DOI: 10.14529/jsfi250101

Recent Technology Trends in High-Performance Computing: Quantifying the Divergence of Top and the Rest

***Sergey S. Konyukhov*[1], *Alexander A. Moskovsky*[1,2]**

Exploratory analysis methods were used to study basic characteristics of computing systems from TOP500_Lists. One of the peculiarities of the distribution of computing systems by performance is that it sufficiently well obeys an analog of the empirical Zipf's law, in which logarithm of performance is reciprocal to the rank of computing system. Based on this observation we can divide all systems from the lists into several performance classes: top, high, base, and entry levels. Our analysis also revealed differences between these classes in other characteristics besides, the computational performance, e. g., such as power consumption. For all performance classes, trends in evolution of the basic characteristics of TOP500 computing systems were described and, where possible, comments were provided to explain their behavior. Performance and energy efficiency of the TOP500_List computing systems in the next 5–10 years were estimated using simple linear models obtained by the least-square method. We have found that energy consumption needed for entry-level supercomputers to surpass the threshold value of performance and to enter into TOP500_List will decrease during this period.

Keywords: TOP500_List, *energy efficiency, supercomputer technology trends, Zipf's law.*

Introduction

Data archives of the world-wide known TOP500_List present an indispensable source of information about technology trends, international politics, economy, sociology, and other subjects concerning the area of high-performance computation. The GREEN500 rating emerged more recently with such important aspect as measurement of energy consumed by large high-performance computing systems. As average energy consumption increased, this aspect became one of the most important characteristics of computing systems along with their performance, so the latest TOP500_Lists contain information about energy efficiency of computing systems. Our work is dedicated to the combined analysis of performance and energy efficiency for computing systems from TOP500_Lists.

TOP500_List consists of 500 entries sorted by maximal performance in HPL test, R_{max}. The rating has been updated twice a year since 1993. In each edition, Erich Strohmaier, Jack Dongarra, and other rating authors present an analysis of new data and trends. They analyze long-term trends of R_{max} on TOP1, TOP100, TOP500, or other sublists, and different aggregates such as sums of R_{peak} or R_{max} of all TOP500 entries.

While R_{max} of Rank_1 position for both GREEN500 and TOP500 lists demonstrate sharp, step-like increase over the last years, most aggregate numbers change more smoothly. Decades-long trends are clearly visible in the analysis of R_{max} (R_{peak}) of Rank_1 or Rank_500 positions what allowed to make the long-term predictions such as the development of petaflops machines n the early 2000s or the emergence of the exascale supercomputer by 2018 back in 2007–2008.

By the mid 2010s the fast pace of the performance increasing slowed down, driven by slower progress of microelectronics[3] and budget limitations. By the end of the decade, the number of new systems in each TOP500_List edition also declined dramatically to 40–50 from more than 150 in the 2000s.

[1]RSC Group, Moscow, Russian Federation
[2]Lomonosov Moscow State University, Moscow, Russian Federation
[3]Proclaimed Moore's Law "death".

During the same period, energy efficiency of the Rank_1 computing systems from GREEN500_List grew by the factor of 20 over the last 10 years. However, this rapid progress was enrolled with diminishing returns from new microelectronic technology generations compared to six decades of development of computer architectures. We think that it would be useful to characterize quantitatively how such progress of the high-end technology is supported by a "common" supercomputer system from TOP500_List because the most performant TOP500 systems cannot rely on bleeding edge technology – they need mature solutions that can be deployed at scale. At the same time GREEN500 rating is a better indicator of new technology trends as it demonstrates the best solutions viable enough to run efficiently the complex program stack of LINPACK benchmark. In other words, we could suppose energy efficiency became the most important characteristic of computing system.

In our work we noted and measured the recent decline in the "TOP500 entry power ticket", i. e., minimal energy consumption between all TOP500 systems, $Min(P)$. We suppose that it demonstrates an important trend in the current technology development: divergence of the most energy efficient architectures from a "general case" represented in low-rank TOP500 systems. In other words, the cutting edge advances in CPU/GPU architecture become available to the majority of practical users only after a significant period of time.

The article is organized as follows. Section 1 is devoted to the review of related works. In Section 2 we describe our method by which we analyze data from TOP500_Lists. Section 3 contains the discussion of obtained results. Conclusion summarizes the study and points directions for further work.

1. Related Works

TOP500_List was initiated as a source of information to identify the current trends in the field of high-performance computing and continue Mannheim List statistics proposed by Hans Meuer [8]. Since 2001 in addition to brief analyses performed by the rating authors in each TOP500 issue, papers have been regularly published, in which their authors tried also to reveal technology trends and tendencies by statistical methods [1–4, 6, 7].

The scope of analysis was broad: from low-level detailed analysis of various subsystem designs to market share of HPC vendors and application areas. In our paper we focus on energy efficiency of supercomputers, or more specifically, on how much power would be enough to run supercomputer having minimal performance sufficient to appear on TOP500_List.

The work [7] presents "strawman" design solutions of future supercomputers and uses the TOP500 trends to estimate relative performance and energy efficiency of future designs, and especially to carry out detailed analysis of future design of memory and interconnect subsystems. Although in this work physical and infrastructure limits were taken into consideration, economic limitations were underexplored. In our work we analyze relations between performance and energy efficiency in a more formal way.

Out of the most recent publications, the papers [4] and [6] seem to be the most relevant to the subject of our study. The former article overviews the performance and energy efficiency trends and tries to give reasons for them by observing features of new hardware architectures. For these purposes heterogeneous systems and top-rank systems were carefully studied.

The authors of the later paper made a unique comparison of actual trends of the 2010s with the past assessment of "Exascale Report-2008" [5], and the "Frontier" supercomputer characteristics were compared with "strawman" models presented in 2015. As a result this work gives an idea how accurate such long-term projections can be. While the works [4] and [6] focus mostly on top-rank systems, our work investigates trends in the entry-level systems of the TOP500_Lists.

Continuing a series of their works, Abramov and Abramov [1] analyze the TOP500_Lists of the latest years from different points of view. Together with other researchers they determine the concentration of the most part of the aggregate performance by the top-ranks systems in complete agreement with our observations.

2. Method

In this work we have analyzed data collected from *top500.org* website back to June 2013 since the first GREEN500 data became available. We used methods of exploratory data analysis and try to figure out evolution trends for the basic characteristics of TOP500 systems.

We used the following primary data (i. e., measured data):

- maximal performance, R_{max};
- peak performance, R_{peak};
- energy consumption, power, P;
- total number of cores, N_{cores};
- number of accelerator/co-processor cores, $N_{accores}$;

and secondary data (i. e., derived quantities):

- resource usage efficiency, $RUE = R_{max}/R_{peak}$;
- energy efficiency, $EE = R_{max}/P$.

First of all, we were interested in total energy consumption of a computational system having performance compared with the performance of Rank_500 system because the cost of electricity and the need to create special infrastructure required for the operation of computational equipment with high energy consumption can limit the usage of high-performance systems in practice.

Due to the specifics and quality of the used data[4], we applied the following procedure for our analysis:

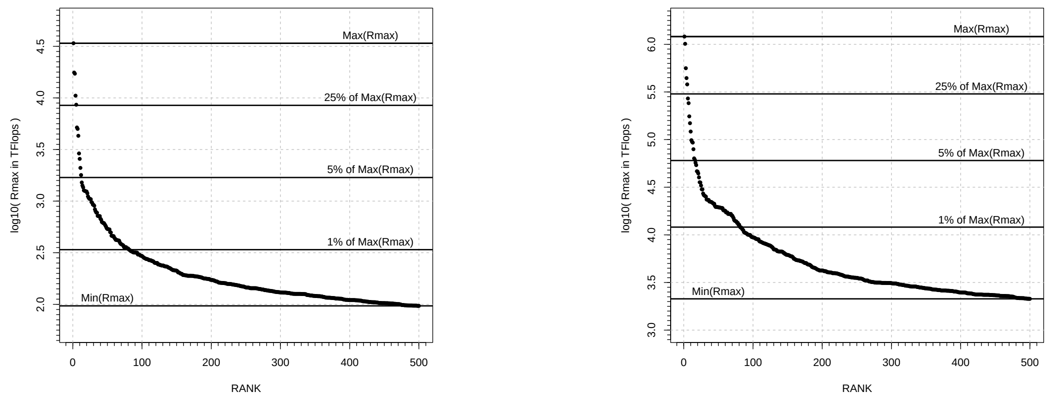
1. divide computational systems from the TOP500_Lists by R_{max} into several classes;
2. for each class deduce extrapolation formulae (models) describing behavior of R_{max} and P as functions of time (more precisely, as functions of TOP500_List number);
3. for two classes with strongly correlated data, transfer the model from the class having more reliable statistical description to the class with less reliable data.

3. Discussion

The distinctive feature of the TOP500 systems is a significant difference between characteristics such as maximal performance, energy consumption, or total cores number for Rank_1 computing systems and the vast majority of other computing systems in a list.

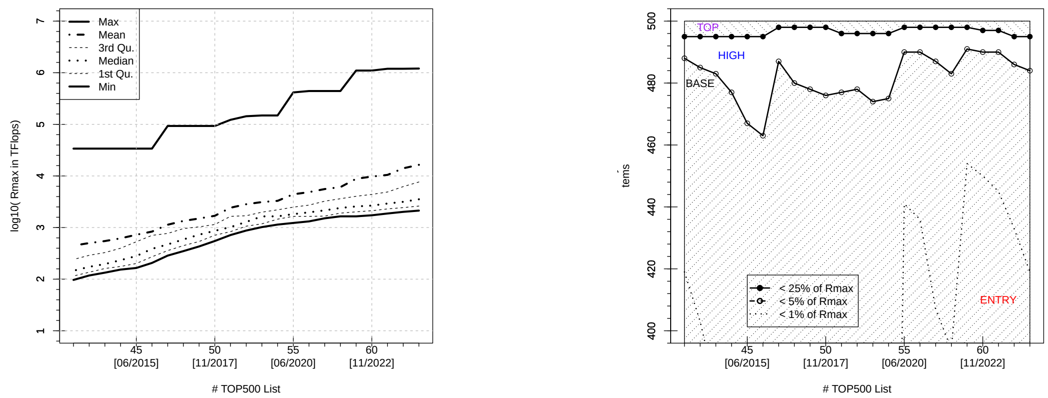
The distribution of R_{max} by system ranks on a semi-logarithmic scale can be described as an analogue of Zipf's empirical law (see Fig. 1) and this behaviour is more or less the same for all studied lists (see Fig. 2a).

[4]While the publication of R_{max} is evidentally mandatory for TOP500 systems, such data as energy consumption are practically absent for the most part of TOP500_List.



(a) List#41 [June/2013] (b) List#63 [June/2024]

Figure 1. Stratification of TOP500 systems by maximal performance



(a) Statistical distribution of R_{max} (b) Total number systems in different classes

Figure 2. Distribution of TOP500 systems by R_{max}

The data presented in Fig. 1 show that it is quite difficult to provide an objective criterion to divide all systems from TOP500_LIST into separate classes, so we formally divided the dynamic range of $\log(R_{max})$ values into 4 approximately equal intervals, giving preference to the ease of perception of information compared to the exact values. Thus, denoting R_{max} for Rank_1 system as $Max(R_{max})$ we use as class boundries $Max(R_{max})/4, Max(R_{max})/20$, and $Max(R_{max})/100$, $(25\%, 5\%$, and 1% of $Max(R_{max}))$, instead of

$$\log\bigl(Max(R_{max})\bigr) - \frac{\log\bigl(Max(R_{max})\bigr) - \log\bigl(Min(R_{max})\bigr)}{4} \times K, \quad K = 1, 2, 3;$$

what, as we believe, should not greatly change our results and conclusions. Moreover, at $R_{max} = Max(R_{max})/4$ the distribution changes its character, the density of points noticeably increases and the slope of the curve decreases, i. e., the division of systems by this boundary seems to us more or less objective and justified.

So, for the purpose of our analysis we distinguish the following types of computing systems:

- top class with $R_{max} > 25\%$ of $Max(R_{max})$;
- high class with $R_{max} > 5\%$ of $Max(R_{max})$ and $R_{max} < 25\%$ of $Max(R_{max})$;

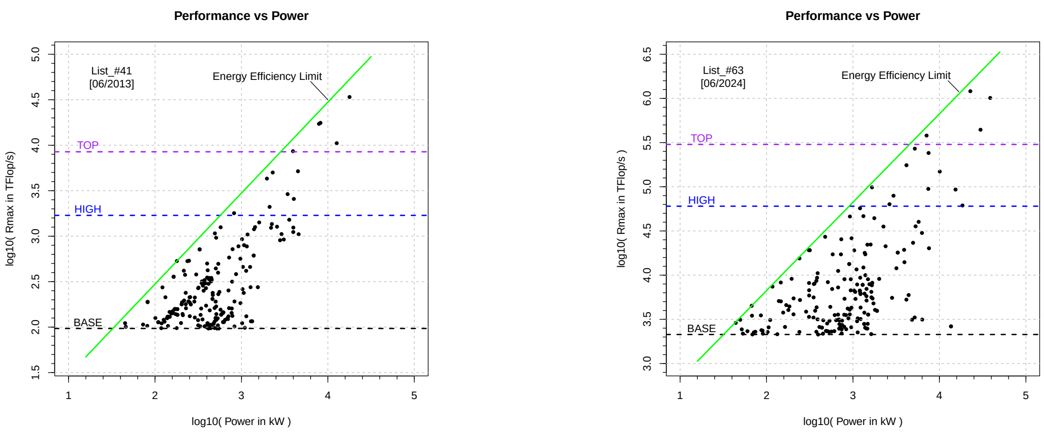
- base class with $R_{max} > 1\%$ of $Max(R_{max})$ and $R_{max} < 5\%$ of $Max(R_{max})$;
- entry class with $R_{max} < 1\%$ of $Max(R_{max})$.

We refer further to these classes as TOP, HIGH, BASE, and ENTRY classes, respectively.

The TOP class usually comprised 2–5 computing systems (see Fig. 2b). In the HIGH class approx. 7–33 computing systems resided. Total number of computing systems in the BASE and ENTRY classes, i. e., with $R_{max} < 5\%$ of $Max(R_{max})$, was never less than 460 during the reviewed period (June 2013 – June 2024), so the BASE and ENTRY classes were never less than 92% of the total number of computing systems in TOP500_List.

The ENTRY class is not only the most widespread class, but also the most diverse one in its design and technical characteristics (see, e. g., Fig. 3). Since the number of computing systems from the ENTRY class reached up to 90% of the total number of systems within the BASE and ENTRY classes, they are described further as one BASE class (as it denoted in Fig. 2b), and the term *entry-level systems* is reserved for computing systems with R_{max} close to R_{max} of Rank_500 systems, which we will denote further as $Min(R_{max})$.

Each of the aforementioned classes has its own statistical characteristics. The TOP and HIGH classes include relatively few computing systems; therefore, to describe their typical properties, it is necessary to use such robust estimators as median. Although the BASE class is the most representative class in terms of the number of computing systems, their properties change over a wide range of values what from a statistical point of view can be interpreted as the presence of "outliers", and the usage of robust estimators like median is also preferable in this case.

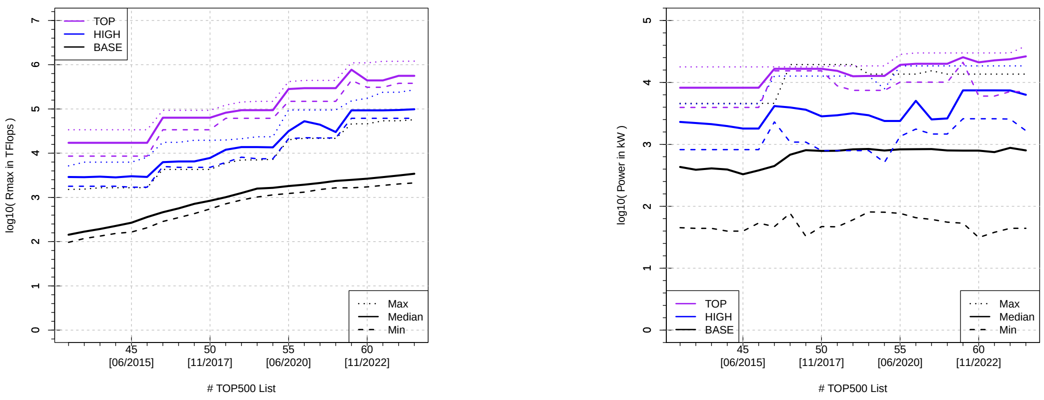


(a) List#41 [June/2013] (b) List#63 [June/2024]

Figure 3. Scatterplots based on energy consumption and maximal performance data for TOP500 systems

Segregation of computing systems from TOP500_List into three classes is observed not only by their maximum (or peak) performance, but also by other extensive characteristics, first of all, by energy consumption and total number of computing cores (see Fig. 4–Fig. 6).

The behavior of intensive quantities defined either as ratio of two different extensive quantities (e. g., energy efficiency) or as fraction (e. g., resource usage efficiency) is more complex. For each TOP500]_List maximum energy efficiency among all systems almost always was attained by not top-level, but basic-level systems (see Fig. 5a); which is understandable because the most

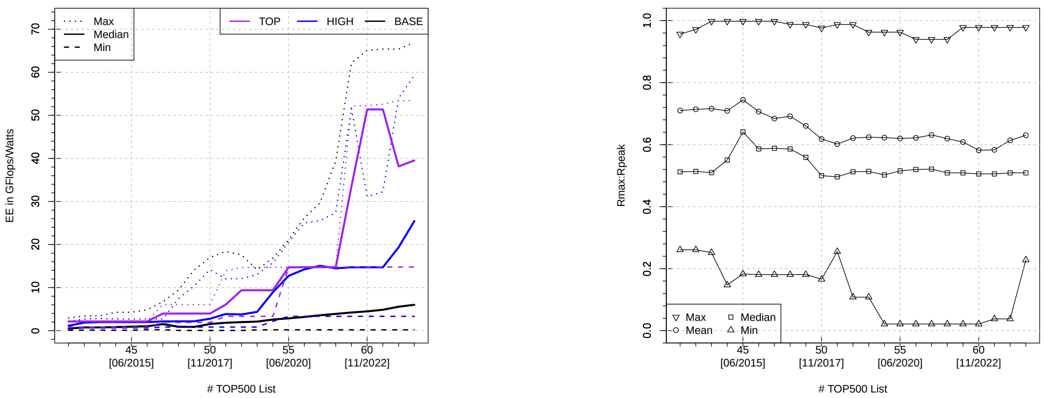


(a) Maximal performance, R_{max} (b) Power, P

Figure 4. Basic characteristics of TOP500 systems for different performance classes [Max/Median/Min – maximal/median/minimal within a class]

energy-efficient technologies, due to their novelty, are first tested on relatively small computing systems.

During the observed period the ratio of co-processor cores to total number of cores was approximately the same for all three classes (see Fig. 6b) and was ≈ 0.7 until mid-2016, and then increased more or less abruptly to 0.9. Therefore, it can be assumed that the difference between classes is due to absolute value of total core number and performance per core, rather than a fundamental difference in architecture.



(a) Energy efficiency, EE (b) Resource usage efficiency, RUE

Figure 5. Computational efficiency of TOP500 systems [Max/Median/Min – maximal/median/minimal within a class]

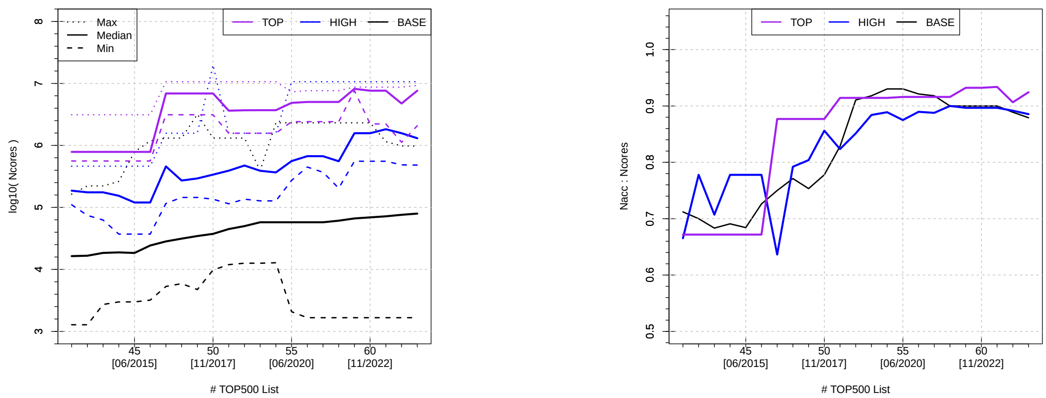
Similarly, the resource usage efficiency for all three classes has the same distribution with no significant difference between the mean and median values (see Fig. 5b). Possibly it can be explained by the fact that this characteristic depends more strongly on software settings and HPL test running conditions rather than on hardware parameters and specifics. More precisely, the increasing of GPU cores number in computing system hinders their usage efficiency. Indirectly it is confirmed by the fact that starting from Lists#48–52 there was a significant growth in the proportion of co-processor cores used in computing systems accompanied by a fall in the resource usage efficiency (compare Fig. 5b and Fig. 6b).

Considering the evolution of the main extensive characteristics we may conclude that:

- all characteristics have revealed growth except the minimal energy consumption, $Min(P)$, and energy consumption of BASE class, $P(BASE)$, the behavior of which is more complicated;
- gaps between characteristics of the top and base or entry level systems also have widened during the observation time.

Data given in Tab. 1 and Tab. 2 illustrate these trends. Meanwhile some comments would probably be appropriate.

For the TOP and HIGH classes the average number of cores in computing system correlated with the average performance (see Fig. 6a). However, for the BASE class the dependence between these two characteristics was extremely complicated. Indeed, the maximum number of cores for the base-level systems was comparable to the minimum number of computing cores for the top-level systems, meanwhile the performance between the base-level and top-level systems could be different by 2–3 orders of magnitude (compare Fig. 4a and Fig. 6a).



(a) Total cores number, N_{cores}

(b) Accelerator/Total cores numbers ratio, $N_{accores}/N_{cores}$

Figure 6. Formal performance characteristics of TOP500 systems [Max/Median/Min – maximal/median/minimal within a class]

While the maximum energy consumption has increased steadily over time, the minimum energy consumption could both increase and decrease (see more details later, Fig. 7b). It can be arguably explained as follows. Usually the computing system with the maximum performance has energy consumption near to maximum values, and therefore it sets the value of the maximum energy consumption for the lists during its almost entire service life. Here, the most prominent example are Tianhe-2/Tianhe-2A supercomputers.

Meanwhile, for the computing systems with minimal energy consumption and performance the time of presence in the TOP500_Lists is mainly determined by growth rate of the entry threshold value of R_{max}[5].

[5] On average, energy consumption was not reported for 250–300 systems out of 500, what evidently complicates the analysis; one can assume technical difficulties in the reliable measuring of this parameter but by and large we have no explanation for this fact, because, e. g., List#49 contains data for all systems.

Table 1. Median values of basic characteristic for TOP500 systems

		TOP500_List 41 June/2013	TOP500_List 63 June/2024
Total Cores, N_{cores}	TOP	786 432	7 630 848
	HIGH	186 368	1 305 600
	BASE	16 388	79 524
Maximal Performance, R_{max}, TFlop/s	TOP	17 173.2	561 200
	HIGH	2 897	98 510
	BASE	143.4	3 430
Power, P, kW	TOP	8 209	26 343
	HIGH	2 301	6 316
	BASE	431.3	798.3
Energy Efficiency, EE, GFlop/s per W	TOP	2.14	39.54
	HIGH	1.15	25.44
	BASE	0.49	6.01
$N_{cores}(TOP)/N_{cores}(BASE)$		48	96
$R_{max}(TOP)/R_{max}(BASE)$		120	164
$P(TOP)/P(BASE)$		19	33
$EE(TOP)/EE(BASE)$		4.37	6.58

Table 2. Growth factors of basic characteristics for TOP500 systems between two time snapshots: List#41 [June, 2013] and List#63 [June, 2024]

		Growth
Total Cores, N_{cores}	TOP	9.7
	HIGH	7.0
	BASE	4.8
Maximal Performance, R_{max}, TFlop/s	TOP	32.68
	HIGH	34.0
	BASE	23.92
Power, P, kW	TOP	3.2
	HIGH	2.7
	BASE	1.8
Energy Efficiency, EE, GFlop/s per W	TOP	18.5
	HIGH	22.1
	BASE	12.2
		Gap Growth
$N_{cores}(TOP)/N_{cores}(BASE)$		2.0
$R_{max}(TOP)/R_{max}(BASE)$		1.4
$P(TOP)/P(BASE)$		1.7
$EE(TOP)/EE(BASE)$		1.5

The data in Tab. 1 and Tab. 2 show that computing systems from different classes evolved differently; however, some general patterns common for all classes can be identified. For example, the comparison of growth factors for number of cores in computing system and energy consumption shows that, on average, energy consumption per computing core has decreased: more for the TOP and HIGH class systems and less for the BASE class systems. For example, theoretical growth in $EE(TOP)$ can be roughly estimated as $32.68/3.2 = 10.2$, but empirical growth is 18.5, i.e., 1.8 times more. For $EE(BASE)$ these values are $23.92/1.8 = 13.3$ and 12.2, respectively.

The comparison of the growth in the total number of cores with the growth in performance shows that the greatest increase in performance per one core has occurred in the BASE class, $32.68/9.7 = 3.37$ and $34.0/7.0 = 4.85$ vs. $23.92/4.8 = 4.98$ (see Tab. 2). Complementing this observation with the contrary tendency for energy consumption per computing core, it can be assumed that the greater rate of the growth in the performance per one core for the BASE class systems in comparison with this growth for the TOP class systems was due to the usage of computational cores from the top-level systems of the previous generations in the present base-level systems because, this hardware is usually already available on the market and can be used widely. While the growth in the performance per one core for the top-level systems is possibly connected with the usage of new types of computing cores, which require more time for their development and production.

Thus, it seems that the growth in the total performance was mainly related with the growth of the total number of cores for the TOP/HIGH class systems and with the growth of the performance per core for the BASE class systems.

Studying the behavior of energy efficiency, EE, we can make the following observations (see Fig. 5a):

- maximum energy efficiencies for the computing systems from all three classes were relatively close, which indicates, the simultaneous emergence of the most energy efficient platforms in all three classes (green lines on Fig. 3 and dotted lines on Fig. 5a);
- minimum energy efficiency of the computing systems of a higher performance class was close to average energy efficiency of a lower performance class what probably reflects the natural "aging" of computing systems when less energy-efficient computing systems are replaced by more energy-efficient ones and pass into a lower performance class.

Distribution of resource usage efficiency, RUE, remained almost without significant changes except for the sharp drop in the minimum values in the last few years. This is probably caused by an appearance in the Top500_Lists of computing systems not initially intended for massively parallel calculations, or perhaps due to the fact that HPL test benchmarking was executed not using all computing resources available in such systems (see Fig. 5b).

To summarize, it can be argued that the gap in performance between the TOP and BASE class systems was mainly widening due to the usage in the TOP class systems of large number of increasingly more powerful computing equipment having increasingly greater energy efficiency. No matter how the facts listed above might be trivial, taken together, they show the degree of heterogeneity in the evolution of computing systems from TOP500_Lists.

The trends of the basic characteristics presented above show that the maximum/minimum values of the characteristics have the greatest variability and the tendency to sharp transitions that are difficult to predict, so more reliable information can be obtained by considering the median values. In addition, it should be noted that extrapolation of derived quantities such as, for example, energy efficiency, EE, has an even lower degree of reliability compared to extrapolation of primary (i. e., measured) characteristics, since derived quantities depend on implementation of not one, but several (at least two) random variables.

On the basis of the presented information, we try to predict the energy consumption of a typical computing system from three main classes in 5 and 10 years, as well as what the lowest value of the energy consumption of a system, the computational performance of which corresponds to the threshold level to enter to TOP500_List.

Our choice of a forecast horizon is based on the Bell's empirical law [8], which states that complete mass replacement of computing technologies occurs within 10 years on average.

The data from the TOP500_Lists on computing performance and energy consumption show that, as a rule, changes of these two quantities had complex abrupt nature, so for their extrapolation we used the simplest linear extrapolation designed first of all to highlight the general global trend.

In case of the TOP or HIGH classes, such approximation seems quite sufficient. In case of our particular interest, i. e., for the data on computational performance and energy consumption of the BASE and ENTRY class systems, it seems reasonable for the sake of more accurate description to take the nonlinear component into account in addition to the global linear trend (see Fig. 7).

Thus, for quantities $R = Max(R_{max}), R_{max}(TOP), R_{max}(HIGH)$, in TFlop/s, the following formula was used:

$$Y = A \cdot X + B, \tag{1}$$

where $Y = \log_{10}(R)$ and X is TOP500_List number, and similarly, for quantities $P = Max(P), P(TOP), P(HIGH)$, in kW, the following formula was used:

$$\tilde{Y} = A \cdot X + B, \tag{1'}$$

where $\tilde{Y} = \log_{10}(P)$ and X is TOP500_List number.

To take into account the nonlinear effects having local nature, a nonlinear term was added to the linear trend (1/1'):

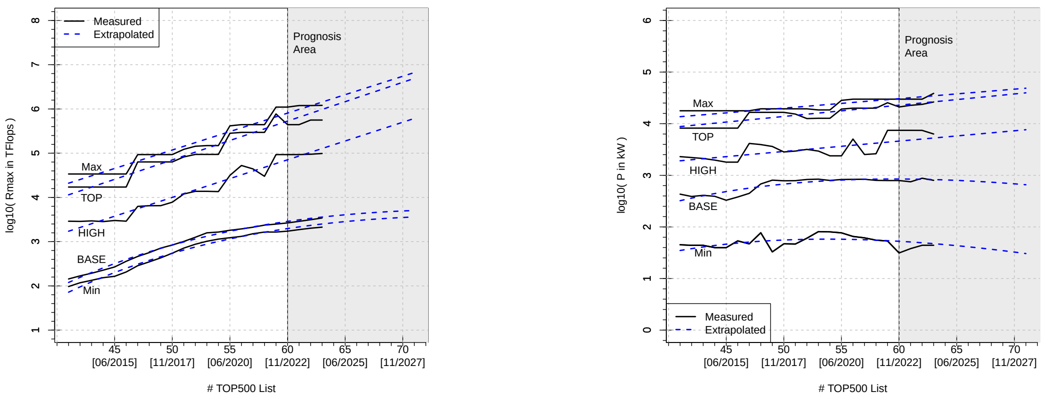
$$Y = C \cdot \log_{10}(X) + D \cdot X + E, \text{ and} \tag{2}$$

$$\tilde{Y} = C \cdot \log_{10}(X) + D \cdot X + E, \text{ respectively.} \tag{2'}$$

We used the least squares method to obtain the coefficient values for all our models with data from Lists#41–60 as training data and data from Lists#61–63 as test data (see Tab. 3).

In Tab. 4 and Tab. 5 the results of predicting the values of R_{max} and P using the formulas (1–2) are given. Figure 8 shows the energy efficiency values calculated as the ratio R_{max}/P, where R_{max} and P were estimated by formulas (1) and (2), along with the corresponding measured values. The dotted lines in Fig. 8 indicate the hypothetical boundaries of energy efficiency (the red line is for the minimum and the green line is for the maximal values) estimated by the following formulae:

$$Max(EE)_{estimated} = \left(f_{pol}(X) \cdot f_{exp}(X)\right)^{1/2},$$
$$f_{pol}(X) = 458.5 - 20.3X + 0.2269X^2,$$
$$f_{exp}(X) = 10^{(0.064X - 2.13)};$$
$$Min(EE)_{estimated} = 0.014X - 0.62, \text{ where } X \text{ is TOP500\_List number.}$$



(a) Maximal performance, R_{max} (b) Power, P

Figure 7. Measured and extrapolated values for basic characteristics of TOP500 systems

Table 3. The coefficients for models (1/1') and (2/2')

	A	**B**	**C**	**D**	**E**
$Max(R_{max})$	0.08337	0.90365	—	—	—
$R_{max}(TOP)$	0.08773	0.46187	—	—	—
$R_{max}(HIGH)$	0.08499	−0.24883	—	—	—
$R_{max}(BASE)$	—	—	24.4634	−0.1402	−31.6292
$Min(R_{max})$	—	—	25.2348	−0.1439125	−32.9403375
$Max(P)$	0.01406	3.61809	—	—	—
$P(TOP)$	0.02465	2.89387	—	—	—
$P(HIGH)$	0.01878	2.52791	—	—	—
$P(BASE)$	—	—	15.6505	−0.1141	−18.0546
$Min(P)$	—	—	15.0905	−0.1219405	−17.79493

To derive this formulae, we performed a standard statistical trial-error procedure including:

1. the choice basis function to cope with nonlinearity;
2. the least-square procedure.

The similar behavior of the calculated energy efficiency for all three performance classes can imply the reasonable correctness of the values of R_{max} and P obtained by formulae (1) and (2). Therefore, it can be assumed that model (2) for $Min(R_{max})$ and $Min(P)$ also correctly describes the behavior of R_{max} and P for systems of the entry level. As the main result we could suppose that energy consumption for the computing system with the performance of the TOP500 entry threshold will decrease in the next 5–10 years (see Tab. 5).

Conclusion

1. The emphasis placed in this work on the difference between the top-level and base-level systems made it possible not only to show the heterogeneity of the development of the HPC area, but also served as a methodological basis for identifying the reasons for such heterogeneous development; whereas the knowledge of such causes and relationships between various characteristics allows to build more accurate models to describe their behavior.

Table 4. Measured and predicted by formula (1) and (2) maximal performance for TOP500 systems

TOP500 List	Maximal Performance, R_{max}, TFlop/s				
	Max(R_{max})	R_{max}(TOP)	R_{max}(HIGH)	R_{max}(BASE)	Min(R_{max})
	Measured				
61	1 194 000	442 010	93 015	2 878	1 872
62	1 194 000	561 200	94 640	3 131	2 015
63	1 206 000	561 200	98 510	3 430	2 130
	Predicted				
61	975 483.7	650 728.8	86 210.47	3 117.726	2 154.970
62	1 181 925.2	796 397.7	104 845.55	3 360.430	2 332.026
63	1 432 056.0	974 675.4	127 508.75	3 599.048	2 507.113
64	1 735 122	1 192 861	155 070.8	3 830.921	2 678.261
65	2 102 326	1 459 890	188 590.6	4 053.448	2 843.518
70	5 489 718	4 008 390	501 729.9	4 945.058	3 519.391
	Relative Errors, %				
61	−18	47	−7	8	15
62	−1	42	11	7	16
63	19	74	29	5	18

Table 5. Measured and predicted by formula (1) and (2) energy consumption for TOP500 systems

TOP500 List	Power, P, kW				
	Max(P)	P(TOP)	P(HIGH)	P(BASE)	Min(P)
	Measured				
61	29 899.2	22 703	7 438	749	38
62	29 899.2	23 695	7 421	877.87	44.07
63	38 698.4	26 343	6 316	798.27	44.07
	Predicted				
61	31 905.1	23 997.7	4 814.93	844.51	51.08
62	33 285.1	25 232.5	5 042.43	837.58	49.30
63	34 724.8	26 530.8	5 280.68	827.33	47.40
64	36 226.8	27 895.9	5 530.19	813.99	45.40
65	37 793.8	29 331.2	5 791.49	797.81	43.32
70	46 705.7	37 694.7	7 295.25	684.08	32.56
	Relative Errors, %				
61	7	6	−35	13	34
62	11	6	−32	−5	12
63	−10	1	−16	4	8

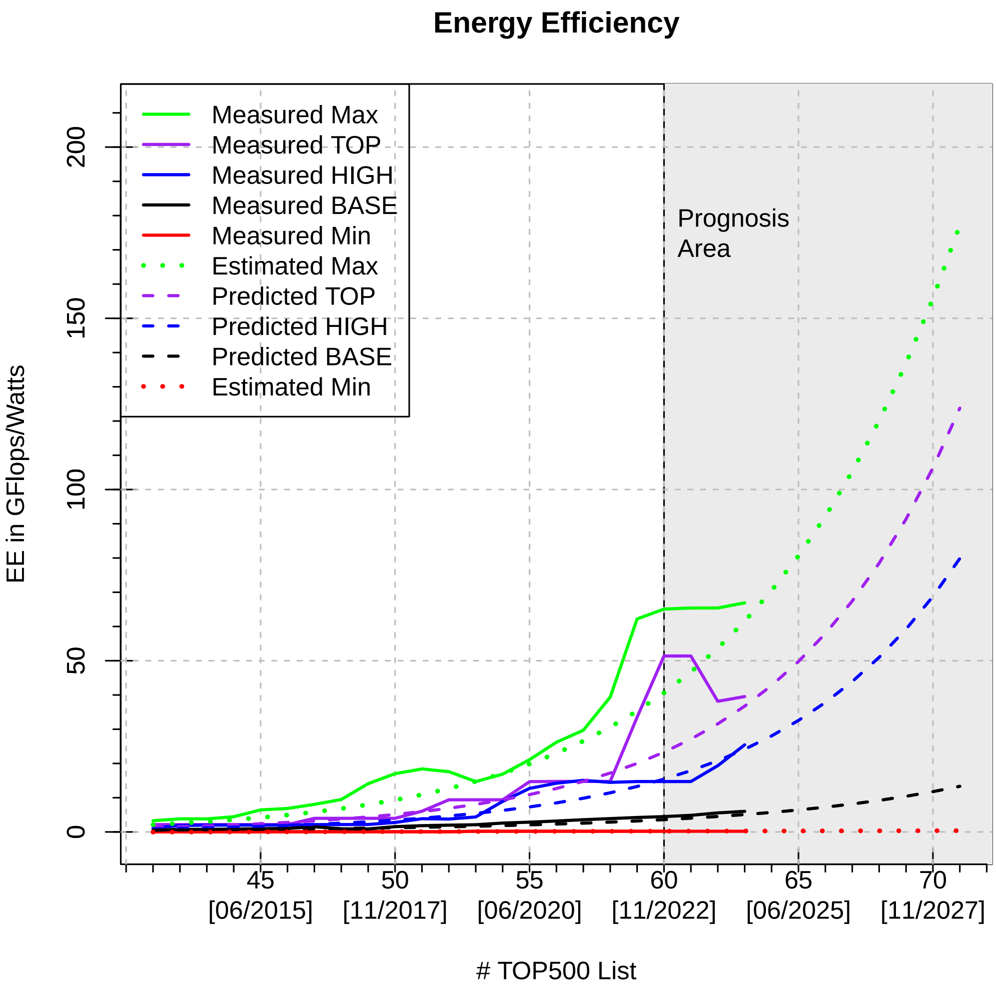


Figure 8. Measured and predicted values along with hypothetical estimations of energy efficiency for computing systems from TOP500_LIST

2. The proposed procedure for the estimation of energy consumption of Rank_500 system allowed us to make prognosis for the near future. Despite the increase in the threshold value of R_{max} required to include a computing system in TOP500_LIST, the total cost of the energy entry ticket will decrease due to the rapid growth in the energy efficiency of computing equipment. According to our estimations, it will decrease by 2 times in the next 5–10 years. We expect this trend to continue in the next few years, as adoption of thin-cored, vector and tensor-operation architectures for general numerical simulation. If the trend continues for a few years, we could install TOP500 system in facilities such as International Space Station or Antarctic research sites (under 10 kW budget).
3. The decision to limit ourselves to the consideration of only quantitative characteristics, related directly to the functioning of computing systems, allowed us to carry out the analysis at a relatively simple level. However, a more complicated approach based on the inclusion qualitive characteristics describing computing system design (for example, processor/co-processor model, type of interconnect and so on) into consideration can increase the accuracy of the forecast provided that machine learning methods are used that make it possible to uniformly take into account both quantitative and qualitative characteristics of the objects being studied.

Acknowledgements

S.S.K. is especially grateful to Maria S. Timoshik for her incomparable help with the translation of the article.

References

1. Abramov, N.S., Abramov, S.M.: Top500 list overview. Supercomputing Frontiers and Innovations 10(2), 4–17 (2023). https://doi.org/10.14529/jsfi230201

2. Austin, B., Daley, C.S., Doerfler, D., *et al.*: A metric for evaluating supercomputer performance in the era of extreme heterogeneity. In: 2018 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems, PMBS@SC 2018, Dallas, TX, USA, November 12, 2018. pp. 63–71. IEEE (2018). https://doi.org/10.1109/PMBS.2018.8641549

3. Davis, N.E., Robey, R.W., Ferenbaugh, C.R., *et al.*: Paradigmatic shifts for exascale supercomputing. J. Supercomput. 62(2), 1023–1044 (2012). https://doi.org/10.1007/S11227-012-0789-3

4. Khan, A., Sim, H., Vazhkudai, S.S., *et al.*: An analysis of system balance and architectural trends based on Top500 supercomputers. In: Hwang, S., Yeom, H.Y. (eds.) HPC Asia 2021: The International Conference on High Performance Computing in Asia-Pacific Region, Virtual Event, Republic of Korea, January 20–21, 2021. pp. 11–22. ACM (2021). https://doi.org/10.1145/3432261.3432263

5. Kogge, P., Borkar, S., Campbell, D., *et al.*: Exascale computing study: Technology challenges in achieving exascale systems. Defense Advanced Research Projects Agency Information Processing Techniques Office (DARPA IPTO), Techinal Representative 15 (01 2008)

6. Kogge, P.M., Dally, W.J.: Frontier vs the exascale report: Why so long? and are we really there yet? In: IEEE/ACM International Workshop on Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems, PMBS@SC 2022, Dallas, TX, USA, November 13–18, 2022. pp. 26–35. IEEE (2022). https://doi.org/10.1109/PMBS56514.2022.00008

7. Kogge, P.M., Dysart, T.J.: Using the TOP500 to trace and project technology and architecture trends. In: Lathrop, S.A., Costa, J., Kramer, W. (eds.) Conference on High Performance Computing Networking, Storage and Analysis, SC 2011, Seattle, WA, USA, November 12–18, 2011. pp. 28:1–28:11. ACM (2011). https://doi.org/10.1145/2063384.2063421

8. Meuer, H.W.: The TOP500 project: Looking back over 15 years of supercomputing experience. Inform. Spektrum 31(3), 203–222 (2008). https://doi.org/10.1007/S00287-008-0240-6