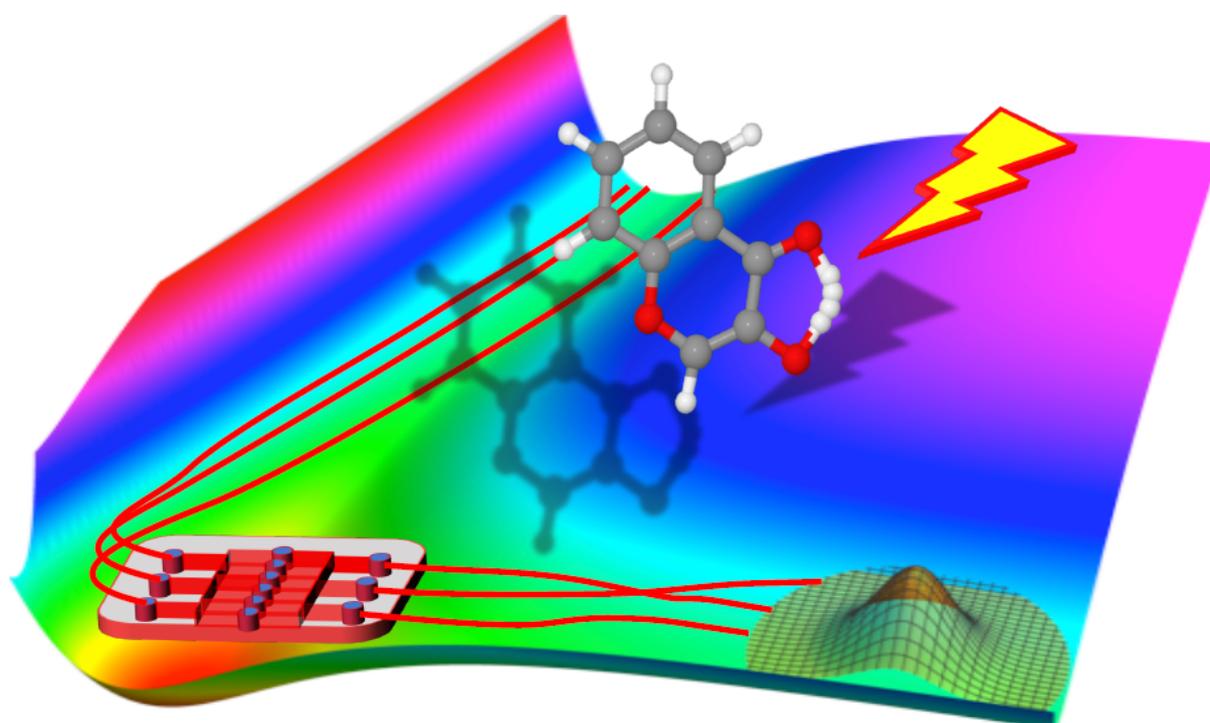


# MLQC4Dyn - Week 2

## Abstracts



# Classical and Quantum Dynamics on Machine-Learned Potential Energy Surfaces

Markus Meuwly  
University of Basel

In this talk I will discuss recent developments and applications of machine learning (ML) potential energy surfaces (PESs) to chemical and biological processes in the gas and in the condensed phase. Particular emphasis is put on effects that can not be easily captured from more conventional empirical energy functions. It will be shown that rigorous sampling of conformational space and converged observables can be obtained from ML MD and ML/MM MD simulations. Applications that will be discussed range from chemical reactions in the gas phase to the spectroscopy of eutectic mixtures for all of which experimental data is available for direct comparison and validation.

# Predicting material properties with the help of machine learning

Bingqing Cheng  
*Institute of Science and Technology (IST) Austria*

A central goal of computational physics and chemistry is to predict material properties using first-principles methods based on the fundamental laws of quantum mechanics. However, the high computational costs of these methods typically prevent rigorous predictions of macroscopic quantities at finite temperatures, such as chemical potential, heat capacity and thermal conductivity.

In this talk, I will first discuss how to enable such predictions by combining advanced statistical mechanics with data-driven machine learning interatomic potentials. As an example, we computed the phase diagram of water from density functional theory at the hybrid level, accounting for thermal fluctuations, proton disordering and nuclear quantum effects. As applications in high-pressure physics, we simulated the high-pressure hydrogen system and found supercritical behaviour above the melting line, mapped the phase diagram of superionic water, and determined the thermodynamic phase boundaries for diamond nucleation from hydrocarbon mixtures. Besides thermodynamic properties, I will talk about how to compute the heat conductivities of liquids just from equilibrium molecular dynamics trajectories.

## Machine learning for excited-state molecular dynamics

Julia Westermayr  
University of Warwick

Nonadiabatic molecular dynamics simulations are powerful tools to study the light-induced breaking and formation of bonds in molecules and materials and can provide a deeper understanding of the underlying mechanisms that is also crucial to deliver sustainable photocatalysis and efficient energy materials. However, existing limitations of conventional techniques, such as high computational costs mean that we currently cannot address the most interesting problems in materials science, photochemistry, and biology [1].

In this talk, I will show how machine learning techniques can be used to overcome current limitations of nonadiabatic dynamics simulations: On the example of the nonadiabatic molecular dynamics of the amino acid tyrosine I will show that photodynamics can be sped up by machine learning to allow for simulations on experimentally relevant time scales. In this way, unexpected reaction mechanisms are discovered that provide new insights into the photochemistry of biological systems [2,3]. Finally, I will discuss the description of excited states of multiple molecules to enable high-throughput screening and the design of novel molecular systems [4,5].

[1] J. Westermayr, P. Marquetand *Chem. Rev.* 121(16), 9873-9926 (2021).

[2] J. Westermayr, M. Gastegger, P. Marquetand *J. Phys. Chem. Lett.* 11(10), 3828-3834 (2020).

[3] J. Westermayr, M. Gastegger, D. Vörös, L. Panzenböck, F. Jörg, L. González, P. Marquetand, arXiv:2108.04373 (2022).

[4] J. Westermayr, R. Maurer, *Chem. Sci.* 12, 10755-10764 (2021).

[5] J. Westermayr, J. Gilkes, R. Barrett, R. J. Maurer, unpublished (2022).

# Unraveling patterns in nonadiabatic molecular dynamics data with machine learning

Max Pinheiro Jr

Aix-Marseille University

Trajectory-based nonadiabatic molecular dynamics (NAMD) can be viewed as robust data generators. Within the wealth of output generated by NAMD simulations, a common and challenging task is to find the key active coordinates that drive the molecular system through critical regions of the potential energy surfaces, which, in turn, can trigger the photoinduced reactions. Such photochemical processes can unfold into a cascade of possible reaction pathways accessible through the NAMD simulations, thereby adding an extra layer of difficulty to the data exploration problem. In this scenario, unsupervised machine learning (ML) can bring an automated solution for the in-depth analysis of NAMD data, facilitating the interpretation and understanding of the underlying photo-dynamical processes. To contribute to this solution, we have developed the *Unsupervised Learning Analysis of Molecular Dynamics* (ULaMDyn) program that provides a complete data analysis pipeline, going from data curation to molecular representations, dimension reduction, and clustering analysis. The unsupervised learning methods implemented in ULaMDyn aim to surpass existing barriers for chemists to extract insights from NAMD simulations regardless of the complexity of the molecular system under study. In this talk, I will present the main features of ULaMDyn by showcasing practical examples of dimensionality reduction and clustering for analyzing NAMD data. We expect that the development of a new program based on unsupervised ML will pave the way for conceptual breakthroughs in the understanding of photochemical phenomena, as these methodologies provide an objectively improved analysis tool for discovering patterns in excited-states molecular dynamics without requiring prior knowledge of the underlying chemical reaction mechanisms.

# On Electrons and Machine Learning Force Fields

Igor Poltavsky  
University of Luxembourg

Machine Learning Force Fields (MLFF) should be accurate, efficient, and applicable to molecules, materials, and interfaces thereof. The first step toward ensuring broad applicability and reliability of MLFFs requires a robust conceptual understanding of how to map interacting electrons to interacting "atoms". Here I discuss two aspects: (1) how electronic interactions are mapped to atoms with a critique of the "electronic nearsightedness" principle, and (2) our developments of symmetry-adapted gradient-domain machine learning (sGDML) framework for MLFFs generally applicable for modeling of molecules, materials, and their interfaces. I highlight the key importance of bridging fundamental physical priors and conservation laws with the flexibility of non-linear ML regressors to achieve the challenging goal of constructing chemically-accurate force fields for a broad set of systems. Applications of sGDML will be presented for small and large (bio/DNA) molecules, pristine and realistic solids, and interfaces between molecules and 2D materials. In particular, we highlight the effects of quantum interactions on the dynamics of these systems.

[Refs] *Sci. Adv.* 3, e1603015 (2017); *Nat. Commun.* 9, 3887 (2018); *Comp. Phys. Comm.* 240, 38 (2019); *J. Chem. Phys.* 150, 114102 (2019); *Sci. Adv.* 5, eaax0024 (2019).

# Solving the Schrödinger Equation in the Configuration Space with Generative Machine Learning

Basile Herzog, Bastien Casier, Sébastien Lebègue, and Dario Rocca  
*Université de Lorraine & CNRS, Nancy (France)*

The configuration interaction approach provides a conceptually simple and powerful approach to solve the Schrödinger equation for realistic molecules and materials but is characterized by an unfavourable scaling, which strongly limits its practical applicability. Effectively selecting only the configurations that actually contribute to the wavefunction is a fundamental step towards practical applications. We propose a machine learning approach that iteratively trains a generative model to preferentially generate the important configurations. By considering molecular applications it is shown that convergence to chemical accuracy can be achieved much more rapidly with respect to random sampling or the Monte Carlo configuration interaction method. This work paves the way to a broader use of generative models to solve the electronic structure problem.

## **Speeding-up quantum chemistry, and thus quantum molecular dynamics, with machine learning**

Nicolas SISOURAT

Laboratoire Chimie Physique Matière et Rayonnement (LCPMR), Sorbonne Université – CNRS, UMR 7614

Knowledge of the electronic properties of a molecular system is a prerequisite of any quantum molecular dynamics simulation. However, computing *ab initio* these properties is highly demanding, impeding the size of the systems that can be tackled and/or limiting the accuracy of the quantum molecular dynamics results. In this talk, I will present two theoretical approaches relying on machine learning techniques that we have recently developed to improve the efficiency and accuracy of electronic structure calculations.

In the first approach, we have employed a principal components analysis (PCA) to prepare an optimal set of molecular descriptors and to build an efficient Neural Network (NN) to predict molecular potential energy surfaces (PESs). Furthermore, the PCA provides a means to reduce the size of the input space (i.e. number of descriptors) without losing accuracy. This novel approach to the computation of high-dimensional PESs is illustrated on the example of the keto-enol tautomerism reaction occurring in the acetone molecule.

In the second approach, we have built and train NN models to predict the two-electron integrals of molecular systems. Computing and handling an enormous number of two-electron integrals is a common bottleneck of various electronic structure calculations. Therefore, there is a continuous search for always faster methods to compute these integrals. As will be shown on several examples, the use of NNs to predict these integrals can speed-up the computations by several orders of magnitude, opening unprecedented possibilities.

# Neural-network wave functions for quantum chemistry

Jan Hermann

*Freie Universität Berlin*

I will review variational quantum Monte Carlo as applied to arbitrary Hilbert spaces and Hamiltonians, and how wave-function ansatzes based on neural networks can be easily incorporated both in the first- and second-quantization formalisms. I will then demonstrate two applications to Hamiltonians relevant for quantum chemistry: First, an ab-initio electronic Hamiltonian for molecules in first quantization is solved with an antisymmetric ansatz that combines a physical baseline with a Jastrow factor and backflow parametrized by neural networks. Ground and excited states can be obtained with accuracy rivaling established quantum-chemistry methods. Second, a model exciton-phonon Hamiltonian in mixed first and second quantization is solved with an off-the-shelf convolutional neural network, improving upon state-of-the-art methods.

## Machine-learning-driven molecular dynamics for inorganic materials chemistry

Volker L. Deringer  
*University of Oxford*

Understanding the links between structures and properties of materials is an important research challenge. Molecular-dynamics simulations based on density-functional theory (DFT) have played important roles in this, but are typically limited to system sizes of a few hundred atoms. Machine learning (ML) based interatomic potential models are a rapidly emerging approach that aims to overcome this limitation: being "trained" on a suitably chosen set of DFT reference data, they achieve close-to-DFT accuracy whilst giving access to substantially larger system sizes.

In this presentation, I will showcase some recent advances in the modelling and understanding of materials that have been enabled by ML-driven molecular dynamics. I will argue that ML potentials are particularly useful for studies of structurally complex inorganic solids, for example, of non-crystalline (amorphous) phases. Examples of our ML-driven studies range from the structural characterisation of amorphous elements at ambient and high pressure [1, 2] to the computational modelling of materials for practical applications – for example, of battery electrodes. I will also discuss envisioned future directions: for the improved construction of reference databases, for the physically-guided validation of ML potentials, and for combining these emerging simulation tools with electronic-structure computations and experimental approaches.

[1] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, S. R. Elliott, *Nature* 2021, 589, 59.

[2] Y. Zhou, W. Kirkpatrick, V. L. Deringer, *Adv. Mater.* 2022, 34, 2107515.

# MLPT – an effective method for determining free energies at multiple electronic structure levels

Tomáš Bučko,<sup>1,2,\*</sup> Monika Gešvandtnerová,<sup>1</sup> Dario Rocca<sup>3,4</sup>

<sup>1</sup> Comenius University in Bratislava, Ilkovičova 6, 84115 Bratislava, Slovakia

<sup>2</sup> Slovak Academy of Sciences, Dúbravská cesta 9, 84236 Bratislava, Slovakia

<sup>3</sup> Université de Lorraine, LPCT, UMR 7019, F-54506 Vandoeuvre-lès-Nancy, France

<sup>4</sup> CNRS, LPCT, UMR 7019, F-54506 Vandoeuvre-lès-Nancy, France

\* Email : tomas.bucko@uniba.sk

The free energy belongs to the most important thermodynamic quantities used to characterize chemical reactions. Knowledge of the free energy of activation, for example, allows for a determination of the rate constant through the Eyring-Polanyi equation [1], thus providing access to reaction kinetics. Accurate calculations of free energies of chemical transformations are complicated by the necessity to sample the parts of the configuration space located between the reactant and the product, in which some relevant configurations receive Boltzmann weights that may be many orders of magnitude lower than those for configurations of stable states. This latter fact precludes direct use of standard molecular dynamics or Monte Carlo approaches, and therefore specialized simulation techniques, such as umbrella sampling [2], bluemoon ensemble technique [3], or metadynamics [4], must be employed instead. As all these techniques are very time consuming, the performance of various electronic structure methods in the free energy calculations is usually not examined. In this work, we propose a simulation protocol that combines the free energy perturbation theory [5] with a machine learning-based method MLPT introduced recently by Rocca et al. [6] that allows us to efficiently recompute the free energy determined at a certain level of theory using another electronic structure method. Importantly, such a calculation can be accomplished, without a significant loss of accuracy, by performing only a few tens of extra single-point energy calculations executed at the target electronic structure level, which may be computationally expensive, and for which energy gradients may not be available. The use of our method is exemplified on calculation of free energy of activation for the proton exchange in chabazite and methanol carbonylation in mordenite.

[1] P. Atkins and J. de Paula, Atkins' physical chemistry (Oxford university press: San Diego, 2002), pp. 956–963.

[2] G. Torrie and J. Valleau, J. Comput. Phys. 23, 187 (1977).

[3] E. Carter, G. Ciccotti, J. Hynes, and R. Kapral, Chem. Phys. Lett. 156, 472 (1989).

[4] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. 99, 12562 (2002).

[5] R. W. Zwanzig, The Journal of Chemical Physics 22, 1420 (1954).

[6] B. Chehaibou, M. Badawi, T. Bučko, T. Bazhurov, and D. Rocca, J. Chem. Theory Comput. 15, 6333 (2019).

# Observing the birth of crystals using machine-learning assisted simulations

Jacek Goniakowski,<sup>1</sup> Sarath Menon,<sup>2</sup> Gaetan Laurens,<sup>3</sup> and Julien Lam<sup>4</sup>

<sup>1</sup>Institut des NanoSciences de Paris, Sorbone Université, CNRS

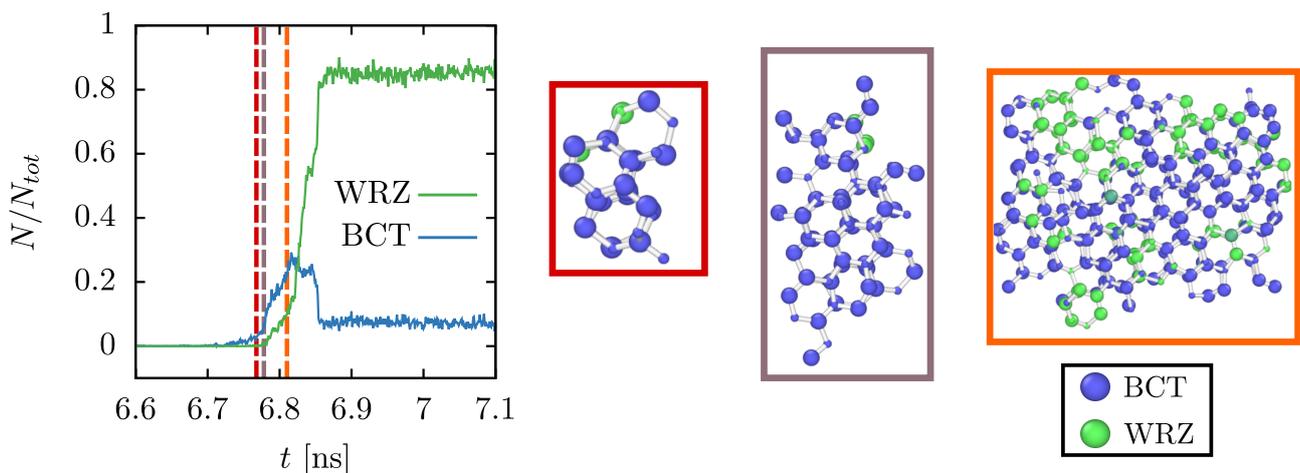
<sup>2</sup>ICAMS, Ruhr-Universität Bochum, Bochum, Germany

<sup>3</sup>Institut Lumière Matière, Université Lyon 1, CNRS,

<sup>4</sup>Centre d'élaboration des Matériaux et d'Etudes Structurales, CNRS

While crystals in material science are ubiquitous, the mechanisms of their formation which spans from nucleation to crystal growth remain one of the most intriguing process in nature. Better understanding crystallization would allow for a rational control of material engineering and possibly the development of novel functional materials and technological applications. From the fundamental point of view, numerous works have been dedicated to elucidating the emergence of the nucleation core and its role in controlling the final crystal structure. For instance, it is now possible to observe the crystal birth with electron microscopy, and colloidal science has also provided numerous experimental results on nucleation. Yet, while numerical simulations should have been a pivotal instrument to investigate crystallization at the atomistic level, studying nucleation require large scale simulations in order to observe the phase transition. As such, most works based on simulations have only focused on simple model materials thus preventing from targeting specific technological applications.

In this presentation, we will introduce innovative numerical tools based on machine-learning approaches and show how they can be employed to probe the nucleation processes in more complex materials. In particular, we focused on zinc oxide crystallization for which we developed a linear-based machine-learning interaction potential that can surpass all of the current classical models. Then, we combined large scale simulations and data-driven structure analysis to study the crystal nucleation. One key aspect of our work was to demonstrate that predictions of non-classical nucleation that were obtained with model materials including Lennard-Jones and hard spheres systems are still valid. Altogether, our results seems to demonstrate a universality of the crystallization processes ranging from soft to condensed matter.



**Figure:** Nucleation and growth of zinc oxide crystals

# Nonadiabatic dynamics in the long timescale: the next challenge in computational photochemistry

Mario Barbatti

*Aix Marseille University, CNRS, ICR, Marseille, France*

*Institut Universitaire de France, 75231 Paris, France*

*The Light and Molecules Group*

www.barbatti.org | mario.barbatti@univ-amu.fr | @MarioBarbatti

Nonadiabatic dynamics simulations in the long timescale (much longer than 10 ps) are the next challenge in computational photochemistry.<sup>1</sup> In this talk, we will explore the scope of what we expect from methods to run such simulations: they should work in full nuclear dimensionality, be general enough to tackle any molecule, and not require unrealistic computational resources. We will examine the main methodological challenges we should venture into to advance the field, including the computational costs of the electronic structure calculations, stability of the integration methods, the accuracy of the nonadiabatic dynamics algorithms, and software optimization. Based on simulations designed to shed light on these issues, we show how machine learning may be a crucial element for long-timescale dynamics, either as a surrogate for electronic structure calculations or aiding the parameterization of model Hamiltonians. We also show that conventional methods for integrating classical equations should be adequate for extended simulations up to 1 ns and that surface hopping agrees semi-quantitatively with wavepacket propagation in the weak-coupling regime. Finally, we describe our optimization of the Newton-X program to reduce computational overheads in data processing and storage.

[1] Mukherjee, et al. *Philos Trans R Soc A* **2022**, *380*, 20200382. (DOI: 10.1098/rsta-2020-0382)

# Accelerating and improving quantum chemistry and dynamics with machine learning

Pavlo O. Dral

*State Key Laboratory of Physical Chemistry of Solid Surfaces, Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Department of Chemistry, and College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China*

E-mail: [dral@xmu.edu.cn](mailto:dral@xmu.edu.cn)

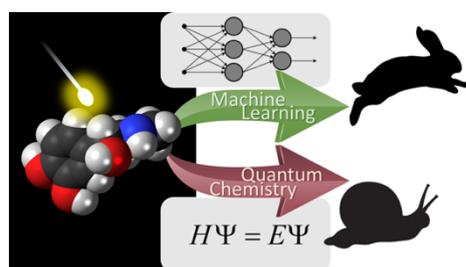


Figure 1. Machine learning speeds up quantum mechanical simulations.[1]

I will present our recent developments in applying machine learning to enhance quantum chemistry and (quantum) molecular dynamics simulations. One of these developments is general-purpose, artificial intelligence-enhanced quantum mechanical method 1 (AIQM1),[2] which approaches the accuracy of golden-standard, traditional CCSD(T)/CBS approach for closed-shell, neutral organic molecules in their ground state at the speed of semiempirical quantum mechanical methods while retaining good accuracy for charged systems and excited states.

This method does not need retraining and enables us to perform simulations we have not been able to do with either traditional quantum chemical approaches or with experimental techniques. Another development is our artificial intelligence-based quantum dynamics (AI-QD) approach which does not require iterative trajectory propagation.[3] Finally, I will talk about creating atomistic artificial intelligence models in 4D spacetime as an efficient tool for investigating molecular dynamics.[4] AIQM1 along with many other methods are implemented in our *MAtom* program package for user-friendly atomistic machine learning simulations which can be run online using our *MAtom @ XACS* (Xiamen atomistic computing suite) cloud-based service.[5]

- [1] P. O. Dral, M. Barbatti. Molecular Excited States Through a Machine Learning Lens. *Nat. Rev. Chem.* **2021**, 5, 388–405.
- [2] P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, P. O. Dral. Artificial Intelligence-Enhanced Quantum Chemical Method with Broad Applicability. *Nat. Commun.* **2021**, 12, 7022.
- [3] A. Ullah, P. O. Dral. Predicting the future of excitation energy transfer in light-harvesting complex with artificial intelligence-based quantum dynamics. *Nat. Commun.* **2022**, 13, 1930.
- [4] F. Ge, L. Zhang, A. Ullah, P. O. Dral. Four-dimensional spacetime atomistic artificial intelligence models. **2022**, preprint on ChemRxiv: <http://doi.org/10.26434/chemrxiv-2022-qf75v>
- [5] P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro Jr, J. Huang, M. Barbatti. *MAtom 2: An Integrative Platform for Atomistic Machine Learning*. *Top. Curr. Chem.* **2021**, 379, 27. See <http://mlatom.com> @ <http://xacs.xmu.edu.cn>.

# Progress and Challenges in Modelling Photoinduced Processes

Nada Došlić

<sup>1</sup>*Department of Physical Chemistry, Ruđer Bošković Institute, 10000 Zagreb, Croatia*

*E-mail: nadja.doslic@irb.hr*

Over the years, two main families of methods to solve the time-dependent Schrödinger equation have been developed: fully quantized electron-nuclear dynamics such as multi-configurational time-dependent Hartree (MCTDH) and mixed quantum-classical dynamics such as trajectory surface hopping (TSH). Since MCTDH requires high-accuracy pre-calculated PESs and nonadiabatic coupling matrices, the method is suitable for describing comparatively small systems. To examine molecular dynamics of polyatomic systems in full dimensionality, TSH is the method of choice. A comparison between these methods is made difficult by the fact that MCTDH calculations are run in the basis of diabatic states, while TSH simulations are performed in the adiabatic basis.[1,2] A possible approach would be the computation of the diabatic populations in TSH simulations. We will discuss algorithmic and computational challenges encountered when computing diabatic populations in TSH simulations

New light sources have allowed the investigation of light-induced processes in molecules with a range of spectral probes. Due to the complexity of the measured signals, the interpretation of these experiments requires accurate simulations. Here we present a protocol for simulating time resolved photoionization spectra of polyatomic molecular systems, which combines TSH simulations with the doorway-window (DW) approximation for the evaluation of spectroscopic signals, and apply it to systems of current interest. [3,4]

## References

- [1] W. Xie *et al.* *J. Chem. Phys.* **150**, 154119 (2019)
- [2] J. Coonjobeeharry *et al.* *Philos. Trans. Royal Soc. A* **380**, 20200386 (2022)
- [3] M. F. Gelin *et al.* *JCTC*, **17**, 2394 (2021)
- [4] T. Piteša *et al.* *JCTC*, **17**, 5098 (2021)

## **Full quantum (thermo)dynamics with integrated machine learning models**

Michele Ceriotti - EPFL

Machine learning is finding applications to more and more tasks, in science as much as in everyday life. In this talk I will focus on how atomic and molecular simulations are being transformed by the use of statistical regression models, that make it possible to approximate accurately and efficiently atomistic properties computed from a few reference electronic-structure calculations.

I will argue about the advantages that are brought about by an integrated framework, that makes it possible to access not only the interatomic potential, but all sort of properties from NMR chemical shieldings to the electronic charge density, and provide examples from molecular to condensed-phase systems.

In particular, I will show how integrated models facilitate computing the (thermo)dynamic properties of condensed phases incorporating fully the quantum mechanical behavior of both electrons and nuclei.

# Computational materials engineering with active learning

Milica Todorovic  
*University of Turku*

Data-driven materials science based on artificial intelligence (AI) algorithms has facilitated breakthroughs in materials optimization and design. Of particular interest are active learning algorithms, where datasets are collected on-the-fly in the search for optimal solutions. We encoded such a probabilistic algorithm into the Bayesian Optimization Structure Search (BOSS) Python tool for materials research. We utilized this versatile tool to study molecular surface adsorbates, thin film growth, solid-solid interfaces, molecular conformers and even optimise experimental outcomes. New algorithm developments will allow us to harness the power of exascale computing platforms for next-generation materials engineering.

# FCInet : An iterative machine learning classification neural network for a smartly selective CI.

Bastien Casier<sup>†</sup>, Basile Herzog<sup>†</sup>, Dario Rocca<sup>†</sup>

<sup>†</sup> *Univ. de Lorraine, CNRS, LPCT, UMR 7019, F-54000 Nancy, France*

Resolving in a straight manner the Full Configuration Interaction (FCI) equation stays an impossible way. The impracticable nature of this problem comes from the exponential growth of the number of Slaters determinants according to the number of orbitals. For instance, the size of the Hilberts space relative to the ethane (C<sub>2</sub>H<sub>6</sub>) molecule reaches more than one million of determinants in a minimal basis set (e.g. STO-3G). Consequently, a simple solution consists in limiting the full expansion of the FCI wave function to simply and doubly excited determinants (CISD). Nevertheless, this truncation, resulting from an arbitrary choice, may be too severe, especially at dissociation limits where a large amount of determinants becomes degenerated [1].

Another strategy consists in selecting the most important Slaters determinants. To do so, the most popular method is the Configuration Interaction using a Perturbative Selection done Iteratively (CIPSI) [2]. This approach presents some outstanding results at the equilibrium geometries and around. However, the perturbative nature of the method yields a lack of accuracy for large interatomic distances.

In this work we introduce an iterative machine learning classification algorithm to smartly select the most important determinants. The learnable information is encoded in a binary representation of the spin-orbital populations of each Slaters determinant, while the learning is based on the minimization of the binary cross-entropy. This method presents very promising results for small molecules (e.g. CO, N<sub>2</sub>, H<sub>2</sub>O, NH<sub>3</sub>, C<sub>2</sub>H<sub>6</sub>, etc.), as FCI accurate energies have been obtained along all the dissociation curve (see Figure 1) by only using a small fraction of the Hilberts space. This primary study is a proof of principle of the binary classification for the selective CI. Our end goal would be to realize a global learning on the integral curves and to define the first interatomic potentials at the FCI.

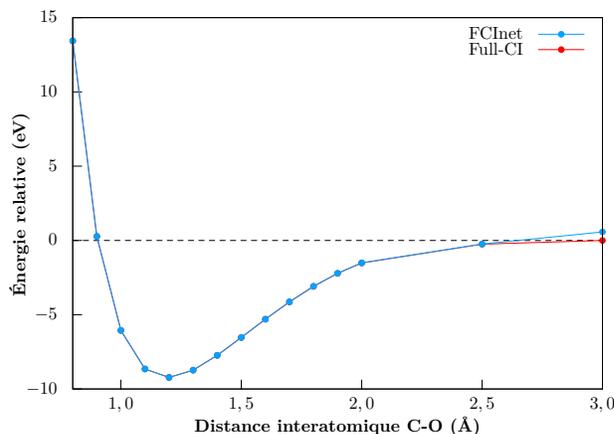


Figure 1: Dissociation curve for carbon monoxide.

## References

- [1] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. First. Dover Publications, Inc., 1996.
- [2] B. Huron, J. P. Malrieu, and P. Rancurel. "Iterative perturbation calculations of ground and excited state energies from multiconfigurational zerothorder wavefunctions". In: *The Journal of Chemical Physics* 58.12 (1973), pp. 5745–5759.

# Predicting electronic structure properties with equivariant and long-range machine-learning methods

Andrea Grisafi<sup>1</sup>

<sup>1</sup>*PASTEUR, Département de Chimie, École Normale Supérieure, 75005 Paris, France*

The accurate data-driven prediction of electronic-structure properties requires machine-learning models that incorporate the fundamental symmetries of the physical target while presenting a consistent treatment of long-range interactions. In the first part of the talk, I will introduce a family of equivariant structural representations that can be used in the regression of electronic response tensors, such as the polarizability of a system, paving the way to the quantum-level simulation of absorption spectra. I will then show how to use the same family of equivariant descriptors to set up the highly-transferable regression of the electronic charge density expanded on a basis of atom-centered spherical harmonics, in such a way that the densities of both molecular and condensed-phase systems can be accurately predicted with linear-scaling cost. When used in conjunction with density functional theory, the predicted densities can be used to access extremely accurate electronic energies as well as any other ground-state property at the cost of a single Kohn-Sham diagonalization step. In the second part of the talk, I will introduce a family of long-range structural representations that can be used to complement local machine-learning potentials by rigorously incorporating electrostatic interactions while also capturing other kind of non-local effects. I will finally show how to use a finite-field extension of these long-range descriptors to simulate the electronic response of model electrodes under an applied electric field.