LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-404088

Simulating the Birth of the Universe on a PetaFlop

T. Luu, R. Soltz, P. Vranas

May 22, 2008

Computing in Science and Engineering

Disclaimer

PETAFLOP ESSAY CONTEST WINNER

Simulating the Birth of the Universe on a Petaflop

By Thomas Luu, Ron Soltz, and Pavlos Vranas

IN THE BEGINNING, WHEN THE UNIVERSE WAS LESS THAN ONE MICROSECOND OLD AND MORE THAN ONE TRILLION DEGREES HOT, IT TRANSFORMED FROM A PLASMA OF QUARKS AND GLUONS INTO BOUND STATES OF QUARKS WE REFER TO AS PROTONS AND NEUTRONS, THE FUNDAMENTAL BUILDING BLOCKS OF NUCLEAR MATTER THAT MAKE UP MOST OF THE visible universe. We believe this happened because the theory of quantum chromodynamics (QCD), which governs the interactions of the strong nuclear force, predicts it should happen when such conditions occur. Recent experiments at the Relativistic Heavy Ion Collider at Brookhaven National Laboratory have provided direct evidence of the existence of this phase transition in collisions between gold nuclei at the highest attainable energies.

But calculating the properties of this phase transition has been notoriously difficult and computationally challenging. What we've been able to discern via this theory has come by situating space and time onto a 4D grid of lattice points, on a volume no bigger than the size of a large nucleus. This discrete formulation of QCD known as lattice QCD (LQCD) can be tamed in a way that's numerically amenable to massively parallel machines. Indeed, our LQCD code scaled with perfect speedup all the way to the 131,072 CPU cores on the world's fastest supercomputer, the BlueGene/L (BG/L) system at Lawrence Livermore National Laboratory (LLNL). Efforts are currently under way to calculate this phase transition's properties using tens of teraflops spread across many of the world's fastest computers.

However, the lattice discretization used in our calculations of space-time breaks one of the most crucial properties of the underlying theory: the basic symmetry of chiral rotations. Chiral symmetry can't be fully restored on the lattice, but it can be restored and controlled to a high degree by using the method of domain wall fermions (DWF).

DWF introduces an extra fifth dimension, in which chiral symmetry is restored with only a small amount of breaking, which decreases as the fifth dimension's size increases. In particular, researchers have estimated that a fifth dimension of roughly 64 lattice points is sufficient to restore symmetry with a systematic error of only a few percent, even around the transition temperature. Studies have also shown that DWF exhibits excellent fidelity in incorporating the transition's basic driving forces. But the computational cost of this calculation is roughly $2 \times 64 = 128$ times higher than the cost of current methods that don't preserve this symmetry on the lattice, and thus it falls outside the realm of current terascale supercomputing. With petascale supercomputing, however, this restriction is lifted.

If we had petascale supercomputing for our calculation, we could use computational techniques similar to those used on BG/L to calculate the thermodynamic properties—or more precisely, the equation of state—of the quark-gluon plasma just as it begins to form protons and neutrons as it did in the very early universe. The basic level of parallelism again comes from the division of the 4D grid into subgrids assigned to nodes on a massively parallel machine, thereby providing perfect speedup. DWF's fifth dimension provides yet another level of parallelism: given that the nodes of a massively parallel petascale supercomputer will likely consist of several CPU cores per chip, it's natural to fully map the fifth dimension along the node chip's CPU cores. On a 16-CPU core chip, for example, we could assign four fifth-dimension "slices" per CPU core for a total of 64. In this way, the new level of parallelism in the application matches the new level of parallelism in the hardware. The communication patterns along the fifth dimension are nearest-lattice-neighbor only and can be handled via shared memory or any other on-chip data transfer mechanism, thus alleviating off-chip data transfers while exploiting the hardware's native capability.

Based on current calculations using tens of teraflops on the LLNL BG/L supercomputer, our proposed simulation would require 128x more computing power—in other words, several petaflops for approximately three months.

Petaflop Contest for Supercomputing 2007

I'm sure some interesting comment on how these people won the contest will be placed here. Plus it will help bottom out the article.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

This level of computing could provide a glimpse of the early universe as we've never seen it before and lead to a dramatic improvement in our understanding of the interactions of nature's smallest particles, quarks, which also comprise its largest visible structures.

Acknowledgments

This work is supported under the auspices of the US Department of Energy by the University of California, Lawrence Livermore National Laboratory, under contract W-7405-Eng-48.

Thomas Luu is a staff physicist at Lawrence Livermore National Laboratory. His research interests include many-body nuclear physics, cold-atom physics, and lattice quantum chromodynamics applied to few-body nuclear systems. Luu has a PhD in physics from the University of Washington. Contact him at tluu@llnl.gov.

Ron Soltz is a staff physicist at Lawrence Livermore National Laboratory. His research interests include relativistic heavy ion collisions, lattice quantum chromodynamics, and investigating methods for detecting nuclear materials. Soltz has a PhD in physics from MIT. Contact him at soltz@llnl.gov.

Pavlos Vranas is a staff physicist at Lawrence Livermore National Laboratory. His research interests include lattice quantum chromodynamics and physics beyond the standard model and supercomputing. Vranas has a PhD in physics from the University of California, Davis. Contact him at vranasp@llnl.gov.

Petaflop Essay Contest Runnerup

Petaflop-Enabled Simulation of Molecular Evolution at an Atomic Scale

By Christopher M. Frenz

ONE OF THE MOST PROMINENT USES OF COMPUTATIONAL METHODS IN THE BIOPHYSICAL SCIENCES IS TO HELP US UNDERSTAND THE FUNCTION OF BIOLOGICAL molecules at the atomic level. Researchers often use techniques that employ varying degrees of classical and quantum mechanical methods to predict molecular motions and other physical characteristics from atomic interactions.[1] Although these simulations often provide insights into the structural basis for molecular functionality, they're extremely computationally intensive. Simulations of large protein systems that comprise multiple proteins, and hence tens of thousands of atoms and atomic interactions, can therefore only be conducted in finite time by the most powerful computing environments. Computational time can be a limiting factor even for smaller protein systems, making it difficult to detect characteristics such as large-scale molecular motions in a protein.[2]

The availability of a computing environment with petaflop calculation capabilities, however, would make it feasible not only to perform atomic-scale calculations of protein systems but would let us simulate an entire population of cells at an atomic level. The large increase in computational power would likely even make it feasible to observe large-scale molecular interactions and catalytic reactions *in silico*.

Such simulation capabilities would provide an interesting environment for studying molecular evolution as well. The introduction of mutations into an organism's genome might be a random event, but recent evidence suggests that

the physics of the proteins expressed from those genes plays a strong role in whether mutations remain. Recent studies have demonstrated correlations between amino acid residue conservation and protein physical properties such as electrostatics, energetics, and residue packing.[3–5] Computational methods exist for computing each of these properties, so it's computationally feasible that petascale computing could help us develop a way to predict a mutation's viability over evolutionary time. Moreover, we could use other computational methods such as docking to assess the actual functional impact a given mutation would have on a biological system, such as by predicting a change in ligand binding or in the potential for molecular interactions.[6]

Conducting an atomic-level simulation of a cell would be a major breakthrough for the biological sciences in and of itself because it would let us observe the metabolic pathways and all the biochemical steps and the interactions that comprise them with an unprecedented level of detail. Yet, by simulating a population of such cells and bestowing on them a reasonable degree of genetic diversity and the ability to replicate, the possibilities are even more profound. We could expose such a simulation system to some type of environmental stressor and then monitor it to see what types of mutations enabled certain cells to survive and breed while others died out. These findings, in turn, could help scientists come closer to answering some of the most pressing questions humans have faced, including how we got here and where we come from.

The evolutionary insights gained from developing a simulation of this nature are only a fraction of its utility, however—such a simulation system would also have practical applications to medicine and biotechnology. A common problem that the healthcare field faces is that many viruses and bacteria have developed drug-resistant strains that curtail the effect certain pharmaceuticals such as antibiotics pose to them.[7] With petascale-enabled simulations, we could expose a simulated population of cells or viruses to a pharmaceutical stressor and try to predict *in silico* what types of drug resistance will likely develop, the biological effects of the resistance mutations, and how long it might take for such strains to appear. Moreover, much of the biotechnology industry is concerned with developing enzymes or metabolic systems that can carry out novel metabolic reactions, such as biodegrading oil spills.[8] With a petascale-based simulation methodology, we could expose a cell population to the novel metabolite as an environmental stressor. The cells that evolved the ability to process the metabolite would be the ones fit enough to survive the simulation, allowing for the potential introduction of this mutation into actual biological cells and thereby bestowing the "evolved" capabilities on them. Thus, the simulations could serve as an extremely sophisticated method of bioengineering via directed evolution.

This type of cellular evolution simulation at an atomic level of detail would be one of the most beneficial uses for a petaflop computing system. It would not only have an insurmountable value for deepening our understanding of one of biology's greatest phenomena—evolution—but it would also be of great practical value for both medicine and biotechnology.

References

1. D.A. Case et al., "The Amber Biomolecular Simulation Program," *J. Computational Chemistry*, vol. 26, no. 16, 2005, pp. 1668–1688.
2. C.M. Frenz, "Possibilities and Limitations of Computer Simulation," *IEEE Potentials*, vol. 26, no. 2, 2007, pp. 30–33.
3. C.M. Frenz, "Interrelationship between Protein Electrostatics and Evolution in HCV and HIV Replicative Proteins," *Proc. 2007 Int'l Conf. Bioinformatics and Computational Biology* (Biocomp 07), CSREA Press, 2007, pp. 91–98.
4. D.R. Livesay et al., "Conservation of Electrostatic Properties within Enzyme Families and Superfamilies," *Biochemistry*, vol. 42, no. 12, 2003, pp. 3464–3473.
5. H. Liao et al., "Protein Sequence Entropy Is Closely Related to Packing Density and Hydrophobicity," *Protein Eng. Design and Selection*, vol. 18, no. 2, 2005, pp. 59–64.
6. D.S. Goodsell and A.J. Olson, "Automated Docking of Substrates to Protein by Simulated Annealing," *Proteins: Structure, Function, Genetics*, vol. 8, no. 3, 1990, pp. 195–202.
7. C.F. Higgins, "Multiple Molecular Mechanisms for Multidrug Resistant Transporters," *Nature*, vol. 446, no. 7137, 2007, pp. 749–757.
8. A. Yamatsu et al., "Isolation and Characterization of a Novel Poly(vinyl alcohol)-Degrading Bacterium, Sphyngopyxis sp. PVA3," *Applied Microbiological Biotechnology*, vol. 72, no. 4, 2006, pp. 804–811.

Christopher M. Frenz is an instructor in the Department of Computer Engineering Technology at the New York City College of Technology (CUNY) and the author of the books *Visual Basic and Visual Basic .NET for Scientists and Engineers* (Apress, 2002) and *Pro Perl Parsing* (Apress, 2005). His research interests include modeling protein structure and function relationships as well as the development of machine learning and artificial intelligent approaches to protein engineering. Contact him at cfrenz@gmail.com.