20 Years Supercomputer Market Analysis

Erich Strohmaier,
Future Technology Group, Lawrence Berkeley National Laboratory, 50A1182,
Berkeley, CA 94720;
e-mail: estrohmaier@lbl.gov
May 2005

Abstract

Since the very beginning of the International Supercomputer Conference (ISC) series, one important focus has been the analysis of the supercomputer market. For 20 years, statistics about this marketplace have been published at ISC. Initially these were based on simple market surveys and since 1993, they are based on the TOP500 project, which has become the accepted standard for such data. We take the occasion of the 20th anniversary of ISC to combine and extend several previously published articles based on these data. We analyze our methodologies for collecting our statistics and illustrate the major developments, trends and changes in the High Performance Computing (HPC) marketplace and the supercomputer market since the introduction of the Cray 1 system.

The introduction of vector computers started the area of modern 'Supercomputing'. The initial success of vector computers in the seventies and early eighties was driven by raw performance. In the second half of 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. 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'. In the lower and medium segments of the market MPPs were replaced by microprocessor based symmetrical multiprocessor systems (SMP) 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 HPC application areas in the TOP500 on all ranges of performance. However, the success of 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. Built with low performance but very low power components, it allows a tight integration of an unprecedented number of processors to achieve surprising performance levels for suitable applications. The DARPA HPCS program has the declared goal of building a PetaFlops computer by the end of the decade using novel computer architectures.

1. Introduction

"The Only Thing Constant Is Change" — Looking back on the last decades this 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. 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 [1]. If we plot the peak performance of various computers, which could have been called the 'supercomputers' of their time [2,3], over the last six decades we indeed see in Fig. 1 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.

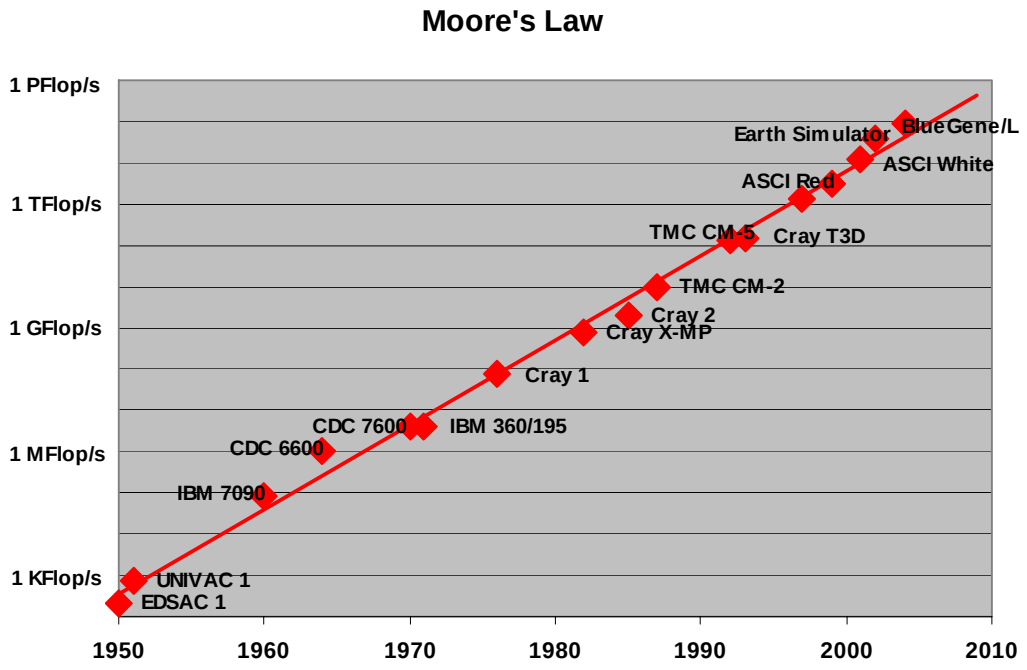

Fig. 1: Performance of the fastest computer systems for the last six decades compared to Moore's Law.

In this paper, we analyze the major trends and changes in the HPC market for the last three decades. For this we focus on systems, which had at least some commercial relevance and illustrate our findings using market data obtained from various sources including the Mannheim Supercomputer Seminar and the TOP500 project. We also analyze the procedures used to obtain these market data and the limits of their usability. This paper extends previous analyses of the HPC market in [4,5,6]. Historical overviews with different focus are found in [7,8].

2. Summary

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 increased mainly due to improved chip technologies and the usage of 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" microprocessors 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 had shifted towards the larger 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.

In the early 2000s, Clusters built with components off the shelf gained more and more attention also with end-users of HPC computing systems. Since 2004, this group of clusters represents the majority of 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 either by a few traditional large HPC manufacturers, such as IBM or HP, or by numerous small, specialized integrators of such systems.

In 2002, a system 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 built by NEC is based on the NEC vector technology and showed unusual high efficiency on many applications. It demonstrated that many scientific applications could benefit greatly from other computer architectures. This fact invigorated discussions about future architectures for high-end scientific computing systems. At the end of 2004, another system built with an entirely different design focus took the #1 spot from the Earth Simulator. IBMs BlueGene/L system is still built with mostly conventional off the shelf components. Its design focuses on building a system with an unprecedented number of processors using a power efficient design with high-density packaging while sacrificing main memory size. The DARPA High Productivity Computing Systems (HPCS) program has the declared goal of building a computer system by the end of the decade, which can sustain PetaFlop/s performance levels on real applications.

3. 1976–1985: The first Vector Computers

If one had to pick one person associated with Supercomputing, it would be without doubt Seymour Cray. Coming from Control Data Corporation (CDC), where he had designed the CDC 6600 series in the sixties, he had started his own company 'Cray Research Inc.' in 1972. The delivery of the first Cray 1 vector computer in 1976 to the Los Alamos Scientific Laboratory marked the beginning of the modern area of 'Supercomputing'. The Cray 1 was characterized by a new architecture, which gave it a performance advantage of more than an order of magnitude over scalar systems at that time. Beginning with this system high-performance computers had a substantially different architecture from mainstream computers. Before the Cray 1, systems, which sometimes were called 'Supercomputer' like the CDC 7600, still had been scalar systems and did not differ in their architecture to this extent from competing mainstream systems. For more than a decade, supercomputer was a synonym for vector computer. Only at the beginning of the nineties would the MPPs be able to challenge or outperform their MP vector competitors.

3.1. Cray 1

The architecture of the vector units of the Cray 1 was the basis for the complete family of Cray vector systems into the nineties including the Cray 2, Cray X-MP, Y-MP, C-90, J-90 and T-90. Common feature was not only the usage of vector instructions and vector register but especially the close coupling of the fast main memory with the CPU. The system did not have a separate scalar unit but integrated the scalar functions efficiently in the vector CPU with the advantage of high scalar computing speed as well. One common remark about the Cray 1 was that it was not only the fastest vector system but it was also the fastest scalar system of its time. The Cray 1 was also a true Load/Store architecture, which was a new concept, which later

entered mainstream computing with the RISC processors. In the X-MP and follow on architecture, three simultaneous Load/Store operations per CPU were supported in parallel from main memory without using caches. This gave the systems exceptionally high memory to register bandwidth and facilitated the ease of use greatly.

The Cray 1 was well accepted in the scientific community and 65 systems were sold until the end of its production in 1984. In the US, the initial acceptance was largely driven by government laboratories and classified sites for which raw performance was essential. Due to its potential, the Cray 1 soon gained great popularity in general research laboratories and at universities.

3.2. Cyber 205

The main competitor for the Cray 1 was a vector computer from CDC, the Cyber 205. This system was based on the design of the Star 100 of which only four systems had been built after its first delivery in 1974. Neil Lincoln designed the Cyber 203 and Cyber 205 systems as memory-to-memory machines not using any registers for the vector units. The system also had separate scalar units. The system used multiple pipelines to achieve high peak performance and a virtual memory in contrast to Cray's direct memory. Due to the memory-to-memory operation, the vector units had rather long startup phases, which allowed it to achieve high performance only on long vectors.

CDC had been the market leader for high performance systems with its CDC 6600 and CDC 7600 models for many years, which gave the company the advantage of a broad existing customer base. The Cyber 205 was first delivered in 1982 and about 30 systems were sold altogether.

3.3. Japanese Vector Systems

At the end of the seventies, the main Japanese computer manufacturers (Fujitsu, Hitachi and NEC) started to develop their own vector computer systems. First models were delivered in late 1983 and mainly sold in Japan. Fujitsu had early decided to sell their vector systems in the USA and Europe through their mainframe distribution partners Amdahl and Siemens. This was the main reason that Fujitsu VP100 and VP200 systems could be found in decent numbers early on in Europe. NEC tried to market their SX1 and SX2 systems by themselves and had a much harder time to find customers outside of Japan. From the beginning, Hitachi had decided not to market the S810 system outside of Japan. Common feature of the Japanese systems were separate scalar and vector units and the usage of large vector registers and multiple pipelines in the vector units. The scalar units were IBM 370 instruction compatible which made the integration of these systems in existing computer centers easy. In Japan, all these systems were well accepted and especially the smaller models were sold in reasonable numbers.

3.4. Vector Multi-Processor

At Cray Research, the next steps to increase performance were not only to increase the performance and efficiency of the single processors but also to build systems with multiple processors. Due to diverging design ideas and emphasis, two design teams worked parallel in Chippewa Falls.

Steve Chen first designed and introduced the Cray X-MP system with two processors in 1982. The enlarged model with four processors was available in 1984. The systems were designed as symmetrical shared memory multi processor systems. The main emphasis of the development was the support of multiple processors. Great effort went into the design of an effective memory access subsystem, which was able to support multiple processors with high bandwidth. While the multiple processors were mainly used to increase the throughput of computing centers, Cray was one of the first to offer a means for parallelization within a user's program using features such as Macrotasking and later on Microtasking.

At the same time, Seymour Cray focused at Cray Research on the development of the Cray 2. His main focus was on advanced chip technology and new concepts in cooling. The first model was delivered in 1985. With its four processors, it promised a peak performance of almost 2 GFlop/s, more than twice as much as a four processor X-MP did. As its memory was built with DRAM technology, the available real main memory reached the unprecedented amount of 2 GByte. This memory size allowed for long running programs not feasible on any other systems.

The Japanese Vector computer manufacturers decided to follow a different technology path. They increased the performance of their single processors by using advanced chip technology and multiple pipelines. Later the Japanese manufacturers announced multiprocessor models typically with two or at most four processors.

3.5. Early Market Growth

Hard data about the early development of the supercomputer market before the first ISC conference (then called "*Mannheim Supercomputer Seminar")* are difficult to find. Access to data preceding the modern Web is in general harder to get but the lack

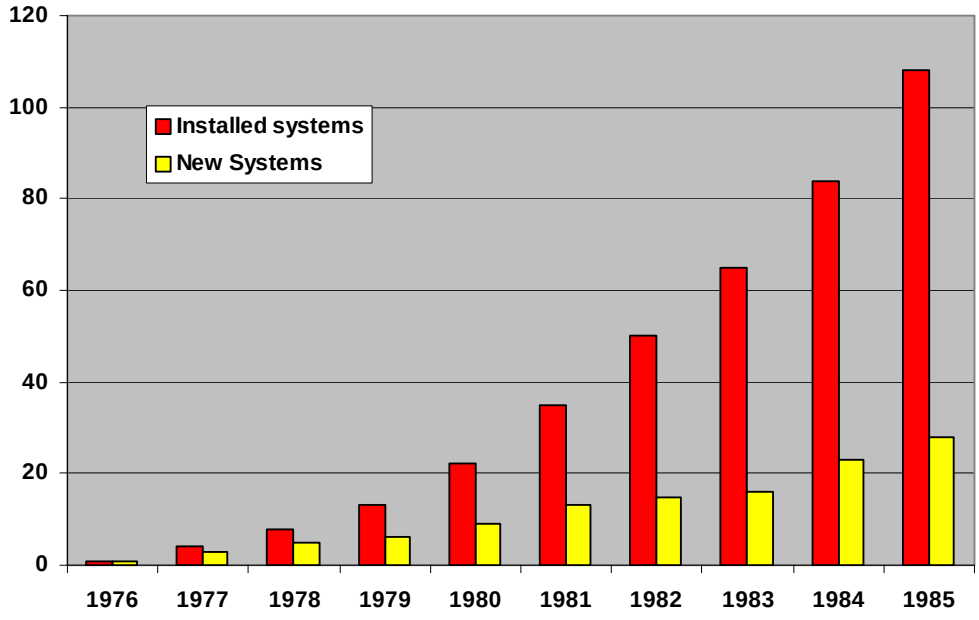

Fig. 2: Number of total installed systems and new delivered systems for Cray before the Mannheim statistics.

of data is in our opinion mostly due to the lack of publicly available data compilations in the first place. Some data from the Charles Babbage Institute[1] about the number of Cray systems installed can still be found on the Web and are shown in Fig. 2. After a slow start in the seventies, Cray enjoyed a steady growth in the early eighties and was the undisputed market leader during this period.

From 1980 to 1985, the number of installed systems grew by 37% annually. While this demonstrated a healthy growth rate for Cray Research, it would have been naïve to assume that it reflected a sustainable growth rate for the industry. This was merely a rate driven by the early adoption of a new technology. Once suitable customers adopted a new technology, the expectable growth rate would be limited by the emergence of new customers and consequently drop lower.

4. 1985–1990: The Golden Age of Vector Computers

The class of symmetric multi-vector processor systems dominated the supercomputing arena due to its commercial success in the eighties. This allowed the compilation of realistic statistics on the supercomputer market by focusing on this class of system exclusively, which was done by Hans W. Meuer at the Mannheim Supercomputer Seminar. Nevertheless, the new class of mini-supercomputers offered a very lucrative niche-market by providing an order of magnitude of smaller systems, mostly vector processor based systems to customers, which did not need a full-scale vector mainframe. Their emergence was the first sign that the performance gap between normal mainframes and vector-supercomputers would eventually close. This period also saw first attempts on building massively parallel computer systems. This architectural class would eventually close the performance gap and largely replace vector processor based systems. However, despite the later success of MPPs, none of the early pioneering MPP companies survived.

4.1. The Mannheim Supercomputer Statistics

Every year since 1986, Hans W. Meuer published statistics on the supercomputer market at the Mannheim Supercomputer Seminar [4], which later was renamed the International Supercomputer Conference (ISC). The statistics were primarily based on system counts of the major vector computer manufacturers. All the data used in these statistics, which were compiled from 1986 to 1992, had mostly been collected by market surveys from the manufacturers about the numbers of systems installed in different countries. The data obtained from these sources were crosschecked with each other (which from a European perspective was due to geographic distance particularly important for Japanese systems) and appropriate averages were derived.

Figure 3 shows the development of the overall market for vector supercomputers during the available time frame of 1986 to 1992. In 1986, the vector computer technology was already proven and mature. Vector computers were 10 years old, very sophisticated auto-vectorizing Fortran-compilers were available from both US vendors, Cray Research and Control Data, and the three Japanese vendors: Fujitsu,

[1] See: http://www.cbi.umn.edu

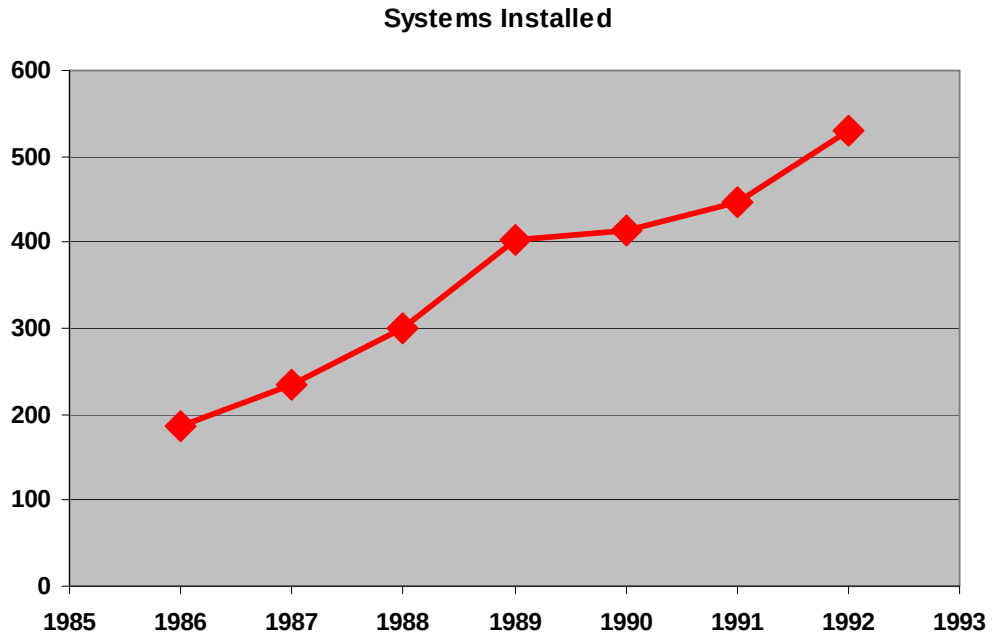

Fig. 3: Number of vector supercomputers installed 1986 to 1992

NEC and Hitachi. Last but not least, a variety of software packages, e.g. Nastran, Gaussian, PAM-crash, were on the market, making vector computers attractive not only to universities and research laboratories but also to the industry, especially to the automotive industry. If we look at the market growth in Fig. 3, we see that the vector computer had an average annual growth of 15%, which was quite a bit above the growth of the market for general computer systems at the same time. Nevertheless, the growth of supercomputer installations had slowed down from the high 37% annual increase Cray had enjoyed in the first half of the decade (see Fig. 2)

Fig 4 shows the geographical distribution of supercomputers over the years 1986 to 1992. Two major trends are recognizable:

The share of US-installations decreased steadily in contrast to the increase of the Japanese share. By 1992, the worldwide dominance of Japan was agreed upon more or less by all experts in the field, but in 1993, the first TOP500 showed the USA far ahead in supercomputers [3]. This was mainly because a bigger number of relatively small entry-level systems of Japanese vendors have been installed in Japan. In the older Mannheim statistics, these systems counted with equal weights to larger systems while in the TOP500 system are counted based on their Linpack performance relative to their peer systems. There was also an early orientation of US researchers towards MPP systems, which were not included in the Mannheim statistics but were allowed for the TOP500. The TOP500 list showed indeed that the shares of the USA and Japan were almost the same in 1993 as they had been in 1986 and that the USA clearly dominated as the consumer of supercomputer systems.

The European portion kept rather stable over these six years with a little less than 25%. The share of the UK in Europe decreased from 8% in 1986 down to 4% in

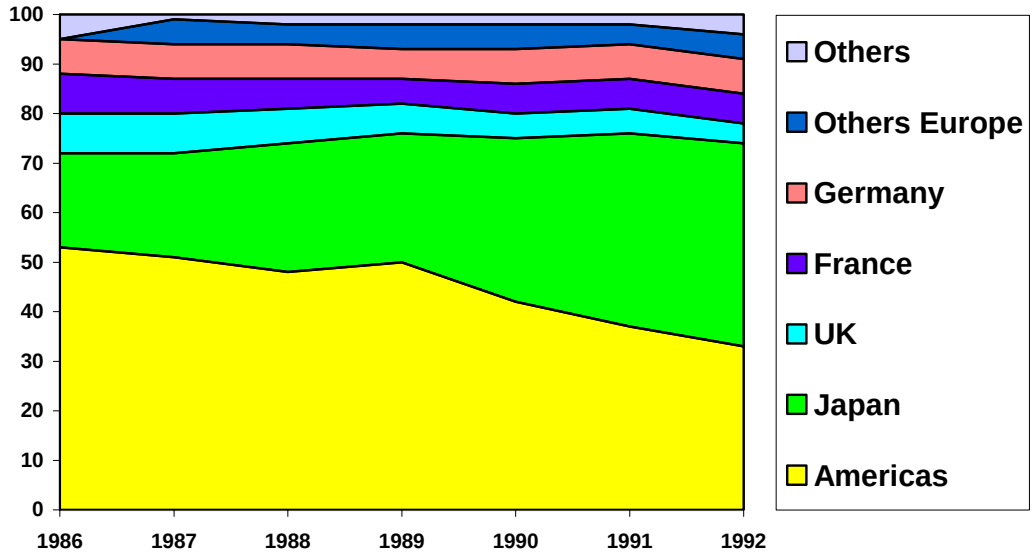

Fig. 4: Geographic distribution of vector supercomputers 1986 to 1992 in percent

1992, which is remarkable and can be explained by the funding policy of the UK government. The share of France slightly decreased from 8% to 6%, while the portion of Germany kept stable over the years with 7%. Other countries in Europe, e.g. Italy, Spain, the Netherlands and Denmark had 5% of the worldwide supercomputer distribution.

4.2. Cray Y-MP - Success in Industry

The follow up of the Cray X-MP, the Cray Y-MP, was a typical example for the sophistication of the memory access subsystems, which was one of the major reasons for the overall very high efficiency achievable on these systems. With this product line later including the C-90 and T-90, Cray Research followed the very successful path to higher processor numbers always trying to keep the usability and efficiency of their systems as high as possible. The Cray Y-MP first delivered in 1988 had up to eight processors, the C-90 first delivered in 1991 up to 16 processors and the T-90, first delivered in 1995, up to 32 processors. All these systems were produced in ECL chip technology.

Beginning of the eighties the acceptance of the Cray 1 systems was strongly helped by the easy integration in computing center environments of other vendors and by standard programming language support (Fortran77). After 1984, a standard UNIX operating system, UNICOS, was available for all Cray systems, which was quite an innovation for a mainframe at that time. With the availability of vectorizing compilers in the second half of the eighties, more independent software vendors started to port their key applications on Cray systems, which was an immense advantage to sell Cray systems to industrial customers. Due to these reasons, Cray vector systems started to have success in industries such as automotive industry and oil industry. Success in these markets ensured the dominance of Cray Research in the overall supercomputer market for more than a decade.

Table 1 shows the number of worldwide installed vector systems based on the Mannheim statistics. The dominance of Cray during this time with a constant market share of 60% is quite evident. This was later confirmed by the first TOP500 list from June 1993 [3], which included not only vector but also MPP systems. Cray had an overall share of 40% of all the installed systems, which was equivalent to 60% of the included vector systems.

Year	Cray	CDC	Fujitsu	Hitachi	NEC	Total
1986	118	30	31	8	2	187
1987	148	34	36	9	8	235
1988	178	45	56	11	10	300
1989	235	62	72	17	18	404
1990	248	24	87	27	27	413
1991	268		108	36	35	447
1992	305		141	44	40	530

Table 1: Vector computer installations worldwide.

4.3. Cray 3

Seymour Cray left Cray Research Inc. in 1989 to start Cray Computer and to build the follow up to the Cray 2 the Cray 3. Again, the idea was to use the most advanced chip technology to push single processor performance to its limits. The choice of GaAs technology was however ahead of its time and lead to many development problems. In 1992, a single system was delivered. The announced Cray 4 system was never completed.

4.4. ETA

In 1983 CDC decided to spin of its supercomputer business in the subsidiary 'ETA Systems Inc'. The ETA10 system was the follow up to the Cyber 205 on which it was based. The CPU's had the same design and the systems had up to eight processors with a hierarchical memory. This memory consisted of a global shared memory and local memories per processor, all of which were organized as virtual memory. CMOS was chosen as basic chip technology. To achieve low cycle times the high-end models had sophisticated cooling system using liquid nitrogen. First systems were delivered in 1987. The largest model had a peak performance of 10 GFlop/s well beyond the competing model of Cray Research.

ETA however seem to have overlooked the fact that raw performance was no longer the only or even most important selling argument. In April 1989, CDC terminated ETA and closed its supercomputer business. One of the main failures of the company was that they overlooked the importance of standard operating system and standard development environments. This was a mistake, which brought not only ETA but also CDC itself down. The main impact of the CDC/ETA withdrawal from the market was the relatively small overall growth of the market (Fig. 3) in 1990 and 1991. Many of the former CDC installations did not switch to one of the other big vector computer manufacturers like Cray Research. It seemed, that in many cases they bought mini-supercomputers like Convex or Alliant systems.

4.5. Mini-Supercomputer

Due to the limited scalability of existing vector systems there was a gap in performance between traditional scalar mainframes and the vector systems of the Cray class. This market was targeted by some new companies who started in the early eighties to develop the so-called mini-supercomputer. Design goal was one third of the performance of the Cray class supers but only one tenth of the price. The most successful of these companies was Convex founded by Steve Wallach in 1982. They delivered the first single processor system Convex C1 in 1985. In 1988, the multiprocessor system C2 followed. Due to the wide software support, these systems were well accepted in industrial environments and Convex sold more than 500 of these systems worldwide.

4.6. MPP - Scalable Systems and the Killer Micros

In the second half of the eighties a new class of system started to appear - parallel computers with distributed memory. Supported by the Strategic Computing Initiative of the US Defense Advanced Research Agency (DARPA – 1983), a couple of companies started developing such systems early in the eighties. Basic idea was to create parallel systems without the obvious limitations in processor number shown by shared memory designs of the vector multiprocessor.

First models of such massive parallel systems (MPP) were introduced in the market in 1985. At the beginning, the architectures of the different MPPs were still quite diverse. Major exception was the connection network as most vendors choose a hypercube topology. Thinking Machine Corporation (TMC) demonstrated their first SIMD system the Connection Machine 1 (CM1). Intel showed their iPSC/1 hypercube system using the Intel 80286 processor and Ethernet based connection network. nCube produced the first nCube/10 hypercube system with scalar Vax-like custom processors. While these systems still were clearly in the stage of experimental machines, they formed the basis for broad research on all issues of massive parallel computing. Later generations of these systems were then able to compete with vector MP systems.

Due to the conceptual simplicity of the global architecture the number of companies, building such machine grew very fast. This included the otherwise rare European efforts to produce supercomputer hardware. Companies who started to develop or produce MPP system in the second half of the eighties include: TMC, Intel, nCube, FPS (Floating Point Systems), KSR (Kendall Square Research), Meiko, Parsytec, Telmat, Suprenum, MasPar, BBN, and others.

4.7. Thinking Machines

After demonstrating the CM-1 (Connection Machine) in 1985, TMC soon introduced the follow on CM-2, which became the first major MPP, designed by Danny Hillis. In 1987 TMC started to install the CM-2 system. The Connection Machine model were single instruction on multiple data (SIMD) systems. Up to 64k single-bit processors connected in a hypercube network together with 2048 Weitek floating-point units could work together under the control of a single front-end system on a single problem. The CM-2 was the first MPP system which was not only successful in the

market but which also could challenge the vector MP systems of its time (Cray Y-MP), at least for certain applications.

The success of the CM-2 was great enough that another company, MasPar, which started producing SIMD systems as well. Its first system the MasPar MP-1 using 4-bit processors was first delivered in 1990. The follow on model MP-2 with 8-bit processors was first installed in 1992.

Main disadvantage of all SIMD system however proved to be the limited flexibility of the hardware, which limited the number of applications and programming models, which could be supported. Consequently, TMC decided to design their next major system the CM-5 as MIMD system. To satisfy the existing customer base this system could run data-parallel program as well.

4.8. Early MPPs

Competing MPP manufacturer had from the start decided to produce MIMD systems. The more complex programming of these systems was more than compensated by their much greater flexibility to support different programming paradigms efficiently.

Intel built systems based on the different generations of Intel microprocessors. The first such system, the iPSC/1, was introduced in 1985 and used Intel 80286 processors with an Ethernet based connection network. The second model iPSC/2 used the 80386 and already had a circuit switched routing network. The iPSC/860 introduced in 1990 finally featured the i860 chip. For Intel massive parallel meant up to 128 processors which was the limit due to the maximum dimension of the connection network.

In contrast to using standard off-the-shelf microprocessor nCube had designed their own custom processor as basis for their nCube/10 system introduced in 1985 as well. The design of the processor was similar to the good old Vax and therefore a typically CISC design. To compensate for the relatively small performance of this processor the maximal number of processors possible was however quite high. Limitation was again the dimension 13 of the hypercube network, which would have allowed up to 8096 processors. The follow-up nCube/2 again using this custom processor was introduced in 1990.

5. 1990–1995: MPP come to age

Beginning of the 1990s while the MP vector systems reached their widest distribution, a new generation of MPP system came on the market with the claim to be able to substitute of even surpass the vector MPs. The increased competitiveness of MPPs made it less and less meaningful to compile ‘Supercomputer’ statistics by just counting vector computers. This was together with the increasing importance of min-supercomputer and entry level models of larger vector system the major reason for starting the TOP500 project [3]. In this project, we list twice a year the 500 most powerful installed computer systems ranked by the best LINPACK performance [9]. In the first TOP500 list in June 1993, there were already 156 MPP and SIMD systems present (31% of the total 500 systems).

The hopes of all the MPP manufacturers to grow and gain market share however did

not come true. The overall market for HPC systems did grow only slowly and mostly in directions not anticipated by the traditional MP vector or MPP manufacturers. The attack of killer micros went into its next stage. This time the large-scale architecture of the MPPs seen before would be under attack.

One major side effect of the introduction of MPP system in the market for supercomputer was that the sharp performance gap between supercomputers and mainstream computers no longer existed. MPPs could (almost by definition) be scaled by one or two magnitudes of order bridging the gap between high-end multi-processor workstations and supercomputers. Homogeneous and monolithic architectures designed for the very high end of performance would have to compete with clustered concepts based on shared memory multi-processor models from the traditional UNIX server manufacturer. These cluster offered another level of price/performance advantage due to the large supporting industrial and commercial business market in the background and due to the reduced design cost not focusing on the very high end of performance any more. This change towards the usage of standard components widely used in the commercial marketplace actually widened the market scope for such MPPs in general. As a result the notion of a separated market for floating point intensive high-performance computing no longer holds. The HPC market is nowadays the upper end of a continuum of systems and usage in all kind of applications.

5.1. TOP500

Early in the nineties, the limitations of the approach of the Mannheim statistics became evident and the author together with Hans Meuer experimented for three year with different new approaches to compile better statistics about supercomputers. These approaches included counting systems and processors and compiling different lists of systems. The outcome of our studies was the TOP500 project [3]. Basic idea was to give any type of system the possibility to be counted as a supercomputer if it could demonstrate performance levels worthy of such a label. The actual performance level necessary for this label would have to be adjusted over time as general performance levels increased. This could be done in an automatic and very elegant way by compiling a list of systems with the largest performance values and cutting it of after a predetermined number of entries. This would ensure that only the very largest system would be counted at any given time.

Because we knew from the Mannheim statistics (table 1), that more than 500 vector systems were installed worldwide, a cut-off after 500 systems appeared a good choice. For ranking purposes we decided not to use peak-performance, but actual measured performance values to avoid listing any non-functional or even fictional systems. For practical reasons the only benchmark usable was the Linpack benchmark, as it was the benchmark with by far the largest number of results available for almost all relevant systems [9]. With this the TOP500 was borne. Ever since June 1993, we assemble twice a year a list of the 500 most powerful computer systems installed.

Together with the computer system, the installation site, and the Linpack benchmark performance, we are recording a variety of information such as the number of processors, the customer segment, the major application area, the year of installation or last major update, the processor and interconnect technologies, and the operating system used. Keeping all this information in a database enables us to easily answer different statistical questions.

One drawback of our new methodology is however the fact, that we set the number of supercomputers at an arbitrary level, 500. With this, we can no longer directly follow the overall size of the market and its development. This could only be avoided by using a different definition of the term supercomputer such as its price tag. We are often asked about the prices of the supercomputer and the price/performance of a system is a very useful metric. However, list prices differ substantial from actual prices paid and we found over the years that real prices are very hard to come by, very hard to track, and very unreliable. We therefore never started to collect any pricing information about the system we track.

The TOP500 is based on information obtained from manufacturers, customers and users of such systems. We ensure the quality of the information by crosschecking different sources, and by the comments of experts in the field who are willing to proofread the list before publication. Errors are still bound to exist, which is especially true for classified installations as the nature of such sites makes it difficult to obtain any information about them. From the responses we received, we are very confident that the average accuracy and quality of the TOP500 is quite high.

5.2. Attack of the Killer Micros

Beginning of the nineties one phrase showed up on the front pages of several magazines: “The Attack of the Killer Micro”. Coined by Eugene Brooks from Livermore National Laboratory this was the synonym for the greatly increased performance levels achievable with microprocessors after the RISC/super-scalar design revolution. The performance of microprocessors seemed to have reached comparable level to the much more expensive ECL custom mainframe computer. However, not only the traditional mainframes started to feel the heat, even the traditional supercomputer the vector multi-processors got under attack.

Another slogan for this process was widely used when Intel introduced its answer to the RISC processors, the i860 chip: “Cray on a chip” Even as sustained performance values did not always come up to PAP (peak advertised performance) values the direction of the attack was clear. The new generation of microprocessors manufactured relatively cheap in great numbers for the workstation market offered the much better price/performance ratios. Together with the scalable architecture of MPPs, the same high performance levels as with vector multiprocessor could be achieved for a better price.

Aggravated was this contrast greatly by the fact that most mainframe manufacturers had not seen early enough the advantage of CMOS chip technology and were still using the (little) faster but much more expensive ECL technology. Cray Research was no exception in this respect. Under great pressure, all mainframe manufacturers started to switch from ECL to CMOS. At the same time, they also started to produce their own MPP systems to have competing products with the up and coming new MPP vendors.

Hidden behind these slogans were actually two different trends working together, both of which effects can clearly be seen in the TOP500 data. The replacement of ECL chip technology by CMOS is shown in Fig. 5. The rapid decline of vector-based systems in the nineties can be seen in the TOP500 in Fig. 6. This change in the processor architecture however was not as generally accepted as the change form ECL to

CMOS. NEC and Cray continue to produce vector-based systems to this day.

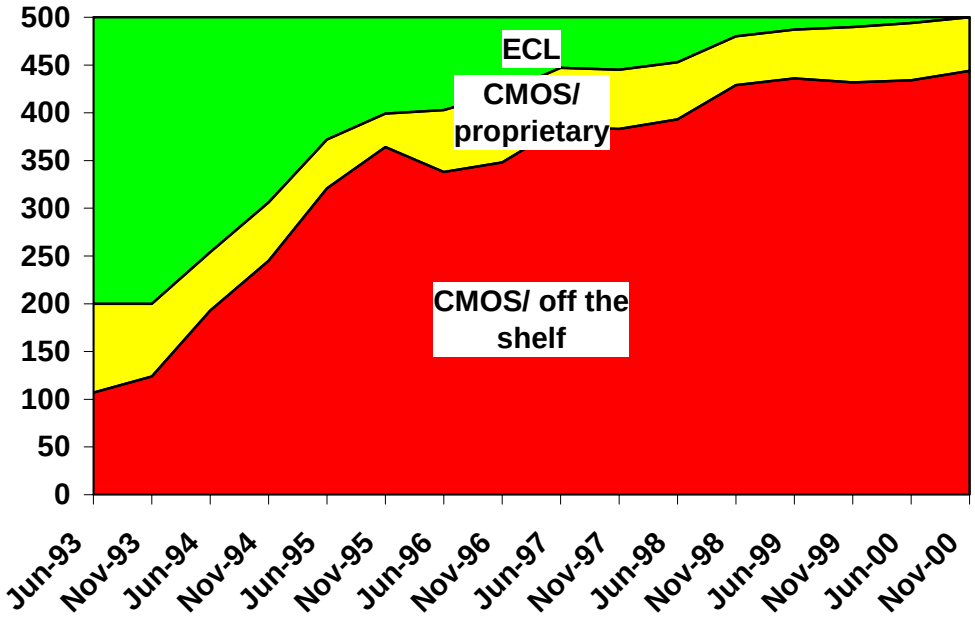

Fig. 5: Chip technologies usage as seen in the TOP500.

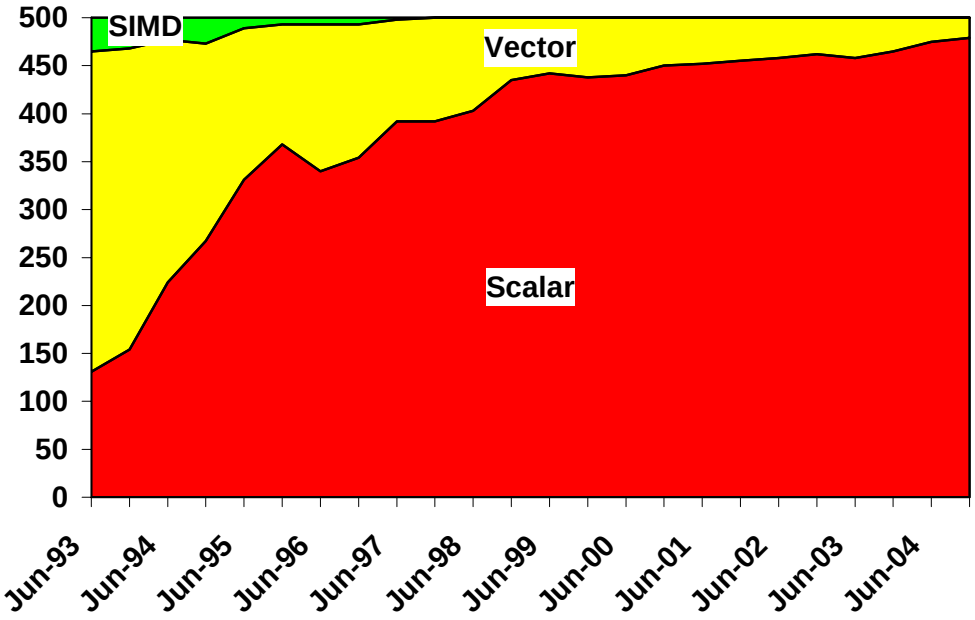

Fig 6: CPU design usage as seen in the TOP500.

5.3. Playground for Manufacturers

With the largely increased number of MPP manufacturers, it was evident that a "shake-out" of manufactures was unavoidable. In table 2, we list vendors, which have been active at some point in the HPC market [10,11]. In Fig. 7, we try to visualize the

Status	Vendors
Out of business	Alliant, American Supercomputer, Ametek, AMT(Cambridge), Astronautics, BBN Supercomputer, Biin, CDC/ETA Systems, Chen Systems, CHoPPCHoPP, Cogent, Cray Computer, Culler, Cydrome, Denelcor, Elexsi, Encore, E&S Supercomputers, Flexible, Goodyear, Gould/SEL, Intel Supercomputer Division, IPM, iP-Systems, Key, Kendall Square Research, Multiflow, Myrias, Pixar, Prevec, Prisma, Saxpy, SCS, SDSA, Suprenum, Stardent (Stellar and Ardent), Supercomputer Systems Inc., Synapse, Thinking Machines, Trilogy, VItec, Vitesse, Wavetracer
Merged	Celerity, Compaq (with Hewlett-Packard), Convex (with Hewlett-Packard), Cray Research (with SGI - temporarily), DEC (with Compaq), Floating Point Systems (with Cray Research), Key, MasPar (with DEC), Meiko, Supertek (with Cray Research), Tera (with Cray)
Changed market	nCUBE, Parsytec Siemens
Currently active	Cray, Fujitsu, Hewlett-Packard, Hitachi, IBM, NEC, SGI, Sun, and a large number of small cluster integrators such as Linux Networks, Atipa and Lenovo

Table 2. Commercial HPC Vendors

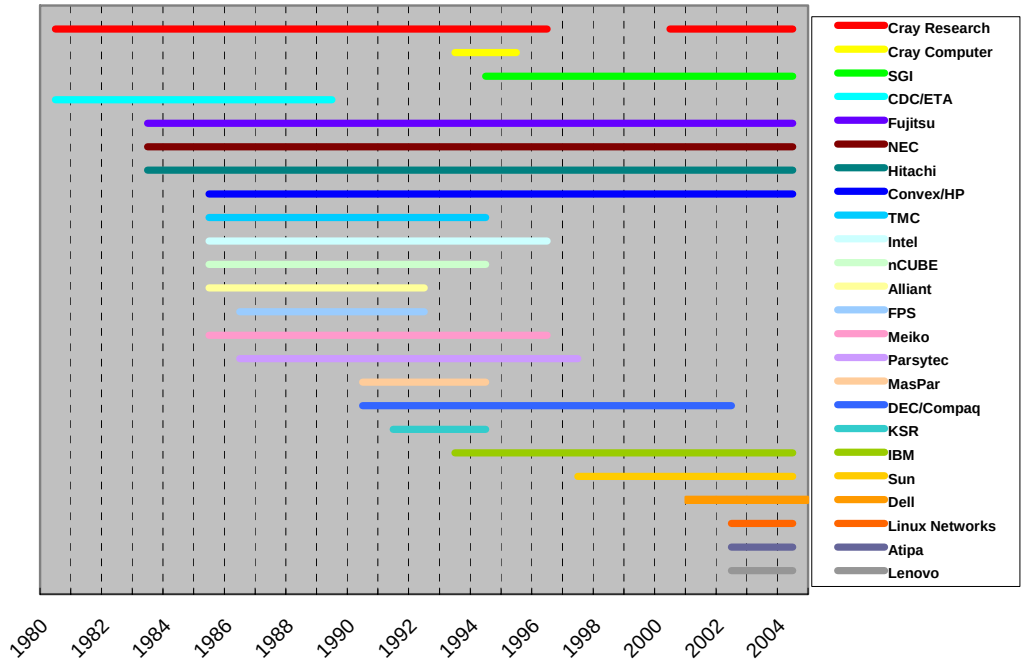

Fig. 7: Main manufacturer active in the HPC market.

historic presence of companies in the HPC market. After 1993, we included only companies, which had at least once two entries in the TOP500 and were actively manufacturing. There are more than twenty additional small companies, which had at the most two entries in the TOP500.

Of the 14 major companies in the early nineties, only four survived the decade on their own. These were the three Japanese vector manufacturer (Fujitsu, Hitachi, and NEC) and IBM, which due to its marginal HPC presence at the beginning of the nineties could even be considered a newcomer. Four other companies entered the HPC market either by buying some companies or by developing their own products (Silicon Graphics, Hewlett-Packard, Sun, Compaq). None of these was a former HPC manufacturer. All were typical workstation manufacturer, which entered the HPC market (at least initially) from the lower end with high-end UNIX-server models. Their presence and success already indicated the change in focus from the very high end to markets for medium size HPC systems.

5.4. Kendall Square Research

In 1991, first models of a quite new and innovative system were installed, the KSR1 from Kendall Square Research. The hardware was built similar to other MPPs with distributed memory but gave the user the view of a shared memory system. The custom design hardware and the operating system software were responsible for this virtual shared memory appearance. This concept of virtual shared memory could later on be found in other systems such as the Convex SPP series and lately in the SGI Origin series. The KSR systems organized the complete memory on top of the VSM as cache only memory. By this, the data had no fixed home in the machine and could freely room to the location where they were needed. ‘Management mistakes’ brought the operations of KSR to an abrupt end in late 1994.

5.5. Intel

In 1992 Intel started to produce the Paragon/XP series after delivering the Touchstone Delta system to Caltech in 1991. Still based on the i860 chips the interconnection network was changed to a two dimensional grid which now allowed up to 2048 processors. Several quite large system were installed and in June 1994 a system at Sandia National Laboratory which achieved 143 GFlop/s on the LINPACK benchmark was the number one in the TOP500. Intel decided to stop its general HPC business in 1996 but still built the ASCI Red system afterwards.

5.6. Thinking Machines

In 1992, TMC also started to deliver the CM-5 a MIMD system designed for the very high end. Theoretically, systems up to 16k processors could be built. In practice, the largest configurations reached 1056 processors at the Los Alamos National Laboratory. This system was at the number one spot of the very first TOP500 list in June 1993 achieving almost 60 GFlop/s. A Sun SPARC processor was chosen as basic node processor. Each of these processors had four vector coprocessors to increase the floating-point performance of the system. Initial programming paradigm was the data-parallel model familiar from the CM-2 predecessor. The complexity of this node design however was more than the company or customers could handle. Due

to the design point being the very high end, the smaller models also had problems competing with models from other companies, which did not have to pay for the overhead of supporting such large systems in their design. The CM-5 would be the last TMC model before the company had to stop the production of hardware in 1994.

The raw potential of the CM-5 was demonstrated by the fact that in June 1993 the position 1–4 were all held by TMC CM-5 systems ahead of the first MP vector system an NEC SX-3/44R. The only other MPP system able to beat a Cray C-90 at that time was the Intel Delta system. The performance leadership however had started to change.

In June 1993, still five of the first 10 systems were MP vector systems. This number decreased fast and the last MP vector system, which managed to make the top 10, was a NEC SX-4 with 32 processors in June 1996 with 66.5 GFlop/s. Later only systems with distributed memory made the top 10 list. Japanese MPPs with vector processors managed to keep their spot in the top 10 for some time. In the November 1998 list however, the top 10 positions were for the first time all taken by microprocessor based ASCI or Cray T3E systems. The first system with vector CPU's was a NEC SX-4 with 128 processor at number 18.

5.7. IBM

In 1993, IBM finally joined the field of MPP producers by building the IBM SP1 based on their successful workstation series RS6000. While this system was often mocked being a workstation cluster and not a MPP, it set the ground for the reentry of IBM in the supercomputer market. The follow on SP2 with increased node and network performance was first delivered in 1994. Contrary to other MPP manufacturers IBM was focusing on the market for small to medium size machines especially for the commercial UNIX server market. Over time, this proved to be a very profitable strategy for IBM who managed to sell models of the SP quite successful as a database server. Due to the design of the SP2, IBM is able to constantly offer new nodes based on the latest RISC system available.

5.8. Cray Research

In 1993, Cray Research finally started to install their first MPP system, the Cray T3D. As indicated by the name the network was a three dimensional torus. Cray had chosen the DEC alpha processor as CPU. The design of the node was completely done by Cray itself and was substantially different from a typical workstation using the same processor. This had advantages and disadvantages. Due to their closely integrated custom network interface, the network latencies and the bandwidth reached values not seen before and allowed very efficient parallel processing. The computational node performance itself was however greatly affected by the missing 2nd level cache. The system was immediately well accepted at research laboratories and was even installed at some industrial customer sites. The largest configuration known is installed at a classified government site in the USA with 1024 processors and just breaking the 100 GFlop/s barrier on the LINPACK.

5.9. Convex

In 1994, Convex introduced its first true MPP, the SPP1000 series. This series was

also awaited with some curiosity, as it was after KSR the second commercial system featuring a virtual shared memory. The architecture of the system was hierarchical. Up to 8 HP microprocessors were connected to a shared memory with crossbar technology similar to the one used in the Convex vector series. Multiple of these SMP units would then be connected in a distributed memory fashion. The operating system and the connection hardware would provide the view of a non-uniform shared memory over the whole machine. In the following years a series of follow on models was introduced the SPP1200 in 1995, the SPP1600 in 1996, and the SPP2000 renamed by HP as Exemplar X-Class in 1997.

5.10. The role of MP vector systems during 1990–1995

Cray Research continued building their main MP vector system in traditional style. The Triton, known as T-90, was introduced in 1995 and built in ECL very much along the line of the Y-MP and C-90 series. The maximum number of processors was increased to 32. This gave a full system a peak performance of 58 GFlop/s. Realizing that it needed a product for lower market segment Cray had bought the company Supercomputer which had developed Cray Y-MP compatible vector systems in CMOS technology. It was marketed by Cray starting in 1993. The next system in this series developed by Cray Research itself was the J-90 introduced in 1995 as well. With up to 32 processors it reached a peak performance of 6.4 GFlop/s, which was well below the ECL systems from Cray and unfortunately not much above the performance of best microprocessor available.

Convex introduced the C3 series in 1991 and the C4 in 1994 before the company was bought by Hewlett-Packard the same year. After this merger, the unit focused on its MPP products.

In Japan, Fujitsu had introduced the single processor VP2600 in 1990 and would market this series until the introduction of the CMOS based VPP500 in 1994. NEC introduced the SX-3 series in 1990 as well. With its up to four processors this system reached a peak performance of 26 GFlop/s. NEC subsequently implemented their vector series in CMOS and introduced the NEC SX-4 in 1994. Up to 32 processors can be installed as conventional shared memory MP vector system. Beyond this up to 16 of these units can be clustered in a distributed memory fashion. The largest configurations known to be installed have 128 processors with which they gained positions 29 and 30 in the June 1999 TOP500 list. These are at present the largest vector based systems with traditional design.

5.11. Fujitsu's MPP-Vector Approach

Fujitsu decided to go its own way into the world of MPPs. They built their commercial MPP system with distributed memory around the node and the processors of their successful VP2600 vector computer series. However, Fujitsu was the first Japanese company who implementing their vector design in CMOS, the VPP500. A first 'pre-prototype' was developed together with the National Aerospace Laboratories (NAL). The installation of this system named the Numerical Wind Tunnel (NWT) started in 1993. Due to its size this system managed to gain the number 1 position in the TOP500 an unchallenged four and number 2 position three times from November 1993 to November 1996. Delivery of VPP500 systems started in 1993.

6. 1995-2000: SMP based Systems and Commercial Customers

The year 1995 saw some remarkable changes in the distribution of the systems in the TOP500 for the different types of customer (academic sites, research labs, industrial/commercial users, vendor installations, and confidential sites) (see Fig. 8).

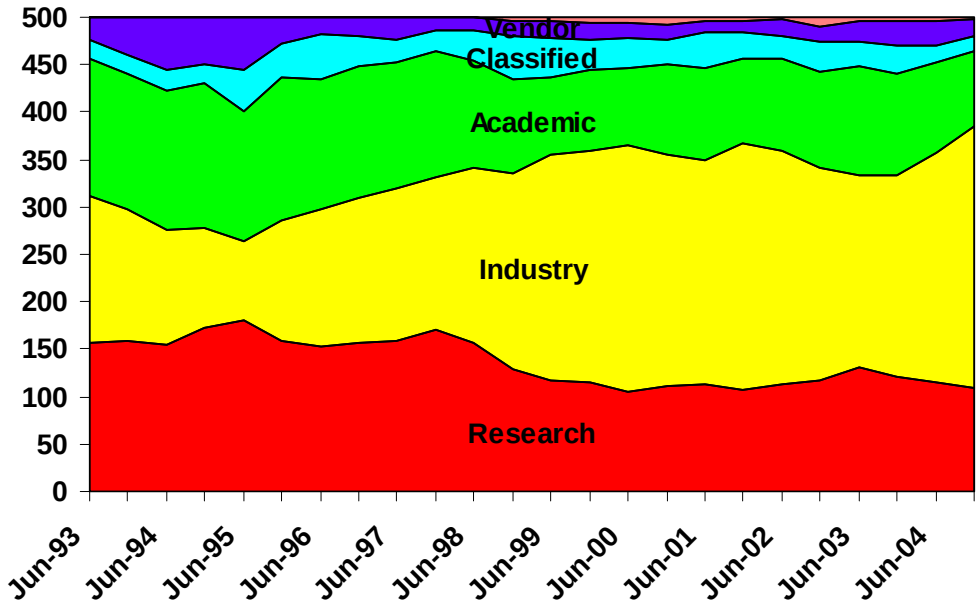

Fig. 8: The number of systems on the different types of customers over time.

Until June 1995, the major trend seen in the TOP500 data was a steady decrease of industrial customers, matched by an increase in the number of government-funded research sites. This trend reflected the influence of the different governmental HPC programs that enabled research sites to buy parallel systems, especially systems with distributed memory. Industry was understandably reluctant to follow this step, since systems with distributed memory had often been far from mature or stable. Hence, industrial customers stayed with their older vector systems, which gradually dropped off the TOP500 list because of low performance.

Beginning in 1994, however, companies such as SGI, Digital, and Sun started to sell symmetrical multiprocessor (SMP) models of their major workstation families. From the very beginning, these systems were popular with industrial customers because of the maturity of these architectures and their superior price/performance ratio. At the same time, IBM SP2 systems started to appear at a reasonable number of industrial sites. While the SP initially was sold for numerically intensive applications, the system began selling successfully to a larger market, including database applications, in the second half of 1995.

Subsequently, the number of industrial customers listed in the TOP500 increased from 85, or 17%, in June 1995 to about 241, or 48.2%, in June 1999. We believe that this was a strong new trend because of the following reasons.

- The architectures installed at industrial sites changed from vector systems to a substantial number of MPP systems. This change reflected the fact that parallel systems are ready for commercial use and environments.

- The most successful companies (Sun, IBM and SGI) were selling well to industrial customers. Their success was built on the fact that they were using standard workstation technologies for their MPP nodes. This approach provided a smooth migration path for applications from workstations up to parallel machines.

- The maturity of these advanced systems and the availability of key applications for them made the systems appealing to commercial customers. Especially important were database applications, since these could use highly parallel systems with more than 128 processors.

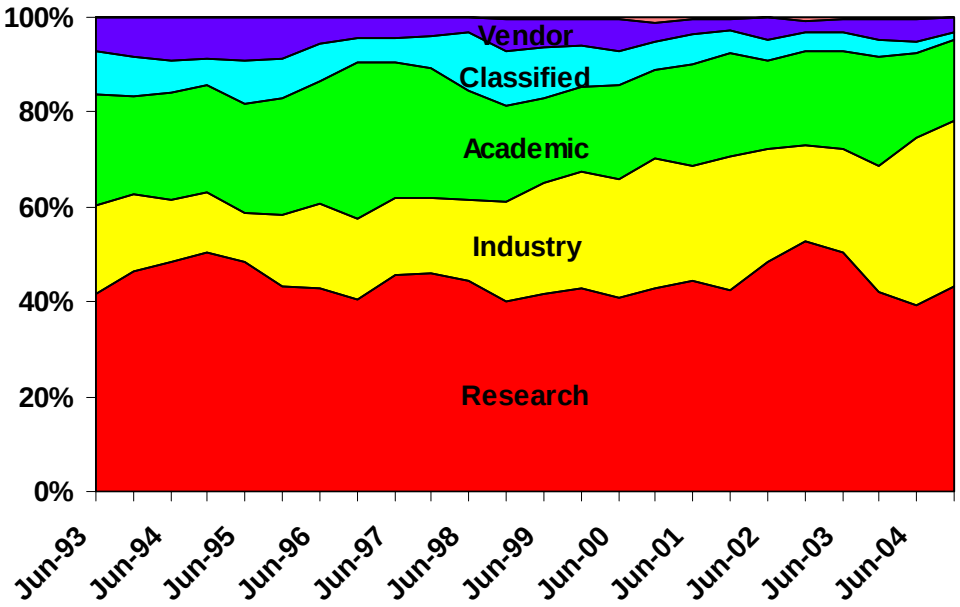

Fig. 9: The accumulated performance of the different types of customers over time.

Fig. 9 shows that the increase in the number of systems installed at industrial sites was matched by a similar increase in the installed accumulated performance. The relative share of industrial sites rose from 8.7% in June 1995 to 24.0% in June 1999. Thus, even though industrial systems were typically smaller than systems at research laboratories and universities, their average performance and size were growing at the same rate as at research installations. The strong increase in the number of processors in systems at industrial sites was another major reason for the rise of industrial sites in the TOP500. The industry was ready to use bigger parallel systems than in the past.

6.1. Architectures

The changing share of the different system architectures in the HPC market as reflected in the TOP500 is shown in Fig. 10. Besides the fact that no single processor systems were any longer powerful enough to enter the TOP500 at the end of the

decade, the major trend was the growing number of MPP systems. The number of clustered systems was also growing and at the end of the decade, we saw a number of PC or workstation based 'Network of Workstations' in the TOP500. It was an interesting and open question, which share of the TOP500, such NOWs would eventually capture in the future.

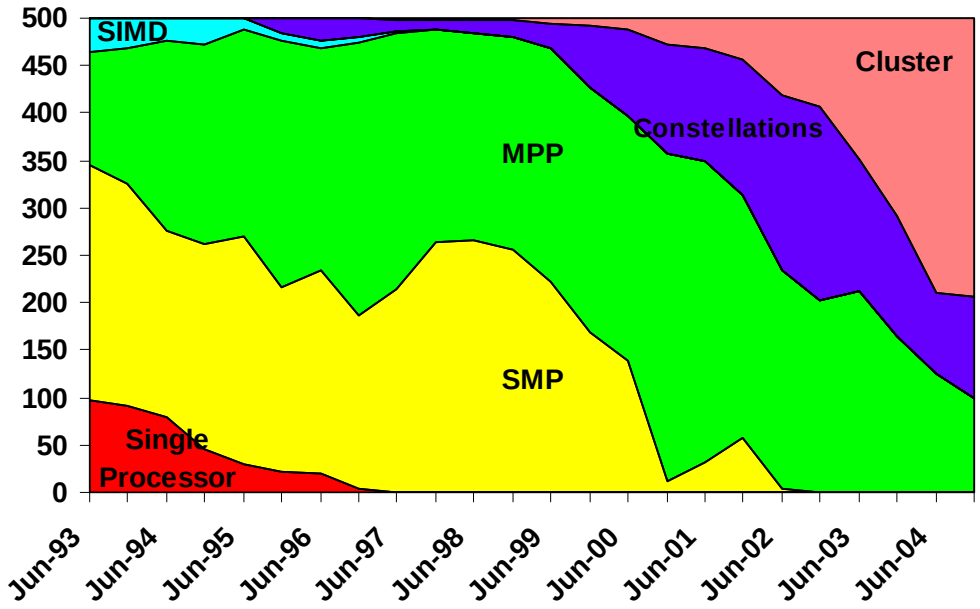

Fig. 10. Main Architectural Categories seen in the TOP500.

6.2. Vector based Systems

Cray Research introduced their last ECL-based vector system the T-90 series in 1995. Due to the unfavorable price/performance of this technology, the T-90 was not an economical success for Cray Research. One year later in 1996, SGI bought Cray Research. After this acquisition, the future of the Cray vector series was in doubt. The joint company announced plans to produce a joint macro architecture for its microprocessor and vector processor based MPPs. In mid 1998, the SGI SV1 was announced as future vector system of SGI. The SV1 was the successor to both the CMOS based Cray J-90 and the ECL based Cray T-90. The SV1 was CMOS based which meant that SGI was finally following the trend set in by Fujitsu (VPP700) and NEC (SX-4) a few years earlier. First user shipments happened end of the decade and it was not clear, if the SV1 would be able to compete with the advanced new generation of Japanese vector systems, especially the NEC SX-5 and the Fujitsu VPP5000.

Fujitsu continued along the line of the VPP system and introduced in 1996 the VPP700 series featuring increased single node performance. For the lower market segment the VPP300 using the same nodes but a less expandable interconnect network was introduced. The next generation model VPP5000 was again a distributed-memory vector system, where four up to 128 (512 by special order) processors could be connected via a fully distributed crossbar. The theoretical peak performance ranged from 38.4 GFlop/s up to 1.229 TFlop/s, and in special

configurations even 4.915 TFlop/s.

NEC had announced the SX-4 series in 1994 and continued to produce systems along this architectural line. The SX-4 featured shared memory up to a maximum of 32 processors. Larger configurations were built as cluster using a proprietary crossbar switch. In 1998, the follow up model SX-5 was announced for first delivery in early 1999. In 1998, the follow up model SX-5 was announced for first delivery in late 1998 and early 1999 and first system were listed in the November 99 TOP500 list. In contrast to its predecessor, the SX-4, the SX-5 was not offered anymore with faster, but more expensive SRAM memory. The SX-5 systems were exclusively manufactured with synchronous DRAM memory. The multi-frame version of the SX-5 could host up to 512 processors with 8 GFlop/s peak performance each, resulting in a theoretical peak performance of 2 TFlop/s. More information about all these current architecture can be found in [12].

6.3. Traditional MPPs

Large scale MPPs with homogeneous system architectures had matured during the nineties with respect to performance and usage. Cray finally took the leadership here as well with the T3E system series introduced in 1996 just before the merger with SGI. The performance potential of the T3E can be seen by the fact that in June 1997, 6 of the top 10 positions in the TOP500 were occupied by T3Es. End of 1998 the top 10 consisted only of ASCI systems and T3Es.

Hitachi was one of the few companies introducing large-scale homogeneous system in the late nineties. It announced the SR2201 series in 1996 and tried to sell this system for the first time outside of Japan as well.

The first of the ASCI system, the ASCI Red at Sandia National Laboratory, was delivered in 1997. It took immediately the first position in the TOP500 in June 1997 being the first system to exceed 1 TFlop/s LINPACK performance. ASCI Red also ended several years during which several Japanese systems ranked as number one.

IBM developed their SP computer series further and introduced new nodes and faster interconnects. One major innovation here was the usage of SMP nodes as building blocks, which further demonstrates the proximity of the SP architecture to clusters. This design with SMP nodes was also chosen for the ASCI Blue Pacific systems.

6.4. SMPs and their Successors

Beginning in 1994, however, companies such as SGI, Digital, and Sun started to sell symmetrical multiprocessor (SMP) models of their major workstation families. From the very beginning, these systems were popular with industrial customers because of the maturity of these architectures and their superior price/performance ratio. At the same time, IBM SP2 systems started to appear at a reasonable number of industrial sites. While the SP initially was sold for numerically intensive applications, the system began selling successfully to a larger market, including database applications, in the second half of 1995.

SGI made a strong appearance in the TOP500 in 1994 and 1995. Their PowerChallenge systems introduced in 1994 sold very well in the industrial market

for floating point intensive applications. Cluster built with these SMPs appeared in a reasonable number at customer sites.

In 1996, the Origin2000 series was announced. With this system SGI took the step away form the bus based SMP design of the Challenge series. The Origin series featured a virtual memory system built with distributed memory nodes up to 128 processors. To achieve higher performance these systems could be clustered again. This was the basic design of the ASCI Blue Mountain system.

Digital was for a long time active as producer of clustered systems for commercial customers. In 1997, the Alpha Server Cluster was introduced which was targeted towards floating point intensive applications as well. One year later Compaq acquired Digital giving Compaq as first PC manufacturer an entry in the HPC market.

Hewlett-Packard continued producing systems along the line of the former Convex SPP systems targeting mainly the midsize business market where the company had good success. The very high end - which had never been a target for Convex or HP - still seemed to of minor interest to HP.

Sun was the next company who entered the TOP500. After the merger of SGI and Cray Research in 1996, Sun bought the former business server division of Cray which produced SPARC based SMP systems for several years. In 1997, Sun introduced the HPC 10000 series. This SMP system was built around a new type of switched bus, which allowed integrating up to 64 processors in an efficient way. Due to its wide customer base and good reputation in the commercial market, Sun was able to sell these SMPs very well especially to commercial and industrial customers. For the very high end market clusters built with these SMP were introduced in 1998.

6.5. New Application Areas

For research sites or academic installations, it is often difficult — if not impossible — to specify a single dominant application. The situation is different for industrial installations, however, where systems are often dedicated to specialized tasks or even to single major application programs. Since the very beginning of the TOP500 project, we have tried to record the major application area for the industrial systems in the list. We have managed to track the application area for almost 90% of the industrial systems over time.

Since June 1995, we saw many systems involved in new application areas entering the list. Fig. 11 shows the total numbers of all industrial systems which is made up of three components, traditional floating point intensive engineering applications, new non floating point applications, and unknown application areas.

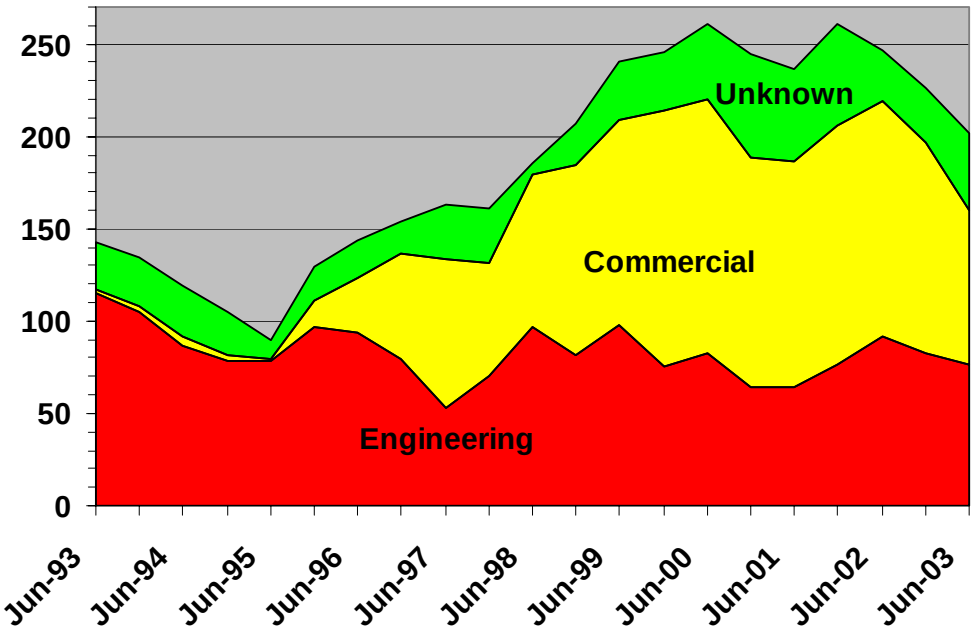

Fig. 11: The total number of systems at industrial sites together with the numbers of sites with traditional engineering applications, new emerging application areas and unknown application areas.

In 1993, the applications in industry typically were numerically intensive applications, for example,

- Geophysics and oil applications,
- Automotive applications,
- Chemical and pharmaceutical studies,
- Aerospace studies,
- Electronics, and
- Other engineering including energy research, mechanical engineering etc.

The share of these areas from 1993 to 1996 remained fairly constant over time. In the second half of the nineties, however, industrial systems in the TOP500 have been used for new application areas. These include

- Database applications,
- Finance applications, and
- Image processing.

The most dominant trend was the strong rise of database applications after November

1995. These applications included on-line transaction processing as well as data mining. The HPC systems, which were sold and installed for such applications were large enough to enter the first hundred positions—a clear sign of the maturity of these systems and their practicality for industrial usage.

It is also important to notice that industrial customers were buying not only systems with traditional architectures, such as the SGI PowerChallenge or Cray T-90, but also MPP systems with distributed memory, such as the IBM SP2. Distributed memory was no longer a hindrance to success in the commercial marketplace.

6.6. Government programs

The high end of the HPC market was always the target for government program all over the world to influence the further development of new systems. In the USA, there were and are currently several government projects on the way to consolidate and advance the numerical HPC capabilities of US government laboratories and the US research community in general. The most prominent of these was for some time the 'Accelerated Strategic Computing Initiative (ASCI)'. Goal of this program is "to create the leading-edge computational modeling and simulation capabilities that are essential for maintaining the safety, reliability, and performance of the U.S. nuclear stockpile and reducing the nuclear danger" [2].

Three main laboratories were selected as sites for deploring parallel computing systems of the largest scale technically possible. The system 'ASCI Red' was installed at Sandia National Laboratory. This system was produced by Intel and had 9472 Intel Pentium Xeon processors. It was the first system to exceed the 1 TFlop/s mark on the LINPACK benchmark in 1997 and remained the number 1 on the TOP500 for some time. ASCI Red is currently being replace by the Cray co-developed Opteron based Red Storm system with about 30 TFlop/s peak performance. 'ASCI Blue Mountain' a cluster of Origin2000 systems with a total of 6144 processors was produced by SGI. It was installed at the Los Alamos National Laboratory and achieved 1.6 TFlop/s Linpack performance. It was replace by ASCI-Q, a Compaq/HP built cluster with 8192 alpha processors and 13.9 TFlop/s Linpack. 'ASCI Blue Pacific' an IBM SP system with a total of 5856 processors was installed at the Lawrence Livermore National Laboratory. It was replaced by another IBM system called ASCI White with 8192 Power processors and 7.3 TFlop/s Linpack. In the near future LLNL will install a replacement system from IBM called ASCI Purple with a peak performance of about 100 TFlop/s

The Japanese government decided to fund the development of an 'Earth Simulator' to simulate and forecast the global environment. In 1998, NEC was awarded the contract to develop a 30 TFlop/s system to be installed by 2002.

---

[2] www.llnl.gov/asci/

7. 2000-2005: Cluster, Intel Processors, the Earth-Simulator, and BlueGene

In the early 2000s, Clusters built with components off the shelf 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 group of clusters represented the majority of systems on the TOP500 in a broad range of application areas. One major consequence of this trend was the rapid rise in the utilization if 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. Cluster in the nineties were mostly self-made system designed and built by small groups of dedicated scientist or application experts. This changed rapidly as soon as clusters based on PC technology became an important market. Nowadays the large majority of TOP500-class cluster is manufactured and integrated by either a traditional large HPC manufacturer such as IBM or HP or a small, specialized integrator 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 even so it had been announced 4 years earlier. The Earth Simulator built by NEC was based on the NEC vector technology and showed unusual high efficiency on many applications. It demonstrated that many scientific applications can benefit greatly from other computer architectures and helped to invigorate the discussions about future architectures for high-end scientific computing systems. In the U.S. the DARPA HPCS program has the declared goal of building a Petaflops computer system by the end of the decade and a variety of radically new architecture are investigated within this program.

In the meantime, the IBM BlueGene/L system has entered the supercomputer scene with a similar splash as the Earth Simulator. It is one example of a shifting design focus for large-scale system as it is build with low performance but very low power components. This allows a tight integration of an unprecedented number of processors to achieve surprising performance levels for suitable applications. Even so this system is still very early in its life cycle, it has already taken the #1 spot on the TOP500 without problem and is expected to hold on to it for some time to come. Again, there are serious open questions about the generality of its design, especially when considering its comparable small memory, but IBM might remedy some of the shortcomings of the first generation BlueGene in future years.

7.1. 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 10). 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.

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). In fact Linux is the operating system used on over 60% of the machines on the Top500. 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.

7.2. 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. 12).

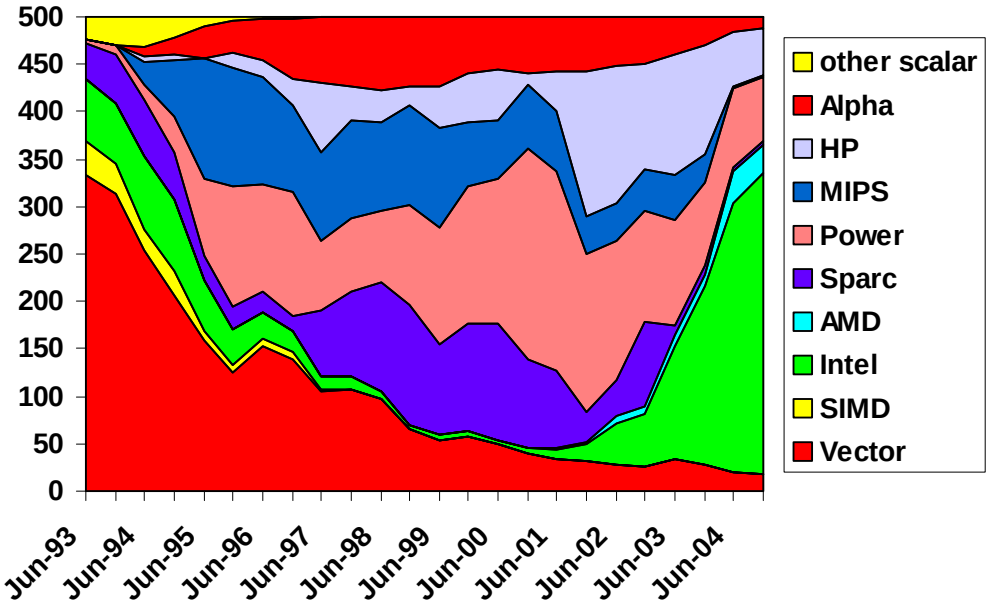

Fig. 12: Main Processor Families seen in the TOP500.

7.3. The Impact of the Earth-Simulator

The Earth-Simulator (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 high-end 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 commercially 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) [13]. This debate also led to a NRC study on "The Future of Supercomputing" [14].

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 the 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.

7.4. 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 new computer architectures 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.

7.5. New Benchmark and Performance Measure

The benchmark used in the TOP500 project is the Linpack benchmark, which solves a dense system of linear equations. Since this problem is very regular, the performance achieved is quite high, and the performance numbers give in most cases only a minor correction to the theoretical peak performance of a system. This leniency in performance requirements might explain its popularity but does not explain its longevity, which is greatly facilitated by a continuously scalable problem size. The ability to scale problems to arbitrary size and the performance property, that performance grows steadily with larger problem sizes explain a good deal of the ongoing popularity of the Linpack benchmark.

An unfortunate side effect of this is however, that the performance requirements of the Linpack benchmark become lower and lower as systems grow with time. Nowadays the Linpack performance on many system reaches 70% to 90% of peak performance, while at the same time real application performance is often in the range of 5% to 20% and seems to even further decline with newer generations of systems. Linpack is no longer a good representative of applications performance and the

TOP500 ranking based on it is therefore subject to distortions caused by this discrepancy.

This became very clear in the first half of this decade and a few new initiatives emerged trying to replace Linpack with a more demanding benchmark, which would be more representative of real HPC applications. However finding a benchmark with a potential longevity such as Linpack is by no means trivial. Ongoing research projects in these directions include the APEX-Map project [15,16] and the HPC-Challenge project [17].

8. 2005 and beyond

Three decades after the introduction of the Cray 1, the HPC market has changed 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 produce 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, there is a strong interest in new computer architectures.

8.1. Dynamic of the Market

The HPC market is by its very nature very dynamic. This is not only reflected by the coming and going of new manufacturers but especially by the need to update and replace systems quite often to keep pace with the general performance increase. This general dynamic of the HPC market is well reflected in the TOP500. In table 3, we show the number of systems, which fall off the end of the list within 6 month due to the increase in the entry-level performance. We see an average replacement rate of about 180 systems every half year or more than half the list every year. This means that a system, which is at position 100 at a given time, will fall off the TOP500 within two to three years.

When we devised the methodology behind the TOP500, we did not anticipate such a large turnover on such a list. Considering the simplicity of our approach and despite all the limitations an overly simplified performance measure such as the Linpack benchmark has, we can say that the TOP500 approach has worked very well for a long time.

List	Last System on the List Rank #500	Processors	Entry Level R_{max} [GFlop/s]	Replaced Systems
6/1993	Fujitsu VP200	1	0.422	
11/1993	Fujitsu VP200EX	1	0.472	84
6/1994	Cray X-MP	4	0.822	123
11/1994	Cray Y-MP M98	4	1.114	115
6/1995	SGI Power Challenge	8	1.955	216
11/1995	Cray C94	3	2.489	144
6/1996	Convex SPP1000	32	3.306	137
11/1996	SGI Power Challenge	18	4.620	183
6/1997	NEC SX-4	4	7.670	244
11/1997	Sun HPC 10000	22	9.513	129
6/1998	Sun HPC 6000	30	13.390	179
11/1998	Sun HPC 10000	40	17.120	164
6/1999	SGI T3E900	38	24.730	193
11/1999	Sun HPC 10000	48	33.090	222
6/2000	Sun HPC 10000	64	43.820	189
11/2000	IBM SP Power3	52	55.300	232
6/2001	IBM SP Power3	64	67.780	132
11/2001	Cray T3E1200	116	94.300	160
6/2002	Pentium3 Cluster	208	134.300	220
11/2002	HP SuperDome	128	195.000	184
6/2003	HP SuperDome (750 MHz)	128	245.100	234
11/2003	Dell PowerEdge Cluster	128	403.400	210
6/2004	IBM xSeries Cluster	290	634.000	259
11/2004	HP SuperDome (875 MHz)	416	850.600	190
Average				180

Table 3: The replacement rate in the TOP500 defined as number of systems omitted because of their performance being too small.

8.2. Consumer and Producer

The dynamics of the HPC market is well reflected in the rapidly changing market shares of the used chip or system technologies, of manufacturers, customer types or application areas. If we however are interested in where these HPC systems are installed or produced, we see a different picture.

Plotting the number of systems installed in different geographical areas in Fig. 13, we see a rather steady distribution. The number of system installed in the US is slightly increasing over time while the number of systems in Japan is slowly decreasing.

Looking at the producers of HPC system in Fig. 14, we see an even greater dominance of the US, which actually slowly increases it share over time. European manufacturers do not play any substantial role in the HPC market at all. Even the introduction of new architectures such as PC clusters has not changed this picture.

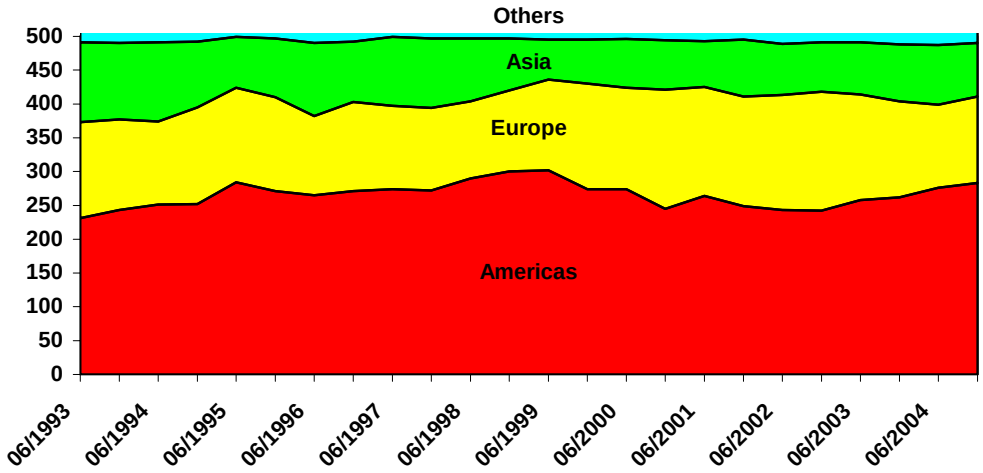

Fig. 13: The consumers of HPC systems in different geographical regions as seen in the TOP500.

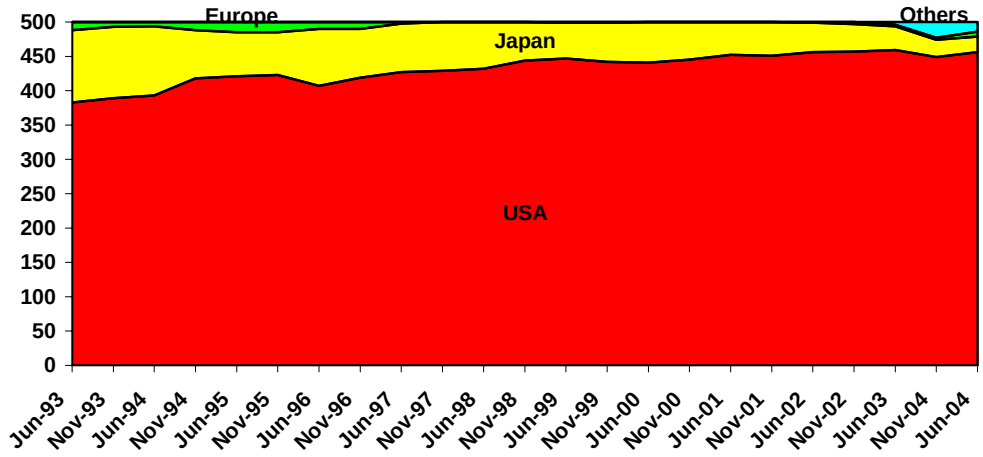

Fig. 14: The producers of HPC systems as seen in the TOP500.

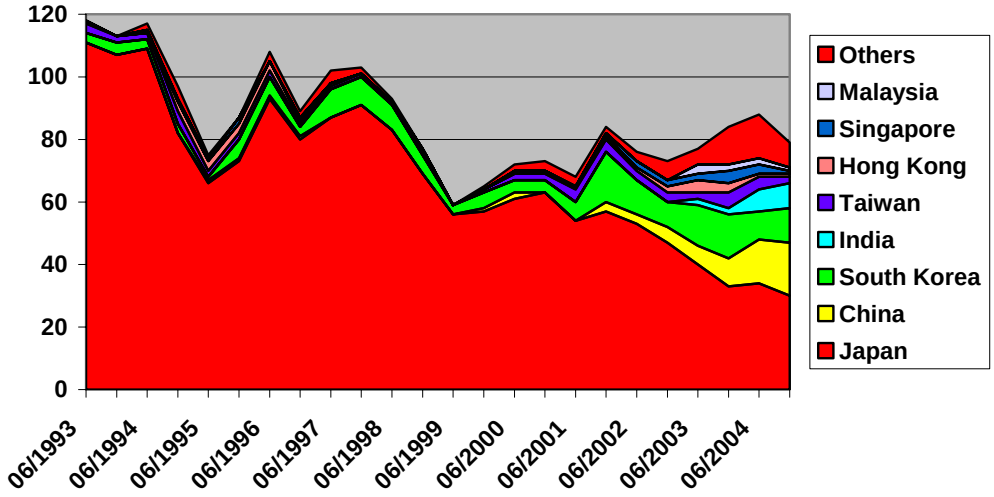

Fig. 15: The consumers of HPC systems in Asia as reflected in the TOP500.

During the last few years, a new geographical trend with respect to the countries using supercomputers is emerging. An increasing number of supercomputers is being installed in upcoming Asian countries such as China, South Korea and India as shown in Fig. 15. 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.

8.3. 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 laws 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. 16, 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.

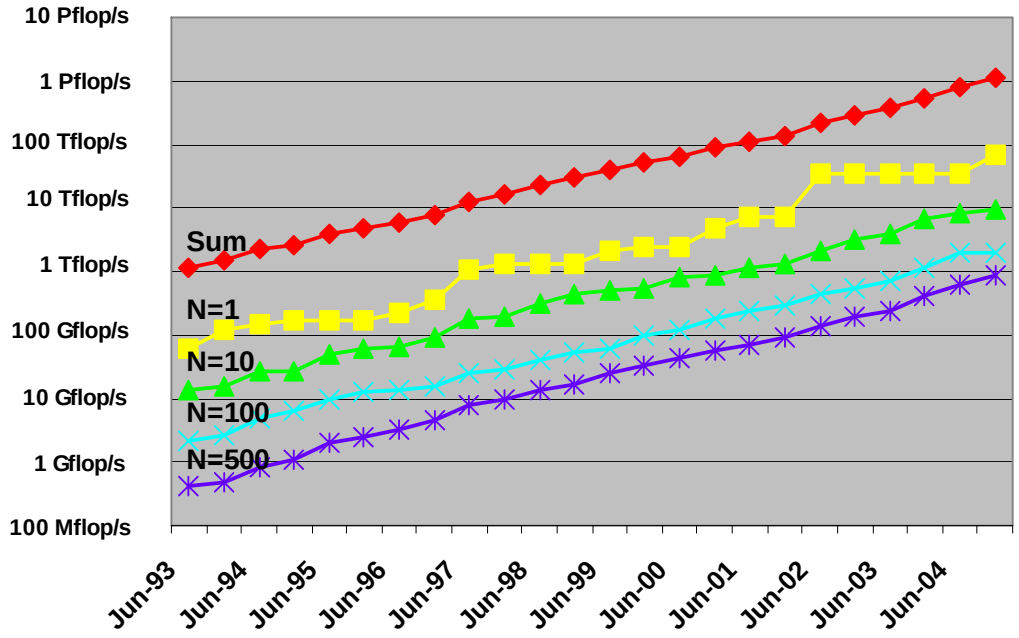

Fig. 16: Overall growth of accumulated and individual performance as seen in the TOP500.

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. 17, 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.

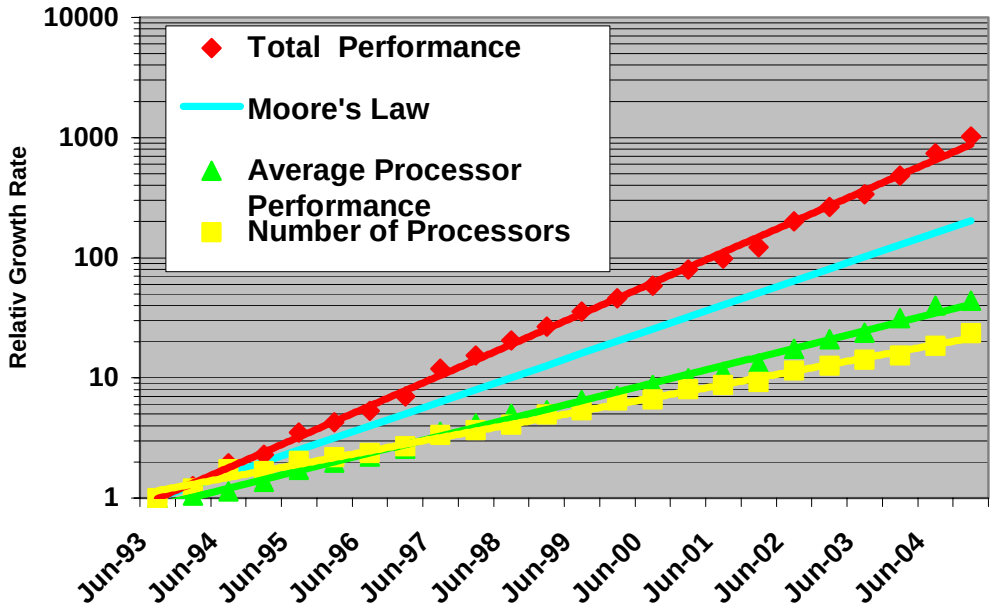

Fig. 17: The growth rate of total accumulated performance, total number of processors and single processor performance for non-SIMD systems as seen in the TOP500.

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 were 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.

8.4. 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. 18, 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.

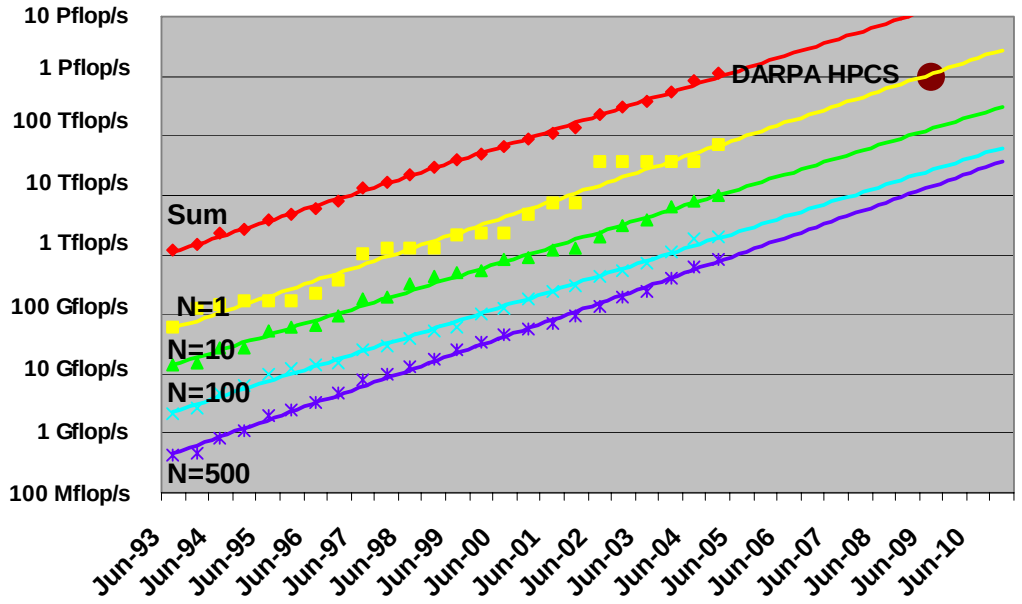

Fig. 18: Extrapolation of recent growth rates of performance seen in the TOP500.

This simple curve fitting of the data shows surprisingly consistent results. In 1999, based on a similar extrapolation [5], 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 certain to be fulfilled this year. Extrapolating over another five-year period to 2010, we expected to see the first PetaFlops system at about 2009 [5] and our current extrapolation is still the same. This coincides with the declared goal of the DARPA HPCS program.

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 three decades ago, there is no end in sight for these rapid cycles of re-definition. And at the end, we still can say that in the High-Performance Computing Market "The Only Thing Constant Is Change".

9. References

[1] G. E. Moore, *Cramming more components onto integrated circuits*, Electronics, Volume 38, Number 8, April 19, 1965

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

[3] H. W. Meuer, E. Strohmaier, J. J. Dongarra, and Horst D. Simon, TOP500, www.top500.org

[4] H. W. Meuer, The Mannheim Supercomputer Statistics 1986–1992, *TOP500 Report 1993*, (University of Mannheim, 1994), 1–15

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

[6] E. Strohmaier, J. J. Dongarra, H. W. Meuer, and H. D. Simon, *Recent Trends in the marketplace of high-performance computing*, Parallel Computing (2005) to appear

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

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

[9] See: www.netlib.org/benchmark/performance.ps

[10] H. D. Simon, High Performance Computing in the U.S., *TOP500 Report 1993*, (University of Mannheim, 1994), 116–147

[11] G. Bell, The Next Ten Years of Supercomputing, *Proceedings 14th Supercomputer Conference*, Mannheim, June 10 -12, 1999, Editor: Hans Werner Meuer, CD-ROM (MaxionMedia), ISBN Nr. 3-932178-08-4

[12] A. J. van der Steen and J. J. Dongarra, Overview of Recent Supercomputers, www.phys.uu.nl/~steen/overv99-web/overview99.html.

[13] 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/)

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

[15] Erich Strohmaier and Hongzhang Shan, "Architecture Independent Performance Characterization and Benchmarking for Scientific Applications", International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems, Volendam, The Netherlands, Oct. 2004

[16] Application Performance Characterization benchmarking, http://ftg.lbl.gov

[17] HPC Challenge Benchmark, http://icl.cs.utk.edu/hpcc/

DISCLAIMER