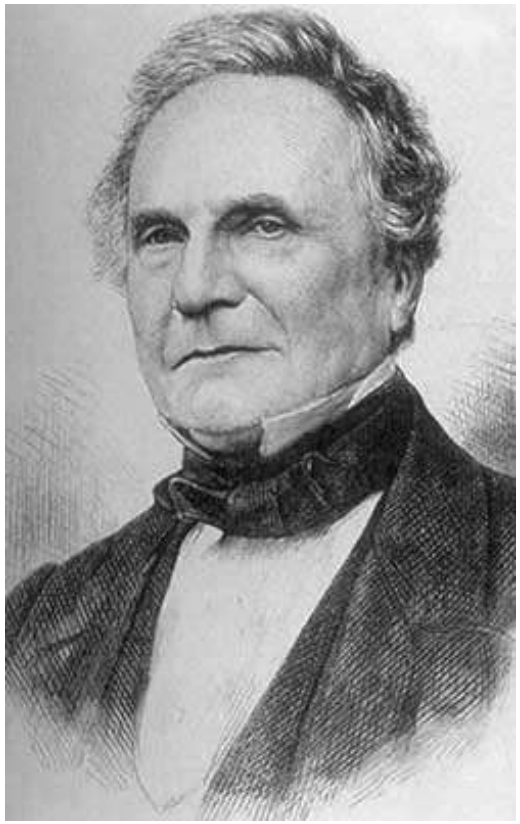
Параллельные вычислительные системы

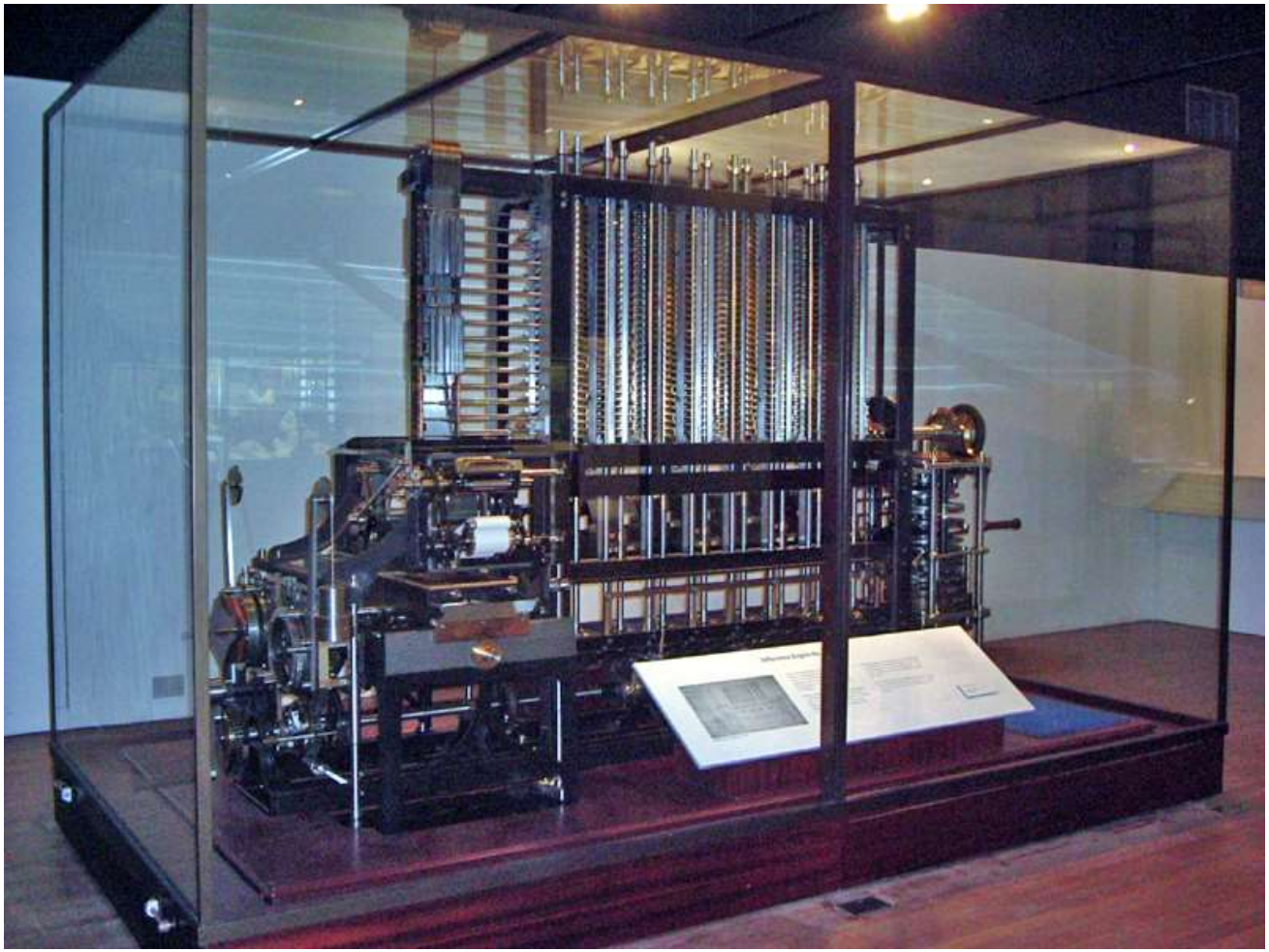
Введение

Чарльз Бэббидж: первое упоминание о параллелизме



- " В случае выполнения серии идентичных вычислений, подобных операции умножения и необходимых для формирования цифровых таблиц, машина может быть введена в действие с целью выдачи нескольких результатов одновременно, что очень существенно сократит весь объем процессов"

Чарльз Бэббидж: вычислительная машина



Определение параллелизма

- ***А.С. Головкин***

Параллельная вычислительная система - вычислительная система, у которой имеется по меньшей мере более одного устройства управления или более одного центрального обрабатывающего устройства, которые работают одновременно.

Определение параллелизма

- ***П.М. Коуги***

Параллелизм - воспроизведение в нескольких копиях некоторой аппаратной структуры, что позволяет достигнуть повышения производительности за счет одновременной работы всех элементов структуры, осуществляющих решение различных частей этой задачи.

Определение параллелизма

- ***Хокни, Джесссхоуп***

Параллелизм - способность к частичному совмещению или одновременному выполнению операций.

Развитие элементной базы и рост производительности параллельных вычислительных систем

Период	Элементная база	Задержка	Быстрод-е элементной базы	Быстрод-е ЭВМ
1940-1950	Лампы	1 мкс	Рост В 1000 раз	Рост в 100000 раз
Начало 1960гг	Дискретные германиевые транзисторы	0,3 мкс		
Середина 1960 гг	Биполярные ИС малой степени интеграции	0,1 мкс=10 нс		
Середина 1970 гг	- «» -	До 1 нс		
Конец 1970	Переход к МОП	До 10 нс	Снижение	Рост

Области применения параллельных вычислительных систем

- предсказания погоды, климата и глобальных изменений в атмосфере;
- науки о материалах;
- построение полупроводниковых приборов;
- сверхпроводимость;
- структурная биология;
- разработка фармацевтических препаратов;
- генетика;

Области применения параллельных вычислительных систем

- квантовая хромодинамика;
- астрономия;
- транспортные задачи;
- гидро- и газодинамика;
- управляемый термоядерный синтез;
- эффективность систем сгорания топлива;
- геоинформационные системы;

Области применения параллельных вычислительных систем

- разведка недр;
- наука о мировом океане;
- распознавание и синтез речи;
- распознавание изображений;
- военные цели.

- Ряд областей применения находится на стыках соответствующих наук.

Оценка производительности параллельных вычислительных систем

- Пиковая производительность - величина, равная произведению пиковой производительности одного процессора на число таких процессоров в данной машине.

Параллельные вычислительные системы

Классификация

Классификация Флинна

- Основана на том, как в машине увязываются команды с обрабатываемыми данными.

- ***Поток -*** последовательность элементов (команд или данных), выполняемая или обрабатываемая процессором.

Классификация Флинна

- ***ОКОД (SISD)***

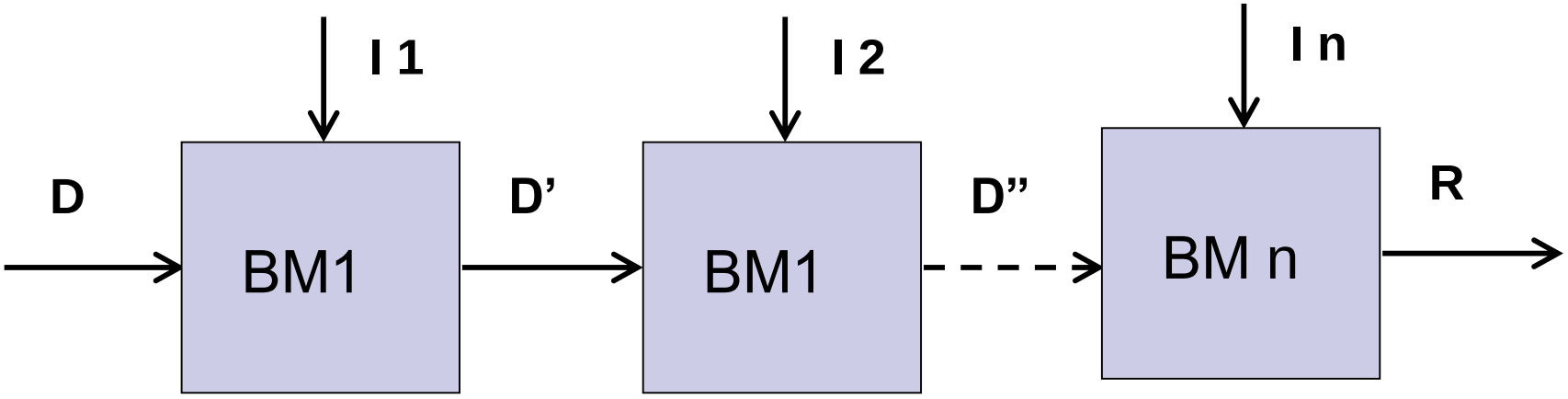
  один поток команд, много потоков данных
- ***МКОД (MISD)***

  много потоков команд, один поток данных
- ***ОКМД (SIMD)***

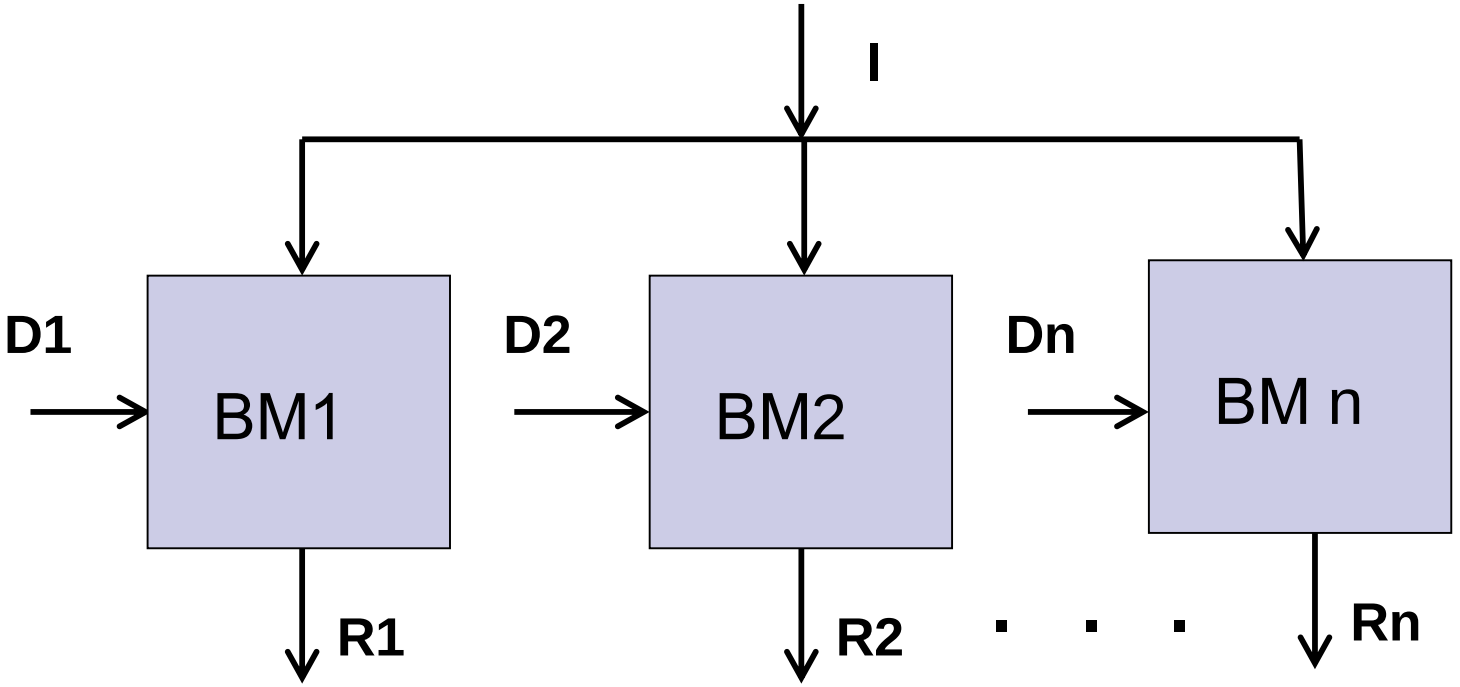
  один поток команд, много потоков данных
- ***МКМД (MKMD)***

  много потоков команд, много потоков данных

МКОД – Конвейерные ПВС



ОКМД – Процессорные матрицы



Классификация Флинна - МКМД

- ***SMP*** –

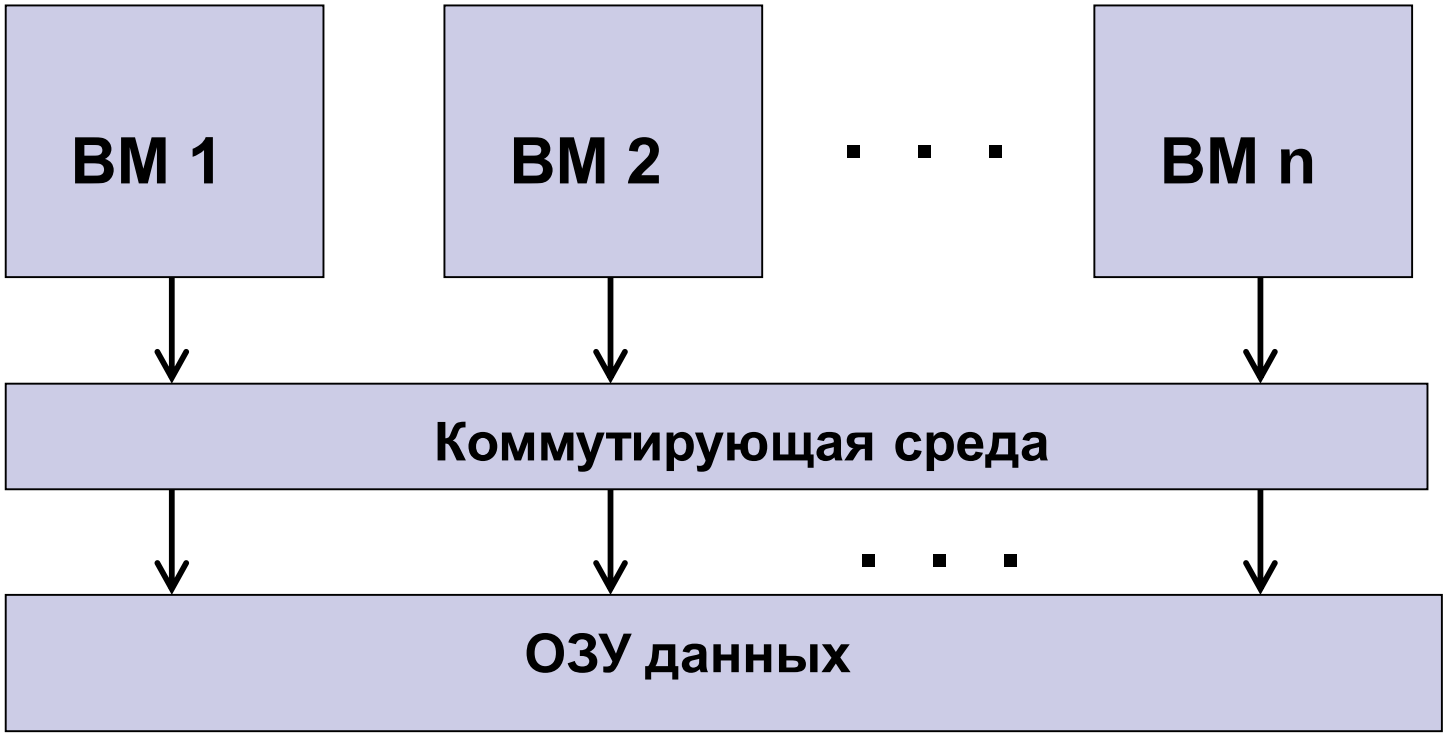
  симметричные мультипроцессорные системы
- ***Кластерные вычислительные системы***
  - Специализированные кластеры
  - Кластеры общего назначения
- ***MPP*** –

  массивно-параллельные системы

Симметричные мультипроцессоры (SMP)

- ***Симметричные мультипроцессоры (SMP)*** - состоят из совокупности процессоров, обладающих одинаковыми возможностями доступа к памяти и внешним устройством и функционирующих под управлением единой ОС.

SMP - симметричные мультипроцессорные системы



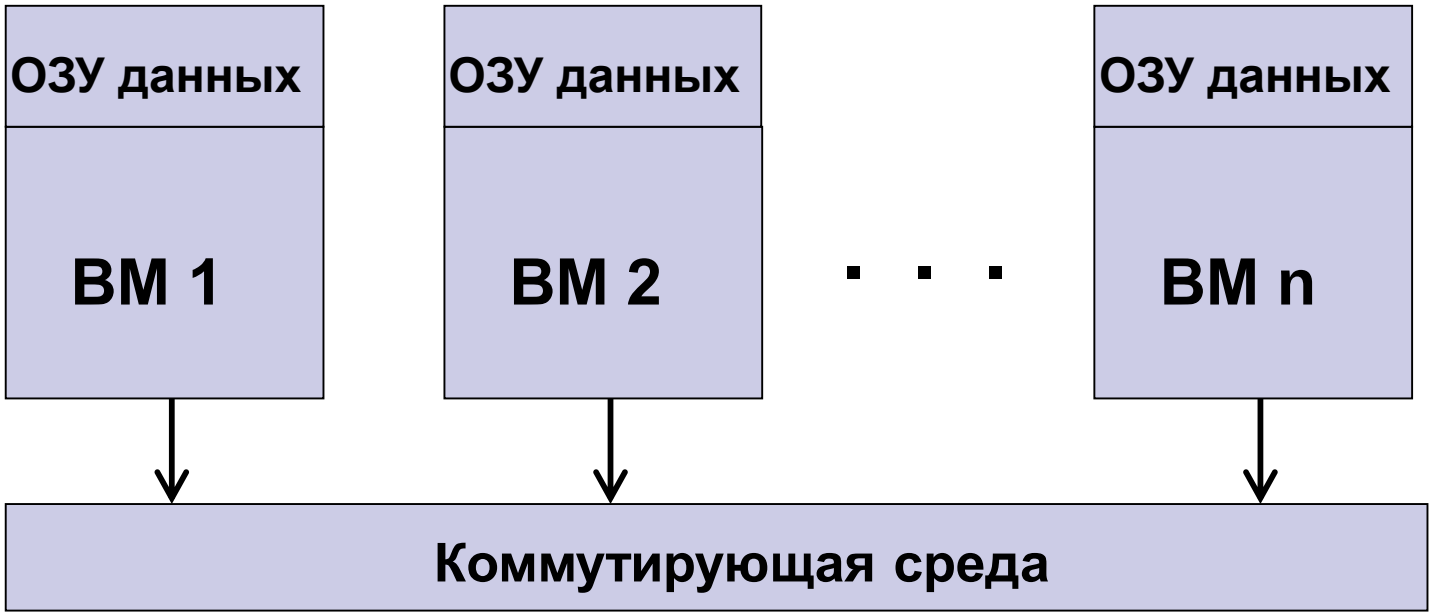
Кластеры

- ***Кластерная система*** – параллельная вычислительная система, создаваемая из модулей высокой степени готовности, объединенных стандартной системой связи или разделяемыми устройствами внешней памяти.

Массивно-параллельная система MPP

- ***Массивно-параллельная система*** – высокопроизводительная параллельная вычислительная система, создаваемая с использованием специализированных вычислительных модулей и систем связи.

Кластеры и массивно-параллельные системы (MPP)



Параллельные вычислительные системы

Конвейерные ВС

Конвейерные ВС

- ***Конвейеризация*** - метод проектирования, в результате применения которого в вычислительной системе обеспечивается совмещение различных действий по вычислению базовых функций за счет их разбиения на подфункции.

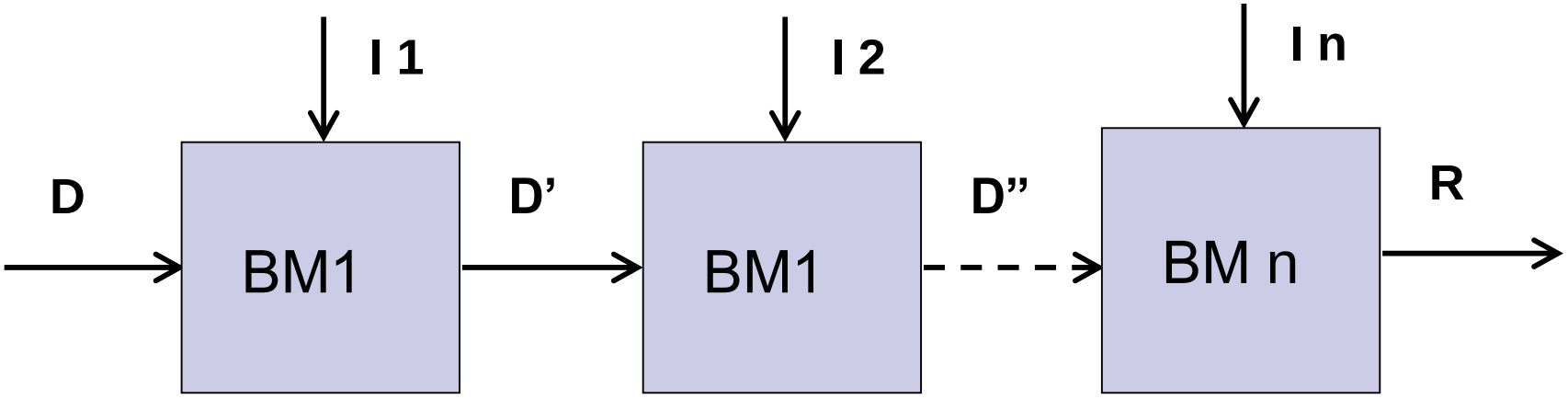
Конвейерные ВС – Условия конвейеризации

- вычисление базовой функции эквивалентно вычислению некоторой последовательности подфункций;
- величины, являющиеся входными для данной подфункции, являются выходными величинами той подфункции, которая предшествует данной в процессе вычисления;
- никаких других взаимосвязей, кроме обмена данными, между подфункциями нет;

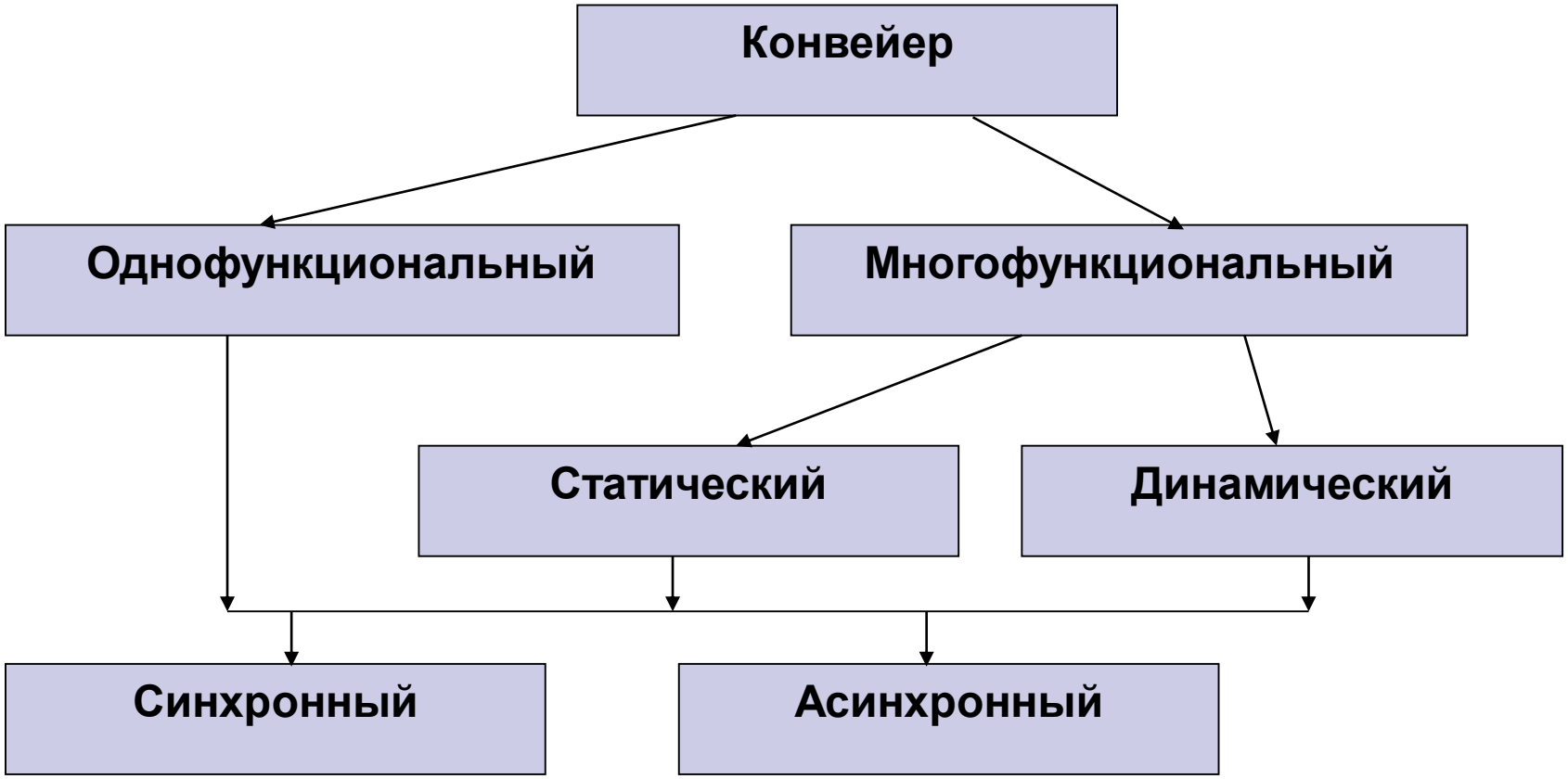
Конвейерные ВС – Условия конвейеризации

- каждая подфункция может быть выполнена аппаратными блоками;
- времена, необходимые для реализации аппаратными блоками своих действий, имеют один порядок величины.

Конвейерные ВС - Архитектура



Конвейерные ВС - Классификация



Конвейерные ВС – Таблица занятости

Время (такт) / Ступень	0	1	2	3	4	…
1	*					
2		*	*			
3				*	*	
…						

Конвейерные ВС – Задача управления

- обеспечение входного потока данных (заполнение конвейера)
- задача диспетчеризации - определение моментов времени, в которые каждый элемент входных данных должен начинать свое прохождение по конвейеру.

Конвейерные ВС – Проблемы управления

- разный период времени обработки данных на разных ступенях;
- обратная связь от текущей ступени к какой-либо из предыдущих;
- множественные пути от текущей ступени к последующим;
- подача элемента данных более чем на одну ступень одновременно (элемент распараллеливания обработки);
- существование между входными элементами зависимостей, которые принуждают к определенному упорядочению связанных с ними вычислений;

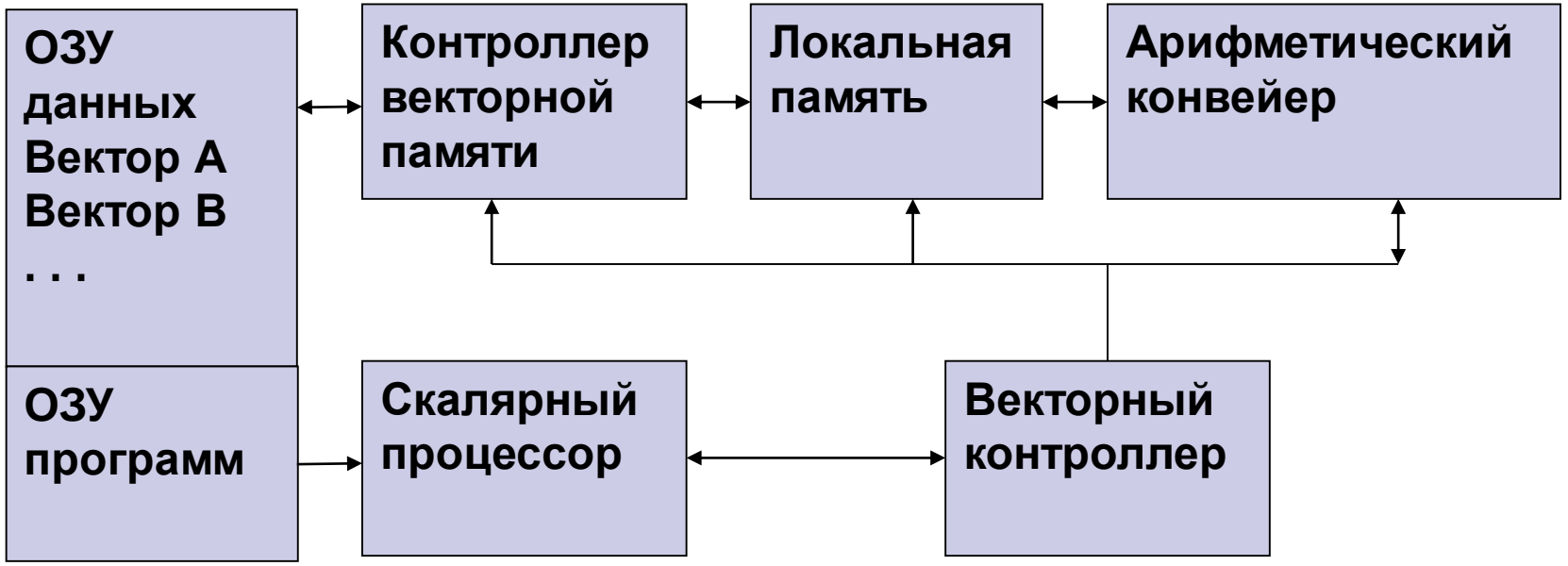
Конвейерные ВС – Стратегия управления

- ***Стратегия управления*** - процедура, которая выбирает последовательность латентностей.
- ***Жадная стратегия*** - выбирает всегда минимально возможную латентность между данной и следующей инициацией без учета каких бы то ни было следующих инициаций.
- ***Оптимальная стратегия*** - обеспечивает минимальную достижимую среднюю латентность.

Конвейерные ВС – Векторно-конвейерные процессоры

- ***Вектор*** - набор данных, которые должны быть обработаны по одному алгоритму.
- ***Векторные команды*** - команды, предназначенные для организации эффективной обработки векторных данных.
- ***Векторные процессоры*** - процессоры, предназначенные для реализации эффективной обработки векторных данных.

Векторно-конвейерные процессоры - Типичная архитектура



Векторно-конвейерные процессоры - Cray - 1

Компания Cray Research в 1976г. выпускает первый векторно-конвейерный компьютер CRAY-1:

- время такта 12.5нс,
- 12 конвейерных функциональных устройств
- пиковая производительность 160 миллионов операций в секунду,
- оперативная память до 1Мслова (слово - 64 разряда),
- цикл памяти 50нс.

Развитие векторных процессоров - Параллельно-векторные процессоры (PVP)

- ***Архитектура.*** PVP-системы строятся из векторно-конвейерных процессоров, в которых предусмотрены команды однотипной обработки векторов независимых данных.
- Как правило, несколько таких процессоров (1-16) работают одновременно над общей памятью (аналогично SMP) в рамках многопроцессорных конфигураций. Несколько таких узлов могут быть объединены с помощью коммутатора (аналогично MPP).

Развитие векторных процессоров - Параллельно-векторные процессоры (PVP)

- ***Примеры.*** NEC SX-4/SX-5, линия векторно-конвейерных компьютеров CRAY: от CRAY-1, CRAY J90/T90, CRAY SV1, CRAY X1, серия Fujitsu VPP.
- ***Модель программирования.*** Эффективное программирование подразумевает векторизацию циклов и их распараллеливание (для одновременной загрузки нескольких процессоров одним приложением).

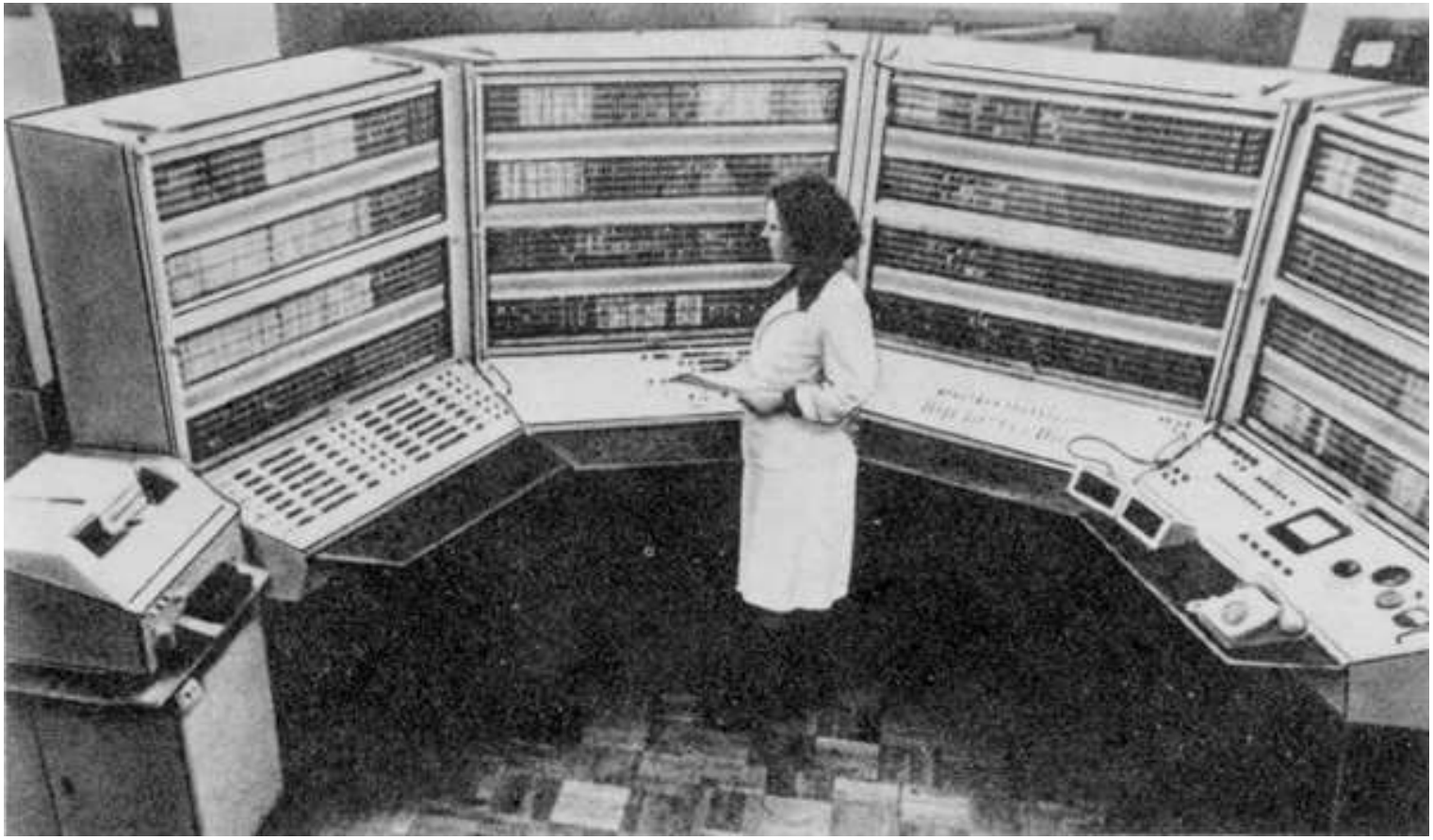
Параллельные вычислительные системы

Конвейеризация однопроцессорных ЭВМ

Конвейеризация однопроцессорных ЭВМ

- ***Конвейеризация*** - метод проектирования, в результате применения которого в вычислительной системе обеспечивается совмещение различных действий по вычислению базовых функций за счет их разбиения на подфункции.

Конвейеризация однопроцессорных ЭВМ БЭСМ-6

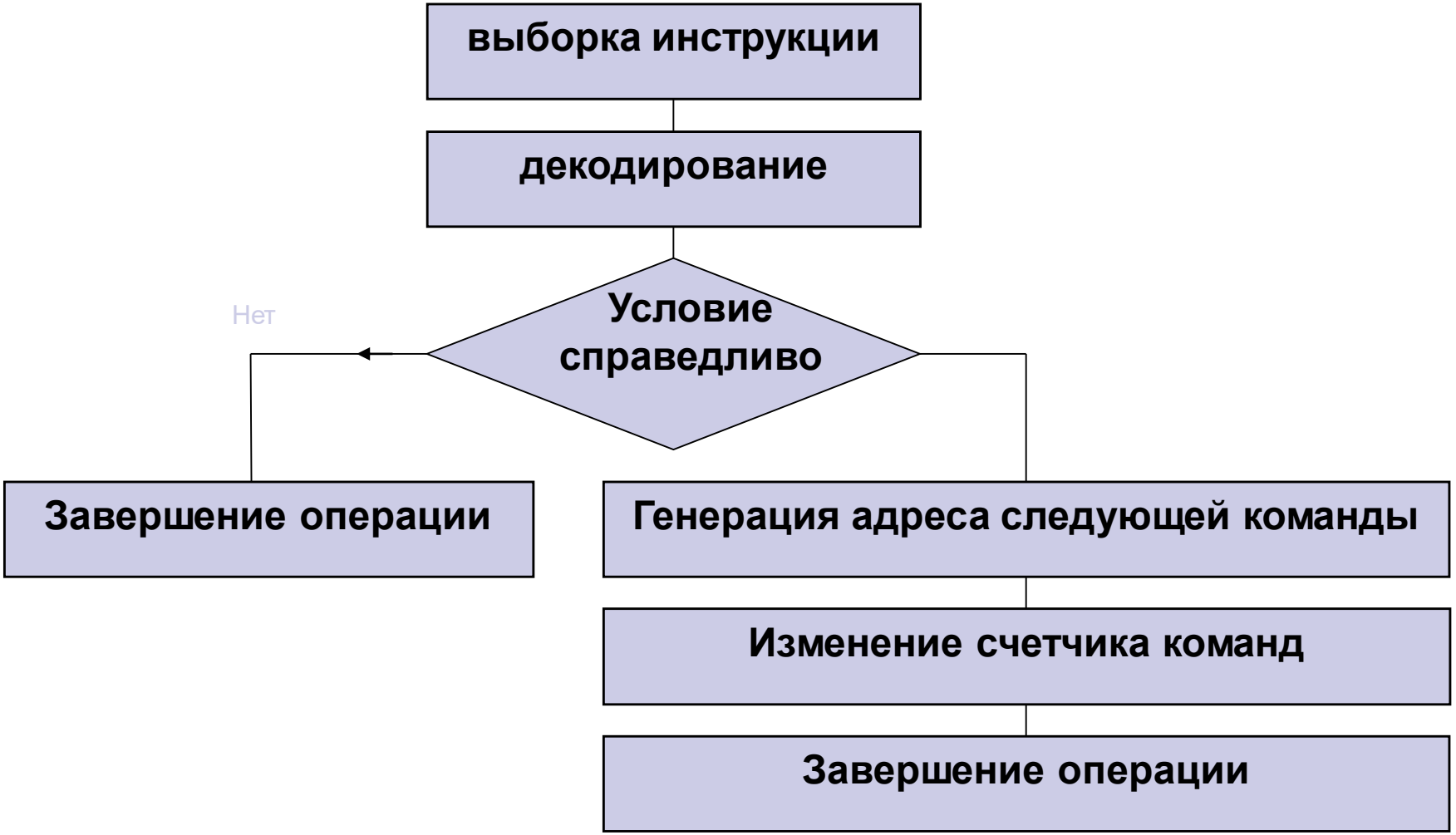


Конвейеризация однопроцессорных ЭВМ. Первый этап – предварительная выборка

- ***Предварительная (опережающая) выборка команд*** - выборка следующей команды во время завершения текущей.
- Введение модифицированного метода предварительной выборки позволяет повысить производительность реальных ЭВМ в среднем на 24% по сравнению с неконвейеризованными ЭВМ.

Конвейеризация однопроцессорных ЭВМ. Второй этап – конвейеризация ЦП.

Конвейеризация однопроцессорных ЭВМ. Второй этап – конвейеризация ЦП.



Конвейеризация однопроцессорных ЭВМ. Второй этап – конвейеризация ЦП.

При проектировании конвейера для процессора машины с архитектурой ОКОД требуются следующие данные:

- разбиения всех типов команд, включенных в систему команд процессора;
- время исполнения каждой ступенью конвейера всех типов разбиений команд в общих (часто условных) единицах времени;
- смесь команд, на которую должен ориентироваться разработчик

Конвейеризация однопроцессорных ЭВМ. Помехи.

Помеха возникает, когда к одному элементу данных (ячейке памяти, регистру, разряду слова состояния) обращаются две или более команд, которые расположены в программе настолько близко, что при выполнении происходит их перекрытие в конвейере.

Конвейеризация однопроцессорных ЭВМ. Помехи.

Три класса помех:

- чтение после записи (RAW);
- запись после чтения (WAR);
- запись после записи (WAW).

Конвейеризация однопроцессорных ЭВМ. КЭШ-память.

- Введение в систему кэш-памяти можно рассматривать, как еще один вариант конвейеризации с целью повышения быстродействия.

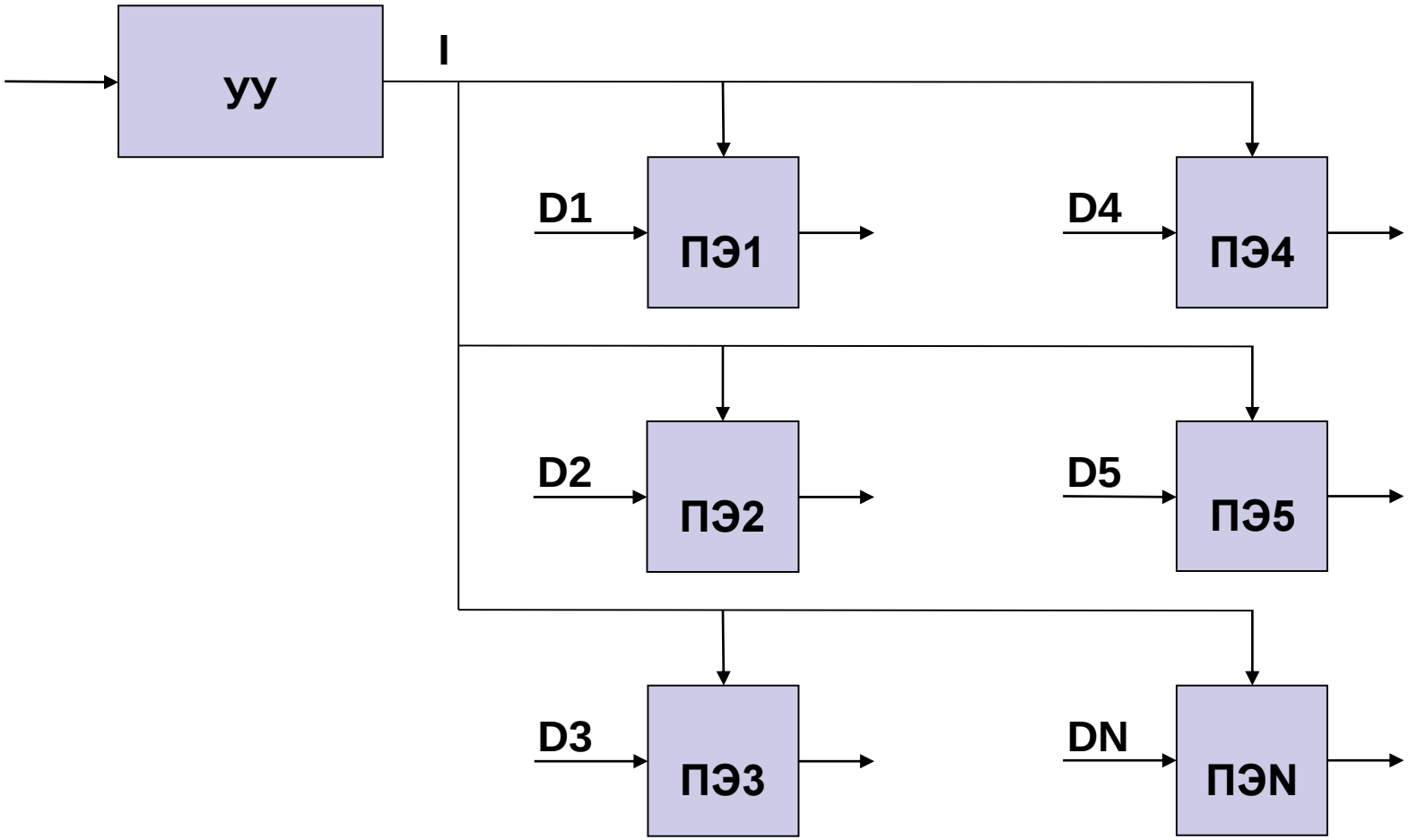
Параллельные вычислительные системы

Класс ОКМД

Параллельные ВС класса ОКМД

- ***Один поток команд – много потоков данных, ОКМД (single instruction – multiple data, SIMD)*** - в таких системах исполняется один поток команд, распределяемый между несколькими исполняющими устройствами (процессорными элементами).

Параллельные ВС класса ОКМД



ОКМД – Процессорная матрица

- ***Процессорная матрица*** - группа одинаковых процессорных элементов, объединенных единой коммутационной сетью, как правило, управляемая единым устройством управления и выполняющая единую программу.

ОКМД – Процессорная матрица ILLIAC - IV

ОКМД – Процессорная матрица ПС - 2000

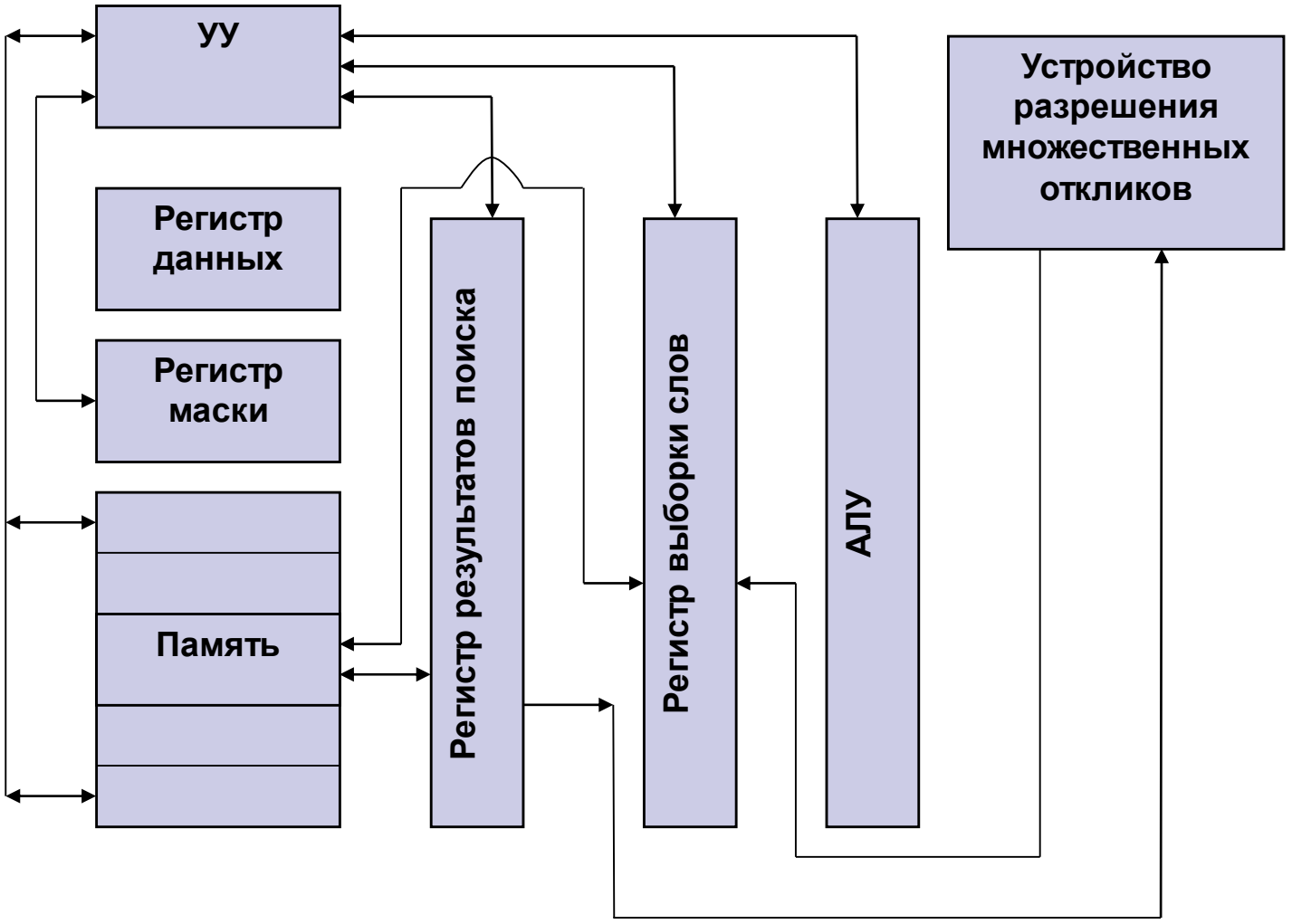
ОКМД – Однородная вычислительная среда

- ***Однородная вычислительная среда*** - регулярная решетка из однотипных процессорных элементов (ПЭ).
- Каждый ПЭ может как обладать алгоритмически полным набором операций, так и реализовывать один вид операций, жестко заданный в структуре микросхемы на этапе проектирования, а также операциями обмена или взаимодействия с другими ПЭ.

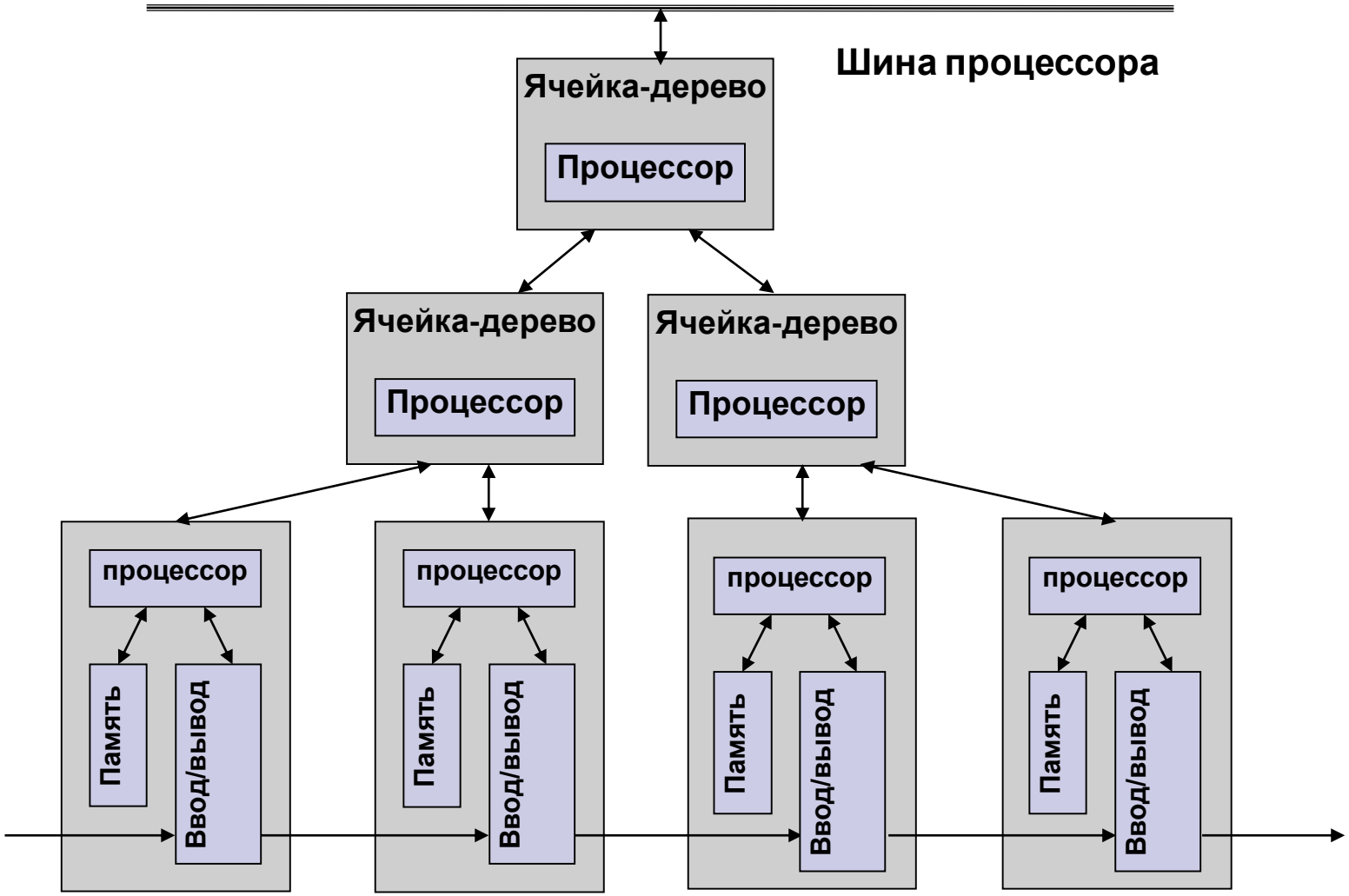
ОКМД – Однородная вычислительная среда

- ***Систолическая матрица*** - реализация однородной вычислительной среды на СБИС.
- Систолическая матрица представляет собой регулярный массив процессорных элементов, выполняющих на протяжении каждого такта одинаковые вычислительные операции с пересылкой результатов вычислений своим ближайшим соседям.

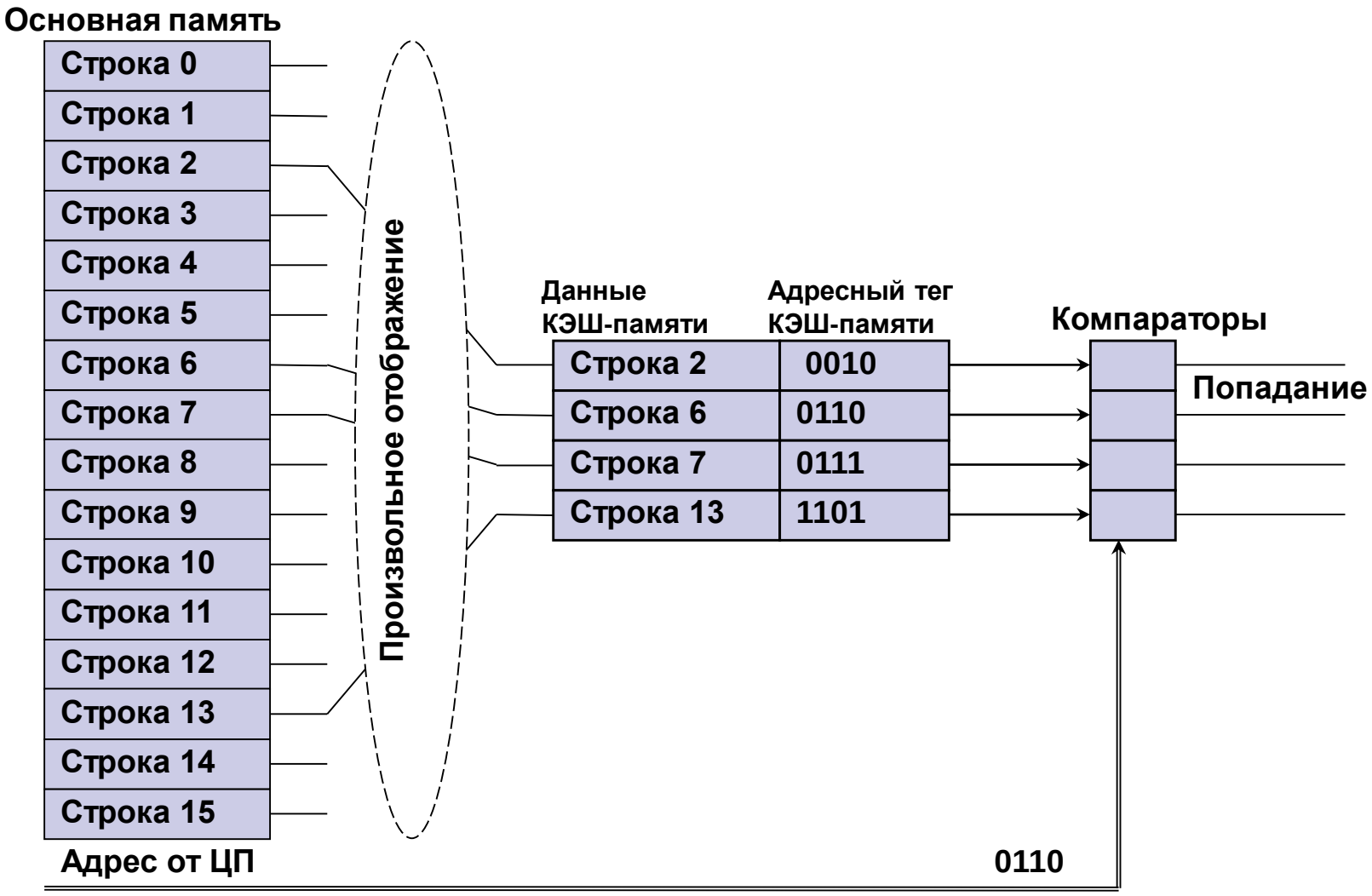
Архитектура ассоциативной ВС



Архитектура ассоциативной ВС



Полностью ассоциативная КЭШ-память



Параллельные вычислительные системы

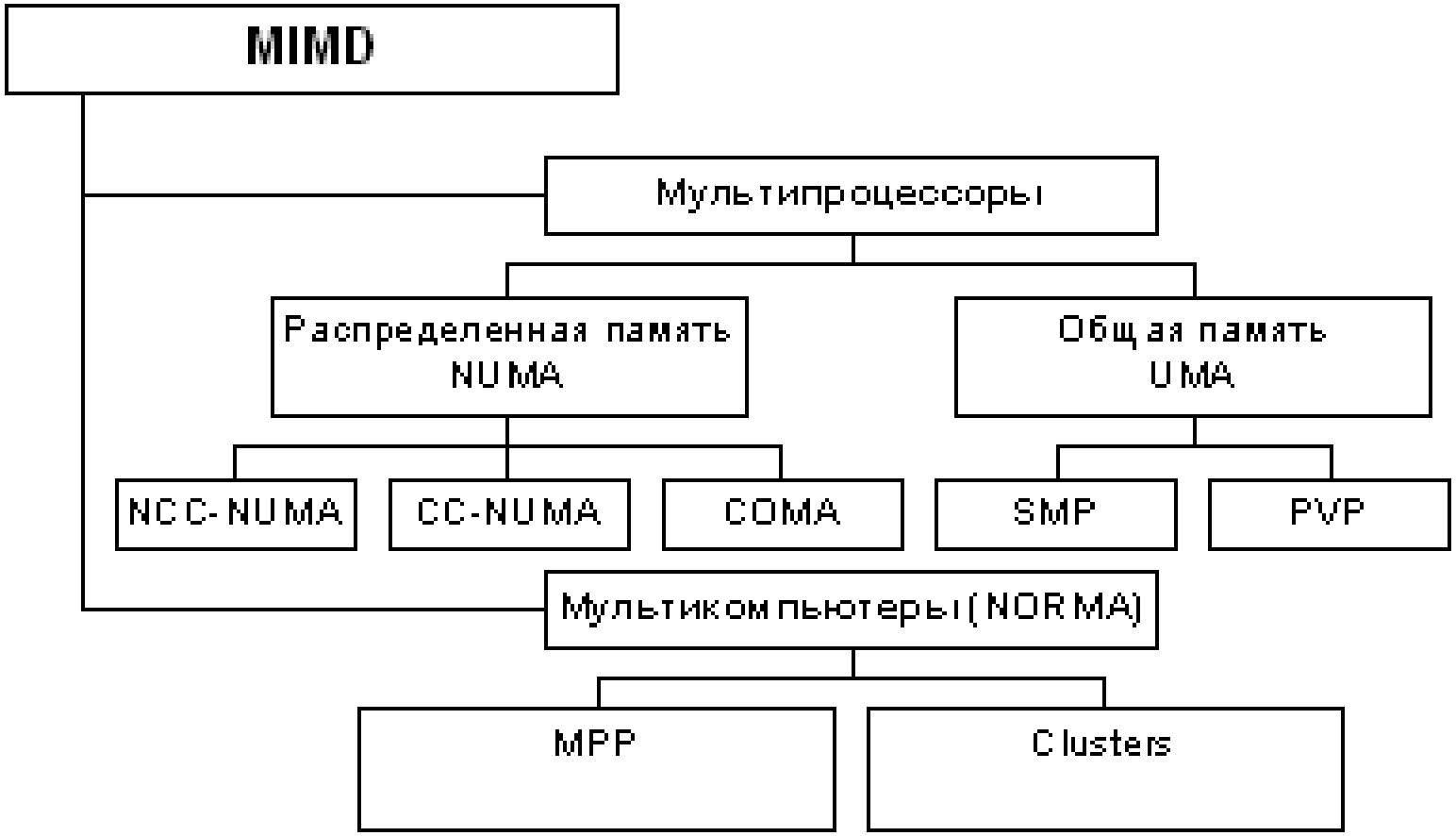
Класс МКМД (MIMD) Мультипроцессоры

Параллельные ВС класса МКМД

Один из основных недостатков систематики Флинна - излишняя широта класса МКМД.

Практически все современные высокопроизводительные вычислительные системы относятся к этому классу.

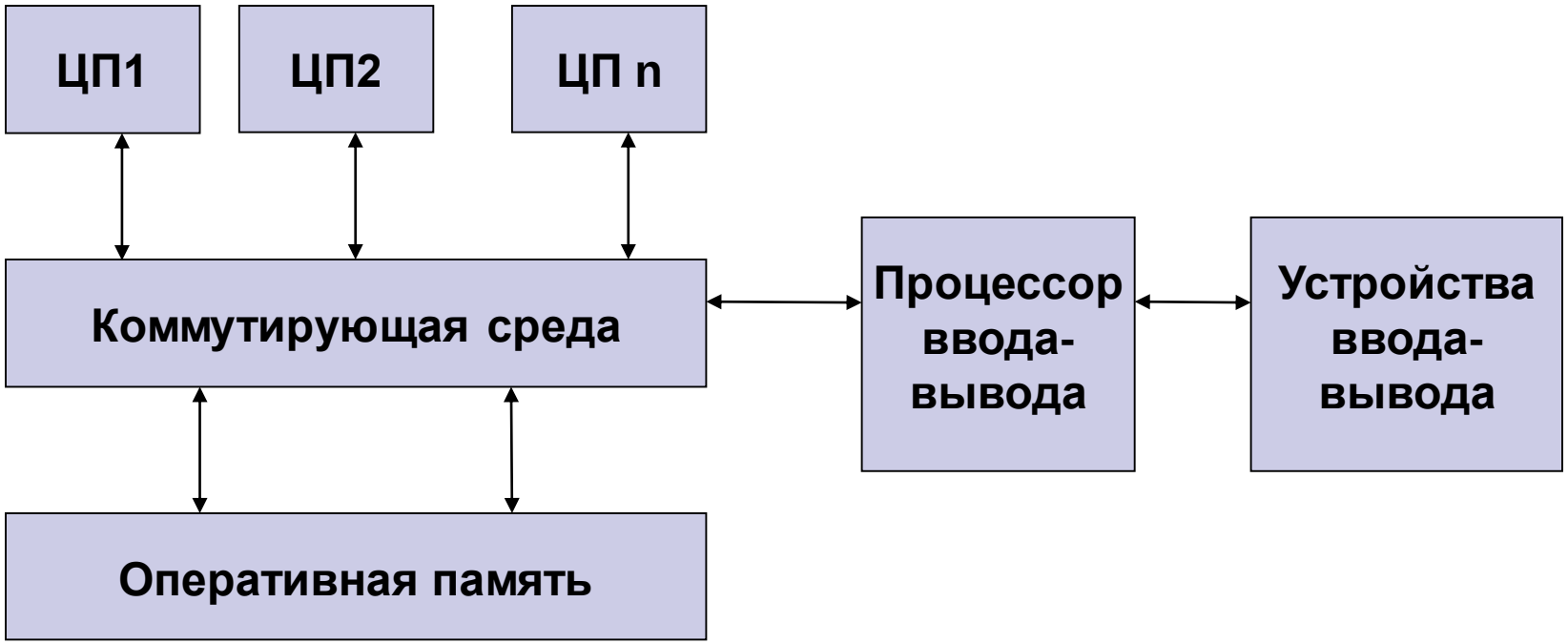
Параллельные ВС класса МКМД (MIMD)



Параллельные ВС класса МКМД
Симметричные мультипроцессоры - SMP

SMP (Symmetric MultiProcessing) – симметричная многопроцессорная архитектура. Главной особенностью систем с архитектурой SMP является наличие общей физической памяти, разделяемой всеми процессорами.

Параллельные ВС класса МКМД
Симметричные мультипроцессоры - SMP



Параллельные ВС класса МКМД Симметричные мультипроцессоры - SMP

Примеры. HP 9000 V-class, N-class; SMP-сервера и рабочие станции на базе процессоров Intel.

Масштабируемость. Наличие общей памяти упрощает взаимодействие процессоров между собой, однако накладывает сильные ограничения на их число - не более 32 в реальных системах.

Параллельные ВС класса МКМД Симметричные мультипроцессоры - SMP

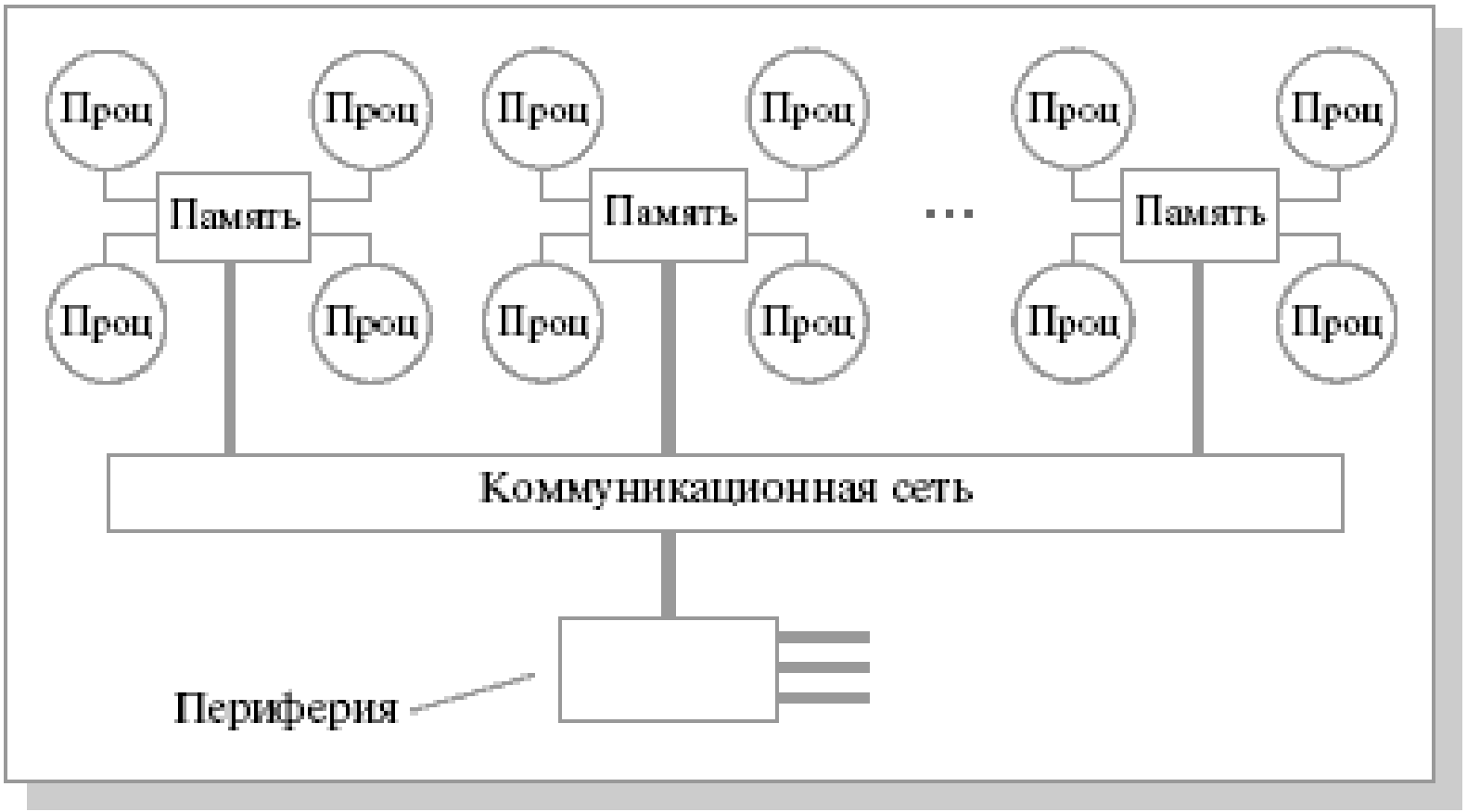
Операционная система. **С**истема работает под управлением единой ОС (обычно UNIX-подобной, но для Intel-платформ поддерживается Windows NT). ОС автоматически распределяет процессы/нити по процессорам; но иногда возможна и явная привязка.

Модель программирования – с обменом данными через общую память (POSIX threads, OpenMP).

МКМД – Мультипроцессоры с распределенной памятью (NUMA)

- ***Cache-Only Memory Architecture, COMA*** - для представления данных используется только локальная кэш-память имеющихся процессоров.
- ***Cache-Coherent NUMA, CC-NUMA*** - обеспечивается однозначность локальных кэш-памятей разных процессоров.
- ***Non-Cache Coherent NUMA, NCC-NUMA*** - обеспечивается общий доступ к локальной памяти разных процессоров без поддержки на аппаратном уровне когерентности кэша.

Мультипроцессоры с распределенной памятью (NUMA) – схема «Бабочка»



Параллельные вычислительные системы

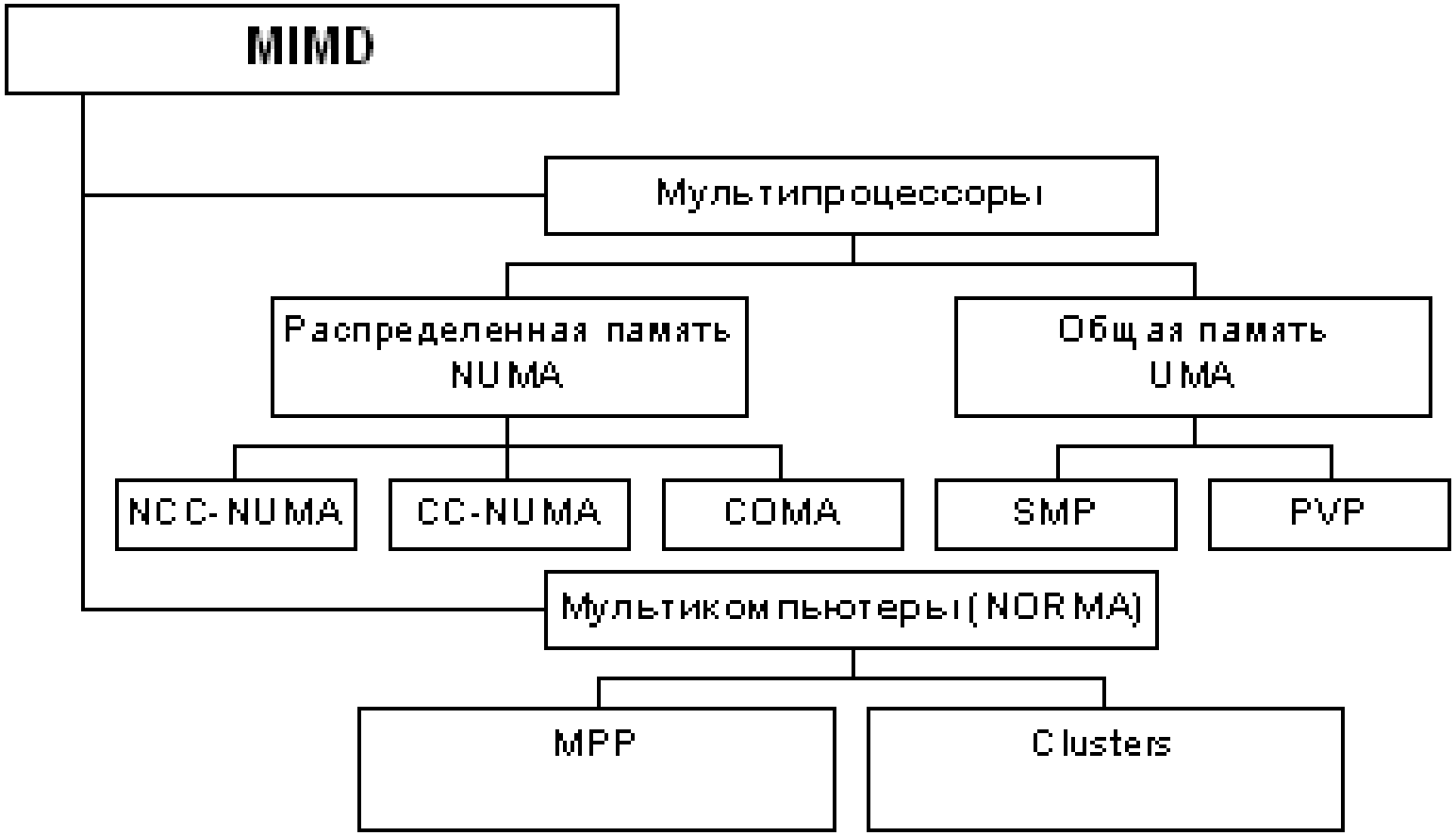
Класс МКМД (MIMD) Мультипроцессоры

Параллельные ВС класса МКМД

Один из основных недостатков систематики Флинна - излишняя широта класса МКМД.

Практически все современные высокопроизводительные вычислительные системы относятся к этому классу.

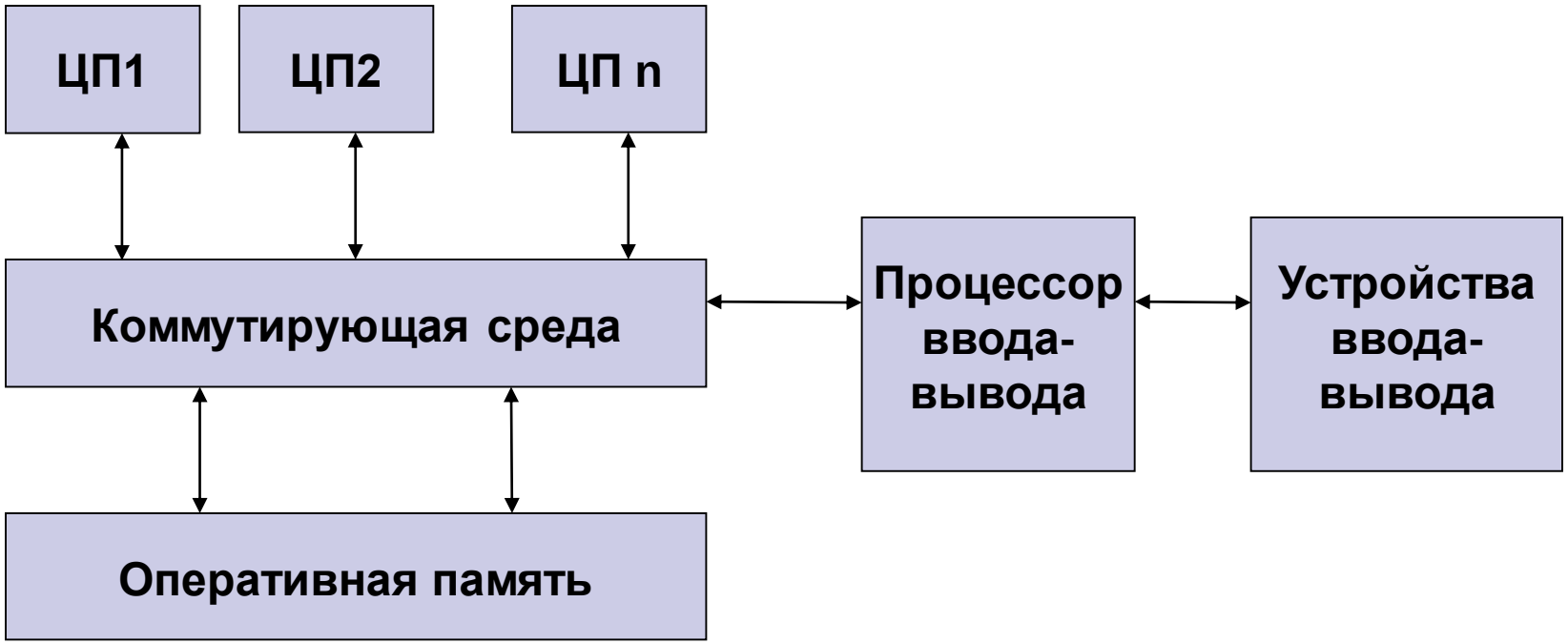
Параллельные ВС класса МКМД (MIMD)



Параллельные ВС класса МКМД
Симметричные мультипроцессоры - SMP

SMP (Symmetric MultiProcessing) – симметричная многопроцессорная архитектура. Главной особенностью систем с архитектурой SMP является наличие общей физической памяти, разделяемой всеми процессорами.

Параллельные ВС класса МКМД
Симметричные мультипроцессоры - SMP



Параллельные ВС класса МКМД Симметричные мультипроцессоры - SMP

Примеры. HP 9000 V-class, N-class; SMP-сервера и рабочие станции на базе процессоров Intel.

Масштабируемость. Наличие общей памяти упрощает взаимодействие процессоров между собой, однако накладывает сильные ограничения на их число - не более 32 в реальных системах.

Параллельные ВС класса МКМД Симметричные мультипроцессоры - SMP

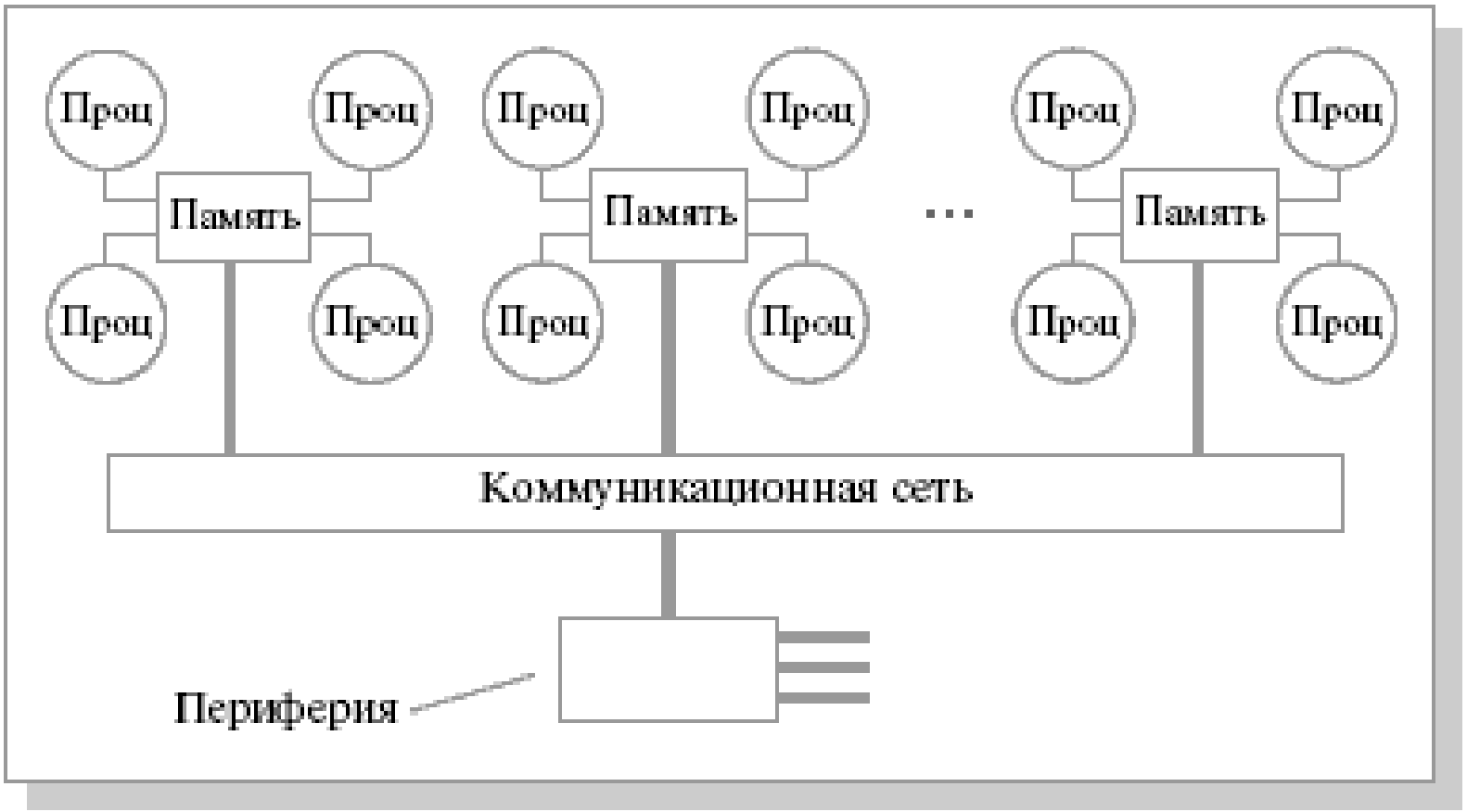
Операционная система. **С**истема работает под управлением единой ОС (обычно UNIX-подобной, но для Intel-платформ поддерживается Windows). ОС автоматически распределяет процессы/нити по процессорам; но иногда возможна и явная привязка.

Модель программирования – с обменом данными через общую память (POSIX threads, OpenMP).

МКМД – Мультипроцессоры с распределенной памятью (NUMA)

- ***Cache-Only Memory Architecture, COMA*** - для представления данных используется только локальная кэш-память имеющихся процессоров.
- ***Cache-Coherent NUMA, CC-NUMA*** - обеспечивается однозначность локальных кэш-памятей разных процессоров.
- ***Non-Cache Coherent NUMA, NCC-NUMA*** - обеспечивается общий доступ к локальной памяти разных процессоров без поддержки на аппаратном уровне когерентности кэша.

Мультипроцессоры с распределенной памятью (NUMA) – схема «Бабочка»



Параллельные вычислительные системы

СуперЭВМ

СуперЭВМ

Впервые термин ***суперЭВМ*** был использован в начале 60-х годов, когда группа специалистов Иллинойского университета (США) под руководством доктора Д. Слотника предложила идею реализации первой в мире параллельной вычислительной системы.

Суперкомпьютер – это …

- Компьютер с производительностью свыше 10 000 млн. теоретических операций в сек.
- Компьютер стоимостью более 2 млн. долларов.
- Штучно или мелкосерийно выпускаемая вычислительная система, производительность которой многократно превосходит производительность массово выпускаемых компьютеров.
- Вычислительная система, сводящая проблему вычислений любого объема к проблеме ввода/вывода.

Суперкомпьютеры

29-я редакция Top500 от 27.06.2007

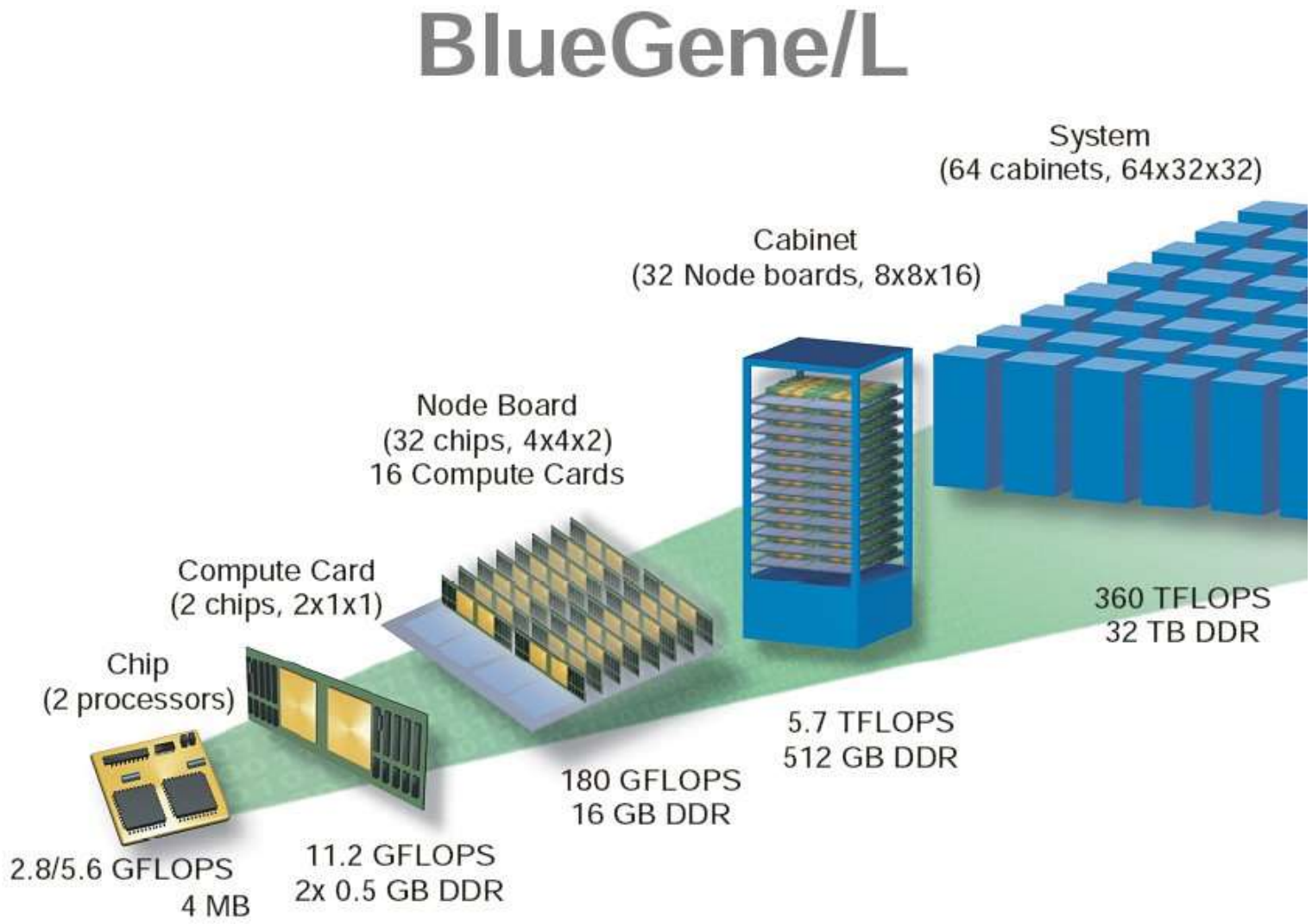
1 - прототип будущего суперкомпьютера IBM BlueGene/L с производительностью на Linpack 280.6 TFlop/s.

2 - Cray XT4/XT3, установленный в Oak Ridge National Laboratory, производительность на тесте Linpack составила 101.7 TFlop/s.

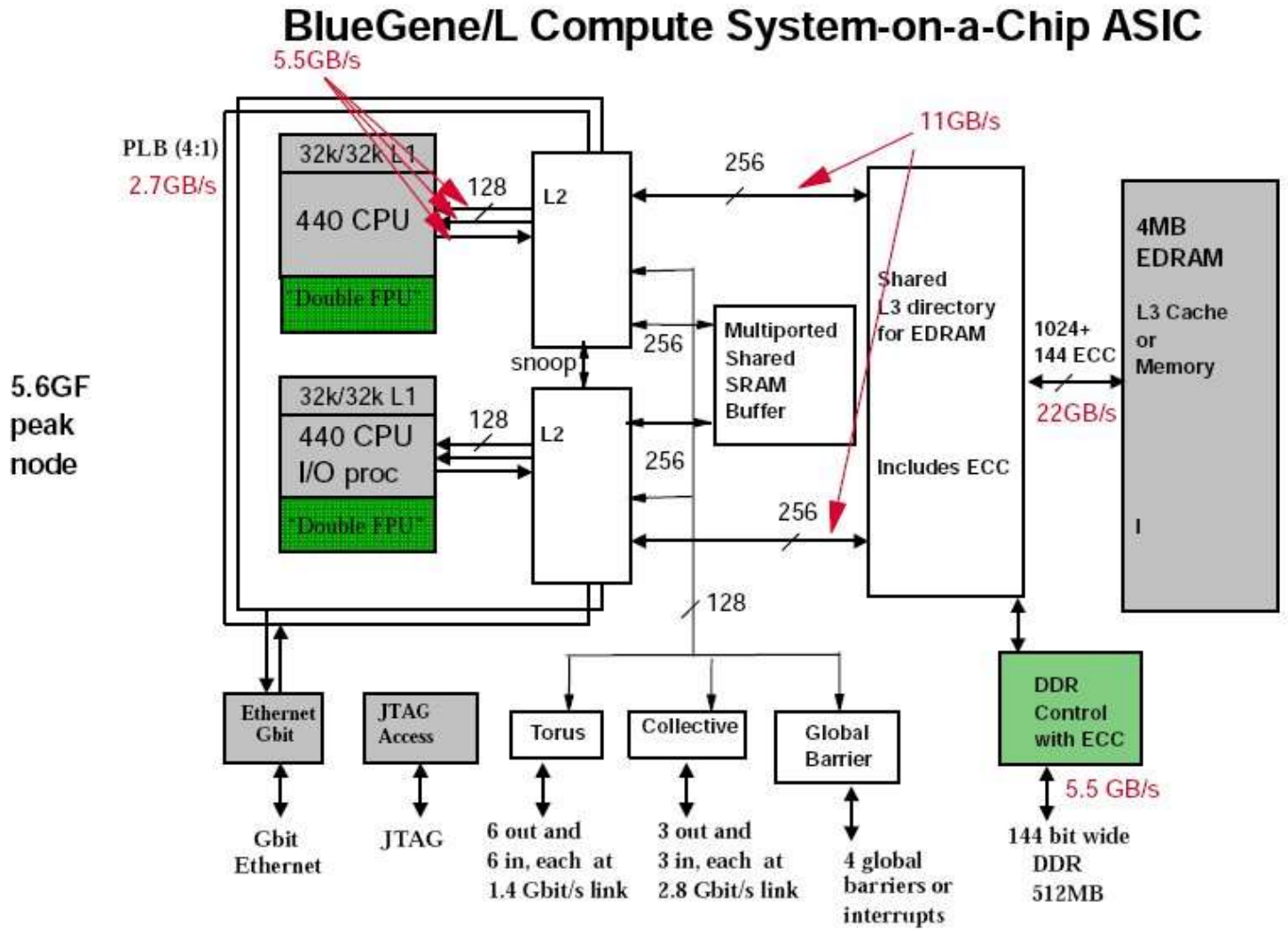
3 - Cray Red Storm с производительностью 101.4 TFlop/s на тесте Linpack.

Суперкомпьютер Blue Gene

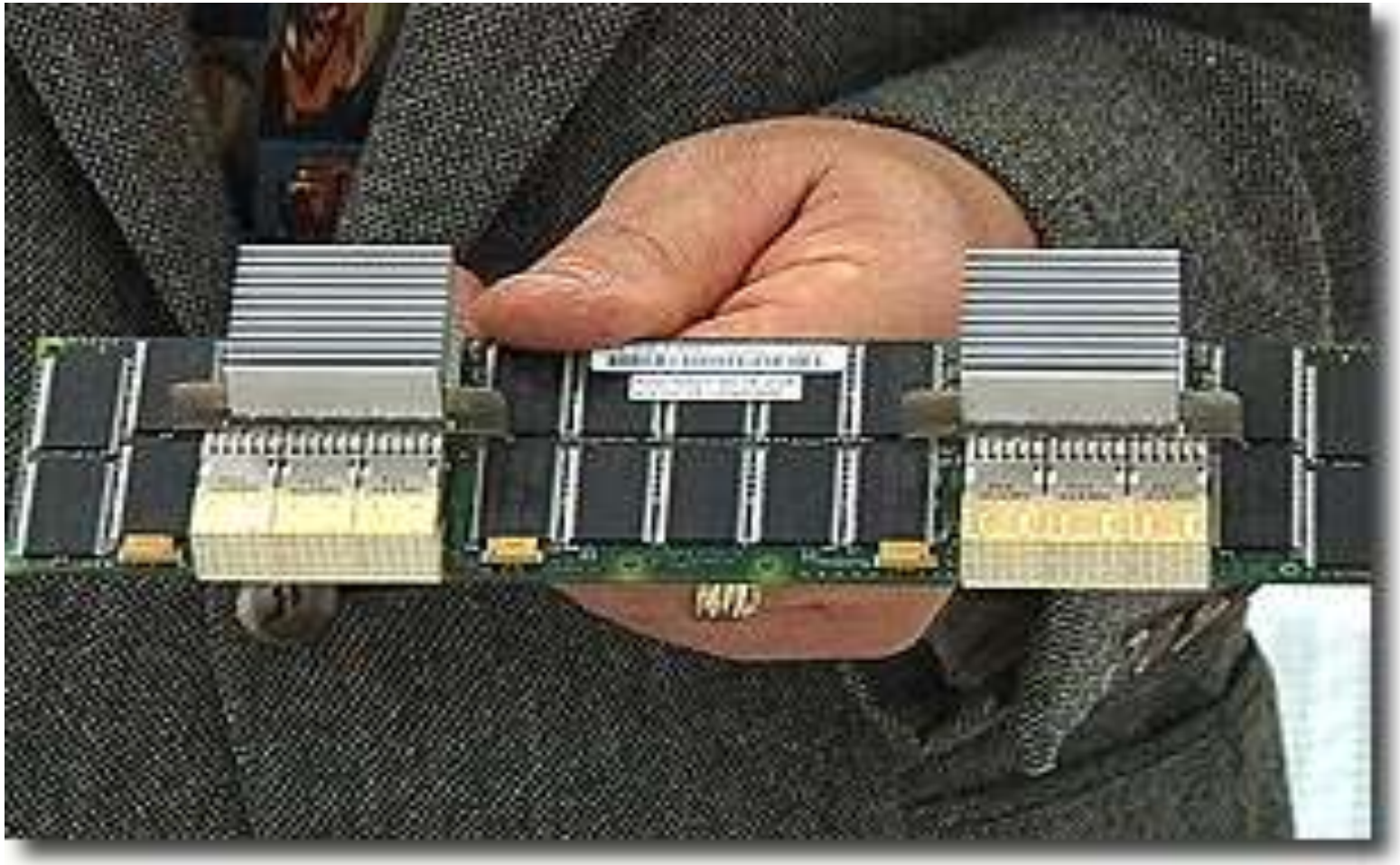
Суперкомпьютер Blue Gene
Архитектура



Суперкомпьютер Blue Gene
Архитектура



Суперкомпьютер Blue Gene
Базовый компонент (карта)



Параллельные вычислительные системы

Элементная база Микропроцессоры

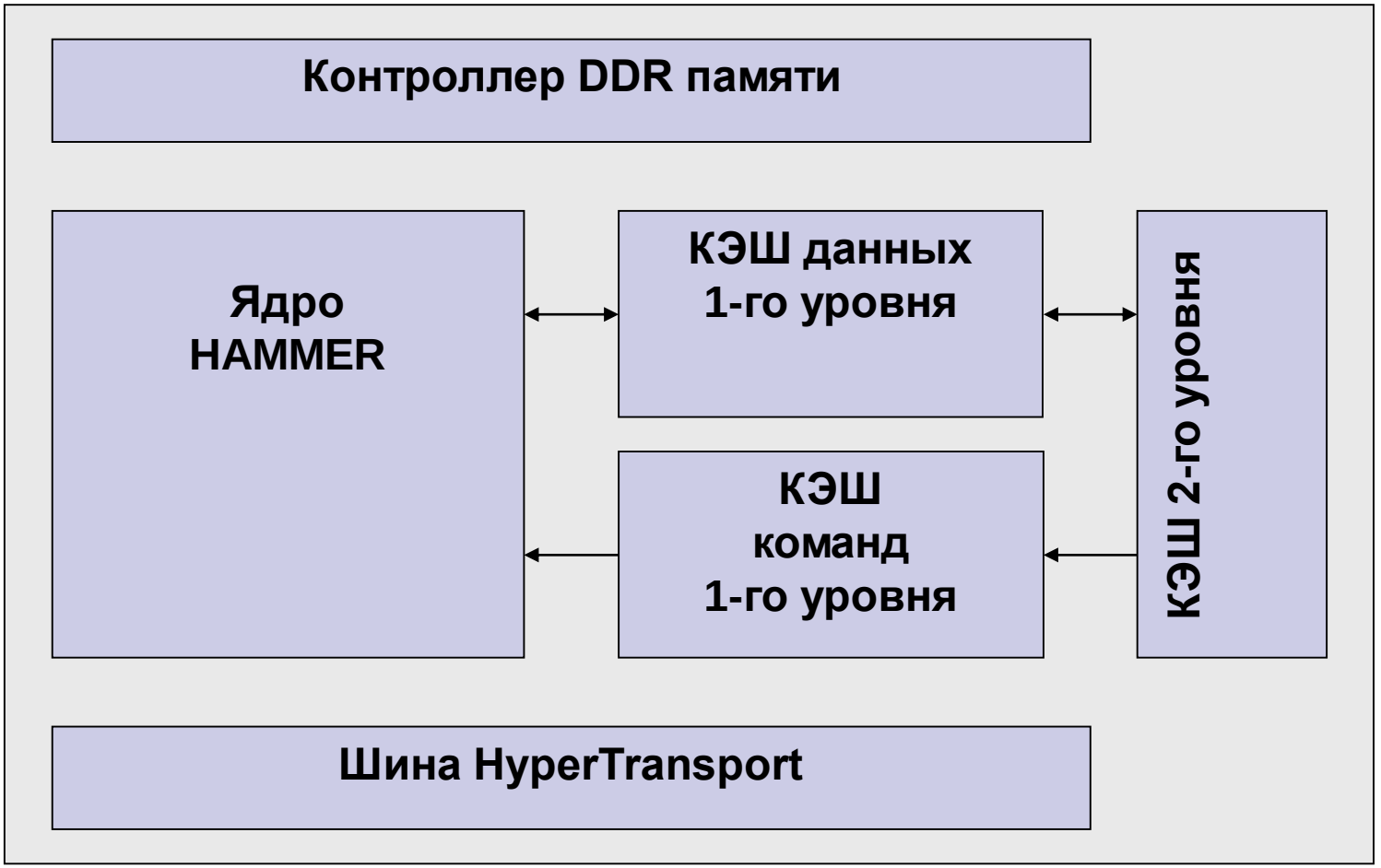
Элементная база параллельных ВС Микропроцессоры

Основные требования к микропроцессорам, используемым в параллельных ВС:

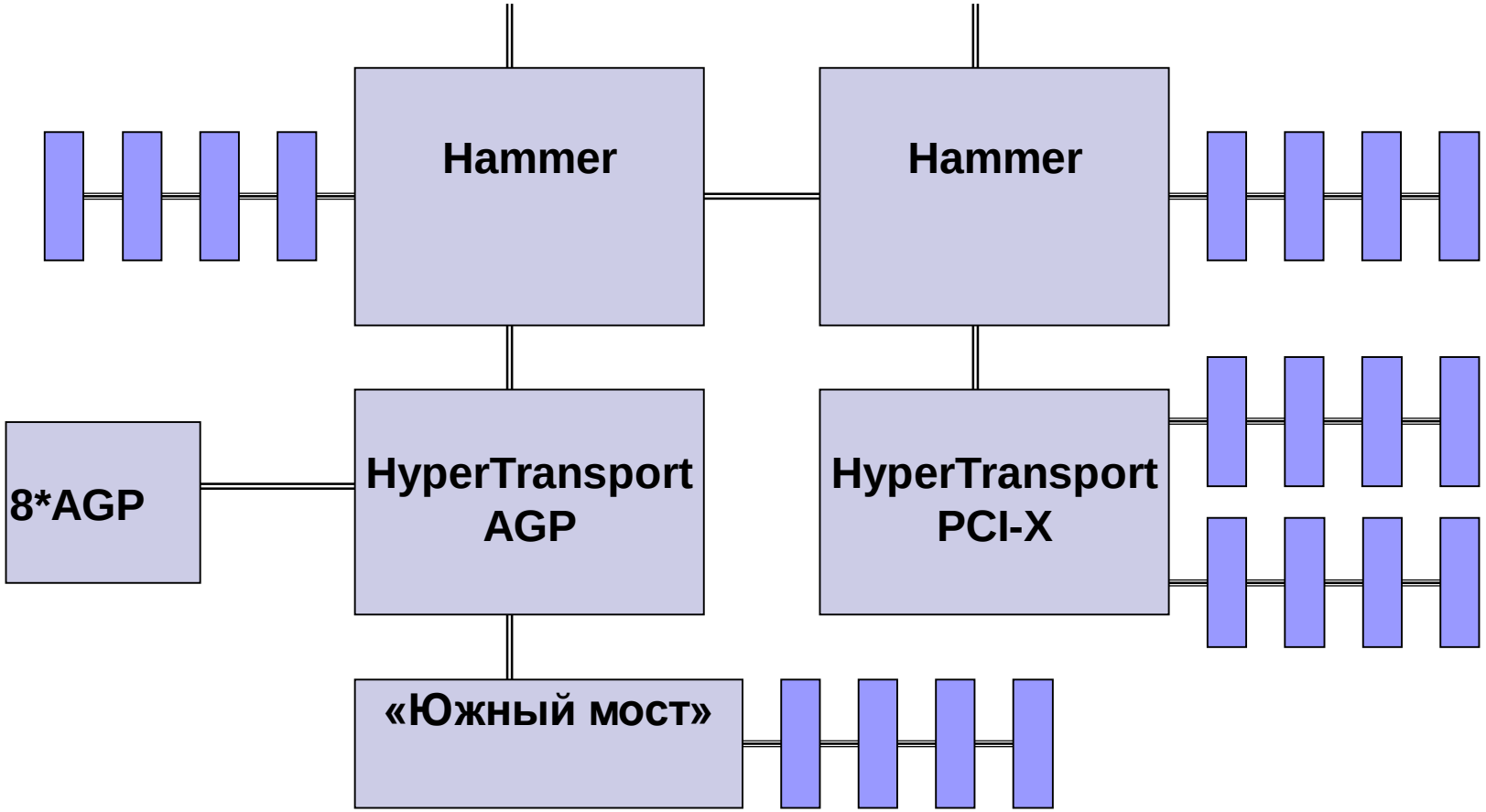
- высокая производительность
- развитые средства обмена
- низкая рассеиваемая мощность

Элементная база параллельных ВС

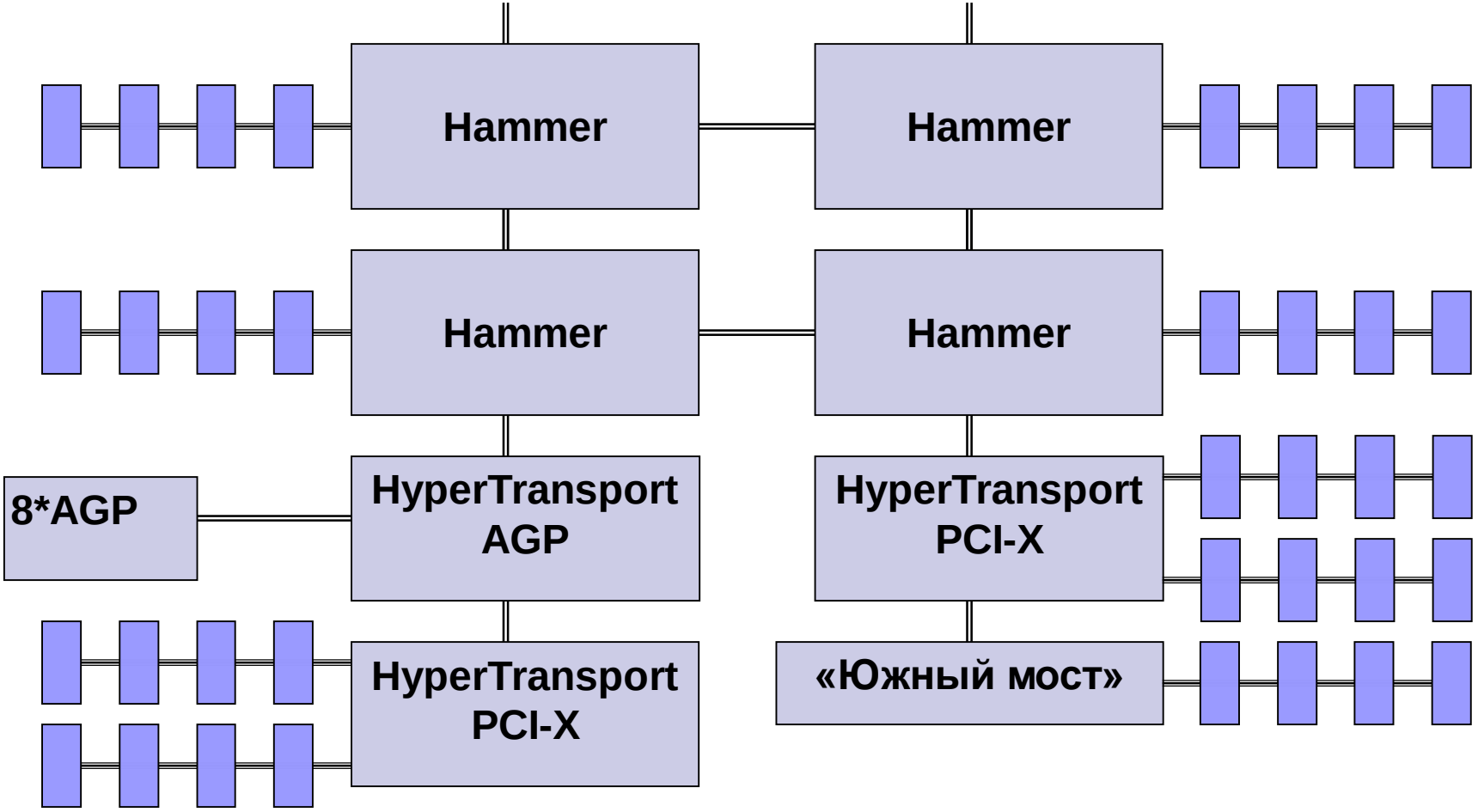
*Микропроцессор **AMD Opteron***



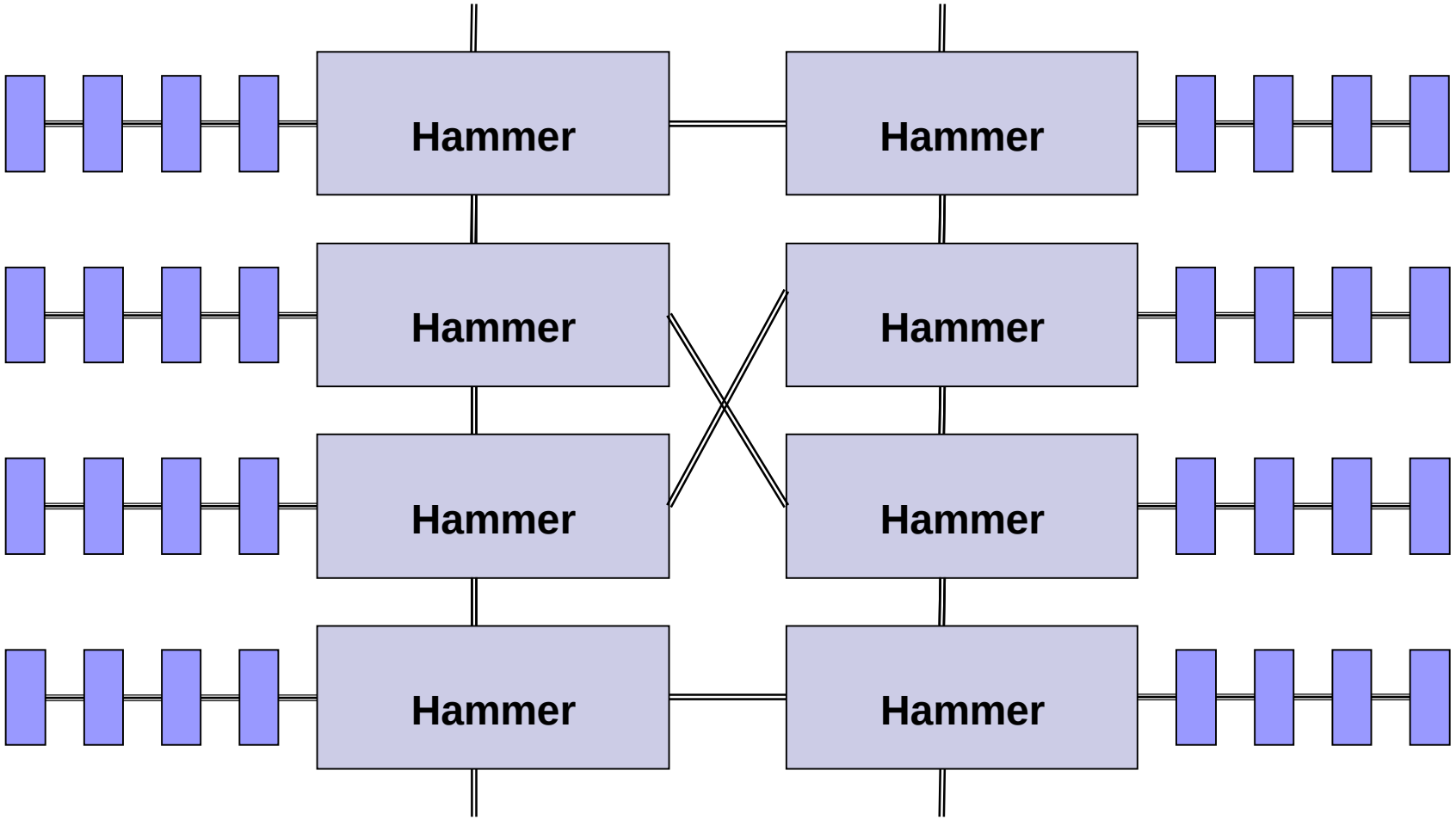
*Микропроцессор **AMD Opteron***
Варианты объединения – 2 процессора



*Микропроцессор **AMD Opteron***
Варианты объединения – 4 процессора



*Микропроцессор **AMD Opteron***
Варианты объединения – 8 процессоров

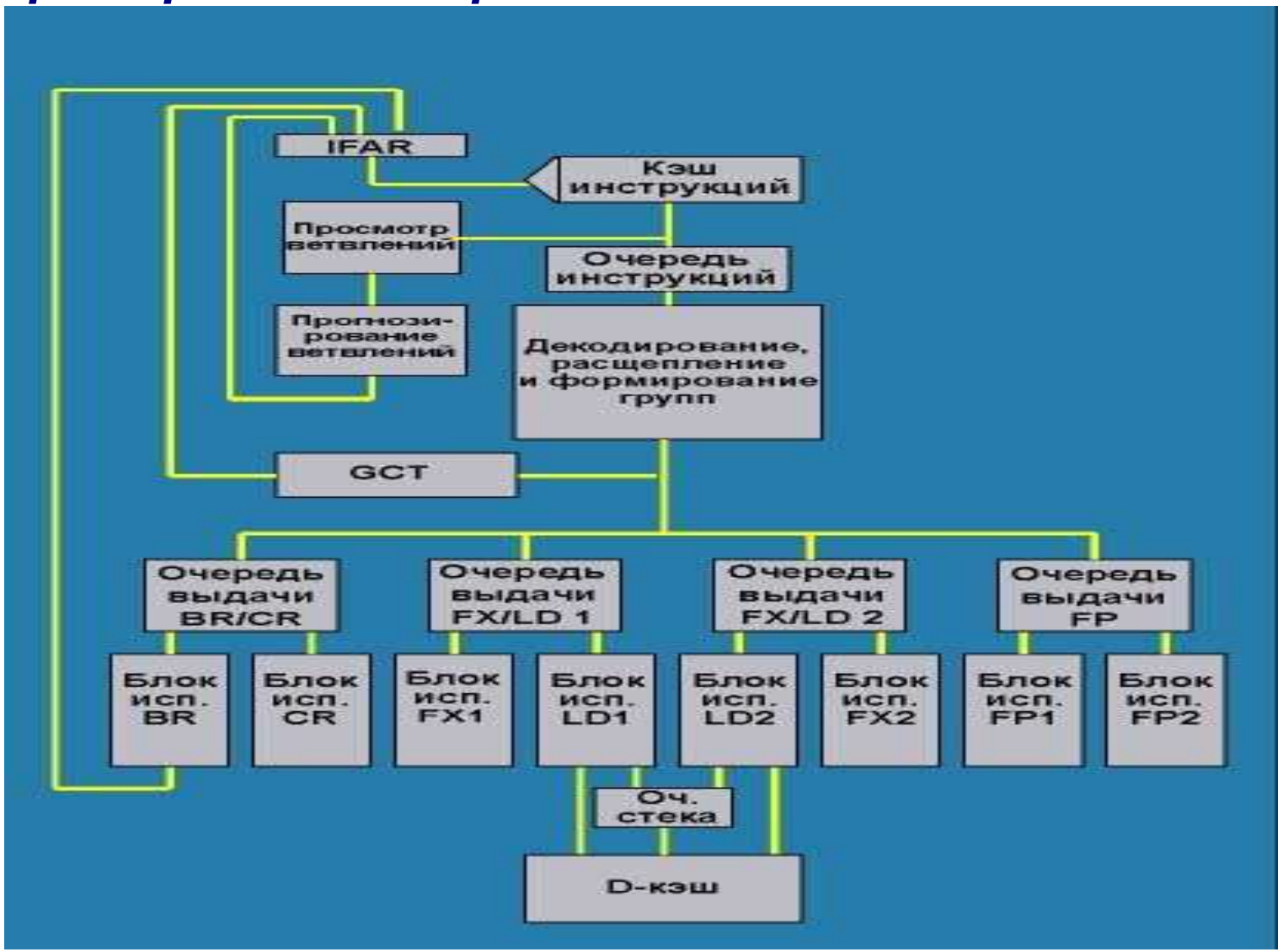


Элементная база параллельных ВС

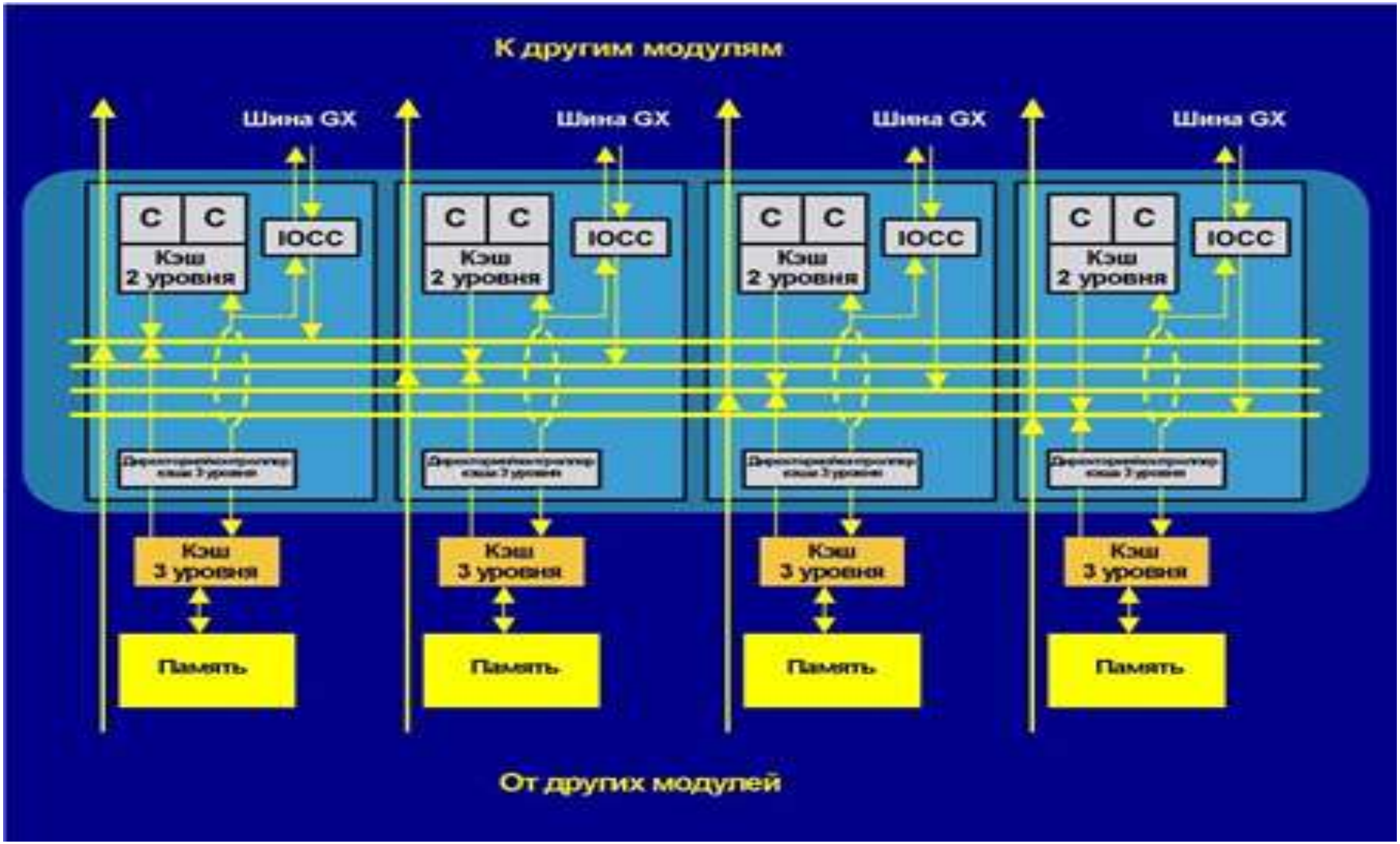
*Микропроцессор **AMD Opteron***

- 10 сентября 2007 года Компания AMD представила процессор ***Quad-Core AMD Opteron*** (ранее известный под кодовым названием ***Barcelona***), по словам производителя, «самый передовой x86-процессор из когда либо созданных и производимых, и первый настоящий четырехъядерный x86-микропроцессор»

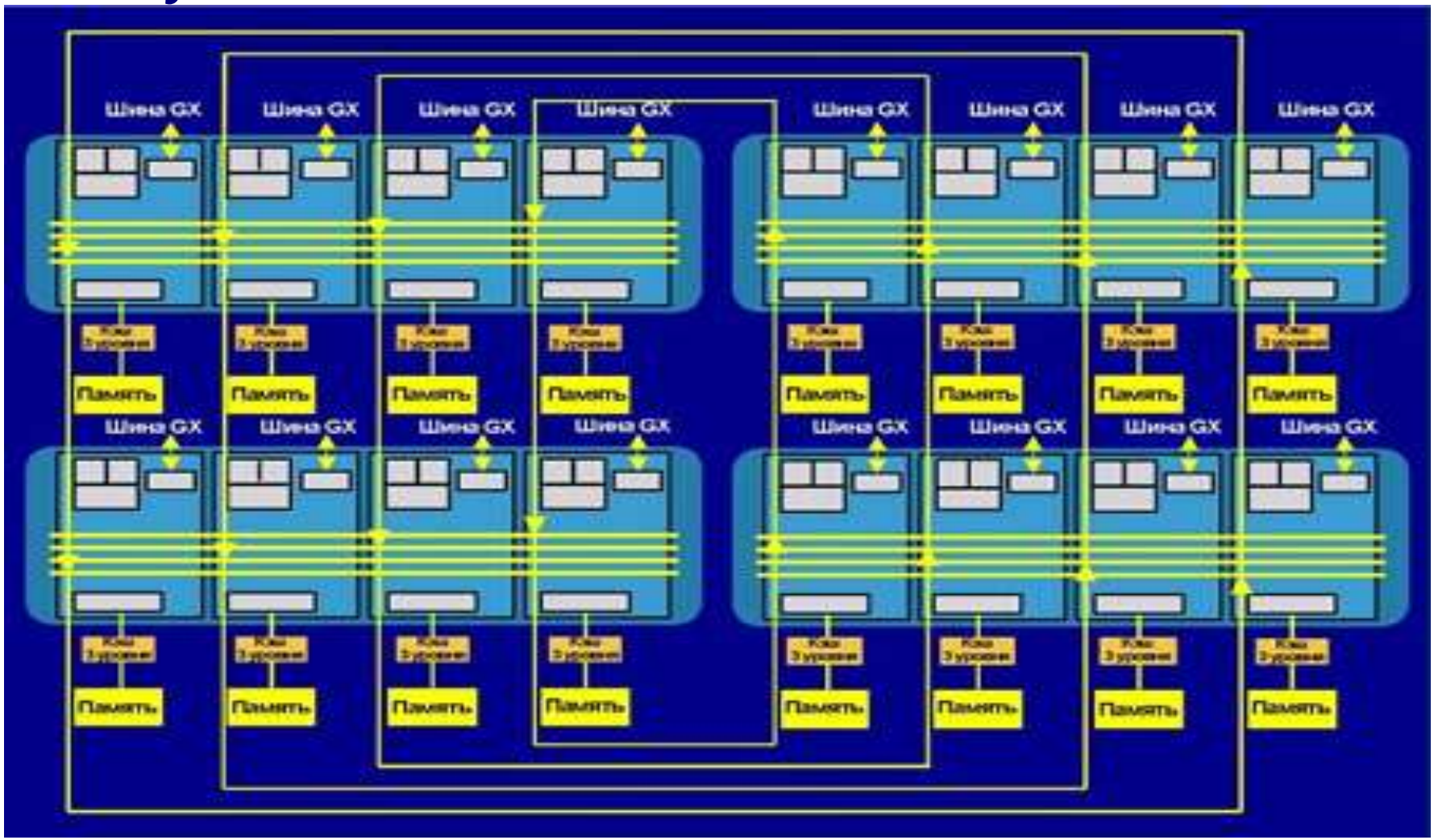
Элементная база параллельных ВС
*Микропроцессор **IBM Power4***



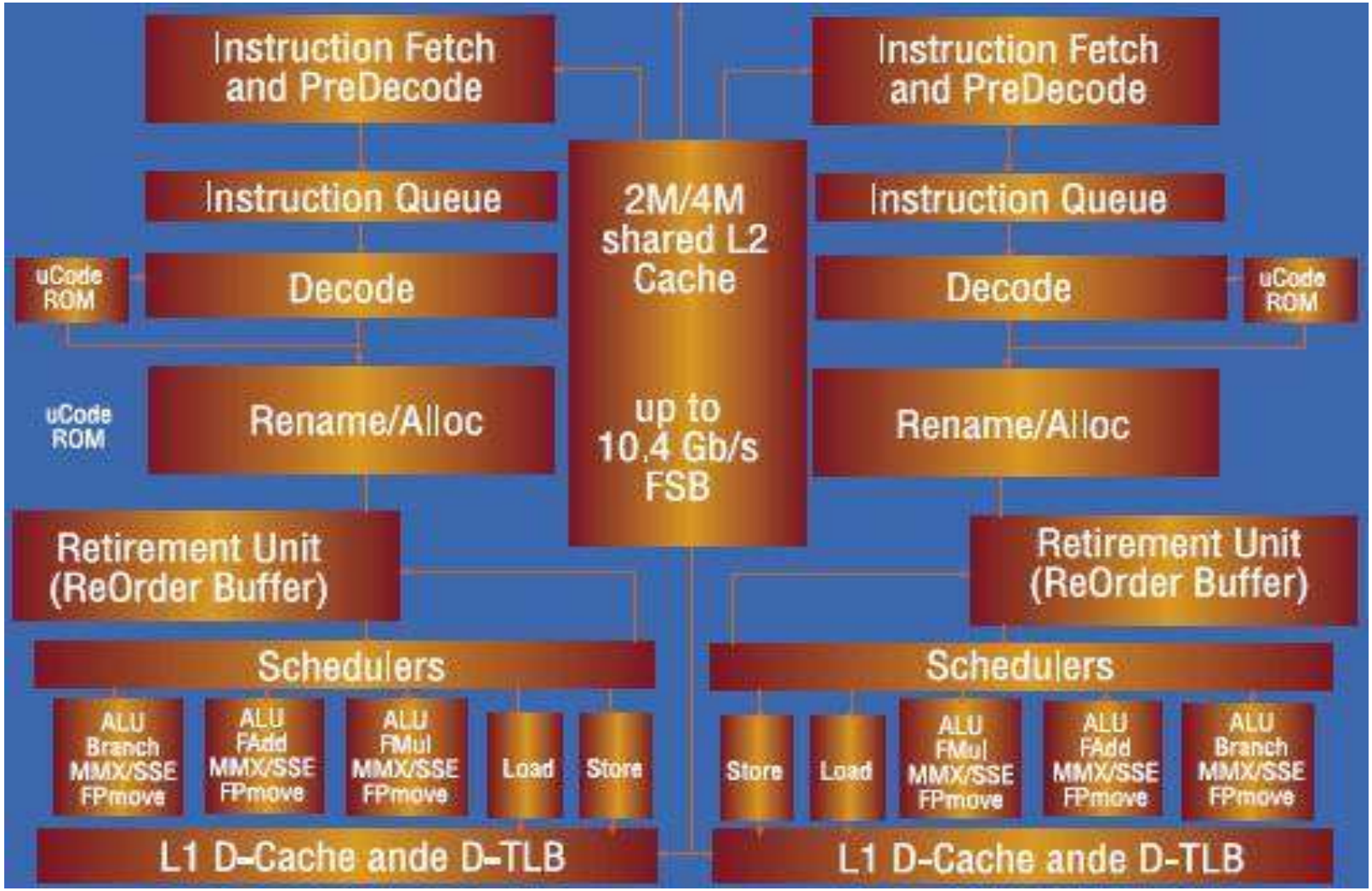
*Микропроцессор **IBM Power4***
Многокристальный модуль – 4 процессора



*Микропроцессор **IBM Power4** Объединение многокристальных модулей*



Элементная база параллельных ВС
Микропроцессор Intel Core2 Duo



Параллельные вычислительные системы

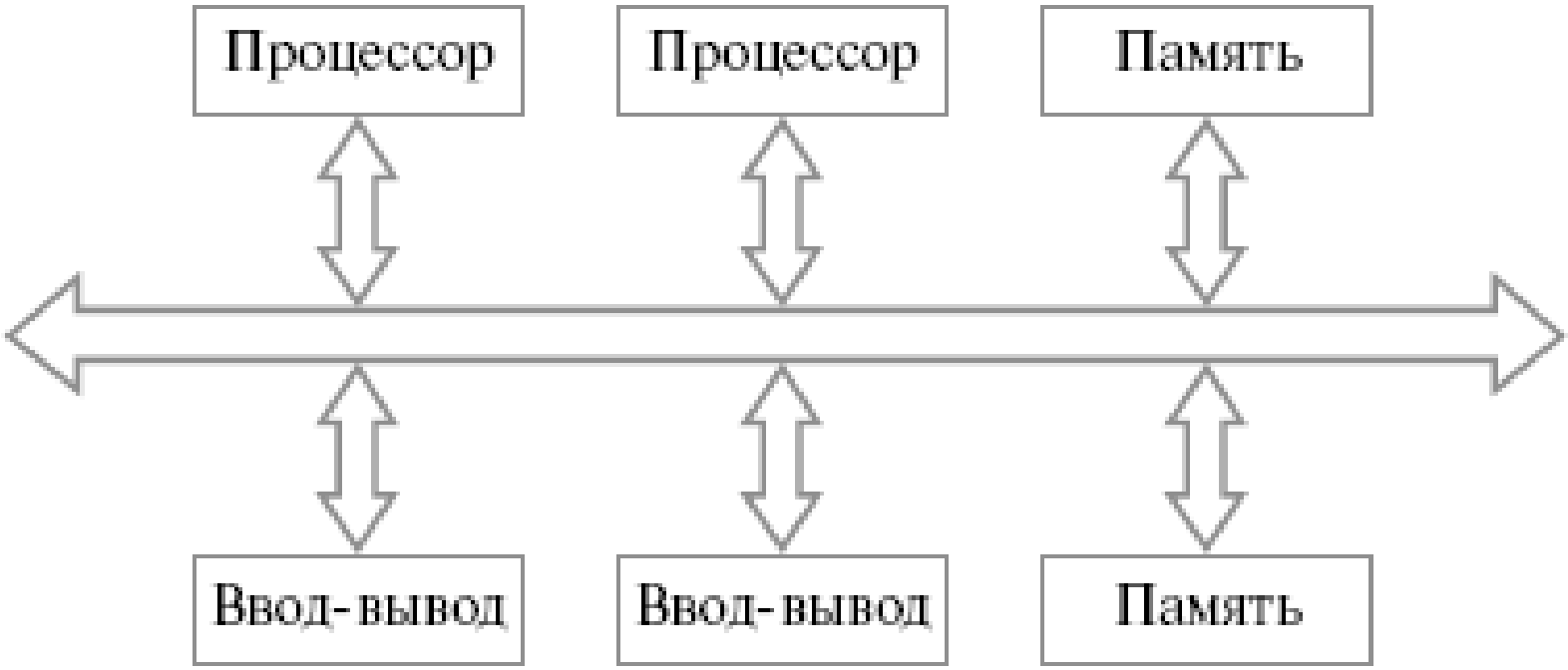
Элементная база. Коммутаторы и топология

Коммутирующие среды параллельных ВС
Простые коммутаторы

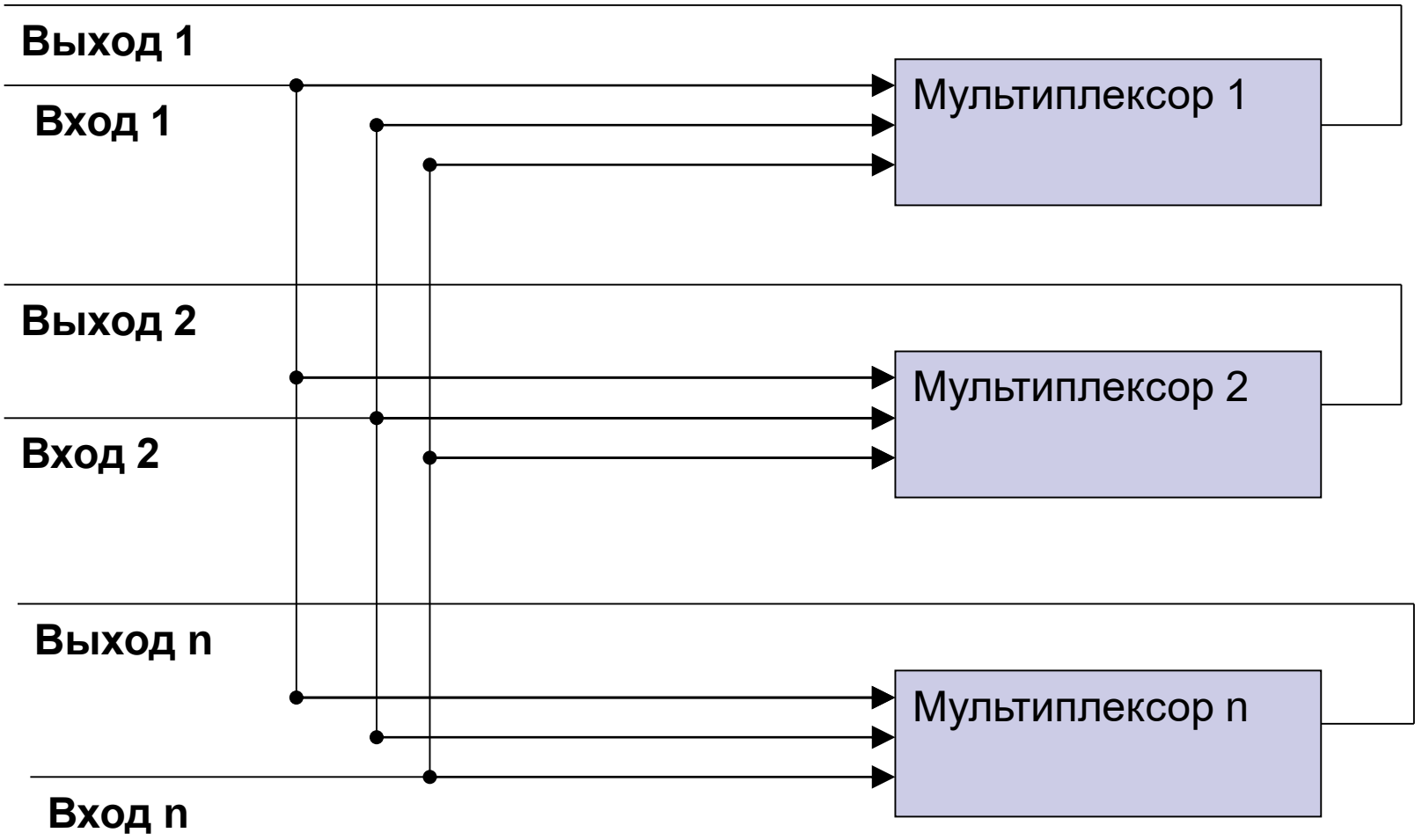
Типы простых коммутаторов:

- с временным разделением;
- с пространственным разделением.

Простые коммутаторы с временным разделением - шины

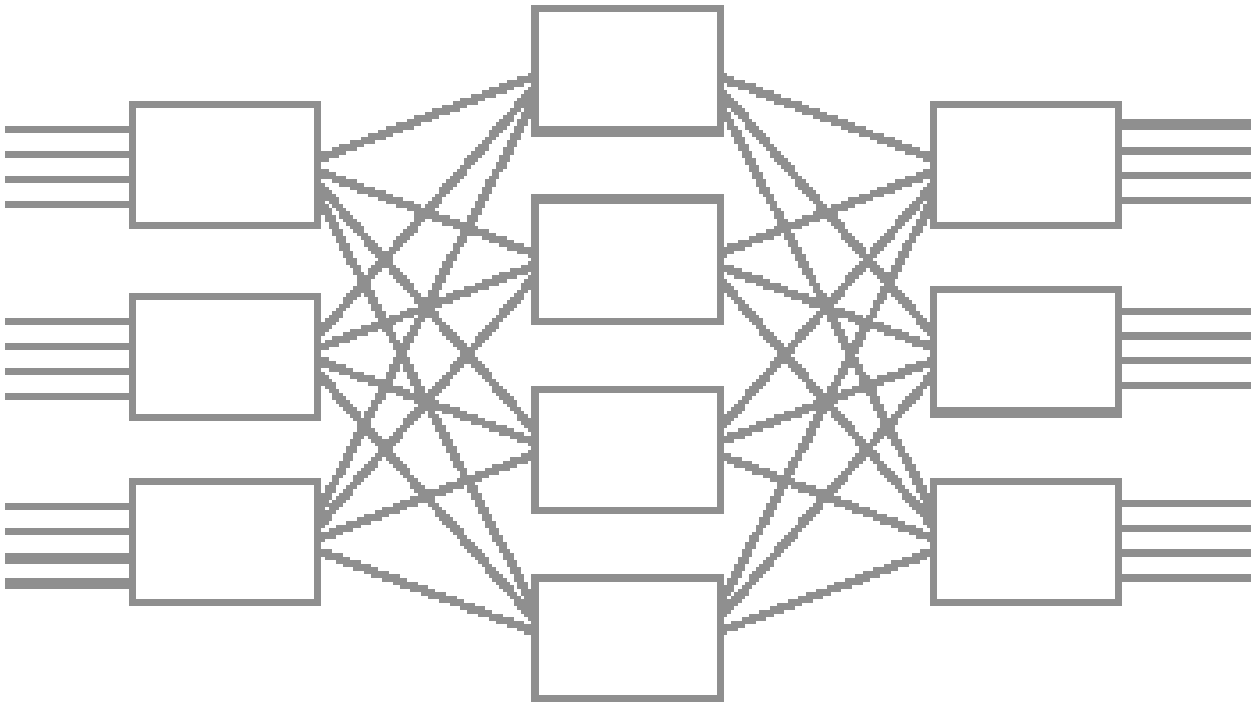


Простые коммутаторы с пространственным разделением

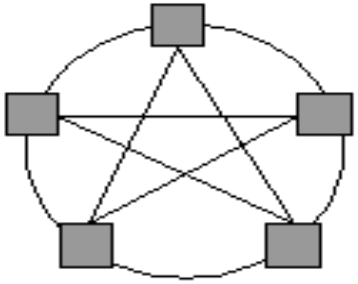


Составные коммутаторы

Коммутатор Клоза

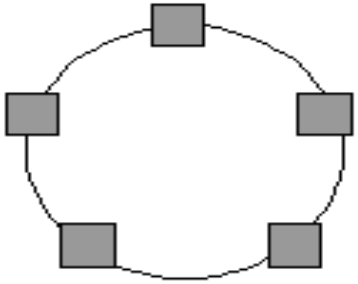


Топологии параллельной ВС

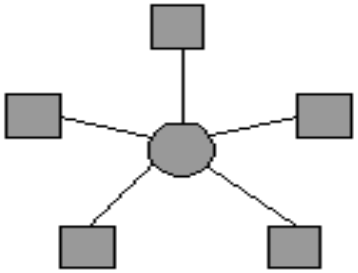


1) полный граф

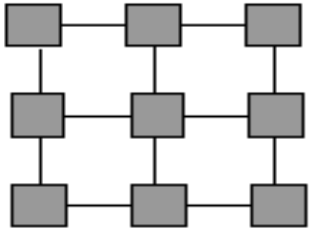
2) линейка



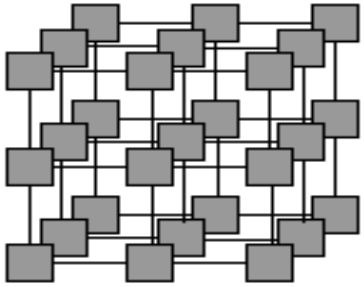
3) кольцо



4) звезда

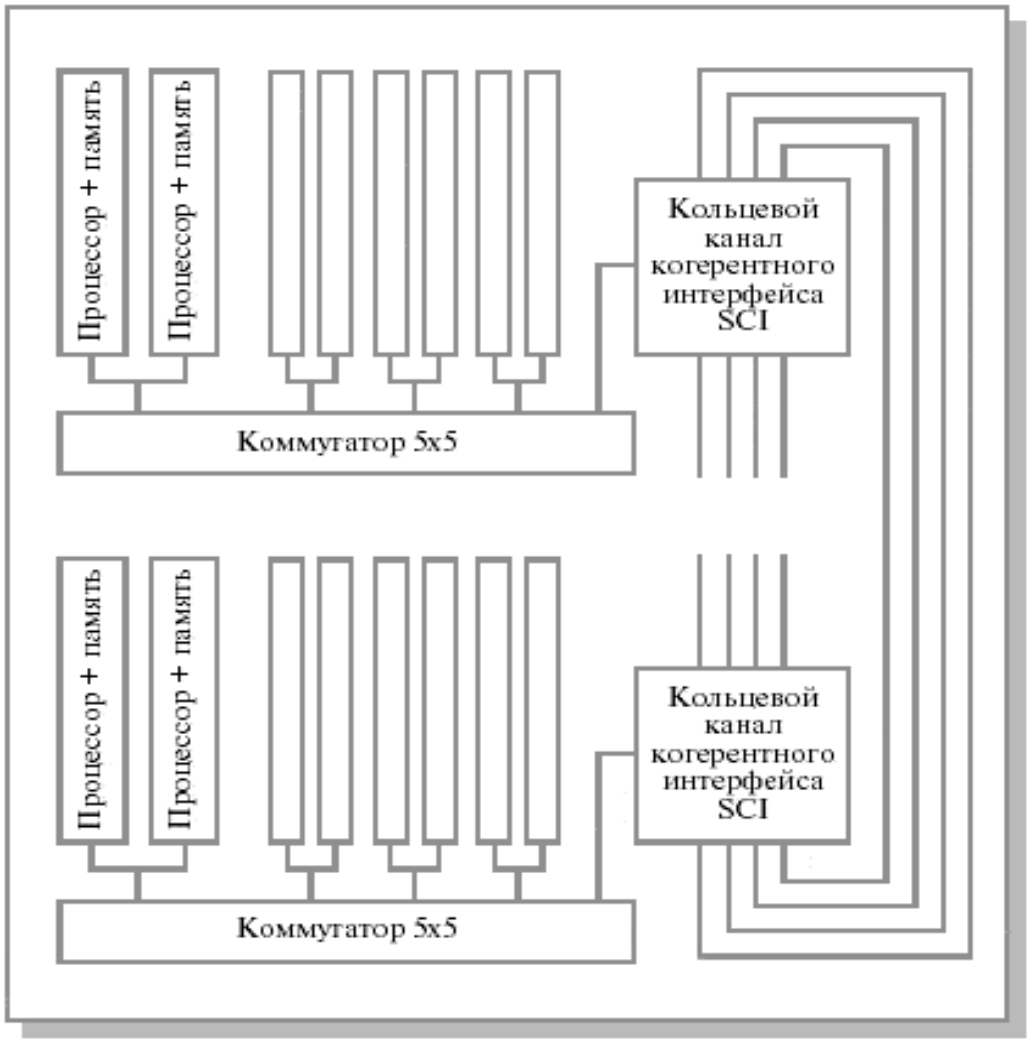


5) 2-мерная решетка

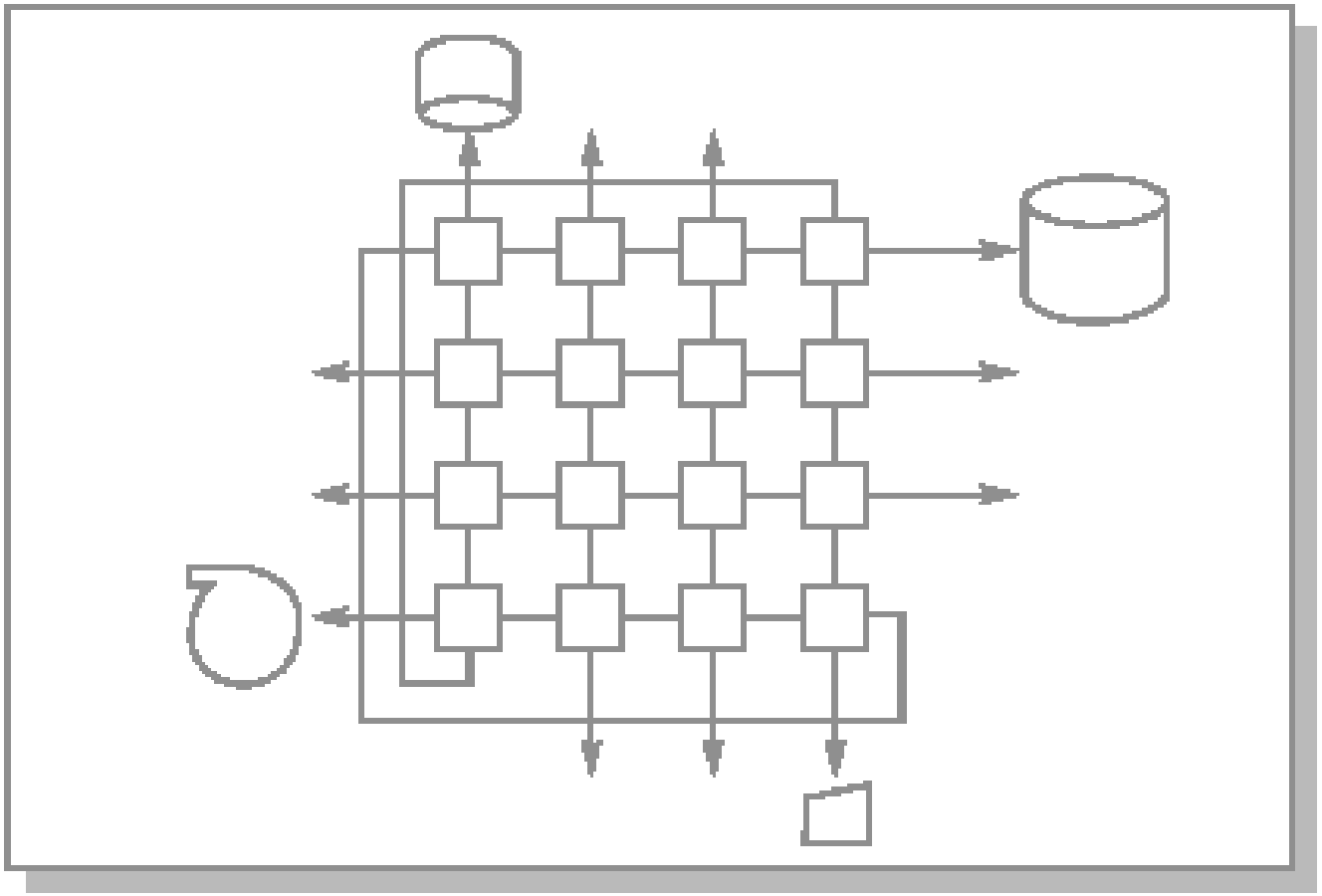


6) 3-мерная решетка

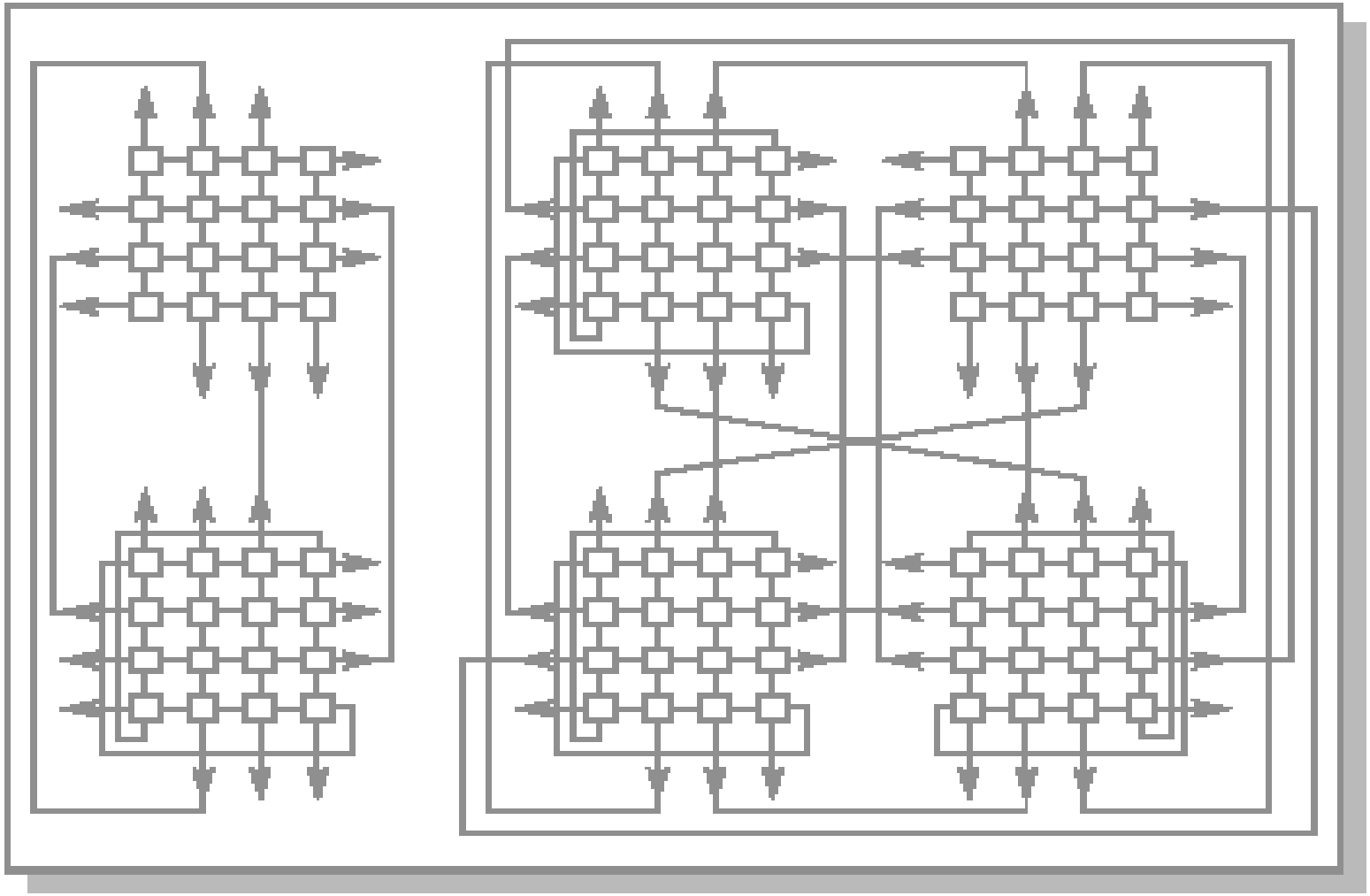
Топологии параллельных ВС
Convex Exemplar SPP1000



Топологии параллельных ВС
Модуль МВС-100



Топологии параллельных ВС - МВС-100
Варианты соединения модулей



Параллельные вычислительные системы

Элементная база. Коммутирующие среды

Коммутирующие среды параллельных ВС ***Myrinet***

Достоинства Myrinet:

- широкое распространение и высокая надежность;
- небольшое время задержки;
- хорошее соотношение цена/производительность.

Коммутирующие среды параллельных ВС ***Myrinet***

Недостатки Myrinet:

- нестандартное решение, поддерживаемое всего одним производителем;
- ограниченная пропускная способность — не более 2 Гбит/с (в ближайшее время ожидается появление варианта 10 Гбит/с);
- сложная структура кабельной проводки при максимуме 256 узлов;
- отсутствие возможности подключения к сетям хранения и глобальным сетям;
- отсутствие систем хранения с поддержкой этой технологии.

Коммутирующие среды параллельных ВС ***Infiniband***

Достоинства Infiniband:

- стандарт Infiniband Trade Assotiation (IBTA);
- несколько производителей;
- небольшое время задержки;
- пропускная способность 2, 10, 30 Гбит/с;
- поддержка приоритезации Quality of Service;
- наличие сдвоенных адаптеров 2 х 10 Гбит/с.

Коммутирующие среды параллельных ВС ***Infiniband***

Недостатки Infiniband:

- сложность изменения физической и логической структуры;
- необходимость применения дополнительного шлюза для подключения к магистральной сети или глобальной сети;
- сложная и дорогостоящая кабельная проводка;
- ограничения на дальность передачи (17 м в случае применения электропроводных кабелей);

Коммутирующие среды параллельных ВС

Ethernet

Достоинства Ethernet:

- наличие развитого инструментария для управления и отладки;
- простая и дешевая кабельная проводка;
- высокая эксплуатационная надежность;
- высокая собственная динамика при построении сетей хранения на базе IP с iSCSI;
- возможность формирования структуры из нескольких удаленных кластеров (Grid);
- низкие вычислительные затраты в случае интеграции сетевого адаптера на системную плату;

Коммутирующие среды параллельных ВС ***Ethernet***

Недостатки Ethernet:

- наличие задержки (сокращение времени задержки за счет применения TOE и RDMA должно получить свое практическое подтверждение);
- высокая стоимость 10-гигабитного интерфейса (в ближайшем будущем ожидается снижение цены).

Параллельные вычислительные системы

Технологии GRID

Параллельные ВС
GRID

Технология GRID подразумевает слаженное взаимодействие множества ресурсов, гетерогенных по своей природе и расположенных в многочисленных, возможно, географически удаленных административных доменах.

Параллельные ВС
GRID

Параллельные ВС
***GRID** – предпосылки возникновения*

- Необходимость в концентрации огромного количества данных, хранящихся в разных организациях
- Необходимость выполнения очень большого количества вычислений в рамках решения одной задачи.
- Необходимость в совместном использовании больших массивов данных территориально разрозненной рабочей группой,

Параллельные ВС
***GRID** – предпосылки возникновения*

“Вероятно, мы скоро увидим распространение “компьютерных коммунальных услуг”, которые, подобно электричеству и телефону, придут в дома и офисы по всему миру”.

Лен Клейнрок, 1969г.

Параллельные ВС
Метакомпьютинг и GRID

Метакомпьютинг - особый тип распределенного компьютинга, подразумевающего соединение суперкомпьютерных центров высокоскоростными сетями для решения одной задачи.

Параллельные ВС
*Свойства **GRID***

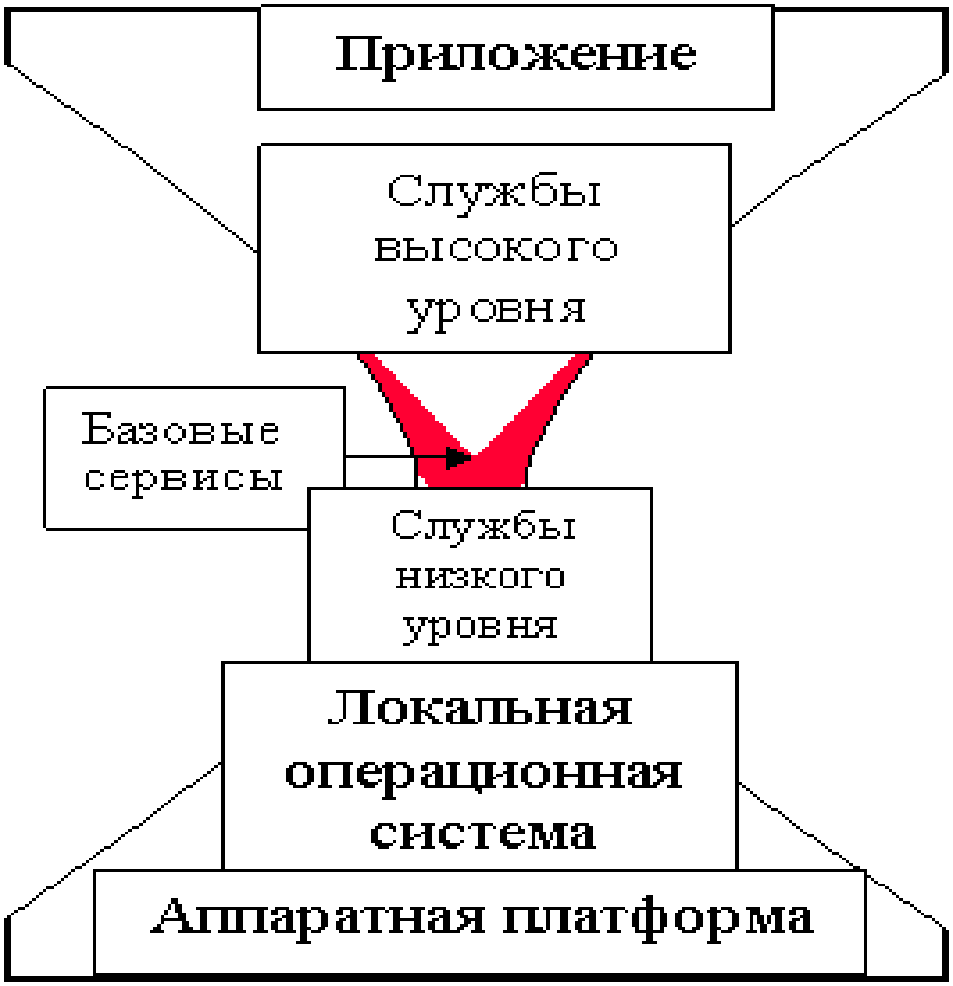
- масштабы вычислительного ресурса многократно превосходят ресурсы отдельного компьютера (вычислительного комплекса)
- гетерогенность среды
- пространственное (географическое) распределение информационно-вычислительного ресурса;
- объединение ресурсов, которые не могут управляться централизованно (не принадлежат одной организации);
- использование стандартных, открытых, общедоступных протоколов и интерфейсов;
- обеспечение информационной безопасности.

Параллельные ВС
*Области применения **GRID***

- массовая обработка потоков данных большого объема;
- многопараметрический анализ данных;
- моделирование на удаленных суперкомпьютерах;
- реалистичная визуализация больших наборов данных;
- сложные бизнес-приложения с большими объемами вычислений.

Параллельные ВС

*Архитектура **GRID** – модель «песочных часов»*



Параллельные ВС

Архитектура протоколов **GRID**

Параллельные вычислительные системы

Прикладное программное обеспечение

Параллельные ВС
Прикладное программное обеспечение

Проблемы разработки параллельного ПО

- проблема распараллеливания
- проблема отладки и верификации
- проблема наращиваемости
- проблема переносимости

Параллельные ВС
Прикладное ПО – закон Амдала

$$S \le \frac{1}{f+(1-f)/p}$$

- ***S*** **–** ускорение программы по сравнению с последовательным выполнением
- ***p*** **–** количество процессоров
- ***f*** **–** доля последовательного кода в программе ($0 \le f \le 1$)

Параллельные ВС
Прикладное ПО – подходы к созданию

- Написание параллельной программы «с нуля»
- Распараллеливание (автоматическое) существующих последовательных программ
- Смешанный подход – автоматическое распараллеливание с последующей оптимизацией

Параллельные ВС
Прикладное ПО – подходы к созданию

Написание параллельной программы «с нуля»

Достоинства:

- Возможность получения эффективного кода

Недостатки:

- Высокая трудоемкость подхода
- Высокие требования к квалификации программиста
- Высокая вероятность ошибок в коде, трудность отладки ПО

Параллельные ВС
Прикладное ПО – подходы к созданию

Автоматическое распараллеливание последовательной программы

Достоинства:

- Использование наработанного (последовательного) программного обеспечения
- Высокая надежность кода

Недостатки:

- Низкая эффективность распараллеливания

Параллельные ВС
Прикладное ПО – подходы к созданию

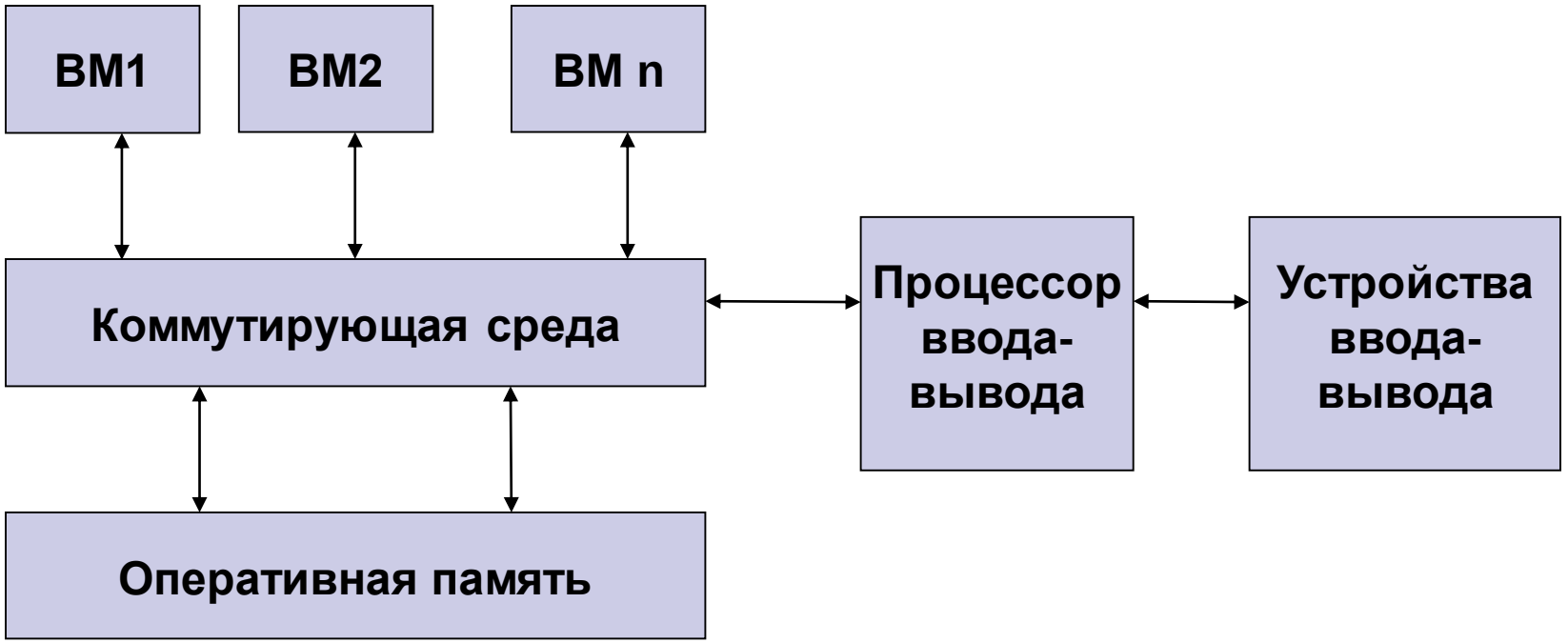
Смешанный подход – автоматическое распараллеливание с последующей оптимизацией

- Этот подход в равной мере обладает и достоинствами, и недостатками обеих методов, описанных ранее.
- Его применение требует обширного набора инструментальных программных средств.

Параллельные вычислительные системы

Программирование параллельных ВС с разделяемой памятью

Параллельные ВС класса МКМД
Системы с разделяемой памятью



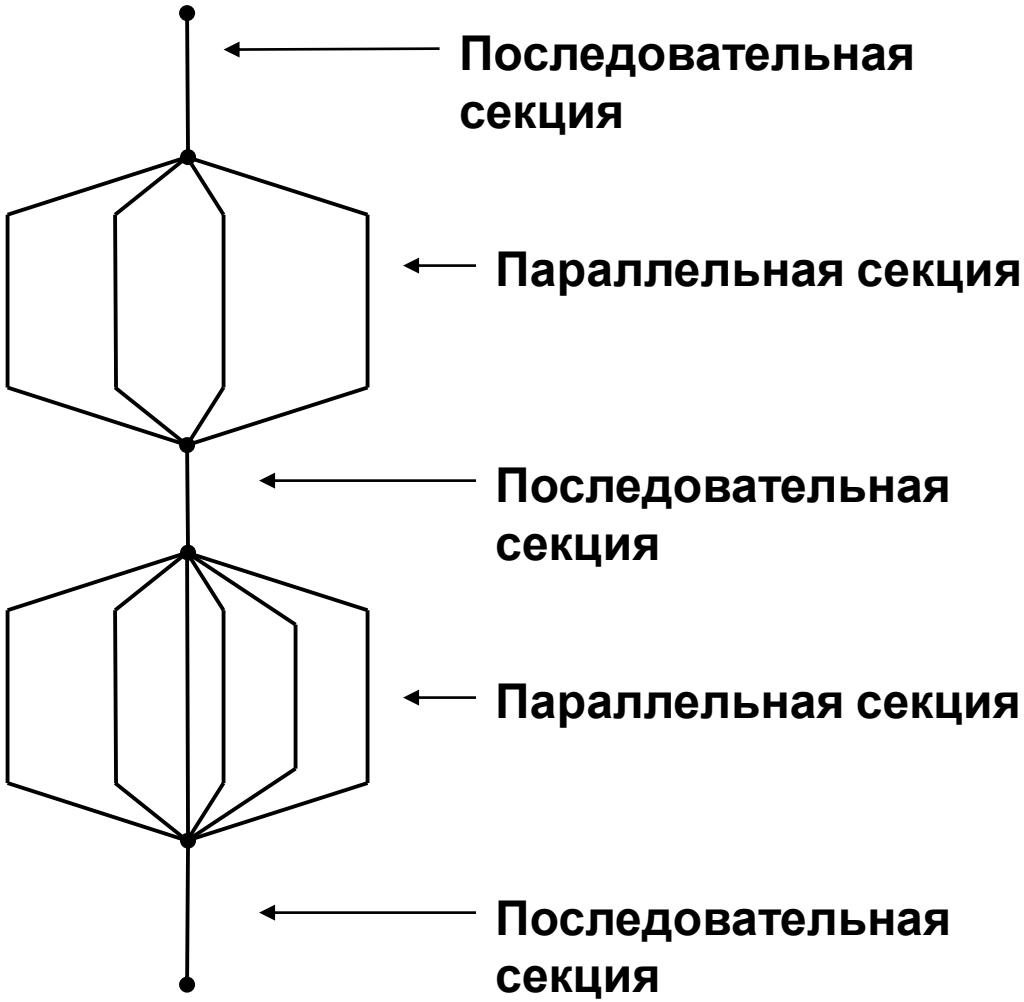
Программирование параллельных ВС
Системы с разделяемой памятью

Программирование систем с разделяемой памятью осуществляется согласно модели обмена через общую память

Инструментальные средства: POSIX threads, OpenMP.

Для подобных систем существуют сравнительно эффективные средства ***автоматического распараллеливания***.

Программирование параллельных ВС OpenMP – структура программы



Программирование параллельных ВС OpenMP – структура программы

- Основная нить и только она исполняет все последовательные области программы.
- При входе в параллельную область порождаются дополнительные нити.
- После порождения каждая нить получает свой уникальный номер, причем нить-мастер всегда имеет номер 0.
- Все нити исполняют один и тот же код, соответствующий параллельной области.
- При выходе из параллельной области основная нить дожидается завершения остальных нитей, и дальнейшее выполнение программы продолжает только она.

Программирование параллельных ВС OpenMP – переменные

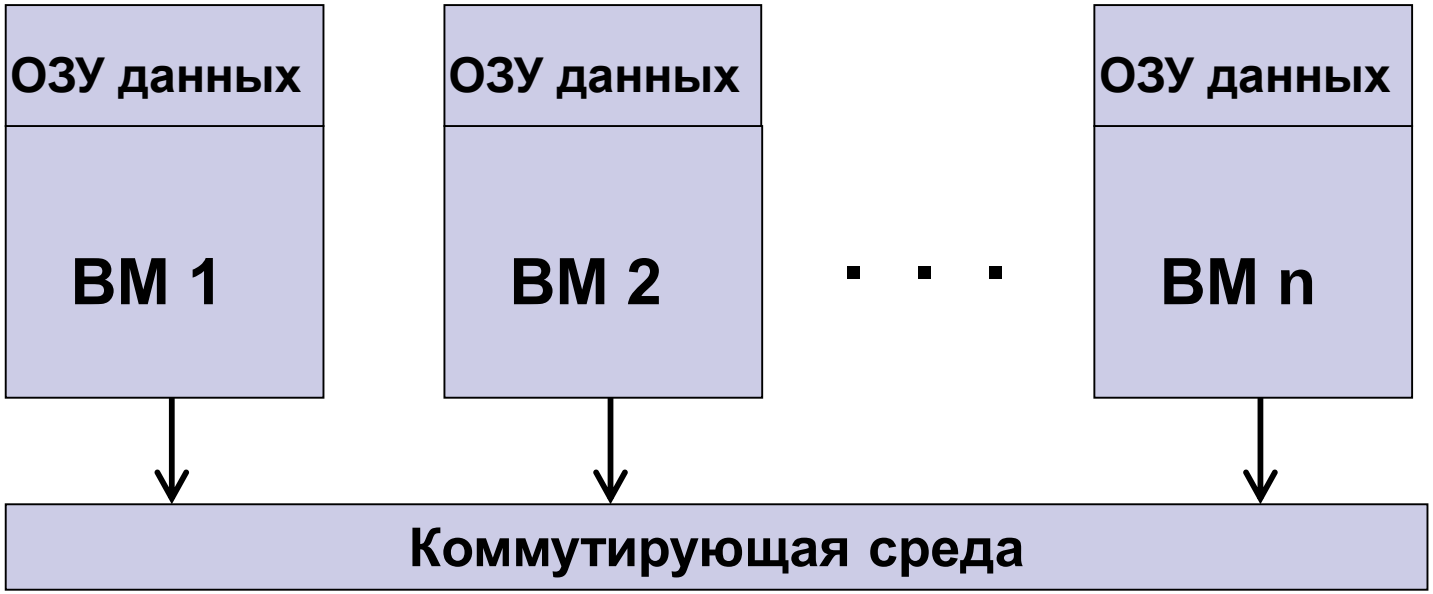
В параллельной области все переменные программы разделяются общие (***SHARED***) и локальные (***PRIVATE***).

- ***Общая переменная*** всегда существует в одном экземпляре и доступна всем нитям под одним и тем же именем.
- Объявление ***локальной переменной*** вызывает порождение своего экземпляра данной переменной для каждой нити.
- Изменение нитью значения своей ***локальной переменной*** никак не влияет на значения этой же локальной переменной в других нитях.

Параллельные вычислительные системы

Программирование кластерных и MPP параллельных ВС

Параллельные ВС класса МКМД
Кластерные и массивно-параллельные ВС



Программирование параллельных ВС Кластеры и MPP

- Программирование кластерных и MPP параллельных ВС осуществляется в рамках модели ***передачи сообщений.***
- ***Инструментальные средства:*** MPI, PVM, BSPlib.
- Стандартом, использующимся при разработке программ, основанных на передаче сообщений, является стандарт ***MPI*** (***Message Passing Interface*** – Взаимодействие через передачу сообщений).

*Программирование параллельных ВС **MPI***

- При запуске MPI-программы создается несколько ***ветвей***;
- Все ветви программы запускаются загрузчиком одновременно как процессы;
- Ветви объединяются в ***группы*** - это некое множество взаимодействующих ветвей;
- Каждой группе в соответствие ставится ***область связи***;
- Каждой области связи в соответствие ставится ***коммуникатор***.

*Программирование параллельных ВС **MPI***

Библиотека MPI состоит примерно из 130 функций, в число которых входят:

- функции инициализации и закрытия MPI-процессов;
- функции, реализующие коммуникационные операции типа точка-точка;
- функции, реализующие коллективные операции;

*Программирование параллельных ВС **MPI***

Библиотека MPI состоит примерно из 130 функций, в число которых входят:

- функции для работы с группами процессов и коммуникаторами;
- функции для работы со структурами данных;
- функции формирования топологии процессов.

MPI - Функции инициализации и завершения

int MPI_Init(int* argc, char* argv)**

- Инициализация параллельной части приложения. Все MPI-процедуры могут быть вызваны только после вызова *MPI_Init*. Возвращает: в случае успешного выполнения - *MPI_SUCCESS*, иначе - код ошибки.

int MPI_Finalize(void)

- MPI_Finalize - завершение параллельной части приложения. Все последующие обращения к любым MPI-процедурам, в том числе к MPI_Init, запрещены.

MPI – информационные функции

int MPI_Comm_size(MPI_Comm comm, int* size)

- Определение общего числа параллельных процессов в группе *comm*.
  - *comm* - идентификатор группы
  - OUT *size* - размер группы

int MPI_Comm_rank(MPI_comm comm, int* rank)

- Определение номера процесса в группе *comm*. Значение, возвращаемое по адресу *&rank*, лежит в диапазоне от 0 до *size_of_group-1*.

MPI –функции обмена «точка-точка»

int MPI_Send(void* buf, int count, MPI_Datatype datatype, int dest, int msgtag, MPI_Comm comm)

Блокирующая посылка сообщения.

- *buf* - адрес начала буфера посылки сообщения
- *count* - число передаваемых элементов в сообщении
- *datatype* - тип передаваемых элементов
- *dest* - номер процесса-получателя
- *msgtag* - идентификатор сообщения
- *comm* - идентификатор группы

MPI –функции обмена «точка-точка»

int MPI_Recv(void* buf, int count, MPI_Datatype datatype, int source, int msgtag, MPI_comm comm, MPI_Status *status)

Прием сообщения.

- OUT *buf* - адрес начала буфера приема сообщения
- *count* - максимальное число элементов в принимаемом сообщении
- *datatype* - тип элементов принимаемого сообщения
- *source* - номер процесса-отправителя
- *msgtag* - идентификатор принимаемого сообщения
- *comm* - идентификатор группы
- *status* - параметры принятого сообщения

MPI – *аргументы – «джокеры» функций обмена «точка-точка»*

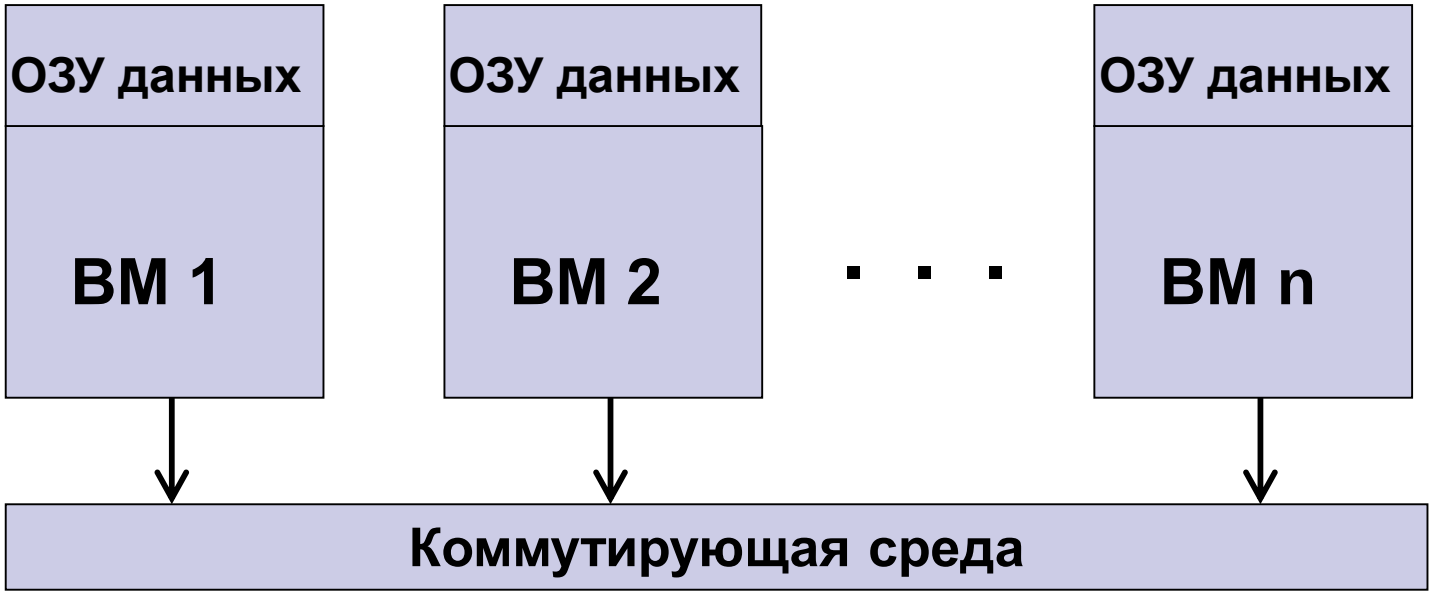
MPI_ANY_SOURCE – заменяет аргумент «номер передающего процесса»; признак того, что подходит сообщение от любого процесса.

MPI_ANY_TAG – заменяет аргумент «идентификатор сообщения»; признак того, что подходит сообщение с любым идентификатором.

Параллельные вычислительные системы

Программирование кластерных и MPP параллельных ВС

Параллельные ВС класса МКМД
Кластерные и массивно-параллельные ВС



*Программирование параллельных ВС **MPI***

Библиотека MPI состоит примерно из 130 функций, в число которых входят:

- функции инициализации и закрытия MPI-процессов;
- функции, реализующие коммуникационные операции типа точка-точка;
- функции, реализующие коллективные операции;

*Программирование параллельных ВС **MPI***

Библиотека MPI состоит примерно из 130 функций, в число которых входят:

- функции для работы с группами процессов и коммуникаторами;
- функции для работы со структурами данных;
- функции формирования топологии процессов.

MPI – коллективные функции

Под термином "***коллективные***" в MPI подразумеваются три группы функций:

- функции коллективного обмена данными;
- барьеры (точки синхронизации);
- распределенные операции.

MPI – коллективные функции

int MPI_Barrier(MPI_Comm comm);

Останавливает выполнение вызвавшей ее задачи до тех пор, пока не будет вызвана изо всех остальных задач, подсоединенных к указываемому коммуникатору. Гарантирует, что к выполнению следующей за MPI_Barrier инструкции каждая задача приступит одновременно с остальными.

MPI –функции коллективного обмена

Основные особенности и отличия от коммуникаций типа "точка-точка":

- на прием и/или передачу работают все задачи-абоненты указываемого коммуникатора;
- коллективная функция выполняет одновременно и прием, и передачу; она имеет большое количество параметров, часть которых нужна для приема, а часть для передачи; в разных задачах та или иная часть игнорируется;
- как правило, значения параметров (за исключением адресов буферов) должны быть идентичными во всех задачах;

MPI –функции коллективного обмена

int MPI_Bcast(void *buf, int count, MPI_Datatype datatype, int source, MPI_Comm comm)

Рассылка сообщения от процесса *source* всем процессам, включая рассылающий процесс.

buf - адрес начала буфера посылки сообщения

- *count* - число передаваемых элементов в сообщении
- *datatype* - тип передаваемых элементов
- *source* - номер рассылающего процесса
- *comm* - идентификатор группы

MPI –функции коллективного обмена

MPI_Gather ("совок") собирает в приемный буфер задачи root передающие буфера остальных задач.

MPI_Scatter ("разбрызгиватель") : части передающего буфера из задачи root распределяются по приемным буферам всех задач.

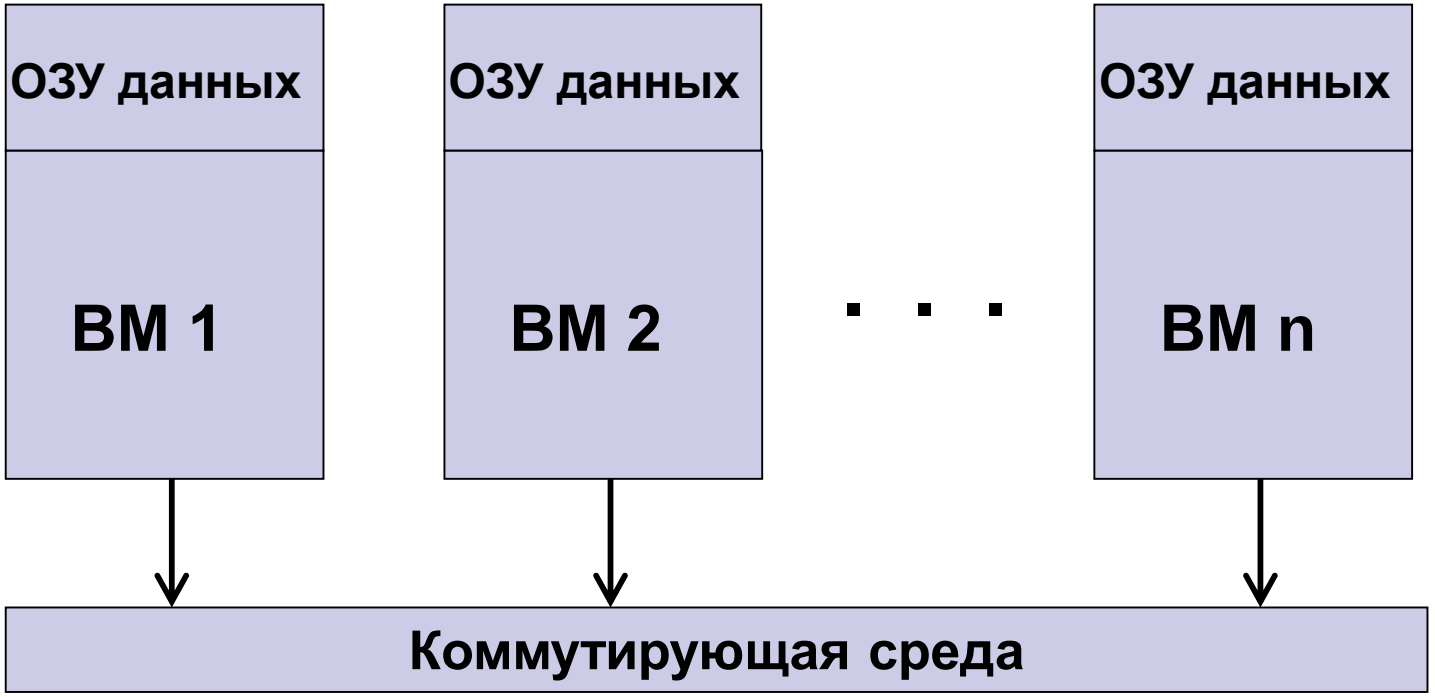
MPI_Allgather аналогична *MPI_Gather*, но прием осуществляется не в одной задаче, а во ВСЕХ: каждая имеет специфическое содержимое в передающем буфере, и все получают одинаковое содержимое в буфере приемном.

MPI_Alltoall : каждый процесс нарезает передающий буфер на куски и рассылает куски остальным процессам.

Параллельные вычислительные системы

Проектирование кластера

Параллельные ВС класса МКМД: Кластеры



Параллельные ВС класса МКМД Кластеры

- ***Архитектура*.** Набор элементов высокой степени готовности, рабочих станций или ПК общего назначения, объединяемых при помощи сетевых технологий и используемых в качестве массивно-параллельного компьютера.
- ***Коммуникационная среда***. Стандартные сетевые технологий (Fast/Gigabit Ethernet, Myrinet) на базе шинной архитектуры или коммутатора.

Параллельные ВС класса МКМД Кластеры

- При объединении в кластер компьютеров разной мощности или разной архитектуры, говорят о ***гетерогенных*** (неоднородных) кластерах.
- Узлы кластера могут одновременно использоваться в качестве пользовательских рабочих станций (***кластер WOB***)

Параллельные ВС класса МКМД Кластеры

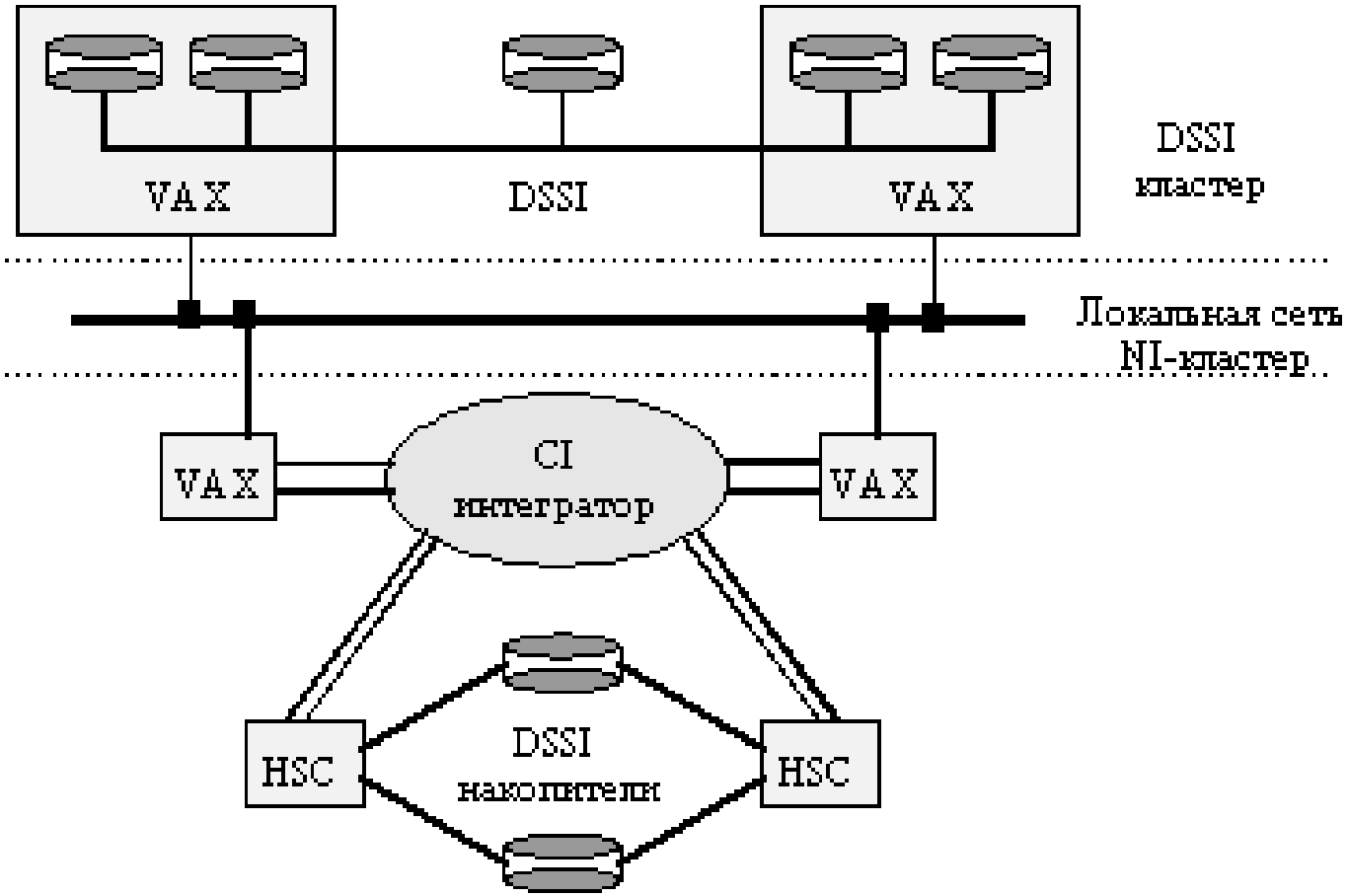
- ***Операционная система*** **-** стандартные ОС - Linux/FreeBSD, вместе со средствами поддержки параллельного программирования и распределения нагрузки.
- ***Модель программирования*** **-** с использованием передачи сообщений (PVM, MPI).
- ***Основная проблема*** - большие накладные расходы на взаимодействие параллельных процессов между собой, что сильно сужает потенциальный класс решаемых задач.

Кластеры высокой надежности

- в случае сбоя ПО на одном из узлов приложение продолжает функционировать или автоматически перезапускается на других узлах кластера;
- выход из строя одного из узлов (или нескольких) не приведет к краху всей кластерной системы;
- профилактические и ремонтные работы, реконфигурацию или смену версий программного обеспечения можно осуществлять в узлах кластера поочередно, не прерывая работы других узлов.

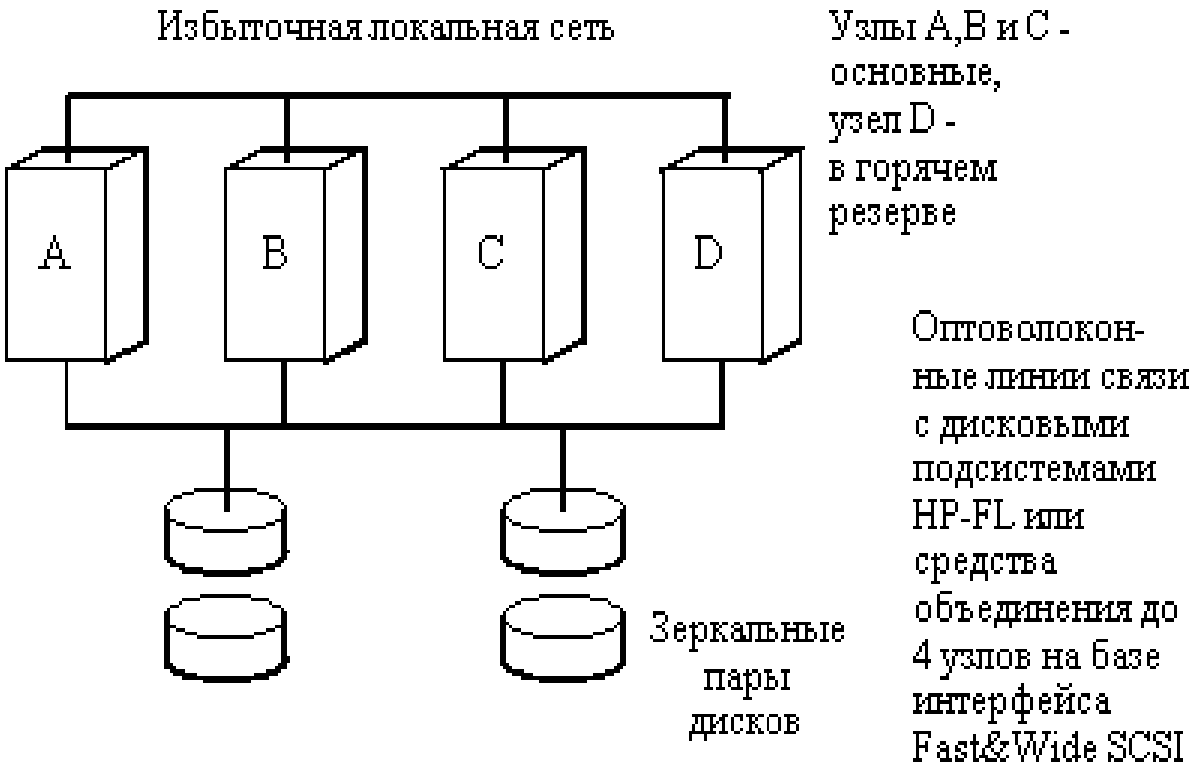
Кластеры высокой надежности

VAX/VMS кластер



Кластеры высокой надежности Switchover/UX компании Hewlett Packard

- *Средства автоматического восстановления*
- *Средства прозрачного переназначения сетевого адреса*
- *Средства прозрачного восстановления прикладных систем*
- *Средства маскирования одиночных отказов и сбоев*
- *Поддержка управления логическими томами*

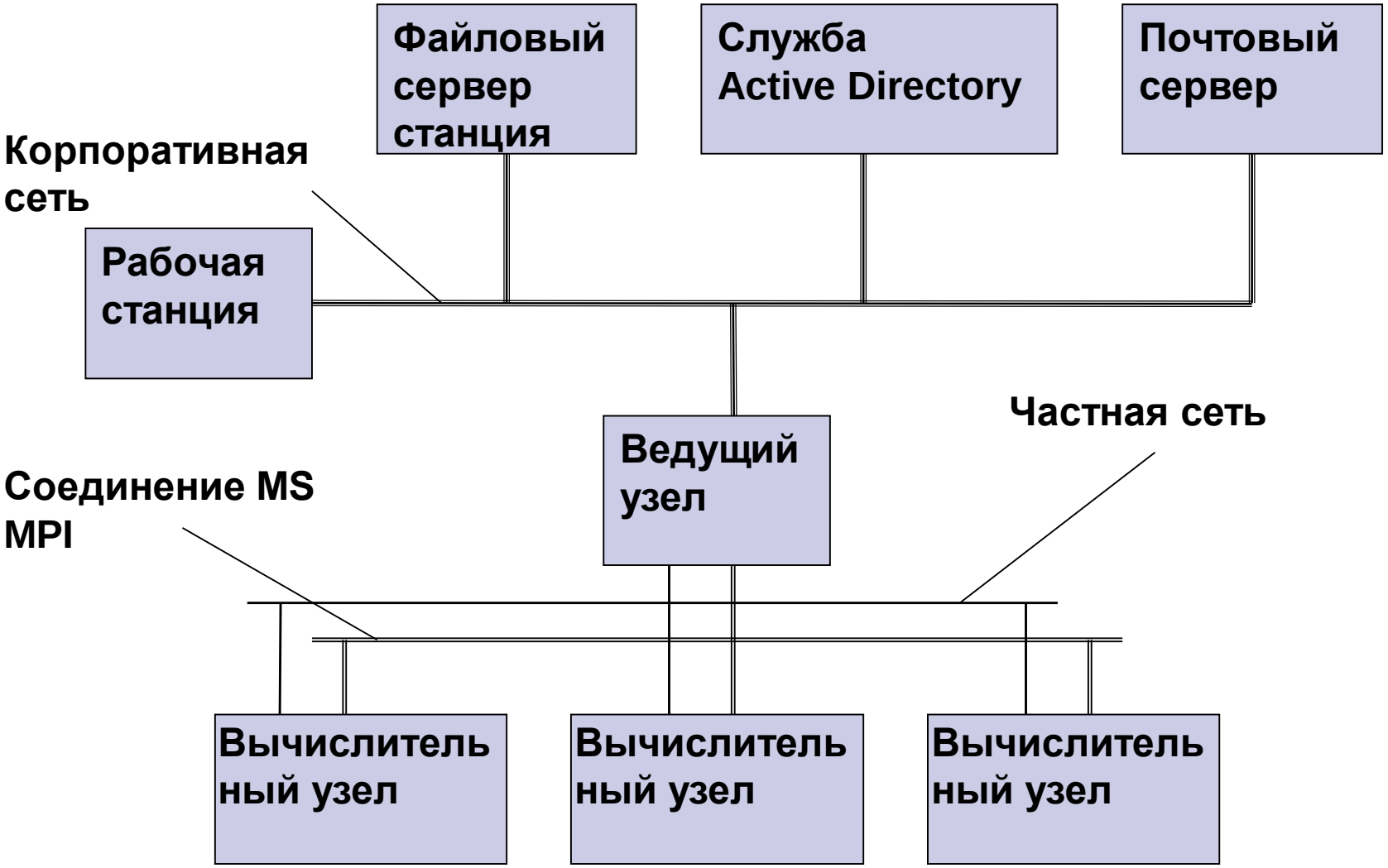


Высокопроизводительные кластеры

Высокопроизводительный кластер -

- параллельная вычислительная система с распределенной памятью;
- построенная из компонент общего назначения;
- с единой точкой доступа;
- однородными вычислительными узлами;
- специализированной сетью, обеспечивающей эффективный обмен данными.

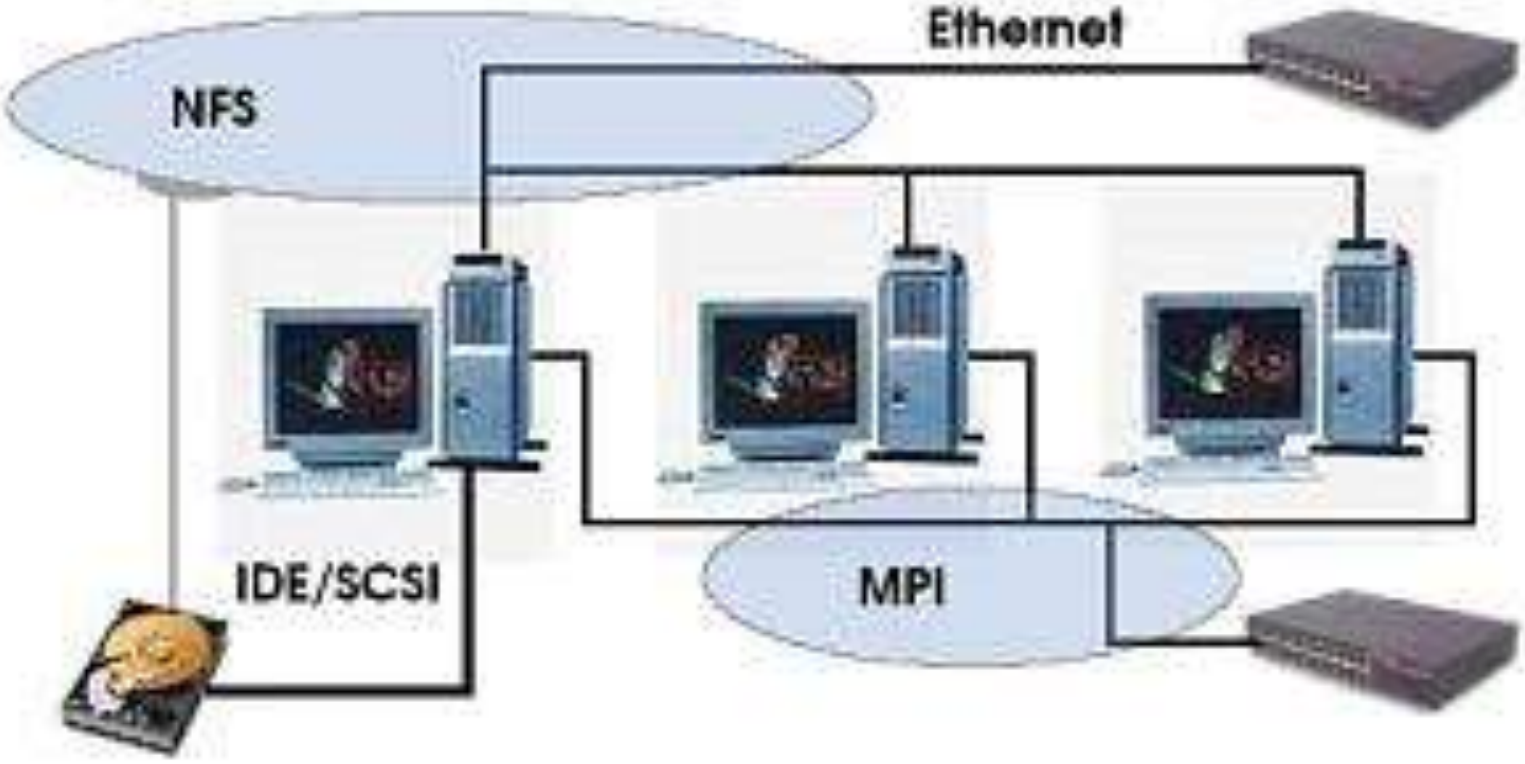
Высокопроизводительные кластеры



Характеристики коммутирующих сред

Сетевое оборудование	Пиковая пропускная способность	Латентность
FastEthernet	12.5 Mbyte/sec	150 sec
GigabitEthernet	125 Mbyte/sec	150 sec
Myrinet	160 Mbyte/sec	5 sec
SCI	400 Mbyte/sec (реально 100)	2.3 sec
cLAN	150 Mbyte/sec	30 sec

Кластеры на основе локальной сети (Cluster Of Workstations – COW)



Параллельные вычислительные системы

Системное ПО кластера

Кластеры - Системное ПО
Windows Compute Cluster Server 2003

- Упрощенная настройка параметров безопасности и проверки подлинности за счет использования существующих экземпляров Active Directory.
- Управление обновлениями для узлов с помощью Microsoft Systems Management Server (SMS).
- Управление системой и заданиями с помощью Microsoft Operations Manager (MOM).
- Использование оснасток из состава консоли управления Майкрософт (MMC).
- CCS совместим с ведущими приложениями в каждой из целевых групп. Это позволяет развертывать серийные приложения, пользуясь разнообразными вариантами поддержки.

Кластеры - Системное ПО
Solaris (Sun Microsystems)

Коммерческая верся ***UNIX***.

- поддержка до 1 млн. одновременно работающих процессов;
- до 128 процессоров в одной системе и до 848 процессоров в кластере;
- до 576 Гбайт физической оперативной памяти;
- поддержка файловых систем размером до 252 Тбайт;
- наличие средств управления конфигурациями и изменениями;
- встроенная совместимость с Linux.

Кластеры - Системное ПО

HP-UX (Hewlett-Packard)

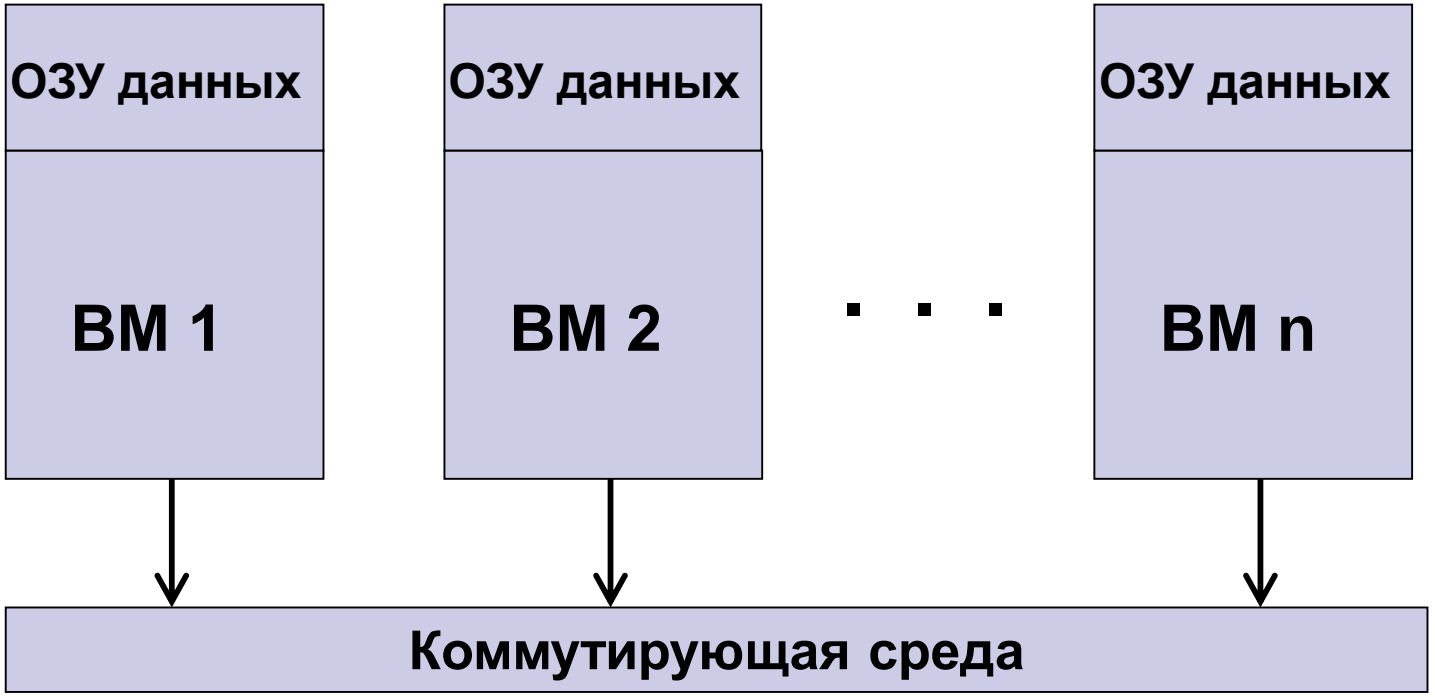
Потомок ***AT&T System V.***

- поддерживает до 256 процессоров;
- поддерживает кластеры размером до 128 узлов;
- подключение и отключение дополнительных процессоров, замену аппаратного обеспечения, динамическую настройку и обновление операционной системы без перезагрузки;
- резервное копирование в режиме on-line и дефрагментацию дисков без выключения системы.

Параллельные вычислительные системы

Кластер на основе локальной сети

Параллельные ВС класса МКМД: Кластеры



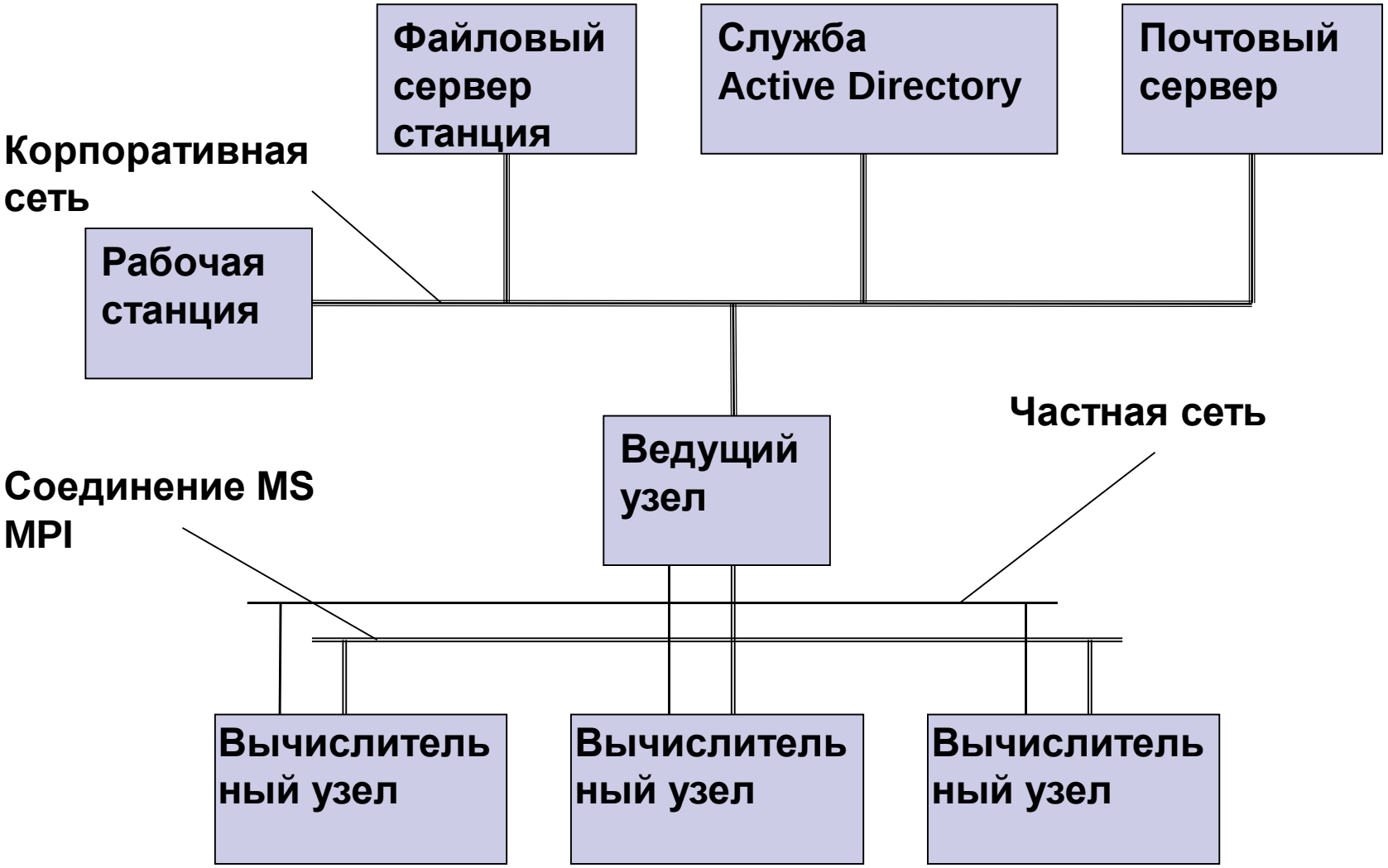
Кластеры на основе локальной (корпоративной) сети

- При объединении в кластер компьютеров разной мощности или разной архитектуры, говорят о ***гетерогенных*** (неоднородных) кластерах.
- Узлы кластера могут одновременно использоваться в качестве пользовательских рабочих станций (***кластер WOB***)

Кластеры на основе локальной (корпоративной) сети

- ***Операционная система*** **-** стандартные ОС - вместе со средствами поддержки параллельного программирования и распределения нагрузки.
- ***Модель программирования*** **-** с использованием передачи сообщений (PVM, MPI).
- ***Основная проблема*** - большие накладные расходы на взаимодействие параллельных процессов между собой, что сильно сужает потенциальный класс решаемых задач.

Кластер COW


Файловый сервер станция
Служба Active Directory
Почтовый сервер
Корпоративная сеть
Рабочая станция
Ведущий узел
Частная сеть
Соединение MS MPI
Вычислитель ный узел
Вычислитель ный узел
Вычислитель ный узел

Кластеры на основе локальной сети (Cluster Of Workstations – COW)