

THE TOP500 PROJECT: LOOKING BACK OVER 15 YEARS OF SUPERCOMPUTING EXPERIENCE

Hans Werner Meuer, University of Mannheim & Prometheus GmbH, Germany
January 20, 2008

Abstract

The TOP500 project was launched in 1993 to provide a reliable basis for tracking and detecting trends in high performance computing. Twice a year, a list of the sites operating the world's 500 most powerful computer systems is compiled and released. The best performance on the Linpack benchmark is used as the measurement for ranking the computer systems. The list contains a variety of information including the systems' specifications and major application areas. Information on all 30 TOP500 lists issued to date is available at: www.top500.org

1. Mannheim Supercomputer Statistics 1986–1992 and TOP500 Project Start in 1993

From 1986 through 1992, the Mannheim supercomputer statistics were presented to participants of the Supercomputer Seminars at Mannheim University, and we noticed an increased interest in this kind of data from year to year.^[1] The statistics simply counted the vector computer systems installed in the U.S., Japan and Europe, since in the mid-80s a supercomputer was synonymous with a vector computer. Counting the vector computers installed worldwide primarily depended on the input provided by the manufacturers of the systems, which made the statistics less reliable. Whereas we knew well which vector systems existed in the U.S. and Europe, information regarding systems in Japan was much more difficult to collect. We therefore contacted the three Japanese vector computer manufacturers – Fujitsu, NEC and Hitachi – for information on all systems installed in Japan and used their data as the basis for our yearly estimations.

In 1992, we released the last Mannheim statistics, counting 530 supercomputers installed worldwide. **Figure 1** shows the result of our 7-year activity regarding the share of the different manufacturers in the supercomputer market. Cray clearly led with a constant share of about 60%; the second U.S. manufacturer, CDC (Control Data Corporation), had been doing rather well with just under 10% – until the end of the 80s when their share started to drop, and they were completely out of the supercomputer business in 1991. The Japanese vector computer manufacturers Fujitsu, NEC and Hitachi entered into our statistics in 1986 with a combined share of 20%, and were able to expand their share to about 40% in 1992, with Fujitsu clearly in the lead at 30% of all vector computers installed worldwide.

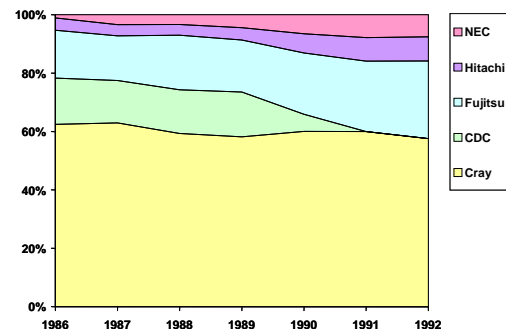


Figure 1: Manufacturers' shares

Figure 2 illustrates the shares of vector computer installations by country. The U.S. clearly led in 1986 with a share of 50%, but which dropped to 35% in 1992, however. In Japan, the situation developed in the opposite direction, with a share of about 20% in 1986 and already 40% in 1992, surpassing the U.S. share. Europe had a constant share of between 25% and 30% over the seven years, with Germany leading slightly ahead of France and the U.K.

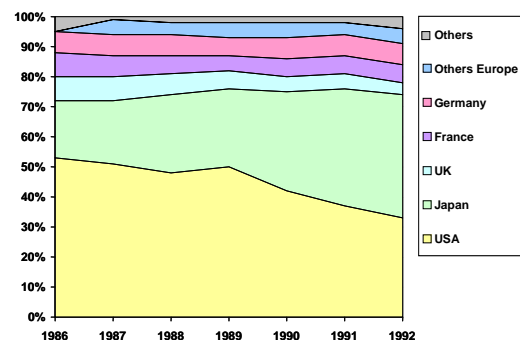


Figure 2: Countries' shares

Though useful, the Mannheim supercomputer statistics were not perfect, as they lacked a reliable database. Additionally, the so-called entry level vector computer systems such as Fujitsu's VP30/50 became more and more popular in Japan. But were these systems really supercomputers in terms of performance? And how should mini-supercomputers such as the Convex C1/2 from the U.S. be rated? We had to carefully consider which systems qualified as supercomputers and therefore should be listed in the Mannheim statistics. From the early 90s on, vector computers were no longer the only supercomputer architecture; massively parallel systems such as the CM2 of Thinking Machines (TMC) had entered the market. What we therefore needed was a method to define what constituted a "supercomputer" and could be updated on a yearly basis.

This is why Hans Werner Meuer and Erich Strohmaier started the TOP500 project at the University of Mannheim/Germany, in spring 1993. Here are its simple guiding principles:

- Listing of the 500 most powerful computers in the world
- Rmax, the best Linpack performance, is used as the benchmark^[2]
- The TOP500 list is updated and published twice a year, in June at ISC in Germany and in November at SC in the U.S.
- All TOP500 data is publicly available at www.top500.org

There are some immediate questions that we would like to answer here:

- Why is it “the 500 most powerful computers”? One reason is that the last time we counted the supercomputers worldwide in 1992, we ended up with 530. And another reason surely is the (emotional) influence of the Forbes 500 lists, e.g. of the 500 richest men or the 500 biggest corporations in the world.
- “Most powerful” is defined by a common benchmark, for which we had chosen Linpack. But why Linpack? Linpack data, above all Rmax, are well known and easily available for ALL systems in question. Strictly speaking, TOP500 lists computers only by their ability to solve a set of linear equations, $Ax = b$, using a dense random matrix A.
- An alternative to updating the TOP500 list twice a year would be to continuously update the list. Why don't we do this? First, updating the TOP500 list is a time-consuming and complex process. Second, we thought that a biannual publication would be a much better way to show significant changes, which the HPC community is primarily interested in, and this has proven to be true over the years.

TOP500 authors are Hans Werner Meuer, Erich Strohmaier, now Lawrence Berkeley National Laboratory (LBNL), USA, and Jack Dongarra, the “Father of Linpack”, University of Tennessee, USA. The fourth author, Horst Simon, LBNL, had supported the TOP500 project from the very beginning and joined the project officially in 2000. In 1999, only six years after starting the project, the authors published their experiences with TOP500.^[3]

30 TOP500 lists have been published up to now:

- First TOP500 list was published on June 24, 1993, at ISC'93 in Mannheim, Germany
- 29th TOP500 list was published on June 27, 2007, at ISC'07 in Dresden, Germany
- 30th TOP500 list was published on November 12, 2007, at SC07 in Reno, USA

The release dates of the next three TOP500 lists are:

- 31st TOP500 list will be published on June 18, 2008, in Dresden, Germany
- 32nd TOP500 list will be published on November 18, 2008, in Austin, USA
- 33rd TOP500 list will be published on June 24, 2009, in Hamburg, Germany

After 15 years and 30 lists, we have managed to establish TOP500 among HPC users, manufacturers and the media as THE instrument for analyzing the HPC market.

2. Competition between Countries, Manufacturers and Sites

One of the most important reasons for TOP500's success is that we foster competition between countries, manufacturers and computing sites.

2.1 Competition between Countries

From our 7th Mannheim supercomputer statistics published at the Mannheim Supercomputer Seminar in 1992, we expected a neck-and-neck race between the U.S. and Japan for our first TOP500 list (see **Figure 3**). However, the “Japanese danger” was grossly overestimated, as the first TOP500 list showed the U.S. clearly leading with 45% of all TOP500 installations, and Japan was far behind with only 22%.

If we look at the 30th TOP500 list published in November 2007 at SC in Reno/USA, we see that the dominance of the U.S. is even bigger today than 15 years ago: Now they have a share of 56.6 % of all systems installed, and Japan holds a share of only 4%. Even the U.K., with a 9.6% share, and Germany, with a 6.2% share, are ahead of Japan, which is followed closely by France with 3.4%.

The overall development of the various countries' share through the past 30 TOP500 lists is also very interesting (see **Figure 4**). In 1993, the U.S. started with a huge share of 45%, which they have even managed to expand slightly. Japan, however, started with a 22% share but has fallen back significantly. In Europe, Germany, which had always clearly been ahead of the U.K., is now far behind the U.K.

Figure 5 illustrates the development of the supercomputer installations in Asia since 1993. It shows the rapid drop in Japan's share and indicates that China and India will enter the HPC market as new players in the medium term. But we will have to wait until the next TOP500 lists to see how this plays out.

1st TOP500 List, 06/1993

Country	Count	Share
USA	225	45.0%
Japan	111	22.2%
Germany	59	11.8%
France	26	5.2%
U.K.	25	5.0%
Australia	9	1.8%
Italy	6	1.2%
Netherlands	6	1.2%
Switzerland	4	0.8%
Canada	3	0.6%
Denmark	3	0.6%
Korea	3	0.6%
Others	20	4.0%
Total	500	100.0%

30th TOP500 List, 11/2007

Country	Count	Share
USA	283	56.6%
Japan	20	4.0%
Germany	31	6.2%
France	17	3.4%
U.K.	48	9.6%
Australia	1	0.2%
Italy	6	1.2%
Netherlands	6	1.2%
Switzerland	7	1.4%
Canada	5	1.0%
Denmark	1	0.2%
Korea	1	0.2%
China	10	2.0%
India	9	1.8%
Others	55	11.0%
Total	500	100.0%

Figure 3: 1st and 30th TOP500 lists – Countries

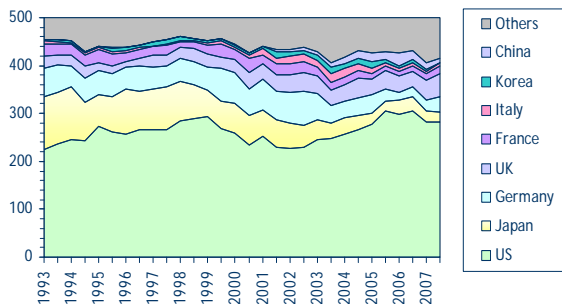


Figure 4: Supercomputer installations worldwide

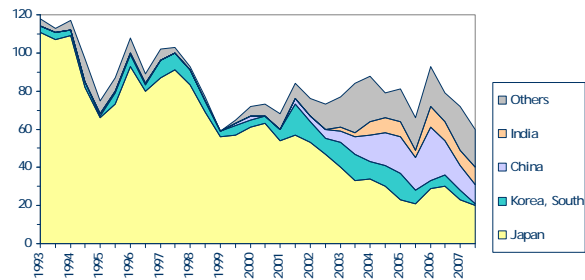


Figure 5: Supercomputer installations in Asia

2.2 Competition between Manufacturers

If we focus on the manufacturers (see **Figure 6**), Cray Research was the clear leader on our first TOP500 list with a 41% share, ahead of Fujitsu with 14%. Third place was already held by TMC – a non-vector supercomputer manufacturer – with 10.8%, ahead of Intel with 8.8%. At that time, Intel still had its Supercomputer Division, which also produced non-vector supercomputers. Surprisingly, today’s leading HPC manufacturers, IBM and Hewlett-Packard, were not represented on the first TOP500 list at all.

1st TOP500 List, 06/1993

Manufacturer	Count	Share
Cray Research	205	41.0%
Fujitsu	69	13.8%
Thinking Machines	54	10.8%
Intel	44	8.8%
Convex	36	7.2%
NEC	32	6.4%
Kendall Square Res.	21	4.2%
MasPar	18	3.6%
Meiko	9	1.8%
Hitachi	6	1.2%
Parsytec	3	0.6%
nCube	3	0.6%
Total	500	100.0%

30th TOP500 List, 11/2007

Manufacturer	Count	Share
Cray Inc.	14	2.8%
Fujitsu	3	0.6%
Thinking Machines	–	–
Intel	1	0.2%
Hewlett-Packard	166	33.2%
NEC	2	0.4%
Kendall Square Res.	–	–
MasPar	–	–
Meiko	–	–
Hitachi/Fujitsu	1	0.2%
Parsytec	–	–
nCube	–	–
IBM	232	46.4%
SGI	22	4.4%
Dell	24	4.8%
Others	35	7.0%
Total	500	100.0%

Figure 6: 1st and 30th TOP500 lists – Manufacturers

In the 30th TOP500 list of November 2007, IBM has the clear lead with a 46.4% share. The second position is held by Hewlett Packard with 33.2%, and the leader of 1993, Cray Research (now Cray Inc.), is now down to 2.8%.

If we look at the development of the manufacturers since 1993 (see **Figure 7**), we notice that the HPC market has been very dynamic: in only 15 years, the market has seen a complete transformation. Cray has turned from the clear market leader in the general HPC market, including the industrial customer segment, into a niche player for high-end government research laboratories and academic customers. IBM on the other hand, which was of virtually no importance in the HPC market in the early 90s, has become the dominant market leader in all market segments, including industrial and commercial customers. Hewlett-Packard – once a small HPC manufacturer represented in the first TOP500 lists only by Convex, which they later took over – has established itself as number two, right after IBM. Sun Microsystems, which used to be number two among the HPC manufacturers a couple of years ago, has fallen back dramatically in the TOP500. But Sun is now trying to catch up with the other HPC manufacturers. And also a re-invigorated Cray might be back in the general HPC arena again: They currently have three hybrid supercomputers in the TOP10, which shows that they have successfully left behind their approach of pure vector computing. With only very few systems in the overall TOP500 list, however, Cray will have to work hard to replicate its earlier success.

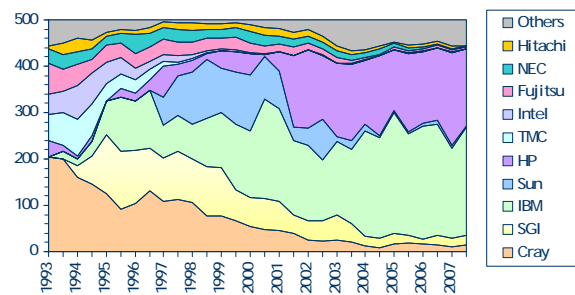


Figure 7: Manufacturers / Systems

2.3 Competition between Sites

Figure 8 lists the 20 most powerful sites through 30 TOP500 lists. The percentage in the right-hand column is a site’s relative contribution to the Rmax total of the average list of 30. In this list, the U.S. leads with two-thirds of the sites (14) ahead of Japan, with four centers (20%). The fact that the U.S. has the four most powerful sites in the world also shows its dominance as a consumer and producer of HPC systems. Europe is represented by Germany (Forschungszentrum Jülich, FZJ) at position 18 and by the U.K. (ECMWF) at position 15. (Note that ECMWF is a European and not purely a U.K. site.)

Rank	Site	Country	Over time
1	Lawrence Livermore National Laboratory	USA	5.39%
2	Sandia National Laboratories	USA	3.70%
3	Los Alamos National Laboratory	USA	3.41%
4	Government	USA	3.34%
5	The Earth Simulator Center	Japan	1.99%
6	National Aerospace Laboratory of Japan	Japan	1.70%
7	Oak Ridge National Laboratory	USA	1.39%
8	NCSA	USA	1.31%
9	NASA/Ames Research Center/NAS	USA	1.25%
10	University of Tokyo	Japan	1.21%
11	NERSC/LBNL	USA	1.19%
12	Pittsburgh Supercomputing Center	USA	1.15%
13	Semiconductor Company (C)	USA	1.11%
14	Naval Oceanographic Office (NAVOCEANO)	USA	1.08%
15	ECMWF	U.K.	1.02%
16	ERDC MSRC	USA	0.91%
17	IBM Thomas J. Watson Research Center	USA	0.86%
18	Forschungszentrum Jülich (FZJ)	Germany	0.84%
19	Japan Atomic Energy Research Institute	Japan	0.83%
20	Minnesota Supercomputer Center	USA	0.74%

Figure 8: TOP20 through 30 TOP500 lists

3. My Favorite Supercomputer in all TOP500 Lists so far

We have published 30 TOP500 lists with a total of 15,000 systems, and our data base has even twice as many entries. So it might sound strange that I have just one favorite system. I would like to emphasize, however, that there are many systems that impressed me over the past 15 years. Before I reveal my favorite supercomputer, I would like to highlight another one here first: It was number 259 on our 9th TOP500 list published at the Mannheim Supercomputer Seminar (ISC'97) in 1997. This system, named "Deep Blue", was installed at the IBM Watson Research Center in Yorktown Heights and had a best Linpack performance of 11.37 Gigaflop/s; it was an IBM SP2 P2SC with 32 processors and a clock rate of 120 MHz. But the floating point performance was not what really mattered: Each of the 32 processors was equipped with 15 special-purpose VLSI chess chips. Deep Blue was the first chess computer to beat a reigning world chess champion, Garry Kasparov.^[4] Ten years after this event, no chess player stands a chance against any kind of computer, not even against a simple home computer. One year ago, in November/December 2006, Deep Fritz played a six-game match against reigning world chess champion Wladimir Kramnik in Bonn. Deep Fritz won 4–2.^[5]

3.1 My All-time Favorite: Intel's ASCI Red

During the 1996 acceptance tests at Sandia National Laboratories in Albuquerque/USA, Intel's ASCI Red showed an Rmax performance of 1,068 Gigaflop/s. Thus it was the first Teraflop/s computer to enter the HPC arena, and it immediately grabbed the first place on our 9th TOP500 list of June 1997 (see **Figure 9**).

"ASCI" stands for "Accelerated Strategic Computing Initiative". This initiative, under the auspices of the Department of Energy (DOE), was the U.S. response to France's nuclear weapons tests at the Mururoa atoll, where France conducted 41 atmospheric and 147 underground nuclear tests between 1966 and 1996. DOE's "Accelerated Strategic Computing Initiative" applies advanced capabilities in scientific and engineering computing to one of the most complex challenges in the nuclear era: maintaining the performance, safety, and reliability of the U.S. nuclear weapons arsenal without physical testing. ASCI was established in 1996 in response to the government's commitment to pursue a comprehensive ban on nuclear weapons testing.

Rank	Manufacturer	Computer	Rmax [GF/s]	Site	Country	Year	#Proc
1	Intel	ASCI Red	1,068	Sandia National Laboratories	USA	1996	7264
2	Hitachi	CP-PACS/2048	368.2	Center for Computational Science	Japan	1996	2048
3	Fujitsu	Numerical Wind Tunnel	229	National Aerospace Laboratory	Japan	1996	167
4	Hitachi	SR2201/1024	220.4	University of Tokyo	Japan	1996	1024
5	Cray	T3E	176	Forschungszentrum Jülich	Germany	1996	512
6	Cray	T3E	176	Government	USA	1996	512
7	Cray	T3E	176	Max-Planck-Gesellschaft MPI/IPP	Germany	1996	512
8	Cray	T3E	176	NASA/Goddard Space Flight Center	USA	1996	512
9	Cray	T3E	176	Pittsburgh Supercomputer Center	USA	1996	512
10	Cray	T3E	176	University of Stuttgart	Germany	1996	512

Figure 9: Top 10 sites of the 9th TOP500 list, June 1997

ASCI Red was the last supercomputer designed and assembled solely by Intel; Intel's Supercomputer Division had already been closed down when ASCI Red was launched.^[6]

ASCI Red was the last supercomputer designed and assembled solely by Intel; Intel's Supercomputer Division had already been closed down when ASCI Red was launched.^[6]

3.2 Eight-year Forecast

At the opening session of the 12th Supercomputer Conference in Mannheim in June 1997, we released our 9th TOP500 list with the new number one system: ASCI Red (see **Figure 10**). Considering the measured data from the nine TOP500 lists from June 1993-June 1997 and assuming that the increases in performance would continue as before, we extrapolated the performance of future systems using linear regression on the logarithmic scale. We fit exponential growth to these three levels of performance, the sum of all 500 systems, the number one system and the entry level system.

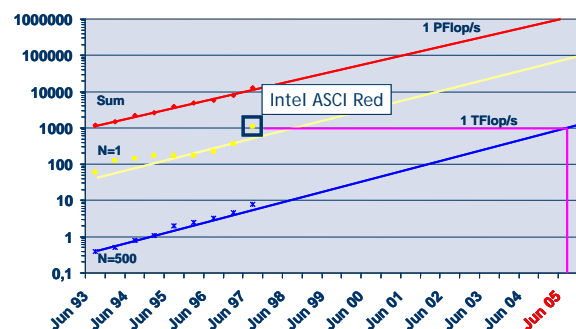


Figure 10: Performance [Gigaflop/s]

Based on the extrapolation of these fits, I announced at ISC'97 that, eight years from then, i.e. in 2005, there would be only Teraflop/s systems on the TOP500 list, even though the 1997 list had only one: Intel's ASCI Red.

Probably many of the ISC'97 attendees thought I was crazy to make a forecast eight years out for such a dynamic market. And I do have to admit that I did not feel too comfortable myself – going out on a limb like that in front of many of the world's top HPC experts. You certainly can imagine that I was extremely anxious to see the next TOP500 lists.

Number of Teraflop/s Systems in the TOP500 Lists 1997–2005

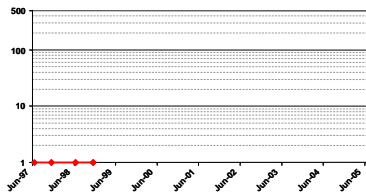


Figure 11a: 1997–1998

The TOP500 lists of November 1997, June 1998 and November 1998 had still only one Teraflop/s system: ASCI Red. I started to get really nervous...

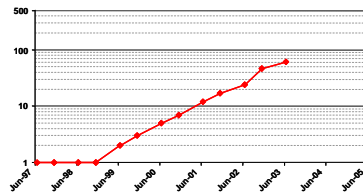


Figure 11b: 1997–2003

... but fortunately, the TOP500 list of June 1999 showed a new Teraflop/s system: SGI's ASCI Blue Mountain at Los Alamos National Laboratory. It was the second system worldwide to break the Linpack Teraflop/s barrier, and with 1.6 Teraflop/s, it took second place on the June 1999 TOP500 list. ASCI Red was slightly over 2 Teraflop/s then, as Intel had managed to increase the number of processors to more than 9,000.

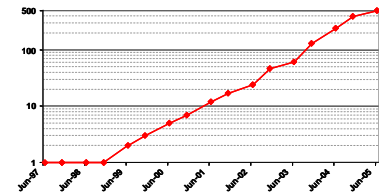


Figure 11c: 1997–2005

And in June 2005 – as I had predicted eight years earlier – there were only Teraflop/s systems in the TOP500 list. The entire Teraflop/s market had gained momentum since 1999. Whereas the June 2001 TOP500 list showed only a few more than 10 Teraflop/s systems, the 2003 list already had more than 100 such systems.

3.3 Why Intel's ASCI Red is my Favorite System in the TOP500

Here are the reasons why ASCI Red is my favorite system:

1. We saw the CDC 7600 (a predecessor model of the later vector computers) break the Megaflop/s barrier with 1.24 Megaflop/s in 1971, and we also saw the legendary Cray 2 exceed the Gigaflop/s barrier in 1986, boasting 2 GB of memory and a best Linpack performance Rmax of 1.7 Gigaflop/s. But what impressed me the most was when Intel's ASCI Red broke the Teraflop/s barrier in 1997. Of course, I hope to see the breaking of the Petaflop/s barrier this year (2008). I will thus have personally witnessed an increase in supercomputer performance of 10 orders of magnitude over the years.
2. Intel's ASCI Red marked the beginning of a new supercomputer era. In the mid-90s when vector computers started to become less important, DOE's ASCI initiative, which focused on defense applications, opened up a completely new source of funds. ASCI Red was the first product of this initiative and laid the foundation for the U.S. dominance in the production and implementation of supercomputers. ASCI Red was also a remarkable supercomputer from a technical point of view:

It was a mesh-based (38 X 32 X 2) MIMD massively parallel machine initially consisting of 7,264 compute nodes, 1,212 gigabytes of total distributed memory and 12.5 terabytes of disk storage. The original incarnation of this machine used Intel Pentium Pro processors, each clocked at 200 MHz. These were later upgraded to Pentium II OverDrive processors. The system was upgraded to a total of 9,632 Pentium II OverDrive processors, each clocked at 333 MHz. It consisted of 104 cabinets, taking up about 2,500 square feet (230 m²). The system was designed to use commodity mass-market components and to be very scalable.

3. In June 1997, I predicted that, eight years later, there would be only Teraflop/s systems on the TOP500 list, even though just a single Teraflop/s computer existed at that time, and this was ASCI Red. Never before – and probably never again – has one of my forecasts proven so accurate. ASCI Red was retired from service in September 2005, after having been on 17 TOP500 lists over eight years. It was the fastest computer on the TOP500 list from June 1997 to June 2000 and was replaced as the number one by IBM's ASCI White at Lawrence Livermore National Laboratory on the November 2000 list.
4. And finally, ASCI Red was the reason I got invited to give an interview on German TV (ZDF, Zweites Deutsches Fernsehen). Shortly before Christmas 1996, Nina Ruge, a German TV journalist, interviewed me on her late-night TV show "Heute Nacht" about the first computer to exceed the Teraflop/s barrier: Intel's ASCI Red.

4. Highlights of the 30th TOP500 List and Bell's Law

4.1 Highlights of the 30th List^[7]

Rank	Manufacturer	Computer	Rmax [TF/s]	Site	Country	Year	#Cores
1	IBM	BlueGene/L eServer Blue Gene	478.2	DOE/NNSA/LLNL	USA	2007	212,992
2	IBM	JUGENE BlueGene/P Solution	167.3	Forschungszentrum Jülich	Germany	2207	62,536
3	SGI	SGI Altix ICE 8200	126.9	New Mexico Computing Center	USA	2007	14,336
4	HP	Cluster Platform 3000 BL460c	117.9	Computational Research Laboratories, TATA SONS	India	2007	14,240
5	HP	Cluster Platform 3000 BL460c	102.8	Swedish Government Agency	Sweden	2007	13,728
6	Sandia/Cray	Red Storm Cray XT3	102.2	DOE/NNSA/Sandia	USA	2006	26,569
7	Cray	Jaguar Cray XT3	101.7	DOE/ORNL	USA	2007	23,016
8	IBM	BGW eServer Blue Gene	91.3	IBM Thomas Watson	USA	2005	40,960
9	Cray	Franklin Cray XT4	85.4	NERSC/LBNL	USA	2007	19,320
10	IBM	New York Blue eServer Blue Gene	82.2	Stony Brook/BNL	USA	2007	36,864

Figure 12: Top 10 sites of the 30th TOP500 list, November 2007

Among the 10 top sites on the 30th TOP500 list are five new systems and one substantially upgraded system (marked yellow in **Figure 12**). The main changes have taken place among the first five places. Place number one again goes to BlueGene/L, a joint development of IBM and the Department of Energy's (DOE) National Nuclear Security Administration (NNSA), which is installed at DOE's Lawrence Livermore National Laboratory in Livermore, CA/USA. BlueGene/L had been in first place since November 2004; however, the current system has been significantly upgraded so that it now achieves a Linpack benchmark performance of 478.2 Teraflop/s (trillions of calculations per second) compared to a performance of 280.6 Teraflop/s six months ago, before its upgrade.

Place number two is held by a brand-new first installation of a newer version of the same type of IBM system. It is a BlueGene/P system installed at the Forschungszentrum Jülich (FZJ), Germany, with a performance of 167.3 Teraflop/s.

The number three system is not only a new one, but also the first system of a new supercomputing center: the New Mexico Computing Applications Center (NMCAC) in Rio Rancho/USA. The system, built by SGI and based on the Altix ICE 8200 model, reaches a speed of 126.9 Teraflop/s.

For the first time ever, India has been able to place a system in the Top10, at number four. The Computational Research Laboratories, a wholly owned subsidiary of Tata Sons Ltd. in Pune/India, installed a Hewlett-Packard Cluster Platform 3000 BL460c system. They integrated this system with their own innovative routing technology and achieved a performance level of 117.9 Teraflop/s.

The number five system is also a new Hewlett-Packard Cluster Platform 3000 BL460c system, installed at a Swedish government agency. It was measured at 102.8 Teraflop/s.

The last new system in the Top10 is a Cray XT4 system installed at the National Energy Research Scientific Computing Center (NERSC) at DOE's Lawrence Berkeley National Laboratory in Berkeley, CA/USA. With a Linpack performance of 85.4 Teraflop/s, it ranks ninth.

Processor Architecture / Systems

Figure 13 illustrates that vector computers are on the retreat in the TOP500: On the 30th TOP500 list of November 2007, there are only four vector computer systems, two from Cray Inc. and two from NEC, among them the worldwide number one of 2002–2004, the Earth Simulator in Yokohama/Japan, which has fallen to number 30.

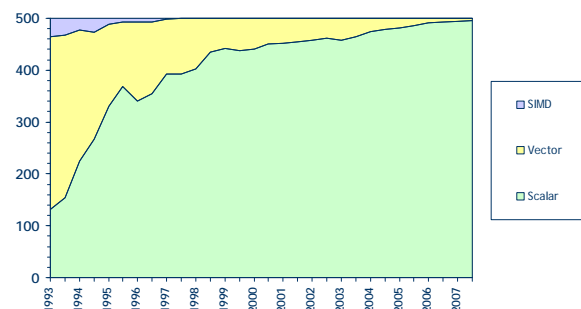


Figure 13: Processor architecture / Systems

Operating Systems / Systems

Up to a couple of years ago, UNIX in all its variations was the prevalent operating system on supercomputers, but now Linux has taken over this role. Despite Microsoft's effort to break into this market, Windows plays no role. See **Figure 14**.

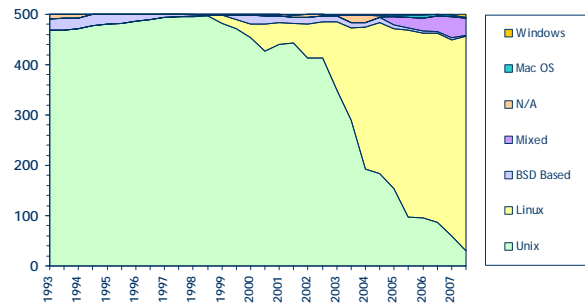


Figure 14: Operating systems / Systems

Processor Generations / Systems

354 out of 500 systems (70.8%) use Intel processors, whereas six months ago, only 289 systems (57.8%) had Intel processors. This is the largest share for Intel chips in the TOP500 ever. Especially successful are the Dual-Core Woodcrest and the Quadcore Clovertown processors with a share of 43% and 20.4% respectively. The AMD Opteron family, which left the IBM Power processors behind a year ago, still remains the second-most-common processor family, even though the number of systems using this processor went down from 105 (21%) to 78 (15.6%). 61 systems (12.2%) run on IBM Power processors, compared to 85 systems (17%) half a year ago. See **Figure 15**.

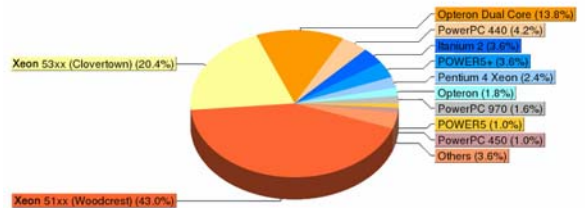


Figure 15: Processor generations / Systems

Interconnect Family / Systems

Due to its widespread use by industrial customers, Gigabit Ethernet is still the most widely utilized internal system interconnect technology (270 systems). It is followed by InfiniBand technology with 121 systems. Myrinet, which dominated the market a couple of years ago, has fallen even further back. See **Figure 16**.

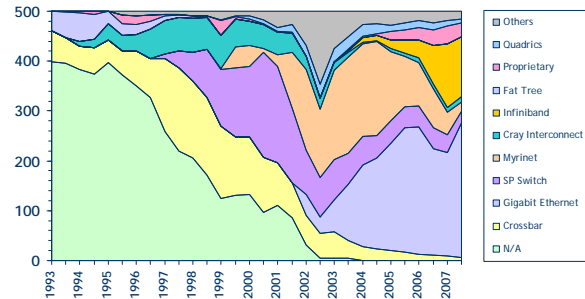


Figure 16: Interconnect family / Systems

Architectures / Systems

The most widely used architecture is the cluster architecture, as 406 out of 500 systems (81.2%) are labeled as clusters. Times are long gone when clusters only appeared in the second half of the TOP500 list. There are even two cluster systems among the TOP10, including the most powerful system in Asia at number four, the Cluster Platform system of HP. Constellations have dropped to 0.6%. MPPs hold an 18.2% share, with eight systems in the TOP10. See **Figure 17**.

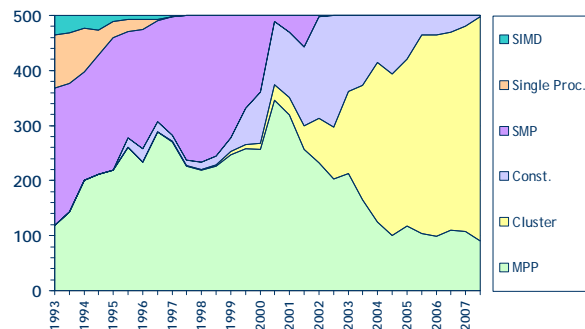


Figure 17: Architectures / Systems

4.2 Bell's Law (1972)

Bell's Law of Computer Class formation was discovered about 1972.^[8] It states that technology advances in semiconductors, storage, user interfaces and networking take place approximately every decade enabling a new, usually lower-priced computing platform to form. Once formed, each class is maintained as a quite independent industry structure. This explains mainframes, minicomputers, workstations and personal computers, the web, emerging web services, palm and mobile devices and ubiquitous interconnected networks. We can expect home and body area networks to follow this path. Bell's Law states that important classes of computer architectures come in cycles of about ten years. It takes about a decade for each of the phases:

- Early research
- Early adoption and maturation
- Prime usage
- Phase out past its prime

When I was preparing my presentation for the IDPT conference in Antalya in June 2007,^[9] Erich Strohmaier and I asked ourselves: Can we use Bell's Law to classify computer architectures in the TOP500?

We make an attempt by introducing the following architectures/computer classes:

- **Data parallel systems**
Vector (Cray Y-MP and X1, NEC SX etc.)
SIMD (CM-2 etc.)
- **Custom scalar systems**
MPP (Cray T3E and XT3, IBM SP etc.)
Scalar SMPs and Constellations (Cluster of big SMPs)
- **Commodity clusters**
NOW, PC cluster, Blades etc.
- **Power-efficient systems**
BG/L or BG/P as first examples of low-power systems. They might be able to form a new class, but we do not know this yet.

When analyzing all TOP500 lists from the very beginning in 1993 up to now, we find the following computer classes over time, as shown in **Figures 18 and 19**:

Computer classes in HPC based on the TOP500

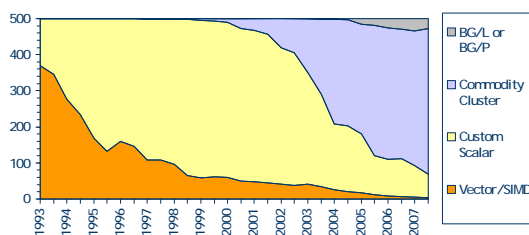


Figure 18: Computer classes / Systems

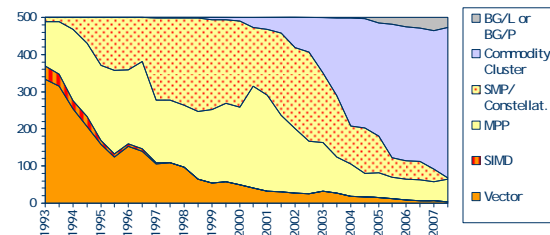


Figure 19: Computer classes – refined / Systems

The HPC computer classes “Data parallel systems”, “Custom scalar systems” and also “Commodity clusters” follow nicely the 10-year cycle of Bell's Law and confirm this law for HPC computer architectures, see **Figure 20**. For simplicity, we have left out the “Early research” phase of Bell's Law. Of course, we have to wait and see whether or not the 10-year “Past prime usage” phase of the “Commodity clusters” class will really start around 2010. And there is even more speculation about which way the “Power-efficient systems” class, represented by IBM's BG series, will go. A couple of smaller such systems were delisted from the 30th TOP500 list due to poor performance. The next TOP500 lists will show whether or not the “Power-efficient systems” class will really be a class of its own.

Class	Early adoption starts	Prime use starts	Past prime usage starts
Data parallel systems	Mid 70's	Mid 80's	Mid 90's
Custom scalar systems	Mid 80's	Mid 90's	Mid 2000's
Commodity clusters	Mid 90's	Mid 2000's	Mid 2010's ???
BG/L or BG/P	Mid 2000's	Mid 2010's ???	Mid 2020's ???

Figure 20: HPC computer classes and Bell's Law

5. Performance Development and Performance Projections

In **Figure 21**, we have plotted the performance over the last 15 years at position N=500 (entry level performance), at the leading position (number one), as well as the total accumulated performance, the sum of all 500 systems. As can easily be seen, all these curves show an exponential growth. The scale on the left-hand side is a logarithmic scale.

If we compare this growth with Moore's Law, we find that, even though Moore's Law assumes a doubling in performance every 18 months for microprocessors, our growth is larger. We have a doubling for the sum in approximately 14 months, for the number one position in approximately 13 months and even for the number 500 position in a little less than 13 months. There are two main reasons for this larger growth in performance: processor performance and number of processors used.

Also note that the curves at positions one and 500 are quite different: At number one, we typically see a step function. Once a system has made number one, it remains there in the next couple of TOP500 lists. That was true for the "Numerical Wind Tunnel – NWT", Intel's ASCI Red and also for the "Earth Simulator", which ranked first from June 2002 through June 2004. And it also proves true for the current number one supercomputer (since November 2004), IBM's BlueGene/L at Lawrence Livermore National Laboratory (LLNL), holding this position at different stages of expansion.

If we include a powerful notebook in this figure, we notice that its performance has reached 7 Gigaflop/s now and has thus grown by a factor of 10 within three years.

Again, as when discussing Intel's ASCI Red, we have done a projection into the future, based on 30 lists of real data, by a least square fit on the logarithmic scale. For a powerful notebook, for example, this means that it will have a Teraflop/s performance in the year 2014, i.e. in less than 18 years after the first Teraflop/s system, ASCI Red, entered the HPC arena.

Generally, it will take six to eight years for any system to move from position one to 500 and eight to ten years to move from position 500 to notebook level.

The Linpack Petaflop/s threshold will be reached in 2008. One of the hot candidates for the first Petaflop/s system to enter the TOP500 list is IBM's RoadRunner at Los Alamos National Laboratory, USA. In 2015, there will be only Petaflop/s systems in the TOP500 list. Our projection also shows that the first Exaflop/s computer will enter the TOP500 list in 2019, and only one year later, in 2020, there will be the first notebooks with a performance of 100 Teraflop/s.

The rule seems to be that system performance increases by a factor 1,000 every eleven years: Cray 2 broke the Gigaflop/s barrier in 1986 (in pre-TOP500 times); Intel's ASCI Red exceeded the Teraflop/s barrier in 1997; the first Petaflop/s system will enter the TOP500 list this year (2008); and, according to our projection, the Exaflop/s threshold will be reached in 2019. See **Figure 22**.

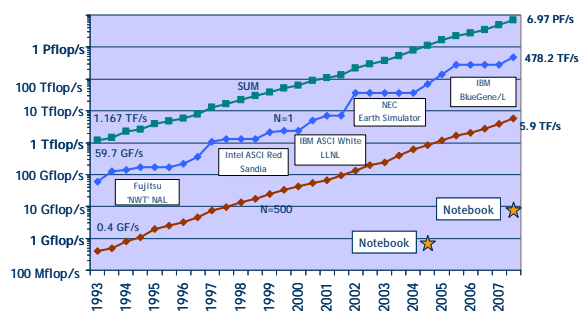


Figure 21: Performance development

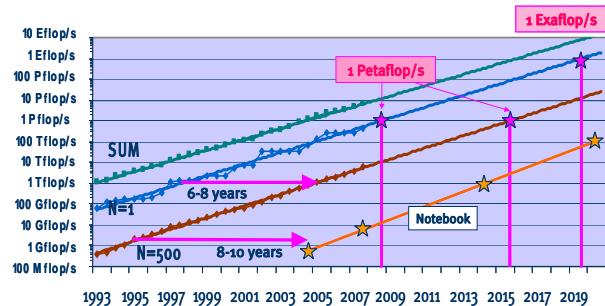


Figure 22: Performance projection

6. Top500 in the Future

6.1 The TOP500 Website

The results of ALL 30 TOP500 lists and a variety of additional information are available on our TOP500 website: www.top500.org, see **Figure 23**. This site draws remarkable traffic – more than 20K page impressions per day and is kept up to date by Anas Nashif, our Technical Manager, who is also responsible for the TOP500 data base. The website has been improved and relaunched recently and offers many interesting features: access to sublists, list charts, list statistics and up-to-date information on the HPC market in general. The TOP500 project is financed – but not influenced – by advertising on the website.



Figure 23: TOP500 website

6.2 Summary after 15 Years of Experience

The TOP500 corrected the deficits of the Mannheim supercomputer statistics, which we had used for seven years at the ISC conferences of 1986-92, and has proven to be a reliable tool ever since. Its simple but successful approach based on the Linpack benchmark, though often criticized, is able to get trends right, as far as processors, architectures, manufacturers, countries and sites are concerned. And as shown in Chapter 3, “My Favorite Supercomputer in all TOP500 Lists so far”, its performance predictions turn out remarkably correct, even over such a long period as eight years.

The TOP500 lists should only be seen as a source of information for general trends. As such, they are extremely useful and much more reliable than the predictions of market research companies such as IDC, Diebold, etc. However, we have always advised people not to use the TOP500 lists to answer specific questions such as: Is system X at position 254 better suited for a certain application than system Y at position 344? For these cases, you would have to run your own benchmarks and applications sets on the systems in question.

With the TOP500 lists, it is also not possible to estimate the size of the HPC market (e.g., in US \$), as we just do not know how much the systems at the different positions cost. We have often been asked for the systems’ prices, and we have to admit that this information would be of great interest to the HPC community. But at the very beginning of the TOP500 project we decided not to include this kind of more or less unreliable and vague data in our lists.

When analyzing the TOP500 lists, we find that systems of the upper half of the list used to remain there for a couple of periods and smooth out seasonal fluctuations. And there are the short time entries, which are often on a list for only six months, since the turnover is very high. **Figure 24** illustrates that, on average, approximately 200 systems drop out after six months, not making it into the following list due to poor performance.

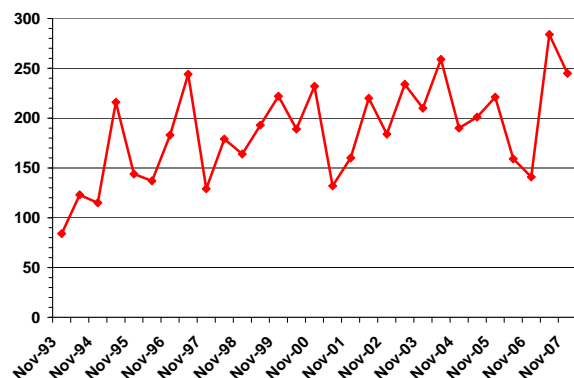


Figure 24: TOP500 replacement rate

6.3 Motivation for Additional Benchmarks

Figure 25 summarizes the pros and cons of the Linpack benchmark:

Pros	Cons
<ul style="list-style-type: none"> • Only one figure of merit • Simple to define and rank • Allows problem size to change with machine and over time • Allows competition 	<ul style="list-style-type: none"> • Emphasizes only “peak” CPU speed and number of CPUs • Does not stress local bandwidth • Does not stress the network • No single figure of merit can reflect the overall performance of an HPC system

Figure 25: Linpack benchmark – Pros and cons

6.4 HPC Challenge Benchmark

We clearly need something more than Linpack for the TOP500, e.g., HPC Challenge Benchmark and others. At ISC’06 in Dresden/Germany, Jack Dongarra gave the Friday keynote presentation on “HPC Challenge Benchmarks and the TOP500”.^[10] The conference attendees voted his excellent talk one of the two best ISC’06 presentations.

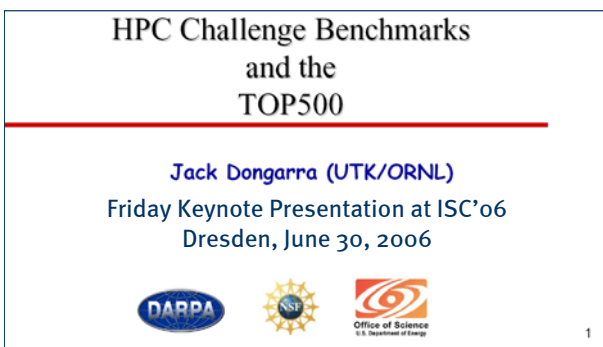


Figure 26: ISC’06 Keynote Presentation

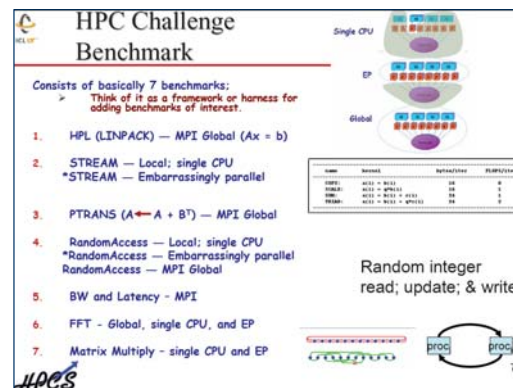


Figure 27: HPC Challenge Benchmark

The HPC Challenge Benchmark basically consists of seven different benchmarks, each stressing a different part of a computer system. Of course HPL, the High Performance Linpack benchmark, is also part of these seven benchmarks and stands for the CPU. We do not have the advantages of a single figure of merit any longer, and the results of the HPC Challenge Benchmark are much more complex so that so-called Kiviart charts are needed. With these charts, it will be much harder for journalists, for example, to report on new systems entering the HPC arena than when they are evaluated only with Linpack.

Dongarra’s conclusion is that we will certainly still see Linpack as the TOP500 benchmark for a while. However, it needs to be expanded to produce lists using another yardstick. The HPC Challenge Benchmark could become a standard in the U.S. when it comes to selecting an HPC system.

6.5 The Green500 List

The Green500 list, overseen by Wuchun Feng and Kirk W. Cameron of Virginia Tech/USA, is another approach to ranking supercomputers.^[11] Its purpose is to list the most energy-efficient supercomputers in the world and serve as a complementary view to the TOP500. However, the latest Green500 list is far from being complete, as it does not even include all TOP500 systems. The Green500's 10 most energy-efficient supercomputers are all IBM systems, see **Figure 28**. This is probably one of the reasons why IBM strongly supports the project.

The TOP500 authors basically support the idea of a Green500 list, but they reserve the right to launch their own independent and more thorough project, if necessary.

Green500 Rank	MFlop/s/W	Site	Computer	Total Power (kW)	TOP500 Rank
1	357.23	Science and Technology Facilities Council – Daresbury Laboratory	Blue Gene/P Solution	31.10	121
2	352.25	Max-Planck-Gesellschaft MPI/IPP	Blue Gene/P Solution	62.20	40
3	346.95	IBM – Rochester	Blue Gene/P Solution	124.40	24
4	336.21	Forschungszentrum Jülich (FZJ)	Blue Gene/P Solution	497.60	2
5	310.93	Oak Ridge National Laboratory	Blue Gene/P Solution	70.47	41
6	210.56	Harvard University	eServer Blue Gene Solution	44.80	170
7	210.56	High Energy Accelerator Research Organization / KEK	eServer Blue Gene Solution	44.80	171
8	210.56	IBM – Almaden Research Center	eServer Blue Gene Solution	44.80	172
9	210.56	IBM Research	eServer Blue Gene Solution	44.80	173
10	210.56	IBM Thomas J. Watson Research Center	eServer Blue Gene Solution	44.80	174

Figure 28: Green500's Top10

7. Conclusion

The TOP500 project was launched in 1993 to improve and renew the Mannheim supercomputer statistics, which had been in use for seven years. Our simple TOP500 approach does not define “supercomputer” as such, but we use a benchmark to rank systems and to decide on whether or not they qualify for the TOP500 list. The benchmark we decided on was Linpack, which means that systems are ranked only by their ability to solve a set of linear equations, $Ax = b$, using a dense random matrix A . Therefore, any supercomputer – no matter what its architecture is – can make it into the TOP500 list, as long as it is able to solve a set of linear equations using floating point arithmetic. We have been criticized for this choice from the very beginning, but now, after 15 years, we can say that it was exactly this choice that has made TOP500 so successful – Linpack therefore was a good choice. And there was, and still is, no alternative to Linpack. Any other benchmark would have been similarly specific, but would not have been so easily available for all systems – a very important factor, as compiling the TOP500 lists twice a year is a very complex process.

One of Linpack's advantages is also its scalability in the sense that it has allowed us in the past 15 years to benchmark systems that cover a performance range of 10 orders of magnitude. It is true that Linpack delivers performance figures that occupy the upper end of any other application performance. In fact, no other realistic application delivers a better efficiency (R_{max}/R_{peak}) of a system. But using the peak performance instead of Linpack, which “experts” have often recommended to us, does not make any sense. We have seen a lot of new systems that were not able to run the Linpack test because they were not stable enough. For example, it takes more than six hours to run the Linpack test on the current number two system on the TOP500 list, BlueGene/P at Forschungszentrum Jülich (FZJ). Therefore, running Linpack to measure the performance is kind of a first reliability test for new HPC systems.

The misinterpretation of the TOP500 results has surely led to a negative attitude towards Linpack. Politicians, for example, tend to see a system's TOP500 rank as a general rank that is valid for all applications, which of course is not true. The TOP500 rank only reflects a system's ability to solve a linear set of equations, and it does not tell anything about its performance with respect to other applications. Therefore, the TOP500 list is not a tool for selecting a supercomputer system for an organization. In this case, you would have to run your own benchmark tests that are relevant to your own applications. In this context, an approach such as the “HPC Challenge Benchmark” consisting of seven different benchmarks, which test different parts of a supercomputer, is critical. As experts run the “HPC Challenge Benchmark” tests and interpret their results, it is not a problem not to have only one single figure of merit. For this reason, the HPC Challenge Benchmark has already become a certain criteria in the U.S. when it comes to buying HPC systems.

The TOP500 lists' success lies in compiling and analyzing data over time. Despite relying solely on Linpack, we have been able to correctly identify and track ALL developments and trends over the past 15 years, covering

manufactures and users of HPC systems, architectures, interconnects, processors, operating systems, etc. And above all, TOP500's strength is that it has proved an exceptionally reliable tool for forecasting developments in performance.

It is very unlikely that another benchmark will replace Linpack as basis for the TOP500 lists in the near future. And in any case we would stick to the concept of a single benchmark because this is the easiest way to trigger competition between manufacturers, countries and sites, which is extremely important for the overall acceptance of the TOP500 lists. Of course, we appreciate it if alternative benchmarks are introduced to complement Linpack. In fact, we are working on this already and encourage other HPC experts to come up with constructive suggestions, too.

Acknowledgement

I would like to thank my colleague, Erich Strohmaier, for his support, his advice and his inspiration, as well as for preparing the data for the figures in this paper and the TOP500 project in general. My thanks go also to my two colleagues, TOP500 authors Jack Dongarra and Horst Simon, for the many interesting discussions on TOP500 topics over the past 15 years, which have been very helpful for this paper, too. I would also like to thank Reiner Hartenstein, University of Kaiserslautern, for inviting me to give a talk on this subject at the World Conference on Integrated Design & Process Technology (IDPT) in Antalya/Turkey on June 7, 2007. He is the person who suggested using "Bell's Law" to classify computer architectures in the TOP500 lists. My thanks go to Hans Günther Kruse, University of Mannheim, for checking the manuscript and for his suggestions for improvement. And last but not least, I would like to thank Heike Walther, Prometheus GmbH, for translating most of this paper and Jon Bashor, LBNL, for giving it the final polish.

References

- [1] H. W. Meuer, The Mannheim Supercomputer Statistics 1986–1992, TOP500 Report 1993, University of Mannheim, 1994, p. 1–15
- [2] See: <http://www.top500.org/project/linpack>
- [3] E. Strohmaier, J. J. Dongarra, H. W. Meuer, H. D. Simon, The Marketplace of High-Performance Computing, Parallel Computing 25, 1999, p. 1517–1544
- [4] See: http://en.wikipedia.org/wiki/IBM_Deep_Blue
- [5] See: <http://www.spiegel.de/netzwelt/tech/o,1518,452735,00.html>
- [6] See: <http://www.sandia.gov/ASCI/Red/>
- [7] See: <http://www.top500.org/>
- [8] See: <http://research.microsoft.com/~GBell/Pubs.htm>
- [9] World Conference on Integrated Design & Process Technology (IDPT), Antalya/Turkey, June 3–8, 2007 <http://www.fpl.uni-kl.de/IDPT/IDPT2007-final.pdf>
- [10] See: <http://www.netlib.org/utk/people/JackDongarra/SLIDES/isc-talk-2006.pdf>
- [11] See: <http://www.green500.org/>