 нГн. Для упрощения анализа предположим вначале, что инвертор действует как источник тока с постоянной силой тока, идущей на зарядку (разрядку) емкости нагрузки. Для получения времен нарастания и спада, равного 1 нс, требуется средний ток 20 мА:

$$I_{avg} = (10 \text{ пФ} \times (0,9 - 0,1) \times 2,5 \text{ В})/1 \text{ нс} = 20 \text{ мА}.$$

Описанный сценарий реализуется, когда на вход буфера поступает резкая ступенчатая функция. На рис. 9.23, *слева* показана смоделированная эволюция во времени выходного напряжения, силы тока и напряжения индуктора для $t_f = 50$ пс. Резкое изменение силы тока вызывает на индукторе резкий всплески напряжения, величиной до 0,95 В. Фактически скачок напряжения был бы больше, если бы само по себе изменение не замедляло переходы и не уменьшало потребность в токе. Тем не менее источники питания, демонстрирующее такие отклонения, неприемлемы.

В действительности зарядный ток редко является постоянным. Гораздо лучшее приближение мы получим, предположив, что сила тока линейно нарастает до максимального значения, после чего падает до нуля (также линейно). Таким образом, распределение тока со временем моделируется треугольной функцией. Такая модель соответствует ситуации, когда сигнал на входе буфера нарастает и спадает медленно. В этой модели скачок напряжения на индукторе (оценка) выражается следующим образом:

$$\nu_L = L di_L/dt = (2,5 \text{ нГ} \times 40 \text{ мА})/((1,25/2) \times 1 \text{ нс}) = 133 \text{ мВ}.$$

Пиковый ток, равный (приблизительно) 40 мА, возникает из-за того, что суммарный перенесенный заряд (интеграл от I_L) фиксирован, следовательно, максимум треугольной функции распределения тока равен удвоенному значению прямоугольной функции. Знаменатель приведенного выше выражения задает время, требуемое для перехода от нулевого до максимального тока. Результаты соответствующего моделирования для $t_f = 800$ пс показаны в правой части рис. 9.23. Видно, что распределение тока действительно становится треугольным для медленно изменяющегося входного сигнала. Смоделирован-

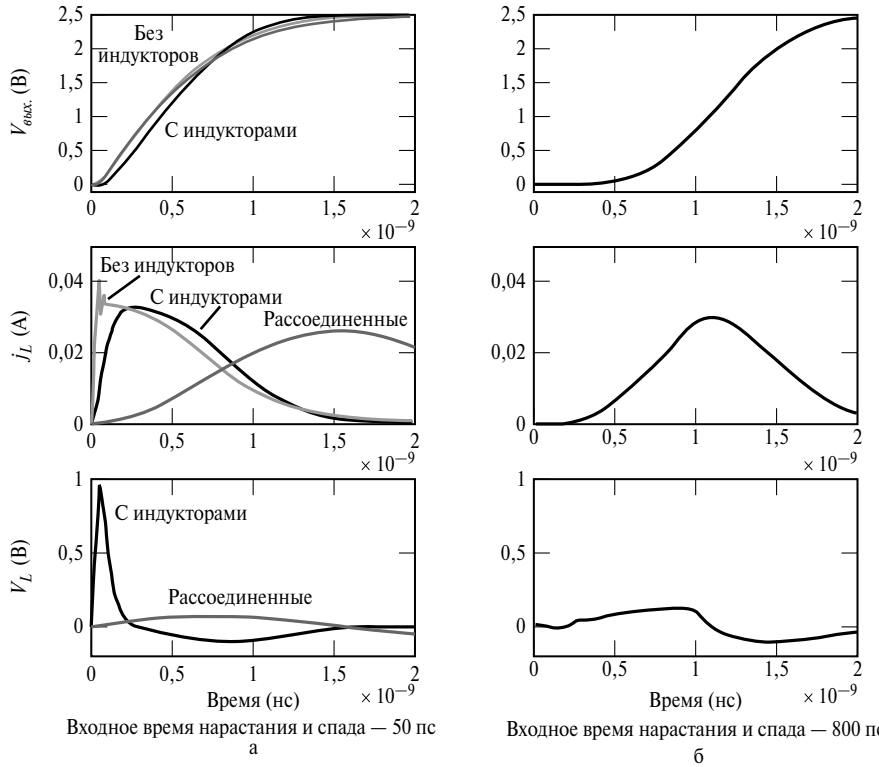


Рис. 9.23. Смоделированные сигналы для выходного задающего устройства, соединенного с контактными площадками для времени нарастания/спада входного сигнала 50 пс (*а*) и 800 пс (*б*) соответственно: *а* — сигналы, соответствующие случаю (1) отсутствия индукторов в схеме разводки питания (идеальный случай); *б* — наличие индукторов; *в* — с дополнительным развязывающим конденсатором емкостью 200 пФ

ное падение напряжения на индукторе составляет менее 100 мВ. Следовательно, можно сделать вывод, что использование на входе задающих устройств медленно изменяющихся сигналов существенно снижает эффекты Ldi/dt.

В реальной схеме один вывод источника питания обслуживает несколько логических элементов или выходных задающих устройств. Одновременное переключение этих устройств может вызвать даже еще худшие переходные токи и скачки напряжения. В результате значения внутренних напряжений питания существенно отклоняются от внешних значений. Например, одновременное переключение 16 выходных задающих устройств выходной шины вызовет скачок напряжения не менее 1,1 В, если соединения всех буферов с источником питания производится через один и тот же вывод корпуса.

Улучшения технологий корпусирования приводят к увеличению числа выводов на один корпус. В настоящее время существуют корпусы, число выводов которых доходит до 1000. Одновременное переключение значительного

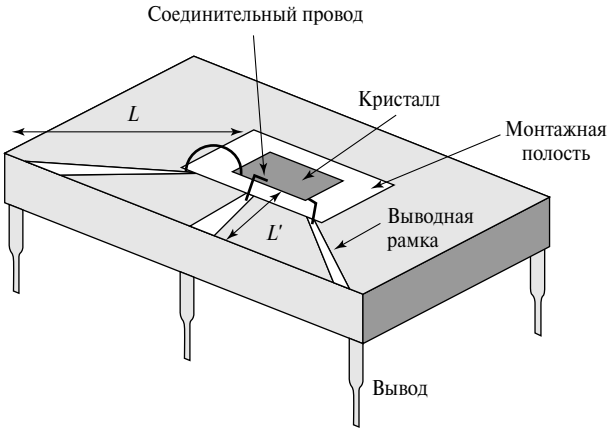


Рис. 9.24. Индуктивность соединения монтажного провода и вывода зависит от положения вывода

числа этих выводов приводит к большим всплескам в шинах питания, что, разумеется, нарушает работу внутренних цепей и внешних компонентов, соединенных с теми же шинами питания.

Технологии проектирования

Разработчик может решить проблему $L(di/dt)$ с помощью нескольких подходов.

1. **Отделить выводы питания от контактных площадок ввода-вывода и ядра кристалла.** Поскольку задающие устройства ввода-вывода требуют максимальных токов переключения, они также вызывают наибольшие изменения тока. Следовательно, ядро кристалла (в котором происходит большинство действий) разумно изолировать от задающих устройств, обеспечив отдельные выводы питания и земли.
2. **Создать несколько выводов питания и земли.** Чтобы уменьшить значение di/dt на один вывод питания, мы можем ограничить число задающих устройств ввода-вывода, соединенных с одним выводом питания. Обычно используются 5–10 задающих устройств на вывод питания. Помните, что это число сильно зависит от характеристик переключения задающих устройств (например, числа одновременно переключаемых элементов и времени нарастания и спада).
3. **Тщательный выбор положения выводов питания и земли на корпусе.** Индуктивность соединительных проводов, связанных с выводами, расположенными в углах корпуса, существенно выше (рис. 9.24).
4. **Увеличить время нарастания/спада** внекристальных сигналов до максимально допустимого значения и распределить эти сигналы по всему кристаллу (особенно если речь идет о шинах данных). Из предыдущего

примера видно, что выходные задающие устройства с завышенными требованиями к времени нарастания/спада не только являются дорогими с точки зрения площади, но и влияют на функционирование и надежность схемы. В разделе 9.2.2 мы отмечали, что лучшим задающим устройством с точки зрения площади является устройство с заданной (а не с минимальной) задержкой. Если учитывать шум, то лучшим будет устройство, дающее заданную задержку с максимальным допустимым временем нарастания и спада на выходе.

Задача 9.1. Проектирование выходного задающего устройства с уменьшенным временем нарастания/спада

Дано каскадное выходное задающее устройство, спроектированное под данную задержку, которая порождает чрезмерный шум в цепях питания. Определите наилучший подход к решению проблемы шума: (*а*) масштабирование всех каскадов буфера; (*б*) масштабирование только последнего каскада; (*в*) масштабирование всех каскадов, кроме последнего.

---

5. **Распределите расходующие ток переходы** так, чтобы они не происходили одновременно. Например, переключение набора выходных задающих устройств можно разнести во времени, немного сместив их управляющие входы.
6. **Используйте современные технологии корпусирования** — поверхностный монтаж или гибридные технологии, существенно снижающие емкость и индуктивность из расчета на один вывод. Например, из табл. 2.2 видно, что контактная индуктивность кристалла, смонтированного в стиле перевернутого кристалла на подложке с использованием технологии столбиков припоя, уменьшается до 0,1 нГн, что в 50–100 раз меньше, чем для стандартных корпусов.
7. **Добавить на плату развязывающую емкость.** Данные емкости, которые необходимо добавлять для каждого вывода питания, действуют как локальные источники питания и стабилизируют напряжение источника питания, получаемое кристаллом. Они отделяют индуктивность монтажного провода от индуктивности межсоединений платы (рис. 9.25). Развязывающий конденсатор, дополненный индуктивностью, действует как фильтр нижних частот, отсекающих высокочастотные компоненты напряжения переходных процессов в линиях питания.

Пример 9.9. Влияние развязывающих емкостей

Между выводами питания буферной схемы, рассмотренной в примере 9.8, добавлен развязывающий конденсатор емкостью 200 пФ. Его влияние смоделировано на рис. 9.23. Если время нарастания входного сигнала равно 50 пс, пики в шинах питания уменьшаются с 0,95 В (без развязывающей емкости)

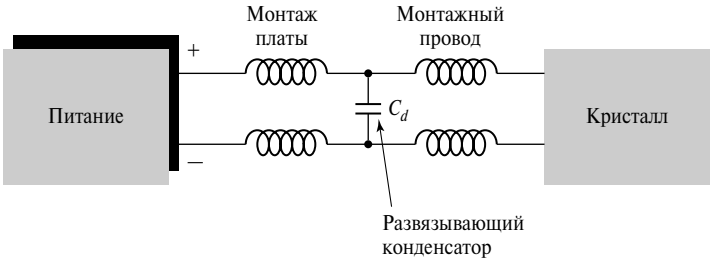


Рис. 9.25. Развязывающие конденсаторы отделяют индуктивность платы от индуктивности монтажной проводки и выводов

приблизительно до 70 мВ. Кроме того, на рисунке показано, как разряжающий конденсатор изначально заряжает выходную емкость. Позже, во время перехода, ток постепенно поглощается из сети питания посредством индукторов в монтажной проводке.

8. **Добавить развязывающую емкость на кристалл.** В быстродействующих схемах с большой скоростью переключения и резкими переходами сигналов развязывающие емкости часто интегрируют в кристалл, поскольку это обеспечивает более чистое напряжение источника питания. Внутрикристальные развязывающие конденсаторы уменьшают требования к пиковому току до средних значений. Чтобы ограничить пульсации напряжения значениями до 0,25 В, каждый модуль, содержащий 50 000 логических элементов (0,25-микронный КМОП-процесс), должен иметь емкость порядка 12,5 пФ [Dally98]. Данная емкость обычно реализуется с использованием тонкого слоя подзатворного оксида. По сути, тонкий оксидный конденсатор представляет собой МОП-транзистор с совмещенными стоком и истоком. С аналогичной целью могут использоваться и диффузионные конденсаторы.

Пример 9.10. Внутрикристальные развязывающие емкости в семействе процессоров Alpha (Compaq) [Herrick00]

Уже упоминавшееся в этой главе семейство процессоров Alpha представляет собой наибольшее достижение в сфере быстродействующих микропроцессоров за последнее десятилетие. Увеличение тактовой частоты до предела возможностей поставило серьезные требования к сетям разводки питания. Соответственно процессоры Alpha представляют наиболее перспективные устройства с точки зрения внутрикристальных развязывающих емкостей. В табл. 9.4 указаны некоторые характеристики последовательных процессоров этого семейства.

Для данной емкости управляющего электрода 4 фФ/мкм^2 ($t_{ox} = 9,0$ нм) развязывающая емкость 320 нФ процессора EV6 требует 80 мм^2 площади

Таблица 9.4. Внутрикристальные развязывающие емкости семейства процессоров Alpha

Процессор	Технология	Тактовая частота	Суммарная переключаемая емкость	Внутрикристальная развязывающая емкость
EV4	0,75-микронная КМОП	200 МГц	12,5 нФ	128 нФ
EV5	0,5-микронная КМОП	350 МГц	13,9 нФ	160 нФ
EV6	0,35-микронная КМОП	575 МГц	34 нФ	320 нФ

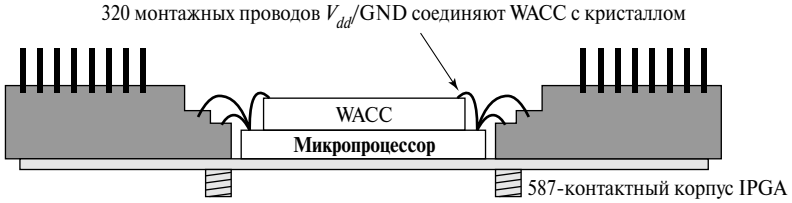


Рис. 9.26. Схема бескорпусного конденсатора, подключенного с помощью проводных соединений (Wire-Bond-Attached Chip Capacitor — WACC), предлагающего дополнительную развязывающую емкость рядом с кристаллом. В качестве примера использована реализация WACC в микропроцессоре EV6 (Compaq) [Gieseke97], соединенном с внешним миром посредством 320 сигнальных выводов и 198 выводов питания/земли

кристалла, т.е. 20% всей площади кристалла. Чтобы минимизировать это значение, емкость была размещена под основными шинами. Несмотря на то что под развязывающую емкость выделяется существенная часть площади, разработчикам емкости не хватило и они вынуждены были прибегнуть к некоторых инновационным технологиям, позволяющим получить больше развязывающей емкости поблизости кристалла. Проблема была решена с помощью проводного соединения с кристаллом конденсатора 2 мФ, 2 $см^2$ и соединения электрической сети с емкостью посредством 160 проводных пар V_{dd}/GND. Схема принципа *WACC* (Wire-Bond-Attached Chip Capacitor — бескорпусный конденсатор, подключенный с помощью проводных соединений) показана на рис. 9.26. Данный пример хорошо иллюстрирует задачи, с которыми сталкиваются разработчики электрической сети для быстродействующей структуры. Существуют и другие решения, не используемые в настоящее время: технологии корпусирования с монтажом по методу перевернутого кристалла, которые позволяют привязывать сеть питания к развязывающим конденсаторам, интегрированным в корпус. Данный подход, например, был использован в процессорах Intel Pentium 4™ [Pentium02].

Наконец, помните, что взаимоиндукция соседних проводов также приводит к перекрестным помехам. Данный эффект не сильно существен в КМОП, но при высоких скоростях переключения он, определенно, будет представлять серьезную проблему [Johnson93].

Емкость (проводов и развязывающих конденсаторов) и индуктивность в сети разводки питания образует резонансную цепь, которая может осциллировать. Существует несколько возможных резонансных состояний, важнейшее из которых определяется индуктивностью корпуса L и развязывающей емкостью C_d (резонансная частота равна $f = 1/(2\pi\sqrt{L \cdot C_d})$). Если сеть питания резонирует с тактовой частотой, в сети питания возможны очень опасные флуктуации. В прошлом резонансная частота сети питания была намного выше тактовой частоты. В последние годы резонансная частота постоянно уменьшалась (поскольку емкость C_d с поколениями устройств увеличивалась, индуктивность L оставалась приблизительно постоянной), а тактовая частота процессоров при этом только повышалась. В настоящее время обычной является ситуация, когда резонансная частота сети питания существенно ниже тактовой частоты. Следовательно, осцилляции в сети питания являются довольно серьезной проблемой.

Достаточное подавление данных осцилляций абсолютно необходимо. Фактически это единственное место во всей области цифрового проектирования, где сопротивление оказывается полезным. Развязывающие емкости часто включаются последовательно с контролируемой дозой сопротивления. Тем не менее это приводит к увеличению омического скачка напряжения и скачка, вызванного резонансом. Помимо этого, правильный выбор и распределение развязывающих емкостей иногда помогает сместить резонансную частоту в области, которые никогда не проявятся на практике.

9.4.2. Индуктивность и быстродействие — эффекты линии передачи

Когда соединительные проводники становятся достаточно длинными, или когда схемы становятся достаточно быстрыми, фактором, определяющим поведение задержки, становится индуктивность, и в таких случае нужно рассматривать эффекты линии электропередачи. Это соответствует случаю, когда время нарастания и спада сигнала сравнимо со временем прохода сигнала через линию (определяется скоростью света). До недавнего времени такое условие выполнялось только для наиболее быстрых цифровых разработок, реализованных на уровне печатной платы, или таких экзотических разработок, как структуры на основе GaAs и SiGe. Поскольку для современных технологий характерно увеличение длины линии и скорости переключения, данная ситуация постепенно становится обычной и для наиболее быстрых КМОП-схем, поэтому разработчику следует учитывать в своей работе эффекты линии передачи.

В данном разделе мы обсудим некоторые технологии, позволяющие минимизировать или смягчить влияние линии передачи. Прежде всего необходимо выбрать правильную оконечную нагрузку. Впрочем, оконечная нагрузка эффективна только при хорошем пути возврата тока. Кроме того, необходимо экранирование проводов, подверженных эффектам линии электропередачи[3].

[3] Ряд впечатляющих анимированных представлений эффектов линии передачи и влияния оконечной нагрузки можно найти на Web-сайте П. Рестла (P. Restle) из IBM [Restle01].

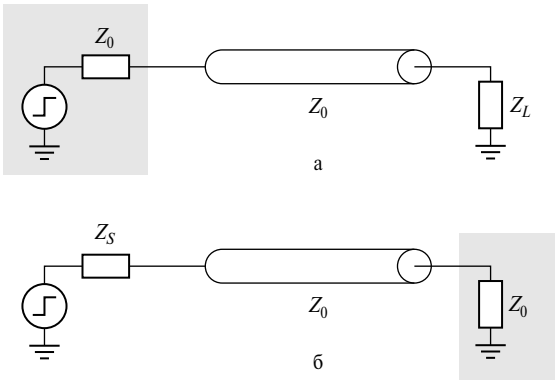


Рис. 9.27. Сценарии согласования оконечной нагрузки для проводников, ведущих себя как линии передачи: *а* — последовательная оконечная нагрузка на входе; *б* — параллельная оконечная нагрузка на выходе

Оконечная нагрузка

В нашем обсуждении линий передач (см. главу 4) отмечалось, что наиболее эффективным путем минимизации задержки является правильный выбор оконечной нагрузки. Согласовывая импеданс нагрузки с характеристическим импедансом линии передачи, вы получите наиболее быстрый отклик. Таким образом, можно сформулировать следующее правило разработки.

> **Чтобы избежать отрицательных эффектов линии передачи, подобных "звону" или большим задержкам распространения, линию необходимо снабдить оконечной нагрузкой либо в начале (последовательная оконечная нагрузка), либо в конце (параллельная оконечная нагрузка), причем сопротивление нагрузки должно согласовываться с характеристическим импедансом линии Z_0.**

Два названных сценария (последовательный и параллельный) представлены на рис. 9.27. Последовательная оконечная нагрузка требует, чтобы импеданс источника сигнала согласовывался с импедансом соединительного провода. Данный подход приемлем для многих КМОП-структур, в которых нагрузка конца линии является чисто емкостной. Импеданс задающего инвертора можно согласовать с импедансом линии путем аккуратного масштабирования транзисторов. Питание линии с сопротивлением 50 Ом, например, требует n-канального полевого транзистора шириной 53 мкм и p-канального полевого транзистора шириной 135 мкм (как всегда, мы предполагаем использование 0,25-микронной КМОП-технологи), дающих номинальный выходной импеданс 50 Ом.

Импеданс задающего устройства необходимо аккуратно согласовать с импедансом линии (обычно требуется соответствие в пределах 10% или даже меньше, если желательно избежать чрезмерных отражений бегущих волн). К сожалению, сопротивление канального транзистора в открытом состоянии

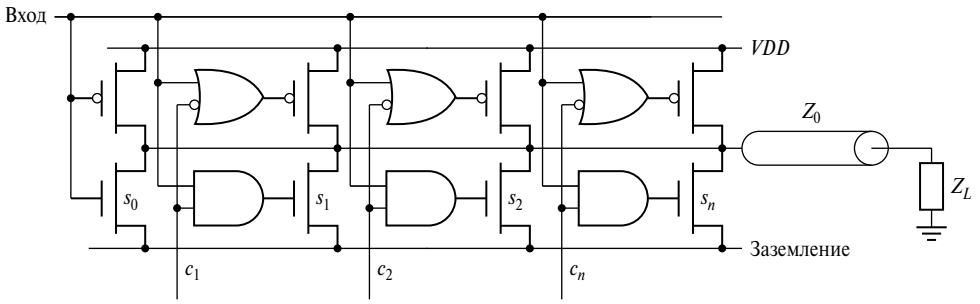


Рис. 9.28. Настраиваемое сегментированное задающее устройство предлагает согласованную последовательную нагрузку для линии передачи

может менять свое значение на 100% в зависимости от процесса, напряжения и температуры. Это можно компенсировать, реализовав электрическую регулировку сопротивлением транзисторов, как показано на рис. 9.28. Каждый транзистор задающего устройства заменяется сегментированным задающим устройством, причем сегменты включаются и выключаются с помощью шин управления от c_1 до c_n (целью является максимальное согласование импеданса задающего устройства с импедансом линии). Каждый сегмент имеет свое сопротивление, определяемое форм-фактором s_i (обычно сопротивления идут по степеням 2). Параллельно с настраиваемыми элементами включается фиксированный элемент (s_0), что ограничивает диапазон настройки и позволяет получать более точную настройку с меньшим числом битов. Шины управления обычно управляются цепью обратной связи, которая сравнивает внутрикристальный эталонный транзистор задающего устройства с фиксированным внешним эталонным транзистором [Dally98].

Подобные соображения справедливы и тогда, когда оконечная нагрузка добавляется в конце линии (в таком случае говорят о *параллельной оконечной нагрузке*). Помните, что описанный подход приводит к появлению тока холостого хода и что постоянное протекание этого тока может привести к неприемлемо большому рассеянию мощности. Параллельная оконечная нагрузка очень популярна в высокоскоростных межкристальных соединениях на уровне печатных плат, где оконечная нагрузка часто реализуется путем введения заземленного резистора возле входного контакта (корпуса) на конце провода. Данная внекристальная оконечная нагрузке не учитывает паразитные нагрузки корпуса и внутренние цепи и может вводить неприемлемо большие отражения в сигнальной шине. Более эффективным подходом является включение нагрузочного резистора внутри корпуса после паразитных нагрузок корпуса.

КМОП-процесс производства не дает нам средства, позволяющего получать точные, не чувствительные к температуре транзисторы. Соединительные материалы или полевые транзисторы могут использоваться как резисторы, однако для преодоления их зависимости от температуры, напряжения источника питания и процесса необходима статическая или динамическая настройка. Для достижения этой цели можно применять цифровую

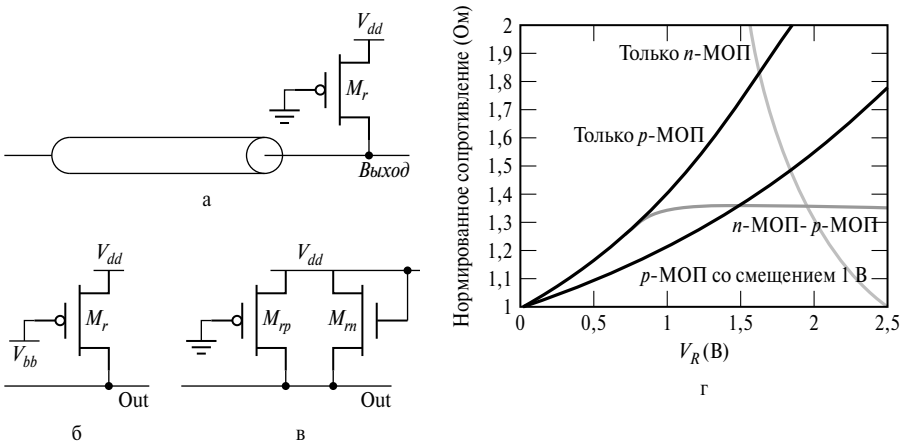


Рис. 9.29. Параллельное окончание линии передачи с использованием транзисторов в качестве резисторов: *а* — заземленный p-МОП-транзистор; *б* — p-МОП-транзистор с отрицательным напряжением смещения; *в* — комбинация p-МОП-n-МОП; *г* — моделирование

настройку, подобную предложенной ранее в этом разделе. Другой проблемой использования канальных полевых униполярных МОП-транзисторов в качестве резисторов является обеспечение линейной характеристики устройства в требуемой рабочей области. Поскольку p-МОП-транзисторы обычно имеют бо́льшие области линейного режима, чем их n-МОП-аналоги, именно их предпочитают использовать для реализации оконечных сопротивлений. Предположим, что в качестве согласованной нагрузки сопротивлением 50 Ом используется p-МОП-транзистор, показанный на рис. 9.29, *а*. Согласно результатам моделирования (см. рис. 9.29, *г*), видим, что сопротивление практически постоянно для небольших значений V_R, однако быстро увеличивается после насыщения транзистора. Линейную область можно расширить, увеличив напряжение смещения p-МОП-транзистора (см. рис. 9.29, *б*). Это потребует дополнительного напряжения питания, что в большинстве случаев непрактично. Лучшим решением является добавление включенного параллельно p-МОП-устройству через диод n-МОП-транзистора (см. рис. 9.29, *в*). Комбинация двух устройств дает почти постоянное сопротивление для всего диапазона изменения напряжений (как было показано в главе 6 при обсуждении передаточных элементов). Существуют и более хитрые схемы, которые в данной книге рассматриваться не будут. Если вас интересует этот вопрос — обращайтесь к [Dally98].

Экранирование

По сути, линия передачи представляет собой двухпортовую (четырехполюсную) сеть. Хотя мы преимущественно обращаем внимание на тракт сигнала, *цепь обратной связи* игнорировать не стоит. Согласно законам Кирхгофа, при входе тока i в проводник сигнала в обратной цепи должен присутствовать полный ток $-i$. Характеристики линии передачи (например, ее харак-

теристический импеданс) сильно зависят от обратного тока. Следовательно, если мы хотим контролировать поведение проводника, ведущего себя как линия передачи, нам нужно тщательно планировать и управлять обратным током. Удачным примером линии передачи с хорошим поведением является коаксиальный кабель, где сигнальный провод окружен цилиндрическим заземленным экраном. Чтобы получить подобные эффекты на плате или кристалле, разработчики часто окружают сигнальный провод заземленными слоями и экранирующими проводниками. Хотя такое решение дорого с точки зрения используемого места, из-за добавления экранирования поведение и задержка межсоединения становятся немного более предсказуемыми. Тем не менее даже в этом случае разработчику быстродействующих схем потребуются мощные средства моделирования.

Пример 9.11. Проектирование выходного задающего устройства (*продолжение*)

Чтобы завершить данный раздел, мы смоделируем выходное задающее устройство, содержащее большинство паразитных эффектов, введенных в этой главе. Схема устройства показана на рис. 9.30, *а*. В расчеты включены индуктивность выводов питания, земли и сигналов (приблизительно 2,5 нГн), а также поведение линии передачи вне платы (предполагается, что длина провода равна 15 см — типичному значению для печатных плат). Изначально задающее устройство проектировалось из расчета, что время нарастания и спада равно 0,33 нс для суммарной емкости нагрузки 10 пФ. Для получения требуемой средней силы тока в 60 мА ширины n-МОП и p-МОП-транзисторов последнего каскада задающего устройства должны составлять 120 и 275 мкм. На первом графике (см. рис. 9.30, *б*) показаны смоделированные сигналы для данной конфигурации. Видно, что имеют место значительные затухающие колебания. Следовательно, сопротивление задающего устройства компенсируется недостаточно (отсюда большие выбросы и большое время установки).

Для устранения указанных эффектов схема была перепроектирована. Прежде всего, были уменьшены размеры транзисторов, чтобы их импедансы соответствовали характеристическому импедансу линии передачи (до 65 и 155 мкм для n-МОП- и p-МОП-транзисторов соответственно). К источнику питания устройства была добавлена развязывающая емкость 200 пФ. Наконец, для получения более реалистичного представления в модель были добавлены диоды защиты входа выходного разветвления. Из смоделированного графика видно, что данные модификации оказались достаточно эффективными, в результате чего схема стала и быстрее и удобнее.

9.5. Улучшенные техники межсоединений

До этого момента мы обсудили несколько технологий, позволяющих в какой-то степени решить проблемы паразитных емкостей, индуктивностей и со-

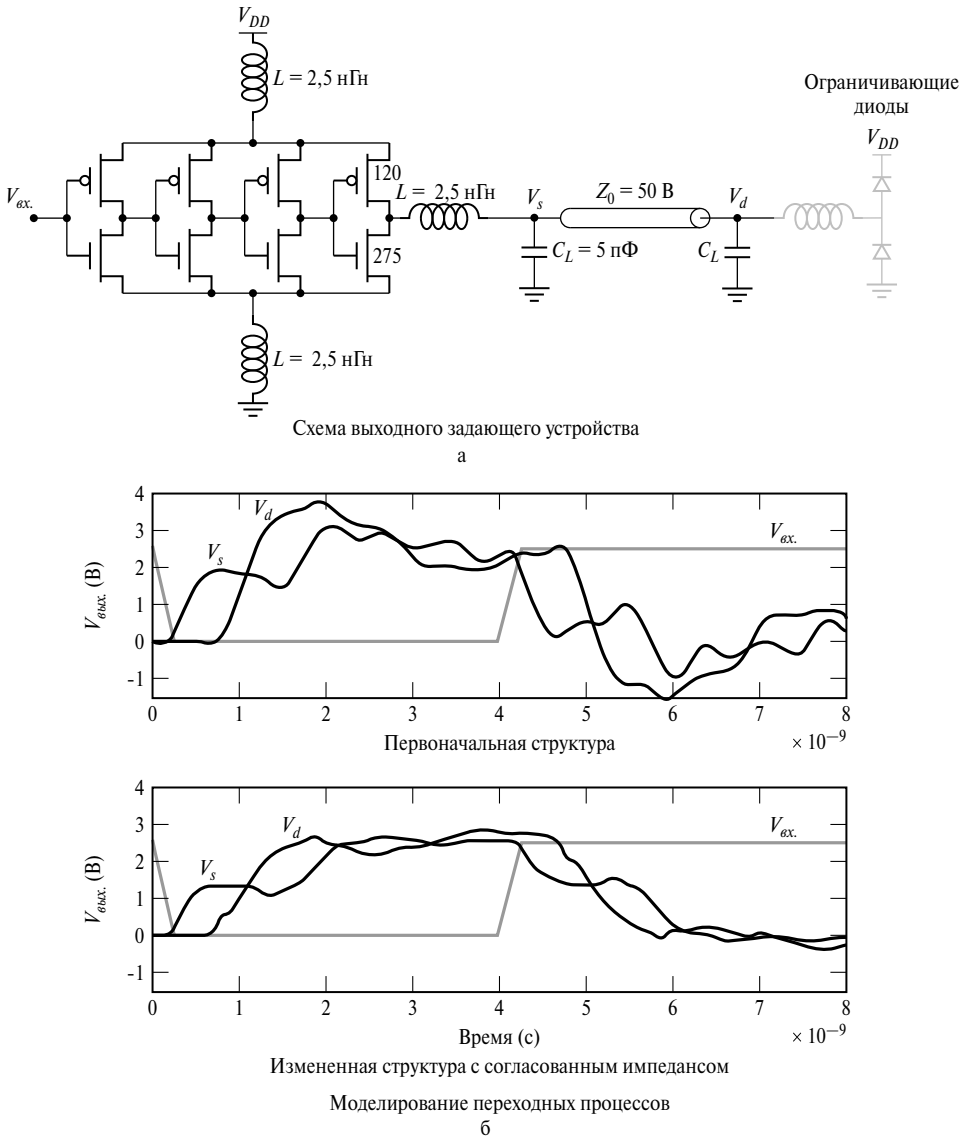


Рис. 9.30. Моделирование выходного задающего устройства для различных оконечных нагрузок

противлений, сопутствующих соединительным проводам. Чтобы представить какие-нибудь перспективы на будущее, в данном разделе мы обсудим более сложные схемы, появившиеся в последние годы. Точнее, мы рассмотрим, как уменьшенный размах сигнала помогает уменьшить задержку и рассеяние мощности при питании проводников с большими емкостями, кроме того, мы опишем несколько технологий реализации уменьшенного размаха.

9.5.1. Цепи с уменьшенным размахом

Увеличение размера задающего транзистора (а следовательно, увеличение средней силы тока I_{av} при переключениях) — это только один из способов борьбы с задержкой, вызванной большой емкостью нагрузки. Другой подход — уменьшить размах сигнала на выходе задающего устройства и на емкости нагрузки. Интуитивно понятно, что уменьшение заряда, который требуется перенести, может оказаться выгодным для быстродействия логического элемента. Чтобы это представить, вернемся к уравнению (5.16), описывающему задержку распространения:

$$t_p = \int_{\nu_1}^{\nu_2} \frac{C_L(\nu)}{i(\nu)} d\nu. \tag{9.13}$$

Предполагая, что емкость нагрузки постоянна, выражение для задержки можно упростить:

$$t_p = \frac{C_L V_{swing}}{I_{a\nu}}. \tag{9.14}$$

где $V_{swing} = \nu_2 - \nu_1$ — размах сигнала на выходе; $I_{a\nu}$ — средний ток зарядки/разрядки. Из приведенных выражений видно, что задержка уменьшается линейно с размахом сигнала при условии, что уменьшение размаха не влияет на ток зарядки (разрядки). Простого снижения общего напряжения питания недостаточно: это хотя и уменьшает размах, но также пропорционально уменьшает и ток. Сказанное иллюстрировалось ранее на рис. 3.28, где показано, что эквивалентное сопротивление транзистора зарядки (разрядки) остается приблизительно постоянным для большого диапазона напряжения источника питания. Меньший размах сигнала не только предлагает потенциально высшее быстродействие, но и снижает динамическое потребление мощности, что может быть существенным выигрышем при большой емкости нагрузки.

С другой стороны, меньший размах сигнала дает меньшую помехоустойчивость, следовательно, отрицательно влияет на достоверность и надежность сигнала. Более того, КМОП-элементы не особенно эффективны в деле детектирования и реагирования на небольшие изменения сигнала из-за сравнительно малой крутизны характеристики прямой передачи МОП-устройства. Для правильной работы и достижения высокого быстродействия в схемы с уменьшенным размахом необходимо вводить усилители, задачей которых является восстановление сигнала до полного размаха за минимальное время и при минимальном потреблении энергии. Служебные издержки, вызванные добавочными усилителями, оправданы только для узлов цепи с большим разветвлением по входу, где цепь может совместно использоваться несколькими входными элементами. В качестве типичных примеров таких узлов можно привести шины данных или адресные шины микропроцессора или шины данных матрицы памяти. В первом случае

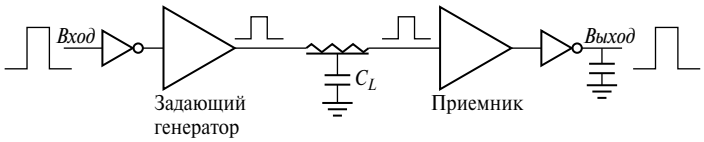


Рис. 9.31. Соединительная цепь с уменьшенным размахом. Задающий генератор уменьшает нормальный размах напряжения, а приемник детектирует сигнал и восстанавливает его до нормального размаха

усилитель часто называется *приемником* (receiver); во втором — *усилителем считывания* (sense amplifier)[4].

На рис. 9.31 показана типичная схема сети с уменьшенным размахом, которая состоит из задающего устройства, соединительного провода с большим сопротивлением и емкостью, а также принимающей схемы. Существует множество различных вариантов задающих устройств и приемников [Zhang00], но в данной книге рассматривается только несколько из них. Вообще, схемы с уменьшенным размахом делят на две основные категории — *статические* и *динамические* (или с предварительной зарядкой). Кроме того, цепи различают по принципу передачи сигналов. Большинство приемников используют *односторонний* (*несимметричный*) подход, когда приемник детектирует абсолютное изменение напряжения в одном проводе. В других схемах применяется *дифференциальный* (или *симметричный*) принцип, когда передаются сигнал и его дополнение, а приемник детектирует относительное изменение напряжения между двумя проводами. Несмотря на то что второй подход требует значительно больше места для проводки, он надежнее при наличии шума.

Статические сети с уменьшенным размахом

Задача разработки сети с уменьшенным размахом существенно упрощается при наличии второй шины питания с более низким напряжением V_{DDL}. На рис. 9.32 показана простая и устойчивая односторонняя задающая схема, реализованная с использованием второго источника напряжения. Проблемы возникают только при проектировании приемника. Просто использовать инвертор нельзя: малый размах на входе n-МОП-транзистора дает малый ток утечки, а следовательно, медленное изменение уровня сигнала от высокого к низкому на выходе; более того, малого значения V_{DDL} недостаточно до выключения p-МОП-транзистора, что еще сильнее ухудшает быстродействие и приводит к появлению статического рассеяния мощности.

На рис. 9.32 показана более удачная схема приемника, построенная на основе DCVSL-элемента, рассмотренного в главе 6. Для генерации локального дополнения входного сигнала в приемнике применяется низковольтный инвертор. Теперь приемник действует как дифференциальный усилитель. Нагрузочные транзисторы с перекрестными обратными связями гарантируют,

[4] Подробнее об усилителях считывания рассказывается в главе 12, посвященной полупроводниковым запоминающим устройствам.

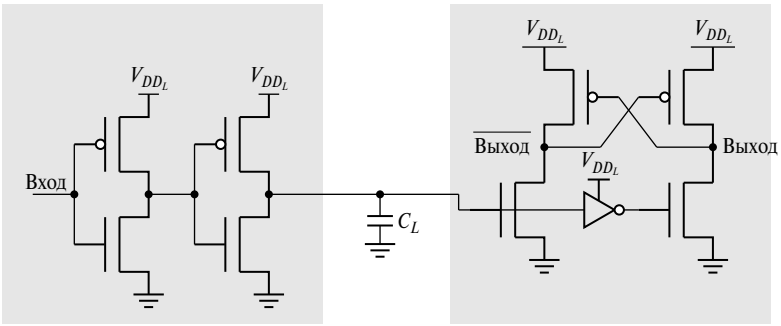


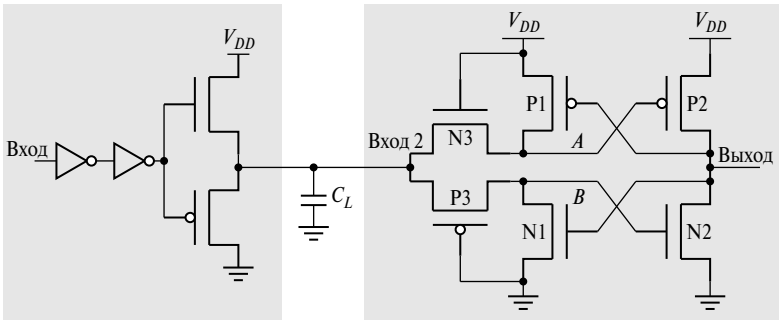
Рис. 9.32. Задающий генератор и приемник односторонней статической цепи с уменьшенным размахом напряжения

что выход будет восстановлен до V_{DD} и что в установившемся состоянии статическая мощность потребляться не будет. Положительная обратная связь помогает ускорить переходы. Недостаток приведенной схемы заключается в том, что для малого размаха она становится неприемлемо медленной.

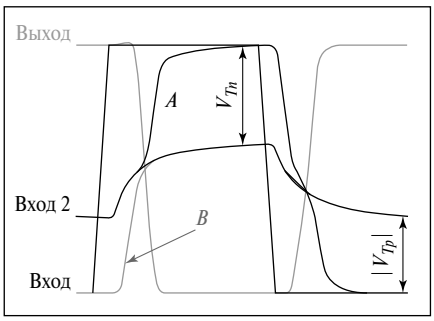
На рис. 9.33 показана схема, не использующая вторую шину питания. Поменяв местами n-МОП- и p-МОП-транзисторы последнего каскада задающего устройства, мы ограничили размах сигнала в межсоединении с $|V_{Tp}|$ до $V_{DD} - V_{Tn}$, или приблизительно на два пороговых значения меньше V_{DD}[5]. Симметричный приемник/преобразователь уровней состоит из двух пар транзисторов с перекрестными обратными связями (P1–P2 и N1–N2), а также из двух транзисторов (N3–P3), изолирующих межсоединение с уменьшенным размахом от выходного сигнала с полным размахом.

Чтобы понять работу приемника, предположим, что на узле $In2$ происходит переход от низкого уровня сигнала к высокому, или от $|V_{Tp}|$ к $V_{DD} - V_{Tn}$. Изначально узлы A и B имеют напряжение $|V_{Tp}|$ и GND соответственно. Во время перехода узлы N3 и P3 проводят, A и B заряжаются до $V_{DD} - V_{Tn}$, как показано на рис. 9.33, *б*. Следовательно, N2 включается и напряжение на узле Out начинает падать. Транзистор обратной связи P1 еще больше заряжает A до V_{DD}, полностью выключая P2. Узлы $In2$ и B сохраняют напряжение $V_{DD} - V_{Tn}$. Пути тока холостого хода от V_{DD} к GND через N3 не существует, хотя напряжение затвора-истока транзистора N3 почти равно V_{Tn}. Поскольку схема является симметричной, те же рассуждения справедливы и для изменения уровня сигнала от высокого к низкому. Внимательные читатели отметят, что транзисторы P1 и N1 действуют как восстановители уровней (как отмечалось ранее для логики на проходных транзисторах) Следовательно, они могут быть очень слабыми, а это минимизирует их состязание с задающим устройством. Задержка чувствительности приемника не превышает двух задержек инвертора.

[5] Недостаток данного подхода состоит в том, что размах напряжения в соединительной линии зависит от параметров технологии, например, порогового напряжения и влияния подложки, а следовательно, будет отличаться для разных кристаллов.



а



б

Рис. 9.33. Симметричный истоковый задающий генератор с преобразователем уровней: *а* — схема; *б* — смоделированные сигналы

Задача 9.2. Потребление энергии приемника с уменьшенным размахом

Предполагая, что основной вклад в емкость дает емкость нагрузки проводника, выведите уменьшение энергии на один переход, которое предлагает схема, изображенная на рис. 9.33, по сравнению со схемой с полным размахом.

Все демонстрировавшиеся выше системы межсоединений являются односторонними (несимметричными). На рис. 9.34 показана дифференциальная (симметричная) схема. Задающий генератор с помощью второй шины питания генерирует два комплементарных сигнала. Приемник представляет собой просто регистр, основанный на усилителе считывания (см. раздел 7.4.2). Дифференциальный подход предлагает возможность существенного подавления помех, подобных шуму в цепях питания и перекрестным помехам. Таким образом, размах сигнала можно уменьшить до очень низких уровней — на практике уже реализованы схемы с размахом 200 мВ. Задающий генератор использует n-МОП-транзисторы в нагрузочных и разгрузочных сетях. Основным недостатком дифференциального подхода является удвоение числа

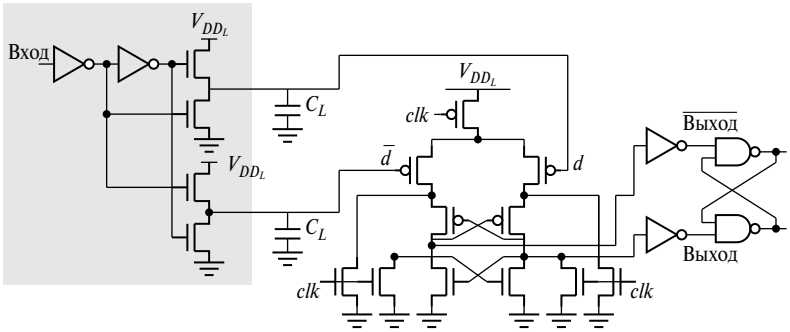


Рис. 9.34. Дифференциальная соединительная система с уменьшенным размахом, использующая в задающем генераторе дополнительную шину питания. Приемник представляет собой тактируемый дифференциальный триггер [Burd00]

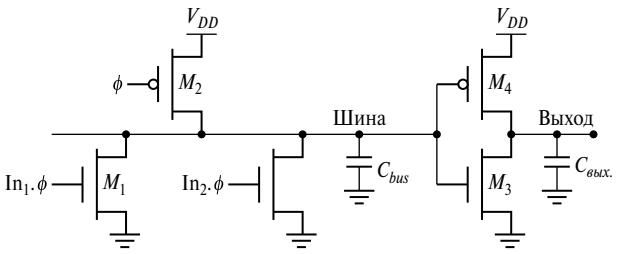
проводов, что во многих случаях представляет серьезную проблему. Кроме того, к дополнительным служебным издержкам относится дополнительный тактовый сигнал.

Динамические сети с уменьшенным размахом

Другой подход к ускорению реакции цепей с большим разветвлением по входу (например, шин) заключается в использовании предварительной зарядки, один из примеров которой показан на рис. 9.35. При $\phi = 0$ шина предварительно заряжается до V_{DD} через транзистор M_2. Поскольку данное устройство совместно используется всеми входными элементами, его можно сделать достаточно большим, обеспечив малое время предварительной зарядки. При $\phi = 1$ емкость шины условно разряжается одним из согласующих транзисторов. Данная операция является медленной, поскольку большую емкость C_{bus} требуется разрядить с помощью маленького устройства M_1.

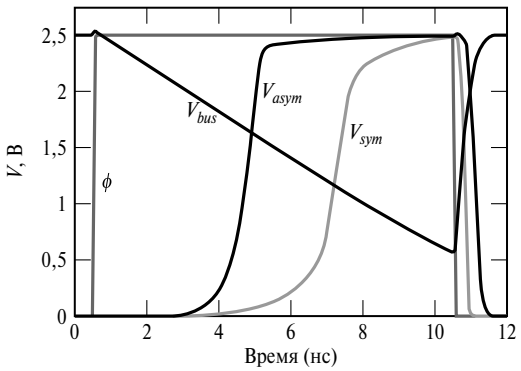
Быстродействие схемы можно повысить за счет помехоустойчивости, отметив, что в фазе оценки все транзисторы шины выполняют переход из высокого в низкое состояние. Наибольшее быстродействие достигается при поднятии порога переключения последующего инвертора, что дает асимметричный элемент. При проектировании инвертора согласно традиционным подходам транзисторы M_3 и M_4 масштабируются так, чтобы t_{pHL} и t_{pLH} были идентичны, а порог переключения (V_M) инвертора составлял порядка $0,5V_{DD}$. Сказанное означает, что перед переключением инвертора напряжение на шине V_{bus} должно упасть ниже $V_{DD}/2$. Этого можно избежать, увеличив размер p-МОП-устройства и V_M, что позволит выходному буферу начать переключение раньше.

Подход, основанный на предварительной зарядке, может привести к существенному ускорению при работе с линиями с большой емкостью. Тем не менее он также обладает всеми недостатками динамических схем — перераспределением заряда, утечкой и неизбежными потерями заряда из-за перекрестных помех и шума. Перекрестные помехи, наводимые соседними проводниками, представляют серьезную проблему (особенно в плотных схемах



Принципиальная схема

а



Смоделированный отклик для симметричного и несимметричного считывающих инверторов. Емкость нагрузки равна 1 пФ. В асимметричном случае M_4 в 15 раз шире

б

Рис. 9.35. Шина с предварительной зарядкой

разводки шин). Из-за уменьшенной помехоустойчивости NM_H асимметричного приемника описанная схема становится особо чувствительной к паразитным эффектам. Следовательно, при проектировании больших цепей с предварительной зарядкой необходимы особая внимательность и подробное моделирование. Если желательно сделать схему более устойчивой, для начала можно превратить ее в псевдостатическую, добавив небольшое устройство восстановления уровня.

На рис. 9.35, *б* показана смоделированная характеристика шины с предварительной зарядкой. Выходной сигнал показан и для симметричного, и для асимметричного выходного инверторов. Сдвиг вверх порога переключения уменьшает задержку распространения более чем на 2,5 нс. Это позволяет дополнительно снизить фазу оценки, уменьшив размах напряжения шины на 0,6 В и потребление энергии на 18%. Кроме того, чтобы выжать из описанной схемы максимум возможного, потребуется аккуратное тактирование сигналов предварительной зарядки/оценки.

Динамическая реализация рассмотренного принципа представлена на рис. 9.36, где изображен задающий генератор, управляемый импульсами. Его идея состоит в том, чтобы так контролировать время зарядки/разрядки за-

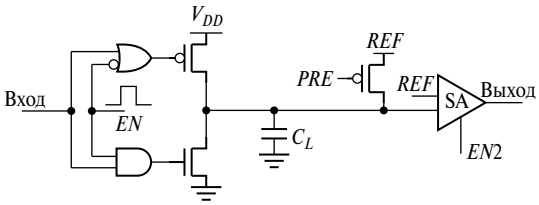


Рис. 9.36. Управляемый импульсами задающий генератор с усилителем считывания

дающих генераторов, чтобы в межсоединении получался искомый размах напряжения. Провод межсоединения предварительно заряжается до эталонного напряжения REF, обычно равного $V_{DD}/2$. Приемник состоит из псведодифференциального усилителя считывания, который сравнивает напряжение на межсоединении с REF. Поскольку усилитель потребляет статическую мощность, его можно активизировать только на небольшое время. Достоинство данной схемы заключается в том, что ширину импульса можно подстроить так, чтобы добиться очень малого размаха, когда дополнительное напряжение источника питания не требуется. Описанная концепция широко используется при разработке запоминающих устройств. Тем не менее она хороша только тогда, когда заранее известны емкостные нагрузки. Более того, при выключенном задающем генераторе межсоединение получается незаземленным, а следовательно, более чувствительным к шуму.

Описанная выше структура представляет лишь небольшую подгруппу из множества разработанных схем с уменьшенным размахом. Например, используя концепцию перераспределения заряда, можно создавать шины, работающие при меньшем размахе напряжения при перебросе заряда из одного проводника в другой. Если вас интересует эта тема — обращайтесь к соответствующим справочникам, например [Zhang00]. Исследуя область соединительных цепей с уменьшенным размахом, следует всегда помнить о сопутствующих компромиссах: мощность и быстродействие против достоверности сигнала и надежности.

9.5.2. Техники передачи на переключателях тока

Во всех описанных подходах предполагается, что данные, передаваемые через проводник, представлены набором уровней напряжения, определенных относительно напряжения шин питания. Хотя такой подход согласуется с использованием логических уровней в цифровых логических схемах, он не обязательно представляет лучшее решение с точки зрения быстродействия, мощности и надежности. Рассмотрим задачу передачи бита по длинной линии передачи с характеристическим импедансом 50 Ом. На рис. 9.37, *а* изображена традиционная реализация, основанная на переключателях напряжения. Задающее устройство с импедансом R_{out} переключает линию между двумя шинами питания, которые представляют соответственно **1** и **0**. Приемник представляет собой КМОП-инвертор, сравнивающий входное напряжение с

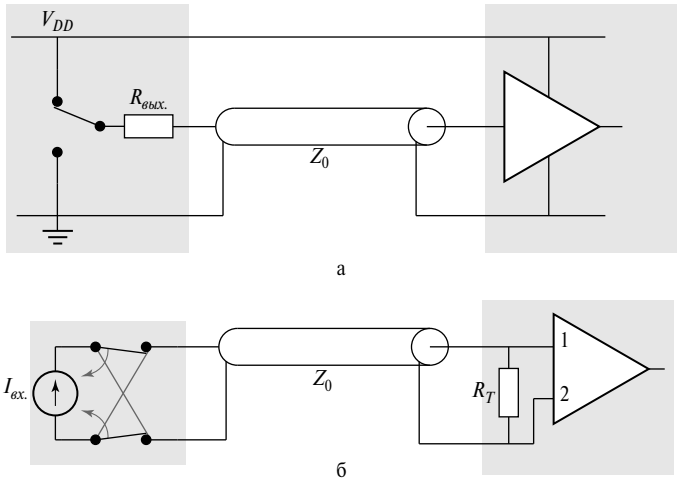


Рис. 9.37. Системы передачи, основанные на переключателях напряжения и тока. Преимуществом последних является более высокая устойчивость к шуму в цепях питания

эталонным пороговым напряжением, обычно представляющим среднее между напряжениями двух шин питания. Размах сигнала ограничен снизу по соображениям, касающимся шума. Точнее, шум в цепях питания сильно влияет как на уровни сигнала, так и на порог переключения приемника (который, кстати, сильно зависит от колебаний параметров технологического процесса).

Альтернативой является использование системы передачи с малым размахом, основанной на переключателях тока (см. рис. 9.37, *б*). Задающий генератор вводит в линию ток I_{in}, представляющий **1**, и обратный ток $-I_{in}$, представляющий **0**. Это индуцирует в линии передачи напряжение $2 \times I_{in} \times Z_0$, которое в конечном итоге поглощается параллельной оконечной нагрузкой с сопротивлением R_T. Время сведения зависит от того, насколько хорошо оконечное сопротивление согласовано с характеристическим импедансом линии передачи. Для детектирования изменения напряжения на R_T применяется дифференциальный усилитель. Обратите внимание на то, что сигнал и его путь возврата изолированы от шин питания и сопутствующего шума, в результате чего все шумы источника питания дифференциальный приемник воспринимает как общие синфазные помехи. Разработчики аналоговых систем с радостью подтвердят, что шумы такого типа легко подавляются в любой достаточно серьезной разработке с дифференциальным усилителем.

Хотя с помощью обоих подходов можно добиться приблизительно одинакового быстродействия, использование переключателей тока предпочтительнее с точки зрения (динамического) рассеяния мощности. Из-за нечувствительности к шуму источника питания такие схемы могут работать при гораздо меньшей помехоустойчивости, чем сети на основе переключателей напряжения, а следовательно, и при гораздо меньшем размахе. Максимальное напряжение сигнала, распространяющегося по линии передачи, может

не превышать 100 мВ. Основной проблемой разработки эффективной цепи на переключателях тока является статическое потребление мощности. Впрочем, в супервысокоскоростных сетях, где основную роль играет динамическая мощность, этот момент не представляет особых проблем.

КМОП-системы передачи на основе переключателей тока становятся достаточно популярными в области высокоскоростных внекристальных соединений. Из-за огромной потребности в пропускной способности ввода-вывода в быстродействующих процессорах и запоминающих устройствах данная тема приобрела первостепенную важность. Читатели, интересующиеся высокоскоростной внутрикристальной передачей сигналов, могут почерпнуть много полезной информации из [Dally98] и [Sidiropoulos01].

9.6. Перспектива: сети на кристалле

Поскольку размеры кристаллов и тактовая частота постоянно увеличиваются, а размер элемента становится все меньше, понятно, что проблема межсоединения вскоре встанет перед нами довольно остро. Физические соображения (скорость света и тепловой шум) налагают ограничения на то, насколько быструю и надежную связь можно установить на "больших" расстояниях. Как узкие места связи были основным фактором, сдерживавшим быстродействие гиперсуперкомпьютеров, теперь подобные ограничения начинают ограничивать быстродействие интегральных схем (систем на кристалле). Новые решения, зародившиеся на уровне технологии и схем, помогают только временно отсрочить проблемы. В конечном счете решением является рассмотрение межсоединений на кристалле как задачи связи и применение техник и подходов, многие годы позволявших крупномасштабным системам связи и сетям работать правильно и надежно в мировом масштабе. Например, вас не удивляет правильная работа Интернета при его масштабе и количестве соединяемых точек. Секрет успеха этой сети заключается в хорошо продуманном наборе протоколов, которые изолируют и ортогонализируют различные функциональные возможности, факторы, влияющие на быстродействие и надежность. Таким образом, вместо того чтобы рассматривать внутрикристальные соединения как провода, соединяющие две точки, их нужно считать каналами связи, по которым данные передают согласно настройкам "качества обслуживания", задающим условия на пропускную способность, время ожидания и точность [Sgroi01].

Приведем такой пример: современные технологии внутрикристальной передачи через межсоединения спроектированы с запасом помехоустойчивости, достаточным для обеспечения правильной передачи бита при любых условиях. В принципе, можно предположить, что последующие разработки будут более экстремальными, возможно, они будут жертвовать целостностью сигнала для выигрыша в энергии/быстродействии и позволять ошибкам появляться в переданных сигналах. Эти ошибки можно исправлять с помощью другой схемы, спроектированной для коррекции ошибок, и (или) за счет повторной передачи поврежденных данных.

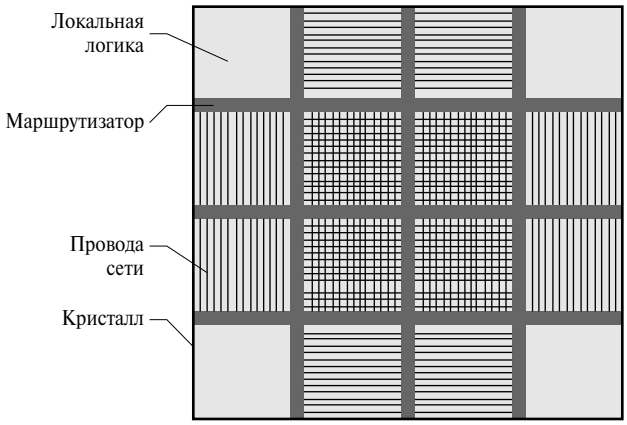


Рис. 9.38. Сеть на кристалле объединяет в себе несколько процессоров и их соединительную сеть

На более высоком уровне абстракции мы уже наблюдаем появление целых сетей на кристалле (рис. 9.38). Вместо обычного статического соединения источника данных с адресатом данные в таких сетях пакетами вводятся в сеть проводов, переключателей и маршрутизаторов, и именно сеть динамически определяет, как и куда направить эти пакеты через свои сегменты [Dally01]. В конечном итоге это единственный подход, который может быть надежным при существующем несоответствии размеров устройств и внутрикристальных расстояний.

9.7. Резюме

В данной главе описано несколько технологий, позволяющих бороться с влиянием межсоединений на быстродействие и надежность цифровых интегральных схем. Представленные паразитные нагрузки имеют двоякое влияние на работу схемы: они вводят шум и они увеличивают задержку распространения и рассеяние мощности. Итак, подведем итоги.

- *Емкостные перекрестные помехи* в плотных сетях влияют на надежность работы системы и, следовательно, на ее быстродействие. В таких условиях необходимо тщательное проектирование с использованием устройств с правильной структурой или применение сложных средств автоматизации разработки. Проводники, подобные шинам данных или цепям разводки тактовых сигналов, должны надлежащим образом экранироваться.
- Быстрое управление большими емкостями в КМОП-структурах требует введения *каскада буферов* с правильно выбранными размерами. К более сложным техникам относятся понижение размаха сигнала в длинных проводниках и использование переключателей тока.

- Удельное сопротивление влияет на надежность схемы, вызывая *омические скачки напряжения*. Особенно это важно для *сетей питания*, где размеры проводов имеют большое значение.
- Дополнительная задержка, вводимая RC-эффектами, минимизируется с помощью *добавления инверторов* и использования лучших технологий межсоединений.
- *Индуктивность межсоединения* играет главную роль при высоких скоростях переключения. В настоящее время один из наибольших вкладов в индуктивность вносит корпус кристалла. При переходе к более быстрым технологиям потребуются лучшие технологии корпусирования.
- *Отражение от земли*, вводимое скачком напряжения Ldi/dt в шинах питания, является одним из важнейших источников шума в современных интегральных схемах. Отражение от земли можно уменьшить, обеспечив достаточное количество выводов источника питания и контролируя наклоны внекристальных сигналов.
- Эффекты линии передачи становятся все более и более серьезной проблемой в сверхгигагерцовых разработках. Единственным средством борьбы с задержкой линии передачи является введение правильной оконечной нагрузки.
- В конечном счете при решении проблем межсоединений мы должны выбирать упреждающие подходы. На конструктивной основе, используя решения на архитектурном и системном уровнях, можно добиться очень многого.

9.8. Для любознательных

Прекрасный обзор вопросов, связанных с межсоединениями в цифровых разработках, можно найти в [Bakoglu90], [Dally98] и [Chandrakasan01].

9.9. Список использованной литературы

[Adler00] V. Adler and E. Friedman, “Uniform Repeater Insertion in RC Trees”, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 47, no. 10, Oct. 2000.

[Apollo] Apollo-II, High-Performance VDSM Place and Route System for Soc Designs, `http://www.synopsys.com/products/avmrg/apolloII_ds.html`.

[Bakoglu90] H. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, Addison-Wesley, 1990.

[Burd00] T. Burd, T. Pering, A. Stratakos, and R. Brodersen, “A Dynamic Voltage Scaled Microprocessor System”, *IEEE ISSCC Dig. Tech. Papers*, Feb. 2000, p. 294–295.

[Cadence-Power] R. Saleh, M. Benoit, and P. McCrorie, "Power Distribution Planning", `http://www.cadence.com/whitepapers/powerdistplan.html`, Cadence Design, 2001.

[Cadence-X Initiative] The X Initiative, `http://www.cadence.com/industry/x2.html`, Cadence Design, 2001.

[Chandrakasan01] A. Chandrakasan, W. Bowhill, and F. Fox, Ed., *Design of High-Performance Microprocessor Circuits*, IEEE Press, 2001.

[Cong99] J. Cong and D.Z. Pan "Interconnect Estimation and Planning for Deep Submicron Designs", *Proc. 36th ACM/IEEE Design Automation Conf.*, New Orleans, LA., June, 1999, p. 507–510.

[Dally98] B. Dally, *Digital Systems Engineering*, Cambridge University Press, 1998.

[Dally01] W. Dally, "Route Packets, Not Wires: On-Chip Interconnection Networks", *Proceedings Design Automation Conference*, June 2001, p. 684–689, Las Vegas.

[Etter93] D. Etter, "Engineering Problem Solving with Matlab", Prentice Hall, 1993.

[Gieseke97] B. Gieseke et al., "A 600 MHz superscalar RISC microprocessor with out-of-order execution", *IEEE ISSCC Digest of Technical Papers*, Febr. 1997, p. 176-177.

[Herrick00] B. Herrick, "Design Challenges in Multi-HHz Microprocessors", *Proceedings ASPDAC 2000*, Yokohama, January 2000.

[Johnson93] H. Johnson and M. Graham, *High-Speed Digital Design — A Handbook of Black Magic*, Prentice Hall, 1993.

[Khatri01] S. Khatri, R. K. Brayton, A. L. Sangiovanni-Vincentelli, *Cross Talk Immune VLSI Design Using Regular Layout Fabrics*, Kluwer Academic Publishers, June 2001.

[Ma94] S. Ma and P. Franzon, "Energy Control and Accurate Delay Estimation in the Design of CMOS Buffers", *IEEE Journal of Solid-State Circuits*, Sept. 1994, vol. 29, p. 1150–1153.

[Pentium02] Intel Pentium 4 Processor Home Page, `http://www.intel.com/products/desk_lap/processors/desktop/pentium4`.

[RailMill] "RailMill Datasheet.Power Network Analysis to Assure IC Performance", `http://www.synopsys.com/products/phy_syn/railmill_ds.html`, Synopsys, Inc.

[Restle01] P. Restle Home Page, `http://www.research.ibm.com/people/r/restle`, IBM Research, 2001.

[Sgroi01] M. Sgroi, M. Sheets, A. Mihal, K. Keutzer, S. Malik, J. Rabaey, A. Sangiovanni-Vincentelli, "Addressing the System-on-a-Chip Interconnect Woes Through Communication-Based Design", *Proceedings Design Automation Conference*, Las Vegas, June 2001, p. 678–683.

[Sidiropoulos01] S. Sidiropoulos, C. Yang, and M. Horowitz, "High Speed Inter-Chip Signaling", Chapter 19, p. 397–425, in [Chandrakasan01].

[Sotiriadis01] P. Sotiriadis and A. Chandrakasan, "Reducing Bus Delay in Submicron Technology Using Coding", *Proceedings ASPDAC Conference 2001*, Yokohama, January 2001.

[Stan95] M. Stan and W. Burleson, "Bus-Invert Coding for Low-Power I/O", *IEEE Transactions on VLSI*, March 1995, p. 49–58.

[Sylvester98] D. Sylvester and K. Keutzer, "Getting to the Bottom of Deep Submicron", *Proceedings ICCAD Conference*, November 1998, p. 203, San Jose.

[Tao94] J. Tao, N. Cheung, and C. Hu, "An Electromigration Failure Model for Interconnects under Pulsed and Bidirectional Current Stressing", *IEEE Trans. on Devices*, April 1994, vol. 41, no. 4, p. 539–545.

[Zhang00] H. Zhang, V. George, J. Rabaey, "Low-swing on-chip Signaling Techniques: Effectiveness and Robustness", *IEEE Transactions on VLSI Systems*, June 2000, vol. 8, no. 3, p. 264–272.

9.9.1. Упражнения

Последнюю версию сборника задач и упражнений можно найти по адресу `http://bwrc.eecs.berkley.edu/IcBook`.

ГЛАВА

10

Синхронизация в цифровых схемах

Влияние расфазировки тактовых импульсов и дрожания на быстродействие и функциональные возможности

Альтернативные методологии тактирования

Синхронизация в цифровых интегральных схемах

Генерация тактовых импульсов

10.1. Введение
10.2. Классификация цифровых систем с позиции синхронизации
10.2.1. Синхронное межсоединение
10.2.2. Мезохронное межсоединение
10.2.3. Плезиохронное межсоединение
10.2.4. Асинхронное межсоединение
10.3. Синхронное проектирование — глубокий анализ перспектив
10.3.1. Основы синхронного тактирования
10.3.2. Источники расфазировки и дрожания
10.3.3. Технологии разводки тактовых сигналов
10.3.4. Тактирование на основе защелок*
10.4. Проектирование схем с внутренней синхронизацией*
10.4.1. Логические схемы с внутренней синхронизацией — асинхронный подход
10.4.2. Генерация сигнала завершения

10.4.3. Передача сигналов с самосинхронизацией
10.4.4. Практические примеры логики с внутренней синхронизацией
10.5. Схемы синхронизации и арбитры*
10.5.1. Схемы синхронизации — концепция и реализация
10.5.2. Арбитры
10.6. Синтез тактовых сигналов и синхронизация с использованием ФАПЧ*
10.6.1. Суть
10.6.2. Основные составляющие схем ФАПЧ
10.7. Перспективы
10.7.1. Распределенное тактирование и использование систем автоподстройки по частоте
10.7.2. Оптическое распределение тактового сигнала
10.7.3. Синхронное или асинхронное проектирование
10.8. Резюме
10.9. Для любознательных

10.1. Введение

Все последовательные схемы обладают одним общим свойством: для правильной работы схемы необходимо обеспечить строго определенный порядок переключений. Если это условие не выполняется, в запоминающие устройства могут записываться неправильные данные, что приведет к функциональному сбою. Эффективным и популярным способом реализации этого упорядочения является *синхронный* подход, при котором все запоминающие устройства системы обновляются одновременно с использованием глобального периодического синхронизирующего сигнала (т.е. глобального тактового сигнала). Требуемые функциональные возможности обеспечиваются путем наложения определенных строгих условий на генерацию тактовых сигналов и их доставку к запоминающим устройствам, расположенным по всему кристаллу; несогласованность часто приводит к нарушению функционирования.

Данная глава начинается с обзора различных методологий синхронизаций. Бо́льшая ее часть посвящена популярному *синхронному подходу*. Мы проанализируем влияние пространственных изменений тактового сигнала, называемых *расфазировкой синхронизирующих импульсов* (clock skew), и временны́х колебаний тактового сигнала, именуемых *дрожанием тактовых импульсов* (clock jitter), и введем технологии, позволяющие решить обе эти проблемы. Эти колебания фундаментально ограничивают быстродействие, которого можно добиться с использованием обычных методов проектирования.

На другом конце спектра находится подход, называемый *асинхронным проектированием* и позволяющий вообще избежать проблемы неопределенности тактовых сигналов, сняв потребность в распределяемых глобально

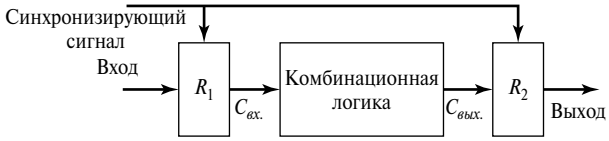


Рис. 10.1. Методология организации синхронного межсоединения

тактовых сигналах. После обсуждения основ проектирования асинхронных схем мы проанализируем сопутствующие этому подходу служебные издержки и определим некоторые практические приложения. Глубокого рассмотрения также заслуживает важный вопрос синхронизации *областей с различными тактовыми сигналами* и *сопряжения* асинхронных и синхронных систем. Наконец, мы представим основы внутрикристальной генерации тактовых сигналов с использованием обратной связи и рассмотрим текущие тенденции в сфере синхронизации.

10.2. Классификация цифровых систем с позиции синхронизации

Сигналы в цифровых системах можно классифицировать в зависимости от их связи с локальным тактовым сигналом ([Messerschmitt90], [Dally98]). Сигналы, выполняющие переходы только в предопределенные промежутки времени, можно отнести к *синхронным*, *мезохронным* или *плезиохронным* по отношению к системному тактовому генератору. С другой стороны, сигнал, который может выполнять переходы в любой момент времени, называется *асинхронным*.

10.2.1. Синхронное межсоединение

Синхронным называется сигнал, имеющий точно такую же частоту, как и локальный тактовый сигнал, и сохраняющий известную фиксированную разность фаз с этим сигналом. При такой структуре системы тактирования сигнал синхронизируется с тактовым генератором, и данные можно дискретизировать без какой-либо неопределенности. В сфере проектирования цифровых логических схем синхронные системы представляют наиболее простой и понятный тип межсоединений. Как показано на рис. 10.1, поток данных в такой цепи жестко пошагово синхронизирован с тактовым сигналом.

В данном случае входной сигнал данных In дискретизируется регистром R_1 и дает сигнал C_{in}, синхронизированный с системным тактовым генератором. Далее этот сигнал передается блоку комбинационной логики. По прошествии необходимого периода установки становится доступен выход C_{out}. Его значение выбирается регистром R_2, который синхронизирует выход с тактовым сигналом. По сути, *период определенности* сигнала C_{out} (время, в течение которого данные остаются достоверными) синхронизируется с системными

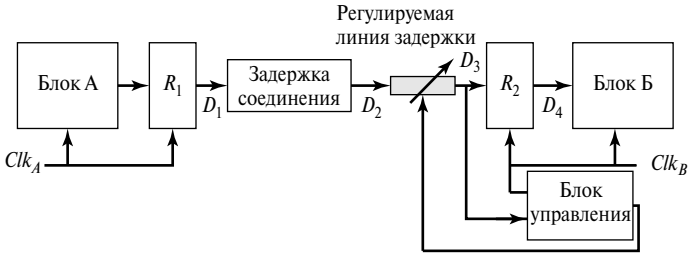


Рис. 10.2. Мезохронная связь с использованием регулируемой линии задержки

часами. Это позволяет регистру R_2 выбирать данные с полной уверенностью в их правильности. Продолжительность *периода неопределенности* (времени, в течение которого данные недостоверны) определяет верхнюю границу того, насколько быстро можно переключать синхронную систему.

10.2.2. Мезохронное межсоединение

Мезохронным ("мезо" в переводе с греч. — "середина") называется сигнал, который имеет ту же частоту, что и локальный тактовый сигнал, но сдвиг фаз которого относительного этого сигнала неизвестен. Например, если данные передают между двумя областями с различными тактовыми сигналами сигнал, передаваемый от первого модуля, может иметь неизвестную разность фаз с тактовым сигналом второго модуля. В такой системе из-за неопределенности разности фаз невозможно напрямую выбрать выход на принимающем модуле. Как показано на рис. 10.2, для синхронизации сигнала данных с тактовым сигналом принимающего модуля можно использовать (мезохронный) синхронизатор, который подстроит фазу принятого сигнала и обеспечит правильную выборку.

Рассмотрим рис. 10.2. Сигнал D_1 синхронизирован с Clk_A, а сигналы D_1 и D_2 мезохронные относительно Clk_B (поскольку разность фаз между Clk_A и Clk_B неизвестна, кроме того, неизвестна задержка тракта между блоками A и B). Задачей синхронизатора является такая настройка регулируемой линии задержки, чтобы сигнал D_3 (запаздывающая версия D_2) был выровнен с системными часами блока B. В данном примере для настройки элемента с переменной задержкой измеряется разность фаз между полученным сигналом и локальным тактовым сигналом. Регистр R_2 выбирает поступающие данные во время периода определенности, получающийся в результате сигнал D_4 синхронизирован с Clk_B.

10.2.3. Плезиохронное межсоединение

Плезиохронный сигнал имеет частоту, номинально совпадающую с частотой локального тактового сигнала, но все же немного отличающуюся. (В переводе с греч. "плезио" означает "почти".) В результате разность фаз сигнала данных и тактового сигнала изменяется со временем. Такой сценарий легко может

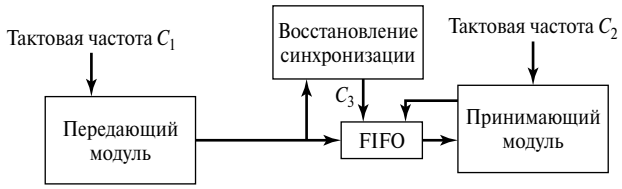


Рис. 10.3. Плезиохронная связь с использованием буфера FIFO

реализоваться в случае, когда два взаимодействующих модуля имеют независимые кварцевые генераторы тактовых импульсов. Поскольку переданный сигнал может поступать на принимающий модуль со скоростью, отличной от задаваемой локальным генератором, для обеспечения приема всех данных необходимо использовать какую-нибудь схему буферизации. Как правило, плезиохронные соединения возникают только в распределенных системах, содержащих протяженные линии связи, поскольку схемы, реализованные на уровне кристалла или даже печатной платы, для получения локальных тактовых сигналов обычно используют один общий генератор. Одно из возможных решений проблемы плезиохронных соединений показано на рис. 10.3.

Согласно данной схеме, передающий модуль выдает данные с некоторой неизвестной частотой C_1, почти синхронной (плезиохронной) C_2. Блок восстановления синхронизации отвечает за извлечение из потока данных тактовой частоты C_3 и буферизацию данных в буфере FIFO (“first in, first out” — “первым пришел, первым обслужен”). В результате сигнал C_3 будет синхронизирован с данными на входе буфера и будет мезохронным с сигналом C_1. Поскольку тактовые частоты передающего и принимающего модулей не согласованы, данные, возможно, придется отбросить (если скорость передачи выше, чем скорость приема) или можно будет продублировать (если скорость передачи ниже, чем скорость приема). Впрочем, если сделать буфер FIFO достаточно большим и периодически восстанавливать исходное состояние системы при возникновении переполнения, вполне реально можно реализовать надежную связь.

10.2.4. Асинхронное межсоединение

Асинхронные сигналы могут совершать переходы в любой момент времени — они не привязаны ни к какому локальному тактовому сигналу. В результате эти произвольные переходы не так просто отобразить в синхронный поток данных. Асинхронные сигналы можно синхронизировать, детектируя события и вводя в поток данных, синхронизированный с локальным тактовым сигналом, периоды ожидания. Впрочем, есть и более естественный путь обработки асинхронных сигналов — просто отказаться от использования локальных тактовых сигналов и задействовать принцип самосинхронизации. При таком подходе связь между модулями контролируется посредством протокола установления связи, гарантирующего правильное упорядочение операций.

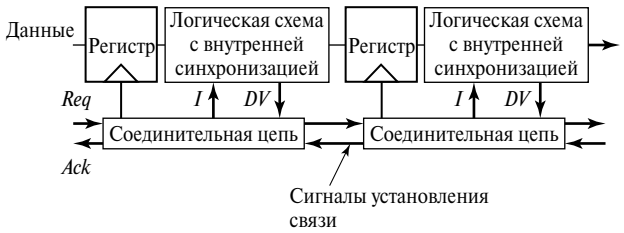


Рис. 10.4. Асинхронная методология проектирования для простого конвейерного соединения

Когда логический блок завершает операцию (рис. 10.4), он генерирует сигнал завершения DV, сообщающий о доступности выходных данных. Далее с помощью сигналов установления связи инициируется передача данных следующему блоку, который захватывает новые данные и начинает новый цикл обработки, генерируя сигнал инициализации I. Подобные асинхронные структуру удобны тем, что все вычисления выполняются с "родной" скоростью логического блока, и как только становятся доступны новые данные, становится возможной блочная обработка. Никаких проблем с *расфазировкой* не возникает, предлагаемая методология проектирования позволяет использовать удобный модульный подход, при котором взаимодействие блоков происходит с помощью протокола установления связи. Впрочем, подобные протоколы усложняют структуру, привносят в нее дополнительные служебные издержки, что, в свою очередь, сказывается на быстродействии.

10.3. Синхронное проектирование — глубокий анализ перспектив

10.3.1. Основы синхронного тактирования

Практически во всех системах, разрабатываемых в настоящее время, используются *синхронизирующие* тактовые сигналы. Генерация и распределение тактового сигнала существенно влияет на быстродействие и рассеяние мощности системой. Предположим пока, что мы рассматриваем систему, *управляемую положительным фронтом*, в которой нарастание тактового сигнала обозначает начало и завершение такта. В идеальном случае фаза тактового сигнала (т.е. положение фронта относительно эталонного сигнала) в различных точках системы одинакова (при этом предполагается, что доставка тактового сигнала от центрального распределительного пункта к каждому регистру сбалансирована идеально). На рис. 10.5 показана базовая структура синхронного конвейерного информационного тракта. В идеальном случае тактовые сигналы на регистрах 1 и 2 имеют одинаковый период и одновременно выполняют переходы.

Предположим, что известны следующие временны*ые* параметры последовательной схемы.

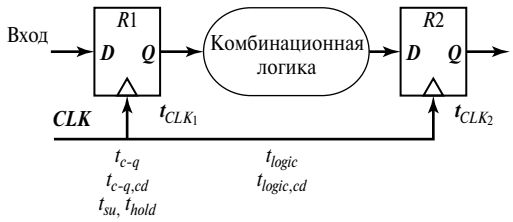


Рис. 10.5. Конвейерный информационный тракт и временны*е* параметры

- Минимальная ($t_{c-q,cd}$) и максимальная (t_{c-q}) задержка распространения регистра.
- Время установки (t_{su}) и удержания (t_{hold}) регистров.
- Минимальная ($t_{logic,cd}$) и максимальная (t_{logic}) задержка комбинационной логики.
- Положение нарастающих фронтов тактовых сигналов CLK_1 и CLK_2 (t_{clk1} и t_{clk2} соответственно) относительно глобального опорного сигнала.

В идеальном случае, когда $t_{clk1} = t_{clk2}$, минимальный такт данной последовательной схемы определяется исключительно наибольшей задержкой распространения. Такт должен быть достаточно большим, чтобы до следующего нарастающего фронта данные прошли через регистры и логику и задали правильный уровень сигнала в последнем регистре. Как было показано в главе 7, данное условие выражается следующим образом:

$$T > t_{c-q} + t_{logic} + t_{su}. \tag{10.1}$$

При этом время удержания последнего регистра должно быть меньше минимальной задержки распространения сигнала через логическую схему:

$$t_{hold} < t_{c-q,cd} + t_{logic,cd}. \tag{10.2}$$

К сожалению, выполненный анализ чересчур упрощен, поскольку тактирование никогда не бывает идеальным. Различные временны*е* события обычно и не идеально периодические, и не одновременные. В результате колебаний параметров процесса и изменений под влиянием внешней среды тактовые сигналы характеризуются пространственными и временны*м*и колебаниями, что приводит к снижению быстродействия и/или неверной работе схемы.

Расфазировка тактовых импульсов

Пространственные колебания времени прибытия перехода тактового сигнала в интегральную схему принято называть *расфазировкой тактовых импульсов* (clock skew). Фазовый сдвиг между двумя точками i и j интегральной схемы находится как $\delta(i,j) = t_i - t_j$, где t_i и t_j — положения нарастающего фронта тактового сигнала относительно глобального опорного сигнала. Рас-

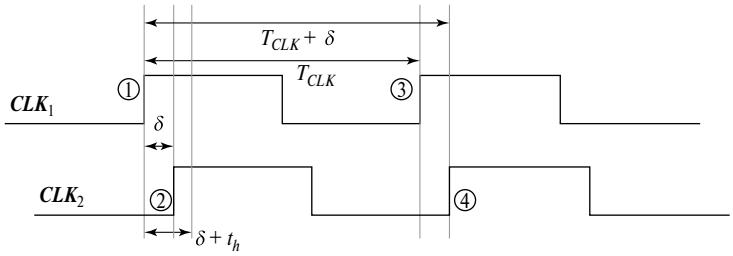


Рис. 10.6. Временная диаграмма, демонстрирующая влияние расфазировки тактовых импульсов ($\delta > 0$) на быстродействие и функционирование схемы

смотрим передачу данных между регистрами $R1$ и $R2$ (рис. 10.5). Фазовый сдвиг тактовых импульсов может быть положительным или отрицательным в зависимости от направления трассировки и положения источника тактового сигнала. На рис. 10.6 показана временная диаграмма для положительного фазового сдвига. На ней показано, что прибытие нарастающего фронта тактового сигнала на второй регистр задерживается на положительную величину δ.

Расфазировка тактовых импульсов порождается статическим рассогласованием цепей разводки тактовых сигналов и нагрузки на тактовый сигнал. По определению в последующих тактах фазовый сдвиг остается таким же, т.е. если в одном такте CLK_2 запаздывал относительно CLK_1 на δ, то в следующем такте он будет запаздывать на такую же величину. Здесь важно отметить, что расфазировка тактовых импульсов не приводит к изменению времени такта — только к сдвигу фазы.

Явление расфазировки тактовых импульсов очень сильно влияет как на быстродействие, так и на функционирование последовательных систем. Прежде всего рассмотрим влияние расфазировки на быстродействие. На рис. 10.6 видно, что новый вход In, выбранный регистром $R1$ в момент нарастания тактового сигнала ①, пройдет через комбинационную логику и будет выбран регистром $R2$ в момент нарастания тактового сигнала ④. При положительном сдвиге фаз время, в течение которого сигнал должен пройти от $R1$ до $R2$, увеличивается на величину δ. Выход комбинационной логики должен в течение промежутка времени, равного времени установки, иметь правильное значение до начала нарастания сигнала CLK_2 (точка ④). Таким образом, получаем следующее условие на минимальный размер такта:

$$T + \delta \geq t_{c-q} + t_{logic} + t_{su} \text{ или } T \geq t_{c-q} + t_{logic} + t_{su} - \delta. \qquad (10.3)$$

Из приведенного уравнения следует, что расфазировка тактовых импульсов может в действительности повысить быстродействие схемы, т.е. минимальный такт, требуемый для надежной работы цепи, уменьшается с увеличением сдвига фаз! Да, это действительно так, но, к сожалению, увеличение сдвига фаз делает цепь более восприимчивой к состязанию, что может разрушить правильную работу последовательных систем.

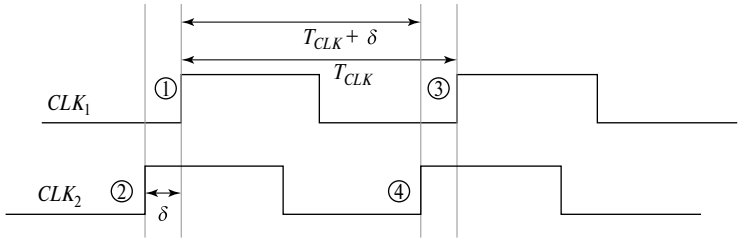


Рис. 10.7. Временная диаграмма для случая $\delta < 0$. Положительный фронт сигнала CLK_2 предшествует фронту CLK_1

Сказанное можно проиллюстрировать на следующем примере: снова предположим, что вход In выбирается регистром $R1$ при нарастании тактового сигнала CLK_1 (①). Новое значение на выходе регистра $R1$ должно пройти через комбинационную логику и быть доступным до момента ①. Однако, если минимальная задержка комбинационной логики *мала*, входы регистра $R2$ могут успеть измениться до момента ①, что приведет к неверной оценке значения. Чтобы избежать подобных состязаний, мы должны гарантировать, что минимальная задержка распространения через регистр и логику достаточно велика, чтобы выходы регистра $R2$ оставались неизменными после момента ② в течение промежутка времени, равного времени удержания. Формально данное условие можно записать следующим образом:

$$\delta + t_{hold} < t_{(c-q,cd)} + t_{(logic,cd)} \text{ или } \delta < t_{(c-q,cd)} + t_{(logic,cd)} - t_{hold}. \qquad (10.4)$$

На рис. 10.7 показана временн*а*я диаграмма для случая $\delta < 0$, когда нарастание сигнала CLK_2 происходит до нарастания сигнала CLK_1. На нарастающем фронте сигнала CLK_1 регистр $R1$ выбирает новый вход. Новые данные проходят через комбинационную логику и выбираются регистром $R2$ на нарастающем фронте сигнала CLK_2, что соответствует фронту ④. Как явно видно на рис. 10.7 и из уравнения (10.3), отрицательная разность фаз отрицательно влияет на быстродействие последовательной системы. Тем не менее, предполагая, что $t_{hold} + \delta < t_{c-q,cd} + t_{logic,cd}$, мы увидим, что система никогда не будет сбоить, поскольку фронт ② всегда предшествует фронту ①!

Для примера на рис. 10.8 показаны различные сценарии, при которых реализуется положительный и отрицательный фазовый сдвиг.

- Случай $\delta > 0$ соответствует трассировке тактового сигнала в том же направлении, в котором через конвейер идет поток данных (см. рис. 10.8, *а*). В этом случае расфазировку нужно тщательно контролировать согласно уравнению (10.4). Если данное условие нарушается, система функционирует **независимо от величины такта**. Уменьшение тактовой частоты схемы, управляемой фронтом, не помогает обойти проблемы расфазировки! Следовательно, в такой ситуации необходимо удовлетворить условию на время удержания во время проектирования. С другой стороны, положительный сдвиг фаз увеличивает пропускную способность схемы, как следу-

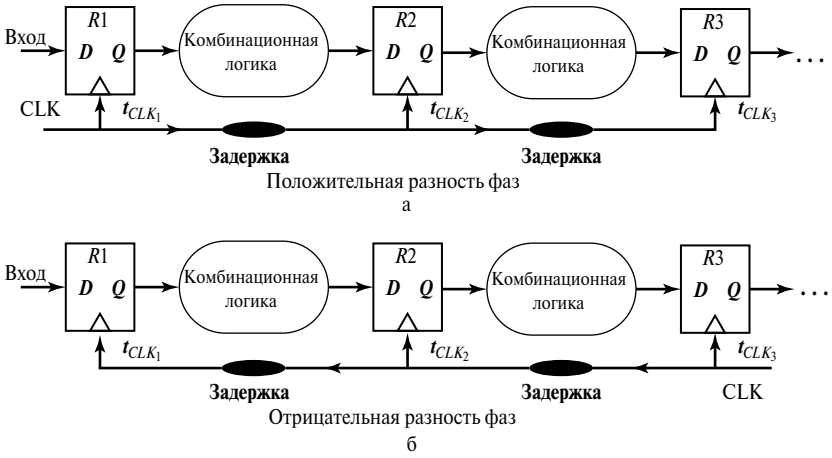


Рис. 10.8. Сценарии реализации положительного и отрицательного фазового сдвига

ет из уравнения (10.3). Величину такта можно уменьшить на δ. Подобное улучшение возможно лишь до определенного предела, поскольку большие значения δ рано или поздно приведут к нарушению условия (10.4).

- Случай $\delta < 0$. Когда тактовый сигнал трассируется в направлении, противоположном направлению потока данных (см. рис. 10.8, *б*), разность фаз отрицательна и обеспечивает существенную защиту от состязаний; если время удержания равно нулю или отрицательно, состязания вообще невозможны, поскольку система удовлетворяет уравнению (10.4) в любом случае! Расфазировка уменьшает время, доступное для реальных вычислений, так что такт необходимо увеличить на величину $|\delta|$. Короче говоря, если тактовый сигнал трассируется в направлении, противоположном потоку данных, это позволяет избежать сбоев, однако ухудшает быстродействие схемы.

К сожалению, поскольку в типичной логической схеме поток данных может идти в обоих направлениях (например, в цепях с обратной связью), данная возможность устранения состязаний доступна не всегда. На рис. 10.9, например, показано, что разность фаз может быть как положительной, так и отрицательной в зависимости от направления передачи данных. В такой ситуации разработчик должен учитывать наихудший вариант расфазировки. Вообще, трассировка тактовых сигналов, допускающая только отрицательную разность фаз, нереальна. Следовательно, требуется разрабатывать сеть разводки тактовых сигналов, допускающих только незначительную разфазировку.

Пример 10.1. Оценка максимальной и минимальной задержки распространения

Рассмотрим логическую схему, показанную на рис. 10.10. Определите минимальную задержку и задержку распространения сети для данной наиболь-

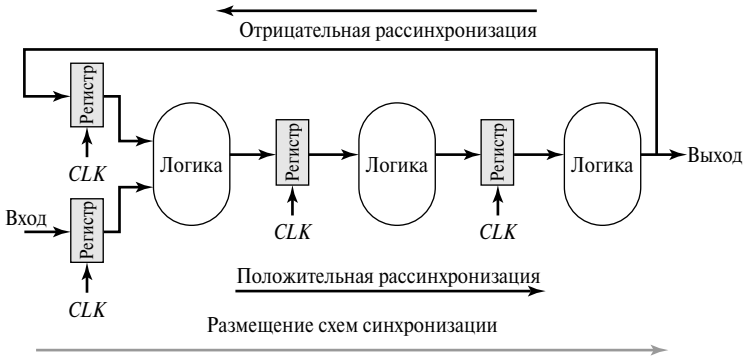


Рис. 10.9. Информационный тракт с обратной связью

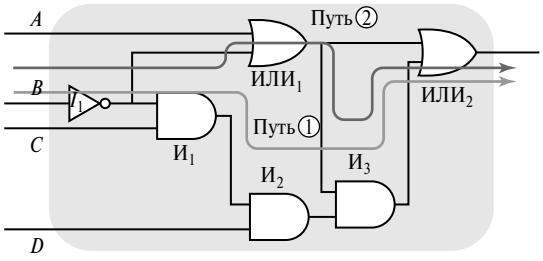


Рис. 10.10. Логическая сеть для расчета быстродействия

шей задержки элемента t_{gate}. Считайте, что максимальная и минимальная задержки элементов идентичны.

Найти минимальную задержку довольно просто — это задержка распространения сигнала через элементы ИЛИ$_1$ и ИЛИ$_2$, и она равна $2t_{gate}$. С другой стороны, вычислить наибольшую задержку распространения не так просто. На первый взгляд кажется, что наихудшим сценарием является прохождение сигнала по пути ①, и задержка в этом случае составляет $5t_{gate}$. Однако, анализируя зависимости логических элементов по данным, мы находим, что данный путь никогда не будет задействован. Поэтому путь ① называется *ложным*. Если $A = 1$, критический путь проходит через ИЛИ$_1$ и ИЛИ$_2$. Если $A = 0$ и $B = 0$, критический путь проходит через I_1, ИЛИ$_1$ и ИЛИ$_2$ (соответствует задержке $3t_{gate}$). В случае, когда $A = 0$ и $B = 1$, наиболее длинным является путь через I_1, ИЛИ$_1$, И$_3$ и ИЛИ$_2$. Другими словами, для данной простой (но запутанной) сети выход вообще не зависит от входов C и D (т.е. они являются *избыточными*). Следовательно, реальная задержка распространения равна $4t_{gate}$. Далее, зная задержку распространения и минимальную задержку сети, можно легко вычислить минимальную и максимальную расфазировку тактовых импульсов.

Предупреждение

Из-за наличия *ложных путей* расчет наибольшей задержки распространения для комбинационной логики нельзя выполнять простым суммированием задержек распространения отдельных логических элементов. Критический путь сильно зависит от топологии схемы и зависимостей по данным.

Дрожание тактовых импульсов

Дрожанием тактовых импульсов (clock jitter) называется временное колебание длительности такта в данной точке кристалла, т.е. длительность последовательных тактов может увеличиваться или уменьшаться. Данная мера неопределенности является чисто временно*й* и часто задается для конкретной точки. *Дрожание* можно измерить и охарактеризовать множеством различных способов, и оно *представляет собой случайную переменную с нулевым средним. Абсолютным дрожанием* (t_{jitter}) называется наибольшее отклонение (амплитуда) фронта синхроимпульса в данной точке относительно фронта идеального эталонного тактового сигнала. *Межтактовым дрожанием* (cycle-to-cycle jitter) (T_{jitter}) обычно называют зависящие от времени отклонения времени одного такта от времени идеального эталонного такта. Для данной точки в пространстве i справедливо соотношение $T^i_{jitter}(n) = t^i_{clk,n+1} - t^i_{clk,n} - T_{CLK}$, где $t^i_{clk,n+1}$ и $t^i_{clk,n}$ представляют время прибытия $(n+1)$-го и n-го фронта тактового сигнала на узел i соответственно, а T_{CLK} — это номинальная продолжительность такта. В худшем случае амплитуда межтактового дрожания равна удвоенному абсолютному дрожанию ($2t^i_{jitter}$).

Дрожание прямо влияет на быстродействие последовательных систем. Рассмотрим, например, рис. 10.11, на котором представлена номинальная продолжительность такта, а также изменение продолжительности такта. В идеальном случае такт начинается на фронте ② и заканчивается на фронте ⑤; его продолжительность равна T_{CLK}. Однако при реализации худшего сценария передний фронт текущего такта задерживается из-за дрожания (фронт ③), а передний фронт следующего текста, опять-таки из-за дрожания, наступает раньше (фронт ④). В результате общее время, доступное для завершения операции, в худшем случае уменьшается на $2t_{jiiter}$ и выражается следующим образом:

$$T_{CLK} - 2t_{jitter} \geq t_{c-q} + t_{logic} + t_{su} or T \geq t_{c-q} + t_{logic} + t_{su} + 2t_{jitter}. \qquad (10.5)$$

Из уравнения (10.5) видно, что дрожание прямо снижает быстродействие последовательной схемы. Если быстродействие схемы имеет значение, это влияние необходимо свести к минимуму.

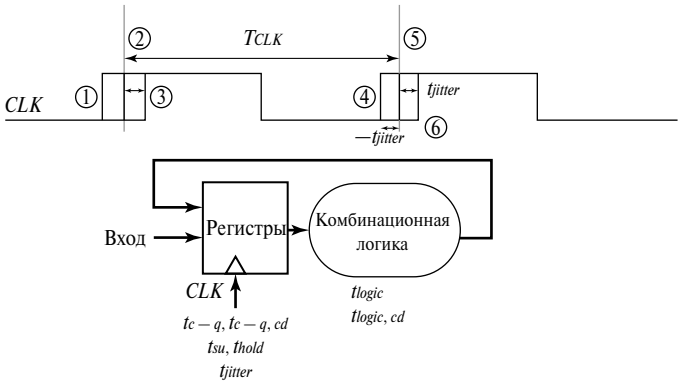


Рис. 10.11. Схема для изучения влияния дрожания на быстродействие

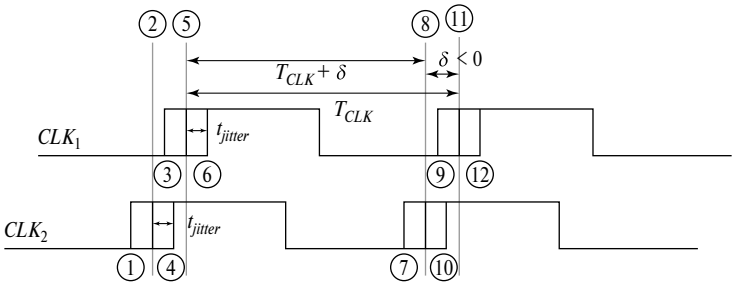


Рис. 10.12. Последовательная схема с отрицательным фазовым сдвигом тактовых импульсов δ. Предполагается, что фазовый сдвиг больше, чем величина дрожания

Объединенное влияние расфазировки и дрожания

В данном разделе для обычной синхронизации, управляемой фронтами, изучается объединенное влияние расфазировки и дрожания. Итак, рассмотрим последовательную схему, показанную на рис. 10.12.

Предположим, что в результате распределения тактового сигнала возник статический фазовый сдвиг δ между тактовыми сигналами в двух регистрах (допустим, что $\delta > 0$). Более того, оба тактовых сигнала испытывают дрожание t_{jitter}. Чтобы определить условие на минимальную продолжительность такта, необходимо найти минимальное доступное время, требуемое для выполнения необходимых вычислений. Наихудший сценарий реализуется, когда передний фронт текущего такта сигнала CLK_1 наступает поздно (фронт ③), а передний фронт следующего такта CLK_2 — рано (фронт ⑩). В результате получаем следующее условие:

$$T_{CLK} + \delta - 2t_{jitter} \geq t_{c-q} + t_{logic} + t_{su}$$

или

$$T \geq t_{c-q} + t_{logic} + t_{su} - \delta + 2t_{jitter}. \qquad (10.6)$$

Из данного условия видно, что положительный фазовый сдвиг может быть выгодным с точки зрения быстродействия. С другой стороны, *дрожание* всегда отрицательно сказывается на минимальной продолжительности такта[1].

Чтобы сформулировать условие минимальной задержки, рассмотрим ситуацию, когда передний фронт такта CLK_1 прибывает на узел рано (фронт ①), а передний фронт текущего такта CLK_2 прибывают поздно (фронт ⑥). Время между фронтами ① и ⑥ должно быть меньше, чем минимальная задержка распространения через сеть. Таким образом, получаем следующее:

$$\delta + t_{hold} + 2t_{jitter} < t_{(c-q,cd)} + t_{(logic,cd)}$$

или (10.7)

$$\delta < t_{(c-q,cd)} + t_{(logic,cd)} - t_{hold} - 2t_{jitter}.$$

Из данного соотношения видно, что приемлемый фазовый сдвиг уменьшается на величину, равную *дрожанию* обоих сигналов.

Рассмотрим теперь случай, когда фазовый сдвиг отрицательный ($\delta < 0$), как показано на рис. 10.12. Предположим, что $|\delta| > t_{jitter}$. Можно показать, что при реализации худшего сценария получаются такие же временны́е соотношения, как и в предыдущем анализе, только с отрицательным δ, т.е. отрицательный фазовый сдвиг также снижает быстродействие.

10.3.2. Источники расфазировки и дрожания

Идеальным тактовым сигналом называется периодический сигнал, одновременно активизирующий различные запоминающие устройства кристалла. Однако из-за множества причин тактовые сигналы являются неидеальными. Чтобы проиллюстрировать источники расфазировки и дрожания, рассмотрим показанное на рис. 10.13 упрощенное представление сети генерации и разводки тактового сигнала. Высокочастотный тактовый сигнал либо поставляется извне кристалла, либо генерируется в самом кристалле. В любом случае из некоторой центральной точки тактовый сигнал распределяется с использованием множества *согласованных* путей к низкоуровневым последовательным элементам. На рисунке показаны два таких пути. Пути разводки тактовых сигналов состоят из проводки и связанных с ней распределенных буферов, которые требуются для управления межсоединениями и нагрузками. Главное, что вы должны понять относительно разводки тактовых сигналов: абсолютная задержка распространения сигнала в сети разводки не важна; важным является относительное время прибытия сигналов в точки в конце каждого пути. Тактовому сигналу может требоваться несколько тактов для прохождения от центрального распределительного пункта к низкоуровнево-

[1] Приведенный анализ описывает наихудший сценарий. Мы предполагали, что дрожание на источнике и адресате информации описывается независимыми статистическими переменными. В реальной жизни фронты тактовых сигналов, задействованные в анализе времени удержания, выводятся из одного фронта тактового сигнала и являются статистически зависимыми. Если учитывать эту зависимость, временны́е условия существенно ослабляются.

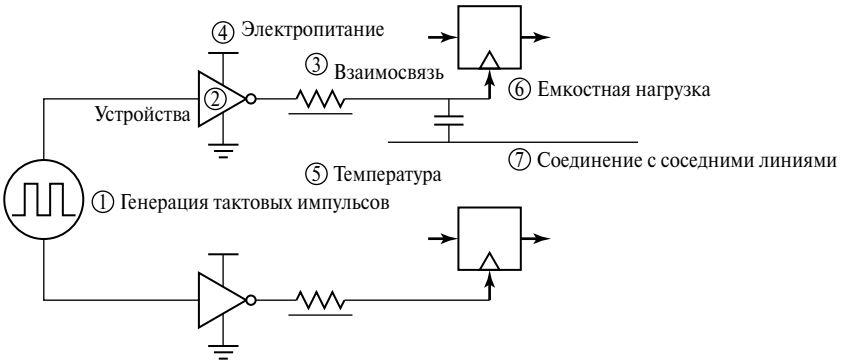


Рис. 10.13. Источники расфазировки и дрожания в синхронных сетях разводки тактовых сигналов

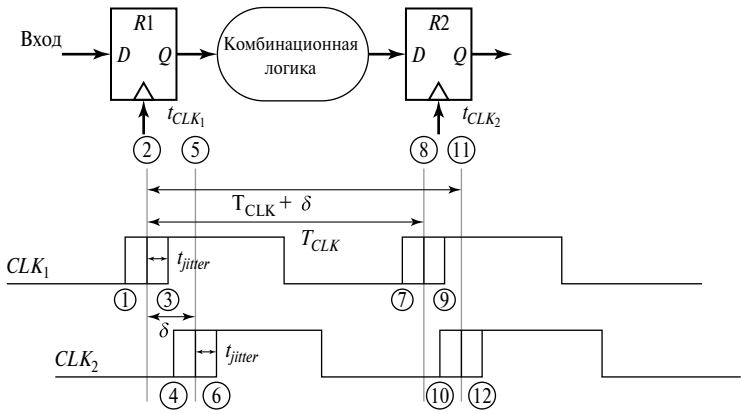


Рис. 10.14. Последовательная схема для изучения влияния расфазировки и дрожания на системы, управляемые фронтом. В данном примере предполагается, что фазовый сдвиг δ положительный

му регистру, однако все тактовые сигналы обязательно должны прибывать одновременно на все регистры кристалла.

Существует множество причин, по которым два параллельных пути имеют различные задержки. Источники неопределенности значения тактового сигнала можно классифицировать по-разному. Например, источники тактовой частоты можно разделить на две категории: *систематические* и *случайные*. *Систематические* ошибки номинально идентичны для всех кристаллов и предсказуемы (например, колебания общей емкости нагрузки каждого пути тактового сигнала). В принципе такие ошибки можно смоделировать и устранить еще на этапе проектирования (разумеется, при наличии хороших моделей и инструментов моделирования). Помимо этого, систематические ошибки можно найти из измерений, выполненных на партии кристаллов, и соответствующим образом модифицировать структуру схемы. *Случайные* ошибки вызываются колебаниями процесса производства, который сложно смодели-

ровать и устранить (например, флуктуации легирующей примеси, которые приводят к колебаниям пороговых значений).

Рассогласование можно также охарактеризовать как *стационарное* или *нестационарное*. На практике необходимо разделять изменения, происходящие медленнее и быстрее, чем рассматриваемая постоянная времени. Например, характерное время изменения градиентов температур на кристалле измеряется миллисекундами. Сеть разводки тактового сигнала, настраиваемая путем однократной калибровки, чувствительна к нестационарному рассогласованию, вызванному переменными градиентами температур. С другой стороны, изменения температуры кажутся практически стационарными для цепи обратной связи с полосой пропускания шириной в несколько мегагерц. В качестве другого примера можно рассмотреть шум в цепях питания. Сеть разводки тактовых сигналов обычно является наибольшей сетью кристалла, а одновременные переходы тактовых сигналов наводят шумы в цепях питания. Данный высокоскоростной эффект не приводит к появлению нестационарного рассогласования, поскольку он повторяется в каждом такте и одинаково влияет на каждый положительный фронт тактового сигнала. Разумеется, данная импульсная помеха в цепях питания может вызывать стационарное рассогласование, если она не одинакова для всех точек кристалла. Ниже мы подробное рассмотрим все источники расфазировки и дрожания, представленные на рис. 10.13.

Генерация тактового сигнала (1)

Генерация тактового сигнала сама по себе уже порождает *дрожание*. Типичный внутрикристальный тактовый генератор, подобный описанному в конце главы, получает низкочастотный опорный тактовый сигнал и генерирует высокочастотный глобальный эталонный сигнал для процессора. Основой такого генератора является *ГУН — генератор, управляемый напряжением* (Voltage-Controlled Oscillator — VCO). Это аналоговая схема, чувствительная к внутренним шумам устройства и изменениям в цепи питания. Основной проблемой здесь является связь через подложку с окружающими зашумленными цифровыми схемами. Особенно эта проблема существенна для современной технологии производства, в которой используется слаболегированная эпитаксия на сильнолегированной подложке. Это приводит к тому, что шум распространяется в кристалле на большие расстояния [Maneatis00]. Вообще, такие источники шума вызывают временны*е* колебания тактового сигнала и приводят к изменению продолжительности такта.

Вариации параметров устройств (2)

Неотъемлемой частью цепей разводки тактовых сигналов являются распределенные буферы. Они необходимы для управления как нагрузкой регистра, так и глобальными и локальными межсоединениями. Для минимизации временно*й* неопределенности критически важно согласование устройств в буферах, расположенных вдоль различных трактов разводки тактовых сигналов. К сожалению, из-за вариаций параметров процесса параметры устройств

в буферах отличаются для разных трактов, что приводит к *статической расфазировке*. Существует множество источников вариаций, например вариации окисла (влияют на усиление и порог) вариации легирующих примесей и размеров (ширины и длины). Вариации легирующих примесей могут влиять на глубину *pn*-переходов и профили распределения примеси, что приводит к колебаниям электрических параметров (подобных порогу устройства и паразитным емкостям).

Некоторое влияние на параметры устройства может также иметь ориентация поликремния. Кроме того, чрезвычайно важны вариации критического размера поликремния, поскольку он прямо влияет на длину канала МОП-транзистора, управляющий ток и параметры переключения. Пространственные вариации обычно представлены вариациями на уровне пластины и вариациями на уровне кристалла. По крайней мере, часть этих вариаций относится к систематическим, а следовательно, их можно смоделировать и подавить. Тем не менее существуют еще случайные вариации, которые в конечном счете ограничивают возможности согласования и определяют минимальный уровень расфазировки, избавиться от которого невозможно.

Вариации межсоединений (3)

Вариации вертикальных и горизонтальных размеров приводят к колебаниям емкости и сопротивления межсоединения в разных точках кристалла. Поскольку эти вариации статичны, они порождают разность фаз между тактовыми сигналами в различных трактах. Одним из важных источников вариаций межсоединений являются колебания толщины *межслойного диэлектрика* (Inter-Layer Dielectric — ILD). При формировании алюминиевого межсоединения между слоями с рисунком металлизации вводятся слои диоксида кремния. Оксид размещается поверх слоя металлических элементов, обычно создавая некоторые остаточные перепады высот или колебания топографии поверхности. Для сглаживания поверхности и удаления топографии, порожденной расположением элементов и вытравливанием, применяется *химико-механическая полировка* (как описано в главе 3 и показано на рис. 10.15, *а*). Если на уровне элементов (т.е. в отдельном металлическом межсоединении) химико-механическая полировка позволяет добиться идеального сглаживания, на глобальном уровне у этого метода становятся заметными определенные ограничения. Это вызвано в основном вариациями скорости полировки, которая зависит от плотности топологии схемы и интерференционных эффектов. Сказанное иллюстрируется на рис. 10.15, *б*, где показано, что для областей с меньшей пространственной плотностью скорость полировки выше, а следовательно, толщина диэлектрика будет ниже, а емкость выше.

Вариации в разработке и производстве полупроводниковых устройств жизненно важно оценить и проконтролировать. Поэтому было потрачено много усилий на развитие аналитических моделей, позволяющих оценивать колебания толщины межслойного диэлектрика на основе пространственной плотности. Поскольку данную величину часто можно предсказать по топологии схемы, систематическую ее составляющую можно в принципе скорректировать в процессе проектирования (например, введя подходящие задерж-

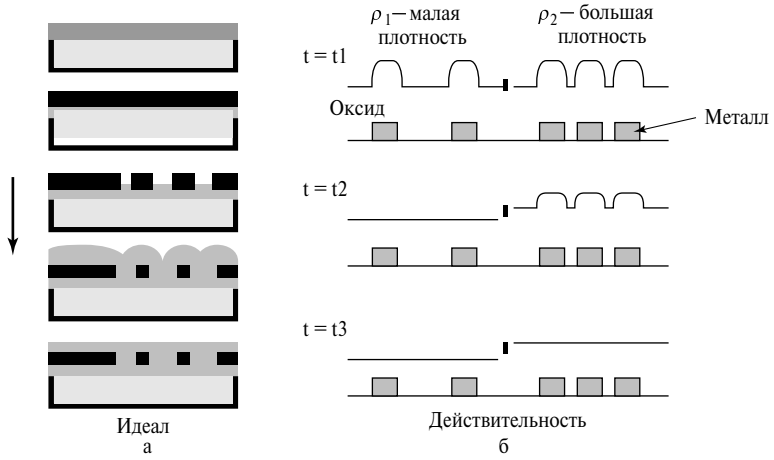


Рис. 10.15. Колебания толщины межслойного диэлектрика, вызванная колебаниями плотности. (Перепечатано с разрешения Дуан Бонинг (Duane Boning).)

ки или сделав плотность более однородной, добавив “ложные заливки”). На рис. 10.16 показана функция распределения плотности и толщины межслойного диэлектрика для быстродействующего микропроцессора. На графике четко видна корреляция плотности и толщины диэлектрика. Следовательно, в цепях разводки тактовых сигналов данная информация должна использоваться для уменьшения расфазировки тактовых импульсов.

Помимо названных, к вариациям межсоединений относят колебания толщины проводников и расстояния между дорожками, порождаемые зависимостью структуры от процессов фотолитографии и травления. На нижних слоях металлизации литографические эффекты играют более важную роль, а эффекты травления, зависящие от ширины и топологии, доминируют на более высоких слоях. Ширина также является критическим параметром, поскольку она непосредственно влияет на сопротивление дорожки, а расстояние между дорожками влияет на емкость между проводами. Если вас интересует эта тема, подробный обзор вариаций параметров устройств и межсоединений представлен в [Boning00]. В современных процессорах используются медные межсоединения, в которых колебания толщины дорожки сильно зависят от шаблона из-за эффектов химико-механической полировки и эрозии [Park00].

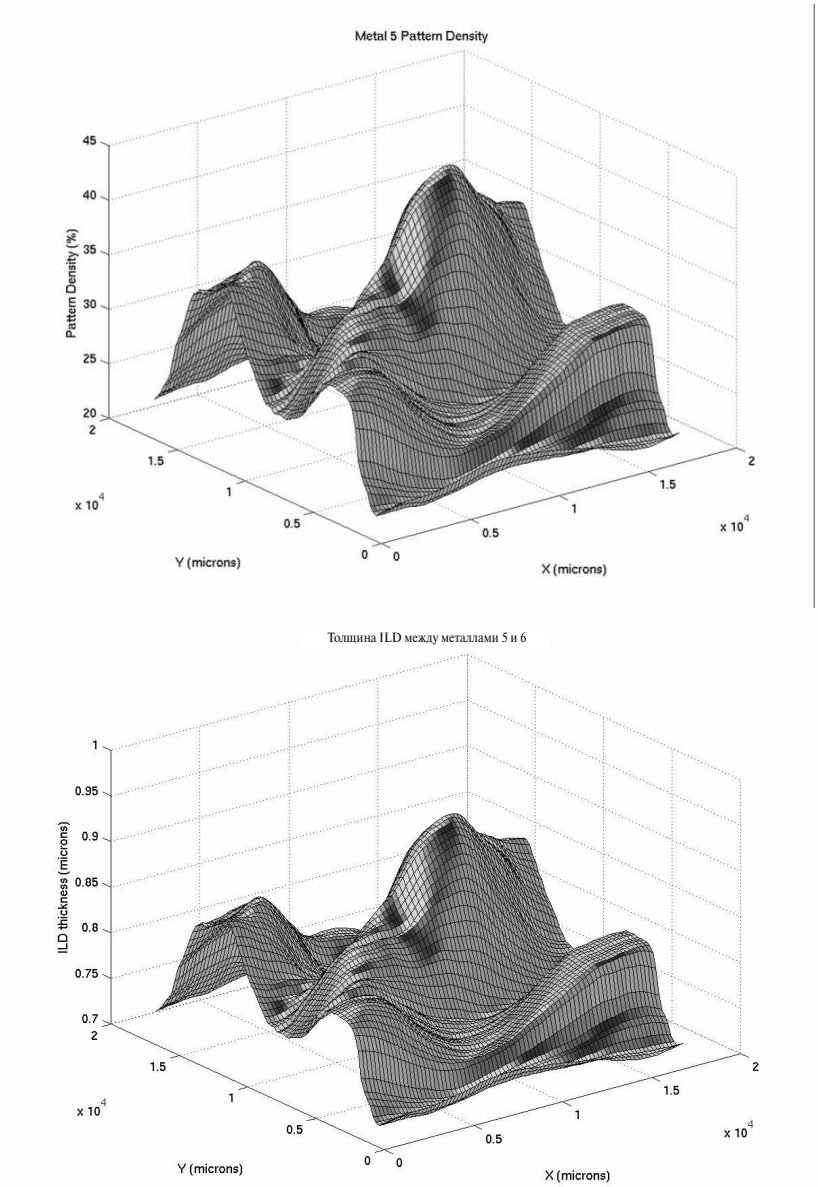


Рис. 10.16. Изменение плотности шаблона и толщины межслойного диэлектрика для быстродействующего процессора. (Перепечатано с разрешения Дуан Бонинг (Duane Boning).)

Изменения под влиянием внешней среды (4 и 5)

Изменения под влиянием внешней среды вносят, пожалуй, самый весомый вклад в рассинхронизацию и дрожание. Двумя основными источниками изменений под влиянием внешней среды являются *температура* и *источник электропитания*. Наличие в кристалле градиентов температур объясняется рассеянием энергии кристаллом. Данные градиенты могут быть достаточно большими, как показано на рис. 10.17, где представлен моментальный снимок температуры поверхности микропроцессора DEC 21064. Колебания температуры играет весьма важную роль при *стробировании тактовых импульсов*, когда часть кристалла находится в холостом режиме, а часть — работает на полную мощность. Стробирование тактовых импульсов (clock gating) стало популярным не так давно и используется как средство минимизации рассеяния мощности в неработающих модулях (см. ниже). Отключение частей кристалла приводит к большим колебаниям температур. Поскольку параметры устройства (например, порог и подвижность) сильно зависят от температуры, задержка цепи разводки тактового сигнала, вызванная буферизацией, может сильно отличаться для разных трактов. Мало того, данный компонент является нестационарным, поскольку температура изменяется в зависимости от активности логики схемы. Следовательно, проблему нельзя решить простым моделированием сетей разводки тактовых сигналов для крайних точек температурного диапазона, считая это наихудшим сценарием; на самом деле необходимо смоделировать наихудший сценарий колебания температуры. Интересным также представляется вопрос, в какое явление вносит вклад колебание температуры — в расфазировку или в дрожание. Понятно, что разность температур — это величина нестационарная, однако изменения происходят сравнительно медленно (типичное характерное время изменения температуры составляет порядка нескольких миллисекунд). Поэтому считается, что колебания температур определяют расфазировку и анализ выполняется исходя из наихудших условий. К счастью, с помощью обратной связи температуру можно откалибровать и компенсировать данный эффект.

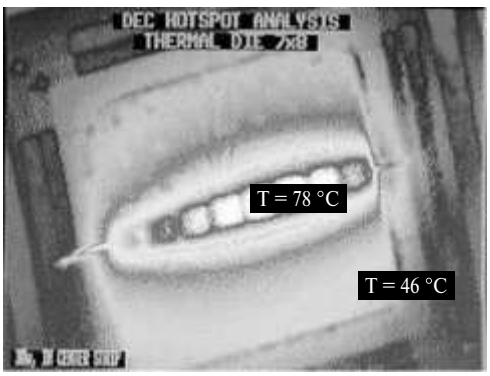


Рис. 10.17. Колебания температуры (моментальный снимок) в микропроцессоре DEC 21064. Наивысшую температуру имеет центральный формирователь тактовых импульсов [Herrick00]

С другой стороны, изменение параметров источника питания является основным источником дрожания в цепях разводки тактовых сигналов. Задержка прохождения через буферы сильно зависит от характеристик источника питания. В свою очередь, напряжение источника питания сильно зависит от активности переключения. Следовательно, для разных трактов задержка буфера будет иметь значительно отличающиеся значения. Изменения параметров источника питания можно разделить на низко- и высокочастотные. Статические изменения параметров источника питания могут вызываться фиксированными токами, идущими от различных модулей, а высокочастотные изменения порождаются мгновенными омическими падениями напряжения в электрической сети, вызванными флуктуациями активности переключения. Кроме того, серьезную проблему представляют индуктивные эффекты в цепях питания, поскольку они вызывают флуктуации напряжения. Дополнительно усугубляет эту проблему стробирование тактовых импульсов, поскольку логические переходы между холостым и активным состояниями может существенно изменить ток, поступающий от источника питания. Поскольку источник питания может быстро изменять свое состояние, период тактового сигнала модулируется потактово, что приводит к дрожанию. Дрожание в двух различных точках цепи разводки тактовых сигналов может коррелировать или не коррелировать, в зависимости от конфигурации сети питания и "рисунка" переключений. К сожалению, высокочастотные изменения в цепях питания очень сложно компенсировать (даже задействовав обратную связь). Следовательно, *шум в цепях питания ограничивает быстродействие цепей разводки тактовых сигналов*. Чтобы минимизировать колебания в цепях питания, вокруг формирователей тактовых импульсов в быстродействующих схемах вводится развязывающая емкость.

Емкостная связь (6 и 7)

Во временну*ю* неопределенность вклад вносят также изменения емкостной нагрузки. Существуют два основных источника колебаний емкостной нагрузки: связь между шинами тактовых импульсов и соседними сигнальными проводами, а также колебания емкости управляющего электрода. В цепь разводки тактовых импульсов входят емкость соединений и управляющих электродов защелок и регистров. Любая связь между проводником тактового сигнала и соседним проводником сигналов дает временну*ю* неопределенность. Поскольку соседний сигнал может совершать произвольные переходы в любое время, его связь с цепью тактовых импульсов изменяется на протяжении тактов, вызывая *дрожание*. Другим серьезным источником неопределенности тактовой частоты являются колебания емкости управляющих электродов присоединенных последовательных элементов. Емкость нагрузки описывается нелинейной функцией и зависит от приложенного напряжения. Для многих защелок и регистров нагрузка на тактовый сигнал является функцией текущего состояния защелки/регистра (т.е. значений, записанных во внутренних узлах схемы), а также следующего состояния. Это приводит к колебаниям задержки распространения через буферы тактовых импульсов в зависимости от такта, что также вызывает *дрожание*.

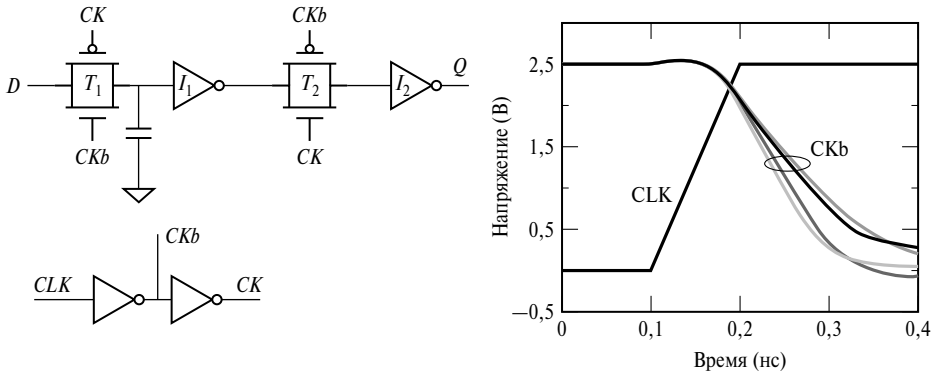


Рис. 10.18. Влияние нагрузки на тактовый сигнал, зависящей от данных, на дрожание тактового сигнала

Пример 10.2. Информационно-зависимое дрожание тактовых импульсов

Рассмотрим схему, показанную на рис. 10.18, где буфер тактовых сигналов минимального размера управляет регистром. (В действительности каждый буфер управляет четырьмя регистрами, но на рисунке показан только один из них.) С помощью моделирования получаем CKb — выход первого инвертора для четырех возможных переходов ($0 \rightarrow 0$, $0 \rightarrow 1$, $1 \rightarrow 0$ и $1 \rightarrow 1$). На рисунке видно, что дрожание тактового сигнала зависит от данных. Вообще, единственный способ решить эту проблему — использовать регистры, нагрузка которых не сильно изменяется в зависимости от данных, например дифференциальный регистр на усилителях считывания (см. главу 7).

10.3.3. Технологии разводки тактовых сигналов

Из сказанного выше понятно, что расфазировка и дрожание тактовых импульсов в цифровых схемах являются весьма серьезными проблемами, которые могут ограничивать быстродействие цифровой системы. Следовательно, разработчику необходимо так спроектировать цепь разводки тактовых импульсов, чтобы влияние обоих факторов было минимальным. Проектируя эту цепь, следует внимательно следить за соответствующих рассеянием мощности. В большинстве высокоскоростных цифровых процессоров основным источником рассеяния мощности является именно цепь разводки тактовых сигналов. Для снижения рассеяния требуется поддержка возможности выключения фрагментов цепи разводки тактовых импульсов. К сожалению, как описывалось ранее, подобная селекция дополнительно увеличивает неопределенность тактовой частоты.

В данном разделе мы рассмотрим основные принципы конструирования высокоэффективных цепей разводки тактовых сигналов, а также изучим распределение тактовых сигналов в микропроцессорах Alpha. В плане выбора

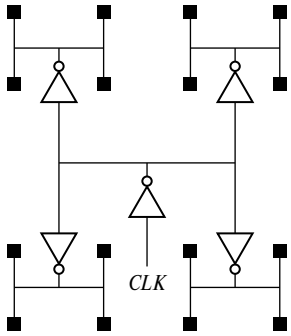


Рис. 10.19. Пример H-образного дерева — цепи распределения тактовых импульсов для 16 узлов-"листьев"

структуры цепи разводки тактовых импульсов разработчику доступно много степеней свободы, в том числе он может выбирать тип материалов проводников, топологию и иерархию схемы, относительные размеры проводов и буферов, время нарастания и спада, а также секционирование емкостей нагрузки.

Фундамент тактирования

Схемы разводки тактовых сигналов обычно включают цепь, которая применяется для доставки *глобального опорного сигнала* в различные части кристалла, и последнего каскада, отвечающего за *локальное распределение* тактовых сигналов с учетом локальных колебаний нагрузки. В большинстве схем распределения тактовых сигналов используется то, что абсолютная задержка распространения от центрального источника тактовых импульсов несущественна, значение имеет лишь разность фаз между двумя синхронизируемыми точками. Следовательно, одним из наиболее распространенных подходов к распределению тактовых сигналов является использование сбалансированных путей (*деревьев*).

Наиболее распространенным типом схемы распределения тактовых импульсов является *H-дерево*, показанное на рис. 10.19 для массива процессоров 4×4. В первую очередь тактовый сигнал подается в центр кристалла. Затем по сбалансированным путям, включающим согласованные соединения и буферы, в различные узлы-"листья" распространяется эталонный сигнал. В идеальном случае, если путь отлично сбалансирован, расфазировка тактовых импульсов равна нулю. Хотя на распространение сигнала от центральной точки к каждому конечному узлу может требоваться несколько тактов, на все узлы тактовые сигналы поступают одновременно. К сожалению, в реальной жизни из-за колебаний процесса и условий окружающей среды в схеме возникают расфазировка и дрожание.

Конфигурация "H-образное дерево" особенно полезна для схем, основанных на регулярных матрицах, где все элементы идентичны и разводку тактового сигнала можно сделать в виде двоичного дерева. Описанную концепцию можно обобщить, и тогда говорят о *согласованных RC-деревьях* — компо-

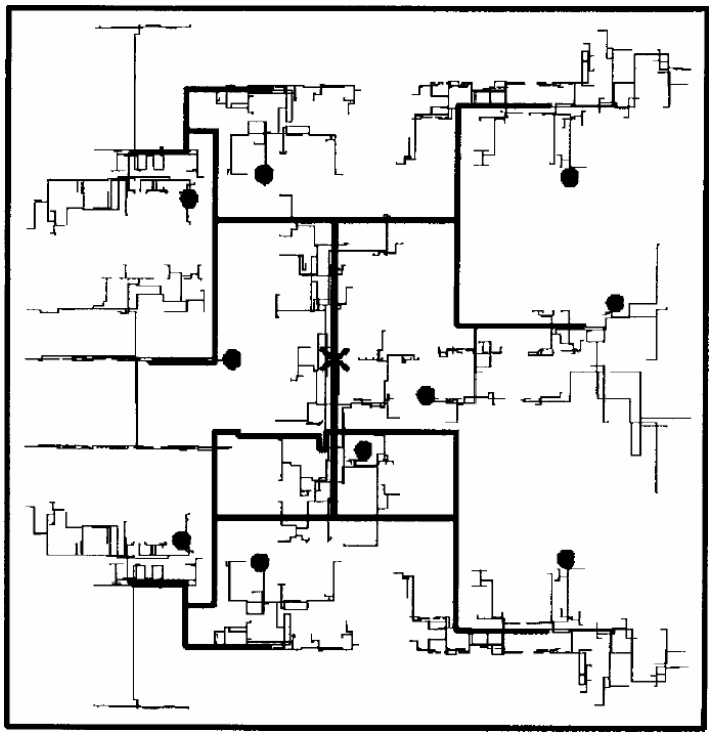
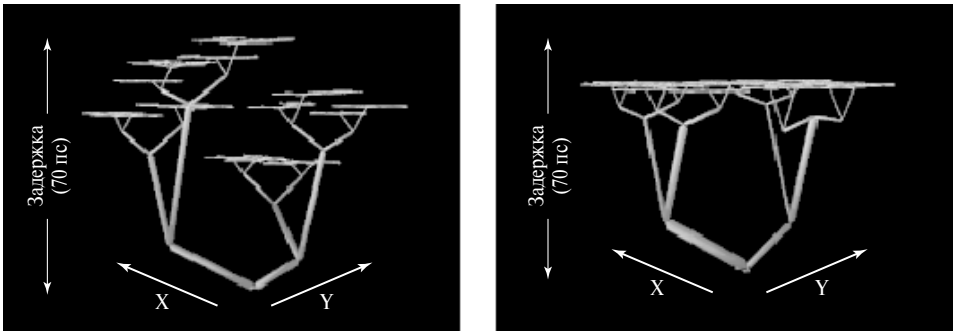


Рис. 10.20. Пример RC-согласованного распределения для микропроцессора IBM [Restle98]

новочном плане, позволяющем так распространить тактовые сигналы, чтобы межсоединения, по которым проходят тактовые сигналы к функциональным подблокам, имели одинаковую длину, т.е. общий подход не предполагает наличие правильной физической структуры. Пример согласованного RC-дерева показан на рис. 10.20. Кристалл разбит на 10 сбалансированных сегментов (*плиток*). Глобальный генератор тактовых импульсов передает тактовый сигнал всем генераторам плиток, расположенных в местах, обозначенных на рисунке точками. Низкоуровневое RC-согласованное дерево используется для питания 580 дополнительных генераторов внутри каждой плитки. На рис. 10.21 показана трехмерная визуализация задержки тактовых сигналов в древовидной сети.

Альтернативный подход к распределению тактовых сигналов представляет сетчатая структура, показанная на рис. 10.22 [Bailey00]. Сети обычно используются в последнем каскаде сети разводки тактовых импульсов для распределения тактовых сигналов синхронизируемым локальным нагрузкам. Данный подход в корне отличается от подхода с использованием сбалансированных RC-деревьев. Основное отличие заключается в том, что задержки оконечного усилителя, вызванные всеми нагрузками, не согласовываются. Вместо этого минимизируется абсолютная задержка в предположении, что размер сетки мал. Основным достоинством подобной сетчатой структуры яв-



а б

Рис. 10.21. Визуализация задержки тактовых сигналов в древовидной сети, питающей различные нагрузки. Оси X и Y представляют кристалл, а ось Z — задержку. Ширина линий пропорциональна расчетной ширине проводника. Несбалансированная нагрузка дает очень большую расфазировку, что видно на рис.10.21, *а*. С помощью аккуратного выбора ширины проводов балансируется нагрузка, благодаря чему минимизируется расфазировка, как показано на рис.10.21, *б* [Restle01]

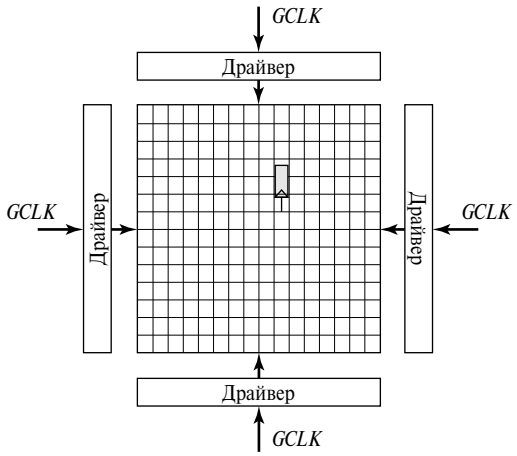


Рис. 10.22. Сетчатые структуры позволяют распределять тактовый сигнал с малой расфазировкой и создавать гибкие физические структуры за счет рассеяния мощности [Bailey00]

ляется то, что она позволяет вносить в проект изменения на более поздних стадиях разработки, поскольку тактовый сигнал легко провести в любую точку кристалла. К сожалению, платой за это является сравнительно большое рассеяние мощности, поскольку структура имеет много лишних межсоединений. В заключение отметим, что помимо схем, рассмотренных ранее, были разработаны и другие подходы к распределению тактовых сигналов (см., например, [Young97]).

Распределение тактового сигнала необходимо рассмотреть на начальных фазах разработки сложной схемы, поскольку это может повлиять на форму компоновочного плана кристалла. Разработчики любят игнорировать цепь

разводки тактовых сигналов на ранних фазах проекта и рассматривают ее только в конце цикла разработки, когда основная часть топологии кристалла уже зафиксирована. В результате получаются неразумные цепи разводки тактовых импульсов и множество временны*ы*х условий, которые снижают быстродействие конечной схемы. При внимательном планировании разработчик может избежать многих подобных проблем, и распространение тактовых сигналов становится вполне управляемой операцией.

Примеры распределения тактовых сигналов — цифровые микропроцессоры Alpha

В данном разделе мы подробно рассмотрим стратегии распространения тактовых сигналов для трех поколений микропроцессоров Alpha. Данные процессоры всегда находились на переднем крае технологии, а следовательно, позволяют проследить эволюцию концепции распределения тактовых сигналов [Herrick00].

Процессор Alpha 21064. Микропроцессор Alpha первого поколения (21064 или EV4) производства компании Digital Equipment Corporation, в котором использован один глобальный генератор тактовых импульсов [Dobberpuhl92]. Распределение емкости нагрузки по различным функциональным блокам показано на рис. 10.23. Суммарная емкость нагрузки на тактовый сигнал составляет 3,25 нФ! В процессоре использована однофазная методология тактирования, и тактовый сигнал с частотой 200 МГц подается на двоичное дерево с пятью уровнями буферизации. Входы генераторов тактовых импульсов закорачиваются с целью сглаживания асимметрии входных сигналов. Последний выходной каскад, располагающийся в середине кристалла, питает цепь тактовых импульсов. Генератор тактовых импульсов и соответствующие предварительные генераторы учитывают 40% эффективной переключаемой емкости (12,5 нФ), дающей значительное рассеяние мощности. Общая ширина генератора тактовых импульсов при 0,75-микронной технологии составляла порядка 35 см. Подробное моделирование расфазировки тактовых импульсов с учетом колебаний процесса показывает, что неопределенность тактовой частоты не превышает 200 пс (меньше 10%).

Процессор Alpha 21164. Микропроцессор Alpha второго поколения (EV5) работает с тактовой частотой 300 МГц и использует 9,3 миллиона транзисторов на кристалле $16,5 \times 18,1$ мм (0,5-микронная КМОП-технология) [Bowhill95]. В процессоре снова задействована однофазная методология тактирования, в проекте широко используется динамическая логика, что дало существенную нагрузку на тактовый сигнал — 3,75 нФ. Система распределения тактового сигнала потребляет 20 Вт, что составляет 40% от общей мощности, потребляемой процессором.

Вначале входной тактовый сигнал направляется через один шестикаскадный буфер, расположенный в центре кристалла. Получающийся в результате сигнал распределяется по третьему металлическому слою в левый и правый блоки конечных генераторов тактовых сигналов, расположенных между кэшем второго уровня и внешней стороной операционного блока (рис. 10.24, *а*).

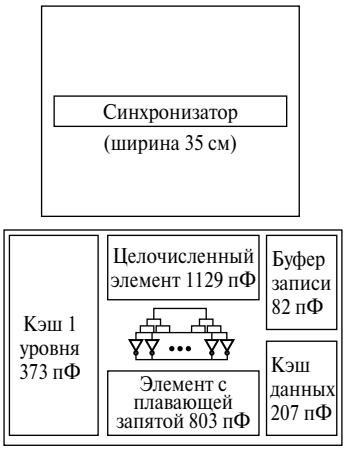
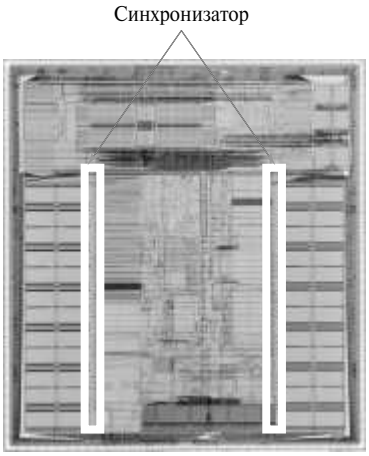


Рис. 10.23. Распределение емкости нагрузки на тактовый сигнал для процессора Alpha 21064

Сгенерированный тактовый сигнал подается на сетку металлических проводов третьего и четвертого слоев. Ширина эквивалентного транзистора инвертора последнего генератор равна 58 см! Чтобы обеспечить целостность сетки разводки тактовых импульсов, эта сетка была отделена от топологии и смоделирована получающаяся RC-цепь. Трехмерное представление результатов моделирования показано на рис. 10.24, *б*. Как видно на графике, на выходе левого и правого генераторов расфазировка равна нулю. Максимальное значение абсолютного сдвига фаз не превышает 90 пс. Все критичные операционные блоки получают тактовый сигнал в пределах 65 пс.

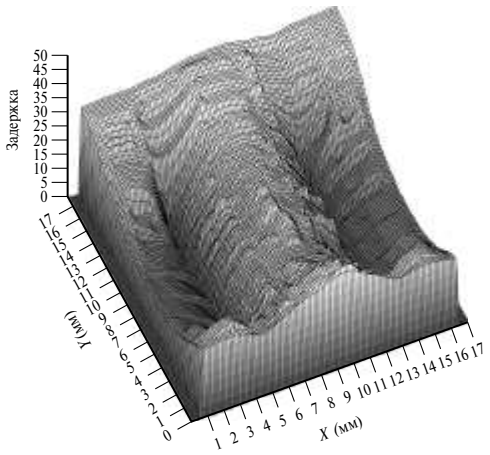
Проблемы расфазировки тактовых сигналов и состязания были решены с помощью подхода "смешивания и согласования". Для снятия проблемы расфазировки необходимо либо трассировать тактовые сигналы в направлении, противоположном потоку данных (платой за это является небольшое снижение быстродействия), либо гарантировать, что данные не смогут обогнать тактовый сигнал. Для реализации процессора использовалась стандартизованная библиотека уровневых защелок на передаточных вентилях. Чтобы снять проблему состязаний, разработчики следовали таким директивам.

- С помощью аккуратного масштабирования буферов локальных тактовых сигналов была минимизирована их разность фаз.
- По крайней мере один логический элемент должен быть вставлен между соединяющими защелками. Этот элемент, который может быть частью логической функции или просто инвертором, гарантирует минимальную задержку распространения. Для проверки выполнения данного правила для всего кристалла были разработаны специальные инструменты верификации проектов.



Микрофотография кристалла с указанными положениями синхронизаторов

а



Моделирование рассинхронизации

б

Рис. 10.24. Распределение тактового сигнала и расфазировка для 300 МГц микропроцессора. (*Перепечатано с разрешения Digital Equipment Corporation.*)

Чтобы повысить однородность межслойного диэлектрика, между широко разнесенными линиями были введены присадочные многоугольники, как показано на рис. 10.25. Хотя это может увеличить емкость с соседними сигнальными шинами, бо́льшая однородность снижает колебания параметров и расфазировку. Ложное заполнение вводится автоматические и подключается к одной из шин питания (V_{DD} или GND). В настоящее время данная техника используется во многих современных процессах для регулировки расфазировки.

Процессор Alpha 21264. В 600-мегагерцовом процессоре Alpha 21264 (EV6) (0,35-микронная КМОП-технология) использовалась иерархическая схема тактирования, показанная на рис. 10.26 [Bailey98]. Выбор такого подхода стал одним из основных отличий данного поколения процессоров от предыдущих, в которых не существовало иерархии тактовых сигналов, помимо выделения глобальной сети тактовых сигналов. Использование иерархического подхода позволяет играть на компромиссах между потребляемой мощностью и расфазровкой. Мощность можно уменьшить благодаря избирательному отключению цепей разводки тактовых импульсов для отдельных блоков. Изучая микропроцессоры предыдущего поколения, видим, что мощность, уходящая на генерацию и разводку тактовых сигналов, составляет бо́льшую часть общей потребляемой мощности. Кроме того, гибкость, вызванная наличием локальных тактовых сигналов, дает разработчику больше свободы в плане выбора стилей проектирования схем на модульном уровне. К недостаткам использования иерархических тактовых сигналов можно отнести более сложный контроль над расфазировкой. Тактовые сигналы для различных локальных регистров могут идти по совершенно различным путям, что может ска-

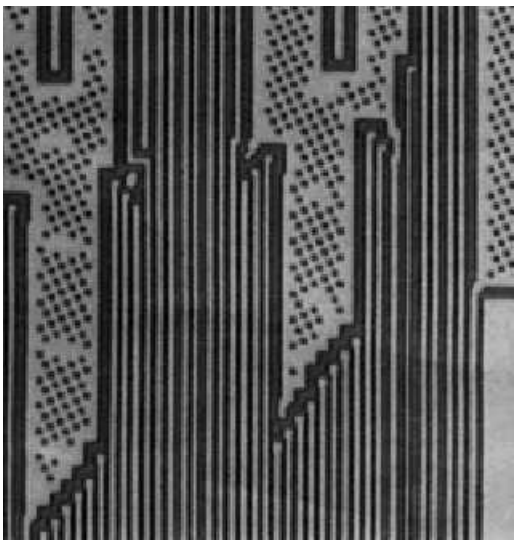


Рис. 10.25. Ложное заполнение уменьшает колебания толщины межслойного диэлектрика и расфазировку

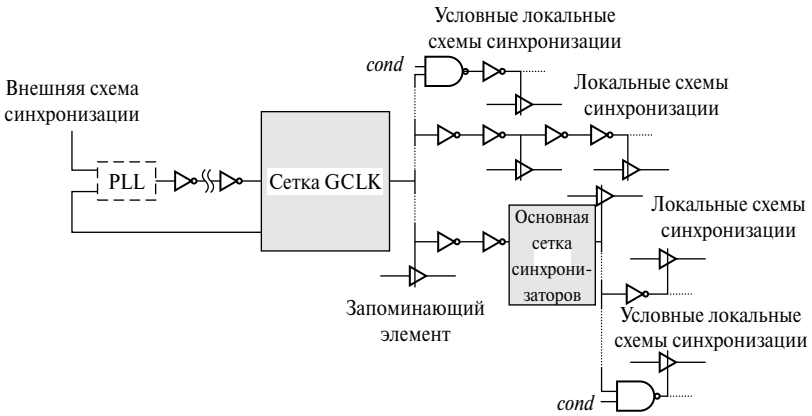
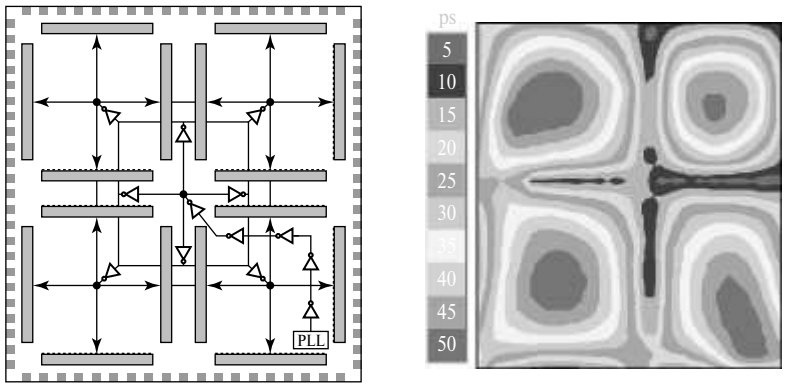


Рис. 10.26. Иерархия тактовых сигналов для процессора Alpha 21264

заться на разности фаз между этими сигналами. Впрочем, используя средства временно́й верификации, расфазировку можно загнать в приемлемые границы с помощью настройки генераторов тактовых сигналов.

Иерархическая структура тактовых сигналов состоит из *глобальной сети тактовых сигналов* ($GCLK$), охватывающей весь кристалл. В $GCLK$ имеются элементы, запоминающие состояния, и точки тактирования. Тактовый сигнал, генерируемый внутри кристалла, подается в центр кристалла и оттуда с помощью древовидной структуры разводится 16 распределенным генераторам тактовых сигналов (рис. 10.27). В глобальной цепи разводки тактовый



а б

Рис. 10.27. Глобальная цепь разводки тактовых импульсов, построенная по принципу секционирования: *а* — структура секции; *б* — фазовый сдвиг тактовых импульсов в кристалле

сигнал делится на четыре области, что уменьшает расстояние от генераторов до нагрузки, а следовательно, снижает расфазировку. Каждая такая секция питается с четырех сторон, что уменьшает зависимость от колебания параметров процесса. Кроме того, это помогает решить проблемы электропитания и колебаний температуры, поскольку генераторы разносятся в разные точки кристалла.

Самый низкий уровень иерархии формируется локальными тактовыми сигналами, которые по мере необходимости генерируются из любого другого тактового сигнала. Обычно их можно настроить согласно локальным временны*ы*м условиям. Локальные тактовые сигналы предлагают хорошую гибкость в плане проектирования логических блоков, но в то же время существенно затрудняют контроль над расфазировкой. Более того, локальные тактовые сигналы более чувствительны к связи с шинами данных, поскольку они не экранируются так, как сигналы глобальной сетки. В результате распределение локальных сигналов сильно зависит от локальных межсоединений, а следовательно, проектировать их нужно очень аккуратно.

Технологии проектирования: борьба с расфазировкой и дрожанием

Из сказанного выше можно сделать несколько полезных выводов по минимизации расфазировки и дрожания.

1. Чтобы минимизировать расфазировку, необходимо *сбалансировать пути распространения тактовых сигналов* от центрального источника к отдельным синхронизируемым элементам, используя H-деревья или более общие древовидные структуры. При использовании протрассированных деревьев разводки тактовых сигналов необходимо выравнять эффективные нагрузки на тактовый сигнал всех путей.

2. Использование *локальных сеток разводки тактовых сигналов* (вместо протрассированных деревьев) может уменьшить расфазировку за счет увеличения емкостной нагрузки и рассеяния мощности.
3. Если информационно-зависимые колебания нагрузки на тактовый сигнал вызывают существенное дрожание, следует использовать *дифференциальные регистры*, имеющие независящую от данных нагрузку на тактовый сигнал. Использование тактовых сигналов с избирательным отключением приводит к тому, что нагрузка на тактовый сигнал зависит от передаваемых данных, а дрожание в цепи увеличивается. В цепях разводки тактовых сигналов с большой постоянной нагрузкой (например, в тактовых сетках) колебания, зависящие от данных, могут быть незначительными.
4. Если поток данных идет в одном направлении, то *данные и тактовые сигналы необходимо пускать в разных направлениях*. Таким образом, за счет скорости работы можно устранить состязания.
5. Чтобы избежать шума, зависящего от данных, *провода тактовых сигналов необходимо экранировать* от соседних сигнальных проводов. Располагая шины питания (V_{DD} или GND) рядом с проводами тактовых сигналов, можно минимизировать или совершенно устранить их связь с соседними сетями передачи сигналов.
6. Колебания емкости межсоединений, вызванные колебаниями толщины межслойного диэлектрика, можно существенно уменьшить, использовав *ложные заполнения* (dummy fills). Ложные заполнения широко используются для снижения расфазировки путем повышения однородности межслойного диэлектрика. Систематические вариации необходимо смоделировать на этапе проектирования схемы и каким-либо образом компенсировать.
7. Наличие разности температур для разных точек кристалла вызывает колебания задержки буферов тактовых сигналов. Используя *цепи обратной связи на основе систем автоматической подстройки по задержке* (подробнее об этом — ниже), колебания температуры можно компенсировать.
8. Существенный вклад в дрожание вносят колебания параметров цепи питания, поскольку они влияют на изменение задержки прохождения сигнала через буферы тактовых импульсов. Высокочастотные колебания параметров источника питания можно уменьшить, добавив *внутрикристальные развязывающие конденсаторы*. К сожалению, развязывающие конденсаторы требуют большой площади, для уменьшения которой необходимы эффективные технологии корпусирования.

---

*10.3.4. Тактирование на основе защелок**

Использование регистров в последовательной схеме является устойчивой и надежной методологией проектирования. Разделяя логические блоки про-

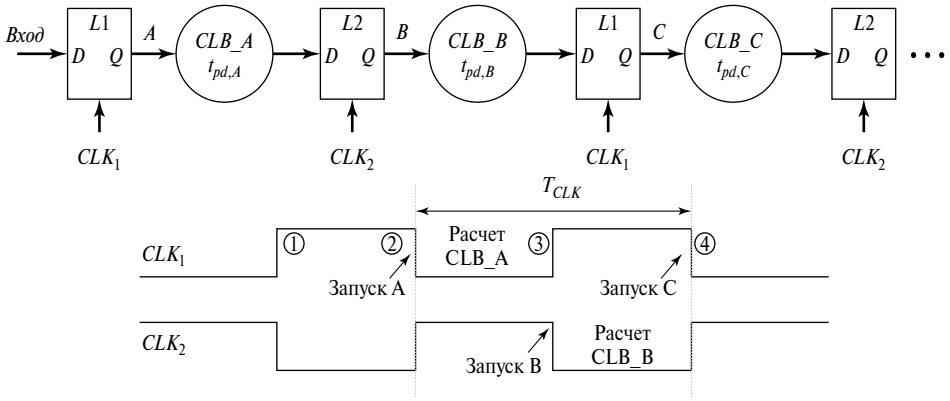


Рис. 10.28. Структура, основанная на защелках, в которой прозрачные защелки разделяются блоками комбинационной логики

зрачными защелками, можно добиться существенного повышения быстродействия схемы. В системе, управляемой фронтом, минимальная продолжительность такта для всей системы определяется наихудшим логическим путем между двумя регистрами. Если логический блок завершает свою работу до конца такта, он должен ожидать начала следующего фронта тактового сигнала. Используя методологию, основанную на защелках, можно получить более гибкую схему тактирования, разрешив одним каскадам одалживать время у других каскадов или "дарить" им избыточное время. Дополнительная гибкость также повышает общее быстродействие. Обратите внимание на то, что методологию, основанную на защелках, можно рассматривать как добавление логических блоков между защелками триггеров "master–slave".

Рассмотрим изображенную на рис. 10.28 систему на основе защелок, в которой использована двухфазная схема тактирования. Предположим для простоты, что оба тактовых сигнала идеальны и что один является инвертированной версией второго. В данной конфигурации на вход блока комбинационной логики A стабильно (на спаде сигнала CLK_1; фронт ②) подается сигнал CLB_A. Максимальное время оценки этого сигнала равно $T_{CLK}/2$. На спаде сигнала CLK_2 (фронт ③) выход CLB_A фиксируется и начинается вычисление CLK_B. Сигнал CLB_B вычисляется в низкой фазе сигнала CLK_2, и выход защелки доступен на спаде сигнала CLK_1 (фронт ④). С временно́й точки зрения данный сценарий выглядит эквивалентным системе, управляемой фронтом, где CLB_A и CLB_B каскадно передаются между двумя регистрами, активизируемыми фронтом (рис. 10.29). В обоих случаях кажется, что время, доступное для получения CLB_A и CLB_B, равно T_{CLK}.

Тем не менее существует важное отличие двух приведенных схем с точки зрения быстродействия. В системе, основанной на защелках, логический блок может использовать время, оставшееся от предыдущего логического блока. Данное явление, возможное благодаря прозрачности защелки в период ее включения, называется *резервным одалживанием* (slack borrowing) [Bernstein98]. Оно не требует никаких явных изменений проекта, поскольку передача резерва от одного блока к другому происходит автоматически.

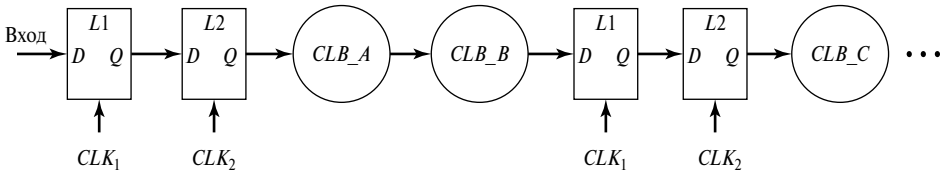


Рис. 10.29. Конвейер, управляемый фронтом (каскадное включение защелок для регистров, управляемых фронтом) логической схемы, приведенной на рис. 10.28

Ключевым преимуществом резервного одалживания является то, что оно позволяет логическим блокам иногда использовать больше времени, чем разрешает длительность такта, если при этом удовлетворяются общие условия на продолжительность такта. Другими словами, если последовательная система работает с определенной тактовой частотой и общая задержка логики для всего такта больше продолжительности такта, то неиспользуемое время (*резерв*) неявно одалживается у предыдущих каскадов. Формально передача резерва происходит, если $T_{CLK} < t_{pd,A} + t_{pd,B}$, и логика функционирует правильно (для простоты мы игнорируем задержку, связанную с защелками). Из сказанного следует, что тактовая частота может быть выше, чем позволяет критический путь в наихудшем случае!

Как отмечалось ранее, передача резерва возможна благодаря уровневой природе защелок. Вернемся к рис. 10.28. Вход CLB_A должен иметь действительное значение на спаде сигнала CLK_1 (фронт ②). Что произойдет, если блок комбинационной логики предыдущего каскада рано завершит свою операцию и сможет предоставить действительные входные данные для CLB_A до фронта ②? Поскольку защелка прозрачна в течение всей высокой фазы тактового сигнала, входные данные для CLB_A действительны сразу же, как только предыдущий каскад завершил свою операцию. При этом подразумевается, что максимальное время, доступное для сигнала CLB_A, — это его "собственное время" (т.е. время, соответствующее низкой фазе CLK_1) плюс все время, оставшееся от предыдущих вычислений.

Рассмотрим показанную на рис. 10.30 систему на основе защелок. В данном примере сигнал a (вход CLB_A) действителен до фронта ②. Это подразумевает, что предыдущий блок не полностью использовал свою долю времени, создав резерв, обозначенный затененной областью. По построению CLB_A может начинать вычисляться, как только сигнал станет действительным. Благодаря одолженному времени вычисления завершаются намного раньше предполагаемого срока (фронт ③). Поскольку L_2 — это прозрачная защелка, имеющая действительное значение в высокой фазе сигнала CLK_2, то CLB_B начинает вычисляться с использованием резерва, предоставленного CLB_A. Как и ранее, вычисление CLB_B завершается до предполагаемого срока (фронт ④), и выигранное резервное время передается в следующий такт. Как видно на рисунке, общая задержка такта (т.е. сумма задержек для CLB_A и CLB_B) больше продолжительности такта. Однако, поскольку конвейер работает правильно, происходит передача резерва и достигается более высокая пропускная способность.

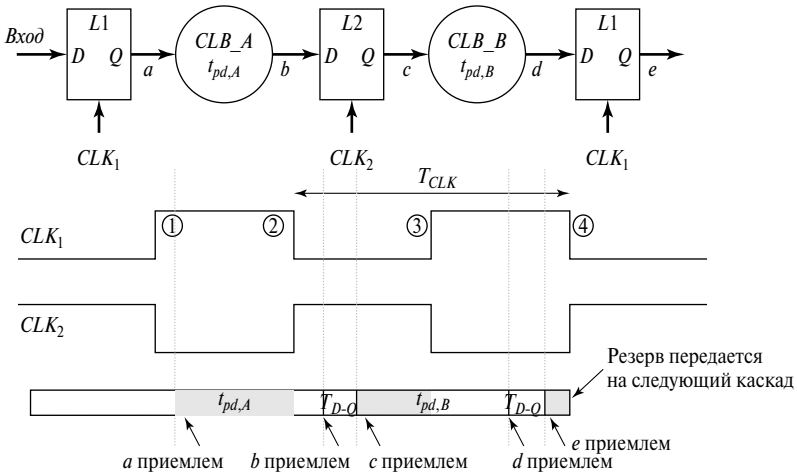


Рис. 10.30. Пример одалживания резерва

С одалживанием резерва связан один важным вопрос: какой максимальный резерв можно передавать между тактами. На рис. 10.30 видно, что сигнал CLB_A может начать вычисляться не раньше момента времени ①. Это происходит в том случае, если предыдущий логический блок вообще не использовал выделенное ему время (высокая фаза CLK_1) или если он завершил вычисления за счет резерва, переданного с предыдущих каскадов. Следовательно, максимальное время, которое можно одолжить у предыдущего каскада, равно половине такта, или $T_{CLK}/2$. Аналогично, CLB_B должен завершить свою операцию до фронта ④. Отсюда следует, что максимальная задержка логического блока равна $1,5 \times T_{CLK}$. Однако обратите внимание на то, что для n-каскадного конвейера общая задержка логических блоков не может превышать время $n \times T_{CLK}$.

Пример 10.3. Пример передачи резерва времени

Прежде всего рассмотрим изображенный на рис. 10.31 конвейер, управляемый отрицательным фронтом. Предположим, что первичный вход In действителен почти сразу же после нарастания тактового сигнала. Можно показать, что минимальная продолжительность такта равна 125 нс. Время ожида-

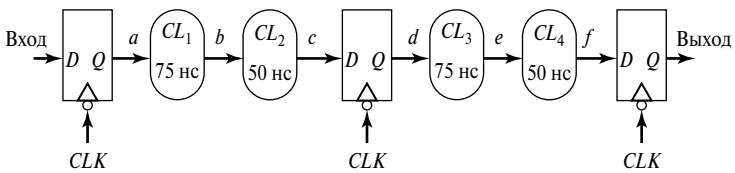


Рис. 10.31. Последовательный конвейер, управляемый фронтом

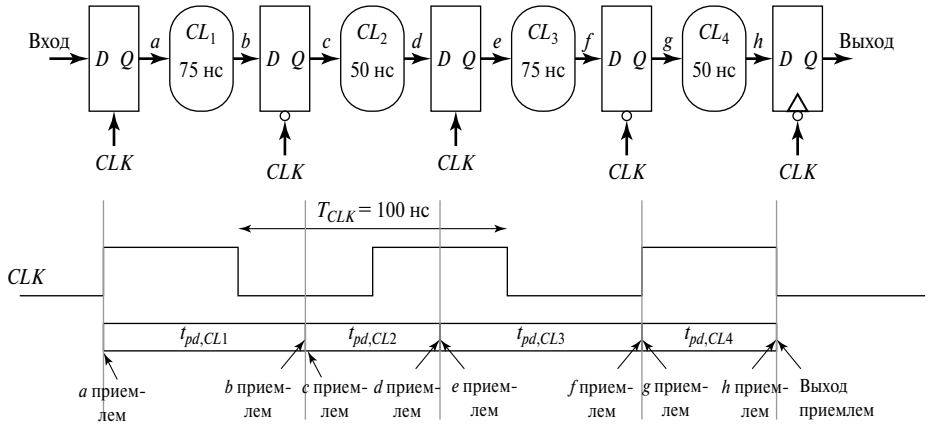


Рис. 10.32. Конвейер на основе защелок

ния составляет два такта. (На самом деле выход действителен 2,5 такта после установки входа.) Обратите внимание на то, что для первого каскада конвейера половина такта — холостое состояние, поскольку входные данные доступны для CL_1 после спада тактового сигнала. В системе, основанной на защелках, это время может использоваться для выполнения каких-либо операций.

На рис. 10.32 показана та же схема, но реализованная на защелках. Как видно на временной диаграмме, аналогичной работы можно добиться и при продолжительности такта 100 нс. Это стало возможным благодаря одалживанию резерва между логическими блоками.

Если вас заинтересовала тема одалживания резерва, обратитесь к [Bernstein98], где представлен великолепный количественный анализ данного привлекательного (хотя и достаточно сложного) подхода.

10.4. Проектирование схем с внутренней синхронизацией*

10.4.1. Логические схемы с внутренней синхронизацией — асинхронный подход

При рассмотренном в предыдущих разделах синхронном подходе к синхронизации предполагается, что ритм всех событий цепи задает центральный генератор тактовых импульсов. В целом подобные генераторы выполняют две функции.

- Они гарантируют, что удовлетворяются *физические условия относительно тактирования*. Следующий такт может начаться только тогда, когда завершились все логические переходы и система пришла в установивше-

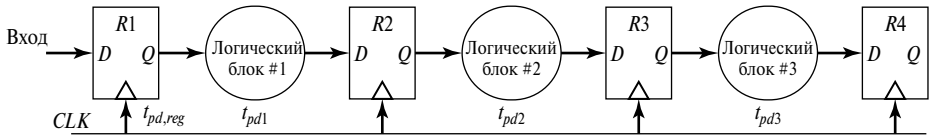


Рис. 10.33. Конвейерный синхронный информационный канал

еся состояние. Таким образом, в следующем цикле вычислений будут задействованы только правильные логические значения. В целом тактовые сигналы учитывают наихудшие значения задержек логических вентилей, последовательных логических элементов и проводки.

- Тактирование используется как *механизм логического упорядочения* глобальных системных событий. Генератор тактовых импульсов предоставляет временну*ю* ось, позволяющую определить, что и где происходит. При каждом переходе тактового сигнала инициируется несколько операций, которые меняют состояние последовательной цепи.

Изучим конвейерный информационный канал, показанный на рис. 10.33. В этой схеме данные проходят через логические каскады под управлением тактовых сигналов. Здесь важно отметить, что продолжительность такта выбрана большей наибольшей задержки всех каскадов конвейера, или $T > \max(t_{pd1}, t_{pd2}, t_{pd3}) + t_{pd,reg}$. Таким образом, мы удовлетворяем физическому условию. При каждом переходе тактового сигнала выбирается новый набор входных сигналов и начинается новый цикл вычислений. Пропускная способность системы (которая равна числу выборок данных, обработанных за секунду) эквивалентна тактовой частоте. Момент времени, в который начинается выборка нового входа, и доступность выхода зависят от *логического упорядочения* системных событий и, естественно, в данном примере управляются тактовыми сигналами.

Синхронная методология проектирования имеет ряд очевидных преимуществ. Она представляет структурированный детерминистический подход к проблеме управления множеством событий, происходящих в цифровых структурах. Подход является надежным и легко поддающимся модификации, что объясняет его невероятную популярность. Тем не менее, он имеет и ряд недостатков.

- Предполагается, что все тактовые события происходят одновременно во всей схеме. В действительности это не так из-за расфазировки и дрожания.
- Поскольку все тактовые сигналы схемы одновременно совершают переходы, в очень короткий промежуток времени создается существенный ток (из-за большой емкостной нагрузки). В свою очередь, из-за наличия индуктивности корпуса и сопротивления сети питания это приводит к появлению больших проблем с шумом.
- Очевидно, одновременное наложение физических и логических ограничений оказывает существенное влияние на быстродействие. Например, про-

пускная способность системы, приведенной на рис. 10.33, непосредственно связана с наихудшей задержкой наиболее медленного элемента конвейера. В среднем задержка любого каскада конвейера меньше этого значения. Тот же конвейер может поддерживать среднюю пропускную способность, которая значительно выше, чем у синхронной реализации. Например, задержка распространения 16-битового сумматора сильно зависит от проходящих через него данных: сложение двух 4-битовых чисел требует гораздо меньшего времени, чем сложение двух 16-битовых величин.

Чтобы избежать данных проблем, можно принять *асинхронный* подход и вообще отказаться от тактовых сигналов. Проектирование чисто асинхронной схемы представляет собой нетривиальную и потенциально опасную задачу, требующую тщательного временн*о*го анализа цепи. Фактически логическое упорядочение событий диктуется структурой сети транзисторов и относительными задержками сигналов. Наложение временн*ы*х ограничений с помощью манипуляций с логической структурой и длинами путей прохождения сигналов требует интенсивного использования средств автоматизированного проектирования и рекомендуется только в тех случаях, когда без него нельзя обойтись.

Более надежной технологией является использование внутренней синхронизации (self-timing), которая представляет локальное решение проблемы тактирования [Seitz80]. На рис. 10.34 данный подход иллюстрируется на примере конвейерного информационного канала. Предполагается, что каждая комбинационная функция может сообщить, что она завершила вычисление для определенного блока данных. Обработка логического блока инициируется сигналом *Start* (*Старт*). Блок комбинационной логики обрабатывает входные данные и в зависимости от них (с учетом физических условий) генерирует метку *Done* (*Готово*) сразу после завершения вычислений. Кроме того, операторы должны сообщить один другому, что либо они готовы получать следующее входное слово, либо они предоставляют на выходе слово, готовое к дальнейшим вычислениям. Таким образом обеспечивается логическое упорядочение событий. Дополнительно могут использоваться сигналы *Ack* (*Подтверждение*) и *Req* (*Запрос*). Если использовать в качестве примера конвейерный информационный канал, последовательность действий можно представить следующим образом.

1. Прибывает входное слово, и в блок $F1$ отправляется запрос *Req*. Если в это время блок $F1$ не активизирован, он передает данные и подтверждает этот факт входному буферу, который после этого может извлекать следующее слово.
2. Блок $F1$ включается с помощью сигнала *Start*. Через некоторое время (которое зависит от данных и режима работы) сигнал *Done* выполняет переход в высокое состояние, сообщая о завершении расчетов.
3. Для модуля $F2$ генерируется сигнал *Req*. Если данная функция свободна, сигнал *Ack* переходит в высокое состояние, передается выходное значение и блок $F1$ может начинать следующий цикл вычислений.

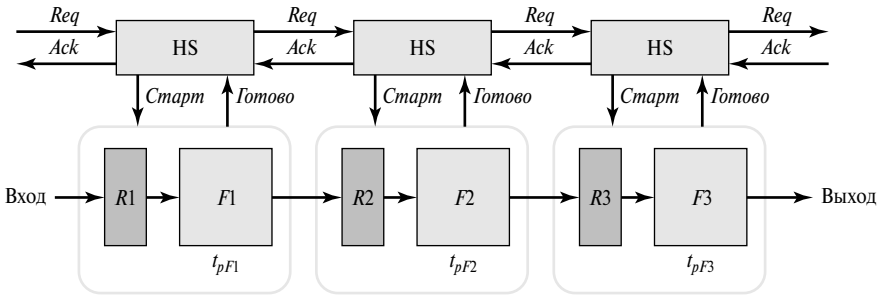


Рис. 10.34. Конвейерный информационный канал с внутренней синхронизацией

Внутренняя синхронизация позволяет эффективно разделить физическое и логическое упорядочение действий. Сигнал завершения *Done* гарантирует, что физические временны*е* условия будут удовлетворяться, а перед получением нового входа схема будет находиться в установившемся состоянии. Логическое упорядочение операций достигается с помощью схемы "подтверждение"–"запрос", часто называемой *протоколом установления связи*. Описанный выше протокол является лишь одним из множества возможных, выбор протокола, используемого в конкретной ситуации, существенно влияет на быстродействие и устойчивость схемы.

По сравнению с синхронным подходом внутренняя синхронизация имеет несколько привлекательных особенностей.

- В отличие от синхронного подхода с глобальной централизацией, синхросигналы генерируются *локально*, что решает все проблемы и снимает все служебные издержки, связанные с распределением высокоскоростных тактовых сигналов.
- Разделение физического и логического упорядочения может потенциально повысить быстродействие. В синхронных системах длительность такта должна увеличиваться так, как определяет наиболее медленный сигнал. В системах с внутренней синхронизацией слово данных не должно ждать начала очередного такта, чтобы перейти на следующие этапы обработки.
- Автоматическое выключение неиспользуемых блоков позволяет снижать энергопотребление. Кроме того, можно уменьшить потребление мощности, связанное с генерацией и распределение высокоскоростных тактовых сигналов. Как обсуждалось ранее, эти служебные издержки могут быть значительными.
- По своей природе схемы с внутренней синхронизацией являются устойчивыми вне зависимости от колебаний процесса производства и условий эксплуатации (например, температуры). Синхронные системы имеют минимальное быстродействие в крайних точках диапазона рабочих условий, а быстродействие систем с внутренней синхронизацией определяется фактическим техническим состоянием.

Таблица 10.1. Избыточное представление сигналов, включающее информацию о переходе

B	$B0$	$B1$
Состояние перехода (или сброса)	0	0
0	0	1
1	1	0
Недопустимое состояние	1	1

К сожалению, платой за данные полезные свойства являются существенные служебные издержки на уровне схем, объясняющиеся необходимостью генерирования сигналов завершения и потребностью в логике установления связи, которая будет на локальном уровне упорядочиваться события схемы (блок HS на рис. 10.34). Обе эти темы подробно рассмотрены ниже.

10.4.2. Генерация сигнала завершения

Необходимым компонентом логики с внутренней синхронизацией является схема, сообщающая, когда определенный участок цепи завершил свои вычисления текущего блока данных. Сгенерировать сигнал завершения можно двумя распространенными и надежными способами.

Двухшинное кодирование

Одним из популярных подходов к генерации сигнала завершения является использование *двухшинного кодирования* (dual-rail coding). В действительности такое кодирование требует введения в представление данных избыточности, которая позволит сообщить, что определенный бит находится в состоянии перехода или в стационарном состоянии. Рассмотрим избыточную модели данных, представленную в табл. 10.1. Для представления одного бита данных B применяются два бита ($B0$ и $B1$). Чтобы данные были приемлемыми или вычисления могли завершиться, схема должны находиться в состоянии 0 ($B0 = 0$, $B1 = 1$) или 1 ($B0 = 1$, $B1 = 0$). Условие ($B0 = 0$, $B1 = 0$) свидетельствует о том, что данные неверны и схема находится либо в состоянии сброса, либо в состоянии перехода. Состояние ($B0 = 1$, $B1 = 1$) является недопустимым и в реальной схеме не должно встречаться никогда.

На рис. 10.35 показана схема, на практике реализующая избыточное представление, — динамическая версия DCVSL-логики, в которой тактовый сигнал заменен сигналом *Start* [Heller84]. Когда сигнал *Start* находится в низком состоянии, схема предварительно заряжается p-МОП-транзисторами, и выход ($B0, B1$) переходит в состояние *Reset-Transition* $(0, 0)$. При переходе сигнала *Start* в высокое состояние (это сообщает о начале вычислений), начинает работу n-МОП-транзисторная цепь PDN, понижающая уровень одного из предварительно заряженных узлов. Один из узлов, $B0$ или $B1$ (но не оба одновременно), переходят в высокое состояние, что приводит к переходу в высокое состояние сигнала *Done* и сообщает о завершении вычислений.

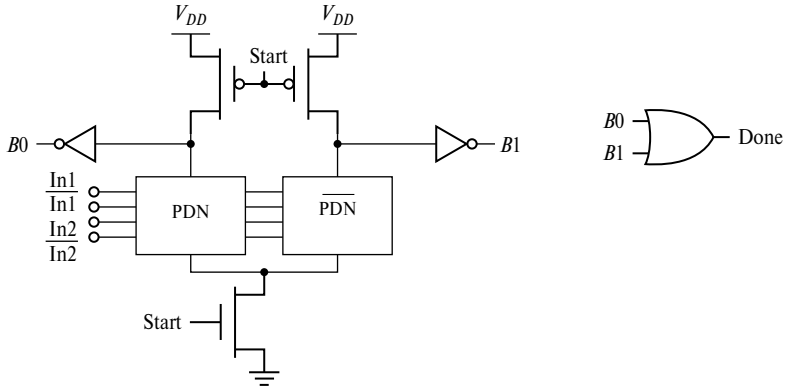


Рис. 10.35. Генерация сигнала завершения в DCVSL

Можно придумать и другие избыточные представления, помимо указанного в табл. 10.1. Единственным необходимым элементом является наличие *переходного состояния*, обозначающего, что схема находится в режиме оценки и выходные данные еще не доступны.

Пример 10.4. Сумматор с внутренней синхронизацией

На рис. 10.36 показана эффективная реализация сумматора с внутренней синхронизацией [Abnous93]. Для повышения быстродействия схемы использована дифференциальная схема переноса. Предлагаемый подход основан на

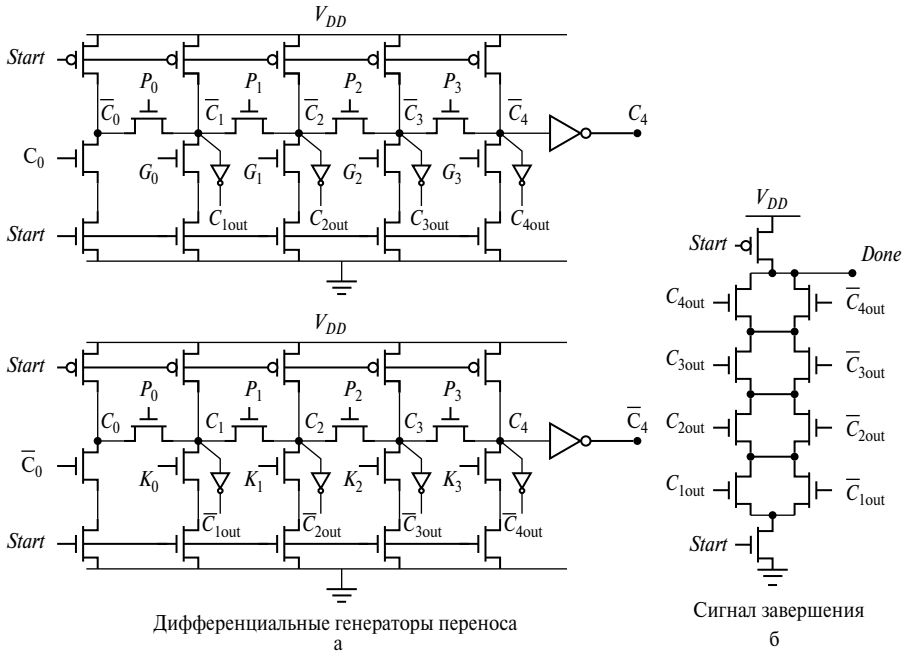


Рис. 10.36. Схема с дифференциальным представлением сигналов

наблюдении, что наибольший вклад в задержку сумматора дает путь передачи сигнала переноса. Следовательно, дифференциальную передачу сигналов достаточно использовать только на пути передачи сигнала переноса (см. рис. 10.36, *а*). Для получения сигнала завершения можно эффективно использовать сигналы переноса различных каскадов (см. рис. 10.36, *б*). При этом вполне обоснованно можно предположить, что генерация суммы, которая зависит от прибытия сигнала переноса, выполняется быстрее, чем генерация сигнала завершения. Преимуществом предложенного подхода является то, что генерация сигнала завершения начинается рано и выполняется параллельно генерации суммы, что уменьшает критический временно́й путь. Все другие сигналы, например, P (распространение), G (генерация), K (прекращение), и S (сумма), не требуют генерации сигнала завершения, поэтому их можно реализовать с помощью однотактных схем. Единственным отличием дифференциальной схемы от обычной является замена сигнала G сигналом K. Анализируя предложенные схемы, видим, что этого достаточно для инверсии сигнала переноса, а следовательно, для получения дифференциальной схемы.

Задержка копии

Описанное выше двухшинное кодирование позволяет отслеживать статистику по сигналу за счет повышенного рассеяния мощности. Любой логический элемент должен выполнять переход для каждого нового входного вектора независимо от значения последнего. Чтобы снизить служебные издержки по детектированию завершения, можно, как показано на рис. 10.37, использовать *копию критического пути*, настроенную как элемент задержки. Чтобы начать вычисления, сигнал *Start* переводится на высокий уровень и инициируются вычисления логической схемы. В то же время сигнал запуска подается в копию линии задержки, которая отслеживает критический путь логической схемы. Здесь важно так настроить копию, чтобы в ней были невозможны ложные импульсные переходы. Когда выход линии задержки совершает переход, это свидетельствует о том, что логика завершила свою работу, поскольку линия задержки имитирует критический путь.

Рис. 10.37. Генерация сигнала завершения с использованием модуля задержки

Преимущество данного подхода заключается в том, что логику можно реализовать с использованием стандартного неизбыточного подхода (например, на КМОП-схемах). Кроме того, если расчеты параллельно выполняют несколько логических блоков, служебные издержки, связанные с задержкой, делятся на эти блоки. Обратите внимание на то, что при описанном подходе сигнал завершения генерируется по прошествии времени, равного наихудшей задержке распространения сигнала по цепи. В результате статистические свойства входных данных не используются. Тем не менее мы можем отследить локальные эффекты колебаний характеристик процесса и параметров окружающей среды (например, температуры или колебаний параметров цепи питания). Данный подход широко используется для внутренней синхронизации полупроводниковых запоминающих устройств, поскольку в них технология внутренней синхронизации используется довольно широко.

Пример 10.5. Альтернативная схема детектирования сигнала завершения с использованием токового считывания

В идеальном случае логику нужно реализовывать с использованием неизбыточных КМОП-схем (например, статических КМОП-элементов), при этом схема детектирования сигнала завершения должна отслеживать зависимости от данных. Подход, позволяющий реализовать эти принципы, показан на рис. 10.38 [Dean94]. Последовательно с комбинационной логикой включен датчик тока, который отслеживает ток, проходящий через логику. Когда ток через логику не проходит (т.е. блок работает в холостом режиме), выход датчик тока имеет низкий уровень, при переключении комбинационной логики уровень становится высоким. Данный сигнал позволяет эффективно определить, когда логический блок завершает свой цикл обработки. Обратите внимание на то, что данный подход позволяет отслеживать время вычислений в зависимости от данных, — если переключаются только биты младших разрядов сумматора, поток тока прекратится, как только младшие разряды переключатся в окончательное значение. Если входной вектор данных не инициирует изменений состояния логики при переходах между тактами, статическая КМОП-логика не потребляет тока от источника питания. В этом

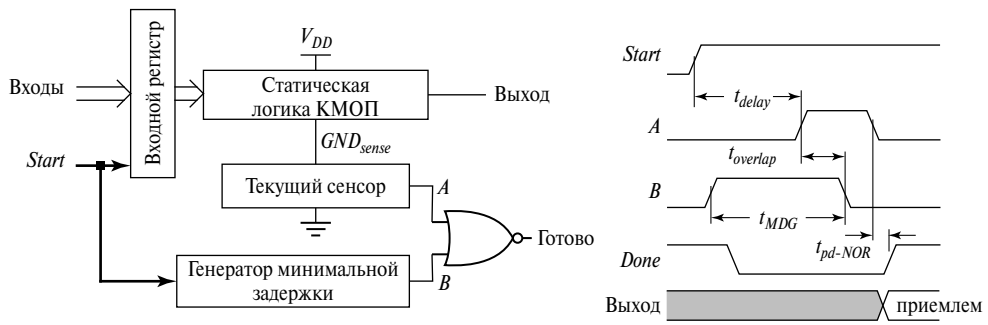


Рис. 10.38. Генерация сигнала завершения с использованием токового считывания

случае для сообщения о завершении работы используется элемент задержки, отслеживающий минимальную задержку прохождения сигнала через логику, и датчик тока. После этого выходы датчика тока и элемента задержки объединяются.

Описанный подход интересен, однако требует внимательного аналогового проектирования. Основной проблемой является обеспечение надежности при минимальных служебных издержках. Из-за этого данный подход используется очень редко, хотя и имеет очевидный большой потенциал.

10.4.3. Передача сигналов с самосинхронизацией

Помимо генерирования сигналов завершения реализация внутренней синхронизации требует протокола установления связи, который позволяет логически упорядочить события схемы и избежать состязаний и сбоев. Функционирование логики передачи сигналов (или установления связи) иллюстрируется на рис. 10.39, где показан отправляющий модуль, передающий данные получателю [Sutherland89]. Передатчик размещает данные в шине данных ① и генерирует событие — управляющий сигнал *Req* меняет полярность ②. В одних случаях событие запроса представляет собой переход в более высокое состояние, в других — обратный переход; в описанном протоколе данные различия не важны. Получив запрос, приемник принимает данные (когда это возможно) и генерирует событие, — переход сигнала *Ack*, сообщающий, что данные были приняты ③. Если приемник занят или его входной буфер полон, событие *Ack* не генерируется и передатчик блокируется, пока приемник не освободится. Как только событие *Ack* сгенерировано, передатчик выдает следующее слово данных ④. Далее четыре описанных события (*изменение данных*, *запрос*, *принятие данных* и *подтверждение*) циклически повторяются. Отметим, что в зависимости от времени, требуемого на генерацию или получение данных, циклы могут иметь разную продолжительность.

Правильная работа системы "передатчик–приемник" требует строгого упорядочения событий, что указано стрелками на рис. 10.39. За поддержание этого порядка отвечает логика установления связи, которая, по сути, выполняет логические манипуляции с событиями. Необходимым компонентом практически любого модуля установления связи является *C-элемент Мюллера*. Данный логический элемент, схематическое изображение и таблица истинности которого представлены на рис. 10.40, применяет к событиям операцию логического И. Выход *C*-элемента — это копия его входов, если они одинаковы. Если входы отличаются друг от друга, выход сохраняет свое предыдущее значение. Иными словами, чтобы выход *C*-элемента Мюллера поменял свое состояние, необходимо наличие событий в обоих входах этого элемента. Если данное условие не выполняется, выход не меняется и выходное событие не генерируется. Реализация *C*-элемента построена на основе защелки, что не удивительно, учитывая сходство их таблиц истинности. Реализация описанной функции в виде статического и динамического элементов показана на рис. 10.41.

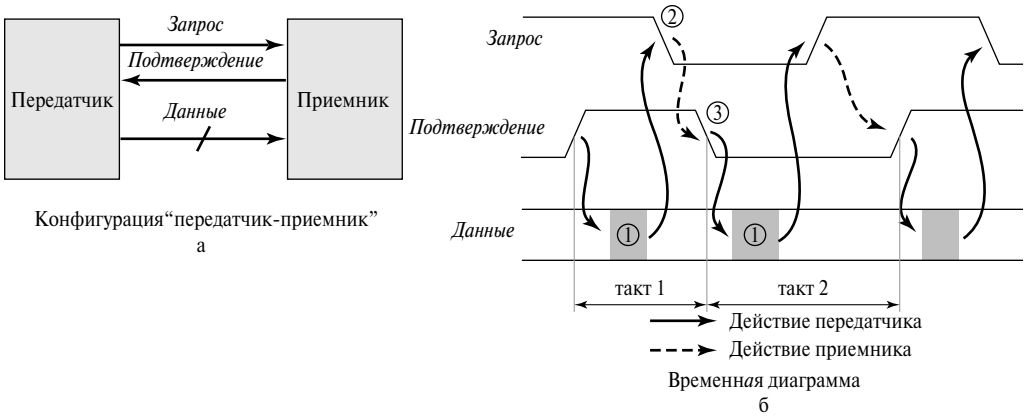
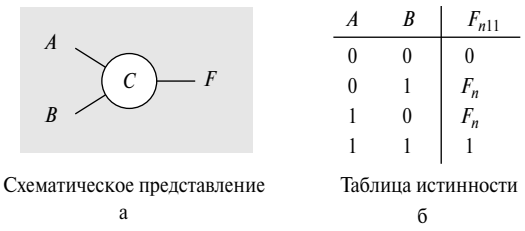


Рис. 10.39. Двухфазный протокол установления связи



A	B	F_{n11}
0	0	0
0	1	F_n
1	0	F_n
1	1	1

Рис. 10.40. *C*-элемент Мюллера

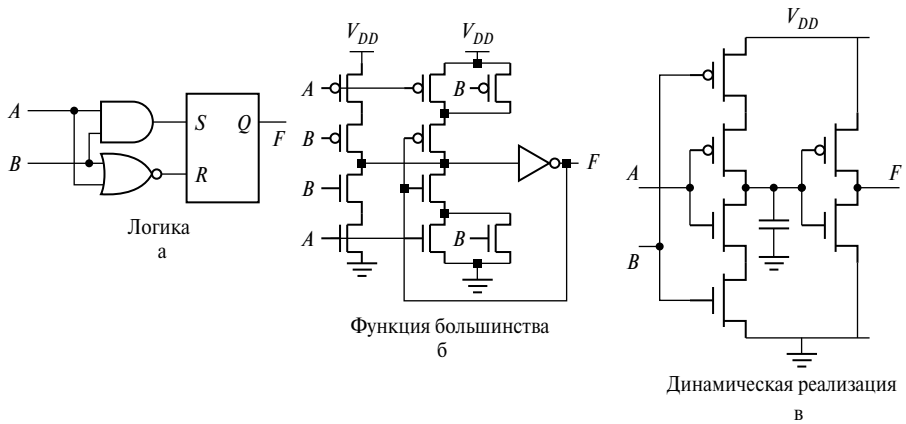


Рис. 10.41. Реализации *C*-элемента Мюллера

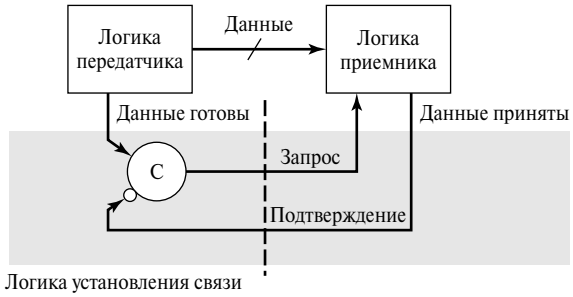


Рис. 10.42. C-элемент Мюллера, реализующий двухфазный протокол установления связи. Кружочек возле нижнего входа C-элемента Мюллера обозначает инверсию

На рис. 10.42 показано, как использовать данный компонент для обеспечения двухфазного протокола установления связи в системе "передатчик–приемник". Предположим, что *Req*, *Ack* и *DataReady* изначально имеют значение 0. Когда передатчик желает отправить следующее слово, сигнал *DataReady* устанавливается равным 1, что активизирует C-элемент, поскольку теперь оба его входа равны 1. Сигнал *Req* переходит на высокий уровень (далее это будет обозначаться *Req* $\uparrow$). Теперь передатчик находится в режиме ожидания и управление передается приемнику. C-элемент блокируется, и никакие новые данные в шину данных не передаются (*Req* остается в высоком состоянии), пока не завершится обработка приемником переданных данных. После этого сигнал *Dataaccepted* переводится на высокий уровень. Это можно сделать множеством различных способов, возможно, задействовав другие C-элементы, сообщающиеся с последующими блоками. Далее происходит событие *Ack* $\uparrow$, которое разблокирует C-элемент и возвращает управление передатчику. Происходит событие *Dataready* $\downarrow$ (возможно, оно уже произошло до *Ack* $\uparrow$), это инициирует событие *Req* $\downarrow$, и цикл повторяется.

Задача 10.1. Двухфазный буфер FIFO с внутренней синхронизацией

На рис. 10.43 показана двухфазная самосинхронизирующаяся реализация буфера FIFO (First-In First-Out — "первым пришел, первым обслужен") с тремя регистрами. Предполагая, что регистры принимают слова данных при повышении и понижении уровня сигналов *En* и что сигнал *Done* представляет собой просто запаздывающую версию сигнала *En*, изучите работу буфера, построив временну*ю* диаграмму всех важных сигналов. Как определить, что буфер FIFO пуст? Полон? (Подсказка: определите необходимые условия для сигналов *Ack* и *Req*.)

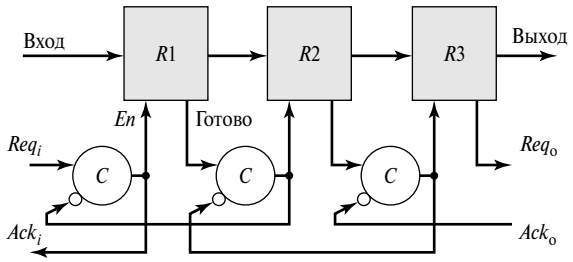


Рис. 10.43. Трехкаскадный буфер FIFO с внутренней синхронизацией, использующий двухфазный протокол передачи сигналов

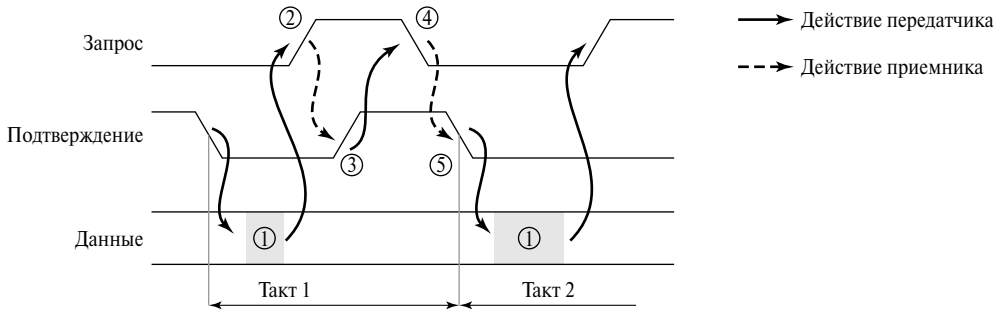


Рис. 10.44. Четырехфазный протокол установления связи

Достоинствами двухфазного протокола являются простота и быстродействие. Тем не менее такой протокол требует детектирования переходов, которые могут происходить в любом направлении. Большинство логических устройств, реализованных по МОП-технологии, чувствительны к уровням или переходам в одном определенном направлении. Логика, управляемая событиями, требует дополнительных цепей, а также хранения информации о состоянии в регистрах и вычислительных элементах. Поскольку нам важно знать направление перехода, необходима инициализация всех *C*-элементов Мюллера в подходящем состоянии. Если этого не сделать, возможно блокирование схемы в определенном состоянии, когда заблокированы все элементы и ничего произойти не может. Если вас интересует данная тема, подробное исследование реализации логики, управляемой событиями, можно найти в [Sutherland89].

Единственной альтернативой является принятие другого подхода к передаче сигналов, например *четырехфазного*, или *с возвращением к нулю* (Return-To-Zero — RTZ). Данный класс систем передачи сигналов требует, чтобы все управляющие сигналы возвращались в исходное состояние перед началом следующего цикла обработки. Проиллюстрируем эту схему на примере системы из передатчика и приемника (рис. 10.44).

В начале цикла мы имеем ту же ситуацию, что и для двухфазного протокола, однако сигналы *Req* и *Ack* находятся в нулевом состоянии. Как только

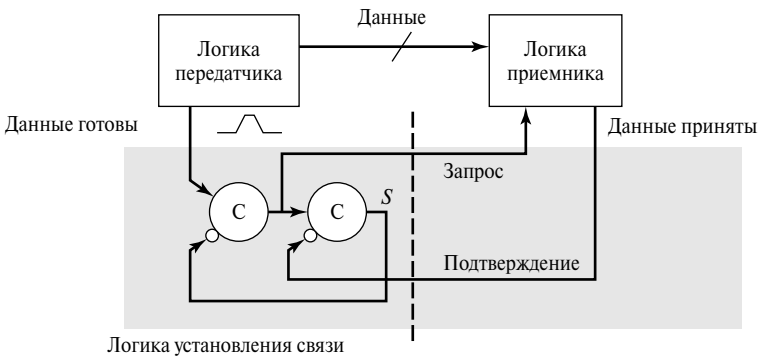


Рис. 10.45. Реализация четырехфазного протокола с использованием *C*-элементов Мюллера

новое слово данных поступает в шину ①, сигнал *Req* переходит на высокий уровень (*Req* ↑ или ②) и управление передается приемнику. Приемник, когда будет готов, примет данные и переведет сигнал *Ack* на высокий уровень (*Ack* ↑ или ③). До этого места ситуация ничем не отличается от двухфазного протокола. Далее сигналы *Req* (*Req* ↓ или ④) и *Ack* (*Ack* ↓ или ⑤) опускаются на первоначальный низкий уровень. Только после этого передатчику разрешается помещать в шину новые данные ①. Описанный протокол называется *четырехфазным* из-за четкого наличия в цикле четырех зон: две связаны с действиями передатчика, две — с действиями приемника. Первые две фазы выглядят так же, как и для двухфазного протокола, а последние две применяются для возврата системы в исходное состояние. На рис. 10.45 показана реализация данного протокола на основе *C*-элементов Мюллера. Интересно отметить, что четырехфазный протокол требует последовательного соединения двух *C*-элементов (поскольку требуется представить четыре состояния). Сигналы *Данные готовы* и *Данные приняты* должны быть импульсами, а не просто одиночными переходами.

Задача 10.2. Четырехфазный протокол

Постройте временну*ю* диаграмму для сигналов, показанных на рис. 10.45. Предположите, что сигнал *Данные готовы* является импульсом, а сигнал *Данные приняты* — это запаздывающая версия сигнала *Req*.

Недостатком четырехфазного протокола является чрезмерная сложность и медленная работа, поскольку для каждой передачи требуется наличие двух событий сигналов *Req* и *Ack*. С другой стороны, он более надежен. По этой причине четырехфазное установление связи предпочтительнее в большинстве современных схем с внутренней синхронизацией. Двухфазный протокол выбирается в основном тогда, когда передатчик и приемник удалены друг от друга, так что управляющие проводники (сигналы *Ack* и *Req*) характеризуются существенной задержкой.

Пример 10.6. Конвейерный информационный канал (продолжение)

Пришло время завершить рассмотрение примера конвейерного информационного канала. Рассмотрим рис. 10.46, на котором представлена схема этого канала с внутренней синхронизацией и временны̀м управлением. Логические функции $F1$ и $F2$ реализованы с использованием двухшинной дифференциальной логики.

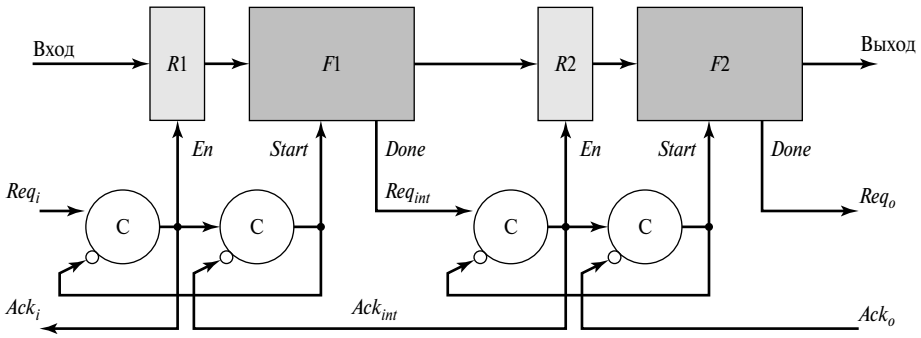


Рис. 10.46. Конвейерный информационный канал с внутренней синхронизацией (окончательный вариант)

Чтобы понять работу данной схемы, предположим, что все сигналы Req и Ack, включая внутренние, переведены в 0 (нулевая активность в информационном канале). Все сигналы $Start$ имеют низкий уровень и все логические схемы находятся в состоянии предварительной зарядки. Входной запрос ($Req_i \uparrow$) активизирует первый C-элемент. Далее повышается уровень сигнала En регистра $R1$, что фиксирует входные данные в регистре (предполагается реализация с управлением положительным фронтом или чувствительная к уровню). Сигнал $Ack_i \uparrow$ подтверждает прием данных. Активизируется второй C-элемент, поскольку Ack_{int} имеет низкий уровень. Это, в свою очередь, повышает уровень сигнала $Start$ и запускает вычисление $F1$. После завершения

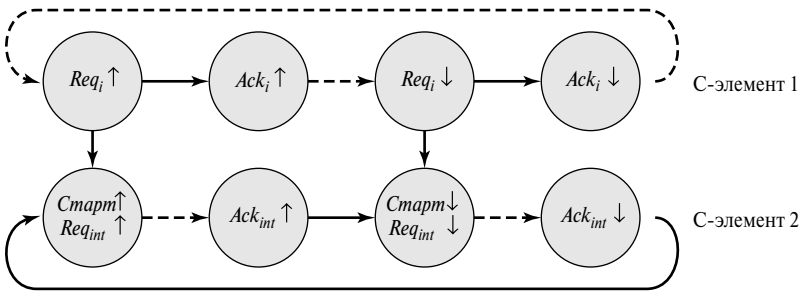


Рис. 10.47. Диаграмма переходов для первого каскада конвейера. Узлы представляют события, а стрелки — зависимости. Пунктирные стрелки обозначают действия предыдущего или следующего каскада

вычислений выходные данные помещаются в шину, и инициируется запрос ко второму каскаду ($Req_{int} \uparrow$), который подтверждает получение этого запроса, повышая уровень сигнала Ack_{int}.

В это время каскад 1 заблокирован для последующих вычислений. Тем не менее входной буфер может реагировать на событие $Ack_i \uparrow$, сбрасывая Req_i в нулевое состояние ($Req_i \downarrow$). В свою очередь, это понижает уровень сигналов En и Ack_i. Получив сигнал $Ack_{int} \uparrow$, сигнал *Start* переходит на низкий уровень, начинается фаза предварительной зарядки, и теперь регистр $F1$ готов для обработки новых данных. Обратите внимание на то, что данная последовательность действий соответствует описанному ранее четырехфазному механизму установления связи. Взаимозависимости событий представлены на *диаграмме переходов* (State Transition Diagram — STG), показанной на рис. 10.47. Подобные диаграммы могут быть очень сложными, часто для их получения применяются компьютерные средства, гарантируются правильное построение диаграммы и оптимизирующее быстродействие представляемой схемы.

10.4.4. Практические примеры логики с внутренней синхронизацией

Схемы с внутренней синхронизацией позволяют существенно повысить быстродействие. К сожалению, из-за наличия значительных служебных издержек самосинхронизация не получила должного распространения в универсальных приложениях цифровой обработки сигналов. Однако ключевые концепции схем с внутренней синхронизацией используются в нескольких практических приложениях. Ниже мы представим несколько примеров, иллюстрирующих использование этих концепций либо для экономии энергии, либо для повышения быстродействия.

Подавление импульсных помех с использованием внутренней синхронизации

Основным источником ненужной емкости переключения в длинных информационных каналах (например, разрядно-модульных сумматорах и умножителях) являются паразитные переходы, вызванные *распространением импульсных помех*. Несбалансированность логических схем приводит к тому, что входы логических элементов или блоков прибывают в разное время, порождая импульсные переходы. В больших комбинационных блоках (например, умножителях) переходы происходят "волнами" по мере того, как первичный вход проходит через логику. Чтобы снизить или устранить импульсные переходы, логический блок можно включать только тогда, когда установятся значения на всех его входах. Чтобы реализовать эту идею, можно разбить вычислительные блоки на фазы и ввести между каждой парой фаз буферы с тремя состояниями, которые не позволят импульсным разрядам распространяться по информационному каналу, как показано на рис. 10.48. Для произвольной логической схемы логично предположить, что выходы логического

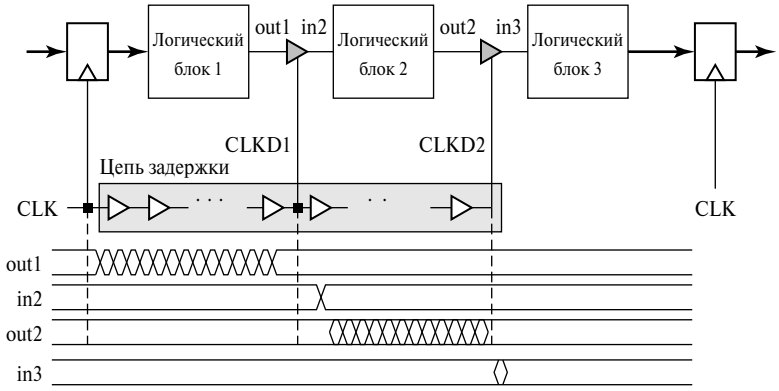


Рис. 10.48. Применение внутренней синхронизации для уменьшения импульсных помех

блока 1 не будут синхронизированы. Активизировав на выходе логического блока 1 буфер с тремя состояниями, мы позволим продолжить работу логическому блоку 2. Чтобы уменьшить интенсивность импульсных переходов, буфер с тремя состояниями нужно активизировать только тогда, когда выходы логического блока 1 имеют установившееся верное значение. Управлять буфером с тремя состояниями можно посредством самосинхронизирующегося сигнала включения, для генерирования которого системный тактовый сигнал пропускается по цепочке задержки, моделирующей критический путь процессора. Далее в цепочке задержки в точках, соответствующих положению буферов, снимаются сигналы и распространяются по всему кристаллу. Описанная техника позволяет существенно уменьшить переключаемую емкость таких блоков комбинационной логики, как умножители (даже с учетом служебных издержек, вводимых цепочкой задержки, избирательным распространением сигналов и буферами) [Goodman98].

Логика с постзарядкой

Интересной формой логики с внутренней синхронизацией является *КМОП-схема с самовозвратом* (self-resetting CMOS). Данная цепь использует управляющую структуру, отличающуюся от привычного конвейера с автосинхронизацией. Вместо того чтобы ожидать, пока все логические каскады завершат свою работу, а только после этого переходить к возврату в исходное состояние, работа схемы построена на предварительной зарядке логического блока сразу после завершения им операции. Поскольку предварительная зарядка происходит после операции, а не перед вычислениями, часто используется термин *логика с постзарядкой* (postcharge logic). Блок-схема логики с постзарядкой показана на рис. 10.49 [Williams00]. Видно, что предварительная зарядка блока L1 происходит, когда следующий за ним каскад завершает работу и уже не требует подачи входа. Вообще, блок можно предварительно заряжать и после завершения выдачи его собственного выхода, однако следует внимательно следить за тем, чтобы каскад, следующий за данным, успел правильно оценить входной сигнал и больше в нем не нуждался.

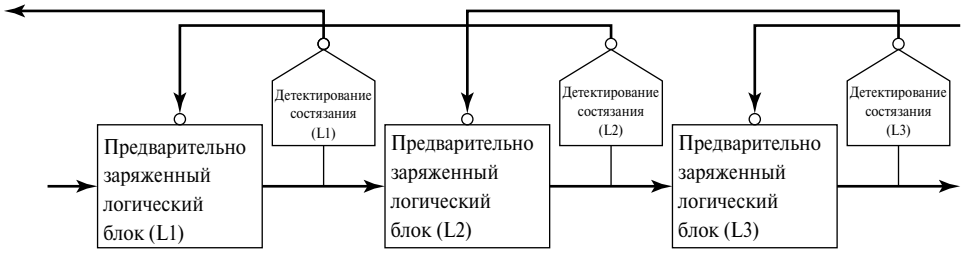


Рис. 10.49. Логика с самовозвратом

Следует отметить, что в отличие от других стилей проектирования логических схем в данном случае сигналы представляются импульсами и действительны только на протяжении некоторого времени. Ширина импульса должна быть меньше, чем задержка возврата схемы к исходному состоянию, иначе между устройством с предварительной зарядкой и переключателями оценки возникнет состязание. Отметим, что хотя описанный стиль разработки логических схем и предлагает потенциальные преимущества с точки зрения быстродействия, следует очень внимательно следить за правильным упорядочением событий. Кроме того, требуется цепь, превращающая уровневые сигналы в импульсы, и наоборот. Пример логики с самовозвратом показан на рис. 10.50 [Bernstein98]. Предположим, что все входы имеют низкий уровень, а сигнал *int* предварительно заряжен. Если на входе A произойдет переход с низкого уровня на высокий, сигнал *int* перейдет на низкий уровень, а *out* — на высокий. В результате логический элемент будет предварительно заряжаться. Чтобы избежать состязания, при активном p-МОП-устройстве предварительной зарядки входы должны находиться в состоянии возврата к исходным уровням.

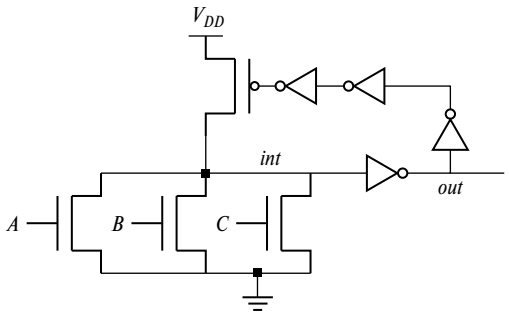


Рис. 10.50. Пример схемы с самовозвратом — логический элемент ИЛИ с тремя входами

Схема “домино” с задержкой тактового сигнала

Одной из интересных сфер применения схем с внутренней синхронизацией, использующей концепцию согласования задержек, является *схема “домино” с задержкой тактового сигнала* (clockdelayed domino) [Yee96, Bernstein00]. Это разновидность динамической логической схемы, в которой нет глобального тактового сигнала. Вместо этого тактовый сигнал для одного каскада выводится из сигналов предыдущего каскада. Простой пример каскада “домино” с задержкой тактового сигнала показан на рис. 10.51. С помощью задержки двух инверторов и передаточного вентиля в тракте тактового сигнала эмулируется наихудшая задержка распространения сигнала через динамический логический элемент. Передаточный вентиль всегда включен, и точное значение задержки тракта тактового сигнала настраивается с помощью масштабирования устройств. Схема “домино” с задержкой тактового сигнала использовалась в микропроцессорах IBM с тактовой частотой 1 ГГц, кроме того, она широко применяется в высокоскоростных логических схемах “домино”. Подобные схемы могут предлагать как инвертированные, так и неинвертированные сигналы. Это хорошо, поскольку основным ограничением обычных техник “домино” является возможность создания только неинвертирующих логических схем. Наличие инвертора после цепи PDN не является необходимым, поскольку тактовый сигнал прибывает на следующий каскад только после завершения оценки на текущем каскаде. С другой стороны, добавление инвертора позволяет обойтись без транзистора оценки (см. главу 6).

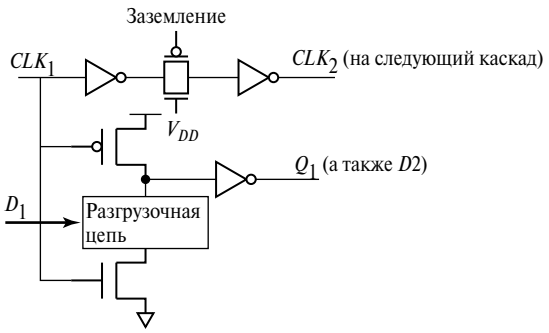


Рис. 10.51. Использование автосинхронизации в схеме “домино” с задержкой тактового сигнала

10.5. Схемы синхронизации и арбитры*

10.5.1. Схемы синхронизации — концепция и реализация

Хотя всю систему целиком можно спроектировать как синхронную, ей все равно придется каким-то образом общаться с внешним миром, в общем случае асинхронным. Как показано на рис. 10.52, асинхронный вход может изменять значение в любой момент, независимо от переходов тактовых сигналов.

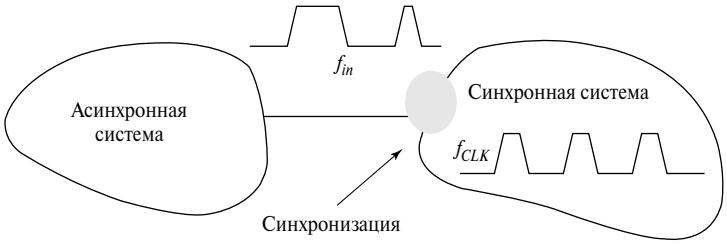


Рис. 10.52. Асинхронно-синхронный интерфейс

Рассмотрим типичный персональный компьютер. Все операции в системе строго контролирует центральный генератор тактовых импульсов, предоставляющий эталонный временной сигнал. Данные "сигналы точного времени" определяют, что произойдет в компьютерной системе в любой момент времени. Такой синхронный компьютер должен общаться с человеком посредством мыши или клавиатуры, причем пользователь не имеет ни малейшего понятия о тактовых импульсах и может нажимать клавиши в любой момент времени. Как синхронная система справляется с подобным асинхронным вводом? Она с определенной периодичностью опрашивает устройства ввода и проверяет их значения. Если частота опроса (или дискретизации) достаточно велика, ни один переход не будет пропущен — в сфере цифровой связи это условие называется *критерием Найквиста*. Тем не менее может случиться так, что сигнал будет опрошен в середине перехода. Получающееся в результате значение не является ни низким, ни высоким — оно неопределенное. В этот момент еще не ясно, нажата клавиша или нет. Подача неопределенного сигнала в компьютер может стать источником всевозможных неприятностей, особенно если он будет направлен не той функции или на тот логический элемент, который не сможет правильно интерпретировать его. Например, одна функция может решить, что нажата определенная клавиша, и инициировать определенное действие, а другая функция может решить иначе и инициировать противоположную команду. Это приведет к конфликту и потенциальному сбою. Следовательно, неопределенное состояние необходимо каким-то образом разрешить, а лишь потом интерпретировать. В действительности нет никакой разницы, какое решение принято, главное, чтобы оно давало однозначный результат. Например, требуется заключить, что клавиша еще не была нажата, и тогда возможное нажатие будет выявлено при следующем опросе клавиатуры, либо считать, что клавиша уже нажата.

Таким образом, асинхронный сигнал необходимо разрешить в высокое или низкое состояние перед подачей в синхронную среду. Цепь, реализующая подобную функцию принятия решений, называется *схемой синхронизации*. А теперь плохие новости: **идеальной схемы синхронизации, всегда дающей правильный ответ, не существует** [Chaney73] [Glasser85]! Схеме синхронизации требуется некоторое время, чтобы принять решение, а в некоторых случаях это время может быть сколь угодно большим. Следовательно, асинхронно-синхронный интерфейс всегда подвержен ошибкам, называемым

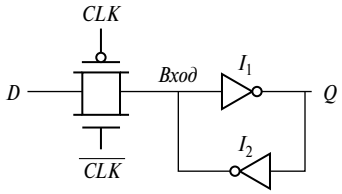


Рис. 10.53. Простая схема синхронизации

сбоям синхронизации. Задача разработчика обеспечить, чтобы вероятность подобного сбоя была такой незначительной, чтобы она не могла повлиять на нормальное поведение системы. Обычно эта вероятность экспоненциально уменьшается при увеличении времени ожидания перед принятием решения. В приведенном примере с клавиатурой это не проблема, но в общем случае ожидание влияет на быстродействие системы, и по возможности прибегать к нему не следует.

Чтобы проиллюстрировать, как ожидание снижает частоту ошибок схемы синхронизации, рассмотрим цепь, показанную на рис. 10.53. Данная схема представляет собой защелку, пропускающую сигнал в низкой фазе тактового сигнала и выбирающую вход при нарастании тактового сигнала CLK. Однако, поскольку выбранный (дискретизованный) сигнал не синхронизирован с тактовым сигналом, существует конечная вероятность того, что время установки или время удержания защелки не будет выдержано. (Вероятность сильно зависит от частоты переходов входа и тактового сигнала.) В результате, как только уровень тактового сигнала начинает расти, увеличиваются шансы того, что выход защелки очутится где-то в неопределенной переходной зоне. Поскольку защелка имеет всего два устойчивых состояния, со временем дискретизованный сигнал обязательно разрешится в 0 либо 1.

Пример 10.7. Траектории синхронизаторов

На рис. 10.54 показаны смоделированные траектории выходных узлов пары статических КМОП-инверторов с перекрестными обратными связями, начальное состояние которой близко к метастабильной точке. Инверторы сконструированы из устройств минимального размера.

Если вход выбирается так, что пара инверторов с перекрестными обратными связями начинает работу в метастабильной точке, при отсутствии шума напряжение будет оставаться метастабильным бесконечно долго. Если данные выбираются с небольшим смещением (положительным или отрицательным), дифференциальное напряжение будет изменяться экспоненциально с характерным временем, зависящим от мощности транзисторов и паразитных емкостей. Время, необходимое сигналу для достижения напряжения в зоне, допускающей разрешение (с приемлемой точностью) в один из логических уровней, зависит от исходного расстояния выбираемого сигнала от метастабильной точки.

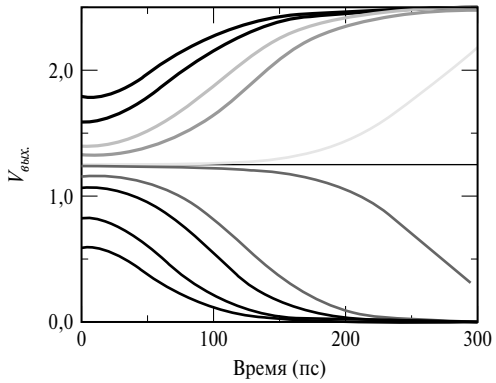


Рис. 10.54. Смоделированные траектории простого синхронизатора

Чтобы определить необходимое время ожидания, мы построим математическую модель поведения бистабильного элемента и используем результаты моделирования для определения вероятности сбоя синхронизации как функции времени ожидания [Veendrick80].

Переходной режим в окрестности метастабильной точки пары инверторов с перекрестными обратными связями точно моделируется системой с одним главным полюсом. В таком приближении переходную характеристику бистабильного элемента после выключения стробирующего тактового сигнала можно смоделировать следующим образом:

$$v(t) = V_{MS} + (v(0) - V_{MS})e^{t/\tau}, \tag{10.8}$$

где V_{MS} — метастабильное напряжение защелки; $v(0)$ — исходное напряжение после выключение стробирующего сигнала; τ — временн*а*я постоянная системы. По результатам моделирования можно сказать, что данная модель первого порядка является достаточно точной.

Теперь мы можем с помощью этой модели рассчитать диапазон значений $v(0)$, которые порождают ошибку (или напряжение в области неопределенности), после периода ожидания T. Сигнал называется неопределенным, если его значение располагается между V_{IH} и V_{IL}:

$$V_{MS} - (V_{MS} - V_{IL})e^{-T/\tau} \leq \nu(0) \leq \ V_{MS} + (V_{IH} - V_{MS})e^{-T/\tau}. \tag{10.9}$$

В уравнении (10.9) заложено важное сообщение: **диапазон входных напряжений, приводящих к ошибке синхронизации, экспоненциально снижается с увеличением периода ожидания** T. Увеличение периода ожидания с 2τ до 4τ уменьшает интервал (а следовательно, и шансы на ошибку) приблизительно в 7,4 раза.

Чтобы вычислить вероятность ошибки, требуется еще некоторая информация об асинхронном сигнале. Предположим, что $V_{\text{вх.}}$ — периодический сиг-

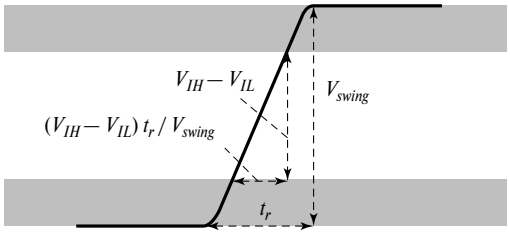


Рис. 10.55. Линейная аппроксимация наклона сигнала

нал со средним периодом T_{signal} между переходами и одинаковыми временами нарастания и спада t_r. Предположим также, что наклоны сигналов в неопределенной области можно аппроксимировать линейной функцией, как показано на рис. 10.55. Используя данную модель, мы можем оценить вероятность P_{init} того, что $v(0)$ (значение $V_{\text{вх.}}$ в момент выборки) попадает в неопределенную область:

$$P_{init} = \frac{\left(\frac{V_{IH} - V_{IL}}{V_{swing}}\right) t_r}{T_{signal}}. \tag{10.10}$$

Шансы появления ошибки синхронизации зависят от частоты синхронизирующего сигнала ϕ. Чем больше число выборок, тем выше шансы на ошибку. Это означает, что без использования синхронизатора среднее число ошибок синхронизации в секунду $N_{sync}(0)$ равно:

$$N_{sync}(0) = \frac{P_{init}}{T_\phi}, \tag{10.11}$$

где T_ϕ — период дискретизации.

Из уравнения (10.9) видно, что ожидание в течение времени T до наблюдения выхода экспоненциально уменьшает вероятность того, что сигнал по-прежнему будет неопределенным:

$$N_{sync}(T) = \frac{P_{init} e^{-T/\tau}}{T_\phi} = \frac{(V_{IH} - V_{IL}) e^{-T/\tau}}{V_{swing}} \frac{t_r}{T_{signal} T_\phi}. \tag{10.12}$$

Таким образом, устойчивость асинхронно-синхронного интерфейса определяется следующими параметрами: частотой переключения сигнала и временем нарастания, частотой дискретизации и временем ожидания T.

Пример 10.8. Схемы синхронизации и среднее время наработки на отказ

Рассмотрим следующий проект: $T_\phi = 5$ нс, что соответствует тактовой частоте 200 МГц; $T = T_\phi = 5$ нс; $T_{signal} = 50$ нс; $t_r = 0,5$ нс и $\tau = 150$ пс. Из передаточной характеристики типичного КМОП-инвертора можно найти,

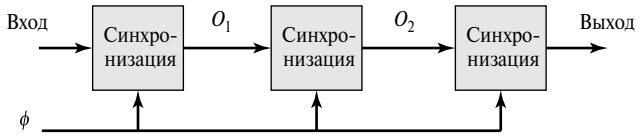


Рис. 10.56. Каскадные (активизируемые фронтом) синхронизаторы уменьшают среднее время наработки на отказ

что для размаха напряжения 2,5 В значение $V_{IH} - V_{IL}$ приблизительно равно 0,5 В. Подставляя наши числа в уравнение (10.12), получим вероятность ошибки $1,38 \times 10^{-9}$ ошибок/с. Величина, обратная к N_{sync}, называется *средним временем наработки на отказ* (mean time to failure) и в данном случае равна 7×10^8 с, или 23 года. Если же не использовать синхронизатор, то среднее время наработки на отказ составит всего 2,5 мс!

Вопрос проектирования

Разрабатывая синхронно-асинхронный интерфейс, необходимо учитывать следующие моменты.

- Приемлемая частота отказов системы зависит от множества экономических и социальных факторов, а также от области применения системы.
- Из-за экспоненциальной зависимости в уравнении (10.12) частота сбоев крайне чувствительна к значению τ. Определить точное значение τ непросто, поскольку они различаются для разных кристаллов, а кроме того, зависят от температуры. Таким образом, вероятность появления ошибки может флуктуировать в широком диапазоне даже для одной и той же структуры. Если частота сбоев для наихудшего сценария превышает определенное значение, заданное в виде критерия, ее можно уменьшить, увеличив значение T. Когда T превышает период дискретизации T_ϕ, следует ожидать появления проблем. Избежать их можно с помощью каскадного соединения нескольких синхронизаторов (организации конвейера), как показано на рис. 10.56. Каждый из этих синхронизаторов имеет период ожидания, равный T_ϕ. Обратите внимание на то, что при таком упорядочении импульс ϕ должен быть достаточно коротким, чтобы не возникло состязания. Глобальный период ожидания равен сумме значений T_s всех схем синхронизации. Отметим, что полученное в результате увеличение среднего времени наработки на отказ происходит за счет увеличения времени ожидания.
- Ошибки синхронизации очень сложно отследить из-за их случайной природы. Даже сделав среднее время наработки на отказ очень большим, вы не застрахуетесь от ошибок. Следовательно, **число схем синхронизации в системе не должно быть большим**. Рекомендуется, чтобы в системе присутствовало не более одного-двух синхронизаторов.

10.5.2. Арбитры

Помимо схем синхронизации, следует упомянуть еще одних их "близких родственников", именуемых *арбитрами*, или взаимно-исключающими схемами. Арбитр — это элемент, решающий, какое из двух событий произошло первым. Например, подобные компоненты позволяют нескольким процессорам обращаться к одному ресурсу, например, к большому совместно используемому запоминающему устройству. По сути, схема синхронизации представляет собой частный случай арбитра, поскольку он определяет, когда произошло изменение уровня сигнала — до или после перехода тактового сигнала. Следовательно, синхронизатор — это арбитр, один из входов которого связан с тактовым сигналом.

Рассмотрим рис. 10.57, где показан взаимоисключающий элемент. Он, согласно четырехфазному протоколу, обрабатывает два сигнала с запросами на ввод, т.е. до запуска нового сигнала Req (запрос) текущий сигнал необходимо вернуть в исходное состояние. Выход элемента состоит из двух сигналов Ack (подтверждение), которые должны быть взаимоисключающими. Хотя сигналы Req могут одновременно переходить на высокий уровень, в любой момент времени допускается переход только одного из сигналов Ack на высокий уровень. Чтобы представить себе работу данной схемы, начнем с момента, когда оба входа имеют низкий уровень (ни одно из устройств не подает запроса), узлы A и B имеют высокий уровень, а оба сигнала Ack имеют низкий уровень. Событие, происходящее на одном из входов (например, $Req_1 \uparrow$), приводит к переключению триггера, переходу узла A в низкое состояние и $Ack_1 \uparrow$. Одновременные события на обоих входах переводят триггер в метастабильное состояние, и некоторое время сигналы A и B могут быть неопределенными. Благодаря перекрестному включению выходов значения на выходах будут находиться в низких состояниях, пока один из выходов НЕ-И не станет достаточно (определяется пороговым значением V_T) отличным от другого. При таком подходе появление импульсных помех на выходе невозможно.

10.6. Синтез тактовых сигналов и синхронизация с использованием ФАПЧ*

Существует множество цифровых приложений, требующих внутрикристальной генерации периодического сигнала. В синхронных схемах нужен глобальный эталонный периодический тактовый сигнал, позволяющий "дирижировать" последовательностью событий. Современные микропроцессоры и быстродействующие цифровые схемы требуют, чтобы тактовая частота принадлежала гигагерцовому диапазону. Для справки: кварцевые генераторы могут точно генерировать только тактовые сигналы с частотой от нескольких десятков мегагерц до приблизительно 200 МГц. Для генерации более высокой частоты, требуемой цифровыми схемами, обычно используется схема *фазовой автоподстройки частоты*, ФАПЧ (Phase-Locked Loop — PLL). ФАПЧ принимает внешний низкочастотный сигнал, сгенерированный

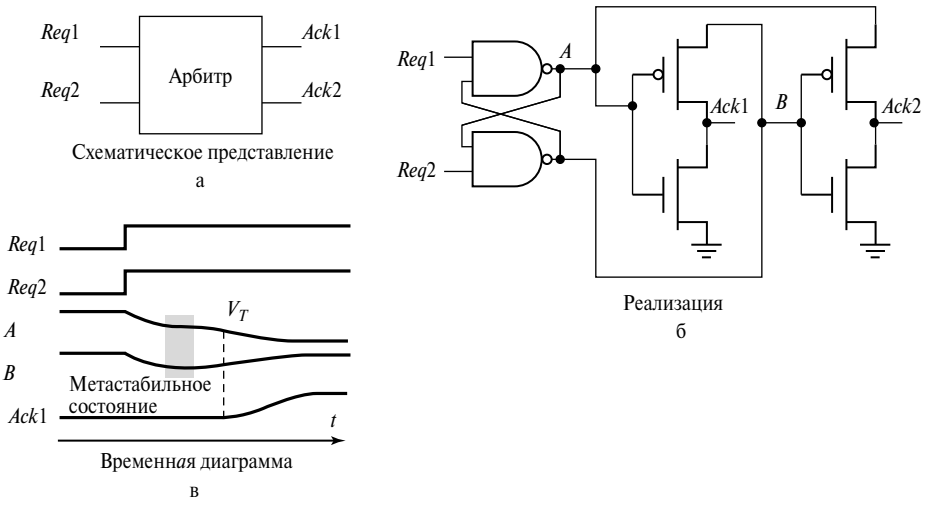


Рис. 10.57. Взаимоисключающий элемент (или арбитр)

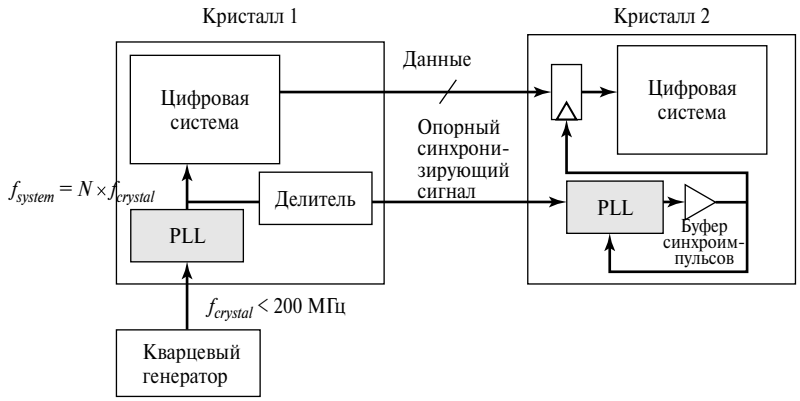


Рис. 10.58. Применения систем фазовой автоподстройки частоты (ФАПЧ)

кварцевым генератором, и умножает его частоту на рациональное число N (см. рис. 10.58, *слева*).

Другой не менее важной функцией ФАПЧ является синхронизация связи между кристаллами. Как показано на рис. 10.58, параллельно с данными отправляется эталонный тактовый сигнал. (В данном примере показан только путь передачи от кристалла 1 к кристаллу 2.) Поскольку связь между кристаллами чаще всего происходит с частотой, меньшей частоты внутрикристального тактового сигнала, частота эталонного тактового сигнала уменьшается в несколько раз, но при этом остается синфазной частоте системного тактового сигнала. В кристалле 2 для синхронизации всех входных триггеров применяется эталонный тактовый сигнал, которыqй может порождать боль-

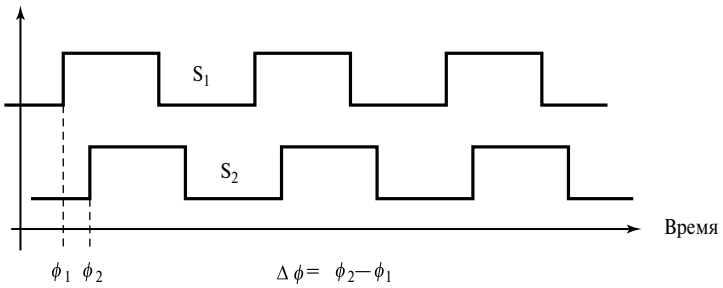


Рис. 10.59. Относительная и абсолютная фаза двух периодических сигналов

шую нагрузку на тактовый сигнал в случае использования широких шин данных. К сожалению, реализация буферов тактовых сигналов, которые позволили бы справиться с этой проблемой, приводит к расфазировке данных и тактового сигнала. Схема ФАПЧ выравнивает (т.е. убирает рассинхронизацию) выход буфера тактовых сигналов и данные. Кроме того, схема ФАПЧ может умножать частоту входного эталонного тактового сигнала, позволяя ядру второго кристалла работать с частотой, большей частоты входного эталонного сигнала.

10.6.1. Суть

Как показано на рис. 10.59, набор из нескольких периодических сигналов одинаковой частоты легко определяется, если мы знаем один из этих сигналов и его фазу относительно других сигналов. На рисунке ϕ_1 и ϕ_2 представляют фазу двух сигналов. Относительная фаза (разность фаз) представляет собой разность этих двух величин.

Схема ФАПЧ представляет собой сложный нелинейный контур обратной связи. Чтобы понять принцип его работы, удобно рассмотреть схему, приведенную на рис. 10.60 [Jeong87]. *Генератор, управляемый напряжением*, ГУН (Voltage-Controlled Oscillator — VCO) принимает аналоговый управляющий входной сигнал и генерирует тактовый сигнал требуемой частоты. В общем случае выходная частота является нелинейной функцией управляющего напряжения (v_{cont}). Для получения системной тактовой частоты с определенным значением необходимо установить требуемое значение управляющего напряжения. Это значение является функцией остальных блоков и контура обратной связи в схеме ФАПЧ. Контур обратной связи жизненно важен для отслеживания колебаний параметров процесса и среды. Кроме того, обратная связь позволяет умножать частоту.

В общем случае эталонный тактовый сигнал генерируется вне кристалла. Он сравнивается с "масштабированной" версией системного тактового сигнала (т.е. с локальным тактовым сигналом) с помощью фазового детектора, который определяет разность фаз между сигналами и генерирует сигнал *Up* или *Down* в случае, если локальный тактовый сигнал запаздывает по отношению к эталонному сигналу (или опережает его). Он определяет, какой из

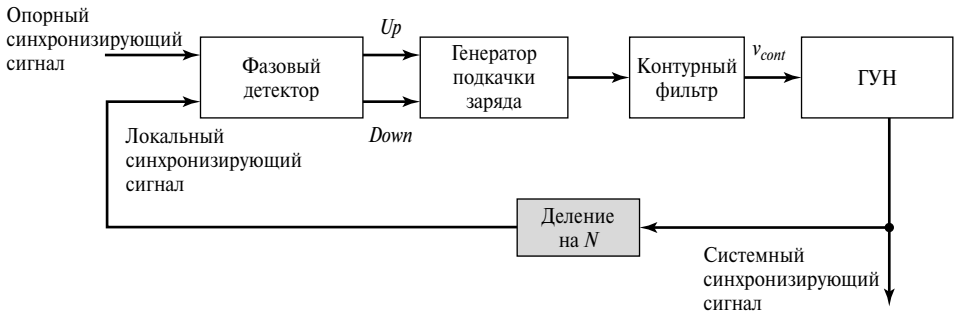


Рис. 10.60. Функциональный состав схемы фазовой автоподстройки частоты (ФАПЧ)

двух входных сигналов поступил раньше, и генерирует соответствующий выходной сигнал. Сигналы Up и $Down$ подаются в генератор подкачки заряда, который переводит закодированную в цифровом виде управляющую информацию в аналоговое напряжение [Gardner80]. Сигнал Up повышает значение управляющего напряжения и ускоряет ГУН, что приводит к согласованию локального сигнала с эталонным тактовым сигналом. С другой стороны, сигнал $Down$ замедляет генератор и устраняет опережение по фазе, имеющееся у локального тактового сигнала.

Передача выхода генератора подкачки заряда непосредственно в ГУН создает дрожащий тактовый сигнал. Фронт локального тактового сигнала постоянно прыгает туда-сюда и осциллирует возле искомого положения. Как обсуждалось ранее, дрожание тактового сигнала крайне нежелательно, поскольку оно уменьшает время, доступное для логических вычислений, а следовательно, дрожание необходимо удерживать в разумных пределах. Отчасти этому помогает введение в цепи обратной связи *контурного фильтра*. Данный фильтр нижних частот устраняет высокочастотные компоненты из сигнала управляющего напряжения ГУН и сглаживает его отклик, что приводит к уменьшению дрожания. Обратите внимание на то, что структура схемы ФАПЧ соответствует структуре контура обратной связи. Добавление лишних сдвигов фаз, которые вводятся фильтрами высоких порядков, может привести к неустойчивости. Схемы ФАПЧ характеризуются такими важными свойствами, как диапазон входных частот, в котором схема может выполнять свои функции; время, которое требуется схеме ФАПЧ для привязки к данному входному сигналу; и дрожание. В рабочем режиме системная тактовая частота в N раз превышает эталонную тактовую частоту.

Схема ФАПЧ является аналоговой, а следовательно, она чувствительна к шуму и помехам. Особенно это относится к контурному фильтру в ГУН, для которого наведенный шум непосредственно влияет на получающееся дрожание тактового сигнала. Основным источником помех является шумовая связь между шинами питания и подложкой. Это представляет серьезную проблему в цифровой среде, имеющей множество источников шума. Следовательно, желательно использовать аналоговые схемы с хорошими цепями подавления

шумов питания (например, дифференциальными ГУН) [Kim90]. Подведем итоги: введение высокочувствительного компонента во враждебную цифровую среду является нетривиальной задачей и требует дополнительных аналоговых цепей. Более подробное описание различных компонентов схем ФАПЧ представлено в следующих разделах.

10.6.2. Основные составляющие схем ФАПЧ

Генератор, управляемый напряжением (ГУН)

ГУН генерирует периодический сигнал с частотой, являющейся линейной функцией входного управляющего напряжения v_{cont}. Другими словами, частоту ГУН можно выразить следующим образом:

$$\omega = \omega_0 + K_{vco} \cdot v_{cont}. \tag{10.13}$$

Поскольку фаза — это интеграл по времени от функции, представляющей изменение частоты, выходная фаза блока ГУН записывается следующим образом:

$$\phi_{\text{вых.}} = \omega_0 t + K_{vco} \cdot \int_{-\infty}^{t} v_{cont} dt, \tag{10.14}$$

где K_{vco} — коэффициент усиления ГУН, выражаемый в рад/с/В; ω_0 — фиксированный сдвиг частоты; v_{cont} контролирует частоту, центрированную на ω_0. Выходной сигнал имеет следующую форму:

$$x(t) = A \cdot \cos\left(\omega_0 t + K_{vco} \cdot \int_{-\infty}^{t} v_{cont} dt\right). \tag{10.15}$$

О различных стратегиях реализации генераторов, управляемых напряжением, см. в главе 7. В настоящее время предпочтительным является метод дифференциальных кольцевых генераторов.

Фазовые детекторы

Фазовый детектор определяет относительную разность фаз между двумя входами и выдает сигнал, пропорциональный данной разности фаз. Один из входов фазового детектора представляет собой эталонный тактовый сигнал, который обычно генерируется вне кристалла, а другой вход — это масштабированная версия сигнала ГУН. В настоящее время широко используются два основных типа фазовых детекторов — *элементы ИСКЛЮЧАЮЩЕЕ ИЛИ* и *фазочастотный детектор* (Phase Frequency Detector — PFD).

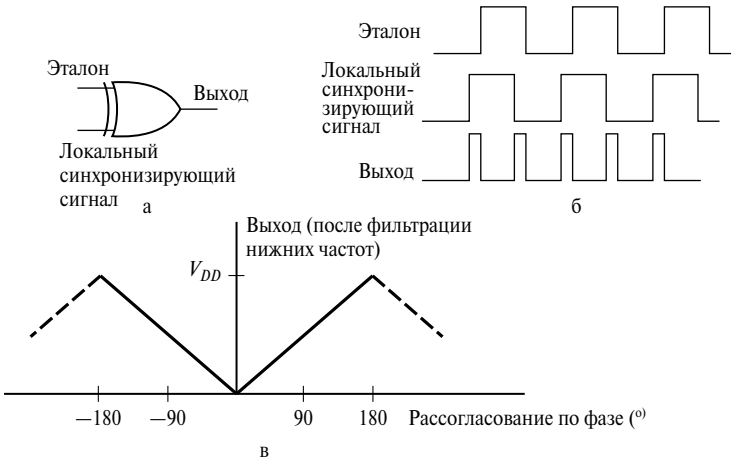


Рис. 10.61. Элемент ИСКЛЮЧАЮЩЕЕ ИЛИ в роли фазового детектора: *а* — схема; *б* — выход перед фильтрацией нижних частот; *в* — передаточная характеристика

Фазовый детектор ИСКЛЮЧАЮЩЕЕ ИЛИ

Элемент ИСКЛЮЧАЮЩЕЕ ИЛИ (рис. 10.61, *а*), являющийся простейшим из двух вариантов, удобно использовать в роли фазового детектора по следующей причине: относительная разность фаз между входами отражается на времени, в течение которого входы отличаются. На рис. 10.61, *б* представлен результат применения функции ИСКЛЮЧАЮЩЕЕ ИЛИ к двум сигналам. Передача данного выхода на фильтр нижних частот дает сигнал, пропорциональный разности фаз (см. рис. 10.61, *в*).

Для данного детектора отклонение в положительном или отрицательном направлении от идеального синфазного состояния (т.е. когда рассогласование по фазе равно нулю) порождает идентичное изменение коэффициента заполнения, что дает такое же среднее напряжение. Следовательно, область, в которой фаза линейно зависит от напряжения, составляет всего 180°. В стационарном состоянии схема ФАПЧ фиксируется так, что между двумя входами сохраняется квадратурное соотношение по фазе (смещение равно $\frac{1}{4}$ периода).

Фазочастотный детектор

Фазочастотный детектор (Phase Frequency Detector — PFD) представляет собой наиболее распространенную форму фазового детектора, не имеющую некоторых недостатков, характерных для детектора рассмотренного выше [Dally98]. Как следует из названия, выход PFD зависит как от разности фаз, так и от разности частот сигналов, поступающих на его вход. Соответственно, он не может зафиксироваться на неправильной частоте, кратной искомой. Как показано на рис. 10.62, PFD принимает на вход два тактовых сигнала и генерирует два выхода, UP и DN.

Фазочастотный детектор представляет собой конечный автомат с тремя состояниями. Предположим, что оба выхода UP и DN изначально имеют низкий уровень. Когда вход A опережает вход B, выход UP активизируется

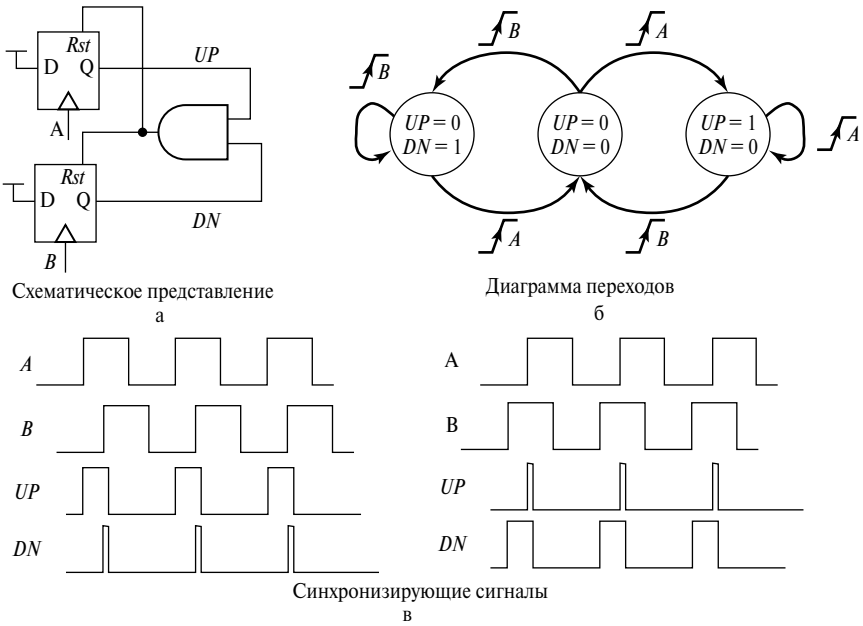


Рис. 10.62. Фазочастотный детектор

на нарастающем фронте входа A. Сигнал UP остается в этом состоянии до тех пор, пока на входе B не произойдет изменение уровня сигнала от низкого к высокому. В это время активизируется выход DN, после чего посредством асинхронного сигнала сброса оба триггера возвращаются в исходное состояние. Обратите внимание на небольшой импульс в выходе DN, длительность которого равна задержке прохождения сигнала через элемент И и задержке регистра *Reset-Q*. Ширина импульса UP равна рассогласованию по фазе двух сигналов. В случае, когда вход A запаздывает по отношению к входу B, сигналы меняются ролями. Импульс, пропорциональный разности фаз, генерируется в выходе DN. Если контур синхронизован, небольшие импульсы будут генерироваться и в выходе UP, и в выходе DN.

Описанная схема также действует как детектор частоты, позволяя измерять ошибку по частоте (рис. 10.63). В случае, когда сигнал A имеет более высокую частоту, чем B, фазочастотный детектор генерирует больше импульсов UP (средний импульс пропорционален разности частот), тогда как число импульсов DN в среднем близко к нулю. В случае, когда сигнал B имеет частоту, бо́льшую, чем A, имеем противоположную ситуацию — в выходе DN генерируется гораздо больше импульсов, чем в выходе UP.

На рис. 10.64 показана фазовая характеристика фазочастотного детектора. Обратите внимание на то, что линейная область увеличилась до 4π.

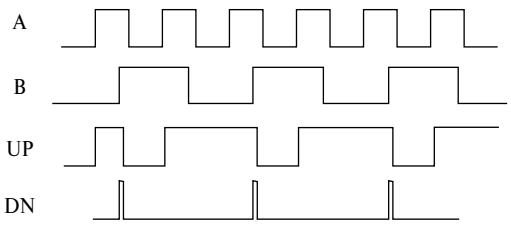


Рис. 10.63. Настройка фазочастотного детектора, измеряющего ошибку по частоте

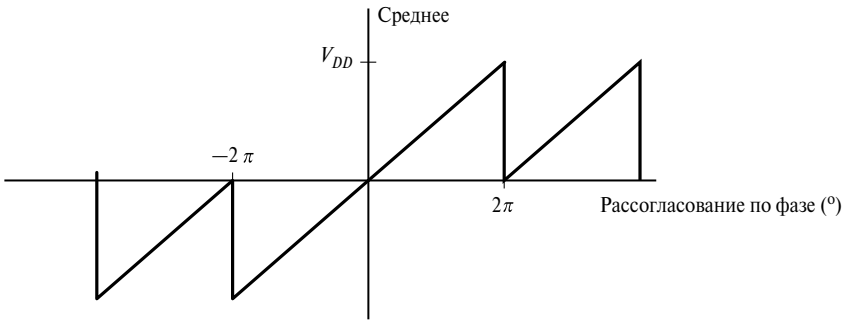


Рис. 10.64. Фазовая характеристика фазочастотного детектора

Генератор подкачки заряда

Импульсы UP и DN необходимо преобразовать в аналоговое напряжение, которое управляет ГУН. Одна из возможных реализаций соответствующей схемы показана на рис. 10.65. Импульс в сигнале UP добавляет в конденсатор заряд, пропорциональный ширине импульса UP, тогда как импульс в сигнале DN снимает заряд, пропорциональный ширине импульса DN. Если ширина импульса UP больше, чем ширина импульса DN, имеем суммарное увеличение управляющего напряжения, что, в свою очередь, увеличивает частоту ГУН.

Пример 10.9. Переходная характеристика схемы ФАПЧ

Под временем установления схемы ФАПЧ понимают время, требуемое для выхода на стационарный режим. Длительность начального переходного процесса сильно зависит от ширины полосы контурного фильтра. На рис. 10.66 показаны результаты моделирование в SPICE схемы ФАПЧ, реализованной по 0,25-микронной КМОП-технологии. Для ускорения моделирования использовался идеальный ГУН. В данном примере была выбрана эталонная частота 100 МГц, а схема ФАПЧ умножала эту частоту на 8 (800 МГц). На рис. 10.66, *а* показаны переходная характеристика и процесс установления управляющего напряжения на входе ГУН. Как только управляющее напряжение достигает конечного значения, выходная частота (тактовая частота

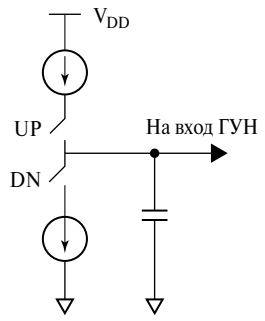


Рис. 10.65. Генератор подкачки заряда

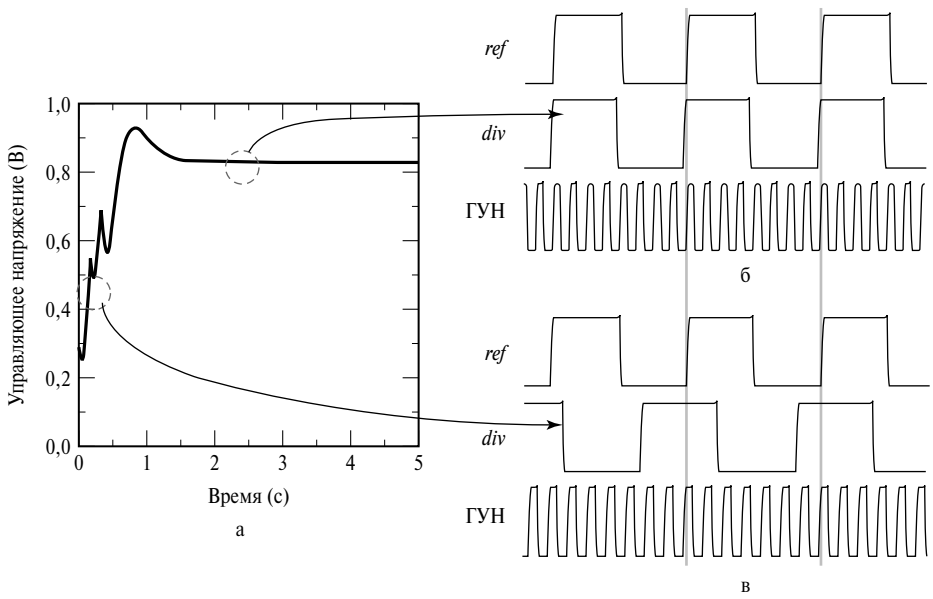


Рис. 10.66. Моделирование в SPICE схемы ФАПЧ. Вместе с сигналами до и после синхронизации показано управляющее напряжение ГУН

цифровой системы) выходит на стационарный режим. На рис. 10.66, *б* и *в* показаны эталонная частота, выход делителя и выходная частота в стационарном режиме. На рис. 10.66, *б* показана схема ФАПЧ, синхронизированная на частоте $f_{out} = 8 \times f_{ref}$, а на рис. 10.66, *в* — сигналы в процессе синхронизации, когда выход еще не синхронизирован с входом.

Резюме

Довольно быстро схемы ФАПЧ стали необходимым компонентом любой высокоскоростной цифровой структуры. Проектирование таких структур требует множества специальных навыков, интеграции аналоговых схем во "враждебную" цифровую среду. Тем не менее данные условия вполне выполнимы и позволяют получать новые и лучшие решения.

10.7. Перспективы

В данном разделе рассмотрены некоторые тенденции в области оптимизации тактирования, направленные на получение более высокого быстродействия и низкого рассеяния мощности.

10.7.1. Распределенное тактирование и использование систем автоподстройки по задержке

Не так давно в сфере высокоскоростного тактирования появилась тенденция к использованию одной из разновидности схем ФАПЧ — *систем автоподстройки по задержке* (Delay Locked Loop — DLL). Схема такой системы показана на рис. 10.67, *а* [Maneatis00]. Ключевым компонентом структуры является *управляемая напряжением линия задержки* (Voltage-Controlled Delay Line — VCDL). Она состоит из каскада настраиваемых элементов задержки. Идея системы заключается в **такой задержке выходного тактового сигнала, чтобы он идеально согласовался с эталонным сигналом**. В отличие от использования ГУН в данной схеме нет генератора тактовых сигналов. Эталонная частота подается на вход VCDL. Подобно ситуации со структурой ФАПЧ, фазовый детектор сравнивает эталонную частоту с выходом линии задержки (F_O) и генерирует сигнал ошибки $UP - DN$. Обратите внимание на то, что требуется только детектирование фазы (а не фазы и частоты). После синхронизации ошибки между двумя тактовыми сигналами отсутствуют. Функцией обратной связи является такая настройка задержки распространения через VCDL, чтобы совпали нарастающие фронты входного эталонного тактового сигнала (f_{REF}) и выходного тактового сигнала (f_O).

Качественное представление сигналов линии автоподстройки по задержке показано на рис. 10.67, *б*. Изначально линия рассинхронизирована. Поскольку первый фронт сигнала на выходе появляется раньше фронта эталонного сигнала, генерируется импульс UP, ширина которого равна ошибке между двумя сигналами. Роль генератора подкачки заряда заключается в генерации заряда, пропорционального ошибке, что увеличивает управляющее напряжение VCDL. Это приводит к тому, что в следующем такте фронт выходного сигнала задерживается. (В описанной реализации VCDL предполагается, что большее напряжение дает бо́льшую задержку.) Через несколько тактов фазовая ошибка будет исправлена и два сигнала синхронизированы. Обратите внимание на то, что DLL не изменяет частоту входного эталонного сигнала, а просто настраивает его фазу.

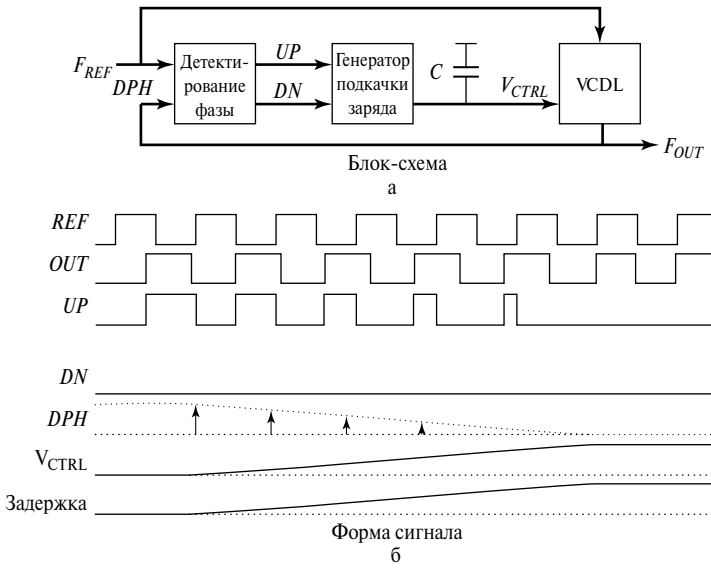


Рис. 10.67. Система автоподстройки по задержке [Maneatis00]

На рис. 10.68 показано использование структуры DLL в сети распределения тактовых сигналов. Кристалл разбит на множество небольших областей. Глобальный тактовый сигнал распределяется в каждую ячейку с небольшой рассинхронизацией; этого можно добиться с помощью соответствующего пакета или схемы трассировки с низкой рассинхронизацией. Для простоты на рисунке показан кристалл из двух ячеек, но, разумеется, описанный принцип легко обобщить на большее число областей. Внутри каждой ячейки глобальный тактовый сигнал буферизируется перед подачей на цифровую нагрузку. Перед каждым буфером находится линия VCDL. Целью сети распределения тактовых сигналов является доставка сигнала в цифровую схему с почти нулевыми расфазировкой и дрожанием. Как отмечалось ранее, статические и динамические колебания в буфере приводят к тому, что разность фаз между буферизированными тактовыми сигналами не равна нулю и изменяется со временем. Находящаяся в каждой ячейке цепь обратной связь так настраивает управляющее напряжение VCDL, что буферизированный выход синхронизируется по фазе с глобальным входным тактовым сигналом. Контур обратной связи компенсирует как статические колебания параметров процесса, так и медленные динамические вариации (например, температуры). Подобные конфигурации широко распространены в высокоскоростных цифровых микропроцессорах. Описанный выше подход к распределению тактовых сигналов с использованием нескольких линий автоподстройки по задержке можно обобщить на использование нескольких распределенных линий на кристалле [Gutnik00]. Распределенные линии автоподстройки по задержке предлагают потенциальные преимущества с точки зрения быстродействия, однако для их устойчивой работы требуется внимательный системный анализ.

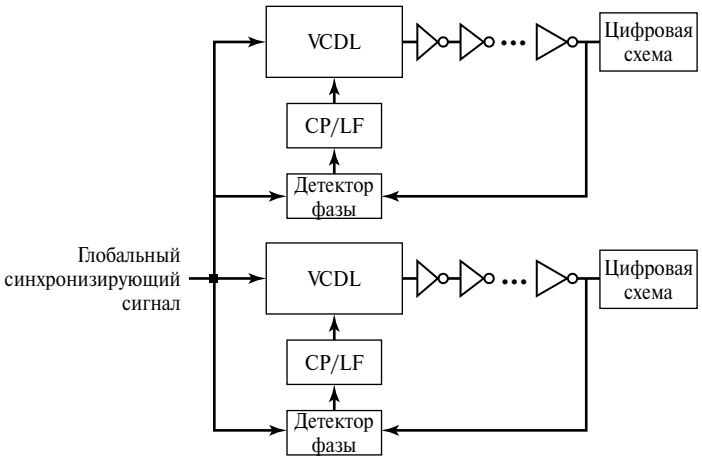


Рис. 10.68. Распределение тактового сигнала с использованием линии автоподстройки по задержке

10.7.2. Оптическое распределение тактового сигнала

Теперь вы должны понимать, что будущие высокоскоростные системы, работающие с частотой несколько гигагерц, будут иметь ряд фундаментальных проблем с синхронизацией. Быстродействие цифровой структуры принципиально ограничивается колебаниями параметров процесса и окружающей среды. Даже при использовании агрессивных активных схем управления тактовыми сигналами (DLL и PLL) колебания параметров цепей питания и нагрузки на тактовый сигнал приводят к неприемлемой неопределенности тактовых сигналов. В настоящее время ученые продолжают поиск инновационных решений данных проблем. Одним из альтернативных подходов, получившим в настоящее время много внимания, является использование оптики для синхронизации системы. Прекрасный обзор сути оптических межсоединений и их плюсов и минусов по сравнению с электрическими представлен в [Miller00].

К потенциальным достоинствам использования оптической технологии при распределении тактового сигнала относят то, что задержка оптических сигналов нечувствительна к температуре и что фронты тактовых сигналов не искажаются при распространении на большие расстояния (т.е. десятки метров). Фактически оптический тактовый сигнал можно доставить на расстояние в несколько десятков метров с неопределенностью, не превышающей 10–100 пс. Оптические тактовые сигналы можно распространять внутри кристалла посредством волноводов или через свободное пространство. Для примера на рис. 10.69 представлена архитектура оптической сети распространения тактовых импульсов с использованием волноводов. Сигналы от внешнего источника переносятся в кристалл и распределяются посредством волноводов, после чего преобразуются и вводятся в локальные электрические цепи распределения тактовых сигналов. Кристалл делится на небольшие ячейки, и глобальный тактовый сигнал распространяется от источника фотонов че-

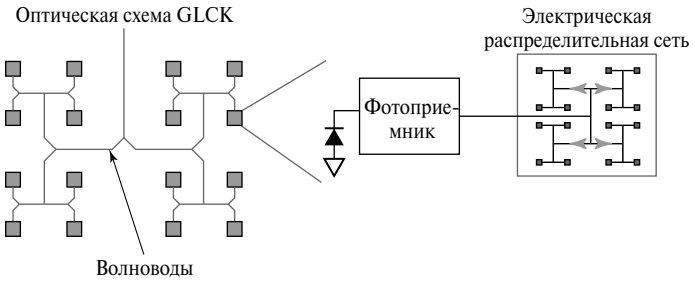


Рис. 10.69. Архитектура оптической цепи распределения тактовых сигналов. (*Перепечатано с разрешения Лайонела Кимерлинга (Lionel Kimerling).*)

рез волноводы с разветвителями, которые отводят сигнал к каждому участку. Обратите внимание на то, что при распределении оптического тактового сигнала применяется H-дерево. В конце волновода находится фотодетектор, который можно реализовать с использованием кремния или германия. После поступления на детектор оптические импульсы, представляющие глобальный тактовый сигнал, преобразуются в импульсы тока, которые имеют очень маленькую амплитуду (десятки миллиампер), и подаются на усилитель, усиливающий сигнал до уровня, подходящего для цифровой обработки. После этого полученный электрический тактовый сигнал распространяется к локальной нагрузке с использованием обычных технологий [Kimerling00].

Преимуществом оптического подхода является то, что расфазировка тактового сигнала при поступлении на фотодетекторы практически равна нулю. Разумеется, небольшие отличия во времени поступления оптических сигналов на детекторы имеют место, однако они минимальны. Например, колебания, вызванные изгибами волноводов, могут привести к тому, что на разных путях будет разным рассеяние энергии. Дополнительным преимуществом оптики является еще и отсутствие многих сложностей, связанных с распространением электромагнитных волн, например перекрестных помех и индуктивной связи.

С другой стороны, проблематичным является разработка оптического приемника. Чтобы усилить и преобразовать слабые импульсы тока в разумные сигналы напряжения, требуется несколько усиливающих каскадов. Здесь возникают проблемы колебаний параметров процесса и окружающей среды, а следовательно, создаются условия для расфазировки. В результате возникают практически те же проблемы, что и при использовании обычного подхода. Впрочем, радует то, что проблема привязана только к приемникам, т.е. она более управляема.

Возможно, в будущем оптическое тактирование займет достойное место в высокоскоростных системах. Чтобы это стало реальностью, в первую очередь необходимо решить проблему колебаний параметров процесса в оптоэлектронных системах.

10.7.3. Синхронное или асинхронное проектирование

Теоретически автоматическая синхронизация позволяет решить проблему растущих сетей распределения тактовых сигналов. При принятии такого подхода глобальный тактовый сигнал переводится в несколько локальных, соответственно, и проблемы переходят на локальный уровень. Независимость от физических временны*х* условий достигается с помощью сигналов завершения. Для обеспечения логического упорядочения событий цепи и устранения состязаний требуется логика установления связи. Это, в свою очередь, требует реализации определенного протокола, который обычно состоит из двух или четырех фаз.

Несмотря на все свои достоинства, автосинхронизация используется лишь в нескольких ситуациях. В частности, это сфера обработки сигналов [Jacobs90], быстрые арифметические устройства (например, делители с внутренней синхронизацией) [Williams87], простые микропроцессоры [Martin89] и запоминающие устройства (статические ОЗУ, буферы FIFO). В общем случае синхронная логика и быстрее, и проще, поскольку в ней отсутствуют служебные издержки, связанные с генерацией сигналов завершения и реализацией логики установления связи. Разработка безопасной сети из блоков, поддерживающих протокол установления связи, которая была бы устойчивой к состязаниям и тупикам, является нетривиальной задачей, которая требует хороших специализированных средств автоматического проектирования.

С другой стороны, распределение тактового сигнала с высокой скоростью становится все более и более сложным. Управление расфазировкой требует сложного моделирования и аккуратного анализа, а также тщательного последующего проектирования. Описанную методологию совсем не просто расширить на следующее поколение разработок. С увеличением временн*о*й неопределенности в синхронных схемах, похоже, наиболее привлекательным становится подход автосинхронизации [Sutherland02]. Данное наблюдение уже подтверждается тем фактом, что трассировочная сеть для последнего поколения суперкомпьютеров с массовым параллелизмом полностью реализована с использованием автосинхронизации [Seitz92]. Но чтобы данный подход встал на поток (если это вообще когда-либо случится), необходимы инновационные разработки в сфере технологий передачи сигналов и организации цепей, а также методологии проектирования.

10.8. Резюме

В данной главе исследовались вопросы синхронизации и тактирования последовательных цифровых схем.

- Представлен подробный анализ синхронных цифровых схем и подходов к тактированию. Расфазировка тактовых импульсов и дрожание существенно влияют на функционирование и быстродействие системы. Важными параметрами цифровых схем являются выбранная схема тактирования и природа сетей генерации и распределения тактовых сигналов.

- С точки зрения борьбы с проблемами распределения тактовых сигналов привлекательны альтернативные подходы к тактированию, например автосинхронизация. В схемах с автосинхронизацией используются сигналы завершения и логика установления связи, что позволяет отделить физические условия на тактирование от упорядочения событий.
- Соединение синхронных и асинхронных компонентов вводит опасность сбоя синхронизации. Введение схем синхронизации позволяет уменьшить эту опасность, но не способно устранить ее полностью.
- Важным элементом в наборе инструментов разработчика цифровых схем становятся схемы фазовой автоподстройки частоты. Они используются для генерации высокочастотных тактовых сигналов в самом кристалле. Однако из-за аналоговой природы схем ФАПЧ их разработка является довольно сложной.
- Еще одной важной тенденцией в сфере разработки сетей распределения тактовых импульсов является использование систем автоподстройки по задержке, позволяющих активно настраивать задержки в кристалле.
- Основная мысль данной главы заключается в том, что синхронизация и тактирование являются одними из самых интересных и сложных задач, с которыми сталкиваются разработчики современных цифровых схем.

10.9. Для любознательных

Хотя тема временны*х* связей в системе является весьма важной, хорошего исчерпывающего справочника по ней до сих пор нет. Одним из лучших источников информации, по мнению авторов, является глава, написанная Чаком Сейцом (Chuck Seitz) для [Mead80]. Можно также порекомендовать обзоры, представленные в [Bakoglu90], [Johnson93], [Bernstein98] и [Chandrakasan00]. В [Friedman95] также представлен набор статей о цепях разводки тактовых сигналов. Кроме того, данной теме посвящено множество публикаций в профильных журналах; часть достойных внимания статей приведена ниже.

10.10. Список использованной литературы

[Abnous93] A. Abnous and A. Behzad, "A High-Performance Self-Timed CMOS Adder", in *EE241 Final Class Project Reports*, by J. Rabaey, Univ. of California-Berkeley, May 1993.

[Bailey98] D. Bailey and B. Benschneider, "Clocking Design and Analysis for a 600-MHz Alpha Microprocessor", *IEEE Journal of Solid-State Circuits*, November 1998, vol. 33, p. 1627–1633.

[Bailey00] D. Bailey, "Clock Distribution", in [Chandrakasan00].

[Bakoglu90] H. Bakoglu, Circuits, *Interconnections and Packaging for VLSI*, Addison-Wesley, 1980, p. 338–393.

[Boning00] D. Boning, S. Nassif, "Models of Process Variations in Device and Interconnect", in [Chandrakasan00].

[Bowhill95] W. Bowhill et al., "A 300 MHz Quad-Issue CMOS RISC Microprocessor", *Technical Digest of the 1995 International Solid State Circuits Conference*

(ISSCC), San Francisco, February 1995, p. 182-183.

[Bernstein98] K. Bernstein et al., *High Speed CMOS Design Styles*, Kluwer Academic Publishers, 1998.

[Bernstein00] K. Bernstein, "Basic Logic Families", in [Chandrakasan00].

[Chandrakasan00] A. Chandrakasan, W. Bowhill, and F. Fox, *Design of High-Performance Microprocessor Circuits*, IEEE Press, 2000.

[Chaney73] T. Chaney and F. Rosenberger, "Anomalous Behavior of Synchronizer and Arbiter Circuits", *IEEE Trans. on Computers*, April 1973, vol. C-22, p. 421-422.

[Dally98] W. Dally and J. Poulton, *Digital Systems Engineering*, Cambridge University Press, 1998.

[Dean94] M. Dean, D. Dill, and M. Horowitz, "Self-Timed Logic Using Current-Sensing Completion Detection", *Journal of VLSI Signal Processing*, February 1994, p. 7–16.

[Dobberpuhl92] D. Dopperpuhl et al., "A 200 MHz 64-b Dual Issue CMOS Microprocessor", *IEEE Journal on Solid State Circuits*, November 1992, vol. 27, no. 11, p. 1555–1567.

[Friedman95] E. Friedman, ed., *Clock Distribution Networks in VLSI Circuits and Systems*, IEEE Press, 1995.

[Gardner80] F. Gardner, "Charge-Pump Phase-Locked Loops", *IEEE Trans. on Communications*, November 1980, vol. COM-28, p. 1849–1858.

[Glasser85] L. Glasser and D. Dopperpuhl, *The Design and Analysis of VLSI Circuits*, Addison-Wesley, 1985, p. 360–365.

[Goodman98] J. Goodman, A. P. Dancy, A. P. Chandrakasan, "An Energy/Security Scalable Encryption Processor Using an Embedded Variable Voltage DC/DC Converter", *IEEE Journal of Solid State Circuits*, 1998, p. 1799–1809.

[Gutnik00] Gutnik, V., A. P. Chandrakasan, "Active GHz Clock Network Using Distributed PLLs", *IEEE Journal of Solid State Circuits*, November 2000, p. 1553.

[Heller84] L. Heller et al., "Cascade Voltage Switch Logic: A Differential CMOS Logic Family", *IEEE International Solid State Conference Digest*, San Francisco, February 1984, p. 16-17.

[Herrick00] B. Herrick, "Design Challenges in Multi-GHz Microprocessors", *Proceedings ASPDAC 2000*, Yokohama, January 2000.

[Jacobs90] G. Jacobs and R. Brodersen, "A Fully Asynchronous Digital Signal Processor", *IEEE Journal on Solid State Circuits*, December 1990, vol. 25, no. 6, p. 1526–1537.

[Jeong87] D. Jeong et al., "Design of PLL-Based Clock Generation Circuits", *IEEE Journal on Solid State Circuits*, April 1987, vol. SC-22, no. 2, p. 255–261.

[Johnson93] H. Johnson and M. Graham, *High-Speed Digital Design.A Handbook of Black Magic*, Prentice Hall, 1993.

[Kim90] B. Kim, D. Helman, and P. Gray, "A 30 MHz Hybrid Analog/Digital Clock Recovery Circuit in 2 μm CMOS", *IEEE Journal on Solid State Circuits*, December 1990, vol. SC-25, no. 6, p. 1385–1394.

[Kimerling00] L. Kimerling, "Photons to the Rescue: Microelectronics becomes Microphotonics", *Interface*, 2000, vol. 9, no. 2.

[Martin89] A. Martin et al., "The First Asynchronous Microprocessor: Test Results", *Computer Architecture News*, June 1989, vol. 17, no. 4, p. 95–110.

[Maneatis00] J. Maneatis, "Design of High-Speed CMOS PLLs and DLLs", in [Chandrakasan00].

[Mead80] C. Mead and L. Conway, *Introduction to VLSI Design*, Addison-Wesley, 1980.

[Messerschmitt90] D. Messerschmitt, "Synchronization in Digital Systems", *IEEE Journal on Selected Areas in Communications*, October 1990, vol. 8, no. 8, p. 1404–1419.

[Miller00] D. Miller, "Rationale and Challenges for Optical Interconnects to Electronic Chips", *Proceedings of the IEEE*, June 2000, p. 728–749.

[Park00] T. Park, T. Tugbawa, and D. Boning, "An Overview of Methods for Characterization of Pattern Dependencies in Copper CMP", *Chemical Mechanical Polish for ULSI Multilevel Interconnection Conference (CMP-MIC 2000)*, March 2000, p. 196–205.

[Restle98] P. Restle, K. Jenkins, A. Deutsch, P. Cook, "Measurement and Modeling of On-Chip Transmission Line Effects in a 400 MHz Microprocessor", *IEEE Journal of Solid State Circuits*, April 1998, p. 662–665.

[Restle01] P. Restle, "Technical Visualizations in VLSI Design", *Proceedings 38th Design Automation Conference*, Las Vegas, June 2001.

[Seitz80] C. Seitz, "System Timing", in [Mead80], 1980, p. 218–262.

[Seitz92] C. Seitz, "Mosaic C: An Experimental Fine-Grain Multicomputer", in *Future Tendencies in Computer Science*, *Control and Applied Mathematics*, Proceedings International Conference on the 25th Anniversary of INRIA, Springer-Verlag, 1992, p. 69–85.

[Sutherland89] I. Sutherland, "Micropipelines", *Communications of the ACM*, June 1989, p. 720–738.

[Sutherland02] I. Sutherland and J. Ebergen, "Computers Without Clocks", *Scientific American*, August 2002, vol. 287, no. 2, p. 62–69.

[Veendrick80] H. Veendrick, "The Behavior of Flip Flops Used as Synchronizers and Prediction of Their Failure Rates", *IEEE Journal of Solid State Circuits*, vol. SC-15, no. 2, April 1980, p. 169–176.

[Williams87] T. Williams et al., "A Self-Timed Chip for Division", in *Proceedings of Advanced Research in VLSI 1987*, Stanford Conference, March 1987, p. 75–96.

[Williams00] T. Williams, "Self-Timed Pipelines", in [Chandrakasan00].

[Yee96] G. Yee et al., "Clock-Delayed Domino for Adder and Combinational Logic Design", *Proceedings IEEE International Conference on Computer Design (ICCD)*, 1996, p. 332–337.

[Young97] I. Young, M. Mar, and B. Brushan, "A 0,35 μm CMOS 3-880 MHz PLL N/2 Clock Multiplier and Distribution Network with Low Jitter for Microprocessors", *IEEE International Solid State Conference Digest*, 1997, p. 330-331.

10.10.1. Упражнения

Последнюю версию сборника задач и упражнений можно найти по адресу `http://bwrc.eecs.berkley.edu/IcBook`.

Вставка

Ж

Верификация проекта

Моделирование и верификация

Электрическая и временная верификация

Формальная верификация

До этого момента мы очень полагались на моделирование как на основной метод обеспечения точности проектирования и нахождения критических параметров (например, скорости и рассеяние энергии). Однако следует помнить, что моделирование — это лишь один из инструментов разработчика. В общем случае полезно бывает различать *моделирование* и *верификацию*.

При *моделировании* значение параметра проекта (запас помехоустойчивости, задержка распространения или рассеиваемая энергия) определяется в выбранной модели путем подачи на вход цепи набора управляющих сигналов и извлечения параметров из получаемых сигналов. Данный подход очень гибкий, но обладает одним недостатком: его результаты зависят от выбора управляющих сигналов. Например, условие перераспределения заряда в динамическом логическом элементе нельзя обнаружить, не подав на вход именно ту последовательность сигналов, которая активизирует перераспределение заряда. Аналогично, задержка сумматора варьируется в широких пределах в зависимости от входного сигнала. Определение наихудшей ситуации также требует внимательного выбора управляющих сигналов. Другими словами, разработчик должен хорошо представлять внутреннюю структуру схемы и ее работу. Если это не так, моделирование даст бессмысленные результаты.

С другой стороны, при *верификации* разработчик пытается извлечь параметры системы непосредственно из описания схемы. Например, критический путь сумматора можно определить, изучив схему цепи или ее модель. Преимуществом данного подхода является то, что результат не зависит от выбора пробных сигналов и, предположительно, наиболее надежен. С другой сто-

роны, он основан на ряде неявных предположений, касающихся технологий и методологий проектирования. Возьмем, например, сумматор: определение задержки распространения требует понимания работы логических цепей, составляющих сумматор (например, какая это логика — динамическая или статическая), и определения термина *задержка распространения*. А нахождение минимальной тактовой частоты синхронной схемы требует определения элементов регистра. Таким образом, существующие инструменты верификации имеют ограниченную область применения и позволяют учесть только часть стилей разработки схем (например, рассмотреть класс однофазных синхронных схем). В данной вставке мы проанализируем данную тему подробнее.

Электрическая верификация

Имея транзисторную схему цифровой структуры, можно проверить (верифицировать), удовлетворяет ли она нескольким базовым правилам. Ниже приводится ряд примеров типичных правил, иллюстрирующих данную концепцию. Еще больше правил можно сформулировать, исходя из материала предыдущих глав.

- Число инверсий между двумя тактируемыми КМОП-элементами должно быть четным.
- В псевдо-n-МОП-элементе для обеспечения достаточной нижней границы запаса помехоустойчивости NM_L необходимо строго определенное соотношение размеров задающих p-МОП- и n-МОП-устройств.
- Чтобы удержать время нарастания и спада сигналов в определенных пределах, необходимо зафиксировать минимальные размеры задающих транзисторов как функцию разветвления по выходу.
- Максимальная величина перераспределения заряда в динамическом проекте должна согласовываться с верхней границей запаса помехоустойчивости NM_H.

Просто руководствуясь здравым смыслом, можно сформулировать множество правил, которым должен удовлетворять проект. Применение этих правил требует глубокого понимания структуры схемы. Следовательно, программа электрической верификации начинается с выделения известных подструктур в общей схеме. В качестве типичных шаблонов используются простейшие логические элементы, проходные транзисторы и регистры. Программа верификации обходит получающуюся в результате цепь, последовательно применяя все правила. Поскольку некоторые электрические правила применимы только к определенным стилям проектирования, они должны формулироваться так, чтобы их было просто модифицировать. Например, *экспертные системы на основе продукционных правил* позволяют легко обновлять базу правил [DeMan85]. Отдельные правила могут быть сложными и даже требовать применения программы моделирования схемы для небольших участков цепи. Короче говоря, электрическая верификация — это полезный инструмент, который существенно снижает опасность неверной работы схемы.

Временная верификация

Поскольку схемы постоянно становятся все более и более сложными, бывает проблематично точно определить, какие пути цепи критичны с точки зрения тактирования. Одним из решений данной проблемы является запуск подробных программ моделирования SPICE, что может потребовать довольно много времени. И даже при этом они не гарантируют, что определенный критический путь является наихудшим, поскольку задержка пути зависит от поданных на вход сигналов. Программа временной верификации обходит электрическую цепь и упорядочивает по рангам различные пути на основе их задержек. Задержка может определяться множеством способов. Например, можно создать RC-модель цепи и рассчитать граничные значения задержки получающейся пассивной цепи. Для получения более точных результатов многие программы вначале, основываясь на RC-модели, извлекают информацию о наиболее длинном пути и моделируют сокращенную схему, вычисляя более точную оценку. В качестве примеров ранних систем временной верификации можно привести системы Crystal [Ousterhout83] и TV [Jouppi84].

Проблемой многих ранних систем было определение *ложных путей*, т.е. критических путей, которые при нормальной работе схемы никогда не будут задействованы. Например, один такой путь существует в сумматоре, рассмотренном в главе 11 и в упрощенной форме показанном на рис. Ж.1. Анализируя только топологию схемы, можно решить, что критический путь данной цепи проходит через сумматор и мультиплексор, как показано стрелкой. Более внимательно изучив работу схемы, мы находим, что такой путь невозможен. Чтобы пройти через весь сумматор, все отдельные биты сумматора должны находиться в режиме прохождения и поступать на вход In. Однако в таком случае включается сигнал *обхода* и выбирается нижний путь через мультиплексор. Следовательно, реальный критический путь короче, чем можно подумать, проведя анализ в первом приближении. Вообще, выявлять ложные пути непросто, поскольку это требует понимания функционирования схемы на уровне логических элементов. Современные программы временной верификации более успешно справляются с этой проблемой и существенно облегчают жизнь разработчику высокоскоростных схем [Devadas91].

Рис. Ж.1. Пример ложного пути при временной верификации

Пример Ж.1. Пример временной верификации

На рис. Ж.2 показан результат выполнения программы статической временной верификации PathMill от Synopsys [PathMill]. На вход процесса верификации подается таблица соединений транзисторов, но можно использовать и модели на уровне логических элементов или блоков. При анализе учитываются емкостные и резистивные паразитные нагрузки, вычисленные по транзисторной схеме или топологии.

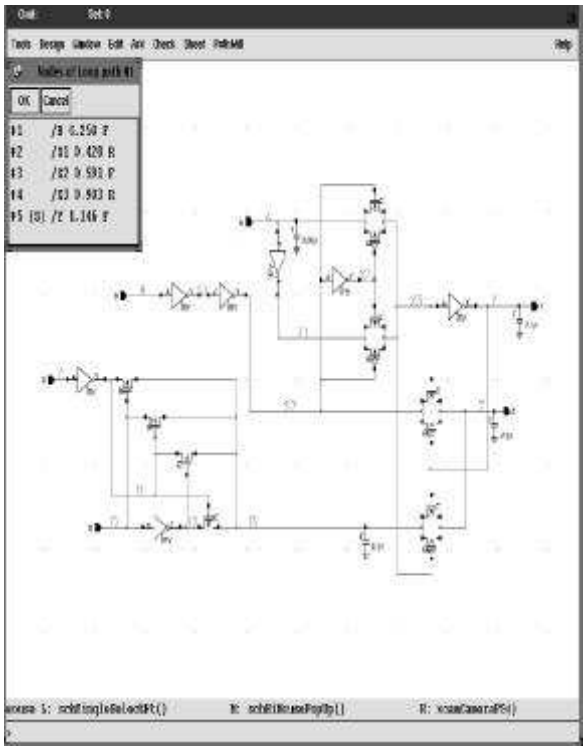


Рис. Ж.2. Результат выполнения программы статической временной верификации PathMill. (*Перепечатано с разрешения Synopsys, Inc.*) Для представления результатов использовалось средство *Cadence DFII*. Критический временной путь проходит от узла B к узлу Y через узлы $S1$, $S2$ и $X3$

Результатом временного анализа является упорядоченный список критических временных путей. В примере, рассмотренном на рис. Ж.2, наиболее длинный путь проходит от узла B к узлу Y через узлы $S1$, $S2$ и $X3$. Предсказываемая задержка равна 1,14 нс. Кроме того, к критическому близки следующие пути (упорядочены по уменьшению задержки): $B(F) \rightarrow Y(R)$ (1,11 нс), $B(R) \rightarrow Y(F)$ (1,04 нс), $C(R) \rightarrow Z(R)$ (1,04 нс) и $D(R) \rightarrow Y(R)$ (1,00 нс). Обратите внимание на то, что схема содержит несколько цепочек транзисторов, которые усложняют анализ и повышают шансы нахождения ложного пути. Например, ложным является путь $AF \rightarrow X1 \rightarrow X3 \rightarrow Z(R)$ для $B = 0$.

Функциональная (или формальная) верификация

Каждый компонент (транзистор, логический элемент или функциональный блок) схемы можно описать по его поведению как функцию входов и внутреннего состояния. Объединяя описания этих элементов, можно сгенерировать модель всей схемы, в символической форме описывающей поведение рассматриваемой структуры. При формальной верификации данное расчетное поведение сравнивается с исходной спецификацией разработчика. Хотя два данных описания и не идентичны, они должны быть математически эквивалентными.

Формальная верификация — это предел мечтаний разработчика, которого он желает добиться от средств автоматизированного проектирования: *доказательство, что схема будет работать именно так, как задано*. К сожалению, до настоящего момента не было создано ни одного универсального средства верификации, которое получило бы широкое распространение. Сложность данной проблемы иллюстрируется следующим аргументом: чтобы доказать, что два описания схемы идентичны, необходимо сравнить их выходы для всех возможных входных последовательностей сигналов. Проблема в том, что время требуемых для этого расчетов экспоненциально растет в зависимости от числа входов и состояний.

Тем не менее сказанное не означает, что формальная верификация относится к области фантазий. Для определенных классов схем соответствующие принципы верификации были предложены и успешно реализованы. Например, существенно сузить пространство поиска позволяет предположение о синхронности схемы. Одним из основных направлений исследований является доказательство эквивалентности между конечными автоматами; определенный прогресс в данной сфере уже наблюдается (см., например, [Coudert90]). В общем, формальная верификация делится на два основных класса.

- *Доказательство эквивалентности.* Проектирование часто следует по пути уточнения, когда исходное высокоуровневое описание последовательно уточняется до более подробного. Средства доказательства эквивалентности позволяют гарантировать, что получаемый результат функционально эквивалентен первоначальному описанию. Например, подобное средство может установить, что транзисторная схема статического КМОП-элемента НЕ-И с N входами действительно реализует задуманную функцию НЕ-И.
- *Доказательство свойств.* Разработчику часто нужно знать, сохраняются ли в процессе проектирования некоторые заданные свойства. Возможно, к разработке применяется следующее требование: “В схеме никогда не должны возникать состязания”. Добиться поставленной цели можно с помощью хорошего средства доказательства свойств; кроме того, средство может сообщить, почему условие не удовлетворяется.

Одним из необходимых условий реализации формальной верификации является описание проекта, не допускающее неоднозначности и имеющее четко определенное значение (или семантику). При выполнении этого условия техника формальной верификации весьма эффективна и полезна. Например,

областью, в которой формальная верификация достигла существенных успехов, являются конечные автоматы (Finite State Machine — FSM). Секрет этих успехов кроется в хорошо определенной математической модели описания.

Резюме

Хотя функциональная верификация еще не получила должного признания в сфере автоматизированного проектирования, она может стать одним из наиболее важных инструментов в арсенале разработчика цифровых систем. Чтобы этот прогноз подтвердился, необходима большая работа в сфере спецификации и интерпретации проектов.

Список использованной литературы

[Coudert90] O. Coudert and J. Madre, *Proceedings ICCAD 1990*, November 1990, p. 126–129.

[DeMan85] H. De Man, I. Bolsens, E. Vandenmeersch, and J. Van Cleynenbreugel, "DIALOG: An Expert Debugging System for MOS VLSI Design", *IEEE Trans. on CAD*, July 1985, vol. 4, no. 3, p. 301–311.

[Devadas91] S. Devadas, K. Keutzer, and S. Malik, "Delay Computations in Combinational Logic Circuits: Theory and Algorithms", *Proc. ICCAD-91*, Santa Clara, 1991, p. 176–179.

[Jouppi84] N. Jouppi, *Timing Verification and Performance Improvement of MOS VLSI Designs*, Ph. D. diss., Stanford University, 1984.

[Ousterhout83] J. Ousterhout, "Crystal: A Timing Analyzer for nMOS VLSI Circuits", *Proc. 3rd Caltech Conf. on VLSI* (Bryant ed.), Computer Science Press, March 1983, p. 57–69.

[PathMill] PathMill: Transistor-Level Static Timing Analysis, `http://www.synopsys.com/products/analysis/pathmill_ds.html`, Synopsys, Inc.

ГЛАВА

11

Проектирование арифметических блоков

Проектирование сумматоров, умножителей и сдвиговых регистров с акцентом на быстродействие, площадь или мощность

Оптимизация логики и системы для информационных каналов

Компромиссы между мощностью и задержкой

11.1. Введение
11.2. Информационные каналы в архитектурах цифровых процессоров
11.3. Сумматор
11.3.1. Двоичный сумматор: определения
11.3.2. Полный сумматор: вопросы проектирования схем
11.3.3. Двоичный сумматор: проектирование логики
11.4. Умножитель
11.4.1. Умножитель: определения
11.4.2. Генерация частичных произведений
11.4.3. Накопление частичных произведений
11.4.4. Конечное суммирование
11.5. Сдвиговый регистр
11.5.1. Многорегистровое циклическое сдвиговое устройство
11.5.2. Логарифмический сдвиговый регистр
11.6. Другие арифметические операторы

11.7. Компромиссы между мощностью и скоростью в информационных каналах*
11.7.1. Технологии уменьшения мощности (время проектирования)
11.7.2. Управление мощностью во время выполнения
11.7.3. Уменьшение мощности в холостом (или спящем) режиме
11.8. Перспектива: проектирование как поиск компромиссов
11.9. Резюме
11.10. Для любознательных

11.1. Введение

После подробного изучения тем проектирования и оптимизации стандартных цифровых элементов пришло время проверить полученные знания на более крупных объектах с перспективой перехода на системный уровеньпроектирования.

Технологии, рассмотренные в предыдущих главах, мы собираемся использовать для проектирования нескольких схем, часто используемых в информационных трактах микропроцессоров и процессоров цифровой обработки сигналов. Точнее, мы рассмотрим разработку характерного набора модулей — сумматоров, умножителей и сдвиговых регистров. Быстродействие и мощность данных элементов часто определяет быстродействие всей системы. Следовательно, необходима грамотная оптимизация проекта. Довольно быстро вам станет ясно, что прямолинейное проектирование подобных схем невозможно. Каждый модуль можно реализовать с помощью множества эквивалентных логических представлений и топологий, каждая из которых имеет свои плюсы и минусы с точки зрения площади, быстродействия и мощности.

Хотя наш анализ нельзя назвать полным, он поможет вам понять основные компромиссы, из которых необходимо выбирать при проектировании цифровой схемы. Мы покажем, что оптимизация всего на одном уровне (например, выбор правильных размеров транзисторов) приводит к плохим последствиям для проекта в целом. Следовательно, необходимо четко представлять глобальную картину. Разработчики цифровых схем фокусируют свое внимание на логических элементах, цепях или транзисторах, имеющих наиболее сильное влияние на целевые функции схемы. Некритические части схемы можно проектировать в плановом порядке, не утруждаясь особыми решениями. Кроме того, можно создавать модели производительности первого порядка, которые помогают лучше представить поведение модуля. Помимо этого мы обсудим, как упростить и автоматизировать некоторые фазы процесса проектирования с помощью компьютерных средств.

Прежде чем мы приступим к анализу проектирования арифметических модулей, стоит вкратце обсудить роль информационных каналов в цифровом процессоре. Это не только поможет выделить конкретные требования к проекту, касающиеся информационных каналов, но и позволит лучше представить перспективы развития других технологий, рассмотренных в данной

книге. Другие модули процессора (блок ввода-вывода, контроллер, запоминающие устройства) имеют другие требования, рассмотренные в главе 8. Проанализировав компромиссы между площадью, мощностью и задержкой, реализующиеся при проектировании сумматоров, умножителей и сдвиговых регистров, мы используем те же структуры, чтобы проиллюстрировать некоторые из введенных в главе 6 подходов к минимизации мощности. Глава завершается коротким обзором перспектив проектирования информационных каналов и сопутствующих компромиссов.

11.2. Информационные каналы в архитектурах цифровых процессоров

В главе 8 мы ввели концепцию цифрового процессора, состоящего из информационного канала, памяти, управляющего блока и блоков ввода-вывода. Центральным элементом процессора является информационный канал, в котором происходят все вычисления. Другие блоки процесса являются вспомогательными элементами, которые либо хранят результаты, сгенерированные информационным каналом, либо помогают определить, что произойдет в следующем такте. Типичный информационный канал состоит из соединенных между собой основных комбинаторных функций, реализованных арифметическими операторами (сложение, умножение, сравнение и сдвиг) или логическими операторами (И, ИЛИ и ИСКЛЮЧАЮЩЕЕ ИЛИ). В данной главе рассматривается проектирование арифметических операторов. В некоторых случаях (например, в персональных компьютерах) основным параметром цифровой схемы является скорость обработки. В большинстве других приложений важна максимальная энергия, которую можно рассеять, или минимальная энергия, с которой можно провести необходимые вычисления при поддержании требуемой пропускной способности.

Информационные каналы часто имеют *побитовую* организацию, как показано на рис. 11.1. Вместо обработки однобитовых цифровых сигналов данные на процессор подаются *словами*. В типичных микропроцессорах слова

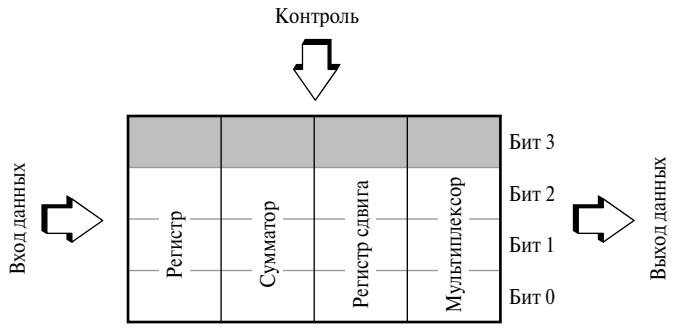


Рис. 11.1. Побитовая организация информационного канала

имеют размер 32 или 64 бита, а в специализированных каналах обработки сигналов (например, в DSL-модемах, накопителях на магнитных дисках или проигрывателях компакт-дисков) слова могут иметь произвольную длину, обычно от 5 до 24 бит. Например, 32-битовый процессор работает с 32-битовыми словами. Эти нюансы отражаются в структуре информационного канала. Поскольку довольно часто ко всем битам слова требуется применить одну и ту же операцию, информационный канал состоит из 32 разрядных секций, каждая из которых обрабатывает свой бит (отсюда и термин *побитовая* организация или обработка). Разрядные секции бывают идентичными или имеют сходную структуру для всех битов. Разработчик информационного канала может сконцентрировать на проектировании одной секции, а затем ввести в схему 32 ее копии.

11.3. Сумматор

Наиболее распространенной арифметической операцией является суммирование. Кроме того, сумматор также часто является основным элементом, ограничивающим быстродействие схемы. Следовательно, к разработке и оптимизации сумматора необходимо относиться очень внимательно. Эту оптимизацию можно выполнить либо на уровне логических элементов, либо на уровне схемы. Как правило, при оптимизации на уровне логики разработчики пытаются так записать булевы уравнения, чтобы схема получилась более быстрой или с минимальным размером. В качестве примера подобной оптимизации логики можно привести рассмотренный далее в этой главе *сумматор с ускоренным переносом*. С другой стороны, при оптимизации на уровне схемы размеры транзисторов и топология схемы выбираются так, чтобы получить максимальное быстродействие. Прежде чем мы рассмотрим оба типа оптимизации, остановимся вкратце на основных определениях схемы сумматора (в том виде, в котором их можно найти в любой книге по проектированию логических схем, например [Katz94]).

11.3.1. Двоичный сумматор: определения

В табл. 11.1 приведена таблица истинности двоичного полного сумматора, где A и B — входы сумматора; C_i — входной сигнал переноса; S — выходной сигнал суммы; C_o — выходной сигнал переноса. Булевы выражения для S и C_o представлены в уравнении (11.1):

$$\begin{aligned} S &= A \oplus B \oplus C_i = \\ &= A\bar{B}\bar{C}_i + \bar{A}B\bar{C}_i + \bar{A}\bar{B}C_i + ABC_i; \\ C_0 &= AB + BC_i + AC_i. \end{aligned} \tag{11.1}$$

С точки зрения реализации S и C_o часто полезно определить как функции некоторых промежуточных сигналов G (“сгенерировать”), D (“удалить”)

Таблица 11.1. Таблица истинности полного сумматора

A	B	C_i	S	C_0	Состояние сигнала переноса
0	0	0	0	0	Удалить
0	0	1	1	0	Удалить
0	1	0	1	0	Передать
0	1	1	0	1	Передать
1	0	0	1	0	Передать
1	0	1	0	1	Передать
1	1	0	0	1	Сгенерировать/передать
1	1	1	1	1	Сгенерировать/передать

и P ("передать")[1]. $G = 1$ ($D = 1$) гарантирует *генерацию* (*удаление*) разряда переноса в C_o независимо от C_i, а $P = 1$ гарантирует передачу в C_o входного сигнала переноса. Выражения для этих сигналов можно получить, просто изучив таблицу истинности:

$$\begin{aligned} G &= AB; \\ D &= \bar{A}\bar{B}; \\ P &= A \oplus B. \end{aligned} \tag{11.2}$$

Сигналы S и C_o мы можем переписать как функции P и G (или D):

$$\begin{aligned} C_0(G, P) &= G + PC_i; \\ S(G, P) &= P \oplus C_i. \end{aligned} \tag{11.3}$$

Обратите внимание на то, что G и P зависят только от A и B и не зависят от C_i. Аналогичным образом мы можем вывести выражения для $S(D, P)$и $C_o(D, P)$.

N-битовый сумматор можно построить, соединив каскадом N полных сумматоров (Full Adder — FA), соединяющих $C_{o,k-1}$ с $C_{i,k}$ для k от 1 до $N - 1$ и первый входной сигнал переноса $C_{i,0}$ с 0 (рис. 11.2). Данная конфигурация называется *сумматором со сквозным переносом* (ripple-carry adder), поскольку бит переноса "проносится" с одного каскада на другой. Задержка распространения сигнала через схему зависит от количества логических каскадов и сигналов, поданных на вход. Для одних входных сигналов перенос вообще не происходит, а для других бит переноса необходимо пронести от самого младшего до самого старшего разряда. Задержка распространения подобной структуры (также именуемой *критическим путем*) определяется как наихудшая задержка распространения для всех возможных наборов входных сигналов.

[1] Обратите внимание на то, что сигнал передачи иногда определяется как функция ИЛИ входов A и B; данное условие гарантирует, что входной сигнал переноса будет передан на выход и при $A = B = 1$. Если мы будем использовать данное определение, это будет оговариваться явно.

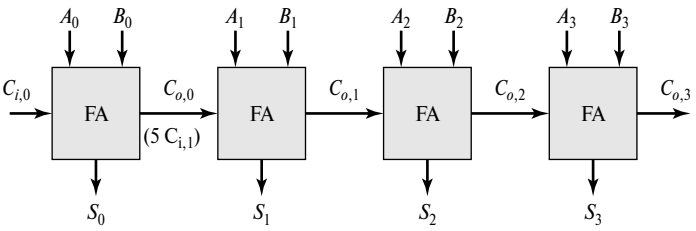


Рис. 11.2. Топология четырехбитового сумматора со сквозным переносом

Для сумматора со сквозным переносом наихудшая задержка наблюдается, когда сигнал переноса, сгенерированный в самом младшем разряде, проходит до самого старшего разряда. На последнем каскаде данный сигнал переноса поглощается и генерируется сумма. Таким образом, задержка распространения пропорциональна числу битов в входных словах N, и ее можно аппроксимировать следующим образом:

$$t_{adder} \approx (N-1)t_{carry} + t_{sum}, \tag{11.4}$$

где t_{carry} и t_{sum} равны задержкам распространения сигналов от C_i до C_o и S соответственно[2].

Пример 11.1. Задержка распространения сумматора со сквозным переносом

Выведем значения A_k и B_k $(k = 0 \ldots N-1)$, чтобы найти наихудшую задержку распространения сумматора со сквозным переносом.

Реализация наихудшего сценария требует, чтобы сигнал переноса генерировался в самом младшем разряде. Поскольку входной сигнал переноса первого полного сумматора C_{i0} всегда равен 0, то сигналы A_0 и B_0 должны равняться 1. Все остальные каскады должны находиться в режиме передачи. Следовательно, один из сигналов, A_i или B_i, должен иметь высокий уровень. Наконец, нам желательно физически измерить задержку перехода на старшем бите суммы. Предполагая, что изначально S_{N-1} равно 0, мы должны получить переход $0 \rightarrow 1$. Для этого A_{N-1} и B_{N-1} необходимо установить в значение 0 (или 1), что даст нам высокое значение бита суммы (при условии, что входной сигнал переноса равен 1).

Например, для 8-битового сумматора наихудший сценарий с точки зрения задержки распространения реализуется при следующих значениях A и B:

$$A : 0000001; \quad B : 01111111.$$

[2] В уравнении (11.4) предполагается, что задержки распространения входных сигналов A_0 (или B_0) до $C_{o,0}$ для самого младшего разряда и задержка распространения от C_i до C_o для всех остальных битов равны t_{carry}

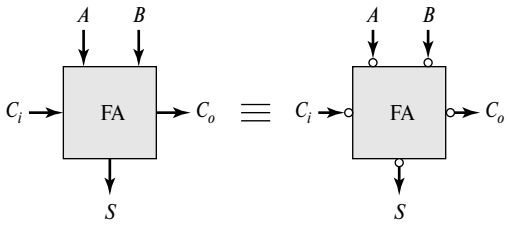


Рис. 11.3. Свойство инверсии полного сумматора. Кружками на схеме обозначены инверторы

Чтобы реализовать наихудший с точки зрения задержки переход, все входы могут быть постоянными, а сигнал A_0 — выполнять переход $0 \rightarrow 1$.

В данном двоичном представлении крайний слева бит является старшим разрядом. Обратите внимание на то, что мы рассмотрели лишь один из множества худших наборов условий. В данном случае в конечной сумме происходит переход $0 \rightarrow 1$. Приведите другие наихудшие ситуации, в которых реализуются переходы $0 \rightarrow 1$ и $1 \rightarrow 0$.

Из уравнения (11.4) можно сделать два важных вывода.

- Задержка распространения сумматора со сквозным переносом *линейно* пропорциональна N. Данное свойство особенно важно при проектировании сумматоров для широких информационных каналов ($N = 16 \ldots 128$), требуемых в современных и будущих компьютерах.
- Разрабатывая ячейку полного сумматора для быстрого сумматора со сквозным переносом, гораздо важнее оптимизировать t_{carry}, чем t_{sum}, поскольку последнее время лишь незначительно влияет на общее значение t_{adder}.

Прежде чем мы начнем подробное обсуждение темы проектирования ячеек полного сумматора, стоит упомянуть одно важное логическое свойство полного сумматора: *инвертирование всех входов полного сумматора приводит к инверсии значений всех выходов*. Данное свойство, также называемое *свойством инверсии*, выражается двумя уравнениями:

$$\begin{aligned} \bar{S}(A, B, C_i) &= S(\bar{A}, \bar{B}, \bar{C}_i); \\ \bar{C}_0(A, B, C_i) &= C_0(\bar{A}, \bar{B}, \bar{C}_i). \end{aligned} \tag{11.5}$$

Особенно оно полезно при оптимизации скорости сумматора со сквозным переносом. Из него следует, что схемы, приведенные на рис. 11.3, являются идентичными.

11.3.2. Полный сумматор: вопросы проектирования схем

Статический сумматор

Один из способов реализации схемы полного сумматора заключается в трансляции логических уравнений (11.1) непосредственно в КМОП-схему. Число требуемых транзисторов можно уменьшить с помощью определенных логических манипуляций. Например, полезным является совместное использование некоторых логических элементов подсхемами генерации сигналов суммы и переноса, если это не замедляет генерацию сигнала переноса (как отмечалось ранее, это наиболее критичный путь). Пример подобного переупорядочения членов уравнения приведен ниже.

$$C_0 = AB + BC_i + AC_i$$

и (11.6)

$$S = ABC_i + \bar{C}_0(A + B + C_i.)$$

Легко проверить, что данная пара уравнений эквивалентна исходной. Схема соответствующего сумматора, реализованного на КМОП, показана на рис. 11.4; она требует 28 транзисторов. Помимо того, что данная схема имеет слишком большую площадь, она является очень медленной.

- В схемах генерации сигнала переноса и суммы присутствуют высокие стеки p-МОП-транзисторов.

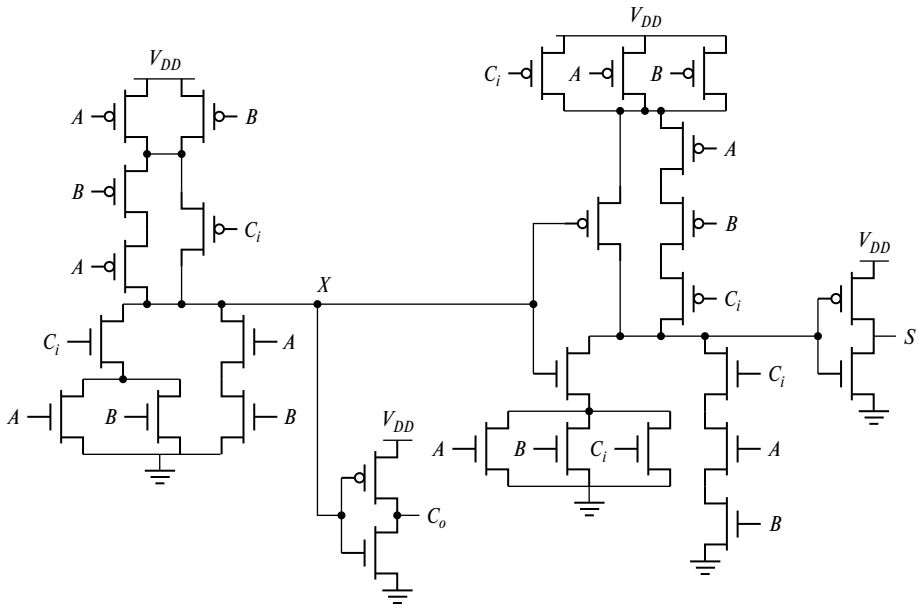


Рис. 11.4. Реализация полного сумматора на КМОП-схемах

- Внутренняя емкость нагрузки сигнала C_o велика и состоит из двух диффузионных емкостей, емкостей шести логических элементов и емкости проводки.
- Сигнал распространяется в схеме генерации сигнала переноса через два инвертирующих каскада. Как отмечалось ранее, разработчик быстродействующего сумматора должен обратить внимание именно на минимизацию задержки пути передачи сигнала переноса. При наличии на выходе цепочки переноса небольшой нагрузки (разветвления по выходу) два логических каскада — это слишком много (к тому же это приводит к дополнительному увеличению задержки).
- Генерация суммы требует одного дополнительного логического каскада, однако это не так важно, поскольку данный фактор всего лишь раз влияет на задержку распространения сумматора со сквозным переносом, представленного уравнением (11.4).

Данная схема хотя и является медленной, но в ней все же реализован ряд интересных технических решений. Обратите внимание на то, что первый элемент цепи генерации сигнала переноса спроектирован с сигналом C_i в меньшем p-МОП-стеке, что понижает логические трудозатраты до 2. Кроме того, n-МОП- и p-МОП-транзисторы, соединенные с C_i, расположены максимально близко к выходу вентиля. Такое решение является прямым применением техники оптимизации схемы, рассмотренной в разделе 4.2, — транзисторы на критическом пути должны располагаться максимально близко к выходу вентиля. Например, в каскаде k сумматора сигналы A_k и B_k доступны задолго до поступления сигнала $C_{i,k}$ $(= C_{o,k-1})$, проносимого через предыдущие каскады. Таким образом, емкости внутренних узлов цепочки транзисторов заблаговременно предварительно заряжаются или разряжаются. После поступления $C_{i,k}$ остается зарядить (разрядить) только емкость узла X. Размещение транзисторов $C_{i,k}$ близко к V_{DD} и GND потребует зарядки (разрядки) не только емкости узла X, но и внутренних емкостей.

Быстродействие данной схемы можно повысить, использовав некоторые свойства сумматора, рассмотренные в предыдущем разделе. Прежде всего, число инвертирующих каскадов на пути сигнала переноса можно уменьшить, использовав свойство инверсии, — инверсия всех входов полного сумматора инвертирует и все выходы. Как показано на рис. 11.5, данное правило позволяет убрать инвертор в цепочке передачи сигнала переноса.

Проектирование зеркального сумматора

На рис. 11.6 показана улучшенная схема сумматора — так называемый *зеркальный сумматор* (mirror adder) [Weste93], действие которого основано на уравнении (11.3). Цепь генерации сигнала переноса стоит проанализировать отдельно. Во-первых, как предлагалось в предыдущем разделе, в цепи отсутствует элемент инвертирования сигнала переноса. Во-вторых, сети PDN и PUN не являются дуальными, а формируют интересную реализацию функции распространения/генерации/удаления, — когда сигнал D либо G имеет

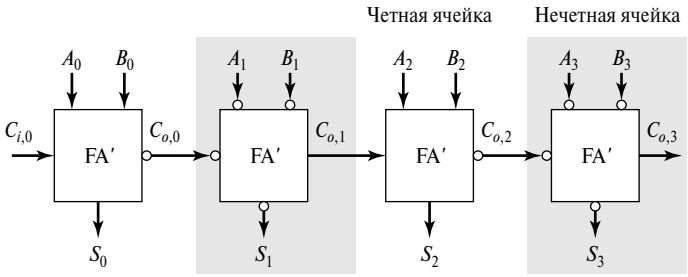


Рис. 11.5. Удаление инвертора с пути сигнала переноса. Под FA′ подразумевается полный сумматор без инвертора на пути сигнала переноса

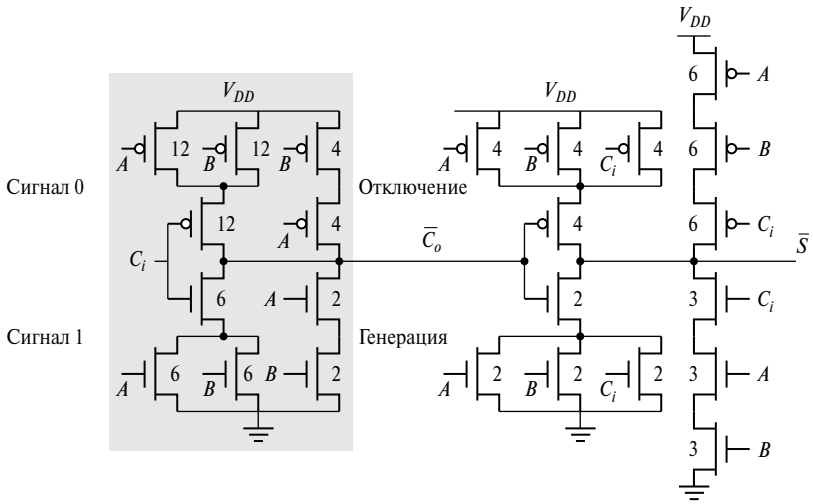


Рис. 11.6. Зеркальный сумматор

высокий уровень, сигнал $\bar{C}_o$ устанавливается равным V_{DD} или GND соответственно. При выполнении условия генерации сигнала распространения (или когда $P = 1$)[3] входной сигнал распространяется (в инвертированном виде) в $\bar{C}_o$. В результате существенно уменьшается задержка и используемая площадь. Проанализировать схему суммирования читателю предлагается самостоятельно. В заключение отметим несколько моментов.

- Данная ячейка полного сумматора требует всего 24 транзистора.
- Цепочки n-МОП- и p-МОП-транзисторов являются полностью симметричными, что не мешает правильной работе из-за *самодуальности* функций суммирования и переноса. В результате схема генерации сигнала переноса содержит максимум два последовательно включенных транзистора.

[3] В данном случае сигнал распространения определяется как $P = A + B$.

- Транзисторы, соединенные с C_i, располагаются близко к выходу вентиля.
- Оптимизацию скорости необходимо выполнять только для транзисторов каскада переноса. Все транзисторы каскада суммы могут иметь минимальный размер. При проектировании топологии ячейки наиболее критическим вопросом является минимизация емкости на узле $\bar{C}_o$.
- В сумматоре, показанном на рис. 11.4, инвертор можно масштабировать независимо, исходя из характеристик следующего за ним входа C_i каскада суммирования. Если схема переноса на рис. 11.6 симметрична, то все входы имеют логические затраты, равные 2. Это означает, что оптимальное разветвление по выходу, масштабированное для получения минимальной задержки, должно быть равно $(4/2) = 2$. Однако выход данного каскада питает две внутренние емкости и шесть емкостей логических элементов в связанной с ним ячейке суммирования. Правильным решением возникшей проблемы, позволяющим использовать транзисторы одинакового размера во всех каскадах, является увеличение размера каскада переноса в три-четыре раза по сравнению с каскадом суммирования. Это позволит сохранить оптимальное разветвление по выходу, равное 2. Получающиеся в результате размеры транзисторов обозначены на рис. 11.6, где предполагается, что отношение размеров p-МОП-транзисторов к n-МОП-транзисторам равно 2.

Сумматор на передаточных элементах

Полный сумматор можно спроектировать так, чтобы он мог использовать мультиплексоры и схемы ИСКЛЮЧАЮЩЕЕ ИЛИ. Хотя в КМОП-реализации это решение непрактично, оно становится гораздо привлекательнее при реализации мультиплексоров и схем ИСКЛЮЧАЮЩЕЕ ИЛИ как передаточных элементов. Полный сумматор, реализованный на основе данного подхода, показан на рис. 11.7 (использовано 24 транзистора). Данное устройство основано на модели *распространения/генерации*, введенной в уравнении (11.3). Сигнал распространения, представляющий собой результат применения операции ИСКЛЮЧАЮЩЕЕ ИЛИ к входам A и B, используется для выбора обычного или комплементарного значения входного сигнала переноса в качестве новой выходной суммы. На основе сигнала распространения выходной сигнал переноса устанавливается равным либо входному сигналу переноса, либо одному из входов A или B. Одна из интересных особенностей подобного сумматора заключается в том, что он имеет похожие задержки выходов каскадов суммирования и переноса.

Сумматор с манчестерской (дифференциальной) цепочкой переноса

Схему распространения сигнала переноса (см. рис. 11.7) можно упростить, добавив сигналы генерации и удаления, как показано на рис. 11.8, *а*. Путь распространения не меняется, он по-прежнему передает сигнал C_i на выход C_o, когда сигнал распространения $(A_i \oplus B_i)$ имеет значение True. Если условие распространения не удовлетворяется, выход либо опускается на низкий логический уровень с помощью сигнала D_i, либо поднимается на высокий уровень

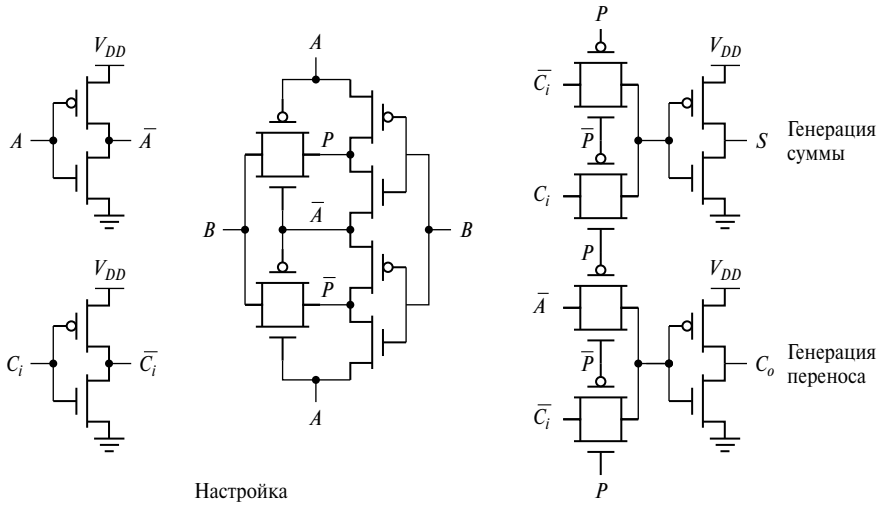


Рис. 11.7. Полный сумматор на передаточных элементах, задержки цепей суммирования и переноса которого приблизительно равны (согласно [Weste93])

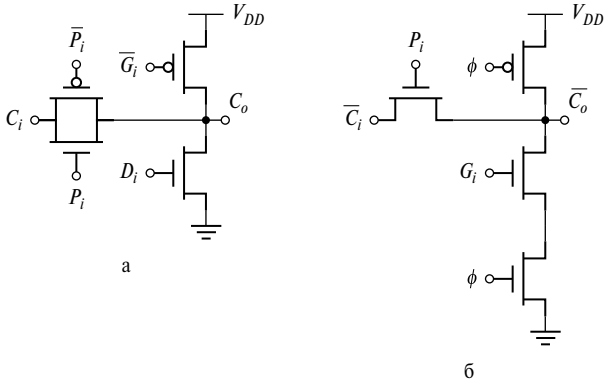


Рис. 11.8. Манчестерские (дифференциальные) элементы переноса: а — статическая реализация, использующая сигналы распространения, генерации и остановки; б — динамическая реализация, использующая только сигналы распространения и генерации

с помощью сигнала $\bar{G}_i$. Динамическая реализация (см. рис. 11.8, *б*) позволяет еще больше упростить схему. Поскольку переходы в динамических схемах являются монотонными, передаточные элементы можно заменить проходными n-МОП-транзисторами. Используя предварительную зарядку выхода, мы избавляемся от сигнала остановки (для случая, когда в цепочке переноса распространяются комплементарные значения сигналов переноса).

В сумматоре с дифференциальной цепочкой переноса данная цепочка реализуется с помощью каскада проходных транзисторов [Kilburn60]. На

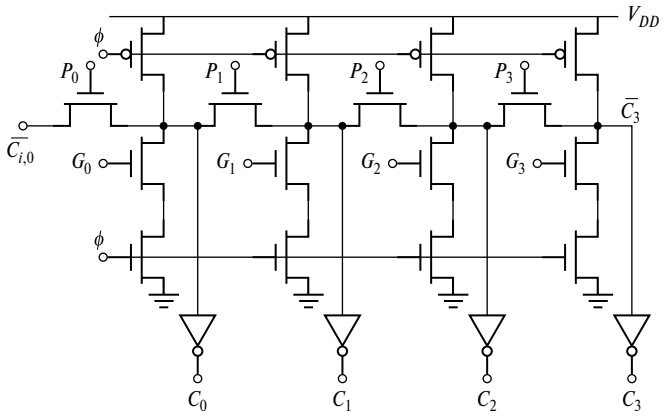


Рис. 11.9. Сумматор с дифференциальной цепочкой переноса, реализованный на динамической логике (четырехбитовая секция)

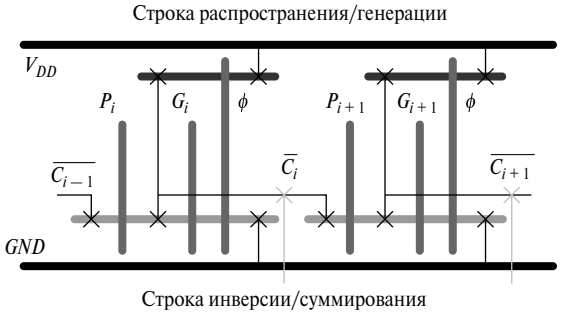


Рис. 11.10. Штриховая диаграмма двух битов дифференциальной цепочки переноса

рис. 11.9 показан пример схемы, относящейся к данному типу и построенной на динамических цепях. В фазе предварительной зарядки ($\phi = 0$) все промежуточные узлы цепочки переноса на проходных транзисторах предварительно заряжаются до V_{DD}. В фазе оценки узел A_k разряжается при наличии входного сигнала переноса и сигнала распространения P_k в высоком состоянии или при наличии в высоком состоянии сигнала генерации для каскада k (G_k).

На рис. 11.10 приведен пример топологии дифференциальной цепочки переноса, представленной в виде штриховой диаграммы. Топология информационного канала состоит из трех строк ячеек с побитовой организацией: в верхней строке ячеек вычисляются сигналы распространения и генерации, в средней — слева направо распространяется сигнал переноса, а в нижней — генерируются конечные суммы.

Наихудшая задержка цепочки переноса сумматора, показанного на рис. 11.9, моделируется линеаризованной RC-цепью, показанной на рис. 11.11.

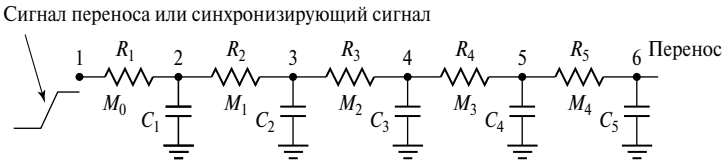


Рис. 11.11. Эквивалентная схема, используемая для определения задержки распространения цепочки переноса

Как отмечалось в главе 4, задержка распространения подобной цепи равна:

$$t_p = 0,69 \sum_{i=1}^{N} C_i \left(\sum_{j=1}^{i} R_j \right) = 0,69 \frac{N(N+1)}{2} RC, \tag{11.7}$$

где все $C_i = C$ и $R_j = R$.

Пример 11.2. Масштабирование дифференциальной цепочки переноса

Емкость на один узел цепочки переноса равна сумме четырех диффузионных емкостей, одной входной емкости инвертора и емкости монтажа, пропорциональной размеру ячейки. Инвертор и p-МОП-транзистор предварительной зарядки могут иметь единичный размер. Вместе с емкостью проводки фиксированную емкость можно оценить как 15 фФ (для нашей технологии). Если транзистор единичного размера с шириной W_0 имеет сопротивление 10 кОм и диффузионную емкость 2 фФ, то RC-постоянная времени для цепочки транзисторов шириной W запишется следующим образом:

$$RC = \left(6 \text{ фФ} \cdot \frac{W}{W_0} + 15 \text{ фФ} \right) \cdot 10 \text{ кОм} \cdot \frac{W_0}{W}.$$

Увеличение ширины транзистора уменьшает данную постоянную времени, но при этом больше нагружает логические элементы предыдущего каскада. Следовательно, размер транзистора ограничивается емкостью входной нагрузки.

К сожалению, поскольку цепочка переноса сходна по природе с распределенной RC-цепью, ее задержка распространения квадратично зависит от числа битов N. Чтобы избежать этого, в схему необходимо вводить инверторы, буферизирующие сигнал. Оптимальное количество каскадов на буфер зависит от эквивалентного сопротивления инвертора, а также от сопротивления и емкости проходных транзисторов (см. главу 9). В нашей технологии (и в большинстве практических ситуаций) данное число лежит между 3 и 4. Благодаря добавлению инвертора общая задержка распространения становится линейной функцией N, как для сумматоров со сквозным переносом.

11.3.3. Двоичный сумматор: проектирование логики

Единственным практическим устройством для реализации суммирования с относительно небольшой длиной слов является сумматор со сквозным переносом. В большинстве настольных компьютеров используется длина слова 32 бита, серверы обычно требуют 64 бит; очень быстрые компьютеры (мэйнфреймы, суперкомпьютеры или мультимедийные процессоры, подобные Sony PlayStation2 [Suzuoki99]) требуют, чтобы слова имели длину до 128 бит. Поскольку скорость сумматора линейно зависит от числа битов, использование сумматоров со сквозным переносом становится непрактичным. Следовательно, необходима оптимизация логики, которая бы позволила получить сумматоры с $t_p < O(N)$. Ниже мы кратко рассмотрим ряд соответствующих технологий. Основное внимание будет уделено использованию данных технологий в проектировании цифровых схем, поскольку большинство из представленных структур хорошо известны в традиционной литературе по проектированию логических схем.

Сумматор с обводом переноса

Рассмотрим четырехбитовый сумматор, изображенный на рис. 11.12, *а*. Предположим, что значения A_k и B_k $(k = 0 \ldots 3)$ выбраны так, чтобы все сигналы распространения P_k $(k = 0 \ldots 3)$ имели высокий уровень. При таких условиях входной сигнал переноса $C_{i,0} = 1$ распространяется через всю цепочку сумматора и приводит к тому, что выходной сигнал переноса $C_{o,3} = 1$. Другими словами,

$$\text{если } (P_0P_1P_2P_3 = 1), \text{ то } C_{0,3} = C_{i,0}, \qquad (11.8)$$
$$\text{иначе имеет место сигнал DELITE или GENERATE.}$$

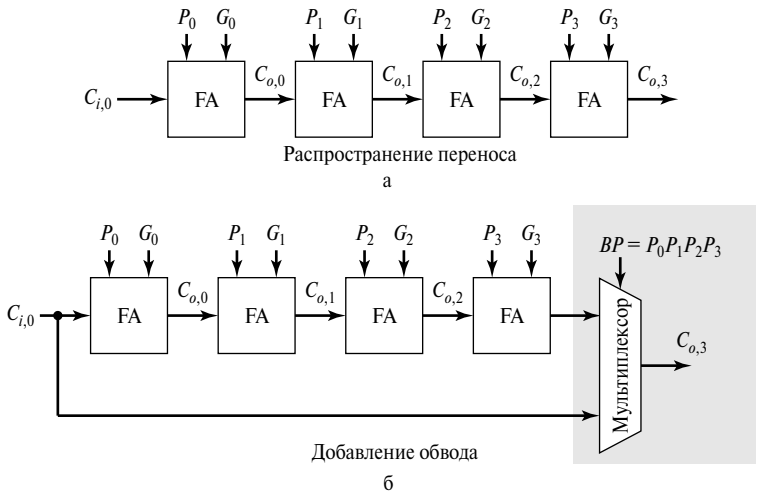


Рис. 11.12. Структура обвода переноса (основной принцип)

Используя данную информацию, можно ускорить работу сумматора, как показано на рис. 11.12, *б*. Когда $BP = P_0P_1P_2P_3 = 1$, входной сигнал переноса перенаправляется непосредственно на следующий блок с помощью обводного транзистора M_b; отсюда и название — *сумматор с обводом переноса* (carry-bypass adder) или *сумматор с пропуском переноса* (carry-skip adder) [Lehman62]. Если указанное условие не выполняется, сигнал переноса проходит по обычному маршруту.

Пример 11.3. Обвод сигнала переноса в сумматоре с дифференциальной цепочкой переноса

На рис. 11.13 показаны возможные пути распространения сигнала переноса для полного сумматора, реализованного по дифференциальному принципу. Видно, что обвод ускоряет операцию сложения: сигнал переноса либо распространяется через обводной путь, либо генерируется где-то в цепи. В обоих случаях задержка получается меньше, чем в обычной конфигурации со сквозным переносом. Дополнительная площадь, требуемая на реализацию обвода, мала и обычно составляет 10–20% от общей площади схемы. Тем не менее добавление обвода нарушает правильную побитовую структуру сумматора (см. рис. 11.10).

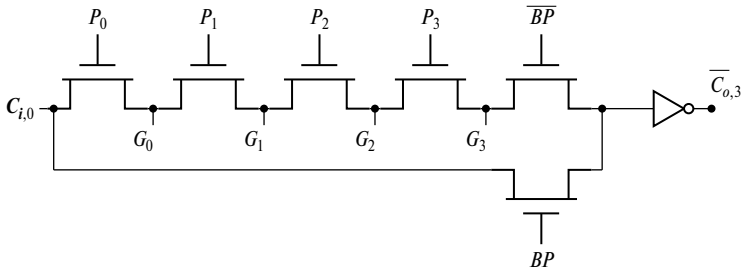


Рис. 11.13. Реализация обводного сумматора с дифференциальной цепочкой переноса

Далее мы рассчитаем задержку N-битового сумматора. Прежде всего предположим, что весь сумматор разбит на (N/M) каскадов равной длины, каждый из которых содержит M бит. Используя рис. 11.14, *а* и уравнение (11.9), можно вывести приблизительную формулу для общего времени распространения:

$$t_p = t_{setup} + Mt_{carry} + \left(\frac{N}{M} - 1\right) t_{bypass} + (M-1)t_{carry} + t_{sum}. \qquad (11.9)$$

В данное уравнение входят следующие параметры:

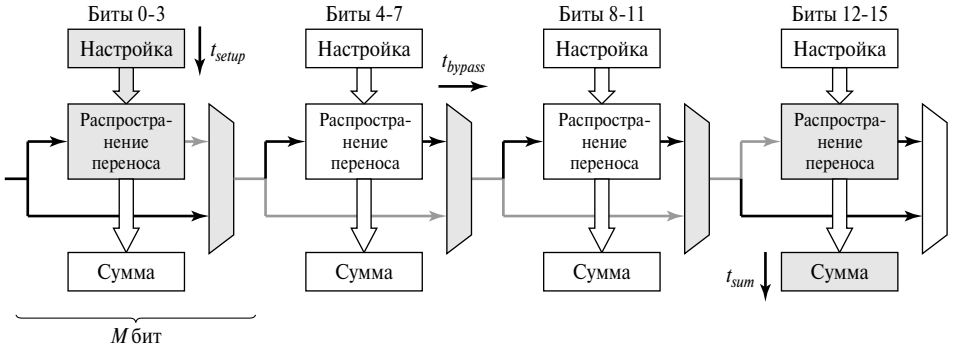


Рис. 11.14. Структура сумматора с обводом сигнала переноса ($N = 16$). Наихудший с точки зрения задержки путь обозначен серым цветом

- t_{setup} — фиксированное время, необходимое для создания сигналов генерации и распространения;
- t_{carry} — задержка распространения сигнала, приходящаяся на один бит; наихудшая задержка распространения сигнала переноса через один M-битовый каскад примерно в M раз больше t_{carry};
- t_{bypass} — задержка распространения через обводной мультиплексор одного каскада;
- t_{sum} — время, требуемое для генерации суммы на последнем каскаде.

На блок-схеме, приведенной на рис. 11.14, критический путь обозначен серым цветом. Из уравнения (11.9) следует, что t_p по-прежнему линейно зависит от числа битов N, поскольку в худшем случае сигнал переноса генерируется в первом разряде, проходит через первый блок, пропускает $((N/M-2))$ каскада с помощью обводного пути и поглощается в последнем разряде, не генерируя выходной сигнал переноса. Оптимальное число битов на блок, который можно пропустить, определяется технологическими параметрами (дополнительной задержкой мультиплексора, выбирающего обводной путь, требованиями к буферизации в цепочке переноса и отношением задержек на сквозном и обводном пути).

Хотя зависимость задержки от числа битов N по-прежнему является линейной, как показано на рис. 11.15, наклон функции задержки является более приемлемым, чем для сумматора со сквозным переносом. Для больших сумматоров данная разница может быть значительной. Обратите также внимание на то, что для небольших значений N сумматор со сквозным переносом в действительности является более быстрым, поскольку в этом случае служебные издержки на введение мультиплексоров себя не окупают. Точка пересечения двух функций зависит от выбранной технологии и обычно находится между четырьмя и восемью битами.

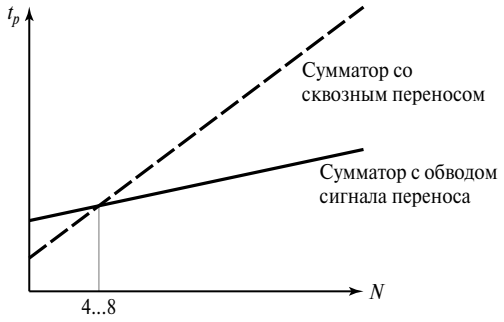


Рис. 11.15. Задержка распространения сумматоров со сквозным переносом и с обводом переноса

Задача 11.1. Задержка сумматора с пропуском переноса

Определите входной сигнал, реализующий наихудшую задержку в 16-битовом (4×4) сумматоре с обводом переноса. Предполагая, что $t_{carry} = t_{setup} = t_{skip} = t_{sum} = 1$, определите задержку и сравните ее с задержкой обычного сумматора со сквозным переносом.

Линейный сумматор с выбором сигнала переноса

В сумматоре со сквозным переносом любая ячейка полного сумматора должна ожидать поступления входного сигнала, а лишь затем генерировать выходной сигнал переноса. Чтобы нарушить эту линейную зависимость, можно заранее рассчитать результат для обоих возможных значений входного сигнала переноса. Как только станет известно фактическое значение входного сигнала, правильный результат быстро выбирается с помощью простого мультиплексорного каскада. Реализация данной идеи, получившей название *сумматора с выбором сигнала переноса* (carry-select adder) [Bedrij62], показана на рис. 11.16. Рассмотрим блок сумматоров, который складывает биты от k до $k+3$. Вместо того чтобы ожидать поступления выходного сигнала переноса бита $k-1$, анализируются оба его возможных значения — 0 и 1. С точки зрения схемы это означает реализацию двух путей сигнала переноса. Как только значение сигнала $C_{o,k-1}$ окончательно установится, мультиплексор выбирает результат, соответствующий пути 0 или 1, что происходит с минимальной задержкой. Как видно на рис. 11.16, дополнительные аппаратные издержки реализации сумматора с выбором сигнала переноса ограничиваются дополнительным путем передачи сигнала переноса и мультиплексором и составляют порядка 30% относительно структуры со сквозным переносом.

Как показано на рис. 11.17, полный сумматор с выбором сигнала переноса формируется путем последовательного соединения нескольких суммирующих каскадов равной длины, как в рассмотренном выше сумматоре с обводом переноса. Критический путь показан на рисунке серым цветом. Изучая схему,

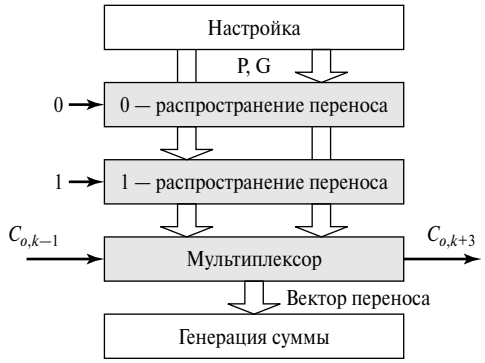


Рис. 11.16. Топология четырехбитового модуля выбора сигнала переноса

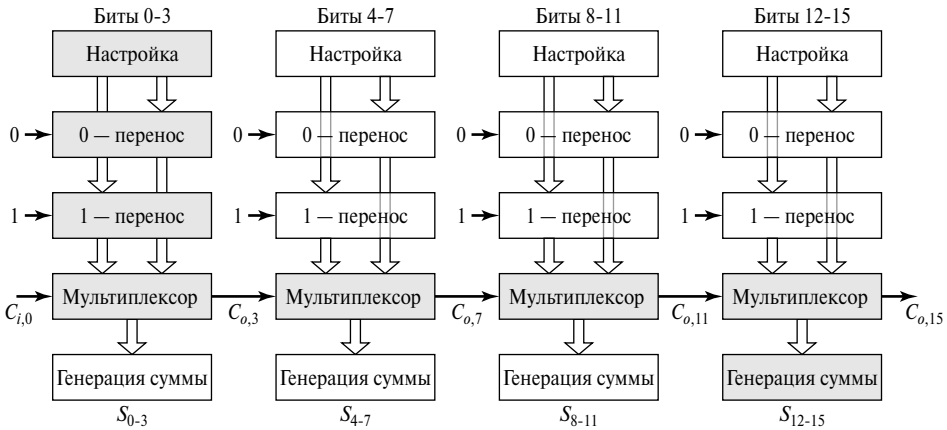


Рис. 11.17. Шестнадцатибитовый линейный сумматор с выбором сигнала переноса. Критический путь обозначен серым цветом

можно вывести модель первого порядка для наихудшей задержки распространения модуля:

$$t_{add} = t_{setup} + Mt_{carry} + \left(\frac{N}{M}\right) t_{mux} + t_{sum}, \tag{11.10}$$

где t_{setup}, t_{sum} и t_{mux} — фиксированные задержки; N и M представляют общее число битов и число битов на каскад соответственно; t_{carry} — задержка сигнала переноса при прохождении через ячейку полного сумматора. Задержка сигнала переноса при прохождении через один блок пропорциональна длине этого каскада или равна Mt_{carry}.

Как и ранее, задержка распространения сумматора линейно пропорциональна N (уравнение (11.10)). Причина такой линейной зависимости заключается в том, что в худшем случае сигнал *выбора блока*, разделяющий

решения, которые соответствуют нулевому и единичному значению сигнала переноса, по-прежнему нужно передавать через все каскады.

Задача 11.2. Задержка линейного сумматора с выбором сигнала переноса

Определите задержку 16-битового линейного сумматора с выбором сигнала переноса, полагая, что все ячейки имеют единичную задержку. Сравните результат с ответом задачи 11.1. Сравните результаты для различных конфигураций блоков.

Сумматор с выбором сигнала переноса (характеристика типа квадратного корня)

На примере следующей структуры показано, как опытный разработчик может повлиять на результаты проектирования. Для оптимизации структуры вначале жизненно необходимо определить критический временной путь. Рассмотрим 16-битовый сумматор с линейный выбором сигнала переноса. Для упрощения рассуждений предположим, что ячейки полных сумматоров и мультиплексоров имеют идентичные задержки распространения, равные нормированному значению 1. На рис. 11.18, *а* показаны наихудшие времена прибытия сигналов на различные узлы сети относительно времени подачи сигнала на вход. Из приведенного анализа следует, что критический путь сумматора последовательно проходит через сеть мультиплексоров последующих каскадов.

В такой ситуации существует одна замечательная возможность. Рассмотрим мультиплексор в последнем каскаде сумматора. Входами этого мультиплексора являются две цепочки сигналов переноса и сигнал мультиплексора предыдущего каскада. Мы видим существенное расхождение между временем прибытия сигналов. Результаты цепочек переноса устанавливаются задолго до поступления сигнала мультиплексора. Следовательно, имеет смысл выровнять задержки распространения по обоим путям. Это реально, если каждый последующий каскад сумматора будет иметь больше битов, чем предыдущий (тогда генерация сигналов переноса будет требовать больше времени). Например, первый каскад может суммировать 2 бита, второй — 3, третий — 4 и т.д. (см. рис. 11.18, *б*). Из обозначенного времени прибытия сигналов видно, что сумматор с такой топологией быстрее, чем сумматор с линейной структурой. Фактически та же самая задержка распространения получается и для 20-битового сумматора. Отметим, что при описанном подходе расхождение во времени поступления сигналов на мультиплексор устраняется.

По сути, простой трюк (последовательное увеличение размеров каскадов) приводит к тому, что задержка сумматора теперь выражается сублинейной функцией. Сказанное можно доказать с помощью следующих рассуждений. Предположим, что N-битовый сумматор содержит P каскадов и первый каскад суммирует M бит. Каждый последующий каскад суммирует

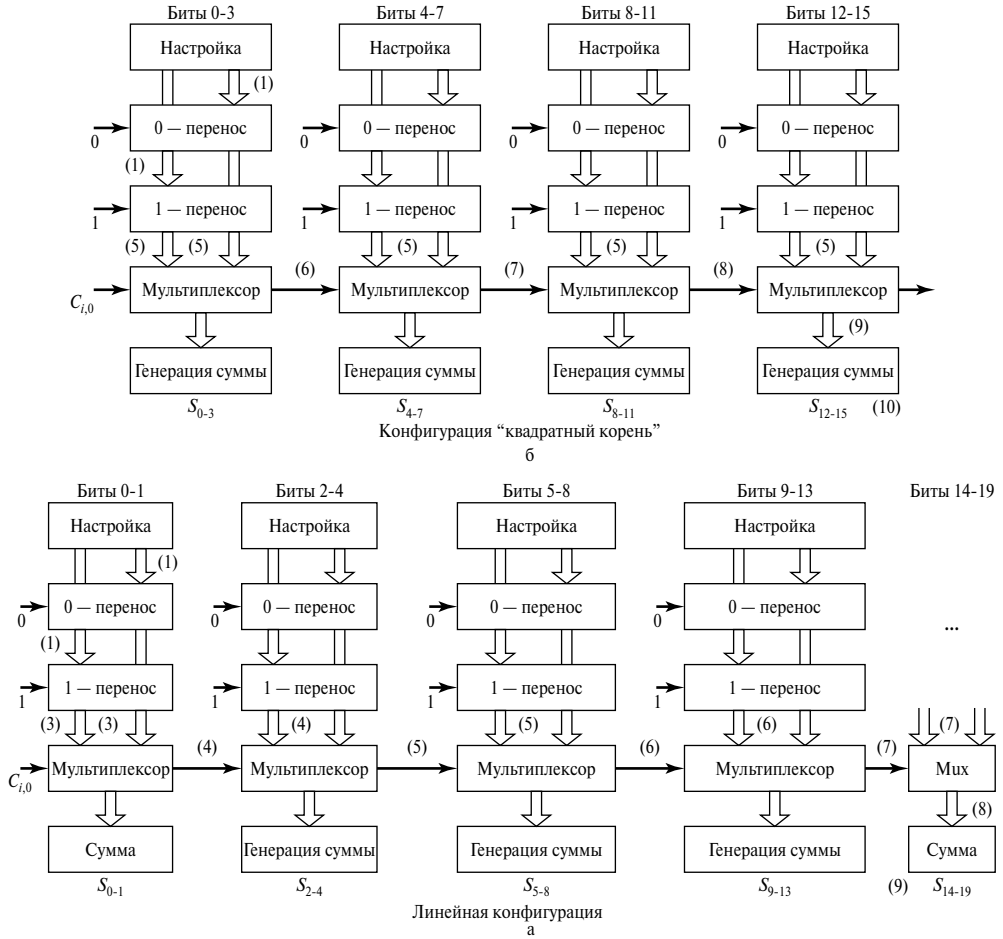


Рис. 11.18. Сумматоры с выбором сигнала переноса. В скобках указаны наихудшие времена прибытия сигналов

на один бит больше, чем предыдущий. В таком случае справедливо следующее соотношение:

$$N = M + (M+1) + (M+2) + (M+3) + \ldots + (M+P-1) = \\ = MP + \frac{P(P-1)}{2} = \frac{P^2}{2} + P\left(M - \frac{1}{2}\right). \tag{11.11}$$

Если $M << N$ (например, $M = 2$ и $N = 64$), то основной вклад в результат дает первое слагаемое, и уравнение (11.11) можно упростить до следующего вида:

$$N \approx \frac{P^2}{2} \tag{11.12}$$

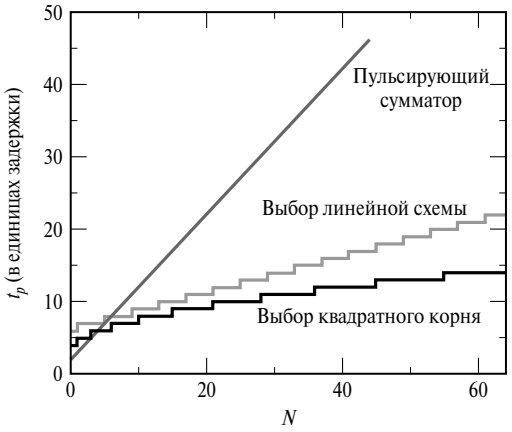


Рис. 11.19. Задержка распространения сумматора с выбором сигнала переноса (характеристика типа квадратного корня) в сравнении с линейными характеристиками сумматоров со сквозным переносом и выбором сигнала переноса. Для моделирования задержек ячеек применялась единичная модель задержки

или

$$P \approx \sqrt{2N}. \tag{11.13}$$

Переписав уравнение (11.10) с использованием формулы (11.13), t_{add} можно выразить как функцию N:

$$t_{add} = t_{setup} + Mt_{carry} + (\sqrt{2N})t_{mux} + t_{sum}. \tag{11.14}$$

Для больших сумматоров ($N >> M$) задержка пропорциональна $\sqrt{N}$, или $t_{add} = O(\sqrt{N})$. Для сравнения на рис. 11.19 показана зависимость от N задержек трех рассмотренных типов сумматоров — со сквозным переносом, с линейным выбором переноса и выбором переноса с характеристикой типа квадратного корня. Видно, что для больших значений N t_{add} становится практически постоянной величиной.

Задача 11.3. Неравные группы в сумматоре с обводом переноса

Возможно, вас интересует вопрос, а нельзя ли применить описанную выше технику к сумматорам с обводом сигнала переноса? Ранее мы показывали, что задержка таких сумматоров является линейной функцией числа битов. Можно ли, изменив размер групп, модифицировать этот сумматор так, чтобы получить лучшую задержку?

Кажется, что в данном случае также можно сделать группы последовательно увеличивающегося размера. Однако технику, использованную в сум-

маторах с выбором сигнала переноса, нельзя применить к данной ситуации, и последовательное увеличение размеров каскадов на самом деле увеличивает задержку. Рассмотрим сумматор с обводом сигнала переноса, в котором последний каскад является самым большим: сигнал переноса, распространяющийся через этот каскад и поглощаемый в старшем разряде (без возможности обвода), находится на критическом для генерации суммы пути. Увеличение размера последней группы проблему не решает.

Отталкиваясь от сказанного и предполагая, что элементы переноса и обвода имеют одинаковые задержки, изобразите цепь обвода сигнала переноса, дающую зависимость задержки, лучшую линейной.

Сумматор с ускоренным переносом*

Интегральный сумматор с ускоренным переносом

При проектировании еще более быстрых сумматоров необходимо както побороть эффект сквозного распространения сигнала переноса, который в том или ином виде присутствует и в сумматорах с обводом сигнала переноса, и в сумматорах с выбором сигнала переноса. Одним из возможных решений этой проблемы является использование принципа *упреждающей выборки*, или *ускорения*, переноса (carry-lookahead) [Weinberger56], [MacSorley61]. Как отмечалось ранее, для каждого разряда N-битового сумматора справедливо следующее соотношение:

$$C_{0,k} = f(A_k, B_k, C_{0,k-1}) = G_k + P_k C_{0,k-1}. \tag{11.15}$$

Зависимость между $C_{o,k}$ и $C_{o,k-1}$ можно устранить, расписав $C_{o,k-1}$:

$$C_{0,k} = G_k + P_k(G_{k-1} + P_{k-1}C_{0,k-2}). \tag{11.16}$$

Окончательно получаем:

$$C_{0,k} = G_k + P_k(G_{k-1} + P_{k-1}(\ldots + P_1(G_0 + P_0 C_{i,0}))), \tag{11.17}$$

где $C_{i,0}$ обычно равно 0.

На основе подобного развернутого уравнения можно реализовать N-битовый сумматор. Для каждого бита выходные сигналы переноса и суммы не зависят от предыдущих битов. Таким образом, эффект сквозного переноса эффективно устранен, и время суммирования не должно зависеть от числа битов. Общая блок-схема сумматора с ускоренным переносом показана на рис. 11.20.

Подобная высокоуровневая модель содержит несколько скрытых зависимостей. При изучении подробной схемы сумматора становится ясно, что постоянное время сложения — не более чем иллюзия и что реальная задержка по меньшей мере линейно растет с увеличением числа битов. Сказанное иллюстрируется на рис. 11.21, где для случая $N = 4$ представлена возможная реализация уравнения (11.17) на схемном уровне. Обратите внимание на то,

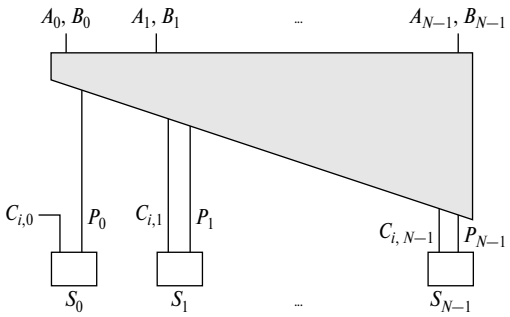


Рис. 11.20. Блок-схема сумматора с ускоренным переносом

что в схеме использована самодуальность и рекурсивность уравнения ускоренного переноса, на основе чего создается зеркальная структура, похожая на однобитовый полный сумматор, показанный на рис. 11.6[4]. Из-за большого разветвления по входу данная схема невероятно медленная для больш*их* значений N. Реализация ее на более простых элементах требует использования нескольких логических уровней. В любом случае задержка распространения увеличивается. Более того, чрезмерно возрастают некоторые разветвления по выходу, еще больше замедляя сумматор. Например, сигналы G_0 и P_0 фигурируют в уравнении для каждого последующего бита. Следовательно, емкость в этих линиях достаточно велика. Наконец, площадь реализации растет вместе с N. Следовательно, структура с ускоренным переносом, предлагаемая уравнением (11.16), полезна только для небольших значений N (≤ 4).

Логарифмический сумматор с ускоренным переносом

Для группы из N бит с ускоренным переносом транзисторная реализация имеет $N+1$ параллельную ветвь, причем число транзисторов в стеке может доходить до $N+1$. Поскольку широкие элементы и большие стеки характеризуются низкой производительностью, на практике вычисления с ускоренным переносом необходимо ограничивать двумя–четырьмя битами. Чтобы построить очень быстрые сумматоры, генерацию и распространение сигнала переноса необходимо организовать в форме рекурсивных деревьев. Более эффективной является реализация с иерархическим разложением распространения переноса на подгруппы из N бит:

$$
\begin{aligned}
C_{o,0} &= G_0 + P_0 C_{i,0};\\
C_{o,1} &= G_1 + P_1 G_0 + P_1 P_0 C_{i,0} = (G_1 + P_1 G_0) + (P_1 P_0) C_{i,0} = G_{1:0} + P_{1:0} C_{i,0};\\
C_{o,2} &= G_2 + P_2 G_1 + P_2 P_1 G_0 + P_2 P_1 P_0 C_{i,0} = G_2 + P_2 C_{o,1};\\
C_{o,3} &= G_3 + P_3 G_2 + P_3 P_2 G_1 + P_3 P_2 P_1 G_0 + P_3 P_2 P_1 P_0 C_{i,0} =\\
&\quad = (G_3 + P_3 G_2) + (P_3 P_2) C_{o,1} = G_{3:2} + P_{3:2} C_{o,1}.
\end{aligned}
\tag{11.18}
$$

[4] Подобно зеркальному сумматору, данная схема требует, чтобы сигнал распространения определялся как $P = A + B$.

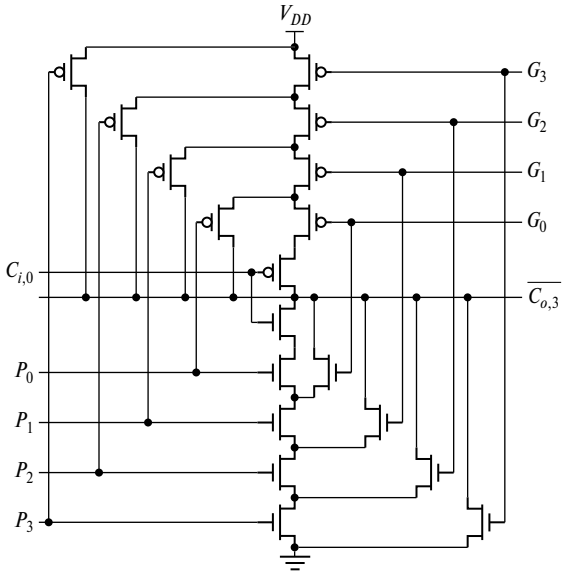


Рис. 11.21. Принципиальная схема зеркальной реализации четырехбитового сумматора с ускоренным переносом (согласно [Weste85])

В уравнении (11.18) процесс распространения сигнала переноса раскладывается на подгруппы из двух бит. Символами $G_{i:j}$ и $P_{i:j}$ обозначены соответственно функции генерации и распространения для *группы битов* (с разряда i до разряда j). Поэтому далее будем называть их *блочными* сигналами *генерации* и *распространения*. $G_{i:j}$ равно 1, если, независимо от входного сигнала переноса, группа генерирует сигнал переноса. Блочный сигнал распространения $P_{i:j}$ имеет значение true, если входной сигнал переноса распространяется через всю группу. Данное условие эквивалентно рассмотренному ранее условию обвода сигнала переноса. Например, $G_{3:2}$ равно 1, когда сигнал переноса либо генерируется в разряде 3, либо генерируется в разряде 2 и проносится через разряд 3, или $G_{3:2} = G_3 + P_3G_2$. $P_{3:2}$ имеет значение true, когда входной сигнал переноса проходит через оба разряда, или $P_{3:2} = P_3P_2$.

Обратите внимание на то, что формат нового выражения для сигнала переноса эквивалентен исходному, только на этот раз вместо сигналов генерации и распространения использовались блочные сигналы генерации и распространения. Запись $G_{i:j}$ и $P_{i:j}$ является обобщением исходных уравнений для сигнала переноса, поскольку $G_i = G_{i:i}$ и $P_i = P_{i:i}$. Еще одно обобщение можно получить, рассматривая функции генерации и распространения как пару $(G_{i:j}, P_{i:j})$, а не как отдельные функции. Далее можно ввести новый булев оператор "точка" $(\cdot)$, который действует на пары функции и обозначает следующую операцию с битами:

$$(G, P) \cdot (G', P') = (G + PG', PP'). \tag{11.19}$$

Используя данный оператор, можно разложить $(G_{3:2}, P_{3:2}) = (G_3, P_3) \cdot (G_2, P_2)$. Оператор "точка" обладает свойством ассоциативности, но не является коммутативным.

Пример 11.4. Сумматор со сквозным переносом, представленный с помощью оператора "точка"

Используя оператор "точка", мы можем следующим образом переписать четырехбитовый сумматор со сквозным переносом:

$$(C_{o,3}, 0) = [(G_3, P_3) \cdot (G_2, P_2) \cdot (G_1, P_1) \cdot (G_0, P_0)] \cdot (C_{i,0}, 0).$$

Используя свойство ассоциативности, данную функцию можно переписать и представить $C_{o,3}$ как функцию двух групповых сигналов переноса:

$$\begin{aligned}(G_{3:0}, P_{3:0}) &= [(G_3, P_3) \cdot (G_2, P_2)] \cdot [(G_1, P_1) \cdot (G_0, P_0)] = \\ &= (G_{3:2}, P_{3:2}) \cdot (G_{1:0}, P_{1:0}).\end{aligned}$$

Используя свойство ассоциативности оператора "точка", можно построить дерево, эффективно вычисляющее сигналы переноса во всех $2^i - 1$ позициях (т.е. 1, 3, 7, 15 и т.д.) для $i = 1 \ldots \log_2(N)$. Основное достоинство такого подхода заключается в том, что вычисление сигнала переноса в позиции $2^i - 1$ требует всего $\log_2(N)$ шагов. Другими словами, выходной сигнал переноса N-битового сумматора можно вычислить за время $\log_2(N)$. Это существенное улучшение по сравнению с рассмотренными выше сумматорами. Например, для 64-битового сумматора задержка распространения линейного сумматора пропорциональна 64. Для сумматора с характеристикой типа квадратного корня эта величина составляет всего 8, а для логарифмического сумматора коэффициент пропорциональности равен 6. Сказанное иллюстрируется на рис. 11.22, где показана блок-схема 16-битового логарифмического сумматора. Сигнал переноса в позиции 15 вычисляется по результатам, выданным блоками (0:7) и (8:15). В свою очередь, эти результаты также раскладываются иерархически. Например, (0:7) — это результат обработки сигналов (0:3) и (4:7), а (0:3) состоит из (0:1) и (2:3) и т.д.

Очевидно, расчета сигналов переноса всего в $2^i - 1$ позиции недостаточно. Помимо этого требуется знать сигналы переноса в промежуточных позициях. Для решения этой задачи можно продублировать дерево в каждом разряде, как показано на рис. 11.22 для $N = 16$. Например, сигнал переноса в позиции 6 вычисляется путем объединения результатов блоков (6:3) и (2:0). Полученная в ходе подобных манипуляций структура часто называется *деревом Когга–Стоуна* (Kogge–Stone tree) [Kogge73] и относится к классу деревьев *по основанию 2*. "По основанию 2" означает, что дерево является двоичным: на каждом уровне иерархии оно объединяет два слова, состоящих из сигналов переноса. Реализация общего сумматора, представленного с помощью оператора "точка", требует 49 сложных логических элементов.

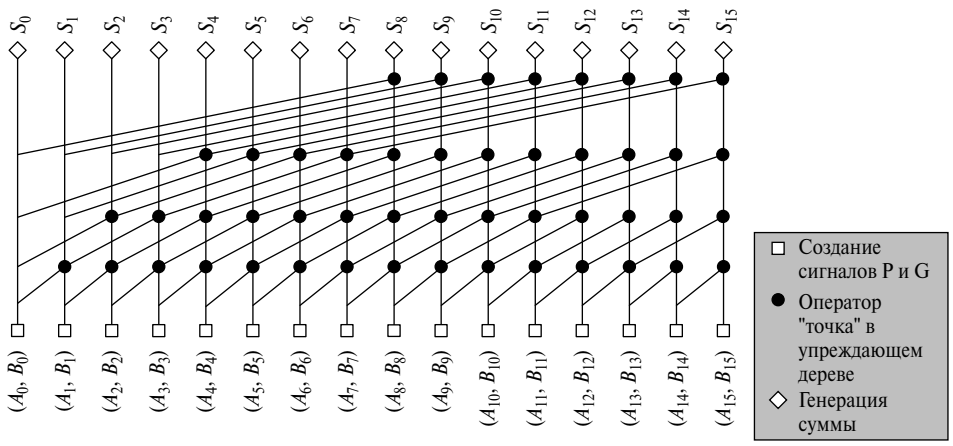


Рис. 11.22. Принципиальная схема 16-битового логарифмического сумматора с ускоренным переносом Когга–Стоуна

Кроме того, для генерации сигналов распространения и генерации на первом уровне (P_i и G_i) требуется 16 логических модулей, также необходимы 16 элементов генерации суммы.

Реализация сумматора с ускоренным переносом на динамической логике

Когда основной целью проектирования является очень высокое быстродействие, идеальным кажется сочетание техник ускоренного переноса (Carry-Lookahead — CLA) и динамической логики. Следовательно, полезно будет подробно рассмотреть принципы динамического ускоренного переноса.

Первый модуль генерирует сигналы распространения и генерации, как показано на рис. 11.23. Небольшой хитростью является добавление отдельного инвертора, управляющего устройством восстановления уровня. Такой подход выгоден в элементах, питающих большие разветвления по выходу. Отделяя управляющий элемент от разветвления по выходу, мы разрешаем быстрое отключение устройства восстановления уровня после начала перехода. С другой стороны, инвертор, питающий последующие логические элементы, оптимизирован для питания разветвления по выходу, равного двум (для выходов G) или трем (для выходом P) n-МОП-цепям перевода в низкое состояние.

Каждая черная точка на рис. 11.22 представляет два логических элемента, вычисляющих блочные сигналы генерации и распространения, как показано на рис. 11.24. Поскольку данные элементы располагаются не в начале конвейера, транзистор оценки является необязательным (см. главу 6). Такой подход широко используется в динамических информационных каналах. В фазе предварительной зарядки все выходы элементов “домино” гарантированно будут низкими, отключая любой возможный путь разрядки последующих каскадов “домино”. Удаление транзисторов оценки в любом каскаде, кроме первого, приводит к снижению логических трудозатрат элемен-

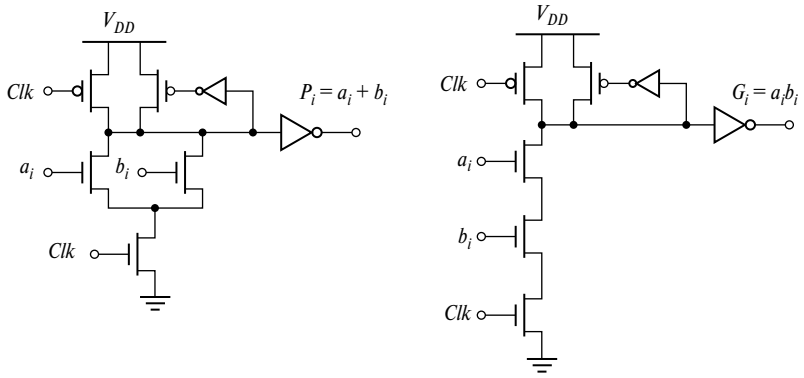


Рис. 11.23. Динамическая реализация сигналов распространения и генерации

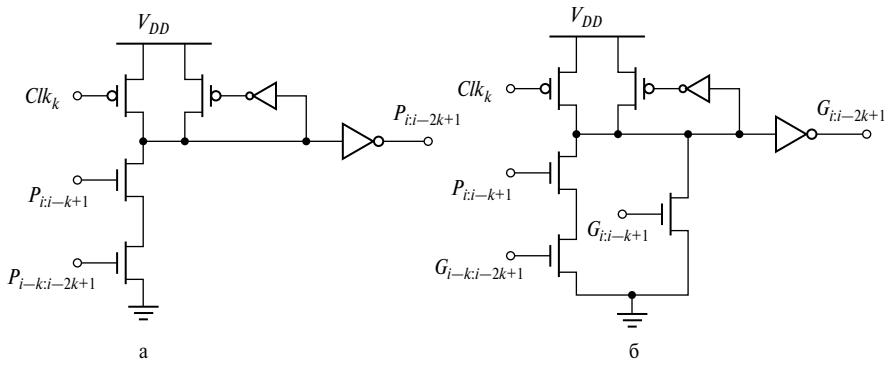


Рис. 11.24. Реализация оператора “точка” на динамической логике: а — логика распространения; б — логика генерации в каскаде k и разряде i (см. рис. 11.22)

тов и ускоряет оценку. Например, элемент распространения, показанный на рис. 11.24, имеет логические трудозатраты, равные $2/3$, а не 1 (предполагается, что инвертор симметричный). Тем не менее у данного метода есть недостаток. В фазе предварительной зарядки во всех элементах без транзистора оценки ток короткого замыкания существует до тех пор, пока входы этих элементов не будут предварительно заряжены. Чтобы убрать ток короткого замыкания, тактовые сигналы каждого каскада Clk_k задерживаются по сравнению с предыдущим каскадом. Данный подход называется *схемой “домино” с задержкой тактового сигнала* (см. главу 10). Обратите внимание на то, что тактовый сигнал задерживается на одно и то же время для всех разрядов каскада, что упрощает реализацию описанного подхода.

Собирая в единое целое семь каскадов логики (генерация сигналов P и G плюс шесть операторов “точка”), можно построить 64-битовый сумматор. Единственным каскадом, которого не хватает для получения динамического сумматора, является конечный блок генерации суммы. Генерация суммы

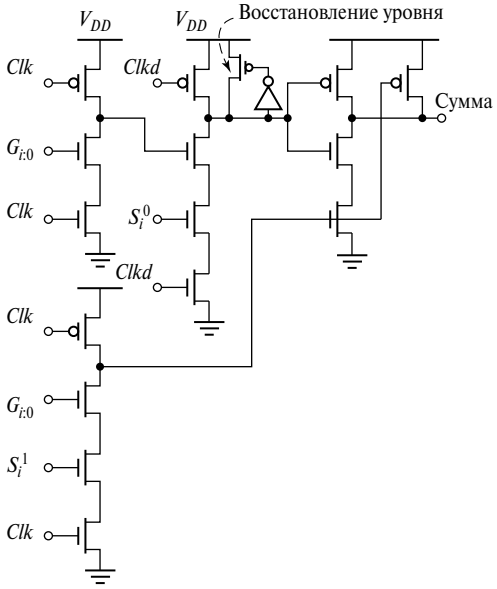


Рис. 11.25. Выбор суммы в динамической логике

требует использования функции ИСКЛЮЧАЮЩЕЕ ИЛИ, которую не так просто реализовать на логике "домино". Разумеется, можно использовать статические элементы ИСКЛЮЧАЮЩЕЕ ИЛИ, но это приведет к появлению немонотонных переходов, а следовательно, выход такой реализации нельзя будет подавать на другие элементы "домино". В принципе с этим можно смириться, поскольку генерация суммы часто является последним этапом операции сложения. Однако защелка, следующая за генератором суммы, не может быть прозрачной, поскольку это может привести к нарушению правила переходов для последующих элементов "домино".

Один из способов реализации суммы на логике "домино" — использовать *выбор суммы*, когда вычисляются два варианта суммы, $S_i^0 = \overline{a_i \oplus b_i}$ и $S_i^1 = a_i \oplus b_i$. Затем с помощью динамического элемента, показанного на рис. 11.25, после получения входного сигнала переноса происходит выбор одной из двух возможностей.

Реализация мультиплексора требует наличия трех логических уровней, поскольку в логике "домино" нет комплементарного сигнала переноса. Как обычно, во всех динамических узлах необходимо разместить устройства восстановления уровня, но самый критичный из них показан на рисунке. Два первых динамических каскада (вверху) нарушают правила проектирования динамических узлов: два динамических элемента соединены каскадом без промежуточного инвертора. На выходе второго элемента возможно появление импульсных помех, если значения обоих элементов оцениваются с помощью одного тактового сигналом. Решить эту проблему помогает задержка тактового сигнала для второго элемента ($Clkd$).

Следует отметить, что, когда разработчик решает использовать описанные техники проектирования, необходимо обращать особое внимание на устойчивость схемы. Все тактовые сигналы подвержены случайной расфазировке, и запаздывающий тактовый сигнал должен иметь достаточный запас прочности, чтобы справиться с наихудшей расфазировкой. Если $Clkd$ прибывает на узел рано, сумматор будет работать неверно. Впрочем, при правильном масштабировании устройства восстановления уровня можно обеспечить вполне достаточную надежность проекта. Таким образом можно избежать импульсных помех на выходе второго каскада при раннем поступлении тактового сигнала. Устройство восстановления уровня должно быть достаточно большим, чтобы минимизировать указанные помехи, но не настолько большим, чтобы сильно повлиять на быстродействие.

Другая возможность заключается в реализации сумматора на дифференциальной логике "домино", когда доступны сигналы с обеими полярностями и проблема инверсии исчезает. Следует, правда, отметить, что при такой реализации вся система может быть достаточно энергоемкой.

Задача 11.4. Проектирование дерева статического сумматора

Спроектируйте дерево 16-битового сумматора с ускоренным переносом на статических КМОП-схемах. Разработайте дерево ускоренного переноса, используя инвертированные логические элементы и не добавляя инверторы. Обозначьте критический путь данного сумматора. Чему равны логические трудозатраты элементов, расположенных на критическом пути? Как с помощью их масштабирования получить минимальную задержку?

Логарифмический сумматор с ускоренным переносом — альтернативы

Разработчики быстрых сумматоров иногда возвращаются к старым стилям древовидных структур, если с помощью какого-нибудь компромисса требуется получить определенную площадь, мощность или быстродействие. Ниже мы кратко рассмотрим две основные альтернативные структуры — *сумматор Брента–Канга* (Brent–Kung adder) и *четверичный сумматор* (radix-4 adder).

Дерево Когга–Стоуна (см. рис. 11.22) обладает некоторыми интересными свойствами. Во-первых, это правильная структура межсоединений, что существенно облегчает реализацию. Во-вторых, разветвление по выходу в разных точках дерева практически одинаковое (особенно это относится к критическим путям). Соответственно упрощается задача масштабирования транзисторов с целью получения оптимального быстродействия. В то же время дублирование деревьев переноса с целью генерации промежуточных сигналов переноса является достаточно дорогим с точки зрения площади и мощности. Иногда за счет задержки разработчики добиваются более выгодных площади и мощности, выбирая менее сложные деревья. При более простой структуре дерева вычисляются только сигналы переноса разрядов, являющихся степенями двойки [Brent82], как показано на рис. 11.26 для $N = 16$.

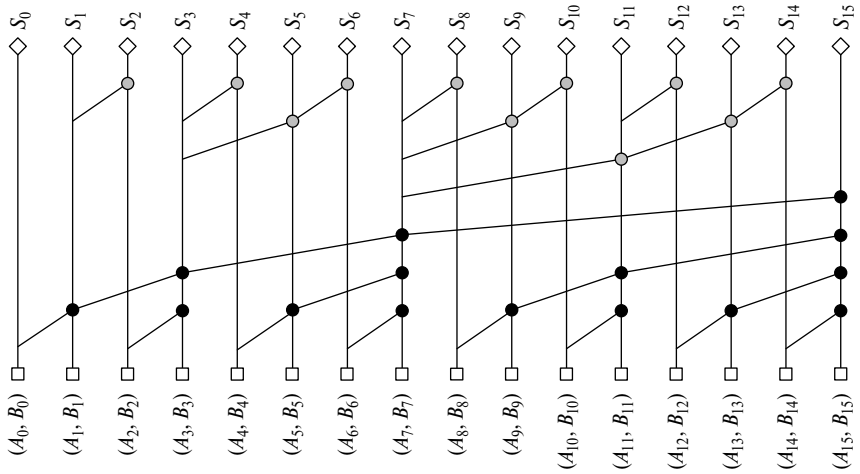


Рис. 11.26. 16-битовое дерево Брента–Канга. Обратное двоичное дерево обозначено серым цветом

Прямое двоичное дерево реализует сигналы переноса только в позициях $2^N - 1$:

$$
\begin{aligned}
&(C_{o,0}, 0) = (G_0, P_0) \cdot (C_{i,0}, 0);\\
&(C_{o,1}, 0) = [(G_1, P_1) \cdot (G_0, P_0)] \cdot (C_{i,0}, 0) = (G_{1:0}, P_{1:0}) \cdot (C_{i,0}, 0) = \\
&(C_{o,3}, 0) = [(G_{3:2}, P_{3:2}) \cdot (G_{1:0}, P_{1:0})] \cdot (C_{i,0}, 0) = (G_{3:0}, P_{3:0}) \cdot (C_{i,0}, 0);\\
&(C_{o,7}, 0) = [(G_{7:4}, P_{7:4}) \cdot (G_{3:0}, P_{3:0})] \cdot (C_{i,0}, 0) = (G_{7:0}, P_{7:0}) \cdot (C_{i,0}, 0).\\
&\quad \ldots
\end{aligned}
\tag{11.20}
$$

Структуры прямого двоичного дерева недостаточно для генерирования полного набора битов переноса. Для реализации остальных битов требуется *обратное двоичное дерево* (серые линии на рис. 11.26), которое на основе промежуточных результатов получает оставшиеся биты сигнала переноса. Читателю предлагается самостоятельно проверить, что данная структура выдает правильные выражения для всех битов сигнала переноса. Получившуюся в целом структуру часто называют *сумматором Брента–Канга*; она использует 27 элементов "точка", или почти половину из 49 элементов, требуемых для реализации полного двоичного дерева (разумеется, соединений в данном случае также требуется меньше). При этом проводная структура получается нерегулярной с разными разветвлениями по выходу для разных узлов, что несколько усложняет оптимизацию быстродействия. Особенно большие проблемы связаны с разветвлением по выходу промежуточного узла ($C_{0,7}$), которое в данном случае равно одной сумме и пяти операциям "точка". Из-за этого сумматор Брента–Канга не подходит для очень больших сумматоров (больше 32 бит).

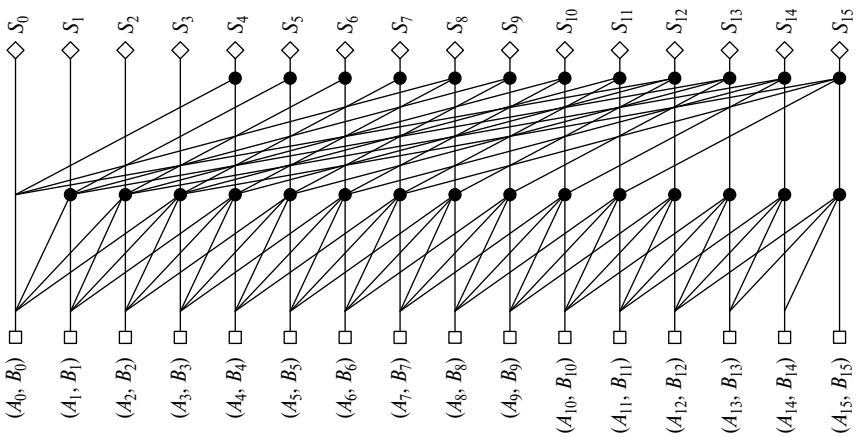


Рис. 11.27. Четверичное дерево Когга–Стоуна для 16-битовых операндов

Чтобы уменьшить глубину дерева, можно последовательно на каждом уровне иерархии объединять по четыре сигнала. Получающееся в результате дерево относится к классу деревьев *по основанию 4*, поскольку использует блоки четвертого порядка, как показано на рис. 11.27. Реализация 16-битового сложения требует всего двух каскадов логики переноса. Учтите, что в данном случае каждый элемент сложнее, чем в описанных выше схемах, поэтому наличие меньшего числа логических каскадов не обязательно даст более быструю работу (см. главу 5).

Задача 11.5. Четверичный оператор "точка" в динамической логике

Разработайте четверичные операторы "точка" в динамической логике. В качестве отправной точки возьмите схемы по основанию 2. Как реализуется выбор суммы по основанию 4? Используя метод логических трудозатрат, сравните задержки двоичных и четверичных структур 16-битового сумматора. Какой вариант быстрее?

В среднем размеры сумматора с ускоренным переносом в несколько раз больше, чем размеры сумматора со сквозным переносом, однако для больших операндов имеет существенно большее быстродействие. Благодаря логарифмической характеристике задержки для больших значений N данный сумматор выгоднее, чем сумматоры с обводом или выбором сигнала переноса. Точное значение точки характеристик этих устройств зависит в основном от технологии и факторов, связанных с проектированием схем.

Наше обсуждение сумматоров получилось далеко не полным. Из-за своего влияния на быстродействие вычислительных структур проектированию быстрых сумматоров посвящено довольно много работ. В принципе можно даже построить сумматор, задержка распространения которого *не будет зависеть от числа битов*. В качестве примеров подобных структур можно приве-

сти схемы с записью сигнала переноса и избыточной двоичной арифметикой [Swartzlander90]. Такие сумматоры требуют этапов численного кодирования и декодирования, задержка которых является функцией N. Следовательно, они интересны только при внедрении в большие структуры, подобные умножителям или высокоскоростным процессорам сигналов.

11.4. Умножитель

Умножение представляет собой дорогую и медленную операцию. Быстродействие многих вычислительных задач часто определяется скоростью, с которой можно выполнять умножение. Поэтому в современных цифровых микропроцессорах и процессорах сигналов происходит интеграция полных блоков умножения.

По сути, умножители представляют собой сложные массивы сумматоров. Следовательно, подавляющее большинство тем, рассмотренных в предыдущем разделе, важны и в контексте изучения умножителей. Анализ умножителя позволяет лучше понять, как оптимизировать быстродействие (или площадь) схем со сложной топологией. После краткого обсуждения операции умножения мы рассмотрим стандартный матричный умножитель. Кроме того, мы обсудим различные подходы к генерации частичных произведений, их накопления и конечного суммирования.

11.4.1. Умножитель: определения

Рассмотрим два двоичных числа *без знака* X и Y, имеющие ширину M и N бит соответственно. Чтобы ввести операцию умножения, полезно выразить X и Y в двоичном представлении:

$$X = \sum_{i=0}^{M-1} X_i 2^i \quad Y = \sum_{j=0}^{N-1} Y_j 2^j, \tag{11.21}$$

где $X_i, Y_j \in 0{,}1$. Операция умножения определяется следующим образом:

$$\begin{aligned} Z = X \times Y = & \sum_{k=0}^{M+N-1} Z_k 2^k = \\ & = \left(\sum_{i=0}^{M-1} X_i 2^i \right) \left(\sum_{j=0}^{N-1} Y_j 2^j \right) = \sum_{i=0}^{M-1} \left(\sum_{j=0}^{N-1} X_i Y_j 2^{i+j} \right). \end{aligned} \tag{11.22}$$

Наиболее простой способ выполнения операции умножения — это использовать один сумматор с двумя входами. Для входов шириной M и N бит умножение с использованием N-битового сумматора требует M тактов. Представленный алгоритм реализации умножения в виде *смещения/суммирования* заключается в суммировании M *частичных произведе-*

Рис. 11.28. Пример двоичного умножения

ний, каждое из которых генерируется путем умножения множимого на один бит множителя (по сути, это операция И) и смещения результата согласно разряду множителя.

Более быстрый способ реализации умножения заключается в использовании подхода, подобного реализованному при умножении в столбик. Все частичные произведения генерируются одновременно и записываются в виде таблицы. Для вычисления конечного произведения используется одновременное суммирование всех промежуточных значений. Данный подход иллюстрируется на рис. 11.28. Описанный набор операций можно реализовать непосредственно на аппаратном уровне. Получающаяся в результате структура называется *матричным умножителем* (array multiplier) и объединяет следующие три функции: *генерацию частичных произведений*, *накопление частичных произведении* и *конечное суммирование*.

Задача 11.6. Умножение на константу

Довольно часто требуется реализовать умножение операнда на константу. Опишите, как бы вы решали эту задачу.

11.4.2. Генерация частичных произведений

Частичные произведения получаются при применении логического И к множимому X и биту множителя Y_i (рис. 11.29). Каждая строка матрицы частичного произведения представляет собой либо копию множимого, либо строку нулей. Аккуратно оптимизируя процесс генерации частичных произведений, можно существенно уменьшить задержку и площадь, требуемую для реализации. Обратите внимание на то, что в большинстве случаев матрицы частичных произведений имеют много нулевых строк, которые никак не влияют на результат, а следовательно, их суммирование представляет собой пустую трату времени. Если множитель состоит из одних единиц, существуют все частичные произведения, а если из одних нулей, то нет ни одного частичного произведения. Данное наблюдение позволяет вполовину уменьшить число генерируемых частичных произведений.

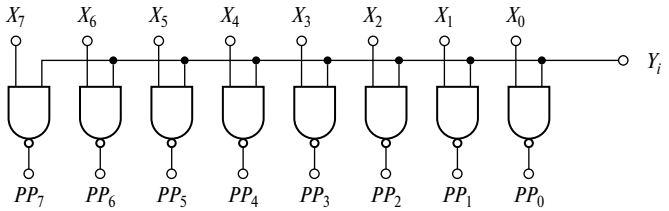


Рис. 11.29. Логика генерации частичных произведений

Рассмотрим, например, восьмибитовый множитель вида 01111110, который дает шесть ненулевых частичных произведений. Число ненулевых строк можно существенно уменьшить, записав соответствующее число $(2^7+2^6+2^5+2^4+2^3+2^2)$ в другом формате. Можете проверить, что то же самое число представляется в виде $10000\bar{1}0$, где $\bar{1}$ — сокращенная запись -1. Используя данный формат, нам останется просуммировать всего два частичных произведения, но конечный сумматор должен уметь выполнять операцию вычитания. Преобразование подобного типа называется *перекодированием Бута* (Booth's recoding) [Booth51]; оно может уменьшать число частичных произведений вдвое (максимум). Данное кодирование гарантирует, что для любых двух последовательных битов по крайней мере один будет равен 1 или -1. Уменьшение числа частичных произведений эквивалентно уменьшению числа операций сложения, что не только ускоряет процесс умножения, но и уменьшает площадь умножителя. Формально такое преобразование эквивалентно переводу слова-множителя в систему по основанию 4 вместо использования обычного двоичного формата:

$$Y = \sum_{j=0}^{(N-1)/2} Y_j 4^j, \text{ где } (Y_j \in \{-2,-1,0,1,2\}). \qquad (11.23)$$

Обратите внимание на то, что самым плохим входом умножителя является 1010...10, поскольку он генерирует наибольшее число частичных произведений (половину от максимального числа). Отметим, что умножение не $\{0,1\}$ эквивалентно операции И, а умножение на $\{-2,-1,0,1,2\}$ требует наличия логики инверсии и смещения. Для необходимого кодирования требуется всего несколько простых логических элементов.

Наличие матрицы частичных произведений, имеющей переменный размер, неудобно при проектировании умножителей, поэтому, как правило, используется *модифицированное кодирование Бута* [MacSorley61]. Множитель разбивается на трехбитовые группы, причем соседние группы перекрываются на один бит. Каждая группа записывается, как показано в табл. 11.2, и формирует одно частичное произведение. Получающееся число частичных произведений составляет половину ширины множителя. В процессе записи задействуются два текущих бита плюс верхний бит из следующей группы; данное "окно" последовательно проходит от старшего до младшего разряда.

Таблица 11.2. Модифицированное кодирование Бута

Таблица выбора частичных произведений	
Биты множителя	Записанные биты
000	0
001	+множитель
010	+множитель
011	$+2\times$ множитель
100	$-2\times$ множитель
101	−множитель
110	−множитель
111	0

Проще говоря, при модифицированном кодировании Бута множитель последовательно, начиная со старшего и заканчивая младшим разрядом, исследуется на предмет наличия строк из одних единиц, первая из которых заменяется 1, а последняя — −1. Например, 011 понимается как начало строки, состоящей из единиц, а следовательно, заменяется единицей в первой позиции (или 100), а 110 понимается как конец строки и заменяется −1 в младшем разряде (или $0\bar{1}0$).

Пример 11.5. Модифицированное кодирование Бута

Рассмотрим восьмибитовое двоичное число 01111110. Его можно разбить на четыре перекрывающихся группы из трех битов, идущих от старшего до младшего разрядов: 01 (1), 11 (1), 11 (1), 10 (0). Используя табл. 11.2, записываем: 10 ($2\times$), 00 ($0\times$), 00 ($0\times$), $\bar{1}0$ ($-2\times$), или $1000000\bar{1}0$, т.е. получаем тот же результат, что и ранее.

Задача 11.7. Устройство кодирования Бута

Используя табл. 11.2, спроектируйте комбинационную логику, реализующую модифицированное кодирование Бута для параллельного умножителя. Сравните полученную реализацию на КМОП-логике и логике на проходных транзисторах.

11.4.3. Накопление частичных произведений

После генерации частичных произведений их нужно просуммировать. Наиболее прямым способом накопления частичных произведений является использование нескольких сумматоров, формирующих массив; соответствующая реализация получила название *матричного умножителя*. Существуют и более сложные процедуры, например с представлением суммирования в древовидном формате.

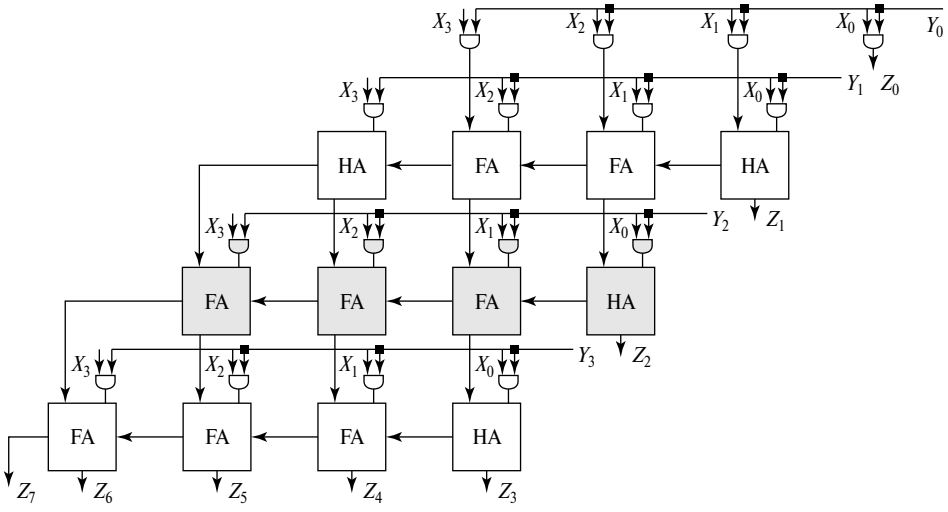


Рис. 11.30. Структура матричного умножителя чисел без знака (4×4 бит). Буквами HA обозначен *полусумматор* (half adder), или *одноразрядный сумматор с двумя входами*. Серым цветом выделена аппаратная реализация генерации и сложения одного частичного произведения

Матричный умножитель

Структура матричного умножителя показана на рис. 11.30. Существует взаимно-однозначное соответствие между данной аппаратной структурой и схематическим представлением умножения, показанным на рис. 11.28. Генерация N частичных произведений требует $N \times M$ двухбитовых элементов И (подобных показанным на рис. 11.29)[5]. Бо́льшая часть площади умножителя уходит на реализацию сложения N частичных произведений, что требует $N - 1$ M-битовых сумматоров. Смещение частичных произведений, дающее правильное их выравнивание, выполняется с помощью трассировки и не требует никаких логических схем. Общую структуру легко вместить в прямоугольник, что является довольно эффективным с точки зрения топологии.

Из-за матричной структуры схемы определить ее задержку распространения не так просто. Рассмотрим реализацию, приведенную на рис. 11.30. Сумматоры частичных сумм реализованы в виде структур со сквозным переносом. Для оптимизации быстродействия требуется вначале определить критический временной путь. В данном случае эта задача является нетривиальной. Фактически можно выделить большое число путей практически одинаковой длины. Два из них показаны на рис. 11.31. Внимательно изучив данные пути, мы получим приблизительное выражение для задержки распространения (в данном случае выведенное для критического пути 2):

[5] Именно в этой реализации кодирование Бута не задействовано. Впрочем, добавление кодирования не сильно меняет реализацию. Число сумматоров сокращается вдвое, а генерация частичных произведений немного усложняется.

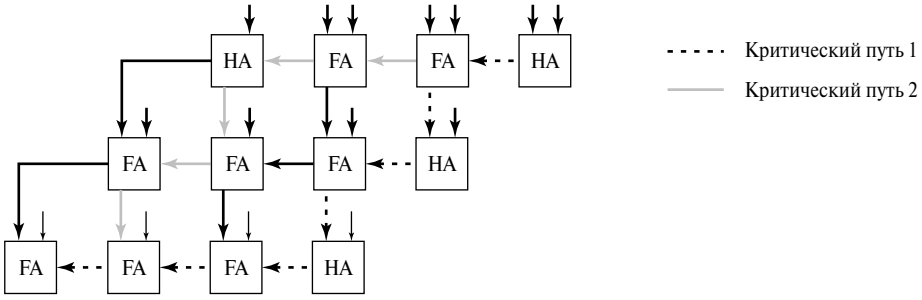


Рис. 11.31. Умножитель 4×4, основанный на сумматорах со сквозным переносом (упрощенная диаграмма). Обозначены два возможных критических пути

$$t_{mult} \approx [(M-1)+(N-2)]t_{carry} + (N-1)t_{sum} + t_{AND}, \qquad (11.24)$$

где t_{carry} — задержка распространения между входным и выходным сигналом переноса; t_{sum} — задержка между входным сигналом переноса и битом суммы полного сумматора; t_{AND} — задержка элемента И.

Поскольку все критические пути имеют одинаковую длину, ускорение одного из них (например, путем замены одного сумматора более быстрой реализацией — сумматором с выбором сигнала переноса) не имеет никакого смысла с точки зрения проекта в целом. Необходимо одновременно улучшить все критические пути. Из уравнения (11.24) следует, что минимизация t_{mult} требует минимизации t_{carry} и t_{sum}. В данном случае выгодно, чтобы t_{carry} равнялось t_{sum}. Это противоречит требованиям, которые предъявлялись к рассмотренным ранее сумматорам, когда важно было только минимальное время t_{carry}. Пример полного сумматора с приблизительно одинаковыми задержками t_{sum} и t_{carry} показан на рис. 11.7.

Задача 11.8. Двоичный умножитель величин со знаком

Умножитель, представленный на рис. 11.30, обрабатывает только числа без знака. Измените его структуру так, чтобы он мог работать с числами в дополнительном двоичном коде.

Умножитель с записью переноса

Из-за большого числа идентичных критических путей увеличение быстродействия структуры, приведенной на рис. 11.31, посредством масштабирования транзисторов приносит минимальную пользу. Более эффективную реализацию можно получить, заметив, что результат умножения не меняется, когда биты выходного сигнала переноса передаются не только вправо, но и по диагонали вниз, как показано на рис. 11.32. Для генерации конечного результата мы ввели дополнительный сумматор, называемый *сумматором слияния векторов* (vector-merging adder). Полученное в результате устройство име-

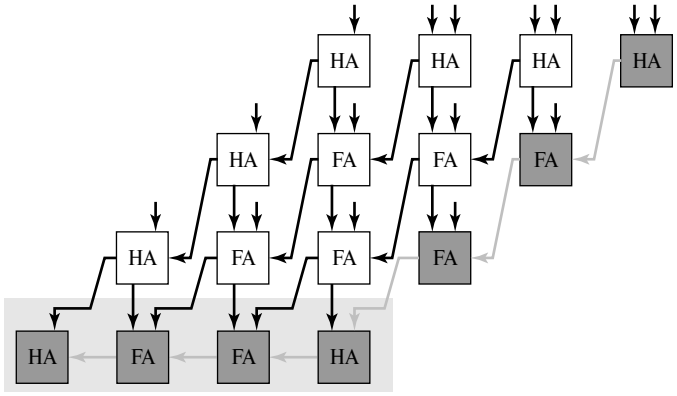


Рис. 11.32. Умножитель 4×4 с записью сигнала переноса. Критический путь показан серым цветом

нуется *умножителем с записью переноса* (carry-save multiplier) [Wallace64], поскольку биты сигнала переноса не добавляются незамедлительно, а сохраняются для следующего суммирующего каскада. В последнем каскаде сигналы переноса и суммы объединяются, например, с помощью сумматора с ускоренным переносом. Отметим, что хотя описанная структура имеет несколько бо́льшую площадь (один дополнительный сумматор), она выгоднее описанной выше, поскольку имеет единственный (к тому же более короткий) наихудший критический путь. Данный путь обозначен на рис. 11.32, а наибольшая задержка распространения сигнала по этому пути выражается следующим образом:

$$t_{mult} = t_{AND} + (N-1)t_{carry} + t_{merge} \tag{11.25}$$

(здесь по-прежнему предполагается, что $t_{add} = t_{carry}$).

Пример 11.6. Умножитель с записью переноса

При реализации показанного на рис. 11.32 умножителя с записью переноса на полупроводниковых структурах необходимо принять во внимание некоторые топологические нюансы. Для облегчения интеграции умножителя в кристалл рекомендуется, чтобы контур модуля был приблизительно прямоугольным. Архитектурный план умножителя с записью переноса, согласующегося с данной рекомендацией, показан на рис. 11.33. Обратите внимание на правильную топологию — благодаря ей генерацию структуры можно автоматизировать.

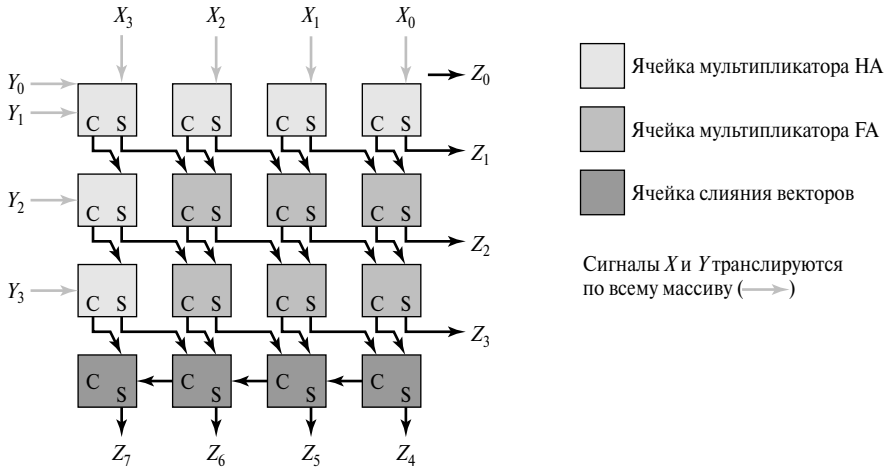


Рис. 11.33. Прямоугольный архитектурный план умножителя с записью переноса. Различные ячейки обозначены оттенками серого. Сигналы X и Y перед суммированием подвергаются операции И. Крайний левый столбец ячеек является избыточным, поэтому его можно удалить

Древовидный (каскадный) умножитель

Сумматоры частичных сумм также можно организовать в виде древовидной структуры, уменьшающей и критический путь, и число требуемых суммирующих ячеек. Рассмотрим простой пример, показанный на рис. 11.34, *а*: четыре частичных произведения по четыре бита шириной каждое. Число полных сумматоров, требуемых для данной операции, можно уменьшить, заметив, что суммировать четыре бита должен только третий столбец матрицы. Все остальные столбцы реализуют гораздо более простые операции. Сказанное иллюстрируется на рис. 11.34, *б*, где первоначальная матрица частичных произведений реорганизована в дерево. Наша задача состоит в том, чтобы получить матрицу с минимальной глубиной и минимальным числом суммирующих элементов. Первый тип операторов, которые можно использовать для заполнения матрицы, — это полный сумматор, принимающий три входа и выдающий два выхода: сумму, поступающую в тот же столбец, и сигнал переноса, переходящий в следующий столбец. Кроме того, имеется еще полусумматор, принимающий два входных бита в столбце и выдающий два выхода. На схеме полусумматор обозначен кружочком, охватывающим два бита.

Чтобы получить минимальную реализацию, мы последовательно закрываем дерево полными сумматорами и полусумматорами, начиная с наиболее плотной его части. На первом шаге мы вводим полусумматоры в столбцы 4 и 3 (см. рис. 11.34, *б*). На рис. 11.34, *в* показано сокращенное дерево. На втором шаге сокращения получается дерево с глубиной 2 (см. рис. 11.34, *г*). В процессе сокращения задействованы только три полных сумматора и три полусумматора по сравнению с шестью полными сумматорами и шестью полусумматорами умножителя с записью переноса, показанного на рис. 11.32! На последнем шаге остается всего один сумматор с двумя входами, для реали-

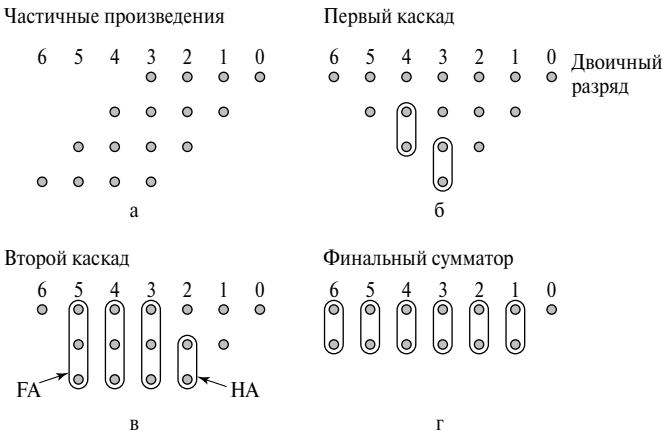


Рис. 11.34. Преобразование дерева частичных произведений (а) в дерево Уолласа (б–г) с использованием итеративного процесса охвата. В качестве примера использован четырехбитовый операнд

зации которого можно использовать любую удобную схему, как обсуждается в следующем разделе.

Представленная структура называется *умножителем в форме дерева Уолласа* (Wallace tree multiplier) [Wallace64], ее реализация показана на рис. 11.35. В таком виде умножитель позволяет существенно сэкономить аппаратные ресурсы для больших умножителей. Кроме того, при такой реализации уменьшается задержка распространения. Фактически можно показать, что задержка распространения сигнала по дереву равна $O(log_3/2(N))$. Хотя для умножителей с большой длиной слова умножитель Уолласа работает существенно быстрее, чем структуры с записью сигнала переноса; его недостатком является неправильная структура, которая затрудняет задачу разработки эффективной топологии. Эту нерегулярность можно видеть даже в четырехбитовой реализации, показанной на рис. 11.35.

Существует множество других способов реализации дерева частичных произведений. В литературе предлагается несколько схем сжатия, которые в данной книге мы рассматривать не будем. Все они основаны на использовании особенностей полных сумматоров (сжатия "3 в 2", когда каждый каскад умножителя уменьшает число частичных произведений в 1,5 раза). В принципе можно пойти и дальше, разработав схемы сжатия "4 в 2" (или более высокого порядка), подобные предложенным в [Weinberger81], [Santoro89]. Фактически по такому принципу построено множество современных высокоскоростных умножителей.

11.4.4. Конечное суммирование

Последним этапом умножения является объединение промежуточных результатов в конечном сумматоре. Быстродействие данной операции "слияния век-

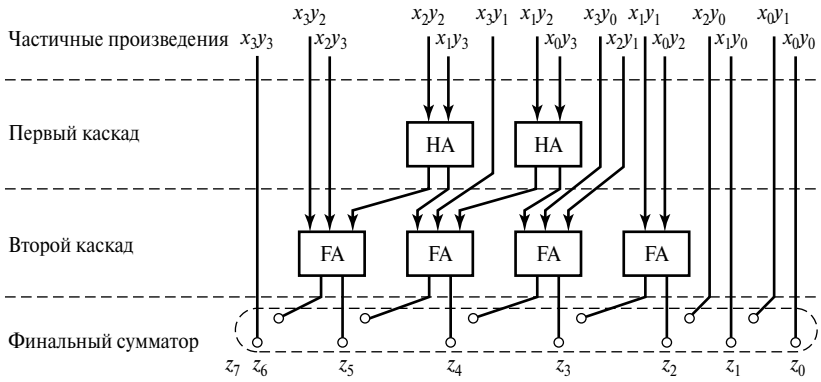


Рис. 11.35. Дерево Уолласа для четырехбитового умножителя

торов" имеет ключевое значение. Выбор стиля конечного сумматора зависит от структуры матрицы накопления промежуточных произведений. Если все входные биты сумматора прибывают одновременно, предпочтительным вариантом является сумматор с ускоренным переносом, поскольку он дает наименьшую возможную задержку. Такая ситуация возможна, когда каскад конвейера располагается непосредственно перед конечным сумматором. Вообще, технология конвейерной обработки довольно широко используется в высокоскоростных умножителях. В умножителях без конвейера время прибытия сигналов на конечный сумматор довольно сильно отличается из-за переменной глубины дерева умножителя. В такой ситуации часто выгодно выбрать другую топологию сумматора, например использовать выбор сигнала переноса; иногда это позволяет существенно уменьшить стоимость аппаратной реализации и получить хорошее быстродействие [Oklobdzija01].

11.4.5. Резюме

Все представленные технологии можно объединить и получить умножители с очень высоким быстродействием. Например, можно объединить кодирование Бута и дерево Уолласа с помощью сжатия "4 в 2" в реализации на проходных транзисторах и использовать в конечном сумматоре смешанную топологию с выбором сигнала переноса и ускоренным переносом. В таком случае реализованный по 0,25-микронной КМОП-технологии умножитель 54×54 будет иметь задержку распространения, равную всего 4,4 нс [Ohkubo95]. Подробнее об этом и других умножителях можно узнать в приведенной ниже литературе (например, [Swartzlander90], [Oklobdzija01]).

11.5. Сдвиговый регистр

Еще одной необходимой арифметической операцией, требующей адекватной аппаратной поддержки, является операция сдвига. Она широко используется

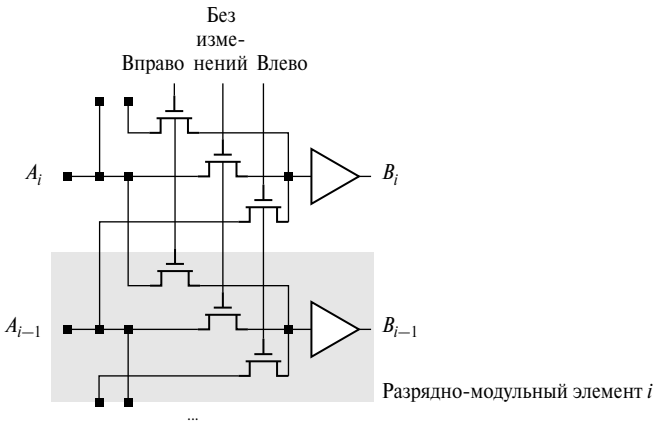


Рис. 11.36. Однобитовый (лево-правый) программируемый сдвиговый регистр

в модулях для выполнения операций с плавающей точкой, счетчиках и блоках, реализующих умножения на константы (которые, кстати, можно реализовать с помощью операций сложения и сдвига). Сдвиг слова данных влево или вправо на постоянную величину представляет собой тривиальную аппаратную операцию, реализовать которую можно с помощью правильной трассировки сигналов. С другой стороны, программируемый сдвиговый регистр является более сложным и требует наличия активных цепей. По сути, подобный регистр — это сложный мультиплексор. Рассмотрим представленный на рис. 11.36 простой однобитовый лево-правый сдвиговый регистр. В зависимости от управляющих сигналов входное слово либо смещается влево или вправо, либо остается без изменений. Объединяя несколько подобных блоков, можно создавать многобитовые сдвиговые регистры. Такой подход быстро становится чересчур сложным, неразумным и в конечном счете слишком медленным для больших сдвигов. Следовательно, желателен более структурированный подход. Ниже мы рассмотрим две распространенные структуры сдвиговых регистров, *многорегистровое циклическое сдвиговое устройство* (barrel shifter) и *логарифмический сдвиговый регистр* (logarithmic shifter).

11.5.1. Многорегистровое циклическое сдвиговое устройство

Структура многорегистрового циклического сдвигового регистра показана на рис. 11.37. Он состоит из матрицы транзисторов, в которой число строк равняется длине слова данных, а число столбцов равно максимальной ширине сдвига. В данном случае обе величины равны 4. Контрольные проводы проложены диагонально через матрицу. Основным достоинством предложенного сдвигового регистра является прохождение сигнала по большей мере через один передаточный вентиль. Другими словами, задержка распространения теоретически постоянна и не зависит от смещения или размера регистра.

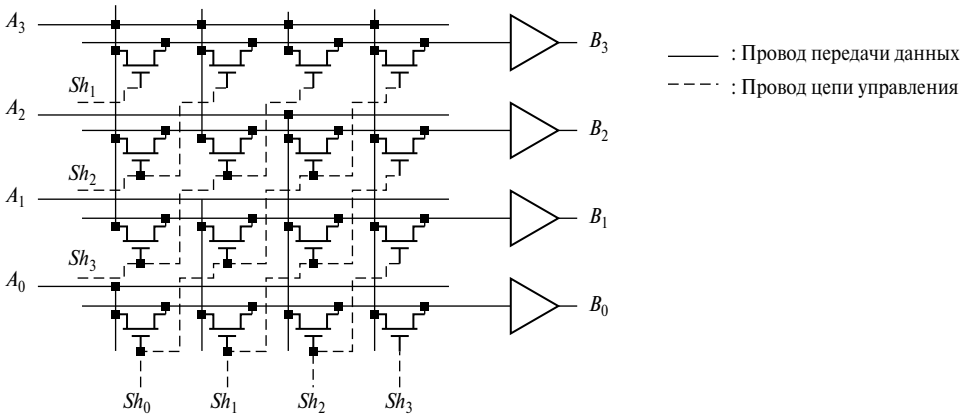


Рис. 11.37. Многорегистровое циклическое сдвиговое устройство с программируемым сдвигом вправо на величину от нуля до трех битов. Структура поддерживает автоматический повтор бита знака (А3)

К сожалению, в действительности это не так, поскольку емкость на входе буферов линейно растет с шириной сдвига.

Данная схема обладает одним важным свойством — основное влияние на ее размер оказывают не активные транзисторы (в отличие от всех остальных арифметических схем), а число проводников, проложенных через ячейку. Точнее, размер ячейки ограничивается шагом между металлическими проводниками!

Другой важный момент, связанный с выбором сдвигового регистра, — это формат, в котором нужно представлять величину сдвига. Изучая принципиальную схему, представленную на рис. 11.37, видим, что регистр требует наличия контрольного провода для каждого бита сдвига. Например, в четырехбитовом регистре сдвига нужны четыре контрольных сигнала. Чтобы выполнить сдвиг на три бита, сигналы $Sh_3 : Sh_0$ должны иметь значение 1000, т.е. высокий уровень имеет только один сигнал. В процессоре требуемая величина сдвига обычно записывается в закодированном двоичном формате — намного более компактном представлении. Например, кодированному контрольному слову требуется всего два контрольных сигнала; смещение на три бита, например, представляется как 11. Чтобы перевести закодированное представление в исходный формат (с одним единичным битом), требуется дополнительный модуль — *декодер*. (Подробно декодеры рассмотрены в главе 12.)

Задача 11.9. Сдвиговый регистр с дополнением до двух

Объясните, почему сдвиговый регистр, показанный на рис. 11.37, реализует сдвиг с дополнением до двух.

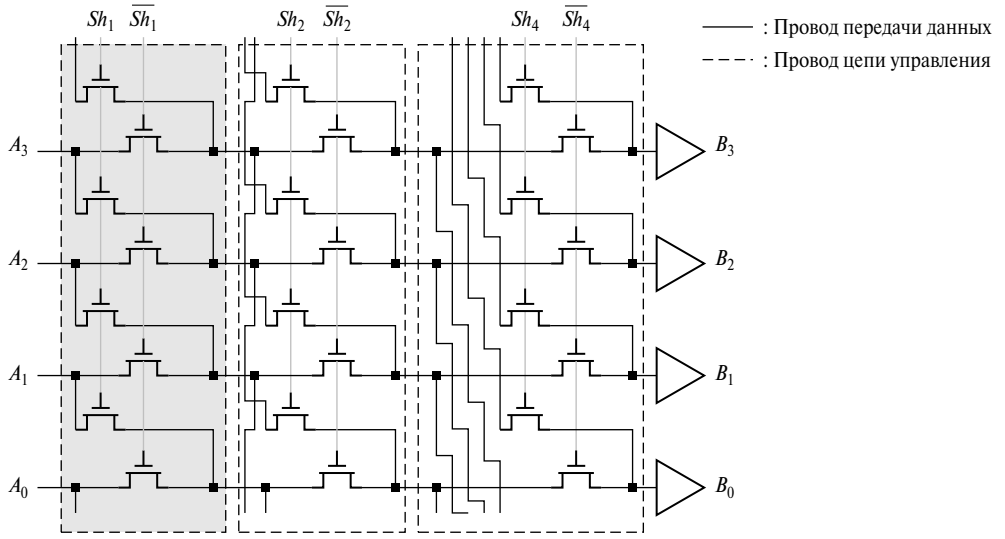


Рис. 11.38. Логарифмический сдвиговый регистр с максимальной величиной сдвига вправо, равной 7 битам. (Показаны только четыре младших разряда.)

11.5.2. Логарифмический сдвиговый регистр

В рассмотренном выше устройстве весь сдвиговый регистр был реализован в виде одной матрицы проходных транзисторов, а в логарифмическом сдвиговом регистре использован каскадный подход. Общая величина сдвига разбивается на меньшие сдвиги, представляемые степенями 2. Сдвиговый регистр максимальной ширины M состоит из $\log_2 M$ каскадов, причем i-й каскад либо сдвигает данные на 2^i позиции, либо передает их без изменений. На рис. 11.38 приведен пример сдвигового регистра с максимальной величиной сдвига, равной 7 битам. Например, чтобы выполнить сдвиг на 5 битов, первый каскад устанавливается в режим сдвига; второй — в режим передачи без изменений; третий — снова в режим сдвига. Обратите внимание на то, что контрольное слово для данного сдвигового регистра уже представлено в закодированном формате и дополнительный декодер не требуется.

Скорость логарифмического сдвигового регистра логарифмически зависит от величины смещения, поскольку M-битовый сдвиговый регистр требует $\log_2 M$ каскадов. Более того, последовательное соединение проходных транзисторов будет замедлять схему для бо́льших сдвигов. Следовательно, требуется аккуратно добавить в схему промежуточные буферы (см. главу 6).

В целом можно заключить, что многорегистровая циклическая реализация подходит для небольших сдвиговых регистров. При больших сдвигах более эффективен (с точки зрения площади и скорости) логарифмический сдвиговый регистр. Более того, логарифмический сдвиговый регистр легко параметризовать, т.е. возможно его автоматическое проектирование. Самое главное, что вы должны вынести из этого раздела: используя правильность структуры арифметического оператора, можно получить плотную и высокоскоростную реализацию схемы.

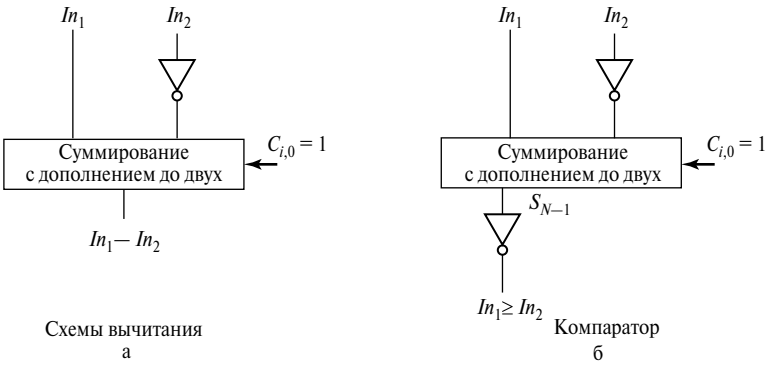


Рис. 11.39. Арифметическая структура, полученная из полного сумматора

11.6. Другие арифметические операторы

В предыдущих разделах мы рассмотрели только небольшое подмножество огромного набора арифметических цепей, требуемых при проектировании микропроцессоров и процессоров сигналов. Помимо сумматоров, умножителей и сдвиговых регистров часто требуются другие операторы — компараторы, делители, счетчики и тригонометрические операторы (синус, косинус, тангенс). Полный анализ данных схем выходит за рамки рассмотрения данной книги, поэтому, если вас интересует данная тема, обратитесь к прекрасным справочникам по этим вопросам (например, [Swartzlander90], [Oklobdzija01]).

Обращаем ваше внимание на то, что большинство идей проектирования, введенных в этой главе, применимы и к другим операторам. Например, компараторы можно спроектировать с различными зависимостями задержки от числа битов (линейной, типа квадратного корня, логарифмической). Фактически некоторые операторы представляют собой простые производные представленных ранее структур сумматора или умножителя. Например, на рис. 11.39 показано, как можно реализовать схему вычитания с дополнением до двух, объединив сумматор с дополнением до двух и инвертирующий каскад, или использовать схему вычитания для реализации $A \geq B$.

Задача 11.10. Компаратор

Спроектируйте логическую диаграмму компаратора, реализующего следующие логические функции: $\geq$, $=$ и $\leq$

Проектирование арифметико-логического устройства (АЛУ)

Ядром любого микропроцессора или микроконтроллера является *арифметико-логическое устройство*, АЛУ (Arithmetic-Logic Unit — ALU). В АЛУ сложение и вычитание объединяются с другими операциями, например сдвигом и побитовыми логическими операциями (И, ИЛИ и ИСКЛЮЧАЮЩЕЕ ИЛИ).

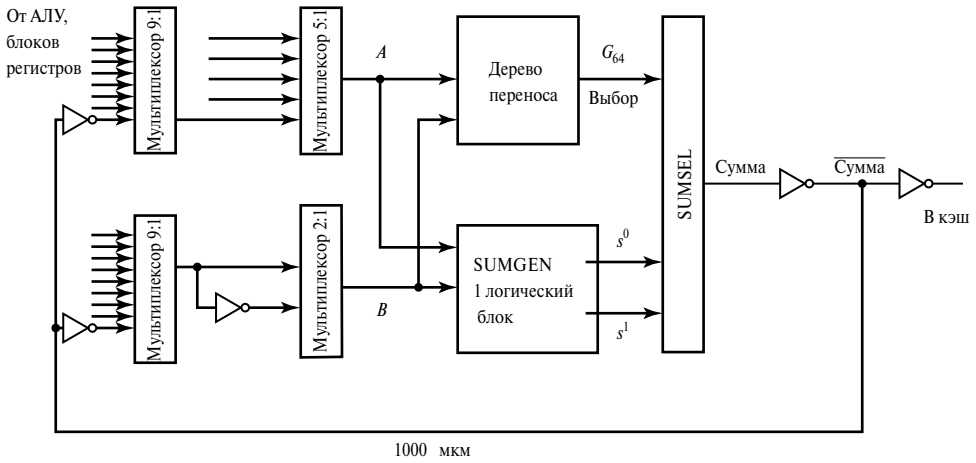


Рис. 11.40. 64-битовое арифметико-логическое устройство

На рис. 11.40 показано АЛУ 64-битового высокопроизводительного микропроцессора [Mathew01]. Оно состоит из двух уровней широких мультиплексоров, 64-битового сумматора, логического устройства, мультиплексора слияния операторов и драйвера шины с обратной записью. Первая пара мультиплексоров 9:1 выбирает данные из девяти различных входных источников, три из которых поставляют массивы регистров и кэш. Шесть остальных входов приходят непосредственно от шести АЛУ, — рассматриваемый микропроцессор может за один такт выдавать шесть целочисленных команд. В действительности по одной из этих линий выход АЛУ возвращается на собственный вход элемента, в результате чего создается контур, определяющий критический путь. Второй мультиплексор 5:1 на входе A выполняет частичный сдвиг оператора, а мультиплексор 2:1 на входе B инвертирует операнд для реализации вычитания. Сумматор выполнят сложение или вычитание операндов A и B с дополнением до двух. Вычитание с дополнением до двух выполняется путем инвертирования всех битов операнда B (с помощью мультиплексора 2:1) и присвоения значения 1 входному сигналу переноса. Блок SUMSEL не только реализует выбор суммы, но и выполняет слияние выходов арифметических и логических устройств. В конце схемы находится мощный буфер, который управляет шиной обратной связи, представляющей большую нагрузку для АЛУ. Например, для процессора Intel/HP Itanium [Fetzer02] шину нужно соединить со всеми шестью блоками, что при 0,18-микронной технологии означает трассировку более 2 мм проводников. Если к этому добавить нагрузку от блока регистров, то получим более 0,5 пФ емкостной нагрузки.

АЛУ представляет собой типичный пример разрядно-модульного проектирования. Все модули в каждом блоке согласуются по шагу, благодаря чему минимизируется вертикальная трассировка. На рис. 11.41 показан компоновочный план 64-битового АЛУ. Трассировка между ячейками выполняется горизонтально по третьему металлическому слою. Исключением являются

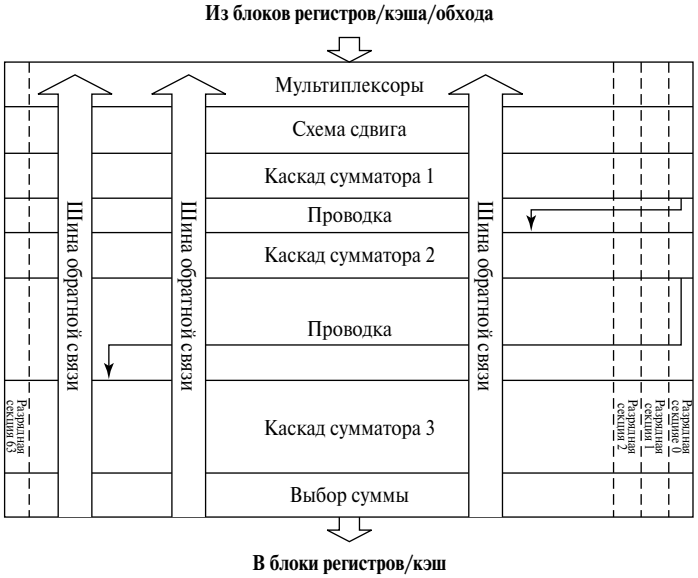


Рис. 11.41. Компоновочный план 64-битового АЛУ

только длинные горизонтальные проводники между 1-м и 2-м и между 2-м и 3-м каскадами сумматора, которые проходят по третьему и четвертому металлическим слоям. Например, если сумматор реализован как четверичный сумматор с ускоренным переносом, наиболее длинный провод проходит через 48 ячеек.

Проводка обратной связи прокладывается по верхнему металлическому слою (например, пятому). Если микропроцессор обслуживает несколько блоков, то данные шины распространяются по всем АЛУ, как показано на рис. 11.42. В нашем примере на вход АЛУ должно вернуться девять 64-битовых шин. Мало того, что данные проводники длинные, они еще разрабатываются с удвоенной шириной и, по крайней мере, с двойным шагом, чтобы уменьшить сопротивление и емкость.

11.7. Компромиссы между мощностью и скоростью в информационных каналах*

Обсуждая сумматоры, умножители и сдвиговые регистры, мы в основном акцентировали внимание на компромиссах между скоростью и площадью, не обращая внимания на вопросы потребления мощности. В данном разделе мы кратко изучим третье измерение пространства анализа разработки. Поскольку большинство подходов к минимизации потребляемой мощности уже рассматривалось в главе 6, приведенное ниже обсуждение является в основном иллюстрацией рассмотренных концепций.

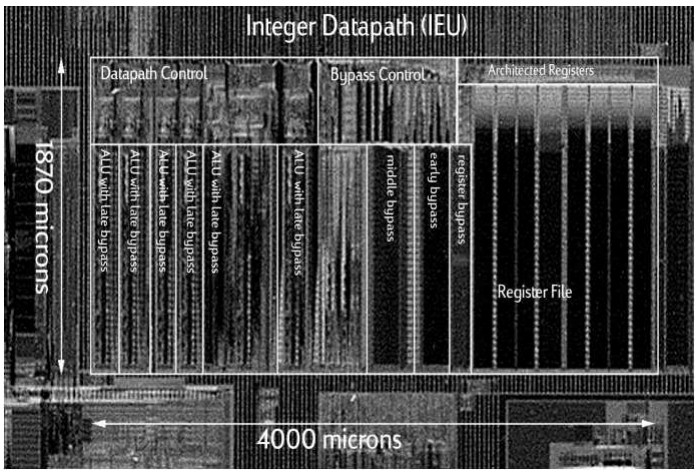


Рис. 11.42. Микрофотография процессора Itanium, на которой видна разрядно-модульная структура шести АЛУ, обводов, массива регистров и управляющей логики. (*Перепечатано с разрешения компании Intel.*)

В типичных цифровых разработках обычно накладываются ограничения либо на *время ожидания*, либо на *пропускную способность*. Структуры с условиями на время ожидания должны завершить вычисления до указанного предельного срока (например, приложения интерактивной связи и игры). С другой стороны, структуры с условиями на пропускную способность должны поддерживать указанную пропускную способность данных. Например, соединение 1000BaseT Gigabit Ethernet должно поддерживать постоянную пропускную способность в 1 Гбит/с. Исходя из условий к разработке, отличаются и техники архитектурной оптимизации, доступные разработчику. Конвейерная организация или параллельная обработка, например, эффективны для сценария с ограниченной пропускной способностью, но могут оказаться неприемлемыми для системы с условиями на время ожидания. Например, пропускная способность в 1 Гбит/с при передаче по медным проводам достигается в Gigabit Ethernet благодаря параллельной обработке четырех потоков данных, идущих со скоростью 250 Мбит/с.

При фиксированной архитектуре информационного канала существует множество компромиссов между скоростью, площадью и мощностью, реализуемых при выборе напряжения (напряжений) питания, порогов транзисторов и размеров устройств. В результате было разработано множество технологий минимизации мощности, сведенных в табл. 11.3. Данные технологии можно классифицировать следующим образом.

- **Время реализации.** Некоторые технологии разработки реализуются (или активизируются) *во время проектирования*. Например, во время проектирования фиксируется ширина и длина транзисторов. Напряжение питания и пороги транзисторов, наоборот, можно либо выбрать статически в фазе разработки, либо изменять динамически *во время выполнения*.

Таблица 11.3. Технологии минимизации мощности

	Постоянная пропускная способность	Переменная пропускная способность	
	Время проектирования	Спящий режим	Время выполнения
Активное рассеяние	Понижение V_{DD}, несколько V_{DD}, масштабирование транзисторов, оптимизация логики	Избирательное отключение	Динамическое масштабирование напряжения
Утечка	Несколько V_{Th}+ активные методы	Спящие транзисторы, переменное V_{Th}	Переменное V_{Th}+ активные методы

В других технологиях существует возможность перевода функции или модуля в холостой режим (режим пониженного энергопотребления). В таких случаях логично потребовать, чтобы рассеяние мощности модуля, находящегося в *спящем режиме*, было минимальным.

- **Целевой источник рассеяния.** Другая классификация технологий управления мощностью связана с источниками рассеяния мощности, и здесь различают активную (динамическую) мощность и утечку (статическую мощность). Например, очень привлекательной техникой является понижение напряжения питания: оно не только квадратично уменьшает энергию, потребляемую за один переход (правда, за счет быстродействия), но и снижает точки утечки. С другой стороны, повышение порогового напряжения влияет в основном на утечку.

Спящий режим работы заслуживает особого внимания. Если в холостом режиме цифровой блок по-прежнему получает тактовый сигнал, то соответствующая сеть распределения тактовых импульсов (плюс присоединенные триггеры) продолжает потреблять энергию (при том, что никакие вычисления не выполняются). Напомним, что сеть распределения тактовых импульсов потребляет примерно треть суммарной энергии цифровой системы. Наиболее распространенным методом снижения потребления мощности в холостом режиме является технология стробирования тактовых сигналов (см. главу 10). Согласно этому подходу, место подвода тактового сигнала к модулю отключается (стробируется), когда блок находится в холостом режиме. При этом данная схема не уменьшает мощность утечки блока, находящегося в холостом режиме. Для понижения мощности, потребляемой в холостом режиме, следует применять более сложные схемы, например увеличение порогов транзисторов или отключение шин электропитания.

В следующих разделах мы подробно обсудим все названные выше технологии времени проектирования и времени выполнения.

11.7.1. Технологии уменьшения мощности (время проектирования)

Уменьшение напряжения питания

Снижение напряжения питания приводит к квадратичной экономии мощности, а следовательно, является весьма привлекательным подходом. К его недостаткам относится то, что задержка КМОП-элементов обратно пропорциональна напряжению питания. Правда, на уровне информационного канала данное снижение быстродействия можно компенсировать иными средствами, например с помощью логической и архитектурной оптимизации. Например, сумматор со сквозным переносом можно заменить более быстрой структурой, скажем сумматором с ускоренным переносом. Эта реализация больше и сложнее, т.е. имеет бо́льшую физическую и переключаемую емкость. Как правило, это вполне приемлемая плата за то, что более быстрый сумматор сможет работать при более низком напряжении питания, показывая при этом то же самое быстродействие. Кроме того, как показано в примере 11.7, можно выполнить архитектурную оптимизацию, которая позволит компенсировать эффект снижения напряжения V_{DD}.

Пример 11.7. Минимизация потребления мощности с помощью параллельной обработки

Чтобы проиллюстрировать, как архитектурная оптимизация позволяет компенсировать сниженную скорость, проанализируем простой восьмибитовый информационный канал, состоящий из сумматора и компаратора, предполагая, что используется 2-микронная КМОП-технология [Chandrakasan92]. Как показано на рис. 11.43, входы A и B суммируются, а результат сравнивается с входом C. Предположим, что наихудшая задержка прохождения сигнала через сумматор, компаратор и защелку приблизительно равна 25 нс при напряжении питания 5 В. В лучшем случае система допускает длительность такта $T = 25$ нс. Если требуется максимальная пропускная способность, очевидно, что рабочее напряжение уменьшать нельзя, поскольку дополнительное увеличение задержки недопустимо. Используем показанную топологию информационного канала в качестве отправной точки и рассчитаем, как можно улучшить его энергопотребление. Средняя мощность, потребляемая исходным информационным каналом, выражается следующим образом:

$$P_{ref} = C_{ref} V_{ref}^2 f_{ref}, \tag{11.26}$$

где C_{ref} — суммарная эффективная емкость, переключаемая за один такт. Эффективную емкость можно определить, усреднив энергию по равномерно распределенной последовательности случайных входных сигналов.

Один из способов сохранения пропускной способности при снижении напряжения питания — использовать параллельную архитектуру. Как показано на рис. 11.44, можно параллельно соединить два идентичных информационных канала “сумматор–компаратор”, позволив каждому блоку работать с по-

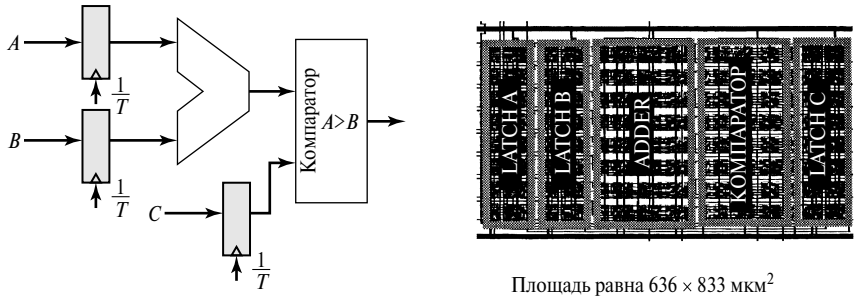


Рис. 11.43. Простой информационный канал и соответствующая топология

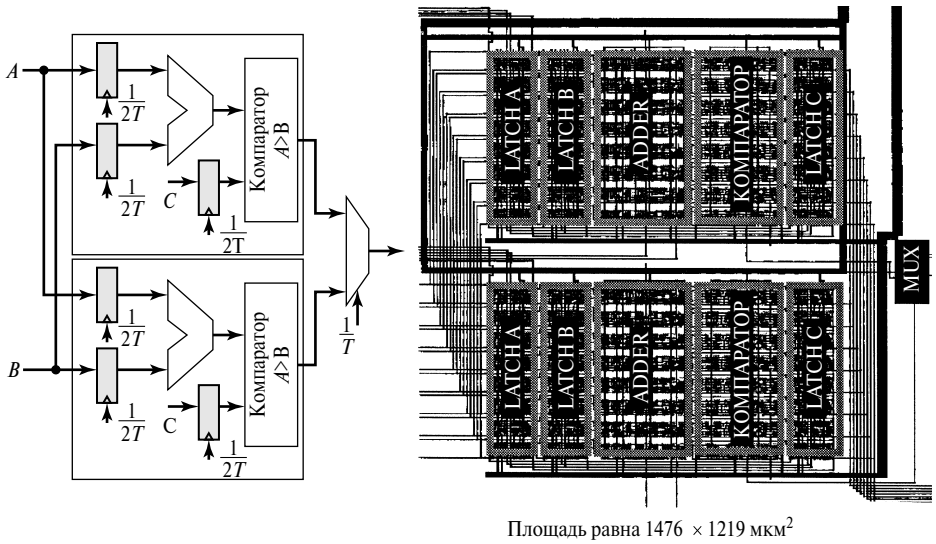


Рис. 11.44. Параллельная реализация простого информационного канала

ловиной исходной тактовой частоты и сохранив исходную пропускную способность. Поскольку требования к скорости сумматора, компаратора и защелки снизились с 25 до 50 нс, напряжение можно понизить с 5 до 2,9 В (напряжение, при котором удваивается задержка). Хотя мы вдвое увеличили емкость информационного канала, вдвое была уменьшена и рабочая частота. К сожалению, наблюдается еще небольшое увеличение суммарной эффективной емкости, вызванное дополнительной трассировкой и уплотнением данных. Это приводит к увеличению емкости в 2,15 раза. В результате мощность параллельного информационного канала равна:

$$P_{par} = C_{par}V_{par}^2 f_{par} = (2,15C_{ref})(0,58\ V_{ref})^2 \left(\frac{f_{ref}}{2}\right) \approx 0,36P_{ref}. \qquad (11.27)$$

Представленный подход позволяет *снизить мощность за счет увеличения площади* (примерно в 3,4 раза по сравнению с исходным вариантом). Его рекомендуется использовать только в том случае, если в проекте не предусмотрены жесткие условия на площадь. Более того, параллелизм чреват большими расходами на трассировку, которая может вызвать дополнительное рассеяние мощности. В общем, для минимизации этого недостатка потребуется еще один цикл оптимизации.

Параллелизм — не единственный способ, позволяющий компенсировать потерю быстродействия. Этой же цели можно добиться с помощью других архитектурных подходов, например конвейерной обработки. Самое важное, что нужно вынести из сказанного выше: если рассеяние мощности является критичным параметром схемы, то наиболее эффективным средством борьбы с ним будет понижение напряжения питания. Сопутствующее снижение быстродействия можно компенсировать за счет увеличения площади. Довольно часто этот компромисс является приемлемым, поскольку площадь — не такой серьезный параметр благодаря существенному увеличению за последние десять лет числа уровней интеграции.

Задача 11.11. Уменьшение напряжения питания с помощью конвейера

В информационном канале, показанном на рис. 11.44, между сумматором и компаратором введен конвейерный каскад. Исходя из схемы, можно предположить, что это примерно вдвое уменьшает задержку распространения логического блока, при этом увеличивая емкость примерно на 15%. Очевидно, что дополнительный конвейерный регистр требуется также на входе C. Определите, какой экономии мощности можно добиться с помощью этого подхода, учитывая, что пропускная способность должна оставаться такой же, как у исходного информационного канала. Прокомментируйте получившееся увеличение площади.

Использование нескольких напряжений питания

Снижение напряжения питания является достаточно понятным подходом, который, впрочем, не всегда дает оптимальные результаты. Уменьшение напряжения питания в конечном счете понижает рассеяние мощности во всех логических элементах, увеличивая при этом их задержку. Гораздо лучшим подходом является избирательное уменьшение напряжения питания лишь в следующих элементах:

- в блоках, находящихся на быстрых путях и рано завершающих вычисления;
- в блоках, логические элементы которых управляют большими емкостями и больше всего выигрывают от увеличения задержки.

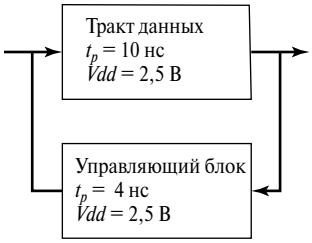


Рис. 11.45. Структура с разными длинами критических путей

Отметим, что данный подход требует использования нескольких напряжений питания. Вообще, в современных интегральных схемах такое решение применяется довольно широко. Со схемами ввода-вывода (по соображениям совместимости) соотносится одно значение напряжения питания, а в кольцевом регистре ввода-вывода применяются транзисторы, имеющие более толстый слой подзатворного окисла, рассчитанный на высокие напряжения. Логическое ядро питается более низким напряжением, и в нем используются транзисторы с более тонким слоем окисла. Данный подход позволяет понизить рассеяние мощности в схеме. Например, любой модуль может выбирать наиболее подходящее напряжение из двух (или большего числа) имеющихся. Мало того, выбор подходящего напряжения питания можно реализовать даже на уровне отдельных логических элементов.

Выбор напряжения на уровне модулей

Рассмотрим, например, цифровую систему показанную на рис. 11.45 и состоящую из информационного канала с критическим путем 10 нс и управляющего блока с более коротким критическим путем, 4 нс, которая питается напряжением 2,5 В. Кроме того, предположим, что блок информационного канала имеет фиксированные время задержки и пропускную способность и для понижения напряжения источника питания нельзя использовать никакие архитектурные преобразования. Поскольку контрольный блок рано завершает свою работу (другими словами, имеет небольшой *резерв времени*), его напряжение питания можно понизить. При уменьшении напряжения до $V_{DDL} = 1$ В задержка критического пути увеличивается до 10 нс, а рассеяние мощности сокращается более чем в пять раз[6]. Таким образом, мы эффективно использовали разницу в длинах критических путей разных модулей (*резерв времени*) для избирательного понижения потребляемой мощности модулей с бо́льшим резервом времени.

При использовании в кристалле нескольких напряжений питания следует помнить о *преобразователях уровня*, которые нужны в каждой точке, где модуль с низким напряжением питания подает сигнал на вход элемента с более высоким напряжением питания. Если элемент, питаемый напряжением V_{DDL}, управляет элементом с напряжением V_{DDH}, p-МОП-транзистор нико-

[6] Здесь мы для упрощения предположили, что задержка распространения обратно пропорциональная напряжению питания.

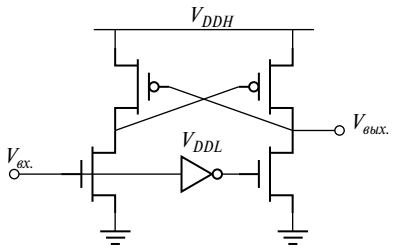


Рис. 11.46. Преобразователь уровня

гда не выключится, что, как показано на рис. 11.46, приведет к появлению статического тока и снизит общий размах напряжения на выходе. Решить эти проблемы можно с помощью преобразования уровня, выполняемого на границе раздела между блоками с разным напряжением. На рис. 11.46 показан асинхронный преобразователь уровня, основанный на структуре DVSL (см. главу 6) и подобный элементу с малым размахом, изображенному на рис. 9.32. Пара p-МОП-транзисторов преобразует уровень за счет положительной обратной связи. Задержка данного преобразователя уровня довольно чувствительна к масштабированию транзисторов и напряжения питания. Показанные на схеме n-МОП-транзисторы работают с пониженной перегрузкой, $V_{DDL} - V_{Th}$, по сравнению с p-МОП-устройствами. Чтобы они могли перегрузить положительную обратную связь, их нужно сделать бо́льшими. Для низкого значения V_{DDL} задержка может стать довольно большой. Кроме того, из-за сниженной перегрузки схема также весьма чувствительна к колебаниям напряжения питания. Обратите внимание на то, что для переходов с понижением напряжения преобразователи уровня не требуются.

Служебные издержки, связанные с реализацией преобразования уровня, можно иногда уменьшить, заметив, что большинство преобразований выполняется на границе регистра. Например, все управляющие входы информационного канала, показанного на рис. 11.45, выбираются в целом одним регистром. Следовательно, в таком случае преобразование уровня логично реализовать внутри регистра. Такой регистр с преобразованием уровня показан на рис. 11.47. Он представляет собой обычную реализацию на передаточных вентилях пары защелок “master–slave”, где в защелке “slave” с помощью пары перекрестно-включенных p-МОП-транзисторов выполняется преобразование уровня.

Использование нескольких напряжений питания внутри блока

Описанный выше подход можно применить и к гораздо меньшим элементам, установив индивидуальное напряжение питания для каждой ячейки внутри блока [Usami95], [Hamada01]. Изучая гистограмму задержек критических путей для типичного цифрового блока, можно обнаружить, что критическими (или близкими к ним) являются несколько путей, а подавляющее множество путей обладает гораздо меньшими задержками. По сути, более короткие пути растрачивают энергию, поскольку никакого “приза” за быстрое завершение работы не существует. На каждом из таких путей напряжение

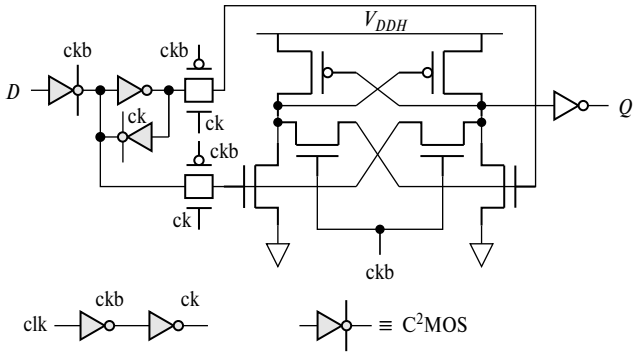


Рис. 11.47. Регистр с преобразованием уровня. Напряжение питания затененных элементов равно V_{DDL} [Usami95]

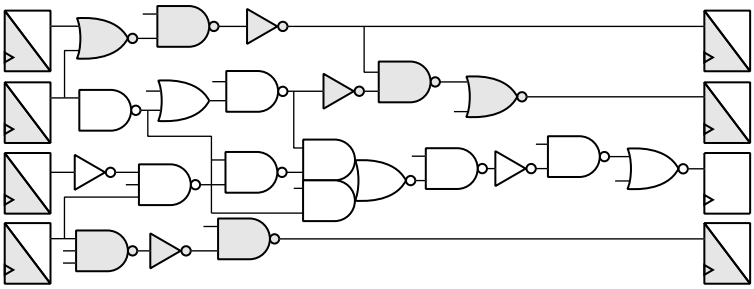


Рис. 11.48. Структура с двумя напряжениями питания, в которой реализовано кластерное масштабирование напряжения. Каждый путь начинается с высокого напряжения и переключается на низкое напряжение (серые логические элементы) при наличии резерва времени. Преобразование уровней выполняется в регистрах

питания следует уменьшать до оптимального уровня. Минимальное потребление энергии достигается тогда, когда критическими становятся все пути. Впрочем, добиться этого не так просто, поскольку многие логические элементы используются совместно несколькими путями, а, кроме того, генерировать и распространять слишком большое число разных напряжений питания непрактично. На рис. 11.48 показана гораздо более логичная реализация, в которой задействовано всего два напряжения питания. Согласно технологии *кластерного масштабирования напряжения*, каждый путь начинается с источника высокого напряжения и переключается на источник с низким напряжением при появлении резерва времени. Преобразование уровней происходит в регистрах в конце пути с помощью схем, подобных показанной на рис. 11.47. Обратите внимание на то, что преобразование уровня необходимо только тогда, когда логический блок, следующий за данным, никак не может отказаться от низкого напряжения.

Влияние этого подхода на быстродействие схемы иллюстрируется на рис. 11.49, где показано распределение длин путей для типичного логиче-

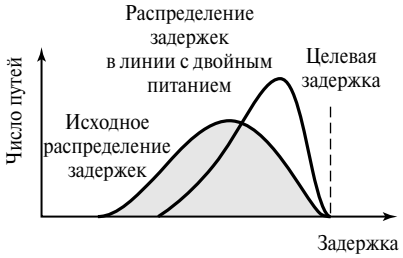


Рис. 11.49. Типичное распределение задержек путей до и после введения второго уровня напряжения

ского блока с одним и двумя уровнями напряжения питания. При использовании одного источника напряжения значительная часть путей существенно короче критического значения. Добавление второго уровня напряжения питания смещает профиль распределения задержек ближе к целевой задержке, т.е. больше путей становятся критическими. Если исходная структура имела незначительное число критических путей и обработка в большинстве путей завершалась рано, в результате модификации можно получить существенную экономию энергии. С другой стороны, если большинство путей внутри блока уже были критическими, введение второго источника напряжения не даст никакого заметного результата. Возможно, вы думаете, что введение третьего или даже четвертого напряжения питания может дать еще большую экономию мощности. К сожалению, в многочисленных исследованиях было показано, что для типичных распределений задержек по путям добавление большего числа уровней напряжения дает весьма незначительный выигрыш [Hamada01], [Augsburger02].

Подобный образом, может возникнуть вопрос, дает ли техника двойного питания больший выигрыш, чем равномерное уменьшение напряжения питания. При использовании кластерного масштабирования напряжения двойное питания более эффективно, если ближе к концу блока сосредоточены большие емкости (как в буферных цепочках) [Stojanovic02]. Данное наблюдение согласуется с концепцией шин с низким размахом (см. главу 9). Обычно шина представляет собой наибольшую емкость в цепи, поэтому выгоднее всего понизить напряжение ее задающего генератора, чем напряжение любого другого логического элемента. Поскольку шина имеет наибольшую емкость, в этом случае мы получим наибольшую экономию мощности при том же увеличении задержки.

Необходимость распределения в кристалле нескольких уровней напряжения питания усложняет проектирование схемы разводки питания и мощности и часто является непосильной задачей для традиционных средств размещения элементов и трассировки соединений. Как неоднократно было показано выше, минимизировать это негативное влияние можно с помощью структурированного подхода. Например, можно разместить ячейки с различными напряжениями питания в разных строках стандартной топологии ячеек (рис. 11.50). Второй уровень напряжения можно подвести извне кристалла,

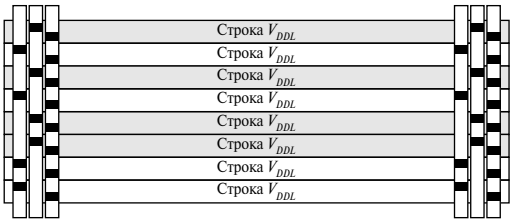


Рис. 11.50. Топология логического блока с двойным питанием, в котором все ячейки, относящиеся к разным источникам питания, располагаются в разных строках

или же он может генерироваться внутри кристалла с помощью внутреннего преобразователя постоянного тока. Эффективность подобных преобразователей превышает 90% и сопровождается минимальными служебными издержками. Разумеется, схема разводки напряжения V_{DDL} также должна удовлетворять всем требованиям к разработке, например, использовать развязывающие элементы для минимизации перепадов напряжения и повышения сопротивления электромиграции.

Использование устройств с несколькими порогами

Другой способ понижения мощности за счет скорости — использование устройств с несколькими порогами. Большинство КМОП-технологий с размером элемента меньше 0,25-мкм предлагает два типа транзисторов n- и p-типа, пороги которых отличаются примерно на 100 мВ. Эти устройства с более высокими порогами характеризуются током утечки, примерно на порядок меньшим, чем у устройств с низкими порогами. Платой за это является снижение активного тока примерно на 30%. Следовательно, устройства с низким порогом желательно использовать на критичных по времени путях, а устройства с высоким порогом — во всех остальных случаях. Подобное распределение можно выполнить на уровне ячеек, а не на уровне отдельных транзисторов [Wei99], [Kato00]. Обратите внимание на то, что использование нескольких порогов не требует преобразователей уровня или любых других специальных схем (рис. 11.51). Кластеризация логики также не требуется; заменять устройства с низким порогом устройствами с высоким порогом можно до тех пор, пока не будет использован весь резерв времени. Правильно решив эту задачу, можно уменьшить мощность утечки на 80–90%.

Хотя основной задачей использования устройств с несколькими порогами является борьба с током утечки, в качестве дополнительного преимущества данный подход немного уменьшает активную мощность. В основном это вызвано снижением емкости затвор-канал в закрытом состоянии и небольшим уменьшением размаха сигнала на внутренних узлах элемента ($V_{DD} - V_{ThH}$). Подобное уменьшение частично компенсируется увеличением емкостей боковых переходов истока и стока. В целом общее уменьшение активной мощности составляет порядка 4%.

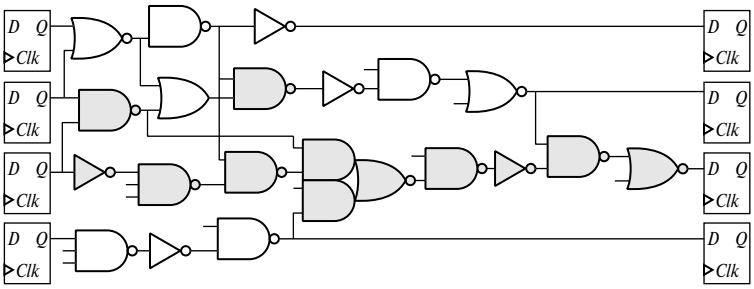


Рис. 11.51. Использование элементов с двумя порогами. Элементы, закрашенные серым цветом, используют транзисторы с низким порогом и применяются только на критических путях

Уменьшение переключаемой емкости с помощью масштабирования транзисторов

Входная емкость комплементарного МОП-элемента прямо пропорциональна его размеру, а следовательно, и его скорости. В главе 6 мы исследовали оптимальное масштабирование логических элементов и нашли, что каждый элемент цепочки логических схем должен масштабироваться так, чтобы его эффективное разветвление по выходу составляло порядка 4, поскольку это дает минимальную задержку для этого пути. В связи с этим возникает интересный вопрос: как масштабировать схему, чтобы получить минимальную энергию при незначительном повышении допустимой задержки?

Известно, что для цепочки инверторов с данной нагрузкой и числом каскадов минимальная задержка достигается, когда размер каждого инвертора в цепи является геометрическим средним его соседей:

$$s_i^2 = s_{i+1}s_{i-1}. \tag{11.28}$$

Если минимальная задержка, получаемая таким способом, меньше требуемой, можно сформулировать задачу оптимизации, которая минимизирует переключаемую емкость за счет допустимого увеличения задержки. Другой вариант: можно уменьшить количество каскадов и увеличить масштабный коэффициент, как предлагалось в главе 9. Вообще, у данной задачи существует аналитическое решение [Ma94], из которого следует, что оптимальным является выбор масштабного коэффициента каскадов по следующей формуле:

$$s_i^2 = \frac{s_{i+1}s_{i-1}}{1 + \mu s_{i-1}}. \tag{11.29}$$

Параметр μ представляет собой неотрицательное число, зависящее от величины доступного резерва времени (относительно цепочки, масштабированной для получения минимальной задержки). Интуитивно данное решение довольно понятно — поскольку последние каскады цепочки инверторов являются самыми большими, они и будут первыми кандидатами на уменьшение. Уменьшение размеров любого инвертора в цепи в некоторое

число раз дает такое же увеличение задержки. Следовательно, лучше всего это сделать в каскаде, который даст нам наибольший выигрыш энергии (т.е. в наибольшем).

Тот же принцип применим и к цепочкам логических элементов. Если оптимизировать ее для получения минимальной задержки, то задержки всех каскадов будут одинаковыми (причем собственные задержки логических элементов могут отличаться). Следовательно, при последующей оптимизации нужно в первую очередь увеличивать задержку самого энергоемкого элемента. Данную идею можно расширить на оптимизацию энергии/задержки произвольного блока комбинационной логики. Следует, правда, помнить, что ее применение на практике не является тривиальным. Уменьшение размеров элементов одного пути влияет на задержку всех путей, имеющих с данным путем общие логические элементы. Из-за этого изолировать и оптимизировать какой-то конкретный путь достаточно трудно. Кроме того, многие пути произвольного комбинационного блока являются сходящимися.

Пример 11.8. Компромиссы между энергией и задержкой при проектировании сумматоров

Рассмотрим оптимизацию энергии и задержки 64-битового сумматора Когга–Стоуна. Существует множество путей через сумматор, причем далеко не все они сбалансированы. Возникает важный вопрос: как определить пути, подлежащие масштабированию? Один из вариантов — выбрать все пути сумматора, равные критическому. Поскольку пути через сумматор, грубо говоря, соответствуют различным разрядам, сопоставим каждый логический элемент сумматора с разрядом. Всего имеется 64 разряда и девять каскадов логики. На рис. 11.52, *а* для такого разбиения показано распределение энергии, соответствующее реализации с минимальной задержкой. Видно, что сумматор потребляет наибольшее количество энергии вдоль наиболее длинных путей переноса. На рис. 11.52, *б* показано распределение энергии того же сумматора после масштабирования, приведшего к увеличению задержки на 10%. При этом суммарное рассеяние энергии было уменьшено на 54%. Для сравнения, использование двух уровней напряжения питания дает выигрыш всего в 27%, а простое уменьшение напряжения питания — 22% [Stojanovic02]. Следовательно, **масштабирование является эффективным методом снижения мощности в информационных каналах, где основная энергия уходит на потребление внутри блока, а не на управление внешней нагрузкой.**

Уменьшение переключаемой емкости с помощью логической и архитектурной оптимизации

Поскольку эффективная емкость — это произведение физической емкости и активности переключения, рекомендуется минимизировать оба фактора. Уменьшить активность без ущерба для быстродействия можно с помощью

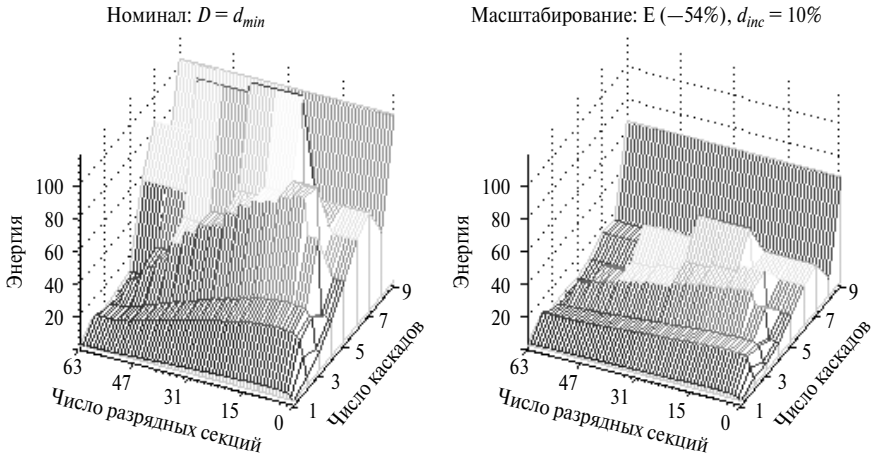


Рис. 11.52. Распределение энергии в 64-битовом сумматоре: *а* — масштабированного для получения минимальной задержки; *б* — масштабированного для получения минимальной энергии при увеличении задержки на 10%

огромного числа методов логической и архитектурной оптимизации. Некоторые из этих методов уже вводились в главе 6. Для краткости мы ограничимся рассмотрением лишь небольшого числа характерных методов оптимизации. Более подробный обзор можно найти в [Rabaey96] и [Chandrakasan98].

Уменьшение активности переключения с помощью распределения ресурсов

Уплотнение нескольких операций в одном аппаратном блоке существенно влияет на потребление мощности. Помимо увеличения физической емкости, это также может повысить активность переключения. Сказанное иллюстрирует эксперимент, представленный на рис. 11.53, где сравнивается потребление мощности двух счетчиков, запущенных одновременно. В первом случае оба счетчика запущены на разных аппаратных блоках, а во втором они уплотняются в одном устройстве. На рис. 11.53, *б* показано число переключений как функция расфазировки сигналов двух счетчиков. Без уплотнения характеристика всегда получается лучше, за исключением единственного случая, когда оба счетчика работают абсолютно синхронно. Уплотнение несколько рандомизирует сигналы данных, подаваемые на функциональный блок, что приводит к увеличению активности переключения. Если для схемы важен вопрос потребляемой мощности, то желательно избегать чрезмерного повторного использования ресурсов. Обратите внимание на то, что при работе в холостом режиме аппаратные КМОП-устройства потребляют очень небольшие объемы мощности. Реализация специализированных операторов приводит лишь к увеличению площади, улучшая при этом быстродействие и потребляемую мощность.

Уменьшение импульсных помех с помощью балансировки путей

Динамические, или импульсные, помехи являются основным источником рассеяния энергии в таких сложных структурах, как сумматоры и умножи-

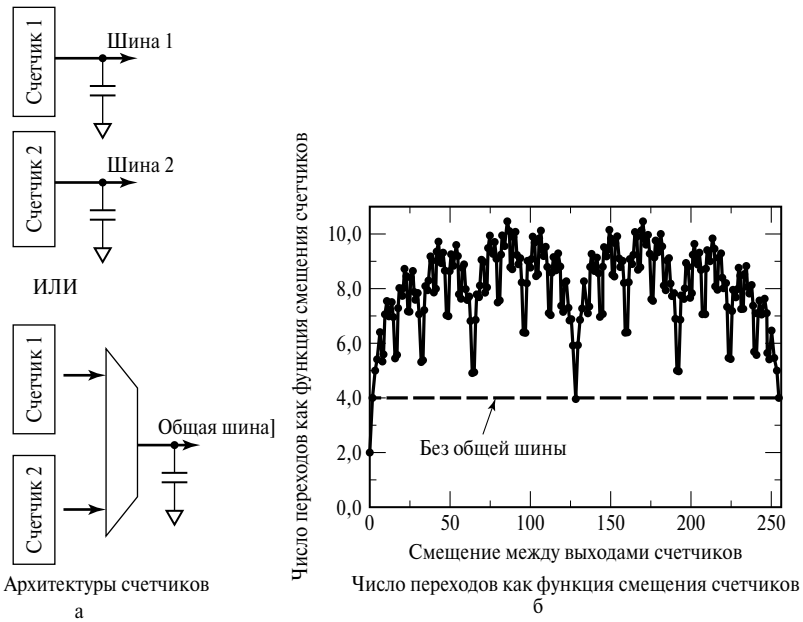


Рис. 11.53. Уплотнение увеличивает активность переключения

тели. Большое расхождение в длинах путей сигналов может в некоторых случаях стать источником паразитных импульсных помех. Сказанное демонстрируется на рис. 11.54, где показана смоделированная характеристика 16-битового сумматора со сквозным переносом, на всех входах которого одновременно происходит переход $0 \rightarrow 1$. Число битов суммы показано как функция времени. Сигналы суммы должны быть нулевыми для всех битов. К сожалению, во всех выходах присутствует значение 1, поскольку сигналу переноса требуется значительное время для прохождения от первого до последнего бита. Обратите внимание на то, как помеха становится все более явной по мере распространения по цепочке.

Существенно снизить активность импульсных помех можно, выбрав структуры со сбалансированными путями сигналов. Данным свойством обладают каскадные сумматоры с ускоренным переносом (например, Когга–Стоуна) и каскадные умножители; следовательно, с точки зрения потребляемой мощности они являются более приемлемыми даже при наличии большой физической емкости. Изучая структуру, показанную на рис. 11.22, находим, что временны*е* пути ко входам операторов "точка" имеют схожую длину, хотя из-за отличий в нагрузке и разветвлении по выходу могут наблюдаться некоторые расхождения.

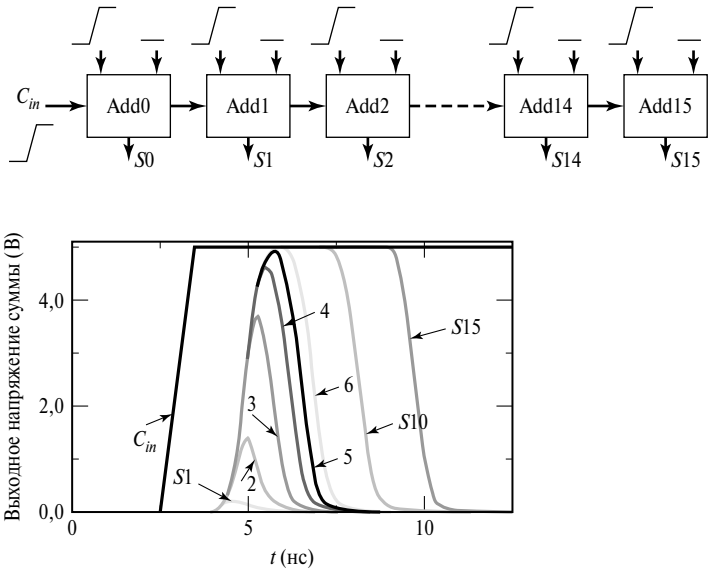


Рис. 11.54. Импульсная помеха в битах суммы 16-битового сумматора со сквозным переносом

11.7.2. Управление мощностью во время выполнения

Динамическое масштабирование напряжения питания

Как говорилось выше, статическое уменьшение напряжения питания понижает расход энергии на одну операцию и увеличивает срок действия батарей за счет производительности. Данный недостаток часто является неприемлемым, особенно в приложениях с ограниченным временем ожидания. Впрочем, существенное снижение мощности можно получить, заметив, что максимальная производительность требуется не всегда. Рассмотрим, например, универсальный процессор, используемый в переносных структурах (ноутбуках, органайзерах и мобильных телефонах). Вычислительные функции подобного процессора делятся на три основные категории: задачи, требующие интенсивных вычислений; задачи, выполняемые с низкой скоростью; работа в холостом режиме. Задачи, требующие интенсивных вычислений и малого времени ожидания, требуют, чтобы процессор работал с полной пропускной способностью (например, воспроизведение видео в формате MPEG и сжатие аудио). Задачи, выполняемые с низкой скоростью и допускающие большое время ожидания (обработка текста, ввод данных, резервное копирование памяти), требуют лишь части максимальной пропускной способности микропроцессора. Никакой пользы от быстрого завершения подобных вычислений нет, наоборот, это свидетельствует о пустой трате энергии. Наконец, подобные процессоры проводят значительную часть времени в холостом режиме, ожидая действий пользователя или внешнего запускающего события. В общем, вычислительная пропускная способность и время ожидания, требуемые от процессора, являются переменными в очень широком диапазоне.

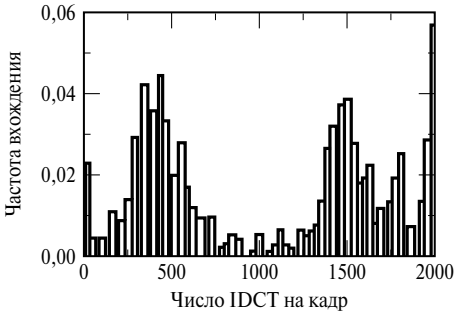


Рис. 11.55. Типичная гистограмма ОДКП для декодирования MPEG

Даже операции, требующие интенсивных вычислений, подобные декодированию MPEG, характеризуются переменными требованиями в зависимости от обрабатываемых данных. Например, то, сколько раз декодер MPEG вычисляет обратное дискретное косинус-преобразование (ОДКП) на один кадр видео, существенно зависит от количества движения на сценах. Сказанное иллюстрируется на рис. 11.55, на котором представлено распределение числа ОДКП на кадр для типичной видеопоследовательности. Процессор, выполняющий данный алгоритм, испытывает разную нагрузку в зависимости от обрабатываемого кадра.

Снижение тактовой частоты при выполнении задач, требующих меньшей нагрузки, уменьшает потребляемую мощность, однако не экономит энергию, — любая операция по-прежнему выполняется при высоком уровне напряжения. Тем не менее, если одновременно снизить напряжение питания и частоту, потребление энергии будет уменьшено. Чтобы поддерживать требуемую пропускную способность для высокой нагрузки и минимизировать энергию при низкой нагрузке, напряжение питания и частоту нужно динамически изменять, согласно требованиям выполняемого в данный момент приложения. Данная технология получила название *динамического масштабирования напряжения* (Dynamic Voltage Scaling — DVS). Суть концепции иллюстрируется на рис. 11.56 [Burd00], [Gutnik97] и заключается в том, что функция всегда должна **работать при самом низком напряжении питания, удовлетворяющем временны́м условиям**.

Концепция динамического масштабирования напряжения стала возможной благодаря следующему наблюдению: задержки большинства КМОП-схем и функций имеют очень похожее значение в типичном диапазоне напряжений питания, что является необходимым условием для работы системы с переменным напряжением питания. На рис. 11.57 показаны задержки нескольких характерных КМОП-блоков (элементы НЕ-И, кольцевые генераторы, массивы регистров и статические ОЗУ) в диапазоне напряжений питания от 1 до 4 В [Burd00]. Видим, что быстродействие всех устройств описывается очень похожими функциями. Обратите внимание на то, что некоторые семейства схем (например, логика на проходных n-МОП-транзисторах) не имеют указанного поведения во всем диапазоне напряжений питания.

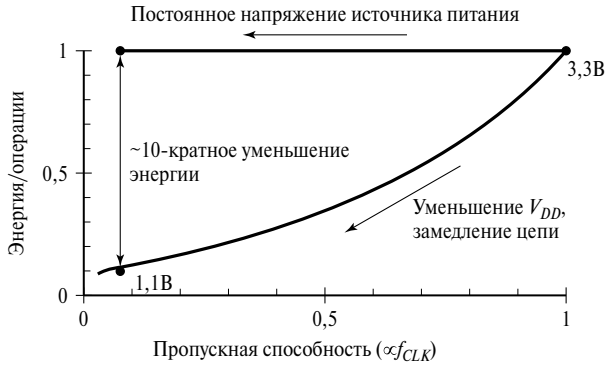


Рис. 11.56. Зависимость энергии на одну операцию от пропускной способности ($1/T$) для работы при постоянном и динамически меняющемся напряжении

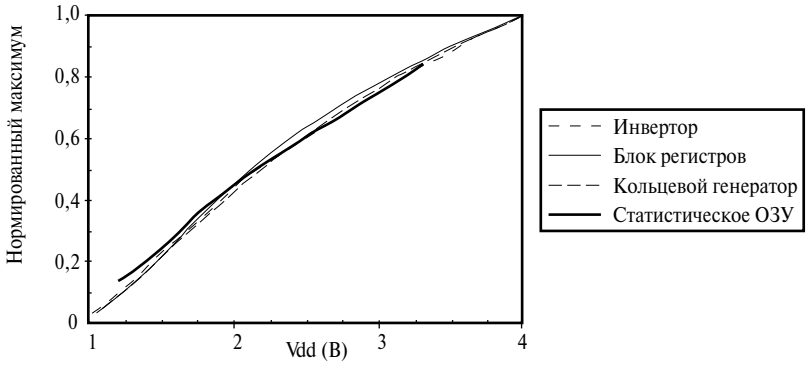


Рис. 11.57. Задержка типичных КМОП-функций как функция напряжения питания для 0,6-микронной КМОП-технологии

Теперь практическая реализация системы с динамическим масштабированием напряжения состоит из следующих компонентов:

- процессора, который может работать в широком диапазоне напряжений питания;
- цепи регулирования напряжения, устанавливающей минимальное напряжение, требуемое для работы при заданной частоте;
- операционной системы, рассчитывающей требуемые частоты для получения заданной пропускной способности и крайних сроков завершения задач.

Одна из возможных реализаций подобной системы показана на рис. 11.58. Ядром системы является кольцевой генератор, частота генерации которого согласуется с характеристиками критического пути микропроцессора. При включении в контур управления напряжения питания данный кольцевой ге-

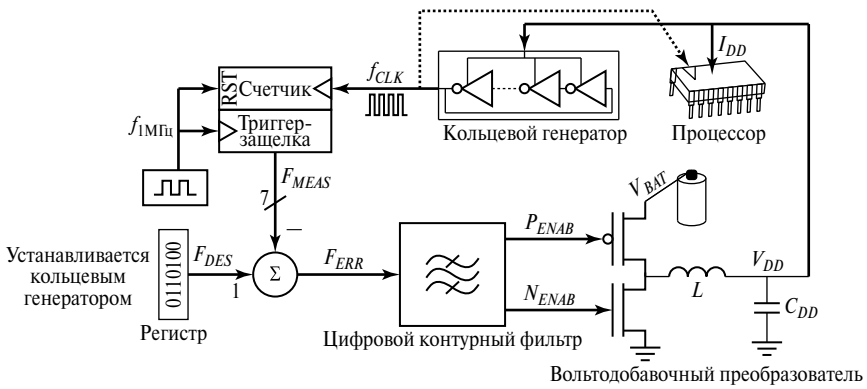


Рис. 11.58. Блок-схема системы с динамическим масштабированием напряжения

нератор обеспечивает перевод напряжения питания в тактовую частоту. Операционная система задает требуемую частоту (F_{DES}). Текущее значение частоты генератора измеряется и сравнивается с требуемой частотой. Разница используется как ошибка в цепи обратной связи. Настраивая напряжение питания, управляющий контур изменяет частоту кольцевого генератора, пока ошибка не станет равна 0.

Задачей планировщика (или операционной системы реального времени) является динамическое определение оптимальной частоты (или напряжения) как функции вычислительных требований всех активных задач системы. В более сложном случае (универсальный процессор) с каждой задачей должно соотноситься предельное время выполнения (например, частота смены видеокадров) или желательная частота выполнения. После этого планировщик напряжения оценивает число тактов процессора, необходимых для выполнения каждой задачи, и вычисляет оптимальную частоту процессора [Pering99]. Если нужно спланировать одну задачу с переменными требованиями к производительности, то для определения вычислительной нагрузки и соответствующей настройки напряжения можно использовать очередь. Сказанное иллюстрируется на рис. 11.59, где для установки напряжения питания (и частоты) потокового процессора сигналов используется глубина входной очереди [Gutnik97].

Динамическое масштабирование порога

По аналогии с динамическим изменением напряжения питания возникает соблазн динамически настраивать пороговое напряжение транзисторов. Для вычислений с малым временем ожидания порог следует понижать до минимального значения; для вычислений с низкой скоростью его можно повысить; в режиме ожидания его необходимо установить равным максимально возможному значению, чтобы минимизировать ток утечки. “Ручкой управления”, позволяющей динамически изменять пороговые напряжения, является смещение подложки. Для реализации этого транзисторы необходимо исполь-

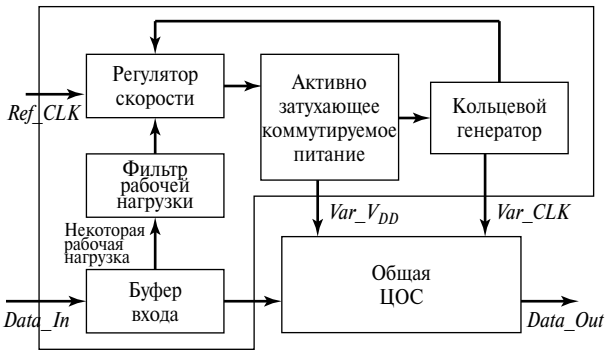


Рис. 11.59. Определение с помощью очереди рабочей нагрузки на процессор сигналов [Gutnik97]

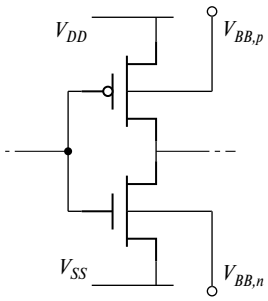


Рис. 11.60. Инвертор с выводами подложки, используемыми для управления порогом

зовать как четырехполюсники. Это возможно только в процессе с тремя карманами, поскольку, как показано на рис. 11.60, требуется возможность независимого управления всеми четырьмя полюсами n- и p-устройств. Смещение подложки можно реализовать для всего кристалла или выборочно для некоторых блоков или ячеек. Отметим, что реализация смещения подложки для отдельных ячеек существенно повышает стоимость топологии.

Подобно схеме динамического масштабирования напряжения питания, схема переменного порогового напряжения основана на контуре обратной связи, который может применяться для решения следующих задач:

- понижение утечки в режиме ожидания;
- компенсация колебаний порога в кристалле при нормальной работе схемы;
- регулирование пропускной способности схемы для понижения активной мощности и мощности утечки на основе требований к рабочим характеристикам.

Поскольку электрический ток в подложке гораздо меньше, чем ток в цепях питания, реализация динамического масштабирования порога характе-

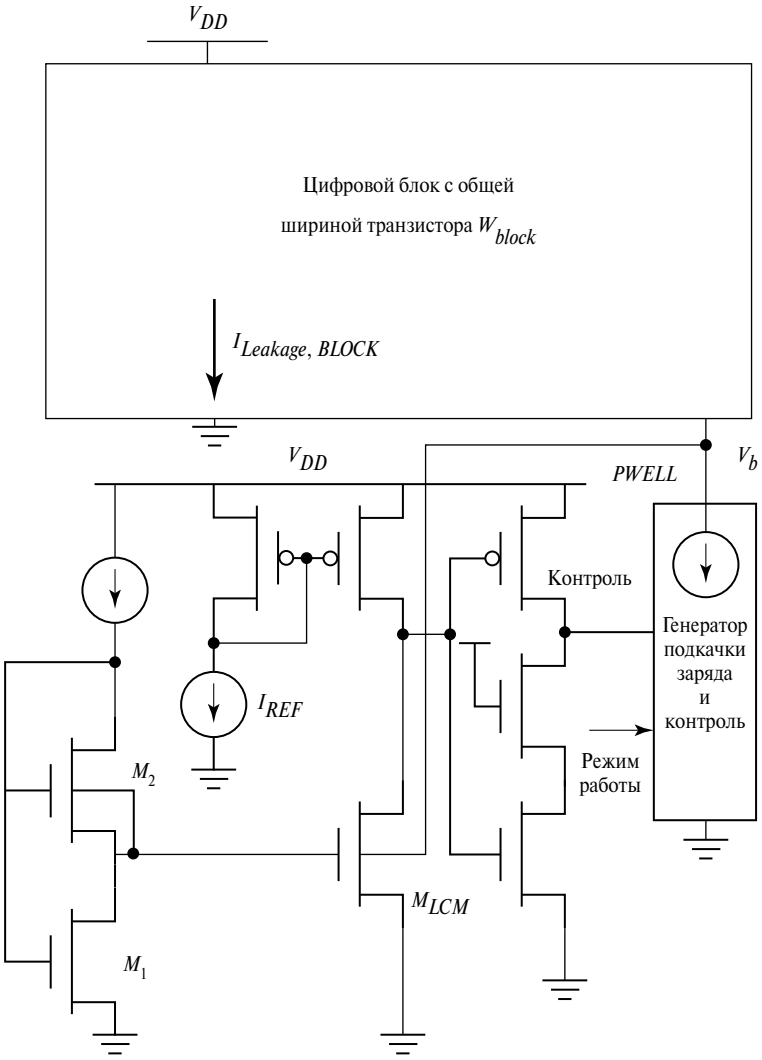


Рис. 11.61. Схема управления переменным порогом, реализованная с помощью монитора контроля тока утечки [Kuroda96]

ризуется меньшими служебными издержками, чем реализация динамического масштабирования напряжения. На рис. 11.61 показана система обратной связи, спроектированная для управления утечкой в цифровом модуле [Kuroda96], [Kuroda01]. Она состоит из монитора контроля утечки и генератора подкачки заряда смещения подложки, добавленных к искомому цифровому блоку.

Для реализации данной схемы критически важен монитор тока утечки. Транзисторы M_1 и M_2 смещаются в подпороговую область. Когда n-МОП-транзистор находится в подпороговой области, его ток выражается следующим образом:

$$I_D = I_S/W_0 \cdot W \cdot 10^{V_{GS}/S}, \tag{11.30}$$

где S — крутизна характеристики в подпороговой области. Выход генератора напряжений смещения V_b равен:

$$V_b = S \cdot \log(W_2/W_1). \tag{11.31}$$

Зная суммарную ширину всех транзисторов блока, за которыми требуется наблюдать, W_{BLOCK}, можно найти общий коэффициент масштабирования силы тока:

$$\frac{I_{LCM}}{I_{BLOCK}} = \frac{W_{LCM}}{W_{BLOCK}} \cdot 10^{\frac{V_b}{S}} = \frac{W_{LCM}}{W_{BLOCK}} \cdot \frac{W_2}{W_1}. \tag{11.32}$$

Чтобы минимизировать штраф мощности (мощность, уходящую на наблюдение), монитор тока утечки должен быть как можно меньше. Однако, если он будет слишком маленьким, наблюдение за утечкой будет медленным и смещение подложки не будет оперативно отслеживать изменения. Следовательно, оптимальное значение не может быть слишком маленьким, нормальная система наблюдения должна составлять порядка 0,001% общей ширины транзисторов кристалла.

Обычный генератор подкачки заряда вводит ток в подложку и из нее. Управляющая схема наблюдает за током утечки. Если он превышает предопределенное значение (согласно режиму работы, задается пользователем извне или определяется операционной системой), генератор подкачки заряда увеличивает отрицательное обратное смещение, откачивая ток из подложки. Когда ток утечки достигает искомого значения, генератор выключается. Из-за утечек через проход и ударной ионизации напряжение смещения подложки со временем снова повысится, а это снова активизирует контур обратной связи. Поскольку данная схема не генерирует постоянно больших токов питания, реализовать ее проще, чем динамическое масштабирование напряжения.

К сожалению, эффективность адаптивного смещения подложки уменьшается при продолжающемся масштабировании технологии. Это вызвано собственными факторами эффекта подложки и повышенной утечкой через проход, присущей межзонным переходам.

11.7.3. Уменьшение мощности в холостом (или спящем) режиме

Холостой режим представляет крайнюю точку пространства динамического управления мощностью. Поскольку активного переключения нет, рассеяние мощности происходит только за счет утечки. Чтобы уменьшить утечку в холостом режиме, можно использовать представленную в предыдущем разделе технологию динамического масштабирования порога. В более простой схеме пониженного энергопотребления применяются большие спящие транзисторы, позволяющие отключать цепи питания, когда схема находится в спящем ре-

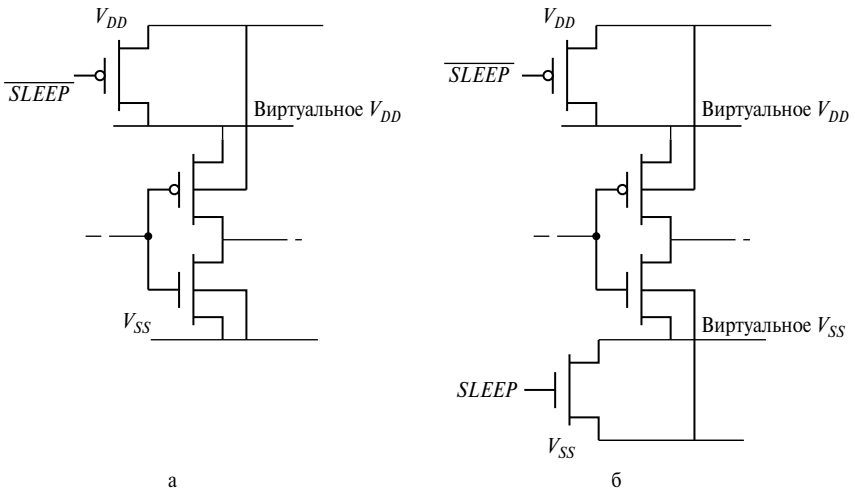


Рис. 11.62. Спящие транзисторы: *а* — используемые только в цепях питания; *б* — используемые в цепях питания и земли

жиме. Данный прямолинейный подход существенно уменьшает ток утечки, но повышает сложность структуры. Реализовать его можно, использовав выключатель электропитания в шине питания или, как показано на рис. 11.62, в шинах питания и земли (лучшая реализация).

В нормальном режиме работы сигнал *SLEEP* имеет высокий уровень и спящие транзисторы должны представлять максимально малое сопротивление. Из-за конечного сопротивления данных транзисторов возникает шум в цепях питания, свойственный изменениям тока питания, получаемого логическими схемами. Размеры и пороги спящих транзисторов необходимо выбирать, исходя из имеющихся компромиссов и требований к проекту. Чтобы минимизировать флуктуации напряжения питания, спящие транзисторы должны иметь очень низкое сопротивление в открытом состоянии, следовательно, быть очень широкими. Однако увеличение их размеров чревато большими служебными издержками. Если есть возможность использовать транзисторы с более высоким порогом, можно получить лучшее подавление утечки. Но следует помнить, что устройства с высоким порогом должны быть еще больше, чтобы дать то же сопротивление, что и устройства с низким порогом. Отметим, что принципы снижения утечки являются прямым расширением принципов, введенных в разделе 6.4.2. Добавление спящего транзистора эффективно увеличивает высоту стека транзисторов, что приводит к уменьшению утечки в десятки (для выключателей с низким порогом) или тысячи (для выключателей с высоким порогом) раз.

В отличие от простого отключения тактового сигнала, выключение питания приводит к стиранию состояний регистров внутри блока. В некоторых приложениях это допустимо, например, когда после выхода из спящего режима система будет решать абсолютно новую задачу. Тем не менее, если состояние нужно сохранить, потребуется дополнительная работа. Одним из реше-

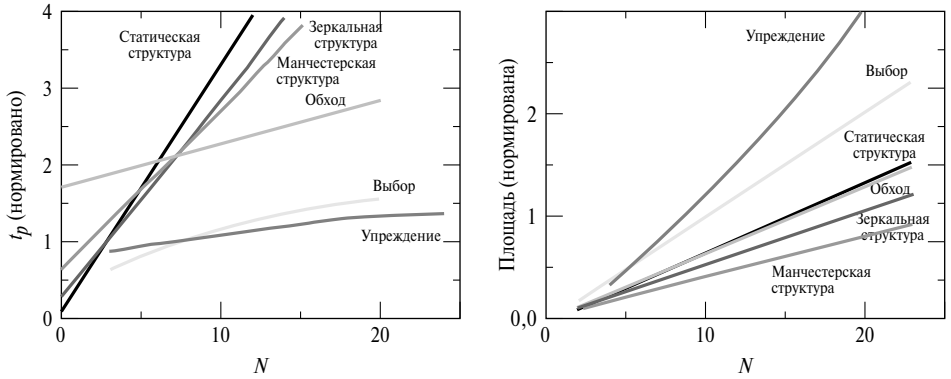


Рис. 11.63. Площадь и задержка распространения различных суммирующих структур как функция числа битов N. Графики построены на основе результатов, приведенных в [Vermassen86]

ний является соединение всех регистров с неотключаемыми цепями питания V_{DD} и V_{SS} (см. главу 7 и рис. 7.18). Альтернативой является использование операционной системы для записи состояния всех регистров в энергонезависимую память.

11.8. Перспектива: проектирование как поиск компромиссов

Из анализа схем сумматора и умножителя снова становится понятно, что проектирование цифровых схем — это постоянный поиск компромиссов между площадью, скоростью и потребляемой мощностью. Это демонстрируется на рис. 11.63, где показана зависимость нормированных площади и скорости для некоторых сумматоров, рассмотренных ранее как функции числа битов[7]. Доминирующие факторы определяются исходя из целей проекта в целом и отдельных наложенных условий.

Площадь кристалла существенно влияет на стоимость интегральной схемы. Чем больше размер кристалла, тем меньше частей поместится на одной пластине (см. главу 1). Повысить жизнеспособность продукта можно за счет уменьшения площади. В современных микропроцессорах основной акцент делается на максимально возможном быстродействии, а на продажи мобильных телефонов серьезно влияет пониженное до минимума энергопотребление. Следовательно, для правильной игры на компромиссах необходимо понимание рынка продукта. Помните: все условия, наложенные на проект (скорость, мощность и площадь), влияют на провал или успех продукта на рынке.

В таком контексте желательно привести некоторые важные концепции, которые вводились в данной главе.

[7] Учтите, что данные результаты относятся к конкретным реализациям по определенным технологиям. Экстраполировать результаты на другие технологии следует очень осторожно.

1. Наиболее важным правилом является выбор *верной структуры*, и лишь после этого можно начинать различную оптимизацию. Если вы будете пытаться получить оптимальное быстродействие сложной структуры, оптимизируя топологии и размеры транзисторов, то, скорее всего, получите не самый лучший результат. Как правило, гораздо лучше провести оптимизацию на высоких уровнях абстракции, например на уровне логики или архитектуры. Увидеть глобальную картину и понять плюсы и минусы предложенной структуры можно с помощью простых расчетов в первом приближении.
2. Определите *критический временной путь* через схему и сосредоточьте усилия на данной части схемы. Помимо анализа вручную, вы можете использовать средства автоматизированного проектирования, которые помогут определить критические пути и соответствующим образом выбрать размеры транзисторов. Помните, что некоторые некритические пути можно масштабировать с целью уменьшения потребления мощности.
3. Размер схемы определяется не только числом и размером транзисторов, но и другими факторами: проводкой, числом сквозных отверстий и контактов. Данные факторы приобретают еще большее значение при уменьшении размеров или в погоне за необычайно высокой скоростью работы.
4. Хотя с помощью загадочной оптимизации можно иногда получить и лучший результат, относитесь с подозрением к структурам с неправильной и запутанной топологией. Помните, что лучшие друзья разработчика — регулярность и модульная структура.
5. Выбирая размеры элементов, напряжения питания и пороги транзисторов, можно найти требуемый компромисс между мощностью и скоростью.

11.9. Резюме

В этой главе мы изучили реализацию операторов арифметического информационного канала с точки зрения быстродействия, площади и мощности. Особое внимание было уделено разработке *моделей задержек первого порядка*, позволяющих быстро проанализировать и сравнить различные логические структуры до ухода в глубокую оптимизацию на уровне транзисторов.

- Лучшей реализацией информационного канала является *разрядно-модульная*. Один модуль топологии последовательно используется для всех битов слова данных. Подобный регулярный подход облегчает проектирование и позволяет получать быстрые и плотные топологии.
- Задержка *сумматора со сквозным переносом* линейно пропорциональна числу битов. Существует множество схем оптимизации этого сумматора, которые направлены на уменьшение задержки на критическом пути. В главе рассмотрено несколько возможных реализаций топологии, показано, как с помощью тщательной оптимизации топологии и размеров транзисторов уменьшить емкость бита переноса.

- В других структурах сумматоров для увеличения быстродействия применяются методы оптимизации логики. Задержка сумматоров с *обводом переноса*, *выбором переноса* и *ускоренным переносом* зависит от числа битов линейно, как квадратный корень и логарифмически соответственно. Платой за повышение быстродействия является увеличение площади.
- *Умножитель* представляет собой набор соединенных каскадом сумматоров. Он имеет гораздо более сложный критический путь, поэтому оптимизация быстродействия может идти совершенно различными путями. При реализации техники записи переноса массив сумматоров превращается в регулярную структуру с четко определенным критическим путем, который позже можно легко оптимизировать. Сложность и величину задержки больших умножителей можно уменьшить с помощью кодирования Бута и накопления частичных произведений.
- Быстродействие и площадь программируемого сдвигового регистра определяются в первую очередь *проводными соединениями*. Минимизировать влияние соединительных проводников можно с помощью регулярных структур, например многорегистрового циклического сдвигового устройства и логарифмического сдвигового регистра.
- *Потребление мощности* можно существенно снизить за счет правильного выбора логической структуры, архитектуры или структуры схемы. Платой за это может быть увеличение площади реализации, однако в эпоху субмикронных устройств это не должно быть критичным.
- Разработчик имеет в своем арсенале достаточно техник, позволяющих во время проектирования и выполнения минимизировать потребление мощности. Во *время проектирования* можно реализовать приемлемые значения мощности и задержки, выбрав напряжения питания и пороги, а также оптимизировав логику и размеры транзисторов. Понизить напряжение питания, сохранив при этом пропускную способность, можно с помощью параллельной и конвейерной обработки. *Эффективную емкость* можно уменьшить, исключив расходы, связанные, например, с чрезмерным уплотнением.
- Некоторые приложения работают с переменными требованиями к пропускной способности или времени ожидания. Используя *изменяющиеся напряжения питания и пороги транзисторов*, можно понизить активную мощность или мощность утечки в таких системах. Минимизация потребления энергии в холостом состоянии необходима, например, портативным устройствам, питающимся от батарей.

11.10. Для любознательных

Об арифметических устройствах и элементах вычислительных машин написано довольно много. Самые современные разработки обычно представляются в *Proceedings of the IEEE Symposium on Computer Arithmetic*, *IEEE Transactions on Computers* и *IEEE Journal of Solid-State Circuits*. Великолепную

подборку наиболее значимых работ в этой области можно найти в некоторых репринтных изданиях IEEE Press [Swartzlander90]. Для повышения уровня рекомендуем ознакомиться с [Omondi94], [Koren98] и [Oklobdzija01].

11.11. Список использованной литературы

[Augsburger02] S. Augsburger, "Using Dual-Supply, Dual-Threshold and Transistor Sizing to Reduce Power in Digital Integrated Circuits", M.S. Project Report, University of California, Berkeley, April 2002.

[Bedrij62] O. Bedrij, "Carry Select Adder", *IRE Trans. on Electronic Computers*, 1962, vol. EC-11, p. 340–346.

[Booth51] A. Booth, "A Signed Binary Multiplication Technique", *Quart. J. Mech. Appl. Math.*, 1951, vol. 4., part 2.

[Brent82] R. Brent and H.T. Kung, "A Regular Layout for Parallel Adders", *IEEE Trans. on Computers*, March 1982, vol. C-31, no. 3, p. 260–264.

[Burd00] T.D. Burd, T.A. Pering, A.J. Stratakos, R.W. Brodersen, "A Dynamic Voltage Scaled Microprocessor System", *IEEE Journal of Solid-State Circuits*, November 2000, vol. 35, no. 11.

[Chandrakasan92] A. Chandrakasan, S. Sheng, and R. Brodersen, "Low Power CMOS Digital Design", *IEEE Journal of Solid State Circuits*, April 1992, vol. SC-27, no. 4, p. 1082–1087.

[Chandrakasan98] A. Chandrakasan and R. Brodersen, Ed., *Low-Power CMOS Design*, IEEE Press, 1998.

[Fetzer02] E.S. Fetzer, J.T. Orton, "A Fully-Bypassed 6-Issue Integer Datapath and Register File on an Itanium Microprocessor", 2002 *IEEE International Solid-State Circuits Conference*, Digest of Technical Papers, San Francisco, CA, February 2002, p. 420-421.

[Gutnik97] V. Gutnik, and A. P. Chandrakasan, "Embedded Power Supply for Low-Power DSP", *IEEE Transactions on Very Large Integration (VLSI) Systems*, December 1997, vol. 5, no. 4, p. 425–435.

[Hamada01] M. Hamada, Y. Ootaguro, and T. Kuroda, "Utilizing Surplus Timing for Power Reduction", *Proceedings of the IEEE 2001 Custom Integrated Circuits Conference*, San Diego, CA, May 2001, p. 89–92.

[Kato00] N. Kato et al., "Random Modulation: Multi-Threshold-Voltage Design Methodology in Sub-2-V Power Supply CMOS", *IEICE Transactions on Electronics*, November 2000, vol. E83-C, no. 11, p. 1747–1754.

[Katz94] R.H. Katz, *Contemporary Logic Design*, Benjamin Cummings, 1994.

[Kilburn60] T. Kilburn, D.B.G. Edwards and D. Aspinall, "A Parallel Arithmetic Unit Using a Saturated-Transistor Fast- Carry Circuit", *Proceedings of the IEE*, Pt. B, November 1960, vol. 107, p. 573–584.

[Kogge73] P.M. Kogge and H.S. Stone, "A Parallel Algorithm for the Efficient Solution of a General Class of Recurrence Equations", *IEEE Transactions on Computers*, August 1973, vol. C-22, p. 786–793.

[Koren98] I. Koren, *Computer Arithmetic Algorithms*, Brookside Court Publishers, 1998.

[Kuroda96] T. Kuroda and T. Sakurai, "Threshold Voltage Control Schemes Through Substrate Bias for Low-Power High-Speed CMOS LSI Design", *Journal on VLSI Signal Processing*, 1996, vol. 13, no. 2/3, p. 191–201.

[Kuroda01] T. Kuroda, and T. Sakurai, "Low-Voltage Technologies", in *Design of High-Performance Microprocessor Circuits*, Chandrakasan, Bowhil, Fox (eds.), IEEE Press, 2001.

[Lehman62] M. Lehman and N. Burla, "Skip Techniques for High-Speed Carry Propagation in Binary Arithmetic Units", *IRE Transactions on Electronic Computers*, December 1962, vol. EC-10, p. 691–698.

[Ma94] S. Ma, P. Franzon, "Energy Control and Accurate Delay Estimation in the Design of CMOS Buffers", *IEEE Journal of Solid-State Circuits*, September 1994, vol. 29, no. 9, p. 1150–1153.

[MacSorley61] O. MacSorley, "High Speed Arithmetic in Binary Computers", *IRE Proceedings*, 1961, vol. 49, p. 67–91.

[Mathew01] S. K. Mathew et al., "Sub-500-ps 64-b ALUs in 0–18-μm SOI/Bulk CMOS: Design and Scaling Trends", *IEEE Journal of Solid-State Circuits*, November 2001, vol. 36, no. 11, p. 1636–1646.

[Ohkubo95] N. Ohkubo et al., "A 4–4 ns CMOS 54*54-b Multiplier Using Pass-Transistor Multiplexer", *IEEE Journal of Solid-State Circuits*, March 1995, vol. 30, no. 3, p. 251–257.

[Oklobdzija01] V. G. Oklobdzija, "High-Speed VLSI Arithmetic Units: Adders and Multipliers", in *Design of High-Performance Microprocessor Circuits*, Chandrakasan, Bowhil, Fox (eds.), IEEE Press, 2001.

[Omondi94] A. Omondi, *Computer Arithmetic Systems*, Prentice Hall, 1994.

[Pering00] T. A. Pering, *Energy-Efficient Operating System Techniques*, Ph.D. Dissertation, University of California, Berkeley, 2000.

[Rabaey96] J. Rabaey, and M. Pedram, Ed., "Low-Power Design Methodologies", Kluwer Academic Publishers, 1996.

[Stojanovic02] V. Stojanovic, D. Markovic, B. Nikolic, M. A. Horowitz, and R. W. Brodersen, "Energy-Delay Tradeoffs in Combinational Logic using Gate Sizing and Supply Voltage Optimization", *European Solid-State Circuits Conference*, ESSCIRC'2002, Florence, Italy, 2002, September 24–26.

[Suzuoki99] M. Suzuoki et al., "A Microprocessor with a 128-bit CPU, Ten Floating-Point MAC's, Four Floating-Point Dividers, and an MPEG-2 Decoder", *IEEE Journal of Solid-State Circuits*, November 1999, vol. 34, no. 11, p. 1608–1618.

[Swartzlander90] E. Swartzlander, ed., *Computer Arithmetic — Part I and II*, IEEE Computer Society Press, 1990.

[Usami95] K. Usami, and M. Horowitz, "Clustered Voltage Scaling Technique for Low-Power Design", *Proceedings, 1995 International Symposium on Low Power Design*, Dana Point, CA, April 1995, p. 3–8.

[Vermassen86] H. Vermassen, "Mathematical Models for the Complexity of VLSI", (на датском языке), Master's Thesis, Katholieke Universiteit, Leuven, Belgium, 1986.

[Wallace64] C. Wallace, "A Suggestion for a Fast Multiplier", *IEEE Trans. on Electronic Computers*, 1964, EC-13, p. 14–17.

[Weinberger56] A. Weinberger and J. L. Smith, "A One-Microsecond Adder Using One-Megacycle Circuitry", *IRE Transactions on Electronic Computers*, 1956, vol. 5, p. 65–73.

[Wei99] L. Wei, Z. Chen, K. Roy, M.C. Johnson, Y. Ye, and V. K. De, "Design and Optimization of Dual-Threshold Circuits for Low-Power Applications", *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, March 1999, vol. 7, no. 1, p. 16–24.

[Weinberger81] A. Weinberger, "A 4:2 Carry-Save Adder Module", *IBM Technical Disclosure Bulletin*, January 1981, vol. 23.

[Weste85&93] N. Weste and K. Eshragian, *Principles of CMOS VLSI Design: A Systems Perspective*, 2d ed., Addison-Wesley, 1985 and 1993.

11.11.1. Упражнения

Последнюю версию сборника задач и упражнений можно найти по адресу `http://bwrc.eecs.berkley.edu/IcBook`.

ГЛАВА

12

Проектирование запоминающих устройств и матричных структур

Классификация и архитектура запоминающих устройств

Ячейки памяти для постоянных, оперативных и энергонезависимых ЗУ

Периферийные схемы — усилители считывания, декодеры, задающие устройства и генераторы тактовых сигналов

Потребление мощности и надежность

12.1. Введение
12.1.1. Классификация запоминающих устройств
12.1.2. Архитектуры и основные составляющие запоминающих устройств
12.2. Сердечник запоминающего устройства
12.2.1. Постоянные запоминающие устройства
12.2.2. Энергонезависимые оперативные запоминающие устройства
12.2.3. Оперативные запоминающие устройства
12.2.4. Ассоциативные запоминающие устройства
12.3. Периферийные схемы запоминающих устройств*
12.3.1. Декодеры адреса
12.3.2. Усилители считывания
12.3.3. Опорное напряжение
12.3.4. Задающие устройства/буферы
12.3.5. Синхронизация и контроль
12.4. Надежность и выход годных запоминающих устройств*

12.4.1. Отношение сигнал/шум
12.4.2. Выход годных запоминающих устройств
12.5. Рассеяние мощности в запоминающих устройствах*
12.5.1. Источники рассеяния мощности в запоминающих устройствах
12.5.2. Секционирование памяти
12.5.3. Решение проблемы рассеяния активной мощности
12.5.4. Рассеяние хранимой информации
12.5.5. Резюме
12.6. Разбор примеров запоминающих устройств
12.6.1. Программируемая логическая матрица (ПЛМ)
12.6.2. Статическое ОЗУ емкостью 4 Мбит
12.6.3. Флэш-память емкостью 1 Гбит на основе элементов НЕ-И
12.7. Перспективы: тенденции и эволюция в сфере полупроводниковых ЗУ
12.8. Резюме
12.9. Для любознательных

12.1. Введение

Существенная часть кремниевой области многих современных цифровых устройств предназначена для хранения данных и команд программ. Более половины транзисторов в современных высокоскоростных микропроцессорах являются частью кэша, причем предполагается, что в будущем их доля будет расти. На уровне системы ситуация получается еще более драматической. Высокопроизводительные рабочие станции и компьютеры содержат несколько гигабайтов полупроводниковой памяти, и это значение также постоянно растет. С появлением полупроводниковых аудио- (MP3) и видеоплейеров (MPEG4) спрос на энергонезависимые запоминающие устройства взлетел до небес. Из сказанного очевидно, что плотные схемы хранения данных по-прежнему представляют существенный интерес для разработчика цифровых схем или систем. Общий рынок полупроводниковых запоминающих устройств в 2003 году составил более 45 млрд. долл., что вдвое больше, чем было в 1998 году.

В главе 7 мы описали средства хранения булевых значений на основе положительной обратной связи или емкостного накопителя энергии. Хотя полупроводниковые запоминающие устройства созданы по аналогичному принципу, простое использование ячейки регистра для накопления информации налагает чрезмерные требования к площади реализации. Поэтому ячейки памяти объединяются в большие матрицы, что минимизирует служебные издержки, связанные с периферийными цепями, и увеличивает плотность размещения информации в памяти. Из-за большого размера и сложности данных матричных структур появляется множество проблем разработки, некоторые из которых рассмотрены в этой главе.

Для начала введем основные архитектуры запоминающих устройств и их составляющие. Затем проанализируем различные ячейки памяти и их свойства. Структура и топология ячейки определяются в основном доступной технологией, т.е. не совсем тем, на что может повлиять разработчик цифровой системы. С другой стороны, периферийные цепи имеют огромное влияние на устойчивость, быстродействие и потребляемую мощность запоминающего устройства. Следовательно, при их проектировании необходим тщательный анализ всех доступных альтернатив. Кроме того, разработчику полупроводниковых запоминающих устройств необходимо уделять особое внимание вопросам надежности и рассеяния мощности, поэтому они довольно подробно рассмотрены в данной главе.

Изучая главу, учтите, что приводимые соображения применимы к большому числу технологий разработки схем, введенных в предыдущих главах. По сути, проектирование запоминающих устройств можно рассматривать как практический пример разработки высокоскоростной плотной схемы с низким потреблением мощности.

12.1.1. Классификация запоминающих устройств

Электрические запоминающие устройства имеют различные форматы и стили. Тип ячейки памяти, наиболее подходящий для данного приложения, зависит от требуемой емкости памяти, времени, необходимого для обращения к записанным данным, схемы доступа, приложения и технических требований к системе.

Емкость

В зависимости от уровня абстракции для представления емкости запоминающего устройства могут использоваться различные средства. Разработчики *схем* предпочитают определять емкость памяти в *битах*, что эквивалентно числу отдельных ячеек (триггеров или регистров), требуемых для хранения данных. Проектировщики *кристаллов* выражают емкость памяти в *байтах* (группах из 8 или 9 бит), килобайтах (Кбайт), мегабайтах (Мбайт), гигабайтах (Гбайт) и терабайтах (Тбайт). Разработчики *систем* любят указывать требования к запоминающим устройствам в *словах*, представляющих основную единицу вычислений. Например, группа из 32 бит представляет слово в компьютере, оперирующем 32-битовыми данными.

Временны*е* параметры

Временны*е* свойства ячейки памяти схематически показаны на рис. 12.1. Время, требуемое для извлечения (чтения) информации из памяти, называется *временем доступа на чтение*; оно равно задержке между запросом на чтение и моментом появления данных на выходе. Это время отличается от *времени доступа для записи* — интервала между моментом подачи запроса на запись и окончательным внесением данных в память. Наконец, еще одним важным параметром является *продолжительность такта* (чтения или за-

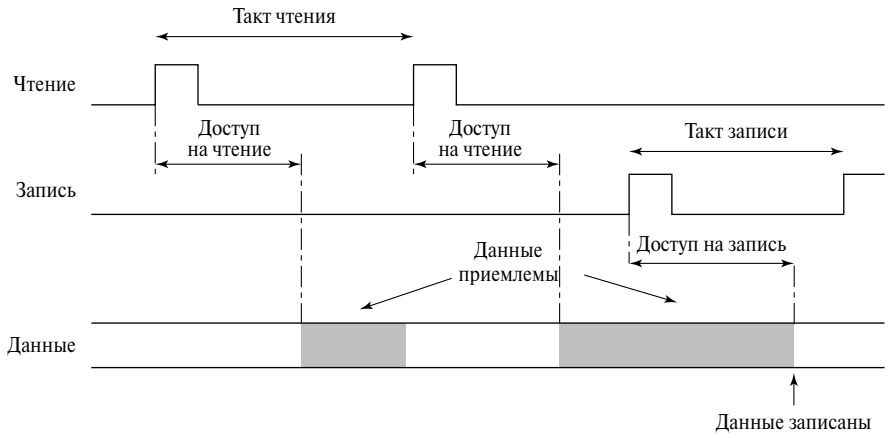


Рис. 12.1. Определения временны*х* параметров

писи) — минимальное время, проходящее между двумя последовательными операциями чтения или записи. Обычно это время больше времени доступа (причины этого объясняются ниже). Такты чтения и записи не обязательно имеют одинаковую продолжительность, однако при разработке системы они для простоты считаются равными.

Функция

Чаще всего полупроводниковые запоминающие устройства классифицируются согласно функциональным возможностям ячеек памяти, схемам доступа и природе механизма хранения информации. Кроме того, следует различать *постоянные* (Read-Only Memory — ROM, ПЗУ) и *оперативные* (Read-Write Memory — RWM, ОЗУ) *запоминающие устройства*. ОЗУ-структуры предлагают возможности чтения и записи с приблизительно одинаковым временем доступа и являются наиболее гибкими запоминающими устройствами. Данные в них хранятся либо в триггерах, либо в виде заряда конденсатора. Согласно классификации, введенной при обсуждении последовательных схем, данные ячейки памяти называются соответственно *статическими* и *динамическими*. Статические запоминающие устройства сохраняют свои данные до тех пор, пока на ячейку подается питание, а динамические требуют периодического обновления, компенсирующего потерю заряда из-за утечки. Поскольку в ОЗУ для хранения информации применяются активные цепи, они относятся к классу *энергозависимых* запоминающих устройств, в которых данные теряются при выключении напряжения питания.

С другой стороны, ПЗУ кодируют информацию в топологии схемы, например с помощью добавления или удаления транзисторов. Поскольку такая топология является фиксированной, данные нельзя модифицировать; их можно только считывать. Кроме того, ПЗУ-структуры принадлежат к классу *энергонезависимых* запоминающих устройств (отключение источника питания не приводит к потере данных).

Новым классом запоминающих устройств являются энергонезависимые ячейки памяти, предлагающие возможности чтения и записи. Как правило, запись в них занимает гораздо большее времени, чем чтение. Мы называем их *энергонезависимыми оперативными запоминающими устройствами* (Nonvolatile Read-Write Memory — NVRWM). К данному классу относятся *стираемые программируемые постоянные запоминающие устройства*, СППЗУ (Erasable Programmable Read-Only Memory — EPROM), *электрически стираемые программируемые постоянные запоминающие устройства*, ЭСППЗУ (Electrically Erasable Programmable Read-Only Memory — E^2PROM) и *флэш-память*. Благодаря появлению в последнее десятилетие новых дешевых и плотных энергонезависимых технологий этот подход к хранению данных стал наиболее быстро развивающимся направлением в сфере полупроводниковых запоминающих устройств.

Схема доступа

Второй способ классификации элементов памяти основан на порядке, в котором можно обращаться к данным. Большинство запоминающих устройств относятся к классу устройств с *произвольным доступом*, т.е. считывать информацию из ячеек памяти или записывать ее в них можно в произвольном порядке.

Некоторые типы запоминающих устройств обычно ограничивают порядок доступа, что позволяет сократить время доступа, уменьшить площадь или получить память с определенными функциональными возможностями. В качестве примеров таких устройств можно привести буферы *FIFO* (First-In First-Out — "первым прибыл, первым обслужен"), *LIFO* (Last-In First-Out — "последним прибыл, первым обслужен", чаще всего применяется в стеках) и *сдвиговые регистры*. Кроме того, к этому классу принадлежит *видеопамять*. В сфере обработки видео данные извлекаются и выдаются последовательно, поэтому произвольный доступ не требуется. Еще одним классом запоминающих устройств без произвольного доступа является *ассоциативная память* (Contents-Addressable Memory — CAM). Для нахождения данных в ней применяется не адрес, а само слово данных, форматируемое в виде запроса. Когда входные данные согласуются со словом, записанным в матрице памяти, включается сигнал *MATCH*. Если в памяти нет данных, соответствующих входному слову, сигнал *MATCH* остается на исходном (нулевом) уровне. Ассоциативная память является очень важным компонентом кэша многих микропроцессоров.

Краткий обзор рассмотренных классов запоминающих устройств представлен на рис. 12.2. Реализации всех названных структур будут рассмотрены в последующих разделах. Мы покажем, что природа запоминающего устройства влияет не только на выбор базовой ячейки памяти, но и на состав периферийных устройств.

RWM		NVRWM	ROM
С произвольным доступом	**Без произвольного доступа**	EPROM E^2PROM FLASH	Программируемые постоянные запоминающие устройства (ППЗУ)
SRAM DRAM	FIFO LIFO Регистр сдвига CAM		

Рис. 12.2. Классификация полупроводниковых запоминающих устройств

Архитектура ввода-вывода

Последний способ классификации полупроводниковых ЗУ основан на *числе портов ввода и вывода данных*. Подавляющее большинство ячеек памяти предоставляют единственный порт, который совместно используется для ввода и вывода данных. Однако запоминающие устройства с более высокими требованиями к полосе пропускания часто имеют несколько портов ввода и вывода и называются *многопортовой памятью*. В качестве примера многопортовых запоминающих устройств можно привести блоки регистров, используемые в микропроцессорах компьютеров с сокращенным набором команд (Reduced Instruction Set Computer — RISC). Отметим, что увеличение числа портов обычно усложняет структуру ячейки памяти.

Приложение

В конце прошлого столетия большинство запоминающих устройств большой емкости выпускались в виде самостоятельных интегральных схем. С развитием систем на кристалле и интеграцией в одном кристалле большого числа функций все больше запоминающих устройств в настоящее время интегрируется в тот же кристалл, что и логические схемы. Память такого типа называется *встроенной* (или *интегрированной*). Совмещение разных функций существенно влияет на проектирование запоминающих устройств — не только на общую архитектуру памяти, но и на используемую технологию.

Если в системе требуется существенный объем памяти (более нескольких гигабитов), то полупроводниковые запоминающие устройства становятся слишком дорогими. В таких случаях выгодно использовать более экономичные технологии, например запись на магнитных и оптических дисках. Хотя соответствующие устройства дают достаточно большую емкость за низкую цену, они обычно бывают слишком медленными или предлагают ограниченные возможности доступа. Например, магнитная лента обычно допускает только последовательный доступ. По этим причинам такие запоминающие устройства не связываются прямо с вычислительным процессором, а сопрягаются с ним через несколько более быстрых полупроводниковых запоминающих устройств. Такие устройства называются памятью *второго* или *третьего уровня* и в данной книге не рассматриваются.

12.1.2. Архитектуры и основные составляющие запоминающих устройств

При реализации памяти, оперирующей N словами по M бит[1], наиболее интуитивным подходом является линейное выстраивание последовательных слов памяти в стек, как показано на рис. 12.3, *а*. Если предположить, что данный модуль является однопортовым запоминающим устройством, в любой момент времени с помощью бита *выбора* (от S_0 до S_{N-1}) можно выбрать слово для чтения или записи. Другими словами, в любой момент времени высокий уровень может иметь только один сигнал S_i. Предположим для простоты (временно), что каждая ячейка памяти представляет собой D-триггер, а для активизации ячейки применяется сигнал выбора. Хотя такой подход сравнительно прост и хорошо подходит для анализа небольших запоминающих устройств, его использование для больших устройств не рекомендуется из-за множества сопутствующих проблем.

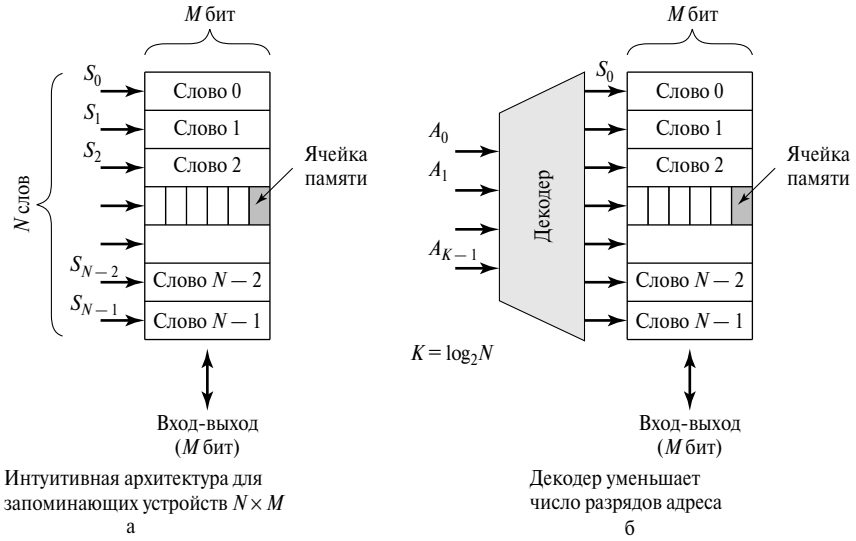


Рис. 12.3. Архитектуры запоминающего устройства на N слов по M бит

Предположим, что нам требуется реализовать запоминающее устройство, содержащее 1 миллион ($N = 10^6$) 8-битовых ($M = 8$) слов. Следует помнить, что с позиции реальной памяти 1 миллион — это упрощение, поскольку емкость запоминающего устройства всегда выражается степенью двойки. В данном случае реальное число слов было бы равно $2^{20} = 1024 \times 1024 = 1\ 048\ 576$.

При реализации данной структуры с использованием стратегии, приведенной на рис. 12.3, *а*, мы быстро обнаруживаем, что требуется 1 милли-

[1] Длина слова может составлять от 1 до 128 бит. В коммерческих запоминающих устройствах длина слова обычно равна 1 биту, 4 или 8 битам.

он сигналов выбора (по одному на каждое слово). Поскольку эти сигналы обычно поставляются извне кристалла или из другой части кристалла, мы получаем непреодолимые проблемы с трассировкой и (или) корпусированием. Поэтому для уменьшения числа сигналов выбора используется *декодер* (см. рис. 12.3, *б*). Слово памяти выбирается посредством двоичного *адресного слова* (от A_0 до A_{K-1}). Декодер переводит этот адрес в $N = 2^K$ шин выборки, только одна из которых может быть активизированной в любой момент времени. В нашем примере такой подход уменьшает число внешних адресных шин с 1 миллиона до 20 ($\log_2 2^{20}$), что практически снимает проблемы трассировки и корпусирования. Декодер обычно проектируется так, чтобы его размер согласовывался с размером ячейки памяти, поэтому соединения между декодером и памятью (в частности, S сигналов на рис. 12.3, *б*) не вводят никаких дополнительных служебных издержек с точки зрения площади. В ценности описанного подхода можно убедиться, изучив рис. 12.3, *б* как архитектурный план модуля памяти. Отметим, что S проводников можно сделать очень короткими, т.е. длинные трассировочные каналы не потребуются.

Итак, мы справились с проблемой выбора, но ничего не сделали с характеристическим отношением запоминающего устройства. Изучая размеры матрицы памяти, видим, что ее высота приблизительно в 128 тысяч раз больше ширины ($2^{20}/2^3$), при этом предполагается, что элементарная ячейка памяти имеет квадратный размер (на практике это почти всегда так). Очевидно, в реальной жизни такую структуру реализовать невозможно. Кроме того, помимо специфической формы, предлагаемое запоминающее устройство получается очень медленным. Вертикальные проводники, соединяющие ячейки памяти с входами/выходами, становятся слишком длинными, а, как говорилось ранее, задержка межсоединения в лучшем случае линейно растет с его длиной.

Чтобы решить описанную проблему, ячейки памяти выстраиваются так, чтобы вертикальные и горизонтальные размеры имели приблизительно одинаковый порядок величины; следовательно, характеристическое отношение приблизительно равно единице. В одном ряде хранятся несколько слов, которые извлекаются одновременно. Чтобы получить на входе/выходе нужное слово, требуется дополнительная схема, называемая *декодером столбцов*. Данная концепция иллюстрируется на рис. 12.4. Адресное слово разбивается на *адрес в столбце* (от A_0 до A_{K-1}) и *адрес в строке* (от A_K до A_{L-1}). Адрес в строке открывает одну строку памяти для чтения/записи, а адрес в столбце позволяет выбрать нужное слово из данной строки.

Пример 12.1. Организация запоминающего устройства

Альтернативным вариантом является организация сердечника запоминающего устройства в виде матрицы 4000 × 2000 ячеек (точнее, 4096 × 2048). В каждой из 4000 строк хранится 256 8-битовых слов. В таком случае адрес в строке состоит из 12 бит, а адрес в столбце — из 8 бит. Можно проверить, что все адресное пространство представляется 20 битами.

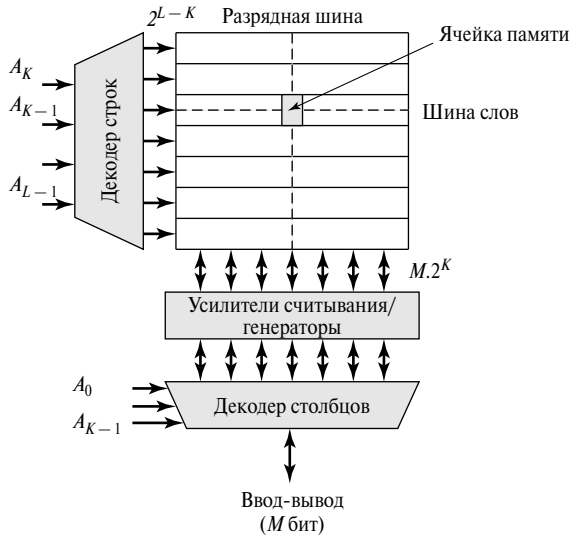


Рис. 12.4. Матричная организация запоминающего устройства

На рис. 12.4 указаны термины, используемые при проектировании данных запоминающих устройств. Горизонтальная шина выбора, активизирующая одну строку ячеек, называется *числовой шиной*, а проводник, соединяющий ячейки одного столбца со схемой ввода-вывода, — *разрядной шиной*.

Основную часть площади больших модулей памяти занимает сердечник запоминающего устройства. Следовательно, составляющие его элементарные ячейки должны иметь как можно меньший размер. Например, для реализации оперативной памяти (с возможностью чтения и записи) мы можем использовать один из регистров, введенных в главе 7. Подобная ячейка вполне может использовать более десяти транзисторов на один бит, поэтому ее использование в больших запоминающих устройствах сразу приведет к чрезмерным требованиям к площади. Из-за этого полупроводниковые ячейки памяти уменьшают площадь ячейки за счет некоторых полезных свойств цифровых схем — запаса помехоустойчивости, размаха логических уровней, изоляции входа от выхода, разветвления по выходу или быстродействия. Хотя ухудшение некоторых из указанных параметров допускается в определенной области сердечника запоминающего устройства, где уровень шума можно контролировать довольно строго, при сопряжении с внешними схемами оно недопустимо. Следовательно, на выходе из сердечника необходимо восстановить допустимые значения параметров, использовав для этого подходящие периферийные цепи.

Например, размах напряжения в разрядных шинах обычно уменьшают до уровня, существенно меньшего напряжения питания. В результате уменьшаются задержка распространения и потребляемая мощность. В матрицах памяти возможен строгий контроль перекрестных помех и других возмущений, что гарантирует достаточный запас помехоустойчивости даже при небольших

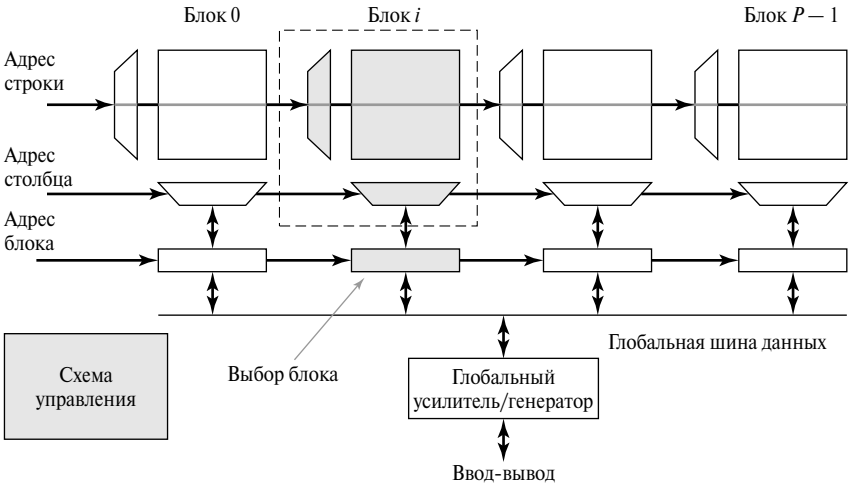


Рис. 12.5. Иерархическая архитектура памяти. Добавлены цепи, позволяющие выбрать требуемый блок

размахах сигналов. С другой стороны, сопряжение с внешним миром требует усиления внутреннего размаха сигналов до полного значения (определяемого напряжением питания). Для этого применяются *усилители считывания*, показанные на рис. 12.4. Структура этих периферийных схем рассматривается в разделе 12.3. Понижение необходимого уровня названных выше параметров цифровых схем позволяет уменьшить число транзисторов в одной ячейке памяти до 1–6 штук!

Архитектура, приведенная на рис. 12.4, удачна для запоминающих устройств с емкостью 64–256 Кбит. В бо́льших блоках такая архитектура характеризуется серьезным снижением быстродействия, поскольку длина, емкость и сопротивление разрядных и числовых шин становятся чересчур большими. Поэтому в более емких запоминающих устройствах в адресное пространство добавляется еще одно измерение, как показано на рис. 12.5.

Запоминающее устройство разбивается на P меньших блоков, имеющих идентичную структуру (см. рис. 12.4). Нужное слово выбирается согласно адресам строки и столбца, которые рассылаются всем блокам. Конкретный блок для чтения или записи выбирается с помощью дополнительного *адреса блока*. В целом данный подход имеет два преимущества.

1. Длина локальных числовой и разрядной шины (т.е. длина шин внутри блоков) является приемлемой, дающей малое время доступа.
2. *Адрес блока* можно использовать для активизации только нужного блока. Неактивизированные блоки входят в энергосберегающий режим с отключенными усилителями считывания и декодерами строк и столбцов. В результате получаем существенную экономию мощности, а это весьма хорошо в условиях большого рассеяния мощности в больших запоминающих устройствах.

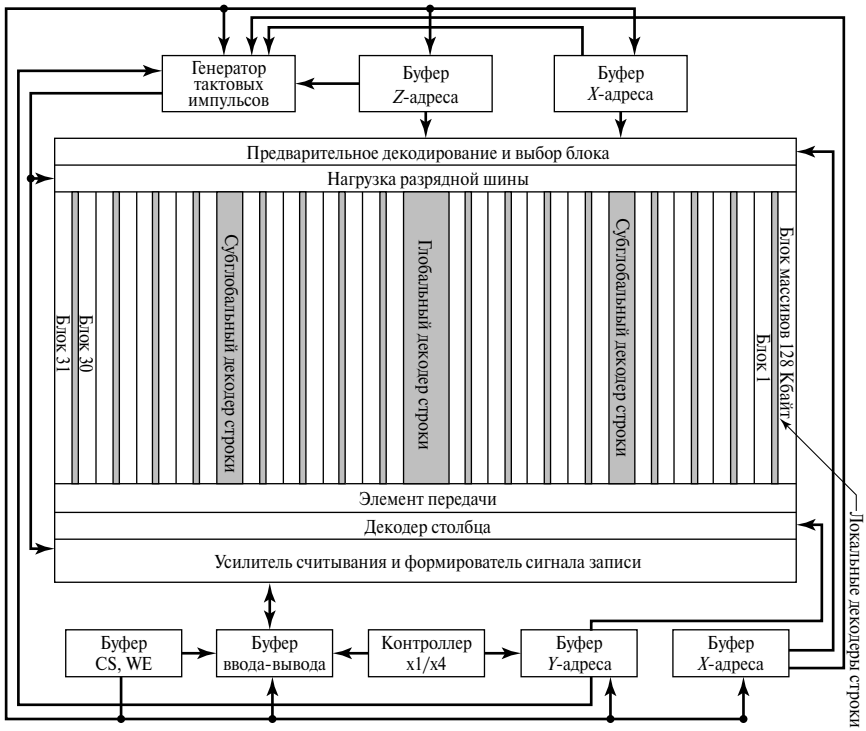


Рис. 12.6. Блок-схема памяти 4 Мбит (согласно [Hirose90])

Пример 12.2. Иерархическая архитектура памяти

Рассмотрим рис. 12.6 [Hirose90], где изображено статическое ОЗУ емкостью 4 Мбит, представленное как набор из 32 блоков по 128 Кбит. Каждый блок структурирован в виде матрицы из 1024 строк и 128 столбцов. Адрес строки (X), адрес столбца (Y) и адрес блока (Z) имеют размер 10, 7 и 5 бит соответственно.

Возможно несколько вариантов данной архитектуры. Варьировать можно расположение усилителей считывания, разделение на числовые и разрядные шины, а также стили используемых декодеров. Суть остается той же — большое запоминающее устройство удобно разбить на меньшие блоки, что позволит уменьшить задержку, присущую очень длинным шинам. Получаемый в результате выигрыш в быстродействии и потребляемой мощности является вполне достаточным, чтобы компенсировать служебные издержки, вводимые дополнительным разбиением.

Природа *последовательных* и *ассоциативных запоминающих устройств* естественным образом порождает разнообразие архитектур и состава блоков памяти. На рис. 12.7 показан пример ассоциативной памяти на 512 слов, которая поддерживает три режима работы: чтение, запись и нахождение со-

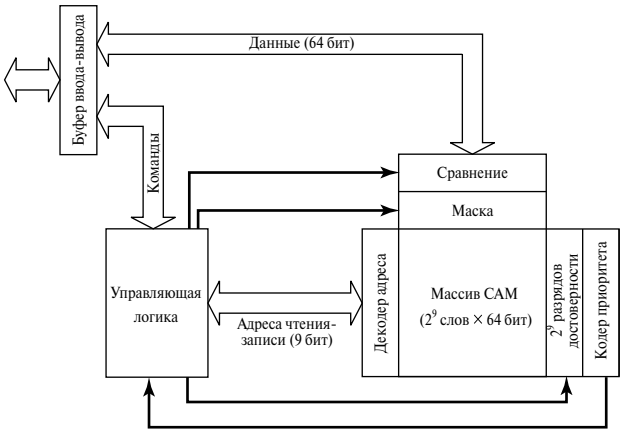


Рис. 12.7. Архитектура ассоциативного запоминающего устройства на 512 слов

ответствия. В режимах чтения и записи происходит доступ к данным и их обработка (как в обычном запоминающем устройстве). Режим установления соответствия существует только в ассоциативных запоминающих устройствах. *Блок сравнения* заполняется структурами, используемыми в операции нахождения соответствия, а слово-*маска* указывает, какие биты следует учитывать. Например, чтобы найти в ассоциативном ЗУ все слова, имеющие в старших разрядах структуру 0×123, в блок сравнения загружается шаблон 0×12300000, а в блок маски — шаблон $0 \times$ FFF00000. После этого 12 старших битов всех 512 строк ассоциативного ЗУ сравниваются с шаблоном. Все строки, согласующиеся с шаблоном, передаются в блок достоверности. Поскольку нас не интересуют строки, содержащие данные, не согласующиеся с шаблоном, на *шифратор приоритетов* подаются только строки, согласующиеся с шаблоном. Если таких строк несколько, то для их упорядочения используется строка адреса массива ассоциативного ЗУ. В этом случае шифратор приоритетов рассматривает все 512 строк матрицы, выбирает из них строку с наибольшим адресом и кодирует ее в двоичном формате[2]. Поскольку в матрице ассоциативного ЗУ имеется 512 строк, для указания старшей строки требуется 9 бит. Кроме того, применяется дополнительный бит "соответствие найдено", поскольку существует вероятность того, что ни одна из строк не соответствует шаблону.

Наконец, рассмотрим компонент запоминающего устройства, о котором часто забывают, — *интерфейс ввода-вывода и схема управления*. Природа интерфейса ввода-вывода существенно влияет на глобальное управлению памятью и синхронизацию. Проиллюстрируем это, сравнив поведение типичных динамического и статического оперативных запоминающих устройств с соответствующей временно*й* структурой.

[2] Самый большой адрес обычно соответствует последней записи кэша, которая чаще всего и является искомой.

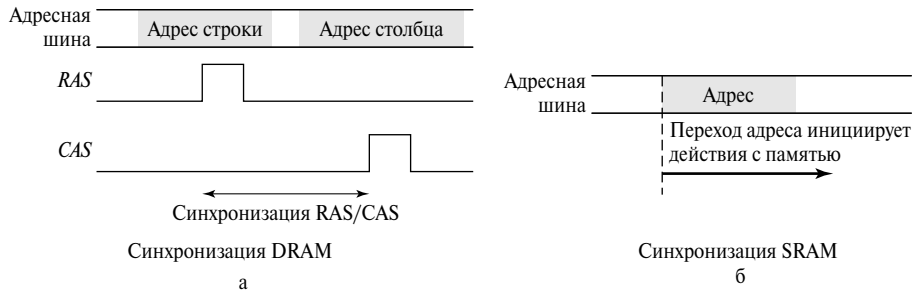


Рис. 12.8. Интерфейс ввода-вывода динамических (DRAM) и статических (SRAM) оперативных запоминающих устройств, а также их влияние на управление памятью

С первых дней своей работы разработчики динамических ОЗУ предпочитали использовать уплотненную схему адресации. В данной модели верхняя и нижняя половины адресных слов последовательно представляются в одной адресной шине. Такой подход уменьшает число выводов корпуса ИС, поэтому он применялся во всех последующих поколениях запоминающих устройств. Динамические ОЗУ обычно имеют большую емкость. Уменьшение числа штырьковых выводов снижает размер и стоимость микросхемы за счет понижения быстродействия. Добавление нового адресного слова обеспечивается за счет увеличения уровня нескольких стробирующих сигналов (рис. 12.8, *а*.) Увеличение уровня *стробирующего сигнала адреса строки* (Row-Access Strobe — RAS) гарантирует наличие в адресной шине *старших битов адреса* и возможность инициации процесса декодирования слова. После этого задействуются *младшие биты адреса* и активизируется стробирующий сигнал адреса столбца (Column-Access Strobe — CAS). Для обеспечения правильной работы памяти требуется аккуратная настройка интервала RAS–CAS. Фактически сигналы RAS и CAS действуют как тактовые входы модуля памяти и используются для синхронизации таких событий, как декодирование, доступ к сердечнику ЗУ и считывание.

С другой стороны, разработчики статических ОЗУ предпочитают подход с использованием автосинхронизации (см. рис. 12.8, *б*). Все слово адреса предоставляется сразу, кроме того, реализована схема, автоматически детектирующая все переходы в шине. Никаких внешних синхронизирующих сигналов не требуется. Все внутренние временны*е* события (активизация декодеров и усилителей считывания) синхронизируются по внутреннему сгенерированному сигналу переходов. Преимущество такого подхода — длительность такта статического ОЗУ близка к времени доступа (разумеется, для динамического ОЗУ такое утверждение неверно). Соответственно общее быстродействие вычислительных систем, использующих для хранения информации динамические ОЗУ, должно расти приблизительно такими же темпами. Поэтому было разработано несколько новых подходов, повышающих быстродействия динамического ОЗУ для операций чтения. В качестве примеров получающихся устройств можно привести *синхронное динамическое ОЗУ* (Synchronous DRAM — SDRAM) и *динамическое ОЗУ Rambus* (Rambus DRAM — RDRAM).

Основные изменения этих новых архитектур, рассмотренных ниже в этой главе, связаны не с сердечником ЗУ, а со способом связи с внешним миром.

Такое проектирование схем управления и синхронизации, чтобы запоминающее устройство функционировало в широком диапазоне производственных допусков и рабочих температур, является достаточно сложной задачей, требующей тщательного моделирования и оптимизации разработки. Это неотъемлемая часть процесса разработки запоминающего устройства (о которой, правда, часто забывают), существенно влияющая на его надежность и быстродействие.

12.2. Сердечник запоминающего устройства

В данном разделе рассматривается проектирование сердечника запоминающего устройства и составляющих его ячеек для различных типов полупроводниковых запоминающих устройств. Хотя основным требованием при разработке больших запоминающих устройств является минимальный размер базовой ячейки памяти, не следует забывать и о других важных метриках качества проекта — быстродействии и надежности. Ниже мы рассмотрим сердечники *энергонезависимой* памяти, а также запоминающих устройств, допускающих *только чтение* (ПЗУ) и *чтение/запись* (ОЗУ). В конце раздела кратко рассматривается ассоциативная ячейка памяти.

12.2.1. Постоянные запоминающие устройства

Хотя идея запоминающего устройства, информацию с которого можно только считывать, может на первый взгляд показаться странной, при внимательном изучении у таких устройств находится множество потенциальных сфер применения. Программы для процессоров, обслуживающих устройства с конечным набором действий (стиральные машины, калькуляторы и игровые автоматы), после разработки и отладки требуют только чтения.

Ячейки ПЗУ — обзор

Фиксация содержимого ячейки ПЗУ существенно упрощает ее проектирование. Ячейку следует разработать так, чтобы после активизации числовой шины в разрядной шине записывалось значение 0 или 1. Несколько способов реализации данного принципа показано на рис. 12.9.

Рассмотрим вначале элементарную ячейку, показанную на рис. 12.9, *а*, — ячейку ПЗУ на основе диода. Предположим, что разрядная шина BL заземлена через сопротивление, т.е. при отсутствии любого возбуждения или входов сигнал BL переводится на низкий уровень через резистор, соединенный с землей. Таким образом, имеем поведение, характерное для ячейки 0. Поскольку физического соединения между числовой шиной WL и разрядной шиной BL нет, значение BL соответствует низкому логическому уровню независимо от значения WL. С другой стороны, когда в числовую шину ячейки 1 подается высокое напряжение V_{WL}, диод отпирается и сигнал числовой шины поднима-

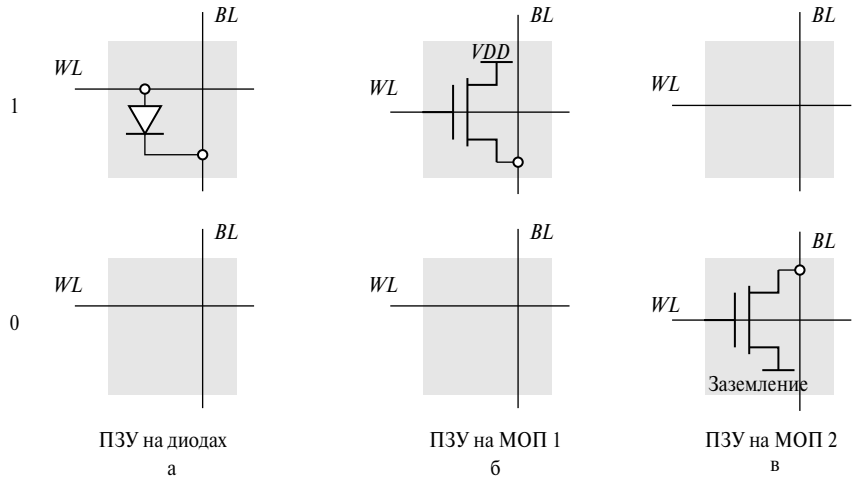


Рис. 12.9. Различные подходы к реализации ячеек 1 и 0 ПЗУ

ется до высокого уровня $(V_{WL} - V_{D(on)})$, т.е. в разрядной шине регистрируется значение 1. Короче говоря, наличие или отсутствие диода между WL и BL определяет, какое значение хранится в ячейке ПЗУ, — 1 или 0 соответственно.

Недостатком диодной ячейки является то, что разрядная шина не изолирована от числовой шины. Весь ток, требуемый для зарядки емкости разрядной шины (а для больших ЗУ может требоваться достаточно большой ток), необходимо пропустить через числовую шину и ее задающие генераторы; следовательно, такой подход приемлем только для небольших запоминающих устройств. Гораздо лучше использовать в ячейке активное устройство, как показано на рис. 12.9, *б*. Вместо диода теперь располагается переход затвор-исток n-МОП-транзистора, сток которого соединен с источником напряжения. Работа такого устройства подобна работе диодной ячейки с одним существенным отличием: весь ток, управляющий выходом, поставляется МОП-транзистором самой ячейки. Задающий генератор числовой шины отвечает только за зарядку и разрядку емкости числовой шины.

Платой за улучшенную изоляцию служит повышенная сложность и бо́льшая площадь ячейки. Увеличение площади происходит в основном за счет дополнительного контакта с источником питания. Этот контакт необходимо реализовать в каждой ячейке памяти, чтобы шину питания можно было провести через всю матрицу. Например, на рис. 12.10 показана матрица 4×4. Обратите внимание на то, как служебные издержки, связанные с шинами питания, уменьшились за счет совместного использования этих шин соседними ячейками. Такое решение требует зеркального отражения нечетных ячеек относительно горизонтальной оси (вообще, этот подход широко используется во всех сердечниках ЗУ).

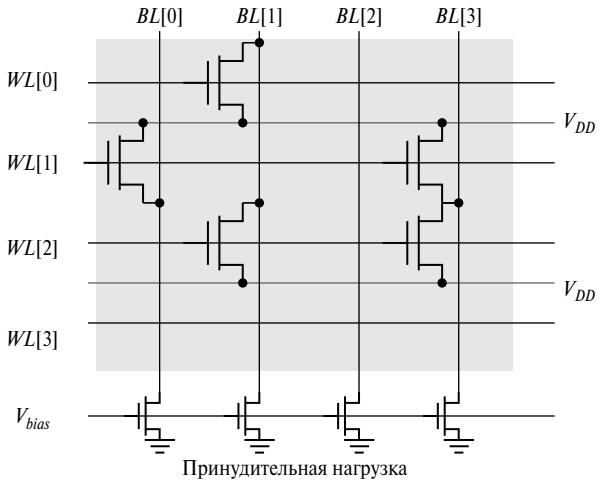


Рис. 12.10. Матрица ПЗУ 4×4 на элементах ИЛИ

Задача 12.1. Матрица ПЗУ

Определите, какие данные хранятся по адресам 0, 1, 2 и 3 ПЗУ, показанного на рис. 12.10.

Еще одной альтернативной реализацией является МОП-ячейка, изображенная на рис. 12.9, *в*. Для работы данной ячейки требуется, чтобы разрядная шина была через сопротивление связана с напряжением питания (или чтобы по умолчанию значение на выходе было равно 1). В таком случае отсутствие транзистора между WL и BL означает, что в ячейке хранится значение 1. Для реализации ячейки 0 между разрядной шиной и землей вводится МОП-устройство. При подаче высокого уровня напряжения на числовую шину устройство отпирается, что, в свою очередь, опускает уровень сигнала в разрядной шине до уровня земли (GND). На рис. 12.11 приведен пример МОП-матрицы ПЗУ 4×4. Если ни одно из присоединенных к цепи n-МОП-устройств не открыто, для повышения уровня в разрядных шинах применяется p-МОП-нагрузка.

Задача 12.2. Матрица ПЗУ на МОП-элементах НЕ-ИЛИ

Определите, какие данные хранятся по адресам 0, 1, 2 и 3 ПЗУ, приведенного на рис. 12.11.

Программирование ПЗУ

Возможно, в последнем примере вы обратили внимание на то, что объединение разрядной шины p-МОП-устройств нагрузки (принудительного перевода

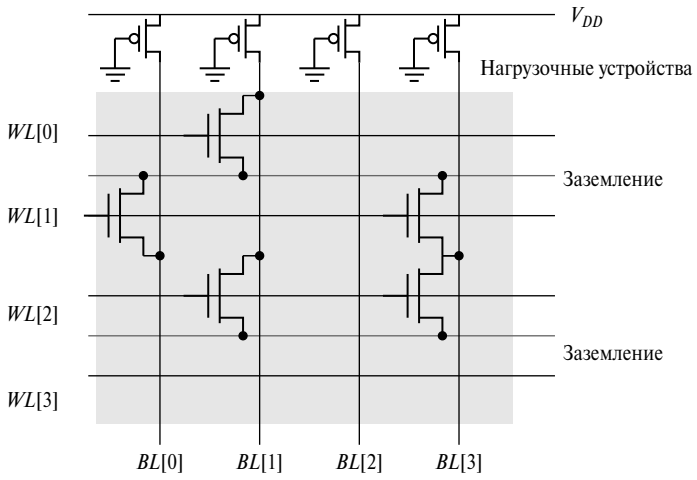


Рис. 12.11. ПЗУ 4×4 на МОП-элементах НЕ-ИЛИ

на уровень 1) и n-МОП-устройств разгрузки (принудительного перевода на уровень 0) дает не что иное, как псевдо-n-МОП-элемент НЕ-ИЛИ, входами которого являются числовые шины. Постоянное запоминающее устройство $N \times M$ можно рассматривать как объединение M элементов НЕ-ИЛИ максимум с N входами (для полностью заполненных столбцов). В таком случае часто говорят о ПЗУ на элементах НЕ-ИЛИ (NOR ROM)[3]. При нормальной работе в высоком состоянии находится только одна числовая шина и включено максимум одно устройство принудительной разгрузки. Из-за этого возникает несколько интересных моментов, связанных с масштабированием ячейки и разгрузочных транзисторов.

- Чтобы размер ячейки и емкость разрядной шины были маленькими, устройство разгрузки должно быть как можно ближе к минимальному размеру.
- С другой стороны, сопротивление нагрузочного устройства должно быть больше, чем сопротивление разгрузочного устройства, чтобы обеспечить достаточный низкий уровень. В главе 6 мы вывели, что для псевдо-n-МОП-элементов сопротивления должны отличаться минимум в четыре раза. Такое большое сопротивление существенно влияет на изменение уровня сигнала от низкого к высокому в разрядных шинах. Емкость разрядной шины определяется вкладами всех присоединенных к ней структур и для больших запоминающих устройств может составлять порядка нескольких пикофарад.

[3] Соответственно, запоминающая структура, приведенная на рис. 12.10, именуется ПЗУ на элементах ИЛИ (OR ROM).

Из-за этих моментов проектирование запоминающих и логических устройств существенно отличается. Для увеличения быстродействия и емкости памяти некоторые стандарты качества, характерные для цифровых логических элементов, немного модифицируются. В ПЗУ на элементах НЕ-ИЛИ мы можем немного пожертвовать запасом помехоустойчивости, чтобы повысить быстродействие, позволив напряжению V_{OL} в разрядной шине иметь более высокое значение (например, 1–1,5 В при напряжении питания 2,5 В). В таком случае нагрузочное устройство можно сделать шире, улучшив характеристики изменения уровня сигнала от низкого к высокому. Уменьшение запаса помехоустойчивости в сердечнике ЗУ допускается, поскольку здесь условия возникновения шума и интерференцию сигналов можно контролировать. Переход к внешним словам требует восстановления полного размаха напряжения. Для этого применяются периферийные устройства (в данном случае — усилитель считывания). Например, полный размах сигнала восстанавливает подача сигнала разрядной шины в КМОП-инвертор с должным образом настроенным порогом переключения.

На рис. 12.12 показаны две возможные топологии ПЗУ-матрицы 4×4 на элементах НЕ-ИЛИ, приведенной на рис. 12.11. Матрицы строятся путем многократного повторения одной ячейки по горизонтали и по вертикали, а также зеркального отражения нечетных ячеек (для совместного использования проводника GND). Две предложенные структуры отличаются способом программирования. В структуре, изображенной на рис. 12.12, *а*, память записывается (персонализуется) путем избирательного добавления транзисторов там, где это требуется. Это выполняется с помощью единственного диффузионного слоя (маска ACTIVE в процессе производства). При втором подходе (см. рис. 12.12, *б*) память программируется путем избирательного добавления контактов металл/диффузионный слой. Наличие металлического контакта с разрядной шиной создает ячейку 0, а отсутствие — ячейку 1. Обратите внимание на то, что в этом случае для программирования памяти применяется также один слой фотошаблона — CONTACT.

Сравнивая реализации ACTIVE и CONTACT, основанные на идентичных правилах проектирования, находим, что первая из них характеризуется экономией площади порядка 15%. С другой стороны, преимуществом стратегии CONTACT является то, что контактный слой используется в процессе производства позже активного слоя. Это позволяет отложить собственно программирование памяти до более поздних циклов процесса. Пластины можно довести до состояния, предшествующего наложению маски CONTACT, и сдать на склад. Оставшиеся этапы производства быстро выполняются, как только будет определена конкретная программа, что уменьшает промежуток времени между поступлением заказа и его выполнением. В многослойных процессах программирование часто выполняется с использованием одной из масок VIA. Конкретный способ программирования выбирается в зависимости от приоритетов заказа (размер, быстродействие или скорость выполнения заказа).

Кроме того, вы могли отметить, что для трассировки сигнала GND применяется диффузия. Вообще, такой подход совершенно недопустим. Тем не менее в запоминающих устройствах его постоянно применяют из соображе-

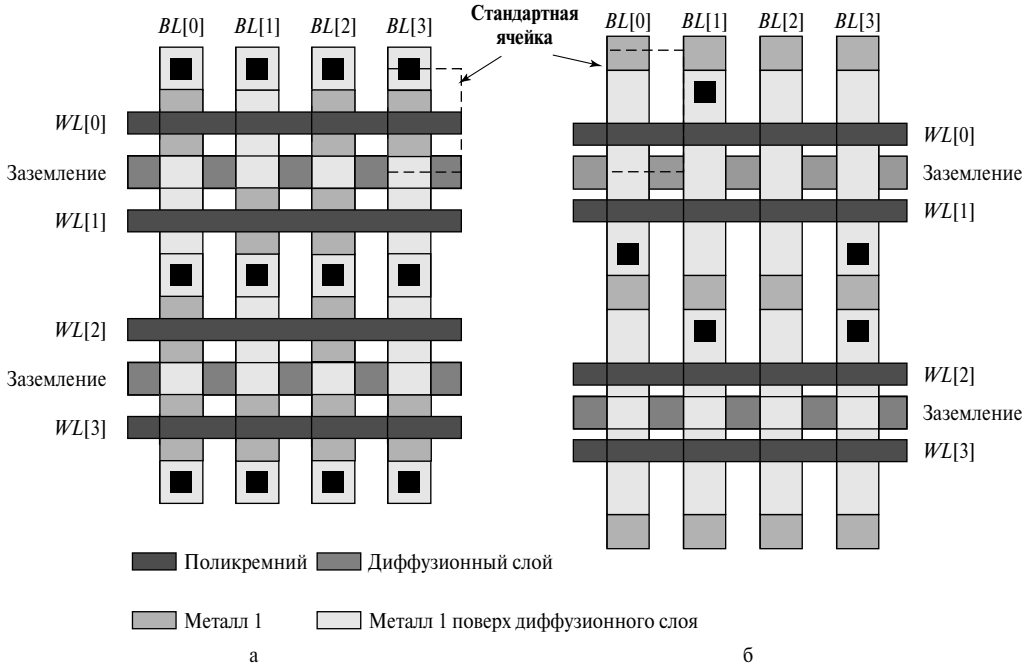


Рис. 12.12. Возможная топология ПЗУ 4×4 на МОП-элементах НЕ-ИЛИ. Разрядные шины реализованы на первом металлическом слое и проложены поверх диффузионного слоя. Шины GND распределяются горизонтально по диффузионному слою. Для программирования памяти используется слой ACTIVE (получающийся размер ячейки $9,5\lambda \times 7\lambda$ (*а*); слой CONTACT (размер ячейки $11\lambda \times 7\lambda$) (*б*)

ний повышения плотности. Для удержания скачков напряжения в проводе заземления в четко определенных пределах применяется металлический обвод с равномерно расположенными перемычками.

Здесь важно отметить, что размер транзистора составляет лишь малую часть общего размера ячейки (порядка $70\lambda^2$). В действительности размер транзистора можно даже немного увеличить, не изменяя размер ячейки. В примерах, показанных на рис. 12.12, использован транзистор с характеристическим отношением $4/2$. Значительная часть ячейки выделена под контакты разрядной шины и соединение с землей. Чтобы сократить эти служебные издержки, можно использовать другую организацию запоминающего устройства. На рис. 12.13 показана матрица ПЗУ 4×4, основанная на конфигурации НЕ-И. Все транзисторы, составляющие столбец, соединяются последовательно. Основным свойством логического элемента НЕ-И является то, что для получения низкого значения должна работать цепочка разгрузки. Для реализации этого числовые шины должны работать в реверсном режиме. По умолчанию высокий сигнал имеют все числовые шины, кроме одной строки, в которой устанавливается значение 0. Следовательно, транзисторы во всех остальных строках открыты. Предположим теперь, что на пересечении рассматриваемого столбца и строки нет транзистора. Поскольку выбраны

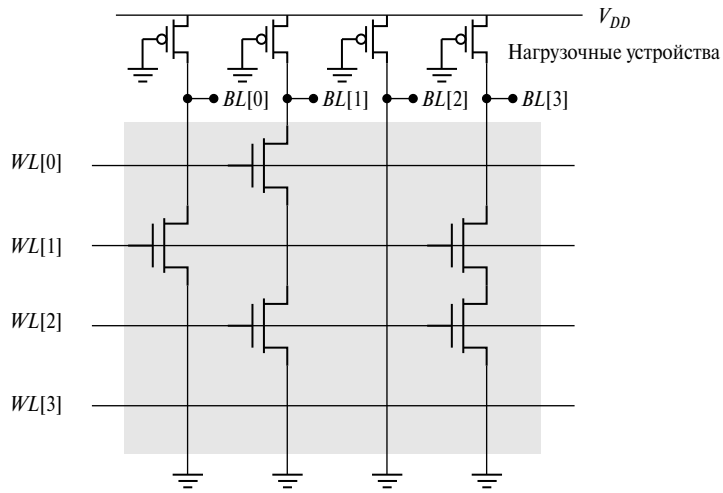


Рис. 12.13. Матрица ПЗУ 4×4 на МОП-элементах НЕ-И

все остальные транзисторы последовательной цепочки, выход переводится на низкий уровень, и значение, записанное в ячейке памяти, равно 0. С другой стороны, когда соответствующая числовая шина переводится в низкое состояние, транзистор, находящийся на пересечении, запирается. В результате выход имеет высокий уровень, что эквивалентно считыванию значения 1.

Задача 12.3. ПЗУ на МОП-элементах НЕ-И

Определите, какие данные хранятся по адресам 0, 1, 2 и 3 ПЗУ, приведенного на рис. 12.13.

Основным достоинством структуры НЕ-И является то, что элементарная ячейка — это всего лишь транзистор (или отсутствие транзистора), при этом не требуется никаких соединений с линиями питания. Это позволяет существенно уменьшить размер ячейки, что видно по топологиям, приведенным на рис. 12.14. Как и ранее, подходы к программированию ПЗУ могут быть разными. Во-первых, можно использовать первый металлический слой для избирательного закорачивания транзисторов (см. рис. 12.14, *а*). В таком случае суммарный размер ячейки составляет примерно на 15% меньше минимального размера ячейки ПЗУ на элементах НЕ-ИЛИ. Еще более существенное уменьшение площади можно получить, если для программирования памяти можно задействовать дополнительный этап имплантации. С помощью понижающего порог имплантанта с примесями n-типа устройство превращается в обедненный транзистор, который всегда находится в открытом состоянии независимо от напряжения, приложенного к числовой шине. Следовательно, этот транзистор эквивалентен короткозамкнутой цепи. Площадь получающейся в результате ячейки ($30\lambda^2$) более чем в два раза меньше, чем у экви-

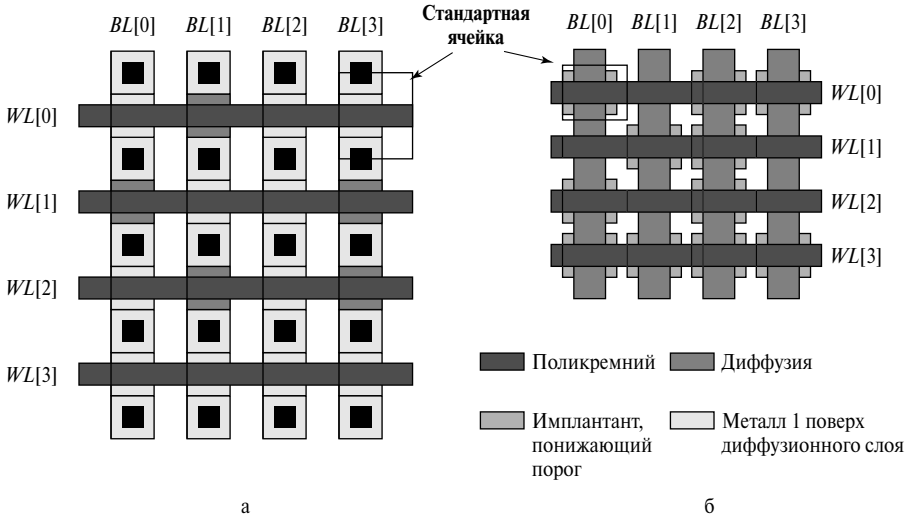


Рис. 12.14. Матрица ПЗУ 4×4 на МОП-элементах НЕ-И: *а* — с использованием программирования на первом металлическом слое (размер ячейки: $8\lambda \times 7\lambda$); *б* — с использованием имплантантов, понижающих порог (размер ячейки: $5\lambda \times 6\lambda$ в предположении, что имплантант должен распространиться на расстояние 1λ во всех направлениях от логического элемента)

валентной ячейки ПЗУ на элементах НЕ-ИЛИ. Тем не менее все имеет свою цену. В следующем разделе мы покажем, что конфигурация НЕ-И характеризуется существенным снижением быстродействия и подходит только для небольших матриц памяти. Кроме того, дополнительный этап имплантации обычно недоступен.

Пример 12.3. Размах напряжения в ПЗУ на элементах НЕ-ИЛИ и НЕ-И

Предполагая, что топологии, приведенные на рис. 12.12 и рис. 12.14, реализованы с использованием стандартной 0,25-микронной КМОП-технологии, определите размер p-МОП-устройства нагрузки, чтобы в самом худшем случае значение V_{OL} не превышало 1,5 В (напряжение питания равно а 2,5 В). В результате размах сигнала в разрядной шине будет равен 1 В. Определите соответствующие значения для матриц 8×8 и 512×512.

1. **ПЗУ на элементах НЕ-ИЛИ.** Поскольку в любой момент времени в открытом состоянии может быть не более одного транзистора, значение напряжения V_{OL} не зависит от размера матрицы и способа ее программирования. Чтобы вычислить низкое выходное напряжение, необходимо проанализировать псевдо-n-МОП-инвертор с одним разгрузочным устройством, имеющим характеристическое отношение $4/2$. Данная цепь подробно анализировалась в главе 6 для несколько иной рабочей области. Можно проверить, что для упомянутых выше условий смещения p-МОП-

и n-МОП-транзисторы находятся в режиме насыщения скорости:

$$\frac{(W/L)_p}{(W/L)_n} = \frac{k'_n[(V_{DD} - V_{Tn})V_{DSATn} - V^2_{DSATn}/2](1 + \lambda_n V_{OL})}{-k'_p[(-V_{DD} - V_{Tp})V_{DSATp} - V^2_{DSATp}/2](1 + \lambda_p(V_{OL} - V_{DD}))}.$$

Подставляя значения $V_{DD} = 2,5$ В и $V_{OL} = 1,5$ В, получаем, что отношение размеров p-МОП/n-МОП равно 2,62, или характеристическое отношение размером p-МОП-устройства равно $(W/L)_p = 5,24$.

2. **ПЗУ на элементах НЕ-И.** Из-за последовательного соединения в цепочку значение V_{OL} зависит как от емкости памяти (числа строк), так и от способа программирования. Наихудшая ситуация реализуется, когда все биты столбца имеют значение 1, т.е. когда N транзисторов соединены последовательно в разгрузочную цепь. Предположив для упрощения анализа, что N транзисторов можно заменить одним устройством в N раз большего размера, для матриц 8×8 и 512×512 находим соответствующие значения $(W/L)_p$. Обратите внимание на то, что структура, изображенная на рис. 12.14, *б*, использует устройства минимального размера с характеристическим отношением (3/2):

$$(8 \times 8) : (W/L)p = 0,49;$$
$$(512 \times 512) : (W/L)p = 0,0077.$$

Видим, что хотя первая конфигурация по-прежнему дает приемлемые результаты для p-МОП-устройства, вторая требует чересчур длинного нагрузочного транзистора, что неприемлемо. По этой причине ПЗУ на элементах НЕ-И очень редко используются для матриц, превышающих 8 или 16 строк.

Характеристика переходного режима ПЗУ

Переходная характеристика матрицы памяти определяется как промежуток времени, проходящий от момента переключения числовой шины до момента, когда сигнал разрядной шины пройдет определенный размах напряжения ΔV. Поскольку сигнал разрядной шины обычно подается на усилитель считывания, ему совсем не обязательно проходить полный размах напряжения. Изменения напряжения на ΔV вполне достаточно для того, чтобы заставить среагировать усилитель считывания. Как правило, ΔV составляет порядка 0,5 В.

Существует одно важное отличие между анализом задержки распространения логического элемента и задержки распространения матрицы памяти: в последнем случае основной вклад дают паразитные нагрузки межсоединений. Следовательно, необходимо аккуратное моделирование данных паразитных нагрузок. Сказанное иллюстрируется в приведенном ниже примере, где рассчитываются паразитное сопротивление и емкость числовой шины и разрядной шины рассмотренных ранее матриц ПЗУ, основанных на элементах НЕ-ИЛИ и НЕ-И. Мы рассмотрим этот пример достаточно подробно, поскольку

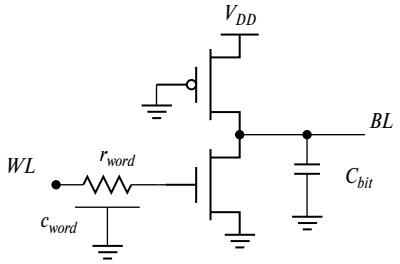


Рис. 12.15. Эквивалентная переходная модель ПЗУ на элементах НЕ-ИЛИ

аналогичный подход также применяется для анализа других запоминающих устройств (например, статических и динамических ОЗУ).

Пример 12.4. Паразитные нагрузки числовой и разрядной шины

В данном примере мы вначале выведем эквивалентную модель матриц памяти, изображенных на рис. 12.12, *б* и рис. 12.14, *б* соответственно. Если сразу учитывать все транзисторы матрицы, то мы быстро придем к "неподъемным" уравнениям и моделям (особенно для больших запоминающих устройств). Следовательно, необходимо сделать какие-то упрощения. Кроме того, мы ограничимся рассмотрением матрицы 512×512.

1. **ПЗУ на элементах НЕ-ИЛИ.** На рис. 12.15 показана модель, подходящая для анализа задержки числовой и разрядной шины ПЗУ на элементах НЕ-ИЛИ. Числовую шину лучше всего смоделировать в виде распределенной RC-сети, поскольку она реализуется в поликремнии с относительно высоким поверхностным сопротивлением. Определенно более целесообразным является использование силицидированного поликремния. С другой стороны, разрядная шина делается из алюминия, и ее сопротивление важно только при большой длине. В данном примере логично предположить, что достаточно чисто емкостной модели и все емкостные нагрузки, соединенные с проводами, можно собрать в один элемент. Паразитные нагрузки числовой шины (для матрицы памяти, изображенной на рис. 12.12, *б*) можно выразить следующим образом.

Сопротивление на ячейку: $(7/2) \times 5$ Ом $= 17,5$ Ом (согласно табл. 4.5)
Емкость провода на ячейку: $(3\lambda \times 2\lambda)(0,125)^2 0,088 + 2 \times (3\lambda \times 0,125) \times$
$\times 0,054 = 0,049$ фФ (согласно табл. 4.2)
Емкость затвора на ячейку: $(4\lambda \times 2\lambda)(0,125)^2 6 = 0,75$ фФ.

Паразитные нагрузки разрядной шины выражаются следующим образом.

Сопротивление на ячейку: $(11/4) \times 0,1$ Ом $= 0,275$ Ом
(пренебрежимо мало).

Емкость провода на ячейку: $(11\lambda \times 4\lambda)(0,125)^2 0,041 +$
$+ 2 \times (11\lambda \times 0,125) \times 0,047 = 0,09$ фФ

Емкость стока на ячейку: $((5\lambda \times 4\lambda)(0,125)^2 \times 2 \times 0,56 + 14\lambda \times 0,125 \times$
$\times 0,28 \times 0,6 + 4\lambda \times 0,125 \times 0,31 = 0,8$ фФ.

Последнее значение стоит объяснить подробнее. *Емкость стока*, в которую дают вклады все ячейки, соединенные с разрядной шиной, состоит из емкостей нижнего соединения, боковой стенки и емкости перекрытия. Можно предположить, что все ячейки, кроме переключаемой в данный момент, находятся в закрытом состоянии, поэтому учитывается только та часть емкости затвора, которая связана с перекрытием. Кроме того, предполагается, что размах напряжения в разрядной шине составляет от 1,5 до 2,5 В. В таком случае коэффициент K_{eq} равен 0,56 и 0,6 для поверхности и боковой стенки соответственно. Все остальные значения, использованные в расчетах, взяты из табл. 3.5.

2. **ПЗУ на элементах НЕ-И.** Как и в случае с ПЗУ на элементах НЕ-ИЛИ, мы можем разработать эквивалентную модель для анализа задержки структуры НЕ-И. Модель числовой шины будет такой же, как и ранее, а вот смоделировать поведение разрядной шины будет несколько сложнее из-за длинной цепочки последовательно соединенных транзисторов. Наихудший сценарий здесь реализуется, когда переключается транзистор внизу цепочки, а весь столбец заполнен транзисторами. Модель, аппроксимирующая поведение устройства для данной ситуации, показана на рис. 12.16. Каждый из последовательно соединенных транзисторов (разумеется, все они находятся в открытом состоянии) моделируется парой сопротивление–емкость. Для больших запоминающих устройств всю цепочку можно смоделировать как распределенную rc-цепь.

   Паразитные нагрузки числовой шины (для матрицы памяти, изображенной на рис. 12.14) выражаются следующим образом.

Сопротивление на ячейку: $(6/2) \times 5$ Ом $= 15$ Ом.

Емкость провода на ячейку: $(3\lambda \times 2\lambda)(0,125)^2 0,088 + 2 \times (3\lambda \times 0,125) \times$
$\times 0,054 = 0,049$ фФ.

Емкость затвора на ячейку: $(3\lambda \times 2\lambda)(0,125)^2 6 = 0,56$ фФ.

Паразитные нагрузки разрядной шины выражаются следующим образом.

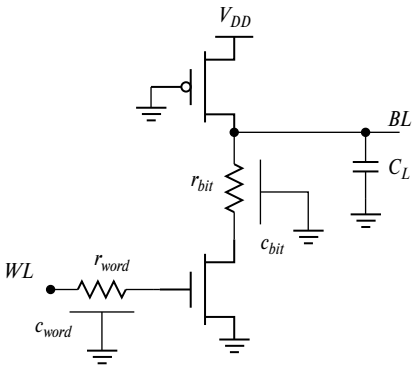


Рис. 12.16. Эквивалентная модель числовой шины и разрядной шины ПЗУ на элементах НЕ-И

Сопротивление на ячейку: $13/1,5 = 8,7$ кОм.

Емкость провода на ячейку: включена в диффузионную емкость.

Наконец, емкость истока/стока на одну ячейку выражается следующим образом:

$$(3\lambda \times 3\lambda)(0,125)^2 \times 2 \times 0,56 + 2 \times 3\lambda \times 0,125 \times 0,28 \times 0,6 + \\ + (3\lambda \times 2\lambda)(0,125)^2 \times 6 = 0,85 \text{ фФ}.$$

Емкость истока/стока должна включать емкости затвора/истока и затвора/стока, т.е. представлять полную емкость элемента. Здесь ситуация отличается от анализа ПЗУ на элементах НЕ-ИЛИ, где учитывалась только емкость перекрытия стока-истока.

Определить среднее сопротивление транзистора несколько сложнее. Фактически сопротивление отличается для разных устройств в зависимости от их положения в цепи (это вызвано различными напряжениями затвора-истока и влиянием подложки). При анализе в первом приближении мы используем эквивалентное сопротивление, выведенное в главе 3. Тем не менее в главе 6 (пример 6.4) с помощью такой аппроксимации мы получили достаточно разумные результаты для элемента НЕ-И с четырьмя входами.

Возможно, вы не понимаете, зачем нам вообще нужны аналитические расчеты, если с помощью моделирования можно получить более точные результаты. Дело в том, что модули памяти могут содержать тысячи и даже миллионы транзисторов, из-за чего моделирование выполняется невероятно

медленно. Кроме того, аналитические расчеты помогают лучше понять поведение запоминающего устройства. Компьютерное моделирование не покажет, где находятся узкие места структуры и как от них избавиться.

С помощью вычисленных данных и эквивалентной модели мы можем теперь вывести приблизительные значения задержки распространения сердечника запоминающего устройства и ее составляющие.

Пример 12.5. Задержка распространения ПЗУ 512×512 на элементах НЕ-ИЛИ

1. Задержку числовой шины, представленной как распределенная rc-цепь из M ячеек, можно аппроксимировать с помощью выражений, выведенных в главе 4:

$$t_{word} = 0,38(r_{word}\times c_{word})M^2 = 0,38(17,5\Omega\times(0,049+0,75)\text{ фФ})512^2 = 1,4\text{ нс}.$$

2. Для разрядной шины время реакции будет зависеть от направления перехода. Предполагая, что разгрузочное устройство характеризуется размерами (0,5/0,25), а нагрузочный транзистор — размерами (1,3125/0,25), как в примере 12.3, можем с помощью привычных методов вычислить задержку распространения:

$$C_{bit} = 512\times(0,8+0,09)\text{ фФ} = 0,46\text{ пФ};$$
$$t_{HL} = 0,69(13\text{ кОм}/2||31\text{ кОм}/5,25)0,46\text{ пФ} = 0,98\text{ нс}.$$

С помощью похожего подхода можно вычислить время перехода сигнала с низкого уровня на высокий:

$$t_{LH} = 0,69(31\text{ кОм}/5,25)0,46\text{ пФ } = 1,87\text{ нс}.$$

Изучая полученные результаты, видим, что основной является задержка числовой шины, выражаемая через большое сопротивление поликремниевого провода. Следовательно, здесь нам пригодятся некоторые техники уменьшения задержки распределенных rc-цепей, введенные в главе 9. Для начала проблему задержки числовой шины можно смягчить, подав питание на оба конца адресной шины и использовав металлические обходные линии (часто именуемые *глобальными числовыми шинами*) (см. рис. 9.18). Наиболее эффективным на данное время подходом является аккуратное разбиение памяти на субблоки подходящего размера, которые позволяют выравнять задержки числовой и разрядной шины. Кроме того, разбиение помогает снизить энергопотребление, связанное с питанием и переключением числовых шин.

Если требуется, можно также уменьшить задержку разрядной шины. Наиболее часто для этого уменьшается размах напряжения на разрядной

шине, а восстановление выходного сигнала до полного размаха выполняется с помощью усилителя считывания. Как правило, уменьшенный размах напряжения составляет около 0,5 В.

Пример 12.6. Задержка распространения ПЗУ на элементах НЕ-И

Используя технологии, подобные представленным в примере 12.5, определим задержку числовой и разрядной шины матрицы ПЗУ 512 × 512 на элементах НЕ-И.

1. Задержка числовой шины получается примерно такой же, как и для реализации на элементах НЕ-ИЛИ:

$$t_{word} = 0,38(r_{word} \times c_{word})M^2 = 0,38(15\Omega \times (0,049 + 0,56)\ \text{фФ})512^2 = 1,3\ \text{нс}.$$

2. Для задержки разрядной шины наихудшая ситуация реализуется, когда весь столбец, за исключением одной ячейки, заполнен битами 0, и включенным является самый нижний транзистор. С помощью распределенной модели находим хорошую аппроксимацию задержки (хотя здесь мы игнорируем влияние нагрузочного транзистора):

$$t_{HL} = 0,38 \times 8,7\ \text{кОМ} \times 0,85\ \text{фФ}\ \times 511^2 = 0,73\ \text{мкс}.$$

Наихудший случай для изменения уровня сигнала от низкого к высокому реализуется, когда заперт самый нижний транзистор. Используя формулу Элмора, находим:

$$t_{LH} = 0,69(31\ \text{кОи}/0,0077)(511 \times 0,85\ \text{фФ}) = 1,2\ \text{мкс}.$$

Очевидно, в большинстве случаев такие задержки неприемлемы. Следовательно, единственным реальным выходом является разбиение запоминающего устройства на меньшие модули.

Потребление мощности и предварительно заряженные матрицы памяти

Предложенные структуры на основе элементов НЕ-И и НЕ-ИЛИ обладают всеми недостатками псевдо-n-МОП-элементов, рассмотренных в главе 6.

1. **Относительная логика**. Напряжение V_{OL} определяется отношением размеров нагрузочных и разгрузочных устройств. Это может привести к неприемлемым характеристикам транзисторов, как было показано на примерах выше.

2. **Статическое потребление мощности**. При низком значении выходного сигнала между шинами питания существует ненулевой ток покоя. Это может породить серьезные проблемы с рассеянием мощности.

Пример 12.7. Статическое рассеяние мощности ПЗУ на элементах НЕ-ИЛИ

Рассмотрим матрицу ПЗУ 512×512 на элементах НЕ-ИЛИ. Логично предположить, что в среднем 50% выходов имеют низкий уровень. Ток холостого хода для структуры, описанной в примере 12.4, приблизительно равен 0,21 мА (выходное напряжение равно 1,5 В). При этом суммарное статическое рассеяние составляет $(512/2) \times 0,21$ мА $\times 2,5$ В $= 0,14$ Вт, и эта мощность расходуется даже в том случае, когда в схеме ничего не происходит. Разумеется, это недопустимо.

Чтобы решить две названные выше проблемы, можно воспользоваться теми же принципами, что и при проектировании цифровых логических элементов. Например, можно использовать комплементарные элементы НЕ-И или НЕ-ИЛИ. Правда, с точки зрения требуемой площади данный подход выглядит непривлекательно из-за большего числа транзисторов и соединения с обеими шинами питания. Гораздо лучше использовать предварительно заряженную логику, как показано на рис. 12.17. Данный подход позволяет избавиться и от статического рассеяния мощности, и от относительной логики, не усложняя при этом общую структуру ячейки. Поскольку логическая структура ПЗУ на элементах НЕ-И и НЕ-ИЛИ проста и имеет глубину всего в один слой, можно гарантировать, что в процессе предварительной зарядки все разгрузочные цепи выключены. Благодаря этому мы можем убрать n-МОП-транзистор внизу разгрузочной цепи.

Динамическая архитектура позволяет независимо управлять синхронизацией разгрузочной и нагрузочной цепи. Например, p-МОП-устройство предварительной зарядки можно сделать настолько большим, насколько это требуется. Помните, однако, что данный транзистор нагружает генератор тактового сигнала, который из-за этого может быть сложно спроектировать.

Задача 12.4. ПЗУ с предварительной зарядкой на элементах НЕ-И

Хотя предварительная зарядка очень хорошо подходит для ПЗУ на элементах НЕ-ИЛИ, при использовании в разработке элементов НЕ-И возникают серьезные проблемы. Объясните, почему.

Благодаря своим особенностям подход с использованием предварительной зарядки является наиболее предпочтительным при разработке запоминающих устройств. Практически все современные большие блоки памяти (в том числе энергонезависимые и оперативные запоминающие устройства) реализуются с использованием динамической предварительной зарядки.

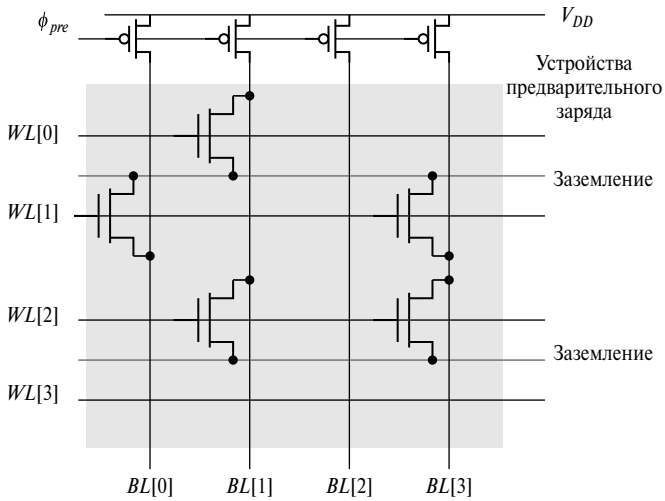


Рис. 12.17. Предварительно заряженная МОП-матрица ПЗУ 4×4 на элементах НЕ-ИЛИ

ПЗУ: перспективы с точки зрения пользователя

Следует помнить, что большинство проблем, упоминавшихся в предыдущих разделах, универсальны по своей природе и присущи всем другим архитектурам запоминающих устройств. Прежде чем рассматривать некоторые из описанных структур, стоит обсудить классификацию модулей ПЗУ и подходов к программированию. К первому классу модулей ПЗУ относятся так называемые *специализированные* (application-specific) ПЗУ, когда модуль памяти является частью большего заказного проекта и программируется только для этого приложения. При работе над такими проектами разработчику дается полная свобода действий и для программирования устройства он может использовать любой фотошаблон или комбинацию фотошаблонов.

Второй класс модулей ПЗУ формируют *стандартные* (commodity) кристаллы ПЗУ, характеризуемые тем, что производитель массово создает модули памяти, которые позже настраиваются согласно спецификациям заказчика. В подобных случаях необходимо, чтобы число этапов процесса, задействованных в программировании, было минимальным и их можно было выполнить в последней фазе процесса производства. При таком подходе можно заблаговременно заготовить большие партии незапрограммированных кристаллов. Как показывалось на примерах выше, при использовании описанного метода *программируемого фотошаблона* для программирования памяти чаще всего используется контактный (или металлический) фотошаблон. В программировании модуля ПЗУ участвует производитель кристалла, что создает нежелательную задержку в процессе разработки изделия. В сфере систем на кристалле существует одна модификация данного подхода, в которой предварительной обработке подвергается практически весь кристалл. Программирование с помощью фотошаблона предусматривается только на

небольшой части кристалла (как правило, при этом задействуются верхние металлические слои). Такой подход можно использовать, например, для программирования микроконтроллеров, встроенных в кристалл (применяются во многих приложениях).

Для клиента наиболее удобен подход, когда он может самостоятельно запрограммировать запоминающее устройство. Одна из технологий, предлагающих такую возможность, — *программируемое запоминающее устройство* (Programmable Read-Only Memory ROM — PROM). Данная структура позволяет заказчику один раз запрограммировать память. Чаще всего такая возможность реализуется с помощью *плавких перемычек* (нихром, поликремний или другие проводники) в ячейке памяти. В фазе программирования некоторые из этих перемычек пережигаются посредством подачи высокого тока, в результате чего присоединенные к ним транзисторы отключаются.

Хотя структуры PROM привлекательны с точки зрения возможности программирования пользователем, они не очень удобны из-за записи в ходе одной фазы. Например, из-за единственной ошибки в процессе программирования устройство становится непригодным для использования. Из-за этого в настоящее время большей популярностью пользуются устройства, допускающие перепрограммирование (хотя и медленное). Реализация данного принципа описывается в следующем разделе.

12.2.2. Энергонезависимые оперативные запоминающие устройства

Архитектура энергонезависимых оперативных запоминающих устройств практически идентична архитектуре ПЗУ. Сердечник ЗУ состоит из матрицы транзисторов, расположенных в сетке числовых/разрядных шин. Память программируется путем избирательного включения или выключения данных устройств. В ПЗУ для этого применяется перестройка на уровне фотошаблонов. В энергонезависимых ОЗУ часто используется модифицированный транзистор, порог которого можно менять с помощью электрического тока. В модифицированном состоянии порог сохраняется бесконечно долго (или, по крайней мере, очень долго), даже при выключенном напряжении питания. Чтобы перепрограммировать память, запрограммированные значения необходимо *стереть*, а после этого задать новые значения. Различные классы перепрограммируемых энергонезависимых устройств отличаются именно методом стирания информации. Отметим также, что процесс программирования памяти обычно выполняется на порядок медленнее, чем операция чтения.

В начале данного раздела мы опишем транзистор с плавающим затвором — элемент, являющийся сердцем подавляющего большинства перепрограммируемых запоминающих устройств. Остальная часть раздела посвящена различным модификациям этого элемента, в основном отличающимся реализациями процесса стирания информации. В результате развития данных модификаций возникли различные семейства энергонезависимых ОЗУ. В конце раздела описываются некоторые развивающиеся классы энергонезависимых ОЗУ.

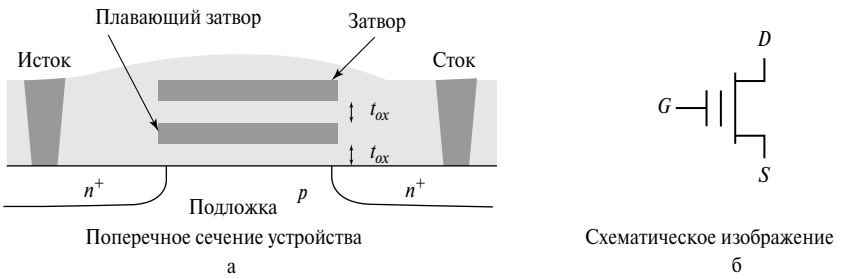


Рис. 12.18. МОП-транзистор с плавающим затвором и лавинной инжекцией заряда (floating-gate avalanche-injection metal-oxide-semiconductor — FAMOS)

Транзистор с плавающим затвором

Попытки создать устройство, характеристики которого можно было бы менять с помощью электрического тока и которое было бы достаточно надежным для реализации нескольких циклов перезаписи, продолжались достаточно долго. Например, многообещающей разработкой казался МНОП-транзистор (транзистор со структурой металл-нитрид-оксид-полупроводник), к сожалению, не нашедший никакого практического применения до настоящих дней. В этом устройстве электроны, модифицирующие порог, захватывались в слое Si_3N_4, расположенном поверх затвора SiO_2 [Chang77]. Более приемлемым оказалось решение, реализованное в транзисторе с плавающим затвором (рис. 12.18), который и стал основой практически всех энергонезависимых устройств, разрабатываемых в наши дни. Структура данного транзистора похожа на структуру традиционного МОП-устройства, только между затвором и каналом вводится дополнительная поликремниевая полоска. Данная полоска ни к чему не присоединяется и называется *плавающим затвором*. Наиболее очевидное последствие введения дополнительного затвора — удвоение толщины подзатворного окисла t_{ox}, что приводит к уменьшению крутизны характеристики прямой передачи транзистора и увеличению порогового напряжения. Обе эти особенности являются нежелательными. Со всех остальных точек зрения устройство действует как обычный транзистор.

Однако гораздо важнее то, что данное устройство обладает одним интересным свойством: программируемым пороговым напряжением. Прикладывая высокое напряжение (более 10 В) между истоком и затвором-стоком, мы получим сильное электрическое поле, вызывающее лавинную инжекцию. Электроны получают достаточно энергии, чтобы пройти через первый защитный слой окисла, в результате чего они захватываются плавающим затвором. Данное явление возможно при толщине оксида 100 нм, т.е. производство соответствующих устройств не представляет особых проблем. Из-за особенностей механизма программирования транзистор с плавающим затвором иногда называют *МОП-транзистором с плавающим затвором и лавинной инжекцией заряда* (floating-gate avalanche-injection metal-oxide-semiconductor — FAMOS) [Frohman74].

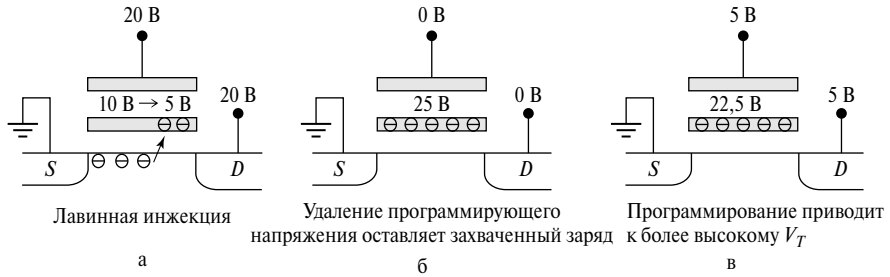


Рис. 12.19. Программирование транзистора с плавающим затвором

Захват электронов на плавающем затворе эффективно понижает напряжение на этом затворе (см. рис. 12.19, *а*.) Данный процесс является самосинхронизирующимся — отрицательный заряд, накапливаемый на плавающем затворе, уменьшает электрическое поле на оксиде, которое в конечном счете не может больше ускорять горячие электроны. После снятия напряжения индуцированный отрицательный заряд остается, порождая отрицательное напряжение на промежуточном затворе (см. рис. 12.19, *б*). С точки зрения устройства это эффективно увеличивает пороговое напряжение. Чтобы включить устройство, необходимо подать более высокое напряжение, которое компенсирует влияние индуцированного отрицательного заряда (рис. 12.19, *в*). Как правило, получающееся в результате пороговое напряжение составляет порядка 7 В; следовательно, для включения транзистора недостаточно обычного напряжения между затвором и истоком 5 В и устройство не задействовано. Заряд, инжектированный в плавающий затвор, эффективно смещает вольт-амперную характеристику транзистора, как показано на рис. 12.20. Кривая A соответствует состоянию “0”, а кривая B — транзистору со смещением V_T, который представляет состояние “1”. Можно показать, что величина смещения V_T выражается следующим образом:

$$\Delta\ V_T = (-\Delta Q_{FG})/C_{FC}, \tag{12.1}$$

где C_{FC} — емкость между внешним контактом затвора и плавающим затвором; ΔQ_{FG} — заряд, инжектированный в плавающий затвор.

Поскольку плавающий затвор окружен SiO_2 (прекрасным изолятором), захваченный заряд может храниться многие годы даже при отключенном напряжении питания. Таким образом мы получаем энергонезависимый механизм хранения информации. Одной из основных проблем транзисторов с плавающим затвором является потребность в высоком напряжении для программирования. Экспериментируя с примесями, технологи смогли понизить требуемое напряжение с первоначальных 25 В приблизительно до 12,5 В (современные запоминающие устройства).

В настоящее время практически все энергонезависимые запоминающие устройства основаны на подходе с использованием плавающих затворов. Отличаются они только механизмом, применяющимся для стирания памяти.

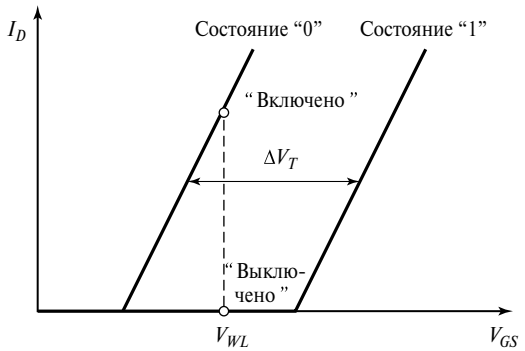


Рис. 12.20. Вольт-амперная характеристика, получаемая в результате программирования горячими электронами. Подача на числовую шину напряжения V_{WL} приводит к тому, что ток либо проходит (состояние “0”), либо не проходит (состояние “1”)

Стираемое программируемое постоянное запоминающее устройство (СППЗУ)

Информация СППЗУ стирается при облучении ячеек ультрафиолетом через прозрачное окошко в корпусе. Под воздействием ультрафиолетового излучения оксид начинает проводить ток из-за прямой генерации в материале электронно-дырочных пар. Процесс стирания происходит медленно — от нескольких секунд до нескольких минут — в зависимости от интенсивности ультрафиолетового излучения. Программирование выполняется со скоростью примерно 5–10 мс/слово. Еще одной проблемой, связанной с данным подходом, является *ограниченная износостойкость*, — число циклов стирания/программирования обычно не превышает одной тысячи (в основном из-за использования ультрафиолета). Кроме того, не все гладко с надежностью. Чем больше прошло циклов программирования, тем выше шансы на то, что пороги устройств изменятся. Из-за этого большинство СППЗУ содержат встроенные схемы, во время программирования удерживающие пороги в заданных диапазонах. Наконец, инжекция всегда влечет за собой большой ток канала (0,5 мА при напряжении затвора 12,5 В), а из-за этого в процессе программирования возникает большое рассеяние мощности.

С другой стороны, ячейка СППЗУ является крайне простой и имеет плотную структуру, благодаря чему достаточно дешево можно произвести большое количество таких запоминающих устройств. Поэтому СППЗУ были весьма привлекательны в приложениях, не требующих регулярного перепрограммирования. В настоящее время на смену СППЗУ пришла флэш-память.

Электрически стираемые перепрограммируемые постоянные запоминающие устройства (ЭСППЗУ)

Основным недостатком СППЗУ является то, что процедура стирания должна происходить “вне системы”. Это означает, что для перепрограммирования память необходимо снять с платы и поместить в соответствующий програм-

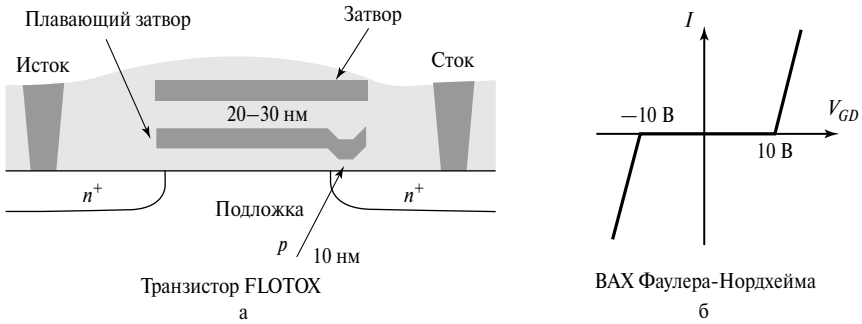


Рис. 12.21. FLOTOX-транзистор, программируемый с использованием туннелирования Фаулера–Нордхейма

матор. Данный недостаток был устранен в *электрически стираемых перепрограммируемых постоянных запоминающих устройствах*, ЭСППЗУ (Electrically Erasable Programmable Read-Only Memory — EEPROM, E²PROM). В этом подходе описанная трудоемкая и неудобная процедура не требуется благодаря использованию другого механизма инжекции или снятия заряда с плавающего затвора — *туннелирования*. Модифицированное устройство с плавающим затвором, получившее название *FLOTOX-транзистора* (Floating-Gate Tunneling Oxide transistor — транзистор с плавающим затвором с туннелированием в окисле), поддерживает процедуру электрического стирания [Johnson80]. Поперечное сечение структуры FLOTOX-транзистора показано на рис. 12.21, *а*. Внешне он напоминает устройство FAMOS, только диэлектрическая прослойка, отделяющая плавающий затвор от канала и стока, теперь уменьшена до толщины порядка 10 нм и даже меньше. При подаче на этот тонкий слой диэлектрика напряжения, примерно равного 10 В (эквивалентно электрическому полю порядка 10^9 В/м), электроны проходят через плавающий затвор в обе стороны благодаря *туннелированию Фаулера–Нордхейма* (Fowler–Nordheim tunneling) [Snow67]. Вольт-амперная характеристика туннельного перехода показана на рис. 12.21, *б*.

Основным достоинством описанного подхода к программированию является его обратимость, т.е. стирание выполняется путем подачи напряжения, обратного к тому, которое было приложено в процессе записи. Инжекция электронов в плавающий затвор увеличивает порог, а обратная операция понижает V_T. Однако подобная симметрия порождает проблему контроля порога: если снять с плавающего затвора слишком большой заряд, получим обедненное устройство, которое нельзя запереть с помощью стандартных сигналов числовой шины. Обратите внимание на то, что получаемое пороговое напряжение зависит от исходного заряда затвора, а также от приложенного программирующего напряжения. Кроме того, оно сильно зависит от толщины окисла, которая может существенно колебаться в пределах кристалла. Чтобы решить эту проблему, в ячейку ЭСППЗУ вводится дополнительный транзистор, последовательно соединяемый с транзистором с плавающим затвором. Новый транзистор действует как устройство доступа в процессе чтения, а

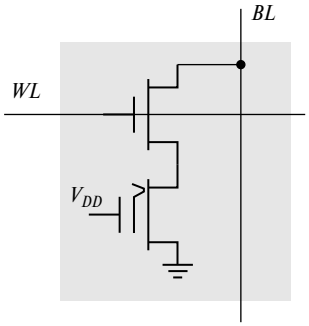


Рис. 12.22. Конфигурация ячейки EEPROM в процессе чтения. При программировании порог транзистора FLOTOX выше, чем V_{DD}, в результате чего ячейка отключается. В остальное время транзистор представляет собой запертый ключ

для хранения информации задействуется FLOTOX-транзистор (рис. 12.22.) Напомним, что в ячейке СППЗУ транзистор FAMOS и применялся для программирования, и выступал как устройство доступа.

Ячейка ЭСППЗУ, содержащая два транзистора, больше по размеру аналогичной ячейки СППЗУ. Данное увеличение площади дополнительно усугубляется тем фактом, что устройство FLOTOX само по себе больше транзистора FAMOS из-за дополнительного объема туннелирующего оксида. Кроме того, производство очень тонкого слоя оксида является достаточно трудным и дорогим. Из-за этого емкость компонентов ЭСППЗУ составляет меньше битов за бо́льшую цену, чем у компонентов СППЗУ. Хорошая новость заключается в том, что компоненты ЭСППЗУ предлагают большую эксплуатационную гибкость и обычно служат дольше — до 10^5 циклов стирания/записи[4]. Многократное программирование вызывает смещение порогового напряжения из-за необратимого захвата электронов в SiO_2. В конечном итоге это приводит к неправильной работе или невозможности перепрограммировать устройство.

Флэш-ЭСППЗУ

Концепция электрически-стираемых перепрограммируемых флэш-ПЗУ появилась в 1984 году и быстро стала одной из наиболее популярных архитектур энергонезависимых запоминающих устройств. Флэш-память объединяет в себе плотность СППЗУ и универсальность ЭСППЗУ, а цена и функциональные возможности являются средними между данными архитектурами.

Технически флэш-память — это гибрид СППЗУ и ЭСППЗУ. В большинстве флэш-устройств для программирования применяется метод лавинной инжекции горячих электронов. Стирание выполняется с использованием тун-

[4] Тонкие оксидные пленки — не единственный способ реализации туннелирования электронов. Существуют и другие подходы, например локальное усиление поверхностного поля с использованием текстурных поверхностей [Masuoka91].

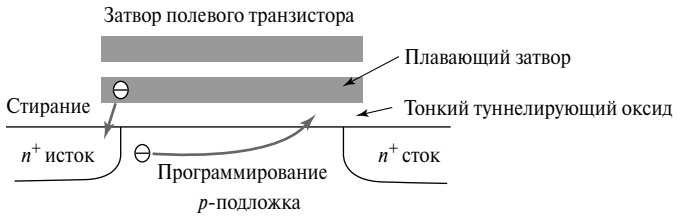


Рис. 12.23. Устройство ETOX, используемое во флэш-памяти

нелирования Фаулера–Нордхейма, как и для ячеек ЭСППЗУ. Основное отличие состоит в том, что стирание происходит по всему кристаллу или только для одного участка памяти. Хотя это снижает универсальность устройства, при этом можно обойтись без дополнительного транзистора доступа, требуемого в ЭСППЗУ. Стирание всего сердечника ЗУ позволяет тщательно отслеживать характеристики устройства в процессе стирания и гарантировать, что незапрограммированный транзистор будет действовать как устройство обогащения. Находящееся в кристалле памяти устройство управления наблюдением регулярно проверяет значение порога в процессе стирания и динамически меняет время стирания. Отметим, что такой подход практичен только при стирании за раз больших фрагментов памяти (отсюда и название flash — "вспышка"). Благодаря более простой структуре ячейки запоминающее устройство в целом является более компактным и допускает б*о*льшую плотность интеграции.

Например, на рис. 12.23 показана разработанная Intel флэш-память ETOX [Pashley89], представляющая одну из множества возможных альтернатив. ETOX похожа на элемент FAMOS, только имеет очень тонкий слой туннелирирующего подзатворного окисла (10 нм). Для программирования и стирания используются различные области подзатворного окисла. Программирование выполняется путем прикладывания высокого напряжения (12 В) к затвору и стоку при заземленном истоке, а стирание происходит при заземленном затворе и напряжении истока 12 В.

На рис. 12.24 показано, как можно ввести данную ячейку в структуру ПЗУ на элементах НЕ-ИЛИ. Цикл программирования начинается с *операции стирания* (см. рис. 12.24, *а*): на затвор подается напряжение 0 В, а на исток — напряжение 12 В. Электроны, если они есть на плавающем затворе, инжектируются в исток посредством туннелирования. Все ячейки стираются одновременно. В конце операции стирания из-за различных исходных значений пороговых напряжений ячейки и колебаний толщины окисла могут проявиться колебания порогового напряжения. С этим борются двумя способами: (1) перед подачей стирающего импульса все ячейки матрицы программируются так, чтобы все пороги изначально имели приблизительно одинаковые значения; (2) после этого подается стирающий импульс управляемой длительности. Затем для проверки правильности стирания считывается информация со всей матрицы. Если операция прошла неудачно, подается следующий стирающий импульс, а после этого снова выполняется считывание.

Описанный алгоритм повторяется до тех пор, пока пороговое напряжение во всех ячейках не опустится ниже требуемого уровня. Как правило, цикл стирания занимает от 100 мс до 1 с. В *цикле записи* (*программирования*) (см. рис. 12.24, *б*) на затвор выбранного устройства подается высоковольтный импульс. Если в это время на сток подан сигнал "1", генерируются горячие электроны и инжектируются в плавающий затвор, повышая порог транзистора (устройство переходит в разряд всегда запертого). В противном случае плавающий затвор остается в предыдущем состоянии "0" (соответствующем отсутствию электронов). Чтобы получить необходимое смещение порога в 3–3,5 В, обычно требуется импульс длительностью 1–10 мкс. Операция чтения (см. рис. 12.24, *в*) происходит так же, как и в любой другой ПЗУ-структуре на элементах НЕ-ИЛИ. Чтобы выбрать ячейку, напряжение в ее числовой шине повышается до 5 В, что приводит к разрядке разрядной шины.

Архитектура элементов НЕ-ИЛИ позволяет получить при произвольном доступе малое время чтения. При этом процессы стирания и программирования происходят достаточно медленно из-за необходимости тщательного контроля за порогами. Из-за данных особенностей флэш-память не очень подходит для приложений, подобных хранению программного кода. В других приложениях, например хранения видео- или аудиофайлов, произвольный доступ не требуется, а важны большая плотность упаковки информации (уменьшает стоимость ЗУ), быстрое стирание и программирование, а также быстрый последовательный доступ. Этим требованиям легко удовлетворить с помощью описанной ранее архитектуры ПЗУ на элементах НЕ-И. Поэтому многие производители флэш-памяти используют именно эту топологию. В данном случае, как показано на рис. 12.25, стандартный модуль состоит из 8–16 соединенных последовательно транзисторов с плавающим затвором. Данная цепочка соединяется с разрядной шиной и землей с помощью двух транзисторов выбора. Из-за отсутствия контактов между числовыми шинами размер получающейся ячейки примерно на 40% меньше, чем у ячейки на элементах НЕ-ИЛИ. В показанном на рисунке запоминающем устройстве операции программирования и стирания реализуются с использованием туннелирования Фаулера–Нордхейма. В процессе стирания все ячейки модуля программируются как обедненные устройства (т.е. транзисторы с отрицательным порогом). Для этого в разрядную шину и шину питания подается 20 В, а на числовую шину — напряжение 0 В. В процессе программирования режим транзисторов выбора устанавливается так, чтобы разрядная шина была включена цепь, а шина питания — изолирована. Для записи значения "1" разрядная шина заземляется и на искомую числовую шину подается высокое напряжение (20 В). Электроны туннелируют в плавающий затвор, вызывая повышение порога. Для записи значения "0" порог ячейки остается без изменений (сигнал в разрядной шине остается на высоком уровне). Операция чтения выполняется так же, как в ПЗУ на элементах НЕ-И, когда активизированы оба транзистора выбора. Использование туннелирования в качестве основного механизма уменьшает требуемый ток и позволяет реализовать параллельное программирование большого числа модулей при контролируемом потреблении мощности. Время программирования из расчета на один байт со-

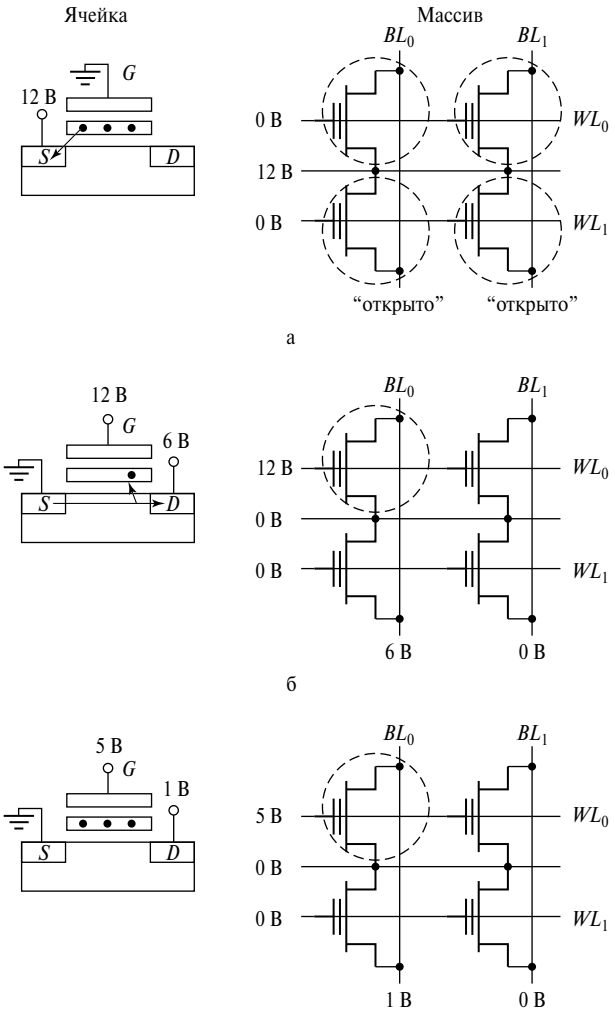


Рис. 12.24. Базовые операции флэш-памяти на элементах НЕ-ИЛИ (согласно [Itoh01]): *а* — стирание; *б* — запись; *в* — чтение

ставляет порядка 200–400 нс. Обрабатывая за раз весь модуль, можно добиться быстрого и надежного стирания, поэтому время стирания всего кристалла составляет порядка 100 мс.

Было определено и множество других архитектур флэш-памяти, в которых выбирались различные соотношения между скоростью чтения, стирания и программирования, а также плотностью интеграции. Все они основаны на описанных выше принципах. Более подробное описание флэш-памяти можно найти в [Itoh01]. Нам же хотелось бы остановиться еще на одном важном вопросе. Все энергонезависимые запоминающие устройства требуют высо-

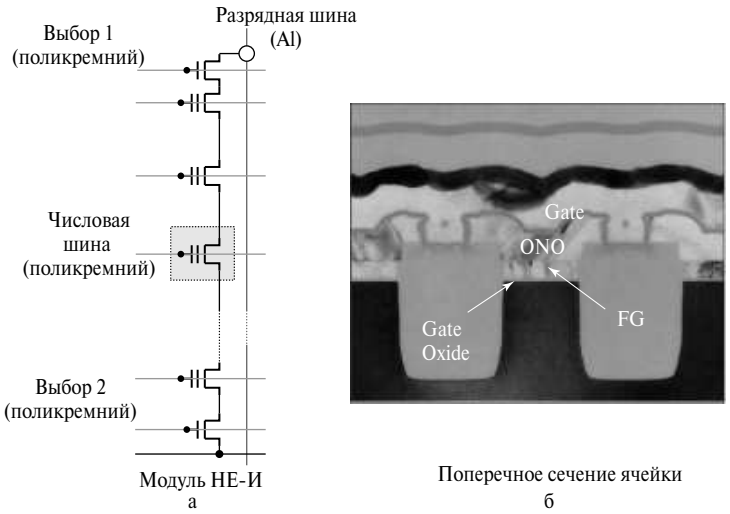


Рис. 12.25. Флэш-модуль на основе элементов НЕ-И [Nakamura02]. Во втором затворе использован сильный диэлектрик ("оксид-нитрид-оксид"), увеличивающий емкость плавающего затвора C_{FG}, а следовательно, влияние инжекции заряда на пороговое напряжение V_T

кого напряжения (больше 10 В) при программировании и стирании. Такие напряжения могут генерироваться внутри кристалла с использованием генераторов подкачки заряда (особенно если ток требуется небольшой, как при туннелировании Фаулера–Нордхейма). С другой стороны, инжекция горячих электронов подразумевает большой ток (более 0,5 мА), поэтому необходим внешний источник питания, что повышает стоимость и сложность системы. Исходя из этого, в настоящее время продолжается поиск энергонезависимых запоминающих устройств, не требующих высокого напряжения и внешних источников питания.

Новые тенденции в энергонезависимых ЗУ

Все рассмотренные выше энергонезависимые запоминающие устройства основаны на концепции транзистора с плавающим затвором. Существует также множество альтернативных подходов. Хотя бо́льшая их часть не получила развития и не попала на коммерческий рынок, некоторые структуры ячеек в настоящее время тщательно изучаются и могут в обозримом будущем оказать существенное влияние на развитие всей отрасли. На момент написания книги наиболее перспективными кажутся *ферроэлектрическое ОЗУ* (Ferroelectric RAM — FRAM) и *магниторезистивное ОЗУ* (Magnetoresistive RAM — MRAM). Тем не менее говорить о конце эры устройств с плавающим затвором еще равно. В то же время инновационные подходы (например, многоуровневые ячейки) создают возможности для еще более плотного энергонезависимого хранения информации. Ниже мы кратко рассмотрим основные концепции наиболее впечатляющих разработок.

Многоуровневые энергонезависимые запоминающие устройства

Все рассмотренные выше ячейки памяти являются бистабильными, т.е. они хранят значение 0 или 1. Если бы в одной ячейке можно было хранить больше двух состояний, емкость запоминающего устройства была бы гораздо выше. Основные проблемы при реализации данной идеи связаны с необходимостью существенного улучшения отношения сигнал/шум ячейки памяти. Если в ячейке памяти хранится k бит, то доступный сигнал ослабляется в $1/(2^k-1)$ раз. К счастью, ячейка флэш-памяти обладает внутренним усилением. Кроме того, используя возможности настройки напряжения V_T для каждой ячейки, можно вообще снять проблему плохого отношения сигнал/шум. В настоящее время уже существуют коммерческие разработки, позволяющие хранить по два бита в ячейке [Baur95]. Тем не менее при наблюдающейся тенденции к постоянному уменьшению напряжения питания многоуровневые запоминающие устройства вряд ли получат широкое распространение.

Ферроэлектрическое ОЗУ

Некоторое время с ферроэлектрическим ОЗУ (Ferroelectric RAM — FRAM) связывались весьма оптимистические ожидания. После ряда фальстартов стало казаться, что ячейки FRAM наконец-то готовы оправдать возлагаемые на них надежды. В отличие от программируемого транзистора, используемого в современных энергонезависимых ОЗУ, концепция FRAM основана на программируемом конденсаторе. Диэлектрический материал, используемый в ферроэлектрическом конденсаторе, относится к классу материалов, именуемых *перовскитами* (Perovskite crystals). Данные кристаллы поляризуются при поступлении на них электрического поля. Данная поляризация поддерживается и после снятия электрического поля. Для деполяризации диэлектрика необходимо приложить электрическое поле с противоположной напряженностью.

Ячейка памяти, созданная согласно этой концепции, очень похожа на однотранзисторную схему динамического ОЗУ, которая рассматривается в следующем разделе. В ходе записи ячейка конденсатора каким-то образом поляризуется. Для чтения содержимого к ячейке прикладывается электрическое поле. В зависимости от состояния конденсатора генерируется ток, который можно детектировать с помощью усилителя считывания.

Достоинством ферроэлектрического ОЗУ является очень высокая плотность хранения информации, из-за чего размеры ячеек получаются гораздо меньшими, чем у динамических ОЗУ, и при этом гарантируется энергонезависимость. Кроме того, число циклов перезаписи ферроэлектрических ОЗУ на порядок больше, чем у ЭСППЗУ и флэш-памяти. Благодаря низкому энергопотреблению данные запоминающие устройства весьма привлекательны в таких приложениях, как смарт-карты и радиочастотные метки. Основной проблемой данной технологии является создание больших запоминающих устройств.

Магниторезистивное ОЗУ

Магниторезистивные оперативные запоминающие устройства (Magnetoresistive Random Access Memory — MRAM) представляют метод хранения битов данных с использованием магнитных, а не электрических зарядов (как

Таблица 12.1. Сравнительная характеристика энергонезависимых ЗУ ([Itoh01]). $V_{DD} = 3,5$ или 5 В; $V_{PP} = 12$ или $12,5$ В

	Кол-во транзисторов в ячейке	Площадь ячейки (по сравнению с СППЗУ)	Механизм		Внешний источник питания		Циклы перезаписи
			Стирание	Запись	Запись	Чтение	
ПЗУ с фотошаблоном	1 (НЕ-И)	0,35–5	–	–	–	V_{DD}	0
СППЗУ	1	1	УФ-облучение	Горячие электроны	V_{PP}	V_{DD}	~ 100
ЭСППЗУ	2	3–5	Туннелирование	Туннелирование	$V_{PP}(int)$	V_{DD}	10^4–10^5
Флэш-память	1	1–2	Туннелирование	Горячие электроны	V_{PP}	V_{DD}	10^4–10^5
			Туннелирование	Туннелирование	$V_{PP}(int)$	V_{DD}	10^4–10^5

в динамических ОЗУ). Металл называется *магниторезистивным*, если при помещении в магнитное поле он немного меняет электрическое сопротивление. Концепция, на основе которой построено устройство, очень похожа на ту, которая использовалась в памяти на магнитных сердечниках в мэйнфреймах на заре развития электроники. Основные отличия заключаются в масштабе (в миллион раз меньше) и используемых материалах. Магниторезистивные ОЗУ возникли на стыке разработок двух научных направлений: (1) спиновой электроники — науки, давшей миру гигантские магниторезистивные головки, используемые в накопителях на магнитных дисках; (2) магнитного сопротивления туннелированию, которое по идее должно стать основой будущих элементов MRAM. В 2000 году разработчики из IBM продемонстрировали MRAM-память емкостью 1 Кбит. Каждая ячейка состоит из магнитного туннельного перехода и полевого транзистора, что при 0,25-микронном КМОП-процессе дает ячейку общей площадью 3 мкм^2 [Scheuerlein00]. Время чтения и записи составляет до 10 нс. Предсказывать будущее данного устройства пока что слишком рано, но, похоже, сфера энергонезависимых запоминающих устройств (да и полупроводниковых запоминающих устройств в целом) будет сильно меняться под воздействием новых материалов и подходов.

Резюме

Чтобы лучше представить общую картину сферы ЗУ, можно обратиться к цифрам. В табл. 12.1 показаны данные по энергонезависимым запоминающим устройствам на момент написания книги. Из таблицы видно, что гибкость структуры ЭСППЗУ возможна за счет плотности и быстродействия. Устройства СППЗУ и флэш-СППЗУ имеют сравнимые характеристики с точки зрения плотности и скорости. Быстрый захват флэш-памятью позиций на рынке объясняется ее универсальностью.

Большое количество соображений, приведенных касательно ПЗУ, справедливы и для энергонезависимых ОЗУ. Кроме того, для структур СППЗУ (ЭСППЗУ) характерна сложность цепей программирования и стирания. Помните, что все предложенные структуры требуют наличия высоковольтных сигналов (12–20 В) в разрядной и числовой шинах в процессе программирования, тогда как в режиме чтения напряжение в шинах должно составлять стандартные 3 или 5 В. Генерирование и распределение данных сигналов требуют использования интересных решений, которые, к сожалению, в данной книге не рассматриваются.

12.2.3. Оперативные запоминающие устройства (ОЗУ)

Создание ячейки памяти с приблизительно одинаковой скоростью чтения и записи требует использования более сложных структур. В то время как содержимое ПЗУ и СППЗУ вводится в топологию ячейки или программируется на уровне характеристик устройства, хранение информации в оперативном запоминающем устройстве основано либо на положительной обратной связи, либо на емкостном заряде (как в подходах, представленных в главе 6). Данные схемы идеально подошли бы в качестве ячеек памяти с возможностью чтения и записи, но, как правило, они занимают слишком много места. В данном разделе представим несколько упрощений, которые позволят уменьшить площадь либо за счет быстродействия, либо за счет электрической надежности. В зависимости от концепции, используемой для хранения информации, данные ячейки называются статическими или динамическими ОЗУ.

Статическое оперативное запоминающее устройство

Общая схема ячейки *статического ОЗУ* (Static Random-Access Memory — SRAM) показана на рис. 12.26. Визуально она очень похожа на статическую SR-защелку, показанную на рис. 7.21. Для хранения одного бита в данной версии статического ОЗУ требуется шесть транзисторов. Доступ к ячейке открывает числовая шина, которая заменяет собой шину тактовых импульсов и управляет двумя проходными транзисторами, M_5 и M_6, совместно используемыми в ходе операций чтения и записи. В отличие от ячейки ПЗУ требуется две шины, по которым передается записанный сигнал и обратный к нему. Хотя наличие двух полярностей не обязательно, оно, как будет показано ниже, позволяет повысить запас помехоустойчивости при чтении и записи.

Задача 12.5. КМОП-ячейка статического ОЗУ

Потребляет ли ячейка статического ОЗУ, изображенная на рис. 12.26, мощность в режиме хранения? Объясните свой ответ. Изобразите эквивалентную псевдо-n-МОП-реализацию. Потребляет ли она мощность в режиме хранения?

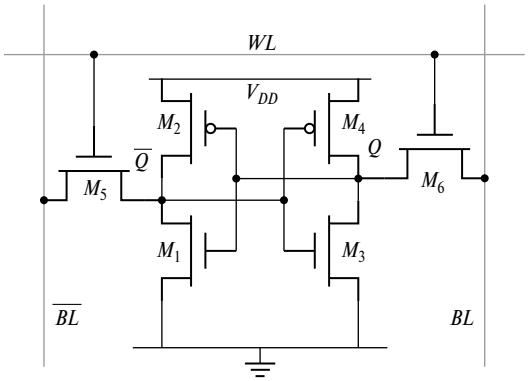


Рис. 12.26. Шеститранзисторная КМОП-ячейка статического ОЗУ

Работа ячейки статического ОЗУ

Чтобы добиться высокой емкости ЗУ, ячейка статического ОЗУ должна иметь минимальный размер. Однако надежная работа ячейки возможна только при выполнении некоторых условий относительно размеров.

Чтобы понять работу ячейки памяти, рассмотрим последовательно операции чтения и записи и выведем условия на размеры транзисторов.

Пример 12.8. КМОП-ячейка статического ОЗУ. Операция чтения

Предположим, что в транзисторе Q хранится значение 1. Кроме того, будем считать, что перед операцией чтения обе разрядных шины предварительно заряжены до значения 2,5 В. Цикл чтения начинается с активизации после первоначальной задержки числовой шины транзисторов M_5 и M_6. В ходе операции чтения значения, записанные в Q и $\overline{Q}$, передаются в разрядную шину; для этого BL сохраняет значение, полученное при предварительной зарядке, а $\overline{BL}$ разряжается через M_1–M_5. Чтобы избежать случайной записи в ячейку значения 1, необходимо аккуратное масштабирование транзисторов.

Сказанное иллюстрируется на рис. 12.27. Рассмотрим сторону $\overline{BL}$ ячейки. Для больших запоминающих устройств емкость разрядной шины составляет порядка нескольких пикофарад. Следовательно, после инициации операции чтения ($WL \rightarrow 1$) значение $\overline{BL}$ сохраняет значение V_{DD}, полученное при предварительной зарядке. Последовательное соединение двух n-МОП-транзисторов разряжает $\overline{BL}$ до уровня земли. Для небольших ячеек данные транзисторы должны иметь как можно меньшие размеры, поскольку это приведет к очень медленной разрядке большой емкости разрядной шины. По мере увеличения разницы между BL и $\overline{BL}$ активизируется усилитель считывания, который ускоряет процесс чтения.

Изначально при нарастании WL напряжение на промежуточном узле между двумя n-МОП-транзисторами ($\bar{Q}$) доводится до значения, соответствующего состоянию предварительной зарядки $\overline{BL}$. Данное увеличение напряжения $\bar{Q}$ должно быть достаточно низким, чтобы не породить большого тока через инвертор M_3–M_4 (что в худшем случае может перевести ячей-

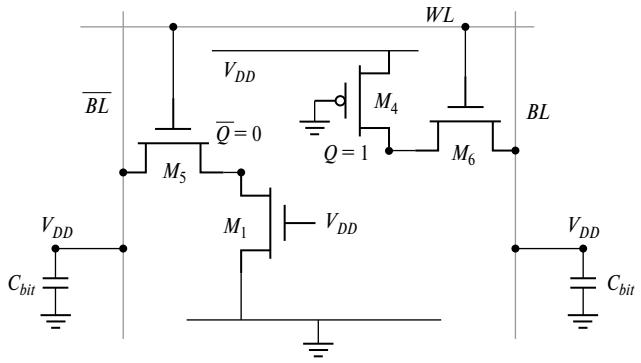


Рис. 12.27. Упрощенная модель КМОП-ячейки статического ОЗУ в процессе чтения ($Q = 1$, $V_{precharge} = V_{DD}$)

ку в противоположное состояние). Чтобы этого не случалось, важно, чтобы сопротивление транзистора M_5 было большем, чем у M_1.

Решив уравнение тока при максимально допустимом значении пульсации напряжения ΔV, мы можем найти граничные условия на размеры устройств. Для простоты будем игнорировать влияние подложки на транзистор M_5 и запишем:

$$k_{n,M5}\left((V_{DD} - \Delta - V_{Tn})V_{DSATn} - \frac{V_{DSATn}^2}{2}\right) = \tag{12.2}$$
$$= k_{n,M1}\left((V_{DD} - V_{Tn})\Delta - \frac{\Delta^2}{2}\right).$$

Данное уравнение можно упростить до следующего вида:

$$\Delta = \tag{12.3}$$
$$= \frac{V_{DSATn} + CR(V_{DD} - V_{Tn}) - \sqrt{V_{DSATn}^2(1 + CR) + CR^2(V_{DD} - V_{Tn})^2}}{CR},$$

где CR называется *характеристическим размером ячейки* и определяется следующим образом:

$$CR = \frac{W_1/L_1}{W_5/L_5}. \tag{12.4}$$

На рис. 12.28 показано, на сколько увеличивается напряжение (ΔV) в зависимости от CR для 0,25-микронной технологии. Чтобы напряжение узла не поднималось выше порога транзистора (порядка 0,4 В), характеристический размер ячейки должен быть выше 1,2. Для больших матриц памяти размер ячейки желательно делать как можно меньше, при этом поддерживая устойчивость операции чтения. Если транзистор M_1 имеет минимальный размер,

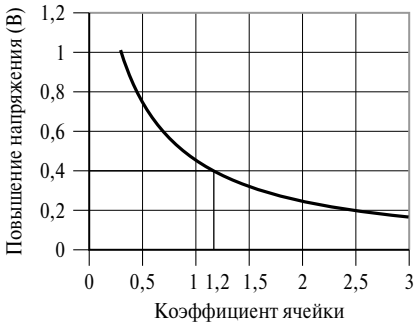


Рис. 12.28. Повышение напряжение внутри ячейки в процессе чтения в зависимости от характеристического размера ячейки (отношения M_1/M_5). При $CR > 1,2$ напряжение внутри ячейки не поднимается выше порога

то проходной транзистор доступа M_5 нужно сделать слабее за счет увеличения его длины. Это нежелательно, поскольку увеличивает нагрузку на разрядную шину. Оптимальное решение в такой ситуации — минимизировать размер проходного транзистора и увеличить ширину n-МОП-транзистора M_5, удовлетворив таким образом условия устойчивости. Отметим, что это немного увеличивает минимальный размер ячейки. Чтобы гарантировать устойчивую работу ячейки в любых условиях, разработчик должен аккуратно смоделировать все возможные ситуации [Preston01].

Выше мы проанализировали наихудшую из возможных ситуаций. Вторая разрядная шина BL фиксирует Q на значении V_{DD}, в результате чего случайное переключение пары инверторов с перекрестными обратными связями становится менее вероятным (это является одним из основных преимуществ архитектуры двойной разрядной шины).

Ошибочное переключение можно предотвратить не только с помощью подбора размеров транзисторов ячейки, но и посредством предварительной зарядки разрядных шин до некоторого значения, например $V_{DD}/2$. Это не позволяет Q достигнуть порога переключения связующего инвертора. Предварительная зарядка до значения, равного половине диапазона изменения напряжения, также положительно сказывается на быстродействии, поскольку ограничивает размах напряжения в разрядных шинах.

Пример 12.9. КМОП-ячейка статического ОЗУ. Операция записи

В данном примере мы выведем условия на устройство, необходимые для обеспечения правильного выполнения операции записи. Предположим, что в ячейке хранится значение 1 (или $Q = 1$). Для записи в ячейку значения 0 необходимо сделать сигнал $\overline{BL}$ равным 1, а BL — равным 0, что идентично подаче импульса сброса на SR-защелку. В результате при правильном подборе размеров устройств триггер меняет состояние.

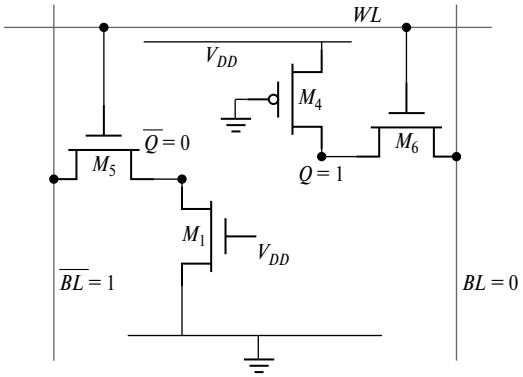


Рис. 12.29. Упрощенная модель КМОП-ячейки статического ОЗУ в процессе записи ($Q = 1$)

При инициации операции записи схему ячейки статического ОЗУ можно упростить до модели, показанной на рис. 12.29. Разумно предположить, что до начала переключения на затворы транзисторов M_1 и M_4 подается напряжение V_{DD} и GND соответственно. Хотя данное условие нарушается, как только триггер начинает менять состояние, для анализа вручную данной модели вполне достаточно.

Обратите внимание на то, что напряжение стороны $\bar{Q}$ ячейки нельзя поднять достаточно высоко, чтобы обеспечить запись значения 1. Из-за условия, налагаемого на размеры из соображений устойчивости, данное напряжение на может подняться выше 0,4 В. Следовательно, новое значение ячейки нужно записывать через транзистор M_6.

Надежная запись ячейки будет гарантирована, если мы достаточно понизим напряжение на узле Q, т.е. опустим его ниже порогового значения транзистора M_1[5]. Чтобы вывести условия, при которых это возможно, запишем уравнение постоянного тока для требуемой пороговой точки:

$$k_{n,M6}\left((V_{DD}-V_{Tn})V_Q-\frac{V_Q^2}{2}\right)= \tag{12.5}$$
$$=k_{p,M4}\left((V_{DD}-|V_{Tp}|)V_{DSATp}-\frac{V_{DSATp}^2}{2}\right).$$

Выражая отсюда V_Q, получим следующее:

$$V_Q=V_{DD}-V_{Tn}- \tag{12.6}$$
$$-\sqrt{(V_{DD}-V_{Tn})^2-2\frac{\mu_p}{\mu_n}PR\left((V_{DD}-|V_{Tp}|)V_{DSATp}-\frac{V_{DSATp}^2}{2}\right)},$$

[5] Напряжение на узле Q достаточно опустить ниже порога переключения инвертора, сформированного транзисторами M_1 и M_2, что гарантирует начало переключения. С точки зрения устойчивости к шуму безопаснее требовать, чтобы напряжение на узле Q понижалось ниже порога M_1.

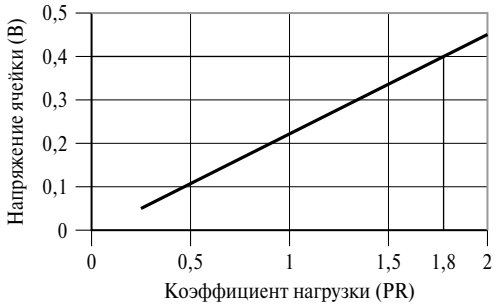


Рис. 12.30. Напряжение, записываемое в ячейку, в зависимости от коэффициента нагрузки ячейки (отношение размеров p-МОП-транзистора и транзистора доступа). Значение PR должно быть меньше 1,8

где *коэффициент нагрузки ячейки* (pull-up ratio — PR) определяется как отношение размеров p-МОП- и n-МОП-транзисторов:

$$PR = \frac{W_4/L_4}{W_6/L_6}. \tag{12.7}$$

На рис. 12.30 показана зависимость V_Q от коэффициента PR для 0,25-микронного процесса. Чем ниже PR, тем меньше значение V_Q. Если нам требуется опустить напряжение на узле ниже V_{Tn}, то значение PR должно быть ниже 1,8.

Данное условие часто удовлетворяется за счет использования в качестве p-МОП-транзистора M_4 и n-МОП-транзистора доступа M_6 устройств минимального размера. Тем не менее разработчик должен убедиться, что условие возможности записи действительно удовлетворяется для всего диапазона рабочих условий. Наихудший случай представлен сильным p-МОП-устройством, слабым n-МОП-устройством и запоминающим устройством, работающим при напряжении питания, превышающем номинальное [Preston01].

Напомним, что изначально мы предположили, что транзисторы M_1 и M_2 не участвуют в процессе записи. На практике это не совсем справедливо. Как только одна сторона ячейки начинает переключение, за ней вскоре последует вторая сторона, образуя положительную обратную связь.

Быстродействие ячейки статического ОЗУ

При анализе переходного режима ячейки статического ОЗУ можно обнаружить, что критической с точки зрения времени является операция чтения. Она требует зарядки (разрядки) большой емкости разрядной шины через два маленьких транзистора выбранной ячейки. Например, в случае, приведенном на рис. 12.27, $C_{\overline{BL}}$ необходимо разрядить через последовательное соединение транзисторов M_5 и M_1. Время записи определяется в основном задержкой распространения пары инверторов с перекрестными обратными связями, по-

Таблица 12.2. Сравнение КМОП-ячеек статического ОЗУ, используемых в памяти емкостью в 1 Мбит (согласно [Takada91])

	КМОП	Активная нагрузка	Тонкопленочная ячейка
Число транзисторов	6	4	4 (+2 TFT)
Размер ячейки	$58,2$ мкм^2 ($0,7$-микронная технология)	$40,8$ мкм^2 ($0,7$-микронная технология)	$41,1$ мкм^2 ($0,8$-микронная технология)
Ток холостого хода (на одну ячейку)	10^{-15} А	10^{-12} А	10^{-13} А

скольку задающие генераторы, доводящие BL и $\overline{BL}$ до требуемых значений, можно сделать достаточно большими. Для ускорения процесса чтения в статических ОЗУ применяются усилители считывания. По мере увеличения разности напряжений между BL и $\overline{BL}$ активизируется усилитель считывания, который быстро разряжает одну из разрядных шин.

Улучшенные МОП-ячейки статического ОЗУ

Шеститранзисторная ячейка статического ОЗУ хотя и является простой и надежной, занимает слишком большую площадь. Помимо собственно транзисторов, ее реализация требует подводки сигналов к двум разрядным шинам, числовой шине и обеим шинам питания и организации соответствующих соединений. Существенно увеличивает требуемую площадь и введение двух p-МОП-транзисторов в n-карман. Возможная топология такой ячейки показана на рис. 12.31. Ее размеры определяются преимущественно проводкой и межслойными контактами (одиннадцать с половиной штук; верхние и нижние учитываются с коэффициентом $1/2$, так как они используются совместно двумя соседними ячейками), а также требованиями к минимальному расстоянию, налагаемыми на карман.

Таким образом, разработчики больших матриц памяти должны использовать другие структуры ячеек, не только основанные на проверенных транзисторных топологиях, но и учитывающие наличие специальных устройств и доступность более сложной технологии.

Рассмотрим приведенную на рис. 12.32 схему ячейки, именуемой ячейкой статического ОЗУ с *активной нагрузкой* (также известна как четырехтранзисторная ячейка). Характерной особенностью данной ячейки является замена пары КМОП-инверторов с перекрестными обратными связями парой n-МОП-инверторов с активной нагрузкой. Вместо p-МОП-транзисторов используются резисторы, что существенно упрощает проводные соединения. Это, в свою очередь, уменьшает размер ячейки статического ОЗУ приблизительно на треть, как показано в табл. 12.2 на примере запоминающего устройства емкостью в 1 Мбит.

Анализируя возможность записи в ячейку статического ОЗУ, находим, что существует только одно ограничение на нижний предел нагрузочного сопротивления. Поскольку разрядные шины предварительно заряжаются извне, ячейка не задействована в процессе перевода напряжения на высокий уровень и никаких ограничений на наличие очень больших нагрузок не су-

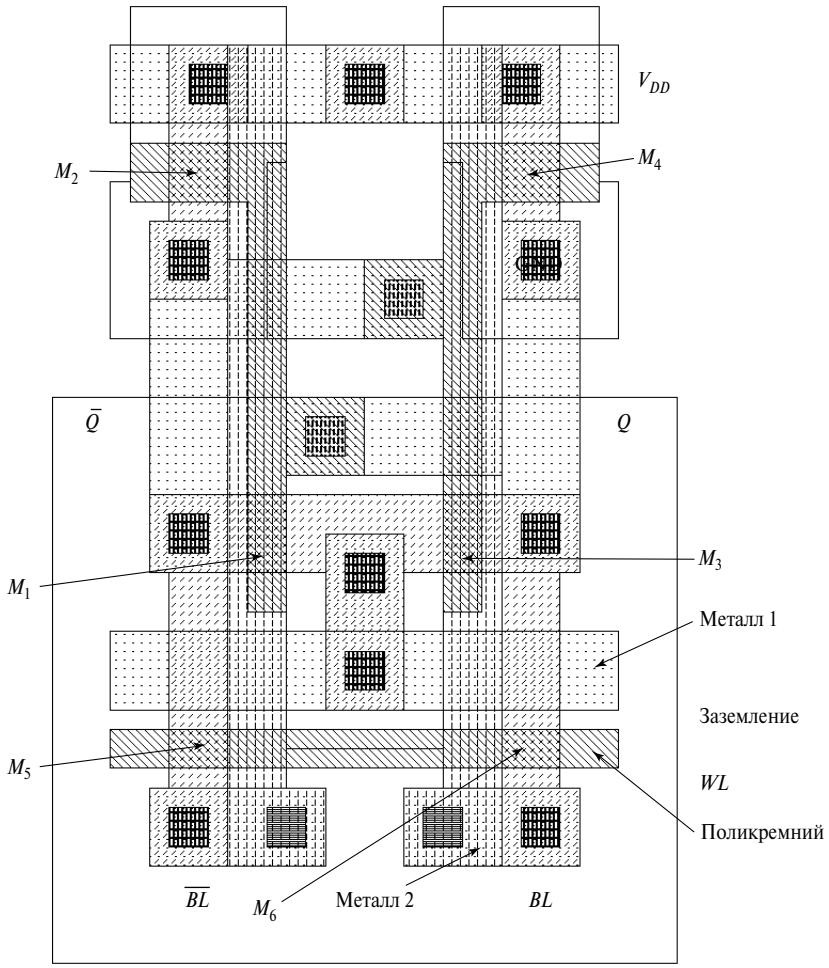


Рис. 12.31. Топология шеститранзисторной КМОП-ячейки памяти статического ОЗУ

ществует (в отличие от обычных логических схем). Следовательно, при улучшенной реализации статического ОЗУ применяются специальные резисторы, которые делаются насколько возможно большими, чтобы минимизировать статическое потребление мощности.

Основным приоритетом при разработке ячеек статических ОЗУ является минимизация статического рассеяния мощности на одну ячейку. Рассмотрим запоминающее устройство емкостью в 1 Мбит, которое работает при напряжении питания 2,5 В и использует в качестве инверторной нагрузки резистор 10 кОм. Поскольку каждая ячейка имеет ток покоя 0,25 мА, общее рассеяние в холостом режиме составляет 250 Вт! Следовательно, единственное решение этой проблемы — сделать как можно большим сопротивление нагрузки. Компактный транзистор с очень большим сопротивлением можно произвести из

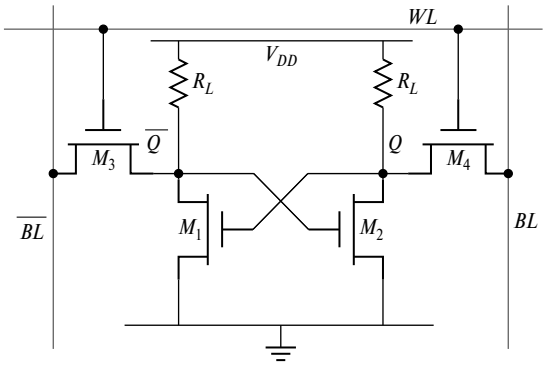


Рис. 12.32. Ячейка статического ОЗУ с активной нагрузкой

нелегированного поликремния, удельное сопротивление которого составляет несколько тераом на мм^2 (тера — 10^{12}!). Единственное дополнительное условие на нагрузочные транзисторы — поддерживать состояние ячейки, т.е. компенсировать токи утечки, обычно составляющие порядка 10^{-15} А на ячейку [Takada91]. В современных условиях малая утечка достигается за счет использования устройств с более высокими порогами. Для реализации этого ток через резистор должен быть по крайней мере на два порядка больше, или $I_{load} > 10^{-13}$ А. Таким образом, получаем условие на верхний предел нагрузочного сопротивления.

Поняв, что нагрузочные устройства нужны только для компенсации потери заряда, получаем обновленную версию шеститранзисторной ячейки памяти, изображенной на рис. 12.26. Вместо использования традиционных дорогих p-МОП-устройств нагрузочные транзисторы с помощью тонкопленочной технологии реализуются как паразитные устройства, располагающиеся поверх структуры ячейки. Эти p-МОП-*тонкопленочные транзисторы* (Thin-Film Transistor — TFT) обладают более удачным набором свойств, чем обычные устройства, и характеризуются током, приблизительно равным 10^{-8} А и 10^{-13} во включенном и выключенном режимах соответственно для напряжения между затвором и истоком 5 В [Sasaki90, Ootani90]. Комплементарная природа ячейки повышает ее надежность и делает менее чувствительной к утечке и нерегулярным ошибкам, при этом имея более низкий ток в режиме хранения, чем у ячейки с активной нагрузкой.

Обратите внимание на то, что поликремний, обладающий высоким сопротивлением, и тонкопленочные транзисторы являются дополнительными инструментами, недоступными при использовании стандартных процессов. Следовательно, внедренные ячейки статического ОЗУ (например, используемые в кэше микропроцессоров) реализуются по стандартной технологии и содержат шесть транзисторов (см. рис. 12.26).

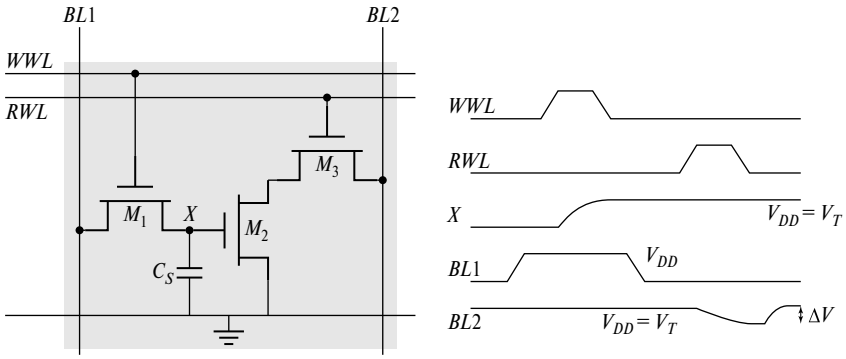


Рис. 12.33. Трехтранзисторная динамическая ячейка памяти и форма сигналов в процессе чтения и записи

Динамические оперативные запоминающие устройства

При обсуждении ячейки статического ОЗУ с активной нагрузкой мы отмечали, что единственная функция нагрузочных резисторов — восстановить заряд, теряемый вследствие утечки. Эти нагрузки можно вообще убрать, тогда потерю заряда придется компенсировать путем периодической перезаписи содержимого ячеек. Данная операция *обновления*, состоящая из двух этапов, чтения содержимого ячейки и операции обновления, должны выполняться достаточно часто, чтобы содержимое ячеек не успевало разрушаться из-за утечки. Как правило, обновление должно выполняться каждые 1–4 мс. Для больших запоминающих устройств снижение сложности ячеек с лихвой окупает расходы на добавление цепей, отвечающих за обновление. Получаемые согласно описанной концепции запоминающие устройства называются *динамическими*, поскольку в их основе лежит хранение заряда на конденсаторе.

Трехтранзисторная динамическая ячейка памяти

Первая разновидность динамической ячейки памяти получается путем удаления нагрузочных резисторов со схемы, изображенной на рис. 12.33. Четырехтранзисторную ячейку можно дополнительно упростить, отметив, что в ячейке хранятся как информационные биты, так и комплементарные к ним; следовательно, информация избыточна. Если убрать еще одно устройство (например, транзистор M_1), избыточность исчезнет и мы получим трехтранзисторную ячейку, приведенную на рис. 12.33 [Regitz70]. Такие ячейки стали ядром первых популярных запоминающих МОП-устройств, например памяти емкостью 1 Кбит производства Intel [Hoff70]. В более современных запоминающих устройствах применяются ячейки, более эффективно использующие площадь, но тем не менее данная концепция, благодаря простоте структуры и удобству использования, все еще применяется во многих устройствах, внедряемых в специализированные интегральные схемы.

Для записи информации в ячейку необходимые данные помещаются в $BL1$ и на высокий уровень переводится сигнал *шины записи чисел* (Write-Word Line — WWL). После того как сигнал в шине WWL опустится на низкий уровень, данные будут храниться в виде заряда на емкости C_S. При чтении содержимого ячейки на высокий уровень переводится сигнал *шины чте-*

ния чисел (Read-Word Line — RWL). Транзистор-хранилище M_2 находится в открытом или запертом состоянии в зависимости от записанного значения. Разрядная шина $BL2$ либо получает значение V_{DD} с помощью нагрузки (например, заземленного p-МОП- или насыщенного n-МОП-транзистора), либо предварительно заряжается до V_{DD} или $V_{DD} - V_T$. Первый подход требует аккуратного выбора размеров транзисторов и характеризуется статическим потреблением мощности. Следовательно, в общем случае более предпочтительным является использование предварительной зарядки. Последовательное соединение транзисторов M_2 и M_3 позволяет перевести сигнал $BL2$ на низкий уровень при хранении значения 1. В противоположном случае сигнал $BL2$ остается на высоком уровне. Обратите внимание на то, что ячейка является инвертирующей, т.е. в разрядной шине детектируется значение, обратное к записанному в ячейке. Наиболее распространенным способом обновления ячейки является чтение записанных данных, подача инвертированного значения в $BL1$ и перевод сигнала WWL на высокий уровень.

Описанная ячейка существенно проще статической. Это иллюстрируется на рис. 12.34. Общая площадь ячейки равна $576\lambda^2$ по сравнению с $1092\lambda^2$ элемента статического ОЗУ, изображенного на рис. 12.31. При этом мы не учитывали потенциальное уменьшение площади, которое получается за счет совместного использования элементов соседними ячейками. В основном уменьшение площади происходит за счет удаления контактов и устройств.

Дальнейшее упрощение структуры ячейки возможно за счет более сложной работы схемы. Например, разрядные шины $BL1$ и $BL2$ можно объединить в один провод. Циклы чтения и записи можно реализовать так же, как ранее. Цикл обновления (чтения–детектирования–записи) нужно существенно видоизменить, поскольку данные, считываемые из ячейки, являются комплементарными по отношению к записанным данным. Для этого требуется, чтобы разрядная шина в течение одного цикла принимала оба значения. Кроме того, можно объединить шины RWL и WWL (это не сильно скажется на работе ячейки). Операция чтения автоматически выполняется путем обновления содержимого ячейки. Чтобы предотвратить запись ячейки до чтения реального значения в процессе обновления, необходимо очень тщательно контролировать напряжение числовой шины.

Наконец, стоит упомянуть следующие интересные свойства трехтранзисторной ячейки.

1. В отличие от ячейки статического ОЗУ никаких ограничений на *размеры устройств* не существует (вообще, это характерно для всех динамических схем). Выбор размеров устройств основывается на соображениях быстродействия и надежности. Обратите внимание на то, что данное утверждение *не справедливо* при реализации подхода со статической нагрузкой разрядной шины.
2. В отличие от других ячеек динамического ОЗУ чтение содержимого трехтранзисторного элемента является *недеструктивным*, т.е. данные, хранимые в ячейке, в процессе чтения не модифицируются.

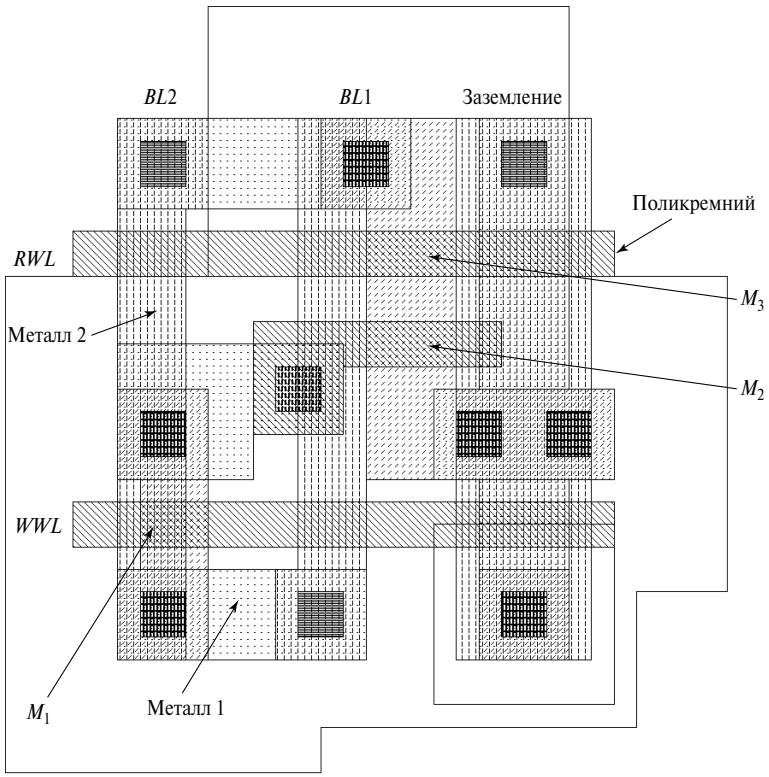


Рис. 12.34. Пример топологии трехтранзисторной динамической ячейки памяти

3. Никаких специальных технологических операций не требуется. Емкость запоминающего устройства представляет собой емкость затвора считывающего устройства. Этот момент отличает рассматриваемое устройство от других ячеек динамических ОЗУ, которые будут рассмотрены ниже, и делает трехтранзисторную ячейку привлекательной с точки зрения внедренных приложений.
4. Значение, записанное на запоминающем узле X при записи 1, равно $V_{WWL} - V_{Tn}$. Такое снижение порога уменьшает ток, проходящий через M_2 в ходе операции чтения, и увеличивает время доступа для чтения. Чтобы избежать этого, некоторые разработчики усиливают напряжение в числовой шине, или, другими словами, повышают V_{WWL} до значения, превышающего V_{DD}.

Однотранзисторная динамическая ячейка памяти

Еще одна возможность для существенного снижения сложности ячейки — пожертвовать некоторыми ее свойствами. Получающаяся в результате структура, именуемая *однотранзисторной ячейкой динамического ОЗУ*, несомненно, является наиболее распространенным устройством в коммерческих моду-

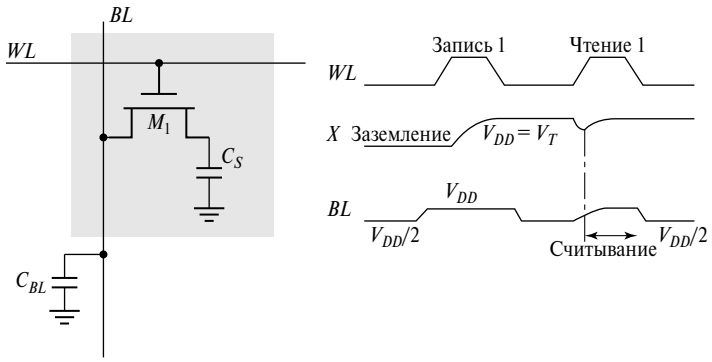


Рис. 12.35. Однотранзисторная ячейка динамического ОЗУ и соответствующие формы сигналов в процессе чтения и записи

лях памяти[6]. Схема этой ячейки показана на рис. 12.35 [Dennard68]. Принцип работы данного устройства крайне прост. В процессе записи данные подаются в разрядную шину BL, а сигнал в числовой шине WL переводится на высокий уровень. В зависимости от данных емкость ячейки заряжается или разряжается. Перед выполнением операции чтения разрядная шина предварительно заряжается до напряжения V_{PRE}. После перехода сигнала числовой шины на высокий уровень происходит перераспределение заряда между разрядной шиной и запоминающей емкостью. Из-за этого происходит изменение напряжения в разрядной шине, направление которого определяет значение записанного бита данных. Величина изменения напряжения находится следующим образом:

$$\Delta \ = V_{BL} - V_{PRE} = (V_{BIT} - V_{PRE})\frac{C_S}{C_S + C_{BL}}, \tag{12.8}$$

где C_{BL} — емкость разрядной шины; V_{BL} — потенциал разрядной шины после перераспределения заряда; V_{BIT} — исходное напряжение, поданное на емкость ячейки C_S. Поскольку емкость ячейки обычно на один-два порядка меньше емкости разрядной шины, данное изменение напряжения очень мало, как правило, порядка 250 мВ для запоминающих устройств, существовавших на момент написания книги [Itoh90]. Отношение $C_S/(C_S{+}C_{BL})$ называется *коэффициентом переноса заряда* (charge-transfer ratio) и составляет от 1 до 10%.

Если ячейка должна иметь требуемые функциональные возможности, необходимо усиление ΔV до полного размаха напряжения. Таким образом, получаем первое серьезное отличие однотранзисторной ячейки от трехтранзисторной (а также других ячеек динамического ОЗУ).

[6] Можно разработать и динамическую ячейку, содержащую всего два транзистора. Однако поскольку она не дает никаких особых преимуществ по сравнению с одно- и трехтранзисторными устойствами, то на практике используется редко.

1. Однотранзисторная ячейка динамического ОЗУ требует *наличия усилителя считывания* для каждой разрядной шины. Это требование вызвано считыванием на основе перераспределения заряда. Операция чтения всех ячеек, рассмотренных ранее, основывалась на переключении тока. Усилитель считывания требуется не для обеспечения работы, а только для ускорения считывания информации. Кроме того, стоит отметить, что ячейки динамического ОЗУ являются *несимметричными* в отличие от ячеек статического ОЗУ, которые выдают и бит данных, и комплементарное значение. Это несколько усложняет проектирование усилителя считывания (подробно об этом — ниже, в разделе, посвященном периферийным устройствам).

Пример 12.10. Считывание информации в однотранзисторной ячейке динамического ОЗУ

Предположим, что емкость разрядной шины равна 1 пФ, а напряжение предварительной зарядки разрядной шины составляет 1,25 В. Подача на емкость ячейки C_S (50 фФ) напряжения 1,9 В и 0 В обозначает значение 1 и 0 соответственно. Таким характеристикам соответствует эффективность переноса заряда 4,8% и следующий размах напряжения в разрядной шине в процессе чтения:

$$\Delta\ V(0) = -1,25\ \text{В} \times \frac{50\ \text{фФ}}{50\ \text{фФ} + 1\ \text{фФ}} = -60\ \text{мВ}$$
$$\Delta\ V(1) = 31\ \text{мВ}$$

Кроме того, следует отметить такие важные отличия.

2. Считывание информации из однотранзисторной ячейки динамического ОЗУ является *деструктивным*. Это означает, что в процессе чтения меняется заряд, запасенный в ячейке. После успешной операции чтения необходимо восстановить исходное значение. Следовательно, в однотранзисторном ОЗУ операции чтения и обновления связаны между собой. Обычно в процессе чтения выход усилителя считывания накладывается на сигнал разрядной шины. Если удерживать сигнал WL на достаточно высоком уровне, то заряд ячейки в это время будет восстановлен. Сказанное иллюстрируется на рис. 12.36, где показана зависимость напряжения типичной разрядной шины от времени в процессе чтения.
3. В отличие от трехтранзисторной ячейки, принцип действия которой основан на хранении заряда на емкости затвора, однотранзисторная ячейка требует наличия дополнительной емкости, которую необходимо явно включить в проект. Из соображений надежности коэффициент переноса заряда делается большим при минимальной емкости порядка 30 фФ. “Втискивание” такой большой емкости в маленькую площадь является,

Рис. 12.36. Напряжение разрядной шины в процессе чтения (для битов 1 и 0)

пожалуй, одной из наиболее сложных задач в сфере проектирования динамических ОЗУ. Некоторые способы ее решения кратко описаны в следующем разделе.

4. Обратите внимание на то, что при записи в ячейку значения 1 пороговое напряжение теряется, что уменьшает доступный заряд. Эту потерю заряда можно компенсировать с помощью самозагрузки числовых шин значениями, превышающими V_{DD}. В современных запоминающих устройствах обычно так и поступают.

Первый подход к проектированию однотранзисторной ячейки динамического ОЗУ показан на рис. 12.37. Основным преимуществом данной структуры является возможность реализации по общей КМОП-технологии. Запоминающий узел в этой ячейке состоит из емкости затвора, расположенной между поликремниевой пластиной и обращенным слоем, внедренным в подложку с помощью положительного напряжения смещения, приложенного к поликремниевой пластине. При записи в ячейку значения 0 потенциальная яма запоминающего узла заполняется электронами и конденсатор заряжается. Если в ячейку записывается значение 1 (путем подачи в разрядную шину высокого напряжения), то электроны удаляются с индуцированного обращенного слоя и поверхность обедняется. Напряжение, поступающее на ячейку, уменьшается. Обратите внимание на то, что в данном случае ситуация обратна к описанной выше: конденсатор заряжается при записи значения 0 и разряжается при записи значения 1.

Реализация ячеек с большей емкостью требует модификации процесса производства. Первое, что требуется, — добавить второй поликремниевый слой, который действует как вторая обкладка конденсатора (первую формирует первый поликремниевый слой). В погоне за большей емкостью запоминающих устройств разработчики начали интенсивно использовать трехмерные структуры, где емкость запоминающего устройства реализуется вертикально на подложке или поверх транзистора доступа [Lu89]. Некоторые распространенные ячейки, реализованные по такому принципу, показаны на рис. 12.38. На рис. 12.38, *а* представлено поперечное сечение *ячейки с канальным конденсатором*. В данной структуре вертикальная канавка глубиной до 5 мкм

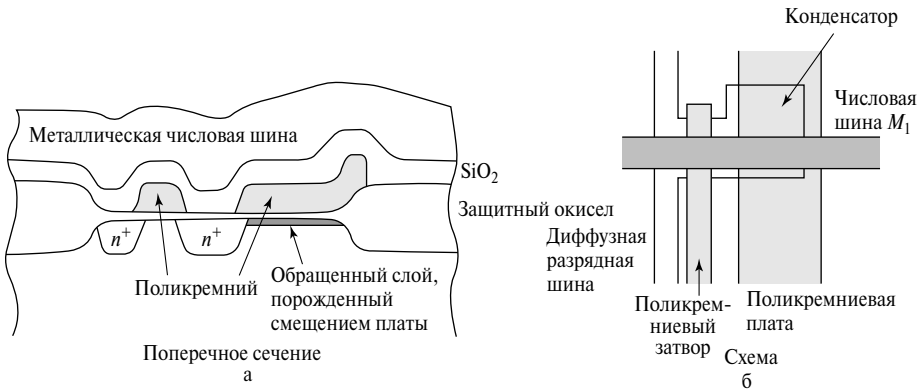


Рис. 12.37. Однотранзисторная ячейка динамического ОЗУ, реализованная с использованием в качестве запоминающего узла поликремниево-диффузионной емкости (согласно [Dillinger88]). Контакт между числовой шиной и поликремниевым затвором проходит через соседнюю ячейку

вытравливается в подложке. Боковые стенки и низ канавки используются под емкостный электрод, в результате получается, что большая поверхность пластины занимает небольшую площадь кристалла. На рис. 12.38, *б* показано поперечное сечение *ячейки с многоярусным конденсатором* (Stacked-Capacitor Cell — STC), где емкость накладывается поверх транзисторов доступа и разрядных шин. Для реализации подобных конденсаторов задействуется до четырех поликремниевых слоев, что позволяет уменьшить эффективную площадь ячейки. При использовании описанных подходов площадь ячейки динамического ОЗУ емкостью 64 Мбит составляет между 1,5 и 2,0 мкм^2, т.е. при 0,4-микронной технологии емкость запоминающего устройства составляет от 20 до 30 фФ! Очевидно, такое уменьшение емкости не может быть бесплатным: производство описанных устройств является весьма сложным и дорогостоящим. Тем не менее возможность производства полупроводниковых запоминающих устройств емкостью более 256 Мбит или 1 Гбит определяется не столько физическими ограничениями или проблемами реализации, сколько экономическими соображениями.

12.2.4. Ассоциативные запоминающие устройства

Концепция ассоциативной памяти уже обсуждалась в этой главе. Ассоциативная память (Contents-Addressable Memory — CAM) представляет собой особое запоминающее устройство, которое не только хранит данные, но и может сравнивать все записанные данные с поступающей информацией. На рис. 12.39 показана одна из возможных реализаций матрицы ассоциативного ЗУ. Устройство объединяет традиционную шеститранзисторную ячейку ОЗУ ($M4$–$M9$) с дополнительной схемой, выполняющей однобитовое цифровое сравнение ($M1$–$M3$). При записи информации в ячейку в разрядные шины подаются комплементарные данные, а числовая шина активизируется как стандартная ячейка статического ОЗУ. В режиме сравнения записанные дан-

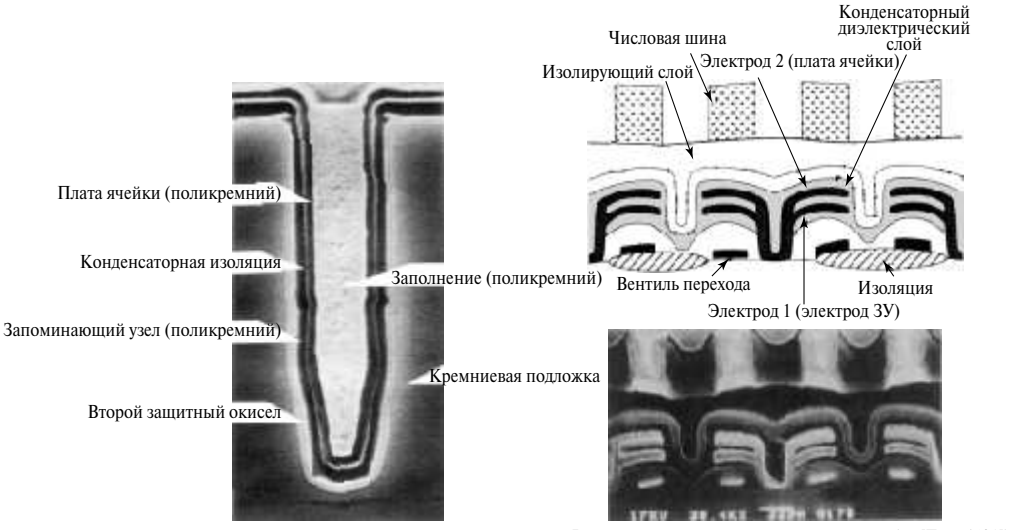


Рис. 12.38. Улучшенная однотранзисторная ячейка динамического ОЗУ

ные (S и комплементарное значение $\bar{S}$) сравниваются с входными данными, поступающими по комплементарным разрядным шинам (Бит и $\overline{\text{Бит}}$). Шина *соответствие* связана со всеми ячейками ЗУ в данной строке и предварительно заряжена до V_{DD}. Если значения в шинах S и Бит согласуются, внутренний узел *int* разряжается и транзистор $M1$ запирается, так что сигнал в шине *соответствие* сохраняет высокий уровень. Тем не менее, если записанный и входной биты отличаются, узел *int* заряжается до значения $V_{DD} - V_T$, что вызывает разрядку шины соответствия. Например, если Бит $= V_{DD}$ и $S = 0$, узел *int* заряжается через транзистор $M3$. Легко проверить, что данная схема реализует обычную функцию ИСКЛЮЧАЮЩЕЕ НЕ-ИЛИ (или является компаратором).

Здесь важно отметить, что разгрузочные устройства в компараторе соединяются с каждой ячейкой ассоциативной памяти в строке по типу монтажного соединения ИЛИ, т.е. даже если один из битов в данной строке не соответствует запросу, сигнал строки соответствия переходит на низкий уровень. В запоминающем устройстве, состоящем из N, большинство строк (несоответствий) будут в данном такте иметь низкий уровень. Очевидно, что с точки зрения рассеяния мощности это не очень привлекательно. Вообще, ассоциативная память не славится эффективным распределением мощности! Логические структуры можно переупорядочить так, чтобы переключалась только та строка, которая соответствует запросу, однако такое решение возможно за счет существенного снижения быстродействия.

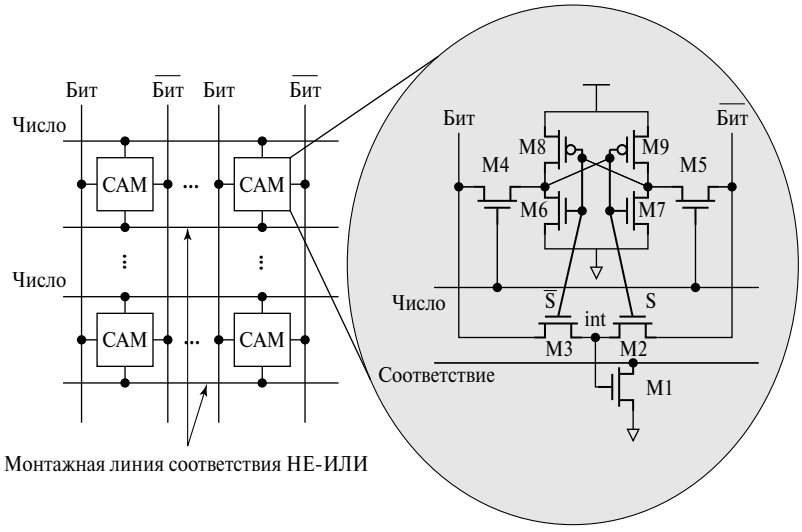


Рис. 12.39. Девятитранзисторная ячейка ассоциативной памяти

Пример 12.11. Использование ассоциативной памяти в кэше

На рис. 12.40 показано применение ячейки ассоциативной памяти в полностью ассоциативном кэше. *Кэш* — это небольшое, но очень быстрое запоминающее устройство, в котором хранится небольшая часть общей информации, требуемой в цифровой системе. Как правило, кэш используется в качестве временного хранилища данных. Например, если в определенной точке программы происходит обращение к данным, существует большая вероятность того, что эти данные снова потребуются в ближайшем будущем. Кэш располагается между процессором и большой и емкой основной памятью (динамическое ОЗУ). Он берет на себя расходы (время и мощность), связанные с получением данных из большого ЗУ. Матрица кэша состоит из двух частей: матрицы ассоциативной памяти, в которой хранятся адреса, и обычной матрицы статического ОЗУ, в которой хранятся данные. Когда процессору требуется записать информацию, в ассоциативную память записывается адрес, а в статическое ОЗУ — данные. Когда требуется записать новый адрес и новую информацию, а кэш полон, необходимо избавиться от одной существующей записи. Для этого применяется специальная политика замены кэша. Например, хорошим кандидатом на удаление является наименее часто используемое слово данных. В режиме чтения происходят следующие действия: адрес запрошенных данных передается матрице ассоциативной памяти, после чего происходит параллельный поиск. При наличии соответствия мы узнаем, что данные действительно находятся в кэше. В таком случае сигнал соответствия используется так же, как сигнал активизации числовой шины в статическом ОЗУ и в конечном счете позволяет получить данные. При отсутствии соответствия (т.е. сигналы во всех шинах соответствия имеют низкий уровень) происходит *промах кэша* и данные необходимо извлекать из внешней медленной систем-

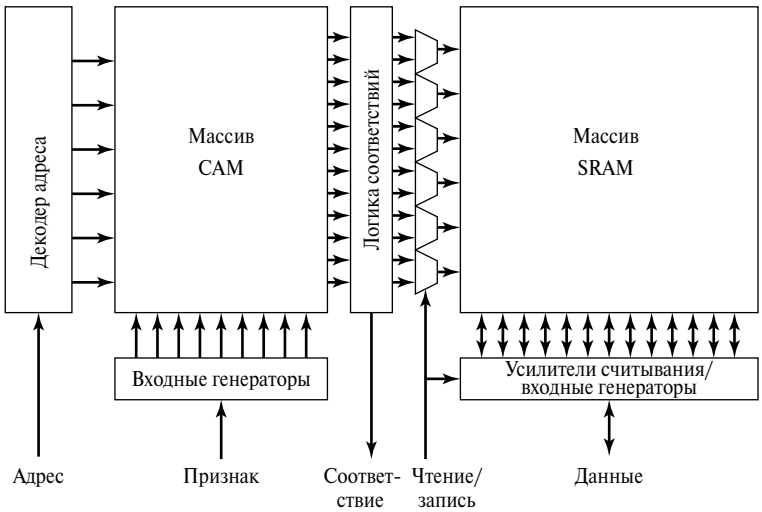


Рис. 12.40. Применение ячеек ассоциативной памяти: высокоскоростной внутрикристальный кэш

ной памяти. Более подробно о кэше можно узнать в различных учебниках, посвященных архитектурам вычислительных систем (например, [Hennessy02]).

12.3. Периферийные схемы запоминающих устройств*

Поскольку в сердечнике запоминающего устройства уменьшение площади происходит за счет быстродействия и надежности, разработчик ячеек памяти должен уделять должное внимание периферийным схемам, которые восстанавливают скорость и обеспечивают достоверность данных. Отметим, что проектирование сердечника ЗУ в основном определяется технологическими соображениями, на которые схемотехник обычно повлиять не может, но вот при проектировании периферийных устройств хороший разработчик может показать себя во всей красе. В данном разделе мы обсудим декодеры адреса, генераторы/буферы ввода-вывода, усилители считывания, а также вопросы тактирования и управления памятью.

12.3.1. Декодеры адреса

Если запоминающее устройство допускает выборку информации по произвольному адресу, то необходимо наличие декодеров адреса. Структура данных декодеров существенно влияет на скорость и потребляемую мощность запоминающего устройства. В разделе 12.1.2 были описаны два класса декодеров — декодеры строк, которые активизируют одну из 2^M строк памяти,

и декодеры столбцов и блоков, которые можно описать как мультиплексоры с 2^K входами (M и K — ширина соответствующих полей в адресном слове). При проектировании этих декодеров важно держать в уме общую топологическую структуру конечного запоминающего устройства. Декодеры очень тесно связаны с сердечником памяти, поэтому необходимо аккуратно согласовать геометрию декодеров и сердечника. Если этого не сделать, можно получить проект, в котором будет слишком много проводных соединений, а это скажется на задержке и рассеянии мощности. Несколько примеров согласованных по размерам декодеров и матриц памяти представлено в конце данной главы.

Декодеры строк

Декодер, выбирающий один элемент из 2^M, представляет собой совокупность 2^M сложных логических элементов с M входами. Рассмотрим 8-битовый декодер адреса. Каждый из выходов WL_i — это логическая функция восьми входных сигналов адресации (от A_0 до A_7). Например, строки с адресами 0 и 127 активизируются с помощью следующих логических функций:

$$\begin{aligned} WL_0 &= \bar{A}_0\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5\bar{A}_6\bar{A}_7; \\ WL_{127} &= \bar{A}_0A_1A_2A_3A_4A_5A_6A_7. \end{aligned} \tag{12.9}$$

Эту функцию можно реализовать в два этапа, используя элемент НЕ-И с восьмью входами и инвертор. Возможна и однокаскадная реализация, но для этого функцию необходимо преобразовать с использованием правил Де Моргана:

$$\begin{aligned} WL_0 &= \overline{A_0A_1A_2A_3A_4A_5A_6A_7}; \\ WL_{127} &= \overline{A_0\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5\bar{A}_6\bar{A}_7}. \end{aligned} \tag{12.10}$$

По сути, для реализации данной логической функции на строке требуется элемент НЕ-ИЛИ с восьмью входами. В результате на структуру налагается несколько условий. Во-первых, вся топология элемента НЕ-ИЛИ должна согласовываться с геометрией числовой шины. Во-вторых, большое разветвление по входу данного логического элемента отрицательно влияет на быстродействие. Задержка распространения декодера очень важна, поскольку она добавляется ко времени доступа для чтения и записи. Кроме того, данный элемент НЕ-ИЛИ должен питать большую нагрузку, представленную числовой шиной, не перегружая при этом входные адреса. Наконец, необходимо следить за рассеянием мощности декодером. Возможные статические и динамические реализации, учитывающие указанные моменты, мы обсудим в следующих разделах.

Структура статического декодера

Реализация широкой функции НЕ-ИЛИ по КМОП-технологии непрактичная. Одним из возможных решений возникающих проблем является возврат к псевдо-n-МОП-структурам, что позволит эффективно реализовать

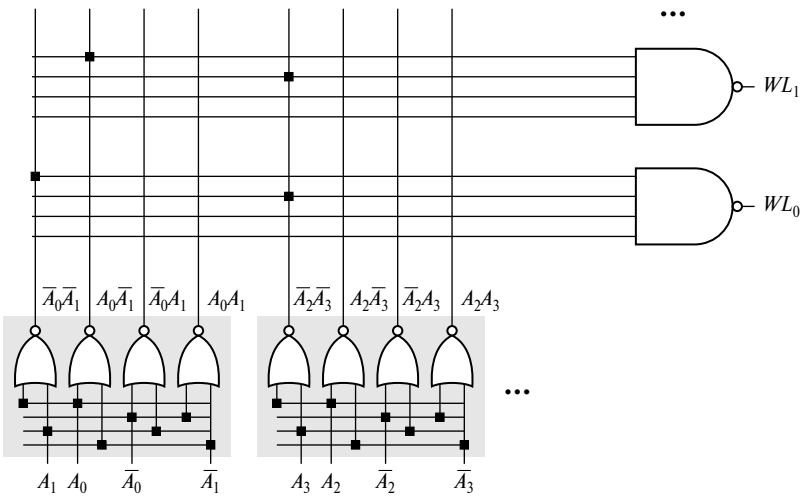


Рис. 12.41. Декодер НЕ-И, использующий предварительные декодеры с двумя входами

широкий элемент НЕ-ИЛИ. Тем не менее в современном мире данный подход не очень приветствуется из-за соображений рассеяния мощности.

К счастью, здесь нам помогают некоторые принципы, введенные в главе 6. Расщепление сложного логического элемента на несколько логических уровней часто дает более быструю и дешевую реализацию. Данная концепция разложения позволяет с помощью КМОП-технологии создать быстрые декодеры, эффективно использующие площадь. Поэтому она интенсивно используется в большинстве современных запоминающих устройств. Сегменты адреса декодируются на первом логическом уровне, именуемом *предварительным декодером* (predecoder). После этого на втором уровне логических элементов генерируются окончательные сигналы числовой шины.

Рассмотрим декодер НЕ-И с восьмью входами. Выражение для WL_0 можно переписать следующим образом:

$$\begin{aligned} WL_0 &= \overline{\bar{A}_0\bar{A}_1\bar{A}_2\bar{A}_3\bar{A}_4\bar{A}_5\bar{A}_6\bar{A}_7} = \\ &= \overline{\overline{(A_0 + A_1)(A_2 + A_3)(A_4 + A_5)(A_6 + A_7)}}. \end{aligned} \tag{12.11}$$

В данном случае адрес разбивается на два двухбитовых блока, которые заблаговременно декодируются. Затем получаемые в результате сигналы объединяются с использованием элементов НЕ-И с четырьмя входами и дают полностью декодированный сигнал числовой шины. Схематически получаемая структура показана на рис. 12.41.

Использование предварительного декодера имеет ряд неоспоримых преимуществ.

- Он уменьшает число требуемых транзисторов. Если предположить, что предварительный декодер реализован на комплементарных статических МОП-схемах, то число активных устройств в декодере с восемью входами получается равным $(256\times8)+(4\times4\times4) = 2,112$, или 52% от однокаскадного декодера, который потребовал бы 4 096 транзисторов.
- Поскольку число входов элементов НЕ-И сокращается вдвое, задержка распространения уменьшается примерно в четыре раза (напоминаем, что задержка пропорциональна квадрату разветвления по входу).

Тем не менее даже при такой структуре элемент НЕ-И с четырьмя входами управляет числовой шиной, которая представляет собой большую нагрузку. Самым лучшим задающим устройством для больших емкостей является инвертор, следовательно, выход элемента НЕ-И необходимо буферизировать. Чтобы максимально приблизить структуру декодера к оптимальной, можно прямо задействовать правила логических трудозатрат. В главе 6 было показано, что при питании очень больших нагрузок выгодно иметь больше каскадов логики. Выполняя преобразования, подобные продемонстрированным в уравнении (12.11), декодер можно разбить на еще большее число логических уровней, каждый из которых будет состоять из элементов НЕ-И с двумя входами, элементов НЕ-ИЛИ с двумя входами или инверторов.

Пример 12.12. Проектирование декодера

При проектировании декодера запоминающего устройства хорошей задачей является минимизация его задержки. Общую емкость числовых шин можно рассчитать, а максимальная емкость, нагружающая входные адресные шины, ограничивается согласно характеристикам проекта. В результате находится эффективное разветвление по выходу тракта декодера, F. Отметим, что существует множество возможных логических структур, реализующих функцию И с восемью входами, одна из которых показана на рис. 12.42. Выбранная структура состоит из инверторов и элементов НЕ-И с двумя входами. Вначале буферизуются входные адреса и группы из 4 бит предварительно кодируются. Эти предварительно кодированные сигналы трассируются вертикально и соединяются с последним уровнем декодеров. Поскольку декодеры последнего уровня состоят всего лишь из элементов НЕ-И с двумя входами, они прекрасно согласуются с геометрией числовой шины.

Логические трудозатраты данной цепочки равны $G = (4/3)^3 = 2,4$, а внутренняя задержка — $P = 12$. Суммарное ветвление B равно 128 $(= 2 \times 4 \times 16)$, как показано на рис. 12.42. Задавая эффективное разветвление по выходу F, можно определить минимальную задержку и оптимальные размеры устройств.

В приведенном анализе мы пренебрегали емкостью проводки C_W между предварительным и конечными декодерами, которая для больших запоминающих устройств может быть значительной. Сказанное станет очевидным при изучении полной архитектуры памяти, подобной показанной на рис. 12.6.

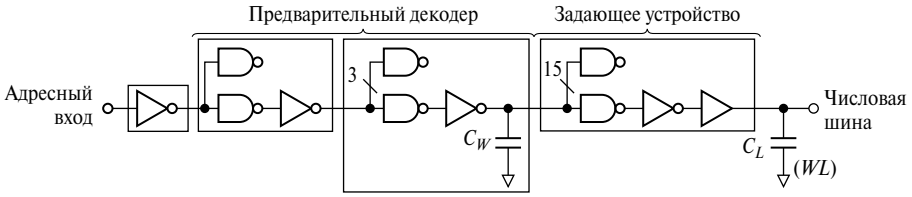


Рис. 12.42. Логическая цепочка восьмибитового декодера

Учитывая при оптимизации указанную емкостную нагрузку, мы получим более сложную, но решаемую задачу [Amrutur01].

Чтобы реализовать окончательную оптимизацию быстродействия, необходимо учесть, что сигнал в числовых шинах обычно имеет низкий уровень. Следовательно, оптимизировать нужно только переход на высокий уровень. Логические трудозатраты всех элементов цепочки можно уменьшить с помощью масштабирования, выгодного только одному переходу.

Все большие декодеры создаются с использованием по крайней мере двухуровневой реализации. Данная конфигурация "предварительный декодер/конечный декодер" обладает еще одним преимуществом. Добавление к каждому из предварительных декодеров сигнала отбора позволяет отключать декодер, когда соответствующий блок памяти не задействован. Понятно, что такая возможность очень полезна с точки зрения экономии энергии.

Динамические декодеры

Поскольку скорость декодера определяется всего одним переходом, интересно будет оценить другие реализации схемы. Ранее, например, мы рассматривали вариант использования псевдо-n-МОП-логики (и отказались от него). Лучшей альтернативой является динамическая логика, например, реализация, представленная на рис. 12.43, где показаны транзисторная диаграмма и концептуальная схема декодера 2:4. Обратите внимание на то, что данная структура геометрически идентична матрице ПЗУ на элементах НЕ-ИЛИ, отличаются только структуры данных.

Подобным образом мы можем создать декодер и в виде матрицы НЕ-И, эффективно реализовав инверсию функций, представленных в формуле (12.9). В этом случае все входы матрицы по умолчанию имеют высокий уровень, за исключением выбранной строки, имеющей низкий уровень. Эта связь посредством "активного низкого уровня сигнала" согласуется с требованиями к числовой шине ПЗУ не элемента НЕ-И, как обсуждалось в разделе 12.2.1. Обратите внимание на то, что интерфейс между декодером и запоминающим устройством часто включает буфер/задающий генератор, который при необходимости можно сделать инвертирующим. Для примера на рис. 12.44 показана конфигурация декодера 2:4 на элементах НЕ-И.

Все соображения, касающиеся быстродействия и емкости запоминающих устройств, изложенные при обсуждении матриц ПЗУ на элементах НЕ-И и НЕ-ИЛИ, действительны для обоих случаев. Декодеры на элементах НЕ-

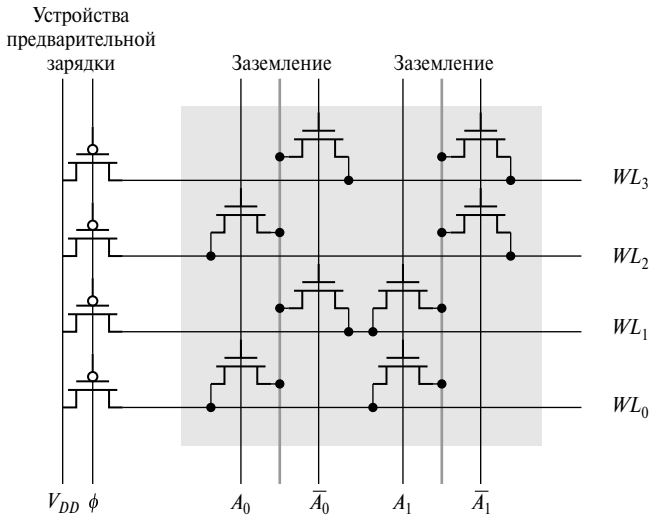


Рис. 12.43. Динамический декодер НЕ-ИЛИ 2:4

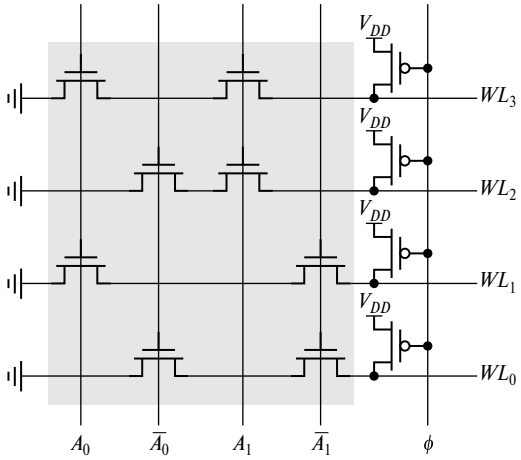


Рис. 12.44. Динамический МОП-декодер 2:4 на элементах НЕ-И. В данной реализации предполагается, что в фазе предварительной зарядки все сигналы адресации имеют низкий уровень. Альтернативный подход — использовать транзисторы оценки внизу каждой цепочки транзисторов

ИЛИ существенно быстрее, однако имеют бо́льшую площадь, чем аналогичные разработки на элементах НЕ-И, и потребляют гораздо большую мощность. Это очевидно, например, из следующего наблюдения: в декодере НЕ-И после предварительной зарядки необходимо понизить уровень только одной числовой шины, тогда как в структуре на элементах НЕ-ИЛИ только одна шина остается на высоком уровне. Как и в случае со статическими декодерами, большие декодеры создаются с использованием многоуровневого подхода.

Декодеры столбцов и блоков

Декодеры столбцов должны согласовываться с общей геометрией матрицы памяти (особенно разрядной шины). Для описания функциональных возможностей декодеров столбцов и строк эти устройства лучше всего рассматривать как мультиплексоры с 2^K входами, где K — размер адресного слова. Для матриц, допускающих чтение и запись, может проектироваться либо один общий набор мультиплексоров, либо два набора, отвечающих отдельно за чтение и запись. При чтении мультиплексоры должны обеспечивать путь разрядки от предварительно заряженных разрядных шин к усилителю считывания. Кроме того, мультиплексоры должны уметь понижать уровень сигнала в разрядной шине, поскольку это потребуется для записи значения 0 в ячейку памяти.

В настоящее время широко применяются две реализации описанной функции уплотнения. Выбор конкретного варианта зависит от соображений, касающихся площади, быстродействия и архитектуры.

Одну реализацию, основанную на КМОП-мультиплексоре на проходных транзисторах, мы ввели в главе 6 (см. рис. 6.46). Управляющие сигналы проходных транзисторов генерируются с использованием предварительного декодера $K : 2^K$, реализованного вдоль шин, описанных в предыдущем разделе. На рис. 12.45 показана схема декодера столбцов 4:1, использующего только n-МОП-транзисторы. Если подобные мультиплексоры совместно используются при чтении и записи, необходимо применять комплементарные передаточные элементы, что позволит обеспечить полный размах напряжения в обоих направлениях. Основным достоинством такого подхода является его быстродействие. На пути прохождения сигнала располагается только один проходной транзистор, дающий минимальное дополнительное сопротивление. Декодирование столбца — это одно из последних действий, выполняемых при чтении, поэтому параллельно с другими операциями (доступ к памяти, считывание) можно выполнить предварительное декодирование. Следовательно, задержка распространения данного элемента не добавляется к общему времени доступа к запоминающему устройству. Иногда бывают приемлемы даже более медленные реализации, например декодеры на элементах НЕ-И. Недостатком подобной структуры является большое число транзисторов. Для декодера с 2^K входами требуется $(K + 1)2^K + 2^K$ устройств. Например, декодер столбцов 1024:1 требует 12 288 транзисторов. Следует также помнить, что емкость (а значит, и переходная характеристика) на узле D пропорциональна числу входов мультиплексора.

Более эффективной реализацией является *древовидный декодер*, использующий двоичную схему сокращения, как показано на рис. 12.46. Обратите внимание на то, что в таком случае предварительный декодер не требуется. Число устройств также существенно уменьшается согласно приведенному ниже уравнению (для декодера с 2^K входами):

$$N_{tree} = 2^k + 2^{k-1} + \ldots + 4 + 2 = 2 \times (2^k - 1). \qquad (12.12)$$

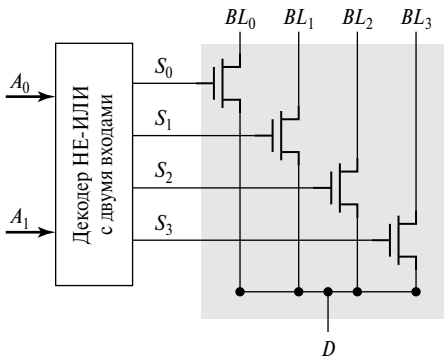


Рис. 12.45. Декодер столбцов на основе проходных транзисторов с четырьмя входами, использующий предварительный декодер НЕ-ИЛИ

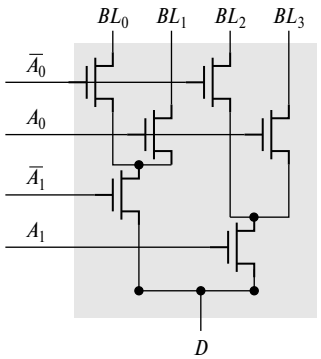


Рис. 12.46. Древовидный декодер столбцов 4:1

Это означает, что декодер 1024:1 требует всего 2046 активных устройств, т.е. в шесть раз меньше! С другой стороны, в тракт сигнала вводится цепочка из K последовательно соединенных проходных транзисторов. Поскольку задержка квадратично растет с увеличением числа каскадов, для больших декодеров древовидное решение получается очень медленным. Эту проблему можно решить, например, с помощью введения промежуточных буферов. Другой вариант — использовать прогрессивное масштабирование транзисторов, когда размер транзисторов последовательно увеличивается снизу вверх. Наконец, можно объединить древовидный подход с проходными транзисторами. В таком случае часть адресного слова (например, старшие разряды) будет предварительно декодироваться, а оставшиеся биты будут декодироваться с помощью дерева. Отметим, что такой подход позволяет уменьшить как число транзисторов, так и задержку распространения.

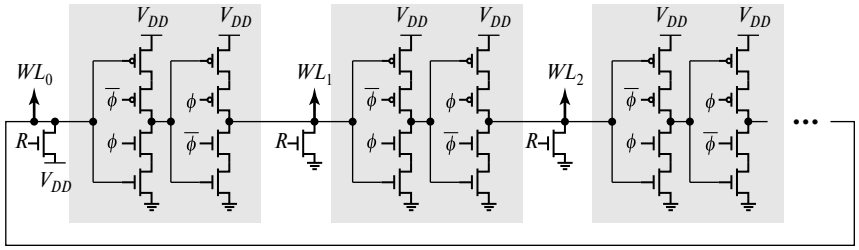


Рис. 12.47. Декодер для циклического сдвигового регистра. Сигнал R сбрасывает указатель в первую позицию

Пример 12.13. Декодеры столбцов

Рассмотрим декодер 1024:1. Если предварительно кодировать 5 бит, потребуется следующее количество транзисторов:

$$N_{dec} = N_{pre} + N_{pass} + N_{tree} = 6;$$
$$2^5 + 2^{10} + 2(2^5 - 1) = 1278!$$

Число последовательных проходных транзисторов уменьшилось до шести.

Декодеры для ЗУ без произвольной выборки

Ячейки памяти, не относящиеся к классу устройств с произвольной выборкой, не требуют полнофункционального декодера. В памяти с последовательной выборкой (например, видеопамяти) декодер вырождается в M-битовый сдвиговый регистр, где M — число строк. В любой момент времени в высоком состоянии может находиться только один бит, именуемый *указателем*. При следующей выборке указатель переходит на соседнюю позицию запоминающего устройства. Пример подобного вырожденного декодера, реализованного с использованием тактируемого КМОП-D-триггера, показан на рис. 12.47. Подобные подходы можно разработать и для других классов запоминающих устройств, например буферов ввода-вывода.

12.3.2. Усилители считывания

Усилители считывания являются главными устройствами, определяющими функционирование, быстродействие и надежность запоминающей схемы. В частности, они выполняют следующие функции.

- *Усиление.* В определенных запоминающих структурах, например однотранзисторной ячейке динамического ОЗУ, для правильного функционирования требуется усиление, поскольку размах напряжения в типичной схеме ограничен значением 100 мВ [Itoh01]. В других запоминающих устройствах усиление позволяет разрешать данные с небольшим размахом напряжения разрядной шины, что уменьшает рассеяние мощности и задержку.

- *Уменьшение задержки.* Усилитель компенсирует ограниченную нагрузочную способность по выходу ячейки памяти, ускоряя переходы в разрядной шине или детектируя и усиливая переходы с малой амплитудой в разрядной шине до сигналов с большим размахом.
- *Снижение энергопотребления.* Уменьшение размаха сигнала в разрядных шинах может существенно сократить рассеяние мощности, связанное с зарядкой и разрядкой разрядных шин.
- *Восстановление сигнала.* Поскольку в однотранзисторных ячейках динамических ОЗУ функции чтения и обновления неразрывно связаны, после считывания сигнал в разрядных шинах необходимо восстановить до полного размаха.

Топология усилителя считывания сильно зависит от типа запоминающего устройства, уровней напряжения и общей архитектуры памяти. Усилители считывания по своей природе являются аналоговыми устройствами, и работа с ними требует наличия опыта обращения с аналоговыми схемами. Следовательно, в данной книге мы очень кратко остановимся на проектировании подобных устройств. Более подробное обсуждение проектирования усилителей можно найти в соответствующих учебниках, например [Gray01], [Sedra87].

Дифференциальные усилители с датчиком напряжения

Дифференциальный усилитель принимает дифференциальные входы при малом уровне сигнала (т.е. низком напряжении разрядной шины) и усиливает их до несимметричного выходного сигнала с большим уровнем. Такая схема работы обычно называется дифференциальным подходом и обладает рядом достоинств, одним из которым является *ослабление синфазных сигналов*. Другими словами, данный усилитель ослабляет шум, равномерно добавляющийся к обоим входам. Такая возможность особенно привлекательна в запоминающих устройствах, в которых точное значение сигнала разрядной шины отличается для разных кристаллов и даже для разных точек одного кристалла. Другими словами, абсолютное значение сигнала 1 или 0 не известно точно и может варьироваться в довольно широком диапазоне. Картина еще больше усложняется из-за наличия нескольких источников шума, например всплесков напряжения питания и емкостных перекрестных помех между разрядными и числовыми шинами. Влияние таких шумов может быть значительным, особенно если вспомнить, что амплитуда считываемого сигнала обычно мала. Эффективность дифференциального усилителя характеризуется его способностью ослаблять общий шум и усиливать действительную разницу сигналов. Сигнал, общий для обоих входов, подавляется на выходе усилителя. Степень его ослабления называется *коэффициентом ослабления синфазного сигнала* (Common-Mode Rejection Ratio — CMRR). Подобным образом, подавление всплесков напряжения в цепях питания характеризуется *коэффициентом подавления шумов питания* (Power-Supply Rejection Ratio — PSRR).

К сожалению, дифференциальный подход можно прямо применить только к статическим ОЗУ, поскольку только эти ячейки памяти дают действи-

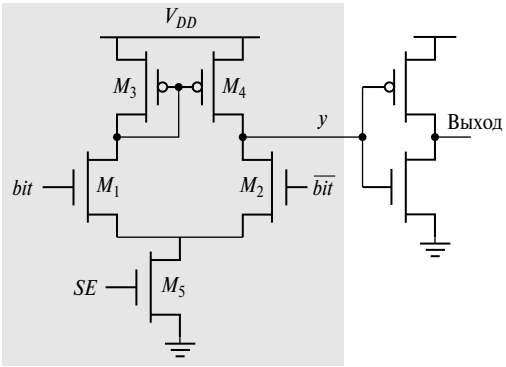


Рис. 12.48. Базовая схема дифференциального усилителя считывания

тельно дифференциальный выход. Рассмотрим показанную на рис. 12.48 базовую схему дифференциального усилителя считывания. Усиление выполняется за счет одного каскада на основе концепции токового зеркала. Выходные сигналы (bit и $\overline{bit}$) сильно нагружаются и управляются ячейкой памяти статического ОЗУ. Размах сигнала в этих шинах довольно мал, поскольку небольшая ячейка памяти управляет большой емкостной нагрузкой. Эти входы подаются на дифференциальные устройства ввода (M_1 и M_2), а транзисторы M_3 и M_4 действуют как зеркальные нагрузки активного тока. Усилитель активизируется разрешающим сигналом усилителя считывания SE. Изначально входы предварительно заряжаются и выравниваются; сигнал SE при этом имеет низкий уровень, отключая схему считывания. Как только инициируется операция чтения, одна из разрядных шин переходит в низкое состояние. Как только дифференциальный сигнал достигнет приемлемого уровня, активизируется сигнал SE и усилитель начнет работу.

Коэффициент усиления описанного устройства выражается следующим образом:

$$A_{sense} = -g_{m1}(r_{o2}||r_{o4}), \tag{12.13}$$

где g_{m1} — крутизна характеристики прямой передачи входных транзисторов; r_o — сопротивление транзистора в режиме малого сигнала. В области насыщения МОП-транзистора значение r_o очень высоко. Крутизну характеристики входных устройств можно увеличить, увеличив либо ширину транзисторов, либо ток смещения. В последнем случае также уменьшится выходное сопротивление транзистора M_2, что ограничивает полезность данного подхода. В целом коэффициент усиления может доходить до 100. Тем не менее усиление усилителей считывания обычно устанавливается равным порядка 10. Напомним, что основной задачей усилителя считывания является быстрая генерация выходного сигнала. Следовательно, усиление имеет не такое большое значение, как время реакции. Также отметим, что для получения желаемого сигнала с полным размахом обычно требуется несколько каскадов.

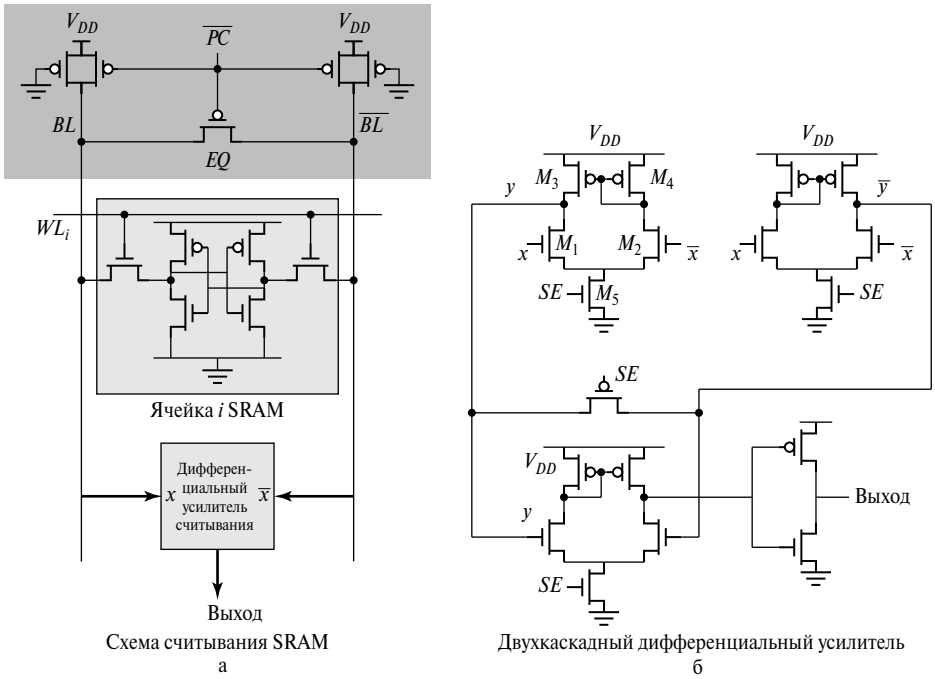


Рис. 12.49. Дифференциальное усиление применимо к столбцу статического ОЗУ

На рис. 12.49 показана реализация полностью дифференциального считывания и структура разрядного столбца ячейки статического ОЗУ. Разрядные шины соединяются с входами x и $\overline{x}$ двухкаскадного дифференциального усилителя. Цикл считывания выглядит следующим образом.

1. На первом этапе разрядные шины предварительно заряжаются до напряжения V_{DD} (для этого сигнал $\overline{PC}$ переводится на низкий уровень). Одновременно с этим включается p-МОП-транзистор EQ, гарантируя идентичность исходных напряжений в обеих разрядных шинах. Данная операция, именуемая *выравниванием*, необходима для того, чтобы усилитель считывания не выполнял ошибочного изменения напряжения в момент своего включения. На практике все дифференциальные сигналы в запоминающем устройстве выравниваются перед операцией чтения. Выравнивание критично, когда разрядные шины предварительно заряжаются через n-МОП-транзисторы, поскольку из-за вариаций порогов устройств величины предварительного заряда в разных шинах могут отличаться.
2. Собственно операция чтения начинается с деактивизации устройств предварительной зарядки и выравнивания и активизации одной из числовых шин. Одна из разрядных шин с помощью выбранной ячейкой памяти переводится на низкий уровень. Обратите внимание на то, что заземленная p-МОП-нагрузка, подключенная параллельно с транзистором предвари-

тельной зарядки, ограничивает размах разрядной шины и ускоряет следующий цикл предварительной зарядки.

3. Как только сигнал станет достаточно сильным (обычно порядка 0,5 В), посредством повышения уровня SE включается усилитель считывания. Дифференциальный входной сигнал в разрядных шинах усиливается с помощью двухкаскадного усилителя, и в конечном счете на выходе инвертора создается сигнал с полным размахом.

На рис. 12.49, *б* показана схема двухкаскадного усилителя считывания. Хотя данные структуры используют p-МОП-нагрузки и n-МОП-устройства ввода, в зависимости от условий смещения также довольно регулярно используется дуальная конфигурация с p-МОП-устройствами ввода и n-МОП-нагрузкой. Обратите внимание на то, что управляющий сигнал SE активизируется лишь на короткие промежутки времени, что позволяет уменьшить статическую мощность усилителя считывания. Кроме того, следует отметить, что один усилитель считывания совместно используется несколькими столбцами, для чего между ячейками памяти и усилителем вводятся проходные транзисторы декодера столбцов (в таком случае входные уровни усилителя уменьшаются на величину, равную порогу устройства). Это позволяет уменьшить площадь и потребляемую мощность.

Представленный ранее усилитель считывания развязывает входы и выходы, т.е. размах разрядной шины определяется ячейками статического ОЗУ и статической p-МОП-нагрузкой. Совершенно иной подход к усилению считывания реализован в схеме, приведенной на рис. 12.50, где в качестве усилителя считывания используется пара КМОП-инверторов с перекрестными обратными связями. Как отмечалось в главе 5, при работе в переходной области КМОП-инвертор характеризуется большим усилением. Чтобы триггер можно было использовать как усилитель считывания, его необходимо инициализировать в метастабильной точке, выравняв сигналы в разрядных шинах. В процессе чтения между разрядными шинами формируется разность напряжений. Как только разность напряжений станет достаточно большой, с помощью повышения уровня сигнала SE активизируется усилитель считывания. В зависимости от входа пара инверторов с перекрестными обратными связями перейдет в одну из стабильный рабочих точек. Благодаря положительной обратной связи переход выполняется достаточно быстро.

Хотя данный триггерный усилитель считывания является простым и быстрым, входы и выходы в нем не разделяются, так что переход сигнала в разрядных шинах имеет полную амплитуду. Это очень удобно для однотранзисторного динамического ОЗУ, где для обновления содержимого ячеек необходимо восстановление уровней сигнала в разрядных шинах. Поэтому ячейки с перекрестными обратными связями присутствуют практически во всех динамических ОЗУ. В следующем разделе мы покажем, как превратить несимметричную структуру (например, ячейку динамического ОЗУ) в дифференциальную.

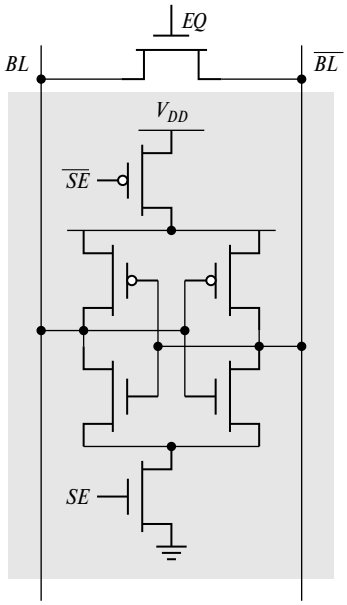


Рис. 12.50. КМОП-инверторы с перекрестными обратными связями, используемые как усилитель считывания

Несимметричное считывание

Хотя дифференциальное считывание, несомненно, является наиболее удачным решением, ячейки памяти, используемые в ПЗУ, СППЗУ, ЭСППЗУ и динамических ОЗУ, по своей природе несимметричны. Одно из возможных решений этой проблемы — прибегнуть к несимметричному усилению. Поскольку разрядные шины обычно предварительно заряжаются, мы можем использовать показанный на рис. 9.35 инвертор с асимметричным смещением. Кроме того, в небольших запоминающих устройствах часто используется интересное устройство, именуемое *усилителем с перераспределением заряда* (charge-redistribution amplifier) (рис. 12.51). Принцип его действия основан на использовании дисбаланса между большой емкостью C_{large} и гораздо меньшим компонентом C_{small}.

Данные конденсаторы разделяются транзистором M_1 [Heller75]. Узлы L и S (исходные напряжения V_{L0} и V_{S0}) предварительно заряжаются до значений $V_{ref} - V_{Tn}$ и V_{DD}, для чего узел S соединяется с шиной питания. Из-за падения напряжения на M_1 V_L предварительно заряжается только до $V_{ref} - V_{Tn}$. После включения одного из разгрузочных устройств (например, M_2) узел L начинает медленно разряжать свою большую емкость. До тех пор, пока $V_L \geq V_{ref} - V_{Tn}$, транзистор M_1 заперт и напряжение V_S остается постоянным. Как только V_L падает ниже отпираюшего напряжения транзистора ($V_{ref} - V_{Tn}$), M_1 отпирается. Начинается перераспределение заряда, и узлы L и S выравниваются. Благодаря маленькой емкости второго узла этот про-

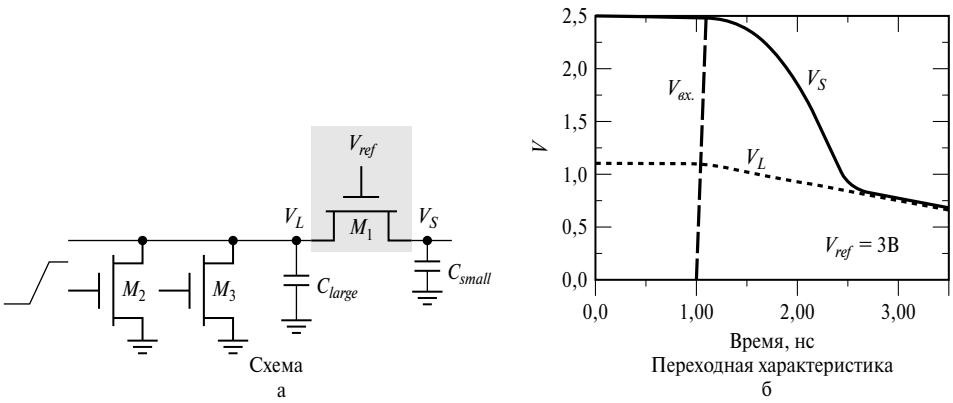


Рис. 12.51. Усилитель с перераспределением заряда

цесс может быть очень быстрым. Как видно из смоделированной переходной характеристики, изображенной на рис. 12.51, *б*, небольшие колебания напряжения на узле L приводят к большим скачкам напряжения на узле S. Следовательно, описанная схема действует как усилитель. Получаемый на ее выходе сигнал можно подать на инвертор, порог переключения которого выше, чем $V_{ref} - V_{Tn}$, и получить сигнал с полным размахом напряжения.

Схема усилителя с перераспределением заряда, используемого в запоминающих устройствах, показана на рис. 12.52. Данная структура была очень популярна в стираемых программируемых постоянных запоминающих устройствах (СППЗУ). К недостаткам усилителя с перераспределением заряда относится его слабая помехоустойчивость. Небольшое изменение напряжения на узле L, вызванное шумом или утечкой, может вызвать ошибочную разрядку узла S (см. рис. 12.51). Следовательно, к вопросу разработки подобных устройств нужно подходить очень внимательно.

Преобразование несимметричного входа в дифференциальный вход

При проектировании запоминающих устройств большой емкости (больше 1 Мбит), которые очень сильно подвержены влиянию шумов, разработчики прибегают к преобразованию несимметричных усилителей считывания в дифференциальные. Основной принцип преобразования иллюстрируется на рис. 12.53. Дифференциальный усилитель считывания с одной стороны соединяется с несимметричной разрядной шиной, а с другой — с источником опорного напряжения, расположенного между уровнями 0 и 1. В зависимости от значения BL усилитель переключается на одно или другое напряжение. Отметим, что создать хороший источник опорного напряжения не так просто, как кажется, поскольку уровни напряжения обычно различаются для разных кристаллов и даже для разных точек одного кристалла. Следовательно, опорный источник должен отслеживать эти вариации. Одно из популярных решений этой проблемы представлено на рис. 12.54 для однотранзисторного динамического ОЗУ. Матрица памяти разделена пополам дифференциаль-

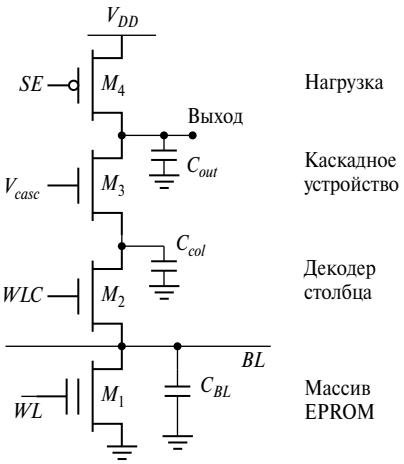


Рис. 12.52. Усилитель с перераспределением заряда, используемый в СППЗУ

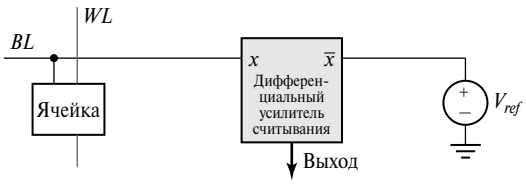


Рис. 12.53. Преобразование несимметричного входа в дифференциальный

ным усилителем. С обеих сторон добавлен столбец так называемых *балластных ячеек* (dummy cells) — однотранзисторных ячеек памяти, похожих на другие, но используемых только в качестве эталонных элементов. В целом данный подход часто называют *открытой архитектурой разрядной шины*.

При повышении уровня сигнала EQ обе разрядные шины BLL и BLR предварительно заряжаются до напряжения $V_{DD}/2$. Одновременная активизация L и $\bar{L}$ гарантирует, что балластные ячейки зарядятся до напряжения $V_{DD}/2$. В цикле чтения активизируется одна из числовых шин. Предположим, что посредством повышения уровня сигнала WL_0 выбирается ячейка из левой половины матрицы, что вызывает изменение напряжения в шине BLL. Соответствующее опорное напряжение генерируется путем одновременного выбора балластной ячейки из другой половины матрицы, для чего повышается уровень сигнала L. Предполагая, что правая и левая половины матрицы идеально сбалансированы, получаем напряжением в шине BLR, являющееся промежуточным между уровнями 0 и 1 и инициирующее переключение защелки считывания. Обратите внимание на важность поддержания идеальной симметрии. Одновременное повышение уровня числовых шин WL_0 и L превращает емкостную связь между разрядной и числовой шиной в синфазный

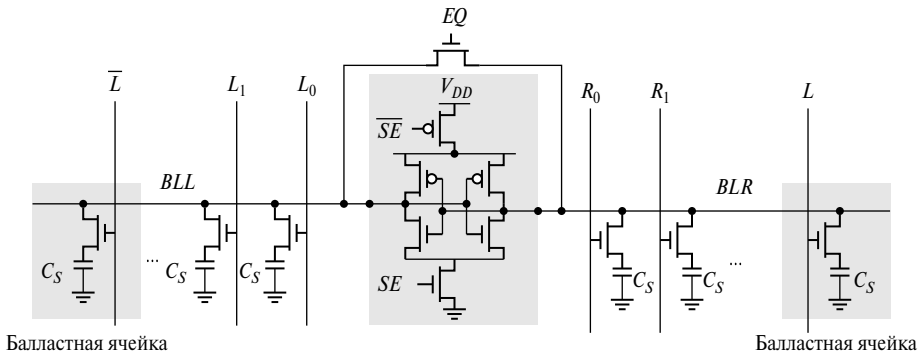


Рис. 12.54. Открытая архитектура разрядной шины с балластными ячейками

сигнал-помеху, который эффективно блокируется усилителем считывания. Обратите внимание также на то, что деление разрядных шин на две половины эффективно уменьшает емкость разрядной шины. В свою очередь, это удваивает коэффициент переноса заряда и улучшает отношение сигнал/шум.

Пример 12.14. Считывание в однотранзисторном динамическом ОЗУ

Операция чтения однотранзисторного динамического ОЗУ, реализованного с использованием открытой архитектуры разрядной шины и триггерного усилителя считывания, моделируется с помощью SPICE. Диффузионная емкость берется равной 50 фФ, а емкость разрядной шины — равной 0,5 пФ. Такой выбор параметров эквивалентен коэффициенту переноса заряда, равному 9%. Предположив, что разрядная шина предварительно заряжена до напряжения 1,25 В, перераспределение заряда приводит к падению напряжения на 110 мВ для сигнала “0”. Из-за снижения порога проходного транзистора напряжение ячейки, соотнесенное с сигналом “1”, равно 1,9 В, так что после активизации числовой шины повышение напряжения в разрядной шине составляет всего около 60 мВ. По результатам моделирования данные значения равны 110 и 45 мВ соответственно до включения усилителя считывания (рис. 12.55).

12.3.3. Опорное напряжение

Большинство запоминающих устройств в той или иной форме требуют внутрикристальной генерации напряжения. Работа сложной ячейки памяти требует нескольких опорных напряжений и уровней питания, часть из которых приведена ниже.

- *Добавочное напряжение числовой шины.* В обычной однотранзисторной ячейке динамического ОЗУ, использующей проходной n-МОП-транзистор,

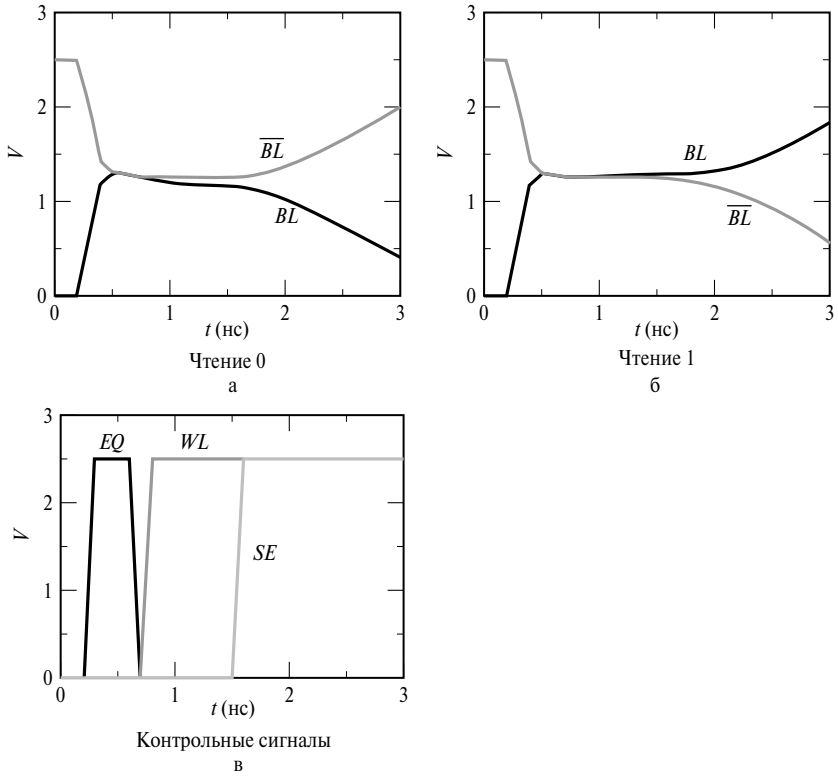


Рис. 12.55. Моделирование процесса чтения в открытой архитектуре разрядной шиной с ложными ячейками. Ложные ячейки соединены с разрядной шиной BL ($C_S = 50$ фФ; $C_{BL} = 0,5$ пФ)

максимальный уровень напряжения, который можно записать в ячейку, равен $V_{DD} - V_T$, что отрицательно сказывается на надежности запоминающего устройства. Поднимая напряжение числовой шины выше V_{DD} (точнее, до значения $V_{DD} + V_{Tn}$), можно записать сигнал с полным размахом. При таком подходе для получения повышенного напряжения обычно задействуется генератор подкачки заряда.

- *Половина V_{DD}.* Как обсуждалось ранее, разрядные шины динамического ОЗУ предварительно заряжаются до напряжения $V_{DD}/2$. Это напряжение должно генерироваться внутри кристалла.
- *Пониженное внутреннее напряжение источника питания.* Большинство запоминающих устройств работает при напряжении питания, меньшем напряжения внешнего источника. Для генерирования требуемых напряжений динамические ОЗУ (и другие ячейки памяти) используют внутренние регуляторы напряжения.
- *Отрицательное напряжение смещения подложки.* Эффективным средством контроля над пороговыми напряжениями внутри запоминающего

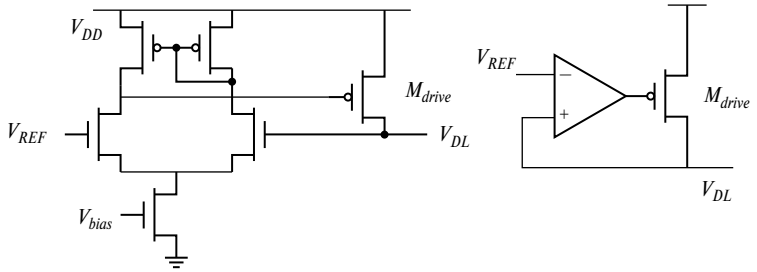


Рис. 12.56. Регулятор напряжения и его эквивалентное представление

устройства является подача отрицательного напряжения смещения подложки и использование контура управления. Данный подход применяется во всех новых поколениях динамических ОЗУ.

Разработка опорных напряжений относится к сфере аналогового проектирования. Поскольку данная книга посвящена в основном цифровым схемам, тему опорных схем мы рассмотрим очень кратко.

Устройства понижения напряжения

Устройства понижения напряжения используются для создания низких внутренних источников напряжения, позволяя схемам сопряжения работать при более высоком напряжении. Следовательно, внутренние схемы могут использовать агрессивные современные техники масштабирования напряжения, при этом оставаясь совместимыми с внешним миром. Фактически уменьшение напряжения необходимо для того, чтобы избежать пробоев в устройствах, реализованных по глубокосубмикронным технологиям. Преобразователи уровня, рассмотренные в главе 11, действуют как интерфейсы между внутренним сердечником и внешними схемами. Регуляторы также служат для установки стабильного внутреннего напряжения в системах, работающих от батарей, принимая широкий диапазон неконтролируемых входных напряжений. Это помогает решить проблему изменения напряжения батареи со временем.

На рис. 12.56 показана базовая структура устройства понижения напряжения (также именуемого *линейным регулятором*). Его работа построена на операционном усилителе, описанном в предыдущем разделе. Схема использует большой выходной задающий p-МОП-транзистор, управляющий нагрузкой запоминающего устройства (можно использовать и выходное n-МОП-устройство). Для установки выходного напряжения V_{DL} равным опорному применяется отрицательная обратная связь.

Конвертер должен давать напряжение, устойчивое к колебаниям условий эксплуатации. Медленные изменения (например, колебания температуры) можно компенсировать с помощью контура обратной связи. Обратите внимание на то, что неправильно спроектированный контур обратной связи может быть неустойчивым. В частности, ток нагрузки меняется со временем в довольно широких пределах, и конвертер необходимо спроектировать так, чтобы он приспосабливался к этим изменениям.

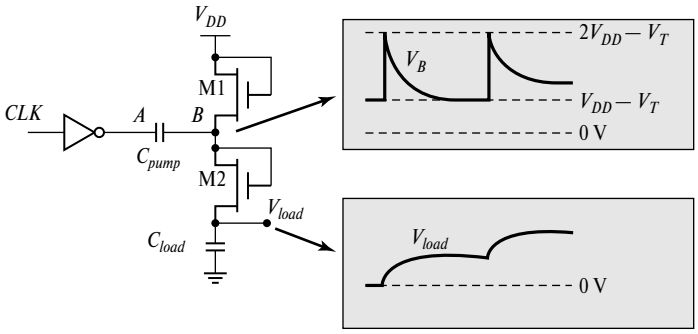


Рис. 12.57. Простой генератор подкачки заряда и его сигналы

Генераторы подкачки заряда

Технологии, подобные усилению числовой шины и смещению карманов, часто требуют источников напряжения, превышающего напряжение питания, но не потребляющих слишком большой ток. Идеальным решением в таком случае является генератор подкачки заряда. Данную концепцию можно проиллюстрировать с помощью простой схемы, приведенной на рис. 12.57. Транзисторы M_1 и M_2 соединяются как диоды. Предположим, что изначально тактовый сигнал CLK имеет высокий уровень. В этой фазе напряжение на узле A равно GND, а на узле B — $V_{DD} - V_T$. Заряд, накопленный на конденсаторе, выражается следующим образом:

$$Q = C_{pump}(V_{DD} - V_T). \tag{12.14}$$

Во второй фазе сигнал CLK переходит на низкий уровень, повышая напряжение на узле A до значения V_{DD}. Вместе с этим повышается напряжение на узле B, эффективно запирая транзистор M_1. Когда B на один порог выше V_{load}, M_2 начинает проводить ток и заряд переносится на C_{load}. В следующих тактах генератор подкачки заряда продолжает поставлять заряд V_{load}, пока на выходе не будет получено максимальное напряжение $2(V_{DD} - V_T)$. Ток, который можно получить от генератора, определяется в основном размером конденсатора и тактовой частотой. Эффективность генератора (ток, который не используется в течение каждого такта) составляет от 30 до 50%. Представленная на рис. 12.57 схема является достаточно простой и не очень эффективной. Существует довольно много более сложных генераторов подкачки заряда, характеризуемых бо́льшими диапазонами напряжения и лучшей эффективностью.

Опорное напряжение

Точное и устойчивое опорное напряжение является важной составляющей устройства понижения напряжения. Предполагается, что опорное напряжение сравнительно не зависит от колебаний напряжения питания и температуры. На рис. 12.58 приведен пример генератора опорного напряжения V_T.

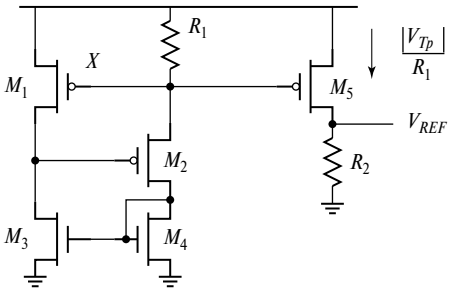


Рис. 12.58. Простой генератор опорного напряжения V_T

Нижние устройства (M_3 и M_4) действуют как токовое зеркало, пропуская такой же ток через сток транзистора M_1 и резистор R_1. Как видно из представленного ниже уравнения, увеличивая устройство M_1 и поддерживая достаточно низкий ток, напряжение истока-затвора транзистора M_1 можно сделать приблизительно равным $|V_{Tp}|$.

$$|V_{GS,M1}| = |V_{Tp}| + \sqrt{\left|\frac{2I_{M1}}{k_{p,M1}}\right|}. \tag{12.15}$$

Кроме того, ток, проходящий через резистор, и ток стока транзистора M_1 равны $|V_{Tp}|/R_1$. Обратите внимание на то, что устройство M_2 действует как транзистор смещения. Поскольку устройства M_1 и M_5 характеризуются таким же напряжением затвора-истока, ток стока транзистора M_1 зеркально отображается в M_5. Следовательно, опорное напряжение равно:

$$V_{REF} = |V_{Tp}| \cdot \frac{R_2}{R_1}. \tag{12.16}$$

Колебания порогового напряжения можно компенсировать с помощью лазерной подгонки резисторов. V_{REF} также характеризуется хорошей устойчивостью к колебаниям температуры. Выбирая подходящие материалы для реализации R_1 и R_2, температурную зависимость V_{TP} можно исключить с помощью температурной зависимости R_2/R_1. Такое решение использовалось, например, в динамическом ОЗУ емкостью 16 Мбит [Hidaka92] и характеризовалось прекрасной устойчивостью: $\Delta V_{REF}/\Delta T = 0,15$ мВ/°С и $\Delta V_{REF}/\Delta V_{DD} = 10$ мВ/В.

Разумеется, приведенный анализ далеко не полный. Более скрупулезное обсуждение опорных напряжений, к сожалению, выходит за рамки рассмотрения данной книги. Если вас интересует данная тема, обратитесь к [Itoh01] и [Gray01].

12.3.4. Задающие устройства/буферы

Длина числовой и разрядной шины увеличивается с увеличением размеров запоминающих устройств. Хотя отчасти сопутствующее ухудшение быстродействия можно смягчить путем секционирования матрицы памяти, значительная доля времени доступа на чтение и запись приходится на задержки проводов. Поэтому значительная часть площади периферийных устройств приходится на задающие устройства, в частности, буферы адресов и формирователи ввода-вывода. (О проектировании каскадных буферов см. в главе 5. Все сделанные там выводы справедливы и в контексте данного анализа.)

12.3.5. Синхронизация и контроль

Из приведенного выше обсуждения видно, что модуль памяти является довольно сложным объектом, работа которого управляется четко определенной последовательностью действий: фиксирование адреса, декодирование числовой шины, предварительная зарядка и выравнивание разрядной шины, активизация усиления считывания и формирование выхода. Для обеспечения нормальной работы устройства необходимо, чтобы данная последовательность действий реализовывалась во всем диапазоне рабочих условий при всех технологических параметрах и характеристиках устройств. Для получения максимального быстродействия необходима аккуратная синхронизация различных событий. Хотя схемы синхронизации и контроля занимают минимальную долю площади, их проектирование является неотъемлемой и очень важной частью всего процесса разработки запоминающего устройства. Данные цепи требуют аккуратной оптимизации и интенсивного последовательного моделирования средствами SPICE для широкого диапазона рабочих условий.

Тем временем возникают различные подходы к синхронизации запоминающих устройств, которые можно отнести к *тактируемым* и *самосинхронизируемым*. Чтобы проиллюстрировать их отличие от традиционных подходов, мы последовательно рассмотрим типичные представители каждого класса.

Динамическое ОЗУ — тактируемый подход

С самого начала своего развития динамические ОЗУ подходили для уплотненной схемы адресации, когда для экономии выводов корпуса адрес строки и адрес столбца последовательно представлялись в одной адресной шине. Данный подход пережил множество поколений запоминающих устройств и попрежнему используется в современном мире, хотя и считается громоздким и несоответствующим системным требованиям.

В уплотненной схеме адресации пользователь должен предоставить два основных управляющих сигнала — RAS (Row-Address Strobe — строб адреса строки) и CAS (Column-Address Strobe — строб адреса столбца), — которые свидетельствуют о наличии адресов строки и столбца соответственно (см. рис. 12.8). Еще один управляющий сигнал (W) указывает, чем является предполагаемая операция — чтением или записью. Данные сигналы можно интерпретировать как внешние тактовые сигналы и использовать для син-

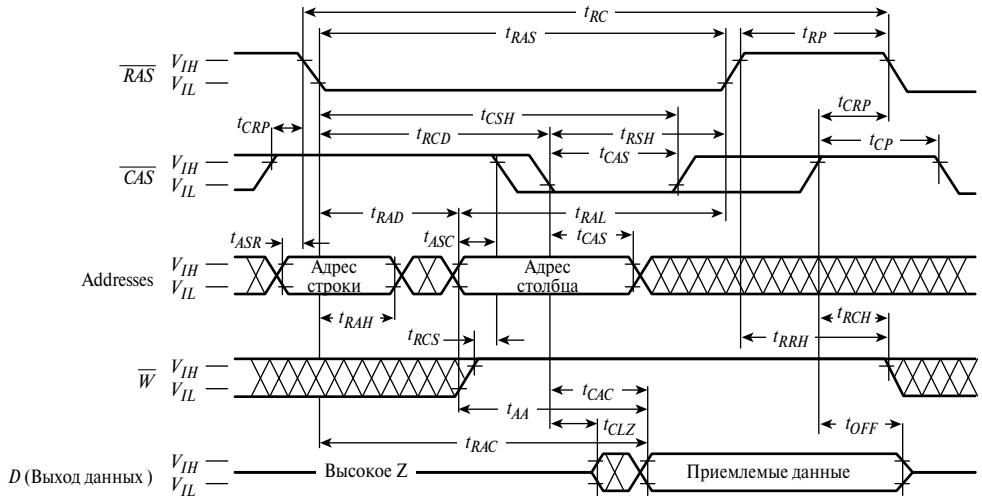


Рис. 12.59. Временная диаграмма цикла чтения для динамического ОЗУ 1×4 Мбит [Mot91]. Минимальная продолжительность цикла чтения t_{RC} равна 110 нс. Расстояние между задними фронтами сигналов RAS и CAS должно составлять от 20 до 40 нс. В целом число временны*х* условий равно 24

хронизации внутренних событий запоминающего устройства. Так же как при синхронном подходе к тактированию, сигналы RAS и CAS должны достаточно отделяться друг от друга, чтобы все инициируемые ими операции успевали завершиться. На рис. 12.59 показана упрощенная временн*а*я диаграмма динамического ОЗУ (1×4) Мбит и некоторые из наложенных временн*ы*х ограничений. Полная спецификация содержит больше 20 временн*ы*х параметров, которые все должны находиться в жестких рамках, обеспечивая надлежащую работу схемы. Очевидно, сгенерировать правильные тактовые сигналы для таких запоминающих устройств совсем непросто!

Все внутренние тактовые сигналы (EQ, PC и SE) являются производными от внешних управляющих сигналов. Чтобы максимизировать быстродействие, некоторые из этих сигналов должны попадать в строго определенные интервалы. Например, сигнал SE нельзя подавать слишком рано после декодирования числовой шины, поскольку это увеличивает чувствительность к шуму. С другой стороны, слишком запоздалая реакция замедляет скорость работы запоминающего устройства. Для разделения этих событий используется задержка числовой шины, которая может зависеть от колебаний температуры и параметров процесса. Довольно часто разработчики настраивают временной интервал с рабочими параметрами, включая в схему генератора тактовых сигналов модель числовой шины (*фиктивную числовую шину*). Данный "элемент задержки" имеет точно такую же задержку, как и реальная числовая шина, поэтому запоминающее устройство становится более надежным, сохраняя при этом прежнее быстродействие. Таким образом, можно утверждать, что динамические ОЗУ объединяют синхронный подход на глобальном уровне с техниками самосинхронизации, которые применяются для генерирования некоторых локальных сигналов.

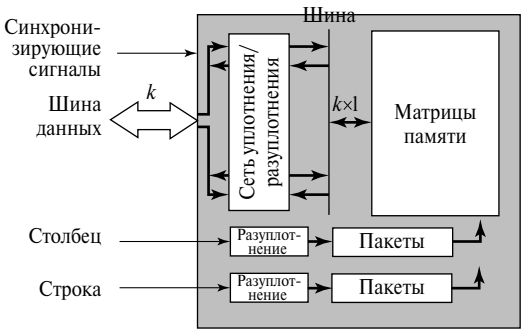


Рис. 12.60. Блок-схема ячейки RDRAM

Синхронные динамические ОЗУ

Одной из основных проблем динамических ОЗУ является большое время доступа и низкая пропускная способность данных. При постоянно повышающейся скорости микропроцессоров и их статических ОЗУ-кэшей увеличивается разница в быстродействии между ядром процессора и основной памятью DRAM. Это является одной из основных проблем в сфере проектирования быстродействующих систем. Для ее решения применяются высокоскоростные синхронные динамические ОЗУ, например синхронные динамические ОЗУ (Synchronous DRAM — SDRAM) [Yoo95] и RDRAM (Rambus DRAM) [Crisp97]. Разработчики данных запоминающих устройств отошли от традиционной модели синхронизации с помощью сигналов RAC/CAS. Оставив ядро памяти практически без изменений, разработчики использовали тот факт, что сердечник ЗУ является, по сути, параллельным устройством; другими словами, для каждого данного цикла чтения и записи большое число битов можно считывать/записывать одновременно, хотя и с относительно низкой скоростью. Перенос этих битов в память или извлечение их из памяти можно реализовать с помощью высокоскоростного асинхронного интерфейса, скорость работы которого близка к тактовой частоте процессора. Платой за реализацию данной идеи являются дополнительные защелки и буферы в интерфейсе с сердечником запоминающего устройства, а также схема, поддерживающая интерфейс ввода-вывода с высокой скоростью передачи данных. Тем не менее, несмотря на эти дополнительные расходы, описанный подход является крайне эффективным, если данные записываются большими последовательными блоками.

Рассмотрим в качестве примера ячейку RDRAM. Входные/выходные данные последовательно в течение нескольких тактов передаются по узкой шине. При этом шина работает с очень высокой скоростью и использует эффективные протоколы пакетной передачи данных. Следовательно, большую порцию данных можно передать за очень короткий промежуток времени. С шиной можно соединить несколько кристаллов памяти. На рис. 12.60 схематически показана схема ввода-вывода получающегося в результате устройства RDRAM.

Статические ОЗУ — самосинхронизация

В отличие от динамических ОЗУ статические ОЗУ исторически разрабатываются из расчета глобальной статической перспективы. Работой запоминающего устройства управляют события в адресной шине или сигнал R/W (см. рис. 12.8, *б*). Никаких дополнительных управляющих или синхронизирующих сигналов не требуется. Кажется, что подобный подход подразумевает полностью статическую реализацию всех схем (декодеров, усилителей считывания и т.д.). В таком случае изменение одного из входных сигналов (информационной или адресной шины, R/W) будет просто "проноситься" через все последующие уровни схемы.

Тем не менее из сказанного в предыдущих разделах понятно, что такой полностью статический подход неприменим для больших запоминающих устройств как с точки зрения площади, так и с позиции рассеяния мощности. Поэтому в статических оперативных запоминающих устройствах используется более разумный подход, называемый *детектированием изменения адресов* (Address-Transition Detection — ATD), для автоматического генерирования таких внутренних сигналов, как PC и SE, при детектировании изменений внешней среды.

Цепь детектирования изменения адресов играет важную роль в архитектуре модулей статических ОЗУ и ПЗУ. Она выступает источником большинства тактовых сигналов и неотъемлемой частью критического временн*о*го пути. Следовательно, ее быстродействие имеет первостепенную важность. Одна из возможных реализаций схемы детектирования адресных переходов показана на рис. 12.61. Она состоит из нескольких одновибраторов, активизируемых переходами (см. рис. 7.50) (по одному на каждый входной бит), соединенных в псевдо-n-МОП-конфигурацию НЕ-ИЛИ. Переходы в любом из входных сигналов заставляют сигнал ATD переходить на низкий уровень на время t_d. Получающийся в результате импульс действует как основной эталонный временной сигнал для остальных элементов ЗУ, что дает огромное разветвление по выходу. Следовательно, рекомендуется использовать подходящую схему буферизации. Более элегантная схема тактирования статического ОЗУ рассмотрена в примере в конце главы.

12.4. Надежность и выход готовых запоминающих устройств*

Как статические, так и динамические оперативные запоминающие устройства работают в условиях низкого отношения сигнал/шум. Для обеспечения стабильной работы памяти необходимо максимизировать сигналы и минимизировать помехи. Еще одной проблемой, затрудняющей разработку ячеек памяти, является низкий выход готовых устройств, вызванный структурными и периодическими дефектами. В данном разделе мы кратко обсудим некоторые из проблем, а также их возможные решения.

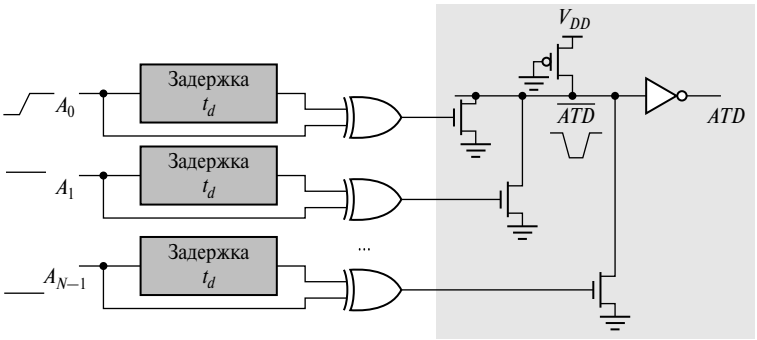


Рис. 12.61. Схема детектирования изменения адресов. Линии задержки обычно реализуются в виде цепочек инверторов

12.4.1. Отношение сигнал/шум

Разработчики сделали многое, чтобы создать ячейки памяти, генерирующие максимально сильный сигнал на единицу площади. Тем не менее, несмотря на эти усилия, качество генерируемого сигнала постепенно ухудшается с увеличением плотности схем (рис. 12.62). Например, емкость ячейки динамического ОЗУ снизилась примерно с 70 фФ для памяти емкостью 16 Кбит до менее чем 30 фФ для запоминающих устройств нынешнего поколения. Одновременно с этим уменьшались уровни напряжения (как из соображений энергопотребления, так и с позиции надежности). В современных запоминающих устройствах нормальными являются напряжения, не превышающие 1 В. Следовательно, заряд, запасаемый в конденсаторе, при переходе от памяти емкостью 64 Кбит к памяти емкостью 64 Мбит уменьшился в 5 раз. То, что уменьшение оказалось не таким стремительным (при том, что размер ячейки за это время уменьшился в 100 раз), объясняется только введением инновационных технологий.

В то же время растущая плотность интеграции повышает уровень шума, вызванного интерференцией сигналов. Хотя проблемой емкостной связи между числовой и разрядной шинами разработчики занимаются с начала 1980-х, актуальной она стала при более тесном расположении шин. Кроме того, этому способствовало повышение требований к скорости, что приводит к увеличению помех переключения в каждом следующем поколении, и случайные ошибки, индуцированные α-частицами, которые могут случайным образом изменить состояние высокоимпедансного узла и представляют собой дополнительный источник помех. Последняя проблема традиционно была актуальна только для динамических запоминающих устройств, но в настоящее время она присуща и статическим ОЗУ из-за быстрого увеличения импеданса запоминающих узлов статических ОЗУ. На рис. 12.63 перечислены источники шумов, влияющих на работу однотранзисторной ячейки динамического ОЗУ. Большинство этих факторов подробно рассмотрены в следующих разделах.

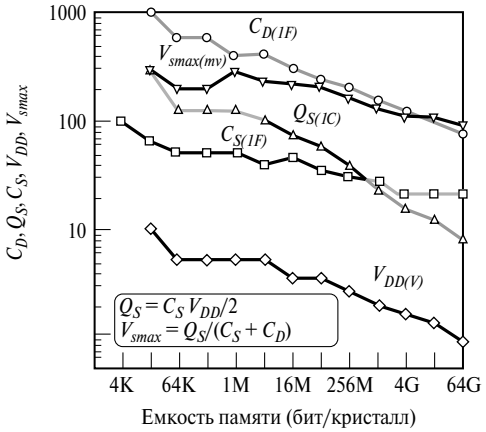


Рис. 12.62. Тенденции изменения параметров динамических ОЗУ (согласно [Itoh01]). C_D, V_{smax}, Q_s и C_s обозначают емкость разрядной шины, напряжение сигнала считывания, заряд ячейки и емкость ячейки соответственно

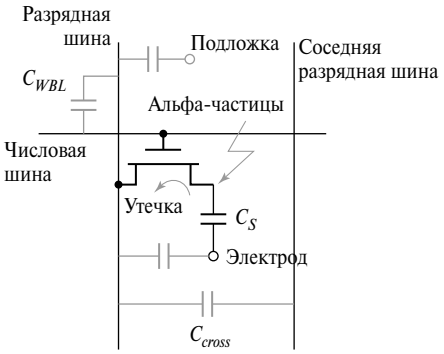


Рис. 12.63. Источники шума в однотранзисторном динамическом ОЗУ

Емкостная связь между числовой и разрядной шинами

Рассмотрим открытую конфигурацию разрядной шины, приведенную на рис. 12.64. В какой-то момент времени выбирается числовая шина WL_0, что приводит к введению некоторого заряда в левую разрядную шину. Мы предполагаем, что балластные ячейки справа, выбранные с помощью WL_D, не вызывают никакого перераспределения заряда, поскольку шина предварительно заряжена до значения $V_{DD}/2$. Наличие переходной емкости C_{WBL} между числовой и разрядной шинами приводит к перераспределению заряда, причем величина переносимого заряда примерно равна $\Delta_{WL} \times C_{WBL}/(C_{WBL} + C_{BL})$, где Δ_{WL} — размах напряжения в числовой шине, а C_{BL} — емкость разрядной шины. Если бы обе стороны матрицы памяти были абсолютно симметричными, наведенный шум разрядной шины (Δ_{BL}) был бы одинаковым с обеих

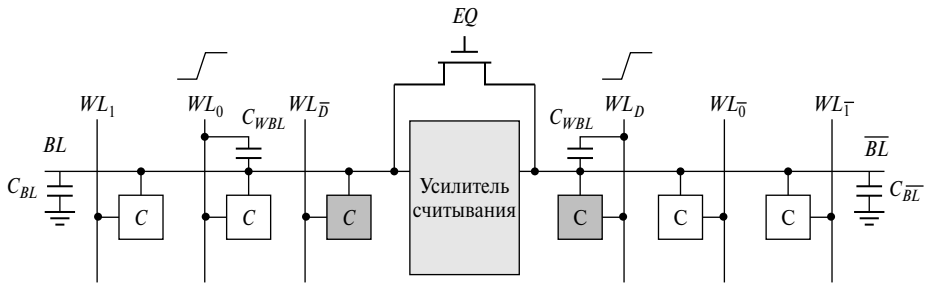


Рис. 12.64. Емкостная связь между числовой и разрядной шинами при открытой архитектуре разрядной шины

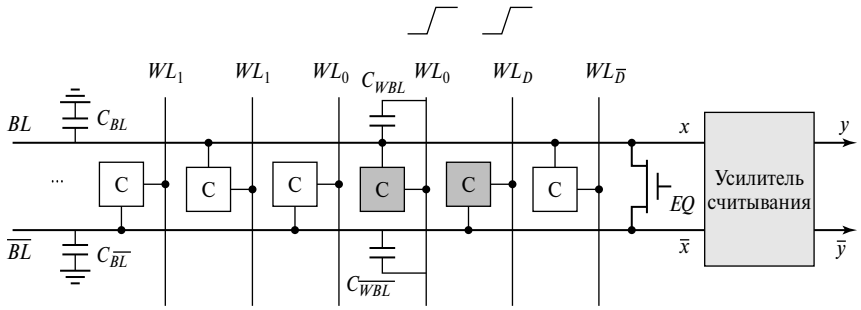


Рис. 12.65. Петлеобразная разрядная шина в однотранзисторном динамическом ОЗУ

сторон и усилитель считывания воспринимал бы его как синфазный сигнал. К сожалению, такая ситуация невозможна, поскольку в пределах матрицы памяти переходная емкость и емкость разрядной шины могут варьироваться.

Данную проблему можно решить с помощью *петлеобразной архитектуры разрядной шины*, показанной на рис. 12.65. Размещая усилитель считывания в конце матрицы памяти и прокладывая рядом шины BL и $\overline{BL}$, получим более согласованные паразитную емкость и емкость разрядной шины. Сигналы числовой шины (WL_0 и WL_D) пересекают обе разрядные шины. Следовательно, шум перекрестного соединения проявляется как синфазный сигнал, даже если формы кривых напряжения в шинах WL_0 и WL_D существенно отличаются. Отметим, что хотя петлеобразная архитектура разрядной шины имеет очевидные преимущества с точки зрения шумоподавления, она увеличивает длину разрядной шины, а следовательно, и ее емкость. Впрочем, правильно чередуя ячейки, данные дополнительные расходы можно свести к минимуму.

Представленный выше сценарий охватывает лишь часть общей картины. Мы можем описать и более сложный случай емкостной связи между числовой и разрядной шинами. В процессе чтения в разрядной шине BL происходит изменение напряжения на ΔV. Этот переход образует связь с невыбранными числовыми шинами, находящимися в режиме предварительной зарядки, а

следовательно, представляющими высокоимпедансные узлы. В свою очередь, данный сигнал шума образует связь с другими разрядными шинами. Получающиеся в результате перекрестные помехи зависят от структуры данных, записанной в памяти. Предположим, что большинство разрядных шин считывает значение 0. Это приводит к отрицательным переходам в невыбранных числовых шинах, что, в свою очередь, понижает уровень сигнала в разрядных шинах. Такая ситуация особенно опасна для разрядных шин, считывающих значение 1. Из-за того что данный сигнал шума зависит от структуры данных, его очень сложно отследить. Следовательно, тестирование запоминающего устройства на предмет устойчивости к такой помехе требует перебора множества наборов данных. Благодаря петлеобразной архитектуре разрядной шины такие помехи можно подавить.

Связь между разрядными шинами

Влияние межпроводникового перекрестного соединения увеличивается при уменьшении размеров элементов, как подробно описывалось в главе 9. Особенно это актуально для матриц памяти, где на небольшом расстоянии друг от друга прокладываются чувствительные к шуму разрядные шины. Задача разработчика — превратить сигналы шума в синфазные сигналы для усилителя считывания. Весьма остроумным решением данной проблемы является *скрещенная* (или *скрученная*) архитектура, показанная на рис. 12.66. На рис. 12.66, *а* показана прямолинейная реализация данной концепции. Обе шины, BL и $\overline{BL}$, соединены с соседним столбцом (или строкой) посредством конденсатора C_{cross}. В наихудшем случае размах сигнала, детектируемого усилителем считывания, может быть уменьшен на величину

$$\Delta\, V_{cross} = 2\frac{C_{cross}}{C_{cross} + C_{BL}} V_{swing}, \tag{12.17}$$

где V_{swing} — размах сигнала в разрядных шинах. Из-за этой интерференции может теряться до одной четвертой уже слабого сигнала [Itoh90].

В *скрещенной архитектуре разрядной шины* (см. рис. 12.66, *б*) данный источник возмущения исключен с помощью деления разрядных шин на сегменты, соединенные перекрестно. Данная модификация представляет сигнал помехи приблизительно одинаково для обеих шин, BL и $\overline{BL}$, т.е. превращает его в синфазный сигнал. Возможно и альтернативное решение — использовать обкладку конденсатора как дополнительный экранирующий слой между шинами данных.

Утечка

Утечка заряда быстро становится одним из основных источников помех в запоминающих устройствах. Это утверждение справедливо уже не только для динамических, но и для статических запоминающих устройств, в которых становится все больше высокоимпедансных узлов. В однотранзисторном динамическом ОЗУ существуют два механизма утечки, влияющих на запоминающий узел невыбранной ячейки: утечка заряда через pn-переход на под-

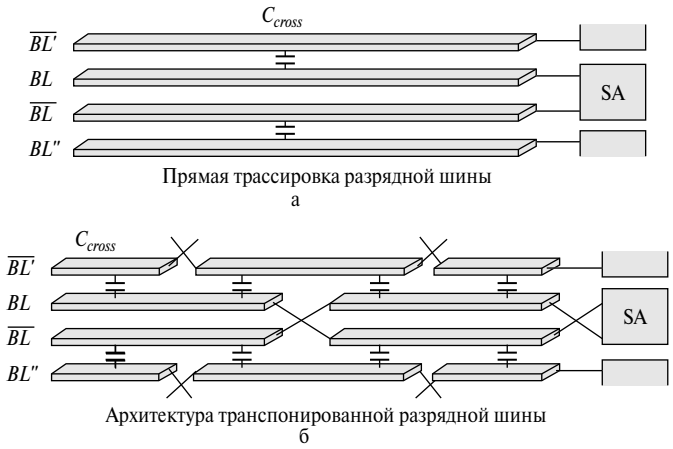


Рис. 12.66. Связь между разрядными шинами

ложку и предпороговый ток в разрядную шину. Данная утечка вызывает постепенное ухудшение сигнала запоминающего узла. Следовательно, задав минимальное допустимое отношение сигнал/шум, получаем условие максимального времени обновления t_{REFmax}. Операции обновления требуют энергии, а следовательно, невыгодны с точки зрения системы, поскольку уменьшают возможную пропускную способность запоминающего устройства. Это неизбежное зло можно попытаться уменьшить с помощью различных архитектурных модификаций. Как правило, обычная операция обновления заключается в последовательном переключении всех числовых шин с одновременным обновлением всех ячеек, соединенных с этими шинами. Таким образом, количество строк в матрице памяти выгодно свести к минимуму. Более того, разбив запоминающее устройство на несколько блоков, можно будет одновременно обновлять несколько строк. Впрочем, даже при таких архитектурных улучшениях вопрос подавления утечки становится все более и более актуальным с каждым новым поколением запоминающих устройств.

С утечкой через переход можно бороться двумя способами: (1) улучшать качество процесса производства, например меняя профиль легирования; (2) уменьшать температуру места перехода, поскольку ток через переход сильно зависит от этой температуры. Здесь свою эффективность доказали маломощные техники, уменьшающие теплоотдачу, и использование корпусов с низким тепловым сопротивлением. Теоретические расчеты показывают, что утечка через переход может быть меньше 3,5 аА при 25°С для памяти емкостью 64 Гбит!

Еще бо́льшую проблему представляет подпороговая утечка. Единственное разумное ее решение заключается в увеличении порога транзистора доступа (и это при том, что все остальные напряжения уменьшаются!). Разумеется, в таких условиях поддерживать достаточный ток выборки и быстродействие довольно сложно. Единственное приемлемое решение — агрессивно масштабировать период обновления!

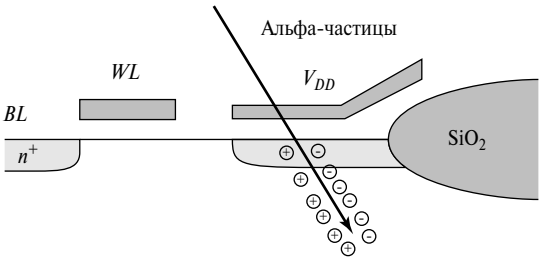


Рис. 12.67. Альфа-частицы, вызывающие нерегулярные ошибки

Критический заряд

Разработчики первых запоминающих устройств терялись перед лицом *нерегулярных ошибок* (soft errors), т.е. единовременных и непостоянных ошибок в динамических ОЗУ, которые нельзя объяснить шумом в цепях питания, утечкой или перекрестными помехами. Отметим, что нерегулярные ошибки оказывают сильное влияние на проектирование системы. Если разработчик не найдет достаточной защиты от них, они могут вызвать невосстановимый сбой компьютерной системы, так что отладчик будет практически бессилен. Первой значимой работой, в которой была разобрана роль нерегулярных ошибок в проектировании системы, была [May79]. В ней основными виновниками этих ошибок были признаны *альфа-частицы*.

Альфа-частицы представляют собой ядра He^{2+} (два протона, два нейтрона), излучаемые в процессе распада радиоактивными элементами. Следы таких элементов неизбежно присутствуют в материалах, из которых производится корпус устройства. При энергии излучения порядка 8-9 МэВ альфа-частицы могут проходить до 10 мкм вглубь кремния. При этом они интенсивно взаимодействуют с кристаллической структурой и генерируют в подложке порядка 2×10^6 электронно-дырочных пар. Проблема нерегулярных ошибок возникает тогда, когда траектория одной из этих частиц пересекается с запоминающим узлом ячейки памяти. Рассмотрим, например, ячейку, показанную на рис. 12.67. Когда в ней хранится значение 1, потенциальная яма пуста. Электроны и дырки, генерируемые налетающей частицей, диффундируют через подложку. Электроны, достигающие края обедненной области до рекомбинации, захватываются электрическим полем запоминающего узла. Если собирается достаточное количество электронов, то записанное значение может измениться на 0. Процент электронов, попадающих в ячейку, называется *эффективностью собирания*.

Недавно было установлено, что еще одним важным источником нерегулярных ошибок являются нейтроны, порожденные космическими лучами. Исследователи доказали, что в современных статических КМОП-ОЗУ частота ошибок, порожденных космическими лучами, имеет тот же порядок, что и частота ошибок, порожденных альфа-частицами. Даже больше, похоже, что основным источником нерегулярных ошибок в КМОП-защелках являются именно нейтроны [Tosaka97].

Частоту появления нерегулярных ошибок можно уменьшить, поддерживая заряд ячейки на уровне выше *критического* Q_C. В таком случае мы получаем условия на минимальную емкость запоминающего узла и напряжение ячейки. Например, емкость 50 фФ, заряженная до потенциала 3,5 В, содержит $1,1 \times 10^6$ электронов. Одна альфа-частица, налетающая на такую ячейку с эффективностью собирания 55%, может стереть весь заряд. Это одна из основных причин, объясняющих, почему даже в наиболее плотных запоминающих устройствах емкость ячейки превышает 30 фФ. Другой подход заключается в уменьшении эффективности собирания, которая, как показывают расчеты, обратно пропорциональна длине по диагонали обедненной области ячейки. Таким образом, величина критического заряда уменьшается в ходе масштабирования технологии! Например, для длины 2,5 мкм критический заряд равен 30 фКл, что соответствует запоминающему устройству емкостью 64 Мбит. Незаменимым средством, позволяющим уменьшить число альфа-частиц, является покрытие кристалла и повышение чистоты материала. Например, чтобы защитить кристалл запоминающего устройства от альфа-излучения, его можно дополнительно покрыть полиимидом[7].

Хотя с помощью внимательного проектирования появление нерегулярных ошибок можно свести к минимуму, совсем избавиться от них довольно сложно. Еще одним их источником являются свободные нейтроны, порождаемые космическими лучами, проходящими через атмосферу. Взаимодействие одного из этих нейтронов с ядром кремния (маловероятное, но реальное событие) может полностью изменить записанные данные, поскольку нейтроны генерируют примерно в 10 раз больше зарядов, чем альфа-частицы.

Таким образом, случайной ошибки очень сложно избежать, тем не менее такая ошибка не обязательно должна быть фатальной. Обнаружить и исправить большинство ошибок можно с помощью различных технологий системного уровня. Похоже, что в будущих запоминающих устройствах такие техники коррекции ошибок будут просто незаменимыми, поэтому мы кратко рассмотрим их в следующем разделе.

12.4.2. Выход годных запоминающих устройств

Несмотря на улучшение процесса производства, с увеличением размера кристалла и плотности интеграции следует ожидать уменьшения выхода годных запоминающих устройств. Причинами аппаратных сбоев детали могут быть как дефекты материала, так и колебания параметров процесса. Для повышения выхода годной продукции и снижения стоимости сложных компонентов разработчики запоминающих устройств используют две концепции: избыточность и исправление ошибок (эта техника помогает бороться и с нерегулярными ошибками).

[7] Стоит отметить, что нерегулярные ошибки также представляют проблему и в современных статических ОЗУ, больше зависящих от емкостных накопителей энергии на высокоимпедансных узлах.

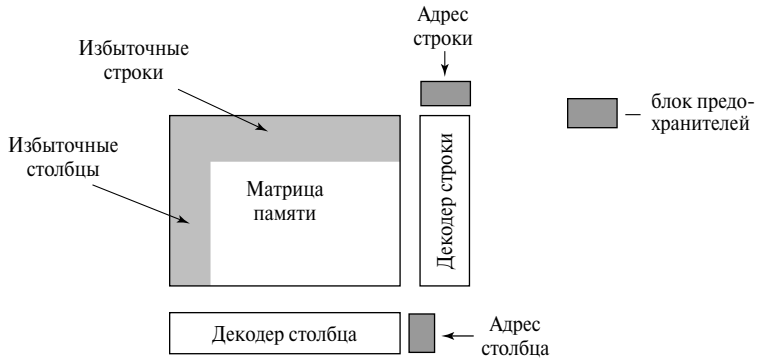


Рис. 12.68. Выход годных матриц памяти повышается за счет использования избыточности

Избыточность

Запоминающие устройства хороши тем, что имеют очень правильную структуру, поэтому внедрение *избыточных цепей* реализуется довольно просто. Дефектные разрядные или числовые шины матрицы памяти легко заменить избыточными (рис. 12.68). Когда в процессе тестирования блока памяти обнаруживается дефектный столбец, он заменяется запасным столбцом путем программирования блока предохранителей, соединенного с декодером столбца. Как правило, для этого предохранительную перемычку пережигают с помощью лазера или импульса тока. Лазерное программирование оказывает минимальное воздействие на быстродействие памяти, а его реализация требует небольшой области кристалла. При этом оно требует специального оборудования и увеличивает время обработки полупроводниковых пластин. Подход с использованием импульса тока можно реализовать с помощью стандартных инструментов, но он требует большей подготовки. Аналогичные подходы применяются и для обработки дефектных числовых шин. При обнаружении сбоящей числовой шины система избыточности задействует запасную числовую шину. В современных запоминающих устройствах запасными можно заменить более 100 дефектных элементов, при этом дополнительные издержки на реализацию этой возможности составляют меньше 5%. В настоящее время избыточность задействована даже во внедренных статических ОЗУ, используемых в системах на кристалле.

Коррекция ошибок

Избыточность помогает исправлять сбои, затрагивающие большой участок запоминающего устройства, например дефектные разрядные или числовые шины. При этом она неэффективна при наличии рассеянных точечных ошибок, например локальных сбоев, вызванных дефектами материала. Чтобы в таких условиях добиться разумного охвата дефектных участков, требуется очень сильная избыточность и существенное увеличение площади структуры. Поэтому гораздо лучшим способом решения таких проблем является

коррекция ошибок. Принцип действия данной схемы основан на том, чтобы с помощью избыточного представления данных детектировать и даже исправлять ошибочные биты. Например, добавив к данным бит четности, можно обнаружить (но не исправить) ошибку. Мы не будем углубляться в подробное исследование техник коррекции ошибок, а просто приведем пример, иллюстрирующий данную концепцию.

Пример 12.15. Коррекция ошибок с использованием кодов Хэмминга

Рассмотрим 4-битовое число (B_i), закодированное с использованием трех контрольных битов (P_i):

$$P_1 P_2 B_3 P_4 B_5 B_6 B_7,$$

где P_i выбирается так, что

$$P_1 \oplus B_3 \oplus B_5 \oplus B_7 = 0;$$
$$P_2 \oplus B_3 \oplus B_6 \oplus B_7 = 0;$$
$$P_4 \oplus B_5 \oplus B_6 \oplus B_7 = 0.$$

Предположим теперь, что B_3 стал ошибочным. Из-за этого в первых двух уравнениях будет получаться не 0, а 1, а в последнем по-прежнему будет получаться 0. В двоичной кодировке это означает, что ошибочным является бит 011 (или 3). Этой информации достаточно, чтобы исправить данную ошибку. Вообще, для исправления одной ошибки должно выполняться следующее условие:

$$2^k \geq m + k + 1, \tag{12.18}$$

где m и k — число информационных и контрольных битов соответственно. Например, для реализации коррекции ошибки в одном бите из 64 нужно 7 контрольных битов, т.е. суммарная длина слова должна составлять 71 бит.

Здесь важно отметить, что коррекция ошибок позволяет не только бороться с технологическими сбоями, но и эффективно решать проблему нерегулярных ошибок и нестационарных сбоев. Например, коррекция ошибок очень эффективна при борьбе с колебаниями порога в ЭСППЗУ.

Коррекция ошибок и избыточность с разных сторон подходят к проблеме выхода готовых запоминающих устройств. Чтобы охватить все возможности, необходимо использовать оба подхода. Чтобы понять справедливость сказанного, достаточно посмотреть на рис. 12.69, где показан процент выхода годных динамических оперативных запоминающих устройств емкостью 16 Мбит при использовании отдельно коррекции ошибок и избыточности, а также при объединении этих технологий [Kalter90].

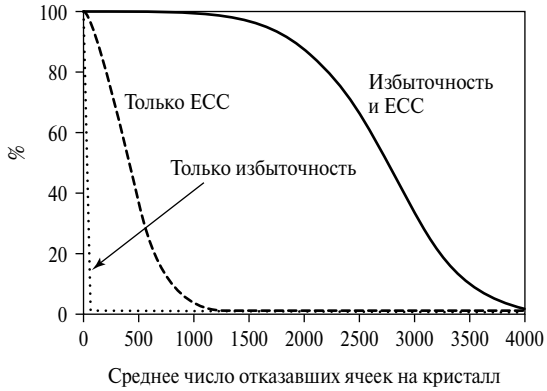


Рис. 12.69. Процент выхода годных устройств (динамическое ОЗУ емкостью 16 Мбит) при использовании кодов коррекции ошибок (ECC) и избыточности разрядной шины [Kalter90]

12.5. Рассеяние мощности в запоминающих устройствах*

Как и в большинстве сфер цифровых разработок, уменьшение рассеяния мощности в запоминающих устройствах имеет первостепенную важность. Разработчики матриц памяти очень постарались свести к минимуму рассеяние даже при том, что за последние тридцать лет емкость запоминающих устройств увеличилась на 6 порядков. Тем не менее проблема продолжает оставаться довольно актуальной. Например, в сфере портативных приложений величина потребляемой мощности довольно жестко ограничена. В то же время масштабирование технологии, сопровождающееся понижением напряжений питания и порогов, приводит к увеличению потребления мощности в стационарном режиме. Следовательно, внедрение новых энергосберегающих технологий в сфере запоминающих устройств действительно является насущной необходимостью.

12.5.1. Источники рассеяния мощности в запоминающих устройствах

Существуют три основных источника потребления мощности в кристалле памяти: матрица ячеек памяти, декодеры (строки, столбца, блока) и периферийные устройства. Обобщенное уравнение активной мощности для современной КМОП-памяти, состоящей из m столбцов и n строк, было выведено в [Itoh01] (рис. 12.70). Для нормального цикла чтения имеем следующие выражения:

$$\begin{aligned} P &= V_{DD} I_{DD}; \\ I_{DD} &= I_{array} + I_{decode} + I_{periphery} = \\ &= [m i_{act} + m(n-1) i_{hld}] + [(n+m) C_{DE} V_{int} f] + [C_{PT} V_{int} f + I_{DCP}]. \end{aligned} \tag{12.19}$$

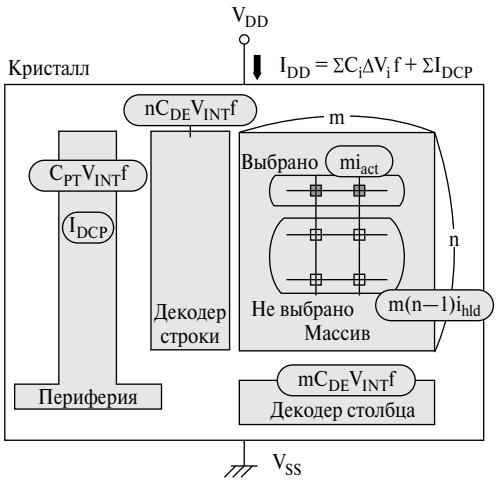


Рис. 12.70. Источники рассеяния мощности в полупроводниковых запоминающих устройствах [Itoh01]

В данном уравнении использованы такие параметры:

- i_{act} — эффективный ток выбранных (или активных) ячеек;
- i_{hld} — ток хранения данных неактивных ячеек;
- C_{DE} — емкость выходного узла каждого декодера;
- C_{PT} — общая емкость КМОП-логики и периферийных схем;
- V_{int} — внутреннее напряжение питания;
- f — рабочая частота;
- I_{DCP} — статический (или квазистатический) ток периферийных устройств. Основными источниками данного тока являются усилители считывания и схемы столбцов. Кроме того, источником тока являются внутрикристальные генераторы напряжения.

Как и следовало ожидать, рассеяние мощности пропорционально размеру матрицы памяти (n, m). Чтобы уменьшить рассеяние, необходимо разбивать запоминающее устройство на подматрицы и использовать небольшие значения n и m. Очевидно, что данный подход эффективен только тогда, когда в стационаром режиме неактивные ячейки памяти имеют очень низкое рассеяние мощности.

Вообще, основной вклад в рассеяние мощности запоминающего устройства дает матрица памяти. Активное рассеяние мощности периферийными схемами мало по сравнению с другими компонентами. Тем не менее мощность стационарного режима может быть достаточно высокой, поэтому усилители считывания, когда они не требуются, должны выключаться. Кроме того, в современных ОЗУ ток зарядки декодера также мал, особенно если разработчик проследил, чтобы в каждом цикле заряжался только один из n или m узлов.

Теме уменьшения рассеяния мощности в запоминающих устройствах можно посвятить не одну главу. К сожалению, объем книги ограничен, поэтому мы лишь перечислим некоторые наиболее распространенные технологии. Более подробное и глубокое обсуждение данной темы приводится в [Itoh01].

12.5.2. Секционирование памяти

Разработчики, разбивающие запоминающие устройства на модули, не сразу смогли ограничить рассеяние активной мощности избранными областями всей матрицы. Блоки запоминающего устройства, не используемые в данный момент, должны потреблять ровно столько мощности, сколько требуется для хранения данных. Секционирование памяти выполняется путем уменьшения m (числа ячеек в числовой шине) и (или) n (числа ячеек в разрядной шине). Разбивая числовую шину на несколько подшин, которые активизируются только при обращении к ним, можно уменьшить суммарную переключаемую емкость. В некотором смысле данная схема представляет собой многокаскадный иерархический декодер строк. Вообще, описанный подход весьма популярен в статических ОЗУ, как видно на примере рассмотренного в следующем разделе устройства емкостью 4 Мбит.

Подобным образом, секционирование разрядной шины уменьшает емкость, переключаемую при каждой операции чтения/записи. В динамических ОЗУ часто применяется метод частичной активизации разрядной шины. Разрядная шина разбивается на несколько участков, которые совместно используют общий усилитель считывания, декодер столбцов и модуль ввода-вывода. Этот подход помогает уменьшить переключаемую емкость со значения, превышающего 1 пФ (динамическое ОЗУ емкостью 16 Кбит), приблизительно до 200 фФ (динамическое ОЗУ емкостью 64 Мбит).

12.5.3. Решение проблемы рассеяния активной мощности

Как и для логических схем, одной из наиболее эффективных техник уменьшения рассеяния мощности в запоминающих устройствах является понижение уровней напряжения. Впрочем, в отличие от логических схем масштабирование напряжения может очень быстро исчерпать свои возможности. Из-за соображений, касающихся хранения данных и надежности, масштабирование напряжения до уровней, существенно меньших 1 В, представляет собой довольно серьезную проблему. Кроме того, необходим тщательный контроль над емкостью и активностью переключения, а также минимизация времени включения периферийных компонентов.

Уменьшение активной мощности статических ОЗУ

Чтобы добиться быстрого чтения, размах напряжения в разрядной шине делается как можно меньшим — обычно от 0,1 до 0,3 В. Получаемый в результате сигнал передается на усилитель считывания для восстановления. Поскольку этот сигнал получается в результате совместной работы нагрузки разрядной шины и транзистора ячейки, ток проходит через разрядную шину, пока

активизирована числовая шина (Δt). Ограничивая Δt и размах разрядной шины, можно свести к минимуму активное рассеяние ячеек статических ОЗУ.

Для операции записи ситуация получается несколько хуже, поскольку шины BL и $\overline{BL}$ должны отработать по полной программе. Единственным выходом является уменьшение базового напряжения питания (плюс, разумеется, описанное выше секционирование разрядной шины). В конечном счете уменьшение базового напряжения ограничивается рассогласованием парных МОП-транзисторов в ячейке статического ОЗУ. Даже если они разрабатывались как совершенно идентичные устройства, колебания параметров процесса и устройств приводят к тому, что МОП-транзисторы в ячейке будут отличаться. Еще более важной является несогласованность порогов V_T, которая часто вызывается неравномерностью примесей, колебаниями длины и ширины канала и даже случайными микроскопическими флуктуациями легирующих атомов в очень маленьких устройствах. Несоответствие транзисторов делает ячейку несимметричной, смещая ее к состоянию 1 или 0. В результате ячейка становится существенно менее надежной при наличии помех. Если исправить ситуацию с помощью контроля над процессом производства невозможно, необходимо задействовать техники избыточности и коррекции ошибок.

Уменьшение напряжения питания также влияет на время доступа к памяти. Понижение порогов устройств допустимо только тогда, когда разработчик способен эффективно побороть получаемое в результате увеличение тока утечки (особенно в неактивных ячейках).

Уменьшение активной мощности динамических ОЗУ

Поскольку процесс считывания ячейки динамического ОЗУ является деструктивным, необходимо последовательное применение к выбранным ячейкам действий считывания, усиления и восстановления. Следовательно, в каждом цикле чтения разрядные шины заряжаются и разряжаются до полного размаха напряжения (ΔV_{BL}). В таком случае необходимо следить за тем, чтобы уменьшить заряд, рассеиваемый разрядной шиной $mC_{BL}\Delta V_{BL}$, поскольку именно он определяет активную мощность. Уменьшение C_{BL} (емкость разрядной шины) выгодно как с точки зрения потребляемой мощности, так и с точки зрения отношения сигнал/шум (уравнение (12.8)), однако для больших запоминающих устройств это сделать не очень просто. Уменьшить ΔV_{BL}, конечно, выгодно с точки зрения мощности, но при этом пострадает отношение сигнал/шум. Это непосредственно следует из уравнения (12.8), из которого видно, что заряд, запасенный в ячейке (а следовательно, и сигнал считывания), прямо пропорционален размаху напряжения, приложенного в цикле записи или обновления. Следовательно, уменьшение напряжения должно сопровождаться либо увеличением размера конденсатора и (или) подавлением шума. Здесь вам могут помочь техники, описанные ниже.

- *Предварительная зарядка разрядных шины до* V_{DD} помогает уменьшить рассеяние активной мощности почти в два раза.
- *Добавочное напряжение числовой шины* в процессе записи позволяет избежать падения напряжения на транзисторе доступа, что существенно увеличивает хранимый заряд.

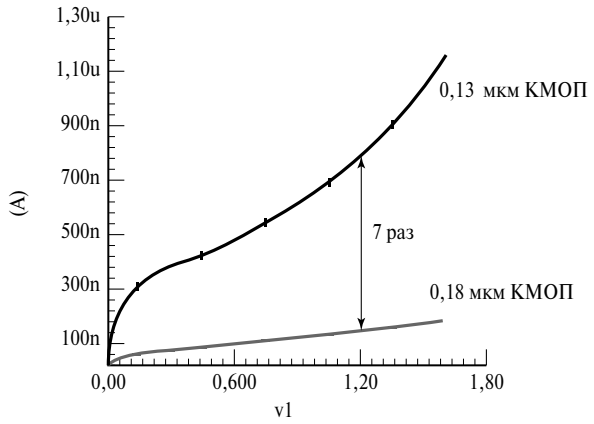


Рис. 12.71. Ток утечки во внедренном статическом ОЗУ емкостью 8 Кбит как функция напряжения

- *Увеличение площади или емкости транзистора.* Вертикальные транзисторы позволяют весьма эффективно увеличить емкость ячейки (хотя и за счет дополнительной обработки и повышения стоимости производства).
- *Увеличение размера ячейки.* В конечном счете в динамических ОЗУ, работающих при очень низком напряжении, возможно, придется пожертвовать эффективным использованием площади (особенно, если речь идет о памяти, встраиваемой в систему на кристалле). В таком случае желательно выяснить, какие ячейки являются наиболее разумной альтернативой (например, трехтранзисторные динамические ОЗУ).

12.5.4. Рассеяние хранимой информации

Хранение данных в статических ОЗУ

В принципе матрица статического ОЗУ вообще не должна давать статического рассеяния мощности (i_{hld}). Тем не менее ток утечки транзисторов запоминающего устройства становится основным источником тока хранения. Хотя данный ток всегда был небольшим, в последних поколениях внедренных запоминающих устройств наблюдается существенное его увеличение, вызванное подпороговой утечкой. Сказанное иллюстрируется на рис. 12.71, где показана зависимость тока в режиме хранения для встроенной памяти емкостью 8 Кбит, реализованной по 0,18- и 0,13-микронному КМОП-процессу. Видим, что при одном и том же напряжении питании значения отличаются в 7 раз.

Следовательно, нужны технологии уменьшения тока в режиме хранения для статических ОЗУ. В частности, можно предложить такие возможности.

- **Отключить неиспользуемые блоки памяти**. Различные функции памяти (например, кэш) бо́льшую часть времени не используют всю доступную емкость запоминающего устройства. Отключив неиспользуемые блоки от шин питания с помощью ключей с высоким порогом, мы уменьшим

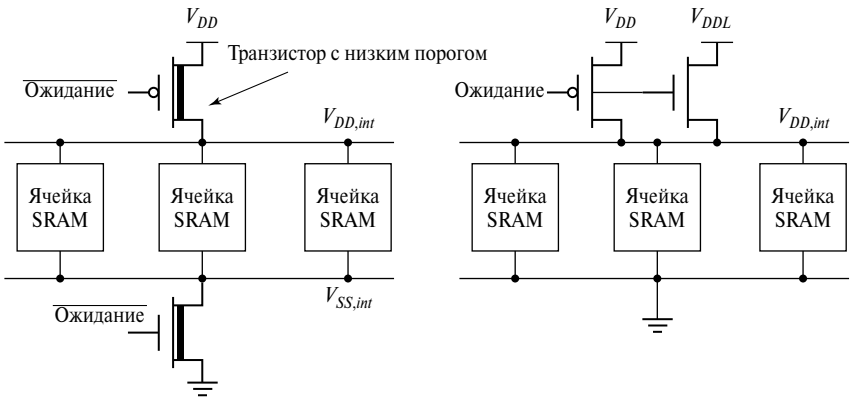


Рис. 12.72. Технологии подавления тока утечки для статических ОЗУ: *а* — введение дополнительного сопротивления; *б* — понижение напряжения питания (V_{DDL} может составлять 100 мВ)

их ток утечки до очень низких значений. Очевидно, что при таком подходе данные, хранимые в памяти, будут потеряны.

- **Повысить пороги, используя смещение подложки**. Отрицательное смещение неактивных ячеек повышает пороги устройств и уменьшает ток утечки.
- **Увеличить дополнительное сопротивление в канале утечки**. При необходимости хранения данных введение в канал утечки ключа с низким порогом позволяет уменьшить ток утечки, не повреждая данные (см. рис. 12.72, *а*). Низкопороговое устройство само по себе дает утечку, которой достаточно для поддержания состояния памяти. В то же время падение напряжения на ключе дает "эффект тяги" в присоединенных к нему ячейках памяти: снижение V_{GS} и отрицательное значение V_{BS} существенно понижает ток утечки.
- **Понизить напряжение питания**. На рис. 12.71 видно, что ток утечки сильно зависит от V_{DD}. Эффективным средством уменьшения утечки в режиме хранения является понижение напряжения шины питания до значения, ограничивающего утечку и обеспечивающего хранение данных (см. рис. 12.72, *б*). Минимальное допустимое напряжение хранения данных (напряжение, при котором хранимое значение еще поддерживается) определяется параметрами устройства. При стандартном 0,13-микронном КМОП-процессе для обеспечения хранения достаточно напряжение питания 100 мВ (без запаса помехоустойчивости). Вообще, одновременное уменьшение напряжения питания и тока утечки весьма облегчает решение проблемы рассеяния мощности в режиме хранения для статических ОЗУ.

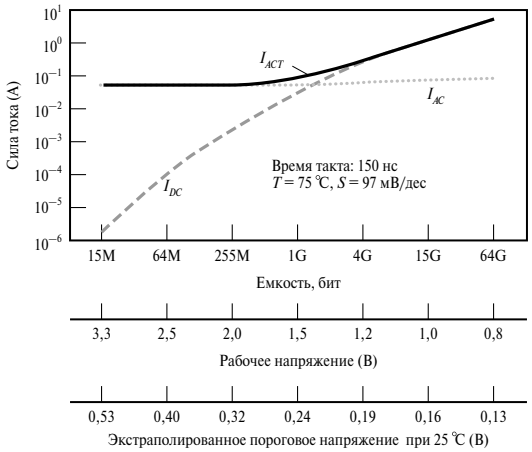


Рис. 12.73. Оценка распределения токов для последовательных поколений динамических ОЗУ (согласно [Itoh01])

Хранение данных в динамических ОЗУ

В динамических ОЗУ мощность в режиме хранения потребляется также из-за утечки. На этом, правда, сходство со статическими ОЗУ и заканчивается. Чтобы побороть утечку и потерю сигнала, динамические ОЗУ в режиме хранения данных необходимо постоянно обновлять. Операция обновления выполняется посредством считывания информации из m ячеек, соединенных с числовой шиной, и ее восстановления. Данная операция повторяется для всех n шин. Таким образом, мощность в режиме хранения пропорциональна заряду, рассеиваемому разрядной шиной, и *частоте обновления* (которая, в свою очередь, сильно зависит от интенсивности утечки). Отметим, что весьма эффективным средством удержания утечки в разумных пределах является поддержание низкой температуры pn-перехода.

Уменьшение размеров ячеек и заряда, хранимого в ячейке, а также уменьшение напряжений приводит к увеличению мощности динамических ОЗУ в режиме хранения. Сказанное иллюстрируется на рис. 12.73, где показана эволюция токов в активном режиме и режиме хранения для различных поколений динамических ОЗУ. Ток хранения данных догнал активный ток в поколении динамических ОЗУ емкостью 1 Гбит и с тех пор является доминирующим источником рассеяния мощности. Предполагается, что такое положение дел может изменить какая-нибудь новая эффективная техника подавления утечки.

Ключом к минимизации утечки в динамических ОЗУ является контроль над V_T. Это можно реализовать во время проектирования (зафиксировать V_T) или динамически (переменное V_T). Одна из возможностей уменьшения утечки через транзистор доступа в ячейке динамического ОЗУ заключается в аппаратном отключении устройства путем подачи в числовую шину неактивных ячеек отрицательного напряжения ($-\Delta V_{WL}$). Кроме того, мы можем

повысить на определенную величину ($+\Delta V_{BL}$) напряжение разрядной шина для неиспользуемых ячеек. Данный подход приводит к отрицательному напряжению затвор-исток и повышению порогового напряжения.

Технологии переменного V_T повышают пороги транзисторов доступа в неактивных ячейках посредством смещения карманов. Управление можно реализовать на разных уровнях (на уровне кристалла, модуля или отдельных транзисторов). При этом следует внимательно следить за генерированием напряжения карманов, поскольку это может привести к неустойчивости.

12.5.5. Резюме

С точки зрения мощности статические ОЗУ имеют очевидное преимущество перед динамическими. Впрочем, увеличивающееся влияние утечки, а также снижение напряжений питания могут несколько уменьшить разрыв между данными устройствами. В то же время, если основным фактором является мощность, потребляемая в стационарном режиме, наиболее жизнеспособной и привлекательной альтернативой являются энергонезависимые запоминающие устройства.

12.6. Разбор примеров запоминающих устройств

В предыдущих разделах мы ввели отдельные компоненты полупроводниковых запоминающих устройств. Теперь пришло время разобрать несколько примеров готовых приборов, чтобы понять, как эти элементы вписываются в общую картину. Для этого мы рассмотрим программируемую логическую матрицу и статическое ОЗУ емкостью 4 Мбит.

12.6.1. Программируемая логическая матрица (ПЛМ)

В нашем обсуждении стратегий реализации проектов в главе 8 вводилась концепция программируемой логической матрицы, ПЛМ (Programmable Logic Array — PLM). ПЛМ стала популярной в начале 1980-х как структурный подход к реализации нерегулярной логики.

Теперь перейдем к причине, по которой вопрос двухуровневой реализации логики поднимается снова в главе, посвященной запоминающим устройствам. При обсуждении сердечника постоянного запоминающего устройства мы отмечали, что матрицы ПЗУ — это набор элементов НЕ-ИЛИ (НЕ-И) с большим разветвлением по входу, который реализован в виде очень правильной структуры. Подобный образом, декодер адресов — это также набор элементов НЕ-ИЛИ (НЕ-И). Следовательно, один и тот же подход можно использовать для реализации любой двухуровневой функции. Единственное отличие между ПЗУ и ПЛМ заключается в том, что в декодере (плоскость И) перечислены все возможные конъюнктивные формы (m входов дает 2^m форм), а плоскость И ПЛМ реализует ограниченный набор конъюнктивных

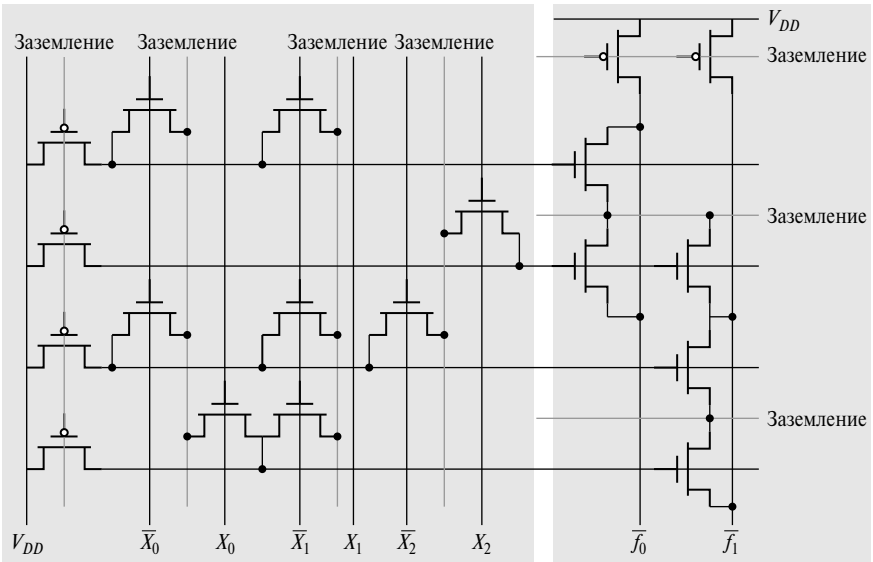


Рис. 12.74. Псевдо-n-МОП ПЛМ

форм, как следует из логических уравнений. С точки зрения же топологии названные структуры идентичны.

Реализация ПЛМ в виде структуры НЕ-ИЛИ/НЕ-ИЛИ с использованием псевдо-n-МОП-схем показана на рис. 12.74. Хотя данное устройство является компактным и быстрым, из-за его рассеяния мощности такая реализация непривлекательна для больших ПЛМ (здесь более удачным является динамический подход). Прямое каскадное соединение мы даже не будем рассматривать. Возможным решением является введение инвертора между плоскостями в стиле логики "домино" или реализация плоскости ИЛИ с помощью p-МОП-транзисторов и использование предварительной разрядки в стиле np-КМОП-схем. Первое решение приводит к проблемам с согласованием размеров, а второе замедляет структуру при использовании p-МОП-устройств минимального размера.

Рассмотрим подробнее вопрос тактирования запоминающего устройства. На рис. 12.75 показана более сложная схема тактирования, решающая проблему соединения плоскостей [Weste93]. Предположим, что нагрузочные транзисторы плоскости И и ИЛИ тактируются сигналами ϕ_{AND} и ϕ_{OR} соответственно. Данные сигналы определены так, что такт начинается с предварительной зарядки обеих плоскостей. По прошествии времени, достаточного для завершения предварительной зарядки плоскости И, плоскость И начинает оценивать сигнал (для чего повышается уровень ϕ_{AND}), а плоскость ИЛИ остается деактивизированной (рис. 12.76, *а*). Как только выходы плоскости И установятся, можно включать плоскость ИЛИ, повышая уровень сигнала ϕ_{OR}.

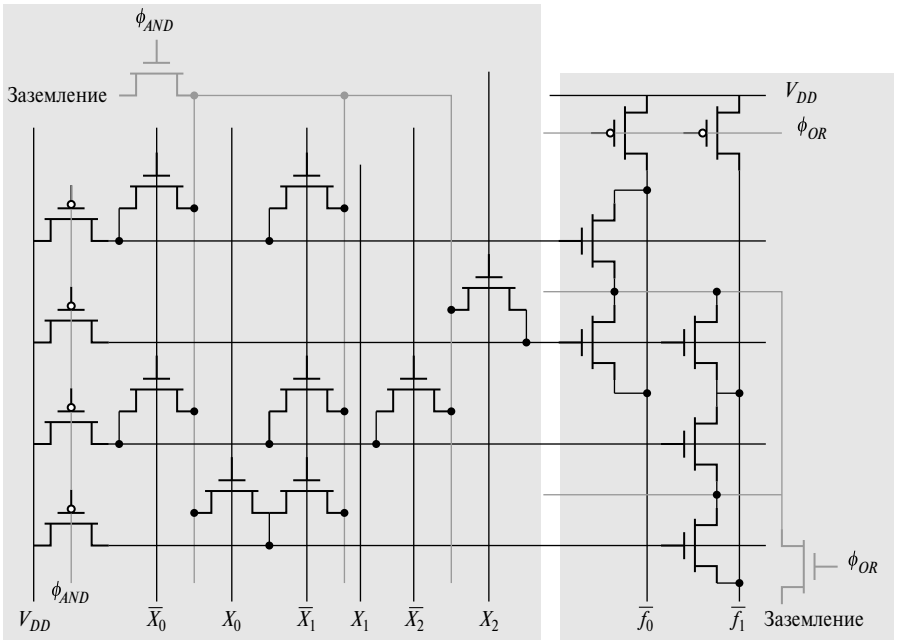


Рис. 12.75. Динамическая реализация ПЛМ

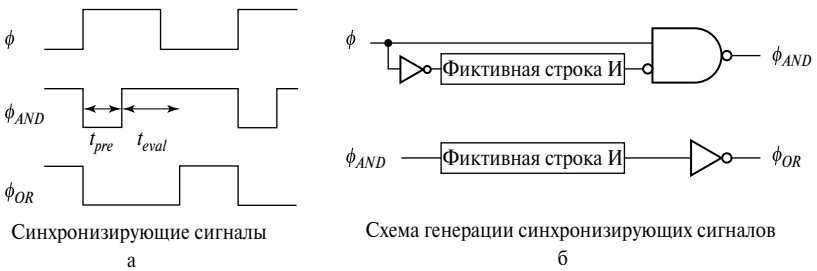


Рис. 12.76. Генерирование тактовых сигналов для динамической ПЛМ с автосинхронизацией. Время предварительной зарядки и оценки строки И в наихудшем случая равно t_{pre} и t_{eval} соответственно. Для получения такого времени задержки используются фиктивные строки И (полностью заполненные); в цепи НЕ-ИЛИ включены все транзисторы, кроме одного (такая ситуация представляет наихудший сценарий разрядки)

Синхронизация данных событий является ненадежной, поскольку зависит от размера ПЛМ и колебаний программных и технологических параметров. Хотя тактовые сигналы можно подводить извне, наиболее рекомендованным подходом является автосинхронизация, которая позволяет при минимальном риске достичь максимального быстродействия. На рис. 12.76, *б* показано, как по одному тактовому сигналу ϕ можно получить подходящие синхронизирующие сигналы. Тактовый сигнал ϕ_{AND} выводится из ϕ с ис-

пользованием моностабильного одновибратора. Элемент задержки состоит из фиктивной строки И с максимальной нагрузкой (полностью заполненной строки, дающей наиболее пессимистичную оценку требуемого времени предварительной зарядки). Сигнал ϕ_{OR} выводится аналогично с использованием фиктивной строки И, синхронизируемой по сигналу ϕ_{AND}. В результате получаем наиболее пессимистичную оценку времени разрядки плюс небольшой запас. Хотя такой подход требует введения дополнительных логических элементов, он гарантирует надежную работу матрицы в широком диапазоне рабочих условий.

12.6.2. Статическое ОЗУ емкостью 4 Мбит

На рис. 12.6 приводилась блок-схема статического ОЗУ емкостью 4 Мбит [Hirose90]. Данное устройство имеет время доступа 20 нс для напряжения питания 3,3 В и производится по 0,6-микронной КМОП-технологии. Память структурирована в 32 блока, причем каждый блок насчитывает 1024 строки и 128 столбцов. Адрес строки (X), адрес столбца (Y) и адрес блока (Z) состоят из 10, 7 и 5 бит соответственно.

Для обеспечения быстрого и маломощного декодирования строк в запоминающем устройстве использован интересный подход, называемый *схемой иерархического декодирования числовых шин* (рис. 12.77). Вместо транслирования декодированного адреса X всем блокам через поликремний, сигнал распространяется по металлу и называется *глобальной числовой шиной*. Локальная числовая шина ограничивается одним блоком и активизируется только тогда, когда этот блок выбирается с использованием адреса блока. Для дальнейшего повышения быстродействия и снижения потребляемой мощности вводится дополнительный уровень иерархии числовых шин — *субглобальная числовая шина*. В результате задержка декодирования строки составляет всего 7 нс.

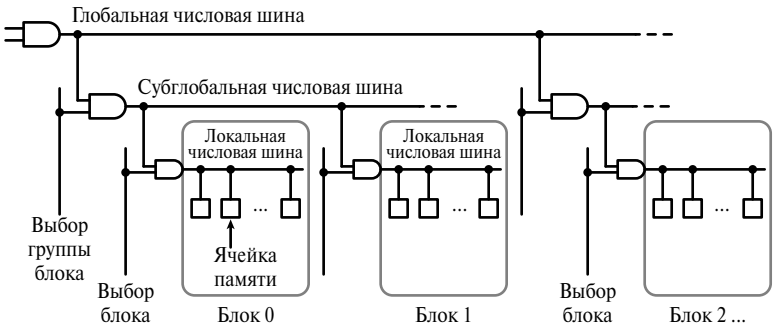


Рис. 12.77. Схема иерархического выбора числовой шины

На рис. 12.78 представлены периферийные схемы разрядной шины. Ячейка памяти представляет собой элемент активной нагрузки, состоящий из четырех транзисторов и занимающий площадь 19 мкм^2. Сопротивление нагрузки равно 10 ТОм. Импульс ATD вызывает предварительную зарядку и выравнивание разрядных шин выбранного блока с помощью контрольного сигнала BEQ. Обратите внимание на то, что нагрузка разрядных шин образуется статическими и динамическими устройствами. Понижение BEQ начинает процесс чтения. Активизируется одна из числовых шин и соответствующие разрядные шины начинают разряжаться. Эти шины соединятся с усилителями считывания после первого слоя декодеров столбцов (CD). В описанной схеме блоку из 128 столбцов требуется всего 16 усилителей считывания. Каждый усилитель состоит из двух каскадов (см. рис. 12.78, *б*). Первый — это каскад с перекрестными связями, который обеспечивает минимальное усиление и действует как схема сдвига уровня. Это позволяет второму усилителю (токовому зеркалу) работать в точке максимального усиления. Двухтактный выходной каскад управляет информационными шинами большой емкости, которые ведут к задающим генераторам ввода-вывода с тремя состояниями. Приблизительные формы соответствующих сигналов показаны на рис. 12.78, *в*. Видно, что для работы большого запоминающего устройства с высокой скоростью нужны тщательно продуманные стратегии тактирования и синхронизации. Для записи информации в память соответствующее значение и дополнительное к нему подаются в шину I/O и $\overline{I/O}$, после чего активизируются нужные адреса строки, столбца и блока. Кстати, стоит отметить, что запоминающее устройство потребляет всего 70 мА при тактовой частоте 40 МГц. Ток в режиме хранения равен 1,5 мА.

12.6.3. Флэш-память емкостью 1 Гбит на основе элементов НЕ-И

На рис. 12.79 показана блок-схема флэш-памяти емкостью 1 Гбит на основе элементов НЕ-И [Nakamura02]. Запоминающее устройство структурировано как два блока по 500 Мбит. В качестве базового модуля памяти (в основном из соображений, касающихся размера ячейки) была выбрана структура НЕ-И в стиле, подобном показанному на рис. 12.25, с 32 битами на блок. Каждая разрядная шина соединяется 1024 такими модулями. Блок емкостью 500 Мбит объединяет в себе 16 895 таких разрядных шин, соединенных параллельно. Числовые шины управляются с обеих сторон блока. Емкость страницы данного запоминающего устройства (т.е. число битов, которые можно считать/записать за один цикл) равна 2 Кбайт. Большой размер страницы позволяет получить высокую скорость программирования — 10 Мбайт/с. Для еще большего ускорения программирования разрядная шина дополняется ячейкой-"кэшем", которая позволяет считывать в память новые данные, пока предыдущая порция информации проходит этап записи/верификации.

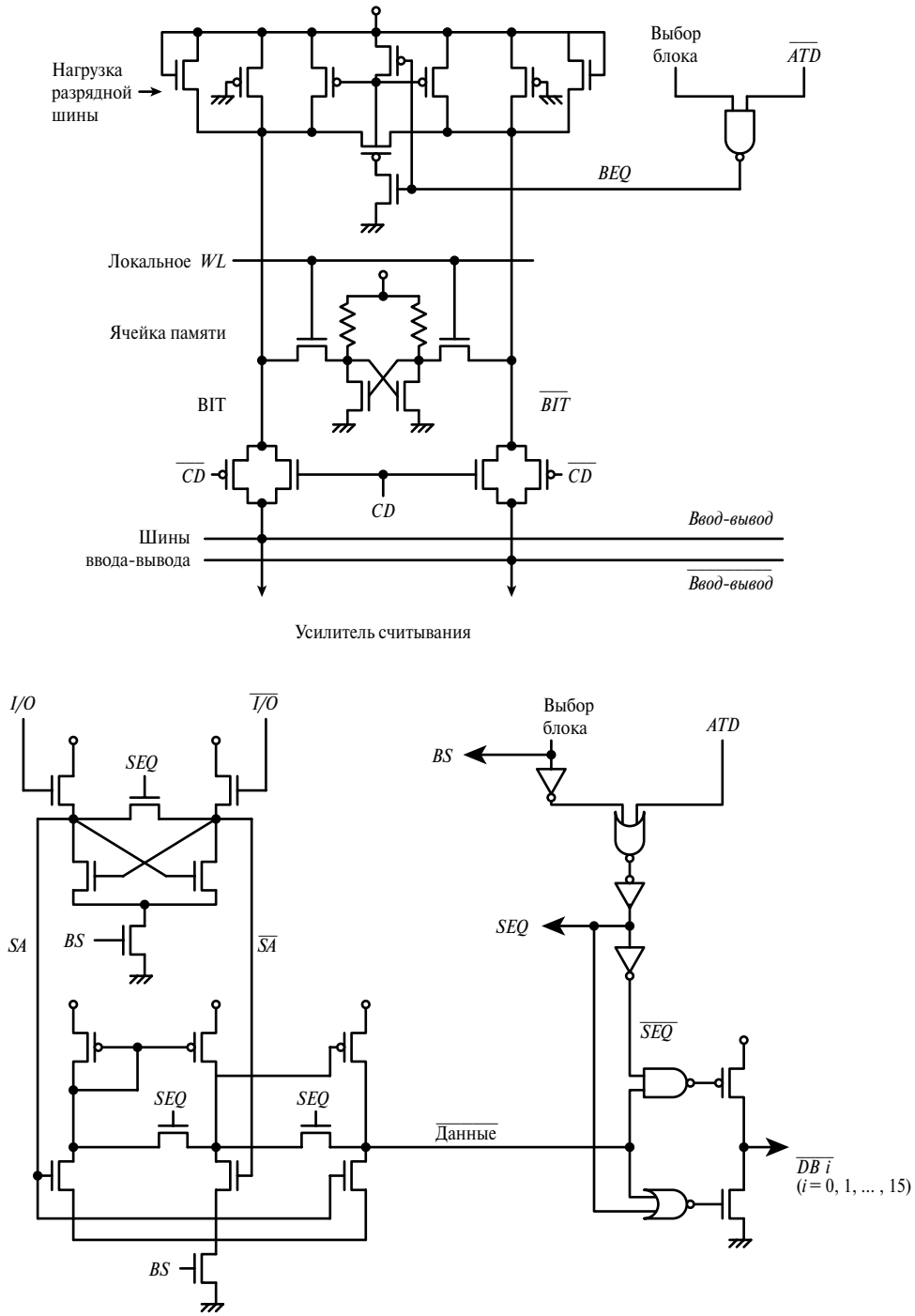


Рис. 12.78. Периферийные схемы разрядной шины и соответствующие сигналы

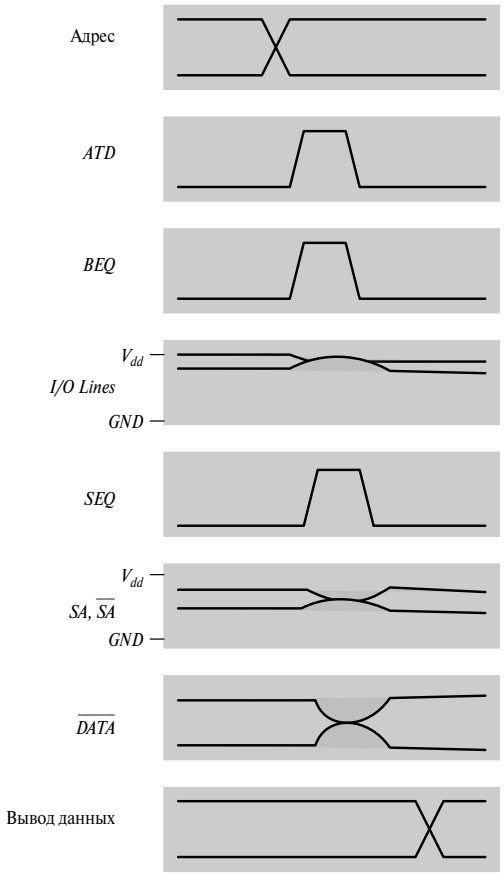


Рис. 12.78. (продолжение)

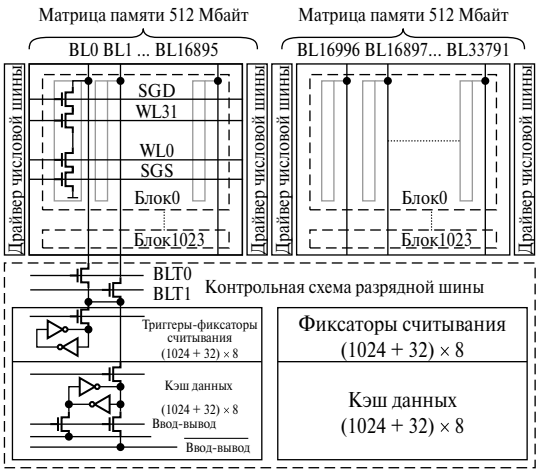


Рис. 12.79. Архитектура кристалла флэш-памяти емкостью 1 Гбит

На рис. 12.80 показан процесс стирания/записи флэш-памяти. При стирании все биты в одном блоке программируются как устройства истощения. В процессе записи порог выбранных устройств повышается с использованием цикла программирования/верификации. Чтобы все пороги были выше 0,8 В, требуется не меньше четырех таких циклов. Распределение порогов после четырех циклов показано на рис. 12.80, *б*.

Микрофотография модуля памяти и перечень основных характеристик устройства показаны на рис. 12.81. Память реализована по 0,13-микронной КМОП-технологии с тройным карманом с использованием одного поликремниевого, одного полисидного, одного вольфрамового и двух алюминиевых слоев.

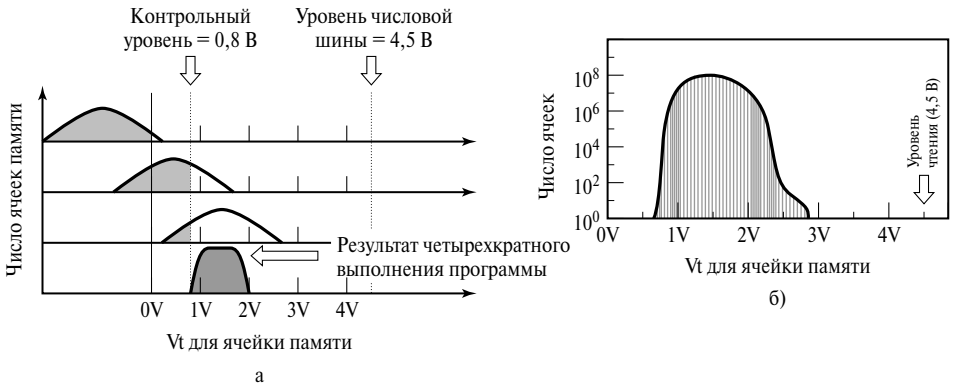
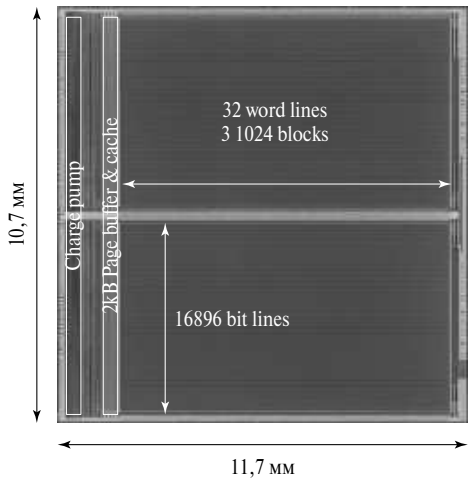


Рис. 12.80. Эволюция порогов ячейки в процессе записи (*а*) и конечное распределение порогов (*б*) после четырех циклов программирования



Характеристики	
Размер ячейки	0,077 мкм^2
Размер кристалла	125,2 мм^2 в 0,13 мкм КМОП
Структура	2112 × 8 байт×64 страницы ×блок 1 Кбайт
Энергопитание	2,7-3,6 В
Время такта	50 нс
Время чтения	25 мкс
Время программирования	200 мкс
Время стирания	2 мс/блок

Рис. 12.81. Микрофотография кристалла и характеристики флэш-памяти емкостью 1 Гбит [Nakamura02]

12.7. Перспективы: тенденции и эволюция в сфере полупроводниковых ЗУ

В завершение данной главы стоит обсудить некоторые эволюционные тенденции в сфере запоминающих устройств. В первых полупроводниковых запоминающих устройствах, представленных в 1960-х, использовалась биполярная технология, а сами устройства могли хранить не больше 100 бит. Основной технологический прорыв произошел в начале 1970-х, когда было продемонстрировано первое динамическое МОП-ОЗУ емкостью 1 Кбит [Hoff70]. Этот момент отметил начало быстрого развития целой области полупроводниковых разработок. На момент написания второго издания данной книги существовали запоминающие устройства емкостью 1 Гбит. Это означает, что за тридцать лет объем памяти, который можно интегрировать в один кристалл, увеличился на шесть порядков! В связи с этим справедливым будет высказывание, что число ячеек памяти, которые можно интегрировать в один кристалл, учетверяется примерно каждые три года (этот срок примерно соответствует частоте смены поколений запоминающих устройств). Схематический обзор тенденций в плане емкости запоминающих устройств различных стилей показан на рис. 12.82.

Изучая данную диаграмму, интересно понять, что порождает эти впечатляющие улучшения. На рис. 12.83 показана эволюция размеров ячеек для разных поколений запоминающих устройств. Ячейки динамических (DRAM) и статических (SRAM) ОЗУ миниатюризируются с коэффициентом 1/50 за каждые десять лет (или уменьшаются примерно в три раза за каждые три года). Кроме того, размер кристаллов увеличивается со скоростью 1,4 на каждое поколение. Объединяя эти два коэффициента, получим, что плотность

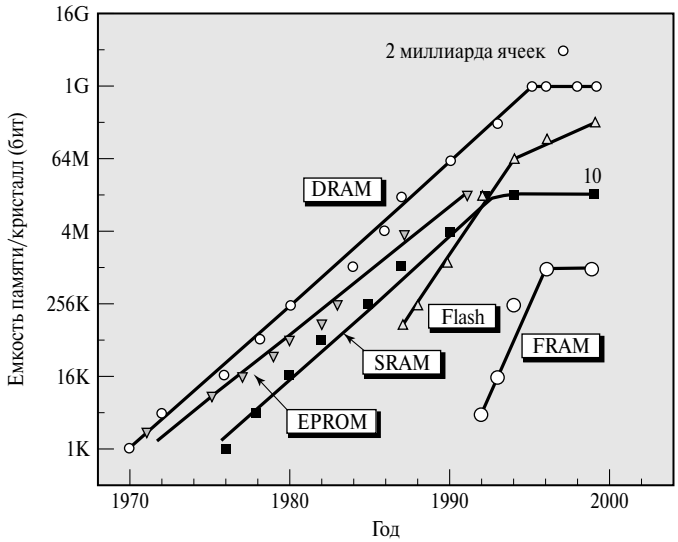


Рис. 12.82. Тенденции в сфере емкости запоминающих устройств [Itoh01]

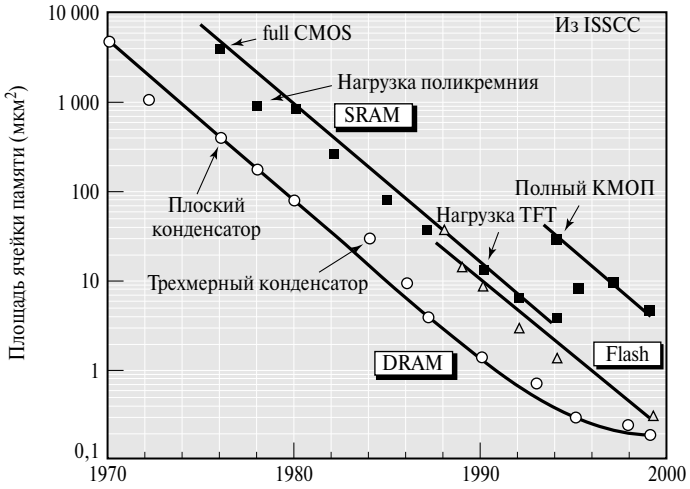


Рис. 12.83. Тенденции в сфере СБИС-ячеек памяти [Itoh01]

последовательных поколений запоминающих устройств отличается примерно в четыре раза. Возможно, вы считаете, что уменьшение размеров ячеек происходит только за счет масштабирования технологии, однако это справедливо лишь отчасти. Например, размер L_{eff} транзисторов, используемых в статическом ОЗУ емкостью 4 Кбит, равнялся 3 мкм и уменьшался от поколения к поколению, так что в статическом ОЗУ емкостью 64 Мбит применялись транзисторы с эффективной длиной 0,25 мкм. Отсюда следует, что за счет масштабирования размер ячейки уменьшился при смене поколений всего в $1,5 \times 1,5 = 2,25$ раза, что явно меньше, чем показывает существующая тенденция. Дополнительное уменьшение получается за счет еще большего усложнения технологии производства ячеек памяти. В этой области можно отметить такие важные события, как введение поликремниевых затворов, более плотная изоляция с использование технологии LOCOS (Local Oxidation of Silicon — технология МОП ИС с толстым защитным слоем оксида кремния) и U-канавок, поликремниевые нагрузочные резисторы для ячеек статических ОЗУ, а также трехмерные ячейки динамических ОЗУ. Если текущие тенденции в сфере интеграции будут сохраняться, то в данных областях потребуется введение еще более сложных технологий.

Это может быть не так и просто, поскольку в ближайшем будущем могут появиться новые проблемы и препятствия. Предпосылки к этому можно видеть на рис. 12.82, где представлено характерное насыщение емкости статических и динамических ОЗУ к концу 1990-х годов. Данное насыщение в основном объясняется крайней дороговизной и сложностью дальнейшей миниатюризации ячеек памяти. Впрочем, это не единственная преграда, существуют и другие соображения, которые нужно учитывать. Ниже мы кратко перечислим основные моменты, влияющие на дальнейшую интеграцию полупроводниковых запоминающих устройств.

- В сфере запоминающих устройств высокой емкости требуются специализированные технологии, сильно отличающиеся от процессов, существующих для логических устройств. Разработка таких технологий сопряжена с определенными финансовыми рисками.
- Поддерживать точную работу статических и динамических ОЗУ при пониженном напряжении питания и пороговом напряжении очень сложно. В частности, до сих пор нормально не решены проблемы утечки и нерегулярных ошибок. В сфере разработки ячеек памяти необходимы новые решения. Например, в качестве альтернативных вариантов можно рассмотреть энергонезависимые ферроэлектрические (FRAM) и магниторезистивные (MRAM) ОЗУ.
- Как отмечалось ранее, потребление мощности является одним из основных факторов, ограничивающих емкость памяти, которую можно интегрировать в один кристалл.
- Отметим еще более важный момент, связанный с изменениями бизнес-климата. Беспрецедентный спрос на память в основном направляется компьютерной индустрией. После смещения прицела основных производителей полупроводниковых устройств в сторону потребительских и мультимедийных приложений, а также сферы связи сформулировались новые требования к запоминающим устройствам. Например, мобильные полупроводниковые видео- и аудиопроигрыватели требуют энергонезависимой памяти. Кроме того, из соображений размера и цены в данных приложениях наиболее желательными являются интегрированные решения типа "систем на кристалле". Большинство запоминающих устройств необходимо встраивать в кристалл вместе с процессором и логическими компонентами. В результате требуются технологии запоминающих устройств, совместимые со стандартными логическими процессами.

12.8. Резюме

В этой главе мы подробно обсудили проектирование полупроводниковых запоминающих устройств. Хотя искусство проектирования ячеек памяти очень отличается от разработки традиционных логических схем, очевидно, что многие из проблем, с которыми сталкиваются разработчики запоминающих устройств, подобны проблемам, которые возникают у проектировщиков логических систем. Более того, с возникновением систем на кристалле встроенная память стала обязательным компонентом практически всех сложных разработок. Наконец, из-за того, что проектирование запоминающих устройств неразрывно связано с наиболее передовыми технологическими разработками (например, уменьшением размеров элемента), многие проблемы и решения, рассмотренные в данной сфере, могут со временем пригодиться в области проектирования логических схем. Например, одной из наиболее серьезных проблем, встающих перед разработчиками запоминающих устройств большой емкости, является рассеяние мощности. Одним из ее решений, реализо-

ванных в современных запоминающих устройствах, стала работа при низком (порядка 1,0 В или меньше) напряжении. Кроме того, можно использовать самосинхронизацию и локальное снижение энергопотребления. Большинство из этих технологий с равным успехом можно использовать и в мире логических схем.

В целом проблемы проектирования запоминающих устройств можно сформулировать следующим образом.

- *Архитектура запоминающего устройства* существенно влияет на легкость использования памяти, ее надежность, быстродействие и потребляемую мощность. Обычно запоминающие устройства выглядят как матрицы ячеек. Отдельная ячейка характеризуется адресом блока, столбца и строки.
- *Ячейку памяти* следует спроектировать так, чтобы при минимальной площади она давала максимальный сигнал. Хотя структура ячейки в основном определяется технологическими соображениями, хороший проект схемы помогает максимизировать сигнал и оптимизировать переходную характеристику. Топологии ячеек не сильно изменились за последние десятилетия. Большинство улучшений в плане повышения емкости запоминающего устройства связаны с масштабированием технологии и использованием улучшенных процессов производства. В главе были рассмотрены ячейки ПЗУ, энергонезависимой памяти и ОЗУ.
- Из-за слабого сигнала для надежной работы запоминающего устройства с разумным быстродействием необходимы *периферийные схемы*. Неотъемлемой частью любого запоминающего устройства являются декодеры, усилители считывания и буферы ввода-вывода. Соображения, приведенные в данной главе, представляют лишь поверхностный взгляд на текущее положение дел. Помимо описанных функций требуется еще самонагрузка, генерирование напряжения в ЭСППЗУ и стабилизация напряжения ([Itoh01]). Хотя данные темы являются весьма интересными и интригующими, их обсуждение чересчур увело бы нас в сторону от основной темы книги. Достаточно сказать, что проектирование периферийных схем запоминающего устройства по-прежнему требует от разработчика цифровых схем применения всех его способностей.
- Запоминающее устройство должно надежно работать во всем диапазоне рабочих условий. Чтобы повысить плотность интеграции, разработчики памяти жертвуют отношением сигнал/шум. В результате проект уязвим к целому ряду шумовых сигналов, которые обычно не порождают проблем в сфере проектирования логических элементов. Первое, что нужно сделать для повышения *надежности запоминающего устройства*, — это определить потенциальные источники сбоев и разработать подходящие средства их устранения. Например, типичными инструментами борьбы с потенциальными сбоями являются избыточность и коррекция ошибок.

12.9. Для любознательных

В ведущих журналах по проектированию интегральных схем публикуется множество статей по разработке запоминающих устройств. Основные передовые технологии представляются либо в *Journal of Solid State Circuits* (особенно в ноябрьских выпусках), материалах конференций *ISSCC* и симпозиумов по СБИС. Тем не менее до недавнего времени хороших книг по данной теме было не так и много. В последние годы было опубликовано несколько интересных базовых работ, некоторые из которых приведены ниже. Кроме того, к счастью, существует несколько очень хороших книг по полупроводниковым запоминающим устройствам. Если вас интересует тема проектирования ячеек памяти, в [Itoh01] и [Keeth01] вы найдете массу полезной информации.

12.10. Список использованной литературы

[Amrutur01] B.S. Amrutur, M.A. Horowitz, "Fast Low-Power Decoders for RAMs", *IEEE Journal of Solid-State Circuits*, October 2001, vol. 36, no. 10, p. 1506–1515.

[Baur95] M. Baur et al., "A Multilevel-Cell 32-Mb Flash Memory", *ISSCC Digest of Technical Papers*, February 1995, p. 13–33.

[Chang77] J. Chang, "Theory of MNOS Memory Transistor", *IEEE Trans. Electron Devices*, 1977, ED-24, p. 511–518.

[Crisp97] R. Crisp, "Direct Rambus Technology: The New Main Memory Standard", *IEEE Micro*, November/December 1997, p. 18–28.

[Dennard68] R. Dennard, "Field-effect transistor memory", U.S. Patent 3 387 286, June 1968.

[Dillinger88] T. Dillinger, *VLSI Engineering*, Chapter 12, Prentice Hall, 1988.

[Frohman74] D. Frohman, "FAMOS — A New Semiconductor Charge Storage Device", *Solid State Electronics*, 1974, vol. 17, p. 517–529.

[Gray01] P. Gray, P. Hurst, S. Lewis, and R. Meyer, *Analysis and Design of Analog Integrated Circuits*, 4th ed., John Wiley and Sons, 2001.

[Heller75] L. Heller et al., "High-Sensitivity Charge-Transfer Sense Amplifier", *Proceedings ISSCC Conf.*, 1975, p. 112-113.

[Hennessy02] J. Hennessy, D. Patterson, and D. Goldberg, *Computer Architecture — A Quantitavie Approach*, 2d ed, Morgan Kaufman Publishers, 2002.

[Hidaka92] H. Hidaka et al., "A 34-ns 16-Mb DRAM with Controllable Voltage-Down Converter", *IEEE Journal of Solid State Circuits*, July 1992, vol. 27, no. 7, p. 1020–1027.

[Hirose90] T. Hirose et al., "A 20-ns 4-Mb CMOS SRAM with Hierarchical Word Decoding Architecture", *IEEE Journal of Solid State Circuits*, October 1990, vol. 25, no. 5, p. 1068–1074.

[Hoff70] E. Hoff, "Silicon-Gate Dynamic MOS Crams 1024 Bits on a Chip", *Electronics*, 1970, August 3, p. 68–73.

[Itoh90] K. Itoh, "Trends in Megabit DRAM Circuit Design", *IEEE Journal of Solid State Circuits*, June 1990, vol. 25, no. 3, p. 778–798.

[Itoh01] K. Itoh, *VLSI Memory Chip Design*, Springer-Verlag, 2001.

[Johnson80] W. Johnson et al., "A 16 Kb Electrically Erasable Nonvolatile Memory", *ISSCC Digest of Technical Papers*, February 1980, p. 152-153.

[Kalter90] H. Kalter et al., "A 50-ns 16-Mb DRAM with a 10-ns Data Rate and On-Chip ECC", *IEEE Journal of Solid State Circuits*, October 1990, vol. 25, no. 5, p. 1118–1128.

[Keeth01] B. Keeth and R. Baker, *DRAM Circuit Design — A Tutorial*, IEEE Press, 2001.

[Lu89] N. Lu, "Advanced Cell Structures for Dynamic RAMs", *IEEE Circuits and Devices Magazine*, January 1989, p. 27–36.

[Mano87] T. Mano et al., "Circuit Technologies for 16-Mbit DRAMs", *ISSCC Digest of Technical Papers*, February 1987, p. 22-23.

[Masuoka91] F. Masuoka et al., "Reviews and Prospects of Non-Volatile Semiconductor Memories", *IEICE Transactions*, April 1991, vol. E74, no. 4, p. 868–874.

[May79] T. May and M. Woods, "Alpha-Particle-Induced Soft Errors in Dynamic Memories", *IEEE Transactions on Electron Devices*, January 1979, ED-26, no. 1, p. 2–9.

[Mot91] Motorola, "Memory Device Data", *Specification Data Book*, 1991.

[Nakamura02] H. Nakamura et al., "A 125-mm2 1-Gb NAND Flash Memory with 10-Mb/s Program Troughput", *ISSCC Digest of Technical Papers*, February 2002, p. 106-107.

[Ootani90] T. Ootani et al., "A 4-Mb CMOS SRAM with PMOS Thin-Film-Transistor Load Cell", *IEEE Journal of Solid State Circuits*, October 1990, vol. 25, no. 5, p. 1082–1092.

[Pashley89] R. Pashley and S. Lai, "Flash Memories: The Best of Two Worlds", *IEEE Spectrum*, December 1989, p. 30–33.

[Preston01] R.P. Preston, "Register Files and Caches", in A. Chandrakasan, W.J. Bowhill and F. Fox (eds.), *Design of High-Performance Microprocessor Circuits*, Piscataway, NJ: IEEE Press, 2001.

[Regitz70] W. Regitz and J. Karp, "A Three-Transistor Cell, 1,024-bit 500 ns MOS RAM", *ISSCC Digest of Technical Papers*, 1970, pp 42-43.

[Sasaki90] K. Sasaki et al., "A 23-ns 4-Mb CMOS SRAM with 0–2-mA Standby Current", *IEEE Journal of Solid State Circuits*, October 1990, vol. 25, no. 5, p. 1075–1081.

[Scheuerlein00] R. Scheuerlein et al., "A 10-ns Read-and-Write Non-Volatile Memory Array using a Magnetic Tunnel Junction and FET Switch in Each Cell", *ISSCC Digest of Technical Papers*, February 2000, p. 128-129.

[Sedra87] A. Sedra and K. Smith, *Microelectronic Circuits*, 2d ed., Holt, Rinehart and Winston, 1987.

[Snow67] E. Snow, "Fowler-Nordheim Tunneling in SiO2 Films", *Solid State Communications*, 1967, vol. 5, p. 813–815.

[Taguchi91] M. Taguchi et al., "A 40-ns 64-Mb DRAM with 64-b Parallel Data Bus Architecture", *IEEE Journal of Solid State Circuits*, November 1991, vol. 26, no. 11, p. 1493–1497.

[Takada91] M. Takada and T. Enomoto, "Reviews and Prospects of SRAM Technology", *IEICE Transactions*, April 1991, vol. E74, no. 4, p. 827–838.

[Tosaka97] Y. Tosaka et al., "Cosmic Ray Neutron-Induced Soft Errors in Sub-Half Micron CMOS Circuits", *IEEE Electron Device Letters*, March 1997, vol. 18, no. 3, p. 99–101.

[Weste93] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design*, 2d ed., Addison-Wesley, 1993.

[Yoo95] H. Yoo et al., "A 159-MHz 8-Banks 256-M Synchronous DRAM with Wave Pipelining Methods", *ISSCC Digest of Technical Papers*, 1995, p. 374-375.

12.10.1. Упражнения

Последнюю версию сборника задач и упражнений можно найти по адресу `http://bwrc.eecs.berkley.edu/IcBook`.

Вставка

3

Приемочные испытания произведенных схем

Производственные испытания

Проектирование с возможностью тестирования

Генерирование тестовых шаблонов

Введение

Хотя разработчики могут часами заниматься анализом, оптимизацией и совершенствованием топологии своих схем, один вопрос часто остается нерешенным: когда компонент поступит с фабрики, как убедиться, что он действительно работает? Отвечает ли он функциональным спецификациям, дает ли необходимое быстродействие? Потребитель ожидает, что компонент будет функционировать так, как описано в спецификации. После установки детали в систему будет довольно грустно (и дорого) выяснить, что она нерабочая. Чем позже выявлен сбой компонента, тем выше стоимость исправления ошибки. Например, замена компонента в проданном телевизоре означает замену всей платы плюс стоимость работы. Отправки неработающих или частично работающих устройств потребителю следует избегать всеми возможными средствами.

Правильное проектирование не гарантирует, что произведенный компонент будет правильно функционировать. В процессе производства возможно появление множества дефектов, вызванных либо недостатками исходного материала (например, примесями в кристалле кремния), либо колебаниями процесса производства (например, разрегулировкой производственного обо-

рудования). Кроме того, повреждения возможны в ходе нагрузочных испытаний, которые выполняются после процесса производства. В ходе данных тестов деталь подвергается воздействию температурных и механических нагрузок, чтобы проверить, будет ли она работать в широком диапазоне рабочих условий. Типичными дефектами, выявляемыми в ходе этих испытаний, являются короткое замыкание между проводниками или слоями, а также разрывы связей.

Проверить, что готовая деталь правильно работает при всех входных условиях, не так просто, как кажется на первый взгляд. Анализируя поведение схемы в фазе проектирования, разработчик располагает неограниченным доступом ко всем узлам в цепи. Он может подавать на входы любые комбинации сигналов и наблюдать за получающейся характеристикой на любом узле структуры. Когда устройство произведено, такая свобода действий невозможна. Теперь доступ к схеме возможен только через выводы входов-выходов. Сложный компонент (например, микропроцессор) составлен из сотен миллионов транзисторов и может находиться в невероятно огромном числе возможных состояний. Перевести такой компонент в какое-то конкретное состояние и исследовать характеристику схемы очень сложно (если вообще возможно). Кроме того, оборудование для тестирования устройств довольно дорого, и каждая секунда, которую устройство проводит на испытательном стенде, повышает цену этого устройства.

Таким образом, желательно учитывать тестирование на ранних этапах процесса проектирования. Вводя в схему небольшие модификации, можно предусмотреть возможность проверки на отсутствие дефектов. Данный подход к проектированию получил название *проектирования с возможностью тестирования*, или *контролепригодности* (*Design For Testability* — DFT). Хотя многие разработчики схемы с презрением относятся к данной методологии, предпочитая концентрировать внимание на более впечатляющих аспектах проектирования (транзисторной оптимизации), она является неотъемлемой частью процесса проектирования, и реализовать ее следует как можно раньше. "*Если вы это не протестируете, работать оно не будет! (Гарантированно.)*" [Weste93]. Стратегия DFT имеет две составляющих.

1. Обеспечение необходимых *цепей*, позволяющих быстро и комплексно протестировать устройство.
2. Обеспечение необходимых *тестовых шаблонов* (векторов возбуждения), применяемых в ходе тестирования. С точки зрения цены желательно, чтобы тестовый шаблон был как можно короче, но при этом обеспечивал максимальный охват возможных дефектов.

В последующих разделах мы кратко опишем некоторые важнейшие вопросы, возникающие в обеих названных областях. Перед этим мы рассмотрим типичную процедуру тестирования.

Процедура тестирования

В зависимости от поставленной цели производственные испытания можно отнести к одной из указанных ниже категорий.

- **Диагностическое испытание** используется при отладке кристалла или печатной платы. В ходе этого текста разработчик по дефектной детали пытается найти узел, являющийся причиной сбоев.
- В ходе **функциональной проверки** (теста "*прошло/не прошло*") определяется, является ли (или не является) произведенный компонент функциональным. Данный процесс проще диагностического испытания, поскольку требует ответа "да" или "нет". Поскольку данной проверке необходимо подвергнуть любой произведенный кристалл (а это влияет на его цену), она должна быть максимально простой и быстрой.
- При **параметрическом тесте** проверяется несколько недискретных параметров (запас помехоустойчивости, задержки распространения, максимальная тактовая частота) при различных рабочих условиях (переменная температура и напряжение питания). Параметрические тесты обычно делят на статические (при постоянном токе) и динамические (при переменном токе).

Типичные производственные испытания происходят следующим образом. В тестирующее устройство загружаются предопределенные тестовые шаблоны, после чего активизируется тестируемое устройство и собираются соответствующие отклики. Тестовые шаблоны определяются в *программе испытаний*, описывающей подаваемые на входы сигналы, уровни напряжения, тактовую частоту и ожидаемую реакцию. Для соединения выходов и входов тестирующего устройства с соответствующими выводами кристалла или корпуса требуется зондовая плата, или плата тестируемого устройства.

Новая часть автоматически подается в тестирующее устройство, которое выполняет программу испытаний, подавая на вход тестируемого устройства ряд шаблонов и сравнивая получаемую характеристику с ожидаемой. При наличии расхождений часть помечается как бракованная (например, с помощью пятнышка чернил) и зонды автоматически переходят на следующий кристалл пластины. В процессе гравировки, когда пластина разбивается на отдельные кристаллы, отмеченные части будут автоматически отброшены. Если проверке подвергается часть в корпусе, то после завершения испытания она будет вынута из тестовой платы и, в зависимости от результата, перенесена в ящики для годных и бракованных деталей. Вся процедура занимает несколько секунд на одну часть, что позволяет одному тестирующему устройству проверять за часть несколько тысяч деталей.

На рис. З.1 показано современное устройство для тестирования логических схем. Автоматические тестирующие устройства являются довольно дорогими. Ситуация усложняется всевозрастающими требованиями к быстродействию, устанавливаемыми современными интегральными схемами, из-за чего цена оборудования для тестирования растет невероятными темпа-

Рис. 3.1. Автоматическое тестирующее устройство

ми. Наиболее эффективным средством снижения стоимости тестирования является уменьшение времени, которое кристалл проводит на испытательном стенде. К сожалению, с увеличением сложности интегральных схем наблюдается обратная тенденция. Следовательно, весьма желательны подходы к проектированию, которые позволили бы уменьшить расходы на тестирование.

Проектирование с возможностью тестирования

Проблемы проектирования с возможностью тестирования

Как отмечалось ранее, быстродействующее тестирующее устройство, позволяющее проверять современные компоненты, имеет астрономическую цену. Увеличить пропускную способность этого устройства (и стоимость тестирования в целом) можно за счет уменьшения времени испытания одного компонента. Изучая возможность тестирования на ранних этапах процесса проектирования, можно упростить весь процесс верификации. Некоторые подходы, позволяющие добиться этой цели, мы опишем в данном разделе. Правда, перед подробным изучением этих техник необходимо вначале понять некоторые сложность проблемы тестирования.

Рассмотрим комбинационную схему, изображенную на рис. З.2, *а*. Правильность цепи можно проверить путем подачи на вход всех возможных комбинаций сигналов и изучения получаемых откликов. Для схемы с N входами это означает применение 2^N входных шаблонов. При $N = 20$ потребуется более одного миллиона шаблонов. Если применение и изучение одного шаблона занимает 1 мкс, полное тестирование модуля потребует 1 с. При изучении последовательного модуля, изображенного на рис. З.2, *б*, ситуация становится более драматичной. Выход схемы зависит не только от поданных входов, но и от состояния. Следовательно, полное тестирование данного конечного автомата требует применения 2^{N+M} входных шаблонов, где M — число регистров состояний [Williams83], [Weste93]. Для конечного автомата среднего размера (например, $M = 10$) это означает необходимость оценки 1 миллиарда шаблонов, что потребует 16 минут на нашем оборудовании с пропускной

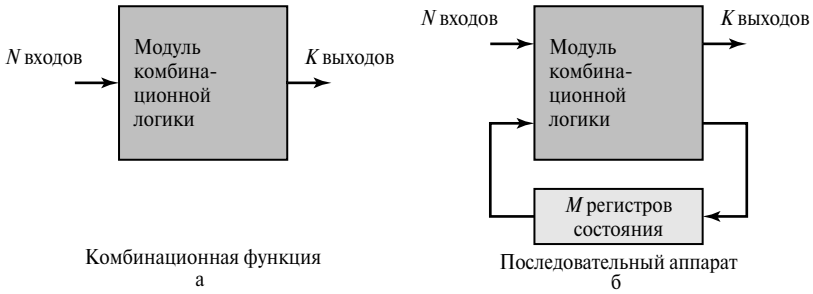


Рис. 3.2. Тестирование комбинационных и последовательных устройств

способностью 1 мкс/шаблон. Если моделировать современный микропроцессор как конечный автомат, эквивалентная модель будет состоять более чем из 50 регистров состояний. Полное тестирование подобного устройства потребует более миллиарда лет!

Очевидно, нужен альтернативный подход, основанный на следующих предпосылках.

- Полное перечисление всех возможных входных шаблонов является чересчур *избыточным*, т.е. один сбой в схеме проявляется при нескольких входных наборах сигналов. Чтобы обнаружить такой сбой, достаточно одного из этих шаблонов, остальные являются ненужными.
- Число шаблонов можно существенно уменьшить, если не требовать детектирования всех сбоев. Например, обнаружение последнего процента сбоев может требовать чрезмерного числа дополнительных шаблонов, и стоимость реализации этой процедуры может оказаться выше цены замены бракованного элемента. Типичные тестовые процедуры обеспечивают покрытие возможных неисправностей на уровне 95–99%.

Избавляясь от избыточности и обеспечивая меньшее покрытие возможных неисправностей, большинство блоков комбинационной логики можно протестировать с помощью ограниченного набора входных векторов. Впрочем, это не решает проблемы последовательных схем. Для тестирования конкретного дефекта в конечном автомате недостаточно просто подать на вход нужные сигналы; предварительно нужно перевести устройство в нужное состояние. Это требует применения последовательности входных шаблонов. Кроме того, передача реакции схемы на эти шаблоны к одному из выходных выводов может требовать применения другой последовательности шаблонов. Другими словами, тестирование конкретной ошибки в конечном автомате требует последовательности векторов. Разумеется, это повлечет за собой удорожание процесса тестирования.

Один из способов решения данной проблемы заключается в преобразовании последовательной цепи в комбинационную путем разрыва в процессе тестирования контура обратной связи. Данный принцип является одной из основ описанной ниже методологии *скан-теста*. Возможен и другой под-

ход — позволить схеме протестировать себя самостоятельно. Такая проверка не требует внешних векторов и может выполняться довольно быстро. Концепция *самоконтроля* будет подробно рассмотрена ниже. Вообще, при изучении возможности тестирования схем наибольшую важность представляют два свойства.

1. **Контролируемость**, измеряющая легкость перевода узла схемы в данное состояние с помощью только входных выводов. Узел называется легкоконтролируемым, если его можно перевести в любое состояние с помощью одного вектора входов. Если узел (или схема) является слабоконтролируемым, перевод в желаемое состояние может требовать длинной последовательности векторов. Понятно, что в проектах, допускающих возможность тестирования, узлы должны контролироваться как можно легче.
2. **Наблюдаемость**, измеряющая легкость наблюдения значения узла не выходных выводах. Прежде чем состояние узла с низкой наблюдаемостью будет доступно на выходе, пройдет несколько тактов. Учитывая сложность схемы и ограниченное число выходных выводов, тестируемая схема должна обладать высокой наблюдаемостью. Методы, позволяющие этого добиться, рассмотрены в следующих разделах.

Комбинационные схемы характеризуются хорошей наблюдаемостью и контролируемостью, поскольку любой узел можно проконтролировать и изучить в пределах одного такта.

Методы *проектирования с возможностью тестирования* можно разбить на три категории: специальные (ad hoc), основанные на сканировании и подходы с самоконтролем.

Специальное тестирование

Как видно из названия, специальное тестирование — это набор трюков и методов, которые можно использовать для увеличения наблюдаемости и контролируемости проекта и которые обычно применяются исходя из специфики конкретного приложения.

Пример подобной техники приведен на рис. З.3, *а*, где изображен простой процессор с памятью для хранения данных. При обычной конфигурации память доступна только через процессор. Чтение (запись) данных из одной секции памяти требует нескольких тактов. Добавив мультиплексоры к шине данных и адресной шине, контролируемость и наблюдаемость памяти можно существенно улучшить (см. рис. З.3, *б*). В нормальном режиме работы данные селекторы связывают порты запоминающего устройства с процессором. В процессе проверки порты данных и адреса соединяются непосредственно с выводами входов-выходов и тестирование памяти может выполняться эффективнее. В данном примере иллюстрируются некоторые важные концепции проектирования с возможностью тестирования.

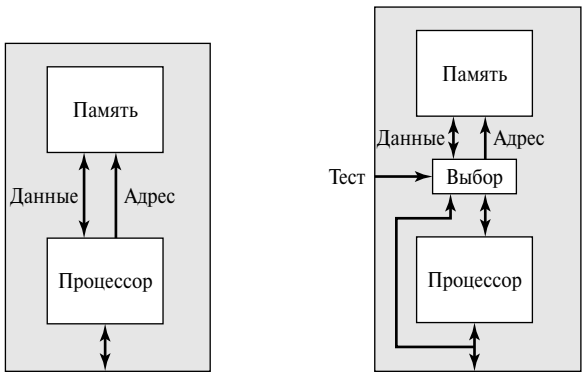


Рис. 3.3. Улучшение возможности тестирования с помощью мультиплексоров

- Довольно часто выгодно ввести *дополнительное аппаратное обеспечение*, никак не влияющее на функциональные возможности, но улучшающее возможности тестирования. Обычно разработчики готовы немного пожертвовать площадью и быстродействием, если в результате у проекта существенно улучшится наблюдаемость и контролируемость.
- Проектирование удоботестируемых схем часто означает введение дополнительных выводов входов-выходов, помимо обычных функциональных выводов входов-выходов. На рис. З.3, *б* таким дополнительным выводом является порт *тест*. Для уменьшения числа дополнительных контактных площадок тестовые сигналы и функциональные можно уплотнять и передавать через одни и те же площадки. Например, шина ввода-вывода на рис. З.3, *б* используется как шина данных в нормальном режиме работы, а в процессе тестирования через нее передаются тестовые шаблоны.

В настоящее время существует большой набор специальных техник тестирования: секционирование больших конечных автоматов, добавление контрольных точек, создание нулевых состояний, введение тестовых шин и т.д. Хотя большинство из этих техник довольно эффективны, их применимость зависит от конкретного приложения и использованной архитектуры. Целесообразность использования названных решений в конкретных проектах определяется на основе множества факторов (в первую очередь — опыта разработчика), и ее довольно сложно автоматизировать. Поэтому на практике более желательны структурированные и автоматизируемые подходы.

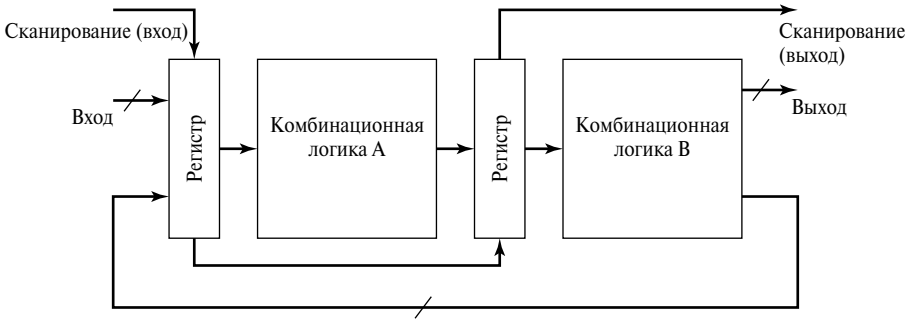


Рис. 3.4. Тест с последовательным сканированием

Тесты, основанные на сканировании

Один из способов решения проблемы тестирования последовательной схемы заключается в преобразовании всех регистров в элементы, загружаемые и считываемые извне. В таком случае тестируемая схема превращается в комбинационную. Для управления узлом создается подходящий вектор, который затем загружается в регистры и проходит через логические цепи. Результат возбуждения доходит до регистров и фиксируется, после чего передается во внешний мир. К сожалению, соединение всех регистров структуры с тестовой шиной неизбежно приводит к неприемлемым служебным издержкам. Гораздо более элегантный подход предлагает схема последовательного сканирования, показанная на рис. З.4.

Регистры были модифицированы так, чтобы поддерживать два режима работы. В нормальном режиме они действуют как N-битовые тактируемые регистры. В тестовом режиме регистры последовательно соединяются в один последовательный сдвиговый регистр. Теперь процедура тестирования происходит следующим образом.

1. Вектор возбуждения для логического модуля А (и/или В) вводится через вывод *ScanIn* и сдвигается в регистры, управляемые генератором тестовых тактовых импульсов.
2. Возбуждение применяется к входам логических блоков и проходит до их выходов. Результат фиксируется в регистрах.
3. Результат сдвигается из схемы через вывод *ScanOut* и сравнивается с ожидаемыми данными. В это же время на вход можно подавать следующий вектор возбуждения.

Данный подход вводит минимальные дополнительные издержки. Последовательная природа цепочки сканирования снижает общие затраты на трассировку. Традиционные регистры можно довольно легко модифицировать, добившись поддержки технологии сканирования, как показано на рис. З.5, где изображен 4-битовый регистр, дополненный цепочкой сканирования. Единственное дополнение — еще один мультиплексор на входе. При низком

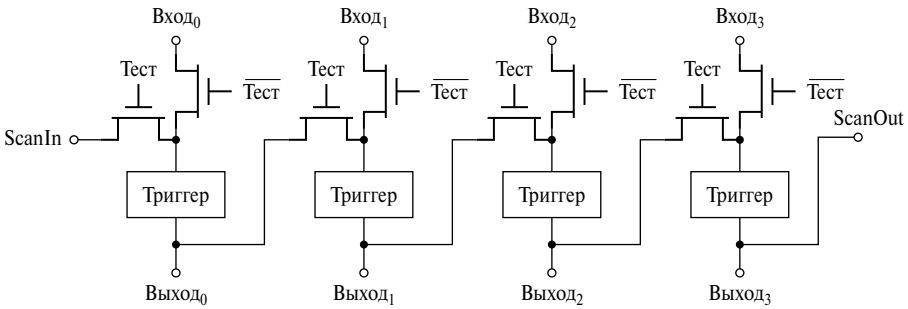


Рис. З.5. Регистры, дополненные цепочкой последовательного сканирования

уровне сигнала *Тест* схема находится в нормальном режиме работы. Повышение уровня сигнала *Тест* приводит к выбору входа *ScanIn* и соединяет регистры в цепочку сканирования. Выход регистра *Out* соединяется с логикой разветвления по выходу и дублируется на выводе *ScanOut*, соединяемом с входом *ScanIn* соседнего регистра. Дополнительные издержки (площадь и быстродействие) можно ограничить до уровня менее 5%.

Задача З.1. Проектирование сканирующего регистра

Модифицируйте статический двухфазный регистр "master-slave", изображенный на рис. 7.10, реализовав поддержку последовательного сканирования.

На рис. З.6 показана временная последовательность событий, которую можно реализовать в схеме, изображенной на рис. З.4, при использовании двухфазного тактирования. Допустим, что цепочка состоит из N регистров. При повышении уровня сигнала *Тест* генерируется N тактовых импульсов и регистры загружаются. Далее уровень сигнала *Тест* понижается и генерируется общая тактовая последовательность, фиксирующая состояние комбинационной логики в регистрах. Наконец, N дополнительных импульсов (при Тест = 1) передают полученный результат на выход. Еще раз обратите внимание на то, что подача результатов сканирования на выход может перекрываться с вводом следующего вектора.

Существует множество вариаций метода последовательного сканирования. Одна из наиболее популярных (кстати, самая первая из существующих) была введена IBM и называется *методом сканирования с опросом состояний* (Level-Sensitive Scan Design — LSSD) [Eichelberger78]. Основной составляющей подхода LSSD является *защелка сдвигового регистра* (Shift-Register Latch — SRL), показанная на рис. З.7. Она состоит из двух защелок, $L1$ и $L2$, причем $L2$ используется только при тестировании. При нормальной работе схемы сигналы D, Q ($\bar{Q}$) и C используются как вход защелки, ее выход и тактовый сигнал соответственно. В этом режиме тактовые сигналы тестового

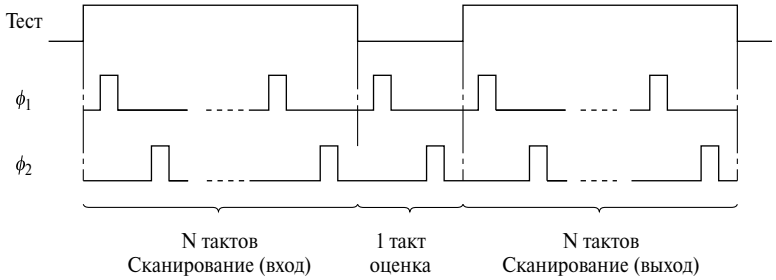


Рис. 3.6. Временная диаграмма тестовой последовательности. N представляет число регистров в тестовой цепочке

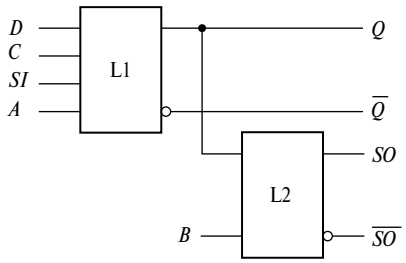


Рис. 3.7. Защелка сдвигового регистра

режима A и B находится на низком уровне. В режиме сканирования входом и выходом являются SI и SO. Тактовый сигнал имеет низкий уровень C, а тактовые сигналы A и B действуют как неперекрывающиеся двухфазные таковые сигналы тестового режима.

Метод LSSD представляет не только стратегию тестирования, но и целую философию синхронизации. Строго следуя правилам, вводимым данной методологией, можно автоматизировать значительную часть процесса генерации тестовых кодов и верификации временны*ы*х параметров. Именно поэтому в IBM долгое время использовали только метод LSSD. Основным недостатком этого подхода является сложность защелки.

Отметим, что делать все регистры проекта тестируемыми часто бывает не нужно. Рассмотрим, например, конвейерный информационный канал, изображенный на рис. З.8. В данном случае регистры конвейера используются только для повышения быстродействия и не дают никакого вклада в состояние схемы. Следовательно, сканируемыми достаточно сделать только входные и выходной регистры. При генерировании тестовых последовательностей сумматор и компаратор можно рассматривать как единый комбинационный блок. Единственное отличие заключается в том, что в процессе тестирования требуется два такта для распространения эффектов вектора возбуждения на выходной регистр. Такой подход называется *частичным сканированием*, и он часто реализуется, когда важно быстродействие схемы. Недостаток этого подхода заключается в том, что не всегда можно быстро определить, какие регистры достаточно сделать сканируемыми.

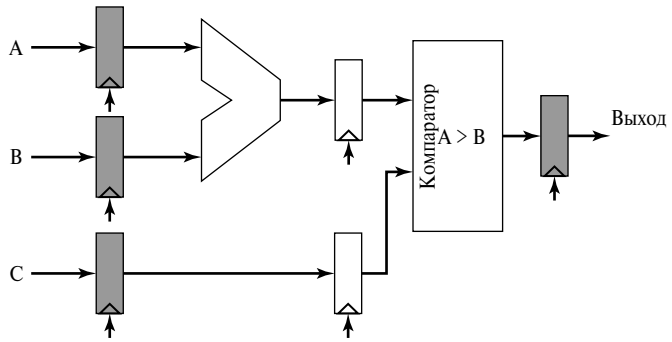


Рис. 3.8. Конвейерный информационный тракт с частичным сканированием. В цепочку включены только закрашенные регистры

Релизация периферийного сканирования

До недавнего времени в сфере интегральных схем наиболее сложной была именно проблема тестирования. Измерительные монтажные платы совершенствовались за счет большого количества контрольных точек. Благодаря штырьковому монтажу на задней стороне платы присутствовал каждый вывод корпуса. Для тестирования было достаточно поместить плату в набор измерительных приставок (“пробники” или “ложе гвоздей”), подать интересующие сигналы и изучить отклик. Такая картина полностью изменилась с появлением более совершенных технологий корпусирования, например технологии поверхностного монтажа или многокристальных модулей (см. главу 2). Теперь реализация контролируемости и наблюдаемости уже не настолько проста, поскольку число контрольных точек существенно уменьшилось. Впрочем, данную проблему можно решить, применив принцип тестирования путем сканирования на уровне компонентов и плат.

Получаемый в результате подход, названный *периферийным сканированием* (boundary scan), был стандартизован для обеспечения совместимости между различными поставщиками ([IEEE1149]). По сути, принцип его действия заключается в соединении выводов входов-выходов компонентов на плате в одну последовательную цепочку сканирования, как показано на рис. З.9. В нормальном режиме работы площадки периферийного сканирования действуют как обычные устройства ввода-вывода. В режиме тестирования векторы можно исследовать (сканировать) на входах и выходах площадок, что обеспечивает контролируемость и наблюдаемость на границах компонентов (отсюда и название). Процесс тестирования происходит по схеме, описанной в предыдущем абзаце. Различные режимы управления позволяют тестировать отдельные компоненты, а также соединительную плату. Служебные издержки, связанные с реализацией описанного подхода, включают несколько более сложные площадки ввода-вывода и дополнительный внутрикристальный тестовый контроллер (конечный автомат с 16 состояниями). В настоящее время периферийное сканирование реализовано в большинстве распространенных компонентов.

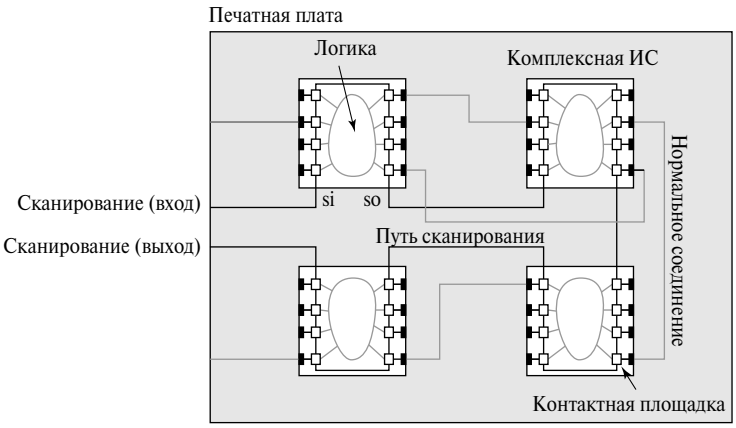


Рис. 3.9. Периферийное сканирование

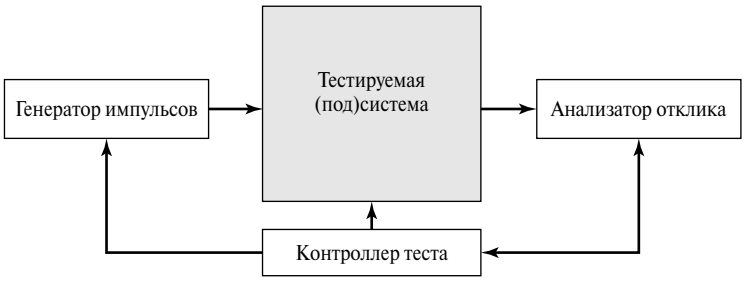


Рис. 3.10. Общий формат встроенной структуры самопроверки

Встроенный самоконтроль (Built-In Self-Test — BIST)

Альтернативным и весьма привлекательным подходом к реализации возможности тестирования является генерирование тестовых сигналов внутри самой схемы без применения внешних сигналов [Wang86]. Еще лучше, если сама схема сможет определять, правильны ли результаты, полученные в ходе такого тестирования. В зависимости от природы схемы реализация такой идеи может требовать дополнительных цепей, отвечающих за генерирование и анализ тестовых сигналов. Возможно, часть необходимых цепей уже реализована в схеме, и их можно будет задействовать в тестировании, поэтому общее увеличение площади схемы будет минимальным.

Общий формат встроенной структуры самопроверки иллюстрируется на рис. 3.10 [Kornegay92]. Структура содержит средства подачи тестовых сигналов на тестируемое устройство и средства для сравнения отклика устройства с известной правильной последовательностью.

Существует множество способов генерирования входных сигналов. Наиболее распространенными являются *исчерпывающий* и *случайный* подходы. При исчерпывающем подходе длина теста составляет 2^N, где N — число входов схемы. Исчерпывающая природа теста означает, что будут обнаружены

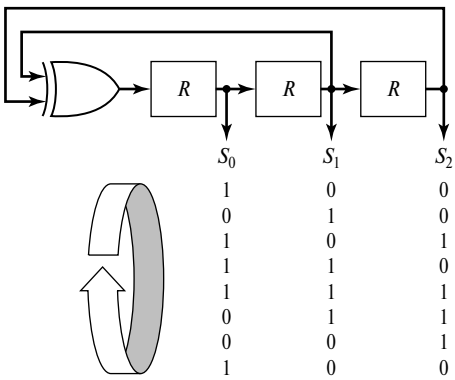


Рис. 3.11. Трехбитовый сдвиговый регистр с линейной обратной связью и генерируемая им последовательность

все сбои, которые можно обнаружить. Хорошим примером исчерпывающего генератора сигналов является N-битовый счетчик. Для схем с большими значениями N время полного цикла проверки может быть довольно большим. Поэтому часто используется альтернативный вариант — проверка с применением случайных кодов, подразумевающая применение случайным образом выбираемого подмножества из 2^N возможных наборов входных сигналов. Данное подмножество следует выбирать так, чтобы получить разумное покрытие неисправностей. Примером псевдослучайного генератора входных наборов является *сдвиговый регистр с линейной обратной связью* (Linear-Feedback Shift Register — LFSR), показанный на рис. З.11. Он состоит из последовательно соединенных однобитовых регистров. К некоторым выходам применяется операция ИСКЛЮЧАЮЩЕЕ ИЛИ, после чего они подаются обратно на вход сдвигового регистра. Перед повтором последовательности N-битовый регистр походит через $2^N - 1$ состояний, поэтому можно говорить, что он генерирует псевдослучайную последовательность. Инициализация регистров с данным начальным числом определяет, какая последовательность входных сигналов будет генерироваться далее.

Анализатор отклика схемы можно реализовать на основе сравнения сгенерированного отклика с ожидаемой реакцией, хранимой во внутрикристальной памяти, однако такой подход чреват слишком большими служебными издержками и поэтому непрактичен. Гораздо более дешевой техникой является сжатие откликов перед их сравнением. Хранение сжатого отклика правильной схемы требует минимального объема памяти, особенно при большой степени сжатия. В таком случае анализатор отклика будет состоять из цепи, динамически сжимающей выход тестируемой схемы и компаратора. Сжатый выход часто называют *сигнатуторой* (буквально — "подпись") схемы, а весь подход получил название *сигнатурного анализа*.

Пример сигнатурного анализатора, сжимающего однобитовый поток данных, показан на рис. З.12. Видим, что данная схема просто подсчитывает число переходов $0 \rightarrow 1$ и $1 \rightarrow 0$ во входном потоке. Отметим, что такое сжатие не гарантирует, что полученная последовательность будет правильной; т.е. су-

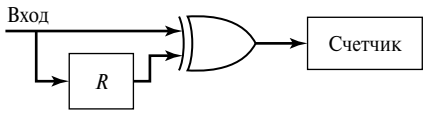
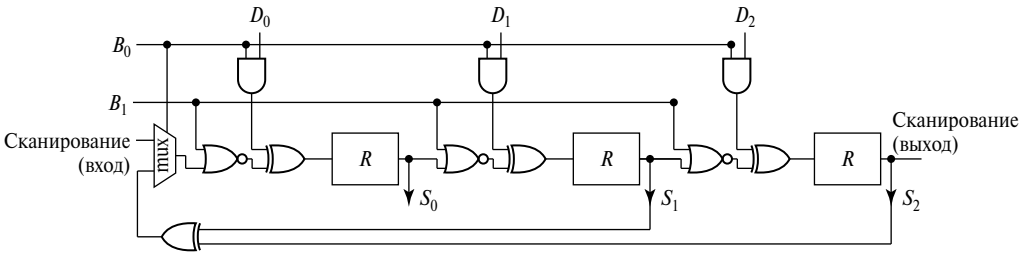


Рис. З.12. Сигнатурный анализ однобитового потока данных

ществует множество различных последовательностей с одинаковым числом переходов. Впрочем, поскольку шансы попасть на одну из таких последовательностей довольно малы, такой риск является допустимым.

Еще одна техника иллюстрируется на рис. З.13, *а*. Здесь показана модификация сдвигового регистра с линейной обратной связью, преимущество которой заключается в том, что одно и то же аппаратное обеспечение можно задействовать и для генерации входных кодов, и для сигнатурного анализа. Каждое входное слово данных подвергается операции ИСКЛЮЧАЮЩЕЕ ИЛИ с содержимым LFSR. В конце тестовой последовательности LFSR будет содержать сигнатуру, или *синдром*, этой последовательности данных, который можно сравнить с синдромом правильной схемы. Схема не только реализует генератор случайных кодов и сигнатурный анализатор, ее еще можно использовать как нормальный регистр и сканирующий регистр в зависимости от значений контрольных сигналов B_0 и B_1 (см. рис. З.13, *б*). Данный подход к тестированию, объединяющий ряд различных техник, называется *встроенным наблюдением логического блока* (Built-In Logic Block Observation — BILBO) [Koeneman79]. На рис. З.13, *в* показано типичное применение контроллера BILBO. Вначале в регистр A схемы BILBO вводится затравочное число, при этом регистр B инициализируется. Затем регистры A и B работают в режимах генерирования случайных кодов и сигнатурного анализа соответственно. В конце тестовой последовательности с помощью режима сканирования из регистра B считывается сигнатура.

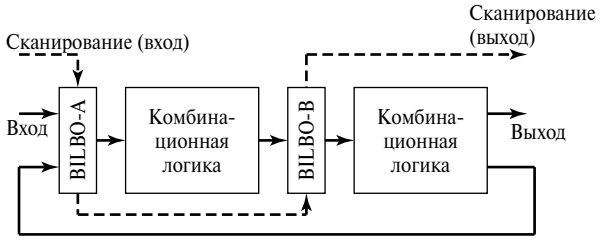
Наконец, следует отметить, что самотестирование наиболее выгодно при проверке регулярных структур, подобных запоминающим устройствам. Проверить, что такая последовательная схема, какой является память, не имеет ошибок, довольно сложно. Задача дополнительно усложняется тем, что из-за перекрестных помех и других паразитных эффектов на данные, считываемые из ячейки (или записываемые в нее), могут влиять значения, хранимые в соседних ячейках. Таким образом тесты памяти включают запись множества различных кодов в запоминающее устройство и их чтение с использованием переменных последовательностей адресации. В качестве обычных кодов применяются последовательности, состоящие из одних нулей или единиц либо чередующихся нулей и единиц. Схема адресации может включать запись всей памяти с последующим считыванием или другие схемы чтения/записи. Служебные издержки, связанные с реализацией этого подхода, малы по сравнению с размером памяти, поэтому данную схему тестирования можно вообще встроить в интегральную схему, как показано на рис. З.14. Такой подход существенно уменьшает время тестирования и минимизирует внешний контроль.



Трехбитовый регистр BILBO
а

B_0 B_1	Режим работы
1 1	Нормальный
0 0	Сканирование
1 0	Генерация шаблонов или сигнатурный анализ
0 1	Сброс

Режимы BILBO
б



Использование BILBO
в

Рис. 3.13. Встроенный контроллер логического блока

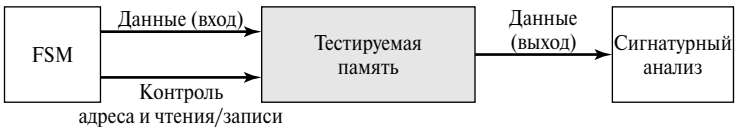


Рис. 3.14. Самопроверка памяти

Развитие систем на кристалле совсем не облегчает тестирование интегральных схем. Одна интегральная схема может содержать микропроцессоры и процессоры сигналов, несколько внедренных запоминающих устройств, модули специализированных интегральных схем, программируемые пользователем матрицы логических элементов и внутрикристальные шины и сети. Каждый из таких модулей имеет собственный предпочтительный способ тестирования, поэтому тестирование схемы в целом является достаточно проблематичным. Единственная возможность — использовать встроенную самопроверку, как показано на рис. З.15 [Krstic01]. Каждый модуль, составляющий систему, соединяется с внутрикристальной сетью через "обертку" — настраиваемый интерфейс, расположенный между блоком и сетью, который поддерживает такие функции, как синхронизация и передача данных. В данную "обертку" можно включить и модуль тестирования. Например, для модуля специализированной интегральной схемы, включающего цепочку сканирования, модуль тестирования обеспечивает интерфейс к цепочке сканирования

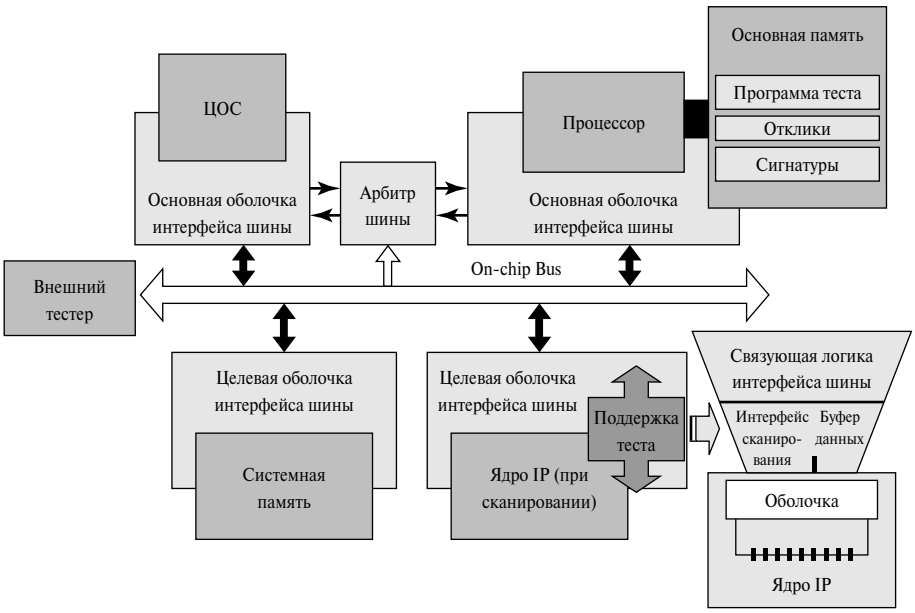


Рис. 3.15. Методология тестирования системы на кристалле

и буфер для тестовых кодов. В этот буфер можно непосредственно записывать информацию (а также считывать ее из него) через системную шину. Аналогично, модуль запоминающего устройства можно оборудовать генератором текстовых кодов и сигнатурным анализатором. В то же время всего этого может оказаться недостаточно, если не реализовать грамотное управление тестом схемы как целого. К счастью, большинство систем на кристалле имеют программируемый процессор, который можно использовать во время запуска для управления тестом и верификации других модулей.

Генерирование тестовых кодов

В предыдущих разделах мы обсудили, как модифицировать проект так, чтобы он допускал эффективное применение тестовых кодов. При этом мы не рассматривали, как определить, какие коды необходимо применять для обеспечения хорошего покрытия неисправностей. В прошлом данный процесс был очень сложным, и специалист по испытаниям (не разработчик!) должен был создавать тестовые векторы после завершения проектирования.

В данном разделе мы несколько глубже затронем тему автоматического генерирования тестовых кодов и представим техники, позволяющие оценивать качество тестовой последовательности. До этого, правда, мы должны более подробно проанализировать концепцию неисправности.

Модель неисправностей

Существует множество возможных производственных дефектов, проявляющихся в виде коротких замыканий между сигналами, коротких замыканий с шинами питания и т.п. Чтобы оценить эффективность схемы тестирования и концепции хорошей или плохой схемы, мы должны связать эти дефекты с моделью схемы или, другими словами, создать *модель неисправности*. Наиболее популярным подходом к решению этой задачи является *модель с константными неисправностями* (stuck-at model). В большинстве средств тестирования единственная предусмотренная неисправность — короткое замыкание с шинами питания. Дефекты подобного рода называются неисправностями типа *константный нуль* (stuck-at-zero, sa0) и *константная единица* (stuck-at-one, sa1) и обозначают короткое замыкание с GND и V_{DD} соответственно.

Можно показать, что модель sa0-sa1 не покрывает все возможные неисправности, которые могут возникнуть в современной интегральной схеме, поэтому необходимо вводить неисправности типа *постоянно открытый транзистор*, или *обрыв цепи* (stuck-at-open), и *постоянно закороченный транзистор* (stuck-at-short). Следует, правда, отметить, что добавление этих неисправностей существенно усложняет процесс генерирования тестовых кодов. Более того, значительная часть таких неисправностей покрывается моделью sa0-sa1. Чтобы проиллюстрировать данное наблюдение, рассмотрим МОП-элемент с резистивной нагрузкой, изображенный на рис. З.16. Все короткие замыкания с шинами питания моделируются путем введения неисправностей sa0 и sa1 в узлах A, B, C, Z и X. На рисунке отмечены некоторые неисправности типа *постоянно открытый транзистор* (β) и *постоянно закороченный транзистор* (α,γ). Можно видеть, что данные дефекты уже покрыты неисправностями sa0 и sa1 на различных узлах. Например, неисправность α покрывается неисправностью A_{sa1}, β — неисправностью A_{sa0} или B_{sa0}, а неисправность γ эквивалентна Z_{sa1}.

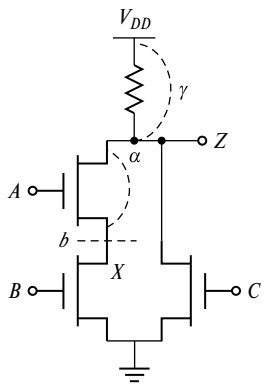
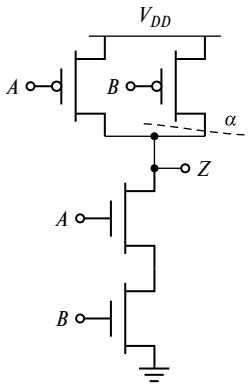


Рис. 3.16. Логический элемент с резистивной нагрузкой. На рисунке обозначены некоторые неисправности типа постоянно открытого транзистора (β) и постоянно закороченного транзистора (α, γ)



A	B	Z
0	2	1
1	1	0
1	0	Z_{t-1}

Рис. 3.17. Комплементарный МОП-вентиль НЕ-И с двумя входами и его таблица истинности при неисправности типа “постоянно открытый транзистора”

Даже при этих условиях неисправности типа короткого замыкания и обрывов могут порождать некоторые интересные артефакты, которые не покрываются моделью sa0-sa1 и которые достойны особого внимания. Рассмотрим изображенный на рис. З.17 логический элемент НЕ-И с двумя входами, в котором имеется неисправность типа *обрыв цепи*, α. Кроме того, на рисунке показана таблица истинности неисправной схемы. Для входного сигнала ($A = 1$, $B = 0$) выходной узел сохраняет свое предыдущее значение, тогда как правильным является значение 1. В зависимости от предыдущего вектора возбуждения данную неисправность можно обнаружить или не обнаружить. Фактически схема ведет себя как последовательная сеть. Для детектирования неисправности необходимо последовательно применить два вектора. Первый из них переводит выход в состояние 0 (или $A = 1$ и $B = 1$), а второй применяет код $A = 1$, $B = 0$. Отметим также, что неисправности типа постоянно закороченного транзистора являются источником проблем в КМОП-схемах, поскольку для определенных входных значений могут порождать постоянный ток между шинами питания, что, в свою очередь, приводит к неопределенным выходным напряжениям.

Хотя модель sa0-sa1 не идеальна, ее простота и сравнительно большое покрытие неисправностей сделала ее стандартной моделью де-факто. Довольно часто ее дополняют другими техниками, например функциональным тестированием, *тестом IDDQ*, который измеряет изменения тока в рабочей точке, вызванные короткими замыканиями, и тестом с учетом задержек. Никакая из существующих моделей тестирования не застрахована полностью от ошибок, поэтому часто требуется объединять несколько методов тестирования [Bhavsar01].

Автоматическая генерация тестовых шаблонов

Задача процесса автоматической генерации тестовых шаблонов (Automatic Test-Pattern Generation — ATPG) — определить минимальный набор векторов

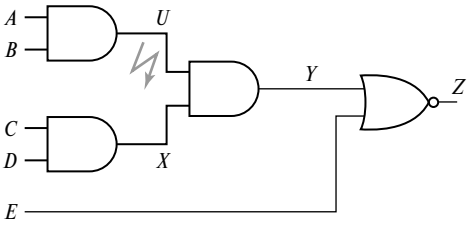


Рис. 3.18. Простая логическая цепь с ошибкой sa0 на узле U

возбуждения, охватывающий достаточную часть набор сбоев, определяемых принятой моделью неисправностей. Один из возможных подходов к этому вопросу заключается в использовании случайного набора тестовых векторов. В таком случае по результатам моделирования определяется, сколько потенциальных неисправностей было выявлено. На основе этих данных принимается решение о том, сколько и каких тестовых векторов необходимо добавить или удалить. Альтернативный и потенциально более привлекательный подход заключается в использовании функциональных возможностей булевой сети для расчета подходящего тестового вектора для данной неисправности. Чтобы проиллюстрировать данную концепцию, рассмотрим пример, изображенный на рис. З.18. Допустим, что наша цель — определить входное возбуждение, которое позволяет выявить на выходе сети Z неисправность sa0 на узле U. Первое требование к такому возбуждению: оно должно инициировать срабатывание неисправности (снова возвращаемся к *контролируемости*). В данном случае мы ищем код, который при нормальной работе переведет узел U в состояние 1. Единственное возможное решение — $A = 1$ и $B = 1$. Далее неправильный сигнал должен дойти до выходного узла Z, чтобы его можно было *наблюдать*. Данная фаза называется *активизацией пути*. Чтобы любые изменений на узле U распространились далее, необходимо, чтобы узел X имел значение 1, а узел E — 0. Таким образом, получаем единственный возможный тестовый вектор U_{sa0} для выявления указанной неисправности: $A = B = C = D = 1$, $E = 0$.

Данный пример очень прост; вывести минимальный набор тестовых векторов для более сложных цепей обычно гораздо сложнее. Впрочем, существует множество прекрасных подходов, позволяющих решить эту проблему. Наиболее значимыми результатами в этой области стала разработка алгоритмов D [Roth66] и PODEM [Goel81], которые в той или иной степени используются во множестве современных средств автоматического генерирования тестовых кодов. Достаточно сказать, что данная область является одним из основных направлений разработок в сфере автоматизации проектирования и многие производители предлагают множество мощных инструментов.

Моделирование неисправностей

Средство моделирования неисправностей измеряет качество тестовой программы. Оно определяет *покрытие возможных неисправностей* (общее число неисправностей, детектируемых тестовой последовательностью, деленное на удвоенное число узлов в цепи, поскольку на каждом узле возможно возникновение ошибки sa0 и sa1). Естественно, покрытие возможных неисправностей имеет смысл только в контексте принятой модели неисправностей. В модели sa0-sa1 некоторые мостиковые неисправности и отказы вследствие короткого замыкания не охвачены, поэтому они не проявятся в статистике покрытия.

Наиболее распространенным подходом к моделированию неисправностей является техники параллельного моделирования, в которой одновременно с годной схемой моделируется несколько дефектных, в каждой из которых реализована одна ошибка. Результаты моделирования сравниваются, и если выходы отличаются, считается, что неисправность была обнаружена для данного тестового вектора. Такое описание является чересчур упрощенным, и в большинстве средств моделирования для ускорения процесса применяются несколько техник (например, вначале выбираются неисправности с высокой вероятностью появления). Кроме того, существуют аппаратные ускорители процесса моделирования, основанные на параллельной обработке и существенно сокращающие время процесса [Agrawal88, p. 159–240].

Для любознательных

Более подробно тема проектирования с возможностью тестирования освещена в [Agrawal88] и [Abramovic91]. Прекрасный обзор проблем тестирования (а также их решений) для быстродействующих микропроцессоров можно найти в [Bhavsar01].

Список использованной литературы

[Abramovic91] M. Abramovic, M. Breuer, and A. Friedman, *Digital Systems Testing and Testable Design*, IEEE Press, Piscataway, NJ, 1991.

[Agrawal88] V. Agrawal and S. Seth, Eds. *Test Generation for VLSI Chips*, IEEE Computer Society Press, 1988.

[Bhavsar01] D. Bhavsar, "Testing of High-Performance Microprocessors", in *Design of High-Performance Microprocessor Circuits*, A. Chandrakasan et al., Ed., p. 523–544, IEEE Press, 2001.

[Eichelberger78] E. Eichelberger and T. Williams, "A Logic Design Structure for VLSI Testability", *Journal on Design Automation of Fault-Tolerant Computing*, May 1978, vol. 2, p. 165–178.

[Goel81] P. Goel, "An Implicit Enumeration Algorithm to Generate Tests for Combinational Logic Circuits", *IEEE Trans. on Computers*, June 1981, vol. C-30, no. 3, p. 26–268.

[IEEE1149] IEEE Standard 1149.1, "IEEE Standard Test Access Port and Boundary-Scan Architecture", IEEE Standards Board, New York.

[Koeneman79] B. Koeneman, J. Mucha, and O. Zwiehoff, "Built-in Logic-Block Observation Techniques", in *Digest 1979 Test Conference*, October 1979, p. 37–41.

[Kornegay92] K. Kornegay, *Automated Testing in an Integrated System Design Environment*, Ph. D dissertation., Mem. No. UCB/ERL M92/104, Sept. 1992.

[Krstic01] A. Krstic, W. Lai, L. Chen, K. Cheng, and S. Dey, "Embedded Software-Based Self-Testing for SOC Design", `http://www.gigascale.org/pubs/139.html`, December 2001.

[Roth66] J. Roth, "Diagnosis of Automata Failures: A Calculus and a Method", *IBM Journal of Research and Development*, 1966, vol. 10, p. 278–291.

[Wang86] L. Wang and E. McCluskey, "Complete Feedback Shift-Register Design for Built-in Self Test", *Proc. ICCAD 1986*, November, 1986, p. 56–59.

[Weste93] N. Weste and K. Eshraghian, *Principles of CMOS VLSI Design — A Systems Perspective*, Addison-Wesley, 1993.

[Williams83] T. Williams and K. Parker, "Design for Testability — A Survey", *Proceedings IEEE*, Jan. 1983, vol. 71, p. 98–112.

ЧАСТЬ

IV

Приложения

Приложение

A

Ответы на задачи

Задача 3.1

Сместить транзистор в желательный рабочий режим, подав подходящее напряжение на затвор. Подать ток I на сток. Ток, протекающий через емкость стока I_C, можно найти, вычитая из I ток транзистора I_D. Зная V_{DS} и $I_C(V_{DS})$, легко найти емкость.

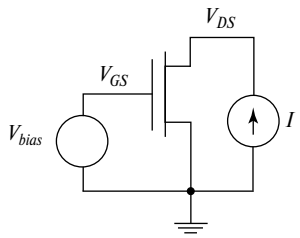


Рис. A.1. Моделирование емкости стока n-МОП-транзистора

Задача 3.2

Предполагая, что ток является квадратичной функцией напряжения, получаем следующую таблицу.

Масштабирование	**Полное**	**Общее**	**С постоянным напряжением**
I_{SAT}	$1/S$	S/U^2	S
R_{ON}	1	U/S	$1/S$
t_p	$1/S$	U/S^2	$1/S^2$
P	$1/S^2$	S/U^3	S

Задача 4.1

См. рис. А.2.

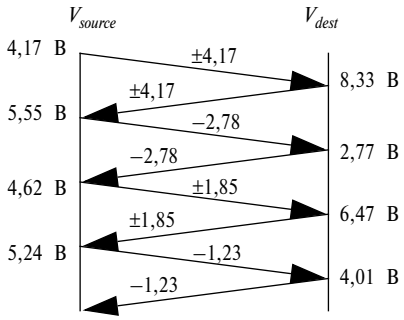


Рис. А.2. Характеристика линии передачи

Задача 5.1

Когда V_{DD} значительно больше V_T, p- и n-МОП-транзисторы находятся в режиме насыщения вблизи рабочей точки V_M. Чтобы найти выражение для V_M, необходимо приравнять токи, проходящие через транзисторы. Модуляция длины канала не учитывается.

$$\frac{k_n}{2}(V_M - V_{Tn})^2 = -\frac{k_p}{2}(V_M - V_{DD} - V_{Tp})^2$$

Задача 5.2

Предположим, что p- и n-МОП-транзисторы находятся в режиме насыщения вблизи рабочей точки V_M. Записываем для этого случая уравнение тока, дифференцируем его и находим dV_{out}/dV_{in}. В результате получаем следующее:

$$\frac{dV_{out}}{dV_{in}} = -\frac{k_n(V_{in} - V_{Tn})(1 + \lambda_n V_{out}) + k_p(V_{in} - V_{DD} - V_{Tp})(1 + \lambda_p V_{out} - \lambda_p V_{DD})}{(\lambda_n k_n(V_{in} - V_{Tn})^2 + \lambda_p k_p(V_{in} - V_{DD} - V_{Tp})^2)/2}.$$

Пренебрегая некоторыми членами второго порядка и полагая $V_{in} = V_M$, получаем:

$$g = -\frac{k_n(V_M - V_{Tn}) + k_p(V_M - V_{DD} - V_{Tp})}{I_D(V_M)(\lambda_n - \lambda_p)}.$$

Зная g, с помощью уравнения (5.7) легко найти V_{IH} и V_{IL}.

Задача 5.3

Используя уравнение (3.39) для тока МОП-устройства в подпороговой области, игнорируя модуляцию длины канала и предполагая, что p- и n-МОП-устройства симметричны, усиление инвертора находится как $-(dI_D/dV_{GS})/(dI_D/dV_{DS})$. В результате сразу приходим к уравнению (5.12).

Задача 5.4

Для краткости ограничимся t_{pHL}. Предполагая, что используется источник тока, t_{pHL} можно аппроксимировать следующим образом:

$$t_{pLH} = \frac{C_L(V_{DD}/2)}{I_{av}},$$

где I_{av} — средний ток разрядки. I_{av} мы будем вычислять как среднее между рассматриваемыми точками, т.е. $V_{out} = V_{DD}$ и $V_{out} = V_{DD}/2$. Поскольку n-МОП-транзистор находится в режиме насыщения скорости бо́льшую часть цикла разрядки, получаем следующее:

$$I_{av} = \frac{I_{Dsat}(1 + \lambda V_{DD}) + I_{Dsat}(1 + \lambda(V_{DD}/2))}{2} = I_{Dsat}\left(1 + \lambda\frac{3V_{DD}}{4}\right),$$

где $I_{Dsat} = k_n((V_{DD} - V_{Tn})V_{Dsat} - V_{Dsat}^2/2)$.

В результате получаем следующее выражение для задержки распространения, которое достаточно близко к уравнению (5.22):

$$t_{pLH} = \frac{C_L V_{DD}}{2I_{Dsat}\left(1 + \lambda\frac{3V_{DD}}{4}\right)}.$$

Задача 5.5

Дифференцируя уравнение (5.32) по входным емкостям трех каскадов и учитывая дополнительное разветвление по выходу, равное 4, для каскадов 1 и 2, получаем следующее соотношение:

$$\frac{4C_{g,2}}{C_{g,1}} = \frac{4C_{g,3}}{C_{g,2}} = \frac{C_L}{C_{g,3}}.$$

Кроме того, $C_L = 64C_{g,1}$. Полученную систему уравнений легко решить и найти следующие значения: $s_1 = 1$; $s_2 = 2^{4/3} = 2,52$; $s_3 = 2^{8/3} = 6,35$. Согласно этим данным, минимальная задержка распространения между входом и выходом равна $33,24t_{p0}$ (предполагается, что $\gamma = 1$).

Задача 5.6

Влияние наклона входного сигнала уменьшается с уменьшением напряжения питания (для фиксированных порогов). Уменьшение напряжения питания

сокращает временной интервал, в течение которого оба устройства включены одновременно, следовательно, уменьшается и теряемый ток.

Задача 6.1

Критическими являются транзисторы в последовательных соединениях. Когда N транзисторов соединено последовательно, мы умножаем их ширину на N и получаем сложное устройство с сопротивлением в открытом состоянии, равным сопротивлению одного транзистора. Транзисторы, соединенные параллельно, не модифицируются, поскольку в худшем случае их сопротивление равно сопротивлению отдельных устройств. Данная техника рекурсивно используется для всех ветвей цепи. Размеры устройств выбираются такими, как показано на рис. А.3. С точки зрения задержки получающийся прибор эквивалентен инвертору с n-МОП-транзистором с размерами 0,5/0,25 мкм и p-МОП-транзистором с размерами 1,5/0,25 мкм.

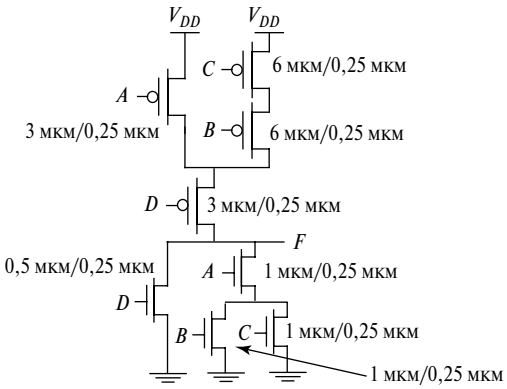


Рис. А.3. Масштабирование транзисторов в сложной КМОП-цепи

Задача 6.2

Трудозатраты на ветвление равны: $b_1 = 4$, $b_2 = 4$, $b_3 = 1$. Следовательно, $B = 16$. Эффективное разветвление тракта по выходу составляет $F = 64$. Логические трудозатраты тракта $G = 1$. Следовательно, общие трудозатраты тракта равны $H = GFB = 1024$. Трудозатраты логического элемента, минимизирующие задержку, равны $(H)^{1/N}$, т.е. 10,079. $C_{in3} = 64C_{g,1}/10,079 = 6,35C_{g,1}$ и $C_{in2} = 4 \times 6,35C_{g,1}/10,079 = 2,52C_{g,1}$. В результате получаем $s_1 = 1$, $s_2 = 2,52$, $s_3 = 6,35$.

Задача 6.3

Поскольку входы элемента ИСКЛЮЧАЮЩЕЕ ИЛИ с N входами не коррелируют и распределены равномерно, вероятность перехода выражается следующим образом:

$$\alpha_{0\to 1} = \frac{N_0}{2^N} \cdot \frac{N_1}{2^N} = \frac{\frac{2^N}{2} \cdot \frac{2^N}{2}}{2^{2N}} = \frac{1}{4}$$

Задача 6.4

Данные вероятности перехода легко вывести из следующих рассуждений. Вероятность того, что элемент И находится в состоянии "1", равна вероятности того, что вход A и вход B имеют высокий уровень, или $P_{AND}(1) = P_A \cdot P_B$. Отсюда можно вычислить вероятность перехода: $P_{0\to 1} = (1 - P(1))P(1)$. Подобным образом вероятность того, что выход элемента ИСКЛЮЧАЮЩЕЕ ИЛИ имеет высокий уровень, равна вероятности того, что или A, или B имеет высокий уровень (но не одновременно). Сказанное записывается следующим образом: $P_{XOR}(1) = P_A + P_B - 2P_AP_B$. Обратите внимание на то, что вероятности перехода для комплементарных логических элементов (НЕ-И, НЕ-ИЛИ, НЕ-ИСКЛЮЧАЮЩЕЕ ИЛИ) идентичны неинвертирующим элементам.

Задача 6.5

В отличие от комплементарных МОП-элементов структура НЕ-ИЛИ больше подходит для псевдо-n-МОП-устройств. В структуре НЕ-ИЛИ совсем не используются последовательные транзисторы.

Задача 6.6

Последовательное соединение шести транзисторов увеличивает сопротивление разгрузки в 6 раз по сравнению с одним устройством, имеющим размеры 0,5/0,25 мкм. Чтобы довести напряжение на узле x до порога переключения инвертора ($V_{DD}/2$), эффективное соотношение W/L должно равняться 1. Поскольку минимальная ширина устройства равна 0,375 мкм, максимальное отношение W/L равно 0,375/0,375 мкм (длину нужно увеличить, чтобы создать некоторый резерв).

Задача 6.7

На рис. А.4 показаны смоделированные сопротивления n- и p-МОП-устройства. Эффективное сопротивление приблизительно равно 4 кОм (усреднено по крайним точкам).

Задача 6.8

Вероятность того, что выход разряжается, равна вероятности того, что все входы имеют высокий уровень. Это также равно вероятности перехода $0 \to 1$, поскольку в фазе низкого уровня тактового сигнала выход всегда предварительно заряжен.

$$\alpha_{0\to 1} = p_A \cdot p_B \cdot p_C \cdot p_D = 0,012$$

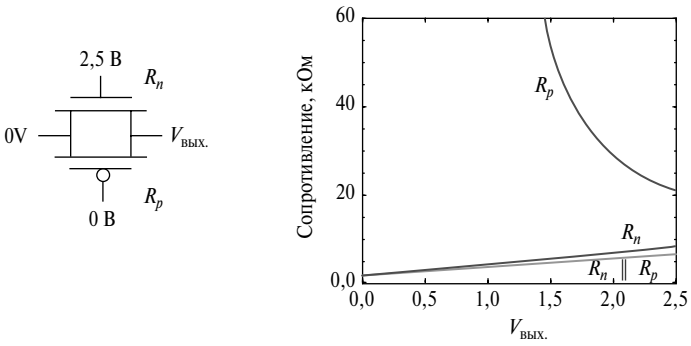


Рис. А.4. Сопротивление разрядки передаточного вентиля

Задача 6.9

Приравнивая два тока утечки, получаем:

$$I_S 10^{\frac{0}{S}} \left(1 - 10^{-\frac{nV_x}{S}}\right) = I_S 10^{\frac{-V_x}{S}} \left(1 - 10^{-n\frac{(V_{DD}-V_x)}{S}}\right).$$

Предполагая, что V_x мало, вторым членом в правой части уравнения можно пренебречь. В результате выражение упрощается до следующего вида:

$$\left(1 - 10^{-\frac{nV_x}{S}}\right) = 10^{\frac{-V_x}{S}} \text{ или } \exp^{-\frac{V_x}{nV_{th}}} = 1 - \exp^{-\frac{V_x}{V_{th}}}.$$

Обозначая $\alpha = \frac{V_x}{V_{th}}$ и предполагая, что α мало, раскладываем правую часть в ряд Тейлора в окрестности нуля и получаем:

$$1 - \frac{\alpha}{n} = 1 - (1 - \alpha) = \alpha.$$

Переупорядочивая элементы данного уравнения, получаем следующее:

$$1 - \alpha = 1 - \frac{n}{n+1} = \frac{1}{n+1}.$$

Отсюда, снова используя разложение экспоненты в ряд Тейлора и беря логарифм от левой и правой части, получаем: $V_x = V_{th} \ln(1+n)$.

Задача 7.1

С точки зрения функциональных возможностей два инвертора можно убрать. Входной инвертор изолирует входы от внутренних узлов, а также обеспечивает триггеру фиксированную входную нагрузку.

Задача 7.2

Временная диаграмма неперекрывающихся тактовых сигналов показана на рис. А.5. Чтобы увеличить время неперекрывания, можно добавить задержки t_{d1} и t_{d2}.

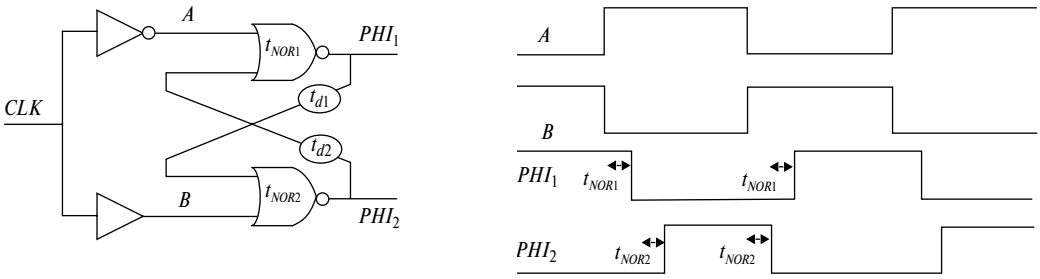


Рис. А.5. Генерирование неперекрывающихся тактовых сигналов

Задача 7.3

На рис. А.6 показаны ситуации, когда из защелки удаляется n- или p-МОП-устройство с высоким порогом. В обоих случаях существует паразитный путь утечки. Следовательно, для устранения утечки нужны оба устройства: и n-, и p-МОП.

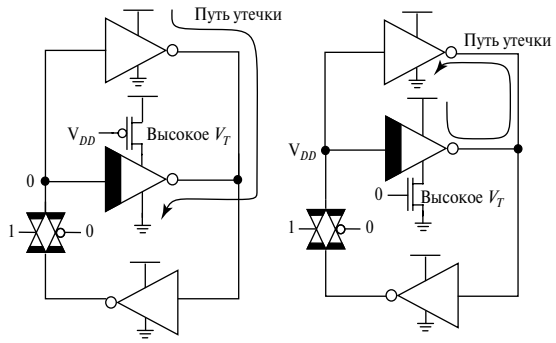


Рис. А.6. Регистры с двумя порогами, используемые для работы при меньшем напряжении

Задача 7.4

Таблица истинности для триггеров на основе элементов НЕ-И показана на рис. А.7. Вход 00 представляет собой запрещенное состояние.

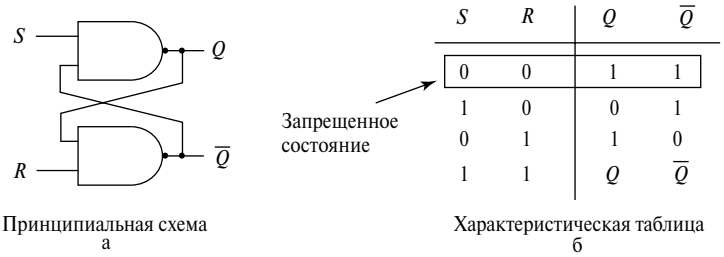


Рис. А.7. Триггер на основе элемента НЕ-И

Задача 7.5

При фиксированной пропускной способности тактовую частоту системы, активизируемой двумя фронтами, можно уменьшить вдвое. В первом приближении это даст вдвое меньшую мощность цепи разводки питания. С помощью аккуратного анализа можно показать, что физическая нагрузка на тактовый сигнал для триггера, активизируемого двумя фронтами, удваивается. В результате цепь разводки тактовых сигналов необходимо сделать больше, что несколько уменьшает выигрыш в мощности. Мощность, используемая для питания нагрузки, остается без изменений, поскольку она определяется удвоенной физической емкостью, переключаемой с половинной частотой.

Задача 7.6

Импульсный регистр с сигналом *Enable* получается, как показано на рис. А.8, путем простого добавления трех устройств с входом D (данные).

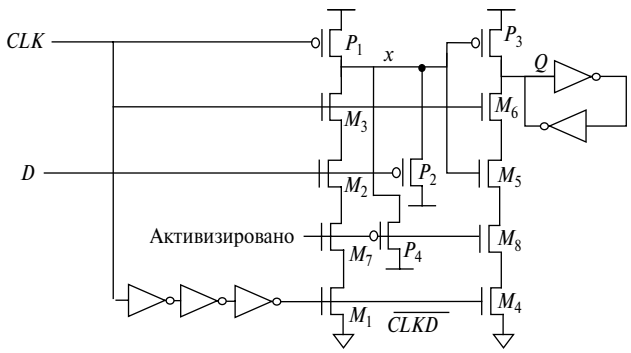


Рис. А.8. Импульсные регистр с сигналом $Enable$

Задача 7.7

Для перехода с низкого уровня на высокий транзистор M_2 отпирается только в том случае, когда на узле X выходное напряжение инвертора M_1-M_5 до-

стигает следующего значения: $V_X = V_{in} - V_{Tn}$. Используем это значение V_{in} как аппроксимацию V_{M-}. Игнорируя влияние подложки и предполагая, что оба устройства находятся в режиме насыщения, можно получить следующее выражение для V_{M-}:

$$k'_n \left(\frac{W}{L}\right)_1 \left((V_{M-} - V_{Tn})V_{DSATn} - \frac{V^2_{DSATn}}{2}\right) =$$
$$= k'_n \left(\frac{W}{L}\right)_5 \left((V_{DD} - (V_{M-} - V_{Tn}) - V_{Tn})V_{DSATn} - \frac{V^2_{DSATn}}{2}\right);$$
$$V_{M-} = \frac{V_{Tn} \cdot V_{DSATn} + \frac{V^2_{DSATn}}{2} + \left((V_{DD} \cdot V_{DSATn}) - \frac{V^2_{DSATn}}{2}\right)(W_5/W_1)}{(1 + W_5/W_1) \cdot V_{DSATn}}.$$

Аналогично можно найти V_{M+}:

$$V_{M+} = \frac{(V_{DD} + V_{Tp}) \cdot V_{DSATp} + \frac{V^2_{DSATp}}{2} + \left(-\frac{V^2_{DSATp}}{2}\right)(W_6/W_4)}{(1 + W_6/W_4) \cdot V_{DSATp}}.$$

Задача 8.1

$$\overline{f_0} = \overline{\overline{x_0 x_1} \cdot x_2};$$
$$\overline{f_1} = \overline{\overline{x_0 x_1 x_2} \cdot \bar{x_2} \cdot \overline{\bar{x}_0 x_1}}.$$

Задача 9.1

Правильный подход — масштабировать только последний каскад (вариант ответа б). Именно здесь происходит наибольшее изменение тока (в основном из-за эффекта Ldi/dt). Влияние масштабирования одного каскада на быстродействие минимально.

Задача 9.2

Потребление энергии для перехода 0-1-0 равно $C_L V_{DD}(V_{DD}) - V_{Tn} + V_{Tp}$. Следовательно, уменьшение энергии по сравнению с использованием полного размаха равно $(V_{DD} - V_{Tn} + V_{Tp})/V_{DD}$. Для $F_{DD} = 2,5$ В и порогов устройства, равных 0,4 В, экономия составляет порядка 32%.

Задача 10.1

Предположим, что буфер FIFO изначально пуст и сигналы $En1$, $En2$, $En3$ и $Ack0$ равны 0. Мы будем вводить новые данные, не удаляя старые, пока буфер не заполнится. Обратите внимание на то, что каждое слово данных автоматически опускает вглубь буфера, если может это сделать.

$$Req_i \uparrow\rightarrow En1 \uparrow\rightarrow Done1 \uparrow\rightarrow En2 \uparrow\rightarrow Done2 \uparrow\rightarrow En3 \uparrow\rightarrow Req0 \uparrow$$
$$\rightarrow Ack_i \uparrow\rightarrow Req_i \downarrow\rightarrow En1 \downarrow \rightarrow Done1 \downarrow\rightarrow En2 \downarrow\rightarrow Done2 \downarrow$$
$$\rightarrow Ack_i \downarrow\rightarrow Req_i \uparrow \rightarrow En1 \uparrow\rightarrow Done1 \uparrow$$
$$\rightarrow Ack_i \uparrow \rightarrow Req_i \downarrow$$

Видим, что буфер FIFO заполняется, если сигналы En (или сигналы Ak следующего каскада) чередуют свои значения. В показанном выше случае имеем: $En1 = 1$, $En2 = 0$, $En3 = 1$ и $Ack0 = 0$, а $Req_i = 0$. С другой стороны, буфер FIFO опустошается, если все сигналы En равны между собой (имеют значение 0 или 1).

Задача 10.2

См. рис. А.9.

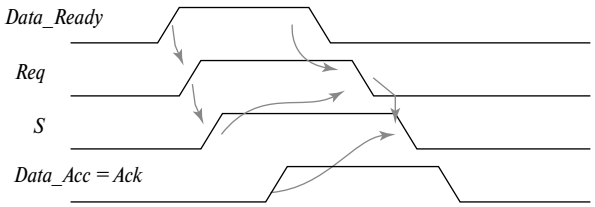


Рис. А.9. Четырехфазный протокол

Задача 12.1

(0) 0100; (1) 1001; (2) 0101; (3) 0000.

Задача 12.2

(0) 1011; (1) 0110; (2) 1010; (3) 1111.

Задача 12.3

(0) 0100; (1) 1001; (2) 0101; (3) 0000.

Задача 12.4

Основной проблемой при реализации ПЗУ на элементах НЕ-И является перераспределение заряда. Указанная реализация возможна, но разработчик должен быть очень внимательным.

Задача 12.5

В режиме хранения не расходуется мощность. Этим ситуация отличается от реализации псевдо-n-МОП-ячейки, где в любой момент времени одна из сторон ячейки проводит ток.

Приложение

Б

Справочная информация

Значения некоторых физических констант и коэффициентов материалов

Название	Символ	Значение	Единицы
Комнатная температура	T	$300(=27°C)$	К
Константа Больцмана	k	$1,38 \times 10^{-23}$	Дж/К
Заряд электрона	q	$1,6 \times 10^{-19}$	Кл
Температурное напряжение	$\phi_T = kT/q$	26	мВ (при 300 К)
Внутренняя концентрация носителей заряда (кремний)	n_i	$1,5 \times 10^{10}$	$см^{-3}$ (при 300 К)
Диэлектрическая проницаемость кремния (Si)	ε_{si}	$1,05 \times 10^{-12}$	Ф/см
Диэлектрическая проницаемость оксида кремния (SiO_2)	ε_{ox}	$3,5 \times 10^{-13}$	Ф/см
Сопротивление алюминия (Al)	ρ_{Al}	$2,7 \times 10^{-8}$	Ом·м
Сопротивление меди (Cu)	ρ_{Cu}	$1,7 \times 10^{-8}$	Ом·м
Магнитная проницаемость вакуума	μ_0	$12,6 \times 10^{-7}$	Вб/Ам
Скорость света в вакууме	c_0	30	см/нс
Скорость света в SiO_2	c_{ox}	15	см/нс

Формулы и уравнения

Диод

$$I_D = I_S(\exp^{V_D/\phi_T} - 1) = Q_D/\tau_T;$$

$$C_j = \frac{C_{j0}}{(1 - V_D/\phi_0)^m};$$

$$K_{eq} = \frac{-\phi_0^m}{(V_{high} - V_{low})(1-m)} \times [(\phi_0 - V_{high})^{1-m} - (\phi_0 - V_{low})^{1-m}].$$

МОП-транзистор

$$V_T = V_{T0} + \gamma(\sqrt{|\ -2\phi_F + V_{SB}\ |} - \sqrt{|\ -2\phi_F\ |});$$
$$I_D = \frac{k'_n}{2}\frac{W}{L}(V_{GS} - V_T)^2(1 + \lambda V_{DS});$$
$$I_D = \upsilon_{sat} C_{0X} W (V_{GS} - V_T - \frac{V_{DSAT}}{2})(1 + \lambda V_{DS});$$
$$I_D = k'_n \frac{W}{L}\left((V_{GS} - V_T)V_{DS} - \frac{V_{DS}^2}{2}\right);$$
$$I_D = I_S \exp^{\frac{V_{GS}}{nKT/q}}\left(1 - \exp^{-\frac{V_{DS}}{KT/q}}\right).$$

Модель МОП-транзистора в глубокой субмикронной области

$$I_D = 0 \text{ при } V_{GT} \le 0;$$
$$I_D = k' \frac{W}{L}\left(V_{GT}V_{min} - \frac{V_{min}^2}{2}\right)(1 + \lambda V_{DS}) \text{ при } V_{GT} \ge 0;$$
$$V_{min} = \min(V_{GT}, V_{DS}, V_{DSAT}) \text{ и } V_{GT} = V_{GS} - V_T.$$

Модель МОП-транзистора как ключа

$$R_{eq} = \frac{1}{2}\left(\frac{V_{DD}}{I_{DSAT}(1 + \lambda V_{DD})} + \frac{V_{DD}/2}{I_{DSAT}(1 + \lambda V_{DD}/2)}\right) \approx \frac{3}{4}\frac{V_{DD}}{I_{DSAT}}\left(1 - \frac{5}{6}\lambda V_{DD}\right).$$

Инвертор

$$V_{OH} = f(V_{OL}) \quad V_{OL} = f(V_{OH});$$
$$V_M = f(V_M);$$
$$t_p = 0,69 R_{eq} C_L = \frac{C_L(V_{swing}/2)}{I_{avg}};$$
$$P_{dyn} = C_L V_{DD} V_{swing} f \quad P_{stat} = V_{DD} I_{DD}.$$

Статический КМОП-инвертор

$$V_{OH} = V_{DD} \quad V_{OL} = GND;$$
$$V_M \approx \frac{rV_{DD}}{1 + r} \text{ при } r = \frac{k_p V_{DSATp}}{k_n V_{DSATn}};$$
$$V_{IH} = V_M - \frac{V_M}{g} \quad V_{IL} = V_M + \frac{V_{DD} - V_M}{g}$$
$$\text{при } g \approx \frac{1 + r}{(V_M - V_{Tn} - V_{DSATn}/2)(\lambda_n - \lambda_p)};$$

$$t_p = \frac{t_{pHL} + t_{pLH}}{2} = 0,69 C_L \left(\frac{R_{eqn} + R_{eqp}}{2} \right);$$
$$P_{av} = C_L V_{DD}^2 f.$$

Межсоединение

Сосредоточенная RC-цепь: $t_p = 0,69RC$.

Распределенная RC-цепь: $t_p = 0,38RC$.

Последовательность RC-блоков: $\tau_N = \sum_{i=1}^{N} R_i \sum_{j=1}^{N} C_j = \sum_{i=1}^{N} C_i \sum_{j=1}^{i} R_j$.

Отражение в линии передач: $\rho = \frac{V_{refl}}{V_{inc}} = \frac{I_{refl}}{I_{inc}} = \frac{R - Z_0}{R + Z_0}$.

Модели КМОП-устройств

0,25-микронная КМОП-технология (унифицированная модель)

Параметр	V_{T0}, В	γ, $\text{В}^{0,5}$	V_{DSAT}, В	k', $\text{А}/\text{В}^2$	λ, В^{-1}
n-МОП	0,43	0,4	0,63	115×10^{-6}	0,06
p-МОП	−0,4	−0,4	−1	-30×10^{-6}	−0,1

0,25-микронная КМОП-технология (модель ключа)

V_{DD}, В	1	1,5	2	2,5
n-МОП, кОм	35	19	15	13
p-МОП, кОм	115	55	38	31

Комбинационная КМОП-логика

Масштабирование транзисторов с использованием логических трудозатрат

$$F = \frac{C_L}{C_{g1}} = \prod_1^N \frac{f_i}{b_i}; \qquad G = \prod_1^N g_i; \qquad D = t_{p0} \prod_{j=1}^N \left(p_j + \frac{f_j g_j}{\gamma} \right);$$
$$B = \prod_1^N b_i; \qquad H = FGB; \qquad D_{min} = t_{p0} \left(\prod_{j=1}^N p_j + \frac{N(\sqrt[N]{H})}{\gamma} \right).$$

Модели межсоединений

Поверхностная и краевая емкость для типичного 0,25-микронного КМОП-процесса

Строки таблицы представляют верхнюю, а столбцы — нижнюю обкладку конденсатора. Поверхностная емкость выражена в аФ/мкм^2, а краевая емкость (в четных строках) — в аФ/мкм.

	Защитный слой	Активный слой	Поликремний	Al1	Al2	Al3	Al4
Поликремний	88						
	54						
Al1	30	41	57				
	40	47	54				
Al2	13	15	17	36			
	25	27	29	45			
Al3	8,9	9,4	10	15	41		
	18	19	20	27	49		
Al4	6,5	6,8	7	8,9	15	35	
	14	15	15	18	27	45	
Al5	5,2	5,4	5,4	6,6	9,1	14	38
	12	12	12	14	19	27	52

Удельное поверхностное сопротивление для типичного 0,25-микронного КМОП-процесса

Материал	Удельное поверхностное сопротивление (Ом/□)
Диффузия n- или p-кармана	1000–1500
Диффузный слой n^+, p^+	50–150
Диффузный слой n^+, p^+ с силицидом	3–5
Поликремний n^+, p^+	150–200
Поликремний n^+, p^+ с силицидом	4–5
Алюминий	0,05–0,1

Предметный указатель

F

FIFO, *755*
FLOTOX-транзистор *784*

N

N-карман, *69*
N-МОП-транзистор, *77*
N-область, *115*
N-тип, *76*; *115*
NORA–КМОП, *445*

P

P-карман, *69*
P-МОП-транзистор, *77*
P-область, *115*
P-тип, *76*; *115*
pn-переход, *115*; *131*

R

RC-дерево, *207*
RC-задержка, *562*
RC-модель
  распределенная, *207*
  сосредоточенная, *207*
RC-цепь
  SPICE-модель, *226*
  распределенная, *211*
RC-цепь первого порядка, *60*
  рассеивание энергии, *62*
RC-эффект, *562*

S

SPICE-модель
  BSIM3V3, *165*
  диода, *127*
  линии передачи, *228*
  МОП-транзисторов, *165*
  проводов, *226*
  транзистора, *167*
SR-триггер, *422*; *499*
  статический, *422*

A

Абстрагирование, *35*
Автоколебательная система, *453*
Адрес блока, *760*
Адрес в столбце, *758*
Адрес в строке, *758*
Адресное слово, *758*
Активизация пути, *885*
Активная нагрузка, *798*
Активная область, *76*; *82*
Активность переключения, *278*; *279*; *324*
Акцепторная примесь, *115*
Альфа-частицы, *840*
Арбитр, *652*
Арифметико-логическое устройство, *720*
Архитектура запоминающего устройства, *862*
Асинхронное межсоединение, *599*
Асинхронное проектирование, *596*; *665*

B

Вариации
  межсоединений, *611*
  параметров устройств, *610*
Верификация топологии, *486*
Видеопамять, *755*
Восстановление сигнала, *819*
Временная диаграмма работы, *44*
Временной инструмент, *386*

Временные метрики, *405*
Время
  доступа, *753*
  пробега, *221*
  проектирования
  такта, *58*
  удержания, *405*; *406*
  удержания регистра, *525*
  установки, *405*
  установки регистра, *525*
  фронта нарастания, *59*
  фронта спада, *59*
Входной сигнал
  время нарастания/спада, *271*
  наклон, *273*
Выборочный слой, *84*
Вывод
  линейный, *224*
Выпечатывание, *487*
Выравнивание, *821*
Выход
  дифференциальный, *337*
  комплементарный, *337*

Д

Двухкарманная технология, *77*
Двухкарманный подход, *69*
Двухшинное кодирование, *633*
Декодер, *718*; *758*
  столбцов, *758*
Детектирование изменения адресов, *834*
Диаграмма
  переходов, *643*
  ступенчатая, *221*
  штриховая, *399*
Динамический логический элемент
  рассеяние мощности, *359*
  скорость, *359*
Динамический регистр, *426*
Диод, *115*
  SPICE-модель, *127*
  идеальный, *118*
  обратносмещенный, *364*
  реальный, *125*
Дифференциальная логика, *342*
Диффузионная емкость, *157*; *254*
Диффузионная имплантация, *74*
Диффузионная область, *82*
Диффузия, *74*; *115*
  боковоая, *153*
Диэлектрик, *97*
Диэлектрическая проницаемость, *97*
Донорная примесь, *115*
Древовидный декодер, *816*
Дрейф во времени, *163*
Дрожание, *596*; *606*

Е

Емкостная нагрузка, *548*
Емкостная связь, *615*
Емкостные параметры, *159*
Емкость, *89*; *150*; *191*
  в КМОП, *544*
  диффузионная, *157*; *254*
  диффузионная
  заряжающаяся, *275*
  затвора разветвления по выходу, *255*
  затвора-стока, *252*
  КМОП-инвертора, *256*
  канала, *153*
  краевая, *192*
  МОП-структуры, *153*
  МОП-транзистора, *158*
  обедненной области, *122*
  перекрытия, *153*
  перехода, *124*; *157*
  разводки, *255*
  разряжающаяся, *275*
  расчет, *252*
  среднее распределение, *155*
  средняя, *125*
  стока, *774*
  эффективная, *278*

З

Задержка
  внутренняя, *266*; *522*
  копии, *635*
  ненагруженная, *266*
  ненагруженная собственная, *314*
  провода, *522*
  распространения, *58*; *258*; *333*; *406*; *525*
Закон Мура, *33*
Закон Рента, *85*
Запас помехоустойчивости, *245*; *333*
  КМОП-инвертора, *247*
Запоминающие устройства, *754*
Затвор, *76*; *129*
Защелка, *407*

динамическая, *425*
низковольтная статическая, *420*
основанная на мультиплексоре, *411*
отрицательная, *408*
положительная, *408*
статическая, *409*
Защитные кольца, *554*
Защитный окисл, *76*
ЗУ
динамические, *754*
энергозависимые, *754*

И

$И^2Л$-логика, *32*
Идеальный тактовый сигнал, *608*
Идеальный цифровой вентиль, *56*
Избыточные цепи, *842*
Изоляция, *491*
Изоляция канавками, *76*
Импеданс, *357*
выходной, *238*
характеристический, *217*
Инверсия, *132*
Инвертор
выбор размеров, *265*
емкостное рассеяние энергии, *278*
метрики, *294*
поглощающий ток, *454*
симметричный, *265*
Индуктивность, *89*; *202*
паразитная, *568*
Интегральная схема, *31*; *96*
корпусирование, *84*
Ионная имплантация, *74*
Исток, *129*

К

Карман, *68*; *83*
Каскад
буферный, *270*
динамических элементов, *370*
цепочки инверторов, *269*
Кассета, *166*
Кластерное масштабирование напряжения, *730*
КМОП ИС, *68*; *301*
КМОП-инвертор, *167*
быстродействие, *252*
выбор размеров, *264*
емкость, *256*
минимальное напряжение питания, *251*
порог переключения, *243*
расчет устойчивости, *241*
статический, *236*
КМОП-процесс, *76*; *149*
КМОП-схема
статическая комплементарная, *382*
тактируемая, *428*
КМОП-устройство
статическое, *301*
КМОП-элемент
динамический, *356*
НЕ-И с четырьмя входами, *311*
потребление мощности, *324*
Кодирование
двухшинное, *633*
состояний, *531*
Кольцевой генератор, *59*
Компилятор модуля, *481*
Конвейерная обработка, *534*
КонвейерныЙ информационный канал, *458*
Конвейерный подход, *442*
Конденсатор
“плавающий”, *194*
с параллельными обкладками, *191*
Конечный автомат, *404*
Контакт, *82*
нормы проектирования, *83*
Контактная разность потенциалов, *116*
pn-перехода, *118*
Контурный фильтр, *655*
Концепция стандартных ячеек, *397*
Копия критического пути, *635*
Корпус, *87*
Корпусирование, *68*; *94*
интегральной схемы, *84*
Коррекция ошибок, *843*
Корреляция сигналов, *326*
Косвенная стоимость, *45*
Коэффициент
активности, *412*
влияния подложки, *133*
модуляции длины канала, *137*
нагрузки ячейки, *797*
ослабления синфазного сигнала, *819*
отражения, *218*
переноса заряда, *804*
плавности, *125*

спада, *144*; *379*
усиления, *53*; *135*
чувствительности к переходу, *523*
Кремний-на-диэлектрике, *97*
Кремний-на-кремнии, *93*
Кристалл, *45*
Критерий Найквиста, *647*
Критический заряд, *841*
Критический путь, *316*; *679*
Кэш, *809*

Л

Линейное масштабирование, *81*
Линейный вывод, *224*
Линейный регулятор, *828*
Линия
возбуждаемая, *542*
микрополосковая, *194*
нагрузки, *239*
слабонагруженная, *540*
Линия передач
SPICE-модель, *228*
без потерь, *216*
модель, *215*
с потерями, *216*; *223*
Логарифмический сдвиговый регистр, *719*
Логика
“домино”
на дуальных шинах, *373*
с несколькими выходами, *375*
составная, *375*
динамическая, *356*; *382*
дифференциальная, *336*
комбинационную, *300*
на проходных транзисторах, *382*
программируемая, *497*
с внутренней синхронизацией, *643*
с постзарядкой, *644*
со сниженным потреблением напряжения, *379*
стандартизованная, *332*; *382*
Логическая (или булева) переменная, *49*
Логические ячейки, *522*
Логический перепад, *49*
Логический узел, *48*

М

Макроэлемент, *480*
Макроячейка, *480*; *499*
жесткая, *480*
мягкая, *480*
Маска, *762*
Масштабирование, *172*
короткоканальных приборов, *173*
напряжения источника питания, *249*
с постоянным напряжением, *173*
сети инверторов, *271*
технологии, *171*; *236*; *294*
транзистора, *280*; *733*
цепочки инверторов, *267*; *280*
Масштабируемая норма проектирования, *81*
Матрица
выводов, *90*
логических элементов, *489*
с предварительной диффузией, *489*
с предварительным монтажом, *495*
шариковых выводов, *92*
Межсоединение, *82*
асинхронное, *599*
глобальное, *228*
задержка инвертора, *274*
локальное, *228*
мезохронное, *598*
плезиохронное, *598*
программируемое, *504*
синхронное, *597*
уровни, *87*
Мезохронное межсоединение, *598*
Метастабильная точка, *409*
Метод перевернутого кристалла, *89*
Метод поверхностного монтажа, *91*
Микрон, *82*
Микросхема, *177*
Миниатюрный корпус, *92*
Минимальная ширина линии, *81*
Многокристальный модуль, *92*
Многопортовая память, *756*
Многорегистровое циклическое сдвиговое устройство, *717*
МОП-транзистор, *161*
в статических условиях, *130*
отклонение процесса изготовления, *169*
полевой, *129*
МОП-элемент
комплементарный, *306*
Моделирование, *155*
на уровне логических элементов, *389*

на уровне переключателей, *388*
функциональное, *389*
Модель, *35*; *114*
линии передачи, *215*
масштабирования с фиксированным напряжением, *174*
с единичными задержками, *388*
с нулевой задержкой, *388*
Модуляция длины канала, *136*
Море элементов, *489*
Мощность, *275*
анализ потребления, *292*
динамическая, *358*
динамическое потребление, *275*
потребление, *383*
рассеянье, *333*
технологии уменьшения, *725*
уменьшение в холостом режиме, *743*
управление во время выполнения, *737*
Мультивибратор, *407*
ждущий, *448*; *451*
Мультиплексор, *411*

Н

Накопление частичных произведений, *710*
Направленность, *55*
Напряжение, *48*
гистерезиса, *449*
порога вентиля, *50*
пороговое, *130*
пробоя, *126*
смещения, *178*
Насыщение скорости, *137*
Носитель кристаллов, *92*

О

Одновибратор, *448*
ОЗУ, *754*
Оконечная нагрузка, *218*; *576*
Оптическая коррекция масок, *73*
Оптическая маска, *68*; *100*
Осаждение, *74*
Отжиг, *74*
Отмывка, *72*
Охранное кольцо, *164*
Ошибка проектирования, *40*

П

Память, *407*
динамическая, *407*; *425*
статическая, *407*; *425*
фоновая, *407*
Паразитная индуктивность, *568*
Паразитное сопротивление, *557*
Перемычка; *506*
Переходные процессы, *121*
ПЗУ, *754*
Планаризация, *75*
Пластмассовый носитель кристалла, *92*
Плата, *90*
Плезиохронное межсоединение, *598*
Побитовая организация, *677*
Повторитель, *564*
Подзатворный окисл, *76*
Подложка, *129*
нормы проектирования контактов, *83*
Поликремний, *75*; *82*
Помехоустойчивость, *48*
Последовательная схема, *50*
небистабильная, *448*
с одним устойчивым состоянием, *451*
Потенциал Ферми, *132*
Предпороговая область, *143*
Пробой
зеннеровский, *126*
лавинный, *126*
Проводимость, *143*
обратная, *417*
Проводник
идеальный, *205*
электрический, *204*
Программируемая логика, *497*
матричная, *498*
Программируемая матрица межсоединений, *509*
Программируемое логическое устройство, *499*
Прокол, *162*
Процент выхода годных кристаллов, *45*
Процесс самосовмещения, *77*

Р

Размах сигнала, *49*
Рассеяние
вызванное токами прямого пути, *282*
динамическое, *275*
емкостное, *278*
Регенеративность, *51*
Регистр, *407*

"master–slave", *413*
двухфронтовый, *431*
динамический, *425*
дифференциальный, *625*
импульсный, *437*
однофазный тактируемый, *432*
основанный на усилителях считывания, *440*
псевдостатический, *419*
сдвиговый, *716*
статический, *409*
ТКМОП, *428*
управляемый отрицательным фронтом, *408*
управляемый положительным фронтом, *408*
Резервное одалживание, *626*
Резистивная область, *135*
Резистивный режим, *133*
Реконфигурация, *497*
Релаксационный подход, *386*

С

Сигнатурный анализ, *879*
Силицид, *96*
Силицидация, *160*
Синдром, *880*
Синхронизация, *44*; *646*; *652*
Синхронное динамическое ОЗУ, *763*
Синхронное межсоединение, *597*
Синхронное проектирование, *665*
Синхронное тактирование, *600*
Система
автоматизированного проектирования, *37*
автоподстройки по задержке, *661*
межсоединений, *44*
разводки питания, *42*
разработка, *753*
распределения тактовых импульсов, *40*
синхронная, *404*
управляемая положительным фронтом, *600*
Сквозное отверстие, *82*
Сопротивление, *150*; *198*
входное, *238*
истока/стока, *159*
контактное, *199*
межсоединений, *198*
паразитное, *557*
перехода, *199*
поверхностное, *160*
тракта, *208*
удельное поверхностное, *198*
Соответствие, *808*
Состояние, *404*
Состязание, *419*
СППЗУ, *755*
Среднее время наработки на отказ, *651*
Стабилизатор, *374*
Стандартная ячейка, *38*
Статическая составляющая, *62*
Статическая характеристика, *49*
Статические ОЗУ, *792*
Степень интеграции, *31*; *96*
Сток, *129*
Стробирование тактовых импульсов, *614*
Структурное описание, *529*
Ступенчатый переход, *115*
Сумматор, *678*
Брента–Канга, *704*
двоичный, *678*
зеркальный
линейный
на передаточных элементах, *685*
полный, *682*
с выбором сигнала переноса, *694*
с манчестерской (дифференциальной) цепочкой переноса, *685*
с обводом переноса, *689*
с пропуском переноса, *690*
с ускоренным переносом, *678*;
слияния векторов, *712*
со сквозным переносом, *679*
статический, *682*
четверичный, *704*
Суммирование
конечное, *715*
Сушка, *72*
Схема
автоколебательная, *407*; *448*
бистабильная, *407*; *410*
восстановления уровня, *344*
входа-выхода, *466*
динамическая, *301*
моностабильная, *407*
нерегенеративная, *300*
псевдостатическая, *428*
разработка, *753*

с внутренней синхронизацией, *629*
синхронизации, *646*; *647*
статическая, *301*
тактирования, *456*
цифровая интегральная, *463*
электростатической защиты, *555*
Схема “домино”
с задержкой тактового сигнала, *646*
Схема ФАПЧ, *656*

Т

Такт, *405*
Тактирование
на основе защелок, *625*
синхронное, *600*
Тактируемая КМОП-схема, *428*
Тактовый сигнал, *417*
оптическое распределение, *663*
Тепловой потенциал, *117*
Тепловой поток, *95*
Теплопроводность, *95*
Тест, *873*; *875*
IDDQ, *884*
Тестовый шаблон, *868*
Тонкопленочные транзисторы, *800*
Топология, *84*
для сложных логических элементов, *397*
с использованием путей Эйлера, *398*
согласно Уэйнбергеру, *397*
Травление, *72*; *75*
Транзистор, *31*; *48*
вертикальный, *176*
нормы проектирования, *83*
описание, *166*
проходной, *338*
Триггер, *409*; *411*
характеристическая таблица, *422*
Шмитта, *448*
Трудозатраты
логические, *318*
на ветвление, *321*
электрические, *318*
Туннелирование Фаулера–Нордхейма, *784*

У

Умножитель, *707*; *747*
в форме дерева Уолласа, *715*
древовидный (каскадный), *714*
матричный, *708*; *711*
определения, *707*
с записью переноса, *712*
Утечка
заряда, *363*
подпороговая, *364*

Ф

Фазовая автоподстройка частоты (ФАПЧ), *652*
Ферми потенциал, *132*
Ферроэлектрическое ОЗУ, *789*
Флэш-память, *755*
Фотолитография, *70*
Фоторезист, *70*
негативный, *71*
проявление и дубление, *72*
удаление, *72*
Фотошаблон, *68*; *100*

Э

Экранирование, *578*
Экспонирование, *71*
Электромиграция, *560*
Электростатический разряд, *555*
Элемент
относительный логический, *238*
передаточный логический, *348*
скомпилированный, *478*
статический, *342*
с одним устойчивым состоянием, *451*
Элмора формула, *207*
Энергия, *62*
задержка, *275*
режим пониженного потребления, *289*
Эпитаксиальный слой, *70*
Эффект узкого канала, *162*
Эффективная длина канала, *135*
Эффективная емкость, *747*
Эффективная ширина канала, *135*
Эффективность
использования энергии, *468*
собирания, *840*
Эффективный коэффициент
разветвления по выходу, *268*

Научно-популярное издание

Жан М. Рабаи, Ананта Чандракасан, Боривож Николич

Цифровые интегральные схемы
2-е издание

Литературный редактор *И.А. Попова*
Верстка *А.В. Назаренко*
Художественные редакторы *Е.П. Дынник, Т.А. Тараброва*
Корректор *А.В. Назаренко*

Издательский дом "Вильямс"
127055, г. Москва, ул. Лесная, д. 43, стр. 1

Подписано в печать 20.06.2007. Формат 70x100/16.
Гарнитура Times. Печать офсетная.
Усл. печ. л. 73,53. Уч.-изд. л. 52,18.
Тираж 2000 экз. Заказ № 0000.

Отпечатано по технологии CtP
в ОАО "Печатный двор" им. А. М. Горького
197110, Санкт-Петербург, Чкаловский пр., 15.