Часть I
Лекция 1 (08.02.16)

1. Преподаватель: Ковалев Сергей Николаевич (kovalyev@mirea.ru)
2. Информация о семестре
   (a) На каждом практическом занятии пишутся контрольные работы (по желанию)
      i. Можно получить экзамен автоматом (средняя оценка за контрольные)
      ii. На контрольных можно использовать что угодно
   (b) Курсовая работа
      i. Задание нужно утвердить до 5 недели включительно
   (c) Экзамен

1 Основные понятия

1. **Вычислительной машиной** будем называть комплекс тенических и программных средств, предназначенных для подготовки и решения задачи пользователей.
2. **Вычислительной системой** будем называть совокупность взаимосвязанных и взаимодействующих вычислителей (процессоров или вычислительных машин), периферийного оборудования и программного обеспечения, предназначенную для подготовки и решения задачи пользователей.
3. **Архитектурой ЭВМ** будем называть основополагающую концепцию технического устройства и организации логического взаимодействия компонентов вычислительной машины.

1.1 Структура вычислительной машины

1. Рассматривая структуру вычислительной машины следует определиться со степенью глубины оценки тех или иных аспектов устройства и работы вычислительной машины. Эту степень углубленности принято называть **уровнем детализации**.

2. Первый уровень (**уровень черного ящика**): Некий объект, способный хранить и обрабатывать информацию и обмениваться ею с пользователем и другими системами

3. На следующем уровне детализации, называемом **уровнем общей архитектуры**, вычислительная машина рассматривается как комплекс основных подсистем, обычно включающий в себя устройства ввода и вывода, память, центральный процессор и средство взаимосвязи этих подсистем.

4. На третьем уровне рассматривается устройство не всей вычислительной машины, а её отдельной конкретной подсистемы.

5. На следующем уровне детализируется устройство отдельного компонента подсистемы.

1.2 Концепции машины в рамках вычислительных программ

1. **Вычислительная машина** — совокупность технических и программных средств, служащих для обработки данных по заданному алгоритму

2. **Алгоритмом** будем называть конечный набор предписаний, определяющий решение задач посредством конечного количества операций.

   (a) Дискретность — алгоритм разбивается на определенные действия, которые тоже дискретны

   (b) Определенность — в алгоритме указано всё, что необходимо сделать, причем ни одно из действий не может трактоваться двояко

   (c) Массовость — применимость алгоритма к множеству наборов исходных данных, а не к отдельным уникальным значениям

   (d) Результативность — возможность получения результата за конечное число шагов

3. Свойства алгоритма предопределяют возможности их реализации на вычислительной машине. При этом процесс, порождаемый алгоритмом, называется вычислительным процессом.

Часть II
Лекция 2 (15.02.16)

1. В течение продолжительного времени архитектура вычислительных машин основывалась на представлении алгоритма решения задачи в виде программы последовательных вычислений. На сегодняшний день основные архитектурные изменения коснулись перехода к параллельным вычислениям, при которых ряд действий, предусмотренных выполняемой программой, выполняется одновременно.

2. Согласно стандарту, **программа** вычислительной машины определяется как упорядоченная последовательность команд, подлежащих обработке.

3. Вычислительная машина, в памяти которой определенным образом закодированы команды программы, получила название **вычислительной машины с хранимой в памяти программой**. Сущность фон-неймановской вычислительной машины сводится к четырем принципам:

   (a) **Двоичное кодирование**: вся информация (и команды и данные), обрабатываемая на вычислительной машине, кодируется нулями и единицами. Каждый вид информации представляется последовательностью двоичных символов (битов) и имеет свой формат.

      i. **Кодирование** — способ представления информации сигналами или состояниями объектов той или иной физической природы.

      ii. Последовательность битовых значений в форматах, имеющая определенный смысл, называется **полем**.

         A. В формате команды, как правило, выделяется два поля: поле кода операции и поле адресов

   (b) **Программное управление**: Все действия, предусмотренные алгоритмом решения задачи, должны быть представлены в виде программы, состоящей из последовательности управляющих слов — команд. Каждая команда предписывает некоторую операцию из набора операций, реализуемых конкретной вычислительной машиной.

      i. Команды программы заносятся и хранятся в ячейках памяти вычислительной машины и выполняются в естественной последовательности, т.е. в порядке их расположения в программе. При необходимости, естественная последовательность выполнения команд может быть изменена с помощью специальных команд, при-

чем решение об изменении последовательности может приниматься либо на основании анализа результатов выполнения предыдущих команд, либо безусловно.

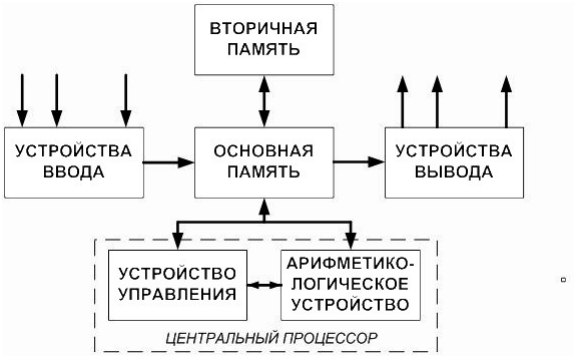
(c) **Принцип однородности памяти**: И команды и данные хранятся в одних и тех же областях памяти и извне неразличимы. Различаются команды и данные только по способу использования соответствующей информации.

i. Архитектура вычислительной машины с однородной памятью разрабатывалась в Принстонском университете и получила название Принстонской архитектуры. Параллельно разрабатывалась архитектура с раздельной памятью, в которой команды и данные хранятся в разных областях (Гарвардская архитектура). В данный момент Гарвардская архитектура реализуется в отдельных видах памяти вычислительных машин.

(d) **Принцип адресности**: Память вычислительной машины состоит из пронумерованных ячеек, причем центральному процессору в любой момент времени доступно содержимое любой ячейки памяти.

i. Номера ячеек иначе называют **адресами**.

2 Структура фон-неймановской вычислительной машины



1. Вся информация, обрабатываемая вычислительной машиной, прежде, чем стать доступной для выполенния над ней определенных действий, должна быть размещена в основной памяти. В основную память информация

поступает с внешних (периферийных) устройств через устройства (модули) ввода. Информация, размещенная в основной памяти, обрабатывается центральным процессором. Результат обработки вновь возвращается в основную память и оттуда, через устройства (модули) вывода, подается на внешние устройства.

2. Особенностью основной памяти является её энергозависимость, т.е. информация в её ячейках сохраняется только в процессе работы.

3. Информация, предназначенная для долговременного хранения с целью её последующего использования, через основную память переносится во вторичную (дополнительную, внешнюю) память. Устройства внешней памяти энергонезависимы.

4. **Устройство управления** решает две ключевые задачи:

   (a) управляет процессом обработки пользовательской информации

   (b) синхронизирует взаимодействие всех структурных элементов вычислителльной машины в составе единого комплекса

5. **Арифметико-логическое устройство** осуществляет непосредственное преобразование информации согласно выполняемым программам.

6. Зачастую, давая общую характеристику вычислительной машины, устройства ввода и вывода объединяют в единую подсистему **ввода-вывода**.

7. В современных вычисллительных машинах устройство управления и арифметико логическое устройство чаще всего реализуется в едином конутруктивном узле, получившем название **центрального процессора**.

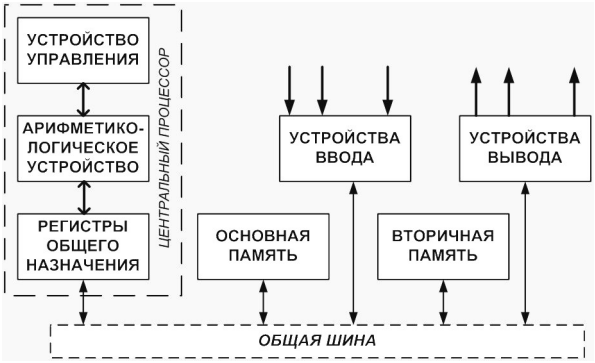
8. Взаимодействие структурных компонентов вычислительной машины возможно реализовать по-разному. Способ организации взаимодействия называют **структурой связей**.

   (a) В классической фон-неймановской вычислительной машине реализована структурой с **непосредственными связями**.

      i. Основное достоинство такой структуры заключается в возможности обеспечения высокой производительности за счет использования индивидуальных каналов связи между различными структурными подсистемами.

      ii. Принципиальный недостаток заключается в невозможности модификации отдельных связей, не затрагивая остальные связи.

(b) Альтернативой структуре с непосредственными связями выступает еще одна организация взаимодействия подсистем на основе использования **шин**. При такой структуре связи взаимодействие между всеми подсистемами осуществляется через информационную магистраль, называемую **общей шиной**.

i. 

ii. Такой подход обеспечивает высокую гибкость в плане возможности отдельной поэтапной модификации конкретных связей, не затрагивая систему в целом.

iii. Принципиальный недостаток заключается в ограниченной пропускной способности общей шины, примерно 90% которой захватывает обмен между центральным процессором и основной памятью.

(c) В технике современных вычислительных машин (персональных компьютеров в том числе) реализуется связь не на основе единственной шины, а на основе иерархии шин разного функционального назначения.

Часть III

Лекция 3 (29.02.16)

3 Устройство управления вычислительной машиной (УУ)

1. **Устройство управления** реализует функцию управления в вычислительных процессах, обеспечивая автоматической выполнение команд про-

граммой.

2. **Вычислительный процесс** представляет собой последовательность машинного цикла. В ходе типового машинного цикла устройством управления реализуется определенное количество функций из типового набора. Эти функции получили название **целевых функций** устройства управления.

3.1 Целевые функции устройства управления

1. Каждая целевая функция реализуется в ходе некоторого условного этапа. Машинный цикл разбит на очередность этапов.

2. Этапы типового цикла:

   (a) Выборка команды из памяти

      i. Подэтап: Декодирование операционной части команды

   (b) Формирование адреса следующей команды

   (c) Формирование исполнительного адреса операнда, либо адреса перехода для отдельного вида команд

      i. Имеет столько модификаций, сколько способов адресации предусмотрено в системе команд отдельной машины.

   (d) Выборка операнда из памяти по исполнительному адресу, сформированному на предыдущем этапе

   (e) Выполнение операции, соответствующей выполняемой команде

      i. Имеет столько модификаций, сколько операций предусмотрено в системе команд отдельной машины

3. В зависимости от выполняемых команд может меняться количество этапов и количество функций. Минимальное число реализуемых функций составляет 3 целевых функции

3.2 Модель устройства управления

1. Центральным звеном устройства управления является **микропрограммный аппарат**.

2. Для выполнения своих функций, устройство управления должно иметь:

   (a) входы, на которые поступают сигналы, позволяющие оценить текущее состояние управляемой системы

   (b) выходы, на которые подаются сигналы, управляющие поведением системы

3. Из оперативной памяти выполняемая команда заносится в **регистр команды**.

4. **Флаги** — сигналы, используемые устройством управления для оценки состояния центрального процессора и результата управления предыдущей операции.

5. **Тактовые индексы** — инициируют выполнение одной или нескольких микроопераций

6. **Сигналы из системной шины** — служебная информация, характеризующая текущее состояние всех систем вычислительной машины, за исключением центрального процессора, в том числае сигналы прерываний, подтверждений и приоритетов.

   (a) **Прерыванием** называется сигнал (последовательность сигналов), изменяющий естественный ход выполнения команд программы.

      i. Аппаратные прерывания — сигналы от конкретных узлов или систем вычислительной машины (клавиатура).

      ii. Программные прерывания — сигналы, формируемые как результат выполнения очередной команды программы.

      iii. Логические прерывания — сигналы, формируемые как результат логической обработки некоторого множества выполненных операций.

   (b) **Сигналы приоритетов** формируются в ситуациях, когда несколько устройств или систем одновременно хотят изменить текущее состояние вычислительной машины. Для предотвращения возможных конфликтов, используется система приоритетов.

3.3 Обобщенная структура устройства управления

1. Вычислительный процесс состоит из последовательностей элементарных действий, выполняемых в узлах вычислительной машины. Такие элементарные действия по преобразованию информации выполняются в течение одного такта сигнала синхронизации и называются **микрооперациями**.

2. Тактовые импульсы генерируются специальным источником, называемым **тактовым генератором**, высокостабильной частоты сигнала.

3. Совокупность сигналов управления, вызывающих одновременно выполняемые микрооперации, образует **микрокоманды**. В свою очередь, последовательность микрокоманд, определяющая содержание и порядок реализации конкретного машинного цикла, называется **микропрограммой**.

4. Микрокоманды обрабатываются центральным узлом устройства управления.

5. Микропрограммы реализаций тех или иных целевых функций устройства управления инициируются задающим оборудованием, которое вырабатывает требуемую последовательность сигналов управления и входит в состав управляющей части устройства управления.

6. Микропрограммы выполняются исполнительным оборудованием, входящим в состав основной памяти для реализации целевых функций выборки команды и выборки операнда. Другая часть исполнительного оборудования входит в состав операционных устройств АЛУ (арифметико-логических устройств). Для целевой функции формирования адреса следующей команды и формирования исполнительного адреса исполнительным оборудованием выступает адресная часть устройства управления.

7. В управляющую часть устройства управления входят

   (a) Регистр команд

      i. Адресная часть
      ii. Операционная часть

   (b) Микропрограммный автомат

   (c) Узел прерываний и приоритетов

Часть IV
Лекция 4 (14.03.16)

4 Арифметико-логическое устройство

4.1 Операционные устройства вычислительной машины

1. **Арифметико-логическое устройство** (АЛУ) классической вычислительной машины выполняет функции арифметической и логической обработки данных.

2. Учитывая разнообразие выполняемых операций и множество типов обрабатываемых данных, на практике речь следует вести не оедином устройстве, а о комплексе специализированных устройств. Каждое из этих операционных устройств реализует определенное подмножество арифметических или логических операций, предусмотренных системой команд конкретной вычислительной машины.

3. С этих позиций принято выделять операционные устройства:

   (a) целочисленной арифметики

   (b) реализации логических операций

   (c) десятичной арифметики

   (d) для чисел с плавающей точкой

4. На текущем этапе в реализации вычислительных машин две первые группы объединяются в одно устройство.

5. Специализированные операционные устройства десятичной арифметики сейчас практически не используются, посколько данные, представленные в двоично-десятичном формате, успешно обрабатываются средствами двоичной арифметики.

6. Таким образом, арифметико-логическое устройство как часть центрального процессора, представлено двумя видами операционных устройств:

   (a) целочисленной арифметики и логических операций

   (b) для чисел с плавающей точкой

7. В минимальном варианте операционные устройства в составе АЛУ должны содержать схемы, реализующие весьма органиченный набор операций. Опираясь на этот набор, можно программным способом обеспечить выполнение множества других операций над различными типами данных.

8. Основная проблема в плане реализации минимально необходимого аппаратурного базиса в сочетании с программными решениями заключается в выборе рационального соотношения между ними, позволяющего, с одной стороны, обеспечить высокую производительность, а с другой снизить избыточность схемного оборудования.

4.2 Структура операционного устройства

1. Набор элементов, на основе которых реализуются операционные устройства, принято называть **структурным базисом**. Структурный базис операционных устройств включает в себя:

   (a) Регистры для хранения машинных слов

      i. **Регистром** принято называть особый вид запоминающих устройств, предназначенных для кратковременного (на время рабочего цикла) хранения выполняемой команды, обрабатываемых данных и результата выполнения.

   (b) Управляемые линии передачи сигналов

   (c) Комбинационные схемы, реализующие микрооперации или логические условия по сигналам устройства управления

2. Среди множества структур, которые можно синтезировать на основе того или иного структурного базиса, отдельно выделяют **каноническую структуру**.

   (a) В такой структуре каждой реализуемой микрооперации соответствует вполне определенный элемент структурного базиса.

   (b) Имея наибольшую производительность среди всех структур возможных структур, каноническая обладает огромной избыточностью и на практике никогда не реализуется.

3. Реализуемые на практике структуры операционных устройств можно разделить на две принципиально разные группы:

   (a) Операционные устройства **с жесткой структурой**

i. Все операционные схемы жестко распределены между имеющимися регистрами, т.е. к каждому регистру относится свой набор комбинационных схем, позволяющих реализовать соответствующий набор микроопераций.

(b) Операционные устройства **с магистральной структурой**

i. Все регистры объединены в отдельный узел, который принято называть **регистрами общего назначения**. А все комбинационные схемы объединены в **операционный блок**.

ii. Для взаимодействия регистров общего назначения с комбинационными схемами используется информационная магистраль, т.е. линия передачи сигналов от регистров общего назначения к операционным блокам.

iii. РОН имеет множество выходов (по числу регистров в его составе), которые подключаются к магистрали через электронный коммутатор, называемый **мультиплексором**. С помощью мультиплексора к магистрали может быть подключен выход любого регистра. С другой стороны магистрали через другой коммутатор (демультиплексор) подключены входы различных комбинационных схем.

iv. Использование магистрали с (де)мультиплексорами дает возможность подключить к любому регистру любой набор комбинационных схем.

v. Операционные устройства с магистральной структурой обладают высокой универсальностью и регулярностью, что наряду с рациональным соотношением аппаратных и программных компонентов позволяет проще реализовать аппаратные составляющие на существующей элементой базе.

5 Память вычислительных машин

5.1 Характеристики памяти

1. В любой вычислительной машине, вне зависимости от ее архитектуры, команды и данные хранятся в памяти.

2. Функции памяти обеспечиваются запоминающими устройствами (ЗУ), предназначенными для фиксации, хранения и выдачи информации.

   (a) Процесс фиксации принято называть **записью**.

(b) Процесс выдачи принято называть **чтением**.

(c) Совместно оба этих процесса принято характеризовать как процесс обращения к запоминающему устройству.

3. Запоминающее устройство в составе вычислительной машины характеризуется множеством признаков и показателей. Среди них наиболее часто используются следующие:

   (a) Место расположения:

      i. Процессорные ЗУ, которые, функционально являясь элементами памяти, архитектурно выполняются как неотъемлемая часть процессора.

      ii. Внутренние ЗУ, которые реализованы в элементах и узлах основной памяти.

      iii. Внешние ЗУ, которые реализованы в элементах и узлах внешней памяти.

   (b) Методы доступа: Конкретный метод доступа во многом определяет скорость работы запоминающего устройства.

      i. **Последовательный** доступ: применяется в запоминающих устройствах, информация в которых хранится в виде последовательности логически завершенных информационных единиц, называемых **записями**.

         A. В ЗУ, хранящих двоичную дискретную информацию, запись имеет фиксированный размер и состоит из набора последовательно нумеруемых (адресуемых) запоминающих элементов.

         B. При обращении к запоминающему элементу, хранящему требуемую информацию, предварительно просматриваются все запоминающие элементы во всех записях, предшествующих искомому.

         C. Время доступа не фиксировано и зависит от места расположения искомого запоминающего элемента.

         D. В нынешних вычислительных машинах практически не используется. Ранее применялся во внешней памяти, в запоминающих устройствах на магнитной ленте и магнитной проволоке.

      ii. **Прямой** доступ: ЗУ так же содержит информацию в виде записей (формат нефиксированной записи), но начало каждой записи имеет собственный физический адрес.

A. Обращение по прямому методу происходит в два этапа: на первом по физическому адресу ищется запись, содержащая искомый запоминающий элемент, а на втором внутри этой записи последовательным перебором ищется запоминающий элемент с искомой информацией.

B. Время обращения не фиксировано, но изменяется в небольшом диапазоне в зависимости от положения искомого элемента внутри записи.

C. Метод применяется, в основном, в дисковых запоминающий устройствах внешней памяти.

iii. **Произвольный** доступ (RAM): Все запоминающие ячейки имеют уникальный физический адрес, по которому любая из них доступна для обращения со стороны процессора в любое время и в произвольном порядке.

A. Время обращения фиксированное.

B. Метод применяется во всех запоминающих устройствах основной памяти за небольшим исключением запоминающих устройств с ассоциативным доступом. Также применяется в полупроводниковых устройствах внешней памяти.

iv. **Ассоциативный** доступ: одновременный просмотр содержимого всех ячеек памяти для обнаружения той, в которой информация побитно совпадает с некоторым заданным шаблоном.

A. Время обращения фиксировано и наименьшее в сравнении с другими методами.

B. Сегодня используется только в кэш-памяти низких уровней.

(c) Примечание:

i. Запоминающий элемент — устройство памяти, предназначенное для хранения одного битового значения (0 или 1).

ii. Запоминающая ячейка — устройство памяти, расчитанное, в общем случае, на хранение несольких битовых значений, т.е. содержит несколько запоминающих элементов.

A. В частном случае запоминающая ячейка может иметь один запоминающий элемент.

(d) Помимо перечисленных характеристик памяти используютс и другие, например:

i. Емкость (объем)

ii. Физический тип

iii. Единицы пересылки
iv. Энергозависимость
v. ...

Часть V
Лекция 5 (21.03.16)

5.2 Структура памяти вычислительной машины на примере памяти ПК

1. Основная память представлена тремя составляющими:
   (a) Процессорные регистры — самые малые по объему хранимой информации и самые быстродействующие устройства памяти любой вычислительной машины.
   (b) Оперативная память (ОЗУ) имеет две составляющие
      i. Динамическая (DRAM), составляющая большую часть (более 90%) оперативной памяти.
      ii. Статическая (Кэш)
         A. При том, что быстродействие и процессора и устройства динамической оперативной памяти постоянно возрастает, разрыв между темпами нарастания производительности неуклонно увеличивается. Именно такой разрыв в нарастании быстродействия обусловил появление статической составляющей оперативной памяти, много меньшей по объему, чем динамическая, но обладающей быстродействием, сопоставимым с быстродействием процессора. Поэтому кэш используется в роли буферной памяти, согласующей работу высокопроизводительного процессора и медленной динамической оперативной памяти.
   (c) Постоянная память (ПЗУ)
      i. Современные ПЗУ, как правило, допускают перепрограммирование (перепрошивку) , но этот процесс принципиально отличается по физике работы от считывания и протекает медленно.
      ii. ПЗУ персональных компьютеров является частично энергонезависимой. Часть информации сохраняется при отключении питания. Информация о пользовательских настройках хранится в

энергозависимой части ПЗУ и при отсутствии питания обнуляется.

2. Вторичная память:
   (a) С магнитным носителем
   (b) С оптическим носителем
   (c) Полупроводниковая

5.3 Организация обращения к устройству оперативной памяти

1. На протяжении развития вычислительных систем использовалось множество подходов к организации обращения к ячейкам динамической оперативной памяти, в том числе:
   (a) **Последовательное** считывание (не путать с последовательным методом доступа)
   (b) **Конвеерное** считывание
   (c) **Регистровое** считывание
   (d) ...

2. Современные реализации динамической памяти в своей основе имеет страничную структуру и метод считывания (обращения) к такой памяти имеет такое же название (страничный).

3. Большинство составляющих оперативной памяти в основе своей логической реализации имеют матричную структуру, в узлах которой расположены запоминающие ячейки. Каждая отдельная запоминающая ячейка содержит набор запоминающих элементов, число которых обычно равно длине машинного слова (1 или 2 байта).

4. Запоминающие ячейки, расположенные в одной строке матрицы, называются **страницами**.

5. В основе страничной организации лежит тот факт, что при доступе к ячейкам со смежными адресами (согласно принципу локальности, такая ситуация наиболее вероятна), причем при доступе к таким, где все адресуемые ячейки принадлежат одной странице, доступ ко второй и последующей адресуемым ячейкам можно организовать значительно быстрее, чем к первой адресуемой ячейке.

6. Если адрес строки (номер страницы) при следующем обращении остается прежним, то все временные затраты, связанные с повторным занесением адреса строки в соответствующий регистр микросхемы данных, дешифровкой этого адреса и другими вспомогательными операциями, можно исключить, сохраняя прежний адрес и номер страницы и задавая только новое значение номера столбца.

7. Наряду с простым страничным методом существует его модификация, получившая название **быстрого страничного доступа**. Отличается от рассмотренного только иным подходом к формированию сигналов.

8. Существует еще одна модификация страничного метода, получившая название **пакетного режима обращения**. В этом режиме, при обращении к некоторой запоминающей ячейке, в выходной регистр микросхемы поступает не только информация из запрошенной ячейки, но и из нескольких следущих с последовательной нумерацией адресов. Такая организация возможна потому, что в большизнстве выходных регистров микросхем памяти разрядность в несколько раз превышает число разрядов одной запоминающей ячейки.

9. В динамической оперативной памяти в последние годы при обращении используется режим **удвоенной скорости передачи данных** (DDR). Сущность метода заключается в том, что данные передаются не один раз за такт синхронизации, а дважды: по переднему и заднему фронтам тактового импульса.

5.4 Асинхронные и синхронные запоминающие устройства

1. ЗУ основной памяти подразделяются на асинхронные и синхронные

   (a) В **асинхронной** памяти момент начала очередной операции определяется только моментом завершения предыдущей операции.

   (b) В **синхронной** памяти процессы чтения и записи выполняются в момент поступления тактовых сигналов синхронизации, вырабатываемых специальным контроллером памяти. В отличие от асинхронной памяти, здесь между окончанием некоторой операции и началом следующей может образовываться некоторая временная задержка, связанная с ожиданием поступления очередного тактового импульса.

2. Поскольку контроллер памяти всегда работает в синхронном режиме и именно он задает начало очередной операции, в общем случае, временная

задержка в асинхронной памяти оказывается значительно больше, чем в синхронной.

3. Все совррменные вычислительные машины используют оперативную память синхронного типа.

4. Для динамичекой оперативной памяти с удвоенной скоростью передачи данных (DDR) различают реальную и эффективную частоты работ.

   (a) Реальная частота — количество тактов, которые успевает реализовать модуль памяти за одну секунду

      i. Отражает действительные возможности микросхемы памяти

   (b) Эффективная частота — удвоенное значение реальной частоты, связанное с удвоенной скоростью передачи данных в памяти DDR.

      i. В большей степени носит маркетинговый характер

5. Тайминги (latency) — понятие, характеризующее быстродействие. Значение указывается в виде трех (реже — четырех) целочисленных значений, разделенных дефисом.

   (a) Первое значение обозначает время, выраженное количеством тактов процессора, затрачиваемых между моментом запроса к памяти о стороны процессора и моментом, когда память сделает доступной для считывания информацию из первой ячейки.

   (b) Второе значение обозначает количество тактов процессора, затрачиваемых на обращение к последующим смежным ячейкам этой же страницы.

   (c) Третье значение обозначает количество тактов процессора, затрачиваемых на регенерацию информации, хранящейся в ячейках памяти.

   (d) Четвертое значение представляет собой суммарное количество тактов процессора, затрачиваемых на один цикл обращения (сумма первых трех чисел)

5.5 Постоянное запоминающее устройство (ROM)

1. В современных вычислительных машинах ROM в основе организации имеет также матричную структуру. В этой структуре в некоторых узлах матрицы расположены перемычки, связывающие линии столба и линию строки. Перемычки могут выполняться в виде проводников, диодов, конденсаторов, либо транзисторов. Запоминащая ячейка расположенная в узле

с перемычкой, хранит единицу, а ячейка в узле без перемычки — ноль. Считывание сводится к определению наличия перемычки и выполняется со скоростью, сопоставимой с динамической памятью. ПЗУ, допускающие возможность перепрограммирования, выполняют запись очень редко. Сам процесс физически отличается от считывания и протекает очень медленно.

Часть VI
Лекция 6 (4.04.16)

6 Системы ввода/вывода

1. Системы ввода/вывода призваны обеспечить обмен информацией между ядром вычислительной машины и разнообразными внешними устройствами.

2. Технические и программные средства системы ввода/вывода несут ответственность за физическое и логическое сопряжение ядра вычислительной машины с внешними устройствами.

3. Технически, системы ввода/вывода в рамках конкретной вычислительной машины реализуются комплексом **модулей** ввода/вывода. Модуль ввода/вывода реализует коммуникационные операции между ядром вычислительной машины и внешшними устройствами. Каждый модуль выполняет две функции:

   (a) Обеспечение интерфейса с центральным процессором и памятью (большой интерфейс)

   (b) Обеспечение интерфейса с одним или несколькими внешними (периферийными) устройствами (малый интерфейс)

6.1 Адресное пространство системы ввода/вывода

1. Операция ввода/вывода, аналогично операциям обращения к памяти, предполагает наличие некоторой системы адресации, позволяющей выбрать конкретные модули ввода/вывода, а также одно из подключенных к этому модулю внешних устройств.

2. Адрес конкретного модуля и конкретного внешнего устройства является составной частью некоторой команды, в то время, как расположение дан-

ных на внешнем устройстве определяется пересылаемой на этой устройство через модуль ввода/вывода служебной информацией.

3. Существует два подхода к организации адресного пространства системы ввода/вывода:

   (a) **Совмещенное** адресное пространство — совмещено с общей памятью вычислительной машины

      i. Для адресации модулей ввода/вывода отводится определенная область адресов в системе адресации конкретной вычислительной машины.

      ii. Операции с модулем ввода/вывода организуются с использованием регистров, входящих в его состав (управления, состояния и данных).

      iii. Физически, операция ввода/вывода сводится к записи информации в одни регистры модуля и считыванию её из других. Такой подход позволяет рассматривать регистры модулей ввода/вывода, как ячейки основной памяти. Это дает возможность работать с модулями ввода/вывода, используя обычные команды обращения к памяти. При этом часто в системе команд вычислительной машины сперциальные команды ввода/вывода вообще отсутствуют.

      iv. Учитывая, что операции ввода/вывода составляют очень незначительную часть вычислительных машинных операций, такая организация адресного пространства обладает преимуществом, особенно в больших и средних вычислительных машинах.

   (b) **Выделенное** адресное пространство

      i. Для обращения к модулям ввода/вывода используются специальные команды и отдельная система адресации. Это позволяет разделить информационные магистрали ядра вычислительной машины и системы ввода/вывода, давая им возможность работать параллельно и условно независимо. Кроме этого, выделенное адресное пространство может быть использовано в полном объеме.

      ii. Выделенное адресное пространство используется в малых вычислительных машинах и в персональных компьютерах.

6.2 Методы управления вводом/выводом

1. В вычислительных машинах находит применение три способа организации ввода/вывода:

(a) **Программно управляемый** ввод/вывод — все операции ввода/вывода происходят по инициативе и под управлением центрального процессора.

i. Центральный процессор, в рамках выполняемой программы, организует ввод/вывод информации, включая

A. проверку состояния внешнего устройства

B. выдачу команд на ввод/вывод информации

C. ожидание завершения операции ввода/вывода: вызывает неоправданно длительные простои процессора и снижение производительности вычислительной машины.

(b) Ввод/вывод **по прерываниям** — усовершенствованный управляемый ввод/вывод.

i. Улучшения связаны с тем, что центральный процессор, выдав команду на ввод или вывод информации, не ожидает завершения её выполнения, а приступает к обработке следующей команды программы. Это продолжается до тех пор, пока процессор не получит сигнал от узла прерываний, свидетельствующий о завершении выполнения команды ввода/вывода, либо о возникновении какого-либо сбоя.

(c) Ввод/вывод **с прямым доступом к памяти** (DMA) — основная память и модули ввода/вывода обмениваются информацией напрямую, минуя центральный процессор.

i. Управление прямым доступом к памяти выполняет одноименный контроллер (контроллер DMA).

7 Шины вычислительной маШины

1. С точки зрения архитектуры вычислительных машин, **шиной** принято называть информационную магистраль, включающую в себя 4 элемента:

   (a) Линии передачи сигналов адресов

   (b) Линии передачи сигналов данных

   (c) Линии передачи сигналов управления

   (d) Управляющее устройство (контроллер шины)

2. Физически, линии передачи сигналов перечисленных видов могут реализовываться как раздельно для каждого вида, так и в виде единой физической среды передачи сигналов.

3. Шина, связывающая только два каких-либо устройства, либо две системы, называется **портом**.

4. Операции, выполняемые на шине, принято называть **транзакциями**. Основными транзакциями на шине являются

   (a) Транзакция **чтения** (считывания)

   (b) Транзакция записи

5. Транзакция — совокупность операций рассматриваемых как единое действие.

   (a) Пересылка адреса

   (b) Пересылка данных

Часть VII
Лекция 7 (11.04.16)

1. При обмене информацией по шине между двумя устройствами, одно из них должно инициировать обмен и управлять им.

   (a) Устройство, инициирующее захват шины для передачи информации называют **ведущим**. Устройство, принимающее информацию, называется **ведомым**.

   (b) Правом захвата шины для управления обменом обладают не все подключенные к шине устройства. Поэтому, в некоторых случаях, в интересах таких устройств захват шины осуществляет какое-либо иное устройство, наделенное правом ведущего.

2. Принимать информацию из шины в любой момент времени может любое из подключенных к шине устройств. Передавать информацию в каждый момент времени может только одно из устройств, наделенных правом ведущего.

3. Для предотвращения возможных конфликтов при попытке захвата шины двумя или более потенциальными ведущими, действует система приоритетов.

   (a) **Статические** приоритеты, приваиваемые каждому потенциально ведущему устройству и остающиеся неизменными.

(b) **Динамические** приоритеты, изменяемые у потенциальных ведущих в ходе сеанса. Такой подход выравнивает шансы ведущих устройств на захват шины.

7.1 Группы шин персонального компьютера различного функционального назначения

1. В иерархии шин современного ПК можно выделить следующие группы шин разного функционального назначения:

   (a) Группа шин «**процессор-память**»

      i. Шина **переднего плана** — связывает центральный процессор с динамической оперативной памятью. Не имеет внешних точек подключения.

      ii. Шина **второго плана** — связывает центральный процессор со статической оперативной памятью. Самая скоросная шина. Не имеет внешних точек подключения.

      iii. ... обе шины этой группы стремятся выполнить максимально короткий путь. Это связано с тем, что при высоких частотах, характерных для современных процессоров, длительность такта оказывается соизмеримой с временем распространения сигналов по шине. При большой длине шины сигнал, отправленный процессором, может не достичь адресата за время такта, что приводит к «перекосу» сигналов и последующей рассинхронизации.

   (b) Группа шин «**ввода/вывода**» — все шины, связывающие внешние (периферийные) устройства с ядром вычислительной машины. Шины имеют точки внешнего подключения. Передача информации может быть организована как в параллельном (побайтном) режиме, так и последовательном. Шины этой группы имеют относительно низкое быстродействие. Ограничение на длину шины не устанавливается.

   (c) Шина **графического адаптера** — в отличие от других шин, связывающих ядро машины с модулями ввода/вывода, высоко требовательна к производительности, а значит и к скорости работы, что, в большей степени, присуще к шинам группы «процессор-память».

   (d) **Системная шина** (не путать с общей или объединительной шиной, которую в прошлом тоже называли системной) — служебная шина, по которой передаются сигналы состояния, сигналы управления и т.п.

      i. Информация, связанная непосредственно с решением пользовательских задач, по системной шине не передается.

7.2 Выделенные и мультиплексируемые шины

1. Магистрали (линии) передачи сигналов конкретного назначения (сигналы адресов, данных, управления) технически могут реализовываться двумя разными способами:

   (a) **Выделенные** линии — линии передачи конкретного вида сигналов не изменяемого в процессе работы.

   (b) **Мультиплексируемые** линии — линии, используемые для передачи сигналов разного вида.

      i. Использование таких линий позволяет значительно сократить число проводников шины и упростить технологию её реализации.

      ii. Недостаток таких линий заключается в снижении общей пропускной способности для каждого вида сигналов.

2. Мультиплексируемые линии в современныхх персональных компьютерах используются только в шинах ввода/вывода. В группах шин «процессор-память» и шинах видеосистемы применяются исключительно выделенные линии, обеспечивающие максимальную пропускную способность.

7.3 Протоколы шины

1. **Протокол** шины ВМ — метод информирования о достоверности передачи информации по шине.

2. Существует два класса протоколов:

   (a) **Синхронные** протоколы — все сигналы привязаны к импульсам единого тактового генератора.

      i. Требуют меньше сигнальных линий, проще для реализации, доступнее для тестирования.

      ii. Менее гибкие, поскольку привязаны к конкретной тактовой частоте, во многом определяемой уровнем технологии.

      iii. По синхронному протоколу работают шины группы «процессор-память».

   (b) **Асинхронные** протоколы — для каждой группы линий шины формируетс свой сигнал подтверждения достоверности.

      i. По асинхронному протоколу работают шины группы ввода/вывода.

8 Параллелизм как основа высокопроизводительных вычислений

1. До недавнего времени в основе архитектуры большинства ВМ лежало представление алгоритма решения задач в виде программы последовательных вычислений. На текущий момент идеи ускорения последовательных вычислений, основанные на совершенствовании технологий и архитектурных изменениях при современной элементной базе, оказались практически исчерпаны.

Часть VIII
Лекция 8 (18.04.16)

8.1 Конвейеризация вычислений

1. В отсутсвие конвейера некоторый функциональный блок выполняет обработку команды, получаемой из входного регистра. Результаты обработки команды заносятся в выходной регистр. Время, затраченное функциональным блоком равняется t_{max} (максимальному времени).

2. Переход к конвейеру возможен и целесообразен, если обработку команды можно разделить на несколько частей, обрабатываемых независимо друг от друга, и при этом время, затрачиваемое на обработку каждой части, примерно одинаковое и составляет t_{max}/n, где n — число частей команды. В этом случае обрработку каждой части можно поручить отдельному функциональному блоку, работающему независимо от других.

3. При переходе к конвейеру, первый функциональный блок, получив команду из входного регистра, обрабатывает свою часть, а оставшуюся часть команды через промежуточный регистр передает следующему функциональному блоку и, не дожидаясь окончания обработки команды остальными функциональными блоками, приступает к обработке своей части следующей команды.

8.2 Уровни параллелизма

1. Методы и средства реализации паралеллизма во многом зависят от того, на каком уровне он обеспечивается. Обычно, выделяют следующие уровни:

(a) **Уровень заданий** — несколько отдельных заданий, сформулированных в результате декомпозиции некоторой сложной (комплексной) задачи, выполняются на разных вычислителях (процессорах), которые время от времени обмениваются результатами вычислений в рамках общей задачи.

i. Этот уровень свойственен крупным вычислительным системам, созданным на основе нескольких вычислительных машин или даже нескольких менее масштабных вычислительных систем.

(b) **Уровень программ**

i. Об этом уровне параллелизма можно говорить в двух случаях:

A. Если в алгоритме решения задачи есть независимые участки, которые допустимо выполнять параллельно.

B. Если в пределах отдельного программного цикла в нем присутсвуют итерационные вычисления, не зависящие друг от друга.

ii. Реализуется как на многопроцессорных вычислительных системах, так и на отдельных вычислительных машинах с многоядерной архитектурой процессоров.

(c) **Уровень команд** — имеет место, когда обработка нескольких команд или выполнение разных этапов одной и той же команды перекрываются во времени.

i. Параллелизм уровня команд реализуется на отдельной вычислительной машине.

(d) **Битовый (арифметический) параллелизм** — имеет место при бит-параллельном выполнении операций, т.е. когда отдельные биты одного машинного слова обрабатываются параллельно.

i. Реализуется архитектурными блоками отдельных процессоров.

2. К понятию уровня параллелизма тесно примыкает понятие **гранулярности** — меры отношения объема вычислений, выполненных в рамках решения отдельной задачи отдельным вычислителем к объему информации, переданной другим вычислителям в рамках решения общей задачи.

(a) Степень гранулярности варьируется от мелкозернистой до крупнозернистой.

i. При **крупнозернистом** параллелизме, каждое выполняемое вычисление мало зависит от остальных и между вычислителями

происходит относительно редкий обмен результатами. На программном уровне, крупнозернистый параллелизм обеспечивается операционной системой вычислительного комплекса (вычислительной системы).

ii. **Среднезернистый** параллелизм имеет дело с единицами обработки по программам, включающим сотни команд. Организуется либо программистом (составителем программы), либо компилятором.

iii. При **мелкозернистом** параллелизме, параллельные вычисления включают в себя обработку до нескольких десятков команд. Характерной особенностью мелкозернистого параллелизма является примерное равенство объемов обрабатываемой и обмениваемой информации.

8.3 Классификация вычислительных систем

1. **Классификация Флинна**: В основу классификации положено понятие **потока** — последовательности команд или данных, обрабатываемых вычислителем. В зависимости от количества потоков команд и потоков данных по Филинну выделяются четыре класса систем:

   (a) SI SD (Single Instruction Single Data)

      i. Типичный представитель: классическая фон Неймановская вычислительная машина и всё, что на ней основано.

   (b) MI SD (Multiple Instruction Single Data)

      i. В ахритектуре такой вычислительной системы присутствует множество процессоров, обрабатывающих один и тот же поток данных.

   (c) SI MD (Single Instruction Multiple Data)

      i. В архитектуре такой вычислительной системы присутсвтует матрица процессоров, т.е. над множеством данных выполняется одна и та же операция. При этом сами данные обычно рассматриваются как элементы (координаты) своего рода многомерного вектора. Все процессоры такой матрицы получают из памяти одну и ту же команду и выполняют её по отношению к смоим локальным наборам данных.

   (d) MI MD (Multiple Instruction Multiple Data) — весьма разнообразен в плане технической реализации и включает в себя всевозможные мно-

гопроцессорные и многомашинные вычислительные системы. Существует отдельная классификация входящив в него вычислительных систем.

i. Каждый процессорный элемент выполняет свою программу и делает это относительно независимо от других процессорных элементов. При этом, с точки зрения технической реализации процессорных элементов, вычислительные системы делятся на две группы:

   A. Мультипроцессоры: с общей памятью (SMP (симметричные мультипроцессорные системы), PVP (параллельные векторные процессы)) и с распределенной памятью (ncc-NUMA (с некогерентным кэшем), cc-NUMA (с когерентным кэшем), COMA). Процессоры довольно тесно взаимодействуют в ходе выполнения параллельных вычислений, поэтому этот класс вычислительных систем часто называют **сильносвязанным**.

   B. Мультикомпьютеры: системы с массовым параллелизмом (MPP) и кластерные системы. Отновятся к **слабосвязанным** системам, так как обмен данными между вычислителями в этих системах происходит лишь эпизодически.

8.4 Симметричные мультипроцессорные системы (SMP)

1. **Симметричной и мультипроцессорной** можно назвать вычислительную систему, обладающую следующими характеристиками:

   (a) Система содержит два или более процессоров сопоставимой производительности.

   (b) Процессоры совместно используют основную память и работают в едином виртуальном и физическом адресном пространстве.

   (c) Все процессоры связаны между собой посредством коммуникационной системы, обеспечивающей для них одинаковое время доступа к основной памяти. Эта коммуникационная система чаще всего реализуется в виде общей шины.

   (d) Все процессоры разделяют доступ к устройствам ввода-вывода либо через одни и те же каналы, либо через раздельные каналы, но обеспечивающие доступ к любому из внешних устройств.

   (e) Все процессоры способны выполнять одинаковые функции.

   (f) Любой из процессоров может обслуживать аппаратные прерывания.

(g) Вычислительная система управляется интегрированной операционной системой, организующей и координирующей взаимодействие между процессорами на всех реализуемых уровнях параллельных вычислений.

2. В этих системах весьма часто реализуется мелкозернистый параллелизм, что подчеркивает сильную связность таких систем.

3. Понятие симметричности здесь относится как к самой архитектуре, так и поведению операционной системы. Однако в определенных ситуациях в поведении процессоров возникает некоторый перекос. Это касается только этапа загрузки операционной системы, в течение которого один из процессоров получает статус ведущего. Но эта характеристика касается только времени загрузки операционной системы, после которой все процессоры становятся абсолютно равноправными.

4. Операционная система планирует вычислительные процессы сразу по всем процессорам. При этом для пользователя многопроцессорный характер системы является скрытым.

5. В сравнении с однопроцессорными системами SMP имеют следующие преимущества:

   (a) Выигрыш в производительности при решении задач, поддающихся разбиению на отдельные составные части

   (b) Готовность: отказ одного или нескольких (но не всех) процессоров системы не ведет к отказу всей вычислительной системы. Функции отказавшего процессора перераспределяются операционной системой между остальными, сохранившими работоспособность.

   (c) Расширяемость: в большинстве случаев, количество используемых процессоров может быть сравнительно легко увеличено как технически, так и программно.

   (d) Масштабируемость: позволяет варьировать число задействованных в вычислении процессоров для максимальной эффективности.

6. Виды SMP систем по использованию кэш-памяти:

   (a) С раздельным кэшем: Каждый процессор имеет собственный локальный кэш. Согласованность содержимого кэшей (когерентность) всех процессоров здесь обеспечивается на аппаратном уровне.

   (b) С общим кэшем: Все процессоры используют один и тот же кэш.

7. В обоих рассмотренных вариантах реализации SMP, Для всех процессоров обеспечивается равноправный доступ к основной памяти и модулям ввода-вывода. Обычно, процессоры взаимодействую между собой через основную память. В некоторых редких случаях реализуется прямой обмен информацией между процессорами, минуя основную память.

8. Существует несколько вариантов реализации коммуникационной системы, обеспечивающей такое взаимодействие:

   (a) Архитектура SMP с общей шиной

      i. Достоинства и недостатки систем на базе общей шины те же, что присущи и однопроцессорным вычислительным машинам с единственной объединительной шиной и связаны с разделением времени использования общей шины в интересах каждого из процессоров.

      ii. С одной стороны, общая шина позволяет легко подключать дополнительные процессоры в систему. С другой стороны, при большом числе процессоров, как правило, пропускной способности общей шины оказывается недостаточно.

   (b) Архитектура с коммуттатором типа «кроссбар»

      i. Эта архитектура ориентирована на модульное построение основной памяти и системы ввода-вывода и нацелена на разрешение проблемы ограничения пропускной способности коммуникационной системы, присущей, например, общей шине.

      ii. Коммутатор обеспечивает множественность путей между процессорами и модулями памяти, причем архитектура кроссбара может быть как двумерной, так и объемной.

   (c) Архитеектура с многопортовой памятью

      i. Применяется в запоминающих устройствах основной памяти, использующих множественные каналы ввода и вывода информации. Каждый выделенный канал обеспечивает взаимодействие конкретного процессора к общему банку памяти.

      ii. При таком подходе усложняется логика взаимодействия процессора и памяти, но существенно повышается производительность системы в целом.

      iii. При необходимости, конкрретные модули памяти могут «закрепляться» за конкрретными процессорами в качестве локальной памяти.

(d) Архитектура с централизованным устройством управления

i. Такое централизованное УУ замыкает на себя и распределяет все информационные потоки между всеми подсистемами.

ii. Архитектура широко использовалась в первых поколениях вычислительных машин. В настоящее время практически не применяется из за чрезвычайной сложности в организации и сравнительно нихкой производительности.