

UNIVERSITY OF WESTMINSTER



WestminsterResearch

<http://www.wmin.ac.uk/westminsterresearch>

Extreme-scale computing—where 'just more of the same' does not work.

**Adolfy Hoisie
Vladimir Getov**

School of Electronics and Computer Science

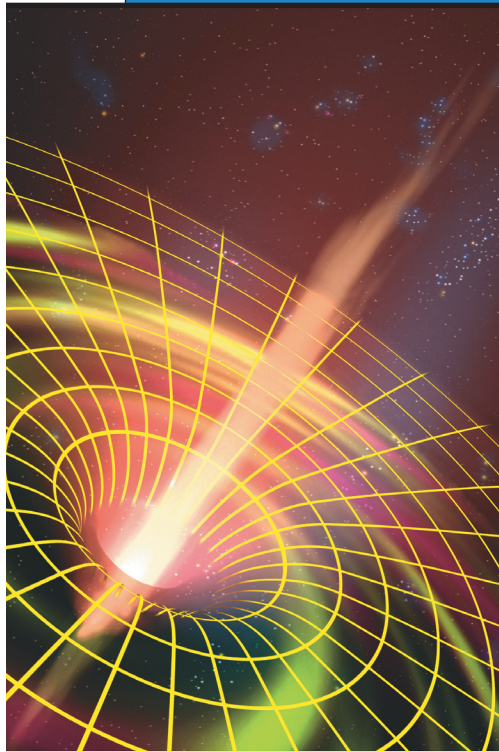
Copyright © [2009] IEEE. Reprinted from Computer, 42 (11). pp. 24-26. ISSN 0018-9162.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Westminster's products or services. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners. Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of the University of Westminster Eprints (<http://www.wmin.ac.uk/westminsterresearch>).

In case of abuse or copyright appearing without permission e-mail wattsn@wmin.ac.uk.



EXTREME-SCALE COMPUTING— WHERE ‘JUST MORE OF THE SAME’ DOES NOT WORK

Adolfy Hoisie, *Los Alamos National Laboratory*
Vladimir Getov, *University of Westminster*

In addition to enabling science through simulations at unprecedented size and fidelity, extreme-scale computing serves as an incubator of scientific and technological ideas for the computing area in general.

The leading edge of high-performance computing (HPC), an area of considerable growth and pace of progress, extreme-scale computing relates directly to the hardware, software, and applications enabling simulations in the petascale performance range and beyond. Moreover, extreme-scale computing acts as a scientific and technological driver for computing in general. In addition to enabling science through simulations at unprecedented size and fidelity, extreme-scale computing serves as an incubator of scientific and technological ideas for the computing area. As such, its rapid development significantly impacts several neighboring areas such as loosely coupled distributed systems, grid infrastructures, cloud computing, and sensor networks.

The complexity of computing at extreme scales is increasing rapidly, now matching the complexity of the

simulations running on them. Therefore, the quest for higher processing speed has become only one of many challenges when designing novel high-end computer systems. This complexity arises from the interplay of various factors such as level of parallelism (systems in this range currently use hundreds of thousands of processing elements and are envisioned to reach millions of threads of parallelism), availability of parallelism in algorithms, design and implementation of system software, deep memory hierarchies, heterogeneity, reliability and resilience, and power consumption, just to name a few.

IT'S ALL ABOUT SCALABILITY

Achieving high levels of sustained performance in applications is a dauntingly challenging task. To respond to this never-ending demand for higher and higher performance, extreme-scale computing incorporates in a single topic area several research and development challenges related to scalability. The questions that have been attracting attention from the professional community at large include the following:

- Are there limits to manageable levels of parallelism? Are millions of threads tractable? What are the programming models that support application de-

velopment within reasonable levels of effort, while allowing high performance and efficiency?

- Is there a limit to the number of cores that can be used for building a single computer? What is the significance of heterogeneity and hybrid designs in this respect?
- Are there fundamental limits to an increasing footprint of the interconnect? What are the performance/reliability tradeoffs?
- What are the factors that hinder high levels of sustained performance? What are the best ways to assess, model, and predict performance in extreme-scale regimes?
- What are the system software challenges, limitations, and opportunities? Can we develop system software that harnesses heterogeneity and asynchronous designs?
- What are design considerations for the I/O and storage subsystems given the vast amounts of data generated by such simulations?
- What are the main characteristics and challenges in providing high-level quality of service by current and future extreme-scale systems? Given the size and complexity of the systems enabling extreme-scale computing, can we overcome the intrinsic limitations in reliability and resilience?
- Is it inevitable that extreme-scale supercomputers will be delivered together with an associated power plant? Can we reduce as much as possible the power consumption to save energy for a greener planet but also enable the design of even faster computers?


IN THIS ISSUE

In this special issue, we explore some of the salient aspects of extreme-scale computing. The selected articles cover a significant cross-section of the questions listed above.

In "Architectures for Extreme-Scale Computing," Josep Torrellas outlines the main architectural challenges of extreme-scale computing and describes potential paths forward to ensure the same fast pace of progress that this area sustained in the past decade. Key technologies such as near-threshold voltage operation, nonsilicon memories, photonics, 3D die stacking, and per-core efficient voltage and frequency management will be key to energy and power efficiency. Efficient, scalable synchronization and communication primitives, together with support for the creation, commit, and migration of lightweight tasks will enable fine-grained concurrency. A hierarchical machine organization, coupled with processing-in-memory will enhance locality. Resiliency will be addressed with a combination of techniques at different levels of the computing stack. Finally, programming the machine with a high-level

data-parallel model and using an intelligent compiler to map the code to the hardware will ensure programmability and performance. Finally, the author outlines Thrifty, a novel extreme-scale architecture.

In "Tofu: A 6D Mesh/Torus Interconnect for Exascale Computers," Yuichiro Ajima, Shinji Sumimoto, and Toshiyuki Shimizu describe their recently developed high-speed interconnect architecture for next-generation supercomputers that operate beyond 25 petaflops. The first such system, which will be one of the world's largest supercomputers, is scheduled to begin operation in 2011. The network topology of Tofu is a fault-tolerant 6D mesh/torus, and each link has 10 Gbytes of bidirectional bandwidth. Each of the computation nodes employs four communi-



Extreme-scale computing incorporates in a single topic area several research and development challenges related to scalability.

cation engines with an integrated collective function. The Tofu interconnect is designed to run a 3D torus application even if there are some faulty nodes inside the system's submesh. A user can specify a 3D Cartesian space for a job, and the system allocates nodes to parallel processes of the job and ensures that a neighboring node of the application's Cartesian space is also a neighbor in the physical 6D space. Since there are several combinations of physical coordinates for folding application coordinates, the system can provide a suitable submesh shape from the available free nodes, which greatly improves system utilization. Additionally, system availability has been further improved by using a newly developed graceful degradation technique that allows a 3D Cartesian space to become available within a faulty 6D submesh.

As supercomputing applications and architectures grow more complex, researchers need methodologies and tools to understand and reason about system performance and design. "Using Performance Modeling to Design Large-Scale Systems" by a team of authors from the Los Alamos National Laboratory, New Mexico, is dedicated to this important topic area. Existing petascale systems contain sufficient hardware complexity to make it impossible for application developers, hardware designers, and system buyers to have an intuitive "feeling" for those factors that have a bearing on performance; as we march toward exascale systems this problem will only get worse. In this article, the authors present a proven, highly accurate quasi-analytical performance modeling methodology that puts performance analysis tools in the hands of applications

and systems researchers. As a case in point, the article demonstrates how performance modeling can accurately predict application performance on IBM's Blue Gene/P system, one of today's largest parallel machines, for three large-scale applications in application domains including shock hydrodynamics, deterministic particle transport, and plasma fusion modeling. Using this system as a baseline, a performance look-ahead is shown for the near-term future, theorizing how these applications will perform on potential future systems incorporating improved compute and interconnection network performance.

In "Parallel Scripting for Applications at the Petascale and Beyond," Michael Wilde and colleagues characterize the applications that can benefit from extreme-scale scripting, discuss the technical obstacles that such applications raise for the system and application architect, and present results achieved with parallel script execution on the extreme-scale computers available today. They show examples of the science that can be achieved with this approach, the scale that extreme machines make possible, the performance of applications at these scales, the systems and architectural challenges that were overcome to make this feasible, and the challenges and opportunities that remain. The article concludes by exploring the relationships—and promising connections—between parallel scripting and traditional memory.

In "Energy-Efficient Computing for Extreme-Scale Science," David Donofrio and colleagues describe the Green Flash project, which aims to deliver an order-of-magnitude increase in efficiency, both computationally and in cost-effectiveness. The main idea is based on offering a many-core processor design with novel alternatives to

cache coherence that enable far more efficient interprocessor communication than a conventional symmetric multiprocessing approach coupled with autotuning technologies to improve kernels' computational efficiency. Application-driven HPC design represents the next transformational change for the industry and will be enabled by leveraging existing embedded ASIC design methods, autotuning for code optimization, and emerging hardware emulation environments for performance evaluation. Looking beyond climate models, the Green Flash approach could allow future exaflops-class systems to be defined by science rather than have the science artificially constrained by generic machine characteristics.

In June 2008, the world entered the petaflops era with the Roadrunner supercomputer installation at Los Alamos. It is widely anticipated that systems with millions of threads, capable of achieving tens of petaflops, will be in existence in just a couple of years. Exascale computing is now within reach.

Development in this area attracts support from funding agencies all around the globe, including the US, Asia (Japan, China, and India, most notably), Europe, and Australia. The main reasons for this are the strategically important application domains and the incubator role that this field has for computing in general. Extreme-scale computing, and HPC in general, is an exciting and fast-developing area with sizable contributions coming from different professional categories, including research and development, industry, education, and end users.

We hope you will enjoy reading the articles in this special issue. 

Adolfy Hoisie is the leader of the Center for Advanced Architectures and Usable Supercomputing (CAAUS) and of the Computer Science for High-Performance Computing group at the Los Alamos National Laboratory. His research interests are performance analysis and modeling of large-scale systems and applications, system architecture, and extreme-scale computing in general. He is a past recipient of the Gordon Bell award and of other awards for research excellence. Contact him at hoisie@lanl.gov.

Vladimir Getov is a professor of distributed and high-performance computing at the University of Westminster, London. His research interests include parallel architectures and performance, autonomous distributed computing, and high-performance programming environments. He received a PhD and DSc in computer science from the Bulgarian Academy of Sciences. He is a member of the IEEE and the ACM and is Computer's area editor for high-performance computing. Contact him at v.s.getov@westminster.ac.uk.



IEEE Annals of the History of Computing

From the analytical engine to the supercomputer, from Pascal to von Neumann—the *IEEE Annals of the History of Computing* covers the breadth of computer history. The quarterly publication is an active center for the collection and dissemination of information on historical projects and organizations, oral history activities, and international conferences.

www.computer.org/annals



Selected CS articles and columns are available for free at <http://ComputingNow.computer.org>