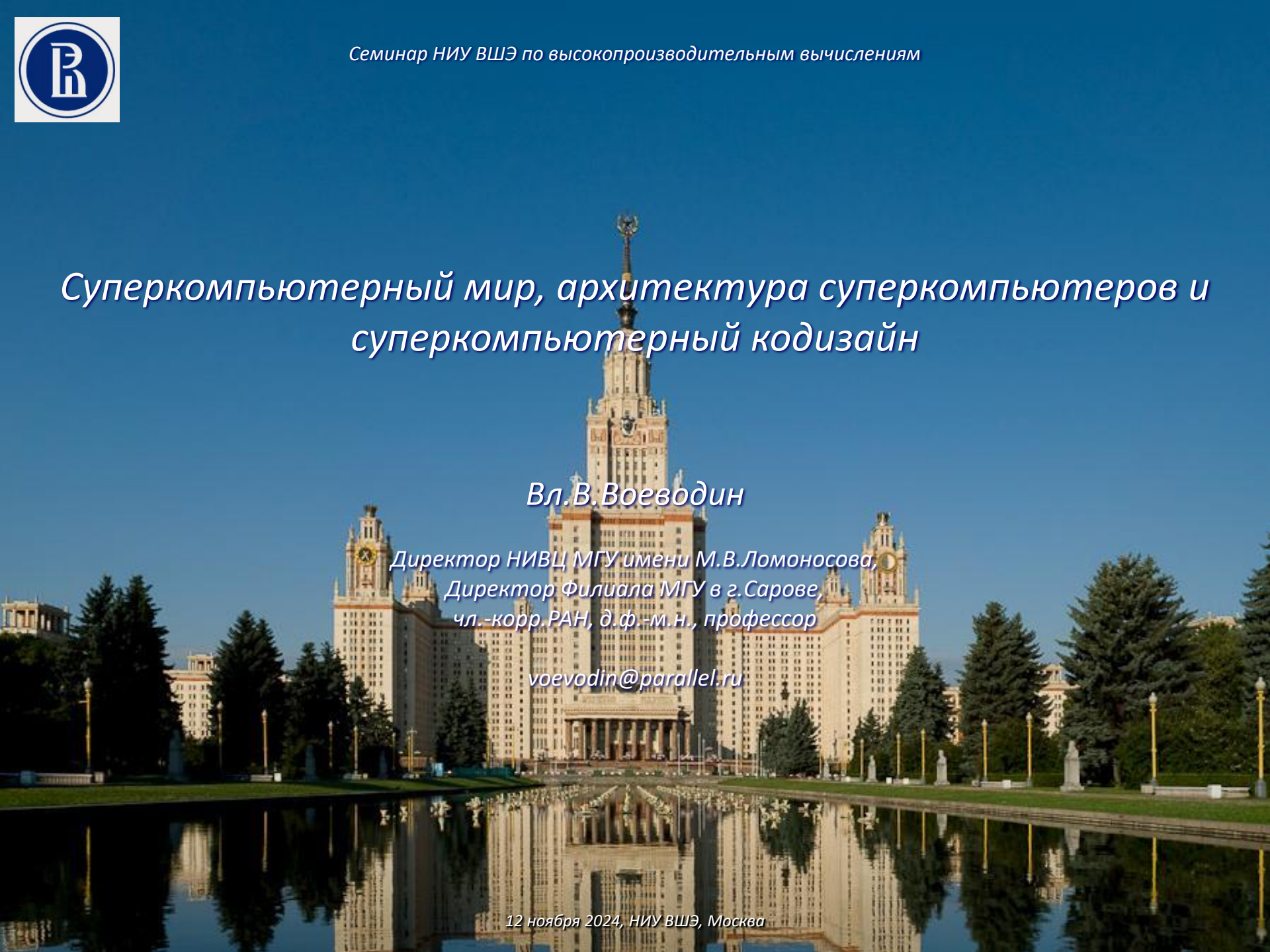


Суперкомпьютерный мир, архитектура суперкомпьютеров и суперкомпьютерный кодизайн

Вл.В.Воеводин

Директор НИВЦ МГУ имени М.В.Ломоносова,
Директор Филиала МГУ в г.Сарове,
чл.-корр.РАН, д.ф.-м.н., профессор

voevodin@parallel.ru

12 ноября 2024, НИУ ВШЭ, Москва

https://eurohpc-ju.europa.eu/index_en

An official website of the European Union How do you know?

EuroHPC Joint Undertaking

Search

The European High Performance Computing Joint Undertaking (EuroHPC JU)

Home | About | Supercomputers | Access to Our Supercomputers | Research & Innovation | News & Events | Media | Documents | Contact

Leading the Way in European Supercomputing

EuroHPC JU is a joint initiative between the EU, European countries and private partners to develop a World Class Supercomputing Ecosystem in Europe.




EuroHPC JU – программа развития суперкомпьютерной инфраструктуры в Европе (https://eurohpc-ju.europa.eu/)

Название компьютера	Страна	Поставщик	Rpeak (Pflops)	Rmax (Pflops)	Top500 (май-2024)	Место в Европе
Lumi	Finland	HPE	539,0	386,0	5	1
Leonardo	Italy	Atos	315,0	249,0	7	2
MareNostrum 5	Spain	Bull/Lenovo	295,0	215,0	8	3
Meluxina	Luxembourg	Atos	18,3	12,8	89	22
Karolina	Chechia	HPE	12,9	9,6	135	37
Discoverer	Bulgaria	Atos	5,9	4,5	188	59
Vega	Slovenia	Atos	5,4	3,8	226	68
Deucalion	Portugal	Fujitsu/Atos	5,0	3,9	219	67

EuroHPC
Joint Undertaking

EuroHPC JU – программа развития суперкомпьютерной инфраструктуры в Европе
(https://eurohpc-ju.europa.eu/)

Название компьютера	Страна	Поставщик	Rpeak (Pflops)	Rmax (Pflops)	Top500 (май-2024)	Место в Европе
Lumi	Finland	HPE	539,0	386,0	5	1
Leonardo	Italy	Atos	315,0	249,0	7	2
MareNostrum 5	Spain	Bull/Lenovo	295,0	215,0	8	3
Meluxina	Luxembourg	Atos	18,3	12,8	89	22
Karolina	Chechia	HPE	12,9	9,6	135	37
Discoverer	Bulgaria	Atos	5,9	4,5	188	59
Vega	Slovenia	Atos	5,4	3,8	226	68
Deucalion	Portugal	Fujitsu/Atos	5,0	3,9	219	67
Jupiter	Germany	ParTec-Eviden		1000+		

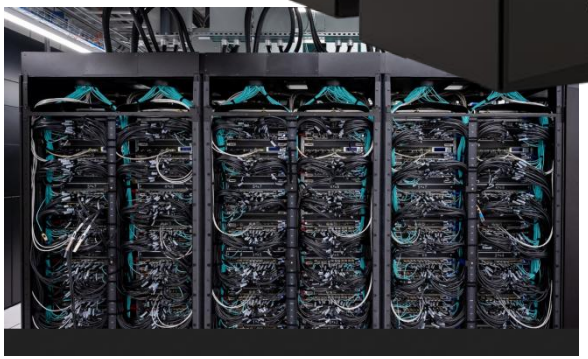
Начало поставки – 2024 год

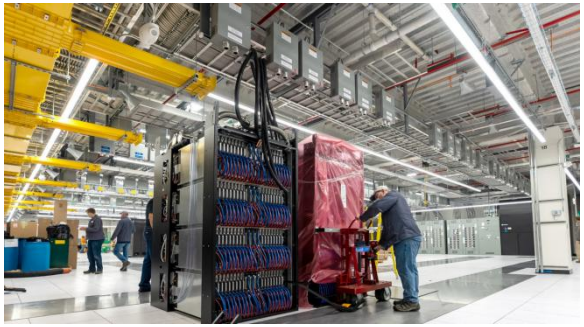
EuroHPC Joint Undertaking

Ключевые особенности архитектуры для обеспечения высокой производительности современных суперкомпьютеров

- Высокая степень параллелизма на всех уровнях архитектуры суперкомпьютеров (параллелизм, конвейерность).
- Обеспечение эффективной работы с иерархией памяти.

Суперкомпьютер Frontier, США
(#1 Top500 в 2022-2024 г.)





9 408 вычислительных узлов,
в каждом узле:
1 x CPUs (AMD “Trento”, 64 ядра, 2GHz)
4 x GPU (AMD Radeon MI250X)
8 699 904 ядра

Производительность:
Пик (теория): 1.71 Eflop/s
Тест Linpack: 1.2 Eflop/s (70%)

Оперативная память = 9.2 PBytes
HDD = 716 PB (+37 PB на узлах)

22.7 MW (всего 29 MW) – 52.2 Gflops/Watt
(1 стойка – 62.68 Gflops/Watt)

Так ли важно, кому доступен этот “Экзафлопс”?
(Экзафлопс, экономика, безопасность…)

HPCwire

Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them

- Home
- Topics
- Sectors
- Exascale
- Specials
- Resource Library
- Podcast
- Events
- Solution Channels
- Job Bank
- About
- Subscribe

US Bars Nvidia and AMD from Selling Top GPUs into China

By Tiffany Trader

September 1, 2022

New trade restrictions levied by the United States against China limit the sale of cutting-edge HPC and AI technologies from Nvidia and AMD to the world’s second-largest economy.

In an SEC filing, Nvidia revealed it was prohibited from exporting its A100 and forthcoming H100 GPUs to China and Russia, effective immediately. The stated purpose of the licensing requirements is to prevent “military end use” by these nations.

AMD reported it had also received instructions from U.S. authorities to halt sales of its top GPU chip, the Instinct MI250, to China and Russia. A variant of that chip, the MI250X, powers the U.S. Department of Energy’s Frontier supercomputer, which became the first officially ranked exascale supercomputer earlier this year.

HPCwire

Search… Go

Since 1987 - Covering the Fastest Computers in the World and the People Who Run Them

- Home
- Topics
- Sectors
- Exascale
- Specials
- Resource Library
- Podcast
- Events

Fearing China, US Blocks the Sale of Nvidia GPUs to the Middle East

By Agam Shah

September 1, 2023

The U.S. has put more curbs that block the sale of certain Nvidia GPUs to the Middle East, largely under fears that the technology will be accessible to China.

The setback for Nvidia was revealed in the company’s 10K filing with the U.S. Securities and Exchange Commission.

In the filing, Nvidia stated that the U.S. government, in the most recent quarter, “informed us of an additional licensing requirement for a subset of A100 and H100 products destined to certain customers and other regions, including some countries in the Middle East.”

Nvidia is already banned from selling the A100 and H100 GPUs to China because of export controls restricting U.S. companies from exporting advanced A.I. and chips. Nvidia devised the

HPCwire

Search… Go

White House Mulls Expanding AI Chip Export Bans Beyond China

By Ali Azhar

October 31, 2024

The Biden administration is reportedly considering capping sales of advanced artificial intelligence (AI) chips from US-based manufacturers like AMD and Nvidia to certain countries, including those in the Middle East. The move represents concerns over the use of advanced AI chips for surveillance and military applications. AI technologies can be used to enhance the capabilities of autonomous drones, cyber warfare, sophisticated mass monitoring systems, and several other applications.

AMD and Nvidia, the two leading AI chip manufacturers, could face restrictions on their export licenses for advanced AI chips.

Reuters

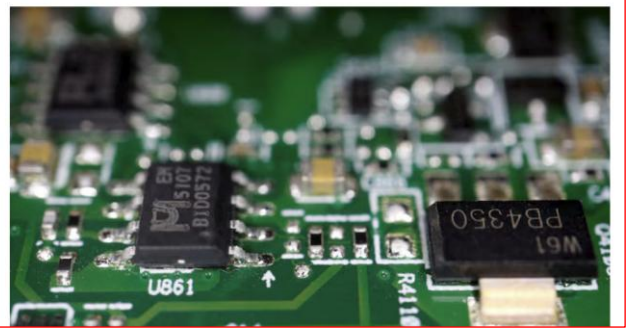
World ⌄ US Election Business ⌄ Markets ⌄ More ⌄

US sets new rule that could spur AI chip shipments to the Middle East

By Karen Freifeld

September 30, 2024 9:56 PM GMT+3 · Updated a month ago

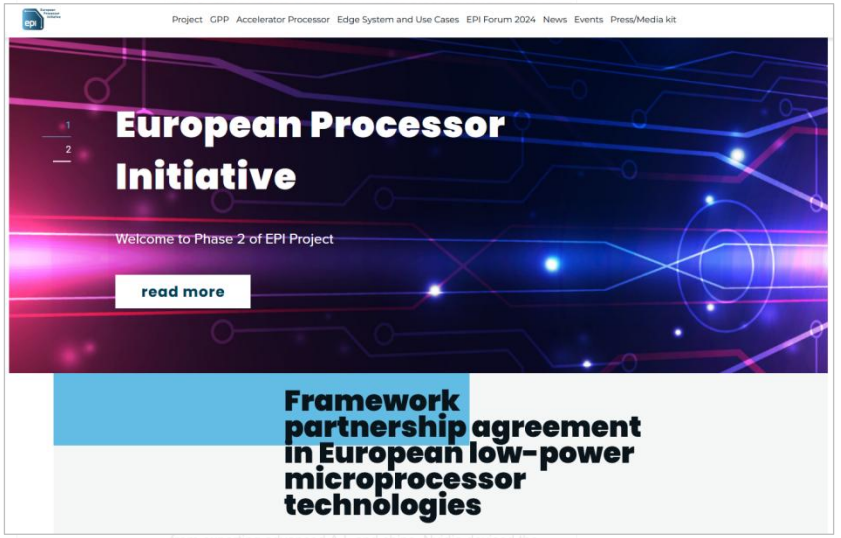
Aa



Так ли важно, кому доступен этот “Экзафлопс”?
(Экзафлопс, экономика, безопасность…)

HALO
HPC-AI LEADERSHIP ORGANIZATION

Similarly, countries are beginning to invest in the idea of “HPC nationalism.” This term refers to the tendency of countries to develop their own HPC technologies, infrastructure, and expertise to achieve technological independence or superiority. Again, international cooperation is impacted when countries have vested interests in developing their own processor chips, systems, and software stacks.



Индия представила 96-ядерный Arm-процессор собственной разработки — это первый местный HPC-чип

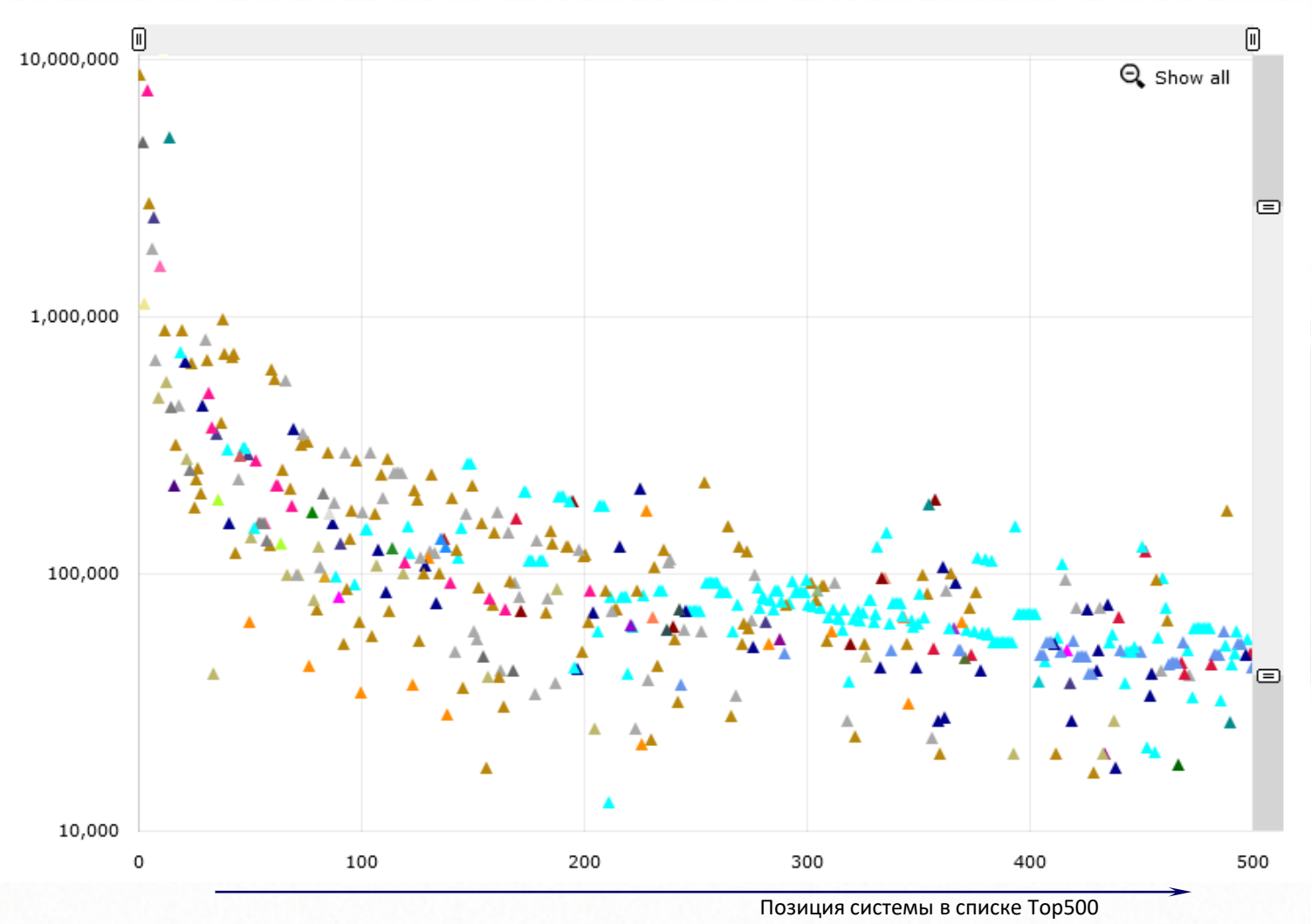
19.05.2023 [19:08], Матвей Филькин

На этой неделе индийский Центр развития передовых вычислений (C-DAC) анонсировал первый в стране разработанный самостоятельно процессор для высокопроизводительных вычислений (HPC). Первый индийский чип, названный Aum, основан на архитектуре Neoverse V1 Zeus (Arm v8.4) и может масштабироваться до 96 ядер. Ожидается, что он появится на рынке уже в 2024 году и будет выпускаться по техпроцессу 5 нм на мощностях TSMC.

	Fujitsu A64FX	C-DAC AUM HPC Processor
Fabrication Technology	7nm FF TSMC	5nm FF
Core Configuration	(48+4)-Cores, 2.2 GHz (typical)	96-Cores, 3.0 GHz (typical) 3.5+ GHz (turbo)
DDR Configuration	No DDR	16-Channels (32 bit) DDR5-5200 BW = 332.8 GB/s
HBM	32-GB HBM2 (4-Controllers) BW = 1 TB/s	64-GB HBM3 (4-Controllers) BW = 2.87 TB/s
PCIe	16 PCIe Gen3 Lanes	64 PCIe Gen5 Lanes
Power	Not Known	300 W (TDP)
Performance (DP)	2.7 TFLOPS per socket	4.6+ TFLOPS per socket
Bytes/FLOPS	0.38	0.7

Число ядер в системах списка Top500

(http://top500.org, ноябрь, 2023 г.)



Средняя степень параллельности в системах списка Top500 – 212 627 вычислительных ядер.

Ключевые особенности архитектуры для обеспечения высокой производительности современных суперкомпьютеров

- Высокая степень параллелизма на всех уровнях архитектуры суперкомпьютеров (параллелизм, конвейерность).
- Обеспечение эффективной работы с иерархической памятью.
- Цель создания суперкомпьютерных систем – быстрое решение вычислительно сложных задач и обработка высокой вычислительной нагрузки.
- Без учета особенностей архитектуры вычислительных систем высокой производительности не достичь.

Ключевое понятие – **суперкомпьютерный кодизайн**, предполагающий обеспечение соответствия структуры и особенностей всех этапов решения задач с использованием суперкомпьютерных систем.

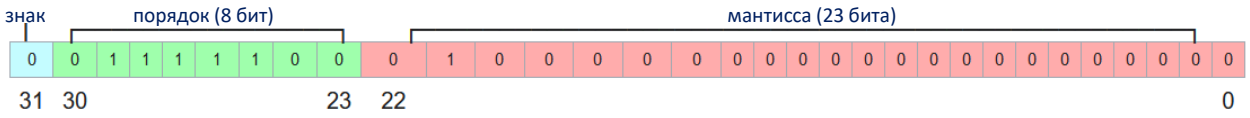
Эффективность вычислительных систем
(рейтинги Top500, HPCG500, Green500, Graph500)

Место в Top500	Система	Число ядер	Rmax (Pflops)	Rpeak (Pflops)	Power (kW)	Эффект-ть Top500 (%)	Место в HPCG500	Эффект-ть HPCG (%)	Место в Green500	Graph500 BFS	Graph500 SSSP
1	**Frontier** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC	8 699 904,00	1 206,00	1 714,81	22 786,00	70,33	2	0,80	13	3	-
2	**Aurora** - HPE Cray EX - Intel Exascale Compute Blade, Xeon	9 264 128,00	1 012,00	1 980,01	38 698,00	51,11	3	0,28	42	5	-
3	**Eagle** - Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA	2 073 600,00	561,20	846,84		66,27	-	-	294	-	-
4	**Supercomputer Fugaku** - Supercomputer Fugaku, A64FX	7 630 848,00	442,01	537,21	29 899,00	82,28	1	2,91	68	1	3
5	**LUMI** - HPE Cray EX235a, AMD Optimized 3rd Generation EPYC	2 752 704,00	379,70	531,51	7 107,00	71,44	4	0,84	12	-	-
6	**Alps** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200	1 305 600,00	270,00	353,75	5 194,00	76,33	5	1,01	14	-	-
7	**Leonardo** - BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz,	1 824 768,00	241,20	306,31	7 494,00	78,74	6	0,99	28	-	-
8	**MareNostrum 5 ACC** - BullSequana XH3000, Xeon	663 040,00	175,30	249,44	4 159,00	70,28	12	0,45	15	8	-
9	**Summit** - IBM Power System AC922, IBM POWER9 22C	2 414 592,00	148,60	200,79	10 096,00	74,01	7	1,42	72	-	-
10	**Eos NVIDIA DGX SuperPOD** - NVIDIA DGX H100, Xeon Platinum	485 888,00	121,40	188,65	-	64,35	-	-	311	-	-
189	**JEDI** - BullSequana XH3000, Grace Hopper Superchip 72C 3GHz,	19 584,00	4,50	5,13	67,00	87,72	94	1,15	1	-	-
128	**Isambard-AI** phase 1 - HPE Cray EX254n, NVIDIA Grace 72C	34 272,00	7,42	9,29	117,00	79,87	-	-	2	-	-
95	**Earth Simulator** -SX-Aurora TSUBASA - SX-Aurora TSUBASA	43 776,00	9,99	13,45	1 391,00	74,28	18	5,56	91	-	-
61	**AOBA-S** - SX-Aurora TSUBASA C401-8, Vector Engine Type 30A	64 512,00	17,22	19,82	1 389,00	86,88	13	5,49	193	-	-

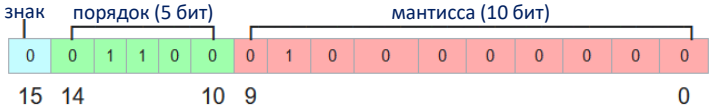
Рейтинг MLPerf: расширение в 2024 году набора ML и AI тестов данного рейтинга моделями Llama 2 (70 млрд. параметров) и Stable Diffusion XL (2.6 млрд.параметров, генерация изображений).

Форматы представления вещественных чисел
(кодизайн: производительность, компактность)

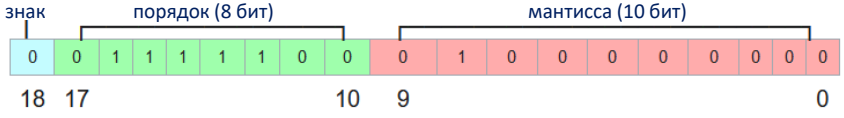
IEEE 754 single-precision 32-bit float



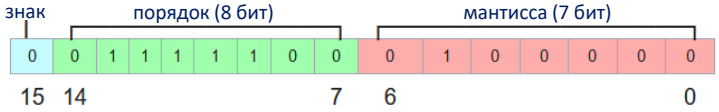
IEEE half-precision 16-bit float



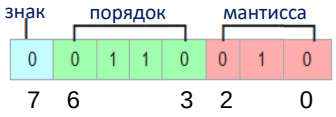
NVIDIA's Tensor float (19 бит)



bfloat16



8-bit float (1.4.3 minifloat)



Форматы и разрядность современных компьютеров:
- вещественные числа,
- целые числа,
- 128/64/32/24/19/16/8/4 бит

Суперкомпьютер Fugaku:
Int: 8,16,32,64 бит;
Float: 16,32,64 бит

	Hopper	Ampere	Turing	Volta
Операции, поддерживаемые тензорными ядрами	FP64, TF32, bfloat16, FP16, FP8, INT8	FP64, TF32, bfloat16, FP16, INT8, INT4, INT1	FP16, INT8, INT4, INT1	FP16
Операции, поддерживаемые ядрами CUDA*	FP64, FP32, FP16, bfloat16, INT8	FP64, FP32, FP16, bfloat16, INT8	FP64, FP32, FP16, INT8	FP64, FP32, FP16, INT8

Поколения архитектур и парадигмы программирования
(кодизайн: метод – программирование – архитектура)

Суперкомпьютер Cray-1

Векторно-конвейерные компьютеры

Середина 70-х годов.

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

Программирование: векторизация самых внутренних циклов.

Поколения архитектур и парадигмы программирования
(кодизайн: метод – программирование – архитектура)

Суперкомпьютер Cray X-MP

Суперкомпьютер Cray Y-MP

Векторно-параллельные компьютеры

Начало 80-х годов.

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.

Поколения архитектур и парадигмы программирования
(кодизайн: метод – программирование – архитектура)

Суперкомпьютер Cray T3D

Суперкомпьютер Intel Paragon XPS140

Массивно-параллельные компьютеры

Начало 90-х годов.

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.

Поколения архитектур и парадигмы программирования
(кодизайн: метод – программирование – архитектура)

DEC AlphaServer

Суперкомпьютер Sun StarFire

Параллельные компьютеры с общей памятью

Середина 90-х годов.

Особенности архитектуры: сотни процессоров объединяются над общей памятью.

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

Поколения архитектур и парадигмы программирования
(кодизайн: метод – программирование – архитектура)

Суперкомпьютер МГУ “Чебышев”

“К” суперкомпьютер

Кластеры из узлов с общей памятью

Начало 2000-х.

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.

Поколения архитектур и парадигмы программирования
(кодизайн: метод – программирование – архитектура)

Суперкомпьютер МГУ “Ломоносов”

Суперкомпьютер Tianhe-2

Кластеры из узлов с общей памятью с ускорителями

С конца 2000-х до настоящего времени.

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+OpenCL/CUDA; AMD ROCm/HIP

EuroHPC JU – программа развития суперкомпьютерной инфраструктуры в Европе (https://eurohpc-ju.europa.eu/)

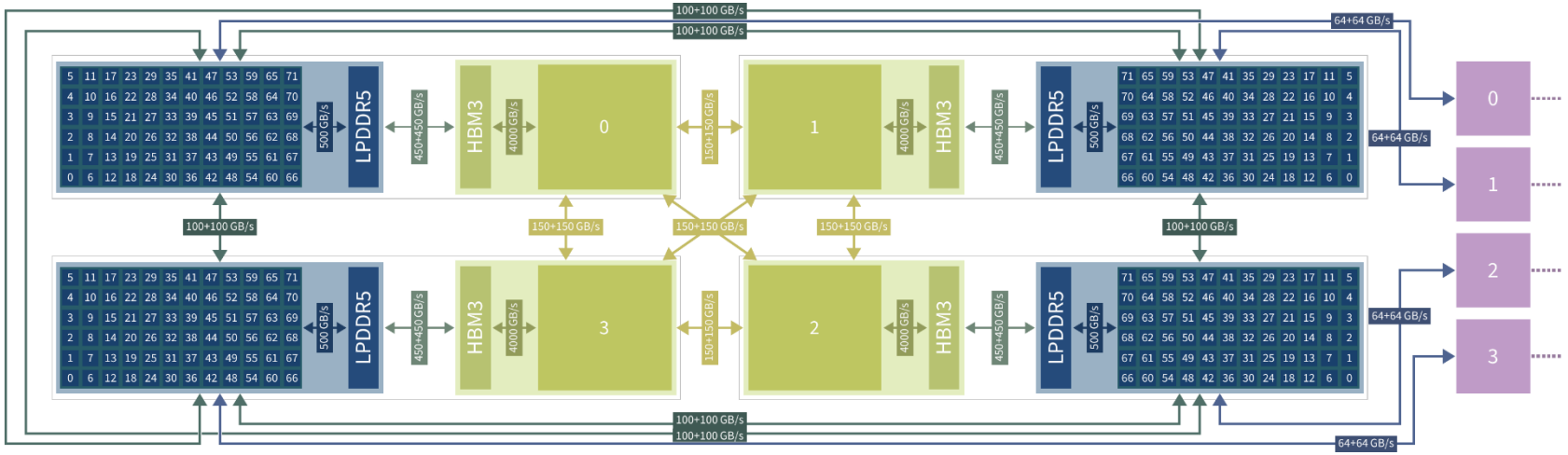
Название компьютера	Страна	Поставщик	Rpeak (Pflops)	Rmax (Pflops)	Top500 (май-2024)	Место в Европе	Вычислительные разделы	Разделы СХД
Lumi	Finland	HPE	539,0	386,0	5	1	GPU-pertition (AMD Instinct GPU) CPU-partition (AMD EPYC CPU) Data analytics partition Container cloud partition	7PB ultra-fast flash storage 80PB traditional storage
Leonardo	Italy	Atos	315,0	249,0	7	2	GPU-pertition (NVIDIA Ampere) CPU-partition (Intel Sapphire Rapids)	5PB flash/NVMe 100PB traditional storage
MareNostrum 5	Spain	Bull/Lenovo	295,0	215,0	8	3	General Purpose (Intel Sapphire Rapids Accelerated partition (NVIDIA Hopper)	248PB SSD/flash/HD 402PB tape storage
Meluxina	Luxembourg	Atos	18,3	12,8	89	22	GPU-pertition (NVIDIA Ampere A100) CPU-partition (AMD EPYC) Accelerator - FPGA	20PB flash storage Tape storage
Karolina	Chechia	HPE	12,9	9,6	135	37		
Discoverer	Bulgaria	Atos	5,9	4,5	188	59		
Vega	Slovenia	Atos	5,4	3,8	226	68		
Deucalion	Portugal	Fujitsu/Atos	5,0	3,9	219	67		
Jupiter	Germany	ParTec-Eviden		1000+			Booster (NVIDIA Grace-Hopper) Cluster (Rhea - EPI project, ARM-based)	21PB low-latency flash storage 29PB NVMe storage 300PB traditional storage 700PB tape storage

EuroHPC Joint Undertaking

Суперкомпьютер Jupiter, Germany, Julich

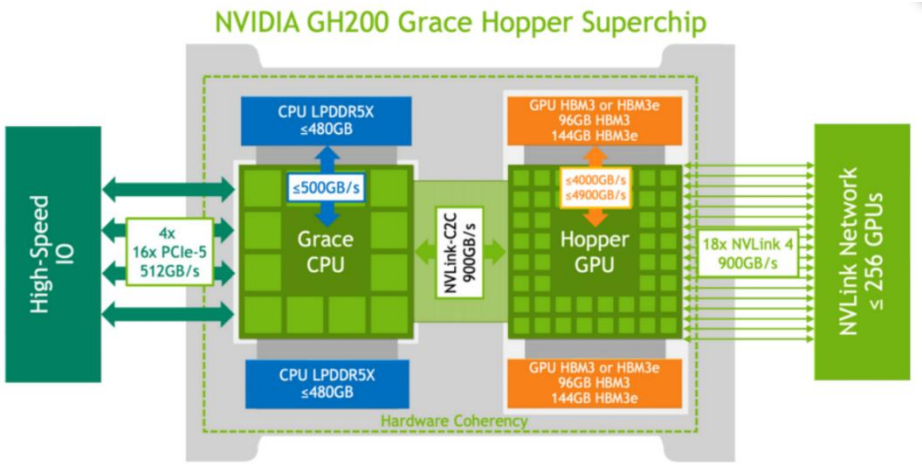
Jupiter – первый суперкомпьютер класса Exascale в Европе. Программа EuroHPC JU.
Стоимость – 500 млн.евро, включая расходы на приобретение и эксплуатацию.

Jupiter, две основные части: **Booster** и **Cluster**.
Booster: 6000 выч.узлов, 1 Eflops (fp64, HPL), 70+ Eflops (8-bits),
Выч.узел – 4 * NVIDIA Grace-Hopper; 4 * InfiniBand NDR (200 Gbit/s),
(4 * CPU Grace (ARM, 72 cores, LPDDR5X, 500GB/s)) * CPU NVLink (100GB/s bi-directional bw)
(4 * GPU Hopper H100, HBM3, 4TB/s) * NVLink 4 (150GB/s per direction)
Cluster: 1300 выч.узлов, 5 Pflops (fp64, HPL),
Выч.узел – 2 * (CPU Rheal, 80 ARM cores), 2*64GB HBM+512GB DDR5; InfiniBand NDR (200 Gbit/s)
Interconnect: InfiniBand NDR (200 Gbit/s), DragonFly+ groups (FatTree inside each group) for Booster, Cluster, storage, administrative infrastructure.



Booster: структура вычислительного узла

NVIDIA Grace Hopper Superchip
(co-design is not simple)



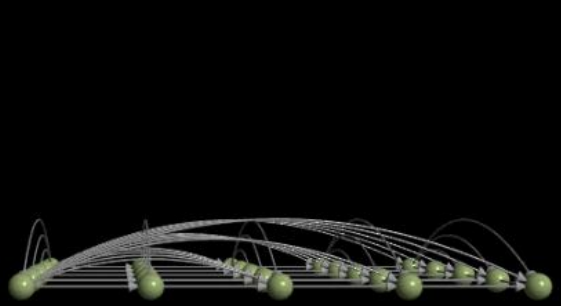
FP64	34 teraFLOPS
FP64 Tensor Core	67 teraFLOPS
FP32	67 teraFLOPS
TF32 Tensor Core	989 teraFLOPS* | 494 teraFLOPS
BFLOAT16 Tensor Core	1,979 teraFLOPS* | 990 teraFLOPS
FP16 Tensor Core	1,979 teraFLOPS* | 990 teraFLOPS
FP8 Tensor Core	3,958 teraFLOPS* | 1,979 teraFLOPS

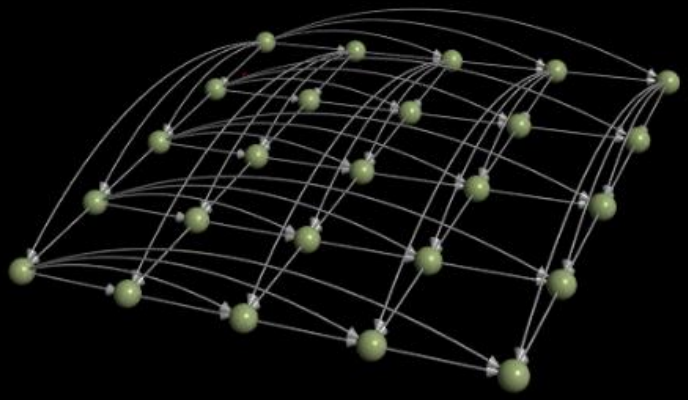
- Using the STREAM benchmark, the researchers achieved 3.4 TB/s H100 of HBM3 memory bandwidth, short of Nvidia's theoretical claim of 4 TB/s.
- Nvidia claims LPDDR5X bandwidth speeds of 500 GB/s, and the researchers measured real-world performance of 486 GB/s.
- The NVLink-C2C interconnect bandwidth was 375 GB/s for transfers from host-to-device and 297 GB/s for device-to-host. That adds up to a total bandwidth of 672 GB/s, well short of the 900 GB/s two-way transfers claimed by Nvidia.

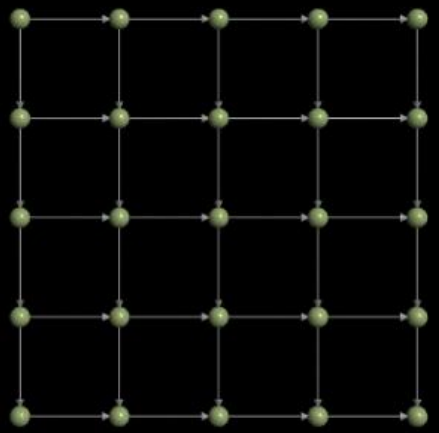
G. Schieffer, J. Wahlgren, J. Ren, J. Faj, I. Peng, "Harnessing Integrated CPU-GPU System Memory for HPC: a first look into Grace Hopper", arXiv:2407.07850.

Свойства метода и алгоритма – элемент кодизайна

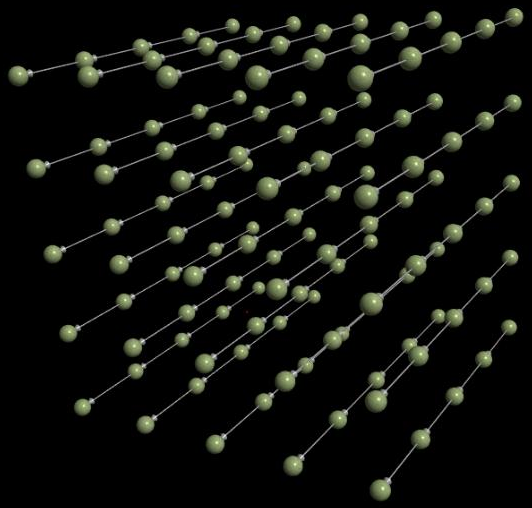
(Информационная структура – элемент описания алгоритма)

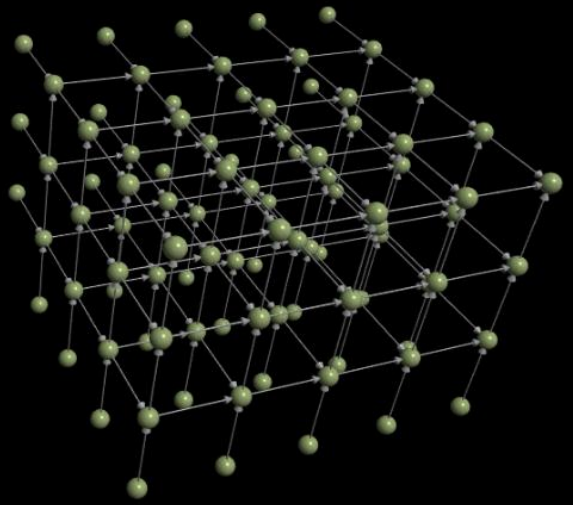






Типовые алгоритмические структуры





Огромный ресурс параллелизма

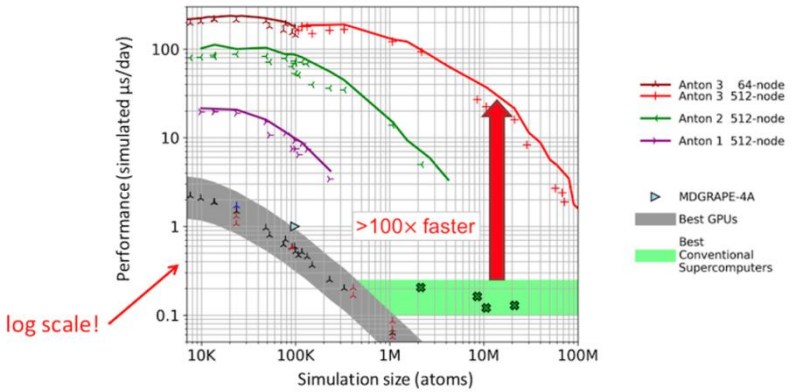
От свойств алгоритмов к архитектуре компьютеров
(Суперкомпьютерный кодизайн: проектирование вычислительных систем)

1. Общее описание алгоритма
2. Математическое описание алгоритма
3. Вычислительное ядро алгоритма
   (операция, форматы данных, структуры данных)
4. Макроструктура алгоритма
   (макрооперации, типовые вычислительные структуры)
5. Схема реализации последовательного алгоритма
6. Последовательная сложность алгоритма
   (арифметические операции, операции чтения/записи)
7. Информационный граф алгоритма
8. Ресурс параллелизма алгоритма
   (параллельная сложность алгоритма, имеющийся параллелизм)
9. Входные и выходные данные алгоритма
10. Свойства алгоритма
    (вычислительная мощность, устойчивость, детерминированность, с
11. Локальность данных, локальность вычислений
12. Возможные способы и особенности параллельной р
    (features, properties, programming technologies…)
13. Возможные препятствия для масштабируемости алгор
    (ограниченный ресурс параллелизма, дисбаланс вычислений, синхронизация…)
14. Возможные особенности динамических характеристик в реализации а
15. Существующие реализации алгоритма

Ландшафт доступных компьютерных архитектур
(Насколько разнообразен компьютерный мир сегодня?)

И мы, безусловно, должны думать о правильном подборе архитектуры…
А меняются ли алгоритмы при изменении архитектуры?…

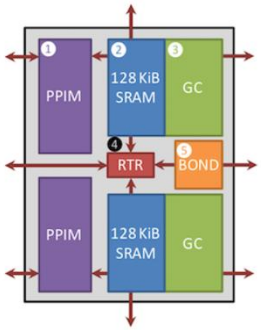
Суперкомпьютер Anton – апофеоз кодизайна для моделирования молекулярной динамики
(D.E. Shaw Research company)



Молекулярная динамика на разных платформах.

	ANTON	ANTON 2	ANTON 3
Tape-out	2007	2012	2020
CPU cores	8+4+1	66	528*
PPIMs	32	76	528*
Flex SRAM	0.125 MiB	4 MiB	66 MiB*
Atoms / node	460	8,000	110,000*
Clock frequency	0.485/0.970 GHz	1.65 GHz	2.8+ GHz
Channel bandwidth	0.607 Tbps	2.7 Tbps	5.6+ Tbps
Process node	90 nm	40 nm	7 nm
Transistors	0.2 G	2.0 G	31.8 G
Die size	299 mm²	410 mm²	451 mm²
Power	30 W	190 W	360 W

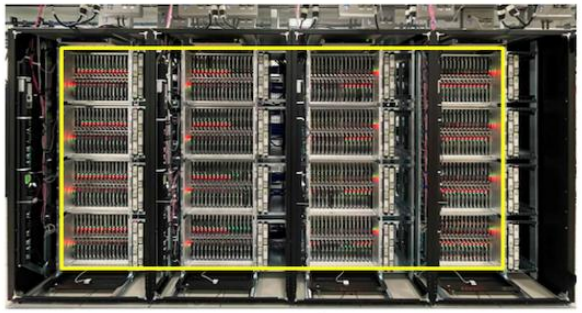
Развитие процессоров семейства Anton.



- Evolutionary changes
  - ① Support additional functional forms
  - ② Increase memory capacity
  - ③ Tune instruction set for MD application
  - ③ Increase code density
- **Revolutionary changes**
  - ④ Co-locate compute resources
  - ⑤ Specialize bonded force computation
  - ① Double effective density of pairwise interaction calculation
  - ②④ Implement fine-grained synchronization within memory and network

8×8 nodes

2×64 nodes



512 nodes

D E Shaw Research

Суперкомпьютер Anton – апофеоз кодизайна для моделирования молекулярной динамики
(D.E. Shaw Research company)

Hot Chips Conference 2021…
Two interesting questions came from Jeffrey Vetter at Oak Ridge National Labs, who asked if the system could scale beyond 512 nodes for larger MD simulations, and whether other types of molecular dynamics applications were being considered, such as those focused on material science.

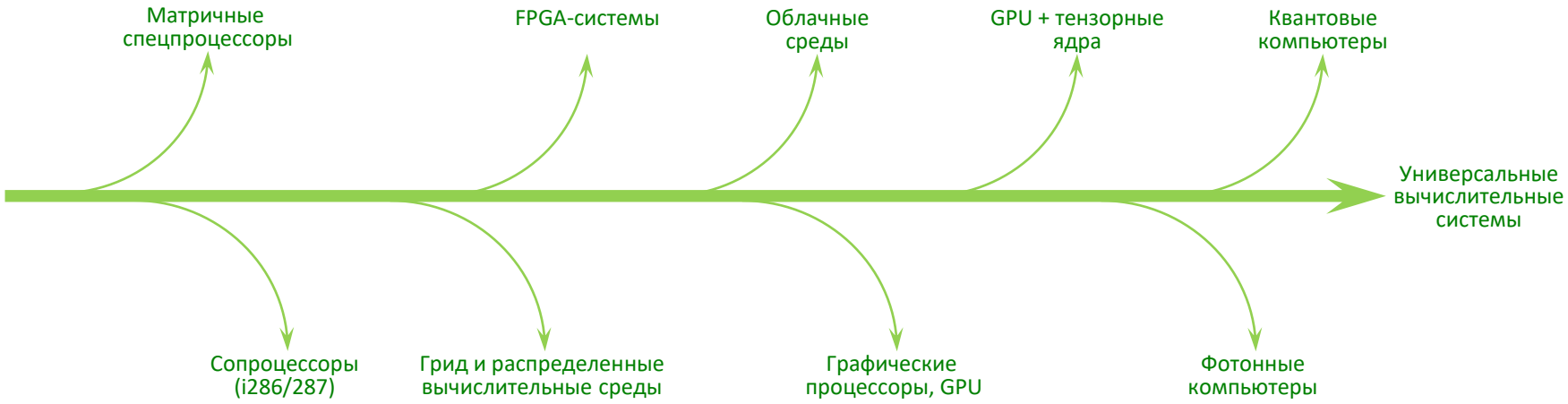
Adam Batson, D.E. Shaw Research, said, “I can tell you that the hardware is physically capable of scaling larger than 512 nodes in terms of the network in the link layer. But the machine is definitely designed to operate normally for where a single molecular dynamic simulation runs on at most 512 nodes...

“For the other applications, in particular material science, it’s possible there are applications that would benefit from Anton. We’ve looked at this somewhat within D.E. Shaw Research, but not a whole lot. Like I said before, you know, we’re not a computer company, our focus is on curing diseases and easing human misery and we just don’t spend a lot of time working on things outside of that scope.”

Развитие архитектур вычислительных систем для эффективного решения классов задач
(кодизайн и специализация)

Возможные причины появления ВС специальной архитектуры:

- Эффективность
- Производительность
- Цена/производительность
- Энергоэффективность
- …



Суперкомпьютер Alice Recoque, Exascale, Europe, CEA

Alice Recoque – второй суперкомпьютер класса Exascale в Европе. Программа EuroHPC JU.

Стоимость – 544 млн.евро. Сдача в эксплуатацию – 2027-2028.

- Europe has many goals with the second supercomputer. “We are looking around and trying to understand what level of ambition we can have there, but the aim is for the system … to have a further increase in European technology compared to what we managed to do in Jupiter…”

- One native European technology will be SiPearl’s ARM-based Rhea-2 chip, which will succeed the Rhea-1 chip in Jupiter.

- Alice Recoque will be for AI and high-precision HPC.

- Like Jupiter, Alice Recoque has a modular design. A system with Rhea-2 CPUs will act as a nerve center onto which modules such as GPU-accelerated and quantum computing systems can be added.

- EuroHPC JU also wants to use accelerators developed by EPI (European Processor Initiative) such as EPAC, which is a RISC-V based vector accelerator.

- Europe has no plans yet for a third exascale system. EuroHPC JU leaders are instead focusing on “post-exascale systems”. Zettascale systems were never part of EuroHPC JU plans, and it seems they understand the technological limitations of exascale systems.

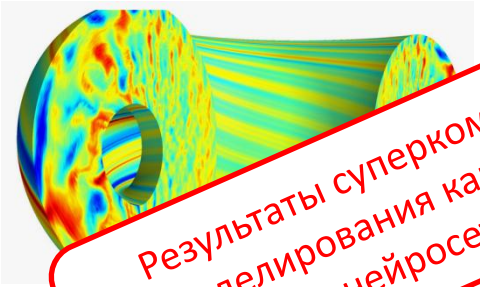
- EuroHPC JU is funneling more money into quantum computing systems, which will be installed at six computing sites. The organization also connects all 29 high-performance systems in Europe.

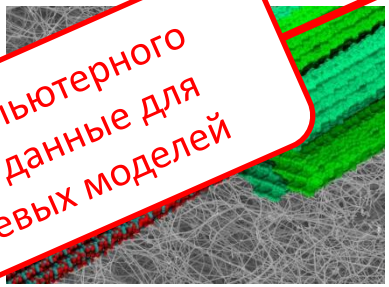
- Europe is also making significant upgrades to existing supercomputers.

Изменение парадигмы вычислительных наук

Традиционное использование высокопроизводительных вычислений (HPC)

Входные параметры HPC-модели → Расчет (часы/дни/недели) → Получение ответа



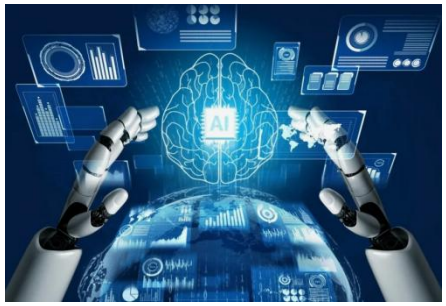


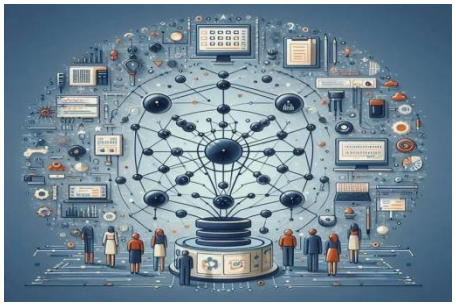
Результаты суперкомпьютерного моделирования как данные для обучения нейросетевых моделей

Цикл научного исследования

Высокопроиз… …ные вычисления, дополненные технологиями ИИ

Сбор/генерация данных для обучения LLM → Обучение LLM (недели/месяцы) → Работа с LLM, inference (секунды/минуты) → Получение ответа





Строительство больших центров для поддержки технологий ИИ

servernews
Все самое свежее из мира больших мощностей

«ИИ-гигафабрика» xAI разместится в гигантском дата-центре в Теннеси

07.06.2024 [15:42], Руслан Авдеев

ИИ-стартап xAI, курируемый Илоном Маском (Elon Musk), намерен построить гигантский дата-центр с самым производительным в мире ИИ-суперкомпьютером. По данным Datacenter Dynamics, ЦОД разместится в окрестностях Мемфиса (штат Теннесси), а пока ожидает одобрения властей.

В обозримом будущем компания должна получить сотни тысяч ускорителей для обучения новых моделей, в частности, чат-бота Grok, предлагаемого, например, по подписке в социальной сети X (Twitter). Ранее в Сеть утекла информация, что NVIDIA передаст xAI чипы, изначально предназначавшиеся для Tesla — Маск весьма вольно распоряжается активами подконтрольных ему бизнесов, часто вызывая недовольство инвесторов.

Пока проект ожидает окончательного разрешения от местного бизнес-инкубатора Memphis Shelby County Economic Development Growth Engine (EDGE), а также муниципальных и федеральных властей. Впрочем, гораздо важнее дождаться одобрения энергетической компании Tennessee Valley Authority (TVA). Реализация проекта сулит появление высокооплачиваемых рабочих мест и увеличение доходов штата, что поможет поддерживать и модернизировать местную инфраструктуру.

b200
h100
hardware
hpc
nvidia
xai
ии
суперкомпьютер
сша

Квантовые компьютеры в суперкомпьютерной инфраструктуре

- IBM announced a 10-year, $100 million initiative with the University of Tokyo and the University of Chicago to develop a quantum-centric supercomputer powered by 100,000 qubits.

- In collaboration with the Leibniz Supercomputing Centre (LRZ), the Q-Exa consortium has integrated a 20-qubit quantum computer into a supercomputer, SuperMUC-NG in Germany. The aim of the project was to connect quantum processing units (QPU) based on superconducting circuits to a supercomputer and to develop interfaces and control tools for this purpose.

- The primary motivation for colocating quantum and classical computers lies in the complementary nature of their computational strengths.

- Unlike classical computers, which operate at room temperature, many types of quantum computers need extremely low temperatures, often near absolute zero, to maintain quantum coherence.

- We’re having to learn a whole new set of operational programs. We have to worry about all these other external factors – humidity, temperature, electromagnetic radiation, electromagnetic interference and vibrations.

- Another critical requirement is the maintenance and calibration of quantum computers. However, the required skills to maintain a quantum computer might be different than a classical HPC.

- On the flip side, quantum computers are very power-efficient, requiring at most tens of kilowatts to operate, as opposed to megawatts required from classical HPCs.

- Today’s quantum computers are incapable of performing long calculations. These perform rapid alternation between a classical and a quantum part, thus trying to overcome the limitations of the quantum machines. Such rapid back and forth between quantum and classical makes data transfer latency of greater importance.

HPC, AI, Quantum…

NCSA RESEARCH EXPERTISE PEOPLE NEWS & EVENTS ABOUT

On the academic research side, there's a lot of interest in understanding the strengths and weaknesses of AI systems – "explainable AI." There's great interest in figuring out how we can build AI systems that don't produce false outputs. NCSA's partnership with the National Deep Inference Fabric provides

How do you increase the speed at which you can do that while reducing the amount of computing and energy it takes? There's a huge amount of research that needs to be done that doesn't require the enormous scale that you find in the commercial sectors. This is one of the goals of the National Artificial

I think there are some very interesting long-term research questions since the methods we're currently using are somewhat brute force. Because of the

but it's not clear that AI is going to be a solution in their work. There's a lot of concern that there is too much focus on AI or quantum – that's another one – and that some of the fundamentals will get lost. We can't allow that to happen,

DELTA OFFERS:

- **Eight utility nodes will provide login access, data transfer capability and other services**
- **200 Gb/s HPE SlingShot network fabric**
- **7 PB of disk-based Lustre storage**
- **3 PB of flash based storage for data intensive workloads**

100 quad A40 GPU nodes consisting of:

- Single AMD 64-core 2.55 GHz Milan processor
- 256 GB DDR4-3200 RAM
- 1.6 TB NVMe solid-state disk
- Four NVIDIA A40 GPUs with 48 GB GDDR6 RAM

One MI100 GPU node consisting of:

- Dual AMD 64-core 2.55 GHz Milan processors
- 2 TB DDR4-3200 RAM
- 1.6 TB NVMe solid-state disk
- Eight AMD MI100 GPUs with 32 GB HBM2 RAM

124 CPU nodes consisting of:

- Dual AMD 64-core 2.55 GHz Milan processors
- 256 GB DDR4-3200 RAM
- 800 GB NVMe solid-state disk

100 quad A100 GPU nodes consisting of:

- Single AMD 64-core 2.55 GHz Milan processor
- 256 GB DDR4-3200 RAM
- 1.6 TB NVMe solid-state disk
- Four NVIDIA A100 GPUs with 40 GB HBM2 RAM and NVLink

Five eight-way A100 GPU nodes consisting of:

- Dual AMD 64-core 2.55 GHz Milan processors
- 2 TB DDR4-3200 RAM
- 1.6 TB NVMe solid-state disk
- Eight NVIDIA A100 GPUs with 40 GB HBM2 RAM and NVLink

DELTA AI OFFERS:

- 456 H100 NVIDIA GPUs
- 200 Gb/s HPE SlingShot network fabric
- Two Lustre file systems (based on HDD and NVME, respectively) shared with Delta to support both block and small file IO.
- Access to project space on the "Taiga" Lustre based center wide project file system
- Home directories provisioned on the "Harbor" VAST based center wide home directory system.

114 CPU-GPU nodes consisting of:

- 4 Grace Hopper GH200 superchips per node
- Each GH200 superchip has one H100 GPU and a 72-core Grace ARM CPU.
- Each H100 has 96GB HBM3
- Each Grace ARM CPU has 120GB of LPDDR5X memory
- 4 Slingshot11 network connections: 1 per Grace Hopper superchip
- One 3.5 TB NVME drive per node

Intersect360 RESEARCH

Addison Snell,
доля рынка:
HPC only, no AI – 10%
AI only, no HPC – 10%
Both, HPC and AI – 80%

Sarov.msu.ru

Msu.ru

Благодарю за внимание !

voevodin@parallel.ru