
ELSEVIER

Parallel Computing 25 (1999) 1517–1544

PARALLEL COMPUTING

www.elsevier.com/locate/parco

The marketplace of high-performance computing

Erich Strohmaier[a,*], Jack J. Dongarra[a,b], Hans W. Meuer[c], Horst D. Simon[d]

[a] *Computer Science Department, University of Tennessee, 107 Ayres Hall, Knoxville, TN 37996, USA*
[b] *Mathematical Science Section, Oak Ridge National Lab., Oak Ridge, TN 37831, USA*
[c] *Computing Center, University of Mannheim, D-68131 Mannheim, Germany*
[d] *NERSC, Lawrence Berkeley Laboratory, 50A, Berkeley, CA 94720, USA*

Received 28 July 1999

Abstract

In this paper we analyze the major trends and changes in the High-Performance Computing (HPC) market place since the beginning of the journal 'Parallel Computing'. The initial success of vector computers in the 1970s was driven by raw performance. The introduction of this type of computer systems started the area of 'Supercomputing'. In the 1980s the availability of standard development environments and of application software packages became more important. Next to performance these factors determined the success of MP vector systems, especially at industrial customers. MPPs became successful in the early 1990s due to their better price/performance ratios, which was made possible by the attack of the 'killer-micros'. In the lower and medium market segments the MPPs were replaced by microprocessor based symmetrical multiprocessor (SMP) systems in the middle of the 1990s. There success formed the basis for the use of new cluster concepts for very high-end systems. In the last few years only the companies which have entered the emerging markets for massive parallel database servers and financial applications attract 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 seems to be no longer sufficient for survival in the market.

Keywords: High-performance computing; HPC market; Supercomputer market; HPC technology; Supercomputer technology

* Corresponding author. Tel.: +1-423-974-0293; fax: +1-423-974-8296.
E-mail addresses: erich@cs.utk.edu (E. Strohmaier), dongarra@cs.utk.edu (J.J. Dongarra), meuer@rz.uni-mannheim.de (H.W. Meuer), simon@nersc.gov (H.D. Simon)

0167-8191/99/$ - see front matter
PII: S0167-8191(99)00067-8

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 (HPC) systems. This market was always characterized by a rapid change of vendors, architectures, technologies and applications. Despite all these changes, however, the evolution of performance on a large scale seems to be a very steady and continuous process. Moore's Law is often cited in this context. If we plot the peak performance of various computers of the last five decades in Fig. 1 which could have been called the 'supercomputers' of their time [4,2] we indeed see how well this law holds for almost the complete lifespan of modern computing. On an average we see an increase in performance of 2 orders of magnitude every decade.

In this paper we analyze the major trends and changes in the HPC market for the last three decades. To do this we focus on systems which have had at least some commercial relevance. Historical overviews with different focus can be found in [8,9].

In the second half of the 1970s the introduction of vector computer systems marked the beginning of modern supercomputing. These systems offered a performance advantage of at least an 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 1980s the integration of vector system in conventional computing environments became more important. Only the manufacturers which provided standard programming environments, operating systems and key applications were successful in getting industrial customers and survived. Performance was mainly increased by

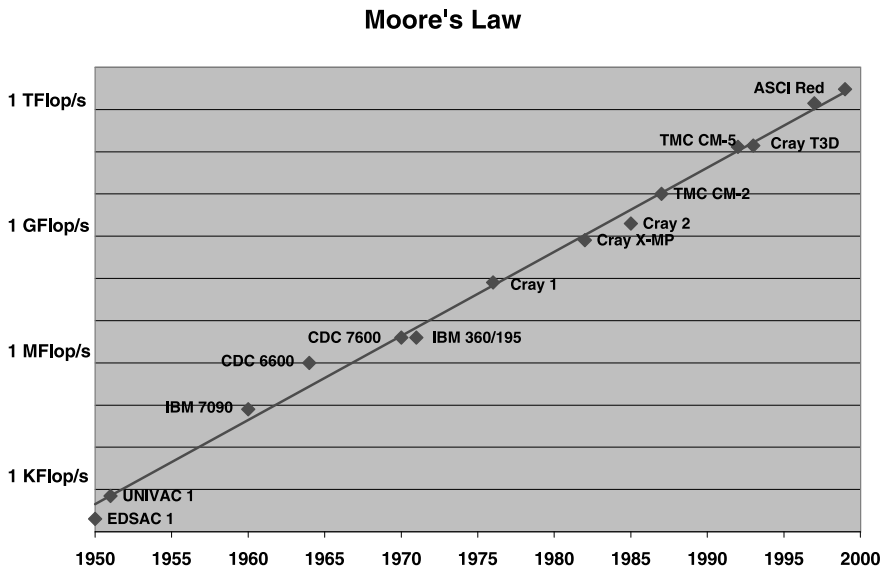

Fig. 1. Performance of the fastest computer systems for the last five decades compared to Moore's Law.

improved chip technologies and by producing shared memory multiprocessor systems.

Fostered by several Government programs, massive parallel computing with scalable systems using distributed memory became the focus of interest at the end of the 1980s. Overcoming the hardware scalability limitations of shared memory systems was the main goal. The increase of 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 use 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 1990s 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. Because of the vertical integration of performance it was no longer economically feasible to produce and focus on the highest end of computing power alone. The design focus for new systems shifted to 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 stability of system, continuity of the manufacturer, and price/performance. Success in commercial environments is now an important new requirement for a successful supercomputer business. 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 are currently replacing homogeneous systems at the very high end of performance. Clusters built with components off the shelf gained more and more attention.

2. 1976–1985: The first vector computers

If one had to pick a single person associated with supercomputing it would without doubt be Seymour Cray. Coming from Control Data Corporation (CDC), where he had designed the CDC 6600 series in the 1960s, 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 main stream 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 main stream systems. For more than a decade supercomputer was a synonym for vector computer. Only at the beginning

of the 1990s would the MPPs be able to challenge or outperform their MP vector competitors.

2.1. 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 1990s including the Cray 2, Cray X-MP, Y-MP, C-90, J-90, T-90 and SV-1. Common feature was not only the usage of vector instructions and vector registers, 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 often that it was not only the fastest vector system but especially also the fastest scalar system of its time. The Cray 1 was also a true Load/Store architecture, a 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 untill 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.

2.2. Cyber 205

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 4 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 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.

2.3. Japanese vector systems

At the end of the 1970s 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 considerable numbers early on in Europe. NEC tried to market their SX1 and SX2 systems by themselves and had a much harder time finding customers outside of Japan. From the beginning Hitachi had decided not to market the S810 system outside of Japan. Common features 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.

2.4. Vector multiprocessor

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

Steve Chen designed the Cray X-MP system first introduced with 2 processors in 1982. The enlarged model with 4 processors was available in 1984. The systems were designed as symmetrical shared memory multiprocessor 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 at Cray Research focused 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 4 processor X-MP. 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 a long running programs not feasible on any other system.

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 manufactures announced multiprocessor models which typically had 2 or at most 4 processors.

3. 1985–1990: The golden age of vector computers

The class of symmetric multivector processor systems dominated the supercomputing arena due to its commercial success in the 1980s.

3.1. Cray Y-MP – Success in industry

The follow-up of the Cray X-MP, the Cray Y-MP, was widely known for the sophistication of its memory access subsystems, which was one of the major reasons for the overall very high efficiency achievable on these systems. With this product line, which later included 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 8 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 using ECL chip technology.

In the beginning of the 1980s the acceptance of the Cray 1 systems was strongly aided by the ease of 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 1980s more independent software vendors started to port their key applications to Cray systems, which was an immense advantage in selling Cray systems to industrial customers. For these reasons Cray vector systems started to have success in various 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.

Looking at the Mannheim supercomputer statistics [3] in Table 1 which counted the worldwide installed vector systems we see the dominance of Cray with a constant market share of 60%. This was confirmed by the first TOP500 list from June 1993 [2] which included MPP systems as well. Cray had an overall share of 40% of all the installed systems which was equivalent to 60% of the included vector systems.

3.2. Cray 3

Seymour Cray left Cray Research 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 led to many development

Table 1
Vector computer installations worldwide [3]

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

problems. In 1992 a single system was delivered. The announced Cray 4 system was never completed.

3.3. ETA

In 1983 CDC decided to spin of its supercomputer business into the subsidiary 'ETA Systems Inc'. The ETA10 system was the follow-up to the Cyber 205 on which it was based. The CPUs had the same design and the systems had up to 8 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 system was delivered in 1987. The largest model had a peak performance of 10 GFlop/s well beyond the competing model of Cray Research.

ETA however seemed 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 ignored the importance of a standard operating system and standard development environments, a mistake which eventually brought CDC itself down.

3.4. Mini-supercomputers

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 supercomputers 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.

3.5. MPP – Scalable systems and the killer-micros

In the second half of the 1980s a new class of systems 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 system early in the 1980s. The 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. A major exception was that the connection network that most vendors

chose used a hypercube topology. Thinking Machine Corporation (TMC) demonstrated their first SIMD system the Connection Machine-1 (CM-1). Intel showed their iPSC/1 hypercube system using Intel 80286 CPU and Ethernet based connection network. nCube produced the first nCube/10 hypercube system with scalar Vax-like custom processors. While these systems were still clearly 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 machines grew very fast. This included the otherwise rare European efforts to produce supercomputer hardware. Companies who started to develop or produce MPP systems in the second half of the 1980s include: TMC, Intel, nCube, Floating Point Systems (FPS), Kendall Square Research (KSR), Meiko, Parsytec, Telmat, Suprenum, MasPar, BBN, and others.

3.6. Thinking Machines

After demonstrating the CM-1 in 1985 TMC soon introduced the follow-on CM-2 which became the first major MPP and was designed by Danny Hillis. In 1987 TMC started to install the CM-2 system. The Connection Machine models 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.

The main disadvantage of all SIMD systems, 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 programs as well.

3.7. Early message passing MPPs

From the start competing MPP manufacturers had decided to produce MIMD systems. The more complex programming of these systems was more than compensated for by the much greater flexibility in supporting different programming paradigms efficiently.

Intel built systems based on the different generations of Intel microprocessors. The first system 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 microprocessors, 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 had a typical CISC design. To compensate for the relatively low performance of this processor, the maximal number of processors possible was quite high. The 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 a follow on custom designed processor, was introduced in 1990.

3.8. Attack of the killer-micros

Beginning of the 1990s 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. But not only the traditional mainframes started to feel the heat. Even the traditional supercomputers the vector multiprocessors 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 though sustained performance values did not nearly come up to peak advertised performance (PAP) values, the direction of the attack was clear. The new generation of microprocessors, manufactured relatively inexpensively 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 vector multiprocessors could be achieved for a better price.

This contrast was greatly aggravated by the fact that most mainframe manufacturers had not seen early enough the advantage of CMOS chip technology and were still using the (somewhat) 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 in order to compete with the up and coming new MPP vendors.

Hidden behind these slogans were actually two different trends working together. Both effects can clearly be seen in the TOP500 data. The replacement of ECL chip technology by CMOS is shown in Fig. 2. The rapid decline of vector based systems in the 1990s can be seen in the TOP500 in Fig. 3. This change in the processor architecture is however not as generally accepted as the change from ECL to CMOS. The Japanese manufacturers together with SGI still continue to produce vector based systems.

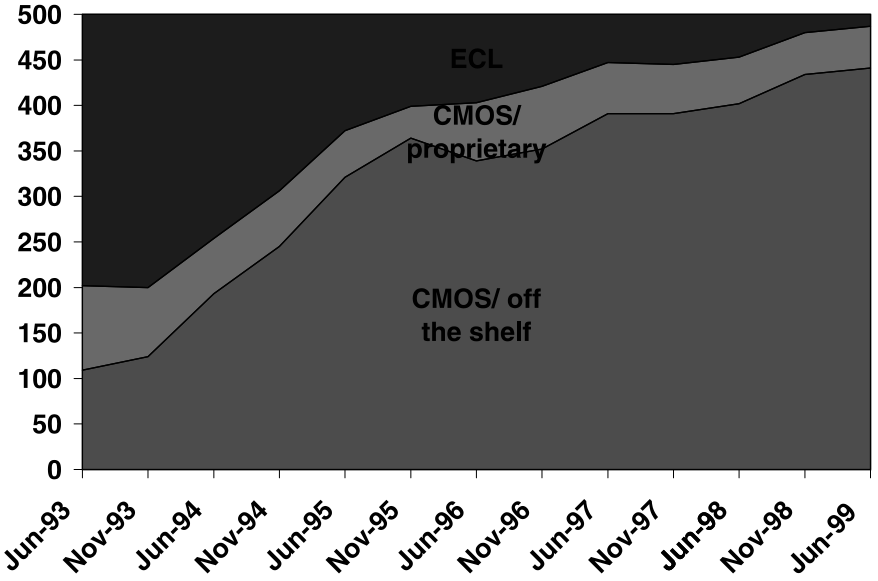

Fig. 2. Chips technologies used as seen in the TOP500.

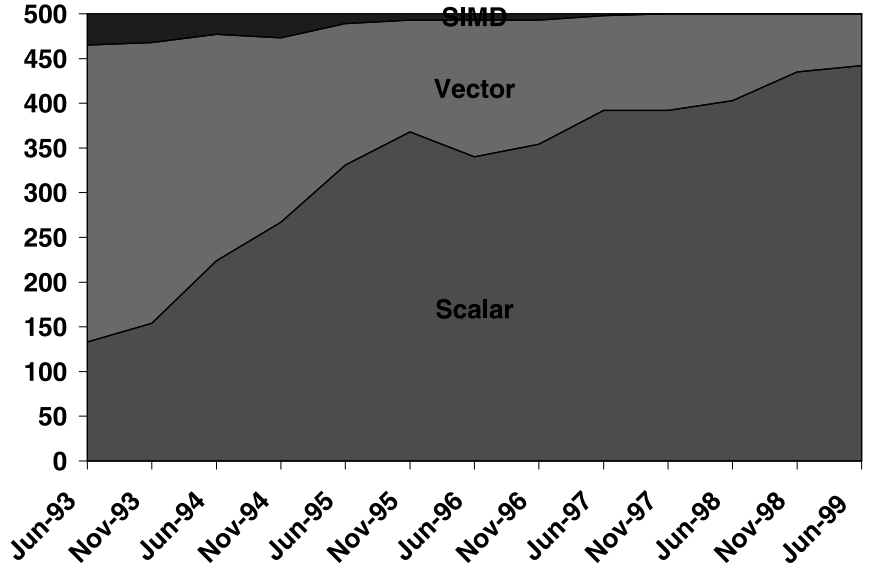

Fig. 3. CPU design used as seen in the TOP500.

4. 1990–1995: MPP comes to age

In the beginning of the 1990s, while the MP vector systems reached their widest distribution, a new generation of MPP systems came on the market claiming to be able to substitute or 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 the major reason for starting the TOP500 project [2]. Here we list since 1993 twice a year the 500 most powerful installed computer systems ranked by the best LINPACK performance [5]. 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, but 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 earlier MPPs would be under attack.

One major side-effect of the introduction of the MPP systems in the supercomputer market was that the sharp performance gap between supercomputers and mainstream computers no longer existed. MPPs could – by definition – be scaled by 1 or 2 orders of magnitude, bridging the gap between high-end multiprocessor workstations and supercomputers. Homogeneous and monolithic architectures designed for the very high end of the performance spectrum would have to compete with clustered systems based on shared memory multiprocessor models from the traditional UNIX server manufacturers. These clusters offered another level of price/performance advantage: They had a large supporting industrial and commercial business market in the background and they could reduce design costs by not focusing on the very high end of performance any more. This change towards the use of standard components widely used in the commercial market place actually widened the market scope for such MPPs in general. As a result, the notion of a separated market for floating point intensive HPC no longer applies. The HPC market is now the upper end of a continuum of systems which run in all kinds of applications.

4.1. Playground for manufacturers

With the increased number of MPP manufacturers it was evident that a "shake-out" of manufacturers was unavoidable. In Table 2 we list vendors which have been active at some point in the HPC market [6,1]. In Fig. 4 we try to visualize the historical presence of companies in the HPC market. After 1993 we include only companies which had entries in the TOP500 and were actively manufacturing.

From the 14 major companies in the early 1990s only 4 survived this decade on their own. These are the 3 Japanese vector manufacturer (Fujitsu, Hitachi, and NEC) and IBM, which due to its marginal HPC presence at the beginning of the 1990s could even be considered a newcomer. Four other companies entered the HPC market either by buying some companies or by developing their own products

Table 2
Commercial HPC vendors

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, KSR, Multiflow, Myrias, Pixar, Prevec, Prisma, Saxpy, SCS, SDSA, Suprenum, Stardent (Stellar and Ardent), Supercomputer Systems, Synapse, Thinking Machines, Trilogy, VItec, Vitesse, Wavetracer
Merged	Celerity (with FPS), Convex (with Hewlett-Packard), Cray Research (with SGI), DEC (with Compaq), Floating Point Systems (with Cray Research), Key (with Amdahl), MasPar (with DEC), Meiko, Supertek (with Cray Research)
Changed market	nCUBE, Parsytec
Currently active	Compaq, Fujitsu, Hewlett-Packard, Hitachi, IBM, NEC, SGI, Siemens, Sun, Tera

(Silicon Graphics, Hewlett-Packard, Sun, Compaq). None of these is a former HPC manufacturer. All are 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 indicate the change in focus from the very high end to the market for medium size HPC systems.

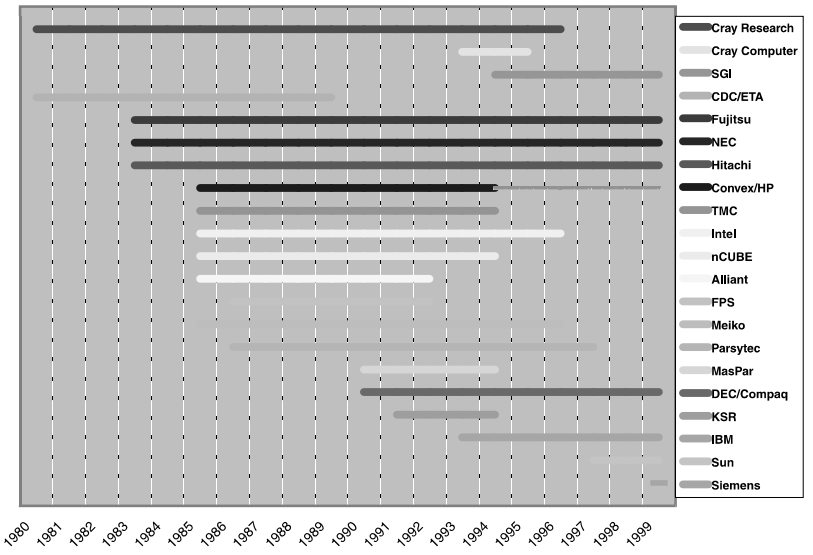

Fig. 4. Main manufacturer active in the HPC market.

4.2. KSR

In 1991 the first models of a quite new and innovative system was installed, the KSR1 from Kendall Square Research (KSR). The hardware was built like other MPP with distributed memory but gave the user the view of a shared memory system. The custom designed hardware and the operating system software were responsible for this virtual shared memory appearance. This concept of virtual shared memory would later appear 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. This way the data had no fixed home in the machine and could freely roam to the location where it was needed. 'Management mistakes' brought the operations of KSR to an abrupt end in late 1994.

4.3. Intel

In 1992 Intel started to produce the Paragon/XP series, having delivered the Touchstone Delta system to Caltech in 1991. Still based on the i860 chips, the interconnection network was changed to a 2-dimensional grid which now allowed up to 2048 processors. Several quite large systems 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.

4.4. Thinking Machines

Also in 1992 TMC started to deliver the CM-5 an 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 number one on the very first TOP500 list in June 1993, achieving almost 60 GFlop/s. A Sun SPARC processor was chosen as basic node CPU. Each of these processors had 4 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 the customers could handle. Because the design point was at 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 positions 1–4 were all held by TMC CM-5 systems, ahead of the first MP vector system, a 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 5 of the first 10 system were MP vector systems. This number decreased quickly and the last MP vector system which managed to make the top 10 was an 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 CPU 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 CPUs was an NEC SX-4 with 128 processor at number 18.

4.5. 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 as being a workstation cluster and not an MPP, it set the ground for the re-entry 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, which managed to sell models of the SP quite successfully as database server. Due to the design of the SP2, IBM is able to constantly offer new nodes based on the latest RISC system available.

4.6. 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 3-dimensional torus. As CPU Cray had chosen the DEC alpha processor. The design of the node was completely done by Cray 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 this allowed for very efficient parallel processing. The computational node performance however was greatly affected by the missing second level cache. The system was immediately well accepted at research laboratories and was even installed at some industrial sites. The largest configuration known (1024 processors) is installed at a classified government site in the USA and just breaks the 100 GFlop/s barrier on the LINPACK.

4.7. 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 provided the view of a non-uniform shared memory over the whole machine. In the succeeding 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.

4.8. 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 a 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 started 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 till the introduction of the CMOS based VPP500 in 1994. NEC introduced the SX-3 series in 1990 as well. With 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 a 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 have been 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.

4.9. 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 CPU of their successful VP2600 vector computer series. However Fujitsu was the first Japanese company which implemented 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 4, and number 2 position three times from November 1993 to November 1996. Delivery of VPP500 systems started in 1993.

5. After 1995: Old and new 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. 5).

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 reflects 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 have 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 was initially 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 is a strong new trend for the following reasons:

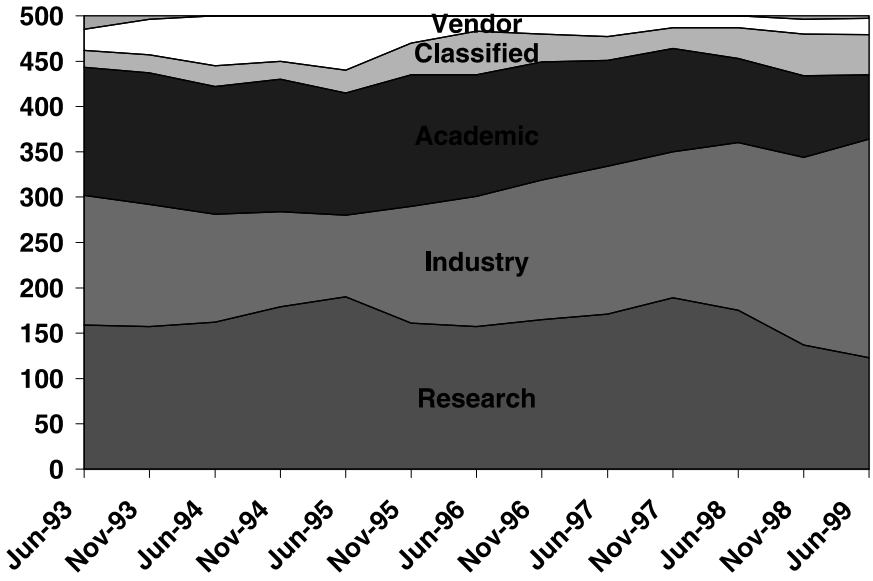

Fig. 5. The number of systems on the different types of customers over time.

- The architectures installed at industrial sites changed from vector systems to a substantial number of MPP systems. This change reflects the fact that parallel systems are ready for commercial use and environments.
- The most successful companies (Sun, IBM and SGI) are selling well to industrial customers. Their success is built on the fact that they are using standard workstation technologies for their MPP nodes. This approach provides 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 make the systems appealing to commercial customers. Especially important are database applications, since these can use highly parallel systems with more than 128 processors.

Fig. 6 shows that the increase in the number of systems installed at industrial sites is 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 are typically smaller than systems at research laboratories and universities, their average performance and size are growing at the same rate as at research installations. The strong increase in the number of processors in systems at industrial sites is another major reason for the rise of industrial sites in the TOP500. Industry is ready to use bigger parallel systems than in the past.

5.1. Architectures

The changing share of the different system architectures in the HPC market as reflected in the TOP500 is shown in Fig. 7. Besides the fact that no single processor

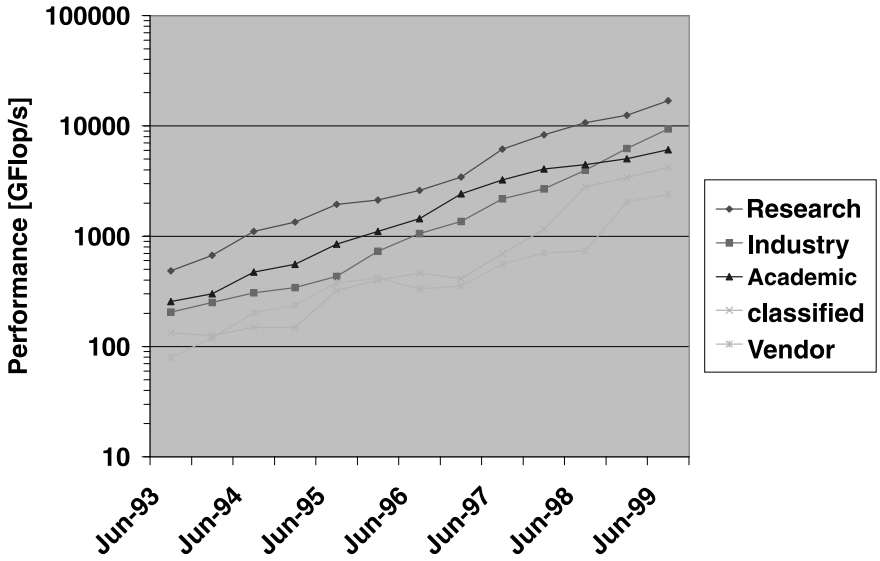

Fig. 6. The accumulated performance of the different types of customers over time.

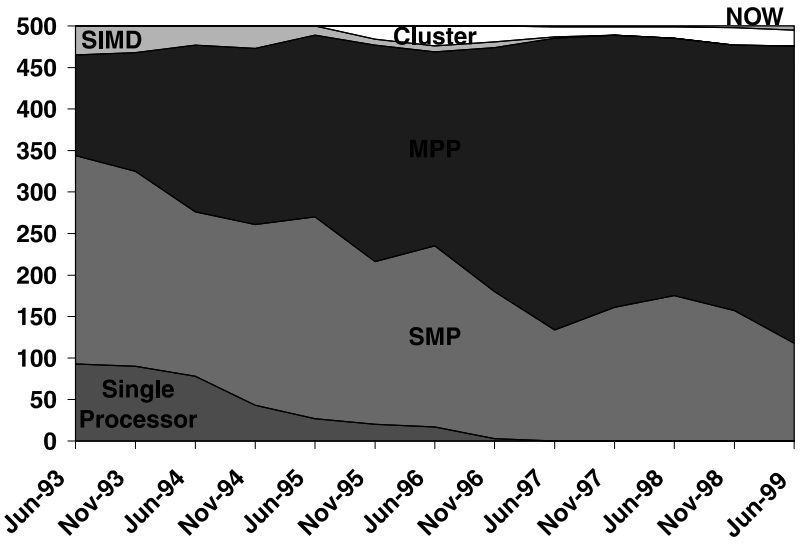

Fig. 7. Architectures of systems in the TOP500.

systems are any longer powerful enough to enter the TOP500 the major trend is the growing number of MPP systems. The number of clustered systems is also growing and for the last two years we have seen a number of PC or workstation based 'Network of Workstations' in the TOP500 It is an interesting and open question how large a share of the TOP500 such NOWs or PC clusters will capture in the future.

5.2. 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 economic 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 macroarchitecture for its microprocessor and vector processor based MPPs. In mid-1998 the SGI SV1 was announced as future vector system of SGI. The SV1 is the successor both to the CMOS based Cray J-90 and the ECL based, Cray T-90. The SV1 is CMOS based, which means that SGI is finally following the trend set in by Fujitsu (VPP700) and NEC (SX-4) a few years earlier. First user shipments are be expected by the middle of 1999. Only time will tell whether the SV1 can 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 in 1996 introduced 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 recently announced VPP5000 is again a distributed-memory vector system, where from 4 to 128 processors (512 by special order) may be connected via a fully distributed crossbar. The theoretical peak performance ranges 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 continues along this line of systems. The SX-4 features shared memory up to a maximum of 32 processors. Larger configurations are built as clusters using a proprietary crossbar switch. In 1998 the follow-up model SX-5 was announced for first delivery in early 1999. In the June 1999 TOP500 list the SX-5 was not yet represented. In contrast to its predecessor, the SX-4, the SX-5 is not offered anymore with faster, but more expensive SRAM memory. The SX-5 systems are exclusively manufactured with synchronous DRAM memory. The multiframe version of the SX-5 can 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 architectures can be found in [7].

5.3. Traditional MPPs

Large scale MPPs with homogeneous system architectures had matured during the 1990s 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. At the end of 1998 the top 10 consisted only of ASCI systems and T3Es.

Hitachi was one of the few companies introducing large scale homogeneous systems in the late 1990s. It announced the SR2201 series in 1996 and tried to sell this system outside of Japan as well.

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

IBM followed the path of their SP series introducing new nodes and faster interconnects as well. 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.

5.4. SMPs and their successors

Beginning in 1994, however, companies such as SGI, Digital, and Sun started to sell 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 was initially 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. Clusters built with these SMPs appeared in a reasonable number at customer sites.

In 1996 the Origin2000 series was announced. With this system SGI took a step away form the bus based SMP design of the Challenge series. The Origin series features a virtual memory system built with distributed memory nodes up to 128 processors. To achieve higher performance these systems can be clustered again. This is 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, making Compaq the first PC manufacturer with 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 has had good success. The very high end – which had never been a target for Convex or HP – still seems to be of minor interest to HP.

Sun was the latest 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 is built around a new type of switched bus which allows it to integrate up to 64 processors in an efficient way. Because of 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. Clusters built with these SMPs for the very high-end market were introduced in 1998.

5.5. 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 a single major application program. 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 over time to track the application area for almost 90% of the industrial systems.

Since June 1995 we have seen many systems involved in new application areas entering the list. Fig. 8 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.

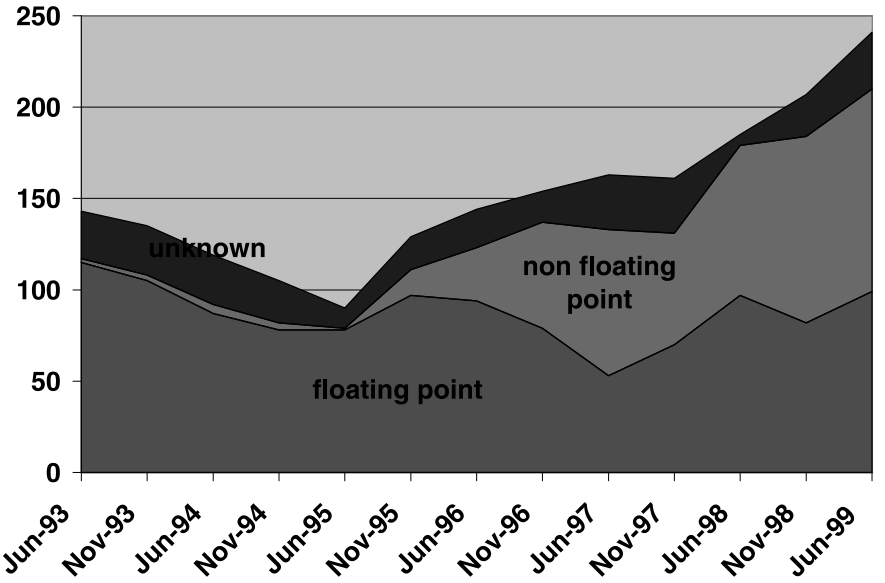

Fig. 8. The total number of systems at industrial sites together with the numbers of sites with traditional engineering applications, new emerging application areas and unknown application areas.

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. Recently, however, industrial systems in the TOP500 have been used for new application areas. These include

- database applications,
- financial applications, and
- image processing and rendering.

The most dominant trend is the strong rise of database applications since November 1995. These applications include on-line transaction processing as well as data mining. The HPC systems being sold and installed for such applications are large enough to enter the first 100 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 are 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 is no longer a hindrance to success in the commercial marketplace.

6. 2000 and beyond

Two decades after the introduction of the Cray 1 the HPC market has changed its face quite a bit. It used to be a market for systems clearly different from any other computer systems. Nowadays the HPC market is no longer an isolated niche market for specialized systems. Vertically integrated companies produce systems of any size. Similar software environments are available on all of these systems. This was the basis for a broad acceptance at industrial and commercial customers.

The increased market share of industrial and commercial installations had several very critical implications for the HPC market. In the market for small to medium size HPC systems the manufacturers of supercomputers for numerical applications face the strong competition from manufacturers selling their systems in the very lucrative commercial market. These systems tend to have better price/performance ratios due to the larger production numbers of systems purchased by commercial customers and the reduced design costs of medium size systems. The market for the very high-end systems itself is relatively small and does not grow. It cannot easily support specialized niche market manufacturers. This forces the remaining manufacturers to change the design for the very high end away from homogeneous large scale systems towards cluster concepts based on medium size 'off-the-shelf' systems.

Currently the debate about whether we need a new architecture for the very high-end supercomputer, such as the multithreaded design of Tera, still goes on. At the same time we see a new architecture appearing in growing numbers in the TOP500 – the 'Network of Workstations' including PC based systems. Depending on where one draws the line between a cluster of SMPs and a network of workstations, we have about 6 such NOWs in the June 1999 edition of the TOP500. As there is not a single manufacturer who provides LINPACK measurements for such systems we certainly underestimate the actual number of NOW large enough to fit in the TOP500 The potential of this class of HPC architecture with excellent price/performance ratios should not be underestimated and we expect to see more in the future. It is an open question, however, if they however will seriously challenge the clusters of large SMPs for a wide range of application.

6.1. 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 general performance increases. 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 months due to the increase in the entry level performance. We see an average replacement rate of about 160 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 2–3 years.

Table 3
The replacement rate in the TOP500, defined as the number of systems omitted because their performance has become too low

List	Last system Rank #500	Entry level R_{max} [GFlop/s]	Replaced systems
6/93	Fujitsu VP200	0.422	
11/93	Fujitsu VP200EX	0.472	83
6/94	Cray X-MP/4	0.822	121
11/94	Cray Y-MP/M4	1.114	115
6/95	SGI PC/8	1.955	216
11/95	Cray C94/3	2.489	144
6/96	Convex SPP1000/32	3.306	137
11/96	SGI PC/18	4.620	183
6/97	IBM SP2/32	7.680	245
11/97	Sun Ultra HPC 10000	9.513	129
6/98	Sun HPC 10000	13.390	180
11/98	Sun HPC 10000	17.120	162
6/99	SGI T3E900	24.730	194
Average			159

6.2. Consumer and producer

The dynamic 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, however, we 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. 9 we see a rather steady distribution. The number of systems in 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. 10 we see an even greater dominance of the US, which in fact slowly increases over time. European manufacturers do not play any substantial role in the HPC market at all.

6.3. Government programs

The high end of the HPC market was always the target for government programs all over the world to influence the further development of new systems. Currently in the USA there are several government projects attempting to consolidate and advance the numerical HPC capabilities of US government laboratories and the US research community in general. The most prominent of these is the 'Accelerated Strategic Computing Initiative (ASCI)'. The 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 US nuclear stockpile and reducing the nuclear danger".[1]

[1] www.llnl.gov/asci/

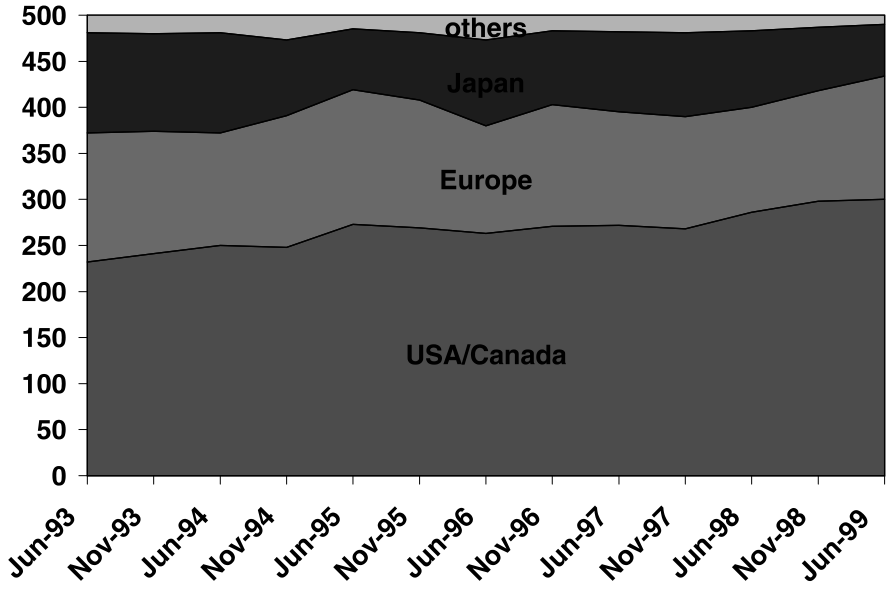

Fig. 9. The consumers of HPC systems in different geographical regions as seen in the TOP500.

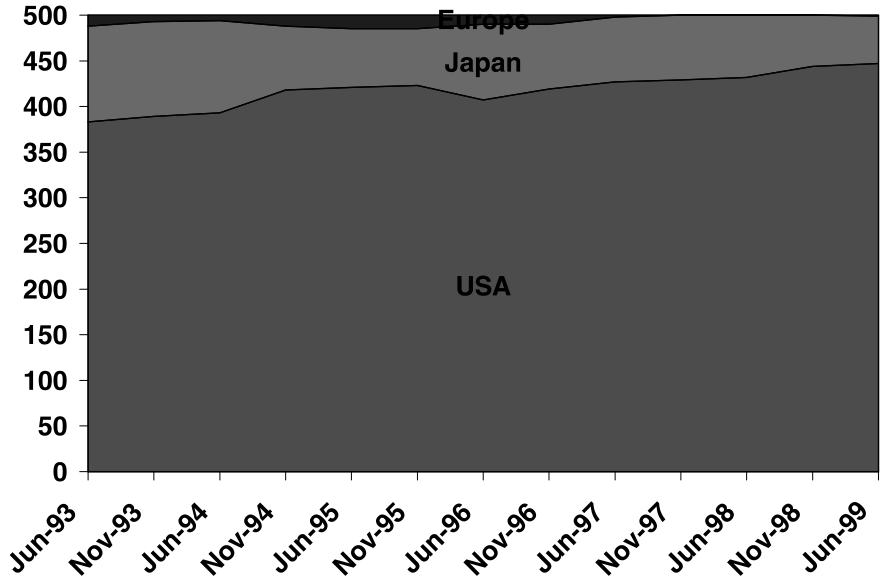

Fig. 10. The producers of HPC systems as seen in the TOP500.

Three main laboratories were selected as sites for deploying parallel computing systems of the largest size technically possible. Currently the system 'ASCI Red' is installed at Sandia National Laboratory. This system is produced by Intel and currently has 9472 Intel Pentium Xeon processors in 1997. It was the first system to exceed the 1 TFlop/s mark on the LINPACK benchmark. Since then ASCI Red has been the number 1 on the TOP500 with currently 2.1 TFlop/s. 'ASCI Blue Mountain' installed at the Los Alamos National Laboratory is a cluster of Origin2000 systems produced by SGI with a total of 6144 processors. It achieves 1.6 TFlop/s LINPACK performance. 'ASCI Blue Pacific' installed at the Lawrence Livermore National Laboratory is an IBM SP system with a total of 5856 processors. The future plans of the ASCI program call for a 100 TFlop/s system by April 2003.

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.

6.4. 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 in Section 1. The TOP500 provides an ideal data set to verify an observation like this. 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. 11 the curve of position 500 shows on the average an increase of a factor of two within one year. All other curves show a growth rate of 1.8 ± 0.07 per year.

To compare these growth rates with Moore's Law we now separate the influence from the increasing processor performance and from the increasing number of processor per system on the total accumulated performance. To get meaningful numbers we exclude the SIMD systems for this analysis since they tend to have extremely high processor numbers and extreme low processor performance. In Fig. 12 we plot the relative growth of the total processor number and of the average processor performance defined as the quotient 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 1.82. The processor number grows per year on the average by a factor of 1.30 and the processor performance by 1.40 compared 1.58 of Moore's Law.

The average growth in processor performance is lower than we expected. A possible explanation is that during the last few years powerful vector processors got replaced in the TOP500 by less powerful super-scalar RISC processors. This effect might be the reason why the TOP500 currently does not reflect the full increase in RISC performance. Once the ratio of vector to super-scalar processors in the TOP500 stabilizes we should see the full increase in processor performance reflected in the TOP500 The overall growth of system performance is however larger than

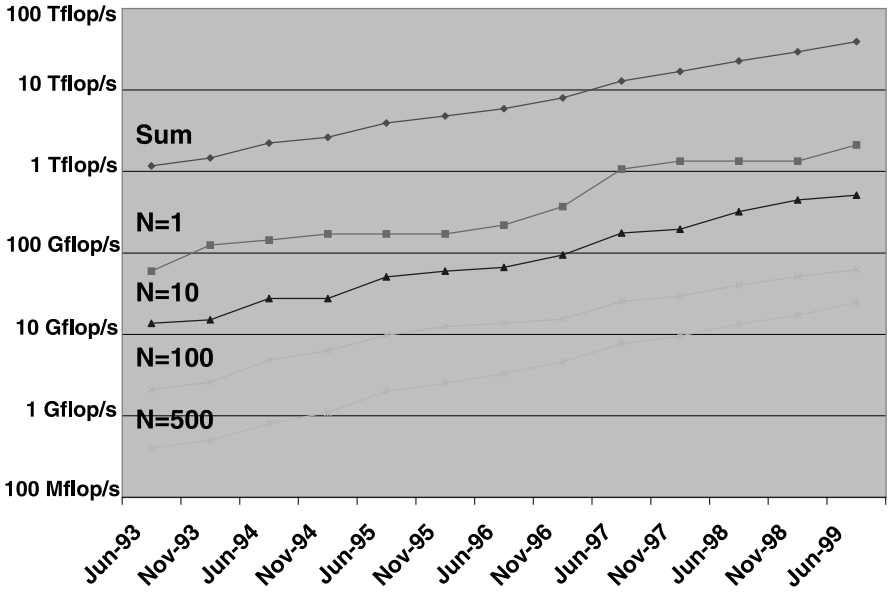

Fig. 11. Overall growth of accumulated and individual performance as seen in the TOP500.

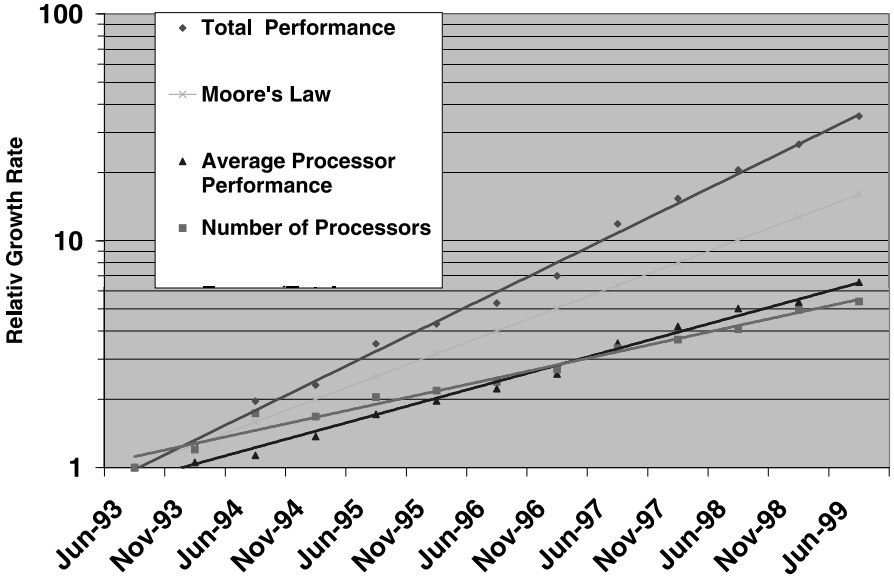

Fig. 12. The growth rate of total accumulated performance, total number of processors and single processor performance for non-SIMD systems as seen in the TOP500.

expected from Moore's Law. This comes from growth in both dimensions: processor performance and the number of processors used.

6.5. Projections

Based on the current TOP500 data, which cover the last 6 years, and the assumption that the current performance developments will continue for some time to come, we can now extrapolate the observed performance and compare these values with the goals of the above mentioned government programs. In Fig. 13 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. This simple fitting of the data shows surprisingly consistent results. Based on the extrapolation from these fits we can expect to have the first 100 TFlop/s system by 2005 which is about 1–2 years later than the ASCI path forward plans. By 2005 also no system smaller then 1 TFlop/s should be able to make the TOP500 any more.

The only performance level which deviates with its growth rate from all others is the performance of position 100. In our opinion this is actually a sign of a lower number of centers being able to afford very high-end systems. This concentration of computing power in fewer centers has already been seen in the US for a couple of years in basically all government programs.

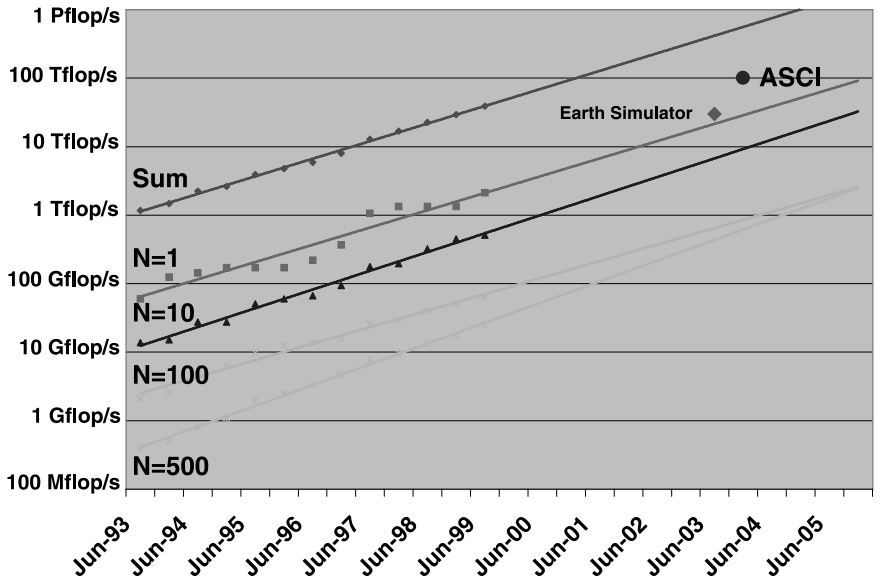

Fig. 13. Extrapolation of recent growth rates of Linpack performance seen in the TOP500, together with peak performance of planned systems.

Looking even further in the future we speculate that based on the current doubling of performance every year, the first PetaFlop system should be available around 2009. Due to the rapid changes in the technologies used in HPC systems, however, there is, at this point in time no reasonable projection possible for the architecture of such a system at the end of the next decade. Even though the face of the HPC market has changed 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 we still can say that in the HPC market The Only Thing Constant Is Change.

References

[1] G. Bell, The next ten years of supercomputing, in: Hans Werner Meuer (Ed.), Proceedings of the 14th Supercomputer Conference, Mannheim, 10–12 June 1999, CD-ROM (MaxionMedia), ISBN Nr. 3-932178-08-4.
[2] J.J. Dongarra, H.W. Meuer, E. Strohmaier, TOP500, www.TOP500.org.
[3] H.W. Meuer, The Mannheim supercomputer statistics 1986–1992, TOP500 Report 1993, University of Mannheim, 1994, pp. 1–15.
[4] R.W. Hockney, C. Jesshope, Parallel Computers II: Architecture, Programming and Algorithms, Adam Hilger, Bristol, UK, 1988.
[5] See: www.netlib.org/benchmark/performance.ps.
[6] H.D. Simon, High performance computing in the US, TOP500 Report 1993, University of Mannheim, 1994, pp. 116–147.
[7] A.J. van der Steen, J.J. Dongarra, Overview of recent supercomputers, www.phys.uu.nl/ steen/overv99-web/overview99.html.
[8] G.V. Wilson, Chronology of major developments in parallel computing and supercomputing, www.unipaderborn.de/fachbereich/AG/agmadh/WWW/GI/History/history.txt.gz.
[9] P.R. Woodward, Perspectives on supercomputing, Computer 29 (10) (1996) 99–111.