

**Allocation:** Innovation & Exploration/525 knh

Gregory Bauer

University of Illinois at Urbana-Champaign

Biology, Chemistry & Health

IMPROVING NWCHEM SCALABILITY USING THE DATASPACE FRAMEWORK

Research Challenge

The present computational challenge is handling a combinatorial explosion in problem size that accompanies the task of identifying the global minimum in parameter space. The solution is to employ parallelization; however the frequent read / write operations of the large number of parallel processes can overload the filesystem. The aggregate distributed array holds billions of records for an average parameter optimization problem. Sorting an array of such size to determine the promising parameter sets for further analysis represents a practically unsolvable problem that requires a creative solution. Another traditional problem encountered in parameter optimization efforts is the limited availability of training data. Extracting unique molecular clusters from high-resolution X-ray crystals offers a powerful solution but requires the use of computationally very expensive electronic structure computations to generate the corresponding data set.

Methods & Codes

- * Target data for parameter optimization is generated by a set of in-house scripts.
- * Electronic structure computations employ previously optimized CCSD(T) method in the NWChem package.
- * Developed a scalable tool to optimize Lennard-Jones parameters in the AMBER (Assisted Model Building with Energy Refinement) force field.

Results & Impact

This project introduces a procedure for systematic improvement of classical force field by determining the global minimum in the parameter space for an expandable set of the training data. The beneficiary of the optimized parameter set is the entire molecular dynamics community. The developed fractional parallel sorting procedure drastically reduces time spent in sorting as well as the required RAM per node.

Why Blue Waters

With its fast interconnect and large memory per core, Blue Waters is unique in its ability to conduct CCSD(T) computations of molecular systems encountering a thousand basis functions, which is vital for the success of the developed parameter optimization procedure. Since the parameter optimization procedure is extremely resource demanding, the availability of large numbers of nodes is essential for the exhaustive exploration of parameter space.