

## Two CRD Staff in SIAM Leadership Positions

Juan Meza, head of CRD's High Performance Computing Research Department, and Esmond Ng, leader of the Scientific Computing Group, were recently named to leadership positions in the Society for Industrial and Applied Mathematics.



Juan Meza

Meza has been appointed to the Board of Trustees of the SIAM, which was founded in 1951 and currently has more than 10,000 members.

The goals of SIAM are to:

- advance the application of mathematics and computational science to engineering, industry, science, and society;
- promote research that will lead to effective new mathematical and computational methods and techniques for science, engineering, industry, and society;
- provide media for the exchange of information and ideas among mathematicians, engineers, and scientists.

Ng was elected to a two-year term as vice chair of the SIAM Activity Group on Supercomputing. The activity group provides a forum for computational mathematicians and scientists, computer scientists and computer architects to exchange ideas



Esmond Ng

on mathematical algorithms and computer architecture needed for high-performance computer systems. The group also organizes a biennial conference on parallel processing for scientific computing.

## CRD Report

CRD Report is published every other month, highlighting recent achievements by staff members in Berkeley Lab's Computational Research Division. Distributed via email and posted on the Web at <http://crd.lbl.gov/DOEResources>, CRD Report may be freely distributed. CRD Report is edited by Jon Bashor, [JBashor@lbl.gov](mailto:JBashor@lbl.gov) or 510-486-5849.

## Berkeley Lab Researchers Analyze Performance, Potential of Cell Processor

Though it was designed as the heart of the upcoming Sony PlayStation3 game console, the STI Cell processor has created quite a stir in the computational science community, where the processor's potential as a building block for high performance computers has been widely discussed and speculated upon.

***“Overall results demonstrate the tremendous potential of the Cell architecture for scientific computations in terms of both raw performance and power efficiency.”***

The paper, “The Potential of the Cell Processor for Scientific Computing,” was written by Samuel Williams, Leonid Oliker, Parry Husbands, Shoaib Kamil and Katherine Yelick of Berkeley Lab's Future Technologies Group and John Shalf from NERSC.

To evaluate Cell's potential, computer scientists at the U.S. Department of Energy's Lawrence Berkeley National Laboratory evaluated the processor's performance in running several scientific application kernels, then compared this performance against other processor architectures. The results of the group's evaluation were presented in a paper at the ACM International Conference on Computing Frontiers, held May 2-6, 2006, in Ischia, Italy.

“Overall results demonstrate the tremendous potential of the Cell architecture for scientific computations in terms of both raw performance and power efficiency,” the authors wrote in their paper. “We also conclude that Cell's heterogeneous multi-core implementation is inherently better suited to the HPC environment than homogeneous commodity multicore processors.”

Cell, designed by a partnership of Sony,

*(continued on page 2)*

## New Microbial Metagenome Data Analysis System in Production

An experimental metagenomics data management and analysis system co-developed by the Biological Data Management and Technology Center (BDMTC) at Berkeley Lab with the Genome Biology Program and Microbial Ecology Program at DOE's Joint Genome Institute and released earlier this year is also already helping produce scientific discoveries.

Called IMG/M, the system extends the Integrated Microbial Genomes (IMG) system with the ability to integrate and analyze metagenome data, and has provided immediate support for metagenomics studies at JGI. “IMG/M is the first publicly available metagenome data management and analysis system—the first of its kind,” said Victor Markowitz, head of BDMTC.

“Most research in microbial genome analysis focuses on individual organisms,” said Nikos Kyrpides, head of JGI's Genome Biology Program. “The application of high throughput sequencing to environmental samples has revealed a new universe of microbial community genomes including mostly

microorganisms unknown to science. Microbial community genome analysis, also known as metagenomics, focuses on how the entire community functions together.”

Thus far, IMG/M has been used by Phil Hugenholtz, head of the Microbial Ecology Program, and his colleagues at JGI for completing the analysis of enhanced biological phosphorus removing sludge communities and for studying the metagenomes of several key microbial communities recently sequenced by DOE JGI, including the lignocellulose-hydrolyzing communities in termite hindguts.

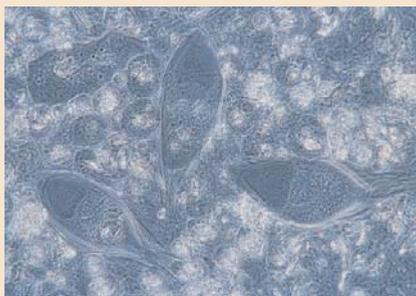
“IMG/M has proven to be an extremely useful resource and tool for analyzing our metagenomic data,” said Jared R. Leadbetter, associate professor of environmental microbiology at the California Institute of Technology and collaborator on the termite hindgut microbial community for bioenergy project. “Such datasets are large, complex, and potentially unwieldy. Importantly, IMG/M is more than just an excellent tool to analyze

*(continued on page 2)*

## IMG/M Proving Useful for Analyzing Complex Metagenomic Data *(continued from page 1)*

data. The manner in which the results of that analysis are organized and made accessible through a user-friendly interface allows the researcher to rapidly move in a number of different intellectual directions. As a result, the user becomes better educated with and gets a real feel for the data in a manner that would not otherwise be possible on such short time scales."

IMG/M was presented by Markowitz at the recent Keystone Symposium on Microbial Community Genomics in Animals and the Environment, organized by DOE JGI Director Edward M. Rubin and Edward F. DeLong, professor in the Division of Biological Engineering and Department of Civil and Environmental Engineering at the Massachusetts Institute of Technology.



IMG/M is being used to study microbial communities, such as these protists found in tirmite guts.

"IMG/M provides an intuitive interface, and nice complement to IMG for comparing gene content and phylogenetic profiles of microbial genomes, and relating them to the large

microbial community datasets now accumulating," DeLong said. "These are great and sorely needed data-exploration tools."

IMG/M was demonstrated at a workshop on April 1, as part of the DOE JGI First Annual User Meeting. Additionally, BDMTC staff members will be presenting the project this summer at workshops and symposia in England, Brazil and Austria.

"We have an experimental system which has proved to be immediately useful in helping scientists to conduct their studies effectively," Markowitz said.

IMG/M is accessible to the public at <http://img.jgi.doe.gov/m>.

## Cell Processor Analysis Shows Stay Potential for HPC Applications *(continued from page 1)*

Toshiba and IBM, is a high performance implementation of software-controlled memory hierarchy in conjunction with the considerable floating point resources that are required for demanding numerical algorithms. Cell takes a radical departure from conventional multiprocessor or multi-core architectures. Instead of using identical cooperating commodity processors, it uses a conventional high performance PowerPC core that controls eight simple SIMD (single instruction, multiple data) cores, called synergistic processing elements (SPEs), where each SPE contains a synergistic processing unit (SPU), a local memory, and a memory flow controller.

Despite its radical departure from mainstream general-purpose processor design, Cell is particularly compelling because it will be produced at such high volumes that it will be cost-competitive with commodity CPUs. At the same time, the slowing pace of commodity microprocessor clock rates and increasing chip power demands have become a concern to computational scientists, encouraging the community to consider alternatives like STI Cell. The authors examined the potential of using the forthcoming STI Cell processor as a building block for

future high-end parallel systems by investigating performance across several key scientific computing kernels: dense matrix multiply, sparse matrix vector multiply, stencil computations on regular grids, as well as 1D and 2D fast Fourier transformations.

According to the authors, the current implementation of Cell is most often noted for its extremely high performance single-precision (32-bit) floating performance, but the majority of scientific applications require double precision (64-bit). Although Cell's peak double precision performance is still impressive relative to its commodity peers (eight SPEs at 3.2 GHz = 14.6 Gflop/s), the group quantified how modest hardware changes, which they named Cell+, could improve double-precision performance.

The authors developed a performance model for Cell and used it to show direct comparisons of Cell with the AMD Opteron, Intel Itanium2 and Cray X1 architectures. The performance model was then used to guide implementation development that was run on IBM's Full System Simulator in order to provide even more accurate performance estimates.

The authors argue that Cell's three-level

memory architecture, which decouples main memory accesses from computation and is explicitly managed by the software, provides several advantages over mainstream cache-based architectures. First, performance is more predictable, because the load time from an SPE's local store is constant. Second, long block transfers from off-chip DRAM can achieve a much higher percentage of memory bandwidth than individual cache-line loads. Finally, for predictable memory access patterns, communication and computation can be effectively overlapped by careful scheduling in software.

While their current analysis uses hand-optimized code on a set of small scientific kernels, the results are striking. On average, Cell is eight times faster and at least eight times more power efficient than current Opteron and Itanium processors, despite the fact that Cell's peak double-precision performance is fourteen times slower than its peak single-precision performance. If Cell were to include at least one fully utilizable pipelined double-precision floating point unit, as proposed in their Cell+ implementation, these speedups would easily double.

The full paper can be read at: <http://www.cs.berkeley.edu/~samw/projects/cell/CF06.pdf>.

### DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California. Ernest Orlando Lawrence Berkeley National Laboratory is an equal opportunity employer.