**7. Организация памяти современного процессора (L1, L2, L3 Cache, TLB, Virtual memory)**

Процессор — быстрый, память — медленная. Чем быстрее память, тем она дороже.

Кэш — память — быстрая, но маленького объема. В нее надо помещать наиболее часто используемые слова.

Кэш базируется на принципе локальности:

* Временная — данные, которые были использованы скорее всего будут использованы еще раз.
* Пространственная — соседние с использованным элементом данные скорее всего будут использоваться

Если требуемого слова нет в кэше, оно подгружается из основной памяти вместе со соседними, вытесняя при этом часть загруженных в кэш слов. Один из принципов вытеснения — Least recently used — вытсенение самых старых данных из кэша.

Префетчинг — подвыборка данных, программа начинает их читать раньше, чем из

используют.

Хорошая статья:

<http://iproc.ru/parallel-programming/lection-7/2/>

N — ассоциативный кэш:

Память состоит из кусочков - «банков». Кэш состоит из отдельных ячеек — слов. N — количество ячеек памяти, сопоставляющихся одному банку. При увеличении N увеличивается число проводов => увеличивается их длина. Нужен компромисс.

Несколько уровней кэша, работающих с разной скоростью

* L1 — 3
* L2 — 9 - 15
* Память — 30 - 180

Виртуальный ресурс(виртуальная память?) — абстракция, создающая иллюзию владения одним процессом всего адресного пространства

Виртуальная память — метод, отображающий адреса памяти, используемые программой, названные виртуальными адресами в физические адреса в памяти компьютера. Для программы адресное пространство выглядит как доступное и непрерывное адресное пространство, либо как набор непрерывных сегментов

Страничная организация виртуальной памяти: Оперативная память делится на страницы: области памяти фиксированной длины (например, 4096 байт), которые являются минимальной единицей выделяемой памяти (то есть даже запрос на 1 байт от приложения приведёт к выделению ему страницы памяти). Исполняемый процессором пользовательский поток обращается к памяти с помощью адреса виртуальной памяти, который делится на номер страницы и смещение внутри страницы. Процессор преобразует номер виртуальной страницы в адрес соответствующей ей физической страницы при помощи буфера ассоциативной трансляции (TLB). Если ему не удалось это сделать, то требуется дозаполнение буфера путём обращения к таблице страниц (так называемый Page Walk), что может сделать либо сам процессор, либо операционная система (в зависимости от архитектуры). Если страница была выгружена из оперативной памяти, то операционная система подкачивает страницу с жёсткого диска в ходе обработки события Page fault. При запросе на выделение памяти операционная система может «сбросить» на жёсткий диск страницы, к которым давно не было обращений. Критические данные (например, код запущенных и работающих программ, код и память ядра системы) обычно находятся в оперативной памяти (исключения существуют, однако они не касаются тех частей, которые отвечают за обработку аппаратных прерываний, работу с таблицей страниц и использование файла подкачки).

TLB(translation lookaside buffer) - это специализированный кэш центрального процессора, используемый для ускорения трансляции адреса виртуальной памяти в адрес физической памяти. TLB используется всеми современными процессорами с поддержкой страничной организации памяти. TLB содержит фиксированный набор записей (от 8 до 4096) и является ассоциативной памятью(Особый вид машинной памяти, используемый в приложениях очень быстрого поиска). Каждая запись содержит соответствие адреса страницы виртуальной памяти адресу физической памяти. Если адрес отсутствует в TLB, процессор обходит таблицы страниц и сохраняет полученный адрес в TLB, что занимает в 10—60 раз больше времени, чем получение адреса из записи, уже закэшированной TLB. Вероятность промаха TLB невысока и составляет в среднем от 0,01 % до 1 %.