

TERASCALE COMPUTING IN ACCELERATOR SCIENCE AND TECHNOLOGY*

R. Ryne, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA
K. Ko, Stanford Linear Accelerator Center, Menlo Park, CA 94025 USA

Abstract

We have entered the age of “terascale” scientific computing. Processors and system architecture both continue to evolve; hundred-teraFLOP computers are expected in the next few years, and petaFLOP computers toward the end of this decade are conceivable. This ever-increasing power to solve previously intractable numerical problems benefits almost every field of science and engineering and is revolutionizing some of them, notably including accelerator physics and technology. At existing accelerators, it will help us optimize performance, expand operational parameter envelopes, and increase reliability. Design decisions for next-generation machines will be informed by unprecedentedly comprehensive and accurate modeling, as well as computer-aided engineering; all this will increase the likelihood that even their most advanced subsystems can be commissioned on time, within budget, and up to specifications. Advanced computing is also vital to developing new means of acceleration and exploring the behavior of beams under extreme conditions. With continued progress it will someday become reasonable to speak of a complete numerical model of all phenomena important to a particular accelerator.

1 HOW FAR WE’VE COME

The challenge of scientific computing tools has been elegantly summarized:

With the advent of everyday use of elaborate calculations, speed has become paramount to such a high degree that there is no machine on the planet today capable of satisfying the full demand of modern computational methods. The most advanced machines have greatly reduced the time required for arriving at solutions to problems which might have required months or days by older procedures. This advance, however, is not adequate for many problems encountered in modern scientific work...

That quotation would be at home in an article about the latest massively parallel cluster, but it

came from the 1947 patent application for ENIAC, which at 4×10^9 TFLOP was among the wonders of its age even without the ability to store programs. [1] You may literally have more computing power in your wristwatch than existed in the world back then; the advent of the integrated-circuit microprocessor brought us into the domain of “Moore’s Law,” which, in its high-performance-computing (HPC) formulation, observes that performance is increasing roughly a hundredfold per decade. The only thing increasing faster than the power of scientific computers is the scope of their users’ dreams.

The desire—in many situations, the need—for ever-more-comprehensive, faithful, and detailed modeling and visualization will always create a demand for faster processing, bigger storage, and broader connectivity. Nonetheless, terascale computing—which provides trillions of floating-point operations per second, trillions of bytes of data management capability, etc.—enables us to do important things that were unfeasible just a few years ago. At the intersection of the accelerator and the HPC communities are users ready to take advantage of this capability.

2 THINGS BECOME DIFFERENT ON THE TERASCALE

Accelerator science is a diverse field—a modern accelerator is a microcosm of many-body dynamics and electromagnetic technology—so it presents a variety of computational challenges. They fall into three broad categories: problems requiring high accuracy, problems of scale, and discovery opportunities in accelerator and beam physics. (The new-physics goals of accelerator users are also highly dependent on HPC, but that is beyond the scope of this paper.)

2.1 High Accuracy

Accelerators include many components whose designs are critical in detail as well as in overall aim. A prime example is the three-dimensional electromagnetic and thus the physical design of rf components, which in some cases must achieve a frequency accuracy of 1 part in 10,000. Computer modeling and computer-aided engineering are not only helpful but necessary. Figure 1 shows an example: an rf acceleration

cavity with higher-order-mode damping for the PEP-II low-energy ring.

The upper two figures show 3-D models of the cavity. The upper left figure shows a model based on a structured mesh, and the upper right shows a model based on an unstructured formulation. The use of unstructured grids and multiprocessor computer systems allows for the high-accuracy simulations that are needed to guide the cooling channel design. The lower row of figures, from left to right, shows adaptively refined grids that were used to obtain convergence of the predicted wall loss distribution.

This cavity proved highly successful in bench measurements and in PEP-II. Future accelerators, including the NLC, neutrino factories, muon colliders, etc., can benefit from HPC for solving electromagnetic problems that require extreme accuracy in complicated geometries and ability to deal with large-scale problems in a reasonable time.

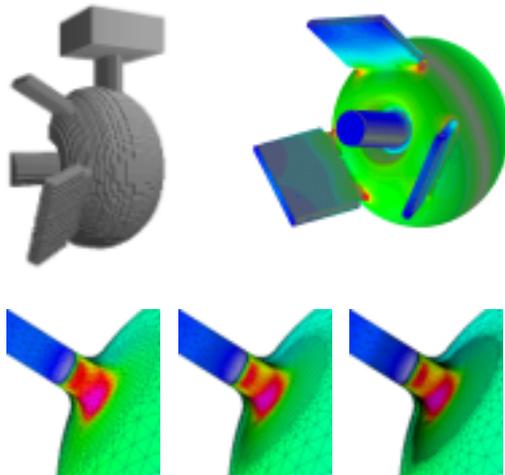


Figure 1: RF acceleration cavity with higher-order-mode damping for PEP-II. Shown are structured (*top left*) and an unstructured (*top right*) 3-D models, and an adaptively refined region showing the wall loss (*bottom row*).

2.2 Problems of Scale

High-performance computing comes into its own in classes of problems whose scale has thus far made them intractable. A prime recent example is the first-ever 3-D Fokker-Planck simulation, performed in 2000. One author claimed as recently as 1997 that it would be “completely impractical... in terms of number of particles, computation time, and statistical

fluctuations to actually compute [the Rosenbluth potentials] as multiple integrals.” [2] Thinking that this was impossible, researchers had previously resorted to approximate calculations. Now the feasibility of a first-principles approach that yields self-consistent results has been demonstrated on highly parallel systems at LBNL’s National Energy Research Supercomputing Center (NERSC) and LANL’s Advanced Computation Laboratory. [3]

High-intensity beam dynamics is another class of problems of scale important in accelerator R&D. In the 1980s, this typically involved two-dimensional simulations with 10^4 particles. By the early 1990s, serial simulations in two dimensions were still the norm, though the state of the art for the number of particles (which, along with the number of grid points, governs the resolution of the simulations) was $10^5 - 10^6$. Today it is *routine*—an overnight job for a 256-processor parallel supercomputer—to simulate the beam dynamics of an rf linac in three dimensions with 10^8 particles.

It is well worth noting that HPC (generally implying massive parallelism in this context) is necessary rather than merely convenient. Ordinary single-processor personal computers have made impressive progress within their realm, but as beam-dynamics simulation engines, even the fastest of them have not moved much beyond that early-90s level. That 10^8 -particle 3-D simulation would take months of CPU time on a personal computer if it were possible at all.

The next level of problems of scale is exemplified by 3-D simulation of intense beams in rings such as the Fermilab Booster, the Alternating Gradient Synchrotron at Brookhaven, the Proton Storage Ring at Los Alamos, and the accumulator that will be built for the Spallation Neutron Source at Oak Ridge. Compared to linac simulations like those described above, these problems are at least 100 to 1000 times more computationally challenging. There are several reasons. The particles circulate for many turns, making the accelerator significantly “longer” from a simulation standpoint. There are additional beam phenomena to be modeled than in an ion linac (e.g., wakefield effects). Further, some of the familiar problems are more challenging to model than they are in a linac; for example, calculating the self-fields is more difficult for a long beam with a high longitudinal-to-transverse aspect ratio than for a beam made up of shorter bunches.

Some problems span these genre boundaries. Figure 2 illustrates a problem of both accuracy and scale in a challenging future accelerator: the rf acceleration structures in the Next Linear Collider (NLC). The structure is not only large, but also quite complicated in three dimensions.

At the lowest level the structure consists of cells whose 3-D shape—determined by an 11-parameter optimization procedure—maximizes shunt impedance while controlling potentially disruptive wakefields. The geometrical complexity is further complicated by the addition of a damping manifold that is used to reduce long-range wakefields and improve vacuum conductance. The individual cells must be accurate to $\pm 0.01\%$ in the accelerating frequency, requiring very high resolution modeling. (The upper-portion portion of Figure 2 shows a close-up view of one quadrant of the mesh used to model 1.5 cells of the structure, with the actual prototype shown next to it. The mesh model also displays decomposition of the geometry domain for parallel processing).

Beyond the need for high accuracy in individual cells, longer structures consisting of many cells must be modeled to predict dipole frequencies in order to verify wakefield suppression. The upper

right portion of Figure 2 shows a schematic of the input section, consisting of the input coupler together with a front portion of the actual structure. A six-cell stack was fabricated on a computer-controlled milling machine, and the measured frequencies of the lowest three dipole bands were found to be in excellent agreement with the parallel simulations. Calculations for full structures, which may involve 50 to 200 cells, have now begun, thus necessitating very-large-scale simulations.

The lower portion of Figure 2 shows a 47-cell model. Such simulations are challenging in part because the associated spectrum of modes is extremely dense. However, as illustrated at the bottom of the figure, using HPC it is now possible to achieve convergence for one of the dipole modes in the actual structure, showing, for the first time, self-consistent fields in both the main accelerating cavity and the damping manifold.

Calculations such as these would have been impossible without the availability of HPC hardware and research in the associated software and algorithms.

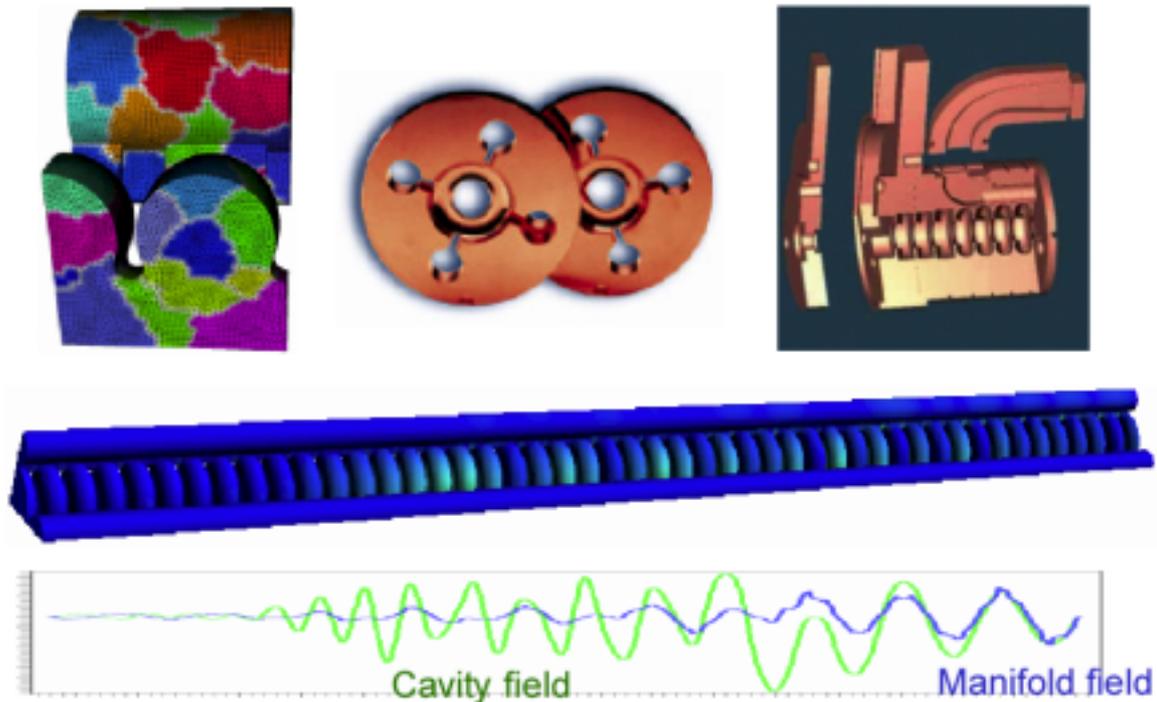


Figure 2. The accelerating structures at the NLC and other future colliders will be geometrically complex as well as large, and even at the level of individual cells their shapes will have important subtleties, so problems of both accuracy and scale will face us in HPC.

2.3 New Physics

Physics may well be entering a new golden age in which exciting discoveries are made that change the way we view the universe. Advances that could change the way we accelerate particles are also in the works, and HPC is vital to them. An example is plasma wakefield acceleration (e.g., Figure 3), which is being investigated at several institutions. Extremely high accelerating gradients—up to a thousand times greater than those achievable with conventional rf-based technology—have been produced in laboratory-scale experiments, but only across very short distances. As researchers attempt to scale up these experiments and someday produce useful accelerators, they must link multiple stages and learn how to obtain and preserve beam quality.

The science and technology challenges are great, but the potential payoff is even greater because accelerators could become much smaller and less expensive. This would be tremendously beneficial both at the energy frontier, where it is imperative to control the size and cost of future generations of colliders, and in applied research, where many scientific disciplines and industrial applications would welcome affordable “tabletop” GeV-class accelerators. To achieve this promise will require synergistic interaction of theory, experiment, and high-performance computing.

3 WHY HPC IS NOT ONLY ADVANTAGEOUS BUT IMPERATIVE

It is inevitable that the accelerator community will embrace HPC for problems of precision and scale that are beyond the reach of desktop and midrange computing. HPC modeling and simulation are the ante these days in complicated, high-cost works of technology; one would no longer design a skyscraper or a jetliner or a race car without the most thorough simulation that one could afford.

As an example of the benefit that simulation can give us, compare the Superconducting Super Collider with today’s efforts. When the SSC was designed in the mid to late 1980s, gigaFLOP performance was impressive and the kinds of simulation we take for granted these days ranged from difficult to impossible. At one point, questions arose about whether a 4-cm beampipe aperture (with its immediate implications for

magnet size) would suffice in the collider rings. Without the ability to definitively answer the question on the scale of a 52-mile-circumference ring, project management felt compelled to increase the aperture to 5 cm. This change added a billion then-year US dollars to a project whose cost was already problematic. By contrast, improvements in accelerator-structure design, like those described earlier, are expected to save an estimated \$100 million in a notional Next Linear Collider by increasing efficiency and thus decreasing the need for rf power sources and all the components associated with them.

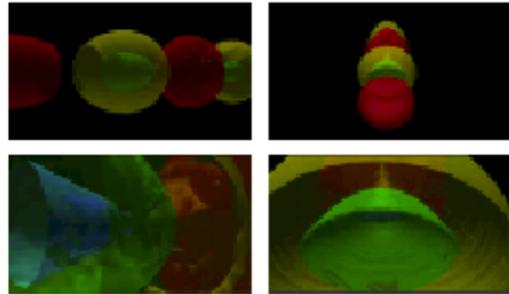


Figure 3. Simulation and visualization combine in this rendering of isosurface contours in a plasma-wakefield acceleration structure performed by a UCLA/USC collaboration.

Today’s machines (and, to an even greater extent, tomorrow’s) present challenges for the accelerator designer and in turn for HPC. Some of these problems are well known but take on new importance in higher-energy or higher-intensity machines; others have only recently been encountered. Space charge, beam-beam effects, various instabilities, halos, collisions, multi-species effects, surface effects, ionization, trapped modes, “dark current,” and wakefields are among the phenomena that must be accurately modeled today. Simulation also aids the technology as well as the physics of accelerators, the search for intense hot spots in an rf structure being a prime example.

4 SOME SPECULATIONS ON THE FUTURE OF HPC AND ACCELERATORS

Though there are great challenges ahead, the future looks bright for continued performance gains in HPC. Extrapolating from the TOP500

list, [4] we have some confidence in the ability to reach the petaFLOP regime by the end of the decade. However, advances in most or all aspects of high-performance computers will be required in order to achieve that level of performance.

Increased parallelism with higher-performance processors, faster interprocessor communication, more-efficient system architectures (perhaps implying more problem-specific architectures with memory and mass storage tailored to the problem in question), and new algorithmic concepts are all necessary parts of the answer. Given progress in these areas, it seems reasonable that the HPC version of Moore's Law will be observed for 10 or perhaps 20 more years.

This author's crystal ball is not nearly so clear on the development of energy-frontier accelerators. There are no approved machines beyond the LHC, nor even consensus on how we might surpass its equivalent-physics energy range. It is apparent that the next generation of HEP accelerators will be formidable in size, cost, and technological challenges. But whatever shape they take, HPC will be vital to them.

To take the best advantage of HPC's capabilities, it must be treated not merely as a tool, nor as something done in a different department, but as a partner (Figure 4). This model is integrated into a DOE Office of Science program called SciDAC. [5]

Under SciDAC, computer scientists, applied mathematicians, and other information technology experts sponsored by the Office of Advanced Scientific Computing Research are partners with physical scientists (both theoretical and applied) who are funded by other DOE program offices. One SciDAC project, whose physical-science sponsor is the Office of High Energy and Nuclear Physics, has the goal of developing a comprehensive terascale accelerator-modeling capability. [6]

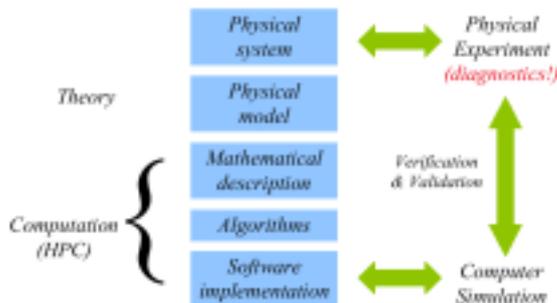


Figure 4. Theory, experimentation, and computation are synergistic when HPC is properly integrated into an R&D program.

5 CONCLUSIONS

Given the complexity and importance of accelerators, it is imperative that the most advanced HPC tools be brought to bear on their design, optimization, commissioning, and operation. Continuing the exciting progress of accelerator physics and technology into a new century will require performance assurance (sometimes in the face of unexplored parameter regimes or even new physics) and also cost containment. Whether they are based on novel principles or advanced applications of existing technology, we must know that these machines will work properly, and we must keep them affordable. HPC is a powerful new tool for solving both these problems.

6 REFERENCES

- [1] See <http://www.computerhistory.org>
- [2] Wallace M. Manheimer, Martin Lampe, and Glenn Joyce, *J. Comp. Phys.* Vol. 138, No. 2, pp. 563-584 (Dec. 1997).
- [3] J. Qiang, S. Habib, and R. Ryne, "Self-Consistent Langevin Simulation of Coulomb Collisions in Charged-Particle Beams," in *Proceedings of the IEEE/ACM SC2000 Conference*, IEEE Press Order #PE00899/ACM Order #415000 (2000).
- [4] See <http://www.top500.org>
- [5] DOE's SciDAC program is described at <http://www.science.doe.gov/scidac/>
- [6] See http://www-afrd.lbl.gov/scidac_accel.pdf

7 ACKNOWLEDGEMENTS

The authors are grateful for the assistance of many colleagues in the preparation of this manuscript and the thinking that led to it. Let us especially thank David Bailey and Horst Simon of LBNL's National Energy Research Scientific Computing Center. The Computer Museum History Center website was invaluable as we prepared the PAC presentation. The writing of this paper was aided by Joe Chew of LBNL.

This work was funded by the Office of Science of the U.S. Department of Energy under Contract. No. DE-AC03-76SF0098. Support came through the Office of High Energy and Nuclear Physics and the Office of Advanced Scientific Computing Research, as well as through the Office of Science's sponsorship of the National Energy Research Scientific Computing Center (NERSC), whose resources were used in this effort.