PROJECT INFORMATION

Title: QMCDB: A Living Database to Accelerate Worldwide Development & Usage of Quantum Monte Carlo Methods
PI: Elif Ertekin, University of Illinois
Co-PI(s): Lucas Wagner (University of Illinois), Ray Plante (NIST)
Corresponding Author: Elif Ertekin, ertekin@illinois.edu

EXECUTIVE SUMMARY

Blue Waters has enabled us to carry out automated, high-throughput quantum Monte Carlo calculations of condensed matter systems as part of our development of QMCDB (Quantum Monte Carlo DataBase), a database of materials properties calculated via the highly-accurate quantum Monte Carlo technique. The systems simulated are to be incorporated into our database, which will be made available to the international materials modeling community. This platform for easy, searchable data exchange will accelerate the knowledge base around the use of QMC for materials modeling and enable its evolution from a physics approach to a tool for real engineering materials design. This work would not be possible without Blue Waters, which allows the calculation of a large class of materials ranging from classic to exotic semiconductor materials, photovoltaics, thermoelectrics, and metallic systems; we can take advantage of the near-linear scaling of our methods and code up to several thousand nodes.

DESCRIPTION OF RESEARCH ACTIVITIES AND RESULTS

Key Challenges. Quantum Monte Carlo (QMC) methods are a suite of tools for direct stochastic solution of the many-body interacting Schrodinger Equation. Although QMC methods are considered to be one of the highest-accuracy first-principles materials modeling methods available, and demonstrate a long and distinguished history of benchmark calculations, their usage for materials design and discovery has historically been limited by their large computational cost. With the high performance computing capabilities of Blue Waters, however, it is now possible to extend this method to the realm of high-throughput materials computation and discovery. The goal of our work, therefore, is to develop the first database of materials computations based on quantum Monte Carlo results. We expect that this database will serve as a shared community resource to accelerate the use of this high-accuracy method and advance the community's knowledge of best practices in the application of QMC to real materials design and discovery.

Why It Matters. The prospect of materials design using high performance computation is one of the most exciting for future technologies. However, the properties of modern materials are rather complex, and current numerical methods can fail to describe them quantitatively. Today, QMC is a state-of-the-art suite of tools for high-accuracy *ab initio* modeling. It shows great promise for high-accuracy materials modeling and is already well established in the physics community (model systems, effective Hamiltonians), but its application to real materials with chemical identity remains fairly young. Our goal is to accelerate the development of the QMC community's collective knowledge base around the use of this method for predictive modeling of real materials. Ultimately, the impact will be to enable quantum Monte Carlo methods to emerge as a standard component of the computational materials modeling toolkit, enabling

unprecendented high-accuracy simulation of complex materials, correlated systems, high-T_c superconductors, and other historically challenging materials.

Why Blue Waters. Blue Waters is the key to carrying out the comprehensive set of QMC materials calculations that populate our database. QMC methods exhibit near-linear scaling on the entire Blue Waters platform, which has allowed us to calculate properties of an extensive set of materials that would otherwise not be possible. Historically, the wide scale adoption of QMC as a method for materials modeling has been limited by its large computational cost, but Blue Waters allows us a first key opportunity to overcome this barrier.

Accomplishments.

- *Autogen.* The implementation of an automated workflow (FIG. 1a) for carrying out QMC calculations of solid materials on Blue Waters. The workflow was used together with Blue Waters to carry out our automated quantum Monte Carlo calculations of semiconductor materials. It is now available to the public via our GitHub pages.
- *Calculation of Thermodynamic Properties of Wide Class of Challenging Materials.* We were able to use the *Autogen* tools to carry out a large, systematic calculation of the thermodynamic properties and formation enthalpies of a large and varied class of materials (FIG. 1b). Our efforts thus far have focused on the calculation of thermodynamic properties (total energies and formation enthalpies) and the band gaps of a wide class of semiconductor materials spanning from conventional (silicon, etc.) to more exotic materials (wide band gap oxides, correlated systems, photovoltaic thin film materials).
- *The establishment of QMCDB*: Our Mongo database "QMCDB" is now active through NDS labs, using labs cloud capabilities. The database, which includes results and provenance, is described in detail on the National Data Service Project Wiki page, at http://wiki.nationaldataservice.org/QMCDBProject. This database will serve as a vehicle to quickly overcome the current expertise hurdle and bring the QMC methodology into the standard computational modeling toolkit. While today there are several competing materials databases for density functional theory, there is no existing QMC database. Thus, Blue Waters have provided us the opportunity to establish this tool for the worldwide QMC community.

Next Generation Work. (i) Next generation Track-1 systems in the 2019-2020 timeframe will enable us to carry out more systematic analyses on a larger class of materials. Our initial work shows that simulations of larger sized systems are important for high accuracy. (ii) Towards the longstanding goal of computational materials design, there is often a tradeoff between speed and accuracy. Current methods are relatively fast, but often not sufficiently accurate to enable true computational materials design. Ultimately, computational materials design will require the incorporation of high-accuracy methods to down select candidate materials. This cannot be accomplished without next generation systems. (iii) Finally, the capability to carry out more systematic studies will enable us to carry out big data assessments of correlated many body wave functions, towards the idea of 'correlated electron genomics'.

LIST OF PUBLICATIONS AND DATA SETS

Publications.

- Joshua A. Schiller, Lucas K. Wagner, Elif Ertekin, "Phase Stability and Properties of Manganese Oxide Polymorphs: Assessment and Insights from Diffusion Monte Carlo", Phys. Rev. B 92, 235209 (2015) DOI: 10.1103/PhysRevB.92.235209
- Jaehyung Yu, Lucas K. Wagner, Elif Ertekin, "Towards a systematic assessment of errors in diffusion Monte Carlo calculations of semiconductors: case study of zinc selenide and zinc oxide", J. Chem. Phys. 143 224707 (2015) DOI: 10.1063/1.4937421
- Jaehyung Yu, Lucas K. Wagner, Elif Ertekin, "Quantum Monte Carlo simulations of nitrogen defects in zinc oxide", manuscript submitted for publication.
- Joshua A. Schiller, Ray Plante, Lucas K. Wagner, Elif Ertekin, "Quantum Monte Carlo Database and Autogen: a shared community tool for many-body statistical simulations of materials", manuscript in preparation
- Jaehyung Yu, Elif Ertekin, "Exciton binding energy of two-dimensional MoS_2 and WS_2 from diffusion Monte Carlo", manuscript in preparation.

Data Sets.

- *Quantum Monte Carlo Database (QMCDB).* We have implemented a dedicated Mongo database, entitled QMCDB, for the purposes of sharing with other researchers the progress and results of our work. The data created in the course of this Blue Waters allocation is included in our database (FIG. 1b). To our knowledge, this comprises the largest QMC data set that has been established to date. Our results are further to be linked to the Materials Data Facility (MDF), a repository that invokes the NIST Materials Data and Curation System (MDCS) to enable rapid citable data publication and discovery by others. We are currently building a publically accessible website, with Django back end, so that the content of our database is available to the materials simulation communities.

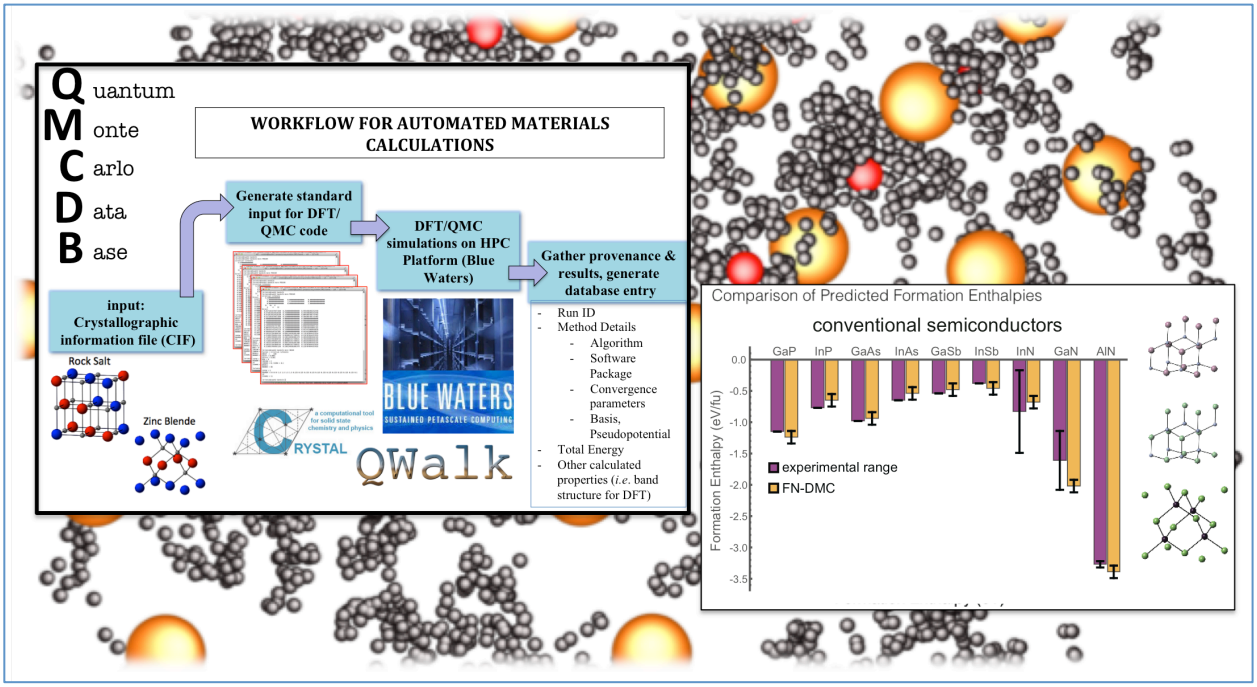

FIG. 1. Background: a snapshot of a configuration of electrons around atomic nuclei in the magnesium oxide solid. Foreground: the *Autogen* workflow for automated quantum Monte Carlo calculations of solids and incorporation into quantum Monte Carlo database, and selected results for calculated formation enthalpies of conventional semiconducting solid materials.