Annual Report for Blue Waters Allocation

- **Project Information**
  - Blue Water Professorship
  - So Hirata, University of Illinois at Urbana-Champaign
  - sohirata@illinois.edu

- **Executive summary (150 words)**

We develop and apply two highly scalable and systematically accurate electronic structure methods. One is the embedded-fragment method for various structural, thermodynamic, and spectroscopic properties of molecular crystals, amorphous solids, and liquids. The other is the Monte Carlo many-body perturbation and Green's function methods as well as grid-based diffusion Monte Carlo method, running efficiently on thousands of CPUs or hundreds of GPUs or more. With Blue Waters, we perform grand-challenge large-scale applications of both methods.

- **Description of research activities and results**

  - *Key Challenges:* Our group have made two breakthroughs in computational chemistry for large molecules, solids, and liquids. (1) One is the method that allows such high-level calculations to be applied to an infinitely extended molecular crystals, amorphous solids, and liquids, including nature's most important solids and liquids such as ice and liquid water. It divides them into fragments embedded in their electrostatic field and treat these fragments by well-developed molecular theories and software in a highly parallel algorithm. (2) The other weds systematic (i.e., *ab initio*) many-body electronic structure methods with quantum Monte Carlo, enabling massively parallel, systematically accurate electronic structure calculations for large molecules and solids. With Blue Waters, we aim to perform grand-challenge applications to demonstrate these methods' full capacity. The applications may include (1) computationally explaining the thermodynamic and phase properties of crystalline and amorphous ices as well as structural and spectroscopic properties of liquid water as well as (2) near-exact calculations of ionization energies and electron affinities of large conjugated molecules, important for advanced organic materials for optoelectronic devices.

  - *Why it Matters:* Project (1) advances geochemistry, astrophysics, and planetary science where probing high-pressure phases of ices of atmospheric species on Earth or other planets are important but experimentally difficult and expensive. The systems studied by project (2) include solids that serve as bases of advanced materials such as bulk heterojunction organic solar cells, batteries, sensors, smart windows, field-effect transistors, and light emitting diodes. The optoelectronic parameters are the quantities of prime importance determining the solids performance and functions, but the usual density-functional approximations are known to be poor for these properties. Here, our new method is uniquely useful

and accurate. The impact of this research is broadly in the area of energy science and technology.

- *Why Blue Waters:* Today's workhorse computational methods for solids (density-functional methods) and liquids (classical molecular dynamics) are routine on a small computer cluster, but with limited accuracy. Therefore, the most meaningful use of Blue Waters in this area is to enhance the accuracy rather than enlarge the system size (which is already formally infinite). In electronic structure theory, this means switching from density-functional methods and empirical force fields to *ab initio* theories, which solve the fundamental equation of motion of chemistry rigorously and in systematic approximations with controlled errors. However, conventional matrix-algebra algorithms of *ab initio* theories are fundamentally non-scalable and extremely expensive. The aforementioned combination of *ab initio* theories with quantum Monte Carlo and the embedded-fragment methods are among the few that may be realistically and usefully deployable on the large number of processors available on Blue Waters, enabling formerly unthinkable predictively accurate calculations for large systems.

- *Accomplishments (2020):*

  We introduced two convergence acceleration schemes for Monte Carlo many-body perturbation and Green's-function methods.

  In the Monte Carlo many-body perturbation (MC-MP) method, the conventional correlation-correction formula, which is a long sum of products of low-dimensional integrals, is first recast into a short sum of high-dimensional integrals over electron-pair and imaginary-time coordinates. These high-dimensional integrals are then evaluated by the Monte Carlo method with random coordinates generated by the Metropolis–Hasting algorithm according to a suitable distribution. The latter algorithm, while advantageous in its ability to sample nearly any distribution, introduces autocorrelation in sampled coordinates, which, in turn, increases the statistical uncertainty of the integrals and thus the computational cost. It also involves wasteful rejected moves and an initial "burn-in" step as well as displays hysteresis. Here, an algorithm is proposed that directly produces a random sequence of electron-pair coordinates for the same distribution used in the MC-MP method, which is free from autocorrelation, rejected moves, a burn-in step, or hysteresis. This direct-sampling algorithm is shown to accelerate second- and third-order Monte Carlo many-body perturbation calculations by up to 222% and 38%, respectively. This work was published in *J. Chem. Phys.* and was selected as JCP Editor's Pick.

  The use of many control variates is proposed as a method to accelerate the second- and third-order Monte Carlo (MC) many-body perturbation (MC-MP2 and MC-MP3) calculations. A control variate is an exactly integrable

function that is strongly correlated or anti-correlated with the target function to be integrated by the MC method. Evaluating both integrals and their covariances in the same MC run, one can effect a mutual cancellation of the statistical uncertainties and biases in the MC integrations, thereby accelerating its convergence considerably. Six and thirty-six control variates, whose integrals are known a priori, are generated for MC- MP2 and MC-MP3, respectively, by systematically replacing one or more two-electron-integral vertices of certain configurations by zero-valued overlap-integral vertices in their Goldstone diagrams. The variances and covariances of these control variates are computed at a marginal cost, enhancing the overall efficiency of the MC-MP2 and MC-MP3 calculations by a factor of up to 14 and 20, respectively. This work was published in *J. Chem. Phys.*

A diffusion Monte Carlo algorithm is introduced that can determine the correct nodal structure of the wave function of a few-fermion system and its ground-state energy without an uncontrolled bias. This is achieved by confining signed random walkers to the points of a uniform infinite spatial grid, allowing them to meet and annihilate one another to establish the nodal structure without the fixed-node approximation. An imaginary-time propagator is derived rigorously from a discretized Hamiltonian, governing a non-Gaussian, sign-flipping, branching, and mutually annihilating random walk of particles. The accuracy of the resulting stochastic representations of a fermion wave function is limited only by the grid and imaginary-time resolutions and can be improved in a controlled manner. This grid-based diffusion Monte Carlo method is tested for a series of model problems including fermions in a harmonic trap as well as the He atom in its singlet or triplet ground state. For the latter case, the energies approach from above with increasing grid resolution and converge within 0.015 E_h of the exact basis-set-limit value with a statistical uncertainty of 10^{-5} E_h without an importance sampling or Jastrow factor. This work was published in *Phys. Rev. E*.

- **List of publications and presentations associated with this work**

S. Y. Willow, M. R. Hermes, K. S. Kim, and S. Hirata, *Journal of Chemical Theory and Computation* **9**, 4396-4402 (2013), "Convergence acceleration of parallel Monte Carlo second-order many-body perturbation calculations using redundant walkers."

S. Y. Willlow, M. A. Salim, K. S. Kim, and S. Hirata, *Scientific Reports* **5**, 14358 (2015) (14 pages), "*Ab initio* molecular dynamics of liquid water using embedded-fragment second-order many-body perturbation theory towards its accurate property prediction."

M. A. Salim, S. Y. Willow, and S. Hirata, *The Journal of Chemical Physics* **144**, 204503 (2016), "Ice Ih anomalies: Thermal contraction, anomalous volume isotope effect, and pressure-induced amorphization."

S. Hirata, K. Gilliard, X. He, M. Keçeli, J. Li, M. A. Salim, O. Sode, and K. Yagi, an invited book chapter in *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems* edited by M. S. Gordon (Wiley, Chichester, 2017), "*Ab initio* ice, dry ice, and liquid water."

A. E. Doran and S. Hirata, *Journal of Chemical Theory and Computation* **12**, 4821-4932 (2016), "Monte Carlo MP2 on many graphical processing units."

C. M. Johnson, A. E. Doran, J. Zhang, E. F. Valeev, and S. Hirata, *The Journal of Chemical Physics* **145**, 154115 (2016), "Monte Carlo explicitly correlated second-order many-body perturbation theory."

C. M. Johnson, S. Hirata, and S. Ten-no, *Chemical Physics Letters* [A. H. Zewail Commemorative Issue] **683**, 247-252 (2017), "Correlation factors in explicitly correlated methods."

C. M. Johnson, A. E. Doran, S. Ten-no, and S. Hirata, *The Journal of Chemical Physics* **149**, 174112 (2018), "Monte Carlo explicitly correlated second-order many-body Green's function theory."

A. E. Doran and S. Hirata, "Monte Carlo second- and third-order many-body Green's function methods with frequency-dependent, nondiagonal self-energy," *Journal of Chemical Theory and Computation* **15**, 6097-6110 (2019).

A. A. Kunitsa and S. Hirata, "Grid-based diffusion Monte Carlo for fermions without the fixed-node approximation," *Physical Review E* **101**, 013311 (2020).

A. E. Doran and S. Hirata, "Convergence acceleration of Monte Carlo many-body perturbation methods by using many control variates," *The Journal of Chemical Physics* **153**, 094108 (2020).

A. E. Doran and S. Hirata, "Convergence acceleration of Monte Carlo many-body perturbation methods by direct sampling," *The Journal of Chemical Physics* **153**, 104112 (2020) [JCP Editor's Pick].

Work benefitted from Blue Waters has been highlighted in invited talks by the PI in the following conferences and university seminars: American Conference on Theoretical Chemistry (2014); Low-Scaling and Unconventional Electronic Structure Techniques Conference (2014); International Symposium on "Quantum Chemistry for Extended Systems" in honor of Professor Jean-Marie André (2014); Seminar, Department of Chemistry and Biochemistry, University of Arkansas (2014); Seminar, Department of Chemistry, Virginia Tech (2014); Blue Waters Symposium (2014); Sanibel Symposium (2015); CECAM meeting (2015); Blue Waters Symposium (2015); RAEST (2015); Pacifichem 2015 (2015); Low-Scaling and Unconventional Electronic Structure Techniques Conference (2016); Molecular Quantum Mechanics 2016; International Symposium of Theoretical Chemical

Physics (2016); Molecular Electronic Structure in Buenos Aires (2016); Southeast Regional ACS Meeting (2016); Seminar, Institute for Molecular Science (2016); Seminar, Kobe University (2016); Seminar, Waseda University (2016); Seminar, Keio University (2016); Sanibel Symposium (2017); ACS National Meeting in San Francisco (2017); Seminar, BU-Harvard-MIT (2017); Molecular Science of Ice Symposium (2017); Department of Energy Chemical Theory and Computation Meeting (2017); Canadian Society of Chemistry National Meeting (2017); ACS National Meeting in Washington DC (2017); WATOC (2017); Waterloo Chemical Physics Symposium (2017); XVI Reunion Mexicana de Fiscoquímica Téorica (2017); The Robert S. Mulliken Lecture, University of Georgia (2018); Seminar, Tsinghua University (2018); ACS National Meeting in San Diego (2019); Utah Workshop on Quantum Methods in Molecular and Solid-State Theory (2019); Seminar, University of Tsukuba (2019); Seminar, RIKEN Center for Computational Science (2019); Sanibel Symposium (2020); PQI 2020 (2020); Warsaw Molecular Electronic Structure Conference (2020).

- **Plan for next year**

  We request 240,000 node-hours (120,000 node-hours on XE and 120,000 node-hours on XK).

  The usage schedule is, Q1: 25%, Q2: 25%, Q3: 25%, Q4: 25%.

  (1) We have also developed MC-MP4. The fourth-order many-body perturbation (MP4) theory is as accurate as the so-called "gold standard" in chemistry, the CCSD(T) method, and is expected to be predictive. Therefore, MC-MP4 completes the hierarchy of ab initio MC methods. We plan to use the next year's allocation to perform the benchmark MC-MP4 calculations for large systems that do not lend themselves to local correlation speedup.

  (2) We have developed the interaction MC-MP2 and MC-MP2-F12 methods, which can directly calculate the van der Waals interaction energies in the complete-basis-set limit (not as small differences of noisy total energies). We have been running these methods for the C_{60} dimer. With the next year's allocation, we plan to complete this grand-challenge calculation.

  (3) We are also developing the four-component (Dirac–Hartree–Fock) relativistic MC-MP methods, which may be more scalable for large (many-electron) but not spatially extended systems. We plan to apply the method to CuH, AgH, AuH as well as U_2.

  Projects (1) and (2) have established parallel codes and calculations already started on Blue Waters and there is almost no risk. Project (3) has a much higher risk, but projects (1) and (2) will use the resources in the event the progress in (3) is too slow.

- **Data migration plan**

  Our group's code is stochastic and there is little intermediate, large-volume data because of the nature of stochastic calculations. So, the only important data on Blue Waters storage system are the final results and codes themselves, which are miniscule. We can migrate them to our laptop computers on a moment's notice.