Exascale Computing and Big Data: The Next Frontier

DANIEL A. REED
University of Iowa
and
JACK DONGARRA
University of Tennessee

For scientific and engineering computing, exascale (10^{18} operations per second) is the next proxy in the long trajectory of exponential performance increases that has continued for over half a century. Similarly, large-scale data preservation and sustainability within and across disciplines, metadata creation and multidisciplinary fusion, and digital privacy and security define the frontiers of big data. Solving the myriad technical, political and economic challenges of exascale computing will require coordinated planning across government, industry and academia, commitment to data sharing and sustainability, collaborative research and development, and recognition that both international competition and collaboration will be necessary.

Categories and Subject Descriptors: C [**Computer System Organization**]: Hardware/Software Interfaces; C.1.2 **(Multiple Data Stream Architectures (Multiprocessors))** – Parallel Processors

Daniel A. Reed acknowledges support from the National Science Foundation under NSF grant ACI-1349521. Jack Dongarra acknowledges support from the National Science Foundation under NSF grant ACI-1339822 and by the Department of Energy under DOE grant DE-FG02-13ER26151Authors' addresses: Daniel Reed, (Current address) Department of Computer Science, University of Iowa, Iowa City, IA 52242 and Jack Dongarra (Current address) Innovative Computing Laboratory, University of Tennessee, Knoxville, TN 37996.

General Terms: Performance, Reliability, Economics

Additional Key Words and Phrases: Exascale computing, big data, exascale computing, computational science and engineering, economics

1. INTRODUCTION

Nearly two centuries ago, the English chemist, Humphrey Davy remarked,

> *Nothing tends so much to the advancement of knowledge as the application of a new instrument. The native intellectual powers of men in different times are not so much the causes of the different success of their labors, as the peculiar nature of the means and artificial resources in their possession.*

Davy's observation that advantage accrues to those who have the most powerful scientific tools is no less true today. Last year, Karplus, Levett and Warshel received the Nobel Prize in chemistry for their work in computational modeling. As the Nobel committee noted, "computer models mirroring real life have become crucial for most advances made in chemistry today" and "computers unveil chemical processes, such as a catalyst's purification of exhaust fumes or the photosynthesis in green leaves."

Whether describing the advantages of high-energy particle accelerators such as the Large Hadron Collider (LHC) and the recent discovery of the Higgs boson; powerful astronomy instruments such as the Hubble Space Telescope and the Planck all-sky survey and insights into the universe's expansion and dark energy, or high throughput DNA sequencers and exploration of metagenomics ecology, ever more powerful scientific instruments continually advance knowledge. Each of these scientific instruments and a host of others are critically dependent on computing for sensor control, data processing and international collaboration and access.

However, computing is much more than simply an augmenter of science. Unlike other tools, which are limited to particular scientific domains, computational modeling and data analytics are applicable to all areas of science and engineering, for they breathe life into the underlying mathematics of scientific models, and they allow researchers to understand nuanced predictions and to shape experiments more efficiently. They also help capture and analyze the torrent of experimental data being produced by a new generation of scientific instruments and sensors, themselves made possible by advances in computing and microelectronics.

Computational modeling can illuminate the subtleties of complex mathematical models and advance science and engineering where time, cost or safety precludes experimental assessment alone. Computational models of astrophysical phenomena, on temporal and spatial scales as diverse as planetary system formation, stellar dynamics, black hole behavior, galactic formation, and the interplay of baryonic and putative dark matter have provided new insights into theories and complemented experimental data. Increasingly sophisticated climate models, which capture the effects of greenhouse gases, deforestation and other planetary changes, have been key to understanding the effects of human behavior on the weather and climate change.

Computational science and engineering also enables multidisciplinary design and optimization, reducing prototyping time and costs. Advanced simulation has already allowed Cummins to build better diesel engines faster and less expensively, Goodyear to design safer tires much more quickly, Boeing to build more fuel-efficient aircraft, and Procter & Gamble to create better materials for home products.

Similarly, "big data," machine learning and predictive data analytics have been hailed as the fourth paradigm of science [5], allowing researchers to extract insights from both scientific instruments and computational simulations. Machine learning has yielded new insights into health risks and the spread of disease via analysis of social networks, web search queries and hospital data. It has also

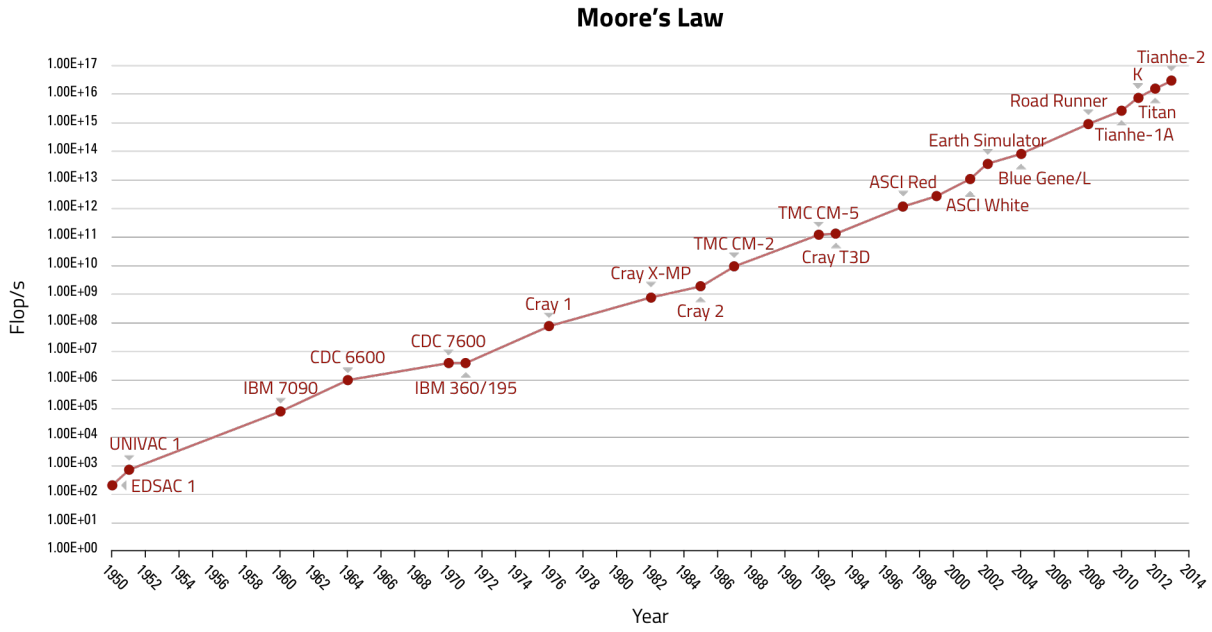

Fig.1 Advanced computing performance measured by the high-performance Linpack (HPL) benchmark

been key to event identification and correlation in domains as diverse as high-energy physics and molecular biology.

As with the successive generations of other large-scale scientific instruments, each new generation of advanced computing brings new capabilities and insights, along with technical design challenges and economic tradeoffs. High-performance computers and big data systems are tied inextricably to the broader computing ecosystem, its designs and its markets. They are also coupled to national security needs and economic competitiveness in ways that distinguish them from most other scientific instruments.

This "dual use" model, together with the rising cost of ever-larger computing and data analysis systems, and a host of new design challenges at massive scale, are raising new questions about advanced computing research investment priorities, design and procurement models, and global collaboration and competition. This paper examines some of these technical challenges, the interdependence of computational modeling and data analytics, and the global ecosystem and competition for leadership in advanced computing. We begin with a primer on the history of advanced computing.

2. ADVANCED SCIENTIFIC COMPUTING AND THE CHALLENGES OF SCALE

By definition, an advanced computing system embodies the hardware, software and algorithms needed to deliver the very highest capability at any given time. In the 1980s, vector supercomputing dominated high-performance computing, as embodied in the eponymously named systems designed by the late Seymour Cray. The 1990s saw the rise of massively parallel processing (MPPs) and shared memory multiprocessors (SMPs), built by Thinking Machines, Silicon Graphics and others. In turn, clusters of commodity (Intel/AMD x86) and purpose-built processors (e.g., IBM's BlueGene), dominated the previous decade. Today, those clusters have been augmented with computational accelerators and GPUs.

Similarly, just a few years ago, the very largest data storage systems contained only a few terabytes of secondary disk storage, backed by automated tape libraries. Today, commercial and research cloud computing systems each contain many petabytes of secondary storage, and individual research laboratories routinely process terabytes of data produced by their own scientific instruments.

2.1 The Leading Edge

Given the rapid pace of technological change, leading edge capability is a moving target – today's smartphone was yesterday's supercomputer, and a personal digital music collection was once enterprise scale storage. Lest this seem an exaggeration, the measured performance of an Apple iPhone5 or Samsung Galaxy S4 on standard linear algebra benchmarks now substantially exceeds that of a Cray-1, which was widely viewed as the first successful supercomputer. That same smartphone has a storage capacity rivaling the text-based content of a major research library.

Just a few years ago, teraflops (10^{12} floating point operations/second) and terabytes (10^{12} bytes of secondary storage) defined state-of-the-art advanced computing. Today, those same values represent a desk side PC with an NVidia or Intel Xeon Phi accelerator and local storage. In 2014, advanced computing is now defined by multiple petaflops (1015 floating operations/second) supercomputing systems and cloud data centers with many petabytes of secondary storage.

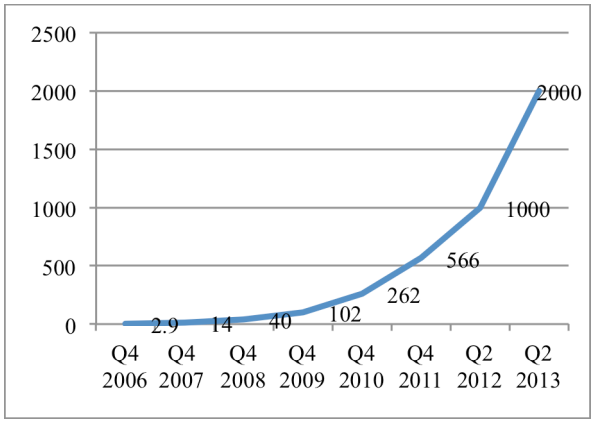

Fig. 2 Growth of Amazon S3 Objects (Billions)

Figure 1 shows this exponential increase in advanced computing capability, based on the widely used High-Performance Linpack (HPL) benchmark [20] and the Top500 list of the world's fastest computers [28]. Although solution of dense linear systems of equations is no longer the best measure of delivered performance on complex scientific and engineering applications, this historical data illustrates how rapidly high-performance computing has evolved. Though high-performance computing has benefited from the same semiconductor advances as commodity computing, sustained system performance has risen even more rapidly, due to the increasing system size and parallelism.

The growth of personal, business, government and scientific data has been even more dramatic, with commercial cloud providers building worldwide networks of data centers, each costing hundreds of millions of dollars, and the volume of scientific data produced annually now challenging the budgets of national research agencies. As an example, Figure 2 shows the exponential growth in the number of objects stored in Amazon's Simple Storage Service (S3).

There are natural technical and economic synergies among the challenges facing data-intensive science and exascale computing, and advances in both are necessary for future scientific breakthroughs. Data-intensive science relies on the collection, analysis and management of massive volumes of data, whether obtained from scientific simulations or experimental facilities. In both cases, national and international investments in "extreme scale" systems will be necessary to analyze the massive volumes of data that are now commonplace in science and engineering.

2.2 The Race to the Future

For scientific and engineering computing, exascale (10^{18} operations per second) is the next proxy in the long trajectory of exponential performance increases that has continued for over half a century. Similarly, large-scale data preservation and sustainability within and across disciplines, metadata creation and multidisciplinary fusion, and digital privacy and security define the frontiers of big data. This multifaceted definition encompasses more than simply quantitative measures of sustained arithmetic operation rates or storage capacity and analysis rates; it is also a relative term encompassing qualitative improvements in the usable capabilities of advanced computing systems at all scales. As such, it is intended to suggest a new frontier of practical, delivered capability to scientific and engineering researchers across all disciplines.

Historically, high-performance computing advances have been largely dependent on concurrent advances in algorithms, software, architecture and hardware that enable higher levels of floating point performance for computational models. Today, advances are also shaped by data analysis pipelines, data architectures, and machine learning that manage large volumes of scientific and engineering data.

However, just as changes in scientific instrumentation scale bring new opportunities, they also bring new challenges, some technical, but others organizational, cultural and economic, and these challenges are not self-similar across scales. Today, exascale computing systems cannot be produced in an affordable and reliable way (i.e., subject to realistic engineering constraints on capital and operating costs, usability, and reliability), and as the costs of advanced computing and data analysis systems have moved from millions to billions of dollars, design and decision processes have necessarily become more complex and fraught with controversy. This is a familiar lesson to those in high-energy physics and astronomy, where particle accelerators and telescopes have become planetary scale instrumentation and the province of international consortia and global politics. Advanced computing is no exception.

The research and development costs to create an exascale computing system have been estimated to exceed $1B, with an annual operating cost of tens of millions of U.S. dollars. Concurrently, there is growing recognition that governments and research agencies have substantially underinvested in data retention and management, as evinced by multi-billion dollar private sector investments in big data and cloud computing. Against this backdrop, U.S. support for basic research is at a decadal low, when adjusted for inflation [6], and both the U.S. and the European Union continue experience weak recoveries from the economic downturn of 2008.

Further exacerbating the challenges, the global race for advanced computing hegemony is convolved with national security desires, economic competitiveness and the future of the mainstream computing ecosystem. The European Union, Japan and China have all launched next-generation computing system research and development projects [4, 10], in competition with the United States. The shift from personal computers to mobile devices and the end of Dennard scaling [16] have also further raised competition between the U.S.-dominated x86 architectural ecosystem and the globally-licensed ARM ecosystem. Concurrently, concerns about national sovereignty, data security and Internet governance have triggered new competition and political concerns around data services and cloud computing operations.

Despite all of these challenges, there is reason for cautious optimism. Every advance in computing technology has driven industry innovations and economic growth, spanning the entire spectrum of computing, from the emerging Internet of Things through ubiquitous mobile devices to the world's most powerful computing systems and largest data archives. These advances have also spurred basic and applied research in every domain of science.

Solving the myriad technical, political and economic challenges will be neither easy nor even possible by tackling them in isolation. Rather, it will require coordinated planning across government, industry and academia, commitment to data sharing and sustainability, collaborative research and development, and recognition that both competition and collaboration will be necessary for success. Nor can the future of big data and analytics be pitted against exascale computing; both are critical to the future of advanced computing and scientific discovery.

3. Scientific and Engineering Opportunities

Researchers in the physical sciences and engineering have long been major users of advanced computing and computational models. The more recent adoption by the biological, environmental and social sciences has been driven in part the rise of big data analytics. In addition, advanced computing is now widely used in engineering and advanced manufacturing. From understanding the subtleties of airflow in turbomachinery through chemical molecular dynamics for consumer products to biomass feedstock modeling for fuel cells, advanced computing has become synonymous with multidisciplinary design and optimization and advanced manufacturing.

Looking forward, just a few examples illustrate the deep and diverse scientific and engineering benefits from advanced computing:

- Biology and biomedicine have been transformed by access to large volumes of genetic data. Inexpensive, high throughput genetic sequencers have enabled capture of organism DNA sequences and have made possible genome-wide association studies (GWAS) for human disease and human microbiome investigations, as well as metagenomics environmental studies. More generally, biological and biomedical challenges span sequence annotation and comparison, protein structure prediction; molecular simulations and protein machines; metabolic pathways and regulatory networks; whole cell models and organs; and organisms, environments and ecologies.
- High Energy Physics (HEP) is both computational and data-intensive. First principles computational models of quantum chromodynamics (QCD) provide numerical estimates and validations of the Standard Model. Similarly, particle detectors require the measurement of probabilities of "interesting" events in large numbers of observations (e.g., in 10^{16} or more particle collisions observed in a year). The Large Hadron Collider (LHC) and its experiments necessitated creation of a worldwide computing grid for data sharing and reduction, driving deployment of advanced networks and protocols, as well as a hierarchy of data repositories. All of these were necessary to identify the long-sought Higgs boson.
- Climate science is also critically dependent on the availability of a reliable infrastructure for managing and accessing large, heterogeneous quantities of data on a global scale. It is inherently a collaborative and multidisciplinary effort that requires sophisticated modeling of the physical processes and exchange mechanisms among multiple Earth realms (atmosphere, land, ocean, and sea ice) and comparison and validation of these simulations with observational data from various sources, all collected over long periods.
- Cosmology and astrophysics are now critically dependent on advanced computational models to understand stellar structure, planetary formation, galactic evolution and other interactions. These models combine fluid processes, radiation transfer, Newtonian gravity, nuclear physics and general relativity (among others). Underlying these is a rich set of computational techniques based on adaptive mesh refinement and particle in cell (PIC), multipole and Monte Carlo methods and smoothed particle hydrodynamics. Complementing computation, whole sky surveys and a new generation of telescopes now routinely generate terabytes of data each day, with data reduction, machine learning and statistical comparisons now essential elements of observational astronomy.
- Experimental and computational materials science is key to understanding materials properties and engineering options. For example, neutron scattering allows researchers to understand the structure and properties of materials, macromolecular and biological systems, and the fundamental physics of the neutron by providing data on the internal structure of materials from the atomic scale (atomic positions and excitations) up to the mesoscale (e.g., the effects of stress).

There are two common themes across these science and engineering challenges. The first is an extremely wide range of temporal and spatial scales and complex, nonlinear interactions across multiple biological and physical processes. These are the most demanding of computational simulations, requiring collaborative research teams and the very largest and most capable computing systems. In each case, the goal is predictive simulation – gleaning insights that test theories, identify subtle interactions and guide new research.

The second theme is the enormous scale and diversity of scientific data, and the unprecedented opportunities for data assimilation, multidisciplinary correlation and statistical analysis. Whether in the biological or physical sciences, engineering or business, big data is creating new research needs and opportunities.

4. TECHNICAL CHALLENGES IN ADVANCED COMPUTING

The scientific and engineering opportunities made possible by advanced computing are deep, but the technical challenges in designing, constructing and operating advanced computing systems of unprecedented scale are just as daunting. To deliver exascale computing and big data analytics to the scientific and engineering communities, many challenges must be overcome.

In a series of recent studies, the U.S. Department of Energy identified ten research challenges (DOE, 2010; Geist & Lucas, 2009; Lucas et al., 2014) in developing a new generation of advanced computing systems. These include:

Energy efficient circuit, power and cooling technologies. With current semiconductor technologies, all proposed exascale designs would consume hundreds of megawatts of power. New designs and technologies are needed to reduce this to a more manageable and economically feasible level (e.g., 20-40 MW, comparable to that used by commercial cloud data centers).

High performance interconnect technologies. In the exascale regime, the cost to move a datum will exceed the cost of a floating point operation, necessitating very energy efficient, low latency, high bandwidth interconnects for fine-grained data exchanges among hundreds of thousands of processors. Even with such designs, locality-aware algorithms and software will be needed to maximize performance and reduce energy needs.

Advanced memory technologies to improve capacity and bandwidth. Minimizing data movement and minimizing energy use are also dependent on new memory technologies, including processor-in-memory, stacked memory, non-volatile memory approaches. Algorithmic determinants of memory capacity will be a significant driver of overall system cost, as the memory per node will necessarily be smaller than in current designs.

Scalable system software that is power and resilience aware. Traditional high-performance computing software has been predicated on the assumption that failures are infrequent; at very large scale,

systemic resilience in the face of regular component failures will be essential. Similarly, dynamic, adaptive energy management must become an integral part of system software, for both economic and technical reasons.

Data management software that can handle the volume, velocity and diversity of data. Whether computationally generated or captured from scientific instruments, efficient *in situ* data analysis will require restructuring of scientific workflows and applications, as well as new techniques for data coordinating and mining. Without these, I/O bottlenecks will limit system utility and applicability.

Programming environments to express massive parallelism, data locality, and resilience. The widely used communicating sequential process (CSP) model (i.e., MPI programming) places the burden of locality and parallelization on application developers. Exascale systems will have billion-way parallelism and frequent faults. More expressive programming models are needed that can deal with this behavior and simplify the developer's efforts.

Reformulation of science problems and refactoring solution algorithms. Many thousands of person-years have been invested in current scientific and engineering codes. Adapting them to billion-way parallelism will require redesigning, or even reinventing, the algorithms, and potentially reformulating the science problems. Understanding how to do these things efficiently and effectively will be key to solving mission-critical science problems at exascale.

Ensuring correctness in face of faults, reproducibility, and algorithm verification. With frequent transient and permanent faults, lack of reproducibility in collective communication, and new mathematical algorithms with limited verification, computation validation and correctness assurance rise dramatically in importance for the next generation of massively parallel systems.

Mathematical optimization and uncertainty quantification for discovery, design, and decision. Large-scale computations are themselves experiments that probe the sample space of numerical models. Understanding the sensitivity of computational predictions to model inputs and assumptions, particularly when involving complex, multidisciplinary applications is dependent on new tools and techniques for application validation and assessment.

Software engineering and supporting structures to enable scientific productivity. Although programming tools, compilers, debuggers, and performance enhancement tools shape research productivity for all computing systems, at exascale, application design and management for reliable, efficient and correct computation is especially daunting. Unless researcher productivity increases, the time to solution may be dominated by application development, not computation.

Similar hardware and software studies (Amarasinghe et al., 2009; Kogge et al., 2008) chartered by the U.S. Defense Advanced Research Projects Agency (DARPA) identified the following challenges, most similar to those cited by the Department of Energy studies:

- Energy efficient operation to achieve desired computation rates subject to overall power dissipation
- Primary and secondary memory capacity and access rates, subject to power constraints
- Concurrency and locality to meet performance targets, while allowing some threads to stall during long latency operations
- Resilience, given large component counts, shrinking silicon feature sizes, low power operation and transient and permanent component failures
- Application scaling subject to memory capacity and communication latency constraints
- Expressing and managing parallelism and locality in system software and portable programming models, including runtime systems, schedulers and libraries
- Software tools for performance tuning, correctness assessment and energy management

Finally, a recent study by the U.S. National Academy of Science (NAS) (Fuller & Millett, 2011) focused on the technology challenges created by the end of Dennard scaling (Dennard et al.) – the ability to shrink transistors while also reducing voltage and current – and the implications for programming models and software. Simply put, the NAS study suggests that, barring a breakthrough, the exponential increases in performance brought by shrinking semiconductor feature sizes and architectural innovations are nearing an end.

These studies suggest that computing technology is poised at important inflection points, both at the very largest scale (i.e., leading edge high-performance computing) and at the very smallest scale (i.e., semiconductor processes). On this, the computing community remains divided, with strong believers that technical obstacles limiting extension of current approaches will be overcome, and others who believe, more radical technology and design approaches, (e.g., quantum or superconducting devices), may be required.

4.1 Hardware and Architecture Challenges

Although a complete description the hardware and software technical challenges just enumerated is beyond the scope of this survey, review of a selected subset is useful to illuminate the depth and breadth of the problems and their implications for the future of both advanced computing and the broader deployment of next-generation consumer and business computing technologies.

4.1.1. Post-Dennard Scaling

Over the past thirty years, Moore's "law" has held true due to the hard work and creativity of a great many people, as well as many billions of dollars of investment in process technology and silicon foundries. It has also rested on the principle of Dennard scaling (Dennard et al.; Kamil, Shalf, & Strohmaier, 2008), which showed that as transistors got smaller the power density remained constant. Thus, decreasing a transistor's linear size by a factor of two, reduced the power by a factor of four (i.e., with both voltage and current halving).

Although transistor sizes continue to decline, with 22 nanometer feature sizes now common, transistor power consumption no longer decreases accordingly. This has led to limits on chip clock rates and power consumption, along with design of multicore chips and the rise of dark silicon – chips with more transistors than can be active simultaneously due to thermal and power constraints (Esmaeilzadeh, Blem, Amant, Sankaralingam, & Burger, 2011).

These semiconductor challenges have been mirrored by an increasingly empty bag of architectural tricks. Most of the techniques once found in supercomputer processor designs – superpipelining, scoreboarding, vectorization and parallelization – are now in mainstream microprocessors. These power constraints have also driven design of function-specific accelerators (e.g., graphics processing units (GPUs) and accelerators such as NVidia's Tesla) and heterogeneous cores that balance power consumption and performance in different ways.

In this new world, hardware/software co-design becomes *de rigeur*, with devices and software systems interdependent. The implications are far fewer general-purpose performance increases, more hardware diversity, elevation of multivariate optimization – power, performance, reliability – in programming models, and new system software resource management challenges.

4.1.2. Resilience and Energy Efficiency at Scale

As advanced computing systems grow ever larger, the assumption of reliable hardware and software also becomes much less credible. Although the mean time before failure (MTBF) for individual components continues to increase incrementally, the large overall component count means the systems themselves will fail more frequently. This has been confirmed by analysis of high-performance computing failure modes (Schroeder & Gibson, 2006).

Moreover, new data from commercial cloud data centers suggests that some long-held assumptions about component failures and lifetimes are incorrect (Gill, Jain, & Nagappan, 2011; Pinheiro, Weber, & Barroso, 2007; Schroeder & Gibson, 2007; Schroeder, Pinheiro, & Weber, 2009). A Google study (Schroeder et al., 2009) showed that DRAM error rates were orders of magnitude higher than previously reported, with more than eight percent of DIMMs affected by errors in a year. Equally surprisingly, these were hard errors, rather than soft, correctable (via ECC) errors.

In addition to resilience, scale also brings new challenges in energy management and thermal dissipation. Today's advanced computing systems consume megawatts of power, and cooling capability and peak power loads both limit where many systems can be placed geographically. As commercial cloud operators have learned, energy infrastructure and power are a substantial fraction of total system cost at scale, necessitating new infrastructure approaches and operating models. These have included low power designs, new cooling approaches, energy accountability and operational efficiencies.

The energy constraints for exascale systems also have profound algorithmic and software implications. Because the energy requirements for DRAM dominate a large-scale system's energy budget, new designs are likely to be "memory starved" relative to current systems. When combined with a reversal of traditional models, where communication is viewed as free relative to computation, a radical reengineering of algorithms and architectures will be essential.

4.2 Software and Algorithmic Challenges

Many of the software and algorithmic challenges for advanced computing are themselves consequences of extreme system scale. Consequently, advanced computing shares many of the scaling problems of web and cloud services, but differs in its price-performance optimization balance, emphasizing high levels of performance, whether for computation or data analysis. This distinction is central to the design choices and optimization criteria.

4.2.1. Locality and Scale

As noted earlier, putative designs for extreme scale systems are projected to require billion-way computational concurrency, with aggressive parallelism at all system levels. Thus, maintaining load balance on all levels of a hierarchy of algorithms and platforms will be the key to an efficient execution. This will likely require dynamic, adaptive run-time mechanisms (Datta et al., 2008) and self-aware resource allocation to tolerate not only algorithmic imbalances but also variability in hardware performance and reliability.

In turn, the energy costs and latencies for communication will place an even greater premium on computation locality than today. Inverting long held models, arithmetic operations will be far less energy intensive and more efficient than communications. This will challenge traditional algorithmic design approaches and comparative optimizations, making redundant computation sometimes preferable to data sharing and elevating communication complexity to parity with computation. It will also necessitate models and methods to minimize and tolerate (hide) latency, optimize data motion and remove global synchronization.

4.2.2. Adaptive System Software

Resource management for today's high-performance computing systems remains rooted in a *deus ex machina* model, with coordinated scheduling and tightly synchronized communication. However, extreme scale, hardware heterogeneity, system power and heat dissipation constraints and increased component failure rates not only influence the design and implementation of applications, they also affect the design of system software in areas as diverse as energy management and I/O. Similarly, as the volume of scientific data grows, it is unclear the traditional file abstractions and parallel file systems used by technical computing will scale to trans-petascale data analysis.

Instead, new system software and operating system designs will need to support management of heterogeneous resources and non-cache-coherent memory hierarchies, provide applications and runtime with more control of task scheduling policies, and manage global namespaces. They will also need to expose mechanisms for finer measurement, prediction and control of power management, allowing schedulers to map computations to function-specific accelerators and manage thermal envelopes and application energy profiles.

4.2.3. Parallel Programming Support

As the diversity, complexity and scale of advanced computing hardware has increased, the complexity and difficulty in developing applications has grown equally rapidly, with many operating functions now subsumed by applications. This has been further exacerbated by the increasingly multidisciplinary nature of applications, which combine algorithms and models that span a wide range of spatio-temporal scales and algorithmic approaches.

Consider the typical single program multiple data (SPMD) parallel programming model, where application data is partitioned and distributed across the individual memories of the computation nodes, and the nodes share data via the message passing interface (MPI). In turn, the application code on each node manages the local, multilevel computation hierarchy – typically multiple, multithreaded, possibly heterogeneous cores and (often) a GPU accelerator – and coordinates I/O, manages application checkpointing, and oversees power budgets and thermal dissipation. This daunting level of complexity and detailed configuration and tuning makes developing robust applications an arcane art accessible to only a dedicated few.

Ideally, future software design, development and deployment would be done with performance and correctness in mind at the outset rather than *ex situ*. Beyond more performance-aware design and development of applications based on integrated performance and correctness models, these tools need to be integrated with compilers and runtime systems, they need to provide support for heterogeneous

hardware and mixed programming models, and they must provide more intelligence in raw data processing and analysis.

4.2.4. Algorithmic and Mathematics Challenges

Given the scale and expected error rates of exascale systems, design and implementation of algorithms must be rethought from first principles. This includes exploration of global synchronization-free (or at least minimal) algorithms, fault oblivious and error tolerant algorithms, architecture-aware algorithms suitable for heterogeneous and hierarchical organized hardware, support for mixed arithmetic, and software for energy-efficient computing.

Moving forward to exascale will put heavier demands on algorithms in at least two areas: the need for increasing amounts of data locality to perform computations efficiently, and the need to obtain much higher factors of fine-grained parallelism as high-end systems support increasing numbers of compute threads. As a consequence, parallel algorithms must adapt to this environment, and new algorithms and implementations must be developed to extract the computational capabilities of the new hardware.

Significant model development, algorithm re-design and science application reimplementation, supported by (an) exascale-appropriate programming model(s), will be required to exploit the power of exascale architectures. The transition from current sub-petascale and petascale computing to exascale computing will be at least as disruptive as the transition from vector to parallel computing in the 1990's.

5. Economic and Political Challenges

The technical challenges of advanced computing are shaped and constrained by other elements of the broader computing landscape. In particular, powerful smartphones and cloud computing services are rapidly displacing the PC and local servers as the computing standard. This shift has also triggered international competition for industrial and business advantage, with countries and regions investing in new technologies and system deployments.

5.1 Computing Ecosystem Shifts

The Internet and web services revolution is now global and U.S. influence, though substantial, is being diluted. Notwithstanding Apple's phenomenal success, most smartphones and tablets are now designed, built and purchased globally, and the annual sales volume of smartphones and tablets already exceeds that of PCs and servers.

This ongoing shift in consumer preferences and markets is accompanied by another technology shift. Smartphones and tablets are based on energy-efficient microprocessors (a key component of proposed exascale computing designs) and systems-on-a-chip (SoCs) using the ARM architecture. Unlike Intel and AMD, which design and manufacture the x86 chips found in today's PCs and most leading edge servers and HPC systems, ARM does not manufacture its own chips. Instead, it licenses the design to others, who incorporate the architecture into custom SoCs that are manufactured by global semiconductor foundries such as Taiwan's TSMC.

5.2 International Exascale Projects

The international competition surrounding advanced computing mixes concerns about economic competitiveness, shifting technology ecosystems (e.g., ARM and x86), business and technical computing (e.g., cloud computing services and data centers), and scientific and engineering research. The European Union, Japan, China and the United States have each launched exascale computing projects, each with differing emphases on hardware technologies, system software, algorithms and applications.

5.2.1. European Union

The European Union (EU) announced the start of its exascale research program in October 2011 with 25 million Euros in funding for three complementary research projects in the EU's Framework 7 effort. The CRESTA, DEEP, and Mont-Blanc projects will each investigate different exascale challenges using a co-design model spanning hardware, system software, and software applications. This initiative represents Europe's first sustained investment in exascale research.

CRESTA brings together four European high-performance computing centers: Edinburgh Parallel Computing Centre (project lead), the High Performance Computing Center Stuttgart, Finland's IT Center for Science Ltd, and Partner Development Center Sweden, as well as the Dresden University of Technology which will lend expertise in performance optimization. In addition, the CRESTA team also consists of application professionals from European science and industry, as well as HPC vendors—including HPC tool developer Allinea and HPC vendor Cray. CRESTA focuses on the use of applications as co-design drivers for software development environments, algorithms and libraries, user tools, and underpinning and crosscutting technologies.

The Mont-Blanc project, led by the Barcelona Supercomputing Center, brings together European technology providers ARM, Bull, Gnodal, and major supercomputing organizations involved with the PRACE project (Juelich, LRZ, GENCI, and CINECA). The project intends to deploy a first generation HPC system built from energy-efficient embedded technologies, and will conduct the research necessary to achieve exascale performance with energy-efficient designs.

The Dynamical Exascale Entry Platform (DEEP) ,led by Forschungszentrum Juelich, seeks to develop an exascale-enabling platform and optimization of a set of grand challenge codes. The DEEP system is based on a commodity cluster and accelerator design (Cluster Booster Architecture) as a proof-of-concept for a 100 petaflop/s PRACE production system. In addition to the lead partner, Juelich, the project partners include Intel, ParTec, **L**eibniz-Rechenzentrum (LRZ), Universität Heidelberg, German Research School for Simulation Sciences, Eurotech, Barcelona Supercomputing Center, Mellanox, EPFL, Katholieke Universiteit Leuven**,** CERFACS, the Cyprus Institute, Universität Regensburg**,** CINCA, and CGGVeritas**.**

5.2.2. Japan

In December 2013, the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) selected RIKEN to develop and deploy an exascale system by 2020.. RIKEN was selected based on its experience developing and operating the K computer, which at 10 petaflop/s, was ranked as the fastest supercomputer in the world in 2011. Estimated to cost 140 billion yen ($1.38B), the exascale system design will be based on a combination of general purpose processors and accelerators, and involves three key Japanese computer vendors (Fujitsu, Hitachi, and NEC), as well as technical support from the University of Tokyo, University of Tsukuba, Tokyo Institute of Technology, Tohoku University, and RIKEN.

5.2.3. China

Today, China's Tianhe-2 system is the world's fastest supercomputer. It contains 16,000 nodes, each with two Intel Xeon processors and three Intel Xeon Phi coprocessors. The system also contains a proprietary high-speed interconnect, called TH Express-2, which was designed by the National University for Defense Technology (NUDT). NUDT, conducts research on processors, compilers, parallel algorithms, and systems. Based on this work, China is expected to produce two 100-petaflop/s systems as early as 2015, one of which will be built entirely from Chinese-made chips and interconnects. Tianhe-2 will also be upgraded from a peak of 55 petaflop/s to 100 petaflop/s in 2015, and a second system based on China's ShenWei processor will be deployed.

5.2.4. United States

Historically, the U.S. Networking and Information Technology Research and Development program (NITRD) has spanned several research missions and agencies, with primary leadership by the Department of Energy (DOE), the Department of Defense (DoD) and the National Science Foundation (NSF). Today, DOE is the most active deplorer of high-performance computing systems and in developing plans for exascale computing. In contrast, NSF and DoD have focused more on broad cyberinfrastructure and enabling technologies research. Although planning continues, the U.S. has not yet mounted an advanced computing initiative similar to those in Europe and Japan.

5.3 International Collaboration

Although the global competition for advanced computing leadership continues, there is active international collaboration. The International Exascale Software Project (IESP) is one such example. With seed funding from governments in the United States, the European Union and Japan, as well as supplemental contributions from industry stakeholders, IESP was formed to empower ultra-high resolution and data-intensive science and engineering research through the year 2020.

In a series of meetings, the international IESP team developed a plan for a common, high-quality computational environment for petascale/exascale systems. The associated roadmap for software development would take the community from its current position to exascale computing (Dongarra et al., 2011).

6. CONCLUSIONS

Computing is at a profound inflection point, economically and technically. The end of Dennard scaling and its implications for continuing semiconductor design advances, the shift to mobile and cloud computing, the explosive growth of scientific, business, government and consumer data and the opportunities for data analytics and machine learning, and the continuing need for more powerful computing systems to advance science and engineering all form the backdrop for the debate over the future of exascale computing and big data analysis. However, certain things are clear.

- High-end data analytics (big data) and high-end computing (exascale) are both essential elements of an integrated computing research and development agenda; neither can be sacrificed or minimized to advance the other.
- Research and development of next-generation algorithms, software and applications is as crucial as investments in semiconductor devices and hardware, and we have historically underinvested in these areas relative to hardware.
- The global information technology ecosystem is in flux, with the transition to a new generation of low power, mobile devices, cloud services, and rich data analytics.
- Both private sector competition and global research collaboration will be necessary to address design, test and deploy exascale class computing and data analysis capabilities.

There are both great opportunities and great challenges in advanced computing. Scientific discovery via computational science and data analytics truly is the "endless frontier" of which Vannevar Bush spoke so eloquently in 1945.The challenges are for us to sustain the research, development and deployment of the high-performance computing infrastructure needed to enable those discoveries.

ACKNOWLEDGMENTS

We are grateful insights and perspectives from the DARPA and DOE exascale hardware, software and application study groups.

REFERENCES

Amarasinghe, S., Campbell, D., Carlson, W., Chien, A., Dally, W., Elnohazy, E., . . . Sterling, T. (2009). Exascale Software Study: Software Challenges in Extreme Scale Systems: Defense Advanced Research Projects Agency (DARPA).

Datta, K., Murphy, M., Volkov, V., Williams, S., Carter, J., Oliker, L., . . . Yelick, K. (2008). Stencil Computation Optimization and Auto-Tuning on State-of-the-Art Multicore Architectures. *Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, 1-12.

Dennard, R. H., Gaensslen, F. H., Yu, H.-n., Rideout, V. L., Bassous, E., & LeBlanc, A. R. Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions. *IEEE Journal of Solid State Circuits, SC-9*(5), 256-268.

DOE. (2010). The Opportunities and Challenges of Exascale Computing: Office of Science, U.S. Department of Energy.

Dongarra, J., Beckman, P., Moore, T., Aerts, P., Aloisio, G., Andre, J.-C., . . . Yelick, K. (2011). The International Exascale Software Project Roadmap. *International Journal of High Performance Computing Applications, 25*(1), 3-60.

Esmaeilzadeh, H., Blem, E., Amant, R. S., Sankaralingam, K., & Burger, D. (2011). Dark Silicon and the End of Multicore Scaling. *Proceedings of the 38th Annual International Symposium on Computer Architecture*, 365-376.

Fuller, S. H., & Millett, L. I. (2011). Computing Performance: Game Over or Next Level? *Computer, 44*(1), 31-38.

Geist, A., & Lucas, R. (2009). Major Computer Science Challenges at Exascale. *International Journal of High Performance Applications, 23*(4), 427-436.

Gill, P., Jain, N., & Nagappan, N. (2011). Understanding Network Failures in Data Centers: Measurement, Analysis, and Implications. *Proceedings of the ACM SIGCOMM 2011, 41*(4), 350-361

Kamil, S., Shalf, J., & Strohmaier, E. (2008). Power Efficiency in High Performance Computing. *High-Performance, Power-Aware Computing (HPPAC 2008)*.

Kogge, P., Bergman, K., Borkar, S., Campbell, D., Carlson, W., Dally, W., . . . Yelick, K. (2008). Exascale Computing Study: Technology Challenges in Achieving Exascale Systems: U.S. Defense Advanced Research Projects Agency (DARPA).

Lucas, R., Ang, J., Bergman, K., Borkar, S., Carlson, W., Carrington, L., . . . Stevens, R. (2014). *Top10 Exascale Research Challenges*. Department of Energy Office of Science.

Pinheiro, E., Weber, W.-D., & Barroso, L. A. (2007). Failure Trends in a Large Disk Drive Population. *5th USENIX Conference on File and Storage Technologies (FAST'07)*.

Schroeder, B., & Gibson, G. A. (2006). A Large-Scale Study of Failures in High-Performance Computing Systems. *Proceedings of the International Conference on Dependable Systems and Networks*, 249-258. doi: 10.1109/dsn.2006.5

Schroeder, B., & Gibson, G. A. (2007). Understanding Disk Failure Rates: What Does an MTTF of 1,000,000 Hours Mean to You? *ACM Transactions on Storage, 3*(3), 8.

Schroeder, B., Pinheiro, E., & Weber, W.-D. (2009). DRAM Errors in the Wild: A Large-Scale Field Study. *Proceedings of the Eleventh International Joint Conference on Measurement and Modeling of Computer Systems, 37*(1), 193-204.