Recent Trends in the Marketplace of High Performance Computing

Erich Strohmaier[1], Jack J. Dongarra[2], Hans W. Meuer[3], and Horst D. Simon[4]

High Performance Computing, HPC Market, Supercomputer Market, HPC technology, Supercomputer market, Supercomputer technology

In this paper we analyze major recent trends and changes in the High Performance Computing (HPC) market place. The introduction of vector computers started the area of 'Supercomputing'. The initial success of vector computers in the seventies was driven by raw performance. Massive Parallel Systems (MPP) became successful in the early nineties due to their better price/performance ratios, which was enabled by the attack of the 'killer-micros'. The success of microprocessor based SMP concepts even for the very high-end systems, was the basis for the emerging cluster concepts in the early 2000s. Within the first half of this decade clusters of PC's and workstations have become the prevalent architecture for many HPC application areas on all ranges of performance. However, the Earth Simulator vector system demonstrated that many scientific applications could benefit greatly from other computer architectures. At the same time there is renewed broad interest in the scientific HPC community for new hardware architectures and new programming paradigms. The IBM BlueGene/L system is one early example of a shifting design focus for large-scale system. The DARPA HPCS program has the declared goal of building a Petaflops computer system by the end of the decade using novel computer architectures.

1. Introduction

"The Only Thing Constant Is Change" — Looking back on the last four decades this

[1] CRD, Lawrence Berkeley National Laboratory, 50F1603, Berkeley, CA 94720; e-mail: estrohmaeir@lbl.gov

[2] Computer Science Department, University of Tennessee, Knoxville, TN 37996 Mathematical Science Section, Oak Ridge National Lab., Oak Ridge, TN 37831; e-mail: dongarra@cs.utk.edu

[3] Computing Center, University of Mannheim, D-68131 Mannheim, Germany ; e-mail: meuer@rz.uni-mannheim.de

[4] Lawrence Berkeley National Laboratory, 50B-4230, Berkeley, CA 94720; e-mail: simon@nersc.gov

seems certainly to be true for the market of High-Performance Computing systems (HPC). This market was always characterized by a rapid change of vendors, architectures, technologies and the usage of systems [1]. Despite all these changes the evolution of performance on a large scale however seems to be a very steady and continuous process. Moore's Law which states that circuit density and in return processor performance doubles every 18 month is often cited in this context [2]. If we plot the peak performance of various computers of the last six decades in Fig. 1 which could have been called the 'supercomputers' of their time [3,4] we indeed see how well this law holds for almost the complete lifespan of modern computing. On average we see an increase in performance of two orders of magnitudes every decade.

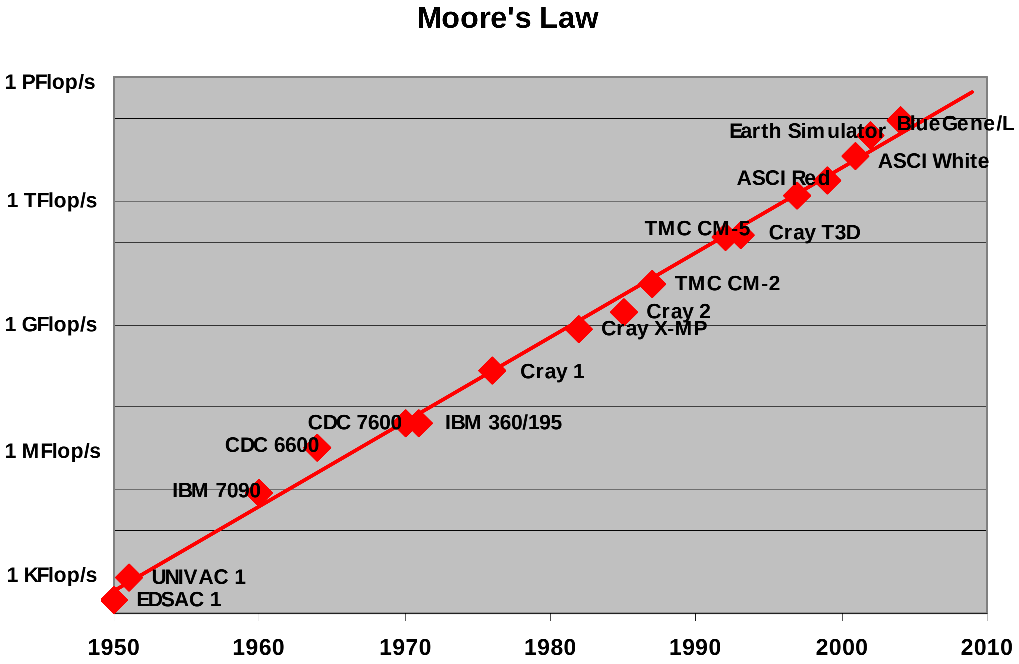

Fig. 1. Performance of the fastest computer systems for the last six decades compared to Moore's Law.

In this paper we analyze recent major trends and changes in the HPC market. For this we focus on systems, which had at least some commercial relevance. This paper extends a previous analysis of the HPC market in [1]. Historical overviews with different focus can be found in [5,6]. Section 2 summarizes our earlier finding [1]. Section 3 analyzes the trend in the first half of this decade and section 4 projects our finding into the future.

The initial success of vector computers in the seventies was driven by raw performance. The introduction of this type of computer systems started the area of 'Supercomputing'. In the eighties the availability of standard development environments and of application software packages became more important. Next to performance these criteria determined the success of MP vector systems especially at industrial customers. MPPs became successful in the early nineties due to their better price/performance ratios, which was enabled by the attack of the 'killer-micros'. In the lower and medium market segments the MPPs were replaced by microprocessor based SMP systems in the middle of the nineties. Towards the end of the nineties only the companies which had entered the emerging markets for massive parallel database servers and financial applications attracted enough business volume to be able to support the hardware development for the numerical high end computing market as well. Success in the traditional floating point intensive engineering applications was no longer sufficient for survival in the market. The success of microprocessor based SMP concepts even for the very high-end systems was the basis for the emerging cluster concepts in the early 2000s. Within the first half of this decade clusters of PC's and workstations have become the prevalent architecture for many application areas in the TOP500 on all ranges of performance. However, the Earth Simulator vector system demonstrated that many scientific applications can benefit greatly from other computer architectures. At the same time there is renewed broad interest in the scientific HPC community for new hardware architectures and new programming paradigms. The IBM BlueGene/L system is one early example of a shifting design focus for large-scale system. The DARPA HPCS program has the declared goal of building a Petaflops computer system by the end of the decade.

2. A Short History of Supercomputers until 2000

In the second half of the seventies the introduction of vector computer systems marked the beginning of modern Supercomputing. These systems offered a performance advantage of at least one order of magnitude over conventional systems of that time. Raw performance was the main if not the only selling argument. In the first half of the eighties the integration of vector system in conventional computing environments became more important. Only the manufacturers which provided standard programming environments, operating systems and key applications were successful in getting industrial customers and survived. Performance was mainly increased by improved chip technologies and by producing shared memory multi processor systems.

Fostered by several Government programs massive parallel computing with scalable systems using distributed memory became the center of interest at the end of the eighties. Overcoming the hardware scalability limitations of shared memory systems was the main goal for their development. The increased performance of standard microprocessors after the RISC revolution together with the cost advantage of large-scale productions formed the basis for the "Attack of the Killer Micros". The transition from ECL to CMOS chip technology and the usage of "off the shelf" micro processors instead of custom designed processors for MPPs was the consequence.

The traditional design focus for MPP systems was the very high end of performance. In the early nineties the SMP systems of various workstation manufacturers as well as the IBM SP series, which targeted the lower and medium market segments, gained great popularity. Their price/performance ratios were better due to the missing overhead in the design for support of the very large configurations and due to cost advantages of the larger production numbers. Due to the vertical integration of performance it was no longer economically feasible to produce and focus on the highest end of computing power alone. The design focus for new systems shifted to the market of medium performance systems.

The acceptance of MPP systems not only for engineering applications but also for new commercial applications especially for database applications emphasized different criteria for market success such as the stability of system, continuity of the manufacturer and price/performance. Success in commercial environments became a new important requirement for a successful Supercomputer manufacturing business towards the end of the nineties. Due to these factors and the consolidation in the number of vendors in the market hierarchical systems built with components designed for the broader commercial market did replace homogeneous systems at the very high end of performance. The marketplace adopted clusters of SMPs readily, while academic research focused on clusters of workstations and PCs.

3. 2000-2005: Clusters, Intel Processors, and the Earth-Simulator

In the early 2000's Clusters built with off-the-shelf components gained more and more attention not only as academic research objects but also as computing platforms for end-users of HPC computing systems. By 2004 these clusters represent the majority of new systems on the TOP500 in a broad range of application areas. One major consequence of this trend was the rapid rise in the utilization of Intel processors in HPC systems. While virtually absent in the high end at the beginning of the decade, Intel processors are now used in the majority of HPC systems. Clusters in the nineties were mostly self-made system designed and built by small groups of dedicated scientists or application experts. This changed rapidly as soon as the market for clusters based on PC technology matured. Nowadays the large majority of TOP500-class clusters are manufactured and integrated by either a few traditional large HPC manufacturers such as IBM or HP or numerous small, specialized integrators of such systems.

In 2002 a system called "Computnik" with a quite different architecture, the Earth Simulator, entered the spotlight as new #1 system on the TOP500 and it managed to take the U.S. HPC community by surprise. The Earth Simulator build by NEC is based on the NEC vector technology and showed unusual high efficiency on many applications. This fact invigorated discussions about future architectures for high-end scientific computing systems.

3.1. Explosion of Cluster Based Systems

At the end of the nineties clusters were common in academia, but mostly as research objects and not primarily as general purpose computing platforms for applications. Most of these clusters were of comparable small scale and as a result the November 1999 edition of the TOP500 listed only seven cluster systems. This changed dramatically as industrial and commercial customers started deploying clusters as soon as applications with less stringent communication requirements permitted them to take advantage of the better price/performance ratio -roughly an order of magnitude- of commodity based clusters. At the same time all major vendors in the HPC market started selling this type of cluster to their customer base. In November 2004 clusters are the dominant architectures in the TOP500 with 294 systems at all levels of performance (see Fig 2). Companies such as IBM and Hewlett-Packard sell the majority of these clusters and a large number of them are installed at commercial and industrial customers.

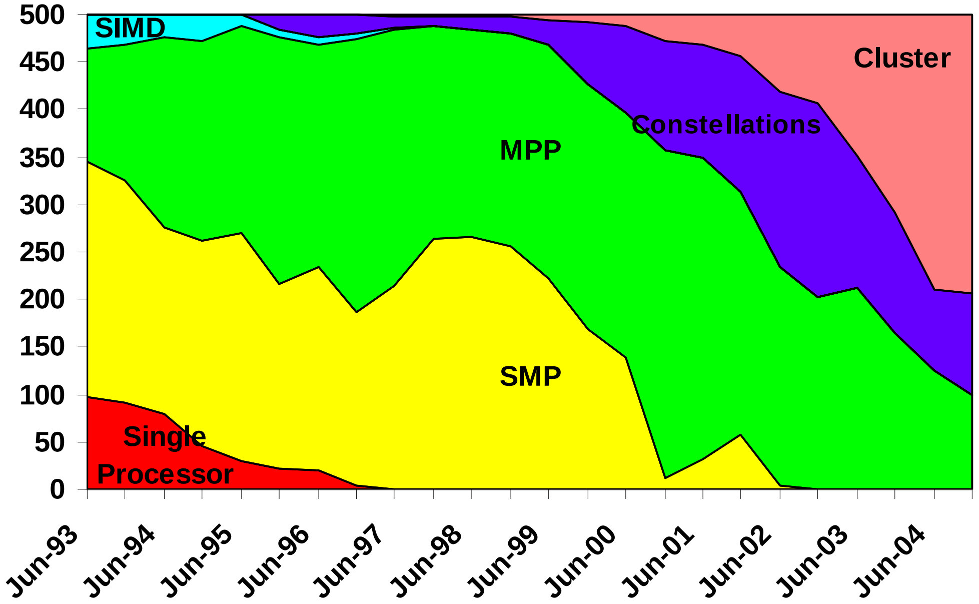

Fig. 2. Main Architectural Categories seen in the TOP500.

In addition, there still is generally a large difference in the usage of clusters and their more integrated counterparts: clusters are mostly used for capacity computing, while the integrated machines are primarily used for capability computing. The largest supercomputers are used for capability or turnaround computing where the maximum processing power is applied to a single problem. The goal is to solve a larger problem, or to solve a single problem in a shorter period of time. Capability computing enables the solution of problems that cannot otherwise be solved in a reasonable period of time (for example, by moving from a 2D to a 3D simulation, using finer grids, or using more realistic models). Capability computing also enables the solution of problems with real-time constraints (e.g., predicting weather). The main figure of merit is time to solution. Smaller or cheaper systems are used for capacity computing, where smaller problems are solved. Capacity computing can be used to enable parametric studies or to explore design alternatives; it is often needed to prepare for more expensive runs on capability systems. Capacity systems will often run several jobs simultaneously. The main figure of merit is sustained performance per unit cost. Traditionally, vendors of large supercomputer systems have learned to provide for this first mode of operation as the precious resources of their systems were required to be used as efficiently and effectively as possible. By contrast, Beowulf clusters are mostly operated through the Linux operating system (a small minority using Microsoft Windows). These operating systems do not have sophisticated tools available to use a cluster efficiently or effectively for capability computing. However, as clusters become on average both larger and more stable, there is a trend to use them also as computational capability servers.

There are a number of choices of communication networks available in clusters. Of course 100 Mb/s Ethernet or Gigabit Ethernet is always possible, which is attractive for economic reasons, but has the drawback of a high latency (~ 100 µs). Alternatively, there are for instance networks that operate from user space, like

Myrinet, Infiniband, and SCI. The communication speeds of these networks are more or less on a par with some integrated parallel systems. So, possibly apart from the speed of the processors and of the software that is provided by the vendors of traditional integrated supercomputers, the distinction between clusters and this class of machines becomes rather small and will without a doubt decrease further in the coming years.

3.2. Intel-ization of the Processor Landscape

The HPC community had started to use commodity components in large numbers in the nineties already. MPPs and Constellations (Cluster of SMP) typically used standard workstation microprocessors even though custom interconnect systems might still be used. There was, however, one big exception: virtually nobody used Intel microprocessors. Lack of performance and the limitations of a 32-bit processor design were the main reasons for this. This changed with the introduction of the Pentium III and especially in 2001 with the Pentium 4, which featured greatly improved memory performance due to its redesigned front-side bus and full 64-bit floating point support. The number of systems in the TOP500 with Intel processors exploded from only 6 in November 2000 to 318 in November 2004 (Fig. 3).

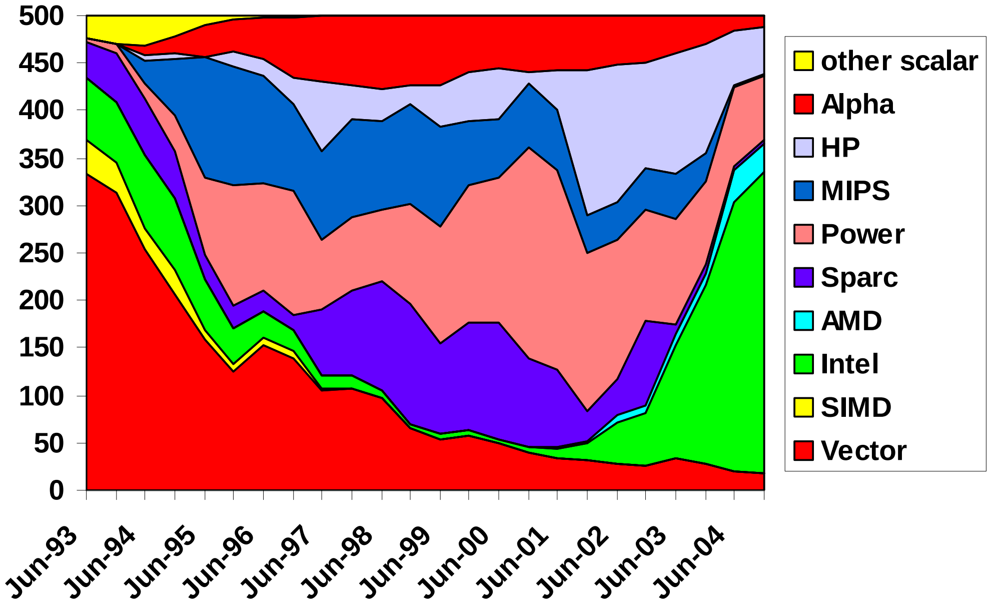

Fig. 3. Main Processor Families seen in the TOP500.

3.3. The Impact of the Earth-Simulator

The ES project was conceived, developed, and implemented by Dr. Hajime Miyoshi who is regarded as the Seymour Cray of Japan. Unlike his peers, he seldom attended conferences or gave public speeches. However, he was well known within the HPC community in Japan for his involvement in the development of the first Fujitsu supercomputers in Japan, and later on of the Numerical Wind Tunnel (NWT) at NAL. In 1997 he took up his post as the director of the Earth Simulator Research &

Development Center (ESRDC) and led the development of the 40 Tflop/s Earth Simulator, which would serve as a powerful computational engine for global environmental simulation. The machine was completed in February 2002 and presently the entire system is working as an end user service.

The launch of the Earth Simulator created a substantial amount of concern in the U.S. that it had lost the leadership in high performance computing. While there was certainly a loss of national pride for the U.S. not to be first on a list of the world's fastest supercomputers, this is certainly not the same as having lost leadership in the field in general. However, it is important to understand the set of issues that surrounded the concerns in the US about the sudden emergence of the ES as the number one system. The development of the ES represents a large investment (approximately $500M, including a special facility to house the system) and a large commitment over a long period of time. While the U.S. has made an even larger investment in HPC, for example in the ASC program in DOE, the funds were not spent on a single platform. Other important differences are:

- ES was developed for basic research and is shared internationally, whereas the largest systems in the U.S. are developed for national security applications and consequently have restricted access.

- A large part of the ES investment directly supported the vendor NEC and the development of their SX-6 technology, which is mostly used for highend engineering and science applications In contrast in the U.S. the approach of the last decade was generally not to provide any direct support for HPC vendors, but to leverage off the commerically successful technology used for business applications.

- ES uses custom vector processors; almost all U.S. high-end systems use commodity processors.

- The ES software technology largely originates from abroad, although it is often modified and enhanced in Japan. For example, significant ES codes were developed using a Japanese enhanced version of HPF. Virtually all software used on high end platforms in the U.S. were developed by U.S. research programs.

These significant differences led in the U.S. to a vigorous debate about the relative merits of the two approaches, and to renewed interest in national programs to revitalize high-end computing (HECRTF) [7]. This debate also led to a NRC study on "The Future of Supercomputing" [8].

Surprisingly, the Earth Simulator's number one ranking on the TOP500 list was not a matter of national pride in Japan. In fact, there is considerable resentment of the Earth Simulator in some sectors of research communities in Japan. Some Japanese researchers feel that the ES is too expensive and drains critical resources from other science and technology projects. Due to the continued economic crisis in Japan and the large budget deficits, it is getting more difficult to justify government projects of this kind.

3.4. New Architectures on the Horizon

Interest in novel computer architectures has always been large in the HPC community, which comes at little surprise as this field was borne and continues to thrive on technological innovations. Some of the concerns of recent years were the ever increasing space and power requirements of modern commodity based supercomputers. In the BlueGene/L development, IBM addressed these issues by designing a very power and space efficient system. BlueGene/L does not use the latest commodity processors available but computationally less powerful and much more power efficient processor versions developed mainly not for the PC and workstation market but for embedded applications. Together with a drastic reduction of the available main memory this leads to a very dense system. To achieve the targeted extreme performance level and unprecedented number of these processors (up to 128,000) are combined using several specialized interconnects. There was and is considerable doubt whether such a system would be able to deliver the promised performance and would be usable as a general purpose system. First results of the current beta-System are very encouraging and the one-quarter size beta-System of the future LLNL system was able to claim the number one spot on the November 2004 TOP500 list.

Contrary to the progress in hardware development, there has been little progress, and perhaps regress, in making scalable systems easy to program. Software directions that were started in the early 90's (such as CM-Fortran and High-Performance Fortran) were largely abandoned. The payoff to finding better ways to program such systems and thus expand the domains in which these systems can be applied would appear to be large.

The move to distributed memory has forced changes in the programming paradigm of supercomputing. The high cost of processor-to-processor synchronization and communication requires new algorithms that minimize these operations. The structuring of an application for vectorization is seldom the best parallelization strategy for these systems. Moreover, despite some research successes in this area, without some guidance from the programmer, compilers are generally able neither to detect enough of the necessary parallelism, nor to reduce sufficiently the inter-processor overheads. The use of distributed memory systems has led to the introduction of new programming models, particularly the message passing paradigm, as realized in MPI, and the use of parallel loops in shared memory subsystems, as supported by OpenMP. It also has forced significant reprogramming of libraries and applications to port onto the new architectures. Debuggers and performance tools for scalable systems have developed slowly, however, and even today most users consider the programming tools on parallel supercomputers to be inadequate.

All these issues prompted DARPA to start a program for High Productivity Computing Systems (HPCS) with the declared goal to develop a new computer architecture by the end of the decade with high performance and productivity. The performance goal is to install a system by 2009, which can sustain Petaflop/s performance levels on real applications. This should be achieved by the combination of a new architecture designed to be easy programmable and combined with a complete new software infrastructure to make user productivity as high as possible.

4. 2005 and Beyond

Three decades after the introduction of the Cray 1 the HPC market has changed its face quite a bit. It used to be a market for systems clearly different from any other computer systems. Nowadays the HPC market is no longer an isolated niche market for specialized systems. Vertically integrated companies produced systems of any size. Components used for these systems are the same from an individual desktop PC up to the most powerful supercomputers. Similar software environments are available on all of these systems.

Market and cost pressure have driven the majority of customers away from specialized highly-integrated traditional supercomputers towards using clustered systems built using commodity components. The overall market for the very high end systems itself is also relatively small and does not grow strongly if at all. It cannot easily support specialized niche market manufacturers, which poses a problem for customers with applications requiring highly integrated supercomputers. Together with reduced system efficiencies, reduced productivity, and a lack of supporting software-infrastructure, this leads to a strong interest in new computer architectures.

4.1. Consumer and Producer

During the last few years a new trend with respect to the countries using supercomputers is emerging. Globally the number of systems installed in the U.S. increased slightly over time, while the number of systems in Japan decreased. As a producer of HPC systems the U.S. dominates with a market share of about 90%, which actually slowly increased over time. European manufacturers have never played a substantial role in the HPC market at all. Even the introduction of new architectures such as PC clusters has not changed this picture.

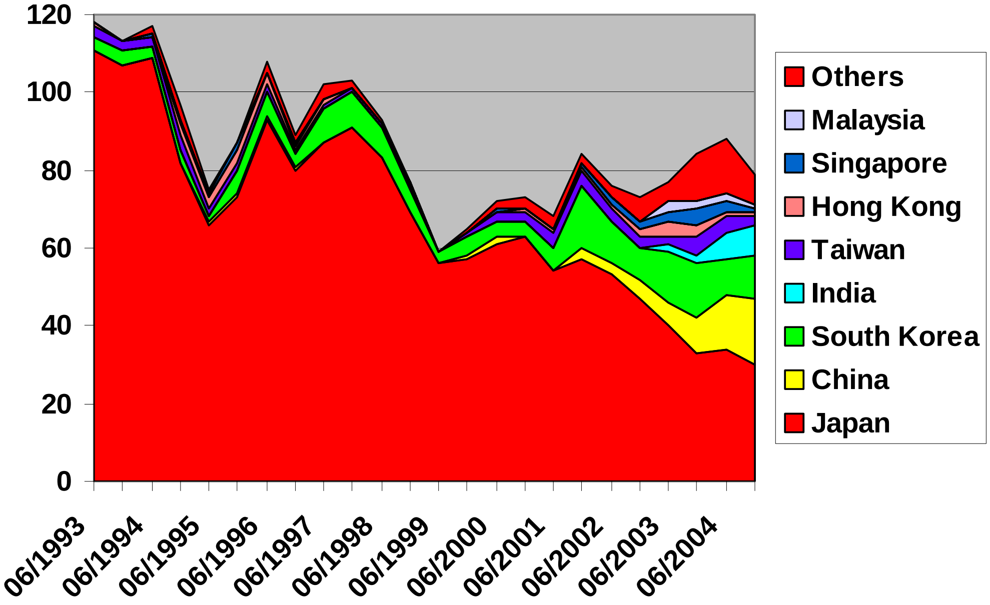

Fig. 4. The consumers of HPC systems in Asia as reflected in the TOP500.

The strongest recent geographical trend is the increasing number of supercomputers being installed in upcoming Asian countries such as China, South Korea and India shown in Fig. 4. While this can be interpreted as a reflection of increasing economical stamina of these countries it also highlights the fact that it is becoming easier for such countries to buy or even build cluster based systems themselves. It is, however, an open question, whether any new Asian manufactures will be able to successfully enter the HPC market. It is interesting, however, to note that the Chinese cluster integrator Lenovo (with two systems on the TOP500 list) just recently acquired IBM's PC business. This hints that Chinese companies such as Dawning and Lenovo, are well positioned for a larger role in the world market for high-end clusters, and could increase their market share in the coming years.

4.2. Performance Growth

While many aspects of the HPC market change quite dynamically over time, the evolution of performance seems to follow quite well some empirical law such as Moore’s law mentioned at the beginning of this article. The TOP500 provides an ideal data basis to verify such an observation. Looking at the computing power of the individual machines present in the TOP500 and the evolution of the total installed performance, we plot the performance of the systems at positions 1, 10, 100 and 500 in the list as well as the total accumulated performance. In Fig. 5 the curve of position 500 shows on the average an increase of a factor of 1.9 per year. All other curves show a growth rate of 1.8 ± 0.05 per year.

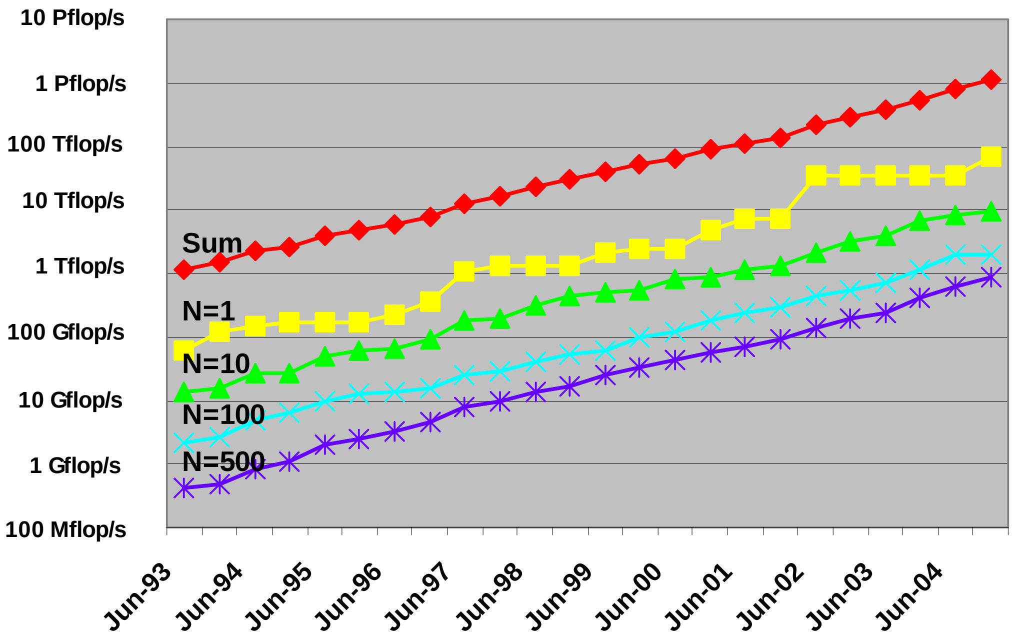

Fig. 5. Overall growth of accumulated and individual performance as seen in the TOP500.

To compare these growth rates with Moore’s Law we now separate the influence of the increasing processor performance and of the increasing number of processor per system on the total accumulated performance. To get meaningful numbers we exclude the SIMD systems from this analysis as they tend to have extremely large numbers of processors with very low processor performance. In Fig. 6 we plot the relative growth of the total number of processors and of the average processor performance defined as the ratio of total accumulated performance by the total processor number. We find that these two factors contribute almost equally to the annual total performance growth factor of 1.80. The number of processors grows with an average growth factor of 1.29 per year. Processor performance increases by a factor of 1.40 compared to the 1.58 of Moore's Law.

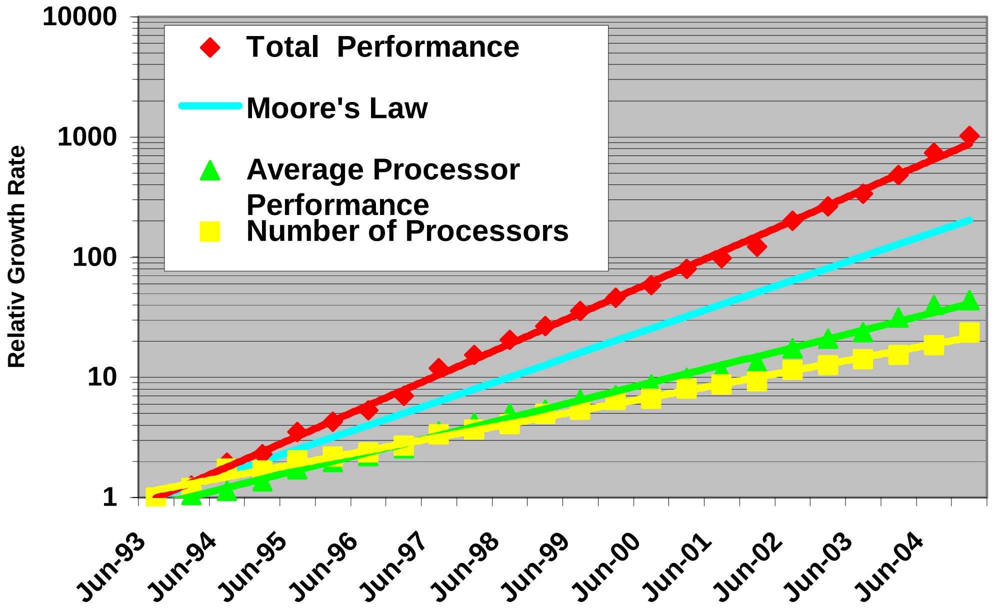

Fig. 6. The growth rate of total accumulated performance, total number of processors and single processor performance for non-SIMD systems as seen in the TOP500.

The average growth in processor performance is lower than we expected. A possible explanation is that during the recoding time of the TOP500 project powerful vector processors got replaced by less powerful super-scalar RISC processors. This effect might be the reason why the TOP500 does not reflect the full increase in RISC performance. The overall growth of system performance is, however, larger than expected from Moore's Law. This results from growth in the two dimensions processor performance and number of processors used.

4.3. Projections

Based on the current TOP500 data, which cover the last twelve years, and the assumption that the current performance development continues for some time to come, one can now extrapolate the observed performance and compare these values with the goals of the mentioned government programs. In Fig. 7 we extrapolate the observed performance values using linear regression on the logarithmic scale. This means that we fit exponential growth to all levels of performance in the TOP500. These simple fitting of the data shows surprisingly consistent results. In 1999 based on a similar extrapolation [1] we expected to have the first 100 TFlop/s system by 2005. We also predicted that by 2005 no system smaller then 1 TFlop/s should be able to make the TOP500 any more. Both of these predictions are basically certain to be fulfilled next year. Extrapolating over another five year period to 2010 we expected to see the first PetaFlops system at about 2009 [1] and our current extrapolation is still the same. This coincides with the declared goal of the DARPA HPCS program.

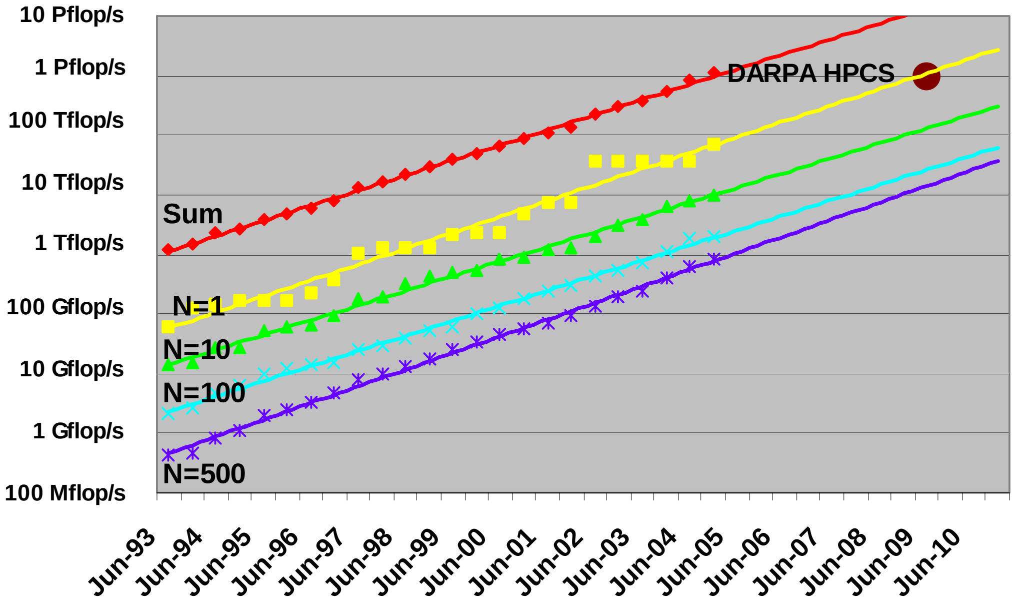

Fig. 7. Extrapolation of recent growth rates of performance seen in the TOP500.

Looking even further in the future we could speculate that based on the current doubling of performance every year the first system exceeding 100 Petaflop/s should be available around or shortly after 2015. Due to the rapid changes in the technologies used in HPC systems there is however again no reasonable projection possible for the architecture of such a system in ten years. The end of Moore's Law as we know it has often been predicted and one day it will come. Whether there might be new technologies such as quantum computing, which would allow us to further extend our computing capabilities is well beyond the capabilities of our simple performance projections. However, even as the HPC market has changed its face several times quite substantially since the introduction of the Cray 1 four decades ago, there is no end in sight for these rapid cycles of re-definition. And we still can say that in the High-Performance Computing Market "The Only Thing Constant Is Change".

5. References

[1] E. Strohmaier, J.J. Dongarra, H.W. Meuer, and H.D. Simon, *The marketplace of high-performance computing*, Parallel Computing 25 (1999) 1517

[2] G. E. Moore, *Cramming more components onto integrated circuits*, Electronics, Volume 38, Number8, April 19, 1965

[3] R. W. Hockney, C. Jesshope, *Parallel Computers II: Architecture, Programming and Algorithms*, Adam Hilger, Ltd., Bristol, United Kingdom, 1988

[4] H. W. Meuer, E. Strohmaier, J. J. Dongarra, and Horst D. Simon, TOP500,

www.top500.org.

[5] G. V. Wilson, Chronology of major developments in parallel computing and supercomputing, www.unipaderborn.de/fachbereich/AG/agmadh/WWW/GI/History/history.txt.gz

[6] P. R. Woodward, Perspectives on Supercomputing, *Computer* (October 1996), no. 10, 99–111.

[7] Federal Plan for High-End Computing: Report of the High-End Computing Revitalization Task Force (HECRTF), May 10, 2004 (available at http://www.itrd.gov/hecrtf-outreach/)

[8] Susan L. Graham, Marc Snir, and Cynthia A. Patterson (editors), Getting up to Speed: The Future of Supercomputing, National Academies Press, 2004.

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.