SAND2011-1505P

Chemistry Exascale Co-design Center (CECC)

Oak Ridge National Laboratory, R.J. Harrison* (Director), E. Aprà and D. Hudson; **Pacific Northwest National Laboratory,** W.A. de Jong, S. Krishnamoorthy, K. Kowalski, and E.J. Bylaska; **Argonne National Laboratory,** R.A. Bair and J.R. Hammond; **Los Alamos National Laboratory,** W.M. Challacombe; **National Renewable Energy Laboratory,** R.E. Larsen and K. Kim; **Sandia National Laboratories,** C.L. Janssen (Co-director); **SLAC National Accelerator Laboratory,** J. Nørskov and F. Abild-Pedersen

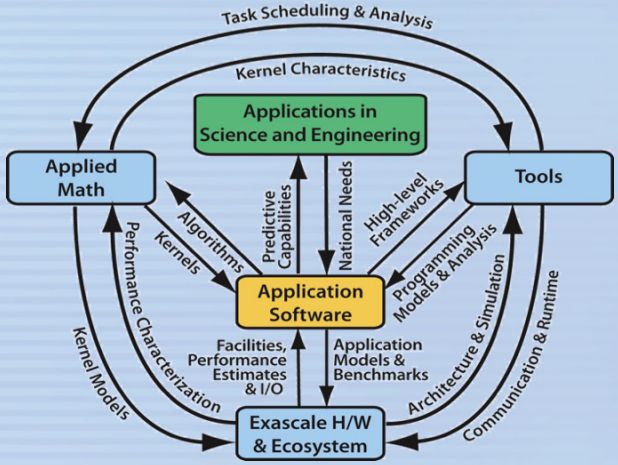
Academic Collaborators: L. Mitas (North Carolina State University), P. Sadayappan and A. Rountev (The Ohio State University), S.B. Baden (University of California, San Diego), E. Deumens (University of Florida), W-M. Hwu and M. Snir (University of Illinois at Urbana-Champaign), J. Dongarra and G. Basilca (University of Tennessee), M. Mavrikakis (University of Wisconsin-Madison), T.D. Crawford and E.F. Valeev (Virginia Tech)

Center co-PIs will be finalized upon DOE funding decision.

The mission of the Chemistry Exascale Co-design Center (CECC) is to establish exascale simulation as a practical and valued tool for discovery and design in the chemical sciences with emphasis on applications in energy production and storage.

We will accomplish this in intimate collaboration with technology developers, system vendors, and other stakeholders through co-design of all elements of the vertically integrated stack (processor, memory, interconnect, I/O, system and application software, numerical algorithms, data structures, programming models and tools, and reformulation of science objectives). High-consequence objectives are radical advances in scientific productivity, programmer productivity, sustained performance, price performance, and energy utilization. We will replace current chemical simulation capabilities that are very limited and inappropriate for exascale, with new tools and algorithms that treat with robust fidelity the full complexity and dynamics of physical systems. Finally, we plan our software and underlying programming models and tools to provide a solid and open foundation for *innovation by others and ourselves* using post-petascale and exascale simulation. Thus, the eventual capabilities of the code will not be limited by the scope of the current project, and through proactive community engagement and outreach we strive for broad and long-lived impact.

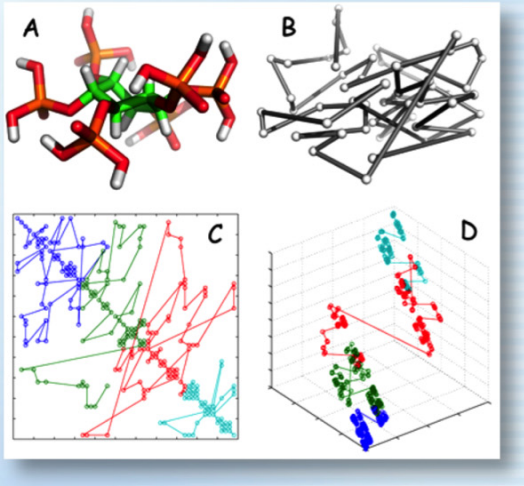
* *harrisonrj@ornl.gov*



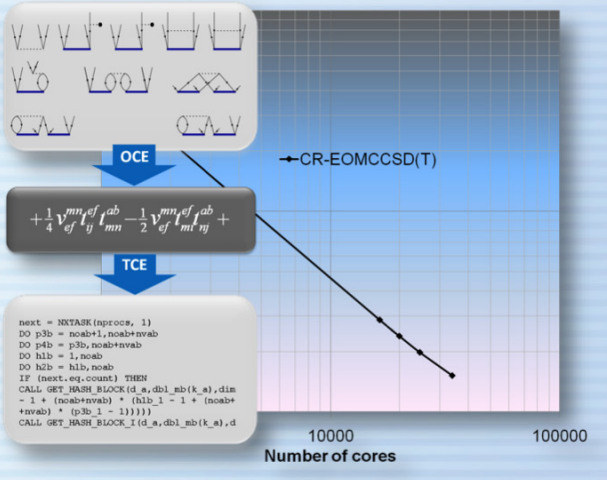
The CECC co-design process

Special Concerns of CECC

- Emphasis on auto-generation from high-level specifications (DSLs) , e.g. kernels (DFT functionals), libraries (integrals), and entire parallel applications (tensor engine). This is essential for scientific productivity, facilitates rapid retargeting and prototyping, injecting instrumentation, and resilience.
- One-sided messaging (RMA and RMI) to support globally addressable sparse data structures and dynamic task-based runtime
- FLOP rich kernels for small/block-sparse arrays
- Fast algorithms that maintain both speed and precision for exascale simulation



Quantum locality can be exploited for data- and load-balancing via space-filling curves, from atoms (A-B) through matrices (C) to the product space (D).



The Tensor Contraction Engine is a domain-specific language for transforming quantum theory into high-performance code

Application Software

Our vision for the exascale is to deliver simulation capabilities that rival experiment in scientific discovery. We will do this through development and deployment of advanced theories and methods, relying heavily on advancements in computer science and pervasive co-design. Methods with computational cost scaling nearly linearly with system size are essential to realize the full potential of exascale simulation through a one-to-one relationship between simulation capability and processing power. *Indeed, while exascale hardware will provide a 1000x speedup in moving from the peta to exascale, we anticipate equal or greater speedup from these new methods that are presently not deployed in any scalable chemistry code.*

Applied Math

Exascale computers will open a new frontier in the scale of both the size of the simulation and the amount of resulting data. Re-evaluation of existing algorithms and the development of novel numerical methods that are reliable, accurate and scalable will be central to our work. End-to-end error analysis of simulation results will be used to control numerical errors and give us confidence in our results. Reduced rank representations will compress our data sets reducing the memory footprint and bandwidth utilization. Data consistency error detection and resilient algorithms will permit efficient utilization of resources in the face of machine faults.

Tools

The significant complexity of the needed science models, coupled with the rapidly increasing difficulty of developing high-performance applications on emerging heterogeneous, multi-level parallel architectures, makes it imperative to develop advanced software development environments that enhance the productivity of developers and the performance of applications. Using the Tensor Contraction Engine as a starting point, the application co-design effort will develop domain-specific representations for use by code transformation tools to generate platform-specific optimized code. Extended run-time capabilities will be used to adapt code execution to a specific system instance and to dynamically schedule work while minimizing global synchronization operations. Performance analysis tools will provide insight into the runtime characteristics of the simulations, and resilience capabilities will provide the end-to-end framework for handling faults. Together, the CECC tools infrastructure will enable computational chemistry applications to make effective use of future exascale class platforms.

Hardware and Exascale Ecosystem

The CECC is structured to have a significant and sustained impact on the exascale hardware and software ecosystem. Team members will be embedded in the Leadership Computing Facilities at Argonne and Oak Ridge National Laboratories and with both exascale technology tracks. Modeling and simulation are a major component of the CECC and will provide the bridge between application software, runtime systems, and hardware platforms well before fully functional components are available.

***The time is right for a completely new generation of codes in chemistry*, and through active dissemination of results, outreach and education, and distribution of open-source software, the CECC aims to transform large elements of entire communities. In particular, we aim to eliminate the gulf between theoretical innovation in small groups and its realization on high-end computers. Many of our computer science tools and related capabilities are highly relevant to other application domains, and we have an explicit agenda of inter-center and inter-disciplinary engagement.**