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

Mikołaj Martyka,[†] Lina Zhang,[‡] Fuchun Ge,[‡] Yi-Fan Hou,[‡] Joanna Jankowska,[*,†]
Mario Barbatti,[*,¶,§] and Pavlo O. Dral[*,‡,‖]

†*Faculty of Chemistry, University of Warsaw, Pasteura 1, Warsaw, 02-093, Poland*
‡*State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Xiamen University, Xiamen, Fujian 361005, China*
¶*Aix Marseille University, CNRS, ICR, Marseille, France*
§*Institut Universitaire de France, 75231 Paris, France*
‖*Institute of Physics, Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University in Toruń, ul. Grudziadzka 5, 87-100 Toruń, Poland*

E-mail: jjankowska@chem.uw.edu.pl; mario.barbatti@univ-amu.fr; dral@xmu.edu.cn

Abstract

We present a robust protocol for affordable learning of the electronic-state manifold to accelerate photophysical and photochemical molecular simulations. The protocol solves several pertinent issues precluding the widespread use of machine learning (ML) in excited-state simulations. We introduce a novel physics-informed multi-state

https://doi.org/10.26434/chemrxiv-2024-dtc1w **ORCID:** https://orcid.org/0000-0002-2975-9876

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 with established ML potentials. We also present gap-driven dynamics for meticulous accelerated sampling of the small-gap regions: which proves crucial for stable surface-hopping dynamics. Put together, multi-state learning and gap-driven dynamics enable efficient active learning furnishing robust models for surface-hopping simulations. 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. The thresholds for uncertainty quantification are automatically chosen based on statistical and physical considerations. The protocol will be made available with the next release of the open-source MLatom as described at `https://github.com/dralgroup/al-namd`.

Introduction

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.[1] The latter provide an invaluable computational tool for investigating complex photoprocesses in the real-time domain. As of today, by far the most popular nonadiabatic dynamics simulations technique is trajectory surface hopping (TSH). It has been successfully used to study a wide range of photoresponsive systems.[2–6]

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 instantenous hoppings, whose probability is evaluated at each integration time step on the bases of 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 for performing hundreds of thousands or millions of single-point electronic-structure calculations. Hence, a significant effort has been put into developing machine learning (ML) protocols to accelerate the TSH simulations of photoprocesses.[7–23]

Indeed, ML allowed breaking through the limitations of the electronic-structure TSH simulations by enabling large-scale computations for longer time, 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 even detected, with the pure electronic-structure calculations.[24–29] Despite all this progress, the state-of-the-art ML-accelerated TSH is still 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 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 2%).

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. The learning of this manifold, along with all the intrinsic correlations with high precision is required for robust ML TSH. The suggested solutions include creating single-state ML models (i.e., one ML model per each electronic state, Figure 1a) 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).[30] 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 often have larger errors compared to single-state models.[30] Learning excited-state PESs across different molecules is another grand challenge, and only a few studies[16,31] attempted to do so.

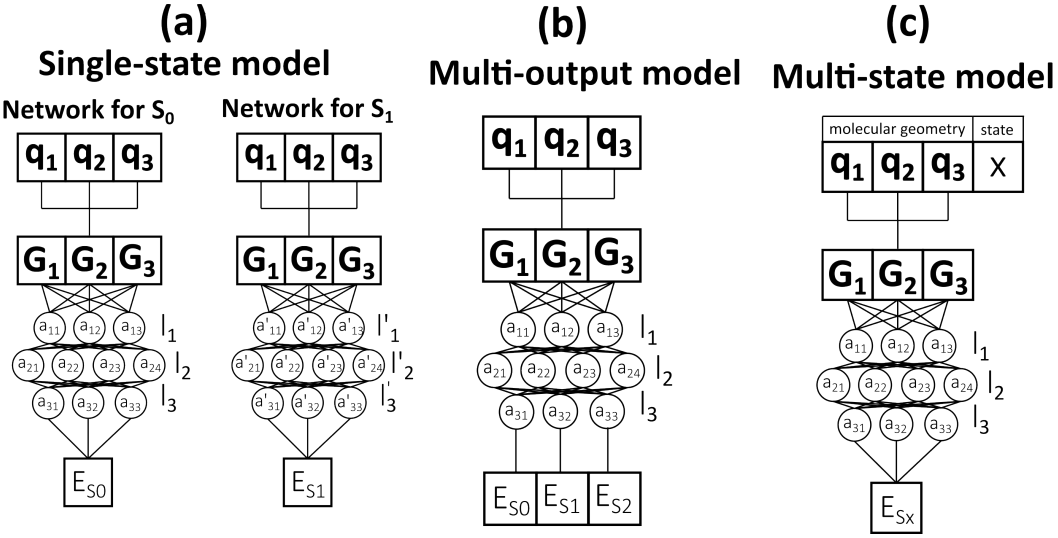


Figure 1: A comparison of different NN models that can be used for excited state properties prediction, using a triatomic molecule with coordinates $\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$: (a) a combination of single output NNs; (b) multi-output NN; (c) multi-state NN. The vectors $\mathbf{G}_i$ are the molecular descriptors of a given atom, l_i are the network layers with nodes a_{ij}, and E_{Sx} is the final predicted energy of adiabatic state x.

Having a good ML model architecture alone is insufficient: 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.[8,10,32] 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.[11,25,26] Even then, these strategies are often insufficient for robust ML TSH and require further interventions, such as interpolation between critical PES points (e.g., between minima and conical intersections).[26,27,33]

Another issue is the need for easy-to-use, end-to-end protocols and software enabling routine simulations, although progress is being made in this direction too.[13,14,34]

In this work, we solve these daunting problems by: i) developing accurate and extendable physics-informed multi-state models (Figure 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 we can obtain the final simulation results within days on commodity hardware. 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 which satisfies the following requirements: 1) learning an arbitrary number of electronic states on equal footing with high accuracy; 2) learning and making predictions for different molecules; 3) capturing the required correlations between states and, especially, correctly reproducing energy gaps between surfaces; 4) being fast to train and evaluate. The core idea is captured in Figure 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 (cf. Figure 1 a and b). The paramount benefit is that this architecture can treat 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. 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.[30] In the multi-output models, on the other hand, only the last output layer differentiates between states, but the correlation between states is implicitly learned to some extent as the weights are shared in the preceding layers.[35] Eventually, it may lead to better performance in ML TSH, despite the larger errors in energies.[30]

In addition, our multi-state models are ideally suited for incorporating crucial physics information, such as the importance of energy gaps between states. This importance is underappreciated in the ML TSH research, but we found that the correct, special treatment of the 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_{gap} for the gap when training the multi-state models:

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

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

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

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

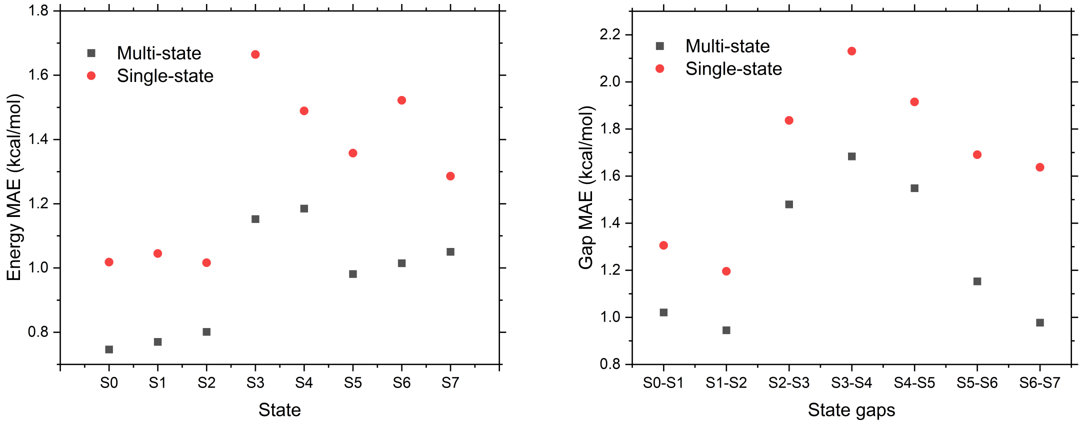
The energy and force losses follow their usual definitions of $||E^{\text{ML}} - E^{\text{ref}}||^2$ and $||\mathbf{F}^{\text{ML}} + \mathbf{F}^{\text{ref}}||^2$, respectively. Each of the loss terms comes with the corresponding weight, ω, with ω_E and ω_{gap} equal to 1, treating the gaps and energies as equally important, and with ω_F set to the standard value of 0.1.

This makes our multi-state model a typical representative of the physics-informed NNs, which were shown to provide more qualitatively accurate behavior in the related quantum dynamics context.[35]

The overall NN architecture is based on the established ANI-type network, which has a good balance between cost and accuracy,[36] 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 call this particular architecture MS-ANI ("MS" for "multi-state").

To demonstrate the performance of the multi-state model, firstly, we evaluate its accuracy in the energies and energy gaps for the first eight electronic states of pyrene. For this, we construct the data set with 4000 conformations generated from Wigner sampling around the S_0 minimum. All calculations were performed with AIQM1[37] using CIS[38] treatment for the excited-state properties. 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) in energies close to ca. 1 kcal/mol (0.04 eV) (Figure 2a) and root-mean-squared errors (RMSEs) below 1.6 kcal/mol (Table S1 of the ESI). The gaps were also described with good accuracy, with MAEs below 1.7 kcal/mol (Figure 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 ANI-type models. We observe that for all electronic states the errors in both, energies and gaps, are substantially higher than in the multi-state model, with differences most pronounced for higher excited states (Figure 2, as well as Tables S1 and S2). 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 requires much less time than training eight separate single-state



(a) (b)

Figure 2: Performance of the multi-state (black squares) and single-state (red dots) models for predicting the energies and energy gaps of the first 8 electronic states of pyrene judged by mean absolute errors (MAEs).

models.

We also observe for the first time that the accuracy in energies of the ground state is improved upon the inclusion of the excited-state information. Also, our multi-state model can learn the excited states 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, sufficient amount of data to sample all relevant PESs regions 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 often have insufficient representation of the critical region in the vicinity of the conical intersections because the TSH trajectories contain only very few points in that region. 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.[8,9,32] To mitigate this issue, specially designed protocols, such as interpolation from minimum energy conical intersection points[25–27] were occasionally employed.

Here, we have developed an accelerated sampling approach to meticulously chart the small-gap regions in the direct vicinity of a conical intersection. We use this approach to improve the robustness, and to accelerate the active learning loop described in the next section, but we envision that it can be used also in other contexts, such as accelerating the search for conical intersections. Our approach samples the points relevant to the TSH via propagating special, gap-driven molecular dynamics (gapMD) trajectories (Figure 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, ΔE, larger than 0.03 Hartree, the trajectories are propagated along the energy gap gradient, $\boldsymbol{\nabla} E_{\text{gap}}$, defined as the difference between energy gradients in the upper, ($\boldsymbol{\nabla} E_{\text{upper}}$), and lower, ($\boldsymbol{\nabla} E_{\text{lower}}$), surfaces:

$$\boldsymbol{\nabla} E_{\text{gap}} = \boldsymbol{\nabla} E_{\text{upper}} - \boldsymbol{\nabla} E_{\text{lower}}, \tag{3}$$

with equivalent expression formulated in terms of forces:

$$\mathbf{F}_{\text{gap}} = \mathbf{F}_{\text{upper}} - \mathbf{F}_{\text{lower}}. \tag{4}$$

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.[39] Once the region of a smaller gaps 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: i) propagating exclusively with the energy gap gradients will sample many irrelevant points, e.g., to complete dissociation of a molecule as the dissociated structures have degenerate energy levels with a zero gap, ii) in actual TSH, trajectories are propagated either on the upper, or lower surface, and we want to sample the relevant phase space. Hence, our trajectories propagated with the energy-gap

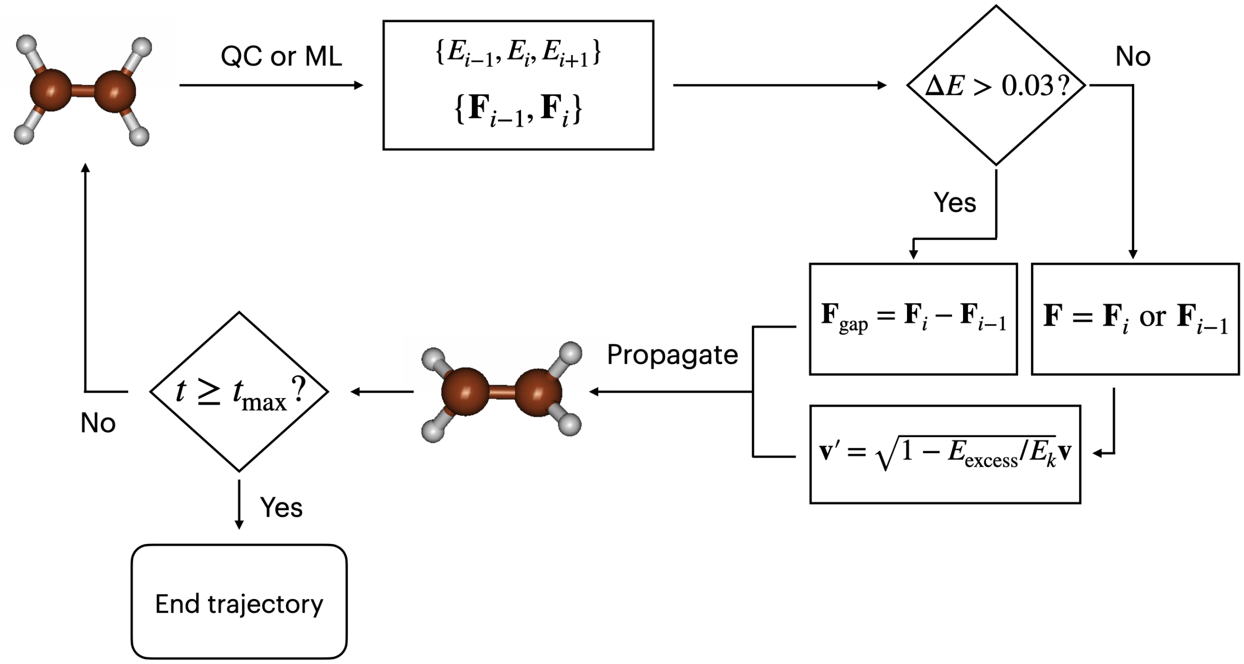


Figure 3: Gap dynamics flowchart.

gradients are only for biasing dynamics to visit the small-gap region faster, to 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_{excess} is greater 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:

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

The updated velocities $\mathbf{v}'$ are used in the further propagation. The excess energy, E_{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, E_{excess}, is greater than the kinetic energy, E_k, then there is not enough kinetic energy to compensate for this excess; hence, in such 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 (Figure 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.

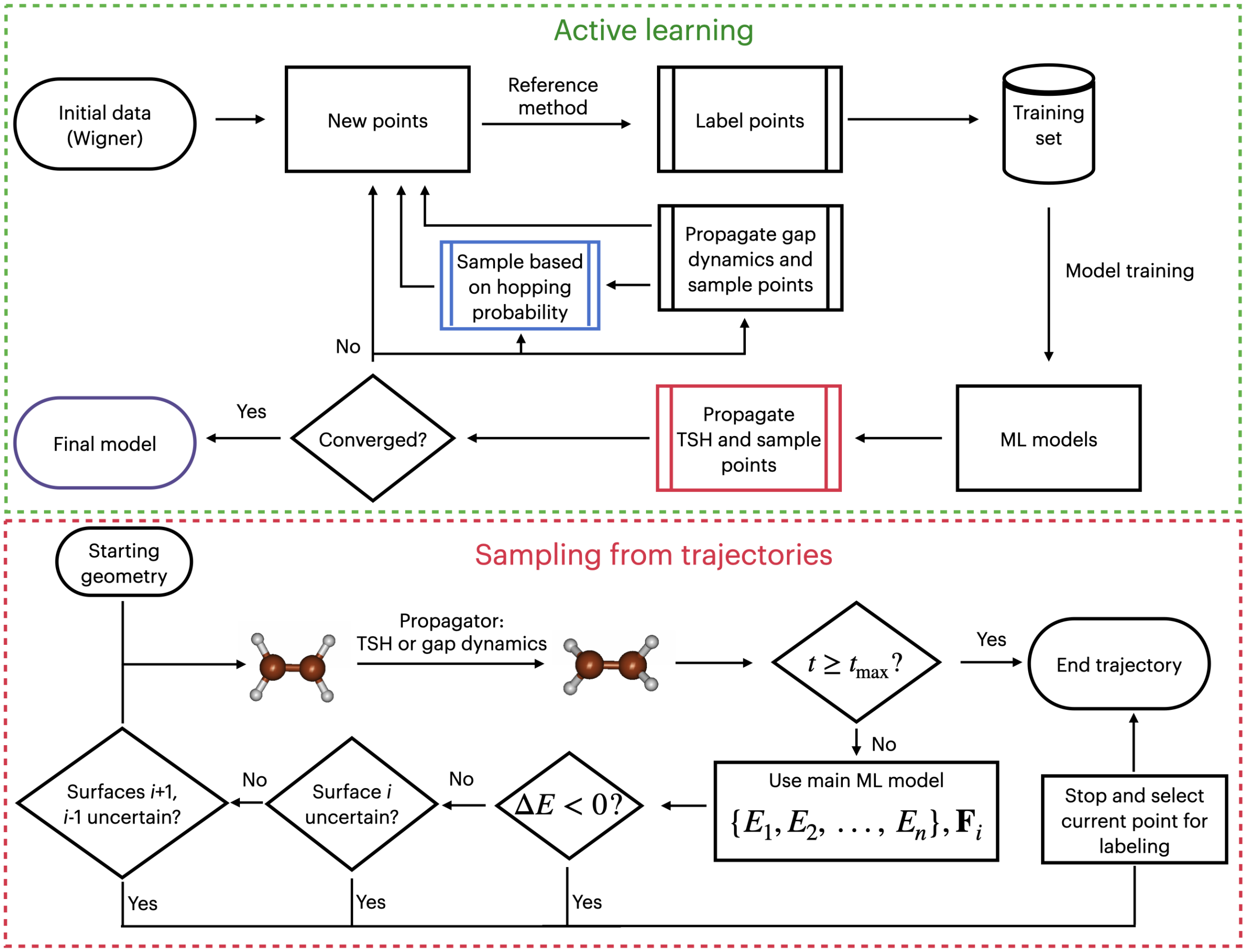


Figure 4: Active learning flowchart.

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.[40] 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,[41] 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 are sampled, we start the AL loop by training our multi-state model on both, energies and energy gradients of all electronic states, and propagating many (50 in this work) ML-TSH trajectories started from the Wigner-sampled initial conditions. 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 (Figure 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 the uncertainty quantification (UQ) metric for the current surface is exceeding the threshold, which ensures the quality of the relevant adiabatic surfaces on which the dynamics is propagated. Finally, we check whether the UQ for the surfaces up and down the current surface are exceeding 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 in a direct way, 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:[40] we evaluate the absolute deviations in energy predictions by 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): $U = |E^{\text{aux}} - E^{\text{main}}|$. 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 was shown before to produce robust active learning for ground-state dynamics.[40,42]

The UQ thresholds are calculated based on statistical considerations for each state separately, in the analogous way as before in the ground-state AL.[40] They are evaluated with the UQ values calculated for the initial validation set (10% of the initial data):

$$\text{UQ}_{\text{threshold}} = M(\mathbf{U}) + 3 \cdot \text{MAD}(\mathbf{U}) \tag{6}$$

where M is the median, and MAD is the median absolute deviation. 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.[11,21,22]

None of the above sampling criteria directly addresses the key factor controlling surface hopping: the hopping probability. This factor is very sensitive to any deviations in the energy gap and having a model, that yields accurate probabilities, is essential for ensuring the 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 models at each time step before the UQ threshold was exceeded. We use the same formula as in the Landau–Zener–Belyaev–Lebedev (LZBL) formulation of TSH:[43–47]

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

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[47,48] 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 are randomly sampled per AL iteration to ensure a balanced training set.

To better sample points in the direct vicinity of the conical intersection, we spawn ML-gapMD trajectories starting from a random time step of each ML-TSH trajectory (see the section on gap-driven dynamics). Starting 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, nor 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 asses 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 has been 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. The performance of the model in ML-TSH shows excellent agreement with the reference quantum-chemical results but in terms of the populations and the distribution of the geometric parameters at the $S_1 \to S_0$ hopping points (Figure 5a, c, and d, as well as Table 1). The latter are classified into three groups, following the Refs. 49,50 based on the $\mathrm{C{=}CH_2}$ bond length, and the mean dihedral angle $\phi_{\mathrm{C=CH_2}}$ between the 5-membered ring and the methylene group:

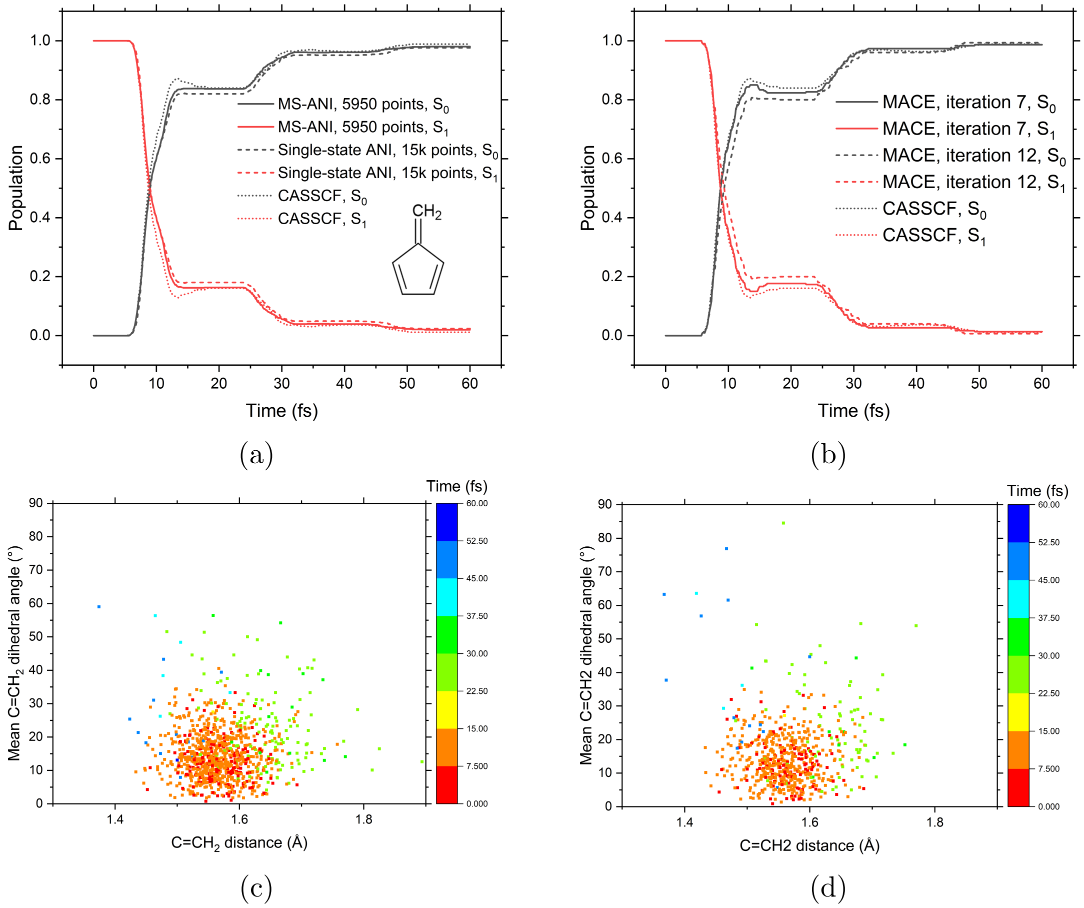
$$\phi_{\mathrm{C=CH_2}} = \frac{1}{4}\Big(|\phi^{\mathrm{cis}_1}_{\mathrm{C=CH_2}}| + |\phi^{\mathrm{cis}_2}_{\mathrm{C=CH_2}}| + |\phi^{\mathrm{trans}_1}_{\mathrm{C=CH_2}}| + \phi^{\mathrm{trans}_2}_{|\mathrm{C=CH_2}}|\Big). \tag{8}$$

The planar group is defined with $\phi_{\mathrm{C=CH_2}} < 30°$, twisted-stretched — $\phi_{\mathrm{C=CH_2}} > 30°$ and $\mathrm{C{=}CH_2} > 1.55$ Å, and twisted shrunk — $\phi_{\mathrm{C=CH_2}} > 30°$, and $\mathrm{C{=}CH_2} < 1.55$ Å.

The obtained data is also in good agreement with previous works,[49,50] 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 a high-quality ML-TSH dynamics? Indeed, we tried many different settings over eight years of research, and



(a) (b)

(c) (d)

Figure 5: Summary of the nonadiabatic molecular dynamics simulations of fulvene. Panel (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). Panel (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). Panels (c) and (d) show correlation plots of the C=CH2 distance and the mean dihedral angle at the $S_1 \rightarrow S_0$ hopping points in ML-TSH dynamics (c) and reference CASSCF dynamics (d).

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
Population at 20 fs (%)	16.3 ± 2.9	16.9 ± 2.3
Population at 40 fs (%)	3.7 ± 1.5	3.9 ± 1.2
Planar hopping (%)	94.0 ± 1.8	93.8 ± 1.5
Twisted-stretched hopping (%)	3.7 ± 1.5	4.1 ± 1.2
Twisted-shrunk hopping	2.3 ± 1.2	2.15 ± 0.90

none showed the 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, but 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 model and without gapMD and probability uncertainty evaluation, we can also obtain a good population plot for fulvene (Figures 5a, no checks for negative gaps were performed either). The problem is, however, that the procedure converged only after a whopping 104 days, 159 iterations, and 15000 training points (using NVIDIA GeForce RTX 3080 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.[51] 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 (Figures 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 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 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 characterizes with a complicated electronic structure due to charge transfer between its three structural units. Amazingly, our AL 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 (Figure 6). Testing of the model was performed using 200 ML-TSH trajectories and compared to 100 reference AIQM1/CIS trajectories.

In this case, propagating one ML-TSH trajectory takes about 5 minutes 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.

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 and the molecular ferro-wire. These data sets were labeled with different electronic-structure methods, CASSCF and AIQM1/CIS, respectively. They also had different numbers of electronic states (two in fulvene, and four in the ferro-wire). Regardless these differences, the model was able to learn from this new data set, and the populations produced with this

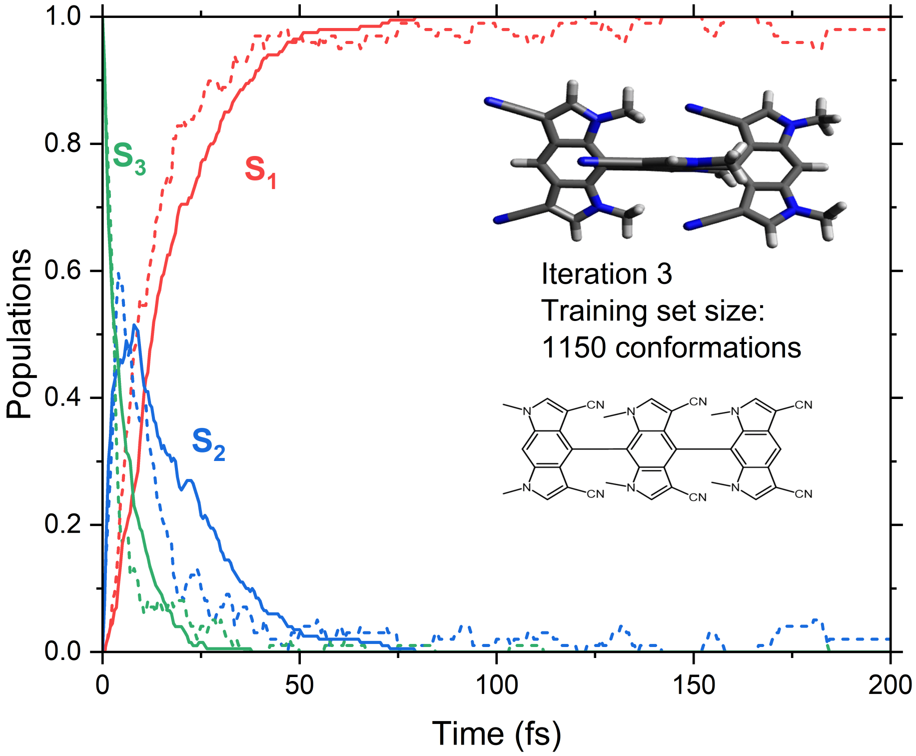
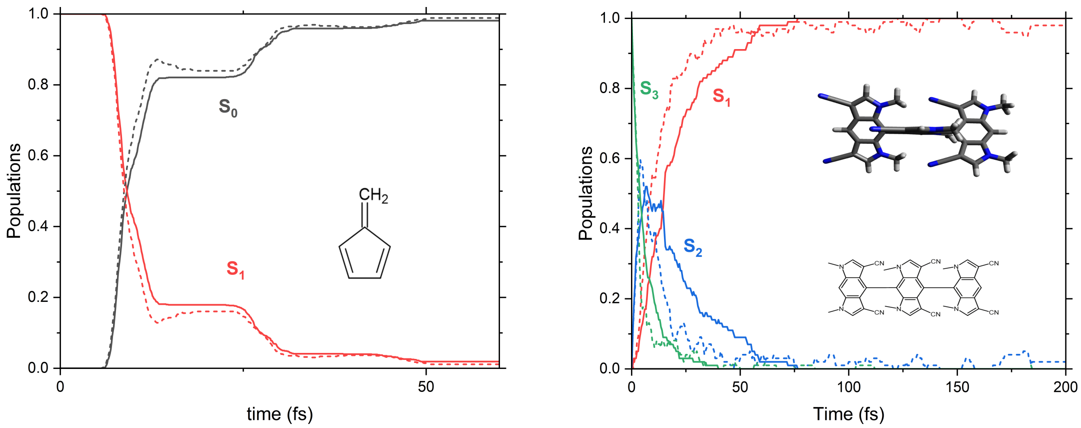


Figure 6: Electronic state population evolution during the dynamics of the molecular ferro-wire, as predicted with the ML model (full lines), and reference AIQM1/CIS calculations (dashed lines).

single multi-state model for each of the molecules are as good as the populations produced with the separate, dedicated multi-state models (Figure 7).



(a) (b)

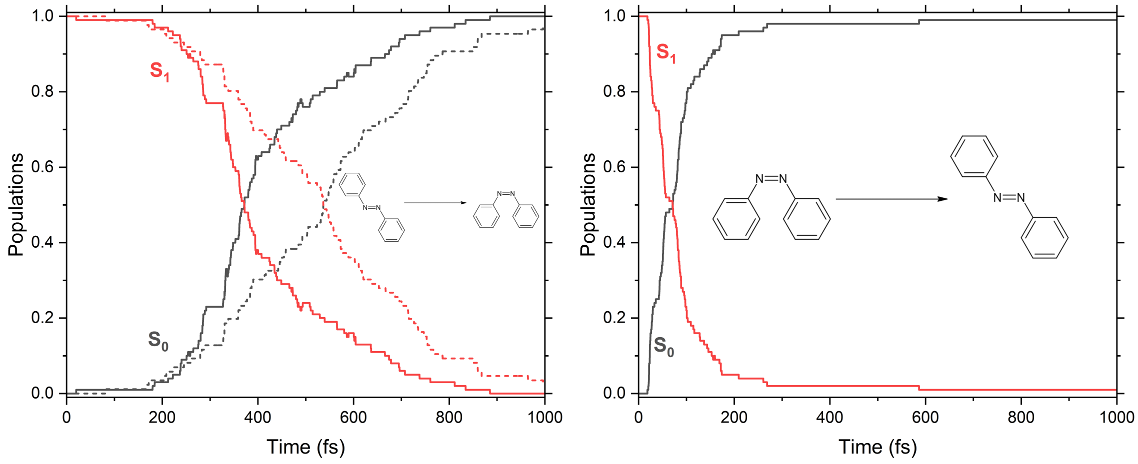
Figure 7: Electronic state population evolution during the dynamics conducted with a combined fulvene plus molecular ferro-wire model for (a) fulvene and (b) molecular ferro-wire. Full lines represent the populations from ML-TSH dynamics, while dashed lines are reference populations.

Can the protocol learn photoreactions?

Finally, we would like to see whether the proposed end-to-end protocol works for a more complicated case of a photochemical transformation. For this, we take as a test case the well-known photoisomerization of azobenzene.[52–54]

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[40] from the AL-based search of conformers in their ground state). After running for 3 days of wall-clock time, with 33 iterations (about 6000 training points), the model is 26% converged – the best convergence rate at the time of writing (we are still running the AL for this system and will update the results). This non-converged model can still predict the photochemistry with reasonable accuracy and below we show the preliminary analysis.

Analyzing the electronic state populations during the dynamics (depicted in Figure 8), we see acceptable agreement between the ML model and reference AIQM1/MR-CISD trajectories in the *trans* → *cis* isomerization reaction. The excited-state lifetimes can be extracted by fitting an exponential function to the ground-state population rise, with $P(t) = A(1 - \exp(-t/\tau))$, where τ is the photoprocess timescale, and A is associated with the infinite-time population limit. After fitting, we obtain: 492 ± 10 fs for the ML-TSH dynamics, and 705 ± 22 fs for reference AIQM1/MRCI-SD, showing that, at the current AL convergence rate, the ML model overestimates the speed of this photoprocess. While we do not have yet the reference trajectories for the *cis* → *trans* process, the overall shape of the decay curves is in reasonable agreement with literature data,[52] showing faster deactivation of the *cis* isomer. Particularly, the fitted timescale of 79 ± 1 fs is in good agreement with Ref. 52, reporting a timescale of 74 fs at the CASSCF level.



(a) (b)

Figure 8: Electronic state population evolution during the isomerization dynamics of azobenzene. Trajectories shown in panel (a) started from the *trans* isomer, while trajectories starting from the *cis* isomer are shown in panel (b). ML-TSH are represented by full lines, while reference AIQM1/MRCI-SD trajectories are dashed. Note that these results are shown for 26% convergence rate of AL and we will update them later.

As the next point in the analysis, we can look at the predicted quantum yields of the photoisomerization processes. The dihedral angle C_1–N_2–N_3–C_4 (as shown in Figure 9) can be used as a convenient descriptor of the formed product, taking the value of approximately

0 degree for *cis*-azobenzene, and 180 degrees for *trans*-azobenzene. The distribution of the final dihedral angles can be found in Figure 9. Taking a dihedral angle of fewer than 60 degrees 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 Φ are calculated as the fraction of the trajectories relaxing to form the photoproduct (the other isomer), N_{reactive}, to the total number of trajectories N_{traj}: $\Phi = N_{\mathrm{reactive}}/N_{\mathrm{traj}}$.

Here, the model shows excellent agreement with results obtained by the reference method, as well as the experimental results.[55,56]The ML-TSH dynamics predict the quantum yields as $\Phi^{\mathrm{ML}}_{trans \to cis} = 0.19 \ \pm\ 0.08$ and $\Phi^{\mathrm{ML}}_{trans \to cis} = 0.19 \ \pm\ 0.08$. 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 ± 0.09), as well as experimental results, predicting a quantum yield of 0.2 to 0.36 for the *trans* to *cis* isomerization, and 0.42 to 0.55 for the reverse process. This remarkable accuracy in the vicinity of the conical intersection can be attributed to the inclusion of the gap-driven trajectories in the sampling process.

Conclusions and outlook

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

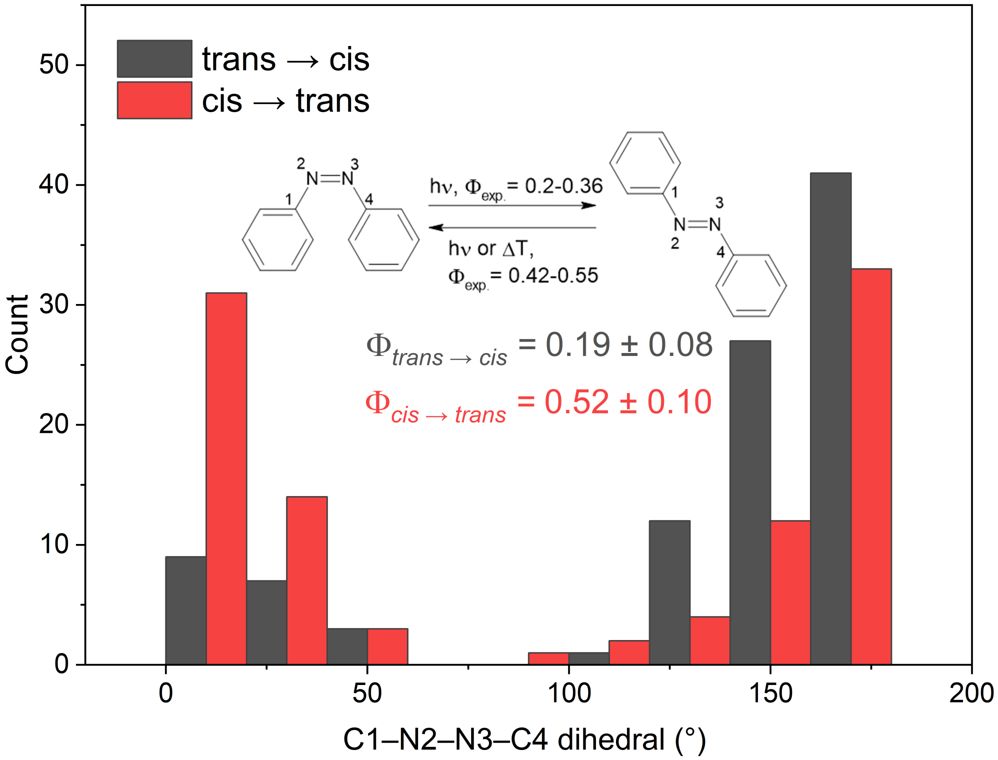


Figure 9: Distribution of the C–N–N–C dihedral angle of azobenzene at the trajectories final points. Points from trajectories starting from the *trans* isomer of azobenzene are marked in black, while geometries starting from the *cis* isomer of azobenzene are marked in red.

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, minimizing arbitrary, user decisions.

This all yields a robust protocol able to conduct nonadiabatic dynamics investigation of photophysical and photochemical processes using affordable time and resources. Additionally, 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.

Computational details

All computations were performed with the development version of MLatom[47,48] and additional scripts which will be integrated in the next release of MLatom. Wigner sampling performed with MLatom routines adapted[47] from Newton-X.[41] Major bulk of computations and implementations reported here were performed on the cloud computing service of the Xiamen Atomistic Computing Suite at `http://XACScloud.com` supporting collaborative work. MS-ANI and ANI models are based on MLatom's interface to the TorchANI[57] package. AIQM1 calculations were performed with the help of MLatom's interfaces to the MNDO[58] (providing the semi-empirical part[59] of AIQM1 and CIS), TorchANI (providing ANI-type NN), and dftd4[60] programs. CASSCF calculations were performed through the interface to the COLUMBUS quantum chemistry package.[61]

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.[50] 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 with a time step of 0.1 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/MR-CISD method.[62] In the initial, semi-empirical part of AIQM1, the half-electron restricted open-shell Hartree-Fock formalism[63] 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.

For Wigner sampling of initial conditions, all frequencies smaller than 100 cm^{-1} were set to 100 cm^{-1}, to avoid sampling unphysical structures.[47] Testing of the model was performed on a set of initial conditions filtered to an excitation window of 2.53 ± 0.3 eV for *trans*-azobenzene, and 2.89 ± 0.3 eV for *cis*-azobenzene, which corresponds to the $S_0 \rightarrow S_1$ transition energies at the AIQM1 optimized ground-state minima.

Data availability

The data (training sets and ML models) will be made available under the open-source MIT license as described at `https://github.com/dralgroup/al-namd`.

Code availability

The code will be made available with the next release of the open-source MLatom under the MIT license as described at `https://github.com/dralgroup/al-namd`.

Authors 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.

Acknowledgement

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 by the National Natural Science Foundation of China (via the Outstanding Youth Scholars (Overseas, 2021) project). 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.

References

(1) Crespo-Otero, R.; Barbatti, M. Recent Advances and Perspectives on Nonadiabatic Mixed Quantum-Classical Dynamics. *Chem. Rev.* **2018**, *118*, 7026–7068.

(2) Martyka, M.; Jankowska, J. Nonadiabatic molecular dynamics study of a complete photoswitching cycle for a full-size diarylethene system. *J. Photochem. Photobiol. A* **2023**, *438*, 114513.

(3) Iwasa, N.; Nanbu, S. Investigating the photo-relaxation mechanism of 6-azauracil through ab initio nonadiabatic molecular dynamics simulations. *Chem. Phys. Lett.* **2024**, *838*, 141088.

(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.* **2024**, *26*, 5296–5302.

(5) Zhou, G.; Lu, G.; Prezhdo, O. V. Modeling Auger Processes with Nonadiabatic Molecular Dynamics. *Nano Lett.* **2021**, *21*, 756–761.

(6) Giannini, S.; Peng, W.-T.; Cupellini, L.; Padula, D.; Carof, A.; Blumberger, J. Exciton transport in molecular organic semiconductors boosted by transient quantum delocalization. *Nat. Commun.* **2022**, *13*, 2755.

(7) Behler, J.; Reuter, K.; Scheffler, M. Nonadiabatic effects in the dissociation of oxygen molecules at the Al(111) surface. *Phys. Rev. B* **2008**, *77*, 115421.

(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.* **2018**, *9*, 2725–2732.

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

(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.* **2018**, *9*, 6702–6708.

(11) Westermayr, J.; Gastegger, M.; Menger, M. F. S. J.; Mai, S.; González, L.; Marquetand, P. Machine learning enables long time scale molecular photodynamics simulations. *Chem. Sci.* **2019**, *10*, 8100–8107.

(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* **2020**, *14*, 10608–10615.

(13) Westermayr, J.; Gastegger, M.; Marquetand, P. Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics. *J. Phys. Chem. Lett.* **2020**, *11*, 3828–3834.

(14) Li, J.; Reiser, P.; Boswell, B. R.; Eberhard, A.; Burns, N. Z.; Friederich, P.; Lopez, S. A. Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chem. Sci.* **2021**, *12*, 5302–5314.

(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.* **2021**, *2*, 035039.

(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.* **2022**, *13*, 3440.

(17) Bai, Y.; Vogt-Maranto, L.; Tuckerman, M. E.; Glover, W. J. Machine learning the Hohenberg-Kohn map for molecular excited states. *Nat. Commun.* **2022**, *13*, 7044.

(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* **2023**, *127*, 2376–2387.

(19) Žugec, I.; Tetenoire, A.; Muzas, A. S.; Zhang, Y.; Jiang, B.; Alducin, M.; Juaristi, J. I. Understanding the Photoinduced Desorption and Oxidation of CO on Ru(0001) Using a Neural Network Potential Energy Surface. *JACS Au* **2024**, *4*, 1997–2004.

(20) Dral, P. O.; Barbatti, M. Molecular excited states through a machine learning lens. *Nat. Rev. Chem.* **2021**, *5*, 388–405.

(21) Westermayr, J.; Marquetand, P. Machine Learning for Electronically Excited States of Molecules. *Chem. Rev.* **2021**, *121*, 9873–9926.

(22) Li, J.; Lopez, S. A. Machine learning accelerated photodynamics simulations. *Chem. Phys. Rev.* **2023**, *4*, 031309.

(23) Zhang, L.; Ullah, A.; Pinheiro Jr, M.; Dral, P. O.; Barbatti, M. In *Quantum chemistry in the age of machine learning*; Dral, P. O., Ed.; Elsevier, 2023; pp 329–353.

(24) Westermayr, J.; Gastegger, M.; Voros, D.; Panzenboeck, L.; Joerg, F.; Gonzalez, L.;

Marquetand, P. Deep learning study of tyrosine reveals that roaming can lead to photodamage. *Nat. Chem.* **2022**, *14*, 914–919.

(25) 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.* **2021**, *143*, 20166–20175.

(26) Li, J.; Reiser, P.; Boswell, B. R.; Eberhard, A.; Burns, N. Z.; Friederich, P.; Lopez, S. A. Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chem. Sci.* **2021**, *12*, 5302–5314.

(27) Li, J.; Lopez, S. A. Excited-State Distortions Promote the Photochemical 4π-Electrocyclizations of Fluorobenzenes via Machine Learning Accelerated Photodynamics Simulations. *Chem. Eur. J.* **2022**, *28*, e202200651.

(28) Wang, L.; Salguero, C.; Lopez, S. A.; Li, J. Machine learning photodynamics uncover blocked non-radiative mechanisms in aggregation-induced emission. *Chem* **2024**, *10*, 2295–2310.

(29) Xu, H.; Zhang, B.; Tao, Y.; Xu, W.; Hu, B.; Yan, F.; Wen, J. Ultrafast Photocontrolled Rotation in a Molecular Motor Investigated by Machine Learning-Based Nonadiabatic Dynamics Simulations. *J. Phys. Chem. A* **2023**, *127*, 7682–7693.

(30) 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 CH2NH2+: From single-state to multi-state representations and multi-property machine learning models. *Mach. Learn.: Sci. Technol.* **2020**, *1*, 025009.

(31) Westermayr, J.; Marquetand, P. Deep learning for UV absorption spectra with SchNarc: First steps toward transferability in chemical compound space. *J. Chem. Phys.* **2020**, *153*, 154112.

(32) 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.* **2024**, *15*, 2325–2331.

(33) Wang, L.; Li, Z.; Li, J. Balancing Wigner sampling and geometry interpolation for deep neural networks learning photochemical reactions. *Artif. Intell. Chem.* **2023**, *1*, 100018.

(34) 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.* **2015**, *11*, 1360–1374.

(35) Ullah, A.; Huang, Y.; Yang, M.; Dral, P. O. Physics-Informed Neural Networks and Beyond: Enforcing Physical Constraints in Quantum Dissipative Dynamics. 2024; `https://arxiv.org/abs/2404.14021`.

(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.* **2017**, *8*, 3192–3203.

(37) Zheng, P.; Zubatyuk, R.; Wu, W.; Isayev, O.; Dral, P. O. Artificial intelligence-enhanced quantum chemical method with broad applicability. *Nat. Commun.* **2021**, *12*, 7022.

(38) Liu, J.; Thiel, W. An efficient implementation of semiempirical quantum-chemical orthogonalization-corrected methods for excited-state dynamics. *J. Chem. Phys.* **2018**, *148*, 154103.

(39) Mukherjee, S.; Barbatti, M. Ultrafast internal conversion without energy crossing. *Results Chem.* **2022**, *4*, 100521.

(40) Hou, Y.-F.; Zhang, L.; Zhang, Q.; Ge, F.; Dral, P. O. Physics-informed active learning for accelerating quantum chemical simulations. *arXiv:2404.11811* **2024**, `https://arxiv.org/abs/2404.11811`.

(41) Barbatti, M. et al. Newton-X Platform: New Software Developments for Surface Hopping and Nuclear Ensembles. *J. Chem. Theory Comput.* **2022**, *18*, 6851–6865.

(42) Yi-Fan Hou, P. O. D., Quanhao Zhang Surprising new dynamics phenomena in Diels–Alder reaction of C_{60} uncovered with AI. *ChemRxiv* **2024**, `https://doi.org/10.1021/acs.jpclett.4c00107`.

(43) Landau, L. On the Theory of Transfer of Energy at Collisions I. *Physikalische Zeitschrift der Sowjetunion* **1932**, *1*, 88.

(44) Landau, L. On the Theory of Transfer of Energy at Collisions II. *Physikalische Zeitschrift der Sowjetunion* **1932**, *2*, 46.

(45) Belyaev, A. K.; Lebedev, O. V. Nonadiabatic nuclear dynamics of atomic collisions based on branching classical trajectories. *Phys. Rev. A* **2011**, *84*, 014701.

(46) Zener, C.; Fowler, R. H. Non-adiabatic crossing of energy levels. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* **1932**, *137*, 696–702.

(47) Zhang, L.; Pios, S. V.; Martyka, M.; Ge, F.; Hou, Y.-F.; Chen, Y.; Chen, L.; Jankowska, J.; Barbatti, M.; Dral, P. O. MLatom Software Ecosystem for Surface Hopping Dynamics in Python with Quantum Mechanical and Machine Learning Methods. *J. Chem. Theory Comput.* **2024**, *20*, 5043–5057.

(48) Dral, P. O. et al. MLatom 3: A Platform for Machine Learning-Enhanced Computational Chemistry Simulations and Workflows. *J. Chem. Theory Comput.* **2024**, *20*, 1193–1213.

(49) T do Casal, M.; Toldo, J. M.; Pinheiro, M., Jr; Barbatti, M. Fewest switches surface hopping with Baeck-An couplings. *Open. Res. Eur.* **2022**, *1*, 49.

(50) Toldo, J. M.; Mattos, R. S.; Pinheiro, M. J.; Mukherjee, S.; Barbatti, M. Recommendations for Velocity Adjustment in Surface Hopping. *J. Chem. Theory Comput.* **2024**, *20*, 614–624.

(51) 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. 2023.

(52) Merritt, I. C. D.; Jacquemin, D.; Vacher, M. cis → trans photoisomerisation of azobenzene: a fresh theoretical look. *Phys. Chem. Chem. Phys.* **2021**, *23*, 19155–19165.

(53) 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* **2022**, *428*, 113869.

(54) Rouxel, J. R.; Keefer, D.; Aleotti, F.; Nenov, A.; Garavelli, M.; Mukamel, S. Coupled Electronic and Nuclear Motions during Azobenzene Photoisomerization Monitored by Ultrafast Electron Diffraction. *J. Chem. Theory Comput.* **2022**, *18*, 605–613.

(55) Zimmerman, G.; Chow, L.-Y.; Paik, U.-J. The Photochemical Isomerization of Azobenzene1. *J. Am. Chem. Soc.* **1958**, *80*, 3528–3531.

(56) Ladányi, V.; Dvořák, P.; Al Anshori, J.; Vetráková, L.; Wirz, J.; Heger, D. Azobenzene photoisomerization quantum yields in methanol redetermined. *Photochem. Photobiol. Sci.* **2017**, *16*, 1757–1761.

(57) 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.* **2020**, *60*, 3408–3415.

(58) 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.

(59) Dral, P. O.; Wu, X.; Thiel, W. Semiempirical Quantum-Chemical Methods with Orthogonalization and Dispersion Corrections. *J. Chem. Theory Comput.* **2019**, *15*, 1743–1760.

(60) Caldeweyher, E.; Ehlert, S.; Grimme, S. DFT-D4, Version 2.5.0. 2020.

(61) Lischka, H. et al. The generality of the GUGA MRCI approach in COLUMBUS for treating complex quantum chemistry. *J. Chem. Phys.* **2020**, *152*, 134110.

(62) 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.* **2003**, *24*, 714–726.

(63) 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.* **1968**, *90*, 1953–1957.

Supporting Information for ”Charting electronic-state manifolds across molecules with multi-state learning and gap-driven dynamics via efficient and robust active learning”

Mikolaj Martyka[1], Lina Zhang[2], Fuchun Ge[2], Yi-Fan Hou[2], Joanna Jankowska[⋆,1],
Mario Barbatti[⋆,3,4], and Pavlo O. Dral[⋆,2,5]

[1]Faculty of Chemistry, University of Warsaw, Pasteura 1, Warsaw, 02-093, Poland
[2]State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, Xiamen University, Xiamen, Fujian 361005, China
[3]Aix Marseille University, CNRS, ICR, Marseille, France
[4]Institut Universitaire de France, 75231 Paris, France
[5]Institute of Physics, Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University in Toruń, ul. Grudziadzka 5, 87-100 Toruń, Poland

Contents

S1 Performance of MS-ANI

Table S1: Performance of the MS-ANI and single-state ANI models for predicting the energies of the first eight electronic states of pyrene, kcal/mol.

State	Single-state RMSE	Multi-state RMSE	Single-state MAE	Multi-state MAE
S_0	1.32538	0.96832	1.01809	0.74647
S_1	1.33954	1.00046	1.04498	0.77016
S_2	1.29239	1.02189	1.01633	0.80134
S_3	2.1493	1.47939	1.6647	1.15242
S_4	1.92178	1.49865	1.48902	1.18519
S_5	1.72372	1.23457	1.35766	0.98109
S_6	1.96994	1.28371	1.52209	1.0147
S_7	1.64697	1.33066	1.28604	1.05033

Table S2: Performance of the MS-ANI and single-state ANI models for predicting the energy gaps of the first eight electronic states of pyrene, kcal/mol.

States	Single-state RMSE	Multi-state RMSE	Single-state MAE	Multi-state MAE
S_0 - S_1	1.72368	1.32452	1.30568	1.02068
S_1 - S_2	1.54319	1.23684	1.1957	0.94517
S_2 - S_3	2.38496	1.89723	1.83622	1.47973
S_3 - S_4	2.73568	2.09397	2.13051	1.68351
S_4 - S_5	2.47677	1.93942	1.91491	1.54836
S_5 - S_6	2.19074	1.42764	1.69054	1.15271
S_6 - S_7	2.1447	1.22756	1.63729	0.97736