Supporting Information

for

"Nonadiabatic Excited-State Dynamics with Machine Learning"

Pavlo O. Dral,[*,†] Mario Barbatti,[*,‡] and Walter Thiel[*,†]

[†] *Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany*

[‡] *Aix Marseille Univ, CNRS, ICR, Marseille, France*

**E-mail: dral@kofo.mpg.de. Website: dr-dral.com.*

**E-mail: mario.barbatti@univ-amu.fr. Website: www.barbatti.org.*

**E-mail: thiel@kofo.mpg.de.*

Table of Contents

1. Methods

1.1. Fewest Switches Surface Hopping

NA-MQC dynamics forms a class of methods to approximately solve the time-dependent Schrödinger equation for an electronically excited molecular system.[1] In these methods, the nuclei are treated as a swarm of classical trajectories, while the electrons are treated quantum-mechanically. Nonadiabatic information (the coupling between different adiabatic electronic states induced by nuclear motion) is provided in diverse ways depending on the specific method. The classical trajectories are usually (but not necessarily[2]) independent from each other. NA-MQC dynamics methods are well suited for on-the-fly propagation, in which electronic properties (potential energies, potential energy gradients, and state couplings) are computed for the nuclear geometry of the classical trajectory as the time propagation goes. Therefore, they do not require precomputed or fitted multidimensional surfaces, which allows considering the dynamics of all nuclear degrees of freedom.

In this paper, we employ one of the most well-known on-the-fly NA-MQC methods, the surface hopping dynamics with decoherence-corrected[3] fewest switches[4] (DC-FSSH). The latest advances in and main limitations of NA-MQC dynamics have been recently reviewed in Ref. 5. Here, we only present the core features of DC-FSSH such that the explanation of its integration with machine learning is self-contained.

In surface hopping, Newton's equations for each nucleus k with mass M_k

$$\frac{d^2\mathbf{R}_k}{dt^2} = -\frac{1}{M_k}\frac{\partial V_L}{\partial \mathbf{R}_k} \tag{S1}$$

are solved on a single Born-Oppenheimer potential energy surface (adiabatic state L with potential energy V_L), while the nonadiabatic transition probabilities between state L and any other state J nearby are simultaneously computed. At each time step, a stochastic algorithm decides whether the propagation should remain on L or switch to J. The sequence of nuclear geometries $\mathbf{R}(t)$ generated in this way is called a trajectory.

The integration of Eq. (S1) requires the force $-\partial V_L / \partial \mathbf{R}_k$ acting on the nuclei α due to energy gradient of state L. In our implementation, this integration is done with the popular velocity-Verlet algorithm[6] at time steps Δt.

In the DC-FSSH version of surface hopping, transition probabilities are computed for each time step by integrating a local approximation of the time-dependent Schrödinger equation (TDSE)

$$\frac{dc_J}{dt} = -\frac{i}{\hbar}V_J c_J - \sum_K \sigma_{JK}^{NAC} c_K \tag{S2}$$

This approximation is *local* because the time-dependent coefficients c_J are computed only at the position $\mathbf{R}(t)$ of the classical trajectory. The integration of the local TDSE requires knowledge of the potential energies of each state J and the couplings between J and K

$$\sigma_{JK}^{NAC}(\mathbf{R}) \equiv \left\langle \psi_J \middle| \frac{\partial \psi_K}{\partial t} \right\rangle = \sum_k \mathbf{F}_{JK}^k \cdot \mathbf{v}_k \tag{S3}$$

where ψ_J is the electronic wave function of state *J*,

$$\mathbf{F}_{JK}^k \equiv \left\langle \psi_J \middle| \frac{\partial \psi_K}{\partial \mathbf{R}_k} \right\rangle \tag{S4}$$

is the nonadiabatic coupling vector, and $\mathbf{v}_k$ is the classical velocity of nucleus k. In the examples discussed later, the local TDSE was integrated with the Butcher algorithm[7] to the 6th order, with time step $\Delta\tau = \Delta t / m_s$, where m_s is an integer. This choice of $\Delta\tau$ is motivated by the fact that the local TDSE requires much smaller time steps than the classical equations. The integration of the local TDSE between two classical steps is done using interpolated electronic quantities.

The locality of the TDSE is known to lead to overcoherence,[8] with a superposition of several electronic states holding over time, rather than selecting a specific state, as expected. Overcoherence may lead to bad probability estimates, and several algorithms have been proposed to fix it.[8-10] Here, we use the simplified decay of mixing (SDM),[3] which transforms the coefficients obtained from Eq. (S2) according to

$$\begin{aligned} c_J^{DC} &= c_J e^{-\Delta t/\tau_{JL}}, \quad \forall J \neq L, \\ c_L^{DC} &= \frac{c_L}{|c_L|}\left[1 - \sum_{K \neq L} \left|c_K^{DC}\right|^2\right]^{1/2}. \end{aligned} \tag{S5}$$

In the SDM, the decoherence time τ_{JL} is given by

$$\frac{1}{\tau_{JL}} = \frac{|V_J - V_L|}{\hbar}\left(C + \frac{\varepsilon}{K_n}\right)^{-1} \tag{S6}$$

where K_n is the classical kinetic energy of the nuclei. C and ε are parameters whose recommended values are 1 and 0.1 Hartree, respectively.[11]

The decoherence-corrected coefficients are then used to compute the FSSH probabilities according to

$$P_{L\to J}^{FSSH} = \max\left[0, \frac{-2\Delta\tau}{\left|c_L^{DC}\right|^2} \mathrm{Re}\left(\sigma_{LJ}^{NAC} c_J^{DC} c_L^{DC*}\right)\right] \tag{S7}$$

The stochastic algorithm samples a uniform [0,1] random number r_t, which is compared to the computed probabilities. The trajectory hops from *L* to *J* if two conditions are satisfied. First,

$$\sum_{K=1}^{J-1} P_{L\to K}^{FSSH} < r_t \leq \sum_{K=1}^{J} P_{L\to J}^{FSSH}. \tag{S8}$$

Second, the total energy after hopping is not larger than before. The second condition ensures that a hop from a lower to a higher surface is not allowed if it requires more energy than available in the system. In our calculations, if a hopping is frustrated by the second condition, the trajectory continues in the same state, with the same momentum. Hops usually occur for finite energy gaps ΔV_{LJ}. To ensure energy conservation, an amount of linear momentum corresponding to this gap is added after hopping, along the direction of the nonadiabatic coupling vectors.

The aim of surface hopping is to emulate the wave packet propagation by a swarm of independent trajectories. Therefore, the procedure outlined above needs to be repeated many times to produce many trajectories, which are then statistically analysed. Because of the stochastic nature of the algorithm, even when starting from the same initial conditions, two trajectories are usually different, as they hop at different times.

The DC-FSSH approach is expected to work well if 1) the light pulse is much shorter than the dynamics; 2) the nuclear momentum is large; and 3) there is no strong interference between different parts of the wave packet.[12]

All DC-FSSH trajectories in this study were run for 2 ps (if not specified otherwise) with time steps $\Delta t = 0.2$ fs, and $m_s = 20$ using the development version of the Newton-X program package.[13-14] The initial conditions (starting geometries and velocities) for the dynamics with the 33-D A-SBH were generated by sampling randomly from a 2 ps ground-state dynamics run.

1.2. Adiabatic Spin-Boson Hamiltonian

Dynamics of dissipative systems can be simulated with the spin-boson Hamiltonian (SBH),[15] which in the diabatic basis is given as

$$H = H_s + H_b + H_{sb}. \tag{S9}$$

This Hamiltonian has contributions from the system (s), the bath (b), and the system-bath coupling (sb). For the two-state problem studied here, these terms are defined as follows.

The system Hamiltonian is

$$H_s = \sigma_z \varepsilon_0 - \sigma_x v_0, \tag{S10}$$

written in terms of the Pauli matrices

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{S11}$$

In the isolated system, the two electronic levels are separated by $2\varepsilon_0$ and coupled by v_0. The bath is defined in terms of N independent harmonic oscillators

$$H_b = \mathbf{I}\frac{1}{2}\sum_{k=1}^{N}\left(\frac{P_k^2}{M_k} + M_k \omega_k^2 R_k^2\right) \tag{S12}$$

each with mass M_k and frequency ω_k, oscillating along the coordinate R_k with momentum P_k. $\mathbf{I}$ is the 2×2 identity matrix. The system and the bath are linearly coupled through

$$H_{sb} = \sigma_z \sum_{k=1}^{N} g_k R_k \,, \tag{S13}$$

where g_k is a coupling constant.

For surface hopping, it is more convenient to treat this model in the adiabatic basis.[16] The potential energy surfaces are

$$V_J = \frac{1}{2}\sum_{k=1}^{N} M_k \omega_k^2 R_k^2 + (-1)^J \left[\eta^2 + v_0^2\right]^{1/2}, \quad (J = 1,2), \tag{S14}$$

where

$$\eta = \left(\sum_{k=1}^{N} g_k R_k + \varepsilon_0\right). \tag{S15}$$

The energy gradient components are

$$\frac{\partial V_J}{\partial R_k} = M_k \omega_k^2 R_k + (-1)^J g_k \left[\frac{\eta}{\left[\eta^2 + v_0^2\right]^{1/2}}\right], \quad (k = 1 \cdots N), \tag{S16}$$

and the nonadiabatic coupling components are

$$F_{12}^{k} = -F_{21}^{k} = -\frac{1}{2} g_k \left[\frac{v_0}{\eta^2 + v_0^2} \right]. \tag{S17}$$

Surface hopping with the spin-boson Hamiltonian is discussed in Refs. 16-19. In our implementation in Newton-X, energies, energy gradients, and nonadiabatic couplings are computed for the set of coordinates $\{R_k\}$ using Eqs. (S14), (S16), and (S17). The parameters ε_0 and v_0, the number of oscillators N, their masses $\{M_k\}$ and frequencies $\{\omega_k\}$, and the system-bath couplings $\{g_k\}$ are provided by the user.

The largest A-SBH model considered here consists of 33 oscillators and has parameters chosen to roughly resemble the topography of the S_1/S_0 states of cytosine. Therefore, the masses and frequencies of this 33 dimensional (33-D) model were taken equal to those calculated for cytosine at B3LYP[20-25]/6-31G(d)[26-37] with Gaussian 09[38] (Table S1). The coupling v_0 was arbitrarily set to 800 cm^{-1} (ca. 0.1 eV), ε_0 was set to an arbitrary value of 12,000 cm^{-1}. g_1 and g_2 were set to 0.012 Hartree/a$_0$, and g_k for all other modes were set to 0.001 Hartree/a$_0$.

Table S1. Normal frequencies and reduced masses of cytosine in the ground state at B3LYP/6-31G(d) and system-bath couplings $\{g_k\}$.

Mode	Frequency (cm^{-1})	Reduced Mass (amu)	g_k (Hartree/a$_0$)
1	138.2378	5.5516	0.012
2	203.879	5.9372	0.012
3	326.1611	1.2151	0.001
4	359.2826	3.4224	0.001
5	393.2188	2.8741	0.001
6	525.6653	1.783	0.001
7	533.3969	3.0695	0.001
8	548.8424	2.7842	0.001
9	577.1353	9.1378	0.001
10	630.4229	1.2068	0.001
11	732.2183	1.7023	0.001
12	760.5861	2.7665	0.001
13	769.6856	7.6244	0.001
14	771.343	9.3183	0.001
15	926.0104	3.3787	0.001
16	955.0711	1.2839	0.001
17	990.7741	4.5736	0.001
18	1104.1897	1.9972	0.001
19	1134.2905	1.4871	0.001
20	1223.8565	1.318	0.001
21	1263.4521	3.8957	0.001
22	1367.2161	1.8242	0.001
23	1453.6857	2.4315	0.001
24	1522.0054	2.9114	0.001
25	1581.3994	5.1055	0.001
26	1659.507	1.3666	0.001
27	1708.1024	4.6716	0.001
28	1820.0429	9.5456	0.001
29	3216.0061	1.0902	0.001
30	3239.3832	1.0975	0.001
31	3589.0573	1.0462	0.001
32	3617.1036	1.0787	0.001
33	3711.0983	1.103	0.001

1.3. Machine Learning

In this study we use an ML approach based on kernel ridge regression (KRR), which is easy to implement and use. It has been successfully applied for predicting many quantum chemical properties with good accuracy.[39-46]

As explained above, performing DC-FSSH nonadiabatic dynamics usually requires energies, energy gradients, and nonadiabatic couplings for a given set of coordinates. Thus, we estimate each of these components (V_J, $\frac{\partial V_J}{\partial R_k}$, F_{12}^k) individually using KRR:

$$Y^{\mathrm{est}}\left(\mathbf{R}_p\right)=\sum_{m=1}^{N_{\mathrm{tr}}}\alpha_m K\left(\mathbf{R}_p,\mathbf{R}_m\right) \tag{S18}$$

where Y^{est} is an estimate of one of the above components for the p-th set of coordinates $\{R_k\}$ defined by a vector $\mathbf{R}_p$, $\{\alpha_m\}$ are regression coefficients corresponding to the m-th set of coordinates $\{R_k\}$ defined by a vector $\mathbf{R}_m$ of the training set, $K\left(\mathbf{R}_p,\mathbf{R}_m\right)$ is the kernel, and N_{tr} is the number of training set points. The dimensionality of the input vectors $\mathbf{R}$ for ML is defined by the number of independent oscillators N. In this study we use the kernel of the Matérn type defined as follows:[47]

$$K\left(\mathbf{R}_p,\mathbf{R}_m\right)=\exp\left(-\frac{\left\|\mathbf{R}_p-\mathbf{R}_m\right\|_2}{\sigma}\right)\sum_{l=0}^{n}\frac{(n+l)!}{(2n)!}\binom{n}{l}\left(2\frac{\left\|\mathbf{R}_p-\mathbf{R}_m\right\|_2}{\sigma}\right)^{n-l} \tag{S19}$$

where n is an integer parameter, σ is the kernel width, $\left\|\mathbf{R}_p-\mathbf{R}_m\right\|_2$ is the Euclidean (L^2) norm of the difference vector $\mathbf{R}_p-\mathbf{R}_m$ (the Euclidean distance). In our tests we found that this kernel generally gives results superior to those obtained with commonly used Gaussian[42] or Laplacian[42] kernels for many oscillators.

The vector $\boldsymbol{\alpha}$ of regression coefficients $\{\alpha_m\}$ is obtained from the well-known expression:[48]

$$\boldsymbol{\alpha}=\left(\mathbf{K}+\lambda\mathbf{I}\right)^{-1}\mathbf{Y}^{\mathrm{ref}} \tag{S20}$$

where $\mathbf{I}$ is the $N_{\mathrm{tr}}\times N_{\mathrm{tr}}$ identity matrix, $\mathbf{K}$ is the kernel matrix with elements calculated with the Matérn kernel for all pairs of training set points, λ is the regularization parameter, and $\mathbf{Y}^{\mathrm{ref}}$ is the vector of the reference values calculated with A-SBH for the training set points.

The only parameters in Eqs. (S19) and (S20) that need to be defined by a user, are n, σ, and λ. We found that $n=1$ provides the best results, while σ and λ were determined by two iterations of a logarithmic grid search[44] that aims to minimize the error in the validation set for KRR models trained on the subtraining set. The validation and subtraining sets are obtained by splitting the training set so that the validation set consists of 20% of the training set points. The splitting was done randomly for the high-dimensional case. In the one-dimensional (1-D) case the splitting was performed uniformly so that validation set points are always in between the boundaries of the training set points defined by R_1.

In principle, it is possible to estimate energy gradients $\frac{\partial V_J}{\partial R_k}$ from ML models trained only on energies V_J. However, this is known to lead to rather poor results[39-40, 49] and our test calculations have confirmed that errors in gradients calculated by taking derivatives of ML models for energies are much worse than gradient components predicted by ML models individually trained for them. Another approach would be to train ML on both energies and gradients[50-52] or only gradients[40] to generate the universal ML model for predicting both energies and gradients. Although this approach is known to improve the accuracy of ML models,[40, 50-51] it very quickly leads to an explosion in the kernel matrix size, which then either does not fit into even large random-access memory or is too slow to invert, because matrix inversion scales as O(N^3). This limits the applicability of the latter approach. Hence, for practical applications, we prefer training ML models in this exploratory study for each gradient component individually as done in other studies.[53]

All KRR calculations and parameter searches were performed with MLatom.[54] During the nonadiabatic dynamics it was called by Newton-X to calculate on-the-fly energies, gradients, and nonadiabatic coupling components.

2. Nonadiabatic Dynamics with Machine Learning

2.1. Complete Machine Learning Model

Machine learning can, in principle, approximate any complex function with arbitrary accuracy if provided with sufficient information and hardware resources. We demonstrate for the 1-D model that it is possible to construct a *complete* ML model, which knows all the information about the photophysical system required to perform FSSH dynamics without loss of accuracy. During the dynamics with the complete model, calculations of additional reference data and re-training of the KRR models are not needed.

We trained ML models on 2^q ($q \geq 2$) points sampled uniformly from $R_1 = [-15.0 \ldots 15.0]$ a_0. The ML model trained on only 2^7 = 128 points could perfectly reproduce the numerically exact A-SBH trajectory with all hopping events (see Figure 1 in the main text). The maximum drift in total energy during 2 ps ML dynamics was much less than 0.005 eV or 0.1 kcal/mol, and could be reduced even further with more training points. Trajectories with ML models trained on fewer points had larger total energy drifts, fewer or no hops, and hops at other times than found in the A-SBH trajectory.

2.2. Approximate 1-D Machine Learning Potentials

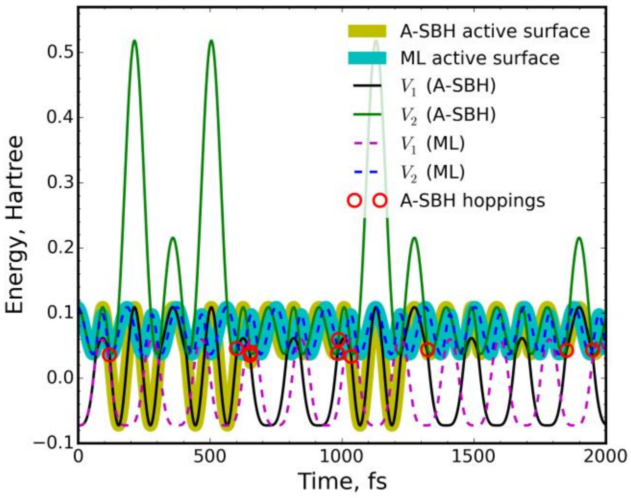

Figure S1. Comparison of A-SBH and ML surface hopping trajectories for 1-D model: The simulations started from the same initial conditions and were run with the same random seed. The ML model was trained on 16 points.

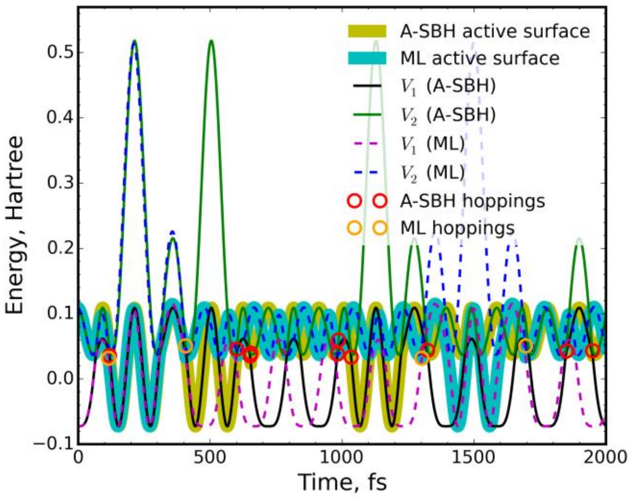

Figure S2. Comparison of A-SBH and ML surface hopping trajectories for 1-D model: The simulations started from the same initial conditions and were run with the same random seed. The ML model was trained on 16 points, and A-SBH nonadiabatic couplings were used throughout during the ML trajectory propagation.

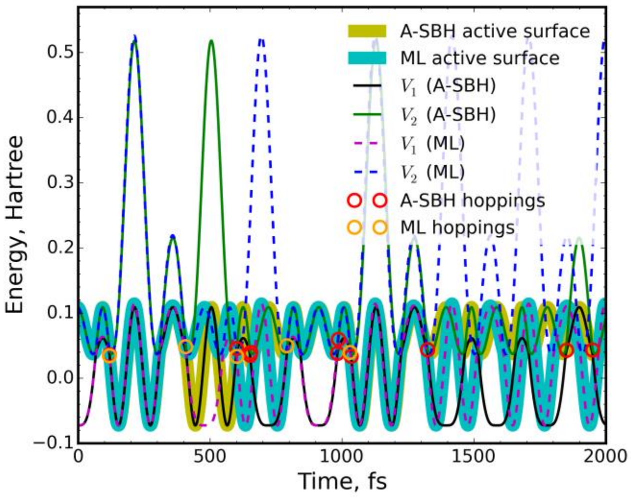

Figure S3. Comparison of A-SBH and ML surface hopping trajectories for 1-D model: The simulations started from the same initial conditions and were run with the same random seed. The ML model was trained on 16 points. During the ML trajectory propagation, A-SBH calculations were performed when the estimated energy gap between S_1 and S_0 was below 0.03 Hartree.

Sparse sampling creates a problem: the total energy is no longer perfectly constant (Figure S4). However, the conservation of total energy is not necessary as long as forces are predicted accurately with ML.[49] For example, in case of the 1-D ML model trained on 16 points, the maximum drift in total energy during 2 ps dynamics was as large as 0.30 eV or 7.0 kcal/mol. Nevertheless, the quality of this trajectory is reasonable (Figure S3). In practice, even during pure QM dynamics the total energy can deviate rather strongly due to various reasons. For example, when nonadiabatic dynamics are run at the semiempirical multi-reference configuration interaction level without adaptive time steps, there may be total energy deviations up to 10 kcal/mol between consecutive time steps and a total energy drift during trajectory propagation of more than 100 kcal/mol.[55] Thus, one may tolerate fluctuations in total energy during ML dynamics caused by incompleteness of the ML model as long as qualitative conclusions are preserved and crucial ensemble properties (*e.g.* excited-state lifetimes, evolution of geometrical properties, etc.) remain realistic. In other words, our approach can be regarded as an approximation to the reference high-level QM dynamics method, which may be more accurate than on-the-fly NA-MQC dynamics with a more approximate QM method. Low-level QM methods can still be useful as a prior to obtain more accurate ML models though,[56] which will be the topic of our future research.

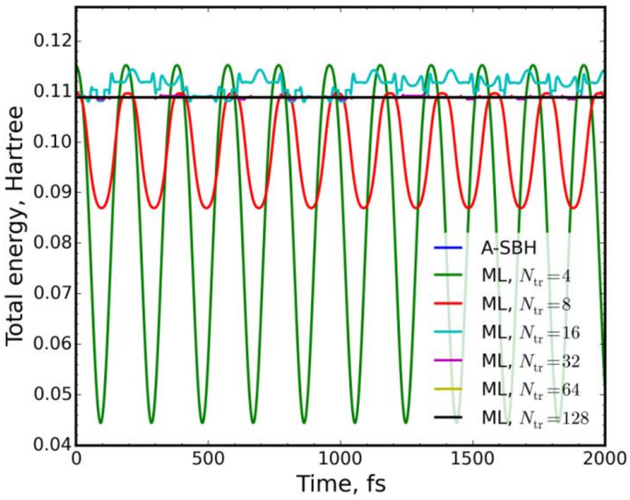

Figure S4. Total energy evolution along A-SBH and ML surface hopping trajectories for 1-D model: The ML models were trained on a different number of points (N_{tr}). The simulations started from the same initial conditions and were run with the same random seed. During the ML trajectory propagation, A-SBH calculations were performed when the estimated energy gap between S_1 and S_0 was below 0.03 Hartree.

2.3. Approximate 33-D Machine Learning Potentials

Simple random sampling will lead to many redundant structures in high-dimensional cases. We avoided this problem by first randomly sampling 50,000 points from the 33-dimensional space and then sampling 1,000 and 10,000 points from these 50,000 points using farthest-point traversal (FPT, with initial two points taken farthest apart), which is similar to structure-based sampling[41] and is known to yield more diverse training sets for more accurate ML models. Energies, gradients, and nonadiabatic couplings were calculated with A-SBH only for the selected 1,000 and 10,000 points. The initial 50,000 points were sampled within boundaries far exceeding those encountered during A-SBH dynamics to avoid extrapolation problems with ML. In our test calculations we observed that dynamics with ML trained on points sampled with FPT is of better quality than dynamics with ML trained on randomly sampled points.

The reduction of the number of required QM calculations can roughly serve as an estimate of the reduction of total computational cost, because the time for training ML models on small numbers of training set points and the time for estimating QM properties with ML (several seconds at most) are negligible compared to the time for typical QM calculations. In our case, A-SBH computations were required for the training points, while in our ML dynamics A-SBH was invoked for 13% (ML trained on 1,000 points) and 16% (ML trained on 10,000 points) of all time steps when the energy gap dropped below 0.03 Hartree, *i.e.* the number of A-SBH calculations was reduced in DC-FSSH dynamics by 84–87% by using ML. It is possible to reduce the fraction of QM calculations invoked during dynamics to less than 10% by decreasing the cutoff for the energy gap. However, we would not recommend going below 0.02 Hartree: this is the value, where hopping often happens in our pure A-SBH trajectories, and ML may well overestimate the gap and would then miss the hopping. Our ML model trained on 1,000 points indeed behaved poorly for a cutoff of 0.02 Hartree, while the model trained on 10,000 points gave hopping times that were similar for both values of the cutoff (0.02 and 0.03 Hartree).

In principle, no energy calculations are necessary for running accurate dynamics as long as accurate forces are available. This is often exploited in ML dynamics simulations, where only forces are predicted with ML.[49, 53] In our simulations we need ML energies only to estimate energy gap and thus to detect regions of higher probability of nonadiabatic transitions. Hence, in our model, total energy conservation is not a crucial issue. As we have shown above for individual trajectories of the 1-D model, relatively large deviations in total energy have little effect on the quality of the trajectories. On the other hand, we consider it useful to monitor the total energy drift to evaluate the stability of the ML approach, while we regard immediate fluctuations in the total energy as being less important.

We first analyze the total energy evolution averaged over 1,000 trajectories for the simulations with A-SBH and two ML models. Increasing the number of ML training points decreases the magnitude of the total energy fluctuations, and also brings the average total energy and the corresponding standard deviation closer to the A-SBH values (Figure S5). The final drift of the average total energy in 2 ps trajectories relatively to the initial time is reasonably small for ML trained on 1,000 points (0.35 eV, 8.03 kcal/mol) and much smaller for ML trained on 10,000 points (0.02 eV, 0.4 kcal/mol). This confirms the stability of the ML approach.

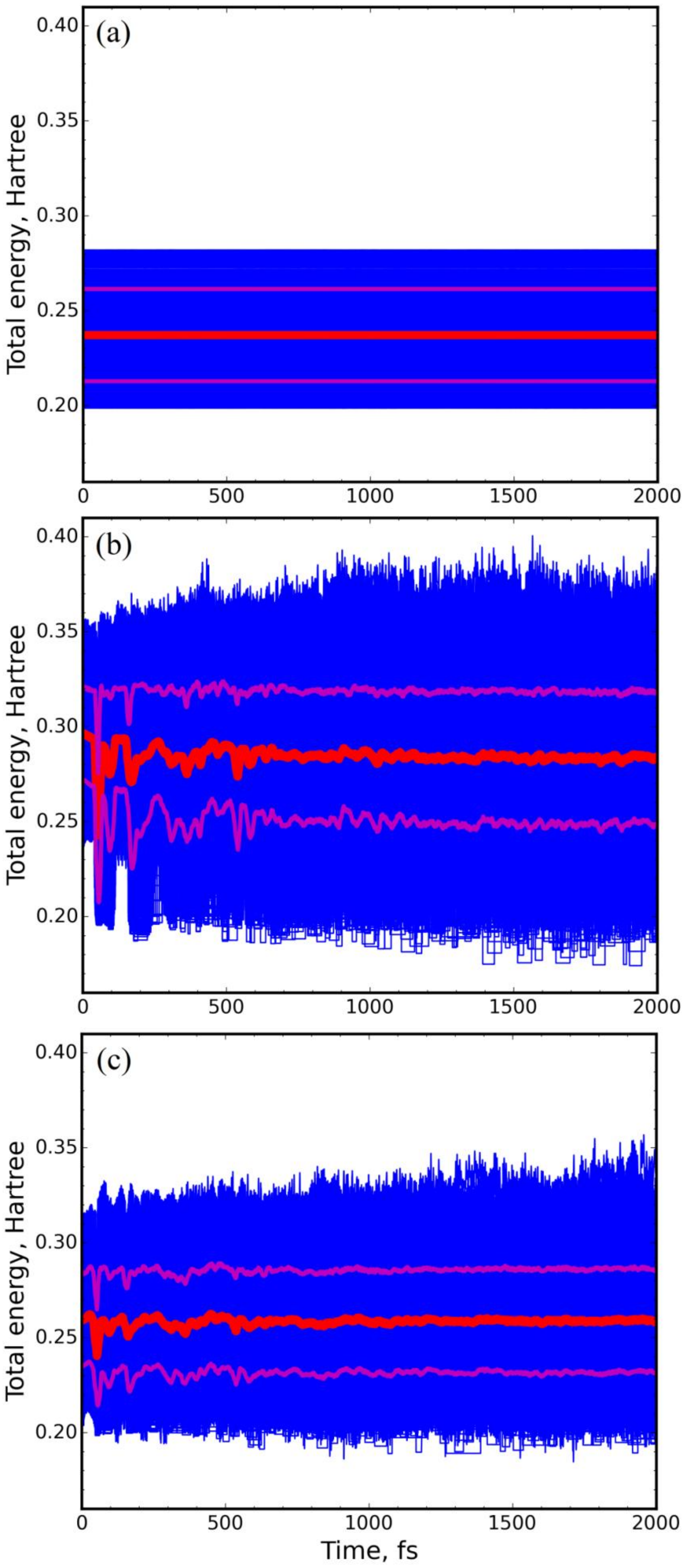

Figure S5. Evolution of the total energies along A-SBH (a) and ML (b,c) surface hopping trajectories for the 33-D model (blue). The ML models were trained on 1,000 (b) and 10,000 (c) points. Also shown are the total energy averaged over 1,000 trajectories (red) and the standard deviations (magenta). During the ML trajectory propagation, A-SBH calculations were performed when the estimated energy gap between S_1 and S_0 was below 0.03 Hartree.

As seen from a selection of randomly chosen individual trajectories (Figure S6) fluctuations of the total energy during ML dynamics are rather large, similar to the case of 1-D dynamics runs with ML trained on a very sparse grids. The total energy fluctuations are smaller for the ML model trained on 10,000 points. The largest total energy jumps between two consecutive steps happen when one step is calculated with ML and the other with A-SBH, *i.e.* at the borders of the region with an energy gap of 0.03 Hartree. For a randomly chosen trajectory, the largest total energy jumps are 2.17 eV (50.1 kcal/mol, ML trained on 1,000 points) and 0.92 eV (21.3 kcal/mol, ML trained on 10,000 points). However, the largest total energy jumps before the first A-SBH calculation is done are reasonably small: 0.13 eV (3.1 kcal/mol, ML trained on 1,000 points) and 0.11 eV (2.6 kcal/mol, ML trained on 10,000 points). As already mentioned, large variations in the ML total energy do not imply that the simulation is not reliable. Another important observation from Figure S6 is that there are clearly visible short periods when the total energy is constant: this happens when A-SBH calculations are requested for energy gaps below the cutoff of 0.03 Hartree. In these periods, the total energy is the true A-SBH energy, which is close to the A-SBH total energy at the start of the corresponding pure A-SBH trajectory. This suggests another way of estimating the quality of ML dynamics. Ideally, the A-SBH total energy should be the same during ML dynamics and pure A-SBH dynamics. To monitor this for our ML trajectories, we calculated the deviations between the A-SBH total energies at the first point of each low-energy-gap period in the ML trajectory and the starting point of the pure A-SBH trajectory with the same initial conditions. These deviations were collected from 1,000 trajectories run with both ML models. The corresponding error histogram is given in Figure S7. The mean deviations are reasonably close to zero, and the standard deviations are also rather small: −0.06±0.27 eV (ML trained on 1,000 points) and +0.06±0.14 eV (ML trained on 10,000 points, *i.e.* with roughly half the standard deviation of the model trained on 1,000 points). This provides further evidence for the stability of our ML approach.

Our smallest ML model contains 1,000 training points. Decreasing this number even further would quickly increase the ML error, which is proportional to $1/N_{\mathrm{tr}}^{b}$, where b is a constant depending on specific conditions and can assume values typically ranging from 0.5 to 2.[41, 57-58] In our simulations with 1,000 training points we have already seen signs of a break-down in terms of a larger drift in total energy after 2 ps and a higher sensitivity to the chosen energy cutoff. Thus, a training set of 1,000 points seems close to the the minimum size that is still applicable for the nonadiabatic dynamics simulations of our 33-D model. Generalizations to other systems are not yet possible at this stage.

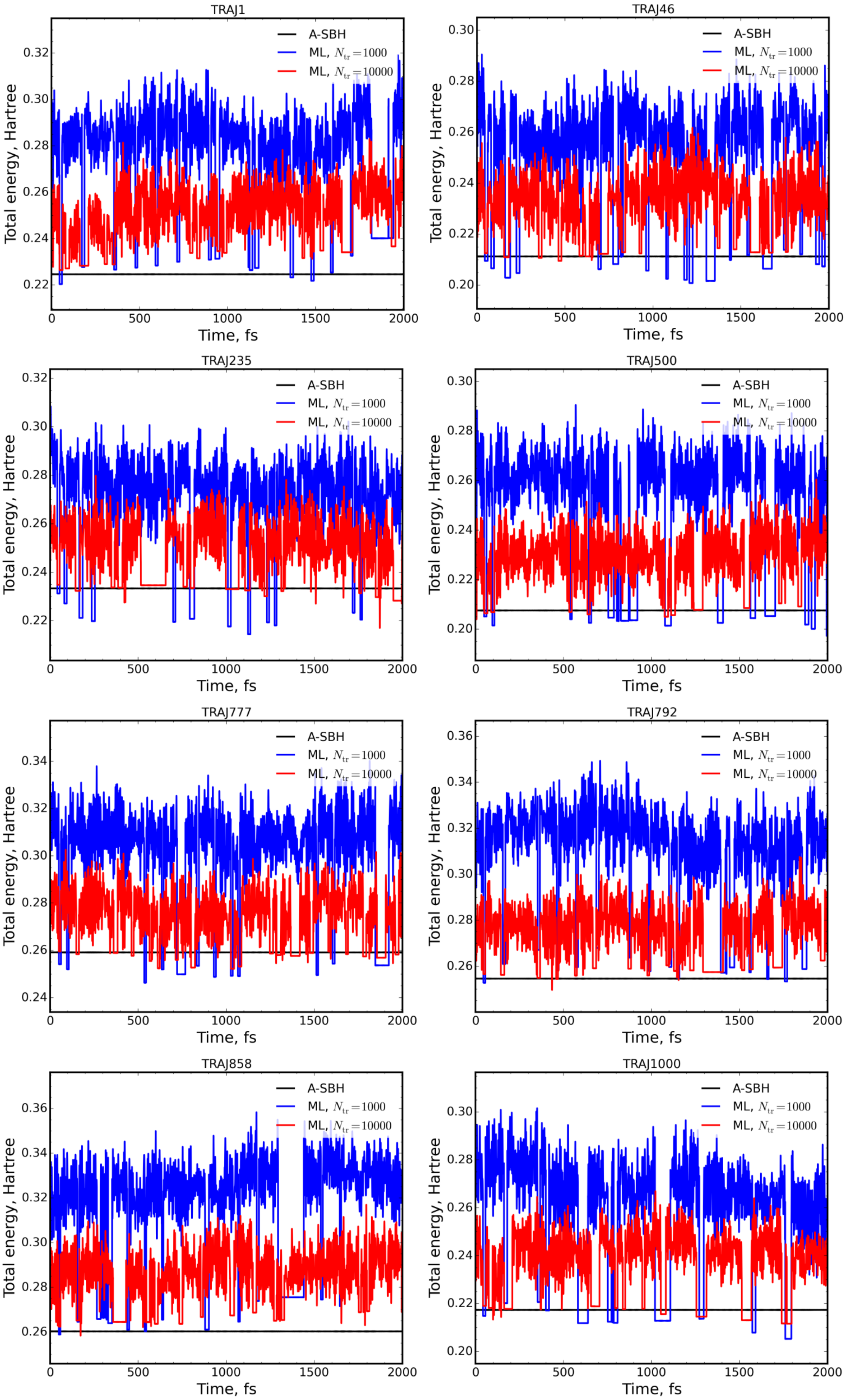

Figure S6. Total energy evolution along several randomly chosen A-SBH and ML surface hopping trajectories for 33-D model: The ML models were trained on 1,000 and 10,000 points. The simulations started from the same initial conditions for each trajectory. During the ML trajectory propagation, A-SBH calculations were performed when the estimated energy gap between S_1 and S_0 was below 0.03 Hartree. As discussed in the text, despite these large fluctuations, the trajectories are still considered reliable.

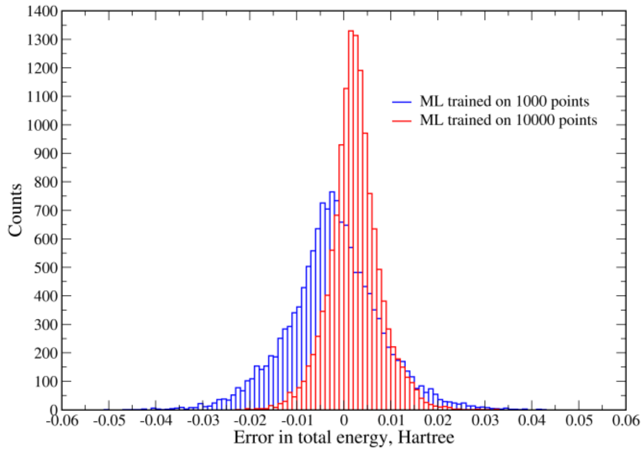

Figure S7. Histogram of the deviations between the A-SBH total energies during ML dynamics (calculated for the first points of periods with ML-estimated S_1–S_0 energy gaps below 0.03 Hartree) and at the starting point of pure A-SBH dynamics with the same initial conditions. The ML models were trained on 1,000 and 10,000 points.

2.4. Lifetime Fitting and Correlation Analysis

The S_1 population as a function of time was fitted with the function

$$f(t) = f_0 + (1 - f_0)\exp\left(-\frac{t - \tau_L}{\tau_D}\right) \quad \text{(S21)}$$

where f_0 is the population asymptotically remaining in the excited state, τ_L is the latency time before which there are no transfers to the ground state, and τ_D is the exponential decay time constant. For the fitting, τ_L was kept fixed. The lifetime of the excited state is $\tau_{ex} = \tau_L + \tau_D$.

	f_0	τ_L (fs)	τ_D (fs)	τ_{ex} (fs)
A-SBH	0.28	44	70	114 ± 1
ML (trained on 1,000 points)	0.28	37	63	100 ± 1
ML (trained on 10,000 points)	0.27	39	66	105 ± 1

The margin of error in the lifetime corresponds to a 95% confidence interval.

There is a high correlation between the results of the pure A-SBH and the ML simulations. This is shown in Figure S8 for the S_1 populations. The Pearson correlation coefficient between A-SBH and ML trained on 1,000 points is 0.988 over the full 2 ps data collection. It is 0.987 between A-SBH and ML trained on 10,000 points.

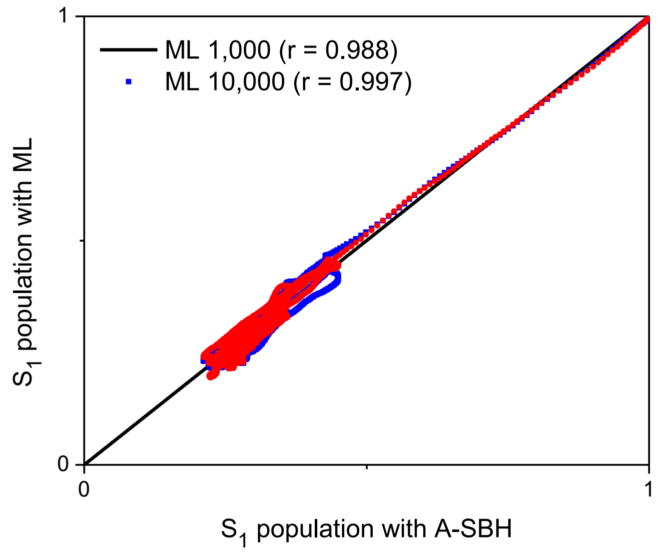

Figure S8. Correlation between the S_1 populations computed with pure A-SBH and with ML for all points up to 2 ps. *r* is the Pearson correlation coefficient.

The correlation tends to degrade as the dynamics evolves. If only points between 0 and 1 ps are considered, *r* = 0.994 between pure A-SBH and each ML set. By contrast, if only points between 1 and 2 ps are considered, the correlation coefficient is 0.477 between pure A-SBH and ML trained on 1,000 points, and 0.611 between pure A-SBH and ML trained on 10,000 points.

References

(1) Tully, J. C. Mixed Quantum-Classical Dynamics. *Faraday Discuss.* **1998,** *110*, 407–419.

(2) Agostini, F.; Abedi, A.; Suzuki, Y.; Gross, E. K. U. Mixed Quantum-Classical Dynamics on the Exact Time-Dependent Potential Energy Surface: A Fresh Look at Non-Adiabatic Processes. *Mol. Phys.* **2013,** *111*, 3625–3640.

(3) Granucci, G.; Persico, M. Critical Appraisal of the Fewest Switches Algorithm for Surface Hopping. *J. Chem. Phys.* **2007,** *126*, 134114.

(4) Tully, J. C. Molecular-Dynamics with Electronic-Transitions. *J. Chem. Phys.* **1990,** *93*, 1061–1071.

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

(6) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982,** *76*, 637–649.

(7) Butcher, J. A Modified Multistep Method for the Numerical Integration of Ordinary Differential Equations. *J. Assoc. Comput. Mach.* **1965,** *12*, 124–135.

(8) Subotnik, J. E.; Jain, A.; Landry, B.; Petit, A.; Ouyang, W.; Bellonzi, N. Understanding the Surface Hopping View of Electronic Transitions and Decoherence. *Annu. Rev. Phys. Chem.* **2016,** *67*, 387–417.

(9) Granucci, G.; Persico, M.; Zoccante, A. Including Quantum Decoherence in Surface Hopping. *J. Chem. Phys.* **2010,** *133*, 134111.

(10) Gao, X.; Thiel, W. Non-Hermitian Surface Hopping. *Phys. Rev. E* **2017,** *95*, 013308.

(11) Zhu, C. Y.; Nangia, S.; Jasper, A. W.; Truhlar, D. G. Coherent Switching with Decay of Mixing: An Improved Treatment of Electronic Coherence for Non-Born-Oppenheimer Trajectories. *J. Chem. Phys.* **2004,** *121*, 7658–7670.

(12) Subotnik, J. E.; Ouyang, W.; Landry, B. R. Can We Derive Tully's Surface-Hopping Algorithm from the Semiclassical Quantum Liouville Equation? Almost, but Only with Decoherence. *J. Chem. Phys.* **2013,** *139*, 214107.

(13) Barbatti, M.; Ruckenbauer, M.; Plasser, F.; Pittner, J.; Granucci, G.; Persico, M.; Lischka, H. Newton-X: A Surface-Hopping Program for Nonadiabatic Molecular Dynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2014,** *4*, 26–33.

(14) Barbatti, M.; Granucci, G.; Ruckenbauer, M.; Plasser, F.; Crespo-Otero, R.; Pittner, J.; Persico, M.; Lischka, H. *NEWTON-X: A Package for Newtonian Dynamics Close to the Crossing Seam (v. 2.0). Available via the Internet at www.newtonx.org*, 2017.

(15) Leggett, A. J.; Chakravarty, S.; Dorsey, A. T.; Fisher, M. P. A.; Garg, A.; Zwerger, W. Dynamics of the Dissipative Two-State System. *Rev. Mod. Phys.* **1987,** *59*, 1-85.

(16) Chen, H.-T.; Reichman, D. R. On the Accuracy of Surface Hopping Dynamics in Condensed Phase Non-Adiabatic Problems. *J. Chem. Phys.* **2016,** *144*, 094104.

(17) Kelly, A.; Markland, T. E. Efficient and Accurate Surface Hopping for Long Time Nonadiabatic Quantum Dynamics. *J. Chem. Phys.* **2013,** *139*, 014104.

(18) Landry, B. R.; Falk, M. J.; Subotnik, J. E. Communication: The Correct Interpretation of Surface Hopping Trajectories: How to Calculate Electronic Properties. *J. Chem. Phys.* **2013,** *139*, 211101.

(19) Landry, B. R.; Subotnik, J. E. How to Recover Marcus Theory with Fewest Switches Surface Hopping: Add Just a Touch of Decoherence. *J. Chem. Phys.* **2012,** *137*, 22A513.

(20) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989,** *157*, 200–206.

(21) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993,** *98*, 5648–5652.

(22) Becke, A. D. A New Mixing of Hartree-Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993,** *98*, 1372–1377.

(23) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab-Initio Calculation of Vibrational Absorption and Circular-Dichroism Spectra Using Density-Functional Force-Fields. *J. Phys. Chem.* **1994,** *98*, 11623–11627.

(24) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic-Behavior. *Phys. Rev. A* **1988,** *38*, 3098–3100.

(25) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988,** *37*, 785–789.

(26) Gordon, M. S. The Isomers of Silacyclopropane. *Chem. Phys. Lett.* **1980,** *76*, 163–168.

(27) Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements. *J. Am. Chem. Soc.* **1980,** *102*, 939–947.

(28) Gordon, M. S.; Binkley, J. S.; Pople, J. A.; Pietro, W. J.; Hehre, W. J. Self-Consistent Molecular Orbital Methods. 22. Small Split-Valence Basis Sets for Second-Row Elements. *J. Am. Chem. Soc.* **1982,** *104*, 2797–2803.

(29) Blaudeau, J.-P.; McGrath, M. P.; Curtiss, L. A.; Radom, L. Extension of Gaussian-2 (G2) Theory to Molecules Containing Third-Row Atoms K and Ca. *J. Chem. Phys.* **1997,** *107*, 5016–5021.

(30) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. 9. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971,** *54*, 724–728.

(31) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; DeFrees, D. J.; Pople, J. A.; Gordon, M. S. Self-Consistent Molecular Orbital Methods. 23. A Polarization-Type Basis Set for 2nd-Row Elements. *J. Chem. Phys.* **1982,** *77*, 3654–3665.

(32) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. 12. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular-Orbital Studies of Organic-Molecules. *J. Chem. Phys.* **1972,** *56*, 2257–2261.

(33) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G* Basis Set for Atoms K through Zn. *J. Chem. Phys.* **1998,** *109*, 1223–1229.

(34) Binning Jr., R. C.; Curtiss, L. A. Compact Contracted Basis-Sets for 3rd-Row Atoms - Ga-Kr. *J. Comput. Chem.* **1990,** *11*, 1206–1216.

(35) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* Basis Set for Third-Row Atoms. *J. Comput. Chem.* **2001,** *22*, 976–984.

(36) Hariharan, P. C.; Pople, J. A. Accuracy of Ah Equilibrium Geometries by Single Determinant Molecular-Orbital Theory. *Mol. Phys.* **1974,** *27*, 209–214.

(37) Hariharan, P. C.; Pople, J. A. Influence of Polarization Functions on Molecular-Orbital Hydrogenation Energies. *Theor. Chem. Acc.* **1973,** *28*, 213–222.

(38) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; J. A. Montgomery, J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford CT, 2013.

(39) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K.-R. Bypassing the Kohn-Sham Equations with Machine Learning. *Nat. Commun.* **2017,** *8*, 872.

(40) Chmiela, S.; Tkatchenko, A.; Sauceda, H. E.; Poltavsky, I.; Schütt, K. T.; Müller, K.-R. Machine Learning of Accurate Energy-Conserving Molecular Force Fields. *Sci. Adv.* **2017,** *3*, e1603015.

(41) Dral, P. O.; Owens, A.; Yurchenko, S. N.; Thiel, W. Structure-Based Sampling and Self-Correcting Machine Learning for Accurate Calculations of Potential Energy Surfaces and Vibrational Levels. *J. Chem. Phys.* **2017,** *146*, 244108.

(42) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013,** *9*, 3404–3419.

(43) 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.

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

(45) Ramakrishnan, R.; Hartmann, M.; Tapavicza, E.; von Lilienfeld, O. A. Electronic Spectra from TDDFT and Machine Learning in Chemical Space. *J. Chem. Phys.* **2015,** *143*, 084111.

(46) Unke, O. T.; Meuwly, M. Toolkit for the Construction of Reproducing Kernel-Based Representations of Data: Application to Multidimensional Potential Energy Surfaces. *J. Chem. Inf. Model.* **2017,** *57*, 1923–1931.

(47) Gneiting, T.; Kleiber, W.; Schlather, M. Matérn Cross-Covariance Functions for Multivariate Random Fields. *J. Am. Stat. Assoc.* **2010,** *105*, 1167–1177.

(48) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed.; Springer-Verlag: 2009.

(49) Li, Z.; Kermode, J. R.; De Vita, A. Molecular Dynamics with On-the-Fly Machine Learning of Quantum-Mechanical Forces. *Phys. Rev. Lett.* **2015,** *114*, 096405.

(50) Denzel, A.; Kästner, J. Gaussian Process Regression for Geometry Optimization. *J. Chem. Phys.* **2018,** *148*, 094114.

(51) Bartók, A. P.; Csányi, G. Gaussian Approximation Potentials: A Brief Tutorial Introduction. *Int. J. Quantum Chem.* **2015,** *115*, 1051–1057.

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

(53) Botu, V.; Ramprasad, R. Learning Scheme to Predict Atomic Forces and Accelerate Materials Simulations. *Phys. Rev. B* **2015,** *92*, 094306.

(54) Dral, P. O. *Mlatom: A Package for Atomistic Simulations with Machine Learning*, developmental versions; Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany (http://MLatom.com, accessed June 7th, 2018), 2013–2018.

(55) Spörkel, L.; Thiel, W. Adaptive Time Steps in Trajectory Surface Hopping Simulations. *J. Chem. Phys.* **2016,** *144*, 194108.

(56) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *J. Chem. Theory Comput.* **2015,** *11*, 2087–2096.

(57) Amari, S.-i.; Fujita, N.; Shinomoto, S. Four Types of Learning Curves. *Neural Comput.* **1992,** *4*, 605–618.

(58) Rupp, M.; Ramakrishnan, R.; von Lilienfeld, O. A. Machine Learning for Quantum Mechanical Properties of Atoms in Molecules. *J. Phys. Chem. Lett.* **2015,** *6*, 3309–3313.