

High Accuracy Predictions of the Bulk Properties of Water (ESP Technical Report: February, 2013)

ALCF-2 Early Science Program Technical Report

Argonne Leadership Computing Facility

About Argonne National Laboratory

Argonne is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC under contract DE-AC02-06CH11357. The Laboratory's main facility is outside Chicago, at 9700 South Cass Avenue, Argonne, Illinois 60439. For information about Argonne and its pioneering science and technology programs, see www.anl.gov.

Availability of This Report

This report is available, at no cost, at <http://www.osti.gov/bridge>. It is also available on paper to the U.S. Department of Energy and its contractors, for a processing fee, from:

U.S. Department of Energy

Office of Scientific and Technical Information

P.O. Box 62

Oak Ridge, TN 37831-0062

phone (865) 576-8401

fax (865) 576-5728

reports@adonis.osti.gov

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor UChicago Argonne, LLC, nor any of their employees or officers, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of document authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, Argonne National Laboratory, or UChicago Argonne, LLC.

High Accuracy Predictions of the Bulk Properties of Water (ESP Technical Report: February, 2013)

ALCF-2 Early Science Program Technical Report

prepared by
Maricris Mayes, Graham Fletcher, Yuri Alexeev
Argonne Leadership Computing Facility, Argonne National Laboratory

May 7, 2013

ESP Technical Report: February, 2013

High Accuracy Predictions of the Bulk Properties of Water

Staff at ALCF

Maricris Mayes (postdoctoral student)
Graham Fletcher (catalyst group member)
Yuri Alexeev (catalyst group member)

Executive Summary

Water is the most important liquid to humanity and a deep understanding of its behavior is of critical importance to national priority scientific issues such as global warming. Despite this, accurate theory and modeling of water is lacking, and the question of the detailed mechanism behind its bulk properties remains one of the outstanding unsolved problems in science. While many simulation studies of liquid water have been undertaken, resource and capability limitations have forced the vast majority to be performed using empirical model potentials with questionable reliability.

This project is the first to simulate water with high accuracy and high precision, employing recently developed advanced techniques that allow first principles *ab initio* quantum mechanical methods to harness hundreds of thousands of processors and scale up to thousands of water molecules.

Overview of numerical methods

A partial integro-differential equation based on Quantum Mechanics - called the Schroedinger Equation - is solved for the electronic structure of a molecular system in a basis of atom-centered functions. Dense linear algebra is involved. Integration over all space is performed analytically, and this consumes the majority of the execution time. Inter-atomic forces are derived from the electronic structure and used to drive dynamical simulations that yield bulk property predictions.

The project code is the General Atomic and Molecular Electronic Structure System (GAMESS), maintained by the Gordon Group at Iowa State University. We employ the Fragment Molecular Orbital (FMO) method in GAMESS together

with the Restricted Hartree-Fock (RHF) and Moeller-Plesset second order perturbation theory (MP2) levels of theory.

What is enabled by Mira over Intrepid

Following on from a 2010 INCITE award on *Intrepid*, this Early Science project on *Mira* focusses on the bulk properties of water at higher levels of accuracy using larger basis sets, larger embedded clusters, and longer dynamics simulations. While benchmark tests were performed on *Intrepid*, the greater computational capability available with *Mira* is critical to establishing *convergence* of the bulk properties with respect to theory, basis set, cluster size, and simulation length.

Code Modifications, new algorithms: RATTLE

This section describes new functionality - an implementation of the RATTLE method - that has been coded especially for the current Early Science Project by Yuri Alexeev of ALCF. This development stands to significantly reduce the cost of the project, to one half or one quarter of its original cost. The motivation for and impact of RATTLE are as follows.

Energy conservation is critical to any dynamical simulation carried out in a series of discrete time steps. With current technology, energy conservation imposes a maximum time step of around 0.3 femtoseconds. In addition, systems with large numbers of hydrogen atoms (such as water) can waste many steps where the trajectory is trapped within local minima corresponding to the vibrational modes of all the hydrogens.

One way to guide the trajectory through the most important regions of phase space, avoid local minima and, in so doing, permit larger steps, is to apply constrained dynamics. Today, almost all classical molecular dynamics simulations employ constraints, particularly where hydrogen atoms are involved. The most popular methods include SHAKE, RATTLE, and SETTLE. RATTLE is the most accurate because the constraints are applied to both distances and velocities simultaneously. RATTLE is a Lagrange-multiplier-based method developed especially for Velocity Verlet, a method for integrating Newton's equations implemented in GAMESS.

Figure 1 shows an example of 2 water molecules 8 Angstroms apart at 0.5fs intervals (time is plotted along the x-axis, and total energy along the y-axis), already the curve with RATTLE (green) is smoother than without (red).

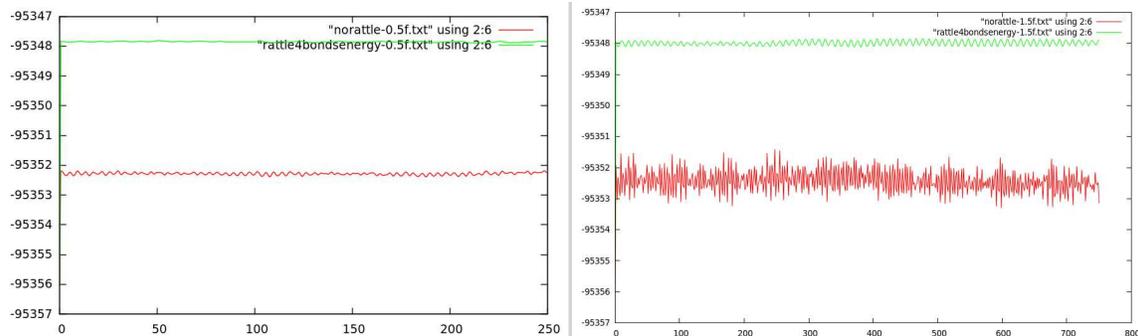


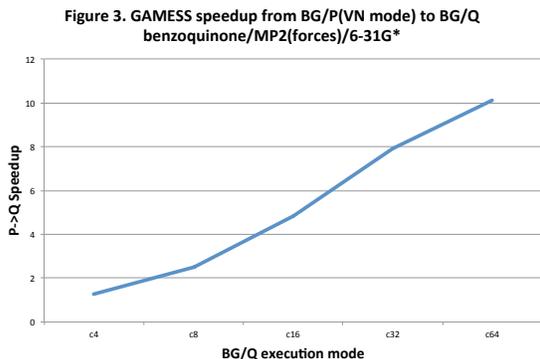
Figure 2, at 1.5fs intervals, shows that without RATTLE (red) the variation in energy can be large and erratic, while with RATTLE (green) the energy is much more stable. In this example, RATTLE allows 2-4x longer time steps to be taken, and we expect similar improvements to be obtained with larger systems. With longer time steps, proportionally fewer steps need to be taken, reducing the overall cost of the simulations.

As it smooths out the trajectory, RATTLE also improves overall efficiency by removing any abrupt shifts in molecule positions that may hinder solution of the Schrodinger Equation for each water. Thus, RATTLE allows fewer, larger, more efficient and reliable time steps to be taken.

Performance on Mira

In this section, the performance on Mira is examined in three ways: first, a comparison with BG/P, second, for increasing relevant problem sizes, and lastly with BG/Q execution mode.

Figure 3 shows the speedup from BG/P to BG/Q of a calculation of the energy and forces at the MP2 level for the benzoquinone molecule with a 6-31G* atomic basis set. The relative improvement on a BG/Q node can be seen to rise monotonically with the execution mode compared to the 'VN' mode on BG/P, reaching a maximum speedup of x10.



The reasons for such a large improvement are several. BG/Q benefits not only from 4x more processor cores per node, and a 2x higher clock rate, but also from the ability in its microkernel to multi-rank the cores (not available on BG/P) in c32 mode, giving two MPI ranks per

core. This is of great benefit to GAMESS, allowing its 'data-server' processes to reside on the same core as the compute processes and, overall, utilize the cores more efficiently. An additional higher execution mode, c64, with four MPI ranks per core, gives the maximum speedup but the resultant amount of memory available to processes is too small (being somewhat less than 250 megabytes each) to be useful in the rest of the current project.

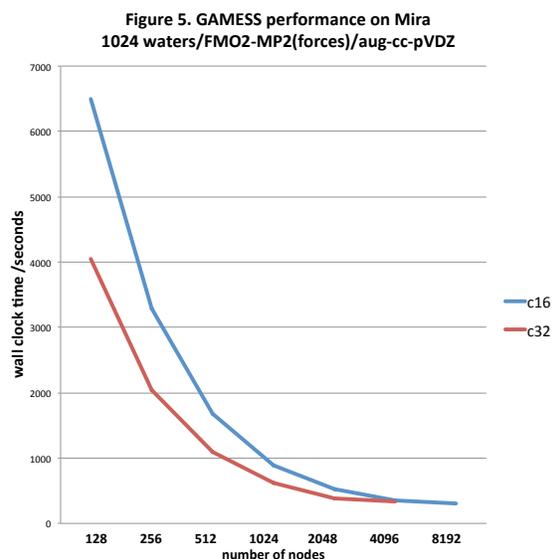
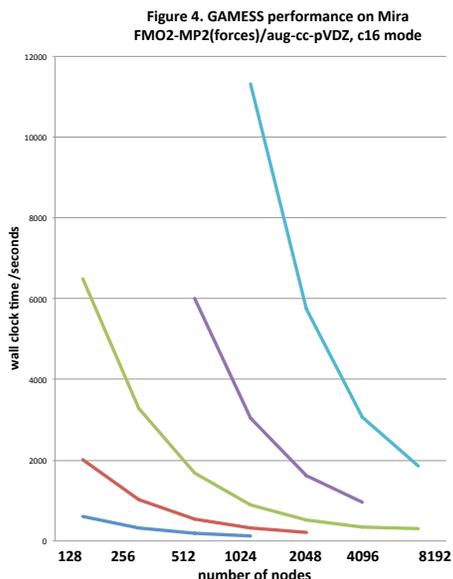


Figure 4 shows the scalability of FMO-MP2 calculations on 128 to 4096 waters with a 6-31G* atomic basis set, in c16 mode. As with c64 mode, some of the larger calculations cannot be accessed from c32 mode due to memory limitations, and c16 allows the greatest range of problem sizes that is useful in the current project to be examined for scalability. It can be seen how scalability improves with problem size, being poorest for the smallest problem (128 waters) and best for the largest problem (4096 waters). In quantum chemistry, where the computational cost is approximately cubic with the problem size while the communication overhead is quadratic, scalability usually improves with problem size.

Figure 5 compares the scalability of the 1024 water calculation from Figure 4 for two execution modes, c16 and c32. It can be seen that while c32 mode is always faster, especially at the lowest node count, the two curves converge as the number of nodes increases. While c32 mode is expected to be more efficient for reasons described above, the convergence with c16 mode toward higher node counts is most likely the result of increased I/O contention as the overall number of ranks increases. We are pursuing the latter hypothesis in our investigation of this phenomenon.



Argonne Leadership Computing Facility

Argonne National Laboratory
9700 South Cass Avenue, Bldg. 240
Argonne, IL 60439

www.anl.gov



Argonne National Laboratory is a U.S. Department of Energy
laboratory managed by UChicago Argonne, LLC