The IBM Blue Gene project

IBM Blue Gene team

This paper provides a short overview of the IBM Blue Gene® project and an introduction to all of the papers in this issue of the IBM Journal of Research and Development.

Introduction and overview

Late in 1999, IBM announced an effort to build a petaFLOPS (PFLOPS; 10^{15} floating-point operations per second) scale supercomputer, with the dual objective of advancing the state of the art in biomolecular simulations, as well as the state of the art in computer design and software for extremely large-scale systems. In this context, the IBM Blue Gene* platform was created at the IBM Thomas J. Watson Research Center. From the start, IBM Research partnered with Columbia University and the University of Edinburgh to use the design philosophy and experience of the QCDSP and QCDOC supercomputers [1], which exhibited superior performance on lattice quantum chromodynamics (QCD) applications using large numbers of low-power processors. IBM next partnered with Lawrence Livermore National Laboratory (LLNL) and later with Argonne National Laboratory (ANL). The former emphasized applications related to national security, and the latter served the science community that needed access to massively parallel machines in general. By involving partners at the beginning, IBM was able to optimize hardware and software design parameters around actual application kernels, carefully chosen to represent a large class of interesting problems. This collaborative approach greatly contributed to the success of the Blue Gene family.

Since its inception, the Blue Gene project has been aimed at expanding IBM's high-performance computing (HPC) platform to unprecedented levels of scale and performance. Of the first generation, the IBM Blue Gene*/L supercomputer [2], in which the largest installation consists of 104 racks (106,496 nodes) at LLNL, rated at a peak performance of 596 teraFLOPS. At the time of its introduction, this was more than an order of magnitude increase in both core count and performance over its competitors. As it expanded over 2004 to 2007, the Blue Gene/L installation at LLNL won the Top500** race [3] seven times consecutively.

Corresponding author: George Chiu, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA; gchiu@us.ibm.com.

Digital Object Identifier: 10.1147/JRD.2012.2220487

The second generation in the Blue Gene series was the IBM Blue Gene*/P supercomputer [4], introduced in 2007. The largest Blue Gene*/P installation consists of 72 racks with 73,728 nodes at the Forschungszentrum Juelich in Germany. It became the first 1-PFLOPS system in Europe, fulfilling the goal set in 1999.

The third generation is the IBM Blue Gene*/Q supercomputer. The largest installations are the Sequoia installation at LLNL, with 96 racks, or 98,304 nodes, rated at a peak performance of 20 PFLOPS, and the Mira system at ANL with 48 racks (49,152 nodes, 10 PFLOPS peak).

This double issue of the *IBM Journal of Research and Development* is dedicated to the Blue Gene/Q project. **Table 1** highlights the key parameters of the three generations of Blue Gene. The substantial increase in node performance and network bandwidth of Blue Gene/Q over previous generations is due to radical new approaches in node and network architecture.

The first paper in this issue, by the IBM Blue Gene team, is an overview of the Blue Gene/Q Compute chip, including its A2 processor core, quad double-precision floating-point unit, cache hierarchy, DDR (double data rate) memory controllers, and chip-to-chip communication logic, all integrated on the same compute chip. The cache hierarchy consists of the instruction and data Level 1 (L1) cache, L1 prefetcher (L1P), and Level 2 (L2) cache. The L1P incorporates a new list prefetcher, jointly developed with the Columbia/Edinburgh team, designed to prefetch repetitive addressing patterns effectively. This foundational paper establishes the chip architectural hardware features.

As shown in Table 1, the core count per chip progressed from 2 to 4 to 16 in three generations, and the corresponding frequency increased from 0.7 GHz to 0.85 GHz to 1.6 GHz while improving the power efficiency as measured in gigaFLOPS per watt. A key differentiator of the Blue Gene family is its low-power design. In order to maximize the performance of a rack in a thermally limited regime, the Blue Gene team chose to use the most power-efficient processor cores. Because a mere 1 W increase per core will result in a 1 MW increase in power for a one-million core system,

0018-8646/13/$5.00 © 2013 IBM

Table 1 Parameters for three generations of the IBM Blue Gene family. Note that the core count and node count in the table represent the maximum counts associated with the largest existing system in each of the three generations; these are not the scaling limits, which could, in principle, be much larger.

Property	*Blue Gene/L*	*Blue Gene/P*	*Blue Gene/Q*
Node properties			
Cores per chip	2 × 440 PowerPC*	4 × 450 PowerPC	16 × A2
Processor frequency	0.7 GHz	0.85 GHz	1.6 GHz
Processor voltage	1.5 V	1.2 V	0.85 V
Coherency	Software managed	SMP	SMP
L3 cache size (shared)	4 MB	8 MB	—
L2 cache size (shared)	—	—	32 MB
Main store bandwidth	5.6 GB/s (16 bytes wide)	13.6 GB/s (2 × 16 bytes wide)	42.6 GB/s (2 × 16 bytes wide)
Peak performance	5.6 gigaFLOPS/node	13.6 gigaFLOPS/node	204.8 gigaFLOPS/node
Torus network			
Dimension	3D	3D	5D
Bandwidth	6 × 2 × 175 MB/s = 2.1 GB/s	6 × 2 × 425 MB/s = 5.1 GB/s	10 × 2 × 2 GB/s = 40 GB/s
System properties			
Peak performance	596 teraFLOPS (104 racks)	1 petaFLOPS (72 racks)	20 petaFLOPS (96 racks)
Core count	106,496	73,728	98,304
Node count	212,992	294,912	1,572,864
Total power	2.3 MW	2.3 MW	7.9 MW

increasing the yearly electricity bill by ∼$1 million at an $0.11/kilowatt-hour) rate, low-power processing nodes are an inherent requirement for highly scalable machines. The success of the low-power design approach resulted in the Blue Gene systems placing at the top of the Green500 list [5] of the most power-efficient supercomputers at the time of their respective introductions.

Blue Gene/Q uses a five-dimensional (5D) torus network to interconnect the compute nodes. The increased dimensions from three (for Blue Gene/L and Blue Gene/P) to five resulted in both a higher nearest-neighbor bandwidth and a higher bisectional bandwidth, and also resulted in fewer hops for a given total node count. These savings allow Blue Gene/Q to remove the separate global barrier and collective communication networks and instead allow the packets for these operations to be carried over the 5D torus network. By integrating the network logic on the compute chip, further architectural and power efficiencies were achieved. Multiple Direct Memory Access (DMA) devices in the chip's network-to-memory interface enable a maximum overlap of computation and communication. The first paper also illuminates the design methodology and bring-up effort of the Blue Gene/Q Compute chip.

Once the low-power approach was adopted, it was the responsibility of the packaging team to amass as many nodes as possible in a rack to realize the rack-level performance. In the second paper, Coteus et al. describe how 1,024 nodes were densely packaged into the Blue Gene/Q rack. The collective power at up to 100 kW/rack required a new look at power supplies and cooling designs. Thus, new ac/dc and dc/dc power supplies are presented in this paper. Water cooling was introduced in the thermal design, with 90% of the heat being removed by water and the remaining 10% by air. The electrical and mechanical packaging design of compute chips, networks, link chip, and input/output (I/O) racks, along with the cabling of ∼100 racks, is discussed. The topic of reliability permeates the paper, as it is an indispensable requirement to scale to multi-PFLOPS systems and beyond. **Figure 1** shows a high-level overview of the Blue Gene/Q system compute and I/O nodes. This figure is referenced in many of the papers in this issue and thus is included here. Illustrated is how a single compute

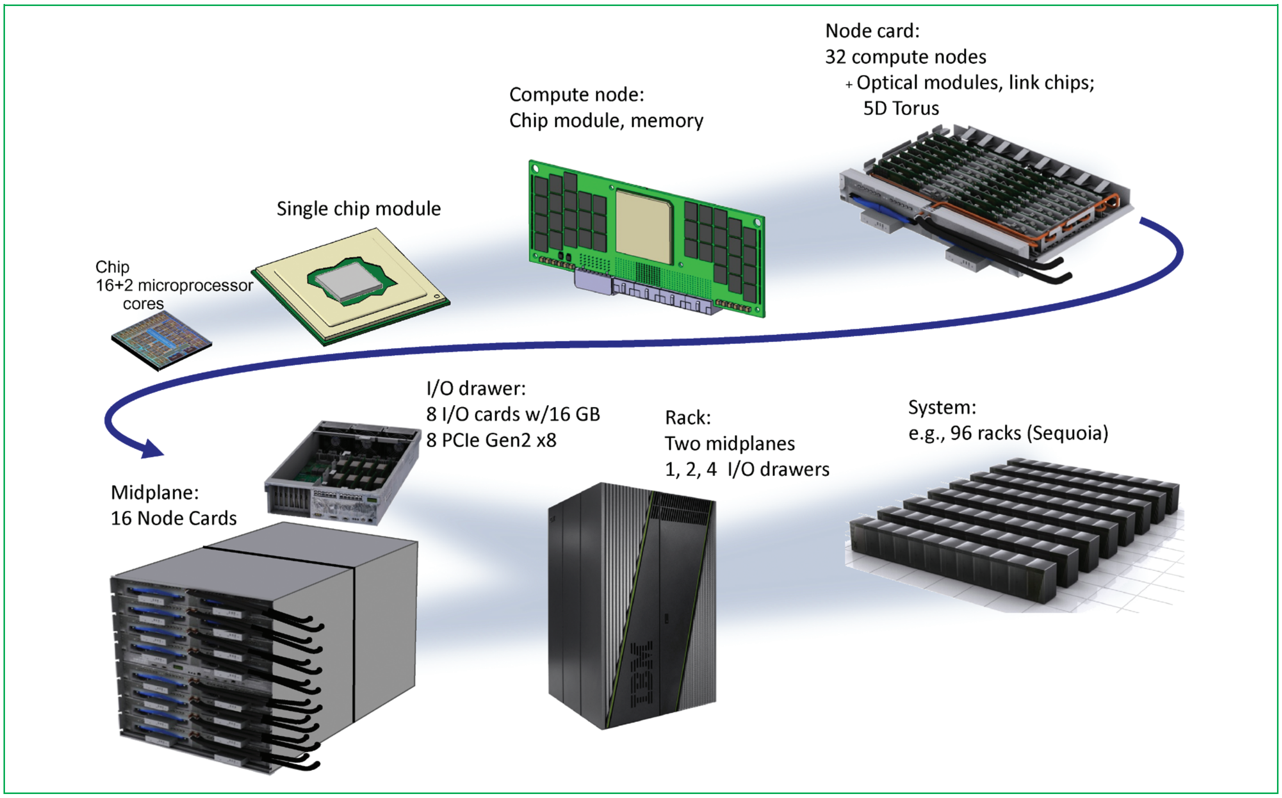

Figure 1

Blue Gene/Q packaging hierarchy: the construction of the 96 racks of Blue Gene/Q processor chips plus I/O drawers, which form the heart of the LLNL Sequoia System. ©Springer-Verlag, 2012. Reprinted with permission.

application-specific integrated circuit (ASIC) is packaged with its associated memory to form a compute node. The aforementioned 5D torus connections are realized on a water-cooled node board connecting $2^5 = 32$ compute nodes with a common power supply. Sixteen such node boards are connected through a midplane circuit board to form a $2^5 \times 2^4 = 512$-node half rack. Midplanes are then connected optically, to form systems of up to 1,024 racks. Each midplane is generally associated with one or more I/O drawers, which allow connections through PCI Express** (PCIe**) adapter cards to storage, tape, and other servers. Much more detail is covered in Paper 2.

Continuing the low-power theme, the third paper, by Sugavanam et al., explains the techniques used to monitor and manage the power of the various subunits of the Blue Gene/Q chip. By using the option of turning off selective subunits, the contributions of leakage power, clocking power, and active power for various subunits were measured. A power characterization technique for the development of application-dependent power projection models is also included in the paper. Differences between the power estimated before chip fabrication versus the measured power are discussed. The architectural-level, design-level, and manufacturing-level techniques used in Blue Gene/Q to achieve high energy efficiency are presented, including multiple frequency domains and single-instruction, multiple-data (SIMD) floating-point, clock gating, the use of multiple threshold voltage devices, and speed binning.

In order to understand application-level power/performance tradeoffs on current computer systems, runtime monitoring capabilities are needed. Specifically, very fine-grained monitoring capabilities are needed to enhance the insights on power and performance behavior. In the fourth paper, Bertran et al. describe a new experimental technique to perform automatic fine-grained power and performance characterization of applications on the IBM Blue Gene/Q platform.

For Blue Gene/Q, the system software group participated in the hardware/software co-design tradeoffs from the beginning. The fifth paper, by Ryu et al., presents the system software from this perspective. As an example, one of the challenging aspects of the Blue Gene/Q system design has been the development of communication software that scales, inside a node, to 16 cores, or 64 threads. The co-design process greatly eased the transition to the high thread count per node. For reliability and scalability reasons, Blue Gene/Q maintained the simple Compute Node Kernel (CNK) on each compute node. Asynchronous background processes are executed on the seventeenth core of the node, further reducing the operating noise on the 16 user cores to a minimum. In addition, this paper also addresses the control system, compiler, and messaging.

In the sixth paper, the IBM Blue Gene team describes the architectural simulator Mambo and the Twinstar FPGA platform that was used to verify logic correctness of the compute chip before the chip was released to manufacturing. Significantly, the thoroughness of the Twinstar simulations allowed a single-pass correct design for the multi-versioned L2 cache, which supports transactional memory (TM) and speculative execution (SE).

In the seventh paper, Ohmacht et al. describe the multi-versioned L2 cache, which includes hardware support for scalable locks, TM, and SE with rollback. Blue Gene/Q is the first commercial processor to implement these advanced memory concepts in hardware, and they will significantly simplify the programming and help efficient execution of many parallel threads. Note that the corresponding TM and SE software is covered by Bellofatto et al. in the fifth paper.

Blue Gene/Q is a homogeneous architecture. Through a crossbar switch between L1P and L2, each thread on each processor core can access any of the 16 slices of the 32 MB shared L2 cache uniformly and coherently. The external SDRAM-DDR3 main memory behind the L2 cache is also uniformly accessible. Unlike graphical processing unit (GPU)-based systems, the symmetric access of all cores to all banks of coherent memory simplifies programming, another major characteristic of the Blue Gene platform.

A Blue Gene/Q node can use either message passing interface (MPI) alone or a mixed MPI/OpenMP** programming model. In the eighth paper, Eichenberger and O'Brien detail the techniques of using OpenMP threads efficiently on a Blue Gene/Q node.

The ninth paper, by Evangelinos et al., is an application survey based on IBM's experience with tuning Livermore benchmarks and codes, and Argonne science application codes on Blue Gene/Q. Also included is an interesting industrial seismic imaging code using the RTM (Reverse Time Migration) algorithm.

Blue Gene/Q shows great promise for business analytics and large data applications. The Graph 500 aims to provide benchmark graph algorithms that will provide useful information on the suitability of supercomputers for data-intensive applications. In November 2011 and June 2012, Blue Gene/Q placed first on the Graph 500 list [6] by a convincing margin. These gratifying results are the theme of the tenth paper, by Checconi and Petrini.

LLNL was the earliest adopter of the Blue Gene platform. From 2004 to 2011, LLNL possessed the largest Blue Gene/L system, and LLNL will also own the largest Blue Gene/Q system. Some of LLNL's applications are described in the eleventh paper, by Carnes et al. They include LLNL's traditional applications such as ALE3D (Arbitrary Lagrangian-Eulerian 3D), ddcMD (classical molecular dynamics), and Qbox (molecular dynamics based on density functional theory), which ran on Blue Gene/L and Blue Gene/P previously and now on Blue Gene/Q. Using Blue Gene/L, ddcMD won the Gordon Bell Prize in 2005 and 2007, and Qbox won this prize in 2006. PF3D (laser-plasma interaction code) is a plasma fusion code that simulates deuterium/tritium fusion energy ignited by intense laser beams. Cardioid (heart modeling) is a newcomer written for the Blue Gene/Q generation for the first time.

In 2008, the Argonne Leadership Computing Facility (ALCF) started with 40 racks of Blue Gene/P. It was the largest HPC system at the time of its deployment dedicated entirely for science applications. The purpose of the ALCF is to provide cycles to a relatively small number of large scientific projects, which run simulations on significant fractions of the entire machine. Argonne fulfilled this mission admirably with Blue Gene/P since 2008, and now with 48 racks of Blue Gene/Q. In the twelfth paper, Coghlan et al. describe Argonne's applications including supernova explosions, cosmology with dark energy, thermal hydraulics for cooling nuclear reactors, density functional theory for materials research, classical molecular mechanics, weather and climate modeling, nuclear physics, plasma simulation of tokamaks, and the like.

The thirteenth paper, by Boyle et al., describes the simulation of QCD on Blue Gene/Q. The authors, from Columbia University led by Professor Norman Christ, and from Edinburgh University led by Peter Boyle, participated in the Blue Gene/Q design team throughout the development cycle. The converged Blue Gene/Q design ensured superior performance for HPC in general and QCD in particular.

Finally, the fourteenth paper, by Alam et al., documents IBM's collaboration with the Forschungszentrum Juelich (FzJ) and the Swiss National Supercomputing Center (CSCS) and describes applications on molecular dynamics (MP2C, CP2K, BigDFT), plasma physics (ORB5), and life sciences (LifeV). Juelich owned the largest Blue Gene/P system in the world and will have the largest Blue Gene/Q in Europe.

IBM is proud to have many repeat customers who challenged us, who collaborated with us in hardware and software, and who enabled IBM to produce machines of superior characteristics to meet their computational needs.

Conclusion

In this double issue of the *IBM Journal of Research and Development*, we describe the main thrust of the Blue Gene/Q supercomputer. From 2004 to 2012, we witnessed a 36-fold increase in performance. The hardware and software focus is on low power, high scalability, and high reliability. In addition, IBM made a conscious effort to move toward open source system software. Compared with cluster-based commercial-off-the-shelf systems, Blue Gene offers superior interconnectivity, power efficiency and reliability. Compared with GPU-based systems, Blue Gene offers ease of programming in addition to better power efficiency and reliability.

Not only has the performance of Blue Gene improved at a dizzying pace, but also application reach has grown in leaps and bounds. The increased amount of nodal memory, the flexibility of the mixed MPI/OpenMP programming model, and the high-performance network all have enabled Blue Gene/Q to become a general-purpose supercomputer.

Acknowledgments
The IBM Blue Gene/Q project has been supported and partially funded by Argonne National Laboratory and the Lawrence Livermore National Laboratory on behalf of the United States Department of Energy, under Lawrence Livermore National Laboratory subcontract no. B554331. The Blue Gene project is a team effort, benefiting from the cooperation of many individuals at IBM Research, IBM Systems and Technology Group, and IBM Software Group.

References
1. P. A. Boyle, D. Chen, N. H. Christ, M. A. Clark, S. D. Cohen, C. Cristian, Z. Dong, A. Gara, B. Joo, C. Jung, C. Kim, L. A. Levkova, X. Liao, G. Liu, R. D. Mawhinney, S. Ohta, K. Petrov, T. Wettig, and A. Yamaguchi, "Overview of the QCDSP and QCDOC computers," *IBM J. Res. & Dev.*, vol. 49, no. 2/3, pp. 351–365, Mar. 2005.
2. A. Gara, M. A. Blumrich, D. Chen, L.-T. Chiu, P. W. Coteus, M. E. Giampapa, R. A. Haring, P. Heidelberger, D. Hoenicke, G. V. Kopcsay, T. A. Liebsch, M. Ohmacht, B. D. Steinmacher-Burow, T. Takken, and P. Vranas, "Overview of the Blue Gene/L system architecture," *IBM J. Res. & Dev.*, vol. 49, no. 2/3, pp. 195–212, Mar. 2005.
3. Top500 List. [Online]. Available: http://www.top500.org
4. IBM Blue Gene team, "Overview of the IBM Blue Gene/P project," *IBM J. Res. & Dev.*, vol. 52, no. 1/2, pp. 199–220, Jan. 2008.
5. Green500 List. [Online]. Available: http://www.green500.org
6. Graph 500 List. [Online]. Available: http://www.graph500.org

Received June 11, 2012; accepted for publication June 13, 2012

IBM Blue Gene team The IBM Blue Gene team spans a wide range of technical disciplines and organizations within IBM. Team members are listed below.

From IBM
Mike Aho, David W. Alderman, Jon Allen, Eric Alter, Eberhard Amann, Anatoli Andreev, James Andrel, Sameh W. Asaad, John Attinella, Vernon Austel, Marcy L. Averill, Michael J. Azevedo, Joern Babinsky, Kit Barton, Darryl J. Becker, J. Richard Behun, Costas Bekas, Ralph Bellofatto, James R. Bentlage, Jeremy Berg, Darcy Berger, Randy Bickford, SuEllen Birkholz, Michael Blocksome, Matthias A. Blumrich, Hans Boettiger, Lynn Boger, Alan Boulter, Thomas C. Brennan, Jeremy J. Brewer, Bernard Brezzo, Arthur A. Bright, Jose Brunheroto, Jay S. Bryant, Nathan C. Buck, Tom Budnik, Daniel Buerkle, Raymond J. Bulaga, Mark Campana, Kenneth M. Caron, Bob Cernohous, Douglas Chartrand, Jeffery D. Chauvin, Fabio Checconi, Daniel C. Chen, Dong Chen, Rui C Chen, Wang Chen, Chen-Yong Cher, George LT Chiu, I-Hsin Chung, Lyman R. Clark, Paul K. Coffman, Evan Colgan, Miguel Comparan, Paul W. Coteus, Alessandro Curioni, Ron Daede, Bruce D'Amora, Kris Davis, Michael Deindl, Brian Deskin, Jun Doi, Marc B. Dombrowa, Roger Dong, Michael L. Eaton, Alexandre E. Eichenberger, Noel A. Eisley, Matthew R. Ellavsky, Robert F. Enenkel, Constantinos Evangelinos, Kahn C. Evans, Sean T. Evans, Steve Faas, Daniel Faraj, George A Fax, Mitchell D. Felton, Andrew Ferencz, Shawn Fetterolf, Joel Y. Ficke, Giovanni Fiorenza, Uwe Fischer, Blake G. Fitch, Ryan A. Fitch, Bruce M. Fleischer, Bill Flynn, Jeffrey Fosmo, Thomas W. Fox, John F. Fraley, Ross L. Franke, Scott Frei, Diego S. Gallo, Michael Gaynes, Yaoqing Gao, Robert Germain, Philip Germann, Mark E. Giampapa, Frank P. Giordano, Emanuel Gofman, Mike P. Good, Thomas Gooding, Nicholas Goracke, Jason Greenwood, Michael K. Gschwind, John A. Gunnels, Shawn Hall, Michael Hamilton, Ruud A. Haring, James S. Harveland, Troy L. Haugen, Charles L. Haymes, Philip Heidelberger, Timothy D. Helvey, Olaf Hendrickson, Brent Hilgart, Misky H. Hillestad, Judith W. Hjortness, Dennis Y. Huang, Todd Inglett, Hans Jacobson, Randy T. Jacobson, Geert Janssen, Mark Jeanson, Anto A. John, Mark C. Johnson, Steven P. Jones, Kirk E. Jordan, Jeffrey N. Judd, Kerry T. Kaliszewski, Robert Kammerer, Mohit Kapur, Michael Kaufmann, Amanda R. Kaufer, Chulho Kim, Kyu-hyoun Kim, David Klepacki, Ravi Komanduri, Gerard V. Kopcsay, Anatoly Koyfman, Brant L. Knudson, Jon Kriegel, Eric Kronstadt, Sameer Kumar, Lih-Chung Kuo, Mark Kupferschmidt, Teodoro Laino, Alphonso P. Lanzetta, Christopher A. Lapkowski, Jay A. Lawrence, David Lawson, Gene Leung, Tak O. Leung, Qing AG Li, Thomas A. Liebsch, Tao T Liu, Yan Liu, Meryl Lo, Ligang Lu, Ray Lucas, Bob Lytle, Scott H. Mack, Serban Maerean, Karen A. Magerlein, David Malone, Amith R. Mamidala, Jim Marcella, Christopher M. Marroquin, John K. Masi, Thilo Maurer, Patrick J. McCarthy, Moyra K. McManus, Mark Megerian, Douglas Miller, Sam Miller, Khaled A. Mohammed, Jaime H. Moreno, Adam Muff, Patrick Mulligan, Roy Musselman, Tom Musta, Indira Nair, Ben Nathanson, Mike Nelson, Hoang N. Nguyen, Carl Nilsen, Kinya Noguchi, Carl Obert, Kathryn O'Brien, Kevin K. O'Brien, Alda S. Ohmacht, Martin Ohmacht, Bitwoded Okbay, Michael R. Ouellette, Bruce Owens, Mike J. Palmer, Benjamin J. Parker, David P. Paulsen, Michael P. Perrone, Fabrizio Petrini, Kerry Pfarr, Huyen Phan, Swetha Pullela, Irina Rada, Don Reed, Michael T. Repede, Dennis Rickert, Thomas Roewer, Bryan S. Rosenburg, Michael G. Rosenfield, Jeff Ruedinger, Kyung Dong Ryu, Yogish Sabharwal, Proshanta K. Saha, Takehito Sakuragi, Valentina Salapura, David Salinas, David L. Satterfield, Jun Sawada, Vipin Sachdeva, Vaibhav Saxena, Paul Schardt, Matthew Scheckel, Brandon Schenck, Heiko Schick, Dietmar Schmunkamp, Bob Schoen, Andrew A. Schram, Brian Schuelke, Alan D. Secor, Faith W. Sell, Woody Sellers, Robert M. Senger, James Sexton, Vinay V. Shah, Raul E. Silvera, Karl Solie, David L. Sparks, Burkhard Steinmacher-Burow, Will Stockdell, Scott Strissel, Craig Stunkel, Krishnan Sugavanam, Yutaka Sugawara, Nobu Suginaka, Christopher Surovic, Corey Swenson, Keith A. Tally, Yoshihisa Takatsu, Todd Takken, Andrew Tauferner, Kiswanto Thayib, John Thomas, Shurong Tian, Barry M. Trager, Ailoan T. Tran, Scott Trcka, Michael M. Tsao, Priya Unnikrishnan, Jim Van Oosten, Jason L. VanVreede, Michael Vaughn, Pascal Vezolle, Ivan Vo, Rebeccah Vossberg, Martha Voytovich, Charles Wait, Robert E. Walkup, Amy Wang, Bryan J. Weatherford, Valery Weber, Shmuel Winograd, Bruce Winter, Bryon Wirtz, Kelvin Wong, Michael Wong, Peng Wu, Hanhong Xue, Brett M. Yokom, Guo Zhong Zhang, Yong Zheng, Ching Zhou,

Jianwei (John) Zhuang, Fred Ziegler, Matthew Ziegler, and
Christian G. Zoellin

From Columbia University
Norman H. Christ

From Columbia University and RIKEN Brookhaven National Laboratory Research Center
Changhoan Kim

From University of Edinburgh
Peter A. Boyle

Formerly at IBM
Jeremy Balster, Thomas Cipolla, James Clemens, Carlos Costa, Paul Curtis, Bill Donegan, Gabor Dozsa, Don Eisenmenger, Matthias Fritsch, Alan Gara, Roberto Gioiosa, Russell Hoover, David J. Iverson, Fariba Kasemkhani, Stefan Koch, Brian Koehler, David E. Lackey, Cory Lappi, Derek Lieber, Dan Littrell, Lurng-Kuo Liu, Matt Logelin, Curt Mathiowetz, Eric O. Mejdrich, Virginia Metayer, Timothy Moe, Mike Mundy, Eldon Nelson, Dennis Olson, Jeff Parker, Rick A. Rand, Joseph Ratterman, Rony Sakuragui, Daniele P. Scarpazza, Marcel Schaal, Ryan J. Schlichting, Robert Shearer, Brian Smith, Jose A. Tierno, Catherine Trammell, Matthew R. Tubbs, Simeon Wahl, Alfred T. Watson, Tim Wensky, Robert W. Wisniewski, and Xiaotong Zhuang