Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Editorial:

Post-exascale supercomputing: research opportunities abound

Zuo-ning CHEN[1], Jack DONGARRA[2], Zhi-wei XU[‡3]

[1]*Chinese Academy of Engineering, Beijing 100088, China*
[2]*Electrical Engineering and Computer Science Department, University of Tennessee, TN 37996, USA*
[3]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China*
E-mail: chenzuoning@vip.163.com; dongarra@icl.utk.edu; zxu@ict.ac.cn

https://doi.org/10.1631/FITEE.1830000

Exascale supercomputing refers to the scientific research efforts and activities to build and use supercomputers that can perform scientific computing at the speed of exaflops, or 10^{18} floating-point (64-bit) operations per second. Exascale supercomputing is a major milestone in surpassing the state-of-the-art standard of petascale supercomputing, i.e., 10^{15} floating-point operations per second, established a decade ago in 2008.

Exascale scientific computing research is already in full bloom worldwide. The USA leads this research and development direction, with federal government funding starting in as early as 2008. Japan, Europe, India, and China soon followed suit. The most recent focus on the field is in Europe, with 1.4 billion euros budgeted for building pre-exascale supercomputers by 2020, and an additional 2.7 billion euros proposed for building an exascale supercomputer by 2023 (Feldman, 2018). It is expected that multiple exascale supercomputers will become operational in the USA, Europe, and Asia by 2020–2024, supporting cutting edge research in many scientific fields.

In this context, the Chinese Academy of Engineering (CAE) organized a special issue of "Post-exascale Supercomputing" in *Frontiers of Information Technology and Electronic Engineering*, by inviting position papers from leading experts inside and outside China. This special issue targets 2020–2030 supercomputing systems that go beyond the existing exascale systems under construction. It focuses on innovative research ideas in systems architecture, processors, memory, storage, interconnects, operating systems, programming languages and compilers, and application frameworks. It foresees the convergence of high-performance computing (HPC) with big data computing, intelligence computing (e.g., deep learning), cloud computing, and edge computing, and encourages consideration of future HPC workloads.

1 Outlook of post-exascale supercomputing

Three trends and five challenges can be identified for HPC in the coming decade. The three main trends are: (1) the growth of the supercomputing community, (2) the expansion of usage modes and workloads, and (3) growing richness of diverse and innovative solutions.

Community growth: Twenty-five years ago, when the supercomputing Top500 list (www.top500.org) was first established, supercomputing was a field populated mainly by scientific computing and systems experts in developed economies, mostly in the USA, Europe, and Japan. The knowledge and technology arising from these communities has since trickled down to other countries, and the community of interest has grown to include countries with developing economies such as India and China (Xu et al., 2016). This trend is likely to continue. Looking back

‡ Corresponding author
© Zhejiang University and Springer-Verlag GmbH Germany, part of Springer Nature 2018

from 2030, we may see that a significant development in supercomputing took place in India and China, two of the most populous developing countries, as they became active participants in the supercomputing community, as their addition not only helps enlarge the user base and address the digital divide, but also brings in new application requirements and fresh developers.

Expanding workloads: This trend is discussed in detail in a community roadmap study paper (Asch et al., 2018). The on-going work in this area has been conducted by scientists from the USA, Europe, Japan, and China. Expanding workloads manifest in two ways. First, supercomputers are processing new types of workloads such as big data, machine learning tasks, and streaming, in addition to traditional large-scale numerical simulation jobs. Second, new usage modes are making their way into supercomputing, such that future supercomputing is likely to be a convergence of on-the-premises supercomputing, cloud computing, big data analytics, and edge computing. This convergence is driven by the very nature of scientific inquiry. Any work of scientific inquiry needs three types of reasoning: deduction, induction, and abduction, which in turn need appropriate capacity for processing and data handling both at the supercomputer center and on the site (edge).

Diverse solutions: With a larger community, richer workloads and usage modes, much more parallelism, and higher demand for energy efficiency, research on post-exascale supercomputers is likely to offer diverse solutions. This diversity manifests not only in processor, memory/storage, and interconnect technology, but also in developments at various layers of the software stack. Domain-specific architectures may appear in the supercomputing field. Such rich diversity encourages innovation and offers more powerful toolkits for scientific users, but may lead to fragmentation that will hinder developers and users. The rapid growth of supercomputing in the past 25 years benefited from a relatively stable supercomputing ecosystem featuring the cluster architecture supported by open-source software. The question of how to gracefully embrace this emerging diversity while maintaining the architectural stability of the supercomputing ecosystem both poses the main challenge and offers the largest opportunity for post-exascale supercomputing.

Looking more closely at the challenge aspect, the five major challenges are: (1) alleviating the energy efficiency bottleneck, (2) achieving order-of-magnitude better devices and components, (3) creating novel systems architectures, (4) effective co-design of software and hardware, and (5) establishing an ecosystem for diverse applications.

Energy efficiency bottleneck: A central challenge is energy efficiency. We need to consider the whole stack, from using new material and devices, 3D structures, and architectures that effectively integrate multiple types of domain-specific accelerators, to energy-aware systems software and application algorithms.

Currently, the number 1 system in the Top500 list is the IBM Summit supercomputer installed at the Oak Ridge National Laboratory, which achieves 122.3 petaflops at a power budget of 8.8 MW, translating to an energy efficiency of 13.9 giga 64-bit floating-point operations per second per watt, or 13.9 gigaflops/W. The most energy-efficient system is the Shoubou system B, at 18 gigaflops/W (Strohmaier, 2018).

How are we to address the energy efficiency bottleneck for post-exascale supercomputing? How much further can we go? Fig. 1 illustrates the improvement potential.

Improvement potential is plotted on two dimensions in Fig. 1. The first dimension is concerned with improvement potential over time. For instance, the US Department of Energy's exascale research program sets a goal of 1 exaflops at 20–40 MW, or 25–50 gigaflops/W, probably around year 2022. The US DARPA's JUMP program sets a more ambitious long-term goal of 3 peta operations per second per watt, or 3 peta operations per joule (POPJ), possibly by around 2035. Here an operation is not necessarily a 64-bit IEEE floating-point operation. In Physics, there is a hard limit proposed by Rolf Landauer (Landauer, 1961; Jun et al., 2014), which states roughly that at room temperature, a one-bit irreversible operation (such as erasing a bit) needs to consume 3×10^{-21} J energy, translating to an energy efficiency of about 0.33×10^{21} operations per joule, or 0.33 zeta operations per joule (ZOPJ). Currently, there is still a gap (and thus improvement space) of more than six orders of magnitude before we reach Landauer's limit.

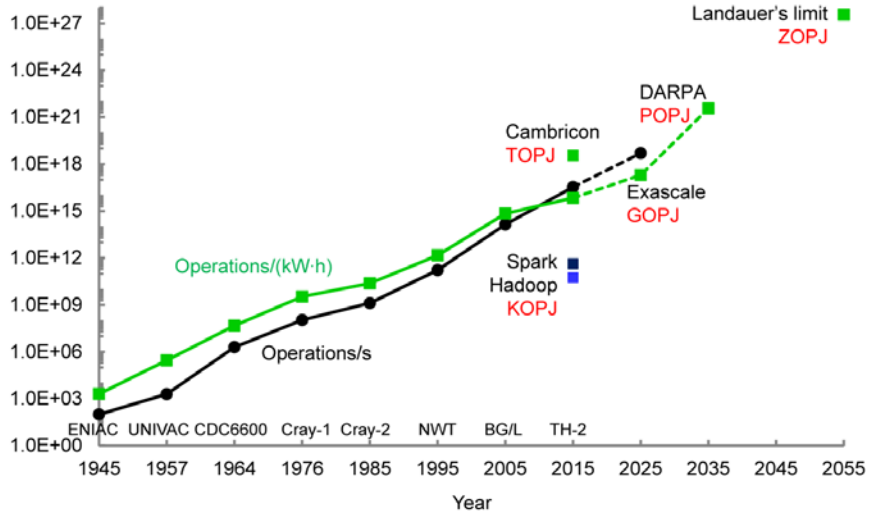

Fig. 1 The energy efficiency improvement potentials for post-exascale supercomputing

Another dimension has to do with the co-design of more efficient hardware-software stack of supercomputers, taking into consideration the convergence of numerical computing, big data, and machine learning. For instance, on current clusters, big data software such as Hadoop and Spark has good scalability, but shows very low energy efficiency of only tens or hundreds of kilo operations per joule (KOPJ). This is orders of magnitude lower than MPI-based parallel programs, which reach GOPJ. On the other hand, a co-designed processor called Cambricon-1A, which was developed at the Institute of Computing Technology of the Chinese Academy of Sciences in 2015 and targets machine learning on small terminals such as smartphones, reached over 1 tera operations per joule (TOPJ). Similar co-design was implemented in the Cambricon MLU-100 machine learning processor for cloud severs (Cambricon, 2018).

Order-of-magnitude better devices: Emerging technologies, such as 3D-stacking, fully optical communication, magnetic semiconductors, and memristors, are challenging mature technologies used in today's supercomputers which are based on CMOS processors, DRAM memory, and hard disks. The question of how to effectively use these new technologies to obtain order-of-magnitude performance improvement will be a challenge for supercomputer designers. For instance, with fully optical communication, we may need to rethink the systems architecture. 3D-stacking may lead to new chips integrating CPUs, accelerators, memory, and high-speed interconnect. Non-volatile memory technology may create new storage systems merging memory and storage.

Novel systems architectures: Systems architecture has played a critical role in the history of modern supercomputing. Architectural innovations, from vector supercomputers, SMP, ccNUMA, and MPP, to clusters, enabled exponential growth of performance and scalability. The architectures in current supercomputers are mainly clusters augmented with heterogeneous accelerators, which have achieved drastic performance gains in certain applications.

But - what next? Novel architectures need to be proposed for post-exascale supercomputers, considering new technology and new applications. It may even be necessary to consider non-von Neuman architectures that are effective and practical.

Effective co-design of software and hardware: Currently, there is a wide gap between the peak performance and the sustained performance that real applications can achieve, especially with new applications with sparsity and irregularity, such as data analytics and complex multi-modal applications. Co-design has been proposed and practiced to bridge this gap. However, with the broadening of applications, the co-design approach to post-exascale supercomputers faces more severe challenges. We need to study systems that seamlessly match software and hardware, as we did with dense scientific computing such as Linpack over MPI over clusters. This time, however, we need to consider a larger scope of

applications, over new software stacks, using new systems architectures and hardware.

Ecosystem for diverse applications: The complexity, heterogeneity, and massive parallelism of modern supercomputers pose numerous difficulties for application development. This problem has been partially overcome by a vibrant supercomputing ecosystem. However, the existing ecosystem has a tradition of scientific and engineering computing, which is not enough for the new diverse applications that converge numeric simulation, big data, and artificial intelligence. We need to build up a new supercomputing ecosystem for application development, which supports the mixed or converged workloads of arithmetic-intensive, data-intensive, and intelligent applications.

2 Scanning the special issue

Eight research teams inside and outside China contributed to this special issue. The topics covered include emerging semiconductor technologies, processors, interconnects, systems architecture, programming, application frameworks, and large-scale applications.

To address the energy efficiency challenge, the US DARPA together with the Semiconductor Research Corporation launched a Joint University Microelectronics Program (JUMP), with the long-term goal of achieving peta operations per second per watt. Hu and Niemier (2018) presented insights into the potential achievement of this ambitious goal. Their main viewpoint is encouraging: the POPJ goal looks feasible, but only with cross-layer design efforts spanning devices, circuits, architectures, and algorithms. They presented concrete benchmarking studies of neural network accelerators for embedded applications, on both CMOS and beyond-CMOS devices, as supporting evidence, and showed that an energy improvement of over 220 times can be achieved with only 0.5% accuracy degradation. They also discussed a model to aid the proposed cross-layer design process.

Xie and Jia (2018) focused on the topic of post-exascale processor architecture, based on evaluating existing processors, including insights gained in designing the SW26010 processor which is used in the Sunway TaihuLight supercomputer. They identified three goals: effective performance scaling, efficient resource utilization, and adaptation to diverse applications. They proposed the Massa architecture, a Many-core processor architecture with scalar processing and application-specific acceleration. An unusual observation is that single instruction multiple data (SIMD) or vector processing hardware present in current processors is difficult to use, and should be eliminated. Instead, application-specific acceleration hardware should be used.

Panda et al. (2018) gave a whole-stack view on the networking and communication challenges for post-exascale systems. They envisioned a post-exascale system as a terabit network connected system, having many dense nodes with a network-centric architecture. The authors then summarized the challenges from multiple angles, including the perspectives of the underlying technology, schemes and protocols, and programming models. The authors foresaw that networking and communication need to support the convergence of HPC, big data, and deep learning workloads, and stressed that co-designing runtime with upper layers is critical for maximum performance and scalability.

Liao et al. (2018) suggested building a zetaflops supercomputer by 2035, based on technology trends and their experiences developing the Tianhe series of supercomputers. The zetaflops system should have a computing speed of 10^{21} 64-bit floating-point operations per second and consume 100 MW power, achieving energy efficiency of 10 TOPJ. In the consideration of this ambitious goal, the authors discussed six major challenges, including manufacturing process limits, power consumption, interconnects and communication, memory and storage, reliability, and programming. They also reviewed evolutionary and revolutionary technologies that help address these challenges.

Sun et al. (2018) offered the viewpoint that high-throughput computing is on the rise and future cloud supercomputing will see more such systems and applications. Future high-throughput computing systems will need to simultaneously satisfy three objectives: high throughput, high systems utilization, and low latency. These can be achieved only when the cloud supercomputing system can significantly reduce the system's performance entropy, or unwanted

disorder and uncertainty. The authors discussed two novel technology ideas towards efficient high-throughput computing, specifically on-chip dataflow architecture and labeled von Neumann architecture.

Mo (2018) addressed an important issue: how to turn a supercomputer's peak performance to real application performance. He focused on application software frameworks for extreme-scale numerical simulation, by discussing three types of bottlenecks and coping strategies of parallel computing that affect computational scale, computing efficiency, and programming productivity. The author also highlighted a five-level quantitative approach for assessing weaknesses in the computing capability, called "supply-side technology."

Zhai and Chen (2018) presented a vision for the important issue of post-exascale programming, based on their experience of programming large-scale systems. Post-exascale supercomputers face three challenges: they are more heterogeneous, need to exploit a much larger amount of parallelism, and require a higher degree of fault tolerance. New programming models, programming frameworks, and domain-specific languages will be needed to address these challenges.

Yang and Fu (2018) presented challenges and research opportunities of post-exascale application software, based on their experiences of developing and optimizing application software on the Sunway TaihuLight supercomputer. They identified three challenges of parallelization over millions of cores, memory wall, and application software migration. They discussed how they overcame these problems to develop applications that won the ACM's prestigious Gordon Bell prize. They also discussed three possible trends: programming efforts shifting from computation to data, precision optimization in mixed-precision supercomputer systems, and programming hardware instead of software.

References

Asch M, Moore T, Badia R, et al., 2018. Big data and extreme-scale computing: pathways to convergence—toward a shaping strategy for a future software and data ecosystem for scientific inquiry. *Int J High Perform Comput Appl*, 32(4):435-479. https://doi.org/10.1177/1094342018778123

Cambricon, 2018. MLU100-Cambricon. https://en.wikichip.org/wiki/cambricon/mlu/mlu100 [Accessed on Oct. 15, 2018].

Feldman M, 2018. Europeans Budget 1.4 Billion Euros to Build Next-Generation Supercomputers. https://www.top500.org/news/europeans-budget-14-billion-euros-to-build-next-generation-supercomputers/ [Accessed on Oct. 15, 2018].

Hu XS, Niemier M, 2018. Cross-layer efforts for energy-efficient computing: towards peta operations per second per watt. *Front Inform Technol Electron Eng*, 19(10):1209-1223. https://doi.org/10.1631/FITEE.1800466

Jun Y, Gavrilov M, Bechhoefer J, 2014. High-precision test of Landauer's principle in a feedback trap. *Phys Rev Lett*, 113(19):190601. https://doi.org/10.1103/PhysRevLett.113.190601

Landauer R, 1961. Irreversibility and heat generation in the computing process. *IBM J Res Devel*, 5(3):183-191.

Liao XK, Lu K, Yang CQ, et al., 2018. Moving from exascale to zettascale computing: challenges and techniques. *Front Inform Technol Electron Eng*, 19(10):1236-1244. https://doi.org/10.1631/FITEE.1800494

Mo ZY, 2018. Extreme-scale parallel computing: bottlenecks and strategies. *Front Inform Technol Electron Eng*, 19(10):1251-1260. https://doi.org/10.1631/FITEE.1800421

Panda DK, Lu XY, Subramoni H, 2018. Networking and communication challenges for post-exascale systems. *Front Inform Technol Electron Eng*, 19(10):1230-1235. https://doi.org/10.1631/FITEE.1800631

Strohmaier E, 2018. Highlights of the 51st TOP500 list. https://www.top500.org/static/media/uploads/top500_ppt_201806.pdf [Accessed on Oct. 15, 2018].

Sun NH, Bao YG, Fan DR, 2018. The rise of high-throughput computing. *Front Inform Technol Electron Eng*, 19(10):1245-1250. https://doi.org/10.1631/FITEE.1800501

Xie XH, Jia X, 2018. Exploring high-performance processor architecture beyond the exascale. *Front Inform Technol Electron Eng*, 19(10):1224-1229. https://doi.org/10.1631/FITEE.1800424

Xu Z, Chi X, Xiao N, 2016. High-performance computing environment: a review of twenty years of experiments in China. *Natl Sci Rev*, 3(1):36-48. http://dx.doi.org/10.1093/nsr/nww001

Yang GW, Fu HH, 2018. Application software beyond exascale: challenges and possible trends. *Front Inform Technol Electron Eng*, 19(10):1267-1272. https://doi.org/10.1631/FITEE.1800459

Zhai JD, Chen WG, 2018. A vision of post-exascale programming. *Front Inform Technol Electron Eng*, 19(10):1261-1266. https://doi.org/10.1631/FITEE.1800442

Zuo-ning CHEN is a vice president of the Chinese Academy of Engineering.

Jack DONGARRA is a professor of the University of Tennessee and a member of the US National Academy of Engineering.

Zhi-wei XU is a professor and the CTO of the Institute of Computing Technology, at the Chinese Academy of Sciences.