

Тема 10. Организации памяти ЭВМ и систем

Основные вопросы:

- 10.1. Введение
- 10.2. Системы памяти – критерии оценки.
- 10.3. Основные характеристики современных запоминающих устройств (ЗУ).

Классификация ЗУ

- 10.4. Полупроводниковые запоминающие устройства: организация ЗУ с произвольным доступом
- 10.5. Постоянные запоминающие устройства: разновидности
- 10.6. Регистровая память, кэш - память

10.1. Введение

Память представляет собой одну из важнейших подсистем ЭВМ, во многом определяющую их производительность. Тем не менее, в течение всей истории развития вычислительных машин она традиционно считается их *"узким местом"*.

Память, запоминающие устройства (ЗУ) (англ. memory, storage) – среда или функциональная часть ЭВМ, предназначенная для приема, хранения и избирательной выдачи данных.

Системы памяти (СП) современных ЭВМ представляют собой **совокупность аппаратных средств**, предназначенных для хранения используемой в ЭВМ информации. На рис. 10.1 представлен возможный состав системы памяти.

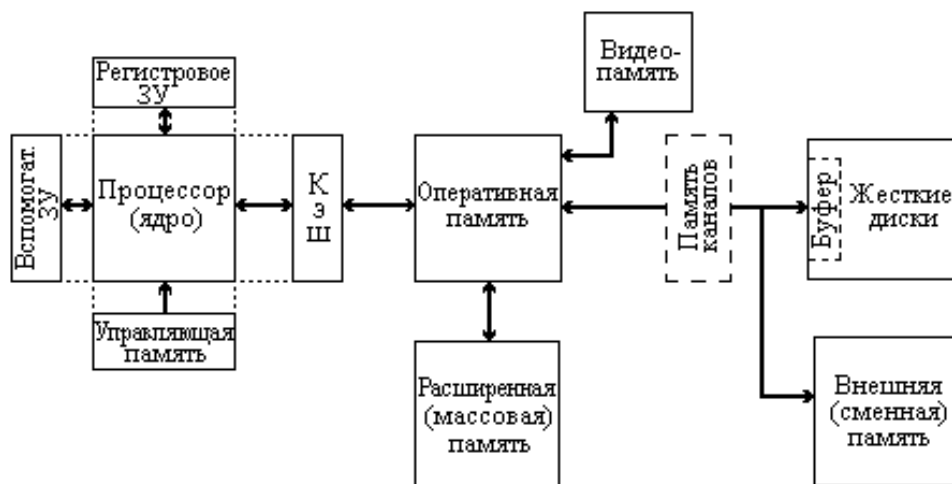


Рис. 10.1. Возможный состав системы памяти ЭВМ

К хранимой в СП информации относятся обрабатываемые данные, прикладные программы, системное программное обеспечение и служебная информация различного назначения. **К системе памяти можно отнести и программные средства**, организующие управление ее работой в целом, а также **драйверы** различных видов запоминающих устройств.

10.2. Системы памяти – критерии оценки

Рис. 10.2 дает представление о группах факторов, влияющих на работоспособность ЗУ, которые определяют информационную и конструктивную надежность и эффективность. При объединении отдельных ЗУ в систему к этим факторам добавляется еще целый ряд, связанный со взаимодействием ЗУ между собой в составе системы памяти.

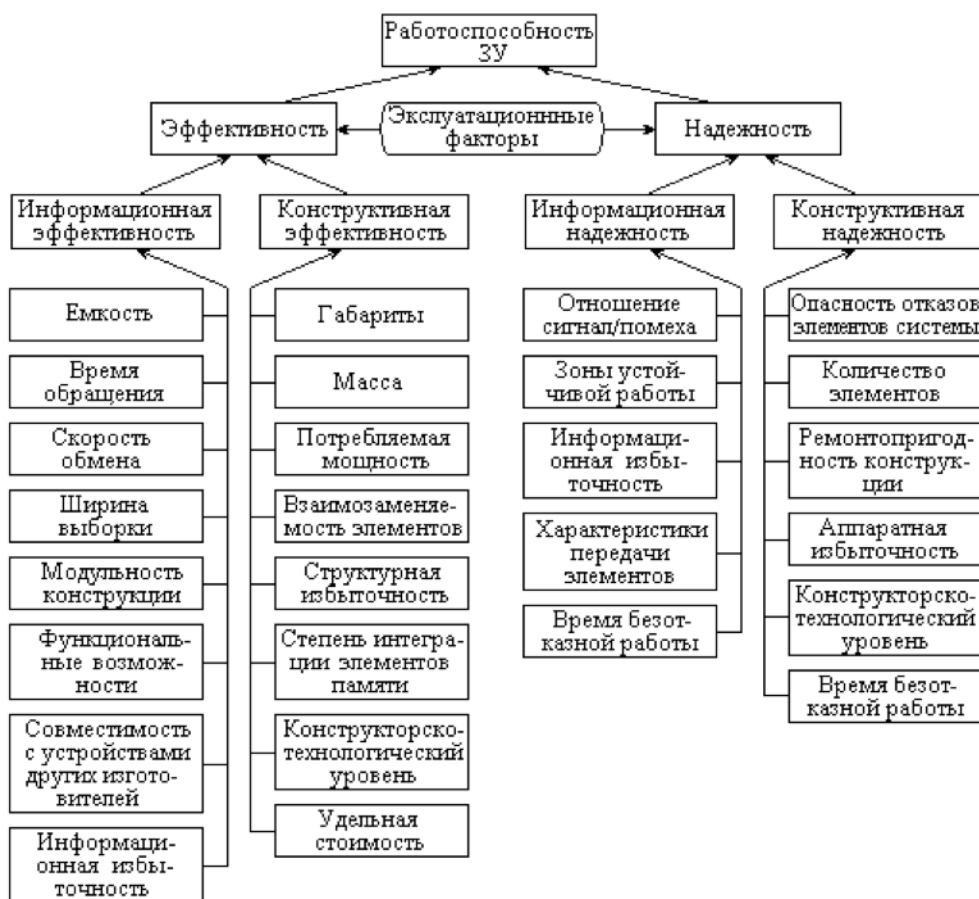


Рис. 10.2. Факторы, определяющие работоспособность ЗУ

С целью более полного учета характера функционирования и окружения СП при выборе критерия ее оценки следует рассматривать эту систему как компоненту вычислительной машины (системы), ориентируясь на назначение последней.

Любой критерий оценки должен включать основные характеристики оцениваемой системы, к которым относятся:

- емкость системы памяти,
- среднее время обращения к ней,
- пропускная способность,
- стоимость,
- надежность.

Ряд характеристик, например, радиационная устойчивость, габариты, масса, энергопотребление, в типовых применениях могут не учитываться. Хотя, если речь идет, например, о мобильных системах, последние три из названных характеристик имеют важное значение.

Емкость $E_{СП}$ системы памяти можно рассматривать в двух аспектах:

- либо как сумму объемов всех ЗУ, входящих в состав СП (технической емкости),
- либо как количество информации (программ и данных/эффективной емкости), которое можно разместить в системе.

В первом случае можно говорить о технической емкости СП, во втором – об эффективной емкости. Понятно, что эффективная емкость всегда меньше технической, так как она определяется не только собственно составом СП, но и методами организации хранения данных, методами управления памятью и др. Например, можно вспомнить о файловых системах, которые накладывают ограничения снизу на место на диске, занимаемое даже самым небольшим файлом.

Среднее время обращения $T_{обр}$ к СП можно определить через частоты обращений к отдельным устройствам системы и времена обращений $t_{обp_i}$ к этим устройствам как

$$T_{обр} = \frac{\sum_i f_i t_{обp_i}}{\sum_i f_i}$$

где f_i есть среднее количество обращений к i -му ЗУ в единицу времени. Очевидно, что в этом случае $T_{обр}$ в значительной степени зависит от относительных частот обращения к различным ЗУ, а не только от времени обращения к ним.

Средняя пропускная способность B системы памяти – количество информации, которое можно передать в СП или извлечь из нее в единицу времени. В общем виде B можно определить, усреднив (взвешенно) пропускные способности отдельных ЗУ, входящих в состав СП.

Стоимость $C_{СП}$ системы памяти определяется как сумма стоимостей всех входящих в ее состав ЗУ, контроллеров и дополнительных аппаратных средств, используемых для управления памятью. Строго говоря, некоторые средства управления памятью, как и сами ЗУ, могут быть интегрированы в процессор или в системные микросхемы (микросхемы чипсета или др.). В этом случае собственно стоимость ЗУ и средств управления ими приходится определять путем сравнения с аналогичными микросхемами, обладающими другими параметрами, или приближенно.

Надежность СП определяется надежностью составляющих ее блоков.

Все характеристики систем памяти взаимосвязаны между собой и имеют противоречивый характер. Например, уменьшение времени обращения к СП связано с использованием более быстродействующих, а, следовательно, и дорогостоящих ЗУ. Увеличения пропускной способности дисковых ЗУ можно достичь, используя аппаратные или алгоритмические методы диспетчеризации, что приводит либо к росту стоимости и снижению надежности СП, либо к увеличению расходов времени на работу операционной системы.

10.3. Основные характеристики современных запоминающих устройств (ЗУ). Классификация ЗУ

Запоминающие устройства (ЗУ) характеризуются рядом параметров, определяющих возможные области применения различных типов таких устройств. К основным параметрам, по которым производится наиболее общая оценка ЗУ, относятся их

- **информационная емкость (E),**
- **время обращения (T) и**
- **стоимость (C).**

Под **информационной емкостью ЗУ** понимают количество информации, измеряемое в байтах, килобайтах, мегабайтах или гигабайтах, которое может храниться в запоминающем устройстве.

Обычно информационная емкость учитывает только полезный объем хранимой информации, который не включает объем памяти, расходуемый на служебную информацию, контрольные разряды или байты, резервные области (например, интервал между концом дорожки диска и ее началом), дорожки синхросигналов и пр.

Время обращения к ЗУ различных типов определяется по-разному.

В качестве примера можно рассмотреть оперативные ЗУ и жесткие диски.

Оперативные ЗУ обычно реализуются как ЗУ с произвольным доступом. Это означает, что доступ к данным, физически организованным в виде двумерного массива (матрицы элементов памяти), производится с помощью схем дешифрации, выбирающих нужные строку и столбец массива по их номерам (адресам), как показано на рис. 10.3. Поэтому время $T_{\text{обр}}$ обращения к ним определяется, в случае отсутствия дополнительных этапов (таких, например, как передача адреса за два такта), временем срабатывания схем дешифрации адреса и собственно временами записи или считывания данных.

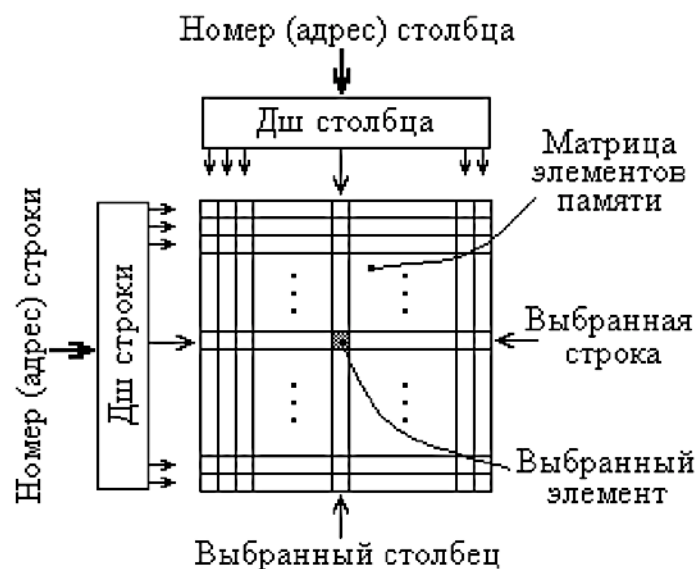


Рис. 10.3. Произвольный доступ к массиву элементов памяти ЗУ

Емкости оперативных ЗУ – порядка 256 Мб – 8 Гб.

Процесс обращения (чтения или записи) к жесткому диску представлен на рис.10.4. Он включает в себя 3 этапа:

1. перемещение блока головок чтения/записи на нужную дорожку (*a*),
2. ожидание подхода требуемого сектора под головки чтения/записи (*б*)
3. собственно, передача данных, считываемых/записываемых.

Каждый из этих этапов занимает определенное время, входящее в общее время обращения к диску. Все этапы так или иначе связаны с механическими перемещениями, поэтому их времена сравнительно велики и составляют величины порядка единиц миллисекунд.

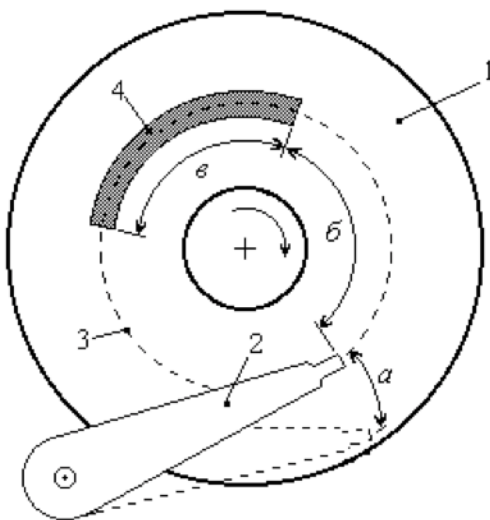


Рис. 10.4. Обращение к жесткому диску

(1 – пластина диска, 2 – блок чт/зп, 3 – дорожка (цилиндр), 4 – файл, *a* – поиск дорожки (перемещение блока головок чт/зп), *б* – ожидание подхода файла под блок головок, *в* – передача данных)

Время перемещения блока головок, обычно называемое изготовителями дисков временем поиска (*seek time*), зависит от количества дорожек, на которое надо переместить блок головок. Минимальное время затрачивается на перемещение блока головок на соседнюю дорожку (цилиндр) – составляет порядка 1–2 мс. Максимальное время требуется на перемещение блока головок от крайней дорожки к центральной или наоборот – может составлять порядка 15–20 мс. Среднее время поиска (перемещения головок) составляет порядка 8–10 мс.

Время ожидания подвода файла (точнее, его первого сектора) под блок головок производители называют также временем задержки (*latency time*). Это время в среднем равно времени половины оборота диска, что, например, при скорости вращения (шпинделя) диска 7200 оборотов/мин, или 120 оборотов/с, составляет 4,2 мс.

Наконец, время передачи данных зависит от количества передаваемых данных (размера файла, если он располагается целиком на последовательных секторах одной дорожки диска) и скорости передачи. Из-за зависимости этого времени от размера файла и его размещения на диске в качестве характеристики диска используют скорость передачи данных (*transfer rate*). Эта скорость определяется как параметрами тракта связи с ЭВМ, так и скоростью считывания данных с диска или записи данных на диск. Обычно пользуются именно этими параметрами, так как

каналы передачи достаточно быстрые, чтобы снижать скорость передачи, а диски имеют буферные ЗУ (кэш диска), скорость обмена данными с которым заметно превышает скорость считывания с диска или записи на диск.

В свою очередь, скорость обмена с диском определяется скоростью его вращения и плотностью записи информации на него. Обе эти величины непрерывно возрастают с развитием технологий изготовления жестких дисков.

Плотность записи информации на диск удваивалась примерно каждый год – полтора.

Стоимость запоминающих устройств также представляет собой важную характеристику. Именно она является одной из причин иерархической организации памяти ЭВМ.

Действительно, хорошо иметь быструю и емкую память. Нужно, чтобы она была и относительно дешевой. Понятно, что эти параметры противоречивы. Поэтому в ЭВМ и строят иерархию памяти.

Определения дорогие и дешевые понимаются не в абсолютном, а в относительном измерении, исходя из стоимости хранения единицы информации (удельной стоимости) в ЗУ. Стоимость хранения 1 Мбайта информации в оперативных ЗУ и на жестких дисках различаются примерно в 100 раз.

Конечно, помимо емкости, времени обращения и стоимости, существуют и другие характеристики памяти такие, как:

- надежность;
- энергопотребление;
- габариты;
- время хранения информации;
- способность сохранять ее при отключении питания и другие.

При определенных условиях эти характеристики могут иметь важное значение. Например, для *ноутбуков* энергопотребление и габариты играют существенную роль, что при обеспечении требуемых значений этих показателей приводит к более высокой стоимости устройств такого класса. Напротив, для *серверов* на первый план выдвигается *требование надежности* сохранения информации.

Перейдем к классификации запоминающих устройств. Существует большое количество различных типов ЗУ, используемых в ЭВМ и системах. Эти устройства различаются рядом признаков:

- принципом действия,
- логической организацией,
- конструктивной и
- технологической реализацией,
- функциональным назначением и т.д.

Большое количество существующих типов ЗУ обуславливает различия в структурной и логической организации (систем) памяти ЭВМ. ***Требуемые характеристики памяти достигаются не только за счет применения ЗУ с соответствующими характеристиками, но в значительной степени за счет особенностей ее структуры и алгоритмов функционирования.***

Память ЭВМ почти всегда является "узким местом", ограничивающим производительность компьютера. Поэтому в ее организации используется ряд приемов, улучшающих временные характеристики памяти и, следовательно, повышающих производительность ЭВМ в целом.

Классификация запоминающих устройств и систем памяти позволяет выделить общие и характерные особенности их организации, систематизировать базовые принципы и методы, положенные в основу их реализации и использования.

Один из возможных вариантов классификации ЗУ представлен на рис. 10.5 В нем устройства памяти подразделяются по двум основным критериям: по **функциональному назначению** (роли или месту в иерархии памяти) и **принципу организации**.

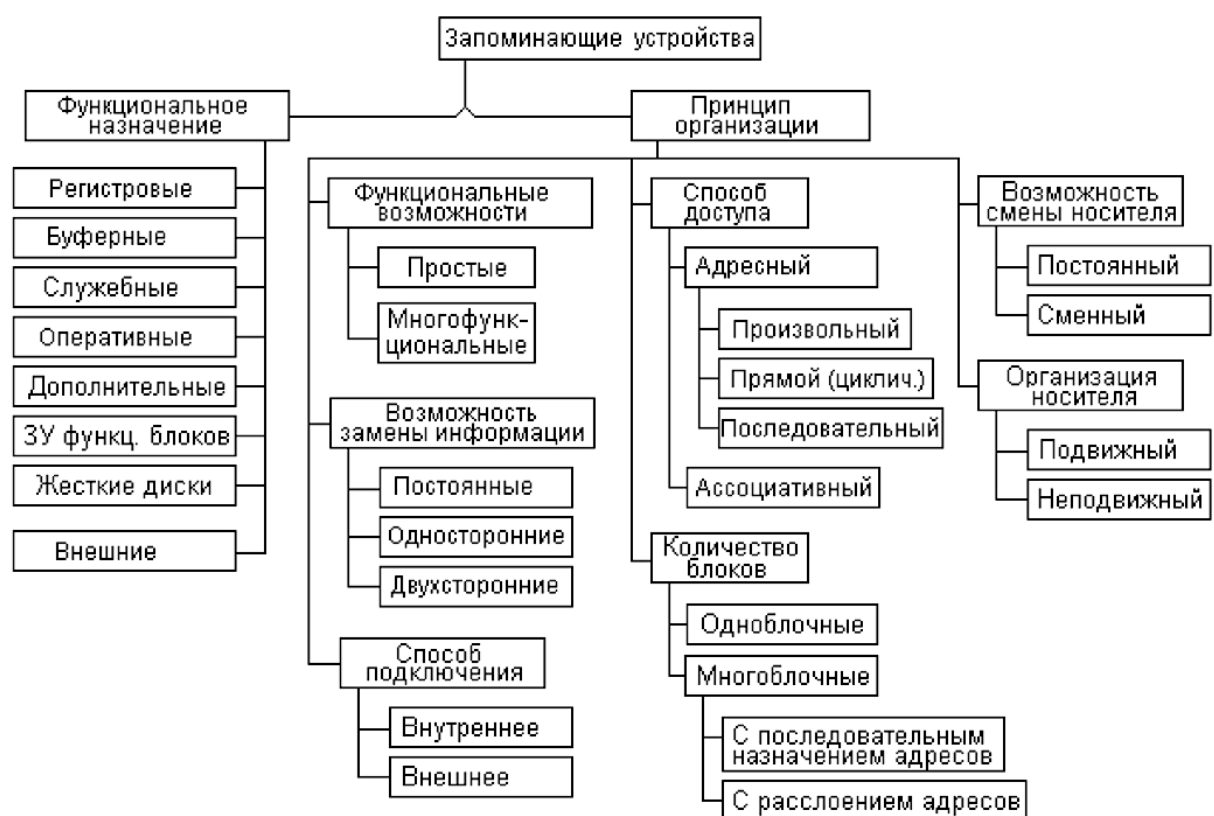


Рис. 10.10. Классификация запоминающих устройств

10.4. Полупроводниковые запоминающие устройства: организация ЗУ с произвольным доступом

Полупроводниковые ЗУ в настоящее время представляют собой большой класс запоминающих устройств, различных по своим функциональным и техническим характеристикам, широко используемых в качестве внутренних ЗУ ЭВМ. Но этим их использование не ограничивается. Подавляющее большинство электронной и бытовой техники перешло на *цифровые методы представления данных* (не только текстовых, но и аудио, графических и видео) и *управления* (использование микроконтроллеров). Различные сферы применения накладывают свои особенности на реализацию полупроводниковых ЗУ, однако это чаще касается их конструктивных особенностей, а принципы построения одинаковы.

Высокое быстродействие полупроводниковых ЗУ обуславливает то, что большинство из них имеет *организацию с произвольным доступом*. Это же высокое быстродействие определяет и основные области применения полупроводниковых ЗУ в ЭВМ: *кэш-память* и *оперативная память*.

В *статических ЗУ* (Static Random Access Memory – SRAM) в качестве элемента памяти используется триггер, поэтому статические ЗУ обладают меньшей плотностью хранения информации.

Однако триггер со времен первых компьютеров был и остается самым быстродействующим элементом памяти. Поэтому статическая память позволяет достичь наибольшего быстродействия, обеспечивая время доступа в единицы и даже десятые доли наносекунд, что и обуславливает ее использование в ЭВМ, главным образом, в высших ступенях памяти – кэш-памяти всех уровней.

Главными недостатками статической памяти являются ее относительно высокие стоимость и энергопотребление.

Конечно, в зависимости от используемой технологии, память будет обладать различным сочетанием параметров быстродействия и потребляемой мощности. Например, статическая память, изготовленная по КМОП-технологии (**КМОП – комплементарный металло-оксидный полупроводник**) может быть светочувствительной.

Основными разновидностями статической памяти **SRAM** с точки зрения организации ее функционирования являются:

- *асинхронная* (Asynchronous) SRAM,
- *синхронная пакетная* (Synchronous Burst) SRAM и
- *синхронная конвейерно-пакетная* (Pipeline Burst) SRAM память.

Синхронная пакетная статическая память (SBSRAM) ориентирована на выполнение пакетного обмена информацией, который характерен для кэш-памяти. Эта память включает в себя внутренний счетчик адреса, предназначенный для перебора адресов пакета, и использует сигналы синхронизации, как и синхронная DRAM память.

Следующим шагом в развитии статической памяти явилась **конвейерно-пакетная память** PBSRAM, обеспечивающая более высокое быстродействие, чем SBSRAM. В нее были введены дополнительные внутренние буферные регистры данных (здесь можно провести аналогию с EDO DRAM памятью) адреса, а в ряде модификаций предусмотрена возможность передачи данных на двойной скорости по переднему и заднему фронтам синхросигнала и используются удвоенные внутренние тракты записи и чтения. Это позволило получить время обращения порядка 2-3 нс и обеспечить передачу данных пакета без задержек на частотах шины более 400 МГц.

Внутренняя логика позволяет переключаться с циклов чтения на циклы записи и наоборот без дополнительных задержек, кроме того, анализируется совпадение адресов записи и чтения для исключения избыточных операций.

Как уже отмечалось, в качестве оперативных ЗУ в настоящее время чаще используются *динамические ЗУ с произвольным доступом* (DRAM). Такое положение обусловлено тем, что недостатки, связанные с необходимостью

регенерации информации в таких ЗУ и относительно невысоким их быстродействием, с лихвой компенсируются другими показателями: малыми размерами элементов памяти и, следовательно, большим объемом микросхем этих ЗУ, а также низкой их стоимостью.

Широкое распространение ЗУ этого типа проявилось также и в разработке многих его разновидностей: *асинхронной*, *синхронной*, **RAMBUS** и других. Перечислим основные разновидности:

1. Синхронная динамическая память DDR SDRAM

обеспечивающая двойную скорость передачи данных (*DDR – Double* или *Dual Data Rate*), в которой за один такт осуществляются две передачи данных – по переднему и заднему фронтам каждого синхроимпульса. Во всем остальном эта память работает аналогично обычной SDRAM памяти. Времена задержек *CAS Latency* для DDR SDRAM могут быть 2 и 2,5 такта

2. Синхронная динамическая память SDRAM

иногда называют SDR SDRAM – *Single Data Rate*.

3. Асинхронная динамическая память DRAM

4. Динамическая память RDRAM

Независимо от того, какова конкретная модификация динамической памяти, запоминающие конденсаторы ее запоминающих элементов разряжаются из-за наличия токов утечки. Постоянная разрядка, как известно, зависит от емкости запоминающего конденсатора и сопротивления цепи тока утечки и может различаться для разных модификаций. Время, в течение которого информация сохраняется в элементе памяти, составляет до нескольких десятков миллисекунд.

Это приводит к необходимости периодического (с периодом не больше, чем время сохранения информации) восстановления зарядов емкостей. Такая процедура и получила название регенерации (*refresh*) динамической памяти. Выполняется она одновременно для целой строки матрицы (банка) элементов памяти, поскольку регенерировать информацию по элементам или по словам (по 8 байт) слишком долго.

Модули динамической полупроводниковой памяти прошли эволюцию от набора микросхем, устанавливаемых на системной плате и заметных по своему регулярному расположению (несколько смежных рядов одинаковых микросхем) до отдельных небольших плат, вставляемых в стандартный разъем (слот) системной платы. Первенство в создании таких модулей памяти обычно относят к фирме IBM.

Основными разновидностями модулей динамических оперативных ЗУ с момента их оформления в виде самостоятельных единиц были:

- 30-контактные однобайтные модули SIMM (DRAM)
- 72-контактные четырехбайтные модули SIMM (DRAM)
- 168-контактные восьмибайтные модули DIMM (SDRAM)
- 184-контактные восьмибайтные модули DIMM (DDR SDRAM)
- 184-контактные (20 из них не заняты) двухбайтные модули RIMM RDDRAM).

Сокращение **SIMM** означает Single In-Line Memory Module – модуль памяти с одним рядом контактов, так как контакты краевого разъема модуля, расположенные в одинаковых позициях с двух сторон платы, электрически соединены. Соответственно **DIMM** означает Dual In-Line Memory Module – модуль памяти с двумя рядами контактов, **RIMM** - Rambus Memory Module – модуль памяти типа Rambus.

10.10. Постоянные запоминающие устройства: разновидности

Постоянные запоминающие устройства (**ПЗУ** или **Read Only Memory - ROM**), которые также часто называют энергонезависимыми (или Non Volatile Storage), обеспечивают сохранение записанной в них информации и при отсутствии напряжения питания. Конечно, под такое определение подпадают и память на жестких и гибких дисках, и компакт диски, и некоторые другие виды ЗУ.

Однако, говоря о постоянных ЗУ, обычно подразумевают устройства памяти с произвольным адресным доступом. Такие ЗУ могут строиться на различных физических принципах и обладать различными характеристиками не только по емкости и времени обращения к ним, но и по возможности замены записанной в них информации.

К началу 2000-х годов наибольшее распространение получили полупроводниковые ПЗУ, элементы памяти которых используют различные модификации диодов и транзисторов и изготавливаются по интегральной технологии.

Непосредственными предшественниками таких ЗУ были магнитные (трансформаторные) ПЗУ, информация в которые записывалась соответствующей прокладкой (прошивкой) проводников ферритовых сердечников. Это обеспечивало при требовавшихся в то время емкостях высокую надежность этих ЗУ в самых тяжелых (в электромагнитном отношении) условиях.

Известны также емкостные и индуктивные ПЗУ, в которых использовались проводники специальной формы, образующие емкостные или индуктивные связи.

В настоящее время исследуются и другие принципы реализации постоянных ЗУ, в некотором смысле возвращающиеся к магнитным и конденсаторным схемам, но на другом уровне развития технологий.

Запись информации в ПЗУ, как правило, существенно отличается от считывания по способу и времени выполнения. Процесс записи для полупроводниковых постоянных ЗУ получил также название “прожига” или программирования, первое из которых связано со способом записи, сводящимся к разрушению (расплавлению, прожигу) соединительных перемычек в чистом ЗУ.

В полупроводниковых ПЗУ в качестве элементов памяти, точнее, в качестве нелинейных коммутирующих и усилительных элементов обычно используются транзисторы. Они объединены в матрицу, выборка данных из которой производится по строкам и столбцам, соответствующим указанному адресу, так же, как и в других ЗУ с произвольным доступом. Один из возможных вариантов структурной схемы полупроводникового ПЗУ, представлен на рис. 10.6. Строго говоря, непосредственно запоминание информации в этом ПЗУ осуществляется плавкой перемычкой, а транзисторы выполняют роль ключей-усилителей. Плавкая перемычка может быть изготовлена из нихрома, поликристаллического кремния или

других материалов. В зависимости от того, как именно работает усилитель считывания (в режиме повторителя или инвертора), наличие перемычки соответствует записи “1” или “0”. Разрушение перемычки (импульсом сильного тока) приводит к записи значения, обратного исходному.

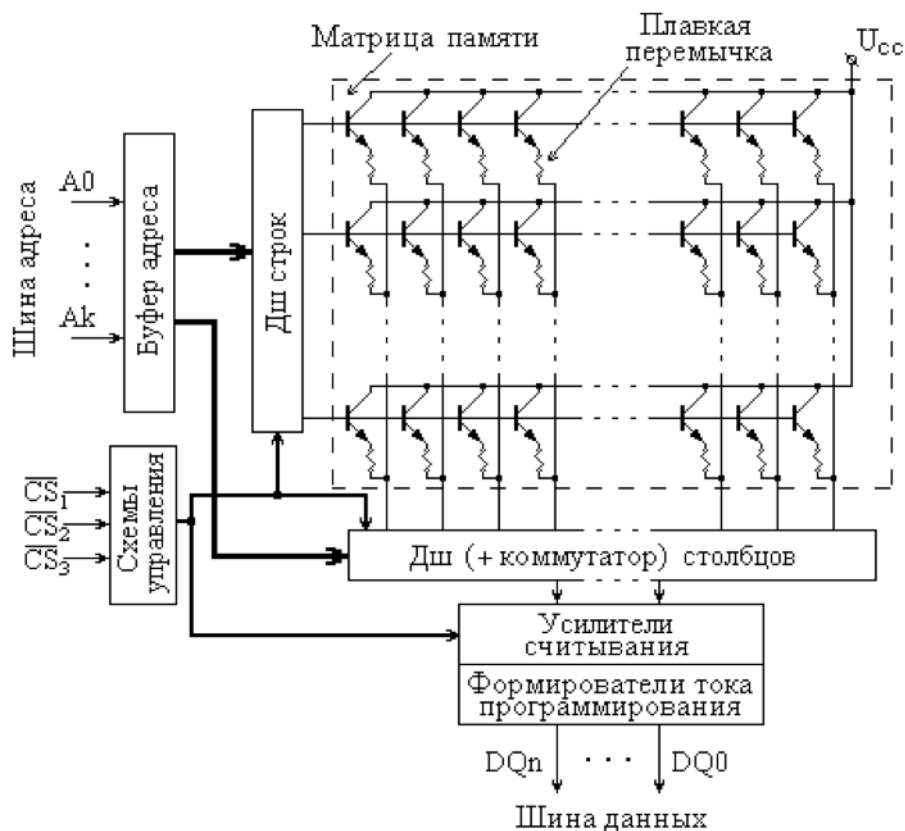


Рис. 10.6. Вариант структурной схемы ПЗУ с однократным программированием

Различают две большие группы ПЗУ: программируемые изготовителем и программируемые пользователем.

ЗУ первой группы, называемые иначе масочными, обычно выпускаются большими партиями. Информация в них заносится в процессе изготовления этих ЗУ на заводах: с помощью специальной маски в конце технологического цикла на кристалле формируется соответствующая конфигурация соединений. Такие ЗУ оказываются наиболее дешевыми при массовом изготовлении. Их обычно используют для хранения различных постоянных программ и подпрограмм, кодов, физических констант, постоянных коэффициентов и пр.

В ПЗУ, программируемые пользователем, информация записывается после их изготовления самими пользователями. При этом существуют два основных типа таких ЗУ: однократно программируемые и перепрограммируемые.

Нетрудно вспомнить, что аналогичные разновидности имеются и у CDROM, которые, по существу, являются ПЗУ (ROM), изготавливаемыми на основе другого физического принципа.

Наиболее простыми являются однократно программируемые ПЗУ. В этих ЗУ запись как раз и производится посредством разрушения соединительных перемычек между выводами транзисторов и шинами матрицы (хотя есть и несколько иные

технологии). Изображение программируемого ПЗУ на функциональной схеме показано на рис. 10.7.

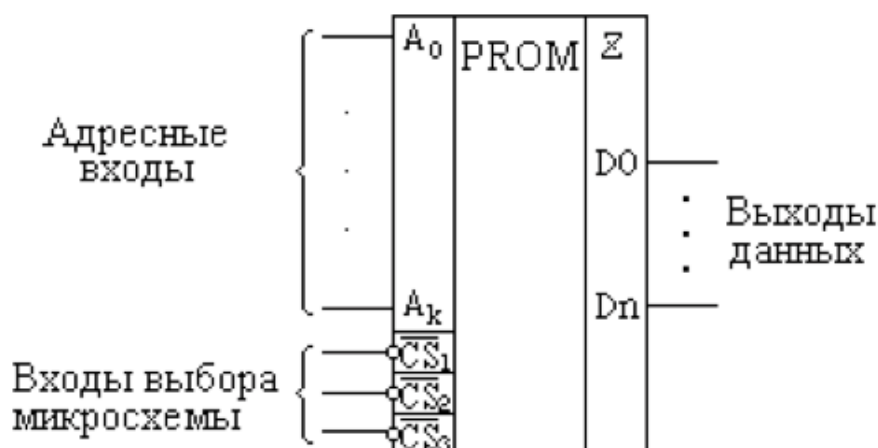


Рис. 10.7. Условное изображение программируемого ПЗУ на функциональных схемах

Перепрограммируемые ПЗУ позволяют производить в них запись информации многократно. Конечно, в таких ЗУ должен использоваться иной принцип, чем разрушение перемычек в процессе записи. Распространенные технологические варианты используют МОП-транзисторы со сложным затвором (составным или “плавающим”), который способен накапливать заряд, снижающий пороговое напряжение отпираания транзистора, и сохранять этот заряд при выключенном питании. Программирование таких ПЗУ и состоит в создании зарядов на затворах тех транзисторов, где должны быть записаны данные (обычно “0”, так как в исходном состоянии в таких микросхемах записаны все “1”).

Перед повторной записью требуется произвести стирание ранее записанной информации. Оно производится либо электрически, подачей напряжения обратной полярности, либо с помощью ультрафиолетового света. У микросхем последнего типа имелось круглое окошечко из кварцевого стекла, через которое и освещался кристалл при стирании.

Параметры постоянных ЗУ соответствуют технологическим нормам своего времени. В начале 2000-х годов типовые емкости микросхем постоянной памяти с масочным программированием составляли порядка 32-128 Мбит, а времена обращения превышали аналогичные показатели оперативной памяти и для различных модификаций достигали до 100 нс.

10.6. Регистровая память, кэш-память

Сверхоперативные ЗУ (исторически первое название регистровой памяти) используются для хранения небольших объемов информации и имеют значительно меньшее время (в 2 - 10 раз) считывания/записи, чем основная память. СОЗУ обычно строятся на регистрах и регистровых структурах.

Регистр представляет собой электронное устройство, способное хранить занесенное в него число неограниченно долго (при **включенном питании**). Наибольшее распространение получили регистры на статических триггерах.

По назначению регистры делятся на *регистры хранения* и регистры сдвига. Информация в регистры может заноситься и считываться либо параллельно, *сразу* всеми разрядами, либо последовательно, через один из крайних разрядов с последующим сдвигом занесенной информации.

Сдвиг записанной в регистр информации может производиться вправо или влево. Если регистр допускает сдвиг информации в любом направлении, он называется **реверсивным**.

Регистры могут быть объединены в единую структуру. Возможности такой структуры определяются способом доступа и адресации регистров.

Если к любому регистру можно обратиться для записи/чтения по его адресу, такая регистровая структура образует **СОЗУ с произвольным доступом**.

Безадресные регистровые структуры могут образовывать два вида устройств памяти: магазинного типа и память с выборкой по содержанию (**ассоциативные ЗУ**).

Память магазинного типа образуется из последовательно соединенных регистров (см. рис. 10.8).

Если запись в регистровую структуру (рис. 10.8,а) производится через один регистр, а считывание — через другой, то такая память является аналогом магазинной памяти и работает по принципу “первым вошел - первым вышел” (FIFO — first input, first output).

Если же запись и чтение осуществляются через один и тот же регистр (рис. 10.8,б), такое устройство называется *стековой памятью*, работающей по принципу “первым вошел — последним вышел” (FILO — first input, last output). При записи числа в стековую память сначала содержимое стека сдвигается в сторону последнего, К-го регистра (если стек был полностью заполнен, то число из К-го регистра теряется), а затем число заносится в вершину стека — регистр 1. Чтение осуществляется тоже через вершину стека, после того как число из вершины прочитано, стек сдвигается в сторону регистра 1.

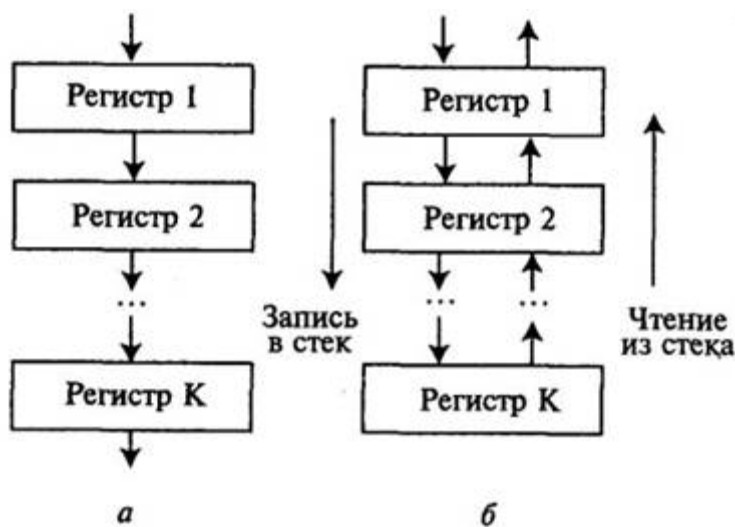


Рис. 10.8. Регистровая структура магазинного типа: а - типа FIFO; б - типа FILO

Стековая память получила широкое распространение. Для ее реализации в ЭВМ разработаны специальные микросхемы. Но часто работа стековой памяти

эмулируется в основной памяти ЭВМ: с помощью программ операционной системы выделяется часть памяти под стек (в IBM PC для этой цели выделяется 64 Кбайта). Специальный регистр микропроцессора (указатель стека) постоянно хранит адрес ячейки ОП, выполняющей функции вершины стека. Чтение числа всегда производится из вершины стека, после чего указатель стека изменяется и указывает на очередную ячейку стековой памяти (т.е. фактически стек остается неподвижным, а перемещается вершина стека). При записи числа в стек сначала номер ячейки в указателе стека модифицируется так, чтобы он указывал на очередную свободную ячейку, после чего производится запись числа по этому адресу. Такая работа указателя стека позволяет реализовать принцип “первым вошел - последним вышел”. В стек может быть загружен в определенной последовательности ряд данных, которые впоследствии считываются из стека уже в обратном порядке, на этом свойстве построена система арифметических преобразований информации, известная под названием “логика Лукашевича”.

Память с выборкой по содержанию является безадресной. Обращение к ней осуществляется по *специальной маске*, которая содержит поисковый образ. Информация считывается из памяти, если часть ее соответствует поисковому образу, зафиксированному в маске. Например, если в такую память записана информация, содержащая данные о месте жительства (включая город), и необходимо найти сведения о жителях определенного города, то название этого города помещается в маску и дается команда **чтение** - из памяти выбираются все записи, относящиеся к заданному городу.

В микропроцессорах ассоциативные ЗУ используются в составе кэш-памяти для хранения адресной части команд и операндов исполняемой программы. При этом нет необходимости обращаться к ОП за следующей командой или требуемым операндом: достаточно поместить в маску необходимый адрес, если искомая информация имеется в СОЗУ, то она будет сразу выдана. Обращение к ОП будет необходимо лишь при отсутствии требуемой информации в СОЗУ. За счет такого использования СОЗУ сокращается число обращений к ОП, а это позволяет экономить время, так как обращение к СОЗУ требует в 2 - 10 раз меньше времени, чем обращение к ОП.

Уровни кеша

Кеш-память уровня $N+1$ всегда больше по размеру и медленнее по скорости обращения, чем кеш-память уровня N .

Самой быстрой памятью является кеш-память первого уровня **L1-cache: L1I, L1D - неотъемлемая часть процессора**, расположена на том же кристалле и входит в состав функциональных блоков. Без нее процессор не сможет функционировать. Память **L1** работает на частоте процессора и в общем случае обращение к ней может производиться каждый такт (зачастую возможно выполнять даже несколько чтений/записей одновременно), латентность доступа обычно равна **2-4 такта ядра**, объем этой памяти обычно невелик — **не более 64Кб**.

Второй по быстродействию является **L2**. В отличие от L1 ее можно отключить с сохранением работоспособности процессора. Кеш второго уровня обычно расположена либо на кристалле, как и **L1**, либо в непосредственной близости от ядра, например, в процессорном картридже (только в слотовых процессорах), в старых процессорах ее располагали на системной плате. Объем **L2** больше — **от**

128Кб до 1—4Мб. Обычно латентность **L2**, расположенной на кристалле ядра, составляет от **8 до 20 тактов ядра**.

Кеш третьего уровня **L3** наименее быстродействующий и обычно расположен отдельно от ядра ЦП, но он может быть очень внушительного размера и всё равно значительно быстрее, чем оперативная память.

Сегодня существует и **L4** уровень кэша, обыкновенно он расположен в отдельной микросхеме. Применение кэша 4 уровня оправдано только для высокопроизводительных серверов и мейнфреймов.

Объемы кеш-памяти: **L1** до 128 Кб,

L2 от 128 Кб до 1-12 Мб,

L3 свыше 24 Мб.

Проблема синхронизации между различными уровнями кэш-памяти (как одного, так и множества процессоров) решается когерентностью кэша.

Существует три варианта обмена информацией между кэш-памятью различных уровней, или, как говорят, кэш-архитектуры:

- **инклюзивная** – предполагает дублирование информации кэша верхнего уровня в нижнем (предпочитает фирма **Intel**);
- **эксклюзивная** - предполагает уникальность информации, находящейся в различных уровнях кэша (предпочитает фирма **AMD**);
- **неэксклюзивная** - уровни кэша связываются между собой произвольным образом.

См. дополнительно информацию о кэш-памяти в файле [Кэш подробно](#).

На рис. 10.9 представлена схема расположения регистровой памяти, кеш-памяти и основной памяти, в которой обозначены место, объем и время доступа к каждой.

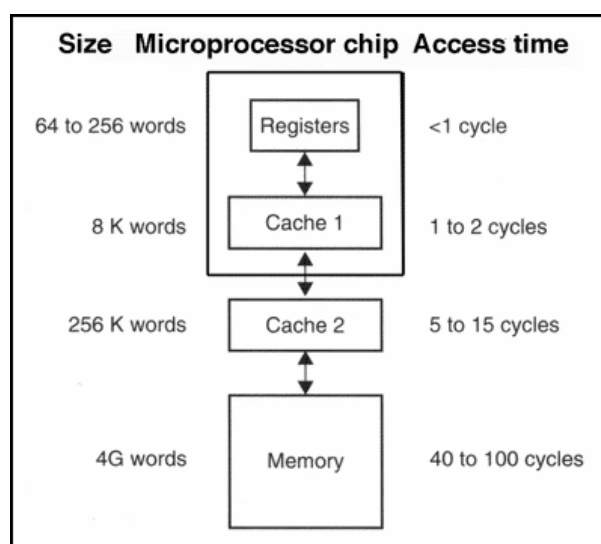
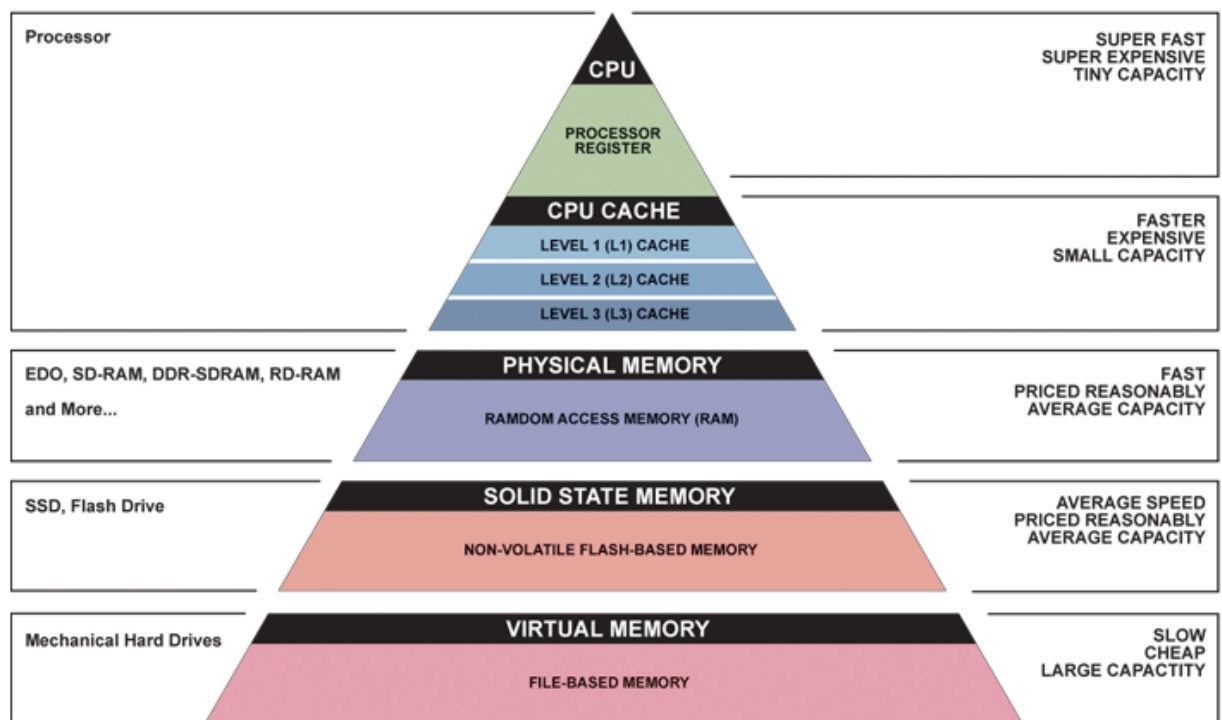


Рис. 10.9. Схема уровней памяти

На рис. 10.10 приведена полная иерархическая структура памяти компьютера с указанием места расположения, типа памяти, ее разновидностей и таких общих характеристик как скорости, стоимости и объема.



▲ Simplified Computer Memory Hierarchy
 Illustration: Ryan J. Leng

Рис. 10.10. Иерархическая структура компьютерной памяти

P.S. См. презентации по теме из папки «Дополнительные материалы»