PHD PROJECT DESCRIPTION
(4000 characters max., including the aims and work plan to de published online)

Project title:

1.1. Project goals

- Accelerating the orbital-optimization (OO) process of pair coupled cluster doubles through machine learning (ML) and the design of the ML-OO-pCCD approach
- Benchmarking of the derived ML-OO-pCCD for large-scale problems found in organic electronics for various interaction types, atomic basis sets, and molecular scales
- Large-scale modeling of organic electronics like light-harvesting materials

1.2. Outline

Experimental advancement of photovoltaic materials heavily relies on a trial-and-error approach. This disadvantage entails large workloads and depletes a lot of resources. A more efficient approach is to exploit quantum chemistry to accelerate the development of novel compounds. The computational models, however, are difficult primarily because conventional highly-accurate approaches are technically limited to small model compounds and demand expert user control. An inexpensive wave-function-based alternative can be found in geminal-based methods, where the electronic wave functions are constructed from electron-pair states, reflecting the intuitive picture of Lewis structures as pairs of electrons. One obstacle for geminal-model chemistry is the large number of orbital optimization steps, which can be rate-determining, prolonging simulation times. This project will remedy this issue by providing a machine-learned recipe to construct the initial guess orbitals. Specifically, we will focus on one computationally inexpensive, but robust geminal-based flavor, namely pCCD. The machine-learned and accelerated orbital-optimization procedure has to take into account a diverse chemical space, comprising various donor, bridge, and acceptor molecules, governed by strong, weak, and dispersion-like interactions. The proposed models will be implemented in PyBEST and will (i) enhance our fundamental understanding of molecules and their properties, (ii) aid the development of black-box quantum chemistry calculations, and (iii) accelerate the discovery of new materials. Thus, this project will shift the current paradigm in theoretical materials design of organic electronics towards novel and systematically improvable approaches (beyond DFT).

1.3. Work plan

The proposed Ph.D. project includes the following work tasks:

- Training data generation for ground-state pCCD (with orbital optimization)
- Definition of appropriate descriptors to efficiently predict the label value of an improved starting guess for pCCD natural orbitals
- Selection of the chemical space and its division into subsystems based on correlation strengths, interaction types, and bonding/conjugation patterns
- Construction of the ML infrastructure and its incorporation into the PyBEST software package

- Benchmarking and application of the derived ML-OO procedure to chemical problems

1.4. Literature (max. 7 listed as a suggestion for a PhD candidate preliminary study)

- *Molecular Electronic-Structure Theory, T. Helgaker, P. J*ørgensen, and J. Olsen, John Wiley & Sons, New York, 2000.
- P. Tecmer and K. Boguslawski, *Phys. Chem. Chem. Phys.*, 2022, 24, 23026–23048.
- K. Boguslawski, A. Leszczyk, A. Nowak, F. Brzęk, P. S. Żuchowski, D. Kędziera and P. Tecmer, *Comput. Phys. Commun.*, 2021, 264, 107933.

1.5. Required initial knowledge and skills of the PhD candidate

The candidate should

- have knowledge of molecular electronic structure theory like Hartree-Fock theory, coupled cluster theory, etc.
- be familiar with Linux, bash, and basic terminal commands.
- have basic programming skills, preferably in Python.
- be familiar with PyBEST.
- be fluent in spoken and written English.

1.6. Expected development of the PhD candidate's knowledge and skills

The candidate will

- gain expertise in unconventional electronic structure methods, like pCCD.
- gain fundamental knowledge in state-of-the-art many-body-theory.
- be actively involved in the development of software written in Python and C++ (using git and GitLab).
- be well-trained in electronic structure calculations (using wave function-based methods) at various scales and in machine-learning techniques.