

Лекция 6

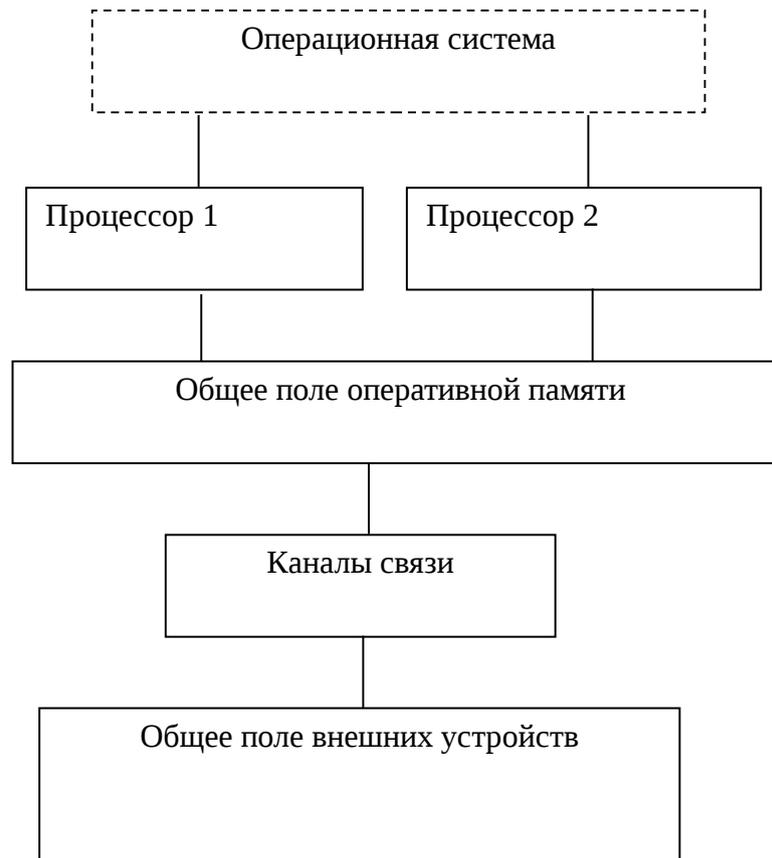
Многопроцессорные
вычислительные системы

Многопроцессорная ВС

К многопроцессорным ВС относятся такие, в которых несколько процессоров информационно взаимодействуют между собой на уровне регистров процессорной памяти, либо на уровне оперативной памяти.

В одних мультипроцессорных системах только определённые процессоры имеют доступ к устройствам ввода-вывода и, следовательно, имеют специальную функцию ввода-вывода. В других мультипроцессорных системах каждый процессор имеет доступ к любому устройству ввода-вывода. *Если все процессоры имеют равный доступ ко всем модулям памяти и всем устройствам ввода-вывода и каждый процессор взаимозаменяем с другими процессорами, то такая система называется SMP (Symmetric Multiprocessor — симметричный мультипроцессор).*

Схема взаимодействия процессоров



Быстродействие

Быстродействие и надежность многопроцессорных ВС по сравнению с многомашинными, взаимодействующими на уровне каналов связи, существенно выше. Это происходит, во-первых, ввиду ускоренного обмена информацией между процессорами и более быстрого реагирования на ситуации, возникающие в системе, во-вторых, вследствие большей степени резервирования устройств системы (система сохраняет работоспособность, пока работоспособны хотя бы по одному модулю каждого типа устройств).

В большинстве случаев многопроцессорные ВС реализованы в суперкомпьютерах.

Классификация по памяти

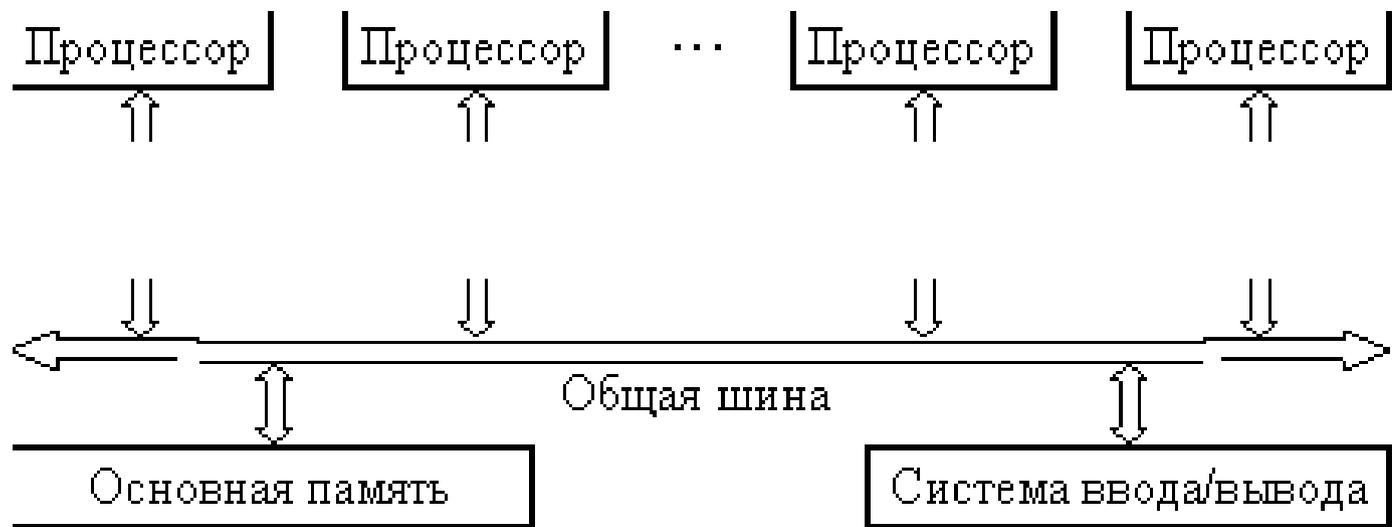
- системы с разделяемой памятью, у которой есть одна большая память и все процессоры имеют одинаковый доступ к данным и командам, хранящимся в этой памяти;
- системы с распределённой памятью, у которых каждый процессор имеет свою локальную память и у других процессоров доступа к этой памяти нет.

Различие

Различие между разделяемой и распределяемой памятью заключается в способе интерпретации адреса. Например, один из процессоров выполняет команду *load r1,i* (загрузить в регистр *r1* данные из ячейки памяти *i*). Если команда выполняется на компьютере с разделяемой памятью, номер ячейки *i* имеет одинаковый смысл для всех процессоров, т.е. *i* является глобальным адресом. В системе с распределенной памятью ячейка *i* разная для различных процессоров и в регистры *r1* по этой команде будут загружены разные значения.

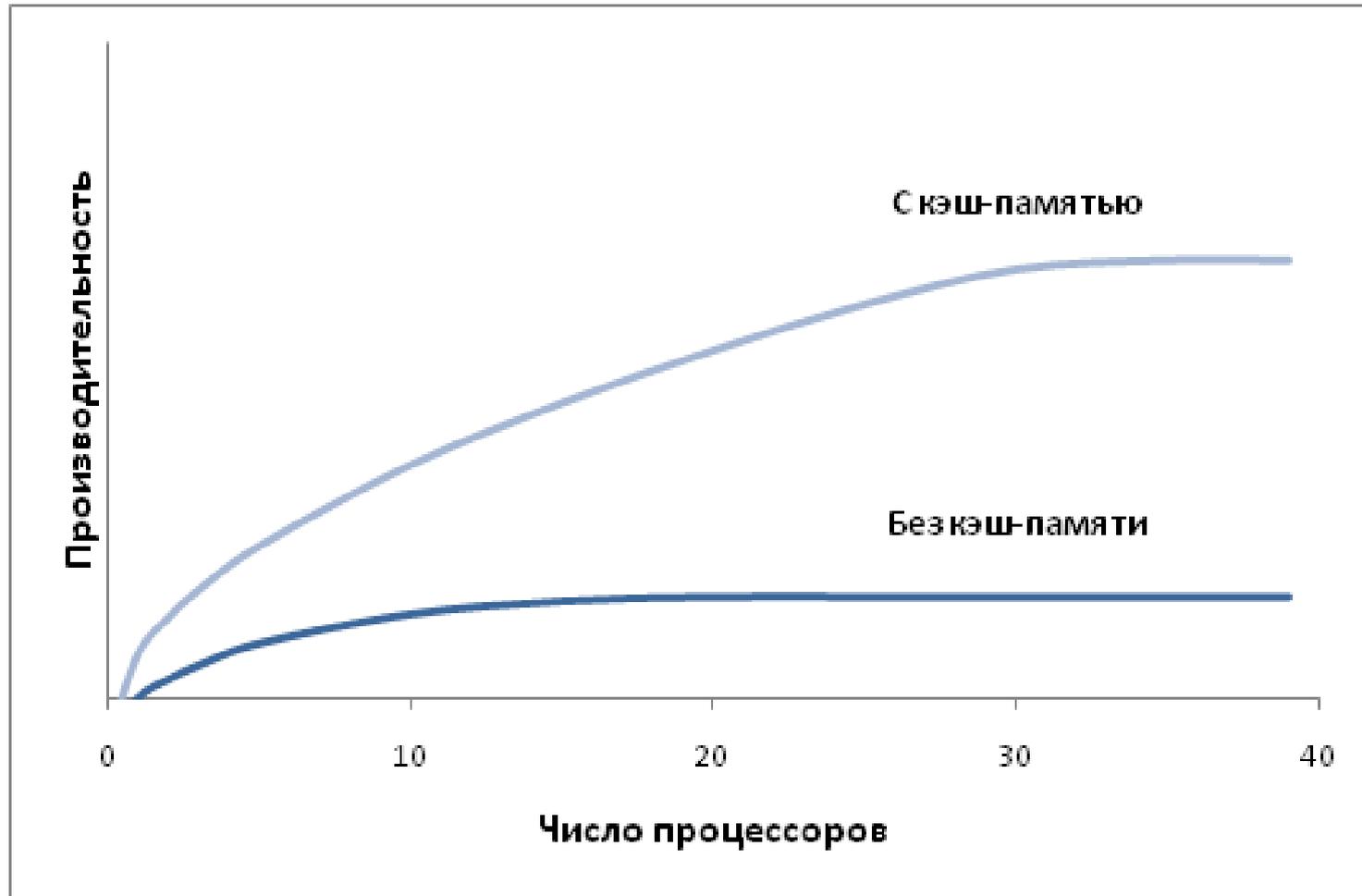
Разделяемая память

Простейшим способом создания разделяемой памяти является соединение нескольких процессоров общей шиной между собой и с общей оперативной памятью. UMA (Uniform Memory Access), до 10 процессоров.



кэш-когерентность

Разделяемая память



Механизмы когерентности кешей

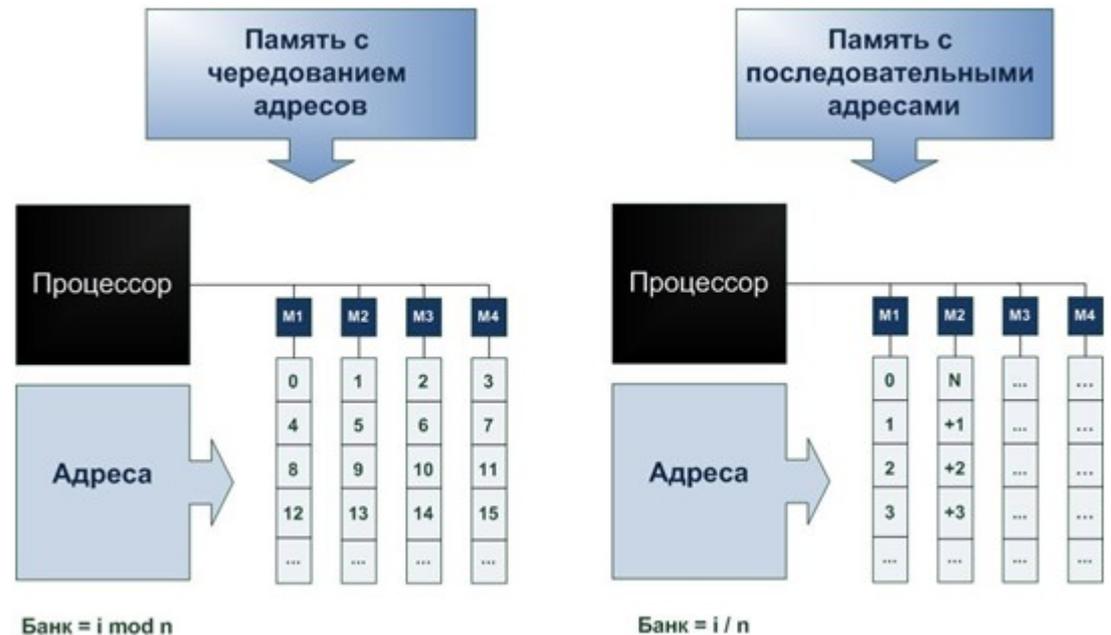
- Когерентность с использованием справочника (directory).
- Когерентность с использованием snooping.
- Snarfing.

Протоколы и модели поддержки когерентности

- Протокол MSI
- Протокол MESI (intel pentium, Core)
- Протокол MOSI
- Протокол MOESI (amd opteron)
- MOWESI
- Протокол MERSI
- Протокол MESIF
- Протокол Write-once
- en:Synapse
- en:Berkeley
- en:Illinois
- en:Firefly (DEC)
- en:Dragon (Xerox)

Чередуемая память

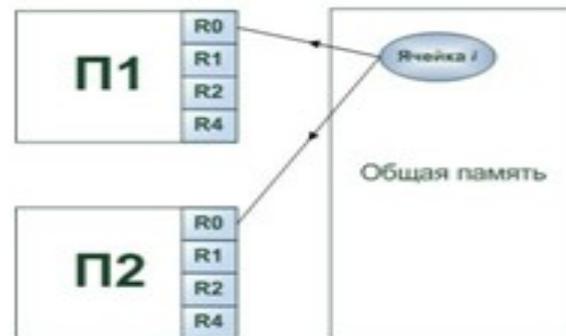
Чередуемая память разделяется на банки памяти. Используется для организации арифметического и оперативного конвейеров. Условно, это – ячейка памяти с номером i находится в банке памяти с номером $i \bmod n$, где n – количество банков памяти, а \bmod – операция вычисления остатков от деления. Таким образом, если имеется 8 банков памяти, то первому банку памяти будут принадлежать ячейки памяти с номерами 0, 8, 16, ..., второму – 1, 9, 17, ..., и так далее.



Распределённая память

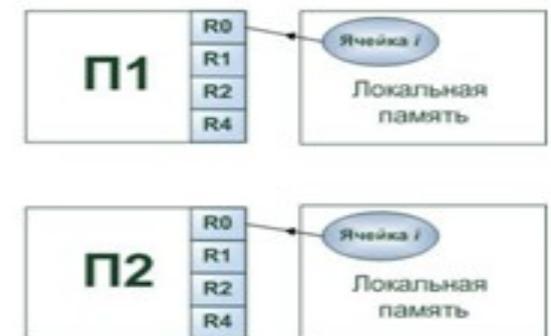
В ВС с распределенной памятью оперативная память имеется у каждого процессора. У процессора есть доступ только к своей памяти, поэтому нет необходимости в шине, в переключателях, нет конфликтов по доступу к памяти, нет ограничений на число процессоров, отсутствуют проблемы с кэш-когерентностью.

Выполнение Load R0, i



ВС с общей памятью

(shared memory,
мультипроцессоры, сильно
связанные, closely coupled systems)



ВС с распределенной памятью

(distributed memory, слабо
связанные, loosely coupled
systems, мультикомпьютеры)

Уровни параллелизма

- параллелизм заданий: каждый процессор загружается своей собственной, независимой от других, задачей;
- параллелизм на уровне программы: вычислительная программа разбивается на части, которые могут выполняться на отдельных процессорах;
- параллелизм команд: организовывается на нижних уровнях архитектуры ВС;
- параллелизм на уровне машинных слов и арифметических операций: в этом случае, например, сложение двух операндов выполняется параллельным сложением всех их двоичных разрядов.

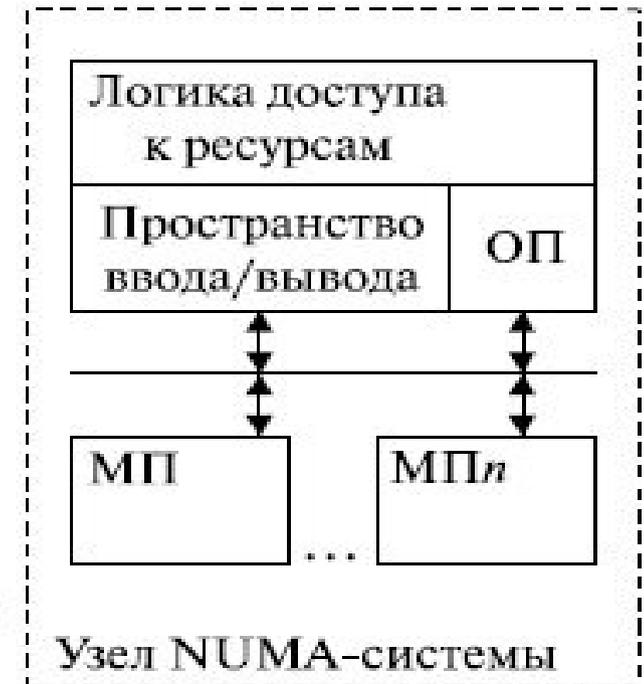
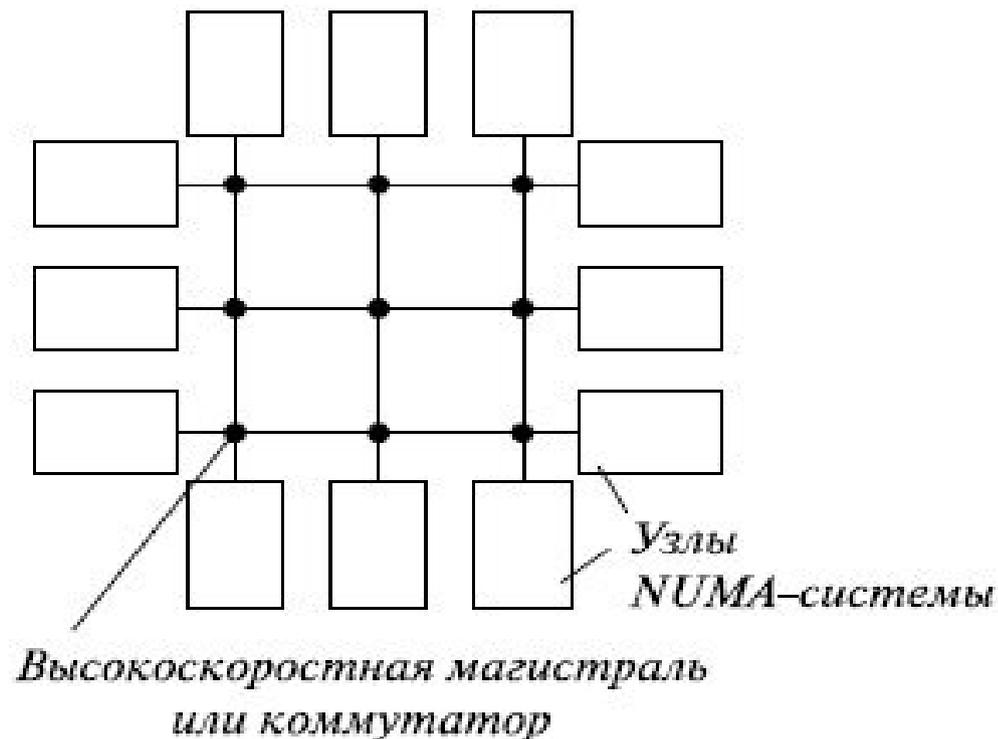
NUMA

Для большей масштабируемости мультимикропроцессоров приспособлена архитектура NUMA (NonUniform Memory Access — с неоднородным доступом к памяти). Как и мультимикропроцессоры UMA, они обеспечивают единое адресное пространство для всех микропроцессоров, но, в отличие от машин UMA, доступ к локальным модулям памяти происходит быстрее, чем к удаленным. Машины NUMA имеют три ключевые характеристики, которыми в совокупности отличают их от других мультимикропроцессоров:

1. Существует одно адресное пространство, видимое для всех микропроцессоров.
2. Доступ к удаленной памяти производится с использованием команд LOAD и STORE.
3. Доступ к удаленной памяти происходит медленнее, чем доступ к локальной памяти. Доступ микропроцессора к собственной Локальной памяти производится напрямую, что намного быстрее, чем доступ к удаленной памяти через коммутатор или сеть.

NUMA

В рамках концепции NUMA реализуется несколько различных подходов, обозначаемых аббревиатурами СОМА, сс-numa и пс-numa.



NUMA

Особенности СОМА

- 1) Локальная память каждого процессора рассматривается как кэш для доступа «своего» процессора.
- 2) Кэши всех процессоров рассматриваются как глобальная память системы, а сама глобальная память отсутствует.
- 3) Данные не привязаны к конкретному модулю памяти и не имеют уникального адреса, остающегося неизменным в течение всего времени существования переменной.
- 4) Данные переносятся в кэш-память того процессора, который последним их запросил. Перенос данных из одного локального кэша в другой не требует участия в этом процессе операционной системы, но подразумевает сложную и дорогостоящую аппаратуру управления памятью.

NUMA

NC-NUMA (No Caching NUMA — NUMA без кэширования)

- 1) Отсутствует кэш-память, это значит, что память гарантированно согласованна
- 2) Каждое слово памяти находится только в одном месте, нет копий.
- 3) От того, в какой памяти находится слово, зависит производительность.
- 4) Имеется страничный сканер, который может перемещать страницы памяти между блоками памяти в зависимости от статистики.

NUMA

CC-NUMA (Cache Coherent Non-Uniform Memory Architecture)

1) Наличие кэша у процессоров.

2) Совместимость кэшей на программном или аппаратном уровне.

NUMA

На сегодня максимальное число процессоров в сс-NUMA-системах может превышать 1000 (серия Origin3000). Один из наиболее производительных суперкомпьютеров - Tera 10 - имеет производительностью 60 Тфлопс и состоит из 544 SMP-узлов, в каждом из которых находится от 8 до 16 процессоров Itanium 2

According to the TOP500 list (as of November 2010), it is currently the 142nd fastest computer in the world running at 52.84 teraFLOPs (52.84 trillion floating point calculations per second) using nearly 10,000 processor cores (in about 4800 dual-core processors). It runs the Linux operating system, with an SMP kernel specially modified to handle very large symmetric clusters currently cadenced by a 8-processors central system.

Tera 100 is a supercomputer built by Bull SA for the French Commissariat à l'Énergie Atomique. It was turned on on May 26, 2010. The computer, which is located in Evry-Courcouronnes is able to sustain around 1 petaFLOPs maximum performance and a peak at 1.25 petaFLOPs. It has 4300 Bullx Series S servers ('Mesca'), 140,000 Intel Xeon 7500 processor cores, 300 TB of memory. The Interconnect is QDR InfiniBand. The file system has a throughput of 500 GB/s and total storage of 20 PB. It uses the SLURM resource manager for scheduling batch jobs. Tera 100 uses Bull XBAS Linux, a Red Hat Enterprise Linux derivative. As of June 2011, TOP500 deems it the ninth fastest supercomputer in the world.

ИСТОЧНИКИ

1. Лекции Кашковского В.В.
2. <http://studdi.ru/lection/avs/lection9.html>
- 3.