

**Параллельное
программирование
для ресурсоёмких задач численного
моделирования в физике**

*В.О. Милицин, Д.Н. Янышев, И.А.
Буткарев*

Лекция № 2

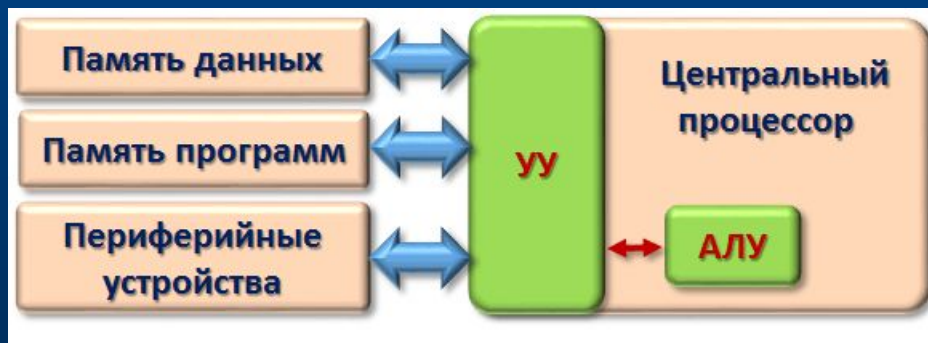
Содержание лекции

- Структуры аппаратного обеспечения
- Принципы построения мультипроцессорных систем
- Многообразии архитектур
- Используемые вычислительные ресурсы

Архитектура фон Неймана



Гарвардская архитектура



Основные архитектуры процессоров

- **CISC (Complex Instruction Set Computing)**
 - Нефиксированным значением длины команды
 - Исполнение операций, таких как загрузка в память, арифметические действия могут кодироваться в одной инструкции (есть микропрограммы)
 - Небольшим числом регистров, каждый из которых выполняет строго определенную функцию
- **RISC (Reduced Instruction Set Computing)**
вычисления с «сокращённым» набором команд
 - Фиксированная длина инструкций и простой формат команды
 - Спец. команды для операций с памятью — только чтения или записи
 - Большое количество регистров общего назначения
- **MISC (Minimal Instruction Set Computer)**
процессор, работающий с минимальным набором длинных команд.
Объединяет в себе суперскалярную и VLIW (Very Long Instruction Word) концепции
...
-

В современных процессорах используют гибридные архитектуры, например:
CISC-процессоры с RISC-ядром

Проблемы увеличения производительности процессора

- Размер кристалла (ограничения техпроцесса, синхронизация узлов)
- Тепловыделение ядра процессора (~ 4-й степени частоты)
- Высокая стоимость производства
- Малая результативность наращивания только тактовой частоты
- Несоответствующий росту производительности рост стоимости процессора

Многоядерные системы

Посмотрим на динамику CPU (Intel **Desktop** Processor):

- 2004 г. - Pentium 4, 3.4 GHz - 1 core
- 2005 г. - Pentium 4, 3.8 GHz - 1 core + HT
- 2005 г. - Pentium 4, 9.2 GHz - canceled
- 2006 г. - Core Duo T2700, 2.33 GHz - 2 core
- 2007 г. - Core 2 Duo E6850, 3 GHz - 2 core
- 2008 г. - Core 2 Quad Q9650, 3.00 GHz - 4 core
- 2009 г. - Core i7 975, 3.33 GHz - 4 core + HT
- 2011 г. - Core i7-990X, 3.46 GHz - 6 core + HT
- 2014 г. - Core i7-5960X, up to 3.5 GHz - 8 core + HT
- 2016 г. - Core i7-6950X, up to 3.5 GHz - 10 core + HT
- 2018 г. - Core i9-9980XE, up to 4.4 GHz - 18 core + HT
- 2019 г. - Core i9-10980XE, up to 4.6 GHz - 18 core + HT

(2019 г. Processor Xeon Platinum 9282M, 2.6 GHz (3.8 GHz) – 56 core + HT)

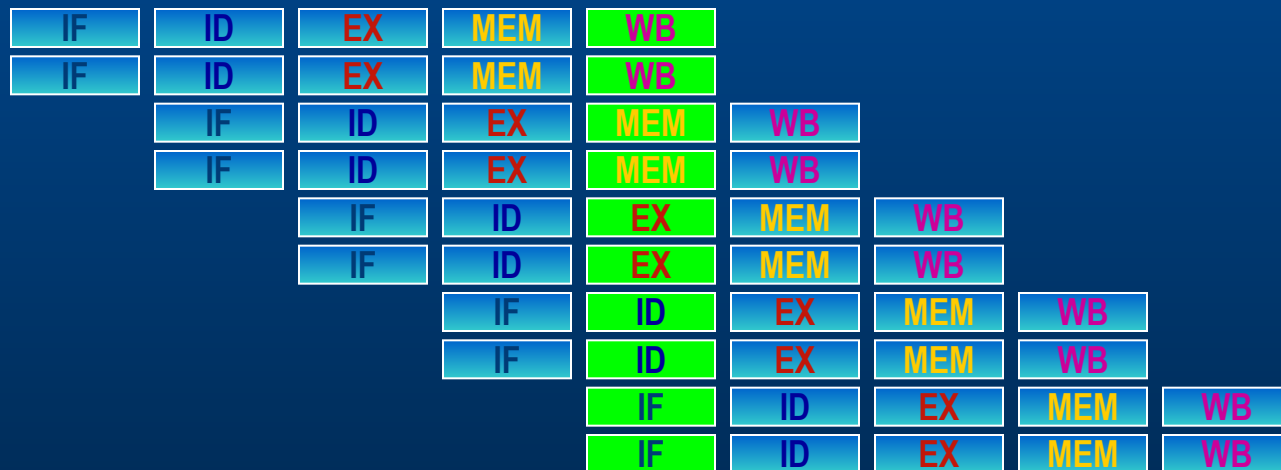
Повышение быстродействия

- ▶ Параллельность
 - Конвейер
 - Multithreading
 - SIMD/Векторизация (SSE / AVX ...)

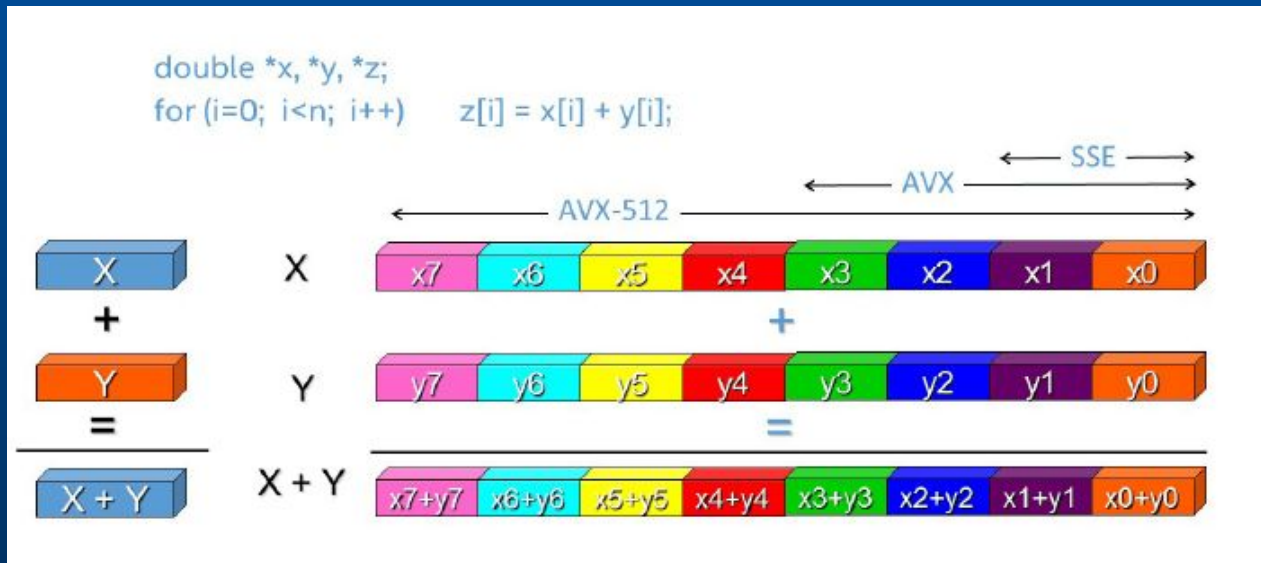
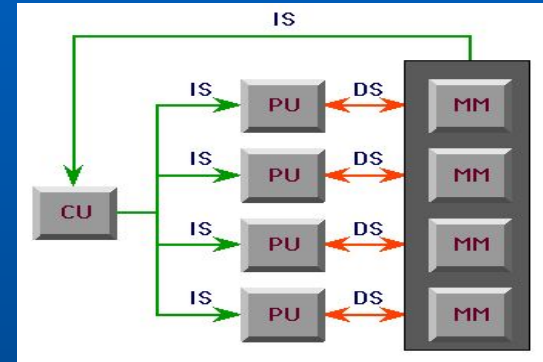
Конвейерная и суперскалярная обработка (Pipelining and Superscalar Execution)

Выполнение типичной команды:

- выборка команды – IF (Instruction fetch);
- декодирование команды/ выборка операндов из регистров – ID (Instruction decode)
- выполнение операции / вычисление эффективного адреса памяти – EX (Execute);
- обращение к памяти – MEM (Memory access);
- запоминание результата – WB (Register write back).



SIMD Extensions



Процессор, кэш и память в современной вычислительной машине



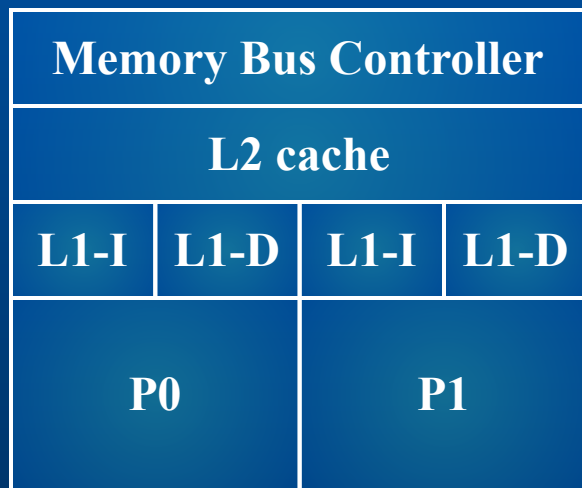
Данные

- ▶ *временная локальность*
- ▶ *пространственная локальность*

Доступ/размер

- ▶ CPU: *1 clk / ~10 register*
- ▶ L1 Cache: *1-2 clk / ~10 kB*
- ▶ L2 Cache: *~10 clk / ~1 MB*
- ▶ RAM: *50-100 clk / ~1GB*

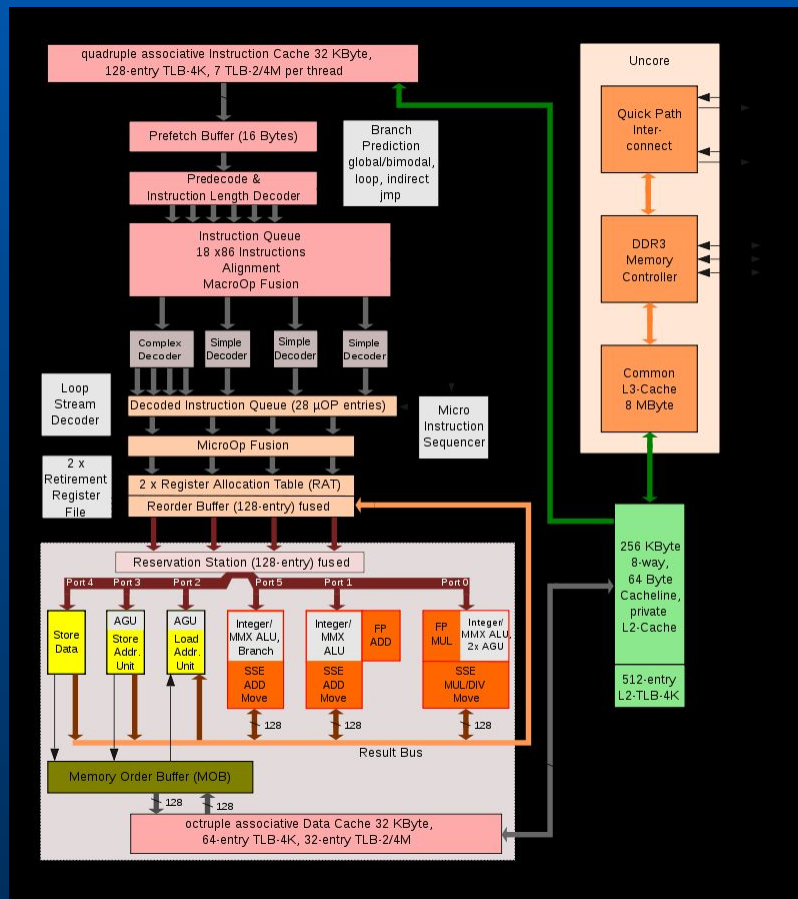
Многоядерные архитектуры



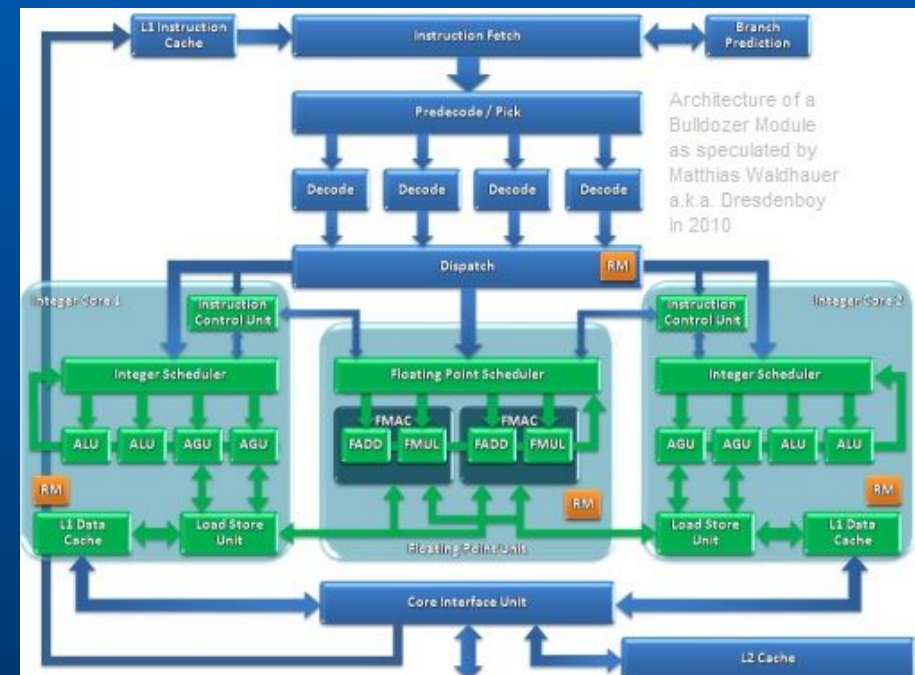
- ▶ 10-100 Кб L1 кэш для каждого ядра
- ▶ 1-10 Мб общий L2 кэш
- ▶ Единый образ памяти для каждого ядра - необходимость синхронизации кэшей

Примеры блок-схем процессоров

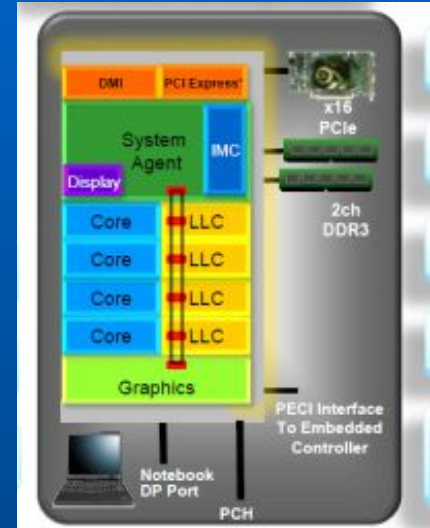
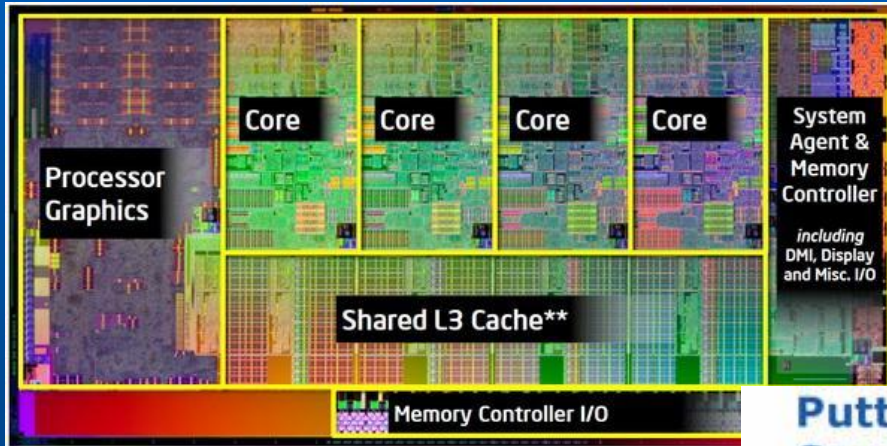
Intel Nehalem



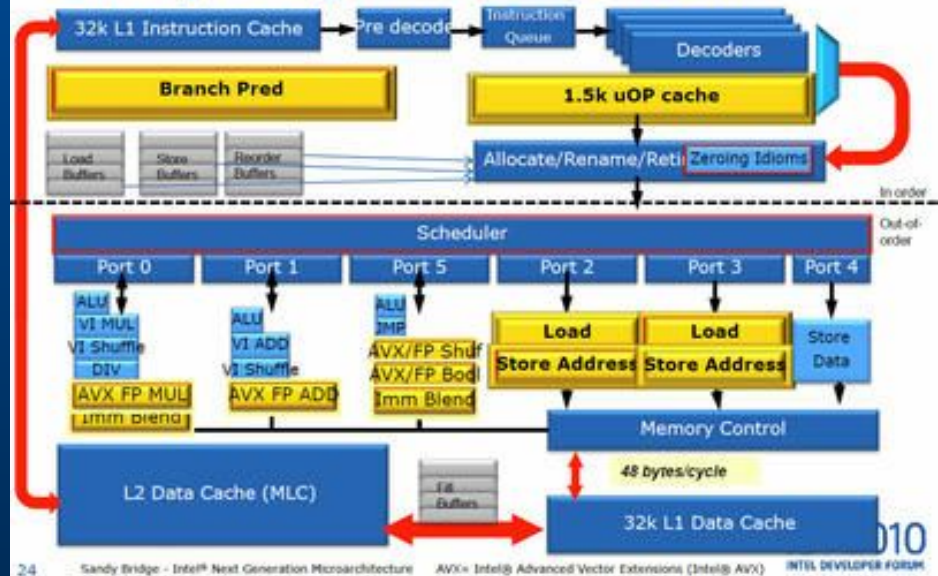
AMD Bulldozer



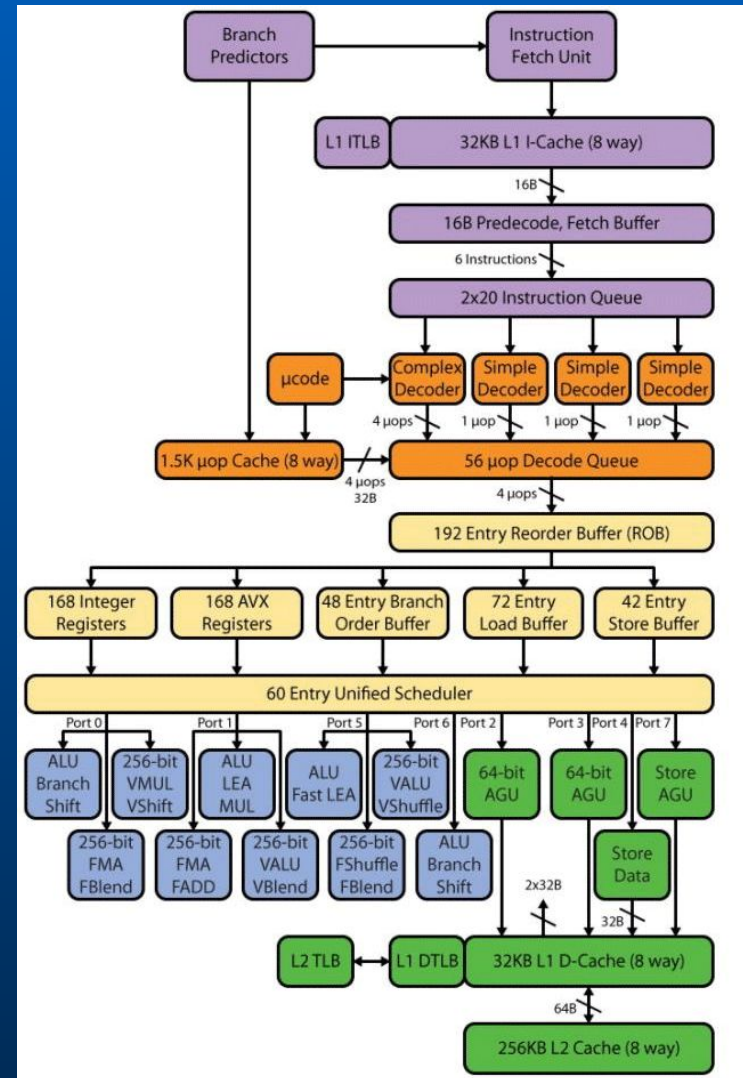
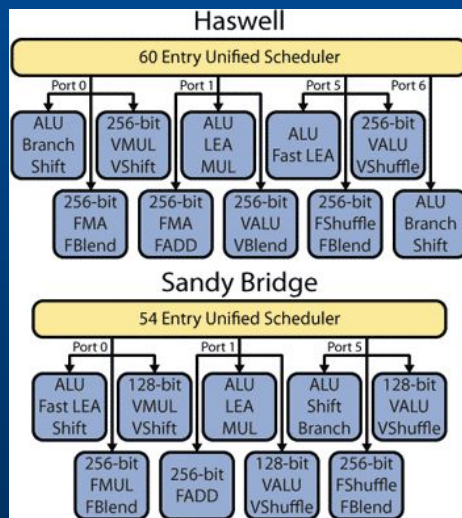
Sandy Bridge



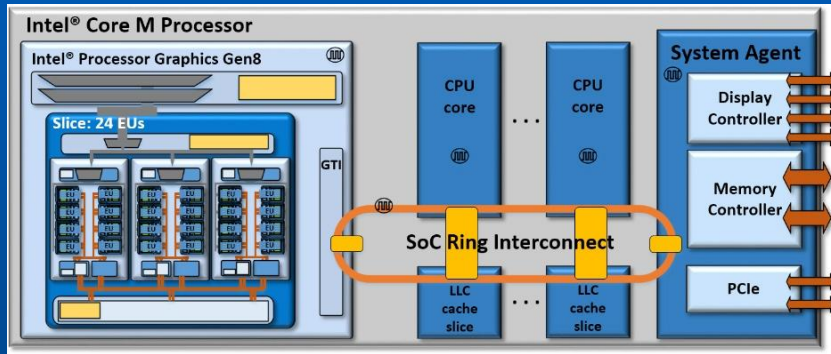
Putting it together Sandy Bridge Microarchitecture



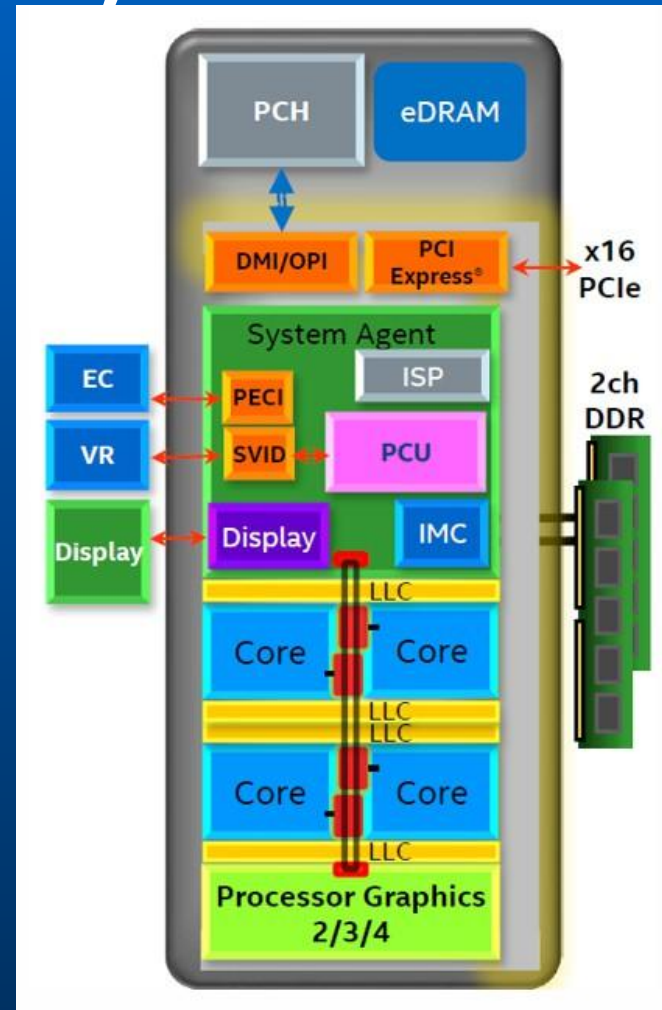
Intel Haswell



Intel Skylake (Kaby Lake)



Расширение общей структуры
(SoC, System on Chip)



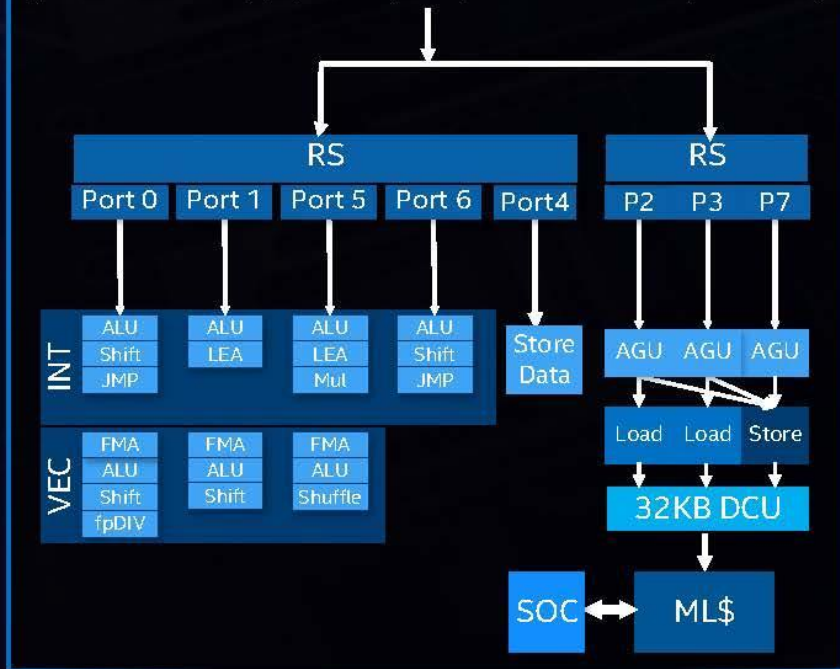
Haswell	Broadwell	Skylake
~500 GFLOPS [DP-F.P. peak]	tbd GFLOPS [DP-F.P. peak]	tbd GFLOPS [DP-F.P. peak]
22nm AVX2 DDR4 PCIe3	14nm	14nm AVX3.2 DDR4 PCIe4

Intel Sunny Cove (Ice Lake)

SKYLAKE REMINDER

Front End

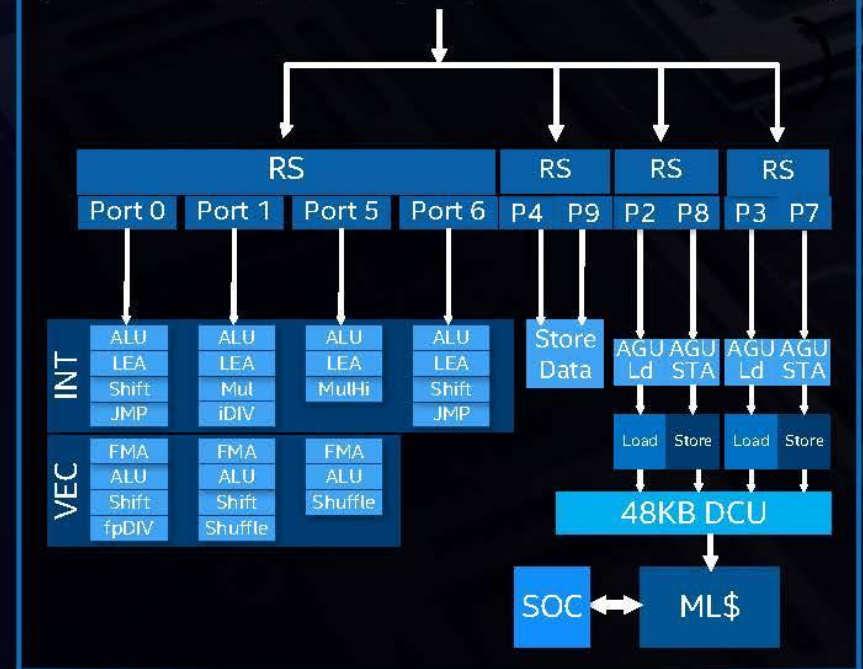
(Branch Predictor, μ op cache, ITLB, Instruction Cache, Decoders)



INTRODUCING SUNNY COVE

Front End

(Branch Predictor, μ op cache, ITLB, Instruction Cache, Decoders)



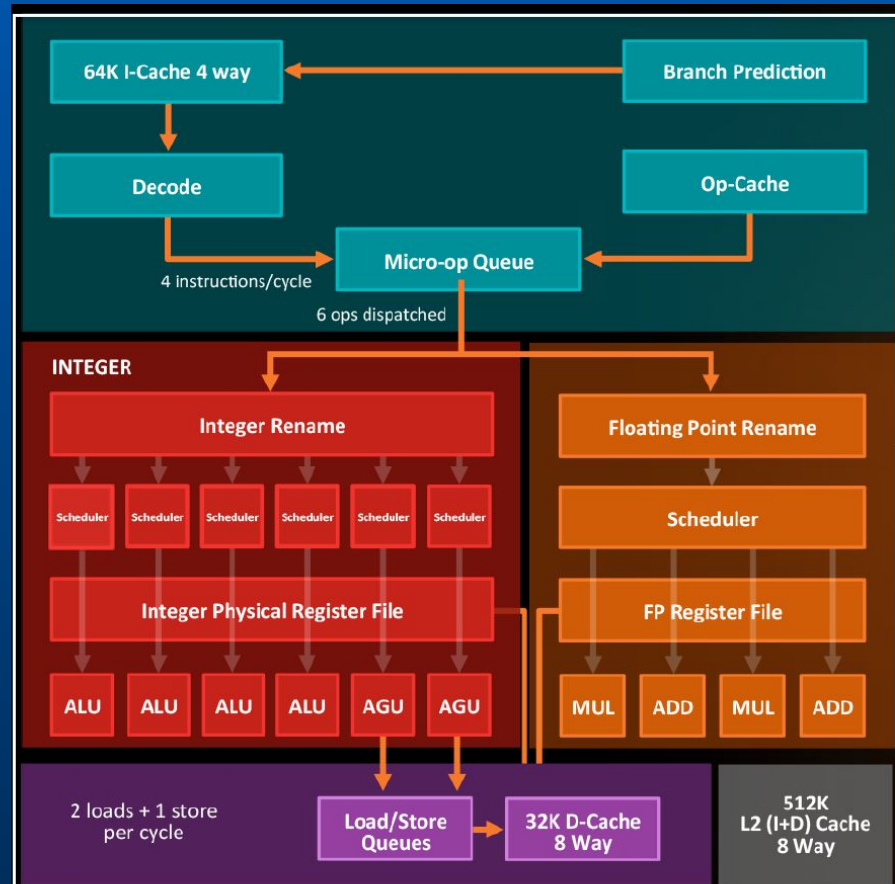
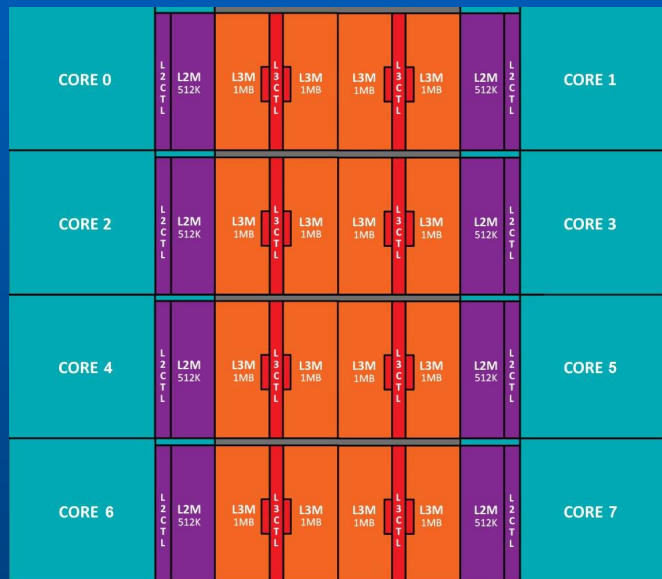
Intel® Confidential. Under embargo.

ARCHITECTURE DAY

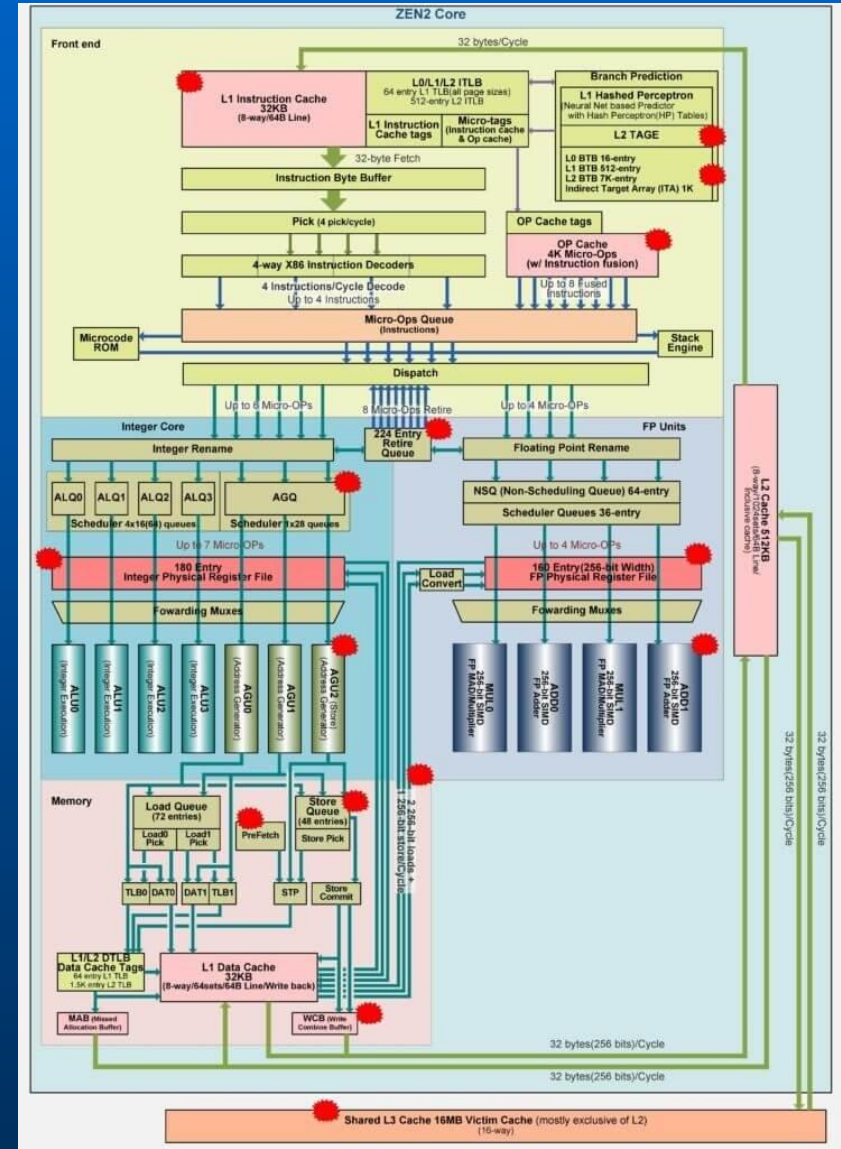
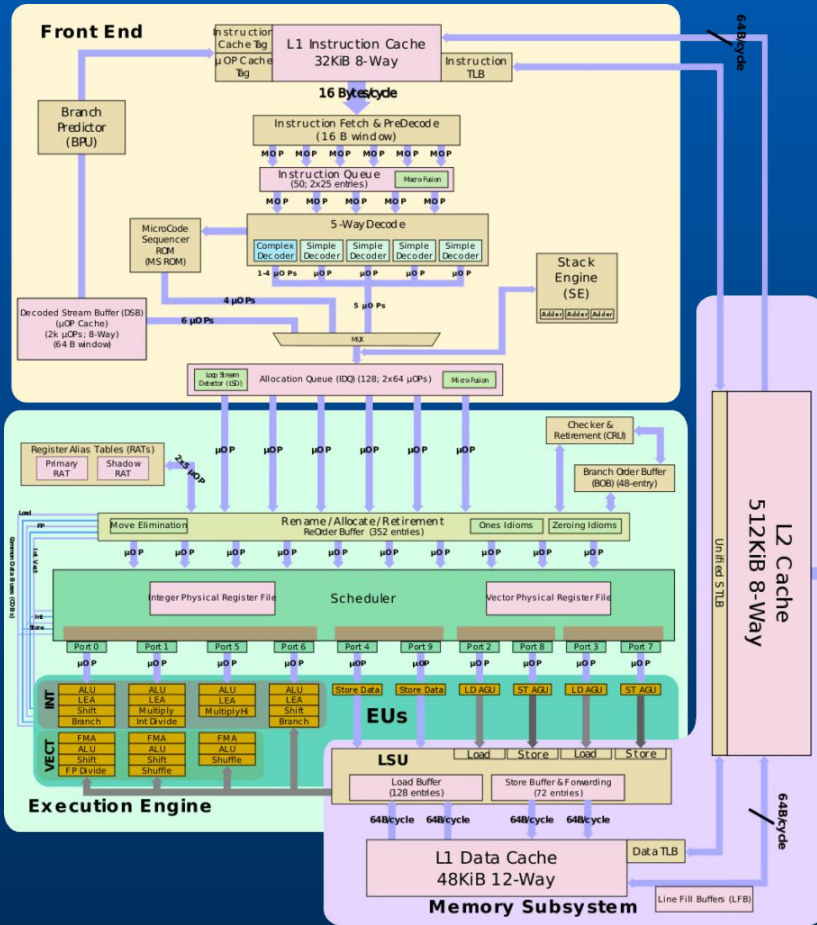


10

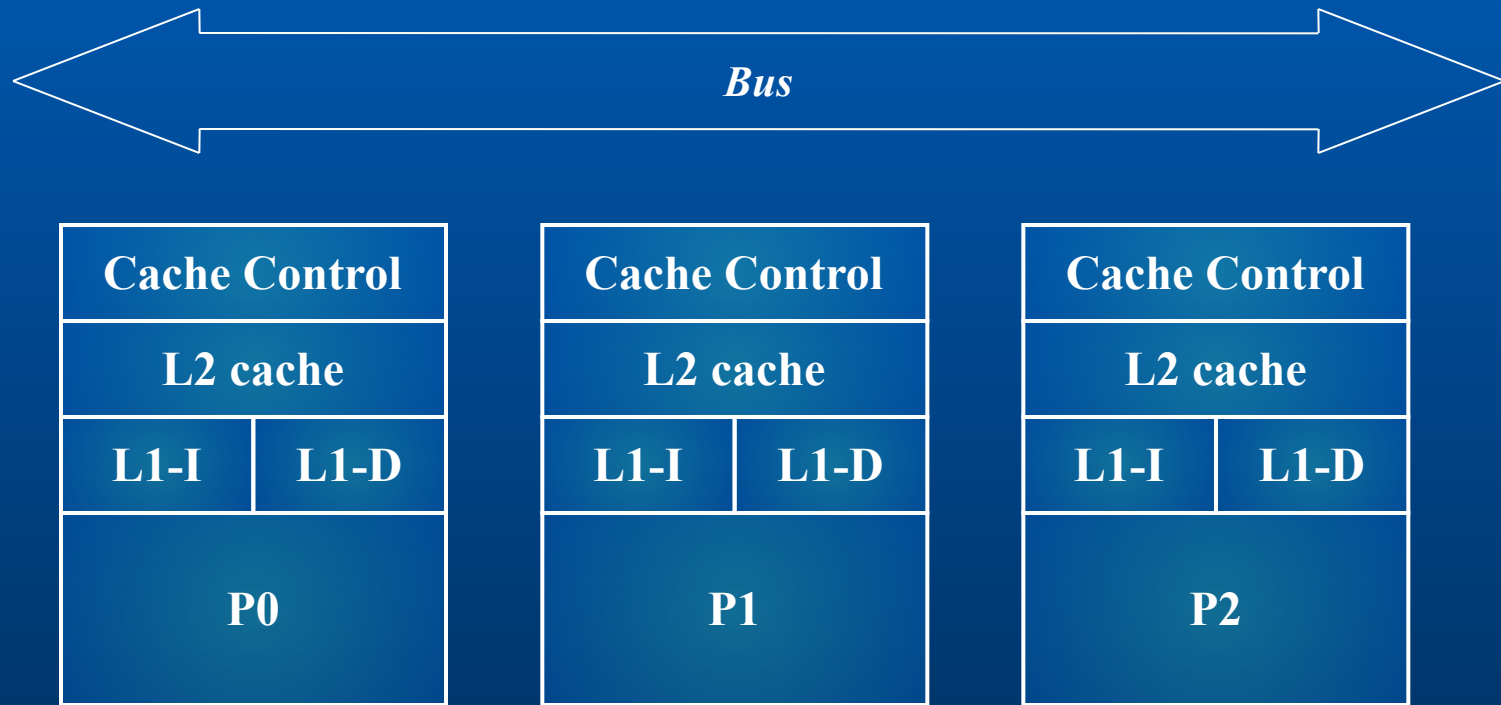
AMD Ryzen/Zen



Intel vs AMD



Symmetric Multiprocessor Architecture (SMP)

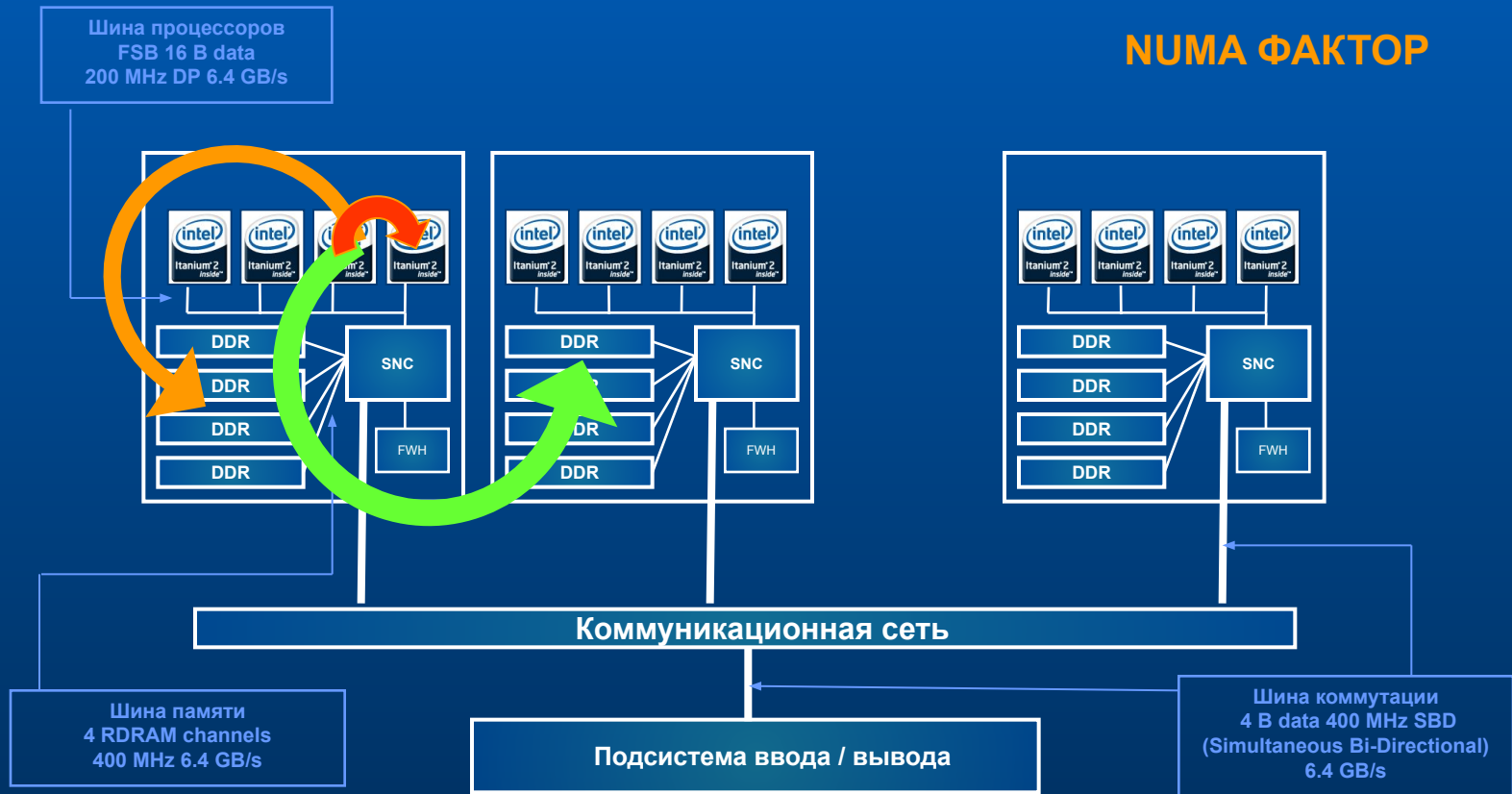


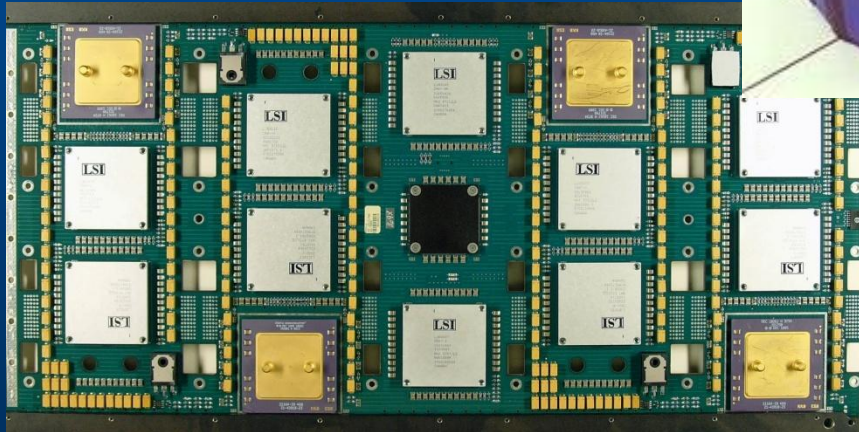
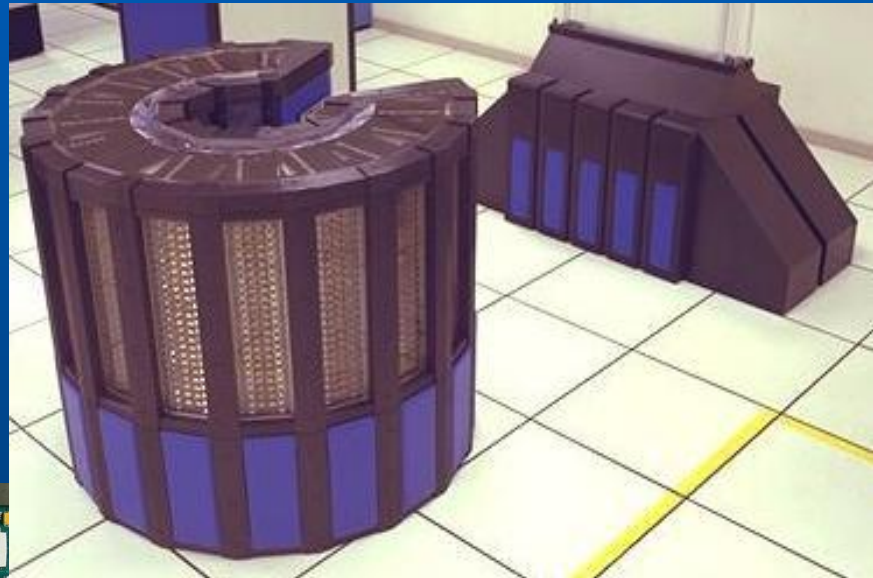
Multi processor motherboard



ccNUMA.

Интервалы при обращении к памяти

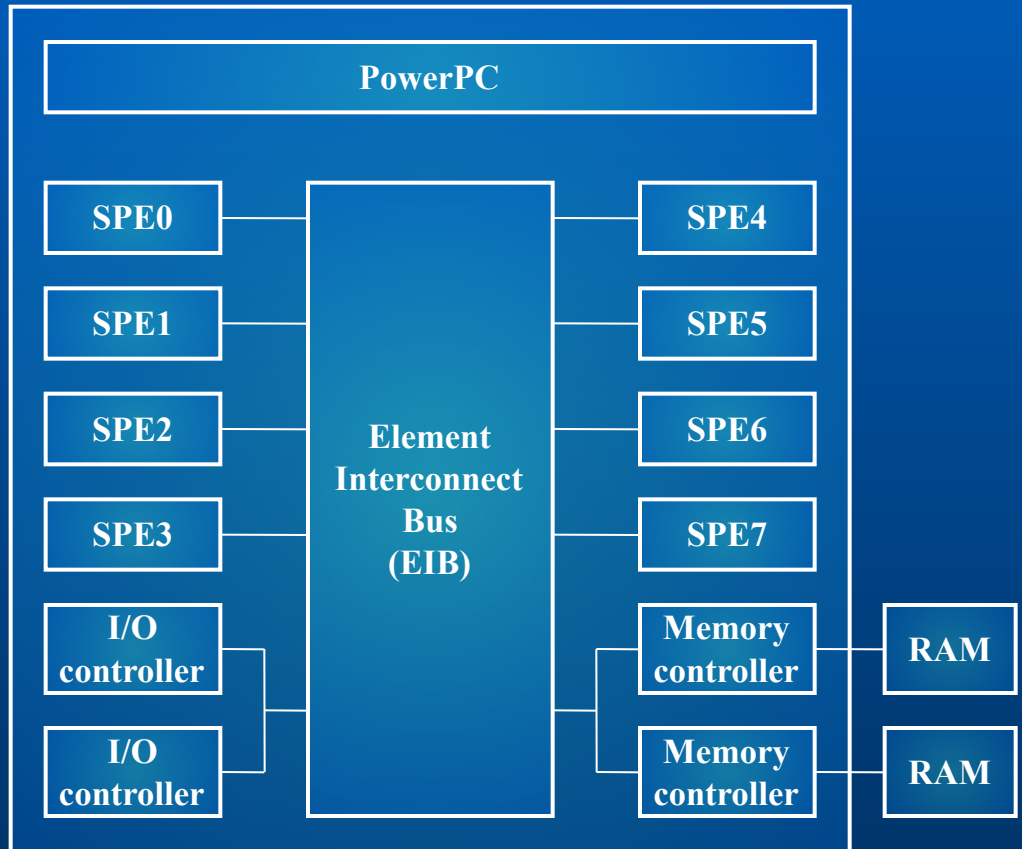




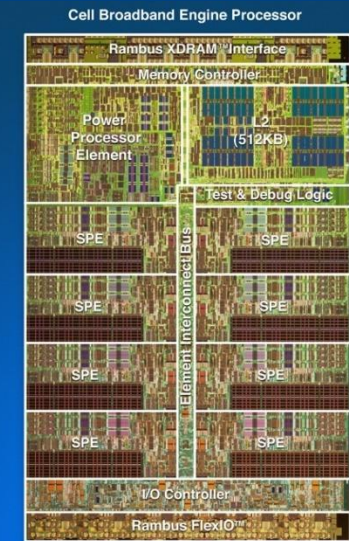
Многообразие архитектур ЭВМ



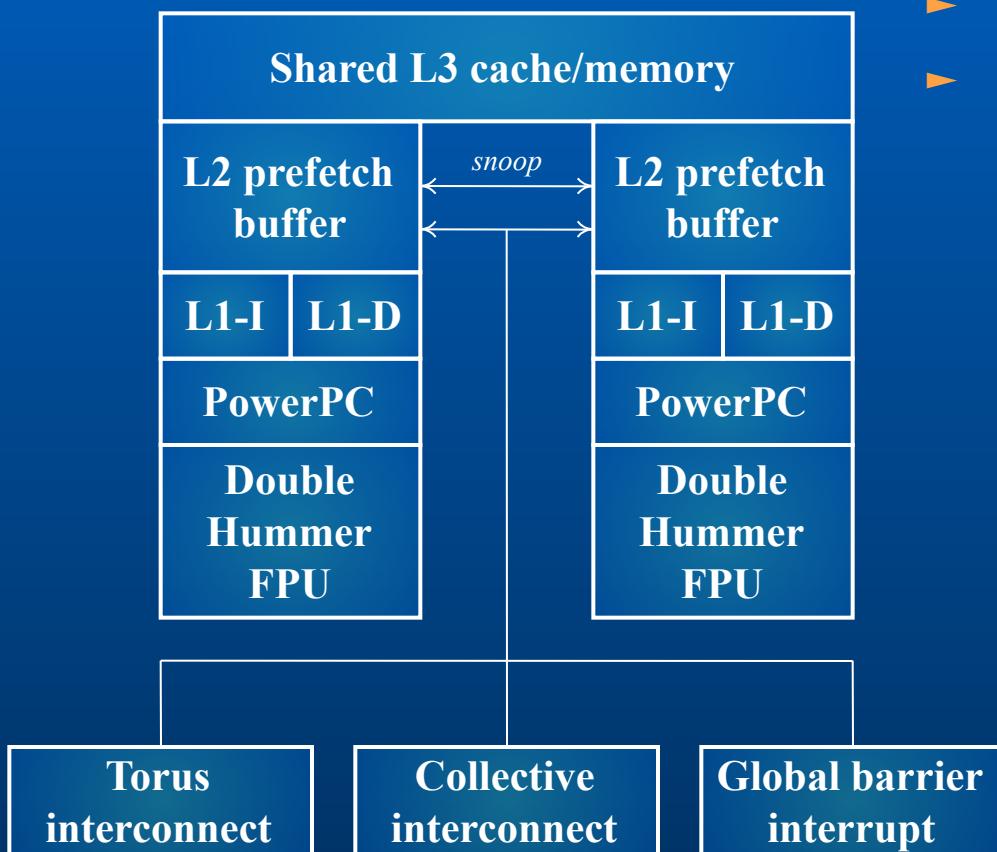
Cell



- ▶ Dual-threaded 64-bit PowerPC
- ▶ 8 Synergistic Processing Elements (SPE)
- ▶ 256 Kb on-chip/SPE

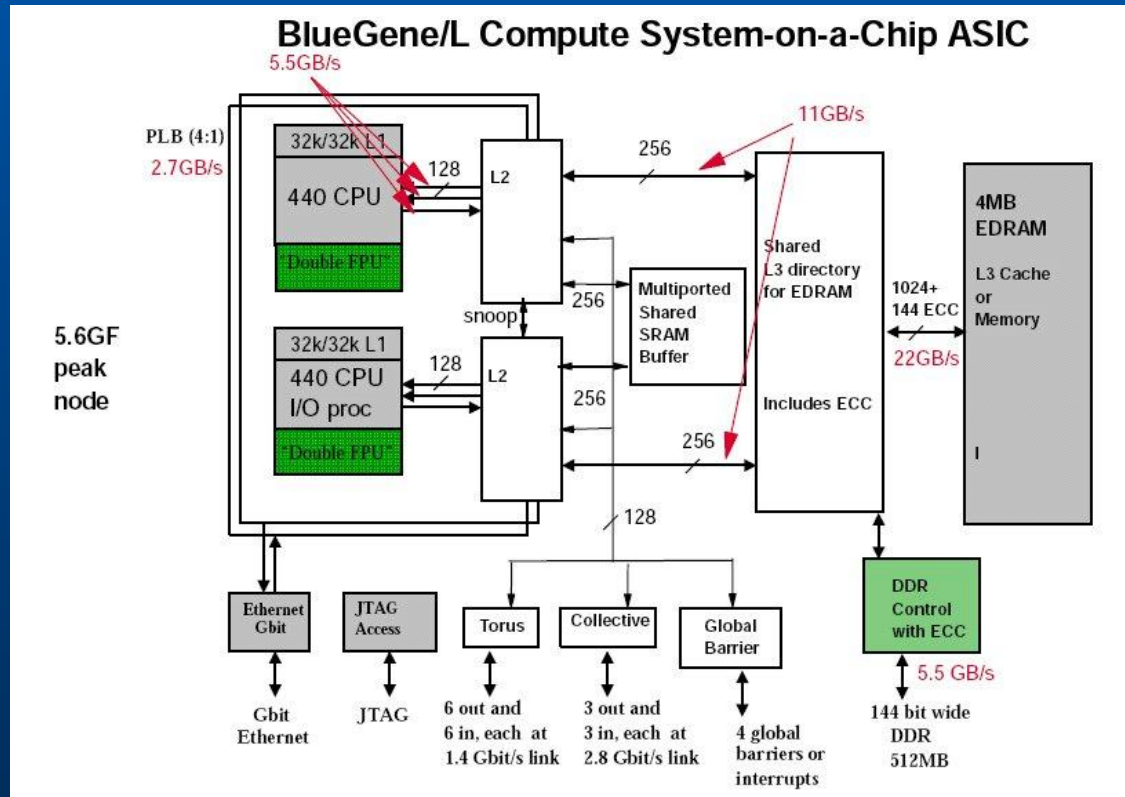
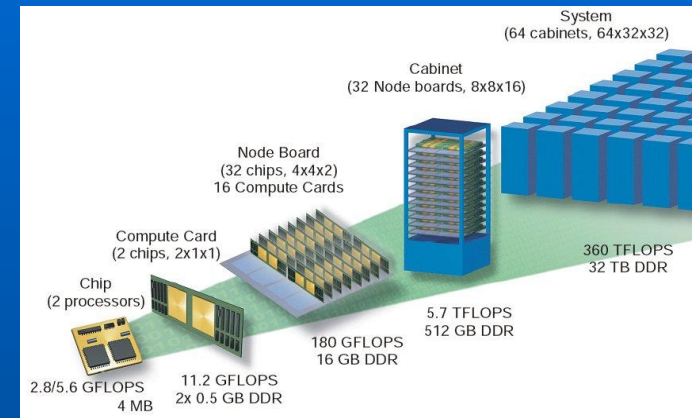


BlueGene/L



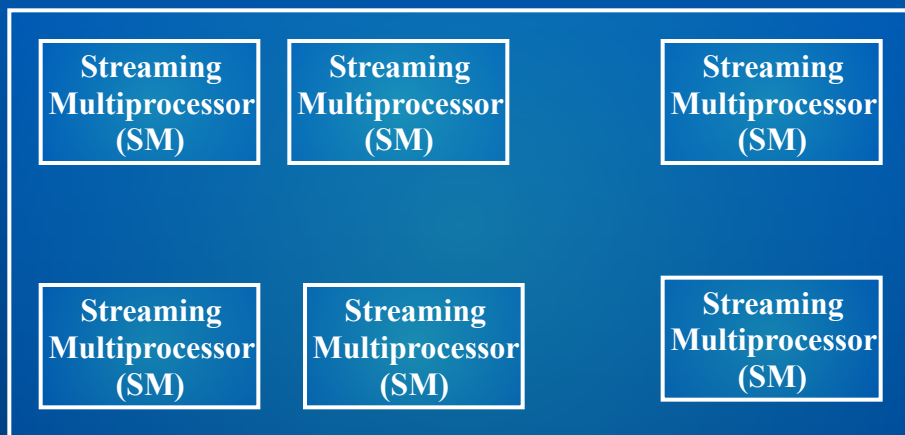
- ▶ 656536 dual-core nodes
- ▶ Node
 - 770 Mhz PowerPC
 - Double Hammer FPU (4 Flop/cycle)
 - 4 Mb on-chip L3 кэш
 - 512 Mb off-chip RAM
 - 6 двухсторонних портов для 3D-тора
 - 3 двухсторонних порта для collective network
 - 4 двухсторонних порта для barrier/interrupt

BlueGene/L

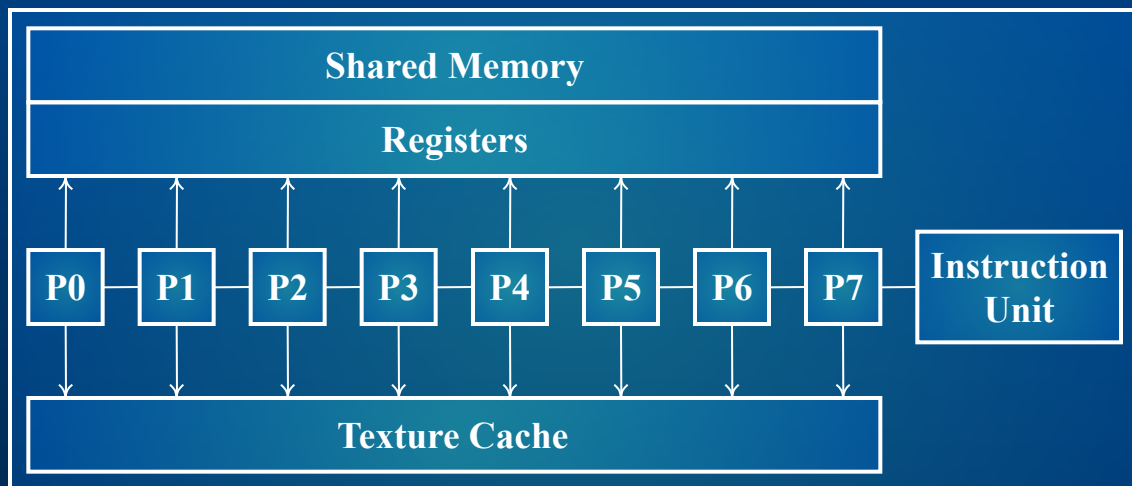


Архитектура G80

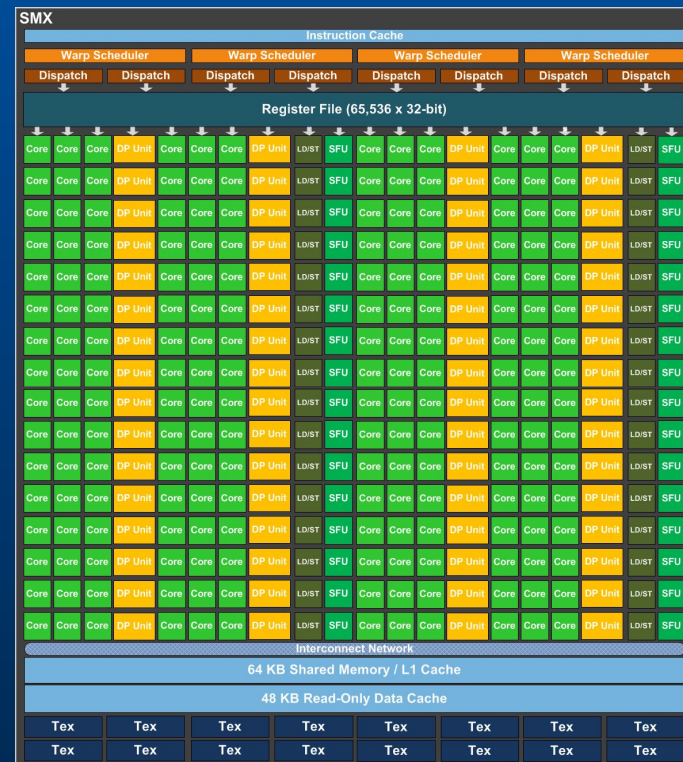
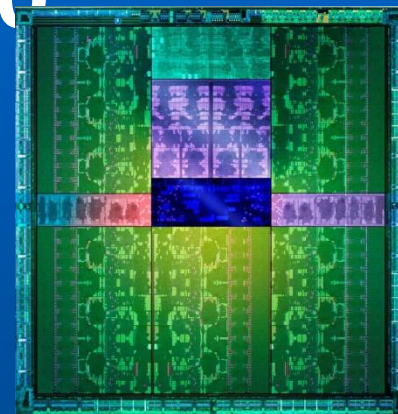
Массив из потоковых мультипроцессоров



Streaming Multiprocessor

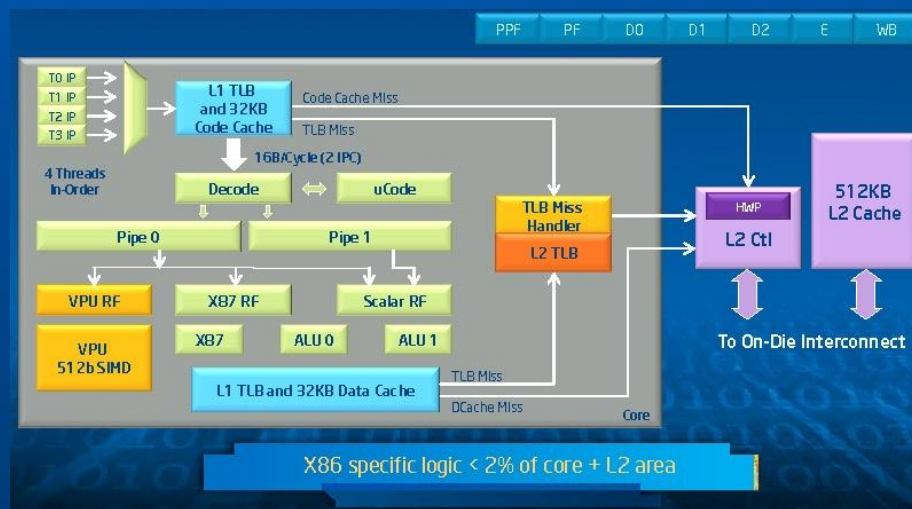
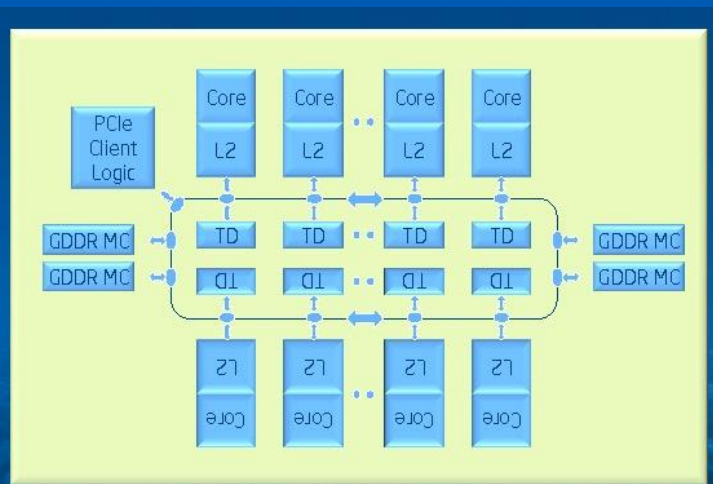
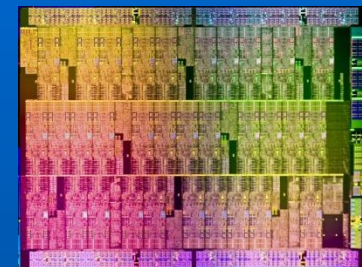


Архитектура KEPLER GK110

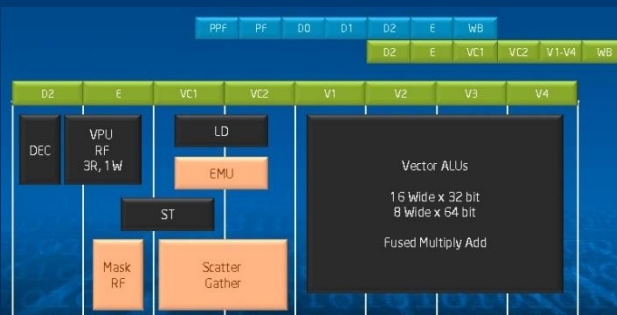


NVIDIA Tesla K20

Архитектура Knight Corner



Intel Xeon Phi



Компьютерная/системная шина

▶ *computer bus*



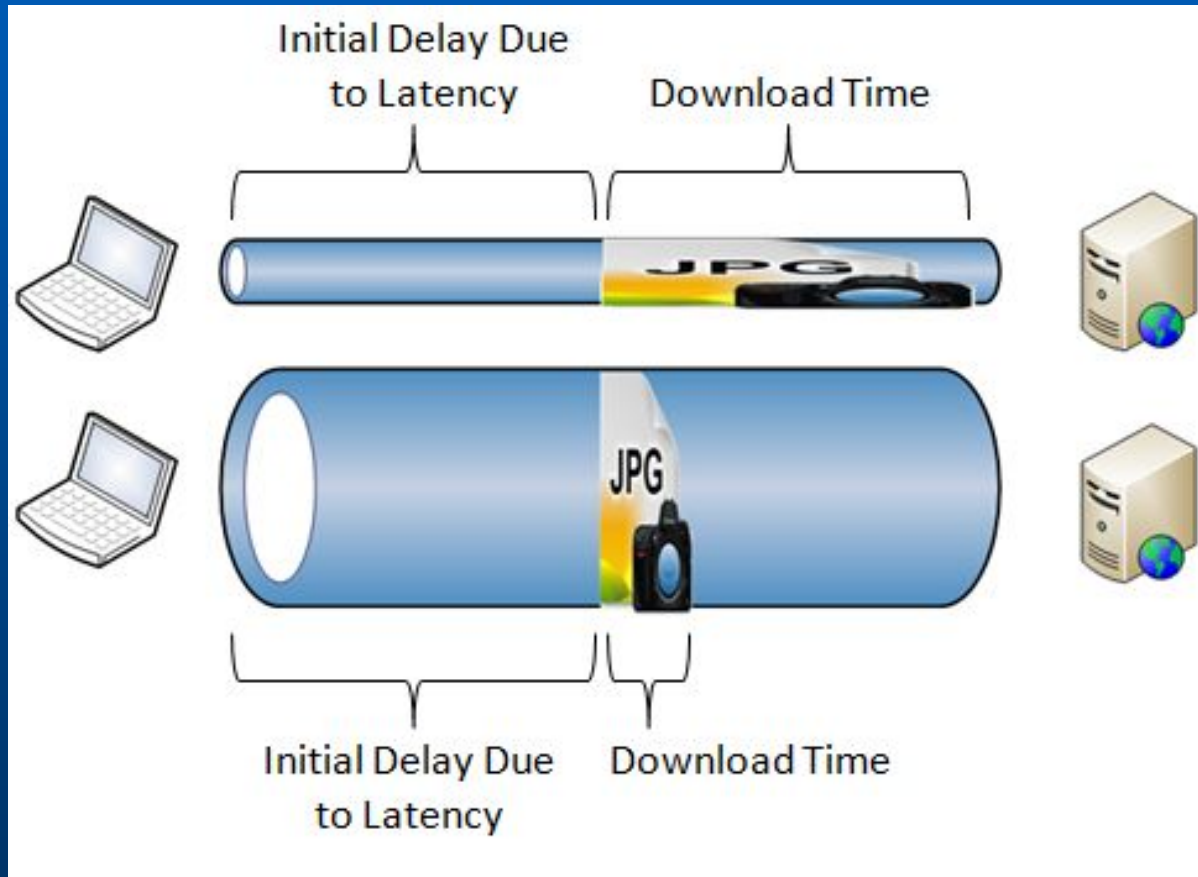
Шина данных (DataBus)

- ▶ PCI 2.1 (66 МГц, 64 bit) – 528 MB/s
- ▶ PCI X 2.0 (до 533 МГц) – 4 GB/s
- ▶ PCI Express – 1x - 0.5 GB/s, 32x – 16 GB/s
- ▶ PCI Express 2.0 – 5 GT/s for lane, 8b/10b, 8.0 GB/s for 16x (spec: 2007)
- ▶ PCI Express 3.0 – 8 GT/s for lane, 128b/130b, 15.75 GB/s for 16x (2010)
- ▶ PCI Express 4.0 – 16 GT/s for lane, 128b/130b, 31.51 GB/s for 16x (2017)

- ▶ HyperTransport (до 2.6 ГГц) – 40 GB/s
- ▶ Infinity Fabric / IF 2.0 – up to 50-100 GB/s

- ▶ FSB front-side bus (до 266 ГГц в Intel Core 2 Quad Pumped Bus) – 8.3 GB/s
- ▶ QuickPath Interconnect – up to 9.6 GT/s
(25.6 GB/s for 20-lane QPI link pair at 3.2 GHz)
- ▶ UPI (Ultra Path Interconnect) – up to 10.4 GT/s

Пропускная способность и латентность

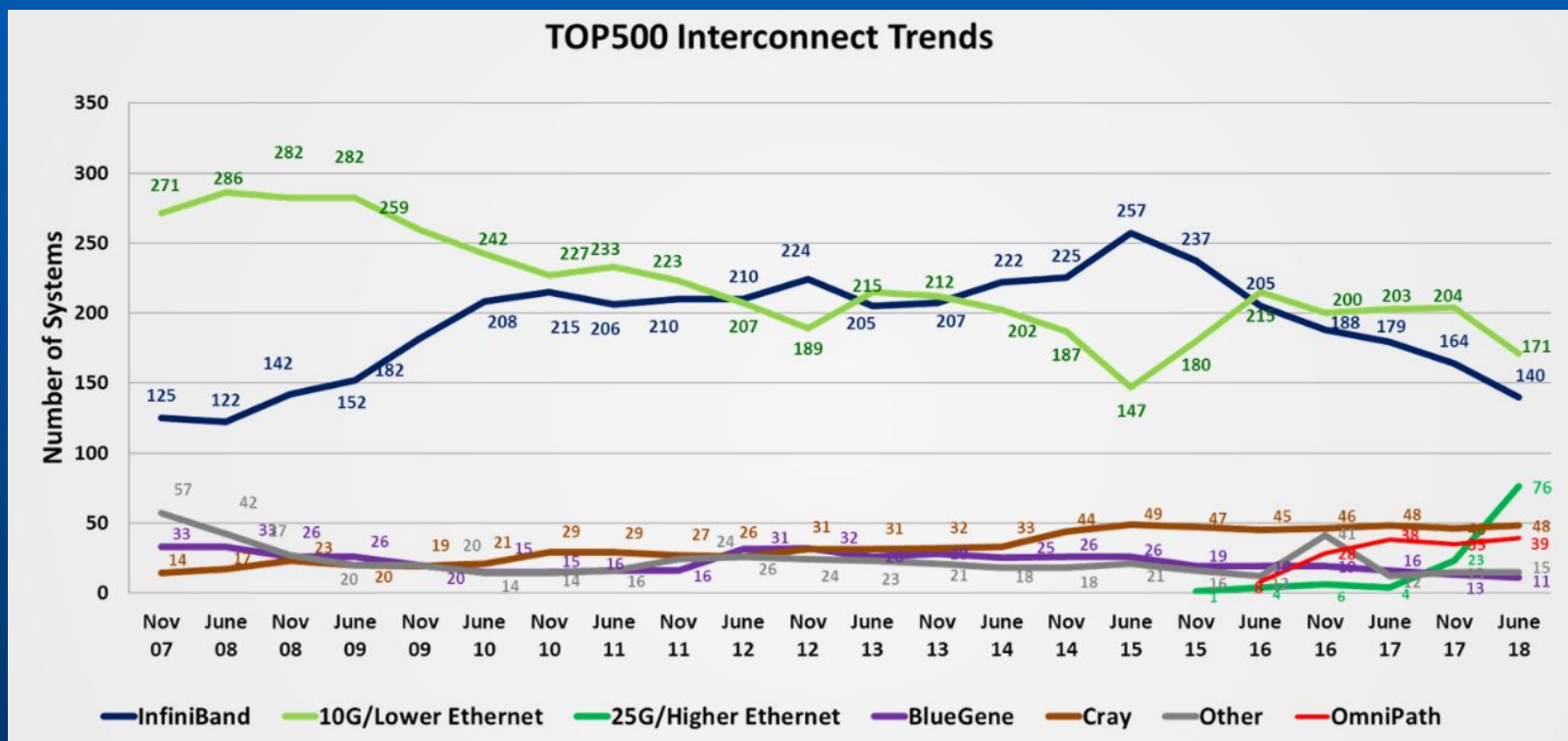


Интерконнект

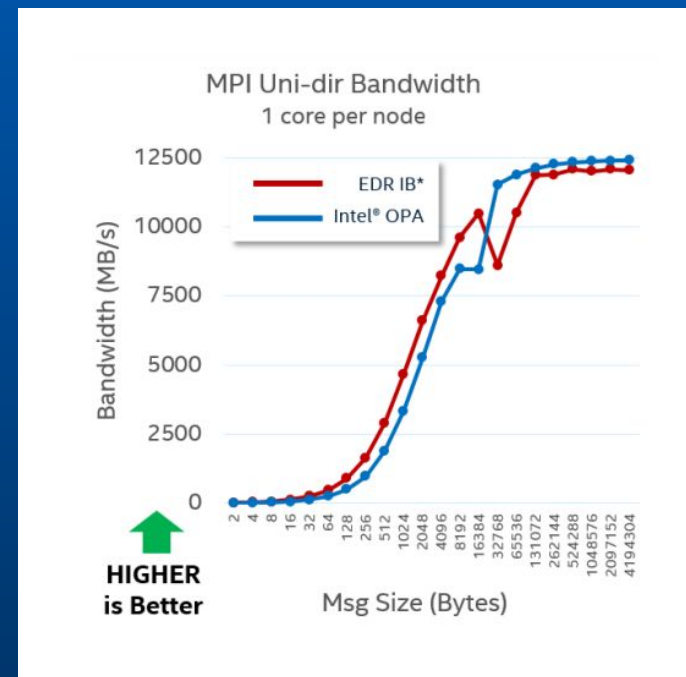
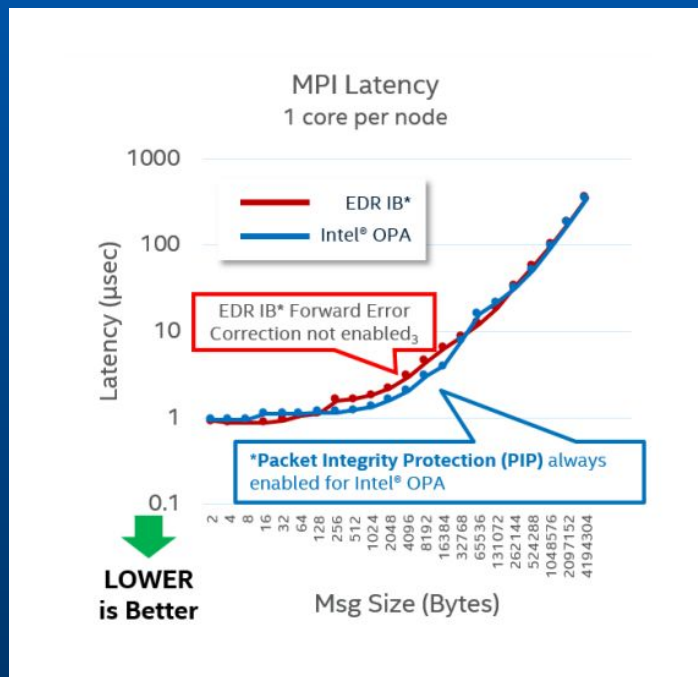
- Gigabit Ethernet (50-100 us / 100 MB/s)
- 40Gb Ethernet (~4 us / 5 GB/s)
- **InfiniBand EDR (<1 us / до 100 GB/s)**
- Omni-Path 100G (<1 us / до 100 GB/s)
- PCI Express (1-2 us / до 128 GB/s)

- уникальные, например IBM BlueGene или Gray
 - *Myrinet (3-10 us / 1 GB/s)*
 - *SCI (1-3 us / 1 GB/s)*
 - *Quadrics (1-3 us / 900 MB/s)*
 - *InfiniPath (1-2 us / 1 GB/s)*

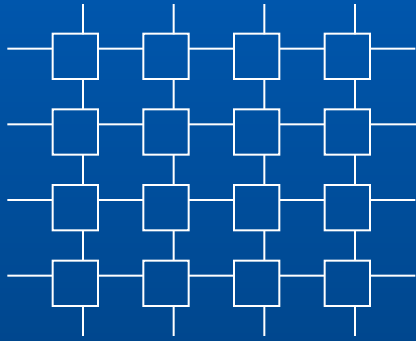
Использование в top500



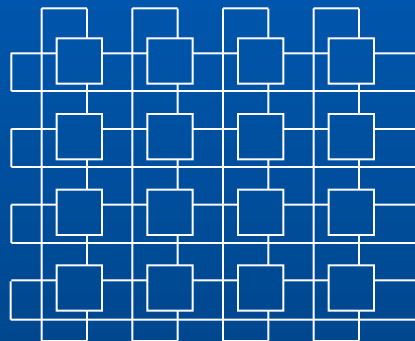
Пример измерения характеристик



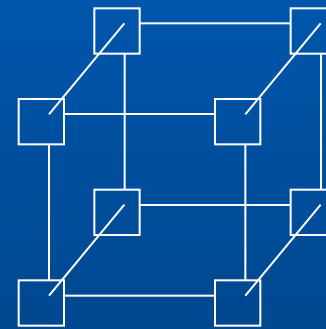
Топологии коммуникационных сетей



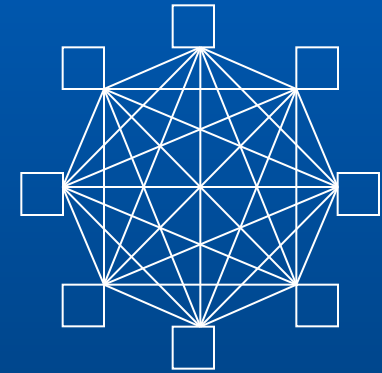
Сетка



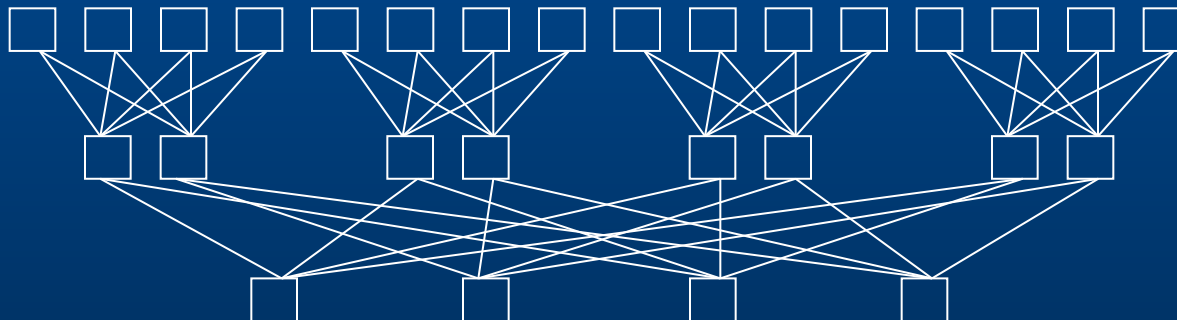
Двухмерный тор



Гиперкуб



Кольцо

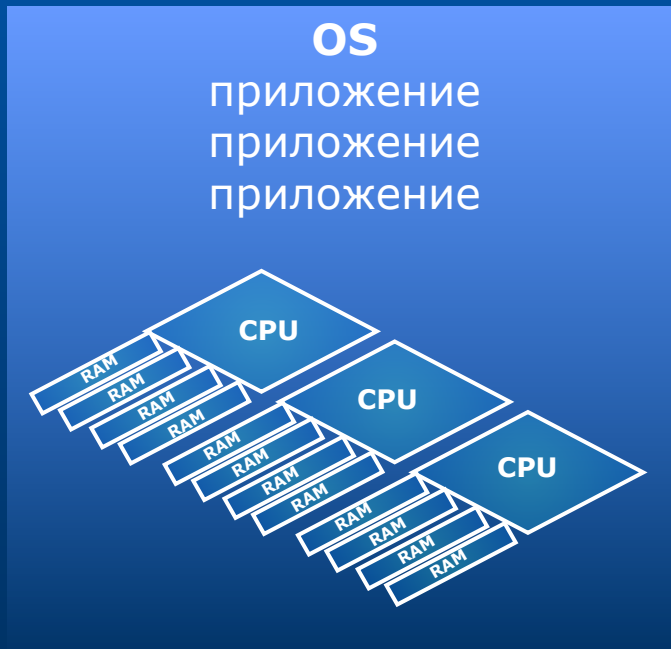


Толстое
дерево

Сравнение SMP и MPP

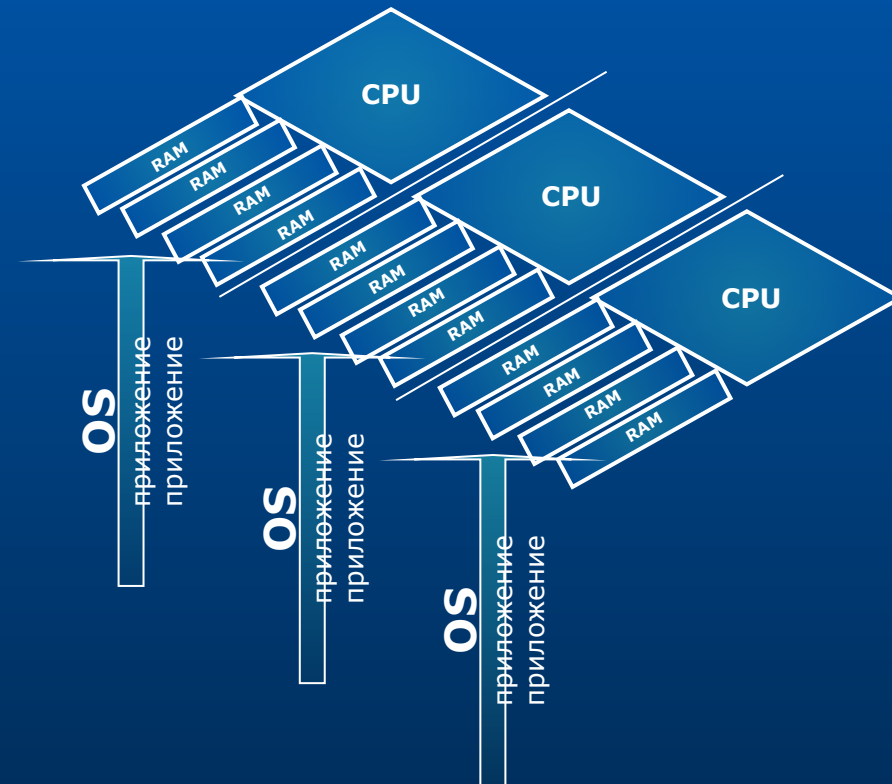
SMP

Архитектура с симметричным параллелизмом



MPP

Архитектура с массовым параллелизмом



Основные технологии параллельного программирования

- ▶ с разделяемой памятью (multi-core, SMP, NUMA)

- Pthread
- OpenMP
- ShMem
- ...

Shared Memory Paradigm

- ▶ с распределенной памятью (MPP)

- MPI
- PVM
- BSPlib
- ...

Massive Passing Programming Paradigm

- ▶ Высокоуровневые средства (языки программирования)

- Разделяемые переменные Java, Ada, SR, Cilk
- Обмен сообщениями Ada, Occam, Java, Fortran M, SR
- Координация Linda, Orca
- Параллельность по данным C*, HPF, NESL, ZPL

Определения

Один из первых архитекторов кластерной технологии **Грегори Пфистер** дал кластеру следующее определение:

«**Кластер** — это разновидность параллельной или распределенной системы,

которая:

- а. состоит из нескольких связанных между собой компьютеров;*
- б. используется как единый, унифицированный компьютерный ресурс».*

Обычно различают следующие основные виды кластеров:

- отказоустойчивые кластеры (High-availability clusters, HA)
- кластеры с балансировкой нагрузки (Load balancing clusters)
- вычислительные кластеры (High-performance clusters, HPC)

Beowulf

Кластер, который состоит из широко распространённого аппаратного обеспечения, работающий под управлением операционной системы, распространяемой с исходными кодами (например, GNU/Linux).



Преимущества Beowulf-систем

- ▶ стоимость системы гораздо ниже стоимости суперкомпьютера;
- ▶ возможность увеличения производительности системы;
- ▶ возможность использования устаревших компьютеров, тем самым увеличивается срок эксплуатации компьютеров;
- ▶ широкая распространённость аппаратного обеспечения;

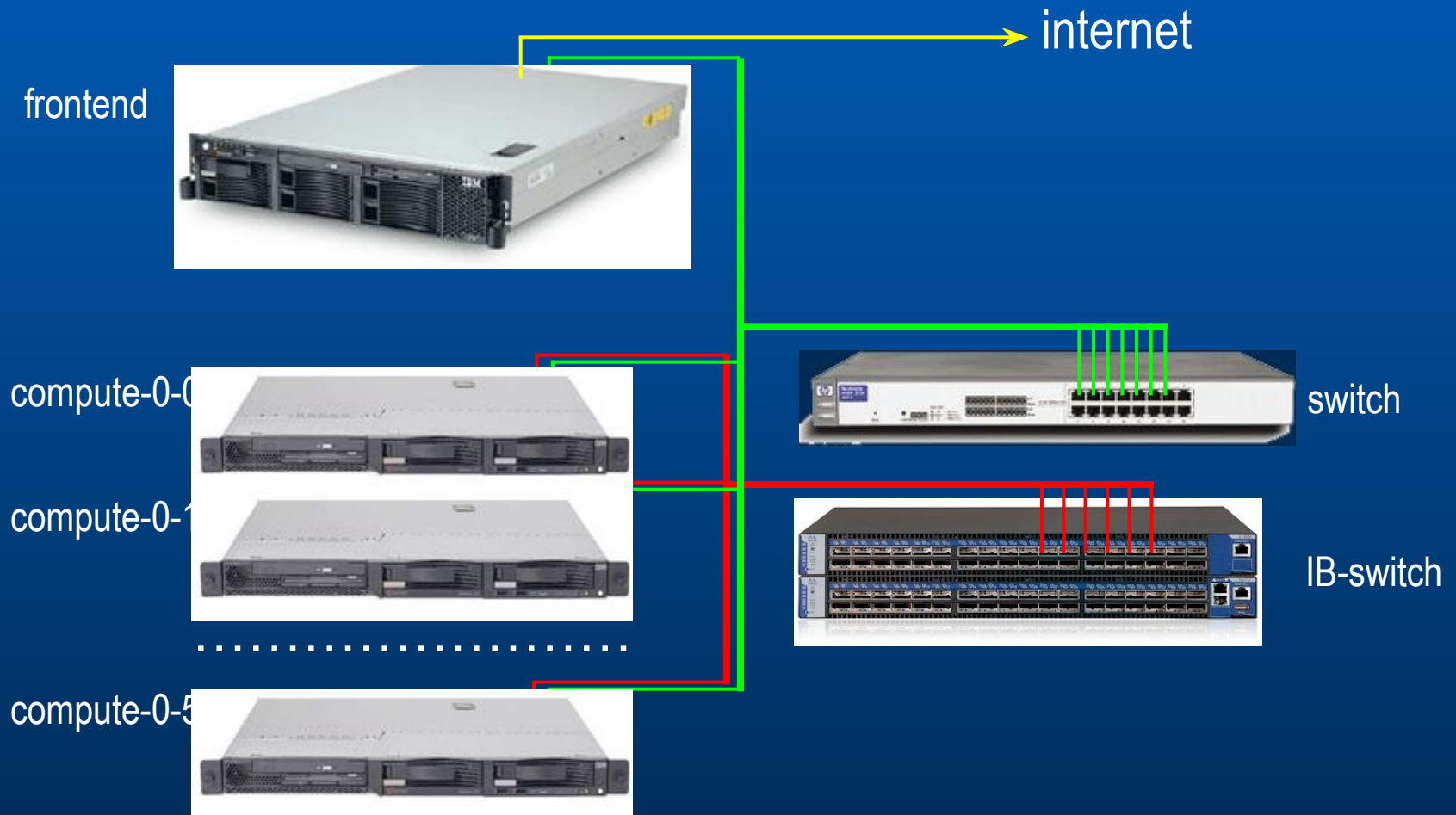


Вычислительный кластер

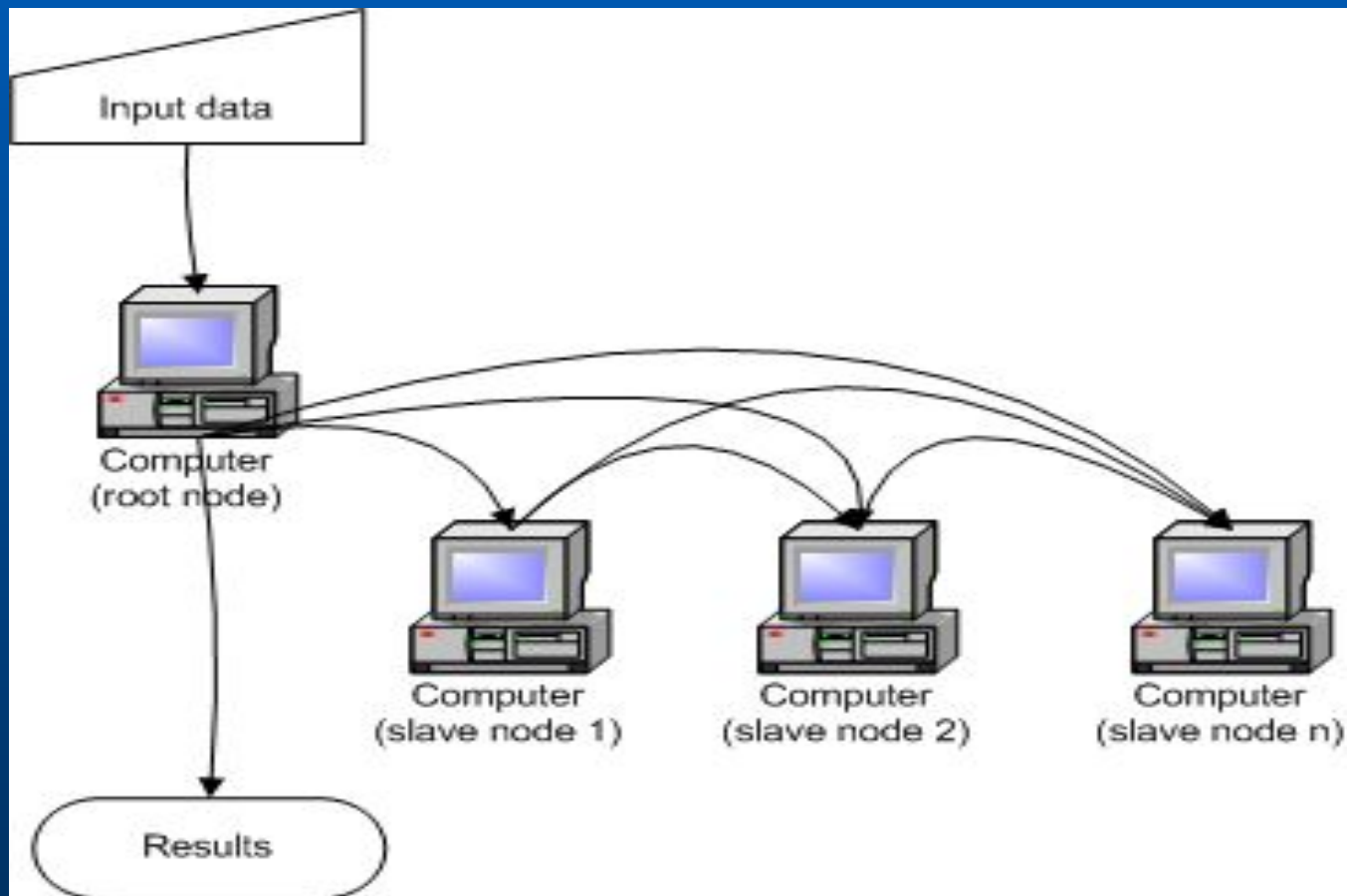


- ▶ IBM eServer 345 -1
- ▶ IBM eServer 335 -6
- ▶ HP Switch – 1 Gb
- ▶ UPS
- ▶ ASM
- ▶ Consol switch

Топология сети



Организация кластера



OpenMP Server

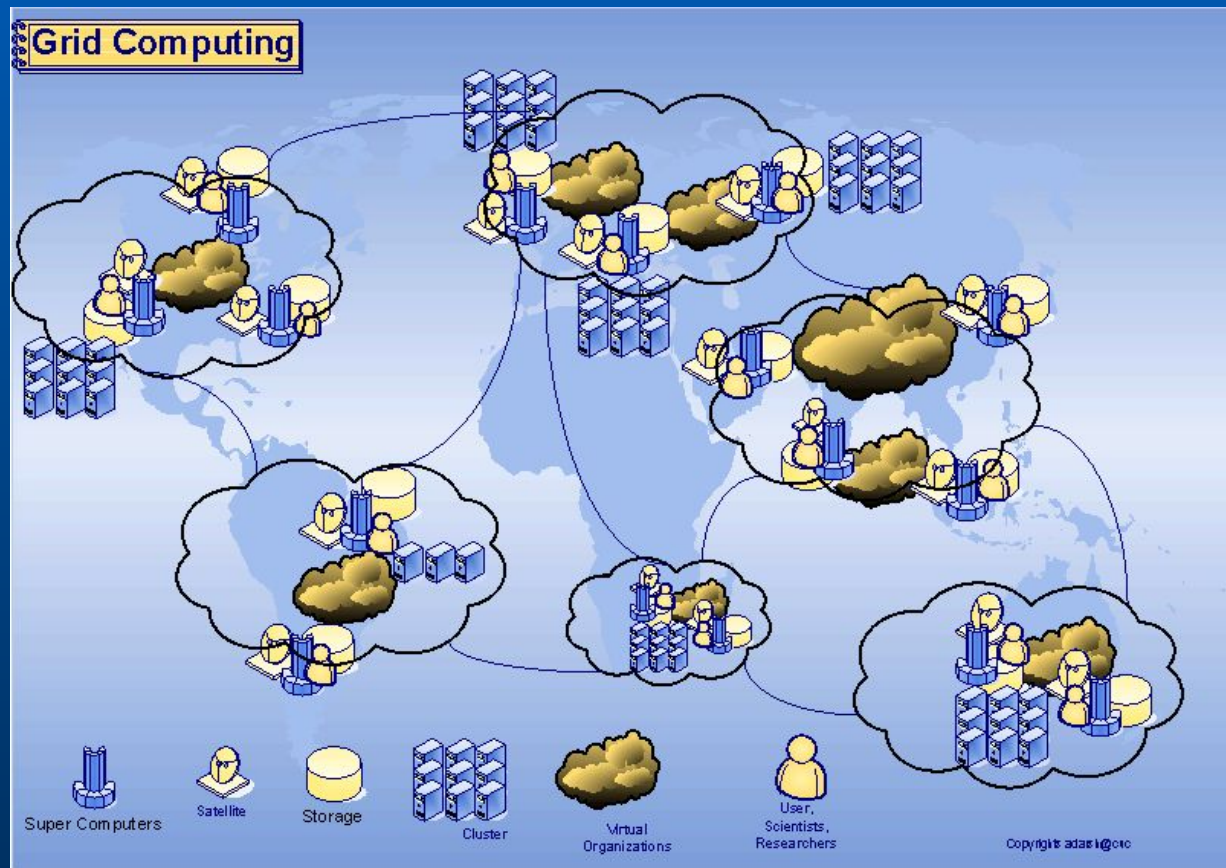
- ▶ openSUSE
- ▶ Intel Xeon X5660 @ 2.80GHz (8 core available)
- ▶ Терминальный Linux сервер (SSH)
- ▶ Компилятор GNU
- ▶ IP - 93.180.53.228

Суперкомпьютер "Ломоносов"

- ▶ <http://parallel.ru/cluster/lomonosov.html>

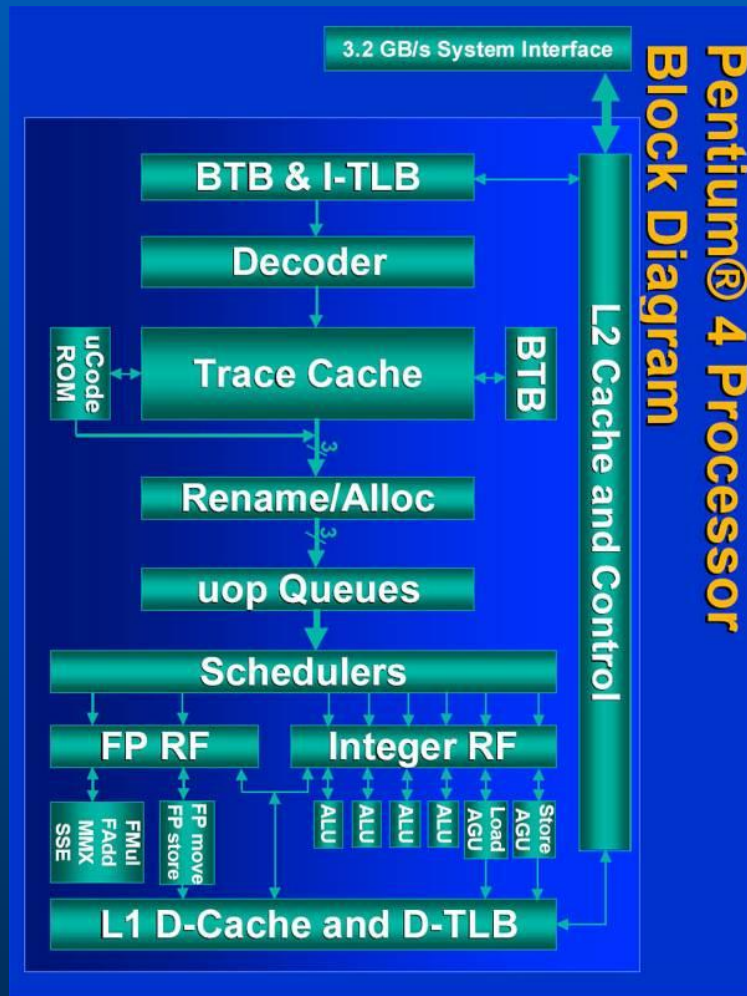


Технология GRID

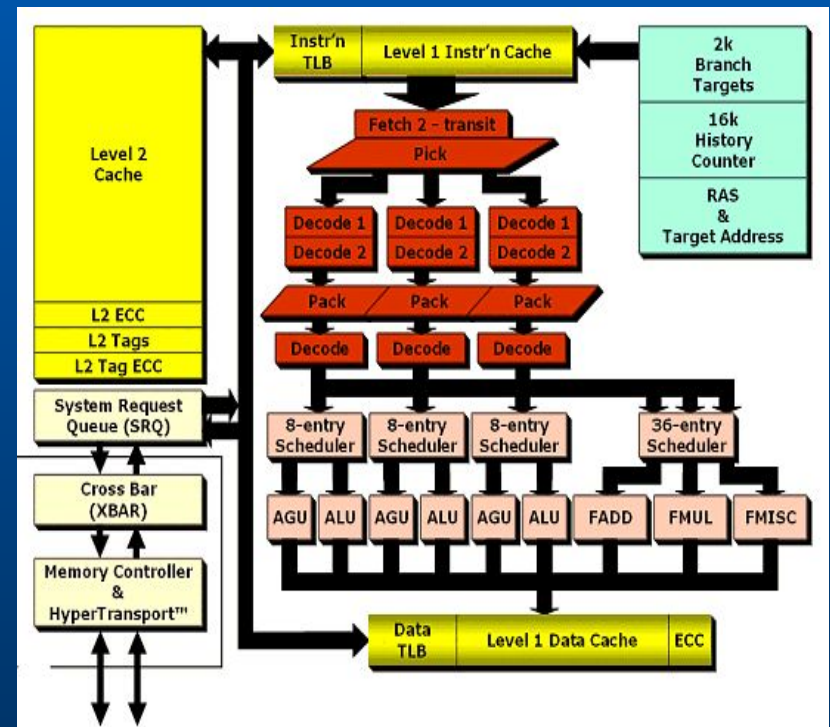




Примеры блок-схем процессоров

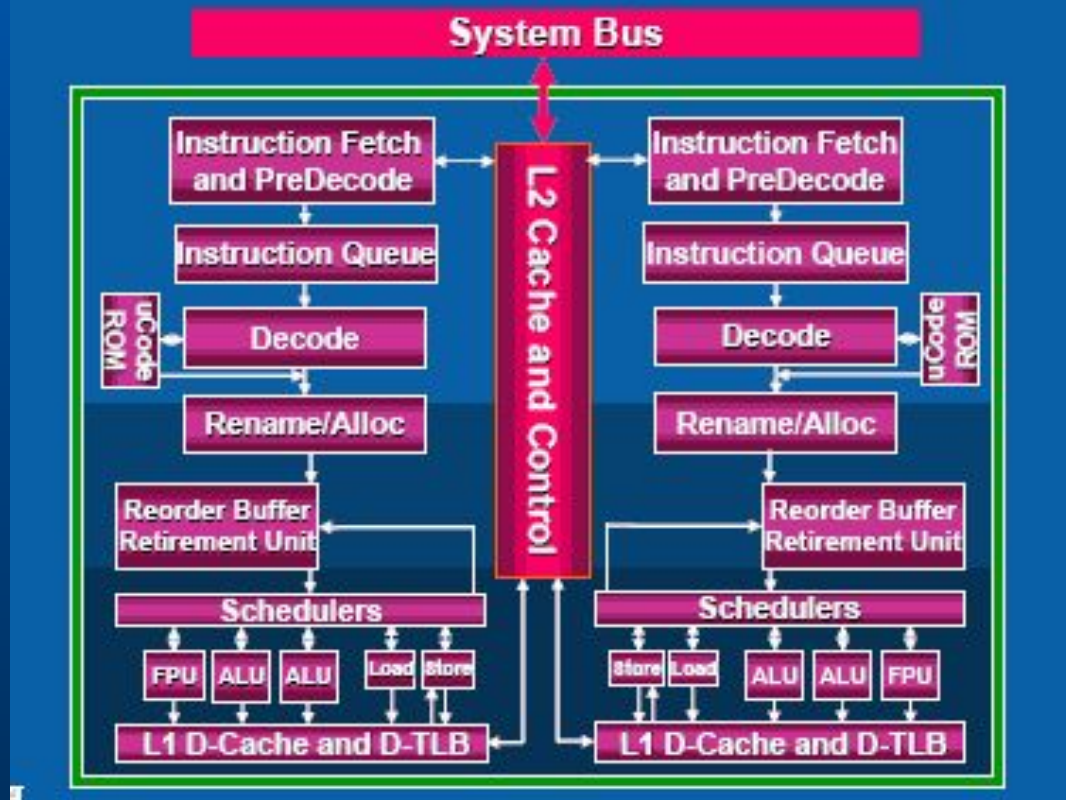


Athlon XP Block Diagram

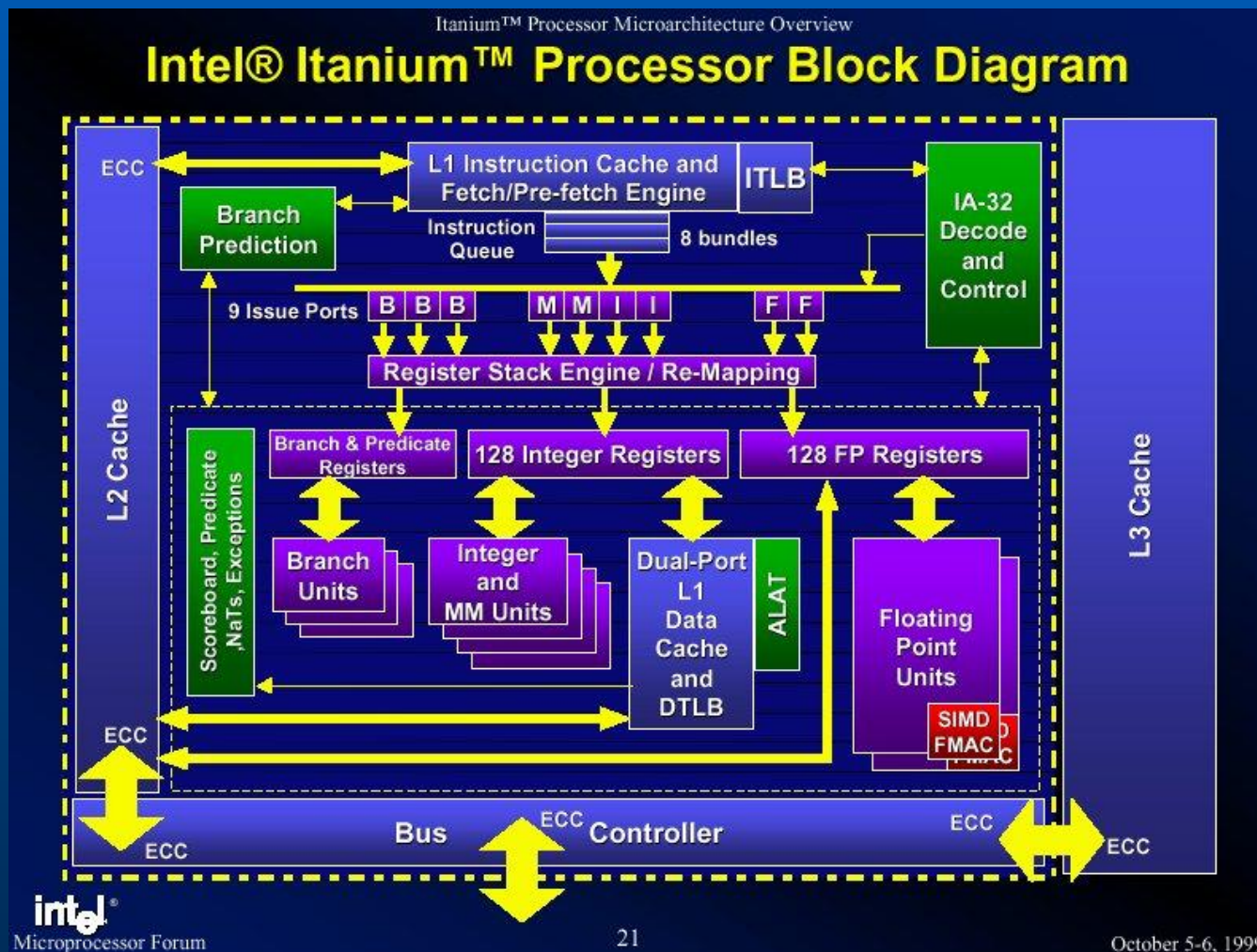


Intel Xeon Dual Core

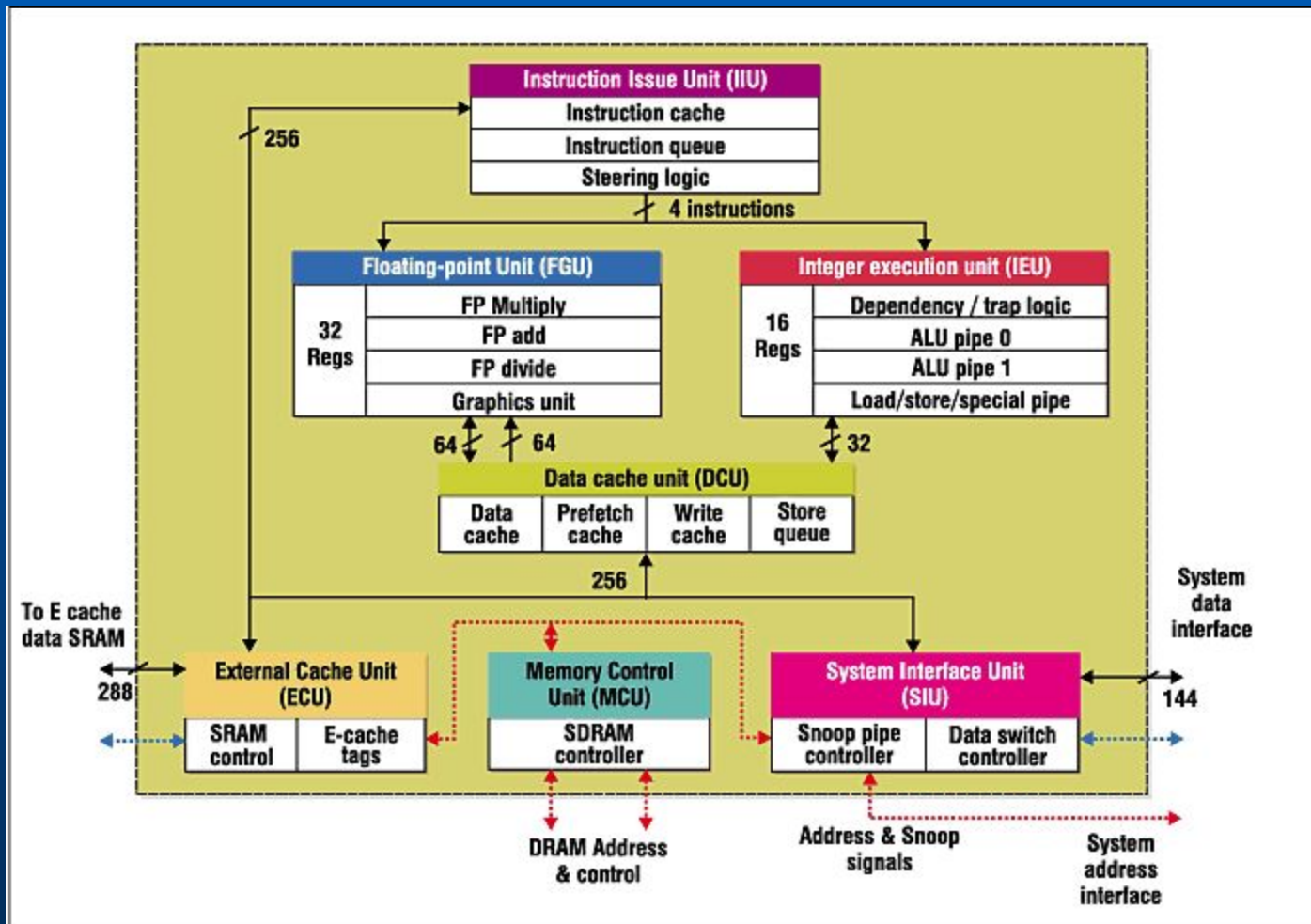
Merom, Conroe and Woodcrest Block Diagram

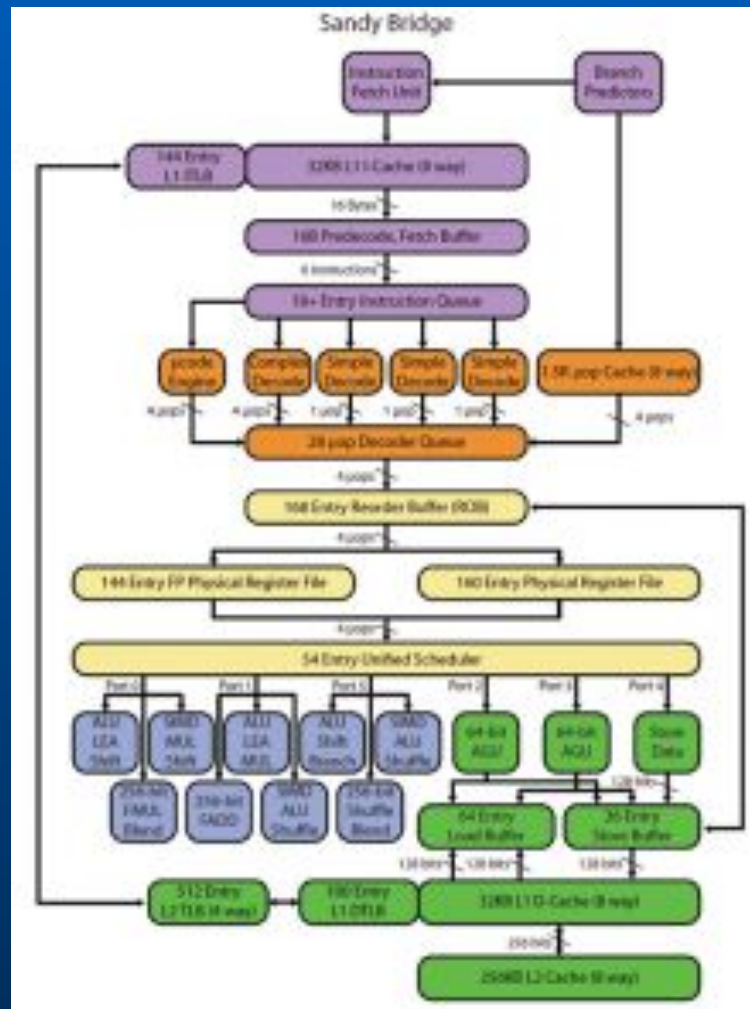


Intel Itanium/Itanium 2



Микропроцессор Sun Ultra SPARC III





GK210 SMX (Unofficial)

