the Exascale Report™
A Global Perspective

ARGONNE TALKS EXASCALE – and MIRA

written by Mike Bernhardt
March 2011

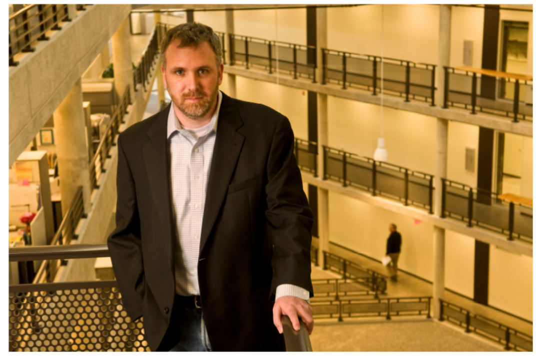
Argonne National Laboratory, established in 1946, owns the prestigious designation as the first national science and engineering research laboratory in the United States.

I first worked with the scientists and researchers at Argonne National Laboratory in the early 1990s, at a time when the global HPC community demonstrated unprecedented levels of collaboration while forging the path to teraFLOPS. It was an exciting time for computational sciences and technology achievement. During that period, Argonne firmly established a globally-recognized leadership position in computational sciences, and, since that time, has continued to provide an environment that fosters strong, open collaboration and innovative thinking.

In 2004, Argonne was instrumental in working with IBM to establish an important architecture-focused technology consortium, the International Blue Gene Consortium, a milestone example of collaboration that was effective in rallying industry and academia around a new architecture and developing momentum ahead of product and technology availability. This is the same type of collaboration challenge we face on a much larger scale today as we approach exascale.

Recently, Argonne has been in the news with "Mira" – the anticipated 10-petaFLOPS IBM Blue Gene/Q supercomputer expected to go online in 2012.

To get a better understanding of what Mira is all about, and what it means for exascale computing, we interviewed Mike Papka, the Deputy Associate Laboratory Director for Computing, Environment, and Life Sciences at Argonne National Laboratory, Acting Division Director of the Argonne Leadership Computing Facility, and Senior Fellow in the Argonne National Laboratory / University of Chicago Computation Institute.



Mike Papka
Argonne National Laboratory

The Exascale Report: ***First of all Mike, would you give our readers an overview of what major systems you have installed today?***

Papka: Argonne's leadership-class system is an IBM Blue Gene/P called Intrepid—a 557-teraflop machine with more than 40,000 quad-core compute nodes. It's one of two petascale machines in the nation dedicated to open science.

The LCF also hosts a single rack Blue Gene system called Surveyor, which has 1,024 quad-core compute nodes, and is used for tool application porting, software testing and optimization, and systems software development. Other ALCF resources include Eureka, a 100-node dual-quad core computer cluster used for visualization and data analytics, and Gadzooks, a 4-node dual-quad compute cluster used for visualization testing and development.

TER: ***How many researchers, directly or indirectly, use the Argonne Leadership Computing Facility on a regular basis?***

Papka: ALCF has roughly 700 unique users spread across four programs. We support 30 INCITE projects, nine ASCR Leadership Computing Challenge Program projects, 16 Early Science Program projects, and numerous Director's Discretionary awards.

TER: ***How much power is currently being used by the facility – and how will that change with the installation of Mira?***

Papka: Last year, the ALCF won an Environmental Sustainability (EStar) award for the innovative energy efficient cooling it designed for its current system, and Argonne researchers anticipate that our Blue Gene Q, called Mira, will be significantly more power friendly.

TER: ***Are there particular application / research areas in which the ALCF is playing a key role?***

Papka: Initially, ALCF is focused on the Early Science program applications. After that, as the machine matures, we expect roughly the same project mix as Intrepid.

The Early Science program aims to bring Mira into full production status in a period of months from when the machine is first accepted. The facility "shakes down" the system with the help of a small community of users running production applications, called the Early Science projects. During this period, users will assist in identifying the root causes of any system instabilities, and will work with ALCF staff to help develop solutions.

The 16 DOE-approved Early Science projects span a diverse range of scientific fields, numerical methods, programming models, and computational approaches—including particle-mesh methods, adaptive meshes, spectral methods, Monte Carlo, molecular dynamics, and ab initio computational chemistry methods. Notably, these applications also represent a large fraction of the ALCF's current and projected computational workload.

Argonne's Blue Gene/P Supercomputer

TER: ***What activity is currently taking place to prepare for this system? What happens this far in advance when planning for a 10 petaFLOPS system?***

Papka: We have completed all the pre-work for power and cooling. As part of the upgrade, we installed a state-of-the-art automatic chiller plant that will automatically swing between "free cooling" and the chillers, allowing us to conserve energy.

We are also getting our Early Science applications ready for Mira. This involves shifting to hybrid programming models from a pure messaging passing model. The testing of such changes is being carried out on our BGP where the software stack is the same as on BGQ hardware.

TER: ***We assume such a system will be installed and tested in phases. Is this the case, and can you describe the installation and deployment phases for Mira?***

Papka: It will be a fairly standard installation—the infrastructure goes in first, then the storage, and finally the compute racks. We have a set of acceptance tests to ensure that the system works as a whole.

TER: ***Can you give us some more details about Mira's specifications?***

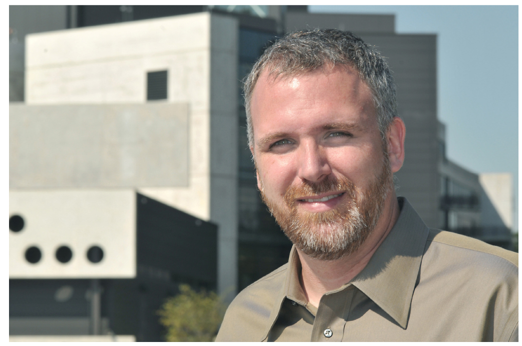
Papka: Mira will have 49,152 16-core compute nodes (786,432 processing cores) and 786 terabytes of memory. Its peak performance will be 10 petaflops. Mira will be used for production scientific and engineering computing. The compute nodes are connected by a 5-dimensional torus network that supports multiple network functions, including high-performance low-latency point-to-point communications, highly-scalable collective operations, and fast barrier operations. GPFS will provide a robust and stable home directory file system as well as the parallel I/O data file system.

TER: ***If the plan is to have Mira up and running in 2012, is it really feasible that we will see exascale-class systems by 2018-2019, only 6-7 years later?***

Papka: It will be a challenge! These systems will likely have billions of cores and tens of billions of threads. It will take a significant investment by the government and industry research labs, and the coordinated efforts of the computational science community. Because scientific computing is such a small part of the IT industry, it requires complex and coordinated R&D efforts to bring down the cost of memory, networking, disks and all of the other essential components of an exascale system.

Computer scientists face at least three challenges in the move to exascale. Exascale programs must harness billion-way parallelism, which is a huge logistical problem.

The second challenge is power. A billion-processor computer made with today's technology would consume more than a gigawatt of electricity. Third, the dramatic increase in the components needed to reach an exaflop will require news methods or approaches to address reliability.



TER: ***How long was the acquisition cycle for Mira – from the time of first issuing an RFP to the time of the award to IBM? Was Argonne able to contribute to design improvements?***

Papka: Over the past decade, Argonne and Lawrence Livermore National Laboratory have partnered with IBM Research to design the Blue Gene architecture family, starting with the Blue Gene L—the first supercomputer to break the 100 teraflops mark. The BGP, introduced in 2008, was optimized based on our experience and user requirements, and has broad, worldwide adoption in the scientific and engineering community. Many of the BGQ improvements are based on the successful deployments of BGP. The BGQ design was also driven by hands-on involvement from users through the international BG Consortium; a group established in 2004 by Argonne and IBM that includes 42 U.S. universities, ten federal laboratories, six members from U.S. industry, and 25 international member sites.

the Exascale Report tm

A Global Perspective

TER: ***What does Mira mean as a stepping stone to achieving exascale-class systems?***

Papka: It's a big step! Today, the main challenge is scaling up applications to achieve increased performance from many more cores while adapting them to the architectures that represent a low-power sustainable future. The BGQ system does both. It is also the first example of a petascale system that embodies many features that will be needed as we continue to push towards exascale, and it provides a stepping-stone to sustainable low-power platforms for computational science and engineering. It was designed to meet the computational needs of the community, dramatically reduce power requirements, and introduce novel I/O and data management strategies that address the demands of both data-intensive and traditional high-performance applications.

TER: ***What points are we missing in this discussion?'***

Papka: The missing point is software. Talking only about the hardware neglects a huge investment in software that needs to be made as we move forward.

TER: Well, we couldn't agree more. And with that, we'll close out this interview. Thanks to Mike Papka for taking the time to speak with us. We plan to hit the topic of software in a future issue of The Exascale Report, and we'll certainly bring the folks from Argonne into that discussion.

<<<<< >>>>>