

Carolin Müller

Illuminating Photodynamics with Machine Learning Techniques

When a molecule absorbs light, it enters a non-equilibrium state and triggers a cascade of non-adiabatic processes. Theoretical modeling of these electronic excitations and the subsequent processes plays a crucial role in advancing fundamental research and driving technological innovation. However, accurate simulation of both, the static properties of multiple electronic states and their dynamics, is limited by the resource-intensive nature of the underlying quantum chemical techniques. To overcome this bottleneck, the emergence of **machine learning** (ML) marks a breakthrough development that offers a vast amount of opportunities to accelerate and facilitate excited state simulations [1-3].

There are various avenues for utilizing ML techniques in describing electronically excited states and their dynamics: On the one hand, ML can be employed in the **analysis** and interpretation of quantum chemical data, such as the result of excited state dynamics simulations [4-7]. On the other hand, when generating the underlying quantum chemical data *via* quantum chemical calculations, ML aids in **streamlining workflows** [8-9] and **accelerating predictions** [10-17]. In the former context, a subset of approaches utilizes ML to identify optimal parameters for quantum chemical computations, such as constructing active spaces for complete active-space self-consistent-field calculations [8]. A more extensive array of approaches focuses on the generation of ML force fields in order to neglect wave function information to achieve higher computational speeds [10-17]. These ML potentials are suitable for describing time-dependent properties in dynamic simulations.

Machine learning photodynamics

Especially the ML approaches that **replace** many **successive quantum chemical calculations** with the faster ML calculations turn out to be particularly beneficial. This is the case when simulating non-adiabatic molecular dynamics (NAMD). These NAMD computations are based on two core components: electronic structure methods that are able to adequately describe the energetics and topology of excited-state potential energy surfaces, and MD algorithms accounting for non-adiabatic phenomena and trajectory propagation.

Jun.-Prof. Dr. Carolin Müller
Department Chemie und Pharmazie
Friedrich-Alexander-Universität Erlangen-Nürnberg
Nägelsbachstraße 25, D-91052 Erlangen
carolin.mueller6@googlemail.com
<https://carolin-m.github.io/>

DOI-Nr.: 10.26125/Oehr-vk47

In the realm of NAMD algorithms, various approaches have been developed, spanning from full quantum treatment of both, nuclei and electrons, to a combination of classical and quantum chemical methods [18]. Of particular interest among these approaches is the latter category, which enables simulations of relatively large molecules containing 10-100 atoms over moderate timeframes, ranging from femtoseconds to picoseconds. This strikes a balance between accuracy and computational resources. One method within the mixed quantum-classical category, known as **surface hopping** (SH), has gained prominence due to its robustness and adaptability for extensions. Recently, it was combined with ML, which led to a considerable acceleration of SH simulations. Thanks to this speed-up, photodynamic studies can be performed within hours, whereas they would otherwise take dozens of years [10, 12].

Incorporating ML into NAMD simulations usually involves the following steps (see Figure 1): First, **training data** is generated using quantum chemistry. This entails computing and storing electronic properties used in surface hopping, such as multiple potential energy surfaces, associated forces on atomic nuclei, and properties arising from interactions between electronic states (known as couplings), for various molecular geometries. In doing so, it is essential to select an appropriate approach from the pool of excited state methods, e.g. multi-reference methods (MRCI, CASSCF, CCSD) or time-dependent density functional theory (TDDFT), to adequately address the specific chemical or physical query of interest. Next, this quantum chemical data is utilized to **train ML models**, which are statistical models that learn the relationships between potential energies and other properties and molecular geometry. For this purpose, neural networks and kernel-based methods are the most widely used techniques [1, 10-17]. Subsequently, the **ML-derived properties** are incorporated into NAMD methods, such as wave packet techniques or surface hopping approaches. This means that instead of performing quantum chemical calculations for specific properties, these properties are obtained by requesting them from the trained ML model.

Nanosecond photodynamics simulations

Typically time steps of around 1 femtosecond are employed in NAMD simulations. This means that around a million calculations are needed to access information in the nanosecond time regime (for a single trajectory). Consequently, conventional trajectory-based NAMD simulations are usually limited to picosecond time scales. In contrast, ML-based dynamics can capture

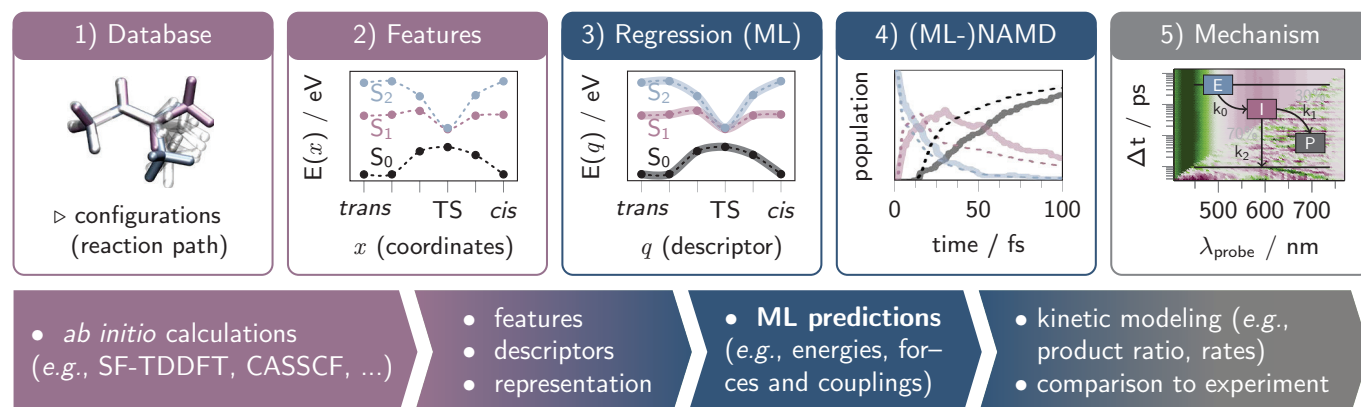


Fig. 1: Schematic representation illustrating a common workflow for calculating excited state phenomena with non-adiabatic molecular dynamics (NAMD) simulations, featuring the integration of machine learning (ML).

processes extending into the nanosecond range, representing a potential acceleration of approximately 1000 times.

For example, Westermayr *et al.* illustrated that they could complete a simulation of surface hopping dynamics for the methylene immonium cation (CH_2NH_2^+) lasting 100 femtoseconds using the *SchNarc* approach on a single CPU (2x Intel Xeon E5-2650 v3 CPU) in just 24 seconds [10]. This achievement enabled them to extend NAMD simulation times to the nanosecond range while maintaining a reasonable balance between efficiency and accuracy [11]. Similarly, Lopez *et al.* conducted a 10 nanosecond simulation of trans-hexafluoro-2-butene in just 2 days, utilizing the machine learning-driven ab initio molecular dynamics package *PyRAI²MD* [12]. Employing the same tool, they were also able to explore the structures of octa-substituted [3]-ladderienes on the subpicosecond timescale, a feat not previously realized [13]. Additionally, Young *et al.* recently presented the *mlp-train* approach that enabled the creation of accurate and efficient machine learning potentials for modelling light-induced Diels-Alder reactions involving up to 50 atoms in just one day (about 100 CPU hours) [14]. Furthermore, Gómez-Bombarelli *et al.* presented the neural network method *DANN*, which relies on diabatic states. This method accelerated the exploration of the photoisomerization of azobenzene by a factor of six. Notably, their model not only exhibited high accuracy but also proved to be transferable across various azobenzene derivatives. This adaptability enabled them to predict reasonable photoisomerization quantum yields for previously unseen species [15].

Identification of photoinduced mechanism

Beyond accelerating dynamic calculations, ML offers opportunities to simulate previously inaccessible systems and identify novel reaction mechanisms. For example, in the case of tyrosine, traditional quantum chemistry methods excel in specific areas of potential energy surfaces but yield inaccurate results elsewhere. However, different QC approaches can complement each other, and by selectively generating training data where QC methods are reliable, ML models were able to describe the system across various structures. This unveiled „roaming“ in

simulations of tyrosine – a phenomenon initially only observed experimentally in formaldehyde [16].

In others studies, ML-NAMD opened avenues for the discovery of photoreaction mechanism. For instance Wen and co-workers explored the **all-light-driven rotation** of a molecular motor on up to 2 picosecond time scale. With the help of these ML-NAMD simulations they discovered that a particular intermediate shows shortcut photoinduced reactions, that share conical intersections similar to those of the photoisomerization process by which the intermediate is formed. Furthermore, they revealed the important role of a ring flip as structural distortion ultimately driving the unidirectional rotation of the motor [17]. In another example, Lopez *et al.* unraveled how substituents mechanistically affect the ultrafast dynamics and mechanism of **[2+2]-photocycloadditions**. Based on 2 picosecond ML-NAMD trajectories, they found that methyl-substituted [3]-ladderdiene undergo 6 π -conrotatory electrocyclic ring-opening, whereas the trifluoromethyl- and cyclopropyl-substituted [3]-ladderdiene prefer the [2+2]-cycloaddition pathways, which could be directly related to experimentally observed reaction yields [13].

Current and future directions

The successes achieved in recent years by ML-based NAMD simulations have opened new avenues for the fast and efficient study of light-induced phenomena across a range of molecules. However, all these efforts highlight a persistent challenge, namely modeling crucial factors such as potential energy surfaces (which require balanced learning across multiple surfaces) and couplings using ML. This requires a careful generation of training data, which is particularly important when applying ML models in dynamics simulations. It is vital that this database encompasses a wide range of critical configurations while maintaining an appropriate dataset size so that over- or under-representation of certain regions is avoided. In addition, careful consideration is required when selecting data points in critical regions such as cone intersections where potential energy surfaces may have discontinuities and non-adiabatic coupling vectors may have singularities.

To overcome these challenges, there is an urgent need to develop comprehensive, robust and diverse databases that thoroughly explore the molecular configuration space [19]. This is crucial for improving the efficiency and accuracy of ML-based NAMD simulations. Our ongoing research is dedicated to this goal. And we are working on developing hierarchical workflows for constructing vast databases in order to achieve a better balance between cost and accuracy in both, training and application of ML force field models for excited states.

In conclusion, ML shows great potential for accelerating diverse simulations, especially in photochemistry and -physics. When applied adeptly, ML can simplify the analysis of light-driven processes, providing profound insights. Though, ML is not a universal solution for all challenges in this field. Various other methods will remain essential, but ML will support and enhance them, solidifying its status as an indispensable tool in theoretical chemistry – and beyond.

References

- [1] P. O. Dral, *Quantum Chemistry in the Age of Machine Learning* 2023, Elsevier, 10.1016/C2020-0-03124-5
- [2] P.O. Dral, M. Barbatti: Molecular excited states through a machine learning lens. *Nat Rev Chem* 2021 **5**, 388 – 405
- [3] J. Westermayr and P. Marquetand: Machine Learning for Electronically Excited States of Molecules, *Chem. Rev.* 2021, **121**, 16, 9873 – 9926
- [4] F. Häse, I. Fdez. Galván, A. Aspuru-Guzik, R. Lindh and M. Vacher: How machine learning can assist the interpretation of ab initio molecular dynamics simulations and conceptual understanding of chemistry, *Chem. Sci.* 2019, **10**, 2298 – 2307
- [5] B. Casier, S. Carniato, T. Miteva, N. Capron, and N. Sisourat: Using principal component analysis for neural network high-dimensional potential energy surface. *J. Chem. Phys.* 2020, **152**, 23, 234103
- [6] S. Bougueroua, Y. Aboulfath, D. Barth and M. Pierre Gaigeot: Algorithmic graph theory for post-processing molecular dynamics trajectories, *Molecular Physics* 2023, **121**, 7 – 8
- [7] Y. Zhu, J. Peng, X. Kang, C. Xu and Z. Lan: The principal component analysis of the ring deformation in the nonadiabatic surface hopping dynamics, *Phys. Chem. Chem. Phys.* 2022, **24**, 24362 – 24382
- [8] J. J. Bao, S. S. Dong, L. Gagliardi, and D. G. Truhlar: Automatic Selection of an Active Space for Calculating Electronic Excitation Spectra by MS-CASPT2 or MC-PDFT, *J. Chem. Theory Comput.* 2018, **14**, 4, 2017 – 2025
- [9] W. Jeong, S. J. Stoneburner, D. King, R. Li, A. Walker, R. Lindh, and L. Gagliardi: Automation of Active Space Selection for Multireference Methods via Machine Learning on Chemical Bond Dissociation, *J. Chem. Theory Comput.* 2020, **16**, 4, 2389 – 2399
- [10] J. Westermayr, M. Gastegger, and P. Marquetand: Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics, *J. Phys. Chem. Lett.* 2020, **11**, 10, 3828–3834
- [11] J. Westermayr, M. Gastegger, M. F. S. J. Menger, S. Mai, L. González, and P. Marquetand: Machine learning enables long time scale molecular photodynamics simulations, *Chem. Sci.* 2019, **10**, 8100-8107
- [12] J. Li, P. Reiser, B. R. Boswell, A. Eberhard, N. Z. Burns, P. Friederich and S. A. Lopez: Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations, *Chem. Sci.* 2021, **12**, 5302 – 5314
- [13] J. Li, R. Stein, D. M. Adrion, and S. A. Lopez: Machine-Learning Photodynamics Simulations Uncover the Role of Substituent Effects on the Photochemical Formation of Cubanes, *J. Am. Chem. Soc.* 2021, **143**, 48, 20166–20175
- [14] T. A. Young, T. Johnston-Wood, H. Zhang, and F. Duarte: Reaction dynamics of Diels–Alder reactions from machine learned potentials, *Phys. Chem. Chem. Phys.* 2022, **24**, 20820 – 20827
- [15] S. Axelrod, E. Shakhnovich, and R. Gómez-Bombarelli: Excited state non-adiabatic dynamics of large photoswitchable molecules using a chemically transferable machine learning potential, *Nat Commun* 2022, **13**, 3440
- [16] J. Westermayr, M. Gastegger, D. Vörös, L. Panzenboeck, F. Joerg, L. González, and P. Marquetand: Deep learning study of tyrosine reveals that roaming can lead to photodamage, *Nat. Chem.* 2022, **14**, 914–919
- [17] H. Xu, B. Zhang, Y. Tao, W. Xu, B. Hu, F. Yan, and J. Wen: Ultrafast Photocontrolled Rotation in a Molecular Motor Investigated by Machine Learning-Based Nonadiabatic Dynamics Simulations, *J. Phys. Chem. A* 2023, **127**, 37, 7682–7693
- [18] T. R. Nelson, A. J. White, J. A. Bjorgaard, A. E. Sifain, Y. Zhang, B. Nebgen, S. Fernandez-Alberti, D. Mozyrsky, A. E. Roitberg, and S. Tretiak: Non-adiabatic Excited-State Molecular Dynamics: Theory and Applications for Modeling Photophysics in Extended Molecular Materials, *Chem. Rev.* 2020, **120**, 4, 2215–2287
- [19] M. Pinheiro, S. Zhang, P. O. Dral, and M. Barbatti: WS22 database, Wigner Sampling and geometry interpolation for configurationally diverse molecular datasets, *Sci Data* 2023, **10**, 95

Jun.-Prof. Dr. Carolin Müller

CV see page 185.



Foto: Sven Döring
(IPHT Jena)