

Архитектура параллельных вычислительных систем



К.Т.Н., доцент

**Костичев Сергей
Валентинович**

Суперкомпьютеры

snenv@mail.ru

Учебные вопросы:

- **Определения суперкомпьютера.**
- **Суперкомпьютеры и классификация Флинна**
- **Топологии связей в массово параллельных системах**
- **Программы развития суперкомпьютеров**



Определения суперкомпьютера

1976 год –появление 1-й векторной системы Cray 1- начало эры суперкомпьютеров

Оксфордский словарь ВТ1986 года:

Суперкомпьютер - компьютер производительностью 10 Mflops

Основано на **производительности** – необходимость постоянно корректировать (сейчас производительность настольных систем на два порядка выше).

В начале 90-х годов граница около 300 MFLOPS

Экономическое: система, цена которой выше 1–2 млн. \$

Философское: компьютер, мощность которого всего на порядок меньше необходимой для решения современных задач



Основные признаки, характеризующие суперкомпьютер

- высокая производительность
- самый современный технологический уровень (например, GaAs (Арсенид галлия) -технология);
- специфические архитектурные решения, направленные на повышение быстродействия (например, наличие операций над векторами);
- цена, обычно свыше 1-2 млн. долл.

Компьютер Cray-1 стоил около 8 млн. долларов



Суперкомпьютеры и классификация Флинна

Классификация Флинна: все компьютеры делятся на 4 класса в зависимости от числа потоков команд и данных.

Класс SISD – суперкомпьютеры отсутствуют

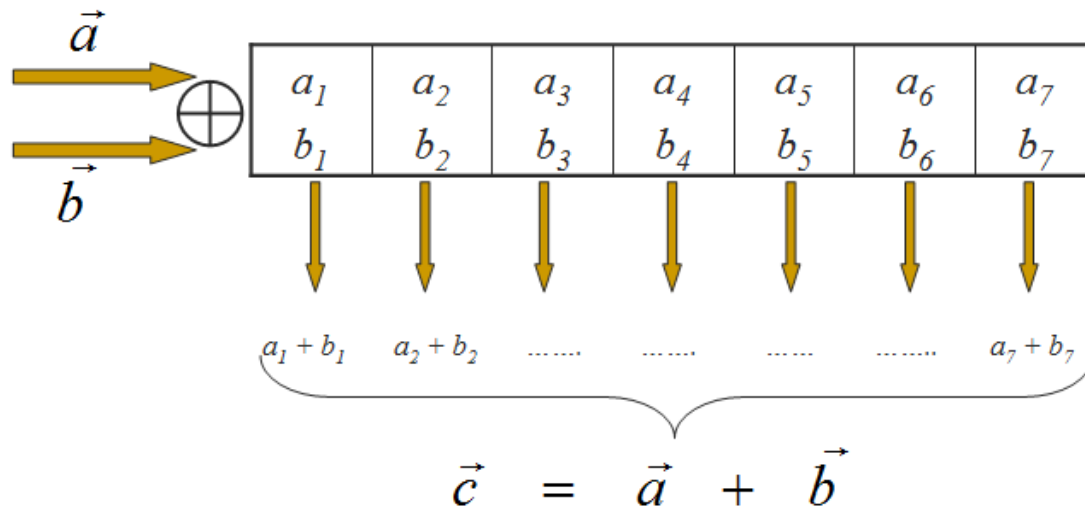


Класс SIMD – однопроцессорные векторные суперкомпьютеры.

- Аппаратная реализация работы с векторами.
- Наличие в системе команд векторных операций
- Наличие регистров маски - векторные команды выполняются над элементами векторов, на которые указывает маска
- Сфера применения ограничена.



Основа векторных вычислений – концепция **конвейеризации**
т.е. явное сегментирование арифметического устройства

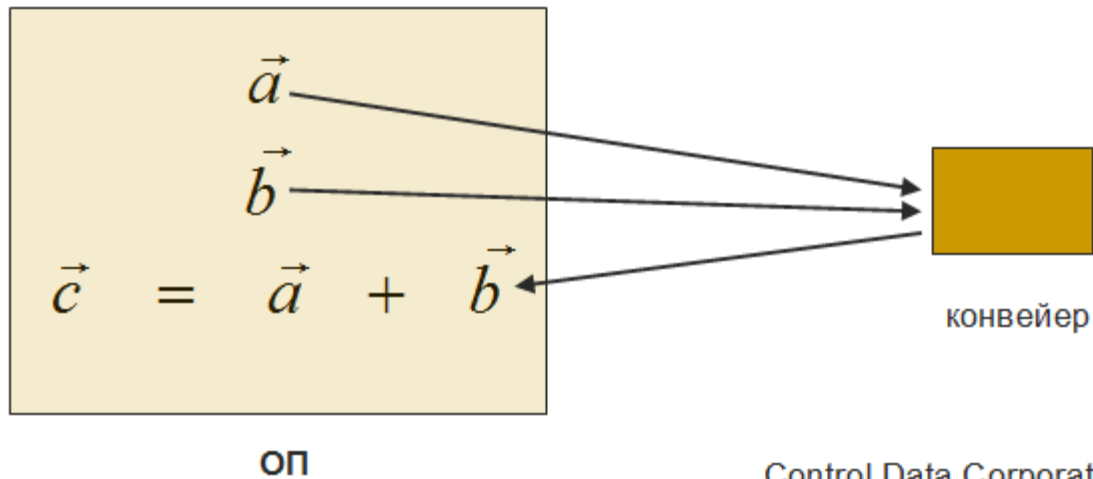


Для реализации возможности ускорения нужно **подавать данные в АУ достаточно быстро.**

Предложены 2 основные формы организации подачи данных на конвейер:

- процессор типа «память-память»
- процессор типа «регистр-регистр»

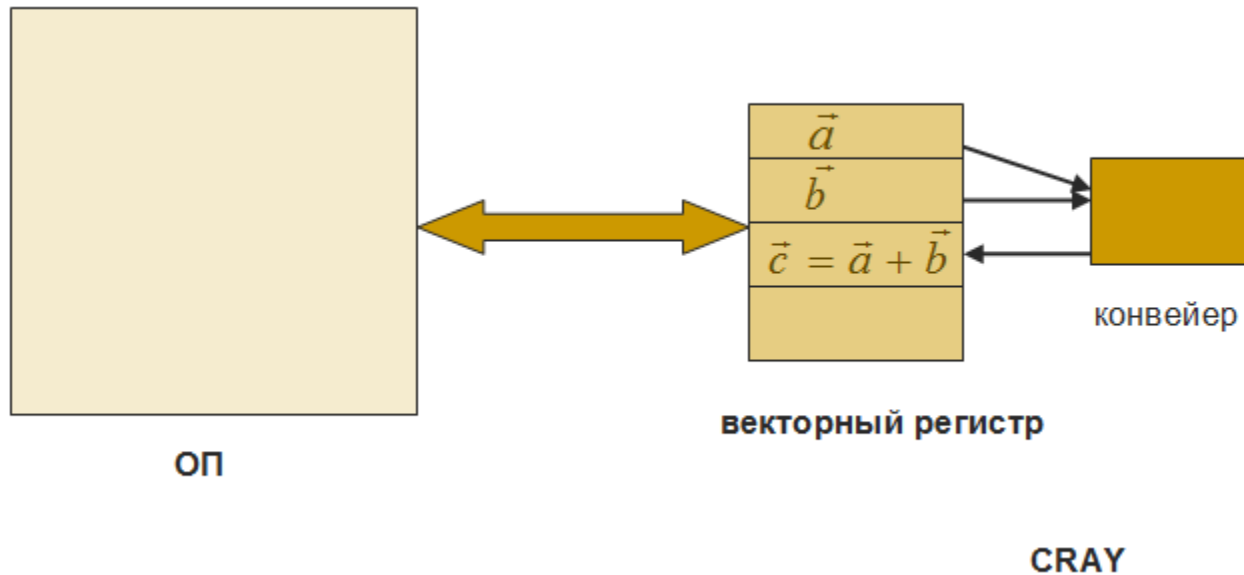
Процессор типа «память-память»



Control Data Corporation – CDC

CYBER 203, CYBER 205

Процессор типа «регистр-регистр»



Усложнение схемы перемещения данных на конвейер привело к появлению **дополнительных требований** к алгоритму

Начиная с Cray-1 векторные суперкомпьютеры имеют средство ускорения векторных вычислений - **зацепление команд**.

Пример:

$$V2=V0*V1$$

$$V4=V2+V3$$

2-я команда не может начать выполняться сразу за 1-й – (1-я команда должна сформировать регистр V2, что требует определенного количества тактов).

Средство зацепления позволяет 2-й команде начать выполнение, не дожидаясь полного завершения 1-й: одновременно с появлением первого результата в регистре V2 его копия направляется в устройство сложения, и запускается 2-я команда.



Класс MIMD – многопроцессорные векторные суперкомпьютеры.

Характеристики:

- симметричность (равноправность) всех процессоров системы
- разделение всеми процессорами общего поля ОП.

Это сильно связанные системы

- **SMP** (симметричные мультипроцессорные системы)
- **PVP** (параллельные векторные системы)

Определение: Сильно-связанные ВС основаны на объединении процессоров на общем поле оперативной памяти



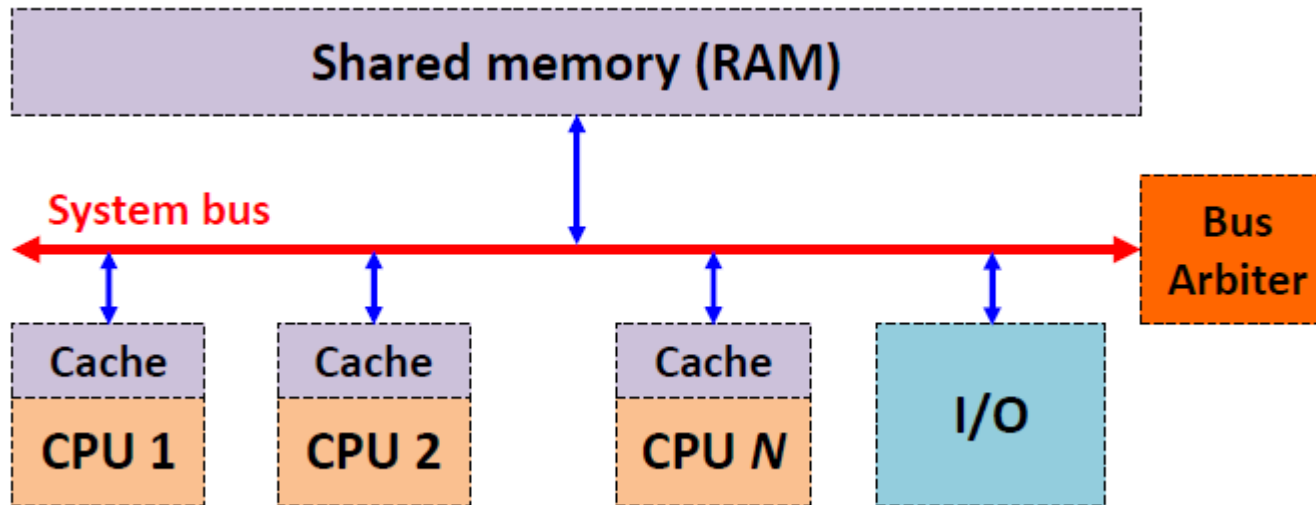
Если в однопроцессорных векторных ЭВМ для создания эффективной программы ее надо **векторизовать**, то в многопроцессорных появляется **задача распараллеливания** программы для ее выполнения одновременно на нескольких процессорах.

Задача распараллеливания является более сложной, (нужно организовать синхронизацию параллельно выполняющихся процессов).



SMP-система

Система состоит из нескольких однородных процессоров и массива общей памяти (обычно из нескольких независимых блоков). Все процессоры имеют доступ к любой точке памяти с одинаковой скоростью.



Система работает под управлением единой ОС (обычно UNIX-подобной). ОС автоматически (в процессе работы) распределяет нити по процессорам, но возможна и явная привязка.

Достоинства SMP-систем:

- простота и универсальность для программирования.
- использование общей памяти увеличивает скорость обмена,
- простота эксплуатации.
- относительно невысокая цена.

Недостатки SMP-систем:

- доступ с разных процессоров к общим данным и **обеспечение**, в связи с этим, **однозначности** (когерентности) содержимого разных кэшей (cache coherence problem).
- системы с общей памятью **плохо масштабируются** (не более 32 процессоров в реальных системах). Контроллер памяти – узкое место



В SMP-системах разработчики стремятся **уменьшить обмен данными на тракте процессоры - память**. Поэтому вместо маленького объема памяти векторных регистров микропроцессоры в суперкомпьютерных SMP-системах **снабжаются кэш - памятью очень большого размера**



PVP (параллельные векторные системы) *система*

Основным признаком PVP-систем является **наличие специальных векторно-конвейерных процессоров**, с командами однотипной обработки векторов независимых данных.

Несколько таких процессоров (1-16) работают одновременно над общей памятью (**аналогично SMP**) в рамках многопроцессорных конфигураций.

Несколько узлов могут быть объединены с помощью коммутатора (**аналогично MPP** - Массово-параллельные системы).

Примеры: линия векторно-конвейерных компьютеров CRAY



**Класс MIMD – массово параллельные
вычислительные системы (МРР)**



История

В середине 90-х годов темп развития сверхпроизводительных ВС упал.

Причины :

- спад государственной поддержки программы развития сверхпроизводительных ВС (прекращение «холодной войны»),
- отсутствие рынка сбыта дорогих ВС (были подходящие мини-суперЭВМ гораздо более дешевые и доступные).



Господство технологии КМОП:

- медленная
- + меньшее энергопотребление
- + значительно большую степень интеграции,
- + отработанность технологии

КМОП: в одной микросхеме несколько миллионов транзисторов -> реализовать **высокопроизводительный суперскалярный процессор** (пример, модели процессоров типа Pentium).

На базе этих процессоров с одной и той же архитектурой строились массовые (и достаточно производительные) персональные компьютеры.



Векторная вычислительная система

Высокая стоимость:

- уникальная архитектура и структура -> надо разрабатывать десятки оригинальных микросхем.
- в качестве элементной базы использовались микросхемы ECL (эмиттерно связанная логика) типа.

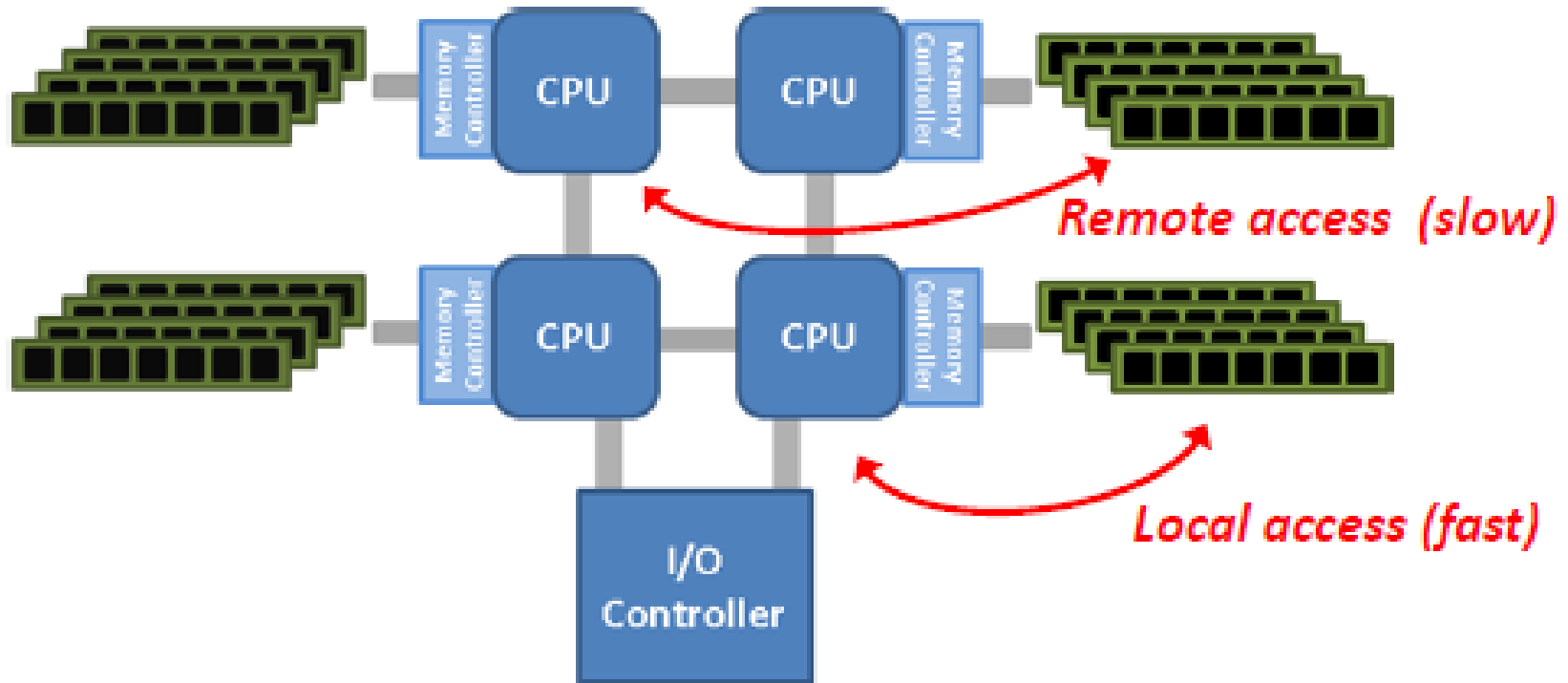
Это была

- + самая быстродействующая элементная база
- самая энергоемкая
- самая малая по степени интеграции

Осознание факта: **из СБИС можно строить структуры с массовым параллелизмом**



Массово параллельные ВС со скалярными вычислительными узлами и общей ОП (NUMA)



Свойства:

- **множество вычислительных узлов** (от нескольких сотен до нескольких десятков тысяч) со скалярной (суперскалярной) структурой;
- вычислительный узел может быть как **однопроцессорным**, так и **многопроцессорным с общей** логически и физически **ОП** (как с однородным, так и неоднородным доступом к ОП узла);
- **много потоков команд** (в соответствии с числом процессоров во всех вычислительных узлах);
- **много потоков данных** (в соответствии с числом процессоров во всех вычислительных узлах);



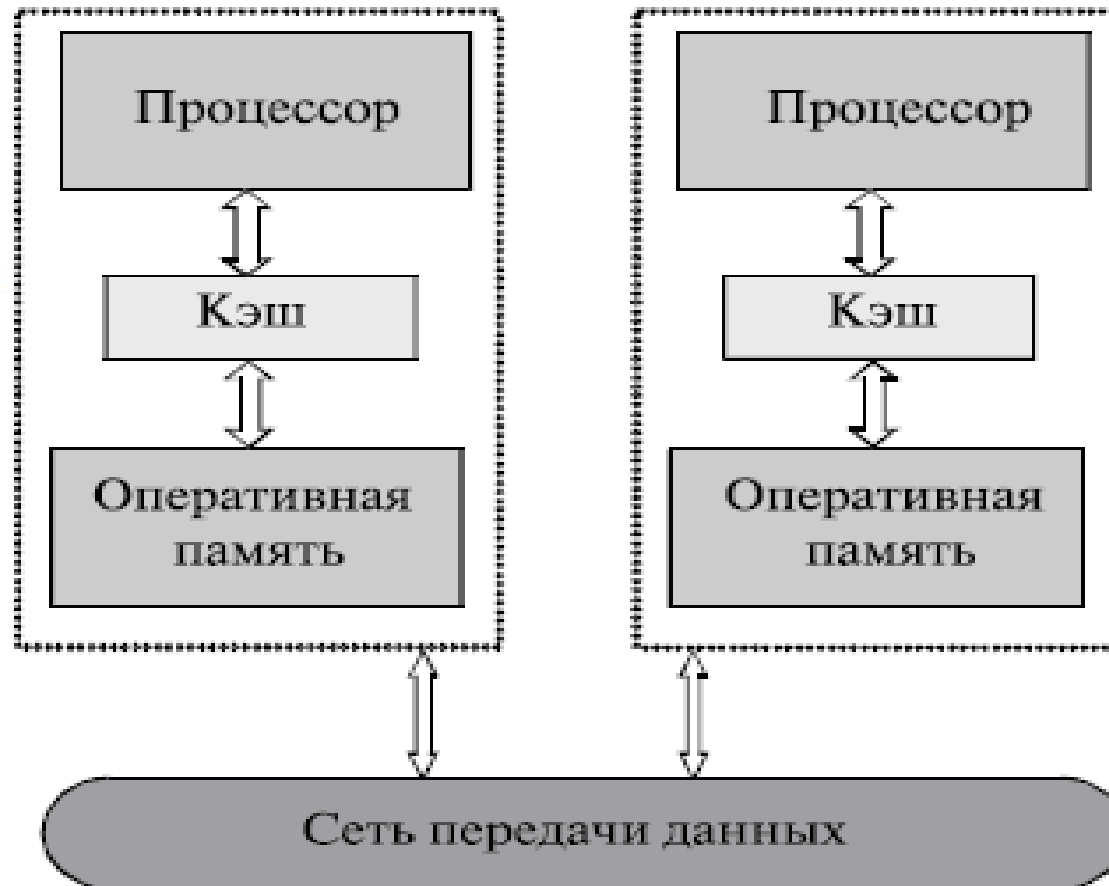
Свойства (продолжение):

- **ОП ВС логически единая, но физически распределена между узлами;**
- **неравноправный доступ процессоров к общей ОП ВС** (быстрый – к локальной ОП узла, более медленный – к ОП других узлов);
- каждый вычислительный узел работает под управлением своей ОС;
- имеется общая ОС, обеспечивающая образ единого вычислительного ресурса для пользователей.

Примеры: ВС CRAY XT4, CRAY XT5 CRAY XE6.



Массово параллельные ВС со скалярными вычислительными узлами и распределенной ОП (МРР)



Свойства:

- **множество вычислительных узлов** (диапазон от нескольких сотен до нескольких сотен тысяч) со скалярной архитектурой;
- вычислительный узел может быть как **однопроцессорным**, так и **многопроцессорным с общей** логически и физически **ОП** (как с однородным, так и неоднородным доступом к оперативной памяти узла);
- **много потоков команд** (в соответствии с числом процессоров во всех вычислительных узлах);
- **много потоков данных** (в соответствии с числом процессоров во всех вычислительных узлах);



- **Свойства (продолжение):**
- **ОП ВС физически распределена между вычислительными узлами;**
- **процессоры вычислительного узла имеют доступ только к локальной ОП узла.**
- **доступ к ОП других узлов осуществляется через коммуникационную подсистему с использованием механизма передачи сообщений.**
- **каждый вычислительный узел работает под управлением своей ОС;**
- **имеется общая ОС, обеспечивающая образ единого вычислительного ресурса для пользователей.**

Примеры: ВС Blue Gene/L, RoadRunner



Благодаря масштабируемости, **MPP-системы являются сегодня лидерами по достигнутой производительности компьютера**

С другой стороны, проблемы распараллеливания в MPP-системах по сравнению с кластерами, содержащими немного процессоров, становятся все более трудно разрешимыми.

Эффективность распараллеливания во многих случаях сильно зависит от деталей архитектуры MPP-системы, например **топологии соединения процессорных узлов.**

Класс MIMD Кластеры

Кластеры - самый дешевый способ наращивания производительности уже инсталлированных компьютеров, это **дешевый вариант MPP систем.**

Кластер - набор из нескольких ЭВМ, соединенных через некоторую коммуникационную инфраструктуру. В качестве такой инфраструктуры может выступать обычная компьютерная сеть, однако для повышения производительности желательно иметь высокоскоростные соединения

Кластеры:

- **гетерогенные** (образованы из различных компьютеров)
- **гомогенные** (образованы из одинаковых компьютеров)



Сравнение кластеров с SMP

- **Преимуществом** кластеров по сравнению с **SMP** является *улучшение возможностей масштабирования*.
В отличие от SMP-архитектуры (ограничение - пропускная способность шины), **добавление компьютеров в кластер позволяет увеличивать пропускную способность ОП и подсистем ввода-вывода**



- В кластерных системах для организации взаимодействия между процессами, выполняющимися на разных компьютерах при решении одной задачи, применяются различные модели обмена сообщениями (PVM, MPI и т.п.).
Задача распараллеливания более сложная, чем в SMP-системах.



Основные **достоинства** кластерных суперкомпьютерных систем:

- высокая суммарная производительность;
- высокая надежность работы системы;
- наилучшее соотношение производительность/стоимость;
- возможность динамического перераспределения нагрузок между серверами;
- легкая масштабируемость

Отличие кластеров от мэйнфреймов :

У **кластеров** все ресурсы обычно направлены на то, чтобы решить одну или в крайнем случае несколько задач **насколько возможно быстро**;

мэйнфреймы, как правило, выполняют довольно **большое число задач, конкурирующих друг с другом.**



Топологии связей в массово параллельных системах

Классификация структур MPP систем с распределенной памятью по типу и пропускной способности коммутирующей сети:

- топология решетка (двухмерная, трехмерная);
- топология гиперкуб;
- Топология «каждый с каждым»
- Топология «распределенный коммутатор».



Структуры с топологией решетки

- Пары из процессора и модуля памяти (процессорный элемент ПЭ) соединены непосредственными связями со своими соседями (4 для двумерной решетки, 6 – для трехмерной решетки).
- Каждый ПЭ имеет одно и то же число подключений к сети вне зависимости от числа процессорных элементов в системе.
- Общая пропускная способность такой сети растет линейно относительно числа ПЭ.



- Для синхронизации параллельно выполняющихся в узлах процессов необходим обмен сообщениями. Сообщения должны проходить между любыми двумя узлами. --> Важной характеристикой является **диаметр системы** - максимальное расстояние между узлами.
- При увеличении числа узлов архитектура гиперкуба является более выгодной, чем решетка (меньше диаметр).



Плоская решётка

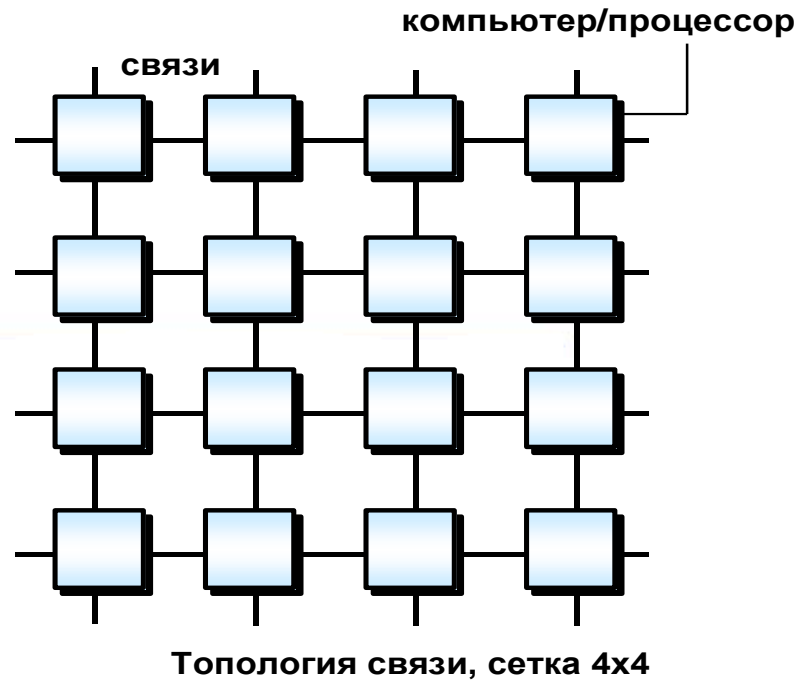
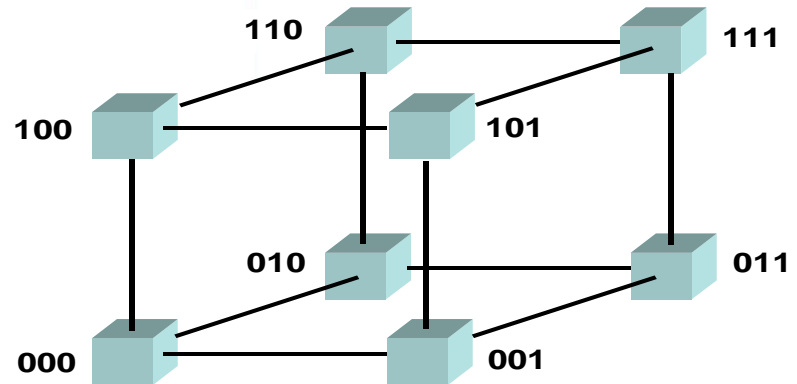
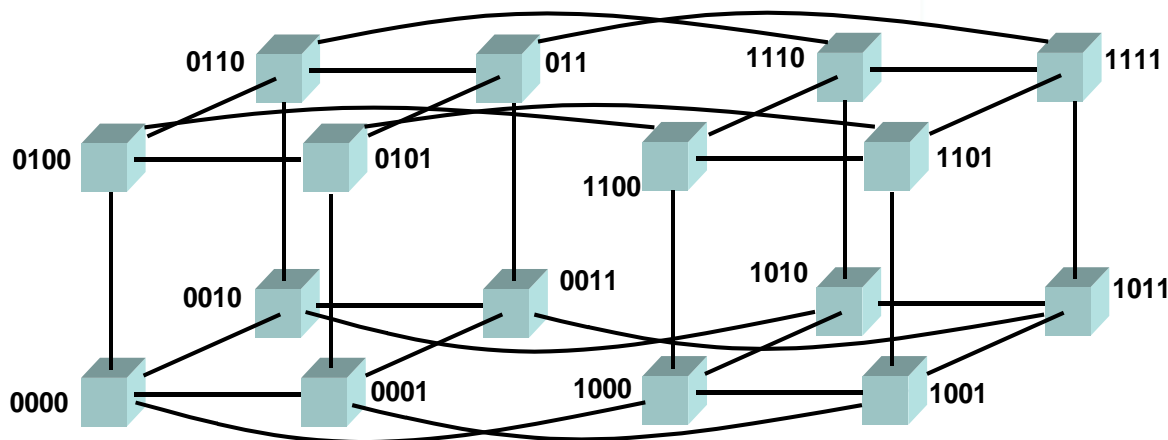


Схема соединения процессоров в виде плоской решётки

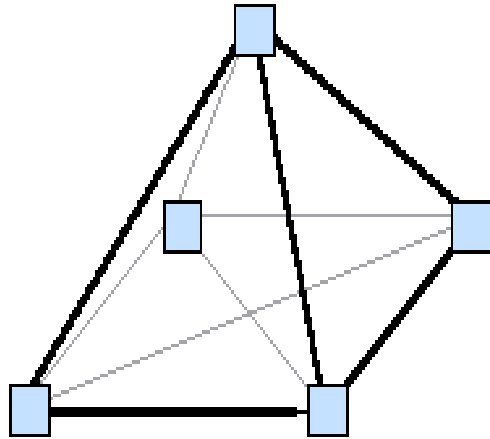
Гиперкуб



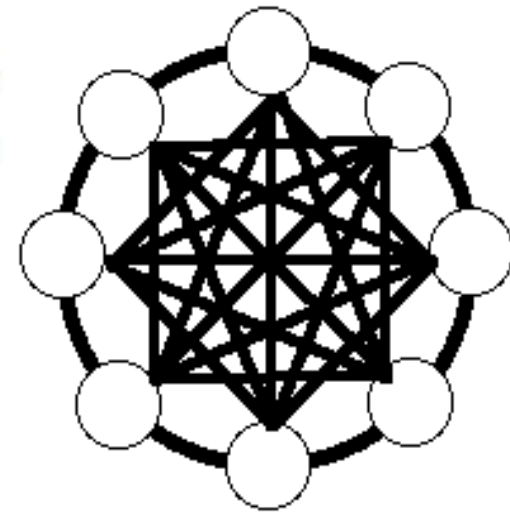
Топология связи, 3-х мерный гиперкуб



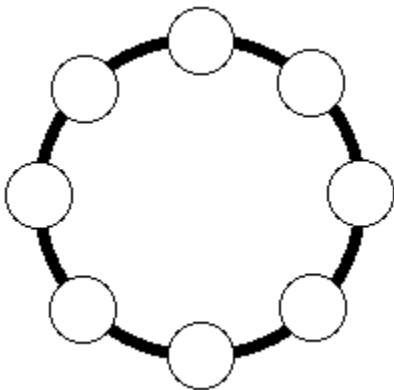
Топология связи, 4-х мерный гиперкуб



Сеть с топологией «клика»



Сеть с топологией «каждый с каждым» - полный граф



Сеть с топологией кольцо

Программы развития суперкомпьютеров

Типовая модель современного суперкомпьютера имеет следующие характеристики:

- высокопараллельная многопроцессорная ВС с быстродействием порядка 100 000 MfloPS и выше;
- емкость оперативной памяти 10 Гбайт и выше;
- емкость дисковой памяти 1-10 Тбайт (1 Тбайт = 1000 Гбайт);
- разрядность 64 -128 бит

В 1996 года Intel создала суперкомпьютер Sandia производительностью 1,4 TFLoPS. Он состоял из 86 шкафов общей площадью 160 кв. м. Имеет 573 Гбайт оперативной и 2250 Гбайт дисковой памяти. Масса компьютера 45 тонн, потребление энергии — 850 кВт
Обеспечение таких технических характеристик обходится довольно дорого - стоимость суперкомпьютеров чрезвычайно высока.



Узнать, какие суперкомпьютеры в настоящее время имеют **максимальную производительность**, можно из официального списка пятисот самых мощных систем мира - Top500 (<http://www.top500.org>), который публикуется два раза в год.



Первые места TOP500 занимают машины с гибридной архитектурой, что подтверждает выводы о том, что суперкомпьютеры нового поколения будут использовать **гибридную схему построения (процессоры + специализированные ускорители)**.

Гибридная архитектура обеспечивает высокое быстродействие и повышает **энергоэффективность** машины.

Для оценки энергоэффективности создан отдельный рейтинг **Green500**, в котором суперкомпьютеры ранжируются по показателю расхода энергии на один миллион операций с плавающей запятой.



Активно развивается также способ повышения **энергоэффективности** за счет использования в суперкомпьютерах **охлаждения горячей** (около 40 градусов) **водой**. Такая технология сейчас активно внедряется, но во избежание протечек требует высокой надёжности всех соединений системы охлаждения.

Повышение **энергоэффективности** также достигается за счет **уменьшение технологических норм** при изготовлении микросхем. По данным Intel, каждый переход на новую технологическую норму снижает энергопотребление примерно на 20%.



**Наличие суперкомпьютеров становится
первостепенной задачей** ведущих мировых держав.
Безусловный лидер - США.

Для обеспечения лидерства в США принимались, принимаются и финансируются государством программы развития суперкомпьютеров (HPCS , УНРС).

Программа **HPCS** закрывала для США технологическую брешь до 2015 года.

Реализовывалась силами IBM, Cray и Sun (проекты PERCS, Cascade и HERO соответственно)



Программа **УНРС** предусматривает создание принципиально новых высокопроизводительных компьютеров **эксафлопсного уровня** и определяет направления работ в области суперкомпьютеров до 2020 года.

Предыдущее **стократное увеличение производительности** для суперкомпьютеров **было достигнуто за 20 лет**, когда закон Мура еще непосредственно влиял на повышение производительности.

В соответствии с программой **УНРС**, следующий рывок по стократному увеличению производительности предстоит выполнить немногим более чем **за пять лет**



В программе **УНРС** выделены следующие основные **проблемы**, которые предстоит решить при создании перспективных ВС различного класса:

- проблема **потребления энергии**;
- проблема **устойчивости ВС** как к обычным сбоям и отказам, так и к информационным атакам;
- проблема **производительности при разработке программ**.



В программе также указаны характерные **черты приложений**, которые предполагается выполнять на системах, созданных в рамках программы :

- обработка массовых потоковых данных от сенсорных устройств;
- работа с большими динамическими графами;
- поддержка принятия решений



Эти приложения соответствуют задачам, предусматривающим интенсивную работу с данными (DIC-задачи). Главное препятствие к эффективному решению таких задач - проблема **преодоления “стены памяти”**, так и не решенная в рамках **программы HPCS**.

Эффективное использование данных давно стало серьезной проблемой, и имеет специальный термин — **«стена памяти»**. Задержка при обращении процессора к ОП для большинства современных систем лежит в диапазоне 150–300 тактов, и пока нет доступных технологий, которые могли бы это значение принципиально уменьшить.



Следующие **обстоятельства** определяют особенность проблем, которые предстоит решить в связи с созданием суперкомпьютеров экзафлопсного уровня :

- **прекращение прямого влияния закона Мура** на производительность ядер процессоров из-за невозможности дальнейшего увеличения тактовой частоты и параллелизма выполнения машинных команд;
- **жесткие ограничения по потребляемой энергии;**
- **проблема «стены памяти».**



Вопросы?

