Usage Report for Fiscal Year 2023

Project Title:

Rigorous higher-order DFT

Name:
○Bun Chan (Nagasaki University), Takahito Nakajima (RIKEN)

Laboratory at RIKEN:
Computational Molecular Science Research Team

1. Background and purpose of the project, relationship of the project with other projects

DFT has been the workhorse of computational chemistry. Yet, its full potential has not been reached. The continuous development of DFT relies on statistical and machine-learning techniques. As with science and technology in other areas that rely on "training" with big data, the quality of the data plays an important role.

Regarding data set quality, DFT methods that are formulated with a narrowly focused set of chemical species and molecular properties often suffers from "overfitting", i.e., they are unable to adequately treat systems that do not belong to the types in the training data. Naturally, this would severely hinder the application of DFT.

In our previous studies within this project on the development of advanced DFT methods, we have found the issue of overfitting to affect even some of the leading DFTs. This is despite them incorporating limitations in the statistical training with the aim to avoid this very issue.

In recent years, we have devised strategies to improve the robustness of new DFT methods by diversifying the range of data used formulation of DFT. Let us emphasize the importance of diversification by noting that major data sets include almost exclusively light p-block species, and they cover just 15% of the periodic table.

We thus aim to expand the data sets to cover the rest of the elements. Our earlier effort includes, e.g., transition-metal species, which is an important class of chemicals that underpins modern industries. In FY2023, we continue our endeavor in the accurate determination of data for a diverse range of chemicals, by using highly accurate quantum chemistry computations including those previously developed in this project and new ones formulated within this past year.

2. Specific usage status of the system and calculation method

This project employs the Gaussian and Q-Chem programs on Hokusai BW, as well as a wide range of standard quantum chemistry software packages such as Molpro, MRCC, and ORCA. Thsy enable us to access a diverse range of quantum chemistry methodologies, including highly accurate coupled-cluster and multi-reference methods with which reference data are obtained, and DFT methods with which insights into the fundamentals of a reliable DFT can be revealed.

Regarding the computational determination of reliable reference data, we have further diversified from light p-block species to cover s-block elements, as well as transition metal species that are more challenging for theoretical methods than the systems that we have previously considered.

In addition, within p-block species, we note that most data sets focus on molecular species, with little attention given to nano-sized and bulk materials. A key reason is the excessive computational resources required for their calculation. As we have previously developed substantially lower-cost yet reliable methods, we apply them to larger nano-sized systems to complement existing molecular data.

With the new data sets, the assessment of DFT is trivial in comparison, and we have carried out such benchmarking in each case. Including in our

assessments are a small but representative collection of DFTs that are orthogonal to one another in their formulations. Such distinct differences in their natures enables straightforward identification of key ingredients that are important for the different classes of chemical systems; this facilitates the future development of more reliable methods.

3. Result

One of our key contributions is the provision of reliable molecular data for a wide range of small to medium-sized ionic species consisting of alkali and alkali-earth metals (Figure 1). There is no such data at all until this point.

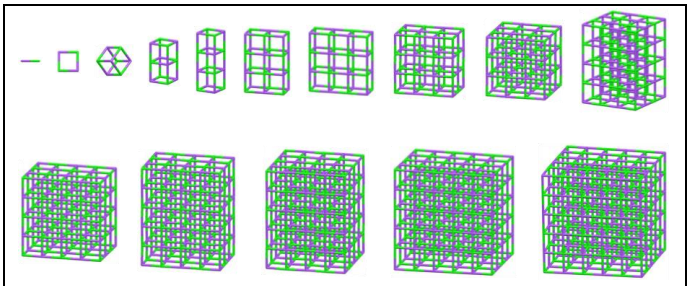

Figure 1. Computed ionic clusters for MX (M = Na, Mg, K, Ca, and X = O, F, S, Cl).

The chemical data that we have obtained including bond lengths, atomization energies and zero-point energies. Using these data, we have found that lower-cost wavefunction methods such as CCSD(T)-F12/VDZ provides an exceptionally accurate means for calculating larger systems, though it is still computationally prohibitive for the calculation of nano-sized species.

For the even-lower-cost DFT methods, the MN15 method, which is one of the most accurate DFT, yields reasonable results. However, another method, ωB97M-V, which is generally as accurate as MN15 or even better, gives errors that are twice as large for the ionic species. The assessment not only reveals the best DFT method for the ionic species, but also enable us to improve its accuracy.

We have used a revised MN15 and some calibrated semi-empirical methods to study nano-sized ionic clusters, thus clarifying the transition from molecules to bulk materials through nano-sized systems (Figure 2).

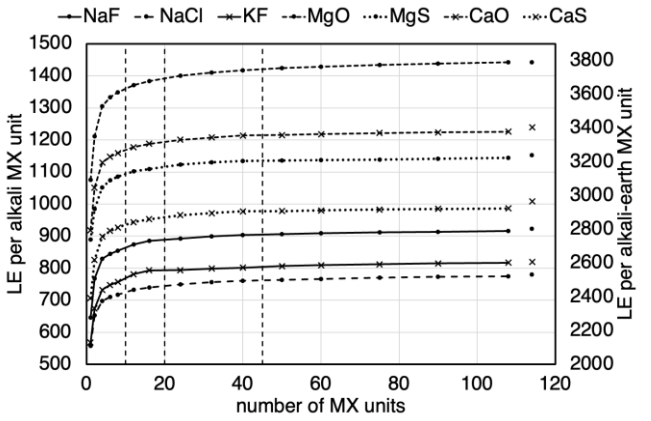

Figure 2. The convergence of bond energies for ionic species from a single molecule to bulk salt.

In other studies, we have obtained accurate thermochemical data for chalcogen nucleophiles, enzyme models, and medium-sized graphene-models. We have also expanded the types of chemical data from thermochemical to electronic properties. Figure 3 shows the expansion of our data from light-p-block to a much wider range of elements.

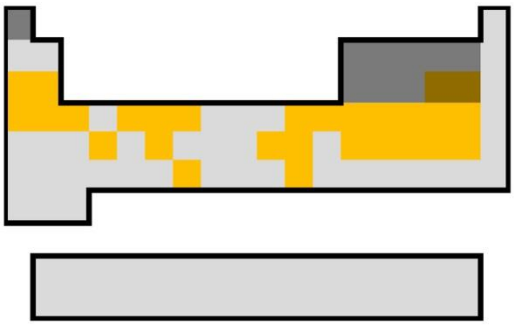

Figure 3. Elements covered by legacy data set (patterned) and those added in this project (yellow).

4. Conclusion

Our provision of high-quality data using quantum chemistry expanded the scope of DFT development. Specifically, they reveal the weaknesses of methods that are generally the most accurate within a more limited scope. These results point towards the need for continual development of DFT methods, and the important role that new and reliable data will play.

5. Schedule and prospect for the future

We will further expand our data sets to cover a larger chemical space. This would ultimately lead to a fully unbiased collection for the development of more reliable low-cost methods.

Usage Report for Fiscal Year 2022

Fiscal Year 2023 List of Publications Resulting from the Use of the supercomputer

[Paper accepted by a journal]

1. Computational Insights into the Singlet–Triplet Energy Gaps, Ionization Energies, and Electron Affinities for a Diverse Set of 812 Small Fullerenes (C_{20}–C_{50}). Chan, B.; Karton, A. *Phys. Chem. Chem. Phys.* **2023**, *25*, 10899.
2. PAH335 – A Diverse Database of Highly Accurate CCSD(T) Isomerization Energies of 335 Polycyclic Aromatic Hydrocarbons. Karton, A.; Chan, B. *Chem. Phys. Lett.* **2023**, *824*, 140544.
3. A Step-by-Step Investigation of Sodium Chloride Clusters: Accurate References, Assessment of Low-Cost Methods, and Convergence from Molecule to Salt. Chan, B. *Mol. Phys.* **2023**, *121*, e2088422.
4. Optimal Small Basis Set and Geometric Counterpoise Correction for DFT Computations. Chan, B. *J. Chem. Theory Comput.* **2023**, *19*, 3958.
5. Compilation of Ionic Clusters with the Rock Salt Structure: Accurate Benchmark Thermochemical Data, Assessment of Quantum Chemistry Methods, and the Convergence Behavior of Lattice Energies. Chan, B. *J. Phys. Chem. A* **2023**, *127*, 5652.
6. The Prospects of Cation Transfer to Chalcogen Nucleophiles. Chan, B.; Shirakawa, S. *Can. J. Chem.* **2023**, *101*, 603.

[Oral presentation]

1. *The 10th Asian Pacific Conference of Theoretical and Computational Chemistry (APATCC2023)*, 2023, Quy Nhon (invited oral presentation).

[Poster presentation]

1. *69th Polarographic and Electroanalytical Chemistry Symposium*, 2023, Fukue (poster presentation).