
Тема 1.10

Работа кэш-памяти

1. Концепции кэш-памяти



Кэш-память является чисто аппаратным средством, *“прозрачным”* для выполняемых программ и представляет собой своеобразный буфер между основной памятью и центральным процессором (*далее ЦП*).

Как правило, соотношение между емкостью кэш-памяти и оперативной памятью (*далее ОП*) составляет $1/100 - 1/1000$ – в зависимости от типа ЭВМ.

Как правило, передача данных между основной памятью и процессором производится через кэш-память, хотя в принципе возможна **прямая передача** (*на рисунке показана пунктиром*).

Кэш-память строится на основе элементов статической памяти (**SRAM – Static Random Access Memory**), а ОП – на элементах динамической памяти (**DRAM – Dynamic RAM**).

Элементы SRAM представляют собой триггеры, а DRAM – по принципу действия подобны конденсатору. В связи со стеканием заряда во времени память типа DRAM требует периодической перезаписи (*refresh*).

Кэш-память и основная память разделяются на блоки одинакового объема, размер которых обычно составляет 8.128 байт. Обычно блок ОП и кэш-памяти называют *строкой (line)*. Обмен между ОП и кэш-памятью носит блочный характер. В кэш-памяти содержатся копии тех блоков ОП, к которым в последнее время выполнялись обращения со стороны ЦП.

При любом обращении процессора к основной памяти определяется наличие блока, к которому производится обращение, в кэш-памяти. При нахождении блока в кэш-памяти (такая ситуация называется *кэш-попаданием – cache-hit*), осуществляется быстрое обращение (чтение или запись) со стороны ЦП в кэш-память;

Обмен между ЦП и кэш-памятью осуществляется не на уровне блоков, а на уровне слов. Под словом в данном случае понимается объём данных, участвующих в одной пересылке между ЦП и кэш-памятью или ЦП и ОП. Разрядность слова определяется разрядностью шины данных между ЦП и кэш-памятью или ЦП и основной памятью;

При отсутствии в кэш-памяти блока, к которому осуществляется обращение (такая ситуация называется *кэш-промахом – cache-miss*), сначала инициируется пересылка блока, содержащего затребованное слово, из ОП в кэш-память, а затем осуществляется обращение к этому слову из кэш-памяти. Как правило, подобный подход имеет место при обращении по чтению. При обращении по записи в случае отсутствия блока в кэш-памяти, запись может производиться и непосредственно в оперативную память без предварительной пересылки блока в кэш-память;

Эффективность использования кэш-памяти определяется так называемым *принципом локальности ссылок* (доступа, обращений). Причем выделяются два вида локальности: *пространственная и временная*. Кроме того, принцип локальности рассматривается как в отношении команд, так и в отношении данных.

- **Пространственная локальность в отношении команд** характеризуется тем, что вероятность выборки команды по следующему адресу, по сравнению с адресом исполняемой команды, намного больше, чем вероятность выборки команды по любому другому адресу.

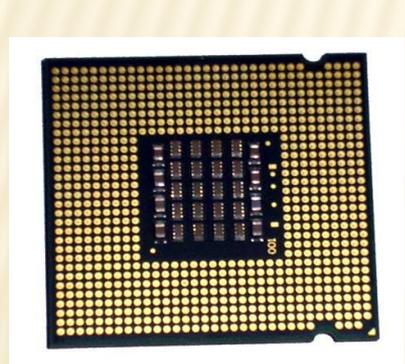
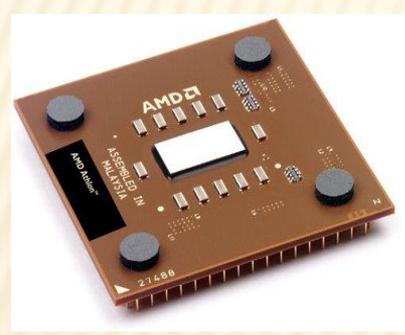
- Этот принцип проявляется на линейных участках программы. По статистическим данным, средняя длина линейных участков большинства программ научно-технического профиля составляет 5-7 машинных команд.

- **Пространственная локальность в отношении данных** выражается в том, что вероятность обращения к слову данных по следующему адресу намного больше вероятности обращения к данным по любому другому адресу. Этот принцип проявляется, например, при обработке массивов данных.

- **Временной аспект принципа локальности обращений в отношении команд** предполагает большую вероятность обращения к команде по одному и тому же адресу в течение небольшого интервала времени. Этот аспект проявляется при выполнении программных циклов.

- **В отношении данных временной аспект принципа локальности обращений** означает большое значение вероятности обращений к одному и тому же слову данных в течение небольшого интервала времени. Этот аспект проявляется при многократной обработке массивов данных.

- В соответствии с принципом локальности ссылок, к слову, однажды прочитанному из ОП в кэш-память, будет выполняться несколько повторных обращений (временной аспект) и, кроме того, так как вместе с одним словом из ОП в кэш-память передается целый блок, состоящий из нескольких последовательных слов, то с большой вероятностью, ряд последующих обращений будет локализован в пределах этого блока (пространственный аспект).



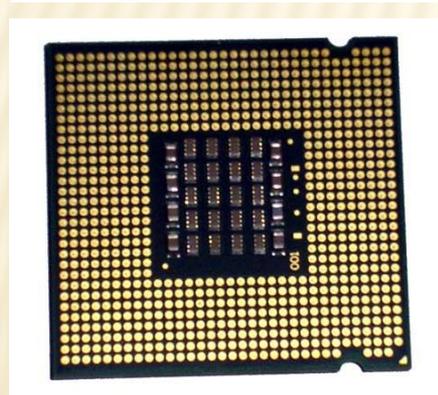
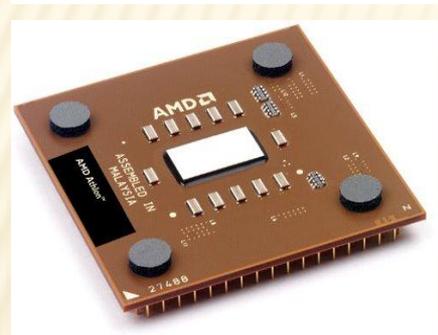


- **Численной оценкой** эффективности принятого принципа построения кэш-памяти является процент удачных обращений (*процент кэш-попаданий*), определяемый как отношение числа обращений к памяти, реализуемых через кэш, к общему числу обращений.

- Как правило, в современных компьютерах процент кэш-попаданий составляет **95.98 %**.

- **При построении кэш-памяти** необходимо решить следующие задачи, определяющие её организацию:

- 1) *выбор принципа отображения блоков основной памяти на блоки кэш-памяти (стратегия отображения \ распределения);*
- 2) *выбор принципа удаления блоков из кэш-памяти (стратегия замещения);*
- 3) *выбор принципа поддержания актуальности копий блоков кэш-памяти в блоках основной памяти (стратегия обновления ОП).*



КОНТРОЛЬНЫЕ ВОПРОСЫ:

1. Что такое когерентность кэша?
2. Что такое триггер?
3. Что такое конденсатор?
4. Заполните таблицу

Вид памяти	Элемент памяти	Преимущества	Недостатки	Сфера использования
Статическая (SRAM)				
Динамическая (DRAM)				