DISSERTATION / DOCTORAL THESIS

Titel der Dissertation /Title of the Doctoral Thesis

„Machine Learning for Excited-State Molecular Dynamics Simulations"

verfasst von / submitted by

Julia Maria Westermayr, MSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Doktorin der Naturwissenschaften (Dr.rer.nat.)

Wien, 2020 / Vienna 2020

In loving memory of my mother.

ACKNOWLEDGMENTS

ABSTRACT

With the rise of machine learning in quantum chemistry, many exciting opportunities have emerged to advance the research field of photochemistry. However, the high complexity and computational efforts associated with a description of coupled electrons and nuclei complicate an application of machine learning techniques, such that this research field remained mainly untouched until the beginning of this work. The main goal of this thesis is to employ machine learning in the simulation of photodynamics. To this aim, machine learning models are developed to describe a manifold of molecular excited-state potentials of different spin multiplicities, their derivatives, and couplings thereof. Central to this thesis is the identification of key obstacles in applying efficient, yet accurate machine learning for photodynamics and overcoming the following hurdles: (1) Excited-state properties carry an arbitrary sign that prohibits a meaningful fitting. These properties can be rendered learnable with a so-called phase correction algorithm developed in this work. As an alternative, a machine-learning intrinsic solution is implemented in a new approach called SchNarc. (2) Couplings between electronic states are often missing in quantum chemistry codes. Therefore, approximations based on machine-learned potentials are provided. (3) Due to the lack of machine learning studies on this topic, many open questions remain on how to best treat excited states. Thus, various models and molecular representations are compared. (4) A cost-effective training set generation is required that still allows for a desired accuracy. In order to meet this requirement, an existing efficient scheme for the ground state is adapted for the excited states. The performance of all these novel strategies and methods is investigated using several small molecules. One big achievement of this work is to enable long time scale photodynamics with neural networks for the methylenimmonium cation. The complex photochemistry of the amino acid tyrosine illustrates the challenges machine learning models still need to overcome, but demonstrates the scopes and possibilities of machine learning for excited electronic states of molecules.

ZUSAMMENFASSUNG

Durch das Aufeinandertreffen der künstlichen Intelligenz und Quantenchemie haben sich eine Vielzahl neuer Wege eröffnet, von denen besonders das Forschungsgebiet der Photochemie profitieren kann. Das Zusammenspiel von Licht, Elektronen und Atomkernen ist jedoch so komplex, dass der hohe Rechenaufwand der mit einer akkuraten Beschreibung verbunden ist, dazu führte, dass dieses Forschungsgebiet bis zu Beginn dieser Arbeit von maschinellem Lernen weitgehend unberührt blieb. Das grundlegende Ziel dieser Arbeit ist es daher, die Techniken des maschinellen Lernens für Photodynamiksimulationen anzuwenden und weiterzuentwickeln, um eine Vielzahl von molekularen elektronischen Zuständen unterschiedlicher Spin-multiplizitäten, deren Ableitungen und Kopplungen zu beschreiben. Im Mittelpunkt dieser Arbeit steht die Identifizierung der Hauptprobleme, die einer Realisierung von effizienten Modellen des maschinellen Lernens für die molekulare Photochemie im Weg stehen, sowie deren überwindung: (1) Eigenschaften, die sich durch die Wechselwirkung unterschiedlicher elektronischer Zustände ergeben, sind durch ein willkürliches Vorzeichen gekennzeichnet, das eine sinnvolle, direkte Anwendung von maschinellen Lerntechniken verhindert. Die Willkür bei Vorzeichen kann mit einem in dieser Arbeit entwickelten Phasenkorrektur-Algorithmus beseitigt werden. Als Alternative dazu ist eine modell-intrinsische Lösung in einem neuen Ansatz namens SchNarc implementiert. (2) Viele quantenchemische Programme können keine Kopplungen zwischen elektronischen Zuständen berechnen. Daher werden Näherungen auf Basis maschinell gelernter Potentiale bereitgestellt. (3) Aufgrund der fehlenden Studien zur Beschreiben der Photochemie mit maschinellem Lernen, bleiben viele Fragen offen, unter anderem, welche Algorithmen und molekularen Repräsentationen für angeregte Zustände am besten geeignet sind. Aus diesem Grund werden mehrere solche Modell des maschinellen Lernens verglichen. (4) Aufgrund des hohen Rechenaufwands quantenchemischer Methoden wird eine kostengünstige Trainingssatzgenerierung benötigt. Daher wird ein effizienter Algorithmus für den elektronischen Grundzustand für die Photochemie erweitert. Die Möglichkeiten und Präzision all dieser neuartigen Strategien und Methoden wird anhand kleiner Moleküle untersucht. Lange Photodynamiksimulationen werden mit neuronale Netzen für das Methylenimmonium-Kation ermöglicht. Die komplexe Photochemie der Aminosäure Tyrosin fasst die Herausforderungen für Methoden des maschinellen Lernens zusammen und demonstriert dessen Möglichkeiten, angeregte elektronische Zustände von Molekülen zu beschreiben.

CONTENTS

Adam   Adaptive Moment Estimation

ADC(2)   Second-Order Algebraic Diagrammatic Construction Scheme

AI   Artificial Intelligence

CASSCF   Complete Active Space Self-Consistent Field

CASPT2   Complete Active Space Perturbation Theory to Second Order

CI   Configuration Interaction

CIS   Configuration Interaction Singles

CISD   Configuration Interaction Singles Doubles

CPU   Central Processing Unit

DFT   Density Functional Theory

FCHL   Faber-Christensen-Huang-Lilienfeld

FCI   Full Configuration Interaction

fs   Femtoseconds

GPU   Graphics Processing Unit

HOMO   Highest Occupied Molecular Orbital

KRR   Kernel Ridge Regression

LR   Linear Model

LUMO   Lowest Unoccupied Molecular Orbital

LVC   Linear Vibronic Coupling

MAE   Mean Absolute Error

MD   Molecular Dynamics

ML   Machine Learning

MLFF   Multi-Layer Feed-Forward

MLMD   Machine Learning Molecular Dynamics

MQC   Mixed Quantum-Classical

MP2   Møller-Plesset Perturbation Theory to Second Order

MR   Multi-Reference

MSE   Mean Squared Error

NAC   Nonadiabatic Coupling

NN   Neural Network

ns   Nanoseconds

PES   Potential Energy Surface

ps   Picoseconds

PT   Perturbation Theory

SCF   Self-Consistent Field

SHARC   Surface Hopping including ARbitrary Couplings

SOC   Spin-Orbit Coupling

SVR   Support Vector Machine for Regression

TDDFT   Time-Dependent Density Functional Theory

UV   Ultraviolet

WFT   Wave Function Theory

INTRODUCTION

The research field of photochemistry focuses on reactions that are inherently associated with light. Since light is ubiquitous, it is not surprising that many processes that are fundamental to nature and the life we know are related to reactions triggered by light or are a direct consequence thereof.[1;2] The most obvious example is photosynthesis, one of the most powerful reactions in nature, which enables plants to convert the greenhouse gas carbon dioxide into sugar molecules and oxygen.[3–5] The ability to harness light energy, which is central to photosynthesis, can be mimicked by nanoparticles or thin films of noble metals, such as gold or silver.[6] The subsequent transfer of energy to attached molecules at their interface has been identified as extremely useful for photocatalysis[7–9] or assays and sensor recognition.[10] Other examples of light-induced reactions are the development of skin cancer due to the photodamage of the genetic material of our body, namely our DNA,[11–14] the ultrafast, radiationless reactions that enable us to see,[15;16] or the reactions that provide the photostability of amino acids that give rise to the multifaceted photochemistry of proteins and peptides.[17–19]

These few examples already indicate that photochemistry of molecules and materials surrounds us almost everywhere and almost anytime. The knowledge of photochemical processes that take place in our close environment and our body is therefore extremely important to understand basic concepts of nature and life. For example, being able to understand what triggers harmful reactions in molecules that can lead to serious diseases such as skin cancer[11;12;14;20] or cataracts[20–22] can help us to find ways to prevent such reactions from happening. In the case of skin cancer, a concrete example is the search for better and more effective sun screens to protect our skin from UV irradiation.[23;24] In addition, the investigation of light-induced reactions can foster the exploration of the relationships between the structure of a system and its photochemical properties. Very early on, structure-property correlations have been identified as very powerful to guide the creation of new compounds with tailored properties.[25;26] A design of novel molecules and materials can be of high value for phototherapy,[27] photocatalysis[28] or photovoltaic devices.[29;30]

As diverse as the aforementioned processes might seem, as many things they have in common: They are all governed by a manifold of excited states that become accessible after enough energy is provided. Obviously, in the case of photochemical reactions, the energy source is light. The basic requirement for any photo-initiated reaction to happen is therefore the ability of a compound to absorb light and to enter a higher electronic state, i.e., to allow for a transition from the ground state potential energy hypersurface (PES) to an excited-state PES. The PESs are functions of nuclear coordinates with local minima and intersections of different PESs. The dimension of one such PES is $3N - 6$, with N referring to the number of nuclei in a system. Hence, the larger the system under investigation, the higher dimensional and more complex the PESs become.[2;26]

A dilemma arises, when the mechanistic details of light-driven reactions that take place in realistic systems are the focus of a study. Dependent on the type of system under investigation, at least several thousands to millions of atoms have to be considered.[31;32] The problem becomes clearer, when taking a look at our body. Our skin as well as our eyes are primary targets of environmental stress, such as

sunlight. One of the functions of our skin is therefore to protect our body from photodamage and other harmful reactions, while our eyes have the main function to enable our vision.[21;33] The investigation of the underlying processes that lead to the photostability of our cells and tissues is not a trivial task as both, our skin and our eyes, are highly complex organizations. Only an approximate picture of the response of our skin to certain stimuli can be obtained with experiments, such as imaging techniques like confocal laser scanning microscopy[34] or fluorescence resonance energy transfer.[35] The UV absorption can be measured for instance with laser optoacoustics[36] in vivo or via UV/visible spectroscopy ex vivo.[37] In case of our eyes, spectral sensitivity measurements can provide information on their response to light.[38] However, many experimentally probed spectra require expert knowledge to allow for a profound interpretation and analysis of the underlying electronic mechanistic details is difficult. Theoretical investigations can help to interpret results, but the large size and complexity of realistic systems make the investigation of their light-induced reactions on a molecular level not only challenging, but in most cases not practicable, if not even impossible.[26;39] In order to get an idea of what leads to photodamage and photostability, it is therefore inevitable to take a closer look at the composition of a system and to find smaller building blocks, i.e., chromophores, that can be considered as the starting points of light-induced reactions.[1;36;40–45]

Getting back to the example of our skin, several such building blocks that can absorb light, exist. As our skin is made up of tissue, that itself is made up of cells, it contains not only a large amount of water and ions, but comprises several macro-molecules. These macro-molecules are mainly carbohydrates, lipids, nucleic acids and proteins that all can absorb UV or visible light and thus play a role in the photochemistry of our skin.[33;37] Not much is known about the exact processes underlying the photo-initiated reactions, but what seems to be clear up to date is that ultrafast transitions occurring in the chromophores of such systems can potentially prevent photodamage and provide -stability.[36;41–44]

Especially proteins are potential targets of light as they make up about 68% of the dry weight of our cells and tissues and are important components of our skin and our eyes. For instance, a cis-trans isomerization of the chromophore retinal of the protein rhodopsin is fundamental for our vision.[15;45] Moreover, the amino acids phenylalanine, histidine, tryptophane and tyrosine can provide the starting points of photochemical reactions that take place in peptides and proteins. Their photochemistry is mainly governed by their ability to transform absorbed UV light into vibrational energy on an ultrashort time scale, usually in the range of femto- (fs) to picoseconds (ps). These ultrafast radiationless transitions are assumed to provide the photostability of these building blocks. If photodamage nevertheless occurs and photoionization or radiative reactions cannot be quenched efficiently, harmful reactions that can lead to the unfolding, aggregation or inactivation of enzymes and peptides are reported. A loss of their function is the consequence, which is related to the alteration of our skin as well as blindness due to cataracts.[21;31;44;46;47]

Experimentally, pump-probe experiments, transient absorption or photoioniza-tion spectroscopy along with ultrafast electron diffraction are powerful techniques to study their light-driven bond cleavage or the lifetime of their electronic excita-tions.[13;20;42;48–53] However, the mechanistic details of most experimental observations are obscured by the measured ensemble averages.[54] No clear picture on the key events that lead to photodamage or -stability can be obtained.

The picture can be completed with theoretical simulations that have the potential to shed light on the missing details underlying light-driven reactions.[26] For exam-

ple, the theoretical simulation of UV spectra as probed by experiments can unveil the relevant states that lead to light absorption.[55–63] As the photochemistry that surrounds us usually concerns the changes a system undergoes, excited-state molecular dynamics (MD) simulations are particularly well suited to study the temporal evolution of a system after light excitation.[14;54;64–71] They can be used to explore the high-dimensional PESs that become accessible via light absorption or to locate relevant structures that lead to ultrafast transitions in molecules. Comprehensive sampling of varying reaction conditions, such as starting geometries, from which the excitation takes place, can provide a qualitative picture of reaction kinetics, reaction probabilities and the accessibility of different reaction channels of a system.[72] Research questions that can be tackled are for example which reactions take place as a consequence of light excitation and with which probability do they occur? What geometrical features are important for photostability and what triggers harmful reactions? The qualitative results can be compared to experiments and together, a far more comprehensive picture can be obtained.[1;2;13;26;31;44;73]

As the previous discussion makes clear, a complementary theoretical investigation of experimental observables is important to provide a better understanding and a prediction of the photostability and damage of molecules. Unfortunately, both, experimental techniques and theoretical simulations, share the characteristic of being extremely expensive.[39] While experiments require elaborate and costly setups, theoretical simulations require high-performance computing facilities. In case of MD simulations the large number of sequential calculations that is required to study the temporal evolution of a system makes this process expensive from a computational point of view. The high associated effort seriously hampers the application of conventional excited-state dynamics simulation methods to long time and large length scales. When treating all degrees of freedom of a system in the most exact way, i.e., exact quantum dynamics combined with high-level electronic structure calculations of a molecular system, photodynamics studies have so far been limited to systems containing less than 5 atoms.[26;74–76]

In order to allow for the simulation of systems comprising more degrees of freedom, approximations become inevitable. The most fundamental approximation underlying almost all existing electronic structure programs is the Born-Oppenheimer approximation, which allows a separation of the nuclear and the electronic degrees of freedom. Another approximation that is important when larger systems are targeted is the neglect of quantum effects in the nuclear motion of molecules. This approximation led to the development of so-called mixed quantum-classical (MQC) MD simulation techniques. As the nuclei are assumed to move classically on the PESs made up by the electrons, MQC methods are capable of describing systems with up to several 100s of degrees of freedom. Still, they can retain some quantum effects. For example, nonadiabatic transitions are allowed between different PESs.[1;2;26;72] With these approximations applied, simulation times of several ps become accessible with *ab initio* accuracy. It should be kept in mind that the classical evaluation of the nuclear motion then requires statistical averaging over many reaction events.[1;2;26]

Nevertheless, even with the approximations applied in MQC photodynamics simulation techniques, the theoretical investigation of systems at experimentally relevant time (in the range of nano- (ns) to milliseconds (ms)) and length scales (in the range of thousands to millions of atoms) is beyond the realms of possibility.[31;39;44] The limitation to much shorter length and time scales is a consequence of the rapidly growing computational efforts of quantum chemical calculations with the number of degrees of freedom of a system. The number of relevant electronic states

additionally complicates the computation and increases computational costs.[1] Even the dynamics simulation of one single amino acid in the gas phase is not feasible with conventional approaches at experimentally relevant time scales, let alone the inclusion of any environmental effects. Even with massively parallel computations, a MQC photodynamics simulation with *ab initio* accuracy for one ns would require at least several years of computation.[77]

As things stand now, the aforementioned limitations of existing MQC photodynamics simulations and the nonlinearly increasing computational costs with the complexity and size of a system under investigation prohibit a viable application of MQC simulations to realistic time and length scales. Not only environmental effects are neglected, also a lot of approximations are applied. Only a small glimpse of reality can be obtained with the existing techniques and computer power available. One research field that might be able to push the boundaries of existing simulation techniques and enable photochemical simulations of larger molecular systems on longer time scales is machine learning (ML). ML belongs to the larger research field of artificial intelligence and offers the possibility to learn from a massive amount of data, which is far too complex to be analyzed with typical human reasoning.[78;79] Without knowing the underlying physics of data, ML models have the potential to explore underlying patterns and fit highly complex relationships of given data.[80–83] Due to the growing availability of computer power and efficient algorithms, ML has gained increasing popularity among various research fields. The powerful combination of efficient ML models with huge amounts of data is even considered to be the fourth big paradigm of science.[84]

Already more than 25 years ago, ML has entered the research field of quantum chemistry,[85;86] but early studies were restricted to model the electronic ground state of a molecular system.[87] Therefore, many novel methods in the research field of ML for quantum chemistry target the electronic ground state of molecular or material systems. Up to date, ML has proven very powerful to advance quantum chemistry for the electronic ground state in the most diverse ways.[81;88;89] For example, ML models exist that make the computation of the solution of the electronic Schrödinger equation more efficient, by e.g. providing an ML approximation to the molecular wavefunction, either in a molecular orbital[90;91] or spin-state basis[92;93], on a grid[94] or within a Monte-Carlo approach.[95–97] Solving the Schrödinger equation can also be skipped by directly fitting PESs of molecules and materials to provide ML-based force fields for energies and dipole moments.[81;90;98–137] ML-fitted functions that relate structures to certain properties further enable the search for new compounds throughout chemical compound space.[28;111;138–141]

In contrast to the exciting opportunities ML models offer for simulations in the electronic ground state, the benefits that ML models can offer for simulations in the electronically excited states are still mostly unexplored. Some early attempts exist and involve (modified) Shepard interpolation[142] for grid-based quantum dynamics simulations, least squares fitting of excited-state PESs[143], and artificial neural networks (NNs)[144;145] for MQC simulations. Besides these few seminal works, not much effort had previously been devoted to develop ML models for the excited electronic states of molecules. Very recently, driven by the aforementioned achievements of ML models for the electronic ground state, the interest to advance photochemical simulation techniques is increasing and ML is revisited for the excited states. Researchers have started to tackle the fitting of energies, their derivatives and excited-state properties mainly in the last few years.[146–152] Pre-fitted ML-PESs are especially tantalizing for photodynamics simulations. Their use offers the possibility to disentangle the high

costs of electronic structure calculations from the MQC MD simulations.[1;153] This separation opens avenues to study reactions that were thought to be infeasible only some years ago and long time scale simulations with *ab initio* accuracy come to the fore. MQC MD simulations in the excited states with ML models have been published mainly in the last three years and comprise small molecular systems or simple reactions.[77;154–160;160–163] At the current stage of research, only a few studies focus on ML for the excited electronic states of molecules. The lack of studies concerning the investigation of the photochemistry of molecules is also one of the main motivations of this thesis.

The reason for the slow development of ML models for the electronic excited states, especially for the simulation of photodynamics, compared to ML models for the electronic ground state and Born-Oppenheimer dynamics is manifold: First of all, an accurate investigation of the photochemistry of a molecular system requires the treatment of many excited states of different spin multiplicity,[1;2;26;164] which makes reference computations generally more expensive. Further, derivatives of different PESs have to be calculated and electron-nuclear couplings between different electronic states arise. The computation of such couplings is extremely costly and complicated, as they show singularities in conformational regions of the excited state PESs, when two electronic states are degenerate. This degeneracy and the strong electron-nuclear coupling lead to the break-down of the Born-Oppenheimer approximation, making it difficult to converge a quantum chemical reference calculation in such critical regions.[165] Furthermore, fitting inconsistent data with singular values exactly is not possible with smooth ML functions. Another problem arises due to the excess of energy as a consequence of light absorption, which results in a higher probability of reactions involved in the formation and breaking of bonds. Many "black-box" quantum chemical methods, which are frequently used for electronic ground state problems, cannot be used for the excited states anymore, as they lack a qualitative correct descriptions of such reactions. In this regard and due to the likely contribution of many electronic configurations for the excited-states, highly accurate and costly *ab initio* methods are required in many cases.[26;164] The search for a proper reference method, that is on one side accurate enough to treat processes involving the formation and breaking of bonds, and on the other hand computationally efficient to provide the amount of reference data needed, is a problem on its own. Many highly precise quantum chemistry methods for the excited states do not allow for a "black-box" use anymore and their use requires expert knowledge. These challenges, which become apparent for the excited-state calculations, but usually do not need to be considered for the ground state, also hamper the application of ML models.

Nevertheless, assuming that a proper method is selected carefully that meets all the requirements for accurate reference simulations, several additional questions arise that must be considered before computing a training set and fitting an ML model. For example, the treatment of the excited state PESs using a finite number of electronic states can lead to inconsistencies in the PESs, which is especially problematic when molecules with a high density of states are treated and in conformational regions of strong electron-nuclear couplings. How to confront these conformational regions with ML is one of the questions to be answered. How can vectorial properties, such as dipole or coupling vectors, be fitted that are still rotationally covariant? What is the best way to treat many electronic states? Should all states and all properties be trained within one ML model or is it more advantageous to treat them independently from each other and in a single-state fashion? How can the molecules be represented to the ML model to accurately fit a manifold of excited-state PESs and properties? Is

it possible to describe molecules using atomic fragments in different chemical and structural environments and to construct excited-state PESs from atomic contributions as it is done for many existing, accurate ground state PESs?

While the enumeration of (open) questions could be continued, the aforementioned questions are identified to be at the forefront of questions to be answered when developing ML PESs for the excited states of molecules. Due to the young age of this research field, most of the questions were unanswered in the beginning of this work and have been targeted while conducting this thesis.

The main goal of this thesis is to provide an answer to the aforementioned questions and to find solutions for fitting the excited-state PESs and related properties of molecules. Special focus is placed on advancing the exciting and highly active research field of photochemistry by providing an integration of ML into existing approaches and enabling long time scale photodynamics simulations. The central model system of this thesis is the methylenimmonium cation, $CH_2NH_2^+$, the smallest member of the family of protonated Schiff bases, to which also retinal, the chromophore of rhodopsin, belongs. $CH_2NH_2^+$ is used theoretically to model the ultrafast rotation around the double bond. This rotation also plays a role in the cis-trans isomerization of retinal, which is fundamental for vision.[15;64] The methylenimmonium cation accompanies almost all method developments carried out in this thesis. Much effort is devoted to develop a method for photodynamics simulations, which is user-friendly and can be used by researchers without expert knowledge on ML. The developed ML methods are implemented in the Surface Hopping including ARbitrary Couplings (SHARC) MD code[72;166] and can potentially evolve as a valuable tool for the study of the photochemistry of molecular systems. Along the way, test systems with different spin-multiplicities are selected. The different photochemistry of these systems mainly serves the method development, rather than the exploration of different reactions.

After the performance of newly developed methods and strategies is assessed using small molecular systems, the excited-states of the amino acid tyrosine are discussed. The investigation of the photochemistry of tyrosine with ML has been planned since the beginning of this thesis, but has been interrupted several times due to existing problems with quantum chemical calculations and ML models. The related problems are illustrated. It is shown, how ML can overcome some problems within quantum chemistry by using different quantum chemical reference methods for the training set generation, which complement each other in terms of accuracy and computational efficiency. Effort is devoted to generate hand-tailored excited-state PESs and correct some problems of quantum chemical methods that prohibit a comprehensive training set generation for the excited states of tyrosine.

THEORY

The following chapter aims to summarize the theoretical background of the techniques and methods that were applied in this thesis. It is divided into 2.1 Quantum Chemistry, 2.2 MD, and 2.3 ML approaches. The execution of ML based MD (MLMD) simulations in the excited states is one goal of this thesis. The term MLMD is used throughout this thesis to indicate that ML models are used for respective MD simulations instead of the reference quantum chemical simulations. To enable MLMD, a comprehensive, yet accurate training set has to be provided that is at the basis of any ML model. Therefore, a short overview of different quantum chemical methods that are evaluated and applied for the computation of the reference data is given. Emphasis is placed on comparison of quantum chemistry methods with respect to their accuracy, efficiency as well as general applicability for the generation of a training set for ML. A short description of MQC simulations follows. The chapter concludes with a brief overview of the different types of ML models and descriptors used. NNs, their high-dimensional variants, and kernel methods are introduced and their respective merits and pitfalls are discussed.

2.1 QUANTUM CHEMISTRY FOR ELECTRONIC EXCITED STATES

In a nutshell, the purpose of electronic structure theory is to provide a framework to enable the computation of the energy of a system and its properties. The Born-Oppenheimer approximation[167] forms the basis of such calculations and is also assumed throughout this thesis. Within the Born-Oppenheimer picture, the electrons move much faster than the nuclei and are therefore considered to move in a constant field of fixed nuclear charges and positions, which is a consequence of their much lighter weight. In this way, the nuclear and electronic degrees of freedom can be separated and electronic structure calculations can be carried out for a set of fixed nuclear positions and a given set of electronic states. Many such independent static calculations with different nuclear positions can then be executed to build the PESs of a system. Having comprehensive and accurate PESs, in the best case for many electronic states including several excited-state properties, is very powerful, because it enables the investigation of the photochemistry of a system in a computationally efficient way and allows for the application of many existing MD simulation methods. However, a bottleneck is not only the search for relevant conformations of a system for which the electronic structure is to be calculated, but also the search for a proper method to calculate the electronic structure of a system.[1;2;168;169]

Existing electronic structure implementations can be differentiated in two main approaches: Wave Function Theory (WFT) and Density Functional Theory (DFT). This classification has been originally introduced by Kohn in his Nobel Lecture[170] after he won the Nobel Prize in Chemistry in 1998 for his developments in DFT. Complementary, Schrödinger won the Nobel Prize in Physics in 1933 for postulating the Schrödinger equation, fundamental to WFT. Both theories have been tremendously important for our understanding of many chemical problems and they have in common, that at least in principle, an exact solution to a chemical problem can be obtained. However, an exact solution is only reachable for systems as simple as H_2^+ in case of WFT and fails even for such simple systems with DFT.[171] The problem

is that the equations are not solvable in case of WFT and that the equations to be solved are not known in case of DFT. At this point, many approximations have been introduced.[172]

2.1.1 *Density Functional Theory (DFT)*

The method of choice for many studies that target complex and large systems is DFT. DFT-based methods are also the preferred workhorse for the calculation of training sets for ML models.[101;129;173–175] "That is because density functional theory, when it works, is the most affordable way to get reasonably reliable and useful accuracy on such problems" (H. S. Yu, *et al.*, *J. Chem. Phys.*, **145**, 130901 (2016)). The most widely implemented approaches of DFT and its time-dependent variant (TDDFT) to compute excited states rely on the Hohenberg-Kohn theorems[176] and the Kohn-Sham approximation[177] and are single-reference in their nature.[171] Within Kohn-Sham DFT, the energy is expressed as a functional of the electron density of a fictitious system. This fictitious system comprises noninteracting electrons and has the same density as the exact system. Assuming that the ground state density of a system is known, the exact ground state energy of a system can be calculated. Unfortunately, the ground state density of a system is generally unknown and approximations have to be imposed to estimate the density and thus the energy of a system. Within an orbital expression, the energy is computed as the sum of kinetic energies of the individual electrons of the fictitious system using the density obtained from a single Slater determinant, the Coulomb energy, and a so-called exchange-correlation energy. The exchange-correlation energy is a functional of the electron density and contains the electron-electron repulsion, the electron exchange, and the corrections to the kinetic energy, i.e., the difference of the real kinetic energy to the one of the fictious system.[171;178]

In theory, a unique exchange-correlation functional (also called Hohenberg-Kohn functional) exists and the holy grail of DFT is to find the correct exchange-correlation functional. So far, hundreds of approximations to the exchange-correlation functional are available, which can be semi-empirical[179;180] and non-empirical[181–186] in nature, but the holy grail has not been found. Some approaches also combine semi-empirical and non-empirical approaches.[187;188] The biggest challenge from a practical point of view is to choose a suitable functional for a given problem. This task is particularly tricky, because a more accurate solution is not guaranteed by simply choosing a more complex functional or a functional associated with higher computational costs.[172;189] The outcome of a DFT calculation thus critically depends on the type of approximated density functional that is applied.

Indeed, DFT-based calculations with conventional exchange-correlation functionals often fail to accurately describe systems, which are characterized by strong static correlation. Static correlation is important in systems, which contain a lot of nearly degenerate orbitals close to the highest occupied molecular orbital (HOMO) and lowest occupied molecular orbital (LUMO). Such quasi-degenerate orbitals are common in transitions states, broken bonds, and excited states. It is therefore not sufficient to describe the electronic structure of such systems with a single Slater determinant, i.e., a single configuration. The required wave function to properly account for static correlation requires more configurations and differs qualitatively from the single-reference wave function of the Kohn-Sham state configuration.[190;191] More

accurate DFT approaches exist, which are better suited for multi-configurational problems, but have not yet entered the mainstream.[192]

2.1.2 *Wave Function Theory (WFT)*

In contrast to DFT, WFT offers a clear-cut hierarchy of methods that are systematically improvable toward the exact but unattainable solution. The time-independent electronic Schrödinger equation[193] is fundamental to WFT:

$$\hat{H}_{el}(\mathbf{R}, \mathbf{r}) \mid \Psi_i(\mathbf{R}, \mathbf{r})\rangle = E_i \mid \Psi_i(\mathbf{R}, \mathbf{r})\rangle. \tag{2.1}$$

$\hat{H}_{el}$ is the electronic Hamilton operator that is applied to the N-electron wave function $\Psi_i(\mathbf{R}, \mathbf{r})$ of electronic state i, which is dependent on the electronic coordinates, $\mathbf{r}$, and parametrically dependent on the nuclear coordinates, $\mathbf{R}$, within the Born-Oppenheimer picture. This representation, in which the potential energy, E_i, of a given state is the eigenvalue of the electronic Schrödinger equation, is called the adiabatic representation.[194] An exact solution can be obtained in theory, but solving this eigenvalue problem is anything but trivial. Approximations to the electronic wave function have to be introduced to allow for a feasible computation of many-electron systems.[2;26;169]

Within the Hartree-Fock theory the energy of a system is computed using a single Slater determinant,[195] ϕ^{HF}, and the variational principle.[196] The Slater determinant represents the ground state electronic configuration that meets the requirements of the antisymmetry principle. The molecular orbitals of this configuration are obtained as linear combinations of atomic orbitals (N one-electron wave functions)[197] and each atomic orbital is itself treated as an expansion of K known basis functions, θ_u :

$$\vartheta_i(r) = \sum_{u=1}^{K} C_{ui}\theta_u \tag{2.2}$$

with $\vartheta_i(r)$ being the spatial part of the atomic orbital. The variational principle guarantees that the energy obtained with an approximated wave function is always higher than the exact energy of the investigated system. In this way, the coefficients weighing the basis functions of the molecular orbitals can be iteratively optimized to minimize the energy of the system. This can be done until self-consistency is reached, giving this approach the name Self-Consistent Field (SCF). The use of a finite basis set to construct the molecular orbitals and the neglect of electron correlation, i.e., instantaneous electron-electron interactions, limit the accuracy of Hartree-Fock theory.[169] Due to the lack of electron correlation, i.e., static and dynamic correlation, the Hartree-Fock method does not allow for a qualitatively and quantitatively correct picture of many chemical problems.[169;198]

A source of error in Hartree-Fock theory is the lack of electron correlation, E_{corr},

$$E_{corr} = E_{HF} - E_{exact}, \tag{2.3}$$

which is the difference of the Hartree-Fock energy, E_{HF}, and the exact energy, E_{exact}, of a system. Besides static correlation, which has already been described, dynamic correlation plays a role. This type of correlation is important to account for the instantaneous interactions of electrons with each other, which are neglected in Hartree-Fock theory, because the electrons are described in a mean field created by all the other electrons. Additional excited-state configurations can be included to

account for dynamic correlation.

In case of single-reference methods, the optimized wave function for the electronic ground state can be used as a reference wave function. The different excited-state configurations are generated by distributing the electrons among the other available molecular spin orbitals. For many systems, the excited states can be sufficiently accurately described with a single reference wave function, which is then obtained as a linear combination of N-electron wave functions:

$$| \Psi_i \rangle = \sum_I C_I | \phi_I \rangle. \tag{2.4}$$

The respective methods, which are obtained by truncation of the configuration space, are known as configuration interaction (CI) methods. The CI coefficients, C_I, are optimized variationally by minimizing the energy of a system and keeping the molecular orbitals fixed. The energy of a system is computed by diagonalization of the Hamiltonian matrix in this N-electron basis. Arranging the electrons in all possible ways and taking all possible Slater determinants into account results in a Full CI (FCI) description. FCI yields the exact solution of a system.

As already mentioned, the exact solution is practically infeasible except for the simplest molecular systems, such as H_2, He, H_3,...[172] Therefore, the approximations of a finite basis and a truncated configuration space are applied. Examples of truncated CI methods are CIS (CI Singles), which considers only single excitations, or CISD (CIS and Doubles), which additionally considers double excitations.[26;169] As long as conformations of a system, where the Born-Oppenheimer approximation is valid, such as the ground state equilibrium structure, are treated, single-reference methods are well suited.[26;164;169]

However, the same cannot simply be said for the excited states, as different electronic states are usually characterized by different configurations defining their distinct nature. An example is a conical intersection, around which two electronic states approach each other and at which point strong electron-nuclear couplings arise leading to the so-called "break-down" of the Born-Oppenheimer approximation. At least two electronic states become degenerate and strong mixing of the different configurations can be observed. Static correlation has to be incorporated in the calculation, which can be achieved with multi-configurational SCF methods.[1;2;26]

Within multi-configurational SCF methods, the molecular orbitals are optimized, either in a state-specific fashion or by averaging over different states. In order to allow for a practicable solution and to reduce the computational effort only some orbitals are chosen, in which the occupation number is allowed to vary. Among the most popular flavours of multi-configurational SCF methods is the Complete Active Space SCF (CASSCF) formalism.[199;200]

Fig. 1 shows an example of a selection of an active space for a CASSCF calculation of an arbitrary system. As can be seen, the orbitals are grouped into inactive and active orbitals. The inactive orbitals are either always double occupied (c) or always empty (a). The occupancy of the active orbitals (b) can vary between 0 and 2. FCI is applied within the active space and all possible configurations are accounted for. The coefficients of the active orbitals are optimized together with the CI coefficients. The active space in Fig. 1 comprises two active orbitals (n) and two active electrons (m), which is denoted as CASSCF(2,2) (CASSCF(m,n)).

CASSCF schemes suffer from a high computational effort that increases exponentially with the number of configurations to be accounted for. State-averaging can be applied to optimize the orbitals within the active space not for each state, but for a

(a) Inactive space (empty)

(b) Active Space (occupancy between 0 and 2)

(c) Inactive space (doubly occupied)
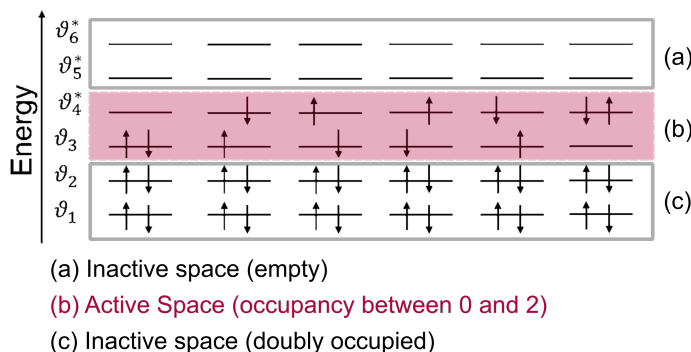
Figure 1: Exemplary scheme of an active space used in CASSCF. (a) Virtually unoccupied orbitals are selected to be inactive. (b) The active space, in which all possible electronic configurations are described contains two active electrons and two active orbitals, resulting in CASSCF(2,2). (c) The inactive part that is lower in energy is always doubly occupied.

set of states that are averaged. Despite allowing for a description of static correlation, excited state configurations have to be added to properly treat dynamic correlation. Schemes like multi-reference CI (MR-CI)[201] that additionally allow excitations out of the active space can account for dynamic and static correlation. The CI coefficients of eq. 2.4 and the molecular orbitals are optimized. Similarly to truncated CI methods, also truncated MR-CI schemes exist. Among the most often applied formalism are MR-CIS and MR-CISD.[169]

Another way to treat dynamic correlation is to apply Multi-Reference Perturbation Theory (MRPT), i.e., applying PT, similar to Møller-Plesset PT to second order (MP2),[202;203] to multi-reference problems. In PT, the Hamiltonian is split into a solvable part and a part acting as a perturbation. As a consequence, the energy and the N-electron wave function are partitioned into a zeroth order and higher-order terms. By applying the perturbation to the known reference zeroth order wave function the correlation energy can be estimated and added as a correction to the SCF energy of the reference system. Dependent on the reference zeroth-order wave function, several variants exist. If the reference zeroth-order wave function is the CASSCF wave function, the resulting method is known as the Complete Active Space PT to Second Order (CASPT2)[204–206] approach. From a practical point of view, higher-order terms are most often not considered as they do not guarantee a more accurate solution, but are computationally more expensive.

Despite the high accuracy that can be achieved with multi-reference methods like CASSCF, CASPT2 or MR-CI schemes, their bottleneck is their complex use and high associated computational effort. Especially the latter makes the treatment of systems that require a large active space computationally impracticable. An improper selection of active orbitals or a generally too small active space can result in inconsistent PESs. Further problematic are intruder states, which are not considered to be relevant at a reference molecular geometry, i.e., they are not described with the reference space, but are close in energy to the ground state at a geometry of interest.[207] In view of preparing a training set for ML models, inconsistencies within the reference PESs pose an additional challenge and are very difficult to identify in systems with many degrees of freedom.[26;77;208]

Another PT scheme for computation of the excited states is the algebraic diagrammatic construction scheme for the polarization operator to second order PT (ADC(2)).[209] ADC(2) has gained popularity for the computation of excited states due

to the good compromise between accuracy and computational efficiency. Compared to DFT, it does not require a proper selection of any functional, but is systematically improvable. Similarly to PT, the second-order correction usually gives the best compromise between accuracy and computational efficiency. This scheme, originally based on Green's function theory, applies the polarization propagator in order to obtain the time evolution of the polarizability of a system.[169;209] The ground state wave function is treated as being externally perturbed and in principle, the relevant excited state information is contained in the polarization propagator. Despite the benefits ADC(2) offers, it is a single-reference method and only dynamic correlation can be accounted for.[209]

2.2 MIXED QUANTUM-CLASSICAL (MQC) PHOTODYNAMICS

MQC photodynamics simulations can provide valuable insights into the processes that take place after light excitation and are thus important for our fundamental understanding of many basic concepts taking place in our closest environment. Their theoretical investigation requires the computation of the electronic PESs of a system, on which the nuclei are considered to move. The underlying quantum chemical computation of the PESs limits the applicability of photodynamics simulations to systems comprising less than some 100s of atoms and on a time scale usually with the upper limit being in the range of a couple of ps. The Born-Oppenheimer approximation allows to decouple the electronic motion from the classical nuclear motion, which renders an on-the-fly evaluation of the PESs at the molecular geometries visited during dynamics simulations possible. On-the-fly methods are particularly powerful because they require the calculation of PESs from a quantum chemical program only for molecular geometries that are relevant for dynamics simulations under the given reaction conditions. In this way, the amount of necessary quantum chemical calculations can be reduced and the PESs do not have to be computed comprehensively in advance, which is not only expensive, but also nontrivial for high-dimensional systems with conventional approaches.[168]

On-the-fly MQC simulations[210] are interesting targets for the application of ML models, because the time limiting step of a simulation is the computation of the electronic PESs. ML models can significantly advance existing simulations by fitting the PESs of systems in advance. Compared to conventional MQC photodynamics simulations with pre-fitted PESs, they do not require the full relevant PESs in advance, but can be adapted efficiently during a dynamics simulations in order to capture all relevant conformations. Compared to on-the-fly MQC simulations with quantum chemical methods, ML models allow for fast inferences at geometries visited during the photodynamics simulations, while retaining the accuracy of the quantum chemical reference method if trained properly.[153;211;212]

Although a quantum dynamical treatment of a system would be more exact, the benefits ML models can offer are limited due to the high costs of evaluating the nuclear motion quantum mechanically at the current stage of research[26] Throughout this thesis, focus will be therefore set on MQC simulations. ML models are used in this framework to replace the quantum chemical calculations of the PESs and corresponding excited-state properties, while keeping the rest of the dynamics as they are. In this thesis, the trajectory surface hopping approach is applied using the SHARC method.[72;166;213;214]

2.2.1 *Trajectory Surface Hopping*

The trajectory surface hopping method is one efficient class of MQC simulation methods to study the photo-initiated reactions of systems on-the-fly.[215] As it is common in MQC approaches, the electrons are treated quantum mechanically and their time evolution is investigated using the time-dependent Schrödinger equation,

$$i\hbar \frac{\partial \Psi(\mathbf{R}, \mathbf{r}, t)}{\partial t} = \hat{H}_{el}(\mathbf{R}, \mathbf{r})\Psi(\mathbf{R}, \mathbf{r}, t), \tag{2.5}$$

with the time-dependent electronic wave function $\Psi(\mathbf{R}, \mathbf{r}, t)$. In contrast, the nuclei are treated classically and follow Newton's equation of motion:[168;216;217]

$$m_A \frac{d^2}{dt^2}\mathbf{R}_A(t) = -\nabla_A E_i(\mathbf{R}_A). \tag{2.6}$$

Due to the number of excited states available, an active state i with potential energy E_i has to be determined, as indicated in equation 2.6. The motion of an atom A with mass m_A is then evaluated according to this active state and the corresponding forces, i.e. derivatives of the potential energy with respect to Cartesian coordinates, $-\nabla_A E_i(\mathbf{R}_A)$, dictate in which direction the atom is pushed. At each time step, the atoms are assumed to move adiabatically on a single electronic state. The Velocity Verlet algorithm[218] can be applied to efficiently integrate this equation and compute the motion of the nuclei. Compared to the quantum chemical calculation of the electrons, the propagation of the nuclei can be considered to be almost for free in terms of computational costs.[72]

In order to go beyond the Born-Oppenheimer approximation and render the description of nonadiabatic effects possible, the active state has to be allowed to change, which comes along with a hop between different states. Within surface hopping a hop refers to a transition of the system from one state to another. The associated hopping probability needs to be determined at each time step.[72;215;217;219]

Several ways exist to approximate this hopping probability. Originally, trajectory surface hopping was developed by Tully[220–222] to treat internal conversion through conical intersections, i.e., transitions between states of same spin multiplicity. The probability for a transition is based on a fewest switches criterion[221] that constrains the number of hops to a minimum so that too many hops do not average the potentials similar to a mean-field treatment. Within this original scheme, the time-dependent coefficients, $c_i(t)$, of the time-dependent electronic wave function are used to estimate the hopping probability and have to be propagated from one time step to another.[222] By expressing the electronic wave function as a linear combination of the eigenfunctions of the electronic Schrödinger equation in the adiabatic basis (as given in equation 2.1),

$$\Psi(\mathbf{R}, r, t) = \sum_i c_i(t)\Psi_i(\mathbf{R}, r)e^{-\frac{i}{\hbar}\int E_i(\mathbf{R})dt}, \tag{2.7}$$

and inserting this ansatz into the time-dependent Schrödinger equation (as given in equation 2.5), the equation to propagate the expansion coefficients can be obtained:[168]

$$\frac{dc_i(t)}{dt} = -\sum_i \left[i\langle \Psi_i \mid \hat{H}_{el} \mid \Psi_j \rangle + \langle \Psi_i \mid \frac{d}{dt} \mid \Psi_j \rangle \right] c_j(t) \tag{2.8}$$

$$= -\sum_i \left[iH_{ij} + K_{ij} \right] c_j(t).$$

The Hamiltonian, H_{ij}, is computed from the electronic structure calculation and $K_{ij} = C_{NAC_{ij}} v_R$ can be obtained from the velocities v_R and the nonadiabatic couplings vectors (NACs), denoted as $C_{NAC_{ij}}$.[198;213;214] By neglecting the second order term, the NACs can be calculated with the following equation:[223;224;224]

$$C_{NAC_{ij}} \approx \langle \Psi_i \mid \frac{\partial}{\partial \mathbf{R}} \Psi_j \rangle \tag{2.9}$$

$$= \frac{1}{E_i - E_j} \langle \Psi_i \mid \frac{\partial H_{el}}{\partial \mathbf{R}} \mid \Psi_j \rangle \quad \text{for } i \neq j.$$

By knowing the coefficients, the probability to hop out of the active state is:[72;222]

$$P_{i \to j} = 1 - \frac{\mid c_i(t + \Delta t) \mid^2}{\mid c_i(t) \mid}. \tag{2.10}$$

In order to obtain the population in the other states, several techniques exist. Within Tully's fewest switches surface hopping scheme, a hop is determined stochastically by comparing the hopping probability from active state i to a new state j, which is computed via:[222]

$$P_{i \to j} = \frac{2\Delta t}{\mid c_i(t) \mid^2} \Re \left\{ c_i^*(t) \sum_j \left[i H_{ij} + K_{ij} \right] c_j(t) \right\}, \tag{2.11}$$

to a random number. If the hopping probability is larger than the random number, a hop is assumed to take place, whereas otherwise the active state is retained. The hopping probabilities computed with SHARC follow the fewest switches criterion, but are based on the propagator matrix, $P = e^{[-(iH_{ij}+K_{ij})\Delta t]}$.[72;213;214]

$$P_{i \to j} = \left(1 - \frac{\mid c_j(t + \Delta t) \mid^2}{\mid c_j(t) \mid^2} \right) \frac{\Re \left[c_i(t + \Delta t) P_{ij}^* c_j(t) \right]}{\mid c_j(t) \mid^2 - \Re \left[c_j(t + \Delta t) P_{jj}^* c_j(t) \right]}. \tag{2.12}$$

Besides the hopping probabilities based on the aforementioned algorithms, other measure to approximate the transitions between different states exist.[194;219;225–229] This makes MQC simulations with the surface hopping procedure not unique.[230] Examples are the Landau-Zener[231;232] or the Zhu-Nakamura[227;229;233] theories. Both formalisms do not require information on the couplings between adiabatic PESs, which leads to less expensive electronic structure calculations. Originally, the Landau-Zener and Zhu-Nakamura formalisms have been developed for two-level systems and they have been derived in the diabatic basis.[234] Subsequently, the validity has been proven also for the adiabatic representation and the respective hops in the adiabatic basis are determined from the energy gaps between different potentials.[233] As a consequence, two different chemical problems with similar adiabatic PESs, but significantly different couplings would yield qualitatively similar results. In addition, systems that are characterized by a high density of states are also insufficiently treated with these algorithms.[157]

Due to the aforementioned failures for some chemical problems, Tully's fewest switches algorithm remains the more accurate scheme, especially for larger systems and systems, whose photochemistry is more complex. Tully's fewest switches hopping algorithm is thus often the method of choice.[1;72]

As a consequence of the classical treatment of the nuclear motion and the stochastic nature of surface hopping, a single trajectory cannot provide statistically significant results. In order to get an estimate of the proper motion of a nuclear wave packet, an ensemble of trajectories has to be considered.[72;215;217;219;222] To this aim, different initial conditions have to be provided from which the trajectories can be started after vertical excitation. This sampling can be carried out from frequencies and normal modes of molecules via a Wigner distribution[235;236] for example, which is the standard method within the applied SHARC approach in this thesis.[72;166] A frequency calculation and single point calculations for each sampled molecular configuration are prerequisites. Considering a large number of initial conditions and consequently a large number of trajectories, the quantum behaviour of the nuclear motion can be approximated at least to some extent. The possibility to propagate the different trajectories independently from each other allows for parallel computations. Different branching channels and reaction kinetics can be computed and the critical regions that lead to ultrafast transitions can be investigated usually within a good compromise between computational efficiency and accuracy.[1;26;215] However, the number of trajectories are often limited to a few hundred of trajectories as the costs of the quantum chemical calculations still remains. For chemical problems, where rare reactions are important that occur with less than 1%, no statistically significant results can be obtained at justifiable costs. ML models can also contribute here and provide a solution to this problem as they allow for efficient sampling of many more trajectories.[1]

2.2.2 *Nonadiabatic Couplings (NACs)*

The outcome of a photodynamics simulation with the surface hopping approach strongly depends on the accuracy of the regions of the PESs, around which hops take place. Those regions are usually called critical regions of the PESs and can be for example two-state conical intersections or interstate crossings that are schematically represented in Fig. 2 in the spin-diabatic basis, which is the direct outcome of a quantum chemical calculation. This basis is often referred to as the adiabatic basis, when only one spin multiplicity is considered because the potentials then never cross each other. However, when states of different spin-multiplicities, e.g. singlet and triplet states, are considered together, they do cross each other and this picture is referred to as spin-diabatic.[165] As can be seen, two states of the same spin multiplicity (singlet states E_i and E_j in Fig. 2) form an avoided crossing and the corresponding NACs (exemplified by their norm using dashed blue lines) show singularities. Thus, this conformational region is also known as a critical region of the PESs.[165;168;237]

An explanation can be provided with equation 2.9, which shows that the NACs are inverse proportional to the energy gap between two electronic states. Obviously, they become very large, when two PESs are close to each other. Especially this characteristic of NACs poses a real challenge to a quantum chemical calculation, as it leads to the break-down of the Born-Oppenheimer approximation. Consequently, it is difficult to properly converge a quantum chemical calculation in such regions. Further, the fitting of such couplings and discontinuous PESs turns out to be challenging. An inaccurate fit in these critical regions of the PESs can lead to qualitatively wrong transition probabilities.[165;168;237]

A way to remove such inconsistencies is to use the diabatic basis instead,[163;238–240] which is illustrated in Figure 3(a).
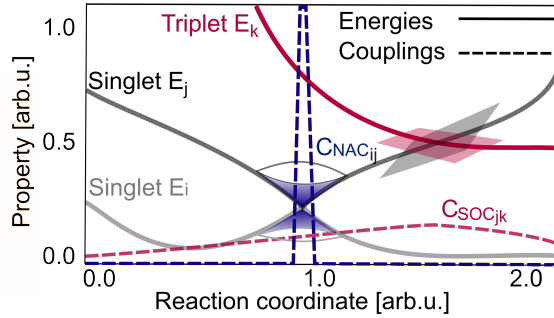
Figure 2: One-dimensional example of an arbitrary reaction coordinate in the spin-diabatic basis with an avoided crossing between singlet states E_i and E_j and a crossing of E_j with triplet state E_k. The planes that cross and the cone illustrates how these critical points look like in two dimensions. The corresponding couplings are plotted using dashed lines. The NACs ($C_{NAC_{ij}}$, blue dashed lines) show singularities at the critical points, whereas SOCs ($C_{SOC_{jk}}$, red dashed lines) yield a smooth function along the reaction coordinate.

Within the diabatic basis, NACs are obtained as smoothly varying potential couplings of nuclear coordinates, removing the singularities at conical intersections. Similarly, the inconsistencies within the PESs of same spin-multiplicities vanish and the states, now ordered by their character, cross. Despite the advantages that diabatic PESs and couplings offer, the weak point of this consideration is that a transformation from the adiabatic to the diabatic basis is not possible for polyatomic systems. Therefore, it is only possible to fit non-unique approximated quasi-diabatic PESs. Finding quasi-diabatic PESs is far from trivial. Moreover, the conformational space of a molecule to be visited during a dynamics simulation usually has to be known in advance, posing another obstacle to properly fit diabatic PESs.[163;238–241] At this point and for the sake of the goals of this thesis, i.e., to provide a generally applicable ML model to simulate photodynamics of polyatomic molecules, the adiabatic basis is preferred over the quasi-diabatic basis. Nevertheless, it should be mentioned that if quasi-diabatic PESs of a system are known, their use is highly favoured. A transformation from quasi-diabatic PESs to adiabatic PESs is almost always possible by diagonalization of the quasi-diabatic Hamiltonian matrix.[241]

2.2.3 Spin-Orbit Couplings (SOCs)

Relativistic effects, i.e., spin-orbit (SO) effects, allow states of different spin-multiplicities to couple with each other via spin-orbit couplings (SOCs). Without an interaction of the orbital angular momentum with the magnetic moment of the electronic spin, the states would be uncoupled.[242–244] At the current stage of research, approximations to the relativistic many-body equation exist, with the Breit equation[245] being among the most popular ones:[245;246]

$$i\hbar \frac{\partial \Psi(\mathbf{R}, r, t)}{\partial t} = \hat{H}_{DCB}\Psi(\mathbf{R}, r, t). \tag{2.13}$$

Instead of the electronic Hamiltonian, the Dirac-Coulomb-Breit Hamiltonian, $\hat{H}_{DCB}$, is applied. This Hamiltonian can be decomposed into a non-relativistic part, a mass-velocity part, and a SOC part, denoted here as $\hat{H}_{DCB-SO}$:[243;245]

$$\hat{H}_{DCB-SO} = \frac{1}{2c^2}\left[\sum_j \sum_A \frac{Z_A}{r_{jA}^3}\left(\hat{r_{jA}} \times \hat{p}_j\right) \cdot \hat{s}_j - \sum_{k \neq j} \frac{1}{r_{jk}^3}\left(\hat{r}_{jk} \times \hat{p}_j\right) \cdot \left(\hat{s}_j + 2\hat{s}_j\right)\right] \tag{2.14}$$
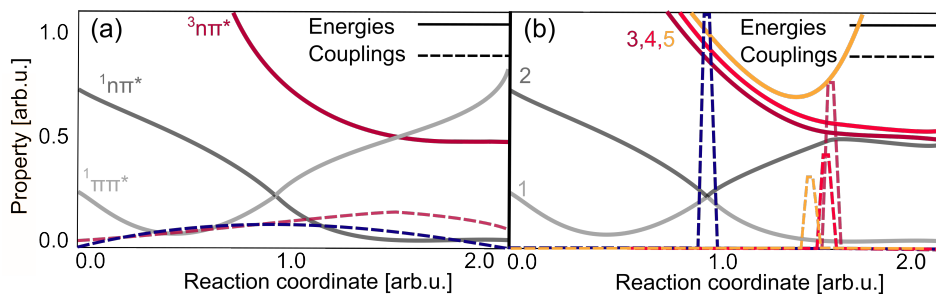
Figure 3: The potential energy curves of Fig. 2 in (a) the diabatic and (b) the diagonal, strictly adiabatic, representation. (a) The singlet and triplet states are ordered by character in the diabatic basis and the couplings (dashed lines) are delocalized. (b) The spin-mixed states never cross and are ordered according to their energy. Couplings are localized at the critical regions of the potential energy curves.

In equation 2.14, c refers to the speed of light and the operators $\hat{r}_j$, $\hat{p}_j$, and $\hat{s}_j$ refer to the position, momentum, and spin operators of electron j, respectively. The cross-products $\hat{r}_{jA} \times \hat{p}_j$ and $\hat{r}_{jk} \times \hat{p}_j$ give the angular momentum of electron j relative to the atom A and the electron k, respectively. The first term of equation 2.14 is related to the one-electron SO term, whereas the second part refers to spin-same-orbit and spin-other-orbit terms.[243–245] As given in the equation, the atomic charge, Z_A, enters and significantly contributes to the SOCs when the mass of the atom is large. This relation led to the widespread assumption that SO effects would almost exclusively be relevant in systems containing heavy atoms.[247;248] Today, it is well known, that SOCs also play a role in other systems and give rise to a multifaceted photochemistry of a large number of chemical systems.[2;213;214;249]

In view of photodynamics simulations, the SOCs are used to determine the rate of intersystem crossing. The SHARC program was the first on-the-fly trajectory surface hopping method based on the fewest switches algorithm to include also such relativistic SO effects.[213;214] To incorporate these effects, the electronic Hamilton operator needs to be adapted with the spin-orbit term, $\hat{H}_{SO}$, to form the total Hamilton operator, $\hat{H}_{tot} = \hat{H}_{el} + \hat{H}_{SO}$. The two-electron integrals present in $\hat{H}_{DCB-SOC}$ makes the computation of SOCs computationally expensive and the use of an effective SO mean-field operator[249;250] is computationally more efficient while providing reasonably accurate results.[213;214;243]

Due to the off-diagonal potential couplings in H_{tot}, the Hamiltonian is not diagonal anymore, which is in contrast to H_{el}, hence the name spin-diabatic representation. The arising potential couplings are the SOCs. As it is illustrated in Fig. 2, in contrast to the singlet-singlet avoided state crossings, the singlet and triplet states cross (as shown by singlet state E_j and triplet state E_k) when computed with standard quantum chemistry. SOCs in this spin-diabatic basis are exemplified by red dashed lines and are smoothly varying complex or real-valued properties with respect to nuclear coordinates. Whether SOCs are obtained as real or complex valued properties depends on the electronic structure program applied. For one described singlet and triplet state, three SOC elements arise that are due to the different magnetic numbers of degenerate triplet components. Real-valued SOCs can be converted into complex-valued SOCs and vice versa.[72;244]

For the determination of the hopping probability between states of different spin multiplicities with SHARC, the spin-diabatic basis is transformed to the diagonal basis by diagonalization of the Hamiltonian matrix. As a result, the previously denoted

singlet and triplet states are then spin-mixed states and all states are eigenvalues of the total Hamilton operator. An example of a strict-adiabatic basis is shown in Fig. 3(b). As it is visible, the degeneracy of the previously degenerate triplet states is lifted.[1;72;230] The obtained kinetic SOCs are localized in this diagonal basis. In view of surface hopping simulations, the use of localized kinetic couplings, as illustrated by reddish dashed lines, is advantageous, because they guarantee a more accurate determination of hopping probabilities between the now spin-mixed states. The use of the directly obtained SOCs from quantum chemical programs could induce hops with equally large probabilities almost anywhere along the conformational space of a molecule and would result in an inaccurate mean-field description of intersystem crossing.[214]

2.2.4 *Arbitrary Phase of the Wave Function*

An important effect to consider for surface hopping simulations is the phase of the wave function. Two different effects can be distinguished that can lead to a change in the phase of the electronic wave function. The Berry phase, also known as the geometric phase, exists.[251−253] The Berry phase gives rise to path-dependent transitions. When a closed path around a conical intersection is undertaken and the original molecular geometry is reached, the electronic wave function differs by a phase factor of π. This effect can have a significant influence on the outcome of a dynamics simulation when the nuclear motion is treated quantum mechanically. In contrast, MQC simulations are assumed to be almost unaffected by the Berry phase in many cases,[254;255] which is demonstrated for example by the Zhu-Nakamura approach, where the phase is neglected.[233;256]

In addition to the Berry phase, the arbitrary phase exists. The electronic wave function is an eigenfunction of the electronic Schrödinger equation, but is only defined up to an arbitrary phase, which can be seen similar to an arbitrary sign as a pre-factor. The electronic wave function can change its phase and is still a valid eigenfunction of the electronic Schrödinger equation.[72;257] As a consequence of this arbitrariness, also excited-state properties that result from two different electronic states carry an arbitrary sign. Independent electronic structure calculations of the same molecular geometry can therefore result in excited-state properties that arbitrarily differ in their sign. This is the case for NACs or SOCs, which are particularly relevant for computing the hopping probabilities, and transition dipole moments.[77;257] This effect is exemplified in Fig. 4 for a SOC value.

As shown in Figure 4, the SOC value, which is represented by dashed lines, can have either positive ($+C_{SOC_{jk}}$) or negative ($-C_{SOC_{jk}}$) values. The red dashed lines illustrate the two optimal possibilities for representing the couplings along this reaction coordinate. However, reality looks differently and most probably more similar to the black dashed line in-between the red dashed lines. Identifying whether the phase changed for a given electronic state at a certain point in time or not is therefore critical for the outcome of a photodynamics simulation. A wrong phase factor can lead to wrong hopping probabilities and consequently wrong surface hopping dynamics simulations.[72]

In many conventional surface hopping dynamics simulations the problem is solved by carrying the phase along a trajectory. A change in a phase factor for a given electronic state from one time step to another can be identified for example by
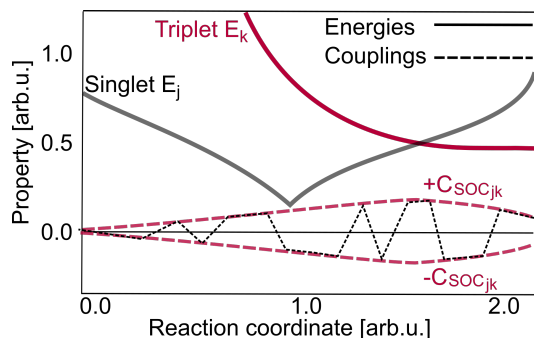
Figure 4: The potential energy curves for the singlet state E_j and the triplet state E_k (as already introduced in Fig. 2) are shown along with their couplings, the SOCs, C_{jk}^{SOC}. The red dashed lines show how couplings will look like if the phase of the electronic wave function is corrected for each electronic state. The direct output obtained with quantum chemistry is exemplified as the black dashed line and illustrates the arbitrariness of the sign of coupling values.

computing wave function overlaps, S, between the electronic wave function of the actual time step, α, and the electronic wave function of the previous time step, β:[258]

$$S = \langle \Psi_\beta \mid \Psi_\alpha \rangle \tag{2.15}$$

Small enough time steps in photodynamics simulation guarantee that the two electronic wave functions at the different molecular geometries are similar enough to obtain overlaps for given electronic states that are either close to +1 or -1, indicating whether a change in the corresponding phase has occurred or not. Large overlap values are obtained as diagonal elements of the overlap matrix in most conformational regions of the PESs. Off-diagonal elements become important, when a conical intersection is in proximity.[72;214;259] Also when the fitting of properties is required or when a diabatization procedure based on NACs[260] or Boys localization[261] is carried out, the phase has to be considered, because inconsistencies in the respective excited-state properties can make the training of an ML model[77] or the diabatization procedure[162;208] fail.[157]

2.3 MACHINE LEARNING (ML) IN QUANTUM CHEMISTRY

ML models are becoming increasingly important in many research fields. Especially deep learning, i.e., the part of ML that is inspired by the human brain, is a powerful technique to fit highly complex structure-property relationships. It allows to create intelligent machines that learn from data without being explicitly programmed.[262]

In 1943, a model based on flexible functions without any direct physical meaning was introduced to investigate the signal processing in the brain.[263] Today, this study is regarded as the birth of NNs[153] and some pioneering works applying different types of ML algorithms followed in the 1950s. After these initial studies a so-called first "Artificial Intelligence (AI) Winter" followed. This period of time was characterized by slow progress and little effort was made to further develop this research field.[262]

Nevertheless, every winter is followed by a summer, and ML models were rediscovered in the 1980s, which is known as the first "AI Summer". This summer was mainly driven by the prospects of generating intelligent machines that can learn from the increasingly available amount of data. However, the bottleneck during this time was

the limited computational power and inefficient algorithms that prohibited the use of deep learning models, leading to a second "AI winter" until the mid 1990s.[262] Since then, with the increasing computer power available, ML started to constantly gain popularity among many diverse research fields. Today, it is part of our everyday lives and used for image recognition,[264] for translation of text and speech[265;266] or for holding conversations with Siri or Alexa.[265;266]

Even though ML models were unnoticed in the research field of quantum chemistry for a long time, their prospects have already been identified more than 25 years ago:[267]

> "The potential in neural networks for the processing of chemical information is very far from being exhausted. In chemistry the task is often to assign objects to certain categories or to predict the characteristics of objects. This accounts for the dominance of the back-propagation algorithm. A whole series of other neural network models exists, however, which could be applied successfully to the field of chemistry. This should be explored more widely in future." (Gasteiger, J. and Zupan, J., *Angew. Chem., Int. Ed. Engl.*, **32**, 503 (1993))

The first ML models in quantum chemistry were almost exclusively applied in the field of quantum dynamics.[85;86] Nowadays ML models are massively booming and have entered many other fields of quantum chemistry as well.[262] For example, they can help to solve the Schrödinger equation by providing approximations to the electronic wave function or density of a system,[90−97;174;174;268−279] they can learn from electronic structure data to fit high-dimensional PESs or properties of molecules and materials,[77;81;86;98−110;112−115;117−130;145;154−156;159;160;162;163;208;212;280−291] and can be used to find relations between molecules and certain properties, e.g. reaction kinetics of photodynamics simulations, energy gaps between orbitals or electronic states or catalytic activities.[28;90;129;139;141;150;292−296] Many newly developed methods in the research field of quantum chemistry incorporate ML models. Old concepts have often been revisited and have been significantly advanced by an integration of ML algorithms. However, ML in quantum chemistry is still far from being routine and method development is pushed toward solutions that can be used without expert knowledge on ML. In this thesis, the idea of advancing existing methods, i.e., MQC dynamics with the surface hopping approach, with ML is maintained.

2.3.1 *Training Set Generation*

Supervised learning techniques can be divided into regression and classification strategies. By applying a regression or classification model, at least in theory, almost any input-output relation can be fitted or patterns within a data set can be explored, respectively. The basis of any ML model for this purpose provides a labeled data set. One data point inside of the training set needs to contain an input that is related to one or more outputs. When using quantum chemistry data, the input is often a molecular structure that is linked to quantum chemical properties computed with an electronic structure method. Nowadays several public repositories exist that provide such data sets. In the research field of quantum chemistry for molecules, examples are the QM7,[280;297] QM8,[146;298] QM9,[173] QMspin,[299] MD17,[300;301] NMD-18[302] or H2O-13[303] data sets that mainly include ground-state properties of molecules or first excitation energies. These data sets are very important for the development of new ML methods and can serve as benchmark sets.[304;305]

However, available data sets that include excitation energies have been generated mainly with the purpose of sampling the chemical space comprehensively. Most often, molecules of different compositions were generated or extracted from existing data bases and were optimized with an electronic structure method. In order to obtain a training set that can be used for photodynamics simulations, a different strategy has to be applied as more data points of the same molecule, but with different structure have to be provided. To be more specific, the conformational space that is visited during the dynamics simulations of at least one system under investigation has to be sampled comprehensively. Obviously, a dilemma arises, because the important conformational regions of a molecule are not known in advance and the transferability of excited-state PESs and couplings is not known. To this aim, an iterative sampling procedure has been proposed by Behler[153] and has been adapted by others to generate training sets for Born-Oppenheimer MLMD simulations for one molecular system. This scheme uses MD simulations with preliminary ML potentials to automatically identify regions of the PESs that are not yet included in the training set. In this way, the training set can be built efficiently and only relevant conformations are included.[100;106;306;307]

The starting point of this automatic selection scheme is a small initial training set that can be based on the equilibrium geometry of the molecular system under investigation. Additional geometries can be sampled with MD simulations or enhanced sampling techniques[308–311] using a cheap electronic structure method. Clustering of data allows to select distinct molecular geometries that serve as an input for electronic structure simulations with the reference method that determines the accuracy of the ML models.[312] Other sampling techniques, such as random sampling or Wigner sampling,[236] can also serve for this purpose.[153]

In order to enlarge the training set and render efficient, yet accurate MLMD simulations possible, N_{ML} preliminary ML models are trained on this training set ($N_{ML} \geq 2$) and MLMD simulations are initiated. At each time step, the mean of the predictions of energies, $\overline{E}_0^{ML}$, (equation 2.16) and forces, $\overline{F}_0^{ML}$, (equation 2.17) of the ML models for the electronic ground-state, S_0, are used to evaluate the classical equations of motion to propagate the nuclei of the molecule. The index J runs over all N_{ML} ML models used in this scheme and is indicated by ML_J. The uncertainty of the PES provided by the ensemble of ML models is computed from the standard deviation of energy predictions, $E_{\sigma_0}^{ML}$, as specified in equation 2.18[106]

$$\overline{E}_0^{ML} = \frac{1}{N_{ML}} \sum_{J=1}^{N_{ML}} E_0^{ML_J} \tag{2.16}$$

$$\overline{\mathbf{F}}_0^{ML} = \frac{1}{N_{ML}} \sum_{J=1}^{N_{ML}} \mathbf{F}_0^{ML_J} \tag{2.17}$$

$$E_{\sigma_0}^{ML} = \sqrt{\frac{1}{N_{ML} - 1} \sum_{J}^{N_{ML}} \left(E_0^{ML_J} - \overline{E}_0^{ML} \right)^2} \tag{2.18}$$

Whenever the deviation goes beyond a pre-defined threshold, the geometry at this time step is used as an input to re-compute the PES with the reference electronic structure method. This information is added to the training set. ML models are re-trained with the extended training set and MLMD simulations are re-started.

This process is computationally expensive, because firstly, data points are added sequentially and secondly, each time a data point is added to the training set, ML models have to be re-trained. The use of several parallel simulations initiated from different starting conditions can reduce the computational effort. Therefore, ML models are only re-trained after all MLMD simulations have arrived at an extrapolative region, relevant data points have been re-computed with quantum chemistry, and the training set has been expanded.[106;313] As soon as the ML models have reached higher accuracy, the threshold is adjusted to smaller values, which is termed adaptive sampling.[106]

The use of an adaptive sampling scheme is advantageous compared to random sampling as the relevant conformational regions of the PESs are scanned comprehensively, whereas a finite number of randomly sampled data points cannot guarantee that the training set is sufficient for dynamics simulations. Less data points are needed when sampled wisely leading to less computational costs with respect to reference electronic structure calculations and final ML training times.

2.3.2 *ML Models*

With a training set at hand, the easiest way to fit a function between a molecular input, X, and a target output, Y, is to assume a linear relation and use a linear function,

$$Y = b + w \cdot X, \tag{2.19}$$

where w are the coefficients, often known as weights, and b is a constant bias. In case a linear model is selected, ordinary least squares regression can be applied to find the coefficients that relate X to Y in the most optimal way. During this process, called training, a loss function, L_x, is evaluated at each step:

$$L_2 = \frac{1}{N_M} \sum_{\beta}^{N_M} \left(Y_{\beta}^{ML} - Y_{\beta}^{QC} \right)^2 . \tag{2.20}$$

This function monitors the error between a predicted property, Y_{β}^{ML}, for a given molecular input, β, and the original quantum chemical value of the same input, Y_{β}^{QC}. N_M refers to the number of compounds used for training. In equation 2.20 the mean squared error (MSE) is shown as an example and the function is called L_2 norm. Besides the widely used L_2 norm, the L_1 norm is frequently applied to monitor the mean absolute error (MAE) during training. In principal, any type of function that can provide a measure for the accuracy of an ML prediction is applicable. The best practice is to report the error of a successfully trained ML model on a test set that has not been shown to the ML model during training.[82;304;314;315] Although many chemical problems cannot simply be fit with a linear model, it is often used as a baseline model. The accuracy that can be obtained with linear regression can therefore be considered as a lower limit.

Kernel Methods

Kernel methods[316] are one of the two main classes that are frequently used in quantum chemistry to solve non-linear problems. Among them, Gaussian processes, support vector regression (SVR), and kernel ridge regression (KRR) are often applied.

Kernel methods rely on the kernel trick to fit data that cannot be represented using a linear function. The basis provides a kernel basis function, K_1, that defines the non-linearity of the model. Most models make use of a Gaussian kernel or Laplacian kernel. This kernel is centered on each compound within the training set and the distance to all other compounds is computed. In other words, kernel methods measure the similarity between data points and save the relevant information in the kernel matrix. Thus, the size of the kernel, i.e., the depth of the ML model, is inherently linked to the amount of data points. Usually the kernel contains one main hyperparameter, which is its width, σ.[280;316] Hyperparameters of ML models are internal parameters that need to be optimized for a given training set under investigation in order to allow for meaningful performance. As kernel methods usually require less hyperparameters to be tuned, they are relatively simple to train compared to NNs, which depend on a large number of hyperparameters.

When fitting a KRR model, a property of an unknown compound X_α can be approximated by summing up the weighted contributions of the kernel functions placed on each compound, X_β, within the training set:

$$Y^{ML}(X_\alpha) = \sum_\beta^{N_M} w_\beta K_1(X_\alpha, X_\beta) \tag{2.21}$$

For many ground-state problems, a fitted property can also be obtained as the sum of atomic contributions. The kernel is decomposed into a sum of weighted kernels representing atoms in their environment:

$$Y^{ML}(X_\alpha) = \sum_\beta^{N_M} \sum_A^{N_A} w_\beta K_1(X_\alpha, X_{\beta A}). \tag{2.22}$$

Note that in the expression in eq. 2.22, the kernel matrix is not symmetric anymore. The need to estimate the similarity of a new input to all other samples within the training set makes inferences relatively slow compared to predictions executed with NNs. Still, PESs can be evaluated much faster than with quantum chemical reference methods.

The fitting parameters of the model that best relate an input to an output can be found with ridge regression that combines linear regression and a regularization.

$$w = (\mathbf{K_1} + \lambda \mathbf{1})^{-1} \mathbf{X}^{QC} \tag{2.23}$$

Noticeably, w is a vector of size N_M and $\mathbf{X}^{QC}$ is a vector that contains all data points included in the training set. The penalty term (or L_2 penalty), λ, has the task to prevent the model from overfitting and remains a second hyperparameter of KRR models that is strongly dependent on the provided training set. The term overfitting will be described in more details in chapter 2.3.4

A benefit of KRR models is their simplicity and fast training process, due to the little amount of hyperparameters, which makes them generally easy to use. Another main advantages is that they enable almost exact reproduction of data points that are within a region represented by the training set.[317] KRR is therefore a frequently applied method in the research field of quantum chemistry.[123] However, compared to more complex and higher flexible models, such as NNs, conventional KRR models can only relate a single output value to a molecular input. Hence vectorial properties need to be fitted with several KRR models. The same accounts for the task of fitting

several electronic states – a separate KRR model has to be used for each electronic state to be fitted with conventional approaches. A recently proposed response formalism exists,[132] which trains and predicts forces and dipole moments together with potential energies of a system. Forces and dipoles are treated by using the corresponding force or dipole operator that treats forces or dipoles as derivatives of potential energies with respect to atomic coordinates or an external electric field, respectively. This formalism enables a prediction of a complete vector in one KRR model. Such models are, however, only applicable to one electronic state and usually suffer from high memory consumption, restricting the amount of data points that can be used for training considerably.[305]

In general, the main drawback of conventional KRR models is that the training set size that can be used is limited due to the fact that it directly influences the kernel matrix size. A possible way to overcome this limitation without suffering from an accuracy-loss is to use only a subset of training data that is mapped to all data points within the training set.[314;318] Another type of kernel methods is based on support vector machine, often applied to classification problems, but similarly applicable for regression problems. SVR models differ from KRR models in their loss function. While KRR applies a minimum loss function, SVR makes use of an ϵ-intensive loss function, L_ϵ:[319]

$$L_\epsilon = \begin{cases} 0 \text{ if } \mid Y_\beta^{ML} - Y_\beta^{QC} \mid \leq \epsilon \\ \mid Y_\beta^{ML} - Y_\beta^{QC} \mid -\epsilon \text{ otherwise} \end{cases} \tag{2.24}$$

As it is given in equation 2.24, a predicted value, whose absolute error is within the range of ϵ, does not influence the training process. This region is also known as ϵ-insensitive region or ϵ-tube and allows for a sparse solution. If the error exceeds this range the data point is taken into account. Support vectors are instances outside the ϵ-tube. A small number of support vectors and combinations thereof can circumvent the huge memory requirement common with large training sets.[320] Consequently, the aim of the training process is to find the tube that includes most data points of the training set.[313;321;322]

Neural Networks (NNs)

In contrast to kernel methods that have a limited number of internal hyperparameters to be fit, NNs contain a flexible number of hyperparameters.[317] In other words, their set of hyperparameters, i.e., their depth, is independent of the training set size. This makes it possible to fit a huge amount of data. The relevant information is stored in the fitting parameters, i.e., the weights of the network. Additionally advantageous are fast inferences of new compounds as no similarity to all data points has to be computed.[82;315]

As already mentioned, NNs were developed in order to mimic the human brain's information processing. Consequently, the structure of an NN is inspired by our human brain. An example is shown in Fig. 5 (reproduced from Ref. 322 with permission from the Royal Society of Chemistry). Illustrated is a multi-layer feed-forward (MLFF) NN, one of the simplest classes of NNs. It contains an input layer, shown here with two input coordinates, X_1 and X_2, to represent a molecular system, X. These input nodes are connected with the weights of the NN to nodes, y^1, of the first hidden layer that are themselves connected to nodes, y^2, of the second hidden layer.
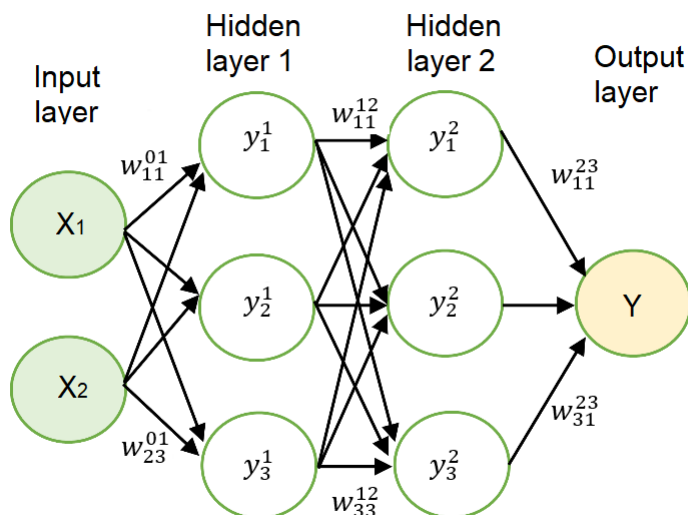
Figure 5: Example of a multi-layer feed-forward NN as used in this thesis. Shown are an input layer comprising two input coordinates, X_1 and X_2, two hidden layers with each hidden layer containing 3 nodes, y, and an output layer that refers to the targeted quantum-chemical property. Weights are denoted as w and a bias connected to each node is hidden from the scheme. (Reproduced from Ref. 322 with permission from the Royal Society of Chemistry.)

Per hidden layer three nodes are exemplified. The number of nodes together with the number of hidden layers can vary. Both are important parameters of the NN and the number of nodes defines its width. The output layer contains the property to be fitted, Y.[153;315] Here, it is shown to be a single value, but due to the high flexibility of the NN, Y could also be a vector composed of several elements. Fitting many states in one ML model is possible – which is in contrast to conventional kernel models.

The training process attempts to find the weights minimizing a loss function with an example given in equation 2.20 The weights, w_{mn}^{rs}, connect node m of hidden layer r to node n of hidden layer s. The function mapping the last hidden layer to the output layer should be linear. Basis functions (also known as activation functions) are introduced to provide NNs with the ability to fit nonlinear relations. They process the weights and give the network flexibility. Common types of such functions are the hyperbolic tangent, sigmoid, Gaussian or the shifted softplus function, $ln(0.5e^x + 0.5)$. The basis functions further need to be differentiable to allow for gradient-based optimization algorithms.[153]

The targeted property can then be computed for the specific example shown in Fig. 5 using the following equation:

$$Y^{ML} = f_1^3 \left\{ b_1^3 + \sum_{o=1}^{3} w_{1o}^{23} f_o^2 \left[b_o^2 + \sum_{n=1}^{3} w_{no}^{12} f_n^1 \left(b_n^1 + \sum_{m=1}^{2} w_{mn}^{01} X_m \right) \right] \right\} \tag{2.25}$$

The nested functional form of the NN in combination with non-linear basis functions and finite numbers of hidden layers and nodes per hidden layer enables the fitting of high-dimensional and complex relations making NNs well suited for many quantum chemical problems. The bias values, b_i^j, not shown in Fig. 5, provide additional flexibility and can shift basis functions. They are connected to each node in every hidden layer. In case no hidden layer is used at all and the activation function is linear, the MLFF NN equals a linear model.[153;315] Whenever several hidden layers

and/or several nodes per hidden layer are used, a deep learning model is the result.

High-dimensional NNs comprise a set of MLFF NNs, where each single MLFF NN represents a different atom type. This type of NN architecture was originally introduced by Behler and Parrinello to fit PESs of the ground state.[323;324] They are based on two approaches: 1) The potential energy, E, is modeled as the sum of atomic contributions, E_A, with N_A being the number of atoms in a molecule:[153]

$$E = \sum_A^{N_A} E_A.$$

(2.26)

2) Each atomic contribution to the energy is modelled in a separate NN. The atom then needs to be represented in it's chemical and structural environment using atom-wise representations.[153] The different ways to represent a molecule to a ML model will be discussed in the next chapter.

2.3.3 *Molecular Descriptors*

The quality of an ML model does not only depend on the type of regressor, but is also sensitive to the input representation. A designed descriptor should fulfill several requirements. In the best way, it should be unique, permutationally, rotationally and translationally invariant, differentiable,[281] and generalizable with respect to other elements.[325] The latter property of a descriptor should enable an ML model to suggest how an atom interacts with another type of atom by using the provided information on interactions from the training set. This ability requires information on the atom type to be incorporated in the descriptor in order to allow for ML generalization across the periodic table. Common input representations used in quantum chemistry calculations, such as xyz-coordinates or internal coordinates, are not well suited for ML models.[153]

In general, descriptors can be classified into two main categories: molecule-based descriptors and atom-wise descriptors.[326] The former type treats molecules as a whole with the matrix of inverse distances or the Coulomb Matrix[280] as examples. The matrix of inverse distances, $\mathbf{D}$, is one of the simplest descriptors that can be used and comprises elements D_{AB} between atoms A and B:[326]

$$D_{AB} = \frac{1}{\| r_A - r_B \|}$$

(2.27)

Elements between the same atom, D_{AA}, are ill-defined and not differentiable and the whole matrix is symmetric. Therefore, only one triangle excluding the main diagonal of the matrix is used for the input vector that is passed to the ML model.[326] Very similar to the matrix of inverse distances is the Coulomb matrix,[280] $\mathbf{C}$, that additionally includes information on atomic charges and encloses diagonal elements:

$$C_{AB} = \begin{cases} 0.5 Z_A^{2.4} & \text{if } A = B \\ \frac{Z_A Z_B}{\|r_A - r_B\|} \end{cases}$$

(2.28)

In contrast to xyz-coordinates, rotational and translational invariance is given. However, these representations are not invariant to permutation as the rows of the matrix change.[325]

Another drawback of such descriptors is their fixed input size. Molecules of arbitrary

sizes cannot simply be processed using standard ML models and descriptors. Defining the size of the input with respect to the largest descriptor that is included in the training set would result in input vectors that contain zero values for all smaller systems. This is only possible, when no preprocessing, such as scaling or centering of the input data, is carried out as this would lead to ill-defined input vectors.[153] Processing zero values otherwise would require unnecessarily large parameters to be fitted for smaller systems and is generally not recommended for NNs.[153]

One solution to enable a description of molecules of arbitrary size and composition is the use of atom-wise representations in combination with high-dimensional NNs or atomistic kernel methods. Atom-wise descriptors allow for a decomposition of the energy of a system into a sum of atomic contributions, which makes them applicable to high-dimensional NNs or atomistic kernel methods. In addition, permutation invariance can be achieved. Examples are the atom-centered symmetry functions,[324] originally developed by Behler to fit PESs with high-dimensional NNs, their weighted variants[327;328] or the FCHL representation,[123] the latter being similar to the descriptors of Behler, but were implemented for KRR and adapted for NNs as well.[305] Atom-centered symmetry functions[324] were the first developed descriptors of this kind. These descriptors are combinations of functions, which describe atoms in local environments using radial and angular atom-centered symmetry functions, which account for radial and angular information to neighbouring atoms, respectively. A cut-off region around a central atom defines the local environment, i.e., the maximum distance at which an atom is allowed to interact with another atom. All atoms that are not inside this sphere are not considered to be a neighbour of an atom when modelling its molecular input.

The FCHL representation is very similar to the symmetry functions of Behler and was developed recently by Faber et al.[123] A set of M-body expansions, usually truncated after the third-body expansion, is used to describe the structural and chemical environment of an atom. An additional first-body term, $A_1(A)$ incorporates information on the stoichiometry. Similar to the symmetry functions, the second-body term describes the distance between atoms and the third-body term additionally includes angular information. When standing alone, the second-body term is equal to a radial distribution function and the third-body term to an angular distribution function.[123]

Representing an atom in its chemical and structural environment can result in highly accurate ML-fitted PESs, at least for ground-state problems, but especially for small systems, the dimension of the input vector is much larger compared to a molecule-based descriptor, leading to longer training times.[324;327] Another drawback is that additional hyperparameters have to be optimized, e.g., the cut-off region or widths of Gaussian functions. In case of the first-order expansion used in the FCHL representation, the values σ_P and σ_G with G and P denoting the group and period of the periodic table, enter:

$$N(\mathbf{x}^{(1)}) = A_1(A) = e^{-\frac{(P_A - \chi_1)^2}{2\sigma_P^2} - \frac{(G_A - \chi_2)^2}{2\sigma_G^2}} ; \tag{2.29}$$

$\mathbf{x}^{(1)} = \{P_A, \sigma_P; G_A, \sigma_G\}$ and χ_1 and χ_2 are dummy variables.

As it is visible, handcrafted representations suffer from a large number of hyperparameters, which have to be tuned manually and an improper selection of such internal hyperparameters can limit the prediction capabilities of an ML model.[111;325] However, the task of tuning these hyperparameters is tedious and requires expert knowledge.[111;326] To this aim, automated ways have been provided as an alternative route to allow for the elimination of the manual definition of a molecular representa-

tions.

Automatic molecular descriptors can be generated for instance with message-passing ML models in an end-to-end fashion with one example being the deep continuous-filter convolutional NN SchNet. As the name already suggests, this model is based on a convolutional NN, which is frequently applied to image and speech recognition. The ability of such NNs to detect patterns makes them well-suited for many different types of applications, such as cancer recognition[329] or autonomous driving.[330] However, applying conventional convolutional NNs for quantum chemical problems would not lead to accurate results, because the positions of atoms in a chemical system cannot simply be placed on a grid. So to speak, the image of the molecule would be pixelated.[325]

Therefore, SchNet uses a filter-tensor instead of discrete filters making this model applicable to chemical systems. The filter tensor is part of a larger architecture of SchNet, which is built on the DeepTensor NN.[300] The molecular input is obtained from an atomistic embedding layer that contains information on the atom types:[111]

$$X_A^0 = a_{Z_A} \tag{2.30}$$

with a_{Z_A} representing a randomly initialized embedding. This input is processed into an atom-wise layer:[111]

$$X_A^1 = w^1 X_A^0 + b^0, \tag{2.31}$$

which is a fully-connected layer with weights and biases. Up to now, only the atom type is described and this preliminary representation is further refined with the interactions between atoms obtained from the filter-generating NN. The filter-tensor in form of a fully-connected NN processes the information of the positions of atom A in form of distances to its neighbours B to obtain the filter values: $W(r_A - r_B)$. A convolutional layer is formed by element-wise multiplication of filter values with atomic representations:[111]

$$X_A^2 = \sum_A^{N_A} X_A^1 \circ W^1(r_A - r_B). \tag{2.32}$$

A shifted softplus function is used to transform the information obtained from the convolutional layer to an atom-wise layer. In the end, the atomic contributions are summed to obtain the total energy of a system.

A drawback of this input generation is the larger computational effort, because it requires additional hyperparameters to be optimized during training.[111] However, the automatic design of the most suitable descriptor for a given data set makes an accurate fitting of complex structure-property relationships possible. As the descriptor is designed by an NN, it offers the additional benefit of being automatically differentiable with the tools provided by the open-source ML framework pytorch,[331] which is used to set up the network architecture.[116] The only human input that has to be made regarding the descriptor is to select a cut-off region. The automatically generated molecular representations by SchNet obey all the initially defined required properties a molecular representation has to fulfill and therefore provide highly accurate solutions to chemical problems.[325]

2.3.4 *Training of ML Models*

A well-defined descriptor and a comprehensive training set in combination with a proper training of an ML model should be sufficient to obtain highly precise ML-fitted

functions. A set of weight parameters is required that provides the best solution for relating an input to an output of a training set under investigation. To find these weight parameters, a loss function, such as the L_2 loss defined in equation 2.20, has to be computed at each step during training, i.e., the error an ML model makes on the training instances is evaluated at each step. The purpose of training is to update the adaptive fitting parameters of the ML model in order to minimize this error. In case of the L_2-norm, the resulting convex minimization problem for linear models,

$$min_w = \sum_{\beta=1}^{N_M} \left((b + wX_\beta) - Y_\beta^{QC} \right)^2,$$
(2.33)

can be solved by setting the gradient with respect to the weights to zero. Equation 2.23 is used to find the weights for KRR models. Unfortunately, in case of NNs, the problem must be solved numerically, as NNs are highly nonlinear, flexible functions. Stochastic gradient descent optimization algorithms can be used to achieve a step-wise update of weight parameters, i.e., from step k to step $k + 1$:

$$w_{k+1} = w_k - l_r \nabla L_2(w)$$
(2.34)

The step size, l_r, also known as the learning rate, is one of the most important hyper-parameters, which is difficult to determine. While large step sizes lead to fast and most often sub-optimal results, small step sizes enable an NN to learn slowly, but might trap it in local minima and generally require longer training times. Therefore, algorithms, such as ADAM (adaptive moment estimation)[332] or AdaGrad,[333] have been developed to automatically adapt the learning rate. Similarly, the same is to be done for the adaptive bias.

A common problem when training an ML model is over-generalization of the training samples, also known as overfitting. In other words, overfitting refers to the case, when data inside of the training set is fitted (almost) exactly, including (numerical) noise. This effect can lead to very small errors on the training instances, but to large error on test samples, limiting the prediction capability mostly to data represented in the training set. One often applied technique to mitigate this effect is to apply regularization. This is for example the role of the parameter λ used in KRR models as given in equation 2.23 that controls the strength of regularization. In case of quantum chemical data, the noise is considered to be small, hence λ is also small in most cases.[123;304;334]

Another way to prevent from overfitting is early stopping,[82] commonly applied when training NNs. To this aim, the data set used for training is split, often in a ratio of 9:1, into a "direct training" set and validation set. The "direct training" set is used to optimize the parameters by minimizing the loss function, as explained below, but additionally the error on the validation set is monitored. As soon as this error starts to rise for several sequential steps, the model is assumed to be overfitting. The training process is then stopped and the set of weights, with which the smallest error on the validation set was obtained, is stored and used for inferences.[82]

Additionally and separately from the fitting parameters that are automatically optimized during training, the internal hyperparameters of an ML should be optimized for a given training set. This can be done via K-fold cross-validation[335] or random grid search[82] for example. All of these methods have in common that a set of hyperparameters is sampled and respective ML models are trained. The ML models that differ in their hyperparameters can be ranked by comparing their error

on a separate validation set that is not used for training. These parameters can then be optimized in several iterations by trying to minimize the error of a trained ML model on this validation set.[314]

The quality of the most optimal ML model, i.e., the model with optimized hyper-parameters and fitting parameters, can be assessed by computation of the error on a separate test set that has neither been used for training, nor for optimization of hyperparameters.[314] The learning efficiency and behaviour of ML models can be reported with learning curves by plotting the number of instances used for training and the corresponding error on the test set in a logarithmic scale. A linear function is obtained when the ML models are trained properly.[336;337]

In order to use an ML model for MD simulations and to guarantee that the energy of a system is conserved, the forces need to be treated as derivatives of ML PESs.[106] In case of NNs, the training process can be adapted by including the forces, F, as NN derivatives into the loss function:

$$L_2 = \frac{1}{N_M} \sum_{\beta}^{N_M} (E_{\beta}^{NN} - E_{\beta}^{QC})^2 + \frac{\eta}{3N_M \cdot N_A} \sum_{\beta}^{N_M} \sum_{A}^{3N_A} (F_{\beta A}^{NN} - F_{\beta A}^{QC})^2. \tag{2.35}$$

The hyperparameter η can additionally be optimized to balance the influence of forces and energies during training. Remarkably, treating the forces as derivatives of NN PESs with properly weighted forces in equation 2.35 does not only improve the accuracy of forces, but also of energies.[106]

In order to compute atomic forces, $\mathbf{F}_A$, as NN derivatives with respect to Cartesian coordinates, $\mathbf{R}_A$, of this atom, the chain rule can be used:

$$\mathbf{F}_A = -\frac{\partial E^{NN}}{\partial \mathbf{R}_A} = -\sum_{\varrho}^{N_D} \frac{\partial E^{NN}}{\partial D_{\varrho}} \cdot \frac{\partial D_{\varrho}}{\partial \mathbf{R}_A} \tag{2.36}$$

The partial derivatives of the PESs with respect to the molecular descriptor of length N_D are combined with the partial derivatives of each descriptor with respect to Cartesian coordinates of an atom. The implementation of this derivation is a rather tedious task that has to be done for each hand-crafted descriptor separately.[106;313] The ability to automatically obtain the forces as derivatives when using end-to-end architectures provides a clear advantage in this regard.[325]

With kernel methods, some procedures to describe forces exist, e.g., with Gaussian Process Regression and SOAP,[117] with the sGDML approach[118;118] or as response properties for KRR and FCHL.[132] The latter applies a force operator, i.e., the negative of the atomic gradient operator:[132]

$$\mathbf{F} = \hat{F} \left[E^{KRR} \right] = \hat{F} \left[\mathbf{K}_1 \right] \mathbf{w}, \tag{2.37}$$

with E^{KRR} being the energy of a system obtained as a sum of atomic contributions. By using this implementation, the accuracy of energy and force predictions increases considerably. As a drawback, in the specific case of KRR with the FCHL representation, also the size of the kernel matrix increases from $N_M \times N_M$ to $3N_A N_M^2 \times 3N_A N_M^2$, thus increasing the computational costs of the training process and also the associated memory requirements.

DEVELOPED METHODS AND STRATEGIES

In this chapter newly developed ML methods, strategies and adaptations of existing algorithms for the description of the excited electronic states of molecules are presented. Since the reference data forms the basis of any developed ML model, this chapter starts with the adaptive sampling scheme for excited states combined with a phase correction algorithm to render excited-state properties learnable. Subsequently, the developed ML models to describe the excited electronic states will be discussed that are mostly built on regressors for the electronic ground state energies and forces of a molecule. The discussion is dedicated to the introduction of single-state, single-property, multi-state, and multi-property ML models with KRR, MLFF NNs and SchNet as regressors. Approximated ML NACs based on first and second derivatives of ML PESs with respect to atomic coordinates are introduced. Reprints of the published articles that provide detailed information and further discussion on the method developments are part of the appendix.

3.1 ADAPTIVE SAMPLING FOR EXCITED STATES OF MOLECULES

The adaptive sampling scheme for excited electronic states of molecules is based on the description of Behler[153] and the algorithm of Gastegger et al.[101] for the electronic ground state of molecular systems (as described in section 2.3.1). An overview of this scheme is illustrated in Fig. 6 with two NNs and the methylenimmonium cation, $CH_2NH_2^+$, as an example. Therefore, the scheme will be discussed with two NNs, but it can be used with any number of NNs greater than one and any other type of regressor.

As can be seen in Fig. 6, the process starts with initial quantum chemical calculations to create a small, preliminary data set. In principle, any sampling technique to generate this initial set of data points can be used. In our case, sampling along different normal modes and combinations thereof obtained from the equilibrium geometry of the molecule turned out to be beneficial. In this way, conformational regions close to the equilibrium structure of a system can be sampled comprehensively and accurate initial ML PESs can be obtained, which is advantageous for the adaption of the training set with MLMD simulations. Additionally, some reaction coordinates that are deemed to be important for the excited-state dynamics of a molecule can be included right from the beginning. These can comprise for instance dissociative reaction channels. Approximately 1,000 initially sampled data points seem to be a good starting point for small to medium-sized molecules.

If excited-state properties, such as NACs or SOCS are included to describe the hopping probability between different electronic states with ML models, these initial quantum chemical data points should be phase corrected to remove inconsistencies of such properties as it is exemplified in Fig. 4 in chapter 2.2.4 as a result of the arbitrary phase of the electronic wave function. This task can be achieved with a phase correction algorithm that was developed during this work and will be described in chapter 3.2.

Assuming a proper initial training set is available, at least two preliminary ML models, denoted as NN1 and NN2 in Fig. 6, are trained on these data points. Excited-state MD simulations are initiated using the ML models instead of the electronic
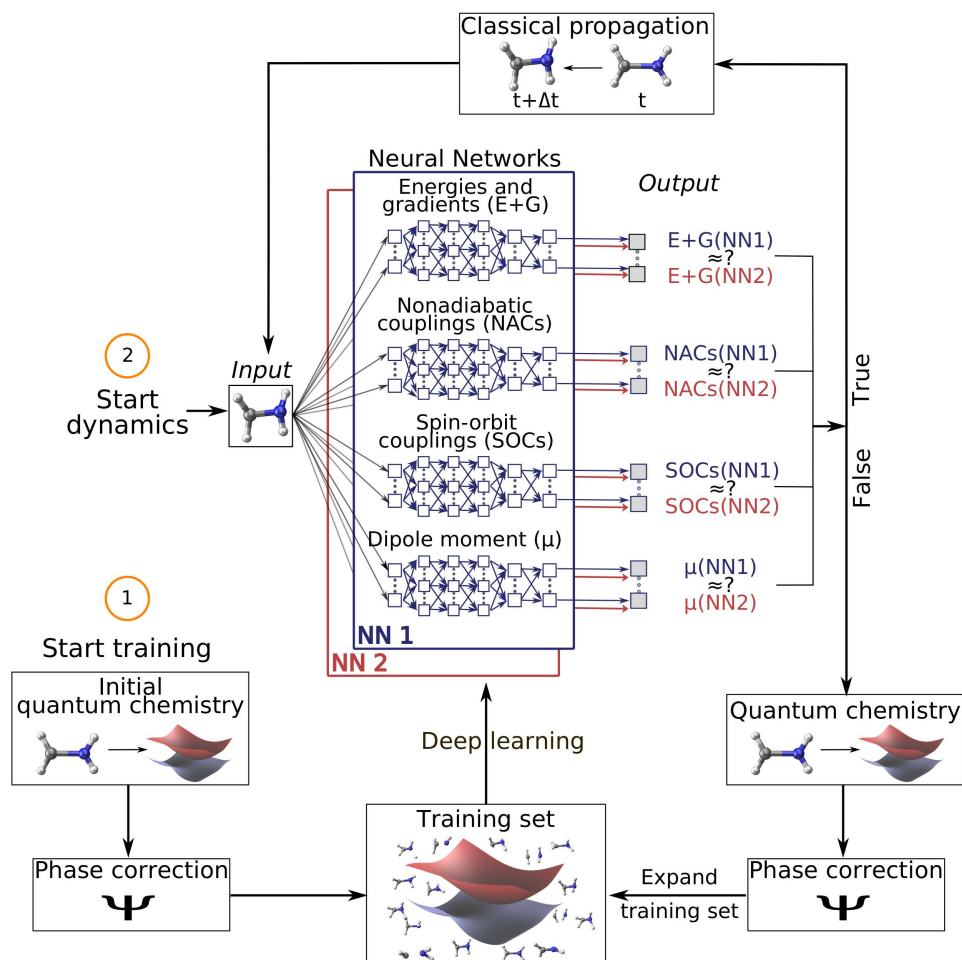
Figure 6: Illustration of the adaptive sampling scheme for excited states with two NNs (NN1 and NN2) and the methylenimmonium cation, $CH_2NH_2^+$, as an example. The scheme starts with an initial training set provided by a quantum chemical reference method that is additionally phase corrected. Deep learning models are trained on this training set and a molecular input obtained from an MD program can be used to start dynamics simulations. The energies (E), gradients (G), NACs, SOCs and transition and permanent dipole moments (μ) are predicted with both NNs separately. The predictions are compared to each other and the standard deviation is computed. Whenever the two NNs predict similar values, the mean of the independent predictions is passed to the MD program to classically propagate the nuclei and generate the molecular input of the next time step that can be given to the NNs again. Whenever the standard deviation of the NNs exceeds a pre-defined threshold for one of the trained properties, a quantum chemical calculation of this data point is carried out, the data point is phase corrected, added to the training set, NNs are re-trained and MLMD simulations are re-started. This loop is carried out until the training set is sufficiently large for the problem under investigation. (Reproduced from Ref. 77 under CC-BY, https://creativecommons.org/licenses/by/3.0/.)

structure method. The starting points for simulations are obtained from the initial sampling of geometries after vertical excitation to an excited state. The molecular input is transformed into the chosen molecular descriptor and the NNs predict the energies, gradients, SOCs, NACs, and, if needed, dipole moments (μ). The mean of the two NNs is passed to the MD program.

Besides the phase correction algorithm, a greater amount of properties and electronic states has to be considered, which is different to the adaptive selection scheme

proposed for the ground state of a molecular system.[106;153] Each property as well as each electronic state should be fitted accurately and should be allowed to influence the sampling procedure. Therefore, the standard deviation of the ensemble of the ML models is computed for energies and gradients as given in equation 3.1 and for properties coupling two different electronic states as given in equation 3.2. N_S refers to the number of electronic states and $Y_{\sigma_i}^{ML_J}$ indicates that a property is a result of a single electronic state of ML model J, which is the case for energies, gradients and permanent dipole moments. $Y_{\sigma_{ij}}^{ML_J}$ indicates that a property results from electronically different states of ML model J, which is the case for NACs, SOCs, and transition dipole moments. The mean of the standard deviations for each electronic state,

$$Y_{\sigma_i}^{ML} = \frac{1}{N_S} \sum_i^{N_S} \left(\sqrt{\frac{1}{N_{ML}-1} \sum_{J=1}^{N_{ML}} \left(Y_i^{ML_J} - \overline{Y}_i^{ML}\right)^2} \right), \tag{3.1}$$

and for each pair of states,

$$Y_{\sigma_{ij}}^{ML} = \frac{1}{2N_S^2} \sum_i^{N_S} \sum_j^{N_S} \left(\sqrt{\frac{1}{N_{ML}-1} \sum_{J=1}^{M} \left(Y_{ij}^{ML_J} - \overline{Y}_{ij}^{ML}\right)^2} \right), \tag{3.2}$$

is considered for computing the standard deviations of the ensemble of the network for excited-states. Consequently, also the mean of the properties, $\overline{Y}_i^{ML}$ and $\overline{Y}_{ij}^{ML}$, has to be computed as the average of the means of NN models resulting from different electronic states and pairs of states, respectively.

A separate threshold is set for each of the fitted properties and is compared to the standard deviation of the respective property. The value of each threshold is slightly larger than the RMSE of ML predictions on the validation set that is used to control overfitting during training of NN models. Whenever one of the thresholds is exceeded, a molecular geometry is recomputed with quantum chemistry and included in the training set. As it is described in Ref. 106, also here an ensemble of independent trajectories is computed in parallel and only after all trajectories have reached a point that belongs to an undersampled or unknown region of the PESs, the ML models are re-trained with the considerably larger training set. After each time that ML models are re-trained and MD simulations are re-started with more robust ML models, the threshold is multiplied with a factor smaller than 1.

There might be cases, when the standard deviation of the ML models with respect to the NACs exceeds the threshold to a large extent. This can happen close to conical intersections, at which point the reference NACs should be singular. The trajectories are aborted mostly due to the problematic NACs in these regions. The threshold for NACs can then be set to a large value so that the NACs do not influence the sampling procedure when passing through a conical intersection. In this way, critical regions can also be sampled and trajectories are allowed to go beyond such critical regions. The threshold for NACs can then be reduced again. Re-starting the adaptive sampling from the geometry at which point the simulation was interrupted, guarantees that a wide range of the PESs is visited.

Whenever the ML predictions for all properties are similar enough to each other, MLMD simulations are carried out in the same way as MD simulations would be carried out with the electronic structure reference method. Based on the active state, the MD program extracts the relevant information from ML predictions and can propagate the nuclei according to classical mechanics. Because Wigner sampling is

often applied to generate the starting geometries for excited-state MD simulations with SHARC, the initial training set obtained from normal mode sampling should be comprehensive enough to allow for accurate ML PESs in the beginning of MLMD simulations. As soon as the PESs are not accurate anymore, the adaptive sampling procedure guarantees that relevant data points are automatically added to the training set.

Several loops of this procedure and re-optimizations of hyperparameters of the NNs might be necessary during this process. This has to be done in principle until the training set is converged. Whether a training set is converged or not depends strongly on the purpose of the study and the molecule under investigation. Flexible molecules with a high density of states often require a lot of data points, because many different reaction channels can be reached during an excited-state MD simulation. Molecules with a less diverse photochemistry can lead to much faster training set convergence. The ability to compute many consecutive time steps with ML models without having to abort a trajectory can indicate that the PESs are converged at least to a large extent. A comparison with the reference dynamics for a short time scale is also helpful to assess the convergence of the ML PESs. Additionally, the prediction of reaction coordinates that are deemed to be important can provide a measure of uncertainty. A good indicator that ML PESs are not converged are fluctuations in the total energy of a system along an MLMD run and the occurrence of unphysical conformations that can fall apart. In this sense, executing the production runs with an ensemble of ML models and comparing their predictions at each time step is beneficial to guarantee an interpolative regime at each time step. Using an ensemble instead of a single ML model is further advantageous, because more robust MLMD simulations can be achieved [106], while the additional costs due to the ensemble of ML models are usually minor. The adaptive sampling scheme for excited states is described in more details in the corresponding publication 77 in chapter A.1.

3.2 PHASE CORRECTION ALGORITHM

Whenever properties are fitted that arise between different electronic states, the arbitrariness of their signs has to be taken into account. This problem, shortly introduced in section 2.2.4, can be solved to a large extent by applying a phase correction algorithm to preprocess the training set. The idea and goal behind this process is summarized in Fig. 7 (reproduced from Ref. 77).

Panel (a) shows five different conformations of $CH_2NH_2^+$ that differ in their bond length between the carbon and the nitrogen atom. This input is given to an electronic structure program. For each molecular geometry, the energy is computed along with the coupling value between the S_1 and S_2 state, $Y_{12}^{QC} = \langle S_1 | \hat{H} | S_2 \rangle$. The wave function carries an arbitrary phase for each electronic state, similar to a positive or a negative sign that does not change the characteristic of the wave function of being an eigenfunction of the electronic Hamilton operator. Illustrating the phase of the wave function is difficult and molecular orbitals are shown instead as replacements of the electronic wave function in the eigenvalue equation, so that $Y_{12}^{QC} = \langle \blacktriangleright | \hat{H} | \blacktriangleright \rangle$. As can be seen in panel (b), the color of the molecular orbitals can be either blue or red, similar to carrying a pre-factor of +1 or -1. As this pre-factor is not uniquely defined, it can switch arbitrarily along the reaction coordinate, leading to an arbitrary sign of the coupling value, which is illustrated at the bottom of the middle panel. A positive sign is the result of: $Y_{12}^{QC} = \langle \blacktriangleright | \hat{H} | \blacktriangleright \rangle = \langle \blacktriangleright | \hat{H} | \blacktriangleright \rangle$, whereas a negative
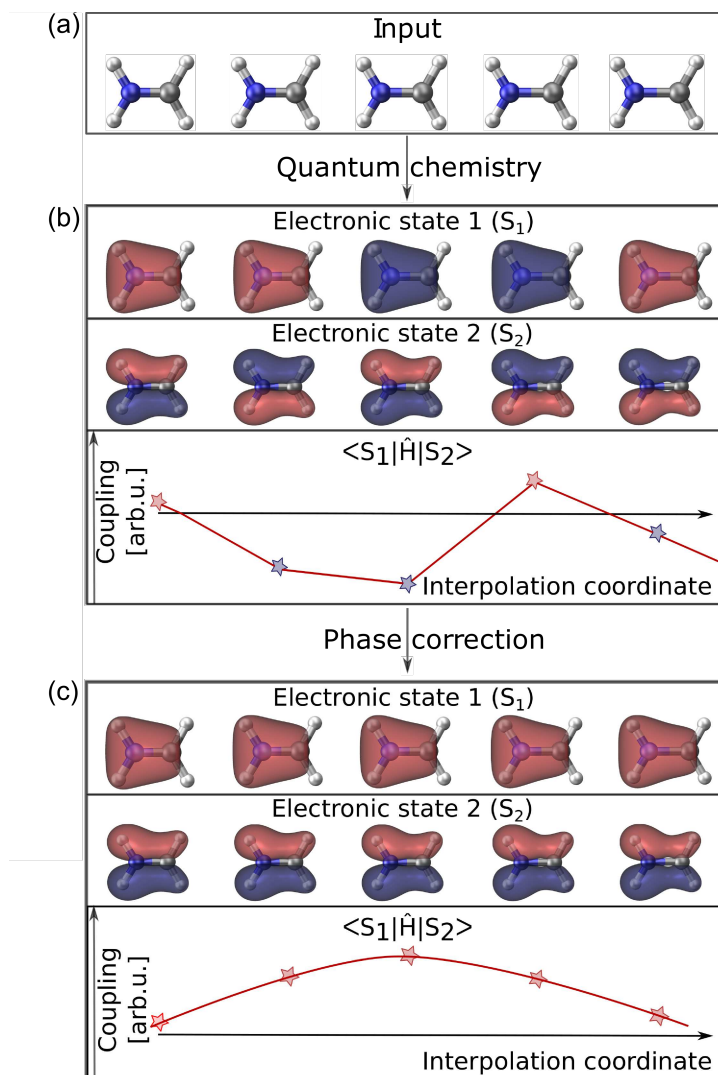
Figure 7: Phase correction algorithm exemplified with molecular orbitals as placeholders for the electronic wave functions of two excited electronic states of the methyenimmonium cation, $CH_2NH_2^+$. Panel (a) shows the molecular input with molecules of slightly different bond lengths between the carbon and nitrogen atom. (b) Along this reaction coordinate, quantum chemical solutions to a coupling value, $\langle S_1|\hat{H}|S_2\rangle$, are shown with the corresponding molecular orbitals. (c) The arbitrariness with respect to the sign of the coupling value can be removed by applying phase correction. (Reproduced from Ref. 77 under CC-BY, https://creativecommons.org/licenses/by/3.0/.)

sign is obtained from: $-Y_{12}^{QC} = \langle$ 🗎 $|\hat{H}|$ 🗎 $\rangle = \langle$ 🗎 $|\hat{H}|$ 🗎 $\rangle$. Compared to the energy of a system, where the same electronic state enters the eigenvalue equation twice, e.g., $E_1 = \langle$ 🗎 $|\hat{H}|$ 🗎 $\rangle = \langle$ 🗎 $|\hat{H}|$ 🗎 $\rangle$, the phase factor does not cancel out for elements of different states.

Indeed, both outcomes, $\pm Y_{12}^{QC}$, are equally likely and the arbitrariness in their signs makes elements resulting from two different electronic states particularly challenging to fit. It even prohibits the use of conventional ML training algorithms[77;208;257;290] and needs to be removed in order to make the data learnable with standard ML models. A consistent phase is further important for the computation of the hopping probabilities in MQC dynamics simulations as arbitrary signs in the coupling values can lead to wrong hopping probabilities.[72] The learning of a consistent coupling

surface is thus indispensable for MLMD surface hopping techniques.

How to correct the coupling values in a one-dimensional space is shown in the bottom of panel (c) of Fig. 7 and might be straightforward in this dimension. In case of two states, only two possible solutions exist to correct the phases, which is clearly visible in Fig. 7. However, the complexity of this problem increases with the number of data points that have to be phase corrected, with the degrees of freedom a system has, and also with the amount of electronic states considered. The number of possible solutions for one data point scales with a factor of 2^{N_S-1}. To eliminate the problem of arbitrary sign jumps in coupling values that rapidly becomes infeasible to account for manually, a phase correction algorithm was developed that automatically corrects the coupling and transition dipole moment values and provides a consistent sign within a data set.

The prerequisite of the phase correction algorithm is a reference computation, e.g., of the equilibrium geometry of the system under investigation. The wave function of this calculation has to be stored and provides the reference to all further data points that should be included in the training set. Every new data point that is added to the training set is then corrected with respect to the reference phases. To this aim, a phase vector is determined for each data point inside of the training set from the overlap matrix, $\mathbf{S}$, that is computed with SHARC according to equation 2.15[258] with respect to the reference wave function. Each entry of a phase vector, $\mathbf{p}$, of length N_S corresponds to a certain electronic state and determines whether a phase change occurred or not:

$$p_i = sgn\left(max\left(|S_{ij}|\right) sgn\left(S_{ij}\right)\right) \forall |S_{ij}| \geq 0.5; \quad i,j = 1,2,...,N_S \tag{3.3}$$

Single- or complex-valued properties can then be corrected by multiplication of each value resulting from state i and j with the values of the phase vector corresponding to state i and j. Similarly, the vectorial properties, such as NACs or transition dipole moments, can be corrected by multiplication of the whole vector with the corresponding values of +1 or -1:

$$\widetilde{Y}_{ij}^{QC} = Y_{ij}^{QC} \cdot p_i \cdot p_j; \quad i,j = 1,2,...,N_S \tag{3.4}$$

Similar molecular geometries often guarantee a proper phase correction. If conformations differ strongly from each other, a phase correction can be prohibited. In such cases, the wave functions do not overlap sufficiently anymore and no information on the phase can be obtained. For every data point with wave function overlaps smaller than a threshold of 0.5, interpolation between the reference geometry and the geometry to be included in the training set is required. A sequence of N_{int} wave function overlap computations from the reference geometry to the targeted molecular conformation has to be carried out and the phase vectors for each molecular input with indices d running from 0 to N_{int} has to be determined. The phase vector, p_{last}, corresponding to the molecular geometry furthest apart from the reference geometry, is obtained by multiplication of all precedent phase vectors:

$$\mathbf{p}_{last} = \prod_{d=0}^{N_{int}} \mathbf{p}_d. \tag{3.5}$$

Importantly, intruder states can influence the phase correction algorithm and can provide a source of error. They play a crucial role in systems that have a high density of states. Electronic states, which are high in energy at the reference geometry so

that they are not considered in the beginning, can enter along a reaction coordinate. Consequently, no phase information on these states can be obtained. The inclusion of additional states for the wave function overlap computation from the beginning can help to identify the phases of the relevant states at another molecular geometry.

A proper phase correction for systems with many close-lying electronic states can easily require about 100 electronic states to be considered. In the form presented here, the phase correction algorithm becomes very expensive in such cases. The strategy of including a large number of electronic states only close to the equilibrium geometry and reducing them along a reaction coordinate as they become irrelevant[338] can be applied to lower the computational costs for accurate phase correction. Saving electronic wave functions of distinct conformations in addition to the electronic wave function of the reference geometry can further reduce the computational costs, but requires large amount of memory or hard disk space.

Nevertheless, with all these tricks in mind, the computation of a training set with the phase correction algorithm is applicable to many molecular systems, i.e., those systems with a moderate density of electronic states in relevant conformational regions of the PESs. The data can be used to train conventional ML models that can be applied to surface hopping excited-state MD simulations. Furthermore, transition dipole moments can be fitted. The phase correction algorithm is published in Ref. 77 and part of the appendix A.1, where a detailed discussion of the problems and the benefits of this algorithm is provided.

3.3 PHASE-FREE TRAINING

Another way to cope with the arbitrariness in the electronic structure calculations of properties obtained from two different electronic states is to apply a phase-free training algorithm. The phase-free training was developed in this work and was implemented in the SchNarc approach for photodynamics[157] that is discussed in chapter 3.4.3.

In order to render a phase-free training possible, the loss function has to be adapted. The conventional L_2 loss as specified in equation 2.20 forms the basis of the new loss-function, termed phase-less loss function, L_{ph}. Several variants of the loss function exist that depend on the number of properties that are included in the training process. The idea is to compute the MSE of a property several times taking all possibly relevant phase combinations into account. In all variants, the phase-less loss function takes the minimum of all computed MSEs, $\{\varepsilon_{ph}\}$, that arise from different possible phase combinations:

$$L_{ph} = min\left(\{\varepsilon_{ph}\}\right) \tag{3.6}$$

In case single- or complex-valued properties are trained (with dimension smaller or equal to 2), i.e., SOCs, the most efficient way is to compute the entries of $\{\varepsilon_{ph}\}$ according to the following equation:

$$\{\varepsilon_{ph}\} = \{\varepsilon_{ph}^+, \varepsilon_{ph}^-\} = ||\ Y_{jk}^{QC} \pm Y_{jk}^{ML}\ ||^2 \quad if\ dim(Y) \leq 2. \tag{3.7}$$

As can be seen from the previous equation, the MSE is computed twice, once assuming a positive and once assuming a negative sign for a coupling value between state j and k. The minimum function guarantees that the sign producing the minimal error is used for the SOC value that influences the training process. The two computed MSEs, i.e., the one resulting from a positive sign and the one resulting from a negative

sign, can further be used in another variant, which is more suitable for vectorial properties, i.e., properties with dimension ≥ 3:

$$\{\varepsilon_{ph}\} = \varepsilon_{ph} = \varepsilon_{ph}^{-} \cdot \zeta^{+} + \varepsilon_{ph}^{+} \cdot (1 - \zeta^{+}) \quad \text{if } dim(Y) \geq 3 \tag{3.8}$$

with

$$\zeta^{+} = \frac{\varepsilon_{ph}^{+}}{\varepsilon_{ph}^{-} + \varepsilon_{ph}^{+}}. \tag{3.9}$$

$\varepsilon_{ph}^{\pm}$ is computed as given in equation 3.7. Noteworthy, the relative signs of the values within one vector remain and have to be described correctly so that the vector does not point into an arbitrary direction.

Whenever MQC dynamics simulations with the surface hopping method should be enhanced with ML, the phases of the NACs and SOCs can not be treated independently from each other as inconsistent phases can potentially lead to wrong hopping probabilities. In this case, a combined loss function should be used that computes the error 2^{N_S-1}-times as given in equation 3.6 with each entry, ε_{ph}^{f}, being computed according to equation 3.11.

$$\{\varepsilon_{ph}\} = \{\varepsilon_{ph}^{1}, \varepsilon_{ph}^{2}, ..., \varepsilon_{ph}^{2^{N_S-1}}\} \tag{3.10}$$

$$\varepsilon_{ph}^{f} = \begin{cases} \frac{1}{N_S^2} \sum_{i}^{N_S} \sum_{j \neq i}^{N_S} \frac{1}{N_A} \sum_{A}^{N_A} || Y_{ij,A}^{QC} - Y_{ij,A}^{ML} \cdot p_i^{f} \cdot p_j^{f} ||^2 & \text{if } dim(Y) \geq 3 \\ \frac{1}{N_S^2} \sum_{j}^{N_S} \sum_{k \neq j}^{N_S} || Y_{jk}^{QC} - Y_{jk}^{ML} \cdot p_j^{f} \cdot p_k^{f} ||^2 & \text{if } dim(Y) \leq 2 \end{cases} \tag{3.11}$$

The phase-less loss function enables the computation of the appropriate phase vector for each data point inside of the training set and can therefore be used to describe NACs together with SOCs for surface hopping MD simulations with ML. The whole procedure can be seen as an ML-based internal phase correction for each data point. The fact that ML models are defined as smooth functions allow for a consistent sign during predictions and MLMD simulations.

3.4 ML MODELS FOR EXCITED ELECTRONIC STATES

Since ML approaches in quantum chemistry were designed almost exclusively for fitting ground-state energies, properties, derivatives thereof, and other single-valued features, such as energy gaps, existing methods could only be used to train one electronic state at a time. The treatment of a manifold of excited electronic states with these conventional models requires as many ML models as electronic states. New models need to be developed that can handle a manifold of excited state at once. As the excited states are inherently linked to each other, ML models can possibly benefit from the information provided about their interplay.

In this chapter, several terms are defined: Single-state, multi-state, single-property, and multi-property. The first term can be used for conventional ML models that can describe a single value. With respect to the energy of a system this is a single electronic state. Similarly, this term is used to indicate that only one coupling value can be fitted with one ML model, hence a vector of $N_S \times N_S$ entries needs $N_S \times N_S$ independent ML models to be fitted. Multi-state then refers to the possibility of learning several values at once, such as the energies of different electronic states

including different spin-multiplicities or properties that couple different electronic states or all values within one vector. Single-property is then used for single-state or multi-state ML models that can only learn one property at once, e.g., the SOCs or the energies of a system including their derivatives. Multi-property models would then be able to treat coupling values in addition to energies and derivatives as well as (transition) dipole moments.

3.4.1 Kernel Ridge Regression

KRR is per standard definition (equation 2.21) a single-state single-property model that can relate a molecular input to a single valued property. In order to enable the prediction of several electronic states at once, another kernel is necessary, i.e., a Gaussian state kernel, $K_2(S_{\alpha'}, S_{\beta'})$, that maps a state of interest, $S_{\alpha'}$, to all available states:

$$Y^{ML}(X_\alpha, S_{\alpha'}) = \sum_{\beta}^{N_M} \sum_{\beta'}^{N_S} w_{\beta\beta'} K_1(X_\alpha, X_\beta) K_2(S_{\alpha'}, S_{\beta'}). \tag{3.12}$$

This state kernel is similar to a conventional Gaussian kernel that maps one molecular geometry to all other geometries in the training set and measures the similarity between them. However, the conventional KRR model only depends on the molecular geometry as an input and can therefore only relate one value corresponding to one state. As it is given in equation 3.12, an encoding of the electronic state by using an additional Gaussian kernel allows for N_S predictions for one molecular input.

To define the state kernel, the states also have to be presented to the model using a state representation, similar to the molecular descriptor. In principle, the representation of the state can be almost any value, but similarly to the conventional kernel for a molecular geometry, the width of the state kernel has to be adapted accordingly and is dependent on the representation used. The KRR models were implemented into the SHARC code using the QML[339] toolkit. A more detailed discussion on the state representation and KRR for excited electronic states of molecules can be found in appendix A.2.

3.4.2 Multi-Layer Feed-Forward Neural Networks (NNs)

MLFF NNs were developed using the python library theano[340] and the inverse distance matrix, D (equation 2.27), as a descriptor for molecular geometries. Due to their flexibility, NNs allow for a straightforward implementation in a multi-state fashion.

One molecular input is mapped to a vector consisting of the energies of several states including also states of different spin-multiplicities. The energies, E, are trained together with forces, F, in order to conserve energy during MD simulations and to improve the accuracy of the energies. The scheme of Ref. 106 explained in

equation 2.35 is adapted to incorporate contributions from different electronic states and the errors corresponding to different electronic states are averaged for training:

$$L_2 = \frac{1}{N_M \cdot N_S} \sum_{\beta}^{N_M} \sum_{i}^{N_S} \left(E_{\beta i}^{ML} - E_{\beta i}^{QC} \right)^2 \tag{3.13}$$

$$+ \frac{\eta}{3 N_M \cdot N_S \cdot N_A} \sum_{\beta}^{N_M} \sum_{i}^{N_S} \sum_{A}^{3 N_A} \left(F_{\beta i A}^{ML} - F_{\beta i A}^{QC} \right)^2 .$$

The SOCs, NACs and dipole moments are fitted with separate NNs, whereas permanent and transition dipole moments are treated together. All values of each property that are descried with one NN are summarized in one vector. The loss function is then the average obtained from MSEs of all values within a vector. The single-property multi-state NNs are part of the publication given in appendix A.1.

In addition to these single-property models, NNs further allow for joint fitting of different properties. The properties are summarized in a straightforward manner in one vector and mapped to one molecular input. The forces can still be treated as derivatives of ML PESs and can be included in the loss function for training as described above. Additionally, a state-representation similar to the state kernel introduced for KRR to treat several excited electronic states is defined for the NNs. This state representation in addition to the multi-property model is summarized and explained further in appendix A.2.

All models based on MLFF NNs can be used to execute surface hopping MQC simulations with SHARC using coupling values to compute the hopping probabilities. The formula given in equation 2.36 is used to compute forces as derivatives of NN PESs with respect to atomic coordinates of a system and is thus expanded for more electronic states. This has been implemented for the inverse distance matrix. Moreover, the computation of the Hessian matrix, $\mathbf{H}$, is implemented and the entries for atoms A and B are calculated using the following equation:

$$\mathbf{H}_{AB} = \frac{\partial^2 E^{NN}}{\partial \mathbf{R}_A \mathbf{R}_B} = \sum_{\varrho}^{N_D} \left(\frac{\partial E^{NN}}{\partial D_\varrho} \cdot \frac{\partial^2 D_\varrho}{\partial \mathbf{R}_A \partial \mathbf{R}_B} \right) + \sum_{\varrho}^{N_D} \sum_{\varsigma}^{N_D} \left(\frac{\partial^2 E^{NN}}{\partial D_\varrho \partial D_\varsigma} \cdot \frac{\partial D_\varrho}{\partial \mathbf{R}_A} \cdot \frac{\partial D_\varsigma}{\partial \mathbf{R}_B} \right) . \tag{3.14}$$

This equation has to be calculated for each electronic state and N_D is the length of the molecular descriptor.

3.4.3 *SchNarc*

The SchNarc approach for photodynamics simulations is based on the deep-learning model SchNet[111] and is made publicly available on github. The toolbox SchNet-Pack[116] was extended to allow for a description of a manifold of excited singlet, doublet, triplet, and quartet states including their first and second derivatives, couplings and dipole moments. The package comprises two main building blocks, i.e., a representation and a prediction block. The former has not been adapted, hence SchNarc profits from the end-to-end architecture of SchNet that automatically determines a well-suited descriptor for a given training set. The derivatives of PESs, i.e. forces and Hessian matrices, can be obtained for the separate electronic states as well.[331]

In SchNarc, each property is treated in a separate prediction block that decomposes a molecular property into atom-wise contributions. The forces are trained as

derivatives of ML PESs. Whenever phase corrected data are provided, the L_2 loss function can be used that combines the MSEs of all different properties. A separate tradeoff-value, t, can be set for each property, i.e., t_E for energies, t_F for forces, t_{SOC} for SOCs, t_{NAC} for NACs, and t_μ for permanent and transition dipole moments:

$$L_2 = t_E||E^{QC} - E^{ML}||^2 + \Lambda t_F||F^{QC} - F^{ML}||^2 +$$
$$t_{SOC}||C_{SOC}^{QC} - C_{SOC}^{ML}||^2 + t_{NAC}||C_{NAC}^{QC} - C_{NAC}^{ML}||^2 + t_\mu||\mu^{QC} - \mu^{ML}||^2, \tag{3.15}$$

The tradeoff-values have to be assessed for each property and training set and determine how strong a property influences the training process. If a tradeoff is set to 0, then the property is not considered at all. The optional value, Λ, that is set to 1 per default, can be defined for each data point within a training set. Whenever a data point does not contain forces, Λ is automatically set to 0. In this way, the computational costs for the training set generation can be reduced by allowing the inclusion of data points without forces. The training sets have to be provided in the format of the atomic simulation environment,[341] which is also compatible with SchNetPack. The default loss functions are the phase-less loss functions defined in equations 3.6-3.11. Depending on the properties available, SchNarc determines automatically the most suitable loss-function, while it always resorts to the simplest possible loss available.

In contrast to the previously mentioned representation of NACs and dipole moments with MLFF NNs and KRR models, these vectorial properties are treated in a rotational covariant way. For completeness of the SchNarc approach, dipole moments can be trained according to the scheme proposed in Ref. 106 using the following relation:

$$\mu_{ij} = \sum_A^{N_A} q_{ij,A} \mathbf{R}_A^{CM} \tag{3.16}$$

with $\mathbf{R}_A^{CM}$ being the distance in Cartesian coordinates from atom A to the center of mass of the molecule under investigation,. The charges, $q_{ij,A}$, are initiated randomly and fitted during the training process indirectly in order to match the reference vectors for permanent and transition dipole moments rather than letting the charges match any reference charges obtained from quantum chemistry,[106] but have not been applied yet.

3.5 ML-FITTED NACS

NACs are treated as single values with KRR, in one vector all together with MLFF NNs and as derivatives of virtual ML properties with respect to Cartesian coordinates with SchNet. The virtual property in the latter representation is the anti-derivative of $\langle \Psi_i | \frac{\partial H_{el}}{\mathbf{R}} | \Psi_j \rangle$. The mathematical derivation of this formula and the experiments proving the validity of this approach can be found in the supporting information of Ref. 157.

Moreover, NACs can be trained in two different ways: They can be trained and predicted as directly obtained from quantum chemistry (as specified in equation 2.9 in chapter 2.1) or can be scaled with the energy gaps between the coupling electronic states. The idea of the latter approach is to remove the sharp peaks of NACs in critical regions of the PESs by multiplication with the corresponding energy gaps, $\Delta E_{ij} = |E_i - E_j|$ for $C_{NAC_{ij}}$ elements:

$$\widetilde{\mathbf{C}}_{NAC_{ij}}^{QC} = \mathbf{C}_{NAC_{ij}}^{QC} \cdot \Delta E_{ij}^{QC}. \tag{3.17}$$

These values are then used for training instead of those obtained directly from quantum chemistry. For prediction, the so-obtained smooth NACs need to be divided by the energy gaps obtained from the ML models to capture their behaviour around critical regions correctly:

$$\mathbf{C}_{NAC_{ij}}^{ML} = \frac{\widetilde{\mathbf{C}}_{NAC_{ij}}^{ML}}{\Delta E_{ij}^{ML}} \tag{3.18}$$

This scheme is explained comprehensively in Ref. 158 and part of the appendix A.2.

3.5.1 Approximations to NACs

As NACs are an expensive part of an electronic structure calculation and are additionally challenging to fit accurately, ML approximations to NACs were developed. These are based on the squared energy-gap Hessians, which characterize the branching space vectors, i.e., the NAC vectors and gradient difference vectors. The squared energy-gap Hessian is computed from ML PESs via the following relation:

$$\frac{\partial^2 (\Delta E_{ij})^2}{\partial R^2} = 2 \left(\Delta E_{ij} \cdot \frac{\partial^2 \Delta E_{ij}}{\partial R^2} + \left(\frac{\partial \Delta E_{ij}}{\partial R} \right)^2 \right). \tag{3.19}$$

The squared energy-gap Hessian can be further decomposed into a sum of symmetric dyads of the two vectors defining the branching space [342–345]:

$$\frac{\partial^2 (\Delta E_{ij})^2}{8 \partial R^2} \approx \frac{\partial \Delta E_{ij}}{2 \partial R} \otimes \frac{\partial \Delta E_{ij}}{2 \partial R} + \widetilde{C}_{\text{NAC}_{ij}} \otimes \widetilde{C}_{\text{NAC}_{ij}}. \tag{3.20}$$

Subtracting the gradient difference vectors, which can be easily obtained from ML models, the NAC vectors can be approximated from the eigenvector, v_{ij}, giving rise to the largest non-zero eigenvalue, λ_{ij}, after singular value decomposition [343–345]:

$$C_{\text{aNAC}_{ij}} = v_{ij} \cdot \frac{\sqrt{\lambda_{ij}}}{\Delta E_{ij}}. \tag{3.21}$$

Approximated NACs are valid close to conical intersections and same-symmetry crossings, which are considered to be most common in polyatomic molecular systems. [344;345] To compute NACs only in regions of PESs, in which they are valid, and to avoid any computational overhead, the SchNarc model uses thresholds between adjacent PESs that are used to form an estimate whether approximated NACs are required or not. In case PESs are further apart than this defined threshold, no approximated NACs are computed. The default values for thresholds between singlet-singlet gaps and triplet-triplet gaps are 0.5 eV and 1.0 eV, respectively.

3.6 MANIPULATION OF DATA POINTS

For many systems, the excited states need to be treated with multi-reference methods. Especially when homolytic bond breaking or bond formation needs to be described, single-reference methods can fail to provide a qualitatively correct picture of the excited states. However, the treatment of a manifold of excited states of medium-sized to large systems with multi-reference methods can become very expensive prohibiting an efficient and comprehensive training set generation. Therefore, the

aim of this section is to describe the idea of using an affordable single-reference method, such as ADC(2) or TDDFT, to compute most parts of a training set, and to use multi-reference methods to correct single-reference data in regions of the PESs, which are problematic for single-reference methods. This approach can be seen similar to a hand-tailored parametrization of empirical force fields.[346;347]

In principle, an ML model should be able to learn from a data set, regardless of how it is generated. The only prerequisite is that the data can be represented using smooth functions. In order to validate the approach of generating a training set for ML-fitted excited-state PESs with multi-reference accuracy, but single-reference costs, the molecule tyrosine serves and its dissociative reaction coordinates.Especially in the dissociation limit, single-reference methods often fail to reflect qualitatively correct physical behaviour.

The training set for tyrosine contains 16654 data points and each data point contains energies of 5 singlet states and 8 triplet states, corresponding forces, dipole moments, and SOCs computed with ADC(2).[209] The equilibrium conformation of tyrosine is taken as a reference geometry for the generation of manually manipulated data points. NACs are missing in ADC(2) codes, at least in open-source codes. Noteworthy the reference for ADC(2) is restricted MP2. Compared to unrestricted MP2 spin contamination can be avoided and SOCs between singlets and the T_1 and T_2 states can be computed more reliably. With respect to homolytic bond breaking, also unrestricted MP2 as a reference for ADC(2) would lead to qualitatively wrong results, as perturbation terms increase in the limit of dissociation.

The equilibrium geometry of tyrosine and the 5 different reaction coordinates that are selected to carry out unrelaxed scans are indicated in Fig 8. The hydrogen atoms of the hydroxy group of the phenyl-ring (1), of two C-atoms (2 and 3), of the amino group (4), and of the carboxyl group (5) are detached along the vectors they are originally pointing to, while keeping the rest of the atoms fixed. A total number of 105 data points are additionally added to the training set. Emphasis is placed on keeping this number as small as possible to reduce the additional computational overhead and manual data fitting, which can become tedious when a large number of electronic states is described. Table 1 shows the equilibrium bond distances, r_{eq}, the largest bond distances used with ADC(2) data, r_{max}, and the bond distances of the manually added artificial data points. It is important to note that these bond-distances have to be determined manually according to quantum chemical reference calculations, while the bond distances for the additional manually generated data points were set randomly. A trend regarding the optimal maximum bond lengths and the bond lengths of the first manually added data point can be observed: For r_{max} a value of 1.4Å seems to be a good starting point. Further, and unsurprisingly, similar functional groups have similar bond distances for the first manually generated data point. Further evaluations are necessary to generalize some reference bond distances for functional groups, e.g. if an ML force field for the excited states is targeted and other building blocks, such as phenol or different amino acids, are to be included.

A comparison of the potential energy curves obtained with different levels of theory, i.e., ADC(2), MS-CASPT2 (for simplicity MS-CASPT2 is referred to as CASPT2 in the following), and MP2, along the bond distance of the hydrogen located (a) at the hydroxy group of the phenyl ring (Ph-OH) and (b) at the amino group (N-H) is shown in Fig. 9 in the left panels. The energies are shown as relative energies with respect to the equilibrium distance for better comparison. As it is visible, the S_0/T_1, S_1/T_2, S_2/T_3, S_3/T_4, and S_4/T_5 states calculated with ADC(2) are not degenerate at large bond distances, which is the case according to MP2, CASPT2, and chemical
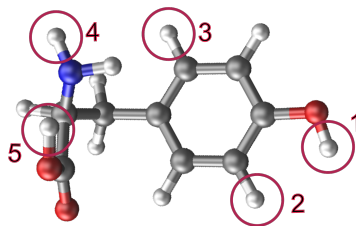
Figure 8: Equilibrium structure of tyrosine optimized at MP2 level of theory. Numbers indicate the bonds of which hydrogen atoms are detached to generate artificial data points.

intuition. Therefore, these potential energy curves are manipulated to reproduce the degeneracy at large bond distances.

The right panels illustrate the ADC(2) potential energy curves and the energies of the artificially generated data points that are used for the training set. The procedure to generate these data points is as follows: First, the energies of the S_0/T_1, S_1/T_2, S_2/T_3, S_3/T_4, and S_4/T_5 pairs of states are computed with CASPT(12,11) and the S_0/T_1 states are further computed with unrestricted MP2 to confirm their degeneracy in the dissociation limit. MP2 defines the energies of the S_0 and T_1 states. The highest triplet states, T_5, T_6, and T_7 relate to the ADC(2) values at r_{max}, which is also done for SOCs. The remaining energies are the sum of the S_0 energy of artificial points and the energy difference of CASPT2 obtained from each of the pairs of states, i.e., the difference between S_1/T_2, S_2/T_3, S_3/T_4, and S_4/T_5 CASPT2 energies and S_0/T_1 CASPT2 energies. Gram-Schmidt orthogonalization is applied to set the forces along the dissociation coordinate to zero.
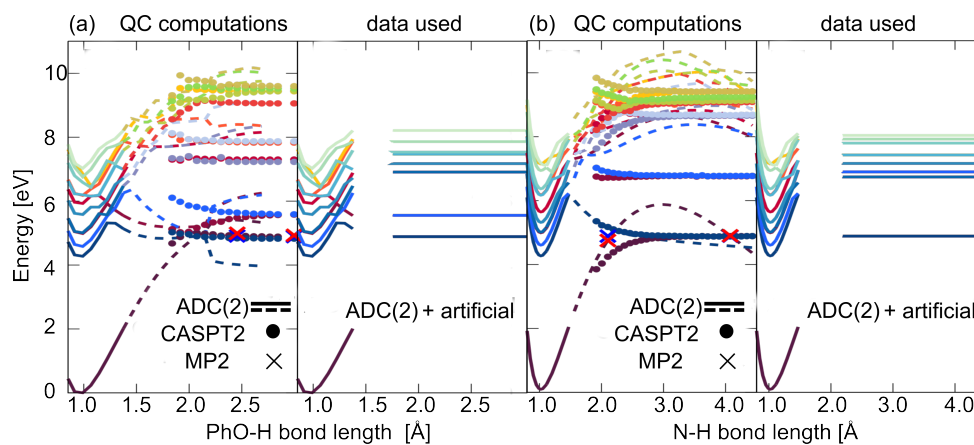
Figure 9: Potential energy curves of 5 singlets (S) and 8 triplets (T) computed with ADC(2) (solid and dashed lines), CASPT2(12,11) (circles), and MP2 (crosses) along the bond distance between hydrogen and oxygen of (a) the hydroxy group located at the phenyl ring and (b) between hydrogen and nitrogen located at the amino group. Panels on the right show the data points used for the training set, which are computed with ADC(2) for short bond distances and manipulated at large bond distances to fit the ADC(2) energies and the qualitatively correct behaviour according to CASPT2 and MP2.

Table 1: Distances between the atoms X and H (X=C,N,O) with numbers as defined in Fig. 8 of the equilibrium geometry, r_{eq}, the geometry computed with ADC(2) corresponding to the largest bond length in the training set (at r_{max}), and the artificially generated data points (# 1-21).

X-H:	1 [Å]	2 [Å]	3 [Å]	4 [Å]	5 [Å]
r_{eq}	0.97	1.10	1.10	0.99	1.02
r_{max}	1.38	1.43	1.43	1.48	1.44
# Data point					
1	1.76	3.47	3.47	2.19	1.66
2	1.96	3.57	3.57	2.37	1.70
3	2.06	3.67	3.67	2.55	1.76
4	2.16	3.77	3.77	2.73	1.84
5	2.26	3.87	3.87	2.91	1.89
6	2.35	3.97	3.97	3.09	2.06
7	2.45	4.06	4.06	3.36	2.11
8	2.55	4.56	4.56	3.45	2.37
9	2.65	5.05	5.05	3.72	2.43
10	2.75	5.55	5.55	4.07	2.49
11	2.85	6.04	6.04	4.34	13.98
12	2.95	7.03	7.03	24.88	22.64
13	3.00	8.02	8.02	28.14	26.59
14	3.50	9.01	9.01	32.10	32.72
15	3.94	10.00	10.00	38.44	44.49
16	5.92	10.99	10.99	49.52	54.58
17	10.87	20.88	20.88	60.31	62.60
18	20.77	40.66	40.66	68.72	76.16
19	40.58	80.22	80.22	75.26	79.13
20	80.19	90.11	90.11	84.06	94.77
21	100.00	100.00	100.00	99.90	97.05

4

RESULTS AND DISCUSSION

This chapter is dedicated to assess the performance of the developed methods and strategies presented in chapter 3. The methylenimmonium cation, $CH_2NH_2^+$, is given in the upper left of Fig. 10 and serves as a test system that accompanies the main developments. In a first attempt, a training set for this molecule is generated using the extended adaptive sampling scheme for excited states and the phase correction algorithm. Simple MLFF NNs and the inverse distance matrix as a descriptor demonstrate that ML can be used for photodynamics simulations by comparison to reference photodynamics simulations and to reference PESs in critical regions. The benefits that ML models offer to advance this research field are shown with efficient computations of thousands of independent trajectories and with long time scale excited-state MD simulations in the ns regime. This proof of concept is followed by a comparison of different types of regressors and descriptors, i.e., NNs and KRR with focus on the FCHL representation and the inverse distance matrix. Single-state and single-property ML models are compared to their multi-state and multi-property counterparts.

The amino acid tyrosine (bottom right corner in Fig. 10) is investigated as a representative for molecules comprising a complex photochemistry, whose description requires highly accurate, but mostly impracticable quantum chemical methods. This molecule highlights the challenges ML models need to overcome in order to allow for an automatized way of improving existing methods with ML models. The possibility to fuse two data sets to combine the efficiency of affordable single-reference methods with the accuracy of multi-reference methods where single-reference methods are qualitatively inaccurate is evaluated using MLFF NNs.

Some obstacles are identified and solved with the developed SchNarc approach for photodynamics simulations that profits from the end-to-end architecture of the underlying deep-learning model SchNet and incorporates a phase-free training algorithm to train data as directly obtained from quantum chemistry. The phase-free training algorithm is tested along with the accuracy of ML-fitted electronic states of different spin-multiplicities, ML-fitted SOCs, and approximated NACs using the molecules sulfur dioxide, SO_2 (right upper corner in Fig. 10), thioformaldehyde, CSH_2 (bottom left corner in Fig. 10), and $CH_2NH_2^+$. The respective studies are reprinted
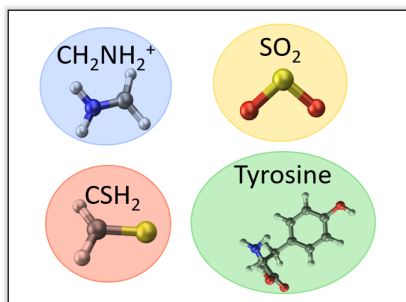

Figure 10: Molecules used in this thesis to evaluate the developed methods and strategies.

in the appendix in sections A.1 to A.3. A contribution to the book titled "Machine Learning in Chemistry: The Impact of Artificial Intelligence" edited by Hugh M.

Cartwright is enclosed in the appendix in section A.4. An accepted perspective on ML for excited state dynamics is enclosed in the appendix in sections A.5.

4.1 PROOF OF CONCEPT: THE EXCITED-STATE DYNAMICS OF CH$_2$NH$_2^+$

The photodynamics of the methylenimmonium cation is used to prove that MLMD simulations in the excited states are possible. This molecule is particularly well suited for this purpose, because its excited-state dynamics happens fast enough and this molecule is small enough, so that highly accurate quantum chemical reference calculations can be carried out to validate the developed ML approach. High accuracy is needed to accurately describe the ultrafast transitions taking place in CH$_2$NH$_2^+$. Consequently, the dynamics are challenging to reproduce also with ML and require precisely fitted PESs.

4.1.1 *Training Set Generation*

The quality of the training set defines the quality of the ML model. Therefore, the development of new methods is often accompanied with the computation of a training set, which is also necessary for the scope of this thesis. For the sake of enhancing MQC simulations with the fewest switches surface hopping algorithm, a wish list is identified for the training set, which is similar to the one for training sets for ground-state MLMD simulations:[106;108;153] (i) The training set should be free of errors, i.e., it should be computed with an accurate electronic structure method that can describe the electronically excited states of the system under investigation and that does not lead to artifacts within the PESs, which is, more often than not, a problem on its own. (ii) The training set should be sampled comprehensively in order to enable an ML model to interpolate the relevant conformational regions for the subsequent excited-state MLMD simulations. (iii) In addition, it should contain as few data points as possible in order to minimize the costs related to the reference simulations and training of ML models, and to remain feasible (e.g., such that the kernel matrix fits into the computer's memory).

The ability of the extended adaptive sampling scheme for the excited states of molecules to meet these requirements is evaluated using the methylenimmonium cation. To properly describe the photochemistry of this system, a highly precise multi-reference quantum chemical method is needed, as homolytic bond breaking can possibly happen after light excitation.[64] Therefore, MR-CISD is selected as a reference method with the aug-cc-pVDZ basis set and an active space of six active electrons in four active orbitals, which is chosen in accordance with literature.[64] The program COLUMBUS[201] is used for these reference computations. Each data point comprises energies and forces of three singlet states, permanent and transition dipole moments, and NACs between each pair of states. Moreover, every data point inside of the training set is phase corrected.

In order to start adaptive sampling, an initial training set has to be provided, which is based on the equilibrium structure of CH$_2$NH$_2^+$ and sampled along its 12 normal modes. Each scan along a normal mode contributes with 100 data points to the initial training set and additionally, 70 data points are sampled by rotation of the molecule along its dihedral angle. 992 data points are then carefully selected from the initially computed 1,270 data points. Emphasis is placed on removing data points that show artifacts in the potential energy curves due to changing active spaces, intruder

states, or not converged reference calculations. This selection is done to guarantee that the development of an ML model capable to describe a manifold of electronic states is not biased by erroneous or inconsistent data points. An example of artifacts within quantum chemical simulations for the excited states is shown in panel (a) of Fig. 11. As it is visible, at the geometry of the avoided crossing between the ground state, S_0, and the first excited singlet state, S_1, a jump in the potential energy of the second excited singlet state, S_2, occurs. Unfortunately, such inconsistencies are quite common.
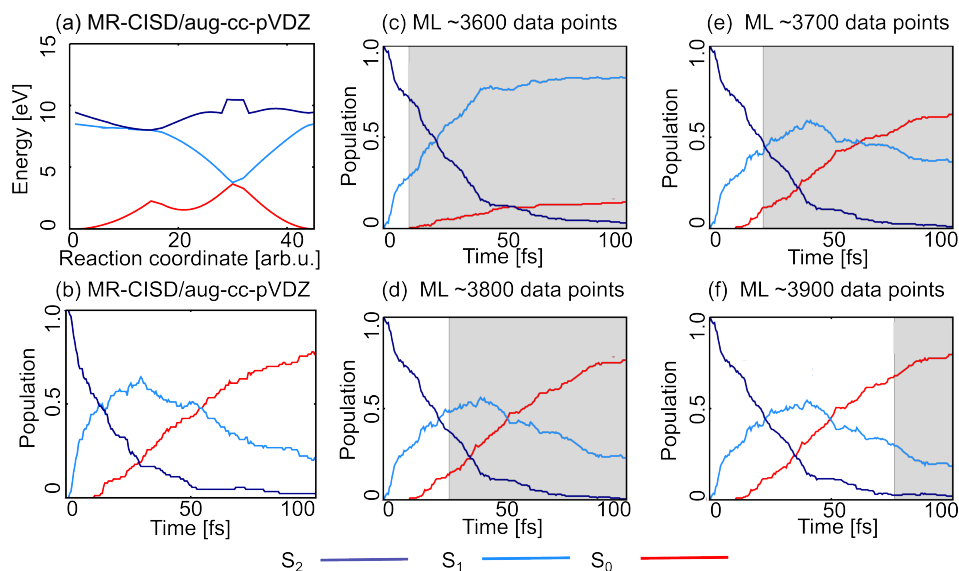
Figure 11: MR-CISD/aug-cc-pVDZ (a) potential energy curves along a reaction coordinate that incorporates two avoided state crossings, showing an artifact in the energy of the second excited singlet state, S_2, and (b) population curves obtained from 90 averaged trajectories initiated in the second excited singlet state, S_2. (c)-(f) The population curves obtained with 100 averaged trajectories with ML models trained on an increasing amount of data points during the adaptive sampling scheme show the convergence of the ML PESs to reproduce the reference dynamics.

The so-obtained 992 data points are used to initiate the adaptive sampling scheme for excited states using the MD program SHARC,[72;166;213] which uses NACs to estimate the hopping probabilities between different electronic states. To this aim, two MLFF NNs are trained on the energies, forces, and NACs using the initial data set and slightly different hyperparameters for each NN. The adaptive sampling scheme then starts with an ensemble of trajectories. These trajectories start from molecular geometries obtained after Wigner sampling[236] that are excited to the S_2 state. After 4,000 data points, the PESs are considered to be converged with respect to reference molecular dynamics simulations given in panel (b) of Fig. 11. In this plot, the population of each singlet state is averaged over 90 trajectories and plotted against the simulation time. It further illustrates that within around 100 fs, most of the population is transferred back to the ground state. The same behaviour can be obtained with ML models trained on around 3,800 data points. The convergence of the ML PESs with increasing number of data points is visible in panels (c)-(f) of Fig. 11. As can be seen, the PESs converge very fast at the end of the adaptive sampling procedure and not much difference can be observed between dynamics obtained from ML models trained on 3,800 data points (panel (e)) and ML models trained on 3,900 data points (panel (f)). Therefore, after 4,000 data points, the adaptive sampling scheme

is considered to be converged. An in depth discussion on the adaptive sampling scheme and the network architecture of the NNs can be found in Ref. 77 reprinted in section A.1.

The final training set of $CH_2NH_2^+$ contains 4,000 data points. With respect to the 12 degrees of freedom of the molecule, about 333 data points are included per degree of freedom. Compared to other recently reported training sets for MQC photodynamics simulations with ML,[154–156] which easily comprise several tens of thousands of data points, this training set is relatively small. Common training set generation approaches are discussed in a perspective, which is reprinted in section A.5.

4.1.2 *Accuracy of NNs*

With the optimal data set that contains 4,000 data points the final NN models are trained. Six hidden layers and 50 nodes per hidden layer are used for each property. The training set is randomly split into 90% training data and 10% validation data to employ an early stopping mechanism to prevent from overfitting. An additional test set of 770 phase corrected data points is computed from linear combinations of normal modes to assess the accuracy of ML-fitted PESs, forces, and NACs. In addition to NNs, SVR and LR (as a baseline model) are used. The MAEs for energies, forces, and NACs on the test set are reported for all models in Table 2, which is reproduced partly from Ref. 77. As expected, LR is less accurate than NN, SVR, and KRR. SVR and KRR are used in their standard implementation, i.e., in a single-state version using scikit-learn[348] and QML,[339] respectively. While the MAE for the energies obtained with an NN is about a factor of 39 lower compared to LR, the MAE of NACs can only be reduced by a factor of about 1.2. This indicates that NACs might profit from better descriptors. However, the accuracy obtained with more complex descriptors, as summarized in Table S4 in Ref. 77, can only be improved slightly with more accurate descriptors, hence the NNs with the inverse distance matrix are considered to be accurate enough for the purpose of this study.

Table 2: Mean absolute errors (MAEs) on energies, gradients, and NACs, averaged over three singlet states and pairs of states, respectively, and reported for LR, SVR, KRR, and NN models using the inverse distance matrix as a molecular representation. (The table is adapted from Ref. 77 under CC-BY.)

Model	MAE Energy [H]	MAE Gradients [H/Bohr]	MAE NACs [a.u.]
NN	0.00237	0.00669	0.328
SVR	0.00618	0.01169	0.382
KRR	0.01650	0.04040	0.401
LR	0.09240	0.13902	0.471

The PESs are further evaluated with two-dimensional potential energy scans around the minimum energy conical intersections obtained with the reference method. It is important to describe critical regions of the PESs accurately with ML, because transitions between states of same spin multiplicity are highly probable around such conformational regions. The plots are given in Fig. 12 (adapted from Ref. 77 under CC-BY) and demonstrate the accuracy of NN fitted PESs. The crossing surfaces between the S_2 and S_1 states are given in the upper panels, (a) and (c), and between the S_1 and S_0 states in the lower panels, (b) and (d), with NNs in the right (panels (c) and (d)) and quantum chemistry in the left (panels (a) and (b)). Importantly, the NNs can reproduce the the conical intersection seams and their curved shape
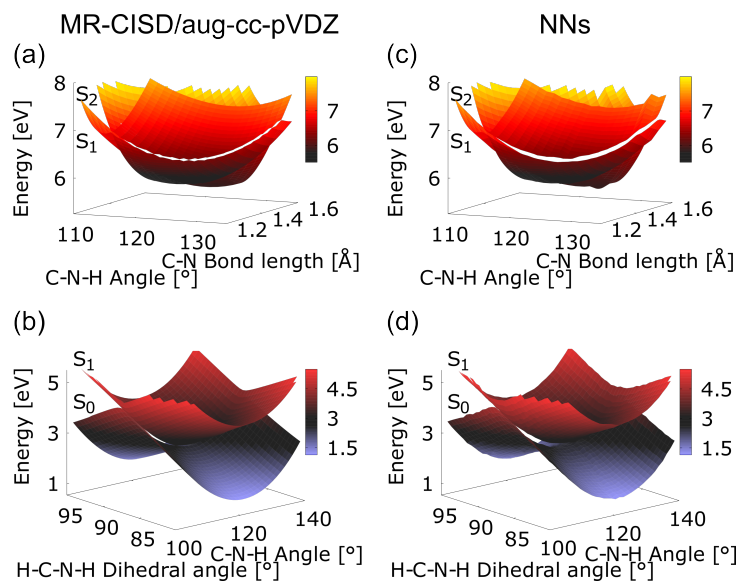
Figure 12: Two-dimensional potential energy scans around minimum energy conical intersections optimized at MR-CISD/aug-cc-pVDZ level of theory computed with (a,b) MR-CISD/aug-cc-pVDZ and (c,d) NNs showing the S_2 and S_1 (upper panels) as well as S_1 and S_0 (lower panels) states. (Adapted from Ref. 77 under CC-BY.)

qualitatively correctly. The energy gaps between the approaching electronic states are slightly larger with NNs than with quantum chemistry. This overestimation of the energy gaps is possibly due to the fact that NN PESs are smooth functions by definition and need to be differentiable at any point in the conformational space of a molecule. Hence, a cone cannot be reproduced exactly, which is however, the theoretically correct shape of a conical intersection. This effect is discussed comprehensively in the enclosed reprint in section A.1. In addition, the accuracy of the fitted PESs and NACs is discussed along with optimized minimum energy conical intersections from NN PESs. The molecular geometries obtained from NNs are similar to those calculated with the reference method.

4.1.3 *Excited-State Dynamics with NNs*

Nevertheless, the ultimate test to prove the accuracy of the NN fitted PESs, forces, and NACs, is a photodynamics simulation. NN PESs should accurately reproduce the results of the reference method and enable the simulation of many more trajectories, while still maintaining computational efficiency. The envisioned goal is to go beyond the time scales of what is feasible with the reference method. The results of the envisioned simulations with NNs are shown in Fig. 13 (adapted from Ref. 349) and is part of the appendix in sections A.1 and A.4.

The results of surface hopping MD simulations with the reference method MR-CISD/aug-cc-pVDZ (abbreviated QC1 and illustrated with dashed lines) are given in panel (a) of Fig. 13. The photodynamics are computed additionally with the same electronic structure method, but with another double-zeta basis set, namely MR-CISD/6-31++G** abbreviated QC2 in panel (a) and illustrated with dotted lines, to form an estimate of what can be deemed similar dynamics. The population plots of photodynamics with 2 NNs are illustrated in panel (b) using solid lines. As it is visible, the NNs can accurately reproduce the ultrafast transitions from the S_2 state to
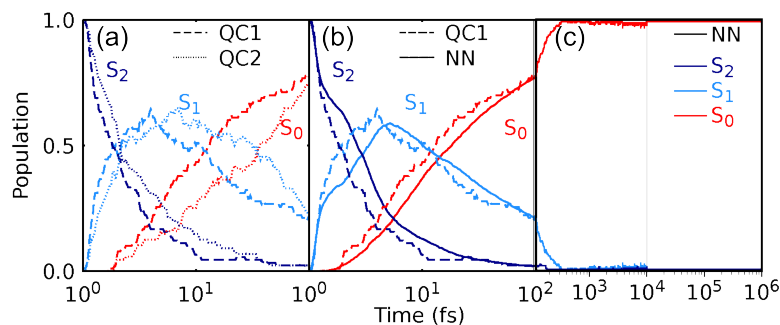
Figure 13: Population curves obtained from dynamics simulations of CH$_2$NH$_2^+$ excited to the second excited singlet state, S$_2$, using (a) 90 trajectories based on MR-CISD(6,4)/aug-cc-pVDZ (QC1) and 88 trajectories based on MR-CISD(6,4)/6-31++G** (QC2), which are compared to results obtained from (b) 3864 trajectories based on NNs, whereof 200 trajectories are continued until 10 ps and (c) 2 trajectories are continued until 1 ns. (Adapted from Ref. 349 with permission from the Royal Society of Chemistry.)

the S$_1$ state and back to the S$_0$ state. The hopping geometries obtained with the three different methods, which lead to transitions during the dynamics simulations, are discussed in-depth in the published article enclosed in the reprint in section A.1, and are similar to each other. With respect to the similarity of QC1 and QC2 population curves and the limited number of trajectories that are computed with these methods, the dynamics of NNs are deemed to be similar to QC1 dynamics and qualitatively correct. These results also demonstrate the validity of the phase correction algorithm. Only with accurate NACs in addition to accurate PESs and forces, dynamics can be reproduced.

Besides the fact that NNs allow for an accurate description of the temporal evolution of CH$_2$NH$_2^+$ in the excited states, the main benefit is their computational efficiency. The fast evaluation of PESs, forces, and NACs enables a computation of thousands of independent trajectories, which is much more than what is typically feasible with quantum chemistry methods. Indeed, the computation of 3864 trajectories for 100 fs with NNs requires about 10 times less central processing unit (CPU) hours than 90 trajectories with the reference method require. The population curves averaged over 3864 trajectories are illustrated in panel (b) for NNs, which are characterized by much smoother population curves.

In addition to the enlarged statistics, longer simulation times can be reached. Panel (c) shows populations curves that are averaged over 200 trajectories until 10 ps and over two trajectories until 1 ns. The trajectories up to 10 ps can be simulated within 6 hours on one core, which is 300 times faster than the respective simulations with the QC1 method on the same core. As can be obtained from the population curves, the population is almost completely transferred to the ground state after about 300 fs. Due to the remaining kinetic energy of the molecule, some transitions still take place, which are visible as noise. For demonstration purposes, only two trajectories are computed up to 1 ns.

Similar to the adaptive sampling scheme, the dynamics are achieved with 2 NNs, whose mean is given to the MD program. The variance of the predictions of the NNs for PESs, forces, and NACs are compared each time step in order to guarantee an interpolative regime and accurate NN predictions. Whenever a conformational region of the PESs is reached that is not or only sparsely sampled in the training set, quantum chemical calculations are carried out to adapt the training set. The NNs are then re-trained and photodynamics simulations are re-started from the point at

which the simulations have been interrupted. The calculation of 1 ns with two NNs considering the additional sampling and re-training steps takes approximately 59 days on one core, whereas the quantum chemical reference dynamics would require computations of estimated 19 years on the same core. Noticeably, due to the fast inferences of NNs, the time limiting step of MD simulations is the file I/O required for the communication of the MD driver with the NNs (or in case of conventional MD simulations the electronic structure program). The molecular geometry of the next time step is written to an ascii-file by the MD program and read in by the NNs, and the same is done for the predictions of the NNs, which are written to an ascii-file and read in by the MD program. This file I/O usually does not affect conventional MQC dynamics due to the high computational effort of quantum chemical calculations. With the conventional SHARC MD program, the simulation of 1 ns with NN PESs would require at least one year of computation. A python wrapper termed pySHARC directly interfaces the NNs with the SHARC dynamics driver, omitting the aforementioned file I/O, and thus enables long time scale photodynamics. This python wrapper is also used for the other implemented ML models in this thesis, namely KRR and SchNet.

4.2 comparison of different regressors and descriptors

The MLFF NN models applied here are most probably the simplest NN models that one can use for photodynamics simulations and the use of the inverse distance matrix might be comparable to a minimal basis in a quantum chemical calculation. Comparable to a quantum chemical calculation, a more precise method (a deeper network architecture in the case of ML) can only compensate for a minimal basis (a minimal descriptor) to a certain extent. The matrix of inverse distances in combination with a deep learning model is accurate enough to allow for a description of a manifold of excited states, energy conservation during dynamics simulations, and an accurate computation of hopping probabilities as well as minimum energy conical intersections. In light of entering the research field of photochemistry with ML and carrying out a proof of concept study, this architecture is well suited. Nevertheless, when a larger molecular system with a more complex photochemistry is investigated, these NN models most probably do not meet the requirements of efficient and accurate photodynamics simulations anymore. Especially the matrix of inverse distances, which is used as a molecular representation, remains problematic, as every data point in the training set needs to be ordered accordingly. Interchanging two hydrogen atoms, e.g., when rotating $CH_2NH_2^+$ along its dihedral angle by 180°, should not change the energy predicted by the networks. However, the energy will differ and permutation invariance is not given, unless the relevant conformations are explicitly encountered in the training set. The computational effort related to the training set generation of a molecule with more atoms that can easily comprise 50-100s of degrees of freedom can potentially become a limiting factor, not even thinking about the inclusion of all possible permutations that would then lead to long training times and huge memory requirements. Furthermore, the MAE of NACs shows that the accuracy achieved with ML models is far from being exhausted, as the MAE is only slightly lower than the one obtained with linear regression.

Due to these limitations and the lack of studies that focus on a description of the excited state PESs, forces, couplings, and dipole moments with ML, a comparison of different types of regressors and molecular representation follows the proof of

concept that ML can advance photodynamics simulations. Yet, it is not known how to best treat a manifold of excited states and the properties between them. Therefore, the following questions are addressed here: (1) Is it advantageous to describe all excited-state PESs within one ML model or can more accurate PESs be obtained with a single-state treatment? (2) How can the accuracy of NACs be improved and what effects does a joint learning on the quality of different properties have?

An additional motivation behind this comparison is that two[154;155] out of three very recent studies that apply ML models for advancing MQC dynamics in the excited states, employ KRR models. However, during dynamics simulations it is required to resort to quantum chemical PESs and couplings in critical regions. The remaining third study,[156] i.e., some kind of a follow-up with one of the authors from a previous KRR study, uses NNs and can replace quantum chemistry completely during photodynamics simulations. This has been achieved for two singlet states of the molecule CH_2NH with the Zhu-Nakamura hopping probabilities, which allow to omit any computation of couplings. These findings leave the questions why some models fail for some excited-state properties open. What are the benefits of different types of regressors and can these benefits be combined similar to a *wide and deep learning*[350] concept?

To address these questions, the training set computed for the methylenimmonium cation comprising the energies and forces of three singlet states as well as NACs and dipole moments between them is used. MLFF NNs and KRR with the QML toolkit[339] are applied and compared to each other. The latter is already implemented with the FCHL representation and allows to treat forces as response properties, which is important for dynamics simulations. The models are evaluated with excited-state MLMD simulations, potential energy scans, and learning curves.

4.2.1 *Photodynamics Simulations and Excited-State Properties of $CH_2NH_2^+$*

In a first attempt, KRR models are trained with the FCHL representation using the standard implementation, i.e., each energy and NAC value is trained with a separate KRR model in a single-state fashion and forces are treated as response properties. Additionally, all properties are treated together in one NN model and forces are trained as derivatives of ML potentials for energies, which is termed multi-property model. The inverse distance matrix is used as a representation for NNs to allow for better comparison to single-property NNs. The optimal hyperparameters for the ML models are obtained using 5-fold cross validation. The learning curves, which show the training set size plotted against the MAE on a separate test set, are computed as averages from 10 independent calculations, providing a measure of uncertainty. An in depth discussion on the ML models and the respective computations can be found in the reprint of the published article enclosed in section A.2.

The learning curves for energies and gradients are illustrated in Fig. 14(a). As it is visible, the MAE obtained with NNs (red curves) and KRR (blue curves) models decreases with increasing training set size, which is the desired behaviour, which indicates that the models can learn from the training set. The NNs with the inverse distance matrix outperform KRR models in their accuracy for energies and gradients, even though the FCHL representation incorporates more information on the molecule than the inverse distance matrix. To evaluate the performance of NNs with the FCHL representation, additional learning curves are computed that show only slight improvements in accuracy and can be found along with an exhaustive comparison
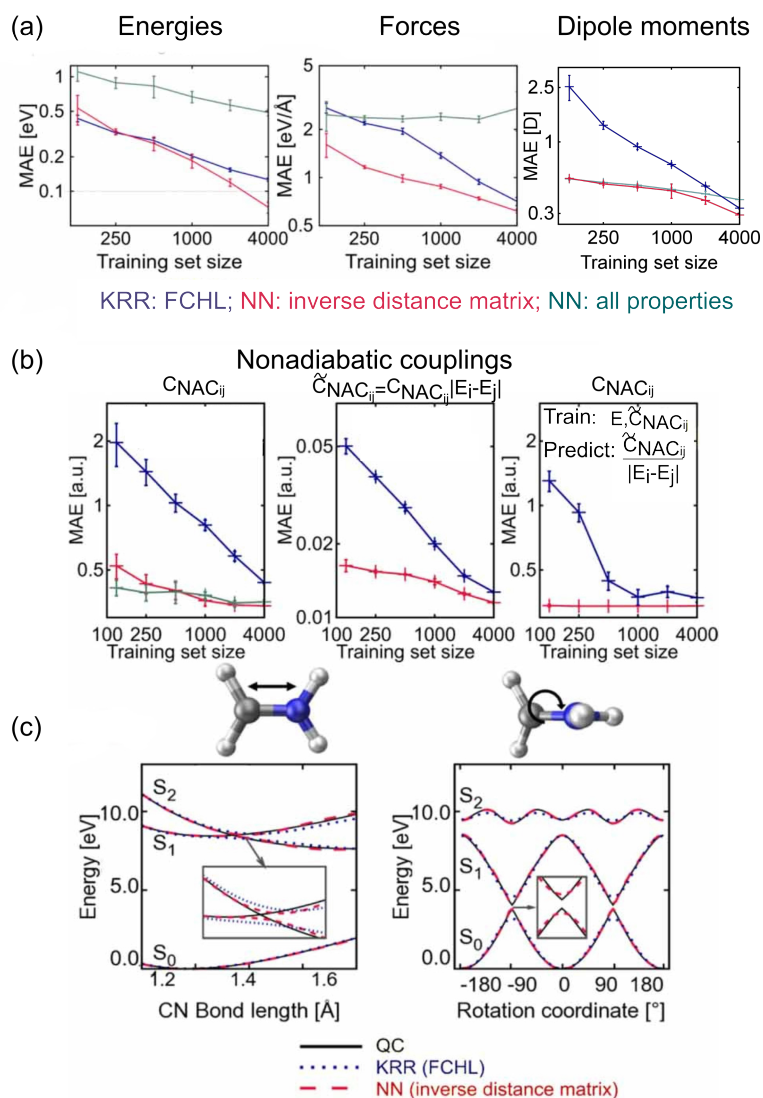
Figure 14: (a) Learning curves of energies, forces, and dipole moments obtained with KRR using FCHL (blue lines) and single-property and multi-property NNs using the inverse distance matrix (red and green lines, respectively) are compared to each other. (b) NACs are predicted as directly obtained from quantum chemistry after phase correction (left panel) and as smooth properties (middle and right panels). (c) The potential energy curves along the C-N bond and the dihedral angle of $CH_2NH_2^+$ further illustrate the accuracy of ML PESs. (Adapted from Ref. 157 under CC-BY.)

of different models and descriptors in the related publication 158 in the supporting information. Regarding the NN model that describes all properties at once (green curves), less accurate energies and gradients are received. In case of gradients, the slope of the learning curve is even positive, revealing an anti-learning behaviour for forces. However, the MAE decreases when all properties, i.e., the whole predicted vector of the multi-property model containing energies, forces, dipole moments, and NACs, is considered. Nevertheless, the results indicate that these models are not sufficient for excited-state MD simulations.

The learning curves for dipole moments are shown in the rightmost panel, showing the same result with respect to the MAE at largest training set size. Remarkably, the learning efficiency, deduced from the slope of the learning curves, is largest for KRR.

This observation indicates that KRR might lead to higher accurate dipole moments for larger training set sizes, possibly due to the relation of the depth of the model and the training set size. A larger network architecture used for NNs does not yield more accurate results and the network architecture is supposed to be converged with six hidden layers.

The learning curves of NACs are illustrated in panel (b), learned as directly obtained from quantum chemistry after phase correction in the leftmost panel. The ML models exhibit a similar learning behaviour as for dipole moments, but the single-property and multi-property NN models are comparable to each other and a joint learning of properties is neither beneficial, nor disadvantageous with respect to NACs. Additionally, it is evaluated, whether learning and predicting smooth NACs, $\widetilde{C}_{NAC_{ij}}$, i.e., their values multiplied with the corresponding energy gap as defined in equation 3.17, leads to an improved training. The learning curves of smooth couplings are illustrated in the middle plot of panel (b). These smooth coupling values are then divided by the energy gap obtained from ML models, shown in the rightmost panel. With regards to KRR, the MAE obtained for a training set size of 4,000 data points is lower, but the learning curve is not linear, indicating that the PESs are not accurate enough to improve the accuracy of NACs considerably. The NN models are comparable to each other and the accuracy obtained with a training set size of 4,000 is comparable to the accuracy of NN models that learn the non-smooth NAC values. This observation raises two assumptions: First, in order to improve from an inclusion of energy gaps into the fitting procedure of NACs, highly accurate ML-fitted PESs are required. Secondly, the NACs cannot simply be predicted as single values, but their rotational covariance needs to be incorporated into the ML model. This could be done for example with the response formalism for KRR as shortly mentioned in chapter 2.21. Neither a single-state, nor a multi-state or multi-property treatment can further increase the accuracy of NACs.

In order to estimate the quality of ML PESs in critical regions, scans along the C-N bond and the dihedral angle of $CH_2NH_2^+$ are carried out with quantum chemistry (QC), NNs, and KRR. The multi-property NN models are not sufficiently accurate for fitting PESs, hence their ability to predict energies in critical regions is not further assessed here. An in-depth discussion can be found in the enclosed reprint A.2. As can be seen in panel (c), the potential energy curves can be fitted accurately and only slight differences are visible between NN PESs and KRR PESs close to avoided crossings between different electronic states. The KRR model is therefore supposed to be accurate enough for photodynamics simulations and evaluated for its applicability in surface hopping MD simulations using NACs from KRR and NNs.

Fig. 15(a) shows the population curves obtained from KRR models for $CH_2NH_2^+$. As it is clearly visible, compared to the reference dynamics and the NN dynamics given in Fig. 13, KRR models are not accurate enough to replace quantum chemical calculations in photodynamics simulations. As they fail to reproduce dynamics completely, the influence of the PESs and NACs is evaluated separately. To this aim, mixed models are used for photodynamics simulations. Panel (b) shows population curves obtained from NNs that predict energies and forces and KRR that predict NACs. As it is visible, the dynamics is similar to the reference dynamics, but happens on a shorter time scale. The photodynamics obtained with NACs from NNs and energies and forces from KRR does not agree to the reference dynamics at all. This observation indicates that energies and forces have a greater influence on the photodynamics than NACs have, that are indeed only used to measure the probability of stochastically determined hops between different states. The results obtained from photodynamics

simulations indicate that the small differences in the PESs of KRR models and NNs lead to such wrong dynamics, even when the same model for NACs is used. A more detailed discussion can be found in chapter A.2.
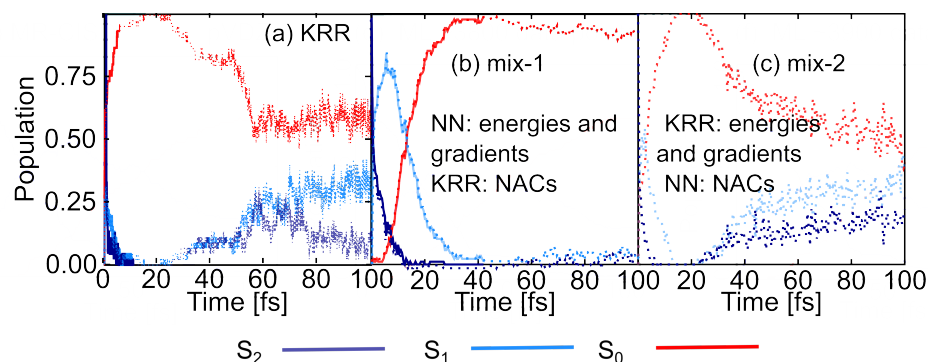
Figure 15: Population curves of $CH_2NH_2^+$ obtained from averaging over 200 initially excited structures to the S_2 state computed with (a) KRR models using the FCHL representation, (b) mixed ML models, where KRR is used to predict NACs and NNs are used to predict energies and gradients, and (c) mixed ML models with KRR energies and forces and NN NACs. Dotted lines refer to population curves that are deemed to be wrong with respect to reference dynamics. (Adapted from Ref. 158 under CC-BY.)

4.2.2 *Multi-State Representation*

Since NNs describe energies, forces, and NACs in a multi-state fashion and KRR models in a single-state fashion, it is sought to investigate the effect of a multi-state treatment for KRR. To this aim, the developed multi-state representation as introduced in chapter 3.12 is used and the performance of respective models to predict energies of $CH_2NH_2^+$ is evaluated. Due to the implementation, the forces are not described as response properties. Also for NNs the electronic state is explicitly encoded in the representation of molecules to allow for a better comparison and assessment of the effect of an enlarged descriptor on the performance of NNs. Therefore, the matrix of inverse distances is multiplied N_S times with a corresponding state number.

The results for the multi-state models are given in Fig. 16. Panel (a) shows the learning curves and corresponding scatter plots for energies obtained with the NN and KRR models applying their standard representation and the state representation. As it is clearly visible, the energies obtained from KRR models increase significantly in their accuracy, while the effect of a state representation for NNs is minor. This observation indicates that the NNs are already converged to their optimal results and the multi-state treatment is sufficient enough using the standard representation, while KRR models are not converged to their optimal result. The potential energy curves compare the standard representation (with forces not described as response properties for KRR) in panel (b) with the state representation in panel (c). Also here, the potential energy curves become more accurate when a state representation is used. The improved accuracy of KRR models can be a result of the larger size of the kernel matrix, i.e., the depth of the KRR model, which increases from $N_M \times N_M$ for the standard representation to $N_M \cdot N_S \times N_M \cdot N_S$ with the state kernel. Nevertheless, KRR models are still slightly less accurate than NNs in critical regions of the PESs.

KRR models could be further improved by including forces as response properties.[132] An improved accuracy for energies is already visible for the standard
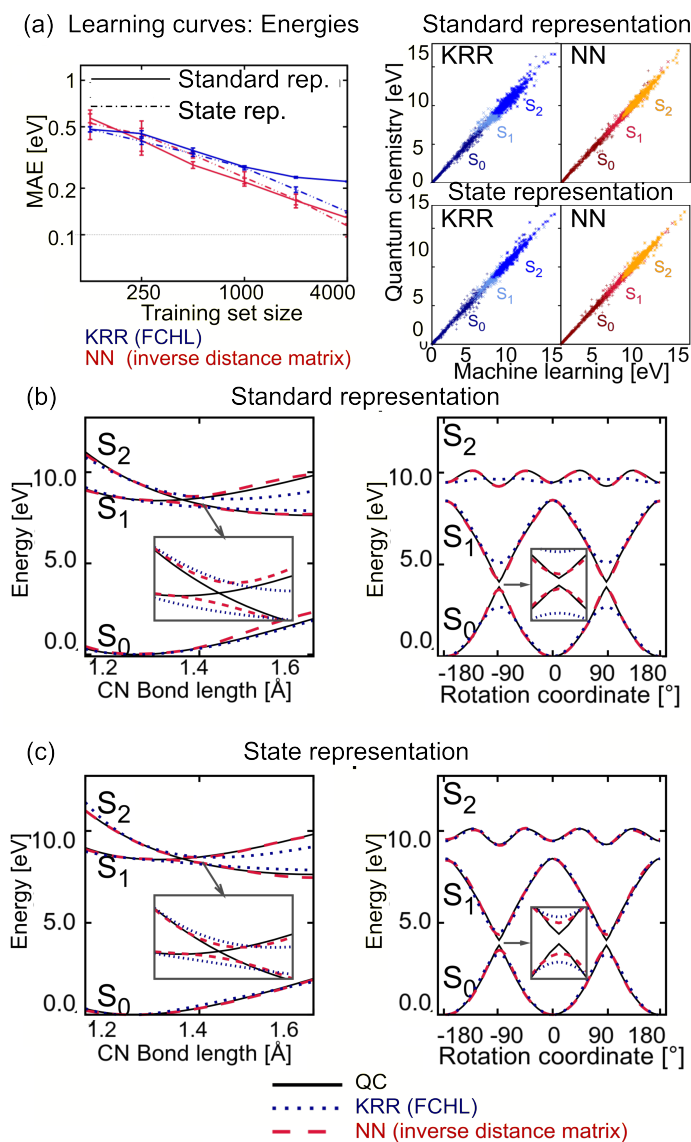
Figure 16: (a) Learning curves for energies computed with KRR (blue) and NNs (red) to compare the standard (solid lines) and state representation (dashed lines). The right panels shows scatter plots for energies using ML models trained on 4,000 data points. Corresponding potential energy curves of $CH_2NH_2^+$ along its C-N bond (left panels) and dihedral angle (right panels) are illustrated in panels (b) and (c) to further compare the standard and the state representation, respectively. (Adapted from Ref. 158 under CC-BY.)

representation, compare Fig. 16(b), in which forces are accounted for, to Fig. 14(c), in which forces are not accounted for with KRR models. This effect is also envisioned when combining the state representation with the kernel that can account for energies and forces in one model. However, the kernel matrix size increases to $3N_AN_M^2N_S \times 3N_AN_M^2N_S$ leading to large memory requirements with increasing training set sizes. While for $CH_2NH_2^+$ about 126 GB are occupied using 4,000 data points and 3 electronic states for training, larger systems would require even more memory. For example, for tyrosine with 24 atoms, 22,000 data points (using a comparable amount of data points per degree of freedom), and 13 states (considered to be relevant, as will be discussed in chapter 4.4), more than 1 million GB are estimated, which is far beyond the usual memory consumption for training. Even when only

a subset of molecules is mapped to the complete training set, the training process remains extremely expensive. The training of $CH_2NH_2^+$ using 4,000 data points required two weeks of computation on an 8x Intel Xeon CPU (E7-8867). The high costs related to the training of KRR models therefore prohibits an efficient application of hyperparameter optimization procedures, such as cross-validation or random grid search, to obtain the optimal width for the state kernel and the Gaussian kernel for molecules. The associated computational efforts also hamper an evaluation of NACs as response properties, as the memory consumption would increase even more due to a pair-wise state representation, which would be required for that purpose.

The comparison of the different ML models indicates that, in their current implementations, NNs are more suitable for advancing photodynamics simulations. A multi-state treatment of properties is favourable for NNs, but encoding of the energy level in the representation cannot improve results significantly. Nevertheless, it cannot be said that one model outperforms the other, as KRR models exhibit much steeper learning curves for excited-state properties. Coupling values are therefore supposed to profit from a multi-state treatment. Further, KRR models benefit from their simplicity and the fact that only a few hyperparameters need to be optimized compared to NN models. At the current stage of research, the huge memory requirements, however, restrict their use to single-state models and also prohibits the use of multi-state models in photodynamics simulations.

4.3 ML DESCRIPTIONS OF NACS

This chapter is dedicated to improving the accuracy of NACs by describing them in a rotationally covariant manner and providing an ML based approximation to them by using the ML-fitted PESs, their Jacobians, and Hessians.

4.3.1 *NACs as Virtual Derivatives*

In order to describe NACs as vectorial properties that are covariant with respect to the rotation of a molecule, the SchNet model is used with the adaptation for excited states. As described in section 3.5, NACs are treated as virtual derivatives of ML PESs. These virtual properties are initiated randomly in the beginning of a training and are optimized so that their derivatives with respect to Cartesian coordinates of a molecule match the NACs of the reference method. The concept of learning and predicting smooth couplings as evaluated in the previous chapter is used. The predicted smooth couplings are subsequently divided by the corresponding energy gaps provided by SchNet. As SchNet for excited states describes every property in a different prediction block, but connects them by combining their losses in one function, accurate predictions of all properties are enabled and a fast communication between the independently described properties is possible. This architecture, which is inspired by the underlying physics of NACs, is supposed to result in higher accurate NACs.

The model is tested using the training set of $CH_2NH_2^+$. The lowest MAE obtained for NACs with this approach is 0.15 a.u., which is a factor of about 3.2, 2.7, and 2.2 lower than the MAEs obtained with LR, KRR, and NNs, respectively. Noticeably, this MAE is obtained by using only 3,000 data points for training, while the previous models require 4,000 data points. The accuracy of this approach is further demonstrated by photodynamics simulations obtained from models with the combined L_2 loss function

and 3,000 trained phase corrected data points. The population curves are shown in Fig. 17 and ML models can accurately reproduce the dynamics of the reference method in the left panel. More details on the performance of NACs can be found in section A.3.

With regards to computational costs, the photodynamics simulation with the SchNarc approach is about 15 times faster than the respective MD simulations with the MLFF NNs. SchNarc further allows for training and inference on graphic processing units (GPUs). However, the training of a SchNet model for energies, forces, and NACs of 3,000 data points takes about 20 hours on a GPU, whereas the MLFF NNs require approximately 4 hours on a CPU. Due to the different benefits of the different ML models, MLFF NNs are more suitable for training set generation, whereas the SchNet architecture is better suited for production runs.
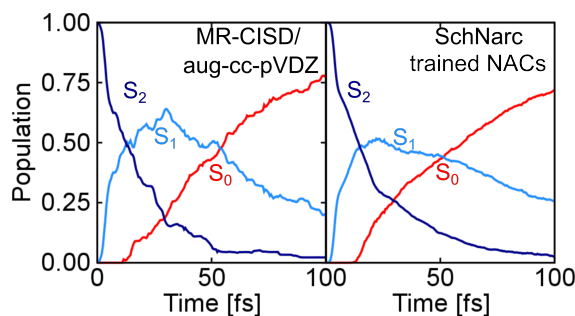
Figure 17: Population curves obtained with MR-CISD/aug-cc-pVDZ for $CH_2NH_2^+$ for 100 fs on the left and respective photodynamics using SchNarc trained on MR-CISD/aug-cc-pVDZ phase-corrected data and trained NACs on the right. (Adapted from Ref. 157 under CC-BY.)

4.3.2 *ML Approximation to NACs*

Another problematic aspect of NACs is their peaky nature, which makes them generally difficult to fit with ML models and also challenging to compute accurately with quantum chemistry. Furthermore, NAC calculations remain one of the most expensive parts of a quantum chemical calculation. The costs of a NAC calculation is comparable to the costs of a force calculation, but the number of coupling vectors is $N_S(N_S - 1)/2$ compared to the N_S required force calculations. During excited-state MD simulations, the number of computations of NAC vectors can be reduced by considering only NACs between electronic states that lie close to the active state. However, all NAC vectors need to be calculated for a training set for ML models.

Another important factor to consider when surface hopping dynamics are to be described with ML is that NACs are often missing in quantum chemistry codes. Consequently, a method that cannot account for NACs, cannot be used to generate a training set for MLMD simulations with Tully's fewest switches surface hopping algorithm based on coupling values. Approximations to surface hopping probabilities exist, but they are limited to systems comprising a low density of states. An alternative way is to approximate hopping probabilities from wave function overlaps of two sequential time steps, which is often employed in conventional MQC simulations.[259] However, providing ML approximations to molecular wave functions is a difficult task and has not been achieved for the excited states until now. The ambition to find another route toward ML-approximated hopping probabilities remains. To make ML applicable for many molecular systems and quantum chemical programs, a strategy

that allows to omit the calculation of NACs for training set generation is required.

Admittedly, an approximation to NACs based on second derivatives of PESs exists, which has been shown to be accurate for 1D systems. This approach is limited due to the high costs of calculating the second derivatives of the required squared energy-gap PESs with quantum chemistry methods. This limitation is not present when using ML models instead, as accurate PESs are available, which can be efficiently derived with respect to Cartesian coordinates. The second order derivatives are implemented in SchNarc and for MLFF NNs. Their performance is assessed by computation of frequencies for $CH_2NH_2^+$, which are further compared to values obtained from quantum chemistry in the supporting information of Ref. 157. Here, the accurate ML PESs provided with SchNet are evaluated for their applicability to approximate NAC vectors. Equation 3.21 and the process explained in chapter 3.5.1 are employed for the computation of squared energy-gap Hessians of ML PESs. This approach is tested using a "one-shot" linear vibronic coupling (LVC) model[241] of SO_2. The advantage of the LVC model is that it approximates the excited-state PESs by harmonic potentials with parameters obtained from a single quantum chemical calculation. The analytic functions of PESs make the computations of surface hopping dynamics highly efficient and at least as cheap as surface hopping dynamics with ML models. They further allow for a precise comparison of analytic NACs obtained from LVC and NACs obtained from SchNet. Hence, the LVC model of SO_2 is well suited for the purpose of this study.

The parameters for the LVC model of SO_2 are obtained from Ref. 241 and is based on MR-CISD. Three singlet and three triplet states are treated. Due to the low costs of the LVC model 200 trajectory surface hopping MD simulations with the LVC model are carried out for 700 fs to generate the training set, resulting in a total amount of 280,200 data points including energies, forces, and NACs. 20,000 data points are used for training. Dynamics with approximated NACs are compared to dynamics with NACs that are trained as derivatives of virtual ML properties. Their accuracy is further validated with photodynamics simulations of $CH_2NH_2^+$.

Figure 18(a) shows potential energy curves along the asymmetric stretching mode of SO_2 of the LVC model (continuous lines) and the SchNet deep learning model (dashed lines) for three singlet states on the left and three triplet states on the right. Due to the symmetry of the molecule NACs arise between the S_1 and S_2 states and between the T_1 and T_3 states and not between any other states. The singlet and triplet states couple via SOCs that are also trained for the purpose of surface hopping MD simulations. The NAC vectors are illustrated by their norms in both plots in panel (a). The trained and rotationally covariantly predicted values are illustrated using dashed lines, whereas dotted lines show norms of approximated NACs from the squared energy-gap PESs. As it is clearly visible, the norms of approximated NACs matches the ones obtained from the LVC model and can reproduce the peaky nature of NACs. The approximated NACs are even closer to the LVC model than the trained NACs. Noticeably, the trained NACs between the triplet states approach zero at the point, where the coupled triplet states are closest to each other. This effect might be due to the geometric phase, which is not considered when NACs are trained, but is supposed to be negligible for MQC simulations with the surface hopping approach in most cases. The approximated NACs are only valid close to conical intersections. Therefore, they are set to zero when the energy gap between adjacent electronic states exceeds a pre-defined threshold and consequently, the approximated NACs approach zero faster than the trained NACs and reference NACs.
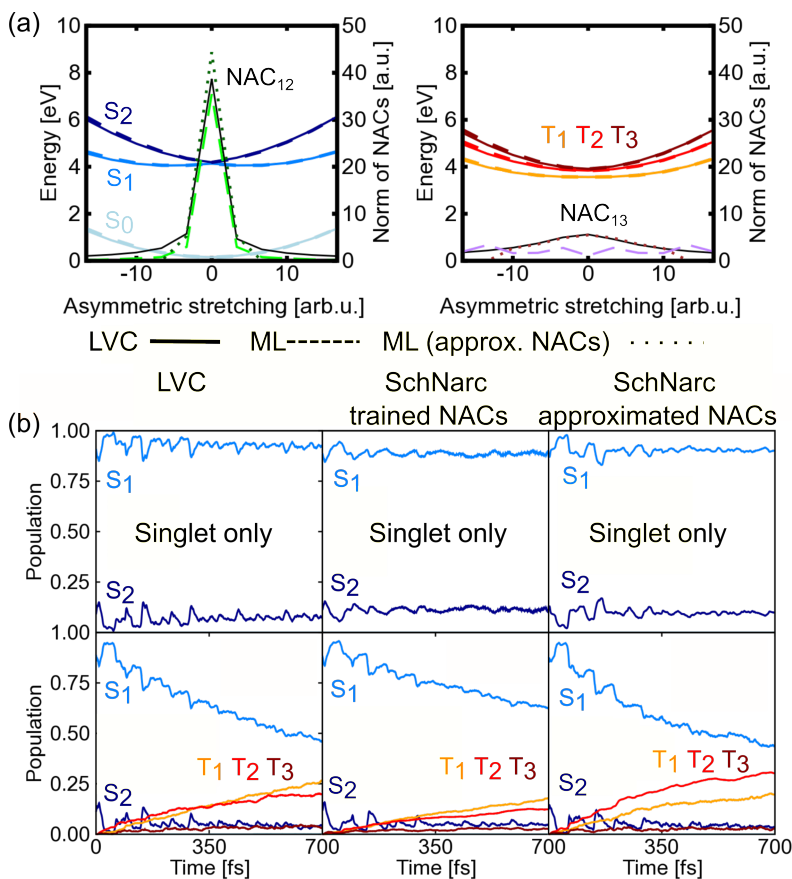
Figure 18: Potential energy curves (a) and population plots (b) of SO_2 computed with an LVC model based on MR-CISD[241] and SchNarc. (a) The potential energy curves are shown for three singlet states on the left and three triplet states on the right along the asymmetric stretching of SO_2. In addition, the norms of the available NAC vectors between the S_1 and S_2 states and between the T_1 and T_3 states are shown. The LVC curves are shown by solid lines and compared to predictions of SchNarc with NACs trained (dashed lines) and approximated from ML PESs (dotted lines). (b) The corresponding population curves of singlet-only dynamics and dynamics considering singlet and triplet states are illustrated. The left, middle, and right plots are obtained from averaging 1,000 trajectories obtained from LVC, SchNarc with trained NACs, and SchNarc with approximated NACs, respectively. (Adapted from Ref. 157 under CC-BY.)

The quantum population curves obtained from surface hopping dynamics of 1,000 initially excited trajectories are illustrated in panel (b). The upper plots show dynamics considering only singlet states, while the lower plots additionally include triplet states, which couple between each other and additionally between the singlet states. The population curves using trained NACs in the middle plots and approximated NACs in the rightmost plots indicate that both ML approaches lead to comparable dynamics as LVC models in the leftmost plots. Hence, both ML NACs variants are assumed to be suitable to compute hopping probabilities for surface hopping MD simulations.

Additionally, it is sought to investigate the accuracy of approximated NACs with photodynamics of $CH_2NH_2^+$, which is also considered to be well suited to evaluate the applicability of approximated NACs for ultrafast population transfer. The respective population curves can be found in Fig. 19 in the right panel and reference dynamics

61
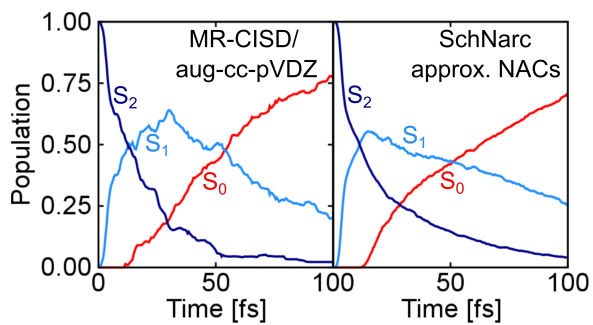
Figure 19: Population curves obtained from photodynamics simulations of $CH_2NH_2^+$ with the reference method MR-CISD/aug-cc-pVDZ in the left panels for 100 fs. SchNarc population curves using approximated NACs and therefore only energies and gradients show the potential of approximated NACs for fast population transfer. (Adapted from Ref. 157 under CC-BY.)

in the left panel. As can be seen, NACs approximated from ML PESs are valid for photodynamics simulations and ultrafast population transfer can be reproduced.

4.4 THE EXCITED STATES OF TYROSINE

After having developed (1) a procedure to efficiently compute a training set, (2) a correction algorithm for making inconsistent quantum chemical data due to the arbitrary phase of the wavefunction learnable, and (3) the SchNarc approach for efficient photodynamics, the necessary ingredients to tackle larger and more complex systems are, in principle, available. To evaluate the performance of the developed methods, the molecule tyrosine is taken as a test case. This molecule is considered a suitable challenge, as there is evidence that intersystem crossing occurs relatively slowly, i.e., in the ns regime.[351;352] Computing the excited-state photodynamics simulations for 1 ns with quantum chemistry is far from being feasible, but have been shown to be feasible with ML for $CH_2NH_2^+$.

Studying the amino acid tyrosine is important, because it is among the essential amino acids that can absorb light and give rise to the multifaceted photochemistry of peptides and proteins.[17;18;41;353−355] Tyrosine can also serve as an example to investigate proton-coupled electron transfer reactions,[356;357] which are common in photosynthesis[358;359] for example.

Several reasons led to the choice of investigating the isolated tyrosine molecule in the gas phase rather than in solvent. The inclusion of an environment with ML as it is done in mixed quantum mechanics/molecular mechanics simulations for example, is associated with some additional challenges. For example, a way to let an ML model feel any environmental changes has to be provided. Up to date, only a mechanical embedding scheme with ML exists for ground state MD simulations[360] and a lot still remains to be done.[361]

Experiments of tyrosine and its chromophores in vacuum exist, which focus on shedding light on the role of the $\pi\sigma^*$ state in the photodynamics of these molecules.[52;362−366] Interestingly, the photodissociation of hydrogen atoms attached to oxygen or nitrogen via a dissociative $\pi\sigma^*$ channel was suggested as an important nonradiative decay mechanism for phenol. Besides, C-H and C-C bond cleavage could not be excluded. Ashfold et. al[367] investigated the photodynamics of tyrosine with total kinetic energy release measurements, but no clear picture could be obtained due to experimental limitations and noise. A qualitative result was nevertheless suggested by the authors, i.e., that the hydrogen dissociation decreases from p-ethylphenol, to tyramine, and to tyrosine. This size-dependency of the photodynamics was further observed by Tseng et. al.[363;364] While phenol shows dissociation after photoexcitation, C-C bond breaking is reported for p-(2-aminoethyl)phenol. The experimental findings and supporting DFT calculations in their study suggest that the dynamics of tyrosine might differ completely from those of phenol questioning the suitability of phenol as a test system for the excited states of tyrosine.

Up to date, theoretical effort has mainly been devoted to the investigation of phenol, see e.g. Refs 352;363−365;368−371, and only a few studies describe tyrosine in its protonated or neutral form.[372−374] The photodynamics of isolated tyrosine are investigated by Tomasello et. al[373] with surface hopping. The study reveals that after photoexcitation to the bright S_1 and S_3 states, the molecule remains trapped in the S_1 state after 200 fs simulation time. No triplet states are considered.

Because of all these considerations and the fact that only a qualitative picture could be obtained from experiments and theoretical investigations of tyrosine are limited to static investigations or short dynamics, the role of the dissociative $\pi\sigma^*$ in tyrosine is not yet fully clear. ML models are particularly well suited for tyrosine, because in principle, a comprehensive training set of tyrosine should be extendable also to

smaller subsystems of tyrosine. In combination with an atom-wise description of the molecules, the investigation of the photochemistry of tyrosine and its subsystem could be possible.

4.4.1 *Choosing the Right Reference Method*

The accuracy of an ML model is dependent on the reference method. Therefore, a proper method to compute the training set has to be selected. As photodissociation and bond-breaking can possibly take place after photo-excitation, a multi-reference method is most suitable for this task. The reference geometry for the training set that serves as a reference for phase correction is selected to be the lowest-energy conformer of tyrosine. The molecular structure is taken from Ref. 375 and optimized with MP2/TZVP along with all other 11 expected lowest energy conformers of tyrosine. Table 3 shows the energies of the optimized conformers. The structures are shown in Fig. 20 and are listed according to Ref. 375 where the structures are analyzed in detail. The ORCA program is used for the geometry optimizations.[376]

Conformer Nr.[375]	Energy [kcal/mol]
1	0.00
2	0.41
3	2.34
4	2.35
5	0.65
6	0.67
7	1.29
8	1.27
9	1.54
10	1.51
11	3.62
12	3.72

Table 3: Relative energies of the conformers of tyrosine with respect to the energetically most favorable conformer optimized at MP2/TZVP level of theory. The numbers of the structures are in accordance to Ref. 375, which was used to obtain the initial structures of the conformers.

The excited states of tyrosine, i.e., those of the energetically most favourable structure, are described with CASSCF(12,11)[199] and CASPT2(12,11)[206] as implemented in openMolcas.[377] For comparison and to evaluate the active space of CASSCF, ADC(2) is used as implemented in Turbomole,[209] because it offers a good compromise between computational efficiency and accuracy. The ano-rcc-pVDZ[378] is selected as a basis set for CASSCF and CASPT2 and cc-pVDZ is used for ADC(2) calculations. The excitation energies of the 5 lowest singlet states and 8 lowest triplet states are given in Table 4, which are additionally compared to energies of TD-DFT/PBE0/SV(P)[373] and experimental values estimated from an absorption spectrum in the gas phase taken from Ref. 379. TD-DFT calculations of Ref. 373 and additional reference calculations show that the S_6 state contributes only negligibly to the absorption spectrum of tyrosine and the S_5 state is a dark state. Therefore singlet states up to the S_4 state are considered to be important for surface hopping dynamics with a total number of 8 triplet states included.
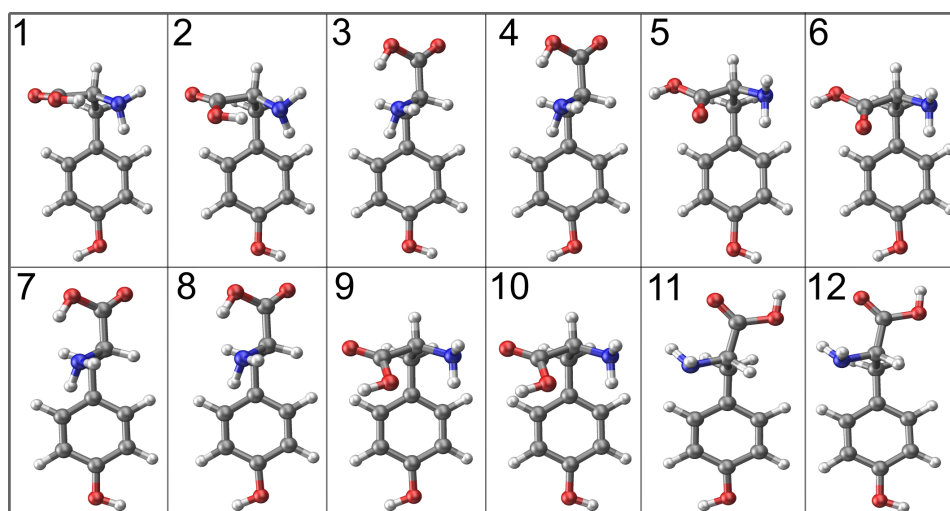
Figure 20: Geometries of the 12 energetically most favorable structures of tyrosine optimized at MP2/TZVP level of theory.

As can be obtained from Table 4, the CASSCF energies are furthest from CASPT2 and experimental values. As expected, the S_3 and S_4 states computed with CASSCF are remarkably high in energy. The experimental values cannot be reproduced accurately with any of the used methods, but a qualitative correct description can be expected from ADC(2) or CASPT2. The TD-DFT energies are very close to each other.

Table 4: ADC(2)/cc-pVDZ, CASSCF(12,11)/ano-rcc-pVDZ, and CASPT2/ano-rcc-pVDZ energies of the MP2-minimum energy conformation of tyrosine in the gas phase compared to values obtained from TD-DFT and an experimental spectrum. The relative energies with respect to the S_0 state are reported. The TD-DFT computation refers to the functional and basis set PBE0/SV(P). Note that S_2 is a dark state according to ADC(2), TD-DFT, CASSCF.

State	ADC(2)	CASSCF(12,11)	CASPT2(12,11)	TD-DFT	Exp. [379]
S_1	5.00	4.94	4.63	5.11	~ 4.5
S_2	5.92	6.43	5.99	5.51	
S_3	6.20	7.87	6.24	6.00	~ 5.5
S_4	7.16	8.19	7.70	6.06	~ 6.5
T_1	4.27	3.86	3.91		
T_2	4.60	4.91	4.34		
T_3	5.01	4.98	4.49		
T_4	5.58	6.10	5.32		
T_5	6.19	6.94	5.65		
T_6	6.75	7.14	7.01		
T_7	6.99	7.31	7.10		
T_8	7.03	7.47	7.15		

Fig. 21(a) shows the 12 molecular orbitals that are closest to the HOMO and LUMO computed with ADC(2). The theoretical absorption spectra are plotted along with the experimental spectrum[379] in panels (b) and (c). The theoretical spectra considering only the most stable conformer of tyrosine and 5 singlet states are scaled according to the lowest experimental peak and the experimentally fit spectrum is shown by dashed lines. Panel (b) contains the ADC(2) absorption spectrum, which is obtained
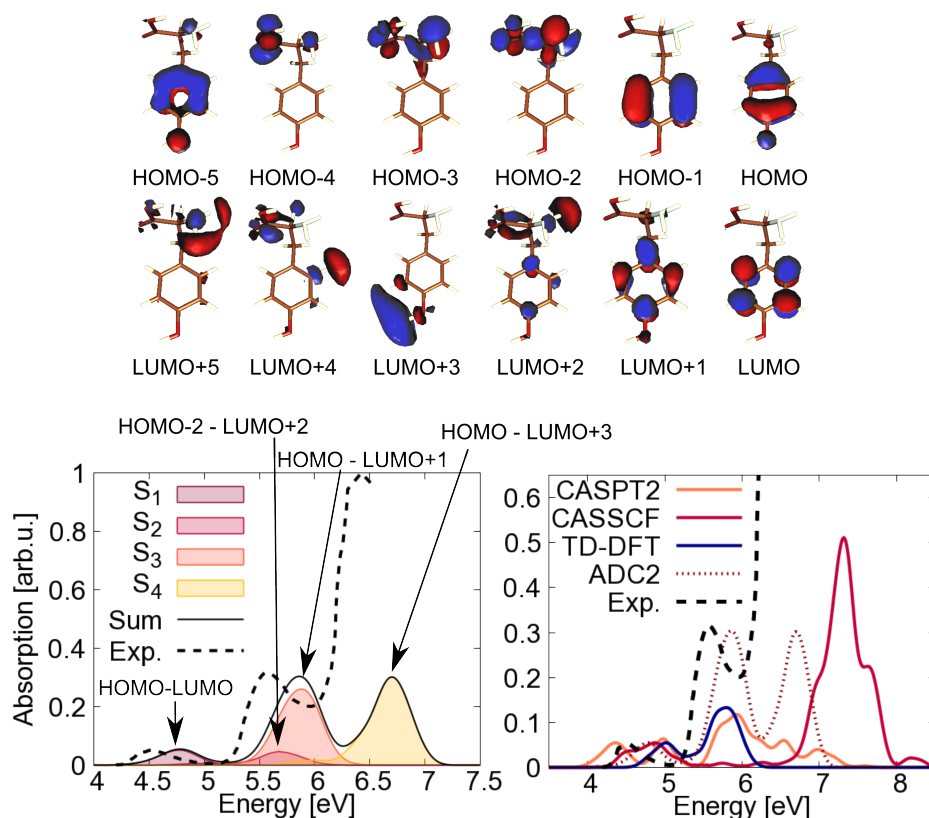
Figure 21: (a) The minimum energy conformer of tyrosine along with molecular orbitals around the HOMO and LUMO computed for 5 singlet states using ADC(2). (b) The respective absorption spectrum computed with ADC(2) for 1000 Wigner-sampled conformations is shown together with the experimental spectrum for tyrosine in the gas phase ("Exp.") obtained from Ref. 379. (c) The absorption spectrum is additionally calculated with TD-DFT/PBE0/SV(P), CASSCF(12,11), and CASPT(12,11) using 100 Wigner-sampled conformations. For comparison, the experimental spectrum and the ADC(2) spectrum is plotted. The full width at half maximum for the Gaussian convolution was 0.2 eV in all theoretical spectra. The absorption peaks are scaled according to the energetically lowest lying experimental absorption peak. The corresponding vertical excitation energies are specified in Table 4.

from the average of 1,000 Wigner sampled molecular conformations. As can be seen, the S_2 state has important contributions from the lone pair orbital in N of the NH_2 group and is a dark state. The S_1 and S_3 states are bright and mainly $\pi\pi^*$ configurations contribute to their character. The S_4 state is mainly characterized by a $\pi\sigma^*$ configuration and dissociation of hydrogen from the hydroxy group of the phenyl ring cannot be excluded.

Compared to the CASPT2, CASSCF and TDDFT spectra, ADC(2) matches the experimental spectrum qualitatively best. TDDFT shows two main peaks in the absorption spectrum compared to three experimental peaks. CASSCF is shifted to higher energies and the peaks of the different states cannot be distinguished in the high-energy region. CASPT2 can correct the latter to some extent and shows three peaks, comparable to the experimental spectrum.

In order to evaluate the relevant active space, single point calculations and scans along different reaction coordinates are carried out with different active spaces and different numbers of states. With respect to CASSCF, the best compromise between computational efficiency and accuracy can be obtained with 12 active electrons in

11 active orbitals. However, the lone pair located at the nitrogen atom of the amino group and σ orbitals are not included in the active space, which are considered to be relevant for the S_2 and S_4 states according to ADC(2). Using a larger active space than 12 active electrons in 11 active orbitals in addition to a larger amount of electronic states does not improve results, but leads to considerably higher computational effort. An active space up to 14 electrons in 18 orbitals is tested and the computation of energies of 6 singlet states without any triplet states and forces already takes about 5 hours and 40 minutes on a 2x Intel Xeon E5-2650 v3 CPU, making a larger active space impracticable, even if better results were obtained. Consequently, an active space of 12 electrons in 11 orbitals in combination with 5 singlet states and 8 triplet states is employed for computing potentials, but the potential energy curves are not smooth. An example is given in Fig. 22, where the scan along the PhO-H bond is shown with the different quantum chemistry methods.
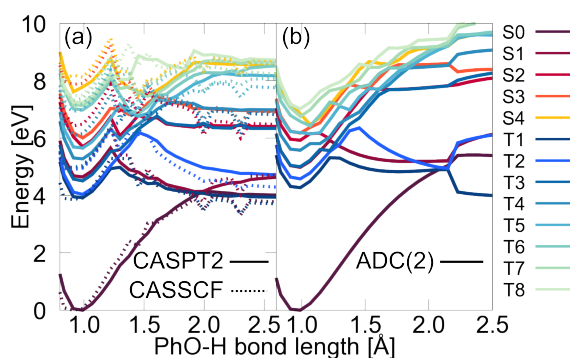
Figure 22: Scan along the PhO-H bond distance of tyrosine for 4 singlet states and 8 triplet states computed with (A) MS-CASPT2(12,11) and CASSCF(12,11) illustrated using solid and dotted lines, respectively. Panel (b) shows potential energy curves computed with ADC(2).

As can be seen in panel (a) in Fig. 22, the 13 described states are very close in energy, which makes inconsistent potential energy curves problematic to fit with ML. CASPT2 (continuous lines) can improve results of CASSCF (dotted lines), but is still deemed to be insufficiently accurate to compute a training set for ML models, while additionally suffering from high computational costs. The potential energy curves obtained with ADC(2) are far better suited for the purpose of generating a training set and can be computed computationally more efficient. The negative side effect of this single-reference method remains, i.e., that ADC(2) leads to qualitatively wrong results in dissociative regions of the PESs, as it is visible in panel (b). These observations led to the idea of using ADC(2) to compute the training set and to correct the data points in conformational regions of the PESs, in which ADC(2) fails to describe the qualitatively correct behaviour. Data points for ML models are envisioned to be manipulated according to results from multi-reference methods to capture the qualitatively correct physics with ML and to combine the benefits of single-reference and multi-reference methods.

4.4.2 *Training Set Generation*

The training set is computed with ADC(2). An additional singlet and triplet state is considered to guarantee sufficiently large wave function overlaps for phase correction. The additional computed states are not included in the training set. Scans along normal modes and combinations thereof are carried out and a scan from the

zwitterionic form to the neutral form of tyrosine result in 698 data points. Two additional conformers are incorporated in the initial training set (conformers 5 and 7 of Fig. 4 in Ref. 375 with energies listed in Table 3). A total number of 1967 data points is obtained to train preliminary ML models.

MLFF NNs are used for the adaptive sampling scheme, because they are less expensive with regard to training times. Training a SchNet model for tyrosine using 5 singlet and 8 triplet states in addition to SOCs takes several days on a GPU, while training a MLFF NN takes a couple of hours on a CPU with the inverse distance matrix as a descriptor (depending on the training set size 1-4 hours). For the sake of the training set generation, MLFF NNs are thus deemed sufficient. As the NACs are missing in ADC(2) codes and approximated NACs are neither evaluated with MLFF NNs, nor deemed sufficiently accurate with this small initial training set at hand, the adaptive sampling scheme is used as follows: 25 trajectories with different initial coordinates and velocities obtained from a Wigner distribution are launched on each of the different excited singlet states, i.e., the S_1, S_2, S_3, and S_4 states. The mean of 6 NNs for energies, forces, and SOCs is used to propagate the nuclei using a time step of 0.5 fs. As SOCs are considered with MLFF NNs, the molecule can enter the triplet states, but no internal conversion is allowed. In the beginning, the average of 2 NNs is used, but led to unstable simulations. In general, simulations are more stable when the average of more NNs is passed to an MD program, which has already been observed for ground-state MD simulations[106]. Hyperparameters of NNs differ slightly in the used learning rate or the random seed to shuffle the data set. After a pre-defined optimized number of steps the learning rate is annealed by multiplication with a factor, usually between 0.9 and 1.0. Random grid search and optimization of the number of hidden layers is carried out to arrive at hyperparameters for the adaptive sampling scheme. The thresholds that are used to decide whether a data point is insufficiently sampled or not are 25 kcal/mol (1.08 eV) for energies, 25 kcal/mol/a.u. for forces, and 0.000022 a.u. (4.83 cm^{-1}) for SOCs. Thresholds are decreased by multiplication with a factor of 0.95 after each iteration of the adaptive sampling scheme. When the training set size is doubled, hyperparameters are optimized again, i.e., when approximately 4,000, 8,000, and 16,000 data points are sampled. After 16654 data points have been sampled, the adaptive sampling scheme is stopped as the adaptive sampling scheme has led toward conformations far from the minimum energy structure. The inclusion of these data points is complicated by the necessity to phase correct them. Along certain reaction coordinates, many intruder states enter, as it is for example visible in Fig. 22, which prohibits an efficient training set generation. Already during the last adaptive sampling steps, more and more data points required the computation of many excited states, which should have been accounted for in the beginning. The additionally computed singlet and triplet state are not sufficient in this case and some conformations even required about 100 different excited singlet and triplet states for proper phase correction. Computing this number of excited states is infeasible for a large number of data points. Nevertheless, this training set is used to introduce hand-tailored data points. 105 hand-tailored manipulated data points as described in 3.6 are added leading to a total amount of 16759 preliminary data points. This training set is assumed to be sufficient to train ML models and evaluate whether artificial data can be used to learn the correct physics in dissociative regions or not.

16500 data points are used to train NNs with and without artificial data points to compare their performance. An early stopping mechanism is employed and hyperparameters are obtained from random grid search. Focus is placed on optimization of the learning rate, the L2 regularization rate, and the number of hidden layers.
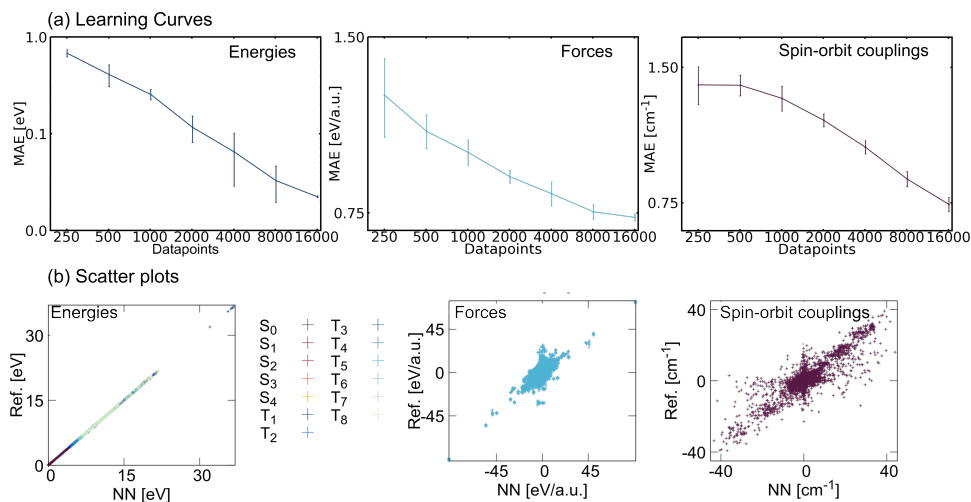
Figure 23: Learning curves (a) and scatter plots (b) for NNs trained on a data set of tyrosine including ADC(2) and manually manipulated data points. Panel (a) shows the learning curves as the mean of five calculations, in which the training and validation set is split differently, similar to 5-fold cross validation. The error bars are obtained from the different NN models and provide a measure of uncertainty. This process is described in more details in the appendix A.2. Panel (b) illustrates the prediction capabilities of NNs trained on 16,500 data points on a small test set of 259 data points, i.e., 3367 energy values, 242,424 force values and about 105,154 SOC values.

Learning curves for energies, forces, and SOCs reported on the validation set that incorporates artificial data points are illustrated in Fig. 23(a) along with corresponding scatter plots in panel (b). The plots show that the NNs can learn from the fused training set. Noticeably, the scatter plots of the forces indicate that more data points might be required for desired photodynamics simulations. Due to the use of the inverse distance matrix data augmentation[380;381] might also lead to improved results.[325] 5-fold cross validation is then used to further optimize the hyperparameters of the NNs. A final number of 8 hidden layers and 50 nodes per hidden layer for training of energies and forces of 13 states simultaneously and 15 hidden layers with 50 nodes per hidden layer for training SOCs, i.e. 406 complex, phase corrected values per data point, are used. The NNs with the lowest error on the remaining test set are used for the following predictions of potential energy scans. The MAEs of energies, forces, and SOCs on the preliminary test set are 0.0227 eV, 0.626 eV/Bohr, and 0.971 cm^{-1}, respectively.

4.4.3 *Performance of NNs in Dissociative Regions*

Fig. 24 shows potential energy curves of NNs trained on 16,500 ADC(2) data points. For comparison, the ADC(2) and artificial reference points are shown using dashed lines. The PhO-H bond and the N-H bond are elongated as an example. As it is visible from panels (a) and (c) an exclusion of the artificial data points leads to wrong energies in the dissociative regions of the PESs and the energies of all states increase rapidly. Panels (b) and (d) indicate that the NNs can learn from the data set that contains 105 hand-tailored data points. Additionally, predictions up to a distance of 1,000 Å are possible, which exemplifies that the NNs predict the degeneracy of certain pairs of states also at large bond distances, which are not explicitly encountered for.

These results demonstrate the possibility of generating hand-tailored potentials for ML models, with which inaccuracies in different methods can be corrected.
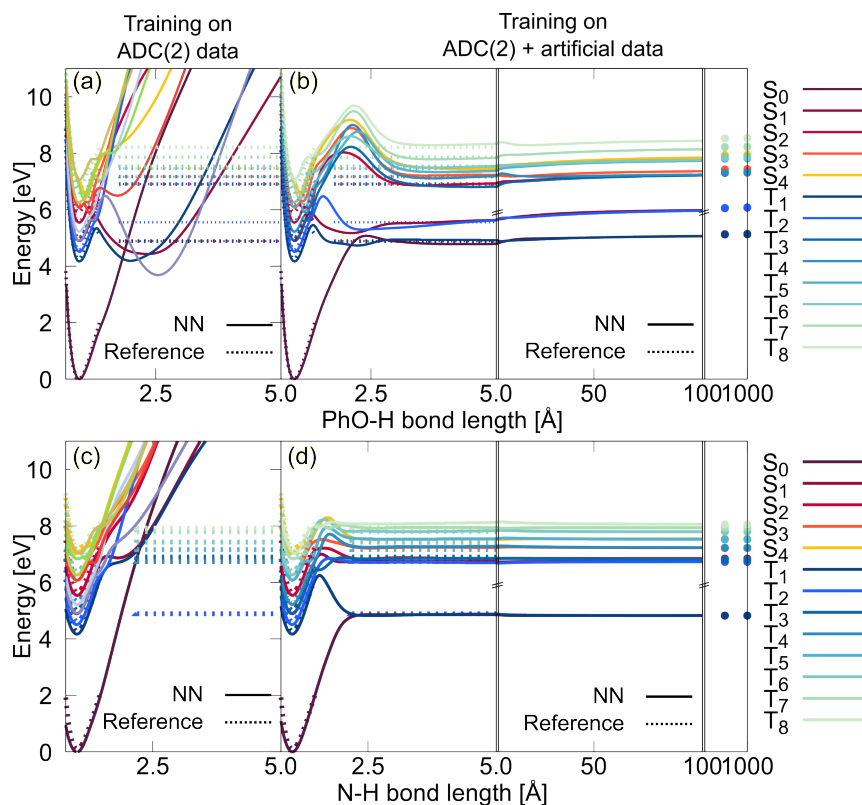
Figure 24: Comparison of potential energy curves predicted with NNs trained on 16,500 data points without (panels (a) and (c)) and with artificial data points (panels (b) and (d)) along the PhO-H and N-H bonds of the minimum energy conformer of tyrosine. NNs are shown with continuous lines and reference data (ADC(2) and artificial points) with continuous lines.

The capability of NNs to predict different reaction coordinates and to go beyond the included reaction coordinates is assessed with the manually tuned potential. Therefore, the hydrogen is detached in different directions and different conformers are tested. The PhO-H and N-H bonds are illustrated as an example in Fig. 25. As it is visible, the NNs predict correctly that singlet and triplet states are degenerate at large bond distances (panels (a) and (c)). However, panels (b) and (d) indicate that NNs are not yet accurate enough to predict other conformers. Noticeably, the dissociation barrier is visible, which could be accurate enough to predict qualitatively correctly whether dissociation takes place after excitation or not. The nonphysically increasing potentials predicted with NNs, which are trained on pure single-reference data (as given in Fig. 24 panels (a) and (c)) can be corrected with the hand-tailored data points. Importantly, the energies are almost degenerate and do not increase also for the different conformers, which are not included in the training set. A better prediction accuracy for other conformers might be possible by augmentation of data points as the current NN models are not invariant to permutation.

The results obtained with NNs, which are trained on a hand-tailored, partially corrected single-reference PES of a manifold of singlet and triplet states show the potential of ML to go beyond the accuracy of single-reference methods. The manipulation of data in some dissociative regions of the molecule offers the possibility to maintain the computational efficiency of single-reference methods and their smooth
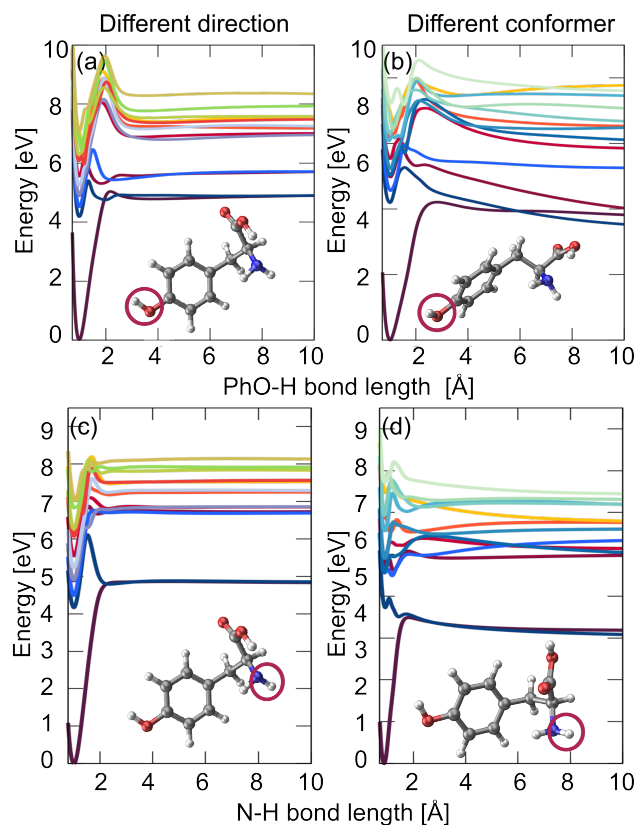
Figure 25: Potential energy curves of NNs trained on 16,500 data points of a data set comprising ADC(2) and artificially generated data points. Reaction curves along the PhO-H and N-H bonds are shown – pointing toward different directions (panels (a) and (c)) and belonging to different conformers (panels (b) and (d)).

PESs, while their inaccurate description of some conformational regions can be corrected according to multi-reference methods in conformational regions, in which single-reference methods fail to describe the qualitatively correct physical behaviour. NNs should therefore be able to learn from fused data sets and predict dissociation barriers as well as degeneracy of states on a qualitatively correct picture.

In order to assess the performance of ML models trained on the hand-tailored excited-state potentials, observables should be computed and compared to experiments. The time constants of the transitions and the dissociation reaction should be computed and compared to experiments.[382] The scatter plots of forces and reaction scans of different conformers of tyrosine indicate that more data points are necessary for ML enhanced photodynamics. However, as already discussed earlier, the phase correction procedure increases the computational effort to further expand the training set considerably due to the amount of excited states that have to be included from the beginning. Reducing the number of excited electronic states along a reaction coordinate can in principle reduce the costs, but makes it difficult to automatize the training set generation as it requires manual interaction to adapt the number of relevant excited states. A straightforward and efficient computation of a training set without the need to account for any phases and as little human interaction as possible would be highly desirable. To this aim, an intrinsic ML-based phase correction procedure as explained in 3.3 is developed and implemented in the SchNarc approach. This approach is termed phase-free training and will be discussed in details in chapter 4.5. The next sub-section is dedicated to the remaining problems

that additionally limit a comprehensive ML-based study of the temporal evolution of tyrosine in the excited states.

4.4.4 *Additional Challenges of Fitting the Excited States of Tyrosine*

Besides the challenge to find the right phases for excited-state properties, there are a number of other challenges that need to be overcome, which are briefly discussed below.

(1) Up to now, the training set is based on one conformer of tyrosine, i.e., the minimum-energy conformer. However, there are many other conformers that are important in order to obtain statistically significant results at a given temperature. If possible, the photodynamics of every conformer should be investigated at different reaction conditions, i.e., different temperature and excitation energies. Until now, the PESs of tyrosine fitted with NNs are mainly limited to conformational regions close to the minimum-energy conformer. It is not known, how many additional data points are required to describe all possible conformers of tyrosine with ML. In addition, the high density of excited electronic states further complicates the fitting of PESs and requires consistent quantum chemical data.

(2) As it is shown for the excited states of $CH_2NH_2^+$ in section 4.1, a better descriptor can reduce the amount of necessary training data for accurate photodynamics simulations and can provide more accurate PESs. However, test runs with SchNet reveal that the respective PESs cannot simply be constructed from atom-wise, local contributions. The cutoff, which defines the environment an atom can perceive, should possibly be large enough to incorporate the whole molecular system.[383] As a consequence of the enlarged molecular "image", more Gaussian functions for the expansion of the atomic distances are required to circumvent a loss in spatial resolution. The resulting larger network architecture leads to longer training times and large memory consumption even for very small mini-batch sizes. These problems limit the applicability of atom-wise representations for the excited states of large systems and indicate that molecule-wise descriptor might be beneficial for the excited states. As a negative side effect of molecule-wise descriptors, more data points are required and data augmentation is necessary to guarantee permutation invariance.

(3) The computation of the excited states of tyrosine along with the forces and SOCs is expensive and it would be beneficial to reuse this training set and the trained ML model also for smaller subsystems of tyrosine, such as phenol or tyramine, with only a few additional reference calculations of these smaller systems. It would also be good to treat the protonated form of tyrosine or tyrosine with capped amino and carboxy groups to mimic this molecule in a protein environment. In order to allow for such a description, an ML model must comply with at least two main requirements, which have not yet been fulfilled for the excited states: Firstly, an atom-wise representation is needed, but a description of the excited states via atom-wise local contributions has not yet been shown to be possible. Secondly, a quantum chemical method is required, which can provide accurate PESs and properties thereof for all types of systems to be treated with an ML model. Multi-reference methods are limited in this respect, unless an impracticable FCI description is employed. For example, if a training set for phenol is calculated, an active space of 10 active electrons in 9 active orbitals might be sufficient to accommodate the molecular orbitals ($3\times \pi\pi^*$ of the phenyl ring, lone pair of oxygen, and $\sigma\sigma^*$ of the hydroxy group) considered to be relevant for the excited states. In contrast, tyrosine requires a much larger active

space, and the inclusion of peptide bonds yet requires an even larger active space than tyrosine and phenol. Either different active spaces or unnecessarily large active spaces for small systems are the result. The use of single reference methods is not only computationally more efficient, but also advantageous for describing different molecular systems in one training set. ADC(2), for example, which does not require the selection of a density functional or an active space, is well suited for this task. The fusion of single reference data with artificially generated data points to obtain correct physics in some conformational regions of PESs is therefore promising to compromise the loss in accuracy due to the single-reference description. As it is shown before, only a small amount of such manipulated data points is needed to obtain a qualitatively correct description in dissociative regions of the PESs, i.e., about 0.6% of the training set. This small number should make the generation of hand-tailored potentials for the excited states feasible.

Tyrosine is a good example to illustrate the existing problems, which still need to be overcome, if a description of molecular excited electronic states with ML is envisioned for more complex and larger systems than e.g. $CH_2NH_2^+$ or SO_2. One of the problems, which has been mentioned before and could be solved in this thesis, is the training of excited-state properties, which carry an arbitrary sign due to the arbitrary phase of the wave function. The performance of the phase-free training algorithm approach, which renders the learning of such inconsistent quantum chemical data possible, is evaluated in the next sub-section.

4.5 PHASE-FREE TRAINING FOR PHOTODYNAMICS: $CH_2NH_2^+$ AND CSH_2

An efficient and affordable training set generation for the excited states requires an alternative treatment of the phase of the wave function to omit the costly and time consuming phase correction procedure. Up to date, training sets that are used to accurately fit the excited states of molecules are either phase corrected or do not include any properties that are influenced by the arbitrariness of the wave function phases. The phase corrected training sets, which are available in the literature, are partly or completely manually phase corrected,[208;290] which is even more tedious than the phase correction procedure applied here. Only one study exists, where phase-inconsistent NACs are fitted, which leads to insufficiently accurate ML models for photodynamics simulations.[155] Therefore, training on inconsistent raw quantum chemical data is assessed using different types of loss functions, i.e., the standard L2 loss function and the phase-less loss function as introduced in chapter 3.3.

Photodynamics simulations of $CH_2NH_2^+$ are used to prove the validity of the phase-free training with the phase-less loss function. In a first attempt, the photodynamics of $CH_2NH_2^+$ are evaluated, because a phase corrected training set already exists that has been shown to be comprehensive to allow for ML-enhanced photodynamics simulations. This training set is recomputed with the same reference method, MR-CISD/aug-cc-pVDZ, but without applying phase correction. Deep learning models are trained with SchNarc on 3,000 data points using the L2 loss function. NACs are trained in a rotationally covariant way as derivatives of virtual ML properties as explained before. In addition, the same ML model, but with the phase-less loss function instead of the L2 loss function, is trained. Photodynamics simulations for 100 fs are carried out after excitation to the S_2 state. Results are illustrated in Fig. 26. Panel (a) shows population curves of the reference method (dotted lines) and of SchNarc models, which are trained using the L_2 loss function (solid lines).

As it is visible, the photodynamics simulations of the reference method cannot be reproduced. Panel (b) illustrates the population curves of SchNarc with ML models trained with the phase-less loss function. As it is visible, the phase free training enables the use of inconsistent quantum chemical data and the ultrafast dynamics is correctly reproduced.
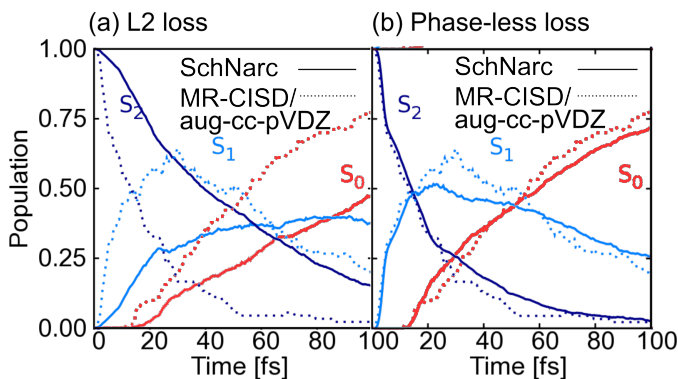
Figure 26: Photodynamics simulations of $CH_2NH_2^+$ considering three singlet states. The reference population curves obtained from MR-CISD/aug-cc-pVDZ are plotted against population curves obtained from SchNarc models trained on NACs with the L2 loss function (panel (a)) and the phase-less loss function (panel (b)). (Adapted from Ref. 157 under CC-BY.)

Additionally, the performance of this approach is evaluated for the photodynamics simulations of CSH_2, which is characterized by slow population transfer. The reference method is CASSCF(6,5)/def2-SVP. The training set for this molecule contains 4703 data points. The initial training set is computed with scans along different normal modes of the molecule. Energies, forces, and SOCs are trained for two singlet states and two triplet states. Adaptive sampling is applied to further expand the training set with MLFF NNs. No phase correction is applied. For photodynamics simulations, the adapted version of SchNet and the phase-free training algorithm are used. NACs are approximated from PESs and their derivatives. The training set generation and the ML models are discussed in more details along with scatter plots and scans along the C=S bond distance in the reprint in section A.3 and the supporting information of Ref. 157. Fig. 27 shows the population curves obtained
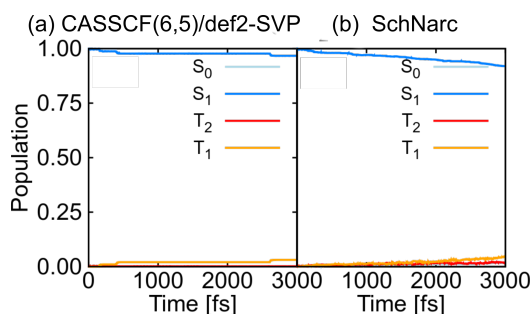
Figure 27: Photodynamics simulations of CSH_2 are carried out with CASSCF(6,5)/def2-SVP and SchNarc for 3 ps considering two singlet states and two triplet states. The phase-less loss function and the NAC approximation form ML PESs is applied. (Reproduced from Ref. 157 under CC-BY.)

with quantum chemistry in panel (a) and SchNarc in panel (b). As it is illustrated in Fig. 27, the photodynamics simulations of the reference method, which are computed for 3 ps, can be reproduced with SchNarc. A negative side effect of the phase-less

loss function are the longer training times, which rise with the number of excited electronic states. More details on training times and the performance of the phase-free training procedure can be found in Ref. 157.

The results shown here demonstrate not only the validity of the intrinsic ML-based phase correction scheme, but also the validity of the NAC approximation from squared energy-gap Hessian of PESs, which are usable to predict hopping probabilities of systems comprising fast as well as slow population transfer. The ML-based intrinsic phase correction is especially important to describe medium-sized to large systems with a manifold of excited electronic states and to enable an efficient and affordable training set generation. With respect to photodynamics simulations, a consistent phase can be obtained as ML models are smooth functions by definition. The phase-free training process thus allows for omitting the phase correction of training data completely when MQC photodynamics simulations are to be described.

5

SUMMARY AND CONCLUSION

The main objective of this thesis was the investigation and development of machine learning (ML) models for the excited electronic states of molecules and their application in photodynamics simulations. The computational efficiency of ML models in combination with the accuracy of the underlying quantum chemical reference methods makes ML models promising to advance the research field of photochemistry, which is severely limited by the high complexity and computational effort involved in the description of coupled electrons and nuclei. Unfortunately, challenges associated with this complexity limit not only mixed quantum-classical photodynamics simulations, but also mixed ML-classical photodynamics simulations. Therefore, the identification of key obstacles that prevent an accurate and efficient photodynamics simulation with ML-fitted coupled electronic excited states of different spin multiplicities, their derivatives and their properties was central to this work, and solutions to overcome these challenges were provided:

(1) The most severe problem with which we have been confronted, which also distinguishes excited-state properties from ground-state properties and impedes the use of conventional ML models, is the arbitrariness in the signs of properties resulting from different electronic states as a consequence of the arbitrary phase of the wave function. The need to guarantee a consistent phase during photodynamics simulations further complicates the fitting of such properties with ML. Two different approaches to tackle this task were developed: First, a phase correction algorithm makes inconsistent quantum chemical data learnable and allows for a reasonable and efficient training of excited-state properties with standard ML models. Second, an intrinsic ML-based solution to learn from inconsistent quantum chemical data was implemented in the new SchNarc approach developed in this thesis. The latter enables an efficient training set generation, which is practicable for many molecular systems and many excited electronic states. While the first approach suffers from an expensive training set generation, the second approach is more expensive with respect to the training of an ML model. Nevertheless, both methods are, to the best of our knowledge, the only available phase correction algorithms that do not require manual adjustment of data points and are therefore an important step to allow for an integration of ML into photochemical simulation methods. They further form the basis of all studies carried out in this thesis, without which our ML-based photodynamics simulation would not have been possible.

(2) Many electronic structure codes do not offer the calculation of couplings between states of same spin multiplicity, which prohibits their use as reference methods for the training set generation if they need to be described. Therefore, an approximation of such coupling values from ML potential energy surfaces, their first and second derivatives, was implemented into the SchNarc approach and the accuracy of the ML approximated coupling vectors was proven with photodynamics simulations of SO_2, $CH_2NH_2^+$ and CSH_2, demonstrating their validity for slow and ultrafast population transfer between different excited electronic states.

(3) Due to the lack of studies that focused on a description of the excited states with ML, many questions regarding the proper choice of an ML model and a molecular descriptor or the treatment of excited states in a multi-state or single-state fashion remained unanswered. To this aim, several ML models and molecular representations

were compared and results led to the following main conclusions: (i) As the excited states of a molecule are inherently linked to each other, a multi-state treatment, i.e., the description of all excited states in an ML model, is advantageous compared to a single-state treatment. (ii) Vectors that couple different electronic states should be represented in a rotationally covariant manner. This was achieved by the definition of virtual ML properties whose derivatives with respect to Cartesian coordinates reflect those coupling vectors. This approach, inspired by the underlying physics, can benefit from the inclusion of the energy gaps of the coupled potential energy surfaces, which must be fitted accurately for this purpose. (iii) At the current stage of research, the benefits of conventional kernel methods are their simplicity and ease of use. However, kernel methods turned out to be less accurate for photodynamics simulations, because they suffer from high memory consumption, which further restricts them to small training set sizes. In contrast, artificial neural networks are more flexible and can efficiently describe a manifold of excited states in one model, leading to more accurate potential energy surfaces. (iv) Besides the influence of the ML model on the accuracy of fitted potential energy surfaces, an precise and tailored molecular representation can further improve results, while reducing the number of necessary training points for subsequent photodynamics simulations. An optimal, ML-generated molecular representation can be provided by SchNet, which was adapted for the excited states and implemented in the SchNarc approach. The automatically created descriptor describes atoms in their chemical and structural environment and eliminates the time-consuming manual definition of a representation for molecules. Unfortunately, it is not yet known, whether the excited states can be described with atom-wise local contributions. The small test systems used in this thesis are not sufficiently large enough to answer this question. Further tests are needed to evaluate whether atom-wise local contributions can be used for excited-state potential energy surfaces, their derivatives, and properties thereof.

(4) In order to enhance photodynamics simulations with ML, a cost effective training set generation is required, which allows for a desired accuracy. Therefore, an efficient procedure originally developed for ground-state molecular dynamics simulations was adapted for the excited states to comprehensively sample the relevant conformational space of a molecule after photo-excitation. This scheme was used to generate training sets for photodynamics of $CH_2NH_2^+$ and CSH_2 with multi-reference methods. ML models trained on these training sets were sufficiently accurate to reproduce photodynamics simulations, while keeping the number of required reference data at a minimum.

All these novel strategies have been successfully applied to the calculation of the photodynamics of SO_2, $CH_2NH_2^+$ and CSH_2, allowing for excellent agreement to the quantum chemical reference methods. Statistically significant populations in the excited states could be obtained by efficient sampling of thousands of independent reaction events. In addition, artificial neural networks were used to simulate the photodynamics of $CH_2NH_2^+$ for one nanosecond. This simulation was achieved in 59 days on a single core, which would have required approximately 19 years of computation on the same core without the use of ML.

The amino acid tyrosine, which possesses a high density of electronic states, provides a perspective on the capabilities of ML approaches developed in this thesis and summarizes the challenges that still have to be overcome: A description of the photochemistry of many molecular systems requires precise multi-reference methods that can accurately treat reactions, such as homolytic bond breaking. However, multi-reference methods suffer from high computational costs and inconsistencies in

potential energy surfaces, which was exemplified for the excited states of tyrosine. Especially the large number of relevant and energetically close-lying excited electronic states, which need to be fitted with ML models, require smooth underlying potential energy surfaces, which are impracticable to calculate with high-level quantum chemistry methods, such as CASPT2. Single-reference methods are much better suited for the computation of a training set for ML. However, such methods cannot provide a physically correct picture in some conformational regions of molecules. It was shown that artificially generated data points can, in principle, correct single reference data in such conformational regions. Therefore, most parts of a training set were computed with an affordable single-reference method and homolytic bond breaking was described with artificially generated data, i.e., data points that were manipulated to describe the multi-reference behaviour in such conformational regions. Artificial neural networks could learn from this fused data set and interpolate the qualitatively correct behaviour in dissociative regions.

The strategies and methods developed in this thesis shed light on how ML models can be used to treat a variety and manifold of excited electronic states of molecules. The overall developments provide a framework for an efficient and accurate study of the photochemistry of molecules with ML and represent an important step toward improving existing photodynamics simulations. Nevertheless, much remains to be done and ML is not yet close to being a routine application in quantum chemistry. Probably the most interesting question that remains unsolved is whether excited-state potential energy surfaces can be described with local atomic contributions. Another question, which is still unanswered, is the transferability of ML models to describe the photochemistry of many molecular systems instead of only one, as it is currently feasible. Especially the latter requires a quantum chemistry method, with which a training set can be computed for many molecular systems under investigation. However, despite the impracticable full configuration interaction method, no method exists, which can accurately describe any molecular system.

The main bottleneck identified that prohibits a practicable application of ML models for photodynamics simulations of many molecules is therefore the underlying quantum chemical calculation. The unfavourable scaling of existing accurate multi-reference methods with the number of atoms and excited electronic states of a molecule limits their application to small molecular systems. It was demonstrated that if a quantum chemical method exists to provide accurate underlying potential energy surfaces, their derivatives, and properties thereof, ML can significantly advance conventional photodynamics methods and can go beyond the time scales of them. However, the development of quantum chemical reference methods for the excited states of larger systems comprising a complex photochemistry cannot keep up with the space of the extremely fast progressing ML models. The potential of ML models to overcome the existing limitations of quantum chemical methods for describing the photochemistry of molecules is particularly interesting in this respect and could provide an efficient solution to obtain a photochemical description of many molecular systems and for many applications.

A

In the following sections, the published and accepted articles are reprinted. In the publication enclosed in section A.1 the proof of concept that ML can advance photo-dynamics simulations is discussed along with the phase correction algorithm, the adaptive sampling scheme for excited states and photodynamics of the methylen-immonium cation for 1 ns. Section A.2 contains the publication, in which different molecular descirptors and NN and KRR models are compared. The state encoding for KRR models is part of this publication. The SchNarc approach for photodynamics simulations is introduced together with the phase-free training algorithm and the ML approximation to kinetic couplings in section A.3. A book chapter on ML for nonadiabatic MD simulations is enclosed in section A.4, which is scheduled for release on July 17, 2020. A just accepted Perspective on the same topic can be found in section A.5.

appendix a.1 machine learning enables long time scale molecular photodynamics simulations

Julia Westermayr, Michael Gastegger, Maximilian F. S. J. Menger, Sebsatian Mai, Leticia González and Philipp Marquetand

Contributions:

Julia Westermayr developed the multi-layer feed-forward neural networks and phase correction algorithm, implemented adaptive sampling for excited states, computed the training set, carried out the quantum chemical reference computations, evaluated the models, performed the production runs, and contributed to the initial draft and final manuscript.

Michael Gastegger helped in setting up the framework for the initial developed machine learning models and helped in identifying the assessment of the accuracy of different machine learning models.

Maximilian F. S. J. Menger developed the python wrapper for the SHARC molecular dynamics program and contributed to the interface of the developed machine learning model.

Sebastian Mai contributed to the discussions concerning the phase correction algorithm and the implementation of the machine learning model into the SHARC molecular dynamics code.

Leticia González helped in identifying the main goals of this study and contributed to the initial and final manuscript.

Philipp Marquetand conceived the main goals of this study, supervised methodological developments, computations, and data analysis, contributed to the code, and the writing of the initial and final manuscript.
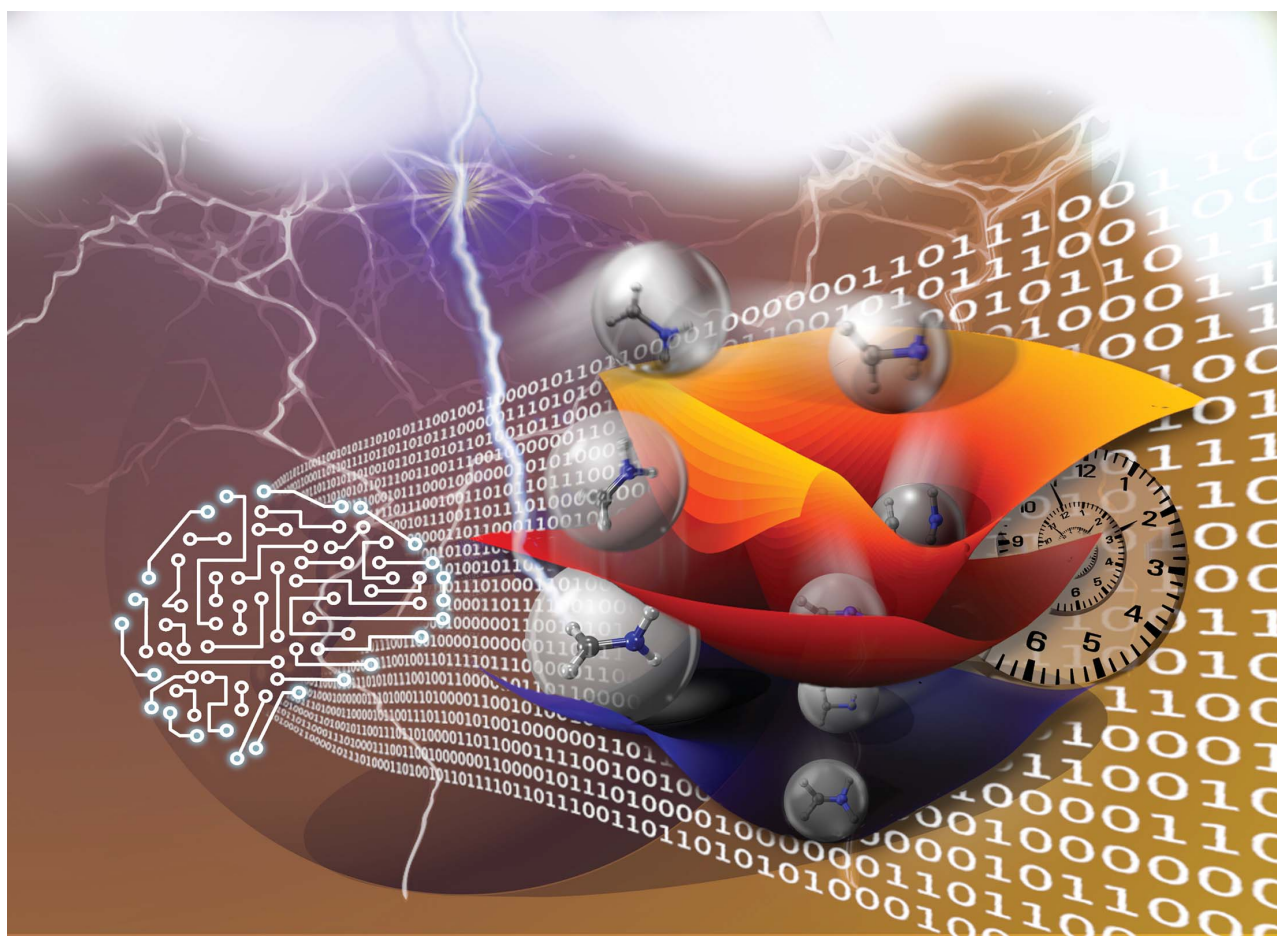
Showcasing research from Philipp Marquetand's laboratory, Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Vienna, Austria.

Machine learning enables long time scale molecular photodynamics simulations

The work illustrates how machine learning is used to perform nonadiabatic excited-state dynamics simulations on a nanosecond time scale. The test molecule is the methylenimmonium cation, for which a nanosecond simulation using an *ab initio* method would have taken approximately 19 years. With our presented method, which is based on neural networks that learn potentials and couplings from precise quantum chemical data, this dynamics simulation could be achieved within 59 days. It shows that machine learning algorithms can outperform standard photodynamics approaches in their computational performance while delivering the same accuracy.

As featured in:

See Philipp Marquetand *et al.*, *Chem. Sci.*, 2019, **10**, 8100.

ROYAL SOCIETY OF CHEMISTRY | Celebrating IYPT 2019

rsc.li/chemical-science

Registered charity number: 207890

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

EDGE ARTICLE

Check for updates

Machine learning enables long time scale molecular photodynamics simulations†

Julia Westermayr, [ID] [a] Michael Gastegger, [ID] [b] Maximilian F. S. J. Menger, [ac] Sebastian Mai, [ID] [a] Leticia González [ID] [a] and Philipp Marquetand [ID] *[a]

Photo-induced processes are fundamental in nature but accurate simulations of their dynamics are seriously limited by the cost of the underlying quantum chemical calculations, hampering their application for long time scales. Here we introduce a method based on machine learning to overcome this bottleneck and enable accurate photodynamics on nanosecond time scales, which are otherwise out of reach with contemporary approaches. Instead of expensive quantum chemistry during molecular dynamics simulations, we use deep neural networks to learn the relationship between a molecular geometry and its high-dimensional electronic properties. As an example, the time evolution of the methylenimmonium cation for one nanosecond is used to demonstrate that machine learning algorithms can outperform standard excited-state molecular dynamics approaches in their computational efficiency while delivering the same accuracy.

1 Introduction

Machine learning (ML) is revolutionizing the most diverse domains, like image recognition,[1] playing board games,[2] or social integration of refugees.[3] Also in chemistry, an increasing range of applications is being tackled with ML, for example, the design and discovery of new molecules and materials.[4–6] In the present study, we show how ML enables efficient photodynamics simulations. Photodynamics is the study of photo-induced processes that occur after a molecule is exposed to light. Photosynthesis and DNA photodamage leading to skin cancer are only two examples of phenomena that involve molecules interacting with light.[7–11] The simulation of such processes has been key to learning structure–dynamics–function relationships that can be used to guide the design of photonic materials, such as photosensitive drugs,[12] photo-catalysts[4] and photovoltaics.[13,14]

Computer simulations of photodynamics typically rely on molecular dynamics simulations of coupled nuclei and electrons. These simulations require the computation of high-dimensional potential energy surfaces (PESs), *i.e.*, the electronic energy levels of the molecule for all possible molecular configurations, using quantum chemistry. The calculation of these PESs is usually the most expensive part of

the dynamics simulations[15] and therefore, different approximations are necessary and ubiquitous. For the electronic ground state, the time-consuming quantum chemical calculations are often replaced with force fields[16] but no standard force fields are available to describe electronically excited states. Another drawback of most conventional force fields is their inability to describe the breaking and formation of chemical bonds. Recently, increasing effort has been devoted to ML potentials,[17,18] where an accurate representation of the ground state PES including bond breaking[19] and formation is promised.[16,20–32] Similarly, modified Shepard interpolation is used to construct PESs in low-dimensional systems and adapt them in out-of-confidence regions.[33,34] However, the problem of obtaining accurate full-dimensional PESs for excited states in order to simulate long time photodynamics has not been solved yet. A few studies have focused on the prediction of excited state dynamics as well as on excited-state properties such as spectral densities with ML.[35–44] The breakdown of the Born–Oppenheimer approximation, leading to critical regions in the coupled excited state PESs,[45] poses yet another obstacle to quantum chemistry (QC) and consequently also ML.[39–41] Among those critical regions are conical intersections (or state crossings), where two PESs get into close proximity. The underlying elements that become important in such areas are nonadiabatic couplings (or spin–orbit couplings). They induce non-radiative transitions between two electronic states of the same (or different) spin-multiplicities involving ultrafast rearrangements of both nuclei and electrons. These challenges led to the need for intermittent quantum chemistry calculations[39,40] or omittance of couplings between different PESs[41] in ML driven

[a]Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, 1090 Vienna, Austria. E-mail: philipp.marquetand@univie.ac.at

[b]Machine Learning Group, Technical University of Berlin, 10587 Berlin, Germany

[c]Dipartimento di Chimica e Chimica Industriale, University of Pisa, Via G. Moruzzi 13, 56124 Pisa, Italy

Edge Article

photodynamics. Hence long time photodynamics are still lacking and the possibility to additionally represent the aforementioned nonadiabatic derivative couplings between PESs fundamental to model photodynamics has not been demonstrated yet. Here we overcome all these different bottlenecks using deep neural networks (NNs) and achieve the simulation of photodynamics for long time scales. We expand on the idea of using ML to obtain potentials for electronic excited states, as well as arbitrary couplings within a framework that combines ML with trajectory surface hopping molecular dynamics (Fig. 1). Our ML approach is fully capable of describing all necessary properties for executing nonadiabatic excited-state molecular dynamics on the order of nanoseconds. These properties include electronic energies, gradients, spin–orbit couplings, nonadiabatic couplings, and dipole moments of molecules. Additionally, the underlying potentials and couplings can be used to optimize critical points of the configurational space, such as potential minima or crossing points, which are important for interpreting photochemical mechanisms.

Fig. 1 Schematic workflow of surface hopping molecular dynamics with deep NNs: the scheme starts from a set of initial quantum chemical calculations, which are pre-processed using a phase-correction algorithm and constitute an initial training set. Using this set, two deep NNs (NN1 and NN2) are trained and replace the quantum chemical calculations of energies (E) and gradients (G), nonadiabatic couplings (NACs), spin–orbit couplings (SOCs) and dipole moments (μ). The dynamics calculation starts with an input geometry, for which the two NNs provide all electronic quantities. If the outcomes of both NNs are sufficiently similar, the configurational space around this input geometry is adequately represented by the training set and the electronic quantities are used for a propagation time step. If not, the nuclear configuration is recomputed with quantum chemistry, phase corrected and included in the training set – a process referred to as adaptive sampling. The NNs are then re-trained and a new dynamics cycle is started.

2 Theoretical background

Nonadiabatic excited-state molecular dynamics simulations are carried out using the Surface Hopping including ARbitrary Couplings (SHARC) method,[46] which is an extension of the fewest switches surface hopping method of Tully.[47] Within surface hopping, the nuclei are propagated according to the classical equations of motion and the electrons are treated quantum mechanically *via* interfaces to external electronic structure program packages. The electronic structure calculations are carried out on-the-fly at the nuclear geometries visited by the classical trajectories. The probability of a molecular system occupying a specific electronic state and population transfer between the different electronic states – in the form of stochastic, instantaneous hops from one electronic state to another – are dependent on the couplings between them.

2.1 Surface hopping molecular dynamics with deep NNs

For surface hopping simulations with NNs, the idea of retrieving electronic properties from an external source stays the same, but instead of a quantum chemical calculation, NNs are used to predict energies, gradients, couplings and dipole moments. The relationships between the nuclear coordinates and the corresponding electronic properties are learned from a training set, in which each data point is one set of nuclear coordinates and its associated set of quantities computed with a reference method. In order to make the procedure usable, the processes for generating NN potentials and their use in photodynamics simulations have been automated in a development version of the program suite SHARC.[15,46,48]

2.2 Training set generation and adaptive sampling for excited states

The combination of quantum chemistry with ML requires a cost-effective generation of a training set that, while it samples the conformational space of a molecular system comprehensively, is small enough to keep demanding quantum chemical reference calculations feasible.[27] With this in mind, we employ an initial training set based on normal mode scans and then switch to an adaptive sampling scheme[21,24,49–51] that automatically identifies untrustworthy regions not covered by the initial training set. The adaptive sampling procedure employs excited-state dynamics simulations using two or more NNs that are independently trained from the same training set. At every time step, the root mean squared error (RMSE) between the predictions of the different NNs of each property is compared to a predefined threshold. A separate threshold is set for each property (initially based on the validation error of the respective NN). Whenever any one of the thresholds is exceeded, *i.e.*, the different NNs make very different predictions, the corresponding geometry is assumed to lie in a conformational region with too few training points, even if the rest of the properties are predicted reliably. It is then necessary to expand the training set by computing the quantum chemistry data for this geometry. Along a dynamics run, the threshold for the error between predictions made by the NNs is adapted by multiplication with
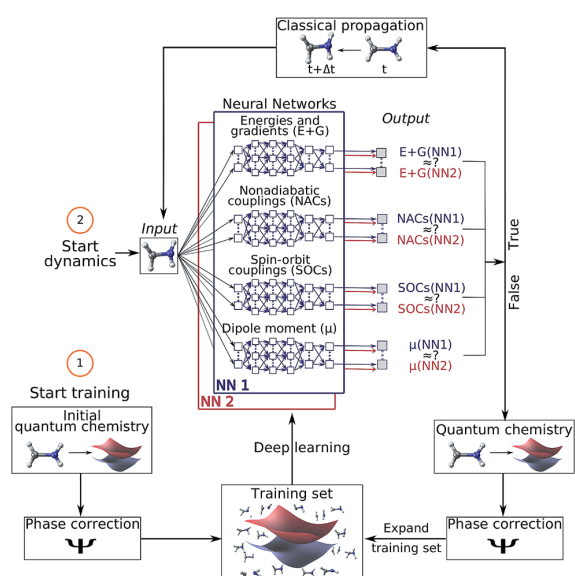
Chemical Science

a factor of 0.95 until the conformational space is sampled sufficiently to make accurate predictions without any additional reference calculations.

An ensemble of two NNs is used not only during the initial adaptive sampling period, but also for the production dynamics simulations in order to check the accuracy of the NN predictions and to discover undersampled regions of conformational space. After 10 ps, the threshold for the RMSE between NN forecasts is not reduced anymore but kept at the previous value when a new data point is added to the training set and NNs are retrained. More details on criteria for the thresholds and iterations are discussed in the ESI.†

2.3 Multi-layer feed-forward NNs

For the sake of making predictions of the quantum chemical properties of molecules, multi-layer feed forward NNs are applied.[49] For training of NNs, we use as input the matrix of inverse distances in order to achieve translational and rotational invariance in the relations established between the predicted properties and the nuclear coordinates. For prediction we use two similarly accurate NNs, with their optimal-network-architecture identified by random grid search[1] of (hyper) parameters. Additional information on network parameters and specifications can be found in Table S1 and Section S1 in the ESI along with NN convergence during training in Fig. S1.† We assessed the quality of the used NNs by comparing them with different ML models and NNs using a different molecular descriptor on an additionally generated test set (see Section S1.3 in the ESI†). Different ML models or descriptors do not lead to a considerable improvement of the accuracy. As a different ML model we choose support vector machine for regression and linear regression as a baseline model, but our NN approaches outperform these regression models. Furthermore, the performance of our NNs is presented in Table S5† for each electronic state, separately. In this context, it is shown how the tendency towards smooth interpolation of the ML models can even correct for discontinuities present in the QC1 method (see Fig. S2†), which demonstrates the utility of our approach.

Quantum chemical properties that were learned with NNs are energies, gradients, permanent as well as transition dipole moments, and NACs. Other quantities like spin–orbit couplings can also be trained (see the analytical model in the ESI†). Although the (transition) dipole moments are not needed for the present dynamics simulation, calculating them on-the-fly enables the computation of pump–probe schemes, static-field interactions, or time-resolved spectra (see for example ref. 52 and 53). While energies are directly used for training purposes in a single NN, forces are predicted as analytical derivatives of the NNs,[54] ensuring energy conservation.[24,32,39] Similarly, permanent dipole moments are directly used in the training. However, couplings (as well as transition dipole moments) need to be pre-processed as they are computed from the wave functions of two different electronic states and therefore depend on the relative phases of these two wave functions. Phase inconsistencies need to be eliminated in order to avoid ill-behaved photodynamics,[55] as is described in the following subsection.

2.4 Phase correction

Electronic wave functions computed with quantum chemistry programs are usually obtained as the eigenfunctions of the electronic Hamiltonian. However, this requirement does not uniquely define an electronic wave function because multiplying it by a phase factor still returns a valid eigenfunction. Thus, in practice two wave functions computed for two very similar geometries might randomly differ in their phase factor. This problem is best visualized using molecular orbitals (see Fig. 2). For different single point calculations along an interpolation coordinate (Fig. 2A), orbitals can arbitrarily switch their sign (illustrated by their color in Fig. 2B) and so does the complete electronic wave function. As energies are obtained from diagonal elements of the general form $\langle \Psi_i | \hat{O} | \Psi_i \rangle$ in matrix notation, the electronic wave function enters twice and any phase is squared, thus canceling out. However, off-diagonal elements, $\langle \Psi_i | \hat{O} | \Psi_j \rangle$, such as couplings involve the wave functions of two different electronic states and different phases do not necessarily cancel out. The example of Fig. 2B shows how the curves of such off-diagonal properties can be discontinuous, impeding correct learning behavior in the NN. It is thus mandatory to track the phases of all wave functions from one reference geometry to every other data point in the training set and apply a phase-correction algorithm that provides smooth
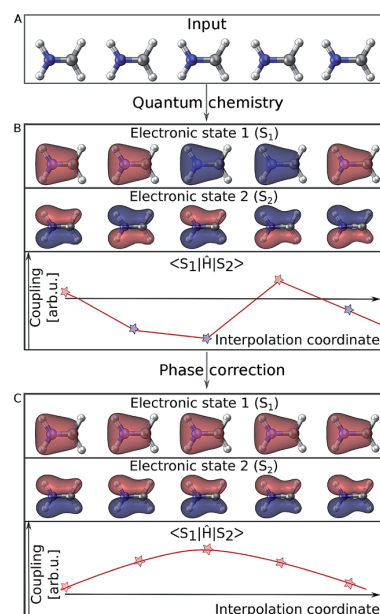
Fig. 2 Molecular orbitals representing two different electronic states of the methylenimmonium cation, $CH_2NH_2^+$. Panel A shows molecular geometries (with slightly different bond lengths) that are given as an input to a quantum chemistry program. The results for properties corresponding to off-diagonal matrix elements of the Hamiltonian are shown in panel B. Random signs are obtained due to random assignments of the phases of the involved wave functions. As can be seen in panel C, these random switches can be removed by phase correction, and smooth relations between a molecular geometry and any property can be found.

84

Edge Article

curves (Fig. 2C). In this way, a virtual global phase convention is applied to all data points within the training set, with the only aim of ensuring correct NN training.

Such a global phase convention is not mathematically possible for general polyatomic molecules due to the existence of the so-called Berry (or geometric) phase.[56] Due to the latter, the phase depends on the path between a given geometry and the reference point.[57] Still the above phase correction is advantageous because it removes phase jumps from almost all parts of configurational space. This is critically necessary to make the data learnable. Only the non-removable phase jumps from the Berry phase remain, but occupy a small volume of configurational space. Hence, our phase correction is assumed to leave the dynamics mostly unaffected. For instance, successful surface hopping algorithms without phase tracking, such as the Zhu–Nakamura theory,[58,59] exist and substantiate the validity of this approximation. In the case of the Zhu–Nakamura theory, dynamics are comparable to conventional surface-hopping molecular dynamics simulations propagated from NAC vectors.[59–63] Note that the approximated phase correction for generation of the training set above cannot be circumvented by learning the absolute value of couplings since the relative sign between nonadiabatic coupling vectors of each atom in the x, y and z directions should be retained.

In order to make off-diagonal elements learnable for ML models, phases are tracked by computing wave function overlaps between adjacent molecular geometries.[55,64,65] If the geometries are close enough, the overlaps will be sufficiently large and contain values close to +1 or −1, allowing a detection of phase changes. In cases where molecular geometries are too far apart, the overlap will generally be close to zero, offering no information about a phase change. In this case, we resort to interpolation between the two molecular geometries and iterative computation of wave function overlaps. In principle, the interpolation can be carried out between the new geometry and any geometry already inside the training set as long as the wavefunction of this previous geometry is stored. Storing the wavefunctions for at least a few geometries and identifying the most suitable one for interpolation *via* root mean square deviations of the geometry should be considered for larger and more flexible molecules.

Especially for large molecules, where many states lie close in energy, the so-called "intruder states" might become problematic. Such states are excluded at the reference geometry, but are included at another geometry due to an energy change, thus leading to small overlaps for the phase tracking algorithm. In such situations, different possibilities for adapting the phase correction algorithm should be considered. For instance, additional electronic states could be computed with QC. Those should not be included in the training data, but only used to continuously track the phase of all relevant states. This process then still stays affordable, since the additional states do not require a computation of gradients or couplings and do not have to be considered further. Additional details on the phase correction algorithm are given in Section S2 in the ESI.†

3 Computational details

The photodynamics simulations have been carried out with a development version of the program suite SHARC.[15,48] Besides the newly developed modules for NN training and prediction, this development version also employs the pySHARC Python wrapper for the SHARC dynamics driver. This wrapper enables communication between the driver and the NN code without any file I/O and thus reduces the runtime of the program substantially.

The reference quantum chemical computations were carried out with COLUMBUS[66] using the accurate multi-reference configuration interaction method including single and double excitations and a double-zeta basis set (abbreviated to MR-CISD/aug-cc-pVDZ and in the following sections labelled as QC1). For comparison, we carried out quantum chemical computations with another basis set, 6-31++G**, using the same MR-CISD method (abbreviated to QC2 in the following sections). NNs were implemented in Python using the numpy[67] and theano[68] packages. They were trained on energies, forces, dipole moments and nonadiabatic couplings, obtained with the QC1 method using the adaptive sampling scheme described above, resulting in about 4000 data points (mean absolute error (MAE) energies among all states: 0.032 eV $\widehat{=}$ 0.73 kcal mol^{-1}; MAE forces among all states: 0.51 eVÅ^{-1} $\widehat{=}$ 11.9 kcal mol^{-1} Å^{-1}; see also Tables S2, S4 and S5 in the ESI as well as Fig. S2† for analysis of different states). Using each method, QC1, QC2, and NNs trained on QC1, we simulated the dynamics of the methylenimmonium cation after excitation from the electronic ground state (S_0) to the second excited electronic state (S_2) over 100 fs using a time step of 0.5 fs.

Optimizations of minima were carried out with the SHARC tools that utilize an external optimizer, ORCA,[69] where the computed energies and gradients[70,71] from the NNs were fed in or those from COLUMBUS for comparison.

4 Results and discussion

First, a one-dimensional model was employed to test our deep learning molecular dynamics approach (see Fig. S3 in Section S3 in the ESI†). In the following, the performance of the method is demonstrated by simulating the full-dimensional photodynamics of the methylenimmonium cation, $CH_2NH_2^+$ – the simplest member of the protonated Schiff bases. Methylenimmonium has been reported to undergo ultrafast switches between different electronic states after excitation with light.[72] A larger member of this family is retinal, which is fundamental for vision[73] but the methylenimmonium cation is an ideal testbed to demonstrate the applicability of NNs in photodynamics, because it is small enough to perform accurate reference photodynamics simulations for short time scales for comparison.

4.1 Nanosecond molecular dynamics simulation

Our NNs were trained on data obtained using the QC1 method (see details on active space in Section S4 and Fig. S4 in the ESI†).

Chemical Science

Independently with the QC1 method and with NNs, we simulated the dynamics of the methylenimmonium cation after excitation to the second excited singlet state, S_2. As can be seen from Fig. 3A, fast population transfer from the S_2 state to the first excited singlet state, S_1, and back to the ground state, S_0, takes place. The population dynamics obtained with the NN potentials and that obtained using the QC1 method agree very well. These results are also in good agreement with the literature.[72] Both methods describe the deactivation to the ground state, S_0, through the correct conical intersections, as will be discussed in the next subsection. A movie of one trajectory over 100 fs along with the NN potential energy curves is part of the ESI Movie S1.†

One of the first advantages of the NN driven dynamics simulations is that due to their very low computational cost, a much larger number of trajectories (3846) was simulated than what is typically possible with standard quantum chemistry (90). This enlarged statistics provides smooth population curves for the NN simulations (a comparison of the curves with an identical number of trajectories for NNs and QC1 can be found in Fig. S5A in the ESI along with analysis of energy conservation in Table S11†).

In order to estimate the magnitude of the error obtained with the NNs, we carried out a second *ab initio* molecular dynamics study with an additional, very similar, quantum chemistry method where only the double-zeta basis set is changed from aug-cc-pVDZ to 6-31++G**. As Fig. 3B shows, the differences between the two levels of theory are of the same order of magnitude as those encountered between NNs and quantum chemistry, indicating that the agreement between the methods is very good. The MAE in population between QC1 and NNs is 0.057 and between QC1 and QC2 it is 0.099. Time constants derived from dynamics with each method also agree well. The time constant from S_2 to S_1 is 18.3 fs according to the QC1 method, which is comparable to the QC2 method with 25.0 fs and to NN driven dynamics with 25.2 fs. The time constant obtained for transitions from S_1 to S_0 is 51.0 fs for the QC1 method, which is very similar to the value obtained with NNs

(52.6 fs), whereas the QC2 method yields a time constant of 73.2 fs.

After nonadiabatic dynamics using deep NNs has been validated for short time scales, we show the major advantage of the method, *i.e.* that it is able to overcome the problem of limited simulation time and predict long excited-state dynamics. Fig. 4 shows the population dynamics of the methylenimmonium cation on a logarithmic scale up to 1 nanosecond (ns), *i.e.*, 10^4 times longer than they were simulated using our quantum chemical reference method. Up to 10 ps, we simulated an ensemble of 200 trajectories with 2 NNs using the adaptive sampling scheme described above in order to correctly predict events not yet learned by the NNs. After that, 2 trajectories are propagated up to 1 ns for demonstration purposes using 2 NNs. The populations are thus averaged over 200 trajectories up to 10 ps and over 2 trajectories from 10 ps up to 1 ns, respectively. As can be seen, the molecule relaxes to the ground state after around 300 fs. Due to the remaining kinetic energy a few hops between different states are recorded and can be regarded as noise. A movie of one trajectory over 10 ps is part of the ESI Movie S2.†

The propagation of a $CH_2NH_2^+$ trajectory for 10 ps can be executed in less than 6 hours on one core, which is 300 times faster than the calculation with the quantum chemical reference method. The propagation of 1 ns took 59 days employing two deep NNs serially, whereas an estimated 19 years of computation would have been required with the quantum chemical reference.

4.2 Conical intersections obtained from NNs

Since NNs can provide energies, gradients, and couplings, they can also be used to optimize important points of the PES, like state minima or conical intersections. The identification of conical intersections is the target of many quantum chemical studies as they are commonly deemed as the most probable geometries for radiationless transitions between electronic states of the same spin multiplicity. Due to their special topology with discontinuous first derivatives, the surroundings of a conical intersection pose serious challenges to the NN
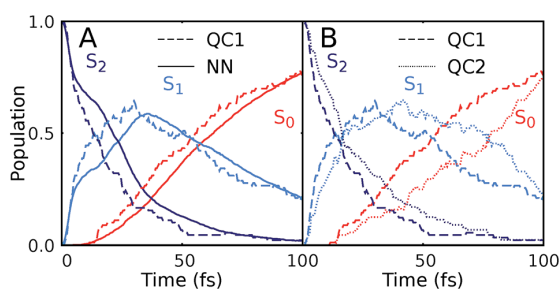
Fig. 3 Population dynamics of $CH_2NH_2^+$ based on deep NNs and traditional quantum chemistry: comparison between results obtained from (A) QC1 (90 trajectories) and NN (3846 trajectories) and (B) QC1 (90 trajectories) and QC2 (88 trajectories). For completeness, the populations from 90 trajectories propagated with NNs are given in Fig. S5A in the ESI along with geometrical analysis along the trajectories in Fig. S5B.†
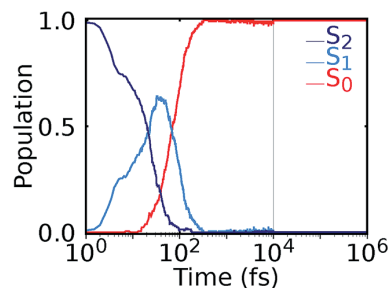
Fig. 4 Nonadiabatic molecular dynamics simulations using deep NNs for one nanosecond. After excitation to the S_2 state, ultrafast internal conversion to the S_1 state takes place, followed by recovery of the S_0 state within 300 fs. Until 10 ps, an ensemble of 200 trajectories is analyzed, followed by the population averaged from 2 trajectories.

training.[45] As the photodynamics critically depends on a correct representation of these surroundings, here we perform some tests to validate their accuracy.

To this aim we optimize two minimum energy conical intersections in $CH_2NH_2^+$, one between the S_2 state and the S_1 state and another one between the S_1 state and the S_0 state. We use the QC1 method and NNs to perform potential energy scans around the minimum energy conical intersections optimized at the QC1 level of theory. As can be seen from Fig. 5A–D, typical curved seams of conical intersections between the S_2 and S_1 states (Fig. 5A (QC1) and 5B (NN)) and the S_1 and S_0 states (Fig. 5C (QC1) and 5D (NN)) are obtained around the minimum energy conical intersections.[74] The NNs get the shape of this seam correct with slightly larger energy gaps between the crossing surfaces due to the fact that NN potentials need to be differentiable at any point. Analysis of 408 (for the S_1/S_0 CI) and 302 (for the S_2/S_1 CI) configurations around the minimum energy conical intersections – identified by an energy gap smaller than 0.8 eV according to the QC1 method – showed that on average, the gaps are overestimated by 0.068 eV for S_1/S_0 and by 0.014 eV for S_2/S_1 by our NNs. As can be seen from Fig. 5, the potentials around the S_1/S_0 CI are flatter than the potentials around the S_2/S_1 CI, indicating that hopping geometries are closer to the CI in the latter case and that the molecules can also hop farther from the CI in the former case.

Fig. 6 shows the scatter plots of the optimized geometries of the minimum energy conical intersections projected along two important coordinates together with the hopping geometries and the geometries contained in the training set. As can be seen, the hopping geometries between the S_2 and S_1 states are mainly located close to the optimized geometry of the minimum energy conical intersection, while the hopping geometries in the case of the S_1/S_0 crossing are more widely distributed around the optimized geometry. As a consequence, the S_2/S_1 crossing is sampled more comprehensively, since more

Fig. 6 Scatter plots showing the distribution of hopping geometries obtained with QC1, QC2, and NN as well as optimized S_1/S_0 (A) and S_2/S_1 (B) minimum energy conical intersections (CIs) along with the geometries that make up the training set with 4000 data points. The actual geometry is depicted on top (geometrical parameters are given in Fig. S7B†). A zoom-in of the regions near the optimized points is shown in Fig. S7A in the ESI† together with a definition of the dihedral and pyramidalization angles.

trajectories pass by near the minimum energy conical intersection. This observation also explains the larger NN energy gap obtained for the second crossing, the S_1/S_0 CI, in Fig. 5.

The optimizations of the minimum energy conical intersections were independently performed with the trained NN, as well as with the QC1 and QC2 methods for comparison. The optimized molecular geometries (shown in Fig. S6 along with Cartesian coordinates in the ESI†) agree well. As can be seen, the driving force for the transition from the S_2 state to the S_1 state is an elongation of the C–N bond in combination with a bipyramidalization. The torsion of the molecule further leads to internal conversion to the ground state, S_0. Additionally, each method results in a comparable distribution of hopping geometries around the optimized points, which in practice is of utmost importance[75] for describing the population transfer in the simulations correctly. There are very few NN hopping geometries at either large pyramidalization angles (S_1/S_0 CI) or long C–N bonds (S_2/S_1 CI), compared to the QC trajectories. This finding correlates with the distribution of training set geometries, which are also absent in these regions of the PES (see the grey circles in Fig. 6). Configurations obtained via sampling of normal modes are clearly visible by a dense alignment of data points. However, the configurations obtained via adaptive sampling are mostly centered in the middle of the plot for the S_1/S_0 CI and close to the optimized CI for the S_2/S_1 crossing, explaining the smaller distribution of NN hopping geometries. Further analysis showed that geometries at large bond lengths
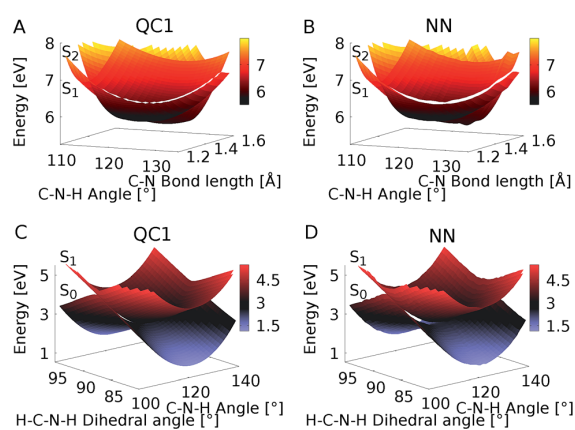

Fig. 5 Potential energy scans around the minimum energy conical intersections obtained with QC1 of the S_2 and S_1 states (A and B) and S_1 and S_0 states (C and D). Panels A and C show the PESs calculated with QC1, while panels B and D illustrate NN potentials. See the caption of Fig. S7 in the ESI† for clarification of the dihedral angle.
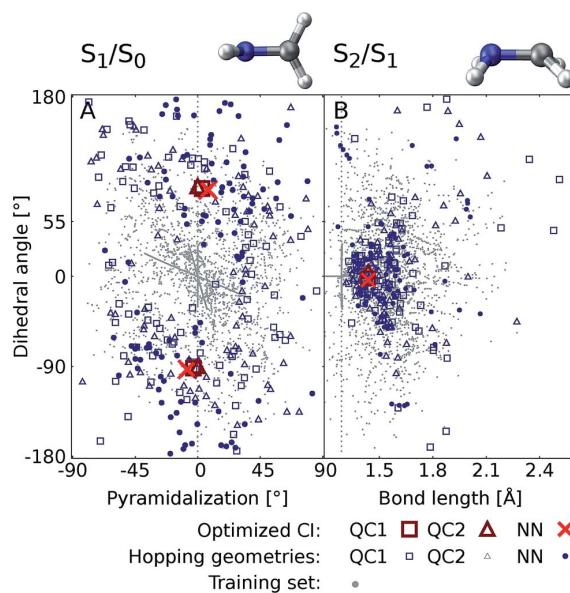
are approximately 4 eV higher in energy than the geometries close to the optimized minimum energy conical intersection in the case of the S_2/S_1 crossing. Therefore, trajectories simulated during adaptive sampling probably did not visit those regions of the PES. In the case of the S_1/S_0 crossing, this effect is less pronounced and the geometries with a large pyramidalization angle are approximately 1–1.5 eV larger in energy than the configurations close to the optimized CI, indicating again the much flatter potential.

5 Conclusions

We demonstrate that deep NNs are able to accelerate nonadiabatic excited-state molecular dynamics simulations by orders of magnitude, thus overcoming the constraints of limited time scales and limited statistics. Our approach offers an automatic learning procedure by implementation of adaptive sampling for excited states, which opens new avenues for studying the photodynamics of complex systems on long time scales relevant for chemistry, biology, medicine, and materials design, for which the PESs cannot be explored in advance with conventional *ab initio* techniques. Offering access to the precision of high-level quantum chemistry methods at only a fraction of the original computational cost, we expect this setup to become a powerful tool in several research fields.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

References

1 I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.

2 D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel and D. Hassabis, *Nature*, 2016, **529**, 484–489.

3 K. Bansak, J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence and J. Weinstein, *Science*, 2018, **359**, 325–329.

4 B. Sanchez-Lengeling and A. Aspuru-Guzik, *Science*, 2018, **361**, 360–365.

5 K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Nature*, 2018, **559**, 547–555.

6 B. R. Goldsmith, J. Esterhuizen, J.-X. Liu, C. J. Bartel and C. Sutton, *AIChE J.*, 2018, **64**, 2311–2323.

7 G. Cerullo, D. Polli, G. Lanzani, S. De Silvestri, H. Hashimoto and R. J. Cogdell, *Science*, 2002, **298**, 2395–2398.

8 T. Schultz, E. Samoylova, W. Radloff, I. V. Hertel, A. L. Sobolewski and W. Domcke, *Science*, 2004, **306**, 1765–1768.

9 W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo-Hernández, V. N. Swaminathan, T. Charell, W. Zinth and B. Kohler, *Science*, 2007, **315**, 625–629.

10 C. Rauer, J. J. Nogueira, P. Marquetand and L. González, *J. Am. Chem. Soc.*, 2016, **138**, 15911–15916.

11 E. Romero, V. I. Novoderezhkin and R. v. Grondelle, *Nature*, 2017, **543**, 355.

12 I. Ahmad, S. Ahmed, Z. Anwar, M. A. Sheraz and M. Sikorski, *Int. J. Photoenergy*, 2016, **2016**, 1–19.

13 S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B. F. E. Curchod, N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M. K. Nazeeruddin and M. Grätzel, *Nat. Chem.*, 2014, **6**, 242.

14 A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi and M. Ceriotti, *Sci. Adv.*, 2017, **3**, e1701816.

15 S. Mai, P. Marquetand and L. González, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, e1370.

16 S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, *Nat. Commun.*, 2018, **9**, 3887.

17 M. Rupp, *Int. J. Quantum Chem.*, 2015, **115**, 1058–1073.

18 O. A. von Lilienfeld, *Angew. Chem., Int. Ed.*, 2018, **57**, 4164–4169.

19 F. Häse, I. F. Galván, A. Aspuru-Guzik, R. Lindh and M. Vacher, *Chem. Sci.*, 2019, **10**, 2298–2307.

20 A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*, 2010, **104**, 136403.

21 Z. Li, J. R. Kermode and A. De Vita, *Phys. Rev. Lett.*, 2015, **114**, 096405.

22 M. Rupp, R. Ramakrishnan and O. A. von Lilienfeld, *J. Phys. Chem. Lett.*, 2015, **6**, 3309–3313.

23 J. Behler, *J. Chem. Phys.*, 2016, **145**, 170901.

24 M. Gastegger, J. Behler and P. Marquetand, *Chem. Sci.*, 2017, **8**, 6924–6935.

25 V. Botu, R. Batra, J. Chapman and R. Ramprasad, *J. Phys. Chem. C*, 2017, **121**, 511–522.

26 J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203.

27 J. Behler, *Angew. Chem., Int. Ed.*, 2017, **56**, 12828–12840.

28 H. Zong, G. Pilania, X. Ding, G. J. Ackland and T. Lookman, *npj Comput. Mater.*, 2018, **4**, 48.

29 A. P. Bartók, J. Kermode, N. Bernstein and G. Csányi, *Phys. Rev. X*, 2018, **8**, 041048.

30 R. Xia and S. Kais, *Nat. Commun.*, 2018, **9**, 4195.

31 H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Y. Chan and

Edge Article

S. K. R. S. Sankaranarayanan, *J. Phys. Chem. C*, 2019, **123**, 6941.

32 A. S. Christensen, F. A. Faber and O. A. von Lilienfeld, *J. Chem. Phys.*, 2019, **150**, 064105.

33 H. M. Netzloff, M. A. Collins and M. S. Gordon, *J. Chem. Phys.*, 2006, **124**, 154104.

34 R. P. A. Bettens and M. A. Collins, *J. Chem. Phys.*, 1999, **111**, 816–826.

35 J. Behler, K. Reuter and M. Scheffler, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2008, **77**, 115421.

36 C. Carbogno, J. Behler, K. Reuter and A. Groß, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2010, **81**, 035410.

37 F. Häse, S. Valleau, E. Pyzer-Knapp and A. Aspuru-Guzik, *Chem. Sci.*, 2016, **7**, 5139–5147.

38 F. Liu, L. Du, D. Zhang and J. Gao, *Sci. Rep.*, 2017, **7**, 1–12.

39 D. Hu, Y. Xie, X. Li, L. Li and Z. Lan, *J. Phys. Chem. Lett.*, 2018, **9**, 2725–2732.

40 P. O. Dral, M. Barbatti and W. Thiel, *J. Phys. Chem. Lett.*, 2018, **9**, 5660–5663.

41 W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral and G. Cui, *J. Phys. Chem. Lett.*, 2018, **9**, 6702–6708.

42 D. M. G. Williams and W. Eisfeld, *J. Chem. Phys.*, 2018, **149**, 204106.

43 C. Xie, X. Zhu, D. R. Yarkony and H. Guo, *J. Chem. Phys.*, 2018, **149**, 144107.

44 Y. Guan, D. H. Zhang, H. Guo and D. R. Yarkony, *Phys. Chem. Chem. Phys.*, 2019, **21**, 14205–14213.

45 W. Domcke, D. R. Yarkony and H. Köppel, *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*, WORLD SCIENTIFIC, 2004.

46 M. Richter, P. Marquetand, J. González-Vázquez, I. Sola and L. González, *J. Chem. Theory Comput.*, 2011, **7**, 1253–1258.

47 J. C. Tully, *J. Chem. Phys.*, 1990, **93**, 1061–1071.

48 S. Mai, M. Richter, M. Ruckenbauer, M. Oppel, P. Marquetand and L. González, *SHARC2.0: Surface Hopping Including Arbitrary Couplings – Program Package for Non-Adiabatic Dynamics*, sharc-md.org, 2018.

49 J. Behler, *Int. J. Quantum Chem.*, 2015, **115**, 1032–1050.

50 V. Botu and R. Ramprasad, *Int. J. Quantum Chem.*, 2015, **115**, 1074–1083.

51 J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. Roitberg, *J. Chem. Phys.*, 2018, **148**, 241733.

52 P. Marquetand, A. Materny, N. E. Henriksen and V. Engel, *J. Chem. Phys.*, 2004, **120**, 5871–5874.

53 F. P. Bonafé, F. J. Hernández, B. Aradi, T. Frauenheim and C. G. Sánchez, *J. Phys. Chem. Lett.*, 2018, **9**, 4355–4359.

54 M. Gastegger and P. Marquetand, *J. Chem. Theory Comput.*, 2015, **11**, 2187–2198.

55 A. V. Akimov, *J. Phys. Chem. Lett.*, 2018, **9**, 6096–6102.

56 I. G. Ryabinkin, L. Joubert-Doriol and A. F. Izmaylov, *Acc. Chem. Res.*, 2017, **50**, 1785–1793.

57 S. Matsika and P. Krause, *Annu. Rev. Phys. Chem.*, 2011, **62**, 621–643.

58 P. Oloyede, G. Mil'nikov and H. Nakamura, *J. Chem. Phys.*, 2006, **124**, 144110.

59 T. Ishida, S. Nanbu and H. Nakamura, *Int. Rev. Phys. Chem.*, 2017, **36**, 229–286.

60 C. Zhu, H. Kamisaka and H. Nakamura, *J. Chem. Phys.*, 2002, **116**, 3234–3247.

61 T. Ishida, S. Nanbu and H. Nakamura, *J. Phys. Chem. A*, 2009, **113**, 4356–4366.

62 A.-H. Gao, B. Li, P.-Y. Zhang and K.-L. Han, *J. Chem. Phys.*, 2012, **137**, 204305.

63 L. Yu, C. Xu, Y. Lei, C. Zhu and Z. Wen, *Phys. Chem. Chem. Phys.*, 2014, **16**, 25883–25895.

64 S. Mai, P. Marquetand and L. González, *Int. J. Quantum Chem.*, 2015, **115**, 1215–1231.

65 F. Plasser, M. Ruckenbauer, S. Mai, M. Oppel, P. Marquetand and L. González, *J. Chem. Theory Comput.*, 2016, **12**, 1207.

66 H. Lischka, R. Shepard, R. M. Pitzer, I. Shavitt, M. Dallos, T. Müller, P. G. Szalay, M. Seth, G. S. Kedziora, S. Yabushita and Z. Zhang, *Phys. Chem. Chem. Phys.*, 2001, **3**, 664–673.

67 S. van der Walt, S. C. Colbert and G. Varoquaux, *Comput. Sci. Eng.*, 2011, **13**, 22–30.

68 Theano Development Team, arXiv 1605.02688 [cs.SC], 2016.

69 F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 73–78.

70 B. G. Levine, J. D. Coe and T. J. Martinez, *J. Phys. Chem. B*, 2007, **112**, 405–413.

71 M. J. Bearpark, M. A. Robb and H. B. Schlegel, *Chem. Phys. Lett.*, 1994, **223**, 269.

72 M. Barbatti, A. J. A. Aquino and H. Lischka, *Mol. Phys.*, 2006, **104**, 1053–1060.

73 J. Herbst, K. Heyne and R. Diller, *Science*, 2002, **297**, 822–825.

74 D. R. Yarkony, *J. Chem. Phys.*, 2005, **123**, 204101.

75 H. R. Hudock, B. G. Levine, A. L. Thompson, H. Satzger, D. Townsend, N. Gador, S. Ullrich, A. Stolow and T. J. Martínez, *J. Phys. Chem. A*, 2007, **111**, 8500–8508.

89

APPENDIX A.2 NEURAL NETWORKS AND KERNEL RIDGE REGRESSION FOR EXCITED STATES DYNAMICS OF $CH_2NH_2^+$: FROM SINGLE-STATE TO MULTI-STATE REPRESENTATIONS AND MULTI-PROPERTY MACHINE LEARNING MODELS

Julia Westermayr, Felix Faber, Anders S. Christensen, O Anatole von Lilienfeld and Philipp Marquetand

Contributions:

Julia Westermayr implemented the state representations, carried out the reference calculations, trained and evaluated the machine learning, and wrote the initial manuscript.

Felix Faber discussed the state representation and possibilities to assess the quality of the machine learning models, i.e., the scatter plots and learning curves.

Anders S. Christensen helped to implement the FCHL representation for neural networks and the response formalism for kernel ridge regression models.

O Anatole von Lilienfeld was supervising this study and conceived the goals of this work. He contributed to the manuscript in its initial and final stages.

Philipp Marquetand conceived the scope of this work and supervised the practical part of the study, i.e., the cross-validation and surface-hopping dynamics simulations. He contributed to the manuscript in its initial and final stages.

Machine Learning: Science and Technology

PAPER • OPEN ACCESS

Neural networks and kernel ridge regression for excited states dynamics of $CH_2NH_2^+$: From single-state to multi-state representations and multi-property machine learning models

To cite this article: Julia Westermayr *et al* 2020 *Mach. Learn.: Sci. Technol.* **1** 025009

View the article online for updates and enhancements.

Recent citations

- Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics
Julia Westermayr *et al*

MACHINE
LEARNING
Science and Technology

PAPER

Neural networks and kernel ridge regression for excited states dynamics of $CH_2NH_2^+$: From single-state to multi-state representations and multi-property machine learning models

Julia Westermayr[1] , Felix A Faber[2], Anders S Christensen[2], O Anatole von Lilienfeld[2] and Philipp Marquetand[1,3,4]

[1] University of Vienna, Faculty of Chemistry, Institute of Theoretical Chemistry, Währinger Str. 17, 1090 Wien, Austria
[2] Institute of Physical Chemistry and National Center for Computational Design and Discovery of Novel Materials (MARVEL), Department of Chemistry, University of Basel, Klingelbergstr. 80, CH-4056 Basel, Switzerland
[3] Vienna Research Platform on Accelerating Photoreaction Discovery, University of Vienna, Währinger Str. 17, 1090 Wien, Austria
[4] University of Vienna, Faculty of Chemistry, Data Science @ Uni Vienna, Währinger Str. 29, 1090 Wien, Austria

E-mail: anatole.vonlilienfeld@unibas.ch and philipp.marquetand@univie.ac.at

Abstract

Excited-state dynamics simulations are a powerful tool to investigate photo-induced reactions of molecules and materials and provide complementary information to experiments. Since the applicability of these simulation techniques is limited by the costs of the underlying electronic structure calculations, we develop and assess different machine learning models for this task. The machine learning models are trained on *ab initio* calculations for excited electronic states, using the methylenimmonium cation ($CH_2NH_2^+$) as a model system. Two distinct strategies for modeling excited state properties are tested in this work. The first strategy is to treat each state separately in a kernel ridge regression model and all states together in a multiclass neural network. The second strategy is to instead encode the state as input into the model, which is tested with both models. Numerical evidence suggests that using the state as input yields the best performance. An important goal for excited-state machine learning models is their use in dynamics simulations, which needs not only state-specific information but also couplings, i.e. properties involving pairs of states. Accordingly, we investigate how well machine learning models can predict the couplings. Furthermore, we explore how combining all properties in a single neural network affects the accuracy. Finally, machine learning predicted energies, forces, and couplings are used to carry out excited-state dynamics simulations. Results demonstrate the scopes and possibilities of machine learning to model excited-state properties.

1. Introduction

Many fundamental processes in nature and life are direct consequences of excitation of molecules by light. For example, photosynthesis [1], vision with photo-receptors in the eye [2, 3], or the root cause of diseases such as skin cancer [4, 5] are all based on a photo-induced process. The excited-state dynamics and kinetics of compounds can give insight into why and how these processes occur or can be used to help designing new drugs [6] or materials [7–9].

After a molecule is irradiated with light, it can enter a higher electronic state. Several processes, radiationless or radiative ones, may follow. These processes give rise to the photostability or photodamage of a molecule, and hence photoreactions in general. Knowing the high-dimensional potential energy surfaces of a molecule makes a comprehensive photochemical study possible. However, it is a challenge to find meaningful and accurate potential energy surfaces in advance to execute excited-state molecular dynamics simulations. A possible solution for this problem is the use of on-the-fly *ab initio* molecular dynamics

simulations [10–12]. The mixed-quantum classical methods—such as the surface-hopping methodology that is used in this work—often remain the methods of choice and are a good compromise between accuracy and computational efficiency. In this way, large molecules, i.e. with up to hundreds of atoms, can be treated. Still, the large number of costly electronic structure calculations limits the simulation times of nonadiabatic dynamics to the range of femto- to picoseconds [10, 13–15].

With the rise of machine learning (ML) and the amount of data and computational power available, the fitting of potential energy surfaces of a molecule was put into spotlight. Although such approaches were used already more than 25 years ago [16–20], the interest to speed up simulations in the field of nonadiabatic molecular dynamics simulations has increased only in the last three years [21–30, 30–36].

The main advantage of ML models is that, at least in principle, they can predict any molecular property, typically with much improved efficiency when compared to their quantum chemical counterpart. This can be achieved by learning relations between a molecular structure (in the form of some translation- and rotation-invariant representation) and some target property (provided by quantum chemistry, usually real-valued or complex numbers) [37, 38]. The application of ML models for dynamics simulations in the electronic ground state [39–45] or excited states [21–30, 30–36, 46, 47] already exist, and show the potential for further developing this research field.

Nevertheless, the challenges to model excited states of molecules have not always been tackled successfully with ML: Not only one potential energy surface has to be learned, but several, including the couplings between them, that should also be treated in an ML model [29]. Independent works applied kernel ridge regression (KRR) and neural networks (NNs) to enhance quantum chemical calculations in nonadiabatic molecular dynamics simulations and show the following trend: Models based on KRR need to resort to intermediate quantum chemistry steps in critical regions to obtain correct dynamics simulations [22, 23], whereas NNs are able to completely replace quantum chemistry during the dynamics [24, 26, 28, 29]. This observation raises the question why some models fail for some excited-state properties and others do not. Our goal is to investigate relevant properties for nonadiabatic molecular dynamics simulations by using KRR and NNs. The aim of this work is thus twofold: 1) a comparison of KRR and NNs using different representations for the molecular structure and 2) possible improvements of existing ML techniques for excited-state property prediction.

The methylenimmonium cation, $CH_2NH_2^+$, is used as model system for this purpose. This cation, like the larger homologue retinal, belongs to the member of protonated Schiff bases, and shows isomerization of the double bond after light excitation [48–51]. Similarly, the absorption of light in retinal leads to an isomerization process, that is fundamental for vision [52]. The ultrafast dynamics is particularly challenging to reproduce with ML models, but, at the same time, it remains feasible to provide quantum chemistry reference simulations. Unlike in the isoelectronic molecule ethylene, the excited-state calculations are not hampered by low-lying Rydberg states [53]. Therefore, the methylenimmonium cation is well suited as a test system for this study.

2. Methods

2.1. Surface-hopping molecular dynamics

The program SHARC (Surface-Hopping including ARbitrary Couplings) [54] is used for surface-hopping molecular dynamics simulations with interfaced ML models. Due to the stochastic nature of this method, reliable results can only be obtained when considering an ensemble of independent trajectories. To this aim, Wigner sampling [55] is used to obtain 200 initial configurations of the methylenimmonium cation, to start dynamics simulations. SHARC determines nonadiabatic transitions from one potential energy surface to another from the NAC vectors between each set of singlet states. Those transitions, or so-called hops in surface-hopping, usually take place in conformational regions of the molecule, where two potential energy surfaces are in close proximity. Those regions are termed conical intersection and are not only difficult to model with ML models [23, 24, 29], but also provide a challenge to converge a quantum chemical calculation. This can be attributed to the inverse proportionality of the NAC vectors between two coupled states i and j, NAC_{ij}, to the energy gap between these two states [14]:

$$NAC_{ij} \approx \frac{\langle \Psi_i \mid \nabla_{\mathbf{R}} \hat{H}_{el} \mid \Psi_j \rangle}{E_i - E_j} \text{ for } i \neq j. \tag{1}$$

Ψ_i and Ψ_j are the eigenstates of the electronic Schrödinger equation with E_i and E_j being the respective eigenvalues, i.e. the potential energies of state i and j. $\nabla_{\mathbf{R}}$ denotes the spatial derivatives with respect to the nuclear coordinates. When the two electronic states are degenerate, NACs become infinitely large and consequently show sharp peaks around conical intersections. In contrast, NAC values are almost zero elsewhere. An example is illustrated in figure 1.

Figure 1. Scan along a reaction coordinate of $CH_2NH_2^+$ that shows a conical intersection of the E_i and E_j states. The corresponding NAC values, NAC_{ij}, taken as the norm of the NAC vectors between those states, are large at the conical intersection, but almost zero elsewhere (red dashed line). Multiplied with the energy-difference, smoother properties can be obtained (blue dashed line).

Multiplication of NACs obtained from quantum chemistry (labelled as QC) with the corresponding energy gaps can remove such sharp spikes and can provide smoother quantities (also referred to as interstate couplings) [28].

$$C_{ij}^{\text{reference}} = NAC_{ij}^{QC} \cdot \mid E_i^{QC} - E_j^{QC} \mid \tag{2}$$

For prediction, it is necessary to divide the learned quantity by the energy gap [28] obtained from ML models.

$$NAC_{ij}^{\text{predict}} = \frac{C_{ij}^{ML}}{\mid E_i^{ML} - E_j^{ML} \mid} \tag{3}$$

The outcome is given to the surface-hopping molecular dynamics program to compute the hopping probability. However, this approach requires very accurate ML potentials for energies and those are also challenging to obtain in regions near conical intersections.

2.2. Training set

The training set used here is taken from reference [29] and represents a conformational subspace of the methylenimmonium cation, $CH_2NH_2^+$. This training set involves three singlet states and covers the relevant conformational space that is visited during excited-state molecular dynamics simulations after excitation to the bright second excited state (S_2, $\pi\pi^*$). It is thus considered to constitute an optimal set for analysis of different ML models as well as common molecular representations. Triplet states are not assumed to play a major role in the excited state dynamics of $CH_2NH_2^+$ [29, 48–51]. In case of triplet states, additional couplings between triplet-triplet states (NAC vectors) and singlet-triplet states (spin-orbit couplings) would arise, that could be similarly modelled as NACs between singlet-singlet states.

The quantum chemical reference method is the multi-reference configuration interaction method accounting for single and double excitations with the basis set aug-cc-pVDZ (MR-CISD(6,4)/aug-cc-pVDZ). The active space consists of 4 electrons in 6 orbitals. The data set for training and validation contains 4000 data points that are obtained by randomly shuffling the complete set of 4770 points. The rest of the data set (770 data points) is held back as a test set. Each data point contains the xyz-coordinates of a molecular structure as well as energies for three singlet states, corresponding gradients, (transition) dipole moments, and NACs between each state. In total, 3 energy values, 54 gradient values, 27 values for (transition) dipole moments, and 54 values for NACs have to be predicted. Noticeably, calculating dipole moments and transition dipole moments to reproduce absolute experimental values is difficult even with high-level *ab initio* methods [56]. However when calculating transition probabilities, e.g. oscillator strengths for the simulation of UV spectra, absolute values are rarely important but rather relative values are of interest. The (transition) dipole moments in this work are modelled in the simplest way—that is the direct fitting of values obtained from quantum chemistry, without an *a posteriori* determination of nuclear charges [57]. For more details on the training set and its generation, see reference [29].

2.3. Machine learning models

Learning curves and scatter plots between reference data and predicted data show the quality of each ML model. By plotting the prediction error against the training set size, N, in logarithmic scale, the learning efficiency can be assessed [58–61].

Multi-layer feed forward NNs and KRR using the quantum machine learning (QML) toolkit [62] are chosen as ML models. 5-fold cross-validation is applied to optimize the hyperparameters of each model with the training set of 4000 data points. The MAE is reported on the test set of 770 data points as the mean of 10 calculations along with the standard deviations obtained from models trained on 90% of the 4000 data points. 10% are necessary to employ early-stopping for the NNs. An analogous procedure is applied for comparability when using KRR. A detailed description can be found in the Supplementary Information (SI), which is available online at (stacks.iop.org/MLST/1/025009/mmedia).

The gradients are treated as derivatives of ML potentials for energies as described in reference [63] for NNs and reference [61] for KRR. This is necessary to conserve the energy in nonadiabatic molecular dynamics simulations. For comparison, gradients are directly trained and predicted too. Details on chosen parameters for KRR and NNs are given in the SI in chapter S2.2 and S2.3, respectively. Learning (transition) dipole moments and NACs in addition to energies and gradients with one model can give insights into the influence of of joint learning of different properties. This is straightforward with NNs, therefore this effect is investigated using NNs.

2.3.0.1. Molecular representations
As a molecular representation, the matrix of inverse distances is chosen, as it gave fair results in reference [29] for NNs and other ML models, see for example references [45, 64, 65]. The FCHL (Faber-Christensen-Huang-Lilienfeld) representation is used for KRR [61, 66] and a development version of the same representation is also tested for NNs [67].

The molecule is treated as a whole with the matrix of inverse distances. The distance of an atom to all other atoms is computed, these distances are inverted and arranged in matrix format. This representation is probably the simplest and cheapest representation to use for ML and can be used for very efficient training and evaluation of ML models. The FCHL representation, however, provides a more accurate description and is computationally more expensive to apply. It does not treat the molecule as a whole, but describes an atom in its chemical and structural environment within a pre-defined cut-off region [66]. Not only the distances from one atom to the other atoms are taken into account as two-body terms, but also one- and three-body terms. These account for chemical composition as well as angular contributions, respectively.

An encoding for the quantum energy level is implemented in addition to the aforementioned representations to predict several electronic state energies at once. Several possibilities are tested to describe the electronic state. For KRR, a representation for each electronic singlet state, $S = \{1, 2, 3\}$, containing simply numbers of 1, 2, and 3 for the three states turns out to be beneficial. Other representations of the state do not result in improved learning and only change the additional hyperparameter, the width of the state kernel. Also for the NNs, several state-encoding representations are tested. Duplication of a a molecular representation N_S-times and multiplication of each copy with the corresponding state-number – 1, 2 or 3 in this case, turns out to be best. Due to the existing implementations, gradients are only treated as response properties for KKR with the FCHL representation [61] and as derivatives of NN potentials for energies with the matrix of inverse distances.

2.3.0.2. Kernel ridge regression (KRR)
In KRR, a kernel basis function is placed on each compound (each molecule) in the training set, $\{M_k\}$, and related to a property of a query compound M, $p^{KRR}(M)$, by:

$$p^{KRR}(M) = \sum_{k=1}^{N_M} \alpha_k K(M, M_k) \tag{4}$$

with N_M being the number of molecules in the training set, K the kernel, and $\{\alpha_k\}$ the regression coefficients, which are obtained through linear regression:

$$\alpha = (K + \lambda I)^{-1} p^{\text{reference}}. \tag{5}$$

The regularizer, λ (multiplied with the unit matrix I), is usually small assuming that the noise in the training set is negligible [38, 66]. Using this standard implementation, only one molecular property can be fitted at a time, which we call single-state fitting. An additional Gaussian kernel, $K_2(S, S_l)$, which relates information from a query state S to the set of available states $\{S_l\}$, is used to extended the representations and

4

to predict all electronic states, N_S, at once. This kernel is subsequently combined with the original kernel (now denoted as K_1) that maps a compound to its property.

$$p^{KRR}(M,S) = \sum_{k=1}^{N_M} \sum_{l=1}^{N_S} \alpha_k^l K_1(M,M_k) K_2(S,S_l). \qquad (6)$$

Here, $\mathbf{p}^{KRR}(M,S)$ is a vector of length $N_M \cdot N_S$, which can be recast as a matrix of size $N_M \times N_S$. In contrast, the predicted property $p^{KRR}(M)$ from equation (4) is only dependent on the molecule and is a vector of length N_M – representing only one electronic state. N_S versions of KRR models with the standard representation in equation (5) have to be used to predict N_S energetic values. With the new state representation, only one KRR model has to be trained.

2.3.0.3. Multi-layer feed forward neural networks (NNs)
Compared to KRR, NNs possess more hyperparameters and, thus, are more difficult to optimize with respect to error convergence. However, due to the NN architecture, a benefit lies in the flexibility and possibility to relate a molecular structure to a many-state output. In principle, this many-state output can be obtained without additional encodings like the state kernel $K_S(S,S_l)$ in the KRR approach. For better comparison to the KRR approach, we also used state-encoding representations as detailed above.

All NN models use the numpy [68] and theano [69] distribution implemented in python. To find optimal hyperparameters of the models to represent the relation between a molecular geometry and its multi-dimensional output, random grid search of different sets of hyperparameters is carried out, see reference [29] for details. In all cases, the stochastic gradient descent optimization algorithm adaptive moment estimation (Adam) [70] is applied and the learning rate is annealed during training.

The NN models are trained by optimizing parameters such that mean squared errors of the predicted properties p^{NN} and the reference properties p^{QC} are minimized. Note that p is a vector that contains three values, corresponding to the electronic states, in case of fitted energies. It contains 84 values if all properties are treated together. We use the scheme of reference [40] to additionally include the forces, F, as NN derivatives in the loss function, L_2:

$$L_2 = \frac{1}{N_M}(p^{NN} - p^{QC})^2 \qquad (7)$$
$$+ \frac{1}{N_M}\frac{1}{3N_a}\sum_a^{3N_a}(F_\alpha^{NN} - F_\alpha^{QC})^2.$$

F_α^{NN} are the values of forces predicted with NNs and F_α^{QC} are corresponding reference values, where a runs over all atoms, N_a.

3. Results

ML models are trained on energies, forces, and NACs to speed up simulations and successfully reproduce surface-hopping molecular dynamics simulations of the reference method. The populations obtained from surface-hopping molecular dynamics simulations with chosen ML models that take gradients as derivatives from ML energy potentials are given in figure 2.

The reference dynamics in figure 2 (A) is taken from reference [29], and so is the population scheme in figure 2 (B), that results from an NN model with the inverse distance matrix as the molecular representation. The average populations of the ensemble show that the time evolution is governed by fast population transfer from the second excited state to the first excited state and back to the ground state. The agreement of these methods is fairly good.

Panel (C) shows dynamics computed with an NN model that treats all properties together. Up to around 10 fs, the dynamics agree to the reference scheme. During this time, the population is transferred from the S_2 state to the S_1 state. Afterwards, there are less transitions than expected and in the end of the simulation time only a small fraction of the population is in the S_0 state. For the KRR model with the FCHL representation given in panel (D), all of the population is transferred to the electronic ground state, S_0, within the first 10 fs. After that, there are hops from lower lying states to higher energetic states. Those hops, especially in cases of large potential energy gaps between states, are considered to be implausible and the trajectories are not reliable anymore. Furthermore, the molecule atomizes during the course of the simulation, which is not the case in the quantum chemistry reference dynamics. The premature population transfer leads the molecule to regions of the conformational subspace that are not visited with the reference method and are also not

Figure 2. Populations of three singlet states as a function of time obtained from surface-hopping molecular dynamics simulations using six different methods. Dynamics based on [A] the quantum chemistry reference[1] (labeled as QC), [B] NN[1] models using the inverse distance as a descriptor, [C] NNs that treat all properties in one model, [D] KRR with the FCHL representation, and [E]-[F] mixed NN-KRR models, where energies and gradients are taken either from the NN model of panel [B] and NACs from KRR of panel [D] or vice versa. Dotted lines are populations that are considered to be wrong with respect to the reference or exhibit large energy fluctuations ([1] taken from reference [29]).

considered in the training set. Panels (E) and (F) show results from mixed models, i.e. KRR and NN models, and will be discussed later.

   The way ML models learn shows why dynamics is erroneous. A correct learning behaviour is given, when the out of sample error of an ML model decreases with increasing training set size, which has been shown by Vapnik and coworkers for KRR [58] and by Müller and coworkers for NNs [59]. The learning curves are given in figure 3(A). While KRR (blue) and the NN model (red) for energies and gradients yield very similar accuracy, the NNs that treat all properties together (green) are far from being accurate.

   The learning curves for the NACs (left panel in figure 3(B)) as directly obtained from quantum chemistry, i.e. as non-smooth properties, show that the learning efficiency of KRR is much higher than the one of the NNs, which is dictated by the slope of the learning curve. The NN models are comparable in their accuracy. The learning of the smooth NACs (as described in equation (2)) is evaluated in the middle panel with KRR and NN models that account for each property separately. The inaccurate potential energies of the NN model that treats all properties together prohibit the prediction of smooth NACs in this case. The accuracy of the actual NACs obtained from the smooth couplings and the corresponding ML energy gaps (as given in panel (A)), is already high for a small training set size and is comparable to the MAE for direct NAC prediction with a large training set. However, increasing the training set size cannot improve prediction accuracy anymore. For KRR, the MAE can be reduced with increasing training set size, but the learning curve is not linear. These findings indicate that the energies must be predicted with high accuracy in order to use smooth NACs, compare also figure S4 panels (B) and (C) in the SI. The learning curves for (transition) dipole moments show similar trends as the ones for energies, gradients and NACs and are given in the SI in figure S5. Furthermore, NNs with the FCHL representation as well as KRR with the inverse distance matrix as molecular representations are discussed along with scatter plots for energies and gradients (see SI figures S1, S2, S4, and S5).

   Having analyzed the learning behavior for energies and NACs separately, their interplay is investigated in the following. We thus return to the dynamics depicted in figure 2 in panels (E) and (F). In these panels, population plots from mixed models are shown—once with energies and gradients obtained from NNs and NACs obtained from KRR and once vice versa. Surprisingly, none of the cases leads to accurate population schemes. Nethertheless, the trend of the populations in panel (E) – with energies and gradients from NNs and NACs from KRR—is similar to the reference scheme. The dynamics only happens on a much shorter time scale, indicating too large NACs. Plotting the potential energy curves along two different reaction

Figure 3. Learning curves showing the mean absolute error (MAE) for the energies and gradients [A] and NACs [B] averaged over all three singlet states (MAEs obtained for the separate energy levels S_0, S_1, and S_2 are shown in the SI in figure S1). [C] Scans along the CN-bond elongation of the methylenimmonium cation (left panel) and the rotation along the dihedral angle (right panel) computed with quantum chemistry (QC, black continuous lines), KRR (blue dotted lines), and NNs (red dashed lines).

coordinates including a critical region (panel C in figure 3) reveals that both ML models, KRR and NNs, can correctly reproduce the shape of the curves. However, there are small deviations in the critical regions of the PESs. A comparison of all these findings implies that an accurate prediction of energies is more important for reproducing the dynamics than an accurate prediction of NACs. Nevertheless, it is intuitively clear that surface-hopping molecular dynamics requires all properties to be accurate enough: Having the correct potentials, but wrong NAC values does also result in wrong dynamics.

3.1. Multi-state representation

A major difference between NN models and KRR models is, that the NNs can predict all electronic states at once, whereas three independent KRR models are used for the three states. However, the PESs are not independent from each other. Consequently, it should also be favourable to learn all states together in one ML model. Encoding of the quantum energy level in an additional state-representation makes multi-state outputs possible. This representation is tested for KRR with the FCHL representation and NNs with the inverse distance matrix.

A comparison of the standard representations and the encoded state representations is given in figure 4(A). All ML models show that learning of all state energies at once is possible and favorable. Encoding of the quantum energy level improves the accuracy of all ML models, whereby this effect is significant for KRR models and small for NNs. With the state representation for KRR, the kernel matrix size increases from $N_m \times N_m$ to $N_S \cdot N_m \times N_S \cdot N_m$. Mapping only a subset of molecules to the complete dataset for training can reduce the memory consumption and can make the training process more efficient without a major loss in accuracy.

Principal component analysis provides an explanation to the above observed results. Therefore, the first principal component is plotted against the second principal component of an ML model in figure 5. The ML models used for surface-hopping dynamics are compared to the ML models with a state representation. Remarkably, a state representation for KRR (figures 5 (B) and (C)) leads to a clear ordering of the data corresponding to different electronic states. Moreover, within one state, a better ordering can be obtained

7

Figure 4. Learning curves showing the mean absolute error (MAE) for the energies obtained from standard and state representations [A] along with corresponding scatter plots [B]. [C] Scans along the CN-bond elongation of the methylenimmonium cation (left panel) and the rotation along the dihedral angle (right panel) computed with quantum chemistry (QC, black continuous lines), KRR (blue dotted lines), and NNs (red dashed lines).

than for KRR without a representation for the electronic states (panel (A)). Similarly, NNs illustrated in panels (D) and (E) show an improved ordering of data, when a representation for the electronic state is used. The application of those models in surface-hopping dynamics, is assumed to lead to even more accurate energy predictions, but requires the gradients as derivatives from ML potentials and as response properties from KRR models. This approach needs further considerations, especially due to the high memory consumption of KRR models, and will be the subject of future developments. Nevertheless, the results clarify how an improved ordering of data can lead to higher prediction accuracy of ML models and thus highlight the importance of the molecular representation in addition to the type of regressor.

4. Conclusion

In this paper, two frequently used ML regressors, namely KRR and NNs, are compared for their application in excited-state molecular dynamics of $CH_2NH_2^+$. The role of the ML model in combination with different representations of the molecular structure for the prediction of energies of the ground state as well as excited states, corresponding forces, NACs between different states, and (transition) dipole moments, is investigated. All ML models are able to learn the relation between a quantum chemical property and the molecular structure, when the properties are treated separately from each other. Learning all properties at once leads to significantly worse results and the learning of single properties can even be impeded when the cost function includes all properties at once.

It is shown that the FCHL representation is in most cases superior to the matrix of inverse distances. Encoding of the quantum energy level in the representation can further improve results and make multiple outputs for KRR possible. The state encoding is shown for three electronic singlet states for KRR as well as NNs. In both cases, the modification of the representation is necessarily accompanied by an enlargement of the ML model, a larger kernel matrix in the case of KRR, and a larger input layer in the case of NNs. Principal component analyses further show that an enhanced ordering of data points is obtained.

Figure 5. Principal component analysis of the kernel matrices using the FCHL representation and the matrices obtained from the last NN-layers with the inverse distance matrix. The left side gives results for KRR, whereas the right side shows results obtained from NNs.
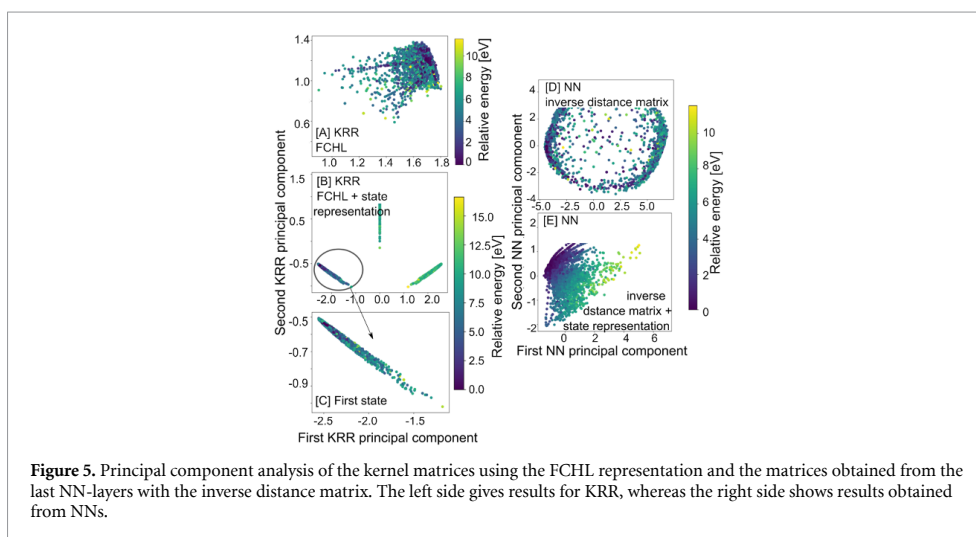
The NAC vectors pose a real challenge for ML models due to their peaked nature. The inclusion of the energy gap of the coupled pair of states can improve the accuracy of KRR models, but does not allow for more accurate NAC vectors in general. This is due to deteriorating effects of the errors in the energy gaps.

Finally, it is not obvious that one ML model outperforms the other. Different options, such as the size of the training set and the computational power available, may favour different models. The faster hyperparameter optimization for KRR and the higher model flexibility of NNs lets us recommend to use KRR for first exploratory runs and NNs for final production runs of excited-state dynamics simulations. The concept of *wide and deep learning* [71] is interesting for future applications, in the sense that different ML models can be applied within one application to combine their distinct benefits.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

The data that support the findings of this study are openly available at
http://dx.doi.org/10.103 9/C9SC01 742A.

Acknowledgments

ORCID iDs

Julia Westermayr ⊙ https://orcid.org/0000-0002-6531-0742
O Anatole von Lilienfeld ⊙ https://orcid.org/0000-0001-7419-0466
Philipp Marquetand ⊙ https://orcid.org/0000-0002-8711-1533

9

References

[1] Vass I, Cser K and Cheregi O 2007 *Ann. NY Acad. Sci.* **1113** 114–22
[2] Schönlein R, Peteanu L, Mathies R and Shank C 1991 *Science* **254** 412–15
[3] Garavelli M, Negri F and Olivucci M 1999 *J. Am. Chem. Soc.* **121** 1023–9
[4] Schreier W J, Schrader T E, Koller F O, Gilch P, Crespo-Hernández C E, Swaminathan V N, Charell T, Zinth W and Kohler B 2007 *Science* **315** 625–9
[5] Rauer C, Nogueira J J, Marquetand P and González L 2016 *J. Am. Chem. Soc.* **138** 15911–16
[6] Ahmad I, Ahmed S, Anwar Z, Sheraz M A and Sikorski M 2016 *Int. J. Photoenergy* **2016** 1–19
[7] Häse F, Valleau S, Pyzer-Knapp E and Aspuru-Guzik A 2016 *Chem. Sci.* **7** 5139–47
[8] Butler K T, Davies D W, Cartwright H, Isayev O and Walsh A 2018 *Nature* **559** 547–55
[9] Liu J and Prezhdo O V 2015 *J. Phys. Chem. Lett.* **6** 4463–9
[10] Domcke W and Sobolewski A L 2013 *Nat. Chem.* **5** 257–8
[11] Liu F, Du L, Zhang D and Gao J 2017 *Sci. Rep.* **7** 1–12
[12] Mai S, Marquetand P and González L 2018 *WIREs Comput. Mol. Sci.* **8** e1370
[13] Mai S, Marquetand P and González L 2016 *J. Phys. Chem. Lett.* **7** 1978–83
[14] Doltsinis N L 2006 *Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum-Classical Approaches* (Jülich, Germany: John von Neuman Institut for Computing) vol 31
[15] Subotnik J E, Jain A, Landry B, Petit A, Ouyang W and Bellonzi N 2016 *Annu. Rev. Phys. Chem.* **67** 387–417
[16] Blank T B and Brown S D 1994 *J. Chemometr.* **8** 391–407
[17] Blank T B, Brown S D, Calhoun A W and Doren D J 1995 *J. Chem. Phys.* **103** 4129–37
[18] Brown D F R, Gibbs M N and Clary D C 1996 *J. Chem. Phys.* **105** 7597–604
[19] Tafeit E, Estelberger W, Horejsi R, Moeller R, Oettl K, Vrecko K and Reibnegger G 1996 *J. Mol. Graphics Modell.* **14** 12–18
[20] No K T, Chang B H, Kim S Y, Jhon M S and Scheraga H A 1997 *Chem. Phys. Lett.* **271** 152–6
[21] Richings G W and Habershon S 2017 *Chem. Phys. Lett.* **683** 228–33
[22] Hu D, Xie Y, Li X, Li L and Lan Z 2018 *J. Phys. Chem. Lett.* **9** 2725–32
[23] Dral P O, Barbatti M and Thiel W 2018 *J. Phys. Chem. Lett.* **9** 5660–3
[24] Chen W K, Liu X Y, Fang W H, Dral P O and Cui G 2018 *J. Phys. Chem. Lett.* **9** 6702–8
[25] Williams D M G and Eisfeld W 2018 *J. Chem. Phys.* **149** 204106
[26] Xie C, Zhu X, Yarkony D R and Guo H 2018 *J. Chem. Phys.* **149** 144107
[27] Richings G W and Habershon S 2018 *J. Chem. Phys.* **148** 134116
[28] Guan Y, Zhang D H, Guo H and Yarkony D R 2019 *Phys. Chem. Chem. Phys.* **21** 14205–13
[29] Westermayr J, Gastegger M, Menger M F S J, Mai S, González L and Marquetand P 2019 *Chem. Sci.* **10** 8100–7
[30] Guan Y, Guo H and Yarkony D R 2020 *J. Chem. Theory Comput.* **16** 302–13
[31] Krems R V 2019 *Phys. Chem. Chem. Phys.* **21** 13392–13410
[32] Richings G W, Robertson C and Habershon S 2019 *J. Chem. Theory Comput.* **15** 857–70
[33] Polyak I, Richings G W, Habershon S and Knowles P J 2019 *J. Chem. Phys.* **150** 041101
[34] Guan Y, Guo H and Yarkony D R 2019 *J. Chem. Phys.* **150** 214101
[35] Wang Y, Xie C, Guo H and Yarkony D R 2019 *J. Phys. Chem. A* **123** 5231–41
[36] Westermayr J, Gastegger M and Marquetand P 2020 *J. Phys. Chem. Lett.* **11** 3828–34
[37] Behler J 2015 *Int. J. Quantum Chem.* **115** 1032–50
[38] Rupp M 2015 *Int. J. Quantum Chem.* **115** 1058–73
[39] Li Z, Kermode J R and De Vita A 2015 *Phys. Rev. Lett.* **114** 096405
[40] Gastegger M, Behler J and Marquetand P 2017 *Chem. Sci.* **8** 6924–35
[41] Botu V, Batra R, Chapman J and Ramprasad R 2017 *J. Phys. Chem. C* **121** 511–22
[42] Artrith N, Urban A and Ceder G 2017 *Phys. Rev. B* **96** 014112
[43] Behler J 2017 *Angew. Chem. Int. Edit.* **56** 12828–40
[44] Chmiela S, Tkatchenko A, Sauceda H E, Poltavsky I, Schütt K T and Müller K R 2017 *Sci. Adv.* **3** e1603015
[45] Schütt K T, Gastegger M, Tkatchenko A, Müller K R and Maurer R J 2019 *Nat. Commun.* **10** 5024
[46] Behler J, Reuter K and Scheffler M 2008 *Phys. Rev. B* **77** 115421
[47] Carbogno C, Behler J, Reuter K and Groß A 2010 *Phys. Rev. B* **81** 035410
[48] Barbatti M, Aquino A J A and Lischka H 2006 *Mol. Phys.* **104** 1053–60
[49] Tapavicza E, Tavernelli I and Rothlisberger U 2007 *Phys. Rev. Lett.* **98** 023001
[50] Tavernelli I, Tapavicza E and Rothlisberger U 2009 *J. Chem. Phys.* **130** 124107
[51] Tavernelli I, Tapavicza E and Rothlisberger U 2009 *J. Mol. Struct.: Theor.* **914** 22–9
[52] Hudock H R, Levine B G, Thompson A L, Satzger H, Townsend D, Gador N, Ullrich S, Stolow A and Martínez T J 2007 *J. Phys. Chem. A* **111** 8500–8
[53] Hollas D, Šištík L, Hohenstein E G, Martínez T J and Slavíček P 2018 *J. Chem. Theory Comput.* **14** 339–350
[54] Mai S, Richter M, Ruckenbauer M, Oppel M, Marquetand P and González L 2018 Sharc2.0: Surface hopping including arbitrary couplings – program package for non-adiabatic dynamics (sharc-md.org)
[55] Wigner E 1932 *Phys. Rev.* **40** 749–50
[56] Tennyson J 2016 *J. Chem. Phys.* **145** 120901
[57] Gastegger M and Marquetand P 2018 arxiv:1812.07676 [physics.chem-ph]
[58] Cortes C, Jackel L D, Solla S A, Vapnik V and Denker J S 1994 Learning curves: Asymptotic values and rate of convergence ed Cowan J D, Tesauro G and Alspector J *Advances in Neural Information Processing Systems 6* (San Mateo, CA: Morgan Kaufmann Publishers) pp 327–34
[59] Müller K R, Finke M, Murata N, Schulten K and Amari S 1996 *Neural Comput.* **8** 1085–106
[60] von Lilienfeld O A 2018 *Angew. Chem. Int. Edit.* **57** 4164–9
[61] Christensen A S, Faber F A and von Lilienfeld O A 2019 *J. Chem. Phys.* **150** 064105
[62] Christensen A, Faber F, Huang B, Bratholm L, Tkatchenko A, Müller K and Lilienfeld O A 2017 *QML: A Python toolkit for quantum machine learning* (https://github.com/qmlcode/qml)
[63] Gastegger M and Marquetand P 2015 *J. Chem. Theory Comput.* **11** 2187–98
[64] Chmiela S, Sauceda H E, Müller K R and Tkatchenko A 2018 *Nat. Commun.* **9** 3887

[65] Schütt K T, Sauceda H E, Kindermans P J, Tkatchenko A and Müller K R 2018 *J. Chem. Phys.* **148** 241722

[66] Faber F A, Christensen A S, Huang B and von Lilienfeld O A 2018 *J. Chem. Phys.* **148** 241717

[67] Christensen A S, Bratholm L A, Faber F A and von Lilienfeld O A 2020 *J. Chem. Phys.* **152** 044107

[68] van der Walt S, Colbert S C and Varoquaux G 2011 *Comput. Sci. Eng.* **13** 22–30

[69] Al-Rfou R *et al* (The Theano Development Team) 2016 arXiv:abs/1605.02688

[70] Kingma D P and Ba J 2014 arXiv:1412.6980 [cs.LG]

[71] Cheng H T, Koc L, Harmsen J, Shaked T, Chandra T, Aradhye H, Anderson G, Corrado G, Chai W, Ispir M, Anil R, Haque Z, Hong L, Jain V, Liu X and Shah H 2016 arXiv:1606.07792

102

Julia Westermayr, Michael Gastegger and Philipp Marquetand

Contributions:

Julia Westermayr implemented SchNarc by providing the interface of SchNet to SHARC, adapted the building blocks of SchNet for excited-state potential energy surfaces and properties, implemented the phase-free training algorithm, developed the approximation of nonadiabatic couplings from derivatives of potential energy surfaces, conducted the reference simulations, carried out the test runs to evaluate the accuracy of machine learning models, computed the training sets, wrote the first draft of the manuscript and contributed to the final version.

Michael Gastegger supervised this work, set up the framework for SchNarc, discussed the idea of phase-free training and derivation of coupling vectors from virtual properties, and contributed to the initial and final writing of the manuscript.

Philipp Marquetand conceived the goal of this work, supervised it, discussed the phase-free training algorithm and linear-vibronic coupling model, and contributed to the initial and final writing of the manuscript.

Letter

Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics

Julia Westermayr, Michael Gastegger,* and Philipp Marquetand*

Cite This: *J. Phys. Chem. Lett.* 2020, 11, 3828−3834

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: In recent years, deep learning has become a part of our everyday life and is revolutionizing quantum chemistry as well. In this work, we show how deep learning can be used to advance the research field of photochemistry by learning all important properties— multiple energies, forces, and different couplings—for photodynamics simulations. We simplify such simulations substantially by (i) a phase-free training skipping costly preprocessing of raw quantum chemistry data; (ii) rotationally covariant nonadiabatic couplings, which can either be trained or (iii) alternatively be approximated from only ML potentials, their gradients, and Hessians; and (iv) incorporating spin−orbit couplings. As the deep-learning method, we employ SchNet with its automatically determined representation of molecular structures and extend it for multiple electronic states. In combination with the molecular dynamics program SHARC, our approach termed SchNarc is tested on two polyatomic molecules and paves the way toward efficient photodynamics simulations of complex systems.
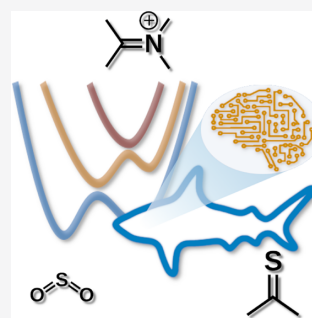
Excited-state dynamics simulations are powerful tools to predict, understand, and explain photoinduced processes, especially in combination with experimental studies. Examples of photoinduced processes range from photosynthesis, DNA photodamage as the starting point of skin cancer, to processes that enable our vision.[1−5] As they are part of our everyday lives, their understanding can help to unravel fundamental processes of nature and to advance several research fields, such as photovoltaics,[6,7] photocatalysis,[8] or photosensitive drug design.[9]

Because the full quantum mechanical treatment of molecules remains challenging, exact quantum dynamics simulations are limited to systems containing only a couple of atoms, even if fitted potential energy surfaces (PESs) are used.[10−26] In order to treat larger systems in full dimensions, i.e., systems with up to hundreds of atoms, and on long time scales, i.e., in the range of several 100 ps, excited-state machine learning (ML) molecular dynamics (MD), where the ML model is trained on quantum chemistry data, has evolved as a promising tool in the last couple of years.[27−33]

Such nonadiabatic MLMD simulations are in many senses analogous to excited-state ab initio molecular dynamics simulations. The only difference is that the costly electronic structure calculations are mostly replaced by an ML model, providing quantum properties like the PESs and the corresponding forces. The nuclei are assumed to move classically on those PESs. This mixed quantum−classical dynamics approach allows for a very fast on-the-fly evaluation of the necessary properties at the geometries visited during the dynamics simulations.

In order to account for nonadiabatic effects, i.e., transitions from one state to another, further approximations have to be introduced.[34] One method, which is frequently used to account for such transitions, is the surface-hopping method originally developed by Tully.[35] A popular extension for this method including not only nonadiabatic couplings (NACs) but also other couplings, e.g., spin−orbit couplings (SOCs), is the SHARC (surface hopping including arbitrary couplings) approach.[36−38] Importantly, NACs, also called derivative couplings, are used to determine the hopping directions and probabilities between states of the same spin multiplicity.[36,37,39−41] The NAC vector (denoted as C^{NAC}) between two states, i and j, can be computed as[39,42,43]

$$C_{ij}^{NAC} \approx \langle \Psi_i | \frac{\partial}{\partial \mathbf{R}} \Psi_j \rangle = \frac{1}{E_i - E_j} \langle \Psi_i | \frac{\partial H_{el}}{\partial \mathbf{R}} | \Psi_j \rangle \text{ for } i \neq j \quad (1)$$

where the second-order derivatives are neglected. As a further difficulty, NACs are often missing from quantum chemistry implementations and hopping probabilities and directions are thus often approximated.[44−52] SOCs (denoted as C^{SOC}) are present between states of different spin multiplicity

3828
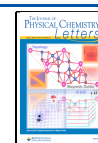
$$C_{ij}^{SOC} = \langle \Psi_i | \hat{H}^{SO} | \Psi_j \rangle \tag{2}$$

and determine the rate of intersystem crossing. They are obtained as off-diagonal elements of the Hamiltonian matrix in standard electronic-structure calculations.[37,53]

Most of the recent studies involving ML dynamics deal with ground-state MD simulations, see e.g. refs [54−78], where one of the most promising ML models is SchNet,[79,80] a deep continuous-filter convolutional-layer neural network. In contrast to popular ML models like RuNNer,[64] n2p2,[81] TensorMol,[82] ANI,[63] or the DeePMD model[83] that require hand-crafted molecular descriptors, SchNet belongs to the class of message-passing neural networks.[84] Other examples of such networks are the DTNN,[85] PhysNet,[86] or HIP-NN.[87] The advantage is that the descriptors of molecules are automatically designed by a deep neural network and are based on the provided data set. In this way, the descriptors are tailored to the encountered chemical environments. Thus, we choose SchNet as a convenient platform for our developments.

An arising difficulty compared to ground-state energies and properties is that for the excited states not only one, but several PESs as well as the couplings between them have to be taken into account. Only a small but quickly increasing number of studies deal with the treatment of excited states and their properties using ML.[15−28,31,33,88]

In addition to the higher dimensionality that can be tackled with ML,[89,90] the learning of couplings proves challenging because properties resulting from electronic wave functions of two different states, Ψ_i and Ψ_j, have their sign dependent on the phase of the wave functions.[32,33,91,92] Because the wave function phase is not uniquely defined in quantum chemistry calculations, random phase jumps occur, leading to sign jumps of the coupling values along a reaction path. Hence, the couplings can not be learned directly as obtained from a quantum chemistry calculation. An option is to use a phase correction algorithm to preprocess data and remove these random phase jumps. Assuming that the effect of the Berry phase remains minor on the training set, smooth properties are obtained that are learnable by ML models.[32] However, this approach is expensive and many quantum chemistry reference computations are necessary to generate the training set. In cases of large polyatomic molecules with many close-lying energetic states, this approach might even be infeasible.

The aim of this Letter is to provide a framework to carry out efficient excited-state MLMD simulations and to combine two popular methods for this purpose: the SHARC approach for photodynamics with states of different multiplicity and SchNet to efficiently and accurately fit potential energies and other molecular properties. We call this combination the SchNarc approach and adapted SchNet for the treatment of excited-state potentials, their forces, and couplings for this purpose. The SchNarc approach can overcome the current limitations of existing MLMD simulations for excited states by allowing (i) a phase-free training to omit the costly preprocessing of raw quantum chemistry data and, to treat (ii) rotationally covariant NACs, which can either be trained or (iii) alternatively be approximated from only ML potentials, their gradients, and Hessians, and to treat (iv) SOCs. When using the phase-free training in combination with the approximated NACs, the costs required for the training set generation can be reduced substantially. Further, each data point can be computed in parallel, which is not possible when phase correction is needed. With all these methodological advances, the SchNarc ML

approach simplifies accelerated nonadiabatic dynamics simulations, broadening the range of possible users and the scope of systems, in order to make long time scales accessible.

To validate our developments, the surface hopping dynamics of two molecules, showing slow and ultrafast excited-state dynamics, are investigated. The first molecule is $CH_2NH_2^+$, of which we take a phase-corrected training set from ref [32]. Using the same level of theory (MR-CISD(6,4)/aug-cc-pVDZ) with the program COLUMBUS,[93] the training set is recomputed without applying phase correction to train ML models also on raw data obtained directly from quantum chemistry programs. This training set should be used to validate our phase-free training approach. The ML models are trained on energies, gradients, and NACs for three singlet states using 3000 data points.

Slow photoinduced processes are present in thioformaldehyde, CSH.[94] The training set is built up of 4703 data points with two singlet states and two triplet states after initial sampling of normal modes and adaptive sampling with simple multilayer feed-forward neural networks according to the scheme described in ref [32]. This scheme applied for the training set generation is based on an uncertainty measure. Two (or more) ML models are trained and dynamics simulations are performed. At each time step, the predictions of the different ML models are compared. Whenever the error between the models exceeds a manually defined threshold, the molecular geometry visited at this time step is recomputed with quantum chemistry and the data added to the training set. Our previously proposed network[32] was used for this purpose because the training set was generated before SchNarc was developed. The sampling procedure is largely independent of the network architecture used and could therefore also be carried out with SchNarc but has not been tested here. The program MOLPRO[95] is used for the reference calculations with CASSCF(6,5)/def2-SVP.

The main novelty of the phase-free training is that it removes the influence of the arbitrary phase during the learning process of an ML model. It can be applied to any existing ML model capable of treating excited states and any existing data set for excited states eliminating an expensive and time-consuming phase correction preprocessing. The chosen ML model here is SchNet, for which the new loss function termed *phase-less loss function* is implemented and tested using the methylenimmonium cation, $CH_2NH_2^+$. The phase-less loss is based on the standard L_2 loss, but here, the squared error of the predicted properties is computed 2^{N_S-1} times, with N_S being the total number of states. The value of each property, L_P (i.e., L_{SOC} and L_{NAC}), that enters the loss function is the minimum function of all possible squared errors ε_P^k:

$$L_P = \min(\{\varepsilon_P^k\}) \text{ with } 0 \leq k \leq 2^{N_S-1} \tag{3}$$

with

$$\varepsilon_P^k = \begin{cases} \frac{1}{N_S^2} \sum_i^{N_S} \sum_{j\neq i}^{N_S} \frac{1}{N_A} \sum_m^{N_A} \left\| P_{ij,m}^{QC} - P_{ij,m}^{ML} \cdot p_i^k \cdot p_j^k \right\|^2 & \text{if } \dim(\mathbf{P}) \geq 3 \\ \frac{1}{N_S^2} \sum_i^{N_S} \sum_{j\neq i}^{N_S} \left\| P_{ij}^{QC} - P_{ij}^{ML} \cdot p_i^k \cdot p_j^k \right\|^2 & \text{if } \dim(\mathbf{P}) \leq 2 \end{cases} \tag{4}$$

for vectorial and nonvectorial properties, respectively. The error ε_P^k for a specific phase is computed as the mean squared error of a property P from quantum chemistry (index QC) and

machine learning (index ML). The property P couples different states, indicated by i and j. Because the wave function of each of the states can have an arbitrary phase, the property P_{ij} that couples state i and j has to be multiplied with a product of the phases for these states, $p_i \cdot p_j$. The phases for all states together form a vector $\mathbf{p}$ with entries of either +1 and −1. Which of the 2^{N_S-1} possible combinations for p is chosen is indicated by the index k, also defined in eq 3. The possible combination that gives the lowest error enters the loss function. This is done for all samples inside of the training set and can be seen as an internal ML-based phase correction. Because of the continuity of ML fitted functions, the sign of properties is consistent for predictions. Note that the relative signs within one vector remain and must be predicted correctly for successful training.

The overall loss function in this work is a combination of such phase-less loss functions and mean squared errors for all properties with a trade-off factor to account for their relative magnitude. The relative magnitude of each property is defined by a manually set trade-off factor (a detailed description of the implementation is given in the Supporting Information).

Results are given in Figure 1, which shows the population schemes of $CH_2NH_2^+$ obtained after excitation to the second
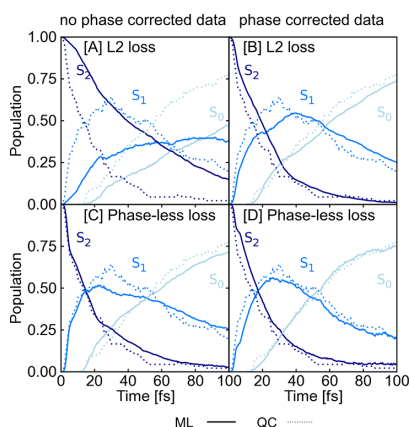

Figure 1. Populations obtained from 90 QC (MR-CISD(6,4)/aug-cc-pVDZ) trajectories are shown by dotted lines and are compared to populations resulting from 1000 trajectories initially excited to the S_2 obtained from SchNet (solid lines) that is trained on (A) not phase-corrected data and takes the L_2 norm as loss function; (B) a similar SchNet model, but trained on phase-corrected data; (C) a SchNet model trained on not phase-corrected data, but using the new phase-less loss function; and (D) a SchNet model trained on phase-corrected data using the new phase-less loss function.

excited singlet state, S_2. The populations obtained from quantum chemistry are shown by dotted lines. SchNarc models are illustrated using solid lines. Panels A and C are obtained from SchNarc models trained on a data set that is not phase-corrected; that is, it contains couplings that can randomly switch their sign. Those are compared to populations obtained from models trained on phase-corrected data (panels B and D). As can be seen, the L_2 loss function, as used in the upper plots, leads to an accurate ML model to reproduce ultrafast transitions only in the case of phase-corrected data (panel B), whereas this loss can not be used when trained on raw quantum chemistry data (panel A). In comparison, a

SchNarc simulation with an ML model that applies the phase-less loss function is successful in reproducing the populations for both training sets (panels C and D).

In those simulations, the NACs are multiplied with the corresponding energy gaps, i.e., $\tilde{C}_{ij}^{NAC} = C_{ij}^{NAC} \cdot \Delta E_{ij}$, to get rid of singularities.[21,33] These smooth couplings $\tilde{C}_{ij}^{NAC}$ are not directly learned, but rather constructed as the derivative of a virtual property, analogously to forces that are predicted as derivatives of an energy-ML model. The virtual property is the multidimensional antiderivative of the rightmost expression in eq 1, $\langle \Psi_i | \frac{\partial H_{el}}{\partial \mathbf{R}} | \Psi_j \rangle$ (a derivation is given in the Supporting Information). Compared to previous ML models for NACs,[29,30,32,33] where NACs are learned and predicted as direct outputs or even single values, this approach provides rotational and translational covariance, which has recently been achieved in a similar way for the electronic friction tensor.[96]

However, even without the need of preprocessing the training set, the costly computations of NAC vectors for the training set generation remain. Approximations of NACs exist and often involve the computation of the squared energy-gap Hessian.[11,97−101] Their use in dynamics simulations is rather impracticable with quantum chemistry methods, especially in the case of complex systems, because of the expenses of computing second-order derivatives.

Here, we take advantage of the efficiency for second-order derivative computation from ML models with respect to atomic coordinates to obtain the Hessians of the fitted PESs:

$$\frac{\partial^2 (\Delta E_{ij})^2}{\partial R^2} = 2\left(\Delta E_{ij} \cdot \frac{\partial^2 \Delta E_{ij}}{\partial R^2} + \left(\frac{\partial \Delta E_{ij}}{\partial R} \right)^2 \right) \tag{5}$$

with R being the atomic coordinates of a molecular system. Note that Hessians are also employed in quantum dynamics simulations,[102,103] which might open further applications for our implementation.

The squared energy-gap Hessian can be further obtained as the sum of two symmetric dyads, which define the branching space.[101] Hence, this Hessian can be employed to obtain the symmetric dyad of the smooth NACs via[11,104]

$$\tilde{C}_{ij}^{NAC} \otimes \tilde{C}_{ij}^{NAC} \approx \frac{\partial^2 (\Delta E_{ij})^2}{8 \partial R^2} - \frac{\partial \Delta E_{ij}}{2 \partial R} \otimes \frac{\partial \Delta E_{ij}}{2 \partial R} \tag{6}$$

After singular value decomposition, the hopping direction can be computed as the eigenvector, v_{ij}, of the largest nonzero eigenvalue[101,104,105] with the corresponding eigenvalue, λ_{ij}, as the squared magnitude of the ML smooth coupling, $\tilde{C}_{ij}^{NAC}$. The final approximated NAC vectors, C_{ij}^{aNAC}, between two states are then

$$C_{ij}^{aNAC} = v_{ij} \cdot \frac{\sqrt{\lambda_{ij}}}{\Delta E_{ij}} \tag{7}$$

The approximated NAC vectors can be employed in the vicinity of a conical intersection; otherwise, the output becomes too noisy. For the latter reason, we define thresholds of 0.5 and 1.0 eV for the energy gaps to compute approximated NACs between coupled singlet−singlet states and triplet−triplet states, respectively. It is worth mentioning that the ML models slightly overestimate the energy gaps at a conical intersection,[32] because ML PESs, in contrast to quantum chemical PESs, are smooth everywhere and can reproduce the cones present in such critical regions only to a certain extent.

In ref 105, approximated NACs were applied for a 1D system, and their usefulness in combination with ML was already anticipated.

We turn this idea into reality and show ML excited-state dynamics with approximated NACs for the methylenimmonium cation, $CH_2NH_2^+$, as presented before, and thioformaldehyde, CSH_2. A detailed analysis on the reference computations, the ab initio methods applied, as well as information on the timing of the Hessian evaluation are given in the Supporting Information in sections S2 and S3.1, respectively. The quality of the approximated NACs that is further compared to learned NACs is assessed using a linear vibronic coupling model of sulfur dioxide. The results are given in the Supporting Information in section S4.1 in Figures S1–S3 and support the validity of this approximation. Scatter plots and scans along a reaction coordinate of $CH_2NH_2^+$ and CSH_2 are further computed for energies, gradients, and couplings (and approximated NACs for $CH_2NH_2^+$) in sections S4.2 (Figures S4 and S5) and S4.3 (Figures S6 and S7), respectively. Computed normal modes from ML Hessians are compared to reference values in Tables S3–S5 and show their accuracy. None of the data points from the ab initio MD simulations, to which we compare our SchNarc models, are included in the training sets, and thus, the dynamics simulations can be seen as an additional test.

The populations of $CH_2NH_2^+$ are given in Figure 2. The reference population (panel A) is compared to SchNarc
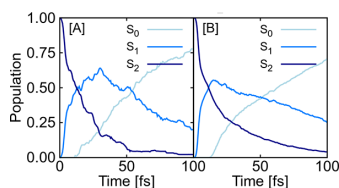
Figure 2. Quantum populations of the methylenimmonium cation obtained from (A) 90 trajectories using MR-CISD/aug-cc-pVDZ (QC) and (B) 1000 trajectories using SchNarc (ML). ML models are trained only on energies and gradients, and NACs are approximated from energies, gradients, and Hessians of ML models; this is in contrast to results from Figure 1, where ML models are also trained on NACs.

simulations trained on energies and gradients (panel B). As is visible, $CH_2NH_2^+$ serves as a test system for ultrafast population transfer after photoexcitation. Transitions from the second excited singlet state back to the ground state take place within 100 fs,[32,106] which can be reproduced only with accurate NACs.[33] Those transitions can be reproduced with SchNarc using also the approximated NACs.

The application of the NAC approximation is further tested on CSH_2, showing slow population transfer. This model also includes triplet states; thus, SOCs are additionally treated with SchNarc. To the best of our knowledge, for the first time, SOCs are trained with ML as directly obtained from quantum chemistry. The population curves are given in Figure 3, where panel A gives the reference population for 3000 fs and panel B the SchNarc populations. In contrast to the methylenimmonium cation, the CSH_2 molecule serves as a test system for slow populations transfer and shows intersystem crossing strongly dependent on the accuracy of the underlying potentials.[94] Inaccurate ML models would thus be unable to
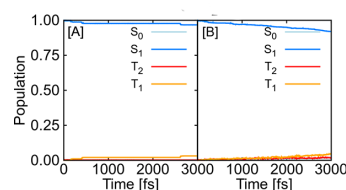
Figure 3. Quantum populations of populations up to 3000 fs of the thioformaldehyde molecule obtained from (A) 100 trajectories using CASSCF(6,5)/def2-SVP (QC) and (B) 9590 trajectories using SchNarc (ML).

reproduce the reference dynamics. Also the slow population can be reproduced accurately, which proves the validity of the ML approach.

In summary, the SchNarc framework combines the SHARC[37] approach for surface hopping and the SchNet[80] approach for ML and introduces several methodological developments, which simplify the use of nonadiabatic ML dynamics substantially. SchNarc takes advantage of SchNet's automatic generation of representations for the molecular structure and extends it to excited states. The training of ML models is facilitated by using the phase-less loss and the NAC approximation, avoiding quantum chemical NAC calculations at all. Thus, photodynamics simulations are possible based on solely ML PESs, their derivatives, and SOCs. Furthermore, this method allows for an efficient computation of the Hessians of all the excited states at each time step. Hence, SchNarc allows for efficient nonadiabatic dynamics simulations of excited states and light-induced processes including internal conversion and intersystem crossing.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpclett.0c00527.

> SchNet for excited states, training sets and reference computations, nonadiabatic couplings, and quality of ML models (PDF)
>
> Calculated frequencies and normal modes of SO_2, $CH_2NH_2^+$, and CSH_2 with ML (readable with standard programs for quantum chemistry) (ZIP)

■ AUTHOR INFORMATION

Corresponding Authors

    Michael Gastegger − *Machine Learning Group, Technical University of Berlin, 10587 Berlin, Germany;* Email: michael.gastegger@tu-berlin.de

    Philipp Marquetand − *Institute of Theoretical Chemistry, Faculty of Chemistry, Vienna Research Platform on Accelerating Photoreaction Discovery, and Data Science @ Uni Vienna, University of Vienna, 1090 Vienna, Austria;* ⓞ orcid.org/0000-0002-8711-1533; Email: philipp.marquetand@univie.ac.at

Author

    Julia Westermayr − *Institute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, 1090 Vienna, Austria*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpclett.0c00527

Notes

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Cerullo, G.; Polli, D.; Lanzani, G.; De Silvestri, S.; Hashimoto, H.; Cogdell, R. J. Photosynthetic Light Harvesting by Carotenoids: Detection of an Intermediate Excited State. *Science* **2002**, *298*, 2395−2398.

(2) Schultz, T.; Samoylova, E.; Radloff, W.; Hertel, I. V.; Sobolewski, A. L.; Domcke, W. Efficient Deactivation of a Model Base Pair via Excited-State Hydrogen Transfer. *Science* **2004**, *306*, 1765−1768.

(3) Schreier, W. J.; Schrader, T. E.; Koller, F. O.; Gilch, P.; Crespo-Hernández, C. E.; Swaminathan, V. N.; Carell, T.; Zinth, W.; Kohler, B. Thymine Dimerization in DNA Is an Ultrafast Photoreaction. *Science* **2007**, *315*, 625−629.

(4) Rauer, C.; Nogueira, J. J.; Marquetand, P.; González, L. Cyclobutane Thymine Photodimerization Mechanism Revealed by Nonadiabatic Molecular Dynamics. *J. Am. Chem. Soc.* **2016**, *138*, 15911−15916.

(5) Romero, E.; Novoderezhkin, V. I.; van Grondelle, R. Quantum Design of Photosynthesis for Bio-Inspired Solar-Energy Conversion. *Nature* **2017**, *543*, 355−365.

(6) Mathew, S.; Yella, A.; Gao, P.; Humphry-Baker, R.; Curchod, B. F. E.; Ashari-Astani, N.; Tavernelli, I.; Rothlisberger, U.; Nazeeruddin, M. K.; Grätzel, M. Dye-Sensitized Solar Cells with 13% Efficiency Achieved Through the Molecular Engineering of Porphyrin Sensitizers. *Nat. Chem.* **2014**, *6*, 242−247.

(7) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine Learning Unifies the Modeling of Materials and Molecules. *Sci. Adv.* **2017**, *3*, e1701816.

(8) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering. *Science* **2018**, *361*, 360−365.

(9) Ahmad, I.; Ahmed, S.; Anwar, Z.; Sheraz, M. A.; Sikorski, M. Photostability and Photostabilization of Drugs and Drug Products. *Int. J. Photoenergy* **2016**, *2016*, 1−19.

(10) Köuppel, H.; Domcke, W.; Cederbaum, L. S. Multimode molecular dynamics beyond the Born-Oppenheimer approximation. *Adv. Chem. Phys.* **2007**, *57*, 59−246.

(11) Köppel, H.; Gronki, J.; Mahapatra, S. Construction Scheme for Regularized Diabatic States. *J. Chem. Phys.* **2001**, *115*, 2377−2388.

(12) Worth, G. A.; Cederbaum, L. S. Beyond Born-Oppenheimer: Molecular Dynamics Through a Conical Intersection. *Annu. Rev. Phys. Chem.* **2004**, *55*, 127−158.

(13) Bowman, J. M.; Carrington, T.; Meyer, H. Variational Quantum Approaches for Computing Vibrational Energies of Polyatomic Molecules. *Mol. Phys.* **2008**, *106*, 2145−2182.

(14) Meyer, H.-D.; Gatti, F.; Worth, G. A. *Multidimensional Quantum Dynamics*; Wiley-VCH Verlag GmbH & Co. KGaA, 2009.

(15) Alborzpour, J. P.; Tew, D. P.; Habershon, S. Efficient and Accurate Evaluation of Potential Energy Matrix Elements for Quantum Dynamics Using Gaussian Process Regression. *J. Chem. Phys.* **2016**, *145*, 174112.

(16) Richings, G. W.; Habershon, S. Direct grid-Based Quantum Dynamics on Propagated Diabatic Potential Energy Surfaces. *Chem. Phys. Lett.* **2017**, *683*, 228−233.

(17) Liu, F.; Du, L.; Zhang, D.; Gao, J. Direct Learning Hidden Excited State Interaction Patterns from ab initio Dynamics and Its Implication as Alternative Molecular Mechanism Models. *Sci. Rep.* **2017**, *7*, 8737.

(18) Richings, G. W.; Habershon, S. MCTDH On-the-Fly: Efficient Grid-Based Quantum Dynamics Without Pre-Computed Potential Energy Surfaces. *J. Chem. Phys.* **2018**, *148*, 134116.

(19) Williams, D. M. G.; Eisfeld, W. Neural Network Diabatization: A New Ansatz for Accurate High-Dimensional Coupled Potential Energy Surfaces. *J. Chem. Phys.* **2018**, *149*, 204106.

(20) Xie, C.; Zhu, X.; Yarkony, D. R.; Guo, H. Permutation Invariant Polynomial Neural Network Approach to Fitting Potential Energy Surfaces. IV. Coupled Diabatic Potential Energy Matrices. *J. Chem. Phys.* **2018**, *149*, 144107.

(21) Guan, Y.; Zhang, D. H.; Guo, H.; Yarkony, D. R. Representation of Coupled Adiabatic Potential Energy Surfaces Using Neural Network Based Quasi-Diabatic Hamiltonians: 1,2 2A' States of LiFH. *Phys. Chem. Chem. Phys.* **2019**, *21*, 14205−14213.

(22) Richings, G. W.; Robertson, C.; Habershon, S. Improved On-the-Fly MCTDH Simulations with Many-Body-Potential Tensor Decomposition and Projection Diabatization. *J. Chem. Theory Comput.* **2019**, *15*, 857−870.

(23) Polyak, I.; Richings, G. W.; Habershon, S.; Knowles, P. J. Direct Quantum Dynamics Using Variational Gaussian Wavepackets and Gaussian Process Regression. *J. Chem. Phys.* **2019**, *150*, No. 041101.

(24) Guan, Y.; Guo, H.; Yarkony, D. R. Neural Network Based Quasi-Diabatic Hamiltonians with Symmetry Adaptation and a Correct Description of Conical Intersections. *J. Chem. Phys.* **2019**, *150*, 214101.

(25) Wang, Y.; Xie, C.; Guo, H.; Yarkony, D. R. A Quasi-Diabatic Representation of the 1,21A States of Methylamine. *J. Phys. Chem. A* **2019**, *123*, 5231−5241.

(26) Guan, Y.; Guo, H.; Yarkony, D. R. Extending the Representation of Multistate Coupled Potential Energy Surfaces to Include Properties Operators Using Neural Networks: Application to the 1,21A States of Ammonia. *J. Chem. Theory Comput.* **2020**, *16*, 302−313.

(27) Behler, J.; Reuter, K.; Scheffler, M. Nonadiabatic Effects in the Dissociation of Oxygen Molecules at the Al(111) Surface. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *77*, 115421.

(28) Carbogno, C.; Behler, J.; Reuter, K.; Groß, A. Signatures of Nonadiabatic O_2 Dissociation at Al(111): First-Principles Fewest-Switches Study. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81*, No. 035410.

(29) Hu, D.; Xie, Y.; Li, X.; Li, L.; Lan, Z. Inclusion of Machine Learning Kernel Ridge Regression Potential Energy Surfaces in On-the-Fly Nonadiabatic Molecular Dynamics Simulation. *J. Phys. Chem. Lett.* **2018**, *9*, 2725−2732.

(30) Dral, P. O.; Barbatti, M.; Thiel, W. Nonadiabatic Excited-State Dynamics with Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 5660−5663.

(31) Chen, W.-K.; Liu, X.-Y.; Fang, W.-H.; Dral, P. O.; Cui, G. Deep Learning for Nonadiabatic Excited-State Dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 6702−6708.

(32) Westermayr, J.; Gastegger, M.; Menger, M. F. S. J.; Mai, S.; González, L.; Marquetand, P. Machine Learning Enables Long Time Scale Molecular Photodynamics Simulations. *Chem. Sci.* **2019**, *10*, 8100−8107.

(33) Westermayr, J.; Faber, F.; Christensen, A. S.; von Lilienfeld, A.; Marquetand, P. Neural Networks and Kernel Ridge Regression for Excited States Dynamics of $CH_2NH_2^+$: From Single-State to Multi-State Representations and Multi-Property Machine Learning Models. *Mach. Learn.: Sci. Technol.* **2020**, in press, DOI: 10.1088/2632-2153/ab88d0.

(34) Ibele, L. M.; Nicolson, A.; Curchod, B. F. E. Excited-State Dynamics of Molecules with Classically Driven Trajectories and Gaussians. *Mol. Phys.* **2019**, 1.

(35) Tully, J. C. Molecular Dynamics with Electronic Transitions. *J. Chem. Phys.* **1990**, *93*, 1061−1071.

(36) Richter, M.; Marquetand, P.; González-Vázquez, J.; Sola, I.; González, L. SHARC: Ab Initio Molecular Dynamics with Surface Hopping in the Adiabatic Representation Including Arbitrary Couplings. *J. Chem. Theory Comput.* **2011**, *7*, 1253−1258.

(37) Mai, S.; Marquetand, P.; González, L. Nonadiabatic Dynamics: The SHARC Approach. *WIREs Comput. Mol. Sci.* **2018**, *8*, No. e1370.

(38) Mai, S.; Richter, M.; Ruckenbauer, M.; Oppel, M.; Marquetand, P.; González, L. SHARC2.0: Surface Hopping Including Arbitrary Couplings − Program Package for Non-Adiabatic Dynamics. sharc-md.org, 2018.

(39) Doltsinis, N. L. *Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum-Classical Approaches*; NIC Series; John von Neumann Institute for Computing, 2006.

(40) Ha, J.-K.; Lee, I. S.; Min, S. K. Surface Hopping Dynamics beyond Nonadiabatic Couplings for Quantum Coherence. *J. Phys. Chem. Lett.* **2018**, *9*, 1097−1104.

(41) Mai, S.; González, L. Molecular Photochemistry: Recent Developments in Theory. *Angew. Chem., Int. Ed.* in press. **2020** DOI: 10.1002/anie.201916381

(42) Baer, M. Introduction to the Theory of Electronic Non-Adiabatic Coupling Terms in Molecular Systems. *Phys. Rep.* **2002**, *358*, 75−142.

(43) Lischka, H.; Dallos, M.; Szalay, P. G.; Yarkony, D. R.; Shepard, R. Analytic Evaluation of Nonadiabatic Coupling Terms at the MR-CI Level. I. Formalism. *J. Chem. Phys.* **2004**, *120*, 7322−7329.

(44) Rubbmark, J. R.; Kash, M. M.; Littman, M. G.; Kleppner, D. Dynamical Effects at Avoided Level Crossings: A Study of the Landau-Zener Effect Using Rydberg Atoms. *Phys. Rev. A: At., Mol, Opt. Phys.* **1981**, *23*, 3107−3117.

(45) Hagedorn, G. A. Proof of the Landau-Zener Formula in an Adiabatic Limit with Small Eigenvalue Gaps. *Commun. Math. Phys.* **1991**, *136*, 433−449.

(46) Wittig, C. The Landau-Zener Formula. *J. Phys. Chem. B* **2005**, *109*, 8428−8430.

(47) Nakamura, H. *Nonadiabatic Transition: Concepts, Basic Theories and Applications*, 2nd edition; World Scientific: Singapore, 2012. DOI: 10.1142/8009.

(48) Zhu, C.; Kamisaka, H.; Nakamura, H. Significant Improvement of the Trajectory Surface Hopping Method by the Zhu−Nakamura Theory. *J. Chem. Phys.* **2001**, *115*, 11036−11039.

(49) Zhu, C.; Kamisaka, H.; Nakamura, H. New Implementation of the Trajectory Surface Hopping Method with Use of the Zhu−Nakamura Theory. II. Application to the Charge Transfer Processes in the 3D DH_2^+ System. *J. Chem. Phys.* **2002**, *116*, 3234−3247.

(50) Kondorskiy, A.; Nakamura, H. Semiclassical Theory of Electronically Nonadiabatic Chemical Dynamics: Incorporation of the Zhu−Nakamura Theory into the Frozen Gaussian Propagation Method. *J. Chem. Phys.* **2004**, *120*, 8937−8954.

(51) Oloyede, P.; Mil'nikov, G.; Nakamura, H. Generalized Trajectory Surface Hopping Method Based on the Zhu-Nakamura Theory. *J. Chem. Phys.* **2006**, *124*, 144110.

(52) Ishida, T.; Nanbu, S.; Nakamura, H. Clarification of Nonadiabatic Chemical Dynamics by the Zhu-Nakamura Theory of Nonadiabatic Transition: From Tri-Atomic Systems to Reactions in Solutions. *Int. Rev. Phys. Chem.* **2017**, *36*, 229−285.

(53) Granucci, G.; Persico, M.; Spighi, G. Surface Hopping Trajectory Simulations with Spin-Orbit and Dynamical Couplings. *J. Chem. Phys.* **2012**, *137*, 22A501.

(54) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Phys. Rev. Lett.* **2010**, *104*, 136403.

(55) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015**, *114*, No. 096405.

(56) Gastegger, M.; Marquetand, P. High-Dimensional Neural Network Potentials for Organic Reactions and an Improved Training Algorithm. *J. Chem. Theory Comput.* **2015**, *11*, 2187−2198.

(57) Rupp, M. Machine Learning for Quantum Mechanics in a Nutshell. *Int. J. Quantum Chem.* **2015**, *115*, 1058−1073.

(58) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145*, 170901.

(59) Gastegger, M.; Kauffmann, C.; Behler, J.; Marquetand, P. Comparing the Accuracy of High-Dimensional Neural Network Potentials and the Systematic Molecular Fragmentation Method: A benchmark Study for All-Trans Alkanes. *J. Chem. Phys.* **2016**, *144*, 194110.

(60) Gastegger, M.; Behler, J.; Marquetand, P. Machine Learning Molecular Dynamics for the Simulation of Infrared Spectra. *Chem. Sci.* **2017**, *8*, 6924−6935.

(61) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *95*, No. 094203.

(62) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511−522.

(63) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192−3203.

(64) Behler, J. First Principles Neural Network Potentials for Reactive Simulations of Large Molecular and Condensed Systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828−12840.

(65) Zong, H.; Pilania, G.; Ding, X.; Ackland, G. J.; Lookman, T. Developing an Interatomic Potential for Martensitic Phase Transformations in Zirconium by Machine Learning. *npj Comput. Mater.* **2018**, *4*, 48.

(66) Bartók, A. P.; Kermode, J.; Bernstein, N.; Csányi, G. Machine Learning a General-Purpose Interatomic Potential for Silicon. *Phys. Rev. X* **2018**, *8*, No. 041048.

(67) Chmiela, S.; Sauceda, H. E.; Müller, K.-R.; Tkatchenko, A. Towards Exact Molecular Dynamics Simulations with Machine-Learned Force Fields. *Nat. Commun.* **2018**, *9*, 3887.

(68) Imbalzano, G.; Anelli, A.; Giofré, D.; Klees, S.; Behler, J.; Ceriotti, M. Automatic Selection of Atomic Fingerprints and Reference Configurations for Machine-Learning Potentials. *J. Chem. Phys.* **2018**, *148*, 241730.

(69) Zhang, L.; Han, J.; Wang, H.; Saidi, W. A.; Car, R.; Weinan, E. End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems. *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*; United States, 2018.

(70) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.

(71) Chan, H.; Narayanan, B.; Cherukara, M. J.; Sen, F. G.; Sasikumar, K.; Gray, S. K.; Chan, M. K. Y.; Sankaranarayanan, S. K. R. S. Machine Learning Classical Interatomic Potentials for Molecular Dynamics from First-Principles Training Data. *J. Phys. Chem. C* **2019**, *123*, 6941−6957.

(72) Christensen, A. S.; Faber, F. A.; von Lilienfeld, O. A. Operators in Quantum Machine Learning: Response Properties in Chemical Space. *J. Chem. Phys.* **2019**, *150*, No. 064105.

(73) Wang, H.; Yang, W. Toward Building Protein Force Fields by Residue-Based Systematic Molecular Fragmentation and Neural Network. *J. Chem. Theory Comput.* **2019**, *15*, 1409−1417.

(74) Chmiela, S.; Sauceda, H. E.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. sGDML: Constructing Accurate and Data Efficient Molecular Force Fields Using Machine Learning. *Comput. Phys. Commun.* **2019**, *240*, 38−45.

(75) Carleo, G.; Cirac, I.; Cranmer, K.; Daudet, L.; Schuld, M.; Tishby, N.; Vogt-Maranto, L.; Zdeborová, L. Machine Learning and the Physical Sciences. *Rev. Mod. Phys.* **2019**, *91*, No. 045002.

(76) Krems, R. V. Bayesian Machine Learning for Quantum Molecular Dynamics. *Phys. Chem. Chem. Phys.* **2019**, *21*, 13392−13410.

(77) Deringer, V. L.; Caro, M. A.; Csányi, G. Machine Learning Interatomic Potentials as Emerging Tools for Materials Science. *Adv. Mater.* **2019**, *31*, 1902765.

(78) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying Machine Learning and Quantum Chemistry with a Deep Neural Network for Molecular Wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.

(79) Schütt, K. T.; Sauceda, H. E.; Kindermans, P.-J.; Tkatchenko, A.; Müller, K.-R. SchNet − A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722.

(80) Schütt, K. T.; Kessel, P.; Gastegger, M.; Nicoli, K. A.; Tkatchenko, A.; Müller, K.-R. SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **2019**, *15*, 448−455.

(81) Singraber, A.; Morawietz, T.; Behler, J.; Dellago, C. Parallel Multistream Training of High-Dimensional Neural Network Potentials. *J. Chem. Theory Comput.* **2019**, *15*, 3075−3092.

(82) Yao, K.; Herr, J. E.; Toth, D.; Mckintyre, R.; Parkhill, J. The TensorMol-0.1 Model Chemistry: A Neural Network Augmented with Long-Range Physics. *Chem. Sci.* **2018**, *9*, 2261−2269.

(83) Wang, H.; Zhang, L.; Han, J.; E, W. DeePMD-kit: A Deep Learning Package for Many-Body Potential Energy Representation and Molecular Dynamics. *Comput. Phys. Commun.* **2018**, *228*, 178−184.

(84) Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E. Neural Message Passing for Quantum Chemistry. *Proceedings of the 34th International Conference on Machine Learning*; International Convention Centre, Sydney, Australia, 2017; pp 1263−1272.

(85) Schütt, K. T.; Arbabzadah, F.; Chmiela, S.; Müller, K. R.; Tkatchenko, A. Quantum-Chemical Insights from Deep Tensor Neural Networks. *Nat. Commun.* **2017**, *8*, 13890.

(86) Unke, O. T.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.

(87) Lubbers, N.; Smith, J. S.; Barros, K. Hierarchical Modeling of Molecular Energies Using a Deep Neural Network. *J. Chem. Phys.* **2018**, *148*, 241715.

(88) Häse, F.; Valleau, S.; Pyzer-Knapp, E.; Aspuru-Guzik, A. Machine Learning Exciton Dynamics. *Chem. Sci.* **2016**, *7*, 5139−5147.

(89) Ramakrishnan, R.; von Lilienfeld, O. A. Many Molecular Properties from One Kernel in Chemical Space. *Chimia* **2015**, *69*, 182−186.

(90) Zubatyuk, R.; Smith, J. S.; Leszczynski, J.; Isayev, O. Accurate and Transferable Multitask Prediction of Chemical Properties with an Atoms-in-Molecules Neural Network. *Sci. Adv.* **2019**, *5*, eaav6490.

(91) Akimov, A. V. A Simple Phase Correction Makes a Big Difference in Nonadiabatic Molecular Dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 6096−6102.

(92) Bellonzi, N.; Medders, G. R.; Epifanovsky, E.; Subotnik, J. E. Configuration Interaction Singles with Spin-Orbit Coupling: Constructing Spin-Adiabatic States and their Analytical Nuclear Gradients. *J. Chem. Phys.* **2019**, *150*, No. 014106.

(93) Lischka, H.; et al. High-Level Multireference Methods in the Quantum-Chemistry Program System COLUMBUS: Analytic MR-CISD and MR-AQCC Gradients and MR-AQCC-LRT for Excited States, GUGA Spin-Orbit CI and Parallel CI Density. *Phys. Chem. Chem. Phys.* **2001**, *3*, 664−673.

(94) Mai, S.; Atkins, A. J.; Plasser, F.; González, L. The Influence of the Electronic Structure Method on Intersystem Crossing Dynamics. The Case of Thioformaldehyde. *J. Chem. Theory Comput.* **2019**, *15*, 3470−3480.

(95) Werner, H.-J. et al. 2012; see http://www.molpro.net.

(96) Zhang, Y.; Maurer, R. J.; Jiang, B. Symmetry-Adapted High Dimensional Neural Network Representation of Electronic Friction Tensor of Adsorbates on Metals. *J. Phys. Chem. C* **2020**, *124*, 186−195.

(97) Thiel, A.; Köppel, H. Proposal and Numerical Test of a Simple Diabatization Scheme. *J. Chem. Phys.* **1999**, *110*, 9371−9383.

(98) Köppel, H.; Schubert, B. The Concept of Regularized Diabatic States for a General Conical Intersection. *Mol. Phys.* **2006**, *104*, 1069−1079.

(99) Maeda, S.; Ohno, K.; Morokuma, K. Updated Branching Plane for Finding Conical Intersections without Coupling Derivative Vectors. *J. Chem. Theory Comput.* **2010**, *6*, 1538−1545.

(100) Kammeraad, J. A.; Zimmerman, P. M. Estimating the Derivative Coupling Vector Using Gradients. *J. Phys. Chem. Lett.* **2016**, *7*, 5074−5079.

(101) Gonon, B.; Perveaux, A.; Gatti, F.; Lauvergnat, D.; Lasorne, B. On the Applicability of a Wavefunction-Free, Energy-Based Procedure for Generating First-Order Non-Adiabatic Couplings Around Conical Intersections. *J. Chem. Phys.* **2017**, *147*, 114114.

(102) Frankcombe, T. J. Using Hessian Update Formulae to Construct Modified Shepard Interpolated Potential Energy Surfaces: Application to Vibrating Surface Atoms. *J. Chem. Phys.* **2014**, *140*, 114108.

(103) Richings, G.; Polyak, I.; Spinlove, K.; Worth, G.; Burghardt, I.; Lasorne, B. Quantum dynamics simulations using Gaussian wave-packets: the vMCG method. *Int. Rev. Phys. Chem.* **2015**, *34*, 269−308.

(104) An, H.; Baeck, K. K. Practical and Reliable Approximation of Nonadiabatic Coupling Terms between Triplet Electronic States Using Only Adiabatic Potential Energies. *Chem. Phys. Lett.* **2018**, *696*, 100−105.

(105) Baeck, K. K.; An, H. Practical Approximation of the Non-Adiabatic Coupling Terms for Same-Symmetry Interstate Crossings by Using Adiabatic Potential Energies Only. *J. Chem. Phys.* **2017**, *146*, No. 064107.

(106) Barbatti, M.; Aquino, A. J. A.; Lischka, H. Ultrafast Two-Step Process in the Non-Adiabatic Relaxation of the CH_2NH_2 Molecule. *Mol. Phys.* **2006**, *104*, 1053−1060.

(107) Hjorth Larsen, A.; et al. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys.: Condens. Matter* **2017**, *29*, 273002.

110

JULIA WESTERMAYR AND PHILIPP MARQUETAND

"Machine Learning for Nonadiabatic Molecular Dynamics" in *Machine Learning in Chemistry: The Impact of Artificial Intelligence*, Hugh M. Cartwright, The Royal Society of Chemistry, in press (2020).
https://pubs.rsc.org/en/content/ebook/978-1-78801-789-3

Contributions:

JULIA WESTERMAYR wrote the initial manuscript and contributed to the writing of the final manuscript.

PHILIPP MARQUETAND conceived the scope of the manuscript and contributed to the initial and final writing of the manuscript.

CHAPTER 4

Machine Learning for Nonadiabatic Molecular Dynamics

JULIA WESTERMAYR(0000-0002-6531-0742)[a] AND
PHILIPP MARQUETAND(0000-0002-8711-1533)*[a,b,c]

[a] University of Vienna, Faculty of Chemistry, Institute of Theoretical
Chemistry, Währinger Str. 17, 1090 Vienna, Austria; [b] Vienna Research
Platform on Accelerating Photoreaction Discovery, University of Vienna,
Währinger Str. 17, 1090 Vienna, Austria; [c] University of Vienna, Faculty
of Chemistry, Data Science @ Uni Vienna, Währinger Str. 29, 1090 Vienna,
Austria
*Email: philipp.marquetand@univie.ac.at

4.1 Introduction

Nonadiabatic molecular dynamics simulations (NAMD) go beyond Born–Oppenheimer molecular dynamics by including two or more electronic states that are coupled. By doing so, many fundamental processes that are at the origin of life and death on earth, such as photosynthesis or DNA photodamage, can be studied.[1–3] To name only a few applications, getting to know the relation between a molecular structure and its photochemical properties can help to design new materials for photovoltaics[4,5] or drugs for phototherapy.[6]

Despite the importance of NAMD simulations, they are limited by the expense of the underlying quantum chemical computation of the high-dimensional potential energy surfaces (PESs), *i.e.,* the energetic states for

76

Machine Learning for Nonadiabatic Molecular Dynamics 77

corresponding molecular configurations that are involved in the targeted re-
action. This computation is the most expensive part of a NAMD simulation and
hampers its application for long time scales. With contemporary approaches,
up to a couple of picoseconds can be simulated and a compromise between
accuracy and computational efficiency usually has to be made.[7–9] To speed up
simulations and make long time scales as well as accurate PESs accessible, the
expensive quantum chemical calculations have to be replaced by more efficient
methods. For the electronic ground state force fields exist,[10–13] but these
introduce a lot of approximations leading to less accurate potentials. Recently,
with the rise of machine learning (ML), ML potentials have been successfully
developed for the electronic ground state[14–39] that allow an accurate repre-
sentation of the PES and the investigation of chemical reactions, which is not
possible with standard force fields. However, no such force fields for excited
states are available, and only a few studies have tackled the development of
ML potentials for excited-state properties.[40–50] The reason for the slow devel-
opment of ML potentials for the excited state to carry out dynamics simu-
lations compared with the ML potentials for the ground state is two-fold. On
one side, not just one PES must be accurately described; more need to be taken
into account. Further, not only the energies and corresponding forces are
necessary to carry out accurate NAMD simulations, but also couplings between
different states; this leads to additional challenges for the regression with ML,
as will be explained below.

  Almost all approaches in quantum chemistry are based on the Born–
Oppenheimer approximation that allows one to separate the degrees of
freedom of the electrons and the nuclei. The Born–Oppenheimer approxi-
mation is valid for many geometries of the electronic ground state, in par-
ticular for the equilibrium of most molecules. However, it breaks down
whenever two or more states are strongly coupled and the electrons and the
nuclei rearrange on an ultra-short time scale.[51–53] Nevertheless, the use of
this approximation is crucial for efficient computations of NAMD. It allows
for mixed quantum-classical dynamics, where the motion of the nuclei on
the PESs made up by the electrons is treated classically, while the electronic
motion follows the time-dependent Schrödinger equation. The underlying
elements that couple different PESs show singularities in regions where PESs
are close to each other and are almost vanishing elsewhere. In the con-
formational space, where states of the same spin multiplicity get into
proximity, nonadiabatic couplings (NACs) are important, while spin–orbit
couplings play a key role for transitions between states of different spin
multiplicities. These critical regions are called conical intersections or
simply state crossings, respectively, and do not only pose an obstacle to
quantum chemistry but consequently also to the fitting of the PESs.[44–46,50]

  The numerical difficulties that arise due to the singularities of NACs at
conical intersections lead to other approaches that focus on the diabatiza-
tion of adiabatic potentials.[48,54–57] Using diabatic potentials is favourable, as
the off-diagonal coupling elements of the Hamiltonian are smoothly varying
properties. Another advantage is that effects due to the Berry (or geometric)

phase vanish and do not influence dynamics simulations. The importance of the Berry phase depends on the dynamics method and the chemical system under study. While the Berry phase can be safely neglected in some cases,[50,58] it plays a vital role in others, leading to path-dependent transition probabilities in the vicinity of conical intersections.[59] For the latter cases, a diabatic approach is desirable, but finding meaningful diabatic potentials is often a tedious task.[48] The output of a quantum chemistry calculation is given in the adiabatic picture, as the diabatic wavefunction is not an eigenfunction of the electronic Hamiltonian. Hence, the diabatic potentials are not unique and it often turns out to be infeasible to find diabatic potentials, especially for larger and more complex molecules. One reason is that the number of electronic states within a certain energy range often increases with the number of atoms in the molecule, therefore increasing the computational effort to provide all desired electronic states. Recently, some approaches were developed to improve the diabatization process by using ML models.[47–49] In large molecules, *i.e.,* in the range of 10s to 100s of atoms, more extensive approximations,[47,60] such as the linear vibronic coupling model,[61] should be considered. As their discussion goes beyond the scope of this chapter, the reader is referred to ref. 54 and 61–64 for more details. Although excited-state dynamics are favourable in the diabatic picture, due to the aforementioned problems and the fact that the computation of diabatic potentials is still a challenging task, on-the-fly NAMD in the adiabatic representation remains often the method of choice.

Within NAMD simulations, different approximations exist to account for nonadiabatic transitions between different adiabatic states. The surface-hopping methodology is often a good compromise between accuracy and efficiency because it computes the PESs and corresponding properties on-the-fly at geometries obtained after classical propagation of the nuclei.[65,66] This method has been the method of choice in recent ML-driven NAMD simulations[44–46,50] and allows for an on-the-fly exploration of the PESs with ML models.[44,50] Transitions, or the so-called hops, between different PESs are determined stochastically. Different algorithms[58,66,67] to compute the hopping probability and its direction are available with the most popular one being based on Tully's fewest switching algorithm.[68] In this case, the hopping probability is dependent on the couplings between adjacent states.[65] Another commonly used algorithm is the Zhu–Nakamura approximation,[58] that omits the computation of any couplings. For more information on hopping algorithms, the reader is referred to ref. 68–77.

Here, we discuss several ML approaches for the computation of NAMD simulations. In the following sections, different algorithms, *i.e.,* a linear model (LR), kernel-ridge regression (KRR), support vector regression (SVR), and artificial neural networks (NNs), will be compared and their advantages and disadvantages will be listed. It is further shown how to efficiently generate a training set by keeping in mind that the number of costly quantum chemical reference calculations should be small. The problem of wave function phases will be discussed and its impact on couplings, and in

general any excited-state properties corresponding to off-diagonal elements in matrix notation, such as transition dipole moments. As an example, the methylenimmonium cation is selected, where the training set is taken from ref. 50 and contains energies, corresponding forces, NACs as well as permanent and transition dipole moments for three singlet states. However, the latter property will not be discussed in the course of this chapter.

4.2   Methods

4.2.1   Machine Learning (ML) Models

As ML models, KRR, SVR, and NNs are discussed and compared with LR as a baseline model. The focus of this chapter is on NNs. In case of LR, SVR, and KRR, each electronic state is independently modelled, so for three singlet states of the methylenimmonium cation containing six atoms, the respective state energies must be determined in three separate fits for one molecular geometry, the gradients (if not fitted as energy derivatives) as single values 18 times, and the NACs as single values 162 times. By using the relations of NAC vectors between two states, i and j, where $NAC_{ij} = -NAC_{ji}$ as well as $NAC_{ii} = 0$, the parameters can be reduced to 54. As NNs can relate all energetic states to one molecular geometry at once, the main focus will be on this model, which we call a multi-state model. The gradients are treated as derivatives of the PESs[20] and NACs are treated together in one additional NN.

4.2.1.1   *Linear Model (LR)*

The central goal of supervised ML, and in general of regression, is to fit a function to find the best possible relation between an input, *i.e.*, the molecular geometry, and its corresponding output,[78] in our case the energies and corresponding forces for each state, as well as NACs between each set of states. In the simplest case, the relationship can be approximated by a linear function,

$$Y = b + w \cdot X, \tag{4.1}$$

where Y is the output property, X is the data to be fitted, w are the coefficients, and b is a constant bias. Ordinary least squares regression is used for the linear model as implemented in scikit-learn.[79] To fit the PESs, a training set is used that contains X as well as corresponding Ys. The learning process is based on finding the parameters, w and b, where the error between a predicted property, Y^p, by the model and the original value, Y, for a given input, is minimal.[80]

The corresponding loss function reads

$$L = \sum_i^N \left(Y_i^p - Y_i\right)^2 \tag{4.2}$$

with i running over all data points N in the training set.[81] In the first step, the parameters w and b are estimated or randomly chosen. They get

optimized during a process termed training, in which the loss function is
minimized,

$$\min_w \sum_i^N \left(\langle w, Y_i^p \rangle - Y_i \right)^2.$$

(4.3)

The resulting problem is thus a minimization problem that can be solved
by following the gradient of the function with respect to the weights.

4.2.1.2   Kernel Ridge Regression (KRR)

Kernel ridge regression (KRR) is derived from ridge regression instead of LR
in combination with the kernel trick. Ridge regression adds a penalty term to
the loss function, that is minimized according to:

$$\min_w \sum_i^N \left(\langle w_i, Y_i^p \rangle - Y_i \right)^2 + \lambda \|w\|^2.$$

(4.4)

λ is a regularizer, also known as L2 penalty, that is added to the weights and
used to control overfitting.[78] As the noise in the training set made up by
quantum chemistry is supposed to be small when calculations are con-
verged, λ is also very small in these cases.[82] The kernel trick is used to map a
training set that has no linear relation between inputs and outputs into a
higher-dimensional space, where it is linearly separable. This leads to a non-
linear model with the non-linearity being determined by the type of the used
kernel, K.[78] In most cases, a Gaussian kernel or Laplacian kernel is chosen,
centred on each compound (molecule), X_i, in the training set. The prediction
of a property, Y^p, is then computed using eqn (4.5). The regression co-
efficients are again found *via* linear regression according to eqn (4.6).

$$Y^p = \sum_i^N w_i K(X, X_i)$$

(4.5)

$$w = (K + \lambda \mathbf{1})^{-1} Y$$

(4.6)

For KRR, the QML toolkit[83] in combination with the FCHL represen-
tation[82] was used relying on a Gaussian kernel function and treating gra-
dients as energy derivatives. Note that a new version of the FCHL
representation[84] already exists, allowing for more efficient computations.
KRR is often the method of choice for quantum chemical ML problems. The
advantage of this method is that it can fit non-linear data sets, while still
profiting from simplicity and ease of use.[82]

4.2.1.3   Support Vector Regression (SVR)

Support vector machines can be used for classification as well as re-
gression problems. Here, the focus is on the latter case, also known as

Machine Learning for Nonadiabatic Molecular Dynamics 81

support vector regression (SVR). Like KRR, SVR makes use of the kernel trick. Input data are mapped into a higher-dimensional feature space, where a linear model can be constructed. However, instead of a minimum loss function, an ϵ-intensive loss function, L_ϵ, is employed, where ϵ is a margin of tolerance:[85]

$$L_\epsilon = \begin{cases} 0 \text{ if } \left|Y_i^p - Y_i\right| \leq \epsilon \\ \left|Y_i^p - Y_i\right| - \epsilon \text{ otherwise} \end{cases} \tag{4.7}$$

The ϵ-insensitive region (where $L_\epsilon = 0$) is often referred to as ϵ-*tube*, where the optimization process is based on finding the tube with most of the training points inside. The support vectors represent the instances outside this tube, which have the most influence in the training process.[86] Hence, the loss function is only affected by data points that are predicted with a larger error than $\pm\epsilon$.[87] The function to be minimized is then a regularized error function:

$$\min_w C \sum_i^N L_\epsilon + \frac{1}{2}\|w\|^2 \tag{4.8}$$

where C is a regularization parameter written in front of the error term.[88]

Common kernel types, or basis functions, that are placed on each training point, are linear kernels, radial basis functions, and polynomials. For the results given here, a radial basis function is used as implemented in scikit-learn.[79] SVR has the advantage of having the capability of good generalization while achieving high prediction accuracy. However, as with the previous models that rely on basis functions that are centred on the training compounds, the depth of the model is directly dependent on the size of the training set.[86]

4.2.1.4 Neural Networks (NNs)

In contrast to LR, KRR, and SVR, neural networks (NNs) have their basic concepts from the information processing in the human brain. As can be seen from Figure 4.1, an NN consists of several layers, with an input layer, that contains the input, X, several hidden layers, each comprised of a finite number of nodes, and an output layer that gives the desired property, Y. Similar to the neurons in our brain that are connected by synapses, the nodes i and k in the hidden layers j and l are connected by weights, w_{ij}^{kl}, that are the fitting parameters during training. The nested functional form of the NN with two hidden layers and three nodes per hidden layer (as given in Figure 4.1) is presented in eqn (4.9).

$$Y^p = f_1^3\left\{b_1^3 + \sum_{l=1}^{3} w_{1k}^{23} f_k^2\left[b_k^2 + \sum_{j=1}^{3} w_{jk}^{12} f_j^1\left(b_j^1 + \sum_{i=1}^{2} w_{ij}^{01} X_i\right)\right]\right\} \tag{4.9}$$
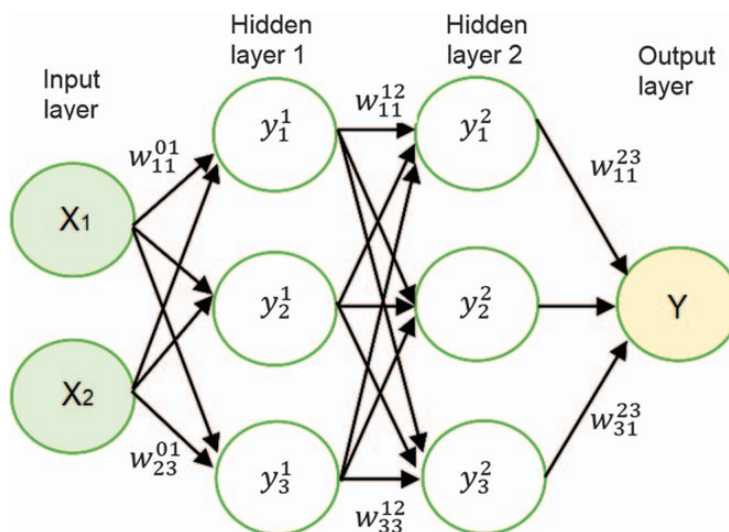
Figure 4.1 A small multi-layer feed-forward NN with an input layer that contains atomic coordinates, $X = \{X_i\}$, two hidden layers with 3 nodes, y, per hidden layer, and an output layer that yields a property, Y. A bias that is connected to each node in the hidden layers and output layer is not illustrated here.

As can be seen from eqn (4.9), the NN model can be referred to as a set of functional transformations. First, the linear combinations of the inputs are generated, $\sum_{i=1}^{2} w_{ij}^{01} X_i + b_j^1$. A non-linear basis function, f, is then employed and used to transform this expression to yield the second layer of the network. Frequently used basis functions are sigmoidal functions, tangent hyperbolic, or a shifted softplus function, such as $\ln(0.5e^x + 0.5)$, that was used for the results presented here. By a series of such transformations, the NN model can be constructed in its simple multi-layer feed-forward (MLFF) architecture. The bias values, b_i^j, are used to achieve more flexibility by allowing basis functions to shift. In the last step, the desired properties are obtained by using a linear basis function. Hence, if no hidden layers are defined, the MLFF NN is equivalent to LR in its simplest form.[88,89] As the number of hidden layers and nodes per hidden layer defines the depth of the network, the complexity and the depth of the models are independent of the training set size.

To find the best possible relation between *X* and *Y*, the loss function to be minimized is the sum of mean squared errors between NN predictions and reference values, as given in eqn (4.3). Several algorithms exist, with Adam (adaptive moment estimation)[90] being one of the most popular stochastic gradient descent optimization algorithms. To prevent the model from overfitting, different techniques can be applied, such as the inclusion of a L1 and L2 regularization, soft weight sharing, dropout or an early stopping

Machine Learning for Nonadiabatic Molecular Dynamics 83

mechanism.[91] The latter can be applied by splitting the whole training set into a training set and a validation set, whereas the validation set is excluded from training, but included for prediction. This means that after each epoch, *i.e.*, when the weights get updated, the error on the validation set is computed and recorded. Once this error increases, the model is assumed to overfit and the training needs to be stopped.[89] The NN results here were obtained using theano.[92]

4.2.1.5 Training Process

Compared with LR, KRR, and SVR, NNs contain a lot more hyperparameters that need to be optimized in order to achieve successful training and to fit the weights to accurately predict the desired property. Hence not only the type of basis function, its width and the regularization terms have to be optimized, as is the case for KRR and SVR, but also the number of nodes per layer, the number of layers and several more hyperparameters. One commonly-used way to search the space of hyperparameters is *via* random grid search.[93] For some implementations and further details, see for example ref. 9 or 50. It is further advantageous to scale inputs and outputs of the training set by subtracting the mean and dividing by the standard deviation. We used such a scaling for our NNs.

As already mentioned, the depth of the NN does not depend on the number of data points, and a deep NN can be built from any training set size. For practical reasons and memory-efficient computations, it is thus favourable to split the training set into the so-called mini batches. The batch size depends on the problem and no rule of thumb on its optimal value can be given. A batch is thus a subset of the training set, which delivers a combined gradient for the optimization of the weights *via* gradient descent. Each time a batch is used for training, the weight parameters get updated. After one pass through the complete training set, an epoch is finished; for practical reasons, an early-stopping mechanism is recommended to prevent overfitting.[93] After one epoch, the weights can be adapted, and also the learning rate, which defines the magnitude of change of the weights after each epoch and is one of the most critical hyperparameters. A large initial learning rate prevents the NN model from getting stuck at a local minimum. It can then be decreased during the optimization process. The number of hidden layers and nodes per hidden layer determines the depth of the NN. Usually, up to a certain number of hidden layers and nodes per hidden layer, the NN gets more accurate, but converges towards an optimal value in terms of accuracy and computational costs. In addition, the L1 and L2 regularization rates have to be selected and searched accordingly.

Owing to this more complex architecture, NNs are more difficult to apply in practice than LR, KRR, and SVR. However, an NN is more flexible and can fit more complex data. It is also straightforward to relate one molecular input to a multi-dimensional output. In detail, one model can predict

the values of each energetic state as well as the corresponding gradients
in one step. In this case, the training improves if the forces, F, are included
in the loss function, L. Another hyperparameter, η, that controls the in-
fluence of forces on the loss function, as described in eqn (4.10), has to
be added.

$$L = \frac{1}{N}\sum_{i}^{N}\left(E_i^p - E_i\right)^2 - \frac{\eta}{N}\sum_{i}^{N}\frac{1}{3M}\sum_{\alpha}^{3M}\left(F_i^p - F_i\right)^2 \qquad (4.10)$$

The same accounts for NACs, where one model can be built to match an
input to all the NAC vectors between different states of the same spin
multiplicity. The corresponding models are thus multi-state models for en-
ergies and gradients and multi-value models for NACs. LR, KRR, and SVR,
are single-state models by definition. To make multi-state predictions pos-
sible, an explicit encoding of the energetic state is necessary, *i.e.*, a separate
kernel for this state must be defined to predict the energies of all states with
one model.

To achieve energy conservation during dynamics simulations, the forces
should be treated as derivatives from energy potentials. This can be done for
KRR by using atomistic Gaussian kernels in combination with the FCHL
representation,[34,84] which will be explained in the following and is imple-
mented in the QML toolkit.[83]

4.2.1.6   Descriptors

To represent a molecule suitably for an ML model, a rotationally and
translationally invariant descriptor must be provided that contains all the
relevant information regarding a molecular geometry to fit a complex
property. Compared to previous ML potentials for the electronic ground
state, where a universal model is aimed for, the excited-state potentials are
molecule-specific. Hence, it is sufficient to present the molecule as a whole
to the ML models. This MLFF NN contrasts to high-dimensional NNs,[94]
where each atom type is represented by one NN, such as given in ref. 15, 20,
89, and 95–97 for example to construct ML potentials for the electronic
ground state. The target in our approach is to fit complex excited-state
properties and aim for long time scale photodynamics simulations. Several
types of descriptors are compared: The inverse distance matrix is the sim-
plest descriptor applied, which contains only information about the dis-
tances of the atoms. The Coulomb matrix[98] additionally contains atomic
charges, and a polynomial descriptor is applied, generated *via* cross-
products of all entries of the inverse distance matrix.[50] For KRR, the FCHL
representation is applied,[82] whereas LR and SVR are trained to relate the
inverse distance matrix to a desired property. In the following, the de-
scriptors will be explained in more detail with the test molecule, methyle-
nimmonium cation, $CH_2NH_2^+$, taken as the reference.

4.2.1.7 Distance-matrix Based Descriptors

The inverse distance matrix, D, will be abbreviated as inv.D. in this chapter. Its entries

$$D_{ij} = \frac{1}{\left| R_i - R_j \right|} \tag{4.11}$$

with R_i being the position of atom i and R_j the position of atom j are very similar to those of the Coulomb matrix,[98]

$$C_{ij} = \begin{cases} \dfrac{Z_i Z_j}{\left| R_i - R_j \right|} & \text{if } \delta_{ij} = 0 \\ \dfrac{1}{2} Z_i^{2.4} & \text{if } \delta_{ij} = 1. \end{cases} \tag{4.12}$$

In the latter, the atomic charges of atoms i and j, Z_i and Z_j, are additionally included. When comparing matrices, the inverse distance matrix contains 15 features for the 6-atom molecule $CH_2NH_2^+$, *i.e.,* the off-diagonal elements of the upper or lower triangular matrix. The diagonal elements of this symmetric matrix are not defined. The Coulomb matrix additionally contains constant values for the diagonal matrix elements resulting in 21 input features as a molecular descriptor for $CH_2NH_2^+$. Considering the aforementioned scaling of inputs, both descriptors are equivalent and should yield comparable results. Therefore, another descriptor is computed, consisting of polynomials of the inverse distances, abbreviated as poly.D. in the following. The dimension of this descriptor is squared compared to inv.D., thus allowing for a more accurate representation of the molecule.

4.2.1.8 FCHL: Faber-Christensen-Huang-Lilienfeld Representation

The FCHL (Faber-Christensen-Huang-Lilienfeld) representation[34,84] has the advantage over distance-matrix based descriptors that it is also permutation invariant. This invariance is achieved by describing the molecule atom-wise in its chemical and structural environment. To this aim, a set of M-body expansions, $A_M(I)$, is defined for each atom, I. Here, only the first three terms will be included, due to minor advances when accounting for higher body terms.[34] Similar to the distance-based descriptors, the FCHL representation contains information about the interatomic distances (second-order expansion, $A_2(I)$), but additionally includes information about the chemical compositions (first-order expansion, $A_1(I)$) and angular distributions (third-order expansion, $A_3(I)$). The representation up to the third-order expansion reads:

$$A_3(I) = N\left(x^{(1)}\right) \sum_{i \neq I} N\left(x_{iI}^{(2)}\right) \sum_{j \neq i,I} N\left(x_{ijI}^{(3)}\right) \xi_3\left(d_{iI}, d_{jI}, \theta_{ij}^I\right) \tag{4.13}$$

with the first-order expansion

$$N\left(x^{(1)}\right) = A_1(I) = e^{-\frac{(P_I - x_1)^2}{2\sigma_P^2} - \frac{(G_I - x_2)^2}{2\sigma_G^2}} \tag{4.14}$$

and $x^{(1)} = \{P_i, \sigma_P; G_i, \sigma_G\}$ accounting for stoichiometry. The corresponding higher terms are similarly generated with $x_{ij}^{(2)} = \{d_{iI}, \sigma_d; P_i, \sigma_P; G_i, \sigma_G\}$ and $x_{ijI}^{(3)} = \{\theta_{ij}^I, \sigma_\theta; P_j, \sigma_P; G_j, \sigma_G\}$ and are similar to the atom-centred symmetry functions for NNs of Behler[99] or their weighted variants.[97,100]

The advantage of this descriptor compared with the previously mentioned distance-based descriptors is that it treats the molecule atom-wise and thus includes more information about the molecule, *i.e.,* the atomic composition and angular distribution. The disadvantage is a higher computational expense, which is optimized in a new version, called FCHL19.[84] The previous version FCHL18, was used to obtain the results presented here.

4.2.2   Training Set Generation

A training set that is applicable for NAMD should contain a molecular geometry and the energies of all energetic states that are included in dynamics simulations, corresponding gradients as well as couplings between the states. Spin–orbit couplings are important between states of different spin-multiplicities, whereas NACs account for transitions between states of same spin-multiplicity. Here, the test system has three active singlet states and thus NACs will be described. Notably, the accuracy of the ML potentials solely relies on the quantum chemistry reference method if ML models are trained appropriately. Hence it is important to choose this method thoughtfully. To keep the number of expensive reference computations minimal, cost-effective generation of the training set is required.[50] Nevertheless, this small training set should cover the relevant conformational space of a molecule that is visited during dynamics simulations. Therefore, it is advisable to build such a training set by starting with an initial, small set that is generated manually and to expand it in a meaningful way *via* adaptive sampling.[15,20,50,89,101,102]

4.2.2.1   *Initial Training Set*

As a starting point for generating a training set for NAMD simulations, the equilibrium geometry of a molecule should be included and taken as a reference point for all subsequent calculations to compute wave function overlaps. Hence, the information of the wave function of this geometry should be saved for further pre-processing (see Section 4.2.3). In principle, every sampling method can be used for computing the rest of the data points for the initial training set. For small molecules, which do not have many degrees of freedom, a good guess is to carry out scans along each normal mode coordinate. An even better idea is to take points from the optimization of critical points like excited-state minima, conical intersections, and state crossing points. For larger systems, some other approaches like Wigner

Machine Learning for Nonadiabatic Molecular Dynamics 87

sampling[103] or sampling *via* molecular dynamics simulations,[104,105] such as umbrella sampling,[106] trajectory-guided sampling,[107] enhanced sampling[108] or metadynamics,[109] *e.g.* with the semiempirical tight-binding based quantum chemistry method GFN2-xTB[110] can be employed, to name only a few recent developments. In any case, it is advantageous to consider critical regions of the PESs or reaction coordinates that might be visited during photodynamics simulations in advance. For example, if it is known from literature or supposed by chemical intuition that photo-excitation can lead to hydrogen dissociation or carbon–carbon bond cleavage, appropriate reaction coordinates should be scanned. As soon as the initial training set is thought to be large enough (so far, we have used approximately 1000 points), ML models can be trained and dynamics simulations can be carried out using adaptive sampling for excited states,[50] as will be explained below.

4.2.2.2 Adaptive Sampling for Excited States

In general, an ML model used to replace quantum chemistry in NAMD simulations is, simply said, interpolation between data points. Therefore, it can be as accurate as the reference method while being extremely fast at the same time in interpolative regions of the PESs. However, it fails dramatically in extrapolative regions.[78] Hence, it is not sufficient to sample a conformational subspace of a molecule in advance, but it is necessary to check the reliability of ML predictions at any point in time along a dynamics run. If a region is detected to be untrustworthy, it is either undersampled or not included in the training set at all. The conformation, where the untrustworthiness was detected, should be added to the training set and the ML models should be re-trained from this expanded training set in order to continue ML-NAMD simulations accurately and on long time scales. The method of choice to expand the training set with unknown conformations on-the-fly and check the accuracy of ML predictions is in most cases a process termed adaptive sampling,[15,20,89,101,102] which is well-established for ground state dynamics, and modified for excited-state dynamics.[50] The process is exemplified in Figure 4.2.
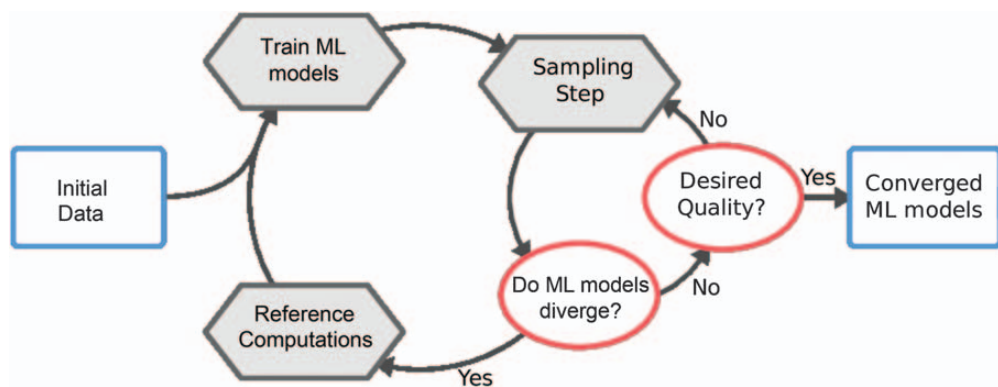
Figure 4.2 Scheme of the adaptive sampling procedure adapted from ref. 32.

In more detail, adaptive sampling for excited states starts with a NAMD simulation, where the excited-state properties, $\tilde{Y}_j^p$, (PESs, forces, and NACs), are taken as the mean, $\bar{Y}^p$, of predictions from at least two ML models, J:

$$\bar{Y}^p = \frac{1}{J}\sum_{j=1}^{J}\tilde{Y}_j^p. \tag{4.15}$$

Those ML models are independently trained from the initial training set but differ slightly in their initial hyperparameters or starting weights. At each time step, the RMSE between the ML model forecasts, Y_σ^p, is computed and compared with a predefined threshold, ε^p, which is set for each excited-state property. The thresholds are defined manually and should be in the range of the validation error of the ML models. A separate threshold for each property is used. Whenever one of these thresholds is exceeded, such that

$$Y_\sigma^p = \sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left(\tilde{Y}_j^p - \bar{Y}^p\right)^2} \geq \varepsilon^p, \tag{4.16}$$

the prediction is termed untrustworthy, and the molecular geometry is presumed to be part of an unexplored conformational space. This geometry is then used as an input to carry out an additional quantum chemistry reference calculation for each property, no matter which property's threshold is exceeded. The training set is expanded with this additional data point and the ML models are retrained from the expanded training set. After one such cycle, the thresholds are adapted by multiplication with a pre-defined factor (*e.g.* 0.95) to reduce the magnitude of the thresholds and sample the conformational space comprehensively for the training set generation. The dynamics simulations with an ensemble of newly-trained ML models can be continued from the unknown configuration, employing the adaptive sampling procedure until the training set covers the relevant space for NAMD.[20,50,89]

For practical reasons and efficient training set generation, it is favourable to start not one trajectory with an ensemble of ML models, but approximately 100 trajectories, as is usually done within NAMD. The ML models should then be retrained only after each trajectory reaches an unexplored conformational region of the PESs and the corresponding data points are all added to the training set. Hence the training of the ML models is carried out only once for all trajectories and not for each of the aborted trajectories separately. In this way, the computational cost for the training of the ML models is reduced.

Furthermore, it is advisable to recalculate the thresholds according to the error of ML models on the training set and to restart the dynamics from the beginning when the size of the training set increases by a factor of about two. Further, the hyperparameters of the ML models should be scanned and optimized again, in order to find the best fit for the expanded training set.

Machine Learning for Nonadiabatic Molecular Dynamics 89

This procedure is advisable, because the trajectories will probably not lead to the exact same conformational region, but might explore different subspaces of the PESs.[32,50]

In general, there is no guide on how large the training set should be for NAMDs. It is rather a matter of the size and flexibility of the targeted molecule as well as the complexity and diversity of the photo-induced processes that can take place. If different reaction channels are accessible after photo-excitation, far more data points might be needed than for a rigid molecule, where most trajectories lead towards the same reaction channel. Another important factor is the number of active states that are included in NAMD. By checking the convergence behaviour of the sampling procedure, an indication can be obtained on how close the training set is to its optimal size.[50]

For all those reference computations, more states might be considered than in the actual ML-NAMD, and this fact should also be kept in mind when choosing the reference method. This is especially important for excited-state properties that result from wave functions of two different electronic states, such as NACs that couple two different states or transition dipole moments. Their sign, *i.e.,* if the coupling value, C_{ij}, is $+C_{ij}$ or $-C_{ij}$, is dependent on the relative phase of the wave function of each electronic state. For very similar geometries, the couplings can arbitrarily switch their sign, leading to inconsistencies in the potentials formed. This makes it difficult for an ML model to find a relation between C_{ij} values and a molecular geometry. To allow for meaningful learning of off-diagonal elements in terms of matrix notation, a further preprocessing, termed phase correction, is carried out such that almost all such inconsistencies are removed, and a proper learning behaviour of the ML models can be achieved. To this aim, it is necessary to compute wave function overlaps between every geometry in the training set and the reference geometry defined at the beginning of the training set generation, that still contains all the information about the wave function.[50,111,112] This contrasts with the generation of a training set for Born–Oppenheimer molecular dynamics and thus will be explained in the following.

4.2.3 Wave Function Phases

The problem of phase inconsistencies throughout a set of quantum chemical reference computations can be exemplified with molecular orbitals of the methylenimmonium cation (Figure 4.3 taken from ref. 50). As can be seen in panel A of Figure 4.3, molecular geometries are the input of a quantum chemical reference computation. The interpolation coordinate here is the elongation along the C–N bond. In panel B, the molecular orbitals obtained from each quantum chemical calculation are given for the S_1 and S_2 states. As shown, they can switch their sign, illustrated by their brightness – that is either dark or light. The same accounts for the electronic wave function, which is a valid eigenfunction of the electronic Hamiltonian. If it is
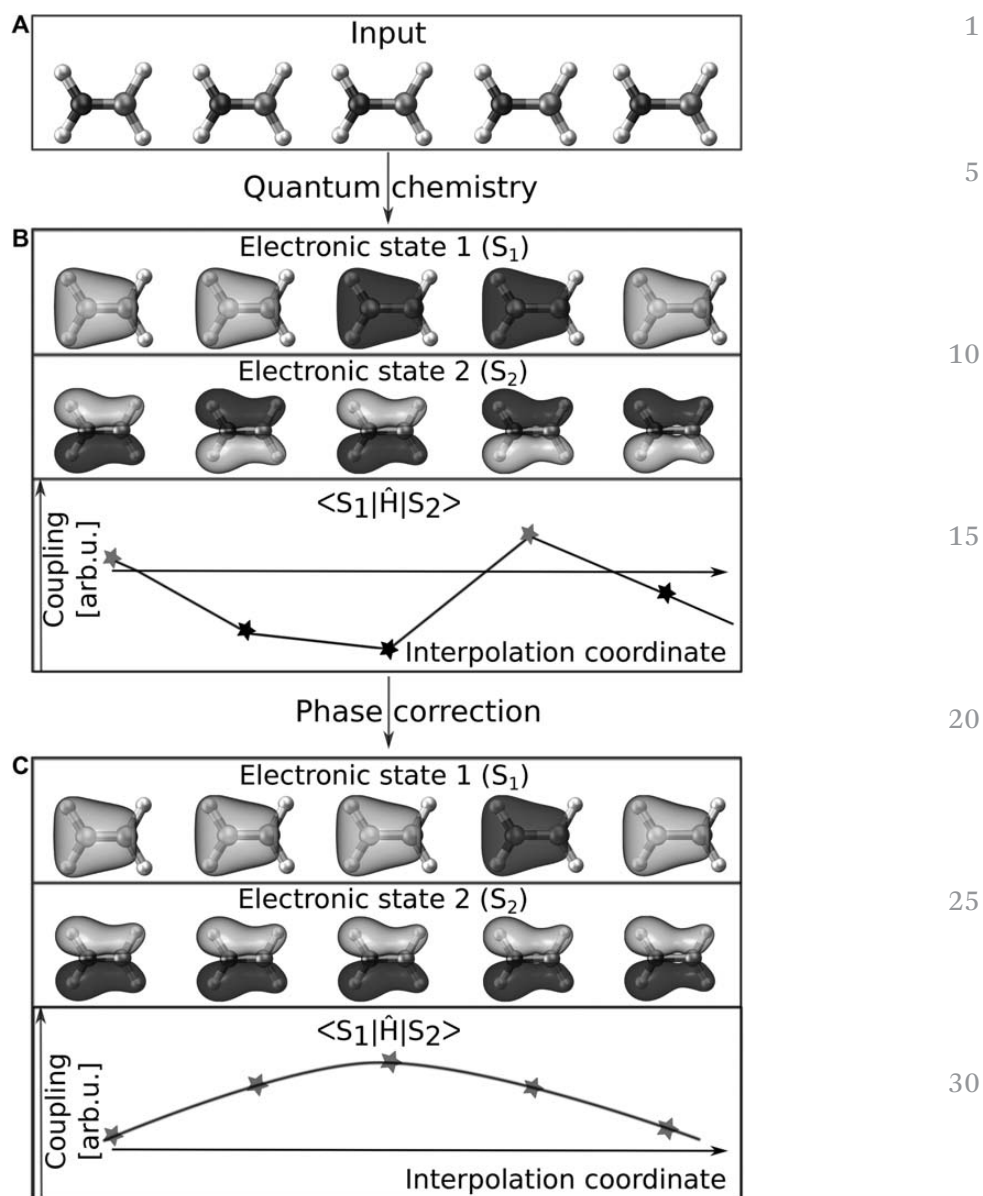
Figure 4.3 A set of quantum chemical calculations of the methylenimmonium cation, $CH_2NH_2^+$, along the C–N bond. The molecular conformations, which are the input of a quantum chemical calculation, are given in panel A, the orbitals of the computed two electronic states, the S_1 and S_2 state as well as the corresponding off-diagonal matrix elements between those two states, $\langle S_1|\hat{H}|S_2\rangle$, are given in panel B. As can be seen, the sign of those values arbitrarily switches. Those sign jumps can be removed by applying a phase correction algorithm. Results are given for those elements in panel C. Reproduced from ref. 50 with permission from the Royal Society of Chemistry.

multiplied by a phase factor, it is still a valid eigenfunction. Hence, the wave function is not uniquely defined. As a consequence, the coupling vectors between two states, NAC_{ij}, and in general any off-diagonal element in terms

of matrix notation, C_{ij}, resulting from the wave functions of those two states, Ψ_i and Ψ_j,

$$C_{ij} = \langle \Psi_i | \hat{O} | \Psi_j \rangle, \tag{4.17}$$

might randomly change sign along an interpolation coordinate. As a result, these elements show discontinuous curves, as can be seen in the bottom of panel B. To remove such phase jumps and make correct learning of the ML models possible, a virtual global phase correction has to be applied.[50] This process allows one to correct the couplings so that smooth curves can be obtained, as given in Figure 4.3 in panel C. This phase correction algorithm, discussed in more detail below, is applied only to ensure a global phase in most parts of the configurational space that is covered within the training set in order to make it learnable and applicable to an ML model.

It is worth mentioning that phase jumps resulting from the Berry phase cannot be removed, see ref. 59 for more details. However, it is assumed that those remain only in a small subspace of the training set and are thus negligible. This is further validated by theories, such as the Zhu Nakamura theory,[58,67,80,113–115] that do not track any phase along a NAMD simulation and still yield accurate results.

4.2.4 Phase Correction Algorithm

To carry out phase correction, the phase of each electronic state must be tracked from a pre-defined reference geometry onward. This can be done by computing wave function overlaps,[74,111,112] S, between two adjacent molecular geometries, k and l, as implemented for example in the WFoverlap code:[111]

$$S = \langle \Psi_k | \Psi_l \rangle. \tag{4.18}$$

The dimension of this matrix is $N^{\text{states}} \times N^{\text{states}}$, with N^{states} being the number of active states included in the calculations. If the two configurations k and l are very similar, the overlaps are obtained as diagonal elements with values close to ± 1. If the wave function overlap of a state is very close to $+1$, no phase change occurred, while a value close to -1 indicates that the phase of the wave function switched. A phase vector, p, can then be defined that contains a value of $+1$ or -1 for each electronic state, which indicates the change of the phase with respect to the previous geometry. Apart from critical regions of the PESs, p usually corresponds to the diagonal matrix elements of S.

In regions of the PESs where adiabatic potential energy curves have an avoided crossing and NACs become large, the off-diagonal elements of S might be larger than the corresponding diagonal elements. Then, the sign of p for each electronic state is obtained from the largest absolute value of a row of S.

During adaptive sampling, there will be cases when the two configurations, *i.e.*, the one that should be included in the training set and the reference geometry, are not similar enough to obtain overlaps larger than an absolute value of 0.5. In such cases, no information about a phase change can be obtained and interpolation between the two geometries in combination with iterative computation of wave function overlaps is necessary. The entries of p can be obtained using the following equation:

$$p_i = sgn\left(\max\left(\left|S_{ij}\right|\right)\right) \ \forall \ \left|S_{ij}\right| \geq 0.5; \ i,j = 1, 2, \ldots, N^{\text{states}}. \tag{4.19}$$

If interpolation is necessary during phase correction, the phase vector of the final geometry, p_n, is computed as the product of all previous phase vectors, p_0 to $pn-1$:

$$p_n = \prod_{\beta=0}^{n-1} p_\beta. \tag{4.20}$$

If the interpolation steps, β, become very large, it is advisable to store the wave functions of different geometries and the corresponding overlap calculations, *i.e.,* p. If a new geometry has to be added to the training set, the root mean square deviations (RMSDs) of this geometry to all other geometries, of which wave function information is available, can be computed and the phases can be tracked back to the geometry with the lowest RMSD. This intermediate calculation can reduce the number of interpolation steps significantly and hence leads to less additional computational cost.[50]

Special care must be taken when molecules are treated that contain many states lying energetically close to each other. So-called *intruder states* can become a source of error for phase correction. These states are so high in energy at the reference geometry that they are not computed, but possess a significantly lower energy at another geometry during a dynamics simulation such that they should have been included. In such cases, the relevant state for tracking the phase is not part of the active states of the geometry that has to be added to the training set and the result of the wave function overlap computation will not provide any information about a phase change of this state. Therefore, it is favourable to include more energetic states in the calculations from the beginning. In the targeted NAMD simulations, those additional states will not be accounted for and hence the forces and couplings that involve those states do not need to be computed. Therefore, computations are still affordable.

When the phase vectors of the configurations in the training set are calculated, the training set can be pre-processed. This should be done during adaptive sampling on-the-fly. For the initial training set, it can be done at any time before training. Phase correction can be employed using eqn (4.21) for the Hamiltonian matrix, H containing *e.g.* SOCs, and eqn (4.22) for the NAC vectors between a set of states, NAC_{ij}. Each matrix element of H is multiplied with the corresponding phase values of each state. In contrast, all

Machine Learning for Nonadiabatic Molecular Dynamics 93

values of a NAC vector between two states, $\mathbf{NAC}_{ij}$, (3 vectors accounting for x, y, and z direction with N^{atoms} entries (N^{atoms} is the number of atoms of the molecule)), is multiplied with the corresponding phase values, p_i and p_j, of those states.

$$H_{ij} = H_{ij} \cdot p_i \cdot p_j; \ 1, \ 2, \ \ldots N^{\text{states}} \tag{4.21}$$

$$\mathbf{NAC}_{ij} = \mathbf{NAC}_{ij} \cdot p_j \cdot p_j \tag{4.22}$$

In contrast to NAMD simulations, ground-state dynamics simulations do not need a pre-processing of the training set and thus also no phase correction. The reason is that energies and gradients of any electronic state are computed as diagonal elements from the wave function of the same state, $\langle \Psi_i | \hat{O} | \Psi_i \rangle$, where the phase factor enters twice and hence cancels out.

4.2.5   Surface Hopping Dynamics

To carry out NAMD simulations, the Surface Hopping including ARbitrary Couplings (SHARC) method[115] was used for the results represented here. SHARC is an extended version of Tully's fewest switches surface hopping method.[68] In surface hopping methods, the nuclei move on the adiabatic PESs that are made up by the electrons and are computed with quantum chemistry. The PESs are coupled by spin–orbit couplings or NACs, that directly influence the hopping probability of a molecule. A hop is a non-adiabatic transition from one electronic state to another. This event is computed stochastically by comparing a random number between 0 and 1 to the computed hopping probability. As surface hopping is a stochastic approach, an ensemble of trajectories is computed to approximate the motion of a wave packet. The results can be displayed by population plots, showing the population transfer between the electronic states along time.

In order to execute ML surface hopping molecular dynamics, the pySHARC Python wrapper for the SHARC dynamics driver[116] can be used, that, compared with the conventional NAMD program SHARC, omits the file I/O and allows for direct communication of ML models and the driver. This reduces the program runtime of SHARC drastically.[50] Usually, conventional NAMD codes are not limited by the duration of file I/O, as the quantum chemical calculations of the electronic properties are generally the time-limiting step in such simulations. However, when they are replaced by ML models that are much faster than quantum chemistry, the file I/O can become the time-limiting step. Moreover, as long dynamics simulations are aimed for, one has to keep in mind that the amount of produced data increases substantially and that the analysis of the outcomes might take longer than the production runs. In these cases, ML models can also be used to efficiently analyze the data and explore new concepts.[9] Further, it is advisable to save only each 10th–100th time step, when the dynamics do not change much and to save every time step, when transitions and important processes take place. For example, when a molecule is trapped for a long time in an excited

state, it not necessary to save each time step of each trajectory, whereas in critical regions of the PESs, where many different processes can take place on a very short time scale, it is important to store the information of every time step for subsequent analysis.

The training set that is used to produce all the results mentioned in this chapter is taken from ref. 50 and comprises a conformational subspace of the methylenimmonium cation, $CH_2NH_2^+$, that is obtained from normal mode scans and adaptive sampling using the SHARC program and two NNs. It contains three active singlet states and was computed with the multi-reference configuration interaction method that accounts for single and double excitations (MR-CISD). The double-zeta basis set aug-cc-pVDZ was used; in total the training set contains 4000 phase corrected data points, where each data point contains the molecular geometry in *xyz*-format, energies of three singlet states, corresponding gradients, and NACs. Permanent and transition dipole moments are also included but are not treated here. To compare NAMD based on ML with NAMD based on quantum chemistry, reference dynamics based on MR-CISD was carried out after excitation to the second excited singlet state for 100 fs using a time step of 0.5 fs. The so-obtained single points were not part of the training set for the ML models and cannot be used as such, as no phase correction was carried out for these points. For more details on reference dynamics, see ref. 50.

4.3 Example: Methylenimmonium Cation

As already introduced, many processes that are important for nature and life depend on light, such as photosynthesis or the ability to see. Concerning the latter, retinal is the molecule in the protein rhodopsin that plays a key role in vision. The photoreactive part is the chromophore retinal, which belongs to the family of protonated Schiff bases. The smallest member of this family is the methylenimmonium cation, $CH_2NH_2^+$, which undergoes ultrafast switches after photo-excitation. The nonadiabatic transition from the first excited singlet state back to the ground state leads to a rotation of the molecule along the HC–N–H dihedral angle.[116–118]

In the following, the photo-induced processes of this test molecule are used to investigate the ability of ML models to reproduce NAMD. The ML models are NNs taken from ref. 50. Here, the methylenimmonium cation is used as a test system, because transitions from the second excited singlet state back to the ground state are occurring on an ultrafast time scale, which are challenging to reproduce. More importantly, this ultrafast nature makes the dynamics computationally accessible for simulations based on the quantum chemistry reference method for comparison.[50]

4.3.1 ML Surface Hopping Dynamics

Figure 4.4 shows the population schemes obtained after executing surface hopping dynamics with different methods. Panel A shows results from two

Machine Learning for Nonadiabatic Molecular Dynamics 95
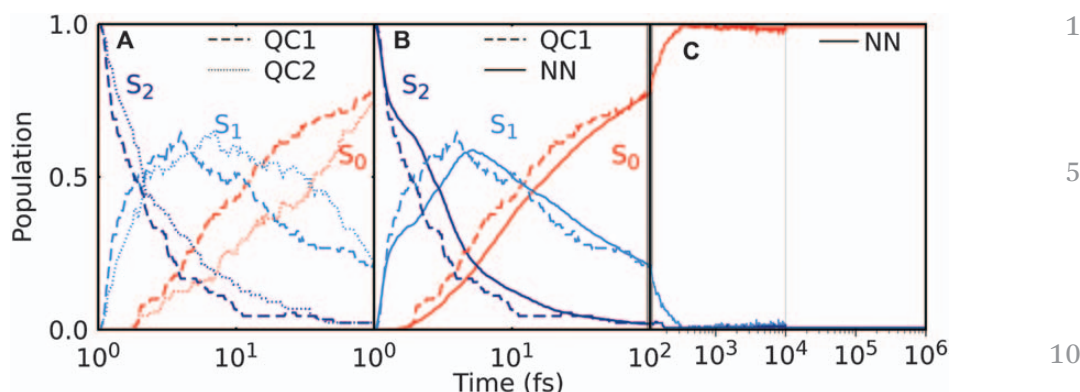
Figure 4.4 Surface hopping dynamics of $CH_2NH_2^+$ after excitation to the S_2 state. (A) Dynamics of 90 trajectories propagated for 100 fs with the QC1 and QC2 method are compared with (B) NNs, which are obtained from 3846 trajectories. (C) Long excited-state dynamics from NNs were obtained from 200 trajectories propagated for 10 ps and 2 trajectories propagated for 1 ns.[50]

slightly different quantum chemistry methods. QC1 refers to the reference quantum chemistry method MR-CISD(6,4)/aug-cc-pVDZ (dashed lines), while QC2 indicates the same method, but with another double-zeta bases set, MR-CISD(6,4)/631 ++ G** (dotted lines). The slightly different quantum chemistry methods are used to demonstrate which population schemes can be regarded as similar. As can be seen from panel A, the two quantum chemistry methods QC1 and QC2 result in very similar population schemes, which are obtained from dynamics simulations of 90 trajectories propagated for 100 fs. Initially, 100% of the population is in the second excited singlet state, S_2, from which fast population transfer to the first excited singlet state, S_1, and back to the ground state, S_0, takes place. This dynamics can be successfully reproduced with NNs trained on the training set of 4000 data points computed with the QC1 method. The results are directly compared with each other in panel B. As can be seen, the population curves obtained from NN computations (continuous lines) are much smoother than those obtained with QC1 (dashed lines). This is because not only 90 trajectories (resulting from 1000 initial conditions from Wigner sampling) were propagated with NNs, but 3846 trajectories (resulting from 20.000 initial conditions from Wigner sampling). This enlarged statistics is the first advantage of NNs and is possible due to the NNs' low computational costs. The overall computational time of 3846 trajectories with NNs ($\approx 19\,000$ min) is still about a factor of 800 lower than the overall computational time of 90 trajectories obtained from QC1 calculations (≈ 15.5 Mio. min.) when using the same central processing unit. Comparing panels A and B further shows that the agreement of NNs and QC1 is akin to the agreement of QC1 and QC2.

The goal of executing long NAMD simulations that are not feasible with the QC1 method anymore is possible with two NNs by employing the aforementioned sampling scheme to ensure accurate potentials for energies,

forces, and NACs along the dynamics run. In contrast to the adaptive sampling scheme described above, an iterative sampling scheme is employed, *i.e.,* the threshold for comparison of the different NNs is not reduced anymore, but is fixed in the beginning and kept constant. An ensemble of 200 trajectories was used to produce the population schemes up to 10 ps (10^4 fs) and 2 trajectories were further propagated for 1 ns (10^6 fs). As can be seen, the population is transferred to the electronic ground state after around 300 fs and stays there for the rest of the simulation time. Hops that still take place, can be regarded as noise and are due to remaining kinetic energy. In 6 hours, a dynamics simulation of 10 ps could be achieved, while it took around 59 days to compute the ns simulation. A simulation for 1 ns would have taken approximately 19 years with the QC1 method.[50]

To make long time scale NAMD simulations possible, it is favourable to include important processes, such as hydrogen atom dissociation or rotations of different molecular groups, into the training set from the beginning. This can be done to reduce the computational costs from additional quantum chemistry calculations, which become necessary if such previously-mentioned events take place during the long dynamics simulations. In the case of hydrogen dissociation, for example, the trajectory might be aborted after each time step the corresponding H-bond gets elongated and no profit would be gained from using NNs instead of quantum chemistry. Indeed, the dynamics would even take longer than the reference dynamics in such cases, as the NNs have to be re-trained after each time step and the phase correction procedure might need additional interpolation steps to properly track the phases of the wave function of each state.

4.3.2   Energy Conservation

Importantly, forces need to be computed as derivatives of NN potentials to preserve energy conservation. To check the energy conservation along NN dynamics, the mean and standard deviation of the total energy along the trajectories is given in Table 4.1 as taken from ref. 50. As can be seen, the total energy is as well conserved with QC1 as with NNs. It is evident that energy conservation becomes slightly worse when hops are forbidden and

Table 4.1   Computation of the MAE and standard deviation (Std.) of the total energy averaged over 90 trajectories (time step for classical propagation of 0.5 fs) and 50 trajectories (time step for classical propagation of 0.05 fs) propagated for 100 fs with the QC1 method and NNs. Values taken from ref. 50.

Method	Time step [fs]	MAE [eV]	Std. [eV]	Hops allowed?
QC1	0.5	10.63	0.047	Yes
QC1	0.5	10.77	0.059	No
QC1	0.05	10.73	0.011	No
NN	0.5	10.72	0.052	Yes
NN	0.5	10.73	0.061	No
NN	0.05	10.80	0.017	No

Machine Learning for Nonadiabatic Molecular Dynamics 97

when trajectories have to stay in the S_2 state for the whole simulation time. This is demonstrated by the larger standard deviation – a trend that is obtained with both methods separately. By reducing the time step of classical nuclei propagation from 0.5 fs to 0.05 fs, the standard deviation lowers considerably, even though hops are forbidden.[50]

4.3.3 Further Tools of ML Models

As shown in the previous section, the NNs can be trained to reproduce excited-state energies, forces, and NACs accurately. Hence, not only NAMD simulations can be carried out, but also optimizations of critical regions, such as state minima or conical intersections. The latter in particular are the target of many quantum chemical studies, and therefore, will also be discussed here.

Conical intersections are places of the PESs, where most of the transitions from one state to another are assumed to take place. The specific geometries at which the hops take place in a surface hopping molecular dynamics simulation are called hopping geometries. At the conical intersections of two coupled states, the NACs show a singularity and are very large. Further, the first derivative of such an avoided state crossing is discontinuous. Therefore, conical intersections pose a challenge to quantum chemistry as well as to ML models. Optimization runs from hopping geometries as starting guesses can be done to find such critical points and to additionally provide a measure of the quality of NN potentials for energies, gradients, and NACs. Figure 4.5, taken from ref. 50, illustrates that NN potentials are accurate enough to find the correct minimum energy conical intersections. It shows the scatter plots of the optimized geometries and the hopping geometries of QC1, QC2, and NNs and further gives the distribution of the training set geometries along two coordinates that are important for each minimum energy conical intersection. Panel A gives the plot of the S_1/S_0 state crossing, which is characterized by a rotation of the molecule along its dihedral angle (see Figure 4.6 first row), and panel B shows the scatter plot of the S_2/S_1 state crossing that leads to an elongated C–N bond in combination with a slight bipyramidalization of the molecule (see Figure 4.6 second row).

As can be seen in the scatter plots, the minimum energy conical intersections found with each method are located very close to each other, which can be also verified in detail from the geometries given in Figure 4.6. In the S_1/S_0 conical intersection (panel A), the hopping geometries are more widely distributed around the minimum energy conical intersection than in the S_2/S_1 state crossing (panel B). This indicates that the potential around the former state crossing is smoother, allowing hops from configurations that are further away from the minimum energy conical intersection. In the latter case, the PESs have a steeper slope and the hopping geometries are more similar to the minimum energy conical intersection. This trend is obtained with each method separately and demonstrates the high accuracy of the ML potentials for energies, gradients, and NACs.
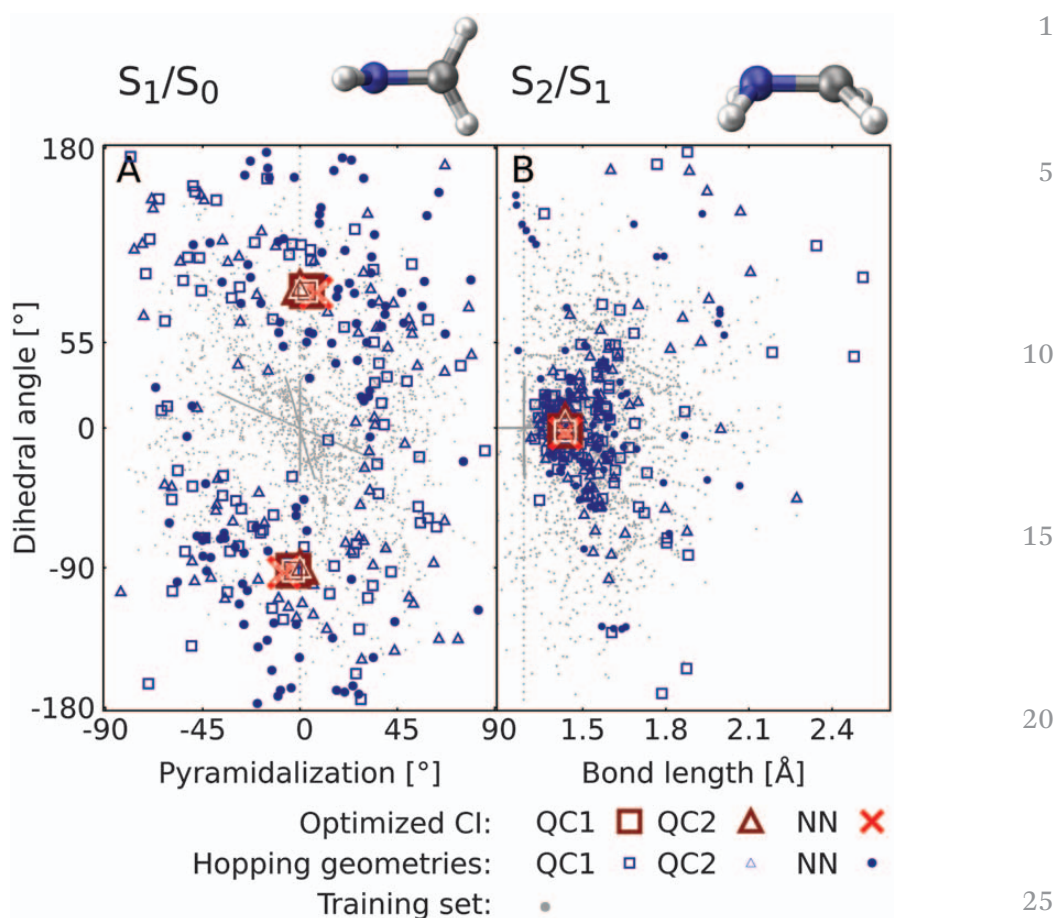
Figure 4.5 The distribution of the hopping geometries around the minimum energy conical intersections obtained with QC1, QC2, and NNs along with the geometries in the training set are shown in panel A and B for the S_1/S_2 conical intersection and the S_2/S_1 conical intersection, respectively. The geometry of the minimum energy conical intersection of QC1 is depicted on top. The definition of the dihedral and pyramidalization angle used to produce the scatter plots is given in panel C.
Reproduced from ref. 50 with permission from the Royal Society of Chemistry.

This assumption can be further tested by potential energy scans around each conical intersection. The scans were independently performed with the QC1 and NN method and the corresponding three-dimensional plots are shown in Figure 4.7. As can be seen, the NN potentials are slightly smoother than the QC1 potentials (compare panels B, A for the S_2/S_1 state crossing and panels D, C for the S_1/S_0 state crossing) and show larger energy gaps between the two approaching surfaces. These larger gaps result from the fact that the ML potentials need to be differentiable at any point in space. Analysis of the conducted potential energy scans around the minimum energy conical intersections further proves the shallower PESs around the S_1/S_0 conical intersection compared with the S_2/S_1 conical intersection and hence explains
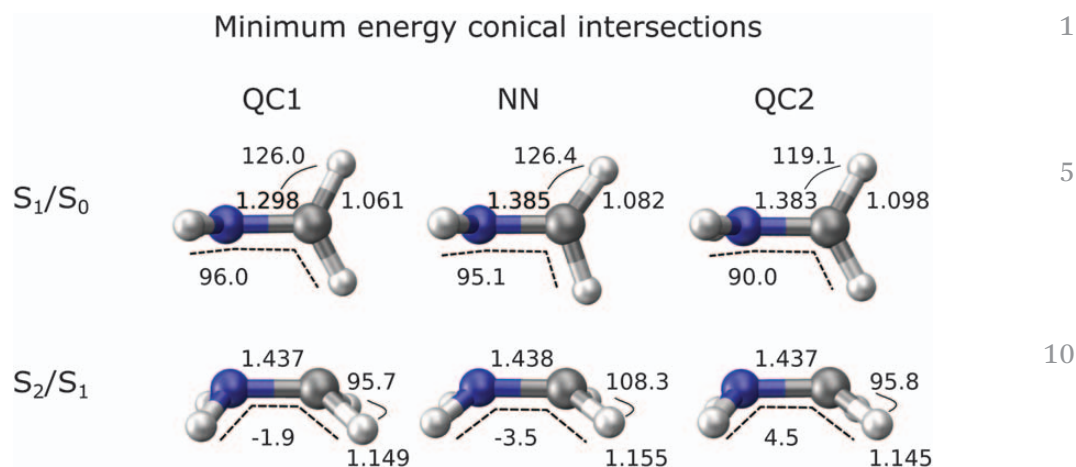
Figure 4.6 For each method, we could find two different conical intersections. QC1 (MR-CISD/aug-cc-pVDZ) is used as the reference method to train neural networks (NNs) and compared with conical intersections obtained with QC2 (MR-CISD/6-31 ++ G**). For the two conical intersections, the bond length between the nitrogen (dark grey) and the carbon (light grey) atom is shown, as well as the bond length between the carbon and one hydrogen atom. Values are given in Angström. A dihedral angle between four atoms marked with the dashed line is given, as well as an angle between the carbon and a hydrogen atom (S_1/S_0 CI) and between two hydrogen atoms (S_2/S_1 CI). Reproduced from ref. 50 with permission from the Royal Society of Chemistry.

the wider distribution of hopping geometries around the S_1/S_0 state crossing. Because hops occur further away from the minimum energy conical intersection, the training set (obtained *via* adaptive sampling) does not cover the S_1/S_0 conical intersection as densely as the S_2/S_1 conical intersection. The error of the NNs to reproduce the former conical intersection is thus also larger.[50]

In addition to the scans around the minimum energy conical intersections, a scan was carried out along a reaction coordinate that contains both of the state crossings. The obtained potential energy curves are computed with QC1, NNs with different representations, LR, SVR, and KRR and are given in Figure 4.8.

The NNs employed for the dynamics shown above (NN1/NN2 with inv.D.) are compared with additional NN models with advanced descriptors and different types of regressors. Panel B and C show NN potential energy curves with the Coulomb matrix (NN3) and the poly.D. matrix (NN4) as a representation, respectively. These lead to similarly accurate results and are comparable to the results obtained with the inv.D. descriptor in panel A. The two avoided crossings between different states are not captured as accurately as the rest of the potentials with each method and larger energy gaps are obtained in these regions compared with the QC1 method.

However, a comparison of the NN results with the result of the baseline model, LR, clearly indicates the higher prediction accuracy of NNs. LR is not able to reproduce the shape of the S_2 state at all, and the avoided crossings
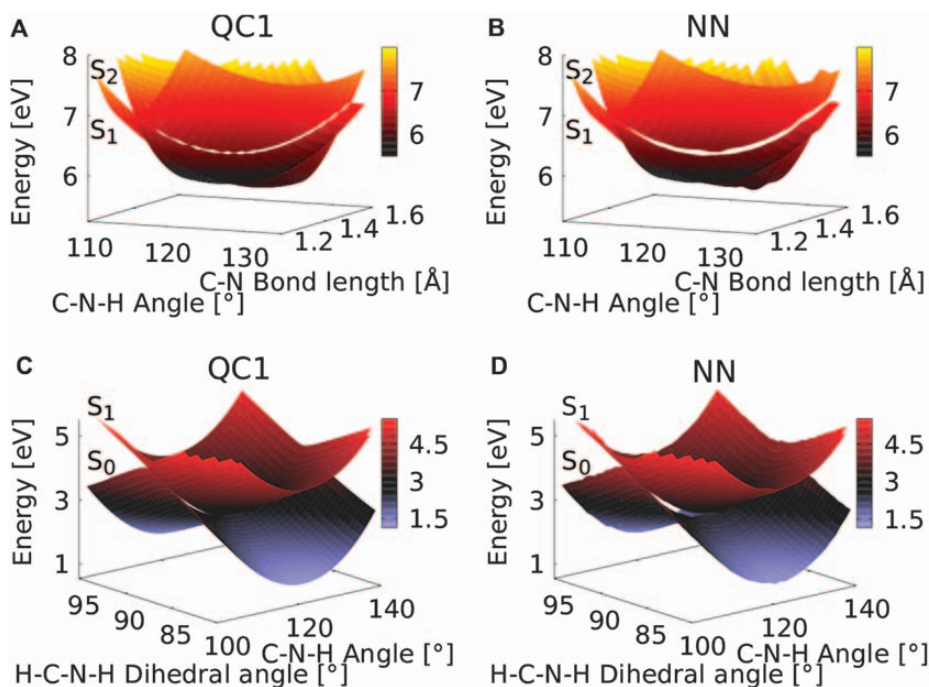
Figure 4.7 The minimum energy conical intersections computed with the QC1 method were used as a centre to compute potential energy scans with the QC1 and NN method around them. The scans around the S_2/S_1 state crossing are given in panel A and B for QC1 and NN, respectively, while the scans around the S_1/S_0 minimum energy conical intersection are given in panel C and D for QC1 and NN, respectively.
Reproduced from ref. 50 with permission from the Royal Society of Chemistry.

show a very large energy gap. SVR in panel E is more accurate than LR, but less accurate than NNs with the same descriptor (inv.D. – compare to panel A). KRR given in panel F yields comparable results to NNs. Besides the distinct regressor, the kernel methods, SVR and KRR, differ in the used representation. As the FCHL[82,84] representation could be used for KRR compared with the inv.D. matrix for SVR, more information about the molecule is encoded for KRR, which finally leads to more accurate potential energy curves of KRR. The kernel size depends on the number of data points, therefore a better descriptor can lead to higher prediction accuracy. Nonetheless, the most accurate results are obtained with the NN models. There, no significant improvement can be obtained by using a better descriptor, because the NNs can compensate deficiencies of the descriptor by a deeper network architecture.

As Figure 4.8 indicates, the potential energy curve of the S_2 state shows an artificial jump in the energy that is located near the avoided crossing between the S_1 and S_0 state. This erroneous discontinuity in the quantum chemical potential energy curve is not reproduced by the ML models. This ability of ML models to provide smooth interpolation between data points is even able to correct errors in this case. Moreover, as is clear in panel A, the
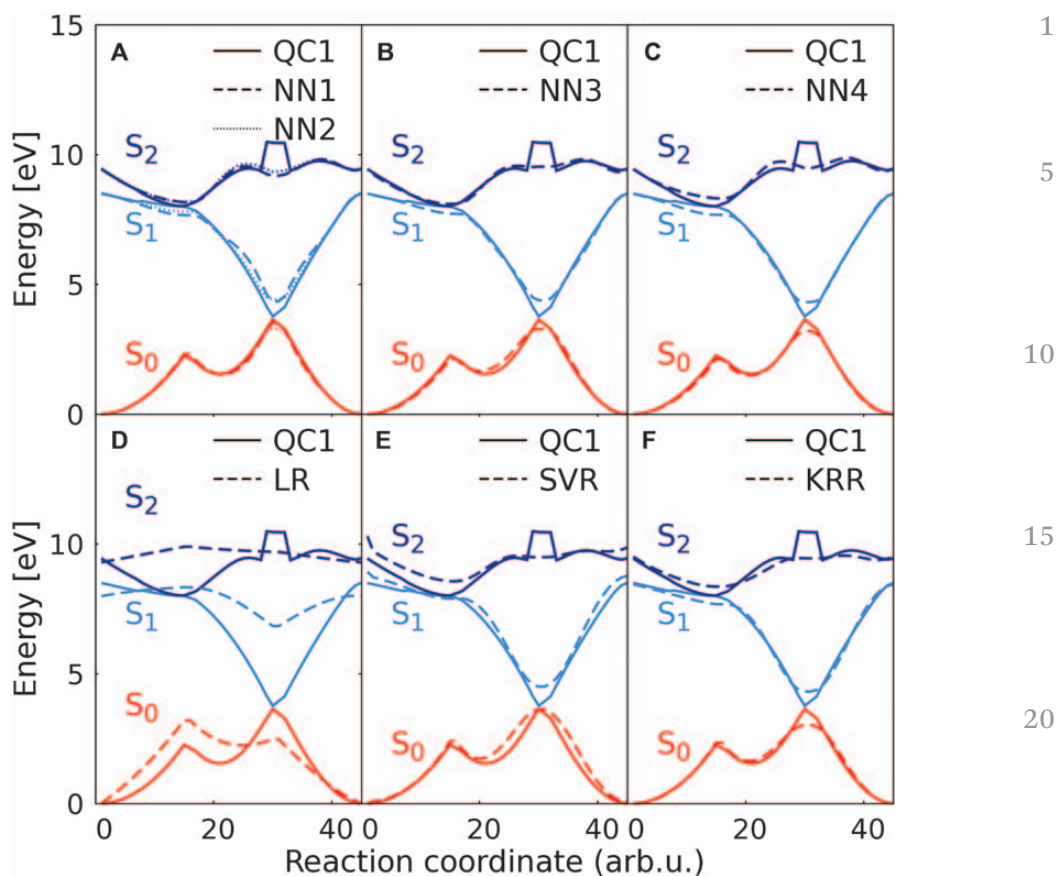
Figure 4.8 Potential energy curves along a reaction coordinate that contains two avoided crossings. One avoided crossing is between the S_2 and S_1 states, and the other one is between the S_1 and S_0 states. The potential energy curves are computed with the QC1 method (continuous line in each panel) and are compared with different ML models (dashed and dotted lines). (A) NN1 and NN2 indicate predictions with the inverse distance matrix using slightly different network parameters, such as it is done in adaptive sampling. (B) NN3 gives results from NNs using the Coulomb matrix as a descriptor and (C) the poly.D. matrix as a descriptor. (E) Curves obtained with LR, (F) SVR, and the inv.D. matrix, and (G) KRR using the FCHL representation that treats gradients as derivatives and provide the most sophisticated representation used here.[34,50]

two slightly different NNs with the inv.D. matrix as a descriptor for molecules (NN1 and NN2) yield very similar potential energy curves along this reaction coordinate. This means that when NN1 and NN2 are used for adaptive sampling, no quantum chemistry reference calculation is required in such a region of larger artificial error of the NN models with respect to the reference method. The trend of larger error for the S_2 state is also revealed by computing the state-specific MAE on a separate test set with 770 data points obtained from scans along the linear combinations of normal modes (see Table 4.2). The MAE for the S_2 state is approximately doubled for energies and gradients compared with the MAE obtained for the S_0 state. The MAE of

Chapter 4

Table 4.2 The MAE of each electronic state separately on a test set containing 770 data points computed with NN1 and NN2 from Figure 4.8 with the inv.D. matrix as a descriptor for molecules. MAEs are taken from ref. 50.

State	MAE energy [H]	MAE gradients [H/Bohr]
S_0	0.00176	0.00444
S_1	0.00200	0.00670
S_2	0.00335	0.00893

Table 4.3 A test set of 770 data points of the methylenimmonium cation was used to compute the MAE on energies [H], gradients [H/Bohr], and NACs [a.u.] averaged over all states for LR, SVR, KRR, and NNs with the FCHL, inv.D., poly.D. and the Coulomb matrix as molecular descriptors. Most values are taken from ref. 50.

Model	MAE energy	MAE gradients	MAE NACs
NN/inv.D. (NN1/NN2)	0.00237	0.00669	0.328
NN/Coulomb (NN3)	0.00238	0.00690	0.314
NN/poly.D. (NN4)	0.00197	0.00617	0.335
LR/inv.D.	0.09240	0.13902	0.471
SVR/inv.D.	0.00618	0.01169	0.382
KRR/FCHL	0.00300	0.00866	0.351

the S_1 state is close to the MAE of the S_0 state for energies and lies between the MAE of the S_0 and S_2 states for gradients.

For the sake of completeness, the MAEs obtained with each ML model on the same test set of 770 data points on energies, gradients, and NACs are given in Table 4.3. Referring to the NN models, the Coulomb matrix and the inv.D. matrix have similar MAEs for each property, which is expected due to the similarity of the descriptors. It can, however, be observed that the poly.D. matrix yields a slightly lower error and thus higher accuracy. The input vector, *i.e.*, the number of input features to represent a molecule, is squared for the poly.D. matrix compared with the inv.D. matrix. As expected, the LR model yields the largest MAEs for each property. The same trend as it is assumed from potential energy curves in Figure 4.8 can be obtained for the SVR model, which achieves higher accuracy than the LR model and lower accuracy than the NN models and KRR.

The MAEs obtained from KRR are slightly larger than those obtained from NNs but smaller than those from SVR and LR. This trend is again in agreement with the accuracy of the potential energy curves given in Figure 4.8.

For all properties, the LR baseline model yields the least accurate results. Using the inv.D. matrix as a descriptor is sufficient for NNs, but it leads to larger errors on energies, gradients, and NACs using SVR. In fact, the MAE on energies is three times as large as it is for NNs, the MAE on gradients is doubled compared with the MAE on gradients obtained with NNs, and the MAE on NACs is also larger. In case of KRR in combination with the FCHL

Machine Learning for Nonadiabatic Molecular Dynamics 103

representation, a significant improvement compared with SVR can be obtained. Still, KRR is slightly worse than NNs, which might be because NNs can be regarded as a *multi-state* model and one molecular structure can be related to all energy levels at once. In contrast, KRR can only relate one molecular geometry to one output value. This gives the NN models more information about the relationship between a geometry and its excited-state properties.

4.4 Conclusion and Outlook

ML can be successfully applied to compute excited-state energies, gradients, and NACs. With these properties at hand, much longer time scales than with quantum chemistry-based dynamics simulations are reachable, due to the low computational costs of ML models and the use of at least two ML models on-the-fly to ensure accurate ML predictions. Further, not only long dynamics simulations can be carried out, but also many more trajectories can be computed than it is usually the case with quantum chemistry methods. Hence, better statistics can be obtained, and rare reaction channels can be analyzed.

To conserve the energies along ML trajectories, the gradients should be derived from ML PESs. As a consequence, the ML PESs are smoother than those obtained with quantum chemistry, and the discontinuities occurring at crossing points of the PESs can never be reproduced completely. This drawback, in turn, becomes an advantage, when erroneous energy jumps arise in the quantum chemistry training data due to the shortcomings of the chosen underlying *ab initio* method. Such wrong discontinuities are then corrected automatically in the smooth ML PESs.

By comparing different ML models and representations of molecules, it can be shown that it is advisable to treat all energy levels that are included in NAMD in one ML model. NNs are employed in this way, and we find that NN potentials are more accurate than the potentials obtained from LR, SVR, or KRR, which can relate the molecular input only to one energy level, when employed straightforwardly. By using a more intricate representation of the molecule, the accuracy of KRR can catch up with that of NNs, while using a better representation for NNs does not improve their performance significantly.

If only energies and gradients are necessary, as is the case for Born–Oppenheimer dynamics, a very efficient training set generation can be achieved from a small initial training set that can include any configurations obtained from manually conducted scans or a sampling method. To enlarge the training set in an efficient way, it is beneficial to start with an adaptive sampling approach as soon as the initial training set is large enough (≈ 1000 data points) to enable molecular dynamics simulations with preliminary, crude ML models. In this way, the relevant conformational space for photodynamics can be sampled. However, if transition dipole moments or couplings between different states are required, it is important to apply a phase correction procedure from one reference geometry to every data point that needs to be added to the training set.

104   *Chapter 4*

To conclude, using the automatic sampling technique in combination   1
with the phase correction algorithm to build up and expand a training set for
NAMD, new possibilities arise to study the excited-state dynamics of complex
chemical systems at nanosecond time scales. With the high accuracy of ex-
pensive quantum chemistry methods and the low computational costs of ML   5
models at the same time, this setup can be used to calculate NAMD that are
not feasible with conventional *ab initio* methods. Thus, ML-based dynamics
simulations allow exploration of new reaction channels and pathways of
molecules that are relevant in several research fields, such as medicine,
biology, chemistry, and materials design.   10

References

1. T. Schultz, E. Samoylova, W. Radloff, I. V. Hertel, A. L. Sobolewski and
   W. Domcke, *Science*, 2004, **306**(5702), 1765.   15
2. W. J. Schreier, T. E. Schrader, F. O. Koller, P. Gilch, C. E. Crespo
   Hernández, V. N. Swaminathan, T. Charell, W. Zinth and B. Kohler,
   Science, 2007, **315**, 625.
3. C. Rauer, J. J. Nogueira, P. Marquetand and L. González, *J. Am. Chem.
   Soc.*, 2016, **138**, 15911.   20
4. S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B. F. E. Curchod,
   N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M. K. Nazeeruddin and
   M. Gräatzel, *Nat. Chem.*, 2014, **6**, 242.
5. A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi
   and M. Ceriotti, *Sci. Adv.*, 2017, **3**(12), e1701816.   25
6. I. Ahmad, S. Ahmed, Z. Anwar, M. A. Sheraz and M. Sikorski, *Int. J.
   Photoenergy*, 2016, **2016**, 1.
7. S. Mai, P. Marquetand and L. González, *J. Phys. Chem. Lett.*, 2016,
   7, 1978.
8. S. Mai, M. Richter, P. Marquetand and L. González, *Chem. Phys.*, 2017, 30
   482, 9.
9. F. Häse, I. F. Galván, A. Aspuru-Guzik, R. Lindh and M. Vacher, *Chem.
   Sci.*, 2019, **10**, 2298.
10. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States,
    S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187.   35
11. A. D. Mackerell, *J. Comput. Chem.*, 2004, **25**, 1584.
12. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev.:
    Comput. Mol. Sci.*, 2013, **3**(2), 198.
13. M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi,
    D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Kräutler, C. Oostenbrink, 40
    C. Peter, D. Trzesniak and W. F. van Gunsteren, *J. Comput. Chem.*, 2005,
    26(16), 1719.
14. A. P. Bartók, M. C. Payne, R. Kondor and G. Csányi, *Phys. Rev. Lett.*,
    2010, **104**, 136403.
15. Z. Li, J. R. Kermode and A. De Vita, *Phys. Rev. Lett.*, 2015, **114**, 45
    096405.

Machine Learning for Nonadiabatic Molecular Dynamics 105

16. M. Gastegger and P. Marquetand, *J. Chem. Theory Comput.*, 2015, **11**(5), 2187.
17. M. Rupp, R. Ramakrishnan and O. A. von Lilienfeld, *J. Phys. Chem. Lett.*, 2015, **6**(16), 3309.
18. J. Behler, *J. Chem. Phys.*, 2016, **145**(17), 170901.
19. M. Gastegger, C. Kauffmann, J. Behler and P. Marquetand, *J. Chem. Phys.*, 2016, **144**(9), 194110.
20. M. Gastegger, J. Behler and P. Marquetand, *Chem. Sci.*, 2017, **8**, 6924.
21. V. L. Deringer and G. Csányi, *Phys. Rev. B*, 2017, **95**, 094203.
22. V. Botu, R. Batra, J. Chapman and R. Ramprasad, *J. Phys. Chem. C*, 2017, **121**(1), 511.
23. J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192.
24. J. Behler, *Angew. Chem., Int. Ed.*, 2017, **56**(42), 12828.
25. H. Zong, G. Pilania, X. Ding, G. J. Ackland and T. Lookman, *Npj Comput. Mater.*, 2018, **4**, 48.
26. A. P. Bartók, J. Kermode, N. Bernstein and G. Csányi, *Phys. Rev. X*, 2018, **8**, 041048.
27. R. Xia and S. Kais, *Nat. Commun.*, 2018, **9**, 4195.
28. S. Chmiela, H. E. Sauceda, K.-R. Müller and A. Tkatchenko, *Nat. Commun.*, 2018, **9**, 3887.
29. G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler and M. Ceriotti, *J. Chem. Phys.*, 2018, **148**(24), 241730.
30. L. Zhang, J. Han, H. Wang, W. A. Saidi, R. Car and E. Weinan, in *Proceedings of the 32Nd International Conference on Neural Information Processing Systems*, NIPS'18, USA, 2018, p. 4441.
31. L. Zhang, J. Han, H. Wang, R. Car and E. Weinan, *Phys. Rev. Lett.*, 2018, **120**, 143001.
32. M. Gastegger and P. Marquetand, *arXiv e-prints*, 2018, arXiv:1812.07676.
33. H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Y. Chan and S. K. R. S. Sankaranarayanan, *J. Phys. Chem. C*, 2019, **123**(12), 6941.
34. A. S. Christensen, F. A. Faber and O. A. von Lilienfeld, *J. Chem. Phys.*, 2019, **150**(6), 064105.
35. H. Wang and W. Yang, *J. Chem. Theory Comput.*, 2019, **15**(2), 1409.
36. S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Müller and A. Tkatchenko, *Comput. Phys. Commun.*, 2019, **240**, 38.
37. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. VogtMaranto and L. Zdeborová, *Rev. Mod. Phys.*, 2019, **91**, 045002.
38. R. V. Krems, *Phys. Chem. Chem. Phys.*, 2019, **21**, 13392.
39. V. L. Deringer, M. A. Caro and G. Csányi, *Adv. Mater.*, 2019, 1902765.
40. J. Behler, K. Reuter and M. Scheffler, *Phys. Rev. B*, 2008, **77**, 115421.
41. C. Carbogno, J. Behler, K. Reuter and A. Groß, *Phys. Rev. B*, 2010, **81**, 035410.
42. F. Häse, S. Valleau, E. Pyzer-Knapp and A. Aspuru-Guzik, *Chem. Sci.*, 2016, **7**, 5139.
43. F. Liu, L. Du, D. Zhang and J. Gao, *Sci. Rep.*, 2017, **7**(8737), 1.

1

5

10

15

20

25

30

35

40

45

44. D. Hu, Y. Xie, X. Li, L. Li and Z. Lan, *J. Phys. Chem. Lett.*, 2018, **9**(11), 2725.

45. P. O. Dral, M. Barbatti and W. Thiel, *J. Phys. Chem. Lett.*, 2018, **9**, 5660.

46. W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral and G. Cui, *J. Phys. Chem. Lett.*, 2018, **9**(23), 6702.

47. D. M. G. Williams and W. Eisfeld, *J. Chem. Phys.*, 2018, **149**(20), 204106.

48. C. Xie, X. Zhu, D. R. Yarkony and H. Guo, *J. Chem. Phys.*, 2018, **149**(14), 144107.

49. Y. Guan, D. H. Zhang, H. Guo and D. R. Yarkony, *Phys. Chem. Chem. Phys.*, 2019, **21**, 14205.

50. J. Westermayr, M. Gastegger, M. F. S. J. Menger, S. Mai, L. González and P. Marquetand, *Chem. Sci.*, 2019, **10**, 8100.

51. N. L. Doltsinis, P. R. L. Markwick, H. Nieber and H. Langer, in *Radiation Induced Molecular Phenomena in Nucleic Acids*, ed. M. K. Shukla and J. Leszczynski, Springer Netherlands, 2008, p. 265.

52. E. Tapavicza, G. D. Bellchambers, J. C. Vincent and F. Furche, *Phys. Chem. Chem. Phys.*, 2013, **15**, 18336.

53. P. Marquetand, J. J. Nogueira, S. Mai, F. Plasser and L. González, *Molecules*, 2017, **22**(1), 49.

54. A. W. Jasper, B. K. Kendrick, C. A. Mead and D. G. Truhlar, in *Modern Trends in Chemical Reaction Dynamics*, ed. X. Yang and K. Liu, World Scientific, vol. 14, 2004, p. 329.

55. D. R. Yarkony, *J. Chem. Phys.*, 2005, **123**(20), 204101.

56. X. Zhu and D. R. Yarkony, *J. Chem. Phys.*, 2016, **144**(4), 044104.

57. N. Wittenbrink, F. Venghaus, D. Williams and W. Eisfeld, *J. Chem. Phys.*, 2016, **145**(8), 184108.

58. T. Ishida, S. Nanbu and H. Nakamura, *Int. Rev. Phys. Chem.*, 2017, **36**(2), 229.

59. I. G. Ryabinkin, L. Joubert-Doriol and A. F. Izmaylov, *Acc. Chem. Res.*, 2017, **50**(7), 1785.

60. S. Gómez, M. Heindl, A. Szabadi and L. González, *J. Phys. Chem. A*, 2019, **123**(38), 8321.

61. H. Köppel, W. Domcke and L. S. Cederbaum, in *Conical Intersections*, ed. W. Domcke, D. R. Yarkony and H. Köppel, World Scientific, vol. 15, 2004, p. 323.

62. D. R. Yarkony, in *Conical Intersections*, ed. W. Domcke, D. R. Yarkony and H. Köppel, World Scientific, vol. 15, 2004, p. 41.

63. G. A. Worth and L. S. Cederbaum, *Annu. Rev. Phys. Chem.*, 2004, **55**, 127.

64. F. Plasser, S. Gómez, M. F. S. J. Menger, S. Mai and L. González, *Phys. Chem. Chem. Phys.*, 2019, **21**, 57.

65. M. Richter, P. Marquetand, J. González-Vázquez, I. Sola and L. González, *J. Chem. Theory Comput.*, 2011, **7**(5), 1253.

66. E. Fabiano, T. Keal and W. Thiel, *Chem. Phys.*, 2008, **349**(1–3), 334.

67. P. Oloyede, G. Mil'nikov and H. Nakamura, *J. Chem. Phys.*, 2006, **124**(14), 144110.

Machine Learning for Nonadiabatic Molecular Dynamics 107

68. J. C. Tully, Molecular dynamics with electronic transitions, *J. Chem. Phys.*, 1990, **93**(2), 1061.

69. L. Zhu, V. Kleiman, X. Li, S. P. Lu, K. Trentelman and R. J. Gordon, *Phys. Rev. Lett.*, 1995, **75**(13), 2598.

70. C. Zhu, H. Kamisaka and H. Nakamura, *J. Chem. Phys.*, 2002, **116**(8), 3234.

71. N. L. Doltsinis, in *Computational Nanoscience: Do It Yourself!*, John von Neumann Institute for Computing, Jülich, vol. 31, 2006, p. 389.

72. G. Granucci and M. Persico, *J. Chem. Phys.*, 2007, **126**(13), 134114.

73. J. P. Malhado, M. J. Bearpark and J. T. Hynes, *Front. Chem.*, 2014, **2**(97), 97.

74. S. Mai, P. Marquetand and L. González, *Int. J. Quantum Chem.*, 2015, **115**, 1215.

75. L. Wang, A. Akimov and O. V. Prezhdo, *J. Phys. Chem. Lett.*, 2016, **7**(11), 2100.

76. J. E. Subotnik, A. Jain, B. Landry, A. Petit, W. Ouyang and N. Bellonzi, *Annu. Rev. Phys. Chem.*, 2016, **67**(1), 387.

77. S. Mai, P. Marquetand and L. González, *WIREs Comput. Mol. Sci.*, 2018, **8**(6), e1370.

78. M. Rupp, *Int. J. Quantum Chem.*, 2015, **115**(16), 1058.

79. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825.

80. D. Uysal and H. A. Güvenir, *Knowl. Eng. Rev.*, 1999, **14**(4), 319.

81. E. I. Im, *Working Paper*, Working Papers 199611, 1996, University of Hawaii at Manoa, Department of Economics.

82. F. A. Faber, A. S. Christensen, B. Huang and O. A. von Lilienfeld, *J. Chem. Phys.*, 2018, **148**(24), 241717.

83. A. Christensen, F. Faber, B. Huang, L. Bratholm, A. Tkatchenko, K. Müller and O. Lilienfeld, https://github.com/qmlcode/qml, 2017.

84. A. S. Christensen, L. A. Bratholm, F. A. Faber, D. R. Glowacki and O. A. von Lilienfeld, *J. Chem. Phys*, 2020, **152**(4), 044107.

85. V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

86. M. Awad and R. Khanna, in *Efficient Learning Machines*, ed. J. Pepper, S. Weiss and P. Hauke, Apress, Berkeley, CA, 2015, p. 67.

87. A. J. Smola and B. Schölkopf, *Stat. Comput.*, 2004, **14**, 199.

88. C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

89. J. Behler, *Int. J. Quantum Chem.*, 2015, **115**, 1032.

90. D. P. Kingma and J. Ba, in *ICLR 2015*, 2014, arXiv:1412.6980.

91. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, *J. Mach. Learn. Res.*, 2014, **15**, 1929.

92. Theano Development Team, *arXiv e-prints*, 2016, arXiv:1605.02688.

108   *Chapter 4*

93. I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, MIT Press, 2016.
94. J. Behler and M. Parrinello, *Phys. Rev. Lett.*, 2007, **98**, 146401.
95. H. Wang and W. Yang, *J. Phys. Chem. Lett.*, 2018, **9**(12), 3232.
96. Y. Huang, J. Kang, W. A. Goddard and L.-W. Wang, *Phys. Rev. B*, 2019, **99**, 064103.
97. M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi and P. Marquetand, *J. Chem. Phys.*, 2018, **148**(24), 241709.
98. M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.
99. J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.
100. J. E. Herr, K. Koh, K. Yao and J. Parkhill, *J. Chem. Phys.*, 2019, **151**(8), 084103.
101. V. Botu and R. Ramprasad, *Int. J. Quantum Chem.*, 2015, **115**(16), 1074.
102. J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev and A. E. Roitberg, *J. Chem. Phys.*, 2018, **148**(24), 241733.
103. E. Wigner, *Phys. Rev.*, 1932, **40**, 749.
104. R. E. Bruccoleri and M. Karplus, *Biopolymers*, 1990, **29**(14), 1847.
105. T. Maximova, R. Moffatt, B. Ma, R. Nussinov and A. Shehu, *PLoS Comput. Biol.*, 2016, **12**, e1004619.
106. J. Kästner, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**(6), 932.
107. G. Tao, *Theor. Chem. Acc.*, 2019, **138**, 34.
108. Y. I. Yang, Q. Shao, J. Zhang, L. Yang and Y. Q. Gao, *J. Chem. Phys.*, 2019, **151**(7), 070902.
109. J. E. Herr, K. Yao, R. McIntyre, D. W. Toth and J. Parkhill, *J. Chem. Phys.*, 2018, **148**(24), 241710.
110. S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**(5), 2847.
111. F. Plasser, M. Ruckenbauer, S. Mai, M. Oppel, P. Marquetand and L. González, *J. Chem. Theory Comput.*, 2016, **12**, 1207.
112. A. V. Akimov, *J. Phys. Chem. Lett.*, 2018, **9**(20), 6096.
113. T. Ishida, S. Nanbu and H. Nakamura, *J. Phys. Chem. A*, 2009, **113**(16), 4356.
114. A.-H. Gao, B. Li, P.-Y. Zhang and K.-L. Han, *J. Chem. Phys.*, 2012, **137**(20), 204305.
115. S. Mai, M. Richter, M. Ruckenbauer, M. Oppel, P. Marquetand and L. González, *sharc-md.org*, 2018.
116. G. J. M. Dormans, G. C. Groenenboom, W. C. A. Van Dorst and H. M. Buck, *J. Am. Chem. Soc.*, 1988, **110**(5), 1406.
117. M. Barbatti, A. J. A. Aquino and H. Lischka, *Mol. Phys.*, 2006, **104**(5–7), 1053.
118. J. Herbst, K. Heyne and R. Diller, *Science*, 2002, **297**(5582), 822.

APPENDIX A.5 MACHINE LEARNING AND EXCITED-STATE MOLECULAR DYNAMICS

JULIA WESTERMAYR AND PHILIPP MARQUETAND

Mach. Learn.: Sci. Technol., accepted manuscript (2020).
https://iopscience.iop.org/article/10.1088/2632-2153/ab9c3e

Contributions:

JULIA WESTERMAYR wrote the initial manuscript and contributed to the final writing of the manuscript.

PHILIPP MARQUETAND conceived the scope of the manuscript and contributed to the initial and final writing of the manuscript.

Machine learning and excited-state molecular dynamics

Julia Westermayr[1] and Philipp Marquetand[1,2,3]

[1] University of Vienna, Faculty of Chemistry, Institute of Theoretical Chemistry,
Währinger Str. 17, 1090 Wien, Austria
[2] Vienna Research Platform on Accelerating Photoreaction Discovery,
University of Vienna, Währinger Str. 17, 1090 Wien, Austria
[3] University of Vienna, Faculty of Chemistry, Data Science @ Uni Vienna,
Währinger Str. 29, 1090 Wien, Austria

E-mail: `philipp.marquetand@univie.ac.at`

Abstract. Machine learning is employed at an increasing rate in the research field of quantum chemistry. While the majority of approaches target the investigation of chemical systems in their electronic ground state, the inclusion of light into the processes leads to electronically excited states and gives rise to several new challenges. Here, we survey recent advances for excited-state dynamics based on machine learning. In doing so, we highlight successes, pitfalls, challenges and future avenues for machine learning approaches for light-induced molecular processes.

Keywords: machine learning, photodynamics, photochemistry, excited states, quantum chemistry, spin-orbit couplings, nonadiabatic couplings.

1. Introduction

Photosynthesis, photovoltaics, the processes that enable our vision or photodamage of biologically relevant molecules, such as DNA or peptides – they all have one thing in common: The underlying processes are governed by a manifold of excited states after the absorption of light [1–18]. They can be studied experimentally via several techniques, such as UV/visible spectroscopy, transient absorption spectroscopy, photoionization spectroscopy or ultrafast electron diffraction [19–27]. However, experimental techniques are to some extent blind to the exact electronic mechanism of photo-induced reactions. In order to get a more comprehensive understanding, theoretical simulations can complement experimental findings and can provide explanations for observed reactions [9]. For instance, simulated UV spectra can be used to unveil the states relevant for photodamage and -stability of molecules [28–36] and the temporal evolution of molecules can be studied via nonadiabatic molecular dynamics (NAMD) simulations [37–46]. The latter gives access to different reaction channels, branching ratios, and excited-state lifetimes and will be the main topic of discussion here.

While experimental techniques require large and costly setups, theoretical simulations require high-performance computing facilities due to expensive electronic structure computations. Especially NAMD simulations are seriously limited by the underlying quantum chemical calculations, making long and experimentally relevant simulation times inaccessible with conventional *ab initio* methods. The larger the molecule becomes, the more electronically excited states are involved in reactions and the more complex their interactions become. This leads to non-linearly increasing costs of quantum chemical calculations and a compromise between accuracy and computational efficiency is indispensable. Relying on such expensive *ab initio* potentials, only a couple of picoseconds can be simulated and the exploration of rare reaction channels is restricted due to bad statistics [17, 43, 47].

Technically, the nuclear part and the electronic part of the calculations can be separated to a large extent. First, the electronic problem is solved leading to potential energies for the nuclei. Afterwards, the nuclei move on these potentials classically or quantum chemically [6, 8, 48–52]. These two subsequent steps can be carried out in every time step (on-the-fly),

if classical trajectories are employed. Alternatively, the two steps are separated as much as possible by precomputing the potential energy surfaces (PESs) and then using these precomputed PESs in the subsequent nuclear dynamics. Experimental observables and macroscopic properties can be obtained in follow-up computations or analysis runs. Machine learning (ML) can accelerate the overall simulation process on different levels and at several points. A broad classification of how to use ML models to replace different parts of quantum chemistry to make simulations more efficient is given in Fig. 1 [53].

The probably most fundamental way is to use
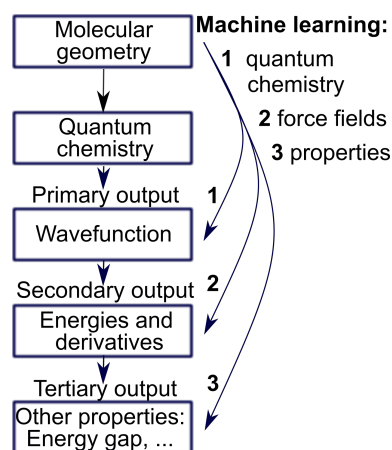
Figure 1. A broad classification of how to use machine learning models to replace different parts of quantum chemistry [53]. Simulations can be enhanced by providing (1) a wavefunction from a machine learning model, (2) a force field by fitting energies and forces or (3) other properties, such as energy gaps or reaction rates by learning the final output of a dynamics simulation directly.machine learning model and overview of the parts of a quantum chemical simulation.

ML to solve the Schrödinger equation. This has been done for the ground state by representing the molecular wavefunction on a grid, in a molecular orbital basis or in a Monte-Carlo approach [54–62] and has recently also been applied for the excited states of a one-dimensional model [63]. ML can also be used to reconstruct the wavefunction from near-field spectra [64] or to bypass the Kohn-Sham equation in density functional theory (DFT) [65]. The external potential, functional, electronic density or local density of states can be learned [53, 65–72]. Very recently, Ceriotti and co-workers further introduced

Machine learning and excited-state molecular dynamics 3

a smooth atomic density by defining an abstract chemical environment [73].

By having access to the molecular wavefunction or the electron density, the secondary output, which are energies and forces for the ground state and additionally couplings for the excited states, can be derived efficiently with ML. The coefficients of the ML wavefunctions or the density can further be used as an input for quantum chemical simulations, reducing the number of SCF iterations substantially [59].

Instead of learning the quantum chemistry of systems, the so-called "secondary outputs" [53] can also be mapped directly to a molecular structure, giving rise to so-called ML force fields. By training an ML model on *ab* initio data, the accuracy of quantum chemistry can be combined with the efficiency of conventional force fields for molecular dynamics (MD) simulations in the ground state [74–109]. For the excited states, only a couple of studies are available [106, 110–129]. Nevertheless, the first NAMD simulation with ML dates back to the year 2008, where the scattering of O_2 on Al(III) was studied in a mixed quantum-classical approach considering singlet-triplet transitions [111, 130].

Having access to the excited state energies, "tertiary properties", such as UV spectra [131], band gaps [132–134], HOMO-LUMO gaps or vertical excitation energies [135–138] of molecules can be derived. Again, this tertiary output can also be fit in a direct fashion, which has been done for instance for a light-harvesting system by learning the excitation energy transfer properties [139] or the output of NAMD simulations to find out about the relations of molecular structures and dynamic properties [140]. Moreover, ML has been successfully applied for the inverse design of molecules and materials featuring specific properties, such as defined HOMO-LUMO gaps or catalytic activities. Examples range from the inverse design of photonic materials, to (photo-)catalysts, solar cells or (photo-active) drugs, to name only a few applications [107, 141–150].

Despite the opportunities of ML for the development of groundbreaking new methodologies, current techniques are often limited to certain molecules or specific problems. Methods exist, that extrapolate throughout chemical compound space, see e.g. Refs. [59, 131, 151–160], but usually models fail to go beyond energies and related properties, such as forces, atomization or excitation energies. Further, it is challenging to predict compounds consisting of atom types strongly different from those inside of the training set. Especially the fitting of the excited-state PESs poses another obstacle, let alone the transferability of excited-state PESs: Not only are ML models restricted to certain molecules or materials [106, 110–125, 127–

129], more often the different energetic states are fit independently from each other with separately trained ML models. As it is clear that the PESs of molecules are not independent of each other, it might also be unsurprising that learning them simultaneously is advantageous for various applications, such as NAMD or spectra predictions. Only a few studies exist that include more energetic states in one ML model and even less treat related properties, such as the vectorial dipole moments or couplings between different PESs in one ML model [85, 118, 131, 154, 161, 162].

However, in our view, the "holy grail" of ML for photochemistry is an ML model that provides all relevant energetic states, forces and properties at once, using derivatives where possible, rather than learning the properties independently. At the very best, this model should be transferable throughout chemical compound space [163] and could be used for molecules of any composition. Given the fact that ML models for the electron wavefunctions of different excited states (or related models within the DFT framework) do not exist for polyatomic systems, this dream has not yet come alive. Hence, we will focus this perspective on ML models that learn the secondary outputs, i.e., excited state PESs, corresponding forces, and nonadiabatic and spin-orbit couplings (NACs and SOCs, respectively) between them. We note that our discussion holds for different spin-multiplicities, although most studies focus on singlet states only. We try to address the recent achievements in the fields of photochemistry using ML and discuss the current challenges and future perspectives to get a step further to a transferable ML model for excited states that treats all properties on the same footing.

We start by discussing the generation of a training set for the treatment of excited state PESs, corresponding forces and couplings and focus on their use in NAMD simulations. Especially, we aim to clarify the differences between excited-state and ground-state properties. We therefore describe the NACs and SOCs that couple different electronic states and highlight their importance for NAMD simulations. Subsequently, state-of-the-art ML models for excited-state PESs are considered along with the challenge of modelling a manifold of energetic states.

2. Generating a training set for excited states

The basis of any successful ML model is a comprehensive and accurate training set that contains the molecular geometries in combination with the corresponding properties that need to be predicted. For the application of ML within NAMD simulations, the training set should contain a molecular geometry and the energies of all energetic states, corresponding forces, and

couplings between the states. It is computed with the quantum chemistry method, whose accuracy one wants to obtain. The choice of the quantum chemistry method is a problem on its own and often requires expert knowledge [18, 164]. Simply said, ML can be seen as efficient interpolation between data points with the accuracy of the reference method.

Before we go into detail on how to efficiently create a training set for excited states, we will first discuss the differences to ground state potentials and properties that need to be considered. A major drawback is the fact that a manifold of excited states and thus also the properties between them have to be accounted for. These are NACs between states of same spin multiplicity and SOCs between states of different spin multiplicity as well as transition dipole moments. The fitting of such properties is problematic due to the arbitrary phase of the wavefunction [118, 161]. Therefore, an additional pre-processing might be necessary. Either a diabatization [106, 115–117, 119–127], a so-called phase correction [118], or a special learning algorithm [165] renders data learnable. The latter two are described in the following while further details on the former are given in section 3.2.

2.1. Making excited-state data learnable

Compared to energies and forces, NACs, SOCs as well as transition dipole moments result from the wavefunctions of two different electronic states. Due to the non-unique definition of the wavefunction itself, i.e., the fact that multiplication of the electronic wavefunction with a phase factor still gives a valid eigenfunction of the electronic Hamiltonian, leads to an arbitrary phase, which is initiated randomly in a quantum chemical calculation. Consequently, also the sign of the couplings, C_{ij}, can be positive or negative. Here, i and j denote the indices of the involved states. The resulting inconsistencies in the coupling hypersurfaces make it challenging to find a good relation between an ML model, which is per definition a smooth function [166], and those discontinuous raw outputs.

This problem can be illustrated with molecular orbitals of the methylenimmonium cation (Fig. 2 reproduced from Ref. [118]). Panel A shows the molecular geometries, which are given as an input to a quantum chemical program. Two molecular orbitals are shown as placeholders for the wavefunctions of two electronic states, the S_1 and S_2 states. The color of the orbitals can be either blue or red and changes arbitrarily throughout the reaction coordinate. In the same way, also the overall wavefunction (which is difficult to plot) for the respective state changes its phase arbitrarily. As a consequence, also the sign of the couplings, where the product of the two wavefunctions'

signs enters, may change randomly (see panel B).
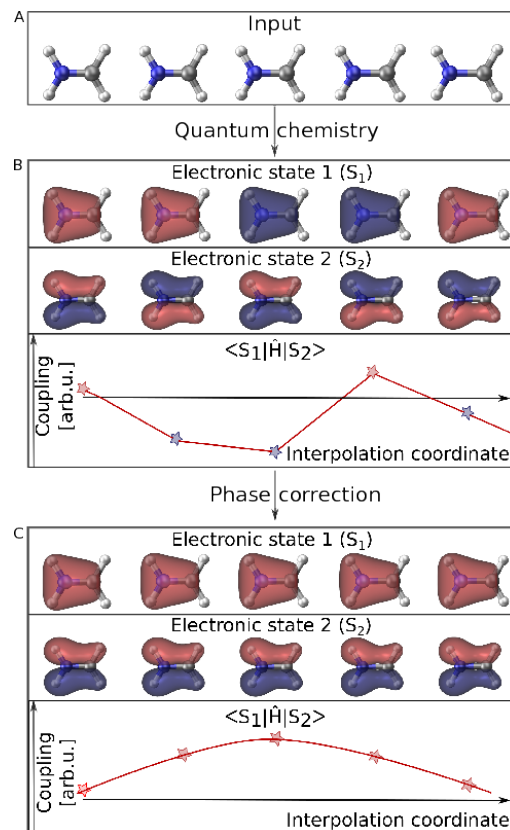
Figure 2. A set of quantum chemical calculations of the methylenimmonium cation, $CH_2NH_2^+$, along the C-N bond. The molecular conformations, which are the input of a quantum chemical calculation, are given in panel A, the orbitals of the computed two electronic states, the S_1 and S_2 state, as well as the corresponding off-diagonal matrix elements between those two states, $\langle S_1 | \hat{H} | S_2 \rangle$, are given in panel B. As can be seen, the sign of those values arbitrarily switches. Those sign jumps can be removed by applying a phase correction algorithm. Results are given for those elements in panel C. Reproduced from Ref. [118] under CC-BY, https://creativecommons.org/licenses/by/3.0/.

In order to allow for a meaningful ML description of elements resulting from two different electronic states, a data pre-processing is helpful. The former process is termed phase correction [118, 167] and is practicable to remove almost all such inconsistencies in the configurational space of the training set. This phase correction makes the use of conventional training algorithms possible.

To carry out phase correction, a wavefunction overlap computation [168], $\boldsymbol{S} = \langle \Psi_\alpha | \Psi_\beta \rangle$, has to be carried out between the wavefunction Ψ_β at every geometry β inside the training set and the wavefunction Ψ_α at a reference geometry α. The phase thus has to be tracked from a pre-defined reference geometry. It often happens that two geometries are dissimilar to each other, so that interpolation between

them is necessary, making this process generally more expensive. So-called intruder states give rise to additional problems, since they are so high in energy at the reference geometry that they usually would not be included in the initial calculation. However, they enter the lower energy region at another geometry visited during the NAMD simulations and thus need to be considered in the current calculation step. Hence, they should have been included from the beginning for the phase correction algorithm to work. As a solution to this problem, many electronic states need to be computed from the start. In some cases, where many energetic states lie close to each other and where the photochemistry is complex, phase correction might even be infeasible. The problem of intruder states was also identified by Robertsson et al. and is well explained for a diabatization procedure in Ref. [169]. For a more detailed discussion on phase correction, the reader is referred to Ref.s [118, 167, 168, 170]. Nevertheless, as given in panel C of Fig. 2, smooth curves are obtained if phase correction is carried out correctly and these phase-corrected properties can be learned with conventional ML models.

Similarly, a small set of data can be corrected manually and afterwards a cluster growing algorithm can be applied [125, 171]. This algorithm uses Gaussian process regression to continuously add data points to the initially phase-corrected data set. This approach has been employed recently to obtain diabatic transition dipole moments [119]. However, in systems containing many degrees of freedom and many electronic states, a manual correction of the sign of couplings is tedious and the approach has only been applied to small systems, yet [119, 126].

In contrast to the phase-correction procedures described above, an ML-based internal phase-correction during training renders the learning of raw quantum chemical data possible and does not require any pre-processing. However, it requires a modification of the training process itself [165]. In a recent study, we applied such a phase-free training using the deep continuous-filter convolutional-layer NN SchNet [156, 172] that we adapted for the treatment of excited states. In contrast to conventional algorithms, where the hyperparameters of the network are optimized to minimize the L_1 or L_2 loss function, here a phase-less loss function is applied. The latter allows the ML model to possess a different phase (or sign) for the learned property than the reference data. Since ML models intrinsically yield smooth curves, the algorithm will then automatically choose a phase for every data point such that smooth coupling curves are produced. This freedom of choice is achieved by calculating the errors between the ML value and all possible phase variations of the reference value and using only the smallest

of these errors. The possibilities for phase conventions scales with 2^{N_S-1}, where N_S is the number of considered states. Since the error is computed more often than in conventional ML training, the phase-less loss training becomes more expensive, when including more electronic states. Mathematically, instead of computing the mean squared error, ε_{L_2}, between reference couplings, C_{ij}^{QC}, and predicted couplings, C_{ij}^{ML},

$$\varepsilon_{L_2} = \|C_{ij}^{QC} - C_{ij}^{ML}\|^2, \tag{1}$$

the phase-free error, ε_{ph}, is computed as the minimum of 2^{N_S-1} computed errors:

$$\varepsilon_{ph} = \|C_{ij}^{QC} \cdot p_i^k \cdot p_j^k - C_{ij}^{ML}\| \text{ with } 0 \leq k \leq 2^{N_S-1}. \tag{2}$$

p_i^k and p_j^k are phase factors, giving rise to the sign of state i and j resulting in a phase vector with index k. This adaption of the loss function can remove the influence of any phase during the training process, making the use of raw data possible. A variation of this approach with reduced cost is possible if only one property, i.e. NACs or SOCs, are trained for NAMD simulations. A detailed discussion can be found in Ref. [165].

It is worth mentioning that, besides the arbitrary phase of the wavefunction, also the Berry phase (or geometric phase) [173] exists. Effects due to the Berry phase can not be accounted for with phase correction. Nevertheless, most often in mixed quantum classical NAMD simulations, the Berry phase can be neglected. As a drawback, some effects, such as interference of nuclear wavefunctions might not be described correctly with such methods, and thus prevents the application of the corresponding ML properties if those effects are important. In some other dynamics methods and reactions, the Berry phase plays a crucial role and can lead to path-dependent transition probabilities close to conical intersections. This effect is important in quantum dynamics simulations, where problems can be circumvented by using diabatic potentials, which will be described in section 3.2.

2.2. Choosing the right reference method

While many ML potentials for ground-state MD simulations are based on DFT training sets, see e.g. Refs. [77, 79, 86, 174, 175], the training sets for the excited states are mainly obtained with multi-reference methods. Examples are the complete active space self-consistent-field (CASSCF) method [112–114, 120, 122, 127, 140, 165] or multi-reference configuration interaction (MR-CI) schemes [115, 117–119, 126, 161, 165, 176–182]. The advantage of multi-reference methods compared to single reference methods is that photo-dissociation, which is likely to occur in many molecules after their excitation by light, can be treated accurately. In contrast, single-reference methods fail to

do so in many cases. However, multi-reference methods are seriously limited by their computational costs [7, 10], calling for an efficient and meaningful training set generation. Therefore, the training set should be as small as possible, but should cover the relevant conformational space of a molecule that is required for accurate NAMD simulations [183]. Accordingly, many recent training sets for MD simulations are built by a so-called "active-learning" [77, 184] or iterative/adaptive sampling scheme [85, 185] that will be described in the following and can be adapted for excited states [118]. From our point of view, it is most favorable to start by computing a small initial training set and to expand it via such an adaptive sampling scheme [77, 85, 118, 184–186].

2.3. Initial training set

If only static calculations are targeted, data bases can be generated efficiently by starting from already existing data sets. As an example, Schwilk et al. [182] constructed a large data set of 13k carbene structures by randomly choosing 4,000 geometries from the QM9 [187] data set (consisting of 130k organic molecular structures). Hydrogen-atoms were abstracted and singlet and triplet states were optimized. The MR-CI method was subsequently used to compute the energies of the singlet and triplet state and a data set of 13,000 different carbene structures, called QMspin, was obtained, opening avenues to investigate important intermediate geometries critical for organic reaction networks.

As a starting point for all following training-set generation schemes aiming to investigate the temporal evolution of a system, the equilibrium geometry of a molecule can be computed and taken as a reference. The initial training set can then be built up by sampling conformations close to this molecular configuration. In general, every sampling method is possible. Since the normal modes of a molecule are generally important for dynamics, scans along these coordinates can be used to sample different conformations. In cases of small molecules with few degrees of freedom, this process might be a good starting guess for an initial training set [118]. It also makes sense to optimize critical points like excited-state minima, conical intersections and state crossings and to include data along such optimization runs into the training set. The same is advisable for larger systems, but in addition, some other approaches like Wigner sampling [188] or sampling via MD simulations [189, 190] can be considered. To name a few approaches, umbrella sampling [191], trajectory-guided sampling [192], enhanced sampling [193] or metadynamics [194], using a cheap electronic structure method like the semi-empirical tight-binding based

quantum chemistry method GFN2-xTB [195]), can be employed.

Further, if literature or chemical intuition indicate that certain reactions, like dissociation, take place after photo-excitation, it is also favorable to include those reaction coordinates right from the beginning. The initial training set can easily comprise on the order of 1,000 data points, which might seem like a lot but is reasonable given the large number of data points in commonly used training sets [112–114, 161]. The quality of the initial ML potentials can be assessed by carrying out short scans along different reaction coordinates, such as combinations of normal modes. As soon as the initial training set is large enough, the training set expansion via an adaptive sampling scheme can be started.

2.4. Adaptive sampling for excited-states

ML models fail to predict regions with scarce training data, i.e., their extrapolation capabilities are faint [196]. Since such regions are likely visited during a dynamics simulation, the initial training set then needs to be expanded. A quality control is needed to detect whether unknown conformational regions of the molecule are visited, such that the corresponding structures afterwards can be added to the training set.

This concept was introduced already in 1992 as query by committee [197] and has been used in chemistry in the so-called GROW algorithm of Collins and coworkers [128, 129] as well as in the iterative sampling of Behler [185]. The latter is nowadays well adapted for the ground state [77, 85, 184, 186] and was recently modified for the excited states [118]. The scheme is described in more detail in the following.

To apply the procedure of adaptive sampling, at least two ML models have to be trained independently, e.g., with slightly different hyperparameters or starting weights. An overview of this procedure with two neural networks (NNs) is given as an example in Fig. 3. At every time step during an MD simulation, the predictions Y_M^p of at least two ML models, M, for a property p (e.g., a potential energy) are compared to each other. To this end, the standard deviation of these predictions with respect to the mean of each property, $\overline{Y}^P$, is computed according to

$$Y_\sigma^p = \sqrt{\frac{1}{M-1} \sum_{m=1}^{J} (Y_m^p - \overline{Y}^p)^2}. \tag{3}$$

This standard deviation is compared to a pre-defined threshold, ε^p, for each trained property. If the standard deviation stays below the threshold, the mean of each property, $\overline{Y}^P$, is forwarded to the MD program to propagate the nuclei. If the threshold is exceeded, the ML prediction is assumed to stem from
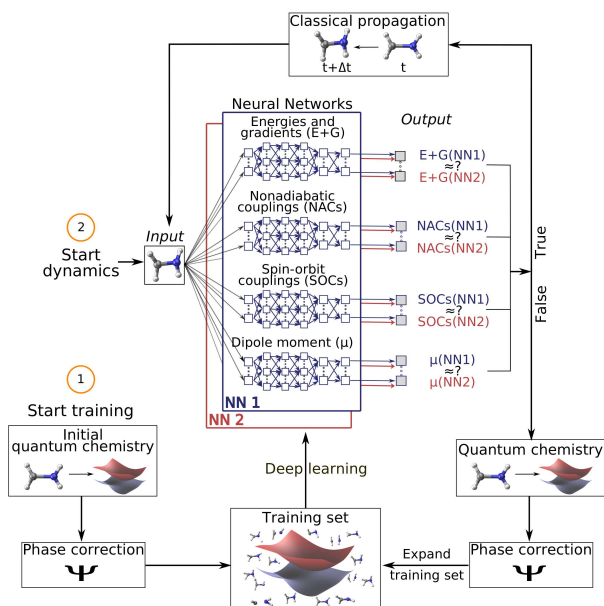
Figure 3. Overview of the adaptive sampling scheme for excited states, reproduced from Ref. [118] under CC-BY. Adaptive sampling is illustrated exemplarily with two NNs for the methylenimmonium cation, $CH_2NH_2^+$. As a starting point, ML models are trained on the initial training set and ML-based dynamics are executed. At each time step, the predictions of the two deep NNs (NN1 and NN2) are compared to each other for energies (E) and gradients (G), nonadiabatic couplings (NACs), spin-orbit couplings (SOCs) and dipole moments (μ). If the difference between the ML models overcomes a pre-defined, adaptive threshold, the geometry visited at this time step is re-computed with quantum chemistry, added to the training set after phase correction and the ML models are re-trained. Subsequently, a new dynamics cycle is started and this process is repeated until the ML models are deemed to be converged.

an undersampled or unknown region of the PESs and is deemed untrustworthy. This conformation has to be included into the training set to guarantee accurate ML PESs. Thus, a quantum chemical reference computation is carried out, the data point is added to the training set and the ML models are re-trained to execute ML-NAMD simulations on longer time scales. It is sensible to choose a large threshold, ε^p, in the beginning and adaptively make it smaller as the robustness of the ML models increases, giving rise to the name adaptive sampling [85].

Adaptive sampling for excited states differs from adaptive sampling in the ground state in the number of properties that are considered. As illustrated in Fig. 3, not only the energies must be accurately predicted, but also the couplings and, if necessary, dipole moments. Since more states are considered, an average standard deviation is taken as the mean of the standard deviations of each state in case of energies

and gradients,

$$Y_\sigma^p = \frac{1}{N_S} \sum_i^{N_S} \left(\sqrt{\frac{1}{M-1} \sum_{m=1}^M (Y_m^{p_i} - \overline{Y}^{p_i})^2} \right), \qquad (4)$$

and as the mean of the standard deviations of each pair of states for couplings or dipole moments,

$$Y_\sigma^p = \frac{1}{2N_S^2} \sum_i^{N_S} \sum_j^{N_S} \left(\sqrt{\frac{1}{M-1} \sum_{m=1}^M (Y_m^{p_{ij}} - \overline{Y}^{p_{ij}})^2} \right). (5)$$

A separate threshold is set for each of these averaged quantities. If any of the quantities is predicted inaccurately, the data point is recomputed with quantum chemistry, phase corrected and added to the training set.

In order to make this process more efficient, not only one MD simulation, but an ensemble of trajectories can be computed. The ML models are only retrained after each of the independent trajectories has reached an untrustworthy region of the PES and after each of the reference calculations has been finished and included in the training set. This makes the parallelization of many trajectories possible [85, 118, 185, 198].

The adaptive sampling scheme should be carried out until the relevant conformational space for photodynamics is sampled sufficiently. However, using more than one ML model for production runs is still favorable. One of us and coworkers observed that the error of predictions decreases with the number of ML models used [85, 198]. We have seen the same trend in a recent study for NAMD simulations. With the adaptive sampling scheme for excited states, we generated a training set of 4,000 data points of the methylenimmonium cation, $CH_2NH_2^+$, to carry out long time-scale NAMD simulations with NNs [118].

2.5. Additional sampling techniques for excited states

Further training sets for NAMD simulations were generated for one-dimensional systems as well as polyatomic molecules. For example, Chen et al. have computed 90,000 data points via Born-Oppenheimer MD simulations and NAMD simulations, where emphasis was placed on the inclusion of geometries after a transition from one state to another took place [114]. Deep NNs were trained on energies and gradients of this training set to accurately reproduce NAMD simulations.

Hu et al. [112] used a very similar approach and obtained around 200,000 data points of 6-aminopyrimidine from Born-Oppenheimer MD simulations. They further carried out NAMD simulations with surface hopping [199, 200], where transitions from one state to another were allowed via so-called hops.

The geometries that were visited shortly before a hop took place were used as a starting guess to optimize conical intersections, i.e. critical points of the PESs, where two states become degenerate. Those data points were included to comprehensively sample the regions around a conical intersection. However, the ML models were not accurate enough for NAMD simulations solely based on ML potentials and the authors had to resort to quantum chemistry calculations in critical regions of the PESs.

Dral et al. [113] generated a training set for a one-dimensional two-state spin-boson model consisting of 10,000 data points with a grid-based method. The training data selection was then based on the structure, rather than on the energy of the molecules. For each data point, a molecular descriptor was computed and the distances of the descriptors were compared. Data points for the training set were chosen to sample the relevant space sufficiently [113, 201]. Compared to random sampling, this method allowed a reduction of training set sizes up to 90 %, which was shown for static calculations of the methyl chloride molecule [201, 202]. A similar structure-based sampling scheme was proposed by Ceriotti et al. [203].

Additionally, a maximum and minimum value can be computed for each representation of a molecule inside the training set. Every new structure that is obtained throughout an MD run can be compared to those values to get a measure of reliability of ML predictions. If the configuration does not lie within the known region, it can be added to the training set [184, 204]. Very recently, an active learning approach has been proposed to construct PESs without the need of running MD trajectories. The difference between two NN potentials was computed and points were iteratively added at the maxima of this difference surface (or, as phrased in the study, at the minima of the negative difference) [205].

It becomes evident from the diversity of approaches and training set sizes that a general guide on how to compute the training set and how large it should be for NAMD simulations can not be given. It is rather a matter of the efficiency that should be achieved and the computational costs that can be justified. Further, the training set strongly depends on the molecule under investigation. Especially its size, flexibility and the complexity of the light-induced dynamics play an important role. A molecule, whose photodynamics can be described as a two-state problem, such as in a simplified case of ethylene [206, 207], or a molecule, which is rigid, where dynamics mostly lead towards one reaction channel, possibly requires less data points than molecules that exhibit several different reaction channels after photo-excitation.

3. Machine learning nonadiabatic molecular dynamics simulations (ML NAMD)

3.1. Beyond Born-Oppenheimer dynamics

With an accurate training set for excited states at hand, NAMD simulations can be enhanced with ML models in order to enable the dynamics on time scales otherwise unfeasible. The most accurate way to study the dynamics of a molecule would be the full quantum mechanical treatment, which is, however, expensive and limited to a few atoms, even if ML PESs are applied [49, 115–117, 119–125, 127, 208–211]. A mixed quantum classical treatment is thus often preferred, where the motion of the nuclei are treated classically on one of the ML PESs. The mixed quantum classical MD simulation can then be interpreted as a mixed MLMD simulation [165]. The Born-Oppenheimer approximation allows to separate the nuclear from the electronic degrees of freedom. However, this approximation is not valid in the vicinity of avoided state crossings of PESs (or conical intersections, as mentioned before), which play an important role in excited-state dynamics.

In these critical regions of the PESs, ultrafast rearrangement of the motions of the electrons and the nuclei takes place due to strong couplings. As already mentioned, the relevant coupling elements are NACs and SOCs. The NACs (denoted as C^{NAC}) are vectorial properties and can be computed as [48, 212, 213]

$$
\begin{aligned}
C_{ij}^{\mathrm{NAC}} &\approx \langle \Psi_i \mid \tfrac{\partial}{\partial \mathbf{R}} \Psi_j \rangle \\
&= \tfrac{1}{E_i - E_j} \langle \Psi_i \mid \tfrac{\partial H_{el}}{\partial \mathbf{R}} \mid \Psi_j \rangle \quad \text{for } i \neq j,
\end{aligned} \tag{6}
$$

neglecting the second order derivatives. Thus, in the vicinity of a conical intersection, the couplings become very large, whereas they are almost vanishing elsewhere. The singularities that arise when two states are degenerate do not only pose an obstacle to quantum chemistry, but consequently also to PESs fitted with ML [112–114, 118]. NACs are nevertheless important properties to determine the direction and probability of internal conversion – a transition from one state to another, where the spin multiplicity does not change [8, 48, 51, 52, 214]. In contrast, the SOCs (denoted as C^{SOC}) are complex-valued properties that determine the rate of intersystem crossing, i.e., the transitions from one state to another, where spin multiplicity does change. In standard quantum chemistry programs, SOCs are given as the off-diagonal elements of the Hamiltonian matrix [8, 52, 215]:

$$
C_{ij}^{\mathrm{SOC}} = \langle \Psi_i \mid \hat{H}^{SOC} \mid \Psi_j \rangle. \tag{7}
$$

3.2. Fitting diabatic potentials

The numerical difficulties that arise due to discontinuous PESs and singularities of couplings at conical

Machine learning and excited-state molecular dynamics 9

intersections can be circumvented by the use of diabatic potentials instead of adiabatic ones [116, 216–220]. In the diabatic basis, the coupling elements are smooth properties and the arbitrary phase of the wavefunction does not have an impact. This favors the use of diabatic PESs. Since the output of a quantum chemistry program is generally given in the adiabatic basis, a quasi-diabatization procedure is necessary. Strictly speaking, a diabatization procedure is not possible because e.g. an infinite number of states is needed for an accurate representation. If using a finite number of states, the term quasi-diabatic is employed. For simplicity, we still use the notation of diabatic potentials for quasi-diabatic potentials. Those have been generated with different methods [221] and for small molecules up to date. Examples are the propagation diabatization [222], diabatization by localization [223] or by ansatz [115, 224], diabatization based on couplings or other properties [225–228], configuration uniformity [229], block-diagonalization [230], CI vectors [169] or (partly) on ML [115, 116, 224, 231–234].

Since several years, (modified) Shepard interpolation is used to fit diabatic potentials [129, 235–238] and also least squares fitting was applied to study the photo dissociation of molecules, such as NH_3 and phenol [239, 240]. In a series, Guo, Yarkony and co-workers developed invariant polynomial NNs [116, 124, 231–234] to address the excited-state dynamics of NH_3 and H_2O by fitting diabatic potential energy matrix elements. Absorption spectra as well as branching ratios could be obtained with high accuracy. The same authors further fit the diabatic $1,2^1A$ dipole moment surfaces of NH_3, which can only be fitted accurately if the topography of the PESs is reproduced correctly, validating the previously fitted diabatic potentials [119].

Habershon, Richings and co-workers used Gaussian process regression (in their notation equal to kernel-ridge regression) to fit diabatic potentials to execute on-the-fly dynamics of the butatrien cation with variational Gaussian wavepackets [127]. In another study, they applied an on-the-fly MCTDH scheme (DD-MCTDH) and carried out 4-mode/2-state simulations of pyrazine [120]. By improving the ML approach with a systematic tensor decomposition of kernel ridge regression, the study of 12-mode/2-state dynamics of pyrazine was rendered possible. This achievement remains a huge improvement over current MCTDH simulations in terms of accuracy and efficiency [241].

For the improvement of the diabatization by ansatz procedure, Williams et al. [115] applied NNs and enabled the fitting of the electronic low lying states of NO_3. The improvement of the diabatization procedure itself is desirable [116, 117], since the generation of meaningful diabatic potentials is often a tedious task and restricts their use tremendously. Up to date,

no rigorous diabatization procedure exists that allows the diabatization of adiabatic potentials of polyatomic systems by non-experts in this field [116, 119]. Especially challenging for larger and more complex systems is the number of electronic states within a certain energy range that have to be considered for successful diabatization. An increasing computational effort to provide all relevant electronic states is the result, making diabatization further challenging [169]. Often, more extensive approximations [115, 169, 242], e.g. the linear vibronic coupling model [243] are applied. We refer the reader to Ref.s [50, 217, 243–245] for more details on such approaches.

Despite the advantages of diabatic potentials, due to the before-mentioned drawbacks and the fact that the direct output of a quantum chemical calculation is given in the adiabatic basis, on-the-fly NAMD in the adiabatic representation is often the method of choice for large polyatomic systems, which will be discussed in the following.

3.3. Fitting adiabatic potentials

In order to execute NAMD simulations in the adiabatic basis, approximations have to be introduced to account for nonadiabatic transitions between different PESs. A good trade-off between accuracy and efficiency can be achieved with the surface hopping methodology [199, 200], which is often applied in ML-based NAMD simulations [112–114, 118]. In surface hopping, the transitions, or so-called hops, between different states, are computed stochastically and a manifold of trajectories needs to be taken into account to analyze different reaction channels and branching ratios [8, 51, 246] Several algorithms [52, 246–248] are frequently used to compute the hopping probability as well as its direction, with Tully's fewest switching algorithm being among the most popular ones [199, 200]. There, the couplings between adjacent states determine the hopping probability [51]. Other frequently applied algorithms to compute hopping probabilities are the Landau-Zener [249, 250] and the Zhu-Nakamura approximations [247, 248, 251, 252]. Those approximations solely rely on the PESs and omit the computation of wavefunction coefficients and couplings. Other flavors to account for transitions exist, which have, however, not been used in ML based NAMD studies yet. We thus refer the reader to Ref.s [48, 51, 52, 170, 199, 246, 251–256] for further information.

NAMD simulations with ML energies and forces

Based on the fewest switches algorithm, one of the first ML NAMD simulation is carried out by Carbogno et al. [130], where the scattering of O_2 at Al(III)

is studied [111, 130, 257]. A set of 3768 carefully selected data points [258] allowed for interpolation of the PESs with NNs [257]. In a first attempt, the authors include a spin-unpolarized singlet PES and a spin-polarized triplet state. Strictly speaking, the output of a quantum chemistry simulation for a singlet and triplet state is spin-diabatic [52, 215] and NAMD simulations ideally should carry out a diagonalization to obtain the spin-adiabatic PESs [52, 215]. The authors took advantage of the adiabatic spin-polarized PES [259] to compute the absolute value of couplings [130]. In this way, the transitions between the states could be approximated using surface hopping omitting the computation of wavefunction dynamics. This study was extended by two-state NAMD simulations of different multiple PESs arising from different spin configurations. Findings suggested a high probability of singlet-to-triplet conversion during scattering experiments with a non-zero probability even at low coupling values [111, 130, 257].

Other studies using the Zhu-Nakamura method [247, 248, 251, 252] to account for nonadiabatic transitions are discussed below. This approximation is based solely on energies and neglects the phase of the wavefunction. As a drawback, PESs are always assumed to couple to each other, when they are close in energy. This holds true for many cases, but one must be aware that strongly and weakly coupled PESs can not be distinguished.

Hu et al. [112] for example trained separate kernel ridge regression models to fit three singlet states of 6-aminopyrimidine. For learning, they used 65,316 data points comprising the molecular structures and energies of 6-aminopyrimidine with gradients not fitted, but computed afterwards. The data points were obtained from Born-Oppenheimer simulations, which were further clustered into sub-groups, from which the training points were selected randomly. As mentioned before, hopping geometries obtained from reference NAMD simulations were taken to find minimum conical intersections and the latter were also included in the training set. In contrast, Chen et al. [114] trained two deep NNs on 90,000 data points of two singlet states of CH_2NH. Again, data points were obtained from Born-Oppenheimer MD simulations and NAMD simulations starting from hopping geometries. In both studies, the NAMD simulations of the reference method could be successfully reproduced.

Instead of approximating the hopping probability, the NACs can also be approximated from PESs, gradients and Hessians [208, 260–264]. We made use of this relation and the fact that ML Hessians can be computed efficiently, and carried out NAMD simulations with the surface hopping method for sulfurdioxide, thioformaldehyde and the methylenimmonium

cation [165].

NAMD simulations with ML energies, forces, and couplings

In addition to energies and forces, SOCs need to be fitted with ML models when states of different spin multiplicities become relevant. Furthermore, when approximative schemes for the computation of hopping probabilities fail, the ML models need to learn NACs. One of the first studies, where NACs were fitted, used 1,000 and 10,000 data points to train kernel-ridge regression models to reproduce NAMD simulations of a one-dimensional system. However, especially in critical regions of the PESs, the ML models could not replace quantum chemical calculations and so 13-16% electronic structure calculations were required during an NAMD simulation [113]. The authors highlighted this as a drawback, because efficient simulations should be performed purely with ML and should not rely on intermediate quantum chemical calculations. Moreover, each entry of the NAC vectors was fitted by a separate kernel-ridge regression model, which turned out to be insufficiently accurate.

As indicated before, we also aimed for reproducing NAMD simulations with ML. We employed multilayer feed-forward NNs trained on 4,000 data points of 3 singlet states of $CH_2NH_2^+$ [118]. Short reference NAMD simulations based on electronic structure calculations could be reproduced. With the ML NAMD, long simulation times on the order of a nanosecond were successfully reached. Significantly different from previous ML NAMD approaches is the smaller size of the training set required to reproduce NAMD simulations. Further, a multi-output ML model was used to fit all NAC vectors between different states of same spin multiplicity at once. We term such models multi-state models. Per definition, kernel-ridge regression, and similar approaches such as linear regression, are single-state models. In order to make multi-state predictions of such models possible, the energetic state has to be encoded explicitly by using for example an additional state kernel. This procedure enables to model several states simultaneously. We studied the use of multi-state descriptors with the QML toolkit [265] for kernel-ridge regression models and showed that a multi-state description is generally superior to a single-state description in terms of accuracy [161].

Lastly, we want to comment on the NACs as vectorial properties. It should be clarified that approaches relating a molecular input directly to NAC values do not provide rotational covariance. This drawback is independent of a single-state treatment, i.e., the use of a separate ML model for each coupling value, or a multi-state treatment, where all values are

represented in one ML model. Very recently, Zhang et al. applied a symmetry-adapted high-dimensional NN [266] and treated the couplings as derivatives of NN representations. In this case, electronic friction was modelled via ML and applied for MD simulations of molecules at metal surfaces to treat the electron-nuclei coupling in a rotationally covariant manner. For the NAC vectors, we applied a similar strategy (similar also to force-only training for potentials), and implemented them as derivatives of virtual properties (i.e., non-existent in quantum chemistry) built by a deep NN [165].

3.4. Choosing the right descriptor

Many of the aforementioned studies use kernel ridge regression models or NNs in combination with distance-based descriptors [99, 112–114, 118, 267] such as the matrix of inverse distances or the Coulomb matrix [76]. It is worth mentioning that the accuracy of the ML PESs also depends on the type of descriptor. Molecular descriptors that represent atoms in their chemical and structural environment are often superior to those who treat complete molecules [101, 154, 155]. The symmetry functions of Behler [268, 269], their weighted counterparts [175, 270] or the FCHL (Faber-Christensen-Huang-Lilienfeld) representation [101, 155] work very well for NNs and the latter also for kernel-ridge regression and additionally provide permutation invariance.

Further improvement can be provided by message passing neural networks [271]. Compared to hand-crafted molecular descriptors, the representation of molecules can be seen as a part of a deep NN and, thus, is generated automatically. For each training set, an accurate descriptor is intrinsically designed, which accounts for the chemical and structural environment of a molecule. Examples for such networks are SchNet [156, 172], the DTNN [133], PhysNet [272], or HIP-NN [273]. For the excited states, the SchNarc [165] approach offers this type of descriptor.

4. Conclusion and outlook

To conclude, ML methods are very powerful and can be used to speed up current MD approaches for the excited states. They have been successfully applied to circumvent existing problems due to the expenses of the underlying electronic structure methods, but there is still a long way ahead of us to make ML applicable for the photodynamics simulations of large and complex systems.

While the fitting of diabatic potentials is generally more favorable, those methods are limited by the challenges that arise in finding meaningful diabatic potentials. Up to date, diabatization procedures are tedious and often not feasible for large and complex systems. ML models have been successfully applied to improve these processes [115–117], but methods to treat large and complex polyatomic systems in the diabatic basis are still lacking.

To investigate the photodynamics of polyatomic molecules, mixed quantum-classical MD simulations in the adiabatic basis thus often remain the method of choice. One advantage is that the direct output of a quantum chemical calculation is given in the adiabatic basis and so the obtained potential energies and forces can be directly fitted with an ML model. By applying approximations for the computation of transition probabilities from one state to another, the photodynamics can be studied efficiently with ML [112, 114, 165]. When approaches aim for additionally fitting the coupling values between different electronic states, inconsistencies in the data need to be considered carefully. Those have to be either removed from the training set or the training process itself has to be adapted in order to achieve successful training. Both approaches have been applied and were used for NAMD simulations [118, 165].

Unfortunately, also mixed quantum-classical MD simulations in the adiabatic basis have been restricted to small, isolated systems or single reactions, when described with ML PESs. At the current stage of research, many challenges remain that need to be tackled when replacing quantum chemical calculations in photodynamics simulations of large and complex polyatomic molecules. The most severe issue is to identify an adequate quantum chemical reference method. Up to date, mainly multi-reference methods have been used to compute training sets for small molecular systems for the excited states. However, the computational effort increases rapidly with the number of atoms and electronic excited states considered making single-reference methods more viable in this regards. The drawback of the latter methods is that their PESs can be even qualitatively wrong in some conformational regions. A tedious exclusion of such regions could in principle allow for a comprehensive training set generation and the accurate fitting of the manifold of energetic states, forces, and couplings between them in the remaining regions. Nonetheless, the resulting ML models might still be inappropriate for dynamics simulations due to the validity in only a restricted conformational space. Furthermore, the PESs from common single-reference methods or approximate methods like Time-Dependent DFT based Tight Binding are likely to be at least quantitatively incorrect. Since even quantitatively small errors in the potential energy can accumulate to completely wrong dynamics over millions of time steps, the validity of ML models based on such approximate methods

Machine learning and excited-state molecular dynamics 12

for long time scale dynamics is questionable. It is unclear whether such quantitatively wrong potentials can still lead to qualitatively correct trends for the temporal evolution of a system in some cases. Studies comparing such long photodynamics simulations with experimental observables could provide an answer to this question but are not yet available.

The large number of data points that have been required with many recent ML approaches for small polyatomic molecules is particularly concerning in the view of treating larger systems. Here, intelligent algorithms for efficient training data set generation with the smallest possible number of points are required. In addition, energies, forces, and couplings were often trained in separate ML models, leading to unsatisfactory accuracy. The development of an ML model that can treat all properties for photodynamics simulations at once and enables a derivation of relevant properties from ML PESs is clearly desirable. The latter can potentially reduce the amount of required training data to precisely fit quantum chemical properties and further allows to omit their quantum chemical computation during dynamics. The ultimate solution would be an ML multi-reference wave function of a molecular system, but this envisioned dream has not yet been realized.

To enable a full exploitation of the advantages of ML, an ML model should optimally be applicable throughout chemical compound space. While current studies struggle with fitting the excited states of one molecule, the transferability of ML potentials for the excited states is far from being achieved. The description of the photochemistry of a biologically more relevant system, such as a DNA strand or peptide chain, from local contributions of single building blocks, i.e., DNA bases or single amino acids, is one of the biggest benefits an ML model can offer. However, the construction of excited-state PESs from local atomic contributions has not yet been shown to hold and a description of a whole molecular system limits the applicability of the ML model to other systems. At the current stage of research, the small molecules that have been successfully described in their excited states with atom-wise descriptors are not large enough to prove the validity of locally constructed excited state PESs. An estimation of the excited-state locality could also pave the way towards excited-state ML/MM simulations, similar to QM/MM (quantum mechanics/molecular mechanics). Further investigations along the aforementioned avenues are needed to gain insights into the possibilities of ML models to describe the excited states of larger and more complex systems and to fit more than one small, isolated molecule.

References

[1] B. Cohen, C. E. Crespo-Hernández, P. M. Hare, and B. Kohler, *Ultrafast Excited-State Dynamics in DNA and RNA Polymers*, ch. Ultrafast Excited-State Dynamics in DNA and RNA Polymers, pp. 463–470. Amsterdam: Elsevier, 2004.

[2] B. G. Levine and T. J. Martínez, "Isomerization through conical intersections," *Annu. Rev. Phys. Chem.*, vol. 58, pp. 613–634, 2007.

[3] N. J. Turro, V. Ramamurthy, and J. C. Scaiano, "Principles of molecular photochemistry: An introduction," in *Principles of Molecular Photochemistry: An Introduction*, University Science Books, 2009.

[4] D. R. Yarkony, "Nonadiabatic quantum chemistry - past, present, and future," *Chem. Rev.*, vol. 112, no. 1, pp. 481–498, 2012.

[5] M. Barbatti, A. C. Borin, and S. Ullrich, "Photoinduced processes in nucleic acids," in *Photoinduced Phenomena in Nucleic Acids I*, vol. 355 of *Topics in Current Chemistry*, pp. 1–32, Springer Berlin Heidelberg, 2014.

[6] L. M. Ibele, A. Nicolson, and B. F. E. Curchod, "Excited-state dynamics of molecules with classically driven trajectories and gaussians," *Mol. Phys.*, vol. 0, no. 0, pp. 1–18, 2019.

[7] 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.*, vol. 0, no. 0, p. null, 0.

[8] S. Mai and L. Gonzlez, "Molecular photochemistry: Recent developments in theory," *Angew. Chem. Int. Ed.*, vol. n/a, no. n/a, 2020.

[9] S. Matsika and A. I. Krylov, "Introduction: Theoretical modeling of excited state processes," *Chem. Rev.*, vol. 118, no. 15, pp. 6925–6926, 2018.

[10] H. Lischka, D. Nachtigallová, A. J. A. Aquino, P. G. Szalay, F. Plasser, F. B. C. Machado, and M. Barbatti, "Multireference approaches for excited states of molecules," *Chem. Rev.*, vol. 118, no. 15, pp. 7293–7361, 2018.

[11] S. Ghosh, P. Verma, C. J. Cramer, L. Gagliardi, and D. G. Truhlar, "Combining wave function methods with density functional theory for excited states," *Chem. Rev.*, vol. 118, no. 15, pp. 7249–7292, 2018.

[12] P. Norman and A. Dreuw, "Simulating x-ray spectroscopies and calculating core-excited states of molecules," *Chem. Rev.*, vol. 118, no. 15, pp. 7208–7248, 2018.

[13] D. Casanova, "Theoretical modeling of singlet fission," *Chem. Rev.*, vol. 118, no. 15, pp. 7164–7207, 2018.

[14] N. J. Hestand and F. C. Spano, "Expanded theory of h- and j-molecular aggregates: The effects of vibronic coupling and intermolecular charge transfer," *Chem. Rev.*, vol. 118, no. 15, pp. 7069–7163, 2018.

[15] T. J. Penfold, E. Gindensperger, C. Daniel, and C. M.

Marian, "Spin-vibronic mechanism for intersystem crossing," *Chem. Rev.*, vol. 118, no. 15, pp. 6975–7025, 2018.

[16] M. Vacher, I. Fdez. Galván, B.-W. Ding, S. Schramm, R. Berraud-Pache, P. Naumov, N. Ferré, Y.-J. Liu, I. Navizet, D. Roca-Sanjun, W. J. Baader, and R. Lindh, "Chemi- and bioluminescence of cyclic peroxides," *Chem. Rev.*, vol. 118, no. 15, pp. 6927–6974, 2018.

[17] R. Crespo-Otero and M. Barbatti, "Recent advances and perspectives on nonadiabatic mixed quantumclassical dynamics," *Chem. Rev.*, vol. 118, no. 15, pp. 7026–7068, 2018.

[18] L. González and R. Lindh, *Quantum Chemistry and Dynamics of Excited States : Methods and Applications*. John Wiley and Sons Ltd, 2020.

[19] D. C. Harris and M. D. Bertolucci, *Symmetry and spectroscopy: an introduction to vibrational and electronic spectroscopy*. New York: Dover Publications, 1989.

[20] C.-Y. Ng, *Vacuum ultraviolet photoionization and photodissociation of molecules and clusters*. World Scientific, 1991.

[21] A. H. Zewail, *Femtochemistry*, pp. 3–22. World Scientific, 1994.

[22] T. Brixner, T. Pfeifer, G. Gerber, M. Wollenhaupt, and T. Baumert, "Optimal control of atomic, molecular and electron dynamics with tailored femtosecond laser pulses," in *Femtosecond Laser Spectroscopy* (P. Hannaford, ed.), pp. 225–266, New York: Springer-Verlag, 2005.

[23] Maria Teresa Neves-Petersen and Søren Klitgaard and Torbjorn Pascher and Esben Skovsen and Tomas Polivka and Arkady Yartsev and Villly Sundström and Steffen B. Petersen, "Flash Photolysis of Cutinase: Identification and Decay Kinetics of Transient Intermediates Formed upon UV Excitation of Aromatic Residues," *Biophys. J.*, vol. 97, no. 1, pp. 211 – 226, 2009.

[24] A. Iqbal and V. G. Stavros, "Active participation of $1\pi\sigma^*$ states in the photodissociation of tyrosine and its subunits," *J. Phys. Chem. Lett.*, vol. 1, no. 15, pp. 2274–2278, 2010.

[25] M. Mališ, Y. Loquais, E. Gloaguen, H. S. Biswal, F. Piuzzi, B. Tardivel, V. Brenner, M. Broquier, C. Jouvet, M. Mons, *et al.*, "Unraveling the mechanisms of nonradiative deactivation in model peptides following photoexcitation of a phenylalanine residue," *J. Am. Chem. Soc.*, vol. 134, no. 50, pp. 20340–20351, 2012.

[26] M. Kowalewski, B. P. Fingerhut, K. E. Dorfman, K. Bennett, and S. Mukamel, "Simulating coherent multidimensional spectroscopy of nonadiabatic molecular processes: From the infrared to the x-ray regime," *Chem. Rev.*, vol. 117, no. 19, pp. 12165–12226, 2017.

[27] S. Soorkia, C. Jouvet, and G. Grégoire, "Uv photoinduced dynamics of conformer-resolved aromatic peptides," *Chem. Rev.*, vol. 0, no. 0, p. null, 0.

[28] A. Tajti, G. Fogarasi, and P. G. Szalay, "Reinterpretation of the UV spectrum of cytosine: Only two electronic transitions?," *ChemPhysChem*, vol. 10, no. 9-10, pp. 1603–1606, 2009.

[29] M. Barbatti, J. J. Szymczak, A. J. A. Aquino, D. Nachtigallová, and H. Lischka, "The decay mechanism of photoexcited guanine – a nonadiabatic dynamics study," *J. Chem. Phys.*, vol. 134, no. 1, p. 014304, 2011.

[30] Y. Lu, Z. Lan, and W. Thiel, "computational modeling of photoexcitation in DNA single and double strands," in *Photoinduced Phenomena in Nucleic Acids II*, vol. 356 of *Topics in Current Chemistry*, pp. 89–122, Springer

Berlin Heidelberg, 2014.

[31] M. Ruckenbauer, S. Mai, P. Marquetand, and L. Gonzlez, "Photoelectron spectra of 2-thiouracil, 4-thiouracil, and 2,4-dithiouracil," *J. Chem. Phys.*, vol. 144, no. 7, p. 074303, 2016.

[32] M. Manathunga, X. Yang, H. L. Luk, S. Gozem, L. M. Frutos, A. Valentini, N. Ferrè, and M. Olivucci, "Probing the photodynamics of rhodopsins with reduced retinal chromophores," *J. Chem. Theory Comput.*, vol. 12, no. 2, pp. 839–850, 2016.

[33] J. J. Nogueira, F. Plasser, and L. González, "Electronic delocalization, charge transfer and hypochromism in the uv absorption spectrum of polyadenine unravelled by multiscale computations and quantitative wavefunction analysis," *Chem. Sci.*, vol. 8, pp. 5682–5691, 2017.

[34] S. Mai, A. Mohamadzade, P. Marquetand, L. González, and S. Ullrich, "Simulated and experimental time-resolved photoelectron spectra of the intersystem crossing dynamics in 2-thiouracil," *Molecules*, vol. 23, no. 11, p. 2836, 2018.

[35] C. Rauer, J. J. Nogueira, P. Marquetand, and L. González, "Stepwise photosensitized thymine dimerization mediated by an exciton intermediate," *Monatsh. Chem.*, vol. 149, p. 19, 2018.

[36] J. P. Zobel, M. Heindl, J. J. Nogueira, and L. González, "Vibrational sampling and solvent effects on the electronic structure of the absorption spectrum of 2-nitronaphthalene," *J. Chem. Theory Comput.*, vol. 14, no. 6, pp. 3205–3217, 2018.

[37] M. Barbatti, A. J. A. Aquino, and H. Lischka, "Ultrafast two-step process in the non-adiabatic relaxation of the CH_2NH_2 molecule," *Mol. Phys.*, vol. 104, no. 5-7, pp. 1053–1060, 2006.

[38] B. F. E. Curchod and I. Tavernelli, "On trajectory-based nonadiabatic dynamics: Bohmian dynamics versus trajectory surface hopping," *J. Chem. Phys.*, vol. 138, no. 18, 2013.

[39] A. V. Akimov, R. Asahi, R. Jinnouchi, and O. V. Prezhdo, "What makes the photocatalytic CO_2 reduction on n-doped ta2o5 efficient: Insights from nonadiabatic molecular dynamics," *J. Am. Chem. Soc.*, vol. 137, no. 35, pp. 11517–11525, 2015.

[40] I. Schapiro, D. Roca-Sanjuán, R. Lindh, and M. Olivucci, "A surface hopping algorithm for nonadiabatic minimum energy path calculations," *J. Comput. Chem.*, vol. 36, no. 5, pp. 312–320, 2015.

[41] C. Rauer, J. J. Nogueira, P. Marquetand, and L. González, "Cyclobutane thymine photodimerization mechanism revealed by nonadiabatic molecular dynamics," *J. Am. Chem. Soc.*, vol. 138, pp. 15911–15916, 2016.

[42] M. Ruckenbauer, S. Mai, P. Marquetand, and L. González, "Revealing deactivation pathways hidden in time-resolved photoelectron spectra," *Sci. Rep.*, vol. 6, p. 35522, 2016.

[43] S. Mai, M. Richter, P. Marquetand, and L. González, "The dna nucleobase thymine in motion intersystem crossing simulated with surface hopping," *Chem. Phys.*, vol. 482, pp. 9 – 15, 2017.

[44] S. Mai and L. González, "Unconventional two-step spin relaxation dynamics of [Re(co)3(im)(phen)]+ in aqueous solution," *Chem. Sci.*, vol. 10, pp. 10405–10411, 2019.

[45] S. L. Horton, Y. Liu, R. Forbes, V. Makhija, R. Lausten, A. Stolow, P. Hockett, P. Marquetand, T. Rozgonyi, and T. Weinacht, "Excited state dynamics of CH_2I_2 and CH_2BrI studied with UV pump VUV probe photoelectron spectroscopy," *J. Chem. Phys.*, vol. 150, no. 17, p. 174201, 2019.

[46] P. Heim, S. Mai, B. Thaler, S. Cesnik, D. Avagliano, D. Bella-Velidou, W. E. Ernst, L. González, and

M. Koch, "Revealing ultrafast population transfer between nearly degenerate electronic states," *J. Phys. Chem. Lett.*, vol. 0, no. 0, pp. 1443–1449, 0.

[47] S. Mai, P. Marquetand, and L. González, "Intersystem crossing pathways in the noncanonical nucleobase 2-thiouracil: A time-dependent picture," *J. Phys. Chem. Lett.*, vol. 7, no. 0, pp. 1978–1983, 2016.

[48] N. L. Doltsinis, *Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum-Classical Approaches*, vol. 31 of *NIC Series*. John von Neuman Institut for computing, 2006.

[49] H. Köppel, W. Domcke, and L. S. Cederbaum, "Multimode molecular dynamics beyond the born-oppenheimer approximation," *Adv. Chem. Phys.*, vol. 57, pp. 59–246, 1984.

[50] G. A. Worth and L. S. Cederbaum, "Beyond born-oppenheimer: Molecular dynamics through a conical intersection," *Annu. Rev. Phys. Chem.*, vol. 55, no. 1, pp. 127–158, 2004.

[51] M. Richter, P. Marquetand, J. González-Vázquez, I. Sola, and L. González, "SHARC: Ab initio molecular dynamics with surface hopping in the adiabatic representation including arbitrary couplings," *J. Chem. Theory Comput.*, vol. 7, no. 5, pp. 1253–1258, 2011.

[52] S. Mai, P. Marquetand, and L. González, "Nonadiabatic Dynamics: The SHARC Approach," *WIREs Comput. Mol. Sci.*, vol. 8, p. e1370, 2018.

[53] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad, "Solving the electronic structure problem with machine learning," *npj Comput. Mater.*, vol. 5, no. 1, p. 22, 2019.

[54] G. Carleo and M. Troyer, "Solving the quantum many-body problem with artificial neural networks," *Science*, vol. 355, no. 6325, pp. 602–606, 2017.

[55] H. Saito, "Solving the bosehubbard model with machine learning," *J. Phys. Soc. Jpn.*, vol. 86, no. 9, p. 093001, 2017.

[56] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, "Restricted boltzmann machine learning for solving strongly correlated quantum systems," *Phys. Rev. B*, vol. 96, p. 205152, Nov 2017.

[57] J. Han, L. Zhang, and W. E, "Solving many-electron schrödinger equation using deep neural networks," 2018.

[58] J. Townsend and K. D. Vogiatzis, "Data-driven acceleration of the coupled-cluster singles and doubles iterative solver," *J. Phys. Chem. Lett.*, vol. 10, no. 14, pp. 4129–4135, 2019.

[59] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, and R. J. Maurer, "Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions," *Nat. Commun.*, vol. 10, p. 5024, 2019.

[60] D. Pfau, J. S. Spencer, A. G. de G. Matthews, and W. M. C. Foulkes, "Ab-initio solution of the many-electron schrödinger equation with deep neural networks," 2019.

[61] J. Hermann, Z. Schätzle, and F. Noé, "Deep neural network solution of the electronic schrödinger equation," 2019.

[62] M. Gastegger, A. McSloy, M. Luya, K. T. Schütt, and R. J. Maurer, "A Deep Neural Network for Molecular Wave Functions in Quasi-Atomic Minimal Basis Representation," *https://arxiv.org/abs/2005.06979*, 2020. preprint.

[63] K. Choo, G. Carleo, N. Regnault, and T. Neupert, "Symmetries and many-body excitations with neural-network quantum states," *Phys. Rev. Lett.*, vol. 121, p. 167204, Oct 2018.

[64] F. Zheng, X. Gao, and A. Eisfeld, "Excitonic wave function reconstruction from near-field spectra using machine learning techniques," *Phys. Rev. Lett.*, vol. 123, p. 163202, Oct 2019.

[65] F. Brockherde, L. Vogt, L. Li, M. E. Tuckerman, K. Burke, and K.-R. Müller, "Bypassing the kohn-sham equations with machine learning," *Nat. Commun.*, vol. 8, no. 1, p. 872, 2017.

[66] G. Hegde and R. C. Bowen, "Machine-learned approximations to density functional theory Hamiltonians," *Sci. Rep.*, vol. 7, p. 42669, 2017.

[67] M. Gastegger, L. González, and P. Marquetand, "Exploring density functional subspaces with genetic algorithms," *Monatsh. Chem.*, vol. 150, no. 2, pp. 173–182, 2019.

[68] J. Nelson, R. Tiwari, and S. Sanvito, "Machine learning density functional theory for the hubbard model," *Phys. Rev. B*, vol. 99, p. 075132, Feb 2019.

[69] L. Cheng, M. Welborn, A. S. Christensen, and T. F. Miller, "A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules," *J. Chem. Phys.*, vol. 150, no. 13, p. 131103, 2019.

[70] X. Lei and A. J. Medford, "Design and analysis of machine learning exchange-correlation functionals via rotationally invariant convolutional descriptors," *Phys. Rev. Materials*, vol. 3, p. 063801, Jun 2019.

[71] Y. Zhou, J. Wu, S. Chen, and G. Chen, "Toward the exact exchangecorrelation potential: A three-dimensional convolutional neural network construct," *J. Phys. Chem. Lett.*, vol. 10, no. 22, pp. 7264–7269, 2019.

[72] B. Kolb, L. C. Lentz, and A. M. Kolpak, "Discovering charge density functionals and structure-property relationships with prophet: A general framework for coupling machine learning and first-principles methods," *Sci. Rep.*, vol. 7, no. 1192, 2017.

[73] M. J. Willatt, F. Musil, and M. Ceriotti, "Atom-density representations for machine learning," *J. Chem. Phys.*, vol. 150, no. 15, p. 154110, 2019.

[74] S. Hobday, R. Smith, and J. Belbruno, "Applications of neural networks to fitting interatomic potential functions," *Modell. Simul. Mater. Sci. Eng.*, vol. 7, no. 3, p. 397, 1999.

[75] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, "Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons," *Phys. Rev. Lett.*, vol. 104, p. 136403, Apr 2010.

[76] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, "Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning," *Phys. Rev. Lett.*, vol. 108, no. 5, p. 058301, 2012.

[77] Z. Li, J. R. Kermode, and A. De Vita, "Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces," *Phys. Rev. Lett.*, vol. 114, p. 096405, Mar 2015.

[78] O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp, and A. Knoll, "Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties," *Int. J. Quantum Chem.*, vol. 115, no. 16, pp. 1084–1093, 2015.

[79] M. Gastegger and P. Marquetand, "High-dimensional neural network potentials for organic reactions and an improved training algorithm," *J. Chem. Theory Comput.*, vol. 11, no. 5, pp. 2187–2198, 2015.

[80] M. Rupp, R. Ramakrishnan, and O. A. von Lilienfeld, "Machine learning for quantum mechanical properties of atoms in molecules," *J. Phys. Chem. Lett.*, vol. 6, no. 16, pp. 3309–3313, 2015.

[81] J. Behler, "Perspective: Machine learning potentials for atomistic simulations," *J. Chem. Phys.*, vol. 145, no. 17, p. 170901, 2016.

[82] N. Artrith and A. Urban, "An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for tio₂," *computational Materials Science*, vol. 114, pp. 135 – 150, 2016.

[83] M. Gastegger, C. Kauffmann, J. Behler, and P. Marquetand, "Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes," *J. Chem. Phys.*, vol. 144, no. 19, 2016.

[84] N. Artrith, A. Urban, and G. Ceder, "Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species," *Phys. Rev. B*, vol. 96, p. 014112, Jul 2017.

[85] M. Gastegger, J. Behler, and P. Marquetand, "Machine learning molecular dynamics for the simulation of infrared spectra," *Chem. Sci.*, vol. 8, pp. 6924–6935, 2017.

[86] V. L. Deringer and G. Csányi, "Machine learning based interatomic potential for amorphous carbon," *Phys. Rev. B*, vol. 95, p. 094203, Mar 2017.

[87] V. Botu, R. Batra, J. Chapman, and R. Ramprasad, "Machine learning force fields: Construction, validation, and outlook," *J. Phys. Chem. C*, vol. 121, no. 1, pp. 511–522, 2017.

[88] A. Glielmo, P. Sollich, and A. De Vita, "Accurate interatomic force fields via machine learning with covariant kernels," *Phys. Rev. B*, vol. 95, no. 21, p. 214302, 2017.

[89] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with dft accuracy at force field computational cost," *Chem. Sci.*, vol. 8, pp. 3192–3203, 2017.

[90] S. Fujikake, V. L. Deringer, T. H. Lee, M. Krynski, S. R. Elliott, and G. Csányi, "Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures," *J. Chem. Phys.*, vol. 148, no. 24, p. 241714, 2018.

[91] J. Behler, "First principles neural network potentials for reactive simulations of large molecular and condensed systems," *Angew. Chem. Int. Edit.*, vol. 56, no. 42, pp. 12828–12840, 2017.

[92] H. Zong, G. Pilania, X. Ding, G. J. Ackland, and T. Lookman, "Developing an interatomic potential for martensitic phase transformations in zirconium by machine learning," *npj comput Mater*, vol. 4, 2018.

[93] M. A. Wood and A. P. Thompson, "Extending the accuracy of the snap interatomic potential form," *J. Chem. Phys.*, vol. 148, no. 24, p. 241721, 2018.

[94] X. Chen, M. S. Jørgensen, J. Li, and B. Hammer, "Atomic energies from a convolutional neural network," *J. Chem. Theory Comput.*, vol. 14, no. 7, pp. 3933–3942, 2018.

[95] A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, "Machine learning a general-purpose interatomic potential for silicon," *Phys. Rev. X*, vol. 8, p. 041048, 2018.

[96] S. Chmiela, H. E. Sauceda, K.-R. Müller, and A. Tkatchenko, "Towards exact molecular dynamics simulations with machine-learned force fields," *Nat. Commun.*, vol. 9, p. 3887, 2018.

[97] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, and M. Ceriotti, "Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials," *J. Chem. Phys.*, vol. 148, no. 24, p. 241730, 2018.

[98] L. Zhang, J. Han, H. Wang, W. A. Saidi, R. Car, and E. Weinan, "End-to-end symmetry preserving interatomic potential energy model for finite and extended systems," in *Proceedings of the 32Nd International conference on Neural Information Processing Systems*, NIPS'18, (USA), pp. 4441–4451, Curran Associates Inc., 2018.

[99] L. Zhang, J. Han, H. Wang, R. Car, and W. E, "Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics," *Phys. Rev. Lett.*, vol. 120, p. 143001, Apr 2018.

[100] H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Y. Chan, and S. K. R. S. Sankaranarayanan, "Machine learning classical interatomic potentials for molecular dynamics from first-principles training data," *J. Phys. Chem. C*, vol. 123, no. 12, pp. 6941–6957, 2019.

[101] F. A. Faber, A. S. Christensen, B. Huang, and O. A. von Lilienfeld, "Alchemical and structural distribution based representation for universal quantum machine learning," *J. Chem. Phys.*, vol. 148, no. 24, p. 241717, 2018.

[102] H. Wang and W. Yang, "Toward building protein force fields by residue-based systematic molecular fragmentation and neural network," *J. Chem. Theory Comput.*, vol. 15, no. 2, pp. 1409–1417, 2019.

[103] N. Gerrits, K. Shakouri, J. Behler, and G.-J. Kroes, "Accurate probabilities for highly activated reaction of polyatomic molecules on surfaces using a high-dimensional neural network potential: Chd₃ + cu(111)," *J. Phys. Chem. Lett.*, vol. 10, no. 8, pp. 1763–1768, 2019.

[104] S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Mller, and A. Tkatchenko, "sgdml: Constructing accurate and data efficient molecular force fields using machine learning," *Comput. Phys. Commun.*, vol. 240, pp. 38 – 45, 2019.

[105] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, "Machine learning and the physical sciences," *Rev. Mod. Phys.*, vol. 91, p. 045002, Dec 2019.

[106] R. V. Krems, "Bayesian Machine Learning for Quantum Molecular Dynamics," *Phys. Chem. Chem. Phys.*, vol. 21, pp. 13392–13410, 2019.

[107] V. L. Deringer, M. A. Caro, and G. Csnyi, "Machine learning interatomic potentials as emerging tools for materials science," *Adv. Mat.*, vol. 31, no. 46, p. 1902765, 2019.

[108] L. Ward, B. Blaiszik, I. Foster, R. S. Assary, B. Narayanan, and L. Curtiss, "Machine learning prediction of accurate atomization energies of organic molecules from low-fidelity quantum chemical calculations," *MRS Commun.*, vol. 9, no. 3, p. 891899, 2019.

[109] F. Noé, A. Tkatchenko, K.-R. Müller, and C. Clementi, "Machine learning for molecular simulation," *Annu. Rev. Phys. Chem.*, vol. 71, no. 1, pp. 361–390, 2020.

[110] J. Behler, K. Reuter, and M. Scheffler, "Nonadiabatic effects in the dissociation of oxygen molecules at the Al(111) surface," *Phys. Rev. B*, vol. 77, no. 11, p. 115421, 2008.

[111] C. Carbogno, J. Behler, K. Reuter, and A. Groß, "Signatures of nonadiabatic O₂ dissociation at Al(111): First-principles fewest-switches study," *Phys. Rev. B*, vol. 81, p. 035410, Jan 2010.

[112] D. Hu, Y. Xie, X. Li, L. Li, and Z. Lan, "Inclusion of machine learning kernel ridge regression potential energy surfaces in on-the-fly nonadiabatic molecular dynamics simulation," *J. Phys. Chem. Lett.*, vol. 9, no. 11, pp. 2725–2732, 2018.

[113] P. O. Dral, M. Barbatti, and W. Thiel, "Nonadiabatic excited-state dynamics with machine learning," *J. Phys. Chem. Lett.*, vol. 9, no. 0, pp. 5660–5663, 2018.

[114] W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral, and G. Cui, "Deep learning for nonadiabatic excited-state dynamics," *J. Phys. Chem. Lett.*, vol. 9, no. 23, pp. 6702–6708, 2018.

[115] D. M. G. Williams and W. Eisfeld, "Neural network diabatization: A new ansatz for accurate high-dimensional coupled potential energy surfaces," *J. Chem. Phys.*, vol. 149, no. 20, p. 204106, 2018.

[116] C. Xie, X. Zhu, D. R. Yarkony, and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces. IV. coupled diabatic potential energy matrices," *J. Chem. Phys.*, vol. 149, no. 14, p. 144107, 2018.

[117] Y. Guan, D. H. Zhang, H. Guo, and D. R. Yarkony, "Representation of coupled adiabatic potential energy surfaces using neural network based quasi-diabatic Hamiltonians: 1,2 2A' states of LiFH," *Phys. Chem. Chem. Phys.*, p. 10.1039/C8CP06598E, 2019.

[118] 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.*, vol. 10, pp. 8100–8107, 2019.

[119] Y. Guan, H. Guo, and D. R. Yarkony, "Extending the representation of multistate coupled potential energy surfaces to include properties operators using neural networks: Application to the 1,21a states of ammonia," *J. Chem. Theory Comput.*, vol. 16, no. 1, pp. 302–313, 2020.

[120] G. W. Richings and S. Habershon, "MCTDH on-the-fly: Efficient grid-based quantum dynamics without pre-computed potential energy surfaces," *J. Chem. Phys.*, vol. 148, no. 13, p. 134116, 2018.

[121] J. P. Alborzpour, D. P. Tew, and S. Habershon, "Efficient and accurate evaluation of potential energy matrix elements for quantum dynamics using gaussian process regression," *J. Chem. Phys.*, vol. 145, no. 17, p. 174112, 2016.

[122] G. W. Richings, C. Robertson, and S. Habershon, "Improved on-the-fly MCTDH simulations with many-body-potential tensor decomposition and projection diabatization," *J. Chem. Theory Comput.*, vol. 15, no. 2, pp. 857–870, 2019.

[123] I. Polyak, G. W. Richings, S. Habershon, and P. J. Knowles, "Direct quantum dynamics using variational gaussian wavepackets and gaussian process regression," *J. Chem. Phys.*, vol. 150, no. 4, p. 041101, 2019.

[124] Y. Guan, H. Guo, and D. R. Yarkony, "Neural network based quasi-diabatic Hamiltonians with symmetry adaptation and a correct description of conical intersections," *J. Chem. Phys.*, vol. 150, no. 21, p. 214101, 2019.

[125] Y. Wang, C. Xie, H. Guo, and D. R. Yarkony, "A quasi-diabatic representation of the 1,21a states of methylamine," *J. Phys. Chem. A*, vol. 123, no. 25, pp. 5231–5241, 2019.

[126] Y. Guan and D. R. Yarkony, "Accurate neural network representation of the ab initio determined spinorbit interaction in the diabatic representation including the effects of conical intersections," *J. Phys. Chem. Lett.*, vol. 11, no. 5, pp. 1848–1858, 2020.

[127] G. W. Richings and S. Habershon, "Direct grid-based quantum dynamics on propagated diabatic potential energy surfaces," *Chem. Phys. Lett.*, vol. 683, pp. 228 – 233, 2017.

[128] H. M. Netzloff, M. A. collins, and M. S. Gordon, "Growing multiconfigurational potential energy surfaces with applications to X+H$_2$ (X=C,N,O) reactions," *J. Chem. Phys.*, vol. 124, no. 15, p. 154104, 2006.

[129] R. P. A. Bettens and M. A. collins, "Learning to interpolate molecular potential energy surfaces with confidence: A bayesian approach," *J. Chem. Phys.*, vol. 111, no. 3, pp. 816–826, 1999.

[130] C. Carbogno, J. Behler, A. Groß, and K. Reuter, "Fingerprints for Spin-Selection Rules in the Interaction Dynamics of O$_2$ at Al(111)," *Phys. Rev. Lett.*, vol. 101, no. 9, p. 096104, 2008.

[131] K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari, and P. Rinke, "Deep learning spectroscopy: Neural networks for molecular excitation spectra," *Adv. Sci.*, vol. 6, no. 9, p. 1801367, 2019.

[132] G. Pilania, J. Gubernatis, and T. Lookman, "Multi-fidelity machine learning models for accurate bandgap predictions of solids," *Comput. Mat. Sci.*, vol. 129, pp. 156 – 163, 2017.

[133] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, and A. Tkatchenko, "Quantum-chemical insights from deep tensor neural networks," *Nat. Commun.*, vol. 8, pp. 13890 EP –, Jan 2017.

[134] Y. Zhuo, A. Mansouri Tehrani, and J. Brgoch, "Predicting the band gaps of inorganic solids by machine learning," *J. Phys. Chem. Lett.*, vol. 9, no. 7, pp. 1668–1673, 2018.

[135] F. Pereira, K. Xiao, D. A. R. S. Latino, C. Wu, Q. Zhang, and J. Aires-de Sousa, "Machine learning methods to predict density functional theory b3lyp energies of homo and lumo orbitals," *J. Chem. Inf. Model.*, vol. 57, no. 1, pp. 11–21, 2017.

[136] O. Isayev, c. Oses, c. Toher, E. Gossett, S. Curtarolo, and A. Tropsha, "Universal fragment descriptors for predicting properties of inorganic crystals," *Nat. Commun.*, vol. 8, p. 15679, 2017.

[137] W. Pronobis, K. R. Schütt, A. Tkatchenko, and K.-R. Müller, "Capturing intensive and extensive dft/tddft molecular properties with machine learning," *Eur. Phys. J. B*, vol. 91, p. 178, 2018.

[138] A. Stuke, M. Todorović, M. Rupp, C. Kunkel, K. Ghosh, L. Himanen, and P. Rinke, "Chemical diversity in molecular orbital energy predictions with kernel ridge regression," *J. Chem. Phys.*, vol. 150, no. 20, p. 204121, 2019.

[139] F. Häse, S. Valleau, E. Pyzer-Knapp, and A. Aspuru-Guzik, "Machine learning exciton dynamics," *Chem. Sci.*, vol. 7, pp. 5139–5147, 2016.

[140] 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.*, vol. 10, pp. 2298–2307, 2019.

[141] N. M. OBoyle, C. M. Campbell, and G. R. Hutchison, "computational design and selection of optimal organic photovoltaic materials," *J. Phys. Chem. C*, vol. 115, no. 32, pp. 16200–16210, 2011.

[142] J. L. Teunissen, F. De Proft, and F. De Vleeschouwer, "Tuning the homolumo energy gap of small diamondoids using inverse molecular design," *J. Chem. Theory Comput.*, vol. 13, no. 3, pp. 1351–1365, 2017.

[143] D. Liu, Y. Tan, E. Khoram, and Z. Yu, "Training deep neural networks for the inverse design of nanophotonic structures," *ACS Photonics*, vol. 5, no. 4, pp. 1365–1369, 2018.

[144] D. C. Elton, Z. Boukouvalas, M. D. Fuge, and P. W. Chung, "Deep learning for molecular design - a review of the state of the art," *Mol. Syst. Des. Eng.*, vol. 4, pp. 828–849, 2019.

[145] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," *Science*, vol. 361, no. 6400, pp. 360–365, 2018.

[146] B. R. Goldsmith, J. Esterhuizen, J.-X. Liu, C. J. Bartel, and C. Sutton, "Machine learning for heterogeneous catalyst design and discovery," *AIChE J.*, vol. 64, no. 7, pp. 2311–2323, 2018.

[147] D. W. Davies, K. T. Butler, O. Isayev, and A. Walsh, "Materials discovery by chemical analogy: role of oxidation states in structure prediction," *Faraday*

Discuss., vol. 211, pp. 553–568, 2018.

[148] O. Anatole von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Exploring Chemical compound Space with Quantum-Based Machine Learning," *arXiv:1911.10084*, 2019.

[149] J. G. Freeze, H. R. Kelly, and V. S. Batista, "Search for catalysts by inverse design: Artificial intelligence, mountain climbers, and alchemists," *Chemical Reviews*, vol. 119, no. 11, pp. 6595–6612, 2019.

[150] M.-H. Lee, "Robust random forest based non-fullerene organic solar cells efficiency prediction," *Org. Electron.*, vol. 76, p. 105465, 2020.

[151] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, "Machine learning of molecular electronic properties in chemical compound space," *New J. Phys.*, vol. 15, no. 9, p. 095003, 2013.

[152] K. Hansen, F. Biegler, R. Ramakrishnan, W. Pronobis, O. A. von Lilienfeld, K.-R. Müller, and A. Tkatchenko, "Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space," *J. Phys. Chem. Lett.*, vol. 6, no. 12, pp. 2326–2331, 2015.

[153] B. Huang and O. A. von Lilienfeld, "communication: Understanding molecular representations in machine learning: The role of uniqueness and target similarity," *J. Chem. Phys.*, vol. 145, no. 16, p. 161102, 2016.

[154] A. S. Christensen, F. A. Faber, and O. A. von Lilienfeld, "Operators in quantum machine learning: Response properties in chemical space," *J. Chem. Phys.*, vol. 150, no. 6, p. 064105, 2019.

[155] A. S. Christensen, L. A. Bratholm, F. A. Faber, and O. Anatole von Lilienfeld, "Fchl revisited: Faster and more accurate quantum machine learning," *J. Chem. Phys.*, vol. 152, no. 4, p. 044107, 2020.

[156] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet – a deep learning architecture for molecules and materials," *J. Chem. Phys.*, vol. 148, no. 24, p. 241722, 2018.

[157] S. A. Tawfik, O. Isayev, M. J. S. Spencer, and D. A. Winkler, "Predicting thermal properties of crystals using machine learning," *Adv. Theory Sim.*, vol. n/a, no. n/a, p. 1900208, 2019.

[158] A. S. Christensen and O. A. von Lilienfeld, "Operator quantum machine learning: Navigating the chemical space of response properties," *CHIMIA*, vol. 73, no. 12, pp. 1028–1031, 2019.

[159] S. Fias, K. Y. S. Chang, and O. A. von Lilienfeld, "Alchemical normal modes unify chemical space," *J. Phys. Chem. Lett.*, vol. 10, no. 1, pp. 30–39, 2019.

[160] G. F. von Rudorff and O. A. von Lilienfeld, "Rapid and accurate molecular deprotonation energies from quantum alchemy," *Phys. Chem. Chem. Phys.*, pp. –, 2020.

[161] J. Westermayr, F. Faber, A. S. Christensen, A. von Lilienfeld, and P. Marquetand, "Neural networks and kernel ridge regression for excited states dynamics of $CH_2NH_2^+$: from single-state to multi-state representations and multi-property machine learning models," *Mach. Learn.: Sci. Technol.*, vol. in press, 2020.

[162] R. Zubatyuk, J. S. Smith, J. Leszczynski, and O. Isayev, "Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network," *Sci. Adv.*, vol. 5, no. 8, 2019.

[163] O. A. von Lilienfeld, "Quantum machine learning in chemical compound space," *Angew. Chem. Int. Edit.*, vol. 57, no. 16, pp. 4164–4169, 2018.

[164] J. W. Park, R. Al-Saadon, M. K. MacLeod, T. Shiozaki, and B. Vlaisavljevich, "Multireference Electron Correlation Methods: Journeys along Potential Energy Sur-

faces," *Chem. Rev.*, vol. 0, no. 0, p. null, 0.

[165] J. Westermayr, M. Gastegger, and P. Marquetand, "Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics," *J. Phys. Chem. Lett.*, vol. in press, no. 0, p. 0, 2020.

[166] I. Goodfellow, Y. Bengio, and A. Courville. MIT Press, 2016.

[167] A. V. Akimov, "A simple phase correction makes a big difference in nonadiabatic molecular dynamics," *J. Phys. Chem. Lett.*, vol. 9, no. 20, pp. 6096–6102, 2018.

[168] F. Plasser, M. Ruckenbauer, S. Mai, M. Oppel, P. Marquetand, and L. González, "Efficient and flexible computation of many-electron wave function overlaps," *J. Chem. Theory Comput.*, vol. 12, p. 1207, 2016.

[169] C. Robertson, J. González-Vázquez, I. corral, S. Díaz-Tendero, and C. Díaz, "Nonadiabatic scattering of NO off Au3 clusters: A simple and robust diabatic state manifold generation method for multiconfigurational wavefunctions," *J. Comput. Chem.*, vol. 40, no. 6, pp. 794–810, 2019.

[170] S. Mai, P. Marquetand, and L. González, "A general method to describe intersystem crossing dynamics in trajectory surface hopping," *Int. J. Quantum Chem.*, vol. 115, no. 18, pp. 1215–1231, 2015.

[171] Y. Shu, J. Kryven, A. G. Sampaio de Oliveira-Filho, L. Zhang, G.-L. Song, S. L. Li, R. Meana-Pañeda, B. Fu, J. M. Bowman, and D. G. Truhlar, "Direct diabatization and analytic representation of coupled potential energy surfaces and couplings for the reactive quenching of the excited $^2\sigma^+$ state of OH by molecular hydrogen," *J. Chem. Phys.*, vol. 151, no. 10, p. 104311, 2019.

[172] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, "Schnetpack: A deep learning toolbox for atomistic systems," *J. Chem. Theory Comput.*, vol. 15, no. 1, pp. 448–455, 2019.

[173] I. G. Ryabinkin, L. Joubert-Doriol, and A. F. Izmaylov, "Geometric phase effects in nonadiabatic dynamics near conical intersections," *Acc. Chem. Res.*, vol. 50, no. 7, pp. 1785–1793, 2017.

[174] J. Behler, R. Martoák, D. Donadio, and M. Parrinello, "Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential," *Phys. Rev. Lett.*, vol. 100, p. 185501, May 2008.

[175] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, and P. Marquetand, "wacsf – weighted atom-centered symmetry functions as descriptors in machine learning potentials," *J. Chem. Phys.*, vol. 148, no. 24, p. 241709, 2018.

[176] W. Koch and D. H. Zhang, "Communication: Separable potential energy surfaces from multiplicative artificial neural networks," *J. Chem. Phys.*, vol. 141, no. 2, p. 021101, 2014.

[177] D. He, J. Yuan, H. Li, and M. Chen, "Global Diabatic Potential Energy Surfaces and Quantum Dynamical Studies for the Li(2p) + $H_2(X^1\Sigma_g^+)$ → LiH($X^1\Sigma^+$) + H reaction," *Sci. Rep.*, vol. 6, no. 25083, 2016.

[178] Y. Guan, B. Fu, and D. H. Zhang, "Construction of diabatic energy surfaces for lifh with artificial neural networks," *J. Chem. Phys.*, vol. 147, no. 22, p. 224307, 2017.

[179] S. Wang, Z. Yang, J. Yuan, and M. Chen, "New Diabatic Potential Energy Surfaces of the NaH_2 System and Dynamics Studies for the Na(3p) + H_2 → NaH + H Reaction," *Sci. Rep.*, vol. 8, no. 17960, 2018.

[180] J. Yuan, D. He, S. Wang, M. Chen, and K. Han, "Diabatic potential energy surfaces of MgH_2^+ and dynamic studies for the Mg^+(3p) +H_2 → MgH^+ + H reaction," *Phys. Chem. Chem. Phys.*, vol. 20, pp. 6638–6647, 2018.

[181] Z. Yin, Y. Guan, B. Fu, and D. H. Zhang, "Two-state diabatic potential energy surfaces of ClH_2 based on nonadiabatic couplings with neural networks," *Phys. Chem. Chem. Phys.*, vol. 21, pp. 20372–20383, 2019.

[182] M. Schwilk, D. N. Tahchieva, and O. A. von Lilienfeld, "Large yet bounded: Spin gap ranges in carbenes," 2020.

[183] P. O. Dral, "Quantum chemistry in the age of machine learning," *J. Phys. Chem. Lett.*, vol. 11, no. 6, pp. 2336–2347, 2020.

[184] V. Botu and R. Ramprasad, "Adaptive machine learning framework to accelerate ab initio molecular dynamics," *Int. J. Quant. Chem.*, vol. 115, no. 16, pp. 1074–1083, 2015.

[185] J. Behler, "Constructing high-dimensional neural network potentials: A tutorial review," *Int. J. Quantum Chem.*, vol. 115, pp. 1032–1050, 2015.

[186] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, and A. E. Roitberg, "Less is more: Sampling chemical space with active learning," *J. Chem. Phys.*, vol. 148, no. 24, p. 241733, 2018.

[187] M. R. Raghunathan Ramakrishnan, Pavlo O. Dral and O. A. von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," *Sci. Data*, vol. 1, no. 140022, 2014.

[188] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Phys. Rev.*, vol. 40, pp. 749–750, 1932.

[189] R. E. Bruccoleri and M. Karplus, "conformational sampling using high-temperature molecular dynamics," *Biopolymers*, vol. 29, no. 14, pp. 1847–1862, 1990.

[190] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, and A. Shehu, "Principles and overview of sampling methods for modeling macromolecular structure and dynamics," *PLOS computational Biology*, vol. 12, pp. 1–70, 04 2016.

[191] J. Kästner, "Umbrella sampling," *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, vol. 1, no. 6, pp. 932–942, 2011.

[192] G. Tao, "Trajectory-guided sampling for molecular dynamics simulation," *Theor. Chem. Acc.*, vol. 138, p. 34, Feb 2019.

[193] Y. I. Yang, Q. Shao, J. Zhang, L. Yang, and Y. Q. Gao, "Enhanced sampling in molecular dynamics," *J. Chem. Phys.*, vol. 151, no. 7, p. 070902, 2019.

[194] J. E. Herr, K. Yao, R. McIntyre, D. W. Toth, and J. Parkhill, "Metadynamics for training neural network model chemistries: A competitive assessment," *J. Chem. Phys.*, vol. 148, no. 24, p. 241710, 2018.

[195] S. Grimme, "Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations," *J. Chem. Theory Comput.*, vol. 15, no. 5, pp. 2847–2862, 2019.

[196] M. Rupp, "Machine learning for quantum mechanics in a nutshell," *Int. J. Quantum Chem.*, vol. 115, no. 16, pp. 1058–1073, 2015.

[197] H. S. Seung, M. Opper, and H. Sompolinsky, "Query by committee," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT 92, (New York, NY, USA), p. 287294, Association for Computing Machinery, 1992.

[198] M. Gastegger and P. Marquetand, "Molecular dynamics with neural-network potentials," *arXiv:1812.07676 [physics.chem-ph]*, 2018.

[199] J. C. Tully, "Molecular dynamics with electronic transitions," *J. Chem. Phys.*, vol. 93, no. 2, pp. 1061–1071, 1990.

[200] J. C. Tully, "Nonadiabatic molecular dynamics," *Int. J. Quantum Chem.*, vol. 40, no. S25, pp. 299–309, 1991.

[201] P. O. Dral, A. Owens, S. N. Yurchenko, and W. Thiel,

"Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels," *J. Chem. Phys.*, vol. 146, no. 24, p. 244108, 2017.

[202] I. M. Sobol', D. Asotsky, A. Kreinin, and S. Kucherenko, "construction and comparison of high-dimensional sobol' generators," *Wilmott*, vol. 2011, no. 56, pp. 64–79, 2011.

[203] M. Ceriotti, G. A. Tribello, and M. Parrinello, "Demonstrating the transferability and the descriptive power of sketch-map," *J. Chem. Theory Comput.*, vol. 9, no. 3, pp. 1521–1532, 2013.

[204] J. Behler, "Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations," *Phys. Chem. Chem. Phys.*, vol. 13, pp. 17930–17955, 2011.

[205] Q. Lin, Y. Zhang, B. Zhao, and B. Jiang, "Automatically growing global reactive neural network potential energy surfaces: A trajectory-free active learning strategy," *J. Chem. Phys.*, vol. 152, no. 15, p. 154104, 2020.

[206] M. Barbatti, M. Ruckenbauer, and H. Lischka, "The photodynamics of ethylene: A surface-hopping study on structural aspects," *J. Chem. Phys.*, vol. 122, no. 17, p. 174307, 2005.

[207] D. Hollas, L. Šištík, E. G. Hohenstein, T. J. Martínez, and P. Slavíček, "Nonadiabatic ab initio molecular dynamics with the floating occupation molecular orbital-complete active space configuration interaction method," *J. Chem. Theory Comput.*, vol. 14, no. 1, pp. 339–350, 2018.

[208] H. Köppel, J. Gronki, and S. Mahapatra, "construction scheme for regularized diabatic states," *J. Chem. Phys.*, vol. 115, no. 6, pp. 2377–2388, 2001.

[209] J. M. Bowman, T. Carrington, and H. Meyer, "Variational quantum approaches for computing vibrational energies of polyatomic molecules," *Mol. Phys.*, vol. 106, no. 16, pp. 2145–2182, 2008.

[210] H.-D. Meyer, F. Gatti, and G. A. Worth, *The Road to MCTDH*, ch. 2, pp. 9–15. Wiley-VCH Verlag GmbH & co. KGaA, 2009.

[211] F. Liu, L. Du, D. Zhang, and J. Gao, "Direct learning hidden excited state interaction patterns from ab initio dynamics and its implication as alternative molecular mechanism models," *Sci. Rep.*, vol. 7, no. 8737, pp. 1–12, 2017.

[212] M. Baer, "Introduction to the theory of electronic non-adiabatic coupling terms in molecular systems," *Phys. Rep.*, vol. 358, no. 2, pp. 75–142, 2002.

[213] H. Lischka, M. Dallos, P. G. Szalay, D. R. Yarkony, and R. Shepard, "Analytic evaluation of nonadiabatic coupling terms at the MR-CI level. I. Formalism," *J. Chem. Phys.*, vol. 120, no. 16, pp. 7322–7329, 2004.

[214] J.-K. Ha, I. S. Lee, and S. K. Min, "Surface hopping dynamics beyond nonadiabatic couplings for quantum coherence," *J. Phys. Chem. Lett.*, vol. 9, no. 5, pp. 1097–1104, 2018.

[215] G. Granucci, M. Persico, and G. Spighi, "Surface hopping trajectory simulations with spin-orbit and dynamical couplings," *J. Chem. Phys.*, vol. 137, no. 22, p. 22A501, 2012.

[216] D. Tannor, *Introduction to Quantum Mechanics: A Time-Dependent Perspective*. Sausalito: University Science Books, 2006.

[217] A. W. Jasper, B. K. Kendrick, C. A. Mead, and D. G. Truhlar, *NON-BORN-OPPENHEIMER CHEMISTRY: POTENTIAL SURFACES, coUPLINGS, AND DYNAMICS*, pp. 329–391. World Scientific, 2004.

[218] D. R. Yarkony, "On the connectivity of seams of conical intersection: Seam curvature," *J. Chem. Phys.*, vol. 123, no. 20, p. 204101, 2005.

[219] X. Zhu and D. R. Yarkony, "Constructing diabatic representations using adiabatic and approximate diabatic data - coping with diabolical singularities," *J. Chem. Phys.*, vol. 144, no. 4, p. 044104, 2016.

[220] N. Wittenbrink, F. Venghaus, D. Williams, and W. Eisfeld, "A new approach for the development of diabatic potential energy surfaces: Hybrid block-diagonalization and diabatization by ansatz," *J. Chem. Phys.*, vol. 145, no. 18, p. 184108, 2016.

[221] W. Domcke, D. R. Yarkony, and H. Köppel, eds., *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy.* Singapore: World Scientific, 2004.

[222] G. W. Richings and G. A. Worth, "A practical diabatisation scheme for use with the direct-dynamics variational multi-configuration gaussian method," *J. Phys. Chem. A*, vol. 119, no. 50, pp. 12457–12470, 2015.

[223] D. Accomasso, M. Persico, and G. Granucci, "Diabatization by localization in the framework of configuration interaction based on floating occupation molecular orbitals (FOMO-CI)," *ChemPhotoChem*, vol. 3, no. 9, pp. 933–944, 2019.

[224] T. Lenzen and U. Manthe, "Neural network based coupled diabatic potential energy surfaces for reactive scattering," *J. Chem. Phys.*, vol. 147, no. 8, p. 084105, 2017.

[225] J. E. Subotnik, S. Yeganeh, R. J. Cave, and M. A. Ratner, "Constructing Diabatic States from Adiabatic States: Extending Generalized MullikenHush to Multiple Charge Centers with Boys localization," *J. Chem. Phys.*, vol. 129, no. 24, p. 244101, 2008.

[226] C. E. Hoyer, K. Parker, L. Gagliardi, and D. G. Truhlar, "The DQ and DQØ electronic structure diabatization methods: Validation for general applications," *J. Chem. Phys.*, vol. 144, no. 19, p. 194101, 2016.

[227] N. Wittenbrink, H. Ndome, and W. Eisfeld, "Toward spinorbit coupled diabatic potential energy surfaces for methyl iodide using effective relativistic coupling by asymptotic representation," *J. Phys. Chem. A*, vol. 117, no. 32, pp. 7408–7420, 2013.

[228] Z. Varga, K. A. Parker, and D. G. Truhlar, "Direct diabatization based on nonadiabatic couplings: the n/d method," *Phys. Chem. Chem. Phys.*, vol. 20, pp. 26643–26659, 2018.

[229] H. Nakamura and D. G. Truhlar, "Direct diabatization of electronic states by the fourfold way. ii. dynamical correlation and rearrangement processes," *J. Chem. Phys.*, vol. 117, no. 12, pp. 5576–5593, 2002.

[230] F. Venghaus and W. Eisfeld, "Block-diagonalization as a tool for the robust diabatization of high-dimensional potential energy surfaces," *J. Chem. Phys.*, vol. 144, no. 11, p. 114110, 2016.

[231] J. Li, B. Jiang, and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces. II. Four-atom systems," *J. Chem. Phys.*, vol. 139, no. 20, p. 204103, 2013.

[232] B. Jiang and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces," *J. Chem. Phys.*, vol. 139, no. 5, p. 054112, 2013.

[233] B. Jiang and H. Guo, "Permutation invariant polynomial neural network approach to fitting potential energy surfaces. III. molecule-surface interactions," *J. Chem. Phys.*, vol. 141, no. 3, p. 034109, 2014.

[234] B. Jiang, J. Li, and H. Guo, "Potential energy surfaces from high fidelity fitting of ab initio points: the permutation invariant polynomial - neural network approach," *Int. Rev. Phys. Chem.*, vol. 35, no. 3, pp. 479–506, 2016.

[235] J. Ischtwan and M. A. Collins, "Molecular potential energy surfaces by interpolation," *J. Chem. Phys.*, vol. 100, no. 11, pp. 8080–8088, 1994.

[236] C. R. Evenhuis and M. A. collins, "Interpolation of diabatic potential energy surfaces," *J. Chem. Phys.*, vol. 121, no. 6, pp. 2515–2527, 2004.

[237] C. Evenhuis and T. J. Martínez, "A scheme to interpolate potential energy surfaces and derivative coupling vectors without performing a global diabatization," *J. Chem. Phys.*, vol. 135, no. 22, p. 224110, 2011.

[238] S. Mukherjee, S. Bandyopadhyay, A. K. Paul, and S. Adhikari, "Construction of Diabatic Hamiltonian Matrix from ab Initio Calculated Molecular Symmetry Adapted Nonadiabatic Coupling Terms and Nuclear Dynamics for the Excited States of Na_3 Cluster," *J. Phys. Chem. A*, vol. 117, no. 16, pp. 3475–3495, 2013.

[239] X. Zhu and D. R. Yarkony, "Quasi-diabatic representations of adiabatic potential energy surfaces coupled by conical intersections including bond breaking: A more general construction procedure and an analysis of the diabatic representation," *J. Chem. Phys.*, vol. 137, no. 22, p. 22A511, 2012.

[240] X. Zhu and D. R. Yarkony, "Fitting coupled potential energy surfaces for large systems: Method and construction of a 3-state representation for phenol photodissociation in the full 33 internal degrees of freedom using multireference configuration interaction determined data," *J. Chem. Phys.*, vol. 140, no. 2, p. 024112, 2014.

[241] G. W. Richings, C. Robertson, and S. Habershon, "Improved on-the-fly MCTDH simulations with many-body-potential tensor decomposition and projection diabatization," *J. Chem. Theory Comput.*, vol. 15, no. 2, pp. 857–870, 2019.

[242] S. Gómez, M. Heindl, A. Szabadi, and L. González, "From surface hopping to quantum dynamics and back. finding essential electronic and nuclear degrees of freedom and optimal surface hopping parameters," *J. Phys. Chem. A*, vol. 123, no. 38, pp. 8321–8332, 2019.

[243] H. Köppel, W. Domcke, and L. S. Cederbaum, *in: conical Intersections (W. Domcke, D. R. Yarkony, H. Köppel, Eds.).* New York: World Scientific, 2004.

[244] D. R. Yarkony, "conical intersections: Their description and consequences," in *conical Intersections (W. Domcke, D. R. Yarkony, and H. Köppel, eds.)*, vol. 15 of *Advanced Series in Physical Chemistry*, World Scientific, 2004.

[245] F. Plasser, S. Gómez, M. F. S. J. Menger, S. Mai, and L. González, "Highly Efficient Surface Hopping Dynamics Using a Linear Vibronic Coupling Model," *Phys. Chem. Chem. Phys.*, vol. 21, pp. 57–69, 2019.

[246] E. Fabiano, T. Keal, and W. Thiel, "Implementation of surface hopping molecular dynamics using semiempirical methods," *Chem. Phys.*, vol. 349, no. 1-3, pp. 334 – 347, 2008.

[247] P. Oloyede, G. Milnikov, and H. Nakamura, "Generalized trajectory surface hopping method based on the zhu-nakamura theory," *J. Chem. Phys.*, vol. 124, no. 14, p. 144110, 2006.

[248] T. Ishida, S. Nanbu, and H. Nakamura, "Clarification of nonadiabatic chemical dynamics by the zhu-nakamura theory of nonadiabatic transition: from tri-atomic systems to reactions in solutions," *Int. Rev. Phys. Chem.*, vol. 36, no. 2, pp. 229–286, 2017.

[249] C. Zener, "Non-adiabatic crossing of energy levels," *Proc. Roy. Soc. Lond. A*, vol. 137, pp. 696–701, 1932.

[250] C. Wittig, "The Landau-Zener Formula," *J. Phys. Chem. B*, vol. 109, no. 17, pp. 8428–8430, 2005.

[251] L. Zhu, V. Kleiman, X. Li, S. P. Lu, K. Trentelman, and R. J. Gordon, "Ultrafast coherent control and destruction of excitons in quantum wells," *Phys. Rev. Lett.*, vol. 75, no. 13, pp. 2598–2601, 1995.

[252] C. Zhu, H. Kamisaka, and H. Nakamura, "New implementation of the trajectory surface hopping method with use of the ZhuNakamura theory. II. Application to the charge transfer processes in the 3D DH$_2$+ system," *J. Chem. Phys.*, vol. 116, no. 8, pp. 3234–3247, 2002.

[253] G. Granucci and M. Persico, "Critical appraisal of the fewest switching algorithm for surface hopping," *J. Chem. Phys.*, vol. 126, no. 13, p. 134114, 2007.

[254] J. P. Malhado, M. J. Bearpark, and J. T. Hynes, "Non-adiabatic dynamics close to conical intersections and the surface hopping perspective," *Front. Chem.*, vol. 2, no. 97, p. 97, 2014.

[255] L. Wang, A. Akimov, and O. V. Prezhdo, "Recent progress in surface hopping: 2011-2015," *J. Phys. Chem. Lett.*, vol. 7, no. 11, pp. 2100–2112, 2016.

[256] J. E. Subotnik, A. Jain, B. Landry, A. Petit, W. Ouyang, and N. Bellonzi, "Understanding the surface hopping view of electronic transitions and decoherence," *Annu. Rev. Phys. Chem.*, vol. 67, no. 1, pp. 387–417, 2016.

[257] J. Behler, S. Lorenz, and K. Reuter, "Representing molecule-surface interactions with symmetry-adapted neural networks," *J. Chem. Phys.*, vol. 127, no. 1, p. 014705, 2007.

[258] J. Behler, *Dissociation of Oxygen Molecules on the Al(111) Surface*. PhD thesis, Technical University Berlin, 2004.

[259] J. Behler, B. Delley, S. Lorenz, K. Reuter, and M. Scheffler, "Dissociation of o$_2$ at al(111): The role of spin selection rules," *Phys. Rev. Lett.*, vol. 94, p. 036104, Jan 2005.

[260] A. Thiel and H. Köppel, "Proposal and numerical test of a simple diabatization scheme," *J. Chem. Phys.*, vol. 110, no. 19, pp. 9371–9383, 1999.

[261] H. Köppel and B. Schubert, "The concept of regularized diabatic states for a general conical intersection," *Mol. Phys.*, vol. 104, no. 5-7, pp. 1069–1079, 2006.

[262] S. Maeda, K. Ohno, and K. Morokuma, "Updated branching plane for finding conical intersections without coupling derivative vectors," *J. Chem. Theory Comput.*, vol. 6, no. 5, pp. 1538–1545, 2010.

[263] J. A. Kammeraad and P. M. Zimmerman, "Estimating the derivative coupling vector using gradients," *J. Phys. Chem. Lett.*, vol. 7, no. 24, pp. 5074–5079, 2016.

[264] B. Gonon, A. Perveaux, F. Gatti, D. Lauvergnat, and B. Lasorne, "On the applicability of a wavefunction-free, energy-based procedure for generating first-order non-adiabatic couplings around conical intersections," *J. Chem. Phys.*, vol. 147, no. 11, p. 114114, 2017.

[265] A. Christensen, F. Faber, B. Huang, L. Bratholm, A. Tkatchenko, K. Müller, and O. Lilienfeld, "Qml: A python toolkit for quantum machine learning." https://github.com/qmlcode/qml, 2017.

[266] Y. Zhang, R. J. Maurer, and B. Jiang, "Symmetry-adapted high dimensional neural network representation of electronic friction tensor of adsorbates on metals," *J. Phys. Chem. C*, vol. 124, no. 1, pp. 186–195, 2020.

[267] H. Wang, L. Zhang, J. Han, and W. E, "DeePMD-kit: A deep learning package for many-body potential energy representation and molecular dynamics," *Comput. Phys. Commun.*, vol. 228, pp. 178 – 184, 2018.

[268] J. Behler and M. Parrinello, "Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces," *Phys. Rev. Lett.*, vol. 98, no. 14, p. 146401, 2007.

[269] J. Behler, "Atom-centered symmetry functions for constructing high-dimensional neural network potentials," *J. Chem. Phys.*, vol. 134, no. 7, p. 074106, 2011.

[270] J. E. Herr, K. Koh, K. Yao, and J. Parkhill, "compressing physics with an autoencoder: Creating an atomic species representation to improve machine learning models in the chemical sciences," *J. Chem. Phys.*, vol. 151, no. 8, p. 084103, 2019.

[271] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML17, p. 12631272, JMLR.org, 2017.

[272] O. T. Unke and M. Meuwly, "PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges," *J. Chem. Theory Comput.*, vol. 15, no. 6, pp. 3678–3693, 2019.

[273] N. Lubbers, J. S. Smith, and K. Barros, "Hierarchical modeling of molecular energies using a deep neural network," *J. Chem. Phys.*, vol. 148, no. 24, p. 241715, 2018.

BIBLIOGRAPHY

[1] S. MAI, L. GONZÁLEZ: Molecular photochemistry: Recent developments in theory, *Angew. Chem. Int. Ed.*, **n/a** (2020).

[2] R. L. LETICIA GONZÁLEZ: *Quantum Chemistry and Dynamics of Excited States : Methods and Applications*, John Wiley and Sons Ltd (2020).

[3] G. CERULLO, D. POLLI, G. LANZANI, S. DE SILVESTRI, H. HASHIMOTO, R. J. COGDELL: Photosynthetic light harvesting by carotenoids: Detection of an intermediate excited state, *Science*, **298**, 2395 (2002).

[4] Y.-C. CHENG, G. R. FLEMING: Dynamics of Light Harvesting in Photosynthesis, *Annu. Rev. Phys. Chem.*, **60**, 241 (2009).

[5] E. ROMERO, V. I. NOVODEREZHKIN, R. V. GRONDELLE: Quantum design of photosynthesis for bio-inspired solar-energy conversion, *Nature*, **543**, 355 (2017).

[6] N. J. HALAS, S. LAL, W.-S. CHANG, S. LINK, P. NORDLANDER: Plasmons in strongly coupled metallic nanostructures, *Chem. Rev.*, **111**, 3913 (2011).

[7] A. GE, B. RUDSHTEYN, J. ZHU, R. J. MAURER, V. S. BATISTA, T. LIAN: Electron–Hole-Pair-Induced Vibrational Energy Relaxation of Rhenium Catalysts on Gold Surfaces, *J. Phys. Chem. Lett.*, **9**, 406 (2018).

[8] Y. ZHANG, R. J. MAURER, H. GUO, B. JIANG: Hot-electron effects during reactive scattering of H_2 from Ag(111): The interplay between mode-specific electronic friction and the potential energy landscape, *Chem. Sci.*, **10**, 1089 (2019).

[9] S. YU, P. JAIN: Plasmonic photosynthesis of C1–C3 hydrocarbons from carbon dioxide assisted by an ionic liquid, *Nat. Commun.*, **10** (2019).

[10] K. SERGELEN, S. FOSSATI, A. TURUPCU, C. OOSTENBRINK, B. LIEDBERG, W. KNOLL, J. DOSTÁLEK: Plasmon Field-Enhanced Fluorescence Energy Transfer for Hairpin Aptamer Assay Readout, *ACS Sens.*, **2**, 916 (2017).

[11] T. SCHULTZ, E. SAMOYLOVA, W. RADLOFF, I. V. HERTEL, A. L. SOBOLEWSKI, W. DOMCKE: Efficient Deactivation of a Model Base Pair via Excited-State Hydrogen Transfer, *Science*, **306**, 1765 (2004).

[12] W. J. SCHREIER, T. E. SCHRADER, F. O. KOLLER, P. GILCH, C. E. CRESPO-HERNÁNDEZ, V. N. SWAMINATHAN, T. CARELL, W. ZINTH, B. KOHLER: Thymine Dimerization in DNA Is an Ultrafast Photoreaction, *Science*, **315**, 625 (2007).

[13] M. MALIŞ, Y. LOQUAIS, E. GLOAGUEN, H. S. BISWAL, F. PIUZZI, B. TARDIVEL, V. BRENNER, M. BROQUIER, C. JOUVET, M. MONS, ET AL.: Unraveling the mechanisms of nonradiative deactivation in model peptides following photoexcitation of a phenylalanine residue, *J. Am. Chem. Soc.*, **134**, 20340 (2012).

[14] C. RAUER, J. J. NOGUEIRA, P. MARQUETAND, L. GONZÁLEZ: Cyclobutane thymine photodimerization mechanism revealed by nonadiabatic molecular dynamics, *J. Am. Chem. Soc.*, **138**, 15911 (2016).

[15] J. HERBST, K. HEYNE, R. DILLER: Femtosecond infrared spectroscopy of bacteriorhodopsin chromophore isomerization, *Science*, **297**, 822 (2002).

[16] E. Tapavicza, I. Tavernelli, U. Rothlisberger: Trajectory Surface Hopping within Linear Response Time-Dependent Density-Functional Theory, *Phys. Rev. Lett.*, **98**, 023001 (2007).

[17] W. Domcke, A. L. Sobolewski: Peptide deactivation: Spectroscopy meets theory, *Nat. Chem.*, **5**, 257 (2013).

[18] D. I. Pattison, A. S. Rahmanto, M. J. Davies: Photo-oxidation of proteins, *Photochem. Photobiol. Sci.*, **11**, 38 (2012).

[19] M. Correia, M. T. Neves-Petersen, P. B. Jeppesen, S. Gregersen, S. B. Petersen: UV-Light Exposure of Insulin: Pharmaceutical Implications upon Covalent Insulin Dityrosine Dimerization and Disulphide Bond Photolysis, *PLoS One*, 7, 1 (2012).

[20] Maria Teresa Neves-Petersen and Søren Klitgaard and Torbjorn Pascher and Esben Skovsen and Tomas Polivka and Arkady Yartsev and Villly Sundström and Steffen B. Petersen: Flash Photolysis of Cutinase: Identification and Decay Kinetics of Transient Intermediates Formed upon UV Excitation of Aromatic Residues, *Biophys. J.*, **97**, 211 (2009).

[21] M. J. Davies, R. J. Truscott: Photo-Oxidation of Proteins and its Role in Cataractogenesis, *J. Photochem. Photobiol., B*, **63**, 114 (2001).

[22] Stefan Löfgren: Solar Ultraviolet Radiation Cataract, *Exp. Eye Res.*, **156**, 112 (2017).

[23] L. A. Baker, M. D. Horbury, S. E. Greenough, F. Allais, P. S. Walsh, S. Habershon, V. G. Stavros: Ultrafast Photoprotecting Sunscreens in Natural Plants, *J. Phys. Chem. Lett.*, **7**, 56 (2016).

[24] S. Heo, H. S. Hwang, Y. Jeong, K. Na: Skin Protection Efficacy from UV Irradiation and Skin Penetration Property of Polysaccharide-Benzophenone Conjugates as a Sunscreen Agent, *Carbohydr. Polym.*, **195**, 534 (2018).

[25] A. Tropsha: *Recent Trends in Quantitative Structure-Activity Relationships*, 49–76, American Cancer Society (2003).

[26] T. R. Nelson, A. J. White, J. A. Bjorgaard, A. E. Sifain, Y. Zhang, B. Nebgen, S. Fernandez-Alberti, D. Mozyrsky, A. E. Roitberg, S. Tretiak: Non-adiabatic excited-state molecular dynamics: Theory and applications for modeling photophysics in extended molecular materials, *Chem. Rev.*, **120**, 2215 (2020).

[27] I. Ahmad, S. Ahmed, Z. Anwar, M. A. Sheraz, M. Sikorski: Photostability and photostabilization of drugs and drug products, *Int. J. Photoenergy*, **2016**, 1 (2016).

[28] B. Sanchez-Lengeling, A. Aspuru-Guzik: Inverse molecular design using machine learning: Generative models for matter engineering, *Science*, **361**, 360 (2018).

[29] S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B. F. E. Curchod, N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M. K. Nazeeruddin, M. Grätzel: Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers, *Nat Chem* (2014).

[30] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, M. Ceriotti: Machine learning unifies the modeling of materials and molecules, *Sci. Adv.*, **3** (2017).

[31] E. Brunk, U. Rothlisberger: Mixed Quantum Mechanical/Molecular Mechanical Molecular Dynamics Simulations of Biological Systems in Ground and Electronically Excited States, *Chem. Rev.*, **115**, 6217 (2015).

[32] C. J. Wilson, A. S. Bommarius, J. A. Champion, Y. O. Chernoff, D. G. Lynn, A. K. Paravastu, C. Liang, M.-C. Hsieh, J. M. Heemstra: Biomolecular Assemblies: Moving from Observation to Predictive Design, *Chem. Rev.*, **118**, 11519 (2018).

[33] J. Berg, B. Häcker, A. Held, J. Tymoczko, L. Stryer, C. Lange, K. Mahlke, G. Maxam, L. Seidler, N. Zellerhoff: *Stryer Biochemie*, Spektrum Akademischer Verlag (2009).

[34] M. A. Ilie, C. Caruntu, D. Lixandru, M. Tampa, S.-R. Georgescu, M.-M. Constantin, C. Constantin, M. Neagu, S. A. Zurac, D. Boda1: In vivo confocal laser scanning microscopy imaging of skin inflammation: Clinical applications and research directions, *Exp. Ther. Med.*, **17**, 1004-1011 (2019).

[35] P. Selvin: The Renaissance of Fluorescence Resonance Energy Transfer, *Nat. Struct. Mol. Biol.*, **7**, 730–734 (2000).

[36] M. Meinhardt, R. Krebs, A. Anders, U. Heinrich, H. Tronnier: Absorption Spectra of Human Skin In Vivo in the Ultraviolet Wavelength Range Measured by Optoacoustics, *Photochem. Photobiol.*, **85**, 70 (2009).

[37] F. Duck: *Physical Properties of Tissue: A Comprehensive Reference Book*, Academic Press (1990).

[38] R. Navarro: The optical design of the human eye: a critical review, *J. Optom.*, **2**, 3 (2009).

[39] S. Ghosh, P. Verma, C. J. Cramer, L. Gagliardi, D. G. Truhlar: Combining wave function methods with density functional theory for excited states, *Chem. Rev.*, **118**, 7249 (2018).

[40] D. Bhattacharya, S. Basu, P. C. Mandal: UV Radiation Effects on Flavocytochrome B_2 in Dilute Aqueous Solution, *J. Photochem. Photobiol., B*, **59**, 54 (2000).

[41] S. P. Maria Teresa Neves-Petersen, G. P. Gajula: *UV Light Effects on Proteins: From Photochemistry to Nanomedicine, Molecular Photochemistry - Various Aspects*, chapter 7, IntechOpen (2012).

[42] S. Soorkia, C. Jouvet, G. Grégoire: Uv photoinduced dynamics of conformer-resolved aromatic peptides, *Chem. Rev.*, **120**, 3296–3327 (2020).

[43] R. Crespo-Otero, M. Barbatti: Recent advances and perspectives on nonadiabatic mixed quantum–classical dynamics, *Chem. Rev.*, **118**, 7026 (2018).

[44] S. Gozem, H. L. Luk, I. Schapiro, M. Olivucci: Theory and Simulation of the Ultrafast Double-Bond Isomerization of Biological Chromophores, *Chem. Rev.*, **117**, 13502 (2017).

[45] F. Plasser, M. Barbatti, A. Aquino, H. Lischka: Electronically Excited States and Photodynamics: A Continuing Challenge, *Theor. Chem. Acc.*, **131** (2012).

[46] J. Xie, M. Qin, Y. Cao, W. Wang: Mechanistic insight of photo-induced aggregation of chicken egg white lysozyme: The interplay between hydrophobic interactions and formation of intermolecular disulfide bonds, *Proteins: Struct., Funct., Bioinf.*, **79**, 2505 (2011).

[47] R. M. Abaskharon, F. Gai: Direct Measurement of the Tryptophan-Mediated Photocleavage Kinetics of a Protein Disulfide Bond, *Phys. Chem. Chem. Phys.*, **18**, 9602 (2016).

[48] D. C. Harris, M. D. Bertolucci: *Symmetry and Spectroscopy: an Introduction to Vibrational and Electronic Spectroscopy*, New York: Dover Publications (1989).

[49] C.-Y. Ng: *Vacuum Ultraviolet Photoionization and Photodissociation of Molecules and Clusters*, World Scientific (1991).

[50] A. H. Zewail: *Femtochemistry*, 3–22, World Scientific (1994).

[51] T. Brixner, T. Pfeifer, G. Gerber, M. Wollenhaupt, T. Baumert: Optimal control of atomic, molecular and electron dynamics with tailored femtosecond laser pulses, in P. Hannaford (editor), *Femtosecond Laser Spectroscopy*, 225–266, Springer-Verlag, New York (2005).

[52] A. Iqbal, V. G. Stavros: Active participation of $1\pi\sigma^*$ states in the photodissociation of tyrosine and its subunits, *J. Phys. Chem. Lett.*, **1**, 2274 (2010).

[53] M. Kowalewski, B. P. Fingerhut, K. E. Dorfman, K. Bennett, S. Mukamel: Simulating coherent multidimensional spectroscopy of nonadiabatic molecular processes: From the infrared to the x-ray regime, *Chem. Rev.*, **117**, 12165 (2017).

[54] M. Ruckenbauer, S. Mai, P. Marquetand, L. González: Revealing deactivation pathways hidden in time-resolved photoelectron spectra, *Sci. Rep.*, **6**, 35522 (2016).

[55] A. Tajti, G. Fogarasi, P. G. Szalay: Reinterpretation of the UV spectrum of cytosine: Only two electronic transitions?, *ChemPhysChem*, **10**, 1603 (2009).

[56] M. Barbatti, J. J. Szymczak, A. J. A. Aquino, D. Nachtigallová, H. Lischka: The decay mechanism of photoexcited guanine – a nonadiabatic dynamics study, *J. Chem. Phys.*, **134**, 014304 (2011).

[57] Y. Lu, Z. Lan, W. Thiel: Computational modeling of photoexcitation in DNA single and double strands, in *Photoinduced Phenomena in Nucleic Acids II*, volume 356 of *Topics in Current Chemistry*, 89–122, Springer Berlin Heidelberg (2014).

[58] M. Ruckenbauer, S. Mai, P. Marquetand, L. González: Photoelectron spectra of 2-thiouracil, 4-thiouracil, and 2,4-dithiouracil, *J. Chem. Phys.*, **144**, 074303 (2016).

[59] M. Manathunga, X. Yang, H. L. Luk, S. Gozem, L. M. Frutos, A. Valentini, N. Ferrè, M. Olivucci: Probing the photodynamics of rhodopsins with reduced retinal chromophores, *J. Chem. Theory Comput.*, **12**, 839 (2016).

[60] J. J. Nogueira, F. Plasser, L. González: Electronic Delocalization, Charge Transfer and Hypochromism in the UV Absorption Spectrum of Polyadenine Unravelled by Multiscale Computations and Quantitative Wavefunction Analysis, *Chem. Sci.*, **8**, 5682 (2017).

[61] S. Mai, A. Mohamadzade, P. Marquetand, L. González, S. Ullrich: Simulated and experimental time-resolved photoelectron spectra of the intersystem crossing dynamics in 2-thiouracil, *Molecules*, **23**, 2836 (2018).

[62] C. Rauer, J. J. Nogueira, P. Marquetand, L. González: Stepwise photosensitized thymine dimerization mediated by an exciton intermediate, *Monatsh. Chem.*, **149**, 1–9 (2018).

[63] J. P. Zobel, M. Heindl, J. J. Nogueira, L. González: Vibrational sampling and solvent effects on the electronic structure of the absorption spectrum of 2-nitronaphthalene, *J. Chem. Theory Comput.*, **14**, 3205 (2018).

[64] M. Barbatti, A. J. A. Aquino, H. Lischka: Ultrafast two-step process in the nonadiabatic relaxation of the CH_2NH_2 molecule, *Mol. Phys.*, **104**, 1053 (2006).

[65] B. F. E. Curchod, I. Tavernelli: On trajectory-based nonadiabatic dynamics: Bohmian dynamics versus trajectory surface hopping, *J. Chem. Phys.*, **138** (2013).

[66] A. V. Akimov, R. Asahi, R. Jinnouchi, O. V. Prezhdo: What makes the photocatalytic CO_2 reduction on n-doped ta2o5 efficient: Insights from nonadiabatic molecular dynamics, *J. Am. Chem. Soc.*, **137**, 11517 (2015).

[67] I. Schapiro, D. Roca-Sanjuán, R. Lindh, M. Olivucci: A surface hopping algorithm for nonadiabatic minimum energy path calculations, *J. Comput. Chem.*, **36**, 312 (2015).

[68] S. Mai, M. Richter, P. Marquetand, L. González: The dna nucleobase thymine in motion – intersystem crossing simulated with surface hopping, *Chem. Phys.*, **482**, 9 (2017).

[69] S. Mai, L. González: Unconventional two-step spin relaxation dynamics of [Re(co)3(im)(phen)]+ in aqueous solution, *Chem. Sci.*, **10**, 10405 (2019).

[70] S. L. Horton, Y. Liu, R. Forbes, V. Makhija, R. Lausten, A. Stolow, P. Hockett, P. Marquetand, T. Rozgonyi, T. Weinacht: Excited state dynamics of CH_2I_2 and CH_2BrI studied with UV pump VUV probe photoelectron spectroscopy, *J. Chem. Phys.*, **150**, 174201 (2019).

[71] P. Heim, S. Mai, B. Thaler, S. Cesnik, D. Avagliano, D. Bella-Velidou, W. E. Ernst, L. González, M. Koch: Revealing ultrafast population transfer between nearly degenerate electronic states, *J. Phys. Chem. Lett.*, **11**, 1443 (0).

[72] S. Mai, P. Marquetand, L. González: Nonadiabatic Dynamics: The SHARC Approach, *WIREs Comput. Mol. Sci.*, **8**, e1370 (2018).

[73] G. Groenhof, M. Bouxin-Cademartory, B. Hess, S. P. de Visser, H. J. C. Berendsen, M. Olivucci, A. E. Mark, M. A. Robb: Photoactivation of the Photoactive Yellow Protein: Why Photon Absorption Triggers a Trans-to-Cis Isomerization of the Chromophore in the Protein, *J. Am. Chem. Soc.*, **126**, 4228 (2004).

[74] T. J. Penfold, E. Gindensperger, C. Daniel, C. M. Marian: Spin-vibronic mechanism for intersystem crossing, *Chem. Rev.*, **118**, 6975 (2018).

[75] M. Ben-Nun, T. J. Martínez: Ab Initio *Quantum Molecular Dynamics*, 439–512, John Wiley & Sons, Ltd (2002).

[76] B. F. E. Curchod, F. Agostini, E. K. U. Gross: An exact factorization perspective on quantum interferences in nonadiabatic dynamics, *J. Chem. Phys.*, **145**, 034103 (2016).

[77] J. Westermayr, M. Gastegger, M. F. S. J. Menger, S. Mai, L. González, P. Marquetand: Machine Learning Enables Long Time Scale Molecular Photodynamics Simulations, *Chem. Sci.*, **10**, 8100 (2019).

[78] V. Yeguas, R. Casado: Big data issues in computational chemistry, in *2014 International Conference on Future Internet of Things and Cloud*, 389–392 (2014).

[79] R. Gómez-Bombarelli, A. Aspuru-Guzik: *Machine Learning and Big-Data in Computational Chemistry*, 1–24, Springer International Publishing, Cham (2018).

[80] L. Himanen, A. Geurts, A. S. Foster, P. Rinke: Data-Driven Materials Science: Status, Challenges, and Perspectives, *Adv. Sci.*, **6**, 1900808 (2019).

[81] J. Behler: First principles neural network potentials for reactive simulations of large molecular and condensed systems, *Angew. Chem. Int. Edit.*, **56**, 12828 (2017).

[82] I. Goodfellow, Y. Bengio, A. Courville: *Deep Learning*, MIT Press (2016).

[83] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller: *Quantum-Chemical Insights from Interpretable Atomistic Neural Networks*, 311–330, Springer International Publishing (2019).

[84] A. Agrawal, A. Choudhary: Perspective: Materials Informatics and Big Data: Realization of the "Fourth Paradigm" of Science in Materials Science, *APL Mat.*, **4**, 053208 (2016).

[85] J. Ischtwan, M. A. Collins: Molecular potential energy surfaces by interpolation, *J. Chem. Phys.*, **100**, 8080 (1994).

[86] R. P. A. Bettens, M. A. Collins: Learning to interpolate molecular potential energy surfaces with confidence: A bayesian approach, *J. Chem. Phys.*, **111**, 816 (1999).

[87] J. Westermayr, P. Marquetand: Machine learning and excited-state molecular dynamics, *Mach. Learn.: Sci. Technol.*, **in press**, doi:10.1088/2632 (2020).

[88] K. T. Schütt, S. Chmiela, O. A. von Lilienfeld, A. Tkatchenko, K. Tsuda, K.-R. Müller (editors): *Machine Learning Meets Quantum Physics*, Springer International Publishing (2020).

[89] O. Anatole von Lilienfeld, K.-R. Müller, A. Tkatchenko: Exploring Chemical compound Space with Quantum-Based Machine Learning, *Nat. Rev. Chem.* (2020).

[90] K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, R. J. Maurer: Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions, *Nat. Commun.*, **10**, 5024 (2019).

[91] M. Gastegger, A. McSloy, M. Luya, K. T. Schütt, R. J. Maurer: A Deep Neural Network for Molecular Wave Functions in Quasi-Atomic Minimal Basis Representation, *arXiv*, **2005.06979** (2020).

[92] G. Carleo, M. Troyer: Solving the quantum many-body problem with artificial neural networks, *Science*, **355**, 602 (2017).

[93] H. Saito: Solving the bose–hubbard model with machine learning, *J. Phys. Soc. Jpn.*, **86**, 093001 (2017).

[94] D. Pfau, J. S. Spencer, A. G. de G. Matthews, W. M. C. Foulkes: Ab-initio solution of the many-electron schrödinger equation with deep neural networks (2019).

[95] J. Hermann, Z. Schätzle, F. Noé: Deep neural network solution of the electronic schrödinger equation (2019).

[96] J. Han, L. Zhang, W. E: Solving Many-Electron Schrödinger Equation using Deep Neural Networks, *J. Comput. Phys.*, **399**, 108929 (2019).

[97] Y. Nomura, A. S. Darmawan, Y. Yamaji, M. Imada: Restricted boltzmann machine learning for solving strongly correlated quantum systems, *Phys. Rev. B*, **96**, 205152 (2017).

[98] S. Hobday, R. Smith, J. Belbruno: Applications of neural networks to fitting interatomic potential functions, *Modell. Simul. Mater. Sci. Eng.*, **7**, 397 (1999).

[99] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi: Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, *Phys. Rev. Lett.*, **104**, 136403 (2010).

[100] Z. Li, J. R. Kermode, A. De Vita: Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces, *Phys. Rev. Lett.*, **114**, 096405 (2015).

[101] M. Gastegger, P. Marquetand: High-dimensional neural network potentials for organic reactions and an improved training algorithm, *J. Chem. Theory Comput.*, **11**, 2187 (2015).

[102] J. Behler: Perspective: Machine learning potentials for atomistic simulations, *J. Chem. Phys.*, **145**, 170901 (2016).

[103] N. Artrith, A. Urban: An implementation of artificial neural-network potentials for atomistic materials simulations: Performance for tio$_2$, *computational Materials Science*, **114**, 135 (2016).

[104] M. Gastegger, C. Kauffmann, J. Behler, P. Marquetand: Comparing the accuracy of high-dimensional neural network potentials and the systematic molecular fragmentation method: A benchmark study for all-trans alkanes, *J. Chem. Phys.*, **144**, 194110 (2016).

[105] N. Artrith, A. Urban, G. Ceder: Efficient and accurate machine-learning interpolation of atomic energies in compositions with many species, *Phys. Rev. B*, **96**, 014112 (2017).

[106] M. Gastegger, J. Behler, P. Marquetand: Machine learning molecular dynamics for the simulation of infrared spectra, *Chem. Sci.*, **8**, 6924 (2017).

[107] V. L. Deringer, G. Csányi: Machine learning based interatomic potential for amorphous carbon, *Phys. Rev. B: Condens. Matter Mater. Phys.*, **95**, 094203 (2017).

[108] V. Botu, R. Batra, J. Chapman, R. Ramprasad: Machine learning force fields: Construction, validation, and outlook, *J. Phys. Chem. C*, **121**, 511 (2017).

[109] A. Glielmo, P. Sollich, A. De Vita: Accurate interatomic force fields via machine learning with covariant kernels, *Phys. Rev. B*, **95**, 214302 (2017).

[110] J. S. Smith, O. Isayev, A. E. Roitberg: ANI-1: An extensible neural network potential with dft accuracy at force field computational cost, *Chem. Sci.*, **8**, 3192 (2017).

[111] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, K.-R. Müller: SchNet – a deep learning architecture for molecules and materials, *J. Chem. Phys.*, **148**, 241722 (2018).

[112] S. Fujikake, V. L. Deringer, T. H. Lee, M. Krynski, S. R. Elliott, G. Csányi: Gaussian approximation potential modeling of lithium intercalation in carbon nanostructures, *J. Chem. Phys.*, **148**, 241714 (2018).

[113] H. Zong, G. Pilania, X. Ding, G. J. Ackland, T. Lookman: Developing an interatomic potential for martensitic phase transformations in zirconium by machine learning, *npj comput Mater*, **4** (2018).

[114] M. A. Wood, A. P. Thompson: Extending the accuracy of the snap interatomic potential form, *J. Chem. Phys.*, **148**, 241721 (2018).

[115] X. Chen, M. S. Jørgensen, J. Li, B. Hammer: Atomic energies from a convolutional neural network, *J. Chem. Theory Comput.*, **14**, 3933 (2018).

[116] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, K.-R. Müller: Schnetpack: A deep learning toolbox for atomistic systems, *J. Chem. Theory Comput.*, **15**, 448 (2019).

[117] A. P. Bartók, J. Kermode, N. Bernstein, G. Csányi: Machine learning a general-purpose interatomic potential for silicon, *Phys. Rev. X*, **8**, 041048 (2018).

[118] S. Chmiela, H. E. Sauceda, K.-R. Müller, A. Tkatchenko: Towards exact molecular dynamics simulations with machine-learned force fields, *Nat. Commun.*, **9**, 3887 (2018).

[119] G. Imbalzano, A. Anelli, D. Giofré, S. Klees, J. Behler, M. Ceriotti: Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials, *J. Chem. Phys.*, **148**, 241730 (2018).

[120] L. Zhang, J. Han, H. Wang, W. A. Saidi, R. Car, E. Weinan: End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems, in *Proceedings of the 32Nd International conference on Neural Information Processing Systems*, NIPS'18, 4441–4451, Curran Associates Inc., USA (2018).

[121] L. Zhang, J. Han, H. Wang, R. Car, W. E: Deep potential molecular dynamics: A scalable model with the accuracy of quantum mechanics, *Phys. Rev. Lett.*, **120**, 143001 (2018).

[122] H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Y. Chan, S. K. R. S. Sankaranarayanan: Machine learning classical interatomic potentials for molecular dynamics from first-principles training data, *J. Phys. Chem. C*, **123**, 6941 (2019).

[123] F. A. Faber, A. S. Christensen, B. Huang, O. A. von Lilienfeld: Alchemical and structural distribution based representation for universal quantum machine learning, *J. Chem. Phys.*, **148**, 241717 (2018).

[124] H. Wang, W. Yang: Toward building protein force fields by residue-based systematic molecular fragmentation and neural network, *J. Chem. Theory Comput.*, **15**, 1409 (2019).

[125] N. Gerrits, K. Shakouri, J. Behler, G.-J. Kroes: Accurate probabilities for highly activated reaction of polyatomic molecules on surfaces using a high-dimensional neural network potential: Chd_3 + cu(111), *J. Phys. Chem. Lett.*, **10**, 1763 (2019).

[126] S. Chmiela, H. E. Sauceda, I. Poltavsky, K.-R. Müller, A. Tkatchenko: sgdml: Constructing accurate and data efficient molecular force fields using machine learning, *Comput. Phys. Commun.*, **240**, 38 (2019).

[127] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, L. Zdeborová: Machine learning and the physical sciences, *Rev. Mod. Phys.*, **91**, 045002 (2019).

[128] R. V. Krems: Bayesian machine learning for quantum molecular dynamics, *Phys. Chem. Chem. Phys.*, **21**, 13392 (2019).

[129] V. L. Deringer, M. A. Caro, G. Csányi: Machine learning interatomic potentials as emerging tools for materials science, *Adv. Mat.*, **31**, 1902765 (2019).

[130] L. Ward, B. Blaiszik, I. Foster, R. S. Assary, B. Narayanan, L. Curtiss: Machine learning prediction of accurate atomization energies of organic molecules from low-fidelity quantum chemical calculations, *MRS Commun.*, **9**, 891–899 (2019).

[131] J. Weinreich, A. Römer, M. L. Paleico, J. Behler: Properties of α-brass nanoparticles. 1. neural network potential energy surface, *J. Phys. Chem. C*, **in press** (2020).

[132] A. S. Christensen, F. A. Faber, O. A. von Lilienfeld: Operators in quantum machine learning: Response properties in chemical space, *J. Chem. Phys.*, **150**, 064105 (2019).

[133] R. Zubatyuk, J. S. Smith, J. Leszczynski, O. Isayev: Accurate and transferable multi-task prediction of chemical properties with an atoms-in-molecules neural network, *Sci. Adv.*, **5** (2019).

[134] S. Wang, Z. Yang, J. Yuan, M. Chen: New Diabatic Potential Energy Surfaces of the NaH_2 System and Dynamics Studies for the Na(3p) + H_2 → NaH + H Reaction, *Sci. Rep.*, **8** (2018).

[135] J. Yuan, D. He, S. Wang, M. Chen, K. Han: Diabatic Potential Energy Surfaces of MgH_2^+ and Dynamic Studies for the Mg^+(3p) +H_2 → MgH^+ + H Reaction, *Phys. Chem. Chem. Phys.*, **20**, 6638 (2018).

[136] Z. Yin, Y. Guan, B. Fu, D. H. Zhang: Two-State Diabatic Potential Energy Surfaces of ClH$_2$ Based on Nonadiabatic Couplings with Neural Networks, *Phys. Chem. Chem. Phys.*, **21**, 20372 (2019).

[137] Y. Guan, B. Fu, D. H. Zhang: Construction of diabatic energy surfaces for lifh with artificial neural networks, *J. Chem. Phys.*, **147**, 224307 (2017).

[138] R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik: Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules, *ACS Cent. Sci.*, **4**, 268 (2018).

[139] J. G. Freeze, H. R. Kelly, V. S. Batista: Search for catalysts by inverse design: Artificial intelligence, mountain climbers, and alchemists, *Chemical Reviews*, **119**, 6595 (2019).

[140] G. F. von Rudorff, O. A. von Lilienfeld: Rapid and accurate molecular deprotonation energies from quantum alchemy, *Phys. Chem. Chem. Phys.*, – (2020).

[141] D. C. Elton, Z. Boukouvalas, M. D. Fuge, P. W. Chung: Deep Learning for Molecular Design – A Review of the State of the Art, *Mol. Syst. Des. Eng.*, **4**, 828 (2019).

[142] C. Evenhuis, T. J. Martínez: A scheme to interpolate potential energy surfaces and derivative coupling vectors without performing a global diabatization, *J. Chem. Phys.*, **135**, 224110 (2011).

[143] X. Zhu, D. R. Yarkony: Fitting Coupled Potential Energy Surfaces for Large Systems: Method and Construction of a 3-State Representation for Phenol Photodissociation in the Full 33 Internal Degrees of Freedom using Multireference Configuration Interaction Determined Data, *J. Chem. Phys.*, **140**, 024112 (2014).

[144] C. Carbogno, J. Behler, A. Gross, K. Reuter: Fingerprints for Spin-Selection Rules in the Interaction Dynamics of O$_2$ at Al(111), *Phys. Rev. Lett.*, **101**, 096104 (2008).

[145] C. Carbogno, J. Behler, K. Reuter, A. Gross: Signatures of nonadiabatic O$_2$ dissociation at Al(111): First-principles fewest-switches study, *Phys. Rev. B*, **81**, 035410 (2010).

[146] R. Ramakrishnan, M. Hartmann, E. Tapavicza, O. A. von Lilienfeld: Electronic spectra from TDDFT and machine learning in chemical space, *J. Chem. Phys.*, **143**, 084111 (2015).

[147] F. Häse, S. Valleau, E. Pyzer-Knapp, A. Aspuru-Guzik: Machine learning exciton dynamics, *Chem. Sci.*, **7**, 5139 (2016).

[148] Häse, Florian and Kreisbeck, Christoph and Aspuru-Guzik, Alán: Machine learning for quantum dynamics: Deep learning of excitation energy transfer properties, *Chem. Sci.*, **8**, 8419 (2017).

[149] Y. Shen, D. R. Yarkony: Construction of quasi-diabatic hamiltonians that accurately represent ab initio determined adiabatic electronic states coupled by conical intersections for systems on the order of 15 atoms. application to cyclopentoxide photoelectron detachment in the full 39 degrees of freedom, *J. Phys. Chem. A*, **124**, 4539 (2020).

[150] K. Ghosh, A. Stuke, M. Todorović, P. B. Jørgensen, M. N. Schmidt, A. Vehtari, P. Rinke: Deep learning spectroscopy: Neural networks for molecular excitation spectra, *Adv. Sci.*, **6**, 1801367 (2019).

[151] S. Ye, W. Hu, X. Li, J. Zhang, K. Zhong, G. Zhang, Y. Luo, S. Mukamel, J. Jiang: A neural network protocol for electronic excitations of n-methylacetamide, *Proc. Natl. Acad. Sci.*, **116**, 11612 (2019).

[152] Y. Zhang, S. Ye, J. Zhang, J. Jiang, B. Jiang: Efficient and accurate spectroscopic simulations with symmetry-preserving neural network models for tensorial properties, *arXiv*, **2004.13605** (2020).

[153] J. Behler: Constructing high-dimensional neural network potentials: A tutorial review, *Int. J. Quantum Chem.*, **115**, 1032 (2015).

[154] D. Hu, Y. Xie, X. Li, L. Li, Z. Lan: Inclusion of machine learning kernel ridge regression potential energy surfaces in on-the-fly nonadiabatic molecular dynamics simulation, *J. Phys. Chem. Lett.*, **9**, 2725 (2018).

[155] P. O. Dral, M. Barbatti, W. Thiel: Nonadiabatic excited-state dynamics with machine learning, *J. Phys. Chem. Lett.*, **9**, 5660 (2018).

[156] W.-K. Chen, X.-Y. Liu, W.-H. Fang, P. O. Dral, G. Cui: Deep learning for nonadiabatic excited-state dynamics, *J. Phys. Chem. Lett.*, **9**, 6702 (2018).

[157] J. Westermayr, M. Gastegger, P. Marquetand: Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics, *J. Phys. Chem. Lett.*, **11**, 3828 (2020).

[158] J. Westermayr, F. A. Faber, A. S. Christensen, O. A. von Lilienfeld, P. Marquetand: Neural networks and kernel ridge regression for excited states dynamics of $CH_2NH_2^+$: From single-state to multi-state representations and multi-property machine learning models, *Mach. Learn.: Sci. Technol.*, **1**, 025009 (2020).

[159] G. W. Richings, S. Habershon: Direct grid-based quantum dynamics on propagated diabatic potential energy surfaces, *Chem. Phys. Lett.*, **683**, 228 (2017).

[160] G. W. Richings, S. Habershon: MCTDH on-the-fly: Efficient grid-based quantum dynamics without pre-computed potential energy surfaces, *J. Chem. Phys.*, **148**, 134116 (2018).

[161] G. W. Richings, C. Robertson, S. Habershon: Can we use on-the-fly quantum simulations to connect molecular structure and sunscreen action?, *Faraday Discuss.*, **216**, 476 (2019).

[162] Y. Guan, H. Guo, D. R. Yarkony: Neural network based quasi-diabatic Hamiltonians with symmetry adaptation and a correct description of conical intersections, *J. Chem. Phys.*, **150**, 214101 (2019).

[163] C. Xie, X. Zhu, D. R. Yarkony, H. Guo: Permutation invariant polynomial neural network approach to fitting potential energy surfaces. IV. coupled diabatic potential energy matrices, *J. Chem. Phys.*, **149**, 144107 (2018).

[164] H. Lischka, D. Nachtigallová, A. J. A. Aquino, P. G. Szalay, F. Plasser, F. B. C. Machado, M. Barbatti: Multireference approaches for excited states of molecules, *Chem. Rev.*, **118**, 7293 (2018).

[165] D. R. Yarkony: conical intersections: Their description and consequences, in W. Domcke, D. R. Yarkony, H. Köppel (editors), *conical Intersections*, volume 15 of *Advanced Series in Physical Chemistry*, World Scientific (2004).

[166] S. Mai, M. Richter, M. Ruckenbauer, M. Oppel, P. Marquetand, L. González: SHARC2.0: Surface Hopping Including ARbitrary Couplings – Program Package for Non-Adiabatic Dynamics, sharc-md.org (2018).

[167] M. Born, R. Oppenheimer: Zur quantentheorie der molekeln, *Ann. Phys.*, **389**, 457 (1927).

[168] N. L. DOLTSINIS: *Molecular Dynamics Beyond the Born-Oppenheimer Approximation: Mixed Quantum-Classical Approaches*, volume 31 of *NIC Series*, John von Neuman Institut for computing (2006).

[169] A. SZABO, N. OSTLUND: *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Books on Chemistry, Dover Publications (2012).

[170] W. KOHN: Nobel Lecture: Electronic Structure of Matter – Wave Functions and Density Functionals, *Rev. Mod. Phys.*, **71**, 1253 (1999).

[171] A. J. COHEN, P. MORI-SÁNCHEZ, W. YANG: Challenges for Density Functional Theory, *Chem. Rev.*, **112**, 289 (2012).

[172] H. S. YU, S. L. LI, D. G. TRUHLAR: Perspective: Kohn-Sham density functional theory descending a staircase, *J. Chem. Phys.*, **145**, 130901 (2016).

[173] M. R. RAGHUNATHAN RAMAKRISHNAN, PAVLO O. DRAL, O. A. VON LILIENFELD: Quantum chemistry structures and properties of 134 kilo molecules, *Sci. Data*, **1** (2014).

[174] F. BROCKHERDE, L. VOGT, L. LI, M. E. TUCKERMAN, K. BURKE, K.-R. MÜLLER: Bypassing the kohn-sham equations with machine learning, *Nat. Commun.*, **8**, 872 (2017).

[175] R. J. MAURER, C. FREYSOLDT, A. M. REILLY, J. G. BRANDENBURG, O. T. HOFMANN, T. BJÖRKMAN, S. LEBÈGUE, A. TKATCHENKO: Advances in density-functional calculations for materials modeling, *Annual Review of Materials Research*, **49**, 1 (2019).

[176] P. HOHENBERG, W. KOHN: Inhomogeneous electron gas, *Phys. Rev.*, **136**, B864 (1964).

[177] W. KOHN, L. J. SHAM: Self-consistent equations including exchange and correlation effects, *Phys. Rev.*, **140**, A1133 (1965).

[178] S. HIRATA, M. HEAD-GORDON: Time-dependent density functional theory within the tamm−dancoff approximation, *Chem. Phys. Lett.*, **314**, 291 (1999).

[179] A. D. BECKE: Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.*, **98**, 5648 (1993).

[180] A. D. BECKE: Density-functional thermochemistry. V. Systematic optimization of exchange-correlation functionals, *J. Chem. Phys.*, **107**, 8554 (1997).

[181] R. D. ADAMSON, P. M. GILL, J. A. POPLE: Empirical Density Functionals, *Chem, Phys. Lett.*, **284**, 6 (1998).

[182] J. M. L. MARTIN, G. SANTRA: Empirical Double-Hybrid Density Functional Theory: A 'Third Way' in Between WFT and DFT, *Isr. J. Chem.*, **doi:10.1002/ijch.201900114** (2019).

[183] J. P. PERDEW, K. BURKE, M. ERNZERHOF: Generalized gradient approximation made simple, *Phys. Rev. Lett.*, **77**, 3865 (1996).

[184] J. TAO, J. P. PERDEW, V. N. STAROVEROV, G. E. SCUSERIA: Climbing the density functional ladder: Nonempirical meta−generalized gradient approximation designed for molecules and solids, *Phys. Rev. Lett.*, **91**, 146401 (2003).

[185] C. ADAMO, V. BARONE: Toward reliable density functional methods without adjustable parameters: The pbe0 model, *J. Chem. Phys.*, **110**, 6158 (1999).

[186] J. P. PERDEW, M. ERNZERHOF, K. BURKE: Rationale for mixing exact exchange with density functional approximations, *J. Chem. Phys.*, **105**, 9982 (1996).

[187] J.-D. CHAI, M. HEAD-GORDON: Systematic optimization of long-range corrected hybrid density functionals, *J. Chem. Phys.*, **128**, 084106 (2008).

[188] J. Sun, R. Haunschild, B. Xiao, I. W. Bulik, G. E. Scuseria, J. P. Perdew: Semilocal and hybrid meta-generalized gradient approximations based on the understanding of the kinetic-energy-density dependence, *J. Chem. Phys.*, **138**, 044113 (2013).

[189] F. Cordova, L. J. Doriol, A. Ipatov, M. E. Casida, C. Filippi, A. Vela: Troubleshooting time-dependent density-functional theory for photochemical applications: Oxirane, *J. Chem. Phys.*, **127**, 164111 (2007).

[190] C. L. Benavides-Riveros, N. N. Lathiotakis, M. A. L. Marques: Towards a formal definition of static and dynamic electronic correlations, *Phys. Chem. Chem. Phys.*, **19**, 12655 (2017).

[191] A. J. Cohen, P. Mori-Sánchez, W. Yang: Insights into Current Limitations of Density Functional Theory, *Science*, **321**, 792 (2008).

[192] N. Mardirossian, M. Head-Gordon: Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals, *Mol. Phys.*, **115**, 2315 (2017).

[193] E. Schrödinger: An Undulatory Theory of the Mechanics of Atoms and Molecules, *Phys. Rev.*, **28**, 1049 (1926).

[194] G. Granucci, M. Persico: Critical appraisal of the fewest switching algorithm for surface hopping, *J. Chem. Phys.*, **126**, 134114 (2007).

[195] J. C. Slater: The theory of complex spectra, *Phys. Rev.*, **34**, 1293 (1929).

[196] V. Fock: Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, *Z. Phys.*, **61**, 126 (1930).

[197] D. R. Hartree: The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods, *Math. Proc. Cambridge,*, **24**, 89–110 (1928).

[198] S. Mai: *Excited-State Dynamics of Nucleobase Analogues*, Ph.D. thesis, University of Vienna (2016).

[199] B. O. Roos, P. R. Taylor, P. E. Siegbahn: A complete active space scf method (casscf) using a density matrix formulated super-ci approach, *Chem. Phys.*, **48**, 157 (1980).

[200] B. O. Roos, P. E. M. Siegbahn: A Direct CI Method with a Multiconfigurational Reference State, *Int. J. Quantum Chem.*, **17**, 485 (1980).

[201] Lischka, Hans and Shepard, Ron and Pitzer, Russell M. and Shavitt, Isaiah and Dallos, Michal and Müller, Thomas and Szalay, Péter G. and Seth, Michael and Kedziora, Gary S. and Yabushita, Satoshi and Zhang, Zhiyong: High-Level Multireference Methods in the Quantum-Chemistry Program System COLUMBUS: Analytic MR-CISD and MR-AQCC Gradients and MR-AQCC-LRT for Excited States, GUGA Spin–Orbit CI and Parallel CI Density, *Phys. Chem. Chem. Phys.*, **3**, 664 (2001).

[202] C. Möller, M. S. Plesset: Note on an Approximation Treatment for Many-Electron Systems, *Phys. Rev.*, **46**, 618 (1934).

[203] R. J. Bartlett: Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules, *Annu. Rev. Phys. Chem.*, **32**, 359 (1981).

[204] K. Andersson, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, K. Wolinski: Second-Order Perturbation Theory with a CASSCF Reference Function, *J. Phys. Chem.*, **94**, 5483 (1990).

[205] K. Andersson, P. Malmqvist, B. O. Roos: Second-order perturbation theory with a complete active space self-consistent field reference function, *J. Phys. Chem.*, **96**, 1218 (1992).

[206] J. Finley, P.-A. Malmqvist, B. O. Roos, L. Serrano-Andrés: The multi-state CASPT2 method, *Chem. Phys. Lett.*, **288**, 299 (1998).

[207] D. Roca-Sanjuán, F. Aquilante, R. Lindh: Multiconfiguration second-order perturbation theory approach to strong electron correlation in chemistry and photochemistry, *WIREs Comput. Mol. Sci.*, **2**, 585 (2012).

[208] Y. Guan, H. Guo, D. R. Yarkony: Extending the representation of multistate coupled potential energy surfaces to include properties operators using neural networks: Application to the $1,2^1a$ states of ammonia, *J. Chem. Theory Comput.*, **16**, 302 (2020).

[209] A. Dreuw, M. Wormit: The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states, *WIREs Comput. Mol. Sci.*, **5**, 82 (2015).

[210] A. Warshel, M. Karplus: Semiclassical Trajectory Approach to Photoisomerization, *Chem. Phys. Lett.*, **32**, 11 (1975).

[211] J. Behler: Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations, *Phys. Chem. Chem. Phys.*, **13**, 17930 (2011).

[212] F. Noé, A. Tkatchenko, K.-R. Müller, C. Clementi: Machine learning for molecular simulation, *Annu. Rev. Phys. Chem.*, **71**, 361 (2020).

[213] M. Richter, P. Marquetand, J. González-Vázquez, I. Sola, L. González: SHARC: Ab initio molecular dynamics with surface hopping in the adiabatic representation including arbitrary couplings, *J. Chem. Theory Comput.*, **7**, 1253 (2011).

[214] S. Mai, P. Marquetand, L. González: A general method to describe intersystem crossing dynamics in trajectory surface hopping, *Int. J. Quantum Chem.*, **115**, 1215 (2015).

[215] M. Barbatti: Nonadiabatic dynamics with trajectory surface hopping method, *WIREs Comput. Mol. Sci.*, **1**, 620 (2011).

[216] I. Newton: *Philosophiae Naturalis Principia Mathematica*, J. Societatis Regiae ac Typis J. Streater (1687).

[217] M. Persico, G. Granucci: An overview of nonadiabatic dynamics simulations methods, with focus on the direct approach versus the fitting of potential energy surfaces, *Theor. Chem. Acc.*, **133**, 1526 (2014).

[218] L. Verlet: Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules, *Phys. Rev.*, **159**, 98 (1967).

[219] J. E. Subotnik, A. Jain, B. Landry, A. Petit, W. Ouyang, N. Bellonzi: Understanding the surface hopping view of electronic transitions and decoherence, *Annu. Rev. Phys. Chem.*, **67**, 387 (2016).

[220] J. C. Tully, R. K. Preston: Trajectory Surface Hopping Approach to Nonadiabatic Molecular Collisions: The Reaction of H^+ with D_2, *J. Chem. Phys.*, **55**, 562 (1971).

[221] J. C. Tully: Molecular dynamics with electronic transitions, *J. Chem. Phys.*, **93**, 1061 (1990).

[222] J. C. Tully: Nonadiabatic molecular dynamics, *Int. J. Quantum Chem.*, **40**, 299 (1991).

[223] M. Baer: Introduction to the theory of electronic non-adiabatic coupling terms in molecular systems, *Phys. Rep.*, **358**, 75 (2002).

[224] H. Lischka, M. Dallos, P. G. Szalay, D. R. Yarkony, R. Shepard: Analytic evaluation of nonadiabatic coupling terms at the MR-CI level. I. Formalism, *J. Chem. Phys.*, **120**, 7322 (2004).

[225] L. Wang, A. Akimov, O. V. Prezhdo: Recent progress in surface hopping: 2011-2015, *J. Phys. Chem. Lett.*, **7**, 2100 (2016).

[226] J. P. Malhado, M. J. Bearpark, J. T. Hynes: Non-adiabatic dynamics close to conical intersections and the surface hopping perspective, *Front. Chem.*, **2**, 97 (2014).

[227] C. Zhu, H. Kamisaka, H. Nakamura: New Implementation of the Trajectory Surface Hopping Method with Use of the Zhu–Nakamura Theory. II. Application to the Charge Transfer Processes in the 3D DH$_2$+ System, *J. Chem. Phys.*, **116**, 3234 (2002).

[228] L. Zhu, V. Kleiman, X. Li, S. P. Lu, K. Trentelman, R. J. Gordon: Ultrafast coherent control and destruction of excitons in quantum wells, *Phys. Rev. Lett.*, **75**, 2598 (1995).

[229] P. Oloyede, G. Mil'nikov, H. Nakamura: Generalized trajectory surface hopping method based on the zhu-nakamura theory, *J. Chem. Phys.*, **124**, 144110 (2006).

[230] G. Granucci, M. Persico, G. Spighi: Surface hopping trajectory simulations with spin-orbit and dynamical couplings, *J. Chem. Phys.*, **137**, 22A501 (2012).

[231] C. Zener: Non-adiabatic crossing of energy levels, *Proc. Roy. Soc. Lond. A*, **137**, 696 (1932).

[232] C. Wittig: The Landau-Zener Formula, *J. Phys. Chem. B*, **109**, 8428 (2005).

[233] T. Ishida, S. Nanbu, H. Nakamura: Clarification of nonadiabatic chemical dynamics by the zhu-nakamura theory of nonadiabatic transition: From tri-atomic systems to reactions in solutions, *Int. Rev. Phys. Chem.*, **36**, 229 (2017).

[234] K. Takatsuka, T. Yonehara, K. Hanasaki, Y. Arasaki: *Breakdown of the Born–Oppenheimer Approximation: Classic Theories of Nonadiabatic Transitions and Ideas behind*, chapter Chapter 4, 59–95, WORLD SCIENTIFIC (2015).

[235] E. Wigner, E. Guth: Gruppentheorie und ihre anwendung auf die quantenmechanik der atomspektren, *Monatsh. Math.*, **39**, A51 (1932).

[236] E. Wigner: On the quantum correction for thermodynamic equilibrium, *Phys. Rev.*, **40**, 749 (1932).

[237] G. A. Worth, L. S. Cederbaum: Beyond born-oppenheimer: Molecular dynamics through a conical intersection, *Annu. Rev. Phys. Chem.*, **55**, 127 (2004).

[238] A. W. Jasper, B. K. Kendrick, C. A. Mead, D. G. Truhlar: *Non-Born-Oppenheimer Chemistry: Potential Surfaces, Couplings, and Dynamics*, 329–391, World Scientific (2004).

[239] H. Köppel, W. Domcke, L. S. Cederbaum: *in: Conical Intersections (W. Domcke, D. R. Yarkony, H. Köppel, Eds.)*, World Scientific, New York (2004).

[240] D. R. Yarkony: On the connectivity of seams of conical intersection: Seam curvature, *J. Chem. Phys.*, **123**, 204101 (2005).

[241] F. Plasser, S. Gómez, M. F. S. J. Menger, S. Mai, L. González: Highly efficient surface hopping dynamics using a linear vibronic coupling model, *Phys. Chem. Chem. Phys.*, **21**, 57 (2019).

[242] K. F. K. G. Dyall: *Introduction to Relativistic Quantum Chemistry*, Oxford University Press (2007).

[243] A. W. M. Reiher: *Relativistic Quantum Chemistry*, Wiley VCH Verlag Weinheim (2009).

[244] C. M. Marian: Spin–Orbit Coupling and Intersystem Crossing in Molecules, *WIREs Comput. Mol. Sci.*, **2**, 187 (2012).

[245] E. E. S. H. A. Bethe: *Quantum Mechanics of One- and Two-Electron Atoms*, Springer, Berlin (1957).

[246] P. A. M. Dirac, R. H. Fowler: The quantum theory of the electron, *Proc. R. Soc. London, Ser. A*, **117**, 610 (1928).

[247] P. Pyykko: Relativistic Effects in Structural Chemistry, *Chem. Rev.*, **88**, 563 (1988).

[248] F. Neese, T. Petrenko, D. Ganyushin, G. Olbrich: Advanced Aspects of *Ab Initio* Theoretical Optical Spectroscopy of Transition Metal complexes: Multiplets, spin-orbit coupling and resonance Raman intensities, *Coord. Chem. Rev.*, **251**, 288 (2007).

[249] F. Neese: Efficient and accurate approximations to the molecular spin-orbit coupling operator and their use in molecular g-tensor calculations, *J. Chem. Phys.*, **122**, 034107 (2005).

[250] B. A. Hess, C. M. Marian, U. Wahlgren, O. Gropen: A mean-field spin-orbit method applicable to correlated wavefunctions, *Chem. Phys. Lett.*, **251**, 365 (1996).

[251] D. R. Yarkony: On the Consequences of Nonremovable Derivative Couplings. I. The Geometric Phase and Quasidiabatic States: A Numerical Study, *J. Chem. Phys.*, **105**, 10456 (1996).

[252] D. R. Yarkony: On the Role of Conical Intersections in Photodissociation. V. Conical Intersections and the Geometric Phase in the Photodissociation of Methyl Mercaptan, *J. Chem. Phys.*, **104**, 7866 (1996).

[253] I. G. Ryabinkin, L. Joubert-Doriol, A. F. Izmaylov: Geometric phase effects in nonadiabatic dynamics near conical intersections, *Acc. Chem. Res.*, **50**, 1785 (2017).

[254] I. G. Ryabinkin, L. Joubert-Doriol, A. F. Izmaylov: When do we need to account for the geometric phase in excited state dynamics?, *J. Chem. Phys.*, **140**, 214116 (2014).

[255] R. Gherib, I. G. Ryabinkin, A. F. Izmaylov: Why Do Mixed Quantum-Classical Methods Describe Short-Time Dynamics through Conical Intersections So Well? Analysis of Geometric Phase Effects, *J. Chem. Theory Comput.*, **11**, 1375 (2015).

[256] C. Zhu, H. Kamisaka, H. Nakamura: Significant improvement of the trajectory surface hopping method by the zhu–nakamura theory, *J. Chem. Phys*, **115**, 11036 (2001).

[257] A. V. Akimov: A simple phase correction makes a big difference in nonadiabatic molecular dynamics, *J. Phys. Chem. Lett.*, **9**, 6096 (2018).

[258] F. Plasser, M. Ruckenbauer, S. Mai, M. Oppel, P. Marquetand, L. González: Efficient and flexible computation of many-electron wave function overlaps, *J. Chem. Theory Comput.*, **12**, 1207 (2016).

[259] F. Plasser, L. González: communication: Unambiguous comparison of many-electron wavefunctions through their overlaps, *J. Chem. Phys.*, **145**, 021103 (2016).

[260] Z. Varga, K. A. Parker, D. G. Truhlar: Direct diabatization based on nonadiabatic couplings: The n/d method, *Phys. Chem. Chem. Phys.*, **20**, 26643 (2018).

[261] J. E. Subotnik, S. Yeganeh, R. J. Cave, M. A. Ratner: Constructing Diabatic States from Adiabatic States: Extending Generalized Mulliken–Hush to Multiple Charge Centers with Boys Localization, *J. Chem. Phys.*, **129**, 244101 (2008).

[262] N. Nilsson: *The Quest for Artificial Intelligence*, Cambridge University Press (2009).

[263] W. S. McCulloch, W. Pitts: A logical calculus of the ideas immanent in nervous activity, *B. Math. Biophys.*, **5**, 115 (1943).

[264] D. Shen, G. Wu, H.-I. Suk: Deep learning in medical image analysis, *Annu. Rev. Biomed. Eng.*, **19**, 221 (2017).

[265] V. Këpuska, G. Bohouta: Next-Generation of Virtual Personal Assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home), in *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 99–103 (2018).

[266] M. B. Hoy: Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants, *Med. Ref. Serv. Q.*, **37**, 81 (2018).

[267] J. Gasteiger, J. Zupan: Neural networks in chemistry, *Angew. Chem., Int. Ed. Engl.*, **32**, 503 (1993).

[268] J. Townsend, K. D. Vogiatzis: Data-driven acceleration of the coupled-cluster singles and doubles iterative solver, *J. Phys. Chem. Lett.*, **10**, 4129 (2019).

[269] K. Choo, G. Carleo, N. Regnault, T. Neupert: Symmetries and many-body excitations with neural-network quantum states, *Phys. Rev. Lett.*, **121**, 167204 (2018).

[270] F. Zheng, X. Gao, A. Eisfeld: Excitonic wave function reconstruction from near-field spectra using machine learning techniques, *Phys. Rev. Lett.*, **123**, 163202 (2019).

[271] G. Hegde, R. C. Bowen: Machine-learned approximations to density functional theory Hamiltonians, *Sci. Rep.*, **7**, 42669 (2017).

[272] M. Gastegger, L. González, P. Marquetand: Exploring density functional subspaces with genetic algorithms, *Monatsh. Chem.*, **150**, 173 (2019).

[273] J. Nelson, R. Tiwari, S. Sanvito: Machine learning density functional theory for the hubbard model, *Phys. Rev. B*, **99**, 075132 (2019).

[274] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, R. Ramprasad: Solving the electronic structure problem with machine learning, *npj Comput. Mater.*, **5**, 22 (2019).

[275] L. Cheng, M. Welborn, A. S. Christensen, T. F. Miller: A universal density matrix functional from molecular orbital-based machine learning: Transferability across organic molecules, *J. Chem. Phys.*, **150**, 131103 (2019).

[276] X. Lei, A. J. Medford: Design and analysis of machine learning exchange-correlation functionals via rotationally invariant convolutional descriptors, *Phys. Rev. Materials*, **3**, 063801 (2019).

[277] Y. Zhou, J. Wu, S. Chen, G. Chen: Toward the exact exchange–correlation potential: A three-dimensional convolutional neural network construct, *J. Phys. Chem. Lett.*, **10**, 7264 (2019).

[278] B. Kolb, L. C. Lentz, A. M. Kolpak: Discovering charge density functionals and structure-property relationships with prophet: A general framework for coupling machine learning and first-principles methods, *Sci. Rep.*, **7** (2017).

[279] M. J. Willatt, F. Musil, M. Ceriotti: Atom-density representations for machine learning, *J. Chem. Phys.*, **150**, 154110 (2019).

[280] M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld: Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning, *Phys. Rev. Lett.*, **108**, 058301 (2012).

[281] O. A. von Lilienfeld, R. Ramakrishnan, M. Rupp, A. Knoll: Fourier series of atomic radial distribution functions: A molecular fingerprint for machine learning models of quantum chemical properties, *Int. J. Quantum Chem.*, **115**, 1084 (2015).

[282] M. Rupp, R. Ramakrishnan, O. A. von Lilienfeld: Machine learning for quantum mechanical properties of atoms in molecules, *J. Phys. Chem. Lett.*, **6**, 3309 (2015).

[283] J. Behler, K. Reuter, M. Scheffler: Nonadiabatic effects in the dissociation of oxygen molecules at the Al(111) surface, *Phys. Rev. B*, **77**, 115421 (2008).

[284] D. M. G. Williams, W. Eisfeld: Neural Network Diabatization: A New Ansatz for Accurate High-Dimensional Coupled Potential Energy Surfaces, *J. Chem. Phys.*, **149**, 204106 (2018).

[285] Y. Guan, D. H. Zhang, H. Guo, D. R. Yarkony: Representation of coupled adiabatic potential energy surfaces using neural network based quasi-diabatic Hamiltonians: $1,2\,^2$A' states of LiFH, *Phys. Chem. Chem. Phys.*, 10.1039/C8CP06598E (2019).

[286] J. P. Alborzpour, D. P. Tew, S. Habershon: Efficient and accurate evaluation of potential energy matrix elements for quantum dynamics using gaussian process regression, *J. Chem. Phys.*, **145**, 174112 (2016).

[287] G. W. Richings, C. Robertson, S. Habershon: Improved on-the-fly MCTDH simulations with many-body-potential tensor decomposition and projection diabatization, *J. Chem. Theory Comput.*, **15**, 857 (2019).

[288] I. Polyak, G. W. Richings, S. Habershon, P. J. Knowles: Direct quantum dynamics using variational gaussian wavepackets and gaussian process regression, *J. Chem. Phys.*, **150**, 041101 (2019).

[289] Y. Wang, C. Xie, H. Guo, D. R. Yarkony: A quasi-diabatic representation of the $1,2^1$a states of methylamine, *J. Phys. Chem. A*, **123**, 5231 (2019).

[290] Y. Guan, D. R. Yarkony: Accurate neural network representation of the *Ab Initio* determined spin−orbit interaction in the diabatic representation including the effects of conical intersections, *J. Phys. Chem. Lett.*, **11**, 1848 (2020).

[291] H. M. Netzloff, M. A. collins, M. S. Gordon: Growing Multiconfigurational Potential Energy Surfaces with Applications to X+H_2 (X=C,N,O) Reactions, *J. Chem. Phys.*, **124**, 154104 (2006).

[292] F. Häse, I. Fdez. Galván, A. Aspuru-Guzik, R. Lindh, M. Vacher: How machine learning can assist the interpretation of *Ab Initio* molecular dynamics simulations and conceptual understanding of chemistry, *Chem. Sci.*, **10**, 2298 (2019).

[293] N. M. O'Boyle, C. M. Campbell, G. R. Hutchison: Computational design and selection of optimal organic photovoltaic materials, *J. Phys. Chem. C*, **115**, 16200 (2011).

[294] J. L. Teunissen, F. De Proft, F. De Vleeschouwer: Tuning the homo−lumo energy gap of small diamondoids using inverse molecular design, *J. Chem. Theory Comput.*, **13**, 1351 (2017).

[295] D. Liu, Y. Tan, E. Khoram, Z. Yu: Training deep neural networks for the inverse design of nanophotonic structures, *ACS Photonics*, **5**, 1365 (2018).

[296] M.-H. Lee: Robust random forest based non-fullerene organic solar cells efficiency prediction, *Org. Electron.*, **76**, 105465 (2020).

[297] L. C. Blum, J.-L. Reymond: 970 million druglike small molecules for virtual screening in the chemical universe database GDB-13, *J. Am. Chem. Soc.*, **131**, 8732 (2009).

[298] L. Ruddigkeit, R. van Deursen, L. C. Blum, J.-L. Reymond: Enumeration of 166 billion organic small molecules in the chemical universe database gdb-17, *J. Chem. Inf. Model.*, **52**, 2864 (2012).

[299] M. Schwilk, D. N. Tahchieva, O. A. von Lilienfeld: Large yet bounded: Spin gap ranges in carbenes (2020).

[300] K. T. Schütt, F. Arbabzadah, S. Chmiela, K. R. Müller, A. Tkatchenko: Quantum-chemical insights from deep tensor neural networks, *Nat. Commun.*, **8**, 13890 EP (2017).

[301] S. Chmiela, A. Tkatchenko, H. E. Sauceda, I. Poltavsky, K. T. Schütt, K.-R. Müller: Machine learning of accurate energy-conserving molecular force fields, *Sci. Adv.*, **3** (2017).

[302] C. Sutton, L. Ghiringhelli, T. Yamamoto, Y. Lysogorskiy, L. Blumenthal, J. R. Hammerschmidt, T. Golebiowski, X. Liu, A. Ziletti, M. Scheffler: Crowd-sourcing materials-science challenges with the NOMAD 2018, *npj Comput. Mater.*, **5** (2019).

[303] A. P. Bartók, R. Kondor, G. Csányi: On representing chemical environments, *Phys. Rev. B*, **87**, 184115 (2013).

[304] M. Rupp: Machine learning for quantum mechanics in a nutshell, *Int. J. Quantum Chem.*, **115**, 1058 (2015).

[305] A. S. Christensen, L. A. Bratholm, F. A. Faber, O. Anatole von Lilienfeld: Fchl revisited: Faster and more accurate quantum machine learning, *J. Chem. Phys.*, **152**, 044107 (2020).

[306] V. Botu, R. Ramprasad: Adaptive machine learning framework to accelerate *Ab Initio* molecular dynamics, *Int. J. Quant. Chem.*, **115**, 1074 (2015).

[307] J. S. Smith, B. Nebgen, N. Lubbers, O. Isayev, A. E. Roitberg: Less is more: Sampling chemical space with active learning, *J. Chem. Phys.*, **148**, 241733 (2018).

[308] R. E. Bruccoleri, M. Karplus: Conformational sampling using high-temperature molecular dynamics, *Biopolymers*, **29**, 1847 (1990).

[309] T. Maximova, R. Moffatt, B. Ma, R. Nussinov, A. Shehu: Principles and overview of sampling methods for modeling macromolecular structure and dynamics, *PLOS computational Biology*, **12**, 1 (2016).

[310] J. Kästner: Umbrella sampling, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **1**, 932 (2011).

[311] Y. I. Yang, Q. Shao, J. Zhang, L. Yang, Y. Q. Gao: Enhanced sampling in molecular dynamics, *J. Chem. Phys.*, **151**, 070902 (2019).

[312] P. O. Dral, A. Owens, S. N. Yurchenko, W. Thiel: Structure-based sampling and self-correcting machine learning for accurate calculations of potential energy surfaces and vibrational levels, *J. Chem. Phys.*, **146**, 244108 (2017).

[313] M. Gastegger: *Artificial Intelligence in Theoretical Chemistry*, Ph.D. thesis, University of Vienna (2017).

[314] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, K.-R. Müller: Assessment and validation of machine learning methods for predicting molecular atomization energies, *J. Chem. Theory Comput.*, **9**, 3404 (2013).

[315] C. M. Bishop: *Pattern Recognition and Machine Learning*, Springer, New York, 1st edition (2006).

[316] T. Hofmann, B. Schölkopf, A. J. Smola: Kernel methods in machine learning, *Ann. Statist.*, **36**, 1171 (2008).

[317] P. O. Dral: Quantum chemistry in the age of machine learning, *J. Phys. Chem. Lett.*, **11**, 2336 (2020).

[318] C. E. Rasmussen: *Gaussian Processes in Machine Learning*, 63–71, Springer Berlin Heidelberg, Berlin, Heidelberg (2004).

[319] V. Vapnik: *Estimation of Dependences Based on Empirical Data: Springer Series in Statistics (Springer Series in Statistics)*, Springer-Verlag, Berlin, Heidelberg (1982).

[320] Y. Engel, S. Mannor, R. Meir: Sparse Online Greedy Support Vector Regression, in T. Elomaa, H. Mannila, H. Toivonen (editors), *Machine Learning: ECML 2002*, 84–96, Springer Berlin Heidelberg, Berlin, Heidelberg (2002).

[321] M. Awad, R. Khanna: *Support Vector Regression*, 67–80, Apress, Berkeley, CA (2015).

[322] H. M. Cartwright (editor): *Machine Learning in Chemistry*, Theoretical and Computational Chemistry Series, The Royal Society of Chemistry (2020).

[323] J. Behler, M. Parrinello: Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces, *Phys. Rev. Lett.*, **98**, 146401 (2007).

[324] J. Behler: Atom-centered symmetry functions for constructing high-dimensional neural network potentials, *J. Chem. Phys.*, **134**, 074106 (2011).

[325] K. Schütt: *Learning Representations of Atomistic Systems with Deep Neural Networks*, Doctoral thesis, Technische Universität Berlin, Berlin (2018).

[326] R. Ramakrishnan, O. A. von Lilienfeld: *Machine Learning, Quantum Chemistry, and Chemical Space*, chapter 5, 225–256, John Wiley & Sons, Ltd (2017).

[327] M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi, P. Marquetand: wacsf – weighted atom-centered symmetry functions as descriptors in machine learning potentials, *J. Chem. Phys.*, **148**, 241709 (2018).

[328] J. E. Herr, K. Koh, K. Yao, J. Parkhill: Compressing physics with an autoencoder: Creating an atomic species representation to improve machine learning models in the chemical sciences, *J. Chem. Phys.*, **151**, 084103 (2019).

[329] M. K. K. Leung, A. Delong, B. Alipanahi, B. J. Frey: Machine learning in genomic medicine: A review of computational problems and data sets, *Proc. IEEE*, **104**, 176 (2016).

[330] C. Chen, A. Seff, A. Kornhauser, J. Xiao: Deepdriving: Learning affordance for direct perception in autonomous driving, in *The IEEE International Conference on Computer Vision (ICCV)* (2015).

[331] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, A. Lerer: Automatic differentiation in pytorch (2017).

[332] D. P. Kingma, J. Ba: Adam: A method for stochastic optimization, *arXiv:1412.6980*, **abs/1412.6980** (2014).

[333] J. Duchi, E. Hazan, Y. Singer: Adaptive Subgradient Methods for Online Learning and Stochastic Optimization, *J. Mach. Learn. Res.*, **12**, 2121 (2011).

[334] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, O. A. von Lilienfeld: Prediction errors of molecular machine learning models lower than hybrid dft error, *J. Chem. Theory Comput.*, **13**, 5255 (2017).

[335] T. Fushiki: Estimation of prediction error by using K-fold cross-validation, *Stat. Comput.*, **21**, 137–146 (2011).

[336] C. cortes, L. D. Jackel, S. A. Solla, V. Vapnik, J. S. Denker: Learning curves: Asymptotic values and rate of convergence, in J. D. cowan, G. Tesauro, J. Alspector (editors), *Advances in Neural Information Processing Systems 6*, 327–334, Morgan-Kaufmann (1994).

[337] K.-R. Müller, M. Finke, N. Murata, K. Schulten, S. Amari: A numerical study on learning curves in stochastic multilayer feedforward networks, *Neural Comput.*, **8**, 1085 (1996).

[338] C. Robertson, J. González-Vázquez, I. corral, S. Díaz-Tendero, C. Díaz: Nonadiabatic scattering of NO off Au3 clusters: A simple and robust diabatic state manifold generation method for multiconfigurational wavefunctions, *J. Comput. Chem.*, **40**, 794 (2019).

[339] A. Christensen, F. Faber, B. Huang, L. Bratholm, A. Tkatchenko, K. Müller, O. Lilienfeld: Qml: A python toolkit for quantum machine learning, https://github.com/qmlcode/qml (2017).

[340] Theano Development Team: Theano: A Python framework for fast computation of mathematical expressions, *arXiv*, **abs/1605.02688** (2016).

[341] A. H. Larsen *et al.*: The Atomic Simulation Environment—a Python Library for Working with Atoms, *J. Phys.: Condens. Matter*, **29**, 273002 (2017).

[342] H. Köppel, J. Gronki, S. Mahapatra: Construction scheme for regularized diabatic states, *J. Chem. Phys.*, **115**, 2377 (2001).

[343] B. Gonon, A. Perveaux, F. Gatti, D. Lauvergnat, B. Lasorne: On the applicability of a wavefunction-free, energy-based procedure for generating first-order non-adiabatic couplings around conical intersections, *J. Chem. Phys.*, **147**, 114114 (2017).

[344] K. K. Baeck, H. An: Practical Approximation of the Non-Adiabatic Coupling Terms for Same-Symmetry Interstate Crossings by Using Adiabatic Potential Energies Only, *J. Chem. Phys.*, **146**, 064107 (2017).

[345] H. An, K. K. Baeck: Practical and reliable approximation of nonadiabatic coupling terms between triplet electronic states using only adiabatic potential energies, *Chem. Phys. Lett.*, **696**, 100 (2018).

[346] G. König, F. Pickard, J. Huang, W. Thiel, A. MacKerell, B. Brooks, D. York: A Comparison of QM/MM Simulations with and without the Drude Oscillator Model Based on Hydration Free Energies of Simple Solutes, *Molecules*, **23**, 2695 (2018).

[347] O. Andreussi, I. G. Prandi, M. Campetella, G. Prampolini, B. Mennucci: Classical Force Fields Tailored for QM Applications: Is It Really a Feasible Strategy?, *J. Chem. Theory Comput.*, **13**, 4636 (2017).

[348] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay: Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, **12**, 2825 (2011).

[349] J. Westermayr, P. Marquetand: *Machine Learning for Nonadiabatic Molecular Dynamics*, chapter 04, Issn Series, Royal Society of Chemistry (2020).

[350] H.-T. Cheng, L. Koc, J. Harmsen, T. Shaked, T. Chandra, H. Aradhye, G. Anderson, G. corrado, W. Chai, M. Ispir, R. Anil, Z. Haque, L. Hong, V. Jain, X. Liu, H. Shah: Wide & deep learning for recommender systems, *arXiv:1606.07792* (2016).

[351] T. Truong, R. Behrsohn, P. Brumer, C. K. Luk, T. Tao: Effect of ph on the phosphorescence of tryptophan, tyrosine, and proteins, *J. Biol. Chem.*, **242**, 2979 (1967).

[352] J. Gu, J. Ma, J. Jiang, L. Yang, J. Yang, J. Zhang, H. Chi, Y. Song, S. Sun, W. Q. Tian: Hydrated electron (e_{aq}^-) generation from phenol/UV: Efficiency, influencing factors, and mechanism, *Appl. Catal., B*, **200**, 585 (2017).

[353] D. V. Bent, E. Hayon: Excited state chemistry of aromatic amino acids and related peptides. i. tyrosine, *J. Am. Chem. Soc.*, **97**, 2599 (1975).

[354] L. H. Jones, A. Narayanan, E. C. Hett: Understanding and applying tyrosine biochemical diversity, *Mol. BioSyst.*, **10**, 952 (2014).

[355] W. Lee, I. Kim, Y. M. Rhee: A proton transfer network that generates deprotonated tyrosine is a key to producing reactive oxygen species in phototoxic KillerRed protein, *Phys. Chem. Chem. Phys.*, **20**, 22342 (2018).

[356] M. Marazzi, U. Sancho, O. Castano, W. Domcke, L. M. Frutos: Photoinduced proton transfer as a possible mechanism for highly efficient excited-state deactivation in proteins, *J. Phys. Chem. Lett*, **1**, 425 (2009).

[357] O. B. Morozova, A. V. Yurkovskaya: Intramolecular electron transfer in the photooxidized peptides tyrosine–histidine and histidine–tyrosine: A time-resolved CIDNP study, *Angew. Chem. Int. Ed.*, **49**, 7996 (2010).

[358] D. L. Jenson, B. A. Barry: Proton-coupled electron transfer in photosystem ii: proton inventory of a redox active tyrosine, *J. Am. Chem. Soc.*, **131**, 10567 (2009).

[359] A. G. Tebo, A. Quaranta, C. Herrero, V. L. Pecoraro, A. Aukauloo: Intramolecular photogeneration of a tyrosine radical in a designed protein, *ChemPhotoChem*, **1**, 89 (2017).

[360] S.-L. J. Lahey, C. N. Rowley: Simulating protein–ligand binding with neural network potentials, *Chem. Sci.*, **11**, 2362 (2020).

[361] Y.-J. Zhang, A. Khorshidi, G. Kastlunger, A. A. Peterson: The potential for Machine Learning in hybrid QM/MM calculations, *J. Chem. Phys.*, **148**, 241740 (2018).

[362] M. N. R. Ashfold, B. Cronin, A. L. Devine, R. N. Dixon, M. G. D. Nix: The Role of $\pi\sigma^*$ Excited States in the Photodissociation of Heteroaromatic Molecules, *Science*, **312**, 1637 (2006).

[363] C.-M. Tseng, Y. T. Lee, M.-F. Lin, C.-K. Ni, S.-Y. Liu, Y.-P. Lee, Z. Xu, M. Lin: Photodissociation dynamics of phenol, *J. Phys. Chem. A*, **111**, 9463 (2007).

[364] C.-M. Tseng, Y. T. Lee, C.-K. Ni, J.-L. Chang: Photodissociation dynamics of the chromophores of the amino acid tyrosine: p-methylphenol, p-ethylphenol, and p-(2-aminoethyl)phenol, *J. Phys. Chem. A*, **111**, 6674 (2007).

[365] Y. Zhang, T. A. A. Oliver, M. N. R. Ashfold, S. E. Bradforth: contrasting the excited state reaction pathways of phenol and *para*-methylthiophenol in the gas and liquid phases, *Faraday Discuss.*, **157** (2012).

[366] Y.-C. Lin, C. Lee, S.-H. Lee, Y.-Y. Lee, Y. T. Lee, C.-M. Tseng, C.-K. Ni: Excited-state dissociation dynamics of phenol studied by a new time-resolved technique, *J. Chem. Phys.*, **148**, 074306 (2018).

[367] M. N. R. Ashfold, M. Bain, C. S. Hansen, R. A. Ingle, T. N. V. Karsili, B. Marchetti, D. Murdock: Exploring the dynamics of the photoinduced ring-opening of heterocyclic molecules, *J. Phys. Chem. Lett.*, **8**, 3440 (2017).

[368] O. P. J. Vieuxmaire, Z. Lan, A. L. Sobolewski, W. Domcke: *Ab Initio* characterization of the conical intersections involved in the photochemistry of phenol, *J. Chem. Phys.*, **129**, 224307 (2008).

[369] H. Y. Lai, W. R. Jhang, C.-M. Tseng: communication: Mode-dependent excited-state lifetime of phenol under the s1/s2 conical intersection, *J. Chem. Phys.*, **149**, 031104 (2018).

[370] I. Sandler, J. J. Nogueira, L. González: Solvent reorganization triggers photo-induced solvated electron generation in phenol, *Phys. Chem. Chem. Phys.*, **21**, 14261 (2019).

[371] C. Xie, B. Zhao, C. L. Malbon, D. R. Yarkony, D. Xie, H. Guo: Insights into the mechanism of nonadiabatic photodissociation from product vibrational distributions. the remarkable case of phenol, *J. Phys. Chem. Lett.*, **11**, 191 (2020).

[372] G. Grégoire, C. Jouvet, C. Dedonder, A. L. Sobolewski: *Ab Initio* study of the excited-state deactivation pathways of protonated tryptophan and tyrosine, *J. Am. Chem. Soc.*, **129**, 6223 (2007).

[373] G. Tomasello, M. Wohlgemuth, J. Petersen, R. Mitrić: Photodynamics of free and solvated tyrosine, *J. Phys. Chem. B*, **116**, 8762 (2012).

[374] G. Féraud, M. Broquier, C. Dedonder, C. Jouvet, G. Grégoire, S. Soorkia: Excited state dynamics of protonated phenylalanine and tyrosine: Photo-induced reactions following electronic excitation, *J. Phys. Chem. A*, **119**, 5914 (2014).

[375] M. Zhang, Z. Huang, Z. Lin: Systematic *Ab Initio* studies of the conformers and conformational distribution of gas-phase tyrosine, *J. Chem. Phys.*, **122**, 134313 (2005).

[376] F. Neese: The ORCA program system, *WIREs Comput. Mol. Sci.*, **2**, 73 (2012).

[377] I. Fdez. Galván, M. Vacher, A. Alavi, C. Angeli, F. Aquilante, J. Autschbach, J. J. Bao, S. I. Bokarev, N. A. Bogdanov, R. K. Carlson, L. F. Chibotaru, J. Creutzberg, N. Dattani, M. G. Delcey, S. S. Dong, A. Dreuw, L. Freitag, L. M. Frutos, L. Gagliardi, F. Gendron, A. Giussani, L. González, G. Grell, M. Guo, C. E. Hoyer, M. Johansson, S. Keller, S. Knecht, G. Kovaçević, E. Källman, G. Li Manni, M. Lundberg, Y. Ma, S. Mai, J. P. Malhado, P.-Å. Malmqvist, P. Marquetand, S. A. Mewes, J. Norell, M. Olivucci, M. Oppel, Q. M. Phung, K. Pierloot, F. Plasser, M. Reiher, A. M. Sand, I. Schapiro, P. Sharma, C. J. Stein, L. K. Sørensen, D. G. Truhlar, M. Ugandi, L. Ungur, A. Valentini, S. Vancoillie, V. Veryazov, O. Weser, T. A. Wesołowski, P.-O. Widmark, S. Wouters, A. Zech, J. P. Zobel, R. Lindh: Open-molcas: From source code to insight, *J. Chem. Theory Comput.*, **15**, 5925 (2019).

[378] B. Roos, R. Lindh, P.-Å. Malmqvist, V. Veryazov, P.-O. Widmark: Main group atoms and dimers studied with a new relativistic ano basis set, *J. Phys. Chem. A*, **108**, 2851 (2004).

[379] F. C. Smith, C. A. L. Evans: The ultra-violet absorption spectra of certain aromatic amino-acids, and of the serum proteins, *Proc. R. Soc. Lond. B.*, **104**, 198 (1929).

[380] A. J. Ratner, H. R. Ehrenberg, Z. Hussain, J. Dunnmon, C. Ré: Learning to compose domain-specific transformations for data augmentation, *arXiv*, **1709.01643** (2017).

[381] M. Kuchnik, V. Smith: Efficient augmentation via data subsampling, *arXiv*, **abs/1810.05222** (2018).

[382] G. M. Roberts, V. G. Stavros: The role of $\pi\sigma^{ast}$ states in the photochemistry of heteroaromatic biomolecules and their subunits: insights from gas-phase femtosecond spectroscopy, *Chem. Sci.*, **5**, 1698 (2014).

[383] W. Pronobis, K. R. Schütt, A. Tkatchenko, K.-R. Müller: Capturing intensive and extensive dft/tddft molecular properties with machine learning, *Eur. Phys. J. B*, **91**, 178 (2018).