



# Charting electronic-state manifolds across molecules with multi-state learning and gap-driven dynamics via efficient and robust active learning



Mikołaj Martyka<sup>1</sup>, Lina Zhang<sup>2</sup>, Fuchun Ge<sup>2</sup>, Yi-Fan Hou<sup>2</sup>, Joanna Jankowska<sup>1</sup>✉, Mario Barbatti<sup>3,4</sup>✉ & Pavlo O. Dral<sup>1,5,6</sup>✉

We present a robust protocol for affordable learning of electronic states to accelerate photophysical and photochemical molecular simulations. The protocol solves several issues precluding the widespread use of machine learning (ML) in excited-state simulations. We introduce a novel physics-informed multi-state ML model that can learn an arbitrary number of excited states across molecules, with accuracy better or similar to the accuracy of learning ground-state energies, where information on excited-state energies improves the quality of ground-state predictions. We also present gap-driven dynamics for accelerated sampling of the small-gap regions, which proves crucial for stable surface-hopping dynamics. Together, multi-state learning and gap-driven dynamics enable efficient active learning, furnishing robust models for surface-hopping simulations and helping to uncover long-time-scale oscillations in *cis*-azobenzene photoisomerization. Our active-learning protocol includes sampling based on physics-informed uncertainty quantification, ensuring the quality of each adiabatic surface, low error in energy gaps, and precise calculation of the hopping probability.

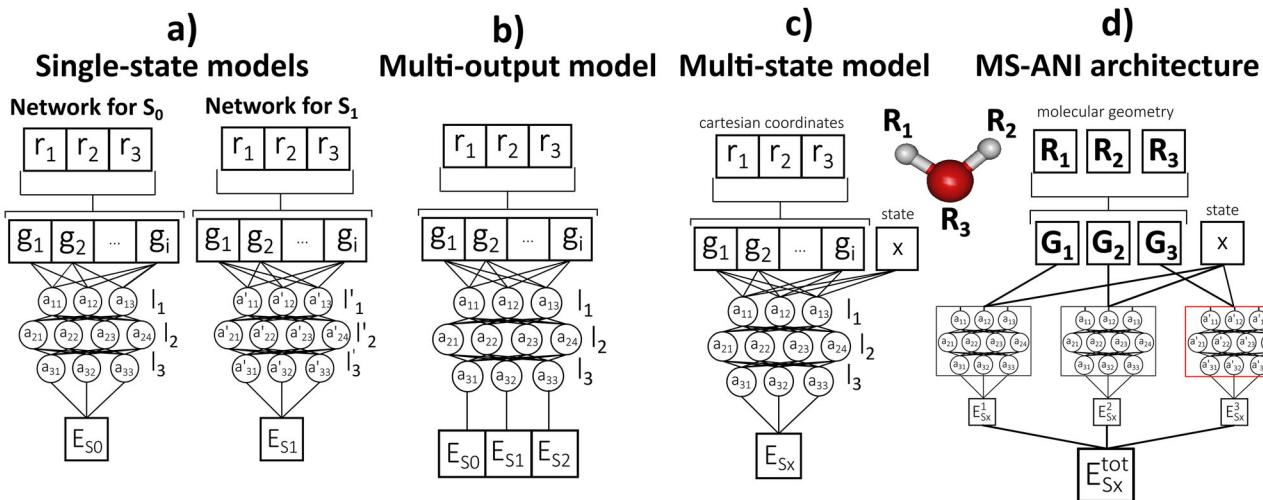
Electronic-structure methods offer unique insight into complex photophysical and photochemical problems, helping to guide and rationalize the experimental results. Unfortunately, these methods come with a steep computational cost, which severely limits their practical applications, particularly in nonadiabatic molecular dynamics simulations<sup>1</sup>. The latter provides an invaluable computational tool for investigating complex photoprocesses in the real-time domain, yet requires performing a large number of expensive excited-state calculations, thereby constraining their applicability. Today, by far the most popular nonadiabatic dynamics simulation technique is trajectory surface hopping (TSH). It has been successfully used to study a wide range of photoresponsive systems<sup>2–6</sup>. TSH simulates the excited-state dynamics of molecules by propagating a swarm of independent classical nuclear trajectories on quantum electronic potential energy surfaces (PESs). Nonadiabatic events are included through interstate instantaneous hoppings, whose probability is evaluated at each integration time step on the basis of the coupling strength between the

starting and the target PES. The swarm of surface hopping trajectories is expected to approximate the quantum nuclear wavepacket. TSH does not need global knowledge of the PESs, only of their values at the classical nuclear geometry. Thus, it is perfectly suited for on-the-fly simulations, in which electronic structure calculations delivering energies, energy gradients, and nonadiabatic couplings are executed as required in the course of the trajectory propagation. The main bottleneck of TSH simulations is their high computational cost due to the need to perform hundreds of thousands or even millions of single-point electronic-structure calculations. Hence, significant effort has been put into developing machine learning (ML) protocols to accelerate the TSH simulations of photoprocesses<sup>7–24</sup>.

As of today, cutting-edge ML models allow breaking through the limitations of the electronic-structure TSH simulations by enabling large-scale computations for longer times and with more quantum-classical trajectories. This research helped to uncover interesting photochemical phenomena, some of which were rather rare to be confidently quantified, or

<sup>1</sup>University of Warsaw, Faculty of Chemistry, 02-093 Warsaw, Poland. <sup>2</sup>State Key Laboratory of Physical Chemistry of Solid Surfaces, Department of Chemistry, College of Chemistry and Chemical Engineering, and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Xiamen University, Xiamen 361005, Fujian, China. <sup>3</sup>Aix Marseille University, CNRS, ICR, Marseille, France. <sup>4</sup>Institut Universitaire de France, Paris, France. <sup>5</sup>Institute of Physics, Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University in Toruń, Toruń, Poland. <sup>6</sup>Aitomistic, Shenzhen 518000, China.

✉e-mail: [jjankowska@chem.uw.edu.pl](mailto:jjankowska@chem.uw.edu.pl); [mario.barbatti@univ-amu.fr](mailto:mario.barbatti@univ-amu.fr); [dral@xmu.edu.cn](mailto:dral@xmu.edu.cn)



**Fig. 1 | A comparison of different NN models that can be used for excited state properties prediction, using a triatomic molecule with coordinates  $R_i = \{r_1, r_2, r_3\}$ .** **a–c** show atomistic NN, highlighting the differences between the approaches used to predict excited-state energies: **a** shows a combination of single-output NNs; **b** – multi-output NN; **c** – multi-state NN. The vectors  $G_i = \{g_1, g_2, \dots, g_i\}$  are the

molecular descriptors of a given atom,  $l_i$  are the network layers with nodes  $a_{ij}$ , and  $E_{Sx}$  is the predicted energy contribution of an atom to adiabatic state  $x$ . **d** shows the architecture of the MS-ANI model, which consists of separate NNs for each atom type, with all atom-wise predictions summed up to yield the total energy of the molecule for adiabatic state  $x$ ,  $E_{Sx}^{tot}$ .

even detected, with the pure electronic-structure calculations<sup>14,25–29</sup>. Despite all this progress, the state-of-the-art ML-accelerated TSH remains an extremely computationally expensive undertaking, with no clear protocols, requiring intensive human-expert supervision. For these reasons, it is still often easier to perform non-ML, pure electronic-structure TSH. This state of affairs precludes the widespread adoption of ML-TSH by the community, which is reflected by the fact that only a few expert groups reported new photochemical phenomena based on ML-accelerated TSH and the fraction of publications using ML in TSH studies remains relatively small compared to the bulk of TSH simulations (circa 3%, as estimated comparing all Web of Science records on “Trajectory Surface hopping” to the number of records that also mention machine learning).

The fundamental issue undermining the advance of ML-TSH is the challenge of predicting a dense manifold of potential energy surfaces, with their complex topography and small interstate energy gaps. Learning this manifold, along with all the intrinsic correlations, with high precision is required for robust ML-TSH. The most popular solutions suggested so far include creating single-state ML models (i.e., one ML model per each electronic state, Fig. 1a), as for example done by Westermayr et al. for the methylenimmonium cation<sup>11</sup>, or Hu et al. for 6-aminopyrimidine<sup>8</sup>, and multi-output ML models (i.e., a single neural network (NN) with the last layer containing as many output neurons as there are electronic states of interest), such as SchNarc<sup>13</sup>, SpaiNN<sup>24</sup>, and PyRAI<sup>2</sup>MD<sup>14</sup>. Both these solutions, unfortunately, have significant disadvantages: the single-state models do not capture correlations between the states, often leading to inferior performance in ML-TSH, while the multi-output models frequently suffer from more significant errors compared to single-state models, as shown later in this work. Learning excited-state PESs across different molecules is another big challenge, and only a few studies<sup>16,30</sup> have attempted to do so until today.

To address these challenges, though, having a good ML model architecture alone is not sufficient: one must precisely map the topography of the electronic-state manifold in the regions visited during TSH. This is an arduous task. Some proposed solutions were based on extensive manual construction of the data sets and on generating data with the pure electronic-structure TSH dynamics<sup>8,10,31</sup>. At the same time, for the sake of universality and efficiency, it is desirable to build data sets from scratch using active learning (AL). However, the reported AL strategies based on ML-TSH exploration required manual adjustment of sampling criteria<sup>11,14,26</sup>. Even then, these strategies were often insufficient for robust ML-TSH and require

further interventions, such as interpolation between critical PES points (e.g., between minima and conical intersections)<sup>26,27,32</sup>. Another critical issue preventing efficient ML-assisted NAMD is the need for easy-to-use, end-to-end protocols and software enabling routine simulations to be run by non-ML experts. However, progress is being made in this direction too<sup>13,14,33</sup>.

In this work, we solve these daunting problems by i) developing accurate and extendable physics-informed multi-state model capable of learning across chemical space (Fig. 1c), ii) proposing accelerated sampling of the critical small-gap regions in the electronic-state manifolds with gap-driven dynamics, and iii) implementing end-to-end, efficient, and robust active learning protocol based on physics and automatically, statistically determined criteria. We show that these new developments make the acceleration of TSH with ML affordable, as for not-so-flexible systems we can obtain the final simulation results within days on commodity hardware. The protocol also works for photoreactions albeit with more computational effort needed. Remarkably, our methods also enable learning an arbitrary number of electronic states not just for a single molecule but also simultaneously across different molecules and different reference electronic-structure levels.

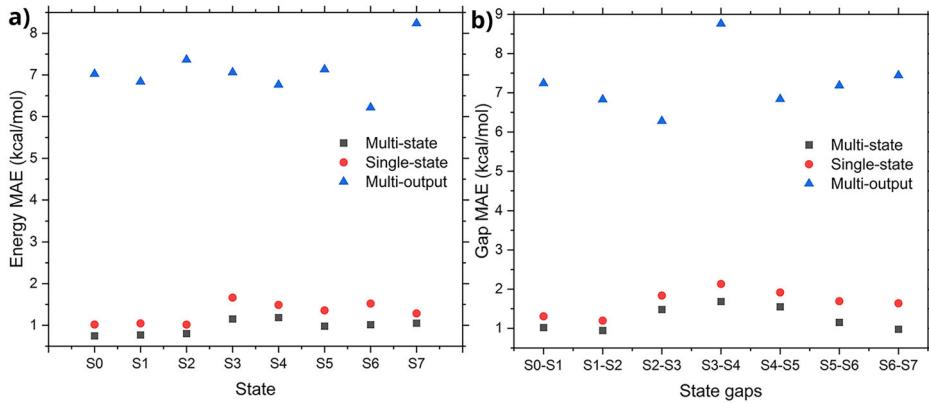
## Results

### Multi-state learning

Here we introduce a novel ML architecture that: 1) can learn an arbitrary number of electronic states on equal footing with high accuracy, with the number of electronic states used independent of the NN structure 2) can make predictions for different molecules, 3) captures the required correlations between states and, especially, correctly reproducing energy gaps between surfaces, and 4) is easy (relatively fast) to train and evaluate. The core idea of this new implementation is captured in Fig. 1c.

The distinguishable feature of our model is the inclusion of information about the state ordering number (e.g., 0 for the ground state, 1 for the first excited state, etc.) in the model-processed features alongside the geometric descriptors. This leads to multiple benefits compared to either single-state or multi-output models reported in the literature previously (cf. Fig. 1a, b). The paramount benefit is that this architecture can handle an arbitrary number of states on equal footing. The information about all the states is passed through all hidden and output neurons of the NN, and these neurons differentiate between different states. This allows not only for electronic state information to be propagated through the entire network in the training process but also training on data with diverse labels (molecules

**Fig. 2 | Performance of the different NN architectures.** Multi-state (black squares), single-state (red dots), and multi-output (blue triangles) models are compared by predicting the energies (a) and energy gaps (b) of the first 8 electronic states of pyrene, and judged by mean absolute errors (MAEs).



with different numbers of labeled electronic states) and transfer learning between very different datasets. In contrast, in the single-state models, only information about one state is learned at a time, i.e., the correlation between states is essentially lost, which might lead to the inferior performance in ML-TSH, as was observed earlier<sup>34</sup>. In the multi-output models, on the other hand, only the last output layer differentiates between states. Still, the correlation between states is implicitly learned to some extent as the weights are shared in the preceding layers<sup>35</sup>. Eventually, it may lead to better performance in ML-TSH, despite the larger errors in energies.

Furthermore, the proposed multi-state model is ideally suited for capturing crucial photophysical information, such as the importance of energy gaps between states. Crucially, we found that the accurate treatment of small-gap regions is the key to the robust performance of ML models in TSH. As one of the solutions to the small-gap problem, we include the special loss term  $L_{\text{gap}}$  taking into account the error in the gaps, which ensures accurate prediction of energy gaps when training the multi-state models:

$$L_{\text{gap}} = \|\Delta E^{\text{ML}} - \Delta E^{\text{ref}}\|^2, \quad (1)$$

where  $\Delta E$  denotes the energy gaps between adjacent adiabatic states, and the superscripts 'ML' and 'ref' correspond to the NN prediction and reference values, respectively.

We include this loss in the total loss,  $L$ , containing the terms for energy,  $L_E$ , and force errors,  $L_F$ :

$$L = \omega_E L_E + \omega_F L_F + \omega_{\text{gap}} L_{\text{gap}}. \quad (2)$$

The energy and force losses follow their usual definitions of  $\|E^{\text{ML}} - E^{\text{ref}}\|^2$  and  $\|\mathbf{F}^{\text{ML}} - \mathbf{F}^{\text{ref}}\|^2$ , respectively. Each of the loss terms comes with the corresponding weight,  $\omega$ , with  $\omega_E$  and  $\omega_{\text{gap}}$  equal to 1, treating the gaps and energies as equally important, and with  $\omega_F$  set to the standard value of 0.1. This makes our multi-state model a representative of the physics-informed NNs, which were shown to provide more qualitatively accurate behavior in the related quantum dynamics context<sup>35</sup>.

The overall NN architecture is based on the established ANI-type network, which has a good balance between cost and accuracy<sup>36</sup>, and which we later compare with the more accurate yet slower equivariant networks. It means that we use the same ANI-type structural descriptors and have separate networks for each element type. The predictions are made for each atom separately and then summed up to yield the total energy for the requested state. The self-atomic energies are computed once and shared for all states of the model. We refer to this architecture as multi-state ANI (MS-ANI).

To demonstrate the performance of the multi-state model, we first evaluate its accuracy in predicting the energies and energy gaps for the first eight electronic states of pyrene. To this end, we constructed a data set of 4000 molecular conformations generated from the Wigner sampling around the  $S_0$  minimum structure. All calculations were performed with

AIQM1<sup>37</sup> using CIS<sup>38</sup> treatment for the excited-state properties. Eventually, we trained the multi-state model on 3000 conformations and evaluated it on the remaining 1000 test conformations. Remarkably, the model yielded mean absolute errors (MAEs) of energies close to ca. 1 kcal/mol (0.04 eV) (Fig. 2a) and root-mean-squared errors (RMSEs) below 1.6 kcal/mol (Table S1 of the ESI). The energy gaps were also described with good accuracy, with MAEs below 1.7 kcal/mol (Fig. 2b), and with the maximum root-mean-squared error (RMSE) below 2.2 kcal/mol (Table S2). Altogether, errors for energies and gaps are falling close to the chemical accuracy margin of 1 kcal/mol.

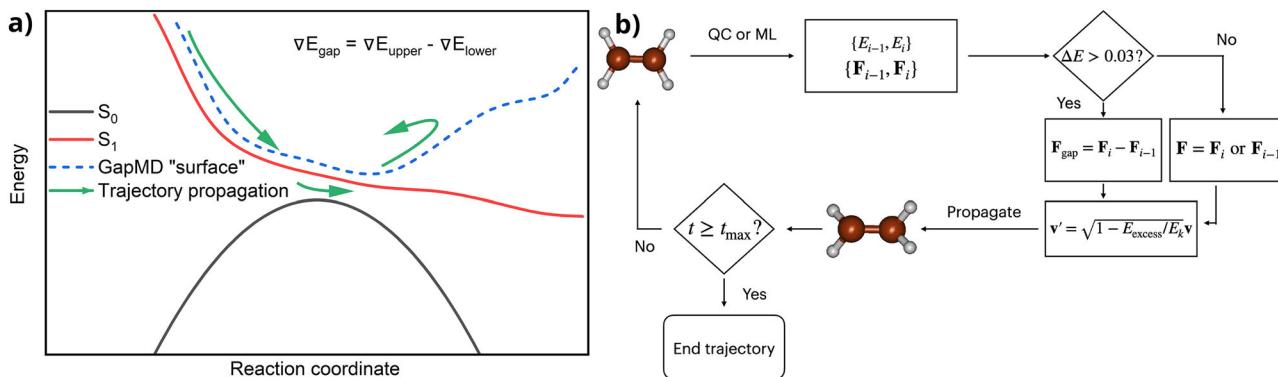
These calculations are better put in perspective when compared to the single-state and multi-output ANI-type models. One can observe that, in the multi-output model, for all electronic states, the errors in energies and gaps are substantially higher than in the multi-state and single-state models, with MAEs between 6.2 and 8.3 kcal/mol for energies, and 6.3 and 8.5 kcal/mol for energy gaps. When comparing the multi-state model with the single state model, we can see that MS-ANI outperforms it for all electronic states energies and gaps, with the differences being most pronounced for higher excited states and strongly coupled  $S_3$ - $S_4$  states (Fig. 2, as well as Tables S1–S3). In all of the single-state models' MAEs and RMSEs are above 1 kcal/mol.

In addition to the multi-state model's superior performance, its training required much less time than training the eight separate single-state models (45 minutes vs 3 hours on an RTX 4090 GPU).

We also observe, for the first time, that the accuracy of energies of the ground state is improved when information regarding the other, i.e., excited states, is included. Moreover, our multi-state model can learn the excited-states properties with the same or higher accuracy than a single-state model can learn the ground state: a previously unachievable result.

#### Mapping small-gap region with gap-driven dynamics

Regardless of how good the model is, a sufficient amount of data to sample all relevant regions of the PESs for the TSH is always a necessity. This data is usually generated based on the TSH trajectories, either propagated with a reference electronic-structure method or with ML in an active learning loop. However, such data may have insufficient representation of the critical regions in the vicinity of the conical intersections because the TSH trajectories typically contain only very few points in that region representing a small fraction of all time steps in the trajectories. It is also known that ML models, in general, struggle with reproducing the PES near conical intersections<sup>39</sup>, hence extra steps must be taken to ensure the correct learning of low-energy-gap regions. This is particularly true for  $S_0$ / $S_1$  conical intersections, usually associated with more drastic geometric changes than excited-state crossings. This explains why some of the previous studies, even after training on extensive data sets (e.g., from the reference TSH trajectories), still could not obtain robust dynamics and needed to switch from ML to the reference electronic-structure calculations in the region of small gaps during the ML-accelerated TSH trajectory propagation<sup>8,9,31</sup>.



**Fig. 3 | Gap dynamics.** **a** Shows a schematic depiction of the gap-driven dynamics trajectory propagation, while **(b)** is an algorithmic flowchart of the propagation scheme.

To mitigate this issue, specially designed protocols, such as interpolation from minimum energy conical intersection points<sup>14,26,27</sup> were sometimes employed.

Here, we have developed an accelerated sampling approach to meticulously chart the small-gap regions in the vicinity of conical intersections. We use this approach to improve the robustness and accelerate the active learning loop described in the next section, but we envision that it can also be used in other contexts, such as accelerating the search for conical intersections themselves. Our approach samples the points relevant to the TSH via propagating special, gap-driven molecular dynamics (gapMD) trajectories (Fig. 3).

The gapMD trajectories involve the back-and-forth switching of propagation along the energy gap gradient and energy gradients of the adiabatic surfaces. In the region of gaps,  $\Delta E$ , larger than 0.03 Hartree, the trajectories are propagated along the energy gap gradient,  $\nabla E_{\text{gap}}$ , defined as the difference between energy gradients in the upper, ( $\nabla E_{\text{upper}}$ ), and lower, ( $\nabla E_{\text{lower}}$ ), surfaces:

$$\nabla E_{\text{gap}} = \nabla E_{\text{upper}} - \nabla E_{\text{lower}}, \quad (3)$$

with equivalent expression formulated in terms of forces:

$$\mathbf{F}_{\text{gap}} = \mathbf{F}_{\text{upper}} - \mathbf{F}_{\text{lower}}. \quad (4)$$

These trajectories, by construction, drive the dynamics into the regions with smaller gaps. The threshold of 0.03 Hartree is chosen because it is a typical range of gaps at which the interstate hoppings commonly happen<sup>40</sup>. Once the region of a smaller gap is reached, we switch to propagating trajectories using the energy gradients of either the upper or lower surface (see below). This is done for several reasons. Firstly, propagating exclusively with the energy gap gradients will sample many irrelevant points, e.g., towards the complete dissociation of a molecule, as the dissociated structures have degenerate energy levels with a zero gap. Furthermore, in the actual TSH, trajectories are propagated either on the upper or lower surface, as opposed to propagating along the gradient difference, which does not correspond to any of the adiabatic PESs. To ensure sampling relevant points of the phase space, the trajectory is switched to one of the surfaces that can be populated in TSH. Hence, our trajectories propagated with the energy-gap gradients are only for biasing dynamics to visit the small-gap region faster and enrich the sampling of this region.

Switching back and forth from energy gap gradients to electronic state gradients makes the gapMD not energy-conserving. To ensure energy conservation and to avoid obtaining dissociated and otherwise nonphysical geometries, if the excess energy  $E_{\text{excess}}$  is smaller than the current kinetic energy  $E_k$  of the system, the atomic velocities  $\mathbf{v}$  are scaled along the

momenta, i.e., according to the formula:

$$\mathbf{v}' = \sqrt{1 - \frac{E_{\text{excess}}}{E_k}} \mathbf{v}. \quad (5)$$

The updated velocities  $\mathbf{v}'$  are used in the further propagation. The excess energy,  $E_{\text{excess}}$ , is defined as a difference between the total energy at the current time step and the initial total energy at the time step zero. If the excess energy is larger than the kinetic energy at the current time step, it means that there is not enough kinetic energy to compensate for this excess in the system. Hence, in such a case, we switch the trajectory propagation to the electronic state surface regardless of the gap value.

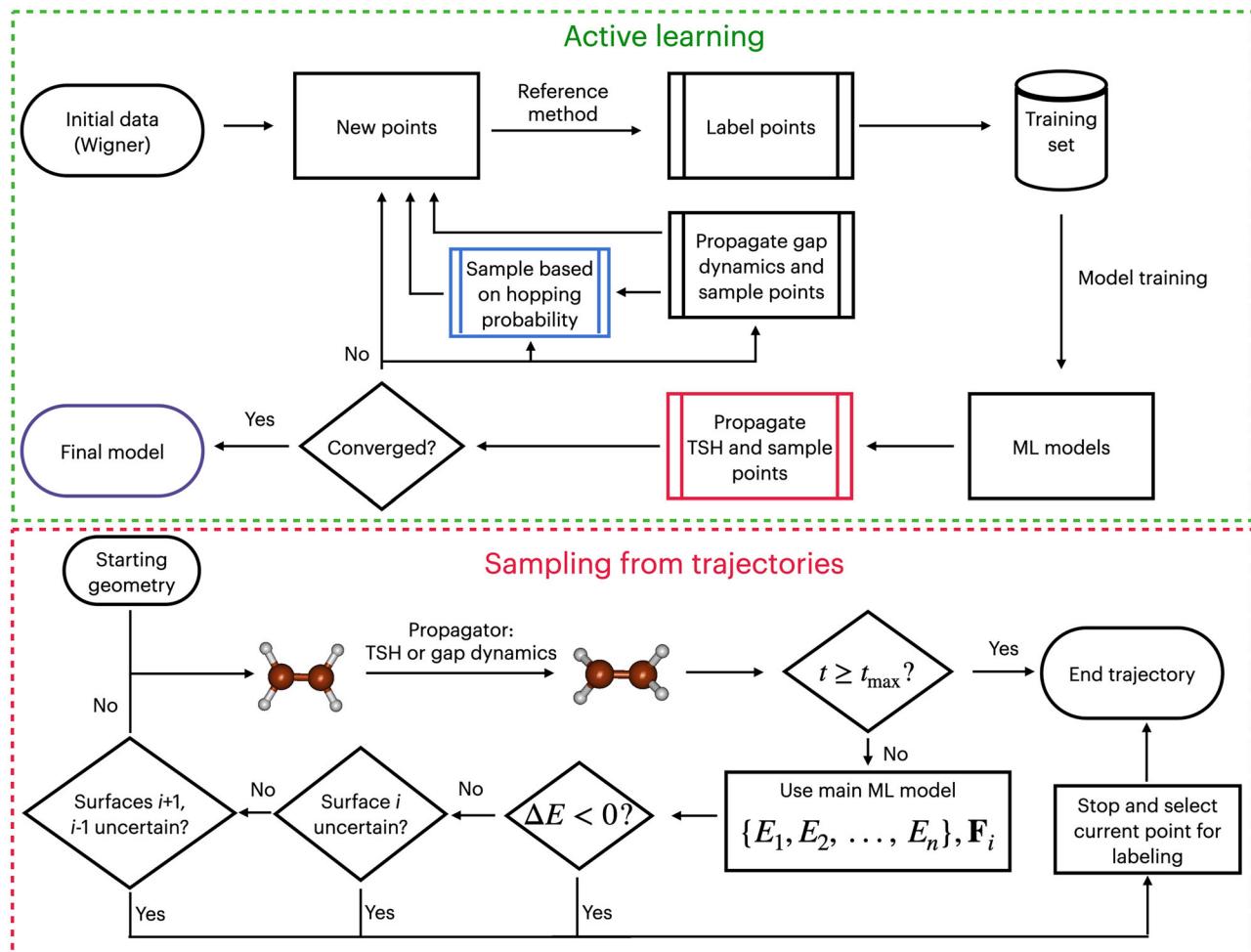
#### End-to-end active learning targeting accurate hopping probabilities

Ultimately, our goal is to design a protocol for robust ML-TSH through meticulous charting of the electronic-state manifolds of new systems from scratch. The multi-state models and gapMD provide useful tools for achieving this goal. Hence, we incorporate both tools into the end-to-end, physics-informed active learning (AL) protocol (Fig. 4). This AL protocol ensures not only the sampling quality of the adiabatic surfaces and small-gap regions but also that the hopping probabilities are accurate.

The core of the AL protocol is running many trajectories to sample new points, labeling them (i.e., calculating the reference electronic-structure properties), training the ML model on the data updated with the newly labeled points, and repeating the procedure until converged. We used our domain knowledge to tune all its stages to obtain robust ML-TSH results as efficiently as possible.

We generate the initial data set using statistical considerations in the same way as described previously for our AL procedure for the ground-state molecular dynamics<sup>41</sup>. In brief, the points are sampled until the validation error does not drop too much (judged by training the ground-state energy-only model). From a ground-state minimum geometry of the studied molecule, 50 conformations are sampled from a harmonic-oscillator Wigner distribution<sup>42</sup>, and the error in energy prediction is calculated using a 5-fold cross-validation. The procedure is repeated until the projected accuracy improvement is less than 10%, as estimated by fitting the learning curve.

Once the initial data is sampled, we start the AL loop by training our multi-state model on both energies and energy gradients of all electronic states and propagating ML-TSH trajectories started from a selected number of Wigner-sampled initial conditions (atomic positions and velocities), which was set to 50 in the current work. At each time step, we check several criteria to evaluate whether to propagate or to stop the trajectory and sample the geometry at this time step (Fig. 4, bottom). The first criterion is whether the ML-predicted energy gap between the current and adjacent surfaces is negative: this check ensures correct state ordering. Then, we check whether



**Fig. 4 | Flowchart describing the active learning procedure.** The upper panel shows the full loop of the procedure, from sourcing the initial dataset, labeling with a reference method of choice, training the main and auxiliary ML models used to propagate TSH simulations, from which points are sampled, later propagation of

gapMD dynamics, and finally sampling points based on hopping probability. The bottom panel details how TSH and gapMD trajectories are propagated, where the three criteria used to sample points: negative energy gaps, current surface uncertainty, and adjacent surface uncertainty.

the uncertainty quantification (UQ) metric for the current surface exceeds the threshold, which ensures the quality of the relevant adiabatic surfaces on which the dynamics are propagated. Finally, we check whether the UQ for the surfaces up and down on the current surface exceeds their respective thresholds. This ensures that the crucial energy gaps are monitored indirectly. It is also possible to check the UQ for the energy gaps involved directly, but our tests showed that this does not enhance the model performance, particularly when gapMD is used (see below).

The UQ values,  $U$ , are obtained using our previous, physics-informed scheme<sup>41</sup>, which was developed for ground-state dynamics. The underlying assumption of this procedure is that the shape of the potential energy surfaces must be contained in the sampled points used for training ML potential. If this is not the case, the model extrapolates from the given data, which can break down and lead to unphysical results. The uncertainty is checked by evaluating the absolute deviations in energy predictions by the main multi-state model trained on more physical information (energies and energy gradients), and the auxiliary multi-state model trained on less information (only on energies):  $U = |E^{\text{aux}} - E^{\text{main}}|$ . If the model starts to extrapolate the shape of the PES using the known energy gradients, then the deviation should grow with respect to the auxiliary model, which does not have access to the energy gradients. The main model is used to propagate the nonadiabatic dynamics simulations, while the predictions of the auxiliary model are only used to judge the quality of the predicted energies. This has been shown to produce robust active learning for ground-state dynamics<sup>41,43</sup>.

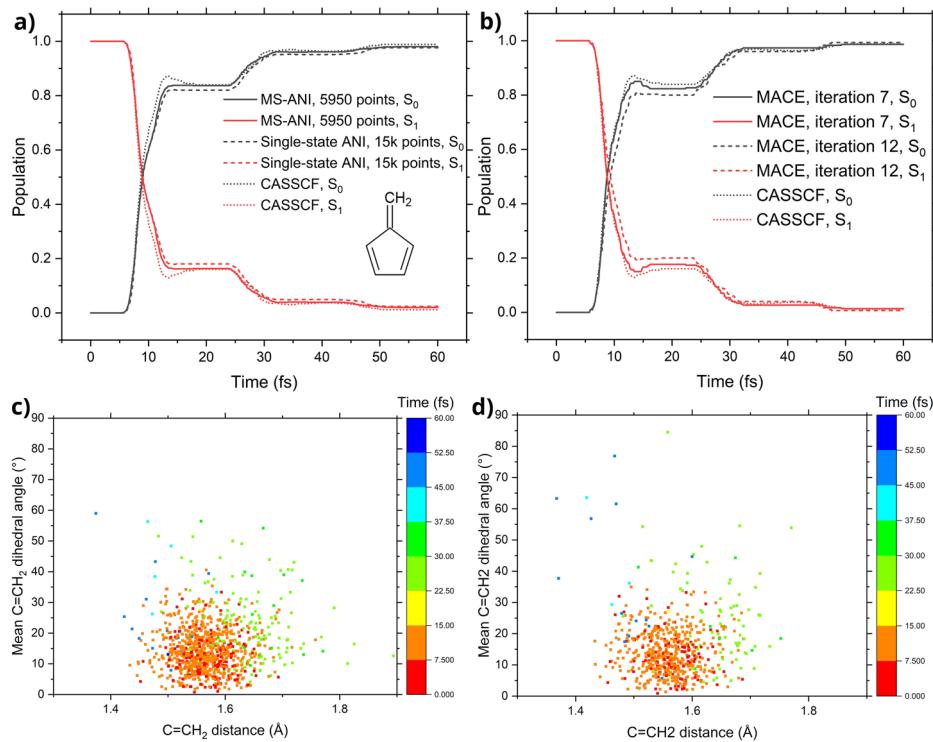
The UQ thresholds are calculated based on statistical considerations for each state separately, in the analogous way as before in the ground-state AL<sup>41</sup>. They are evaluated with the UQ values calculated for the initial validation set (10% of the initial data):

$$UQ_{\text{threshold}} = M(U) + 3 \cdot \text{MAD}(U) \quad (6)$$

where  $M$  is the median, and  $\text{MAD}$  is the median absolute deviation, with  $M + 3 \cdot \text{MAD}$  ensuring a confidence level of 99%. While this threshold can be modified, our testing has indicated that lower values (2 MAD) do not improve the quality of the simulations, increasing only the time it takes to converge the AL procedure, while higher values lead to poor results. The UQ thresholds are calculated using the data available in the initial training set and fixed for the rest of the AL process. This procedure has the benefit that no manual, subjective setting of the thresholds is needed, in contrast to the adaptive sampling widely used in the field<sup>11,21,22</sup>.

None of the above sampling criteria directly addresses the key factor controlling surface hopping: the hopping probability. This factor is susceptible to any deviations in the energy gap, and having a model that yields accurate probabilities is essential for ensuring robust performance in TSH. To further refine the model in terms of hopping probability, after the propagation of each TSH trajectory, we evaluate their uncertainties. The UQ metrics are computed as absolute deviations between the hopping probabilities evaluated using the energies predicted by the main and auxiliary

**Fig. 5 | Summary of the nonadiabatic molecular dynamics simulations of fulvene.** **a** Shows the electronic state population evolution during the dynamics conducted with the MS-ANI model (full lines), single-state ANI models (dashed lines), and reference CASSCF dynamics (dotted line). **b** Shows the electronic state population evolution during the dynamics conducted with the MACE models at iteration 7 (full line), iteration 12 (dashed line), and reference CASSCF populations (dotted line). **c, d** Show correlation plots of the C=CH<sub>2</sub> distance and the mean dihedral angle at the S<sub>1</sub> → S<sub>0</sub> hopping points in ML-TSH dynamics (**c**) and reference CASSCF dynamics (**d**).



models at each time step before the UQ threshold is exceeded. We use the same formula as in the Landau-Zener-Belyaev-Lebedev (LZBL) formulation of TSH<sup>44–48</sup>:

$$P_{j \rightarrow k} = \exp \left( \frac{-\pi}{2\hbar} \sqrt{\frac{Z_{jk}^3}{\ddot{Z}_{jk}}} \right), \quad (7)$$

where  $Z_{jk}$  is the energy gap between adiabatic states,  $j$  and  $k$ , and  $\ddot{Z}_{jk}$  is the second-order time derivative of that gap. The LZBL formalism is also used for all TSH propagations in this work as implemented in MLatom<sup>48,49</sup> due to its simplicity and because it does not require the evaluation of the nonadiabatic couplings. We identify all time steps in all trajectories where the main and auxiliary hopping probabilities deviate by more than 10% (if higher fidelity is needed, the threshold can be reduced). In the end, no more than 15 probability-uncertain points from both ML-TSH and gap-driven trajectories are randomly sampled per AL iteration to ensure a balanced training set.

To better sample points in the direct vicinity of the conical intersections, we spawn ML-gapMD trajectories with initial conditions (geometry and velocities) taken from a random time step of each ML-TSH trajectory (see the section on gap-driven dynamics). Initial conditions for gapMD are only selected from the TSH trajectory before the first uncertain time step (as judged by the negative gap and uncertainty quantification for surfaces). The ML-gapMD are subject to the same sampling and stopping criteria, i.e., UQ of the surface energies, negative gap checks, and hopping probability uncertainty. In each iteration of the active learning procedure, a given number of trajectories, usually equal to the number of ML-TSH trajectories, are spawned, following the gradient difference of two randomly selected adjacent potential energy surfaces. Half of these trajectories are selected to follow the upper potential energy surface in the small-gap region, and half of them follow the lower surface.

The convergence rate of AL is evaluated as the ratio of certain trajectories to the total number of ML-TSH trajectories. Trajectories are considered certain if they are propagated without exceeding the surface UQ thresholds or having negative gaps. Hopping probability UQ is not taken

into account, as it is an exponential function of the energy gap, which is very sensitive to any deviations. If the convergence rate is greater than the desired value (95% in this work), the AL procedure is stopped, and the current model is considered converged.

At an AL iteration, the number of sampled points usually exceeds the number of trajectories because we select them from additional ML-gapMD trajectories and hopping probability conditions. However, all the points sampled in a given AL iteration are included in the training set. Alternatively, to speed up convergence, each AL iteration can proceed until a pre-selected number of sampled points is achieved, after which the models are re-trained.

#### How efficient is the physics-informed AL?

To assess the efficiency of the physics-informed AL protocol, we test it by generating the training data and multi-state model for a fulvene molecule from scratch.

For fulvene, the AL convergence was achieved in three days on a single RTX 4090 GPU and 16 Intel Xeon Gold 6226R CPUs, only with 19 iterations: an unprecedented efficiency not reported before. The final training data set contained relatively few (5950) points. This labeling cost is equivalent to computing only ten quantum mechanical trajectories in terms of CPU time; in terms of the wall-clock time, labeling can be performed much faster as it can be efficiently parallelized. The performance of the model in ML-TSH shows excellent agreement with the reference quantum-chemical results both in terms of the populations and the distribution of the geometric parameters at the S<sub>1</sub> → S<sub>0</sub> hopping points (Fig. 5a, c, and d, as well as Table 1). The latter are classified into three groups, following the refs. 50,51 based on the C=CH<sub>2</sub> bond length, and the mean dihedral angle  $\phi_{C=CH_2}$  between the 5-membered ring and the methylene group:

$$\phi_{C=CH_2} = \frac{1}{4} \left( |\phi_{C=CH_2}^{cis_1}| + |\phi_{C=CH_2}^{cis_2}| + |\phi_{C=CH_2}^{trans_1}| + |\phi_{C=CH_2}^{trans_2}| \right). \quad (8)$$

The planar group is defined with  $\phi_{C=CH_2} < 30^\circ$ , twisted-stretched— $\phi_{C=CH_2} > 30^\circ$  and  $C=CH_2 > 1.55 \text{ \AA}$ , and twisted-shrunk— $\phi_{C=CH_2} > 30^\circ$  and  $C=CH_2 < 1.55 \text{ \AA}$ .

**Table 1 | Mean value and error bars (95% confidence interval) of observables describing the deactivation channels and kinetics of fulvene for ML dynamics and reference CASSCF dynamics**

Observable	CASSCF dynamics	ML dynamics
$S_1$ Population at 20 fs (%)	$16.3 \pm 2.9$	$16.9 \pm 2.3$
$S_1$ Population at 40 fs (%)	$3.7 \pm 1.5$	$3.9 \pm 1.2$
Planar hopping (%)	$94.0 \pm 1.8$	$93.8 \pm 1.5$
Twisted-stretched hopping (%)	$3.7 \pm 1.5$	$4.1 \pm 1.2$
Twisted-shrunk hopping (%)	$2.3 \pm 1.2$	$2.15 \pm 0.90$

The obtained data is also in good agreement with previous works<sup>50,51</sup>, reporting CASSCF dynamics of fulvene with the Baeck-An couplings.

### Do we need all these complications?

The important question is: do we truly need all these complications for high-quality ML-TSH dynamics? Indeed, we tried many different settings over eight years of research, and none showed satisfactory performance in terms of computational efficiency and robustness of the protocol. For example, a common robustness problem observed with less elaborate protocols was that the populations may deteriorate with more AL iterations. Small changes in the AL settings would sometimes lead to completely different results, too. Another issue is that, often, obtaining trajectories with reference electronic-structure method turns out less troublesome than fighting with all the instabilities of ML-TSH. As evidenced by many publications on ML-TSH, these issues can be tolerated and mitigated by experts to solve the problems that are beyond the reach of electronic-structure TSH. Still, they definitely hamper the wider adoption of the technique.

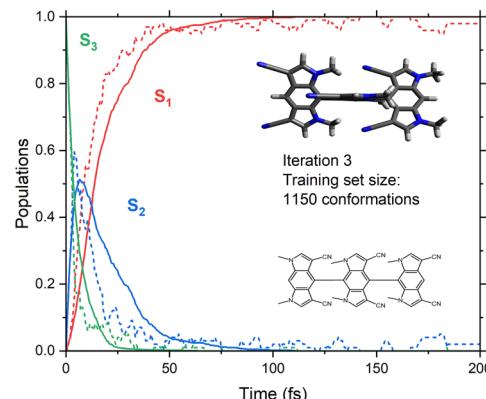
For example, when we use our physics-informed AL developed for the ground state with single-state ANI-type ML models and without gapMD and probability uncertainty evaluation, we can also obtain a good population plot for fulvene (Fig. 5a, no checks for negative gaps were performed either). The problem is, however, that the procedure converged only after a whopping 104 days, 160 iterations, and ca. 15000 training points (using NVIDIA GeForce RTX 3090 for model training and 8 CPU nodes for labeling and MD propagation). Admittedly, the populations started to look good with as few as dozens of iterations, but, in general, it should be considered risky to take such non-converged models if one does not know the reference population in advance.

One can also question whether a better ML model can remove the need for some of our special techniques. To address this, we repeated the above-simplified protocol but with the state-of-the-art, single-state MACE ML model<sup>52</sup>. The model is very slow, which resulted in 28 days of AL to produce 13 iterations and 3850 training points, and it eventually crashed due to memory issues. The AL was not converged either (only 77% trajectories were converged), and the last iteration's model yielded a population that was worse than some of the previous iterations (Fig. 5, panel (b)). The major problem of the MACE model is its high cost, which might be only worth paying in special cases that we have not identified so far.

All of the above anecdotal examples illustrate that the special techniques of our final end-to-end protocol are not bells and whistles but the result of an 8-year-long struggle to obtain an efficient and robust protocol. These examples are just the tip of the iceberg of the thousands of thrown-away experiments and several generations of students and postdocs efforts.

### Does the protocol work for big systems and more states?

Fulvene is a simple system of small size and of only two-state-driven photodynamics. To go beyond such a simple picture, we tested our end-to-end protocol on a molecular ferro-wire made of 80 atoms and featuring four electronic states in its regular photorelaxation dynamics. The chosen molecule is characterized by a complicated electronic structure due to charge transfer between its three structural units. Amazingly, our AL



**Fig. 6 | Electronic state population evolution during the dynamics of the molecular ferro-wire.** Predictions of the ML model are plotted with full lines, and reference AIQM1/CIS calculations with dashed lines.

procedure converged after only three iterations and produced 1150 training points. The final obtained multi-state model agrees well with the reference populations determined at the AIQM1/CIS level (Fig. 6). Testing of the model was performed using 1000 ML-TSH trajectories and compared to 100 reference AIQM1/CIS trajectories.

In this case, propagating one ML-TSH trajectory takes about 5 min on a single Intel Xeon Platinum 8268 CPU, which is a significant speed-up even with respect to fast methods such as AIQM1/CIS (having the speed of a semi-empirical method), where one trajectory takes about 1.5 CPU-hours on average. The computational efficiency of ML-TSH with MS-ANI is reported in Table S3 of the ESI, with trajectories propagated faster than 1.5 ns/day for all studied systems.

### Can the protocol learn photoreactions?

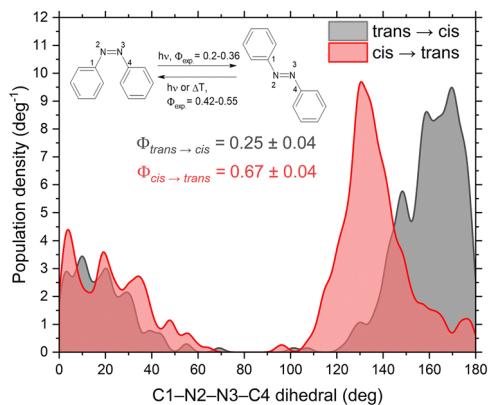
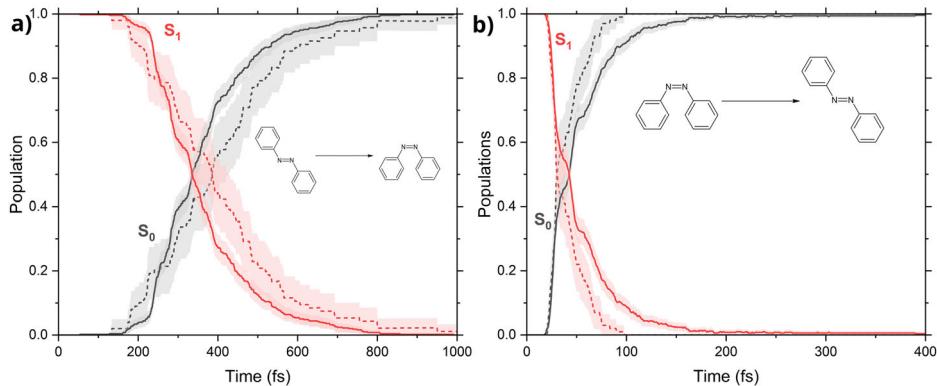
Next, we would like to see whether the proposed end-to-end protocol works for a more complicated case of photochemical transformation. For this, we take as a test case the well-known photoisomerization of azobenzene<sup>53–55</sup>, starting from both *cis* and *trans* isomers.

To do this efficiently, we introduce a slight modification to the AL procedure in this case. Namely, at each iteration, an equal number of trajectories is started from both the *cis* and *trans* isomers of azobenzene. This ensures that we sample both the product and the starting material of each photoprocess, greatly accelerating the model-training convergence (as is known<sup>41</sup> from the AL-based search of conformers in their ground state). After running for about 3 weeks of wall-clock time, with 108 iterations (about 18000 training points), the model was 92% converged for the set of trajectories with initial conditions filtered to an excitation window of  $2.53 \pm 0.3$  eV, corresponding to the  $S_0 \rightarrow S_1$  transition energy at the ground-state minimum, computed at the AIQM1 level of theory. To speed up the calculations, a broadening procedure using Wigner sampling was performed, which consisted of computing frequencies at each point of the training set, using the ML model, and using these frequencies to sample an additional point, effectively doubling the training set size. This increased the convergence to 97% and resulted in a noticeable improvement in the predicted populations. The constructed model was 98% converged for the reverse, *cis*  $\rightarrow$  *trans* photoisomerization process. The final training set consisted of 35071 points. Eventually, 500 trajectories were propagated for both isomers, with the results discussed below.

Analyzing the electronic state populations during the dynamics (depicted in Fig. 7), we see outstanding agreement between the ML model and reference AIQM1/MRCI-SD trajectories in the *trans*  $\rightarrow$  *cis* isomerization reaction, as well as good agreement for the *cis*  $\rightarrow$  *trans* photoprocess.

The excited-state lifetimes can be extracted by fitting an exponential function to the ground-state population rise with  $P(t) = A(1 - \exp(-t/\tau))$ ,

**Fig. 7 | Electronic state population evolution during the isomerization dynamics of azobenzene.** Trajectories shown in (a) started from the *trans* isomer, while trajectories starting from the *cis* isomer are shown in (b). ML-TSH is represented by full lines, while reference AIQM1/MRCI-SD trajectories are dashed. The shaded areas show the statistical uncertainty of the trajectories with a 95% confidence interval.



**Fig. 8 | Distribution of the C-N-N-C dihedral angle of azobenzene at the trajectories final points.** Distribution derived from trajectories starting from the *trans* isomer of azobenzene is marked in gray, while the distribution derived from geometries starting from the *cis* isomer of azobenzene is marked in red.

where  $\tau$  is the photoprocess timescale, and  $A$  is associated with the infinite-time population limit. After fitting for the *trans*  $\rightarrow$  *cis* isomerization, we obtained  $387 \pm 15$  fs for the ML-TSH dynamics and  $437 \pm 16$  fs for reference AIQM1/MRCI-SD, which are in excellent agreement. For the *cis*  $\rightarrow$  *trans* process, the fitted timescale is  $53.0 \pm 0.8$  fs for ML-TSH dynamics and  $41 \pm 1$  fs for reference dynamics. While the predicted decay rates agree acceptably well between the ML model and reference dynamics, overall, the AIQM1 method seems to overestimate the speed of the *trans*  $\rightarrow$  *cis* isomerization process, as literature data<sup>53</sup> report a timescale of 74 fs at the CASSCF level.

As the next point in the analysis, we can look at the predicted quantum yields of the photoisomerization processes. The dihedral angle C<sub>1</sub>-N<sub>2</sub>-N<sub>3</sub>-C<sub>4</sub> (as shown in Fig. 8) can be used as a convenient descriptor of the formed product, taking the value of  $\sim 0^\circ$  for *cis*-azobenzene, and  $180^\circ$  for *trans*-azobenzene. The distribution of the final dihedral angles can be found in Fig. 8, in the form of a density distribution plot. Taking a dihedral angle of fewer than  $60^\circ$  as a fingerprint of the *cis* isomer, and attributing the rest of the population to the *trans* isomer (no other photoproducts were observed), the simulated quantum yields  $\Phi$  are calculated as the fraction of the trajectories relaxing to form the photoproduct (the other isomer),  $N_{\text{reactive}}$ , to the total number of trajectories  $N_{\text{traj}}$ :  $\Phi = N_{\text{reactive}}/N_{\text{traj}}$ .

Here, the model shows excellent agreement with results obtained by the reference method, as well as the experimental results<sup>56,57</sup>. The ML-TSH dynamics predicts the quantum yields of  $\Phi_{\text{trans} \rightarrow \text{cis}}^{\text{ML}} = 0.25 \pm 0.04$  and  $\Phi_{\text{cis} \rightarrow \text{trans}}^{\text{ML}} = 0.67 \pm 0.04$ . The uncertainties of the quantum yields were estimated using the normal approximation interval for a binomial process, assuming a confidence interval of 95%. These values agree with the predictions of the reference method, AIQM1/MRCI-SD ( $0.24 \pm 0.09$  for

*trans*  $\rightarrow$  *cis*, and  $0.78 \pm 0.10$  for *cis*  $\rightarrow$  *trans*), as well as with experimental results, predicting a quantum yield of 0.20–0.36 for the *trans* to *cis* isomerization. The remarkable accuracy at which the proposed model reproduces quantum yields simulated using reference methods can be attributed to the inclusion of the gap-driven trajectories in the sampling process.

#### Discovering oscillations in *cis*-azobenzene photoreaction

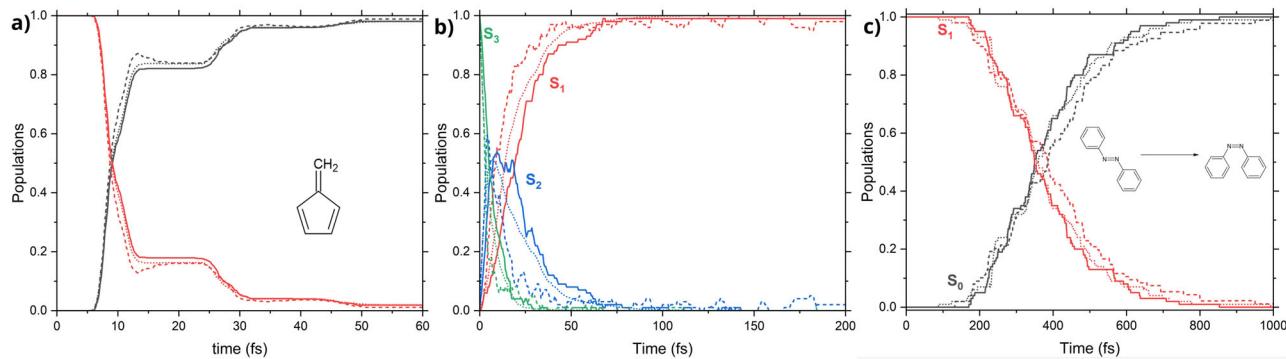
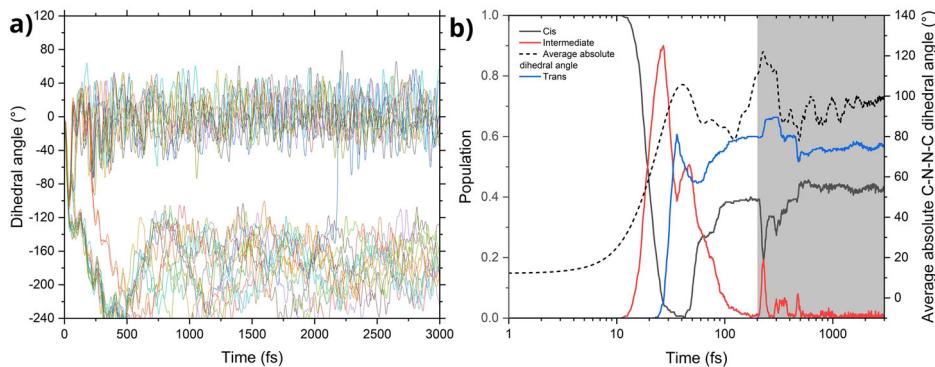
We use the ML-ANI model trained for azobenzene to extend the timescale of the ML-TSH simulations, propagating 300 additional trajectories to 3 ps. These simulations revealed that the *cis*-*trans* photoisomerization has some features not previously reported. We found out that after relaxation to the ground state and isomerization to the *trans* isomer, the molecule is so hot that it often undergoes the reverse process of isomerization to the *cis* isomer (Fig. 9a). This back-and-forth *cis*-*trans* isomerizations have an oscillatory behavior persistent till the end of trajectory propagation, even after 100% of the trajectories decay to the ground state (Fig. 9b). This long-scale thermally induced isomerization process after the decay to the hot-ground state has not been observed in the previous, smaller scale (shorter and with fewer trajectories) NAMD studies<sup>53,54</sup>. These oscillations occur in a much longer time scale than those reported by Weingart et al.<sup>58</sup> (which we could also reproduce with our model).

Machine learning proves to be a convenient tool in this analysis, as it allows us to rapidly compute long trajectories, which are necessary to make this observation. Additionally, it allows us to propagate arbitrarily many trajectories, making this analysis more precise. Propagating a single *cis*-azobenzene trajectory for 3 ps takes about 30 min of CPU time, compared to an estimate of 2 days using reference methods. These oscillations are not just curious observations, they have an impact on the overall quantum yield of the reaction. If one were to cut the simulations short and measure the quantum yield at around 250 fs, right after the deactivation to the ground state, the obtained value would be about 0.65. However, the final quantum yield after equilibration is noticeably different (0.56), which closely matches experimental results (typically reporting 0.42–0.55), which means that the inclusion of longer timescale thermal equilibration is necessary to correctly describe this process.

#### Can the multi-state model work across different systems?

Since, by construction, our multi-state model must be able to learn systems of arbitrary composition, we trained it on a data set combining the AL-produced data sets of fulvene, azobenzene, and the molecular ferro-wire. These data sets were labeled with different electronic-structure methods, CASSCF, AIQM1/MRCI-SD, and AIQM1/CIS, respectively. They also had different numbers of electronic states (two in fulvene and azobenzene, and four in the ferro-wire). Regardless of these differences, the model was able to learn from this new data set, and the populations produced with this single multi-state model for each of the molecules are as good as the populations produced with the separate, dedicated multi-state models, in the cases of fulvene and azobenzene, with small deviation observed in the case of the

**Fig. 9 | Extended dynamics of *cis*-azobenzene.** **a** Evolution of the C–N–N–C dihedral angle during the dynamics, in a random subset of 25 trajectories; **b** temporal evolution of the populations of three isomers of azobenzene: *cis*-azobenzene, C–N–N–C dihedral angle of below 40°; intermediate form, dihedral angle between 40 and 140°; *trans*-azobenzene, dihedral angle above 140°; the dotted line shows the average absolute dihedral angle during the dynamics, and the shaded region of the graph indicates the time after which all trajectories decayed to the ground state. Note the logarithmic scale of the X axis.



**Fig. 10 | Electronic state population evolution during the dynamics conducted with a combined model.** **a** fulvene, **b** molecular ferro-wire, **c** *trans* → *cis* dynamics of azobenzene. Full lines represent the populations from ML-TSH dynamics

performed with the model trained on a combined training set, while dashed lines are reference populations and dotted lines are populations taken from simulations with dedicated ML models.

ferro-wire (Fig. 10). This proves that a single MS-ANI model can learn the underlying photochemistry of vastly different systems, which we utilized recently for constructing a universal ML potential for excited states OMNI-P2x,<sup>66</sup> which is based on an extension of the MS-ANI architecture with the all-in-one learning<sup>67</sup> of different electronic-structure levels.

## Discussion

Using ML methods for simulating nonadiabatic molecular dynamics is a formidable challenge due to the necessity to train ML potentials for several electronic states across a wide range of geometries, as well as due to the intrinsic complexity of photoprocesses, requiring accurate predictions in many different regions of the PES. Owing to this, state-of-the-art ML-accelerated protocols for TSH still require extensive human expertise, as well as extreme computational resources.

Herein, we present an end-to-end protocol for active learning of TSH dynamics that requires minimal user input, delivering results of unprecedented quality for different photochemical and photophysical processes. The heart of this protocol is a new MS-ANI model that is able to make predictions for any number of excited states with unprecedented quality. We leverage the idea of physics-informed active learning combined with accelerated sampling using gap-driven trajectories to explore the vicinity of the conical intersections and sampling based on hopping probability. The thresholds for uncertainty quantification in the AL procedure are based on rigorous statistical considerations that minimize arbitrary user decisions.

This all yields a robust protocol that is able to conduct nonadiabatic dynamics investigations of photophysical and photochemical processes using affordable time and resources. The protocol has enabled a larger scale exploration of the photoisomerization of *cis*-azobenzene that uncovered an oscillatory behavior of this reaction that affects the final quantum yields.

Hence, we show that such explorations require longer timescale propagations which are now doable with ML.

Finally, the presented models can simultaneously predict accurate dynamics across very different chemical species with no noticeable decrease in accuracy, making them extendable across chemical space. It remains to be seen how multi-state learning performs across different systems which are more similar in terms of their structure and excited state energies. Encouragingly, multi-state learning has been recently successfully applied to learn across diverse molecular systems leading to the universal OMNI-P2x potential for excited states.<sup>66</sup>

## Methods

All computations were performed with the development version of MLatom<sup>48,49</sup> and additional scripts, which are available in MLatom version 3.10+. Wigner sampling was performed with MLatom routines adapted<sup>48</sup> from Newton-X<sup>42</sup>. The bulk of the computations and implementations reported here were performed on the cloud computing service of the Xiamen Atomistic Computing Suite at <http://XACScloud.com> that supports collaborative work. MS-ANI and ANI models are based on MLatom's interface to the TorchANI<sup>59</sup> package. AIQM1 calculations were performed with the help of MLatom's interfaces to the MNDO program<sup>60</sup> (providing the semi-empirical part<sup>61</sup> of AIQM1), TorchANI (providing ANI-type NN), and dftd4<sup>62</sup> programs. CASSCF calculations were performed through the interface to the COLUMBUS quantum chemistry package<sup>63</sup>.

## Fulvene

For the active learning of fulvene, initial points were sampled from a harmonic approximation Wigner distribution, with a total of 250 points. The maximum propagation time was set to 60 fs with a time step of 0.1 fs. Velocities after hopping were rescaled in the direction of the momentum

vector, using a reduced kinetic energy reservoir<sup>51</sup>. In each iteration of the AL procedure, 50 ML-TSH trajectories were run, with an additional 50 ML-gap-driven trajectories (25 on each electronic surface). From this set, 15 additional points were sampled based on hopping probability uncertainty. Points were sampled in this manner until a threshold of 300 points was reached in each iteration, after which the new models were trained.

Reference calculations were performed using the CASSCF method, with an active space of 6 electrons in 6 orbitals. The same settings were used for AL loops with the single-state ANI model and the MACE model. However, these runs did not include accelerated sampling with gap-driven trajectories or sampling based on hopping probability.

Testing of the model was performed using 1000 trajectories for the MS-ANI model, as well as the single-state ANI models, and 300 trajectories for the MACE models (due to extensive computational costs). 623 CASSCF(6,6) trajectories were used as reference.

## Ferro-wire

In the active learning procedure of the ferro-wire, initial points were sampled from a Wigner distribution using the standard procedure, with a total of 250 points. The maximum propagation time was set to 200 fs with a time step of 0.5 fs. The velocities after hopping were rescaled in the direction of the momentum vector using a reduced kinetic energy reservoir. Points were sampled using the same procedure as for fulvene. Reference calculations were performed using the AIQM1/CIS method with default settings. All trajectories were initialized in the third excited state,  $S_3$ , without filtering. 100 ML-TSH trajectories were propagated with the model trained on the combined fulvene/ferro-wire dataset.

## Azobenzene

In the active learning procedure of azobenzene, initial points were sampled from the Wigner distribution using the standard procedure, with a total of 250 points. The maximum propagation time was set to 1000 fs for trajectories starting from the *trans* isomer and 400 fs for *cis*-azobenzene, with a time step of 0.5 fs. The velocities after hopping were rescaled in the direction of the momentum vector using a reduced kinetic energy reservoir. In each iteration, two sets of trajectories were propagated, one starting from the *cis* isomer and the other starting from the *trans* isomer. Model training was performed after sampling all points resulting from these trajectories, as well as a set of 50 gap-driven trajectories (25 on each active surface), with 15 points added from probability-based sampling without a maximum number of sampled points. In the active learning procedure, all trajectories were initialized in the first excited state without filtering.

Reference calculations were performed using the AIQM1/MRCI-SD method<sup>64</sup>. In the initial, semi-empirical part of AIQM1, the half-electron restricted open-shell Hartree-Fock formalism<sup>65</sup> was used in the SCF step, with the HOMO and LUMO orbitals singly-occupied. Two additional closed-shell references were added in the MRCI procedure, a HOMO–HOMO configuration and a doubly excited, LUMO–LUMO configuration. The active space consisted of 8 electrons in 10 orbitals (four occupied orbitals, six unoccupied). Single and double excitations within a such-defined active space were allowed. 88 trajectories starting from the *trans* isomer, and 92 starting from *cis*-azobenzene were used as reference.

For Wigner sampling of initial conditions, all frequencies smaller than  $100\text{ cm}^{-1}$  were set to  $100\text{ cm}^{-1}$  to avoid sampling unphysical structures<sup>48</sup>. Testing of the model was performed on a set of initial conditions filtered to an excitation window of  $2.53 \pm 0.3\text{ eV}$  for *trans*-azobenzene and  $2.89 \pm 0.3\text{ eV}$  for *cis*-azobenzene, which corresponds to the  $S_0 \rightarrow S_1$  transition energies at the AIQM1 optimized ground-state minima.

In the 3 ps simulations of azobenzene, 304 trajectories were propagated, with 4 removed from the analysis set due to reaching unphysical, highly distorted geometries, leaving a total of 300 trajectories. The source of this error is most likely extending the timescale of the dynamics beyond the timescale of the AL procedure. Nevertheless, as the fraction of trajectories exhibiting this problem is very small (ca. 1%), we believe it has no impact on the end result. Then, at each point of simulations, the conformation of the

molecule was classified into either: *cis*-azobenzene, for a C–N–N–C dihedral angle of below 40°, an intermediate form with the dihedral angle between 40 and 140°, and *trans*-azobenzene, with the dihedral angle above 140°. The populations of these three forms are then averaged out over all of the trajectories.

## Data availability

The data (training sets and ML models) is available under the open-source MIT license at <https://github.com/dralgroup/al-namd>.

## Code availability

The code is available in the open-source MLatom under the MIT license as described at <https://github.com/dralgroup/mlatom>.

Received: 12 October 2024; Accepted: 10 April 2025;

Published online: 13 May 2025

## References

1. Crespo-Otero, R. & Barbatti, M. Recent advances and perspectives on nonadiabatic mixed quantum-classical dynamics. *Chem. Rev.* **118**, 7026–7068 (2018).
2. Martyka, M. & Jankowska, J. Nonadiabatic molecular dynamics study of a complete photoswitching cycle for a full-size diarylethene system. *J. Photochem. Photobiol. A* **438**, 114513 (2023).
3. Iwasa, N. & Nanbu, S. Investigating the photo-relaxation mechanism of 6-azauracil through ab initio nonadiabatic molecular dynamics simulations. *Chem. Phys. Lett.* **838**, 141088 (2024).
4. Jankowska, J. & Sobolewski, A. L. Photo-oxidation of methanol in complexes with pyrido[2,3-b]pyrazine: a nonadiabatic molecular dynamics study. *Phys. Chem. Chem. Phys.* **26**, 5296–5302 (2024).
5. Zhou, G., Lu, G. & Prezhdo, O. V. Modeling Auger processes with nonadiabatic molecular dynamics. *Nano Lett.* **21**, 756–761 (2021).
6. Giannini, S. et al. Exciton transport in molecular organic semiconductors boosted by transient quantum delocalization. *Nat. Commun.* **13**, 2755 (2022).
7. Behler, J., Reuter, K. & Scheffler, M. Nonadiabatic effects in the dissociation of oxygen molecules at the Al (111) surface. *Phys. Rev. B* **77**, 115421 (2008).
8. 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.* **9**, 2725–2732 (2018).
9. Dral, P. O., Barbatti, M. & Thiel, W. Nonadiabatic excited-state dynamics with machine learning. *J. Phys. Chem. Lett.* **9**, 5660–5663 (2018).
10. 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.* **9**, 6702–6708 (2018).
11. Westermayr, J. et al. Machine learning enables long time scale molecular photodynamics simulations. *Chem. Sci.* **10**, 8100–8107 (2019).
12. Chu, W., Saidi, W. A. & Prezhdo, O. V. Long-lived hot electron in a metallic particle for plasmonics and catalysis: Ab initio nonadiabatic molecular dynamics with machine learning. *ACS Nano* **14**, 10608–10615 (2020).
13. Westermayr, J., Gastegger, M. & Marquetand, P. Combining schnet and sharc: the schnarc machine learning approach for excited-state dynamics. *J. Phys. Chem. Lett.* **11**, 3828–3834 (2020).
14. Li, J., Stein, R., Adrion, D. M. & Lopez, S. A. Machine-learning photodynamics simulations uncover the role of substituent effects on the photochemical formation of cubanes. *J. Am. Chem. Soc.* **143**, 20166–20175 (2021).
15. Posenitskiy, E., Spiegelman, F. & Lemoine, D. On application of deep learning to simplified quantum-classical dynamics in electronically excited states. *Mach. Learn. Sci. Technol.* **2**, 035039 (2021).

16. Axelrod, S., Shakhnovich, E. & Gómez-Bombarelli, R. Excited state non-adiabatic dynamics of large photoswitchable molecules using a chemically transferable machine learning potential. *Nat. Commun.* **13**, 3440 (2022).
17. Bai, Y., Vogt-Maranto, L., Tuckerman, M. E. & Glover, W. J. Machine learning the Hohenberg-Kohn map for molecular excited states. *Nat. Commun.* **13**, 7044 (2022).
18. Sit, M. K., Das, S. & Samanta, K. Semiclassical dynamics on machine-learned coupled multireference potential energy surfaces: application to the photodissociation of the simplest criegee intermediate. *J. Phys. Chem. A* **127**, 2376–2387 (2023).
19. Žugec, I. et al. Understanding the photoinduced desorption and oxidation of Co on Ru (0001) using a neural network potential energy surface. *JACS Au* **4**, 1997–2004 (2024).
20. Dral, P. O. & Barbatti, M. Molecular excited states through a machine learning lens. *Nat. Rev. Chem.* **5**, 388–405 (2021).
21. Westermayr, J. & Marquetand, P. Machine learning for electronically excited states of molecules. *Chem. Rev.* **121**, 9873–9926 (2021).
22. Li, J. & Lopez, S. A. Machine learning accelerated photodynamics simulations. *Chem. Phys. Rev.* **4**, 031309 (2023).
23. Zhang, L., Ullah, A., Pinheiro M. Jr., Dral, P. O. & Barbatti, M. *Excited-state dynamics with machine learning*, 329–353 (Elsevier, 2023).
24. Mausenberger, S. et al. Spainn: equivariant message passing for excited-state nonadiabatic molecular dynamics. *Chem. Sci.* **15**, 15880–15890 (2024).
25. Westermayr, J. et al. Deep learning study of tyrosine reveals that roaming can lead to photodamage. *Nat. Chem.* **14**, 914–919 (2022).
26. Li, J. et al. Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chem. Sci.* **12**, 5302–5314 (2021).
27. Li, J. & Lopez, S. A. Excited-state distortions promote the photochemical  $4\pi$ -electrocyclizations of fluorobenzenes via machine learning accelerated photodynamics simulations. *Chem. Eur. J.* **28**, e202200651 (2022).
28. Wang, L., Salguero, C., Lopez, S. A. & Li, J. Machine learning photodynamics uncover blocked non-radiative mechanisms in aggregation-induced emission. *Chem* **10**, 2295–2310 (2024).
29. Xu, H. et al. Ultrafast photocontrolled rotation in a molecular motor investigated by machine learning-based nonadiabatic dynamics simulations. *J. Phys. Chem. A* **127**, 7682–7693 (2023).
30. Westermayr, J. & Marquetand, P. Deep learning for UV absorption spectra with SchNarc: first steps toward transferability in chemical compound space. *J. Chem. Phys.* **153**, 154112 (2020).
31. Pios, S. V., Gelin, M. F., Ullah, A., Dral, P. O. & Chen, L. Artificial-intelligence-enhanced on-the-fly simulation of nonlinear time-resolved spectra. *J. Phys. Chem. Lett.* **15**, 2325–2331 (2024).
32. Wang, L., Li, Z. & Li, J. Balancing wigner sampling and geometry interpolation for deep neural networks learning photochemical reactions. *Artif. Intell. Chem.* **1**, 100018 (2023).
33. Du, L. & Lan, Z. An on-the-fly surface-hopping program jade for nonadiabatic molecular dynamics of polyatomic systems: Implementation and applications. *J. Chem. Theory Comput.* **11**, 1360–1374 (2015).
34. Westermayr, J., Faber, F. A., Christensen, A. S., von Lilienfeld, O. A. & Marquetand, P. Neural networks and kernel ridge regression for excited states dynamics of  $\text{ch}_2\text{nh}_2^+$ : From single-state to multi-state representations and multi-property machine learning models. *Mach. Learn.: Sci. Technol.* **1**, 025009 (2020).
35. Ullah, A., Huang, Y., Yang, M. & Dral, P. O. Physics-informed neural networks and beyond: enforcing physical constraints in quantum dissipative dynamics. *Digit. Discov.* **3**, 2052–2060 (2024).
36. 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.* **8**, 3192–3203 (2017).
37. Zheng, P., Zubatyuk, R., Wu, W., Isayev, O. & Dral, P. O. Artificial intelligence-enhanced quantum chemical method with broad applicability. *Nat. Commun.* **12**, 7022 (2021).
38. Liu, J. & Thiel, W. An efficient implementation of semiempirical quantum-chemical orthogonalization-corrected methods for excited-state dynamics. *J. Chem. Phys.* **148**, 154103 (2018).
39. Wang, T. Y., Neville, S. P. & Schuurman, M. S. Machine learning seams of conical intersection: a characteristic polynomial approach. *J. Phys. Chem. Lett.* **14**, 7780–7786 (2023).
40. Mukherjee, S. & Barbatti, M. Ultrafast internal conversion without energy crossing. *Results Chem.* **4**, 100521 (2022).
41. Hou, Y.-F., Zhang, L., Zhang, Q., Ge, F. & Dral, P. O. Physics-informed active learning for accelerating quantum chemical simulations. *J. Chem. Theory Comput.* **20**, 7744–7754 (2024).
42. Barbatti, M. et al. Newton-x platform: New software developments for surface hopping and nuclear ensembles. *J. Chem. Theory Comput.* **18**, 6851–6865 (2022).
43. Hou, Y.-H., Dral, P. O. & Zhang, Q. Surprising dynamics phenomena in Diels-Alder reaction of  $\text{c}_{60}$  uncovered with AI. *J. Org. Chem.* **89**, 15041–15047 (2024).
44. Landau, L. On the theory of transfer of energy at collisions I. *Physikalische Z. der Sowjetunion* **1**, 88 (1932).
45. Landau, L. On the theory of transfer of energy at collisions II. *Physikalische Z. der Sowjetunion* **2**, 46 (1932).
46. Belyaev, A. K. & Lebedev, O. V. Nonadiabatic nuclear dynamics of atomic collisions based on branching classical trajectories. *Phys. Rev. A* **84**, 014701 (2011).
47. Zener, C. & Fowler, R. H. Non-adiabatic crossing of energy levels. *Proc. R. Soc. Lond. Ser. A Contain. Pap. A Math. Phys. Character* **137**, 696–702 (1932).
48. Zhang, L. et al. MLATOM software ecosystem for surface hopping dynamics in Python with quantum mechanical and machine learning methods. *J. Chem. Theory Comput.* **20**, 5043–5057 (2024).
49. Dral, P. O. et al. Mlatom 3: A platform for machine learning-enhanced computational chemistry simulations and workflows. *J. Chem. Theory Comput.* **20**, 1193–1213 (2024).
50. T do Casal, M., Toldo, J. M., Pinheiro Jr, M. & Barbatti, M. Fewest switches surface hopping with Baeck-An couplings. *Open. Res. Eur.* **1**, 49 (2022).
51. Toldo, J. M., Mattos, R. S., Pinheiro, M. J., Mukherjee, S. & Barbatti, M. Recommendations for velocity adjustment in surface hopping. *J. Chem. Theory Comput.* **20**, 614–624 (2024).
52. Batatia, I., Kovács, D. P., Simm, G. N. C., Ortner, C. & Csányi, G. MACE: Higher order equivariant message passing neural networks for fast and accurate force fields. *Advances in Neural Information Processing Systems* (2022).
53. Merritt, I. C. D., Jacquemin, D. & Vacher, M. *cis* → *trans* photoisomerisation of azobenzene: a fresh theoretical look. *Phys. Chem. Chem. Phys.* **23**, 19155–19165 (2021).
54. Schnack-Petersen, A. K., Pápai, M. & Møller, K. B. Azobenzene photoisomerization dynamics: Revealing the key degrees of freedom and the long timescale of the *trans*-to-*cis* process. *J. Photochem. Photobiol. A* **428**, 113869 (2022).
55. Rouxel, J. R. et al. Coupled electronic and nuclear motions during azobenzene photoisomerization monitored by ultrafast electron diffraction. *J. Chem. Theory Comput.* **18**, 605–613 (2022).
56. Zimmerman, G., Chow, L.-Y. & Paik, U.-J. The photochemical isomerization of azobenzene1. *J. Am. Chem. Soc.* **80**, 3528–3531 (1958).
57. Ladányi, V. et al. Azobenzene photoisomerization quantum yields in methanol redetermined. *Photochem. Photobiol. Sci.* **16**, 1757–1761 (2017).
58. Weingart, O., Lan, Z., Koslowski, A. & Thiel, W. Chiral pathways and periodic decay in *cis*-azobenzene photodynamics. *J. Phys. Chem. Lett.* **2**, 1506–1509 (2011).

59. Gao, X., Ramezanghorbani, F., Isayev, O., Smith, J. S. & Roitberg, A. E. Torchani: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials. *J. Chem. Inf. Model.* **60**, 3408–3415 (2020).
60. Thiel, Walter, with contributions from M. Beck, S. Billeter, R. Kevorkiants, M. Kolb, A. Koslowski, S. Patchkovskii, A. Turner, E.-U. Wallenborn, W. Weber, L. Spörkel, and P. O. Dral. MNDO2020: a semiempirical quantum chemistry program. Max-Planck-Institut für Kohlenforschung (2020).
61. Dral, P. O., Wu, X. & Thiel, W. Semiempirical Quantum-Chemical Methods with Orthogonalization and Dispersion Corrections. *J. Chem. Theory Comput.* **15**, 1743–1760 (2019).
62. Caldeweyher, E., Ehlert, S. & Grimme, S. DFT-D4, version 2.5.0. Mulliken Center for Theoretical Chemistry, University of Bonn (2020).
63. Lischka, H. et al. The generality of the Guga MRCI approach in Columbus for treating complex quantum chemistry. *J. Chem. Phys.* **152**, 134110 (2020).
64. Koslowski, A., Beck, M. E. & Thiel, W. Implementation of a general multireference configuration interaction procedure with analytic gradients in a semiempirical context using the graphical unitary group approach. *J. Comput. Chem.* **24**, 714–726 (2003).
65. Dewar, M. J. S., Hashmall, J. A. & Venier, C. G. Ground states of conjugated molecules. IX. Hydrocarbon radicals and radical ions. *J. Am. Chem. Soc.* **90**, 1953–1957 (1968).
66. Martyka, M., Tong, X.-Y., Jankowska, J. & Dral, P. O. OMNI-P2x: A Universal Neural Network Potential for Excited-State Simulations. Preprint on ChemRxiv. <https://doi.org/10.26434/chemrxiv-2025-j207x> (2025).
67. Chen, Y. & Dral, P. O. All-in-one foundational models learning across quantum chemical levels. Preprint on ChemRxiv. <https://doi.org/10.26434/chemrxiv-2024-ng3ws> (2024).

## Acknowledgements

M.M. and J.J. acknowledge the Polish Ministry of Education and Science for funding this research under the program “Perły Nauki,” grant number PN/01/0064/2022, amount of funding, and the total value of the project: 239 800,00 PLN, as well as gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Centers: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2024/017363. M.B. thanks the funding provided by the European Research Council (ERC) Advanced grant SubNano (Grant agreement 832237). M.B. received support from the French government under the France 2030 as part of the initiative d’Excellence d’Aix-Marseille Université, A\*MIDEX (AMX-22-IN1-48). P.O.D. acknowledges funding from the National Natural Science Foundation of China (via the Outstanding Youth Scholars (Overseas, 2021) project) and via the Lab project of the State Key Laboratory of Physical Chemistry of Solid Surfaces. The computations were performed using the XACS cloud computing resources. The authors also acknowledge Max Pinheiro Jr, Prateek Goel, and Bao-Xin Xue for many non-published tests of not-so-successful protocols.

## Author contributions

M.M. co-wrote the original manuscript draft, co-designed the AL protocols, did the final implementations (multi-state model, gapMD, AL, modifications to TSH code), and an original implementation of the multi-state model, performed all the simulations with the multi-state model, analyzed results, visualized results. L.Z. made original implementations and tests of AL protocols, performed simulations, and analyzed results with the single-state ANI model. Discussion with F.G. led to the idea of the multi-state model. Y.F.H. performed the simulations with MACE. J.J. contributed to the result analysis and interpretation, co-supervised research, and secured funding. M.B. co-designed and co-conceived the project, co-designed AL protocols and gapMD, contributed to the result analysis and interpretation, co-supervised research, and secured funding. P.O.D. co-wrote the original manuscript draft, co-designed and co-conceived the project, co-designed AL protocols and gapMD, made intermediate implementations and tests of AL protocols, did original implementation of gapMD, contributed to the result analysis and interpretation, co-supervised research, and secured funding. All authors discussed the results and revised the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41524-025-01636-z>.

**Correspondence** and requests for materials should be addressed to Joanna Jankowska, Mario Barbatti or Pavlo O. Dral.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025