

Лекция №2  
по курсу  
«Проектирование и архитектура вычислительных  
систем»

Москва, 2020

# Иерархия памяти

Регистры, расположены прямо в процессоре.

- Кэш память для хранения часто используемых фрагментов памяти.
- Основная оперативная память.
- Дисковая память.
- Память на магнитных лентах (стриммеры).
- В старых ЭВМ активно использовались перфоленты и перфокарты.

# Кэш-память

Кэш-память содержит копии наиболее часто используемых участков основной памяти и строится на быстрой статической памяти.

Кэшируются не отдельные слова, а линейки из последовательности байтов основной памяти (скажем, в архитектуре x86 длина линейки составляет 64 байта).

Кэш-память является своеобразной hash-таблицей, где hash образуется из адреса основной памяти отбрасыванием нескольких младших битов. • Линейка, на которую адрес попал, переносится в начало списка, а линейки, на которые долго не было попадания, выбрасываются из спис

# Принципы работы основной памяти

## Принципы работы основной памяти

1. Статическая память на триггерах, быстрая, но дорогая.
2. Динамическая память на конденсаторах, медленнее, чем статическая, но и значительно дешевле. Поскольку идеальных диэлектриков не бывает, заряд с конденсатора постепенно стекает, поэтому нужно регулярно делать чтение и перезапись (refresh), что еще больше ее замедляет

# Характеристики основной памяти

Характеристики основной памяти

- Latency – задержка. Если 20 лет назад память была медленнее процессора в 3 – 4 раза, то сегодня – в 15-20 раз
  - Cycle time – длительность такта.
- (Модуль 1, Урок «Цифровая логика»)

# Характеристики основной памяти

Bandwidth – пропускная способность, т.е. ширина шины доступа процессор-память.

Пример: первые IBM PC XT базировались на микропроцессоре i8086, процессор и шина 16-ти разрядные. Когда перешли на i8088 с 16-ти разрядным процессором, но с 8-ми разрядной шиной, в несколько раз уменьшили цену персонального компьютера.

# Виртуальная память

Virtualis (лат.) означает «возможный», объект, который не существует, но может возникнуть.

- Страница – фрагмент оперативной памяти фиксированного размера (скажем, 2 Кб).
- Математический адрес – пара  $(N, D)$ , где  $N$  – номер страницы,  $D$  – смещение нужного адреса от начала страницы.

# Таблица страниц

Таблица страниц – таблица, входом которой является номер страницы, а содержимым – адрес страницы в оперативной памяти, если она находится в этой памяти, или информация о ее месте на диске, если этой страницы нет в памяти.



# Главный принцип виртуальной памяти

Главный принцип виртуальной памяти – использование прямых физических адресов запрещено, только относительные математические. Математический адрес – пара: номер страницы и смещение внутри страницы. Физический адрес – настоящий адрес внутри оперативной памяти.

Перед каждым использованием математического адреса для обращения к памяти его нужно преобразовать в физический адрес. По номеру страницы обращаемся к содержимому таблицы страниц.

# Реализация виртуальной памяти

Перед каждым использованием математического адреса для обращения к памяти его нужно преобразовать в физический адрес.

По номеру страницы обращаемся к содержимому таблицы страниц.

Если страница присутствует в памяти, то просто добавляем к ее адресу начала смещение, если же нет – то вызывается процедура операционной системы по подкачке нужной страницы.

Если свободной памяти для этой страницы нет, то приходится откачивать несколько страниц на диск.

# Способы ускорения виртуальной памяти

Если нужной страницы не оказалось в оперативной памяти, говорить о быстродействии не приходится. Оказывается, при работе любой программы быстро формируется рабочее множество страниц, к которым осуществляется 90-95% обращений.

# Способы ускорения виртуальной памяти

Но даже, если страница находится в памяти, преобразование математического адреса в физический адрес сильно замедляет работу.

- TLB (Translation lookaside buffer) – специализированная кэш-память, где находятся адреса часто используемых страниц

# Обработка ошибок в памяти

Ошибки можно компенсировать только избыточной информацией.

Самый простой способ – к каждому байту информации добавить бит четности, если сумма по модулю 2 всех битов, включая дополнительный, равна 0, все в порядке, этот способ помогает только фиксировать ошибку, но никак не помогает в ее исправлении.

# Количество информации

## Количество информации

### Алфавитный

количество информации, которое можно передать с помощью одного двоичного знака (0 или 1)

# 1 бит

### Содержательный

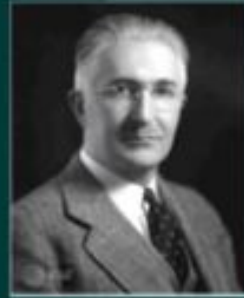
количество информации, уменьшающее неопределенность в два раза

красный 0  
1 бит  
зеленый 1



1) красный  
2) ~~зеленый~~  
1 б

# Количество информации



Р. Хартли

$$i = \log_2 n$$

$n$  - количество равновероятных событий,  
 $i$  - количество информации

Дано:

$$n = 3$$

Найти:

$$L - ?$$

Решение:

$$L = \log_2 3 = 1,58 \text{ бита}$$

$$L \approx 2 \text{ бита}$$

# Количество информации



Какое количество информации несет сообщение, что загорелся желтый сигнал?

$$i_x = \log_2 \left( \frac{1}{p_x} \right)$$

$i_x$  - количество информации при наступлении события  $x$

$p_x$  - вероятность наступления события  $x$

$p_x = \frac{k_x}{n}$   $k_x$  - количество событий  $x$

$n$  - общее количество событий

Дано:

$$k_k = 108$$

$$k_z = 16$$

$$k_m = 4$$

Найти:

$$i_{mk} = ?$$

Решение:

$$p_{mk} = \frac{4}{108 + 16 + 4} = \frac{4}{128} = \frac{1}{32}$$

$$i_{mk} = \log_2 32 = 5$$



# Количество информации



Какое среднее количество информации несет сообщение о том, что загорелся любой из трех сигналов?

$$H = \sum_{x=1}^n p_x \cdot \log_2 \frac{1}{p_x}$$

128 с

H - среднее количество информации, полученной при выборе любого из n возможных вариантов

Дано:  
 $K_K = 108$   
K.

Решение:

# ОСНОВЫ ВВОДА-ВЫВОДА

- Канал – специализированный процессор, отвечающий за обмен данными с определенным устройством.
- Блок данных – порция данных фиксированного размера.
- Запись – единица обработки данных, обычно блок состоит из многих записей

Как уже говорилось, центральный процессор запускает операцию обмена, выполняя команду SIO с параметром – номером канала, после чего центральный процессор и канал работают параллельно

# Буферизация

Пусть есть большой файл, который нужно последовательно обработать.

Чтобы не терять время на чтение блока, можно отвести два или больше буферов, тогда задержка будет только на чтение первого блока файла, затем прочитанный буфер обрабатывается, а следующие параллельно с этим читаются.

Нужно управлять размером и количеством буферов, чтобы время обработки одного буфера было примерно равно времени его чтения, тогда центральный процессор будет постоянно загружен.

# Буферизация

Ввод/вывод по прерыванию – канал по завершении операции вызывает прерывание центрального процессора, во время обработки которого запускается следующая операция обмена, этот способ более эффективен.

# Абстракция

Критически важный принцип управления сложностью системы – абстракция, подразумевающая исключение из рассмотрения тех элементов, которые в данном конкретном случае несущественны

# Принципы разработки электронных систем

- ▶ *Иерархичность* – принцип иерархичности предполагает разделение системы на отдельные модули, а затем последующее разделение каждого такого модуля на фрагменты до уровня, позволяющего легко понять поведение каждого конкретного фрагмента.
- ▶ *Модульность* – принцип модульности требует, чтобы каждый модуль в системе имел четко определенную функциональность и набор интерфейсов и мог быть легко и без непредвиденных побочных эффектов соединен с другими модулями системы.
- ▶ *Регулярность* – принцип регулярности требует соблюдения единообразия при проектировании отдельных модулей системы. Стандартные модули общего назначения, например, такие как блоки питания, могут использоваться многократно, во много раз снижая количество модулей, необходимых для разработки новой системы.