

Архитектура ЭВМ и систем

Лекция № 8

Память. Организация памяти.

План лекции

1. Иерархическая организация памяти
2. Классификация запоминающих устройств
3. Структура основной памяти
4. Память с последовательным доступом
5. Ассоциативная память
6. Организация флэш-памяти
7. Архитектурные способы повышения скорости обмена между процессором и памятью

Память. Организация памяти.

- Память – одна из наиболее важных подсистем ВМ.
- В первую очередь от памяти зависят функциональные возможности ВМ как средства обработки данных.
- Организация и характеристики памяти существенно влияют на общетехнические показатели ВМ: производительность, стоимость, надежность.

Иерархическая организация памяти

Совокупность устройств, обеспечивающих запись, хранение и чтение информации в вычислительной машине, образует **систему памяти**. Основными характеристиками памяти являются:

- 1) *информационная ёмкость* – количество битов или байтов информации, которое может храниться в запоминающем устройстве;
- 2) *время доступа к информации* – промежуток времени между началом обращения процессора за данными и моментом их появления на выходе памяти;
- 3) *стоимость хранения единицы информации* (бита) – отношение стоимости запоминающего устройства к реальному объёму пользовательских данных в битах (на практике часто применяются более крупные единицы информации: Кбайт, Мбайт).

Указанные параметры находятся в противоречии между собой. Поэтому системы памяти обычно имеют **иерархическую многоступенчатую организацию** с использованием различных устройств.

Иерархическая организация памяти

В иерархии памяти можно выделить следующие уровни:

- 1) *регистровая память* – набор регистров процессора, благодаря которым уменьшается число обращений к другим уровням памяти, реализованным вне процессора и требующим большого времени для операций обмена информацией;
- 2) *кэш – память* (сверхоперативная память) – служит для хранения копий информации, используемой в текущих операциях обмена; работа с кэш - памятью высокого быстродействия повышает производительность вычислительных машин;
- 3) *основная память* (оперативная, постоянная, полупостоянная) – работает в режиме обмена информацией с процессором и, по возможности, согласуется с ним по быстродействию; исполняемый в текущий момент фрагмент программы обязательно находится в основной памяти;
- 4) *специализированные виды памяти* – характерны для некоторых специфических архитектур (видеопамять, многопортовая память и т.д.);
- 5) *внешняя память* – позволяет хранить большие объёмы информации; реализуется обычно на основе устройств с подвижным носителем информации (магнитные и оптические диски, магнитные ленты и др.).

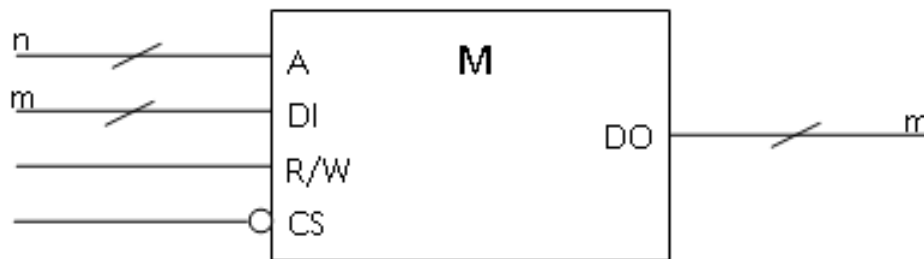
Иногда накопители на магнитных лентах и оптических дисках относят еще к одному уровню – **архивной памяти**.

Иерархическая организация памяти

Чем ниже уровень иерархии, тем выше быстродействие, но при этом также возрастает стоимость запоминающего элемента. На отдельных этапах решения какой-либо задачи в вычислительном процессе задействована относительно небольшая доля команд данных соответствующей программы. В процессе работы эту часть информации требуется обновлять. Это свойство и позволяет использовать в ВМ устройства памяти с различным быстродействием. Выбор информационного объёма устройств памяти и организации обмена между ними проводятся в целях оптимизации соотношения стоимости к быстродействию всей системы памяти.

Далее рассмотрим типичные сигналы простейшего запоминающего устройства и разновидности современных запоминающих устройств (ЗУ).

Классификация запоминающих устройств



Условное обозначение и типовые сигналы простейшего запоминающего устройства

Вначале подаётся адрес, чтобы последующие операции не коснулись какой-либо ячейки, кроме выбранной. Затем разрешается работа микросхемы сигналом *CS* и подаётся сигнал чтения/записи *R/W*. В зависимости от вида операции, на выходе *DO* формируются считываемые данные или на входе *DI* готовятся данные для записи.

M (*Memory*) – устройство памяти.

A – адрес, разрядность которого *n* определяется числом ячеек памяти, а также является номером ячейки, к которой идёт обращение.

CS (*Chip Select*) – сигнал, который разрешает или запрещает работу данной микросхемы.

R/W – сигнал, задающий выполняемую операцию (при единичном значении – чтение (*Read*), при нулевом – запись (*Write*)).

DI (*Data Input*) и *DO* (*Data Output*) – шины входных и выходных данных, разрядность которых *m* определяется разрядностью ячеек запоминающего устройства. В некоторых устройствах памяти эти линии объединены (обозначаются как *DIO*).

Классификация запоминающих устройств

- По способу доступа ЗУ делятся на ***адресные, последовательные и ассоциативные***.

При *адресном доступе* код на адресных входах указывает ячейку, с которой ведётся обмен информацией. В момент обращения все ячейки адресной памяти равнодоступны. Другие виды памяти часто строят на основе адресной памяти с соответствующими модификациями.

В устройствах памяти с *последовательным доступом* записываемые данные образуют некоторую очередь. Считывание происходит по очереди слово за словом либо в порядке записи (FIFO – First Input First Output), либо в обратном порядке (LIFO – Last Input First Output).

Ассоциативный доступ реализует поиск информации по некоторому признаку, а не по её расположению в памяти (адресу или месту в очереди). Основная область применения ассоциативное памяти в вычислительных машинах – кэширование данных.

Классификация запоминающих устройств

- В свою очередь адресные ЗУ по *организации записи* делятся на **оперативные** – ОЗУ (RAM – Random Access Memory) и **постоянные** – ПЗУ (ROM – Read Only Memory).

ОЗУ хранит данные, используемые при исполнении текущей программы, которые могут быть изменены в произвольный момент времени. Является энергозависимым устройством.

В ПЗУ содержимое либо не изменяется, либо изменяется редко и в специальном режиме.

Классификация запоминающих устройств

- По способу хранения информации ОЗУ делятся на **статические** и **динамические**.

В статических ОЗУ (SRAM – Static RAM) запоминающими элементами являются триггеры, сохраняющие своё состояние, пока схема находится под питанием и нет новой записи данных.

В динамических ОЗУ (DRAM – Dynamic RAM) данные хранятся в виде зарядов конденсаторов; при этом конденсаторы должны периодически регенерироваться.

Динамические ОЗУ имеют намного более высокую информационную ёмкость и в несколько раз дешевле статических ОЗУ, которые, в свою очередь, являются более быстродействующими.

В настоящее время именно динамические ОЗУ используются как основная память ВМ. Статические ОЗУ используются для построения кэш-памяти, буферной памяти и т.п.

Классификация запоминающих устройств

- Статические ОЗУ выполняются как **однопортовыми** (возможны одновременные обращения только к одной ячейке), так и **многопортовыми** (возможны одновременные обращения более чем к одной ячейке).
- Кроме того, по возможности синхронизации с процессором статические ОЗУ делятся на **асинхронные** и **синхронные**.

В *асинхронных ОЗУ* после произвольного по времени обращения к памяти до выдачи данных проходит определённое время, которое не синхронизировано с работой процессора. Вследствие этого могут возникать дополнительные задержки обмена данными между памятью и процессором.

В *синхронных ОЗУ* длительности этапов работы памяти жёстко связаны с синхросигналами системы, что позволяет исключить потери времени при обмене данными между памятью и процессором, а также организовать конвейерную обработку данных. Таким образом, синхронность памяти является средством повышения её быстродействия.

Классификация запоминающих устройств

- По виду режима перепрограммирования выделяют следующие виды ПЗУ:

Постоянная масочная память или масочное ПЗУ (ПЗУМ, ROM(M)), является однократно программируемой памятью; информация в неё записывается на промышленных предприятиях с помощью шаблона (маски). В дальнейшем содержимое ПЗУМ не изменяется.

Программируемая пользователем память, или программируемое ПЗУ (ППЗУ, PROM).

- ППЗУ делится на **программируемую однократно** и **программируемую многократно**.

В первом случае информация однократно перезаписывается потребителем в лабораторных условиях с помощью программаторов.

Во втором содержимое может быть изменено либо в лабораторных условиях, либо в специальных режимах.

Популярная в настоящее время FLASH-память относится к многократно программируемому ПЗУ, хотя и обладает рядом особенностей.

Классификация запоминающих устройств

- По способу организации очереди последовательные ЗУ делятся на следующие виды:

Буфер FIFO (Первый пришёл первый ушёл).

Буфер LIFO (Последний пришёл первый ушёл).

Файловые ЗУ. В них записываемые данные объединяются в специальные блоки (файлы). Чтение данных из файлового ЗУ осуществляется в прямом порядке и начинается после обнаружения приёмником символа начала блока.

Циклические ЗУ. Данные доступны одно за другим с постоянным периодом, определяемым ёмкостью памяти. К циклическим ЗУ относятся, например, видеопамять, буфер клавиатуры.

Классификация запоминающих устройств

Среди перспективных ЗУ следует отметить:

1. *ЗУ ферроэлектрического типа (FRAM – Ferroelectric RAM),* имеющие высокие ёмкость и быстродействие, а также обладающие свойствами энергонезависимости.
2. *Магниторезисторные ЗУ (MRAM – Magnetoresistive RAM),* обладающие свойствами естественной энергонезависимости, а также неразрушающего чтения.

Структура основной памяти

Основная память представляет собой вид памяти, к которой процессор может обращаться непосредственно (исключение составляет регистровая память самого процессора).

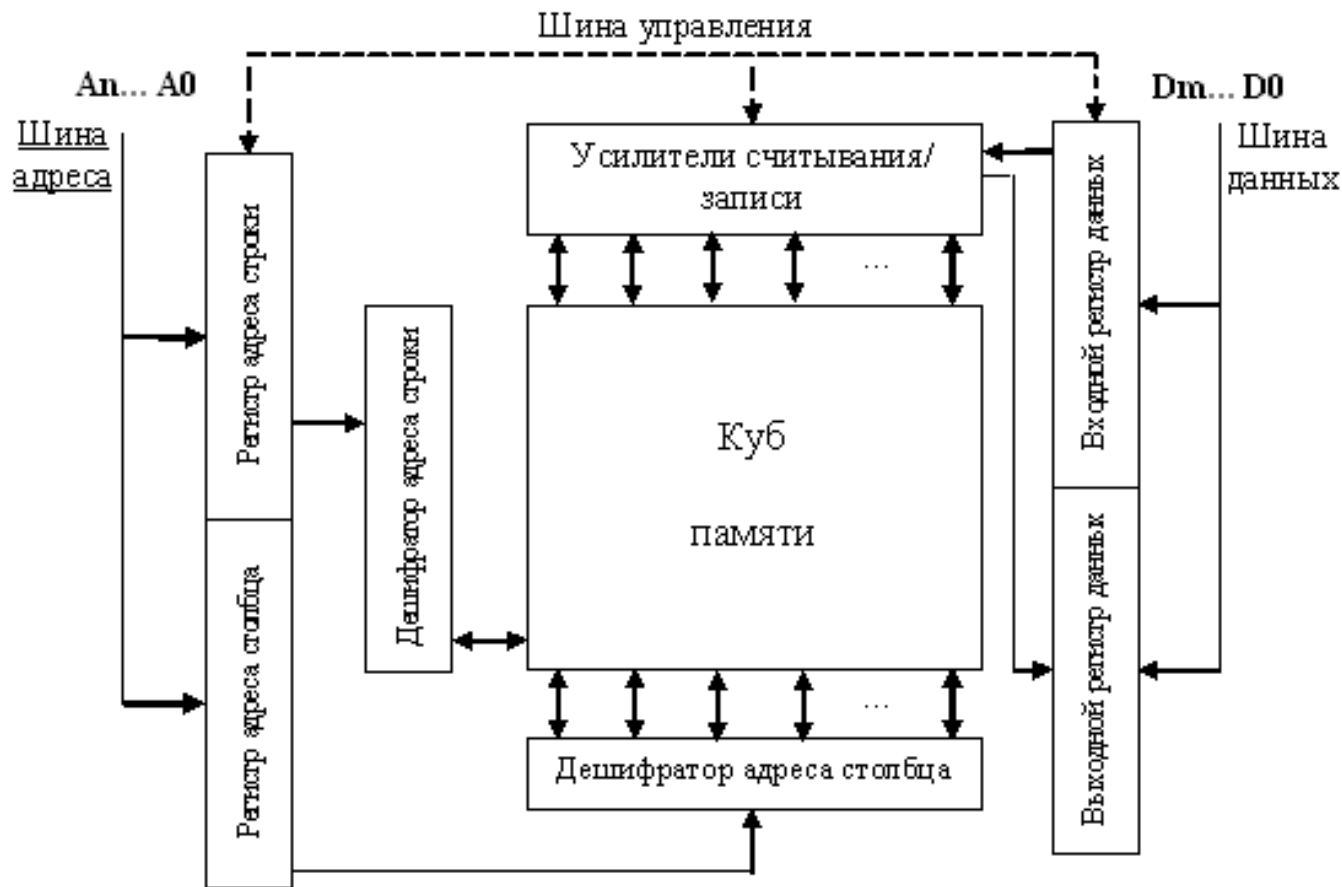
Основную память образуют устройства с произвольным доступом, которые представляют собой массив ячеек; обращение к одной ячейке занимает одно и то же время и может производиться в произвольной последовательности. Каждая ячейка содержит фиксированное число запоминающих элементов и имеет уникальный адрес.

Основная память включает в себя ОЗУ и ПЗУ.

Для запоминающего элемента любой полупроводниковой памяти характерны следующие свойства:

- 1) два стабильных состояния, представляющие двоичные «0» и «1»;
- 2) в запоминающий элемент (хотя бы однажды) может быть произведена запись информации посредством перевода его в одно из двух возможных состояний;
- 3) для определения текущего состояния запоминающего элемента его содержимое может быть считано.

Структура основной памяти



Структура модуля основной памяти

Структура основной памяти

Адрес ячейки, поступающий по шине адреса, разделяется на две составляющие: *адрес строки* и *адрес столбца*, которые запоминаются в соответствующих регистрах микросхемы. Каждый регистр соединён со своим дешифратором. Выходы дешифраторов образуют систему горизонтальных и вертикальных линий, к которым подсоединены запоминающие элементы куба памяти; при этом каждый запоминающий элемент расположен на пересечении одной горизонтальной и одной вертикальной линии.

Запоминающие элементы, объединённые общим «горизонтальным» проводом, называют ***строкой***. Запоминающие элементы, подключённые к общему «вертикальному» проводу, называют ***столбцом***.

Совокупность запоминающих элементов и логических схем, связанных с выбором строк и столбцов, называется ***ядром*** микросхемы памяти.

Количество запоминающих элементов, имеющих один и тот же адрес, называется ***ячейкой***.

Структура основной памяти

Получив значения полуадресов из регистра адреса строки и регистра адреса столбца, дешифраторы адресов строки и столбца определяют положение ячейки для чтения или для записи данных, посылая сигналы по соответствующим выходам к строке и столбцу.

Информация для записи, поступающая по шине данных, сначала заносится во входной регистр данных, а затем – в выбранную ячейку. При выполнении операции чтения информация из ячейки до выдачи на шину данных предварительно помещается в выходной регистр данных. Роль входного и выходного регистров может выполнять один регистр.

Усилители считывания/записи служат для электрического согласования сигналов на линиях данных и внутренних сигналов микросхемы памяти.

Управление операциями с основной памятью осуществляется контроллером памяти, который входит в состав центрального процессора либо реализуется в виде внешнего по отношению к памяти устройства. В последних типах микросхем памяти часть функций контроллера возлагается на саму микросхему. Контроллер памяти является синхронным устройством, срабатывающим исключительно по тактовым импульсам. В общем случае, на каждую операцию с памятью требуется, как минимум, пять тактов.

Структура основной памяти

Конструктивно модуль (микросхема) памяти представляет собой небольшую печатную плату, на которой установлены необходимые интегральные схемы. В настоящее время выделяют следующие типы модулей /2, 3, 5, 6, 9 – 11/:

1. SIMM (Single In-Line Memory Module) – модули с однорядным расположением выводов, в которых установлены асинхронные СБИС DRAM.

2. DIMM (Dual In-Line Memory Module) – модули с двухрядным расположением выводов, в которых могут быть установлены как асинхронные DRAM, так и синхронные DRAM (SDRAM – Synchronous DRAM).

3. RIMM (Rambus In-Line Memory Module) – модули, в которых установлены СБИС фирмы «Rambus» и используется интерфейс Rambus Channel.

В настоящее время среди быстродействующих СБИС DRAM доминируют SDRAM и DRDRAM (Direct Rambus DRAM).

Память с последовательным доступом

Память с последовательным доступом строится либо с использованием продвижения данных по цепочке элементов (по подобию с регистрами сдвига), либо с хранением данных в адресном ЗУ при соответствующем управлении адресом доступа.

Основными представителями такого вида памяти являются видеопамять, буфер FIFO, стек.

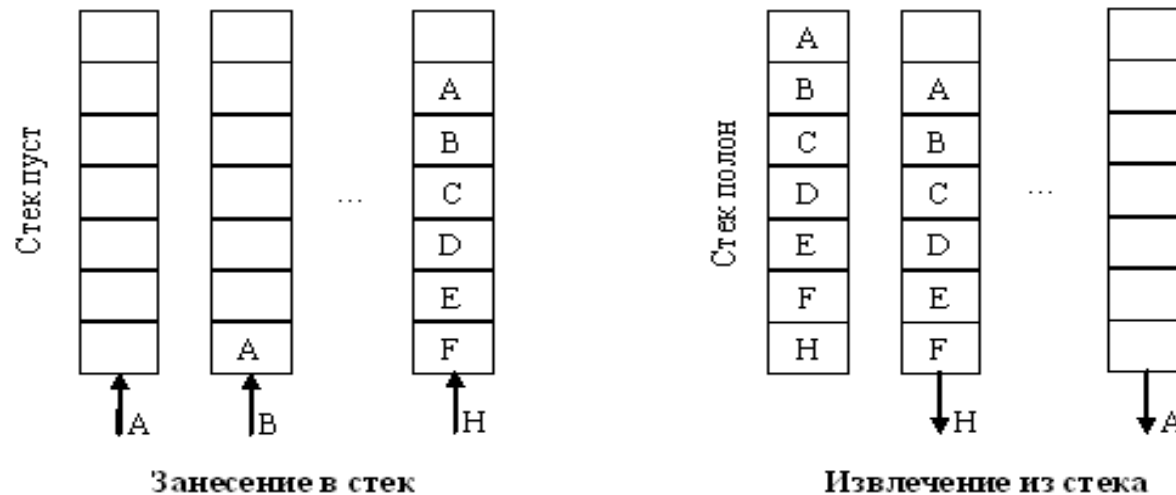
Стек – это однонаправленная очередь, данные в которую помещаются и извлекаются в строго определённом порядке. Стековая память обеспечивает такой режим работы, когда информация записывается и считывается по принципу «последним записан – первым считан» (LIFO – Last Input First Output). Такая память используется для временного хранения данных, например, для запоминания и восстановления регистров процессора (контекста) при обработке подпрограмм и прерываний.

Занесение информации в стек называется **включением**, считывание информации из стека – **извлечением**.

Память с последовательным доступом

Когда слово *A* заносится в стек, то располагается в первой свободной ячейке. Каждое следующее записываемое слово перемещает всё содержимое стека на одну ячейку вверх и занимает освободившуюся ячейку. Запись очередного слова после *H* приводит к переполнению стека, поскольку он рассчитан на 7 слов, и потере кода *A*.

Считывание информации из стека осуществляется в обратном порядке, т.е., начиная с кода *H*, который был записан последним. Доступ к произвольному коду в стеке формально недопустим до извлечения всех данных, записанных позже.



Логика работы стековой памяти

Разработал: Конюхова О.В.

Память с последовательным доступом



Схема организации стека для процессора Intel 8086

Память с последовательным доступом

Наиболее распространённым является внешний, или аппаратно-программный, стек, в котором для хранения информации отводится область оперативной памяти. Обычно под стек отводится участок памяти с наибольшими адресами, а расширяется стек в сторону уменьшения адресов.

Под стек выделяется отдельный сегмент – сегмент стека, начальный адрес которого помещается в соответствующий сегментный регистр – SS. Адресация стека обеспечивается специальным регистром – указателем стека SP, в который предварительно помещается наибольший адрес области основной памяти, отведённой под стек (дно стека). Адрес последнего включённого в стек элемента называется **вершиной стека** (TOS– Top Of Stack).

Для работы со стеком существуют две основные операции: *добавление элемента в вершину стека* (PUSH) и *извлечение элемента из вершины стека* (POP). Команда PUSH имеет один операнд, который может быть непосредственным значением, 2-байтовым регистром или адресом ячейки памяти. При записи в стек данного сначала производится уменьшение на 2 содержимого указателя стека SP (стек оперирует словами), которое затем используется в качестве адреса ячейки, куда и производится запись. Команда POP также имеет один операнд, который может быть 2-байтовым регистром. При считывании слова из стека в качестве адреса этого данного берётся текущее содержимое указателя стека, а после извлечения данного содержимое SP увеличивается на 2.

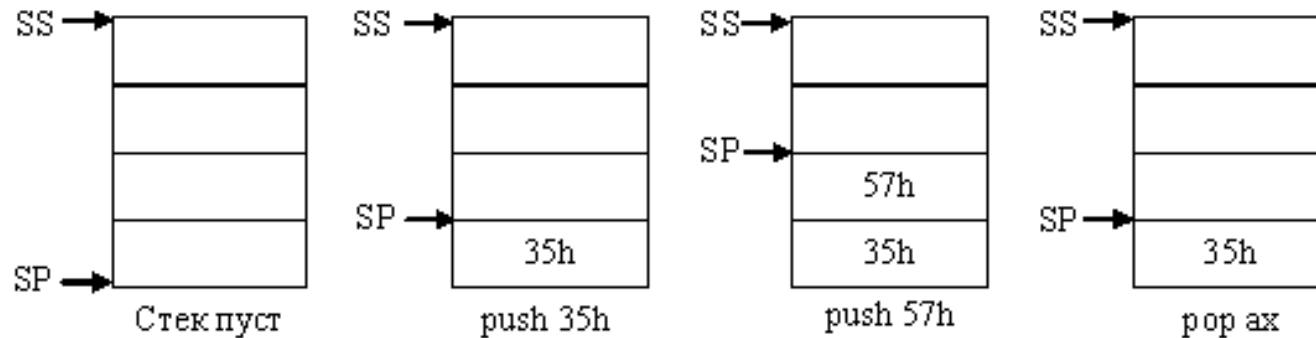
Память с последовательным доступом

Рассмотрим следующий пример:

```
push 35h  
push 57h  
pop ax
```

Изначально стек пуст, и регистр *SP* указывает на ячейку за дном стека. При включении первого значения содержимое *SP* уменьшается на 2 и затем по полученному адресу помещается 35h. Аналогично со вторым значением. При извлечении слова из стека в регистр *AX* помещается значение из вершины стека, т.е., 57h, а содержимое *SP* увеличивается на 2.

Состояние стека при выполнении указанных выше команд иллюстрирует рисунок.



Изменение состояния стека

Разработал: Конюхова О.В.

Ассоциативная память

В ряде случаев значительно удобнее искать информацию не по адресу, а по некоторому характерному признаку, содержащемуся в самой информации. Такой принцип лежит в основе *ассоциативного ЗУ* (АЗУ).

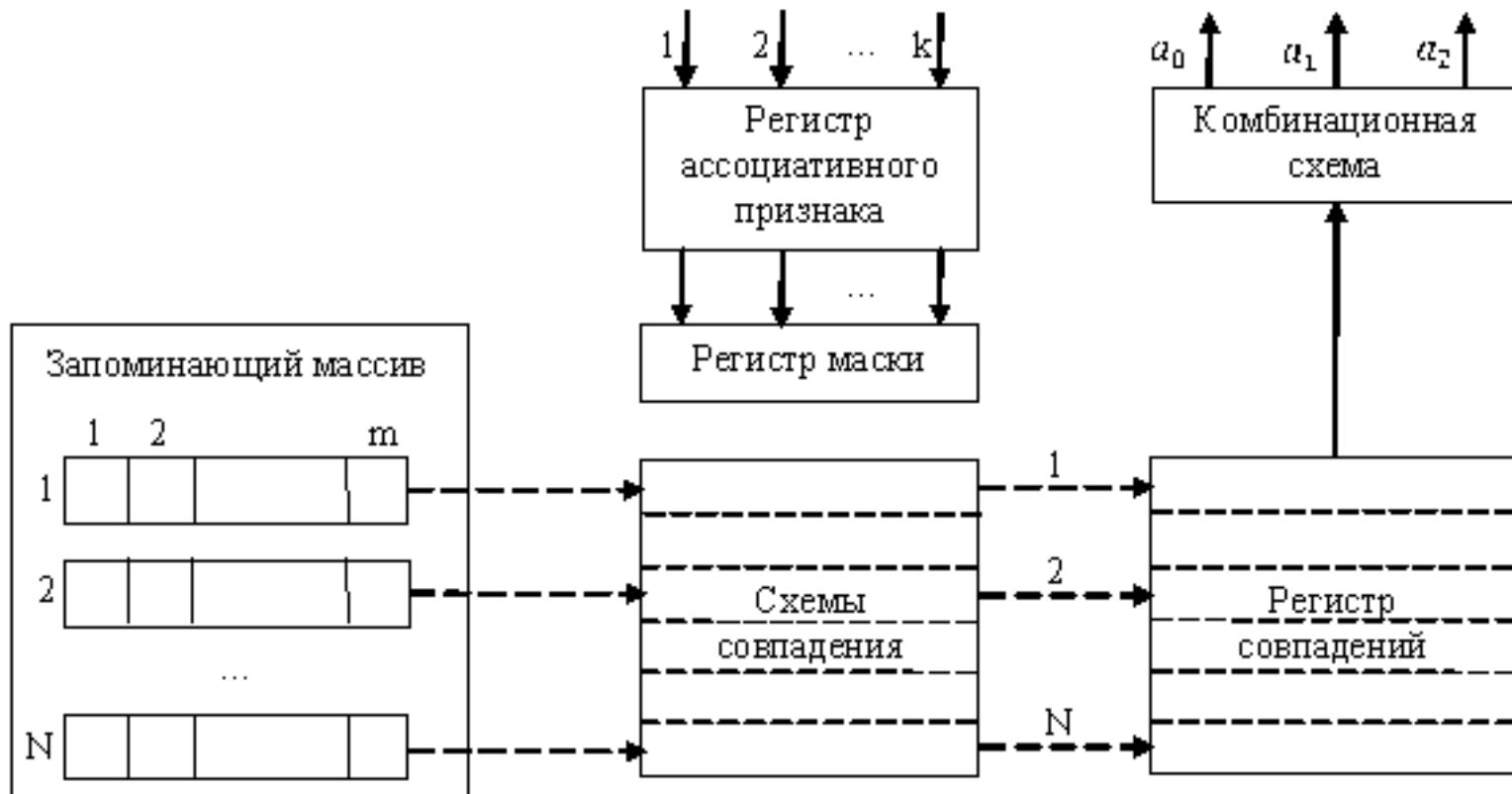
АЗУ – это устройство, способное хранить информацию, сравнивать её с некоторым заданным образцом и указывать на их соответствие или несоответствие друг другу.

Признак, по которому производится поиск информации, называется ***ассоциативным признаком***.

Кодовая комбинация, используемая в роли образца для поиска, называется ***признаком поиска***.

Ассоциативный признак может быть частью искомой информации или дополнительно передаваться ей. В последнем случае его называют ***тэгом*** или ***ярлыком***.

Ассоциативная память



Структура АЗУ

Ассоциативная память

АЗУ включает в себя:

1. Запоминающий массив для хранения N m -разрядных слов, в каждом из которых несколько младших разрядов занимает служебная информация.
2. Регистр ассоциативного признака, куда помещается признак поиска; разрядность регистра k обычно меньше длины слова m .
3. Схемы совпадения, которые используются для параллельного сравнения каждого бита хранимых слов с соответствующим битом признака поиска и выработки сигналов совпадения.
4. Регистр совпадений, где каждой ячейке запоминающего массива соответствует один разряд, в который заносится единица, если все разряды соответствующей ячейки совпали с одноимёнными разрядами признака поиска.
5. Регистр маски, позволяющий запретить сравнение определённых битов.
6. Комбинационную схему, которая на основании анализа содержимого регистра совпадений формирует сигналы, определяющие результаты поиска информации.

Ассоциативная память

При обращении к АЗУ сначала в регистре маски обнуляются разряды, которые не должны учитываться при поиске информации.

Все разряды регистра совпадений устанавливаются в единичное состояние.

В регистр признака заносится код искомой информации; в процессе поиска схемы совпадения одновременно сравнивают соответствующие биты ячеек запоминающего массива с соответствующим (по разряду) битом признака.

Те схемы, которые зафиксировали несовпадение, переводят соответствующие биты регистра совпадений в нулевое состояние. Тогда единицы сохраняются лишь в тех разрядах регистра совпадений, которые соответствуют ячейкам, где найдена искомая информация. Конфигурация единиц в регистре совпадений используется в качестве адресов, по которым производится считывание из запоминающего массива.

Сигналы результата поиска могут принимать следующие значения: a_0 – искомая информация не найдена; a_1 – искомая информация находится в одной ячейке; a_2 – искомая информация содержится в более, чем одной ячейке.

Ассоциативная память

Запись в АЗУ производится без указания конкретного адреса, в первую свободную ячейку. Свободной считается либо пустая ячейка, либо та, которая дольше всего не использовалась.

Главное преимущество АЗУ заключается в том, что время поиска информации зависит только от числа разрядов в признаке поиска и скорости опроса разрядов и не зависит от числа ячеек в запоминающем массиве. Из-за относительно высокой стоимости АЗУ редко используется как самостоятельный вид памяти.

Организация флэш-памяти

Флэш-память – это особый вид энергонезависимой перезаписываемой полупроводниковой твердотельной памяти.

Впервые была разработана компанией «Toshiba» в 1984 году. В 1988 г. компания «Intel» разработала собственный вариант флэш-памяти.

Полное историческое название флэш-памяти – ***Flash Erase Electronically Electrically Programmable ROM – электрически стираемое перепрограммируемое ПЗУ***. Считается, что название Flash было дано компанией «Toshiba» во время разработки первых микросхем флэш-памяти как характеристика скорости стирания информации в микросхеме («in a flash» – мгновенно).

Главной отличительной особенностью флэш является возможность перепрограммирования при подключении к стандартной системной шине микропроцессора. Число циклов репрограммирования флэш-памяти хотя и велико (от 10 000 до 1 000 000 раз), но ограничено.

Внешне флэш-память представляет собой микросхему. В отличие от жёстких дисков, CD- и DVD-ROM, во флэш-носителях нет движущихся частей, поэтому их и называют твердотельными. По оценкам производителей, информация на флэш может храниться от 20 до 100 лет. Благодаря компактным размерам, высокой надёжности и низкому энергопотреблению, флэш-память активно используется в современных ВМ в качестве съёмного носителя информации.

Организация флэш-памяти

Одним из элементов структуры флэш-памяти является накопитель (матрица запоминающих элементов). По организации массива запоминающих элементов различают микросхемы флэш-памяти трёх типов:

1. *Bulk Erase* – стирание допустимо только для всего массива запоминающих элементов.
2. *Boot Block* – массив запоминающих элементов разделён на несколько блоков разного размера, содержимое которых может осуществляться независимо. Среди блоков есть так называемый загрузочный блок, содержимое которого аппаратно защищено от случайного стирания. В нём хранится программное обеспечение для правильной эксплуатации и инициализации микросхемы.
3. *Flash File* – массив запоминающих элементов разделён на несколько равноправных блоков одинакового размера, содержимое которых может стираться независимо. Файловая флэш-память ориентирована на замену жёстких дисков, поэтому её блоки являются аналогами секторов магнитных дисков.

Организация флэш-памяти

Базовым элементом матрицы является флэш-ячейка, которая состоит из транзистора особой архитектуры – полевого двухзатворного транзистора.

Ячейки флэш-памяти можно соединить последовательно (в цепочку) или параллельно. В первом случае логическая организация памяти называется NAND (Not AND, НЕ-И), во втором – NOR (Not OR, НЕ-ИЛИ).

Процесс записи информации для ячеек NOR и NAND различен. В ячейках NOR запись осуществляется *методом инжекции* (когда электронам даётся дополнительная энергия для преодоления потенциального барьера перед изолированным затвором). В ячейках NAND запись производится *путём туннелирования* электронов (электроны переходят сквозь барьер).

Стирание информации в ячейках обоих типов осуществляется механизмом туннельного перехода.

Организация флэш-памяти

В чистом виде флэш-память NOR похожа на обычную оперативную память, а память NAND больше напоминает дисковый накопитель с блочным доступом.

При существенно большей скорости чтения данных память NOR имеет значительно меньшую скорость стирания и несколько меньшую скорость записи по сравнению с памятью NAND. Однако память NOR позволяет исполнять записанный код, а программы из NAND перед исполнением всегда необходимо предварительно загружать в оперативную память.

Ячейка NAND имеет значительно меньшие размеры в сравнении с ячейкой NOR. Хотя первоначально более широко была распространена память NOR в виде оперативной памяти для ВМ и других программируемых устройств, то в настоящее время широкое применение получила память NAND в виде карт памяти и твердотельных носителей.

Однако для работы, требующей побайтового произвольного доступа (например, для хранения программного кода), память NOR предпочтительней.

Следует отметить, что существуют гибридные решения, в одном корпусе объединяющие разные типы памяти.

Организация флэш-памяти

Одна ячейка флэш-памяти содержит один транзистор, который в простейшем случае хранит один бит информации. Такие ячейки называются **одноуровневыми (SLC – Single Level Cell)**. Создание **многоуровневой** ячейки или ячейки **с многоуровневым кодированием (MLC – Multi Level Cell)** позволило хранить на одном транзисторе два бита информации. В качестве опытных образцов существуют 4-битовые ячейки. Компания «Intel» в 1997 г. представила флэш-память с 2-битовыми ячейками, которая получила название Strata Flash.

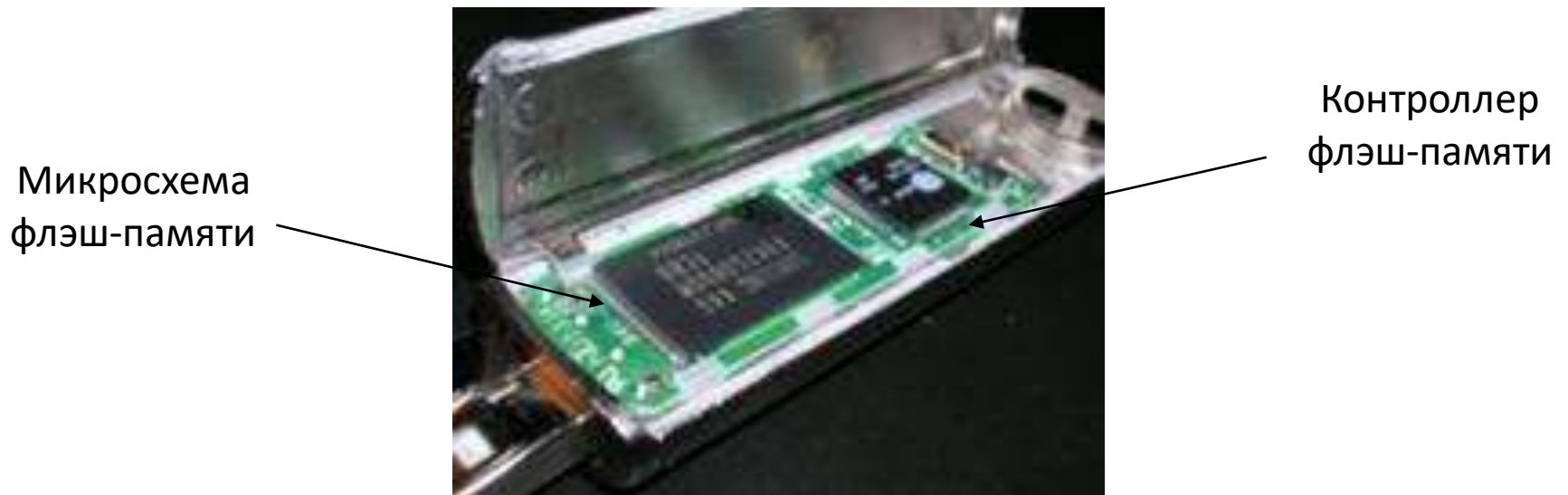
Ячейки MLC применяются в памяти NOR, но наиболее популярны в памяти NAND. Память SLC NAND, в сравнении с MLC NAND, имеет более высокие скорости передачи данных, меньшее энергопотребление и повышенную надёжность. Однако MLC NAND имеет большую ёмкость и меньшую цену.

Современные микросхемы флэш-памяти могут достигать объёма 2 Тбайта.

Организация флэш-памяти

Накопители на флэш-памяти представляют собой микросхему флэш-памяти, дополненную контроллером USB, и подключаются к последовательному USB-порту.

Контроллер может быть выполнен в виде отдельной микросхемы либо встроен в микросхему флэш-памяти.



USB-накопитель на флэш-памяти

Архитектурные способы повышения скорости обмена между ЦП и памятью

Быстродействие СБИС DRAM увеличивается существенно медленнее, чем быстродействие процессоров.

Архитектурное решение – встраивание в структуру ВМ **кэш-памяти**– быстродействующего буфера между основной памятью и регистрами процессора.

В эту буферную память из основной памяти помещаются копии команд и данных обрабатываемого фрагмента программы. Работа кэш-памяти скрыта от пользователя. Кэш-память наиболее эффективна, если встроена внутрь кристалла процессора (кэш-память 1-го уровня – L1). Её объём составляет 16 –32 Кбайт.

В большинстве современных компьютеров используют 2- или 3-уровневую кэш-память, L2 и L3 соответственно. Кэш-память L2 объёмом 256 – 512 Кбайт часто располагается в одном корпусе с процессором и соединяется с ним с помощью специальной локальной шины, работающей на основной или половинной частоте процессора.

Кэш-память L3 объёмом в несколько Мбайт размещается на системной плате компьютера. Обычно всё содержимое кэш L1 находится в кэш L2, а всё содержимое L2 является частью кэш L3.

Архитектурные способы повышения скорости обмена между ЦП и памятью

Вся доступная программе информация размещается в оперативной памяти.

При обращении процессора к памяти вначале проверяется наличие требуемых данных в кэш-памяти. Обнаружение искомой информации фиксируется как кэш-попадание, в противном случае фиксируется кэш-промах.

Обмен данными между кэшем и оперативной памятью осуществляется информационными блоками. В современных микропроцессорных системах используются блоки фиксированного размера, например, 32 байта.

Управляет кэш-памятью специальный контроллер кэша. Если адресуемый операнд находится в кэше, он быстро извлекается из него (при чтении) или результат операции заносится в кэш (при записи). При отсутствии требуемого операнда в кэш-памяти процессор считывает из оперативной памяти блок данных, содержащий искомый операнд и помещает его в кэш.

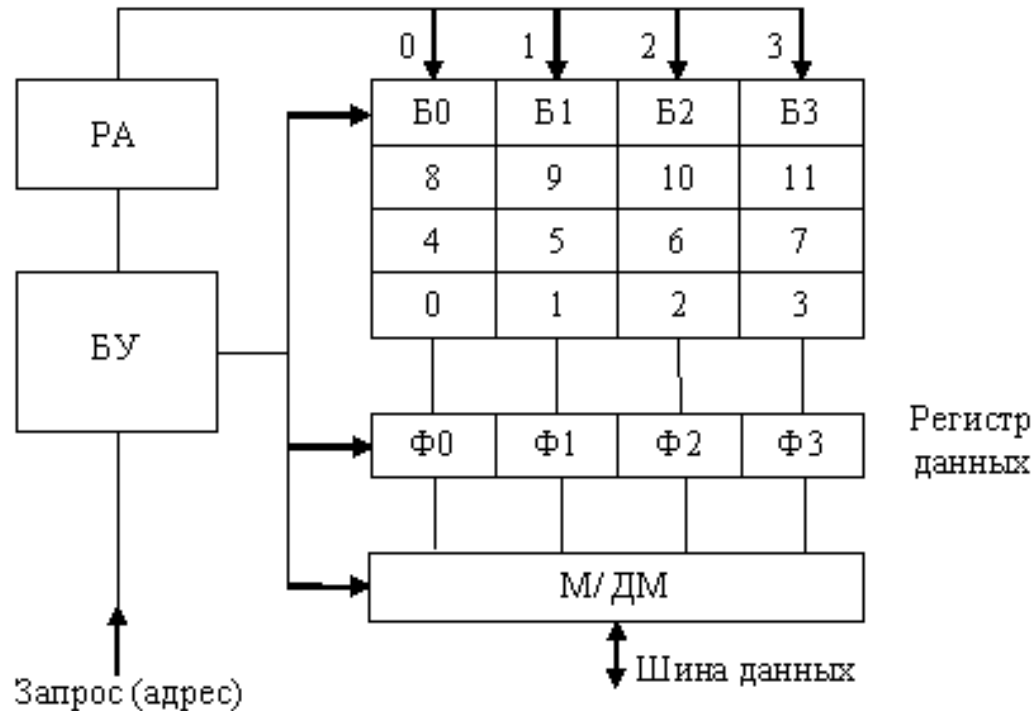
Запись данных, не имеющих копий в кэше, проводится непосредственно в оперативную память.

Архитектурные способы повышения скорости обмена между ЦП и памятью

Можно выделить следующие способы обмена данными процессора с оперативной памятью, построенной на СБИС DRAM:

- *пакетный доступ,*
- *конвейерный доступ и*
- *их сочетание.*

Архитектурные способы повышения скорости обмена между ЦП и памятью



Структура оперативной памяти с пакетным доступом и чередованием банков

Содержит в себе следующие блоки: *накопитель*, разделённый на 4 банка (B0, B1, B2, B3); *регистр данных*, содержащий 4 субрегистра – фиксатора данных (Ф0, Ф1, Ф2, Ф3), соответствующих 4-м банкам; *мультиплексор (демультиплексор) – М/ДМ*; *регистр адреса (РА)*, *блок управления (БУ)*.

Архитектурные способы повышения скорости обмена между ЦП и памятью

Разрядность ячеек каждого банка соответствует разрядности шины данных в системной шине. Адреса присваиваются ячейкам оперативной памяти с чередованием номера банка (на рисунке: 0, 1, ..., 11).

Накопитель (упорядоченный массив из n -разрядных ячеек) подразделяется на m банков: , k – целое. В рассматриваемом примере $k=2$.

При этом адрес i -й ячейки представляется следующим образом:

$$i = dm + b = d \cdot 2^k + b$$

где d – адрес внутри банка, b – номер банка ($b = 0, 1, \dots, m-1$). Такое распределение адресов между m банками называют ***m -кратным чередованием банков***.

При пакетном обмене в РА запоминается адрес внутри банка. БУ обеспечивает считывание целой строки ячеек из накопителя, соответствующей адресу в РА во всех m банках.

Содержимое считанных ячеек фиксируется в регистре данных. Далее данные из Φ_0, Φ_1, \dots по очереди с использованием мультиплексора передаются по системной шине данных.

Архитектурные способы повышения скорости обмена между ЦП и памятью

Пусть, например, осуществляется считывание пакета из ячеек с адресами 4, 5, 6, 7 (рис. на слайде 39). Если обозначить время доступа через $T_{\text{дост}}$, а время передачи через $T_{\text{пер}}$, то оценка времени чтения пакета $T_{\text{чт}}$ без чередования банков вычисляется по формуле

$$T_{\text{чт}} = 4(T_{\text{дост}} + T_{\text{пер}});$$

а с чередованием банков – по формуле

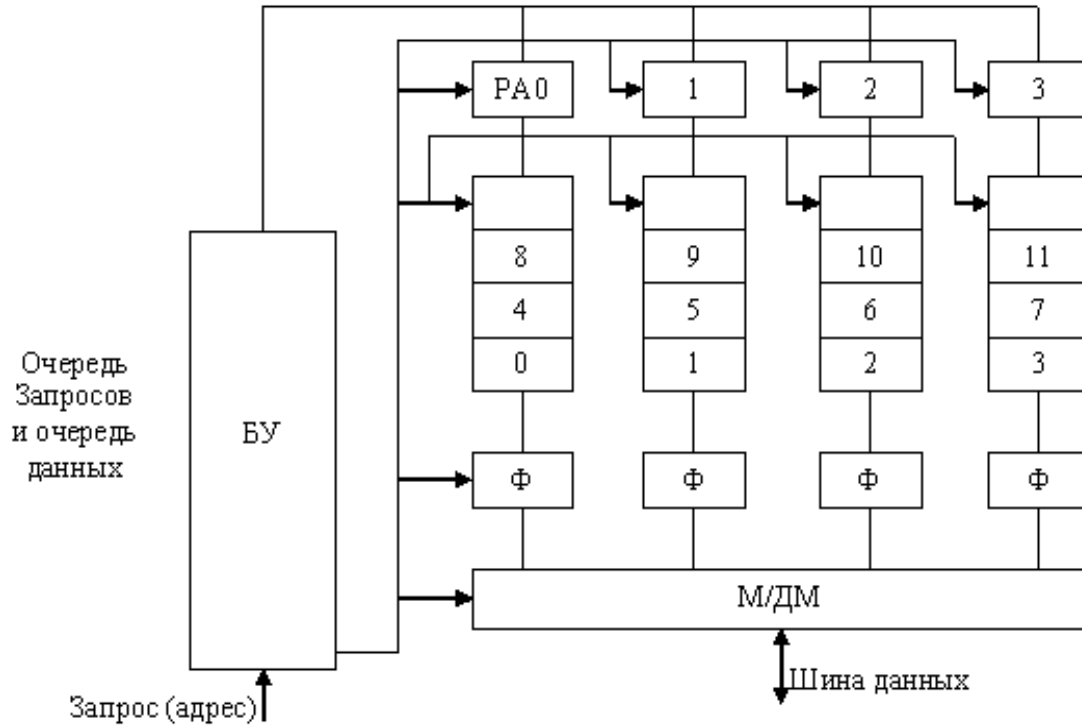
$$T_{\text{чт}} = T_{\text{дост}} + 4T_{\text{пер}}$$

Быстродействие памяти по отношению к быстродействию процессора характеризуется числом тактов ожидания в цикле обращения к памяти. Допустим, что время доступа соответствует трём тактам, а время передачи – двум тактам.

Тогда в случае без чередования банков время передачи пакета из четырёх слов можно охарактеризовать вектором (5, 5, 5, 5), а при использовании чередования банков – (5, 2, 2, 2).

При записи работа организована следующим образом: данные пакета из процессора по очереди записываются в субрегистры Ф0, ..., Ф3, а затем осуществляется запись одновременно в ячейки всех банков.

Архитектурные способы повышения скорости обмена между ЦП и памятью



*Структура
оперативной памяти
с конвейерным
доступом и
чередованием банков*

Она содержит следующие блоки: *накопитель*, разделённый на 4 банка (Б0, Б1, Б2, Б3); 4 *регистра данных* (Φ) с независимым управлением (в отличие от структуры на рис. 43); *мультиплексор* (демультиплексор) – М/ДМ; *регистры адреса* РА0-РА3 для каждого банка; *блок управления* (БУ).

Архитектурные способы повышения скорости обмена между ЦП и памятью

Разрядность ячеек для каждого банка соответствует разрядности данных в системной шине. Адреса присваиваются ячейкам оперативной памяти с чередованием номера банка (на рис.: 0, 1, ..., 11).

Цикл обращения к памяти (например, чтения) содержит три фазы: доступ к ячейке, воспроизведение данных, передача данных от регистров Ф через мультиплексор и системную шину в процессор.

В качестве примера рассмотрим чтение пакета с последовательностью адресов 8, 6, 1, 7. Ограничение на расположение адресов ячеек здесь менее жёсткое, чем при пакетном доступе.

Требуется, чтобы адреса ячеек пакета находились в разных банках.

Если длительность фаз считать одинаковой, равной T , то справедливы следующие оценки для времени чтения пакета из четырёх слов: без конвейеризации – по формуле

$$T_{чм} = (3 \times 4)T$$

с конвейеризацией – по формуле:

$$T_{чм} = 4T + 2T$$

Архитектурные способы повышения скорости обмена между ЦП и памятью

В общем случае возможно сочетание обоих изложенных способов доступа.

При этом накопитель оперативной памяти разбивается на $m \times k$ банков, образующих прямоугольную матрицу. Адреса ячеек располагаются по порядку – сначала в банках первой строки, затем второй и т.д.

Организуется пакетный доступ к строке ячеек и конвейерный доступ к строкам. В каждом интервале, равном $1/m$ цикла памяти, осуществляется доступ к k ячейкам с идущими подряд адресами.