Chemistry and Biochemistry

NWChemEx: Tackling Chemical, Materials, and Biomolecular Challenges in Exascale

A major goal of DOE's advanced biofuels program is to develop fuels that can be distributed using the existing infrastructure and replace existing fuels on a gallon-for-gallon basis. However, producing high-quality biofuels in a sustainable and economically competitive way is technically challenging, especially in a changing global climate. The design of feedstock for the efficient production of biomass and the design of new catalysts for the efficient conversion of biomass-derived intermediates into biofuels are two major science challenges involved in the development of advanced biofuels.

The NWChemEx project directly addresses a Priority Goal in DOE's 2014–2018 Strategic Plan, namely, developing high-performance computational "models demonstrating that biomass can be a viable, sustainable feedstock" for the production of biofuels and other bioproducts. In addition to providing the means to resolve these biofuel challenge problems, NWChemEx will enable exascale computers to be applied toward the development of new materials for solar energy conversion and next-generation batteries, simulation of chemical processes in combustion, prediction of transport and sequestration of energy by-products in the environment, and design of new functional materials.

The NWChemEx project is redesigning and reimplementing NWChem for pre-exascale and exascale computers. NWChemEx is based on NWChem, an open-source, high-performance parallel computational chemistry code funded by the DOE Biological and Environmental Research (BER) program that provides a broad range of capabilities for modeling molecular systems. NWChemEx will support a broad range of chemistry research important to DOE BER and Basic Energy Sciences (BES) on computing systems ranging from terascale workstations and petascale servers to exascale computers.

In Phase I (FY17–20), the NWChemEx project will develop high-performance, scalable implementations of three major physical models: (1) Hartree-Fock (HF) and Density Functional Theory (DFT) methods. These methods are the foundations for the physical models to be incorporated in the NWChemEx framework. Their implementation must be significantly revised to simulate the large molecular systems in the targeted science challenges on exascale computers.

(2) Coupled Cluster methods. A suite of canonical and reduced-scaling coupled cluster (CC) methods will be implemented in NWChemEx. These methods are the gold standard in electronic structure theory and provide the level of fidelity required to address the targeted science challenges.

(3) Density Functional Embedding Theory. Embedding techniques provide a natural and mathematically sound basis for seamlessly integrating subsystems with different electronic structure representations, enabling the active site of interest to be described with high-accuracy CC methods while using a lower fidelity method to describe the impact of the environment on the molecular processes in the active site.

In Phase II (FY21–23), the NWChemEx project will extend NWChemEx to enable it to model the elementary chemical processes involved in the catalytic conversion of propanol to propene in the H-ZSM-5 zeolite. These new capabilities will focus on developing the physical models and algorithms needed to characterize the reactants, products, intermediates, and transition states involved in the conversion process. Two benchmarks have been established for NWChemEx, one at the end of each of the two phases.

Phase I. To illustrate the performance of NWChemEx on biomolecular systems, the ubiquitin molecule was selected as the benchmark for assessing the performance of NWChemEx on ORNL's pre-exascale Summit computer. Ubiquitin is a protein molecule typical of many biomolecular molecules with an abundance of experimental data available from it as well as its fragments. Although it will not be feasible to run canonical coupled cluster calculations on ubiquitin, a 1,231-atom molecule, it will be possible to run reduced-scaling, CC calculations on it. The availability of both implementations along with the sequence of ubiquitin fragments will enable any inaccuracies in the reduced-scaling method to be identified and corrected.

Phase II. A number of elementary chemical transformations have been postulated for the conversion of propanol to propene in the H-ZSM-5 zeolite (basic unit cell: $Si_{96}O_{192}$) based on low-level calculations combined with the available experimental data. To illustrate the performance of the Phase I and II capabilities in NWChemEx on the two exascale computers to be deployed in FY22–23 at Argonne (Aurora) and ORNL (Frontier), reduced-scaling, CC calculations will be used to redefine the structures and energetics of the postulated elementary steps in the conversion of propanol to propene. Depending on the outcome of these calculations, additional work may be required to more fully characterize the mechanism of this conversion.

PI: Thom Dunning, Pacific Northwest National Laboratory

Collaborators: Pacific Northwest National Laboratory, Brookhaven National Laboratory, Ames Laboratory, Argonne National Laboratory, Lawrence Berkeley National Laboratory, Oak Ridge National Laboratory, Virginia Polytechnic Institute and State University

Progress to date

- Designed and implemented a simulation development environment as the framework for initiating and connecting modules. This component includes memoization and caching of results to decrease redundant computations, increase support for scripting, and enable application programming interfaces (APIs) for I/O.
- Implemented the LibChemist module that supports core data associated with quantum chemistry codes (e.g., molecular and basis set information).
- Completed the initial design and implementation on GPUs of the TAMM module for both dense and sparse tensor operations.
- The initial design and implementation of a number of the modules in NWChemEx, e.g., canonical HF, DFT, and CC, have been completed.

NWChemEx is implementing exascale algorithms for computational chemistry that will solve two DOE-recognized grand challenge problems to enable the development of advanced biofuels.