Technical Perspective

Computer Science Takes on Molecular Dynamics

By Bob Colwell

PUT THIS ON your to-do list: read the following paper by researcher David Shaw and colleagues that describes their Anton molecular dynamics (MD) engine. Shaw's Anton engine applies leading-edge computer science concepts to the biologically crucial problem of modeling molecular interactions. In an era when much of our most advanced computer technology is spent creating ever more horrible creatures that we can shoot ever bigger virtual holes in, the idea of productively using this technology to explore nature at its most up-close-and-personal is both exciting and reassuring.

The nature of the computational problem Anton aims to solve, and the unique aspects of the resulting design, are fascinating peeks into a corner of the computer design space we seldom get to visit—even though each of us is a biological machine that relies on the correct functioning of molecular mechanisms. When diseases cause these mechanisms to go awry, medical researchers try to infer the causes and possible remedies from very indirect and error-prone evidence, as they lack direct means of measuring or simulating the molecular underpinnings. David Shaw calls his new instrument a “computational microscope,” and if successful it stands to make the same kind of game-changing impact that Anton van Leeuwenhoek's original optical microscope once did. (Shaw's machine was named in van Leeuwenhoek's honor.)

To appreciate what Shaw's machine is attempting, consider a system containing a realistic protein molecule together with a few layers of water molecules, which might together encompass tens of thousands of atoms. If calculation of the force between any two atoms takes 10 computer operations, then the total ops required per time step would be $(10^4 \text{ atoms}) \times (10^4 \text{ atoms}) \times 10 \text{ ops/atom} = 10^9 \text{ ops}$. Time slices are on the order of femtoseconds

(10^{-15} seconds), and simulations must run for milliseconds (10^{-3} seconds) to capture the biology being modeled. So we'll need to run those 10^9 ops for 10^{12} slices to reach a simulated millisecond—that's 31,000 years. We need six orders of magnitude speedup, roughly three orders of magnitude beyond today's fastest supercomputers.

But even if you weren't a biological unit with a vested interest in this effort, you could still appreciate the Anton design from a computer system perspective. General-purpose computer systems aspire to run everything well, but no one thing spectacularly well. Anton is designed to run a specific molecular dynamics workload spectacularly well. While a well-designed general system can bottleneck 100 different ways on 100 different benchmarks, Anton must try, in essence, to bottleneck everywhere, all at once, on its one workload.

This balancing act must be attempted in the face of imperfect knowledge of that one workload. For example, electrostatic interactions between two atoms that aren't sharing any electrons are considered to be well understood, and are the most numerous, so Anton applies very specific, very parallel, and very inflexible hardware to handling them. Less is known about the infrequent bonded interactions, so those calculations are allocated to a much more flexible subsystem that will allow experimentation with various “force field” models and algorithms.

What might go wrong with the Anton effort? Subtle errors arising from the class of force fields that Anton is designed to handle efficiently may accumulate over the extremely long MD runtimes; in a custom machine with no operational experience, soft errors could strike much more often and substantially slow its performance; quantum effects may turn out to be necessary, beyond the classical force field being modeled here; some clever graduate student may come up with a

software-based approach that reduces Anton's two-orders-of-magnitude performance advantage to only one (which might no longer be enough to justify its hardware expenditure). Or Anton might become a victim of its own success if early learnings point to much better (and much different) MD algorithms that no longer fit well into Anton's overall structure.

But what if things go *right*? Benoit Roux, an MD researcher now at the University of Chicago, said that as soon as Anton has delivered its first verified scientific result he will want an engine of his own, and so will everyone in the entire MD field. Roux points out that molecular biologists must normally have “elaborate strategies to prevent fooling themselves” in their macro-scale experiments. With Anton, “we'll be able to do insane things with unknown problems and two weeks later we'll discover how the molecules actually move. ... Anton will revolutionize molecular biology.”

It is not often that a science reaches a clear tipping point—when it advances very quickly, virtually exploding into a new shape and venue. Our own field of computing has done that several times. Many physicists expect this of the Large Hadron Collider currently being completed in Europe. Shaw and his coworkers are attempting nothing less in the field of molecular dynamics. As a computing professional, I am proud of their efforts, I salute their attempt to drive an extremely important basic science forward, and I heartily recommend their paper.

Bob Colwell (bob.colwell@comcast.net), former chief architect of Intel's IA-32 microprocessors, is now an independent consultant.