

Суперкомпьютеры и параллельная обработка данных

Бахтин Владимир Александрович
*к.ф.-м.н., ведущий научный сотрудник
Института прикладной математики им М.В.Келдыша
РАН
кафедра системного программирования
факультет вычислительной математики и кибернетики
Московского университета им. М.В. Ломоносова*

Тематический план учебной дисциплины

- Введение в предмет
- Архитектура параллельных вычислительных систем
- Методы оценки производительности параллельных вычислительных систем
- Технологии параллельного программирования
- Введение в теорию анализа структуры программ и алгоритмов
- Введение в параллельные методы решения задач

- ❑ Лацис А.О. Параллельная обработка данных: учеб. пособие для студ. вузов. Издательский центр «Академия». 2010. Издательство: Академия
- ❑ Якововский М.В. Введение в параллельные методы решения задач. Учебное пособие. Серия: «Суперкомпьютерное образование». Издательство МГУ. 2013.
- ❑ Вл. В. Воеводин, В. В. Воеводин. Параллельные вычисления — СПб., БХВ-Петербург, 2002, 608 с.
- ❑ Антонов А.С. Технологии параллельного программирования MPI и OpenMP: Учеб. пособие. Предисл.: В.А.Садовничий. - Серия «Суперкомпьютерное образование». М.: Издательство Московского университета, 2012.-344 с.

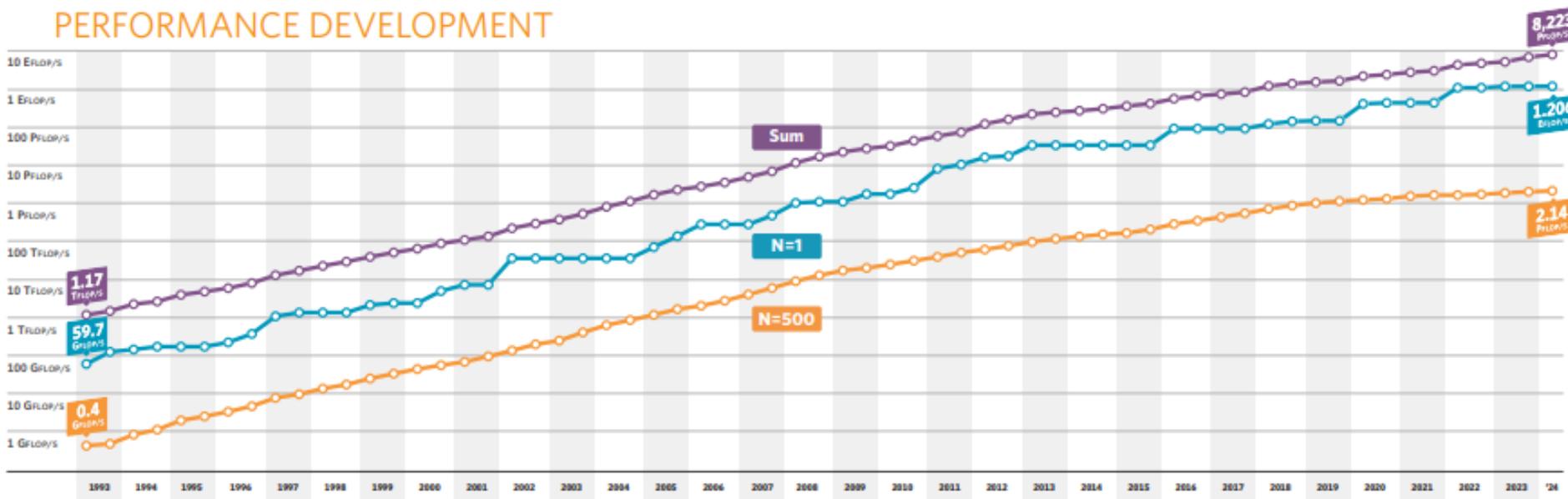
- ❑ OpenMP Application Programming Interface. Version 5.2. November, 2021. URL: <https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-2.pdf>
- ❑ MPI: A Message-Passing Interface Standard. Version 4.0. June 9, 2021. URL: <https://www.mpi-forum.org/docs/mpi-4.0/mpi40-report.pdf>
- ❑ The OpenACC Application Programming Interface. Version 3.2. November, 2021. URL: <https://www.openacc.org/sites/default/files/inline-images/Specification/OpenACC-3.2-final.pdf>
- ❑ Э. Таненбаум, М. ван Стеен. Распределенные системы. Принципы и парадигмы - М.: Питер, 2003 - 876 с. - Классика computer science; ISBN 5-272-00053-6

Суперкомпьютерные системы (Top500)

MAY 2024

			SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Frontier	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	DOE/SC/ORNL	USA	8,699,904	1,206.0	22.7
2	Aurora	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 (52C 2.4GHz), Intel Data Center GPU Max, Slingshot-11	DOE/SC/ANL	USA	9,264,128	1,012.0	38.7
3	Eagle	Microsoft NDv5, Xeon Platinum 8480C (48C 2GHz), NVIDIA H100, NVIDIA Infiniband NDR	Microsoft Azure	USA	1,123,200	561.2	
4	Fugaku	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
5	LUMI	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	EuroHPC/CSC	Finland	2,220,288	379.7	6.01

PERFORMANCE DEVELOPMENT



K computer

- ❑ Японский суперкомпьютер производства компании Fujitsu, запущенный в 2011 году в Институте физико-химических исследований в городе Кобе.
- ❑ В июне 2011 года K computer возглавил список самых производительных суперкомпьютеров мира с результатом в тесте LINPACK в 8,162 петафлопс.
- ❑ По состоянию на июнь 2011 года система имела 68 544 8-ядерных процессора SPARC64 VIIIfx, что составляло 548 352 вычислительных ядра, произведенных компанией Fujitsu по 45-нанометровому техпроцессу.
- ❑ В ноябре 2011 года K Computer был достроен, количество процессоров достигло 88 128, а производительность системы на тесте Linpack достигла 10,51 Пфлопс. Таким образом, K Computer стал первым в истории суперкомпьютером, преодолевшим рубеж в 10 Пфлопс.
- ❑ Стоимость 140 миллиардов йен, или 1,2 миллиарда долларов.



Gyokou Supercomputer



https://youtu.be/_z8ErBIBSo

5 сентября
Москва, 2024

Суперкомпьютеры и параллельная обработка данных

7 из 316

Суперкомпьютеры – что это?

- ❑ Суперкомпьютеры – это компьютеры, которые работают значительно быстрее остальной массы современных компьютеров
- ❑ Суперкомпьютеры – это компьютеры, которые занимают большой зал
- ❑ Суперкомпьютеры – это компьютеры, которые весят больше 1 тонны
- ❑ Суперкомпьютеры – это компьютеры, которые стоят больше 1 млн.долл.
- ❑ Суперкомпьютеры – это компьютеры, которые сводят проблему вычислений к проблеме ввода/вывода
- ❑ Суперкомпьютеры – это компьютеры, мощности которых лишь немного не хватает для решения актуальных вычислительно сложных задач

Суперкомпьютеры – что это?

Суперкомпьютером называется вычислительная система, вычислительное быстродействие которой многократно выше, чем у современных ей компьютеров массового выпуска.

Суперкомпьютеры

Серверы

Персональные компьютеры

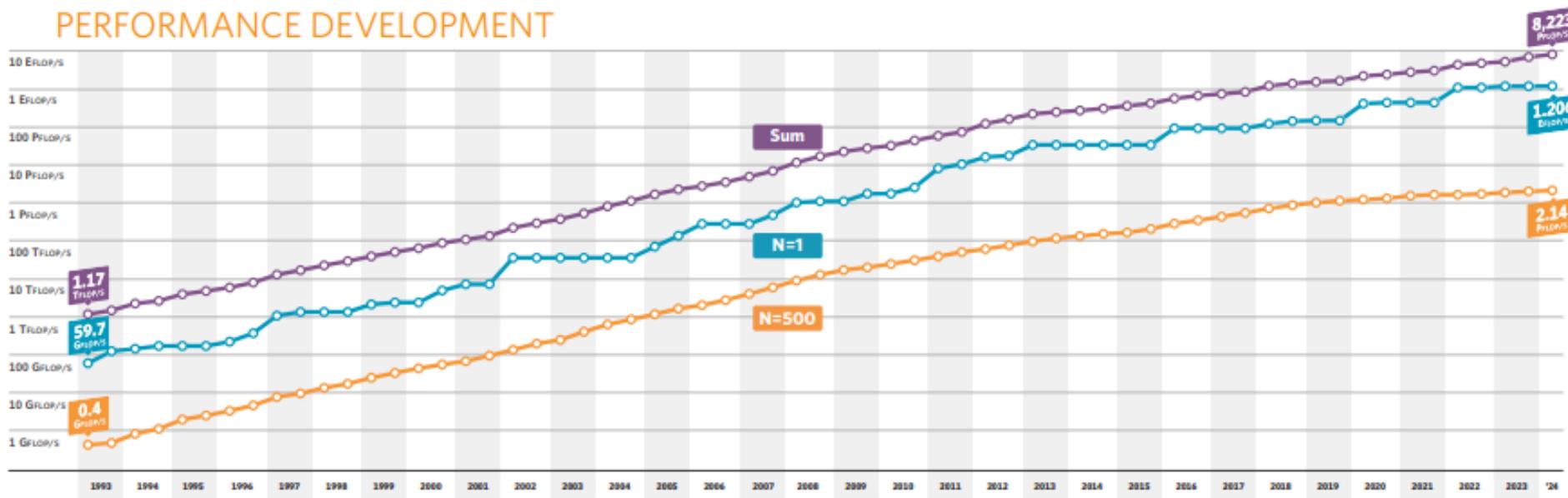
Мобильные компьютерные устройства

Суперкомпьютерные системы (Top500)

MAY 2024

			SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Frontier	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	DOE/SC/ORNL	USA	8,699,904	1,206.0	22.7
2	Aurora	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 (52C 2.4GHz), Intel Data Center GPU Max, Slingshot-11	DOE/SC/ANL	USA	9,264,128	1,012.0	38.7
3	Eagle	Microsoft NDv5, Xeon Platinum 8480C (48C 2GHz), NVIDIA H100, NVIDIA Infiniband NDR	Microsoft Azure	USA	1,123,200	561.2	
4	Fugaku	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
5	LUMI	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	EuroHPC/CSC	Finland	2,220,288	379.7	6.01

PERFORMANCE DEVELOPMENT



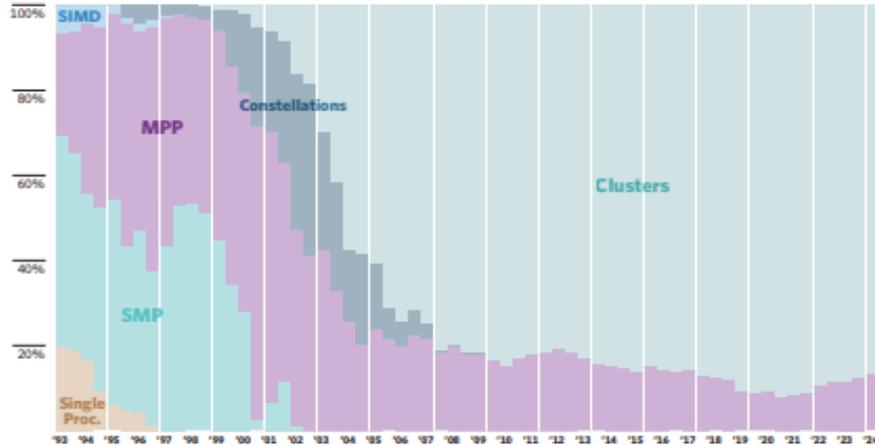
Производительность компьютеров. Тест Linpack

Тест Linpack - решение системы линейных алгебраических уравнений с плотной матрицей.

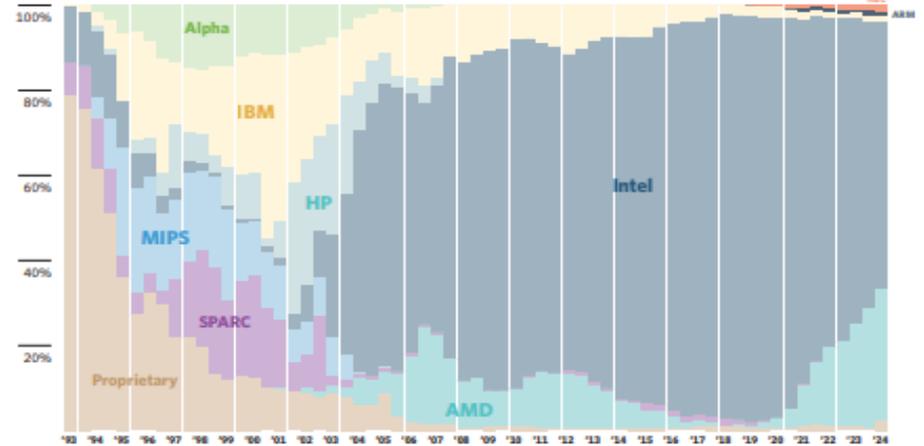
1. Матрица 100×100 , фиксированный текст программы.
2. Linpack TRP: матрица 1000×1000 , можно менять метод и текст программы. Сложность : $\frac{2n^3}{3} + 2n^2$.
3. High Performance Linpack: матрица любого размера, множество дополнительных параметров.

Суперкомпьютерные системы (Top500)

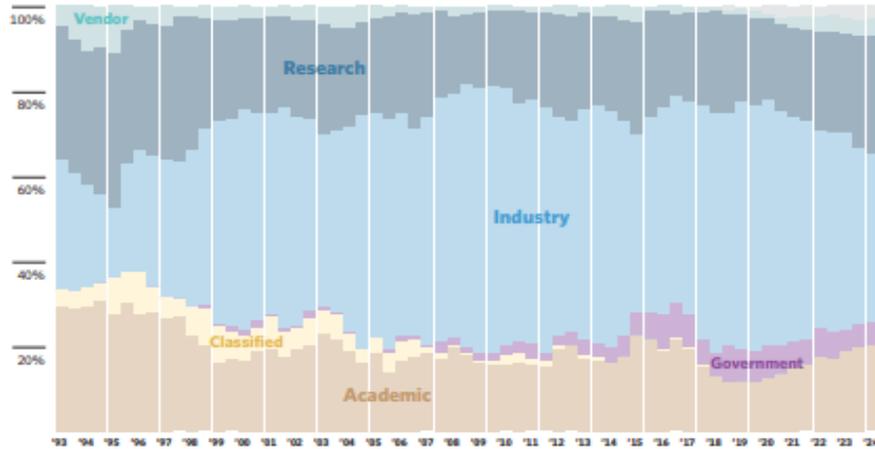
ARCHITECTURES



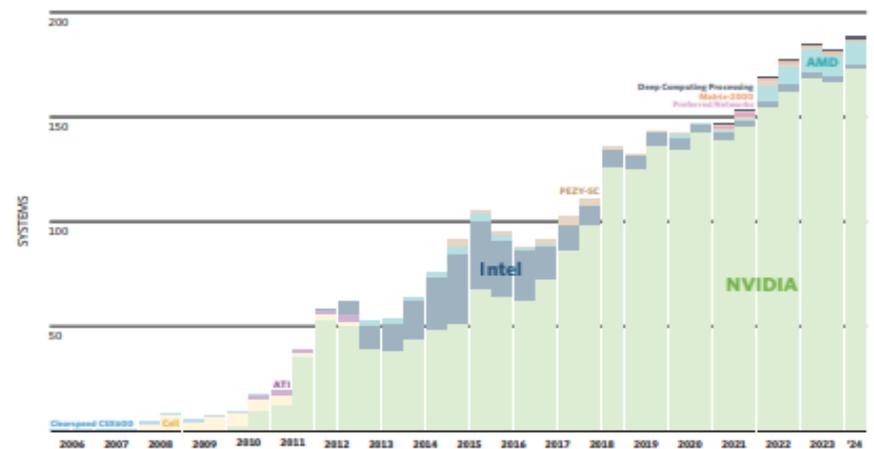
CHIP TECHNOLOGY



INSTALLATION TYPE



ACCELERATORS/CO-PROCESSORS



HPLINPACK

A Portable Implementation of the High Performance Linpack Benchmark for Distributed Memory Computers [FIND OUT MORE AT https://icl.utk.edu/hpl/](https://icl.utk.edu/hpl/)

Суперкомпьютерные системы (Top500)

13 мая 2024 года была опубликована 63-я редакция списка 500 наиболее мощных компьютеров мира Top500.

На первом месте списка остался суперкомпьютер Frontier производства HPE Cray на базе процессоров AMD EPYC 64C и графических ускорителей AMD Instinct 250X, установленный в Oak Ridge National Laboratory (ORNL), с пиковой производительностью 1.715 квинтиллионов ($*10^{18}$) операций с плавающей точкой в секунду (EFlop/s) и производительностью на тесте Linpack 1.206 EFlop/s.

На втором месте списка остался суперкомпьютер Aurora HPE Cray EX на базе процессоров Intel Xeon CPU Max и графических ускорителей Intel Data Center GPU Max, установленный в Argonne Leadership Computing Facility, чья производительность на тесте Linpack выросла до 1.012 EFlop/s. Aurora официально стала вторым экзафлопсным суперкомпьютером в мире.

На третьем месте списка остался суперкомпьютер Eagle на базе процессоров Intel Xeon Platinum 8480C и графических ускорителей NVIDIA H100, установленный в Microsoft Azure Cloud, с производительностью на тесте Linpack 561.2 PFlop/s.

На четвёртом месте списка остался японский суперкомпьютер Fugaku производства Fujitsu на базе процессоров ARM A64FX, установленный в RIKEN Center for Computational Science (R-CCS), с производительностью на тесте Linpack 442 PFlop/s.

На пятом месте списка остался суперкомпьютер LUMI производства HPE Cray на базе процессоров AMD EPYC 64C и графических ускорителей AMD Instinct 250X, установленный в EuroHPC/CSC (Финляндия), чья производительность на тесте Linpack выросла до 379.70 PFlop/s.

На шестое место списка попал новый суперкомпьютер Alps производства HPE Cray на базе Grace Hopper Superchip 72C, установленный в Swiss National Supercomputing Centre (CSCS), чья производительность на тесте Linpack составила 270 PFlop/s.

Суперкомпьютерные системы (Top500)

Суммарная производительность систем в списке составляет 8.21 EFlop/s (7.03 EFlop/s полгода назад). Последняя, 500-ая система в новой редакции списка была бы полгода назад на 457-ом месте. Для того, чтобы попасть в текущий список, потребовалась производительность на тесте Linpack 2.13 PFlop/s (в ноябре 2.02 PFlop/s).

На первом месте по количеству установленных систем, вошедших в список, остаётся компания Lenovo - 163 системы (169), далее HPE - 112 систем (103), EVIDEN - 49 систем (48), DELL EMC - 34 системы, Inspur - 22 системы (34).

По доле суммарной производительности систем на первом месте HPE - 36.2% (34.9%), далее EVIDEN - 9.6% (9.8%), Lenovo - 7.4% (8.6%), Fujitsu - 7% (8.1%).

Среди стран по количеству установленных систем на первом месте США - 171 система (161 полгода назад), далее Китай - 80 (104), Германия - 40 (36), Япония - 29 (32).

Доля производительности американских систем составляет 53.7% от всех систем списка (53%), доля производительности японских систем - 8.2% (9.5%), а китайских систем - 4.3% (5.8%).

Наиболее популярными коммуникационными технологиями остаются InfiniBand - 239 систем (в прошлом списке 219) и Gigabit Ethernet - 195 систем (209). Коммуникационная технология Omni-Path теперь используется в 32 суперкомпьютерах (33).

Количество систем в списке, построенных на процессорах Intel, уменьшилось с 338 до 314, процессоры AMD используются в 157 системах (140). В список входят 9 систем на базе процессоров Arm. 195 систем использует ускорители или сопроцессоры (в ноябре - 186).

Суперкомпьютерные системы (Top500). Россия

На 42-ое место списка с 36-го опустился суперкомпьютер **Chervonenkis** производства IPE, Nvidia и Tian, установленный в Yandex, чья пиковая производительность составляет 29.4 PFlop/s, а производительность на тесте Linpack - 21.5 PFlop/s.

На 69-ое место списка с 58-го опустился суперкомпьютер **Galushkin** производства IPE, Nvidia и Tian, установленный в Yandex, чья пиковая производительность составляет 20.6 PFlop/s, а производительность на тесте Linpack - 16 PFlop/s.

На 79-ое место списка с 64-го опустился суперкомпьютер **Lyapunov** производства NUDT и Inspur, установленный в Yandex, чья пиковая производительность составляет 20 PFlop/s, а производительность на тесте Linpack - 12.8 PFlop/s.

На 83-е место списка с 67-го опустился суперкомпьютер **"Кристофари Нео"** производства NVIDIA, установленный в Сбербанке, чья пиковая производительность составляет 14.9 PFlop/s, а производительность на тесте Linpack - 12 PFlop/s.

На 142-ое место списка с 119-го опустился суперкомпьютер **"Кристофари"** производства NVIDIA, установленный в Сбербанке, чья пиковая производительность составляет 8.79 PFlop/s, а производительность на тесте Linpack - 6.67 PFlop/s.

На 406-ое место списка со 370-го опустился суперкомпьютер **"Ломоносов-2"** производства компании "Т-Платформы", установленный в Научно-исследовательском вычислительном центре МГУ имени М.В.Ломоносова, чья пиковая производительность составляет 4.95 PFlop/s, а производительность на тесте Linpack - 2.48 PFlop/s.

На 472-ое место списка с 433-го опустился суперкомпьютер **MTS GROM** производства NVIDIA, установленный в #CloudMTS, чья пиковая производительность составляет 3.01 PFlop/s, а производительность на тесте Linpack - 2.26 PFlop/s.

Поколения архитектур и парадигмы программирования

Середина 70-х годов.

Векторно-конвейерные компьютеры

Особенности архитектуры: векторные функциональные устройства, зацепление функциональных устройств, векторные команды в системе команд, векторные регистры. Программирование: векторизация самых внутренних циклов.

Cray Fortran первый компилятор с Fortran векторизацией

Суперкомпьютер Cray-1
Пиковая производительность
машины — 133 Мфлопса.



Поколения архитектур и парадигмы программирования

Начало 80-х годов.

Векторно-параллельные компьютеры

Особенности архитектуры: векторные функциональные устройства, зацепление функциональных устройств, векторные команды в системе команд, векторные регистры.

Небольшое число процессоров объединяются над общей памятью.

Программирование: векторизация самых внутренних циклов и распараллеливание на внешнем уровне, единое адресное пространство, локальные и глобальные переменные.



Суперкомпьютеры Cray X-MP, Cray Y-MP

Поколения архитектур и парадигмы программирования

Начало 90-х годов.

Массивно-параллельные компьютеры

Особенности архитектуры: тысячи процессоров объединяются с помощью коммуникационной сети по некоторой топологии, распределенная память.

Программирование: обмен сообщениями, отсутствие единого адресного пространства, PVM, Message Passing Interface. Необходимость выделения массового параллелизма, явного распределения данных и согласования параллелизма с распределением.



Суперкомпьютер Cray T3D,
307 Гфлопс

Поколения архитектур и парадигмы программирования

Середина 90-х годов.

Параллельные компьютеры с общей памятью

Особенности архитектуры: сотни процессоров объединяются над общей памятью.

Программирование: единое адресное пространство, локальные и глобальные переменные, OpenMP.



Dec AlphaServer

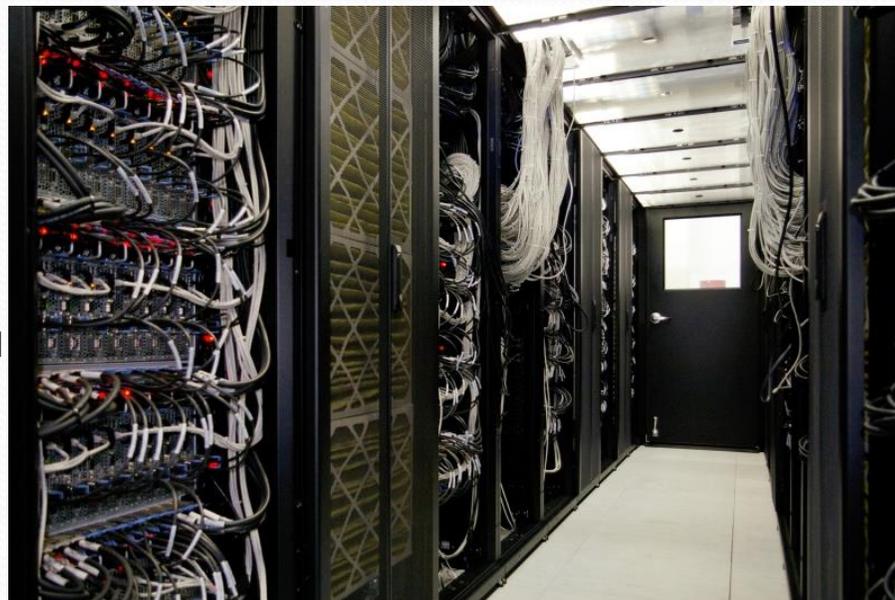
Поколения архитектур и парадигмы программирования

Начало 2000-х.

Кластеры из узлов с общей памятью

Особенности архитектуры: большое число многопроцессорных узлов объединяются вместе с помощью коммуникационной сети по некоторой топологии, распределенная память; в рамках каждого узла несколько (многоядерных) процессоров объединяются над общей памятью.

Программирование: неоднородная схема MPI+OpenMP; необходимость выделения массового параллелизма, явное распределение данных, обмен сообщениями на внешнем уровне; распараллеливание в едином адресном пространстве, локальные и глобальные переменные на уровне узла с общей памятью.



СКИФ МГУ «Чебышев»,
60 Тфлопс

Поколения архитектур и парадигмы программирования

Середина 2000-х.

Кластеры из узлов с общей памятью и ускорителями

Особенности архитектуры: большое число многопроцессорных узлов объединяются вместе с помощью коммуникационной сети по некоторой топологии, распределенная память; в рамках каждого узла несколько (многоядерных) процессоров объединяются над общей памятью; на каждом узле несколько ускорителей (GPU, PHI).

Программирование:

MPI+OpenMP+CUDA/OpenCL



МГУ «Ломоносов», 1.7 Пфлопс

Поколения архитектур и парадигмы программирования

С 1976 года до наших дней:

- ❑ 70-е – Векторизация циклов
- ❑ 80-е – Распараллеливание циклов (внешних) + Векторизация (внутренних)
- ❑ 90-е – MPI
- ❑ середина 90-х – OpenMP
- ❑ середина 2000-х – MPI+OpenMP
- ❑ 2010-е – CUDA, OpenCL, MPI+OpenMP + ускорители (GPU, Xeon Phi)
- ❑ ...

50 самых мощных компьютеров СНГ (top50.supercomputers.ru)

№	Название Место установки	Узлов Проц. Ускор.	Архитектура: кол-во узлов: конфигурация узла сеть: вычислительная / сервисная / транспортная	Rmax Rpeak (Тфлп/с)	Разработчик Область применения
1 <i>new</i>	«Червоненкис» Яндекс, Москва	199 398 1592	199: CPU: 2x AMD EPYC 7702 , 1024 GB RAM Acc: 8x NVIDIA A100 HDR InfiniBand / нд / 100 Gigabit Ethernet	21530.0 29415.17	Яндекс NVIDIA IT Services
2 <i>new</i>	«Галушкин» Яндекс, Москва	136 272 1088	136: CPU: 2x AMD EPYC 7702 , 1024 GB RAM Acc: 8x NVIDIA A100 HDR InfiniBand / нд / 100 Gigabit Ethernet	16020.0 20636.1	Яндекс NVIDIA IT Services
3 <i>new</i>	«Ляпунов» Яндекс, Москва	137 274 1096	137: CPU: 2x AMD EPYC 7662, 512 GB RAM Acc: 8x NVIDIA A100 HDR InfiniBand / нд / 100 Gigabit Ethernet	12810.0 20029.19	NVIDIA Inspur IT Services
4 <i>new</i>	«Кристофари Нео» SberCloud (ООО «Облачные технологии») , СберБанк, Москва	99 198 792	99: CPU: 2x AMD EPYC 7742, 2048 GB RAM Acc: 8x NVIDIA A100 HDR InfiniBand / 10 Gigabit Ethernet / 200 Gigabit Ethernet	11950.0 14908.6	NVIDIA SberCloud (ООО «Облачные технологии») Облачный провайдер
5 ▾	«Кристофари» SberCloud (ООО «Облачные технологии») , СберБанк, Москва	75 150 1200	75: NVIDIA DGX-2 CPU: 2x Intel Xeon Platinum 8168 24C 2.7GHz, 1536 GB RAM Acc: 16x NVIDIA Tesla V100 EDR InfiniBand / 100 Gigabit Ethernet / 10 Gigabit Ethernet	6669.0 8789.76	SberCloud (ООО «Облачные технологии») NVIDIA Облачный провайдер
6 ▾	«Ломоносов-2» Московский государственный университет имени М.В. Ломоносова, Москва	1696 1696 1856	1536: CPU: 1x Intel Xeon E5-2697v3, 64 GB RAM Acc: 1x NVIDIA Tesla K40M 160: CPU: 1x Intel Xeon Gold 6126, 96 GB RAM Acc: 2x NVIDIA Tesla P100	2478.0 4946.79	T-Платформы Наука и образование

Новый суперкомпьютер МГУ (МГУ-270)



5 сентября
Москва, 2024

Применение суперкомпьютеров

- Сокращение времени решения вычислительно сложных задач
- Сокращение времени обработки больших объемов данных
- Решение задач реального времени
- Создание систем высокой надежности

Суперкомпьютеры... Зачем?

- ❑ Неужели есть настолько **сложные задачи**, что для их решения хорошего сервера не хватает?
- ❑ Неужели есть настолько **важные задачи**, которые оправдывают крайне высокую стоимость суперкомпьютеров?

А далеко ли вычислительно сложные задачи?

Задача о числе счастливых билетиков :

```
count = 0;
for ( i1 = 0, ++i1)
  for ( i2 = 0, ++i2)
    for ( i3 = 0, ++i3)
      for ( i4 = 0, ++i4)
        for ( i5 = 0, ++i5)
          for ( i6 = 0, ++i6) {
            if ( i1+i2+i3+i4+i5+i6 == i4+i5+i6 )
              count+1;
          }
        }
      }
    }
  }
```

Intel Core Duo 2.6 ГГц:

8 цифр – 0.1 с

10 цифр – 10 с

12 цифр – 1780 с



А далеко ли вычислительно сложные задачи?

Задача о числе счастливых билетиков :

```
count = 0;
for ( i1 = 0; i1 < 10; ++i1)
  for ( i2 = 0; i2 < 10; ++i2)
    for ( i3 = 0; i3 < 10; ++i3)
      for ( i4 = 0; i4 < 10; ++i4)
        for ( i5 = 0; i5 < 10; ++i5)
          for ( i6 = 0; i6 < 10; ++i6) {
            if( i1+i2+i3 == i4+i5+i6 )
              count = count+1;
          }
```

Поможет ли
оптимизация
программы?



Поможет ли
использование
суперкомпьютера?

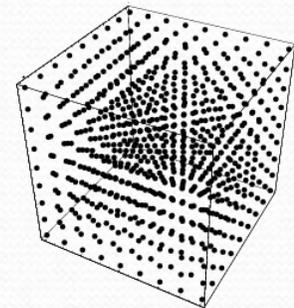
Вычислительная сложность – 10^n операций !

Сверхвысокая производительность - зачем?

Моделирование нефтяных резервуаров:

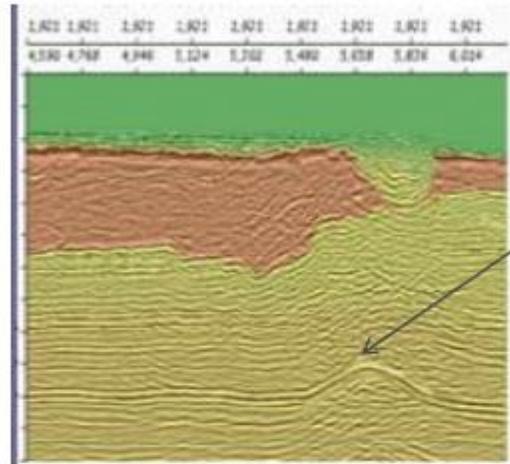
- Нефтеносная область – $100*100*100$ точек
- в каждой точке вычисляется от 5 до 20 функций (скорость, давление, концентрация, температура, ...)
- 200-1000 операций для вычисления каждой функции в каждой точке
- 100-1000 шагов по времени

- Итого: 10^6 (точек сетки) * 10 (функций) *
* 500 (операций) * 500 (шагов) =
= 2500 млрд. операций



Сверхвысокая производительность - зачем?

Ex: Increasing efficiency in Oil & Gas



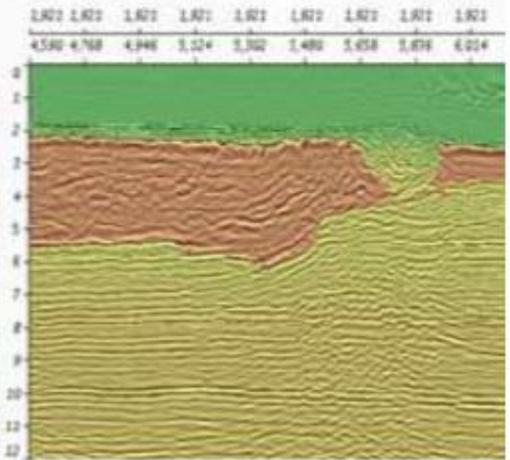
Seismic profiles of a region of the Gulf of Mexico.

The top image, in 2003, on 64 processors,
At the bottom right-hand side, a structure shaped like a bowler hat,
typical of a petroleum zone.

Based on this image, ready to install boring equipment on this site.

Fresh data analysis, on a 13 000 cores supercomputer revealed
that the structure was an artefact.

Thanks to HPC, 80 M\$ saved



In the **mid-90's**, **only 40%** of deposits fulfilled their promises.
Numerical simulations that analyse data obtained by seismic
echography have radically changed the playing field. Armed with
the new supercomputer ... , the Total engineers are **now** hitting
the bull's eye **in 60-70% of cases."**

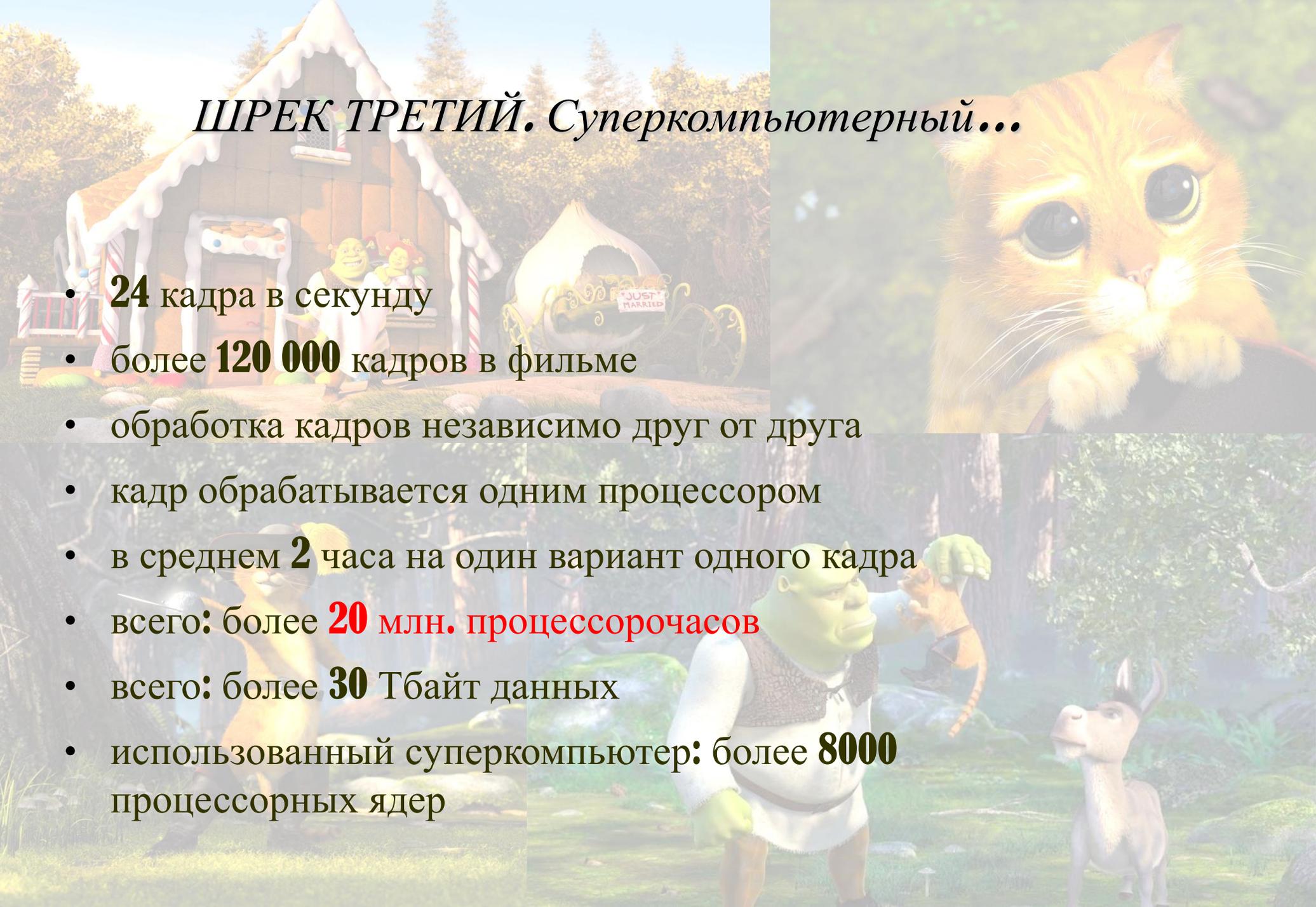
Journal La Recherche, special HPC, July 2009,

Acknowledgements: H. Calandra, Ph. Ricoux (TOTAL)



ШРЕК ТРЕТИЙ. Суперкомпьютерный...

- **24** кадра в секунду
- более **120 000** кадров в фильме
- обработка кадров независимо друг от друга
- кадр обрабатывается одним процессором
- в среднем **2** часа на один вариант одного кадра
- всего: более **20 млн. процессорочасов**
- всего: более **30** Тбайт данных
- использованный суперкомпьютер: более **8000** процессорных ядер



CGF — студия визуальных эффектов и анимации в России

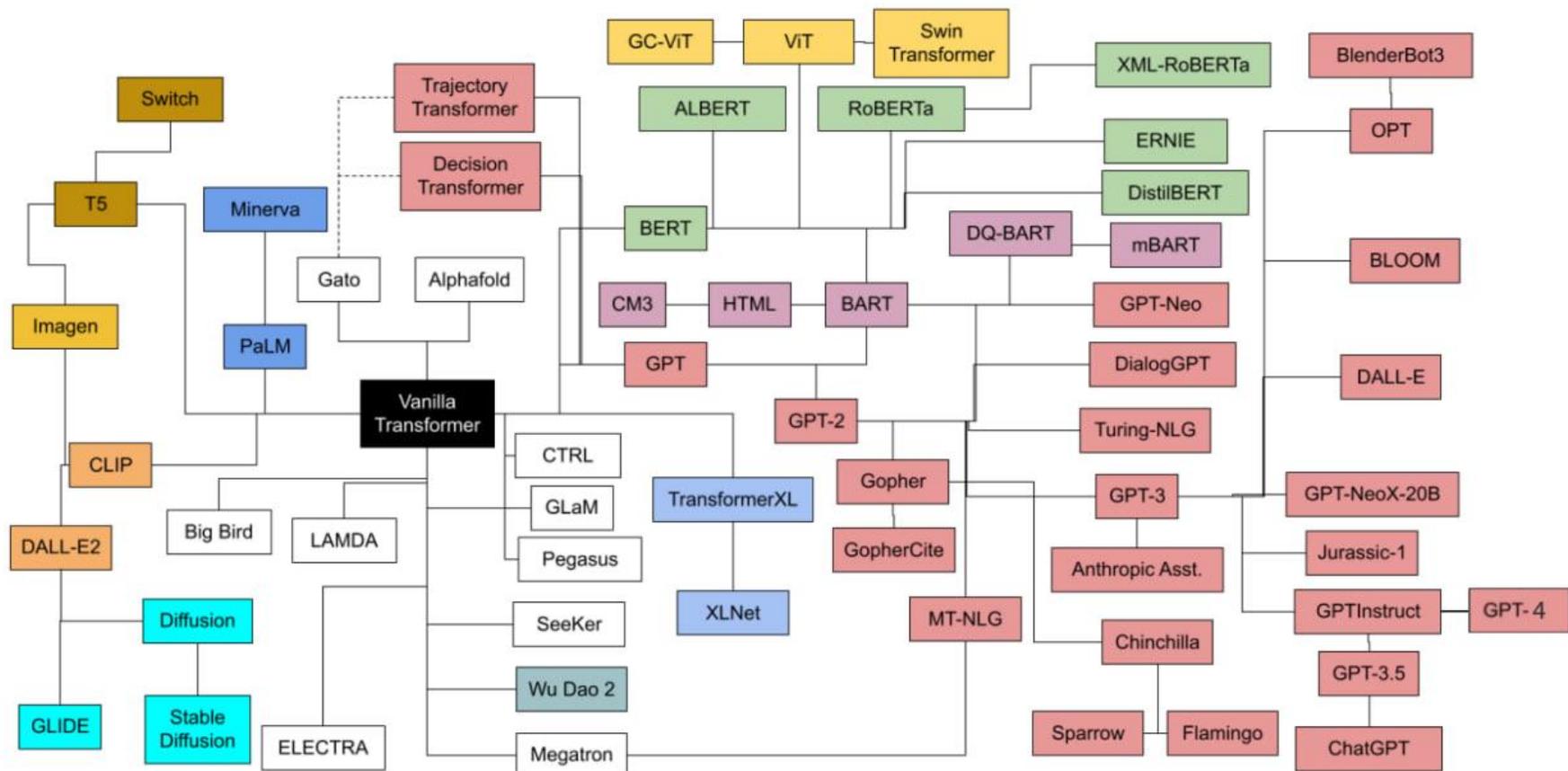


5 сентября
Москва, 2024

Суперкомпьютеры и параллельная обработка данных

33 из 316

Сверхвысокая производительность - зачем?



Transformer models: an introduction and catalog

Xavier Amatriain, Ananth Sankar, Jie Bing, Praveen Kumar Bodigutla, Timothy J. Hazen, Michael Kazi

<https://doi.org/10.48550/arXiv.2302.07730>

Сверхвысокая производительность - зачем?

ruDALL-E - сеть для создания изображения на основе текстового описания на русском языке (<https://rudalle.ru/>)

«шахматная ладья из изумрудного материала»



«кошка, одетая в корону»



Сверхвысокая производительность - зачем?

ruDALL-E - сеть для создания изображения на основе текстового описания на русском языке (<https://rudalle.ru/>)

Создание изображений происходит в три этапа: сначала одна нейросеть принимает текст на вход и генерирует заданное число картинок, затем следующая выбирает наиболее удачные из них и соответствующие описанию, а третья увеличивает их в размере без потери качества.

Два варианта модели:

- ruDALL-E XL, которая содержит 1,3 миллиарда параметров;
- ruDALL-E 12B с 12 миллиардами параметров.

Обучение заняло 23 тысячи GPU-часов на массиве данных из 120 миллионов пар «текст-изображение»

Сверхвысокая производительность - зачем?

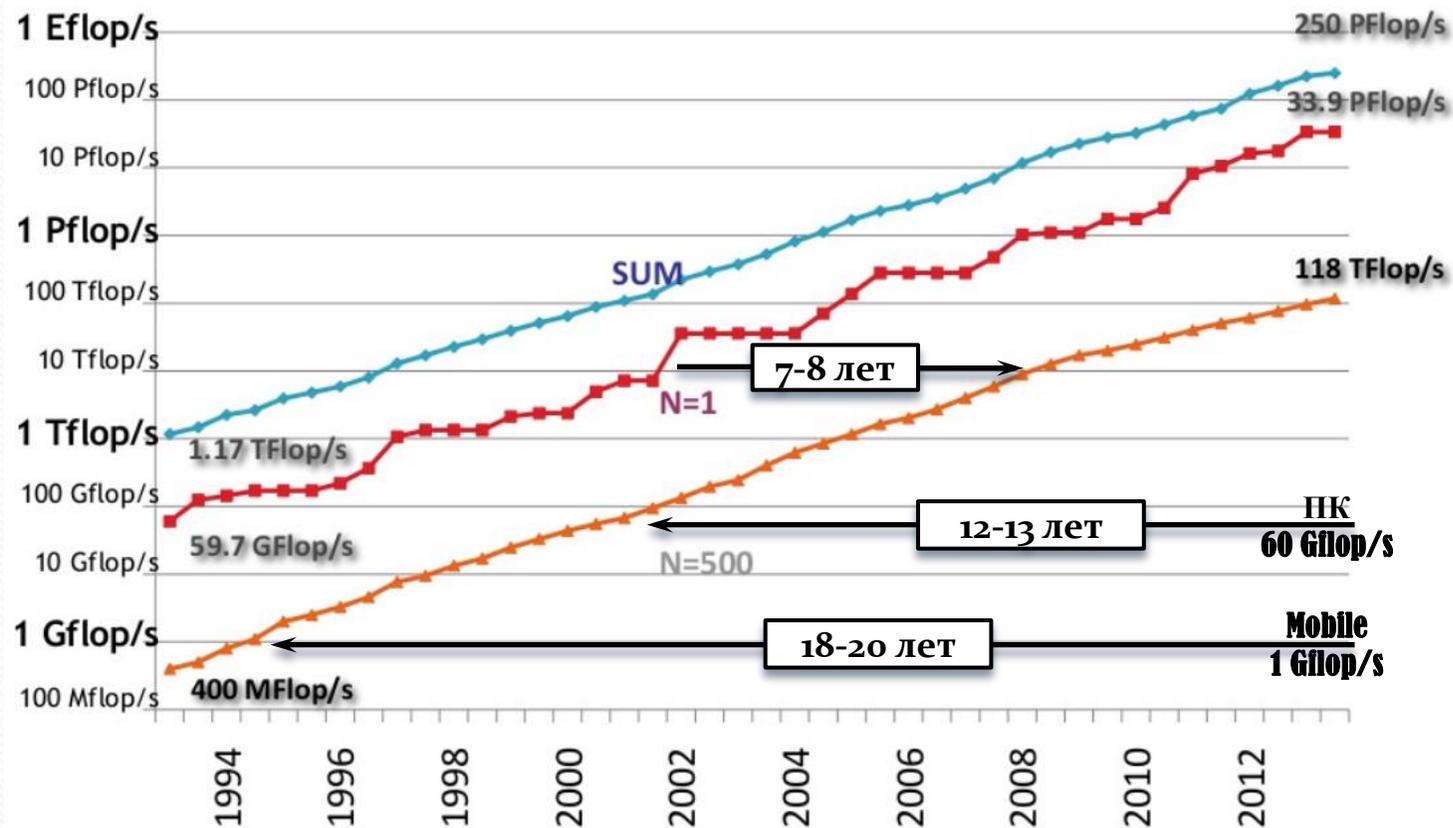
Model	Total train compute (PF-days)	Total train compute (flops)	Params (M)	Training tokens (billions)	Flops per param per token	Mult for bwd pass	Fwd-pass flops per active param per token	Frac of params active for each token
T5-Small	2.08E+00	1.80E+20	60	1,000	3	3	1	0.5
T5-Base	7.64E+00	6.60E+20	220	1,000	3	3	1	0.5
T5-Large	2.67E+01	2.31E+21	770	1,000	3	3	1	0.5
T5-3B	1.04E+02	9.00E+21	3,000	1,000	3	3	1	0.5
T5-11B	3.82E+02	3.30E+22	11,000	1,000	3	3	1	0.5
BERT-Base	1.87E+00	1.64E+20	109	250	6	3	2	1.0
BERT-Large	6.76E+00	5.33E+20	355	250	6	3	2	1.0
RoBERTa-Base	7.74E+01	1.50E+21	125	2,000	6	3	2	1.0
RoBERTa-Large	2.92E+01	4.26E+21	355	2,000	6	3	2	1.0
GPT-3 Small	1.60E+00	2.25E+20	125	300	6	3	2	1.0
GPT-3 Medium	7.42E+00	6.41E+20	356	300	6	3	2	1.0
GPT-3 Large	1.58E+01	1.37E+21	760	300	6	3	2	1.0
GPT-3 XL	2.75E+01	2.38E+21	1,320	300	6	3	2	1.0
GPT-3 2.7B	5.52E+01	4.77E+21	2,650	300	6	3	2	1.0
GPT-3 6.7B	1.39E+02	1.20E+22	6,660	300	6	3	2	1.0
GPT-3 13B	2.68E+02	2.31E+22	12,850	300	6	3	2	1.0
GPT-3 175B	3.64E+03	3.14E+23	174,600	300	6	3	2	1.0



AI-модель для русского языка,
<https://developers.sber.ru/portal/products/rugpt-3>

- Ресурсы Ohio Supercomputer Center (OSC) используются для моделирования формирования звёзд.
https://www.osc.edu/press/supercomputing_becomes_the_star_of_astrophysics_class
- Суперкомпьютеры Frontier, установленный в Oak Ridge Leadership Computing Facility, и Frontera, установленный в Texas Advanced Computing Center, используются для моделирования землетрясений.
https://www.sdsc.edu/News%20Items/PR20240828_earthquake_sim_intern.html
- Суперкомпьютер LUMI используется для прогнозирования наводнений при таянии ледников.
<https://lumi-supercomputer.eu/forecasting-glacier-outburst-floods/>
- Суперкомпьютер Fugaku, установленный в RIKEN, используется для социальных исследований.
<https://tacc.utexas.edu/news/latest-news/2024/08/21/astronomers-use-ai-to-find-elusive-stars-gobbling-up-planets/>
- Суперкомпьютер Expanse, установленный в San Diego Supercomputer Center (SDSC), используется для создания новых методов лечения рака.
https://www.sdsc.edu/News%20Items/PR20240816_chemo_alternative.html
- Суперкомпьютер Piz Daint, установленный в Swiss National Supercomputing Center, используется для климатического моделирования.
<https://www.cscs.ch/science/earth-env-science/how-a-specific-bias-in-climate-models-got-uncovered-and-corrected>

Рост производительности



<http://linpack.hpc.msu.ru/>

Важные сокращения

Мега (Mega) – 10^6 (миллион)

Гига (Giga) – 10^9 (биллион / миллиард)

Тера (Tera) – 10^{12} (триллион)

Пета (Peta) – 10^{15} (квадриллион)

Экса (Exa) – 10^{18} (квинтиллион)

Флоп/с, *Flop/s* – *Floating point operations per second*

$15 \text{ Tflop/s} = 15 * 10^{12}$ арифметических операций в секунду над вещественными данными, представленными в форме с плавающей точкой.

Важные сокращения

Мега (Mega) – 10^6 (миллион)

Гига (Giga) – 10^9 (биллион / миллиард)

Тера (Tera) – 10^{12} (триллион)

Пета (Peta) – 10^{15} (квадриллион)

Экса (Exa) – 10^{18} (квинтиллион)

Флоп/с, Flop/s – *Floating point operations per second*

$15 \text{ Tflop/s} = 15 * 10^{12}$ **арифметических** операций в секунду над **вещественными** данными, представленными в форме с **плавающей точкой**.

Годы, флопсы и степень параллелизма (когда и как был достигнут очередной 'X'flops)

10^6 Mflops	1964 Г.	CDC 6600	10 MHz	1 CPUs
10^9 Gflops	1985 Г.	Cray 2	125 MHz	8 CPUs
10^{12} Tflops	1997 Г.	ASCI Red	200 MHz	9152 CPUs
10^{15} Pflops	2008 Г.	Roadrunner	3,2 GHz	122400 Cores
10^{18} Eflops	2022 Г.	Frontier	2 GHz CPUs (606,208 cores) and 37,888 GPUs (8,335,360 cores)	

Увеличение производительности компьютеров: за счет чего?

EDSAC, 1949 год
год

Cray Titan, 2012

изменение

такт: $2 \cdot 10^{-6}$ с

$\approx 4.4 \cdot 10^3$

$4.5 \cdot 10^{-10}$ с (2.2 GHz)

произв.: 10^2 оп/с

$\approx 1.7 \cdot 10^{14}$

$1.7 \cdot 10^{16}$ оп/с

Время такта = $1/(\text{тактовая частота})$

Увеличение производительности компьютеров: за счет чего?

*EDSAC, 1949 год
год*

Cray Titan, 2012

изменение

такт: $2 \cdot 10^{-6}$ с

$\approx 4.4 \cdot 10^3$

$4.5 \cdot 10^{-10}$ с (2.2 GHz)

произв.: 10^2 оп/с

$\approx 1.7 \cdot 10^{14}$

$1.7 \cdot 10^{16}$ оп/с

Два вывода.

1. Безусловно, без развития элементной базы не было бы такого прогресса в развитии компьютеров.

2. Но основной вклад в увеличении производительности компьютеров – это развитие архитектуры, и прежде всего, за счет глубокого внедрения идей параллелизма

Тенденции развития современных процессоров

В течение нескольких десятилетий развитие ЭВМ сопровождалось удвоением их быстродействия каждые 1.5-2 года. Это обеспечивалось и повышением тактовой частоты и совершенствованием архитектуры (параллельное и конвейерное выполнение команд).

Узким местом стала оперативная память. Знаменитый закон Мура, так хорошо работающий для процессоров, совершенно не применим для памяти, где скорости доступа удваиваются в лучшем случае каждые 5-6 лет.

Совершенствовались системы кэш-памяти, увеличивался объем, усложнялись алгоритмы ее использования.

Для процессора Intel Itanium:

Latency to L1: 1-2 cycles

Latency to L2: 5 - 7 cycles

Latency to L3: 12 - 21 cycles

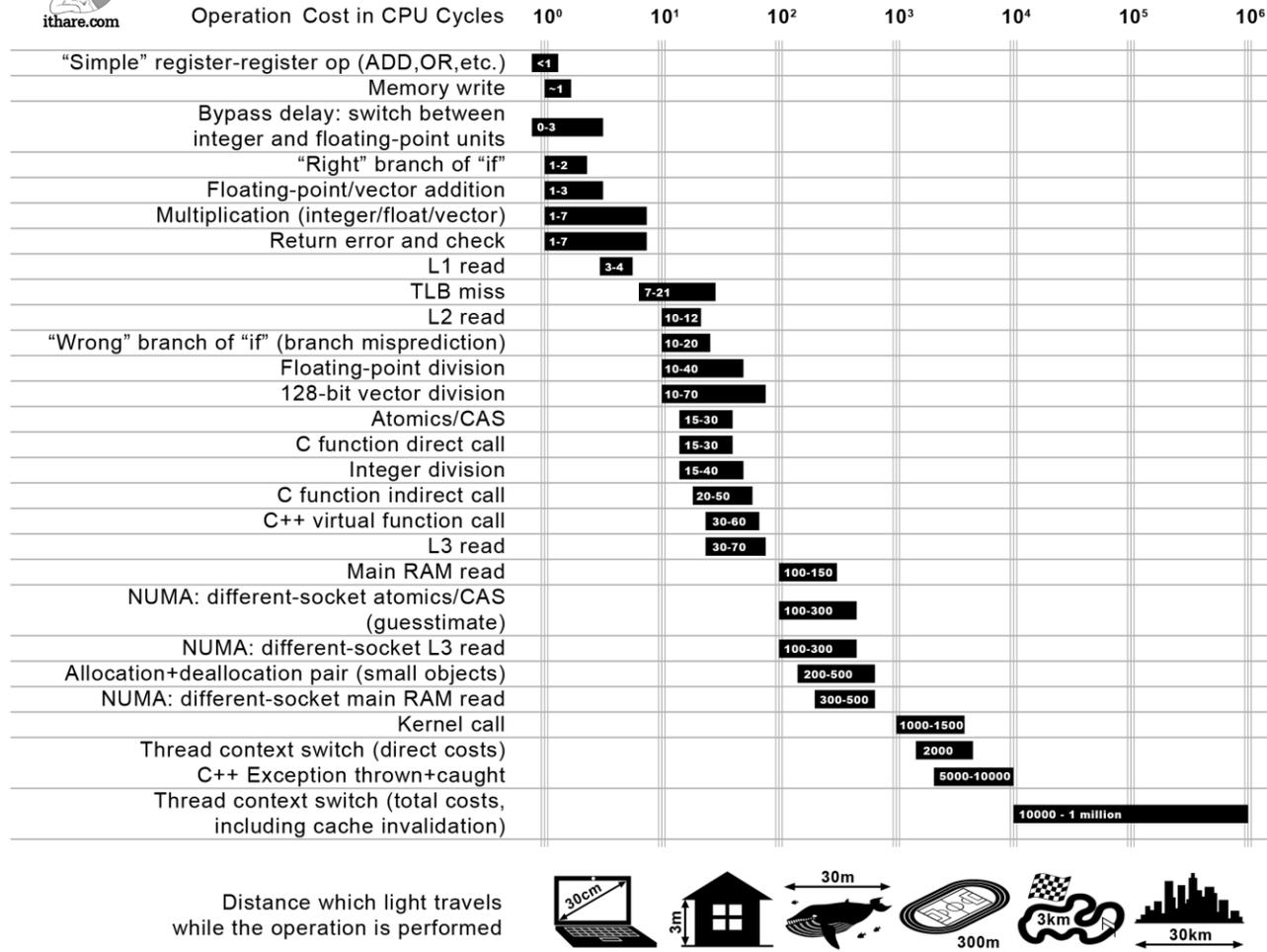
Latency to memory: 180 – 225 cycles

Важным параметром становится - **GUPS** (Giga Updates Per Second)

Тенденции развития современных процессоров



Not all CPU operations are created equal

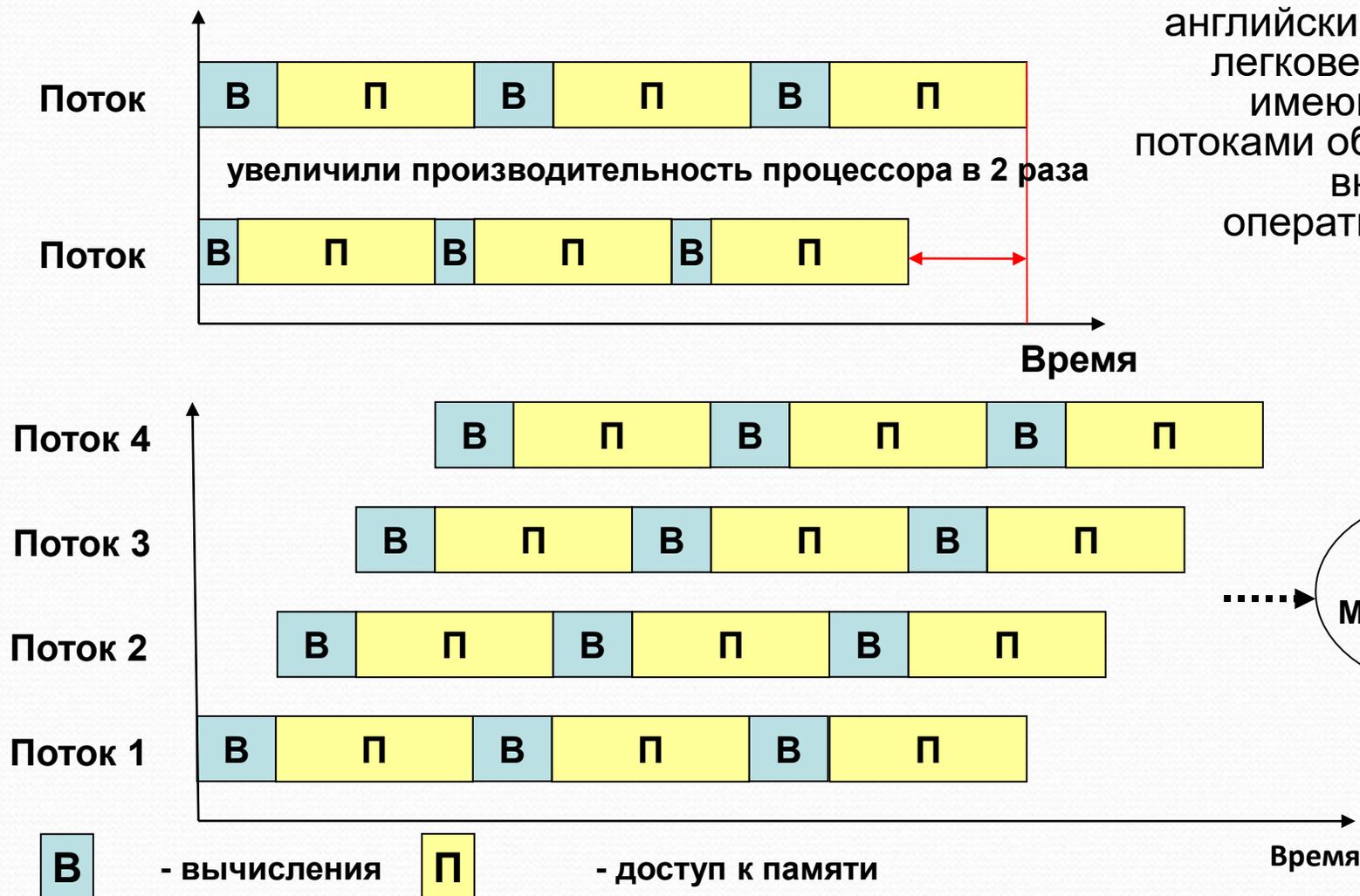


<http://ithare.com/infographics-operation-costs-in-cpu-clock-cycles/>

https://www.agner.org/optimize/instruction_tables.pdf

Тенденции развития современных процессоров

Поток или **нить** (по-английски “thread”) – это легковесный процесс, имеющий с другими потоками общие ресурсы, включая общую оперативную память.

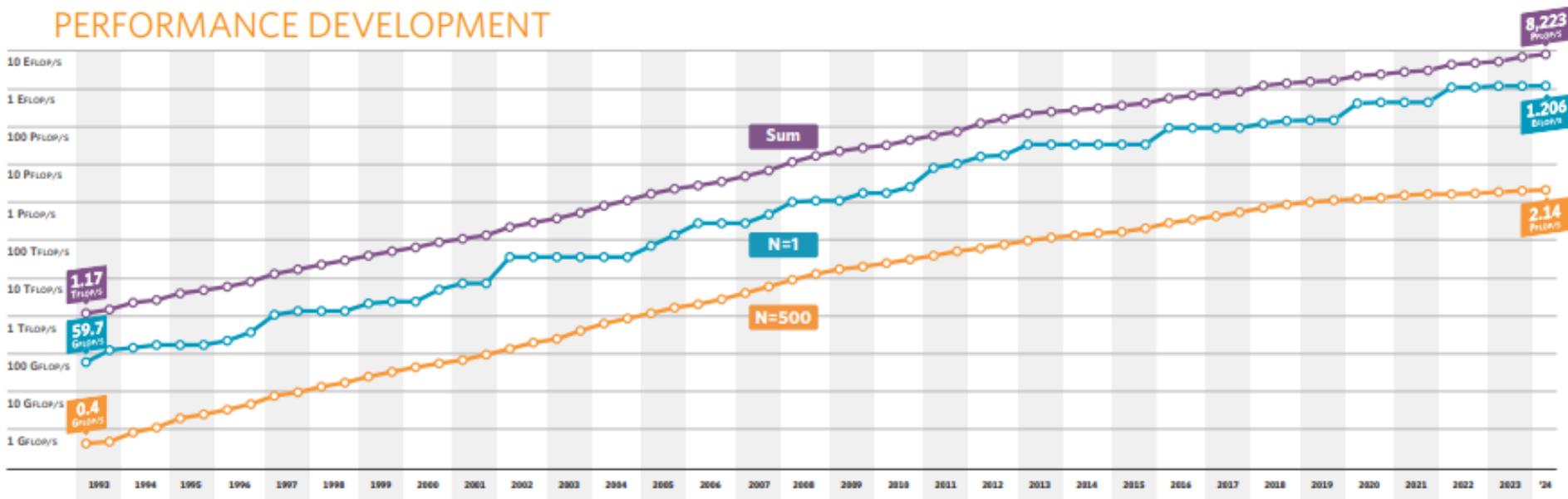


Суперкомпьютерные системы (Top500)

MAY 2024

			SITE	COUNTRY	CORES	RMAX PFL0P/S	POWER MW
1	Frontier	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	DOE/SC/ORNL	USA	8,699,904	1,206.0	22.7
2	Aurora	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 (52C 2.4GHz), Intel Data Center GPU Max, Slingshot-11	DOE/SC/ANL	USA	9,264,128	1,012.0	38.7
3	Eagle	Microsoft NDv5, Xeon Platinum 8480C (48C 2GHz), NVIDIA H100, NVIDIA Infiniband NDR	Microsoft Azure	USA	1,123,200	561.2	
4	Fugaku	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
5	LUMI	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	EuroHPC/CSC	Finland	2,220,288	379.7	6.01

PERFORMANCE DEVELOPMENT



№ 59 в Top 500

Суперкомпьютер MARU, ThinkSystem SD650 V2, Xeon Platinum 8368Q 38C 2.6GHz, Infiniband HDR

- ❑ Пиковая производительность – 25495,14 TFlop/s
- ❑ Число ядер в системе — 306 432
- ❑ Производительность на Linpack - 16753 TFlop/s (65.71 % от пиковой)
- ❑ Энергопотребление комплекса - **15414 кВт**

№ 13 в Top 500

Суперкомпьютер Sunway TaihuLight, Sunway MPP, SW26010 260C 1.45GHz, Custom interconnect

- ❑ Пиковая производительность – 125435.9 TFlop/s
- ❑ Число ядер в системе — 10 649 600
- ❑ Производительность на Linpack – 93014.5 TFlop/s (74.15 % от пиковой)
- ❑ Энергопотребление комплекса - **15371 кВт**

Важным параметром становится – **Power Efficiency (GFlops/watt)**
6,05 VS 1,09

Как добиться максимальной производительности на Ватт => Chip
MultiProcessing, многоядерность.

Тенденции развития современных процессоров



ShenWei SW26010

64-разрядный RISC-процессор с поддержкой SIMD-инструкций и внеочередным исполнением команд

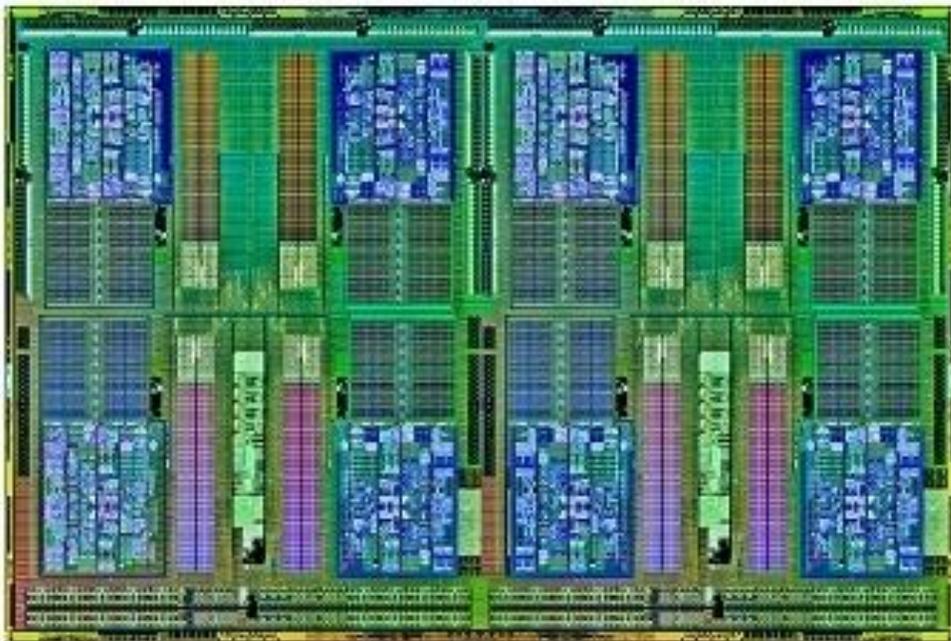
Изготовлен по схеме, предусматривающей использование четырех кластеров с 64 вычислительными ядрами (CPE) и одним управляющим ядром (MPE) в каждом.

В каждом кластере также имеется собственный контроллер памяти, суммарная пропускная способность на один процессорный разъем достигает 136,5 ГБ/с.

На каждое ядро выделено 12 КБ кэш-памяти инструкций и 64 КБ кэш-памяти данных.

Рабочая частота процессора - 1,45 ГГц.

Тенденции развития современных процессоров



AMD Opteron серии 6300

6380 SE 16 ядер @ 2,5 ГГц, 16 МБ L3 Cache

6348 12 ядер @ 2,8 ГГц, 16 МБ L3 Cache

6328 8 ядер @ 3,2 ГГц, 16 МБ L3 Cache

6308 4 ядра @ 3,5 ГГц, 16 МБ L3 Cache

технология AMD Turbo CORE

встроенный контроллер памяти (4 канала памяти DDR3)

4 канала «точка-точка» с использованием HyperTransport 3.0

AMD EPYC 7003 Series Processors

AMD EPYC™ 7763

of CPU Cores 64
of Threads 128
Max Boost Clock Up to 3.5GHz
Base Clock 2.45GHz
Default TDP / TDP 280W

AMD EPYC™ 75F3

of CPU Cores 32
of Threads 64
Max Boost Clock Up to 4.0GHz
Base Clock 2.95GHz
Default TDP / TDP 280W

AMD EPYC™ 7713

of CPU Cores 64
of Threads 128
Max Boost Clock Up to 3.6GHz
Base Clock 2.0GHz
Default TDP / TDP 225W

AMD EPYC™ 7643

of CPU Cores 48
of Threads 96
Max Boost Clock Up to 3.6GHz
Base Clock 2.3GHz
Default TDP / TDP 225W

AMD EPYC™ 7543

of CPU Cores 32
of Threads 64
Max Boost Clock Up to 3.7GHz
Base Clock 2.8GHz
Default TDP / TDP 225W

<https://www.amd.com/en/processors/epyc-7003-series>

Процессоры AMD EPYC серии Milan-X

Процессор	Ядер/Потоков	Базовая частота	Турбо	TDP	Кэш L3 (L3 + 3D V-Cache)
EPYC 7773X	64/128	2,2 ГГц	3,5 ГГц	280 Вт	768 МБ
EPYC 7573X	32/64	2,8 ГГц	3,6 ГГц	280 Вт	768 МБ
EPYC 7473X	24/48	2,8 ГГц	3,7 ГГц	240 Вт	768 МБ
EPYC 7373X	16/32	3,05 ГГц	3,8 ГГц	240 Вт	768 МБ



Процессоры AMD EPYC серии 9004

Name	# of CPU Cores	# of Threads	Max. Boost Clock	Base Clock	L3 Cache	Default TDP
AMD EPYC™ 9754S	128	128	Up to 3.1 GHz	2.25 GHz	256 MB	360W
AMD EPYC™ 9754	128	256	Up to 3.1 GHz	2.25 GHz	256 MB	360W
AMD EPYC™ 9734	112	224	Up to 3 GHz	2.2 GHz	256 MB	340W
AMD EPYC™ 9684X	96	192	Up to 3.7 GHz	2.55 GHz	1152 MB	400W
AMD EPYC™ 9654P	96	192	Up to 3.7 GHz	2.4 GHz	384 MB	360W
AMD EPYC™ 9654	96	192	Up to 3.7 GHz	2.4 GHz	384 MB	360W
AMD EPYC™ 9634	84	168	Up to 3.7 GHz	2.25 GHz	384 MB	290W
AMD EPYC™ 9554P	64	128	Up to 3.75 GHz	3.1 GHz	256 MB	360W
AMD EPYC™ 9554	64	128	Up to 3.75 GHz	3.1 GHz	256 MB	360W
AMD EPYC™ 9534	64	128	Up to 3.7 GHz	2.45 GHz	256 MB	280W
AMD EPYC™ 9474F	48	96	Up to 4.1 GHz	3.6 GHz	256 MB	360W
AMD EPYC™ 9454P	48	96	Up to 3.8 GHz	2.75 GHz	256 MB	290W
AMD EPYC™ 9454	48	96	Up to 3.8 GHz	2.75 GHz	256 MB	290W
AMD EPYC™ 9384X	32	64	Up to 3.9 GHz	3.1 GHz	768 MB	320W
AMD EPYC™ 9374F	32	64	Up to 4.3 GHz	3.85 GHz	256 MB	320W
AMD EPYC™ 9354P	32	64	Up to 3.8 GHz	3.25 GHz	256 MB	280W
AMD EPYC™ 9354	32	64	Up to 3.8 GHz	3.25 GHz	256 MB	280W
AMD EPYC™ 9334	32	64	Up to 3.9 GHz	2.7 GHz	128 MB	210W

Тенденции развития современных процессоров

Intel Xeon Processor серии E5

E5-2699 v4 (55M Cache, 2.20 GHz) 22 ядра, 44 нити

E5-2698 v4 (50M Cache, 2.20 GHz) 20 ядер, 40 нитей

E5-2697 v4 (45M Cache, 2.30 GHz) 18 ядер, 36 нитей

E5-2697A v4 (40M Cache, 2.60 GHz) 16 ядер, 32 нити

E5-2667 v4 (25M Cache, 3.20 GHz) 8 ядер, 16 нитей

Intel® Turbo Boost

Intel® Hyper-Threading

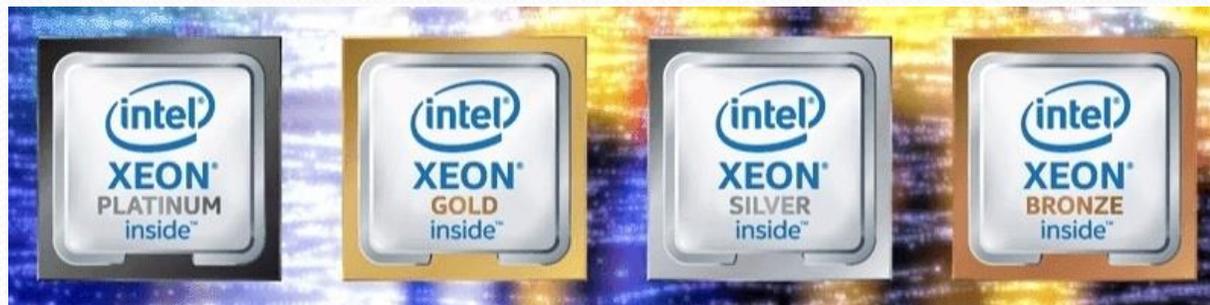
Intel® Intelligent Power

Intel® QuickPath



3rd Generation Intel Xeon Scalable Processors

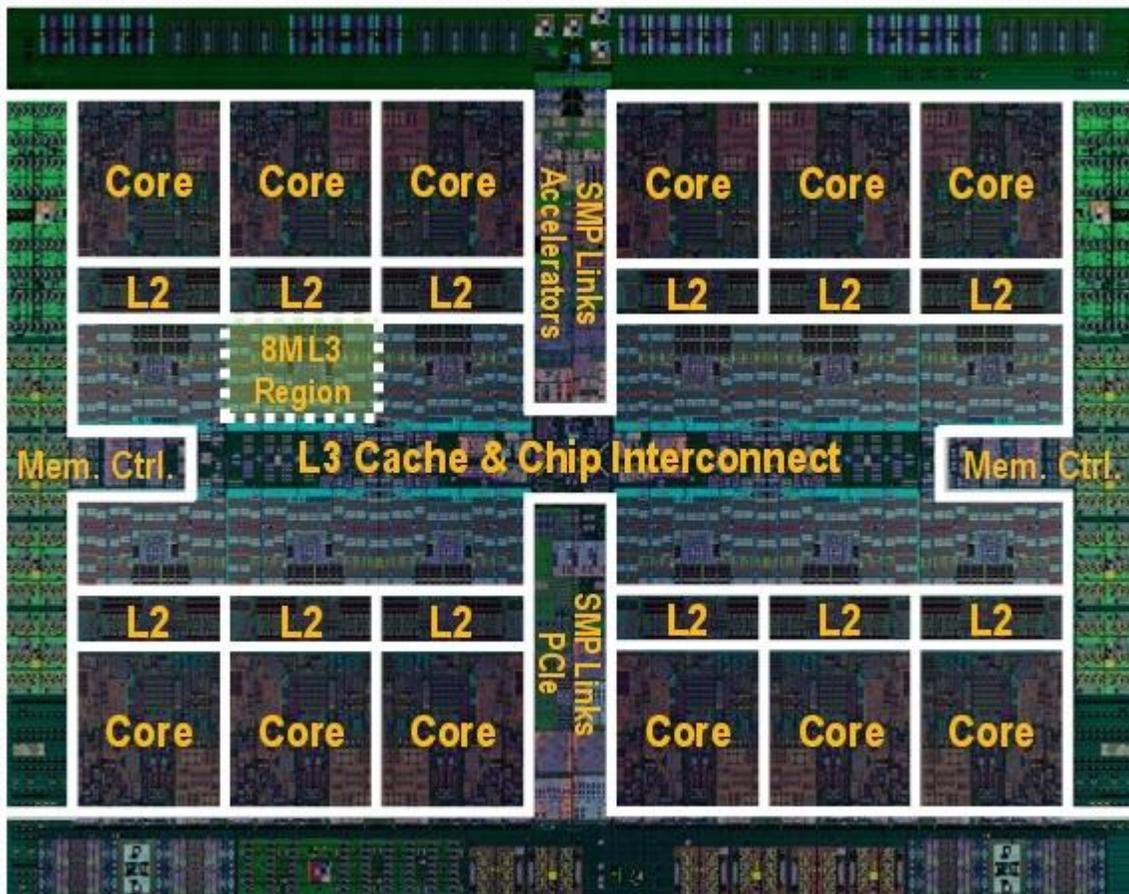
Processor	Launch Date	# of cores	Max Turbo Frequency	Processor Base Frequency	Cache	TDP
Intel Xeon Platinum 8368	Q2'21	38	3.40 GHz	2.40 GHz	57 MB	270 W
Intel Xeon Platinum 8368Q	Q2'21	38	3.70 GHz	2.60 GHz	57 MB	270 W
Intel Xeon Platinum 8380	Q2'21	40	3.40 GHz	2.30 GHz	60 MB	270 W
Intel Xeon Platinum 8360Y	Q2'21	36	3.50 GHz	2.40 GHz	54 MB	250 W
Intel Xeon Platinum 8358	Q2'21	32	3.40 GHz	2.60 GHz	48 MB	250 W
Intel Xeon Platinum 8380H	Q2'20	28	4.30 GHz	2.90 GHz	38.5 MB	250 W



Intel Xeon 6 Processors

Product Name	Total Cores	Max Turbo Frequency	Processor Base Frequency	Cache
Intel® Xeon® 6780E Processor (108M Cache, 2.20 GHz)	144	3 GHz	2.2 GHz	108 MB
Intel® Xeon® 6766E Processor (108M Cache, 1.90 GHz)	144	2.7 GHz	1.9 GHz	108 MB
Intel® Xeon® 6756E Processor (96M Cache, 1.80 GHz)	128	2.6 GHz	1.8 GHz	96 MB
Intel® Xeon® 6746E Processor (96M Cache, 2.00 GHz)	112	2.7 GHz	2 GHz	96 MB
Intel® Xeon® 6740E Processor (96M Cache, 2.40 GHz)	96	3.2 GHz	2.4 GHz	96 MB
Intel® Xeon® 6731E Processor (96M Cache, 2.20 GHz)	96	3.1 GHz	2.2 GHz	96 MB
Intel® Xeon® 6710E Processor (96M Cache, 2.40 GHz)	64	3.2 GHz	2.4 GHz	96 MB

Тенденции развития современных процессоров

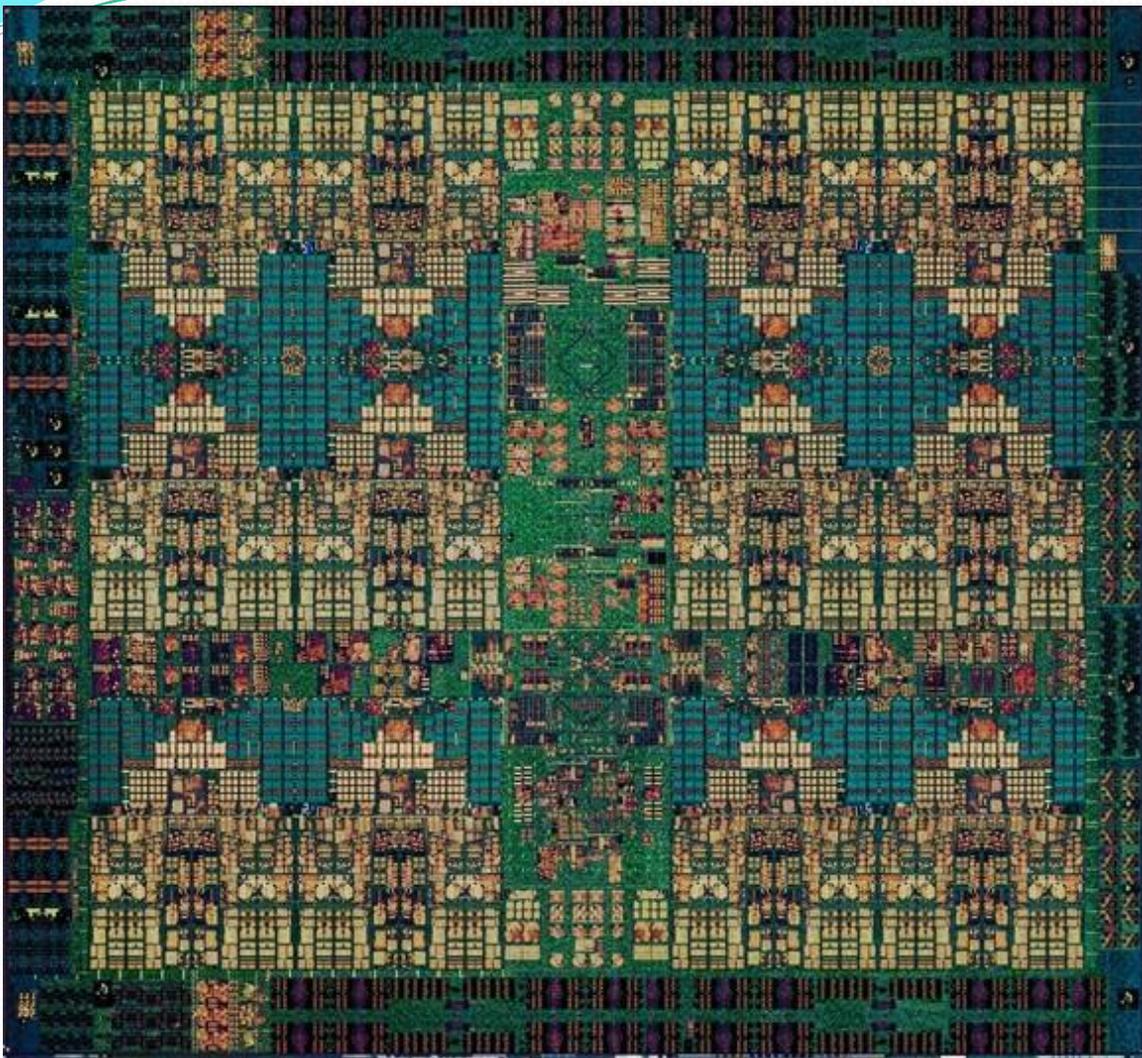


IBM Power8

- 2,75 – 4,2 ГГц
- 12 ядер x 8 нитей
Simultaneous
MultiThreading
- 64 КБ Data Cache +
32КБ instruction Cache
- L2 512 КБ
- L3 96 МБ

www.idh.ch/IBM_TU_2013/Power8.pdf

Тенденции развития современных процессоров

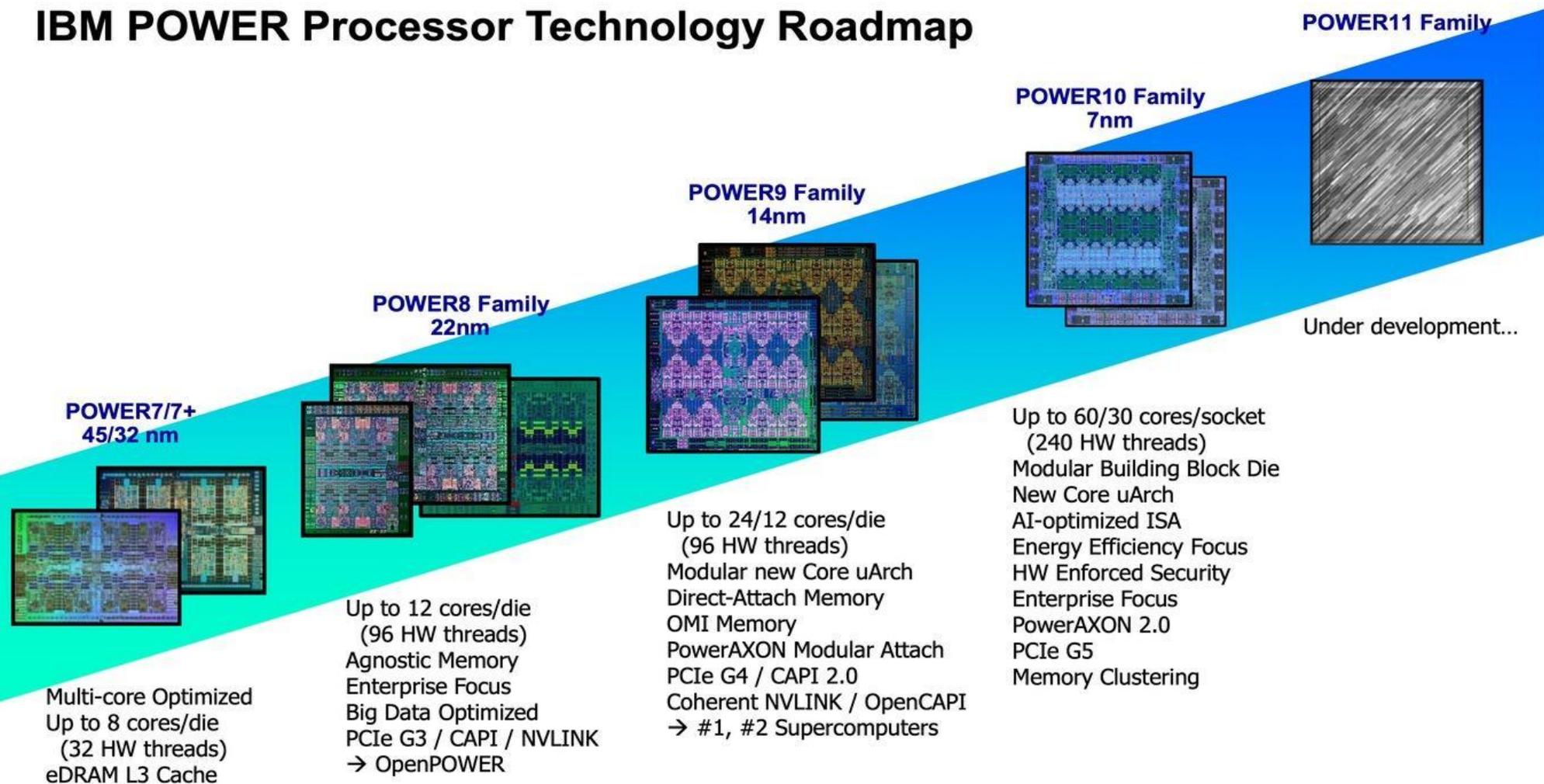


IBM Power9

- 2,75 – 4,2 ГГц
- 12 ядер x 8 нитей
24 ядра x 4 нити
- L2 512 КБ
- L3 120 МБ (10 МБ на 2 ядра)

Тенденции развития современных процессоров

IBM POWER Processor Technology Roadmap



Тенденции развития современных процессоров

POWER10 Processor Chip

Technology and Packaging:

- 602mm² 7nm Samsung (18B devices)
- 18 layer metal stack, enhanced device
- Single-chip or Dual-chip sockets

Computational Capabilities:

- Up to 15 SMT8 Cores (2 MB L2 Cache / core)
(Up to 120 simultaneous hardware threads)
- Up to 120 MB L3 cache (low latency NUCA mgmt)
- 3x energy efficiency relative to POWER9
- Enterprise thread strength optimizations
- AI and security focused ISA additions
- 2x general, 4x matrix SIMD relative to POWER9
- EA-tagged L1 cache, 4x MMU relative to POWER9

Open Memory Interface:

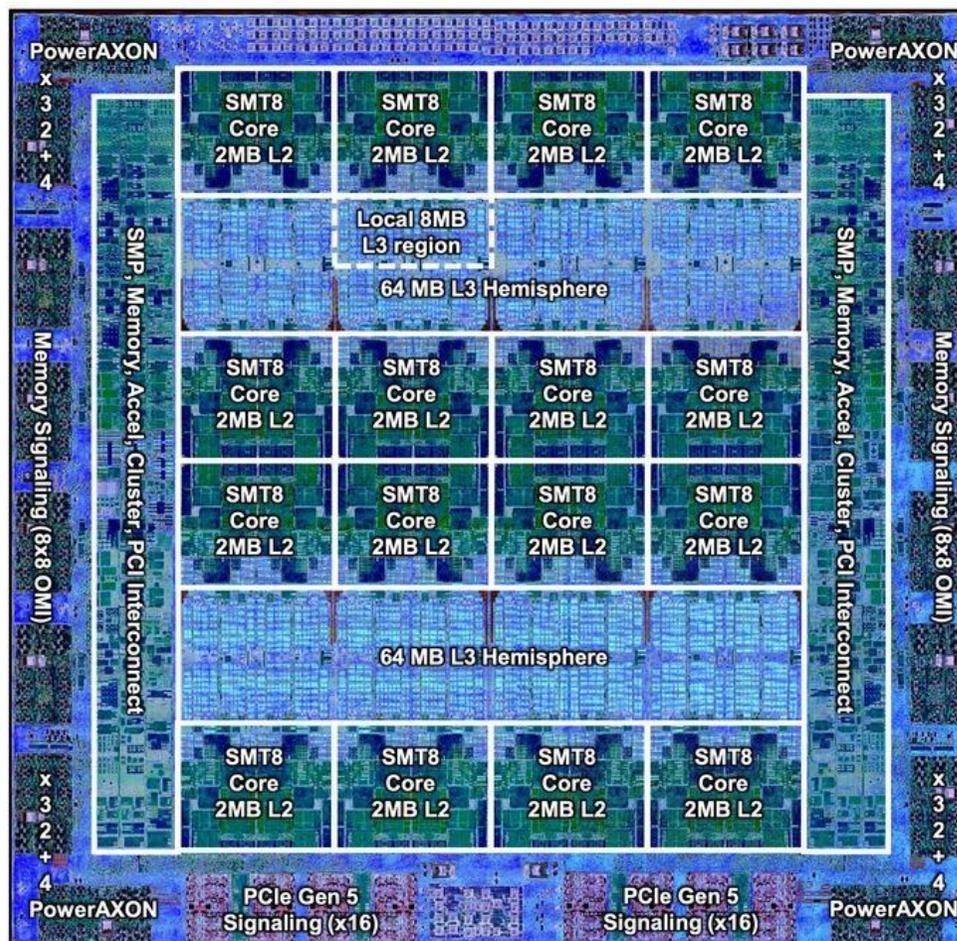
- 16 x8 at up to 32 GT/s (1 TB/s)
- Technology agnostic support: near/main/storage tiers
- Minimal (< 10ns latency) add vs DDR direct attach

PowerAXON Interface:

- 16 x8 at up to 32 GT/s (1 TB/s)
- SMP interconnect for up to 16 sockets
- OpenCAPI attach for memory, accelerators, I/O
- Integrated clustering (memory semantics)

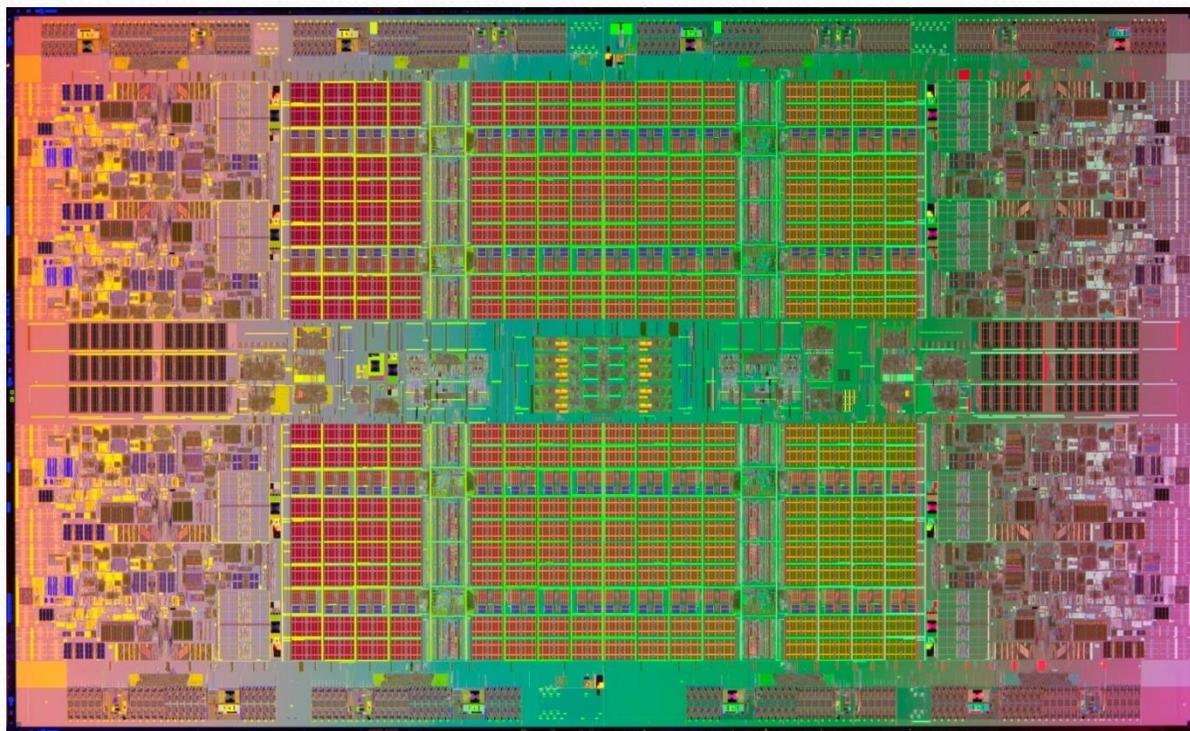
PCIe Gen 5 Interface:

- x64 / DCM at up to 32 GT/s



Die Photo courtesy of Samsung Foundry

Тенденции развития современных процессоров



Intel Itanium серии 9500

9560 8 ядер @ 2,53 ГГц, 16 нитей, 32 МБ L3 Cache

9550 4 ядра @ 2,40 ГГц, 8 нитей, 32 МБ L3 Cache

Тенденции развития современных процессоров

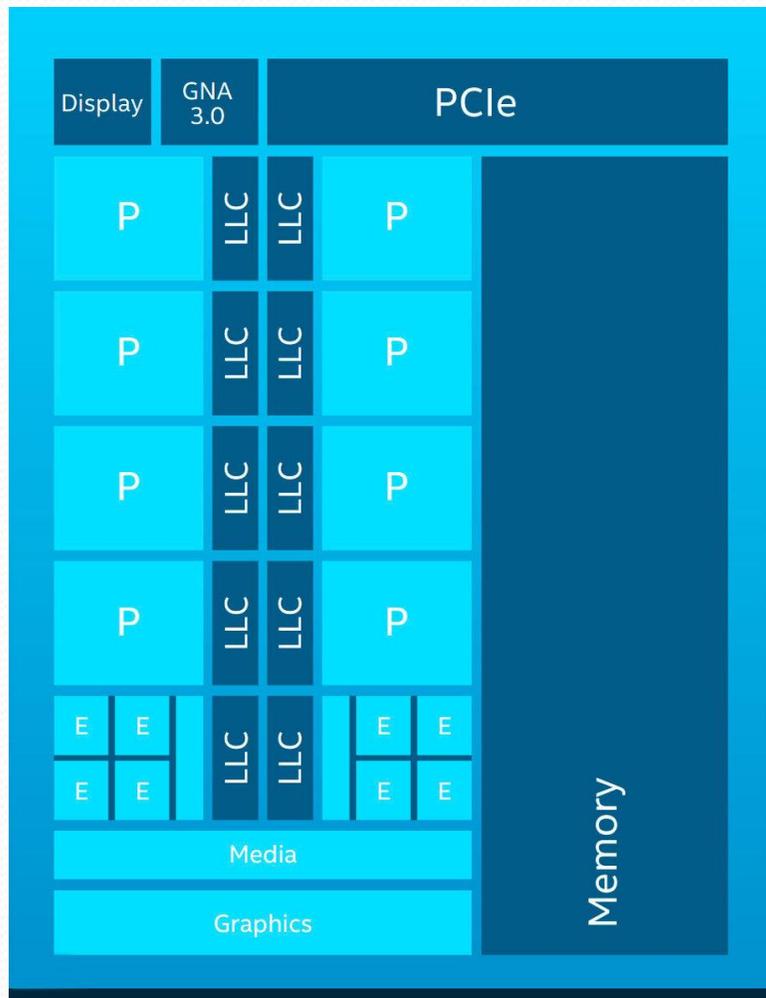
«Huawei Kunpeng 920 is the industry's leading-edge Arm-based server CPU».



Key Features

Architecture	ARMv8.2
Core	Up to 64
Typical Frequency	2.6 GHz / 3.0 GHz
Memory	8 DDR4 channels
I/O	PCIe 4.0, CCIX, 100G, SAS/SATA 3.0
Max Power	180 W
Process	7 nm

Тенденции развития современных процессоров



P — Performance (производительные) ядра предназначены для высокопроизводительных вычислений (могут быть многопоточными) и E — Efficiency (эффективные) ядра, ориентированные на максимальную производительность на ватт

Тенденции развития современных процессоров

13th Gen Intel® Core™ Desktop Processors

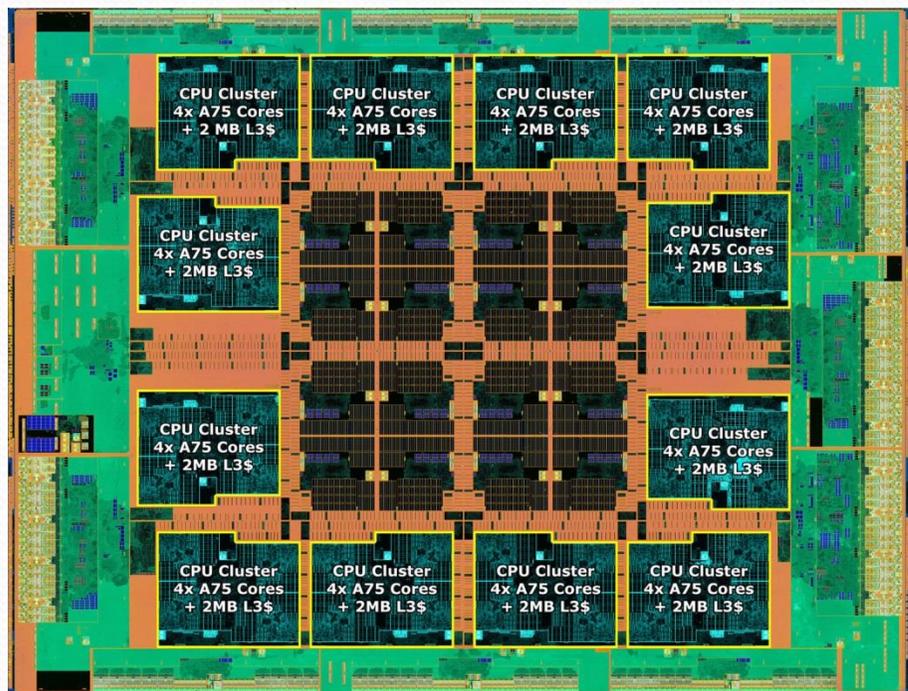
Processor Number	Processor Cores (P+E)	Processor Threads	Intel® Smart Cache (L3)	Total L2 Cache	P-core Max Turbo Frequency (GHz)	E-core Max Turbo Frequency (GHz)	P-core Base Frequency (GHz)	E-core Base Frequency (GHz)	Processor Graphics	Total CPU PCIe Lanes	Max Memory Speed (MT/S)	Memory Capacity	Processor Base Power (W)	Max Turbo Power (W)	RCP (USD\$)
i9-13900	24 (8+16)	32	36MB	32MB	Up to 5.6	Up to 4.2	2.0	1.5	Intel® UHD Graphics 770	20	DDR5 5600 DDR4 3200	128GB	65	219	\$549
i9-13900F	24 (8+16)	32	36MB	32MB	Up to 5.6	Up to 4.2	2.0	1.5	n/a	20	DDR5 5600 DDR4 3200	128GB	65	219	\$524
i7-13700	16 (8+8)	24	30MB	24MB	Up to 5.2	Up to 4.1	2.1	1.5	Intel® UHD Graphics 770	20	DDR5 5600 DDR4 3200	128GB	65	219	\$384
i7-13700F	16 (8+8)	24	30MB	24MB	Up to 5.2	Up to 4.1	2.1	1.5	n/a	20	DDR5 5600 DDR4 3200	128GB	65	219	\$359
i5-13600	14 (6+8)	20	24MB	11.5MB	Up to 5.0	Up to 3.7	2.7	2.0	Intel® UHD Graphics 770	20	DDR5 4800 DDR4 3200	128GB	65	154	\$255
i5-13500	14 (6+8)	20	24MB	11.5MB	Up to 4.8	Up to 3.5	2.5	1.8	Intel® UHD Graphics 770	20	DDR5 4800 DDR4 3200	128GB	65	154	\$232
i5-13400	10 (6+4)	16	20MB	9.5MB	Up to 4.6	Up to 3.3	2.5	1.8	Intel® UHD Graphics 730	20	DDR5 4800 DDR4 3200	128GB	65	148	\$221
i5-13400F	10 (6+4)	16	20MB	9.5MB	Up to 4.6	Up to 3.3	2.5	1.8	n/a	20	DDR5 4800 DDR4 3200	128GB	65	148	\$196
i3-13100	4 (4+0)	8	12MB	5MB	Up to 4.5	n/a	3.4	n/a	Intel® UHD Graphics 730	20	DDR5 4800 DDR4 3200	128GB	60	89	\$134
i3-13100F	4 (4+0)	8	12MB	5MB	Up to 4.5	N/a	3.4	n/a	n/a	20	DDR5 4800 DDR4 3200	128GB	58	89	\$109



Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. The frequency of cores and core types varies by workload, power consumption and other factors. Visit <https://www.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html> for more information. Max Turbo Frequency for P-cores may include Intel® Thermal Velocity Boost and/or Intel Turbo Boost Max 3.0. All SKUs listed above support up to DDR5 (5600 MT/S)/DDR4 (3200 MT/S) memory. See ark.intel.com for more specification details.

Тенденции развития современных процессоров

Байкал-S



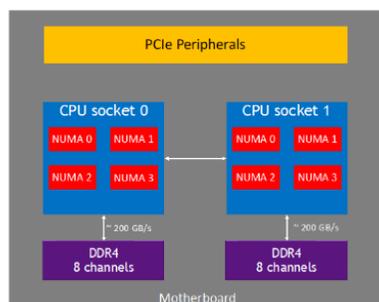
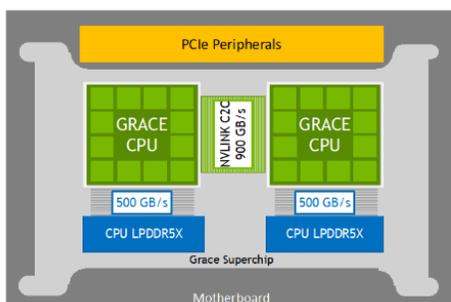
	Байкал-S	Kunpeng 920
Архитектура	aarch64	aarch64
ISA	ARMv8.2-A	ARMv8.2-A
Микроархитектура	Cortex-A75	TaiShan v110
Частота (МГц)	2000	2600
Ядра; Потоки	48	48
Техпроцесс (нм)	16	7
TDP (Вт)	100	150
Тип ОЗУ	DDR4-3200	DDR4-2933
Сокет	FCLGA-3467	BGA
Каналов ОЗУ	6	8
Макс ОЗУ (ГБ)	768	1024
ГФлопс (DP)	384	500
ГФлопс (SP)	768	2000
Год	2021	2019

Тенденции развития современных процессоров

NVIDIA Grace CPU Superchip



	Grace CPU Superchip
Core Architecture	Neoverse V2 Cores: Armv9 with 4x128b SVE2
Core Count	144
Cache	L1: 64KB I-cache + 64KB D-cache per core L2: 1MB per core L3: 234MB per superchip
Memory Technology	LPDDR5X with ECC, Co-Packaged
Raw Memory BW	Up to 1 TB/s
Memory Size	Up to 960GB
FP64 Peak	7.1 TFLOPS
PCI Express	8x PCIe Gen 5 x16 interfaces; option to bifurcate Total 1 TB/s PCIe Bandwidth. Additional low-speed PCIe connectivity for management.
Power	500W TDP with Memory, 12V Supply



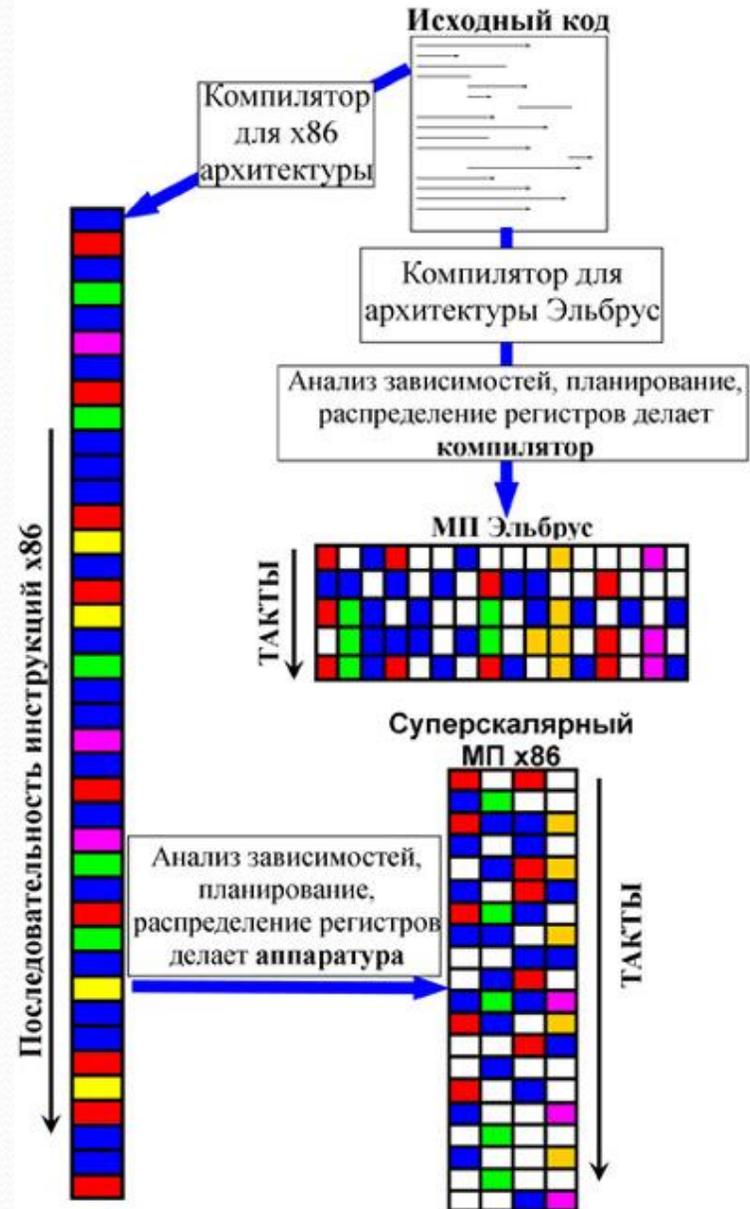
<https://resources.nvidia.com/en-us-grace-cpu/nvidia-grace-cpu-superchip>

Отечественный процессор «Эльбрус-8С»



Количество ядер	8
Кэш-память 2го уровня	8 * 512 КБ
Кэш-память 3го уровня	16 МБ
Рабочая частота	1.3 ГГц
Производительность	~250 ГФлопс
Тип контроллеров памяти	DDR3-1600
Кол-во контроллеров памяти	4
Поддержка многопроцессорных систем	До 4 процессоров
Каналы межпроцессорного обмена (пропускная способность)	3 (16 ГБ/с)
Технологический процесс	28 нм
Площадь кристалла	350 кв. мм
Рассеиваемая мощность на уровне	60 – 90 Вт

Отечественный процессор «Эльбрус-8С»



Отечественный процессор «Эльбрус-16С»



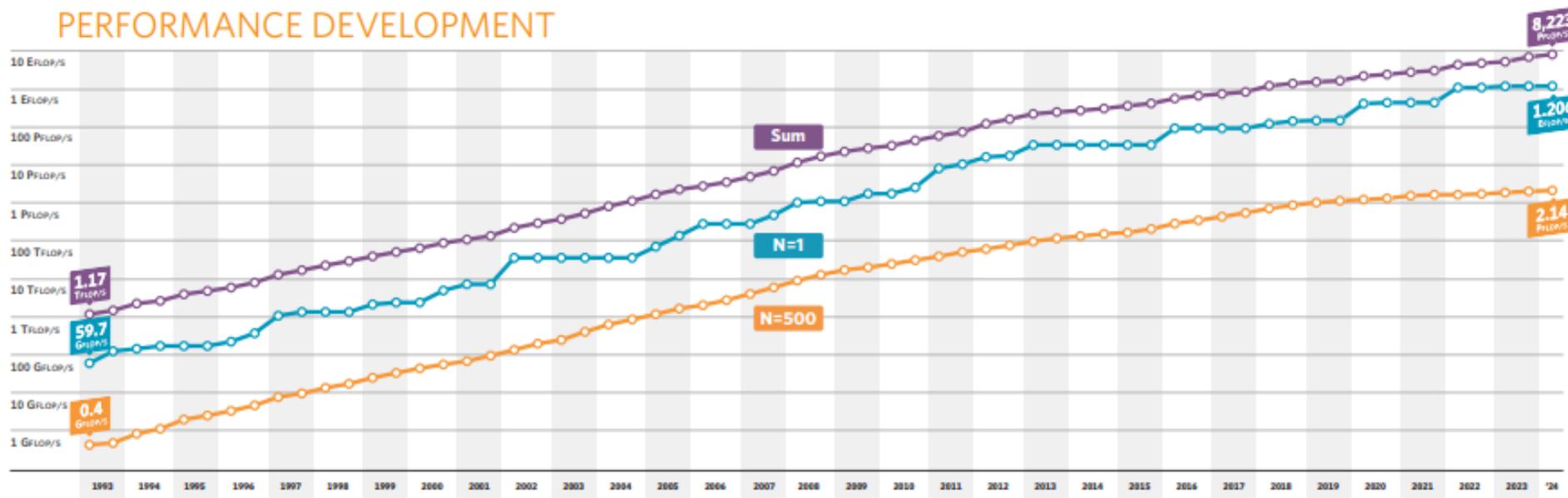
Количество ядер	16
Рабочая частота	2 ГГц
Производительность	~1500 Тфлопс одинарная точность ~750 Тфлопс двойная точность
Тип контроллеров памяти	DDR4-3200
Кол-во контроллеров памяти	8
Поддержка многопроцессорных систем	До 4 процессоров
Технологический процесс	16 нм
Число транзисторов	12 млрд.

Суперкомпьютерные системы (Top500)

MAY 2024

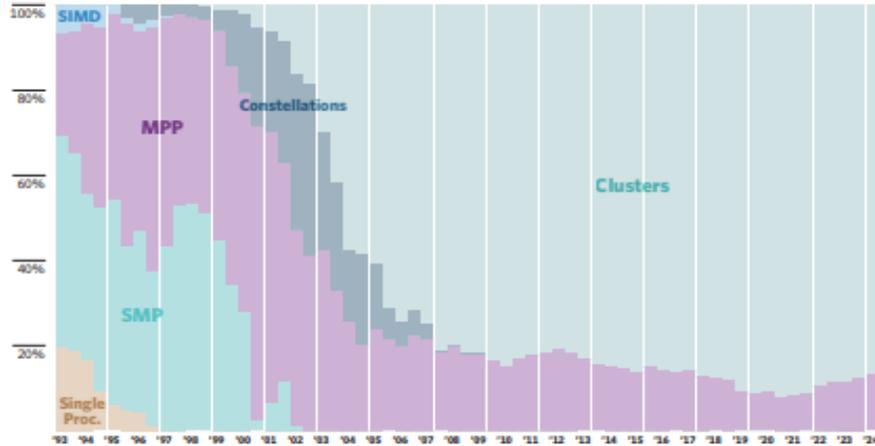
			SITE	COUNTRY	CORES	RMAX PFLOP/S	POWER MW
1	Frontier	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	DOE/SC/ORNL	USA	8,699,904	1,206.0	22.7
2	Aurora	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 (52C 2.4GHz), Intel Data Center GPU Max, Slingshot-11	DOE/SC/ANL	USA	9,264,128	1,012.0	38.7
3	Eagle	Microsoft NDv5, Xeon Platinum 8480C (48C 2GHz), NVIDIA H100, NVIDIA Infiniband NDR	Microsoft Azure	USA	1,123,200	561.2	
4	Fugaku	Fujitsu A64FX (48C, 2.2GHz), Tofu Interconnect D	RIKEN R-CCS	Japan	7,630,848	442.0	29.9
5	LUMI	HPE Cray EX235a, AMD Opt 3rd Gen EPYC (64C 2GHz), AMD Instinct MI250X, Slingshot-11	EuroHPC/CSC	Finland	2,220,288	379.7	6.01

PERFORMANCE DEVELOPMENT

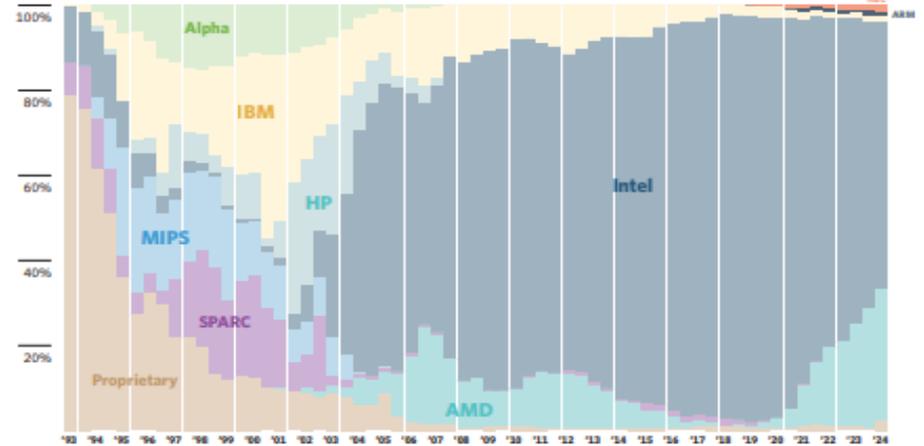


Суперкомпьютерные системы (Top500)

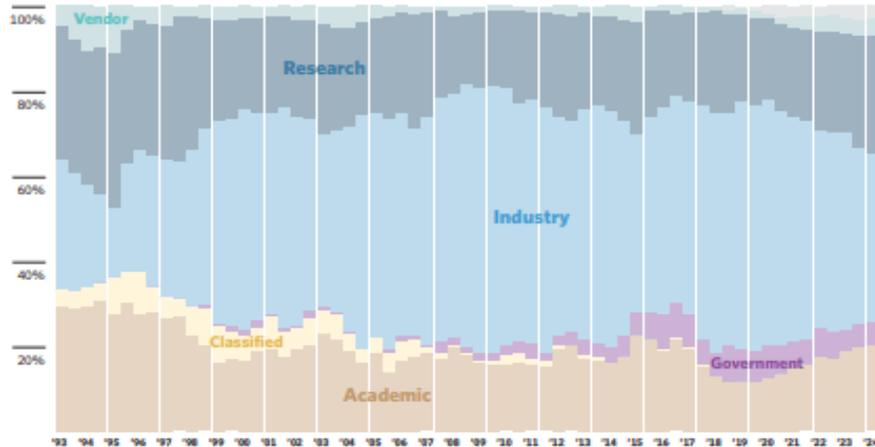
ARCHITECTURES



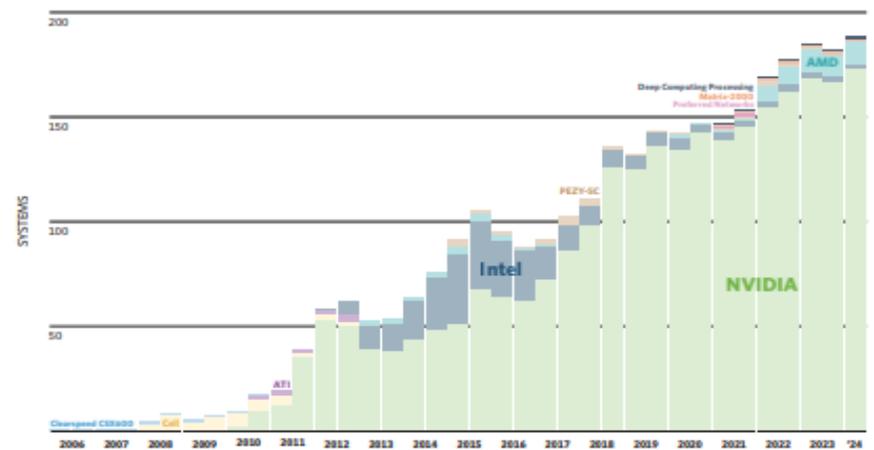
CHIP TECHNOLOGY



INSTALLATION TYPE



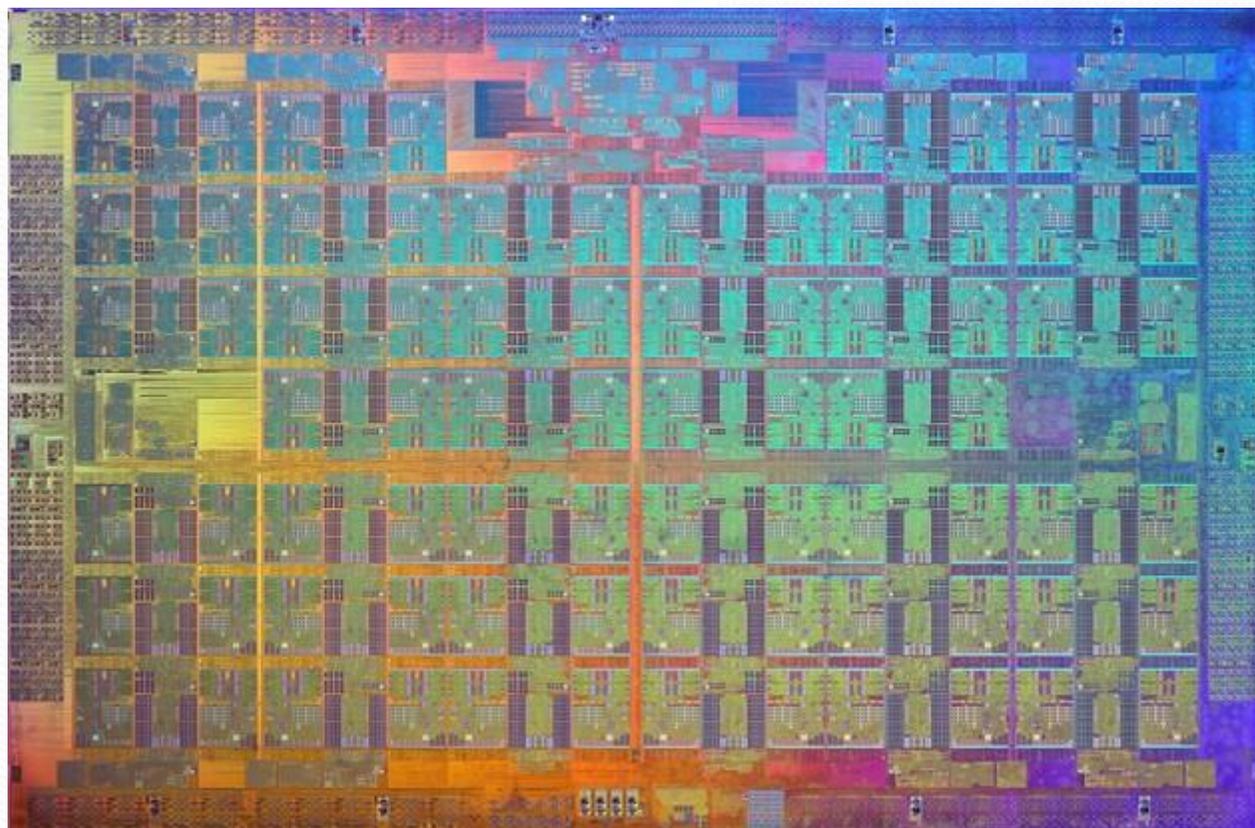
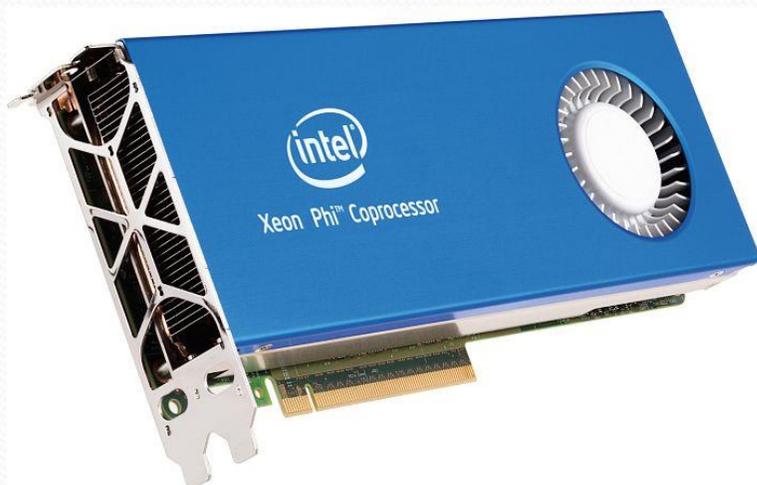
ACCELERATORS/CO-PROCESSORS



HPLINPACK

A Portable Implementation of the High Performance Linpack Benchmark for Distributed Memory Computers [FIND OUT MORE AT https://icl.utk.edu/hpl/](https://icl.utk.edu/hpl/)

Intel Xeon Phi Coprocessor / Processor



Pezy-SC Many Core Processor



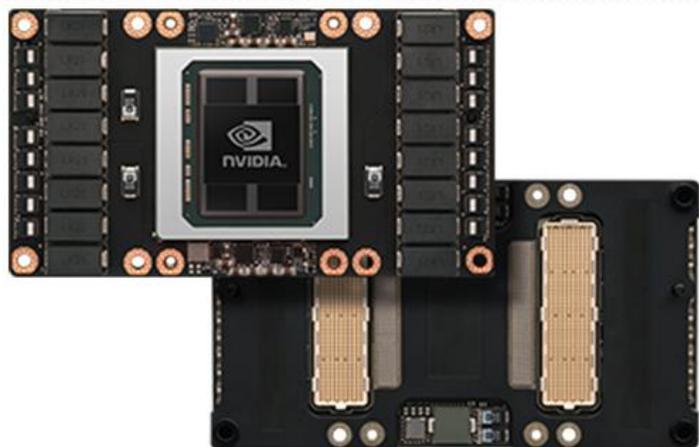
Logic Cores(PE)	1,024
Core Frequency	733MHz
Peak Performance	Floating Point Single 3.0TFlops / Double 1.5TFlops
Host Interface	PCI Express GEN3.0 x 8Lane x 4Port (x16 bifurcation available) JESD204B Protocol support
DRAM Interface	DDR4, DDR3 combo 64bit x 8Port Max B/W 1533.6GB/s +Ultra WIDE IO SDRAM (2,048bit) x 2Port Max B/W 102.4GB/s
Control CPU	ARM926 2core
Process Node	28nm
Package	FCBGA 47.5mm x 47.5mm, Ball Pitch 1mm, 2,112pin

Pezy-SC2 Many Core Processor

Logic Cores(PE)	2,048
Core Frequency	1,000MHz
Peak Performance	Half precision 16.2TFlops / Floating Point Single 8.2TFlops / Double 4.1TFlops
Host Interface	PCIe Gen3/4 x16 * 2CH (x8 * 4CH)
DRAM Interface	DDR4 64bit (ECC) * 4CH / 3,200Mbps BW=100GB/sec
CPU	MIPS64R6 (P6600) L1 I:64KB+D:64KB (each core) L2 2MB
Process Node	16 nm FinFET
Power	130 W



Графический ускоритель Nvidia P100



	Tesla P100 для PCIe серверов	Tesla P100 для серверов с NVLink
Производительность операций двойной точности с плавающей точкой	4,7 Терафлопс	5,3 Терафлопс
Производительность операций одинарной точности с плавающей точкой	9,3 Терафлопс	10,6 Терафлопс
Производительность операций половинной точности с плавающей точкой	18,7 Терафлопс	21,2 Терафлопс
Пропускная способность шины NVIDIA NVLink™	-	160 ГБ/с
Пропускная способность шины PCIe x16	32 ГБ/с	32 ГБ/с
Полоса пропускания стековой памяти CoWoS с HBM2	16 ГБ или 12 ГБ	16 ГБ
Полоса пропускания стековой памяти CoWoS с HBM2	732 ГБ/с или 549 ГБ/с	732 ГБ/с
Улучшенная программируемость с технологией Page Migration Engine	✓	✓
Защита ECC для повышенной надежности	✓	✓
Оптимизация под сервер для развертывания в дата-центре	✓	✓

Графический ускоритель Nvidia H200 Tensor Core GPU



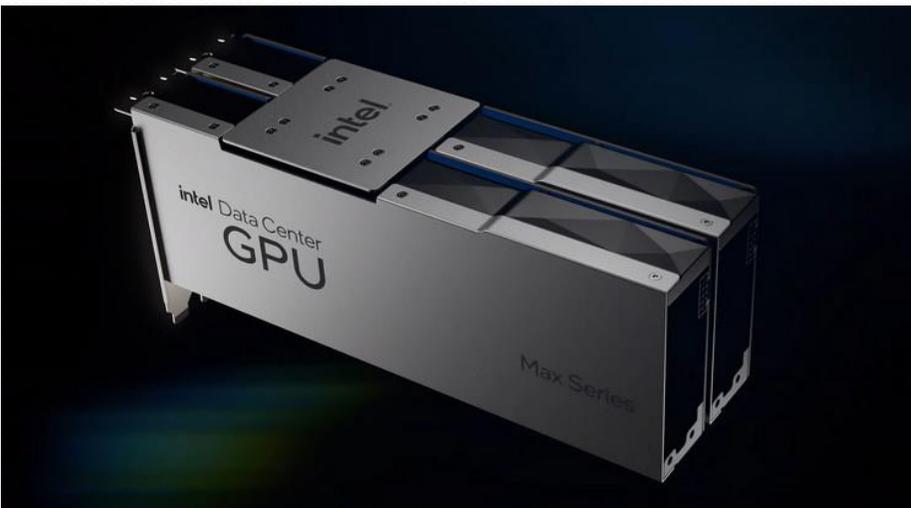
	H200 SXM	H200 NVL
FP64	34 TFLOPS	34 TFLOPS
FP64 Tensor Core	67 TFLOPS	67 TFLOPS
FP32	67 TFLOPS	67 TFLOPS
TF32 Tensor Core ²	989 TFLOPS	989 TFLOPS
BFLOAT16 Tensor Core ²	1,979 TFLOPS	1,979 TFLOPS
FP16 Tensor Core ²	1,979 TFLOPS	1,979 TFLOPS
FP8 Tensor Core ²	3,958 TFLOPS	3,958 TFLOPS
INT8 Tensor Core ²	3,958 TFLOPS	3,958 TFLOPS
GPU Memory	141GB	141GB
GPU Memory Bandwidth	4.8TB/s	4.8TB/s
Decoders	7 NVDEC	7 NVDEC
	7 JPEG	7 JPEG
Confidential Computing	Supported	Supported
Max Thermal Design Power (TDP)	Up to 700W (configurable)	Up to 600W (configurable)
Multi-Instance GPUs	Up to 7 MIGs @18GB each	Up to 7 MIGs @18GB each
Form Factor	SXM	PCIe
Interconnect	NVIDIA NVLink™: 900GB/s PCIe Gen5: 128GB/s	2- or 4-way NVIDIA NVLink bridge: 900GB/s PCIe Gen5: 128GB/s
Server Options	NVIDIA HGX™ H200 partner and NVIDIA-Certified Systems™ with 4 or 8 GPUs	NVIDIA MGX™ H200 NVL partner and NVIDIA-Certified Systems with up to 8 GPUs

Графический ускоритель AMD Instinct™ MI250X



Литография	TSMC 6nm FinFET
Кол-во потоковых процессоров	14,080
Вычислительные блоки	220
Peak Engine Clock	1700 MHz
Пиковая производительность в режиме с половинной точностью (FP16)	383 TFLOPs
Peak Single Precision Matrix (FP32) Performance	95.7 TFLOPs
Peak Double Precision Matrix (FP64) Performance	95.7 TFLOPs
Пиковая производительность в режиме с одинарной точностью (FP32)	47.9 TFLOPs
Пиковая производительность в режиме с двойной точностью (FP64)	47.9 TFLOPs
Peak INT4 Performance	383 TOPs
Peak INT8 Performance	383 TOPs
Total Board Power (TBP)	500Вт 560W Peak
Dedicated Memory Size	128 ГБ
Интерфейс памяти	8192-bit
Memory Clock	1.6 GHz
Пропускная способность памяти	До 3276.8 GB/s
Память с поддержкой ECC	Да (Full-Chip)

Ускоритель Intel Data Center GPU Max



Microarchitecture Xe-HPC **GPU Specifications**

Xe-cores 128

Ray Tracing Units 128

Intel® Xe Matrix Extensions (Intel® XMX)

Engines 1024

Xe Vector Engines 1024

Graphics Max Dynamic Clock 1600 MHz

Graphics Base Clock 900 MHz

Intel® Xe Link Maximum Frequency 53 Gbps

TDP 600 W

PCI Express Configurations Gen 5 x16

Memory Specifications

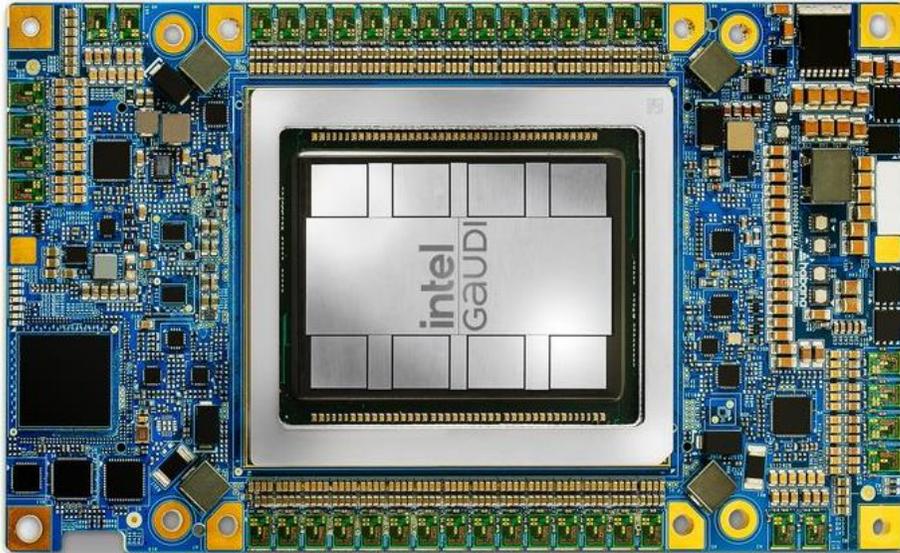
Memory Size 128 GB

Memory Type HBM2e

Graphics Memory Interface 1024 bit

Graphics Memory Bandwidth 3276.8 GB/s

Ускоритель Intel Gaudi 3



Gaudi 3 выполнен по 5-нм техпроцессу TSMC. Ускоритель состоит из двух кристаллов, на которые приходится 64 ядра Tensor Processing Cores (TPC) пятого поколения и восемь матричных математических движков (MME), а также 96 Мбайт памяти SRAM с пропускной способностью 12,8 Тбайт/с.

Установлено 128 Гбайт HBM2e с пропускной способностью 3,7 Тбайт/с.

Gaudi 3 укомплектован 24 контроллерами Ethernet RDMA с пропускной способностью по 200 Гбит/с, которые обеспечивают связь как между ускорителями в одном сервере, так и между разными серверами в одной системе.

intel GAUDI

Universal Baseboard
HLB-325

14.6 PFLOPS FP8	>1ТВ HBM2e	29.6ТВ/с HBM Bandwidth
64 Matrix Multiplication Engines	192 200GbE RDMA NICs	9.6ТВ/с bi-directional networking

intel.vision

Intel Confidential – Embargoed until 4/9/2024 at 8:35am PT

5 сентября
Москва, 2024

Суперкомпьютеры и параллельная обработка данных

81 из 316

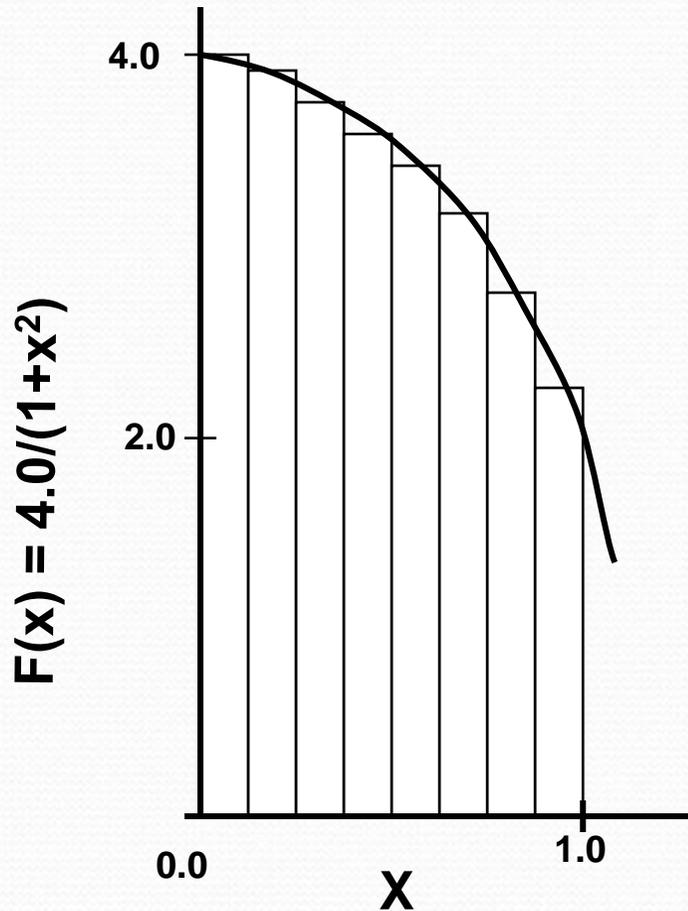
Тенденции развития современных вычислительных систем

- ❑ Темпы уменьшения латентности памяти гораздо ниже темпов ускорения процессоров + прогресс в технологии изготовления кристаллов => CMT (Chip MultiThreading)
- ❑ Пережающий рост потребления энергии при росте тактовой частоты + прогресс в технологии изготовления кристаллов => CMP (Chip MultiProcessing, многоядерность)
- ❑ И то и другое требует более глубокого распараллеливания для эффективного использования аппаратуры

Существующие подходы для создания параллельных программ для современных процессоров/систем

- ❑ Автоматическое / автоматизированное распараллеливание
- ❑ Библиотеки нитей
 - Win32 API
 - POSIX
- ❑ Библиотеки передачи сообщений
 - MPI
 - SHMEM
- ❑ OpenMP
- ❑ CUDA
- ❑ OpenACC
- ❑ DVMH

Вычисление числа π



$$\int_0^1 \frac{4.0}{(1+x^2)} dx = \pi$$

Мы можем аппроксимировать интеграл как сумму прямоугольников:

$$\sum_{i=0}^N F(x_i) \Delta x \approx \pi$$

Где каждый прямоугольник имеет ширину Δx и высоту $F(x_i)$ в середине интервала

Вычисление числа π . Последовательная программа

```
#include <stdio.h>
int main ()
{
    int n =100000, i;
    double pi, h, sum, x;
    h = 1.0 / (double) n;
    sum = 0.0;
    for (i = 1; i <= n; i ++)
    {
        x = h * ((double)i - 0.5);
        sum += (4.0 / (1.0 + x*x));
    }
    pi = h * sum;
    printf("pi is approximately %.16f", pi);
    return 0;
}
```

Автоматическое распараллеливание

Polaris, CAPO, WPP, SUIF, VAST/Parallel, OSCAR, Intel/OpenMP, ParaWise

```
icc -parallel pi.c
```

```
pi.c(8): (col. 5) remark: LOOP WAS AUTO-PARALLELIZED.
```

```
pi.c(8): (col. 5) remark: LOOP WAS VECTORIZED.
```

```
pi.c(8): (col. 5) remark: LOOP WAS VECTORIZED.
```

В общем случае, автоматическое распараллеливание затруднено:

- ❑ косвенная индексация ($A[B[i]]$);
- ❑ указатели (ассоциация по памяти);
- ❑ сложный межпроцедурный анализ.

Автоматизированное распараллеливание

Intel/GAP (Guided Auto-Parallel), CAPTools/ParaWise, BERT77, FORGE Magic/DM, ДВОР (Диалоговый Выскоуровневый Оптимизирующий Распараллеливатель), САПФОР (Система Автоматизации Параллелизации ФОРтран программ)

```
for (i=0; i<n; i++) {  
    if (A[i] > 0) {b=A[i]; A[i] = 1 / A[i]; }  
    if (A[i] > 1) {A[i] += b;}  
}
```

```
icc -guide -parallel test.cpp
```

Автоматизированное распараллеливание

test.cpp(49): remark #30521: (PAR) Loop at line 49 cannot be parallelized due to conditional assignment(s) into the following variable(s): b. This loop will be parallelized if the variable(s) become unconditionally initialized at the top of every iteration. [VERIFY] Make sure that the value(s) of the variable(s) read in any iteration of the loop must have been written earlier in the same iteration.

test.cpp(49): remark #30525: (PAR) If the trip count of the loop at line 49 is greater than 188, then use "#pragma loop count min(188)" to parallelize this loop. [VERIFY] Make sure that the loop has a minimum of 188 iterations.

```
#pragma loop count min (188)
for (i=0; i<n; i++) {
    b = A[i];
    if (A[i] > 0) {A[i] = 1 / A[i];}
    if (A[i] > 1) {A[i] += b;}
}
```

Вычисление числа π с использованием Win32 API

```
#include <stdio.h>
#include <windows.h>
#define NUM_THREADS 2
CRITICAL_SECTION hCriticalSection;
double pi = 0.0;
int n = 100000;
void main ()
{
    int i, threadArg[NUM_THREADS];
    DWORD threadID;
    HANDLE threadHandles[NUM_THREADS];
    for(i=0; i<NUM_THREADS; i++) threadArg[i] = i+1;
    InitializeCriticalSection(&hCriticalSection);
    for (i=0; i<NUM_THREADS; i++) threadHandles[i] =
        CreateThread(0,0,(LPTHREAD_START_ROUTINE) Pi,&threadArg[i], 0, &threadID);
    WaitForMultipleObjects(NUM_THREADS, threadHandles, TRUE,INFINITE);
    printf("pi is approximately %.16f", pi);
}
```

Вычисление числа π с использованием Win32 API

```
void Pi (void *arg)
{
    int i, start;
    double h, sum, x;
    h = 1.0 / (double) n;
    sum = 0.0;
    start = *(int *) arg;
    for (i=start; i<= n; i=i+NUM_THREADS)
    {
        x = h * ((double)i - 0.5);
        sum += (4.0 / (1.0 + x*x));
    }
    EnterCriticalSection(&hCriticalSection);
    pi += h * sum;
    LeaveCriticalSection(&hCriticalSection);
}
```

Взаимное исключение критических интервалов

При взаимодействии через общую память нити должны синхронизовать свое выполнение.

Thread0: $pi = pi + val$; && Thread1: $pi = pi + val$;

Время	Thread 0	Thread 1
1	LOAD R1,pi	
2	LOAD R2,val	
3	ADD R1,R2	LOAD R3,pi
4	STORE R1,pi	LOAD R4,val
5		ADD R3,R4
6		STORE R3,pi

Результат зависит от порядка выполнения команд. Требуется взаимное исключение критических интервалов.