Лекция

- **Система памяти МПС. Состав и основные характеристики системы памяти МПС**
- **Организация и краткие характеристики БИС ПЗУ, ППЗУ**
- **Статическая и динамическая оперативная память. Структура элементарной ячейки, характеристики**
- **Стековая и буферная память. Кэш-память**
- **Flash-память. Надежность ЗУ**

Классификация ЗУ

1. **По использованию**:
   - внешние;
   - внутренние (оперативные).
2. **По назначению**:
   - сверхоперативные;
   - оперативные;
   - постоянные;
   - буферные;
   - внешние.
3. **По физическим принципам действия:**
   - магнитные;
   - полупроводниковые;
   - оптические.

Классификация ЗУ

4. **По характеру обращения**:
- адресные;
- ассоциативные.
5. **По способу доступа к заданной ячейки (для адресных ЗУ):**
- с последовательным доступом;
- с произвольным доступом.

Основные характеристики полупроводниковой памяти

1. **Емкость памяти** определяется числом бит хранимой информации. Емкость кристалла обычно выражается также в битах и составляет 1024 бита, 4К бит, 16 К бит, 64К бит и т.п. Важной характеристикой кристалла является **информационная организация кристалла памяти M x N,** где **M - число слов, N - разрядность слова**.
2. **Временные характеристики памяти**.

Время доступа - временной интервал, определяемый от момента, когда центральный процессор выставил на шину адреса адрес требуемой ячейки памяти и послал по шине управления приказ на чтение или запись данных, до момента осуществления связи адресуемой ячейки с шиной данных.

Время восстановления - это время, необходимое для проведения памяти в исходное состояние после того, как ЦП снял с ША адрес, ШУ - сигнал “чтение” или “запись” и с ШД - данные.

Быстродействие и производительность памяти

Быстродействие памяти определяется временем выполнения операций записи и считывания данных. Основными параметрами любых элементов памяти является минимальное время доступа и длительность цикла обращения. *Время доступа* (access time) определяется как задержка появления действительных данных на выходе памяти относительно начала цикла чтения, *длительность цикла* – как минимальный период следующих друг за другом обращений к памяти, причем циклы чтения и записи могут требовать различных затрат времени. В цикл обращения помимо активной фазы самого доступа входит и фаза восстановления (возврата памяти к исходному состоянию), которая соизмерима по времени с активной фазой. Временные характеристики самих запоминающих элементов определяются их принципом действия и технологией изготовления.

Производительность памяти можно характеризовать как скорость потока записываемых или считываемых данных и измерять в мегабайтах в секунду. Производительность подсистемы памяти наравне с производительностью процессора существенным образом определяет производительность компьютера. Выполняя определенный фрагмент программы, процессору придется, во-первых, загрузить из памяти соответствующий программный код, а во-вторых, произвести требуемые обмены данными, и чем меньше времени потребуется подсистеме памяти на обслуживание этих операций, тем лучше.

Производительность памяти, как основной, так и кэша второго уровня, обычно характеризуют *длительностью пакетных циклов чтения* (memory burst read cycle). Пакетный режим обращения является основным для процессоров, использующих кэш (класса 486 и выше); циклы чтения выполняются гораздо чаще, чем циклы записи (хотя бы потому, что процессору приходится все время считывать инструкции из памяти). Эта длительность выражается в числе тактов системной шины, требуемых для передачи очередной порции данных в пакете. Обозначение вида *5–3–3–3* для диаграммы пакетного цикла чтения соответствует пяти тактам на считывание первого элемента в цикле и трем тактам на считывание каждого из трех последующих элементов. Первое число характеризует латентность (latency) памяти — время ожидания данных, последующие — скорость передачи. При этом, конечно же, оговаривается и частота системной шины.

Режим пакетирования

Средство блочной передачи: 64 бита за 1 раз.

Такт	Конвейеризация (80286)	Пакет (80486)
1	Адрес 1 слова	Адрес 1 слова
2	Передача 1 слова	----------
3	Адрес 2 слова	Передача 1 слова
4	Передача 2 слова	Передача 2 слова
5	Адрес 3 слова	Передача 3 слова
6	Передача 3 слова	Передача 4 слова
7	Адрес 4 слова	----------
8	Передача 4 слова	----------

Производительность подсистемы памяти зависит от *типа* и *быстродействия* применяемых запоминающих элементов, *разрядности* шины памяти и некоторых «хитростей» архитектуры. Современные типы памяти обеспечивают высокую скорость передачи внутри пакета, используя двойную и даже четырехкратную синхронизацию. При этом параметром шины, по которой передаются данные, может быть как *частота тактового сигнала*, так и *частота передачи данных*. Последняя может в 2 (DDR SDRAM) или в 4 (DDR2 SDRAM, шина Pentium 4) раза превышать тактовую частоту. Задержка получения данных чтения процессорным ядром в современных компьютерах может составлять от 45 до нескольких сотен наносекунд в зависимости от способа подключения памяти.

Разрядность шины памяти – это количество байтов (или битов), с которыми операция чтения или записи может быть выполнена одновременно. Разрядность основной памяти обычно согласуется с разрядностью внешней шины процессора (1 байт – для 8088; 2 байта – для 8086, 80286, 386SX; 4 байта – для 386DX, 486; 8 байт – для Pentium и выше). Вполне очевидно, что при одинаковом быстродействии микросхем или модулей памяти производительность блока с большей разрядностью будет выше, чем у малоразрядного. Именно с целью повышения производительности у 32-битных (по внутренним регистрам) процессоров класса Pentium и выше внешняя шина, связывающая процессор с памятью, имеет разрядность 64 бита. У современных процессоров пропускная способность системной шины превышает пропускную способность шины памяти. Это подталкивает к использованию *двухканальной памяти* – удвоению разрядности шины памяти относительно разрядности системной шины процессора.

Банком памяти называют комплект микросхем или модулей (а также их посадочных мест – «кроваток» для микросхем, слотов для SIMM или DIMM), обеспечивающий требуемую для данной системы разрядность хранимых данных. Работоспособным может быть только полностью заполненный банк. Внутри одного банка практически всегда должны применяться одинаковые (по типу и объему) элементы памяти.

В современных компьютерах на процессорах 6–8-го поколений банком является один модуль DIMM или RIMM (подобный модуль может содержать и несколько банков,

Если устанавливаемый объем памяти набирается несколькими банками, появляется резерв повышения производительности за счет *чередования банков* (bank interleaving). Идея чередования заключается в том, что смежные блоки данных (разрядность такого блока данных соответствует разрядности банка) располагаются поочередно в разных банках. Тогда при весьма вероятном последовательном обращении к данным банки будут работать поочередно, причем активная фаза обращения к одному банку может выполняться во время фазы восстановления другого банка, то есть применительно к обоим банкам не будет простоя во время фазы восстановления. Частота передачи данных в системе с чередованием двух банков может быть удвоенной по отношению к максимальной частоте работы отдельного банка. Для реализации механизма чередования чипсет должен обеспечивать возможность перекоммутации адресных линий памяти в зависимости от установленного количества банков и иметь для них (банков) раздельные линии управляющих сигналов. Чем больше банков участвуют в чередовании, тем выше (теоретически) предельная производительность. Чаще всего в чередовании участвуют два банка (two way interleaving), но их может быть и больше. Из разбиения на мелкие банки можно извлечь и другую выгоду. Поскольку современные процессоры способны параллельно выставлять несколько запросов на транзакции с памятью, обусловленные необходимым временем доступа скрытые фазы обработки запросов, относящихся к разным банкам, могут выполняться одновременно.

Основные характеристики полупроводниковой памяти

1. **Удельная стоимость** запоминающего устройства определяется отношением его стоимости к информационной емкости, т.е. **определяется стоимостью бита хранимой информации.**
2. **Потребляемая энергия** (или рассеиваемая мощность) приводится **для двух режимов работы** кристалла: **режима пассивного хранения информации и активного режима**, когда операции записи и считывания выполняются с номинальным быстродействием. Кристаллы динамической МОП - памяти в резервном режиме потребляют примерно в десять раз меньше энергии, чем в активном режиме. Наибольшее потребление энергии, не зависящее от режима работы, характерно для кристаллов биполярной памяти.

Основные характеристики полупроводниковой памяти

3. **Плотность упаковки определяется площадью запоминающего элемента** и зависит от числа транзисторов в схеме элемента и используемой технологии. Наибольшая плотность упаковки достигнута в кристаллах динамической МОП - памяти.
4. **Допустимая температура окружающей среды** обычно указывается **отдельно для активной работы, для пассивного хранения информации и для нерабочего состояния с отключенным питанием**. Указывается **тип корпуса**, если он стандартный, или чертеж корпуса с указанием всех размеров, маркировкой и нумерацией контактов, если корпус новый. Приводятся также **условия эксплуатации***:* рабочее положение, механические воздействия, допустимая влажность и другие.

Постоянная память

Постоянная память используется для энергонезависимого хранения системной информации – BIOS, таблиц знакогенераторов и т. п. Эта память при обычной работе компьютера только считывается, а запись в нее (часто называемая программированием) осуществляется специальными устройствами – программаторами. Отсюда и ее название – *ROM* (Read Only Memory – память только для чтения), или *ПЗУ* (постоянное запоминающее устройство). Требуемый объем памяти этого типа невелик: например, объем BIOS PC/XT составлял 8 Кбайт, в современных компьютерах типовое значение от 128 Кбайт до 2 Мбайт. Быстродействие постоянной памяти обычно ниже, чем оперативной, но этот недостаток может быть исправлен применением теневой памяти

В последние годы постоянную память вытесняют *флэш-память*, запись в которую возможна в самом компьютере в специальном режиме работы, и другие типы энергонезависимой памяти (EEPROM, FRAM).

Масочные постоянные запоминающие устройства (ПЗУ, или ROM) имеют самое высокое быстродействие (время доступа 30–70 нс). Эти микросхемы в PC широкого применения не получили ввиду сложности модификации содержимого (только путем изготовления новых микросхем), они иногда применялись в качестве знакогенераторов в некоторых моделях графических адаптеров CGA, MDA, HGC.

Однократно программируемые постоянные запоминающие устройства (ППЗУ, или PROM) имеют аналогичные параметры и благодаря возможности программирования изготовителем оборудования (а не микросхем) находят более широкое применение для хранения кодов BIOS и в графических адаптерах. Программирование этих микросхем осуществляется только с помощью специальных программаторов, в целевых устройствах они устанавливаются в «кроватки» или запаиваются. Как и масочные, эти микросхемы практически не чувствительны к электромагнитным полям (в том числе к рентгеновскому облучению), и несанкционированное изменение их содержимого в устройстве исключено (конечно, не считая отказа).

Репрограммируемые постоянные запоминающие устройства (РПЗУ, или EPROM) до недавних пор были самыми распространенными носителями BIOS как на системных платах, так и в адаптерах, а также использовались в качестве знакогенераторов. Наиболее популярные микросхемы имеют 8-битную организацию и обозначение вида 27xx-tt или 27Cxx-tt для микросхем CMOS. Здесь xx определяет емкость в килобитах: 2708 — 1K × 8 — родоначальник семейства,

2716/32/64/128/256/512 имеют емкость 2/4/8/16/32/64 Кбайт соответственно, 27010 и 27020 – 128 и 256 Кбайт. Время доступа tt лежит в диапазоне 50–250 нс. 16-битные микросхемы (например, 27001 или 27002 емкостью 64К или 128К 16-битных слов) в PC применяются редко.

Микросхемы EPROM тоже программируются на программаторах, но относительно простой интерфейс записи позволяет их программировать и в устройстве (но не в штатном его режиме работы, а при подключении внешнего программатора). *Стирание* микросхем осуществляется ультрафиолетовым облучением в течение нескольких минут. Специально для стирания микросхемы имеют стеклянные окошки. После программирования эти окошки заклеивают, предотвращая стирание под действием солнечного или люминесцентного облучения. Время стирания зависит от расстояния до источника облучения, его мощности и объема микросхемы (более емкие микросхемы стираются быстрее). Вместо штатных стирающих устройств можно пользоваться и обычной медицинской ультрафиолетовой лампой с расстояния порядка 10 см. Для микросхем 2764 ориентировочное время стирания составляет 5 минут. Стирание переводит все биты в единичное состояние. «Недотертые» микросхемы при программировании могут давать ошибки, передержка при стирании снижает количество возможных циклов перепрограммирования (в пределе – до нуля).

Отметим *основные свойства EPROM*:

- Стирание информации происходит сразу для всей микросхемы под воздействием облучения и занимает несколько минут. Стертые ячейки имеют единичные значения всех битов.
- Запись может производиться в любую часть микросхемы побайтно, в пределах байта можно маскировать запись отдельных битов, устанавливая им единичные значения данных.
- Защита от записи осуществляется подачей низкого (5 В) напряжения на вход V_{PP} в рабочем режиме (только чтение).
- Защита от стирания производится заклейкой окна.

Стирать микросхемы постоянной памяти можно электрическим способом. Однако этот процесс требует значительного расхода энергии, который выражается в необходимости приложения относительно высокого напряжения стирания (10–30 В) и длительности импульса стирания более десятка микросекунд. Интерфейс традиционных микросхем EEPROM имел временную диаграмму режима записи с большой длительностью импульса, что не позволяло непосредственно использовать сигнал записи системной шины. Кроме того, перед записью информации в ячейку обычно требовалось предварительное стирание, тоже довольно длительное. Микросхемы EEPROM относительно небольшого объема широко применяются в качестве энергонезависимой памяти конфигурирования различных адаптеров. Современные микросхемы EEPROM имеют довольно сложную внутреннюю структуру, в которую входит управляющий автомат. Это позволяет упростить внешний интерфейс, делая возможным непосредственное подключение к микропроцессорной шине или иному интерфейсу, и скрыть специфические (и не нужные пользователю) вспомогательные операции типа стирания и верификации. Микросхемы EEPROM позволяют считывать и перезаписывать (стирать) любую ячейку памяти, но перезапись требует довольно много времени.

Оперативные запоминающие устройства (ОЗУ)

Полупроводниковые ЗУ подразделяются на

- **ЗУ с произвольной выборкой** (ЗУПВ):

• оперативные статические запоминающие устройства (СОЗУ),

• динамические оперативные запоминающие устройства (ДОЗУ);

- **ЗУ с последовательным доступом**:

• регистры сдвига,

• приборы с зарядовой связью (ПЗС).

Электронная память применяется практически во всех подсистемах PC, выступая в качестве оперативной памяти, кэш-памяти, постоянной памяти, полупостоянной памяти, буферной памяти, внешней памяти.

Основная, или *оперативная*, *память* (main memory) компьютера используется для оперативного обмена информацией (командами и данными) между процессором, внешней памятью (например, дисковой) и периферийными подсистемами (графика, ввод-вывод, коммуникации и т. п.). Ее другое название – *ОЗУ* (оперативное запоминающее устройство) – примерно соответствует английскому термину *RAM* (Random Access Memory – память с произвольным доступом). Произвольность доступа подразумевает возможность операций записи в любую ячейку ОЗУ или чтения любой ячейки ОЗУ в произвольном порядке.

Требования, предъявляемые к основной памяти:

- большой (для электронной памяти) объем, исчисляемый уже десятками – сотнями мегабайт и даже гигабайтами;
- быстродействие и производительность, позволяющие реализовать вычислительную мощность современных процессоров;
- высокая надежность хранения данных – ошибка даже в одном бите, в принципе, может привести к ошибкам вычислений, к искажению и потере данных, причем иногда и на внешних носителях.

Динамические ОЗУ

Оперативная память персональных компьютеров строится на базе относительно недорогой динамической памяти - **DRAM** (Dynamic Random Access Memory).

За это время сменилось множество поколений интерфейсной логики, соединяющей ядро памяти с "внешним миром". Эволюция носила ярко выраженный преемственный характер - каждое новое поколение памяти так или иначе наследовало архитектуру предыдущего, включая, в том числе, и свойственные ему ограничения. Ядро же памяти (за исключением совершенствования проектных норм таких, например, как степень интеграции) и вовсе не претерпевало никаких принципиальных изменений!

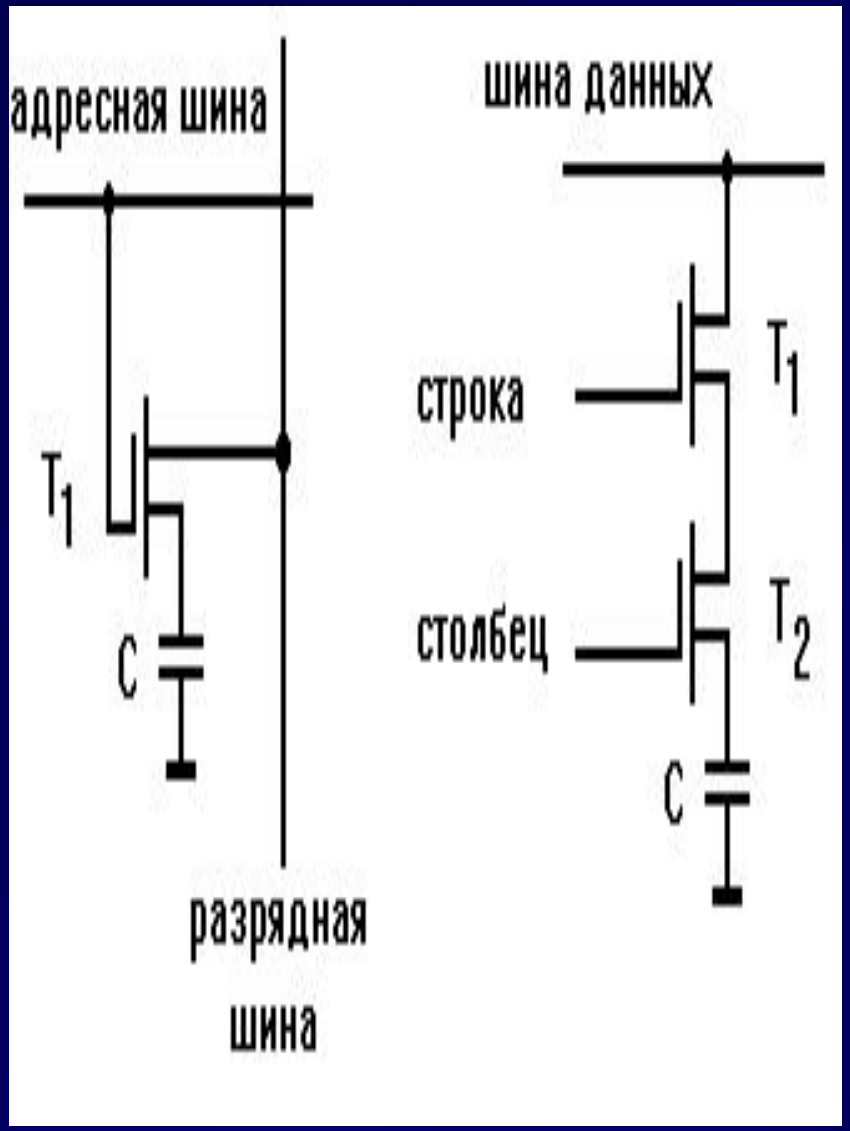
Динамические ОЗУ

Ядро микросхемы динамической памяти состоит из множества ячеек, каждая из которых **хранит всего один бит информации**. На физическом уровне ячейки объединяются в прямоугольную матрицу, горизонтальные линейки которой называются строками (ROW), а вертикальные - столбцами (Column) или страницами (Page).

В динамических ЗУ необходима постоянная регенерация информации, однако при этом для хранения одного бита в ДОЗУ нужны всего 1-2 транзистора и накопительный конденсатор.

Конденсатору отводится роль непосредственного хранителя информации. Правда, хранит он очень немного - всего один бит. Отсутствие заряда на обкладках соответствует логическому нулю, а его наличие - логической единице.

Динамические ОЗУ



Транзистор играет роль "ключа", удерживающего конденсатор от разряда. В спокойном состоянии транзистор закрыт, но, стоит подать на соответствующую строку матрицы электрический сигнал, как спустя мгновение-другое (конкретное время зависит от конструктивных особенностей и качества изготовления микросхемы) он откроется, соединяя обкладку конденсатора с соответствующим ей столбцом.

Чтение/запись отдельно взятой ячейки невозможна! Действительно, открытие одной строки приводит к открытию всех, подключенных к ней транзисторов, а, следовательно, - разряду закрепленных за этими транзисторами конденсаторов.

Чтение ячейки деструктивно по своей природе, поскольку чувствительный усилитель разряжает конденсатор в процессе считывания его заряда. "Благодаря" этому **динамическая память представляет собой память разового действия**. Разумеется, такое положение дел никого устроить не может, и потому во избежание потери информации считанную строку приходится тут же перезаписывать вновь. В зависимости от конструктивных особенностей эту миссию выполняет либо контроллер памяти, либо сама микросхема памяти. Практически все современные микросхемы принадлежат к последней категории.

Ввиду микроскопических размеров, а, следовательно, емкости конденсатора записанная на нем **информация хранится крайне недолго**, - буквально сотые, а-то и **тысячные доли секунды**. Причина тому - **саморазряд конденсатора.** Несмотря на использование высококачественных диэлектриков с огромным удельным сопротивлением, заряд стекает очень быстро, ведь количество электронов, накопленных конденсатором на обкладках, относительно невелико.

Для борьбы с "забывчивостью" памяти прибегают к ее **регенерации - периодическому считыванию ячеек с последующей перезаписью**.

Регенерация – периодическое восстановление исходного напряжения на конденсаторе.

В зависимости от конструктивных особенностей "регенератор" может находиться как в контроллере, так и в самой микросхеме памяти. Сегодня же регенератор чаще всего встраивается внутрь самой микросхемы, причем перед регенерацией содержимое обновляемой строки копируется в специальный буфер, что предотвращает блокировку доступа к информации.

Режимы регенерации

1. Асинхронный – МП и контроллер регенерации обращаются к памяти независимо друг от друга.

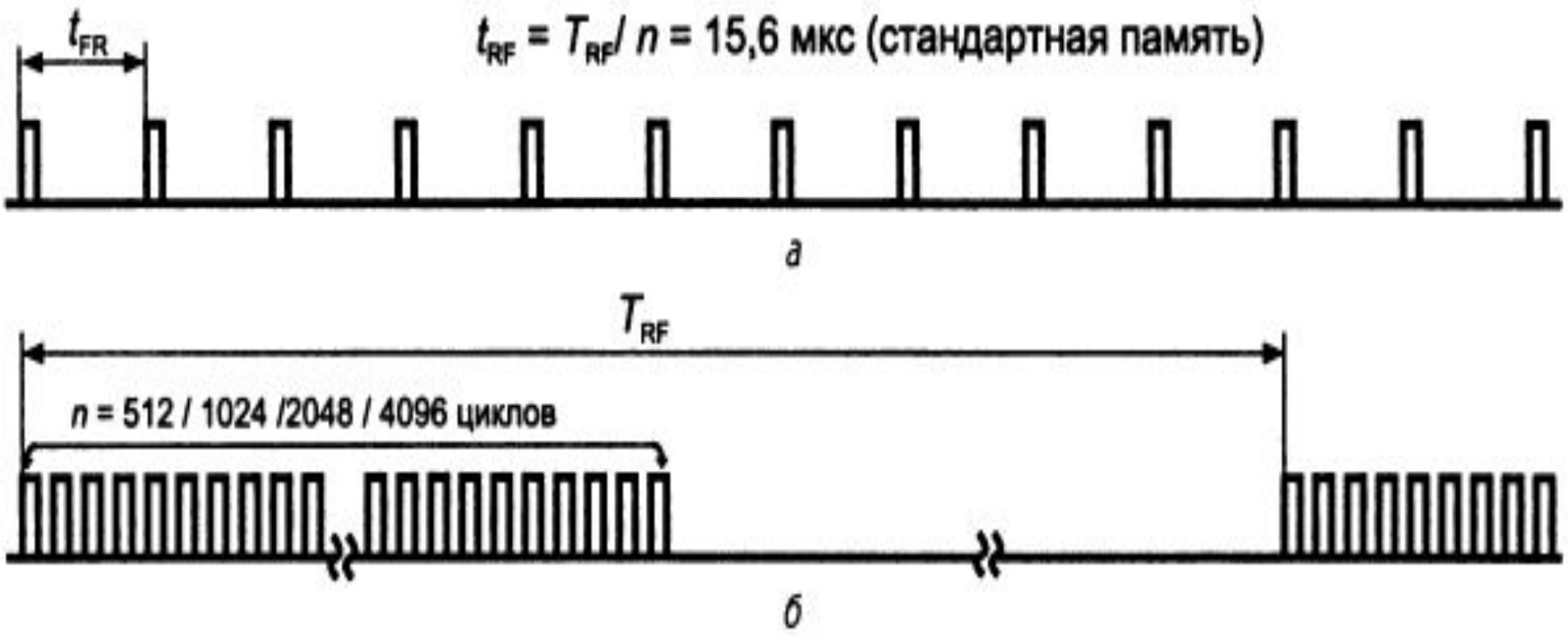
2. Синхронный – регенерации происходит в те циклы, когда МП не обращается к памяти. Нужна схема планировщика (арбитра).

   Арбитр – схема, обеспечивающая упорядочение запросов прерывания от ЦП для считывания или записи данных и запроса на регенерацию от схемы управления.

3. Полусинхронный – синхронно с ГТИ и асинхронно с МП.

Регенерация

Поскольку обращения (запись или чтение) к различным ячейкам памяти обычно происходят в случайном порядке, для поддержания сохранности данных применяется *регенерация памяти* (memory refresh) — регулярный циклический перебор ее ячеек (обращение к ним) с холостыми циклами. Регенерация в микросхеме происходит одновременно по всей строке матрицы при обращении к любой из ее ячеек. Максимальный период обращения к каждой строке T_{RF} (refresh time) для гарантированного сохранения информации у современной памяти лежит в пределах 8–64 мс. В зависимости от объема и организации матрицы для однократной регенерации всего объема требуется 512, 1024, 2048 или 4096 циклов обращений. При *распределенной регенерации* (distributed refresh) одиночные циклы регенерации выполняются равномерно с периодом t_{RF} который для стандартной памяти принимается равным 15,6 мкс. Период этих циклов называют *частотой регенерации* (refresh rate), хотя такое название больше соответствует обратной величине — частоте циклов $f = 1/t_{RF}$. Для памяти с расширенной регенерацией (extended refresh) допустим период циклов до 125 мкс. Возможен также и вариант *пакетной регенерации* (burst refresh), когда все циклы регенерации собираются в пакет во время которого обращение к памяти по чтению и записи блокируется. При количестве циклов 1024 эти пакеты будут периодически занимать шину памяти примерно на 130 мкс, что далеко не всегда допустимо. По этой причине, как правило, выполняется распределенная регенерация, хотя возможен и промежуточный вариант — пакетами по несколько (например, 4) циклов.



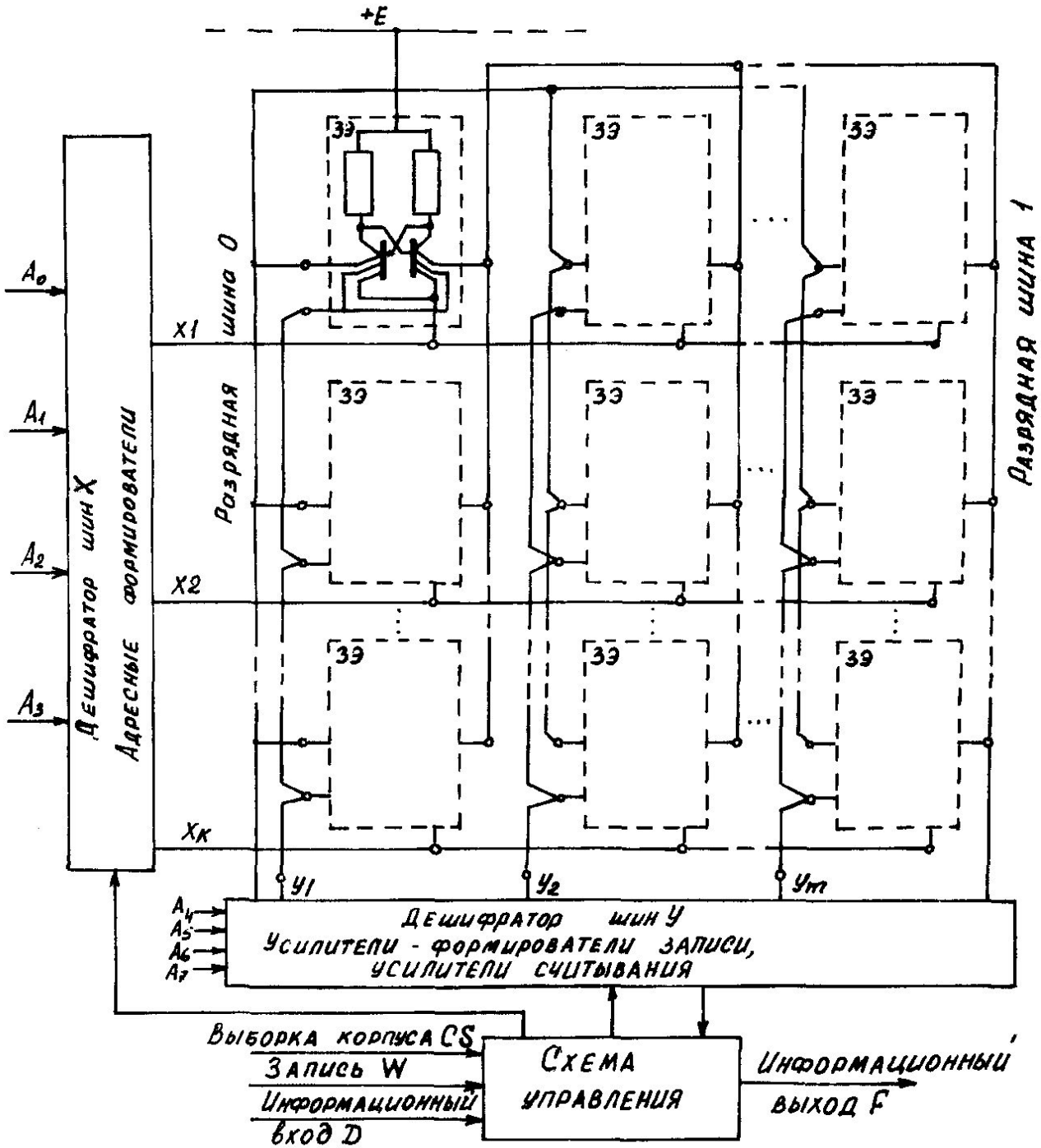
. Регенерация динамической памяти: *а* – распределенная, *б* – пакетная

Статические ОЗУ

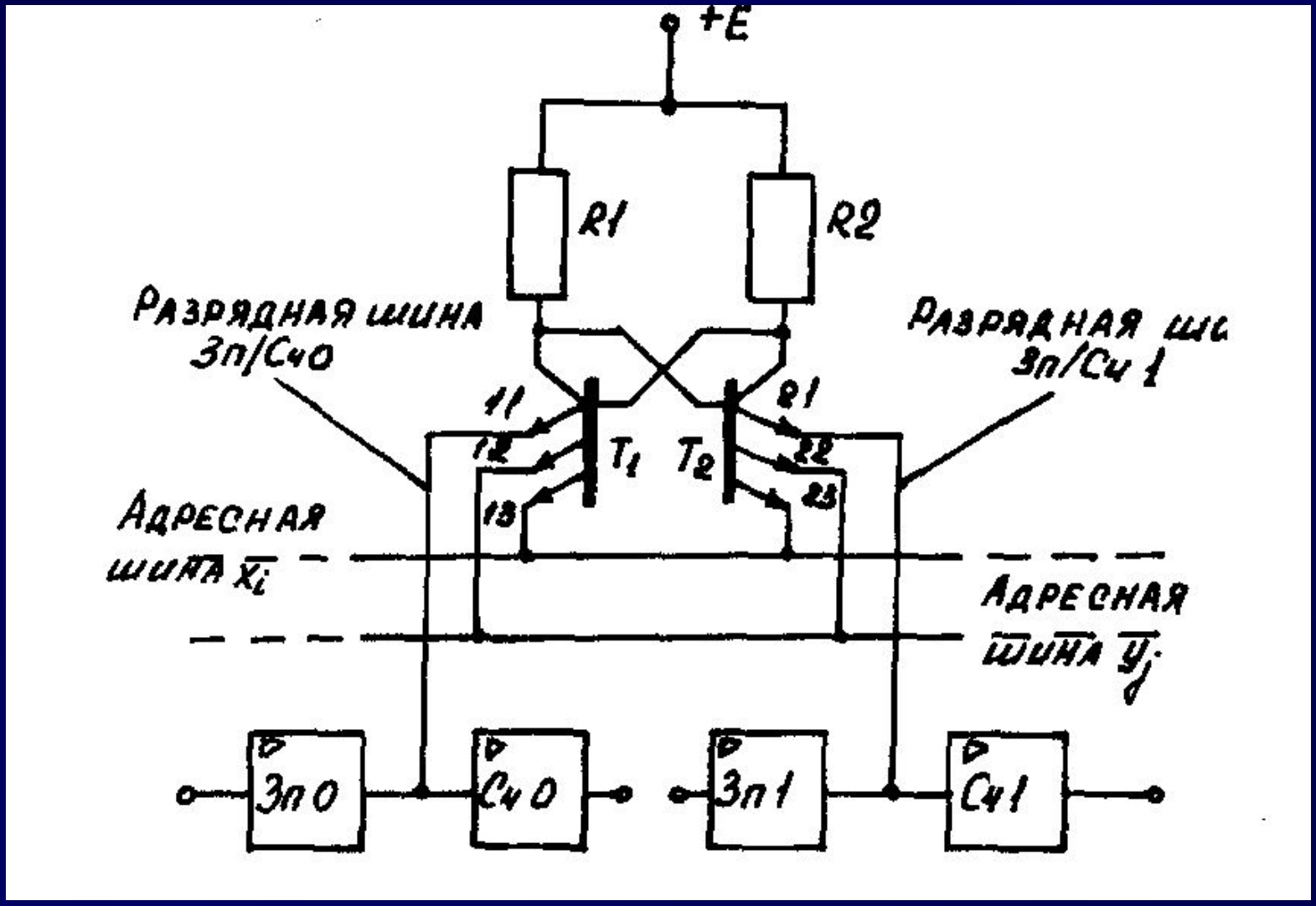
Статическая память (Static Random Access Memory, SRAM), как и следует из ее названия, способна хранить информацию в статическом режиме — то есть сколь угодно долго при отсутствии обращений (но при наличии питающего напряжения). Ячейки статической памяти реализуются на триггерах — элементах с двумя устойчивыми состояниями. По сравнению с динамической памятью эти ячейки более сложные и занимают больше места на кристалле, однако они проще в управлении и не требуют регенерации. Быстродействие и энергопотребление статической памяти определяется технологией изготовления и схемотехникой запоминающих ячеек. Самая экономичная КМОП-память (CMOS memory) имеет значительное время доступа (более 100 нс), но зато пригодна для длительного хранения информации при питании от маломощной батареи и применяется в PC. Самая быстродействующая статическая память имеет время доступа в несколько наносекунд, что позволяет ей работать на частоте системной шины процессора, не требуя от него тактов ожидания. Объем памяти микросхем SRAM уже достиг 32 Мбит. Относительно высокие удельная стоимость хранения информации и энергопотребление при низкой плотности упаковки не позволяют использовать SRAM в качестве основной памяти компьютеров. В PC микросхемы SRAM в основном применяются для построения вторичного кэша; они могут располагаться как на системной плате, так и на картридже процессора.

Статические ОЗУ

Запоминающими элементами статического ОЗУ являются триггерные ячейки, и информация в них хранится до выключения питания. Статические ОЗУ наиболее быстродействующие, но имеют повышенное энергопотребление, невысокую плотность размещения элементов на кристалле и соответственно невысокую емкость (поэтому более дорогие) и используются для организации кэш-памяти.


+E
ЭЭ
Разрядная шина 0
Разрядная шина 1
X1
X2
Xк
A0
A1
A2
A3
Дешифратор шин X
Адресные формирователи
y1
y2
yт
A4
A5
A6
A7
Дешифратор шин У
Усилители - формирователи записи,
усилители считывания
Выборка корпуса CS
Запись W
Информационный вход D
Схема управления
Информационный выход F

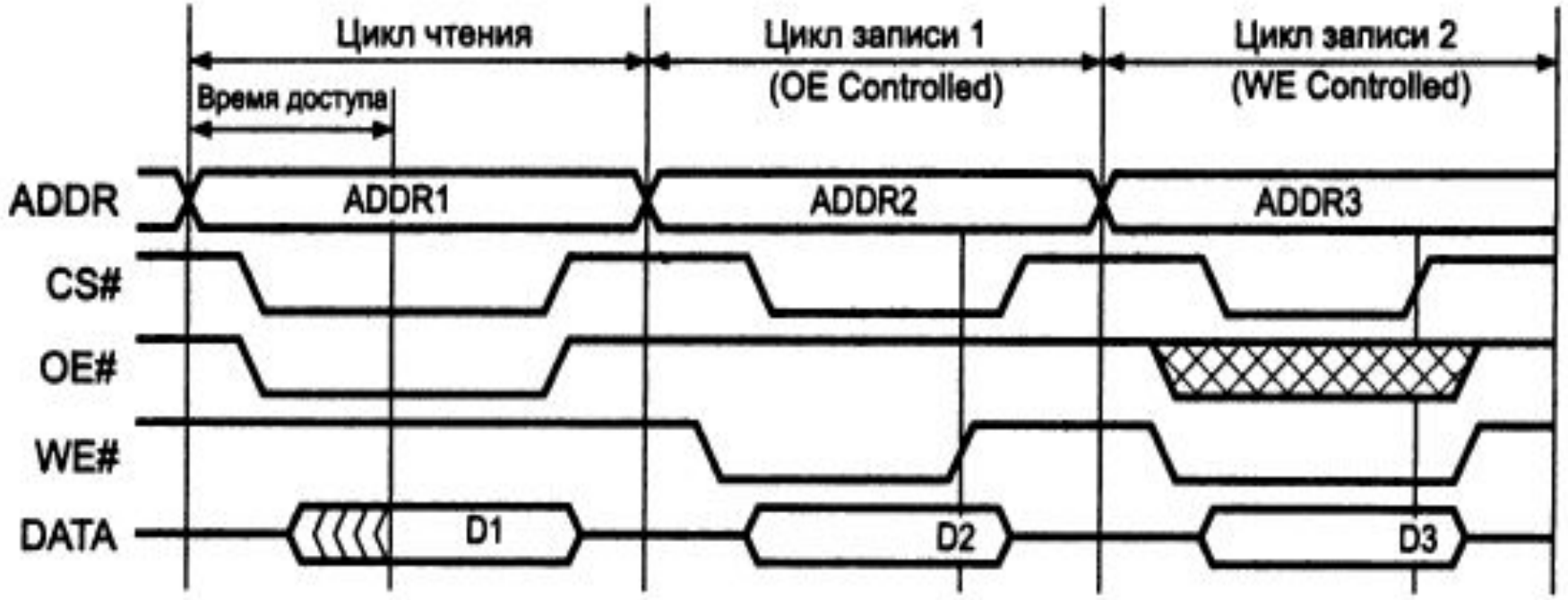
Статические ОЗУ



Разновидности статической памяти

Асинхронная статическая память (Asynchronous SRAM, Async SRAM), она же обычная, или стандартная, подразумевается под термином SRAM по умолчанию, когда тип памяти не указан (до некоторых пор ему действительно не было альтернативы).

Микросхемы этого типа имеют простейший асинхронный интерфейс, включающий шину адреса, шину данных и сигналы управления CS#, OE# и WE#. Микросхема выбирается низким уровнем сигнала CS# (Chip Select), низкий уровень сигнала OE# (Output Enable) открывает выходные буферы для считывания данных, низким уровнем WE# (Write Enable) разрешается запись.



Временные диаграммы чтения и записи асинхронной статической памяти

Время доступа – задержка появления действительных данных на выходе относительно момента установления адреса – у стандартных микросхем SRAM составляет 10, 12, 15 или 20 нс, что позволяет процессору выполнять пакетный цикл чтения *2–1–1–1* (то есть без тактов ожидания) на частоте системной шины до 33 МГц. Наиболее быстрая память имеет время доступа 8 нс. Объем микросхемы SRAM достиг 4 Мбит.

Стековая память

Стековой называют память, доступ к которой организован по принципу: “последним записан - первым считан” (last input first output - LIFO).

С точки зрения реализации механизма доступа к стековой памяти выделяют аппаратный и аппаратно-программный (внешний) стек.

Аппаратный стек представляет собой совокупность регистров, связи между которыми организованы таким образом, что при записи и считывании данных содержимое стека автоматически сдвигается.

Наиболее распространенным в настоящее время и, возможно, лучшим вариантом организации стека в ЭВМ является **использование области памяти**. **Для адресации стека** используется **указатель стека**, который предварительно загружается в регистр и **определяет адрес последней занятой ячейки**.

Кэш-память

Кэш-память представляет собой **быстродействующее ЗУ**, размещенное на одном кристалле с процессором или внешнее по отношению к процессору, и служит высокоскоростным буфером между процессором и относительно медленной основной памятью.

Идея кэш-памяти основана на прогнозировании наиболее вероятных обращений процессора к оперативной памяти. В основу такого подхода положен **принцип временной и пространственной локальности программы**.

Если процессор обратился к какому-либо объекту оперативной памяти, с высокой долей вероятности он вскоре обратится к близлежащим объектам. Эту ситуацию иллюстрирует обращение к массиву данных или любой линейный или циклический алгоритм. Такой концепции соответствует принцип пространственной локальности, когда непрерывные блоки информации переносятся поближе к процессору (в кэш).

В переводе слово «cache» (*кэш*) означает «тайный склад», «тайник», «заначка». Тайна этого склада заключается в его «прозрачности» – адресуемой области памяти для программы он не добавляет. Кэш является дополнительным быстродействующим хранилищем копий блоков информации из основной памяти, вероятность обращения к которым в ближайшее время велика. Кэш не может хранить копию всей основной памяти, поскольку его объем во много раз меньше объема основной памяти. Он хранит лишь ограниченное количество блоков данных и *каталог* (cache directory) – список их текущего соответствия областям основной памяти. Кроме того, кэшироваться может и не вся оперативная память, доступная процессору: во-первых, из-за технических ограничений может быть ограничен максимальный объем кэшируемой памяти; во-вторых, некоторые области памяти могут быть объявлены некэшируемыми (настройкой регистров чипсета или процессора). Если установлено больше оперативной памяти, чем возможно кэшировать, обращение к некэшируемой области ОЗУ будет медленным. Таким образом, увеличение объема ОЗУ, теоретически всегда благотворно влияющее на производительность, может снизить скорость работы определенных компонентов, попавших в некэшируемую память. В ОС Windows память распределяется, начиная с верхних адресов физической памяти в результате в некоторых конфигурациях в некэшируемую область может попасть ядро ОС.

При каждом обращении к памяти контроллер кэш-памяти по каталогу проверяет, есть ли действительная копия затребованных данных в кэше. Если она там есть, то это случай *кэш-попадания* (cache hit) и данные берутся из кэш-памяти. Если действительной копии там нет, это случай *кэш-промаха* (cache miss) и данные берутся из основной памяти. В соответствии с алгоритмом кэширования блок данных, считанный из основной памяти, при определенных условиях замещает один из блоков кэша. От интеллектуальности алгоритма замещения зависит процент попаданий и, следовательно, эффективность кэширования. Поиск блока в списке должен производиться достаточно быстро, чтобы «задумчивостью» в принятии решения не свести на нет выигрыш от применения быстродействующей памяти. Обращение к основной памяти может начинаться одновременно с поиском в каталоге, а в случае попадания – прерываться (архитектура look aside). Это экономит время, но лишние обращения к основной памяти ведут к увеличению энергопотребления. Другой вариант: обращение к основной памяти начинается только после фиксации промаха (архитектура look through); при этом теряется по крайней мере один такт процессора, зато экономится энергия.

Flash-память (**1988** г., фирма **Intel)**

Иногда утверждают, что название Flash применительно к типу памяти переводится как "вспышка". На самом деле это не совсем так. Одна из версий его появления говорит о том, что впервые в 1989-90 году компания Toshiba употребила слово Flash в контексте "быстрый, мгновенный" при описании своих новых микросхем.

Вообще, **изобретателем считается Intel, представившая в 1988 году флэш-память с архитектурой NOR.** Годом позже Toshiba разработала **архитектуру NAND**, которая и сегодня используется наряду с той же NOR в микросхемах флэш. Собственно, сейчас можно сказать, что это два различных вида памяти, имеющие в чем-то схожую технологию производства.

Особенности **Flash-**памяти

Среди **главных достоинств** можно назвать следующие:

- **энергонезависимость**, т.е. способность хранить информацию при выключенном питании (энергия расходуется только в момент записи данных);
- **информация может храниться очень длительное время** (десятки лет);
- **сравнительно небольшие размеры**;
- **высокая надежность хранения данных**, в том числе устойчивость к механическим нагрузкам;
- **не содержит движущихся деталей** (как в жестких дисках).

Особенности **Flash-**памяти

Основные недостатки флэш-памяти:

- **невысокая скорость передачи данных** (в сравнении с динамической оперативной памятью);
- **незначительный объем** (по сравнению с жесткими дисками);
- **ограничение по количеству циклов перезаписи** (хотя эта цифра в современных разработках очень высока – более миллиона циклов).

Флэш-память строится на однотранзисторных элементах памяти с "плавающим" затвором, что обеспечивает высокую плотность хранения информации. Существуют различные технологии построения базовых элементов флэш-памяти, разработанные ее основными производителями. Эти технологии отличаются количеством слоев, методами стирания и записи данных, а также структурной организацией, что отражается в их названии. **Наиболее широко известны NOR и NAND типы флэш-памяти, запоминающие транзисторы в которых подключены к разрядным шинам, соответственно, параллельно и последовательно.**

Архитектура **Flash-**памяти

В настоящее время можно выделить две основные структуры построения флэш-памяти: **память на основе ячеек NOR** (логическая функция ИЛИ-НЕ) и NAND (логическая функция И-НЕ). **Структура NOR состоит из параллельно включенных элементарных ячеек хранения информации**. Такая организация ячеек обеспечивает **произвольный доступ к данным и побайтную запись информации**.

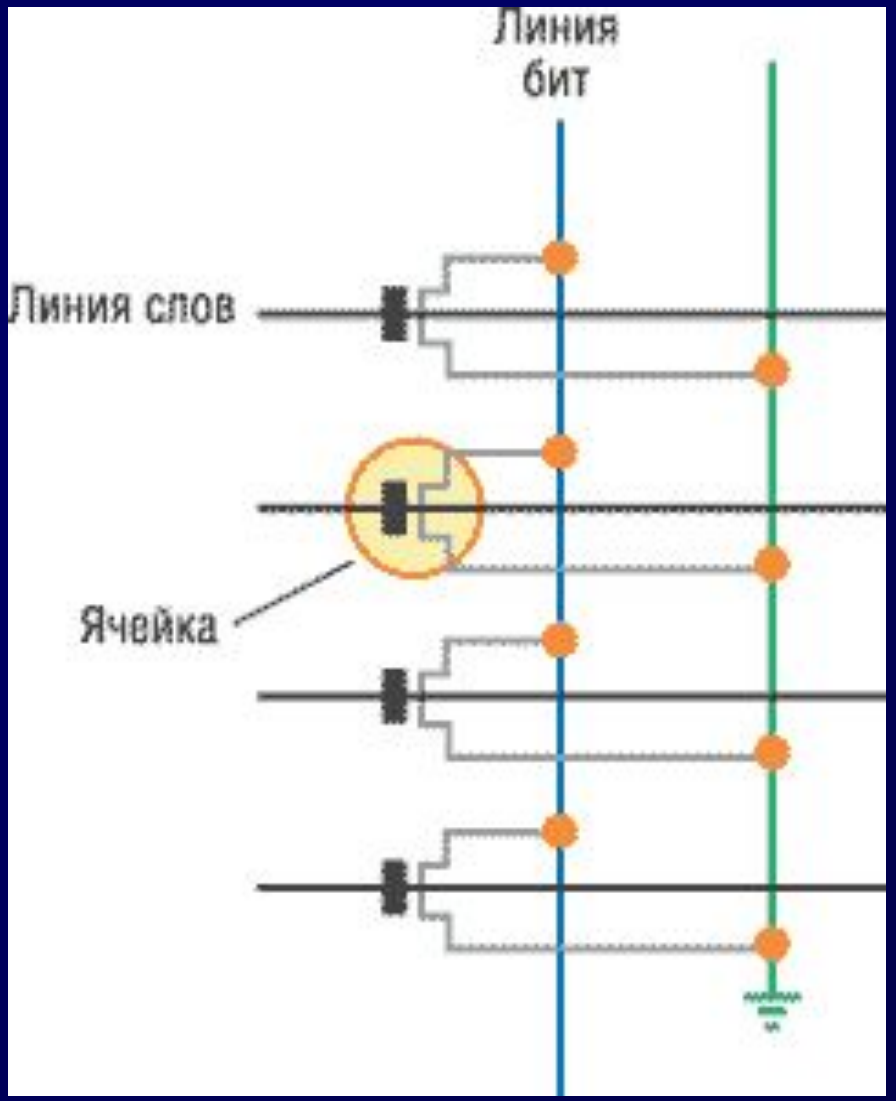


Схема ячейки **NOR**

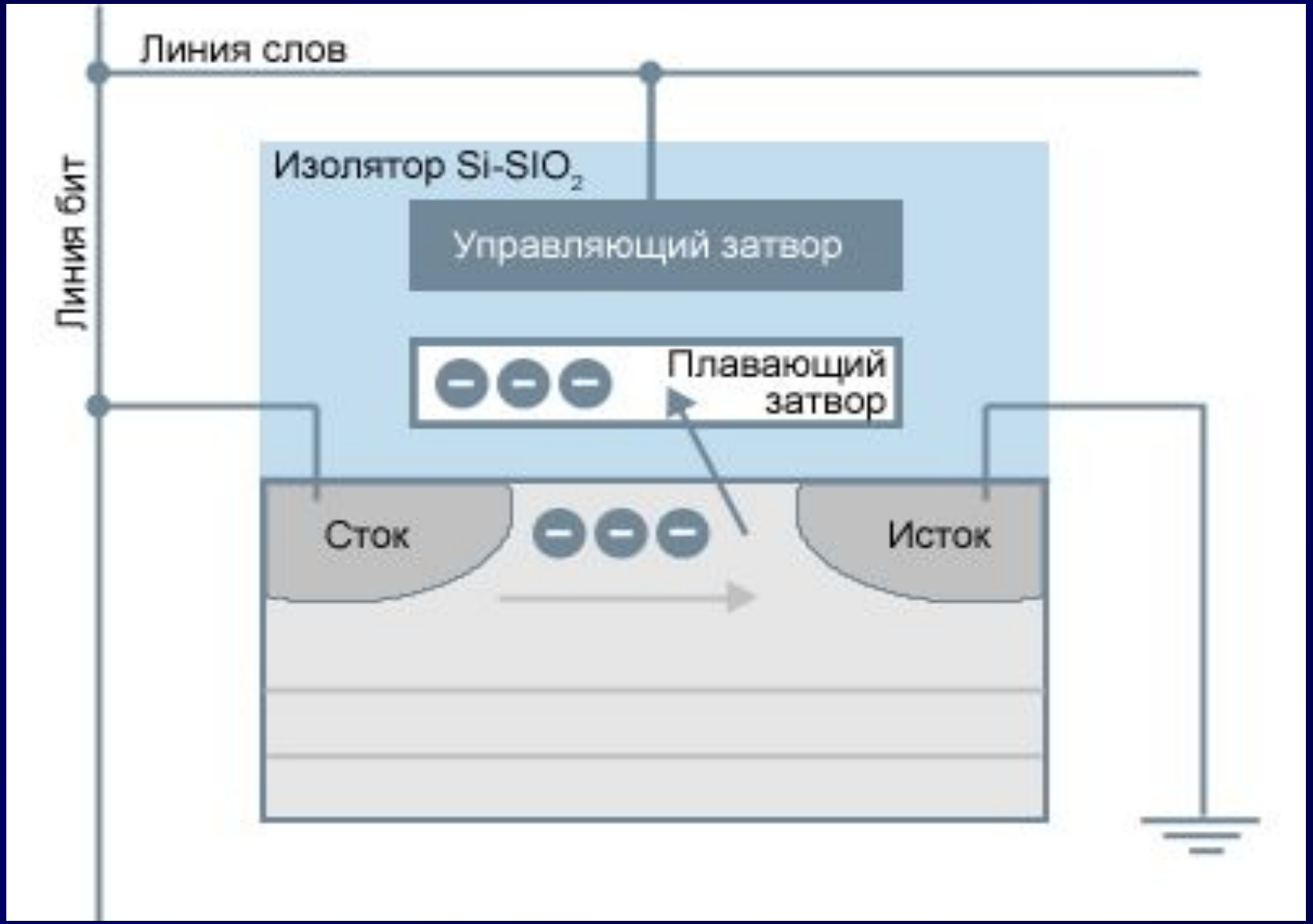
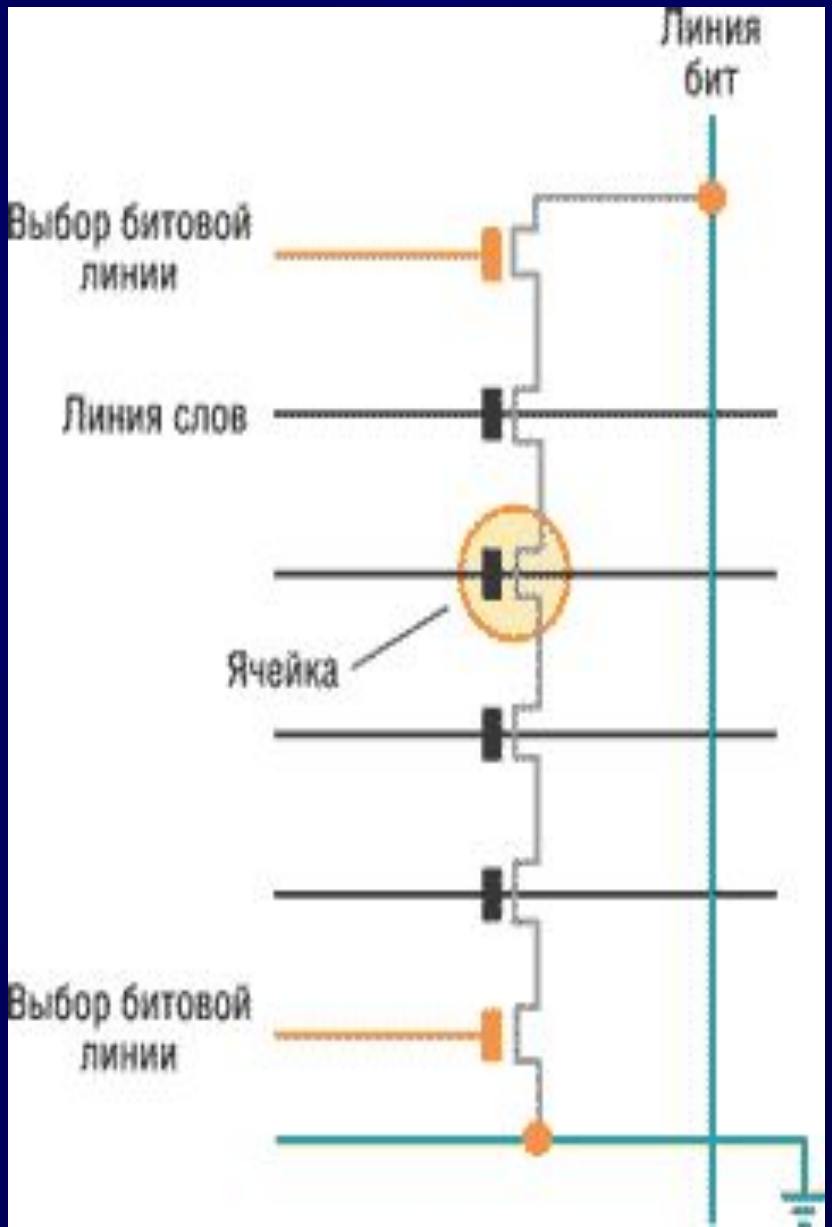


Схема ячейки NOR характерна для большинства флэш-чипов и представляет из себя транзистор с двумя изолированными затворами: управляющим (control) и плавающим (floating). Важной особенностью последнего является способность удерживать электроны, то есть заряд. Также в ячейке имеются так называемые «сток» и «исток». При программировании между ними, вследствие воздействия положительного поля на управляющем затворе, создается канал — поток электронов. Некоторые из электронов, благодаря наличию большей энергии, преодолевают слой изолятора и попадают на плавающий затвор. На нем они могут храниться в течение нескольких лет. **Определенный диапазон количества электронов (заряда) на плавающем затворе соответствует логической единице, а все, что больше его, — нулю**. При чтении эти состояния распознаются путем измерения порогового напряжения транзистора. Для стирания информации на управляющий затвор подается высокое отрицательное напряжение, и электроны с плавающего затвора переходят (туннелируют) на исток. В технологиях различных производителей этот принцип работы может отличаться по способу подачи тока и чтению данных из ячейки.

Схема ячейки **NAND**

В основе структуры NAND лежит принцип последовательного соединения элементарных ячеек, образующих группы (по 16 ячеек в одной группе), которые объединяются в страницы, а страницы - в блоки. При таком построении массива памяти **обращение к отдельным ячейкам невозможно. Программирование выполняется одновременно только в пределах одной страницы,** а при стирании обращение происходит к блокам или к группам блоков.



Различия в организации структуры между памятью NOR и NAND находят свое отражение в их характеристиках. При работе со сравнительно большими массивами данных **процессы записи/стирания** в памяти **NAND выполняются значительно быстрее**, чем в памяти NOR. Поскольку 16 прилегающих друг к другу ячеек памяти **NAND** соединены последовательно, без контактных промежутков, **достигается высокая плотность** размещения ячеек на кристалле, что позволяет получить **большую емкость при одинаковых технологических нормах**. Последовательная организация ячеек обеспечивает высокую степень масштабируемости, что делает **NAND-флэш лидером в гонке наращивания объемов памяти.**

В структуре флэш-памяти для хранения 1 бита информации задействуется только один элемент (транзистор), в то время как в энергозависимых типах памяти для этого требуется несколько транзисторов и конденсатор. Это позволяет существенно уменьшить размеры выпускаемых микросхем, упростить технологический процесс, а следовательно, снизить себестоимость. Но и **1 бит - далеко не предел**.

Еще в 1992 г. команда инженеров корпорации Intel начала разработку устройства флэш-памяти, одна ячейка которого хранила бы более одного бита информации. Еще **в сентябре 1997 г.** была анонсирована **микросхема памяти Intel StrataFlash емкостью 64 Мбит, одна ячейка которой могла хранить 2 бита данных**.

Технология **StrataFlash (Intel)**

В технологии StrataFlash были **использованы элементы двух разных типов флэш-памяти: NAND и NOR**. Доступ к флэш-памяти NOR осуществляется без проверки ошибок, поскольку в этом нет необходимости. Флэш-память NAND не имеет такой надежности, как NOR-память, но она дешевле в производстве, а, кроме того, чтение и запись данных в память NAND происходит намного быстрее, чем в NOR. Это быстродействие дополнительно увеличивается за счет использования в комплекте с этой памятью модулей ОЗУ. **В StrataFlash инженеры Intel объединили два типа флэш-памяти, оптимизировав ее и для хранения данных, и для записи программ.** *Первый модуль памяти StrataFlash состоял из нескольких кристаллов, часть из которых была модулями ОЗУ, а другая представляла собой непосредственно флэш-память*.

Кроме того, **сегодня существуют образцы с 4-битными ячейками**. В такой памяти **используется технология многоуровневых ячеек.** Они имеют обычную структуру, а **отличие заключается в том, что их заряд делится на несколько уровней, каждому из которых в соответствие ставится определенная комбинация битов**. Теоретически прочитать/записать можно и более 4 бит, однако на практике возникают проблемы с устранением шумов и с постепенной утечкой электронов при продолжительном хранении.

Отметим также, что **Intel первой в индустрии наладила выпуск многоуровневых микросхем флэш-памяти класса NOR емкостью 1 Гбит** для мобильных устройств, используя 65-нм производственную технологию.

Флэш-память имеет время доступа при чтении 35–200 нс. Стирание информации (поблочное или во всей микросхеме) у микросхем середины 90-х годов занимает 1–2 с, программирование (запись) байта – порядка 10 мкс. У современных микросхем время стирания и записи заметно сократилось. Процедура записи от поколения к поколению упрощается . Применяются различные методы программной и аппаратной защиты от ошибочного стирания (записи). Программной защитой является ключевая последовательность команд, нарушение которой не позволяет начать операции стирания и записи. Аппаратная защита не дает выполнять стирание и запись, если на определенные входы не поданы требуемые уровни напряжения. Аппаратная защита может защищать как весь массив целиком, так и отдельные блоки.

По организации массива в плане стирания групп ячеек различают следующие архитектуры:

- *Bulk Erase* (BE) – все ячейки памяти образуют единый массив. Запись возможна в произвольную ячейку. Стереть можно только сразу весь массив.
- *Boot Block* (BB) – массив разделен на несколько блоков разного размера, стираемых независимо, причем один из блоков имеет дополнительные средства защиты от стирания и записи.
- *Flash File* – массив разделен на несколько равноправных независимо стираемых блоков обычно одинакового размера, что позволяет их называть микросхемами с симметричной архитектурой (Symmetrical Architecture, SA).

Архитектура BE применялась только в микросхемах первого поколения, ее недостатки вполне очевидны (получается просто аналог EEPROM с более удобным способом стирания и интерфейсом программирования). Все современные микросхемы секторированы (разбиты на отдельно стираемые блоки), так что остается лишь деление на симметричную и несимметричную архитектуры.

В *симметричной архитектуре* (SA), как правило, память разбивается на блоки по 64 Кбайт; один из крайних блоков (с самым большим или самым маленьким адресом) может иметь дополнительные средства защиты.

В *асимметричной архитектуре* один из 64-килобайтных блоков разбивается на 8 блоков по 8 Кбайт. Один из блоков имеет дополнительные аппаратные средства защиты от модификации и предназначается для хранения жизненно важных данных, не изменяемых при запланированных модификациях остальных областей. Эти микросхемы специально предназначены для хранения системного программного обеспечения (BIOS), а привилегированный блок (*Boot Block*) – для хранения минимального загрузчика, позволяющего загрузить (например, с дискеты) и выполнить утилиту программирования основного блока флэш-памяти. В обозначении этих микросхем присутствует суффикс *T* (Top) или *B* (Bottom), определяющий положение загрузочного блока либо в старших, либо в младших адресах соответственно. Первые предназначены для процессоров, стартующих со старших адресов (в том числе x86, Pentium), вторые – для стартующих с нулевого адреса, хотя возможны и противоположные варианты, когда некоторые биты шины адреса перед подачей на микросхему памяти инвертируются.

Для хранения BIOS появились микросхемы *флэш-памяти с интерфейсом LPC*, называемые хабами (firmware hub).

Для некоторых сфер применения требуются специальные меры по блокированию изменения информации пользователем. Так, Intel в некоторые микросхемы вводит однократно записываемые (One-Time-Programmable, OTP) регистры. Один 64-битный регистр содержит уникальный заводской номер, другой может программироваться пользователем (изготовителем устройства) только однажды.

Фирма Intel выпускает микросхемы «Wireless Flash Memory» – за интригующим названием (беспроводная флэш-память) скрывается, конечно же, обычный электрический интерфейс (с проводами). Они ориентированы на применение в средствах беспроводной связи (сотовые телефоны с доступом в Интернет): питание – 1,85 В, наличие регистров OTP для защиты от мошенничества и т. п.

Применение **Flash-**памяти

Современные технологии производства флэш-памяти позволяют использовать ее для различных целей. Непосредственно **в компьютере эту память применяют для хранения BIOS** (базовой системы ввода-вывода), что позволяет, при необходимости, производить обновление последней, прямо на рабочей машине.

Распространение получили, так называемые, USB-Flash накопители, эмулирующие работу внешних винчестеров. Эти устройства подключается, обычно, к шине USB и состоит из собственно флэш-памяти, эмулятора контроллера дисковода и контроллера шины USB. При включении его в систему (допускается "горячее" подключение и отключение) устройство с точки зрения пользователя ведет себя как обычный (съемный) жесткий диск. Конечно, производительность его меньше, чем у жесткого диска.

Флэш-память нашла широкое применение **в различных модификациях карт памяти**, которые обычно используются в цифровых видео- и фотокамерах, плеерах, телефонах.

Необходимо отметить, что надежность и быстродействие флэш-памяти постоянно увеличиваются. Теперь количество циклов записи/перезаписи выражается семизначной цифрой, что позволяет практически забыть о том, что когда-то на карту памяти можно было записывать информацию лишь ограниченное число раз. Современные USB-Flash накопители уже рассчитаны на шину USB 3.0 (и она им действительно необходима). На рынке появляется все больше пылевлагозащищенных устройств. При этом все большее и большее количество производителей встраивают кардридеры в настольные корпуса персональных компьютеров. Это безусловно свидетельствует о том, что данный тип памяти уже стал одним из популярнейших.