GE Global Research
Richard Arthur
Director – Advanced Computing

Position Paper to National Research Council

Committee on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science in 2017-2020
HPCD Workshop, Dallas, TX, January, 2015

ADVANCED COMPUTING AT GENERAL ELECTRIC

GE is a world leader in high-tech industries including Aviation, Energy, Healthcare, Water, Lighting, Industrial Controls and Financial Services. To remain globally competitive in such industries requires expert use of leading-edge technology not only in the products themselves, but also in their design, manufacture, operation, and servicing. Advanced computational modeling is a key technology particularly vital in the design of highly complex systems and the analytics guiding operational performance and reliability. GE has made substantial investments in internal software and computing infrastructure, as well as building strong partnerships to leverage academic, commercial and government facilities and expertise in advanced computing and modeling.

RESEARCH NEEDS/OPPORTUNITIES

Advances in computing hardware performance and development of clever algorithms to tame complexity in processing and data have revolutionized the practice of many areas of science and engineering. This serendipity experienced by fields such as astrophysics, genomics, image processing, computational fluid dynamics, and electromagnetics is not universal, however. Some physical phenomena elude development of deterministic models, relying on methods ranging from stochastic approximations to empirical study. Additionally, many practical problems of interest to industry rely upon integration of multiple physical systems and/or phenomenological scales which until recently have been impractical to implement in models due to computational complexity. But even as hardware evolved to such capability, there remain gaps in the scientific understanding of these multi-physics interactions and thus also the computational modeling software.

For example, materials science continues to present challenges in application of computational modeling in areas of great importance to industry such as corrosion, fatigue, and new materials discovery. Owners of high-value assets such as airplanes, power plants, and medical imaging place a premium on availability and performance. Composition of maintenance schedules and work scope to minimize the costs of planned and unplanned down time relies on factors such as how accurately the health of parts can be predicted. Predictive models of part degradation caused by the asset's operation and environment could be improved through better understanding physical mechanisms underlying processes such as oxidation, creep, and fracture mechanics in alloys. Additionally, the modest progress in model-based discovery of new candidate materials for applications such as turbomachinery, superconducting magnets and energy storage elucidates the need for further investment, pragmatic focus and innovation in programs like the Materials Genome Initiative.

Ultimately the utility of computational modeling to industrial product and service decision-makers is limited by the uncertainty of the numerical results (quantified or not.) To meet product performance and quality demands, designs must compensate for uncertainty in the predicted results of computational models. Thus, in the many subjects where it is impractical or impossible to apply first principles, the rigor employed in physical verification of a model and assigning confidence bounds on reference data and predicted results directly impact real engineering costs and operational performance limitations. Confounding physical verification itself, the state of the art in some fields has evolved beyond the capability of measurement at requisite time granularity, temperature extremes, spatial resolution needed for high-fidelity synthetic mesh geometries, etc.

ADVANCED COMPUTING CAPABILITIES, FACILITIES, REQUIREMENTS

Arguably the most significant impact of Moore's Law on the computing used by industry has been the emergence of the abundant and affordable computing and storage pervasively accessible over the network. This infrastructure is more often associated with "The Cloud" and "Big Data and Analytics" but precisely the same cost and capability advantages manifest in high-performance and leadership computing. Even simply focusing on commodity hardware, this infrastructure that has emerged in the last five years compels rethinking the feasibility of simplifying strategies such as brute force in place of heuristic solutions.

Effectively all algorithms that have prospered in computer science efficiently map to the computer architecture devised by the genius polymath John von Neumann in 1945. Until very recently, the novel capabilities affording opportunity to rethink algorithms were limited to exploiting concurrency through blending of message passing, fine-grained threading, and vectorization. But several non-von Neumann architectures have emerged as commercial processors in the past year. This affords the computer science community a fresh opportunity to revisit shelved algorithms of the past as well as to discover entirely new algorithms that advantageously exploit these architectures (e.g., neuromorphic, processor-in-memory, CPUs coherent with programmable fabric, etc.)

As the costs in capital, operations, and specialized knowledge of high-end computing centers make the products of their use by definition high value assets, these data are also thus targets for theft and information warfare. Safeguards should be put in place not only to protect data from pilfering, but also to ensure continuity of service of computing centers themselves and the integrity of what is processed on the system. Malicious or otherwise unauthorized manipulation of data (reference, input sets, result sets) or solvers (source code, scripts, binaries, libraries) resulting in loss of validity (and thus credibility) of computational methods could be significant – particularly if used in design or operation of critical infrastructure. Fortunately paradigms for establishing trust and integrity have matured in the safety-critical and military contexts and could be adopted in measure.

CHALLENGES AND SUGGESTIONS

The nature of "barrier problems" to science and industry deserves reconsideration. The highest-end computational grant programs of both the NSF and ASCR strongly bias toward scientific breakthrough and discovery. Industry by definition is concerned with reduction to practice, and some pragmatic barriers to exploit the scale afforded by today's leadership computing platforms would not merit "glossy cover story" prestige. For example, the sheer effort generating input data such as validated high-fidelity mesh geometries to Peta+ scale solvers can be tremendously challenging and expensive. Advancing the tools and techniques for mesh generation and validation and offering massively parallel resources simply for their creation would provide substantial value and advance industrial use of leadership-class computational modeling. Similarly, rigorous verification of models against reference sets or regression suites can also be computationally intense and yet a valuable enabler toward adoption and advancement of computational modeling in industry.

While the discussion of means to support scientific community codes themselves is meritorious, the problems with performance portability will persist, as roadmaps from hardware vendors show no sign of convergence. Perhaps this burden now shared by all developers of scalable software begs rethinking the workflow and tool chain employed. One idea would be to restructure the process of compilation, linking, and loading to be more adaptive, employ machine learning techniques, and interact with the human developer. The use of language directives to a batch-form utility could be replaced with a dialogue between the machine and human. The patterns and antipatterns for performance in increasingly complex, heterogeneous architecture configurations could incorporate trial-and-error, evolutionary methods leveraging the strengths of both machine and human.

Last, as massive data become increasingly expensive to replicate, the resources under discussion could become authoritative repositories for ever-improving reference sets (possibly under export license) relating to verified physical dynamics, materials properties, geodata such as weather, etc. In addition to essentially infinite storage, the network performance should meet the needs for both continuous ingest of new data as well as the search and retrieval necessary to maintain the vitality of the data and user community, which should include not only academic, industrial and government research, but also regulatory agencies, manufacturing, etc.