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

Project title:

Harnessing Active Learning Strategies for Simulating Photochemical Reactions

1.1. Project goals

Understanding the excited-states (nonadiabatic) dynamics of a photo-excited molecule is an active area of research because electronic transitions play a vital role in photosynthesis, vision processes, solar cells, photodynamic cancer therapy, and photocatalysis. Nonadiabatic molecular dynamics (NAMD) simulations have become one of the most powerful tools for modeling excited electronic states relaxation mechanisms in photo-active molecular systems. However, on-the-fly NAMD simulations are computationally very expensive. It requires single-point quantum chemistry (QC) calculations for hundreds of thousands of different molecular geometries during its propagation. Recently, the emergence of Machine Learning (ML) approaches in quantum chemistry has created a way to advance the state of the art. A well-trained ML model can predict the energies and forces at a negligible cost, which are the required ingradients in NAMD simulations. The integration of ML models in nonadiabatic molecular dynamics can dramatically speed up the simulation and increase the statistics by producing a large number of trajectories.

This project aims to build a sustainable and robust framework of machine learning assisted excited states molecular dynamics simulation. This proposed protocol will be applied to study the photo-relaxation pathways of biologically important canonical and modified nucleobases.

1.2. Outline

The project exploits the following distinct scientific methodologies:

(a) Exploring the reaction pathways of the photo-excited molecule of interest by static ab initio quantum chemistry calculations.
(b) Simulating the photo-absorption spectra by using the nuclear ensemble approach.
(c) Ab initio NAMD simulations (for verification and) using the Trajectory Surface Hopping methodology.
(d) Active Learning based ML-NAMD simulations.

The project will use and also contribute to developing the following software:

- Newton-X – an open-source, free software for the NAMD simulations. It is interfaced with several quantum chemistry packages (Gaussian16, ORCA, Turbomole, OpenMolCAS, Columbus, PySCF, PyOQP etc.) for electronic structure calculations.
- MLatom – an open-source, free program for developing machine learning models for molecular system.

1.3. Work plan

The project will employ an active learning approach to train ML potentials. Active learning uses a similarity measure to determine the uncertainty of predicted molecular properties during trajectory propagation. The *query-by-committee* strategy will be adapted to assess the model uncertainty, quantified as the maximum standard deviation of the atomic forces $\sigma = max \sqrt{\sum_{\alpha=1}^{N} \left\| F_i^{\alpha} - \bar{F}_i \right\|^2}$. F_i^{α} is the atomic force on the atom i predicted by the ML potential α, and $\bar{F}_i$ is the average force on the atom i over the c. σ provides a measure of uncertainty in the model's predictions, reflecting the variance in atomic force estimations across different

models. Higher values indicate regions where the model is less confident, while lower values suggest greater reliability in the predictions.

The active learning workflow starts with training an initial quantum chemistry dataset to construct independently trained N initial ML potentials. The uncertainty measurement identifies the bad predictions, and the uncertain structures are recomputed with ab initio calculations and added to retrain the ML models. This procedure continues until the error remains approximately the same between two consecutive iterations. As a result of this active learning cycle, the training set is progressively enlarged, enabling the development of robust ML potential capable of predicting atomic energies and forces within chemical accuracy. We will employ the recently developed higher-order equivariant message passing neural network, MACE, for constructing the ML models due to its outstanding success with low training data.

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

1) Crespo-Otero et al. Recent Advances and Perspectives on Nonadiabatic Mixed Quantum-Classical Dynamics. *Chemical Reviews*. (2018) 118, 7026-7068. https://doi.org/10.1021/acs.chemrev.7b00577.
2) Mukherjee et al. Assessing Nonadiabatic Dynamics Methods in Long Timescales. *J. Chem. Theory Comput.* (2025) 21, 29-37. https://doi.org/10.1021/acs.jctc.4c01349
3) Muller et al. Machine learning for nonadiabatic molecular dynamics: best practices and recent progress, *Chem. Sci.* (2025) 16, 17542-17567. https://doi.org/10.1039/D5SC05579B
4) *Quantum Chemistry in the Age of Machine Learning*; Dral, P. O., Ed.; Elsevier, 2023. https://doi.org/10.1016/C2020-0-03124-5
5) Dral et al. MLatom software ecosystem for surface hopping dynamics in Python with quantum mechanical and machine learning methods. *J. Chem. Theory Comput.* (2024) *20*, 5043–5057. https://doi.org/10.1021/acs.jctc.4c00468
6) Dral et al. Charting electronic-state manifolds across molecules with multi-state learning and gap-driven dynamics via efficient and robust active learning. *npj Comput. Mater.* (2025) 11, 132. https://doi.org/10.1038/s41524-025-01636-z
7) Veliz et al. Exploring active learning strategies for excited state dynamics: application to uracil, Phys. Chem. Chem. Phys. 2026. https://doi.org/10.1039/D6CP00203J

1.5. Required initial knowledge and skills of the PhD candidate

Good working knowledge of computational chemistry methods for the description of excited states, experience of machine learning methods in chemistry preferred, Python programming skills, fluent English, high motivation, and ability of the teamwork.

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

This PhD project will offer the candidate a unique opportunity to gain interdisciplinary expertise at the intersection of theoretical and computational chemistry, scientific programming and data science for a versatile career in both academia and industry.

The candidate will:

- Acquire deep scientific knowledge of excited-state phenomena, nonadiabatic molecular dynamics, and photochemical reaction mechanisms, building a strong foundation in computational chemistry.

- Develop advanced computational skills, including high-level Python programming, data handling, and software development relevant to modern chemical simulations.
- Master machine learning techniques, including supervised and unsupervised learning, focusing on practical applications such as model training, feature engineering, and performance evaluation in chemical systems.
- Gain experience in active learning strategies, enabling efficient sampling and model optimization, and deepening their understanding of adaptive data-driven research methodologies.

In addition to technical skills, the candidate will be encouraged to:

- Strengthen communication and collaboration skills through regular presentations at group meetings, workshops, and conferences.
- Hone scientific writing abilities by contributing to high-quality publications in peer-reviewed journals.
- Build project management and critical thinking capabilities, allowing them to independently design, implement, and assess complex computational workflows.

Overall, the project is designed to foster a well-rounded researcher equipped with cutting-edge knowledge and transferable skills that are highly valued in research institutions, tech-driven industries, and interdisciplinary innovation sectors.