



Cite this: *Digital Discovery*, 2025, 4, 1478

Excited-state nonadiabatic dynamics in explicit solvent using machine learned interatomic potentials†

Maximilian X. Tiefenbacher, ^{ab} Brigitta Bachmair, ^{abc} Cheng Giuseppe Chen, ^{cd} Julia Westermayr, ^{ef} Philipp Marquetand, ^{ac} Johannes C. B. Dietschreit ^{*c} and Leticia González ^{*ac}

Excited-state nonadiabatic simulations with quantum mechanics/molecular mechanics (QM/MM) are essential to understand photoinduced processes in explicit environments. However, the high computational cost of the underlying quantum chemical calculations limits its application in combination with trajectory surface hopping methods. Here, we use FieldSchNet, a machine-learned interatomic potential capable of incorporating electric field effects into the electronic states, to replace traditional QM/MM electrostatic embedding with its ML/MM counterpart for nonadiabatic excited state trajectories. The developed method is applied to furan in water, including five coupled singlet states. Our results demonstrate that with sufficiently curated training data, the ML/MM model reproduces the electronic kinetics and structural rearrangements of QM/MM surface hopping reference simulations. Furthermore, we identify performance metrics that provide robust and interpretable validation of model accuracy.

Received 30th January 2025
Accepted 11th April 2025

DOI: 10.1039/d5dd00044k
rsc.li/digitaldiscovery

1 Introduction

Photochemistry lies at the heart of essential natural processes. One example is photosynthesis, where the absorption of light by chlorophyll triggers a cascade of electron transfers, ultimately capturing solar energy and storing it in chemical bonds.^{1–3} Beyond sustaining life, light-driven reactions hold promise for advancements in energy conversion,⁴ molecular electronics,⁵ and the development of photonic materials.⁶ Understanding photochemical mechanisms is thus important to inspire new technologies that harness the power of light. To this end, computational approaches play a key role, offering time-resolved insights that complement and enhance modern ultrafast experimental techniques.^{7,8}

A popular method to carry out excited-state dynamics simulations of molecular systems is trajectory surface hopping (TSH).⁹ In this so-called mixed quantum-classical approach, electrons – responsible for electronic transitions and excited-state properties – are treated quantum mechanically, while the heavier nuclei are described using classical mechanics. At the expense of neglecting nuclear quantum effects, TSH effectively captures the quantum nature of electrons, even if it requires propagating numerous independent classical nuclear trajectories to accurately simulate the behavior of a nuclear wave packet as it splits during nonadiabatic events.^{7,10,11} Despite being attractive, the need for many trajectories makes TSH simulations computationally expensive, especially when the underlying on-the-fly calculations of the coupled potential energy surfaces (PESs) are performed using accurate quantum mechanical methods.

The situation becomes more challenging when simulating photochemical processes in the condensed phase.¹² For example, chromophores in nature are rarely isolated; rather, they are typically embedded within complex, heterogeneous environments that include a variety of molecular interactions, including solvent effects, hydrogen bonding, and intricate structural features. These environmental factors can significantly influence the photochemical behavior of the system, altering deactivation pathways. Implicit solvation models,¹³ which treat the solvent as a continuous medium, are often insufficient to capture detailed interactions. A more accurate treatment is achieved by modeling the environment explicitly. One efficient way to do that is by using hybrid quantum

^aResearch Platform on Accelerating Photoreaction Discovery (ViRAPID), University of Vienna, Währinger Straße 17, 1090 Vienna, Austria. E-mail: leticia.gonzalez@univie.ac.at

^bVienna Doctoral School in Chemistry, University of Vienna, Währinger Straße 42, 1090 Vienna, Austria

^cInstitute of Theoretical Chemistry, Faculty of Chemistry, University of Vienna, Währinger Straße 17, 1090 Vienna, Austria. E-mail: johannes.dietschreit@univie.ac.at

^dDepartment of Chemistry, Sapienza University of Rome, Piazzale Aldo Moro, 5, Rome, 00185, Italy

^eWilhelm-Ostwald Institute, University of Leipzig, Linnéstraße 2, 04103 Leipzig, Germany

^fCenter for Scalable Data Analytics and Artificial Intelligence (ScalDS.AI), Dresden/Leipzig, Humboldtstraße 25, 04105 Leipzig, Germany

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d5dd00044k>

mechanics/molecular mechanics (QM/MM) methods, where the region of interest, such as the excited chromophore, is treated with QM, while the surrounding environment, including solvent or complex heterogeneous scaffolds, is modeled with MM.^{14–16} This approach offers a balance between accuracy and computational cost by restricting expensive QM calculations to the critical region of the system that actually requires a quantum mechanical treatment. The advantages of QM/MM strategies are indisputable, and in combination with TSH, have enabled impressive simulations of time-resolved photochemical processes in large complex systems.^{17–22} However, because the computational expense of a QM/MM calculation is largely determined by the level of theory employed in the QM region, QM/MM dynamical simulations suffer from comparable (or greater) costs than simulations of the isolated chromophore. The situation becomes more prohibitive the more trajectories are needed to achieve statistically meaningful results.¹⁷

Over the years, several strategies have been developed to reduce the cost of TSH simulations. One notable example is the use of vibronic coupling models,²³ which replace the expensive on-the-fly calculations of the PESs by pre-parameterized potentials that are approximated by scaled harmonic oscillators in the simplest case.^{24,25} Recently, TSH simulations using linear vibronic coupling PESs have been extended to include a classical MM environment,²⁶ enabling efficient time-resolved analysis of three-dimensional solvent–solute interactions.²⁷ While this approach reduces the computational cost significantly, it is only applicable to rather rigid molecules, where anharmonic effects, such as large amplitude motions or bond rearrangements, do not play a role in the relaxation dynamics.

An alternative and highly flexible approach to reduce the cost of the underlying electronic structure problem is the use of machine learned (ML) potentials.^{28–31} Trained on high-quality quantum mechanical data, ML potentials have demonstrated their ability to replicate the accuracy of *ab initio* calculations at a fraction of the computational cost.³² ML potentials have already shown considerable success in modeling dynamics in the electronic ground-state.^{33–38} In contrast, their application to excited-state dynamics, which is significantly more expensive than ground state simulations, is currently only feasible for small molecular systems like organic chromophores and is limited by the availability of accurate reference data.^{39–45} This is in contrast to ground-state ML potentials, of which many are transferable between molecular systems and can thus simulate large biomolecules or materials using training data of smaller building blocks. As a consequence, the simulation of large systems in their excited state also requires methods like mixed ML/MM (machine learning/molecular mechanics).⁴⁶ However, the integration of ML potentials with MM for both ground- and excited-state dynamics remains in its infancy.

One of the key challenges in developing an ML/MM approach is to accurately describe the interaction between the ML potential and the surrounding MM environment. In QM/MM simulations, the interactions between the quantum region and the classical region are clearly defined by the Hamiltonian, which ensures that the treatment of the two regions is physically correct.¹⁴ For ML/MM, the interaction

needs to be carefully modeled to ensure that the combined system behaves correctly; however, there appears to be little consensus on the best way to do so.^{47–58}

A recent study by Mazzeo *et al.*⁵⁹ demonstrates the use of Gaussian process regression to describe the excited-state dynamics of a solvated molecule by learning ML/MM energies and forces with kernel models in a two-step process. First, they fit the vacuum PESs and then subsume the differences between pure QM and QM/MM under polarization, which is described by a second model. Their approach is restricted to purely adiabatic dynamics in the excited state, neglecting any coupling between states. In contrast, to the best of our knowledge, ML/MM implementations for nonadiabatic excited-state dynamics using an electrostatic embedding do not exist. In this work, we propose the first ML/MM nonadiabatic excited-state dynamics using electrostatic embedding and a general number of electronic states. We use the FieldSchNet architecture of Gastegger *et al.*,⁵¹ which allows the inclusion of the electric field *via* an additional ML input. The electric field is generated by point charges of the MM environment and alters the different excited states. As an application, we investigate the excited-state dynamics of furan in water (Fig. 1a), including three coupled electronic singlet states. Furan is a small heterocyclic organic molecule that serves as a building block in biologically relevant systems, such as DNA and proteins, and has long been the focus of theoretical and experimental studies.^{60–68} Few also considered the explicit interaction of furan with water forming hydrogen bonds,^{69,70} highlighting the need of QM/MM studies able to capture these interactions and their impact on the excited state relaxation dynamics explicitly.

The remainder of the paper is organized as follows. In Section 2, we present the theory behind a QM/MM setup with electrostatic embedding and describe all the necessary terms when using an ML model. Next, in Section 3 we outline the data collection method, architecture, and training process of the ML interatomic potentials and summarize the numerical

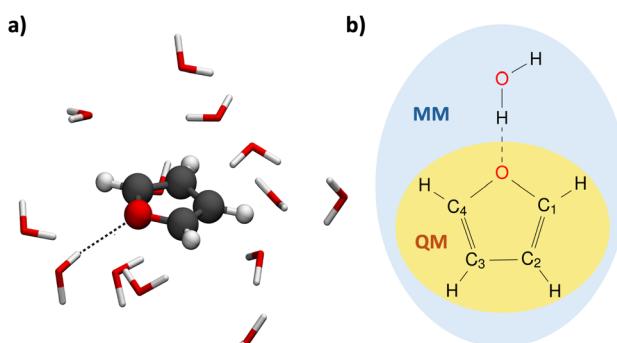


Fig. 1 Depiction of the investigated system, furan, treated quantum mechanically (QM), solvated in water, which is described with molecular mechanics (MM). (a) 3D rendering of furan, depicted as balls and sticks, surrounded by the water molecules of the first solvation shell, shown as sticks. The hydrogen bond present in this configuration is highlighted by a black dashed line. Hydrogen, carbon, and oxygen atoms are colored in white, black, and red, respectively. (b) Lewis structure of furan with numbered carbon atoms.



experiments performed. The results of various training settings, the quality of the obtained PESs, and the ML-driven dynamics are compared to on-the-fly TSH using the quantum chemical reference method used to train our ML potentials in Section 4.

2 Theory

2.1 Electrostatic embedding

In QM/MM simulations, the most relevant region of the chemical system is described with quantum mechanics (QM region), while the surroundings are modeled with computationally more efficient classical force fields (MM region). The total Hamiltonian H_{tot} and total energy E_{tot} of the system within an electrostatic embedding framework can be expressed as,¹⁴

$$H_{\text{tot}} = H_{\text{MM}} + H_{\text{QM}} + H_{\text{QM-MM}}, \quad (1)$$

$$E_{\text{tot}} = E_{\text{MM}} + E_{\text{QM-MM}}^{\text{vdW}} + E_{\text{QM}} + \underbrace{E_{\text{QM-MM}}^{\text{Coulomb}}}_{E_{\text{QM}}^{\text{el-embed.}}}, \quad (2)$$

where the indices MM and QM represent the contributions of each region individually, and QM-MM indicates terms that depend on both regions. Assuming that the border between the QM and MM regions does not cut through any chemical bonds, the only interactions between the QM and MM regions are the van der Waals (vdW) and Coulomb potentials (eqn (2)). In the context of electrostatic embedding, the vdW interactions are computed at the level of the MM region, while the Coulomb interactions are represented by an additional term in the Hamiltonian of the QM region.⁷¹ The term $E_{\text{QM-MM}}^{\text{Coulomb}}$ includes the influence of the surrounding classical region through electrostatic interactions in the description of the QM region. The partial atomic charges of the classical environment $\{q_j^{\text{MM}}\}$ located at $\{\mathbf{R}_j^{\text{MM}}\}$ generate an electrostatic potential at position \mathbf{r} :

$$V_{\text{MM}}(\mathbf{r}) = \sum_j^{N_{\text{MM}}} \frac{q_j^{\text{MM}}}{|\mathbf{r} - \mathbf{R}_j^{\text{MM}}|} \quad (3)$$

Eqn (3) is part of $H_{\text{QM-MM}}$ and is included in the calculation of the quantum subsystem, thus coupling the QM and MM regions.^{14,53}

2.2 Adapted machine-learned gradients for electrostatic embedding

The QM calculation is set to find the wave function, *i.e.* the electron density, corresponding to the given Hamiltonian, which in the case of electrostatic embedding includes the electrostatic potential defined in eqn (3).¹⁴ Hence, as implied by eqn (2), the energy and subsequently also the forces computed for the electrostatic embedding Hamiltonian cannot be separated into the vacuum and the MM polarization contributions without additional QM calculations of the system in vacuum. This is important, because it means that the energy (and forces) of the same arrangement of atoms in the QM region will vary depending on the configuration of the MM region. Therefore, to perform electrostatic embedding simulations using an ML

interatomic potential, it is essential to model the QM region and its interaction with the surrounding MM region simultaneously. This can be accomplished by either passing the positions and charges of the MM atoms to the ML architecture, $\hat{E}_{\text{QM}}^{\text{el-embed.}} = E_{\text{ML}}(\{\mathbf{R}^{\text{QM}}\}, \{\mathbf{R}^{\text{MM}}\}, \{q^{\text{MM}}\})$, or by considering the electric field ϵ created by the surrounding MM atoms at the position of each QM atom, $\hat{E}_{\text{QM}}^{\text{el-embed.}} = E_{\text{ML}}(\{\mathbf{R}^{\text{QM}}\}, \{\epsilon\})$. The latter approach offers the benefit of obviating the need to provide the positions of all MM atoms to the model. Instead, one can conveniently use precomputed electric field values, substantially minimizing the data set file sizes. In this work, we use FieldSchNet,⁵¹ one of the first ML interatomic potentials capable of reproducing QM/MM calculations with electrostatic embedding by using the electric field as an additional input besides the atomic positions. FieldSchNet is based on the SchNet continuous-filter convolutional neural⁷² network architecture that takes the electric field as an additional input to learn a representation of the system.

The electric field ϵ_i at position \mathbf{R}_i^{QM} of QM atom i is the sum over all atoms in the MM region (N_{MM}) and is defined as

$$\epsilon_i = \sum_j^{N_{\text{MM}}} \epsilon_{ij} = \sum_j^{N_{\text{MM}}} q_j^{\text{MM}} \frac{\mathbf{R}_i^{\text{QM}} - \mathbf{R}_j^{\text{MM}}}{|\mathbf{R}_i^{\text{QM}} - \mathbf{R}_j^{\text{MM}}|^3}, \quad (4)$$

where q_j^{MM} is the partial charge of MM atom j with coordinates \mathbf{R}_j^{MM} . The sum in eqn (4) runs over all MM atoms, independently of the short-range cut-off used for the construction of the local graph. The equation highlights that the electric field depends on both the QM and MM atoms.⁵¹ Therefore, energies that depend on this field induce forces that act on both the QM and MM atoms. However, if only the electric field value is passed to the ML model and not the positions and charges of every MM atom, the gradient of the ML energy with respect to the position of ML (or QM) atoms does not include the field-dependent terms. To circumvent this problem, we add the respective derivative to those forces that are computed *via* automatic differentiation. The forces applied to the ML region are then given by

$$-\frac{dE_{\text{ML}}}{d\mathbf{R}^{\text{ML}}} = -\left(\frac{\partial E_{\text{ML}}}{\partial \mathbf{R}^{\text{ML}}} + \frac{\partial E_{\text{ML}}}{\partial \epsilon} \frac{\partial \epsilon}{\partial \mathbf{R}^{\text{ML}}} \right), \quad (5)$$

whereas the contribution of the ML region to the forces acting on the MM atoms is

$$-\frac{dE_{\text{ML}}}{d\mathbf{R}^{\text{MM}}} = -\frac{\partial E_{\text{ML}}}{\partial \epsilon} \frac{\partial \epsilon}{\partial \mathbf{R}^{\text{MM}}}. \quad (6)$$

The partial derivatives in eqn (5) and (6) of the ML energy with respect to the ML atoms and the field are computed *via* back-propagation. The derivatives of the field with respect to the nuclear positions (MM or ML) are obtained analytically, based on eqn (4).

2.3 Augmented loss for training ML interatomic potentials

When training ML interatomic potentials on QM data, the loss function is usually a linear combination of the mean squared error for energies and forces. Accordingly, the loss function for a single configuration and electronic state can be given by



$$\begin{aligned}\mathcal{L} &= w_E \left(E - \hat{E} \right)^2 + \frac{w_F}{3N_{ML}} \sum_{i=1}^{N_{ML}} \sum_{j \in xyz} \left(F_{ij} - \hat{F}_{ij} \right)^2 \\ &= w_E \left(E - \hat{E} \right)^2 + \frac{w_F}{3N_{ML}} \sum_{i=1}^{N_{ML}} \sum_{j \in xyz} \left(F_{ij} + \left[\frac{\partial \hat{E}}{\partial \mathbf{R}} \right]_{ij} \right)^2,\end{aligned}\quad (7)$$

where N_{ML} is the number of ML atoms in a configuration, w_α is the weight for the loss of property α , and the hat “ $\hat{\cdot}$ ” indicates the prediction of the ML model. However, since the ML prediction does not include the field-dependent gradients, we amend the loss to be

$$\begin{aligned}\mathcal{L} &= w_E \left(E_{QM}^{\text{el.-embed.}} - \hat{E}_{ML} \right)^2 \\ &+ \frac{w_F}{3N_{ML}} \sum_{i=1}^{N_{ML}} \sum_{j \in xyz} \left(F_{ij} + \left[\frac{\partial \hat{E}_{ML}}{\partial \mathbf{R}^{\text{ML}}} + \frac{\partial \hat{E}_{ML}}{\partial \epsilon} \frac{\partial \epsilon}{\partial \mathbf{R}^{\text{ML}}} \right]_{ij} \right)^2.\end{aligned}\quad (8)$$

Since the contribution of the field to the nuclear gradient is expected to be small, this term was neglected in the original FieldSchNet paper.⁵¹ We call eqn (8) the “augmented loss” and use it for all ML trainings in this work, to be consistent with the forces used for dynamics simulations, as we found the field-dependent term to be crucial for stable excited-state molecular dynamics simulations.

2.4 Excited-state dynamics simulations using trajectory surface hopping

The excited-state dynamics are carried out in the context of TSH,⁷³ where the nuclei are propagated classically on a single PES at each time step, solving Newton's equations of motion. Still, TSH couples several PESs by allowing stochastic transitions (or “hops”) between electronic states, which are described by the time-dependent Schrödinger equation,

$$\frac{d}{dt} c_k(t) = -\sum_{\ell} \left[\frac{i}{\hbar} H_{k\ell} + \mathbf{d}_{k\ell} \cdot \mathbf{v} \right] c_{\ell}(t). \quad (9)$$

Here, $c_k(t)$ is the time-dependent coefficient of the electronic wavefunction for state k and $H_{k\ell}$ is an element of the electronic Hamiltonian matrix in the molecular Coulomb Hamiltonian basis, where the electronic states are ordered by energy and can change their character.^{74,75} In this representation, $\mathbf{d}_{k\ell} = \langle \phi_k | \nabla_{\mathbf{R}} | \phi_{\ell} \rangle$ is the nonadiabatic coupling vector between the states k and ℓ , and \mathbf{v} is the velocity vector of the nuclei. The explicit calculation of the nonadiabatic couplings is often circumvented by using the local diabatization scheme,^{76,77} which relies on calculating the overlap matrix between two sequential wavefunctions.^{77,78} Alternatively, when neither explicit nonadiabatic coupling vectors nor wavefunctions are available—as is the case when using an adiabatic machine learning model—it is possible to approximate the coupling vectors using the time derivative of the nuclear forces, as in the so-called curvature-driven Tully surface hopping.⁷⁹

The probability of hopping between the electronic states is determined by the fewest switches algorithm,⁹ which minimizes

the number of state transitions. The transition probability between two states n and m is given by

$$P_{n \rightarrow m} = \max \left(0, \frac{2\text{Re} \left\{ c_n^* c_m \left[\frac{i}{\hbar} H_{nm} + \mathbf{d}_{nm} \cdot \mathbf{v} \right] \right\}}{|c_n|^2} \Delta t \right), \quad (10)$$

where Δt is the nuclear time step. After each time step, a uniform random number $\xi \in [0, 1]$ is drawn. If $\xi < P_{n \rightarrow m}$, the system attempts to switch from state n to state m . If a hop occurs, the nuclear kinetic energy is adjusted to conserve total energy.⁸⁰ However, if the kinetic energy after a hop is insufficient to sustain the motion, the hop is rejected, and the system remains on the original PES.⁸¹

3 Computational details

3.1 Molecular dynamics simulations

Initial molecular dynamics (MD) simulations of furan in water were carried out with AMBER2022.⁸² Partial charges for furan were calculated with antechamber using bond charge corrected charges derived from the semi-empirical Austin Model 1 (AM1-BCC charges),^{83,84} which were combined with GAFF2 parameters using parmchk. The furan molecule was solvated in a cubic box of TIP3P water molecules with a side length of 15 Å containing 1365 water molecules.

For energy minimization, heating, equilibration, and production runs, we used the MD engine sander.⁸² Fig. 2 shows a schematic timeline for the MM–MD simulations and how snapshots for the subsequent excited-state QM/MM simulations were collected.

The MM–MD simulations employ a time step of 2 fs, a Langevin thermostat with a friction constant of 2 ps⁻¹, and constrained hydrogen–heavy atom bond distances by means of the SHAKE⁸⁵ algorithm. After the initial setup, the energy of the system was minimized for 2000 steps using steepest descent. After minimization, the system was heated for 20 ps to 300 K through continuous heat transfer from the

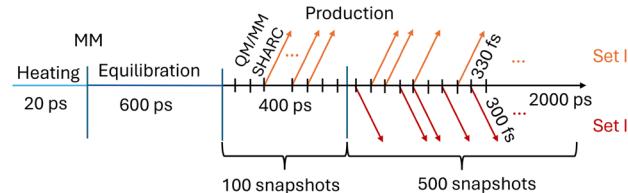


Fig. 2 Timeline of the different molecular dynamics simulations performed for furan in water. Molecular mechanics (MM) is first used for heating (20 ps) and equilibration (600 ps) steps. The production run is divided into two parts, 400 ps and 2000 ps long. Frames are saved every 4 ps, resulting in 100 and 500 snapshots, respectively, which serve as potential initial conditions for two sets of QM/MM-SHARC trajectories. The number of selected trajectories (colored arrows) depends on whether an excitation can occur within a specific energy window (7.3 to 7.5 eV for set I and 6 to 7 eV for set II). For furan in water, this results in 46 (set I, out of 600 initial conditions) and 66 (set II, out of 500 initial conditions selecting only excitations into the S₂) trajectories.

thermostat (the bath temperature is always at 300 K). Subsequently, the system was equilibrated for 600 ps in the isobaric-isothermal ensemble with the Berendsen barostat.⁸⁶ Using the same settings as for the equilibration, we performed a production run with a total of 2.4 ns. Snapshots were recorded every 4 ps, resulting in a total of 600 equidistant frames, which were used to build two sets of initial conditions (sets I and II in Fig. 2) for subsequent excited-state QM/MM simulations – explained next.

3.2 QM/MM trajectory surface hopping simulations

Two sets of nonadiabatic QM/MM-TSH simulations were performed using the QM/MM interface of the SHARC 3.0 (ref. 11, 74, and 75) package to the TINKER MM engine.⁸⁷ Set I was used for model training, validation, and testing. Set II was used to provide reference QM/MM-TSH simulations to compare with ML/MM-TSH dynamics carried out under the exact same conditions (see details below).

The computational details specified here are common to both sets, unless stated otherwise. The interactions of QM and MM regions were described *via* electrostatic embedding. RATTLE⁸⁸ was used to constrain the bond vibrations of the water molecules of the MM region, whereas the hydrogen atoms of furan were left unconstrained. Furan was described using time-dependent density functional theory (TD-DFT) at the BP86/def2-SVP^{89–92} level of theory, as implemented in Orca 5.0.⁹³ This level of theory was chosen based on a benchmark performed by Hieringer *et al.*,⁹⁴ in which a total of 12 different methods were compared, including other DFT functionals, and wave function methods such as CASPT2 or CC3. BP86 showed the best agreement with the experimental excitation energies from Kamada *et al.*⁹⁵

The absorption spectrum of furan in water was calculated using the first 100 frames from the production run of the MM-MD simulations (recall Fig. 2). The spectrum comprises

the lowest ten singlet excited states, covering an energy range up to 9.5 eV, see Fig. 3 and Table S1 of the ESI.[†]

In the simulations set I, we use all 600 snapshots generated from the MM simulation as possible initial conditions (position and velocities) for the subsequent QM/MM-TSH excited-state dynamics, which included the lowest eleven singlet states. Furan is then excited within the energy range of 7.3–7.5 eV, primary targeting the S_3 state. Based on the associated excitation energies and oscillator strengths of the sampled 600 geometries, the initially excited electronic states were selected stochastically,⁹⁶ resulting in 46 initial conditions. Out of those, 39 started in S_3 (the state with a large dipole moment, see Table S1[†]) and 7 started in S_4 .

Following previous studies,^{63,68} which indicate that the relaxation of furan to the ground state occurs within approximately 200 fs, the trajectories of set I were propagated for 330 fs. A nuclear time step of 0.5 fs and an electronic time step of 0.025 fs were employed. Nonadiabatic couplings between singlet states were obtained from the local diabatization algorithm by Granucci *et al.*,⁷⁶ computing the overlap of the wave function between two consecutive time steps.^{77,78} To conserve energy, the velocities of the QM particles were uniformly rescaled after each hop to account for the potential energy difference between the new and old states (no explicit nonadiabatic coupling vectors were used for a projection of the velocity vectors). Since TD-DFT cannot describe conical intersections between S_1 and S_0 due to the degeneracy of the reference state (ground state) and the first excited state,⁹⁷ trajectories were forced to hop to the ground state whenever the energy difference between the states S_1 and S_0 states was smaller than 0.1 eV. Once in the ground state, no back transitions were allowed until the end of the propagation time. The data (set I) were then used for model training, validation, and testing.

Since the local diabatization scheme cannot be used with ML/MM simulations because of the unavailability of the wavefunction, a second set of QM/MM-TSH simulations was performed. In this, so-called set II, the nonadiabatic couplings between the singlet states were determined using the curvature-driven TSH scheme recently developed by Zhao *et al.*,⁹⁸ which relies on the second time derivative of the energies only and thus is accessible in conjunction with ML potentials. In this way, the resulting ML/MM-TSH dynamics are directly comparable with the reference QM/MM curvature-driven TSH ones. The initial conditions for set II of QM/MM simulations were taken from the last 2 ns of the MM trajectory, corresponding to 500 possible initial conditions (see Fig. 2). In order to investigate the transferability of the ML potential, we excite between 6 and 7 eV to have a different excitation window and thus obtain a different set of trajectories with initial conditions and energies different from those used in the training set (set I). The chosen window resulted in 66 excitations to the bright S_2 state. The 26 excitations to the S_1 were not propagated, in order to have all trajectories starting from the same state. This makes the dynamics and kinetic fits easier to interpret. All other settings were identical to those employed in set I.

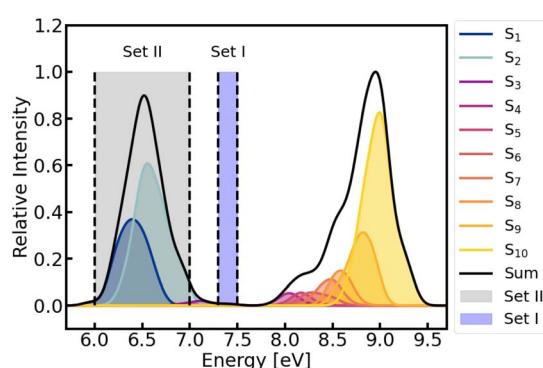


Fig. 3 Absorption spectrum of furan in water (solid black line) calculated at BP86/def2-SVP level of theory from the first 100 snapshots of the MM-MD production run. Contributions from each state are indicated by different colors. Energies and oscillator strengths were convoluted with a Gaussian with a full width at half-maximum of 0.2 eV. Shaded vertical areas denote the two energy windows chosen to initiate the QM/MM-TSH dynamics, using the snapshots of set I and II, respectively.



3.3 Machine-learning setup

To perform ML/MM simulations, we interfaced the graph convolutional neural network FieldSchNet⁵¹ with the SHARC engine,⁹⁹ as schematically depicted in Fig. 4. This neural network models atoms in its chemical and structural environment within a cutoff region. The radial cutoff for the construction of the graph around the central atom was set to 10 Å and we used 50 equidistantly spaced Gaussians in the filter-generating network,¹⁰⁰ which is used in SchNet to featurize the interatomic distances. The atomic features were updated in six message-passing layers, and the length of the atomic feature vectors was 256. Since we did not observe any hops to higher-lying states than S_4 in the QM/MM TSH simulations of set I, only the five lowest-lying singlet states were considered for learning. For the prediction of the lowest five adiabatic energies, we used one dense read-out block with four hidden layers (256-128-64-32-16-5), each halving the length of the previous layer up to a final representation of size 16, followed by a linear output layer that gives the five energies. We used the shifted softplus function¹⁰⁰ for all nonlinearities, as done in

Gastegger *et al.*⁵¹ The forces are computed as the gradient of the energy with respect to the input coordinates (eqn (5) and (6)).

Data set I comprises all points collected from the 46 QM/MM-TSH trajectories, each 330 fs long with a 0.5 fs time step (*i.e.* 46 × 330 × 2 points). However, because a few trajectories ended prematurely, the total amounts to 28 935 data points. This training set was then divided into training, validation and testing using 37, 5 and 4 trajectories, respectively. The validation and test sets were used solely for analyzing the training error on energies and forces, but not for comparing dynamics results, as this cannot be done under identical conditions. We call this approach “split by trajectory”, see Fig. 5a. Furthermore, we performed an alternative set of ML trainings, in which we randomly assigned frames from the 46 trajectories to the three different tasks (train, validation, and test) in a ratio of 80 : 10 : 10. We call this way of mixing the data “random split”, see Fig. 5c. The advantage of “split by trajectory” is that the test error is more likely to reflect the true performance, as it ensures that the model has not seen any frames of the trajectories belonging to the test set. The disadvantage is that the number of trajectories from which the model can learn is smaller than in the “random split” scheme. To further investigate the impact of data partitioning in the dynamics, we also examined the effect of subsampling in time for each partitioning scheme, as consecutive frames that are 0.5 fs apart are likely to be very similar and thus add little new information to the data set. We therefore tested the effect of using only every second (50% of data usage), third (33%), *etc.*, data point. This is schematically indicated in panels b and d of Fig. 5 for the example of 33% of data usage.

The ML trainings were performed using the Adam optimizer.¹⁰¹ We set the starting learning rate to a value of 10^{-5} , and the parameters for the exponential averages of past gradients (momentum term) and squared gradients (raw second moment) to 0.9 and 0.999, respectively. Higher initial learning rates were consistently found to cause numerical instabilities during training. A learning rate scheduler was used to adjust the learning rate. If the model's performance on the validation set did not improve for 20 consecutive epochs, the learning rate was decreased by 20% to avoid overstepping. The maximum number of epochs was set to 5000; alternatively, training was stopped early if the learning rate fell below 10^{-6} . The batch size was 10. We used the augmented mean squared error from eqn (8) as the loss function with $w_E = 1$ and $w_F = 10$.

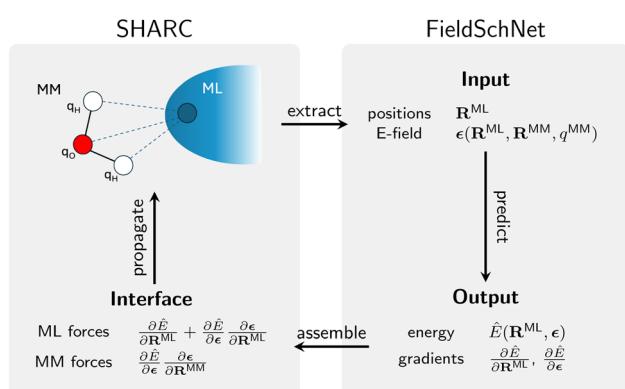


Fig. 4 Schematic representation of the communication between the driver of the nonadiabatic dynamics, the SHARC engine,⁹⁹ and the ML model, FieldSchNet, needed to perform excited state ML/MM dynamics. The SHARC interface extracts the value of the electric field at the position of every ML atom and passes both the field value and the atom positions to the ML model. The FieldSchNet model, predicts the field-dependent energies and gradients and passes them back to the SHARC interface. The system is then propagated in time by the SHARC driver. Here, none of the steps are based on file I/O further accelerating the dynamics.

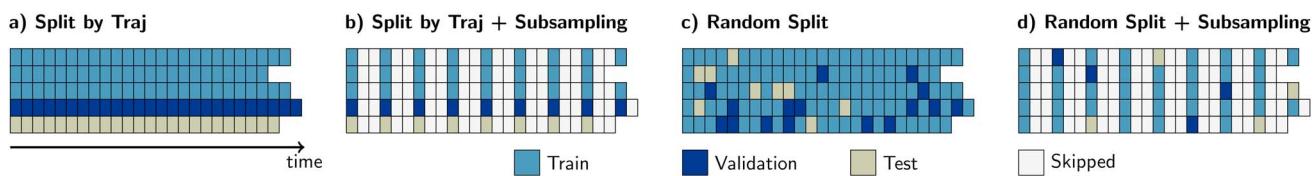


Fig. 5 Strategies to partition data from QM/MM-TSH dynamics, shown for 5 example trajectories of varying lengths. (a) Split by trajectory means that all frames belonging to one trajectory are allocated to the same subset (train, validation, or test), reducing overlap particularly between the train and test sets. (b) To avoid temporal proximity and high correlation of frames, time-based subsampling prior to splitting into train, test, and validation sets is done (exemplary shown for retaining every third frame, *i.e.* 33% data usage). (c) Random split means that the trajectory frames are pooled and randomly allocated to any of the three, train, validation, or test subtests. (d) Random split combined with time-based subsampling, here also using every third frame.



4 Results and discussion

4.1 Learning curves and test performance

The learning curves (change in test performance) of the two partitioning schemes, “random split” and “split by trajectory”, are shown in Fig. 6a, as a function of the equidistant subsampling in time. To that end, the errors of energies and forces are plotted against the amount of data included in the training, averaged over all five electronic states. For every specific combination of split and fraction of data that was retained after subsampling in time, we trained three independent models, which differ with respect to the randomly initialized parameter values (a list of the random seeds together with the data sets can be found in the Zenodo archive associated with this manuscript). As a result, no training with the same hyperparameters produces the exact same result, but is generally assumed to converge to a similarly deep minimum (see Section S2). In general, the performance of the model seems to improve as the training data becomes denser, *i.e.* with more frequent temporal sampling.

The learning curve (Fig. 6a) from the “random split” (circles) forms a nearly perfect line on a log-log graph, suggesting a power-law relationship between the sample count and the error size. This partitioning scheme often places training and testing frames in close temporal proximity. When using the entire data set, the training set probably includes the surrounding time frames of the test frames. This introduces a strong dependency between training and testing errors, raising doubts about whether a small testing error truly reflects the model’s ability to generalize (*i.e.* the power to correctly extrapolate).

In contrast, the “split by trajectory” training approach shows a different pattern. The errors initially decrease before

stabilizing at around 33% data usage, which corresponds to using every third time step. This behavior arises because the training and test sets are less correlated in this scheme. The “split by trajectory” method ensures that test errors directly evaluate the model’s generalization capabilities. Beyond a certain point, adding more closely spaced data does not further improve the test performance. This is due to minimal configuration changes within a time step of 0.5 fs, offering little additional information for extrapolation.

The discrepancy between the “split by trajectory” and random split test errors when using closely spaced trajectory frames is noteworthy. The scatter (parity) plots shown in Fig. S2 and S4[†] for the models using 100% of the training data do not exhibit any trends that would point to a general difference between those models, except that the “split by trajectory” models perform worse, which one would expect based on their relative average test performance from Fig. 6a. However, one should note that the models trained using the “split by trajectory” exhibit a much more homogeneous test performance in comparison to the “random split” models, where some appear to be much better or worse than the other two. If one were to choose settings based on Fig. 6a, then “random split” with 100% and “split by trajectory” with 33% of the data are expected to perform best.

4.2 ML/MM-TSH nonadiabatic dynamics

To further assess the ML models, we performed a second set of nonadiabatic QM/MM dynamics using curvature-driven TSH (set II), which allows us to directly compare both the reference QM/MM and ML/MM dynamics. The initial conditions for the ML/MM trajectories, including geometries, velocity vectors, and random number seeds are the same as those employed for the

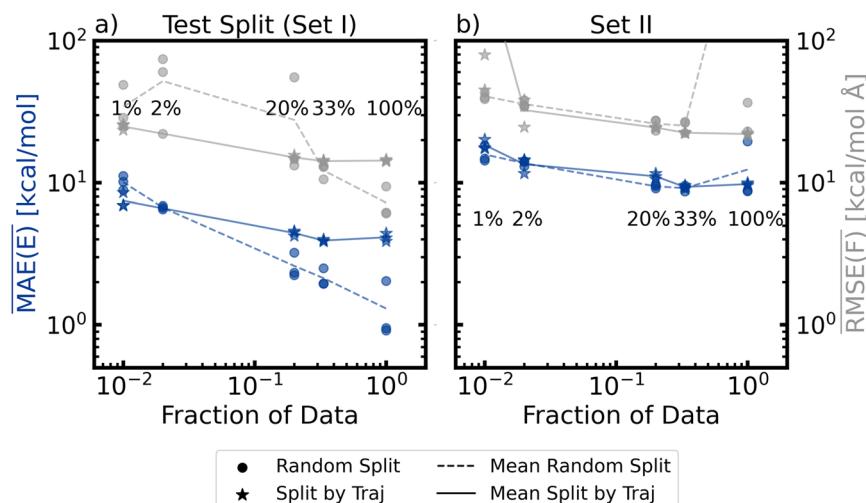


Fig. 6 Change in model test performance on (a) the test set taken from set I and (b) trajectories from set II, as a function of the fraction of data used during model training. Results of “random split” models are shown as circles, “split by trajectory” as stars. Mean absolute error (MAE) of the energy and root mean square error (RMSE) in the forces are shown in blue and grey, respectively. All values correspond to averages over all five electronic states, indicated by the bars over MAE and RMSE. The three symbols for each combination of split type and fraction of data correspond to the three models trained with different random initializations. To show the trends more clearly, two outliers in subfigure (b) have been excluded from this plot, a version that includes these points can be found in the ESI (Fig. S24).[†]



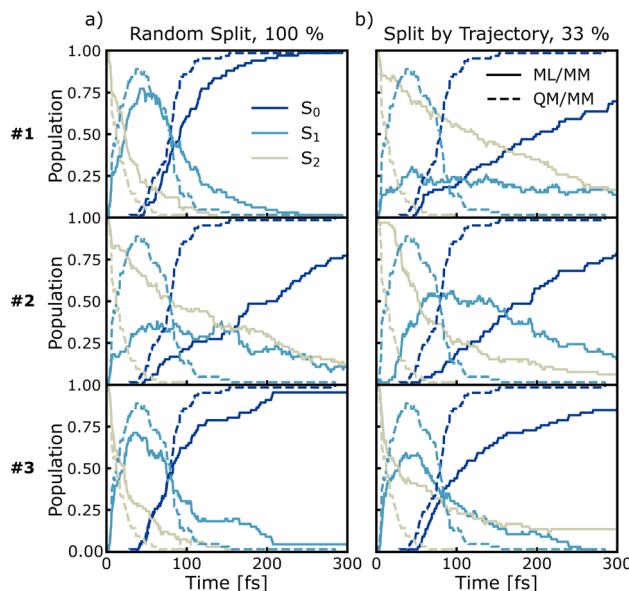


Fig. 7 Excited state occupation population dynamics of furan in water using ML/MM (solid lines) trajectories, trained with random split 100% (a) or split by trajectory 33% (b) partition sets, compared to reference QM/MM (dashed lines) trajectories. The three plots in each partition scheme correspond to three different FieldSchNet models trained with different random seeds.

QM/MM trajectories of set II. Fig. 7 shows the time-resolved electronic populations based on the active state of the reference trajectories, the QM/MM dynamics (dashed lines in each panel). Furan relaxes rapidly from the initially populated S_2 state to the S_1 state, followed by a decay to the electronic ground state S_0 . Within 75 fs about half of the trajectories in set II have relaxed to the ground state and after 300 fs nearly all 66 trajectories have reached S_0 (this is reflected in the average energy for each state, as shown in Fig. S18 of the ESI†). These populations are compared to different ML/MM TSH simulations (solid lines), conducted using the three models trained with the same splitting and subsampling settings. Specifically, we carried out simulations for 100%, 33%, and 1% for both random and trajectory split (from which Fig. 7 shows only random split with 100% and split by trajectory 33%, as these were the best hyperparameters deduced from Fig. 6a). The electronic populations for all ML/MM dynamics can be found in Fig. S16 and S17.†

As can be seen, the electronic populations derived from ML/MM show significant differences depending on which model was used to generate the random split 100% and split by trajectory 33% trajectories, even if those models differ only by their random initialization. The random split 100% models #1 and #3 produce dynamics with very similar population curves as the reference QM/MM dynamics, whereas the dynamics of model #2 show much slower internal conversions. For the split by trajectory 33% models, the visual agreement of the predicted populations increases from model #1 to #3.

The agreement of the ML/MM electronic populations with the corresponding QM/MM reference cannot easily be explained

with the test statistics presented in the previous section. Inspection of the parity plots for the energy gap between neighboring PESs (Fig. S2–S4 of the ESI†) reveals that the split by trajectory models appear to be worse at predicting this energy difference. This is problematic, as the energy gap is used to force hops into the ground state from S_1 . To probe the model performance in regions with small energy gaps further we employ an energy-gap-weighted error measure,

$$\text{wRMSE}(X^j) = \sqrt{\frac{\sum_i^{N_{\text{frames}}} (X_i^j - \hat{X}_i^j)^2 w_i^j}{\sum_i^{N_{\text{frames}}} w_i^j}} \quad (11)$$

with $w_i^j = \frac{1}{\min(|E_i^j - E_i^{j-1}|, |E_i^j - E_i^{j+1}|)}$,

where E_i^j represents the ground truth energy for configuration i in state j and X_i^j denotes either energy or forces for this configuration and state.

To facilitate a more effective comparison across the various models, in addition to visually evaluating the population curves, we fitted a kinetic model to the different steps of the relaxation process. This model includes only two time constants: $\tau_{2 \rightarrow 1}$ for the internal conversion from S_2 to S_1 and $\tau_{1 \rightarrow 0}$ for the transition from S_1 to S_0 . The QM/MM reference simulations predict the first internal conversion to be about five times faster ($\tau_{2 \rightarrow 1}^{\text{QM/MM}} = 16.8$ fs) than the second one ($\tau_{1 \rightarrow 0}^{\text{QM/MM}} = 64.9$ fs). These time constants are comparable to those predicted in the gas phase (9.2 fs and 60 fs, for the lifetimes of the S_2 and S_1 states, respectively), by Fuji *et al.*⁶³ using similar TSH simulations and TD-DFT, which in turn are also consistent with time-resolved photoelectron spectra, recorded by the same authors. The similarity of the time constants indicates that the effect of the solvent is not very pronounced, slightly decelerating the decay of the bright S_2 state.

The different numerical values of the kinetic fits and (weighted) errors, obtained by the ML/MM simulations are collected in Table 1. We observe that none of the ML/MM models produces dynamics that relax as fast as observed in the QM/MM simulations. Furthermore, the relative error for $\tau_{2 \rightarrow 1}$ is almost always larger than for $\tau_{1 \rightarrow 0}$, because of their difference in magnitude. Intriguingly, the kinetics of the ML/MM simulations can differ greatly even for those models that have the same training settings and only differ by weight initialization. In general, models trained using the random split appear to perform better, especially when the training frames are taken at smaller intervals. The reason for this is likely the small set of trajectories available for training, such that splitting by trajectory imposed a stronger limit on the phase space available for training than the random splitting. More trajectories in the training set are expected to remove this trend. Indeed, random split with 100% of the training data delivers the best results.

While it is generally a problem to have a confidence measure for ML-based MD simulations where the behavior of the system is not known, it is gratifying to see that the weighted error, as specified in eqn (11), serves as a metric to predict which of two



Table 1 Time constants for the sequential $S_2 \rightarrow S_1$ and $S_1 \rightarrow S_0$ internal conversions of furan in water, as obtained from ML/MM models trained with 100%, 33%, and 1% of the data from set I (subsampled in time) within the random split and trajectory split schemes. The simulations were started from the initial conditions of set II and are compared with the QM/MM trajectories of this set. The test errors in energies (E) and forces (F) are averaged over the five electronic states (S_0 – S_4). The weighted RMSEs were computed according to eqn (11). All the trainings were performed with the augmented loss. The time constants obtained from the reference QM/MM curvature-driven TSH simulations are $\tau_{2 \rightarrow 1}^{\text{QM/MM}} = 17 \pm 2$ fs and $\tau_{1 \rightarrow 0}^{\text{QM/MM}} = 65 \pm 3$ fs

| Split type | % of data | Model # | $\tau_{2 \rightarrow 1}$ [fs] | $\tau_{1 \rightarrow 0}$ [fs] | $\overline{\text{RMSE}}(E)$ [kcal mol $^{-1}$] | $\overline{\text{wRMSE}}(E)$ [kcal mol $^{-1}$] | $\overline{\text{RMSE}}(F)$ [kcal mol $^{-1}$ Å $^{-1}$] | $\overline{\text{wRMSE}}(F)$ [kcal mol $^{-1}$ Å $^{-1}$] |
|------------|-----------|---------|-------------------------------|-------------------------------|--|---|--|---|
| Random | 100 | 1 | 29 ± 4 | 73 ± 4 | 1.4 | 1.4 | 5.9 | 56.3 |
| | | 2 | 138 ± 28 | 92 ± 12 | 2.7 | 3.2 | 9.1 | 76.2 |
| | | 3 | 28 ± 5 | 73 ± 6 | 1.3 | 1.3 | 6.0 | 54.9 |
| | 33 | 1 | 63 ± 16 | 90 ± 8 | 3.3 | 3.8 | 10.8 | 95.3 |
| | | 2 | 36 ± 8 | 97 ± 8 | 2.6 | 2.7 | 9.8 | 69.5 |
| | | 3 | 66 ± 11 | 84 ± 6 | 5.0 | 5.8 | 19.0 | 126.0 |
| | 1 | 1 | 240 ± 41 | 164 ± 29 | 28.5 | 22.9 | 26.0 | 168.0 |
| | | 2 | 134 ± 18 | 130 ± 16 | 9.5 | 10.0 | 23.6 | 126.7 |
| | | 3 | 97 ± 10 | 116 ± 14 | 11.1 | 11.3 | 23.2 | 124.9 |
| By traj. | 100 | 1 | 153 ± 26 | 95 ± 10 | 6.3 | 6.3 | 13.7 | 81.7 |
| | | 2 | 37 ± 6 | 251 ± 33 | 6.0 | 6.0 | 14.1 | 83.9 |
| | | 3 | 292 ± 61 | 109 ± 17 | 6.0 | 5.9 | 13.9 | 82.6 |
| | 33 | 1 | 67 ± 17 | 67 ± 6 | 5.6 | 5.6 | 13.6 | 80.4 |
| | | 2 | 86 ± 11 | 134 ± 18 | 5.9 | 6.0 | 14.1 | 82.7 |
| | | 3 | 173 ± 32 | 76 ± 9 | 5.7 | 5.8 | 14.0 | 82.0 |
| | 1 | 1 | 231 ± 23 | 70 ± 11 | 10.0 | 10.7 | 23.2 | 122.8 |
| | | 2 | 393 ± 70 | 100 ± 22 | 10.0 | 11.2 | 24.8 | 132.2 |
| | | 3 | 292 ± 51 | 109 ± 18 | 12.1 | 13.4 | 25.4 | 132.6 |

models with identical hyperparameters will perform better. Comparing the normal and weighted RMSE, it can be seen that the energy errors change by a maximum of 1 kcal mol $^{-1}$, except for the models trained with only 1% of the data. We can therefore assume that energy values close to the intersection seams are learned with a similar accuracy as points further away from these important regions. However, the same does not appear to be true for the gradients of the PESs. The increase from non-weighted to energy-gap-weighted RMSE varies, but is approximately a factor of 5 to 10. Weighted energy and force RMSEs taken together suggest that the distance between the different electronic PESs is roughly correct; however, the topology close to the avoided crossings is not. We therefore conclude that the general feature of the avoided crossing is represented correctly, but that the individual ML-PESs are significantly less smooth in these regions, leading to a strong increase in the force errors, albeit not in the energies. Surprisingly, a model with a weighted force RMSE of 50 kcal mol $^{-1}$ Å $^{-1}$ predicts the relaxation dynamics qualitatively correctly (random split, 100%, models no. 1 and 3). Since the weighted error decreases with larger training set sizes, increasing the training set size even further should lead to smaller errors. Focusing on the correct reconstruction of the slope near the intersection seams seems to be especially important. This interpretation is further supported by test calculations where we used the time-derivative of the gradients instead of the energies, which lead to significantly decreased agreement between the ground truth and the ML models. Hence, the computed ML transition rates have the correct order of magnitude, because the couplings that were used in our simulations only depend on the energies and not on the gradients.

To further understand the performance of the different models, we also performed an error analysis regarding energies, gaps, and forces for the data in set II. Parity plots for set II are shown in Section S2.2 of the ESI† and the average errors in Fig. 6b. One can see that on this data – unseen by all models – “random split” and “split by trajectory” models perform similarly. This underlines that test statistics based on the random splitting (see Fig. 6a) provide a strong underestimation of the true error. Furthermore, the parity plots of Section S2.2 in the ESI† show that all models (random and trajectory split) appear to strongly overestimate E_{S_0} for configurations with a ground state energy above 75 kcal mol $^{-1}$ (likely geometries with strongly elongated or broken bonds), which leads to a significant underestimation of the energy gap to the first excited state in these regions. It may appear confusing that the models tend to underestimate this gap when their dynamics is always slower than the QM/MM reference. However, this only happens for configurations with a high ground-state energy, which are those occurring after relaxation to the ground state, and are then prevented from hopping back up (see discussion in Section S3.2 and Fig. S18 of the ESI†).

In general, the errors of the models based on the set II are so large that it is surprising that they can produce reasonable population decays. These errors are based on geometries from the entire 300 fs, where furan displayed ring opening and subsequent bond rearrangements (discussed in detail in Section S4 of the ESI†), which might be difficult to describe with (TD-)DFT. The initial part of the dynamics is much better reproduced by the ML/MM than the latter parts, since it corresponds to regions in configuration space that are reasonably described by (TD-)DFT. To test this hypothesis, we performed



a second round of error analysis on set II, where we only included frames from the first 75 fs. The much better performance of the models on these configurations becomes apparent when looking at the parity plots in Section S2.3 of the ESI.† The errors (MAE and RMSE) of forces, energies, and energy gaps are significantly lower than those on the entire set II. For split by trajectory they are basically identical to the test statistics obtained from set I. This means that the models fit the PESs well in the part of the configuration space that is explored right after irradiation, which is the part needed for the relaxation back to the electronic ground state. The subsequent distortions due to the excess energy of 6–7 eV, are not well described, but these deficiencies are less relevant to the change in electronic populations. This is why most models were able to reproduce the relaxation dynamics of set II even though they cannot extrapolate correctly to the configurations visited in later stages of the QM/MM simulations. We want to acknowledge, however, that FieldSchNet⁵¹ is based on the invariant SchNet architecture.⁷² Hence, we expect an equivariant ML model capable of handling external charges¹⁰² to be more robust, just as such models are in the usual field-independent setting.¹⁰³

4.3 Structural analysis of trajectories

In order to analyze whether the ML/MM simulations show the same structural changes during the dynamics as the reference

QM/MM simulations, we compare the hopping geometries from the QM/MM simulations with those encountered in the ML/MM simulations, here done exemplary for model #1 with 100% of the data and the random split procedure (for a comparison of all 18 ML/MM models, see Section S4.2 of the ESI†). The hopping geometries from the QM/MM simulations (set II) for the S₂–S₁ and S₁–S₀ transitions are shown in Fig. 8a and b, respectively. The geometries responsible of the first internal conversion (Fig. 8a) are very similar because the hops occur shortly after excitation, so they closely resemble the ground-state MM geometries of the Franck–Condon ensemble. The geometries corresponding to the S₁–S₀ deactivation (Fig. 8b) are more diverse. They encompass a smooth interpolation between configurations with a closed and an open ring. For easier visualization, the geometries were aligned so that only one bond appears to open, however, both bonds C₁–O and C₄–O (following the naming convention of Fig. 1b) are equally likely to break.

To analyze the similarities of the hopping structures found in the QM/MM and ML/MM simulations, we performed a principal component analysis (PCA) on the set of QM/MM geometries for the transitions between S₁ and S₀ using a Coulomb matrix representation¹⁰⁴ of furan (Section S4 of the ESI†). We found that the first principal component (PC1), which recovers about 89% of the overall variance, focuses on the distance between the oxygen atom and carbon atoms 1 and 4, whereas

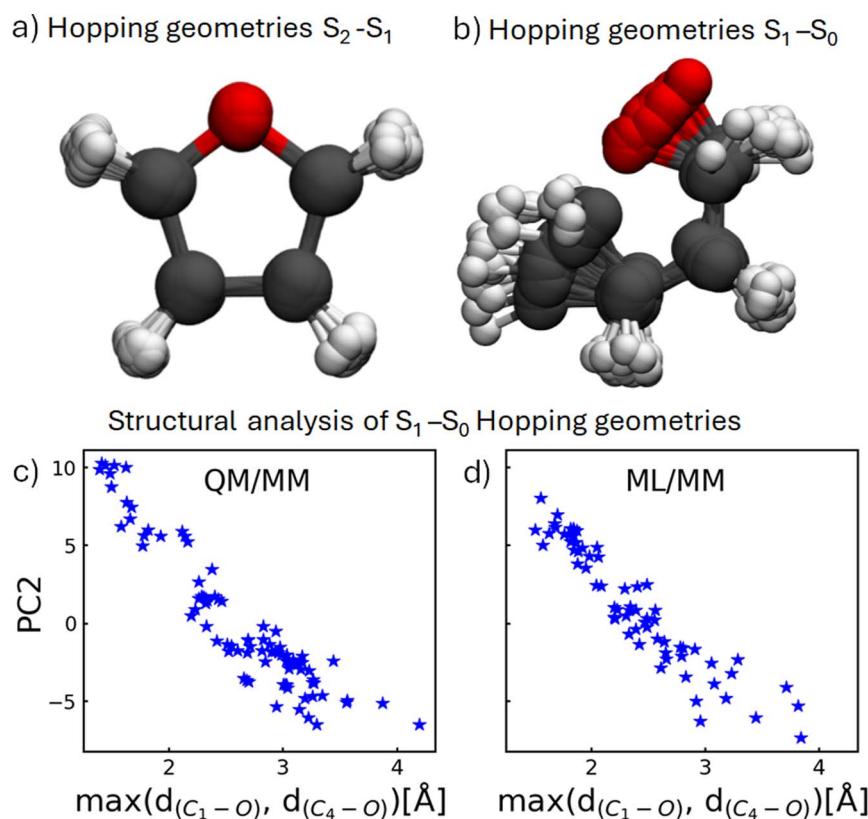


Fig. 8 Aligned QM/MM S₂–S₁ (a) and S₁–S₀ (b) hopping geometries obtained from trajectories in set II. Comparison of S₁–S₀ hopping geometries obtained from QM/MM (c) and ML/MM (d) (random split, 100% data, model #1) by projecting them onto the maximum of either the C₁–O and C₄–O bond distance (see Fig. 1b) as well as the PC2 from a PCA performed solely on the QM/MM frames (Section S5).



PC2, which reflects 9% of the variance, is a linear combination of several interatomic distances. In Fig. 8c we replaced PC1 with the maximum length of the two bonds C₁–O and C₄–O for easier visualization, as PC1 creates an inverted V due to the symmetry of the breaking bonds (see Fig. S19 in Section S4 of the ESI†). One can see that the geometries form one continuous hopping seam.

When applying the same linear transformation to the hopping geometries of the ML/MM simulation (here shown for random split, 100%, model #1) and projecting them onto PC2 and the maximum of the C–O bond distances, they form the same diagonal line, indicating that hops occur in the same region of configuration space. This is not only true for the single model analyzed here, but for most models (Fig. S20 in Section S4.2.1 of the ESI†). Deviations from the rather even distribution seen in Fig. 8c correlate with significantly different relaxation dynamics, *e.g.* some models hop only at certain points along the seam creating visual clusters in the 2D-projection. As the S₂–S₁ hopping geometries are all very similar, the projection onto the first two principal components forms a single cluster (Fig. S21 in Section S4.2.2 of the ESI†). Trajectories with slow internal conversion from S₂ to S₁ show outliers in the 2D-projection (Section S4.2.2 and Fig. S22 of the ESI†).

5 Conclusions

We have integrated the FieldSchNet machine learning interatomic potential into the SHARC software package to enable nonadiabatic ML/MM dynamics simulations in an electrostatic embedding framework, in analogy to the traditional QM/MM counterpart. By developing a training loss function that incorporates the consistent gradients required during simulation, we ensured the inclusion of electric field-dependent components in the nuclear forces, enhancing the accuracy of our approach. Our method was applied to furan in water, trained with almost 30 000 data points obtained from QM/MM nonadiabatic trajectories at the BP86/def2-SVP level of theory. We expect our ML/MM setup to be directly transferable to a boundary across chemical bonds, since link atoms can be used analogously to QM/MM simulations. Moreover, the SHARC software package can already handle QM/MM simulations with link atoms.

We compared the training errors derived from using two distinct data splitting strategies – “random sampling” and “split by trajectory”, depending on whether the data points were sampled randomly from any trajectory, or entire trajectories are used for training, validation and testing. Even though its performance is limited by the smaller number of available trajectories, we recommend to use “split by trajectory”, as it offers cleaner data separation and more realistic error statistics. In general, we observe a wide range in the performance of the ML/MM models when trying to reproduce the nonadiabatic QM/MM dynamics of a set of held out trajectories, even for ML models with the same hyper-parameters. Strong sensitivities of excited state kinetics generated by ML models were already noted in earlier studies.⁴¹ However, well-chosen test statistics

served as reliable indicators of model performance. Energy gap-weighted errors help to highlight discrepancies near the intersection seams. Projecting hopping geometries onto two dimensions provides valuable insights to understand the difference between ground truth dynamics and those derived from the ML model.

Based on our findings, we recommend against using costly nonadiabatic QM/MM dynamics simulations to generate training data, where a single time step of the nonadiabatic QM/MM dynamics takes already 2 minutes using (TD)-DFT for the presented test system. A study that goes beyond a proof of concept would likely use a more expensive level of theory that handles bond dissociations correctly. Most computational effort is spent on geometries that are far away from critical regions of the PESS (the intersection seams). Furthermore, the produced geometries are highly correlated as they are closely spaced in time. Future applications should therefore aim to collect training data through active learning schemes,^{105,106} which has the added benefit that such data are not correlated, which means that no computational effort is wasted obtaining frames that might later be discarded by subsampling in time. We expect models trained on such data to generate better forces near the intersection seams and to reproduce the dynamics of the QM method more reliably.

Data and code availability

Data, code, and processing scripts for this paper, including trained machine learning models, SHARC input scripts, initial conditions, and random seeds for reproducibility, are available on Zenodo at <https://doi.org/10.5281/zenodo.14536035>. Additionally, the archive also contains code to interface FieldSchNet with an upcoming release of SHARC, which is published under the GPL license at <https://github.com/sharemd/>.

Author contributions

MXT (data curation, formal analysis, investigation, methodology, software, visualization, validation, writing – original draft). BB (software and writing – review). CGC (investigation). JW (methodology, writing – review). PM (conceptualization, methodology, supervision, funding acquisition) JCBD (methodology, validation, supervision, visualization, writing – original draft). LG (conceptualization, supervision, funding acquisition, project administration, writing – review).

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work is funded from the University of Vienna in the framework of the research platform ViRAPID. M. X. T. and L. G. appreciate additional support provided by the Austrian Science Fund, W 1232 (MolTag). The Vienna Scientific Cluster is



thanked for generous allocation of computer resources. The authors thank the SHARC development team, the ViRAPID members, and Michael Gastegger for fruitful discussions.

Notes and references

- D. R. Ort and C. F. Yocom, in *Electron Transfer and Energy Transduction in Photosynthesis: An Overview*, Kluwer Academic Publishers, 1996, pp. 1–9.
- J. H. A. Nugent, *Eur. J. Biochem.*, 1996, **237**, 519–531.
- A. Stirbet, D. Lazár, Y. Guo and G. Govindjee, *Ann. Bot.*, 2019, **126**, 511–537.
- D. Gust, T. A. Moore and A. L. Moore, *Acc. Chem. Res.*, 2009, **42**, 1890–1898.
- X. Huang and T. Li, *J. Mater. Chem. C*, 2020, **8**, 821–848.
- M. Butt, S. Khonina and N. Kazanskiy, *Opt Laser. Technol.*, 2021, **142**, 107265.
- S. Mai and L. González, *Angew. Chem., Int. Ed.*, 2020, 2–17.
- Quantum Chemistry and Dynamics of Excited States: Methods and Applications*, ed. R. Lindh and L. González, John Wiley & Sons, 2020.
- J. C. Tully, *J. Chem. Phys.*, 1990, **92**, 1061–1071.
- M. Barbatti, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 620–633.
- S. Mai, P. Marquetand and L. González, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8**, 1–23.
- J. J. Nogueira and L. González, *Annu. Rev. Phys. Chem.*, 2018, **69**, 473–497.
- S. Decherchi, M. Masetti, I. Vyalov and W. Rocchia, *Eur. J. Med. Chem.*, 2015, **91**, 27–42.
- H. M. Senn and W. Thiel, *Angew. Chem., Int. Ed.*, 2009, **48**, 1198–1229.
- G. Groenhof, in *Introduction to QM/MM Simulations*, Humana Press, 2012, pp. 43–66.
- H. Lin, *QM/MM Methods*, American Chemical Society, Washington, DC, USA, 2024.
- E. Brunk and U. Rothlisberger, *Chem. Rev.*, 2015, **115**, 6217–6263.
- C. Schnedermann, X. Yang, M. Liebel, K. M. Spillane, J. Lugtenburg, I. Fernández, A. Valentini, I. Schapiro, M. Olivucci, P. Kukura and R. A. Mathies, *Nat. Chem.*, 2018, **10**, 449–455.
- X. Yang, M. Manathunga, S. Gozem, J. Léonard, T. Andruniów and M. Olivucci, *Nat. Chem.*, 2022, **14**, 441–449.
- M. Bondanza, B. Demoulin, F. Lipparini, M. Barbatti and B. Mennucci, *J. Phys. Chem. A*, 2022, **126**, 6780–6789.
- J. M. Toldo, M. T. do Casal, E. Ventura, S. A. do Monte and M. Barbatti, *Phys. Chem. Chem. Phys.*, 2023, **25**, 8293–8316.
- A. J. Coffman, Z. Jin, J. Chen, J. E. Subotnik and D. V. Cofer-Shabica, *J. Chem. Theory Comput.*, 2023, **19**, 7136–7150.
- C. M. Marian, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 187–203.
- F. Plasser, S. Gómez, M. F. Menger, S. Mai and L. González, *Phys. Chem. Chem. Phys.*, 2019, **21**, 57–69.
- J. P. Zobel, M. Heindl, F. Plasser, S. Mai and L. González, *Acc. Chem. Res.*, 2021, **54**, 3760–3771.
- S. Polonius, D. Lehrner, L. González and S. Mai, *J. Chem. Theory Comput.*, 2024, **20**, 4738–4750.
- S. Polonius, D. Lehrner, L. González and S. Mai, *J. Chem. Theory Comput.*, 2024, **20**, 4738–4750.
- P. O. Dral, *J. Phys. Chem. Lett.*, 2020, **11**, 2336–2347.
- F. Noé, A. Tkatchenko, K.-R. Müller and C. Clementi, *Annu. Rev. Phys. Chem.*, 2020, **71**, 361–390.
- J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K. R. Müller and A. Tkatchenko, *Chem. Rev.*, 2021, **121**, 9816–9872.
- O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko and K. R. Müller, *Chem. Rev.*, 2021, **121**, 10142–10186.
- Y. Yang, S. Zhang, K. D. Ranasinghe, O. Isayev and A. E. Roitberg, *Annu. Rev. Phys. Chem.*, 2024, **75**, 371–395.
- J. Behler, *Angew. Chem., Int. Ed.*, 2017, **56**, 12828–12840.
- Y. Yang, L. Zhao, C.-X. Han, X.-D. Ding, T. Lookman, J. Sun and H.-X. Zong, *J. Mater. Inf.*, 2021, **1**, 10.
- Y. Guan, C. Xie, D. R. Yarkony and H. Guo, *Phys. Chem. Chem. Phys.*, 2021, **23**, 24962–24983.
- H. Dong, Y. Shi, P. Ying, K. Xu, T. Liang, Y. Wang, Z. Zeng, X. Wu, W. Zhou, S. Xiong, S. Chen and Z. Fan, *J. Appl. Phys.*, 2024, **135**, 161101.
- I. Batatia, P. Benner, Y. Chiang, A. M. Elena, D. P. Kovács, J. Riebesell, X. R. Advincula, M. Asta, M. Avaylon, W. J. Baldwin, F. Berger, N. Bernstein, A. Bhowmik, S. M. Blau, V. Cărare, J. P. Darby, S. De, F. D. Pia, V. L. Deringer, R. Elijósius, Z. El-Machachi, F. Falcioni, E. Fako, A. C. Ferrari, A. Genreith-Schriever, J. George, R. E. A. Goodall, C. P. Grey, P. Grigorev, S. Han, W. Handley, H. H. Heenen, K. Hermansson, C. Holm, J. Jaafar, S. Hofmann, K. S. Jakob, H. Jung, V. Kapil, A. D. Kaplan, N. Karimitari, J. R. Kermode, N. Kroupa, J. Kullgren, M. C. Kuner, D. Kuryla, G. Liepuoniute, J. T. Margraf, I.-B. Magdău, A. Michaelides, J. H. Moore, A. A. Naik, S. P. Niblett, S. W. Norwood, N. O'Neill, C. Ortner, K. A. Persson, K. Reuter, A. S. Rosen, L. L. Schaaf, C. Schran, B. X. Shi, E. Sivonxay, T. K. Stenczel, V. Svahn, C. Sutton, T. D. Swinburne, J. Tilly, C. van der Oord, E. Varga-Umbrich, T. Vegge, M. Vondrák, Y. Wang, W. C. Witt, F. Zills and G. Csányi, A foundation model for atomistic materials chemistry, *arXiv*, 2024, preprint, arXiv:2401.00096, DOI: [10.48550/arXiv.2401.00096](https://doi.org/10.48550/arXiv.2401.00096), <https://arxiv.org/abs/2401.00096>.
- O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. Medrano Sandonas, J. T. Berryman, A. Tkatchenko and K.-R. Müller, *Sci. Adv.*, 2024, **10**, eadn4397.
- P. O. Dral, M. Barbatti and W. Thiel, *J. Phys. Chem. Lett.*, 2018, **9**, 5660–5663.
- J. Westermayr, M. Gastegger, M. F. Menger, S. Mai, L. González and P. Marquetand, *Chem. Sci.*, 2019, **10**, 8100–8107.
- J. Westermayr, F. A. Faber, A. S. Christensen, O. A. von Lilienfeld and P. Marquetand, *Mach. Learn.: Sci. Technol.*, 2020, **1**, 025009.



42 J. Westermayr and P. Marquetand, *J. Chem. Phys.*, 2020, **153**, 154112.

43 P. O. Dral and M. Barbatti, *Nat. Rev. Chem.*, 2021, **5**, 388–405.

44 J. Westermayr, M. Gastegger, D. Vörös, L. Panzenboeck, F. Joerg, L. González and P. Marquetand, *Nat. Chem.*, 2022, **14**, 914–919.

45 S. Axelrod, E. Shakhnovich and R. Gómez-Bombarelli, *Nat. Commun.*, 2022, **13**, 1–11.

46 J. Westermayr and P. Marquetand, *Chem. Rev.*, 2020, **121**, 9873–9926.

47 K. Yao, J. E. Herr, D. Toth, R. Mckintyre and J. Parkhill, *Chem. Sci.*, 2018, **9**, 2261–2269.

48 J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev and A. E. Roitberg, *Nat. Commun.*, 2019, **10**, 2903.

49 T. W. Ko, J. A. Finkler, S. Goedecker and J. Behler, *Nat. Commun.*, 2021, **12**, 398.

50 B. Lier, P. Poliak, P. Marquetand, J. Westermayr and C. Oostenbrink, *J. Phys. Chem. Lett.*, 2022, **13**, 3812–3818.

51 M. Gastegger, K. T. Schütt and K.-R. Müller, *Chem. Sci.*, 2021, **12**, 11473–11483.

52 A. Hofstetter, L. Bösel and S. Riniker, *Phys. Chem. Chem. Phys.*, 2022, **24**, 22497–22512.

53 K. Zinovjev, *J. Chem. Theory Comput.*, 2023, **19**, 1888–1897.

54 M. Thürlemann and S. Riniker, *Chem. Sci.*, 2023, **14**, 12661–12675.

55 K. Zinovjev, L. Hedges, R. Montagud Andreu, C. Woods, I. Tuñón and M. W. van der Kamp, *J. Chem. Theory Comput.*, 2024, **20**, 4514–4522.

56 J. Semelak, P. Ignacio, K. Huddleston, J. Olmos, J. S. Grassano, C. Clemente, S. Drusin, M. Martí, M. Lebrero, A. Roitberg and D. Estrin, *ANI Neural Networks Meet Electrostatics: A ML/MM Implementation in Amber*, 2024.

57 J. S. Grassano, I. Pickering, A. E. Roitberg, M. C. González Lebrero, D. A. Estrin and J. A. Semelak, *J. Chem. Inf. Model.*, 2024, **64**, 4047–4058.

58 Y.-K. Lei, K. Yagi and Y. Sugita, *J. Chem. Phys.*, 2024, **160**, 214109.

59 P. Mazzeo, E. Cignoni, A. Arcidiacono, L. Cupellini and B. Mennucci, *Digital Discovery*, 2024, 2560–2571.

60 J. Roebber, D. Gerrity, R. Hemley and V. Vaida, *Chem. Phys. Lett.*, 1980, **75**, 104–106.

61 M. Pastore, C. Angeli and R. Cimiraglia, *Chem. Phys. Lett.*, 2006, **426**, 445–451.

62 N. Gavrilov, S. Salzmann and C. M. Marian, *Chem. Phys.*, 2008, **349**, 269–277.

63 T. Fuji, Y.-I. Suzuki, T. Horio, T. Suzuki, R. Mitić, U. Werner and V. Bonačić-Koutecký, *J. Chem. Phys.*, 2010, **133**, 234303.

64 M. Stenrup and Å. Larson, *Chem. Phys.*, 2011, **379**, 6–12.

65 E. V. Gromov, C. Lévéque, F. Gatti, I. Burghardt and H. Köppel, *J. Chem. Phys.*, 2011, **135**, 164305.

66 E. V. Gromov, V. S. Reddy, F. Gatti and H. Köppel, *J. Chem. Phys.*, 2013, **139**, 234306.

67 R. Spesyvtsev, T. Horio, Y.-I. Suzuki and T. Suzuki, *J. Chem. Phys.*, 2015, **143**, 014302.

68 S. Oesterling, O. Schalk, T. Geng, R. D. Thomas, T. Hansson and R. de Vivie-Riedle, *Phys. Chem. Chem. Phys.*, 2017, **19**, 2025–2035.

69 T. Brupbacher, J. Makarewicz and A. Bauder, *J. Chem. Phys.*, 1998, **108**, 3932–3939.

70 S. P. Lockwood, T. G. Fuller and J. J. Newby, *J. Phys. Chem. A*, 2018, **122**, 7160–7170.

71 D. Bakowies and W. Thiel, *J. Phys. Chem.*, 1996, **100**, 10580–10594.

72 K. T. Schütt, H. E. Saucedo, P. J. Kindermans, A. Tkatchenko and K. R. Müller, *J. Chem. Phys.*, 2018, **148**, 241722.

73 J. C. Tully and R. K. Preston, *J. Chem. Phys.*, 1971, **55**, 562–572.

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

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

76 G. Granucci, M. Persico and A. Toniolo, *J. Chem. Phys.*, 2001, **114**, 10608–10615.

77 F. Plasser, G. Granucci, J. Pittner, M. Barbatti, M. Persico and H. Lischka, *J. Chem. Phys.*, 2012, **137**, 22A514.

78 F. Plasser, M. Ruckenbauer, S. Mai, M. Oppel, P. Marquetand and L. González, *J. Chem. Theory Comput.*, 2016, **12**, 1207–1219.

79 Y. Shu, L. Zhang, X. Chen, S. Sun, Y. Huang and D. G. Truhlar, *J. Chem. Theory Comput.*, 2022, **18**, 1320–1328.

80 A. W. Jasper and D. G. Truhlar, *Chem. Phys. Lett.*, 2003, **369**, 60–67.

81 A. W. Jasper, M. D. Hack and D. G. Truhlar, *J. Chem. Phys.*, 2001, **115**, 1804–1816.

82 D. A. Case, H. M. Aktulga, K. Belfon, D. S. Cerutti, G. A. Cisneros, V. W. D. Cruzeiro, N. Forouzesh, T. J. Giese, A. W. Götz, H. Gohlke, S. Izadi, K. Kasavajhala, M. C. Kaymak, E. King, T. Kurtzman, T.-S. Lee, P. Li, J. Liu, T. Luchko, R. Luo, M. Manathunga, M. R. Machado, H. M. Nguyen, K. A. O’Hearn, A. V. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, A. Risheh, S. Schott-Verdugo, A. Shajan, J. Swails, J. Wang, H. Wei, X. Wu, Y. Wu, S. Zhang, S. Zhao, Q. Zhu, T. E. Cheatham, D. R. Roe, A. Roitberg, C. Simmerling, D. M. York, M. C. Nagan and K. M. Merz, *J. Chem. Inf. Model.*, 2023, **63**, 6183–6191.

83 M. J. S. Dewar, E. G. Zoebisch, E. F. Healy and J. J. P. Stewart, *J. Am. Chem. Soc.*, 1985, **107**, 3902–3909.

84 A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623–1641.

85 J.-P. Ryckaert, G. Ciccotti and H. J. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327–341.

86 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

87 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J.-P. Piquemal, P. Ren and J. W. Ponder, *J. Chem. Theory Comput.*, 2018, **14**, 5273–5289.

88 H. C. Andersen, *J. Comput. Phys.*, 1983, **52**, 24–34.

89 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.



90 J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **33**, 8822–8824.

91 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

92 F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.

93 F. Neese, F. Wennmohs, U. Becker and C. Ripplinger, *J. Chem. Phys.*, 2020, **152**, 224108.

94 W. Hieringer, S. J. A. van Gisbergen and E. J. Baerends, *J. Phys. Chem. A*, 2002, **106**, 10380–10390.

95 K. Kamada, M. Ueda, T. Sakaguchi, K. Ohta and T. Fukumi, *J. Opt. Soc. Am. B*, 1998, **15**, 838–845.

96 M. Barbatti and H. Lischka, *J. Phys. Chem. A*, 2007, **111**, 2852–2858.

97 F. Plasser, R. Crespo-Otero, M. Pederzoli, J. Pittner, H. Lischka and M. Barbatti, *J. Chem. Theory Comput.*, 2014, **10**, 1395–1405.

98 X. Zhao, I. C. D. Merritt, R. Lei, Y. Shu, D. Jacquemin, L. Zhang, X. Xu, M. Vacher and D. G. Truhlar, *J. Chem. Theory Comput.*, 2023, **19**, 6577–6588.

99 S. Mai, D. Avagliano, M. Heindl, P. Marquetand, M. F. S. J. Menger, M. Oppel, F. Plasser, S. Polonius, M. Ruckenbauer, Y. Shu, D. G. Truhlar, L. Zhang, P. Zobel and L. González, SHARC3.0: Surface Hopping Including Arbitrary Couplings—Program Package for Non-Adiabatic Dynamics, 2023, <https://sharc-md.org/>.

100 K. T. Schütt, P. J. Kindermans, H. E. Saucedo, S. Chmiela, A. Tkatchenko and K. R. Müller, *Adv. Neural Inf. Process. Syst.*, 2017, 992–1002.

101 D. P. Kingma and J. Ba, Adam: A Method for Stochastic Optimization, *arXiv*, 2017, preprint, arXiv:1412.6980, DOI: [10.48550/arXiv.1412.6980](https://arxiv.org/abs/1412.6980), <https://arxiv.org/abs/1412.6980>.

102 R. Barrett, J. C. B. Dietschreit and J. Westermayr, Incorporating Long-Range Interactions via the Multipole Expansion into Ground and Excited-State Molecular Simulations, *arXiv*, 2025, preprint, arXiv:2502.21045, DOI: [10.48550/arXiv.2502.21045](https://arxiv.org/abs/2502.21045), <https://arxiv.org/abs/2502.21045>.

103 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nat. Commun.*, 2022, **13**, 2453.

104 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 1–5.

105 B. W. J. Chen, X. Zhang and J. Zhang, *Chem. Sci.*, 2023, **14**, 8338–8354.

106 A. R. Tan, J. C. B. Dietschreit and R. Gómez-Bombarelli, *J. Chem. Phys.*, 2025, **162**, 034114.

