

1. Чем характеризуется программно-управляемый ввод-вывод?

Наиболее простой метод:

- происходит под полным контролем ЦП;
- реализуется специальной процедурой ввода-вывода.

ЦП с помощью команды ввода-вывода сообщает МВВ, а через него и ВУ, о предстоящей операции. МВВ исполняет затребованное действие, после чего устанавливает признак в своем регистре состояния*. ЦП периодически опрашивает и анализирует содержимое регистра состояния МВВ (для определения момента завершения операции или пересылки очередного элемента блока данных).

2. Какие типы команд используются при программно-управляемом вводе-выводе?

- управление;
- проверка;
- чтение;
- запись.

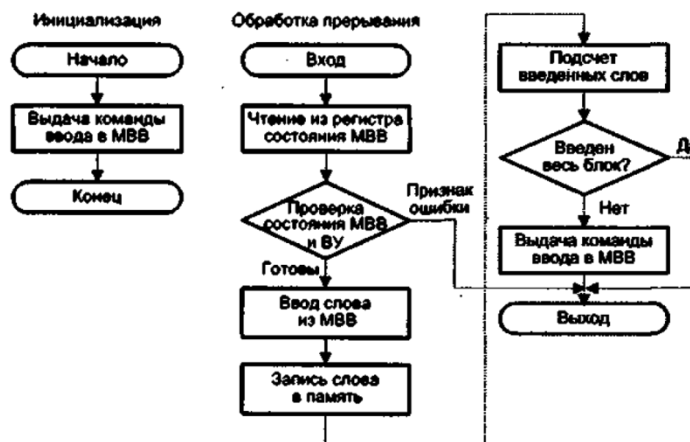
3. Какие задачи должны быть решены при реализации ввода-вывода по прерываниям?

ЦП выдает команду ввода или вывода, а затем продолжает выполнять другую полезную работу.

Когда ВУ готово к обмену данными, оно через МВВ извещает об этом процессору с помощью запроса на прерывание.

ЦП осуществляет передачу очередного элемента данных, после чего возобновляет выполнение прерванной программы.

Ввод данных по прерыванию



ЦП выдает команду чтения, а затем продолжает выполнение других заданий.

Получив команду, МВВ приступает к вводу элемента данных с ВУ.

Когда считанное слово оказывается в РД модуля, МВВ выдает по линии управления запрос прерывания

ЦП.

Выставив запрос, МВВ помещает введенную информацию на ШД, после чего он готов к следующей операции ввода-вывода.

ЦП в конце каждого цикла команды проверяет наличие запросов прерывания.

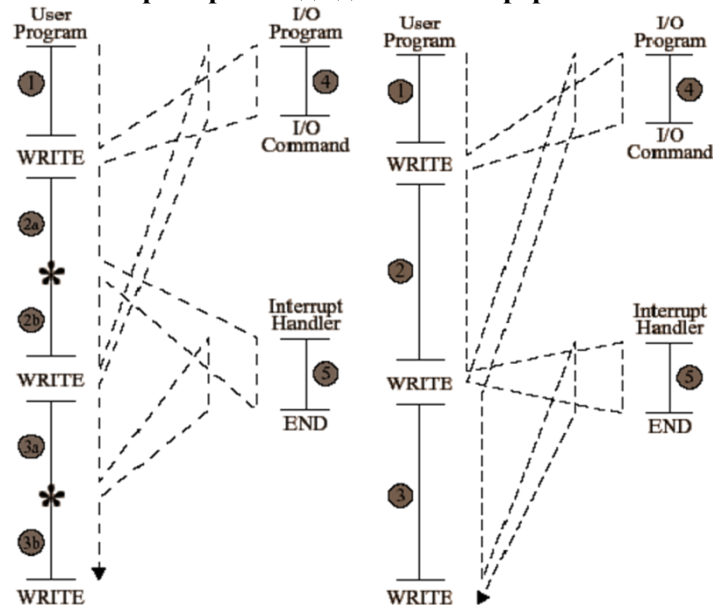
Если получен запрос от МВВ, ЦП сохраняет контекст текущей программы и обрабатывает прерывание*.

В данном случае ЦП:

- читает слово из модуля;
- записывает его в память;
- выдает модулю команду на считывание очередного слова.

ЦП восстанавливает контекст прерванной программы и возобновляет её выполнение.

Пример вывода данных по прерыванию



Подпрограмма WRITE состоит из фрагмента подготовки и собственно команды запуска сеанса обмена.

Подпрограмма завершает свою работу и возвращает управление основной программе, как только отправит на внешнее устройство команду запуска сеанса.

Устройство, получив необходимые данные и команду, приступает к самостоятельной работе, считывая данные из памяти и вывода их на печать.

4. Что [из перечисленного] свойственно вводу-выводу по прерываниям?

- ☺ выше эффективность: устранены ненужные ожидания;
- ☹ обработка прерывания **занимает достаточно много времени ЦП**;
- ☹ каждое слово, пересылаемое из ОП в МВВ или в противоположном направлении проходит через ЦП (как и при программно управляемом методе).

5. Какие существуют методы идентификации устройств, запрашивающих прерывание? Какой является простейшим в реализации, а какой – максимально «гибким»?

- *множественные линии прерывания (наиболее простой подход);*
- *программная идентификация (гибкий);*
- *векторное прерывание.*

6. Объясните термин «вектор прерывания».

Получив подтверждение прерывания от процессора, выставившее запрос устройство выдает на ШД специальное слово – вектор прерывания.

Вектор содержит либо адрес МВВ, либо какой-нибудь другой уникальный идентификатор, который ЦП интерпретирует как указатель на соответствующую программу обработки прерывания.

Свойства:

- ☺ наиболее эффективная процедура идентификации;
- ☺ не требует предварительных действий для определения источника запроса прерывания.

Реализуется с помощью хранящейся в ОП таблицы векторов прерывания (ТВП).

ТВП содержит адреса программ обработки прерываний.

Входом в таблицу служит вектор прерывания.

Начальный адрес таблицы (база) обычно задается неявно (под таблицу отводится определенная область памяти).

7. Укажите варианты идентификации векторных прерываний.

7. Укажите варианты идентификации векторных прерываний.

- цепочечный опрос;
- арбитраж шины.

8. Каким образом реализуется программная идентификация устройств, запрашивающих прерывание?

Обнаружив запрос прерывания, ЦП переходит в общей программе обработки прерывания.

Задачей общего обработчика является опрос всех МВВ с целью определения источника запроса.

Когда источник прерывания установлен, ЦП переходит к программе обработки прерывания, соответствующей этому источнику.

Множество прерываний

В случае одновременного поступления нескольких запросов прерывания существует два различных подхода:

- последовательная обработка прерываний;
- приоритетная обработка прерываний.

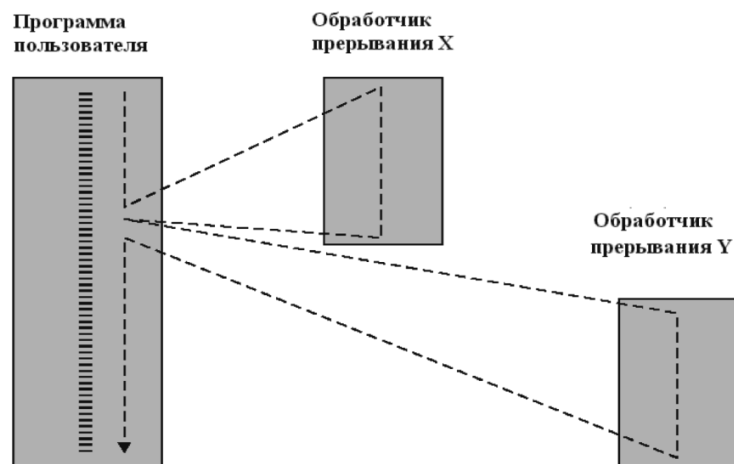
Последовательная обработка прерываний

Процессор игнорирует поступающие запросы до тех пор, пока не завершит обслуживание предыдущего.

Запрет обработки прерывания снимается после завершения предыдущего.

Свойства:

- ☺простота реализации;
- ☹не позволяет «ранжировать» источники прерывания, выделяя среди них более и менее важные.



Приоритетная обработка прерываний

Обслуживание прерываний с низкими приоритетами *может быть прервано* запросами более высокого приоритета.

После завершения обслуживания прерывания с высоким приоритетом процессор *возвращается к дообслуживанию* предыдущего прерывания.

9. При какой дисциплине обслуживание прерываний с низкими приоритетами может быть прервано запросами более высокого приоритета?

Приоритетная обработка

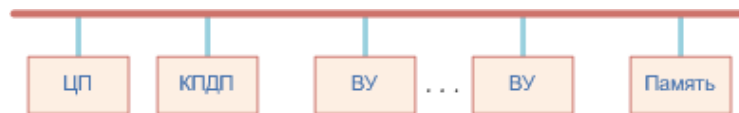
10. Укажите особенности режима ПДП.

- Эффективен при пересылке больших объемов данных.
- Устраняет недостатки предыдущих методов:

В пределах цикла команды имеется несколько «точек» для захвата шины.

Варианты реализации механизма ПДП

1 Совместно используемая общая системная шина.

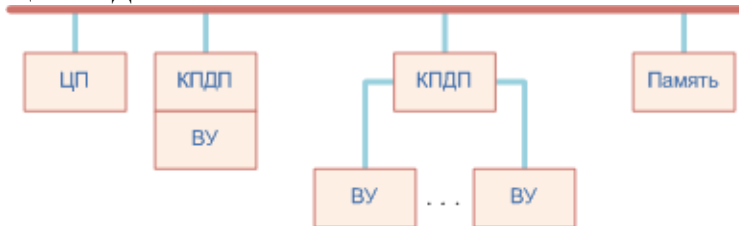


КПДП работает как заменитель ЦП, и обмен данными между памятью и ВУ через КПДП производится через программно управляемый ввод-вывод.

Свойства:

- ☹достаточно дешевый;
- ☹невысокая эффективность (каждая пересылка требует двух циклов шины как и при программно управляемом вводе-выводе).

2 Объединений функций КПДП и ВУ

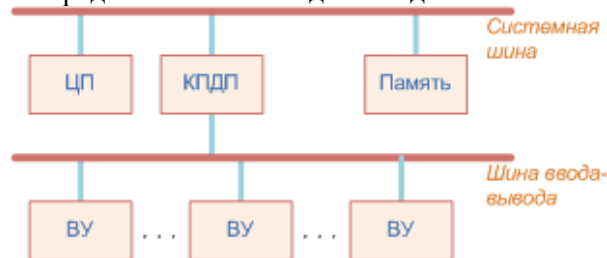


Между КПДП и одним или несколькими ВУ есть другой тракт, не включающий системную шину.

Логика ПДП может быть частью ВУ, либо это может быть отдельный КПДП, управляющий одним или несколькими ВУ.

Число необходимых циклов шины уменьшается.

3 Соединение КПДП с ВУ посредством шины ввода-вывода



Сокращается число интерфейсов В/ВЫВ в КПДП, а конфигурация становится легко расширяемой.

Особенность вариантов 2 и 3

Системная шина задействуется КПДП только для обмена данными о памяти.

Обмен данными между КПДП и ВУ реализуется минуя системную шину.

16. Что означают термины «канал ввода-вывода» и «процессор ввода-вывода»? канал ввода-вывода?

Цель: максимальное высвобождение ЦП от управления процессами ввода-вывода.

Канал ввода-вывода (КВВ). *МВВ с расширенными возможностями и правами процессора со специализированным набором команд, ориентированных на операции ввода-вывода:*

ЦП дает указание КВВ выполнить хранящуюся в памяти компьютера программу ввода-вывода.

КВВ извлекает и исполняет команды этой программы без участия ЦП и прерывает его только после завершения всей программы ввода-вывода.

Различие между каналом и процессором ввода-вывода достаточно условно, поэтому в дальнейшем использован термин «канал».

процессор ввода-вывода

ЦП дает указание КВВ выполнить хранящуюся в памяти компьютера программу ввода-вывода. КВВ извлекает и исполняет команды этой программы без участия ЦП и прерывает его только после завершения всей программы ввода-вывода.

Процессор ввода-вывода (ПВВ) – МВВ с функциональностью КВВ и собственной локальной памятью; при этом возможно управление множеством устройств с минимальным привлечением ЦП.

17. Каким образом функционирует КВВ?

ЦП инициирует ввод-вывод путем:

- инструктирования канала о необходимости выполнить канальную программу, находящуюся в ОП;
- указания начального адреса этой программы в памяти компьютера.

КВВ следует этим указаниям и управляет пересылкой данных.

Пересылка ведется в режиме ПДП.

ВУ взаимодействуют с каналом, получая от него приказы.

Т.о., в компьютерах с КВВ управление вводом-выводом строится иерархическим образом

18. Какие существуют способы организации взаимодействия ПУ с КВВ?

Определяются соотношением быстродействия ОП и ВУ.

По этому признаку ВУ образуют две группы:

- быстродействующие** (накопители на магнитных дисках (НМД), накопители на магнитных лентах (НМЛ));
- медленнодействующие** (дисплеи, печатающие устройства и др.) со скоростями порядка 1 Кбайт/с и менее.

С учетом производительности ВУ в КВВ реализуются два режима работы:

- мультиплексный** (режим разделения времени);
- монопольный**.

Мультиплексный режим

Несколько ВУ разделяют канал во времени, при этом каждое из параллельно работающих с каналом ВУ связывается с КВВ на короткие промежутки времени только после того, как ВУ будет готово к приему или выдаче очередной порции информации (байта, группы байтов и т. д.).

Такая схема принята в мультиплексном канале ввода-вывода.

Байт-мультиплексный канал – в течение сеанса связи пересылается один байт или несколько байтов, образующих одно машинное слово.

Блок-мультиплексный канал – в пределах сеанса связи пересылка данных выполняется поблочно.

Монопольный режим

После установления связи между каналом и ВУ последнее монополизировало канал на все время до завершения инициированной процессором канальной программы и всех предусмотренных этой программой пересылок данных между ВУ и ОП.

На все время выполнения канальной программы канал оказывается недоступным для других ВУ.

Данную процедуру обеспечивает селекторный канал ввода-вывода.

В блок-мультиплексном канале в рамках сеанса связи пересылка блока осуществляется в монопольном режиме.

19. В чем состоит роль ЦП в реализации ввода-вывода на основе каналов?

Все функции ЦП сводятся к запуску и остановке операций в КВВ, а также проверке состояния канала и подключенных к нему ВУ. (т.е. управление)

ЦП дает указание КВВ выполнить хранящуюся в памяти компьютера программу ввода-вывода. КВВ извлекает и исполняет команды этой программы без участия ЦП и прерывает его только после завершения всей программы ввода-вывода.

20. Что свойственно канальной программе?

Реализуется для каждого ВУ, с которым предполагается обмен информацией.

Описывает нужную последовательность операций ввода-вывода.

Хранится в ОП компьютера.

Роль команд в канальных программах выполняют управляющие слова канала (УСК).

21. Что [из перечисленного] характеризует селекторный канал?

Монопольный режим

После установления связи между каналом и ВУ последнее монополизировало канал на все время до завершения инициированной процессором канальной программы и всех предусмотренных этой программой пересылок данных между ВУ и ОП.

На все время выполнения канальной программы канал оказывается недоступным для других ВУ. Данную процедуру обеспечивает селекторный канал ввода-вывода.

22. Какая информация подается на вход модулей ОП?

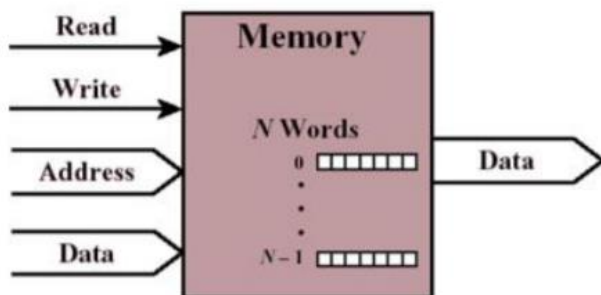
Данные, адрес ячеек, управляющие сигналы: чтение, запись, синхронизация

23. В чем заключаются отличия MBV от ОП?

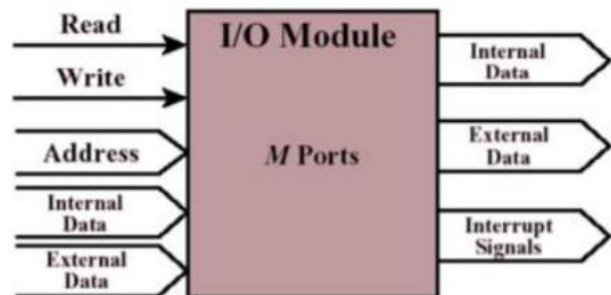
По характеру операций подобен ОП.

Разница в том что MBV содержит и передает данные еще и от внешнего устройства.

Основная память



Модуль ввода-вывода



По характеру операций подобен ОП.

Вход:

- данные;
- адреса ячеек;
- управляющие сигналы:
 - Чтение;
 - Запись;
 - Синхронизация.

Выход:

- данные в ЦП и MBV

Вход:

- данные от одного или нескольких ВУ;
- данные от ЦП;
- управляющие сигналы от ЦП;
- адреса от ЦП (номер порта для идентификации ВУ)

Выход:

- данные к ВУ;
- данные к ЦП;
- управляющие сигналы к ВУ;
- запросы прерываний в ЦП

24. Какое назначение имеют связи ЦП?

Распространенные:

- с непосредственными связями;
- на основе магистрали (шины).

Эволюция:

- непосредственные связи;
- общая шина;
- иерархия шин.

Магистраль или шина (bus) – совокупность электрических связей и обслуживающих устройств, обеспечивающих обмен данными между двумя или более устройствами.

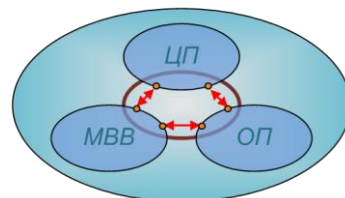
25. В чем заключается специфика архитектуры компьютера с непосредственными связями?

Пример: компьютер IAS

Характеристики связей определяются:

- видом информации;
- характером обмена;
- интенсивностью обмена.

Свойства:



- ☹ возможность развязки «узких мест» за счет улучшения структуры и характеристик отдельных связей (экономическая выгода);
- ☹ плохая реконфигурируемость.

26. Какой тип взаимосвязей наиболее распространен в современных микроЭВМ?

Структура на основе общей шины

27. Что характерно для архитектуры на основе магистральной?

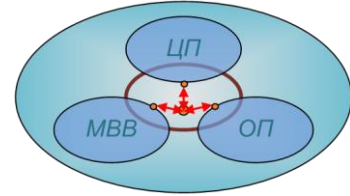
Структура на основе общей шины

Системная шина* (system bus), магистраль, общая шина, центральная шина.

Суть:

- **связывает (логически и физически) основные компоненты;**
- **служит единственным трактом для потоков команд, данных и управления.**

Распространена в мини- и микроЭВМ.



Характеристики шин

- совокупность сигнальных линий;
- спецификации:
 - физические;
 - механические;
 - электрические;
- используемые *сигналы*:
 - арбитража;
 - состояния;
 - управления;
 - синхронизации;
- *правила* взаимодействия устройств (протокол шины).

28. Что является достоинством архитектуры с иерархией шин?

Архитектура с иерархией шин содержит помимо системной шины несколько дополнительных шин, которые могут обеспечивать непосредственную связь между устройствами с наиболее интенсивным обменом, например, ЦП и кэш-памятью.

Наиболее распространена.

«Логическая» системная шина состоит из:

- главной (host) или системной шины;
- дополнительных шин (шины ввода-вывода, шины расширения ввода-вывода).

Суть:

- **непосредственная связь между устройствами с наиболее интенсивным обменом (напр., между ЦП и кэш-памятью);**
- **объединение однотипных УВВ с последующим выходом с дополнительной шины на главную.**

Каждая из шин обеспечивает:

- **упрощение взаимодействия различных подсистем;**
- **высокую пропускную способность;**
- **избыточность (для повышения отказоустойчивости);**
- **эффективность.**

Свойства:

- ☺ снижение нагрузки на главную шину («процессор-память»);
- ☺ повышение производительности компьютера;
- ☺ более эффективное использование пропускной способности главной шины.

29. Какие группы линий составляют системную магистраль?

Первую группу (2-8 линий) образуют линии, по которым передаются сигналы управления транзакциями (тип транзакции, тип адреса, тип применяемого протокола, число байтов, передаваемых по шине данных).

Ко второй группе (1-4 линии) относятся линии передачи информации состояния (статуса). Ведомое устройство может информировать ведущее устройство о своём состоянии или передать код возникшей ошибки.

Третью группу (3-11 линий) образуют линии арбитража. В реальных системах на роль ведущего могут одновременно претендовать сразу несколько из подключённых к шине устройств. Однако управлять шиной в каждый момент времени может только одно из них. Процедура допуска к управлению шиной одного из претендентов называется **арбитражем шины**. Решение обычно принимается на основе приоритетов претендентов.

В **четвёртую группу** (1-2 линии) входят линии прерывания. По ним передаются запросы на обслуживание, посылаемые от ведомых устройств к ведущему.

Пятую группу (1-4 линии) составляют линии, используемые для организации последовательных локальных сетей. Последовательная передача данных протекает значительно медленнее, и сети выгоднее строить, не загружая быстрые линии основных шин адреса и данных. Кроме того, линии этой группы могут использоваться в качестве дополнительного (хотя и медленного) тракта для замены шины адреса и шины данных в случае их отказа.

К **шестой группе** (4-5 линий) относятся линии *позиционного кода*, которые подсоединяются к специальным выводам разъёма. Такой код может быть использован для индивидуальной инициализации материнской или дочерней платы при включении или перезагрузке системы.

В **седьмую группу** (2-6 линий) входят линии *тактирования* и *синхронизации*.

Кроме того, необходимо отметить линии для подвода питающего напряжения и линии заземления.

–**физические;**

–**механические;**

–**электрические;**

– используемые *сигналы*:

– арбитража;

– состояния;

– управления;

– синхронизации;

правила взаимодействия устройств (протокол шины).

30. Что [из перечисленного] свойственно ведущему устройству на шине?

любое устройство, способное взять на себя владение шиной и управлять пересылкой данных;
не обязательно само использует данные;

Взаимодействие устройств

Инициатор передачи данных должен:

a. получить шину в свое распоряжение;

b. передать адрес и данные по шине.

Инициатор приема данных должен:

c. получить шину в свое распоряжение;

d. передать запрос исполнителю (посредством линий управления и адреса);

e. ожидать посылки данных исполнителем.

Количество ведущих, подключенных к шине:

f. потенциальных: ≥ 1 ;

g. активных в любой момент времени: ≤ 1

ДОП_ИНФА

Арбитраж – процедура допуска к управлению шиной только одного из претендентов; предотвращает одновременную активность нескольких ведущих.

Широковещательный режим записи – передача информации от одного ведущего сразу нескольким ведомым; арбитраж не требуется.

31. В чем заключается назначение арбитража шины?

Арбитраж – процедура допуска к управлению шиной только одного из претендентов; предотвращает одновременную активность нескольких ведущих.

Для предотвращения конфликтов предусматриваются:

- механизмы арбитража запросов;
- правила предоставления шины одному из запросивших устройств.

Обычно: решение принимается на основе приоритетов, присваиваемых каждому потенциальному ведущему.

Схемы арбитража

- централизованная схема;
- децентрализованная схема.

Критерии выбора:

- требования к производительности;
- стоимостные ограничения.

32. Какие существуют схемы управления приоритетами?

Схемы приоритетов

Уровни приоритета:

- неизменные (*статический* или *фиксированный* приоритет);
- изменяемые по какому-либо алгоритму (*динамический* приоритет).

Статические приоритеты: высокоприоритетные устройства могут полностью блокировать доступ к шине устройствам с низким уровнем приоритета.

Динамические приоритеты: принцип равнодоступности*.

Динамические приоритеты

Обычно: приоритеты меняются после каждого цикла арбитража.

33. Как реализуется централизованный арбитраж шины?

Центральный арбитр (ЦА)

Центральный контроллер шины (КШ).

Назначение: предоставление доступа к шине только одному из запросивших ведущих.

Варианты:

- самостоятельный модуль;
- часть ЦП.

Один арбитр → единственная точка отказа.

Схемы (по способу подключения ведущих устройств к ЦА):

- *параллельные;*
- *последовательные.*

Централизованный параллельный арбитраж

Текущий ведущий – устройство, управляющее шиной в момент поступления нового запроса.

Запросивший ведущий – устройство, выставившее запрос на управление шиной.

ЗШ – сигнал запроса шины

ПШ – сигнал предоставления шины

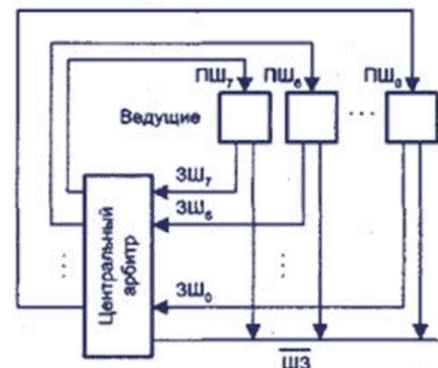
ШЗ – сигнал занятия шины

ЗШ поступают на входы ЦА по индивидуальным линиям.

ПШ по индивидуальной линии возвращается ведущему i , выбранному арбитром.

Текущий ведущий должен сохранять ШЗ и ЗШ активными в течение всего времени использования шины.

Занять шину новый ведущий сможет лишь после того, как текущий ведущий j снимет ШЗ.



Алгоритм централизованного параллельного арбитража

Получив запрос от ведущего, приоритет которого выше, чем у текущего j , арбитр:

- снимает сигнал ПШ j на входе текущего ведущего;
- выдает сигнал ПШ запросившему ведущему.

Текущий ведущий, в ответ на снятие арбитром сигнала ПШ j , снимает свои сигналы ШЗ и ЗШ j .
Запросивший ведущий принимает управление шиной.

Если в момент пропадания сигнала ПШ на шине происходит передача информации, текущий ведущий сначала завершает передачу и лишь после этого снимает свои сигналы.

Централизованный последовательный арбитраж



Для выделения наиболее приоритетного запроса используется один из сигналов, поочередно проходящий через цепочку ведущих (цепочечный или гирляндный арбитраж):

- ПШ (наиболее распротр.);
 - ЗШ;
 - дополнительный сигнал разрешения (РШ).
- Пример: понижение уровней приоритета слева направо.
Основа: статическое распределение приоритетов:
- наивысший – ближайшее к арбитру ведущее устройство;
 - далее приоритеты последовательно понижаются.

Алгоритм централизованного последовательного арбитража

Получив сигнал ЗШ, арбитр анализирует состояние линии ШЗ, и если шина свободна, формирует сигнал ПШ.

Сигнал ПШ *последовательно* переходит по цепочке от одного ведущего к другому.

Если устройство, на которое поступил сигнал ПШ, не запрашивало шину, оно пропускает сигнал дальше по цепочке.

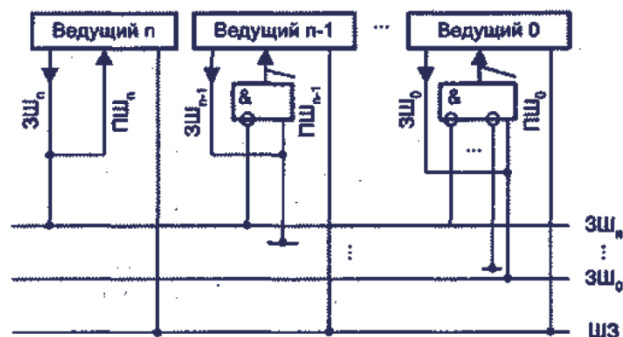
Когда ПШ достигнет самого первого из запрашивших ведущих, последний блокирует дальнейшее распространение сигнала ПШ по цепочке и берет на себя управление шиной.

34. Как реализуется децентрализованный арбитраж шины?

Единый арбитр отсутствует (*распределенный арбитраж*).

По сравнению с централизованной схемой меньшая чувствительность к отказам претендующих на шину устройств.

Децентрализованный параллельный арбитраж



Каждый ведущий:

- имеет уникальный уровень приоритета;
- содержит блок управления доступом к шине (контроллер шины, КШ), формирующий сигналы ПШ и занятия шины.

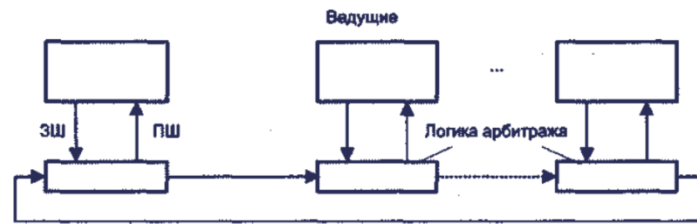
При совместном использовании шины блоки КШ взаимодействуют друг с другом, разделяя между собой ответственность за доступ к шине.

Сигналы запроса от любого ведущего поступают на входы всех остальных ведущих.

Логика арбитража реализуется в КШ каждого ведущего.

Кольцевая схема

Основа: циклическая схема приоритетов



Комбинированные схемы арбитража

Последовательно-параллельные схемы сочетают достоинства обоих методов.

Принцип:

- все ведущие разбиваются на группы;
- арбитраж внутри группы ведется по последовательной схеме, а между группами – по параллельной.

35. В чем заключается отличие параллельного арбитража от последовательного?

36. Как организованы линии управления системной магистрالي.

37. Что составляет шинную транзакцию?

Транзакции – операции на шине. Основные виды: «Чтение»; «Запись»; «Ввод»; «Вывод».

Составляющие части: посылка адреса; посылка/прием данных.

38. Какая информация может передаваться по шине адреса, и на что влияет ширина этой шины?

Назначение: позволяет ведущему устройству выбрать ведомое (источник или приемник данных) и установить соединение с ним.

На ША могут выдаваться: **адреса ячеек памяти; номера регистров ЦП; адреса портов ввода/вывода и т.п.** Вид адреса (в текущей транзакции) уточняется дополнительной информацией, которая:

- может косвенно содержаться в самом адресе;
- или (чаще всего) передается по специальным управляющим линиям шины.

Возможны:

- различные структуры адресов;
- адреса специального вида, обеспечивающие одновременный выбор определенной группы ведомых либо всех ведомых сразу (broadcast).

Ширина шины

- характеризуется числом сигнальных линий, выделенных для передачи адреса;
- определяет максимально возможный размер адресного пространства;
- потенциальную емкость адресуемой памяти;
- число обслуживаемых портов ввода-вывода.

Пример: 16-разрядная шина адреса позволяет адресовать 64К ячеек памяти.

39. Что свойственно понятию пропускной способности шины?

Пропускная способность шины - количество единиц информации, которое допускается передать по шине за единицу времени;

определяется:

- **физическим построением шины;**
- **природой подключаемых к ней устройств.**

40. Что характерно для мультиплексирования шин?

Мультиплексирование шин

Назначение: сократить общее число линий.

Временное мультиплексирование – пересылка адреса и данных по одним и тем же линиям в разных тактах цикла шины. Цикл шины разбит на два временных интервала (для передачи адреса и передачи данных) – режим разделения времени.



Свойства:

- усложняется логика связи с шиной;
- эффективность использования шины ниже, особенно в транзакциях записи (одновременно передать адрес ячейки памяти и записываемые данные невозможно).

41. К какой категории относится [приведенный] пример системы взаимодействия?

42. В чем заключаются особенности архитектуры с тремя видами шин?

Архитектура с «пристройкой» (mezzanine architecture).

Дополнительная высокоскоростная шина расширения – для подключения быстродействующих ВУ. Шины ввода-вывода подключаются к шине расширения.

Свойства:

- еще большее снижение нагрузки на шину «процессор-память».



Пример: магистраль PCI (peripheral component interconnect) в одно- и мультипроцессорной конфигурации, высокоскоростная архитектура соединений (приложение А).

43. Что называют протоколом шины?

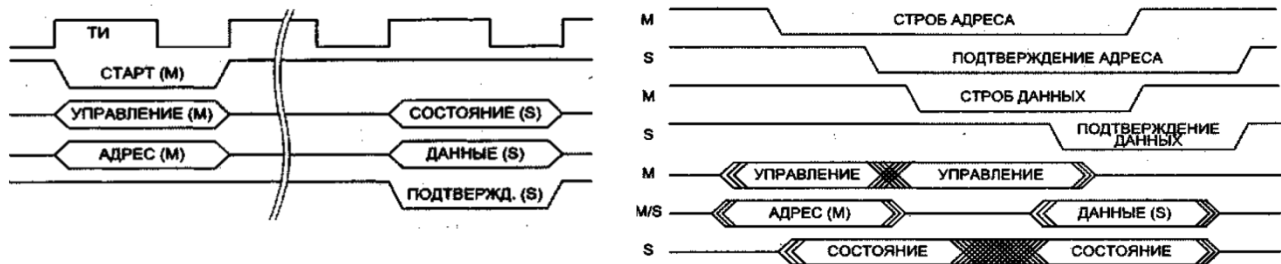
Протокол шины – метод информирования о достоверности адреса, данных, управляющей информации и информации состояния.

44. Какой протокол используется в случае, если для каждой группы линий шины формируется свой сигнал подтверждения достоверности?

асинхронный – для каждой группы линий шины формируется свой сигнал подтверждения достоверности.

45. Какой тип протокола шины соответствует [приведенной] временной диаграмме?

Транзакция чтения на синхронной шине..... На асинхронной



46. Что относится к свойствам синхронных протоколов шины?

- ☺меньшее число сигнальных линий, проще для понимания, реализации и тестирования;
- ☺быстрые и дешевые – дополнительная логика практически не требуется;
- ☹менее гибкие – привязаны к конкретной максимальной тактовой частоте → к конкретному уровню технологии. (Часто не способны реализовать потенциал производительности новых устройств);
- ☹перекос синхросигналов → ограниченная длина.

Пример: шины «процессор-память».

47. Что относится к свойствам асинхронных протоколов шины?

Свойства асинхронных шин

- ☺скорость пересылки данных диктуется ведомым;
- ☺самосинхронизация → возможность совместного использования устройствами с различным быстродействием;
- ☺автоматическая адаптация к требованиям устройств, обменивающихся информацией в данный момент;
- ☺для ускорения не требуется замена старых медленных устройств на быстрые новые;
- ☹некоторое увеличение сложности аппаратуры.

48. Что представляет собой блочный режим шины?

Блочный или пакетный режим (burst mode) – один адресный цикл сопровождается множественными однотипными циклами данных.

Скорость передачи собственно данных увеличивается за счет уменьшения числа передаваемых адресов.

Пример: шина Futurebus+.



Передается адрес только первой ячейки. Последующие адреса генерируются в самой памяти путем последовательно увеличения начального адреса.

Последовательные элементы блока данных располагаются в последовательных ячейках.

Текущий адрес внутри пакета не указывается.

49. Какую особенность шины демонстрирует [приведенный] рисунок?

Пример: шина Futurebus+.



Пакетный режим пересылки информации



Конвейеризация транзакций



Расщепление транзакций

50. Что характеризует методику расщепления транзакций?

Назначение: увеличение эффективной полосы пропускания шины.

Условие: независимые линии адреса и данных.

Обычно: эффективен на транзакциях чтения.

Две части транзакции чтения:

- адресная транзакция:
 - ведущий выставляет на ША адрес ячейки;
 - в памяти начинается процесс поиска и извлечения затребованных данных;
- транзакция данных – по завершении чтения память:
 - запрашивает доступ к шине и становится ведущим устройством;
 - направляет считанные данные по шине данных.



Таким образом:

- от поступления запроса до формирования отклика шина остается незанятой и может быть востребована для выполнения других транзакций;

- на шине имеют место поток запросов и поток откликов.

Отличие от конвейеризации: ответы на запросы могут поступать в произвольной последовательности.

Тег – признак, обеспечивающий соответствие информации на шине данных запросу.

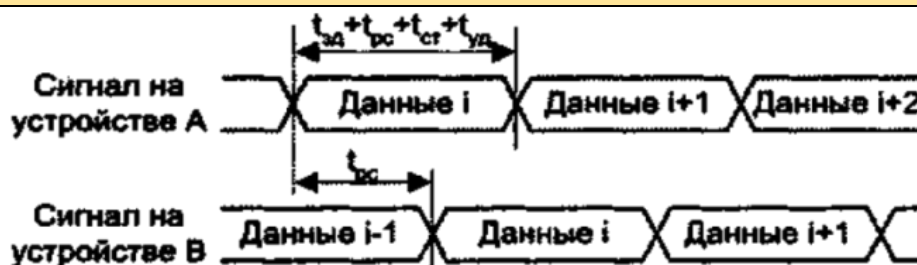
Свойства:

- ☺ более эффективное использование полосы пропускания шины;
- ☹ дополнительная задержка – требуются два подтверждения (при запросе и при отклике);
- ☹ дополнительные затраты – требуется, чтобы транзакции были тегированы и отслеживались каждым устройством.

51. В чем заключается трудность использования расщепления шинных транзакций?

- **дополнительная задержка** – требуются два подтверждения (при запросе и при отклике);
- **дополнительные затраты** – требуется, чтобы транзакции были тегированы и отслеживались каждым устройством.

52. Какой из [приведенных] рисунков соответствует методике конвейеризации транзакций?



53. Что способствует увеличению полосы пропускания шины?

Увеличение полосы пропускания шины

- раздельные шины адреса и данных;
- увеличение ширины шины данных;
- повышение тактовой частоты шины;
- использование блочных(пакетных) транзакций.

54. Объясните смысл термина «память компьютера» и перечислите основные характеристики ЗУ.

Память компьютера – совокупность устройств, служащих для запоминания, хранения и выдачи данных.

В современных компьютерах: одновременно используется несколько типов ЗУ (с различным принципом действия, характеристиками, назначением).

Основные операции – обращение к памяти:

- **запись** – занесение данных в память;
- **чтение** – выборка данных из памяти.

Синонимы: обращение при записи или считывании.

Характеристики систем памяти

Рассматривают применительно к конкретному виду ЗУ.

К основным параметрам, характеризующим запоминающие устройства, относятся *емкость* и *быстродействие*.

2. Емкость

Характеризуется числом единиц информации (битов, байтов), которое может храниться в ЗУ.

На практике: более крупные единицы с добавлением соответствующих приставок.

55. Как классифицируется память по признаку ее расположения?

1.1. Процессорная память – на общем кристалле с ЦП*. Наиболее скоростные виды памяти.

1.2. Внутренняя память – ЗУ, расположенные на системной плате:

- ОП;
- кэш-память второго** и последующих уровней.

1.3. Внешняя память – медленные ЗУ большой емкости (магнитные и оптические диски, магнитные ленты). К ядру компьютера подключаются аналогично устройствам ввода-вывода.

56. Что является пересылаемой информационной единицей при взаимодействии с памятью?

Для ОП: определяется шириной шины данных.

Обычно: единица пересылки равна длине слова (не обязательно).

Для внешней памяти: данные часто передаются блоками – единицами, превышающими размер слова.

57. Объясните особенности различных методов доступа к данным в памяти компьютера.

Существенно влияют на быстродействие памяти; выделяют четыре основных.

4.1. Последовательный доступ (sequential access).

Информация хранится в виде последовательности блоков данных – **записей**.

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;
- позиции элемента внутри данной записи.

Пример: ЗУ на магнитной ленте.

4.2. Прямой доступ (direct access).

Каждая запись имеет **уникальный адрес**, отражающий ее физическое размещение на носителе информации.

Обращение осуществляется как:

- **адресный доступ** к началу записи;
- **последовательный доступ** к определенной единице информации внутри записи.

Время доступа – величина переменная.

Пример: ЗУ на магнитных дисках.

4.3. Произвольный доступ (random access).

Каждая ячейка памяти имеет **уникальный физический адрес**.

Обращение к любой ячейке:

- одинаковое по времени;
- возможно в произвольной очередности.

Пример: запоминающие устройства ОП.

4.4. Ассоциативный доступ (associative access).

Поиск ячеек со значением отдельных битов, совпадающим с одноименными битами в заданном образце.

Сравнение осуществляется параллельно для всех ячеек памяти, независимо от ее емкости.

Пример: ассоциативная кэш-память.

58. Какие параметры используются для оценки быстродействия памяти. Как они рассчитываются?

1) Время доступа

интервал времени от момента поступления адреса до момента, когда данные заносятся в память или становятся доступными.

2) Пропускная способность (скорость передачи)

максимальное количество данных, которое память может (передать?)

эффективная частота * на 8 байт (64 бита) (ширина шины)*

3) Длительность цикла памяти или период обращения (memory cycle, TC).

Для памяти с произвольным доступом (только): минимальное время между двумя последовательными обращениями к памяти.

Включает в себя время доступа плюс некоторое дополнительное время (для затухания сигналов на линиях, восстановления считанной информации и т.д.).

59. Каковы предпосылки внедрения, и в чем заключаются принципы иерархической архитектуры памяти компьютера?

Память – «узкое место» фон-неймановских компьютеров из-за ее серьезного отставания по быстродействию от процессоров. Этот разрыв неуклонно увеличивается (приблизительно на 50% в год).

Закономерность в используемых ЗУ:

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

Цель разработки системы памяти: обеспечить требуемую емкость и высокое быстродействие за приемлемую цену.

Подход (наиболее распространенный): иерархический принцип построения.

Иерархическая система памяти состоит из ЗУ различных типов, которые, в зависимости от характеристик, относят к определенному уровню иерархии.

Более высокий уровень меньше по емкости, быстрее и имеет 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 нижележащем уровне и т.д.

Обобщенная характеристика иерархической памяти

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 ЦП. ☺

60. Объясните термин «локальность по обращению». Какие существуют виды локальности?

Позволяет эффективно реализовывать многоуровневую систему памяти.

Принцип локальности в численной форме представляют в виде **правила «90/10»**: 90% времени работы программы связано с доступом к 10% адресного пространства этой программы.

Существуют три вида локальности.

1. Пространственная локальность программы.

С высокой вероятностью адрес очередной команды программы:

- следует непосредственно за адресом, по которому была считана текущая команда;
- расположен вблизи него.

2. Пространственная локальность данных.

Обрабатываемые данные, как правило, структурированы, и такие структуры обычно хранятся в последовательных ячейках памяти.

3. Временная локальность.

Программы содержат множество небольших циклов и подпрограмм. Небольшие наборы команд могут многократно повторяться в течение некоторого интервала времени.

61. Что является следствием свойства локальности по обращению?

Программу представляют в виде последовательно обрабатываемых фрагментов – *компактных групп команд и данных (кластеров)*.

Помещая кластеры в более быструю память, можно существенно снизить общие задержки на обращение:

- команды и данные, будучи один раз переданы из медленного ЗУ в быстрое, затем могут использоваться многократно;
- 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;
- в процессе обработки активно использовать сравнительно небольшую быструю память, увеличение емкости которой сопряжено с высокими затратами.

Информационный блок

На каждом уровне иерархии информация разбивается на *блоки*. **Блок** – наименьшая информационная единица, пересылаемая между двумя соседними уровнями.

Размер блоков: фиксированный; переменный. При фиксированном размере блока емкость памяти обычно кратна его размеру.

Размер блоков на различных уровнях иерархии:

- чаще всего различен;
- увеличивается от верхних уровней к нижним.

62. Объясните принцип функционирования многоуровневой памяти.

При доступе к командам и данным сначала производится поиск в памяти верхнего уровня. Факт обнаружения нужной информации – *попадание (hit)*, в противном случае – *промах (miss)*.

При промахе:

- производится поиск в ЗУ следующего более низкого уровня, где также возможны попадание или промах;
- после обнаружения необходимой информации выполняется последовательная пересылка блока, содержащего искомую информацию, с нижних уровней на верхние.

Независимо от числа уровней иерархии пересылка информации может осуществляться только между двумя соседними уровнями*.

63. Перечислите и объясните смысл критериев эф-фективности многоуровневой памяти.

Критерии эффективности многоуровневой памяти

- *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; включает в себя:
 - *время доступа (access time)* – время обращения к первому слову блока при промахе (обусловлено задержкой памяти более низкого уровня);
 - *время пересылки (transfer time)* – дополнительное время для пересылки оставшихся слов блока (связано с полосой пропускания канала между ЗУ двух смежных уровней).

64. Перечислите параметры, которыми описывается многоуровневая память.

Описание уровней иерархии ЗУ

- *размещение блока* – допустимое место расположения блока на примыкающем сверху уровне иерархии;

- **идентификация блока** – способ нахождения блока на примыкающем сверху уровне;
- **замещение блока** – выбор блока, заменяемого при промахе с целью освобождения места для нового блока;
- **согласование копий (стратегия записи*)** – обеспечение согласованности копий одних и тех же блоков, расположенных на разных уровнях, при записи новой информации в копию, находящуюся на более высоком** уровне.

Пример

Процессор может обращаться к памяти двух уровней. Время доступа к памяти первого уровня – $T_1 = 0,1$ мкс, второго уровня – $T_2 = 1$ мкс. Если слово читается из памяти второго уровня, то сначала оно записывается в память первого уровня, а затем извлекается в процессор. Вероятность попадания запросов в память первого уровня $P = 0,95$. Необходимо вычислить среднее время обращения к иерархической подсистеме памяти.

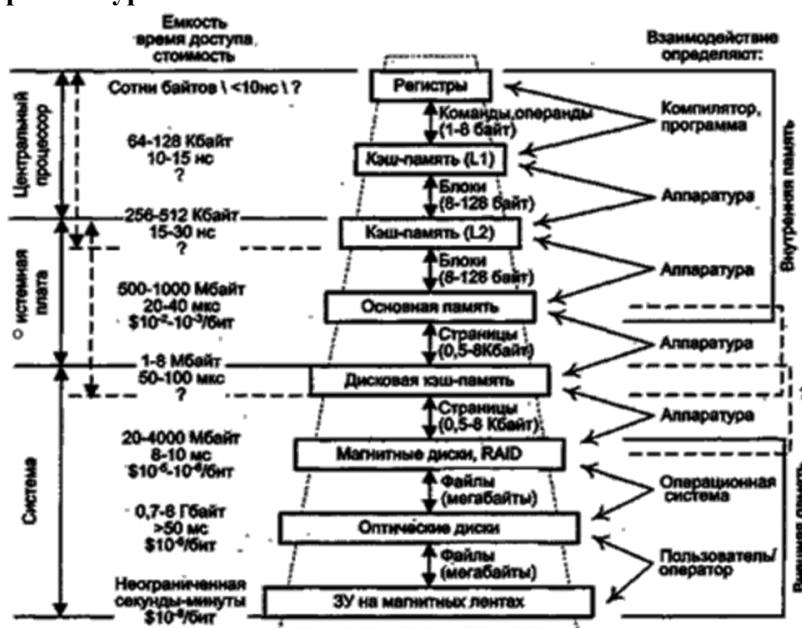
Базовое выражение для расчета среднего времени выборки одного слова из подсистемы памяти:

Тогда:

$$T_{cp} = 0,1 \times 0,95 + (0,1 + 1) \times (1 - 0,95) = 0,15 \text{ (мкс)}$$

65. Как организована внутренняя память компьютера?

Иерархическая архитектура памяти



Память ЦП и внутренняя память

1. **Регистры ЦП** (СОЗУ – сверхоперативное ЗУ):
 - Самый быстрый, но и минимальный по емкости тип памяти;
 - количество регистров невелико* (как правило).
2. **Основная память:**
 - по емкости значительно превосходит регистры ЦП;
 - располагается несколькими уровнями ниже.
3. **Кэш-память:**
 - между регистрами ЦП и основной памятью;
 - проигрывает ОП по емкости, но существенно превосходит по быстродействию (уступая СОЗУ).

Кэш-память

В большинстве современных компьютеров имеется несколько уровней кэш-памяти, которые обозначают буквой L и номером уровня кэш-памяти. При этом все чаще встречается уровень L3; существует мнение о целесообразности добавления уровня L4.

Каждый последующий уровень кэш-памяти по сравнению с предыдущим имеет:

- большую емкость;
- меньшее быстродействие*.

Дисковая кэш-память

Дополнительный уровень иерархии между ОП и дисками. Реализуется в виде самостоятельной ЗУ, включаемого в состав магнитного диска. Существенно повышает производительность при обмене информацией между дисками и основной памятью.

Прочие виды уровней иерархии памяти

Иногда находят применение.

Некоторые модели компьютеров IBM включают в себя *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

66. Что представляет собой основная память компьютера?

Единственный вид памяти (за исключением регистров), к которой ЦП может обращаться непосредственно. Информация, хранящаяся на внешних ЗУ, становится доступной процессору только после того, как будет переписана в ОП.

Аппаратная основа: ЗУ с произвольным доступом.

Элементная база меняется вследствие успехов в области полупроводниковых технологий:

- ранее: электронные лампы, ферритовые кольца, линии задержки;
- сейчас: полупроводниковые микросхемы.

Обычно: ЗУ содержит множество одинаковых запоминающих элементов (ЗЭ), образующих запоминающий массив (ЗМ). Массив разделен на отдельные ячейки.

Каждая ячейка:

- содержит фиксированное число ЗЭ;
- предназначена для хранения двоичного кода, число разрядов в котором определяется шириной выборки памяти.

67. Какие типы ЗУ используются при построении основной памяти?

В зависимости от методов размещения и поиска информации в ЗМ: адресная; ассоциативная; стековая (магазинная).

По типу выполняемых операций:

- **оперативные (ОЗУ, RAM – Random Access Memory, память с произвольным доступом*)**;
- **постоянные (ПЗУ, ROM – Read-Only Memory, память «только для чтения»)**.

Оперативное ЗУ

- допускает запись/считывание информации;
- обе операции выполняются однократно, практически с одной и той же скоростью;
- операции производятся с помощью электрических сигналов.

Преимущественная доля ОП. Большинство типов полупроводниковых ОЗУ – энергозависимые (volatile memory) → используются только как временная память. По сравнению с энергонезависимыми типами ОЗУ (non-volatile memory):

- большая емкость;
- ниже энергопотребление;
- выше быстродействие;
- ниже удельная стоимость.

Подгруппы ОЗУ: динамическая память (DRAM – Dynamic RAM); статическая память (SRAM – Static RAM).

Запоминающие элементы оперативной памяти

Характеристика	Тип	
	SRAM	DRAM
Время хранения информации	не ограничено (при наличии питающего напряжения)	достаточно короткий промежуток → требуется восстановление
Организация ЗЭ	триггер (4 или 6 транзисторов)	1 конденсатор + 1 запирающий транзистор

Регенерация – периодическое (2-8 мс) восстановление заряда ЗЭ.

Постоянное ЗУ

Энергонезависимое ЗУ. Обеспечивает только считывание информации. Некоторые допускают изменение информации, но процесс записи («программирование»):

- сильно отличается от считывания;
- требует значительно большего времени.

Полупроводниковые ИМС по возможностям и способу программирования разделяют на:

- программируемые при изготовлении;

- однократно программируемые (после изготовления);
- многократно программируемые.

Энергонезависимые ОЗУ

NVRAM – Non-Volatile RAM. Несколько типов. Отличие от перепрограммируемых ПЗУ: запись новой информации не требует предварительного стирания → вместо термина «программирование» употребляют стандартный термин «запись».

68. Каким образом организован запоминающий массив, и какие существуют способы организации памяти?

Обычно: ЗУ содержит множество одинаковых запоминающих элементов (ЗЭ), образующих запоминающий массив (ЗМ)

Массив разделен на отдельные ячейки.

Каждая ячейка:

- содержит фиксированное число ЗЭ;
- предназначена для хранения двоичного кода, число разрядов в котором определяется шириной выборки памяти.

69. Объясните функционирование памяти при адресной организации.

Размещение и поиск информации в ЗМ основаны на использовании **адреса хранения слова**.

Адрес = номер ячейки ЗМ.

При записи/считывании слова в ЗМ инициирующая эту операцию команда должна указывать адрес, по которому производится запись/считывание.

70. Как реализуется цикл обращения в памяти с адресной организацией?

Цикл обращения к памяти инициируется поступлением в БУП извне сигнала «Обращение».

Общая часть цикла обращения:

- прием в РГА с шины адреса (ША) адреса обращения;
- прием в БУП и расшифровка управляющего сигнала «Операция», указывающего вид запрашиваемой операции (чтение или запись).

При считывании:

- БАВ дешифрирует адрес, посылает сигналы считывания в заданную адресом ячейку ЗМ;
- код записанного в ячейке слова считывается усилителями считывания БУС и передается в РГИ;
- слово из РГИ поступает на выходную информационную шину ШИВых. При записи:
- прием записываемого слова с входной информационной шины ШИВх в РГИ; ◦ запись слова из РГИ в выбранную БАВ ячейку.

71. С какой целью используется блочная организация памяти, и какие существуют ее виды?

Необходимость объединения нескольких ИМС (интеграл.микросхем) ЗУ возникает также, когда разрядность ячеек в микросхеме ЗУ меньше разрядности слов ВМ.

Виды блочной памяти:

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

Пример: память общей емкостью 512 слов (29) состоит из четырех банков по 128 слов в каждом.

72. Что означает понятие «блочная структура ОП»?

Структура которая содержит несколько банков памяти

Банк – совокупность или 1 модуль памяти

Модуль - совокупность микросхем

73. Объясните смысл понятий «расслоение памяти» и «чередование адресов».

Расслоение памяти

Назначение: обеспечение параллельного доступа к нескольким банкам.

Основа: механизм чередования адресов (address interleaving) – циклическое распределение адресов между банками. Используемое свойство: локальность по обращению.

В каждом такте на ША может присутствовать адрес только одной ячейки → обращение:

- параллельно к нескольким банкам – невозможно;
- со сдвигом на один такт – возможно.

Чередование адресов – циклическая структура. Разряды адреса:

- A1, A0 – выбор банка;
- A8–A2 – выбор ячейки в банке.

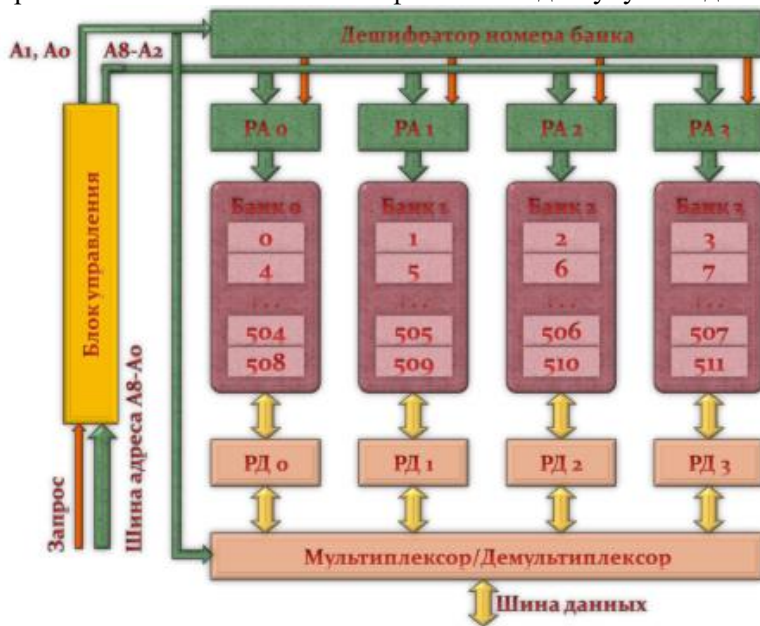
Адрес ячейки запоминается в индивидуальном регистре адреса (РА), и дальнейшие операции по доступу к ячейке в каждом банке протекают независимо.

Эффективность циклической структуры

При большом количестве банков среднее время доступа к ОП сокращается почти в В раз, при условии, что ячейки, к которым производится последовательное обращение, относятся к разным банкам.

Конфликт по доступу – запросы к одному и тому же банку следуют друг за другом → каждый следующий запрос должен ожидать завершения обслуживания предыдущего.

При частом возникновении конфликтов по доступу метод становится неэффективным.



74. Объясните логическую организацию и функционирование блочно-циклической схемы памяти.

Блочно-циклическая структура

Усовершенствованная схема расслоения памяти:

- каждый банк состоит из нескольких модулей, адресуемых по циклической схеме;
- адреса между банками распределены по блочной схеме.

Расслоение памяти в многопроцессорных системах

Традиционные способы эффективны в рамках одной задачи, для которой характерно свойство локальности.

Проблема: в многопроцессорных системах с общей памятью запросы на доступ к памяти достаточно независимы.

Развитие идеи расслоения памяти: несколько контроллеров памяти в системе → отдельные банки работают независимо.

Эффективность: зависит от частоты независимых обращений к разным банкам. Лучший результат – при большом числе банков, т.к. уменьшается вероятность последовательных обращений к одному и тому же банку памяти.

Пример: ОП суперкомпьютера NEC SX/3 состоит из 128 банков.

75. Перечислите виды и укажите назначение специализированных ОЗУ.

К специализированным относят: память для видеоадаптеров; память с множественным доступом (многопортовые ОЗУ); память типа очереди (ОЗУ типа FIFO).

Два последних типа относятся к статическим ОЗУ.

76. Каким образом реализуется многопортовое ОЗУ?

Многопортовые ОЗУ

В n -портовом ОЗУ имеется n независимых наборов шин адреса, данных и управления, гарантирующих одновременный и независимый доступ к ОЗУ n устройствам.

Свойства:

- существенно упрощается создание компьютерных систем с общей (совместно используемой) памятью;
- в рамках отдельного компьютера более эффективный, по сравнению с ПДП, обмен информацией между ЦП и ПУ.

77. Каким образом реализуется ОЗУ типа FIFO?

Алгоритм очереди. Запросы обслуживаются в порядке очереди (FIFO), образовавшейся к моменту начала цикла арбитража. Определенные сложности аппаратной реализации. Используется редко.

Назначение: буферизация потока данных.

Поступление и выборка данных:

- одинаковая последовательность;
- различная скорость (обычно);
- возможность одновременного выполнения.

ИМС – двухпортовое ОЗУ:

- первый порт – занесение информации, второй – считывание;
- механизмы, свойственные двухпортовой памяти (способы арбитража и т.п.).

Особенности:

- отсутствуют адресные входы;
- занесение и считывание данных – через одну входную точку и одну выходную;
- для слежения за состоянием очереди ИМС содержит:
- регистры-указатели адресов начала и конца очереди;
- флаги-индикаторы состояний отсутствия данных и заполненной памяти.

78. Объясните логическую организацию ассоциативной памяти.

Устройство, способное хранить информацию, сравнивать ее с некоторым заданным образцом и указывать на их соответствие или несоответствие друг другу. Поиск информации: не по адресу, а на основании какого-нибудь характерного признака (ассоциативный признак), реализуемого как:

- часть искомой информации;
- дополнительно придаваемый ей (тег или ярлык).

Признак поиска – кодовая комбинация, выступающая в роли образца для поиска.

ЗМ: N ячеек разрядностью $(n+1)$. Разряд n – индикатор занятости ячейки (0 – свободна, 1 – хранит слово).

РГАП – регистр ассоциативного признака. Хранит признак поиска; разрядность $\leq n$.

ССв – схемы совпадения:

- параллельное сравнение каждого бита всех хранимых слов с соответствующим битом признака поиска;
- выработка сигналов совпадения

РгСв – регистр совпадений: каждый разряд соответствует ячейке ЗМ (1 – совпадение всех разрядов ячейки с одноименными разрядами признака поиска).

РгМ – регистр маски. Запрет сравнения определенных битов.

ФС – комбинационная схема, формирующая сигналы совпадения на основании анализа содержимого РгСв.

79. Каким образом функционирует АЗУ?

РгМ: обнуляются разряды, которые не должны учитываться при поиске информации.

РгСв: разряды устанавливаются в 1. РГАП: заносится код признак поиска.

Поиск: ССв одновременно сравнивают первый бит всех ячеек ЗМ с первым битом признака поиска.

Схемы, зафиксировавшие несовпадение, формируют сигнал, переводящий соответствующий бит РгСв в 0.

Поиск повторяется для остальных незамаскированных битов признака поиска.

Итог: единичные разряды РгСв соответствуют ячейкам, содержащим искомую информацию; используются в качестве «адресов», по которым производится считывание из ЗМ.

Результаты поиска могут оказаться неоднозначными → содержимое РгСв подается на комбинационную схему, формирующую осведомительные сигналы о нахождении информации:

- α_0 – ни в одной ячейке;
- α_1 – только в одной ячейке;
- α_2 – более чем в одной ячейке.

Формирование содержимого РгСв и сигналов α_0 , α_1 , α_2 – операция контроля ассоциации:

- является составной частью операций считывания и записи;
- может иметь самостоятельное значение.

Считывание

Сначала: контроль ассоциации по аргументу поиска. Затем:

- при $\alpha_0=1$ считывание отменяется из-за отсутствия искомой информации;
- при $\alpha_1=1$ считывается слово, на которое указывает единица в РгСв;
- при $\alpha_2=1$ сбрасывается самая старшая единица в РгСв и извлекается соответствующее ей слово; операция повторяется последовательно для всех найденных слов.

Запись

Производится в первую свободную ячейку. Поиск свободной ячейки:

- выполняется операция считывания, в которой не замаскированы только служебные разряды, показывающие, как производилось обращение к данной ячейке;
- свободной считается либо пустая ячейка, либо та, которая дольше всего не использовалась.

Свойства ассоциативных ЗУ

Главное преимущество: время поиска:

- зависит только от числа разрядов в признаке поиска и скорости опроса разрядов;
- не зависит от числа ячеек в ЗМ.

Разновидности архитектур АЗУ

Архитектура определяется сочетанием четырех факторов:

- вид поиска информации;
- способ сравнения признаков;
- способ считывания информации при множественных совпадениях;
- способ записи информации.

80. Объясните назначение контроллера ОП.

Контролер выдает сигналы ввода и вывода (Управление операциями с ОП)

81. Чем обусловлена необходимость и каковы на-правления защиты памяти компьютера?

В многопользовательском и многозадачном режимах:

- в ОП одновременно находятся: программы, относящиеся к разным пользователям и/или к различным задачам одного пользователя; фрагменты ОС;
- каждой задаче в ОП выделяется свое адресное пространство.

Задача: предотвращение разрушения одних программ другими

Защита областей ОП:

- одной программы от попыток записи со стороны других программ;
- от чтения (например, при ограничениях на доступ к системной информации);
- от использования содержимого памяти не по назначению.

82. Каким образом реализуется метод колец защиты памяти?

Метод колец защиты

Защита адресного пространства ОС от НСД со стороны пользовательских программ.

Реализуется за счет:

- наличия двух режимов работы ЦП:
 - системного (режим супервизора – «надзирателя»);
 - пользовательского;
- аппаратной поддержки уровней привилегий (системного и пользовательского*).

Кольца защиты – изображение системы защиты в виде концентрических окружностей:

- внешняя – пользовательский режим;
- внутренняя – системный.

Число уровней привилегий (колец защиты) может быть больше двух.

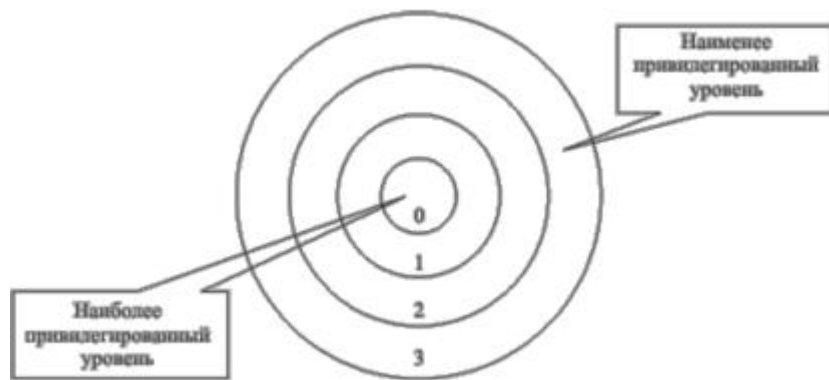
Кольца защиты микропроцессоров Intel x86-x64

0 – ядро ОС;

1 – основная часть программ ОС (утилиты);

2 – служебные программы ОС (драйверы, СУБД и др.);

3 – прикладные программы пользователя.



83. Каким образом реализуется метод граничных регистров памяти?

Метод граничных регистров

Наиболее распространен. В ЦП предусмотрены два регистра, содержимое которых:

- определяет нижнюю и верхнюю границы области памяти, куда программа имеет право доступа;
- формируется ОС при загрузке программы.

При каждом обращении к памяти проверяется, попадает ли используемый адрес в установленные границы.

При нарушении границы: доступ блокируется; формируется запрос прерывания.

Нижнюю границу разрешенной области памяти определяет сегментный регистр.

Верхняя граница подсчитывается ОС в соответствии с размером размещаемого в ОП сегмента.

Требуется наличие двух режимов работы ЦП (привилегированного и пользовательского).

Запись информации в граничные регистры возможна лишь в привилегированном режиме.



84. Как реализуется защита памяти по ключам?

Метод ключей защиты

Позволяет защитить несмежные области памяти.

Реализация:

- память делится на блоки одинакового размера;
- каждому блоку ставится в соответствие некоторый код – ключ защиты памяти;
- каждой программе присваивается ключ защиты программы;
- распределением ключей защиты ведаёт ОС: ключ защиты программы хранится в ССП; ключи защиты памяти хранятся в специальной памяти.



Условия доступа программы к конкретному блоку памяти:

- совпадение ключей защиты памяти и программы (аппаратное сравнение);
- равенство одного из ключей нулю.

Нулевое значение ключа защиты программы:

- разрешает доступ ко всему адресному пространству;
- используется только программами ОС.

При несовпадении реакция зависит от вида запрещенного доступа (запись, чтение или оба); если доступ запрещен, формируется запрос прерывания.

85. Что положено в основу концепции RAID-массивов?

RAID – набор физических дисковых ЗУ:

- управляемых ОС и рассматриваемых как один логический диск;
- распределение данных по физическим дискам;
- избыточность дискового пространства → гарантия восстановления данных в случае отказа диска.

Схемы (уровни) RAID-систем:

- базовые: RAID 1, RAID 2, ..., RAID 5;
- RAID 0, RAID 6 – введены ассоциацией производителей RAID-систем (RAB – RAID Advisory Board);
- комбинации стандартных уровней.

86. За счет чего в RAID-массивах повышается отказоустойчивость и производительность?

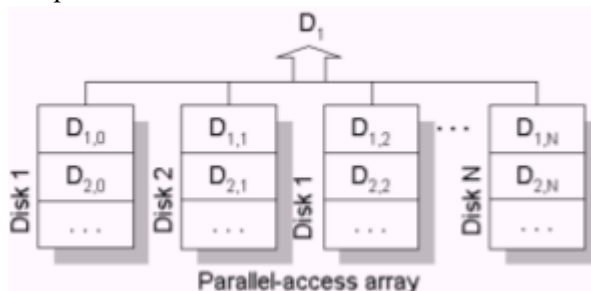
Повышение производительности дисковой подсистемы

Расслоение или расщепление (striping):

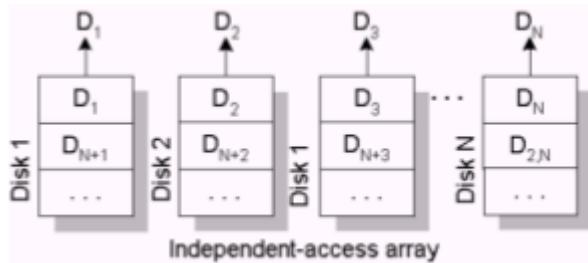
- полосы (strip – узкая полоса) – сегменты, на которые разбиваются данные и дисковое пространство;
- размер (ширина) – бит, байт, физический сектор МД, дорожка;
- распределяются по различным дискам массива;
- пояс (stripe – широкая полоса) – набор полос, одинаково расположенных на каждом диске массива.

Параллельное считывание или запись нескольких полос возможно, если они расположены на разных дисках.

1. Массив с параллельным доступом. Parallel-access array; эффективен при передаче данных большого размера.



2. Массив с независимым доступом. Independent-access array; эффективен при большом количестве малых запросов.



Повышение отказоустойчивости дисковой подсистемы

Среднее время безотказной работы (mean time to failure) массива, при независимых отказах и экспоненциальном законе распределения наработки на отказ:

$$MTTF_{\text{мас}} = MTTF_{\text{МД}} / N_{\text{МД}},$$

где $MTTF_{\text{МД}}$ – среднее время безотказной работы одного диска; $N_{\text{МД}}$ – количество дисков.

Обнаружение и коррекция ошибок за счет избыточного дискового пространства. В RAID – избыточность трех видов.

Дублирование. Экземпляры одних и тех же данных расположены на разных дисках массива. Обычно: разбивка массива на пары дисков. Избыточность – 100%.

Код Хэмминга. Вычисляется для каждого пояса. Корректирующие биты хранятся на дополнительных дисках (один диск на каждый бит). Избыточность – для массива из десяти МД около 30%.

Биты паритета. Для каждого пояса вычисляется контрольная полоса (биты паритета). Отдельный бит паритета – сумма по модулю два одноименных битов всех контролируемых полос. Требуется один дополнительный диск. Избыточность – для массива из четырех МД около 25%.

87. Какие уровни RAID обеспечивают повышение производительности, а какие – надежности?

Повышение производительности

RAID 0

RAID 3

RAID 5

RAID 7

RAID 01

RAID 50

RAID 5E

RAID 5EE

Повышение надежности

RAID 1

RAID 2

RAID 4

RAID 6

RAID 10

RAID 1E