UNIVERSITY OF
FORWARD
THINKING
WESTMINSTER⌗

WestminsterResearch

http://www.westminster.ac.uk/westminsterresearch

The 30th Anniversary of the Supercomputing Conference: Bringing the Future Closer - Supercomputing History and the Immortality of Now

Dongarra, J., Getov, Vladimir and Walsh, K.

The 30th Anniversary of the Supercomputing Conference:

Bringing the Future Closer— Supercomputing History and the Immortality of Now

Jack Dongarra, University of Tennessee, Oak Ridge National Laboratory, and University of Manchester

Vladimir Getov, University of Westminster

Kevin Walsh, University of California, San Diego

A panel of experts discusses historical reflections on the past 30 years of the Supercomputing (SC) conference, its leading role for the professional community and some exciting future challenges.

S upercomputing's nascent era was borne of the late 1940s and 1950s Cold War and increasing tensions between the East and the West; the first installations—which demanded extensive resources and manpower beyond what private corporations could provide—were housed in university and government labs in the United States, United Kingdom, and the Soviet Union. Following the Institute of Advanced Study (IAS) stored program computer architecture, these so-called von Neumann machines were implemented as the MANIAC at Los Alamos Scientific Laboratory, the Atlas at the University of Manchester, the ILLIAC at the University of Illinois, the BESM machines at the Soviet Academy of Sciences, the Johnniac at The Rand Corporation, and the SILLIAC in Australia. By 1955, private industry

joined in to support these initiatives and the IBM User Group, SHARE was formed, the Digital Equipment Corporation was founded in 1957, while IBM built an early wide area computer network SAGE (Semi-Automatic Ground Environment), and the RCA 501, using all transistor logic was launched in 1958.

In 1988, the first IEEE/ACM SC conference in Kissimmee, Florida, was held. At that time, custom-built vector mainframes were the norm; the Cray Y-MP was a leading machine of the day, with a peak performance of 333 Mflops per processor and could be equipped with up the eight processors; users typically accessed the machine over a dumb terminal at 9600 baud; there was no visualization; a single programmer would code and develop everything; and there were few tools or software libraries, and we relied on remote batch job submission.

Today, as we approach SC's 30th anniversary, late commodity massively parallel platforms are the norm. The HPC community has developed parallel debuggers and rich tool sets for code share and reuse. Remote access to several supercomputers at once is made possible by scientific gateways, accessed over 10 and 100 Gbps networks. High performance desktops with scientific visualization capabilities are the chief methods we use to cognitively grasp the quantity of data produced by supercomputers. The Cold War arms race has been eclipsed by an HPC race, and we are living through a radical refactoring of the time it takes to create new knowledge—and, concurrently, the time it takes to learn how much we don't know. The IEEE/ACM SC conference animates the community and allows us to see what knowledge changes, and what knowledge stays stable over time.

To review and summarize the key developments and achievements of

MACHINES OF THE FUTURE

"The advanced arithmetical machines of the future will be electrical in nature, and they will perform at 100 times present speeds, or more. Moreover, they will be far more versatile than present commercial machines, so that they may readily be adapted for a wide variety of operations. They will be controlled by a control card or film, they will select their own data and manipulate it in accordance with the instructions thus inserted, they will perform complex arithmetical computations at exceedingly high speeds, and they will record results in such form as to be readily available for distribution or for later further manipulation."
—"As We May Think," by Vannevar Bush (*The Atlantic*, July 1945)

HPC over the past 30 years, we have invited 6 well-known experts—Gordon Bell, Jack Dongarra, Bill Johnston, Horst Simon, Erich Strohmaier, and Mateo Valero—all of whom offer complimentary perspectives on the past three to four decades of supercomputing. Their histories connect us in community.

COMPUTER: Looking back at the early years of electronic digital computers, what do you see as the turning points and defining eras of supercomputing that help chronicle the growth of the HPC community and the SC conference at its 30th anniversary?

GORDON BELL: In 1961, after I visited Lawrence Livermore National Laboratory with the Univac LARC, and Manchester University with the Atlas prototype, I began to see what supercomputing was about from a design and user perspective—namely, it was designing at the edge of the feasibility envelope using every known technique.

The IBM Stretch was one of these three 1960s computers aimed to achieve over an order of magnitude performance increase over the largest, commercial state-of-the-art computers. Doing everything known to be feasible for performance (for example, parallel units, lookahead, speculative execution). In pre-SC88, there were trials and failures, such as violations of Amdahl's Law[1] by Single Instruction Multiple Data (SIMD) and other architectures or pushing technology that failed to reach a critical production (such as GaAs). At the beginning of the first generation of commercial computing, Seymour Cray joined Control Data Corporation as a founder and quickly demonstrated a proclivity for building the highest performance computer of the day; in essence, he defined and established the tri-decade of supercomputing and the market.

After the initial CDC 1604 (1960) introduction, Seymour proceeded to build computers without peer, including the CDC 6600 in 1965, and CDC 7600 in 1969. He then formed Cray Research to introduce the vector processor Cray 1 (1976), which was followed by the multiprocessor Cray SMP (1985), YMP (1988), and last C90 (1991). From 1965 through 1991, the Cray architectures defined and dominated computer design and the market, which included CDC, Cray Research, Fujitsu, Hitachi, IBM, and NEC.

MATEO VALERO: I see the rise, fall, and resurgence of vector processors as the turning points of supercomputing. Vector processors execute instructions whose operands are complete vectors. This simple idea was so revolutionary, and the implementations were so efficient that the resulting vector supercomputers reigned supreme as the fastest computers in the world from 1975 to 1995. Vector processors exploit data-level parallelism elegantly, they could hide memory latency very well, and they are energy efficient since they did not need to fetch and decode as many instructions. After some early prototypes, the vector supercomputing era started with Seymour Cray and his Cray 1.[2] The company continued with Cray 2, Cray X-MP, and Cray T90, and finalized with the Cray X1 and X1E. Cray Research was building vector processors for 30 years. The implementation was very efficient partially due to the radical technologies that were used at the time such as transistor-based memory instead of magnetic-core and extra fast Emitter-Coupled Logic (ECL) instead of CMOS, which enabled a very high clock rate. This landscape was soon made more heterogeneous with the vector supercomputer implementations from Japan from Hitachi, NEC, and Fujitsu. The vector supercomputers introduced many innovations, from using massively multi-banked high-bandwidth memory systems, to multiprocessors with fast processor synchronization through registers, and to accessing memory by using scatter/gather instructions. For example, for the first TOP500 list in 1993, 310 of the 500 machines listed were vector processors. But by 2007, only 4 vector processors remained in the TOP500 list.

COMPUTER: What were the roots and reasons for starting the TOP500 project?

JACK DONGARRA: The TOP500 project (www.top500.org) has been tracking information about installations of supercomputers since 1993. A list of the 500 largest installations and some of their main system characteristics are published twice a year. Its simplicity has invited many critics but has also allowed it to remain useful during the advent and reigns of giga-, tera-, and petascale computing. Systems are ranked by their performance of the Linpack benchmark,[3] which solves a dense system of linear equations. Over time, the data collected allowed early identification and quantification of many trends related to computer architectures used in HPC.[4,5]

COMPUTER: How was the Linpack benchmark selected as a measure for the TOP500 ranking of supercomputers?

ERICH STROHMAIER: In the mid-1980s, Hans W. Meuer started a small and focused annual conference series about supercomputing, which soon evolved to become the International Supercomputing Conference (www.isc-hpc.com). During the opening sessions of these conferences, he used to present statistics collected from vendors and colleagues about the numbers, locations, and manufacturers of supercomputers worldwide. Initially, it was relatively obvious which systems should be considered as supercomputers. This label was reserved for vector processing systems from companies such as Cray, CDC, Fujitsu, NEC, and Hitachi, which competed in the market and each claimed theirs had the fastest system for scientific computation by some selective measure. However, at the end of that decade the situation became increasingly more complicated as smaller vector systems became available from some of these vendors and new competitors (Convex, IBM), and as massively parallel systems (MPPs) with SIMD architectures (Thinking Machines, MasPar) or MIMD systems based on scalar processors (Intel, nCube, and others) entered the market. Simply counting the installation base for these systems of vastly different scales did not produce any meaningful data about the market. A new criterion for determining which systems could be counted as supercomputers was needed. After two years of experimentation with various metrics and approaches, Hans W. Meuer and I convinced ourselves that the best long-term solution was to maintain a list of systems in question, ranking them based on the actual performance the system had achieved when running the Linpack benchmark. Based on our previous market studies we were confident that we could assemble a list of at least 500 systems that we had previously considered supercomputers. This determined our cutoff.

COMPUTER: There were other drivers beyond Cold War competition by the late 1970s and early 1980s, including the revolution in personal computing. What events spurred the funding of supercomputing in particular?

BELL: In 1982, Japan's Ministry of Trade and Industry established the Fifth Generation Computer Systems research program to create an AI computer. This stimulated DARPA's decade-long Strategic Computing Initiative (SCI) in 1983 to advance computer hardware and artificial intelligence. SCI funded a number of designs, including Thinking Machines, which was a key demonstration for displacing transactional memory supercomputers. Also, in 1982, an NSF/Intel-funded Caltech hypercube-connected multicomputer with 64 Intel microprocessor computers created by Charles Seitz and Geoffrey C. Fox was first operated to demonstrate efficacy and efficiency and stimulate further development, including, in 1985, commercial products from Intel and nCUBE.

In 1987, using a 1024-node nCUBE, Robert E. Benner, John L. Gustafson, and Gary R. Montry at Sandia National Labs won the first Gordon Bell Prize,[6] which was established to recognize progress in parallelism by showing that with sufficiently large problems, the serial overhead time could be

30 YEARS OF SC—ROUNDTABLE PANELISTS

Gordon Bell is a Researcher Emeritus at Microsoft, and he was the vice president of R&D and Digital Equipment Corporation (DEC), where he where he led the development of the first mini- and time-sharing computers. As the first NSF director for computing (CISE), he led the NREN (Internet) creation. Bell has worked on and written articles and books about computer architecture, high-tech startup companies, and lifelogging. He is a member of ACM, the American Academy of Arts and Sciences, IEEE, the National Academy of Engineering, the National Academy of Science, and the Australia Academy of Technological Sciences and Engineering. In 1991, Bell received the US National Medal of Technology. He is a founding trustee of the Computer History Museum in Mountain View. Contact him at gbell@outlook.com.

Jack Dongarra participated as both a panelist and coauthor for this Virtual Roundtable. Please see the "About the Authors" section for his biographical information.

William E. (Bill) Johnston, now retired, was formerly a Senior Scientist and advisor to ESnet—a national network serving the National Research Laboratories and science programs of the US Department of Energy (DOE), Office of Science. Johnston led ESnet from 2003 to 2008, during which time complete reanalysis of the requirements of DOE's science programs that ESnet supports was completed. Johnston has worked in the field of computing for more than 50 years, and he taught computer science at San Francisco State University at both undergraduate and graduate levels. He has a Master's in mathematics and physics from San Francisco State University. Contact him at wej@es.net.

Horst Simon is deputy laboratory director for research and chief research officer (CRO) of Lawrence Berkeley National Laboratory (LBNL). His research interests include the development of sparse matrix algorithms, algorithms for large-scale eigenvalue problems, and domain decomposition algorithms. Simon's recursive spectral bisection algorithm is a breakthrough in parallel algorithms. He has been twice honored with the prestigious Gordon Bell Prize, most recently in 2009 for the development of innovative techniques that produce new levels of performance on a real application (in collaboration with IBM researchers), and in 1988 in recognition of superior effort in parallel processing research (with others from Cray and Boeing). Simon has attended every SC conference, and contributed to many papers, panels, and tutorials. He is also one of the TOP500 authors. Contact him at HDSimon@lbl.gov.

Erich Strohmaier cofounded in 1993 with Prof. Dr. Hans W. Meuer the TOP500 project and has served as coeditor since. He is a Senior Scientist and leads the Performance and Algorithms Research Group at Lawrence Berkeley National Laboratory (LBNL). His research focuses on performance characterization, evaluation, modeling, and prediction for high-performance computing (HPC) systems and on the analysis and optimization of data-intensive large-scale scientific workflows. Strohmaier received a PhD in theoretical physics from the University of Heidelberg. He was awarded the 2008 ACM Gordon Bell Prize for parallel processing research in algorithmic innovation and was named a Fellow of the ISC conference in 2017. He is a member of ACM, IEEE, and the American Physical Society (APS). Contact him at estrohmaier@lbl.gov.

Mateo Valero is a professor at Technical University of Catalonia, UPC, and the director of the Barcelona Supercomputing Center. His research focuses on high performance architectures. He has published over 700 papers, served in the organization of more than 300 international conferences, and given more than 500 invited talks. Valero has been honored with the 2007 IEEE/ACM Eckert-Mauchly Award; the 2015 IEEE Seymour Cray Award; the 2017 IEEE Charles Babbage Award; the 2009 IEEE Harry Goode Award; and the 2012 ACM Distinguished Service Award. He is an IEEE and ACM Fellow; he holds Doctor Honoris Causa from 9 Universities; and he is a member of 8 Academies. In 2018, Valero was honored with "Condecoración de la Orden Mexicana del Águila Azteca," the highest recognition granted by the Mexican Government. Contact him at mateo.valero@bsc.es.
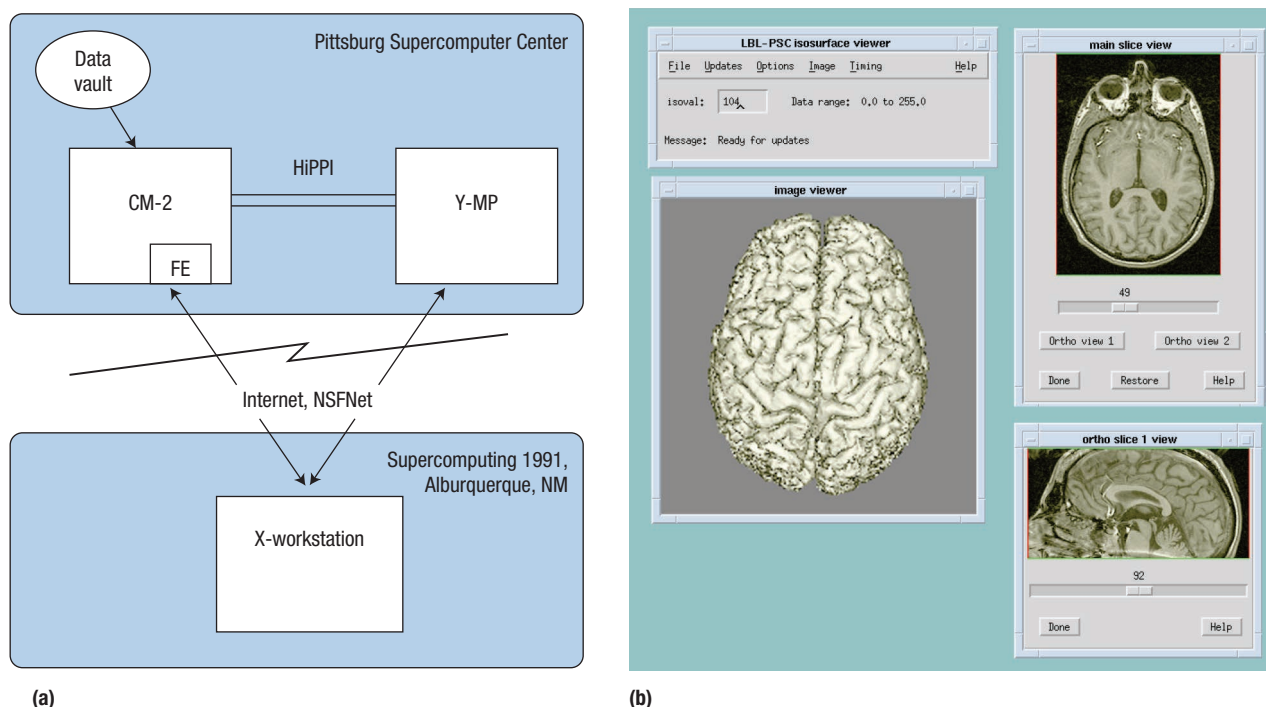
Figure 1. (a) SC91 network topology between Pittsburgh Supercomputer Center (PSC) and Albuquerque, New Mexico, for the remote visualization demonstration; (b) remote (show floor) user interface for a real–time visualization of the human brain using distributed supercomputing resources across the NSFnet between the PSC and Albuquerque at SC91.

proportionally reduced to allow almost perfect speedups. In retrospect, that first SC conference in 1988 started at exactly the right technological time to stimulate, share, and chronicle the development of the "post-Cray" era of computing. Even though the term "supercomputer" appeared in print in the early '70s and by 1980 was understood to be the largest computer of the day, the 1988 conference served to establish the industry as more than a "niche," but, more importantly, it communicated the advances of three decades.

COMPUTER: How have networking and supercomputing evolved over the years?

BILL JOHNSTON: Supercomputing and high-speed networking have evolved sometimes independently— though they inform each other—and sometimes in concert. In the early days (1980s), network access to supercomputers was limited to remote job entry

(a remote card reader) and basic job control at a few hundred bps. This was followed by implementation of the File Transfer Protocol (FTP) on supercomputers in the early 1990s as a means of getting remotely located data to supercomputer centers.

COMPUTER: Is there an event or development in supercomputer networking that stands out as a pivotal achievement?

JOHNSTON: A demonstration at SC91 was arguably the first use of wide area networks to support a high-speed, TCP/IP-based distributed supercomputer application. The overall network topology of this network is shown in Figure 1a. The challenge was real-time remote visualization of a large, complex scientific dataset that was a high-resolution MRI scan of a human brain (see Figure 1b). The approach was to use a Thinking Machines CM-2 and Cray Y-MP at the NSF's Pittsburgh

Supercomputer Center (PSC) to compute the visualization of the dataset based on input from a workstation at SC91 (in Albuquerque). These parameters were sent to PSC where the CM-2 and the Cray produced a visualization. This was then sent through a TCP circuit from the Cray into the just-built NSFNet 45 Mbps Internet backbone. NSFNet had for the first time been extended to the SC show floor, and SCInet was first setup to manage the conference networking. The SCInet LAN connected a Sun workstation, where the images were displayed. The 15 (or so) Mbps that was achieved between the Cray and the Sun was sufficient to display about 10-12 frames/sec on the Sun. Typical of distributed applications, many components had to interoperate to produce a functioning system, an especially difficult task in a wide-area network. Computer scientists at Lawrence Berkeley Laboratory, PSC, and Cray Research addressed the problems of coupling the various processes

running on three different computers, and especially debugging a newly defined TCP option that made high-speed TCP possible in the wide area.[7]

COMPUTER: What were some of the technological developments that created new solutions and new problems to solve?

BELL: In 1988, while the efficacy of large-scale parallelism was demonstrated, the problem of converting programs that ran on a mono-memory, multiprocessor supercomputer to a system running across 1,000 interconnected slower computers. It took a few more years before the circuit speed of CMOS crossed over the speed of ECL. In June 1993, a 1,024 computer Thinking Machines CM5, operating at a peak of 131 Gflops, executed Linpack at 60 Gflops to be the first-place winner of the first TOP500 supercomputer list. In the same year, Cray Research abandoned plans to deliver their evolutionary 32-processor computer that operated at a peak of 64 Gflops. Ironically, months later in November 1993, the second Top500 first-place winner was the Fujitsu Numerical Wind Tunnel (NWT) computer with 140 vector processor computers operated at 120 Gflops. The NWT computer, basically a cluster, held the position through 1996 with 170 processors. The first Intel Sandia cluster with 3680 computers was at the top for the June 1994 list. Thus, 1993 can be marked as the beginning of scalable, clustered computing! In the same year, the first draft of the Message Passing Interface (MPI) standard was introduced. A year later Donald Becker and Thomas Sterling distributed the Beowulf source code that controlled the interconnection and operation of a network of UNIX computers. Thus, at the end of the first five years, all the components were established. Figure 2 shows the situation in terms of parallelism and performance beginning in 1987 with the Bell Prize winners kicking off the transition.
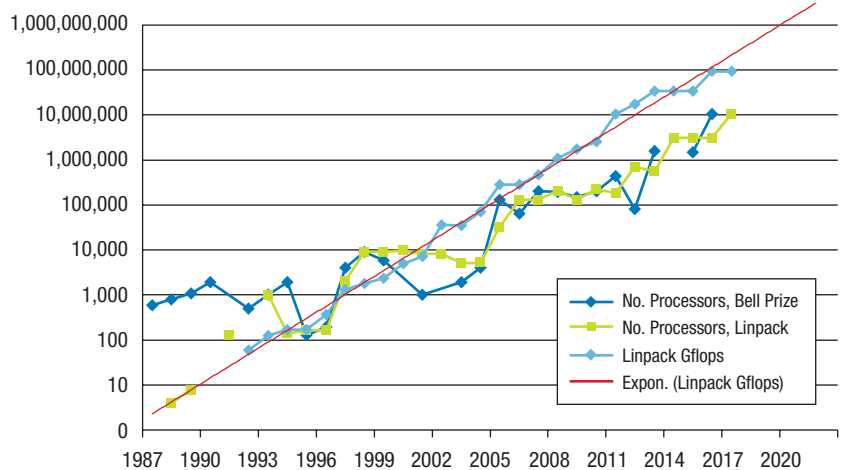

Figure 2. Three decades of performance and parallelism growth. While the Bell Prize winners demonstrated high degrees of parallelism, 1993 was the year a 1024 computer Thinking Machines CM5 dominated performance.

The 1988 through 2018 period can be trivialized by just noting that the number of parallel processing elements or cores went from 1024 nodes in 1988 to 40,960 nodes with 10,650,080 cores. The power required went from a few kW to 15,370 kW. I have argued with members of the community that the names, including single processor, constellation, MPP, and clusters, were essentially the same—multicomputer clusters. Constellation implied multiprocessor nodes, MPP implied a particular vendor network. An early SIMD was tried, made the list, and was abandoned. The SMP category was ambiguous since it included supercomputers with vector processors and multiple microprocessors that I defined as "multis."

Thus, in 2018 every computer is a multicomputer of some kind, and the performance gains come from evolving the computer nodes with some form of accelerators beginning with an attached floating unit. In 2012 a graphics processing unit (GPU) was added to the Cray Titan to establish it as the architecture de jour. Sunway has evolved the powerful node architecture by building nodes with 260 processing elements (cores), managed by a four-way multiprocessor.

COMPUTER: Taking the past as a reference, how do you see the current and future position of vector processors in the HPC space?

VALERO: The so-called "killer micros"[8] wiped out the vector processors from the TOP500 list. This shift to commodity superscalar processors was driven by economics. In particular, the Accelerated Strategic Computing Initiative (ASCI) resulted in supercomputers that were early representatives of this shift: the ASCI Red from Intel, first in the TOP500 from June1997 to November 1999 and the ASCI White from IBM, number one from June 2000 to November 2001. In any case, although there were few vector processors in the TOP500 list in this era, it was very ably represented by the Earth Simulator vector supercomputer from Japan, which dominated the top spot in 2002 and 2003 after the ASCI supercomputers. It should be added that in this supposedly stagnant era, some select companies such as NEC has continued to design pure vector processors "a la Cray" all the way from 1983 to now.

Although few classical vector processors could remain in the TOP500, their design philosophy continued to

influence "killer micro" design and associated accelerators.[9] For example, the inclusion of SIMD execution units in microprocessors could be considered as a pseudo-vector unit. The earliest SIMD units operated on short-vectors of integer data. However, the SIMD units of today are starting to resemble traditional vector processors with their ever-increasing operand size (Intel AVX-512 operates with 512 bits) as well as by their added vector-like functionality, such as support for scatter instructions in the Intel architecture. In parallel with SIMD evolution, the architecture of 3D graphic accelerators in the '90s started evolving from narrow API-driven ASIC accelerators into a more generalized form of compute called SIMT (Single Instruction, Multiple Thread), basically, a marriage between massive simultaneous multithreading and SIMD execution. The requirement to execute a SIMT instruction across multiple threads in lockstep in the GPU back-ends made this execution model quite similar to that employed by vector processors. Compared to NVIDIA GPUs, AMD's GPUs resemble vector processors even more with their internal SIMD-vector units. These vector-processor–inspired GPUs laid the groundwork for an important market spanning 3D graphics, HPC, and, more recently, deep learning. Finally, some select machines with a unique architecture—such as the Roadrunner—borrowed from vector processors too. The Roadrunner was the first petaflops machine, number 1 in the TOP500 list from June 2008 to June 2009, and it featured the Cell microprocessor from IBM.

In the meantime, the classical vector processors staged a comeback by borrowing from the "killer micro" ideas such as out of order processing. Led by pioneering academic designs in UC Berkeley[10] and UPC Barcelona[11]; the idea of designing a "commodity" vector microprocessor became feasible. This then led to multiple tentative proposals by the industry, such as the Tarantula microprocessor from

Compaq in 2000 and, finally, to the current "renaissance" of vector microprocessors. Contemporary examples of vector microprocessors include the Intel Knights family of processors, the NEC SX-Aurora[12] and Fujitsu's Post-K supercomputer design.[13]

COMPUTER: How would you characterize the changes in HPC, especially since the rapid proliferation of microcomputers?

BELL: "Scalability" characterizes this past tri-decade. Clock speed only increased a factor of 10, and gains were achieved by spending more to build by scaling—that is, replicating, adapting, and interconnecting thousands of smaller, powerful computers derived from the off-the-shelf personal computing industry. The net result has been a gain of almost one thousand per decade measured by Linpack—going from 2 Gflops (10^9) in 1988 to a likely 0.12 exaflops (10^{18}) in 2018, or a factor of 60 million with thousands of interconnected computers. The past 30 years is in contrast to the first tri-decade plus (1958–1993) that allowed Cray to focus on building the largest commercially feasible single, shared memory multiple vector processor computer for executing FORTAN. In 1958, the IBM 709 vacuum tube computer operated at roughly 10 Kflops (10^3), for a tri-decade gain of 2 Gflops/10 Kflops or a factor of 200,000 with the benefit of a thousand-fold clock increase.

By 1960, all computers were transistorized enabling higher density and faster clocks. Finally, hardware engineering vs. software and programming challenge delineates the two tri-decades of high performance computing. A summary of the events marking progress over the last 60 years is shown in Figure 3.

COMPUTER: Back in the 1990s, access to expensive supercomputers was a principal driver of the development of TCP/IP and high-speed interconnects.

What sponsored network projects come to mind as being noteworthy during that time?

JOHNSTON: In the Corporation for National Research Initiatives (CNRI) Gigabit Testbeds (~1990–1994) projects supported by the NSF and DARPA, the Casa testbed's goal was direct, long distance, high-speed communication between supercomputers. The Los Alamos National Laboratory (LANL) built an HIPPI (supercomputer local network) to SONET (wide-area optical network) gateway to interconnect supercomputers at LANL and SDSC. The 800 Mb/s HIPPI was stripped across multiple 155 Mbps SONET channels over a network path that was about 2,000-km long.[14]

The focus of the project was to interconnect supercomputers a "metacomputer" built from heterogeneous architecture systems. One goal was to couple an atmospheric circulation model running at one site with an ocean circulation model running at the other site.[15]

COMPUTER: How did the suitability of benchmarks for supercomputing evolve over time?

HORST SIMON: The simplest and most universal ranking metric for scientific computing is floating-point operations per second (flops). This benchmark would not be chosen to represent performance of an actual scientific computing application, but should very coarsely embody the main architectural requirements of scientific computing. We strongly felt that scientific HPC was largely driven by integrated large-scale calculations and therefore decided to avoid any overly simplistic benchmarks, such as embarrassing parallel codes, which could have ranked systems very high, even if they were otherwise unsuited for scientific computing. To encourage participation, we wanted a well-performing code that would showcase the capability of systems while not being overly harsh

Figure 3. Supercomputer evolution events.

or restrictive. Obviously, no single benchmark can ever hope to represent or approximate performance for the majority of scientific computing applications as the space of algorithms and implementations is too vast to allow this. The purpose of using a single benchmark in the TOP500 was never to claim such representativeness, but to collect reproducible and comparable performance numbers.

Linpack is nowadays sometimes criticized as an overly simplistic problem. The HPL (High Performance Linpack) code comes with a self-adjustable problem size, which allowed it to be used seamlessly on systems of vastly different sizes. As opposed to many other benchmarks with variable problem sizes, HPL achieves its best performance for large problems which use all the available memory and not for small problems which fit into the cache. This greatly reduces

the need for elaborate run-rules and procedures to enforce the full usage of computer systems, which is similar to what many applications do. These features made Linpack the obvious choice for our ranking. Having selected a single benchmark for comparability implies several other limitations. In Linpack, the number of operations is not measured but calculated with a simple formula based on the problem size and the computational complexity of the original algorithm. Therefore, the TOP500 cannot provide any basis for research into algorithmic improvements over time. Linpack and HPL could certainly be used for such comparisons of algorithmic improvements, but not in the context of the TOP500 ranking.

COMPUTER: Have the TOP500 data ever shown a change in the performance growth rate of installed systems?

STROHMAIER: While we started the TOP500 to provide statistics about the HPC market at specific dates, it became immediately clear that the inherent ability to track the evolution of supercomputer systems over time in a systematic way was even more valuable. Any edition of the TOP500 includes a mix of new and older installations, systems, and technologies. Figure 4 shows the changes in performance growth since the introduction of the TOP500 list in 1993.

COMPUTER: From a networking perspective, what are some of the challenges that the community has encountered? What models and architectural approaches have been developed within the HPC community to mitigate these issues for the scientific user?

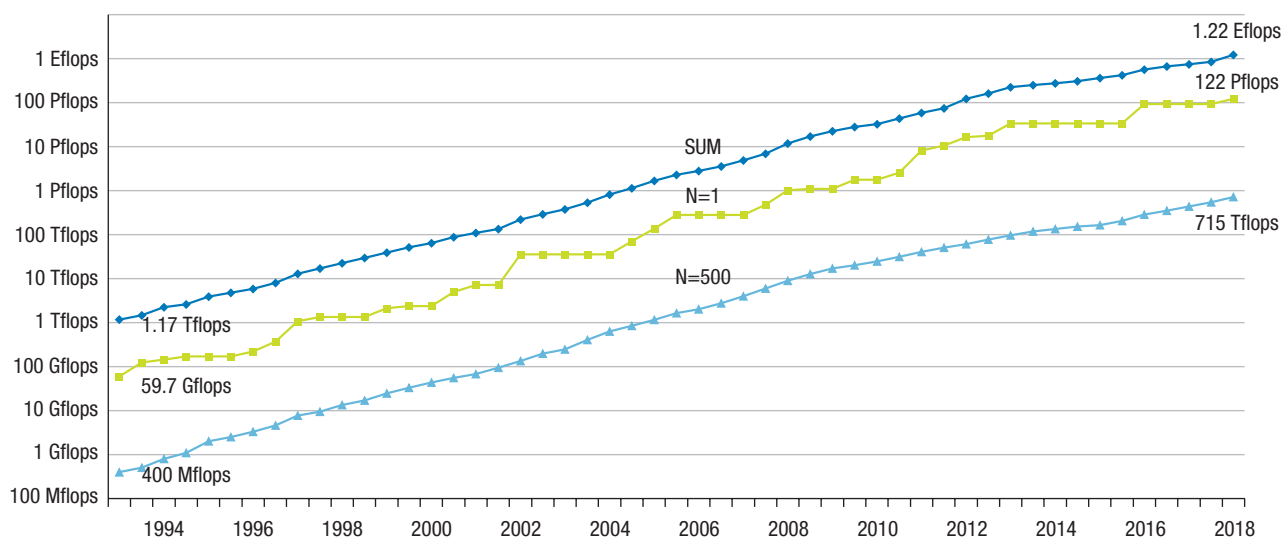**JOHNSTON:** There were several reasons ten years ago why remote user

Figure 4. Performance development of supercomputers as tracked by the TOP500. The green line shows the performance for the highest-performing system on the list, the light blue line for the lowest system (No. 500), and the dark blue line shows the sum of the performance of all systems on the TOP500.

data transfer rates to supercomputers had not significantly increased. LAN network devices are frequently poorly configured for, or even incapable of, receiving high-speed data streams from WAN devices. Storage systems have the ability to move data at high speed in the LAN but are rarely configured to move data at high speed in the WAN environment. Site security at universities and laboratories was typically handled by a (relatively low performance) firewall through which all traffic had to pass to get to computing systems on campus. This is not a problem for thousands of simultaneous small data streams (e.g., web traffic), but is a severe impediment for high-speed, long-duration streams for data-intensive science.

To achieve end-to-end high-speed data throughput for large volume science data all of these issues had to be addressed. Discussions between ESnet (DOE's Office of Science's WAN network) and the NERSC supercomputer center in the early part of 2000 established some basic principles from which ESnet developed a network architecture called the "ScienceDMZ" that addressed the issues.

The ScienceDMZ is a special campus network domain that is built outside the site perimeter but directly adjacent to the site LAN so that it can share LAN connections with the site. It consists of a WAN-capable network device, and a small number of high-performance data transfer systems ("data transfer nodes" [DTN]). The DTNs typically also have a connection to the campus LAN that does not go through the site firewall, but data transfers in either direction have to be initiated from within the site. The control channels for these transfers go through the site firewall. Cybersecurity within the ScienceDMZ is accomplished by well understood server configurations on the DTNs that only run software needed to do data transfer. Access control is managed with an access control list (ACL) on the ScienceDMZ WAN network device. These ACLs restrict access to external sites that are identified as collaborators that have a valid reason to exchange data with a scientist on campus. This concept has been very successful and is now deployed at more than a hundred laboratories, research universities, and supercomputer centers.[16]

The national research and education networks (NRENs) of the Americas, Europe, and Southeast Asia have extended their multi-hundred Gbps backbones across the Atlantic and Pacific oceans, providing high-speed data access internationally. (Transatlantic R&E bandwidth now at record-breaking 740 Gbps.) Such high-speed networks are essential for getting very large amounts of data from instruments to supercomputers.[17]

ESnet has deployed 400 Gbps link technology, providing NERSC computers with remote access to cache disks and mass storage systems. A similar approach is used at CERN where the local disk cache is divided across the CERN Geneva site and the Wigner Data Center in Budapest. This technology, probably at the Tbps level, will also likely be used to connect the next generation Linac Coherent Light Source at the SLAC laboratory to NERSC.

Since the 1980s, network access to supercomputers, and the corresponding ability to move vast amounts of data to and from supercomputers, has increased by more than nine orders of magnitude. This increase is based

on improvements in architectures, software, operating systems, and network technology, much of which was enabled by research and development funding from science-oriented government agencies.

COMPUTER: What are you looking for primarily in an additional complementary benchmark for the TOP500?

DONGARRA: Most requests for new benchmarks usually center on the argument that Linpack is—at least at present—a poor proxy for application performance and that a "better" benchmark is needed. When HPL gained prominence as a performance metric in the early 1990s, there was a strong correlation between its predictions of system rankings and the ranking realized by full-scale applications. In these early years, computer system vendors pursued designs that would increase HPL performance, thus improving overall application function.

However, many aspects of the physical world are modeled with PDEs, which help predictive capability, aiding scientific discovery and engineering optimization. The High-Performance Conjugate Gradients (HPCG) benchmark[18] is a complement to the HPL benchmark and now part of the TOP500 effort. It is designed to exercise computational and data access patterns that more closely match a different yet broad set of important applications, and to encourage computer system designers to invest in capabilities that will impact the collective performance of these applications.

COMPUTER: How do you see the future developments of supercomputer performance rankings?

SIMON: Clearly, the current approach for compiling the TOP500 cannot address truly novel architectures such as neuromorphic systems or quantum computers. Should a market for such systems develop, very domain-specific approaches to benchmarking and

SUPERCOMPUTING HISTORY PRESENTATIONS

» Gordon Bell, "View of History of Supercomputers," presentation, Lawrence Livermore National Lab, 24 April 2013; https://www.youtube.com/watch?v=e5UbGgRGGOk.
» Gordon Bell, "Three Decades of the Gordon Bell Prize," presentation, Frontiers of Computing, March 2017; https://www.youtube.com/watch?v=NZlGOo3_3No.
» Gordon Bell, "Marking 30 Years' History of the Gordon Bell Prize," presentation, SC Conference, Nov. 2017; https://youtu.be/4LCXbpssV1w.
» Jack Dongarra, Erich Strohmaier, and Horst Simon, "Top500: Past, Present, and Future," presentation, SC Conference, Nov. 2017; https://youtu.be/elZZbfrW87M.

ranking would need to be developed, which is very like the situation for data-intensive computing.

The TOP500 collection has enjoyed incredible success as a metric for the HPC community. The trends it exposes, the focused optimization efforts it inspires and the publicity it brings to our community are very important. As we are entering a market with growing diversity and differentiation of architectures, a careful selection of appropriate metrics and benchmarks matching the needs of our applications is more necessary than ever.

HPL encapsulates some aspects of real applications such as strong demands for reliability and stability of the system, for floating point performance, and to some extent network performance, but no longer tests memory performance adequately. Alternative benchmarks as a complement to HPL could provide corrections to individual rankings and improve our understanding of systems but are much less likely to change the magnitude of observed technological trends.

COMPUTER: Do you see any emerging applications outside of the classical HPC domain? How about the applicability of supercomputing ideas to

databases, personalized medicine or deep neural networks?

VALERO: Modern applications such as deep neural networks (DNNs), database management systems (DBMSs) for big data, and personalized medicine (PM) are much more amenable for efficient execution on vector processors. Note that current DNN applications typically feature multiply add operations on huge vectors of data and can benefit from vector architectures as well as DBMSs and PM applications such as gene sequencing that operate on very long vectors with integer operations.

BELL: The massive computer cluster with highly parallel computing nodes describes today's architecture path. Will this general structure be adequate to get to exaflops and beyond, with a clock speed stalled at a few GHz? So far two paths have emerged based on advances in AI including the construction of large neural nets for recognition: specialized chips, such as Google's TPU and FPGAs programmed for the application.

COMPUTER: Networking is just one element of an HPC infrastructure as reflected in the SC conference topic areas

that have come to include not just performance and networking, but also storage, data analytics, and visualization. What project do you consider an exemplar of a current state of the art infrastructure?

JOHNSTON: By far the largest scientific experiment today is the Large Hadron Collider at CERN. Data from the several detectors/experiments on the LHC are distributed to several thousands of scientists at some 200 institutions in more than 40 countries for analysis. This results in petabytes/day of data movement. Some of this involves the use of supercomputers, but even more so the technology and skills needed to accomplish this sort of data management are moving into the supercomputing environment as supercomputers are increasingly used to manage and analyze the vast amounts of data from modern scientific instruments. These instruments are almost always remote from supercomputers and involve collaborations that are widely distributed. Moving petabytes of data into and out of supercomputer centers from remote experiments required new technologies.

COMPUTER: How do you see the holistic approach between applications, programming models, runtime systems and architecture in the future supercomputers?

VALERO: Looking forward, we see three developments that might facilitate the resurgence of vector processors: technology evolution, emergence of modern applications, and runtime-aware architecture. Let us consider each in turn. In a back-to-the-future sense, technological advances, similar to the early vector supercomputer period, drive the new vector renaissance. For example, the memory stacking technology such as HBM, which delivers high bandwidth DRAM systems, is hugely advantageous for vector processor designs since it provides a good technology solution

ABOUT THE AUTHORS

JACK DONGARRA holds an appointment at the University of Tennessee, Oak Ridge National Laboratory, and the University of Manchester. He specializes in numerical algorithms in linear algebra, parallel computing, use of advanced computer architectures, programming methodology, and tools for parallel computers. He was awarded the IEEE Sidney Fernbach Award in 2004; in 2008 he was the recipient of the first IEEE Medal of Excellence in Scalable Computing; in 2010 he was the first recipient of the SIAM Special Interest Group on Supercomputing's award for Career Achievement; in 2011 he was the recipient of the IEEE Charles Babbage Award; and in 2013 he received the ACM/IEEE Ken Kennedy Award. He is a Fellow of the AAAS, ACM, IEEE, and SIAM; he is a foreign member of the Russian Academy of Science, and a member of the US National Academy of Engineering. Contact him at dongarra@icl.utk.edu.

VLADIMIR GETOV is a professor of distributed and high-performance computing (HPC) and leader of the Distributed and Intelligent Systems research group at the University of Westminster. His research interests include parallel architectures and performance, energy-efficient computing, autonomous distributed systems, and HPC programming environments. Getov received a PhD and DSc in computer science from the Bulgarian Academy of Sciences. In 2016 he was the recipient of the IEEE Computer Society Golden Core Award. Getov is a Senior Member of IEEE, a member of ACM, a Fellow of the British Computer Society, and he is Computer's area editor for HPC. Contact him at v.s.getov@westminster.ac.uk.

KEVIN WALSH is a student of the history of HPC. He is the supercomputing history project lead for the 30th anniversary of the IEEE/ACM SC conference in 2018. Previously a systems engineer at the San Diego Supercomputer Center, he is currently at the Institute of Geophysics and Planetary Physics at the Scripps Institution of Oceanography. Walsh received a BA in history of science and an MAS in computer science and engineering at the University of California, San Diego. He is a member of ACM, the IEEE Computer Society, and the Society of History of Technology. Contact him at kwalsh@ucsd.edu.

to the issue of high bandwidth requirements of vector processors. We envisage that instruction set architectures (ISA) supporting operations on long vectors or matrix structures will play an important role in the future. The high semantic level of such operations and their tight coupling to modern runtimes will allow programmers to convey semantic information they already have (on locality, dependences) to the architecture, reducing the need to rediscover as it has been

done in current scalar ISAs. This will allow decoupling the frontend and backend of processors and to explicitly manage locality (long register files, "command vectors") optimizing the memory throughput.

To summarize—vector processors were paramount at the very beginning of supercomputing from the Cray 1 in 1976 to the Convex C4 in 1994. Despite the "Attack of the

Killer Micros," vector processors never disappeared and now they could be the crème de la crème of supercomputers once again. In addition, ISA operations that represent a very large amount of work offer the possibility to keep active a large number of functional units. This will allow the development of energy efficient systems for dedicated highly critical applications such as AI applied to personalized medicine or self-driven vehicles. Programming models and runtime systems will need to adapt and support this new approach driving the supercomputing performance well beyond the exascale barrier. ∎

REFERENCES

1. G.M.Amdahl,"ComputerArchitectureand Amdahl'sLaw,"*Computer*,vol.46,no.12, 2013; pp. 38-46.
2. "Cray-1 Computer System Hardware Reference Manual," Publication 2240004, Rev C, 4 Nov. 1977, Cray Research, Inc.; http://history-computer .com/Library/Cray-1_Reference%20Manual. pdf.
3. J.J. Dongarra, P. Luszczek, and A. Petitet, "The Linpack Benchmark: Past, Present and Future," *Concurrency Computat.: Pract. Exper.*, vol. 15, 2003, pp. 803–820; doi: 10.1002/cpe.728.
4. E. Strohmaier, et al., "The Marketplace of High-Performance Computing," *Parallel Computing*, vol. 25, no. 1517, 1999.
5. E. Strohmaier, et al., "The TOP500 List of Supercomputers and Progress in High Performance Computing," *Computer*, vol. 48, no. 11, 2015, pp. 42–49.
6. G. Bell, et al., "A Look Back on 30 Years of the Gordon Bell Prize," *Int'l J. High Performance Computing Applications*, vol. 31, no. 6, 2017, pp. 469–484.
7. W. Johnston, "High-Speed, Wide Area, Data Intensive Computing: A Ten-Year Retrospective," *Proc. 7th IEEE Symp. on High Performance Distributed Computing*, 1998.
8. J. Markoff, "The Attack of the 'Killer Micros,'" *The New York Times*, 6 May 1991; www.nytimes.com/1991 /05/06/business/the-attack-of -the-killer-micros.html.
9. M. Valero, R. Espasa, and J.E. Smith, "Vector Architectures: Past, Present and Future," *Proc. ACM Int'l Conf. Supercomputing* (ICS 98), 1998, pp. 425-432.
10. K. Asanovic, "Vector microprocessors," PhD Thesis, University of California, Berkeley, 1998; http:// people.eecs.berkeley.edu/~krste /thesis.html.
11. R. Espasa, "Advanced Vector Architectures," PhD Thesis, Universitat Politechnica de Catalunya, 1997; http:// citeseerx.ist.psu.edu /viewdoc/download?doi=10.1.1.51 .9455&rep=rep1&type=pdf.
12. "NEC SX-Aurora TSUBASA—Vector Engine," NEC Corporation, 1994-2018; www.nec.com/en/global /solutions/hpc/sx/vector_engine .html.
13. T. Shimizu, "Post-K Supercomputer with Fujitsu's Original CPU, Powered by ARM ISA," Teratec Forum; www .teratec.eu/library/pdf/forum/2017 /Presentations/04_Toshiyuki _Shimizu_Fujitsu_Forum_Teratec _2017.pdf.
14. "The Gigabit Testbed Initiative, Final Report," CNRI, Dec. 1996; www.cnri .reston.va.us/gigafr.
15. W. Minkowycz, "Advances in Numerical Heat Transfer, Volume 2," CRC Press, 5 Dec. 2000.
16. E. Dart, et al., "The Science DMZ," *Proc. Int'l Conf. High Performance Computing, Networking, Storage and Analysis* (SC 13), 2013.
17. "North Atlantic Network Collaboration Building Foundation for Global Network Architecture," Energy Sciences Network, 17 Apr. 2017; http:// es.net/news-and-publications /esnet-news/2017/north-atlantic -network-collaboration-building -foundation-for-global-network -architecture.
18. J.J. Dongarra, M.A. Heroux, and P. Luszczek, "High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing system," *Int'l J. High Performance Computing*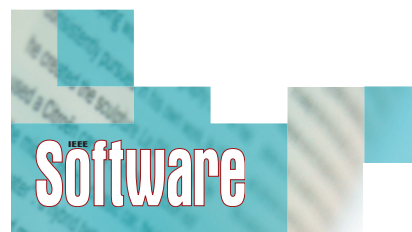 Applications, vol. 30, no. 1, 2015, pp. 3–10.