

Лекция 5: оперативная память.

Память является одним из основных элементов любой вычислительной системы. Ячейки памяти присутствуют в каждом конструктивном модуле компьютера. Основной памятью с которой происходит постоянная работа компьютера является оперативная память. Далее пойдёт рассмотрение именно оперативной памяти. Другие типы памяти (диски, съёмные носители информации и прочее) рассматриваться будут в последующих лекциях. Конструктивно оперативная память является полупроводниковой, которая физически подразделяется на три класса:

- энергонезависимая
- статическая
- динамическая

Так же память подразделяется по типу работы

- синхронная
- асинхронная

Оперативная память временная, то есть данные в ней хранятся только до выключения компьютера. Конструктивно память выполняется в виде модулей, таким образом её легко заменить и — чаще всего — увеличить объем.

Для исполнения любой программы на компьютере её необходимо загрузить в оперативную память. Только после того, как программа загружена в оперативную память она доступна для исполнения. К данным, находящимся в оперативной памяти процессор имеет непосредственный (через контроллер памяти) доступ, а к внешней памяти или памяти периферийных устройств через буфер или порты ввода вывода: области оперативной памяти недоступны для загрузки программ и выделенные специально для общения с устройствами.

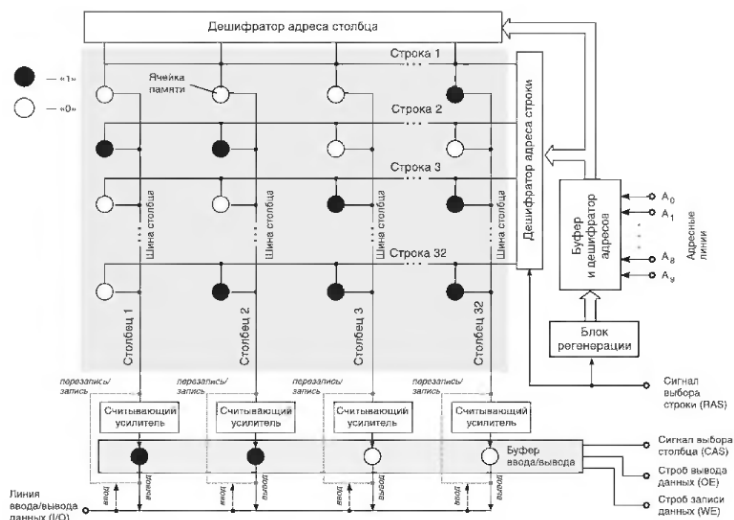
Время доступа к данным в памяти мало — в среднем около 200 нс. Для сравнения, доступ к данным на жёстком диске составляет 12000000 нс, а к флеш-памяти 2000000 нс. Недостаток оперативной памяти в том что содержимое памяти

при отключении питания стирается. Оперативная память относится к категории динамической: её содержимое требует постоянного обновления.

Запоминающим элементом динамической памяти является полупроводниковый конденсатор, который может быть либо заряжен (логическая 1), либо разряжен (логический 0). В идеальном случае, заряд в конденсаторе хранится неограниченное время. В реальности у конденсатора есть ток утечки, то есть с течением времени он теряет заряд и информация, записанная в динамической памяти теряется со временем. Единственным способом восстановить нужный уровень заряда конденсатора является выполнение операции чтения/записи данных. Восстановление нужного уровня заряда в ячейках памяти называется регенерацией оперативной памяти. Регенерация происходит при каждой операции чтения/записи данных, однако нет гарантии, что промежуток между операциями не будет достаточным для того, чтобы содержимое памяти было потеряно. Для этих случаев имеется специальная схема, которая через равные промежутки времени (обычно 2 мс) осуществляет доступ для чтения ко всем строкам памяти. Для работы схемы требуется всего 1 такт шины, но на этот такт процессор переводится в состояние ожидания, чтобы случайно не вмешаться в процесс регенерации.

Принцип работы.

Ячейки организованы в матрицу из строк и столбцов. Когда процессор или контроллер DMA обращается к памяти для чтения данных, на вход микросхемы подаётся сигнал вывода данных OE (output enabled). Затем подаётся номер адреса строки и сигнал RAS (row address strobe). Это даёт команду соединения шин столбцов с ячейками выбранной строки. Для чтения используются усилители на каждой адресной линии, так как заряд ячейки памяти очень мал и разрушается при чтении. Информация со всей строки считывается в буфер ввода/вывода микросхемы.



С незначительной задержкой от сигнала RAS подаётся сигнал CAS (Column Address Strobe) и адрес столбца. Данные выбираются из буфера ввода/вывода и поступают на выход памяти.

При считывании данных из ячеек буфера обмена данных происходит разрушение этой информации, поэтому необходима перезапись информации: выходы буфера снова соединяются с шинами столбцов памяти, чтобы перезаписать считанную информацию. Естественно, происходит зарядка только тех ячеек, где хранится 1. Ячейки содержащие 0 не обновляются. При записи подаётся сигнал WE (write enable) и информация поступает на соответствующую шину столбца не из буфера, а с входа модуля памяти, в соответствии с переданным адресом.

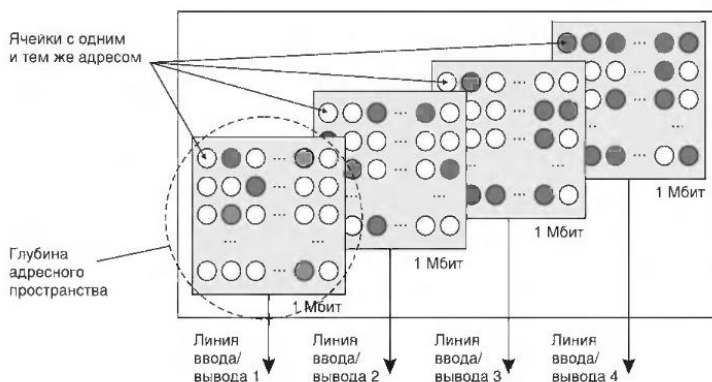
Характеристики микросхем памяти.

Основными характеристиками элементов являются:

- тип
- ёмкость
- разрядность
- быстродействие
- временная диаграмма

Ёмкость и разрядность.

На предыдущем рисунке была представлена схема модуля памяти с одной линией ввода/вывода. Одновременно с такого модуля можно считать и записать только 1 бит информации. Для повышения скорости обмена данными разработаны микросхемы с 4, 8, 16 линиями ввода/вывода. Эти микросхемы имеют соответственно 4, 8 или 16 одинаковых матриц ячеек памяти. Таким образом, при поступлении на входы микросхемы адреса ячейки памяти происходит одновременное чтение (запись) ячеек по данному адресу, находящихся во всех матрицах сразу. Например, если микросхема имеет 8 линий ввода/вывода, то информация читается и записывается в неё по байтово.



Количество линий ввода/вывода определяет **разрядность шины**.

Количество бит информации, которая хранится в ячейках каждой матрицы, называется **глубиной адресного пространства** микросхемы памяти (depth address).

Общая ёмкость микросхемы памяти определяется произведением глубины адресного пространства на количество линий ввода/вывода. Например, ёмкость микросхемы памяти с глубиной адресного пространства в 1 Мб и 4-мя линиями ввода/вывода равна 4 Мб. Обозначается чаще всего 1x4 или 1Mx4.

Быстродействие.

Производительность микросхемы определяется временем между выполнением двух операций чтения или записи. Для выполнения одной операции требуется

1. выбор строки
2. выбор столбца
3. операция чтения или записи.

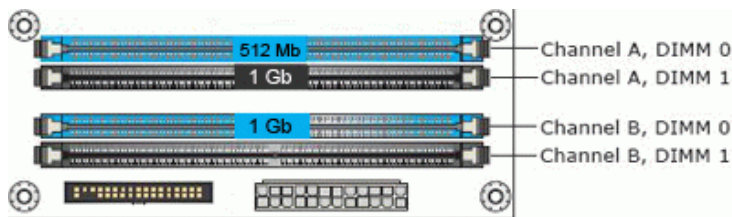
Последовательность этих операций называют рабочим циклом. Время для выполнения чтения или записи по произвольному адресу называется временем доступа (access time). Чаще всего обозначается скоростная характеристика памяти через частоту работы таймера памяти. Запишем характеристики основных типов памяти:

Года выпуска	Технология	Частоты
1987	FPM	Асинхронная, эквивалент 20 МГц
1995	EDO	Асинхронная, эквивалент 20 МГц
1997	PC66 SDRAM	66 МГц
1998	PC100 SDRAM	100 МГц
1999-2000	PC133 SDRAM	133 МГц
1999	RDRAM(Rambus)	800 МГц
2000-2002	DDR1 SDRAM	266-434 МГц
2003-2006	DDR2 SDRAM	500-800 МГц
2007	DDR3 SDRAM	1066-1600 МГц
2012-2013	DDR4 SDRAM	2133-4266 МГц
2020	DDR5 SDRAM	3200-8400МГц

Быстродействие памяти всегда отставало от процессора.

Основным способом повышения быстродействия памяти, на сегодняшний момент, является использование **банков памяти**:

набора микросхем памяти с общей разрядностью равной разрядности шины памяти. То есть, если разрядность шины памяти равна 64 битам и банк собран из двух модулей (двух наборов микросхем), то разрядность каждого модуля должна равняться 32 битам.



В новых процессорах используют двух, трёх, иногда четырёх канальную память. На 2017 год в четырёх канальном режиме работали core i7, начиная с 3-его поколения, core i9, xeon e5 и e7, боги 8 ого поколения, AMD Opteron 6100-6300 и AMD Ryzen (ядро threadripper) В этом случае производительность поднимается за счёт увеличения разрядности шины данных. Если контроллер памяти встроен в процессор, то у каждого канала памяти есть своя прямая шина до процессора. В более старых компьютерах этим управляет чипсет. Например выдает данные в системную шину в два раза быстрее, чем частота шины памяти.

Другим способом повышения быстродействия является **противофаза** синхронизации банков памяти. Если используется, к примеру, два банка памяти, то когда из одного банка памяти идёт чтение, во втором идёт выборка адреса. То есть, если позволяет частота системной шины позволяет (является в два или более раз выше, чем частота шины памяти), то происходит двукратный рост производительности.

Временная диаграмма.

Для чтения или записи в память компьютеру необходимо выполнить ряд операций над памятью:

1. чипсет или процессор выдает адрес нужной ячейки памяти
2. происходит ожидание выборки данных и копирования их

в буфер обмена

Операции требуют времени и так как быстродействие модулей памяти ниже, чем у процессора, то для операций требуются задержки или паузы. Для каждой микросхемы указываются значения задержек в виде цифрового кода (например, 3-2-3 или 2-3-2-6-1), который и называется **таймингом** модуля или **временной диаграммой**.

Цифры обозначают количество тактов шины на выполнение различных этапов выборки данных. Например код 2-3-2-6-1 расшифровывается так:

- 2 – CAS Latency (CL) – задержка между выбором и чтением ряда
- 3 – RAS to CAS (TRDT) – выбор ряда
- 2 – RAS Precharge Delay (tRCP / tRP) – деактивация ряда
- 6 – Row Active Delay (tRA/tRD/tRAS) – число циклов чтения
- 1 – Command Rate (CMD Rate) – задержка адресации

Часто указывают только CL и называют это **латентностью** памяти.

При разгоне памяти повышают не только тактовую частоту шины памяти, но и подбирают новые тайминги. Разгон всегда проводят на повышенном напряжении и часто физически портят микросхемы перегревом. Рекомендуется не превышать указанные производителем модуля частоты и тайминги.

Тип памяти.

Вся современная оперативная память относится к типу DRAM – Dynamic RAM – динамическая память. Это значит, что для сохранности данных в такой памяти необходима регулярная регенерация хранимых данных. В процессе совершенствования оперативной памяти было создано несколько типов памяти

- FPM
- EDO
- SDRAM
- DDR SDRAM

Все они принципиально устроены одинаково: квадратная матрица из элементов памяти, адресуемых при помощи RAS и

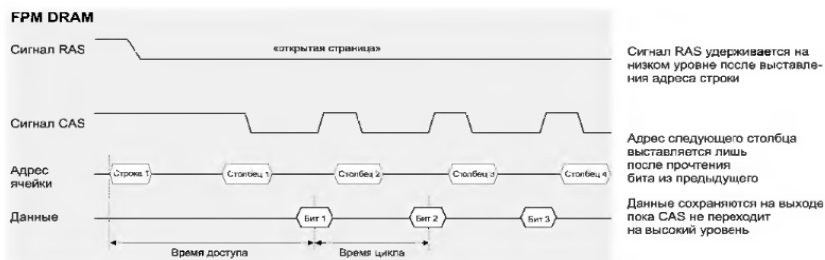
CAS (адреса строки и столбца соответственно). Элементарная ячейка памяти состоит из миниатюрного конденсатора и управляющих транзисторов.



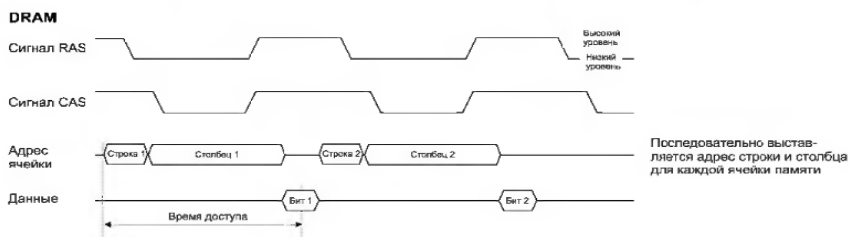
Логическая единица определяется наличием на конденсаторе заряда. Отсутствие заряда означает логический ноль. При активации линии адреса – фактически, при запросе на чтение из конкретной ячейки – на затвор транзистора попадает напряжение и при наличии заряда на конденсаторе на линию данных подается заряд. Таким образом, считывается ноль при отсутствии заряда и единица при его наличии. Из схемы видно, почему чтение данных разрушает их – заряд конденсатора тратится на активацию линии данных. Поэтому и требуется повторный заряд конденсатора при каждом обращении к ячейке. Разумеется, это требуется только если ячейка хранит единицу, но так как контроллер изначально этого не знает, то регенерация происходит в любом случае.

FPM.

Сигнал CAS используется и для указания времени в течении которого будет читаться ячейка данных. Приведём временную диаграмму работы FPM памяти.



FPM DRAM широко использовалась в 486 процессорах. До этого использовалась обычная DRAM имевшая следующую схему работы.



Но несмотря на 40% преимущество над «классической» памятью пределом скорости работы FPM стало 28 МГц.

EDO.

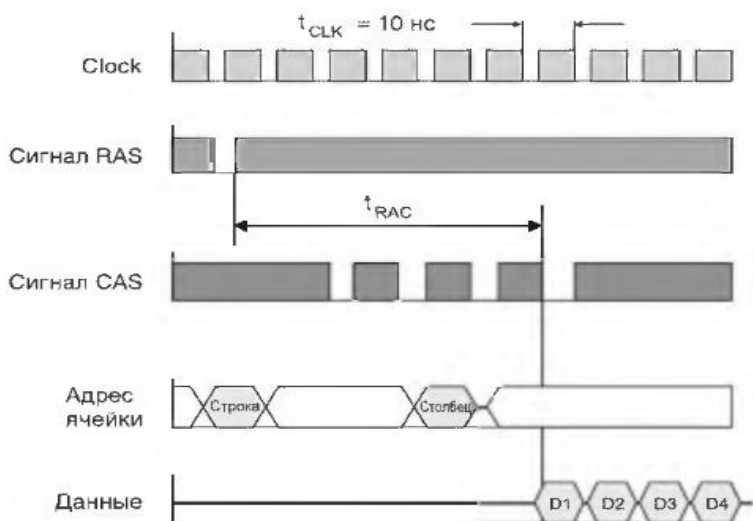
Использовалась в первых Pentium. В отличие от FPM, линии ввода/вывода не отключаются на время установки нового адреса и соответственно позволяют одновременно читать данные и задавать адрес следующей ячейки. Частоты поднялись до 50 МГц, однако при записи данных производительность падала до уровня FPM.

SDRAM.

До 1997г. производились только асинхронные модули памяти, то есть такие, работа которых не была синхронизирована с частотой системной шины. Асинхронные элементы срабатывают как только на адресных и информационных входах появляются сигналы. После их подачи через некоторое время на выходе появляется запрошенная информация. Это приводит к тому, что модули обладают низкой

помехоустойчивостью и часто служат причиной сбоя в работе компьютера (связано с регенерацией содержимого памяти). Для синхронных элементов есть ещё дополнительный вход – сигнал системной шины. Он синхронизирует активность модуля памяти, в том числе и «появление» информации на выходе модуля.

SDRAM и есть Synchronous DRAM – синхронная динамическая память.



Отличие от работы асинхронной памяти заключается в том, что все операции в микросхеме синхронизированы и работают без циклов ожидания. За счёт этого увеличились частоты микросхем памяти, вплоть до 133 МГц. Вместе с конвейерной архитектурой исполнения команд это позволило за один такт системной шины получать до 8 байт. То есть, при частоте системной шины в 100 МГц скорость передачи данных выросла до 800 Мб/с.

Структурно SDRAM предполагает несколько банков памяти внутри чипа.

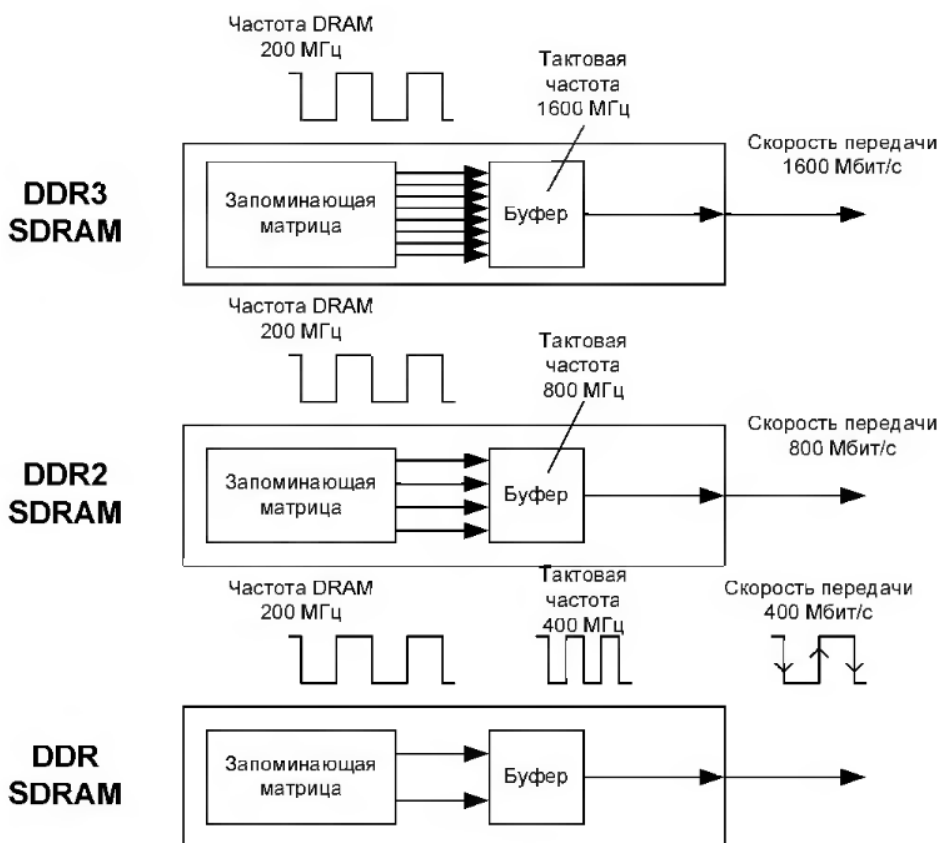
DDR.

Double Data Rate SDRAM – SDRAM с удвоенной частотой данных – позволила передавать данные по обоим фронтам

сигнала – восходящему и нисходящему. Таким образом, объём передаваемой информации возрос в два раза.

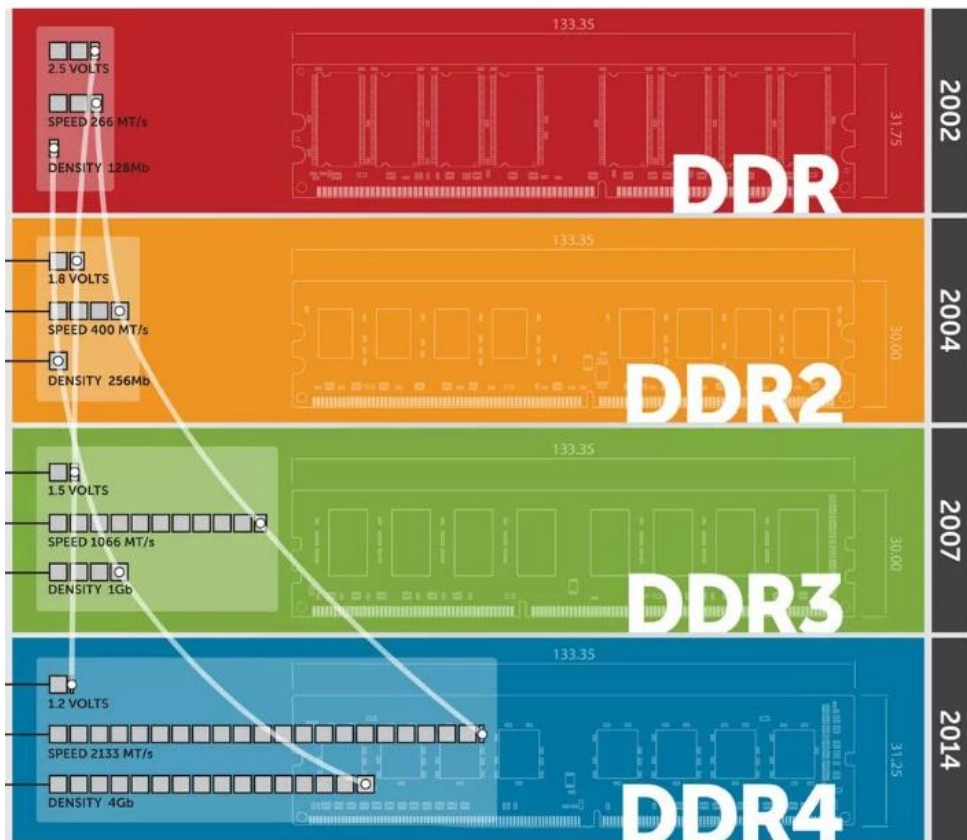
Такой способ передачи данных очень чувствителен к помехам, поэтому для более точной синхронизации применяется дополнительный сигнал DQS. Для точной синхронизации сигналов DQS и DQ (передачи данных) соответствующие проводники имеют одинаковую длину и ёмкость.

Технологии DDR2, DDR3 и DDR4 развили эту схему и позволяют передать по 4, 8 и 16 сигналов за один такт памяти соответственно. Пример для частоты в 200 МГц приведён ниже



Следует отметить, что прочее увеличение производительности происходит в основном за счёт оптимизации поиска ячейки, а

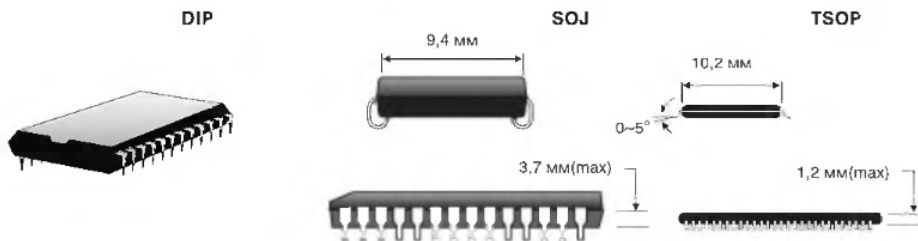
вот производительность запоминающей матрицы меняется слабо, в основном за счёт смены техпроцесса. Однако стоит всё таки отметить прогресс в производительности памяти DDR. Приводим сравнительную графику из статьи на ferra.ru за 2019 год



Видно, что скорость работы памяти за 12 лет выросла в 8 раз.

Форм-факторы модулей памяти.

Первые микросхемы памяти выпускались в DIP-корпусах и размещались непосредственно на материнской плате. Позже стали производить платы с колодками под **DIP-модули**, что позволило наращивать объем памяти самостоятельно.



Память в корпусе SOJ отличается загнутыми ножками, что увеличивает их механическую прочность. TSOP корпуса имеют ножки параллельные корпусу и предназначены для пайки в заводских условиях.

На каждом корпусе имеется маркировка в которой закодированы

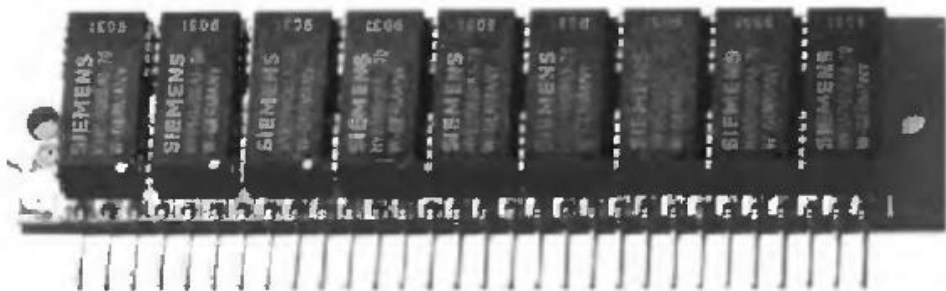
- Фирма производитель чипа (первые 2-3 буквенных символа)
- Ёмкость и прочие характеристики
- Частота (опционально через тире)

К примеру – HYB 514100BJ-60 – Siemens на частоте до 60 МГц,

Модули памяти.

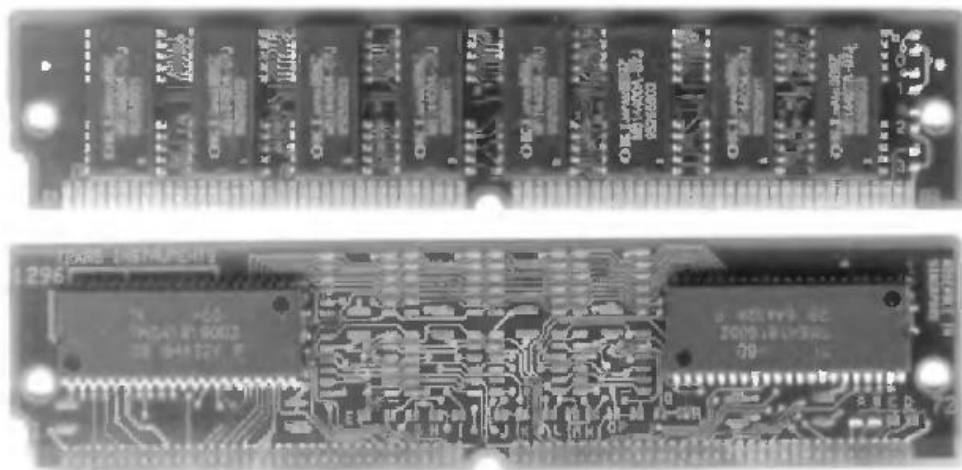
С ростом требований к объёму оперативной памяти корпуса памяти стали занимать слишком много места на материнской плате и потребовалось разработать более компактные способы размещения памяти. Была разработана концепция **модуля памяти**: монтажная плата с напаянными с одной или двух сторон корпусами памяти и вставляющаяся в специальный разъем на материнской плате – **слот** памяти. Концепция осталась и по сей день.

Исторически первыми были SIP модули.



Плата имела 30 игольчатых выводов и применялась в компьютерах на основе 386 процессора.

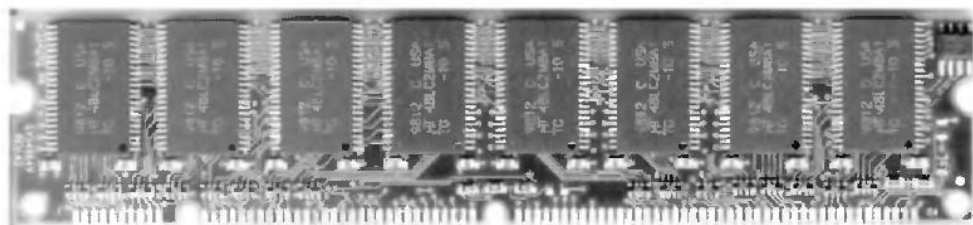
SIMM модули.



Ненадежные игольчатые контакты заменены на площадки с обеих сторон модуля, контакты с противоположных сторон модуля соединены. Выпускались с 30 и 72 контактные модули и использовались в компьютерах на основе 386, 486 и первых Pentium.

На модуле располагалось 8, 16 или 32 корпуса микросхем памяти. На серверных модулях с контролем чётности памяти на каждые 8 корпусов добавляется 1 (очень редко 2) дополнительный корпус для хранения бита чётности.

DIMM модули. Основное отличие от SIMM – независимость контактов с разных сторон платы.



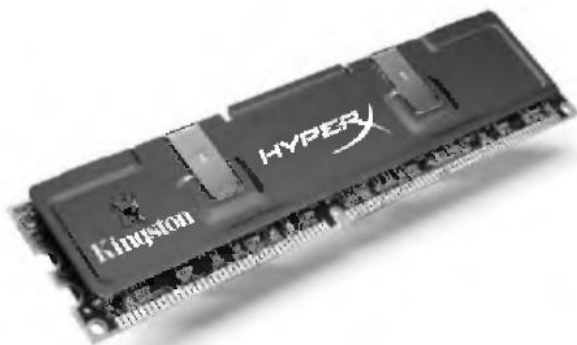
Первые модули, на основе SDRAM памяти имеют по 168 контактов (84 с каждой стороны) и разрядность 64 бита. Для ноутбуков выпускались модули SO-DIMM (Small Outline DIMM) с 144 и 72 контактами.

Впервые на модули добавляется новый компонент – SPD (serial presense detect) : плата энергонезависимой памяти, хранящая конфигурацию модуля и используемая для первоначальной инициализации модуля. Без этой микросхемы при старте компьютера происходило полное обращение ко всем чипам памяти, что занимало значительное время. Сейчас тоже можно отключить использование SPD и увидеть насколько увеличится время первоначальной диагностики системы.

Высокопроизводительные модули так же могут содержать кэш-память.

Для корректной установки модуля в плате есть два пропила – между 10 и 11 контактом (используется для определения типа памяти – EDO или SDRAM) и между 40 и 41 (используется для определения напряжения модуля 5 или 3,3 вольта).

DDR DIMM.



184-контактный модуль DDR400 SDRAM
с радиатором охлаждения микросхем производства компании Kingston



240-контактный модуль DDR2 SDRAM
производства компании Kingston

Модули появились в 1997 году и имели 184 контакта. Один пропил по центру модуля в основном не даёт поставить модуль в слот задом наперёд, что было бы фатально для модуля.

DDR2 получил уменьшенные размеры контактов и – соответственно - большее их количество : 240. **DDR3** имеет столько же контактов, но электрически несовместим с DDR2, поэтому пропилов на модуле сдвинули. **DDR4** модуль имеет 288 контактов.

Для встроенных систем и ноутбуков продолжают выпускаться **SO-DIMM** модули, которые к версии DDR4 получили уже 260 контактов.

Все модули памяти маркируются специальным образом. К примеру:

- KHX 11000 D3 LL K2 / 1G
- KHX – производитель, в данном случае Kingston HyperX

- 11000 – скорость (1375 МГц)
- D3 – тип памяти (DDR3)
- LL – low latency (низкая задержка памяти)
- K2 – набор из двух модулей
- 1G - ёмкость

Маркировка стандартизирована и производители стараются её придерживаться. Однако 4 и 5 группа символов может отличаться. Например: K VX 1066 D3 N 7 / 1G , где N означает без контроля чётности (E – с контролем, S – SoDIMM модуль), а 7 – задержку CAS.

Логическая организация памяти: плоская модель.

От аппаратного обеспечения зависит не только физические характеристики и организация памяти, но и её логическое устройство. Современные операционные системы переключают процессор в работу в защищенном режиме и используют так **называемую плоскую модель памяти** – это метод организации оперативной памяти, где все ячейки памяти пронумерованы от 0 до максимального возможного адреса. Код и данные используют одно и тоже адресное пространство. Для разной разрядности адресной шины возможны разные максимальные значения адресного пространства:

- 16бит – 64 кбайта
- 32бит – 4 Гб
- 64бит – 16 эксаБайт (фактически 256 Тб)

Преимуществами плоской модели является то, модель является максимально простой и в реализации и в поддержке.

Понятие плоской модели связано с понятием **виртуальной памяти**, которая подробно рассматривается в курсе ОС.

Объем оперативной памяти определяется размером установленных в компьютере плат оперативной памяти.

Программы и их данные располагаются в оперативной памяти и по мере необходимости запрашиваются из неё. В разных компьютерах имеется разное количество памяти, что создает следующие трудности для программиста: на этапе создания программы придётся определить количество использованной оперативной памяти. Если задать количество памяти, равное

объему всего кода и данным приложения, то программа будет работать максимально эффективно, но запускаться будет только на компьютерах с нужным, или большим объемом оперативной памяти. Скорее всего, число таких компьютеров будет не велико.

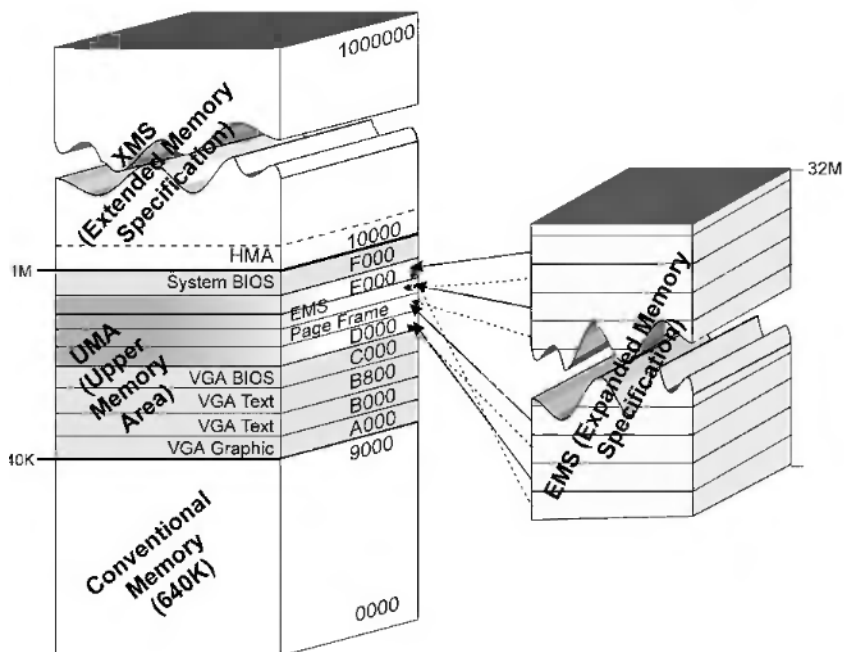
Если же затребовать небольшой объем памяти, то программа будет работать практически на любом компьютере, но медленно, так как постоянно будет использовать внешние носители информации для загрузки в память нужных блоков данных и кода.

Для ликвидации этих проблем разработано понятие виртуальной памяти. Считается, что программа изначально имеет максимально возможный для архитектуры объем памяти. Часть этой памяти располагается в реальной оперативной памяти; остальное хранится на внешнем носителе информации в особом файле – swar-файле. Подгрузкой затребованных областей памяти в реальную память и выгрузкой не нужных управляет операционная система, причём не зависимо от программы и не заметно для неё.

Например, в компьютере есть 1Гб памяти. Из неё 512 Мб занято самой операционной системой. Приложению остаётся всего 512 Мб. Предположим, что приложению необходимо 1Гб памяти. В этом случае приложение получит требуемый объем, но 512 Мб будут в оперативной памяти, а 512 Мб в swar-файле. При обращении к той части программы, что располагается в swar-файле, операционная система приостанавливает работу приложения и копирует нужный объем в реальную память. При этом из памяти стирается не требующаяся в данный момент часть приложения. Алгоритм выбора и механизм копирования зависят от конкретной операционной системы.

Логическая организация памяти: реальная модель.

Плоская модель используется современными операционными системами, но до загрузки системы, или при использовании старых, однозадачных операционных систем, используется другая модель памяти – реальная.



В данной модели можно выделить 5 основных логических областей:

1. Стандартная память (conventional memory)
2. EMS (Expanded Memory Specification)
3. UMA (Upper Memory Area)
4. HMA (High Memory Area)
5. XMS (Extended Memory Area)

Стандартная память.

Стандартная память это адреса памяти с 0000:0000 до 0009:FFFF. Размер в 640кб зарезервирован для использования приложением. В этом объёме выделены следующие адреса:

- Первый килобайт (с адреса 0000:0000) используется таблицей векторов прерываний. Таблица состоит из 256 элементов по 4 байта и содержит адреса загрузки обработчиков прерываний – небольших программ, предназначенных для обработки аппаратных событий.
- Следующие 768 байт используются для хранения данных системы BIOS – часов, таймеров, буфера клавиатуры и

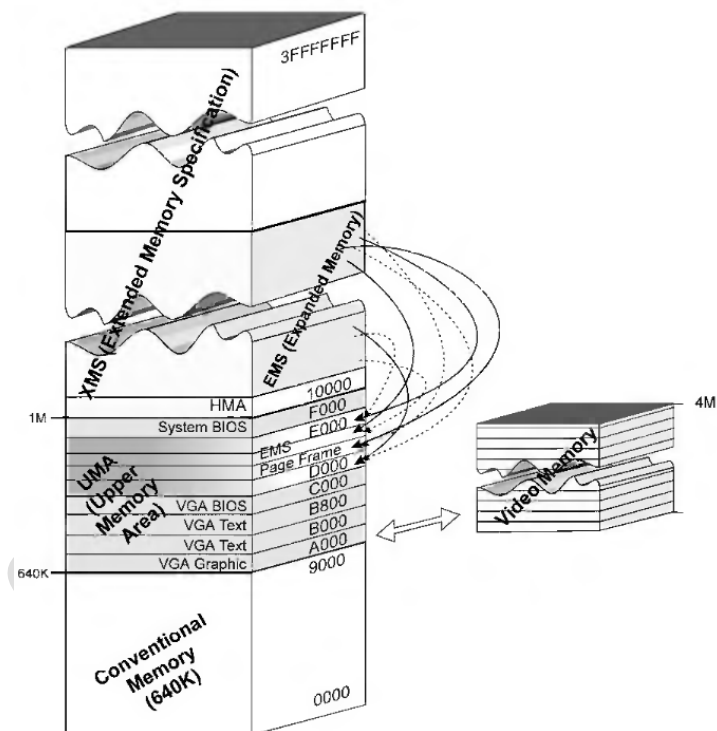
так далее.

- Далее располагается область переменной длины для хранения ядра системы (в DOS например это COMMAND.COM). В последних версиях имелась возможность поместить ядро в другие области памяти и освободить основную память для программы.

Остальная часть основной памяти принадлежит приложению.

UMA.

Часть памяти с адреса 000A:0000 до адреса 000F:FFFF называется UMA и используется для взаимодействия с аппаратурой и служебных функций.



Область UMA не однородна: в ней размещается видеопамять, ROM BIOS и прочее. Некоторые блоки зарезервированы для использования внешними устройствами.

С адреса 000A:0000 до 000B:FFFF располагается

видеопамять. Физически эти 128 Кб находятся на видеокарте, в оперативной памяти располагается их отображение. Фактически, при обработке адресов памяти контроллер анализирует затребованный адрес и, если обнаруживает, что адрес из специального диапазона (в данном случае – из видеопамяти), то передаёт запрос на контроллер соответствующего устройства. Физически, ячейки оперативной памяти в данном диапазоне адресов вообще никогда не используются.

На видеоадаптере может находиться (и находится) больше 128 Кб памяти. В этом случае их можно использовать следующим образом:

в специальные индексные регистры записывается адрес сегмента видеопамяти, размером 64Кб и эти сегменты отображаются на адреса 000A:0000 и 000B:0000. Таким образом, одновременно доступно только 128 Кб видеопамяти, но фактически, с помощью переключений значений регистров можно получить доступ ко всему объему видеопамяти.

С адреса 000C:0000 до адреса 000C:7999 располагается 32 Кб данных VGA BIOS. В этой области описываются инструкции по работе с видеоизображением. Размещены они в виде отображения на память для ускорения работы.

С адреса 000F:0000 до 000F:FFFF располагается 64 Кб область памяти ROM BIOS. Здесь находится информация о конфигурации компьютера и прочие инструкции BIOS.

Область с адреса 000C:8000 и до адреса 000E:FFFF отдана для использования производителями периферийных устройств для размещения отображения на буферы ввода-вывода устройств. Так же зарезервирован 64Кб для использования EMS памяти. Если ничего из этого не используется, то эту область можно использовать для размещения самой операционной системы (для DOS это делается при помощи менеджера EMM386.EXE).

EMS.

Выделенный сегмент UMA разбивается на 4 страницы по 16 Кб каждая. На эти страницы отображаются страницы памяти выше адреса 000F:FFFF. Это позволяет использовать дополнительную оперативную память для крупных приложений. На компьютерах с процессором 80286 требовался специальный

чип на материнской плате для реализации механизма EMS. Начиная с процессора 80386 реализация EMS стала программной (в DOS при помощи EMM386.EXE). В самой последней версии стандарта (4.0) максимальный объем адресуемой таким образом памяти равнялся 32 Мб. Для работы механизма EMS требуется поддержка и со стороны операционной системы, и со стороны приложения.

HMA.

Адреса с 0010:0000 до 0100:FFFF зарезервированы для HMA, которая используется для предотвращения несанкционированного доступа к памяти. Ошибка возникла из-за некорректной реализации механизма обращения к памяти в процессоре 80286. Процессор 8088 был 20разрядным. К стандартной памяти обращались как к набору 64 Кб сегментов, которые вообще то могли перекрываться и начинаться с любого адреса, кратного 16. Если определить начальным адресом сегмента адрес 000F:FFF0, то на процессоре 8088 произойдет переполнение адреса.

Но процессор 80286 уже имеет 32 разрядную шину адреса и переполнения не произойдёт. Выше 1 Мб должна располагаться область данных EMS. Таким образом, можно объявив сегмент стандартной памяти добраться до содержимого EMS памяти, что не допустимо. Потому сегмент памяти HMA и введён для предотвращения таких ситуаций.

XMS.

XMS это расширенная реализация EMS, включающая в себя всю память свыше адреса 0010:FFFF, включая и область EMS памяти. Реализована начиная с процессора 80386. Программная поддержка осуществляется специальным драйвером (в DOS это HIMEM.SYS).

Однако, доступ до XMS может получить только одна программа, совместное использование не допускается.

Кэш-память.

В идее кэширования лежит правило **80/20**: 80% машинного времени используется всего 20% программного кода. При этом, вероятность того, что следующие инструкции и следующие запрошенные данные располагаются в тех же страницах

памяти, что и использующиеся в настоящий момент. Для ускорения доступа к данным и коду используется выделенная память, работающая на больших частотах (например, кэш первого уровня работает на частоте процессора).

В кэш-памяти располагаются код и данные текущего приложения и по мере необходимости они подгружаются из основной оперативной памяти. При каждом обращении к памяти контроллер анализирует адрес в памяти и сверяет его с адресами, помеченными как содержащиеся в кэше. Если адрес не содержится в кэше, то нужная страница оперативной памяти подгружается в кэш поверх наиболее долго не используемой (стратегия стэка). Есть и другие способы выбора данных для удаления, но они не рассматриваются так же в рамках курса Операционные системы.

В современных процессорах применено два способа увеличения эффективности кэширования

1. Раздельный кэш для кода и данных
2. Несколько уровней кэширования: L1 кэш располагается внутри кристалла процессора и работает на его частоте, уровни L2 и L3 располагаются на подложке процессора или на материнской плате. L2 и L3 имеют больший объем, чем L1, но меньшую скорость.

Например, процессор intel Core i7-6700K имеет три уровня кэш-памяти.

1. L1 имеет объем 128 кб для инструкций и 128 Кб для данных
2. L2 имеет объем 1024 Кб
3. L3 имеет объем 8Мб

Кэш память есть и на периферийных устройствах. Используется эта память с такой же целью – ускорение доступа к данным.

Память долговременного хранения.

Модули **NVRAM** (non volatile RAM – не временная RAM) – энергонезависимы и используются для хранения кода BIOS и данных о конфигурации устройств. Существует несколько типов NVRAM, которые различаются по способу записи данных и применяются в разных областях.

ROM – Read Only Memory.

Микросхемы ROM записываются на стадии производства и не подлежат перезаписи или стиранию.

PROM – Programmable ROM.

Микросхемы PROM записываются так же однократно, но не на стадии производства, а в специальном устройстве – программаторе. Как и ROM они практически не чувствительны к электромагнитным помехам.

EPROM – Erasable PROM.

Эти микросхемы могут быть перезаписаны в программаторе.

Стирание же осуществляется путём облучения чипа ультрафиолетом.

EEPROM – Electrically Erasable PROM.

Чипы EEPROM могут быть стёрты электронным образом. Для управления чипами необходима лишь программа, специального устройства не требуется.

В настоящее время EEPROM является основой для микросхем BIOS, располагающихся на материнской плате и содержащих код и данные для первоначальной инициализации компьютера и запуска процесса загрузки операционной системы.

Широко используемая в нынешнее время flash память является разновидностью EEPROM.