1. PHD PROJECT DESCRIPTION (4000 characters max., including the aims and work plan)

Project title: Devising Quantum-Information-Driven Domain-Based Coupled Cluster Models by Means of Machine Learning

1.1. Project goals

- Definition of correlation domains for (frozen) pair coupled cluster (fpCC) methods using quantum information theoretical concepts through machine learning (ML) and the design of the ML correlation-based domain (CD) fpCC approach
- Benchmarking of the derived ML-CD-fpCC 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 guide the synthesis of new materials. The computational models, however, are difficult 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. To approach the ultimate goal of a geminal-model chemistry, several challenges remain. One obstacle is to derive efficient models to account for post-geminal correlation effects. For large systems, the existing corrections may be computationally too expensive, prohibiting theoretical studies. This project will remedy these drawbacks by providing a machine-learned recipe to construct correlation domains for post-geminal calculations based on the orbital-pair mutual information. Specifically, we will focus on specific geminal-based flavors, namely, various fpCC models. The machine-learned correlation domains have to be constructed for a diverse chemical space, comprising large systems featuring 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 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 large-scale modeling of organic electronics towards novel approaches (beyond DFT).

1.3. Work plan

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

- Training data generation for ground-state various (fp)CC models
- Definition of appropriate descriptors to efficiently predict the label value of proper

correlation domains based on pCCD orbital-pair mutual information
- 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
- GPU-acceleration of the ML-CD-fpCC models
- Benchmarking and application of the derived ML-CD-fpCC models 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
- be familiar with Linux, bash, and 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 and 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 machine-learning techniques