

Содержание

Введение.....	6
1. Основные ограничения создания высокопроизводительных микропроцессоров ...	8
1.1. Факторы снижения надежности микросхем	8
1.2. Проблемы повышения потребляемой мощности микросхем	9
1.3. Методы снижения динамического тока потребления	13
1.3.1. Снижение напряжения питания и уменьшение проектных норм.....	13
1.3.2. Динамическое управление частотой и напряжением питания	14
1.3.3. Включение дополнительных тактов останова в конвейер.....	14
1.3.4. Использование нескольких напряжений питания на одном кристалле	15
1.3.5. Применение технологии «кремний на изоляторе»	15
1.3.6. Метод селективного отключения синхросигнала.....	17
1.3.7. Использование средств САПР	19
1.3.8. Принцип понижения энергопотребления адиабатических логических элементов	20
1.3.9. Использование обратимой логики	21
1.4. Синхронная и асинхронная логика, самосинхронные схемы	28
1.5. Ограничения, связанные с предельными технологическими нормами производства микросхем.....	39
2. Основные радиационные эффекты в микросхемах	41
2.1. Эффекты накопленной дозы	41
2.1.1. Эффекты накопленной дозы в подзатворном оксиде.....	42
2.1.2. Эффекты накопленной дозы в оксиде по краям затвора.....	46
2.1.3. Эффекты накопленной дозы в полевом оксиде.....	47
2.2. Одиночные сбои	47
2.3. Множественные сбои.....	49
2.4. Множественные сбои и биполярный эффект.....	50
2.5. Кратковременные переходные процессы	51
2.6. Одиночное защелкивание.....	52
2.7. Модель кратковременного переходного процесса.....	55
3. Основные методики повышения надежности микросхем	56
4. Тестирование и верификация микросхем	59
4.1. Методы функциональной верификации RTL-моделей микропроцессоров	60
4.2. Метод коверификации микропроцессоров.....	72
4.3. Тестирование микропроцессоров	74
4.4. Методы, применяемые для верификации с использованием специализированных аппаратных устройств.....	79
4.5. Валидация кристаллов микросхем	82

4.6. Метрики тестового (функционального) покрытия, применяемые при верификации моделей.....	84
5. Маршрут проектирования микросхем	89
5.1. Классификация микросхем	93
5.2. Маршрут разработки БМК и БК	97
5.3. Маршрут проектирования полузаказных БИС на БМК и БК	98
5.4. Маршрут проектирования заказных схем	99
5.5. Разработка полностью заказных узлов	110
6. Архитектура микропроцессоров	118
6.1. Расположение и порядок битов и байтов	118
6.2. Синхронизация асинхронных сигналов.....	121
6.3. Производительность вычислительных систем.....	123
6.4. Классификация микропроцессоров.....	124
6.5. CISC-процессоры.....	126
6.6. RISC-процессоры.....	126
6.7. VLIW- и EPIC-процессоры.....	131
7. Организация подсистемы памяти.....	134
7.1. Методы доступа к памяти	134
7.2. Иерархия подсистемы памяти	136
7.3. Полностью ассоциативная кеш-память	140
7.4. Индикаторы состояния строки.....	144
7.5. Кеш-память с прямым отображением памяти	146
7.6. Кеш-память, ассоциативная по множеству	148
7.7. Многопортовость кеш-памяти	149
7.8. Когерентность системы памяти	151
7.9. TLB и виртуальная память	153
8. Основные архитектуры параллельных ЭВМ	156
8.1. Симметричные мультимикропроцессорные системы.....	157
8.2. Системы с неоднородным доступом к памяти.....	157
8.3. Системы с массовым параллелизмом.....	158
8.4. Кластерные системы	160
8.5. Неоднородные системы	161
8.6. Программное обеспечение параллельных компьютеров.....	162
8.6.1. Модель SPMD	163
8.6.2. Модель MPMD.....	164
8.6.3. Межпроцессные взаимодействия	165
8.6.4. Стандартные средства поддержки межпроцессных взаимодействий в ОС UNIX	165
9. Компьютерные сети	168
9.1. Классификация компьютерных сетей.....	169
9.2. Основные типы топологий сетей.....	174

9.2.1. Топология «точка-точка»	174
9.2.2. Топология «шина»	174
9.2.3. Топология «звезда»	175
9.2.4. Топология «полносвязная сеть»	176
9.2.5. Топология «частичносвязная сеть»	177
9.2.6. Топология «дерево»	177
9.2.7. Топология «трехмерный тор»	178
9.2.8. Топология «решетка»	178
9.2.9. Комбинация топологий	178
9.3. Протоколы локальных вычислительных сетей	181
9.3.1. Классификация локальных сетей	181
9.3.2. Типы и методы доступа	181
9.3.2.1. Протоколы с разделением канала	182
9.3.2.2. Протоколы случайного доступа	183
9.3.2.3. Протоколы с передачей права	184
9.3.3. Основные сетевые протоколы	185
9.3.3.1. Asynchronous Transfer Mode (ATM)	185
9.3.3.2. Ethernet	185
9.3.3.3. Token Ring	190
9.3.3.4. USB	190
9.3.3.5. IEEE 1394 Fire Wire Serial Bus	206
9.3.3.6. CAN	207
9.4. Основные типы коммуникационных сред	209
9.4.1. HyperTransport (HT)	210
9.4.2. PCI Express. ASI	212
9.4.3. Ethernet 10Gbit	215
9.4.4. VME extensions. VXS	215
9.4.5. StarFabric	216
9.4.6. InfiniBand	217
9.4.7. Myrinet	219
9.4.8. RapidIO	220
Заключение	253
Литература	255

Введение

Вычислительные системы в современном мире во многом определяют прогресс общества, удобство и комфорт жизни людей. От того, насколько надежными, быстродействующими и удобными в пользовании являются вычислительные системы, зависят комфорт и качество нашей жизни. Производительность является основной характеристикой вычислительных систем. Для того чтобы получить требуемые высокопроизводительные системы, требуется решение комплекса мер по поднятию производительности с одновременным учетом таких факторов, как ограничения по потребляемой мощности, стоимости, размеру кристалла, надежности функционирования.

Основным фактором повышения производительности ЭВМ является создание высокопроизводительного микропроцессора. Производительность микропроцессора определяется рядом факторов: частотой его функционирования, архитектурой, возможностью одновременного выполнения нескольких инструкций, обработки нескольких потоков данных, наличием нескольких ядер, архитектурными особенностями (например возможностью аппаратного выполнения сложных функций, таких как $\sin x$, бабочка и пр.), наличием большого объема встроенной кеш-памяти различного уровня (встроенная подсистема памяти) и системой команд.

Следующим фактором является организация подсистемы памяти в целом. Особенно острой является проблема повышения скорости обмена данными с ОЗУ для задач с большими объемами данных, не позволяющих эффективно использовать кеш-память. В таких случаях ускорение возможно прежде всего за счет введения режимов прямого доступа к памяти (*DMA*), предвыборки, увеличения частоты и ширины памяти, создания нескольких каналов памяти.

Для систем реального времени и многопроцессорных систем важны аппаратное выполнение функций синхронизации процессов, организация прерываний и контроля функционирования вычислительной системы и ее отдельных узлов.

Большие потери производительности (до нескольких раз для отдельных задач) вычислительных систем происходят при обмене между отдельными микросхемами в силу ограничений по скорости передачи данных по плате. Поэтому интеграция функций на одном кристалле и создание систем на кристалле приводят не только к повышению надежности системы и уменьшению ее габаритов в силу уменьшения числа компонент, но и к повышению производительности системы в целом.

При создании систем в корпусе (*SiP*) возможна оптимизация выводов кристаллов с учетом их характеристик и числа в микросборке и соответствующее снижение емкости контактов. Это приводит к уменьшению потребления питания и увеличению скорости обмена данными.

Уменьшение числа мостовых схем приводит к уменьшению задержки на передаваемые данные.

Создание комплектов СБИС под вычислительные системы позволяет оптимизировать потребляемую мощность, габариты и повысить производительность системы.

Введение сопроцессоров под выделенные задачи позволяет в несколько раз поднять производительность систем на таких задачах. Для примера: в современных коммуникационных микропроцессорах имеются десятки встроенных сопроцессоров, аппаратно выполняющих функции сжатия данных, криптографии, контроля, *TSP* и пр. [1].

Следующим шагом повышения производительности является создание многопроцессорных модулей и систем. Эффективность многопроцессорной системы во многом зависит от решаемых задач (возможности параллельного выполнения отдельных частей программы) и программного обеспечения, позволяющего осуществлять параллельное выполнение задач. Особая роль в таких системах с точки зрения достижения требуемых параметров производительности отводится коммуникационным системам.

Повышение производительности вычислительных систем сдерживается рядом факторов. Переход на глубокие субмикронные технологические нормы изготовления кристаллов микросхем приводит к снижению надежности, значительному росту токов утечки и статическому потреблению. Создание требуемых высокопроизводительных микропроцессорных систем невозможно без разработки новых методов проектирования микросхем, создания новых систем проектирования и верификации микросхем.

1. Основные ограничения создания высокопроизводительных микропроцессоров

Проблемы создания современных высокопроизводительных микропроцессоров можно разбить на две основные группы:

- усложнение процесса проектирования (повышение сложности проектов приводит к необходимости повышения числа и уровня подготовки специалистов, улучшения качества средств проектирования, формирования новых маршрутов проектирования, включая тестирование и верификацию проектов, создания соответствующей инфраструктуры, включающей мощные сервера, высокочастотные и многоканальные приборы, средства тестирования и верификации, в том числе ускорители, и пр.);

- физические ограничения при создании сложных проектов и переходе на предельные технологические нормы.

Из физических ограничений необходимо выделить три основных фактора: снижение надежности, повышение потребляемой мощности и ограничения, связанные с предельными технологическими нормами производства микросхем. Особенно острой задача становится при создании микропроцессоров для высокопроизводительных серверов и супер-ЭВМ. В США уже сейчас вычислительные центры потребляют 2 % всей энергии. С удвоением этого показателя каждые пять лет суперкомпьютеры в 2100 году теоретически будут нуждаться во всем вырабатываемом в Штатах электричестве.

1.1. Факторы снижения надежности микросхем

При использовании предельных технологических норм изготовления кремниевых пластин, необходимых для создания высокопроизводительных ЭВМ, невозможно добиться высокой надежности функционирования микросхем без использования комплекса мер повышения надежности. Это обусловлено:

- снижением напряжения питания и соответствующим уменьшением возможного разброса порогов логических уровней, повышением требований к качеству питания;

- уменьшением толщин диэлектриков и соответствующим снижением надежности, повышением токов утечки, снижением величин паразитных зарядов (порогов), приводящим к сбою или отказу микросхем;

- уменьшением сечений проводников, что приводит к сокращению времени функционирования микросхем за счет эффектов электромиграции и снижению механической прочности;

- большим количеством элементов, статистически приводящим к повышению вероятности сбоя микросхем;

- увеличением поражающего фактора для одиночных и локальных радиационных эффектов, связанного с существенно меньшими топологическими размерами активных элементов в сравнении с размерами трека поражающей частицы;

- существенным влиянием фактора микродозовых радиационных эффектов на надежность и долговечность схем, обусловленным структурными изменениями в приборных областях схем при воздействии высокоэнергетичных частиц и потоков атмосферных нейтронов.

Как видно, многие проблемы возникают из-за влияния различных радиационных эффектов. Если при технологических нормах 0,13 мкм и выше методы снижения влияния радиационных эффектов использовались только для микросхем специального применения, то сейчас эти методы используются и развиваются и при создании высокопроизводительных микросхем ответственного коммерческого применения (например, для серверов и супер-ЭВМ, управления транспортом и промышленными предприятиями и пр.) с нормами ниже 90 нм. Именно поэтому эти эффекты будут рассматриваться ниже.

1.2. Проблемы повышения потребляемой мощности микросхем

Как было отмечено выше, одним из основных сдерживающих факторов роста производительности микросхем является рост потребляемой мощности. С точки зрения подходов к минимизации энергии, потребляемой при функционировании логических элементов, можно ввести следующую классификацию составляющих частей потребляемой энергии:

- рассеивание энергии из-за токов утечек в статическом режиме;
- рассеивание энергии из-за протекания токов в динамическом режиме;
- энергия формирования логического состояния элемента.

Потребляемую элементом мощность можно рассчитать исходя из следующей формулы:

$$P = P_{in} + P_{st} + P_1, \quad (1.1)$$

где
$$P_1 = \frac{C \times U^2}{2} \times K \times f, \quad (1.2)$$

$$P_{in} = E(t_f C) \times K \times f, \quad (1.3)$$

P_{in} – внутренняя мощность элемента, определяемая сквозными токами и определяемая как энергия переключения элемента $E(t_f C)$, которая зависит от фронта сигнала на входе и от нагрузки, умноженная на коэффициент переключений (K) и на частоту (f);

P_{st} – мощность, потребляемая в статическом режиме;

P_1 – мощность, затрачиваемая на перезаряд нагрузки, определяемая энергией перезаряда емкости нагрузки, умноженная на коэффициент переключений (K) и на частоту (f).

Правильный выбор коэффициента K во многом определяет правильность расчета потребляемой мощности. Он может быть выбран следующими способами:

- $K = const$ и определяется из статистики;
- $K = var$ и определяется из конкретного теста конкретной микросхемы;
- $K = var$ и определяется эвристическими алгоритмами вычисления вероятности переключения, определяемыми САПР [2].

Статическое рассеивание определяется суммарными токами утечки закрытых n - и p -канальных транзисторов при отсутствии тактирующих сигналов, когда схема находится в определенном, стабильном состоянии. Для норм выше 65 нм цифровых КМОП-схем влияние токов в статическом состоянии можно считать пренебрежимо малым и не оказывающим значительного влияния на общую по-

требляемую мощность [3]. Для норм 65 нм токи в статическом состоянии становятся уже заметными, а для норм 45 нм и ниже значения токов утечки становятся значительными.

Токи утечки эффективно снижаются при создании специальных технологических процессов *Low Power*, основанных на изменении режимов работы транзисторов и повышении порога его переключения, соответственно используется специальная библиотека логических элементов на основе транзисторов с разными порогами. МОП-транзисторы с пороговым напряжением, увеличенным на 120–150 мВ, имеют ток утечки в 10–20 раз меньше тока транзисторов с номинальным пороговым напряжением 0,25–0,3 В. Увеличение порогового напряжения достигается с помощью увеличения толщины подзатворного диэлектрика и изменения профиля легирования кармана. Однако это приводит к снижению быстродействия и нагрузочной способности транзисторов.

Схемотехнические методы снижения энергии статического потребления направлены на снижение токов утечек, главным образом за счет последовательного включения в стоковую цепь элементов дополнительных транзисторов, которые ограничивают величину тока стока. Подобные меры по ограничению тока стоковой цепи способствуют увеличению времени переключения элемента из одного логического состояния в другое. Для решения этой проблемы применяются различные схемы управления дополнительными транзисторами, которые в зависимости от режима работы элементов (частоты изменения его выходных логических состояний, продолжительности нахождения в каком-либо одном логическом состоянии) задают повышенный или пониженный ток стока элемента. Подобные подходы имеют наибольшую эффективность снижения энергии потребления в статическом режиме (в 2–5 раз), если частота переключения логических элементов не превышает нескольких сотен мегагерц [4].

Динамическое рассеивание происходит в моменты переключения в логических элементах схемы. Основная причина наличия динамического тока в КМОП-схемах – это перезаряд внутренних емкостей СБИС, величина которых определяется размерами топологических элементов, формируемых в процессе создания транзисторов и межсоединений [5]. Заряд и разряд емкостей осуществляются через коммутируемые *n*- и *p*-канальные транзисторы, в которых происходит рассеивание потребляемой динамической энергии.

На рис 1.1 показана эквивалентная схема заряда и разряда выходной емкости C_L при переключении КМОП-элемента. Ключ моделирует циклы заряда и разряда, V – идеальный источник напряжения питания, R_s и R_d – эквивалентные сопротивления открытых *n*- и *p*-канальных транзисторов.

Если заряжать и разряжать конденсатор C_L с частотой f циклов в секунду, общая мощность рассеивания системы составит

$$P_l \sim C_L \times V^2 \times f. \quad (1.4)$$

Уравнение 1.4 служит для оценки мощности рассеивания цифровых КМОП-схем.

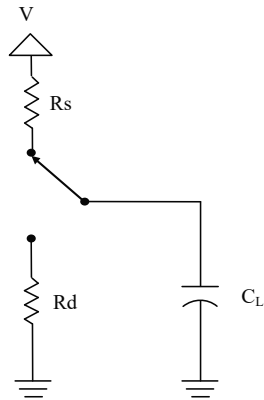


Рис. 1.1. Эквивалентная схема заряда и разряда конденсатора

Необходимо отметить, что при этом были сделаны следующие допущения: емкость нагрузки C_L и напряжение питания V постоянны, емкость успевает полностью зарядиться и разрядиться.

Если частота переключения f и напряжение питания одинаковы для всех этих узлов, общую мощность потребления P можно представить как

$$P_i \sim C_{total} \times V^2 \times f, \quad (1.5)$$

где C_{total} — эквивалентная емкость, определяемая суммарной величиной емкостей всех узлов СБИС, заряжаемая и разряжаемая в процессе функционирования микросхемы.

Все существующие методы снижения энергопотребления цифровых схем направлены на уменьшение величины параметров, входящих в формулу (1.5).

По каждому из варьируемых параметров необходим поиск его оптимального значения, поскольку выигрыш по рассеиваемой мощности сопровождается определенными потерями по другим параметрам микросхем или системным характеристикам.

Например, при уменьшении топологических размеров элементов структур и, соответственно, уменьшении величины емкостей и увеличении степени интеграции уменьшаются коммутируемые заряды и потребляемая динамическая мощность СБИС. Однако уменьшение величины заряда в емкости элемента хранения информации приводит к уменьшению устойчивости данного элемента, так как меньший по величине паразитный заряд вызовет сбой этого элемента. Определение оптимальных для конкретных применений параметров библиотечных элементов, в том числе емкостей элементов памяти, требует проведения исследований условий работы с учетом сбое- и отказоустойчивости СБИС.

Квадратичная зависимость от напряжения питания V определяет высокую эффективность уменьшения V для снижения потребляемой мощности. Это определяет тенденцию уменьшения напряжения питания до значений ниже 1,5 В. Но такое решение возможно только при условии уменьшения величины пороговых напряжений транзисторов, что вызывает снижение помехоустойчивости и, следовательно, ужесточает требования к условиям работы системы.

Величина напряжения питания, необходимая для выполнения операций с минимальной или максимальной скоростью, может отличаться в несколько раз. Это позволяет использовать динамическое управление величиной напряжения питания СБИС или даже ее отдельных узлов в разных режимах работы для уменьшения энергопотребления. Для повышения быстродействия требуемых блоков в ряде случаев используют:

- локальное повышение напряжения питания,
- уменьшение пороговых напряжений транзисторов до предельно допустимых значений, ограниченных возрастанием токов утечек, соизмеримых со значением динамических токов.

Иногда повышение быстродействия СБИС с одновременным снижением величины коммутируемых зарядов, а значит, и уменьшением потребления, достигается уменьшением перепада напряжения заряженной и разряженной емкости (неполного перезаряда).

Отдельное направление занимает снижение энергопотребления с помощью программных средств [6,7], а также с помощью технологических приемов, например за счет применения технологии «кремний на изоляторе» [8], которая характеризуется меньшими значениями емкостей элементов топологических структур, изолированных слоем диэлектрика.

Учитывая вышесказанное, можно провести классификацию методов снижения энергопотребления следующим образом:

- методы, направленные на снижение напряжения питания;
- методы, направленные на снижение суммарной емкости схемы;
- методы, направленные на снижение частоты переключения.

Имеются также методы снижения потребляемой энергии в динамическом режиме, рассеиваемой транзисторами в тепло, которые выделены в отдельное направление проектирования адиабатических элементов. Разработка адиабатических элементов включает в себя специальную схемотехнику и организацию функционирования, благодаря которым исключается протекание сквозных токов в динамическом режиме и снижается количество рассеиваемой энергии при смене логического состояния. Причем на пониженных частотах функционирования (десятки—сотни МГц) эффективность снижения рассеиваемой энергии при изменении логического состояния у адиабатических элементов значительно выше, чем у обыкновенных статических КМОП-элементов.

Одним из физических ограничений на миниатюризацию электронных схем является также принцип Ландауэра, согласно которому логические схемы, не являющиеся обратимыми, должны выделять тепло в количестве, пропорциональном количеству стираемых (безвозвратно потерянных) данных. Возможности же по отводу тепла физически ограничены.

Энергия формирования логического состояния определяется величиной емкости выходного узла, на котором электрический заряд создает разность потенциалов между выходом и шиной питания или общей шиной. Смена логического состояния приводит к рассеиванию этой энергии в тепло на транзисторах элемента. Процесс переключения является необратимым, и для формирования очередного логического состояния требуется такое же количество энергии. Задача по сохранению энергии логического состояния заключается в организации такой

работы логической схемы, при которой информация о входных сигналах не теряется в процессе совершения логических операций и для определения значения входных сигналов не требуется затрачивать дополнительную энергию на формирование новых логических уровней. Логический элемент, обладающий такими функциональными свойствами, называется обратимым. Очевидно, использование обратимой логики имеет и отрицательные стороны, связанные с увеличением числа логических элементов, соответствующим увеличением динамического и статического потребления и площади кристалла.

В отличие от обратимых элементов методы понижения динамического энергопотребления, применяемые в адиабатических элементах, направлены на минимизацию рассеивания энергии логического состояния в тепло, но не на сохранение логического состояния. Это означает, что основная проблема обратимых вычислений заключается в исследовании причин потери логических состояний при совершении логических операций (смена логического состояния с потерей информации об этом логическом состоянии), а основная проблема адиабатических вычислений заключается в минимизации последствий смены логического состояния (рассеивание энергии логического состояния в тепло).

В последние годы идут активные работы по созданию памяти на новых физических принципах, включая резистивную, сегнетоэлектрическую, магнитную память, память на основе фазовых переходов (*ReRAM*, *FeRAM*, *MRAM*, *PCM*) и пр. Исследования направлены на повышение объема памяти, скорости доступа к ней, создание энергонезависимой памяти с характеристиками статической памяти, повышение надежности и радиационной стойкости, снижение энергии формирования логического состояния. Учитывая значительный рост объемов встроенной памяти, создание надежной памяти с малой энергией формирования логического состояния позволит существенно снизить потребляемую мощность микросхем, соответственно, снизится актуальность развития обратимой логики.

1.3. Методы снижения динамического тока потребления

1.3.1. Снижение напряжения питания и уменьшение проектных норм

В соответствии с формулой (1.5) мощность потребления имеет квадратичную зависимость от напряжения питания, поэтому наибольшую эффективность дают методы, ориентированные на снижение напряжения. Однако простое снижение приводит к уменьшению производительности схемы. Так, по данным [9], снижение напряжения питания на 10 % приводит к снижению быстродействия в среднем на 10 %. Поэтому для высокопроизводительных схем всегда необходимо использование комплекса мер по сохранению быстродействия. Самым распространенным методом снижения напряжения при одновременном сохранении и даже повышении быстродействия является переход на улучшенные топологические проектные нормы. Применение новой технологии не только позволяет компенсировать снижение быстродействия, связанное с уменьшением напряжения, но и приводит к существенному повышению производительности СБИС. При переходе на новый уровень технологии уменьшается еще один параметр формулы (1.2) — эквивалентная емкость нагрузки. Это происходит как за счет того, что базовые элементы имеют меньшие размеры и меньшую емкость, так и за счет того, что

общий размер кристалла становится меньше при сохранении выполняемых функций СБИС и, соответственно, уменьшаются длина и емкость межсоединений.

Тем не менее данный метод имеет следующие недостатки:

- стоимость изготовления микросхем значительно увеличивается (примерно в 2 раза при уменьшении технологических норм в 1,4 раза). Особенно это заметно для малых партий, когда изготавливается несколько пластин и стоимость накладных расходов (стоимость масок, подготовка к производству и пр.) получается значительной;
- изменение напряжения питания требует соответствующего изменения схемотехники вычислительных систем, в которых установлена исходная микросхема;
- если оптимизируется микропроцессор, то во многих случаях требуется перепроектирование контроллеров, дополняющих этот микропроцессор в вычислительных системах;
- статический ток заметно возрастает начиная с норм 65 нм.

1.3.2. Динамическое управление частотой и напряжением питания

Динамическое управление частотой и напряжением питания для снижения энергопотребления широко используется при проектировании вычислительных систем на базе микропроцессоров [10]. Метод заключается в том, что специальная программа анализирует текущую загрузку микропроцессора и, если текущая загрузка не требует максимальной производительности, снижает напряжение питания и рабочую частоту. Все наиболее известные производители микропроцессоров для персональных компьютеров используют данный метод в своих системах. Для микропроцессоров компании *Intel* данный метод известен как *SpeedStep* [11], для компании *AMD* – как *AMD PowerNow!™ Technology* [12], для *Transmeta* – как *Transmeta LongRun Power Management* [13]. Эффект, достигаемый применением данного метода, зависит от выполняемой задачи.

Преимущество метода заключается в том, что снижение энергопотребления происходит без потери реальной производительности для пользователя. Данный метод имеет следующие недостатки:

- конструкция микропроцессора должна обеспечить стабильную работу в условиях переменного напряжения питания и тактовой частоты;
- требуются специальные аппаратные ресурсы на уровне вычислительной системы, такие как программно-управляемый источник напряжения. Контроллеры, входящие в состав микропроцессорной системы, должны обеспечивать работу с изменяемой временной диаграммой со стороны микропроцессора и при этом сохранять требуемые временные интервалы для внешних устройств: таймеров, контроллеров последовательных портов и другие;
- требуется специальное программное обеспечение, которое должно определять необходимую достаточность текущих значений напряжения и частоты и, в случае необходимости, управлять их изменением;
- сложно прогнозировать время выполнения программ, что ограничивает применение метода в системах реального времени.

1.3.3. Включение дополнительных тактов останова в конвейер

Оригинальный способ контроля за энергопотреблением применен в микропроцессорах G3 и G4 семейства *PowerPC* [14, 15]. Данные микропроцессоры содер-

жат блок термического сопровождения (*Thermal Assist Unit*), в функцию которого входит контроль за температурой кристалла. В случае превышения допустимой нормы блок дает специальное прерывание в устройство управления, реакцией на которое является формирование тактов останова, что приводит к снижению частоты доступа к кеш-памяти команд. При этом микропроцессор потребляет меньше энергии и, соответственно, начинает охлаждаться. Метод относится к группе методов, направленных на снижение частоты переключения, и имеет следующие преимущества по сравнению с понижением частоты или снижением напряжения питания:

- поскольку рабочая частота не меняется, возможно использование традиционной схемы автоподстройки частоты;
- включение тактов останова происходит без участия программного обеспечения;
- не требуются дополнительных аппаратных ресурсов на уровне вычислительной системы.

Данный метод имеет следующие недостатки:

- требуется включение в конструкцию микропроцессора дополнительных элементов, в том числе аналоговых;
- снижение энергопотребления происходит за счет снижения быстродействия;
- сложно прогнозировать время выполнения программ, что ограничивает применение метода в системах реального времени.

1.3.4. Использование нескольких напряжений питания на одном кристалле

Применение нескольких источников напряжения в одном кристалле, описанное в [16], основано на том, что для любой схемы имеются блоки, которые не являются критичными с точки зрения быстродействия. Соответственно, требования к быстродействию таких блоков ниже, чем для блоков, определяющих производительность СБИС, и для них возможно снижение напряжения питания. В [17] приведен пример СБИС, в которой реализован данный метод. Наиболее критичные блоки памяти команд и регистровый файл используют напряжение 2,5 В, тогда как остальные блоки питаются от пониженного источника 1,75 В.

Для данной СБИС применение отдельных источников позволило снизить энергопотребление на 30 % без снижения производительности.

Метод имеет следующие недостатки:

- значительно повышается трудоемкость разработки топологии микросхемы за счет того, что требуется реализация двух или более шин питания;
- необходима установка дополнительных преобразователей уровней между блоками с разным напряжением питания, что вносит дополнительную задержку;
- предъявляются дополнительные требования к конструкции печатной платы и к источникам питания;
- в процессе проектирования требуется использование нескольких библиотек стандартных ячеек;
- необходимо наличие специального технологического процесса.

1.3.5. Применение технологии «кремний на изоляторе»

Технология «кремний на изоляторе» (КНИ) обладает рядом преимуществ по сравнению с традиционной технологией объемного кремния и получила ши-

рокое распространение в конце 90-х годов. В частности, компании *IBM* и *AMD* выбрали эту технологию как основную для линии своих микропроцессоров. Одним из преимуществ данной технологии является пониженное энергопотребление. На рис. 1.2 показана конструкция транзистора, выполненного по технологии КНИ.

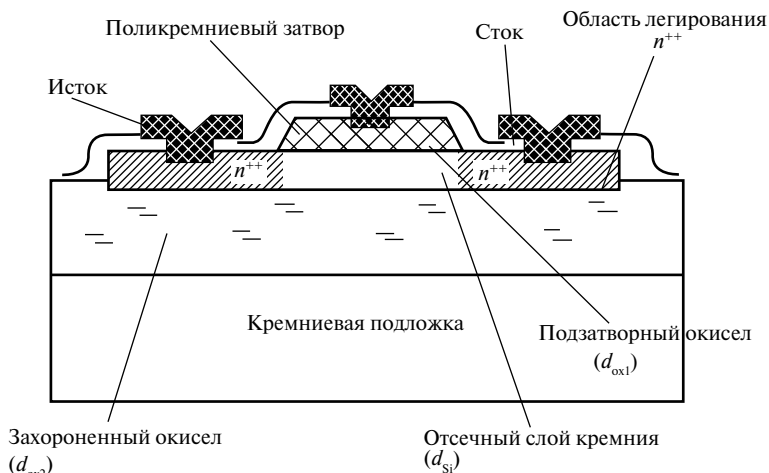


Рис. 1.2. Пример транзистора на структуре КНИ

Площадь изолирующего p - n -перехода для КНИ-транзистора значительно меньше, чем для традиционного транзистора. Соответственно, меньшую емкость требуется заряжать при каждом переключении.

Из литературы следует, что применение КНИ дает до 50 % снижения энергопотребления [18]. Микросхемы, созданные в России по технологии КНИ, потребляют ниже аналогичных микросхем, изготовленных по технологии КМОП, на 20–25 % при увеличении быстродействия до 40 % [19].

Данный метод имеет следующие недостатки:

- высокая стоимость исходного материала: КНИ-пластина стоит на порядок больше пластины на объемном кремнии;
- отсутствие широкодоступных библиотек стандартных ячеек и *IP*-блоков для проектирования цифровых схем для КНИ;
- сложность с размещением заказа на изготовление на зарубежной фабрике.

Следует также отметить, что существует два направления развития КНИ-технологии: создание радиационно-стойких микросхем и создание быстродействующих и низкопотребляющих микросхем. Для радиационно-стойкой технологии необходимо обеспечить устойчивую работу микросхемы при прохождении высокоэнергетической заряженной частицы (ВЧ) (см. рис. 1.3а) и образовании большого количества электронно-дырочных пар (см. рис. 1.3б).

Под воздействием положительного заряда на «тело» транзистора снижается пороговое напряжение. Эффект снижения порогового напряжения N -канального транзистора наблюдается и при воздействии ионизирующего излучения (ИИ) на КНИ СБИС. Это вызывает снижение напряжения пробоя стока-истока (V_{ds})

N -канального транзистора, что, в свою очередь, приводит к росту статического тока потребления КНИ СБИС и к функциональным сбоям.

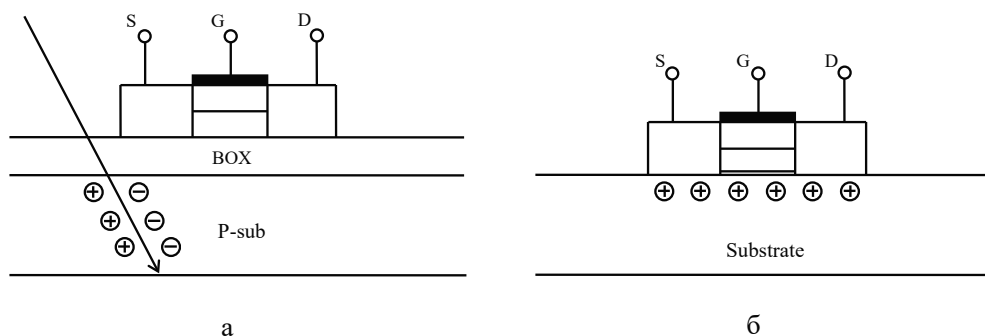


Рис. 1.3. Образование электронно-дырочных КНИ СБИС (а) и положительного заряда в подложке КНС СБИС (б) под воздействием ВЧ

Для снижения этого эффекта, в частности, устанавливается фиксированный потенциал подложки (*substrate*) КНИ СБИС, что не делается в случае создания технологии высокоскоростных схем. Улучшение стойкости к ИИ достигается за счет минимизации накопления заряда в захороненном окисле под действием отрицательного потенциала подложки.

1.3.6. Метод селективного отключения синхросигнала

Метод селективного отключения синхросигнала (*clock gating*) заключается в том, чтобы отключать синхросигнал на более ранней ступени дерева синхросигналов для блоков, которые в данный момент не выполняют полезную функцию. При этом перестают переключаться и рассеивать энергию все узлы отключаемого блока, включая узлы, формирующие сами синхросигналы, расположенные по дереву ниже вентиля отключения. Это дает существенную экономию мощности, поскольку в синхронных схемах дерево синхросигналов работает с наибольшей частотой и имеет большую суммарную емкость. Для микропроцессоров дерево синхросигналов может потреблять свыше трети общей мощности микросхемы [20]. В настоящий момент метод отключения синхросигналов широко используется при проектировании микропроцессоров и достаточно подробно описан в литературе [21]. Имеются средства САПР, позволяющие автоматически встраивать схемы *clock gating* в исходную схему, однако, как правило, это не позволяет добиться максимальной эффективности, то есть не все схемы, которые могли бы отключаться в данный момент времени, отключаются в синтезированной схеме. Для максимального использования данного метода разработчики кода *RTL* (исходный код для синтеза микросхемы) вставляют конструкции, позволяющие при синтезе однозначно вставлять схемы, отключающие синхросигналы.

Вариант перехода от традиционной синхронной схемы к схеме с использованием метода отключения синхросигнала приведен на рис. 1.4 и 1.5.

На схеме, показанной на рис. 1.4, одно и то же значение входной информации перезаписывается в регистр каждый такт, когда управляющий сигнал неактивен ($EN = 0$). Заменяв мультиплексор схемой, позволяющей отключать поступа-

ющий на регистр синхросигнал при $EN = 0$, можно значительно минимизировать энергопотребление данной схемы.

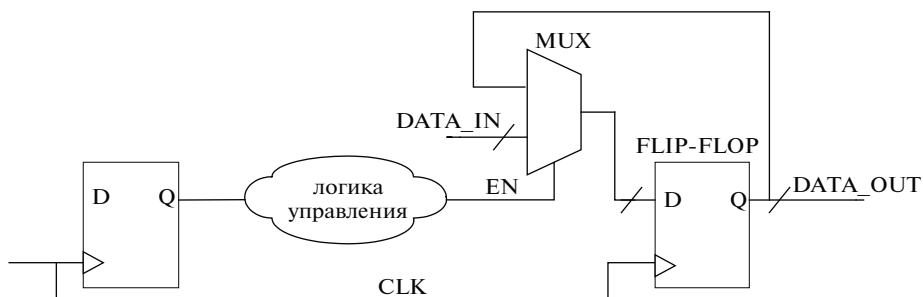


Рис. 1.4. Традиционный способ построения синхронной схемы

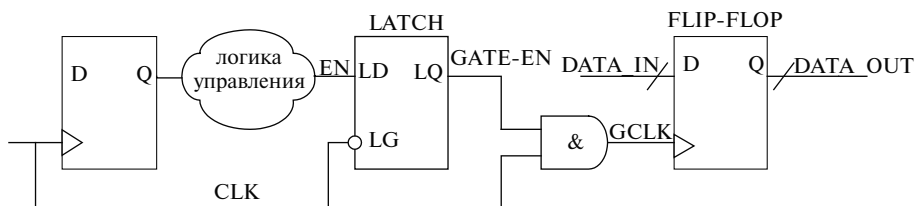


Рис. 1.5. Построение синхронной схемы с отключаемым синхросигналом

На схеме, показанной на рис. 1.5, сигнал EN управляет подачей синхросигнала на элемент типа «регистр». В каждом такте, когда уровень сигнала EN низкий, синхросигнал $GCLK$ не подается на регистр. Элемент «регистр – защелка» предотвращает возможные выбросы (*glitch*) по сигналу управления $GATE-EN$ и обеспечивает стабильный *gate*-сигнал ($GCLK$) во время положительного импульса CLK .

Временная диаграмма, характеризующая работу данной схемы, приведена на рис. 1.6.

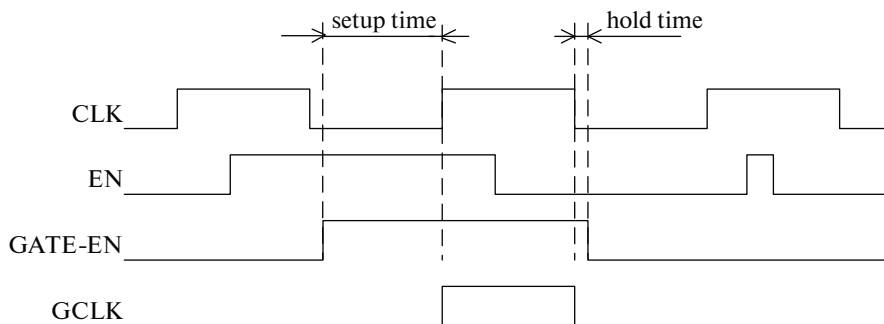


Рис. 1.6. Временная диаграмма работы схемы с отключаемым синхросигналом

Следует отметить, что элементы, используемые для построения дерева синхросигналов, обладают рядом характеристик, которые отсутствуют у остальных

элементов библиотеки. Например, в то время как обычные элементы имеют разное время переключения из состояния логической единицы в ноль, для элементов, формирующих дерево синхросигналов это, наоборот, недопустимо. Поэтому необходимо обеспечить равенство времени задержек для переднего и заднего фронтов во всем диапазоне изменений напряжения питания, температуры и параметров технологического процесса. Особое внимание требуется уделять введению дополнительных буферных элементов в цепь управляющего сигнала (*GATE_EN*), поступающего на объединяющий вентиль, для обеспечения требуемой скважности синхросигнала. Разработчик должен гарантировать, что включение дополнительной управляющей логики в дерево синхросигналов не внесет искажений в длительность фронтов синхросигналов и в их временные характеристики [22].

Таким образом, несмотря на простоту принципа использования механизма отключения синхросигнала, реализация этого метода на практике сопряжена с рядом сложностей. Сама схема управления синхросигналом занимает площадь кристалла, в некоторых случаях значительную, и увеличивает энергопотребление. Это означает, что в каждом конкретном случае следует решать вопрос о поиске наиболее оптимального варианта реализации данного механизма.

1.3.7. Использование средств САПР

В связи с тем что требование снижения энергопотребления современных микропроцессорных систем стало жизненно необходимым, в последние годы появилось достаточное количество САПР, позволяющих автоматизировать процесс построения схем со сниженным энергопотреблением. Наиболее распространенными являются продукты *Power Compiler* (в составе *Design Compiler*) компании *Synopsys* и *RTL Compiler* компании *Cadence*. Оба продукта в качестве исходных данных требуют наличия библиотеки элементов, содержащей данные о мощности потребления каждого элемента в зависимости от емкости нагрузки, и набора временных ограничений, налагаемых на проектируемую схему. Для анализа энергопотребления требуется также информация о входных воздействиях.

Для снижения энергопотребления используются следующие методы [23]:

- анализ схемы на предмет выявления мест, позволяющих применить метод отключения неиспользуемых блоков. На основе анализа автоматически модифицируется схема посредством включения элементов отключения синхросигнала, если это не нарушает временных ограничений;
- оптимизация размещения логических элементов на кристалле в целях уменьшения длин и, соответственно, емкости межсоединений;
- оптимизация схемы на уровне библиотечных ячеек, а именно уменьшение нагрузочной способности элементов;
- перестановка логически эквивалентных, но электрически разных входов библиотечных ячеек;
- ограничение длительности фронтов сигналов посредством добавления буферных элементов;
- замена нескольких библиотечных ячеек одной, более сложной логически эквивалентной ячейкой.

Все вышеперечисленные действия могут быть реализованы, только если соблюдаются заданные временные ограничения.

Использование средств САПР для снижения энергопотребления может быть рекомендовано к включению в маршрут проектирования цифровых схем. Основным ограничением может быть высокая стоимость САПР, а также отсутствие соответствующих библиотек стандартных ячеек, удовлетворяющих требованиям САПР.

Однако возможности САПР не могут заменить возможности человека в тех случаях, когда для снижения показателя энергопотребление/производительность требуется изменение алгоритма функционирования микросхемы.

1.3.8. Принцип понижения энергопотребления адиабатических логических элементов

Переключение статического КМОП цифрового элемента из одного логического состояния в другое сопровождается рассеиванием некоторого количества энергии, которая складывается из энергии, потраченной на протекание сквозных токов через транзисторы при переключении, и энергии, необходимой для изменения потенциала выходного узла схемы. Величина первой составляющей рассеиваемой энергии зависит от времени нарастания или спада входного сигнала, напряжения питания, порогового напряжения КМОП-транзисторов. Величина второй составляющей рассеиваемой энергии рассчитывается по формуле

$$E_{diss} = I^2 RT = \left(\frac{CV_{dd}}{T} \right)^2 RT = \left(\frac{RC}{T} \right) CV_{dd}^2,$$

где I — ток, которым осуществляется перезаряд емкости выходного узла элемента, R — сопротивление элемента, через который проходит ток перезаряда выходной емкости, C — величина эквивалентной емкости выходного узла, V_{dd} — напряжение питания (величина разности уровней двух различных логических состояний), T — время, в течение которого происходит перезаряд выходной емкости (изменение выходного логического состояния).

Эта формула справедлива только для случаев $T \gg RC$.

Задача по снижению динамического энергопотребления с точки зрения принципа адиабатического вычисления заключается в поиске схемотехнических решений и организации функционирования элемента, при которых по отношению к режимам работы транзисторов в элементе должны выполняться два правила:

- транзистор не может переходить из открытого состояния в закрытое, если через него протекает ток;
- транзистор не может переходить из закрытого состояния в открытое, если между стоком и истоком есть разность потенциалов;

Формулировка этих двух условий накладывает ограничения на режим работы транзисторов, при которых ток, проходящий через канал транзистора, вызывает падение напряжения между стоком и истоком, приводя к рассеиванию части энергии носителей заряда в тепло. В обыкновенных КМОП-элементах применяются шины питания с фиксированным потенциалом и при изменении логического состояния элемента на проводящих транзисторах, формирующих выходной уровень, неизбежно падает часть напряжения и рассеивается энергия. Характерной

особенностью адиабатического вычисления является использование на общих шинах элементов импульсов напряжения с линейной зависимостью нарастания и спада от времени. Переключение элемента из одного логического состояния в другое осуществляется в период изменения напряжения на общей шине за счет коммутации выходного узла с общей шиной. Ток перезаряда в период переключения, протекающий через транзисторы, задается разностью потенциалов между общей шиной и выходным узлом. Если время нарастания T линейно изменяющегося напряжения на общей шине значительно превосходит постоянную времени $\tau \approx RC$ данного элемента, то разность потенциалов между выходным узлом и общей шиной в период переключения будет незначительной.

Транзисторы (или диоды), коммутирующие общую шину с выходным узлом в период изменения напряжения на общей шине, пропускают через себя ток, который перезаряжает эквивалентную емкость выходного узла. Элемент переключается в другое логическое состояние, если ток стока проводящих транзисторов I переносит заряд Q , необходимый для изменения выходного напряжения, на величину напряжения питания. Связь тока I и времени изменения напряжения T на общей шине описывается выражением

$$I = Q/T \sim Qf,$$

где f – частота изменения напряжения на общей шине.

Таким образом, ток стока транзисторов, формирующих выходной сигнал, пропорционален частоте импульсов напряжения на общей шине. При повышении быстродействия схемы за счет роста частоты смены импульсов уменьшается время перезаряда выходного узла. Сокращение времени перезаряда вызывает рост тока стока транзисторов, которые осуществляют перезаряд. Повышение тока стока способствует падению напряжения между стоком и истоком и, следовательно, повышению рассеиваемой энергии.

1.3.9. Использование обратимой логики

Для формирования логического состояния требуется потратить некоторое количество энергии. Установлено, что минимальное количество энергии, необходимое для формирования логического уровня, имеет порядок энергии теплового шума [24]:

$$E = K \times T \times \ln 2,$$

K – константа Больцмана, T – абсолютная температура вычислительного устройства.

Именно это минимальное количество энергии рассеивается при потере одного логического состояния. Преодоление этого фундаментального предела осуществимо, если при совершении любой логической операции не происходит потери логических значений входных сигналов и возможно их восстановление на основе выходных. То есть для снижения рассеивания энергии при вычислениях необходимо избегать потери логических состояний сигналов.

Обратимая архитектура включает в себя совокупность методов по снижению потери информации, которые могут применяться на разных уровнях организации вычислений. Например, в программировании некоторые операции являются логически обратимыми ($R1 = R1 + R2$), а некоторые являются логически необрати-

мыми ($R1 = 0$). Логически необратимая операция приводит к потере бита информации в регистре и, соответственно, к дополнительному рассеиванию энергии. Эта проблема решается оптимизацией алгоритма работы компилятора, который по возможности должен избегать в машинном коде таких операций.

Необратимым является процесс рассеивания энергии при потере битов информации в ходе совершения логических операций или записи в элементы памяти. Можно преодолеть этот предел, если в вычислительном устройстве использовать обратимые логические элементы (т.е. логические элементы, в которых происходящие процессы должны быть обратимыми). Эти элементы могут реализовать обратимую булеву функцию и сохранять полное количество битов информации, поэтому энтропия при работе таких элементов остается неизменной. При реализации прямой булевой функции на выходе обратимого элемента кроме информационных битов формируются так называемые мусорные (*garbage*) биты, с помощью которых можно реализовать обратную булеву функцию и восстановить значения входных сигналов.

Примеры обратимых логических элементов. Особенностью всех обратимых логических элементов является возможность однозначно определить состояние входных сигналов по состоянию выходных. Простейшим примером обратимого логического элемента является вентиль НЕ. По состоянию выхода этого элемента можно однозначно определить состояние входа. Вентиль НЕ реализует обратимую логическую функцию, и информация при его работе не теряется.

Другой пример обратимого логического элемента – это УПРАВЛЯЕМОЕ НЕ (см. рис. 1.7).

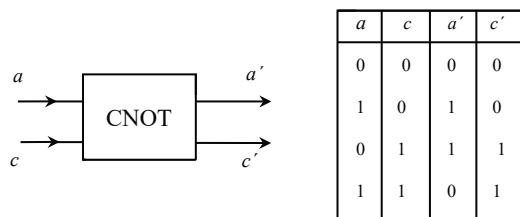


Рис. 1.7. Схематическое изображение вентиля, УПРАВЛЯЕМОЕ НЕ и его таблица истинности

Состояние выхода C' всегда совпадает с входом C , а состояние выхода A' инвертируется при $C = 1$. Как видно из таблицы истинности, этот элемент является обратимым, поскольку каждому выходному состоянию соответствует определенное состояние входов. Логическая функция этого элемента может быть записана следующим образом:

$$a' = a \times \bar{c} + \bar{a} \times c.$$

Стоит отметить, что под свойством консервативности элемента понимается сохранение количества информации при ее прохождении через обратимый логический элемент или цепочки из таких вентилях. Под свойством обратимости элемента также понимается возможность замены его входов на выходы вместе с заменой его логической функции на обратную. При этом количество входов у обратимого элемента должно совпадать с количеством выходов.

Для построения вычислительной системы необходимо, чтобы элемент мог выполнять любую логическую функцию. При этом логический элемент должен реализовывать по крайней мере функции И и ИЛИ, или И и НЕ, или НЕ и ИЛИ и допускать каскадирование. Из одних только элементов УПРАВЛЯЕМОЕ НЕ и НЕ нельзя создать вычислительно-универсальный вентиль, на основе которого можно построить любую булеву функцию. Вычислительно-универсальным обратимым элементом является вентиль Фредкина [25], который можно понимать как вентиль УПРАВЛЯЕМЫЙ ОБМЕН, реализующий следующую функцию:

$$c' = c; y_1 = c \times x_1 + \bar{c} \times x_2; y_2 = \bar{c} \times x_1 + c \times x_2,$$

(см. рис. 1.8).

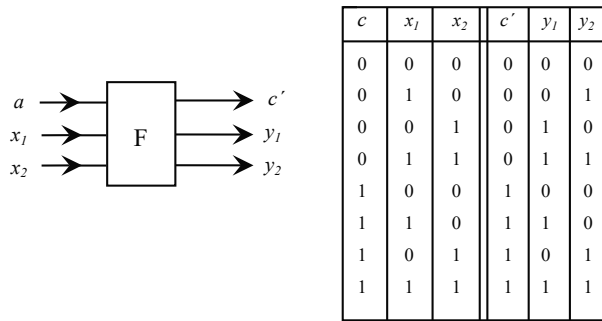


Рис. 1.8. Схематическое изображение вентиля Фредкина и его таблица истинности

На рис. 1.9 показаны различные варианты включения вентиля Фредкина, показывающие его универсальность.

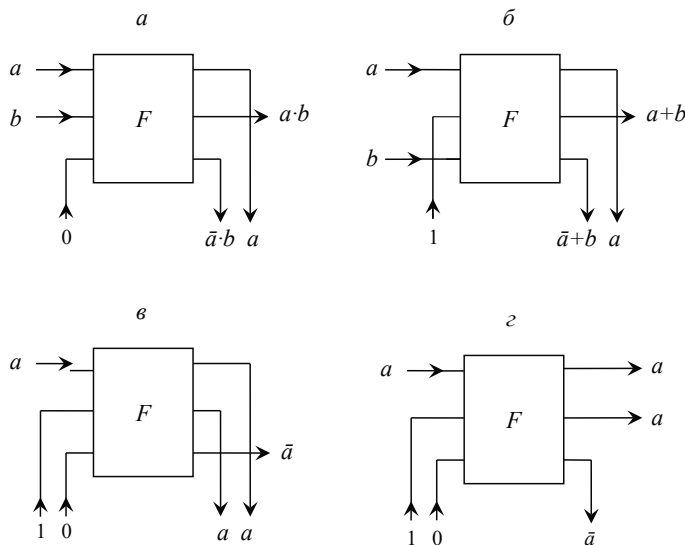


Рис. 1.9. Различные конфигурации обратимых логических вентилях на основе вентиля Фредкина: И (а), ИЛИ (б), НЕ (в) и ВЕТВЛЕНИЕ (г)

С точки зрения оптической реализации можно выделить два логических обратимых и вычислительно-универсальных логических элемента: переключатель Прайса [26] и вентиль ВЗАИМОДЕЙСТВИЕ (*Interaction Gate*) [27]. Переключатель Прайса (см. рис. 1.10) – это управляемый переключатель, представляющий логическую функцию

$$c' = c; y_1 = c \times x; y_2 = \bar{c} \times x.$$

Вход C играет роль адресного.

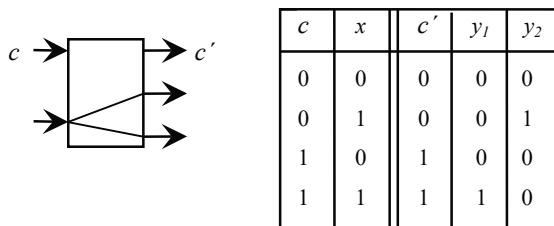


Рис. 1.10. Схематическое изображение и таблица истинности переключателя Прайса

Необходимо обратить внимание, что конфигурация включения вентиль Фредкина, показанная на рис. 1.8а, реализует тот же самый переключатель, если переменную a понимать как адресный сигнал, а переменную b как переключаемый сигнал. Это дополнительный пример универсальности вентиль Фредкина. Вентиль ВЗАИМОДЕЙСТВИЕ (см. рис. 1.11) представляет собой четырехкомпонентную булеву функцию двух аргументов:

$$y_1 = x_1 \times x_2; y_2 = \bar{x}_1 \times x_2; y_3 = x_1 \times \bar{x}_2; y_4 = \bar{x}_1 \times \bar{x}_2;$$

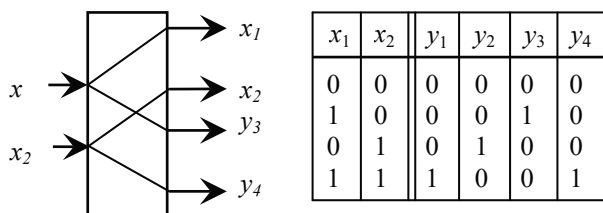


Рис. 1.11. Схематическое изображение и таблица истинности вентиль ВЗАИМОДЕЙСТВИЕ

Для вентиль ВЗАИМОДЕЙСТВИЕ была предложена механическая реализация – модель бильярдных шаров. Суть данной модели можно понять из рис. 1.12. Консервативность этой логики обусловлена обратимостью механики Ньютона, описывающей движение и столкновение абсолютно упругих шаров. Бильярдная реализация переключателя Прайса и вентиль ВЗАИМОДЕЙСТВИЕ показана на рис. 1.13.

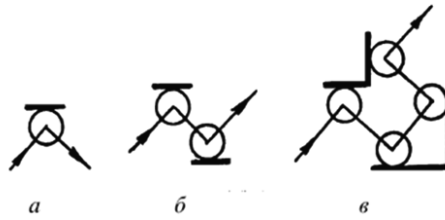


Рис. 1.12. Бильярдная логика: отклонение пути шара (а), смещение пути шара (б) и задержка шара (в)

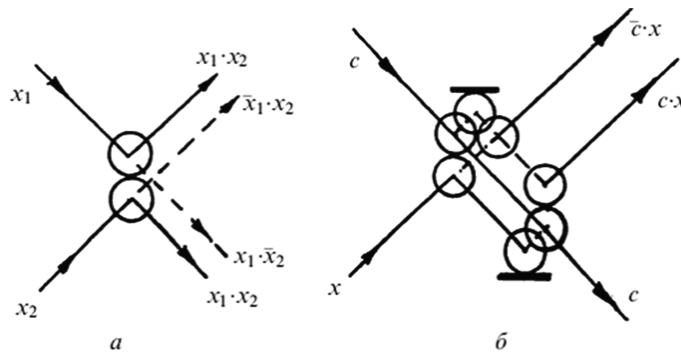


Рис. 1.13. Бильярдная реализация вентилей ВЗАИМОДЕЙСТВИЕ (а) и переключателя Прайса (б)

Предположение, что шары взаимодействуют (сталкиваются) упруго друг с другом и с отражающей стенкой и все столкновения являются центральными, является жестким ограничением, несущественным для теоретического анализа, но выступающим веским препятствием при реальном воплощении.

В зависимости от технологического базиса возможен обоснованный выбор той или иной схемы обратимого вентилей. Перечень схем обратимых элементов не ограничивается элементами, представленными выше. Были предложены и другие схемы обратимых элементов: вентиль Тоффоли [28], вентиль Кернтгофа [29], вентиль Переса [30], вентиль Магроласа [31] и несколько вентилей Де Воса [32].

Схемы цифровых устройств на обратимых элементах

Классические методы синтеза логики не могут быть напрямую применены в схемах, построенных на обратимых логических элементах. В статье [33] говорится, что схемы на обратимых логических элементах проектируются для каждого конкретного устройства с заданной булевой функцией. При этом не существует единых правил, с помощью которых можно было бы спроектировать устройство с произвольной булевой функцией.

Триггеры. В работе [34] была предложена методика построения основных последовательных элементов на обратимых вентилей Фредкина. Согласно описанию триггеры, спроектированные по предложенной методике, имеют лучшие показатели, чем триггеры, предложенные в работе [35]. Триггеры сравнивались по ко-

личеству «мусорных» выходов и по количеству используемых вентиляей. При этом были приведены схемы триггеров без привязки к какой-либо элементной базе.

Сумматоры. В статье [36] предложена схема сумматора на обратимых вентилях Фредкина. В этой статье рассмотрены различные реализации схем сумматоров и обоснована схема на рис. 1.14а и 1.15б по критерию минимального пути прохождения сигналов.

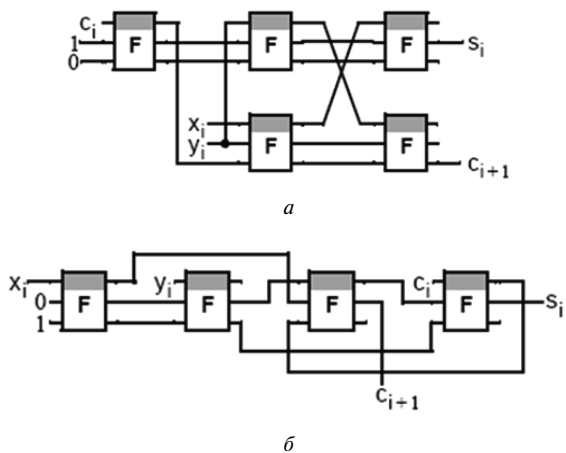


Рис. 1.14. Возможные варианты реализации полного сумматора на обратимых вентилях Фредкина без возможности каскадирования

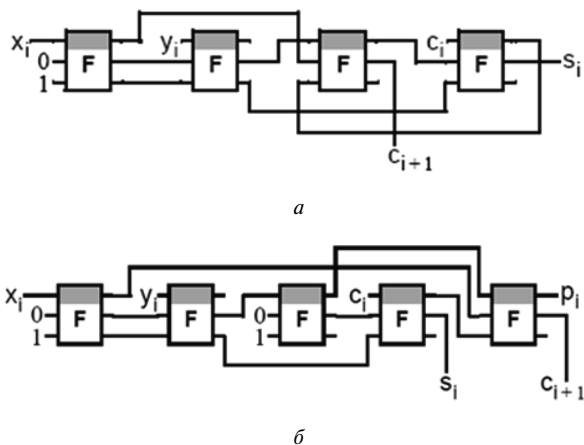


Рис. 1.15. Возможные варианты реализации полного сумматора на обратимых вентилях Фредкина с возможностью каскадирования

Матрицы программируемых логических вентиляей на обратимых элементах

В статье [37] описывается программируемое логическое устройство, построенное на обратимых элементах Фредкина и обратимых элементах ВЗАИМОДЕЙСТВИЕ (вентиль Фейнмана), – *Reversible Programmable Logic Array (RPLA)* (см. рис. 1.16).