

China – A New Power in Supercomputing Hardware

Joshua Alspector, Alfred E. Brenner, Robert F. Leheny, and James N. Richmann
Information Technology and Systems Division
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311

EXECUTIVE SUMMARY

The global high-performance computing (HPC) landscape is evolving rapidly. Long dominated by the United States, many nations now participate in a dynamic HPC development environment with China making the most rapid progress. Driving new developments are fundamental changes in the integrated circuit core computing engines as well as a compelling need for energy efficiency and new computing architectures. This article provides a perspective on the current state of HPC, focusing on hardware development in China where significant new supercomputing progress is occurring.

OVERVIEW

US, Japan, and China: Strong Players in HPC

The global high-performance computing (HPC) competitive environment is very different today from what existed ten years ago. Most notably, China, which was barely visible in HPC in 2002, emerged in November, 2010 with the top machine on the TOP500¹ list of the world's fastest machines. Most recently, in the November, 2012 TOP500 list, China's top entry is now eighth on the list with the US holding the top two positions, and three other entries in the top 10. At the same time, after a concerted, government-financed program, Japan, which fell out of the HPC leadership in the last decade, emerged in 2011 to temporarily recapture the top TOP500 spot, but then slipped in November, 2012 to the number three position. This paper focuses on China as the newly emergent competitor in supercomputing. As a basis for comparison, the paper also considers the impact of new accelerator microchip technology primarily developed by US companies. This competition in supercomputer performance is significant not only for traditional supercomputing markets but also for emerging data center applications and, notably, for defense applications.

China's emergence as a top performer in HPC, and its developments in chips, architecture, interconnect, and cooling technology, reflect its broader challenge to America's leadership in microelectronics and computing. Notably, Chinese HPC developments focus on energy efficiency with an apparent view toward meeting requirements for an exascale (a billion billion operations per second) system with a hundred to a thousand times the computing power of today's most advanced systems. For example, the Sunway BlueLight HPC uses Chinese developed ShenWei CPU microchips, and is capable of nearly a peta (a million billion) double-precision (DP) floating-point-operations-per-second (pflops) while consuming a modest (for supercomputers) 1.1 megawatts of power.

General-purpose graphics processing units (GPGPUs) and the similar Intel Many Integrated Cores (MIC) technology, largely invisible in HPC ten years ago, are beginning to be widely used in today's new system designs as parallel co-processors to accelerate computation and have enabled, along with greater numbers of multiprocessor CPU chips, supercomputer speeds to reach pflops levels.

Overall, while the Chinese producers use some home-grown technologies, they and producers in other countries still depend significantly on US commercial technology, and while the US is still the intellectual and commercial leader in computing, China has developed a substantial domestic HPC technology base. The Japanese, long in competition with the US, have a well-developed and mature sophistication in their products, dominated by NEC and Fujitsu. The Chinese, only newly invested in supercomputing, are less sophisticated, but after a considerable three-decade-long, concerted national investment in science and technology, combined with policies inviting direct foreign investment, especially in technology, are making good progress in catching up. China has a stated goal of reducing dependence on foreign technology and has put into place *indigenous innovation* policies (including export control) to promote this goal. Over the last ten years, Chinese production of microchips increased by an order of magnitude, and, unlike the

equivalent growth in Japan during the 1980s, the internal Chinese market for microchips is very significant; currently roughly equal to the rest of the world.

One issue that hasn't changed (in over ten years) is that developing effective software for parallel architecture machines with large numbers of processors is still the most pressing problem, recently made even more complex with the integration of accelerator chips in the mix, holding back progress in supercomputing. Solving the parallel programming problem is likely the key to capturing future leadership in computer science and computational supercomputing.

The Push for Exascale

Supercomputing leadership is an important contributor to defense leadership. If, as expected by 2020-22, a country can produce an exaflop (a billion billion flops) supercomputer that occupies a large room, it can apply the same technology on a smaller scale to put a petaflop in a box on-board for applications on an aerial platform. A system on this scale can provide the compute power necessary for making sense of video taken from above, for tracking targets, and for autonomous navigation without ground control. This ability will be important for dominance in any theater and especially for large standoff applications where a pilot on-board would limit range and effectiveness. Unmanned Aerial Vehicles (UAVs) can provide long standoff military capabilities on offense and deny such capabilities on defense. These long-standoff systems will be especially important in planning for attack and defense.

Furthermore, there are many software applications, especially in defense system design, where exascale computers can provide crucial system engineering and design capabilities. Combustion and other advanced simulation capabilities will be needed for the next generation of high-performance aircraft. Computer simulations of combustion, important for jet engine design, are beyond current supercomputers capabilities and require exascale machines. These capabilities, applied to the weapons acquisition process, would enable reduced cost and shorten times to deployment. Acquisition of advanced jet engines is high on the shopping list for Chinese military defense planners who were previously reliant on Russian technology.² Another important simulation application is in nuclear weapons design where simulations replace dangerous, costly, and forbidden testing.

For autonomous aircraft and intelligent, autonomous weapon systems, lightweight, low-power, and space-efficient parallel computation systems lead directly to higher performance and lower cost. In the case of airborne or remote surveillance systems, the need to transmit information over a limited bandwidth channel drives the need for on-board processing capabilities. Real-time content understanding will allow for the transmission of just the objects, events, tracks, and other analytic outputs which require much less downlink bandwidth than transmitting every pixel. Furthermore, the need for detailed remote control and analyst communication in real-time will be lessened greatly if enough computational intelligence can be placed on-board. This will also reduce the need for off-line analysis of collected data for intelligence.

For Chinese military planners, especially, catching up to the West and Japan in supercomputing technology, is seen as an important goal. The National University of Defense Technology (NUDT) is a leader in Chinese HPC and is well-funded for this effort. In addition, this goal is

wholly compatible with China’s mercantilist ambitions in high technology semiconductors, software, and systems. Semiconductor computer chips are found in laptops, servers, cell phones, and other high-technology commercial products. These are the same semiconductor chips that in aggregate compose modern supercomputers and represent a dual-use technology. The following sections will discuss the semiconductor hardware that will enable massively parallel exascale systems within the next ten years.

THE STATE OF SUPERCOMPUTING

Traditionally, the TOP500 list has ranked the world’s supercomputers according to their performance on the LINPACK (see <http://www.netlib.org/benchmark/hpl>) benchmark. The US has recently recaptured the top ranking after several years in which first China (2010) and then Japan (2011) held the top spot. China, Japan, and the United States depend to varying degrees on governmental support to develop original hardware technology for HPC. It should be noted that benchmarks other than LINPACK and ranking lists other than TOP500 are receiving increasing attention to reflect the realistic and growing importance on more relevant supercomputer applications and energy efficiency.³ China has 72 machines on the list of the fastest 500 supercomputers, second to the United States at 251 machines. A notable trend is that 62 machines now incorporate parallel computation accelerators such as NVIDIA’s graphical processing units (GPUs) rather than relying solely on CPU cores like those represented by Intel’s Xeon or AMD’s Opteron.

Central Processor Unit (CPU) chips

The traditional measure of flops (floating point operations per second) for speed of computation derives from the computational units for the chips within a supercomputer which use this performance metric. For an application, utilizing all the theoretical compute power of the components is important. In the past decade, these components have evolved from single computational core general-purpose CPU chips to multiple core CPU chips and, most recently, include specialized accelerator chips with even larger numbers of cores. Table 1 below lists general-purpose CPU chips used in HPC. The number of cores on a chip ranges from 4 to 16.

Type of CPU	Freq. (GHz)	Process (nm)	No. of cores	Area of Die (mm ²)	Power consumption (W)	Peak Performance (gflops)	Ratio (gflops/W)
Intel Core i7 980 XE	3.2	32	6	240	130	107.55	0.827
Intel Sandy Bridge (Ivy Bridge)	3-4 (?)	28 (22)	6 (4)	370 (160)	130 (77)	256 (?)	1.96 (3?)

AMD Opteron X12	2.4	45	12	346	130	152	1.16
IBM Power7	3-4.1	45	8	567	100	264.96	2.64
IBM Power X CELL	3.2	45	9	221	80	100	1.25
Fujitsu SPARC64 VIIIfx (IXfx)	2 (1.85)	45 (40)	8 (16)	513 (484)	58 (110)	128 (140)	2.2 (2)
Godson 3B-1500	1.35	32	8	182	40	172.8	4
Shenwei 1600	1.1	65	16	342	70	140	2
Feiteng FT-1000	1	65	8	360	50	8	0.16

Table 1: CPU chips used in high-performance computing. Note that the next Intel CPU design labeled Ivy Bridge (mostly a shrink to 22 nm of the 28 nm Sandy Bridge design) is not yet in production and has approximate figures in parentheses. All performance figures are for double-precision.

It is noteworthy that the Chinese Godson 3B-1500 design has the projected⁴ excellent gflops/watt metric of 4. To achieve exascale performance at acceptable total power levels, this figure will have to rise⁵ to at least 75 gflops/watt and perhaps as much as 1000 gflops/watt. The latter figure will result in an exascale product dissipating about 1 megawatt for just the CPU chips without considering the rest of the machine. Also note that both the Intel and AMD chips are based on complex instruction set computer (CISC) architectures, in particular the Intel x86 architecture. This architecture is based on the 8086 CPU first introduced in 1978. The others, including those from China and Japan, are based on more modern reduced instruction set computer (RISC) architectures.

Graphical Processor Unit (GPU) and Other Processor Chips

The use of GPUs in a hybrid CPU/GPU accelerator design seems to be the current trend, while the programming skills and application software needed to make most efficient use of the increased parallel processing capacity is in short supply. NVIDIA provides the GPU chips used in most hybrid HPC designs. NVIDIA recently announced a Kepler (K) series of GPUs, which are three times as energy-efficient as the current generation, and are likely to have an important impact on HPC designs.⁶ For example, the Kepler K20 compute board containing the K2 chip has been incorporated into the new Titan supercomputer at the Oak Ridge National Laboratory in Tennessee and the Blue Waters system at the National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign. The latter machine is currently the world's fastest on the TOP500 rankings. NVIDIA also supports the Compute Unified Device Architecture (CUDA) language which extends C and C++ to facilitate programming the multi-threaded GPUs.

Table 2 lists performance parameters for the mobile System on Chip (SoC) designs, which include a number of Many Integrated Core (MIC) and GPU chips used in more recent HPC system designs. Note that the GPU and MIC designs, whose compute power derives mainly from

parallelism, use vector concepts such as single instruction multiple data (SIMD) architecture and streaming to boost performance. Existing hybrid HPC designs use NVIDIA’s Fermi series of GPU cores, whose properties are also provided in the table. NVIDIA’s new K series, and especially the 7 billion transistor K2 design, is well-suited for HPC. The new benefits of the NVIDIA hardware optimization features are their promotion of higher utilization of the GPU cores.⁷

As their preferred route to HPC acceleration, Intel has chosen to build a SIMD co-processor, the Xeon Phi, composed of many x86 cores.⁸ The Xeon Phi (which will best operate in concert with a CPU such as the Xeon E5 in a supercomputer system) is the basis for the University of Texas’ Texas Advanced Computing Center Stampede supercomputer. However, regardless of which chip is chosen, porting existing code to a new co-processor is a major challenge.

Type of Processor	Freq. (GHz)	Process (nm)	No. of cores	Area of Die (mm ²)	Power consumption (W)	Peak DP Performance (gflops)	Ratio (gflops/W)
Intel Atom z2460 “Medfield” (“Clover Trail”)	1.6	32	1 (2)	~62	~2 ?	~6 ?	~ 3 ?
NVIDIA Tegra 3 (ARM Cortex A9)	1.3	40	4	~80	~2 ?	~10 ?	~5 ?
Samsung Exynos-5 Dual (ARM Cortex A15 CPU, Mali-T604 GPU)	1.7	32	2	?	5	6 (CPU) 36 (GPU)	8.4
Intel Xeon Phi 5110P MIC Vector co-processor	1.0	22	60	Big ?	~225	~1000	4.4
NVIDIA “Fermi”	1.3	40	512	332 1.96 B xtrs	240	665	2.8
NVIDIA “Kepler” K1	0.75?	28	1536	3.5 B xtrs	185	95 (2290 SP)	0.5
NVIDIA “Kepler” K2	0.6?	28	2880 SP 960 DP	7.1 B xtrs	225	1200	5.9

Table 2: Mobile System on Chip (SoC) chips, which include Many Integrated Core (MIC) and GPU chips, used in recent HPC designs. Note that for some chips the table lists single precision (SP) as well as double precision (DP) figures. Transistors are abbreviated as *xtrs* in the table.

Also listed in Table 2 are chips featuring low power meant for mobile processing. Although these SoC designs incorporate more than just CPUs on a chip, we include them here because their GFlop/watt metrics are very impressive and suggestive of future capabilities. Most smartphone processors use Advanced RISC Machines (ARM) core designs to achieve low power performance, and a representative NVIDIA chip, the NVIDIA Tegra 3, that takes this approach is listed in the table. Samsung Exynos chips using ARM CPU and GPU cores will be the basis for the HPC system being developed at the Barcelona Supercomputer Center (BSC) as part of the European Mont-Blanc project.⁹ The BSC goal is to create technology to build the most energy-efficient supercomputer in the world. In the near term the BSC is aiming to demonstrate the technology for a 200 pflops machine by 2017.

Special-purpose chips like GPUs and low-power processor chips for mobile applications may point the way for energy-efficient application-specific HPC design. For example, there was the recent demonstration of the TI C6678 digital signal processor (DSP) for single-precision matrix multiplication application with a gflops/w performance of 7.4.¹⁰ In another example, Intel has announced its collaboration with Inspur, a Chinese server company to incorporate Xeon Phi accelerators in a new supercomputer. Inspur is significant for its work with the National University of Defense Technology (NUDT) on the Tianhe-1A supercomputer. Also of note is the Samsung Exynos chip which incorporates two low-power ARM cores with a Samsung designed Mali GPU core for increased integration. Because they promote energy efficiency it can be expected that further integration of heterogeneous cores on processor chips will be used in future supercomputers.

Ideally, application software must be optimized for the hardware on which it runs in order to distribute the computational load across the hardware elements in a parallel and efficient manner. Considerations include the use of memory-near-processing to avoid energy-intensive data movement and the decomposition of computation to make best use of specialized hardware such as GPUs. Although it requires a considerable effort, optimization of parallel software is most effective when done in concert with hardware design and tuned for a specific application. Co-design is a promising direction for progress towards exascale systems.¹¹ Further discussion of HPC application software and hardware architectures to support the software is beyond the scope of this paper but is a likely route to future progress in supercomputing.

CHINA'S SUPERCOMPUTER DEVELOPMENT

With China's recent emergence as an important player in HPC, it's interesting to examine in detail how the Chinese planners have developed a complete computing ecosystem over the past ten years. Supercomputer development in China, pushed largely to promote industrial development, has contributed to developing domestic companies capable of producing highly competitive servers, handheld mobile appliances, integrated circuit designs, multicore microprocessors, digital signal processors, secure cryptographic chips, secure operating systems, and HPC software. These accomplishments are the fruit of three decades of focused national investment in science and technology and nationalist industrial policies, and are a key component of the phenomenal economic growth in China during that period. These focused strategic and industrial policies have resulted in substantial technical capabilities in people, institutions, companies, technologies, and products.

The government high-performance computing program has been organized around three main HPC machine efforts: the Sunway Blue Light, the Dawning (now Sugon) 6000, and the Tianhe-1A, which in turn are supported by three RISC multicore microprocessor developments: the ShenWei, the Loongson (Godson), and the the FeiTeng CPU chips, respectively. Though government-funded, the Godson development, with companies such as Lenovo involved, is focused on commercialization, while development of the other chips has been supported largely on defense funding.

A Tsinghua University study¹² attributed the Chinese integrated circuit (IC) industry's rapid rise partially to the globalization of the IC development and manufacturing process. Whereas in the 1970's and 1980's, IC development was largely the province of vertically integrated device manufacturers like Fujitsu and Intel, in the 1990's the advent of specialized IC foundries and IC design houses allowed new entrants to enter the global production chain at any segment (fabless design, foundry fabrication, or package and test), thus lowering the barrier to entry. Furthermore, the mobility of expertise across national borders allowing new firms to easily tap into new technologies accelerated this process.

Hardware

The Tianhe-1A supercomputer, which held the top spot for 6 months in 2011, was developed by the NUDT. Its speed derives primarily from the use of Intel CPUs and NVIDIA GPUs, but its design also incorporates domestically developed FeiTeng FT-1000 CPU chips that utilize the SPARC architecture, combined with domestically-developed interconnect technologies. The Chinese Academy of Sciences (CAS) Institute of Computer Technology (ICT), which has historically been in the forefront of HPC development in China, developed the Dawning 6000 HPC. This machine uses Loongson, also known as Godson, CPU chips based on the MIPS architecture and emphasizing low power dissipation. The Jiangnan Institute of Computer Technology and the server company Inspur have developed the Sunway Blue Light HPC which uses domestic ShenWei SW-1600 CPU chips based on the Alpha architecture. This machine has an advanced water-cooling system and a domestic Infiniband-like interconnect.

The Godson microprocessor development started a decade ago under China's tenth five-year plan and has produced, in a 65 nm manufacturing process, the Godson-3B chip, which is a modern 64-bit, four and eight-core design with advanced features such as superscalar out-of-order instruction pipelines and vector processing. The latest up-grade, the Godson-3B 1500, will likely use a 32 nm process and should be available in 2013. The Godson design has a mesh network interconnecting the cores as opposed to the ring interconnect used in Intel multicore designs. One feature of this approach is it results in less data movement and, consequently, results in the lower power dissipation important for HPC design.

ICT, the leading HPC research institute in China, announced in the 12th Five-year Plan (2011-2015) that it would develop a 100 pflops Dawning 7000 system. This team is developing a many-core processor for the Godson-T project in an approach similar to the Intel Phi chip. The plan¹³ is to provide hardware architectural support for multithreaded programs with goals similar to the new Kepler GPUs discussed previously. The Godson-T processor has 64 cores connected by dual 2D-Mesh networks and a two-level on-chip hierarchical memory. Initially, a 16-core

prototype chip has been manufactured by China's Semiconductor Manufacturing International Corporation (SMIC) using 130nm CMOS technology; the 64-core prototype chip design was scheduled to be complete in 2012, with a 256-core chip design planned for 2015. These chips, similar to the American GPUs and MIC accelerator chips, will be domestically produced.

The Chinese government has also started a program for creation of a national processor architecture. This architecture could become a requirement for use in any project seeking government funding for purchases such as computers or smartphones. Establishing such a standard may facilitate more rapidly deployed and less expensive high performance computers, but it might also stifle innovation. A similar initiative to create a standard for cryptographic processors led to a requirement that all computers in China use the Trusted Cryptography Module (TCM) standard rather than the ISO (International Standards Organization) TPM (Trusted Platform Module) standard for trusted computing. Another development in self-sufficiency is the program to realize the RISC Harmony Unified Processing Architecture by the Chinese firm ICube. In this design each core can process a parallel computing stream with both CPU and GPU-type elements to achieve low power with high efficiency. The first of these chip designs is intended to run the Android operating system.¹⁴

Chinese HPC development strategy displays aggressive use of GPUs. Four of the top five systems combine Intel Xeon and NVIDIA Tesla processors. An exception is the Sunway Blue Light which uses ShenWei domestic CPUs and features indigenous CPU chips and low-power, compact system design. Interconnect systems are mostly Infiniband and gigabit Ethernet but the Tianhe-1A uses a proprietary Galaxy interconnect with two stages and high fanout. About half of today's Chinese supercomputers were built by domestic manufacturers such as Dawning, Inspur, Lenovo, and Sunway using mostly US-designed processor chips. It can be expected that future Chinese HPC development will be increasingly based on domestic CPUs, accelerator chips, and interconnects.

Software

Until recently, there has been little apparent Chinese interest in supporting the development of software, spanning the complete spectrum from operating systems and compilers to applications and support tools. Consequently, almost all software in Chinese systems has been either proprietary systems offered by the non-Chinese computer and chip providers, or applications providers, or from open source developed in the West or Japan.

However, there is good evidence that the Chinese have wide experience with a very large range of commercial applications software. China has imported almost every major commercially-available application package for scientific research, industrial design, finance and business functions, simulations, and diagnostic support, which consequently has educated parts of the Chinese IT community. Chinese national funding organizations are now beginning to support the indigenous development of the required applications software. In many cases, there is encouragement and funding for foreign participation in proposed projects, as long as the principal investigator is Chinese.

HPC applications in China are similar to those in use elsewhere, with energy, industrial, and research areas dominating. There are also applications in telecommunications, weather,

biotechnology, and finance. The industrially focused programs include those for aircraft design, drug discovery, animation, and structural analysis. In almost all cases, the object appears to be to develop *productive approaches* for their designs, capable of scaling to very large numbers of processing elements, indicating that the Chinese HPC research community fully appreciates the hard software problems impeding progress in HPC-delivered performance.

Chinese Science and Technology Research Funding Channels¹⁵

China is a prime example of the power of government to effect industrial change by central planning which in this case has enabled China to achieve a highly competitive capability in enabling microelectronics and HPC technologies. Science and technology research funding by the Chinese central government comes through several organizations with little apparent coordination of their individual planning processes. The Chinese Ministry of Science and Technology (MoST) is responsible for several major S&T programs. The National Basic Research Program (973 Program started in September 1973) supports basic IT research projects at RMB 15 - 50M (there are roughly 6 RMB, also called yuan, per US dollar) each in a continuing series of five-year plans. The current 12th five-year plan (2011 – 2015) in the HPC area focuses on petascale computing.

The National High-Tech Research and Development Program (officially designated as the 863 Program) started in August 1986 and focuses on the application of cutting-edge technologies in key areas in China's National Long-term Scientific and Technological Development Plan (2006-2020). The aim is to strengthen the independent innovation capacity of China in strategic high-tech fields. Applications include fusion, aircraft design, spaceship, drug discovery, animation, structural analysis of large equipment, electromagnetic environment simulation, and new material designs. In 2011, the National Natural Science Foundation of China (NSFC) funded a major research plan for basic algorithm and computational modeling, emphasizing important national demands and promoting hardware-software co-design.

CONCLUSION

At the moment, leadership in supercomputing is being contested among three major players, the United States, China, and Japan, all of whom have held the top spot on the TOP500 list during the last three years. For all three, government funding plays an important role. We have concentrated in this article mainly on the amazing progress in HPC enabling hardware in China but anticipate that, in the future, there is likely to be more importance placed on software, applications, and co-design of hardware in concert with application software in China, Japan, and the United States. Specialized computation chips, interconnect, memory systems, and software will likely define the new era stimulated by constraints on power, speed, and the ability to effectively perform parallel computation.

China's achievements in HPC have been remarkable. There is reasonable agreement on the part of respected computer scientists that the Chinese HPC community is rapidly catching up with the West in the hardware arena but continues to be quite a bit behind in software. Chinese industrial planning is meticulous and focused, and includes extensive support for students and young

professionals for education and research, to include supporting science and engineering students in Western postgraduate universities. The Chinese will not have to trace the fifty years that it has taken the pioneering computer science community to reach this point in the US and can be expected to take its place alongside the United States and Japan as a supercomputing hardware power within the next decade.

References

1. <http://www.TOP500.org>
2. <http://thediplomat.com/2012/12/09/the-long-pole-in-the-tent-chinas-military-jet-engines/>
Andrew Erickson and Gabe Collins The “Long Pole in the Tent”: China’s Military Jet Engines
3. See <http://www.graph500.org/> and <http://www.green500.org/>
4. http://www.computerworld.com/s/article/9233891/China_set_to_launch_own_chip_for_P_Cs_servers
5. [http://www.darpa.mil/Our_Work/MTO/Programs/Power_Efficiency_Revolution_for_Embedded_Computing_Technologies_\(PERFECT\).aspx](http://www.darpa.mil/Our_Work/MTO/Programs/Power_Efficiency_Revolution_for_Embedded_Computing_Technologies_(PERFECT).aspx)
6. http://www.hpcwire.com/hpcwire/2012-05-15/nvidia_announces_kepler-based_tesla_gpus.html
7. http://www.theregister.co.uk/2012/05/15/nvidia_kepler_tesla_gpu_revealed/
8. http://www.pcworld.com/article/257792/intel_invokes_phi_to_reach_exascale_computing_by_2018.html
9. http://www.hpcwire.com/hpcwire/2012-11-13/mont-blanc_project_selects_samsung_exynos_5_processor.html
10. Unleashing the High-performance and Low-power of Multi-core DSPs for General-purpose HPC - Francisco D. Igual, Murtaza Ali, Arnon Friedmann, Eric Stotzer, Timothy Wentz, Robert A. van de Geijn; proceedings SC12
11. Codesign Challenges for Exascale Systems: Performance, Power, and Reliability; Darren J. Kerbyson, Abhinav Vishnu, Kevin J. Barker, and Adolfo Hoisie; IEEE Computer p. 37, November 2011. See also: Towards a Greater Impact of HPC Workloads, David Barkai at SC ‘11 Seattle, WA, USA
12. Ling Chen & Lan Xue - Global Production Network and the Upgrading of China's Integrated Circuit Industry, China & World Economy Volume 18, Issue 6, pages 109–126, November-December 2010 (Chinese Academy of Social Sciences)
13. Godson-T: An Efficient Many-Core Processor Exploring Thread-Level Parallelism, Dongrui Fan, Hao Zhang, Da Wang, Xiaochun Ye, Fenglong Song, Guojie Li, Ninghui Sun, Institute of Computing Technology, Chinese Academy of Sciences
14. <http://www.extremetech.com/computing/113909-new-details-surface-on-the-upu-a-next-generation-cpu-architecture>
15. The data for this section is derived from several sources, not always obviously consistent. The dominant source is HPC China 2011, Asian Technology Information Program (ATIP)12.009 (March 2012).

Author Biographies

Joshua Alspector is a Research Staff Member at the Institute for Defense Analyses where his work includes topics in machine learning, secure information sharing, microelectronics, and high-performance computing. Previously, he was at DARPA where he managed programs in machine learning and distributed information systems. He received the PhD in Physics from MIT. He is a fellow of the IEEE. Contact him at jalspect@ida.org.

Corresponding Author: Joshua Alspector
Information Technology and Systems Division
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311
703 845 6979 (office)
703 845 6848 (fax)
jalspect@ida.org

Alfred E. Brenner is a Project Leader in the Information Technology and Systems Division at the Institute for Defense Analyses (IDA), having started his IDA career as the Director of Algorithms and Applications Research at the Supercomputing Research Center, now the Center for Computing Sciences. He received his PhD in Physics from MIT. He is a member of the IEEE and ACM. Contact him at abrenner@ida.org.

Robert F. Leheny is Assistant Director in the Information Science and Technology Division at the Institute for Defense Analyses. Prior to that, he was Deputy Director of the Defense Advanced Research Projects Agency (DARPA). He received the Dr.Eng.Sci. degree from Columbia University. He is a fellow of the IEEE. Contact him at rleheny@ida.org.

James N. Richmann is a Research Staff Member at the Institute for Defense Analyses (IDA), Information Technology and Systems Division. Prior to joining IDA, Jim was a High Performance Computing Professional at Intel. He has an MS in Industrial Engineering from the US Naval Postgraduate School. He is a member of the IEEE and ACM. Contact him at jrichman@ida.org.

Keywords

Supercomputer, high-performance computing, HPC, parallel computation, GPU, MIC, multicore microelectronics, K computer, Tianhe, Godson, Loongson, Shenwei, Intel, NVIDIA, Samsung, Fujitsu

