

Лекция 6. Подсистема памяти МПС

Подсистема памяти является необходимым компонентом любой микропроцессорной системы. Память используется как для хранения команд, так и данных. Функции памяти обеспечиваются запоминающими устройствами (ЗУ), предназначенными для фиксации, хранения и выдачи информации в процессе работы микропроцессорной системы. Процесс фиксации информации в ЗУ называется записью, процесс выдачи информации – чтением или считыванием, а совместно их определяют как процессы обращения к ЗУ.

Характеристики запоминающих устройств

При рассмотрении запоминающих устройств учитывают следующие характеристики:

- место расположения;
- емкость;
- единица пересылки;
- метод доступа;
- быстродействие;
- физический тип;
- стоимость.

Характеристики запоминающих устройств

По месту расположения ЗУ разделяют на:

- процессорные,
- внутренние
- внешние.

Наиболее скоростные виды памяти – регистры центрального процессора и кэш-память первого уровня (размещена на общем кристалле с процессором). Внутреннюю память образуют ЗУ, расположенные на системной плате. К ней относят основную память, а также кэш-память второго и последующих уровней. К внешней памяти относят ЗУ большой емкости – магнитные и оптические диски, магнитные ленты, так как к ядру микропроцессорной системы они подключаются аналогично устройствам ввода/вывода.

Характеристики запоминающих устройств

Емкость ЗУ характеризуют числом битов или байтов, которое может храниться в запоминающем устройстве.

Важной характеристикой ЗУ является **единица пересылки**. Для основной памяти единица пересылки определяется шириной шины данных, т.е. 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 называемые блоками.

Различают четыре основных **метода доступа**:

- **последовательный доступ**, при котором ЗУ ориентировано на хранение информации в виде последовательности блоков данных, называемых записями. 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**, при котором каждая запись имеет адрес, отражающий ее физическое размещение на носителе информации. Обращение осуществляется как адресный доступ к началу записи, с последующим доступом к определенной единице информации внутри записи. Время доступа является величиной переменной. Такой режим характерен для магнитных дисков.

- **произвольный доступ**, каждая ячейка памяти имеет физический адрес. Обращение к любой ячейке занимает одно и то же время и может производиться в произвольной очередности. Примером могут служить ЗУ основной памяти.

- **ассоциативный доступ** позволяет выполнить поиск необходимых ячеек при совпадении отдельных битов в представлении адреса. По этому принципу построены блоки кэш-памяти.

Быстродействие ЗУ определяется параметрами:

1) время доступа соответствует времени от момента поступления адреса до момента, когда данные заносятся в память или становятся доступными.

2) длительность цикла памяти или период обращения .

Понятие применяется к памяти с произвольным доступом, для которой оно означает минимальное время между двумя последовательными обращениями к памяти.

3) скорость передачи – это скорость, с которой данные могут передаваться в память или из нее. Скорость передачи определяется соотношением

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

T_N - среднее время считывания или записи N битов

T_A - среднее время доступа;

R – скорость пересылки в битах в секунду

R – скорость пересылки в битах в секунду.

Характеристики запоминающих устройств

Физический тип ЗУ определяется технологией изготовления. Применяются три наиболее распространенных технологии: полупроводниковая память, память с магнитным носителем информации, память с оптическим носителем.

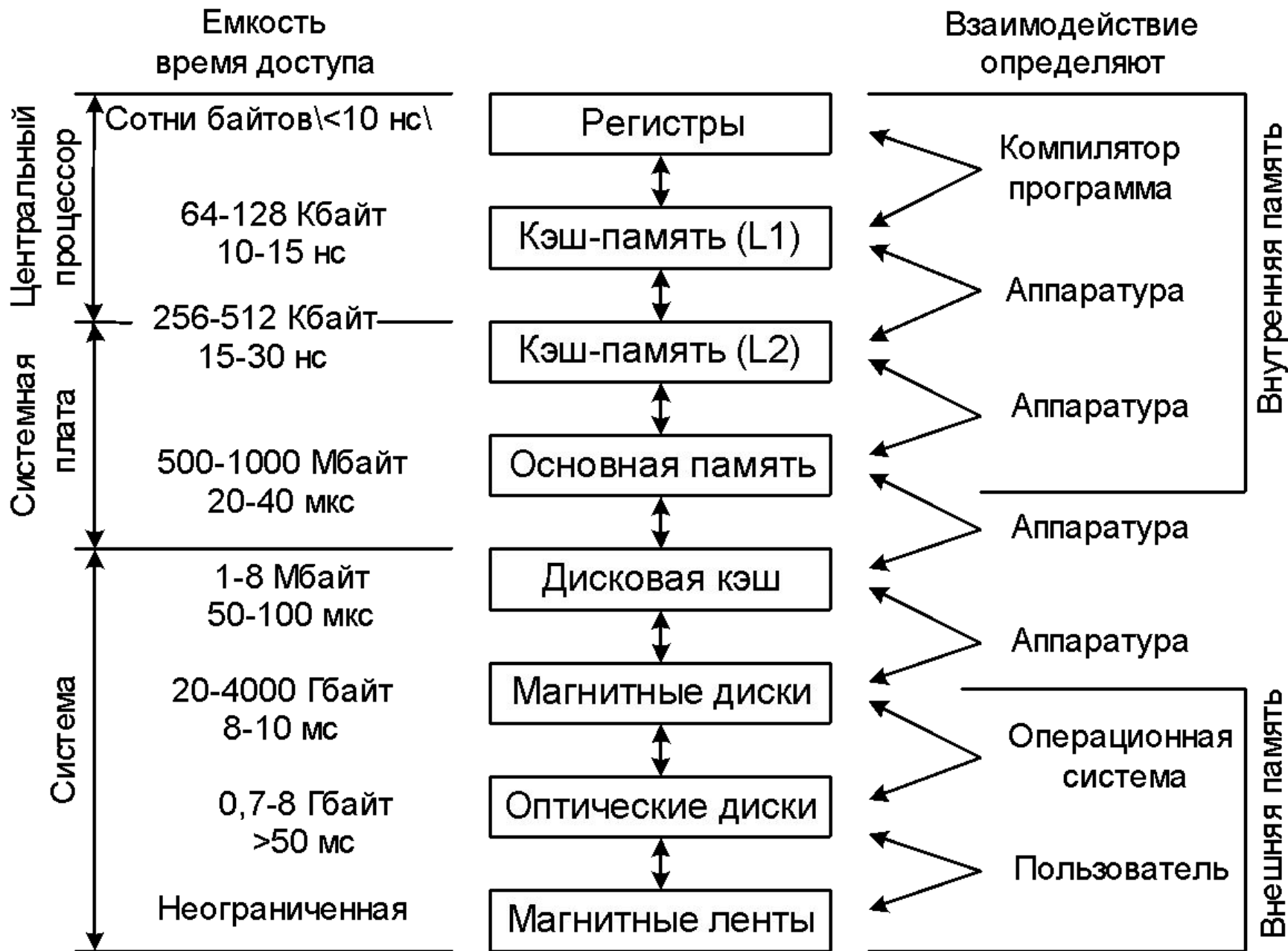
Стоимость ЗУ оценивается стоимостью хранения одного бита информации.

Иерархическая архитектура памяти

Быстродействие микропроцессорных систем в сильной степени зависит от быстродействия подсистемы памяти. Разрыв в быстродействии процессора и памяти возрастает. В настоящее время существует следующая закономерность: чем меньше время доступа, тем выше стоимость хранения бита; чем больше емкость, тем больше время доступа и ниже стоимость хранения бита.

При создании системы памяти приходится решать задачу обеспечения требуемой емкости и высокого быстродействия за приемлемую цену. Разработчики используют более быстродействующую и, соответственно более дорогую память в наиболее “ответственных” местах системы.

Иерархическая архитектура памяти



Иерархическая архитектура памяти

Иерархическая архитектура состоит из запоминающих устройств различного типа, которые, в зависимости от характеристик, относят к определенному уровню иерархии. Более высокий уровень соответствует памяти с большим быстродействием, меньшей емкости, большей стоимости за бит, чем более низкий уровень. Уровни иерархии взаимосвязаны: все данные на одном уровне могут быть также найдены на более низком уровне. Четыре верхних уровня иерархии образуют внутреннюю память, а все нижние – это внешняя или вторичная память.

Иерархическая архитектура памяти

По мере движения вниз по иерархической структуре:

- возрастает емкость
- возрастает время доступа
- уменьшается стоимость хранения бита
- уменьшается частота обращения со стороны центрального процессора.

Самый быстрый, но и минимальный по емкости тип памяти – это внутренние регистры центрального процессора, которые иногда объединяют понятием свероперативное запоминающее устройство – СОЗУ. Как правило, количество регистров невелико, хотя в архитектурах с сокращенным набором команд их число может достигать до нескольких сотен.

Иерархическая архитектура памяти

Основная память, значительно большей емкости, располагается несколькими уровнями ниже. Между регистрами ЦП и основной памятью размещают кэш-память, которая по емкости значительно меньше основной, но существенно превосходит последнюю по быстродействию.

В большинстве современных вычислительных машин имеется несколько уровней кэш-памяти (2-3). Каждый последующий уровень кэш-памяти имеет большую емкость, но одновременно и меньшее быстродействие. Все виды внутренней памяти реализуются на полупроводниковой памяти и в основном являются энергозависимыми.

Иерархическая архитектура памяти

Иерархия может быть дополнена и другими видами памяти. Некоторые модели вычислительных машин включают так называемую расширенную память, выполненную на основе полупроводниковой технологии, но имеющую меньшее быстродействие и стоимость по сравнению с основной памятью. Этот вид памяти представляет собой ответвление от иерархии, так как данные могут передаваться только между расширенной и основной памятью, но не допускается обмен между расширенной и внешней памятью.

Основная память

Основная память представляет собой единственный вид памяти, к которой ЦП может обращаться непосредственно. Информация, хранящаяся на внешних ЗУ, становится доступной процессору только после того, как будет переписана в основную память.

Основную память образуют ЗУ 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 в произвольной последовательности. Каждая ячейка содержит фиксированное число запоминающих элементов и имеет уникальный адрес, позволяющий различать ячейки при обращении к ним для выполнения операций записи и считывания.

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

Большую часть основной памяти составляет оперативная память, называемая так, потому что ОЗУ допускает как запись, так и считывание информации, причем обе операции выполняются односторонне, практически с одной и той же скоростью, и производятся с помощью электрических сигналов. ОЗУ соответствует аббревиатура RAM – Random Access Memory. Для большинства типов полупроводниковых ОЗУ характерна энергозависимость, т.е. даже при кратковременном прерывании питания хранимая информация теряется.

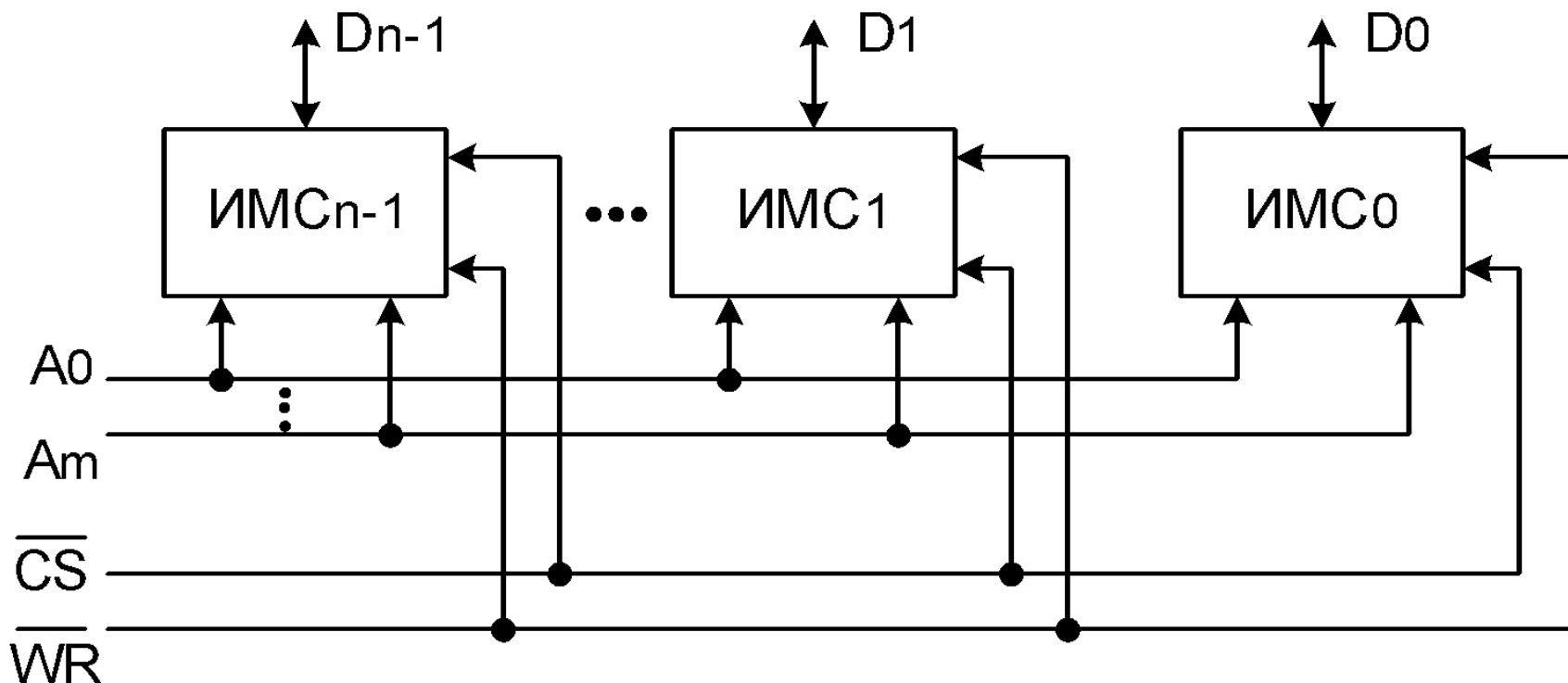
Вторую группу полупроводниковых запоминающих устройств основной памяти образуют энергонезависимые микросхемы ПЗУ (ROM – Read-Only Memory). ПЗУ обеспечивает считывание информации, но не допускает ее изменения в процессе функционирования микропроцессорной системы.

Блочная организация основной памяти

Для реализации основной памяти большого объема (что актуально в современных вычислительных машинах) возникает необходимость объединения нескольких интегральных микросхем запоминающих устройств. Объединение ЗУ также необходимо если разрядность ячеек ЗУ меньше разрядности слов вычислительной машины.

Увеличение разрядности ЗУ реализуется за счет объединения адресных входов объединяемых интегральных микросхем запоминающих устройств. Информационные входы и выходы микросхем являются входами и выходами модуля ЗУ увеличенной разрядности

Увеличение разрядности памяти



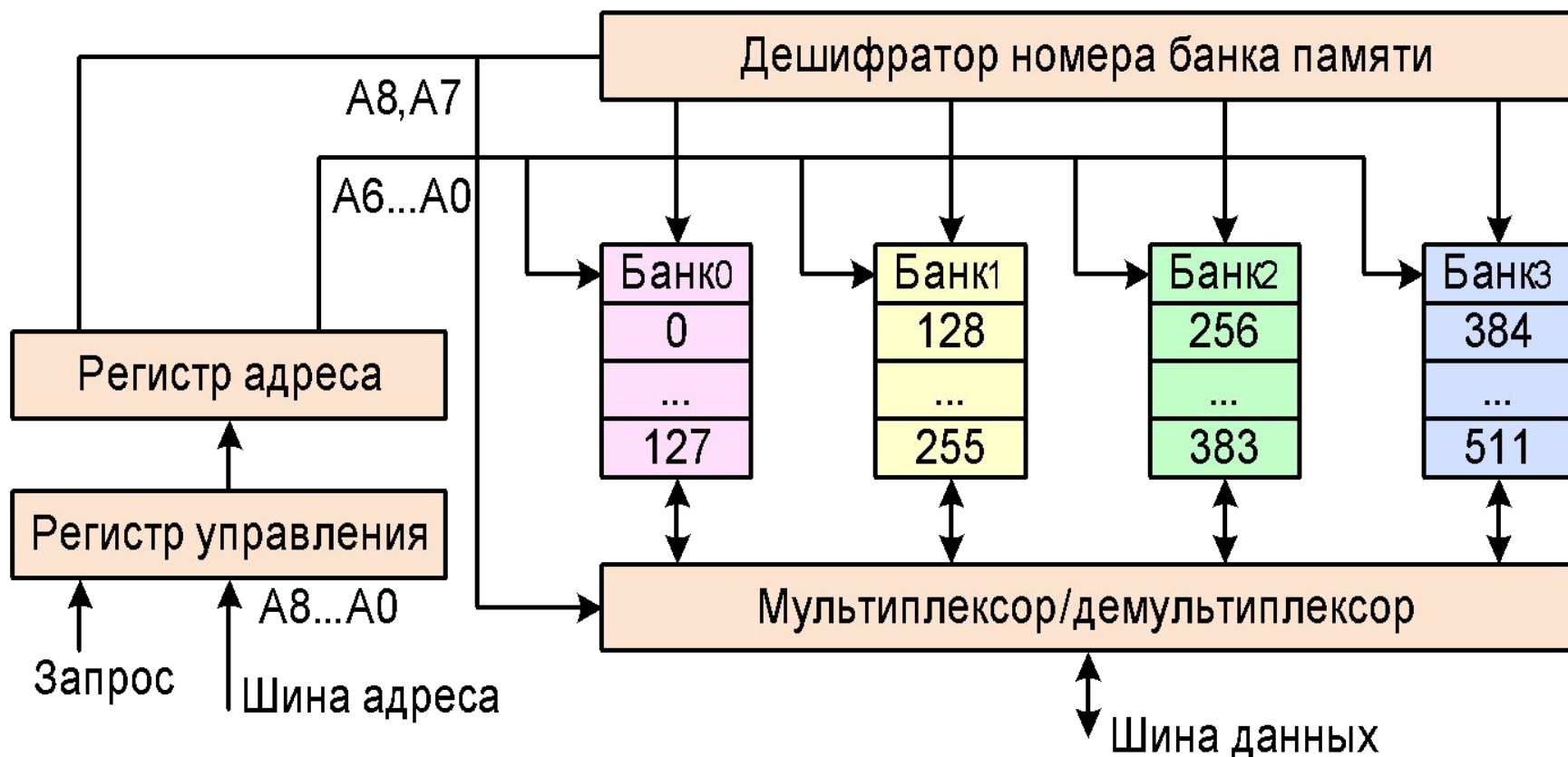
Полученную совокупность микросхем называют модулем памяти. Модулем можно считать и единственную микросхему, если она уже имеет нужную разрядность. Один или несколько модулей образуют банк памяти.

Для получения требуемой емкости нужно определенным образом объединить несколько банков памяти меньшей емкости. В общем случае основная память микропроцессорной системы всегда имеет блочную структуру, т.е. 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 определяет старшие разряда адреса);
- циклическая ($b=A \bmod B$; $w=A \div B$);
- блочно-циклическая (комбинация двух предыдущих схем).

Структура памяти на основе блочной схемы

Рассмотрим блочную структуру памяти на примере памяти емкостью 512 слов, построенной из четырех банков по 128 слов в каждом.



Адресное пространство памяти разбито на группы последовательных адресов, и каждая такая группа обеспечивается отдельным банком памяти. Для обращения к основной памяти используется 9-разрядный адрес, семь младших разрядов которого (A6-A0) поступают параллельно на все банки памяти и выбирают в каждом из них одну ячейку. Два старших разряда адреса (A8, A7) содержат номер банка.

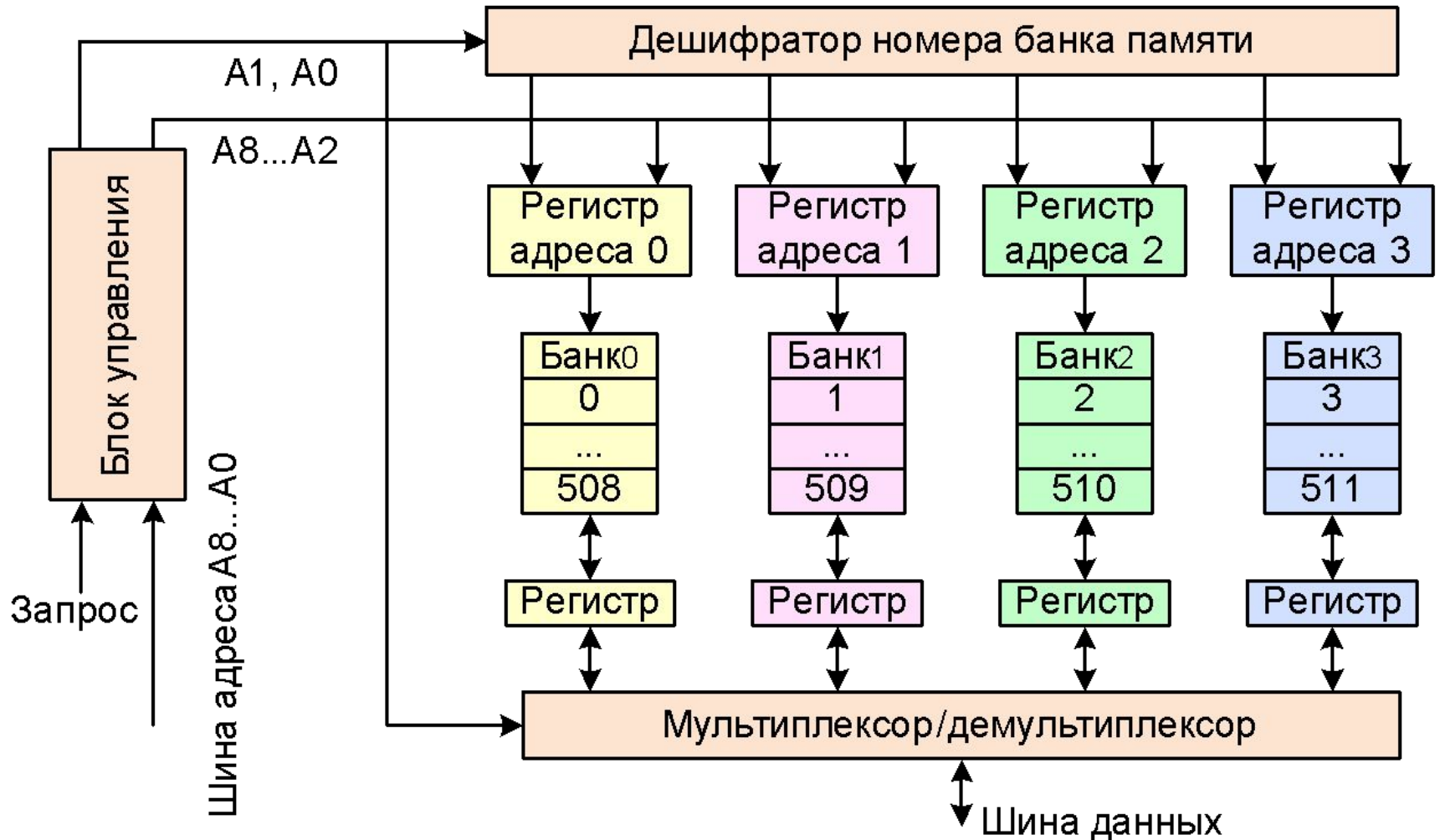
Выбор банка обеспечивается либо с помощью дешифратора номера банка памяти, либо путем мультиплексирования информации (на рисунке 3 показаны оба варианта). В функциональном отношении такая основная память может рассматриваться как единое ЗУ, емкость которого равна суммарной емкости составляющих, а быстродействие - быстродействию отдельного банка.

Расслоение памяти

Блочное построение памяти позволяет не только наращивать емкость, но и сократить время доступа к информации. Это возможно благодаря параллелизму, присущему блочной организации. Скорость доступа увеличивается за счет одновременного доступа ко многим банкам памяти. Одна из используемых для этого методик называется расслоением памяти. В ее основе лежит чередование адресов (address interleaving), которое заключается в изменении системы распределения адресов между банками памяти.

Прием чередования адресов базируется на свойстве локальности по обращению, согласно которому последовательный доступ в память обычно производится к ячейкам, имеющим смежные адреса. Т.е., если в данный момент выполняется обращение к ячейке с адресом 5, то следующее обращение будет к ячейке с адресом 6, затем 7 и т.д. Чередование адресов обеспечивается за счет циклического разбиения адреса.

Блочная память с чередованием адресов по циклической схеме



Рассмотрим пример блочной памяти с чередованием адресов по циклической схеме (рисунок 4), в котором для выбора банка используются два младших разряда адреса (A_1, A_0), а для выбора ячейки в банке – 7 старших разрядов ($A_8 - A_2$). В каждом такте на шине адреса может присутствовать адрес только одной ячейки, поэтому параллельное обращение к нескольким банкам невозможно, но 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 памяти сокращается почти в B раз (B – количество банков), но при условии, что ячейки относятся к разным банкам. Если же запросы к одному тому же банку следуют друг за другом, каждый следующий запрос должен ожидать завершения обслуживания предыдущего. Такая ситуация называется конфликтом по доступу. При частых конфликтах данный метод становится неэффективным.

Блочно-циклическая схема расслоения памяти

Банк0

Модуль0	Модуль1
0	1
2	3
...	...
126	127

Банк1

Модуль2	Модуль3
128	129
130	131
...	...
254	255

Банк2

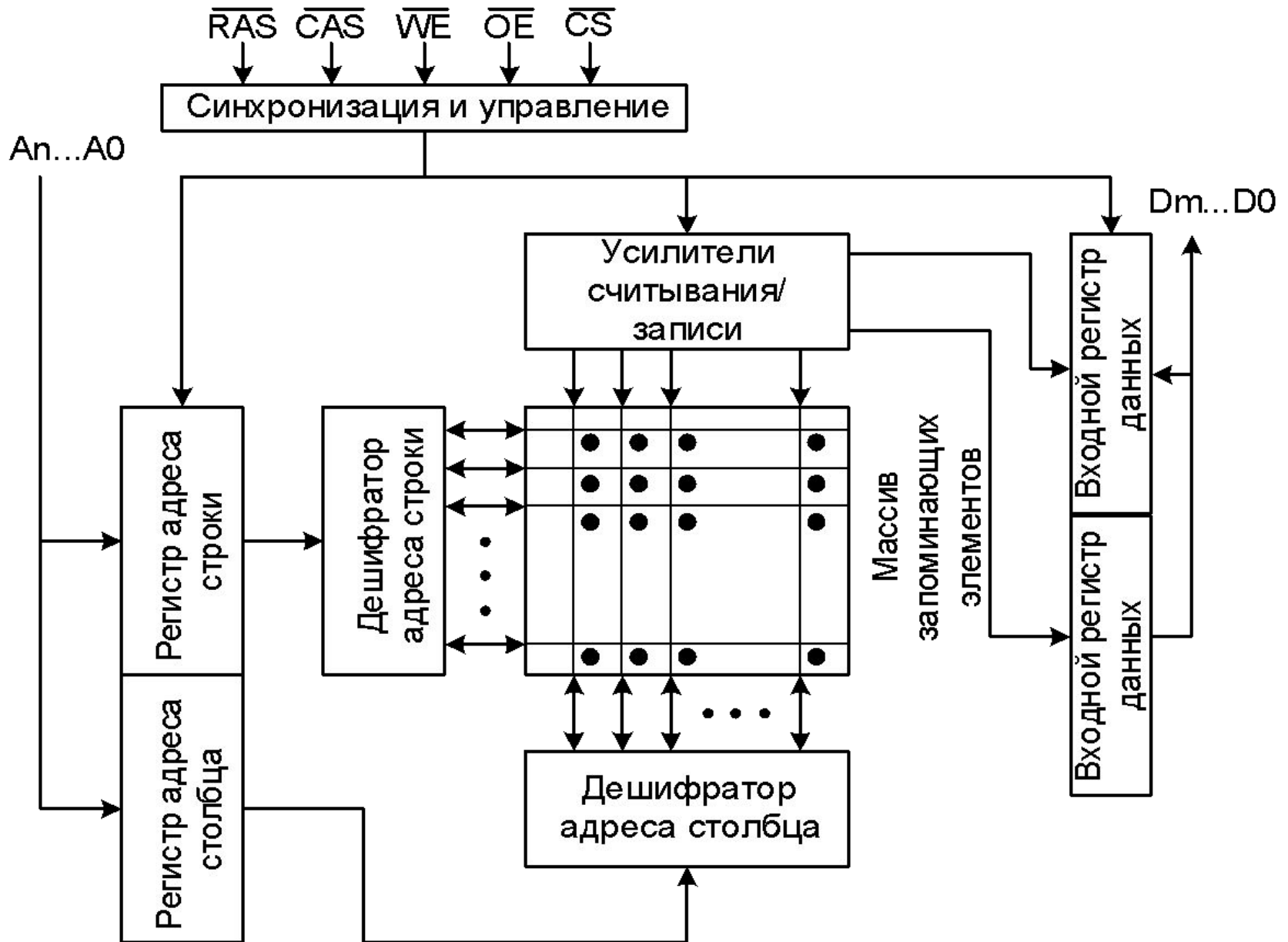
Модуль4	Модуль5
256	257
258	259
...	...
382	383

Банк3

Модуль6	Модуль7
384	385
386	387
...	...
510	511

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. Старшие биты определяют номер банка, следующая группа разрядов адреса указывает на ячейку в модуле, а младшие биты адреса выбирают модуль в банке. Блочно-циклическая схема расслоения памяти показана на рисунке

Структура микросхемы 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 и имеет свой адрес. Каждый запоминающий элемент способен хранить один бит информации. Для запоминающего элемента любой полупроводниковой памяти характерны следующие свойства:

- два стабильных состояния, двоичные 0 и 1;
- в запоминающий элемент может быть произведена запись информации, посредством перевода его в одно из двух возможных состояний;
- содержимое запоминающего элемента может быть считано.

При матричной организации ИМС памяти реализуется координатный принцип адресации ячеек

Адрес ячейки, поступающий по шине адреса, 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 элементы матрицы, при этом каждый запоминающий элемент расположен на пересечении одной горизонтальной и одной вертикальной линии. Запоминающие элементы, объединенные общим «горизонтальным» проводом, называют строкой (row). Запоминающие элементы, объединенные общим «вертикальным» проводом, называют столбцом (column).

Совокупность запоминающих элементов и логических схем, связанных с выбором строк и столбцов, называют ядром микросхемы памяти. Кроме ядра имеется интерфейсная логика, обеспечивающая взаимодействие ядра с внешним миром. В ее задачи входит коммутация нужного столбца на выход при считывании и на вход при записи.

Разрядность микросхемы определяет количество запоминающих элементов, имеющих один и тот же адрес (такая совокупность ЗЭ называется ячейкой), т.е. каждый столбец содержит столько разрядов, сколько есть линий ввода/вывода данных.

Для уменьшения числа контактов ИМС адреса строки и столбца подаются в микросхему через одни и те же контакты последовательно во времени (мультиплексируются) и запоминаются соответственно в регистре адреса строки и регистре адреса столбца микросхемы. Мультиплексирование обычно реализуется внешней схемой.

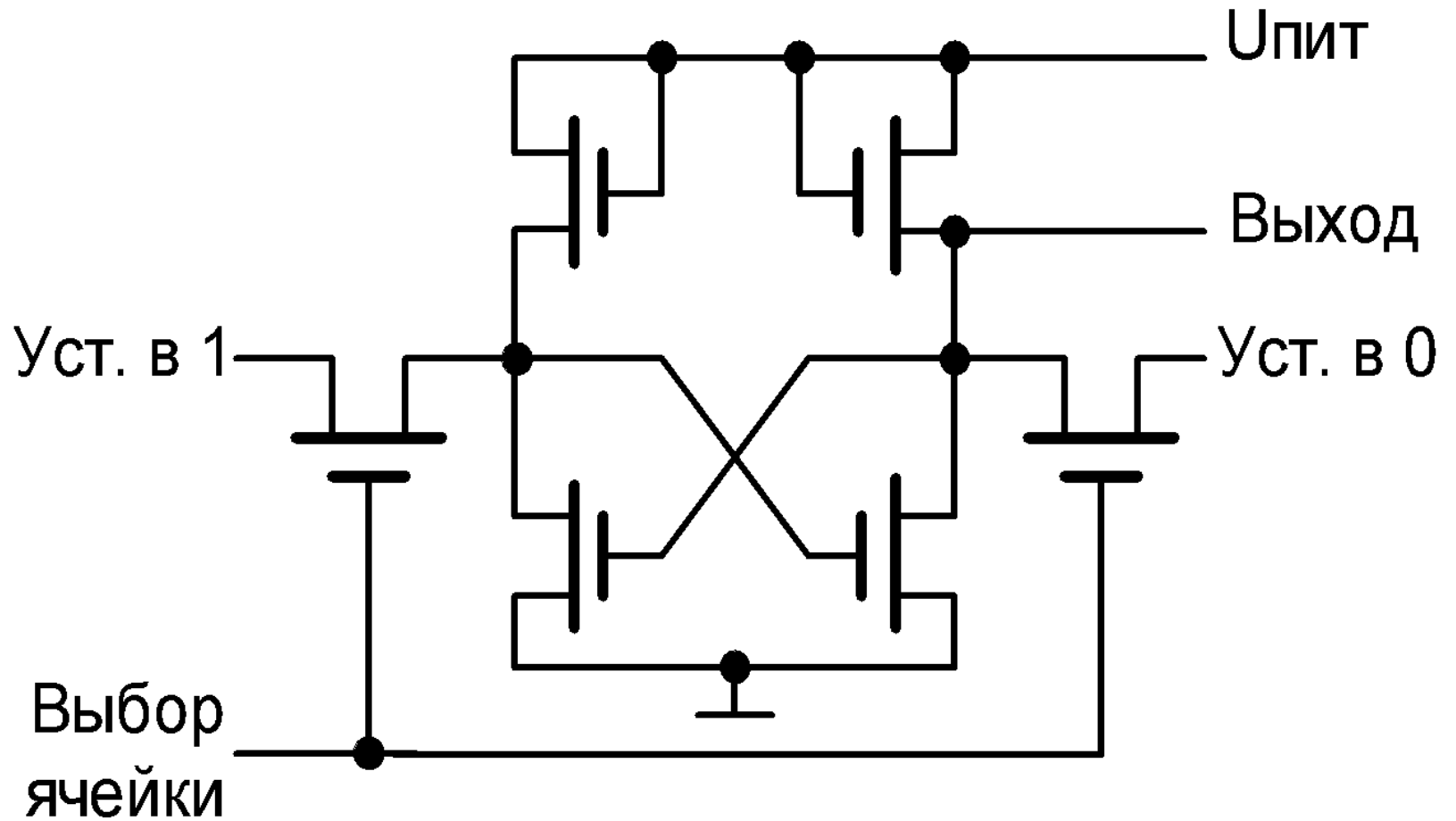
Статическая и динамическая оперативная память

В статических ОЗУ запоминающий элемент может хранить записанную информацию неограниченно долго (при наличии питающего напряжения). Запоминающий элемент динамического ОЗУ способен хранить информацию только в течение достаточно короткого промежутка времени, после которого информацию нужно восстанавливать заново, иначе она будет потеряна. Динамические запоминающие устройства, как и статические, энергозависимы.

Роль запоминающего элемента в статическом ОЗУ выполняет триггер. Такой триггер представляет собой схему с двумя устойчивыми состояниями, обычно состоящую из четырех или шести транзисторов.

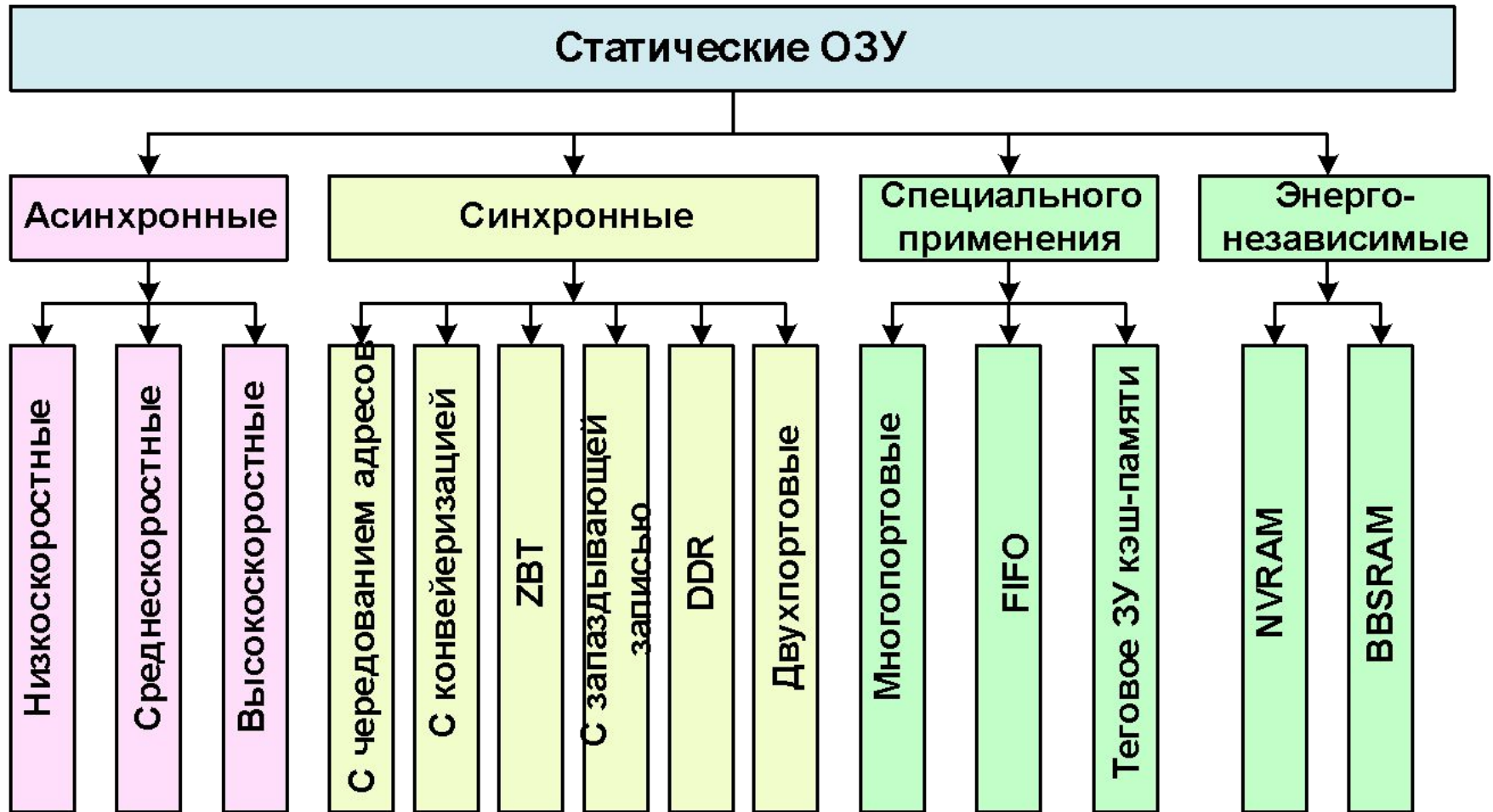
Схема с четырьмя транзисторами обеспечивает большую емкость микросхемы, однако у такой схемы большой ток утечки, а также высокая чувствительность к воздействию внешних источников излучения, которые могут стать причиной потери информации. Наличие двух дополнительных транзисторов позволяет в какой-то мере компенсировать упомянутые недостатки схемы на четырех транзисторах, но, главное — увеличить быстродействие памяти. Схематично запоминающий элемент статического ОЗУ на шести транзисторах показан на рисунке

Запоминающий элемент статического ОЗУ



Запоминающий элемент динамической памяти значительно проще. Он состоит из одного конденсатора и запирающего транзистора. Наличие или отсутствие заряда в конденсаторе интерпретируется как 1 или 0 соответственно. Простота схемы позволяет достичь высокой плотности размещения запоминающих элементов. Главный недостаток подобной технологии связан с тем, что накапливаемый на конденсаторе заряд со временем теряется. Даже при хорошем диэлектрике с электрическим сопротивлением в несколько тераОм (10^{12} Ом), используемом при изготовлении элементарных конденсаторов запоминающих элементов, заряд теряется достаточно быстро. Размеры у такого конденсатора микроскопические, а емкость имеет порядок 10^{-15} Ф. При такой емкости на одном конденсаторе накапливается всего около 40 000 электронов. Среднее время утечки заряда запоминающего элемента динамической памяти составляет сотни или даже десятки миллисекунд, поэтому заряд необходимо успеть восстановить в течение данного отрезка времени, иначе хранящаяся информация будет утеряна. Периодическое восстановление заряда запоминающего элемента называется регенерацией и осуществляется каждые 2-8 мс.

Виды статических ОЗУ



Асинхронные статические ОЗУ:

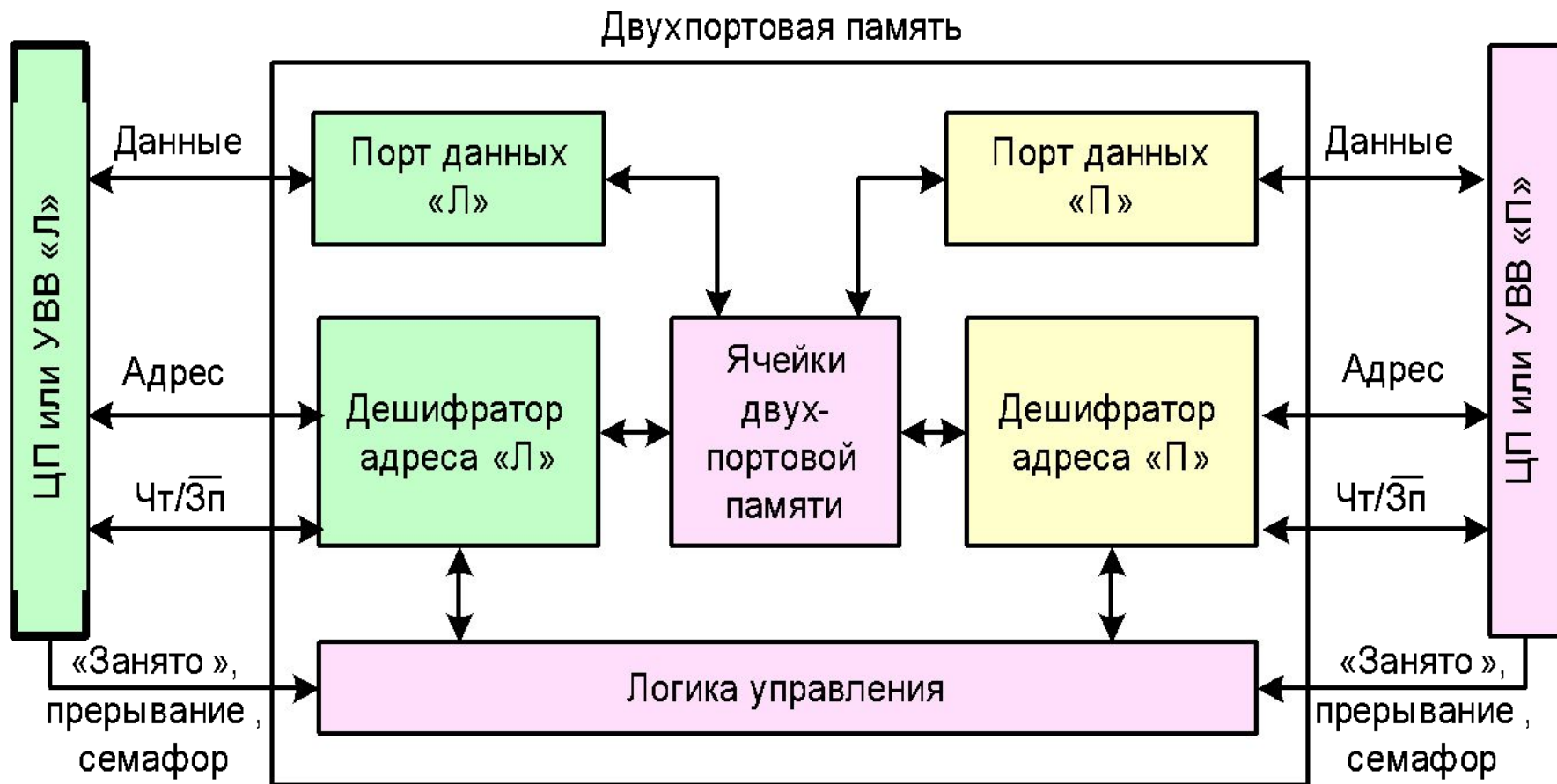
время доступа составляло 15-20 нс (в лучшем случае — 12 нс)

Синхронные статические ОЗУ:SSRAM обеспечивает работу в пакетном режиме с формулой 3-1 -1 -1, но лишь до определенных значений тактовой частоты шины. При более высоких частотах формула изменяется на 3-2-2-2.

PB SRAM. Начиная с Pentium II, взамен SSRAM оснащаются статической оперативной памятью с пакетным конвейерным доступом (PB SRAM — Pipelined Burst SRAM). Время доступа к PB SRAM составляет от 4,5 до 8 нс, при этом формула 3-1-1-1.

Компания IDT (Integrated Device Technology) предложила новую технологию, получившую название **ZBT SRAM** (Zero Bus Turnaround) — нулевое время переключения шины.

Структура двухпортового ОЗУ



Стандартное однопортовое ОЗУ имеет по одной шине адреса, данных и управления и в каждый момент времени обеспечивает доступ к ячейке памяти только одному устройству. В отличие от стандартного в n -портовом ОЗУ имеется n независимых наборов шин адреса, данных и управления, гарантирующий одновременный и независимый доступ к ОЗУ n устройствам. Это свойство позволяет существенно упростить создание многопроцессорных и многомашинных вычислительных систем, где многопортовое ОЗУ выступает в роли общей или совместно используемой памяти. В рамках одной вычислительной машины подобное ОЗУ может обеспечивать обмен информацией между центральным процессором и устройствами ввода/вывода намного эффективней, чем прямой доступ к памяти. В настоящее время серийно выпускаются двух- и четырехпортовые микросхемы, наиболее распространены двухпортовые.

В двухпортовой памяти имеются два набора адресных, информационных и управляющих сигнальных шин, каждый из которых обеспечивает доступ к общему массиву запоминающих элементов. Структура двухпортового ОЗУ показана на рисунке 9. Двухпортовому ОЗУ свойственна симметричная структура, наборы шин называют «левым» (Л) и «правым» (П). В целом организация матрицы запоминающих элементов традиционна.

Доступ к ячейкам возможен как через левую, так и через правую группу шин, причем если Л- и П-адреса различны, никаких конфликтов не возникает. Проблемы возможны, когда Л- и П-устройства одновременно обращаются по одному и тому же адресу и хотя бы одно из этих устройств пытается операцию записи. Для избежания таких случаев предусмотрена схема арбитража с использованием сигналов «Занято». Логика арбитража в микросхеме реализована аппаратными средствами. Схема обеспечивает формирования сигнала «Занято», запрещающего запись в ячейку для той половины, на которой адрес появится позже, а также принятие решения в пользу одного из входных портов при одновременном поступлении адресов.

Помимо возможности доступа к ячейкам с двух направлений, двухпортовая память снабжается средствами для обмена сообщениями между подключенными к ней устройствами: системой прерывания и системой семафоров. Первую из них называют аппаратной, а вторую – программной.

В системе прерываний двухпортовой памяти две последние ячейки микросхемы (с наибольшими адресами) используются в качестве «почтовых ящиков» для обмена сообщениями между устройствами, подключенными к Л- и П-портам. Сообщению от левого устройства выделена ячейка с четным адресом, а от правого – с нечетным. Когда устройство записывает информацию в свой «почтовый ящик», формируется запрос прерывания к устройству, подключенному к противоположному порту. Этот сигнал автоматически сбрасывается, когда адресат считывает информацию из своего «почтового ящика».

Система семафоров – это имеющийся в двухпортовой памяти набор из восьми триггеров, состояние которых может быть прочитано и изменено со стороны любого из портов. Триггеры играют роль программных семафоров или флагов, с помощью которых Л- и П-устройства могут извещать друг друга о каких-либо событиях. Сущность этих событий не зафиксирована и определяется реализуемыми программами. Обычно семафоры нужны для предоставления одному из процессоров монопольного права работы с определенным блоком данных до завершения всех необходимых операций с этим блоком. В этом случае процессор, монополизирующий блок данных, устанавливает один из семафоров в состояние 1, а по завершении – в 0.