DOI: 10.14529/jsfi230201

November 2022 Top500 List Overview

Nikolay S. Abramov[1], *Sergei M. Abramov*[1]

The article is devoted to the analysis of the current state of the supercomputer industry and the prospects for its development. In terms of methodological approach and tools, the work is a continuation of a series of similar analytical reviews by the authors. The main source of information for analysis is the archive of editions (releases) of the world ranking of the five hundred most productive supercomputers in the world. The novelty of this work lies not only in updating the information, taking into account the latest editions of the Top500 list, but also in focusing on the following circumstance: the global supercomputer industry is undergoing a radical restructuring – transition from the "petascale era" to the "exascale era". Technological trends and features of solutions for the most productive systems in the world in recent years are given. The pace of development of supercomputer technologies, development trends are discussed: hybrid architectures, interconnect technologies and changes in the positions of supercomputer manufacturing companies. Based on the results of the analysis, reliable forecasts were made for the coming years about the general appearance of exascale systems.

Keywords: Top500, HPC, supercomputers, hybrid architectures, interconnect.

Introduction

In leading countries, supercomputer technologies (SCT) have been considered by government and society as the only means of ensuring competitive advantages for quite some time [1]. Today, in the era of the digital economy, the roles of the supercomputer industry and the supercomputer cyberinfrastructure (SC infrastructure) in leading countries are becoming increasingly important. It is currently important to have an accurate assessment of the current state of the SC infrastructure in these countries, a reliable comparison of their positions in the supercomputer industry, and a clear understanding of modern trends in SCT development.

This article is a continuation of a series of similar analytical reviews by the authors, which use the same methodological approach and toolkit [2, 3]. The novelty of the material presented in this work lies not only in the updating of information based on the latest revisions to the Top500 list, but also in the focus on the following circumstance.

The global supercomputer industry is currently undergoing a radical restructuring, transitioning from the "petaflops era" to the "exascale era". This is not simply about increasing the performance of supercomputers by 1,000 times. In the process of this transition, solutions had to be found to a large number of problems in both hardware and software [4].

Understanding of recent technological trends will make it possible to identify established solutions that have enabled the transition to exascale systems and apply this knowledge in current and future supercomputer solutions.

The article is organized as follows. Section 1 is devoted to Top500 list. In Section 2 we look at pace of supercomputing technology development. Section 3 contains most powerful public supercomputers from June 2011 to November 2022. Trends in the development of hybrid architectures are considered in Section 4. Section 5 is devoted to trends in interconnect technologies. Conclusion summarizes the study.

[1] Ailamazyan Program Systems Institute of Russian Academy of Sciences, Pereslavl-Zalessky, Russian Federation

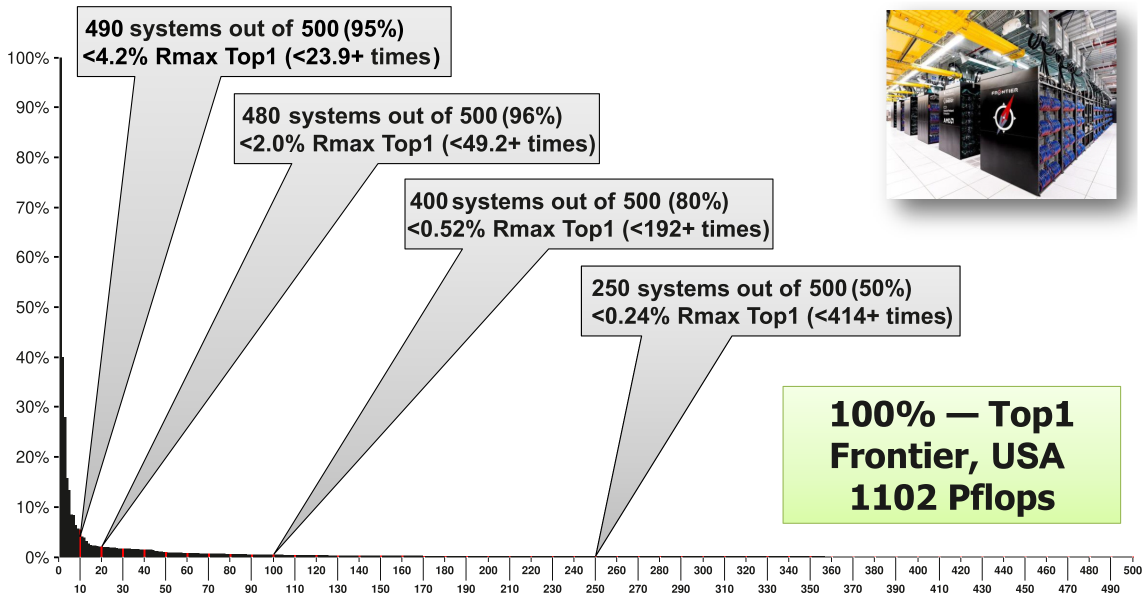

Figure 1. Relative performance of the 500 systems that entered the Top500 ranking in November 2022. The performance of Top1 system is taken as 100%

1. Top500 List

The main source of information for the analysis in this article is the worldwide ranking of the top 500 most powerful supercomputers in the world, known as the Top500 list [5]. The list is updated twice a year, in June and November. Thus, from June 1993 (the first edition) to November 2022, there have been 60 revisions of the list. In the Top500 list, systems are ranked by real performance measured on the Linpack benchmark test – Linpack performance. From now on, we will understand performance only as Linpack – performance, even if it is not specifically emphasized. In the Top500 list, it is denoted as *Rmax*.

Given the fact that some systems are not included in the list for various reasons, the term *supercomputer* here will be defined as follows: *a supercomputer is a computing system that has the performance corresponding to the performance of the machines listed in the corresponding revision of the Top500 list.*

With this definition, the data available on the Top500 website is reliable and multi-dimensional and provides a retrospective description of the state of the supercomputer industry. Moreover, it is detailed, covering all 60 revisions of the list at intervals of six months. Each edition is a table of 500 records. The key aspects of the methodology used to analyze this data set are as follows:

- Supercomputers listed in the Top500 list have a great variety of performance and, as a result, differ significantly in the technical solutions used, price, consumer properties, and so on. The extremely strong "layering" of supercomputers on the example of the November 2022 edition of the Top500 list are illustrated in Fig. 1.
- As a result, it is incorrect to consider supercomputers as "units". It is essential to rely on their key characteristic – their performance. For example, when comparing the equipment of different countries with supercomputer technology, it is necessary to pay more attention to the total performance of the available supercomputers in the country rather than their number.

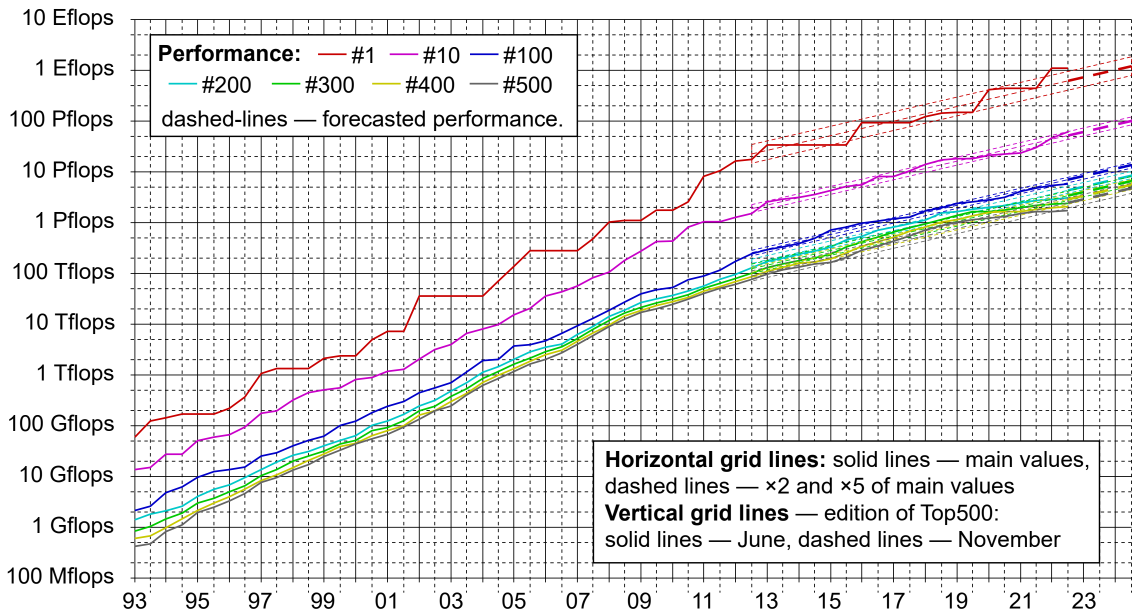

Figure 2. Performance of supercomputers included in various editions of the Top500 ranking

2. Pace of Supercomputing Technology Development

A general assessment of the pace of SCT development can be obtained by analyzing the achieved performance of supercomputers at various points in time. Figure 2 provides the relevant information.

Some explanations are necessary for Fig. 2:

- The X axis shows the times (month and year) of the Top500 editions, and the Y axis shows the performance on a semilogarithmic scale.
- Solid color lines represent performance of systems occupying different positions in the ranking: the top red graph shows the performance of the system in the first place, and the graphs of systems occupying the 10th (lavender), 100th (blue), 200th (light turquoise), 300th (light green), 400th (light grassy), and 500th (dark green) places follow in descending order.
- From June 2012 to November 2024, dotted lines indicate a forecast of the performance of systems occupying the 1st, 10th, 100th, 200th, 300th, 400th, and 500th places in the ranking. The average value of the projection and a 90% confidence interval are indicated. The forecast was calculated using group linear regression (discussed in Section 2.1) over the interval of the last 10 years, from November 2012 to November 2022.
- Colored dots on the graph indicate systems installed in Russia and included in the relevant Top500 edition. Blue dots represent supercomputers purchased abroad, and red dots represent domestic supercomputers.

Graphs similar to those shown in Fig. 2 are often published and serve as an illustration of the exponential pace of development in the SCT industry. Indeed, at first glance, the color lines in Fig. 2 are close to straight lines, which with the logarithmic scale corresponds to an exponential dependence... However, a more careful analysis shows that the graphs are most similar to a broken line consisting of at least two straight segments: before 2008, the linear regression has one slope, and after 2008, another, less steep one.

2.1. Group Linear Regression

Let us describe the methodology of the above-mentioned analysis. Let us number all releases of the Top500 rating by numbers $i \in [1, 60]$. We denote the performance of the system that took a certain place $j \in [1, 500]$ in the Top500 release with number i as $r_{i,j}$. Let us choose a set of places in the Top500 rating $J \subseteq [1, 500]$ and a rating number $k \in [8, 53]$ – here we deviate by 7 releases of the rating to the left and right in the interval $[1, 60]$.

Then by the method of group linear regression for the given k, J, the trend of the performance is determined by finding such a set of parameters a and b_j, where $j \in J$, that the data series $s_j = \{\, (i, \ln r_{i,j}) \,|\, i \in [k-7, k+7] \,\}$ for $j \in J$ is best approximated by the linear functions $f_j(i) = a \cdot i + b_j$ of the variable i. Specifically, the sum of squares is minimized:

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

The minimization problem is solved by the method of least squares (LS) and, in essence, this is a simple modification of the standard linear regression algorithm: the regression is carried out for a group of data sets s_j given by the indices $j \in J$, assuming that the slope of all approximating straight lines (the parameter a) is the same, and the offsets along the Y axis (the parameters b_j) are different.

By the meaning of the solved problem, we can say that at the time of the Top500 release number k, in half a year the performance of systems occupying places $j \in J$, in average, increased e^a times. Thus, this performance will increase on average a thousand times during $y(k, J) = \frac{\ln 1000}{2 \cdot a}$ years.

2.2. Over How Many Years Did the Performance of Supercomputers Increase by a Factor of 1000?

The quantity $y(k, J)$ was calculated for several sets of J and for all $k \in [8, 52]$. Figure 3 shows the results of the calculations. The calendar date of the release of the Top500 rating with number k is labeled on the X axis instead of the quantity k. The value of $y(k, J)$ (in years) is indicated on the Y axis.

It can be seen that on the interval from 1998 to 2008, it was indeed possible to say that Rmax increased by a factor of 1000 in approximately 11 years. This was true for both the leading positions in the supercomputing rating (red and purple lines on the graph) and for the entire ranking as a whole (green line). Later (2009–2016), we observe signs of clear technical difficulties that restrained the growth of technologies in the industry. Recently, the situation has been improving for the higher-ranked systems in the ranking, and the negative trends have persisted for the ranking as a whole, i.e., the *gap is widening* between the higher-ranked systems and all others.

2.3. Main Conclusions from the Analysis of Supercomputing Technology Development Rates

We can state that at the turn of 2008, while maintaining the exponential nature of performance growth, the rate of growth decreased – the base of this very exponent decreased. Until 2008, the following was precisely maintained: the performance of the most powerful system in the world doubled every 18 months; its performance increased by a factor of 1000 in 11 years. Thus,

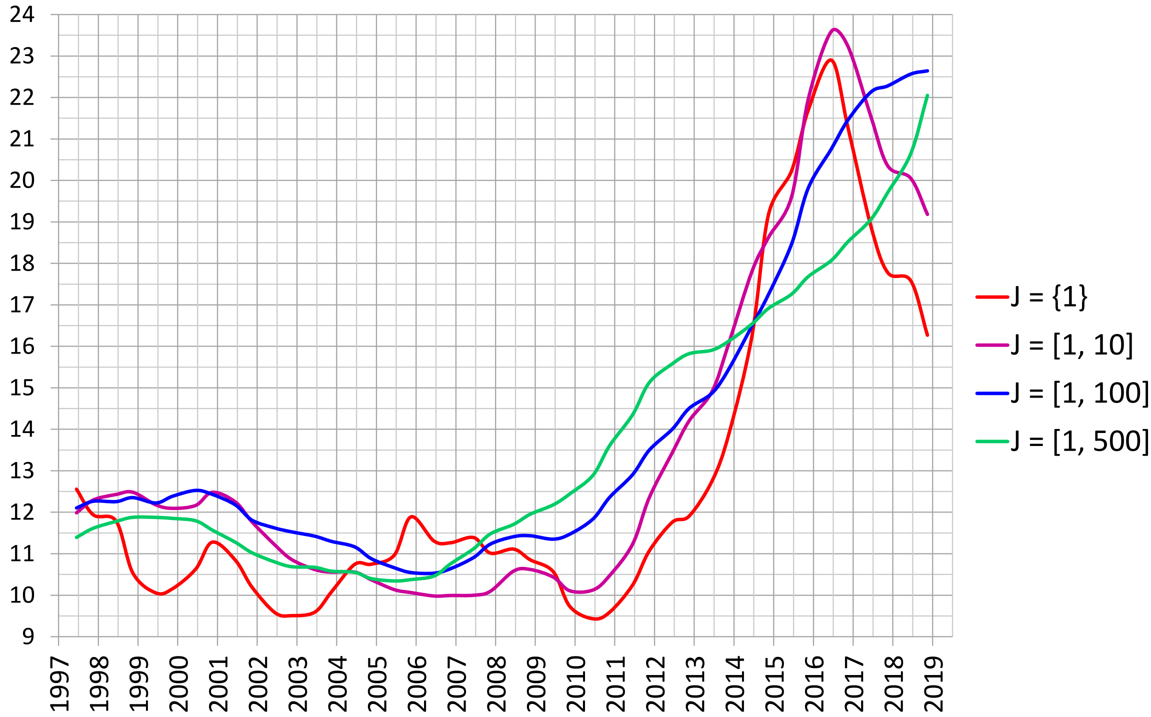

Figure 3. The average forecast for supercomputers occupying positions J is: how long will it take for Rmax to increase by 1,000 times?

the performance milestones were overcome: 1 Mflops (10^6 flops) in 1975, 1 Gflops (10^9 flops) in 1986, 1 Tflops (10^{12} flops) in 1997, and 1 Pflops (10^{15} flops) in 2008. If this trend had persisted, a performance of 1 Eflops (10^{18} flops) would have been achieved in 2019. However, it was achieved in June 2022, see Fig. 2.

Starting from 2008, the rate of growth of achieved maximum performance has clearly changed: instead of an increase of "1000 times in 11 years", we now have an increase of "1000 times in 13–17 years".

Undoubtedly, at the turn of 2008 the world supercomputing industry faced scientific and technical difficulties on the path of SCT development. This led to a review of future performance milestones: 1 Eflops (10^{18} flops) was achieved in June 2022, and one can expect 1 Zflops (10^{21} flops) to be achieved in 2035–2039.

3. Most Powerful Public Supercomputers from June 2011 to November 2022

Table 1 provides brief information about the most powerful supercomputers in the world from June 2011 to November 2022.

Analysis of Tab. 1 reveals obvious trends and allows for reliable (for the next few years) assumptions regarding the general features of exascale systems:

- Number of nodes ranges from 10,000 to 50,000, with 5 to 10 million cores.
- Power consumption ranges from 15 to 30 megawatts.
- Interconnect is, in most cases (87%) of proprietary or customized solutions, or a top standard interconnect with modifications from the manufacturer (13%).
- Cooling systems are closed-water or, in the future, more advanced immersion or boiling.

Table 1. Brief information about the most powerful supercomputers in the world from June 2011 to November 2022

Top500 release, system, ▷ computing subsystem	Rmax, power consumption	Interconnect. Cooling technology
6/2011, K computer, Japan, Fujitsu, RIKEN AICS [6] ▷ **≈0.7M cores**, 88,128 CPU SPARC64 VIIIfx 8C 2.00 GHz 8-cores, 4 processors per node	10.5 Pflops, 12.6 MW	Custom interconnect. Water cooling
06/2012, Sequoia, USA, BlueGene/Q, IBM [7] ▷ **≈1.6M cores**, ≈99K CPU Power BQC 16C, 64-bit RISC, 1.6 GHz 16 cores	16.3 Pflops, 7.9 MW	Custom interconnect. Water cooling
11/2012, Titan, USA, Cray XK7 [8] ▷ **≈0.56M cores**, AMDOpteron – ≈300K cores, NVIDIA K20x – ≈260K cores	17.6 Pflops, 8.2 MW	Custom interconnect. Water cooling
06/2013, Tianhe-2, China, UDT, Inspur [9] ▷ **≈3.1M cores**, 32K Intel Ivy Bridge + 48K Intel Xeon Phi MIC	33.9 Pflops, 24 MW	Custom interconnect. Water cooling
11/2016, Sunway TaihuLight, China, National Supercomputing Center in Wuxi, NRCPC, Inspur [10] ▷ **≈10.65M cores**, CPU Sunway SW26010 260C 1.45 GHz, 40,960×(4+256 cores)	93.0 Pflops, 15.4 MW	Custom interconnect. Water cooling
06/2019, Summit, USA, IBM/NVIDIA [11] ▷ **≈2.4M cores**, IBM POWER9 22C 3.07 GHz – ≈203K cores, NVIDIA Volta GV100 – ≈2.2M cores	148 Pflops, 9.8 MW	Interconnect: Dual-Rail EDR Infiniband. Water cooling
6/2020, Fugaku, Japan, Fujitsu, RIKEN [12] ▷ **≈7.63M cores**, 158,976 CPUs Fujitsu ARM A64FX 48C 2.2GHz 48 cores	442 Pflops, 29.9 MW	Custom interconnect. Water cooling
6/2022, Frontier (OLCF-5, HPE Cray EX235a), USA, DoE, SC/ORNL, HPE [13] ▷ **≈8.7M cores**, 9,248 CPUs AMD Optimized 3rd Generation EPYC 64C 2GHz 64 cores, 36,992 GPUs AMD Instinct MI250X 220 cores	1,102 Pflops, 21.1 MW	Custom interconnect Slingshot-11. Water cooling

- Computer might be homogeneous (50%) or hybrid (50%), depending on the purpose.
- Processors and accelerators are either proprietary (44%) or top standard (56%) and hard available.

4. Trends in Development of Hybrid Architectures

Hybrid supercomputers are systems in which computing nodes are equipped with specialized processors, called accelerators, in addition to standard processors. The idea of accelerators in the supercomputing field has been around for a long time, but hybrid architectures have only become widely adopted in the last decade. Fields detailing the use of accelerators in supercomputers (Accelerator and Accelerator Cores) started appearing in the Top500 list beginning in June 2011. Since then, we have been analyzing hybrid architectures and categorizing each supercomputer into the following classes:

- NONE: accelerators are not used in the supercomputer – homogeneous (non-hybrid) architecture.
- IBM: special-purpose IBM PowerXCell 8i processors are used as accelerators.
- AMD: various specialized processors from AMD are used as accelerators.
- NVIDIA: various specialized processors from Nvidia are used as accelerators.
- Intel: various specialized processors from Intel are used as accelerators – families of Intel MIC / Intel Xeon Phi.
- MIX: various specialized processors from both Nvidia and Intel are used as accelerators.
- PEZY: various specialized processors from PEZY Computing are used as accelerators.
- Other: specialized processors not mentioned above are used as accelerators. Currently, the only accelerators in this category are those developed in China: Matrix/2000 and Deep Computing Processor.

We examined 24 versions of the Top500 ranking from June 2011 to November 2022. Figure 4 shows the proportions of different classes of hybrid supercomputers in the Top500 list for these 24 versions. The labels on the X axis represent years, and the two bar graphs in each year correspond to the two bi-annual versions of Top500 in that year. In the left part of the figure, the proportions are shown in the number of systems (SCs), with each system representing 0.2% of the 500 systems, while in the right part, the proportions are shown in terms of performance (RMax), with the total performance of the entire Top500 list (i.e., ΣRmax) taken as 100%. The following trends can be observed:

- While the number of hybrid supercomputers increased by 12.8% ↗ 35.6% from 2014 to 2022, their contribution to the total Top500 list performance ΣRmax, which is the true industry share, increased by 34.4% ↗ 63.8% over this period.
- IBM's accelerators practically disappeared by 2013.
- Intel's accelerator share has fallen by 18.8% ↘ 0.1% ΣRmax from 2013 to 2022.
- NVIDIA's share has increased by 16.4% ↗ 30.9% ΣRmax from 2014 to 2022.
- AMD's accelerators have practically disappeared by 2017, but then increased by 0.1 ↗ 31.4% ΣRmax.
- Interesting new solutions that deserve attention have emerged (5.4% ΣRmax in the best, 2018) – short and successful projects to create fully proprietary accelerators:
  - PEZY – 2015, PEZY Computing, Japan;
  - Matrix/2000 – 2017, NUDT, China;
  - Deep Computing Processor – 2017, Sugon, China.

4.1. Quintiles of Supercomputers in the Top500 List

It is very important to understand the applicability and suitability of each technology in supercomputers of various performance levels. For example, it is important to understand in which systems a particular interconnect technology or accelerator class is applicable and in demand today. To do this, we will divide the entire Top500 list (which is sorted by supercomputer performance) into five groups: quintile A – several of the most powerful supercomputers in

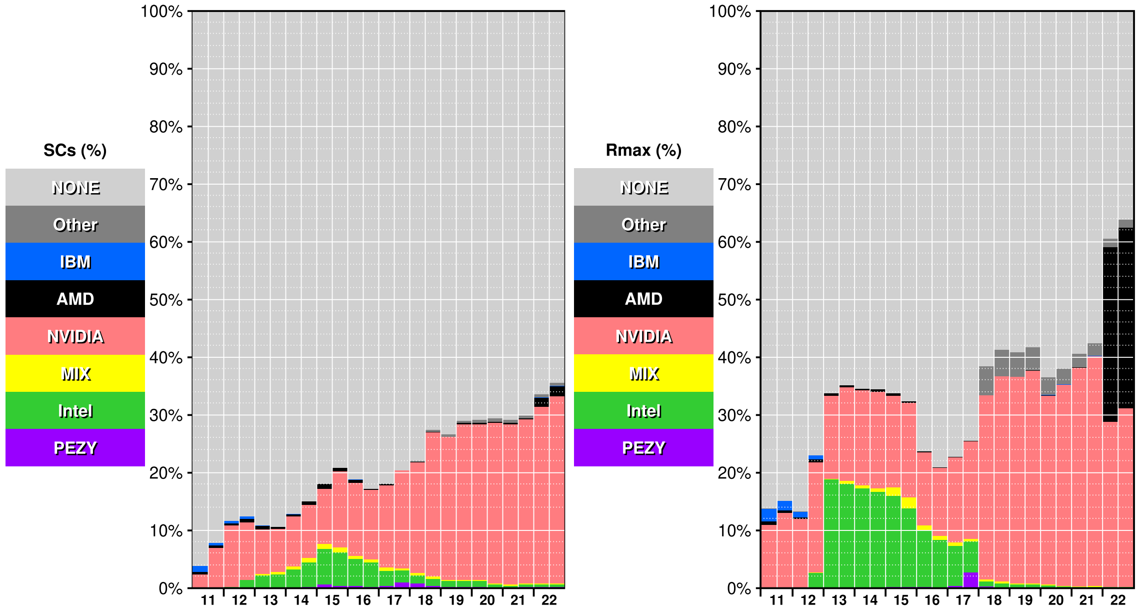

Figure 4. Proportions of different classes of hybrid supercomputers in the Top500 ranking, editions from June 2011 to June 2022

the list, followed by quintiles B, C, D, E. The size of the quintiles will be chosen so that the sum of the performance of all supercomputers in one quintile is as close as possible to 20% of the total performance ΣRmax of the entire Top500 list. That is, quintiles A, B, C, D, and E have (approximately) equal performance, but of course contain different numbers of supercomputers. For each release of the Top500 ranking, the composition of quintiles A, B, C, D, and E (the number of supercomputers in them) is recalculated, based on the requirement of approximating the total performance of each quintile to 20% ΣRmax as accurately as possible.

For the November 2022 edition of the Top500 ranking, the quintile breakdown was as follows:

- A: 1 system, Top1, 22.7% ΣRmax;
- B: 3 systems, Top2–4, 19.0% ΣRmax;
- C: 22 systems, Top5–26, 19.9% ΣRmax;
- D: 111 systems, Top27–137, 20.0% ΣRmax;
- E: 363 systems, Top138–500, 18.4% ΣRmax.

4.2. Distribution of Different Classes of Hybrid Architectures across Quintiles

For each quintile – A, B, C, D, E – let us look at the shares of the None, Other, IBM, AMD, NVIDIA, MIX, Intel, PEZY classes in it (Fig. 5).

By paying attention to the shape and center of gravity of the figures with the corresponding fill color, we can make the following judgments for the November 2022 trends:

- AMD accelerators are predominantly used in the most powerful systems – the entire quintile A and a significant fraction in B;
- NVIDIA accelerators and non-hybrid systems (class NONE) are more or less evenly represented in quintiles B, C, D, E (30%–50% in B, C, D, E);
- other class systems (Chinese accelerators) are mostly represented in quintile C (7%);

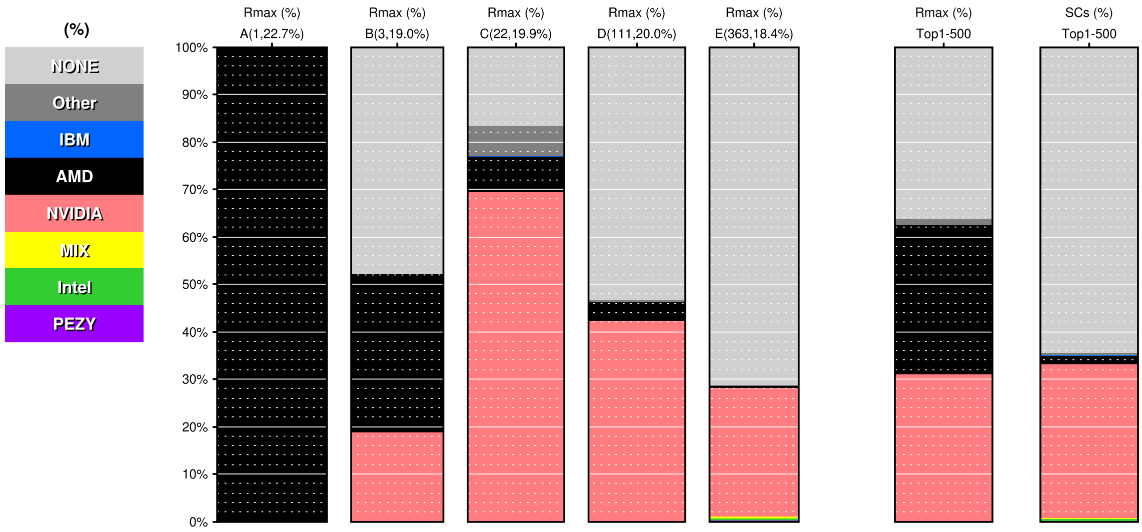

Figure 5. Shares of various classes of hybrid supercomputers in the total performance of quintiles A, B, C, D and E of the Top500 rating November 2022 edition

- MIX, Intel, PEZY accelerators are weakly represented in quintile E – leaving the rating and industry;
- IBM class accelerators have left the industry.

5. Trends in Interconnect Technologies

The interconnect connects all nodes in a supercomputer into a single system. The network technologies used in the interconnect are often different from those used in local, regional, and global computer networks, due to the specific requirements placed on the interconnect: not only high bandwidth, but also minimal latency and maximum message rate, as well as several other specific requirements. In the supercomputing industry, this plays a significant role in determining the interconnect technologies. In the list of 10 critical technological challenges that the industry is facing on the way to exascale systems [14], the development of advanced interconnect technologies occupies a very high (second) place.

Fields that detail the use of various interconnect technologies in supercomputers are present in all editions of the Top500 list. Based on the data in these fields, each supercomputer in all editions of the Top500 list can be classified into the following categories:

- Infiniband: if the supercomputer uses Infiniband technology to implement the interconnect. Such supercomputers have been present in the Top500 list since June 2003 and are still present with several versions of Infiniband having different technical characteristics: Infiniband SDR (8 Gbit/s), Infiniband DDR (16 Gbit/s), Infiniband QDR (32 Gbit/s), Infiniband FDR (54 Gbit/s), Infiniband EDR (100 Gbit/s), Infiniband HDR (200 Gbit/s).
- Ethernet: if the supercomputer uses Ethernet technology to implement the interconnect. Such supercomputers have been present in the Top500 list since June 1996 and with several versions of Ethernet having different technical characteristics ranging from FastEthernet (100 Mbit/s) to 100G Ethernet (100 Gbit/s).

- Myrinet: if the supercomputer uses the Myrinet technology to implement the interconnect. Such supercomputers have been present in the Top500 list since November 1998 until November 2020.
- Quadrics: if the supercomputer uses the Quadrics technology to implement the interconnect. Such supercomputers have been present in the Top500 list since June 1999 until November 2011.
- SCI: if the supercomputer uses the SCI technology to implement the interconnect. Such supercomputers have been present in the Top500 list since June 2002 until November 2004.
- OpTsIntel: if the supercomputer uses the Intel Omni-Path/TrueScale family of solutions to implement the interconnect. Such supercomputers have been present in the Top500 list since November 2013 and are still present.
- Custom: if the supercomputer uses proprietary network solutions or solutions that are not available as a separate commercial product – only available as part of a complete system.

It is important to note the difference between the categories:

- Infiniband, Ethernet, Myrinet, Quadrics, SCI – these interconnect technologies are available as separate commercial products. Any developer of their own supercomputer (for example, with their own architecture of computing nodes, with certain processors) can buy and apply any of these solutions.
- Custom – these interconnect technologies are not available as separate commercial products. They are either unavailable or only available as part of a complete system from a single supplier.
- OpTsIntel – Intel's Omni-Path / TrueScale technologies occupy an intermediate position. On the one hand, these interconnect technologies are available as commercial products. On the other hand, these solutions cannot be applied in a supercomputer if the computing node does not have the appropriate architecture and processor from Intel. And this situation is very close to the situation where the interconnect is only available as part of a complete system from a single supplier – close in terms of the rigidity of technological dependence on a single developer and supplier, and dependence on Intel.

Intel's Omni-Path / TrueScale technology had very high technical characteristics at the time of its appearance, which allowed it to successfully enter the industry and gain a share of 11.0% of the ΣRmax in November 2019. The weak point of the OpTsIntel solution is the rigid technological dependence of the entire system on Intel. And today, this technology is leaving the industry. Figure 6 shows the performance shares of supercomputers with different interconnects in the total performance of the Top500 ranking ΣRmax for 60 editions from June 1993 to November 2022.

Captions on the X axis represent the years, with two bar graphs corresponding to two editions of the list in a particular year.

Analyzing the figure, the following trends of the past 10 years can be noted:

- In recent years, technologies corresponding to classes Myrinet, Quadrics, and SCI are not used in the industry. The figure clearly shows the years of introduction, peak popularity, and decline of these technologies. Today, only 4 classes remain relevant: Infiniband, Ethernet, OpTsIntel, and Custom.
- Since 2012, the highest share of the total performance of the Top500 belongs to supercomputers with Custom interconnects. The dynamics for 2012–2017–2022 are as follows:
  - Custom 54.7% ↘ 42.4% ↗ 53.4%.

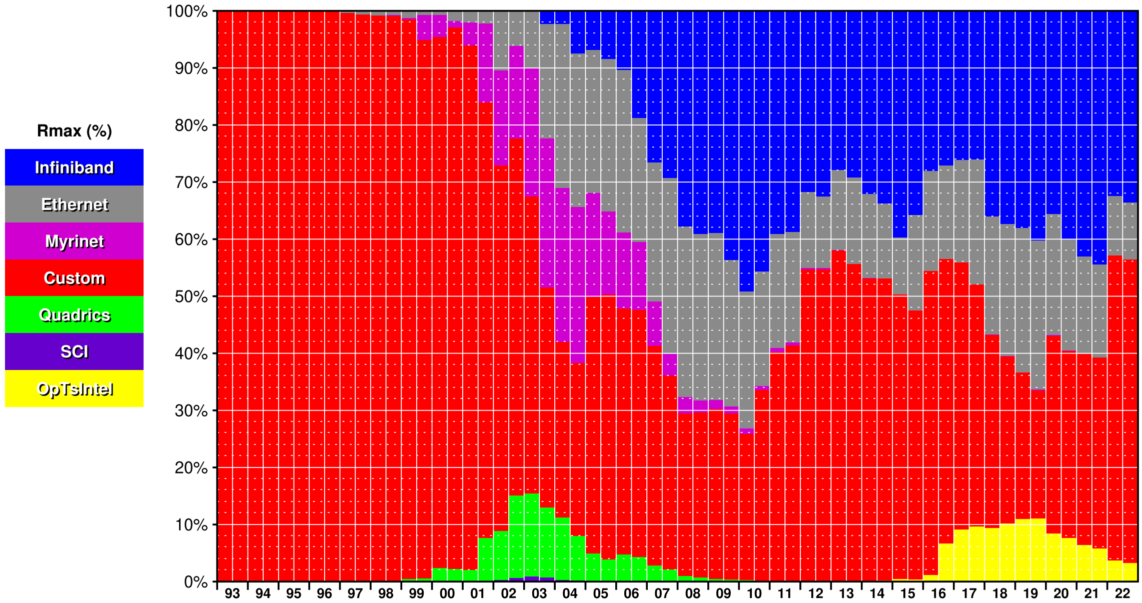

Figure 6. Shares of supercomputers with various interconnects in the total performance of the Top500 ranking, from June 1993 to November 2022

- Next most significant share belongs to Infiniband. The dynamics for 2012–2017–2022 are as follows:
  - Infiniband 32.5% ↘ 26.1% ↗ 33.6%.
- This is followed by Ethernet. The dynamics for 2012–2017–2022 are as follows:
  - Ethernet 13.2% ↗ 21.9% ↘ 10.0%.
- OpTsIntel technologies are leaving the Top500 and the industry. The dynamics for 2012–2017–2022 are as follows:
  - OpTsIntel 0.0% ↗ 9.6% ↘ 3.2%.

5.1. Distribution of Different Interconnect Classes by Quintiles

It is very important to understand the areas of applicability of each interconnect class. To do this, let us look at the distribution of the shares of different interconnect classes by quintiles. The definition of quintiles is given in Section 4.1.

For each quintile – A, B, C, D, E – let us look at the shares of Infiniband, Ethernet, OpTsIntel and Custom interconnect classes – Fig. 7.

It can be seen that:

- Most powerful and mighty supercomputers (quintiles A and B) predominantly use Custom interconnect. The red color fills the triangle, which expands towards the most powerful systems;
- Infiniband technology is predominantly used for medium and lower-level systems – with the largest shares in quintiles C and D. The blue triangle has its center of gravity in this zone;
- Ethernet technology is more often used in the weakest systems (D, E). The grey triangle significantly expands towards the weakest systems;
- OpTsIntel solutions are leaving the list and the industry.

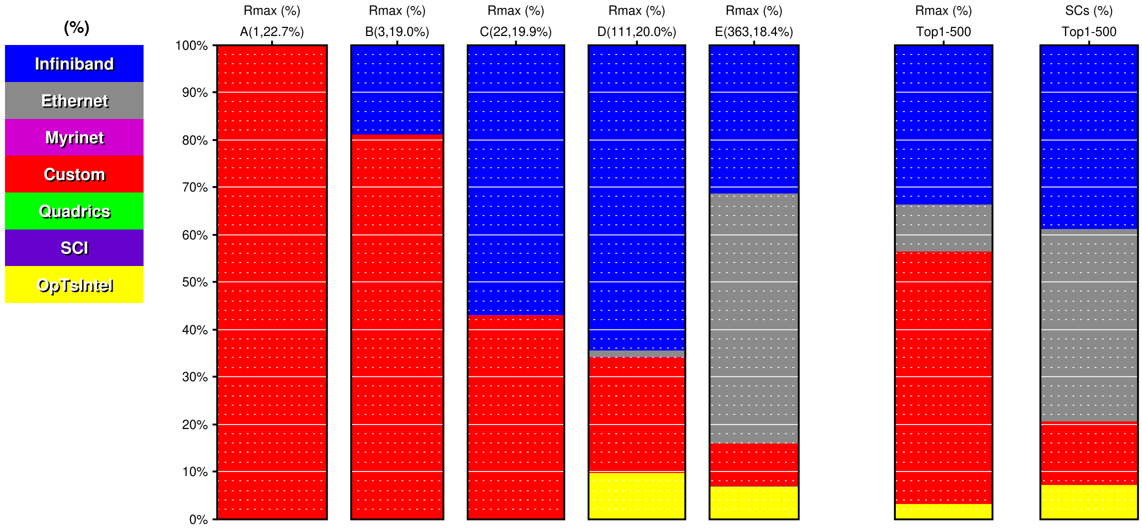

Figure 7. Shares of supercomputers with different interconnect classes in the total performance of quintiles A, B, C, D and E of the Top500 ranking for November 2022 edition

Conclusion

At the turn of 2008 the world supercomputing industry undoubtedly faced scientific and technical difficulties in the development of HPC. Since 2008, the rate of growth of the achieved maximum performance has changed: instead of an increase "1000 times in 11 years", we have an increase "1000 times in 13–17 years". This led to a revision of future forecasts for achieving new performance milestones: 1 Eflops (10^{18} flops) was achieved in June 2022, and 1 Zflops (10^{21} flops) is expected to be achieved in 2035–2039.

Analysis of the most powerful supercomputers of the last decade allows us to make the following generally reliable (for the next few years) assumptions about the overall appearance of exascale-level systems:

- number of nodes – 10–50 thousand, number of cores – 5–10 million;
- power consumption – 15–30 MW;
- interconnect – most likely (7/8 = 87% of cases) Custom (proprietary network, specific solutions or not available as a separate commercial solution), or top (difficult to access) standard interconnect, but with modifications by the manufacturer (1/8 = 13%);
- cooling – either closed water-based or more advanced immersion, boiling, etc.;
- computational subsystem – depending on the purpose – either homogeneous (50%), or hybrid (50% of cases);
- processors and accelerators – either proprietary (44%), or top (limited availability) standard (56% of cases).

Analysis of hybrid architectures in the Top500 rankings indicates that the contribution of hybrid architectures to the overall performance of the Top500 has grown and reached 63.8% ΣRmax. Among the current solutions for hybrid systems, notable accelerators include AMD (31.4% ΣRmax) and NVIDIA (30.9% ΣRmax). Interesting new solutions also deserve attention (5.4% ΣRmax in 2018) – short and successful projects for creating fully proprietary accelerators: PEZY (2015, PEZY Computing, Japan), DeepComputingProcessor (2017, Sugon, China), Matrix-2000 (2017, NUDT, China).

Various interconnect technologies continue to intensively develop in the world. They play an exceptionally important role for the entire HPC complex. Currently, commercially available technologies include Custom (53.4% ΣRmax), Infiniband (33.6% ΣRmax), and Ethernet (10.0% ΣRmax). Since 2012, the largest share of the total performance of Top500 supercomputers is for the interconnects that are not available as individual commercial solutions (Custom class).

For building high-performance systems, top-end (and therefore limited availability) models of the most successful commercial developments such as AMD processors are advanced, as well as proprietary processors based on one of the available architectures, such as ARM (which has proven to be promising for this use in supercomputers) and possibly RISC-V.

Significant changes have taken place among computer manufacturers in recent years. Chinese manufacturers have made a serious intervention in the industry, they accounted for 63.0% of the number of deals and 43.4% of volumes (in terms of ΣRmax) in November 2019. China has the largest share of the entry-level systems.

The following fact is important for building high-performance computing systems:

- they are always used for research and development; for generating new knowledge, technologies, materials; for scientific calculations;
- they are never used directly for real-world economic tasks, such as engineering calculations, financial applications, communication services, internet services, etc.

This has an impact on the class of problems solved: on the cutting edge of scientific research during the lifespan of the record installation (about 5–7 years), special problems may arise that require special mathematical and algorithmic tools for their solution. This should be taken into account when developing architectural solutions, both in terms of hardware and software for such systems.

Currently, the US, China, Japan, and the EU have built a solid HPC infrastructure as a basis for the transition to a digital economy and have significant shares of their countries' overall global supercomputer performance. These shares are greater than their shares in the global GDP (with the exception of China, which hides its achievements in HPC). This ratio illustrates the real progress of countries towards a digital economy.

References

1. Ezell, S.J., Atkinson, R.D.: The Vital Importance of High-Performance Computing to U.S. Competitiveness. The Information Technology and Innovation Foundation (2016). 58 p.

2. Abramov, S.M.: Facts that distort reality. How to analyze Top500? Bulletin of South Ural State University: Series Computational Mathematics and Software Engineering 2(3), 5–31 (2013). https://doi.org/10.14529/cmse130301 (in Russian)

3. Abramov, S.M.: June 2019: analysis of the development of the supercomputer industry in Russia and in the world. Program Systems: Theory and Applications 42, 3–40 (2019). https://doi.org/10.25209/2079-3316-2019-10-3-3-40 (in Russian)

4. United States Department of Energy. Top Ten Exascale Research Challenges. Department of Energy ASCAC Subcommittee Report (2014). 86 p.

5. Top500 Official cite. https://www.top500.org/, accessed: 2023-06-06

6. Supercomputer "K computer" Takes First Place in World. https://www.fujitsu.com/global/about/resources/news/press-releases/2011/0620-02.html, accessed: 2023-06-06

7. With 16 petaflops and 1.6M cores, DOE supercomputer is worlds fastest. https://arstechnica.com/information-technology/2012/06/with-16-petaflops-and-1-6m-cores-doe-supercomputer-is-worlds-fastest/, accessed: 2023-06-06

8. ORNL Debuts Titan Supercomputer. https://www.olcf.ornl.gov/wp-content/themes/olcf/titan/Titan_Debuts.pdf, accessed: 2023-06-06

9. Chinas Tianhe-2 Caps Top 10 Supercomputers. https://spectrum.ieee.org/tianhe2-caps-top-10-supercomputers, accessed: 2023-06-06

10. Dongarra, J.: Sunway TaihuLight supercomputer makes its appearance. National Science Review 3(3), 265–266 (2016). https://doi.org/10.1093/nsr/nww044

11. Oak Ridge National Laboratory, IBM, NVIDIA: Introducing the Summit Supercomputer (2020). 24 p.

12. Dongarra, J.: Report on the Fujitsu Fugaku system. University of Tennessee-Knoxville Innovative Computing Laboratory, Tech Report ICLUT-20-06 (2020). 18 p. https://www.icl.utk.edu/files/publications/2020/icl-utk-1379-2020.pdf, accessed: 2023-06-06

13. Frontier spec sheet. https://www.olcf.ornl.gov/wp-content/uploads/2019/05/frontier_specsheet.pdf, accessed: 2023-06-06

14. Coghlan, S.: Argonnes Aurora Exascale Computer. Smoky Mountain Computational Sciences And Engineering Conference (2019). 17 p. https://smc.ornl.gov/wp-content/uploads/2019/09/Coghlan-presentation-2019.pdf, accessed: 2023-06-06