

Параллельные архитектуры с
распределённой памятью

**Массивно-параллельные
системы (MPP)**

Определение

Массивно-параллельные системы (МРР) – класс архитектур параллельных вычислительных систем, состоящих из вычислительных узлов, связанных через коммуникационную среду. Главная особенность: память физически разделена.



Архитектура

Система состоит из однородных *вычислительных узлов*, включающих:

- один или несколько центральных процессоров (обычно RISC),
- локальную память (прямой доступ к памяти других узлов невозможен),
- коммуникационный процессор или сетевой адаптер
- жесткие диски
- устройства ввода-вывода

К системе могут быть добавлены специальные узлы ввода-вывода и управляющие узлы. Узлы связаны через некоторую коммуникационную среду (высокоскоростная сеть, коммутатор и т.п.)

Операционная система

Существуют два основных варианта:

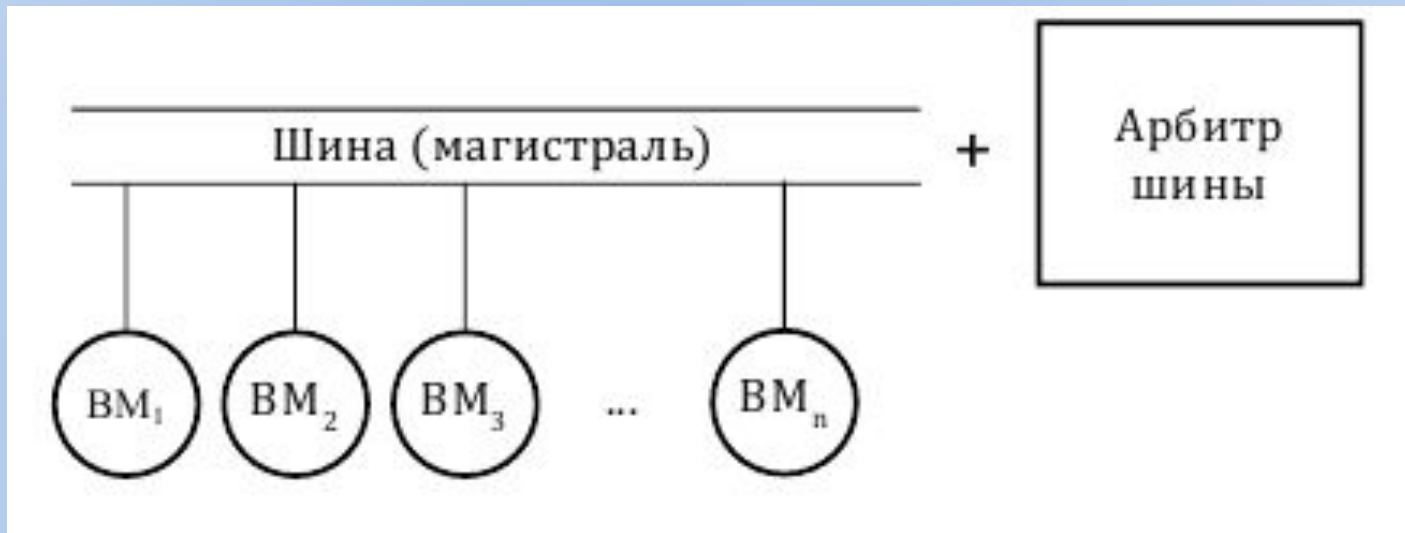
- Полноценная ОС работает только на управляющей машине (front-end), на каждом узле работает сильно урезанный вариант ОС, обеспечивающие только работу расположенной в нем ветви параллельного приложения.
- На каждом узле работает полноценная UNIX-подобная ОС устанавливаемая отдельно на каждом узле (вариант, близкий к кластерному подходу) .

Коммуникационные сети (КС) в МВС с распределённой памятью

Наиболее распространённые решения при построении КС для МВС с распределённой памятью:

- Архитектура со связями через общую шину
- Архитектура со связями через несколько шин
- Архитектура со связями через многоступенчатый коммутатор

Архитектура со связями через общую шину



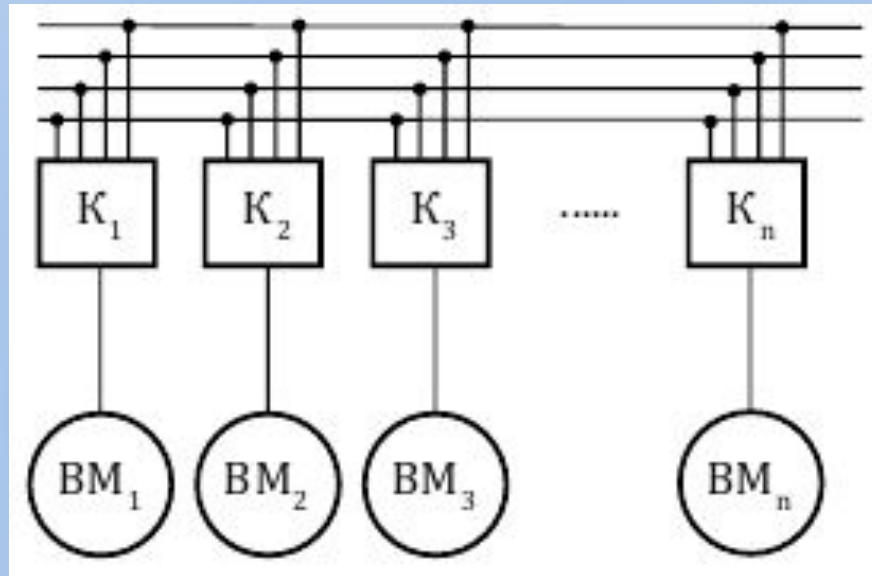
Достоинства:

- Легко и дешево реализуема
- Легко наращивается число вычислительных модулей

Недостатки:

- Требуется арбитр шины – система работает в режиме разделения времени

Архитектура со связями через несколько шин



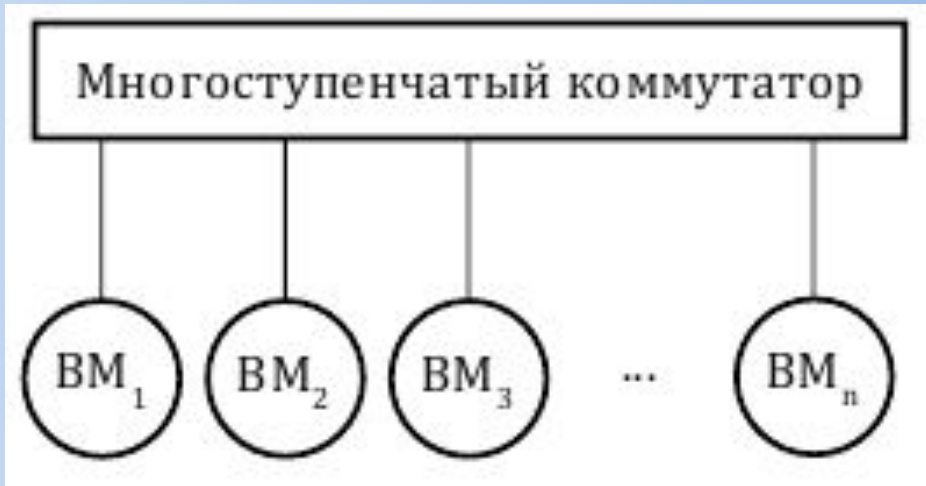
Достоинства:

- Большая производительность и надёжность по сравнению с архитектурой с одной шиной

Недостатки:

- Для организации эффективной работы требуется n коммутаторов (высокие аппаратные затраты)

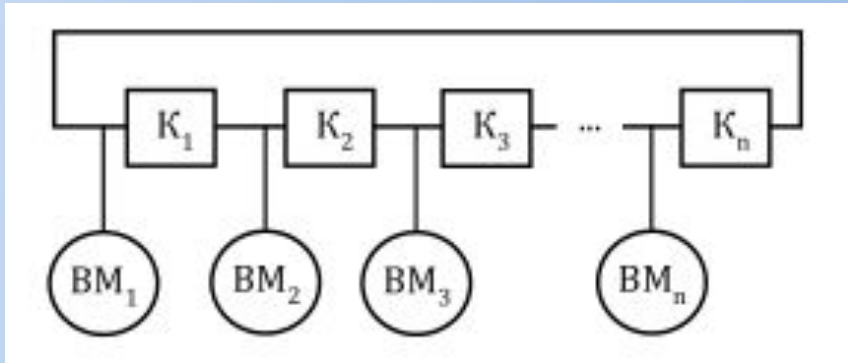
Архитектура со связями через многоступенчатый коммутатор



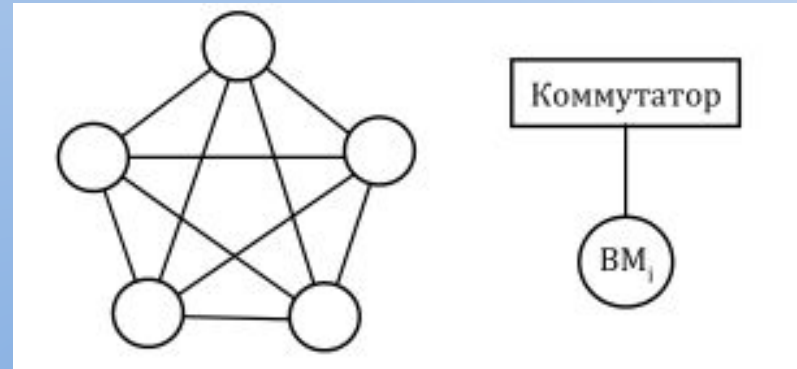
Разновидности данной коммутации:

- Кольцо
- Звезда
- Полная взаимосвязь
- Решётка
- Гиперкуб
- Двойной тор

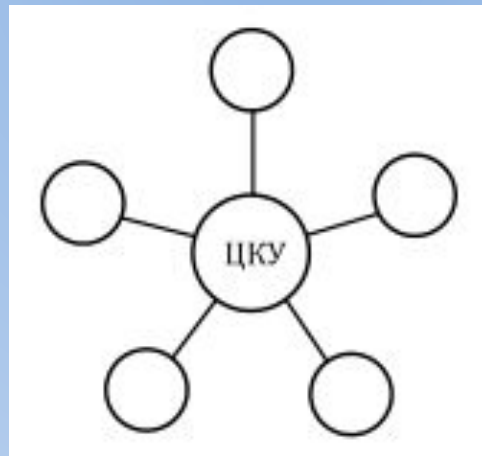
Архитектура со связями через многоступенчатый коммутатор



кольцо

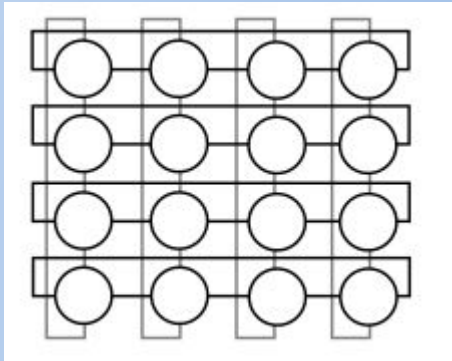


полный набор
связей

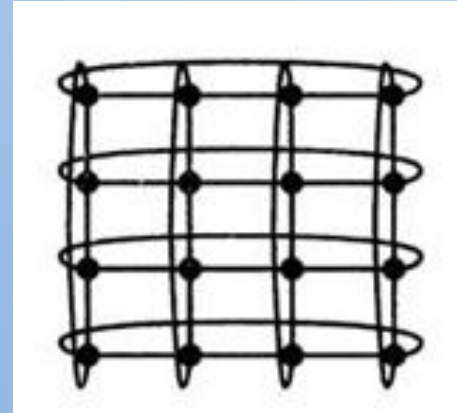


звезда

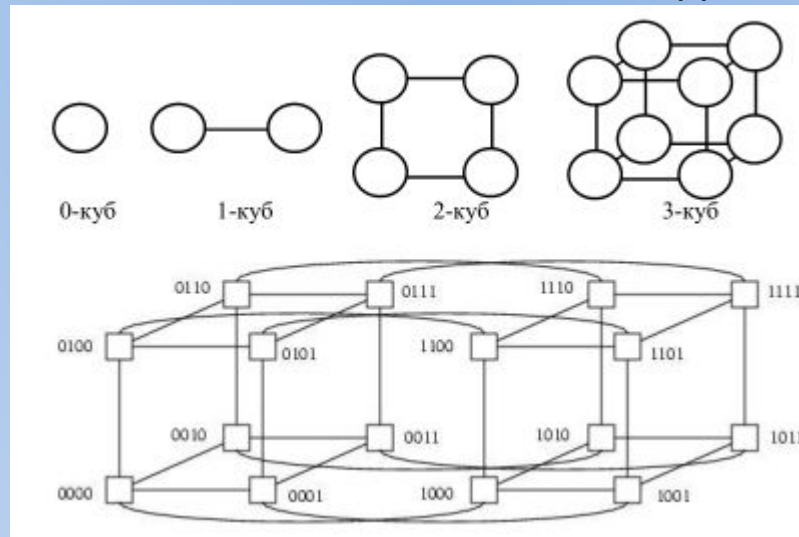
Архитектура со связями через многоступенчатый коммутатор



решётка



двойной тор



гиперкуб

Преимущества архитектуры

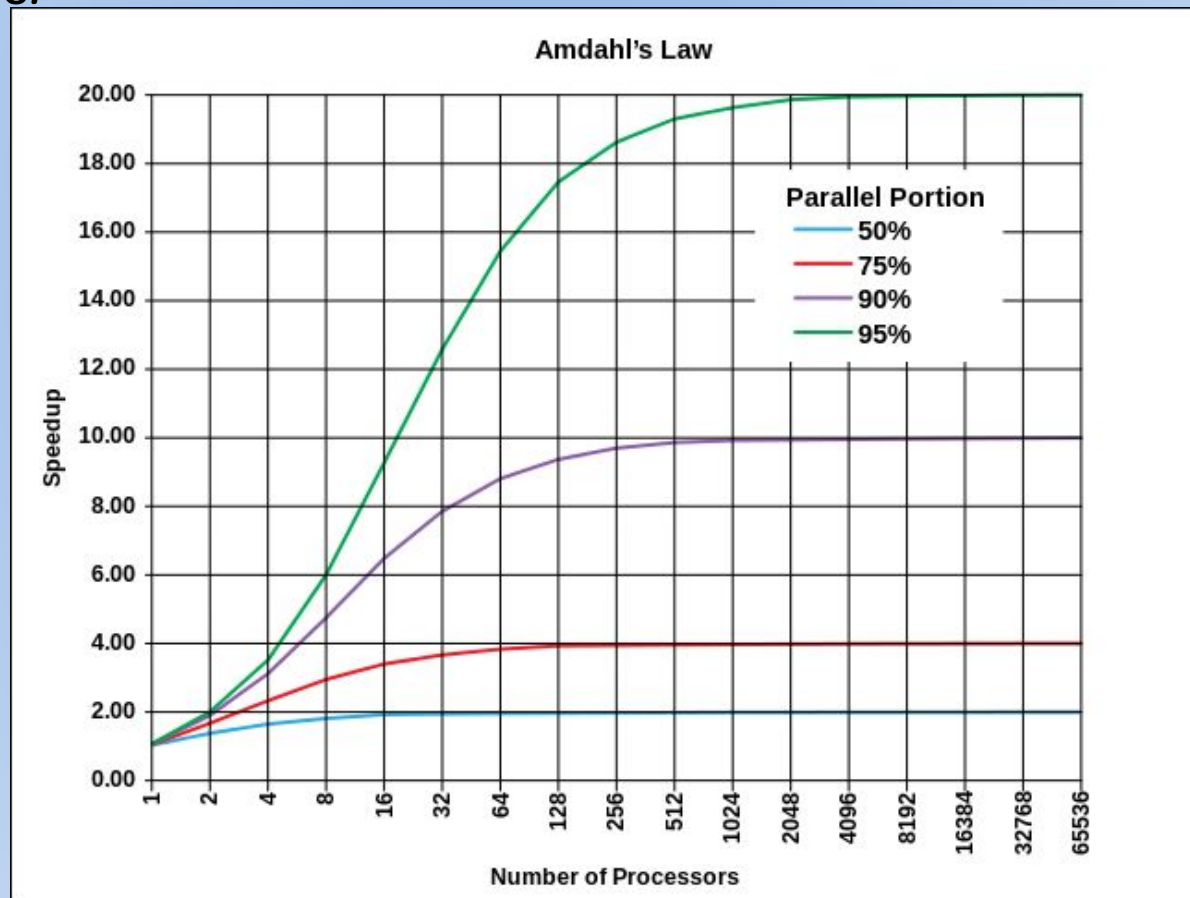
Главным преимуществом систем с раздельной памятью является хорошая масштабируемость: в отличие от SMP-систем, в машинах с раздельной памятью каждый процессор имеет доступ только к своей локальной памяти, в связи с чем не возникает необходимости в потактовой синхронизации процессоров.

Недостатки архитектуры

- Отсутствие общей памяти заметно снижает скорость межпроцессорного обмена, поскольку нет общей среды для хранения данных, предназначенных для обмена между процессорами.
- Требуется специальная техника программирования для реализации обмена сообщениями между процессорами;
- Каждый процессор может использовать только ограниченный объем локальной памяти;
- Вследствие указанных архитектурных недостатков требуются значительные усилия для того, чтобы максимально использовать системные ресурсы.
- Высокая цена программного обеспечения для массивно-параллельных систем с раздельной памятью.

Факторы, влияющие на производительность

Максимальное ускорение работы программы в зависимости от доли последовательных вычислений и числа используемых процессоров.

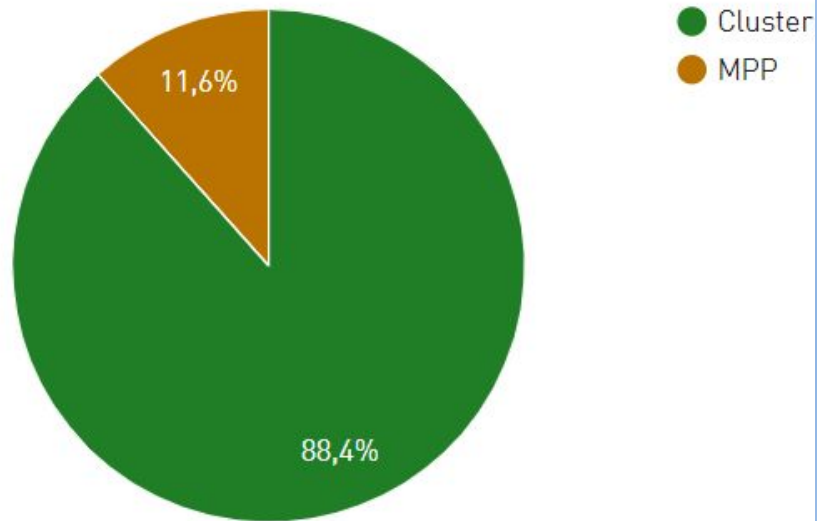


Выдержка из топ 500 за ноябрь 2018

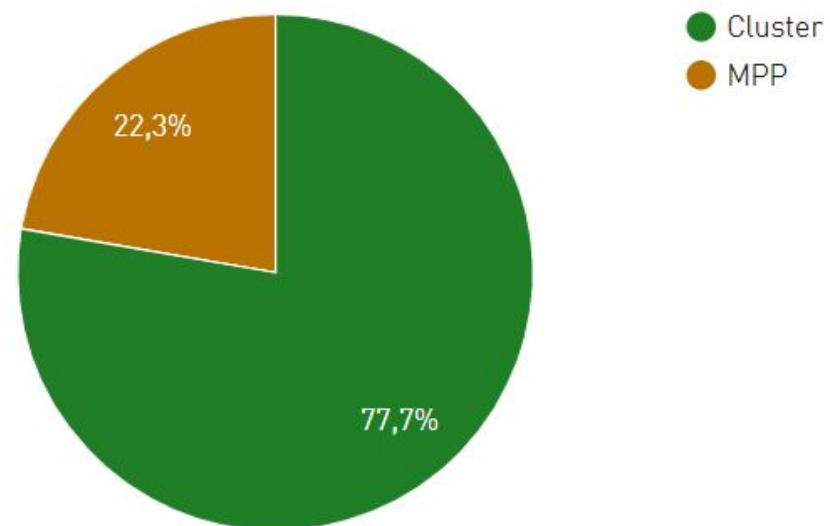
Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
3	Sunway TaihuLight - Sunway MPP, Sunway SW26010 260C 1.45GHz, Sunway , NRCPC National Supercomputing Center in Wuxi China	10,649,600	93,014.6	125,435.9	15,371
5	Piz Daint - Cray XC50, Xeon E5-2690v3 12C 2.6GHz, Aries interconnect , NVIDIA Tesla P100 , Cray Inc. Swiss National Supercomputing Centre (CSCS) Switzerland	387,872	21,230.0	27,154.3	2,384.2
6	Trinity - Cray XC40, Xeon E5-2698v3 16C 2.3GHz, Intel Xeon Phi 7250 68C 1.4GHz, Aries interconnect , Cray Inc. DOE/NNSA/LANL/SNL United States	979,072	20,158.7	41,461.2	7,578.1
9	Titan - Cray XK7, Opteron 6274 16C 2.200GHz, Cray Gemini interconnect, NVIDIA K20x , Cray Inc. DOE/SC/Oak Ridge National Laboratory United States	560,640	17,590.0	27,112.5	8,209
10	Sequoia - BlueGene/Q, Power BQC 16C 1.60 GHz, Custom , IBM DOE/NNSA/LLNL United States	1,572,864	17,173.2	20,132.7	7,890

Архитектуры в топ 500 на ноябрь 2018

Architecture System Share



Architecture Performance Share



Architecture	Count	System Share (%)	Rmax (GFlops)	Rpeak (GFlops)	Cores
Cluster	442	88,4	1,099,791,214	1,767,058,433	38,445,754
MPP	58	11,6	315,164,368	446,250,924	20,640,564

Titan



Sequoia BlueGene/Q

