Глава VI
Серверы и суперкомпьютеры

§1 Архитектуры параллельных компьютеров

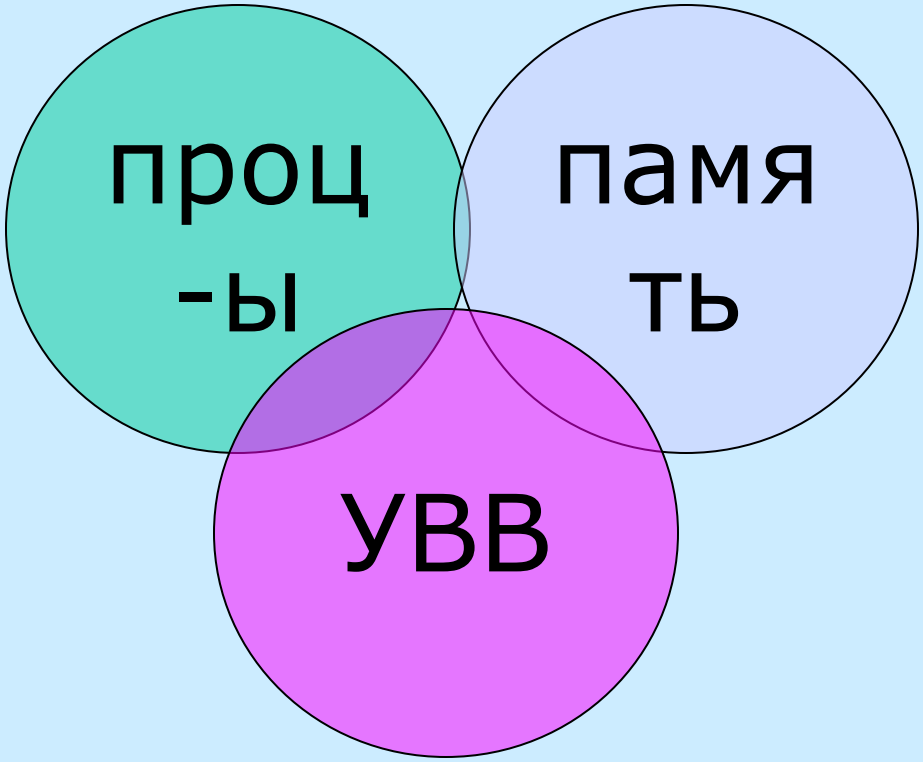
1. За счёт чего выросла производительность

модель	EDSAC (1949)	HP V-class (1999)
такт	2 мкс	2 нс
производ., оп/с	100	$7 \cdot 10^{10}$

- конвейер
- прогноз ветвлений и т. д.
- параллелизм

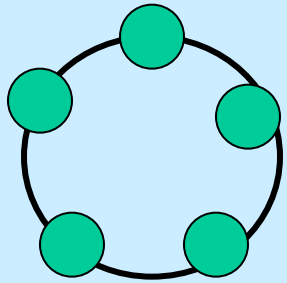
- много процессоров, банков памяти, УВВ
- внутри процессора: много конвейеров, кэшей, буферов

Основная проблема – <u>взаимодействие</u> паралл. работающих устройств



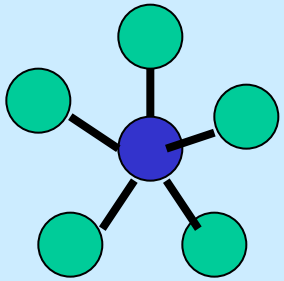
2. Топология

- диаметр – расстояние (в этапах) между наиболее удалёнными узлами

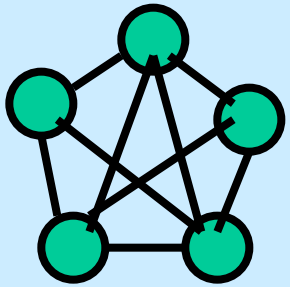


- размерность – число цепочек, пересекающихся в каждом узле

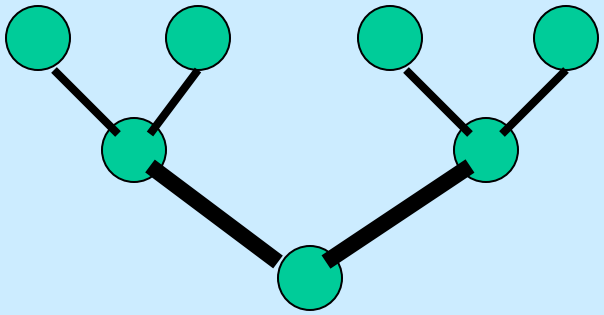
0:



- звезда

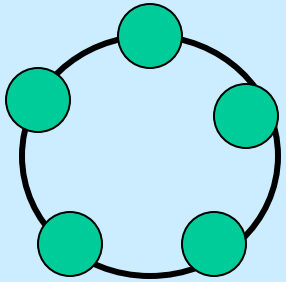


- полное межсоединение



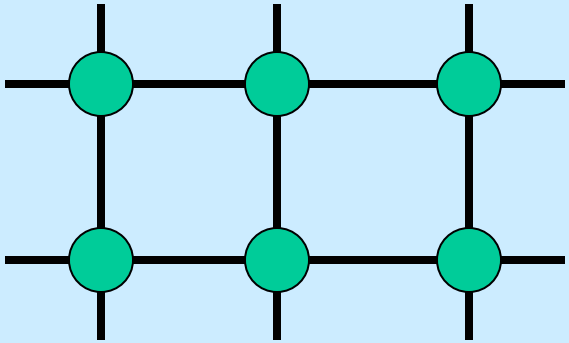
- толстое дерево

1:

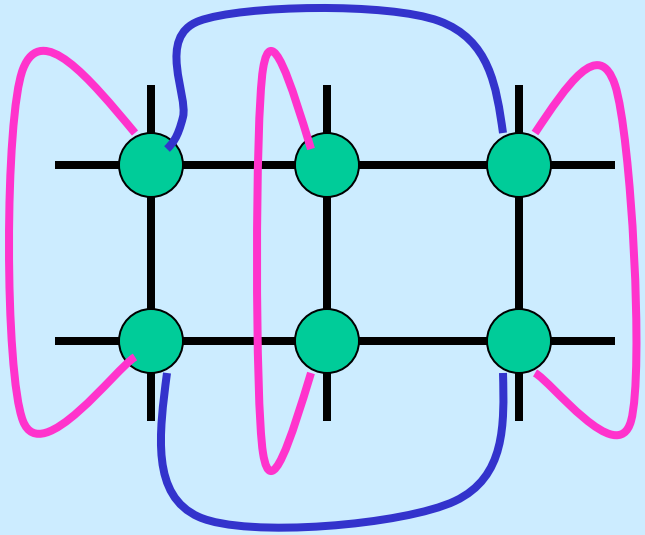


- кольцо

2:

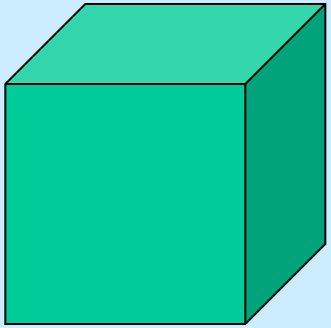


- решётка

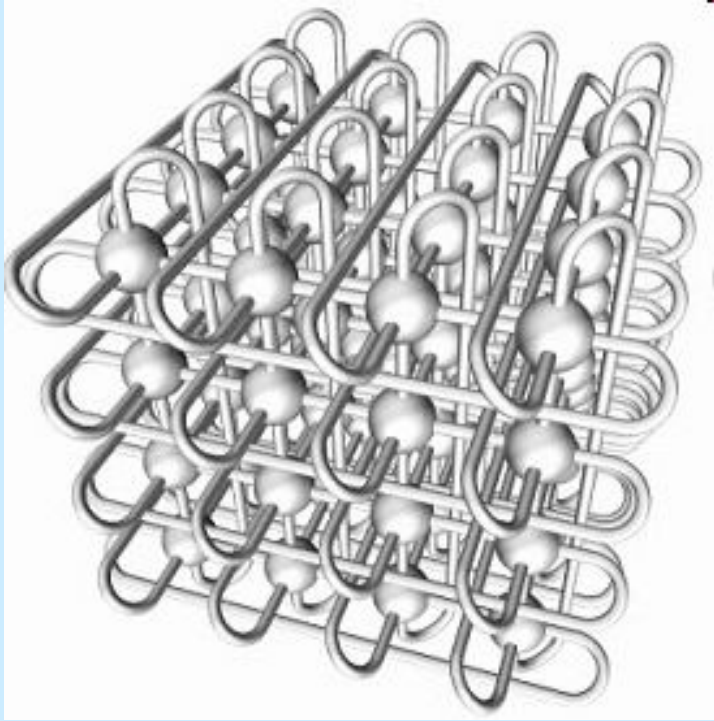


- 2М тор

3:



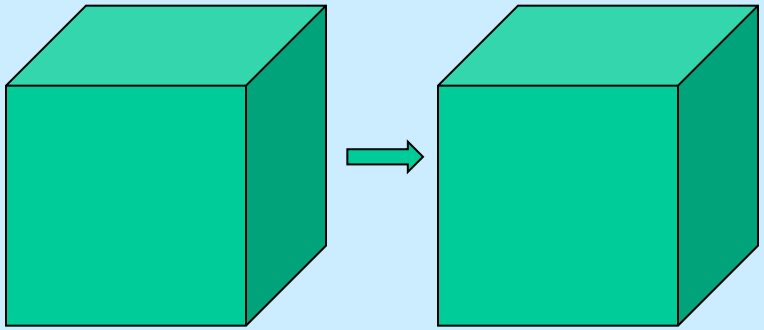
- куб

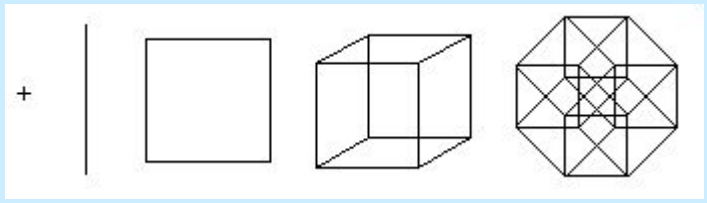


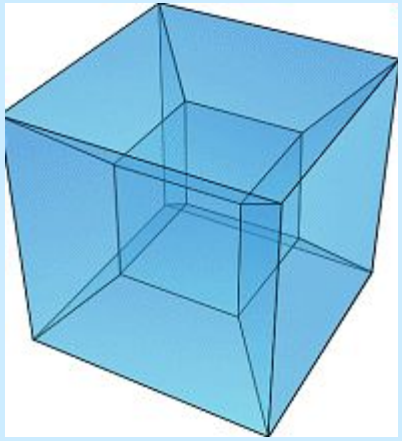
3М тор

4:

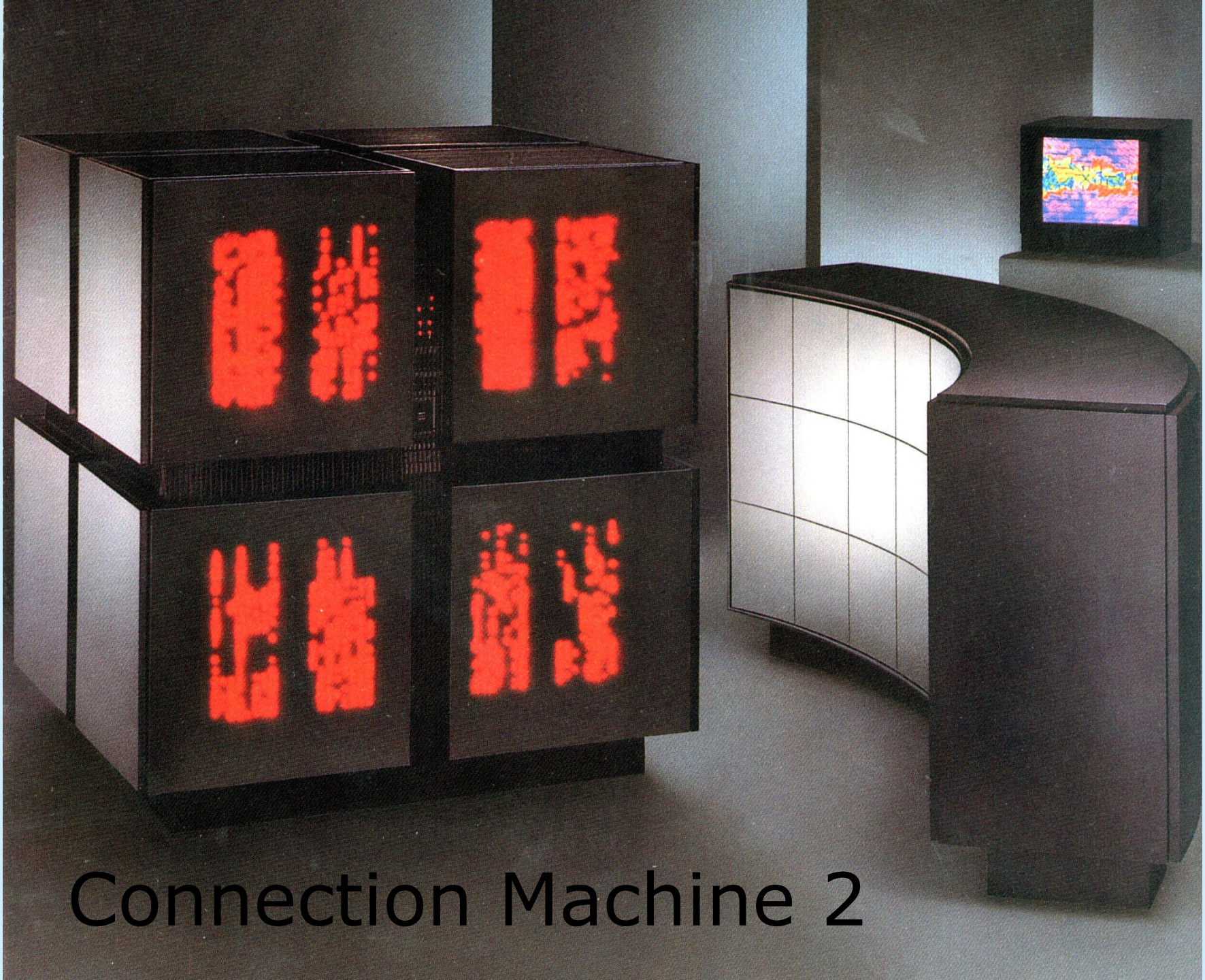
4М куб







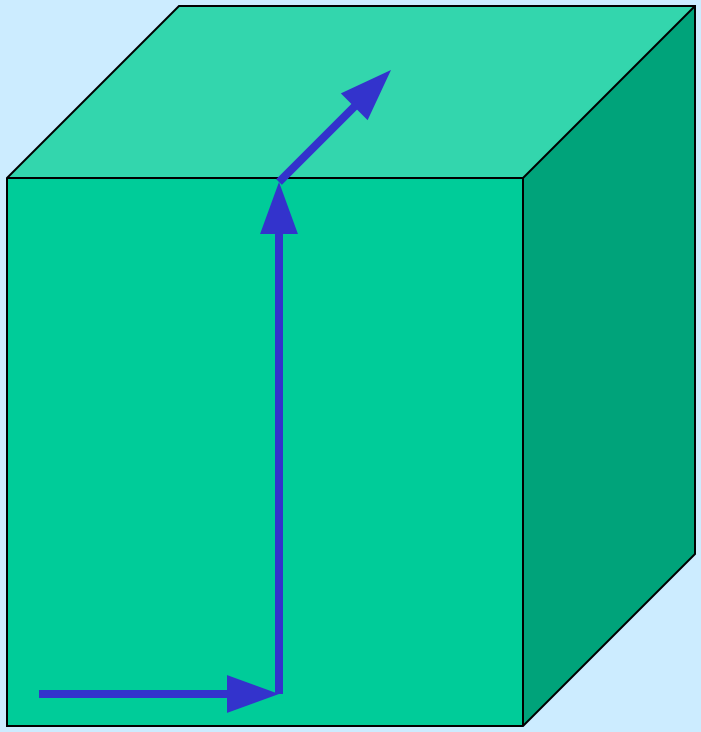
Чем больше размерность, тем меньше задержки, поскольку отношение диаметра к числу узлов уменьшается


Connection Machine 2

3. Маршрутизация

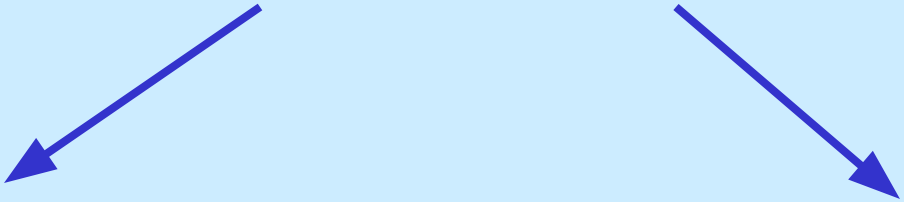
- от источника: источник определяет весь путь заранее и прикрепляет к пакету список номеров портов

- пространственная: по осям на нужное число узлов



не создаёт тупиковых ситуаций

4. Организация памяти



совместная:
единое
физическое
адресное
пространство

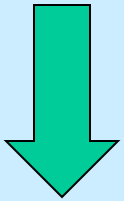
распределённая:
физически
раздельное,
логически единое

Обмен данными при распред. организации:

- проц. определяет, у кого есть нужные ему данные
- посылает запрос
- блокируется до получения ответа
- передача данных
- продолжение работы

Совместную память легко программировать, но трудно сделать (гигабайты)

Распределённую – наоборот



Комбинации

§2 Расширяемый связный интерфейс – РСИ

II

Scalable Coherent Interface – SCI

1. Назначение

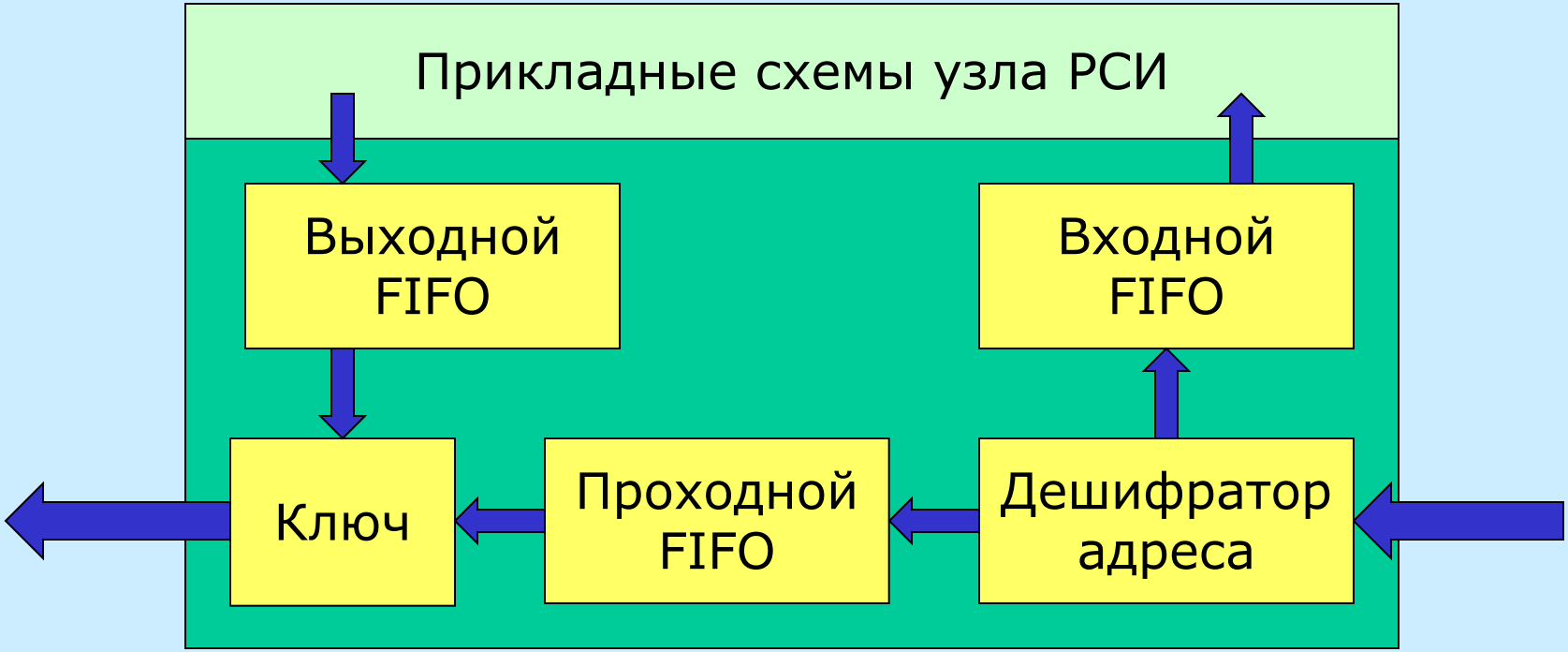
- суперкомпьютеры
- САУ (реального времени)
- сверхнадёжные компьютеры

Примеры:

- управление ядерным реактором
- крылатая ракета
- танк-робот
- комплекс ПВО
- прогноз погоды, землетрясений
- научные расчёты

2. Организация

а) Основной элемент РСИ – «узел»



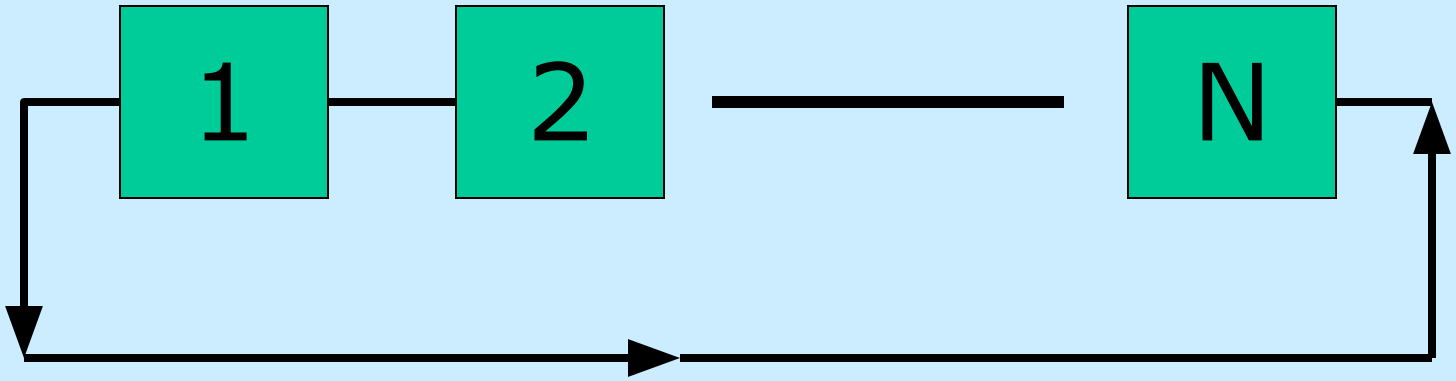
- Пакет поступает на дешифратор адреса

- Если адрес в пакете = адресу узла, то направляем пакет во входной буфер FIFO и затем на обработку

- Иначе пакет попадает в проходной FIFO и, если ключ открыт, выходит из узла

- ключ закрыт, когда прикладная схема выводит созданный ею пакет

б) Простейшая структура РСИ – «колечко»



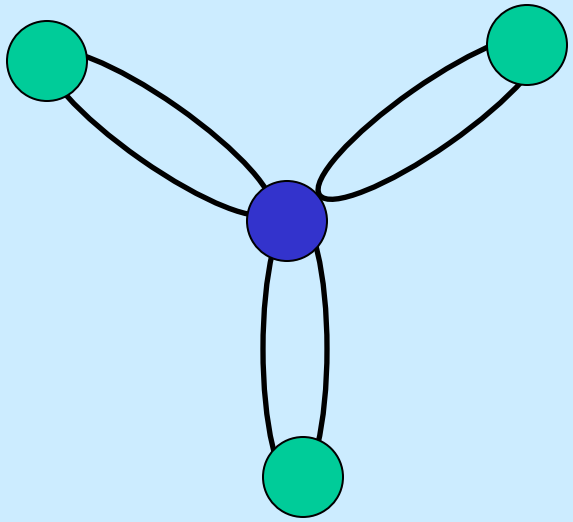
$N \in (2, 65536)$

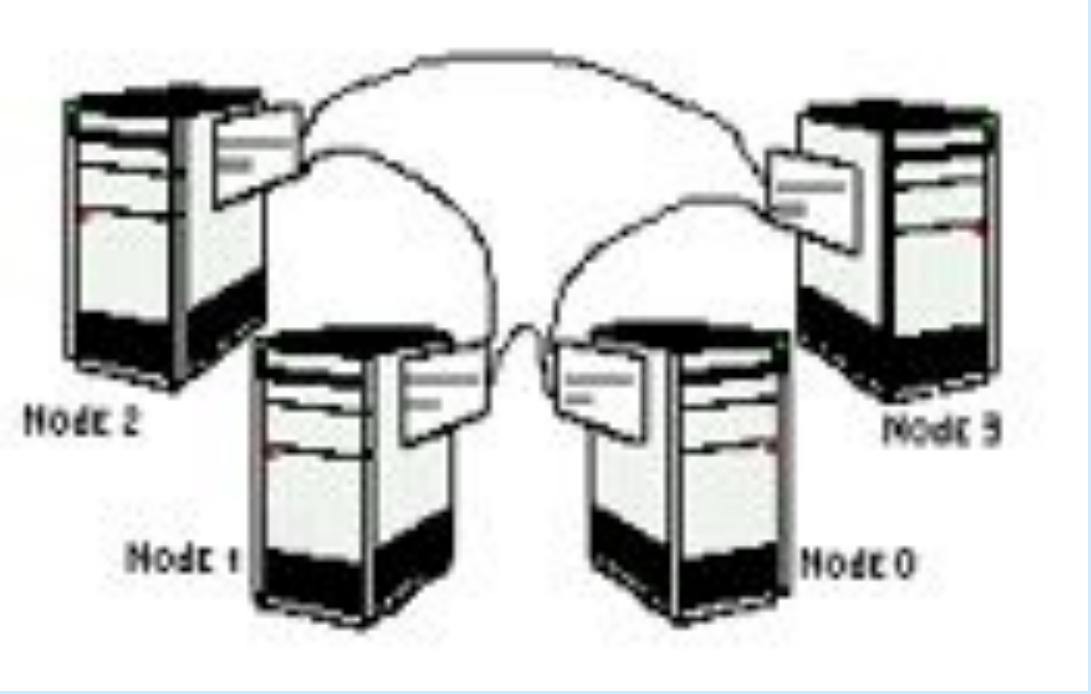
Пакеты бегут в одном направлении

Много узлов в колечке невыгодно

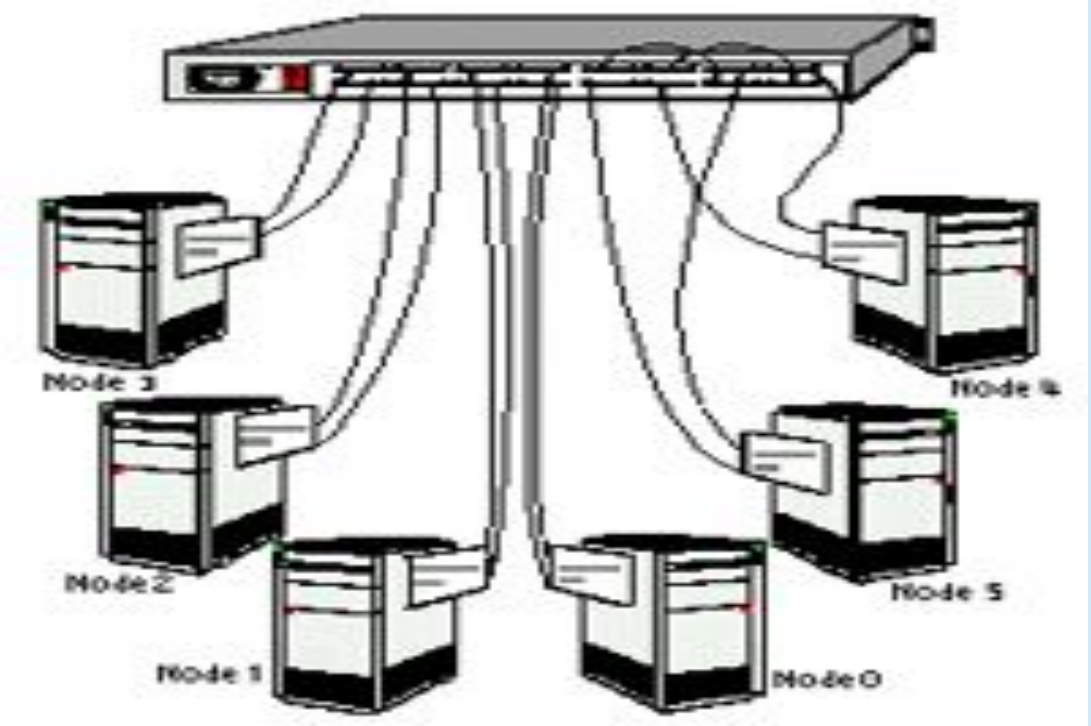
Большие системы состоят из колечек, связанных переключателями

Н-р, «звезда» с N=2



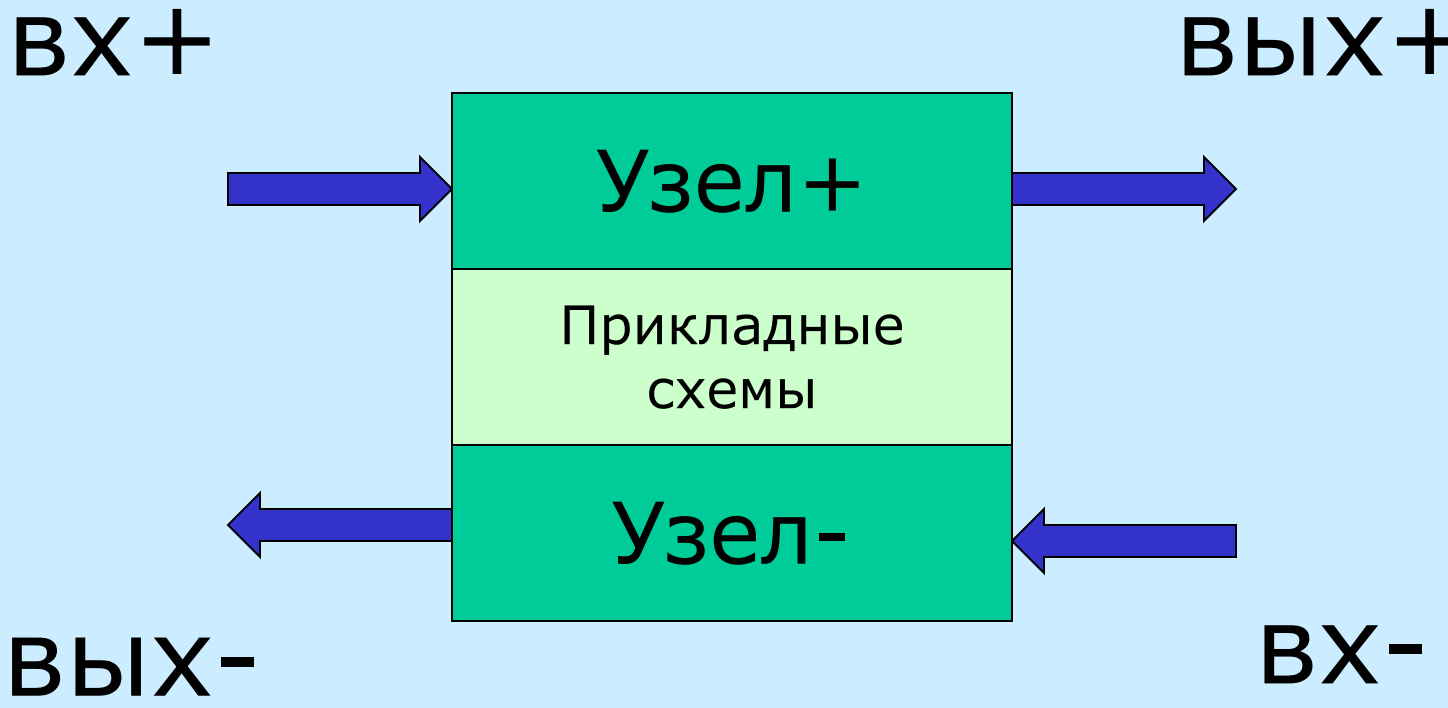


Колечко
N=4

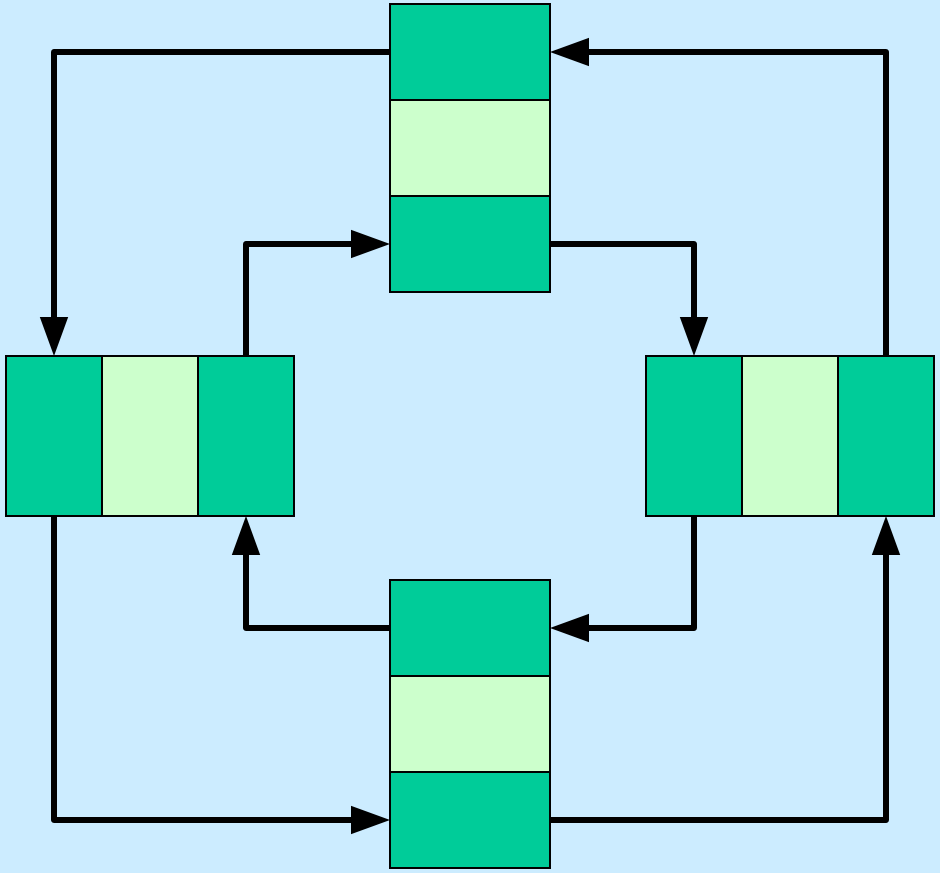


Звезда
N=2

в) Двойной узел

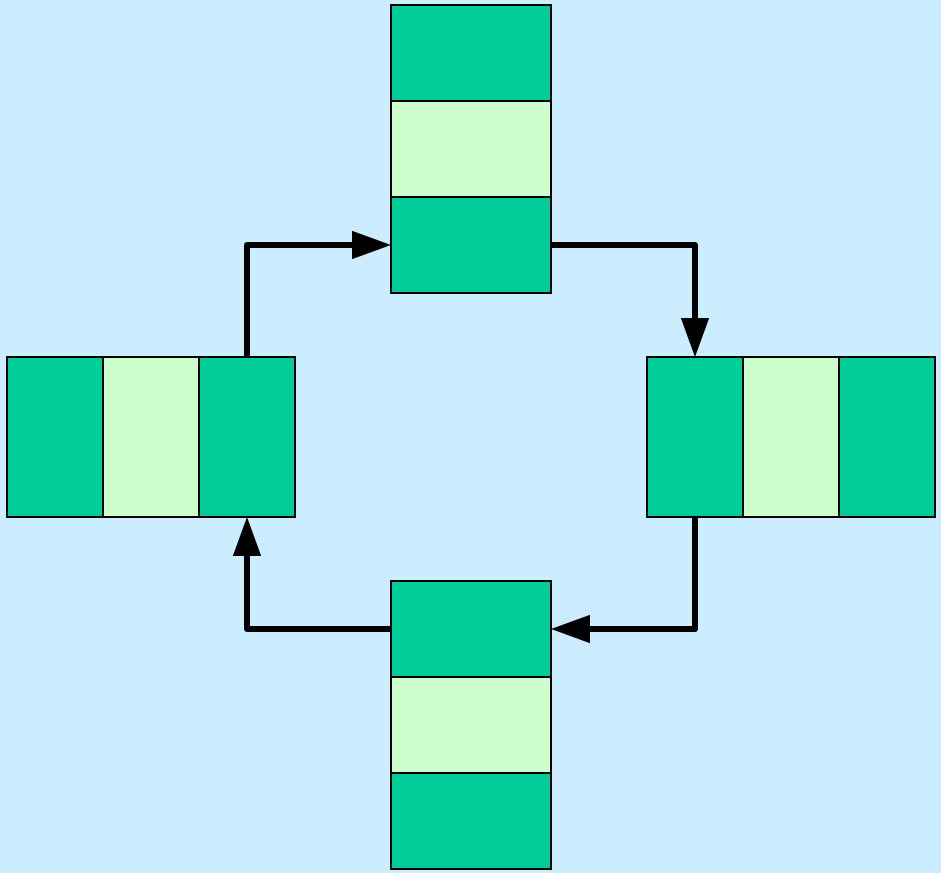


Из двойных узлов компонуют резервированные колечки «Гигаринг»

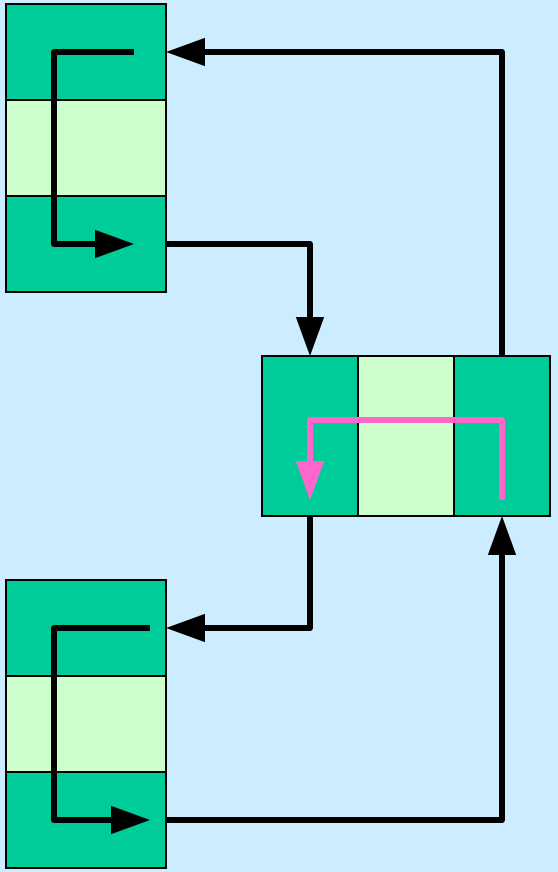


В случае разрыва одного колечка, работает оставшееся

Н-р:



Если разрушены один-два узла, то колечко просто укорачивается



Живучесть системы

г) Дворник колечка

- удаляет повреждённые пакеты
- управляет синхронизацией узлов
- полностью очищает колечко при крупных сбоях

д) Инициализация системы

- при включении питания каждый узел запускает свой тактовый генератор
- в каждом колечке избирается дворник
- он даёт узлам предварительные адреса

- программа высшего уровня активизирует переключатели между колечками
- затем присваивает каждому узлу уникальный адрес

3. InfiniBand

1x = 2 Гбит/с в каждом направлении

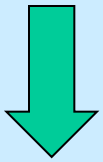
Наследник PCИ:

- ✔ обработка пакетов в узле
- ✔ менеджеры подсетей – дворники
- ✔ менеджер системы

Но колечки не обязательны

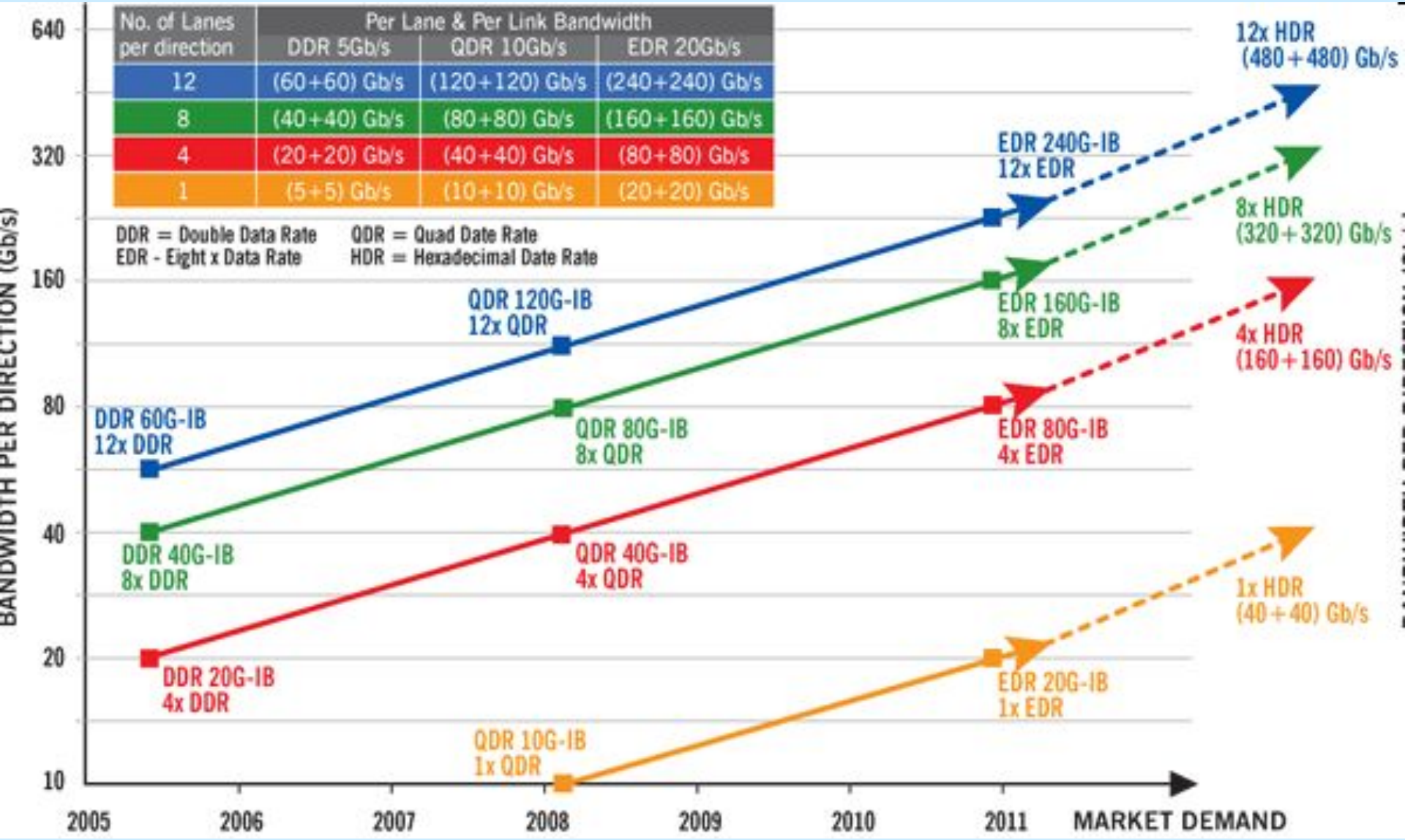
Схема кодирования 8B/10B:
8 разрядов данных + 2 разряда для синхронизации

Распараллеливание на уровне байтов



	Single (SDR)	Double (DDR)	Quad (QDR)
1X	2 Gbit/s	4 Gbit/s	8 Gbit/s
4X	8 Gbit/s	16 Gbit/s	32 Gbit/s
12X	24 Gbit/s	48 Gbit/s	96 Gbit/s

Пропускная способность

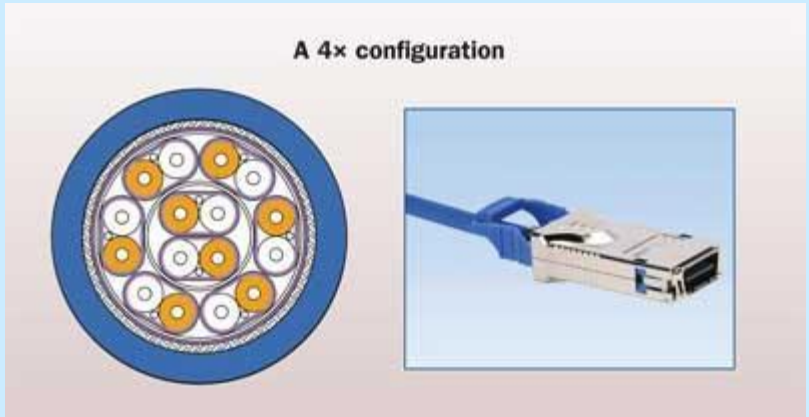


Медные кабели до 17 м

1x-1x

4x-1x

4x: 16 жил



Сетевая карта на 40 Гбит/с

Для PCI Express 2.0: (5 млрд. транзакций/с) ⇒ дуплексный обмен в MPI-приложениях ≈ 6460 МБ/с (по одному порту с задержкой не более 1 мс)

Оптоволокно: сотни метров

Тоньше, легче, «зеленее» (0.05 Вт)