



## Ноябрь 2022: состояние и перспективы развития суперкомпьютерной отрасли в мире и в России

Николай Сергеевич **Абрамов**<sup>1✉</sup>, Сергей Михайлович **Абрамов**<sup>2</sup>

Институт программных систем им. А. К. Айламазяна РАН, Вельково, Россия

<sup>1✉</sup>n-say@nsa.pereslavl.ru

**Аннотация.** Статья посвящена анализу современного состояния суперкомпьютерной отрасли в мире и в России и перспектив ее развития. По методическому подходу и инструментарию данная статья является продолжением серии подобных аналитических обзоров авторов. Новизна данной работы не только в актуализации сведений с учетом последних редакций списка Top500, но и в акценте на следующее: мировая суперкомпьютерная отрасль переживает радикальную перестройку – переход от «петафлопсной эпохи» к «эксафлопсной эпохе». России с некоторым запаздыванием предстоит сделать такой же переход. Поэтому важно понимать технологические тренды и особенности решений для самых мощных систем в мире за последние годы. На это обстоятельство в работе обращается особое внимание и такой акцент при анализе трендов является одной из побудительных целей данной работы.

**Ключевые слова и фразы:** Top500, суперкомпьютеры, гибридные архитектуры, интерконнект

Для цитирования: Абрамов Н. С., Абрамов С. М. *Ноябрь 2022: состояние и перспективы развития суперкомпьютерной отрасли в мире и в России* // Программные системы: теория и приложения. 2023. Т. 14. № 2(57). С. 49–93. [https://psta.psisras.ru/read/psta2023\\_2\\_49-93.pdf](https://psta.psisras.ru/read/psta2023_2_49-93.pdf)

## Введение

В ведущих странах мира суперкомпьютерные технологии (СКТ) властью и обществом достаточно давно рассматриваются как единственный инструмент обеспечения конкурентных преимуществ [1]. Сегодня, в эпоху разворачивания в мире цифровой экономики, усиливаются роли суперкомпьютерной отрасли и суперкомпьютерной киберинфраструктуры (СК-инфраструктуры) ведущих стран. Учитывая планы России совершить рывок и построить в стране цифровую экономику, важными и актуальными становятся истинная оценка сегодняшнего состояния СК-инфраструктуры нашей страны, достоверное сравнение наших позиций в суперкомпьютерной отрасли с позициями ведущих стран мира, точное представление о современных тенденциях развития СКТ.

В настоящее время мировая суперкомпьютерная отрасль переживает радикальную перестройку: идет переход от «петафлопсной эпохи» к «эксафлопсной эпохе». И это связано не просто с увеличением производительности суперкомпьютеров в 1 000 раз. В процессе этого перехода пришлось искать решение большого количества проблем как в области аппаратных средств, так и в части программного обеспечения [2].

России с некоторым запаздыванием предстоит сделать такой же переход. Поэтому очень важно понимать технологические тренды и особенности решений для самых мощных систем в мире за последние годы. Это даст возможность выделить устоявшиеся решения, обеспечившие переход к эксаскейльным системам. На это обстоятельство в работе обращается особое внимание и такой акцент при анализе трендов является одной из побудительных целей данной работы.

По методическому подходу и инструментарию данная статья является продолжением серии подобных аналитических обзоров авторов [3, 4]. Новизна материала данной работы не только в актуализации сведений с учетом последних редакций списка Top500, но и в акценте на отмеченный выше переход к эксафлопсной эпохе.

В следующих разделах анализируются:

- Список Top500 – раздел 1;
- Темпы развития суперкомпьютерных технологий – раздел 2;
- Самые мощные публичные супер-ЭВМ с июня 2011 года по ноябрь 2022 года – раздел 3;
- Тенденции развития гибридных архитектур – раздел 4;
- Тенденции в технологиях интерконнекта – раздел 5;
- Тенденции в технологиях центральных процессоров – раздел 6;

- Компании-производители суперкомпьютеров – раздел 7;
- Области применения суперкомпьютеров – раздел 8;
- Состояние суперкомпьютерной инфраструктуры в мире и в России – раздел 9.

## 1. Список Top500

В настоящее время обсуждение «эксафлопсной эпохи» и анализ всемирного рейтинга пятисот самых производительных суперкомпьютеров в мире – списка *Top500*<sup>[URL](#)</sup> активно ведется учеными. Так, например, в работе [5] показано, что скачок производительности (Flops) в 1000 раз произошел, но при этом пропускная способность памяти увеличилась незначительно. Авторы отмечают, что благодаря, в том числе, решениям с многокристальной компоновкой, лучшими сетями и охлаждением, был достигнут рубеж в 1 EFlops. Работа [6] посвящена исследованию роли Китая в суперкомпьютерной отрасли. Авторы ставят акцент на том, что кроме списка Top500 необходимо учитывать награды за достижения в области высокопроизводительных вычислений – премия *Гордона Белла*<sup>[URL](#)</sup>. При анализе списка Top500 ученым выявлены три основные тенденции в развитии отрасли: широкое распространение гетерогенных архитектур при построении суперкомпьютерных систем; приложения искусственного интеллекта становятся одними из основных, применяемых в HPC-системах; повышение сложности комплексного моделирования различных процессов. Анализ системного баланса и архитектурных тенденций при построении самых мощных суперкомпьютеров приведен в [7]. Рассматриваются все доступные редакции списка Top500 (до 2020 года), исследуются тренды корреляции между производительностью и системными атрибутами (память, ввод-вывод, интерконнект, питание и др.). Так, авторы приводят энергоэффективность вычислений (отношение вычислительной мощности к потребляемой электроэнергии) и отмечают рост корреляции Пирсона между вычислительной мощностью суперкомпьютеров на высоких позициях списка Top500 и их энергоэффективностью.

Основным источником сведений для анализа в настоящей статье также является список Top500. Редакции рейтинга обновляются два раза в год, в июне и в ноябре. Таким образом с июня 1993 года (первая редакция) по ноябрь 2022 года вышло 60 редакций. В списке Top500 системы ранжируются по реальной производительности, измеренной на тесте Linpack – Linpack-производительность. Везде далее мы будем под производительностью понимать только Linpack-производительность, даже когда это явно не подчеркивается. В списке Top500 ее обозначают *RMax*.

Анализ списка проводится в долях  $RMax$ , методология анализа состояния суперкомпьютерной отрасли будет описана ниже.

Имея ввиду тот факт, что некоторые системы по разным причинам не представлены в списке, под определением термина *суперкомпьютер* здесь будем понимать следующее: *суперкомпьютер – это вычислительная система, имеющая такую производительность, которая соответствует производительностям машин, включенных в соответствующую редакцию списка Top500.*

При таком определении данные, доступные на сайте Top500, являются достоверным и многоплановым описанием состояния суперкомпьютерной отрасли. Причем это описание ретроспективное – доступны все шестьдесят редакций рейтинга, с шагом в полгода и весьма подробное. Каждая редакция рейтинга – это таблица из 500 записей, в каждой записи около 40 полей с различными сведениями про каждый суперкомпьютер. Ключевые моменты методики анализа этого массива данных:

- Суперкомпьютеры, входящие в список Top500, очень сильно отличаются друг от друга по своей ключевой характеристике – производительности. И, как следствие, они сильно отличаются по используемым техническим решениям, цене, потребительским свойствам и т. д. Крайне сильное «расслоение» в мире суперкомпьютеров на примере редакции Top500 за ноябрь 2022 года иллюстрирует рисунок 1.

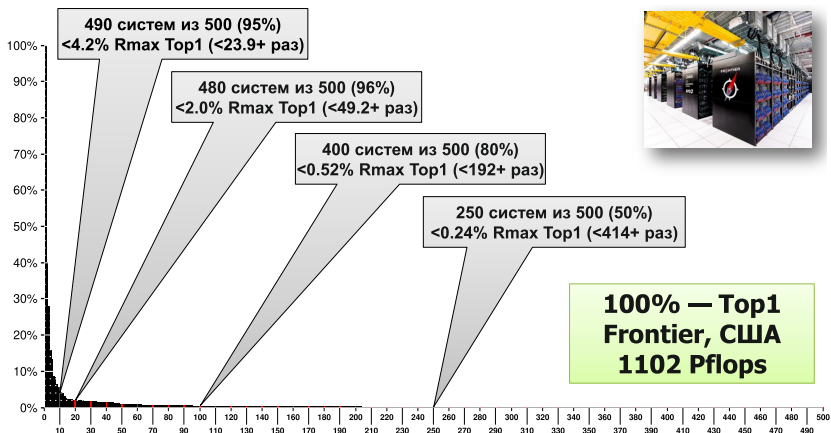
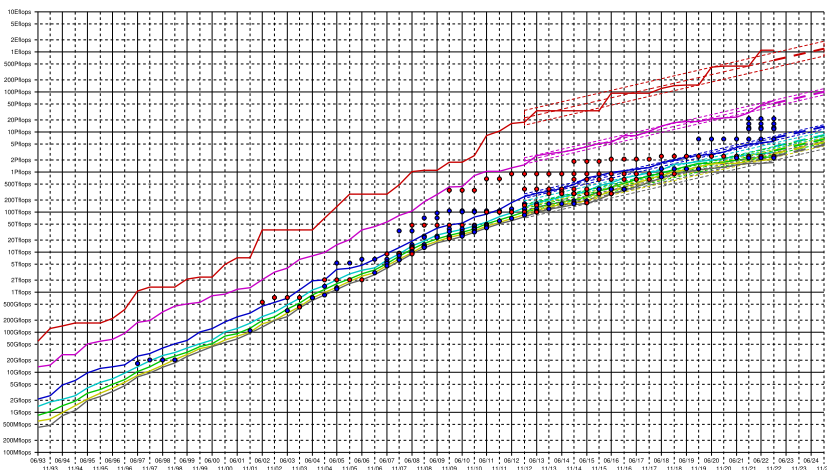


Рисунок 1. Относительная производительность 500 систем, вошедших в рейтинг Top500 в ноябре 2022 года. За 100% принята производительность системы Top1

- Как следствие, ошибочно считать суперкомпьютеры «штуками», необходимо опираться на ключевую характеристику – производительность. Например, сравнивая оснащенность различных стран суперкомпьютерной техникой, необходимо обращать больше внимания на суммарную производительность доступных стране суперкомпьютеров, а не на их количество.

## 2. Темпы развития суперкомпьютерных технологий

Общую оценку темпов развития СКТ можно получить из анализа достигнутой производительности суперкомпьютеров в различные моменты времени. На рисунке 2 приведена соответствующая информация.



Рисунком 2. Производительность суперкомпьютеров, входящих в различные редакции рейтинга Top500

Дадим необходимые пояснения к рисунку 2:

- На оси абсцисс отмечены моменты времени (месяц и год) выхода редакций Top500, на оси ординат – производительность, шкала полулогарифмическая;
- Сплошными цветными линиями показаны графики производительности систем, занимающих различные места в рейтинге: верхний красный график – производительность системы, занимающей первое место; ниже по порядку идут графики производительности

систем, занимающих 10 место (сиреневый), 100 место (синий), 200 место (светло-бирюзовый), 300 место (светло-зеленый), 400 место (светло-травянистый) и 500 место (темно-зеленый);

- На интервале времени с июня 2012 по ноябрь 2024 года пунктирными линиями показан прогноз развития производительности систем, соответствующих 1, 10, 100, 200, 300, 400 и 500 месту в рейтинге – обозначено среднее значение прогноза и интервал 90% достоверности. Для вычисления прогноза использована групповая линейная регрессия (определяется ниже, Раздел 2.1) на отрезке последних 10 лет – с ноября 2012 по ноябрь 2022;
- Цветные точки на графике отмечают системы, установленные в России и попавшие в соответствующую редакцию списка Top500. Точки синего цвета – суперкомпьютеры, закупленные за рубежом, точки красного цвета – отечественные суперкомпьютеры.

Графики, подобные приведенным на рисунке 2, часто публикуются и служат иллюстрацией экспоненциальной скорости развития суперкомпьютерной отрасли. И правда, на первый взгляд цветные линии на рисунке 2 близки к прямым линиям, что с учетом полулогарифмической шкалы соответствует экспоненциальной зависимости. . . Однако, более аккуратный анализ показывает, что более всего графики похожи на ломаную линию из двух (как минимум) прямых отрезков – до 2008 года линейная регрессия имеет один наклон, а после 2008 года другой, более пологий.

## 2.1. Групповая линейная регрессия

Опишем методику упомянутого выше анализа. Занумеруем все выпуски рейтинга Top500 номерами  $i \in [1, 60]$ . Обозначим производительность системы, занявшей некое место  $j \in [1, 500]$  в выпуске Top500 с номером  $i$  через  $r_{i,j}$ . Зададимся неким набором мест в рейтинге Top500:  $J \subseteq [1, 500]$  и неким номером рейтинга  $k \in [8, 53]$  – мы здесь отступили по 7 выпусков рейтинга слева и справа в отрезке  $[1, 60]$ .

Тогда анализом тренда изменения производительности методом групповой линейной регрессии для заданных  $k$ ,  $J$  назовем задачу поиска такого набора параметров  $a$  и  $b_j$ , где  $j \in J$ , что серии данных  $s_j = \{(i, \ln r_{i,j}) \mid i \in [k-7, k+7]\}$  для  $j \in J$ , наилучшим образом приближены линейными функциями  $f_j(i) = a \cdot i + b_j$  от переменной  $i$ . А именно,

минимизируется сумма квадратов:

$$\sum_{i \in [k-7, k+7], j \in J} (a \cdot i + b_j - \ln r_{i,j})^2 \rightarrow \min$$

Задача минимизации решается методом наименьших квадратов (МНК) и, по сути, это несложная модификация стандартного алгоритма линейной регрессии: регрессия проводится для группы наборов данных  $s_j$ , заданных индексами  $j \in J$ , считаем, что угол наклона всех аппроксимирующих прямых линий (параметр  $a$ ) одинаков, а сдвиги по оси  $Y$  (параметры  $b_j$ ) различные.

По смыслу решенной задачи можем сказать, что в момент выпуска Top500 номер  $k$ , за полгода производительность систем, занимающих места  $j \in J$  в среднем, возрастала в  $e^a$  раз. Значит в 1000 раз эта производительность вырастет (в среднем) за  $y(k, J) = \frac{\ln 1000}{2 \cdot a}$  лет.

## 2.2. За сколько лет производительность суперкомпьютеров вырастала в 1000 раз?

Величина  $y(k, J)$  была рассчитана для нескольких наборов  $J$  и для всех  $k \in [8, 52]$ . Рисунок 3 показывает результаты расчётов. На оси  $X$  вместо величины  $k$  подписана календарная дата выпуска рейтинга Top500 с номером  $k$ . По оси  $Y$  указана величина  $y(k, J)$  (годы).

Видно, что на интервале 1998–2008 действительно можно было говорить, что  $R_{\max}$  увеличивался в 1000 раз примерно за 11 лет. И это было верно и для лидирующих мест в суперкомпьютерном рейтинге (красная и сиреневая линия графика), и для всего рейтинга в целом (зеленая линия). Далее (2009–2016) мы наблюдаем признаки явных технических сложностей, которые сдерживали рост технологий в отрасли. В последнее время ситуация стала исправляться для старших систем в рейтинге, и сохранились негативные тенденции для рейтинга в целом – то есть, *нарастает разрыв* между старшими системами и всеми остальными.

## 2.3. Основные выводы из анализа темпов развития суперкомпьютерных технологий

Можем констатировать, что на рубеже 2008 года, при сохранении экспоненциального характера роста производительности, уменьшилась скорость роста – уменьшилось основание этой самой экспоненты. До 2008

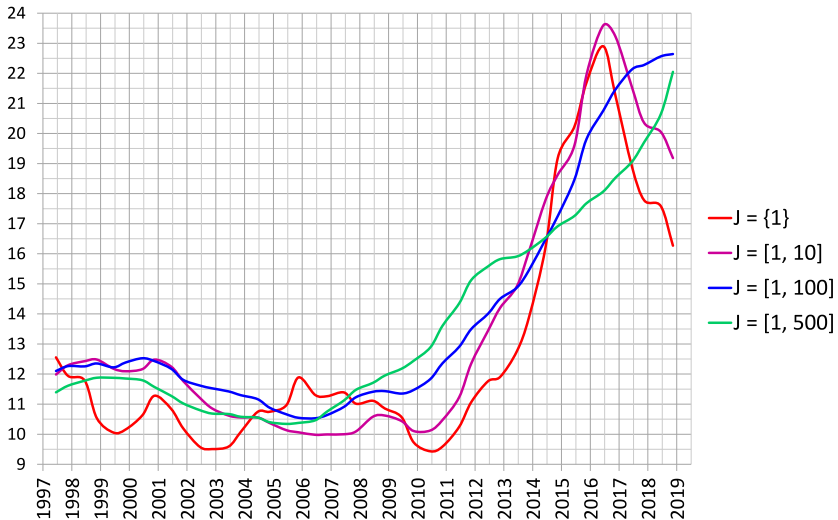


РИСУНОК 3. За сколько лет  $R_{\max}$  суперкомпьютеров, занимающих места  $J$ , вырастет в среднем в 1,000 раз?

года достаточно точно выдерживалось следующее: за 18 месяцев удвоение производительности самой мощной системы в мире; за 11 лет рост ее производительности в 1000 раз. Так, рубежи производительности были преодолены: 1 Mflops ( $10^6$  flops) в 1975 году, 1 Gflops ( $10^9$  flops) в 1986 году, 1 Tflops ( $10^{12}$  flops) в 1997 году, 1 Pflops ( $10^{15}$  flops) в 2008 году. Если бы эта тенденция сохранилась бы, то производительность 1 Eflops ( $10^{18}$  flops) была бы достигнута в 2019 году. Но это произошло в июне 2022 года, см. рисунок 2.

Начиная с 2008 года явно изменилась скорость роста достигнутой максимальной производительности: вместо прироста «в 1000 раз за 11 лет» мы имеем прирост «в 1000 раз за 13–17 лет».

Мировая суперкомпьютерная отрасль на рубеже 2008 года, несомненно, столкнулась с научно-техническими сложностями на пути развития СКТ. Это привело к пересмотру будущих прогнозов достижения новых рубежей производительности: 1 Eflops ( $10^{18}$  flops) был достигнут в июне 2022 года, а 1 Zflops ( $10^{21}$  flops) стоит ожидать в 2035–2039 году.



### 3. Самые мощные публичные супер-ЭВМ с июня 2011 года по ноябрь 2022 года

В таблице 1 собраны краткие сведения о самых мощных суперкомпьютерах в мире с июня 2011 года по ноябрь 2022 года.

ТАБЛИЦА 1. Краткие сведения о самых мощных суперкомпьютерах в мире с июня 2011 года по ноябрь 2022 года

Редакция Top500, система, ▷ вычислительная подсистема	Rmax, электропотребление	Интерконнект. Технология охлаждения
<b>6/2011</b> , K computer, Япония, Fujitsu, RIKEN AICS [8] ▷ ≈ <b>0.7М ядер</b> , 88,128 CPU SPARC64 V8fx 8C 2.00 GHz 8-ядер, 4 процессора в узле	10.5 Pflops, 12.6 МВт	Собственный интерконнект. Водяное охлаждение
<b>06/2012</b> , Sequoia, США, BlueGene/Q, IBM [9] ▷ ≈ <b>1.6М ядер</b> , ≈99К CPU Power BQC 16C, 64-разрядный RISC, 1.6 ГГц 16 ядер	16.3 Pflops, 7.9 МВт	Собственный интерконнект. Водяное охлаждение
<b>11/2012</b> , Titan, США, Cray XK7 [10] ▷ ≈ <b>0.56М ядер</b> , AMDOpteron – ≈300К ядер, NVIDIA K20x – ≈260К ядер	17.6 Pflops, 8.2 МВт	Собственный интерконнект. Водяное охлаждение
<b>06/2013</b> , Tianhe-2, КНР, UDT, Inspur [11] ▷ ≈ <b>3.1М ядер</b> , 32К Intel Ivy Bridge + 48К Intel Xeon Phi MIC	33.9 Pflops, 24 МВт	Собственный интерконнект. Водяное охлаждение
<b>11/2016</b> , Sunway TaihuLight, КНР, National Supercomputing Center in Wuxi, NRCPC, Inspur [12] ▷ ≈ <b>10.65М ядер</b> , CPU Sunway SW26010 260C 1.45 GHz, 40,960×(4+256 ядер)	93.0 Pflops, 15.4 МВт	Собственный интерконнект. Водяное охлаждение
<b>06/2019</b> , Summit, США, IBM/NVIDIA [13, 14] ▷ ≈ <b>2.4М ядер</b> , IBM POWER9 22C 3.07 GHz – ≈203К ядер, NVIDIA Volta GV100 – ≈2.2М ядер	148 Pflops, 9.8 МВт	Интерконнект: Dual-Rail EDR Infiniband. Водяное охлаждение
<b>6/2020</b> , Fugaku, Япония, Fujitsu, RIKEN [15, ?russian-] ▷ ≈ <b>7.63М ядер</b> , 158,976 CPUs Fujitsu ARM A64FX 48C 2.2GHz 48 ядер	442 Pflops, 29.9 МВт	Собственный интерконнект. Водяное охлаждение
<b>6/2022</b> , Frontier <sup>™</sup> (OLCF-5, HPE Cray EX235a), США, DoE, SC/ORNL, HPE [?russian-bib:13] ▷ ≈ <b>8.7М ядер</b> , 9,248 CPUs AMD Optimized 3rd Generation EPYC 64C 2GHz 64 ядра, 36,992 GPUs AMD Instinct MI250X 220 ядер	1,102 Pflops, 21.1 МВт	Собственный интерконнект Slingshot-11. Водяное охлаждение

Анализ таблицы 1 позволяет вскрыть очевидные тенденции и сделать следующие достоверные (на ближайшие годы) предположения об общем облике систем эксафлопсного уровня:

- *число узлов* – 10–50 тысяч, *число ядер* – 5–10 миллионов;
- *электропотребление* – 15–30 МВт;
- *интерконнект* – скорее всего (7/8 = 87% случаев) класса Custom (Proprietary Network, специфичные решения или недоступные как

отдельное коммерческое решение) либо топовый (труднодоступный) стандартный интерконнект, но с доработками производителя (1/8 = 13% случаев);

- *охлаждение* – или закрытое водяное, в будущем – более перспективное иммерсионное или охлаждение кипением;
- *вычислитель* – в зависимости от назначения – или однородный (50%) или гибридный (50% случаев);
- *процессоры и ускорители* – либо собственные (44%) или топовые (труднодоступные) стандартные (56% случаев).

#### 4. Тенденции развития гибридных архитектур

*Гибридными суперкомпьютерами* называют системы, в которых вычислительные узлы в дополнение к основным стандартным процессорам оснащены специализированными процессорами – ускорителями. Идея ускорителей в суперкомпьютерной отрасли использовалась давно, но широкое применение гибридные архитектуры нашли только в последнее десятилетие. Поля, детализирующие использование ускорителей в суперкомпьютерах (Accelerator и Accelerator Cores), появились в записях списка Top500 начиная с июня 2011 года. С этого времени мы и будем анализировать гибридные архитектуры, а каждый суперкомпьютер отнесем к следующим классам по производителям используемых в качестве вычислителей ускорителей:

- NONE: не используются ускорители – однородная (не гибридная) архитектура;
- IBM: спецвычислитель IBM PowerXCell 8i;
- AMD: спецвычислители от компании AMD;
- NVIDIA: спецвычислители от компании NVIDIA;
- Intel: спецвычислители от компании Intel – семейства Intel MIC / Intel Xeon Phi;
- MIX: спецвычислители как от компании NVIDIA, так и от компании Intel;
- PEZY: спецвычислители от компании PEZY Computing;
- Other: используются спецвычислители, не упомянутые выше. Сегодня в этом классе только ускорители, разработанные в Китае: Matrix-2000 и Deep Computing Processor.

Рассмотрим 24 редакции рейтинга Top500 с июня 2011 года по ноябрь 2022 года. На рисунке 4 показаны доли различных классов гибридных суперкомпьютеров в рейтинге Top500 для этих 24 редакций. Подписи

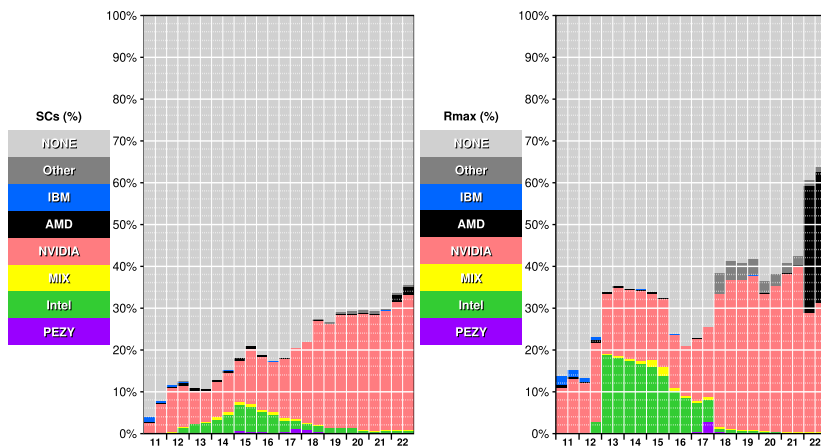


Рисунок 4. Доли различных классов гибридных суперкомпьютеров в рейтинге Top500, редакции с июня 2011 года по июнь 2022 года

по оси абсцисс – года, две столбчатые диаграммы в год соответствуют двум редакциям Top500 в соответствующем году. В левой части рисунка показаны доли в количестве систем (SCs) и за 100% принято 500 систем, а каждая система это 0.2%. В правой части рисунка показаны доли производительности (RMax) и за 100% принята суммарная производительность всего списка Top500 –  $\Sigma R_{max}$ . Можно заметить следующие тенденции:

- по количеству систем в последние годы (2014–2022) гибридные суперкомпьютеры выросли 12.8% ↗ 35.6%, по вкладу в суммарную производительность  $\Sigma R_{max}$  всего списка Top500 (а это и есть истинная доля в отрасли) их доля за этот период выросла 34.4% ↗ 63.8%;
- ускорители компании IBM практически исчезли к 2013 году;
- на пути исчезновения ускорители Intel: доля упала 18.8% ↘ 0.1%  $\Sigma R_{max}$  в период 2013–2022;
- доля NVIDIA выросла 16.4% ↗ 30.9%  $\Sigma R_{max}$  в период 2014–2022;

- ускорители компании AMD практически исчезли к 2017 году, но потом выросли 0.1 ↗ 31.4%  $\Sigma R_{\max}$ ;
- появились интересные новые решения, которые заслуживают внимания (5.4%  $\Sigma R_{\max}$  в лучший, 2018, год) – короткие и успешные проекты создания полностью собственных ускорителей, что должно послужить вдохновляющим примером для России:
  - PEZY – 2015, PEZY Computing, Япония;
  - Matrix-2000 – 2017, NUDT, Китай;
  - Deep Computing Processor – 2017, Sugon, Китай.

#### 4.1. Квинтили суперкомпьютеров в списке Top500

Очень важно понимать применимость и пригодность каждой технологии в супер-ЭВМ различной производительности. Например, важно понимать, в каких системах применима и востребована сегодня та или иная технология интерконнекта, тот или иной класс ускорителей и т. д. Для этого поступим следующим образом: разобьем весь список Top500 (а он отсортирован по производительности суперкомпьютеров) на пять групп: квинтиль А – несколько самых мощных суперкомпьютеров из списка; за нею квинтили В, С, D, Е. Размер квинтилей выберем так, чтобы сумма производительности всех суперкомпьютеров, входящих в один квинтиль, была бы близка (как только возможно) к 20% от суммарной производительности  $\Sigma R_{\max}$  всего списка Top500. То есть, квинтили А, В, С, D и Е имеют (примерно) равные производительности, но, конечно, содержат разное количество суперкомпьютеров. Для каждого выпуска рейтинга Top500 состав квинтилей А, В, С, D и Е (количество суперкомпьютеров в них) вычисляются заново, исходя из требования: суммарную производительность каждого квинтиля – как можно точнее приблизить к 20%  $\Sigma R_{\max}$ .

Для редакции Top500 в ноябре 2022 года деление на квинтили получилось таким:

- А: 1 система, Top1, 22,7%  $\Sigma R_{\max}$ ;
- В: 3 системы, Top2–4, 19,0%  $\Sigma R_{\max}$ ;
- С: 22 системы, Top5–26, 19,9%  $\Sigma R_{\max}$ ;
- D: 111 систем, Top27–137, 20,0%  $\Sigma R_{\max}$ ;
- Е: 363 системы, Top138–500, 18,4%  $\Sigma R_{\max}$ .

### 4.2. Распределение разных классов гибридных архитектур по квинтилям

Для каждого квинтиля – А, В, С, D, Е – посмотрим доли в нем классов None, Other, IBM, AMD, NVIDIA, MIX, Intel, PEZY (рисунок 5).

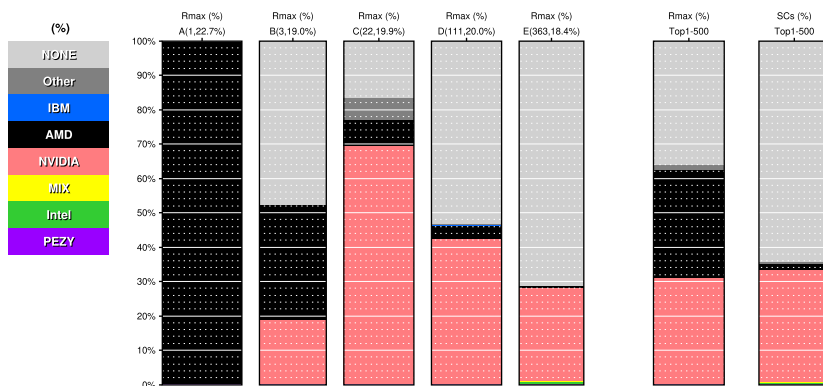


Рисунок 5. Доли различных классов гибридных суперкомпьютеров в суммарной производительности квинтилей А, В, С, D и Е рейтинга Top500 редакции за ноябрь 2022 года

Обращая внимание на форму и центр тяжести у фигур с соответствующим цветом заливки, мы можем сделать следующие суждения для тенденций ноября 2022 года:

- ускорители от AMD используются преимущественно в самых мощных системах – весь квинтиль А и третья часть В;
- ускорители от NVIDIA сосредоточены в большей степени в квинтелях С и D;
- негибридные системы (класс NONE) занимают существенную часть квинтилей В, D, Е;
- класс Other (китайские ускорители) представлены в квинтиле С (6%);
- ускорители классов Intel и MIX совсем слабо представлены в квинтиле Е – покидают рейтинг и отрасль;
- ускорители класса IBM и PEZY покинули отрасль.

## 5. Тенденции в технологиях интерконнекта

Интерконнект связывает все вычислительные узлы суперкомпьютера в единую систему. Сетевые технологии, используемые в интерконнекте, чаще всего отличаются от сетевых технологий, используемых в локальных (региональных и глобальных) компьютерных сетях, в силу специфических требований, предъявляемых к интерконнекту: не только высокая пропускная способность, но и максимально низкая задержка и максимально высокий темп выдачи сообщений, и еще ряд дополнительных, специфических требований.

В суперкомпьютерной отрасли это определяет серьезную роль технологий интерконнекта. В перечне [16] 10 критических технологических вызовов, стоящих перед отраслью на пути к эксафлопсным системам, разработка перспективных технологий интерконнекта занимает весьма высокое (второе) место.

Поля, детализирующие использование различных технологий интерконнекта в суперкомпьютерах, присутствуют во всех редакциях списка Top500. Опираясь на данные в этих полях, каждый суперкомпьютер во всех редакциях списка Top500 отнесем к различным классам в зависимости от семейства решений, используемых для реализации интерконнекта:

- Infiniband, если используются решения семейства Infiniband. Такие суперкомпьютеры встречаются в Top500 с июня 2003 года до сегодня, а технология реализована в нескольких версиях, с различными техническими характеристиками: Infiniband SDR (8 Гбит/с), Infiniband DDR (16 Гбит/с), Infiniband QDR (32 Гбит/с), Infiniband FDR (54 Гбит/с), Infiniband EDR (100 Гбит/с), Infiniband HDR (200 Гбит/с).
- Ethernet, если используются решения семейства Ethernet. Такие суперкомпьютеры встречаются в Top500 с июня 1996 года до сегодня, а технология реализована в нескольких версиях, с различными техническими характеристиками – от FastEthernet (100 мбит/с) до 100G Ethernet (100 Гбит/с).
- Myrinet, если используются решения семейства Myrinet. Такие суперкомпьютеры встречаются в Top500 с ноября 1998 года до ноября 2020 года.
- Quadrics, если используются решения семейства Quadrics. Такие суперкомпьютеры встречаются в Top500 с июня 1999 года до ноября 2011 года.

- SCI, если используются решения семейства SCI. Такие суперкомпьютеры встречаются в Top500 с июня 2002 года до ноября 2004 года.
- OpTsIntel, если используются решения от компании Intel семейств Omni-Path / TrueScale. Такие суперкомпьютеры встречаются в Top500 с ноября 2013 года до сего дня.
- Custom, если используются решения специфичные (Proprietary Network) или недоступные как отдельное коммерческое решение – доступное только в составе законченной системы.

Подчеркнем отличие между классами:

- Infiniband, Ethernet, Myrinet, Quadrics, SCI – эти технологии интерконнекта доступны как отдельные коммерческие продукты. Любой разработчик своего собственного суперкомпьютера (например, со своей архитектурой вычислительных узлов, с теми или иными процессорами) может купить и применить любое из этих решений.
- Custom – эти технологии интерконнекта не доступны как отдельные коммерческие продукты. Либо недоступны вообще, либо доступны только в составе законченной системы от единственного поставщика.
- OpTsIntel – технологии Intel семейств Omni-Path / TrueScale занимают промежуточную позицию. С одной стороны, эти технологии интерконнекта доступны как коммерческие продукты. С другой стороны, эти решения невозможно применить в суперкомпьютере, если вычислительный узел не имеет соответствующую архитектуру и процессор от компании Intel. Эта ситуация очень близка к ситуации «интерконнект доступен только в составе законченной системы от единственного поставщика» – близка по жесткости технологической зависимости от единственного разработчика и поставщика, зависимости от компании Intel.

Технология компании Intel семейств Omni-Path / TrueScale имела в момент появления весьма серьезные технические характеристики, и это ей позволило успешно войти в отрасль и отвоевать в короткие сроки долю в 11.0%  $\Sigma R_{max}$  в ноябре 2019 года. Слабой стороной решения OpTsIntel является жесткая технологическая зависимость всей системы от компании Intel. И сегодня эта технология покидает отрасль.

Рисунок 6 показывает доли производительности суперкомпьютеров с различными интерконнектами в суммарной производительности  $\Sigma R_{max}$  рейтинга Top500 для 60 редакций с июня 1993 года по ноябрь 2022

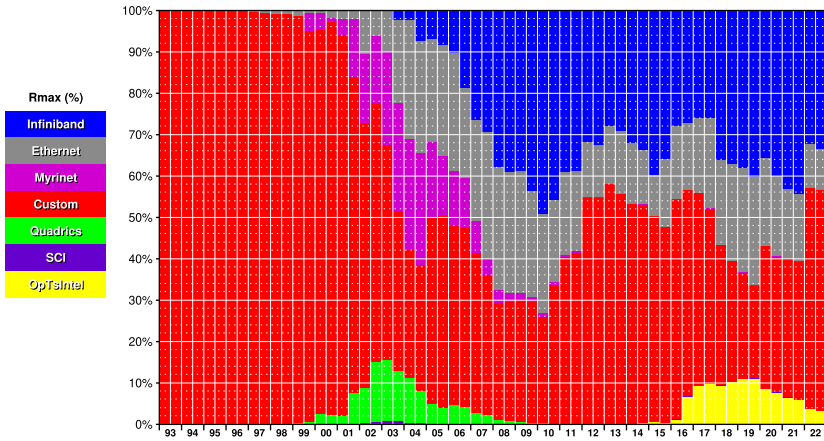


Рисунок 6. Доли суперкомпьютеров с различными интерконнектами в суммарной производительности рейтинга Top500, с июня 1993 по ноябрь 2022 года

года. Подписи по оси абсцисс – года, две столбчатые диаграммы в год соответствуют двум редакциям этого года.

Анализируя рисунок, можно отметить следующие тенденции последних 10 лет:

- технологии, соответствующие классам Myrinet, Quadrics, SCI в последние годы в отрасли не используются. На рисунке хорошо видны годы появления этих решений, годы «рассвета» этих технологий, пик популярности и годы «заката». Сегодня актуально обсуждать только 4 класса: Infiniband, Ethernet, OpTsiIntel и Custom.
- с 2012 года самая большая доля суммарной производительности Top500 у суперкомпьютеров с интерконнектом класса Custom. Динамика, 2012–2017–2022:
  - Custom 54.7% ↘ 42.4% ↗ 53.4%.
- следующей по величине (по значимости) является доля Infiniband. Динамика, 2012–2017–2022:
  - Infiniband 32.5% ↘ 26.1% ↗ 33.6%.
- затем следует Ethernet. Динамика, 2012–2017–2022:
  - Ethernet 13.2% ↗ 21.9% ↘ 10.0%
- технологии OpTsiIntel покидают Top500 и отрасль. Динамика, 2012–2017–2022:



– OpTsIntel 0.0% ↗ 9.6% ↘ 3.2%.

### 5.1. Распределение разных классов интерконнекта по квинтилям

Очень важно понимать области применимости каждого класса интерконнектов. Для этого рассмотрим распределение долей разных классов интерконнекта по квинтилям. Определение квинтилей дано выше в разделе 4.1.

Для каждого квинтиля – А, В, С, D, Е – посмотрим доли классов Infiniband, Ethernet, Myrinet, Custom – рисунок 7.

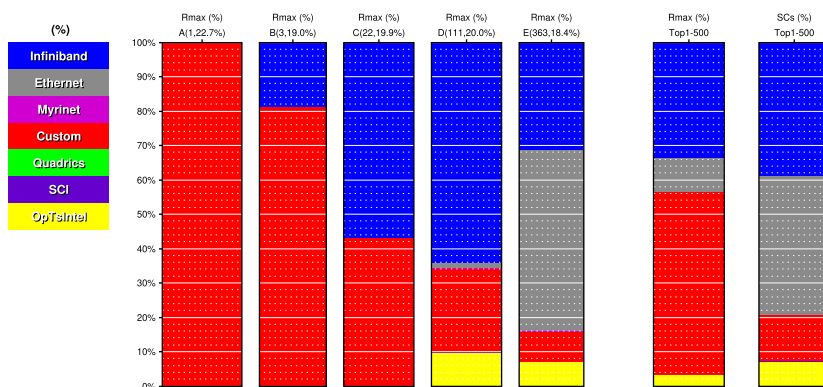


Рисунок 7. Доли суперкомпьютеров с различными интерконнектами в суммарной производительности квинтилей А, В, С, D и Е рейтинга Top500 редакции за ноябрь 2022 года

Видно, что:

- мощнейшие и мощные суперкомпьютеры (квинтили А и В) преимущественно используют интерконнект Custom. Красный цвет заполняет треугольник, который расширяется в направлении самых мощных систем;
- технологию Infiniband преимущественно используют для систем среднего и нижних уровней – самые большие доли в квинтилях С и D. Синий треугольник имеет центр тяжести в этой зоне;
- технологию Ethernet чаще используют в самых слабых системах (D, E). Серый треугольник существенно расширяется в сторону самых слабых систем;
- решения OpTsIntel покидают рейтинг и отрасль.

Несомненным выводом из рисунка 7 является следующее: разработка российских суперкомпьютеров высшей производительности, скорее всего, не сможет обойтись без интерконнекта класса Custom. А так как его невозможно купить, то, с неизбежностью, его придется разработать самим. Это уже не раз в России весьма успешно делалось: ИПС имени А. К. Айламазяна РАН (российские интерконнекты SKIF 3D-torus и Паутина), НИЦЭВТ (Ангара), ВНИИЭФ (СМПО), НИИ «Квант» и ИПМ им. М. В. Келдыша РАН (МВС-Экспресс).

Отметим, что единственный отечественный интерконнект, который был не только разработан, но и был использован в системе, вошедшей в рейтинг Top500 – это SKIF 3D-torus (ИПС имени А. К. Айламазяна РАН).

## 6. Тенденции в технологиях центральных процессоров

Поля, детализирующие использование различных технологий центральных процессоров в суперкомпьютерах, присутствуют во всех редакциях списка Top500. Опираясь на данные в этих полях, каждый суперкомпьютер во всех редакциях списка Top500 отнесем к следующим классам, в зависимости от компании-производителя или от используемых в CPU технологий: Alpha, AMD, ARM, CRAY, Fujitsu, Hitachi, HP, IBM, Intel, MIPS, NEC, SPARC. Если процессор не может быть отнесен к одному из данных классов, то он классифицируется как Proprietary – такое имеет место быть, например, по отношению к современным собственным процессорам Китая.

Рисунок 8 показывает доли производительности суперкомпьютеров с различными центральными процессорами в суммарной производительности SRmax рейтинга Top500 для 60 редакций с июня 1993 года по ноябрь 2022 года. Подписи по оси абсцисс – года, две столбчатые диаграммы в год соответствуют двум редакциям этого года.

Анализируя рисунок 8, можно отметить следующие тенденции:

- с 2020 года в отрасли, по сути, используются процессоры только из пяти классов: IBM, ARM (система Fugaku, Fujitsu), Proprietary (Китай), AMD и Intel;
- решения IBM за период июнь 2020–ноябрь 2022 сильно потеряли свои позиции:
  - IBM 28.7% ↘ 6.4%.
- система Fugaku (компании Fujitsu) показала, что однородные системы на ARM-процессорах могут иметь рекордные мировые показатели и быть эффективными для широкого спектра задач, включая реализацию искусственных нейронных сетей. Отметим, что наиболее близкими разработками в России являются процессоры семейства Байкал. Динамика доли за период июнь 2020–ноябрь 2022:

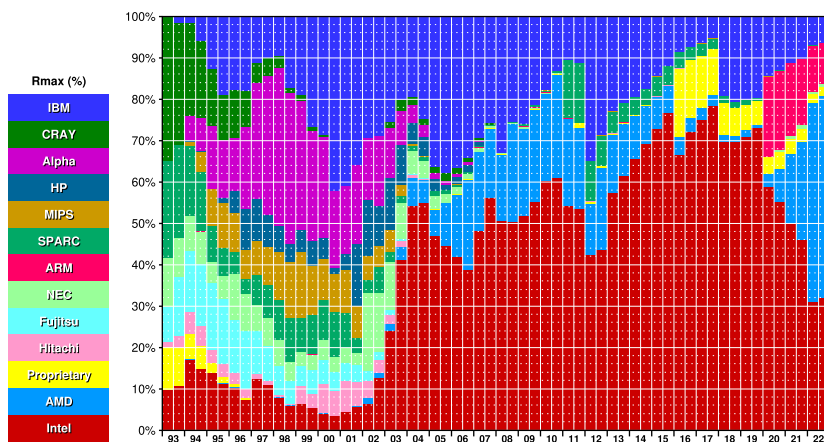


Рисунок 8. Доли суперкомпьютеров с различными центральными процессорами в суммарной производительности рейтинга Top500, редакции с июня 1993 года по ноябрь 2022 года

– ARM 19.2% ↘ 10.0%

- китайские процессоры с 2016 г. совпадают с классом Proprietary. Отметим: по всей видимости, Китай взял курс на сокрытие достаточно большого числа своих крупных установок. В публичной сфере динамика доли китайских процессоров в период июнь 2016–ноябрь 2022:

– Proprietary 16.5% ↘ 2.0%

- ожидаемо самые крупные доли у двух классов центральных процессоров: AMD и Intel. И за период ноябрь 2019–ноябрь 2022 AMD радикально расширил свою долю, отняв ее, по сути, у Intel. Данное событие, как представляется, связано с давней стратегией развития AMD как fabless-компании (в контрасте с Intel) и в нескольких удачных архитектурных разработках AMD. Динамика за ноябрь 2019–ноябрь 2022:

– AMD 0.8% ↗ 48.8%

– Intel 72.9% ↘ 32.0%

### 6.1. Распределение разных классов центральных процессоров по квинтилям

Рассмотрим распределение долей разных классов центральных процессоров по квинтилям. Определение квинтелей дано в разделе 4.1.

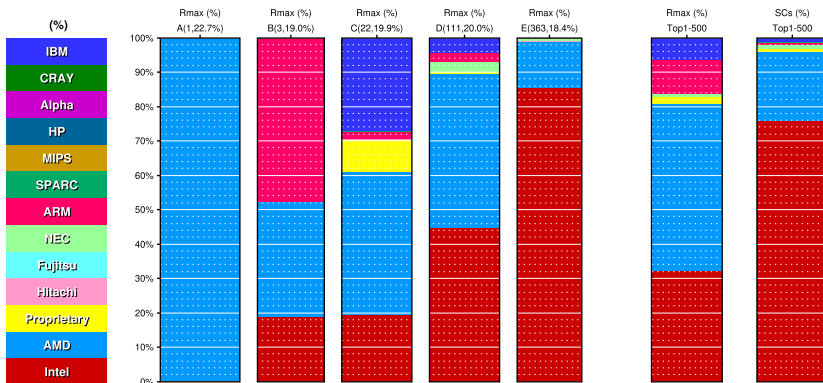


Рисунок 9. Доли суперкомпьютеров с различными центральными процессорами в суммарной производительности квинтилей А, В, С, D и E рейтинга Top500 редакции за ноябрь 2022 года

Рисунок 9 показывает доли центральных процессоров для каждого квинтиля – А, В, С, D, E. Видно, что:

- классы центральных процессоров AMD и ARM по большей части представлены в мощнейших и мощных суперкомпьютерах (квинтили А и В), класс IBM представлен в квинтилях С и D с явным тяготением к старшим системам;
- класс Intel представлен в квинтилях (В, С, D, E), но чаще всего используется в слабых системах – бордовый треугольник существенно расширяется в сторону самых слабых систем;
- остальные классы (Proprietary, NEC) не интересны для серьезного анализа.

Вывод: для построения систем высшей производительности перспективными центральными процессорами являются либо топовые (значит ограничено доступные) модели самых успешных коммерческих разработок, – сегодня это процессоры AMD, – либо собственные процессоры, на базе одной из доступных архитектур, таких как ARM (доказал свою перспективность для использования в суперкомпьютерах) и, возможно, RISC-V.

## 7. Компании-производители суперкомпьютеров

Поле «Manufacturer» присутствует во всех доступных редакциях списка Top500. Опираясь на данные в этом поле, каждый суперкомпьютер

во всех редакциях списка Top500 отнесем к одному из следующих классов, в зависимости от компании-производителя суперкомпьютера: Cray, IBM, Lenovo, Inspur, NRCPC, NUDT, Sugon, Fujitsu, NVIDIA, Atos, HP, Other. Классы выбраны и отсортированы так, чтобы:

- выявлялись все основные участники отрасли, все второстепенные отнесены к классу Other;
- максимально учитывались изменения последних годов в отрасли – например, агрессивное вторжение в отрасль китайских компаний-производителей (Lenovo, Inspur, NRCPC, NUDT, Sugon), на диаграммах эти классы расположены рядом и для них используются цвета «бледной» палитры;
- при переименовании компании на протяжении 1993–2022 гг., класс сохраняет свое имя. Например, Hewlett-Packard до недавнего времени обозначал себя в данных Top500 как HP или Hewlett-Packard, а теперь HPE (Hewlett-Packard Enterprise) – за классом сохранили имя HP.

Рисунок 10 показывает доли производительности суперкомпьютеров от различных производителей в суммарной производительности  $\Sigma R_{max}$  рейтинга Top500 для 60 редакций с июня 1993 года по ноябрь 2022 года. Подписи по оси абсцисс – года, две столбчатые диаграммы в год соответствуют двум редакциям этого года.

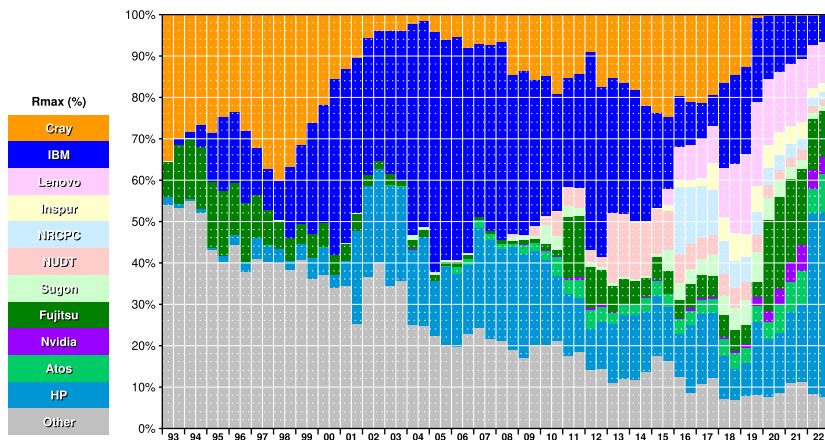


Рисунок 10. Доли суперкомпьютеров различных производителей в суммарной производительности рейтинга Top500, с июня 1993 года по ноябрь 2022 года

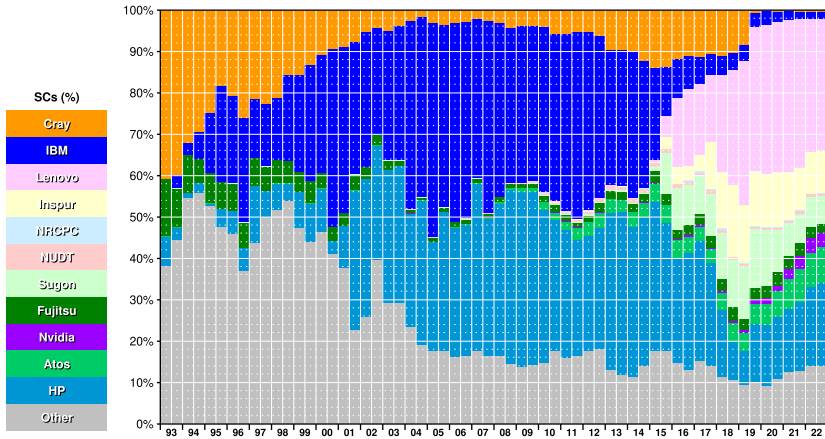


Рисунок 11. Доли количества суперкомпьютеров от различных производителей в общем количестве (500) рейтинга Top500, с июня 1993 года по ноябрь 2022 года

Рисунок 11 показывает доли систем (одна система – это 0.2% от 500) суперкомпьютеров от различных производителей в Top500 для 60 редакций с июня 1993 года по ноябрь 2022 года.

Тем самым, рисунок 10 иллюстрирует реальные объемы продаж, а рисунок 11 – количество продаж. Разница между ними очевидна: много мелких сделок в сумме не дадут большой объем.

Анализируя эти рисунки, можно отметить следующие самые яркие события за последние годы:

- агрессивное вторжение китайских производителей в отрасли с огромным количеством сделок по всему миру и достаточно большим объемом этих сделок. При этом необходимо помнить, что Китай в последние годы закрывает сведения о своих крупных системах. Показатели за июнь 2019–ноябрь 2022:
  - доля SCs 62.2% ↘ 49.6%;
  - доля  $\Sigma R_{max}$  41.3% ↘ 16.6%;
- Hewlett-Packard купил Cray. Как результат, Hewlett-Packard приобрел собственный интерконнект с рекордными характеристиками и в классе HP появились самые крупные системы (чего раньше никогда не было). Доли числа сделок изменились не существенно, а объемы сделок – радикально. Показатели за июнь 2019–ноябрь 2022:
  - доля SCs 8.2% ↗ 20.2%;

- доля  $\Sigma R_{max}$  7.9% ↗ 44.5%;
- Компания NVIDIA купила Mellanox. Как результат, NVIDIA приобрела собственный набор технологий и компетенций для того, чтобы выступать в роли компании-поставщика законченных систем, а не только GPU для них. Показатели за июнь 2019–ноябрь 2022:
  - доля SCs 0.8% ↗ 3.4%;
  - доля  $\Sigma R_{max}$  1.1% ↗ 4.1%.

### 7.1. Распределение компаний-производителей суперкомпьютеров по квинтилям

Рассмотрим распределение долей разных компаний-производителей по квинтилям. Определение квинтилей дано в разделе 4.1. Рисунок 12 показывает доли компаний-производителей для каждого квинтиля – А, В, С, D, Е.

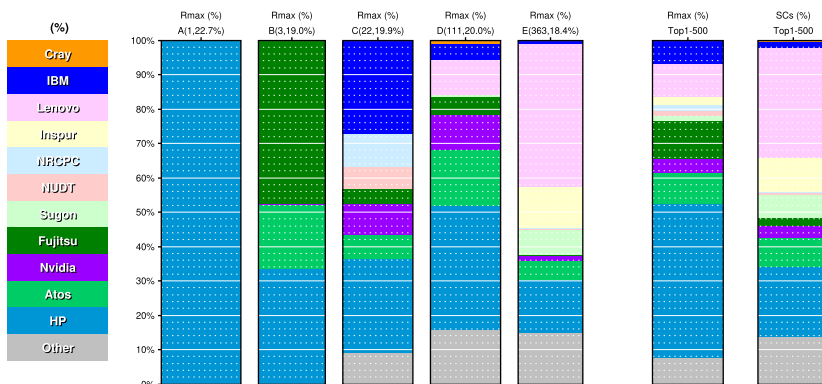


Рисунок 12. Доли суперкомпьютеров различных компаний-производителей в суммарной производительности квинтилей А, В, С, D и Е рейтинга Top500 редакции за ноябрь 2022 года

Видно, что:

- Hewlett-Packard занимает значительные доли не только в квинтилях С, D, Е (что было давно, это традиционная позиция компании), но и в самых высоких сегментах – А, В – что стало возможным после приобретения компании Gray;
- IBM и Fujitsu хорошо представлены в квинтилях среднего сегмента (В, С, D). Их цвета существенно расширяются в сторону крупных систем;

- китайские компании (Lenovo, Inspur, NRCPC, NUDT, Sugon) представлены в начальном сегменте (С, D, E), их цвета существенно расширяются в сторону слабейших систем.

## 7.2. Основные выводы из анализа компаний-производителей суперкомпьютеров

Среди компаний-производителей суперкомпьютеров произошли существенные изменения в последние годы. Китайские производители осуществили серьезную интервенцию в отрасли: на их долю в ноябре 2019 года приходилось 63.0% количества сделок и 43.4% объемов (доли  $\Sigma R_{\max}$ ). Больше всего доля Китая среди систем начального уровня. Для построения систем высшей производительности важен факт консолидации крупнейших участников отрасли:

- Hewlett-Packard купил Cray. Как результат, Hewlett-Packard приобрел собственный интерконнект с рекордными характеристиками, и в классе HP появились самые крупные системы (чего раньше никогда не было);
- компания NVIDIA купила Mellanox. Как результат, NVIDIA приобрела собственный набор технологий и компетенций для того, чтобы выступать в роли компании-поставщика законченных систем, а не только GPU для них.

Доли числа сделок этих двух компаний изменилась значительно (в 2.6 раза), а объемы сделок – вообще радикально (в 5.4 раза), так как эта консолидация повлияла именно на самые крупные проекты, динамика июнь 2019–ноябрь 2022:

- доля SCs 9.0%; ↗ 23.6%;
- доля  $\Sigma R_{\max}$  9.0%; ↗ 48.6%.

## 8. Области применения суперкомпьютеров

Поля («Application Area» и «Segment»), позволяющие определить область применения суперкомпьютеров, присутствуют во всех редакциях списка Top500. Опираясь на данные в этих полях, каждый суперкомпьютер во всех редакциях списка Top500 отнесем к одному из следующих классов:

- RnD – суперкомпьютер используется для исследований и разработок; для генерации новых знаний, технологий, материалов; для научных расчетов в НИИ и университетах;
- Industry – суперкомпьютер используется напрямую для задач реальной экономики: инженерные расчёты, финансовые приложения, услуги связи и интернет-сервисы и т. п.;



- Gov.Mil – суперкомпьютер используется в государственных интересах или для военных нужд;
- Unknown – недостаточно информации для отнесения к одному из классов выше.

В качестве примера использования суперкомпьютерной отрасли в государственных (военных) целях, приведем программу США «Crossroads» (Лос-Аламосская Национальная Лаборатория, создана в интересах Программы управления запасами (SSP) и Национального управления по ядерной безопасности (NNSA), министерство энергетики США) [17]. Ее цель: изучение и использование новых технологий для повышения эффективности приложений и рабочих процессов.

Crossroads, посредством симуляций на супер-ЭВМ, создаст способность прогнозирования ключевых физических явлений, количественную оценку и агрегирование неопределенностей в инструментах моделирования, при этом, обеспечивая точность физики, численную достоверность и геометрическую точность.

Рисунок 13 показывает доли производительности суперкомпьютеров с разными областями их применения в суммарной производительности рейтинга  $\Sigma R_{max}$  рейтинг Top500 для 60 редакций с июня 1993 года по ноябрь 2022 года.

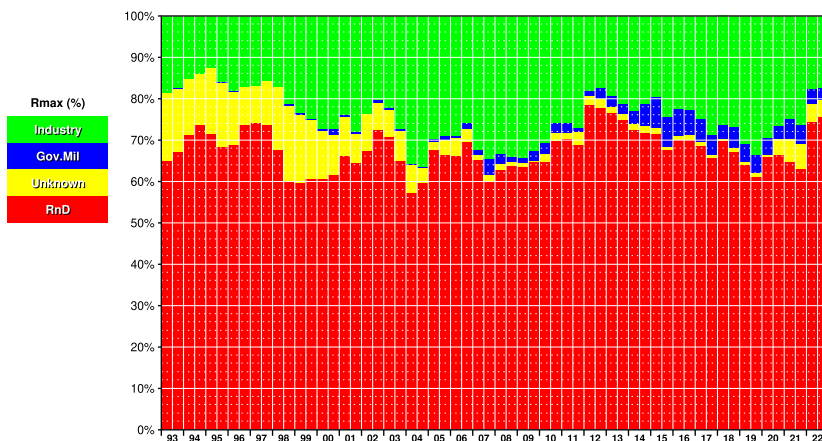


Рисунок 13. Доли суперкомпьютеров с разными областями их применения в суммарной производительности рейтинга Top500, редакции с июня 1993 года по ноябрь 2022 года

Подписи по оси абсцисс – года, две столбчатые диаграммы в год соответствуют двум редакциям списка Top500 этого года.

Рисунок 14 показывает доли систем (доли в SCs – одна система – это 0.2% от 500) суперкомпьютеров с разными областями их применения в рейтинге Top500 для 60 редакций с июня 1993 года по ноябрь 2022 года.

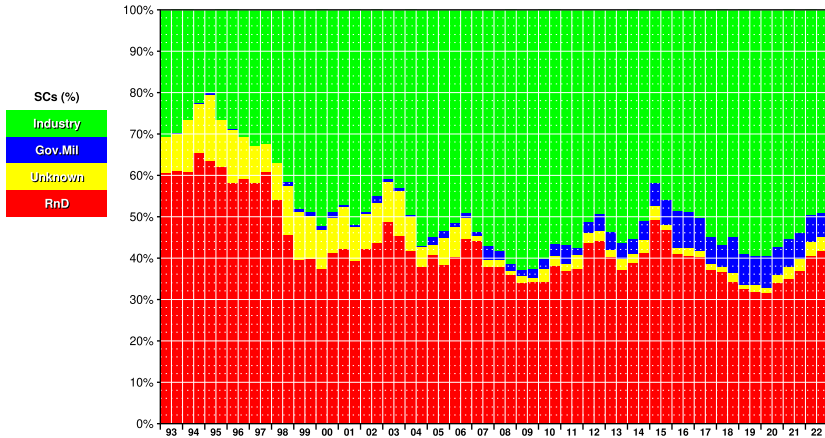


РИСУНОК 14. Доли количества суперкомпьютеров с разными областями их применения в общем количестве (500) рейтинга Top500, редакции с июня 1993 года по ноябрь 2022 года

Анализируя рисунки, можно отметить следующее:

- классы Gov.Mil и Unkown имеют небольшие доли и в SCs, и в  $\Sigma R_{\max}$ . Основные классы применения суперкомпьютеров это Industry и RnD;
- во все времена подавляющая доля производительности (доля  $\Sigma R_{\max}$ ) суперкомпьютеров использовалась для генерации новых знаний (класс RnD), а не для непосредственных прикладных задач реальной экономики (класс Industry). За последние пять лет доля  $\Sigma R_{\max}$  Industry 26.0% в среднем, доля  $\Sigma R_{\max}$  RnD 67.1% в среднем;
- во все времена подавляющая доля количества (доля SCs) суперкомпьютеров использовалась для непосредственных прикладных задач реальной экономики (класс Industry), а не для генерации новых знаний (класс RnD). За последние пять лет доля SCs Industry 55.5% в среднем, доля SCs RnD 35.4% в среднем;
- таким образом, очевидно, что в классе Industry – как правило, слабые системы, а в классе RnD – крупные системы. Более наглядно это можно увидеть при анализе квинтилей.

## 8.1. Распределение областей применения суперкомпьютеров по квинтилям

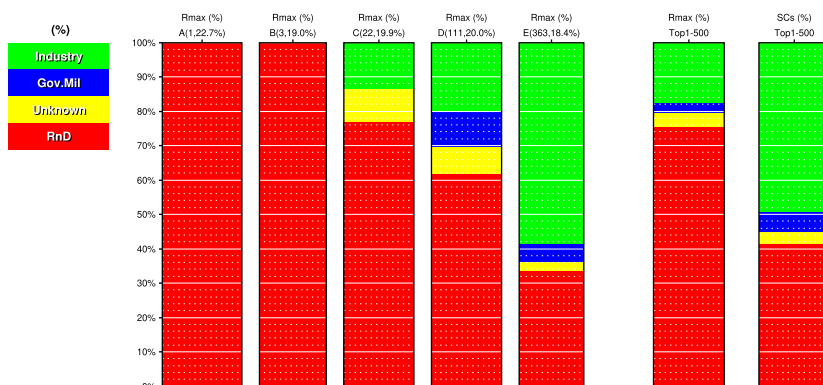


Рисунок 15. Доли областей применения суперкомпьютеров в суммарной производительности квинтилей А, В, С, D и E рейтинга Top500 редакции за ноябрь 2022 года

Рисунок 15 показывает распределение долей разных областей применения суперкомпьютеров по квинтилям А, В, С, D, E в ноябре 2022 года. Видно, что:

- самые крупные системы (квинтили А, В) используются исключительно для генерации новых знаний – класс RnD. Цвет этого класса представлен во всех квинтилях, он явно и существенно расширяется к самым крупным системам;
- цвет класса Industry представлен в квинтилях С, D, E и он явно и существенно расширяется к самым слабым системам. Самые слабые системы (квинтиль E) преимущественно используются для прикладных задач реальной экономики;
- суперкомпьютеры классов Gov.Mil и Other представлены не широко и только в квинтилях С, D, E.

## 8.2. Основные выводы из анализа областей применения суперкомпьютеров

Для построения систем высшей производительности важен следующий факт – такие системы:

- всегда используются для исследований и разработок; для генерации новых знаний, технологий, материалов; для научных расчётов;
- никогда не используются напрямую для задач реальной экономики: инженерных расчетов, финансовых приложений, услуг связи, интернет-сервисов и т. п.

Это накладывает свой отпечаток на класс решаемых задач: по всей видимости, он достаточно широкий и порой непредсказуемый – на самом переднем крае научного поиска за время жизни рекордной установки (около 5–7 лет) могут возникнуть самые неожиданные задачи, требующие для решения самые неожиданные математические и алгоритмические инструменты. Это следует учитывать при разработке архитектурных решений в части и аппаратных, и программных средств таких систем.

## **9. Состояние суперкомпьютерной инфраструктуры в мире и в России**

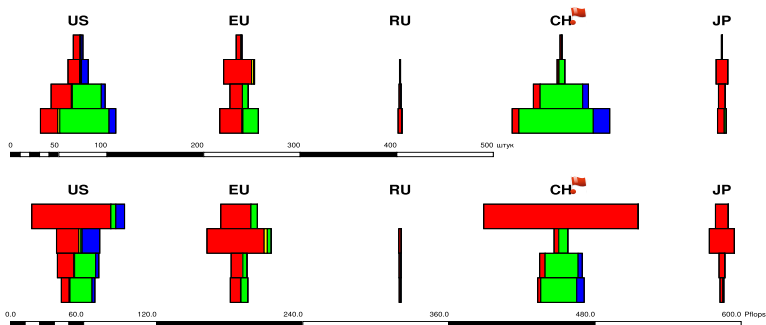
### **9.1. Суперкомпьютерная инфраструктура стран-лидеров и России**

Анализ уровня развития и освоения СКТ в России и в других странах – США, Китай, Япония, Евросоюз – начнем с обзора их суперкомпьютерной инфраструктуры – СК-инфраструктуры.

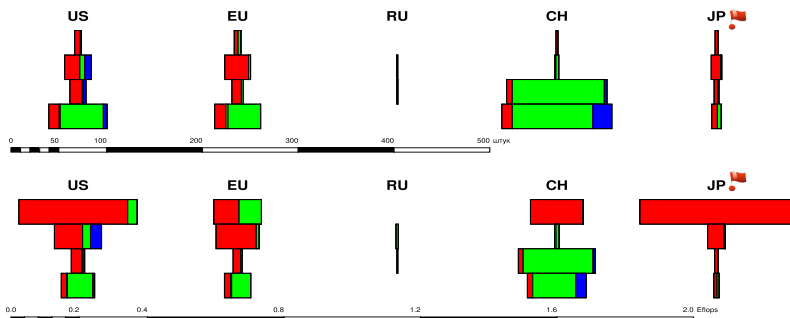
Здесь важно отметить, сколько суперкомпьютеров (мы можем анализировать только публичные суперкомпьютеры) того или иного уровня имеется в стране, какая их суммарная производительность, в каких областях они используются. Будем выделять четыре уровня СК-инфраструктуры:

- суперкомпьютеры уровня Top1–20 (то есть, занимающие первые 20 мест в рейтинге). На практике, эти системы устанавливаются в крупнейших национальных суперкомпьютерных центрах (СКЦ);
- суперкомпьютеры уровня Top21–100 – крупнейшие региональные и отраслевые СКЦ;
- суперкомпьютеры уровня Top101–250, – крупные региональные и корпоративные СКЦ;
- суперкомпьютеры уровня Top251–500, – СКЦ предприятий и научных учреждений.

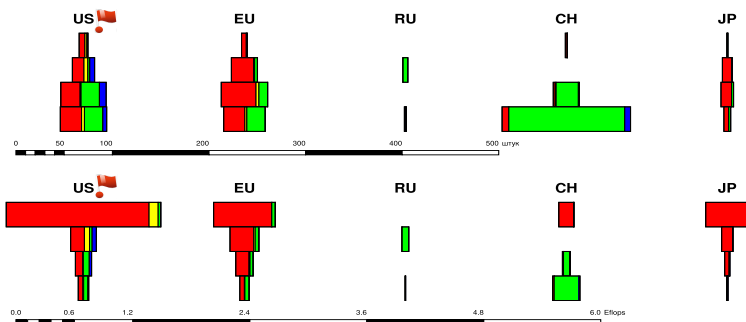
Рисунки 16 показывают характеристики СК-инфраструктур США, Евросоюза, России, Китая и Японии в разные моменты времени.



(а) в июне 2016 года



(б) в июне 2020 года



(в) в ноябре 2022 года

Рисунок 16. Суперкомпьютерная киберинфраструктура США (US), стран Евросоюза (EU), России (RU), Китая (CH) и Японии (JP)

Поясним рисунки 16.

- в верхней части каждого рисунка для каждой страны размер каждого слоя «пирамиды» указывает количество суперкомпьютеров каждого из четырёх уровней, в нижней части показана суммарная производительность суперкомпьютеров этих уровней. Красный флажок отмечает страну расположения системы Top 1;
- цветами выделены доли декларированных областей применения СКЦ: красный цвет – исследования и разработки; зеленый цвет – промышленные применения в реальной экономике; синий цвет – государственные нужды и оборона; желтый цвет – нет информации.

Очевидно следующее:

- все страны (кроме России) имеют сбалансированную СК-инфраструктуру – в заметном количестве присутствуют СКЦ всех 4 уровней;
- Россия, по сути, не обладает никакой СК-инфраструктурой. По крайней мере, такой, которая отвечала бы ее планам по развитию цифровой экономики.

## 9.2. Суперкомпьютеры, установленные в Российской Федерации

На ноябрь 2022 года в списке Top500 имелись 7 машин, установленных в России. Одна из них – суперкомпьютер известной НРС-компании Т-Платформы<sup>1</sup>:

- 290 место – Lomonosov 2 (T-Platform A-Class Cluster, Xeon E5-2697v3 14C 2.6GHz, Intel Xeon Gold 6126, Infiniband FDR, Nvidia K40m/P-100, T-Platforms, Moscow State University – Research Computing Center), 2.48 PFlops – 0.22% от Top1, слабее в 455 раз, чем Top1.

Эта машина входит в список Top500 с 2014 года и *доступна для использования*<sup>URI</sup> в исследовательских проектах в области медицины, инженерных расчетов, астрофизики, материаловедения, энергетики и других областей.

*Три суперкомпьютера компании Яндекс*<sup>URI</sup>:

- 25 место – Chervonenkis (YANDEX, AMD EPYC 7702 64C 2GHz, NVIDIA A100 80GB, Infiniband), 21.53 PFlops – 1.95% от Top1, слабее в 51.3 раза, чем Top1;

---

<sup>1</sup>Новый суперкомпьютер в МГУ вошел в Top500<sup>URI</sup> – новость от 19.11.2014 издательства «Открытые системы»

- 44 место – Galushkin (YANDEX, AMD EPYC 7702 64C 2GHz, NVIDIA A100 80GB, Infiniband), 16 PFlops – 1.45% от Top1, слабее в 69 раз, чем Top1;
- 47 место – Lyapunov (Inspur NF5488A5, AMD EPYC 7662 64C 2GHz, NVIDIA A100 40GB, Infiniband, NVIDIA, Inspur), 12.8 PFlops – 1.16% от Top1, слабее в 86.2 раза, чем Top1.

Данные суперкомпьютеры используются для задач машинного обучения в дата-центрах Яндекса, которые запускаются через разработанную в Яндексе платформу YT для хранения и обработки больших данных. К задачам относятся: нейросетевые модели поиска, сервисы Алисы и другие сервисы компании. На фоне представленных машин в Top500, компания Яндекс продвигает свой сервис для ML-разработки – Yandex DataSphere в рамках инфраструктуры Yandex Cloud. Отметим, что производителями машин Chervonenkis и Galushkin являются компании YANDEX, NVIDIA (США), а машина Lyapunov произведена компаниями NVIDIA, Inspur (Китай).

Две машины компании Сбербанк:

- 50 место – Christofari Neo (NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 80GB, Infiniband, SberCloud), 11.95 PFlops – 1.08% от Top1, слабее в 92.6 раза, чем Top1;
- 86 место – Christofari (NVIDIA DGX-2, Xeon Platinum 8168 24C 2.7GHz, Mellanox InfiniBand EDR, NVIDIA Tesla V100, SberCloud), 6.6 PFlops – 0.61% от Top1, слабее в 164 раза, чем Top1.

Сбербанк использует эти машины как для внутренних задач компании (распознавание речи, генерация голоса и др.), так и для предоставления облачного сервиса ML Space, обеспечивающего работу с ML: решаются задачи государственных и коммерческих заказчиков, а также научно-исследовательских групп. В настоящий момент, на *сайте компании*<sup>URL</sup> можно подать заявку на тестовый доступ к облачной платформе в рамках SberCloud. Производителем машин Christofari Neo и Christofari является компания NVIDIA (США).

Компания МТС представила *одну машину в списке*<sup>URL</sup>:

- 352 место – MTS GROM (NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100 40GB, Infiniband, Nvidia, CloudMTS) 2.26 PFlops – 0.2% от Top1, слабее в 500 раз, чем Top1.

Суперкомпьютер планируется использовать в рамках облачной инфраструктуры компании CloudMTS. Ожидается, что на кластере будут считаться как собственные задачи компании, так и задачи для крупного бизнеса, что позволит кратно сократить время на разработку и внедрение проектов на основе искусственного интеллекта и больших данных. Производителем данной машины является компания NVIDIA (США).

Таким образом, государственная инфраструктура в Российской Федерации слабо развита, большая часть существующих машин из Top500, описанных выше, работают в собственных интересах корпораций. Почти все машины, установленные в России и входящие в Top500 (за исключением Lomonosov 2) – либо совместного производства с зарубежными компаниями (США, Китай), либо полностью зарубежные.

### 9.3. Китай начал скрывать свои суперкомпьютерные разработки

Отдельные слова стоит сказать про Китай. Есть серьезные основания для того, чтобы утверждать, что Китай начал скрывать сведения про свои самые мощные системы. И, по всей видимости, в Китае доля непубличных суперкомпьютеров (и непубличная доля производительности) стали резко превосходить то, что мы имеем для других развитых стран.

Основанием для того, чтобы сделать такой вывод, послужили следующие факты:

- еще летом 2021 года появились сообщения [18] о вероятном существовании в Китае двух суперкомпьютеров с производительностью более 1 EFlops. Официальных подтверждений или опровержений не последовало;
- США внес<sup>2</sup> в «черный список» ряд суперкомпьютерных организаций Китая. Конечно, это стимулирует Китай скрывать свои успехи, чтобы не давать поводов для расширения санкционного списка;
- если сравнить динамику развития СК-инфраструктур ведущих стран за несколько лет на рисунках 16а, 16б и 16в, будет видно, что у Китая была серьезная СК-инфраструктура в 2017–2020 годах – пожалуй, даже более развитая, чем у США. И это все это было достигнуто Китаем всего за 8 лет – в 2009 году практически никакой

---

<sup>2</sup>по сообщению Карасёва С. на [ServerNews<sup>®</sup>](#) от 09.04.2021 «США внесли в чёрный список ряд суперкомпьютерных центров Китая»



СК-инфраструктуры Китая не имел. За восемь лет он догнал и перегнал США. И вдруг, после 2020 года, СК-инфраструктура Китая стала «таять».

### 9.4. Интегральные показатели степени освоения СКТ странами-лидерами и Россией

Важный интегральный показатель степени освоения СКТ страной является суммарная производительность всех суперкомпьютеров страны. Рисунок 17 иллюстрирует этот показатель и наглядно отображает достижения Китая в последние годы и недопустимое отставание России.

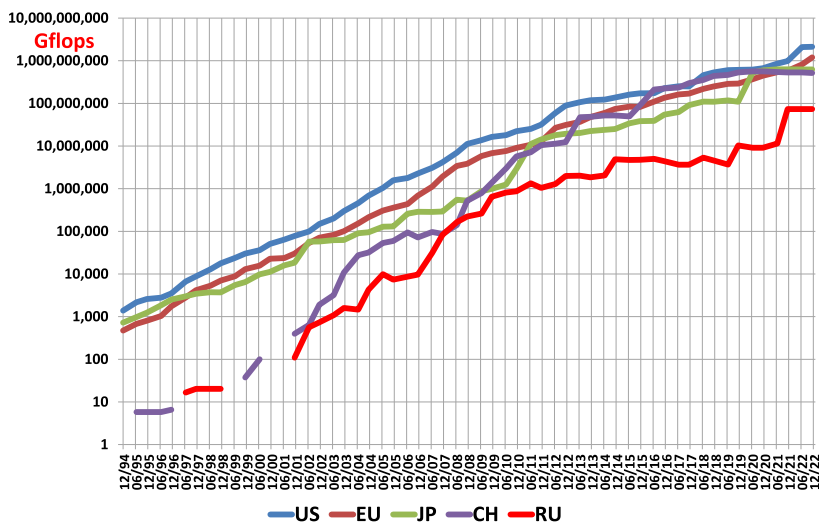


Рисунок 17. Суммарная производительность всех суперкомпьютеров США, стран Евросоюза, России, Китая и Японии за период от июня 1995 года по ноябрь 2022 года

Оценить отставание России (в годах) в области СКТ от других стран (рассмотрим на примере США) можно следующим образом: берем любой момент времени, смотрим суммарную производительность суперкомпьютеров России и дальше смотрим, сколько лет назад такое же (или чуть меньше) значение было у США. Эти вычисления повторяем для каждого момента времени и для всех стран, получая рисунок 18.

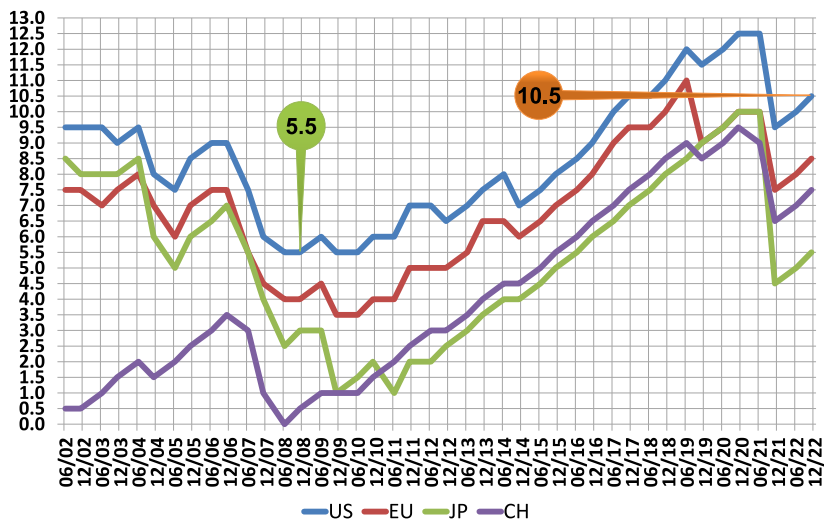


Рисунок 18. Отставание России (в годах) по суммарной производительности суперкомпьютеров от США (US), стран Евросоюза (EU), Китая (CH) и Японии (JP) за период июнь 2002 года по ноябрь 2022 года

Если в 2006–2008 Россия резко сокращала свое отставание от США, Евросоюза, Китая и Японии, то начиная с 2010 года мы практически неуклонно ежегодно наращиваем свое отставание от этих стран. И в июне 2021 года это отставание достигло максимума за весь период наблюдения.

Термин «отставание» требует пояснения. Например, сейчас отставание от США – 10.5 лет. Можем ли мы «догнать Америку» за 10.5 лет? Ответ, формально, да можем, например, при реализации одного из следующих сценариев:

- США в своем развитии полностью останавливается, а Россия 10.5 лет подряд развивается с теми же темпами, что до этого были у США. Через 10.5 лет мы сравняемся;
- США продолжает свое развитие, но Россия 10.5 лет подряд развивается с темпами, вдвое большими, чем они были у США за предыдущие годы. Через 10.5 лет мы сравняемся.

Оба сценария маловероятны. Поэтому, наверное, более правильно использовать термин не «отставание», а «задержка в развитии» – недостаток,

который очень тяжело преодолеть.

Аналогичным образом можно объективно и достоверно рассчитать отставание России (в годах) от уровней технологий. рисунок 19 показывает отставание России от уровня технологий Top1, Top5 и Top10.

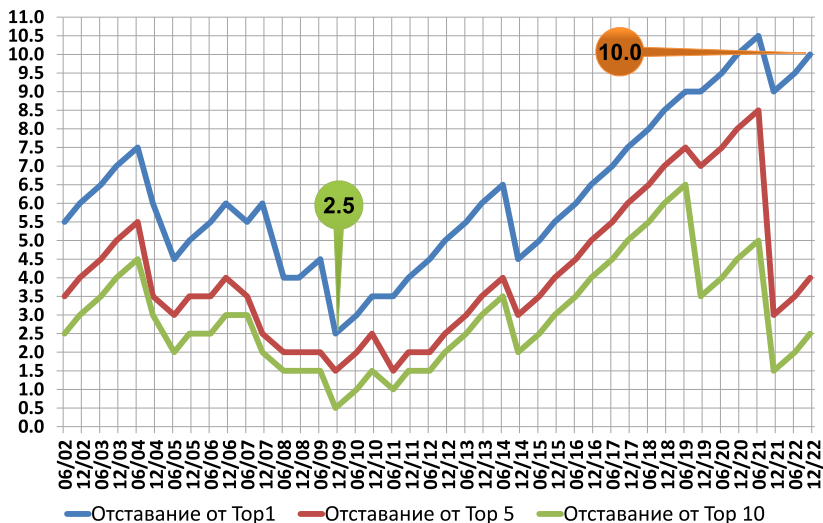


РИСУНОК 19. Отставание (в годах) России от уровней технологий классов Top1, Top5 и Top10 за период июнь 2002 года по ноябрь 2022 года

Рассчитывалось это так: в каждый момент времени рассматриваем самый мощный суперкомпьютер в России и оцениваем, сколько лет назад этот суперкомпьютер занял бы первое (соответственно пятое и десятое) место в рейтинге. Эта величина (в годах) и будет отставанием России от технологий класса Top1 (Top5, Top10).

Видны периоды, когда Россия сокращала технологическое отставание. Было даже время, когда от технологий уровня Top10 нас отделяли всего полгода. Однако в последние годы технологическое отставание неуклонно росло и в июне 2021 года оно достигло максимума за весь период наблюдения.

Наконец, роль страны в мировой суперкомпьютерной отрасли, степень устремленности к цифровой экономике хорошо характеризует доля страны в совокупной мировой вычислительной производительности. То есть,

отношение суммарной производительности  $\Sigma R_{\max}$  суперкомпьютеров, установленных в данной стране, к суммарной производительности всего списка Top500  $\Sigma R_{\max}$ . По своей роли и своему смыслу этот параметр чем-то аналогичен доле экономики страны в мировом ВВП.

Рисунок 20 показывает долю России в совокупной мировой вычислительной производительности с июня 1999 года по ноябрь 2022 года.

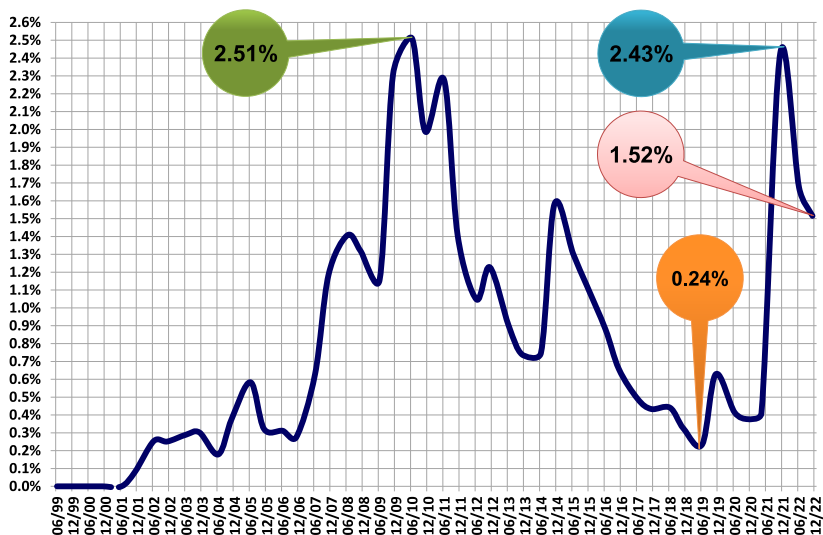


РИСУНОК 20. Доля России в совокупной мировой вычислительной производительности суперкомпьютеров в период с июня 1999 года по ноябрь 2022 года

В лучшие времена (2009 год) эта доля достигала 2.51%, а в последние годы неуклонно сокращалась до 0.24% (июнь 2019 года), но потом за счет закупок шести систем компаниями Яндекс, Сбер и МТС этот показатель немного вырос и сейчас составляет 1.52%.

## 9.5. Индекс цифровизации стран-лидеров и России

В Таблице 2 вводится ещё один важный показатель. В таблице приведены доли различных стран (на ноябрь 2022 года) в совокупной мировой производительности суперкомпьютеров  $\Sigma R_{\max}$  и их доли

в мировой экономике – доли в мировом ВВП, данные о ВВП различных стран взяты из [19], а также отношение этих долей, обозначенное как  $K_{Ц}$  – индекс цифровизации страны. Можно обосновать (в данной работе опущено), что индекс цифровизации страны оценивает величину «средний объём вычислений, затрачиваемый на выпуск единицы продукции в данной стране».

Таблица 2. Индекс цифровизации стран-лидеров и России.  
Ноябрь 2022

<b>Колонки таблицы:</b>				
(A) Доли стран в суммарной производительности супер-ЭВМ мира $\Sigma R_{max}$ ;				
(B) Доли стран в мировом ВВП;				
(C) Индекс цифровизации, $K_{Ц}$ – отношение (A/B);				
(D) Во сколько раз индекс цифровизации данной страны выше индекса цифровизации России.				
Страна	Доля страны в $\Sigma R_{max}$	Доля страны в сумме ВВП	Индекс цифровизации, $K_{Ц}$ страны	Отношение $K_{Ц}$ страны к $K_{Ц}$ России
	(A)	(B)	(C = A/B)	(D)
Китай	12.04%	18.45%	–	–
США	47.35%	23.93%	1.98	2.19
Япония	14.23%	5.14%	2.77	3.06
Евросоюз	18.54%	17.78%	1.04	1.15
Россия	1.67%	1.85%	0.91	–
Мир	100.00%	100.00%	1.00	1.10
Четыре лидера	92.17%	65.30%	1.41	1.56

Как легко заметить, у России этот показатель сильно уступает развитым странам. Индекс цифровизации у России хуже в 2.19 раза – чем у США; в 1.15 раз – чем у Евросоюза; в 3.06 раз – чем у Японии; в 1.56 раз – чем у четырех стран-лидеров в совокупности; в 1.1 раза – чем у мира в среднем.

## Заключение

Просуммируем выводы всех разделов данной статьи.

Мировая суперкомпьютерная отрасль на рубеже 2008 года, несомненно, столкнулась с научно-техническими сложностями на пути развития СКТ. Начиная с 2008 года изменилась скорость роста достигнутой максимальной производительности: вместо прироста «в 1000 раз за 11 лет» мы имеем прирост «в 1000 раз за 13–17 лет». Это привело к пересмотру будущих прогнозов достижения новых рубежей производительности: 1 Eflops ( $10^{18}$  flops) был достигнут в июне 2022 года, а 1 Zflops ( $10^{21}$  flops) стоит ожидать в 2035–2039 году.

Анализ самых мощных суперкомпьютеров последнего десятилетия позволяет сделать следующие общие достоверные (на ближайшие годы) предположения об общем облике систем эксафлопсного уровня:

- число узлов – 10–50 тысяч, число ядер – 5–10 миллионов;
- электропотребление – 15–30 МВт;
- интерконнект – скорее всего ( $7/8 = 87\%$  случаев) класса Custom (Proprietary Network, специфичные решения или недоступные как отдельное коммерческое решение), либо топовый (труднодоступный) стандартный интерконнект, но с доработками производителя ( $1/8 = 13\%$  случаев);
- охлаждение – или закрытое водяное, или более перспективное – иммерсионное, охлаждение кипением и др.;
- вычислитель – в зависимости от назначения – или однородный (50%), или гибридный (50% случаев);
- процессоры и ускорители – либо собственные (44%), либо топовые (труднодоступные) стандартные (56% случаев).

Анализ гибридных архитектур в редакциях Top500 позволяет установить, что вклад гибридных архитектур в общую производительность Top500 вырос и достиг 63.8%  $\Sigma R_{max}$ . Среди актуальных решений для гибридных систем можно отметить ускорители AMD (31.4%  $\Sigma R_{max}$ ) и NVIDIA (30.9%  $\Sigma R_{max}$ ). Заслуживают внимания интересные новые решения (5.4%  $\Sigma R_{max}$  в лучший, 2018, год) – короткие и успешные проекты создания полностью собственных ускорителей, что должно послужить вдохновляющим примером для России: PEZY (2015, PEZY Computing, Япония), DeepComputingProcessor (2017, Sugon, Китай), Matrix-2000 (2017, NUDT, Китай).

В мире продолжают интенсивно развиваться различные технологии интерконнекта. Они играют исключительно высокую роль для всего

комплекса СКТ. На сегодняшний день, из коммерчески доступных технологий используются решения класса Custom (53.4%  $\Sigma R_{max}$ ), Infiniband (33.6%  $\Sigma R_{max}$ ), Ethernet (10.0%  $\Sigma R_{max}$ ). С 2012 года самая большая доля суммарной производительности Top500 у суперкомпьютеров с интерконнектами, которые недоступны, как отдельные коммерческие решения (класс Custom). Для России это означает:

- разработка российских суперкомпьютеров высшей производительности, скорее всего, не сможет обойтись без интерконнекта Custom. А так как его невозможно купить, то, с неизбежностью, его придется разработать самим (задача выполняемая и уже не раз была успешно реализована);
- младшие и средние российские суперкомпьютеры могут использовать для интерконнекта технологии Infiniband и Ethernet.

Из разработок интерконнекта в России можно отметить:

- ИПС имени А.К. Айламазяна РАН – российские интерконнекты SKIF 3D-torus и «Паутина»;
- НИЦЭВТ – Ангара;
- ВНИИЭФ – СМПО;
- НИИ «Квант» и ИПМ им. М.В. Келдыша РАН – МВС-Экспресс.

Отметим, что единственный отечественный интерконнект, который был не только разработан, но и был использован в системе, вошедшей в рейтинг Top500 – это SKIF 3D-torus (ИПС имени А.К. Айламазяна РАН).

Для построения систем высшей производительности перспективными центральными процессорами являются либо топовые (значит, ограниченно доступные) модели самых успешных коммерческих разработок – сегодня это процессоры AMD – либо собственные процессоры, на базе одной из доступных архитектур, таких как ARM (доказал свою перспективность для такого использования в суперкомпьютерах) и, возможно, RISC-V.

Среди компаний-производителей компьютеров произошли существенные изменения в последние годы. Китайские производители осуществили серьезную интервенцию в отрасли: на их долю в ноябре 2019 года приходилось 63.0% количества сделок и 43.4% объемов (доли  $\Sigma R_{max}$ ).

Больше всего доля Китая среди систем начального уровня. Для построения систем высшей производительности важен следующий факт – такие системы:





- всегда используются для исследований и разработок; для генерации новых знаний, технологий, материалов; для научных расчётов;
- никогда не используются напрямую для задач реальной экономики: инженерных расчетов, финансовых приложений, услуг связи, интернет-сервисов и т. п.

Это накладывает свой отпечаток на класс решаемых задач: на самом переднем крае научного поиска за время жизни рекордной установки (около 5–7 лет) могут возникнуть особые задачи, требующие для решения особые математические и алгоритмические инструменты. Это следует учитывать при разработке архитектурных решений в части и аппаратных, и программных средств таких систем.

На сегодня США, Китай, Япония и Евросоюз выстроили добротную СК-инфраструктуру, как базис перехода к цифровой экономике, имеют весомые доли своих стран в совокупной мировой производительности суперкомпьютеров. И доли эти больше, чем доли этих стран в мировом ВВП (исключение – Китай, который скрывает свои успехи в СКТ). Такое соотношение хорошо иллюстрирует реальное продвижение стран на пути к цифровой экономике. Россия на сегодня не имеет никакой осмысленной СК-инфраструктуры, а ее доля в совокупной мировой производительности суперкомпьютеров меньше ее доли в мировом ВВП. В последние годы в области СКТ неуклонно растет отставание России от других ведущих стран и отставание России от мирового уровня развития СКТ.





Все эти обстоятельства следует самым серьезным образом учитывать при формировании государственной политики в области СКТ и цифровых технологий.

## Список литературы

- [1] Ezell S. J., Atkinson R. D. *The vital importance of high-performance computing to U.S. competitiveness.*– Washington, DC, USA: ITIF.– 28.04.2016.– 58 pp.  ↑50
- [2] Lucas R., Ang J., Bergman K., Borkar Sh., Carlson W., Carrington L., Chiu G., Colwell R., Dally W., Dongarra J., et al. *Top ten exascale research challenges*, DOE advanced scientific computing advisory subcommittee (ASCAC) report.– United States: USDOE Office of Science (SC).– 2014.– 86 pp.  ↑50
- [3] Абрамов С. М. *Правда, искажающая истину. Как следует анализировать Top500?* Вестник ЮУрГУ. Сер. Выч. матем. информ.– 2013.– Т. 2.– № 3.– С. 5–31.   ↑50



- [4] Абрамов С. М. *Июнь 2019: анализ развития суперкомпьютерной отрасли в России и в мире* // Программные системы: теория и приложения.– 2019.– Т. **10**.– № 3.– С. 3–40. [doi](#) [URL](#) ↑50
- [5] Kogge P. M., Dally W. J. *Frontier vs the Exascale Report: Why so long? and Are We Really There Yet? 2022 IEEE/ACM International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS)* (13–18 November 2022, Dallas, TX, USA).– IEEE.– 2023.– Pp. 26–35. [doi](#) ↑51
- [6] Zheng W. *Research trend of large-scale supercomputers and applications from the TOP500 and Gordon Bell Prize* // Science China Information Sciences.– 2020.– Vol. **63**.– id. 171001. [doi](#) ↑51
- [7] Khan A., Sim H., Vazhukudai S. S., Butt A. R., Kim Y. *An analysis of system balance and architectural trends based on Top500 supercomputers* // *The International Conference on High Performance Computing in Asia-Pacific Region, HPC Asia 2021*, New York, NY, USA: Association for Computing Machinery.– 2021.– ISBN 978-1-4503-8842-9.– Pp. 11–22. [doi](#) ↑51
- [8] *Supercomputer “K computer” takes first place in World, Achieves world’s best performance of 8.162 petaflops to lead TOP500 list.*– RIKEN, Fujitsu Ltd.– 2011 (revised as of December 17, 2018). [URL](#) ↑57
- [9] Brodtkin J. *With 16 petaflops and 1.6M cores, DOE supercomputer is worlds fastest. Giant cluster built to test nuclear weapons stockpile breaks HPC speed record.*– Ars Technica.– 20.06.2011. [URL](#) ↑57
- [10] Oak Ridge Leadership Computing Facility *ORNL debuts titan supercomputer.*– Oak Ridge National Laboratory.– 26.10.2012. [URL](#) ↑57
- [11] Alba D. *Chinas Tianhe-2 Caps Top 10 Supercomputers, China’s Tianhe-2 outflanks U.S. systems to take No. 1 spot in Top500 list* // IEEE Spectrum.– 17.06.2013. [URL](#) ↑57
- [12] Dongarra J. *Report on the Sunway TaihuLight*, Tech Report UT-EECS-16-742.– Oak Ridge National Laboratory.– 14.06.2016.– 24 pp. [URL](#) ↑57
- [13] Oak Ridge National Laboratory *Introducing the Summit Supercomputer.*– ORNL, IBM, NVIDIA.– 2020.– 24 pp. ↑57
- [14] Vergara L. V. G., Joubert W., Brim M. J., Budiardja R. D., Maxwell D., Ezell M., Zimmer C., Boehm S., Elwasif W., Oral S., Fuson C., Pelfrey D., Hernandez O., Leverman D., Hanley J., Berrill M., Tharrington A. *Scaling the Summit: deploying the World’s fastest supercomputer* // *High Performance Computing* (Frankfurt, Germany, June 16–20, 2019), Lecture Notes in Computer Science.– vol. **11887**, eds. Weiland M., Juckeland G., Alam S., Jagode H., Cham: Springer.– 2019.– ISBN 978-3-030-34355-2.– Pp. 330–351. [doi](#) ↑57
- [15] Dongarra J. *Report on the Fujitsu Fugaku system*, Tech Report ICLUT-20-06.– University of Tennessee-Knoxville Innovative Computing Laboratory.– 2020.– 18 pp. ↑57

- [16] Coghlan S. *Argonnes Aurora Exascale Computer*, Smoky Mountain Computational Sciences And Engineering Conference SMC 2019 (August 29, 2019).– 17 pp.  ↑62
- [17] Grider G. *Update on ASC Crossroads System. History and Plans, From LA-UR-19-24930.*– Los alamos national laboratory.– 2019.– 11 pp.  ↑73
- [18] Prickett N. H. *China has already reached exascale — on two separate systems.*– The Next Platform.– October 26, 2021.  ↑80
- [19] *The global economy by GDP.*– BusinessLIVE.– March 13, 2021.  ↑85

Поступила в редакцию 18.04.2023;  
 одобрена после рецензирования 04.05.2023;  
 принята к публикации 20.05.2023;  
 опубликована онлайн 07.07.2023.


Рекомендовал к публикации

*Анд. В. Климов*

## Информация об авторах:

### Николай Сергеевич Абрамов


к.т.н., вед. научн. сотр. Института программных систем имени А. К. Айламазяна РАН. Научные интересы: компьютерная графика, распознавание образов, обработка изображений, нейронные сети, искусственный интеллект, принятие решений, управление видеокамерой, геометрия.

 0000-0002-1612-3879

**e-mail:** [n-say@nsa.pereslavl.ru](mailto:n-say@nsa.pereslavl.ru)

### Сергей Михайлович Абрамов

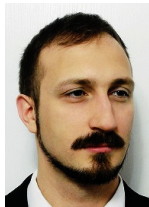
д.ф.-м.н., чл.-корр. РАН, гл. научн. сотр. Института программных систем имени А. К. Айламазяна РАН. Научные интересы: супер-компьютерные технологии, метавычисления, сетевые технологии, интернет вещей.

 0000-0001-6603-7971


**e-mail:** [abram@botik.ru](mailto:abram@botik.ru)

Вклад авторов: *Н. С. Абрамов* – 50% (идея, методология, сбор материала, доработка и редактирование); *С. М. Абрамов* – 50% (идея, методология, анализ списка Top500, иллюстрации и таблицы, визуализация).

*Авторы заявляют об отсутствии конфликта интересов.*



UDC 004.382.2+519.257

 10.25209/2079-3316-2023-14-2-49-93

## June 2022 Top500 List Overview

Nikolai Sergeevich **Abramov**<sup>1</sup>, Sergei Mikhailovich **Abramov**<sup>2</sup>

Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

<sup>1</sup>[n-say@nsa.pereslavl.ru](mailto:n-say@nsa.pereslavl.ru)














**Abstract.** The article is devoted to the analysis of the current state of the supercomputer industry in the world and in Russia and the prospects for its development. In terms of methodological approach and tools, this article is a continuation of a series of similar analytical reviews by the authors. The novelty of this work is not only in updating the information, taking into account the latest editions of the Top500 list, but also in focusing on the following: the global supercomputer industry is undergoing a radical restructuring— the transition from the «petascale era» to the «exascale era». Russia will have to make the same transition with some delay. Therefore, it is important to understand technological trends and features of solutions for the most powerful systems in the world in recent years. This circumstance is given special attention in the work. And such an emphasis on trend analysis is one of the motivational goals of this work. (*In Russian*).





**Key words and phrases:** Top500, supercomputers, hybrid architectures, interconnect

2020 *Mathematics Subject Classification:* 65Y05; 68M20, 62Q05

For citation: Nikolai S. Abramov, Sergei M. Abramov. *June 2022 Top500 List Overview*. Program Systems: Theory and Applications, 2023, **14**:2(57), pp. 49–93. (*In Russ.*). [https://psta.psir.ru/read/psta2023\\_2\\_49-93.pdf](https://psta.psir.ru/read/psta2023_2_49-93.pdf)

## References

- [1] S. J. Ezell, R. D. Atkinson. *The vital importance of high-performance computing to U.S. competitiveness*, ITIF, Washington, DC, USA, 28.04.2016, 58 pp. 
- [2] R. Lucas, J. Ang, K. Bergman, Sh. Borkar, W. Carlson, L. Carrington, G. Chiu, R. Colwell, W. Dally, J. Dongarra, al. et. *Top ten exascale research challenges*, DOE advanced scientific computing advisory subcommittee (ASCAC) report, USDOE Office of Science (SC), United States, 2014, 86 pp. 
- [3] S. M. Abramov. True judgments that distort the real truth. How to analyze the Top500? *Vestnik YuUrGU. Ser. Vych. matem. inform.*, **2:3** (2013), pp. 5–31 (in Russian).  
- [4] S. M. Abramov. “June 2019: Analysis of the Development of the Supercomputer Industry in Russia and Worldwide”, *Program Systems: Theory and Applications*, **10:3** (2019), pp. 3–40 (in Russian). 
- [5] P. M. Kogge, W. J. Dally. Frontier vs the Exascale Report: Why so long? and Are We Really There Yet? *2022 IEEE/ACM International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS)* (13–18 November 2022, Dallas, TX, USA), IEEE, 2023, pp. 26–35. 
- [6] W. Zheng. “Research trend of large-scale supercomputers and applications from the TOP500 and Gordon Bell Prize”, *Science China Information Sciences*, **63** (2020), id. 171001. 
- [7] A. Khan, H. Sim, S. S. Vazhkudai, A. R. Butt, Y. Kim. “An analysis of system balance and architectural trends based on Top500 supercomputers”, *The International Conference on High Performance Computing in Asia-Pacific Region, HPC Asia 2021*, Association for Computing Machinery, New York, NY, USA, 2021, ISBN 978-1-4503-8842-9, pp. 11–22. 
- [8] *Supercomputer “K computer” takes first place in World, Achieves world’s best performance of 8.162 petaflops to lead TOP500 list*, RIKEN, Fujitsu Ltd., 2011 (revised as of December 17, 2018). 
- [9] J. Brodtkin. *With 16 petaflops and 1.6M cores, DOE supercomputer is worlds fastest. Giant cluster built to test nuclear weapons stockpile breaks HPC speed record*, *Ars Technica*, 20.06.2011. 
- [10] Oak Ridge Leadership Computing Facility. *ORNL debuts titan supercomputer*, Oak Ridge National Laboratory, 26.10.2012. 
- [11] D. Alba. “Chinas Tianhe-2 Caps Top 10 Supercomputers”, China’s Tianhe-2 outflanks U.S. systems to take No. 1 spot in Top500 list, *IEEE Spectrum*, 17.06.2013. 
- [12] J. Dongarra. *Report on the Sunway TaihuLight*, Tech Report UT-EECS-16-742, Oak Ridge National Laboratory, 14.06.2016, 24 pp. 

- [13] Ridge National Laboratory Oak. *Introducing the Summit Supercomputer*, ORNL, IBM, NVIDIA, 2020, 24 pp.
- [14] L. V. G. Vergara, W. Joubert, M. J. Brim, R. D. Budiardja, D. Maxwell, M. Ezell, C. Zimmer, S. Boehm, W. Elwasif, S. Oral, C. Fuson, D. Pelfrey, O. Hernandez, D. Leverman, J. Hanley, M. Berrill, A. Tharrington. “Scaling the Summit: deploying the World’s fastest supercomputer”, *High Performance Computing* (Frankfurt, Germany, June 16–20, 2019), Lecture Notes in Computer Science, vol. **11887**, eds. Weiland M., Juckeland G., Alam S., Jagode H., Springer, Cham, 2019, ISBN 978-3-030-34355-2, pp. 330–351. 
- [15] J. Dongarra. *Report on the Fujitsu Fugaku system*, Tech Report ICLUT-20-06, University of Tennessee-Knoxville Innovative Computing Laboratory, 2020, 18 pp.
- [16] S. Coghlan. *Argonnes Aurora Exascale Computer*, Smoky Mountain Computational Sciences And Engineering Conference SMC 2019 (August 29, 2019), 17 pp. 
- [17] G. Grider. *Update on ASC Crossroads System. History and Plans*, From LA-UR-19-24930, Los alamos national laboratory, 2019, 11 pp. 
- [18] N. H. Prickett. *China has already reached exascale — on two separate systems*, The Next Platform, October 26, 2021. 
- [19] *The global economy by GDP*, BusinessLIVE, March 13, 2021. 