

В любой ВМ, вне зависимости от ее архитектуры, программы и данные хранятся в памяти. Функции памяти обеспечиваются запоминающими устройствами (ЗУ), предназначенными для фиксации, хранения и выдачи информации в процессе работы ВМ. Процесс фиксации информации в ЗУ называется *записью*, процесс выдачи информации — *чтением* или *считыванием*, а совместно их определяют как *процессы обращения к ЗУ*.

9.1 Характеристики систем памяти

Перечень основных характеристик, которые необходимо учитывать, рассматривая конкретный вид ЗУ, включает в себя:

- место расположения;
- емкость;
- единицу пересылки;
- метод доступа;
- быстродействие;
- физический тип;
- физические особенности;
- стоимость.

По месторасположения ЗУ разделяют на процессорные, внутренние и внешние. Наиболее скоростные виды памяти (регистры, кэш-память первого уровня) обычно размещают на общем кристалле с центральным процессором, а регистры общего назначения вообще считаются частью ЦП. Вторую группу (внутреннюю память) образуют ЗУ, расположенные на системной плате. К внутренней памяти относят память, а также кэш-память второго и последующих уровней (кэш-память второго уровня может также размещаться на кристалле процессора). Медленные ЗУ большой емкости (магнитные и оптические диски, магнитные ленты) называют внешней памятью, поскольку к ядру ВМ они подключаются аналогично устройствам ввода/вывода.

Емкость ЗУ характеризуют числом битов либо байтов, которое может храниться в запоминающем устройстве. На практике применяются более крупные единицы, а для их обозначения к словам «бит» или «байт» добавляют приставки кило, мега, гига, тера, пета, экса (kilo, mega, giga, tera, peta, exa). Стандартно эти приставки означают умножение основной единицы измерений на 10^3 , 10^6 , 10^9 , 10^{12} ; 10^{18} соответственно. В вычислительной технике, ориентированной на двоичную систему счисления, они соответствуют значениям достаточно близким к стандартным, но представляющим собой целую степень числа 2, то есть 2^{10} , 2^{20} , 2^{30} , 2^{50} , 2^{60} . Во избежание разночтений, в последнее время ведущие международные организации по стандартизации, например IEEE (Institute of Electrical and Electron'* Engineers), предлагают ввести новые обозначения, добавив к основной пристав слово binary (бинарный): kilobinary, megabinary, gigabinary, terabinary, petabinary, exabinary. В результате вместо термина «килобайт» предлагается термин «кибибайт», вместо «мегабайт» — «мебибайт» и т. д.

Для обозначения новых единиц предполагаются сокращения: Ki, Mi, Gi, Ti, Pi и Ei [133].

Важной характеристикой ЗУ является *единица пересылки*. Для основной памяти (ОП) единица пересылки определяется шириной шины данных, то есть количеством битов, передаваемых по линиям шины параллельно. Обычно единица пересылки равна длине слова, но не обязательно. Применительно к внешней памяти данные часто передаются единицами, превышающими размер слова, и такие единицы называются *блоками*.

При оценке быстродействия необходимо учитывать применяемый в данном типе ЗУ *метод доступа* к данным. Различают четыре основных метода доступа:

- **Последовательный доступ.** ЗУ с последовательным доступом ориентировано на хранение информации в виде последовательности блоков данных, называемых записями. Для доступа к нужному элементу (слову или байту) необходимо прочитать все предшествующие ему данные. Время доступа зависит от положения требуемой записи в последовательности записей на носителе информации и позиции элемента внутри данной записи. Примером может служить ЗУ на магнитной ленте.

- **Прямой доступ.** Каждая запись имеет уникальный адрес, отражающий ее физическое размещение на носителе информации. Обращение осуществляется как адресный доступ к началу записи, с последующим последовательным доступом к определенной единице информации внутри записи. В результате время доступа к определенной позиции является величиной переменной. Такой режим характерен для магнитных дисков.

- **Произвольный доступ.** Каждая ячейка памяти имеет уникальный физический адрес. Обращение к любой ячейке занимает одно и то же время и может производиться в произвольной очередности. Примером могут служить запоминающие устройства основной памяти.

- **Ассоциативный доступ.** Этот вид доступа позволяет выполнять поиск ячеек содержащих такую информацию, в которой значение отдельных битов совпадает с состоянием одноименных битов в заданном образце. Сравнение осуществляется параллельно для всех ячеек памяти, независимо от ее емкости ассоциативному принципу построены некоторые блоки кэш-памяти.

Быстродействие ЗУ является одним из важнейших его показателей. Для количественной оценки быстродействия обычно используют три параметра:

Время доступа (T_d). Для памяти с произвольным доступом оно соответствует интервалу времени от момента поступления адреса до момента, когда данные заносятся в память или становятся доступными. В ЗУ с подвижным носителем информации — это время, затрачиваемое на установку головки записи/считывания (или носителя) в нужную позицию.

Длительность цикла памяти или *период обращения* (T_n). Понятие применяется к памяти с произвольным доступом, для которой оно означает минимальное время между двумя последовательными обращениями к памяти. Период обращения включает в себя время доступа плюс некоторое дополнительное время. Дополнительное время может требоваться для

затухания сигналов на линиях, а в некоторых типах ЗУ, где считывание информации приводит к ее разрушению, — для восстановления считанной информации.

Скорость передачи. Это скорость, с которой данные могут передаваться в память или из нее. Для памяти с произвольным доступом она равна $1/\Gamma_{\text{ц}}$. Для других видов памяти скорость передачи определяется соотношением:

$$T_N = T_A + \frac{N}{R},$$

где T_N — среднее время считывания или записи N битов; T_A — среднее время доступа; R — скорость пересылки в битах в секунду.

Говоря о *физическом типе* запоминающего устройства, необходимо упомянуть три наиболее распространенных технологии ЗУ — это полупроводниковая память, память с магнитным носителем информации, используемая в магнитных дисках и лентах, и память с оптическим носителем — оптические диски.

В зависимости от примененной технологии следует учитывать и ряд *физических особенностей* ЗУ, например энергозависимость. В энергозависимой памяти информация может быть искажена или потеряна при отключении источника питания. В энергонезависимых ЗУ записанная информация сохраняется и при отключении питающего напряжения. Магнитная и оптическая память — энергонезависимы. Полупроводниковая память может быть как энергозависимой, так и нет, зависимости от ее типа. Помимо энергозависимости нужно учитывать, приводит и считывание информации к ее разрушению.

Стоимость ЗУ принято оценивать отношением общей стоимости ЗУ к его емкости в битах, то есть стоимостью хранения одного бита информации.

9.2 Иерархия запоминающих устройств

Память часто называют «узким местом» фон-неймановских ВМ из-за ее серьезного отставания по быстродействию от процессоров, причем разрыв этот неуклонно увеличивается. Так, если производительность процессоров ежегодно возрастает вдвое примерно каждые 1,5 года, то для микросхем памяти прирост быстродействия не превышает 9% в год (удвоение за 10 лет), что выражается в увеличении разрыва в быстродействии между процессором и памятью приблизительно на 50% в год.

Если проанализировать используемые в настоящее время типы ЗУ, выявляется следующая закономерность:

- чем меньше время доступа, тем выше стоимость хранения бита;
- чем больше емкость, тем ниже стоимость хранения бита, но больше время доступа.

При создании системы памяти постоянно приходится решать задачу обеспечения требуемой емкости и высокого быстродействия за приемлемую цену. Наиболее распространенным подходом здесь является построение системы памяти ВМ по иерархическому принципу. *Иерархическая память*

состоит из ЗУ различных типов (рис. 8.2.1), которые, в зависимости от характеристик, относят к определенному уровню иерархии. Более высокий уровень меньше по емкости, быстрее и имея большую стоимость в пересчете на бит, чем более низкий уровень. Уровни иерархии взаимосвязаны: все данные на одном уровне могут быть также найдены на более низком уровне, и все данные на этом более низком уровне могут быть найдены на следующем ниже лежащем уровне и т.д.

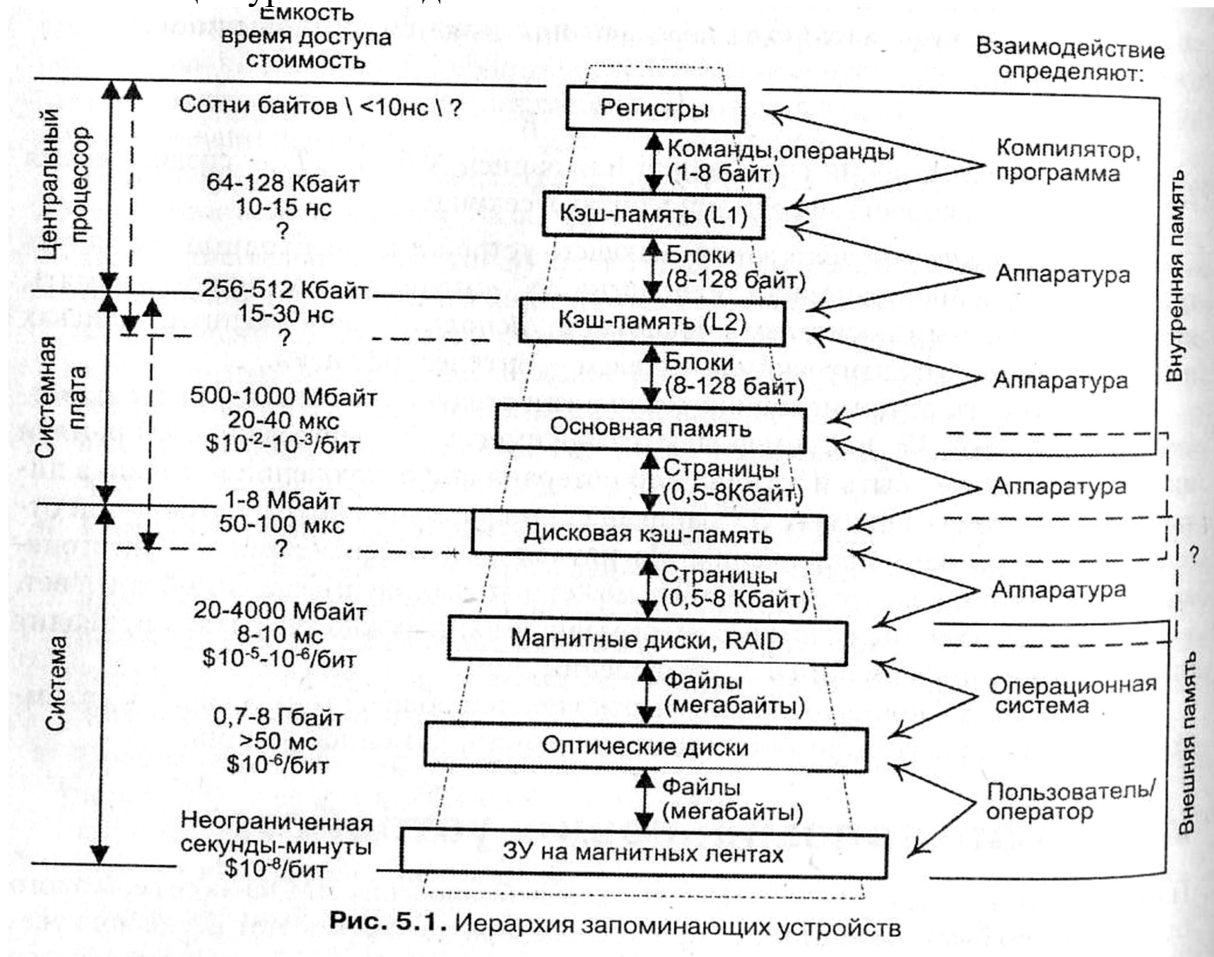


Рис. 5.1. Иерархия запоминающих устройств

Рис. 9.2.1 Иерархия запоминающих устройств

Четыре верхних уровня иерархии образуют *внутреннюю память* ВМ, а все нижние уровни — это *внешняя* или *вторичная* память. По мере движения вниз по иерархической структуре:

1. Уменьшается соотношение «стоимость/бит».
2. Возрастает емкость.
3. Растет время доступа.
4. Уменьшается частота обращения к памяти со стороны центрального процессора.

Если память организована в соответствии с пунктами 1-3, а характер размещения в ней данных и команд удовлетворяет пункту 4, иерархическая организация ведет к уменьшению общей стоимости при заданном уровне производительности.

Справедливость этого утверждения вытекает из принципа *локальности по обращению*. Если рассмотреть процесс выполнения большинства

программ, то можно заметить, что с очень высокой вероятностью адрес очередной команды программы либо следует непосредственно за адресом, по которому была считана текущая команда, либо расположен вблизи него. Такое расположение адресов называется *пространственной локальностью программы*. Обработываемые данные, как правило, структурированы, и такие структуры обычно хранятся в последовательных ячейках памяти. Данная особенность программ называется *пространственной локальностью данных*. Кроме того, программы содержат множество небольших циклов и подпрограмм. Это означает, что небольшие наборы команд могут многократно повторяться в течение некоторого интервала времени, то есть имеет место *временная локальность*. Все три вида локальности объединяет понятие *локальность по обращению*. Принцип локальности часто облекают в численную форму и представляют в виде так называемого правила «90/10»: 90% времени работы программы связано с доступом к 10% адресного пространства этой программы.

Из свойства локальности вытекает, что программу разумно представить в виде последовательно обрабатываемых фрагментов — компактных групп команд и данных. Помещая такие фрагменты в более быструю память, можно существенно снизить общие задержки на обращение, поскольку команды и данные, будучи один раз переданы из медленного ЗУ в быстрое, затем могут использоваться многократно, а среднее время доступа к ним в этом случае определяется уже более быстрым ЗУ. Это позволяет хранить большие программы и массивы данных на медленных, емких, но дешевых ЗУ, а в процессе обработки активно использовать сравнительно небольшую быструю память, увеличение емкости которой сопряжено с высокими затратами.

На каждом уровне иерархии информация разбивается на *блоки*, выступающие в качестве наименьшей информационной единицы, пересылаемой между двумя соседними уровнями иерархии. Размер блоков может быть фиксированным либо переменным. При фиксированном размере блока емкость памяти обычно кратна его размеру. Размер блоков на каждом уровне иерархии чаще всего различен и увеличивается от верхних уровней к нижним.

При доступе к командам и данным, например, для их считывания, сначала производится поиск в памяти верхнего уровня. Факт обнаружения нужной информации называют *попаданием* (hit), в противном случае говорят о *промахе* (miss). При промахе производится поиск в ЗУ следующего более низкого уровня, где также возможны попадание или промах. После обнаружении необходимой информации выполняется последовательная пересылка блока, содержащего искомую информацию с нижних уровней на верхние. Следует отметить, что независимо от числа уровней иерархии пересылка информации может осуществляться только между двумя соседними уровнями.

При оценке эффективности подобной организации памяти обычно используют следующие характеристики: *коэффициент попаданий* (hit rate) —

отношение числа обращений к памяти, при которых произошло попадание, к общему числу обращений к ЗУ данного уровня иерархии;

коэффициент промахов (miss rate) — отношение числа обращений к памяти, при которых имел место промах, к общему числу обращений к ЗУ данного уровня иерархии;

-время обращения при попадании (hit time) — время, необходимое для поиска нужной информации в памяти верхнего уровня (включая выяснение, является ли обращение попаданием), плюс время на фактическое считывание данных

-потери на промах (miss penalty) — время, требуемое для замены блока в памяти более высокого уровня на блок с нужными данными, расположенный в ЗУ следующего (более низкого) уровня. Потери на промах включают в себя: *время доступа* (access time) — время обращения к первому слову блока при промахе и *время пересылки* (transfer time) — дополнительное время для пересылки оставшихся слов блока. Время доступа обусловлено задержкой памяти более низкого уровня, в то время как время пересылки связано с полосой пропускания канала между ЗУ двух смежных уровней.

Описание некоторого уровня иерархии ЗУ предполагает конкретизацию четырех моментов:

-размещения блока — допустимого места расположения блока на примыкающем сверху уровне иерархии;

-идентификации блока — способа нахождения блока на примыкающем сверху уровне;

-замещения блока — выбора блока, заменяемого при промахе с целью освобождения места для нового блока;

-согласования копий (стратегии записи) — обеспечения согласованности копии одних и тех же блоков, расположенных на разных уровнях, при записи новой информации в копию, находящуюся на более высоком уровне. Самый быстрый, но и минимальный по емкости тип памяти — это внутренние регистры ЦП, которые иногда объединяют понятием *сверхоперативное запоминающее устройство* — СОЗУ. Как правило, количество регистров невелико, хотя в архитектурах с сокращенным набором команд их число может достигать до нескольких сотен. Основная память (ОП), значительно большей емкости, располагается несколькими уровнями ниже. Между регистрами ЦП и основной памятью часто размещают кэш-память, которая по емкости ощутимо проигрывает ОП, но существенно превосходит последнюю по быстродействию, уступая в то же время СОЗУ. В большинстве современных ВМ имеется несколько уровней кэш-памяти, которые обозначают буквой L и номером уровня кэш-памяти. На рис. 9.2.1 показаны два таких уровня. В последних разработках все чаще появляется также третий уровень кэш-памяти (L3), причем разработчики ВМ говорят о целесообразности введения и четвертого уровня — L4. Каждый последующий уровень кэш-памяти имеет большую емкость, но одновременно и меньшее быстродействие по сравнению с предыдущим. Как бы то ни было, по «скорости» любой уровень кэш-памяти превосходит

основную память. Все виды внутренней памяти реализуются на основе полупроводниковых технологий и в основном, являются энергозависимыми.

Долговременное хранение больших объемов информации (программ и данных) обеспечивается внешними ЗУ, среди которых наиболее распространены запоминающие устройства на базе магнитных и оптических дисков, а также магнитоленточные ЗУ.

Наконец, еще один уровень иерархии может быть добавлен между основной памятью и дисками. Этот уровень носит название дисковой кэш-памяти и реализуется в виде самостоятельного ЗУ, включаемого в состав магнитного диска. Дисковая кэш-память существенно улучшает производительность при обмене информацией между дисками и основной памятью.

Иерархия может быть дополнена и другими видами памяти. Так, некоторые модели ВМ фирмы IBM включают в себя так называемую расширенную память (expanded storage), выполненную на основе полупроводниковой технологии, но имеющую меньшее быстродействие и стоимость по сравнению с ОП. Строго говоря, этот вид памяти не входит в иерархию, а представляет собой ответвление от нее, поскольку данные могут передаваться только между расширенной и основной памятью, но не допускается обмен между расширенной и внешней памятью.

9.3 Основная память

Основная память (ОП) представляет собой единственный вид памяти, к которой ЦП может обращаться непосредственно (исключение составляют лишь регистры Центрального процессора). Информация, хранящаяся на внешних ЗУ, становится доступной процессору только после того, как будет переписана в основную память.

Основную память образуют запоминающие устройства с произвольным доступом. Такие ЗУ образованы как массив ячеек, а «произвольный доступ» означает, что обращение к любой ячейке занимает одно и то же время и может производиться произвольной последовательности. Каждая ячейка содержит фиксированное число поминающих элементов и имеет уникальный адрес, позволяющий различать ячейки при обращении к ним для выполнения операций записи и считывания.

Следствием огромных успехов в области полупроводниковых технологий стало изменение элементной базы основной памяти. На смену ЗУ на базе ферромагнитных колец пришли полупроводниковые микросхемы, использование которых в наши дни стало повсеместным.

Основная память может включать в себя два типа устройств: *оперативные запоминающие устройства* (ОЗУ) и *постоянные запоминающие устройства* (ПЗУ).

Преимущественную долю основной памяти образует ОЗУ, называемое оперативным, потому что оно допускает как запись, так и считывание информации, причем обе операции выполняются однотипно, практически с

одной и той же скоростью, и производятся с помощью электрических сигналов. В англоязычной литературе ОЗУ соответствует аббревиатура RAM — *Random Access Memory*, то есть «память с произвольным доступом», что не совсем корректно, поскольку памятью с произвольным доступом являются также ПЗУ и регистры процессора. Для большинства типов полупроводниковых ОЗУ характерна энергозависимость даже при кратковременном прерывании питания хранимая информация теряется. Микросхема ОЗУ должна быть постоянно подключена к источнику питания и поэтому может использоваться только как временная память.

Вторую группу полупроводниковых ЗУ основной памяти образуют энергонезависимые микросхемы ПЗУ (ROM — *Read-Only Memory*). ПЗУ обеспечивает считывание информации, но не допускает ее изменения (в ряде случаев информация в ПЗУ может быть изменена, но этот процесс сильно отличается от считывания и требует значительно большего времени).

9.4 Блочная организация основной памяти

Емкость основной памяти современных ВМ слишком велика, чтобы ее можно было реализовать на базе единственной интегральной микросхемы (ИМС). Необходимость объединения нескольких ИМС ЗУ возникает также, когда разрядность ячеек в микросхеме ЗУ меньше разрядности слов ВМ.

Увеличение разрядности ЗУ реализуется за счет объединения адресных входов объединяемых ИМС ЗУ. Информационные входы и выходы микросхем являются входами и выходами модуля ЗУ увеличенной разрядности (рис. 9.4.1). Полученную совокупность микросхем называют *модулем памяти*. Модулем можно считать и единственную микросхему, если она уже имеет нужную разрядность. Один или несколько модулей образуют *банк памяти*.

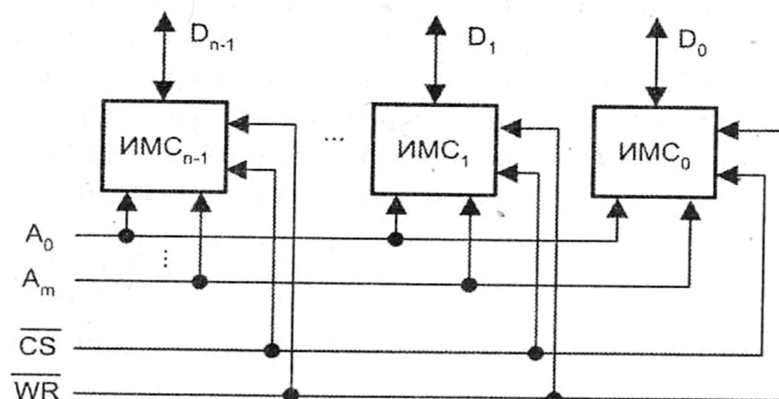


Рис. 5.2. Увеличение разрядности памяти

9.4.1 Увеличение разрядности памяти

Для получения требуемой емкости ЗУ нужно определенным образом объединить несколько банков памяти меньшей емкости. В общем случае основная память ВМ практически всегда имеет блочную структуру, то есть содержит несколько банков.

При использовании блочной памяти, состоящей из B банков, адрес ячейки A преобразуется в пару (b, w) , где b — номер банка, w — адрес ячейки внутри банка. Известны три схемы распределения разрядов адреса A между b и w :

- блочная (номер банка b определяет старшие разряды адреса);
- циклическая ($b = A \bmod B$; $w = A \div B$);
- блочно-циклическая (комбинация двух предыдущих схем).

Рассмотрение основных структур блочной ОП будем проводить на примере памяти емкостью 512 слов (2^9), построенной из четырех банков по 128 слов в каждом. Типовая структура памяти, организованная в соответствии с блочной структурой, показана на рис. 9.4.2.

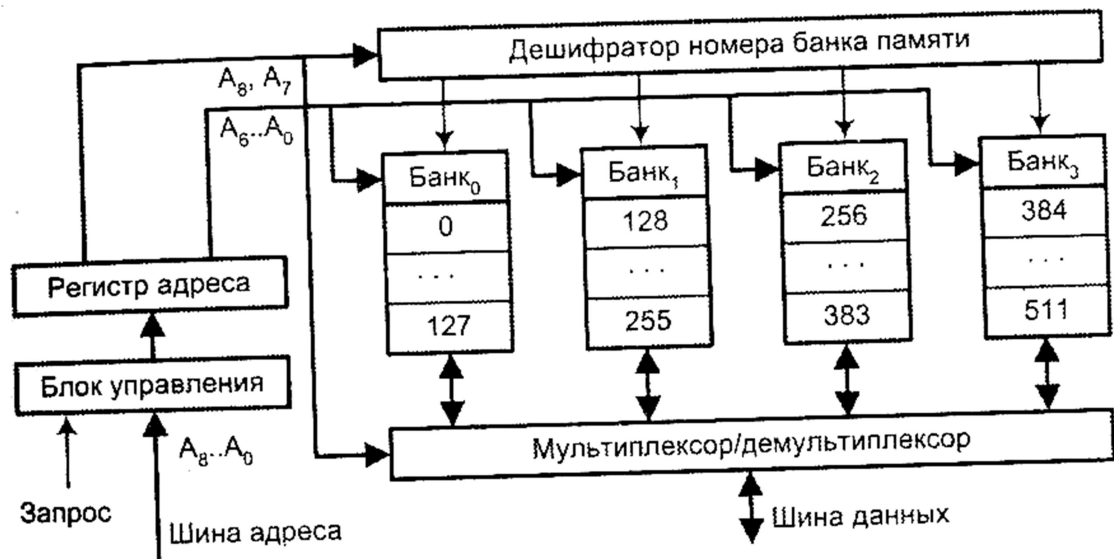


Рис 9.4.2 Структура основной памяти на основе блочной схемы

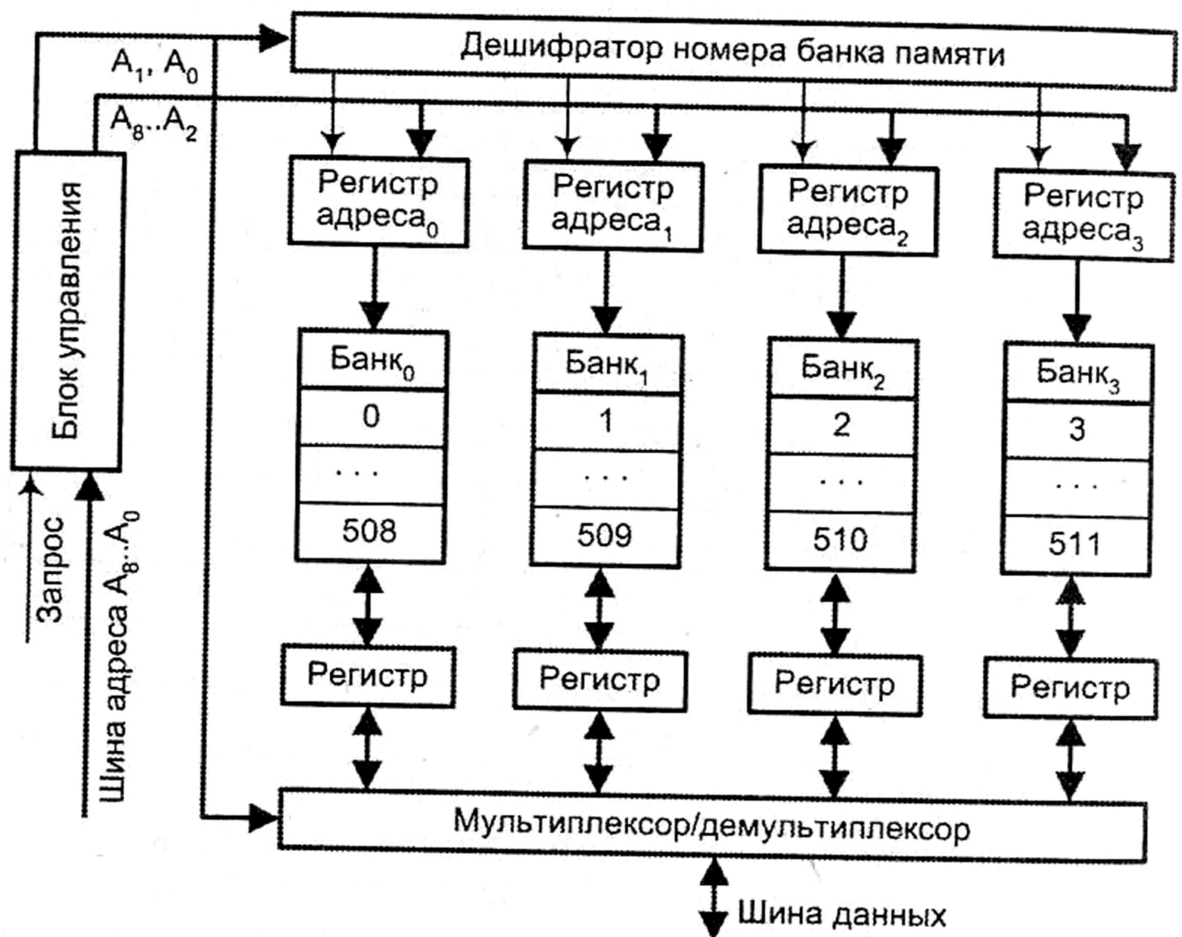
Адресное пространство памяти разбито на группы последовательных адресов, и каждая такая группа обеспечивается отдельным банком памяти. Для обращения к ОП используется 9-разрядный адрес, семь младших разрядов которого (A_6 - A_0) поступают параллельно на все банки памяти и выбирают в каждом из них одну ячейку. Два старших разряда адреса (A_8, A_7) содержат номер банка. Выбор банка обеспечивается либо с помощью дешифратора номера банка памяти, либо путем мультиплексирования информации (на рис. 9.4.2 показаны оба варианта). В функциональном отношении такая ОП может рассматриваться как единое ЗУ, емкость которого равна суммарной емкости составляющих, а быстродействие — быстродействию отдельного банка.

9.5 Расслоение памяти

Помимо податливости к наращиванию емкости, блочное построение памяти обладает еще одним достоинством — позволяет сократить время доступа к информации. Это возможно благодаря потенциальному параллелизму, присущему блочной организации. Большей скорости доступа можно достичь за счет одновременного доступа ко многим банкам памяти.

Одна из используемых для этого методик называется *расслоением памяти*. В ее основе лежит так называемое *чередование адресов* (address interleaving), заключающееся в изменении системы распределения адресов между банками памяти.

Прием чередования адресов базируется на ранее рассмотренном свойстве локальности по обращению, согласно которому последовательный доступ в память обычно производится к ячейкам, имеющим смежные адреса. Иными словами, если в данный момент выполняется обращение к ячейке с адресом 5, то следующее обращение, вероятнее всего, будет к ячейке с адресом 6, затем 7 и т. д. Чередование адресов обеспечивается за счет циклического разбиения адреса. В нашем примере (рис. 9.5.1) для выбора банка используются два младших разряда адреса (A_1, A_0), а для выбора ячейки в банке — 7 старших разрядов (A_8-A_2).



9.5.1 Блочная память с чередованием адресов по циклической схеме

Поскольку в каждом такте на шине адреса может присутствовать адрес только одной ячейки, параллельное обращение к нескольким банкам невозможно, однако оно может быть организовано со сдвигом на один такт. Адрес ячейки запоминается в индивидуальном регистре адреса, и дальнейшие операции по доступу к ячейке в каждом банке протекают независимо. При большом количестве банков среднее время доступа к ОП сокращается почти в B раз (B — количество банков), но при условии, что ячейки, к которым производится последовательное обращение относятся к

разным банкам. Если же запросы к одному и тому же банку следуют друг за другом, каждый следующий запрос должен ожидать завершения обслуживания предыдущего. Такая ситуация называется *конфликтом по доступу*. При частом возникновении конфликтов по доступу метод становится неэффективным.

В блочно-циклической схеме расслоения памяти каждый банк состоит из нескольких модулей, адресуемых по круговой схеме. Адреса между банками распределены по блочной схеме. Таким образом, адрес ячейки разбивается на три части, *n*-ые биты определяют номер банка, следующая группа разрядов адреса указывает на ячейку в модуле, а младшие биты адреса выбирают модуль в банке. Схему иллюстрирует рис.9.5.2

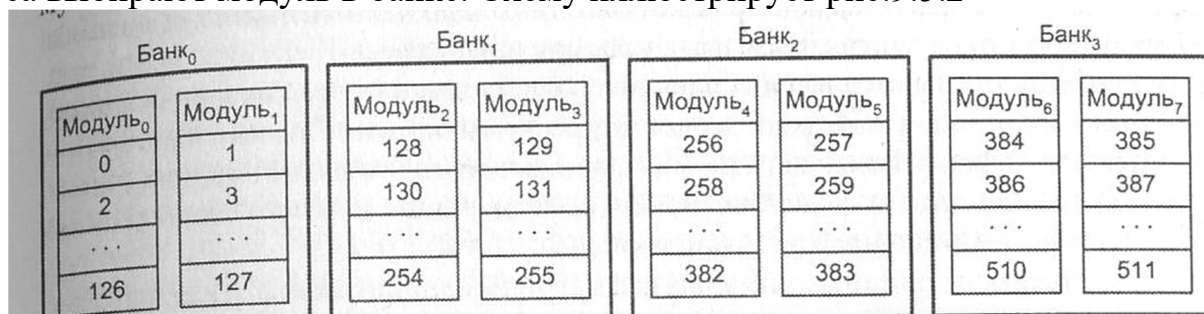


Рис. 9.5.2 Блочно-циклическая схема расслоения памяти

Традиционные способы расслоения памяти хорошо работают в рамках одной задачи, для которой характерно свойство локальности. В многопроцессорных системах с общей памятью, где запросы на доступ к памяти достаточно независимы, не исключен иной подход, который можно рассматривать как развитие идеи расслоения памяти. Для этого в систему включают несколько контроллеров памяти, что позволяет отдельным банкам работать совершенно автономно. Эффективность данного приема зависит от частоты независимых обращений к разным банкам. Лучшего результата можно ожидать при большом числе банков, что уменьшает вероятность последовательных обращений к одному и тому же банку памяти. Так, в суперкомпьютере NEC SX/3 основная память состоит из 128 банков.

9.6 Организация микросхем памяти

Интегральные микросхемы (ИМС) памяти организованы в виде матрицы ячеек, каждая из которых, в зависимости от разрядности ИМС, состоит из одного или более запоминающих элементов (ЗЭ) и имеет свой адрес. Каждый ЗЭ способен хранить один бит информации. Для ЗЭ любой полупроводниковой памяти характерны следующие свойства:

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

При матричной организации ИМС памяти (рис. 9.6.1) реализуется координатный принцип адресации ячеек. Адрес ячейки, поступающий по шине адреса ВМ, пропускается через логику выбора, где он разделяется на две составляющие: адрес строки и адрес столбца. Адреса строки и столбца запоминаются соответственно в регистре адреса строки и регистре адреса столбца микросхемы. Регистры соединены каждый со своим дешифратором. Выходы дешифраторов образуют систему горизонтальных и вертикальных линий, к которым подсоединены запоминающие матрицы, при этом каждый ЗЭ расположен на пересечении одной горизонтальной и одной вертикальной линии.

ЗЭ, объединенные общим «горизонтальным» проводом, принято называть *строкой* (row). Запоминающие элементы, подключенные к общему «вертикальному» проводу, называют *столбцом* (column).

Совокупность запоминающих элементов и логических схем, связанных с выбором строк и столбцов, называют *ядром* микросхемы памяти. Помимо ядра в ИМСИ имеется еще интерфейсная логика, обеспечивающая взаимодействие ядра с внешним миром. В ее задачи, в частности, входят коммутация нужного столбца на выход при считывании и на вход — при записи.

На физическую организацию ядра, как матрицы однобитовых ЗЭ, накладывается логическая организация памяти, под которой понимается разрядность микросхемы, то есть количество линий ввода/вывода. Разрядность микросхемы определяет количество ЗЭ, имеющих один и тот же адрес (такая совокупность запоминающих элементов называется *ячейкой*), то есть каждый столбец содержит столько разрядов, сколько есть линий ввода/вывода данных.

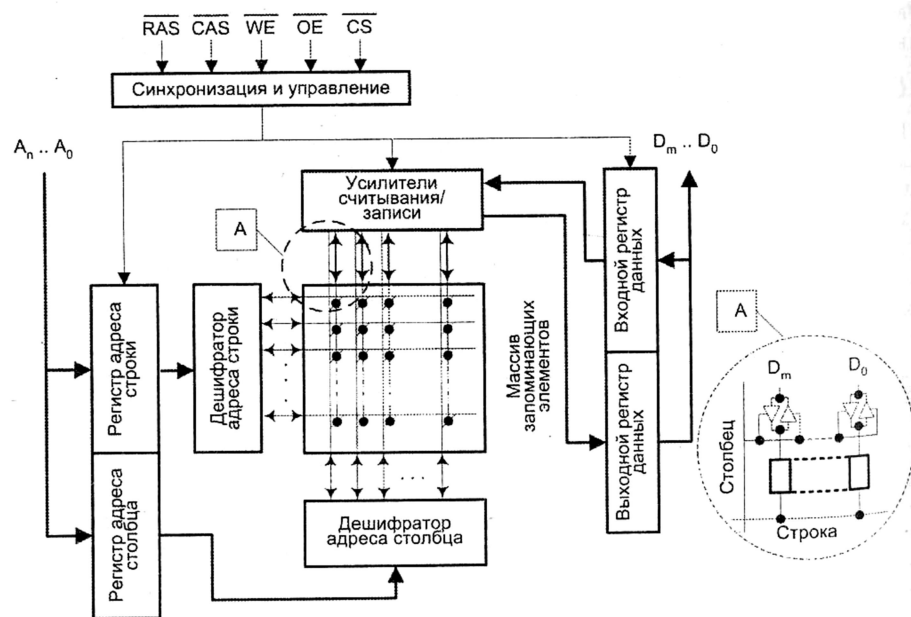


Рис. 9.6.1 Структура микросхемы памяти

Для синхронизации процессов фиксации и обработки адресной информации внутри ИМС адрес строки (RA) сопровождается сигналом RAS (Row Address Strobe — строб строки), а адрес столбца (CA) — сигналом CAS

(Column Address Strobe — строб столбца). Вторую букву в аббревиатурах RAS и CAS иногда расшифровывают как Access — «доступ», то есть имеется строб доступа к строке и строб доступа к столбцу.

Сигнал выбора микросхемы CS (Crystal Select) разрешает работу ИМС и используется для выбора определенной микросхемы в системах, состоящих из нескольких ИМС. Вход WE (Write Enable — разрешение записи) определяет вид выполняемой операции (считывание или запись).

Записываемая информация, поступающая по шине данных, первоначально заносится во входной регистр данных, а затем — в выбранную ячейку. При выполнении операции чтения информация из ячейки до ее выдачи на шину данных буферизируется в выходном регистре данных. Обычно роль входного и выходного выполняет один и тот же регистр. Усилители считывания/записи (УСЗ) служат для электрического согласования сигналов на линиях данных и внутренних сигналов ИМС. Обычно число УСЗ равно числу запоминающих элементов в строке матрицы, и все они при обращении к памяти подключаются к выбранной горизонтальной линии. Каждая группа УСЗ, образующая ячейку, подключена к одному из столбцов матрицы, то есть выбор нужной ячейки в строке обеспечивается активизацией одной из вертикальных линий. На все время пока ИМС памяти не использует шину данных, информационные выходы микросхемы переводятся в третье (высокоимпедансное) состояние. Управление переключением в третье состояние обеспечивается сигналом OE (Output Enable — разрешение выдачи выходных сигналов). Этот сигнал активизируется при выполнении операции чтения.

Управление операциями с основной памятью осуществляется контроллером памяти. Обычно этот контроллер входит в состав центрального процессора либо реализуется в виде внешнего по отношению к памяти устройства. В последних типах ИМС памяти часть функций контроллера возлагается на микросхему памяти. Хотя работа ИМС памяти может быть организована как по синхронной, так и по асинхронной схеме, контроллер памяти — устройство синхронное, то есть срабатывающее исключительно по тактовым импульсам. По этой причине операции с памятью принято описывать с привязкой к тактам. В общем случае на каждую такую операцию требуется как минимум пять тактов, которые используются следующим образом:

1. Указание типа операции (чтение или запись) и установка адреса строки.
2. Формирование сигнала RAS.
3. Установка адреса столбца.
4. Формирование сигнала CAS.
5. Возврат сигналов RAS и CAS в неактивное состояние.

Данный перечень учитывает далеко не все необходимые действия, например регенерацию содержимого памяти в динамических ОЗУ.

Возможности «ускорения» ядра микросхемы ЗУ весьма ограничены и связаны в основном с миниатюризацией запоминающих элементов.

Наибольшие успехи достигнуты в интерфейсной части ИМС, касаются они, главным образом, операции чтения, то есть способов доставки содержимого ячейки на шину данных. Наибольшее распространение получили следующие шесть фундаментальных подходов

- :- последовательный;
- конвейерный;
- регистровый;
- страничный;
- пакетный;
- удвоенной скорости.

Последовательный режим

При использовании *последовательного режима* (Flow through Mode) адрес и управляющие сигналы подаются на микросхему до поступления синхроимпульса.

В момент прихода синхроимпульса вся входная информация запоминается во внутренних регистрах — по его переднему фронту, и начинается цикл чтения. Через некоторое время, но в пределах того же цикла данные появляются на внешней шине, причем момент этот определяется только моментом прихода синхронизирующего импульса и скоростью внутренних цепей микросхемы.

Конвейерный режим

Конвейерный режим (pipelined mode) — это такой метод доступа к данным, при котором можно продолжать операцию чтения по предыдущему адресу в процессе запроса по следующему.

При чтении из памяти время, требуемое для извлечения данных из ячейки, можно условно разбить на два интервала. Первый из них — непосредственно доступ к массиву запоминающих элементов и извлечение данных из ячейки. Второй — передача данных на выход (при этом происходит детектирование состояния ячейки, усиление сигнала и другие операции, необходимые для считывания информации). В отличие от последовательного режима, где следующий цикл чтения начинается только по окончании предыдущего, в конвейерном режиме процесс разбивается на два этапа. Пока данные из предыдущего цикла чтения передаются на внешнюю шину, происходит запрос на следующую операцию чтения. Таким образом, два цикла чтения перекрываются во времени. Из-за усложнения схемы передачи данных на внешнюю шину время считывания увеличивается на один такт, и данные поступают на выход только в следующем такте, но такое запаздывание наблюдается лишь при первом чтении в последовательности операций считывания из памяти. Все последующие данные поступают на выход друг за другом, хотя и с запаздыванием на один такт относительно запроса на чтение. Так как циклы чтения перекрываются, микросхемы с конвейерным режимом могут использоваться при частотах шины, вдвое превышающих допустимую для ИМС с последовательным режимом чтения.

Регистровый режим

Регистровый режим (Register to Latch) используется относительно редко и отличается наличием регистра на выходе микросхемы. Адрес и управляющие сигналы выдаются на шину до поступления синхронизирующего импульса. С приходом положительного фронта синхроимпульса адрес записывается во внутренний регистр микросхемы, и начинается цикл чтения. Считанные данные заносятся в промежуточный выходной регистр и хранятся там до появления отрицательного фронта (спада) синхроимпульса, а с его поступлением передаются на шину. Метод однозначно определяет момент появления данных на выходе ИМС, причем изменяя ширину импульса синхронизации можно менять время появления данных на Шине. Данное свойство часто оказывается весьма полезным при проектировании специализированных ВМ. По быстродействию микросхемы с регистровым режимом идентичны ИМС с последовательным режимом.

Страничный режим

В основе идеи лежит тот факт, что при доступе к ячейкам со смежными адресами (согласно принципу локальности такая ситуация наиболее вероятна), причем к таким, где все ЗЭ расположены в одной строке матрицы, доступ ко второй и последующим ячейкам можно производить существенно быстрее. Действительно если адрес строки при очередном обращении остался прежним, то все временные затраты, связанные с повторным занесением адреса строки в регистр ИМС, дешифровкой, зарядом паразитной емкости горизонтальной линии и т. п., можно исключить. Для доступа к очередной ячейке достаточно подавать на ИМС лишь адрес нового столбца, сопровождая его сигналом CAS. Отметим, что обращение к первой ячейке в последовательности производится стандартным образом — поочередным заданием адреса строки и адреса столбца, то есть здесь время доступа уменьшить практически невозможно. Рассмотренный режим называется *режимом страничного доступа* или просто *страничным режимом* (Page Mode). Под страницей понимается строка матрицы ЗЭ. Микросхемы, где реализуется страничный режим и его модификации, принято характеризовать формулой $x-y-y-y$. Первое число x представляет количество тактов системной шины, необходимое для доступа к первой ячейке последовательности, а y — к каждой из последующих ячеек. Так, выражение 7-3-3-3 означает, что для обработки первого слова необходимо 7 тактовых периодов системной шины (в течение шести из которых шина простаивает в ожидании), а для обработки последующих слов — по три периода, из которых два системная шина также простаивает.

Режим быстрого страничного доступа

Режим быстрого страничного доступа (FPM — Fast Page Mode) представляет собой модификацию стандартного страничного режима. Основное отличие заключается в способе занесения новой информации в регистр адреса столбца. Полный адрес (строки и столбца) передается только при первом обращении к строке. Активизация буферного регистра адреса столбца производится не по сигналу CAS, а по заднему фронту сигнала RAS. Сигнал RAS остается активным на протяжении всего страничного цикла и

позволяет заносить в регистр адреса столбца новую информацию не по спадающему фронту CAS, а как только адрес на входе ИМС стабилизируется, то есть практически по переднему фронту сигнала CAS. В целом же потери времени сокращаются на два такта, которые ранее требовались для передачи адреса каждой строки и сигнала RAS. Реальный выигрыш, однако, наблюдается лишь при передаче блоков данных, хранящихся в одной и той же строке микросхемы. Если же программа часто обращается к разным областям памяти, переходя с одной строки ИМС на другую, преимущества метода теряются. Режим нашел широкое применение в микросхемах ОЗУ, особенно динамического типа.

Пакетный режим

Пакетный режим (Burst Mode) — режим, при котором на запрос по конкретному адресу память возвращает пакет данных, хранящихся не только по этому адресу, но и по нескольким последующим адресам.

Разрядность ячейки памяти современных ВМ обычно равна одному байту, в то время как ширина шины данных, как правило, составляет четыре байта. Следовательно, одно обращение к памяти требует последовательного доступа к четырем смежным ячейкам — пакету¹. С учетом этого обстоятельства в ИМС памяти часто используется модификация страничного режима, носящая название *группового* или *пакетного режима*. При его реализации адрес столбца заносится в ИМС только для первой ячейки пакета, а переход к очередному столбцу производится уже внутри микросхемы. Это позволяет для каждого пакета исключить три из четырех операций занесения адреса столбца и тем самым еще более сократить среднее время доступа.

Режим удвоенной скорости

Важным этапом в дальнейшем развитии технологии микросхем памяти стал режим DDR (Double Data Rate) — удвоенная скорость передачи данных. Сущность метода заключается в передаче данных по обоим фронтам импульса синхронизации, то есть дважды за период. Таким образом, пропускная способность увеличивается в те же два раза.

Помимо упомянутых используются и другие приемы повышения быстродействия ИМС памяти, такие как включение в состав микросхемы вспомогательной кэш-памяти и независимые тракты данных, позволяющие одновременно производить обмен с шиной данных и обращение к матрице ЗЭ и т. д.