

Многопроцессорные вычислительные системы

Лекция 1

1 октября 2018 г.

Задачи для МВС

- Предсказания погоды, климата и изменений в атмосфере
- Науки о материалах
- Построение полупроводниковых приборов
- Структурная биология
- Разработка фармацевтических препаратов
- Генетика человека
- Квантовая хромодинамика
- Астрономия
- Транспортные задачи
- Гидро и газодинамика
- Управляемый термоядерный синтез
- Эффективность систем сгорания топлива
- Разведка нефти и газа
- Распознавание и синтез речи
- Распознавание изображений

Единицы измерения и термины

MBC — многопроцессорные вычислительные системы

HPC (High-performance computing) — высокопроизводительные вычисления

Многие современные процессоры управляются **тактовым генератором**. Процессор внутри состоит из логических элементов и ячеек памяти — триггеров. Когда приходит сигнал от тактового генератора, триггеры приобретают своё новое значение, и логическим элементам требуется некоторое время для декодирования новых значений. Затем приходит следующий сигнал от тактового генератора...

Такт процессора или такт ядра процессора — промежуток между двумя импульсами тактового генератора, который синхронизирует выполнение всех операций процессора.

Выполнение различных элементарных операций (команд или **инструкций**) может занимать от долей такта до многих тактов в зависимости от команды.

Единицы измерения вычислительной мощности компьютера:

MIPS (Million Instructions Per Second) — количество миллионов инструкций процессора в секунду.

FLOPS (FLoating-point Operations Per Second) — количество операций с плавающей точкой в секунду.

Наборы инструкций процессора 1/2

RISC (*Restricted (reduced) Instruction Set Computer*) — архитектура процессора, в которой быстродействие увеличивается за счёт упрощения инструкций, чтобы их декодирование было более простым, а время выполнения — короче.

Упрощение инструкций облегчает повышение тактовой частоты.

CISC (*Complex Instruction Set Computing*) — архитектура процессора, в которой реализуется расширенный набор инструкций.

Данная концепция проектирования процессоров характеризуется следующим набором свойств:

- нефиксированное значение длины команды;
- арифметические действия кодируются в одной команде;
- небольшое число регистров, каждый из которых выполняет строго определённую функцию.

Наборы инструкций процессора 2/2

MISC (**M**inimum **I**nstruction **S**et **C**omputer) — вычисления с минимальным набором команд. Архитектура строится на стековой вычислительной модели с ограниченным числом команд (примерно 20—30 команд).

VLIW (**V**ery **L**ong **I**nstruction **W**ord) —сверхдлинное командное слово. Архитектура процессоров с явно выраженным параллелизмом вычислений, заложенным в систему команд процессора.

Ключевым отличием от CISC-процессоров является то, что для них загрузкой исполнительных устройств занимается часть процессора (планировщик), в то время как загрузкой вычислительных устройств для VLIW-процессора занимается компилятор, на что отводится существенно больше времени.

Неоднозначность понятия «суперкомпьютер»

Несколько попыток определения понятия суперкомпьютер (экономическое, физическое, философское)

Согласно Wikipedia (осень 2016): системы с вычислительной мощностью более 10 терафлопсов (10^{13} или десять триллионов FLOPS) можно причислять к суперкомпьютерам

Производительность системы может сильно зависеть от типа выполняемой задачи

Оценка производительности осуществляется с помощью специализированных тестов

Оценки производительности

Различают **пиковую производительность** (теоретическая величина = производительность одного процессора * число процессоров в системе)

Реальную (фактическую) производительность – величина вычисляется с помощью ряда тестов

Рейтинги МВС:

- TOP500,
- Green500,
- Top50 СНГ.

Основные тесты:

- LINPACK(решение СЛАУ)
- HPL (высокопараллельная реализация Linpack с применением MPI)
- NAMD (решение задач молекулярной динамики)
- HPCC (HPC Challenge Benchmark)
- NAS Parallel Benchmarks (версия 3.3 состоит из 11 тестов)
- 14(24) ливерморских цикла (ориентированы на векторные компьютеры)

Классификации МВС

Классификации архитектур

- Флинна (основная)
- Дополнения Ванга и Бриггса к классификации Флинна
- Фенга (число бит в слове и число слов)
- Шора (типичные способы компоновки)
- Хендлера (по возможности конвейерной и параллельной обработки информации)
- Хокни (по топологии соединения)
- Скилликорна
- Шнайдера

Классификация по Флинну

	Одиночный поток команд (Single Instruction)	Множество потоков команд (Multiple Instruction)
Одиночный поток данных (Single Data)	SISD (ОКОД)	MISD (МКОД)
Множество потоков данных (Multiple Data)	SIMD (ОКМД)	MIMD (МКМД)

Классы SISD и MISD

SISD (Single Instruction stream over a Single Data stream) – один поток команд и один поток данных.

MISD (Multiple Instruction Single Data) – к этому классу ряд исследователей относит конвейерные ЭВМ, однако это не нашло окончательного признания, поэтому можно считать, что реальных систем — представителей данного класса не существует.

SIMD

SIMD (*single instruction, multiple data* — **оди́ночный поток команд, множественный поток данных**) — принцип компьютерных вычислений, позволяющий обеспечить параллелизм на уровне данных.

SIMD-компьютеры состоят из одного командного процессора (управляющего модуля), называемого контроллером, и нескольких модулей обработки данных, называемых процессорными элементами. Управляющий модуль принимает, анализирует и выполняет команды. Если в команде встречаются данные, контроллер рассылает на все процессорные элементы команду, и эта команда выполняется на нескольких или на всех процессорных элементах. Каждый процессорный элемент имеет свою собственную память для хранения данных.

Одним из преимуществ данной архитектуры считается то, что в этом случае более эффективно реализована логика вычислений. До половины логических инструкций обычного процессора связано с управлением выполнением машинных команд, а остальная их часть относится к работе с внутренней памятью процессора и выполнению арифметических операций.

MIMD

MIMD (*multiple instruction, multiple data* — **множественный поток команд, множественный поток данных**) — MIMD компьютер имеет N процессоров, независимо исполняющих N потоков команд и обрабатывающих N потоков данных. Каждый процессор функционирует под управлением собственного потока команд, то есть MIMD компьютер может параллельно выполнять совершенно разные программы.

MIMD архитектуры далее классифицируются в зависимости от физической организации памяти, то есть имеет ли процессор свою собственную локальную память и обращается к другим блокам памяти, используя коммутирующую сеть, или коммутирующая сеть подсоединяет все процессоры к общедоступной памяти. Исходя из организации памяти, различают несколько типов архитектур

Дополнения Ванга и Бриггса

Класс **SISD** разбивается на два подкласса:

- архитектуры с единственным функциональным устройством (PDP-11)
- архитектуры, имеющие в своем составе несколько функциональных устройств (CDC 6600, CRAY-1, FPS AP-120B, CDC Cyber 205, FACOM VP-200)

В классе **SIMD** также вводится два подкласса:

- архитектуры с пословно-последовательной обработкой информации (ILLIAC IV, PEPE, BSP)
- архитектуры с разрядно-последовательной обработкой (STARAN, ICL DAP)

В классе **MIMD**:

- вычислительные системы со слабой связью между процессорами, к которым они относят все системы с распределенной памятью (Cosmic Cube)
- вычислительные системы с сильной связью (системы с общей памятью), куда попадают такие MBC, как C.mmp, BBN Butterfly, CRAY Y-MP, Denelcor HEP.

Классификация Хокни

Классификация машин MIMD-архитектуры:

- **Переключаемые** — с общей памятью и с распределённой памятью. (В данный класс попадают машины, в которых возможна связь каждого процессора с каждым, реализуемая с помощью переключателей — машины с распределённой памятью. Если же память есть разделяемый ресурс, машина называется с общей памятью.)
- **Конвейерные** (В этот класс попадают машины с одним конвейерным устройством обработки, работающим в режиме разделения времени для отдельных потоков.)
- **Сети** — регулярные решётки, гиперкубы, иерархические структуры, изменяющие конфигурацию. (Имеют распределенную память. Разделяются по топологии.)

Классификация Фенга

В 1972 году Фенг (T. Feng) предложил классифицировать вычислительные системы (ВС) на основе двух простых характеристик.

Первая характеристика – **число n бит в машинном слове**, обрабатываемых параллельно при выполнении машинных инструкций.

Вторая характеристика – **число m слов**, обрабатываемых одновременно.

Произведение $P = n \times m$ определяет интегральную характеристику потенциала параллельности архитектуры, которую Фенг назвал максимальной степенью параллелизма ВС.

Иерархия памяти

Иерархия памяти

Различные виды памяти образуют иерархию, на различных уровнях которой расположены памяти с отличающимися

- временем доступа,
- сложностью,
- стоимостью,
- объемом.

Возможность построения иерархии памяти вызвана тем, что большинство алгоритмов обращаются в каждый промежуток времени к небольшому набору данных, который может быть помещен в более быструю, но дорогую и поэтому небольшую, память.

Уровни иерархии памяти

1. Внутренняя память процессора (регистры и кэш нескольких уровней)
2. Оперативная память (ОЗУ)
3. Вторичная память
(жесткие диски, твердотельные накопители)
4. Третичная память
(внешние носители, сетевые и распределенные хранилища)

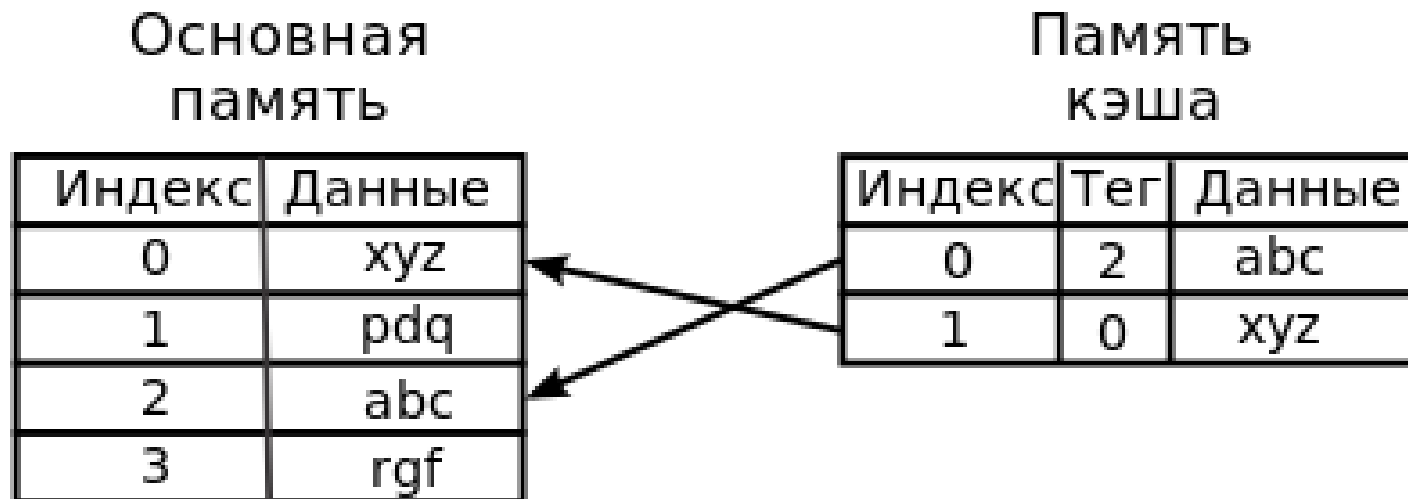
Пирамида иерархии памяти



Виды кэшей

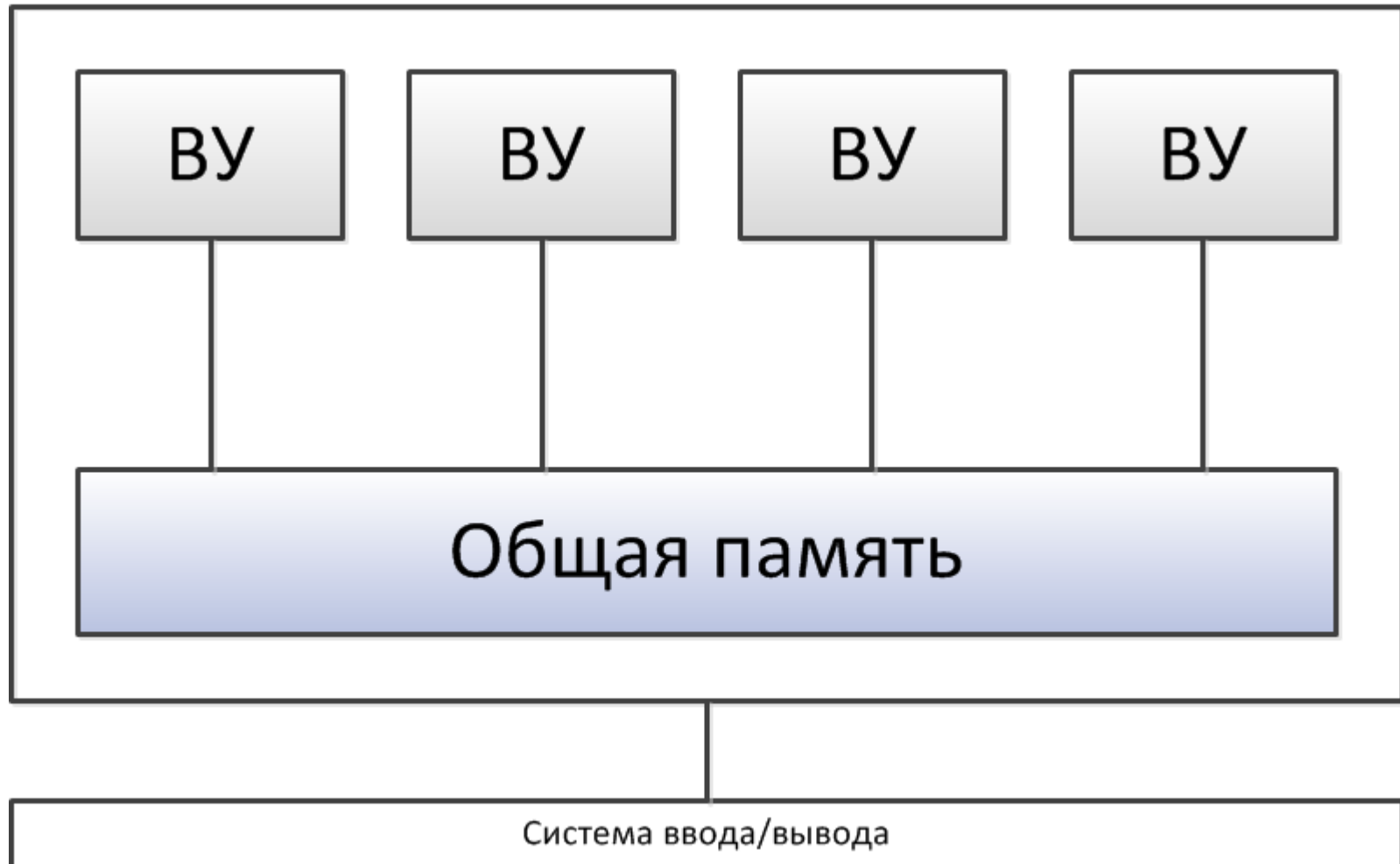
- Кэш данных (как правило 3 уровня)
- Кэш инструкций
- Буфер ассоциативной трансляции (TLB) для ускорения трансляции виртуальных (математических) адресов в физические, как для инструкций, так и для данных

Взаимодействие оперативной памяти и кэш-памяти



Архитектуры МВС

Symmetric multiprocessing (SMP)



Особенности SMP-архитектуры

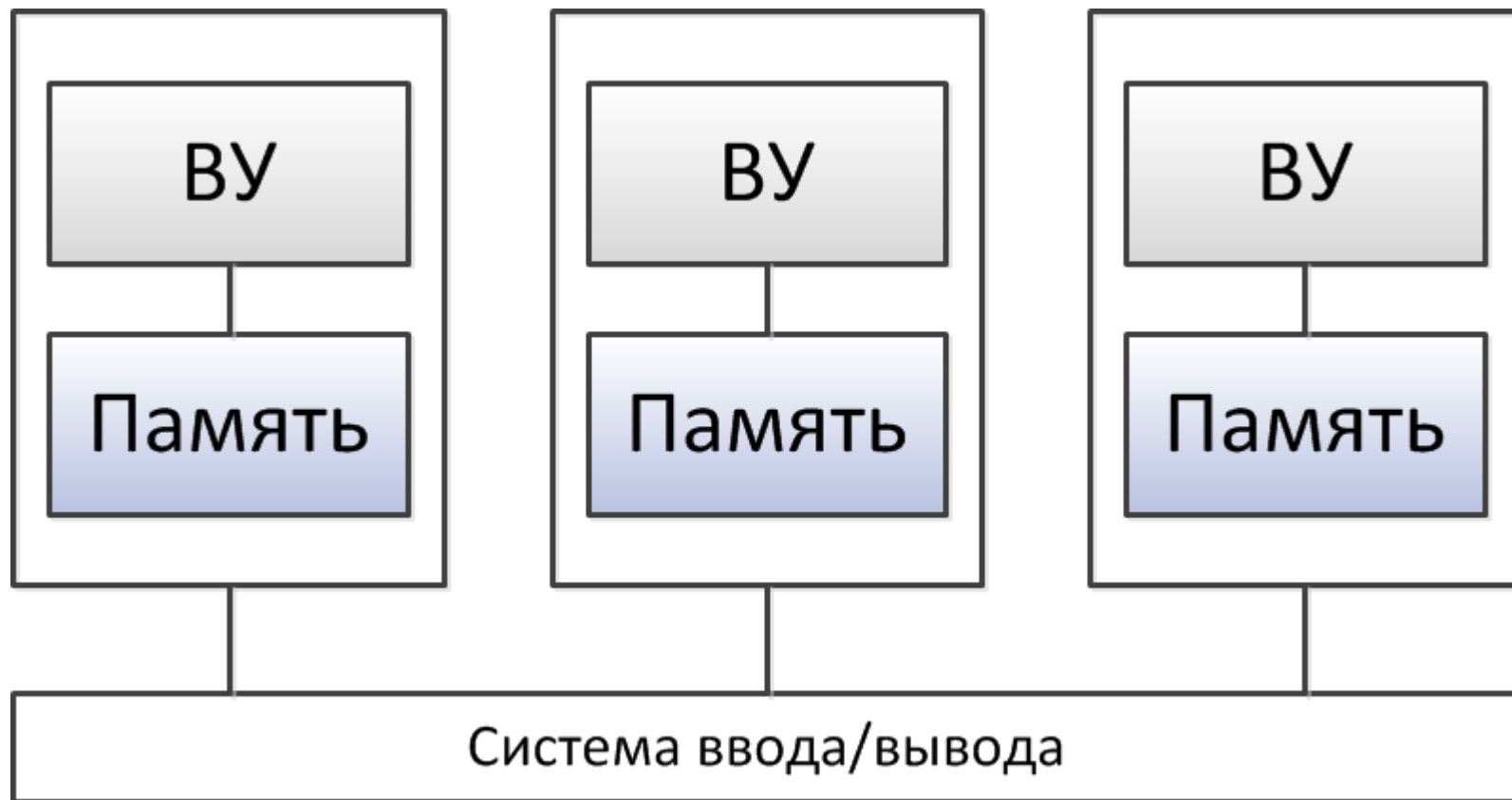
Преимущества

- Простота и универсальность программирования. Нет ограничений на модель программирования
- Простота эксплуатации и тех. обслуживания
- Невысокая цена комплектующих

Недостатки

- Плохая масштабируемость

Massive parallel processing (MPP)



Особенности MPP-архитектуры

Преимущества

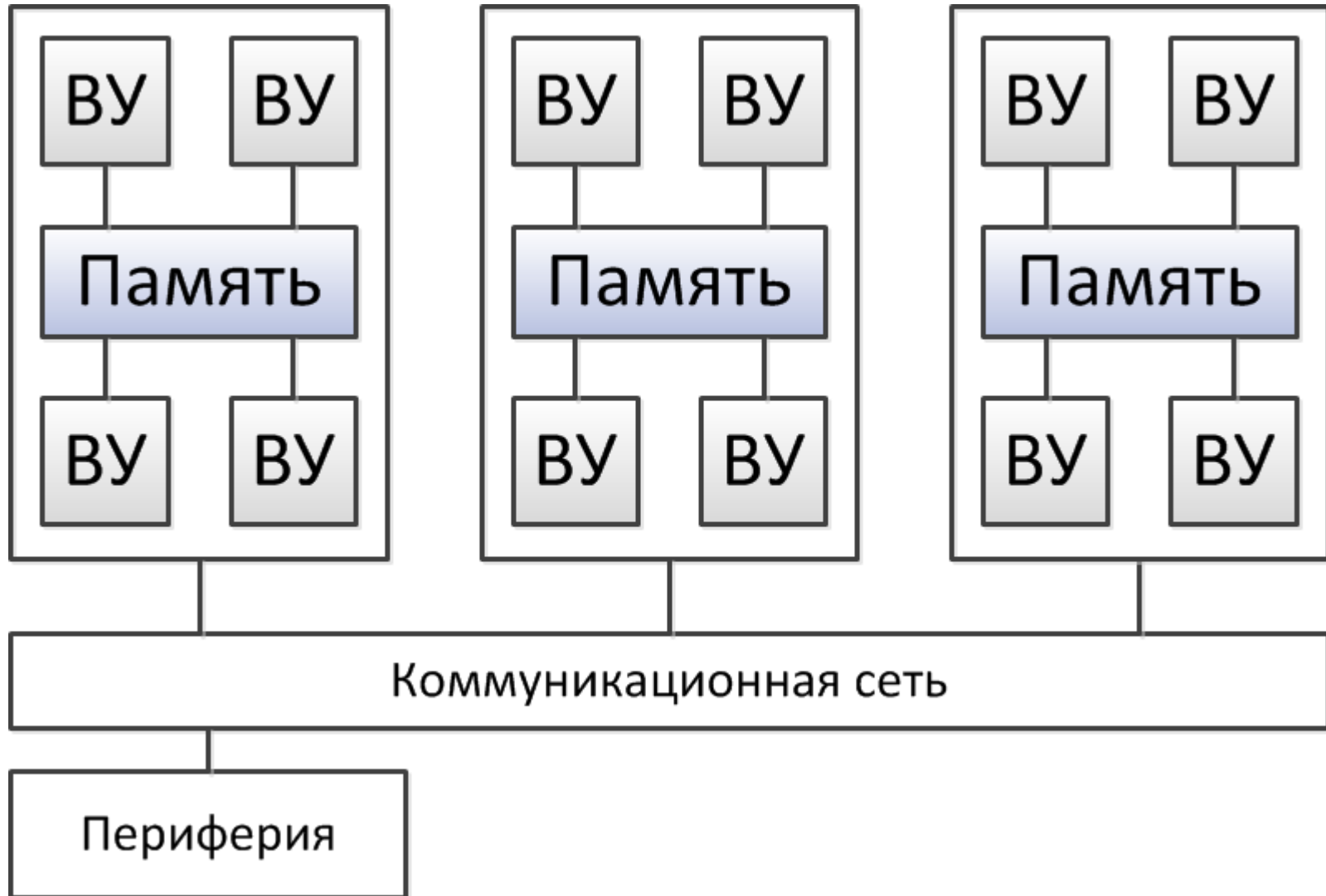
- Хорошая масштабируемость
- Большинство рекордов производительности устанавливаются на машинах MPP-архитектуры

Недостатки

- Отсутствие общей памяти снижает скорость межпроцессорного обмена
- Каждый процессор может использовать только ограниченный объем банка памяти
- Высокая цена программного обеспечения

Гибридная архитектура NUMA

(NonUniform Memory Access)



Когерентность памяти

Когерентность кэшей процессоров:

все процессоры получают одинаковые значения одних переменных в любой момент времени

Когерентность кэшей. Механизмы

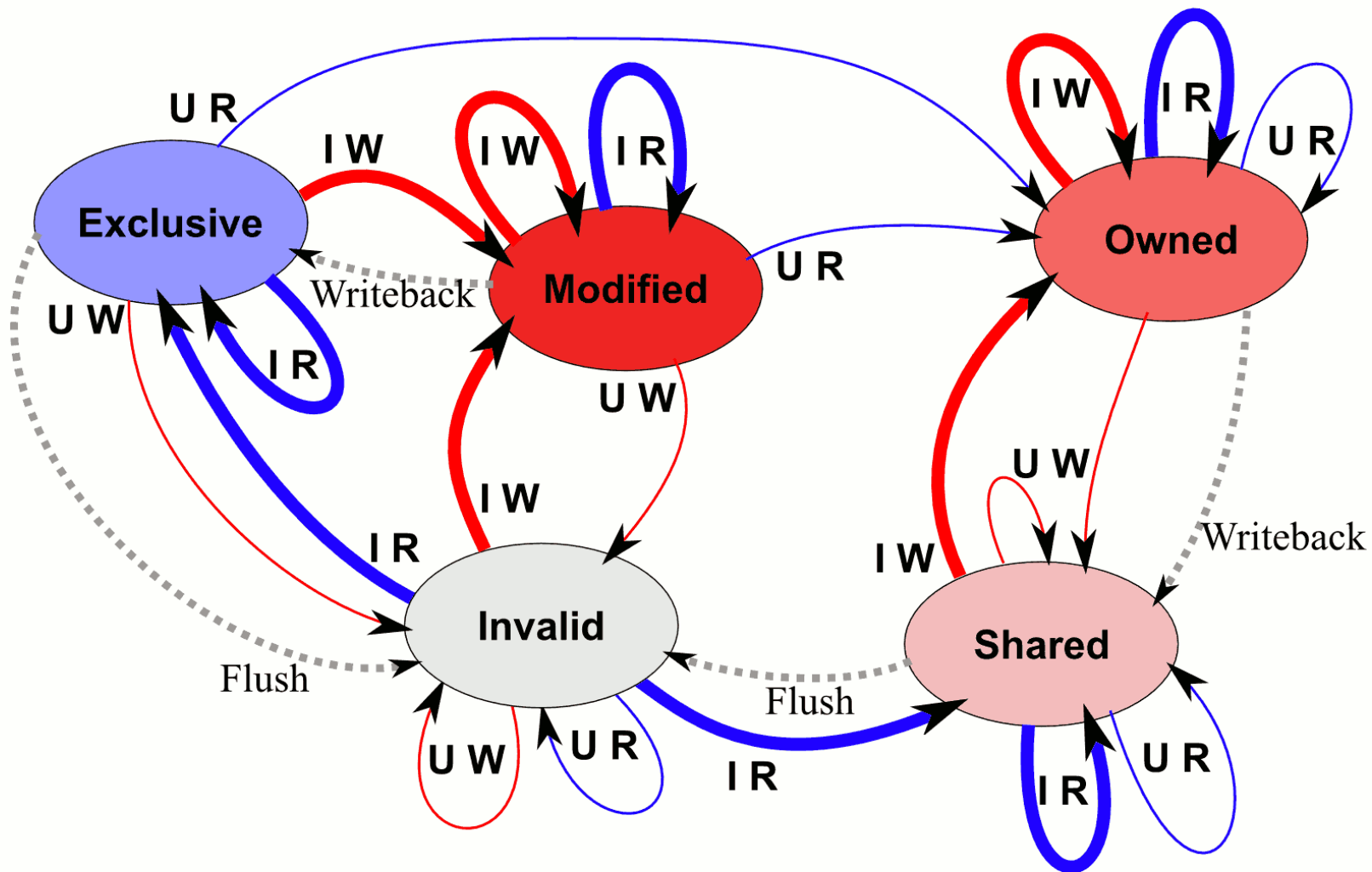
- **Когерентность с использованием справочника (*directory*)**
Информация о состоянии блока физической памяти содержится только в одном месте, называемом справочником (физически справочник может быть распределен по узлам системы).
- **Когерентность с использованием отслеживания (*snooping*)**
Каждый кэш, который содержит копию данных некоторого блока физической памяти, имеет также соответствующую копию служебной информации о его состоянии. Централизованная система записей отсутствует. Обычно кэши расположены на общей (разделяемой) шине и контроллеры всех кэшей наблюдают за шиной (просматривают ее) для определения того, не содержат ли они копию соответствующего блока.
- **Перехват (*snarfing*)**
Когда из какого-либо одного кэша данные переписываются в оперативную память, контроллеры остальных получают сигнал об этом изменении ("перехватывают" информацию об изменении данных) и, если необходимо, изменяют соответствующие данные в своих кэшах.

Алгоритм MOESI когерентности кэшей

Используется в процессорах AMD Opteron

- Modified (Модифицирована)
- Owned (Владелец)
- Exclusive (Эксклюзивное)
- Shared (Разделяемое)
- Invalid (Не актуальное)

Диаграмма состояний (MOESI)



Параллельные векторные системы (PVP)

Основным признаком PVP-систем является наличие специальных векторно-конвейерных процессоров, в которых предусмотрены команды однотипной обработки векторов независимых данных, эффективно выполняющиеся на конвейерных функциональных устройствах.

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).

Кластерные системы

Набор рабочих станций (или даже ПК) общего назначения, используется в качестве дешевого варианта MPP-архитектуры

Для связи узлов используется одна из стандартных сетевых технологий на базе шинной архитектуры или коммутатора (к примеру Ethernet)

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. (несколько функций у узлов кластера)

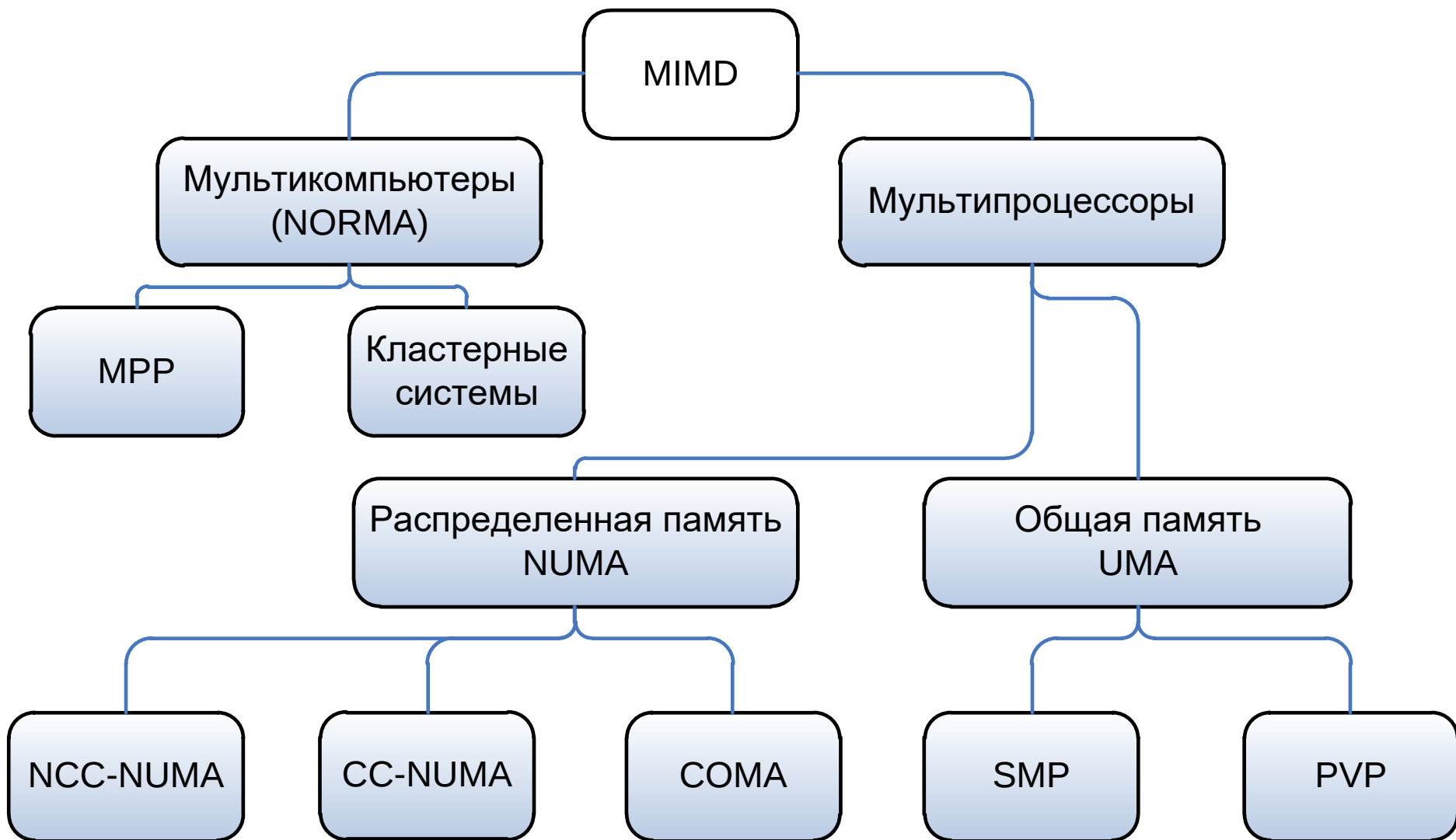
Программирование, как правило, в рамках модели передачи сообщений (MPI), технологии PVM (Parallel Virtual Machine)

Дешевизна подобных систем оборачивается большими накладными расходами на взаимодействие параллельных процессов между собой, что сильно сужает потенциальный класс решаемых задач

Кластер. Тип Beowulf

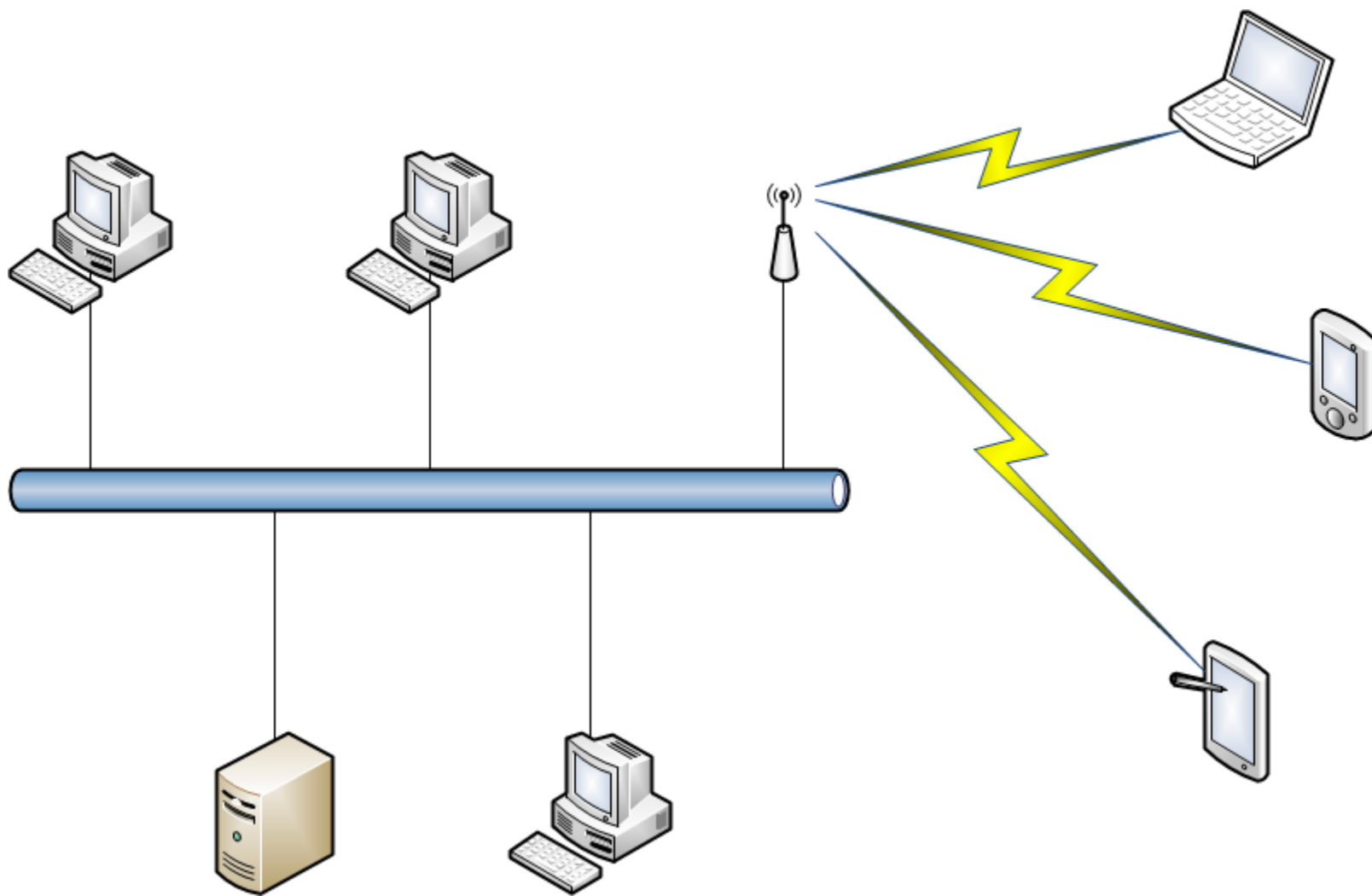


Разделение класса MIMD по типу доступа к памяти

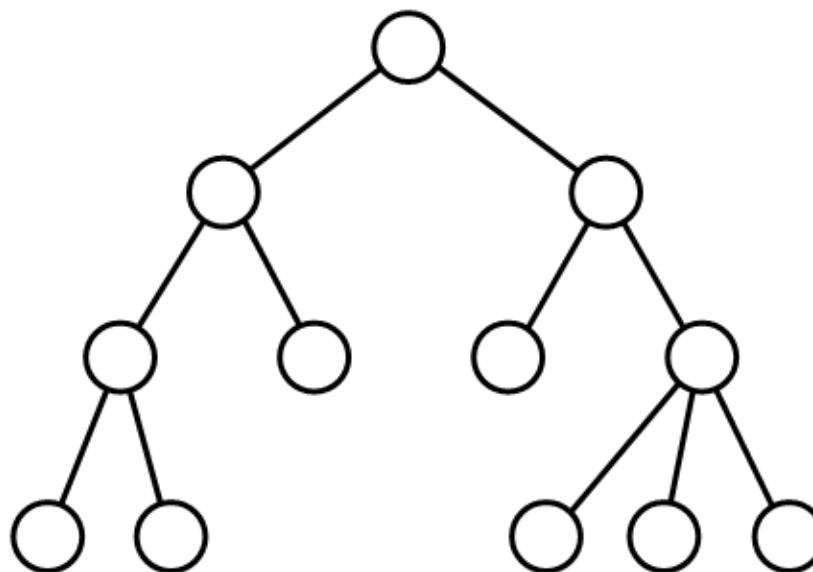
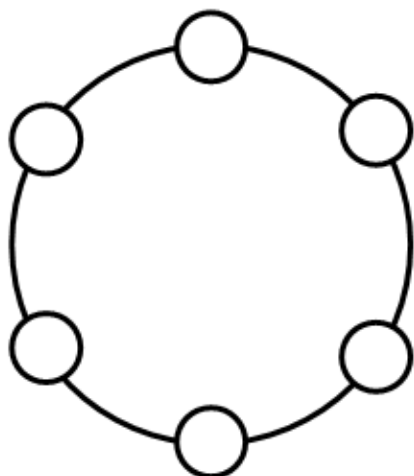
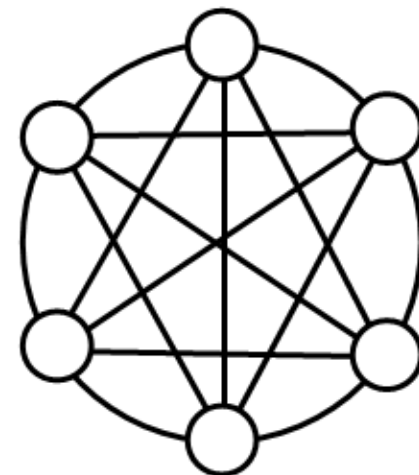
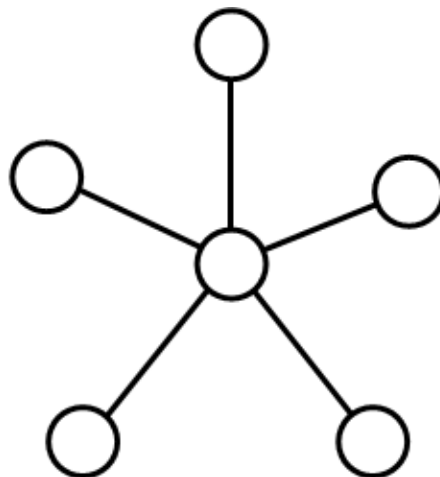
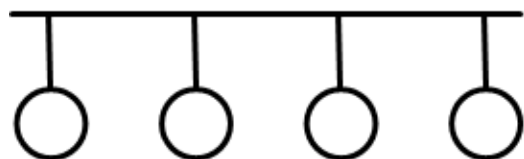
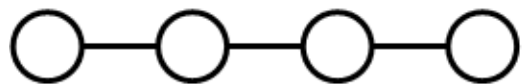


Сетевые топологии МВС

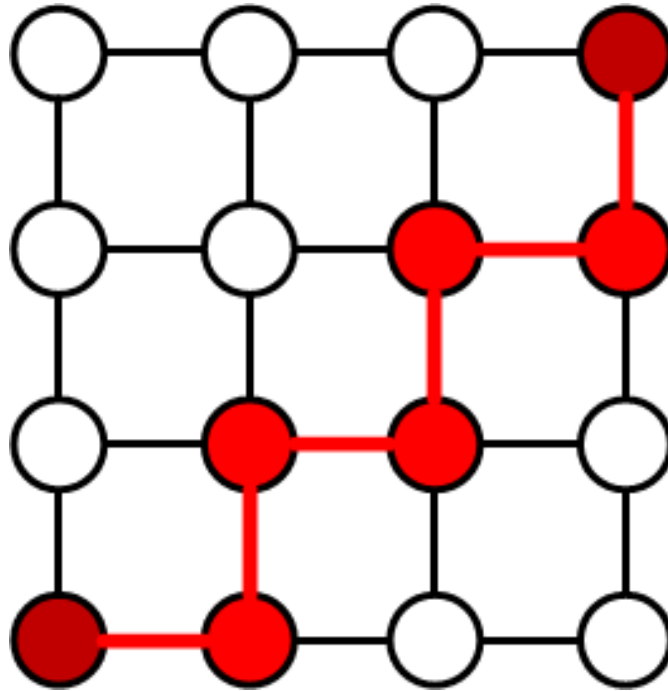
Сетевая топология. Локальная сеть



Типы сетевых топологий

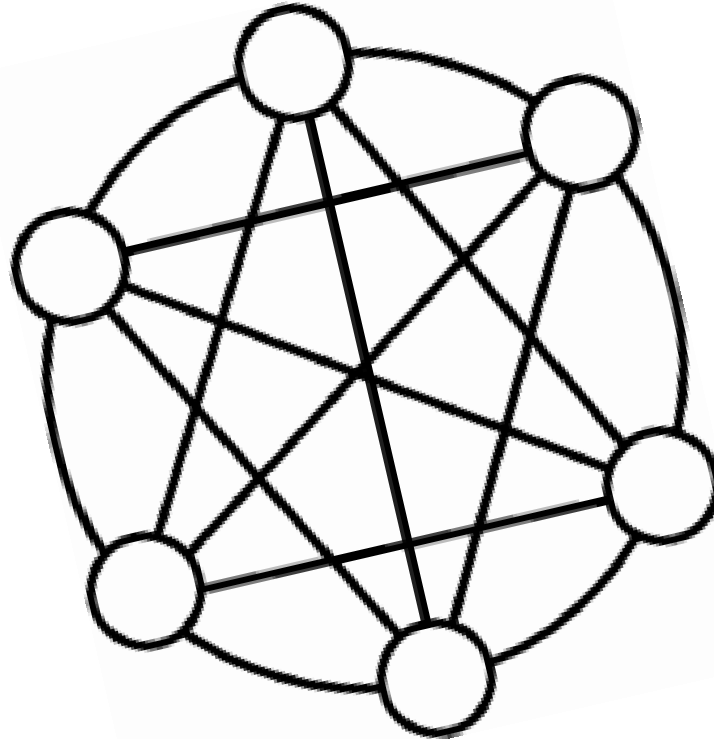


Расстояние между процессорами

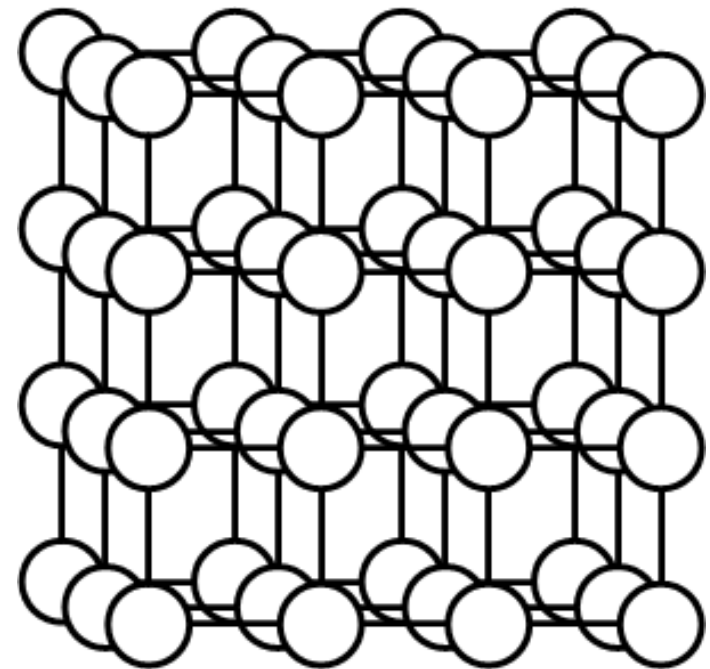
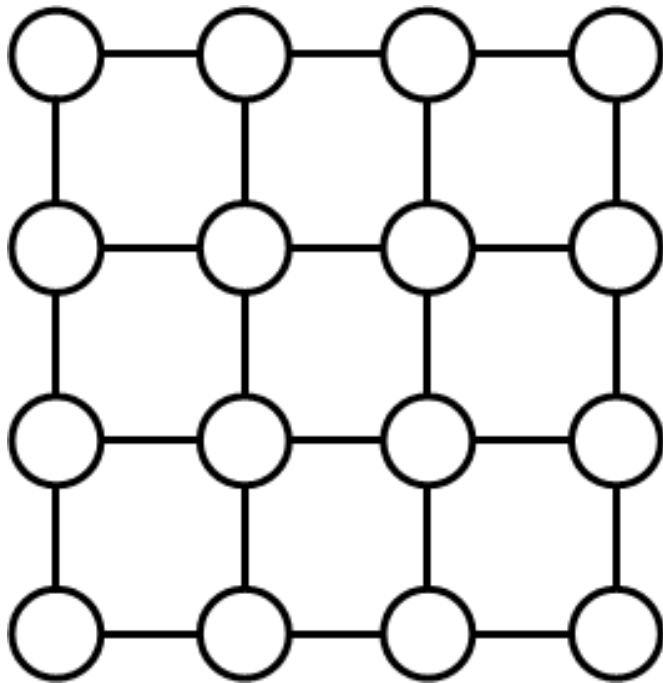


Максимальное расстояние = не более 4 промежуточных связей

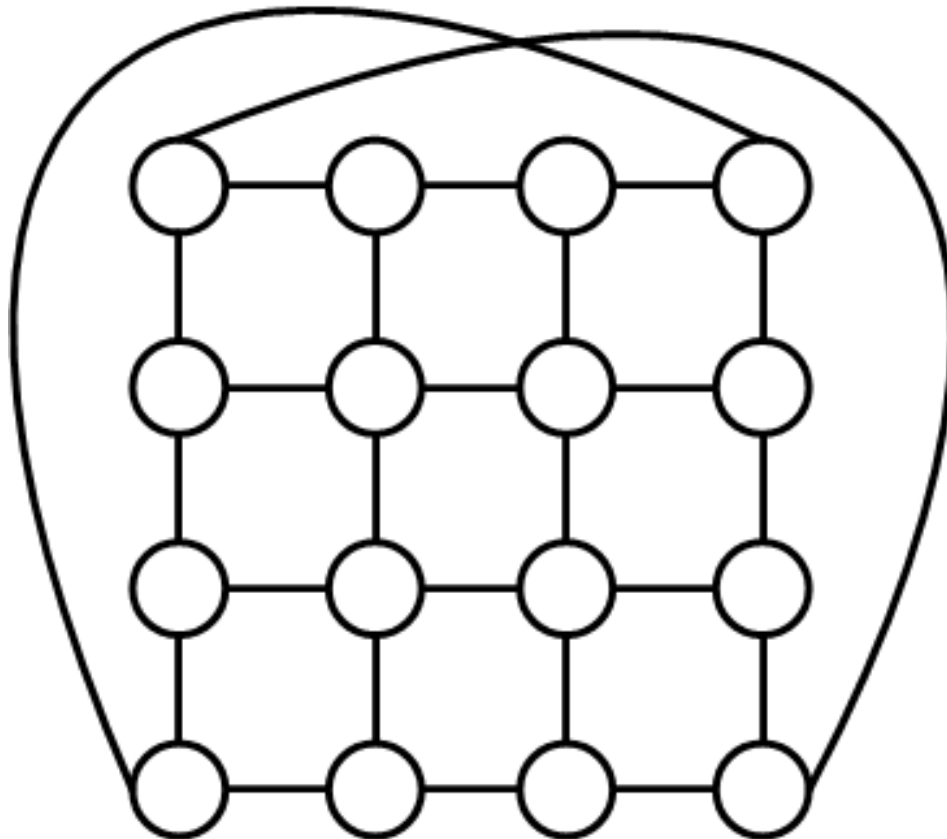
Топологии МВС. Полносвязная



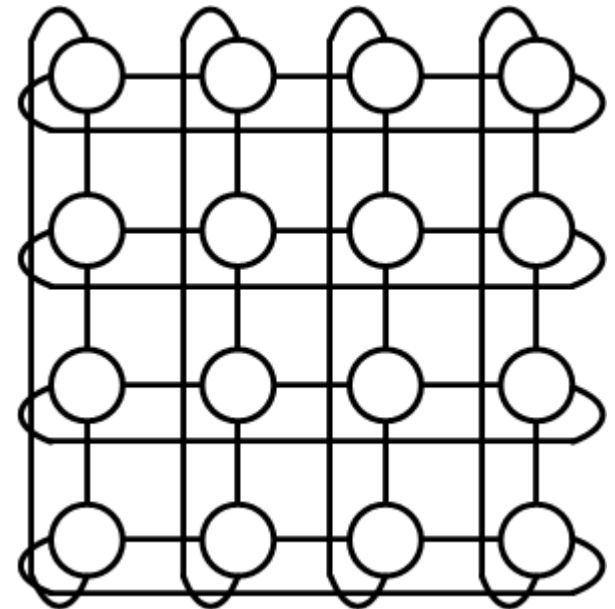
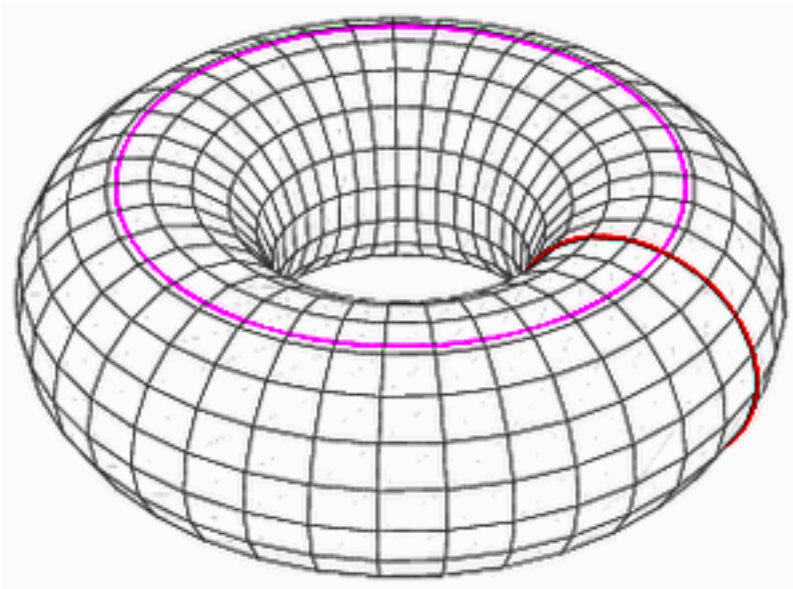
Топологии МВС. Решетка



Топологии МВС. Модификация решетки

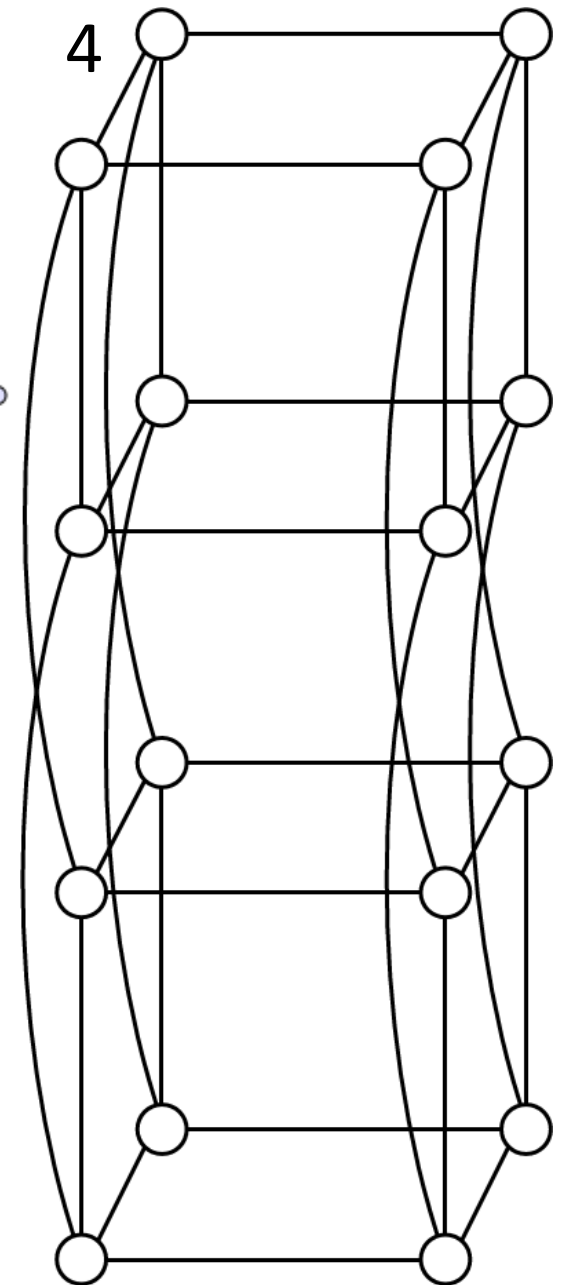
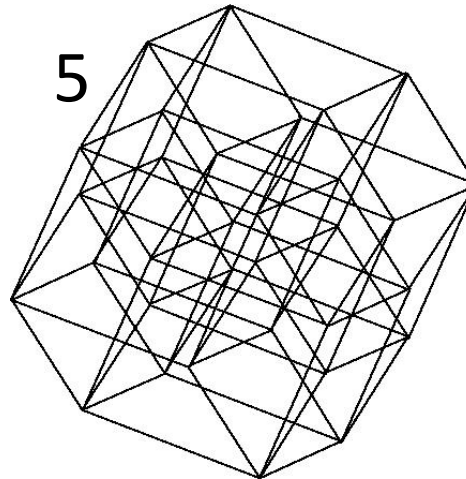
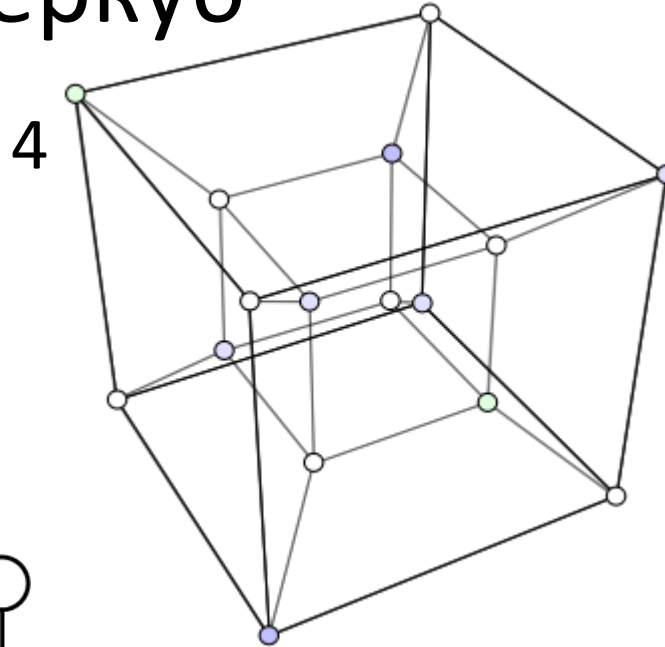
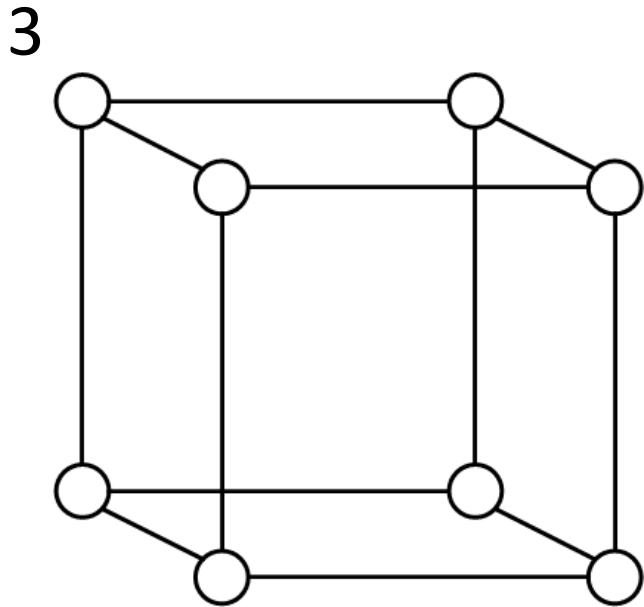
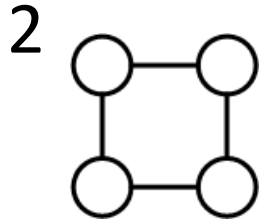


Топологии МВС. Тор

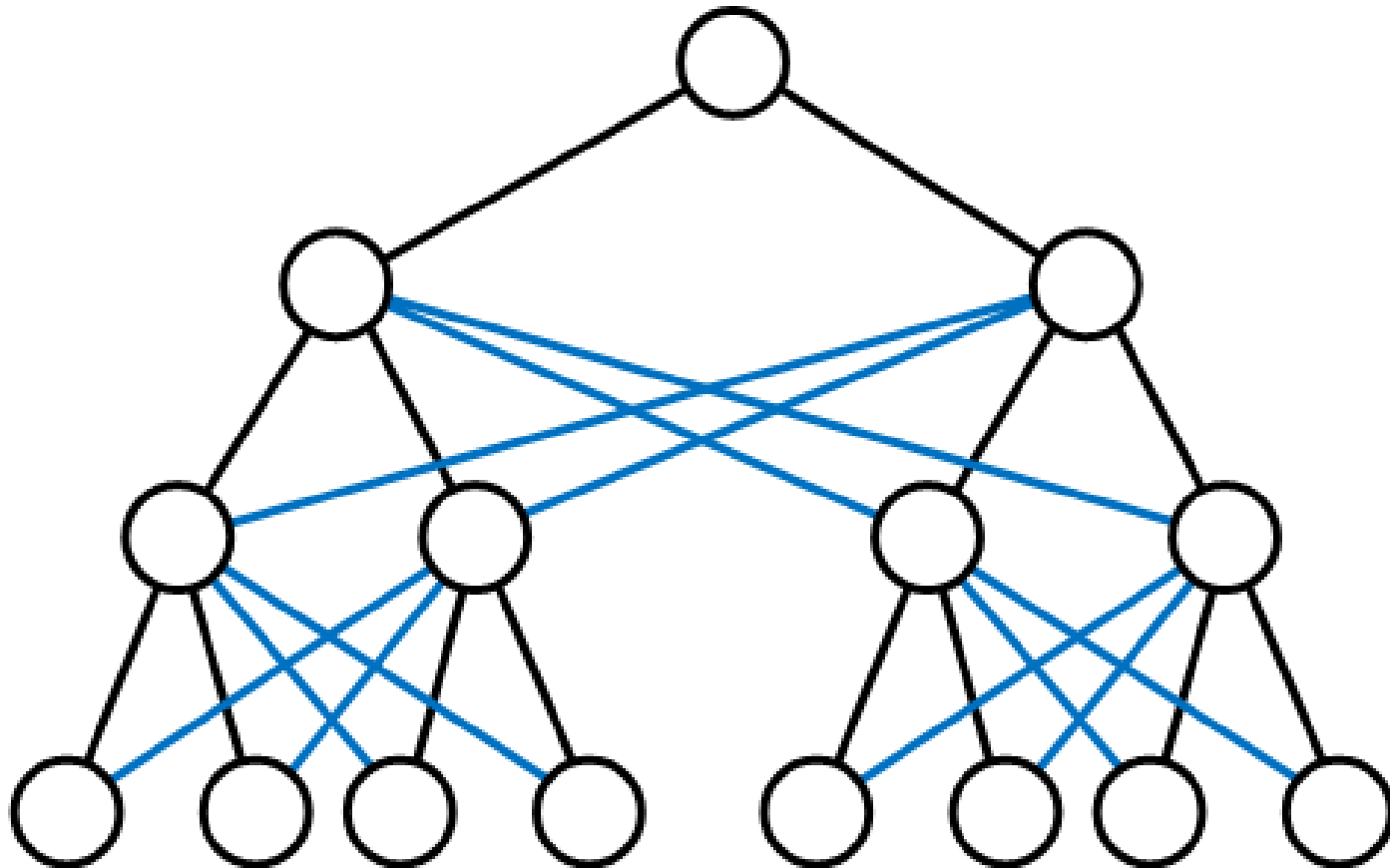


Топологии МВС.

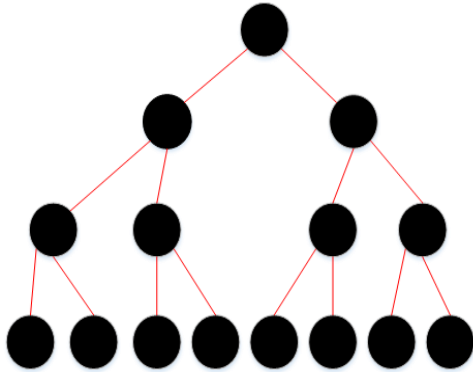
Гиперкуб



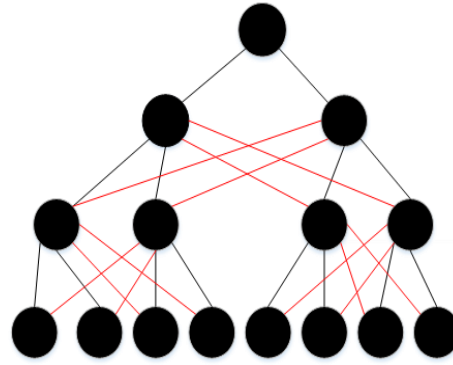
Топологии МВС. Fat tree



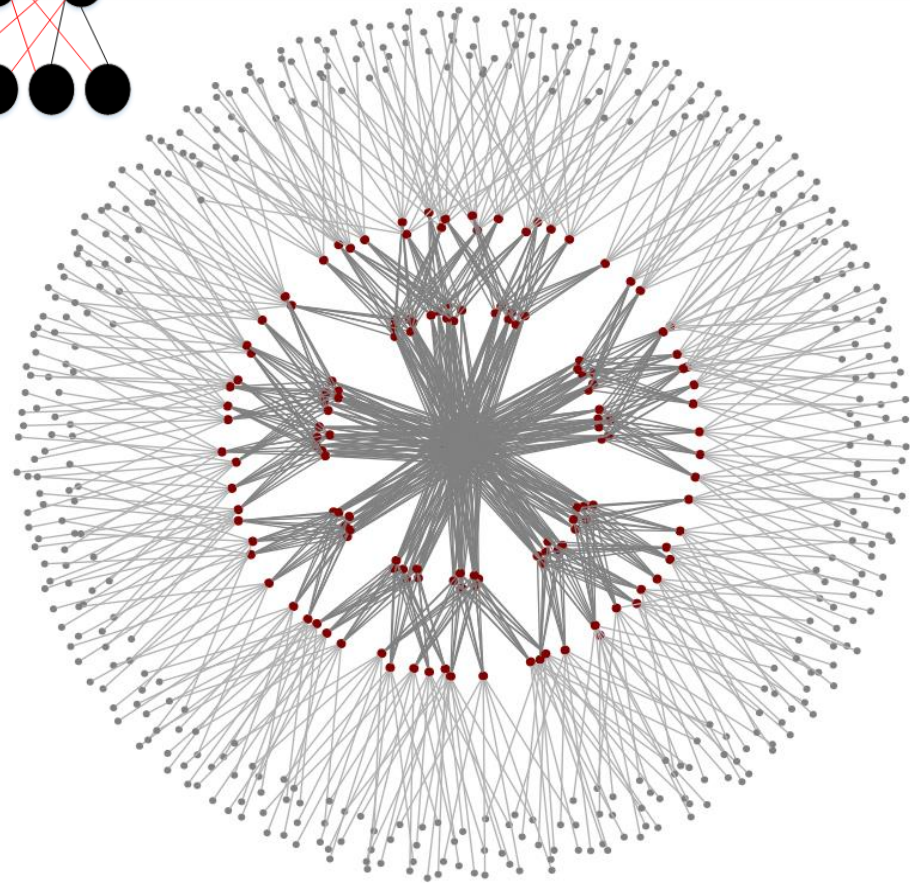
FatTree



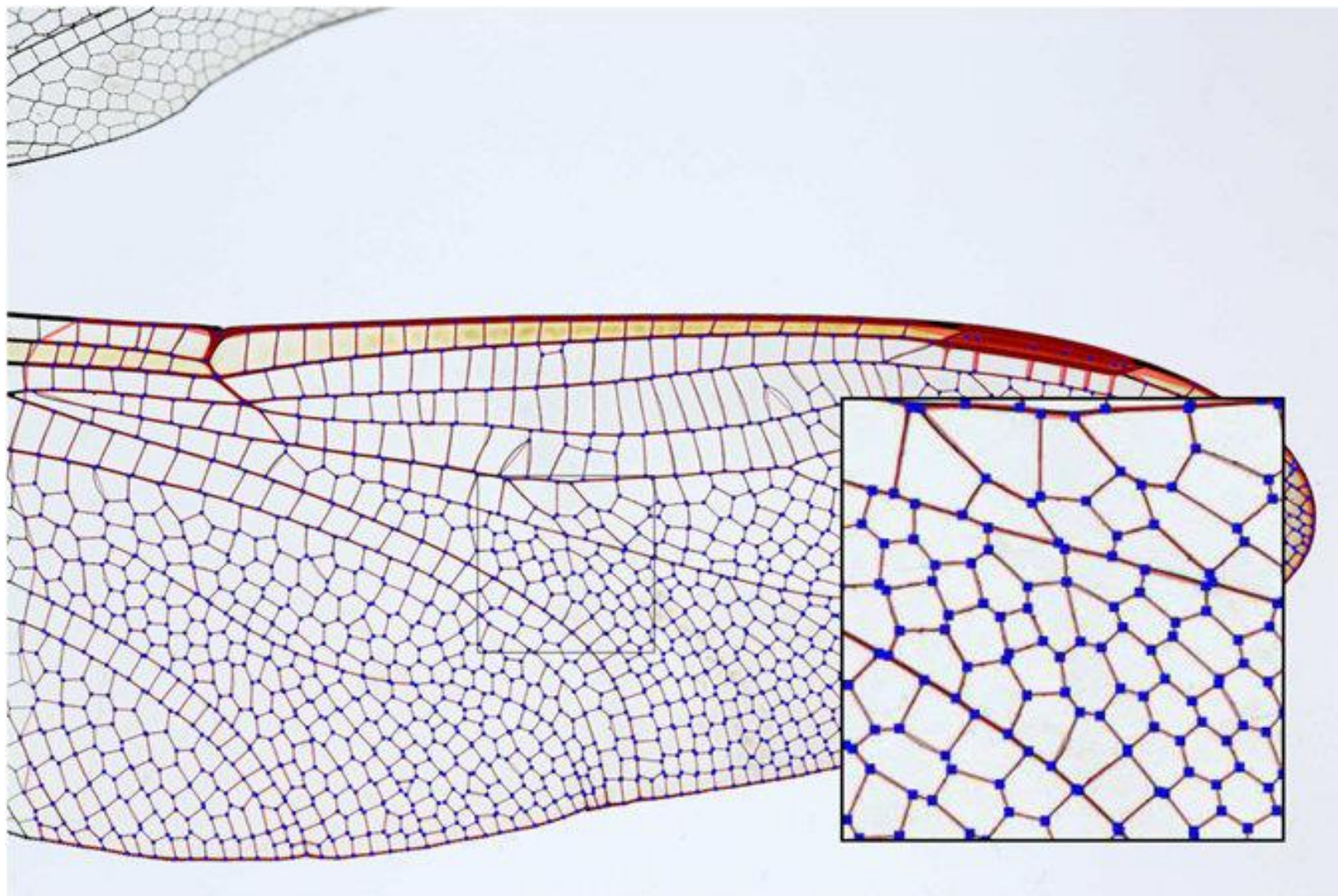
1



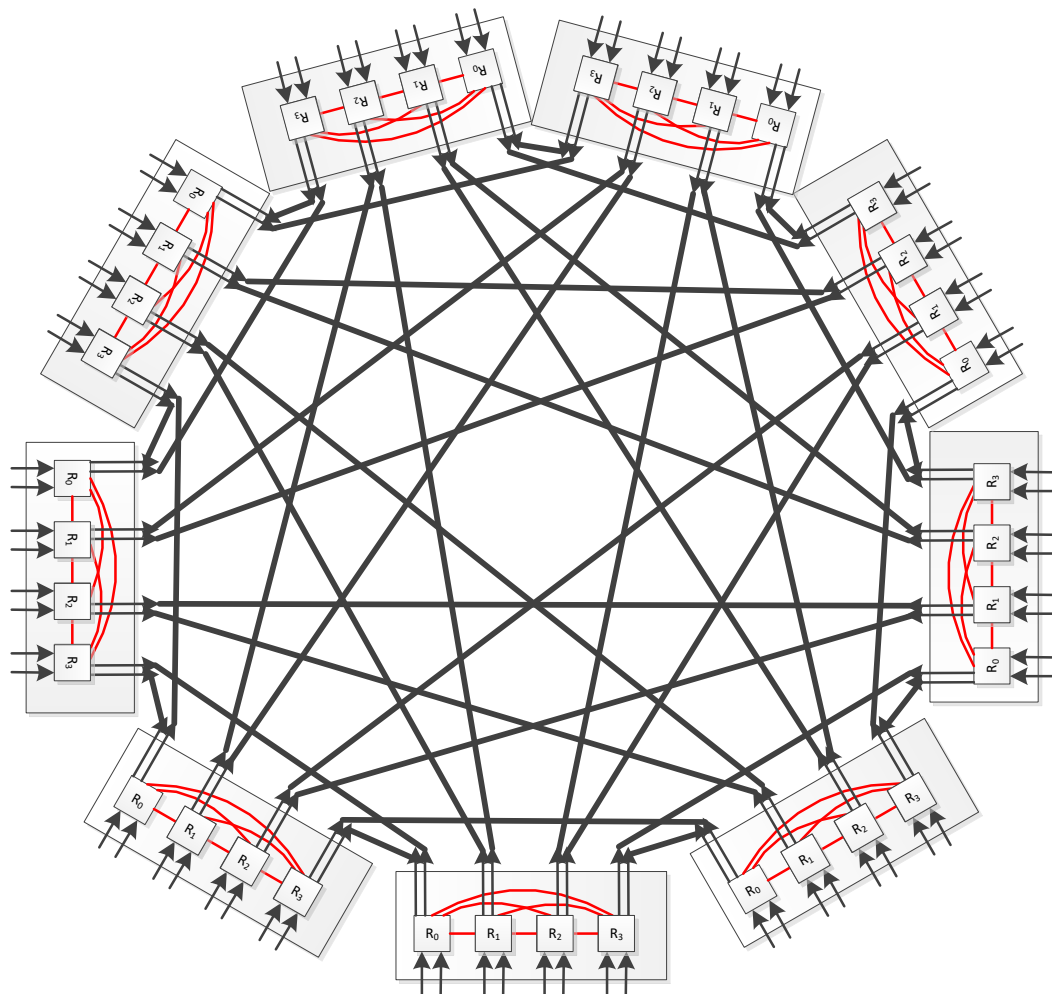
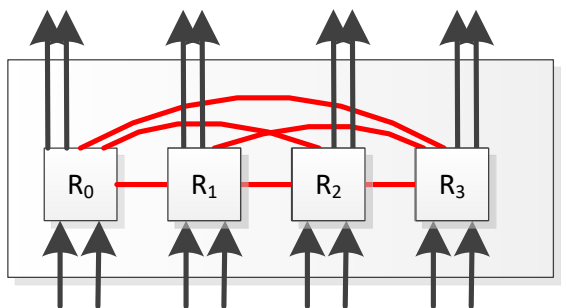
2



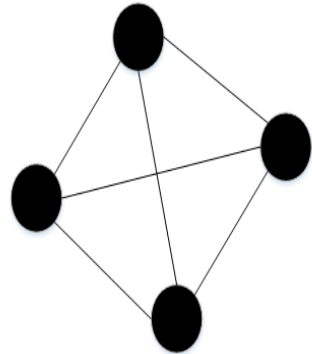
Топологии МВС. Dragonfly



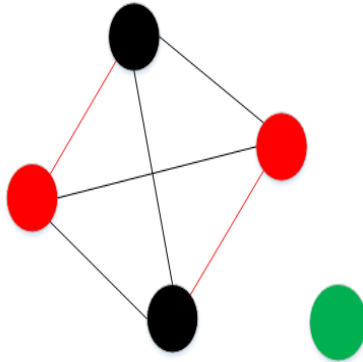
Топологии МВС. Dragonfly



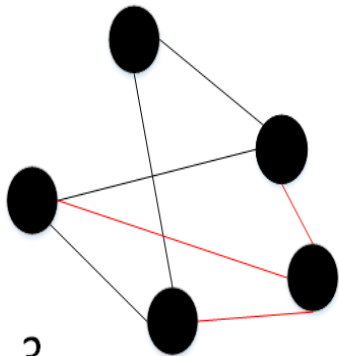
JellyFish



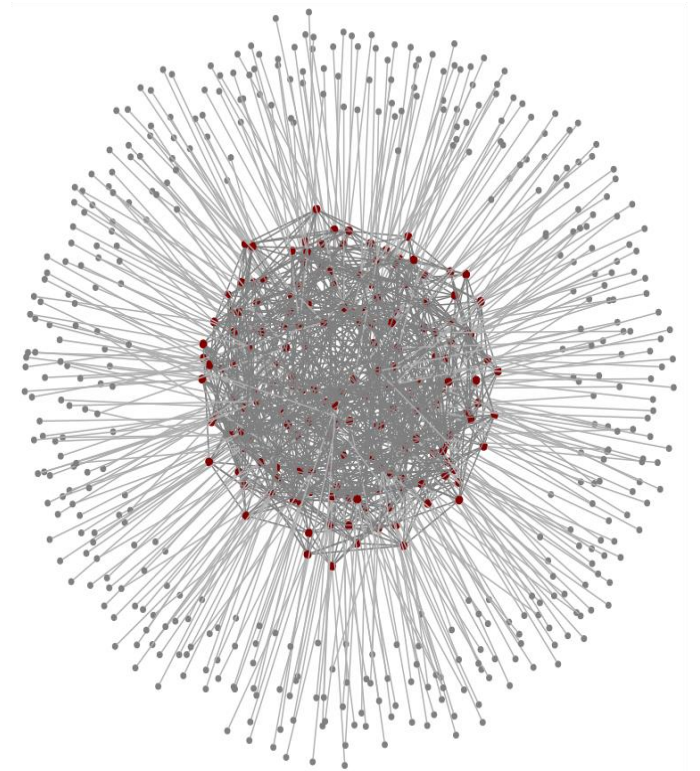
1



2



3



Специализированные процессоры

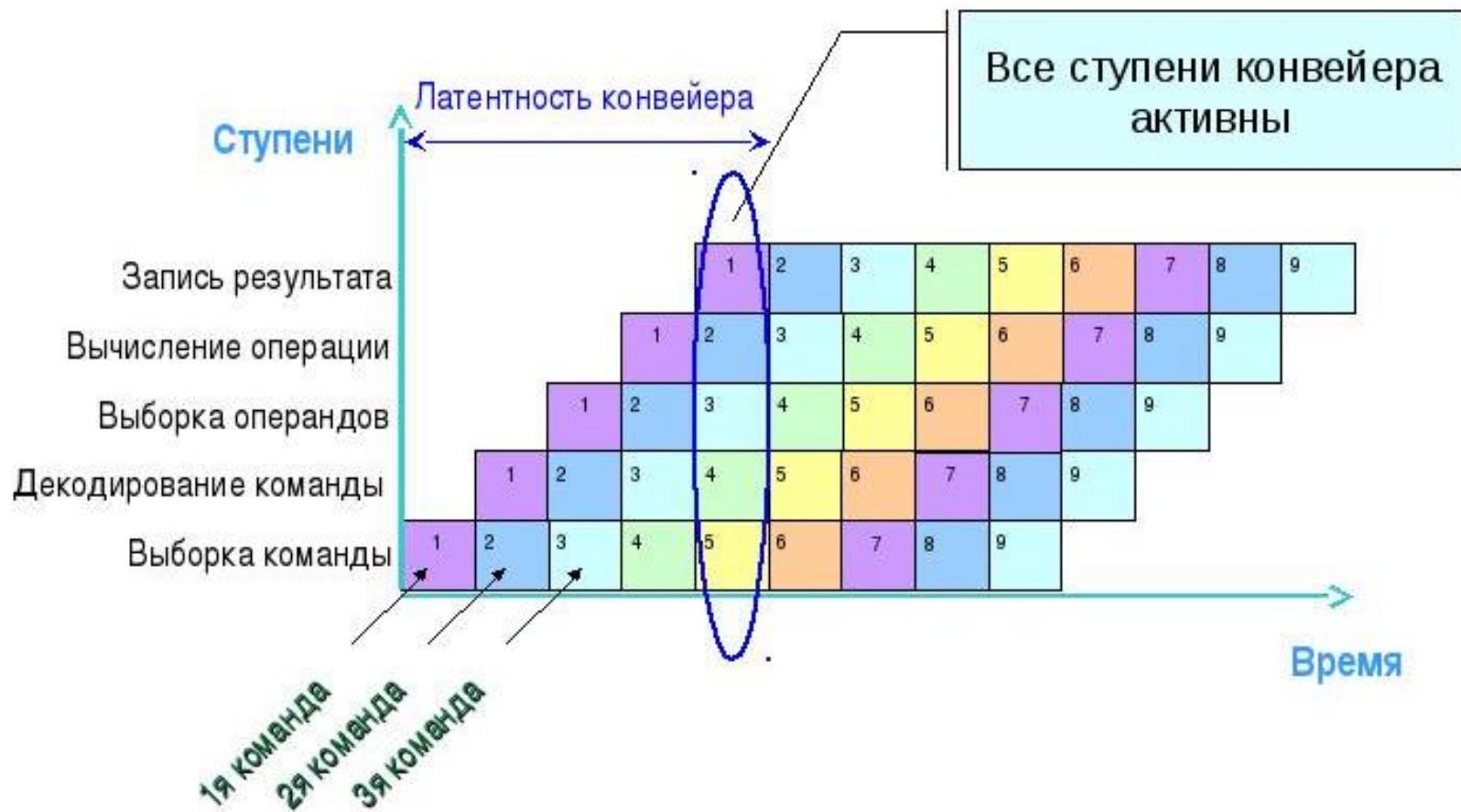
Конвейерные процессоры

- Разбивая последовательности логических элементов на более короткие и помещая триггеры между этими короткими последовательностями, уменьшают время, необходимое логике для обработки сигналов. В этом случае длительность одного такта процессора может быть соответственно уменьшена.

Команда разделяется на несколько этапов выполнения:

1. Выборка команды (*Instruction Fetch*)
2. Расшифровка команды (*Instruction Decode*)
3. Выборка необходимых операндов (*Register fetch*)
4. Выполнение команды (*Execute*)
5. Доступ к памяти (*Memory access*)
6. Запись в регистры (*Register write back*)

Работа конвейера



Коммуникационные процессоры

Среднее между специализированными интегральными микросхемами и гибкими процессорами общего назначения

Нужен для разгрузки основного процессора и оптимизации сетевой работы

Тенденция приближения к универсальному процессору

Матричные процессоры



Архитектура процессора Intel Haswell

ВУ

