Liquid Water: Obtaining the Right Answer for the Right Reasons

Edoardo Aprà  Computer Science and Mathematics Division Oak Ridge National Laboratory Oak Ridge, TN aprae@ornl.gov
Robert J. Harrison  Computer Science and Mathematics Division Oak Ridge National Laboratory Oak Ridge, TN harrisonj@ornl.gov
Alistair P. Rendell  Department of Computer Science Australian National University Canberra, Australia alistair.rendell@anu.edu.au
Wibe A. deJong  Environmental Molecular Sciences Laboratory Pacific Northwest National Laboratory Richland, WA wibe.deJong@pnl.gov
Vinod Tipparaju  Computer Science and Mathematics Division Oak Ridge National Laboratory Oak Ridge, TN tipparaju@ornl.gov
Sotiris S. Xantheas  Chemical and Materials Sciences Division Pacific Northwest National Laboratory Richland, WA sotiris.xantheas@pnl.gov
Ryan M. Olson  Cray, Incorporated 380 Jackson St. Suite 210 St. Paul, MN ryan@cray.com

ABSTRACT

Water is ubiquitous on our planet and plays an essential role in several key chemical and biological processes. Accurate models for water are crucial in understanding, controlling and predicting the physical and chemical properties of complex aqueous systems. Over the last few years we have been developing a molecular-level based approach for a macroscopic model for water that is based on the explicit description of the underlying intermolecular interactions between molecules in water clusters. In the absence of detailed experimental data for small water clusters, highly-accurate theoretical results are required to validate and parameterize classical interaction potentials. As an example of the benchmarks needed for the development of accurate models for the interaction between water molecules, for the most stable structure of the tetrakaidecahedron (T-cage) (H2O)24 cage we ran a coupled-cluster calculation on ORNL’s Jaguar petaflop computer that used around 150 TB of memory for a sustained performance of 1.39 PetaFLOP/s (double precision) on 223,200 processors, lasting for three hours.

The importance of water in nature due to its role as a universal solvent and its significance in sustaining life motivates the development of an extensive knowledge of its properties under various conditions and diverse chemical and biological environments [42]. Important aqueous systems of immediate interest to mankind include the enhancement of functions performed by many biological molecules in general and in particular in systems such as ion channels [24], heterogeneous reactions occurring in polar stratospheric cloud surfaces [22], and catalytic processes targeted towards environmentally benign fuels [43]. Central to a fundamental understanding of processes occurring in these complex aqueous environments is the ability to describe the properties of water at the molecular level (such as in water clusters, in hydrate lattices, in bulk water under ambient and supercritical conditions, and the various forms of crystalline and amorphous ice) as well as its interaction with a variety of molecules, and its behavior at interfaces.

Despite the fact that a universal model for water aiming at describing its anomalous properties [44–46] has been sought for over half a century, such a model has yet to be developed. Most of the existing interaction potentials for water have been parameterized in order to reproduce a subset of its macroscopic structural and thermodynamic properties using classical molecular dynamics (MD) simulations. An alternative philosophy to those effective potentials is the development of classical models that are transferable across different environments. These rely on the accurate description of the Born-Oppenheimer potential energy surface (PES) and the explicit account of the full vibrational motion of the molecular assemblies.

An important component of this approach is the avail-
ability of binding energies of water clusters of various sizes. Clusters of water molecules offer the advantage of probing the relevant interactions at the molecular level and they provide a quantitative picture of the nature and magnitude of the various components of the intermolecular interactions such as exchange, dispersion and induction. Although structural and spectral properties of small water clusters have been recently obtained [10,29,31–33,39], their interaction energies have yet to be measured experimentally. This much-needed information is currently solely obtained from high-level electronic-structure calculations [18,47,48]. To this end, a family of flexible, polarizable classical potentials for water (TTM, TTM2-F, TTM3-F) [4,6,7,20,21] has been parametrized from high-level calculations of selected parts of the water dimer PES [5] and has been shown, in conjunction with nuclear quantum dynamical statistical simulations, to reproduce most of the macroscopic structural, thermodynamic, and transport properties of liquid water [19,37].

This approach has relied on the availability of accurate binding energies of water clusters that are converged with respect to the level of electron correlation and the orbital basis set in an attempt to quantify the magnitude of the many-body interactions in water clusters and to provide a benchmark for the absolute cluster binding energies up to n=20. So far these have been obtained at the second-order perturbation level of theory (MP2, [35]). These previous calculations of the various hydrogen bonded networks present in the families of isomers of (H\textsubscript{2}O)\textsubscript{20} have resolved a long standing issue regarding their relative stability. For the (H\textsubscript{2}O)\textsubscript{20} cluster, based on the arrangement of the oxygen atom network, there are 4 major families of minima as shown in Figure 1: (a) dodecahedron, (b) fused cubes, (c) face-sharing pentagonal prisms and (d) edge-sharing pentagonal prisms. Interest in the relative stabilities of the various networks of these neutral clusters stems from their importance in understanding the transition from the open 3-fold to the compact 4-fold coordinated structures. The relative energetic order of the most stable isomer of each family is predicted from ab-initio calculations to be: (d) (most stable), (c), (b), (a) (less stable). Their corresponding MP2/CBS (Complete Basis Set) binding energies are: -200.1 kcal/mol (dodecahedron), -212.6 kcal/mol (fused cubes), -215.0 kcal/mol (face-sharing pentagonal prisms), -215.0 kcal/mol (edge-sharing pentagonal prisms).

Clusters of water molecules offer the advantage of probing the relevant interactions at the molecular level and they provide a quantitative picture of the nature and magnitude of the various components of the intermolecular interactions such as exchange, dispersion and induction. Although structural and spectral properties of small water clusters have been recently obtained [10,29,31–33,39], their interaction energies have yet to be measured experimentally. This much-needed information is currently solely obtained from high-level electronic-structure calculations [18,47,48]. To this end, a family of flexible, polarizable classical potentials for water (TTM, TTM2-F, TTM3-F) [4,6,7,20,21] has been parametrized from high-level calculations of selected parts of the water dimer PES [5] and has been shown, in conjunction with nuclear quantum dynamical statistical simulations, to reproduce most of the macroscopic structural, thermodynamic, and transport properties of liquid water [19,37].

This approach has relied on the availability of accurate binding energies of water clusters that are converged with respect to the level of electron correlation and the orbital basis set in an attempt to quantify the magnitude of the many-body interactions in water clusters and to provide a benchmark for the absolute cluster binding energies up to n=20. So far these have been obtained at the second-order perturbation level of theory (MP2, [35]). These previous calculations of the various hydrogen bonded networks present in the families of isomers of (H\textsubscript{2}O)\textsubscript{20} have resolved a long standing issue regarding their relative stability. For the (H\textsubscript{2}O)\textsubscript{20} cluster, based on the arrangement of the oxygen atom network, there are 4 major families of minima as shown in Figure 1: (a) dodecahedron, (b) fused cubes, (c) face-sharing pentagonal prisms and (d) edge-sharing pentagonal prisms. Interest in the relative stabilities of the various networks of these neutral clusters stems from their importance in understanding the transition from the open 3-fold to the compact 4-fold coordinated structures. The relative energetic order of the most stable isomer of each family is predicted from ab-initio calculations to be: (d) (most stable), (c), (b), (a) (less stable). Their corresponding MP2/CBS (Complete Basis Set) binding energies are: -200.1 kcal/mol (dodecahedron), -212.6 kcal/mol (fused cubes), -215.0 kcal/mol (face-sharing pentagonal prisms), -215.0 kcal/mol (edge-sharing pentagonal prisms).

Figure 1: The 4 major families of minima of (H\textsubscript{2}O)\textsubscript{20} and the most stable networks (#308 and #316) for (H\textsubscript{2}O)\textsubscript{24}.

are needed in order to obtain fully quantitative agreement and hence to arrive at a fully predictive model potential for liquid water.

In this study we report the results of CCSD(T) calculations of the binding energies of various isomers of (H\textsubscript{2}O)\textsubscript{20} and (H\textsubscript{2}O)\textsubscript{24} [49] to obtain definitive values for the cluster energetics. From prior work on the water dimer [3,9,23,36] we anticipate the CCSD(T) level of theory being essentially converged with respect to many-body effects, though we also plan future calculations up to the CCSDTQ level of theory to quantify this convergence. The computational cost of CCSD(T) grows as the 7th power of the system size (in contrast MP2 scales as the 5th power) and the inverse 4th power of the required accuracy, therefore these calculations are extremely demanding. Hence, another objective of our calculations is to test/validate reduced-scaling methods (e.g., [1,26]) that are applicable to yet larger systems.

In the remainder of this paper, we describe the CCSD(T) method in section 2, while we report the performance and scientific results in section 3 and 4, respectively.

2. ALGORITHM DESCRIPTION

Coupled-cluster (CC) theory is a numerical many-body technique that incorporates the effect of electron correlation on the electronic structure of molecular systems. CCSD(T) is one of several CC variants that estimates the effect of electron correlations by considering single, double and triple excitations; single and double excitations are fully computed with a self-consistent approach, while the contribution of the triple excitations is computed perturbatively [40]. Here we
describe the CCSD(T) module of the NWChem library, its usage of the Global Arrays and discuss how Global Arrays library is structured on the Cray XT5.

2.1 The CCSD(T) module in NWChem

Valence only coupled cluster CCSD(T) [40] calculations have been called the “gold standard of quantum chemistry” [16] for their chemical accuracy in determining molecular energetics. The computational cost of CCSD(T) calculations formally scales as the seventh power of the number of basis functions ($N^7$), hence this is a method that could effectively utilize large supercomputers. A few parallel implementations of CCSD(T) are available [2, 25, 27, 30, 34], however, most of them have been designed either for clusters or for moderate scale parallelism.

In this work we have substantially modified and enhanced the original parallel implementation of CCSD(T) in NWChem of Kobayashi and Rendell [30] that was designed to effectively utilize massively parallel processors and to make minimal use of I/O resources. The focus of the performance numbers reported here is the perturbative triples correction implemented in NWChem following the “$aij kbc$ algorithm” of Rendell and coworkers [41] that makes no use of I/O by storing intermediate quantities (two-electron integrals and coupled-cluster wave function amplitudes) in the global memory managed by Global Arrays.

This use of aggregate memory for storing integrals distinguish the Kobayashi-Rendell implementation from, for example, the Janowski-Pulay algorithm [27] that stores the integrals on the local disks available on clusters. The largest requirements for local memory of the “$aij kbc$ algorithm” are of size $n_v n_o^2$ ($n_o$ being the number of occupied orbitals and $n_v$ being the number of virtual orbitals); this might lead to relative large quantities in excess of 0.5GB for the water clusters here studied, but, since the loop over the $n_o$ occupied orbitals can be split in multiple passes, the memory requirements can be reduced up to $n_v^2$. This $n_v^2$ scaling can potentially become the bottleneck for larger molecules and larger basis sets; the Janowski-Pulay algorithm using local buffer of size $n_v^2$ might be a better option.

The floating-point intensive kernel of this algorithm is a series of calls to the BLAS DGEMM [12] matrix multiplication routine. Several data parallel and one-sided Global Arrays operations are used in the CCSD(T) implementation of NWChem, most notable of these are the $\text{ga}_{\text{get}}()$ and $\text{ga}_{\text{acc}}()$ calls that are used to get/accumulate sections of distributed array.

Our revisions to this algorithm emphasized increased locality to reduce communication and implemented a more careful tiling of intermediates to reduce memory consumption and increase parallelism and load balance. The dynamically load balanced algorithm explicitly considers three levels of the memory hierarchy. It proceeds by tiling the full computation so that intermediate results fit in available global memory, then tiling the nested loops so that data associated with each task fits into local memory. A process access a global shared counter to determine the next task, moves data from global to local memory using the $\text{ga}_{\text{get}}()$ operation, computes, and accumulates results into the result using the $\text{ga}_{\text{acc}}()$ operation. The local computation employs the BLAS DGEMM [12] matrix multiplication routine to optimize for the local memory hierarchy.

2.2 The Global Arrays library on the Cray XT5

The primary component of the Global Arrays library is the distributed arrays layer. This layer does all the necessary translation from the users shared memory styled access to the actual distributed array. It has several components that it uses: a Message Passing library, the ARMCI one-sided communication library and a memory allocator (MA library). The Distributed Arrays layer is mostly implemented on top of the ARMCI one-sided communication library.

2.2.1 Message Passing Library

A message passing library is used in the Global Arrays for process management and for some collective communication. For the Cray XT5 system, we used the MPI library from the Cray Message Passing Toolkit. Version 3.1 of the Cray MPT uses an interface called PMI (process management interface) for scalable process management. This interface also allows for saving space by keeping one copy of the Network-ID/Process-ID information that is necessary for communication. For the CCSD(T) module that we discuss here, the amount of MPI usage is minimal, most of the communication is via ARMCI one-sided communication library using the non-contiguous one-sided communication interfaces it provides. Each of the cores on the octacore XT5 runs an MPI task.

2.2.2 ARMCI one-sided communication library

The ARMCI communication library has interfaces for contiguous, strided and vector communication along with Read-Modify-Write operations. All of the Global Array one-sided calls are translated into ARMCI communication calls. The ARMCI library, underneath, uses the fastest possible mechanism to transmit data. It uses shared memory within the node and, on the Cray XT5 system, uses Portals library for inter-node communication. The Global Arrays library is optimized to overlap intra-node data transfers in shared memory and inter-node data transfers in the network using non-blocking calls. On several networks, ARMCI utilizes “good” communicatabile memory by avoiding any extra copies when transmitting from such memory. Hence it provides memory allocation interfaces which can allocate communicatable memory. On the Cray XT5 system, the Portals library can be used to transmit from any address in the processes virtual memory address space, therefore special memory allocation routines are not required. On the Cray XT5, in addition to $n$ MPI computing tasks (where the value of $n$ can be up to 12), ARMCI spawns a thread to assist with communication (particularly Read-Modify-Write and non-contiguous data transfer operations). This might result in a slight over-subscription, but is necessary to achieve good communication performance.

2.2.3 The Memory Allocator library

The MA library provides dynamic memory allocator routines with debugging and verification support for use by C, FORTRAN, or mixed-language applications. Typically MA layer allows for use of ARMCI memory allocator (by setting an environment variable) but on the Cray XT5 system, this is not necessary. NWChem is written in Fortran77 and exploits the dynamic memory allocation provided by MA in its Fortran code.
3. PERFORMANCE RESULTS

We report performance number by using as base the parallel implementation of CCSD(T) in NWChem of Kobayashi and Rendell [30] that was designed to effectively utilize massively parallel processors and to make minimal use of I/O resources.

Previous CCSD(T) runs with the same NWChem implementation achieved a performance of 6.3 TFLOP/s using 1400 processors [38] on a cluster of Itanium2 processor that uses a Quadrics QnemII network, while more recent runs at PNNL utilized 14000 processors on an Infiniband network of Opteron processors [11]. What distinguishes the benchmark numbers reported here is the unprecedented scale of the calculations and floating-point performance achieved. We run a series of benchmark with the 5.1 version of NWChem [8] using the water cluster (H₂O)₄₄ and a modified cc-pvtz basis set [15] for a total of 1224 Gaussian basis functions¹; wall-time of the CCSD(T) runs as a function of processor number are reported in Figure 2. The last datapoint at 223,200 processors reached a sustained 64-bit floating-point performance of 1.39 PetaFLOP/s. We made use of the PAPI library [13] to compute the Flop/s rate reported in this paper; more specifically we used the PAPIf_flips subroutine to extract the floating-point operations and of the MPI call MPI_Wtime to evaluate the Wall-clock time.

All twelve cores of the twelve-core Cray XT5 node were used in the performance results reported here. In order to alleviate memory usage and improved load-balancing, we had to modify the distribution of the larger Global Arrays used in this algorithm. The effect of last two distribution approaches are shown in the plot² of Figure 3 for the (H₂O)₁₈ benchmark.

We have modified the 5.1 NWChem source code in two main aspects: 1) we moved communications outside of inner loops (sometime at the additional cost of more use of local memory) and 2) we have modified the distribution of the largest Global Arrays used in the triples kernel of the CCSD(T) method. The first, more naïve distribution we attempted (labeled as distribution #1 in the plot), allowed us to scale at 70,000 processors, but it was hitting a wall after 80,000 processor. This first distribution was already an improvement compare to the previous one since the previous one would have required 2GB or more of memory (clearly not an option on a system that has only 2GB of RAM on each core). In the next step, we refined the distribution by having the large Global Arrays being distributed over as many processors as possible; the finer granularity of this second distribution improved load-balancing and allowed us to reach better scaling beyond 80000 processors.

For the larger (H₂O)₂₄ benchmark (reported in Fig. 2), we ran a coupled-cluster calculation on the ORNL’s Jaguar petaflop computer that used around 150 TB of memory for a sustained performance of 1.39 PetaFLOP/s (double precision) in the triples sections on 223,200 processors, lasting for little than 3 hours. This floating-point performance corresponds to 60% of theoretical aggregate performance of the 223,200 2.6GHz AMD Opteron cores. The number of basis functions for this run is equal to 1224 and the number of correlated orbitals is equal to 96. While the 1.39 PetaFLOP/s refers just to the triples section of the calculation, if we consider the run from beginning to end (including I/O operations such as reading the 43GB amplitudes file plus the initial step of transforming the Atomic Orbitals into Molecular orbitals), the floating point results is of 1.27 PetaFLOP/s.

It is important to point out that 64-bit floating precision is a necessary requirement for this kind of calculations; this is due to the vast number of integrals being accumulated and to the high accuracy gaussian basis functions giving rise to linear dependence problems.

The one-sided nature of Global Arrays makes the entire resource of the entire petascale class machine available to a single calculation; this is in contrast with other programming modules (used in recent Gordon Bell award papers) that resort to make use of more cumbersome techniques (e.g. processor groups type parallelism) in order to achieve large scale parallelism.
4. SCIENTIFIC RESULTS AND CONCLUSIONS

At the time of writing we can unambiguously state that the CCSD(T) results being generated by NWChem are correct. So far we have obtained results for two of the several $(H_2O)_{20}$ isomers and the two lowest energy networks of the $(H_2O)_{24}$ isomers with a modified triple zeta quality (cc-pVTZ minus the f-functions) basis set. For concreteness these data are reported in Table 1. However, to draw quantitative conclusions we must complete the calculations on the remaining structures, conduct a study to assess the impact of basis truncation, and finally validate the conclusions by application to other structures including $(H_2O)_{24}$ (we report some initial results in Table 2). It is this system that we anticipate will deliver our highest performance numbers.

As noted in the introduction our approach will also enable the assessment of errors in reduced-scaling methods, and the method of Hirata [26] is particularly relevant. System stability during the early access period limits our current throughput, but this is constantly improving and by summer we anticipate substantial progress towards our scientific goals and will also update these results herein.

<table>
<thead>
<tr>
<th>Edge-sharing Dodecahedron</th>
<th>$E^{\text{tot}}$</th>
<th>B.E.</th>
<th>$E^{\text{tot}}$</th>
<th>B.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCF</td>
<td>-1521.371595</td>
<td>152.87</td>
<td>-1521.363070</td>
<td>147.52</td>
</tr>
<tr>
<td>MP2</td>
<td>-1526.467996</td>
<td>251.02</td>
<td>-1526.444136</td>
<td>236.05</td>
</tr>
<tr>
<td>CCSD</td>
<td>-1526.539945</td>
<td>228.14</td>
<td>-1526.518107</td>
<td>214.43</td>
</tr>
<tr>
<td>CCSD(T)</td>
<td>-1526.709594</td>
<td>245.10</td>
<td>-1526.684946</td>
<td>229.64</td>
</tr>
</tbody>
</table>

Table 1: Total Energies ($E^{\text{tot}}$) in hartrees and Binding energies (B.E.) in Kcal/mol for the edge-sharing pentagonal prism and the dodecahedron $(H_2O)_{20}$ structures

<table>
<thead>
<tr>
<th></th>
<th>$E^{\text{tot}}$</th>
<th>B.E.</th>
<th>$E^{\text{tot}}$</th>
<th>B.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCF</td>
<td>-1825.633230</td>
<td>175.48</td>
<td>-1825.633269</td>
<td>175.50</td>
</tr>
<tr>
<td>MP2</td>
<td>-1831.739872</td>
<td>283.83</td>
<td>-1831.739898</td>
<td>283.84</td>
</tr>
<tr>
<td>CCSD</td>
<td>-1831.822044</td>
<td>257.52</td>
<td>-1831.822066</td>
<td>257.53</td>
</tr>
<tr>
<td>CCSD(T)</td>
<td>-1832.022670</td>
<td>276.02</td>
<td>-1832.022676</td>
<td>276.03</td>
</tr>
</tbody>
</table>

Table 2: Total Energies ($E^{\text{tot}}$) in hartrees and Binding energies (B.E.) in Kcal/mol for two $(H_2O)_{24}$ structures

Our results will provide the “gold standard” – as regards the accuracy - in the binding energies of water clusters up to $n=30$. These are very important, in lieu of the fact that even for the smaller water cluster, the water dimer, there currently does not exist an experimentally measured binding energy.

Our computations will provide definitive results for the significant science question of the structure of liquid water and in particular the associated inter-molecular potentials. The top performing run to date sustained a performance of 1.39 PetaFLOP/s (double precision) on 223,200 processors and around 150 TB of memory. This is a unique performance for computational chemistry applications using a correlated method and is the result of several man-decades of work on Global Arrays, NWChem, and associated algorithms. Also unique is the use of the one-sided Global Arrays programming model that enabled us to scale a single tightly-coupled calculation, rather than many weakly coupled process subgroups. As a forward looking statement, we can predict that algorithm here describe will perform even better on the Cray Gemini network, due to appear in future Cray XT product lines. If we compare Gemini with the current Seastar network, Gemini will result in slightly better latency, less network congestion issues and will enable applications to have some percentage of overlapping of communication and computation; all these characteristics will reduce the walltime to solution of these NWChem runs by decreasing the time spent in communication.

We expect that the CCSD(T) cluster binding energies up to $n=30$ will provide the necessary information needed to re-parametrize the TTM3-F interaction potential for water and obtain the measured enthalpy of liquid water to within a few hundredths of a kcal/mol from quantum statistical mechanical simulations. Our preliminary results for the binding energy of $(H_2O)_{20}$ and $(H_2O)_{24}$ indicate that the CCSD(T) numbers differ from the MP2 ones by about 2.5 to 2.75%, a range that is almost the same between experiment and the computed enthalpy of liquid water based on the MP2 parametrization. It is the combination of the availability of highly accurate electronic structure data for the binding energies of water clusters as the result of the current effort in combination with the potential development that will ultimately lead in obtaining “the right answer for the right reasons” for water, nature’s most important substance.

5. ACKNOWLEDGMENTS

With great sadness, we note the passing of Jarek Nieplocha, a dear friend and colleague, without whom none of this work would have been possible.

This work was supported by the Department of Energy, offices of Basic Energy Science and Advanced Scientific Computing Research as part of the SciDAC program, and was also in part sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL), managed by UT-Battelle, LLC for the U.S. Department of Energy under contract DE-AC05-00OR22725 with Oak Ridge National Laboratory. This research used resources of the National Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Part of this work was supported by the Chemical Sciences, Geosciences and Biosciences Division and the Materials Science and Engineering Division, Office of Basic Energy Sciences, US Department of Energy. Battelle operates the Pacific Northwest National Laboratory for the U.S. Department of Energy.

We would like to thank all the people at Cray Inc who gave valuable suggestions during this work and in particular Kitrick Sheets for his help in addressing several issues we faced along the way. His essential changes to the ARMCI Portal specific code.

6. REFERENCES


