МГУ им. М.В.Ломоносова 26 сентября 2023 г.

Международная научная конференция
"Суперкомпьютерные дни в России 2023"

Внедрение новых суперкомпьютерных технологий в ОИВТ РАН

В. В. Стегайлов

Содержание

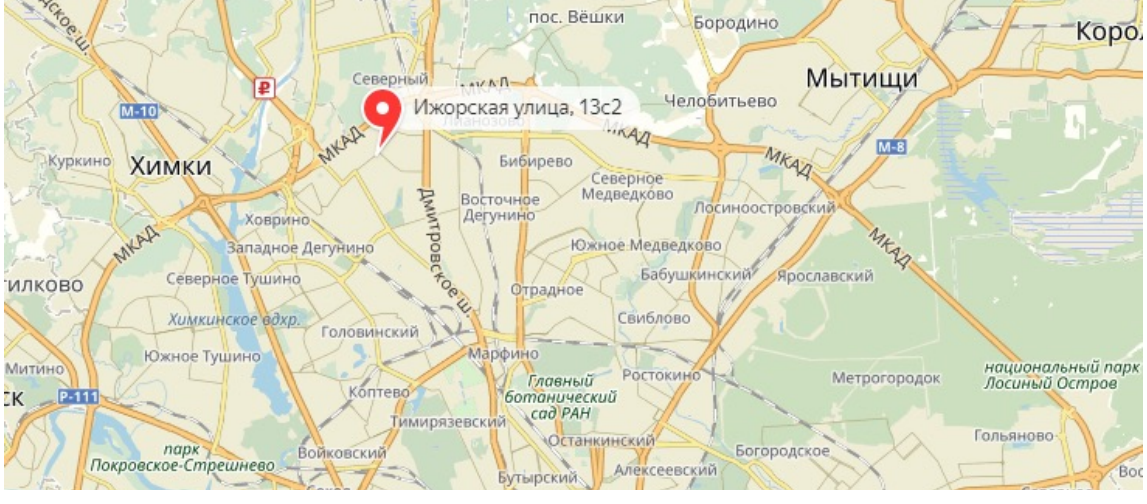
- Краткая история развития Суперкомпьютерного центра ОИВТ РАН
- Зачем нам суперкомпьютеры?

- О тестировании новых технологий
  - Процессоры, графические ускорители
  - Интерконнект Ангара, GPU-aware MPI
  - Файловые системы
  - Ab initio расчеты
  - Энергоэффективность

СУПЕРКОМПЬЮТЕРНЫЙ ЦЕНТР

ОИВТ РАН


ОИВТ РАН


Ижорская улица, 13с2
пос. Вёшки
Бородино
Мытищи
Челобитьево
Северный
МКАД
М-10
Куркино
Химки
Бибирево
Северное Медведково
Восточное Дегунино
Лосиноостровский
М-8
Ховрино
Западное Дегунино
Южное Медведково
Дмитровское ш.
Северное Тушино
Бабушкинский
Ярославский
Отрадное
Химкинское вдхр.
Свиблово
Головинский
Южное Тушино
Марфино
Ростокино
Митино
Коптево
Главный ботанический сад РАН
Метрогородок
национальный парк Лосиный Остров
Р-111
Тимирязевский
Останкинский
Гольяново
парк Покровское-Стрешнево
Войковский
Богородское
Бутырский
Алексеевский

Вычислительная техника в ИВТАНе

1969 г. - введена БЭСМ-4 (20Кфлопс),

в использовании УРАЛ-14 (10-25 Кфлопс).

Вычислительная техника в ИВТАНе

1969 г. - введена БЭСМ-4 (20Кфлопс), в использовании УРАЛ-14.

К 1977 г. - БЭСМ-6 (1Мфлопс), ЕС-1020 (20Кфлопс), HP-2000

В 1978-79 гг. - HP-3000

В 1980-85 гг и позже на БЭСМ-6 работает ИВТАНТЕРМО,
и на HP-3000 работает Банк данных по ТД свойствам.

В 1986 г. – ЕС-1045 (0,8 Мфлопс), М10 (5МФлопс), VAX-6000 (8-63МФлопс)

Академик В.Е.Фортов

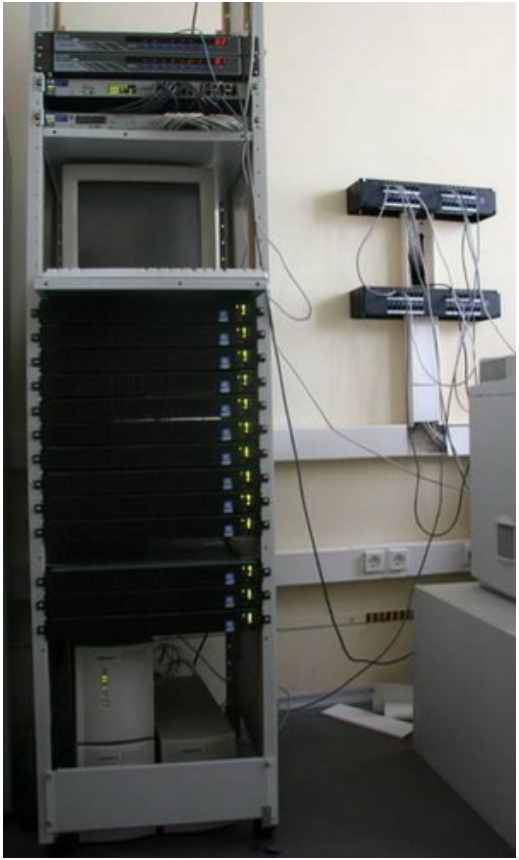
23.01.1946 — 29.11.2020

По инициативе академика Фортова в 1999 году был открыт Межведомственный суперкомпьютерный центр РАН, где был установлен первый российский суперкомпьютер. Это одно из немногих научных мероприятий, которое посетил тогдашний премьер-министр В.В. Путин. На тот момент, что подтверждено сертификатами, кластер вошел в сотню самых мощных компьютеров мира.

Развитие суперкомпьютерного центра

Суперкомпьютер МВС-1000 был поставлен в ОИВТ РАН в 2001 году НИИ «Квант».
Суперкомпьютер имел 16 вычислительных узлов и один управляющий, на каждом узле было установлено по 2 процессора с тактовой частотой 1 ГГц и 1 Гб оперативной памяти.
Пиковая производительность кластера составляла 32 Гфлопс.

Развитие суперкомпьютерного центра



Суперкомпьютер NWO5 был поставлен в ОИВТ РАН в 2005 году на средства РФФИ.
На кластере настроено программное обеспечение для проведения расчетов в GRID-системах.
Суперкомпьютер состоит из 13 вычислительных узлов, имеющих следующие характеристики: 2 процессора Intel Xeon 3.0 ГГц, 2 Гб оперативной памяти на узел, жесткий диск на 160 Гб, сетевые интерфейсы Fast Ethernet и Gigabit Ethernet.
Производительность кластера на тесте LINPACK составляет 108 Гфлопс.

Развитие суперкомпьютерного центра

Суперкомпьютер Т-Платформы TEdge-48 компании Т-Платформы был поставлен в ОИВТ в 2008 году и состоит из 24 вычислительных модулей.
Каждый модуль содержит 2 четырехядерных процессора Intel Xeon 5445 с тактовой частотой 2.33 ГГц и 8 Гбайт оперативной памяти.
Производительность кластера на тесте Linpack составляет 1.4 Tflops.

А ЗАЧЕМ НАМ СУПЕРКОМПЬЮТЕРЫ?

Для каких задач нужны суперкомпьютеры в ОИВТ РАН?

Популярные прикладные пакеты для HPC (с открытым программным кодом):

- Молекулярная динамика (GROMACS, LAMMPS, OpenMM)
- Ab-Initio расчёты (VASP, CP2K, CPMD)
- Газо- и Гидродинамика (FlowVision, OpenFOAM)
- Плазма (PIConGPU, VLPL)

Суперкомпьютер DESMOS

Введен в эксплуатацию в конце 2016 г.

Кластер состоит из 32 гибридных вычислительных узлов и одного головного.

Для объединения узлов используется сеть Ангара в топологии 4-х мерный тор.

2016: прототип на ускорителях **Nvidia GTX1070**,

2018: апгрейд на ускорители **AMD FirePro S9150**
#45 в Топ50 сентября 2018 года – 52.24 (90.75) Тфлопс

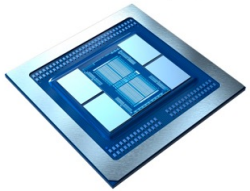
2020: апгрейд на ускорители **AMD Instinct MI50**
#39 в Топ50 марта 2021 года -
85.26 (221.85) Тфлопс

#37 в Топ50 марта 2023 года -
123.38 (221.85) Тфлопс

НИЦ ВТ

NIAGARA
Российские Суперкомпьютеры

Портирование LAMMPS на технологию ROCm HIP



PPAM 2019 Bialystok, Poland, September 8-11, 2019

13th INTERNATIONAL CONFERENCE
ON PARALLEL PROCESSING
AND APPLIED MATHEMATICS

Diploma

PPAM Best Paper Award

The International Conference on Parallel Processing and Applied Mathematics (PPAM) Best Paper Award is given in recognition of the research paper quality, originality and significance of the work in high performance computing.

The PPAM Best Paper was first awarded at PPAM 2019 in Bialystok upon recommendation of the PPAM Chairs and Program Committee.

PPAM 2019 Winner

Evgeny Kuznetsov, Nikolay Kondratyuk, Mikhail Logunov,
Vsevolod Nikolskiy and Vladimir Stegailov

Performance and portability of state-of-art molecular dynamics software on modern GPUs

CHAIR OF PROGRAM COMMITTEE
Roman Wyrzykowski
Prof. Roman Wyrzykowski

VICE-CHAIR OF PROGRAM COMMITTEE
Ewa Deelman
Prof. Ewa Deelman

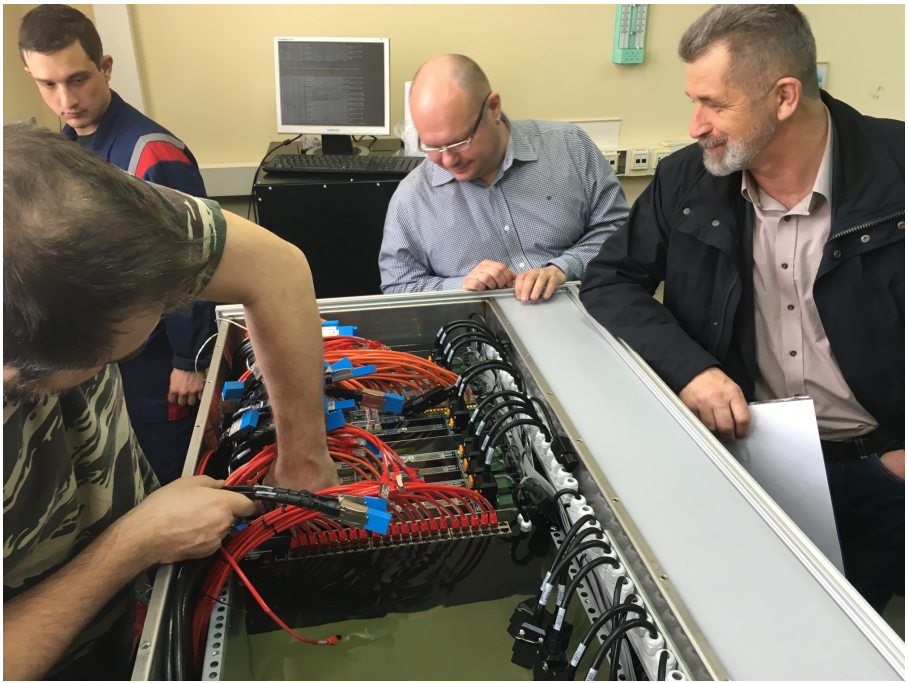
RSF Russian Science Foundation
КПНИ 14-50-00124

Суперкомпьютер ФИШЕР

Суперкомпьютер ФИШЕР введен в эксплуатацию в 2018 г.

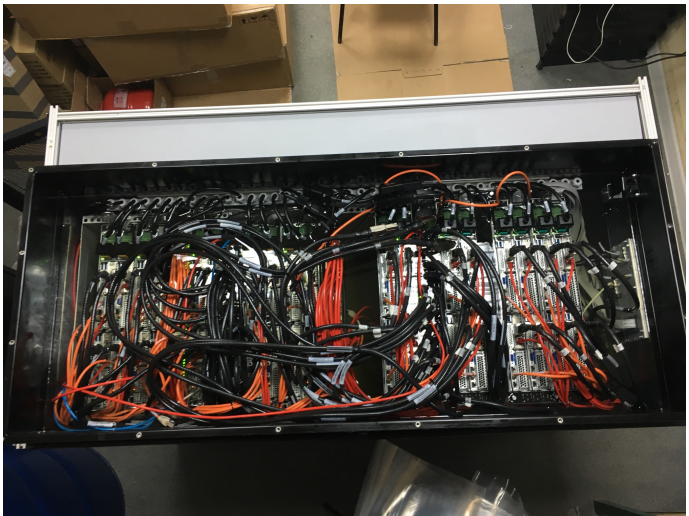
Сегмент с воздушным охлаждением состоит из 18 двухпроцессорных вычислительных узлов на процессорах AMD Epyc и Infiniband FDR.

В марте 2019 года в ОИВТ РАН был запущен сегмент суперкомпьютера ФИШЕР с жидкостным охлаждением (на базе коммутатора Ангара и однопортовых плат PCIe)



Погружная система на процессорах AMD Epyc и интерконекте Ангара

Суперкомпьютер ФИШЕР
ОИВТ РАН



О ТЕСТИРОВАНИИ

И

СРАВНИТЕЛЬНОМ АНАЛИЗЕ

НОВЫХ ТЕХНОЛОГИЙ

ПРОЦЕССОРЫ

Другие новые архитектуры процессоров

ARM®

V Nikolskiy and V Stegailov 2016 J. Phys.: Conf. Ser. 681 012049

Stegailov V., Timofeev A., Dergunov D. //International Conference on Parallel Computational Technologies. – Springer, Cham, 2018. – С. 92-103

PAPER • OPEN ACCESS

Floating-point performance of ARM cores and their efficiency in classical molecular dynamics

V Nikolskiy[1,2] and V Stegailov[2,1]

Published under licence by IOP Publishing Ltd

Journal of Physics: Conference Series, Volume 681, International Conference on Computer Simulation in Physics and Beyond 2015 6–10 September 2015, Moscow, Russia

Citation V Nikolskiy and V Stegailov 2016 *J. Phys.: Conf. Ser.* **681** 012049

ИНТЕРКОННЕКТ

АНГАРА

Суперкомпьютерный интерконнект: международный контекст

NVIDIA NETWORKING

Mellanox TECHNOLOGIES

~~intel~~

~~Intel® Omni-Path Architecture~~

CRAY

Slingshot Interconnect

Atos

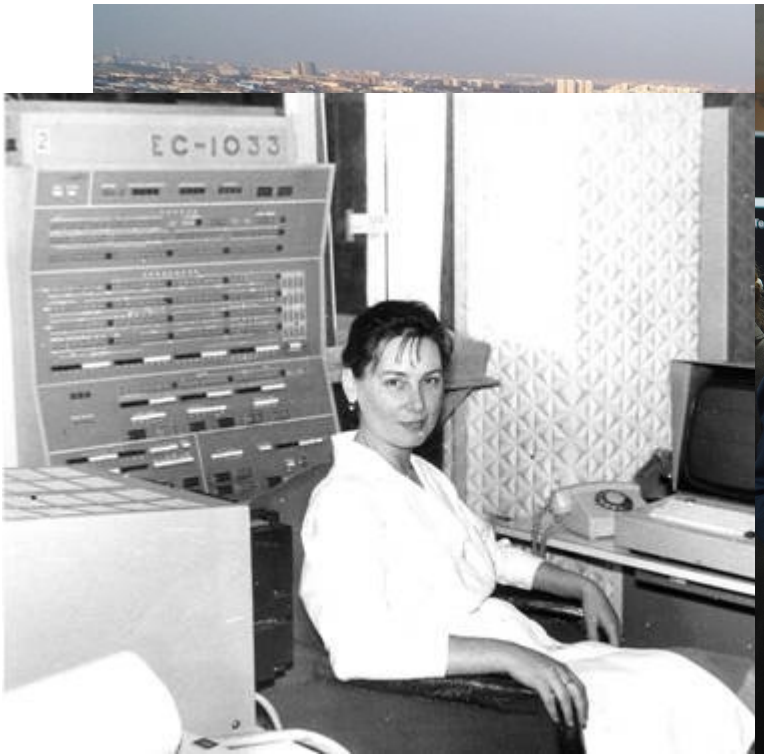
Bull eXascale Interconnect (BXI)

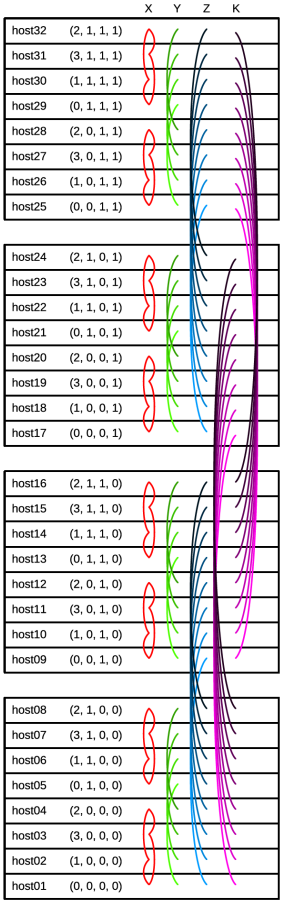
FUJITSU

Tofu D Interconnect

Sunway Interconnect

АО НИЦЭВТ: сеть Ангара



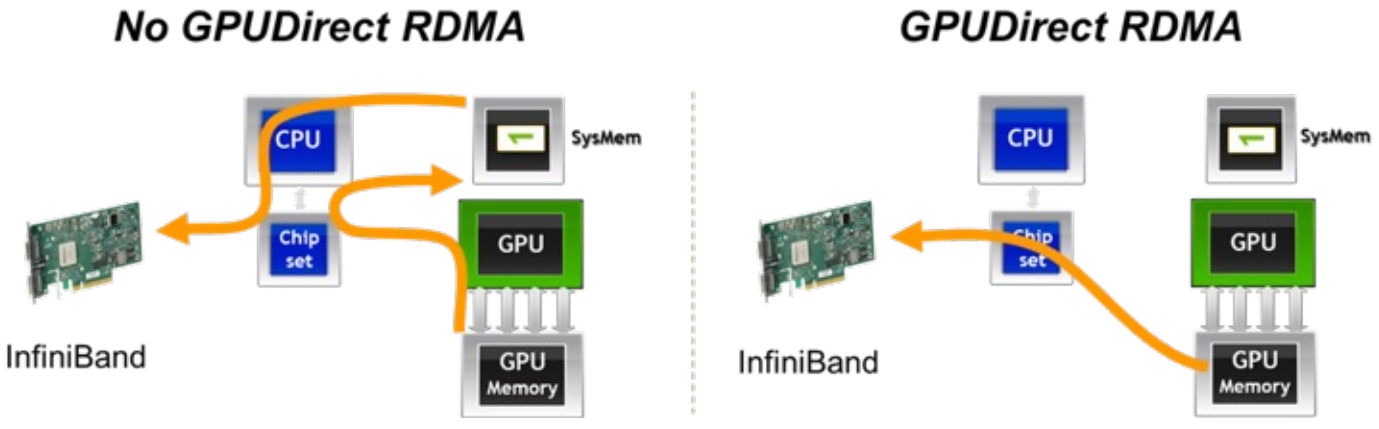


См.: LNCS **10777**, p. 327, 2018

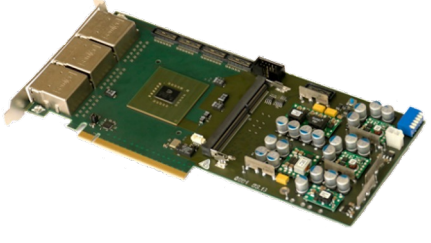
ТЕХНОЛОГИЯ

GPUDIRECT RDMA

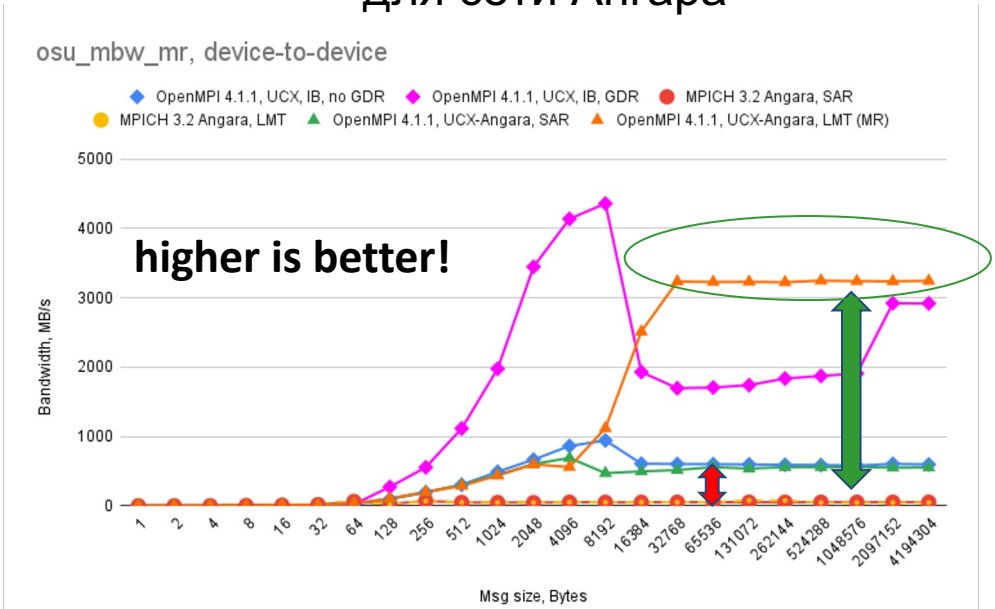
Технология GPUDirect RDMA



Реализация поддержки семейства технологий GPUDirect для сети Ангара и графических ускорителей AMD



Результаты разработки поддержки GPUDirect RDMA для сети Ангара

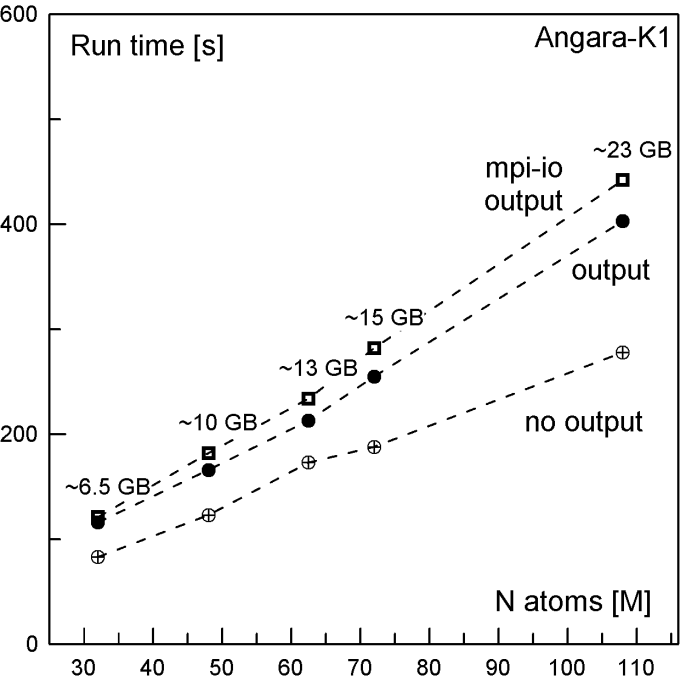


- Использование rocm_copy (GPUDirect) позволяет добиться **ускорения в ~10 раз** при P2P-пересылках между памятью GPU
- Ускорение с использованием протокола Rendezvous в связке с API pkba_reg_mr (GPUDirect RDMA) достигает ~**68 раз**

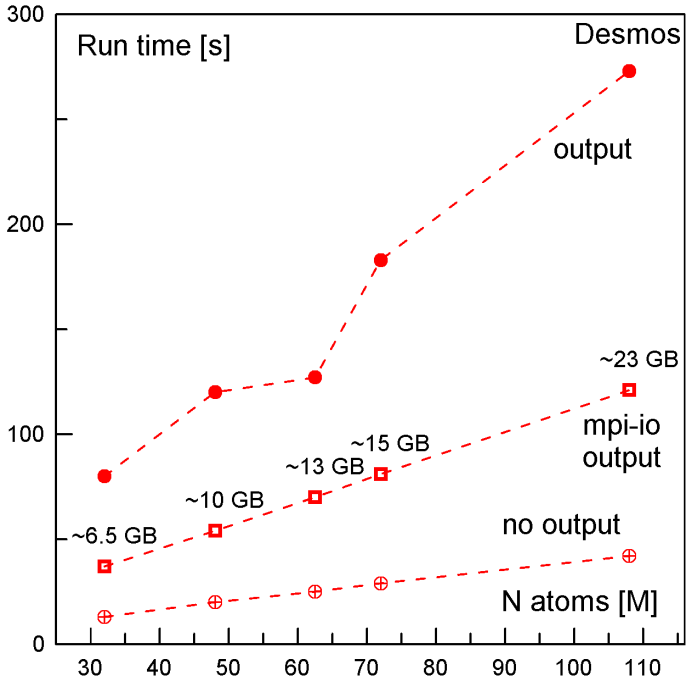
См.: LNCS **13708**, p. 591, 2022

ПАРАЛЛЕЛЬНЫЕ ФАЙЛОВЫЕ СИСТЕМЫ

Slowing down of MD calculations with LAMMPS due to massive data output



Single dedicated storage server (Lustre)



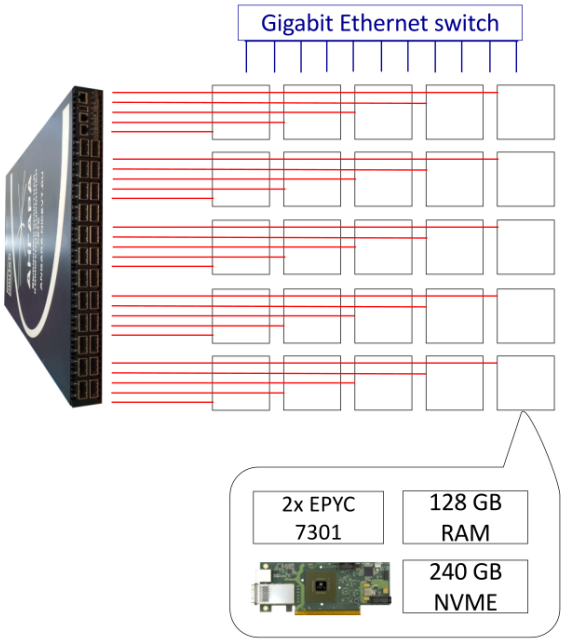
Distributed storage on the nodes (BeeGFS)

См.: CCIS **910**, p. 77, 2018

Сегмент суперкомпьютера Fisher (ОИВТ РАН)

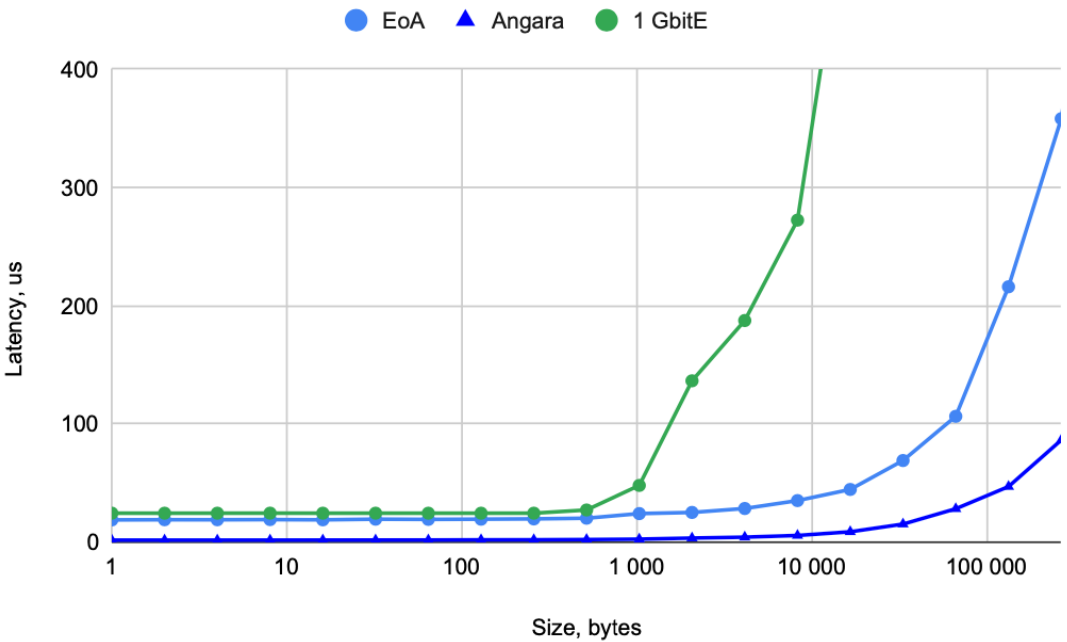
Hardware	
Number of nodes	20
SSD M2 NVMe	Apacer AS2280P2

Software	
OS	OpenSUSE Leap 15.2
Kernel	5.3.18-lp152.87-preempt
Ethernet over Angara	2.1
BeeGFS	7.2.3



См.: LNCS **13708**, p. 423, 2022

Программный стек TCP/IP для интерконекта Ангара



См.: LNCS **13708**, p. 423, 2022

IO500: 16 storage nodes, 4 client nodes, 16 MPI / node

Тест	Характеристика	1 Gbit Ethernet	EoA	EoA / 1 Gbit Ethernet
ior-easy-write	GiB/s	0,459	2,612	5,69
mdtest-easy-write	kIOPS	15,180	17,170	1,13
ior-hard-write	GiB/s	0,171	0,419	2,44
mdtest-hard-write	kIOPS	5,101	3,808	0,75
find	kIOPS	175,189	124,549	0,71
ior-easy-read	GiB/s	0,448	6,498	**14,50**
mdtest-easy-stat	kIOPS	74,599	73,687	0,99
ior-hard-read	GiB/s	0,456	2,908	6,38
mdtest-hard-stat	kIOPS	67,311	70,303	1,04
mdtest-easy-delete	kIOPS	14,203	10,181	0,72
mdtest-hard-read	kIOPS	14,592	15,033	1,03
mdtest-hard-delete	kIOPS	4,187	5,952	1,42
SCORE				
Bandwidth	GiB/s	0,356	2,132	**5,99**
IOPS	kIOPS	22,205	21,042	**0,95**
Total		2,812	6,698	**2,38**

См.: LNCS **13708**, p. 423, 2022

СУПЕРКОМПЬЮТЕРЫ

ДЛЯ AB INITIO РАСЧЕТОВ

Полностью неэмпирический первопринципный метод для конденсированных сред с учетом обмена и корреляции на уровне HF+MP2

Table II. Average time for the evaluation of the energy (single MC cycle) at the various level of theory considered. The computational setups are those specified in this section. Timing measured on a CRAY-XC30 machine, each node is equipped with 8-core CPU and a graphics processing unit (K20X GPU).

	Number of Nodes	Time [s]
GGA	32	7.9
meta-GGA	32	17.4
vdW-DF	32	12.0
hybrid-GGA (ADMM)	48	21.3
meta-hybrid-GGA (ADMM)	48	27.9
hybrid-GGA	128	51.0
PWPB95-D3	200	198
RPA	200	200
MP2 (MC: energy only)	512	163
MP2 (MD: including forces)	2048	257

64 H_2O molecules in a cubic box under periodic boundary conditions

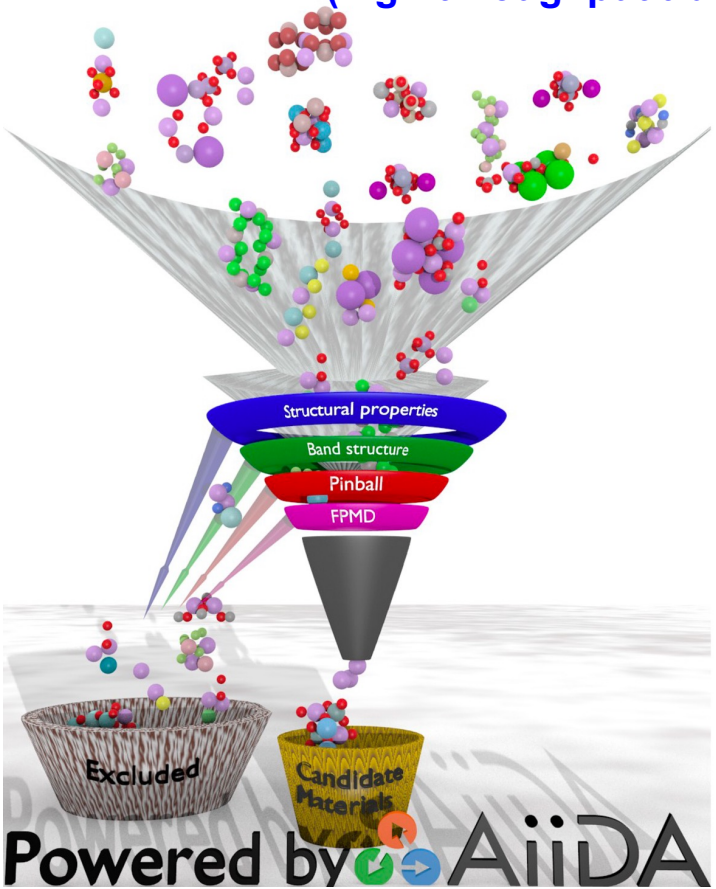
We see the non-trivial prediction that ice floats on water, with a quantitatively correct ratio of liquid and solid density

TABLE II. Equilibrium volumes and energies (at 0 K) for ice I*h* expressed per molecule. The calculated values are obtained from the minimum of the curves of Figure 5. No corrections for the quantum nature of the nuclei and zero point energies (ZPEs) have been considered. Experimental values are from Refs. 6 and 158.

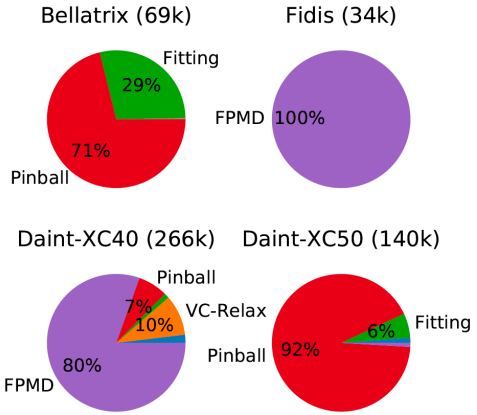
	E_{coh} (kJ/mol)	V_{mol} (Å^3)	ρ (g/ml)
PBE	−62.8	30.69	0.975
MP2	−58.7	31.34	0.955
PWPB95-D3	−58.5	32.15	0.930
(EEX+RPA)@PBE	−52.5	32.37	0.924
Expt.	−58.9	32.05	0.933

Mauro Del Ben et al., 2015

Вычислительный скрининг материалов (high-throughput ab initio calculations)



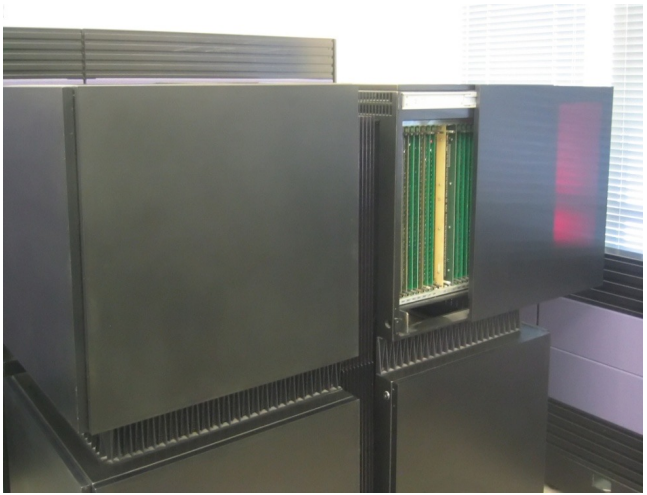
2503 SCF-расчетов
5214 релаксационных расчетов
171370 огрубленных МД расчетов
11525 ab initio МД расчетов



Итого: 509 тыс. узло-часов
= 15 млн. ядро-часов (!)

Kahle, L., Marcolongo, A., & Marzari, N. (2020). High-throughput computational screening for solid-state Li-ion conductors. Energy & Environmental Science, 13(3), 928-948.

Суперкомпьютеры 1986-87 гг. с массово-параллельной архитектурой



Thinking Machines CM-2:
16384 однобитовых процессора совместно с 512 арифметическими ускорителями Weitek

Meiko Computing Surface:
64 транспьютерных узлов с процессорами Intel i860

В. В. Стегайлов, Г. Э. Норман. Проблемы развития суперкомпьютерной отрасли в России: взгляд пользователя высокопроизводительных систем // Программные системы: теория и приложения: электрон. научн. журн. 2014. Т. 5, № 1(19), с. 111–152. URL: http://psta.psiras.ru/read/psta2014_1_111-152.pdf

… и параллельные алгоритмы для первопринципных расчетов

VOLUME 68, NUMBER 9 PHYSICAL REVIEW LETTERS 2 MARCH 1992

Ab Initio Theory of the Si(111)-(7×7) Surface Reconstruction: A Challenge for Massively Parallel Computation

Karl D. Brommer,[1] M. Needels,[2] B. E. Larson,[3] and J. D. Joannopoulos[1]
[1]*Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*
[2]*AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974*
[3]*Thinking Machines, Cambridge, Massachusetts 02139*
(Received 8 November 1991)

An *ab initio* investigation of the Si(111)-(7×7) surface reconstruction is undertaken using the state of the art in massively parallel computation. Calculations of the total energy of an ~700 effective-atom supercell are performed to determine (1) the fully relaxed atomic geometry, (2) the scanning tunneling microscope images as a function of bias voltage, and (3) the energy difference between the (7×7) and the (2×1) reconstructions. The (7×7) reconstruction is found to be energetically favorable to the (2×1) surface by 60 meV per (1×1) unit cell.

PACS numbers: 73.20.−r, 68.35.Bs, 68.35.Md

Thinking Machines CM-2

VOLUME 68, NUMBER 9 PHYSICAL REVIEW LETTERS 2 MARCH 1992

Ab Initio Total-Energy Calculations for Extremely Large Systems: Application to the Takayanagi Reconstruction of Si(111)

I. Štich, M. C. Payne, R. D. King-Smith, and J-S. Lin
Cavendish Laboratory (TCM), University of Cambridge, Madingley Road, Cambridge CB3 0HE, United Kingdom

L. J. Clarke
Edinburgh Parallel Computer Centre, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, United Kingdom
(Received 8 November 1991)

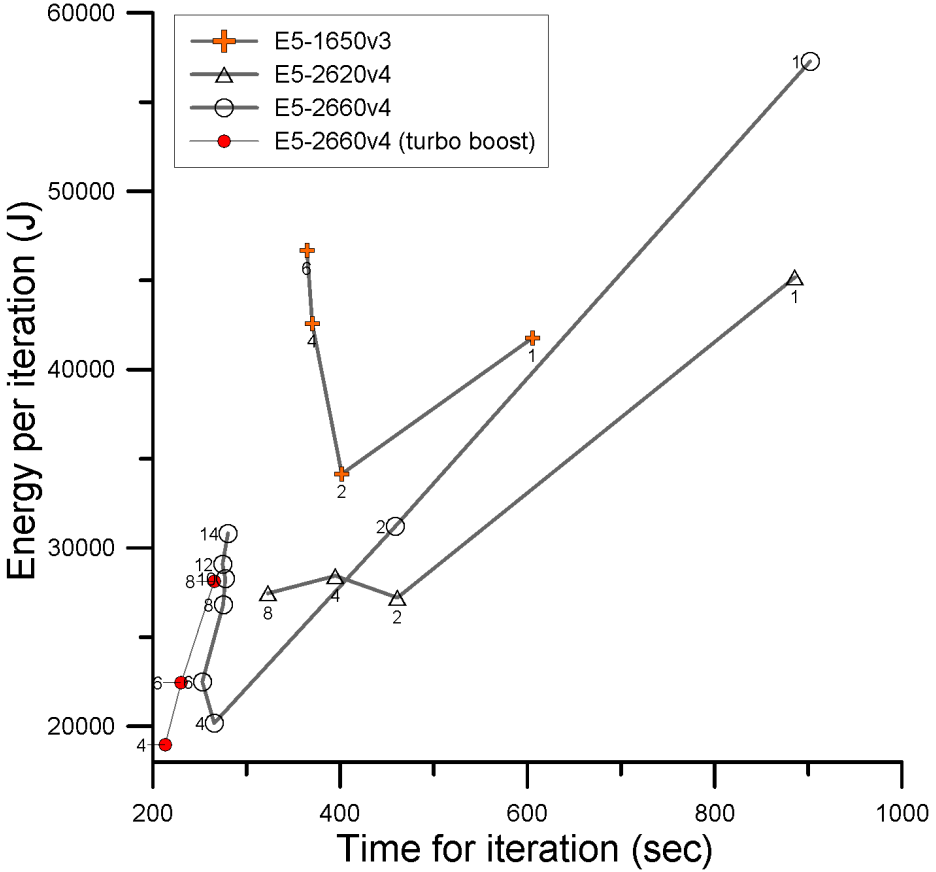
We have implemented a set of total-energy pseudopotential codes on a parallel computer which allows calculations to be performed for systems containing many hundreds of atoms in the unit cell. Using these codes we have calculated the total energies and structures of the 3×3, 5×5, and 7×7 Takayanagi reconstructions of the (111) surface of silicon. We find that the 7×7 structure minimizes the surface energy and observe structural trends across the series which can be correlated with the degree of charge transfer between the dangling bonds on the adatoms and rest atoms.

PACS numbers: 68.35.−p, 31.20.−d, 71.45.Nt

Meiko Computing Surface

ЭНЕРГОЭФФЕКТИВНОСТЬ

Энергоэффективность CPU для кода VASP



Intel Xeon
Haswell/Broadwell
E5-1650v3
6 cores at 3.5 GHz
15Mb of L3 cache
$583

E5-2620v4
8 cores at 2.1 GHz
20 Mb of L3 cache
$417

E5-2660v4
14 cores at 2.0 Ghz
35 Mb of L3 cache
$1445
with turboboost

См.: LNCS **10778**, p. 81, 2018

VASP: speed and energy consumption

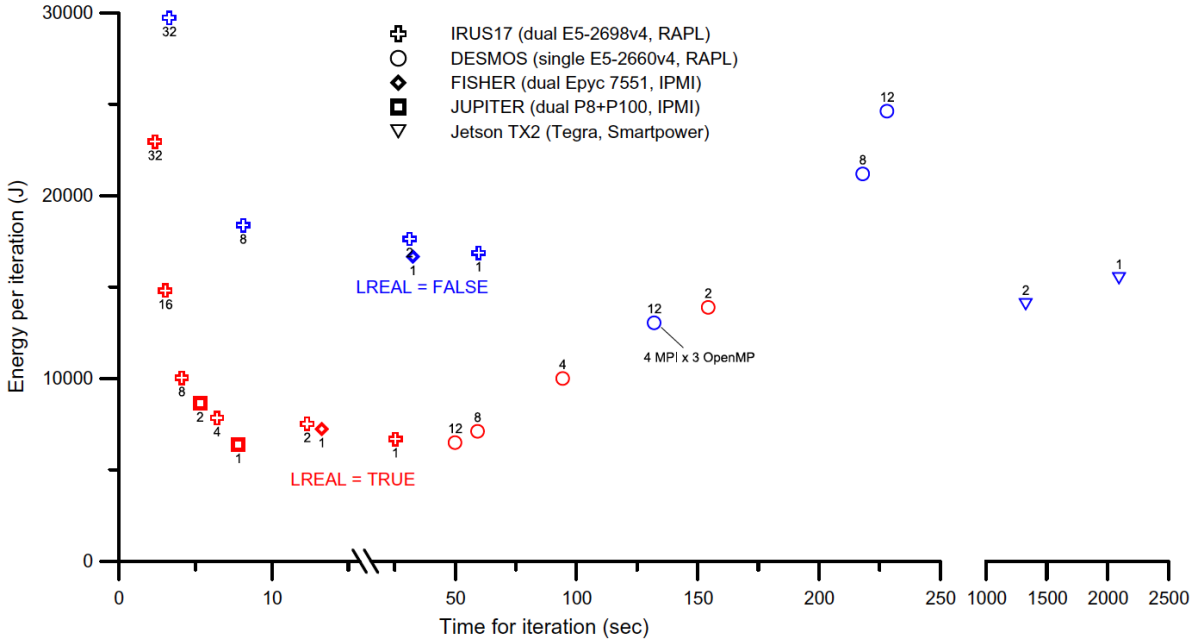


FIGURE 5 The energy-to-solution for the GaAs test model calculation on different platforms: LREAL=TRUE points are shown in red, LREAL=FALSE points are shown in blue. The number of active cores in a single node is shown as a number *above* a data point. A number *below* a data point shows the number of nodes used for the calculation (assuming in each case the optimal configuration of MPI processes and OpenMP threads at the nodes)

Stegailov, V., Smirnov, G., & Vecher, V. (2019). VASP hits the memory wall: Processors efficiency comparison. **Concurrency and Computation: Practice and Experience**, e5136.

Эффект – высокий уровень результатов и признание работ ОИВТ РАН по математическому моделированию

В России

- Премия Президента РФ для молодых учёных
  - Стегайлов В.В., зав.отд. ОИВТ

- 1е место на международном конкурсе молодых ученых «Нефтегазовые проекты: взгляд в будущее-2021» Газпрома.

Коллектив:
Писарев В.В., зав. лаб. ОИВТ,
Кондратюк Н.Д., с.н.с. ОИВТ,
и др.

В мире

- 1е место на международном индустриальном конкурсе по моделированию свойств жидкостей

Коллектив:
Писарев В.В., зав. лаб. ОИВТ, в.н.с.ВШЭ
Кондратюк Н.Д., с.н.с. ОИВТ, зав.лаб. МФТИ

Заключение

- ОИВТ РАН активно ведёт исследования
  - по оптимизации производительности научных кодов на современных компонентах, включая перенос на AMD HIP,
  - по подбору оптимальной конфигурации процессоров (включая EPYC 2), ускорителей и других компонент для научных задач,
  - по разработке поддержки GPUDirect RDMA для сети Ангара,
  - энергоэффективности вычислительных систем с воздушным и иммерсионным охлаждением.

- СКЦ ОИВТ РАН исследует перспективные направления суперкомпьютерных технологий (AMD EPYC2, MI50, Ангара, ARM, Эльбрус и др). Мы готовы делиться опытом и предоставлять тестовый доступ.