EXASCALE COMPUTING CHALLENGES
The issues researchers will encounter on the path to exascale HPC are equally critical for all large-scale computing architectures and facilities, not just the largest ones or only those related to scientific computing. Workloads may differ, but energy challenges are common. Because power is the overriding hardware concern, energy efficiency will be essential across all computing scales. Furthermore, energy issues will affect all levels of the computing system, including processors, interconnects, algorithms, software, and programming models.

Given the complexity of the increasingly daunting constraint space under consideration, successful optimization requires a new tack, a new approach, and a new set of design methodologies. For example, given the overwhelming performance and energy cost of data movement, efficiency requires minimization of data movement—a task for all layers of the stack, from the hardware to the application software.

Similarly, optimization of the performance/power/reliability triad mandates rethinking of algorithms, programming models, and hardware in concert and requires an unprecedented level of collaboration and cooperation in hardware, system architecture, system software, and application codesign. This requires a completely new approach based on concurrent development and engineering in an integrated manner to a set of consistent overall design metrics, employing accurate, quantitative design methodologies.

THE CODESIGN APPROACH—BACKGROUND
For embedded systems, codesign traditionally has meant partitioning concepts in the design process to produce systems meeting stringent performance, verifi-
The key concept is meeting system-level objectives by exploiting tradeoffs between hardware and software through an integrated concurrent design process.

Codesign in embedded systems came about in large part because a variety of factors led to the use of software in systems that had previously been entirely hardware-based. This increased the complexity of that software in microcontrollers, digital signal processors, and even general-purpose processors. Other factors included the decreasing cost of microcontrollers, rapidly increasing numbers of available transistors, the availability of advanced emulation technology, and the improved efficiency of higher-level language compilers for use in embedded systems. A key motivation was the need to support the growing complexity of embedded systems, which has an obvious parallel in exascale computing.

Embedded systems are characterized by running only a few applications that are completely known at design time, not being programmable by end users, and having fixed runtime requirements—meaning that additional dynamic computing power is not useful. Codesign considerations for such systems include cost, power consumption, predictability, and meeting time bounds.

In contrast, general-purpose computing systems are characterized by running a broad class of applications, being programmable by end users, and having the characteristic that faster is always better, which requires including cost and peak speed in their design criteria.

The essence of the codesign challenge for HPC and exascale systems is to use the key design criteria of embedded systems—cost and power consumption—while creating systems that are useful and effective over the broad range of applications needed to advance science. “One-off” exascale systems will suggest failure.

CODESIGN FOR HPC SYSTEMS

In the HPC arena, codesign has also been used recently, and therefore it is not entirely new to exascale computing. Both the IBM BlueGene/L supercomputer ([IBM J. Research and Development, vol. 49, no. 2/3, 2005]) and IBM's PERCS project for DARPA's High-Productivity Computing Systems (HPCS) program have adopted the codesign approach. Two additional excellent examples of codesigned special-purpose supercomputers for molecular dynamics applications are the MDGrape system and the Anton supercomputer built by D.E. Shaw Research.

The IBM RoadRunner pointed the way toward the use of hybrid architectures by its inclusion of coprocessing elements along with general-purpose processors to accelerate a specific workload. The heterogeneity of the resultant architecture, which required a mixture of several programming models, posed significant challenges to ensure the utility of the coprocessor approach for designing HPC systems that are to be truly effective for a wide range of scientific applications. Still, the metric for success in these codesign examples was performance, without regard to power and reliability.

The HPC community currently finds itself needing to apply codesign methods on the path to exascale systems and applications. Therefore, key concepts that apply include:

- employing a high level of abstraction to describe the system;
- using models to allow analysis and exploration of the system architecture, validate assumptions regarding the architecture, explore the design implementation performance parameters, and verify that tradeoffs made using high-level system models were worthwhile; and
- creating codesign methodologies and tools that designers can use to “tinker” with the platform, adding, subtracting, or changing parameters to determine the effect on the architecture and system performance.

A novel exascale concept is related to the necessity of rethinking the application software itself, including optimizing the algorithms and the codes for minimizing data movement for energy efficiency or for implementing resilience mechanisms. Hence, these virtual testbeds need to support initial optimization by both system and appli-
Architectural simulation

“Rethinking Hardware-Software Codesign for Exascale Systems” by John Shalf, Dan Quinlan, and Curtis Janssen describes a set of high-accuracy simulation tools that researchers can employ for low-level hardware and architecture codesign for a simplified application workload.

High-speed interconnects

In “Codesign for InfiniBand Clusters,” Sreeram Potluri and coauthors discuss a codesign approach that takes advanced features from the commodity InfiniBand network, incorporates the design into a state-of-the-art message-passing interface communication library, and then modifies applications to leverage these new features.

High-end systems

“Codesign Challenges for Exascale Systems: Performance, Power, and Reliability” by Darren J. Kerbyson and colleagues describes a comprehensive codesign methodology that uses analytical modeling to achieve maximum performance, power, and reliability for full systems and applications.

In its simplest definition, codesign is about anticipating and changing the future.11 Early intervention in hardware designs, optimizing what is important, influencing the design, redesigning algorithms and system software, devising languages and programming models that reflect abstract machine models, writing code generators and auto-tuners, and modeling all of the above are the essence of the craft. Successfully meeting these challenges is essential for continued progress in computing performance.

References


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 is a member of IEEE and ACM, a Fellow of the BCS, and Computer’s area editor for high-performance computing. Contact him at v.s.getov@westminster.ac.uk.

Adolfy Hoisie is a Laboratory Fellow and director of the Center for Advanced Architectures at the Pacific Northwest National Laboratory. His research focuses on performance analysis and modeling of systems and applications, areas in which he has published extensively. Hoisie is a member of IEEE. Contact him at adolfy.hoisie@pnnl.gov.

Harvey J. Wasserman is a member of the User Services Group at the National Energy Research Scientific Computing Center, the primary computing center for the US Department of Energy’s Office of Science, located at Lawrence Berkeley National Laboratory. Wasserman’s research focuses on workload characterization, benchmarking, and system evaluation. Contact him at hjwasserman@lbl.gov.

For videos related to this topic, see the following:
• “ASCR Discovery—Codesigning for Exascale,” www.youtube.com/watch?v=gXvh7WezxEg
• “Reconfigurable Exascale Computing,” www.youtube.com/watch?v=zTr4bepr5Xc