

Representation of coupled adiabatic potential energy surfaces using neural network based quasi-diabatic Hamiltonians: 1,2 $^2A'$ states of LiFH

Cite this: DOI: 10.1039/c8cp06598e

 Yafu Guan,^a Dong H. Zhang,^b Hua Guo^c and David R. Yarkony^a

An analytic quasi-diabatic representation of *ab initio* electronic structure data is key to the accurate quantum mechanical description of non-adiabatic chemical processes. In this work, a general neural network (NN) fitting procedure is proposed to generate coupled quasi-diabatic Hamiltonians (H^d) that are capable of representing adiabatic energies, energy gradients, and derivative couplings over a wide range of geometries. The quasi-diabatic representation for LiFH is used as a testing example. The fitting data including adiabatic energies, energy gradients and interstate couplings are obtained from a previously fitted analytical quasi-diabatic potential energy matrix, and are well reproduced by the NN fitting. Most importantly, the NN fitting also yields quantum dynamic results that reproduce those on the original LiFH diabatic Hamiltonian, demonstrating the ability of NN to generate highly accurate quasi-diabatic Hamiltonians.

 Received 23rd October 2018,
Accepted 16th November 2018

DOI: 10.1039/c8cp06598e

rsc.li/pccp

1. Introduction

Accurate dynamic simulations of electronically non-adiabatic chemical processes have become a frontier in chemical physics.^{1,2} Such studies require accurate electronic structure data: energies, energy gradients and derivative couplings.^{3,4} Compared to calculating these data “on the fly” in direct dynamics,⁵ fitting methods benefit more from constructing analytic quasi-diabatic representations. A principal advantage of analytic quasi-diabatic representations is that one can employ much more accurate electronic structure methods than are currently practical in direct dynamics. Once constructed, the computational cost of evaluating analytic quasi-diabatic representations is negligible compared to actual *ab initio* calculations. However, it is not a trivial task to construct accurate quasi-diabatic representations from *ab initio* data evaluated exclusively in the adiabatic representation.

The attribute quasi used here, which we shall omit below except as needed for emphasis, indicates that for polyatomic

molecules rigorous diabatic representations do not exist,^{6–8} and the diabatic representation is thus not uniquely defined. As a consequence, a variety of methods to construct diabatic representations have been reported in the literature. These methods can be divided into three categories: property-based methods, diabatization by ansatz and derivative-based methods. The property-based methods use some electronic properties^{9–12} (e.g., dipole moment,¹³ quadrupole moment¹⁴ and transition dipole moment¹⁵) for a number of adiabatic states and construct diabatic states by imposing the condition that the chosen property or properties have to vary smoothly over the entire range of the coordinate space.¹⁶ Diabatization by ansatz is also a widely used method, in which (only) adiabatic energies are reproduced.^{16–18} Simple as these two kinds of methods may be, without using the derivative information, the quality of diabatization is not strictly under control and depends to a large extent on the physical intuition for a specific system.¹⁹ The derivative-based methods are in principle the most accurate, which employ the derivative information of electronic wave functions including energy gradients and derivative couplings. The derivative information is also very important to describe the geometric phase,²⁰ local topology of a conical intersection²¹ and avoided-crossing.²² Despite that much more computational effort is required to obtain all derivative information over an extended range of nuclear coordinates, it is indispensable to obtain accurate dynamic results. Abrol and Kuppermann²³ constructed a global diabatic potential

^a Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, USA. E-mail: yguan15@jhu.edu, yarkony@jhu.edu

^b State Key Laboratory of Molecular Reaction Dynamics and Center for Theoretical Computational Chemistry, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, People's Republic of China. E-mail: zhangdh@dicp.ac.cn

^c Department of Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA. E-mail: hguo@unm.edu

1 energy matrix (PEM) for H_3 based on *ab initio* derivative
couplings by solving the three-dimensional Poisson equation.
Collins and co-workers developed a generally applicable
approach based on modified Shepard interpolation.^{24–26}
5 Recently, Zhu and Yarkony have introduced a diabaticization
procedure, which expresses the diabatic PEM in symmetry
adapted polynomials and simultaneously fits and diabaticizes
ab initio electronic structure data, to produce a coupled quasi-
diabatic representation.^{27–30} A unique feature of this
10 approach is that it uses derivative coupling data to determine
diabatic states, so that the residual derivative coupling can be
determined. This procedure ultimately provides an accurate,
quantifiable diabatic representation of the adiabatic electro-
nic structure data, as demonstrated by, for example, excellent
15 agreement with experimentally measured dynamical attri-
butes in photodissociation of ammonia.^{31–33}

For years, artificial neural networks (ANNs) have been
proven to be a robust and powerful tool to fit accurate
adiabatic potential energy surfaces (PESs) for polyatomic
molecules in the gas phase and for the interaction of small
molecules with metal surfaces.^{34–46} The flexible functional
form of a NN can reproduce smoothly varying *ab initio* data
as accurately as possible. In addition, a NN also provides a
closed analytic form for gradients, and it has also been used
25 for simultaneous fitting of energies and energy gradients.^{47,48}
Very recently, attempts have been made to construct diabatic
representations with NNs to achieve higher fitting accuracy.
For example, Lenzen and Manthe used a NN-based diabatic
PEM to fit adiabatic energies alone within an ansatz based
30 diabaticization framework, in which derivative information is
not used.¹⁸ Guan *et al.* proposed a procedure which fits both
energies and interstate couplings to solve for the mixing angle
for the diabaticization of the lowest two states of LiFH.⁴⁹
However, it is not a general method, but is restricted to an
avoided crossing between only two states. More recently, Xie
35 *et al.* proposed a NN based method to represent elements of a
diabatic PEM and considered the permutation symmetry and
topological features near a conical intersection seam. This
method fits the already determined PEM elements of an
existing diabaticization.⁵⁰ Thus derivative information is **not**
40 included either.

In this work, we propose a more general fitting procedure
with NNs based on the Zhu–Yarkony diabaticization **method**,^{27–30}
which is aimed at reproducing adiabatic energies, energy
gradients and derivative couplings. Considering the fact that
45 the learning capacity of NNs is much greater than that of
polynomials, it can be anticipated that more accurate diabatic
representations can be constructed for larger and more com-
plicated molecules with NNs. In this work, we use a simple
example: the construction of the diabatic Hamiltonian for the
50 two lowest electronic states of LiFH, to demonstrate the accu-
racy of the new NN-based method. The rest of the paper is
organized as following: Section II describes the NN fitting
procedure in detail, the test results for LiFH are presented in
Section III, and the final section contains a brief discussion and
55 offers a future prospect.

II. The quasi-diabatic Hamiltonian and its determination with neural networks

A. The quasi-diabatic Hamiltonian

The quasi-diabatic Hamiltonian \mathbf{H}^d is an $N^{\text{state}} \times N^{\text{state}}$ sym-
metric matrix whose elements are continuous functions of
nuclear coordinates, which takes the following form

$$H_{\alpha,\beta}^d(\mathbf{Q}) \equiv \langle \Psi_\alpha^d(\mathbf{q};\mathbf{Q}) | \mathbf{H}^e(\mathbf{q};\mathbf{Q}) | \Psi_\beta^d(\mathbf{q};\mathbf{Q}) \rangle_{\mathbf{q}} \quad (1)$$

Here, \mathbf{H}^e is the electronic Hamiltonian, \mathbf{q} are the coordinates of
electrons, \mathbf{Q} are the nuclear coordinates that describe molecule
geometries and Ψ_α^d , $\alpha = 1 - N^{\text{state}}$, are the quasi-diabatic
electronic wave functions. This matrix is also referred to as
the diabatic PEM. The eigenvectors of \mathbf{H}^d satisfy the following
electronic Schrödinger equation:

$$[\mathbf{H}^d(\mathbf{Q}) - \mathbf{I}E^{a,I,(m)}(\mathbf{Q})]\mathbf{d}^I(\mathbf{Q}) = 0. \quad (2)$$

Here, \mathbf{I} is the identity matrix and $E^{a,I,(m)}$ is the **corresponding**
eigenenergy. The superscript (m) indicates that the results
come from the model Hamiltonian \mathbf{H}^d , rather than *ab initio*
(*ab*) calculations, and the superscript (a) indicates the adiabatic
representation.

B. Equations defining \mathbf{H}^d

$$E^{a,I,(m)} = \mathbf{d}^I(\mathbf{Q})^\dagger \mathbf{H}^d \mathbf{d}^I(\mathbf{Q}) \quad (3)$$

Eqn (3) is the first equation defining \mathbf{H}^d , which provides the
comparison of adiabatic energies. By differentiating eqn (2), the
defining equations for energy gradients and derivative cou-
plings are obtained:

$$\nabla_k E^{a,I,(m)}(\mathbf{Q}) = \mathbf{d}^I(\mathbf{Q})^\dagger \nabla_k \mathbf{H}^d \mathbf{d}^I(\mathbf{Q}) \quad (4)$$

$$h_k^{I,J,(m)} = [E^{a,J,(m)}(\mathbf{Q}) - E^{a,I,(m)}(\mathbf{Q})] f_k^{I,J,(m)}(\mathbf{Q}) = \mathbf{d}^I(\mathbf{Q})^\dagger (\nabla_k \mathbf{H}^d) \mathbf{d}^J(\mathbf{Q}) \quad (5)$$

where $f_k^{I,J,(m)}$ is derivative coupling and $h_k^{I,J,(m)}$ is defined
as interstate coupling, which is a much smoother function
than $f_k^{I,J,(m)}$ and shows no singularity when a degeneracy
occurs.⁴⁹

In this work, the matrix element of \mathbf{H}^d is expressed by a NN
function which uses \mathbf{Q} as the input. NNs are trained by
minimizing the difference between $E^{a,I,(m)}$, $\nabla_k E^{a,I,(m)}$ and $h_k^{I,J,(m)}$
and the corresponding *ab initio* data.

C. Feed-forward neural networks

In the present work, a feed-forward NN is employed, which is a
powerful and robust fitting tool that can in principle represent
any real and smooth function to an accuracy consistent with
the quality of the data.⁵¹ It also provides a closed analytical
form for both the output and gradient.

Feed-forward NNs consist of several layers. An M -layer NN
can be denoted $R-S^1-S^2-\dots-S^M$, which means that the network
has R elements in the input vector and S^m neurons in the m th
layer. Let $p_{r,q}$ denote the r th element of the q th input vector \mathbf{p}_q ,
where r ranges from 1 to R . Let n_k^m denote the k th element of the

1 network input vector \mathbf{n}^m and a_k^m denote the output of neuron k
 in m th layer. In the M th layer, \mathbf{a}^M is the network output. $w_{i,j}^m$
 represents the element at row i and column j of the weight
 5 matrix between layer $m - 1$ and layer m . The bias of neuron i in
 layer m is denoted by b_i^m . The transfer function in layer m is
 denoted by f^m . The input layer is denoted the zeroth layer. The
 output of one layer becomes the input of the following layer.
 The equations that describe such a NN are

$$10 \quad n_{i,q}^m = \sum_{j=1}^{s^{m-1}} \left(w_{i,j}^m a_{j,q}^{m-1} \right) + b_i^m \quad (6)$$

$$15 \quad a_{i,q}^m = f^m(n_{i,q}^m) \quad (7)$$

The analytic form for the gradient with respect to the input
 can be obtained through the following equations. Eqn (8) is the
 initial start in the zeroth layer, and using eqn (9) and (10)
 iteratively, the gradient of the output in all layers can be
 20 calculated.

$$\frac{\partial a_{i,q}^0}{\partial p_{r,q}} = \delta_{i,r} \quad (8)$$

$$25 \quad \frac{\partial n_{i,q}^m}{\partial p_{r,q}} = \sum_{j=1}^{s^{m-1}} w_{i,j}^m \frac{\partial a_{j,q}^{m-1}}{\partial p_{r,q}} \quad (9)$$

$$30 \quad \frac{\partial a_{i,q}^m}{\partial p_{r,q}} = \frac{\partial a_{i,q}^m}{\partial n_{i,q}^m} \frac{\partial n_{i,q}^m}{\partial p_{r,q}} = f^m(n_{i,q}^m) \frac{\partial n_{i,q}^m}{\partial p_{r,q}} \quad (10)$$

The set of NN parameters θ consists of all weights and
 biases. The training of the NN is the process of finding the
 optimal parameters that can reproduce the *ab initio* data as
 accurately as possible. The derivatives of the NN output and
 its gradients with respect to NN parameters θ also have
 35 closed analytical forms (see Appendices), which make it very
 convenient to apply gradient-based algorithms, such as
 quasi-Newton and conjugate gradient methods, to optimize
 the NN.

40 D. Determination of \mathbf{H}^d from fitting data

In this work, we perform the diabaticization of the two lowest
 states of LiFH as a testing example. A minimization scheme of
 the sum of squared errors or performance index is proposed as
 45 follows:

$$50 \quad P = \sum_{q=1}^Q \sum_{I=1}^{N^{\text{state}}} \left[E_q^{a,I,(m)} - E_q^{a,I,(ab)} \right]^2$$

$$+ \rho_g \sum_{q=1}^Q \sum_{I=1}^{N^{\text{state}}} \sum_{r=1}^R \left[\nabla_r E_q^{a,I,(m)} - \nabla_r E_q^{a,I,(ab)} \right]^2 \quad (11)$$

$$+ \rho_c \sum_{q=1}^Q \sum_{I=1}^{N^{\text{state}}-1} \sum_{J=I+1}^{N^{\text{state}}} \sum_{r=1}^R \left[h_{r,q}^{a,I,J,(m)} - h_{r,q}^{a,I,J,(ab)} \right]^2$$

55 which consists of three parts: sum of squared errors of

energies, gradients and interstate couplings, respectively. Here,
 Q is the number of geometries, R is the number of internal
 coordinates, and ρ_g and ρ_c are connecting weights for the errors
 of gradients and interstate couplings, respectively. We follow
 the choice of Pukrittayakamee and Nguyen-Truong^{47,48} for ρ_g
 and ρ_c :

$$\rho_g = \lambda_g \frac{\max \left\{ E_q^{a,I,(ab)} \right\}_{I=1,2}}{\max \left\{ \nabla_r E_q^{a,I,(ab)} \right\}_{I=1,2}} \quad (12)$$

$$\rho_c = \lambda_c \frac{\max \left\{ E_q^{a,I,(ab)} \right\}_{I=1,2}}{\max \left\{ h_{r,q}^{a,1,2,(ab)} \right\}^2} \quad (13)$$

where $N^{\text{state}} = 2$ in this case.

It is proved that the magnitude $\|\mathbf{f}^{a,I,J,(ab)} - \mathbf{f}^{a,I,J,(m)}\|$ is a
 direct measure of the quality of the quasi-diabatic
 representation.²⁸ If the *ab initio* data $E_q^{a,I,(ab)}$, $\nabla_k E_q^{a,I,(ab)}$ and $h_{k,l}^{a,I,(ab)}$
 are well reproduced, $\|\mathbf{f}^{a,I,J,(ab)} - \mathbf{f}^{a,I,J,(m)}\|$ would be small,
 and hence providing a quantifiable test of the quality of the
 corresponding representation. Therefore, the performance
 index P here is a criterion for the quality of diabaticization:
 the smaller P is, the better quality the corresponding diabatic
 representation has. By including the derivative information,
 energy gradients and interstate couplings, the number of
 points or nuclear configurations at which *ab initio* data are
 required to determine the NN coefficients can be reduced,
 compared to energy based determinations.²⁷

Since the performance index consists of only squared errors,
 the Levenberg–Marquardt algorithm is employed to minimize
 it, which is very robust and converges quickly.⁵² To achieve the
 best results, multiple runs with different initial parameters
 (weights and biases) are performed, from which the fittings
 with smallest performance indexes are selected as the optimal
 35 results.

E. Fitting data for LiFH

The purpose of this work is to demonstrate the ability of NNs as
 a general fitting method to construct the quasi-diabatic repre-
 sentation. Therefore, instead of performing actual *ab initio*
 calculations, the fitting data are generated from the previously
 fitted analytical diabatic PEM of Jasper and co-workers
 (LiFH)⁵³ according to eqn (3)–(5).

The inter-atomic distances R_{LiF} and R_{HF} along with the Li–F–
 H angle γ are used as coordinates to generate fitting data. The
 grid of points at which energies, energy gradients and interstate
 couplings are computed is defined by the following values of
 R_{LiF} , R_{HF} and γ :

$R_{\text{LiF}} = 2.0, 2.25, 2.5, 2.7, 2.96, 3.1, 3.3, 3.5, 3.8, 4.0, 4.5, 5.0,$
 $6.0, 7.0, 8.0, 9.0, 10.0, 11.0$ bohr,

$R_{\text{HF}} = 1.2, 1.4, 1.6, 1.73, 1.8, 2.0, 2.4, 2.8, 3.0, 3.5, 4.0, 4.5, 5.0,$
 $6.0, 7.0, 8.0, 9.0, 10.0, 11.0$ bohr,

$\gamma = 0^\circ, 15^\circ, 35^\circ, 55^\circ, 70^\circ, 90^\circ, 110^\circ, 145^\circ, 180^\circ$.

The geometries with R_{LiH} greater than 12.0 bohr or less than
 2.0 bohr are abandoned. Among all the coordinates listed

above, one can find the equilibrium lengths for LiF (2.96 bohr) and HF (1.73 bohr). This grid offers detailed information of the potentials and interstate couplings, and provides a good description of dynamically relevant regions and is sufficient to generate a satisfactory diabatic PEM.⁴⁹ The energies are read from LiFHJ in eV. For more complex systems, a trajectory-guided point sampling approach should be adopted to saturate data in dynamically relevant regions.²⁸

III. Results

A total of 2552 geometries giving rise to 28 072 least squares terms were assembled. The inverse of three inter-atomic distances R_{LiF} , R_{HF} and R_{LiH} was used as the input of the NN. The NN structure is 3-40-40-3, which has two hidden layers and 1923 parameters. This size of NN can already give satisfactory results. Even though more accurate results can be obtained by increasing the NN size, the fitting may be difficult to converge. The three components of the NN output correspond to H_{11}^d , H_{12}^d (H_{21}^d) and H_{22}^d , respectively. The transfer function in the first and second layers is a hyperbolic tangent function $f(x) = \tanh(x)$; in the third layer, it is a linear function $f(x) = x$. The λ_g and λ_c values were set to 10.0 and 0.1, respectively. The training is stopped after 500 iterations.

Ten fitting results with the smallest performance indices are listed in Table 1. Here, we define the root mean square error (RMSE) of energies, energy gradients and interstate coupling

as:

$$\text{RMSE}(E_{I=1,2}) = \sqrt{\frac{1}{Q} \sum_{q=1}^Q (E_q^{a,I,(m)} - E_q^{a,I,(ab)})^2},$$

$$\text{RMSE}(\nabla E_{I=1,2}) = \sqrt{\frac{1}{QR} \sum_{q=1}^Q \sum_{r=1}^R (\nabla_r E_q^{a,I,(m)} - \nabla_r E_q^{a,I,(ab)})^2},$$

$$\text{RMSE}(h^{1,2}) = \sqrt{\frac{1}{QR} \sum_{q=1}^Q \sum_{r=1}^R (h_{r,q}^{a,1,2,(m)} - h_{r,q}^{a,1,2,(ab)})^2}.$$

The first fitting with the smallest performance index is chosen as the optimal result. Fig. 1 shows the corresponding contour plots of the three diabatic PEM elements at $\gamma = 72.8^\circ$ (bond angle of the saddle point for the ground state adiabatic PES), which exhibit excellent smoothness.

The $\text{RMSE}(E_{I=1,2})$ of these fittings converged around 0.003 eV, which is quite satisfactory. The distributions of the fitting errors for adiabatic energies, which are defined as $E^{a,I,(m)} - E^{a,I,(ab)}$, are shown in Fig. 2. As shown by the figure, the fitting error is typically within the ± 0.01 eV boundaries for both states. Fig. 3 further illustrates the quality of the fitting for both adiabatic PESs. The energy contours of the adiabatic PESs for the Li + HF reaction are plotted as a function of $R(\text{LiF})$ and $R(\text{HF})$, with γ fixed at 72.8° (bond angle of the saddle point). It is

Table 1 NN fitting results for LiFH

No.	P (eV) ²	$\text{RMSE}(E_1)$ (eV)	$\text{RMSE}(E_2)$ (eV)	$\text{RMSE}(\nabla E_1)$ (eV bohr ⁻¹)	$\text{RMSE}(\nabla E_2)$ (eV bohr ⁻¹)	$\text{RMSE}(h^{1,2})$ (eV bohr ⁻¹)
1	0.7807	0.00239	0.00263	0.00421	0.00440	0.00213
2	0.8404	0.00320	0.00260	0.00442	0.00436	0.00232
3	0.8601	0.00281	0.00279	0.00428	0.00480	0.00213
4	0.8909	0.00280	0.00277	0.00444	0.00483	0.00216
5	0.9303	0.00304	0.00272	0.00478	0.00462	0.00230
6	1.0053	0.00345	0.00285	0.00464	0.00507	0.00241
7	1.0172	0.00356	0.00266	0.00441	0.00410	0.00356
8	1.0747	0.00268	0.00259	0.00454	0.00474	0.00335
9	1.2346	0.00408	0.00303	0.00558	0.00525	0.00257
10	1.2864	0.00279	0.00287	0.00484	0.00488	0.00400

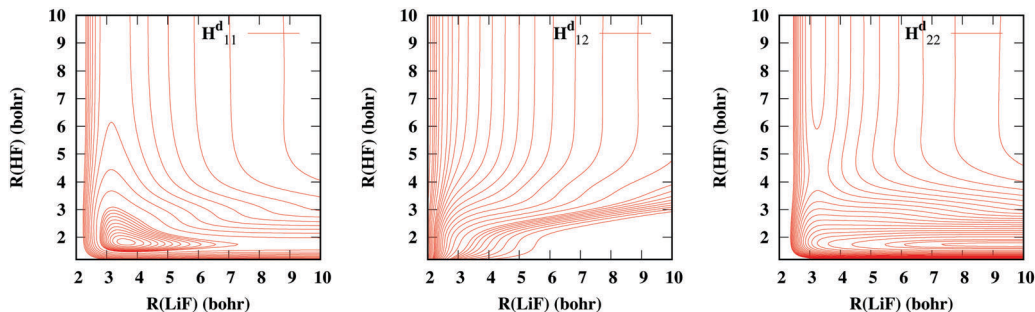


Fig. 1 Contour plots for H_{11}^d , H_{12}^d (H_{21}^d) and H_{22}^d as functions of $R(\text{LiF})$ and $R(\text{HF})$, with γ fixed at 72.8° .

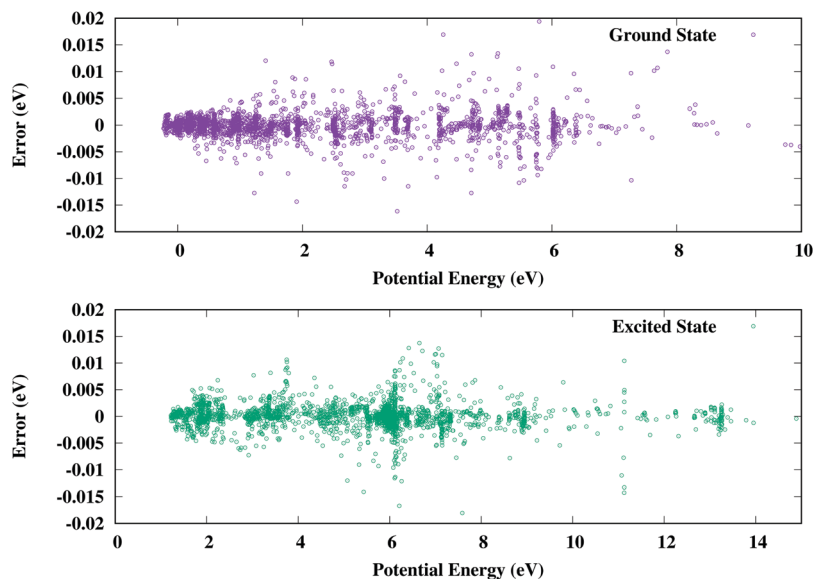


Fig. 2 Error distribution as a function of the adiabatic potential energy for ground and excited states. Error is defined as $E^{a,l,(m)} - E^{a,l,(ab)}$.

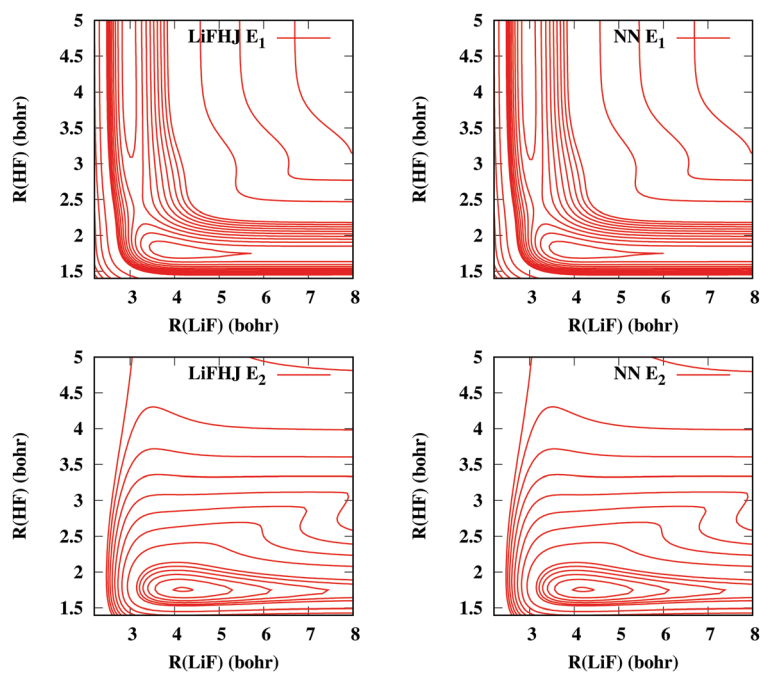


Fig. 3 Comparison of contour plots obtained from the NN PEM and original LiFHJ PEM, as a function of $R(\text{LiF})$ and $R(\text{HF})$, with γ fixed at 72.8° . The upper row shows the comparison of contour plots of the ground state PES, and the lower row compares the contour plots of the excited state PES.

clear that the NN fitting reproduces the original LiFHJ adiabatic PESs very well.

As for the derivative information, $\text{RMSE}(\nabla E_{I=1,2})$ converged around $0.005 \text{ eV bohr}^{-1}$, and $\text{RMSE}(h^{1,2})$ converged around $0.003 \text{ eV bohr}^{-1}$. The fitting errors are quite small compared to the largest absolute values for gradients and interstate couplings: 55.6 and $4.25 \text{ eV bohr}^{-1}$. The interstate couplings play a more important role in the non-adiabatic process. Therefore, we should carefully examine the fitting accuracy of interstate couplings.

As the reaction of $\text{Li} + \text{HF} \rightarrow \text{LiF} + \text{H}$ progresses, $R(\text{HF})$ increases, therefore the $R(\text{HF})$ component of $\mathbf{h}^{1,2}$ could be very important for the non-adiabatic process. In Fig. 4, we compare the $R(\text{HF})$ component of $\mathbf{h}^{1,2}$ obtained from the NN PEM and original LiFHJ PEM, as a function of $R(\text{LiF})$ and $R(\text{HF})$, with γ fixed at 70.0° . As can be seen, the reproduction of the interstate couplings is very accurate. In addition, the plots show that there is no unphysical oscillation in the NN fitting, which also guarantee the smoothness of the resulting

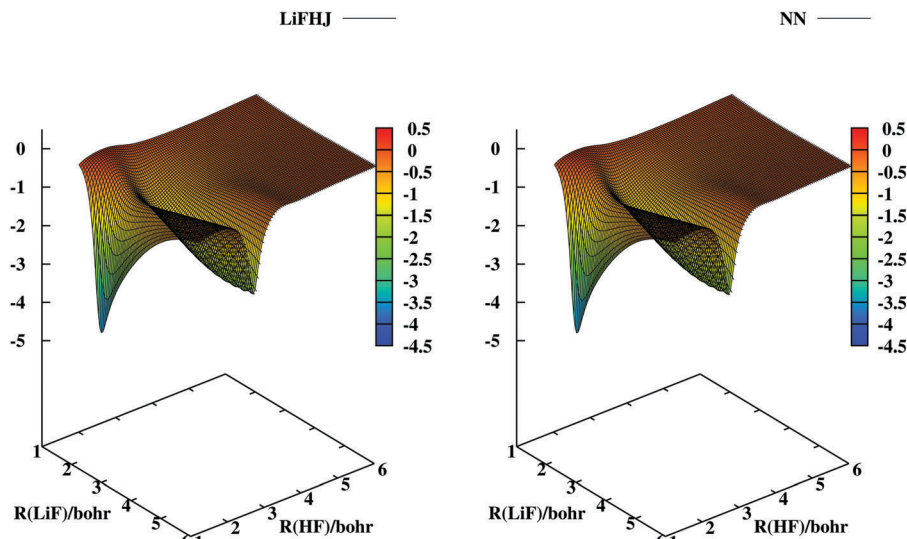


Fig. 4 Comparison of the $R(\text{HF})$ component of $h^{1,2}$ plots obtained from the NN PEM and original LiFHJ PEM, as a function of $R(\text{LiF})$ and $R(\text{HF})$, with γ fixed at 70.0° .

diabatic PEM elements. Fig. 5(a) is an example of avoided crossing between the energy curves and the corresponding interstate coupling is shown in Fig. 5(b). As can be seen, the avoided crossing occurs with $R(\text{HF})$ around 3.2 bohr. Correspondingly, the absolute value of the $R(\text{HF})$ component of interstate coupling reaches its maximum. It is clearly shown in Fig. 5(b) that the NN fitting accurately reproduced the interstate coupling.

To further illustrate the quality of NN fitting, quantum dynamic calculations were performed on both the LiFHJ diabatic PEM and NN diabatic PEM. The total reaction probability for the $\text{Li} + \text{HF} \rightarrow \text{LiF} + \text{H}$ reaction with the total angular

momentum $J = 0$ was calculated using the initial state-selected time-dependent wave packet (TDWP) approach.^{54,55} The numerical parameters for quantum reactive scattering wave packet calculations are shown in Table 2, the definition of which can be found in ref. 54. The initial wave packet was placed on the excited electronic state with HF in its ground rovibrational state to simulate the electronically nonadiabatic dissociation.⁵³ As shown in Fig. 6, despite the slight difference in the low translational energy region, the original LiFHJ PES and NN PES yield almost identical reaction probability over a wide energy range, thus validating the global accuracy of the NN fitting.

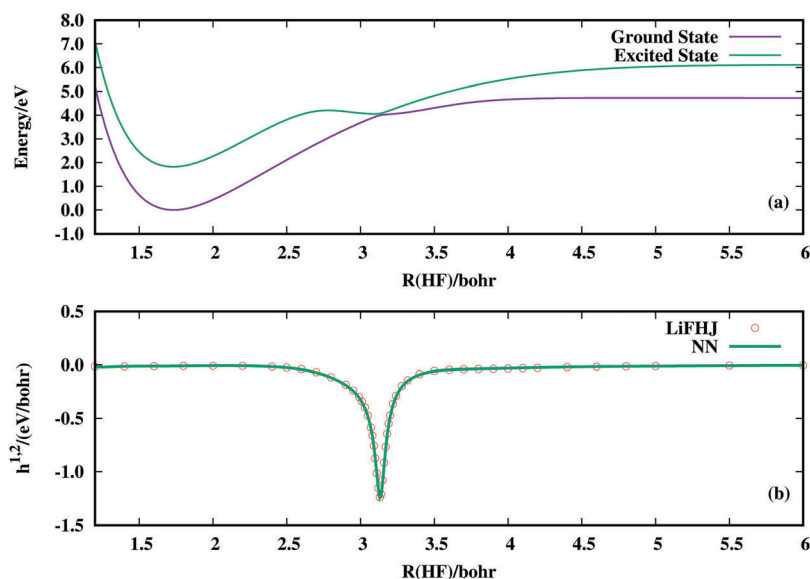


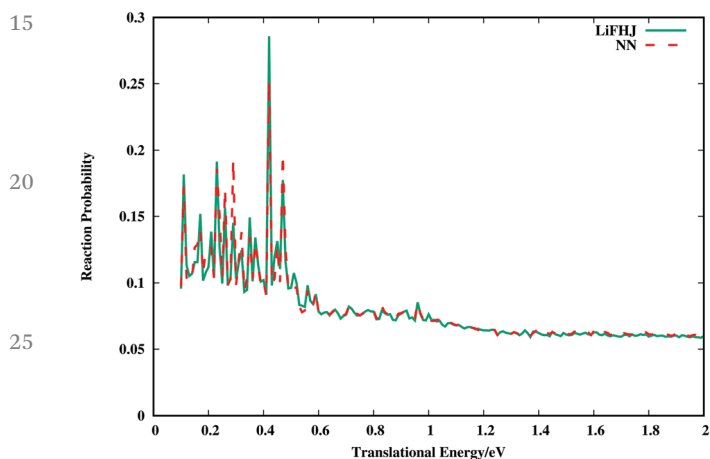
Fig. 5 Avoided crossing on the HF energy curve (a) and the corresponding interstate coupling gradients (b) with $\gamma = 70^\circ$ and $R(\text{LiF}) = 8.0$ bohr. In panel (a), the energy curves for the ground and excited states are shown in green and purple lines, respectively. The minimum energy difference is 0.077 eV. In panel (b), the interstate couplings read from LiFHJ are shown as circles; the green solid line is the corresponding fitted NN curve.

1 **Table 2** Numerical parameters used in the quantum reactive scattering
 wave packet calculations (atomic units are used, unless otherwise stated)

Grid range and size	$R \in [1.0, 14.0]$	$N_R^{\text{rot}} = 255$	$N_R^{\text{int}} = 80$
	$r \in [1.2, 6.5]$	$N_r^{\text{int}} = 40$	$N_r^{\text{asy}} = 10$
5 Rotational basis size	$j_{\text{max}} = 30$		
Initial wave packet	$E_0 = 1.0$ eV	$R_0 = 8.5$	$\delta = 0.08$
Absorbing potential ^a	$C_R = 0.02$	$n_R = 2.0$	$R_a = 12.0$ $R_b = 14.0$
	$C_r = 0.02$	$n_r = 2.0$	$r_a = 5.0$ $r_b = 6.5$
Dividing plane	$r = 4.0$		
Total propagation time	100 000	$\Delta t = 10.0$	

10 ^a Functional form for absorbing potential:

$$V_{\text{abs}}(x) = -iC_x \left(\frac{x - x_a}{x_b - x_a} \right)^{n_x}, \quad x \in [x_a, x_b].$$



15
20
25
30 **Fig. 6** Comparison of total reaction probabilities for the Li + HF \rightarrow LiF + H reaction with total angular momentum $J = 0$ on LiFHJ (green solid line) and the NN PES (red dashed line). The initial wave packet was placed on the excited state.

35 IV. Summary

Analytic quasi-diabatic representations that can reproduce electronic structure data, energies, energy gradients and derivative couplings are crucial to perform accurate fully quantum mechanical simulations of electronically nonadiabatic processes. For years, NNs have been used as an effective and robust tool for the fitting of the adiabatic PESs of molecular systems. Its closed analytical form for both the output and gradient makes it a promising derivative-based method to construct accurate quasi-diabatic representations. In this work, we proposed a general fitting procedure to construct quasi-diabatic representations with NNs. The previously constructed diabatic PEM of the lowest two electronic states LiFH was used as a simple testing example. The NN fitting is demonstrated to accurately reproduce energies, energy gradients and inter-state couplings, and most importantly, it also allows the reproduction of quantum dynamic results generated from the original PEM, which indicates its global accuracy. Even though the testing LiFH case is a simple two-state model, the generalization of the current method to multiple states is straightforward.

In future work, two important issues need to be addressed. One is the description of the vicinity of a conical intersection. It is important to note that the two lowest electronic states of LiFH do not cross to form a conical intersection seam, but form a seam of avoided-crossing instead, which is much easier to cope with. However, when it comes to describe the vicinity of a conical intersection, the orthogonal intersection adapted coordinates based on Yarkony's **g** and **h** vectors must be exploited to better reproduce the local topography.^{56–58} The other important issue to be addressed is the symmetry adaption. To construct the global diabatic representations, one has to consider the complete nuclear permutation inversion (CNPI) symmetry,²⁷ the symmetry induced by interchange of identical nuclei and inversion of the entire molecule, nuclei, and electrons. When the diabatic electronic states carry one-dimensional irreducible representations of the CNPI group, the diagonal elements of the diabatic PEM are invariant with respect to the permutations of identical nuclei (*i.e.*, carry totally symmetric irreducible representation), and the permutation invariant polynomial neural networks (PIP-NNs) can be used to preserve such symmetry.^{38,50} On the other hand, the off-diagonal element of the diabatic matrix may change sign under certain permutations if the two relevant electronic states carry different irreducible representations of the CNPI group. In this case, the off-diagonal element can be expressed as PIP-NN multiplied by a factor that preserves the corresponding non-totally symmetric irreducible representation.⁵⁰ The multiplicative factor can be obtained by applying CNPI group projection operators to basis functions of coordinates.²⁷

To summarize, an important first step in using NNs to construct a quasi-diabatic representation from *ab initio* data: energies, energy gradients and derivative couplings, is reported. The results are quite encouraging. We expect that, through this work, quasi-diabatic representations for larger and more complicated molecules can be constructed by NNs, which would facilitate accurate (quantum) dynamic simulations of nonadiabatic processes and help us to gain better understanding of nonadiabatic chemical processes.

Conflicts of interest

There are no conflicts of interest to declare.

Appendix A: derivatives of NN output with respect to NN parameters

From eqn (6) and (7), by taking the derivatives of $a_{k,q}^M$ with respect to $w_{i,j}^m$ and b_i^m , we have

$$\frac{\partial a_{k,q}^M}{\partial w_{i,j}^m} = \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \frac{\partial n_{i,q}^m}{\partial w_{i,j}^m} = \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} a_{j,q}^{m-1}, \quad (\text{A1})$$

$$\frac{\partial a_{k,q}^M}{\partial b_i^m} = \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \frac{\partial n_{i,q}^m}{\partial b_i^m} = \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \quad (\text{A2})$$

It can be observed that to calculate the derivatives, $\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m}$ must be obtained. In the M th layer (*i.e.*, the output layer), we have

$$\frac{\partial a_{k,q}^M}{\partial n_{i,q}^M} = f^M(n_{i,q}^M) \delta_{i,k} \quad (\text{A3})$$

Then, we apply the chain rule to obtain the recurrence relation between $\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m}$ and $\frac{\partial a_{k,q}^M}{\partial n_{i,q}^{m-1}}$.

$$\begin{aligned} \frac{\partial a_{k,q}^M}{\partial n_{i,q}^{m-1}} &= \sum_{i=1}^{s^m} \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \frac{\partial n_{i,q}^m}{\partial n_{i,q}^{m-1}} \\ &= \sum_{i=1}^{s^m} \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \frac{\partial n_{i,q}^m}{\partial a_{j,q}^{m-1}} \frac{\partial a_{j,q}^{m-1}}{\partial n_{i,q}^{m-1}} \\ &= \sum_{i=1}^{s^m} \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} f^{m-1}(n_{j,q}^{m-1}) w_{ij}^m \end{aligned} \quad (\text{A4})$$

Starting with eqn (A3) and using eqn (A4), all the $\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m}$ can be calculated.

Appendix B: derivatives of the gradient with respect to NN parameters

Similarly, by taking the derivatives of $\frac{\partial a_{k,q}^M}{\partial p_{r,q}}$ with respect to w_{ij}^m and b_i^m , we have

$$\begin{aligned} \frac{\partial}{\partial w_{ij}^m} \left(\frac{\partial a_{k,q}^M}{\partial p_{r,q}} \right) &= \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial w_{ij}^m} \right) \\ &= \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \times \frac{\partial n_{i,q}^m}{\partial w_{ij}^m} \right) \end{aligned} \quad (\text{B1})$$

$$= \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \right) \times a_{i,q}^{m-1} + \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \times \frac{\partial a_{j,q}^{m-1}}{\partial p_{r,q}}$$

$$\begin{aligned} \frac{\partial}{\partial b_i^m} \left(\frac{\partial a_{k,q}^M}{\partial p_{r,q}} \right) &= \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial b_i^m} \right) \\ &= \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \times \frac{\partial n_{i,q}^m}{\partial b_i^m} \right) \\ &= \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \right) \end{aligned} \quad (\text{B2})$$

$\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m}$ have already been calculated using eqn (A3) and (A4), and $\frac{\partial a_{j,q}^{m-1}}{\partial p_{r,q}}$ can be calculated using eqn (8)–(10). It is the

$\frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \right)$ that are left to be calculated.

In the M th layer, it is easy to have the following

$$\frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^M} \right) = \frac{\partial}{\partial p_{r,q}} \left[f^M(n_{i,q}^M) \delta_{i,k} \right] = \delta_{i,k} f^{M'}(n_{i,q}^M) \frac{\partial n_{i,q}^M}{\partial p_{r,q}}, \quad (\text{B3})$$

and the recurrence relation between $\frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \right)$ and

$\frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^{m-1}} \right)$ can be obtained by taking derivative of eqn (A4) with respect with p_r :

$$\begin{aligned} \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \right) &= \sum_{i=1}^{s^m} \frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \right) f^{m-1}(n_{j,q}^{m-1}) w_{ij}^m \\ &\quad + \sum_{i=1}^{s^m} \frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} f^{m-1}(n_{j,q}^{m-1}) \frac{\partial n_{j,q}^{m-1}}{\partial p_{r,q}} w_{ij}^m. \end{aligned} \quad (\text{B4})$$

Therefore, all $\frac{\partial}{\partial p_{r,q}} \left(\frac{\partial a_{k,q}^M}{\partial n_{i,q}^m} \right)$ are obtained, and hence the derivatives of the gradient with respect to the NN parameters.

Appendix C: evaluation of $\frac{\partial \mathbf{d}^I(\mathbf{Q})}{\partial \theta_k}$

Since the expressions for energies, energy gradients and inter-state couplings all contain eigenvector $\mathbf{d}^I(\mathbf{Q})$, in order to perform optimization, its derivatives with respect to NN parameters $\frac{\partial \mathbf{d}^I(\mathbf{Q})}{\partial \theta_k}$ are required. This is obtained from the derivative of eqn (2) as follows:

$$\begin{aligned} \mathbf{d}^I(\mathbf{Q})^\dagger \left[\frac{\partial}{\partial \theta_k} \mathbf{H}^d(\mathbf{Q}) \right] \mathbf{d}^I(\mathbf{Q}) \\ = \left[E^{a,J,(m)}(\mathbf{Q}) - E^{a,I,(m)}(\mathbf{Q}) \right] \mathbf{d}^I(\mathbf{Q})^\dagger \frac{\partial}{\partial \theta_k} \mathbf{d}^I(\mathbf{Q}), \end{aligned} \quad (\text{C1})$$

so

$$\frac{\partial \mathbf{d}^I(\mathbf{Q})}{\partial \theta_k} = \sum_{K \neq I}^{N^{\text{state}}} \mathbf{d}^K(\mathbf{Q}) \frac{\mathbf{d}^K(\mathbf{Q})^\dagger \left[\frac{\partial}{\partial \theta_k} \mathbf{H}^d(\mathbf{Q}) \right] \mathbf{d}^I(\mathbf{Q})}{E^{a,I,(m)}(\mathbf{Q}) - E^{a,K,(m)}(\mathbf{Q})}, \quad (\text{C2})$$

or

$$\frac{\partial \mathbf{d}^I(\mathbf{Q})}{\partial \theta_k} \mathbf{d}^K(\mathbf{Q}) = \frac{\mathbf{d}^K(\mathbf{Q})^\dagger \left[\frac{\partial}{\partial \theta_k} \mathbf{H}^d(\mathbf{Q}) \right] \mathbf{d}^I(\mathbf{Q})}{E^{a,I,(m)}(\mathbf{Q}) - E^{a,K,(m)}(\mathbf{Q})}. \quad (\text{C3})$$

This is valid provided it is not near geometry with degeneracy where $E^{a,J,(m)}(\mathbf{Q}) = E^{a,I,(m)}(\mathbf{Q})$. Since LiFH has no conical intersection, it is safe to use these equations to calculate $\frac{\partial \mathbf{d}^I(\mathbf{Q})}{\partial \theta_k}$.

Acknowledgements

Y. G., H. G. and D. R. Y. acknowledge support of the Department of Energy grant DE-SC0015997.

References

- 1 D. R. Yarkony, *Chem. Rev.*, 2012, **112**, 481.
- 2 D. H. Zhang and H. Guo, *Annu. Rev. Phys. Chem.*, 2016, **67**, 135.
- 3 F. Gatti, *Molecular Quantum Dynamics: from Theory to Applications*, Springer, Berlin, 2014.
- 4 H. Guo and D. R. Yarkony, *Phys. Chem. Chem. Phys.*, 2016, **18**, 26335.
- 5 B. F. E. Curchod and T. J. Martínez, *Chem. Rev.*, 2018, **118**, 3305.
- 6 M. Baer, *Chem. Phys.*, 1976, **15**, 49.
- 7 C. A. Mead and D. G. Truhlar, *J. Chem. Phys.*, 1982, **77**, 6090.
- 8 M. Baer, *Phys. Rep.*, 2002, **358**, 75.
- 9 G. J. Atchity and K. Ruedenberg, *Theor. Chem. Acc.*, 1997, **97**, 47.
- 10 H. Nakamura and D. G. Truhlar, *J. Chem. Phys.*, 2001, **115**, 10353.
- 11 J. E. Subotnik, S. Yeganeh, R. J. Cave and M. A. Ratner, *J. Chem. Phys.*, 2008, **129**, 244101.
- 12 C. E. Hoyer, K. Parker, L. Gagliardi and D. G. Truhlar, *J. Chem. Phys.*, 2016, **144**, 194101.
- 13 H. Werner and W. Meyer, *J. Chem. Phys.*, 1981, **74**, 5802.
- 14 C. Petrongolo, G. Hirsch and R. J. Buenker, *Mol. Phys.*, 1990, **70**, 825.
- 15 M. Perić, R. J. Buenker and S. D. Peyerimhoff, *Mol. Phys.*, 1990, **71**, 673.
- 16 A. Viel and W. Eisfeld, *J. Chem. Phys.*, 2004, **120**, 4603.
- 17 J. Jornet-Somoza, B. Lasorne, M. A. Robb, H.-D. Meyer, D. Lauvergnat and F. Gatti, *J. Chem. Phys.*, 2012, **137**, 084304.
- 18 T. Lenzen and U. Manthe, *J. Chem. Phys.*, 2017, **147**, 084105.
- 19 H. Köppel, *Diabatic representation: Methods for the construction of diabatic electronic states, Conical Intersections*, World Scientific, Singapore, 2011, pp. 175–204.
- 20 D. R. Yarkony, *J. Chem. Phys.*, 1996, **105**, 10456.
- 21 D. R. Yarkony, *Acc. Chem. Res.*, 1998, **31**, 511.
- 22 D. R. Yarkony, *J. Chem. Phys.*, 1990, **92**, 2457.
- 23 R. Abrol and A. Kuppermann, *J. Chem. Phys.*, 2002, **116**, 1035.
- 24 C. R. Evenhuis and M. A. Collins, *J. Chem. Phys.*, 2004, **121**, 2515.
- 25 O. Godsi, C. R. Evenhuis and M. A. Collins, *J. Chem. Phys.*, 2006, **125**, 104105.
- 26 C. Evenhuis and T. J. Martínez, *J. Chem. Phys.*, 2011, **135**, 224110.
- 27 X. Zhu and D. R. Yarkony, *J. Chem. Phys.*, 2010, **132**, 104101.
- 28 X. Zhu and D. R. Yarkony, *J. Chem. Phys.*, 2012, **136**, 174110.
- 29 X. Zhu and D. R. Yarkony, *J. Chem. Phys.*, 2012, **137**, 22A511.
- 30 X. Zhu and D. R. Yarkony, *J. Chem. Phys.*, 2014, **140**, 024112.
- 31 J. Ma, X. Zhu, H. Guo and D. R. Yarkony, *J. Chem. Phys.*, 2012, **137**, 22A541.
- 32 C. Xie, J. Ma, X. Zhu, D. H. Zhang, D. R. Yarkony, D. Xie and H. Guo, *J. Phys. Chem. Lett.*, 2014, **5**, 1055.
- 33 C. Xie, X. Zhu, J. Ma, D. R. Yarkony, D. Xie and H. Guo, *J. Chem. Phys.*, 2015, **142**, 091101.
- 34 T. B. Blank, S. D. Brown, A. W. Calhoun and D. J. Doren, *J. Chem. Phys.*, 1995, **103**, 4129.
- 35 D. F. R. Brown, M. N. Gibbs and D. C. Clary, *J. Chem. Phys.*, 1996, **105**, 7597.
- 36 C. M. Handley and P. L. A. Popelier, *J. Phys. Chem. A*, 2010, **114**, 3371.
- 37 J. Behler, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17930.
- 38 B. Jiang and H. Guo, *J. Chem. Phys.*, 2013, **139**, 054112.
- 39 J. Li, B. Jiang and H. Guo, *J. Chem. Phys.*, 2013, **139**, 204103.
- 40 B. Jiang and H. Guo, *J. Chem. Phys.*, 2014, **141**, 034109.
- 41 S. Manzhos, R. Dawes and T. Carrington, *Int. J. Quantum Chem.*, 2015, **115**, 1012.
- 42 X. Shen, J. Chen, Z. Zhang, K. Shao and D. H. Zhang, *J. Chem. Phys.*, 2015, **143**, 144701.
- 43 T. Liu, Z. Zhang, B. Fu, X. Yang and D. H. Zhang, *Chem. Sci.*, 2016, **7**, 1840.
- 44 K. Shao, J. Chen, Z. Zhao and D. H. Zhang, *J. Chem. Phys.*, 2016, **145**, 071101.
- 45 B. Kolb, X. Luo, X. Zhou, B. Jiang and H. Guo, *J. Phys. Chem. Lett.*, 2017, **8**, 666.
- 46 K. Shakouri, J. Behler, J. Meyer and G.-J. Kroes, *J. Phys. Chem. Lett.*, 2017, **8**, 2131.
- 47 A. Pukrittayakamee, M. Malshe, M. Hagan, L. M. Raff, R. Narulkar, S. Bukkapatnum and R. Komanduri, *J. Chem. Phys.*, 2009, **130**, 134101.
- 48 H. T. Nguyen-Truong and H. M. Le, *Chem. Phys. Lett.*, 2015, **629**, 40.
- 49 Y. Guan, B. Fu and D. H. Zhang, *J. Chem. Phys.*, 2017, **147**, 224307.
- 50 C. Xie, X. Zhu, D. R. Yarkony and H. Guo, *J. Chem. Phys.*, 2018, **149**, 144107.
- 51 K. Hornik, M. Stinchcombe and H. White, *Neural Netw.*, 1989, **2**, 359.
- 52 M. T. Hagan and M. B. Menhaj, *IEEE Trans. Neural Netw.*, 1994, **5**, 989.
- 53 A. W. Jasper, M. D. Hack, D. G. Truhlar and P. Piecuch, *J. Chem. Phys.*, 2002, **116**, 8353.
- 54 D. H. Zhang and J. Z. H. Zhang, *J. Chem. Phys.*, 1994, **101**, 3671.
- 55 D. H. Zhang, S.-Y. Lee and M. Baer, *J. Chem. Phys.*, 2000, **112**, 9802.
- 56 D. R. Yarkony, *J. Phys. Chem. A*, 2004, **108**, 3200.
- 57 D. R. Yarkony, *Faraday Discuss.*, 2004, **127**, 325.
- 58 X. Zhu and D. R. Yarkony, *Mol. Phys.*, 2010, **108**, 2611.